Pacific Journal of Mathematics

IN THIS ISSUE—

William F. Donoghue, Jr., *The lattice of invariant subspaces of a completely continuous quasi-nilpotent transformation* . . . 1031
Michael (Mihály) Fekete and J. L. Walsh, *Asymptotic behavior of restricted extremal polynomials and of their zeros* . . . 1037
Shaul Foguel, *Biorthogonal systems in Banach spaces* . . . 1065
David Gale, *A theorem on flows in networks* . . . 1073
Ioan M. James, *On spaces with a multiplication* . . . 1083
Richard Vincent Kadison and Isadore Manual Singer, *Three test problems in operator theory* . . . 1101
Maurice Kennedy, *A convergence theorem for a certain class of Markoff processes* . . . 1107
G. Kurepa, *On a new reciprocity, distribution and duality law* . . . 1125
Richard Kenneth Lashof, *Lie algebras of locally compact groups* . . . 1145
Calvin T. Long, *Note on normal numbers* . . . 1163
M. Mikolás, *On certain sums generating the Dedekind sums and their reciprocity laws* . . . 1167
Barrett O'Neill, *Induced homology homomorphisms for set-valued maps* . . . 1179
Mary Ellen Rudin, *A topological characterization of sets of real numbers* . . . 1185
M. Schiffer, *The Fredholm eigen values of plane domains* . . . 1187
F. A. Valentine, *A three point convexity property* . . . 1227
Alexander Doniphan Wallace, *The center of a compact lattice is totally disconnected* . . . 1237
Alexander Doniphan Wallace, *Two theorems on topological lattices* . . . 1239
G. T. Whyburn, *Dimension and non-density preservation of mappings* . . . 1243
John Hunter Williamson, *On the functional representation of certain algebraic systems* . . . 1251

Vol. 7, No. 2 **February, 1957**

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 of the editors. All other communications to the editors should be addressed to the managing editor, E. G. Straus 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. 10, 1-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

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

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

THE LATTICE OF INVARIANT SUBSPACES OF A COMPLETELY CONTINUOUS QUASI-NILPOTENT TRANSFORMATION

W. F. DONOGHUE, JR.

An essential result in the study of a continuous linear transformation of a Banach space into itself is the specification of the lattice of proper closed subspaces of the Banach space which are invariant under the transformation. For certain classes of transformations the results which have been obtained in this direction may be regarded as complete, for example, for self-adjoint transformations in Hilbert space. The invariant subspaces for certain isometries in Hilbert space have been found by Beurling [2] whose results have been extended to unitary transformations by the author [3]. In general, however, little is known; in fact, it is not yet known that an arbitrary continuous linear transformation in Hilbert space has nontrivial closed invariant subspaces. A theorem of von Neumann guarantees that a completely continuous transformation in Hilbert space has such subspaces, while more recent work of Aronszajn and Smith [1] establishes the same result for any Banach space. For completely continuous transformations which contain only the point 0 in the spectrum (the quasi-nilpotent transformations), spectral theory can provide no information concerning the invariant subspaces, and the application of the result of Aronszajn and Smith only assures the existence of a nested sequence of closed invariant subspaces. Such a lattice of invariant subspaces is considerably simpler in structure than that usually encountered in spectral theory. It is the purpose of this note to show that more cannot be obtained, and that this very simple lattice does in fact occur. The three examples which follow illustrate this fact; the fourth example shows that not every completely continuous quasi-nilpotent transformation has such a lattice of invariant subspaces.

EXAMPLE 1. Let $\mathscr{H}$ be the Hilbert space consisting of all functions

$$f(z)=\sum_{n=0}^{\infty} a_n z^n$$

analytic in $|z|<1$ with Taylor coefficients in l^2:

$$\sum_{n=0}^{\infty} |a_n|^2=\|f\|^2<\infty .$$

Received June 25, 1956. Paper written under contract with Office of Naval Research, Contract Nonr 58304.

We may also write

$$(f,\ g)=\sum_{n=0}^{\infty} a_n\overline{b}_n=\frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})\overline{g(e^{i\theta})}d\theta$$

where

$$g(z)=\sum_{n=0}^{\infty} b_n z^n .$$

Let V be the transformation defined by $Vf(z)=zf(z/2)$; V is completely continuous and quasi-nilpotent. Let $\mathscr{M}_n$ be the subspace of $\mathscr{H}$ composed of functions $f(z)$ which have a zero of order $\geqq n$ at the origin; it is evident that these subspaces satisfy the relation $\mathscr{M}_{n+1}\subset\mathscr{M}_n$ and that they are closed and invariant under V. It will be shown that these are the only nontrivial closed invariant subspaces.

For the proof, it is enough to consider an element f of $\mathscr{H}$ for which $f(0)\neq 0$ and to show that the sequence $V^nf\ (n\geqq 0)$ spans $\mathscr{H}$. We may suppose $f(0)=1$ and write $f(z)=1+g(z)$ with g in $\mathscr{M}_1$. For $n\geqq 0$ we define $h_n(z)=\sqrt{2^{(n^2-n)}}V^nf$; it is sufficient to show that the sequence h_n spans the space. For this purpose, consider the linear transformation T defined by $T(z^n)=h_n(z)$; it is easy to establish the continuity of T, and it will now be shown that T has a continuous inverse, thereby establishing the completeness of h_n. Note first that

$$\|(T-I)z^n\|^2=\|h_n(z)-z^n\|^2=\|f(z/2^n)-1\|^2=\|g(z/2^n)\|^2\leqq\frac{\|g\|^2}{4^n} .$$

On the other hand, if $f_\nu(z)$ is a sequence in $\mathscr{H}$ weakly converging to 0 then

$$f_\nu(z)=\sum_{n=0}^{\infty} a_n^{(\nu)}z^n$$

and

$$\|(T-I)f_\nu\|\leqq\sum_{n=0}^{\infty}|a_n^{(\nu)}|\ \|h_n-z^n\|\leqq\|g\|\sum_{n=0}^{\infty}|a_n^{(\nu)}|(1/2^n);$$

(the inequality above following in general from its truth for finite sums). From the weak convergence of the f_ν to 0 it is clear that

$$\lim_\nu\sum_{n=0}^{\infty}|a_n^{(\nu)}|(1/2^n)=0 ,$$

hence that $(T-I)f_\nu$ converges strongly to 0 and therefore that $T-I$ is completely continuous. Accordingly 0 is in the resolvent set of T or there exists a finite-dimensional null space for T. Thus T^{-1} exists and is continuous, since the nonexistence of a null space for T is shown as follows: $Tf=0$ for

$$f(z)=\sum_{n=0}^{\infty} a_n z^n$$

means

$$\sum_{n=0}^{\infty} a_n h_n(z)=0$$

whence

$$a_0 h_0(z)+\sum_{n=1}^{\infty} a_n h_n(z)=0\ ;$$

the second term is in $\mathscr{M}_1$, hence the first term is, and since h_0 is not in that subspace it follows that $a_0=0$; from an inductive argument it then follows that all $a_n=0$, as desired.

EXAMPLE 2. Let V^* be the adjoint of the transformation V described above. The invariant subspaces for V^* are the orthogonal complements of the invariant subspaces for V, and form an increasing sequence of finite-dimensional subspaces. Clearly V^* is completely continuous and quasi-nilpotent.

EXAMPLE 3. If $f(t)$ is a function integrable on the interval $0 \leqq t \leqq 1$ let Sf be the indefinite integral

$$Sf(t)=\int_0^t f(x)dx\ .$$

The operator S is the Volterra integral operator and is completely continuous and quasi-nilpotent when considered as a transformation from L^p into itself for $1 \leqq p < \infty$, or when considered on the space $\mathscr{C}$ of continuous functions on the interval. If $\mathscr{M}_s$ denotes the class of all functions in L^p which vanish almost everywhere on the interval $0 \leqq t \leqq s$ it is clear that the $\mathscr{M}_s$ form closed subspaces of L^p which are invariant under S. It will be shown that such subspaces are the only closed invariant subspaces of the operator S.

For the proof, the result is first established for the continuous function space. Let $f(t)$ be continuous on the interval $0 \leqq t \leqq 1$ and define $f=0$ outside that interval to obtain a function defined throughout the axis. If $Y(t)$ is the Heaviside function equal to 1 on the positive half-axis and vanishing on the negative half-axis, then for all x in the unit interval

$$Sf(x)=(Y*f)(x)$$

where $*$ denotes convolution. Similarly, the iterates of S applied to f are given for $0 \leq x \leq 1$ by

$$S^n f(x) = (Y_n * f)(x)$$

where $Y_{n+1}(t)$ vanishes on the left half-axis and equals $t^n/n!$ on the right. Let μ be a measure on the unit interval orthogonal to all $S^n f$; the equation

$$\int_0^1 S^n f(x)\, d\mu = 0$$

may be written $(Y_n * f * \check{\mu})(0) = 0$ where $\check{\mu}$, the reflection of μ through the origin, is given by $\check{\mu}(t) = \mu(-t)$. From the associativity of convolution, then, $(Y_n * (f * \check{\mu}))(0) = 0$ whence it follows that the continuous function $f * \check{\mu}$ is orthogonal to $\check{Y}_n(t) = Y_n(-t)$ for all $n \geq 0$. Accordingly $f * \check{\mu}$ vanishes on the left half-axis. A theorem, the most general version of which is due to J. Lions [4] asserts that for any two distributions on R^n with compact support, the convex hull of the support of the convolution is the vectorial sum of the convex hulls of the supports of the factors. Thus if the convex hull of the support of μ is the interval (c, d) and the convex hull of the support of f is (a, b) it follows that the interval $(a-d, b-c)$ is the convex hull of the support of $f * \check{\mu}$, whence $d \leq a$. Thus the only measures orthogonal to $S^n f$, $n \geq 0$ are measures orthogonal to the subspace $\mathscr{M}_a$ and the closed linear span of that sequence is $\mathscr{M}_a$, unless $a = 0$, in which case the closed linear span will be the whole space if $f(0) \neq 0$. Thus any proper invariant subspace for S in $\mathscr{C}$ is a union of spaces of type $\mathscr{M}_s$ and is therefore a space of that type itself.

For the spaces L^p the same result follows from the observation that the range of S is contained in $\mathscr{C}$. If the smallest interval containing the support of Sf is (a, b), then the sequence $S^n f$, $n \geq 1$ spans the subspace $\mathscr{M}_a$ of $\mathscr{C}$ and its closure in L^p is the corresponding $\mathscr{M}_a$ of that space. Evidently $f(t) = 0$ almost everywhere in $0 \leq t \leq a$, whence $S^n f$, $n \geq 0$ spans $\mathscr{M}_a$ in L^p.

It is of interest to note that our assertion is no longer true for the space L^∞. As above, the transformation S is completely continuous and quasi-nilpotent, however, its range is contained in a separable subspace of the nonseparable L^∞. It is possible to obtain any closed invariant subspace of L^∞ by choosing any closed subspace of a subspace of the type $\mathscr{M}_a$ which contains the corresponding continuous function subspace $\mathscr{M}_a$. Nevertheless, if the word closed were interpreted to mean weak-star closed the result obtained above for $\mathscr{C}$ and L^p would carry over to L^∞.

EXAMPLE 4. Let $\mathscr{H}$ be the separable Hilbert space consisting of functions $f(x, y)$ defined in the unit square for which

$$\|f\|^2=\int_0^1\int_0^1 |f(t, s)|^2\, ds\, dt < \infty .$$

On $\mathscr{H}$ consider the integral operator T defined by

$$Tf(x, y)=\int_0^x\int_0^y f(t, s)\, ds\, dt .$$

T is completely continuous and quasi-nilpotent. Following the methods of the previous example it is not difficult to construct a class of subsets of the square corresponding to a class of invariant subspaces. Obviously the linear subspaces so obtained are not linearly ordered under inclusion; moreover these are not all of the invariant subspaces since the subspace consisting of functions which depend only on the product xy is also a closed invariant subspace for T.

REFERENCES

1. N. Aronszajn and K. T. Smith, *Invariant subspaces of completely continuous operators*, Ann. of Math., **60** (1954), 345–350.

2. A. Beurling, *On two problems concerning linear transformations in Hilbert space*. Acta Math., (1949), 239–255.

3. W. F. Donoghue, *Invariant subspaces of unitary transformations*. Bull. Amer. Math. Soc., **59** (1953), 535.

4. J. L. Lions, *Supports dans la transformation de Laplace*. Journ. d'Analyse Math., **2** (1952), 369–379.

UNIVERSITY OF KANSAS

ASYMPTOTIC BEHAVIOR OF RESTRICTED EXTREMAL POLYNOMIALS AND OF THEIR ZEROS

M. FEKETE AND J. L. WALSH

Introduction. Progress in the study of polynomials has recently been made in two directions: (i) asymptotic properties of sequences of polynomials of least norm on a given set (Leja, [7]; Davis and Pollak, [1]; Fekete, [3]; Walsh and Evans, [10]; Fekete and Walsh, [5]); (ii) geometry of the zeros of polynomials of prescribed degree minimizing a given norm on a given set, where one or more coefficients are preassigned (Zedek, [12]; Fekete, [4]; Walsh and Zedek, [11]; Fekete and Walsh, [6]). The object of the present paper is to combine these two trends, by studying the asymptotic properties of sequences of polynomials of least norm on a given set, where the polynomials are restricted by prescription of one or more coefficients.

If S is a given compact point set and $N[A_n(z), S]$ any norm on S of the polynomial $A_n(z) \equiv z^n + a_{1n}z^{n-1} + \cdots + a_{nn}$ we are interested in the asymptotic relations for (restricted) polynomials $A_n(z, N)$ of least N-norm

$$(1) \qquad \lim_{n\to\infty} \nu_n^{1/n} = \tau(S) , \qquad \nu_n = N[A_n(z, N), S] ,$$

$$(2) \qquad \lim_{n\to\infty} |A_n(z, N)|^{1/n} = |\varphi(z)| ,$$

where $\tau(S)$ is the transfinite diameter of S, $|\varphi(z)| \equiv e^{G(z)}\tau(S)$, $G(z)$ being Green's function with pole at infinity for the maximal infinite region K containing no point of S, and where (2) is considered uniformly on a more or less arbitrary compact set in K.

Part I is devoted primarily to (1); we show for instance that for the unit circle, with the first $k \equiv k(n)$ coefficients a_{jn} of the extremal polynomial $A_n(z, N)$ prescribed and uniformly of the order $O\left(\binom{n}{j}\right)$ in their totality, a necessary and sufficient condition for (1) for all such choices of coefficients is $k = o(n)$, where N is any classical norm. We prove similar results for other sets S. Part II is devoted primarily to (2); first we use as hypothesis the analogue of (1), namely

$$\{N[A_n(z), S]\}^{1/n} \to \tau(S) ,$$

for arbitrary polynomials $A_n(z)$; and then we use (1) as hypothesis, for extremal polynomials $A_n(z, N)$ with k prescribed coefficients and N

Received July 17, 1956. This work was done at Harvard University under contract with the Office of Naval Research.

M. Fekete died on May 13, 1957.

monotonic. If $A_n(z, N)$ has zeros in K, under suitable conditions the corresponding factors of $A_n(z, N)$ can be omitted in whole or in part, and the analogue of (2) is valid for the remaining factor, uniformly on any closed set in K containing no limit point of zeros of that factor; for instance if k is constant we can omit the factors of $A_n(z, N)$ corresponding to the zeros of $A_n(z, N)$ exterior to the inflated convex hull $H_k(S)$, and (2) is valid uniformly on any compact set exterior to $H_k(S)$; as another instance, if $k=1$ and if the prescribed center of gravity of the zeros of $A_n(z, N)$ is fixed and different from the conformal center of gravity of S, then precisely one zero of $A_n(z, N)$ becomes infinite and (2) is valid uniformly on any compact set exterior to the convex hull H_0 of S. Finally, we study (1) for extremal polynomials some of whose zeros are prescribed.

PART I

ASYMPTOTIC PROPERTIES OF THE LEAST N-NORM OF RESTRICTED POLYNOMIALS ON A GIVEN POINT SET

1. **In pursuing the objective indicated,** we start our considerations with remarks relevant to both (1) and (2). Let $A_n = A_n(\gamma_j, 1 \leqq j \leqq k)$ denote the aggregate of all polynomials $A_n(z) \equiv z^n + a_{1n}z^{n-1} + a_{2n}z^{n-2} + \cdots + a_{nn}$ satisfying

$$(3) \qquad a_{jn} = \gamma_j , \qquad 1 \leqq j \leqq k, \ 1 \leqq k \leqq n-1 .$$

The reader may easily prove the existence for each $n (\geqq n^* = n^*(N))$ and for each $\gamma_j = \gamma_j(n)$ and $k = k(n)$ of a polynomial $A_n(z, N)$ in $A_n(\gamma_j, 1 \leqq j \leqq k)$ of least N-norm, provided N belongs to the wide category of quasi-Tchebycheff (q.T.) norms continuous in A_n on S; such norms are broad generalizations of the classical norms, including the (ordinary) Tchebycheff norm

$$M = M[A_n(z), S] = [\max |A_n(z)|, z \text{ on } S] .$$

We recall [5, p. 53] that $N[A_n(z), S]$ is a q.T. norm on S provided for all

$$A_n(z) \equiv z^n + a_{n1}z^{n-1} + \cdots$$

we have

$$(4) \qquad \frac{N[A_n(z), S]}{M[A_n(z), S]} \leqq U(S, N) , \qquad n = 1, 2, \cdots ,$$

$$(5) \qquad \frac{N[A_n(z), S]}{M[A_n(z), S]} \geqq L_n(S, N, \varepsilon) > 0 \qquad \text{for } n = n_0(\varepsilon) ,$$

$$\lim_{n\to\infty} \{L_n(S, N, \varepsilon)\}^{1/n} = f(N, \varepsilon), \qquad \lim_{\varepsilon\to 0} f(N, \varepsilon) = 1 .$$

A norm $N[A_n(z), S]$ is *continuous* in A_n on S provided to an arbitrary $A_n^*(z) \in A_n$ and $\varepsilon\,(>0)$ there corresponds a $\delta = \delta(\varepsilon, A_n^*)\ (>0)$ such that for an arbitrary polynomial $A_n^{**}(z)$ in A_n the inequality $|A_n^*(z) - A_n^{**}(z)| < \delta$ on S implies

$$|N[A_n^*(z), S] - N[A_n^{**}(z), S]| < \varepsilon .$$

Such continuity of $N[A_n(z), S]$ on a certain subset of A_n is also necessary for the existence of a polynomial $A_n(z, N)$ in A_n of least N-norm, since there exist instances of noncontinuous q.T.-norms N for which

$$N[A_n(z), S] > [\inf N[A_n(z), S],\ A_n(z) \in A_n]$$

holds for all $A_n(z) \in A_n$. For the purposes of (1) if N is not continuous one may replace $\min N$ in our considerations by $\inf N$. Henceforth we consider only continuous q.T. norms $N[A_n(z), S]$.

2. The writers have already proved [5, Theorem 2] that the relation

$$\lim_{n\to\infty} \{N_0[A_n(z), S]\}^{1/n} = \tau(S) \tag{6}$$

for an arbitrary compact set S, a given q.T.-norm N_0, and an arbitrary sequence of polynomials $A_n(z) \equiv z^n + \cdots$, implies the relation

$$\lim_{n\to\infty} \{N[A_n(z), S]\}^{1/n} = \tau(S) \tag{7}$$

for any other q.T.-norm on S. There follows

THEOREM 1. *Equation* (1) *holds for every choice of q.T. norm N if and only if* (1) *holds for a particular choice of N, where all polynomials are restricted to A_n.*

From (6) with $A_n(z)$ the polynomials $A_n(z, N_0)$ we deduce (7) involving these same polynomials, and this (7) as a majorant relation proves (1); we use here the consequence of (5) that no matter what the polynomials $A_n(z)$ may be, the first member of (7) is not less than $\tau(S)$. Conversely, if (6) is not valid for a particular N_0 and the polynomials $A_n(z, N_0)$, then (7) is not valid for either the $A_n(z, N_0)$ or the $A_n(z, N)$, so (1) is not valid.

The importance of Theorem 1 for our investigation of (1) is that in the sequel we may instead investigate (6) with $A_n(z) \equiv A_n(z, N_0)$ for a particular N_0 conveniently chosen with respect to S.

3. As a first such application of Theorem 1 we prove

THEOREM 2. *Let $A_n(\gamma_1, \gamma_2, \cdots, \gamma_k)$ be given, and S the unit disc*

$|z| \leqq 1$. *A necessary and sufficient condition for* (1) *with* $A_n(z, N) \in A_n$ *is*

$$\lim_{n\to\infty} \{1+|\gamma_1|^2+|\gamma_2|^2+\cdots+|\gamma_k|^2\}^{1/2n}=1 \ , \tag{8}$$

independently of the q.T.*-norm* N.

In fact, with $A_n(z) \in A_n$ and $n \geqq k+1$, the choice

$$\begin{aligned}\{N_0[A_n(z), S]\}^2 &= \frac{1}{2\pi}\int_{|z|=1} |A_n(z)|^2 |dz| \\ &= 1+|\gamma_1|^2+\cdots+|\gamma_k|^2+|a_{k+1,n}|^2+\cdots\end{aligned} \tag{9}$$

leads to the unique minimizing polynomial

$$A_n(z, N_0) \equiv z^n+\gamma_1 z^{n-1}+\cdots+\gamma_k z^{n-k} \ ,$$

$$\nu_n = N_0[A_n(z, N_0), S] = \{1+|\gamma_1|^2+\cdots+|\gamma_k|^2\}^{1/2} \ . \tag{10}$$

Since $\tau(S)=1$, (1) with N_0 for N is equivalent to (8). To complete the proof we recall Theorem 1.

4. A noteworthy corollary of Theorem 1 is

THEOREM 3. *With the notation and hypothesis of Theorem* 2, *suppose we have*

$$\gamma_j = O\left[\binom{n}{j}\right], \qquad 1 \leqq j \leqq k \ , \tag{11}$$

uniformly in j. *Then a necessary and sufficient condition for* (1) *with arbitrary* q.T.*-norm* N *is*

$$k = o(n) \ . \tag{12}$$

With N_0 of (9) for N, hypothesis (11) in case (12) entails in view of (10)

$$1 \leqq \nu_n = O\left[n\binom{n}{j}\right] \qquad \text{as } n\to\infty \ , \tag{13}$$

since $\binom{n}{j}$ increases with j provided $2j<n$. It is sufficient to prove (1) for every sequence of values $n\to\infty$, so it is sufficient to prove (1) under the alternate assumptions $k(n)\equiv O(1)$ and $k(n)\to\infty$. Relations (13) prove (1) if $k=O(1)$. If however $k=k(n)\to\infty$ still subject to (12), by Stirling's formula

$$\binom{n}{k} \sim \left[\left(\frac{n}{k}-1\right)^k \Big/ \left(1-\frac{k}{n}\right)^n\right]\sqrt{1/2\pi k} \ , \tag{14}$$

$$\binom{n}{k}^{1/n} \sim \left(\frac{n}{k}-1\right)^{k/n} \to 1 ,$$

implying (1) with $N=N_0$, and hence for all q.T.-norms N.

To prove the necessity of (12) for (1) with $N\equiv N_0$ (and hence with an arbitrary q.T.-norm), choose

$$\gamma_j = \binom{n}{j} , \qquad 1 \leqq j \leqq k .$$

Then by (10)

$$\nu_n > \binom{n}{j} , \qquad 1 \leqq j \leqq k = k(n) .$$

If (12) is false, for a suitably chosen sequence we have

$$\lim k(n)/n = \varepsilon , \qquad 0 < \varepsilon \leqq 1 ;$$

in case $0<\varepsilon<1$, by (14) follows

$$\nu_n^{1/n} > \binom{n}{k}^{1/n} \to \frac{1}{(1-\varepsilon)^{1-\varepsilon}\varepsilon^{\varepsilon}} \to 1 ,$$

while in case $\varepsilon=1$ we have

$$\nu_n^{1/n} > \binom{n}{\left[\frac{n}{2}\right]}^{1/n} \to 2 \qquad \text{as } n \to \infty .$$

This contradiction of (1) completes the proof of Theorem 2.

5. **By modifying slightly the above argument,** the reader may easily prove the following proposition, a generalization for S the disc $|z| \leqq R$ of the previous two theorems.

THEOREM 4. *A necessary and sufficient condition for* (1) *with* $S: |z| \leqq R$ *and* $A_n(z, N) \in A_n(\gamma_1, \cdots, \gamma_k)$ *is*

$$\lim \{R^{2n} + |\gamma_1|^2 R^{2n-2} + \cdots + |\gamma_k|^2 R^{2n-2k}\}^{1/2n} = R . \tag{15}$$

With the particular choice (uniformly in j*)*

$$\gamma_j = O\left[R^j \binom{n}{j}\right] , \qquad 1 \leqq j \leqq k , \tag{16}$$

a necessary and sufficient condition for (1) *is* (12).

A word is in order to justify the form of (16). Much of the present paper is devoted to the study of polynomials $A_n(z)$ in $A_n(\gamma_1, \gamma_2, \cdots, \gamma_k)$

where we have

$$\gamma_j = (-1)^j \binom{n}{j} c_j \tag{17}$$

and the numbers c_j are independent of n. For instance the center of gravity of the zeros of $A_n(z)$ is c_1, and (17) may prescribe c_1 independent of n. Here it is significant (Theorem 11, below) that a necessary condition for (1) with the zeros of the $A_n(z)$ bounded is (17) with $c_1 \to c$, where of course c_1 is not necessarily independent of n, and where c is the conformal center of gravity of S, a number depending wholly on S itself.

We shall call the number c_j defined by (17) *the centroid of order* j of the zeros of $A_n(z)$.

Another comment on (16) is that if the z-plane is transformed by a simple stretching $z' = Rz$, the transfinite diameter of every set is multiplied by R, and the jth centroid of the zeros of a polynomial is multiplied by R^j; thus the factor R^j in (16) is appropriate.

6. **We shall shortly indicate** (§§ 7, 8) that Theorems 2, 3, and 4 admit at least partial extensions to arbitrary sets whose boundaries are rectifiable. The usefulness of these extensions in the study of still more general point sets is now to be shown.

If S is an arbitrary compact set, and if the maximal infinite region K belonging to the complement of S is regular in the sense that the classical Green's function $G(z)$ for K with pole at infinity exists, we denote by

$$w = \varphi(z) \equiv \exp\,[G(z) + iH(z) + \log \tau(S)]\ ,$$

where $H(z)$ is conjugate to $G(z)$ in K, a function which maps K onto $|w| > \tau(S)$ with $\varphi(\infty) = \infty$.

The locus C_R: $|\varphi(z)| = R\tau(S)$, $R > 1$, in K consists of a finite number of rectifiable Jordan curves which are mutually exterior except perhaps for a finite number of points each of which may belong to several curves; we denote the sum of the closed interiors of these curves by S_R. As $R \to 1$, the locus C_R approaches the boundary of K.

THEOREM 5. *Let S be a compact set, and let the infinite region K belonging to the complement of S be regular. Let $A_n(\gamma_1, \gamma_2, \cdots, \gamma_k)$ be given and restrict $A_n(z, N)$ to A_n. A necessary and sufficient condition that* (1) *be valid for all* q.T.-*norms on S is that* (1) *be valid for all* q.T.-*norms on all S_R.*

By Theorem 1, we may restrict ourselves to the consideration of the Tchebycheff norms on S and S_R. We denote the respective extremal polynomials by $T_n(z, S)$ and $T_n(z, S_R)$. To prove the sufficiency of the condition, we write

$$M[T_n(z, S_R), S_R] \geqq M[T_n(z, S_R), S] \geqq M[T_n(z, S), S] \geqq \tau(S)^n ,$$

$$\begin{aligned}&\lim \{M[T_n(z, S_R), S_R]\}^{1/n} = \tau(S_R)\\ &\qquad = R\cdot\tau(S) \geqq \limsup \{M[T_n(z, S), S]\}^{1/n} \geqq \tau(S) ,\end{aligned}$$

and $R \to 1$ establishes (1). With this reasoning as given, *it is also sufficient if* (1) *is valid on a sequence of sets* S_m *each containing* S, *with* K *regular or not, provided* $\tau(S_m) \to \tau(S)$; *the sets* S_m *may be taken as the closed interiors of lemniscates.*

Conversely, by use of the generalized Bernstein Lemma [9, p. 77] we have

$$\{\tau(S_R)\}^n \leqq M[T_n(z, S_R), S_R] \leqq M[T_n(z, S), S_R] \leqq M[T_n(z, S), S]R^n ,$$

$$\begin{aligned}&\tau(S_R) \leqq \liminf \{M[T_n(z, S_R), S_R]\}^{1/n} \leqq \limsup \{M[T_n(z, S_R), S_R]\}^{1/n}\\ &\qquad \leqq \lim \{M[T_n(z, S), S]\}^{1/n}\cdot R = \tau(S_R) ,\end{aligned}$$

$$\lim \{M[T_n(z, S_R), S_R]\}^{1/n} = \tau(S_R) .$$

7. Theorem 5 emphasizes the importance in considering (1) of sets with rectifiable boundary, both for their own sake and for the study of more general sets. For the former we have the great advantage of orthogonal polynomials as a tool. Thus we prove[2]

THEOREM 6. *Let the point set* S *consist of a finite number of rectifiable Jordan arcs, and let the polynomials* $P_n(z) \equiv z^n + \cdots$ *of respective degrees* n *be mutually orthogonal on* S, *with*

$$\int_S |P_n(z)|^2 |dz| = p_n .$$

Let the norm N_0 *be defined by*

$$\{N_0[A_n(z)]\}^2 = \int_S |A_n(z)|^2 |dz| ,$$

and let us set

$$B_n(z) \equiv z^n + \gamma_1 z^{n-1} + \cdots + \gamma_k z^{n-k} \equiv d_0 P_n(z) + d_1 P_{n-1}(z) + \cdots + d_n P_0(z) ,$$

$$p_h d_h = \int_S B_n(z)\overline{P_h(z)}|dz| , \qquad d_h = d_h(n),\ 0 \leqq h \leqq n ,$$

[2] An analogous theorem obviously exists if S consists of a finite number of mutually disjoint Jordan regions, and least-square norm is measured by surface instead of line integrals.

where the $\gamma_j = \gamma_j(n)$ *and* $k = k(n)$ *are prescribed. For each* $n \geqq k+1$, $\min N_0[A_n(z)]$ *with* $A_n(z) \in A_n(\gamma_j)$ *is* $N_0[C_n(z)]$ *where*

$$C_n(z) \equiv d_0 P_n(z) + d_1 P_{n-1}(z) + \cdots + d_k P_{n-k}(z) ,$$

and is assumed by no other polynomial. A necessary and sufficient condition for (1) *with* $N = N_0$ *or with* N *an arbitrary* q.T.-*norm on* S *is*

$$\lim_{n \to \infty} \{|d_0|^2 p_n + |d_1|^2 p_{n-1} + \cdots + |d_k|^2 p_{n-k}\}^{1/2n} = \tau(S) .$$

An arbitrary polynomial $A_n(z) \equiv z^n + \cdots$ may obviously be expressed as a linear combination $A_n(z) \equiv b_0 P_n(z) + b_1 P_{n-1}(z) + \cdots + b_n P_0(z)$ by considering successively the coefficients of z^n, z^{n-1}, $\cdots$, 1. Then the coefficients b_h can also be computed by use of the orthogonality relations,

$$\int_S A_n(z) \overline{P_h(z)} \, |dz| = p_h b_{n-h} ,$$

and we have

$$\{N_0[A_n(z)]\}^2 = |b_0|^2 p_n + |b_1|^2 p_{n-1} + \cdots + |b_n|^2 p_0 .$$

The condition $A_n(z) \in A_n(\gamma_j)$ is equivalent to specific prescription of $b_0, b_1, \cdots, b_k$, so it is clear that

$$\min N_0[A_n(z)] = N_0[C_n(z)] = \{|d_0|^2 p_n + |d_1|^2 p_{n-1} + \cdots + |d_k|^2 p_{n-k}\}^{1/2} ,$$

a minimum assumed by no other polynomial than $C_n(z)$ in A_n; the remainder of Theorem 6 follows from Theorem 1.

Both Theorem 2 and the first part of Theorem 4 are clearly generalized in Theorem 6. We proceed to a corresponding generalization of the necessity of condition (16) in the second part of Theorem 4.

8. **The number R plays two roles in** Theorem 4: it is both $\tau(S)$ and a parameter restricting the order of γ_j in (16) if $k = k(n)$ is not bounded.

In extending the second part of Theorem 4 to a compact set S of connected complement K whose boundary B consists of a finite number of rectifiable Jordan arcs or even to a more general set S with regular connected complement K, the second of these roles is kept for R. To be more explicit we shall prove the following.

THEOREM 7. *Let* S *be a compact set of connected regular complement* K *and* R *an arbitrary positive number such that the disc* $|z| \leqq R$ *contains* S *in its interior.*

Suppose that (16) *holds with* $k = k(n) \to \infty$, $k(n) \neq o(n)$. *Then there exist polynomials* $A_n(z, N) \in A_n(\gamma_1, \gamma_2, \cdots, \gamma_k)$ *of least* q.T.-*norm* $N = N[A_n(z), S]$ *on* S *for which* (1) *is not valid.*

We chose $R_1=R_1(R)>1$ so that not only S but also S_{R_1} of § 6 is covered by $|z|<R$ (thus $\tau(S_{R_1})<R$). We know (Cf. § 6) that with $T_n(z, S)\in A_n$, $T_n(z, S_{R_1})\in A_n$,

$$R_1^n M[T_n(z, S), S]\geqq M[T_n(z, S), S_{R_1}]\geqq M[T_n(z, S_{R_1}), S_{R_1}]\ . \tag{18}$$

On the other hand if C_{R_1} denotes the boundary of S_{R_1},

$$\begin{aligned}\{M[T_n(z, S_{R_1}), S_{R_1}]\}^2\int_{C_{R_1}}|dz| &\geqq \int_{C_{R_1}}|T_n(z, S_{R_1})|^2|dz| \\ &\geqq \int_{C_{R_1}}|A_n(z, N_0)|^2|dz|]\ ,\end{aligned} \tag{19}$$

where $A_n(z, N_0)\in A_n$ is of least square norm N_0 on S_{R_1},

$$\{N_0[A_n(z), S_{R_1}]\}^2=\int_{C_{R_1}}|A_n(z)|^2|dz|\ .$$

Fix $\gamma_j=\gamma_j(n)$, $1\leqq j\leqq k=k(n)$, by

$$\begin{aligned}A_n(z)\equiv P_n(z)+R^h\binom{n}{k}P_{n-h}(z)\Big/\binom{n-h}{h} &+ R^k\binom{n}{k}P_{n-k}(z) \\ +\lambda_1 P_{n-k-1}(z)+\cdots &+\lambda_{n-k}P_0(z)\end{aligned}$$

where $h=\left[\dfrac{k}{2}\right]$, and the $P_m(z)=z^m+\cdots$ are mutually orthogonal on C_{R_1}. All zeros of the $P_n(z)$ lie in $|z|\leqq R$ (Fejér, [2]) so (16) is satisfied; indeed $\binom{n}{k}\binom{n-h}{j}=\binom{n-h}{h}\binom{n}{n-h-j}$ for $1\leqq j\leqq h$. For $A_n(z)\in A_n(\gamma_1, \cdots, \gamma_k)$ so chosen we have

$$A_n(z, N_0)=P_n(z)+R^h\binom{n}{k}P_{n-h}(z)\Big/\binom{n-h}{h}+R^k\binom{n}{k}P_{n-k}(z)\ ,$$

$$\{N_0[A_n(z, N_0), S_{R_1}]\}^2=p_n+R^{2h}\binom{n}{k}^2 p_{n-k}\Big/\binom{n-h}{h}^2+R^{2k}\binom{n}{k}^2 p_{n-k}$$

where

$$p_j=\int_{C_{R_1}}|P_j(z)|^2|dz|\ , \qquad\qquad j=0, 1, 2, \cdots\ .$$

Hence

$$N_0[A_n(z, S_{R_1}), S_{R_1}]>\max\left\{R^h\binom{n}{k}p_{n-h}^{1/2}\Big/\binom{n-h}{h},\ R^k\binom{n}{k}p_{n-k}^{1/2}\right\}\ .$$

By Fekete-Walsh [5]

$$p_n^{1/2n} \to \tau(S_{R_1}) \qquad \text{as } n \to \infty .$$

Consider any sequence of n for which $\lim \frac{k}{n} = \varepsilon$ exists with $0 < \varepsilon \leqq 1$. In case $0 < \varepsilon < 1$, (Cf. § 4)

$$\text{(20)} \qquad \liminf \{N_0 [A_n(z, S_{R_1}), S_{R_1}]\}^{1/n} \geqq R^\varepsilon \tau(S_{R_1})^{1-\varepsilon}(1-\varepsilon)^{\varepsilon-1}\varepsilon^{-\varepsilon}$$
$$> R_1\tau(S)(1-\varepsilon)^{\varepsilon-1}\varepsilon^{-\varepsilon};$$

for we have $R > \tau(S_{R_1}) = R_1\tau(S)$.

In case $\varepsilon = 1$, by $\binom{n}{k}\Big/\binom{n-h}{h} \to 1$ as $n \to \infty$, we have

$$\text{(21)} \qquad \liminf \{N_0[A_n(z, N_0), S_{R_1}]\}^{1/n} \geqq R^{1/2}\tau(S_{R_1})^{1/2}\cdot 2^{1/2} \geqq R_1\tau(S)\cdot 2^{1/2} .$$

Combining (20) or (21) with (18) and (19) we obtain

$$\limsup \{M[T_n(z, S), S]\}^{1/n} > \tau(S)\cdot((1-\varepsilon)^{-1}\varepsilon^{-\varepsilon} \text{ or } 1) \geqq \tau(S) .$$

Thus the proof is complete for the classical Tchebycheff norm M, and the theorem follows by Theorem 1.

Theorem 7 can be extended to arbitrary compact sets S of positive transfinite diameter $\tau(S)$ with connected nonregular complement K, the role of R being taken by any positive number such that the disc $|z| < R$ contains a level locus C_{R_1}: $G(z) = \log R_1$ which consists of finitely many Jordan contours and separates S from infinity. The proof is similar to the above one but uses the generalized Bernstein lemma in its extended form. (Walsh, [9], § 4.9). Corresponding extensions to the case of K nonregular can be made for Theorem 8^{bis} and the second part of Theorem 9 below (concerning the respective necessary conditions for the validity of (1)).

9. We have studied in some detail the conditions (12) and (16) singly and in combination, especially if S is a circular disc, and in particular have shown in Theorem 7 for a more general set S that (12) is necessary[3] for (1) provided (16) is assumed with the choice $R > [\max |z|, z \text{ on } S]$. We are not in a position to prove that conversely

[3] Nevertheless $k = k(n) = O(n)$, more precisely $k(n) = n-1$, is compatible with the validity of (1) for k-fold restricted nth degree extremal polynomials $A_n(z, N) \equiv z^n + a_{1n}z^{n-1} + \cdots + a_{n-1,n}z + a_{nn}$ with suitably preassigned coefficients $a_{jn} = \gamma_j = \gamma_j(n)$, $1 \leqq j \leqq k$, fulfilling condition (16) with the new choice $R = [\max |z|, z \text{ on } S]$. In fact, the coefficients $\gamma_j(n)$ of the classical T-polynomial $t_n(z) = z^n + \gamma_1(n)z^{n-1} + \cdots + \gamma_{n-1}(n)z + \gamma_n(n)$ on S, by Fejér's theorem, satisfy (16) with this special choice of R, and the nth degree $(n-1)$-fold restricted T-polynomials $T_n(z, S) \in A_n(\gamma_1(n), \cdots, \gamma_{n-1}(n))$ minimizing the classical T-norm on S obviously coincide with $t_n(z)$, thus satisfy $[\max |T_n(z, S), z \text{ on } S]^{1/n} \to \tau(S)$ as $n \to \infty$. Hence, by Theorem 1, the validity of (1) for all $A_n(z, S) \in A_n(\gamma_1(n), \cdots, \gamma_{n-1}(n))$ of least q. T. norm N on S, subject to (16) with $R = [\max |z|, z \text{ on } S]$ as required, although (12) does not hold.

(12), with the assumption of (16), is sufficient for (1), but now prove for an arbitrary compact set S that a slightly stronger condition on $k(n)$ than (12), namely (22), is sufficient for (1) with a weaker assumption than (16) concerning the centroids c_j defined by (17).

THEOREM 8. *Let S be an arbitrary compact set, and let N be a continuous* q.T.*-norm on S. Let $A_n(z)\equiv z^n+\cdots$ minimize N over $A_n(\gamma_1, \gamma_2, \cdots, \gamma_k)$ with*

$$k=o\left(\frac{n}{\log n}\right), \tag{22}$$

$$|c_j|\leqq\alpha_j\,, \qquad 1\leqq j\leqq k\,, \tag{23}$$

where the α_j are independent of n.

Then (1) *is valid provided the power series*

$$\sum_{h=1}^{\infty}\alpha_h z^h/h!$$

has positive radius of convergences r.[4]

Let $t_h(z)\equiv z^h+a_1^{(h)}z^{h-1}+\cdots+a_h^{(h)}$ be the hth degree classical Tchebycheff polynomial minimizing the norm $M[A_h(z), S]$ on S among all polynomials $A_h(z)$ with the leading terms z^h. Then a majorant of the nth degree generalized Tchebycheff polynomial $T_n(z, S)\in A_n$ can be expressed as the product $t_{n-k}(z)\cdot(z^k+\lambda_1 z^{k-1}+\cdots+\lambda_k)$ where the coefficients $\lambda_j=\lambda_j(k, n)$ satisfy the linear equations

$$\lambda_1+a_1^{(n-k)}=\gamma_1,\ \lambda_2+\lambda_1 a_1^{(n-k)}+a_2^{(n-k)}=\gamma_2,\ \cdots$$

$$\lambda_k+\lambda_{k-1}a_1^{(n-k)}+\lambda_{k-2}a_2^{(n-k)}+\cdots+a_{n-2k}^{(n-k)}=\gamma_k\,.$$

Hence, by Fejér's Theorem [2] on the zeros of $t_h(z)$

$$|\lambda_1|\leqq|\gamma_1|+|a_1^{(n-k)}|\leqq\binom{n}{1}\alpha_1+\binom{n-k}{1}R<nR\left(1+\frac{\alpha_1}{R}\right)$$

where $R\geqq[\max|z|,\ z\in S]+1$. Similarly,

---

[4] While the research here presented was in progress, Professor G. Szegö communicated to the first named author the following result. *Let L be an analytic Jordan curve. Let the positive constants α_j satisfy the condition $|\alpha_j|<ab^j$, $j=0, 1, 2, \cdots$, where a and b are arbitrary positive constants. Let the integer $k=k(n)$ satisfy the condition $k(n)=o(n/\log n)$. There exist polynomials $A_n(z)\equiv z^n+\cdots$ satisfying condition* (23) *such that*

$$\lim_{n\to\infty}\max_{z\in L}|A_n(z)|^{1/n}=\tau(L)$$

where $\tau(L)$ is the transfinite diameter of L.

This communication induced us to study the problem dealt with in Theorem 8.

$$|\lambda_2| \leqq |\gamma_2| + |\lambda_1|\, |a_1^{(n-k)}| + |a_1^{(n-k)}| \leqq \binom{n}{2}\alpha_2 + nR\left(1+\frac{\alpha_1}{R}\right)\binom{n-k}{1}R$$

$$+\binom{n-k}{2}R^2 < 2n^2R^2\left(1+\frac{\alpha_1}{R}+\frac{\alpha_2}{2!R^2}\right),$$

$$\lambda_3 \leqq |\gamma_3| + |\lambda_2|\, |a_1^{(n-k)}| + |\lambda_1|\, |a_2^{(n-k)}| + |a_3^{(n-k)}|$$

$$\leqq \binom{n}{3}\alpha_3 + 2n^2R^2\left(1+\frac{\alpha_1}{R}+\frac{\alpha_2}{2!R^2}\right)\binom{n-k}{1}R$$

$$+nR\left(1+\frac{\alpha_1}{R}\right)\binom{n-k}{2}R^2 + \binom{n-k}{3}R^3 \leqq 2^2n^3R^3\left(1+\frac{\alpha_1}{R}+\frac{\alpha_2}{2!R^2}+\frac{\alpha^3}{3!R^3}\right),$$

and so on; finally

$$|\lambda_k| \leqq 2^{k-1}n^kR^k\left(1+\frac{\alpha_1}{R}+\frac{\alpha_2}{2!R^2}+\cdots+\frac{\alpha_k}{k!\,R^k}\right).$$

Hence, for $R > 1/r$ we have

$$M[T_n(z, S), S] < M[t_n(z), S]\cdot(k+1)2^{k-1}n^kR^k\left(1+\frac{\alpha_1}{R}+\frac{\alpha_2}{2!R^2}\right.$$

$$\left.+\cdots+\frac{\alpha_k}{k!\,R^k}+\cdots \text{ in inf.}\right)$$

and, therefore, by (22) and

$$\lim\, \{M[t_n(z), S]\}^{1/n} = \tau(S),$$

$$\limsup\, \{M[T_n(z, S), S]\}^{1/n} \leqq \tau(S).$$

On the other hand

$$\liminf\, \{M[T_n(z, S), S]\}^{1/n} \geqq \tau(S).$$

Thus the proof is complete for the classical Tchebycheff norm M, and the theorem follows by Theorem 1.

As a converse of the proposition just demonstrated we prove

THEOREM 8 bis. *Let S satisfy the assumptions of Theorem* 7. *Suppose $k=k(n)\to\infty$ as $n\to\infty$ but $k(n)\neq o(n/\log n)$. Then there exist polynomials $\alpha_n(z, N)\in A_n(\gamma_1, \gamma_2, \cdots, \gamma_k)$ of least* q.T.*-norm $N=N[\alpha_n(z), S]$ on S with $|\gamma_j^{(n)}| \leqq \binom{n}{j}\alpha_j$ for $1\leqq j\leqq k(n)$ and* $\limsup\, [\alpha_h/h!]^{1/h} < \infty$ *for which* (1) *is not valid.*

Since Theorem 7 established the existence in question in the case $\alpha_j = R^j$ if $k(n)\neq o(n)$, we assume $\limsup\,(k\log n/n) > 0$ with $k/n\to 0$. Then we fix $\gamma_j=\gamma_j(n)$, $1\leqq j\leqq k=k(n)$, by

$$\alpha_n(z) \equiv P_n(z) + \binom{n}{k}(k!)^{1/2}R^k P_{n-k}(z) + \lambda_1 P_{n-k-1}(z) + \cdots + \lambda_{n-k}P_0(z) ,$$

where $P_n(z)$, C_{R_1}, and R_1 have the same meaning as in § 8; thus $|\gamma_j(n)| < \binom{n}{j} R^j$ for $1 \leqq j \leqq k-1$ and $|\gamma_k(n)| < \binom{n}{k}$ $[1+(k!)^{1/2}]R^k$ and therefore $\alpha_h = 1+(h!)^{1/2}$ satisfies our hypothesis. The nth degree polynomial $\alpha_n(z, N_0)$ of least square norm

$$N_0 = N_0[\alpha_n(z),\ C_{R_1}] = \left\{ \int_{C_{R_1}} |\alpha_n(z)|^2 |dz| \right\}^{1/2}$$

on C_{R_1} among all

$$\alpha_n(z) \equiv z^n + \cdots \in A_n(\gamma_1, \cdots, \gamma_n)$$

is obviously

$$P_n(z) + \binom{n}{k}(k!)^{1/2}R^k P_{n-k}(z) ,$$

and we have

$$\int_{C_{R_1}} |\alpha_n(z, N_0)|^2 |dz| > R^{2k}k! \int_{C_{R_1}} |P_{n-k}(z)|^2 |dz| = R^{2k}k! p_{n-k} .$$

In view of $k(n) = o(n)$, $p_n^{1/2n} \to \tau(C_{R_1}) = R_1\tau(S)$, $\limsup\,[k \log n/n] = \varepsilon(>0)$, we have now

$$\limsup \left\{ \int_{C_{R_1}} |\alpha_n(z, N_0)|^2 |dz| \right\}^{1/2n} \geqq R_1\tau(S) \limsup\,(k!)^{1/2n}$$
$$\geqq R_1\tau(S)e^{\varepsilon/2} > R_1\tau(S) .$$

Combination of this inequality with (18) and (19) yields

$$\limsup\,\{M[T_n(z, S),\ S]\}^{1/2} > \tau(S)$$

for the polynomial $T_n(z, S) \in A_n(\gamma_1, \gamma_2, \cdots, \gamma_k)$ minimizing the classical norm $M[\alpha_n(z), S]$ on S, and the theorem is established.

10. We conclude our investigations concerning the validity of (1) with $A_n(z, N) \in A_n(\gamma_1, \gamma_2, \cdots, \gamma_k)$ by considering the particular case

$$k = k(n) = O(1)$$

nearest to the simplest one: $k = \text{const.}$ Using arguments similar to those just applied in the proof of Theorem 8 we obtain at once, for arbitrary compact set S

$$\begin{aligned}\tau(S)^n &\leqq M[T_n(z, S), S] \\ &\leqq M[t_n(z), S]\cdot(k+1)\cdot 2^{k-1}n^k R^k\left(1+\frac{\alpha_1}{R}+\frac{\alpha_2}{2!R^2}+\cdots+\frac{\alpha_k}{k!R^k}\right)\end{aligned}$$

provided $|\gamma_j| \leqq \binom{n}{j}\alpha_j$, $\alpha_j=\alpha_j(n)$. Hence by the hypothesis $k=O(1)$

$$\begin{aligned}\tau(S) &\leqq \liminf \{M[T_n(z, S), S]\}^{1/n} \leqq \limsup \{M[T_n(z, S), S]\}^{1/n} \\ &\leqq \tau(S)\cdot\max(1, \limsup \{[\max \alpha_j(n), 1 \leqq j \leqq k=k(n)]\}^{1/n}) .\end{aligned}$$

Hence the validity of (1) for the classical Tchebycheff norm M and thus also for each continuous q.T.-norm N provided

$$\limsup \{\alpha_j(n)\}^{1/n} \leqq 1 , \quad 1 \leqq j \leqq \max_{n \geqq 2} k(n) . \tag{24}$$

Conversely, the condition (24) in case $|\gamma_j| \leqq \binom{n}{j}\alpha_j(n)$, $1 \leqq j \leqq k(n)=O(1)$, is also necessary for (1) with $A_n(z, N) \in A_n(\gamma_1, \cdots, \gamma_k)$ for N the classical Tchebycheff norm M and hence for arbitrary q.T.-norms N continuous in A_n on a compact set S of connected regular complement K.

In fact, fix $\gamma_j=\gamma_j(n)$, $1 \leqq j \leqq k=k(n)$, by

$$A_n(z) \equiv P_n(z)+\alpha_k(n)\binom{n}{k}P_{n-k}(z)+\lambda_1 P_{n-k-1}(z)+\cdots+\lambda_{n-k}P_0(z) ,$$

where the $P_m(z)=z^m+\cdots$ are mutually orthogonal on C_{R_1} and R, $R_1(R)$, C_{R_1}, and S_{R_1} are defined as in § 8. Then with $\alpha_j(n) \equiv R^j$ for $1 \leqq j \leqq k-1$, the condition (24) is fulfilled. We shall show that (1) cannot hold under our hypotheses unless (24) is satisfied also for $j=k$. In the proof we may restrict ourselves to the classical Tchebycheff norm M, minimized by the generalized Tchebycheff polynomial $T_n(z, S) \in A_n(\gamma_1, \cdots, \gamma_k)$. Applying the results (18) and (19) of § 8 we can write

$$R_1^n M[T_n(z, S), S]\left\{\int_{C_{R_1}} |dz|\right\}^{1/2} \geqq \left\{\int_{C_{R_1}} |A_n(z, N_0)|^2 |dz|\right\}^{1/2} , \tag{25}$$

where $A_n(z, N_0) \in A_n(\gamma_1, \cdots, \gamma_k)$ is the nth degree polynomial of least square norm $N_0(A_n(z), S_{R_1})=\left\{\int_{C_{R_1}} |A_n(z)|^2 |dz|\right\}^{1/2}$ on C_{R_1}. Our above choice of $A_n(\gamma_1, \cdots, \gamma_k)$ yields

$$A_n(z, N_0) \equiv P_n(z)+\alpha_k(n)\binom{n}{k}P_{n-k}(z) ,$$

whence

$$(26) \qquad \{N_0[A_n(z, N_0), S_{R_1}]\}^2 = p_n + \left\{ \alpha_k(n)\binom{n}{k} \right\}^2 p_{n-k} \, .$$

Combining (25) with (26) leads in view of $\lim p_m^{1/m} = \tau(S_{R_1})$ to

$$R_1 \limsup \{M[T_n(z, S), S]\}^{1/n} \geqq \tau(S_{R_1}) \limsup \{\alpha_k(n)\}^{1/n} \, ,$$

which by $\tau(S_{R_1}) = R_1 \tau(S)$ is equivalent to

$$\limsup \{M[T_n(z, S), S]\}^{1/n} \geqq \tau(S) \limsup \{\alpha_k(n)\}^{1/n} \, .$$

This establishes (24) as a necessary condition for (1) in case of the M-norm and hence also for all q.T.-norms N on any set S of the type considered. We summarize the foregoing results in

THEOREM 9. *Let S be an arbitrary compact set, and let N be a continuous* q.T.*-norm on S. Let $A_n(z) = z^n + \cdots$ minimize N over $A_n(\gamma_1, \cdots, \gamma_k)$ with*

$$(27) \qquad k = k(n) = O(1) \, ,$$

$$(28) \qquad |\gamma_j| \leqq \binom{n}{j} \alpha_j(n) \, , \qquad 1 \leqq j \leqq k \, .$$

Then (1) *is valid provided* (24) *holds. Conversely, in case* (27) *is fulfilled,* (24) *is also necessary for* (1) *provided S is a compact set of regular complement K.*

11. In the previous sections we developed conditions, necessary or sufficient or both, for the validity of (1). By Theorem 1 such conditions are the same for all q.T.-norms N which are defined on the set S considered. If (1) does not hold, there might be two possibilities:

(a) $\lim \{N[A_n(z, N), S]\}^{1/n} = \omega(S, N, \gamma_1, \gamma_2, \cdots, \gamma_k)$, with the polynomials $A_n(z, N)$ of least N-norm restricted to some given $A_n(\gamma_1, \cdots, \gamma_k)$, but ω is different from $\tau(S)$;

(b) the $\{N[A_n(z, N), S]\}^{1/n}$ have no limit as $n \to \infty$ if $A_n(z, N) \in A_n$, that is

$$\liminf \{N[A_n(z, N), S]\}^{1/n} = \alpha(S, N, \gamma_1, \cdots, \gamma_r)$$

is actually smaller than

$$\limsup \{N[A_n(z, N), S]\}^{1/n} = \beta(S, N, \gamma_1, \cdots, \gamma_k) \, .$$

12. It is easy to show that both possibilities (a) and (b) may eventually occur. In the light of this fact the following result has some intrinsic interest:

THEOREM 10. *Let S be an arbitrary compact set and let $N=N[A_n(z), S]$ be any given* q.T.*-norm defined and continuous in $A_n=A_n(\gamma_1, \cdots, \gamma_k)$ on S. Let the least N-norm ν_n on S for polynomials $A_n(z)\in A_n$ satisfy*

$$\liminf \nu_n^{1/n}=\alpha(S, N, \gamma_1, \cdots, \gamma_k) ,$$

$$\limsup \nu_n^{1/n}=\beta(S, N, \gamma_1, \cdots, \gamma_k) .$$

Then α and β are independent of the particular choice of N and, consequently $\lim \nu_n^{1/n}$ *exists or not for all* q.T.*-norms N, the coexisting limits having the same value $\omega(S, \gamma_1, \cdots, \gamma_n)$.*

Applying (5) with $A_n(z)\equiv A_n(z, N)$ the polynomial of least N-norm on S for $A_n(z)\in A_n$, we obtain

$$\begin{aligned}(29)\qquad \liminf \{N[A_n(z, N), S]\}^{1/n} &\geqq \liminf \{M[A_n(z, N), S]\}^{1/n}\\ &\geqq \liminf \{M[A_n(z, M), S]\}^{1/n} ,\\ \limsup \{N[A_n(z, N), S]\}^{1/n} &\geqq \limsup \{M[A_n(z, N), S]\}^{1/n}\\ &\geqq \limsup \{M[A_n(z, M), S]\}^{1/n} ,\end{aligned}$$

while (4) applied to $A_n(z)\equiv A_n(z, M)\in A_n$ yields

$$\begin{aligned}(30)\qquad \liminf \{N[A_n(z, N), S]\}^{1/n} &\leqq \liminf \{M[A_n(z, M), S]\}^{1/n} ,\\ \limsup \{N[A_n(z, N), S]\}^{1/n} &\leqq \limsup \{M[A_n(z, M), S]\}^{1/n} .\end{aligned}$$

Combining (29) with (30) leads to

$$\alpha(S, N, \gamma_1, \cdots, \gamma_k)=\alpha(S, M, \gamma_1, \cdots, \gamma_k) ,$$

and we similarly obtain

$$\beta(S, N, \gamma_1, \cdots, \gamma_k)=\beta(S, M, \gamma_1, \cdots, \gamma_k)$$

for all q.T.-norms N defined and continuous in A_n on S. Thus the proof for the independence of α and β of the choice of N is complete and hence the rest of the theorem follows if $\alpha=\beta$.

PART II

ASYMPTOTIC PROPERTIES OF THE MODULI, AND OF THE ZEROS OF POLYNOMIALS OF LEAST NORM

1. In Part I we have developed primarily sufficient conditions for the validity of (1); we propose now to consider necessary conditions for (1), namely consequences of (1) such as (2) which are significant in the study of restricted extremal polynomials. Our first three theorems are entirely general, without special reference to extremal polynomials.

THEOREM 11. *Let S be a point set of positive transfinite diameter whose complement K is a region containing the point at infinity, and let the zeros of the polynomials $p_n(z) \equiv z^n + a_{n1}z^{n-1} + \cdots + a_{nn}$ be uniformly bounded. Then*

$$\limsup_{n\to\infty} P_n^{1/n} \leqq \tau(S)\ , \qquad P_n = [\max |p_n(z)|,\ z \text{ on } S]\ , \tag{31}$$

implies

$$\lim_{n\to\infty} (a_{n1}/n) = a_1\ , \tag{32}$$

where $-a_1$ is the conformal center of gravity of S. That is, the center of gravity of the zeros of $p_n(z)$ approaches the conformal center of gravity of S.

For any sequence of polynomials $p_n(z) \equiv z^n + \cdots$ we have

$$\liminf_{n\to\infty} P_n^{1/n} \geqq \tau(S)\ , \tag{33}$$

since the corresponding relation holds for the Tchebycheff polynomials of S; thus (31) is equivalent to

$$\lim_{n\to\infty} P_n^{1/n} = \tau(S)\ . \tag{34}$$

If $G(z)$ is the generalized Green's function for K with pole at infinity, a suitably chosen level locus C_R: $G(z) = \log R(>0)$ in K consists of a single Jordan curve containing in its interior both S and all the zeros of the $p_n(z)$. It then follows from (34) that we have exterior to C_R, uniformly on any closed bounded set exterior to C_R,

$$\lim_{n\to\infty} |p_n(z)|^{1/n} = |\phi(z)|\ , \tag{35}$$

where $\phi(z) \equiv \exp[G(z) + iH(z) + \log \tau(S)]$ and $H(z)$ is conjugate to $G(z)$ in K; we need merely apply a previous result [Fekete and Walsh, [5], Theorem 11], where (31) is used to establish (loc. cit.)

$$\limsup_{n\to\infty} [\max |p_n(z)|,\ z \text{ on } C_R]^{1/n} \leqq R\cdot\tau(S)\ ;$$

the closed interior of the present C_R contains all zeros of the $p_n(z)$ and has the transfinite diameter $R\cdot\tau(S)$.

We write (35) in the form

$$\lim_{n\to\infty} |p_n(z)/z^n|^{1/n} = |\phi(z)/z|\ , \tag{36}$$

which holds uniformly in some neighborhood of the point at infinity.

We set

$$\phi(z) \equiv z + a_1 + a_2 z^{-1} + \cdots ; \tag{37}$$

it is of course no loss of generality to choose $\phi'(\infty)=1$, and here $-a_1$ is by definition the conformal center of gravity of S. If the nth root in (36) is suitably chosen, namely with the value unity at infinity, (36) implies

$$\lim_{n\to\infty}\left[1+\frac{a_{n1}}{z^2}+\frac{a_{n2}}{z}+\cdots\right]^{1/n} = 1+\frac{a_1}{z}+\frac{a_2}{z^2}+\cdots \tag{38}$$

uniformly in some neighborhood of the point at infinity. We use here the theory of normal families of functions. Any infinite sequence of the functions in the first member of (38) is bounded and admits a subsequence converging uniformly in the neighborhood of infinity. All limit functions are analytic in this neighborhood, have the same modulus there, and are equal at infinity; hence these limit functions are identical, and the original sequence converges uniformly in a neighborhood of infinity to this limit function. Equation (38) implies (32).

2. **Of course this same reasoning** applies to the higher coefficients in (38); for instance

$$\frac{2a_{n2}-a_{n1}^2}{2n} \to a_2 - \frac{a_1^2}{2},$$

but (32) would seem to be the most interesting of these relations.

Equation (32) has been previously established by Schiffer [8] for the case that K possesses a classical Green's function, and where the $p_n(z)$ are the Fekete polynomials for S, whose zeros lie on S and maximize the discriminant.

In the hypothesis of Theorem 11 we may replace (31) by the corresponding inequality involving an arbitrary quasi-Tchebycheff norm; compare [Fekete and Walsh, [5], Theorem 2].

3. **The significance of Theorem** 11 in the theory of once-restricted and k-fold restricted extremal polynomials is that *if* (31) *is satisfied, then unless* (32) *is also satisfied the zeros of the* $p_n(z)$ *cannot be bounded*; thus (36) cannot be valid uniformly in the neighborhood of infinity, and may not be valid on every compact set in K. Of course (36) is valid uniformly in the neighborhood of infinity for all classical extremal polynomials [Fekete and Walsh, [5], Theorems 11 and 13].

An illustration here is illuminating; we choose S as $|z| \leq 1$ and prescribe merely the (constant) center of gravity c_1 ($\neq 0$) of the zeros of each $p_n(z)$. The extremal polynomials with the least-square norm on the boundary of S are

$$p_n(z) \equiv z^n - nc_1 z^{n-1} \equiv z^{n-1}(z - nc_1) ,$$

and the zeros of these polynomials are not bounded. It is striking that (36) continues to hold, but *nonuniformly*, in the neighborhood of infinity. Moreover, if we replace the prescribed c_1 by $c_1^{(n)}$, where $c_1^{(n)} \to 0$, then the zeros of $p_n(z)$ are bounded if and only if the numbers $nc_1^{(n)}$ are bounded.

Theorem 11 (like later results) does not require that the $p_n(z)$ be defined for every n; it is sufficient if these polynomials are given for an infinite sequence of values of n. Thus, *if the $p_n(z)$ are given and* (32) *is not satisfied, the zeros of the sequence $p_n(z)$ cannot be bounded; if* (32) *is satisfied for no subsequence of indices n, the zeros of no subsequence of the $p_n(z)$ can be bounded; in particular if the $p_n(z)$ are once-restricted extremal polynomials $p_n(z) \equiv z^n + na_1 z^{n-1} + \cdots$ with constant a_1 different from the conformal center of gravity of S, there exists a unique sequence for n sufficiently large of zeros $\mathfrak{z}_n$ of the respective $p_n(z)$, where $\mathfrak{z}_n$ lies exterior to the extended convex hull H_1 of S, with $\mathfrak{z}_n \to \infty$, and all other zeros of $p_n(z)$ lie in H_1*; compare Fekete and Walsh [6; Theorem VII].

4. **Deeper results can be established** concerning the zeros of the $p_n(z)$ which become infinite.

THEOREM 12. *Let S be a point set of positive transfinite diameter whose complement K is a region containing the point at infinity. Let $G(z)$ be the generalized Green's function for K with pole at infinity, and let Γ be a Jordan curve in K which separates S from the point at infinity. Suppose*

$$p_n(z) \equiv z^n + a_{n1} z^{n-1} + \cdots + a_{nn} ,$$

and suppose (31) *is valid. Let us write $p_n(z) \equiv q_{n-\sigma}(z) \cdot r_\sigma(z)$, where $r_\sigma(z) \equiv z^\sigma + \cdots$ is a polynomial whose zeros are precisely the zeros of $p_n(z)$ exterior to Γ. Then we have*

$$\sigma = \sigma(n) = o(n) , \tag{39}$$

$$\lim_{n\to\infty} Q_{n-\sigma}^{1/(n-\sigma)} = \tau(S), \qquad Q_{n-\sigma} = \max\,[|q_{n-\sigma}(z)|,\ z \text{ on } S] . \tag{40}$$

Equation (39) follows at once [Walsh and Evans, [10]], for the number $n-\sigma$ of zeros of $p_n(z)$ on and interior to Γ satisfies $(n-\sigma)/n \to 1$.

Since S is closed, the distance d from S to Γ is positive, so for z on S we have $|r_\sigma(z)| > d^\sigma$. Consideration of a point z of S at which $|q_{n-\sigma}(z)| = Q_{n-\sigma}$ then yields

$$|q_{n-\sigma}(z)|\cdot d^{\sigma}=Q_{n-\sigma}\cdot d^{\sigma}<Q_{n-\sigma}\cdot|r_{\sigma}(z)|\leqq P_n \ . \tag{41}$$

Equation (39) implies $d^{\sigma/n}\to 1$, whence from (41)

$$\limsup Q_{n-\sigma}^{1/n}\leqq \limsup P_n^{1/n}\leqq \tau(S) \ .$$

But we may write also

$$\limsup Q_{n-\sigma}^{1/n}=\limsup Q_{n-\sigma}^{1/(n-\sigma)} \ ,$$

so (40) follows by the analogue of (33).

Of course it is a consequence of (40) that the center of gravity of the zeros of $q_{n-\sigma}(z)$ approaches the conformal center of gravity of S.

It follows from Theorem 12 [Walsh and Evans, [10], p. 335] that on any closed set exterior to Γ we have

$$\lim_{n\to\infty}|q_{n-\sigma}(z)|^{1/(n-\sigma)}=|\phi(z)| \ ,$$

the analogue of (35).

5. **Our main interest lies in** the zeros of $p_n(z)$ which become infinite, but Theorem 12 deals also with also with other zeros. In particular, *if $p_n(z)$ is a k-fold restricted ($k=$const.) extremal polynomial on S for a monotonic quasi-Tchebycheff norm, and if either Γ is the boundary of $H_k(S)$ (supposed to contain S in its interior) or is a curve containing $H_k(S)$ in its closed interior, then at most k zeros of $p_n(z)$ lie exterior to Γ; we have $\sigma(n)\leqq k$. Moreover, equation* (35) *is valid uniformly on any closed bounded set in K containing no limit point of the zeros of the $p_n(z)$; and* (35) *with $p_n(z)$ replaced by $q_{n-\sigma}(z)$ is valid uniformly on any closed bounded set in K containing no limit point of the zeros of the $q_{n-\sigma}(z)$, in particular is valid on any closed bounded set exterior to Γ* [compare Walsh and Evans, [10], p. 335].

6. **Theorem 13 is complementary** to Theorem 12:

THEOREM 13. *Under the conditions of Theorem* 12 *we have*

$$\lim_{n\to\infty}R_{\sigma}^{1/n}=1, \qquad R_{\sigma}=[\max|r_{\sigma}(z)|,\ z \text{ on } S] \ . \tag{42}$$

There exists a number D independent of Z such that for each fixed point Z on or exterior to Γ and as Z_1 and Z_2 range over S, we have for the distances

$$\frac{\max\overline{ZZ_1}}{\min\overline{ZZ_2}}\leqq D \ ,$$

for the first member depends continuously on Z and approaches unity as $Z \to \infty$. The zeros of $r_\sigma(z)$ lie exterior to Γ, so for z on S we have

$$\frac{\max |r_\sigma(z)|}{\min |r_\sigma(z)|} \leqq D^\sigma .$$

From (41) we may write for z on S

$$Q_{n-\sigma} \cdot [\min |r_\sigma(z)|] \leqq P_n ,$$

$$Q_{n-\sigma} \cdot R_\sigma \leqq Q_{n-\sigma} \cdot [\min |r_\sigma(z)|] D^\sigma \leqq P_n \cdot D^\sigma . \tag{43}$$

In equation (40) we may replace the exponent $1/(n-\sigma)$ by $1/n$, so from (43), (31), and (39) we have

$$\limsup_{n\to\infty} R_\sigma^{1/n} \leqq 1 . \tag{44}$$

However, (33) for arbitrary polynomials here shows

$$\liminf_{n\to\infty} R_\sigma^{1/\sigma} \geqq \tau(S) ,$$

$$\liminf_{n\to\infty} \left[\frac{1}{n} \log R_\sigma\right] \geqq 0 , \tag{45}$$

and (42) follows.

7. A result similar to Theorems 12 and 13 is the following.

Under the hypothesis of Theorem 12 *let* Γ_1 *be an arbitrary Jordan curve in K containing on or within it no point of S, and let* $r_\sigma(z) \equiv z^\sigma + \cdots$ *be the polynomial whose zeros are the zeros of* $p_n(z)$ *on and within* Γ_1, *with* $p_n(z) \equiv q_{n-\sigma}(z) \cdot r_\sigma(z)$. *Then we have* (39), (40), (42), *and the relation* $\lim |q_{n-\sigma}(z)|^{1/(n-\sigma)} = |\phi(z)|$ *uniformly on any closed set interior to* Γ_1. Here (39) follows at once [Walsh and Evans, [10]], (40) follows as in the proof of Theorem 12, (42) follows from the boundedness of the zeros of $r_\sigma(z)$ and from (45), and the remaining remark is immediate [Walsh and Evans, [10], p. 335].

8. Theorems 12 **and** 13, **devoted to** arbitrary sequences of polynomials $p_n(z)$ such as those studied in Walsh and Evans [10], yield a precise result for restricted extremal polynomials in the case $k=1$.

THEOREM 14. *Let S be a closed bounded set of positive transfinite diameter, and let* $p_n(z) \equiv z^n + \cdots$ *be the sequence of once-restricted polynomials on S, with constant center of gravity of the zeros different from the conformal center of gravity of S, extremal for a monotonic quasi-Tchebycheff norm. Let C be a (closed) circular disk containing S, let C'* *be a concentric disk whose radius is three times as great, and let*

$$r_\sigma(z) \equiv z^\sigma + \cdots, \qquad \sigma = \sigma(n) = 1 \text{ or } 0,$$

be the polynomial whose zero is the zero of $p_n(z)$ *if any exterior to* C', *otherwise unity. Set*

$$q_{n-\sigma}(z) \equiv z^{n-\sigma} + \cdots, \qquad p_n(z) \equiv q_{n-\sigma}(z) \cdot r_\sigma(z).$$

Then (34), (40), *and* (42) *are valid. Moreover on any closed bounded set* S_1 *exterior to* H *we have uniformly* (*notation of* (35))

$$\lim |p_n(z)|^{1/n} = \lim |q_{n-\sigma}(z)|^{1/n} = |\phi(z)| . \tag{46}$$

The present writers have previously [6] shown that *if a zero of* $p_n(z)$ *lies exterior to* C', *then all other zeros of* $p_n(z)$ *lie in* C. Moreover it is remarked in § 3 that under the present conditions a zero of $p_n(z)$ lies exterior to C' for n sufficiently large. Thus $\sigma(n)=1$ for n sufficiently large. Equation (34) is known [part I, § 10], (40) follows from Theorem 12, and (42) from Theorem 13.

The zeros of the polynomials $p_n(z)$ *and* $q_{n-\sigma}(z)$ *have no* (*finite*) *limit point exterior to* H. Indeed, if $z=\alpha$ is assumed to be such a limit point, let Γ be a circular disc containing H in its interior but to which α is exterior. For n sufficiently large a zero of $p_n(z)$ lies exterior to the disc concentric with Γ whose radius is three times as great, and consequently [Fekete and Walsh, [6], Theorem IX] all other zeros of $p_n(z)$ lie in Γ, which contradicts the assumption of α as a limit point of zeros.

Equation (46) now follows [Walsh and Evans, [10], p. 335]. If S_1 is a closed bounded set exterior to H, for n sufficiently large no zeros of $p_n(z)$ lie on S_1.

9. Under the conditions of Theorem 14 we can obtain some information about the asymptotic behavior of the one zero z_1 among the totality of zeros $(z_1, z_2, \cdots, z_n)$ of $p_n(z)$ which becomes infinite. If α is the prescribed center of gravity and a the conformal center of gravity of S, we have $z_1 + z_2 + \cdots + z_n = n\alpha$, and by Theorem 11

$$\frac{z_2 + z_3 + \cdots + z_n}{n-1} \to a, \qquad \frac{z_1 + z_3 + \cdots + z_n}{n} \to a, \qquad \frac{z_1}{n} \to \alpha - a .$$

10. We are not in a position to extend Theorem 14 to the case of k-fold restricted extremal polynomials, $k > 1$, for with $k > 1$ precise conditions are as yet unknown concerning the number of zeros of $p_n(z)$ which become infinite or indeed lie exterior to H. For instance, if C is $|z|=1$ and we use the least-square norm on C with $k=2$, the twice-restricted extremal polynomial

$$p_n(z) \equiv z^n + 0 \cdot z^{n-1} + \binom{n}{2} c_2 z^{n-2} , \qquad c_2 \neq 0 ,$$

has two zeros $\pm[-n(n-1)c_2/2]^{1/2}$ which become infinite, whereas the twice-restricted extremal polynomial

$$p_n(z) \equiv z^n + nc_1 z^{n-1} + 0 \cdot z^{n-2} , \qquad c_1 \neq 0 ,$$

has but one zero $-nc_1$ which becomes infinite.

Nevertheless, for $k > 1$ *if we know that k zeros of $p_n(z)$ become infinite as $n \to \infty$, the remaining $n-k$ zeros of $p_n(z)$ have no limit point exterior to H.* For let $\phi_1, \phi_2, \cdots, \phi_n$ denote the angles subtended by S at the respective zeros of $p_n(z)$. Zedek's relation (Cf. Walsh and Zedek [11])

$$\phi_1 + \phi_2 + \cdots + \phi_{k+1} \geqq \pi$$

can be written in the form

$$\phi_{k+1} \geqq \pi - (\phi_1 + \phi_2 + \cdots + \phi_k) ,$$

and if $(\phi_1 + \phi_2 + \cdots + \phi_k) \to 0$, then $\phi_{k+1} \to \pi$ if $\phi_{k+1} < \pi$. To be more explicit, if $\phi_j \leqq \varepsilon/k$ for $j = 1, 2, \cdots, k$, then $\phi_j \geqq \pi - \varepsilon$ for $j = k+1, k+2, \cdots, n$, so the $n-k$ corresponding zeros of $p_n(z)$ lie in the locus of points from which S subtends an angle greater than or equal to $\pi - \varepsilon$. Under these conditions (that k zeros of $p_n(z)$ become infinite), the set of zeros of the $p_n(z)$ has no finite limit point exterior to H, and Theorem 14 admits a precise analogue. Even though the set of zeros of the $p_n(z)$ has no finite limit point exterior to H, not all zeros near S need lie in H; compare [Fekete and Walsh, [6], § 13].

11. If we are willing to forego an analogue of (42), we can obtain a further result on extremal polynomials.

THEOREM 15. *Let S be a closed bounded set of positive transfinite diameter, and let $p_n(z) \equiv z^n + \cdots$ be the sequence of k-fold restricted ($k =$ const.) polynomials on S extremal for a monotonic quasi-Tchebycheff norm. Let H_k be the inflated convex hull of S of order k, and let $q_{n-\sigma}(z) \equiv z^{n-\sigma} + \cdots$ be the polynomial whose zeros are the zeros of $p_n(z)$ on H_k. Then on any closed bounded set S_1 exterior to H_k we have uniformly the second of equations* (46).

Let $H_k^{(n)}$ be the (closed) point set consisting of all points at a distance not greater than $1/n$ from H_k, let $r_v(z) \equiv z^v + \cdots$ be the polynomial whose zeros are the zeros of $p_n(z)$ exterior to $H_k^{(n)}$, and let $t_{n-v}(z) \equiv z^{n-v} + \cdots$ be the polynomial whose zeros are the zeros of $p_n(z)$ on $H_k^{(n)}$. At most k zeros of $p_n(z)$ lie exterior to H_k, whence $0 \leqq v \leqq k$. For z

on S we have $|r_v(z)| \geqq 1/n^v \geqq 1/n^k$. We proceed further as in the proof of (41). If z is a point of S at which $|t_{n-v}(z)|$ takes its maximum value T_{n-v}, we have

$$|t_{n-v}(z)|/n^k = T_{n-v}/n^k \leqq T_{n-v}|r_v(z)| \leqq P_n \ .$$

There follows

$$\limsup T_{n-v}^{1/n} \leqq \limsup P_n^{1/n} = \tau(S) \ ,$$

and by (33)

$$\lim T_{n-v}^{1/n} = \tau(S) \ .$$

Thus we have for z on S_1

$$\lim |t_{n-v}(z)|^{1/n} = |\phi(z)| \ .$$

However, $t_{n-v}(z)$ has at most k zeros in H_k but not in $H_k^{(n)}$, so if d_0 (>0) and d_1 denote the shortest and longest distances between S_1 and $H_k^{(n)}$ for n sufficiently large, we have for z on S_1

$$\min\,[1, d_0, \cdots, d_0^k] \leqq \left| \frac{t_{n-v}(z)}{q_{n-\sigma}(z)} \right| \leqq \max\,[1, d_1, \cdots, d_1^k] \ ,$$

so the second of equations (46) follows.

Theorem 15 is not a consequence of Theorem 12, for in Theorem 15 the set S may have points in common with the boundary of H_k.

12. Under the conditions of Theorem 15, if the set S is real, it is known [[6], Part II, § 15] that at most k zeros of a real $p_n(z)$ lie exterior to the convex hull H of S. Precisely the method of proof of Theorem 15 (details are left to the reader) yields

THEOREM 16. *Let the real set S and the real polynomials $p_n(z)$ satisfy the conditions of Theorem* 15. *Let $q_{n-\sigma}(z) \equiv z^{n-\sigma} + \cdots$ be the polynomial whose zeros are the zeros of $p_n(z)$ on H. Then the second of equations* (46) *is valid uniformly on any closed bounded set S_1 exterior to H.*

13. We return to the consideration of asymptotic relations, without specific reference to restricted polynomials. Equation (42) is derived in Theorem 13 as a necessary condition on the polynomials $r_\sigma(z)$, but is sufficient in the following sense. *If S satisfies the conditions of Theorem* 12, *if $\sigma = \sigma(n) = o(n)$, and if for polynomials $q_{n-\sigma}(z) \equiv z^{n-\sigma} + \cdots$ and $r_\sigma(z) \equiv z^\sigma + \cdots$ we have* (40) *and* (42), *then we also have* (31) *with $p_n(z) \equiv q_{n-\sigma}(z) \cdot r_\sigma(z)$.* In fact we may write $p_n \leqq Q_{n-\sigma} \cdot R_\sigma$, whence (31) follows by (40), (42), and (33). In this remark there are no geometric conditions on the zeros of $q_{n-\sigma}(z)$ and $r_\sigma(z)$, but in connection with restricted extremal polynomials the most interesting situation is that the zeros of $q_{n-\sigma}(z)$ are bounded whereas the zeros of $r_\sigma(z)$ are not necessarily bounded.

14. **The polynomials** $r_\sigma(z)$ of Theorem 13 have various interesting properties:

THEOREM 17. *If S is a set of positive transfinite diameter, and if for the polynomials $r_\sigma(z)\equiv z^\sigma+\cdots$ we have* (42) *with $\sigma=\sigma(n)=o(n)$, then on any closed bounded set S' of positive transfinite diameter we have*

$$\lim_{n\to\infty} R_\sigma'^{1/n}=1\ , \qquad R_\sigma'=[\max |r_\sigma(z)|,\ z \text{ on } S']\ . \tag{47}$$

We note that (33) for arbitrary polynomials yields

$$\liminf R_\sigma'^{1/n}\geqq 1\ . \tag{48}$$

The generalized Bernstein Lemma is valid [Walsh, [9], § 4.9] even if we must use the generalized Green's function $G(z)$ instead of the classical Green's function for the maximal subregion K containing infinity of the complement of S. Let ρ (>1) be chosen so large that the level locus Γ_ρ: $G(z)=\log\rho$ in K separates S' from infinity. For z interior to Γ_ρ we have (loc. cit.)

$$|r_\sigma(z)|\leqq R_\sigma\cdot\rho^\sigma\ , \qquad R_\sigma'\leqq R_\sigma\cdot\rho^\sigma\ .$$

Equation (42) yields

$$\limsup_{n\to\infty} R_\sigma'^{1/n}\leqq 1\ ,$$

which with (48) gives (47). We have made no hypothesis on the location of the zeros of $r_\sigma(z)$ relative to S and S'.

Such a sequence as $r_\sigma(z)$ of Theorem 17 may in some respects be considered a "negligible sequence" with respect to n, provided $\sigma=\sigma(n)=o(n)$, in the sense that

(i) its presence or absence as a factor of $q_{n-\sigma}(z)\equiv z^{n-\sigma}+\cdots$ does not alter the value of

$$\lim_{n\to\infty}[\max |q_{n-\sigma}(z)|,\ z \text{ on } S]^{1/n}\ ,$$

or even the value of lim inf or lim sup here, and

(ii) this property of $r_\sigma(z)$ is not dependent on the particular set S (of positive transfinite diameter) chosen. *Any sequence of polynomials $r_\sigma(z)\equiv z^\sigma+\cdots$ whose zeros are bounded is in this sense a negligible sequence with arbitrary $\sigma=\sigma(n)=o(n)$*, for if these zeros and S lie on a point set of diameter D we have

$$R_\sigma \leqq D^\sigma ,$$

which together with (45) implies (42).

15. As an application of this proof of (42) we state

THEOREM 18. *Let S be a closed bounded set not necessarily of positive transfinite diameter, and let an arbitrary bounded set of points $z_1^{(n)}$, $z_2^{(n)}, \cdots, z_\sigma^{(n)}$ be given with $\sigma=\sigma(n)=o(n)$. Then there exists a sequence of polynomials $p_n(z)\equiv z^n+\cdots$ vanishing in the points $z_j^{(n)}$, such that* (1) *is valid.*

Indeed, there exists a sequence of polynomials $q_{n-\sigma}(z)\equiv z^{n-\sigma}+\cdots$ satisfying (40), the polynomials $r_\sigma(z)\equiv(z-z_1^{(n)})\cdots(z-z_\sigma^{(n)})$ satisfy (44), so (1) follows from (33).

16. For any of the classical norms and the polynomials $p_n(z)$ of Theorem 18, the analogue of (1) holds; we state: *the analogue of* (1) *holds for the extremal polynomials for any quasi-Tchebycheff norm, if the polynomials are required to vanish in the points $z_j^{(n)}$ $(j=1, 2, \cdots)$, provided $\sigma=o(n)$.* The polynomials $A_n(z, N)$ of least q.T. N-norm on S with $A_n(z_j^{(n)}, N)=0$ by (4) fulfill

$$N[A_n(z, N), S]\leqq N[P_n(z), S]\leqq U(S, N)P_n ,$$

while by (5) they satisfy

$$N[A_n(z, N), S]\geqq L_n(S, N, \varepsilon)\tau(S)^n$$

for $n\geqq n_0(N, \varepsilon)$. Hence the validity of (1) by virtue of $\lim P_n^{1/n}=\tau(S)$.

Addendum. Using a device, communicated by Prof. Szegö to the first named author after the conclusion of the research above presented we can prove the following counterpart of our Theorem 7:

THEOREM 7 bis. *Let S be an arbitrary compact set and R an arbitrary positive number. Suppose that* (16) *holds with $k=k(n)$ subject to* (12). *Then for all polynomials $A_n(z, N)\in A_k(\gamma_1, \cdots, \gamma_k)$ of least* q.T. *norm $N=N(A_n(z), S)$ on S,* (1) *is valid.*

By Theorem 1 we may restrict the proof of (1) to the particular case $N=M(A_n(z), S)$; thus $A_n(z, N)\equiv T_n(z, S)$, the k-fold restricted Tchebycheff polynomial in $A_n(\gamma_1, \gamma_2, \cdots, \gamma_k)$; and by a remark to Theorem 5 we can reduce this proof to the special case S: $|p(z)|=\rho^m$ with $p(z)\equiv z^m-(p_1z^{m-1}+\cdots+p_m)$, a lemniscate of radius $\delta=\tau(S)$. Then a majorant for $T_n(z, S)$ is the product $z^tp(z)^s(z^k+\delta_1z^{k-1}+\cdots+\delta_k)$, where t and s are nonnegative integers satisfying

$$n-k-ms=t\leqq m-1\ ,$$

while $z^k+\delta_1 z^{k-1}+\cdots+\delta_k$ is the principal part of the Laurent development

$$z^k+\delta_1 z^{k-1}+\cdots+\delta_k+\delta_{k+1}z^{-1}+\delta_{k+2}z^{-2}+\cdots$$

around $z=\infty$ of

$$(z^n+\gamma_1 z^{n-1}+\cdots+\gamma_k z^{n-k})z^{-t}p(z)^{-s}\ .$$

For each $z\neq 0$ with $|p_1|\,|z|^{-1}+\cdots+|p_m|\,|z|^{-m}<1$ we have

$$\begin{aligned}
&(z^n+\gamma_1 z^{n-1}+\cdots+\gamma_k z^{n-k})z^{-t}[p(z)]^{-s}\\
&\quad\equiv(z^k+\gamma_1 z^{k-1}+\cdots+\gamma_k)\{1-(p_1 z^{-1}+\cdots+p_m z^{-m})\}^{-s}\\
&\quad\equiv(z^k+\gamma_1 z^{k-1}+\cdots+\gamma_k)\Big\{1+s(p_1 z^{-1}+\cdots+p_m z^{-m})\\
&\qquad+\binom{s+1}{2}(p_1 z^{-1}+\cdots+p_m z^{-m})^2\\
&\qquad+\binom{s+2}{3}(p_1 z^{-1}+\cdots+p_m z^{-m})^3+\cdots\Big\}
\end{aligned}$$

whence

$$\delta_1=\gamma_1+sp_1,\ \ \delta_2=\gamma_2+\gamma_1 sp_1+\cdots,\ \ \delta_3=\gamma_3+\gamma_2 sp_1+\cdots+\binom{s+2}{3}p_1^3,\ \cdots .$$

Similarly, for the aforesaid values of z,

$$\begin{aligned}
&(z^k+|\gamma_1|z^{k-1}+\cdots+|\gamma_k|)\{1-(|p_1|z^{-1}+\cdots+|p_n|z^{-m})\}^{-s}\\
&\quad\equiv z^k+\Delta_1 z^{k-1}+\cdots+\Delta_k+\Delta_{k+1}z^{-1}+\cdots
\end{aligned}$$

with

$$\begin{aligned}
&\Delta_1=|\gamma_1|+s|p_1|\geqq|\delta_1|\ ,\\
&\Delta_2=|\gamma_2|+|\gamma_1|s|p_1|+\cdots\geqq|\delta_2|,\ \ \Delta_3\geqq|\delta_3|\ ,\ \cdots .
\end{aligned}$$

Hence, for every $r>0$ large enough to satisfy

$$|p_1|r^{-1}+\cdots+|p_m|r^{-m}<1\ ,$$

$$\begin{aligned}
&\max_{|z|=r}|z^k+\delta_1 z^{k-1}+\cdots+\delta_k|\\
&\quad\leqq r^k+\Delta_1 r^{k-1}+\cdots+\Delta_k+\Delta_{k+1}r^{-1}+\Delta_{k+2}r^{-2}+\cdots\\
&\quad=(r^k+|\gamma_1|r^{k-1}+\cdots+|\gamma_k|)\{1-(|p_1|r^{-1}+\cdots+|p_m|r^{-m})\}^{-s}\ ;
\end{aligned}$$

thus in case $|z|\leqq r$ covers the lemniscate $|p(z)|=\rho^m$ we have a fortiori

$$\begin{aligned}
&\max_{|p(z)|=\rho^m}|z^k+\delta_1 z^{k-1}+\cdots+\delta_k|^{1/n}\leqq r^{k/n}(1+|\gamma_1|r^{-1}+\cdots+|\gamma_k|r^{-k})^{1/n}\\
&\qquad\times\{1-(|p_1|r^{-1}+\cdots+|p_m|r^{-m})\}^{\frac{(n-k-t)}{n}\cdot\frac{1}{m}}\ .
\end{aligned}$$

By (16) and (12), for $r > R$ we obtain

$$\limsup \max_{|p(z)|=\rho^m} |z^k + \delta_1 z^{k-1} + \cdots + \delta_k|^{1/n} \leqq \{1 - (|p_1| r^{-1} + \cdots + |p_m| r^{-m})\}^{1/m}$$

which $(r \to \infty)$ yields $\limsup \max_{|p(z)|=\rho^m} |z^k + \delta_1 z^{k-1} + \cdots + \delta_k|^{1/n} \leqq 1$. We therefore have

$$\limsup [\max_{z \in S} |T_n(z, S)|]^{1/n} \leqq \rho = \tau(S).$$

Combining this with $\liminf \max_{z \in S} |T_n(z, S)|^{1/n} \geqq \tau(S)$ leads to the validity of (1) for the special norm and special set considered, whence its validity for arbitrary q.T. norms and arbitrary sets, as stated.

Using the above argument to obtain an upper bound for

$$[\max |z^k + \delta_1 z^{k-1} + \cdots + \delta_k|, \ z \in S \cdot |p(z)| = \rho^m]^{1/n}$$

if $k = k(n)$ is subject to (22), and $\gamma_j = \gamma_j(n)$, for $1 \leqq j \leqq k = k(n)$ subject to (23) with $\limsup \{\alpha_h (h!)^p\}^{1/h} < \infty$, we can prove (1) for all q.T. norms and arbitrary compact sets S provided p is a positive constant. This generalization of our Theorem 8 is due to Professor Szegö.

References

1. P. Davis and H. Pollak, *On an equivalent definition of the transfinite diameter* Proc. Amer. Math. Soc., **3** (1952), 152–158.

2. L. Fejér, *Ueber die Lage der Nullstellen von Polynomen, die aus Minimum-forderungen gewisser Art entspringen*, Math. Ann. **85** (1922), 41–48.

3. M. Fekete, *Asymptotic behaviour of certain polynomials of least deviation on a point set of a given transfinite diameter* (Abstract) Bull. Amer. Math. Soc. **60** (1954), 29.

4. ———, *On the structure of polynomials of least deviation*, Bull. Research Council Israel, **5** (1955), 9–17.

5. M. Fekete and J. L. Walsh, *On the asymptotic behavior of polynomials with extremal properties, and of their zeros*, J. d'analyse math., **4** (1954/55), 49–87.

6. ———, *On restricted infrapolynomials*, J. d'analyse math., **5** (1956-57), 47–76.

7. F. Leja, *Sur les polynomes de Tchebycheff et la fonction de Green*, Ann. Soc. Polon. Math., **19** (1946), 1–6.

8. M. Schiffer, *Sur la variation de la fonction de Green de domains plans quelconques*, C. R. Acad. Sci. Paris, **209** (1939), 980–982.

9. L. J. Walsh, *Interpolation and approximation*, Colloq. Publ. of Amer. Math. Soc. **20** (1935)

10. J. L. Walsh and J. P. Evans, *Note on the distribution of zeros of extremal polynomials*, Proc. Nat. Acad. Sci. **40** (1954), 332–337.

11. J. L. Walsh and M. Zedek, *On generalized Tchebycheff polynomials*, Proc. Nat. Acad. Sci. **42** (1956), 99–104.

12. M. Zedek, *Fejér's theorem on the zeros of extremal polynomials generalized*, Preliminary Report, Bull. Amer. Math. Soc., **61** (1955), 49.

THE HEBREW UNIVERSITY
HARVARD UNVIERSITY

BIORTHOGONAL SYSTEMS IN BANACH SPACES

S. R. FOGUEL

1. Introduction. We shall be interested, in this paper, in the following question: Given a biorthogonal system (x_n, f_n) in a separable Banach space B, under what conditions can one assert that the sequence $\{x_n\}$ constitutes a basis? The system (x_n, f_n) is called a biorthogonal system if

$$x_n \in B,\ f_n \in B^* \quad \text{and} \quad f_n(x_m)=\delta_{nm}\,.$$

We shall assume throughout the paper that $\|x_n\|=1$ and the sequence $\{x_n\}$ is fundamental. When the sequence $\{x_n\}$ constitutes a basis it will be called *regular* otherwise *irregular*.

2. Irregular systems. Let $\{x_n\}$ be an irregular sequence. (For example the trigonometric functions for $C(-\pi, \pi)$). The following definitions will be used.

$$\varphi_n(x)=\sum_{i=1}^{n} f_i(x)x_i$$

$$|||x|||=\sup\{\|\varphi_n(x)\|,\ n=1, 2, 3, \cdots\}$$

Compare [4]

$$E_0=\{x \mid \lim_{n\to\infty}\varphi_n(x)=x\}$$

$$E_1=\{x \mid\ |||x|||<\infty\}$$

$$E_2=\{x \mid \lim_{n\to\infty}\|\varphi_n(x)\|=\infty\}$$

$$E_3=\{x \mid\ |||x|||=\infty\}\,.$$

We have $E_0\subset E_1$ and $E_2\subset E_3$. For regular systems $E_0=E_1=B$ and $E_2=E_3=\phi$ where ϕ is the null set. The system is regular if and only if the sequence $\{\|\varphi_n\|\}$ is bounded [2], and if the sequence $\{\|\varphi_n\|\}$ is not bounded the set

$$\bigcap_{n=1}^{\infty}\{x \mid\ \|\varphi_n(x)\|\leqq K\}$$

is nowhere dense [2], hence for irregular systems the set

$$E_1=\bigcup_{k=1}\bigcap_{n=1}\{x \mid\ \|\varphi_n(x)\|\leqq K\}$$

is of the first category. Also $E_3=B-E_1$ is dense and of the second

Received February 22, 1956. This paper is part of a dissertation presented for the degree of Doctor of Philosophy Yale University.

category. In the case of regular systems there exists a number $K \geqq 1$ such that if $||x||=1$ then $1 \leqq |||x||| \leqq K$. The existence of such a bound, K, is equivalent to the equiboundedness of $\{||\varphi_n(x)||\}$ $||x||=1$ and therefore for irregular systems for any number a, there exists a point x such that $||x||=1$ and $|||x|||>a$, moreover such a point might be found in the linear manifold generated by $\{x_n\}$. (Equiboundedness of $\{||\varphi_n(x)||\}$ on a dense subset of the unit sphere would imply equiboundedness on the unit sphere.) It is interesting to note that for every number $a \geqq 1$ there exists a point x such that $||x||=1$ and $|||x|||=a$. There exists a point y_n satisfying

$$y_n=\sum_{i=1}^{n} a_i x_i, \quad ||y_n||=1, \quad |||y_n|||>a .$$

On the other hand $||x_1||=1$ and $|||x_1|||=1$. Let $0 \leqq t \leqq 1$, then $(1-t)x_1+ty_n \neq 0$. Define

$$g_\nu(t)=\left\|\varphi_\nu\left(\frac{(1-t)x_1+ty_n}{||(1-t)x_1+ty_n||}\right)\right\| \qquad \nu=12\cdots n$$

The functions $g_\nu(t)$ are continuous in t, and so is $g(t)$ where

$$g(t)=\sup\{g_\nu(t) \mid 1 \leqq \nu \leqq n\} .$$

$$g(0)=1$$

and

$$g(1)=\sup\{||\varphi_\nu(y_n)|| 1 \leqq \nu \leqq n\}=|||y_n|||>a .$$

There exists a number t_0 such that

$$0 \leqq t_0 < 1 \quad \text{and} \quad g(t_0)=a .$$

This following generalization of Baire's theorem [1] will be used: *Let $\{u_n(x)\}$ be a sequence of real valued continuous functions defined on a metric space c, and* $\lim u_n(x)=u(x)$, $|u_n(x)| \leqq M$, *then the set of points of discontinuity of u is of the first category.*

THEOREM 1. *The set E_2 is of the first category.*

Proof. Define the functions $u_n(x)$ by

$$u_n(x)=\frac{||\varphi_n(x)||}{1+||\varphi_n(x)||} .$$

We have $0 \leqq u_n(x) \leqq 1$ and if $x \in E_0 \cup E_2$ then

$$\lim u_n(x)=u(x)$$

where $u(x)=1$ for $x \in E_2$ and

$$u(x)=\frac{||x||}{1+||x||} \quad \text{for } x \in E_0 .$$

If E_2 is a set of the second category then there exists at least one point of continuity of u. Let us denote such a point by x_0.

The set E_0 is dense in B. Let $\{y_n\}$ by a sequence of points in E_0 with $\lim y_n=x_0$, then

$$u(x_0)=\lim u(y_n)=\lim\frac{||y_n||}{1+||y_n||}=\frac{||x_0||}{1+||x_0||}$$

The set E_2 is dense in B. If $x \in E_2$ and $y \in E_0$ then $x+y \in E_2$. Let $\{z_n\}$ be a sequence of points in E_2 with $\lim z_n=x_0$, then $u(z_n)=1$ and

$$\frac{||x_0||}{1+||x_0||}=u(x_0)=\lim u(z_n)=1$$

which is absurd.

THEOREM 2. *Let S be a subset of B such that each point $x \in S$ is the limit of some sequence $\{y_n\}$, $y_n \in B$, and the sequence $\{|||y_n|||\}$ is bounded, then S is of the first category.*

Proof. Define the functions $v_n(x)$ by

$$v_n(x)=\frac{|||\varphi_n(x)|||}{1+|||\varphi_n(x)|||}$$

then $0 \leqq v_1(x) \leqq v_2(x) \leqq \cdots < 1$. If $x \in B$ let $\lim v_n(x)=v(x)$. $v(x)=1$ for $x \in E_3$ and the set E_3 is dense, hence $v(x)=1$ at every point of continuity of v. Let x be a point of continuity of v and $\{z_n\}$ a sequence with $\lim z_n=x$, then

$$\lim v(z_n)=v(x)=1$$

therefore the sequence $\{|||z_n|||\}$ is unbounded. Thus the set S is contained in the set of points of discontinuity of v which is a set of the first category by Baire's theorem.

3. General criteria for regularity. From Theorems 1 and 2 we derive the following criteria.

THEOREM 3. *A necessary and sufficient condition for the regularity of the system (x_n, f_n) is.*

$$\sup \{\|\varphi_n(x)\|, n=1, 2, \cdots\} = \infty$$

implies

$$\lim \|\varphi_n(x)\| = \infty . \quad (\text{or } E_2 = E_3) .$$

Proof. If the system is regular, then $E_2 = E_3 = \phi$. On the other hand, if the system is irregular E_2 is of the first category and E_3 of the second category.

Let
$$\psi_n(x) = \sum_{i=1}^{n} a_i^n x_i$$

denote the point nearest to x on the subspace spanned by

$$\{x_1, x_2, x_3, \cdots, x_n\} .$$

THEOREM 4. *The system* (x_n, f_n) *is regular if and only if the sequence* $\{|||\psi_n(x)|||\}$ *is bounded for each* x.

Proof. If the system is regular, then there exists a positive number K, such that $|||x||| \leqq K\|x\|$. Then

$$|||\psi_n(x)||| \leqq K\|\psi_n(x)\| \leqq K(\|x\| + \|x - \psi_n(x)\|) \leqq K2\|x\| ,$$

hence the condition is necessary. Sufficiency is clear by Theorem 2.

4. Biorthogonal systems in Hilbert spaces. In this section we assume that B is a Hilbert space. In order to use Theorem 4 let us compute $|||\psi_n(x)|||$. $\psi_n(x) = \sum_{i=1}^{n} a_i^n x_i$ and the coefficient a_i^n may be computed from the equation

$$(x - \sum_{i=1}^{n} a_i^n x_i, x_k) = 0 \qquad k = 1, 2, \cdots, n$$

or

$$(x, x_k) = \sum_{i=1}^{n} a_i^n (x_i, x_k) \qquad \text{see [5]}.$$

We introduce the following notation

$$(x_i, x_k) = c_{ik}$$

$$C_n = (c_{ik}) \quad 1 \leqq i \leqq n \quad 1 \leqq k \leqq n$$

$$((x, x_1), (x, x_2), \cdots, (x, x_n)) = (\gamma)_n$$

$$(a_1^n, a_2^n, \cdots, a_n^n) = (a)_n$$

Then

$$(\gamma)_n=(a)_n C_n \quad \text{or} \quad (a)_n=(\gamma)_n C_n^{-1}$$

since C_n^{-1} exists. Now

$$\|\sum_{i=1}^{j} a_i^n x_i\|^2=\sum_{i,k=1}^{j} a_i^n \overline{a_k^n} c_{ik}=(a)_n E_j^n C_n E_j^n (a)_n^*$$

where E_j^n is the matrix $(e_{l,m})$ with

$$e_{l,m}\begin{cases}1 & l=m\leqq j\\ 0 & \text{otherwise}\end{cases}$$

$C_n^*=C_n$ and $(a)_n=(\gamma)_n C_n^{-1}$ hence

$$\|\sum_{j=1}^{j} a_i^n x_i\|^2=(\gamma)_n C_n^{-1} E_j^n C_n E_j^n C_n^{-1}(\gamma)_n^*$$

Orthogonalizing the sequence $\{x_n\}$ by Schmidt's process we get the sequence $\{y_n\}$ with

$$x_1=y_1 \quad x_k=\sum_{\alpha=1}^{k} d_{k,\alpha} y_\alpha$$

where

$$d_{k,\alpha}=\begin{cases}(x_k,\ y_\alpha) & \alpha\leqq k\\ 0 & \alpha>k\end{cases}$$

and $d_{11}=1$ see [3].

Let D_n denote the triangular matrix

$$(d_{k,\alpha}) \quad 1\leqq\alpha\leqq n \quad 1\leqq k\leqq n\,.$$

$$(x_i,\ x_j)=\sum_{\alpha} d_{i,\alpha}\overline{d_{j,\alpha}} \text{ or } C_n=D_n D_n^*\,.$$

The matrix D_n can be computed from this relation.

Let
$$(\delta)_n=((x,\ y_1),\ (x,\ y_2),\ \cdots,\ (x,\ y_n))$$

$$(x,\ x_k)=\sum_{\alpha=1}^{k}(x,\ y_\alpha)\overline{d}_{k,\alpha} \text{ or } (\gamma)_n=(\delta)_n D_n^*$$

and hence

$$\|\sum_{i=1}^{j} a_i^n x_i\|^2=(\gamma)_n C_n^{-1} E_j^n C_n E_j^n C_n^{-1}(\gamma)_n^*=(\delta)_n D_n^* C_n^{-1} E_j^n C_n E_j^n C_n^{-1} D_n(\delta)_n^*$$

$$\cdot=(\delta)_n(D_n^{-1}E_n^j D_n)(D_n^* E_j^n (D_n^*)^{-1})(\delta)_n^*$$

Let $A_j^n=D_n^{-1}E_j^n D_n$ then

$$|||\varphi_n(x)||| = \max \{(\delta)_n A_j^n (A_j^n)^* (\delta)_n^* \,|\, 1 \leqq j \leqq n\}$$

The triangular matrix A_j^n is an operator defined on the Hilbert space. If

$$x = \sum_{i=1}^{\infty} \delta_i y_i$$

then

$$A_j^n(x) = (\delta_1, \cdots, \delta_n) A_j^n .$$

By Theorem 4 and the above computation the system is regular if and only if for each x

$$\sup \{ || A_j^n(x) || \,|\, 1 \leqq j \leqq n \;\; n=12 \cdots \} < \infty$$

or by the uniform boundedness theorem.

THEOREM 5. *The system is regular if and only if the double sequence* $\{||A_j^n||\}$ *is bounded, or, in other words, if and only if the set of characteristic roots of* $A_j^n(A_n^n)^*$ *is bounded.*

We shall use Theorem 3 to derive the following theorems.

THEOREM 6. *The system* (x_i, f_i) *is regular and* $\sum_{i=1}^{\infty} |f_i(x)|^2 < \infty$ *if and only if for every* $x \in B$ *there exists a real number* $\alpha = \alpha(x)$ *such that*

$$2\Re \{ \sum_{i=1}^{n} \sum_{j=1}^{i-1} f_i(x) \overline{f_j(x)} c_{ij} \} > \alpha \tag{1}$$

Proof. If the system is regular and

$$\sum_{i=1}^{\infty} | f_i(x) |^2 < \infty \text{ then}$$

$$\begin{aligned} ||x||^2 - \sum | f_i(x) |^2 &= \sum_{i \neq j} f_i(x) \overline{f_j(x)} c_{ij} \\ &= 2\Re \{ \sum_{j<i} f_i(x) \overline{f_j(x)} c_{ij} \} \end{aligned}$$

Therefore the necessity of condition (1) is verified. Assume that condition (1) is satisfied then

$$\begin{aligned} || \varphi_{n+p}(x) ||^2 &= \sum_{ij=1}^{n+p} f_i(x) \overline{f_j(x)} c_{ij} \\ &= || \varphi_n(x) ||^2 + \sum_{i=n+1}^{n+p} | f_i(x) |^2 + 2\Re \{ \sum_{i=n+1}^{n+p} \sum_{j=1}^{i-1} f_i(x) \overline{f_j(x)} c_{ij} \} \\ &\geqq || \varphi_n(x) ||^2 + 2\alpha \end{aligned}$$

Therefore $\sup_n \|\varphi_n(x)\| = \infty$ implies $\lim \|\varphi_n(x)\| = \infty$. According to Theorem 3 the system is regular.
Moreover

$$\sum_{i=1}^{n} |f_i(x)|^2 = \|\varphi_n(x)\|^2 - 2\Re\{\sum_{j<i\leqq n} c_{ij} f_i(x) f_j(x)\}$$
$$\leqq |||x|||^2 - \alpha < \infty .$$

An immediate consequence is the following. The system is regular if $\sum_{i\neq j} |c_{ij}| < \infty$ and the sequence $\{\|f_i\|\}$ is bounded.

Professor R. C. James called my attention to the fact that this may be proved directly and without the assumption of boundedness of the sequence $\{\|f_i\|\}$ as follows. We may assume without loss of generality that $\sum_{i\neq j} |c_{ij}| = r < 1$

$$|\sum_{i\neq j}^{n} a_i \bar{a}_j c_{ij}| \leqq \max |a_i \bar{a}_j| \cdot r \leqq \sum_{i=1}^{n} |a_i|^2 r$$

$$\|\sum_{i=1}^{n} a_i x_i\|^2 = \sum_{i=1}^{n} |a_i|^2 + \sum_{i\neq j}^{n} a_i \bar{a}_j c_{ij} \leqq 2 \sum_{i=1}^{n} |a_i|^2$$

Hence

$$\|\sum_{i=1}^{n+p} a_i x_i\|^2 = \sum_{i=1}^{n+p} |a_i|^2 + \sum_{i\neq j}^{n+p} a_i \bar{a}_j c_{ij}$$

$$\geqq \sum_{i=1}^{n+p} |a_i|^2 (1-r) \geqq \sum_{i=1}^{n} |a_i|^2 (1-r)$$

$$\geqq \frac{1-\mathrm{r}}{2} \left\|\sum_{i=1}^{n} a_i x_i\right\|^2$$

and by [4] the system is regular.

Using the same method as in Theorem 6 we arrive at the following.

THEOREM 7. *The system is regular if and only if for each* x

$$\inf_{n,p} \Re\{\sum_{i=1}^{n} \sum_{j=n+1}^{p} f_i(x) f_j(x) c_{ij}\} > -\infty \tag{2}$$

Proof.

$$\|\sum_{i=1}^{n+p} f_i(x) x_i\|^2 = \|\sum_{i=1}^{n} f_i(x) x_i\|^2 + \|\sum_{i=n+1}^{n+p} f_i(x) x_i\|^2$$
$$+ 2\Re\{(\sum_{i=1}^{n} f_i(x) x_i , \sum_{j=n+1}^{n+p} f_j(x) x_j)\} .$$

If condition (2) is satisfied then according to Theorem 3 the system is regular. If the system is regular then

$$|(\sum_{i=1}^{n} f_i(x)x_i\,,\ \sum_{j=n+1}^{n+p} f_j(x)x_j)| \leqq 2\,|||x|||^2.$$

As a simple application we note the following.

If $c_{ij}=0$ when $|i-j|>N$ then the system is regular if and only if the sequence $\{||f_i||\}$ is bounded.

References

1. R. Baire, *Leçons sur les fonctions discontinues*, Paris, Gauthier Villars (1905).
2. S. Banach, *Théorie des opérations linéaires*, Warszawa (1932).
3. Grunblum and Gourevitch, *Sur une propriété de la base dans l'espace de Hilbert*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **30** (1941), 289–291.
4. M. M. Grunblum, *Sur la théorie des systèmes biorthogonaux*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **55** (1947), 287–290.
5. Kaczmarz and Steinhaus, *Theorie der Orthogonalreihen*, Warszawa (1935) Manografje Matematyczne to VI.

Yale University

A THEOREM ON FLOWS IN NETWORKS

DAVID GALE

1. Introduction. The theorem to be proved in this note is a generalization of a well-known combinatorial theorem of P. Hall, [4].

HALL'S THEOREM. *Let $S_1, S_2, \cdots, S_n$ be subsets of a set X. Then a necessary and sufficient condition that there exist distinct elements $x_1, \cdots, x_n$, such that $x_i \in S_i$ is that the union of every k sets from among the S_i contain at least k elements.*

The result has a simple interpretation in terms of transportation networks. A certain article is produced at a set X of origins, and is demanded at n destinations $y_1, \cdots, y_n$. Certain of the origins x are "connected" to certain of the destinations y making it possible to ship one article from x to y.

PROBLEM. *Under what conditions is it possible to ship articles to all the destinations y?*

An obvious reinterpretation of Hall's theorem shows that this is possible if and only if every k of the destinations are connected to at least k origins.

We shall now give a verbal statement of the generalization to be proved. A more formal statement will be given in the next section.

Let N be an arbitrary network or graph. To each node x of N corresponds a real number $d(x)$, where $|d(x)|$ is to be thought of as the demand for or the supply of some good at x according as $d(x)$ is positive or negative. To each edge (x, y) corresponds a nonnegative real number $c(x, y)$, the capacity of this edge, which assigns an upper bound to the possible flow from x to y.

The demands $d(x)$ are called *feasible* if there exists a flow in the network such that the flow along each edge is no greater than its capacity, and the net flow into (out of) each node is at least (at most) equal to the demand (supply) at that node.

An obviously necessary condition for the demands $d(x)$ to be feasible is the following.

For every collection S of nodes the sum of the demands at the nodes

Received September 24, 1956. The results of this paper were discovered while the author was working as a consultant for the RAND Corporation. A later revision was partially supported by an O. N. R. contract.

of S must not exceed the sum of the capacities of the edges leading into S.

If this condition were not satisfied it would clearly be impossible to satisfy the aggregate demand of the subset S. The principal theorem of this paper shows that conversely, if the above condition is satisfied, then the demands $d(x)$ are feasible.

Hall's theorem drops out as a special case of this result if one applies it to the particular network described in the paragraph above and makes use of the known fact (see [1]) that transportation problems of this type with integral constraints have integral solutions. However, the simple inductive argument which works in [4] does not seem to generalize to yield a proof of our theorem. Our approach is in fact quite different and is based on the "minimum cut" theorem of Ford and Fulkerson, [2], [1].

In the next section we give a formal statement of the problem and prove the principal theorem. The final section is devoted to the treatment of a special case for which the "feasibility criterion" yields a very simple method for computing solutions.

2. The principal theorem.

We proceed to define in a more formal manner the objects to be discussed.

DEFINITIONS. A *network* $[N,c]$ consists of a finite set of *nodes* N and a *capacity function* c on $N \times N$ where $c(x, y)$ is a nonnegative real number or plus infinity.

A *flow* f on $[N, c]$ is a function f on $N \times N$ such that

$$f(x, y)+f(y, x)=0 , \tag{1}$$

$$f(x, y) \leqq c(x, y) \qquad \text{for all } x,\ y \in N . \tag{2}$$

A *demand* d on $[N, c]$ is simply a real valued function on N.

Note that we do not require the function c to be symmetric, thus the maximum allowable flow from x to y need not be the same as that from y to x. Condition (1) above corresponds to the usual convention that the net flow from x to y is the negative of the net flow from y to x.

We shall save writing many summation symbols in what follows by adopting the following convenient notation.

NOTATION. If S is a subset of N and d a function on N, we write

$$d(S)=\sum_{x \in S} d(x) .$$

If S and T are subsets of N and f a function on $N \times N$ we write

$$f(S, T) = \sum_{x \in S,\, y \in T} f(x, y) .$$

From these definitions it follows at once that if U and V are disjoint subsets of N then

$$d(U \cup V) = d(U) + d(V) \tag{3}$$

$$f(S, U \cup V) = f(S, U) + f(S, V) .$$

In particular, denoting the complement of S by S' we have,

$$f(N, T) = f(S, T) + f(S', T) \qquad \text{for all } S \subset N .$$

In this notation (1) and (2) are clearly equivalent to

$$f(A, A) = 0; \tag{1'}$$

and

$$f(A, B) \leqq c(A, B) \qquad \text{for all } A, B \subset N. \tag{2'}$$

The above notation is natural to our problem, for if d is a demand function then $d(S)$ is simply the aggregate demand of the set S, and if f is a flow then $f(S, T)$ represents the net flow from S into T.

DEFINITION. A demand d is called *feasible* if there exists a flow f such that

$$f(N, x) \geqq d(x) \qquad \text{for } x \in N. \tag{4}$$

This condition states that the flow into each node must be at least equal to the demand at that node. However (1) and (4) together imply

$$f(x, N) \leqq -d(x)$$

so that we are also requiring the flow out of each node to be at most equal to the supply at that node (recalling that a negative demand represents a supply).

Finally we note that from (3) it follows that (4) is equivalent to

$$f(N, S) \geqq d(S) \qquad \text{for all } S \subset N. \tag{4'}$$

We can now give a simple statement of our main result.

FEASIBILITY THEOREM. *The demand d is feasible if and only if for every subset $S \subset N$*

$$d(S') \leqq c(S, S') . \tag{5}$$

Proof. The necessity of (5) is obvious, for if d is feasible then there is a flow f such that

$$d(S') \leqq f(N, S') = f(S, S') + f(S', S') = f(S, S') \leqq c(S, S') .$$

The proof of sufficiency depends on the "minimum cut theorem" of Ford and Fulkerson, which we shall now state and prove in our own formulation. While our proof is little more than a translation of the above authors' second proof [3] into our notation, we record it here, nevertheless, both for the sake of completeness and because it is substantially shorter than any proof published heretofore.

DEFINITION. Let $[N, c]$ be a network and let s and s' be two distinguished nodes (s=source, s'=sink). A *flow from s to s'* is a flow such that

$$f(N, x) = 0 \qquad \text{for } x \neq s,\ x \neq s' . \tag{6}$$

Let F denote the set of all flows from s to s'.

A *cut* (S, S') of N with respect to s and s' is a partition of N into sets S and S' such that $s \in S$, $s' \in S'$.

Let Q denote the set of all such cuts.

MINIMUM CUT THEOREM. *For any network* $[N, c]$

$$\max_F f(s, N) = \min_Q c(S, S') ,$$

Proof. First note that for any flow $f \in F$ and cut $(S, S') \in Q$ we have

$$\begin{aligned} f(s, N) &= f(s, N) + \sum_{x \in S - s} f(x, N) = f(s, N) + f(S - s, N) \\ &= f(S, N) = f(S, S) + f(S, S') = f(S, S') \leqq c(S, S') . \end{aligned} \tag{7}$$

Hence, it remains only to show that equality is attained in (7) for some flow and cut.

Let $\bar{f} \in F$ be a flow such that $\bar{f}(s, N)$ is a maximum. Let S consist of s and all nodes x such that there exists a *chain* $\sigma = (x_0, x_1, \cdots, x_n)$ of distinct nodes with $x_0 = s$, $x_n = x$ and $c(x_{i-1}, x_i) - \bar{f}(x_{i-1}, x_i) > 0$, $i = 1, \cdots, n$. Now s' is not in S, for, if it were, there would be a chain σ as above with $x = s'$. But then letting

$$\mu = \min\,[c(x_{i-1}, x_i) - \bar{f}(x_{i-1}, x_i)] ,$$

one could superimpose a flow of μ along the chain σ on top of the flow $\bar{f}$, contradicting the maximality of $\bar{f}$.

The above argument shows that (S, S') is a cut, and we conclude the proof by observing that $\bar{f}(s, N)=c(S, S')$, for if not, then from (7), $\bar{f}(S, S')<c(S, S')$, hence for some $x \in S$ and $y \in S'$ we would have $c(x, y)-\bar{f}(x, y)>0$, but since $x \in S$ there is a chain $\sigma=(s, x_1, \cdots, x)$ which could be extended to a chain $\sigma'=(s, x_1, \cdots, x, y)$, contrary to the fact that $y \in S'$. This completes the proof.

Proof of feasibility theorem. Consider a new network $[\bar{N}, \bar{c}]$ where $\bar{N}$ consists of N plus two additional nodes s and s'. Let $U \subset N$ be all nodes x such that $d(x) \leqq 0$. Then $\bar{c}$ is defined by the rules

$$\begin{aligned}
&\bar{c}(x, y)=c(x, y) && \text{for } x, y \in N,\\
&\bar{c}(s, x)=-d(x) && \text{for } x \in U,\\
&\bar{c}(x, s')=d(x) && \text{for } x \in U',\\
&\bar{c}(x, y)=0 && \text{otherwise.}
\end{aligned}$$

We now assert that the cut $(\bar{N}-s', s')$ is a minimal cut of $[\bar{N}, \bar{c}]$, for let $\bar{S}$, and $\bar{S}'$ be any cut of $[\bar{N}, \bar{c}]$ and let $S=\bar{S}-s$, $S'=\bar{S}'-s'$. From the definition above we have

$$\begin{aligned}
\bar{c}(\bar{S}, \bar{S}')&=c(S, S')+\bar{c}(s, S')+\bar{c}(S, s')\\
&=c(S, S')-d(S' \cap U)+d(S \cap U') ,\\
\bar{c}(\bar{N}-s', s')&=d(U')=d(S' \cap U')+d(S \cap U') ;
\end{aligned}$$

and subtracting we get

$$\begin{aligned}
\bar{c}(\bar{N}-s', s')-\bar{c}(\bar{S}, \bar{S}')&=d(S' \cap U')+d(S' \cap U)-c(S, S')\\
&=d(S')-c(S, S') \leqq 0 ,
\end{aligned}$$

the last inequality being the hypothesis (5), and the assertion is proved.

Now, from the Minimum Cut Theorem, there is a flow $\bar{f}$ from s to s' on $[\bar{N}, \bar{c}]$ such that

$$\bar{f}(\bar{N}-s', s')=\bar{c}(\bar{N}-s', s')=d(U') ,$$

hence

$$\bar{f}(x, s')=d(x) \qquad \text{for all } x \in U'. \tag{8}$$

Let f be $\bar{f}$ restricted to $N \times N$. Then f is clearly a flow and it remains to show that f satisfies (4). If $x \in U'$ then

$$0=\bar{f}(x, \bar{N})=f(x, N)+\bar{f}(x, s')=f(x, N)+d(x) ,$$

hence

$$f(N,\ x)=d(x)\ . \tag{9}$$

If $x \in U$ then

$$0=\overline{f}(\overline{N},\ x)=f(N,\ x)+\overline{f}(s,\ x)\leqq f(N,\ x)+\overline{c}(s,\ x)=f(N,\ x)-d(x)\ ,$$

so

$$f(N,\ x)\geqq d(x)\ , \tag{10}$$

and (9) and (10) together show that f satisfies (4), completing the proof.

REMARK. We wish to call attention to the following important fact. We have at no point in what has been said thus far made use of the assumption that the functions d, c and f were real valued. In fact, all definitions and proofs go through verbatim if the real numbers are replaced by any ordered Abelian group, in particular, the group of integers. One useful consequence of this remark is the fact that if a network with integer valued demand and capacity functions admits a feasible flow then this flow may also be chosen to be integer valued. We shall make use of this fact in the next section.

There is a second formulation of the Feasibility Theorem which is sometimes convenient. In the network $[N,\ c]$ let U be as above the set of nodes x such that $d(x)\leqq 0$.

THEOREM. *The demand d is feasible if and only if for every set $Y \subset U'$ there exists a flow f_Y such that*

$$f_Y(N,\ x)\geqq d(x) \qquad \text{for } x \in U \tag{11}$$

$$f_Y(N,\ Y)\geqq d(Y)\ . \tag{12}$$

Proof. The necessity is obvious. To prove sufficiency we show that (11) and (12) imply (5).

Let $(S,\ S')$ be a partition of N and let $X=U \cap S$, $X'=U \cap S'$, $Y=U' \cap S$, $Y'=U' \cap S'$. Then from (11) there exists $f_{Y'}$ such that

$$d(X')\leqq f_{Y'}(N,\ X')=f_{Y'}(X \cup Y,\ X')+f_{Y'}(Y',\ X')\ ,$$

and from (12),

$$d(Y')\leqq f_{Y'}(N,\ Y')=f_{Y'}(X \cup Y,\ Y')+f_{Y'}(X',\ Y')\ .$$

Adding these inequalities we get

$$d(S')=d(X')+d(Y')=f_{Y'}(X\cup Y,\ X')+f_{Y'}(X\cup Y,\ Y')$$
$$=f_{Y'}(X\cup Y,\ X'\cup Y')=f(S,\ S')\leqq c(S,\ S')\ ,$$

which is exactly (5).

3. An example. As an illustration of the feasibility theorem, consider the following problem.

(I). *Let $a_1, \cdots, a_m$ and $b_1, \cdots, b_n$ be two sets of positive integers. Under what conditions can one find integers $\alpha_{ij}=0$ or 1, such that*

$$\sum_{i=1}^{m}\alpha_{ij}\geqq b_j$$

and

$$\sum_{j=1}^{n}\alpha_{ij}\leqq a_i\ ,$$

for all i and j?

As a concrete illustration, suppose n families are going on a picnic in m busses, where the jth family has b_j members and the ith bus has a_i seats. When is it possible to seat all passengers in such a way that no two members of the same family are in the same bus?

In the case $\sum a_i=\sum b_i$ the problem becomes that of filling an $m\times n$ matrix M with zeros and ones so that the rows and columns shall have prescribed sums.

The feasibility theorem gives a simple necessary and sufficient condition for the problem to have a solution. In order to state if we need the following.

DEFINITION. Let $\{a_i\}$ be a nonincreasing sequence of nonnegative integers $a_1, a_2, \cdots$, such that all but a finite number of the a_i are zero. Let

$$S_j=\{a_i|a_i\geqq j\}$$

where j is a positive integer and let s_j be the number of elements in S_j. The sequence of numbers $\{s_j\}$ clearly satisfies the same conditions as the sequence $\{a_i\}$; it is called the *dual sequence* of the sequence $\{a_i\}$ and is denoted by $\{a_i\}^*$.

It is clear that $\{a_i\}^*$ determines $\{a_i\}$ since the integer a_i occurs exactly $s_{a_i}-s_{a_i+1}$ times in $\{a_i\}$. Actually the correspondence between $\{a_i\}$ and $\{a_i\}^*$ is completely dual in the following sense.

THEOREM. $\{a_i\}^{**}=\{a_i\}$.

This result will not be needed in the sequel and its proof is left as an exercise. However, its validity can be made quite obvious by means of a simple pictorial representation.

Let each number a_i be represented by a row of dots, and write these rows in a vertical array so that a_{i+1} lies under a_i, thus:

$$\begin{array}{l} a_1 \cdot\cdot\cdot\cdot\cdot\cdot \\ a_2 \cdot\cdot\cdot\cdot\cdot \\ a_3 \cdot\cdot\cdot\cdot\cdot \\ a_4 \cdot\cdot\cdot \\ a_5 \cdot \end{array}$$

It is then clear that the dual number s_j is simply the number of dots in the jth column of the array.

We can now give the criterion for the feasibility of Problem I. Henceforth for convenience we shall assume the numbers a_i and b_j are indexed in decreasing order, and shall define $a_i=0$ for $i>m$, $b_j=0$ for $j>n$.

THEOREM. *Let* $\{s_j\}=\{a_i\}^*$. *Then Problem* 1 *is feasible if and only if*

$$\sum_{j=1}^{k} b_j \leqq \sum_{j=1}^{k} s_j \,, \qquad \textit{for all integers } k \,.$$

Proof. We may interpret (I) as a flow problem. Let N be a network consisting of $m+n$ nodes $x_1, \cdots, x_m$ and $y_1, \cdots, y_n$, and let $c(x_i, y_i)=1$ for all i and j, $c=0$ otherwise. Let $d(x_i)=-a_i$ and $d(y_j)=b_j$. One easily verifies that the feasibility of (I) is equivalent to the feasibility of the demand d.

We shall show that d is feasible by applying the second theorem of the previous section. Let Y be a subset of k nodes y_j, say $Y=\{y_{j_1}, \cdots, y_{j_k}\}$. We now compute the maximum possible flow into Y. Because all capacities are unity this maximal flow f_Y is achieved by shipping as much as possible from each node x_i into the set Y. Thus, the flow from x_i to Y is $\min[a_i, k]$ and the total flow into Y is

$$f_Y(N, Y)=\sum_{i=1}^{m} \min[a_i, k] \,.$$

We now assert

$$\sum_{i=1}^{m} \min[a_i, k]=\sum_{j=1}^{k} s_j \,, \tag{13}$$

which is proved by induction on k. It is clear from the definition that

$$\sum_{i=1}^{m} \min [a_i,\ 1]=m=s_1 .$$

Now

$$\min [a_i\ ,\ k+1]=\begin{cases}\min [a_i,\ k] & \text{for } a_i \leqq k\\ \min [a_i,\ k]+1 & \text{for } a_i \geqq k+1, \text{ or } a_i \in S_{k+1},\end{cases}$$

hence,

$$\sum_{i=1}^{m} \min [a_i,\ k+1]=\sum_{i=1}^{m} \min [a_i,\ k]+s_{k+1} ,$$

and (13) follows from the induction hypothesis.

The second feasibility theorem now states that the problem is feasible if and only if

$$\sum_{r=1}^{k} b_{j_r} \leqq \sum_{j=1}^{k} s_j ,$$

and since the b_j are indexed in decreasing order, the conclusion of the theorem follows.

It is interesting that for this particular problem there is a simple "n-step" method for actually filling out the matrix of α_{ij}'s. Such procedures are sufficiently rare in programming theory so that it seems worth while to present it here.

The procedure is the following: If the problem is feasible then $b_1 \leqq s_1$ and hence $a_1, \cdots, a_{b_1} \geqq 1$ (recall that the a_i's are indexed in descending order). Let $\alpha_{i1}=1$ for $i \leqq b_1$, $\alpha_{i1}=0$ for $i > b_1$. Now consider the new problem, (I)$'$, with the matrix M' having m rows and $n-1$ columns, $j=2, \cdots, n$, with $a_i'=a_i-\alpha_{i1}$ and $b_j'=b_j$. We assert that (I)$'$ is again feasible so that by repeating the process we will eventually fill out the whole matrix.

To show that (I)$'$ is feasible we must prove, for any k,

$$\sum_{j=2}^{k+1} b_j \leqq \sum_{j=1}^{k} s_j'=\sum_{i=1}^{m} \min [s_i',\ k] ,$$

where $\{s_i'\}$ is the dual sequence to $\{a_i'\}$. The expression on the right can be rewritten

$$\sum_{i=1}^{m} \min [a_i',\ k]=\sum_{i=1}^{b_1} \min [a_i-1,\ k]+ \sum_{i=b_1+1}^{m} \min [a_i,\ k] .$$

We must now consider two cases.

Case 1. $s_{k+1} \geqq b_1$. Then $a_i-1 \geqq k$ for $i \leqq b_1$ and hence $\min [a_i-1,\ k]$

$=k=\min[a_i, k]$, so that we get

$$\sum_{j=1}^{k} s'_j = \sum_{i=1}^{m} \min[a'_i, k] = \sum_{i=1}^{m} \min[a_i, k] = \sum_{j=1}^{k} s_j \geqq \sum_{j=1}^{k} b_j \geqq \sum_{j=2}^{k+1} b_j .$$

Case 2. $s_{k+1} < b_1$. Then for $i \leqq s_{k+1}$, $a_i \geqq k+1$ so $a_i - 1 \geqq k$ and $\min[a_i - 1, k] = k = \min[a_1, k]$. For $s_{k+1} < i \leqq b_1$, $a_i \leqq k$, so $\min[a_i - 1, k] = \min[a_i, k] - 1$, hence,

$$\sum_{i=1}^{m} \min[a'_i, k] = \sum_{i=1}^{m} \min[a_i, k] - b_1 + s_{k+1} = \sum_{j=1}^{k+1} s_j - b_1 \geqq \sum_{i=2}^{k+1} b_j ,$$

since

$$\sum_{j=1}^{k+1} s_j \geqq \sum_{j=1}^{k+1} b_j$$

by the feasibility condition. The proof is now complete.

In terms of the picnic problem, the n families should be seated in n stages according to the following simple rule: at each stage distribute the largest unseated family among those busses having the greatest number of vacant seats.

References

1. G. B. Dantzig and D. R. Fulkerson, *On the max-flow min-cut theorem of networks*, Ann. of Math. Study No. **38**, *Contributions to linear inequalities and related topics*, edited by H. W. Kuhn and A. W. Tucker, 215–221.

2. L. R. Ford, Jr., and D. R. Fulkerson, *Maximal flow through a network*, Canad. J. Math. **8** (1956), 399–404.

3. ———, *A simple algorithm for finding maximal network flows aud an application to the Hitchcock problem*, Canad. J. Math. **9** (1957), 210–218.

4. P. Hall. *On Representatives of Subsets*, J. London Math. Soc., **10** (1935), 26–30.

THE RAND CORPORATION AND
BROWN UNIVERSITY

ON SPACES WITH A MULTIPLICATION

I. M. JAMES

Introduction. This paper is divided into three parts, together with an appendix.

In the first part we discuss the homotopy theory of mappings into a space with a multiplication, such as a topological group. These spaces are more general than the group-like spaces considered by G. W. Whitehead in [6], and our treatment, as far as it goes, is quite different from his. In the second and third parts we apply the theory to the reduced product spaces of [2] and the loop-spaces of [4]. We arrive at useful new definitions of the Hopf construction and the Whitehead product, such that the relations between them are plainly exhibited. In many respects this completes the theory of the suspension triad as developed in [3].

PART I

HOMOTOPY THEORY OF A SPACE WITH A MULTIPLICATION

1. Preliminary notions. Let S^r denote a topological r-sphere, with basepoint[1] e, where $r \geqq 1$. Let Z be a space with a basepoint, and let $h: S^p \times S^q \to Z$ be a map, where $p, q \geqq 1$. By the *sections* of h we means the maps $f: S^p \to Z$, $g: S^q \to Z$ which are defined by

$$f(x)=h(x, e)\ , \qquad g(y)=h(e, y) \qquad x \in S^p\ , \quad y \in S^q\ .$$

If $h': S^p \times S^q \to Z$ is another map with the same sections as h, then the two maps agree on the set of axes

$$\Sigma = S^p \times e \cup e \times S^q\ ,$$

and since the complement of Σ in $S^p \times S^q$ is an open $(p+q)$-cell the separation element $d(h, h') \in \pi_{p+q}(Z)$ is defined, as in § 10 below. Of course

$$d(h, h)=0\ . \tag{1.1}$$

In particular, let Z be a space with a multiplication; that is to say, there is a continuous product $x \cdot y \in Z$, where $x, y \in Z$, such that $x \cdot z^0 = x$ and $z^0 \cdot y = y$, where z^0 is the basepoint in Z. Let h be as before, and

Received May, 22, 1956.

[1] When we consider a map, or homotopy, of one space into another it is always assumed that the image of the basepoint in the one is the basepoint in the other.

let h' be defined by

$$h'(x, y) = f(x) \cdot g(y) \qquad x \in S^p, \quad y \in S^q,$$

where f, g are the sections of h. Then h' has the same sections as h, and we define

$$(1.2) \qquad \delta(h) = d(h', h) \in \pi_{p+q}(Z).$$

Notice that if $k: S^p \times S^q \to Z$ is another map with the same sections then $k' = h'$ and so

$$(1.3) \qquad \delta(k) = \delta(h) + d(h, k),$$

by the addition formula for separation elements ((10.4) below).

Let $w: Z \to Z'$ be a map, where Z' is a space with a multiplication. We say that w is *multiplicative* if

$$w(x \cdot y) = w(x) \cdot w(y) \qquad x, y \in Z.$$

In that case we have (cf. (10.8))

$$(1.4) \qquad \delta(wh) = w_* \delta(h),$$

where $w_*: \pi_{p+q}(Z) \to \pi_{p+q}(Z')$ denotes the homomorphism which is induced by w.

2. The pairing of $\pi_p(Z)$ with $\pi_q(Z)$ to $\pi_{p+q}(Z)$. Let Z be a space with a multiplication, and let $p, q \geqq 1$. With each pair of elements $\alpha \in \pi_p(Z)$, $\beta \in \pi_q(Z)$ we associate an element $\langle \alpha, \beta \rangle \in \pi_{p+q}(Z)$, as follows. Let $f: S^p \to Z$, $g: S^q \to Z$ be maps which represent α, β, respectively. Let h, $k: S^p \times S^q \to Z$ be the maps which are defined by

$$h(x, y) = f(x) \cdot g(y), \qquad k(x, y) = g(y) \cdot f(x),$$

where $x \in S^p$, $y \in S^q$. Then h and k have the same sections, and we write

$$(2.1) \qquad \langle \alpha, \beta \rangle = d(h, k).$$

We have at once (cf. (10.8))

THEOREM (2.2). *Let $\alpha \in \pi_p(Z)$, $\beta \in \pi_q(Z)$. Let $w: Z \to Z'$ be a multiplicative map. Then*

$$w_* \langle \alpha, \beta \rangle = \langle w_*(\alpha), w_*(\beta) \rangle,$$

where $w_*: \pi_r(Z) \to \pi_r(Z')$ denotes the homomorphism induced by w.

The *type* of a map $h: S^p \times S^q \to Z$ is the pair of elements (α, β), where $\alpha \in \pi_p(Z)$, $\beta \in \pi_q(Z)$ are the homotopy classes of the sections of

h. We prove

THEOREM (2.3). *Let Z be a space with a multiplication, and let*

$$S^p \times S^q \xrightarrow{h} Z \xleftarrow{k} S^q \times S^p$$

be a pair of maps such that $h(x, y)=k(y, x)$, where $x \in S^p$, $y \in S^q$. Let (α, β) be the type of h, where $\alpha \in \pi_p(Z)$, $\beta \in \pi_q(Z)$. Then

$$\langle\alpha, \beta\rangle=\delta(h)-(-1)^{pq}\delta(k) .$$

Proof. Let $f: S^p \rightarrow Z$, $g: S^q \rightarrow Z$ be the sections of h, and let

$$S^p \times S^q \xrightarrow{h'} Z \xleftarrow{k'} S^q \times S^p$$

be the maps which are defined by

$$h'(x, y)=f(x)\cdot g(y) , \qquad k'(y, x)=g(y)\cdot f(x) ,$$

where $x \in S^p$, $y \in S^q$. Then

$$\delta(h)=d(h', h) , \qquad \delta(k)=d(k', k) ,$$

by definition. Let $v: S^p \times S^q \rightarrow S^q \times S^p$ be the map which interchanges the factors. Then $d(k'v, kv)=(-1)^{pq}d(k', k)$, by (10.9), since v has degree $(-1)^{pq}$. Therefore

$$\delta(h)-(-1)^{pq}\delta(k)=d(h', h)-d(k'v, kv)=d(h', k'v) ,$$

by the addition formula for separation elements, since $h=kv$. However

$$k'v(x, y)=k'(y, x)=g(y)\cdot f(x) ,$$

if $x \in S^p$ and $y \in S^q$. Hence $d(h', k'v)=\langle\alpha, \beta\rangle$, by (2.1), since α, β are the homotopy classes of f, g, respectively. Therefore

$$\delta(h)-(-1)^{pq}\delta(k)=\langle\alpha, \beta\rangle ,$$

which proves (2.3).

If we interchange h and k in (2.3), we obtain that

$$\delta(k)-(-1)^{pq}\delta(h)=\langle\beta, \alpha\rangle ,$$

since k is of type (β, α). Hence, and since there exist maps of any given type, we obtain

COROLLARY (2.4). *Let $\alpha \in \pi_p(Z)$, $\beta \in \pi_q(Z)$. Then*

$$\langle\alpha, \beta\rangle=(-1)^{pq+1}\langle\beta, \alpha\rangle .$$

In the next section we shall prove that $\langle\alpha, \beta\rangle$ determines a bilinear pairing of $\pi_p(Z)$ with $\pi_q(Z)$ to $\pi_{p+q}(Z)$.

3. Products of maps. The proof of the following proposition is omitted, since it is the same as in the case of topological groups (see (16.9) of [5]).

THEOREM (3.1). *A space with a multiplication has a commutative fundamental group.*

Let Y be a space and let Z be a space with a multiplication. The product of two maps $u, v\colon Y \to Z$ is the map $u\cdot v\colon Y \to Z$ which is defined by

$$(u\cdot v)(y) = u(y)\cdot v(y) \qquad y \in Y .$$

In view of (3.1), we write $\pi_r(Z)$ additively even when $r=1$. The proof of the following proposition also is the same as in the case of topological groups (see (16.7) of [5]).

THEOREM (3.2). *Let $v, u\colon S^r \to Z$ be maps, where $r \geqq 1$ and Z is a space with a multiplication. Then the homotopy class of $u\cdot v$ is equal to the sum, in $\pi_r(Z)$, of the homotopy classes of u and v.*

The following lemma is an immediate consequence of (3.2) and the definition of separation elements.

LEMMA (3.3). *Let h, k, h', k' be four maps of $S^p \times S^q$ into Z such that h and k have the same sections, and h' and k' have the same sections. Then $h\cdot h$ and $k\cdot k'$ have the same sections, and their separation element is given by*

$$d(h\cdot h', k\cdot k') = d(h, k) + d(h', k') .$$

We use (3.3) to prove

THEOREM (3.4). *Let Z be a space with a multiplication. Let h, h' be maps of $S^p \times S^q$ into Z of type (α, β), (α', β'), respectively, where $\alpha, \alpha' \in \pi_p(Z)$ and $\beta, \beta' \in \pi_q(Z)$. Then*

$$\delta(h\cdot h') = \delta(h) + \delta(h') + \langle\alpha', \beta\rangle .$$

The relation we have to prove is invariant under homotopies of h and h'. Hence there is no real loss of generality if we assume that h and h' are such as to satisfy the following condition. Let (f, g), (f', g') be the sections of h, h', respectively, so that f and f' are maps of S^p, and g and g' are maps of S^q. We assume that f is constant over one hemisphere of S^p and that f' is constant over the other; similarly that

g is constant over one hemisphere of S^q and that g' is constant over the other. Then, if $x \in S^p$ and $y \in S^q$, the expressions

$$f(x)\cdot f'(x)\cdot g(y)\cdot g'(y)\ , \qquad f(x)\cdot g(y)\cdot f'(x)\cdot g'(y)$$

do not depend on the order in which the products are taken; it is as though the multiplication on Z were associative. This is expressed more concisely as follows. Let

$$S^p \xleftarrow{u} S^p \times S^q \xrightarrow{v} S^q$$

denote the canonical projections, and define four maps F, F', G, G' of $S^p \times S^q$ into Z by

$$F = fu\ , \quad F' = f'u\ ; \qquad G = gv\ , \quad G' = g'v\ .$$

Then the product maps $F\cdot F'\cdot G\cdot G'$ and $F\cdot G\cdot F'\cdot G'$ are well-defined.

After these preliminaries, we proceed to prove (3.4). Let $k = F\cdot G$, so that $\delta(h) = d(k, h)$, and let $k' = F'\cdot G'$, so that $\delta(h') = d(k', h')$. Then

$$(3.5) \qquad d(k\cdot k',\ h\cdot h') = \delta(h) + \delta(h')\ ,$$

by (3.3). Let $H = (F\cdot F')\cdot(G\cdot G')$. Then $\delta(h\cdot h') = d(H, h\cdot h')$, by definition, and hence

$$\delta(h\cdot h') = d(H,\ k\cdot k') + d(k\cdot k',\ h\cdot h')\ ,$$

by the addition formula for separation elements. Hence

$$(3.6) \qquad \delta(h\cdot h') = \delta(h) + \delta(h') + d(H,\ k\cdot k')\ ,$$

by (3.5). However,

$$\begin{aligned} \langle\alpha', \beta\rangle &= d(F'\cdot G,\ G\cdot F'), \text{ by definition}\ , \\ &= d(F, F) + d(F'\cdot G,\ G\cdot F') + d(G', G'), \text{ by (1.1)}\ , \\ &= d(F\cdot F'\cdot G,\ F\cdot G\cdot F') + d(G', G'), \text{ by (3.3)}\ , \\ &= d(F\cdot F'\cdot G\cdot G',\ F\cdot G\cdot F'\cdot G'), \text{ by (3.3)}\ , \\ &= d(H,\ k\cdot k')\ , \end{aligned}$$

by definition. Hence it follows from (3.6) that

$$\delta(h\cdot h') = \delta(h) + \delta(h') + \langle\alpha', \beta\rangle\ ,$$

which proves (3.4).

As an application of (3.4) we prove[2].

THEOREM (3.7). *Let Z be a space with a multiplication, and let $p, q \geqq 1$. Then the transformation $(\alpha, \beta) \to \langle\alpha, \beta\rangle$ determines a bilinear*

[2] This can also be proved directly.

pairing of $\pi_p(Z)$ *with* $\pi_q(Z)$ *to* $\pi_{p+q}(Z)$.

We first show that

$$\langle \alpha+\alpha', \beta\rangle = \langle \alpha, \beta\rangle + \langle \alpha', \beta\rangle, \tag{3.8}$$

where $\alpha, \alpha' \in \pi_p(Z)$, $\beta \in \pi_q(Z)$. For let $f, f' : S^p \to Z$ be maps which represent α, α', respectively, such that f is constant over one hemi-sphere of S^p and f' is constant over the other. Let $g: S^q \to Z$ represent β, and let h, h' be the maps of $S^p \times S^q$ into Z which are defined by

$$h(x, y) = g(y) \cdot f(x), \qquad h'(x, y) = f'(x),$$

where $x \in S^p$, $y \in S^q$. Then

$$(h \cdot h')(x, y) = (g(y) \cdot f(x)) \cdot f'(x) = g(y) \cdot (f(x) \cdot f'(x)),$$

and so $\delta(h \cdot h') = \langle \gamma, \beta\rangle$, by (2.1), where γ denotes the homotopy class of $f \cdot f'$. But $\gamma = \alpha + \alpha'$, by (3.2), and so

$$\begin{aligned} \langle \alpha+\alpha', \beta\rangle &= \delta(h \cdot h') \\ &= \delta(h) + \delta(h') + \langle \alpha', \beta\rangle, \text{ by (3.4)}, \\ &= \langle \alpha, \beta\rangle + \langle \alpha', \beta\rangle, \end{aligned}$$

since $\delta(h') = d(h', h') = 0$, by (1.1). This proves (3.8). Linearity on the right follows from (3.8) and (2.4). Hence the proof of (3.7) is complete.

Part II

Application to Reduced Product Complexes

4. The reduced product complex. Throughout this part of the paper, A will denote a countable CW-complex with precisely one 0-cell, say a^0. Let A_∞ denote the reduced product complex of A, as defined in [2]. We recall that A_∞ is a countable CW-complex which contains A as a subcomplex, and that A_∞ carries an associative multiplication with a^0 as unit element. Let I denote the interval $0 \leqq t \leqq 1$. Let $\hat{A}$ denote the suspension of A, that is the space which is obtained from the topological product $A \times I$ by identifying $A \times \dot{I} \cup a^0 \times I$ to a point. The points of $\hat{A}$ are represented by pairs (a, t), where $a \in A$ and $t \in I$, with the identification being tacitly understood. We also identify each point $a \in A$ with $(a, \frac{1}{2}) \in \hat{A}$, so that A is embedded in $\hat{A}$. The suspension triad of A is the triad

$$(\hat{A};\ C_+,\ C_-),$$

in which C_-, C_+ are the half-cones where $t \leqq \frac{1}{2}$, $t \geqq \frac{1}{2}$, respectively, so that

$$\hat{A}=C_+\cup C_- , \qquad A=C_+\cap C_- .$$

The relation between the reduced product complex and the suspension triad is expressed in the following commutative diagram, where ϕ denotes the canonical isomorphism which is defined in § 10 of [3].

$$
\begin{array}{ccccccccc}
\cdots \to & \pi_r(A) & \xrightarrow{j} & \pi_r(A_\infty) & \xrightarrow{k} & \pi_r(A_\infty, A) & \xrightarrow{\partial} & \pi_{r-1}(A) & \to \cdots \\
& \downarrow{\scriptstyle\phi} & & \downarrow{\scriptstyle\phi} & & \downarrow{\scriptstyle\phi} & & \downarrow{\scriptstyle\phi} & \\
\cdots \to & \pi_r(A) & \xrightarrow[E]{} & \pi_{r+1}(\hat{A}) & \xrightarrow[i]{} & \pi_{r+1}(\hat{A}\,;\,C_+,\,C_-) & \xrightarrow[\Delta]{} & \pi_{r-1}(A) & \to \cdots .
\end{array}
\tag{4.1}
$$

The top line of the diagram is part of the homotopy sequence of the pair (A_∞, A), so that j, k are injections, and ∂ is the boundary operator. The bottom line is part of the suspension sequence of A, as defined in [3], so that E is the suspension operator, i is the injection, and Δ is the repeated boundary operator. We recall from [3] that ϕ maps $\pi_r(A)$ identically, so that the commutativity of (4.1) is expressed by the following relations (cf. (10.2) of [3]).

$$
\begin{array}{rl}
\text{(a)} & \phi j=E , \\
\text{(b)} & i\phi=\phi k , \\
\text{(c)} & \Delta\phi=\partial .
\end{array}
\tag{4.2}
$$

Let B a countable CW-complex with precisely one 0-cell, say b^0, and let $f\colon A\to B$ be a map such that $f(a^0)=b^0$. Then the induced mapping $f_\infty\colon A_\infty\to B_\infty$, as defined in § 1 of [2], is multiplicative in the sense of § 1. Let $\hat{f}\colon \hat{A}\to\hat{B}$ denote the suspension of f, which is defined by

$$\hat{f}(a, t)=(f(a), t) \qquad a\in A , \quad t\in I .$$

Then $\hat{f}$ maps the suspension triad of A into the suspension triad of B, and hence induces a homomorphism of the suspension sequence of A into the suspension sequence of B. We denote this homomorphism by f_*, and we also denote by f_* the homomorphism of the homotopy sequence of (A_∞, A) into the homotopy sequence of (B_∞, B) which is induced by f_∞. By (10.5) of [3] these homomorphisms are related by

$$\phi f_*=f_*\phi . \tag{4.3}$$

5. The Hopf construction. Let A mean the same as in § 4, and let $p, q\geqq 1$. A pairing of $\pi_p(A)$ with $\pi_q(A)$ to $\pi_{p+q}(A_\infty, A)$ is defined as follows. Let γ denote the positive generator of the infinite cyclic group

$\pi_{p+q}(S^p \times S^q, \Sigma)$, where $\Sigma = S^p \times e \cup e \times S^q$ and the orientations are the same as in [3]. Let f, g be maps of S^p, S^q into A which represent $\alpha \in \pi_p(A)$, $\beta \in \pi_q(A)$, respectively. Let h: $(S^p \times S^q, \Sigma) \to (A_\infty, A)$ denote the map which is defined by

$$h(x, y) = f(x) \cdot g(y) \qquad x \in S^p\ , \quad y \in S^q\ .$$

Then we define $\alpha \times \beta = h_*(\gamma)$, where

$$h_*:\ \pi_{p+q}(S^p \times S^q, \Sigma) \to \pi_{p+q}(A_\infty, A)$$

denotes the homomorphism induced by h. We write

$$\text{(5.1)} \qquad \{\alpha, \beta\} = \phi(\alpha \times \beta) \in \pi_{p+q+1}(\hat{A}\ ;\ C_+, C_-)\ ,$$

and we refer to $\{\alpha, \beta\}$ as the triad Whitehead product of α and β, in accordance with (7.1) of [3].

Let us apply the theory of Part I to the space with multiplication A_∞. By taking representatives we obtain from (2.1) that

$$k\langle j(\alpha), j(\beta)\rangle = \alpha \times \beta - (-1)^{pq} \beta \times \alpha\ ,$$

and hence, by (4.2b) and (5.1), we have

$$\text{(5.2)} \qquad i\phi\langle j(\alpha), j(\beta)\rangle = \{\alpha, \beta\} - (-1)^{pq}\{\beta, \alpha\}\ .$$

Now suppose that there exists a map $h: S^p \times S^q \to A$ of type (α, β). Let h' denote the inclusion of h into A_∞. By taking representatives we obtain at once that $k\delta(h') = \alpha \times \beta$, and so we conclude from (4.2b) and (5.1) that

$$\text{(5.3)} \qquad i\phi\delta(h') = \{\alpha, \beta\}\ .$$

Recall that the Hopf construction, as defined in [3], assigns an element $c(h) \in \pi_{p+q+1}(\hat{A})$ to each map h: $S^p \times S^q \to A$, and is characterized uniquely by the following three properties. First, let h have type (α, β), where $\alpha \in \pi_p(A)$, $\beta \in \pi_q(A)$. Then

$$\text{(5.4)} \qquad ic(h) = \{\alpha, \beta\}\ .$$

Secondly, let B mean the same as in § 4, and let $f: A \to B$ be a map such that $f(a^0) = b^0$. Then

$$\text{(5.5)} \qquad c(fh) = f_* c(h)\ .$$

Thirdly, let $A = S^p \times S^q$, and let h be either of the projections

$$(x, y) \to (x, e)\ , \qquad (x, y) \to (e, y)\ ,$$

where $x \in S^p$, $y \in S^q$. Then

$$c(h)=0 . \tag{5.6}$$

The uniqueness of this characterization follows from (8.2), (8.3) and (8.4) of [3]. We use it to prove

THEOREM (5.7). *Let $h: S^p \times S^q \to A$ be a map, and let h' denote its inclusion into A_∞. Let $c(h)$ denote the element of $\pi_{p+q+1}(\hat{A})$ which is obtained from h by the Hopf construction, and let $\delta(h')$ denote the element of $\pi_{p+q}(A_\infty)$ which is obtained from h' as in § 1. Then $c(h)=\phi\delta(h')$.*

Let $\gamma(h)=\phi\delta(h')$. We check that $\gamma(h)$ satisfies (5.4), (5.5) and (5.6) as well as $c(h)$. For (5.4) follows from (5.3), in the case of $\gamma(h)$, and (5.6) follows from (1.1). Consider (5.5), where we have a map $f: A \to B$. Let $f_\infty: A_\infty \to B_\infty$ denote the multiplicative map which f determines. Then $f_\infty h'$ is equal to the inclusion of fh into B_∞, and so

$$\begin{aligned}\gamma(fh)=\phi\delta(f_\infty h')&=\phi f_*\delta(h'), \text{ by } (1.4) ,\\ &=f_*\phi\delta(h'), \text{ by } (4.3) ,\\ &=f_*\gamma(h), \text{ by definition} .\end{aligned}$$

Therefore $\gamma(h)$ satisfies all three conditions, whence $\gamma(h)=c(h)$. This proves (5.7).

6. **The Whitehead product.** Let X be a space with a basepoint, and let $p, q \geqq 1$. The Whitehead product of a pair of elements (ξ, η), where $\xi \in \pi_{p+1}(X)$, $\eta \in \pi_{q+1}(X)$, is an element of $\pi_{p+q+1}(X)$, and is denoted by $[\xi, \eta]$. In § 9 we shall prove a general theorem about this product which implies the following in case X is the suspension of A, where A is a conplex as in § 4.

THEOREM (6.1). *Let $\lambda \in \pi_p(A_\infty)$, $\mu \in \pi_q(A_\infty)$. Then*

$$\phi\langle\lambda, \mu\rangle=(-1)^p[\phi(\lambda), \phi(\mu)] .$$

Since $\phi j=E$, by (4.2a), we have the following three corollaries in case $\lambda=j(\alpha)$ or $\mu=j(\beta)$, where $\alpha \in \pi_p(A)$, $\beta \in \pi_q(A)$.

COROLLARY (6.2). *Let $\lambda \in \pi_p(A_\infty)$, $\beta \in \pi_q(A)$. Then*

$$\phi\langle\lambda, j(\beta)\rangle=(-1)^p[\phi(\lambda), E(\beta)] .$$

COROLLARY (6.3). *Let $\alpha \in \pi_p(A)$, $\mu \in \pi_q(A_\infty)$. Then*

$$\phi\langle j(\alpha), \mu\rangle=(-1)^p[E(\alpha), \phi(\mu)] .$$

COROLLARY (6.4). *Let $\alpha \in \pi_p(A)$, $\beta \in \pi_q(A)$. Then*

$$\phi\langle j(\alpha), j(\beta)\rangle=(-1)^p[E(\alpha), E(\beta)] .$$

Hence and from (5.2) we obtain the first commutation law for triad Whitehead products (cf. (2.4) of [3]):

COROLLARY (6.5). *Let* $\alpha \in \pi_p(A)$, $\beta \in \pi_q(A)$. *Then*

$$\{\alpha, \beta\}-(-1)^{pq}\{\beta, \alpha\}=(-1)^p i[E(\alpha), E(\beta)] .$$

As defined in [2], A_∞ is filtered by a sequence of subcomplexes $A_0 \subset A_1 \subset \cdots \subset A_m \subset \cdots$, where $A_0=a^0$ and $A_1=A$. The *reduced product filtration* of $\pi_{r+1}(\hat{A})$ is defined as follows (cf. § 13 of [3]). We say that an element $\gamma \in \pi_r(A_\infty)$ has filtration m ($m \geqq 1$) if γ can be represented by a map of S^r whose image is contained in A_m but not by one whose image is contained in A_{m-1}. We also say that the zero element has filtration zero. The reduced product filtration of $\pi_{r+1}(\hat{A})$ is obtained from this by applying the canonical isomorphism ϕ. We supplement the results of [3] by

COROLLARY (6.6). *Let* $\xi \in \pi_{p+1}(\hat{A})$, $\eta \in \pi_{q+1}(\hat{A})$ *be elements with filtrations* m, n, *respectively. Then the filtration of the Whitehead product* $[\xi, \eta] \in \pi_{p+q+1}(\hat{A})$ *does not exceed* $m+n$.

This follows from (6.1). For let $\alpha \in \pi_p(A_\infty)$, $\beta \in \pi_q(A_\infty)$ be elements such that $\varphi(\alpha)=\xi$, $\varphi(\beta)=\eta$. By hypothesis, there exist maps $f\colon S^p \to A_\infty$, $S^q \to A_\infty$, representing α, β, respectively, such that $fS^p \subset A_m$, $gS^q \subset A_n$. Now $A_m \cdot A_n=A_n \cdot A_m=A_{m+n}$, by the definition of A_∞. Hence $h(S^p \times S^q) \subset A_{m+n}$, where h denotes either of the maps

$$f(x)\cdot g(y) \leftarrow (x, y) \rightarrow g(y)\cdot f(x) \qquad x \in S^p , \quad y \in S^q .$$

Therefore $\langle\alpha, \beta\rangle$ can be represented by a map of S^{p+q} into A_{m+n}, so that the filtration of $\langle\alpha, \beta\rangle$, and hence of $\phi\langle\alpha, \beta\rangle$, does not exceed $m+n$. Hence, by (6.1), the filtration of $[\phi(\alpha), \phi(\beta)]$ does not exceed $m+n$, which proves (6.6).

PART III

APPLICATION TO LOOP-SPACES

7. The loop-space (in the sense of Moore). Let X be a space with basepoint x_0. By a *loop* in X we mean a pair (f, s), where $s \geqq 0$ and f is a map of the interval $0 \leqq t \leqq s$ into X such that $f(0)=f(s)=x_0$. The composition of (f, s) with another loop (f', s') is the loop $(f'', s+s')$, where

$$f''(t)=\begin{cases} f(t) & (0 \leqq t \leqq s)\ , \\ f'(t-s) & (s \leqq t \leqq s+s')\ . \end{cases}$$

Let Λ denote the set of loops with the topology defined in § 2 of [4]; we call Λ the *loop-space* of X. The ordinary space of loops, Ω, consists of those loops (f, s) such that $s=1$. Let x^0 denote the loop $(f, 0)$, where $f(0)=x_0$. The product in Λ which is defined by composition of loops is associative, and admits x^0 as a unit element. In § 2 of [4] Moore asserts the following propositions. We omit the proofs, which are straightforward but tedious.

THEOREM (7.1). (a) *The product in Λ is continuous; and* (b) *Ω is a deformation retract of Λ.*

Let us represent S^{r+1} as the suspension of S^r, as in § 1 of [3], so that $(x, t) \in S^{r+1}$ if $x \in S^r$ and $t \in I$. Let h: $S^r \to \Lambda$ be a map, and let $h(x)=(f, s)$, say, where $s \geqq 0$ and f maps the interval $0 \leqq t \leqq s$ into X. Then a map h': $S^{r+1} \to X$ is defined by $h'(x, t)=f(st)$, where $0 \leqq t \leqq 1$. The transformation $h \to h'$ is invariant under homotopy, and therefore it defines a function ψ: $\pi_r(\Lambda) \to \pi_{r+1}(X)$. We prove

THEOREM (7.2). *The function ψ is an isomorphism* (*onto*).

For let i_* denote the injection of $\pi_r(\Omega)$ into $\pi_r(\Lambda)$, which is an isomorphism by (7.1b). By taking representatives we find that $\psi i_*=\theta$, the Hurewicz isomorphism of $\pi_r(\Omega)$ onto $\pi_{r+1}(X)$. Hence ψ is an isomorphism, which proves (7.2). Notice also that ψ is *natural*. To be precise, let X' be a space with a basepoint and let h: $X \to X'$ be a map. If (f, s) is a loop in X, where $s \geqq 0$ and f maps the interval $0 \leqq t \leqq s$ into X, then (hf, s) is a loop in X'. Let $\bar{h}$: $\Lambda \to \Lambda'$ denote the multiplicative map which is defined by $\bar{h}(f, s)=(hf, s)$, where Λ' is the loop-space of X'. Then by taking representatives it follows at once that

$$\psi'\bar{h}_*=h_*\psi\ , \tag{7.3}$$

where ψ' means the same for X' as ψ does for X, and where h_*, $\bar{h}_*$ are the homomorphisms induced by h, $\bar{h}$, as shown in the following diagram:

$$\begin{array}{ccc} \pi_r(\Lambda) & \xrightarrow{\bar{h}_*} & \pi_r(\Lambda') \\ \psi\downarrow & & \downarrow\psi' \\ \pi_{r+1}(X) & \xrightarrow[h_*]{} & \pi_{r+1}(X')\ . \end{array}$$

8. The canonical isomorphism. Let A be a space, with basepoint

a^0, on which a real-valued continuous function d is defined which is positive except that $d(a^0)=0$. Let $\hat{A}$ denote the suspension of A, and let Λ denote the loop-space of $\hat{A}$. Let $x^0 \in \Lambda$ denote the trivial loop at the suspension of a^0. Then a map $u: A \to \Lambda$ is defined as follows. Let $a \in A$ and let $\alpha = d(a)$. We define $u(a^0)=x^0$. Let $a \neq a^0$. Then $\alpha > 0$ and we define $u(a)=(f, \alpha)$, where f is the map of the interval $0 \leqq t \leqq \alpha$ into $\hat{A}$ which is defined by $f(t)=(a, t/\alpha)$. Of course u depends on d, but since the set of functions d is convex it follows that any two maps u are homotopic. The topology of Λ is such that u is a homeomorphism into Λ, so that we regard A as a subspace of Λ.

We now define a homomorphism, ψ, of the homotopy sequence of the pair (Λ, A) into the suspension sequence of A, as shown in the following diagram.

$$
\begin{array}{ccccccccc}
\cdots \to & \pi_r(A) & \xrightarrow{j} & \pi_r(\Lambda) & \xrightarrow{k} & \pi_r(\Lambda, A) & \xrightarrow{\partial} & \pi_{r-1}(A) & \to \cdots \\
 & \psi\downarrow & & \psi\downarrow & & \psi\downarrow & & \psi\downarrow & \\
\cdots \to & \pi_r(A) & \xrightarrow[E]{} & \pi_{r+1}(\hat{A}) & \xrightarrow[i]{} & \pi_{r+1}(\hat{A}\,;\, C_+,\, C_-) & \xrightarrow[\Delta]{} & \pi_{r-1}(A) & \to \cdots .
\end{array} \tag{8.1}
$$

We define ψ as follows. Let[3] V^r denote the convex hull of S^{r-1}, so that points of V^r are represented by pairs (s, x), where $x \in S^{r-1}$ and $0 \leqq s \leqq 1$, such that $(0, x)=e$, $(1, x)=x$. Let V^{r+1} denote the suspension of V^r, so that points of V^{r+1} are represented by pairs (y, t), where $y \in V^r$ and $0 \leqq t \leqq 1$. Let h: $V^r \to \Lambda$ be a map, and let $h(y)=(f, s)$, say, where $s \geqq 0$ and f maps the interval $0 \leqq t \leqq s$ into $\hat{A}$. Let h': $V^{r+1} \to \hat{A}$ be the map defined by $h'(y, t)=f(st)$. Since $h'S^r=a^0$ if $hS^{r-1}=x^0$, we define ψ on $\pi_r(\Lambda)$ to be the homomorphism induced by the transformation $h \to h'$. It is easy to check that ψ means the same here as in (7.2). If $hS^{r-1} \subset A$ then h' maps one hemisphere of S^r into C_+ and the other into C_-, so that the transformation $h \to h'$ also induces a homomorphism of $\pi_r(\Lambda, A)$ into $\pi_{r+1}(\hat{A}; C_+, C_-)$. Thus we define ψ on $\pi_r(\Lambda, A)$, and the definition is completed by setting ψ to be the identity on $\pi_r(A)$. It is easily verified that these definitions make (8.1) commutative, that is, that

$$
\begin{array}{ll}
\text{(a)} & \\
\text{(b)} & \\
\text{(c)} &
\end{array}
\left\{
\begin{array}{l}
\psi j = E\,, \\
i\psi = \psi k\,, \\
\Delta\psi = \partial\,.
\end{array}
\right. \tag{8.2}
$$

Since ψ maps $\pi_r(A)$ identically, by definition, and maps $\pi_r(\Lambda)$ isomorphically, by (7.2), we obtain by an application of the five lemma:

[3] See § 1 of [**3**] for details of these representations.

THEOREM (8.3). *As shown in* (8.1), ψ *is an isomorphism of the homotopy sequence of pair* (Λ, A) *onto the suspension sequence of* A.

Now let A be a complex as in § 4, and consider the reduced product complex A_∞. We extend the inclusion map $u\colon A \to \Lambda$ to a multiplicative map $w\colon A_\infty \to \Lambda$, as follows. Let $a_* \in A_\infty$, so that $a_* = a_1 \cdot a_2 \cdot \cdots \cdot a_n$, say, where $a_1, a_2, \cdots, a_n \in A$. Then we define

$$w(a_*) = u(a_1) \cdot u(a_2) \cdot \cdots \cdot u(a_n) .$$

Notice that w is nonsingular, although A_∞ is not mapped homeomorphically unless $A = a^0$. Let w_* denote the homomorphism of the homotopy sequence of (A_∞, A) into the homotopy sequence of (Λ, A) which is induced by w, and let ϕ denote the canonical isomorphism of the homotopy sequence of (A_∞, A) onto the suspension sequence of A, as in (4.1). It follows from the definition of ϕ in § 10 of [3] that

$$\phi = \psi w_* . \tag{8.4}$$

Hence and from (8.3) we obtain

THEOREM (8.5). *The homomorphism* w_* *maps the homotopy sequence of the pair* (A_∞, A) *isomorphically onto the homotopy sequence of the pair* (Λ, A).

Thus w is an algebraic homotopy equivalence of the pair, in the sense of [2]. Let us also denote by w_* the homomorphism in singular homology which w induces. Then from (8.5) of [2] we obtain

COROLLARY (8.6). *The homomorphism* w_* *maps the singular homology sequence of the pair* (A_∞, A) *isomorphically onto the singular homology sequence of the pair* (Λ, A).

The next section is devoted to proving:

THEOREM (8.7). *Let* Λ *denote the loop-space of a space* X, *and let* $\xi \in \pi_p(\Lambda)$, $\eta \in \pi_q(\Lambda)$, *where* $p, q \geqq 1$. *Then*

$$\psi\langle \xi, \eta \rangle = (-1)^p [\psi(\xi), \psi(\eta)] .$$

We conclude the present section by showing how (6.1) is deduced from (8.7). Let A be a complex as in § 6, and let $X = \hat{A}$ in (8.7). Then if $\lambda \in \pi_p(A_\infty)$, $\mu \in \pi_q(A_\infty)$ are the elements given in (6.1) we have that

$$w_*\langle \lambda, \mu \rangle = \langle w_*(\lambda), w_*(\mu) \rangle ,$$

by (2.2), since w is multiplicative. Moreover,

$$\psi\langle w_*(\lambda), w_*(\mu) \rangle = (-1)^p [\psi w_*(\lambda), \psi w_*(\mu)] ,$$

by (8.7). Hence and from (8.4) we conclude that

$$\varphi\langle\lambda, \mu\rangle=\psi w_*\langle\lambda, \mu\rangle=(-1)^p[\psi w_*(\lambda), \psi w_*(\mu)]$$
$$=(-1)^p[\phi(\lambda), \phi(\mu)] .$$

Thus (6.1) follows from (8.7), and it only remains for us to prove (8.7).

9. Proof of (8.7). Let A be a countable CW-complex with only one 0-cell, and let $u: A\to \Lambda$ be the inclusion map, where Λ denotes the loop-space of $\hat{A}$. We prove first of all

THEOREM (9.1). *Let $h: S^p\times S^q\to A$ be a map, where $p, q\geqq 1$, and let $c(h)$ denote the element of $\pi_{p+q+1}(\hat{A})$ which is obtained from h by the Hopf construction. Let $\delta(uh)$ denote the element of $\pi_{p+q}(\Lambda)$ which is obtained from the inclusion of h into Λ as in* § 1. *Then $c(h)=\psi\delta(uh)$.*

Proof. We have $uh=wh'$, by the definition of w, where h' denotes the inclusion of h into A_∞. Hence

$$\delta(uh)=\delta(wh')=w_*\delta(h') ,$$

by (1.4), since w is multiplicative. Hence

$$\psi\delta(uh)=\psi w_*\delta(h')=\phi\delta(h'),\ \text{by (8.4)} ,$$
$$=c(h),\ \text{by (5.7)} .$$

This proves (9.1). We deduce

COROLLARY (9.2). *Let $h_1, h_2: S^p\times S^q\to A$ be maps which have the same sections, and let $d(h_1, h_2)$ denote their separation element in $\pi_{p+q}(A)$. Then*

$$c(h_2)=c(h_1)+Ed(h_1, h_2) .$$

Proof. We have

$$\delta(uh_2)-\delta(uh_1)=d(uh_1, uh_2),\ \text{by (1.3)} ,$$
$$=jd(h_1, h_2) .$$

by the naturality of the separation element. Therefore

$$\psi\delta(uh_2)-\psi\delta(uh_1)=\psi jd(h_1, h_2)=Ed(h_1, h_2) ,$$

by (8.2a). Hence (9.2) follows from (9.1).

Now take $A=S^p\times S^q$ and let $\alpha\in\pi_p(A)$, $\beta\in\pi_q(A)$ be the homotopy classes of the maps of S^p, S^q into A which are given by

$$x\to(x, e) ,\quad (x\in S^p) ;\qquad y\to(e, y) ,\quad (y\in S^q) ;$$

respectively. We prove that

$$\psi\langle j(\alpha), j(\beta)\rangle = (-1)^p[E(\alpha), E(\beta)] . \tag{9.3}$$

For let $h: S^p \times S^q \to A$ denote the identity map, and let $v: S^q \times S^p \to A$ denote the map which interchanges the factors. Since uh has type $(j(\alpha), j(\beta))$, it follows from (2.3) that

$$\langle j(\alpha), j(\beta)\rangle = \delta(uh) - (-1)^{pq}\delta(uv) .$$

Therefore

$$\begin{aligned}\psi\langle j(\alpha), j(\beta)\rangle &= \psi\delta(uh) - (-1)^{pq}\psi\delta(uv)\\ &= c(h) - (-1)^{pq}c(v), \text{ by (9.1)} ,\\ &= (-1)^p[E(\alpha), E(\beta)] ,\end{aligned}$$

by (2.19) of [3]. This proves (9.3).

We continue to consider $A = S^p \times S^q$, and we denote the set of axes $S^p \times e \cup e \times S^q$ by A'. Then $\hat{A}$ contains $\hat{A}'$, which we identify with $S^{p+1} \times e \cup e \times S^{q+1}$. Let Λ' denote the loop-space of $\hat{A}'$, regarded as a subspace of Λ, and let j', E' and ψ' mean the same in the case of A' as do j, E and ψ in the case of A. Thus, if k_* denotes any of the homomorphisms induced by the inclusion map $k: A' \to A$ we have the relations

$$\text{(9.4)} \qquad \begin{array}{c} \text{(a)} \\ \text{(b)} \\ \text{(c)} \end{array} \begin{cases} k_* j' = j k_* , \\ k_* E' = E k_* , \\ k_* \psi' = \psi k_* . \end{cases}$$

Let $\alpha' \in \pi_p(A')$, $\beta' \in \pi_q(A')$ denote the homotopy classes of the maps of S^p, S^q into A' which are defined by

$$x \to (x, e) , \quad (x \in S^p) ; \qquad y \to (e, y) , \quad (y \in S^q) ;$$

respectively. Since $\alpha = k_*(\alpha')$ and $\beta = k_*(\beta')$ it follows that

$$\begin{aligned}\psi k_*\langle j'(\alpha'), j'(\beta')\rangle &= \psi\langle k_* j'(\alpha'), k_* j'(\beta')\rangle, \text{ by (2.2)} ,\\ &= \psi\langle j(\alpha), j(\beta)\rangle, \text{ by (9.4a)},\\ &= (-1)^p[E(\alpha), E(\beta)], \text{ by (9.3)} ,\\ &= (-1)^p[k_* E'(\alpha'), k_* E'(\beta')], \text{ by (9.4b)} ,\\ &= (-1)^p k_*[E'(\alpha'), E'(\beta')] ,\end{aligned}$$

by the naturality of the Whitehead product. Hence

$$k_*\psi'\langle j'(\alpha'), j'(\beta')\rangle = (-1)^p k_*[E'(\alpha'), E'(\beta')] ,$$

by (9.4c). But since $\hat{A}'$ is a retract of $\hat{A}$, the injection

$$k_*: \pi_{p+q+1}(\hat{A}') \to \pi_{p+q+1}(\hat{A})$$

is an isomorphism into. Therefore we conclude that

$$\psi'\langle j'(\alpha'), j'(\beta')\rangle = (-1)^p[E'(\alpha'), E'(\beta')] . \tag{9.5}$$

Continue with the same meaning for A', Λ' etc., but now let Λ mean the loop-space of X, as in (8.7). Let $\xi \in \pi_p(\Lambda)$, $\eta \in \pi_q(\Lambda)$ be the elements given in (8.7). Let f, g be maps of S^{p+1}, S^{q+1} into X which represent $\psi(\xi)$, $\psi(\eta)$, respectively, and let $h: \hat{A} \to X$ denote the map which is defined by

$$h(x, e) = f(x) , \quad (x \in S^{p+1}) ; \qquad h(e, y) = g(y) , \quad (y \in S^{q+1}) .$$

Let $\bar{h}: \Lambda' \to \Lambda$ denote the map defined by composing loops with h, as in § 7. Consider the induced homomorphisms

$$\bar{h}_*: \pi_r(\Lambda') \to \pi_r(\Lambda) , \qquad h_*: \pi_{r+1}(\hat{A}') \to \pi_{r+1}(X) ,$$

which are related by (7.3). We have

$$\psi(\xi) = h_* E'(\alpha') , \qquad \psi(\eta) = h_* E'(\beta') , \tag{9.6}$$

by the definition of h. By (8.2a) and (7.3), however

$$h_* E' = h_* \psi' j' = \psi \bar{h}_* j' ,$$

and so it follows from (7.2) and (9.6) that

$$\xi = \bar{h}_* j'(\alpha') , \qquad \eta = \bar{h}_* j'(\beta') .$$

Therefore

$$\begin{aligned} \langle \xi, \eta \rangle &= \langle \bar{h}_* j'(\alpha'), \bar{h}_* j'(\beta') \rangle \\ &= \bar{h}_* \langle j'(\alpha'), j'(\beta') \rangle , \end{aligned}$$

by (2.2), since $\bar{h}$ is multiplicative. Since $\psi \bar{h}_* = h_* \psi'$, by (7.3), it follows that

$$\begin{aligned} \psi\langle \xi, \eta \rangle &= h_* \psi' \langle j'(\alpha'), j'(\beta') \rangle , \\ &= (-1)^p h_* [E'(\alpha'), E'(\beta')], \text{ by } (9.5) , \\ &= (-1)^p [h_* E'(\alpha'), h_* E'(\beta')], \text{ by naturality} , \\ &= (-1)^p [\psi(\xi), \psi(\eta)], \text{ by } (9.6) . \end{aligned}$$

This proves (8.7), and completes the proof of the various other theorems which we have deduced from it.

Appendix

10. Separation elements. The notion of a separation element is not exactly a special case of the notion of a separation cochain (see [1]). Hence we provide a brief account in this Appendix.

Let S^r $(r \geqq 1)$ denote the unit sphere in euclidean $(r+1)$-space, and let S^{r-1} denote its equator. Let V^r denote the convex hull of the equator, and let E_+, E_- denote the two hemispheres into which S^{r-1} divides S^r. Let $p, q: V^r \rightarrow S^r$ denote the orthogonal projections of V^r onto E_+, E_-, respectively, (orthogonal to the plane of V^r).

Let K be a CW-complex with a subcomplex L such that $K-L=e^r$, an open r-cell. That is to say, e^r is the topological image of the interior of V^r under a map $f: V^r \rightarrow K$ such that $fS^{r-1} \subset L$. Let $u, v: K \rightarrow X$ be maps which agree on L, where X is a space. Then we define a map $g: S^r \rightarrow X$ by

$$gp=uf\,, \qquad gq=vf\,. \tag{10.1}$$

We define $d(u, v)$, the separation element of u and v, to be the homotopy class of g in $\pi_r(X)$. The following relations are easily verified.

THEOREM (10.2). *Let $u, v: K \rightarrow X$ be maps which agree on L. Then $u \simeq v$, relative to L, if, and only if, $d(u, v)=0$.*

COROLLARY (10.3). *If $u: K \rightarrow X$ is a map then $d(u, u)=0$.*

THEOREM (10.4). *Let $u, v, w: K \rightarrow X$ be maps which agree on L. Then $d(u, w)=d(u, v)+d(v, w)$.*

COROLLARY (10.5). *Let $u, v: K \rightarrow X$ be maps which agree on L. Then $d(u, v)+d(v, u)=0$.*

THEOREM (10.6). *Let an element $\delta \in \pi_r(X)$ and a map $u: K \rightarrow X$ be given. Then there exists a map $v: K \rightarrow X$ which agrees with u on L such that $d(u, v)=\delta$.*

THEOREM (10.7). *Let $u_t, v_t: K \rightarrow X$ be homotopies which agree on L, where $0 \leqq t \leqq 1$. Then $d(u_0, v_0)=d(u_1, v_1)$.*

THEOREM (10.8). *Let $u, v: K \rightarrow X$ be maps which agree on L. Let $h: X \rightarrow Y$ be a map, where Y is a space. Then $d(hu, hv)$ is equal to the image of $d(u, v)$ under the homomorphism induced by h.*

THEOREM (10.9). *Let $k: (K, L) \to (K', L')$ be a map of degree p, where K and K' are CW-complexes with subcomplexes L and L', respectively, which are complements of r-cells in their respective complexes. Let $u, v: K' \to X$ be maps which agree on L'. Then $d(uk, vk) = pd(u, v)$.*

REFERENCES

1. S. Eilenberg, *Cohomology and continuous mappings,* Ann. of Math., **41** (1940), 231–251.
2. I. M. James, *Reduced product spaces,* Ann. of Math., **62** (1955), 170–197.
3. I. M. James, *On the suspension triad,* Ann. of Math., **63** (1956), 191–247.
4. J. C. Moore, *The double suspension* ···, Bol. Soc. Mat. Mexicana (2) **1** (1956), 28–37.
5. N. E. Steenrod, *The topology of fibre bundles,* Princeton University Press, 1951.
6. G. W. Whitehead, *On mappings into group-likc spaces,* Comm. Math. Helv., **28** (1954), 320–328.

THE INSTITUTE FOR ADVANCED STUDY

THREE TEST PROBLEMS IN OPERATOR THEORY

RICHARD V. KADISON AND I. M. SINGER

1. Introduction. In his tract [3] on infinite abelian groups, I. Kaplansky proposes three problems with which to test the adequacy of a purported structure theory for the subject. The problems are general with a certain intrinsic interest, and he comments there that they provide a worthy test in other subjects. In particular, Kaplansky has suggested these problems, suitably rephrased, in conversation as a test of a unitary equivalence theory for operators on a Hilbert space. In the order we treat them they are:

1. If A and B are operators acting on Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$ and the operators $\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}$ and $\begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$, acting in the obvious way on $\mathscr{H} \oplus \mathscr{H}$ and $\mathscr{K} \oplus \mathscr{K}$, are unitarily equivalent, is it true that A and B are unitarily equivalent?

2. If $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ and $\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$ are unitarily equivalent is it true that B and C are unitarily equivalent?

3. If A and B are unitarily equivalent to direct summands of each other (that is, A equivalent to BF and B equivalent to AE, where E and F commute with A and B, respectively), are A and B unitarily equivalent?

A superficial examination provides examples which show that Problem 2 must, in general, be answered negatively. In fact infinite projections for B and C, one with an infinite and the other with a finite-dimensional orthogonal complement, and A an infinite-dimensional projection with an infinite-dimensional complement illustrates this. On the other hand, all three problems have an affirmative answer in the finite-dimensional case—Problem 3, trivially so, since E and F must be the identity operator on simple numerical-dimension grounds, and the other problems not at all trivially so (especially when approached from an elementary viewpoint).

Problem 3 has an affirmative answer, and a simple adaptation of the usual Cantor-Bernstein argument proves this. We shall give this problem no further attention except to note that it can be settled by

Received July 31, 1956. This work was supported by a contract with The Office of Naval Research.

use of ring of operators techniques as well as by the direct argument mentioned. We shall show that Problem 1 can always be answered affirmatively, and Problem 2 has an affirmative answer provided the rings generated by the operators in question are, together with their commutants, of finite type—a most satisfactory result in view of the negative example presented and the finite-dimensional situation. The proofs make use of some of the sophisticated techniques of the theory of rings of operators (and in some sense these techniques must be used). It seems to us a pleasant circumstance that this theory is capable now of solving some of the primitive problems of the subject. Our primary interest in the questions discussed is in their role of test problems, for which reason, we have refrained from dealing with such obvious generalizations as the one obtained from Problem 1 by replacing the two-fold copies of A and B by n-fold copies (even though the proof would suffice).

2. The test questions. The first of the test questions we shall discuss is that of the unitary equivalence of the operators A and B given that $\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}$ and $\begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$ are unitarily equivalent. A large share of the solution to this question is contained in the process of phrasing it properly in the terminology of rings of operators and taking full advantage of the hypotheses in these terms. Let $\mathscr{M}$ be the ring of operators generated by $\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}$ and φ the $*$-isomorphism of $\mathscr{M}$ onto $\mathscr{N}$, the ring generated by $\begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$, determined by $\varphi\left(\begin{bmatrix} A & 0 \\ 0 & A \end{bmatrix}\right)=\begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$. The projections $E'=\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ and $F'=\begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}$ commute with $\mathscr{M}$ and are equivalent in $\mathscr{M}'$ via the partial isometry $\begin{bmatrix} 0 & I \\ 0 & 0 \end{bmatrix}$ (in $\mathscr{M}'$); moreover $E'+F'=I$. These same properties hold for the projections M', N' given by the same matrix description relative to $\mathscr{N}'$. In these terms, our result becomes:

THEOREM 1. *The mapping ψ defined on $\mathscr{M}E'$ by $\psi(TE')=\varphi(T)M'$ is implemented by a unitary transformation when φ is implemented by a unitary transformation.*

Proof. Let U be a unitary transformation which implements φ, and let us denote by φ again the unitary equivalence induced on all bounded operators by U. Clearly then, φ so extended carries E' and F' into projections $\varphi(E')$ and $\varphi(F')$ in $\mathscr{N}'$ such that $\varphi(E')$ and $\varphi(F')$ are equivalent and have sum I. We shall note that $\varphi(E')$ and M' are equivalent under these conditions; but let us assume this for the moment, and let W' be a partial isometry in $\mathcal{N}'$ effecting this equivalence. We assert that the unitary transformation $W'UE'$ of the range of E' onto the range of M' implements ψ. Indeed,

$$(*)\quad W'UE'(TE')E'U^{-1}W'^* = W'UTU^{-1}UE'U^{-1}W'^* = W'\varphi(T)\varphi(E')W'^*$$

$$= \varphi(T)W'\varphi(E')W'^* = \varphi(T)W'W'^* = \varphi(T)M' = \psi(TE') .$$

That $\varphi(E')$ and M' are equivalent may be accepted as a consequence of the elementary comparison theory of projections in a ring of operators (all projections equivalent to their orthogonal complements are equivalent to each other), or may be reduced to more apparent facts of this theory. In fact if $\varphi(E')$ is not equivalent to M' then for some nonzero central projection P in $\mathcal{N}$, we have, say, $P\varphi(E') \prec PM'$. Restricting consideration to the range of P, we may assume that $\varphi(E') \prec M'$ whence $\varphi(F') \prec N'$ and $I = \varphi(E') + \varphi(F') \prec M' + N' = I$, a contradiction. Establishing this last relation in all detail, however, would require in effect an easy but lengthy development of the cardinal-valued dimension function for projections in a ring of operators. We shall let these remarks suffice as an indication of the proof that $\varphi(E')$ and M' are equivalent.

The argument contained in $(*)$ can be applied more generally to prove a fact which will be of later use. We state this fact in:

REMARK 2. If φ is a unitary equivalence carrying $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ onto $\begin{bmatrix} C & 0 \\ 0 & D \end{bmatrix}$ and $\varphi\left(\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}\right)$ is equivalent to $\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ in the commutant of the ring generated by $\begin{bmatrix} C & 0 \\ 0 & D \end{bmatrix}$, then A and C are unitarily equivalent (via the natural restriction of φ). A curious consequence of this remark is the fact that if $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ generates a factor of type III (on a separable space) then the existence of the unitary equivalence φ implies the unitary equivalence of A, C, B, and D.

It might be thought that some simple construction with the unitary transformation which effects the original equivalence alone in Problem 1 might yield the appropriate unitary operator for demonstrating the equivalence of A and B. That this is not the case can be seen by taking A and B to be I, so that an arbitrary unitary transformation effects the original equivalence.

The next test question we take up is that of the unitary equivalence of B and C given the unitary equivalence of $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ and $\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$. We have noted that the unitary equivalence of B and C, under these conditions, does not follow, in general. Our example illustrating this possibility relies upon an "improper mixture of finiteness and infiniteness". The following theorem shows that, when such a mixture is not possible, B is unitarily equivalent to C. This mixture is not possible when the ring of operators $\mathcal{M}$ generated by $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ is finite with finite commutant $\mathcal{M}'$. Our hypothesis tells us that the $*$-isomorphism ψ of $\mathcal{M}$ onto the ring $\mathcal{N}$ generated by $\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$, determined by $\psi\left(\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\right)=\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$ is implemented by a unitary transformation, and, with E' the projection $\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ in $\mathcal{M}'$, F' the projection $\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}$ in $\mathcal{N}'$, the mapping η of $\mathcal{M}E'$ onto $\mathcal{N}F'$ defined by $\eta(TE')=\psi(T)F'$ is a $*$-isomorphism which is implemented by a unitary transformation. We shall denote the unitary equivalences induced on the rings of all bounded operators on $\mathcal{H}$ and $\mathcal{H}E'$ by unitary transformations which implement ψ and η respectively, by ψ and η again, so that it will be meaningful to speak, for example, of $\psi(E')$. In the notation just described our statement becomes:

THEOREM 3. *If $\mathcal{M}$ and $\mathcal{M}'$ are finite the mapping φ of $\mathcal{M}(I-E')$ onto $\mathcal{N}(I-F')$ defined by $\varphi(T(I-E'))=\psi(T)(I-F')$, for T in $\mathcal{M}$, is a $*$-isomorphism which is implemented by a unitary transformation.*

Proof. Note first that the definition of η and the fact that it is a $*$-isomorphism implies that $\psi(C_{E'})=C_{F'}$, in view of [**2**; Lemma 3.1.3], and by this same result, it will suffice to show that $\psi(C_{I-E'})=C_{I-F'}$ in order to establish that φ is a $*$-isomorphism. Now $I-C_{I-E'}$ is the union of all central projections contained in E', whence, from the symmetry of this situation, it will suffice to show that if P is a central projection in $\mathcal{M}$ contained in E' then $\psi(P)\leqq F'$. We make use of the dimension functions in the various rings, and we shall denote these functions by D for $\mathcal{M}$, $\mathcal{M}'$, $\mathcal{N}$, and $\mathcal{N}'$ and by D_0 for $\mathcal{M}E'$, $E'\mathcal{M}'E'$, $\mathcal{M}(I-E')$, $(I-E')\mathcal{M}'(I-E')$, $\mathcal{N}F'$, $F'\mathcal{N}'F'$, $\mathcal{N}(I-F')$, and $(I-F')\mathcal{N}'(I-F')$. By definition $\eta(P)=\eta(PE')=\psi(P)F'$, and η is a unitary equivalence so that

$$\eta[D_0(E'PE')]=\eta(P)=\psi(P)F'=D_0(F'\psi(P)F')$$

$$=\frac{D[\psi(P)]F'}{D(F')}=\frac{\psi(P)F'}{D(F')}$$

(recall that, with G' in $F'\mathcal{N}'F'$, $D_0(G')=F'D(G')/D(F')$). Thus $\psi(P)(D(F')-I)F'=0$, so that $\psi(P)C_{F'}(D(F')-I)=0$, by [**2**; Lemma 3.1.1], and $\psi(P)(D(F')-I)=0$, since $\psi(P)\leqq\psi(C_{E'})=C_{F'}$. It follows that $D(\psi(P)-\psi(P)F')=0$ and $\psi(P)-\psi(P)F'=0$; that is, $\psi(P)\leqq F'$, and φ is a $*$-isomorphism of $\mathscr{M}(I-E')$ onto $\mathscr{N}(I-F')$.

To show that φ is implemented by a unitary transformation, it will suffice, of course, to establish this for each projection of an orthogonal family of central projections in $\mathscr{M}(I-E')$ with sum $I-E'$; whence it suffices to consider the case in which the center of $\mathscr{M}$, and hence $\mathscr{M}$ itself as well as $\mathscr{M}'$, $\mathscr{M}(I-E')$, $(I-E')\mathscr{M}'(I-E')$, $\mathscr{N}$, $\mathscr{N}'$, $\mathscr{N}(I-F')$, $(I-F')\mathscr{N}'(I-F')$, is countably-decomposable. Choose unit vectors x and y such that $M=[\mathscr{M}'x]$, $M'=[\mathscr{M}x]$, $N=[\mathscr{N}'y]$, and $N'=[\mathscr{N}y]$ are maximal cyclic projections in $\mathscr{M}$, $\mathscr{M}'$, $\mathscr{N}$, and $\mathscr{N}'$, respectively. (The existence of such projections follows from [2; Lemma 3.3.7].

Suppose that we can show

$$\psi[D(E')]=D(F')\,, \tag{1}$$

In this case $\psi(E')$ and F' are equivalent, whence, by the finiteness of $\mathscr{N}'$ $\psi(I-E')$ and $I-F'$ are equivalent and our theorem follows from Remark 2. Our task then is to prove (1).

Let G and G' be paired projections (that is, ones having a joint generating vector) in $\mathscr{M}E'$ and $E'\mathscr{M}'E'$, respectively. Then, for each vector z,

$$D_0(G')D_0([E'\mathscr{M}'E'z])=D_0(G)D_0([\mathscr{M}E'z])\,,$$

by The Coupling Theorem (see [1], for example, or [2; Theorem 3.3.8]). From this, we have

$$D_0(G')D([\mathscr{M}'E'z])D(E')=D_0(G)D([\mathscr{M}E'z])\,. \tag{2}$$

Now

$$D([\mathscr{M}'E'z])D(M')=D([\mathscr{M}E'z])D(M)\,, \tag{3}$$

whence, multiplying (2) by $D(M)$ and combining with (3) we have

$$D_0(G')D([\mathscr{M}'E'z])D(E')D(M)=D_0(G)D([\mathscr{M}'E'z])D(M')$$

so that

$$(D_0(G')D(E')D(M)-D_0(G)D(M'))C_{[\mathscr{M}'E'z]}=0\,. \tag{4}$$

Since E' contains a cyclic projection with central carrier $C_{E'}$, z can be so chosen that $C_{[\mathscr{M}'E'z]}=C_{E'}$, and since $C_{E'}D_0(G')=D_0(G')$, $C_{E'}D_0(G)=D_0(G)$, (4) becomes

$$D_0(G')D(E')D(M)=D_0(G)D(M')\,. \tag{5}$$

Similarly,

$$(6) \qquad D_0(\eta(G'))D(F')D(N) = D_0(\eta(G))D(N') ,$$

since η is a unitary equivalence and $\eta(G')$, $\eta(G)$ are paired projections in $F'\mathcal{N}'F'$, $F'\mathcal{N}$. Writing (5) as $D_0(G')D(E')E'D(M)E' = D_0(G)D(M')E'$ and applying η to it we have

$$(7) \qquad D_0(\eta(G'))\psi(D(E'))\psi(D(M)) = D_0(\eta(G))\psi(D(M')) .$$

Since ψ is a unitary equivalence and N, N', M, M' are maximal cyclic, we have $\psi(D(M)) = D(N)$, $\psi(D(M')) = D(N')$, so that, comparing (6) and (7),

$$D_0(\eta(G'))D(N)(D(F') - \psi(D(E'))) = 0 .$$

This being true for each cyclic projection $\eta(G')$ in $F'\mathcal{N}'F'$, $D(N)(D(F') - \psi(D(E'))) = 0$, whence $C_N(D(F') - \psi(D(E'))) = 0$. But $C_N = I$, so that $D(F') = \psi(D(E'))$, and the proof is complete.

References

1. E. L. Griffin, *Some contributions to the theory of rings of operators*, Trans. Amer. Math. Soc., **75** (1953), 471–504.

2. R. V. Kadison, *Unitary invariants for representations of operator algebras*, Ann. Math., **65** (1957).

3. I. Kaplansky, *Infinite Abelian groups*, Ann Arbor, Michigan, 1955.

Columbia University and
Massachusetts Institute of Technology

A CONVERGENCE THEOREM FOR A CERTAIN CLASS OF MARKOFF PROCESSES

MAURICE KENNEDY

1. Introduction. The object of this paper is to generalize, by means of an approach due to S. Karlin [9], a theorem originally obtained by Bellman, Harris and Shapiro [1] which may be stated in the following way:

A system is considered whose state may be described by a point t in the interval $[0, 1]$. A probability measure μ is given for the initial state of the system. At the end of each unit interval of time, one of the transformations A_0, A_1 is applied to the state t with probabilities $\phi_0(t)$, $\phi_1(t)$ respectively, where $\phi_0(t)+\phi_1(t)=1$. The transformations are defined by

$$(1.1) \qquad A_0t=\lambda_0 t\,, \quad A_1 t=\lambda_1 t+(1-\lambda_1)\,, \quad 0\leqq \lambda_0,\ \lambda_1<1^1.$$

The assumption is made that

$$(1.2) \qquad \phi_0(t)=1-t\,, \quad \phi_1(t)=t\,.$$

It is clear that (1.1) and (1.2) ensure that the end-points of the interval $[0, 1]$ are absorbing, that is, if the state of the system is either 0 or 1, it remains so. Let $T\mu$ be the probability measure at the end of the first unit interval. It is then proved that as $n\to\infty$, $T^n\mu$ (that is, the probability distribution for the state of the system at time n) converges in distribution to a distribution concentrated at the points 0, 1 and the form of this limiting distribution which depends on μ is obtained.

The motivation for the consideration of such a system arose from certain learning models introduced by Bush and Mosteller. These are described in detail in their recent book [2]. (Condition (1.2) means that the state of the system may be identified with the probability of applying A_1).

The methods used in [1] to obtain the convergence of $T^n\mu$ are probabilistic. Karlin [9] considers the space of continuous functions on the unit interval and obtains a bounded operator U on this space whose adjoint is T. A convergence theorem is obtained for U^n and the result is translated into the adjoint space (that is, the space of measures) to

Received June 5, 1956. This paper generalizes the main theorem of the author's thesis presented at the California Institute of Technology 1954, being partial requirements for the degree of Doctor of Philosophy. The author wishes to express his sincere thanks to Professor Samuel Karlin for his help and guidance in the preparation of that thesis.

[1] Karlin also considers boundary cases where λ_0, λ_1 may be 1.

obtain the required result.

Karlin [9] also considers cases where (1.2) no longer holds and obtains for a wide class of non-absorbing models the convergence of $T^n\mu$ to a distribution which is independent of μ. These do not concern us here as the object is to consider only a class of absorbing problems, where of course the final distribution depends on the initial distribution.

We conclude this section by stating a well-known theorem [8].

THEOREM 1.1. *Let Ω be a compact Hausdorff space and let $\mathfrak{C}(\Omega)$ denote the Banach space of real-valued continuous functions $x(t)$ defined on Ω with*

$$\|x\| = \max_{t \in \Omega} |x(t)| . \tag{1.3}$$

Let $\mathfrak{M}(\Omega)$ denote the space of all real-valued completely additive regular set functions $\mu(E)$ defined for all Borel sets E of Ω, with

$$\|\mu\| = \sup_{E \subset \Omega} \mu(E) - \inf_{E \subset \Omega} \mu(E) . \tag{1.4}$$

Then $\mathfrak{M}(\Omega)$ is isometric (and lattice isomorphic) to the conjugate space of $\mathfrak{C}(\Omega)$, the correspondence being given by

$$(x, \mu) = \int_\Omega x(t) d\mu(t) . \tag{1.5}$$

2. **Description of the process.** Let Ω be a compact metric space with metric ρ. Since Ω satisfies the second axiom of countability, the concepts of Baire and Borel measures coincide, and thus since the former are always regular [5], we have that the set $\mathfrak{M}(\Omega)$ of Theorem 1.1 consists of all the completely-additive (finite) set functions defined on the Borel sets of Ω.

Let $\{\tau_i\}$ be a countable sets of points in Ω and $\{A_i\}$ a corresponding set of continuous transformations of Ω into itself with the following properties

$$A_i S_i \subset S_i \qquad i = 1, 2, \cdots \tag{2.1}$$

where S_i is any open sphere with centre τ_i, and

$$\lim_{n \to \infty} A_i^n t = \tau_i \qquad i = 1, 2, \cdots \quad \text{for each } t \in \Omega; \tag{2.2}$$

that is, repeated applications of the transformation A_i transforms t in the limit into τ_i and moreover every open sphere with centre τ_i is mapped by A_i into itself. The points $\{\tau_i\}$ will be referred to as boundary points. It follows from (2.2) and the continuity of A_i, that

$$A_i\tau_i=\tau_i\,. \qquad i=1,\ 2,\cdots. \tag{2.3}$$

Consider a system whose state may be described by a point t in Ω. Let $\{\phi_i(t)\}$ be a countable family of continuous functions defined on Ω with the property that

$$0\leq\phi_i(t)\leq 1\,, \qquad i=1,\ 2,\cdots \tag{2.4}$$

and

$$\sum_1^\infty \phi_i(t)=1\,. \tag{2.5}$$

Let $\mu(E)$ be a probability measure defined on the Borel sets of Ω, giving the probability distribution of the initial state of the system. Our process consists in applying at every unit interval of time one of the transformations $\{A_i\}$, A_i being applied with probability $\phi_i(t)$, where $t\in\Omega$ represents the state of the system.

Let

$$T\mu(E)=\sum_{i=1}^\infty\int_{A_i^{-1}E}\phi_i(t)d\mu(t)\,. \tag{2.6}$$

It is easily seen that $T\mu(E)$ is a Borel measure. It represents the probability measure for the state of the system after unit time. $T\mu$ is defined by (2.6) for any Borel measure μ and

$$\|T\mu\|=T\mu(\Omega)=\mu(\Omega)=\|\mu\|\,. \tag{2.7}$$

More generally, if

$$\mu\in\mathfrak{M}(\Omega),\quad \mu=\mu^+-\mu^-\,,$$

(2.6) defines $T\mu$ and

$$T\mu=T\mu^+-T\mu^-\,. \tag{2.8}$$

T is a linear transformation of $\mathfrak{M}(\Omega)$ into itself and

$$\|T\mu\|=\|T\mu^+-T\mu^-\|\leq\|T\mu^+\|+\|T\mu^-\|=\|\mu^+\|+\|\mu^-\|=\|\mu\|\,.$$

Thus we obtain

LEMMA 2.1. *T is a positive linear transformation of $\mathfrak{M}(\Omega)$ into itself of norm* 1.

Now consider $x(t)\in\mathfrak{C}(\Omega)$ (cf. Theorem 1.1). A function $Ux(t)$ is defined on Ω by

$$(2.9) \qquad Ux(t) = \sum_{i=1}^{\infty} \phi_i(t) x(A_i t) \,.$$ [2]

Each term of this series is continuous on Ω and $|x(A_i t)| \leqq ||x||$.

Since $\sum \phi_i(t) = 1$ the convergence being uniform (by Dini's theorem), the series (2.9) is uniformly convergent and hence $Ux(t) \in \mathfrak{C}(\Omega)$. Clearly U is a linear transformation of $\mathfrak{C}(\Omega)$ into itself and $||Ux|| \leqq ||x||$. Thus, since the functions which are constant on Ω, are fixed points of U, we have the following.

LEMMA 2.2. *U is a bounded positive linear transformation of $\mathfrak{C}(\Omega)$ into itself, for which the constant functions are fixed points. Moreover $||U^n|| = 1$ for all positive integers n.*

Theorem 1.1 connects $\mathfrak{C}(\Omega)$ and $\mathfrak{M}(\Omega)$. We now prove

LEMMA 2.3. *T is the adjoint of U, that is,*

$$(2.10) \qquad (Ux,\ \mu) = (x,\ T\mu), \quad \textit{for each } x \in \mathfrak{C}(\Omega) \textit{ and } \mu \in \mathfrak{M}(\Omega)\,.$$

Since $\mu = \mu^+ - \mu^-$, it is clearly sufficient to prove (2.10) for the case $\mu \geqq 0$. Let

$$\nu_i(E) = \int_E \phi_i(t) d\mu(t)$$

It is easy to see that

$$T\mu = \sum_1^{\infty} \nu_i A_i^{-1} \,,$$

the convergence being in the sense of $\mathfrak{M}(\Omega)$. Hence

$$\begin{aligned}(x,\ T\mu) = \sum_1^{\infty} (x,\ \nu_i A_i^{-1}) &= \sum_1^{\infty} \int x(t) d\nu_i A_i^{-1}(t) \\ &= \sum_1^{\infty} \int x(A_i t)\, d\nu_i(t) \\ &= \sum_1^{\infty} \int \phi_i(t) x(A_i t) d\mu(t) \\ &= (Ux,\ \mu)\end{aligned}$$

since the series (2.9) converges uniformly[3].

[2] Operators of the type U have been considered, and both convergence and $\mathfrak{C}-1$ convergence theorems for the iterates U^n obtained by Ocinescu, Mihoc, Doeblin, Fortet, Ionescu Tulcea and Marinescu [**10**, **3**, **4**, **6**, **7**].

[3] This adjointness lemma expresses the fact that if $t_1, t_2, \cdots$ represents the process then $E\{E\{x(t_2)|t_1\}\} = E\{x(t_2)\}$.

3. Absorption assumptions.[4] The first additional assumption to be made is that each of a finite number of the boundary points is an absorbing point, that is, we assume

$$\phi_i(\tau_i)=1 \qquad i=1, 2, \cdots, m. \tag{3.1}$$

This together with (2.3) ensure that τ_i $(i=1, 2, \cdots, m)$ are absorbing points. (Since $\sum\phi_i(t)=1$ and Ω is compact, it is not possible to extend the assumption (3.1) to an infinite number of the boundary points τ_i). The assumption (3.1) is strengthened as follows :

We assume that about each absorbing point τ_i $(1\leqq i\leqq m)$, an open sphere $\sum_i$ may be drawn with centre τ_i on which the infinite product

$$\phi_i(t)\phi_i(A_it)\cdots\phi_i(A_i^nt)\cdots \tag{3.2}$$

converges uniformly (the convergence being in the sense of infinite products that is, the limit is nonzero).

Clearly assumption (3.2) together with (2.2) imply (3.1). Finally, the assumption is made that for each $t\in\Omega-\bigcup_1^m \sum_i$ there is a finite sequence of transformations

$$A_{j_1},\ A_{j_2}, \cdots, A_{j_n} \qquad (1\leqq j_k<\infty)$$

where $n, j_1, j_2, \cdots, j_n$ depend on t, such that $A_{j_n}A_{j_{n-1}}\cdots A_{j_1}t$ is in one of the spheres $\sum_i$ $(1\leqq i\leqq m)$ and such that each term of the sequence

$$\phi_{j_1}(t),\ \phi_{j_2}(A_{j_1}t), \cdots, \phi_{j_n}(A_{j_{n-1}}\cdots A_{j_1}t) \tag{3.3}$$

is greater than zero.

Assumptions (3.2) and (3.3) imply that no matter what the initial state of the system there is always positive probability of reaching an absorbing point after an infinite number of steps. We conclude this section with the following lemma which is a consequence of (3.1).

LEMMA 3.1. *U preserves the values at the absorbing points, that is,*

$$Ux(\tau_i)=x(\tau_i)\,, \qquad i=1, 2, \cdots, m \tag{3.4}$$

where $x(t)\in\mathfrak{C}(\Omega)$.

Proof. Since

$$\sum_1^\infty \phi_i(t)=1 \quad\text{and}\quad \phi_i(t)\geqq 0$$

[4] Cf. discussion Bush and Mosteller, [**2**, pp. 167–169].

for each l, we have by assumption (3.1) that

$$\phi_l(\tau_i)=0 \qquad l\neq i\,,\ 1\leqq i\leqq m,\ 1\leqq l<\infty. \tag{3.5}$$

The result follows by (2.3) from the definition (2.9) of U.

4. Examples.

EXAMPLE 4.1. Let $\{A_i\}$ be a countable set of transformations of Ω into itself with the property that

$$\rho(A_it,\ A_is)\leqq\lambda\rho(t,\ s) \qquad i=1,\ 2,\ \cdots \tag{4.1}$$

for all pairs of points $t,\ s\in\Omega$, where λ is a constant such that $0\leqq\lambda<1$.

It follows from (4.1) that the transformations $\{A_i\}$ are continuous and moreover there exist points $\{\tau_i\}$ such that (2.1) and (2.2) are satisfied.

Let $\{\phi_i(t)\}$ be a family of continuous functions on Ω satisfying the conditions (2.4), (2.5) and the first absorption assumption (3.1). Suppose also for each i $(1\leqq i\leqq m)$ that there exists an open sphere $\sum_i$ with centre τ_i and radius r_i on which $\phi_i(t)>0$ and satisfies a uniform Lipschitz condition[5] that is,

$$|\varphi_i(t)-\phi_i(s)|\leqq k\rho(t,\ s) \qquad t,\ s\in\textstyle\sum_i \tag{4.2}$$

Finally the assumption is made that one of the probability functions, say $\phi_1(t)$, satisfies

$$\phi_1(t)>0 \quad \text{except at the points } \tau_i\ (2\leqq i\leqq m). \tag{4.3}$$

LEMMA 4.1. *The process just described satisfies the absorption assumptions of* § 3.

Proof. We first observe that (3.1) is satisfied by hypothesis. To establish (3.2), let $t\in\sum_i$ $(1\leqq i\leqq m)$ and $\psi_i(t)=1-\phi_i(t)$.

$$\begin{aligned}\psi_i(A_i^nt)&=1-\phi_i(A_i^nt)\\&=\phi_i(A_i^n\tau_i)-\phi_i(A_i^nt) && \text{by (2.3) and (3.1)}\\&\leqq k\rho(A_i^n\tau_i,\ A_i^nt) && \text{by (4.2)}\\&\leqq k\lambda^n\rho(\tau_i,\ t) && \text{by (4.1)}\\&\leqq(kr_i)\lambda^n.\end{aligned}$$

Since $0\leqq\lambda<1$, $\sum\lambda^n$ converges and hence, by a theorem on infinite

[5] These assumptions link up with those given by other authors [**10**, **3**, **4**, **6**, **7**].

products $\prod_{n=0}^{\infty}(1-\phi_i(A_i^n t))$ converges uniformly on $\sum_i$. Thus $\prod_{n=0}^{\infty}\phi_i(A_i^n t)$ converges uniformly on $\sum_i$ and the assumption (3.2) is verified.

It remains to verify (3.3). Let $t \in \Omega - \bigcup_1^m \sum_i$. Since $A_1^n t \to \tau_1$, there exists n_0 such that $A_1^{n_0} t \in \sum_1$. By (4.3), $\phi_1(t) > 0$. Hence we take A_1 as our first transformation. If $A_1 t \in \bigcup_1^m \sum_i$, then (3.3) is already verified. If not $\phi_1(A_1 t) > 0$ and we take A_1 as our second transformation. Proceeding in this manner a finite sequence of A_1's (of length $\leqq n_0$) is obtained which satisfies the assumption (3.3). Hence the lemma is proved.

EXAMPLE 4.2. The example described in § 1 is a particular case of the example just given.

EXAMPLE 4.3. We now consider a generalization from 1 to N dimensions of the learning model considered by Karlin (cf. Bush and Mosteller [2]).

Let Ω be a simplex in E_N (Euclidean space of N dimensions). Any point of Ω is given by its barycentric coordinates $t=(t_1, t_2, \cdots, t_{N+1})$ where $t_i \geqq 0$ and $\sum_1^{N+1} t_i = 1$. The vertices e^i $(i=1, \cdots, N+1)$ have coordinates $e_j^i = \delta_j^i$ (Kronecker delta). Let I denote the $(N+1)\times(N+1)$ unit matrix, and B_i $(1 \leqq i \leqq N+1)$ denote the $(N+1)\times(N+1)$ projection matrix where each element of the ith row is unity and all other elements are zero. Clearly $B_i t = e^i$ for each $t \in \Omega$. Consider the family $\{A_i\}$ of transformations on Ω into itself defined as follows

$$A_i = \lambda_i I + (1-\lambda_i)B_i\,, \qquad 0 \leqq \lambda_i < 1,\ i=1, 2, \cdots, N+1\,, \tag{4.4}$$

that is, for $t \in \Omega$

$$A_i t = (\lambda_i t_1,\ \lambda_i t_2, \cdots, \lambda_i t_i + 1 - \lambda_i,\ \lambda_i t_{i+1} \cdots \lambda_i t_{N+1})$$

Clearly A_i represents a transformation which carries a point P into a point P' on the line PV_i where V_i is the vertex e^i and

$$P'V_i = \lambda_i(PV_i)\ .$$

The transformations $\{A_i\}$ are continuous and satisfy the conditions (2.1) and (2·2) where $\tau_i = e^i$. For the probabilities $\phi_i(t)$ we take

$$\phi_i(t) = t_i \qquad i=1, \cdots, N+1. \tag{4.5}$$

The conditions (2.4) and (2.5) are clearly satisfied. It remains to verify the absorption assumptions of § 3. Since $\phi_i(e^i) = e_i^i = 1$, the condition (3.1) is satisfied. To verify (3.2) we first note that since $B_i^2 = B_i$

$$A_i^n = \lambda_i^n I + (1-\lambda_i^n) B_i$$

so that

$$\phi_i(t) = t_i, \quad \phi_i(A_i^n t) = \lambda_i^n t_i + (1-\lambda_i^n)$$

If $t_i > \varepsilon$, it is easily seen that the infinite product $\prod_0^\infty \phi_i(A_i^n t)$ converges uniformly.

Condition (3.3) is seen to be satisfied by noting that for any point t, one at least of the coordinates is nonzero, say t_i, and hence the ith coordinate of $A_i^n t$ $(n=1, 2, \cdots)$ is also nonzero.

5. Returning to the general absorption process described in §§ 2 and 3, we establish by means of the assumption (3.2), the equicontinuity of the family of functions $\{U^n x(t)\}$ at each of the absorbing points τ_i $(i=1, 2, \cdots, m)$.

LEMMA 5.1. *Let $x(t) \in \mathfrak{C}(\Omega)$ be such that it vanishes at one of the absorbing points τ_i $(1 \leqq i \leqq m)$, then for each $\varepsilon > 0$, there exists a sphere $S_i(\varepsilon)$ with centre τ_i, such that*

$$|U^n x(t)| < \varepsilon \qquad n=1, 2, \cdots$$

for $t \in S_i(\varepsilon)$.

Proof. Without loss of generality we consider the case where $i=1$. Let

$$f_n(t) = \phi_1(t)\phi_1(A_1 t) \cdots \phi_1(A_1^{n-1} t) .$$

$\{f_n(t)\}$ form a nonincreasing sequence of functions which by assumption (3.2) converges uniformly on $\sum_1$ to a function $f(t)$. It follows that $f(t)$ is continuous, and thus since $f_n(\tau_1)=1$ (by (2.3) and (3.1)) and therefore $f(\tau_1)=1$, we have that given any positive number δ $(0<\delta<1)$, there exists a neighbourhood V of τ_1 (contained in $\sum_1$) on which $f(t) > \delta$, which implies $f_n(t) > \delta$ for all n.

Choose $\delta > 1 - \varepsilon/\|x\|$ and let q be a positive integer such that

$$\frac{q-1}{q}\delta > 1 - \varepsilon/\|x\| . \tag{5.1}$$

Since $x(\tau_1)=0$ by hypothesis, there exists a neighbourhood V' of τ_1 such that for $t \in V'$

$$|x(t)| \leqq \frac{1}{q}\|x\| . \tag{5.2}$$

Let $S_1(\varepsilon)$ be an open sphere with centre τ_1 and such that $S_1(\varepsilon) \subset V \cap V'$.

By (2.1) if $t \in S_1(\varepsilon)$, $A_1^n t \in S_1(\varepsilon)$ for all positive integers n. Hence for $t \in S_1(\varepsilon)$

$$|x(A_1^n t)| \leqq \frac{1}{q} ||x|| . \qquad n=1, 2, \cdots , \tag{5.3}$$

and

$$f_n(t) > \delta \qquad n=1, 2, \cdots . \tag{5.4}$$

Now

$$U^n x(t) = \sum_{i_1, i_2 \cdots i_n} \phi_{i_1}(t)\phi_{i_2}(A_{i_1}t) \cdots \phi_{i_n}(A_{i_{n-1}} \cdots A_{i_1}t) x(A_{i_n} \cdots A_{i_1}t) ,$$

$$|U^n x(t)| \leqq \phi_1(t)\phi_1(A_1 t) \cdots \phi_1(A_1^{n-1}t)|x(A_1^n t)|$$
$$+ ||x|| \textstyle\sum' \phi_{i_1}(t)\phi_{i_2}(A_{i_1}t) \cdots \phi_{i_n}(A_{i_{n-1}} \cdots A_{i_1}t) ,$$

where $\sum'$ denotes the summation omitting the term corresponding to

$$i_1=1, \ i_2=1, \cdots, \ i_n=1 .$$

Replacing $|x(A_1^n t)|$ by $-(||x|| - |x(A_1^n t)|) + ||x||$ we obtain

$$|U^n x(t)| \leqq -f_n(t)(||x|| - |x(A_1^n t)|)$$
$$+ ||x|| \textstyle\sum \phi_{i_1}(t)\phi_{i_2}(A_{i_1}t) \cdots \phi_{i_n}(A_{i_{n-1}} \cdots A_{i_1}t) ,$$

or

$$|U^n x(t)| \leqq -f_n(t)(||x|| - |x(A_1^n t)|) + ||x|| , \tag{5.5}$$

since $\sum \phi_i(t) = 1$. Now let $t \in S_1(\varepsilon)$. Then

$$\begin{aligned} |U^n x(t)| &\leqq ||x|| - \delta(||x|| - |x(A_1^n t)|) && \text{by (5.4)} \\ &\leqq ||x|| - \delta\left(||x|| - \frac{1}{q}||x||\right) && \text{by (5.3)} \\ &= ||x|| \left(1 - \frac{q-1}{q}\delta\right) \\ &< \varepsilon . && \text{by (5.1)} \end{aligned}$$

Hence the lemma is proved.

THEOREM 5.1. *If $x(t) \in \mathfrak{C}(\Omega)$, then $\{U^n x(t)\}$ form an equicontinuous family of functions at each of the absorbing points $\{\tau_i\}$ $(i=1, 2, \cdots, m)$.*

Proof. Without loss of generality, we prove the theorem for the point τ_1. It is required to prove that given $\varepsilon > 0$, there exists a sphere $S_1(\varepsilon)$ with centre τ_1 such that for $t \in S_1(\varepsilon)$

$$|U^n x(t) - U^n x(\tau_1)| < \varepsilon \qquad n=0, 1, 2, \cdots ,$$

or equivalently by Lemma 3.1

$$(5.5) \qquad |U^n x(t) - x(\tau_1)| < \varepsilon \qquad n=0, 1, 2, \cdots ,$$

for $t \in S_1(\varepsilon)$.

Let $z(t) = x(t) - x(\tau_1)$. $z(t) \in \mathfrak{C}(\Omega)$ and $z(\tau_1)=0$. Hence Lemma 5.1 may be applied to obtain a sphere $S_1(\varepsilon)$ with centre τ_1 on which

$$|U^n z(t)| < \varepsilon \qquad n=0, 1, 2, \cdots ,$$

but since U preserves constant functions (by Lemma 2.2)

$$U^n z(t) = U^n x(t) - x(\tau_1) .$$

Hence (5.5) is established and the theorem is proved.

6. The convergence theorem in $\mathfrak{C}(\Omega)$. In this section, the assumption (3.3) is applied in conjunction with Theorem 5.1 to obtain the convergence of $U^n x(t)$ in $\mathfrak{C}(\Omega)$.

LEMMA 6.1. *Let $\{S_i\}$ be spheres with centres $\{\tau_i\}$ such that $\bar{S}_i \subset \sum_i$ $(i=1, 2, \cdots, m)$. Then there exists a positive integer n_0 and a number δ, $(0 < \delta \leq 1)$ such that for each $t \in \Omega$, there exists a sequence of n_0 transformations $A_{i_1}, A_{i_2}, \cdots, A_{i_{n_0}}$ (depending on t) which, when applied consecutively, transform t into a point in $\bigcup_{i=1}^{m} S_i$, and such that the probability of the application of each transformation of the sequence is $\geq \delta$, that is, each term in the finite sequence*

$$\phi_{i_1}(t), \ \phi_{i_2}(A_{i_1} t), \ \cdots, \ \phi_{i_{n_0}}(A_{i_{n_0-1}}, \cdots, A_{i_1} t)$$

is $\geq \delta$.

Proof. By assumption (3.2), it is clear that

$$(6.1) \qquad \phi_i(t) > 0 \quad \text{on } \textstyle\sum_i \text{ (and hence on } \bar{S}_i) \quad i=1, 2, \cdots, m.$$

and thus by the continuity of $\phi_i(t)$, there exists δ_0 such that $0 < \delta_0 \leq 1$ and

(6.2) $$\phi_i(t) \geqq \delta_0 \quad \text{for} \quad t \in \bar{S}_i \qquad i=1, 2, \cdots, m$$

Let $t \in \Omega$. If $t \in \Omega - \bigcup_1^m S_i$, we have by the assumption (3.3) together with (2.1), (2.2) and (6.1) that there exists a finite chain of transformations $A_{j_1}, A_{j_2}, \cdots, A_{j_n}$, which when applied consecutively, transform t into a point in one of the spheres, say S_l $(1 \leqq l \leqq m)$. Moreover each term in the sequence $\phi_{j_1}(t), \phi_{j_2}(A_{j_1}t), \cdots, \phi_{j_n}(A_{j_{n-1}} \cdots A_{j_1}t)$ is > 0. If $t \in \bigcup_1^m S_i$, the same result holds for then $t \in S_l$ $(1 \leqq l \leqq m)$ and thus by (2.1) and (6.1), it is sufficient to take a chain consisting of the single transformation A_l.

Consider $A_{j_n}^{-1}S_l$. This is an open set containing $A_{j_{n-1}} \cdots A_{j_1}t$. Since $\phi_{j_n}(A_{j_{n-1}} \cdots A_{j_1}t) > 0$, there exists an open set U_n such that

$$A_{j_{n-1}} \cdots A_{j_1}t \in U_n \subset A_{j_n}^{-1}S_l .$$

and on which $\phi_{j_n}(t) > 0$. By the regularity of Ω, there exists an open set V_n such that

$$A_{j_{n-1}} \cdots A_{j_1}t \in V_n \subset \bar{V}_n \subset U_n \subset A_{j_n}^{-1}S_l .$$

$\bar{V}_n$ is compact and therefore there exists a positive number δ_n such that $\phi_{j_n}(t) \geqq \delta_n$ on $\bar{V}_n$ and hence in particular on V_n.

Now consider $A_{j_{n-1}}^{-1}V_n$. Proceeding as above, we obtain an open set V_{n-1}, such that

$$A_{j_{n-2}} \cdots A_{j_1}t \in V_{n-1} \subset \bar{V}_{n-1} \subset A_{j_{n-1}}^{-1}V_n$$

and a positive number δ_{n-1}, such that $\phi_{j_{n-1}}(t) \geqq \delta_{n-1}$ on V_{n-1}. Proceeding in this manner, we arrive at an open set V_1 which is such that

$$t \in V_1 \subset \bar{V}_1 \subset A_{j_1}^{-1}V_2$$

and such that $\phi_{j_1}(t) \geqq \delta_1$ on V_1, where $\delta_1 > 0$.

Hence, the open set V_1 containing t has the property that each point in it is transformed by the sequence $A_{j_1}, A_{j_2}, \cdots, A_{j_n}$ into a point of S_l $\left(\subset \bigcup_1^m S_i\right)$ and the conditional probabilities of each of the successive transformations being applied are $\geqq \delta_1, \delta_2, \cdots, \delta_n$ respectively.

This process is repeated for every $t \in \Omega$. For each t an open set corresponding to V_1 is obtained. By the compactness of Ω, we have that Ω is covered by a finite number of open sets Ω_l $(l=1, 2, \cdots, k)$, where each set Ω_l has the property that there is a finite chain of transformations of length n_l (that is, n_l=the number of transformations in the chain) which when applied successively transform each point of Ω_l into one of the sphere S_i $(i=1, 2, \cdots, m)$ say S_l, and which has the property that the conditional probabilities of applying the transformations of the chain are respectively $\geqq \delta_{l_1}, \delta_{l_2}, \cdots, \delta_{l_{n_l}}$, where each of these numbers is greater than zero.

Let $n_0 = \max_{1 \leqq l \leqq k} n_l$. The length of the chain n_l for each Ω_l may be extended to n_0 preserving the above properties. For if $t \in S_l$ $(1 \leqq l \leqq m)$, $A_l t \in S_l$ by (2.1) and $\phi_l(t) \geqq \delta_0$ by (6.2).

Let

$$\delta = \min_{\substack{1 \leqq i \leqq n_l \\ 1 \leqq l \leqq k}} (\delta_{l_i}, \delta_0) .$$

With these values of n_0 and δ, the lemma is established.

LEMMA 6.2. *Let $\{x_p(t)\}$ be a sequence of functions in $\mathfrak{C}(\Omega)$ with the following properties*

(6.3) $$\|x_p\| \leqq H \qquad p = 0, 1, 2, \cdots$$

where H is a constant.

(6.4) $$x_p(\tau_i) = 0 \quad \textit{for all } p \qquad i = 1, 2, \cdots, m$$

and

(6.5) *the family of functions $\{U^n x_p(t)\}$ $(n, p = 0, 1, 2, \cdots)$ is equicontinuous at each of the absorbing points τ_i $(i = 1, 2, \cdots, m)$.*

Then, under these conditions

$$\lim_{n \to \infty} \|U^n x_p\| = 0$$

where the convergence is uniform with respect to p.

Proof. Given $\varepsilon > 0$, there exist by (6.5) spheres $S_i(\varepsilon)$ with centres τ_i $(i = 1, 2, \cdots, m)$ such that for $t \in S_i(\varepsilon)$

$$|U^n x_p(t) - U^n x_p(\tau_i)| < \varepsilon/2 \qquad \text{all } n, p .$$

Hence by (6.4) and Lemma (3.1)

(6.6) $$|U^n x_p(t)| < \varepsilon/2 , \qquad t \in \bigcup_1^m S_i(\varepsilon), \text{ all } n, p.$$

There is no loss in generality in assuming the spheres $S_i(\varepsilon)$ so chosen that

(6.7) $$\overline{S_i(\varepsilon)} \subset \textstyle\sum_i \qquad i = 1, 2, \cdots, m.$$

Thus the spheres $S_i(\varepsilon)$ $(i = 1, 2, \cdots, m)$ satisfy the hypothesis of Lemma 6.1. The positive integer n_0 and the positive number δ obtained in the lemma depend here on ε. Let

$$\alpha=1-\frac{1}{2}\delta^{n_0}. \tag{6.8}$$

Since $0<\delta\leqq 1$, it follows that $0<\alpha<1$. We now show that for all p

$$||U^{kn_0}x_p||\leqq\mu_k \qquad k=0,1,2,\cdots \tag{6.9}$$

where $\mu_k=\max(\alpha^k H, \varepsilon)$.

We prove (6.9) by induction. Clearly by (6.3) it is true for $k=0$. Suppose it is true for k.

$$\begin{aligned} U^{(k+1)n_0}x_p(t) &= \sum_{i_1 i_2\cdots i_{n_0}} \phi_{i_1}(t)\phi_{i_2}(A_{i_1}t)\cdots\phi_{i_{n_0}}(A_{i_{n_0-1}}\cdots A_{i_1}t) \\ &\qquad\times U^{kn_0}x_p(A_{i_{n_0}}\cdots A_{i_1}t). \end{aligned} \tag{6.10}$$

Consider t fixed. By Lemma 6.1, there is associated with t a finite sequence of n_0 transformations $A_{j_1}, A_{j_2}, \cdots, A_{j_{n_0}}$ (depending on t) which when applied consecutively transform t into $\bigcup_{i=1}^{m} S_i(\varepsilon)$ and such that each term of the finite sequence $\phi_{j_1}(t)$, $\phi_{j_2}(A_{j_1}t)$, $\cdots$, $\phi_{j_{n_0}}(A_{j_{n_0-1}}\cdots A_{j_1}t)$ is $\geqq\delta$, that is,

$$A_{j_{n_0}}A_{j_{n_0-1}}\cdots A_{j_1}t\in\bigcup_{1}^{m} S_i(\varepsilon) \tag{6.11}$$

and

$$\phi_{j_1}(t)\phi_{j_2}(A_{j_1}t)\cdots\phi_{j_{n_0}}(A_{j_{n_0-1}}\cdots A_{j_1}t)\geqq\delta^{n_0}. \tag{6.12}$$

In (6.10) we take inequalities with absolute values and separate out the term corresponding to the above sequence and proceed as in § 5 (between the relations (5.4) and (5.5)) to obtain by the induction hypothesis

$$\begin{aligned} |U^{(k+1)n_0}x_p(t)| &\leqq -\phi_{j_1}(t)\phi_{j_2}(A_{j_1}t)\cdots\phi_{j_n}(A_{j_{n-1}}\cdots A_{j_1}t) \\ &\qquad\times(\mu_k-|U^{kn_0}x_p(A_{j_n}\cdots A_{j_1}t)|)+\mu_k. \end{aligned} \tag{6.13}$$

(6.6), (6.11) and (6.12) give

$$|U^{(k+1)n_0}x_p(t)|\leqq-\delta^{n_0}(\mu_k-\varepsilon/2)+\mu_k.$$

Since $\mu_k\geqq\varepsilon$, we have $\mu_k-\varepsilon/2\geqq 1/2\mu_k$. Hence

$$|U^{(k+1)n_0}x_p(t)| \leqq \mu_k\left(1-\frac{1}{2}\delta^{n_0}\right)=\alpha\mu_k \qquad \text{by (6.8).}$$

Therefore

$$||U^{(k+1)n_0}x_p|| \leqq \alpha\mu_k \leqq \max(\alpha^{k+1}H,\ \varepsilon)=\mu_{k+1}\,.$$

Hence (6.9) is established. Clearly there exists k_0 sufficiently large such that $\mu_{k_0} \leqq \varepsilon$. Then, since $||U^n||=1$, all n (Lemma 2.2), we have that for $n \geqq n_0k_0$

$$||U^n x_p|| \leqq \varepsilon \qquad \text{all } p.$$

Hence the lemma is proved.

THEOREM 6.1. *U^n converges strongly on $\mathfrak{C}(\Omega)$, that is, there exists a continuous transformation U_∞ of norm 1 of $\mathfrak{C}(\Omega)$ into itself (which preserves constant function) such that*

$$\lim_{n\to\infty} ||U^n x - U_\infty x||=0 \tag{6.14}$$

for each function $x \in \mathfrak{C}(\Omega)$.

Proof. Given $x(t) \in \mathfrak{C}(\Omega)$, let

$$x_p(t)=U^p x(t)-x(t) \qquad p=0,\ 1,\ 2,\cdots\,. \tag{6.15}$$

Clearly $x_p(t) \in \mathfrak{C}(\Omega)$ and $||x_p|| \leqq 2||x||$. Moreover by Lemma 3.1, $x_p(\tau_i)=0$ ($i=1,\ 2,\cdots,\ m$). Hence the hypothesis (6.3) and (6.4) of Lemma (6.2) are verified for the family $\{x_p(t)\}$. It remains to verify (6.5).

Given $\varepsilon > 0$, we have by Theorem 5.1 that there exists spheres $S_i(\varepsilon)$ with centres τ_i ($i=1,\ 2,\cdots,\ m$) such that for $n=0,\ 1,\ 2,\cdots$ we have

$$|U^n x(t) - x(\tau_i)| < \varepsilon/2\,, \qquad t \in S_i(\varepsilon),\ i=1,\ 2,\cdots,\ m.$$

Hence for $t \in S_i(\varepsilon)$ ($1 \leqq i \leqq m$) and all n, p

$$|U^{n+p}x(t) - U^n x(t)| < \varepsilon$$

or

$$|U^n x_p(t)| < \varepsilon \qquad \text{all } n,\ p,\ t \in \bigcup_1^m S_i(\varepsilon)\,. \tag{6.16}$$

Since $x_p(\tau_i)=0$, $U^n x_p(\tau_i)=0$ (Lemma 3.1) and thus it is clear from (6.16) that the hypothesis (6.5) of Lemma (6.2) is verified.

Hence applying Lemma 6.2 to the family $x_p(t)$ as defined by (6.15) we have that given $\varepsilon > 0$, there exists n such that

$$||U^n x_p|| < \varepsilon \qquad p=0,\ 1,\ 2,\cdots\,,$$

or

$$||U^{n+p}x - U^n x|| < \varepsilon \quad \text{for all positive integers } p.$$

Hence since the space $\mathfrak{C}(\Omega)$ is complete, there exists an element $U_\infty x \in \mathfrak{C}(\Omega)$ such that

$$\lim_{n\to\infty} ||U^n x - U_\infty x|| = 0 .$$

U_∞ is clearly a linear transformation of $\mathfrak{C}(\Omega)$ into itself. Since

$$||U_\infty x|| = \lim_{n\to\infty} ||U^n x|| \leqq ||x|| ,$$

it follows that U_∞ is continuous and $||U_\infty|| \leqq 1$. However since U preserves the constant functions on Ω, it is clear that U_∞ does likewise and hence $||U_\infty|| = 1$. Hence the theorem is proved.

7. **The form of $U_\infty x$.** The following lemma is a direct consequence of Lemma 3.1 and Theorem 6.1.

LEMMA 7.1. *U_∞ preserves the values at the absorbing points, that is,*

$$U_\infty x(\tau_i) = x(\tau_i) \qquad i = 1, 2, \cdots, m$$

where $x(t) \in \mathfrak{C}(\Omega)$.

LEMMA 7.2. *If $x(t)$ is a fixed point of U in $\mathfrak{C}(\Omega)$ having the value zero at each of the absorbing points τ_i $(i=1, 2, \cdots, m)$, then $x(t) \equiv 0$. Two continuous fixed points of U which are equal at each τ_i $(i=1, 2, \cdots, m)$ are identical.*

Proof. Let $x(t)$ be a fixed point of U with $x(\tau_i)=0$ $(i=1, 2, \cdots, m)$. We apply Lemma 6.2 to the family of functions consisting of the single function $x(t)$. Since $U^n x = x$ all n, the conditions of the lemma are trivially satisfied and hence $\lim_{n\to\infty} ||U^n x|| = 0$, that is, $||x|| = 0$. Therefore the first part of the lemma is proved.

If $x(t)$, $y(t)$ are two fixed points in $\mathfrak{C}(\Omega)$ such that $x(\tau_i) = y(\tau_i)$ $(i=1, 2, \cdots, m)$ then, applying the first part of the lemma to the function $z(t) = x(t) - y(t)$, we obtain $z(t) \equiv 0$. Hence the lemma is proved.

LEMMA 7.3. *Let $\psi_i(t) = U_\infty \phi_i(t)$ $(1 \leqq i < \infty)$. Then $\psi_i(t)$ is a fixed point of U in $\mathfrak{C}(\Omega)$. If $i > m$, $\psi_i(t) \equiv 0$. If $i \leqq m$, $\psi_i(t)$ is the unique fixed point of U having the value 1 at τ_i and the value zero at each of the other absorbing points τ_j $(j \neq i,\ 1 \leqq j \leqq m)$. Moreover*

$$\sum_1^m \psi_i(t) = 1 . \tag{7.1}$$

Proof. $$U\psi_i = UU_\infty\phi_i = U_\infty\phi_i = \psi_i \qquad (1 \leqq i < \infty)$$

since $UU_\infty = U_\infty$ by Theorem 6.1. Hence ψ_i is a fixed point of U. For $i > m$, $\phi_i(t)$ has the value zero at each of the absorbing points τ_j $(1 \leqq j \leqq m)$ (by 3.5) and hence by Lemma 7.1 $\psi_i(t)$ has the same property and thus, by Lemma 7.2, is identically zero. If $1 \leqq i \leqq m$, then since $\phi_i(t)$ has the value 1 at τ_i (by (3.1)) and the value zero at each τ_j $(j \neq i, 1 \leqq j \leqq m)$ by (3.5), we have by Lemma 7.1 that ψ_i has the same properties and hence since ψ_i is a fixed point of U, by Lemma 7.2 it is the unique fixed point with these values at the vertices.

By (2.5) $\sum_1^\infty \phi_i(t) = 1$. By Dini's theorem, the convergence is uniform so that we have in the sense of $\mathfrak{C}(\Omega)$, $\sum_1^\infty \phi_i = 1$. Hence since U_∞ is continuous $\sum_1^\infty \psi_i = 1$, and since $\psi_i(t) = 0$ $(i > m)$, $\sum_1^m \psi_i(t) = 1$ and the lemma is established.

THEOREM 7.1. *If $x \in \mathfrak{C}(\Omega)$ then*

$$U_\infty x = \sum_{i=1}^m x(\tau_i)\psi_i\,.$$

Proof. Let

$$y(t) = \sum_{i=1}^m x(\tau_i)\psi_i(t)\,.$$

Clearly by Lemma 7.3, y is a fixed point of U such that $y(\tau_i) = x(\tau_i)$ $(i = 1, \cdots, m)$. By Theorem 6.1, $U_\infty x$ is a fixed point of U and by Lemma 7.1 $U_\infty x(\tau_i) = x(\tau_i)$ $(i = 1, \cdots, m)$. Hence by Lemma 7.2, $y = U_\infty x$ and the theorem is proved.

8. The convergence theorem in $\mathfrak{M}(\Omega)$.

THEOREM 8.1. *Let*

$$\mu \in \mathfrak{M}(\Omega)\,,$$

then

$$T^n\mu \rightharpoonup T_\infty\mu$$

where the half-arrow denotes weak-star convergence, that is,

$$\lim_{n\to\infty} (x,\ T^n\mu) = (x,\ T_\infty\mu)\,, \qquad x \in \mathfrak{C}(\Omega) \tag{8.1}$$

where T_∞ is a positive continuous linear tranformation of norm 1 of $\mathfrak{M}(\Omega)$ into itself. T_∞ is the adjoint of U_∞ and

$$T_\infty \mu = \sum_1^m \left(\int_\Omega \psi_i(t) d\mu(t) \right) \delta_i \tag{8.2}$$

where δ_i is the probability measure with all its measure concentrated at the point τ_i.

Proof. Theorem 6.1 gives that for $x \in \mathfrak{C}(\Omega)$, we have $\lim_{n\to\infty} U^n x = U_\infty x$, the convergence being in the sense of $\mathfrak{C}(\Omega)$. It follows that $\lim_{n\to\infty} (U^n x, \mu) = (U_\infty x, \mu)$ for $\mu \in \mathfrak{M}(\Omega)$. Let T_∞ be the adjoint of U_∞ that is, $(U_\infty x, \mu) = (x, T_\infty \mu)$. Hence by (2.10) $\lim_{n\to\infty} (x, T^n \mu) = (x, T_\infty \mu)$ and (8.1) is established. T_∞ being the adjoint of U_∞ is a continuous linear transformation of $\mathfrak{M}(\Omega)$ into itself of norm 1 and it is clearly positive. By Theorem 7.1 $U_\infty x = \sum_1^m x(\tau_i)\psi_i$. Hence for $\mu \in \mathfrak{M}(\Omega)$, $(U_\infty x, \mu) = \sum_1^m x(\tau_i)(\psi_i, \mu)$. Let $T'\mu = \sum_1^m (\psi_i, \mu)\delta_i$. $T'\mu$ is an element of $\mathfrak{M}(\Omega)$ and it is clear that $(U_\infty x, \mu) = (x, T'\mu)$ for all $x \in \mathfrak{C}(\Omega)$. Hence $T' = U_\infty^* = T_\infty$. Thus $T_\infty \mu = \sum_1^m (\psi_i, \mu)\delta_i$ and the theorem is proved.

9. Probability interpretation of $\psi_i(t)$. It is easy to see from the definition (2.9) of U, that $U^n \phi_i(t)$ represents the probability, that given the initial state of the system is t, that at the end of the $(n+1)$st unit time interval the transformation A_i is applied. $\psi_i(t) = \lim_{n\to\infty} U^n \phi_i(t)$ thus represents the limiting probability of applying A_i, given that initially the state of the system is t.

Another point of view is obtained from (8.2). If δ_{t_0} is the probability measure concentrated at the single point t_0, then $T_\infty \delta_{t_0} = \sum_1^m \psi_1(t_0)\delta_i$, so that $\psi_i(t_0)$ gives the probability that if the initial state is t_0, the limiting state is τ_i.

To sum up, we have two probability interpretations for $\psi_i(t)$:

(1) Limiting probability as $n \to \infty$ that at the nth step in the process, the transformation A_i is applied, given that the initial state is t.

(2) Probability that the limiting state is τ_i, given that the initial state is t.

I wish to express my thanks to the referee for some useful comments.

References

1. R. Bellman, T. Harris and H. N. Shapiro. *Studies on functional equations occurring in decision processes*, RM 878, Rand corporation July, 1952.
2. R. R. Bush and F. Mosteller, *Stochastic models for learning*, Wiley 1955.
3. W. Doeblin and R. Fortet, *Sur les châines à liaisons complètes*, Bull. Soc. Math. France **65** (1937), 132–148.
4. R. Fortet, (Thèse), *Sur l'iteration des substitutions algébrique lineáires à une infinité de variables*, Revista, Nos. 424–426, Año 40, Lima, 1938.
5. P. R. Halmos, *Measure theory*, Van Nostrand 1950.
6. C. T. Ionescu Tulcea and G. Marinescu, *Sur certaines châines à liaisons complètes*, C. R. Acad. Sci. Paris, **227** (1948), 667–669.
7. ———, *Theorie ergodique pour des classes d'operations non-completement continues*, Ann. of Math. **52** (1950), 140–147.
8. S. Kakutani, *Concrete representation of abstract (M)-spaces*, Ann. of Math. **42** (1941), 994–1023.
9. S. Karlin, *Some random walks arising in learning models I*, Pacific J. Math. **3** (1953), 725–756.
10. O. Onicescu and G. Mihoc, *Sur les châines de variables statistiques*, Bull. Sci. Math. 2, **59** (1935), 174–192.

UNIVERSITY COLLEGE, DUBLIN
NATIONAL UNIVERSITY OF IRELAND

ON A NEW RECIPROCITY, DISTRIBUTION AND DUALITY LAW

G. KUREPA

Introduction. One knows various operations on sets, e. g. join, intersection, limit, A-operation (Suslin), etc. In the present article we define, as an extension of operations we introduced in another paper (Kurepa [6], [7]) several operations of considerable generality and importance. It turns out that the well-known *distribution law* (cf. § 11) as well as the *De Morgan duality principle* (cf. § 5) are very special cases of our theorems. Moreover, a new *reciprocity phenomenon* occurs (cf. § 12). All depend on the interconnection between *maximal chains* and *maximal antichains* of ordered sets. By considering ordered sets one achieves considerable generality. By their use we get a *synthetic view* on (1) the *analytic operation*; (2) *c-analytic operation* (definition of complements of analytic sets); (3) the *distribution law*; (4) the *duality law*; and moreover, one arrives at (5) a *new reciprocity law*. In particular, in connection with the distributive law, the maximal chains and maximal antichains indicate respectively two distinct ways to reach the same result (cf. Theorems 4.2, 8.1). On the other hand, the parallel considerations of maximal chains and maximal antichains of S give rise to a new kind of interconnection of elements of $P^2 1$ (1 being any set; cf. the k-condition in § 8). This in turn opens a broad way to new investigations by consideration of the elements of $P^\alpha 1$ instead of those of $P^2 1$. Our results may be interpreted in mathematical logic too.

The results of this paper are connected to an idea we expressed in our Thesis [4], 135 n°40 (cf. A. Tarski [11]).

GLOSSARY AND NOTATIONS

Antichain; an ordered set having no couple of distinct comparable points.

Chain; an ordered set having no two distinct incomparable points.

1 or U means universal set.

γT (cf. 10.1)

Disjunctive family; a family composed of pairwise disjoint sets.

ε' denotes " not ε."

Received September 8, 1953 and in Revised form July 24, 1954. The main results of this paper were presented August 28, 1953, in Bruxelles at a Colloquium of Mathematical Logic (Bruxelles, August 18–19, 28–29, 1953).

The author wishes to express his sincere thanks to the referee for his very attentive examination of this paper and for permission to include here Theorems 4.4 and 10.1.

j-connected (cf. 2.1)

k-condition (cf. 3.1., 8.,)

PS denotes the system of all subsets of S; in particular, the void set v is an element of PS; $P^2S=P(PS)$, $P^{\alpha+1}S=P(P^{\alpha}S)$, etc.

$\Omega \in \{O, O'\}$, $\bar{\Omega} \in \{\bar{O}, \bar{O'}\}$

$\bigcap'=\bigcup$, $\bigcup'=\bigcap$.

$\prod$ denotes the combinatorial multiplication.

Ramified set; an ordered set the predecessors of each of whose points form a chain.

Ramified table or tree; an ordered set S with the property that if $x \in S$ then the set $(.,x)_S$ is well-ordered.

ρ being a relation, ρ_1, ρ_2, ρ_3 designates its first part, second part, third part, e. g., in the equality (2) we use $(2)_1$ to designate the first (left) part of (2); $(2)_2$ designates the second part of (2). If (2) is a binary relation for sets, then $(2)_1$ is the set on the left side of (2).

$(x,.)_S$ denotes the set of all the points $y \in S$ such that $x < y$.

$(.,x)_S$ denotes the set of all the points $y \in S$ such that $y < x$.

$\perp$ denotes $\bigcap$ or $\bigcup$.

v=empty set.

1. The operator $(e, \perp, f)$**.** Let $e \in P^2 1$ and $\perp \in \{\bigcap, \bigcup\}$. Let f be any mapping of 1. This means that, for each $x \in 1$, $f(x)$ is a well-determined set; of course it may happen that $f(x)=v$ (void); by f' we denote the mapping $x \to f'(x)$ which to each $x \in 1$ associates the complement $f'(x)$ of the set $f(x)$; the complement is taken in respect to any set $\supseteqq f(x)$ $(x \in 1)$. In the case that $f(x)$ consists of one point, say $f(x)=\{a\}$, we write $f(x)=a$ as well as $f(x)=\{a\}$. Let $\perp$ denote $\bigcup$ or $\bigcap$; let $\bigcup'=\bigcap$, $\bigcap'=\bigcup$.

We put

$$(1.1) \qquad (e, \perp, f)=\underset{e_1}{\perp'}\ \underset{e_0}{\perp}\ f(e_0) \qquad (e_0 \in e_1 \in e).$$

In particular, we put, by convention,

$$(1.2) \quad (v, \bigcap, f)=v, \ (v, \bigcup, f)=\text{universal set} \supseteqq f(e) \text{ for each } e \in 1.$$

More explicitly (1.1) reads

$$(1.3) \qquad (e, \bigcap, f)=\bigcup_{e_1} \bigcap_{e_0} f(e_0), \quad (e, \bigcup, f)=\bigcap_{e_1} \bigcup_{e_0} f(e_0)$$

where $e_0 \in e_1 \in e$. Thus, $e_0 \in 1$, $e_1 \in P1$.

The meaning of $(E, \perp, f')$, $(F, \perp, f')$ is obvious. Thus, $f'(x)$ denotes the complement of $f(x)$. In particular, one has the De Morgan Theorem.

THEOREM 1.1. $(e, \perp, f)' = (e, \perp', f')$.

In what follows, we shall denote by

$$(1.4) \qquad (e, e^*)$$

any ordered pair of elements of $P^2 1$. Given such a pair (e, e^*) we might consider various sets, as e.g.,

$$(1.5) \qquad (e, \cap, f),\ (e, \cap, f'),\ (e, \cup, f),\ (e, \cup, f'),$$

and similarly for e^*. In particular, we shall consider the sets

$$(1.6) \qquad (e, \perp, f),\ (e^*, \perp, f').$$

Obviously, given $e, \perp, f$, the previous sets are well determined. The problem is to know their interconnections.

2. j-connection of (e, e^*).

THEOREM 2.1. *In order that for each* f

$$(2.1) \qquad (e, \cap, f)' \supseteqq (e^*, \cap, f') \text{ or } (e, \cap', f') \supseteqq (e^*, \cap, f'),$$

it is necessary and sufficient that

$$(2.2) \qquad e_1 \cap e_1^* \neq v \qquad (e_1 \in e,\ e_1^* \in e^*).$$

Proof of necessity (2.1)$\Rightarrow$(2.2). Suppose, on the contrary, that (2.2) does not hold; i.e., that there exist

$$(2.3) \qquad e_{1_0} \in e,\ e_{1_0}^* \in e^*, \text{ so that } e_{1_0} \cap e_{1_0}^* = v.$$

Let f be the characteristic function of e_{1_0} such that $f(e_0)=1 \Longleftrightarrow e_0 \in e_{1_0}$. Since $e_{1_0} \in e$ and since $1 \in f(e_0)$ $(e_0 \in e_{1_0})$ one has obviously $1 \in (2.1)_1$. On the other hand, since $e_{1_0}^* \cap e_{1_0} = v$, $f(e_0^*) = v$ $(e_0^* \in e_{1_0}^*)$, thus $f'(e^*) = 1$ $(e_0^* \in e_{1_0}^*)$; in other words, $1 \in (2.1)_2$. Thus (2.3) implies $1 \in (2.1)_2 \setminus (2.1)_1$ which contradicts the hypothesis (2.1).

Proof of sufficiency. (2.2)$\Rightarrow$(2.1), that is, (2.2)$\Rightarrow$($\xi \in (2.1)_2) \Rightarrow \xi \in (2.1)_1$). Now the relation $\xi \in (e^*, \cap, f')$ means that there is a e_1^* such that $\xi \in f'(e_0^*)$ $(e_0^* \in e_1^*)$.

Again, let $e_1 \in e$; since $e_1 \cap e_1^* \neq v$ by hypothesis (2.2), let $z \in e_1 \cap e_1^*$; thus, $\xi \in {}'f(z)$; consequently, for each $e_1 \in e$ there is an $e_0 \in e_1$ such that $\xi \in {}'f(e_0)$. That means $\xi \in {}'(e, \cap, f)$, that is, $\xi \in (e, \cap, f)'$.

Since the condition (2.2) is symmetrical with respect to e, e^*, we get the following.

THEOREM 2.2. *The f-identity* $(e, \bigcap, f)' \supseteqq (e^*, \bigcap, f')$ *is equivalent to the f-identity* $(e^*, \bigcap, f)' \supseteqq (e, \bigcap, f')$.

The last two theorems give rise to the following.

DEFINITION 2.1. An ordered pair (e, e^*) of elements of $P^2 1$ is said to be *j-connected*, symbolically $(e, e^*) \in (j)$ if

$$e_1 \bigcap e_1^* \neq v, \qquad (e_1 \in e, \ e_1^* \in e^*).$$

THEOREM 2.3. *In order that* (2.1) *holds for each* f, *it is necessary and sufficient that the ordered pair* (e, e^*) *be j-connected.*

3. The k-condition. We will prove the following.

THEOREM 3.1. *In order that for each* f *one has*

$$(e, \bigcap, f)' \subseteqq (e^*, \bigcap, f'), \tag{3.1}$$

it is necessary and sufficient that for each $X \subseteqq 1$ *satisfying*

$$X \bigcap e_1 \neq v \qquad (e_1 \in e) \tag{3.2}$$

one has

$$PX \bigcap e^* \neq v, \tag{3.3}$$

that is, that there is an $e_1^* \in e^*$ *such that* $e_1^* \subseteqq X$.

Proof of necessity. Let X satisfy (3.2). Let f be the characteristic function of X. Then (3.2) implies $v \in {}'\bigcap f(e_0)$, $(e_0 \in e_1)$, for each $e_1 \in e$. Thus $v \in (3.1)_1$. As (3.1) holds, one has $v \in (3.1)_2$. Therefore there exists a $e_1^* \in e^*$ satisfying $v \in \bigcap_{e_0^*} f'(e_0^*)$ $(e^* \in e_1^*)$. Consequently, $f(e_0^*)=1$ for each $e_0^* \in e_1^*$, and that means exactly that $e_1^* \subseteqq X$.

Proof of sufficiency. If (3.2)$\Rightarrow$(3.3), then $\xi \in (3.1)_1$ implies $\xi \in (3.1)_2$. Let

$$X = \underset{x \in 1}{E}\,(\xi \in f'(x)), \tag{3.4}$$

that is, X denotes the set of all the $x \in 1$ for which $\xi \in {}'f(x)$. We say that (3.2) holds. In the opposite case, there would be an $e_{1_0} \in e$ such that $e_{1_0} \bigcap X = v$, thus $\xi \in f(e_0)$ $(e_0 \in e_{1_0})$ and therefore $\xi \in {}'(3, 1)_1$, contrary to the hypothesis that $\xi \in (3.1)_1$. The set (3.4) satisfying (3.2), there exists by supposition an element $e_{1_0}^* \in e^*$ such that $e_{1_0}^* \subseteqq X$. That means that $\xi \in f'(e_0^*)$ $(e_0^* \in e_{1_0}^*)$, that is, $\xi \in (3.1)_2$.

DEFINITION 3.1. The ordered pair (e, e^*) of elements of $P^2 1$ is said to satisfy the *k-condition*, symbolically

$$(e, e^*) \in (k), \tag{3.5}$$

provided the system

$$X \subseteqq 1,\ X \cap e_1 \neq v \qquad (e_1 \in e) \tag{3.6}$$

implies

$$PX \cap e^* \neq v. \tag{3.7}$$

Thus Theorem 3.1 may be expressed in the following form.

THEOREM 3.2. *The relation $(e, e^*) \in (k)$ is equivalent to the f-identity*

$$(e, \cap, f)' \subseteqq (e^*, \cap, f').$$

4. First fundamental theorem. Theorems 2.1 and 3.1 enable us to characterize the equality

$$(e, \cap, f)' = (e^*, \cap, f'). \tag{4.1}$$

THEOREM 4.1. *The equality* (4.1) *is equivalent to the relation*

$$(e, e^*) \in (j) \wedge (k). \tag{4.2}$$

(The last relation means that (e, e^*) satisfies both (j) and (k)).

We transform the previous conditions using De Morgan's theorem (c.f. Theorem 1.1). We have $(e, \cap, f)' = (e, \cup, f')$ so that (4.1) reads

$$(e, \cup, f') = (e^*, \cap, f');$$

and considering f' instead of f we obtain

$$(e, \cup, f) = (e^*, \cap, f).$$

Consequently we have the following,

THEOREM 4.2. (FIRST FUNDAMENTAL THEOREM). *Let (e, e^*) be a given ordered pair of elements of $P^2 1$; then the following properties are pairwise equivalent*:

I. $(e, e^*) \in (j) \wedge (k)$

II. For each mapping f of the set 1 the following duality law holds:

$$\Big(\bigcup_{e_1} \bigcap_{e_0} f(e_0)\Big)' = \bigcup_{e_1^*} \bigcap_{e_0^*} f'(e_0^*), \text{ that is, } (e, \cap, f)' = (e^*, \cap, f').$$

III. *For each mapping* f *of the set* 1, *one has the distributive law*

$$\bigcup_{e_1 \in e} \bigcap_{e_0 \in e_1} f(e_0) = \bigcap_{e_1^* \in e^*} \bigcup_{e_0^* \in e^*} f(e_0^*), \text{ that is, } (e, \bigcap, f) = (e^*, \bigcup, f).$$

IV. $(e^*, e) \in (j) \wedge (k)$.

Proof. In fact, I $\Leftrightarrow$ II (Theorem 4.1) and II$\Leftrightarrow$III as was shown by the application of the De Morgan theorem to $(\bigcup, \bigcap, f)'$. It remains to prove that IV is equivalent to I, II and III. First, the implication I $\Rightarrow$ III yields IV $\Rightarrow (e^*, \bigcap, f) = (e, \bigcup, f)$; from here, passing to complement III$'$ of III: $(e^*, \bigcap, f)' = (e, \bigcup, f)'$, that is, $(e^*, \bigcup, f') = (e, \bigcap, f')$. Writting f' instead of f, one gets III. Thus IV $\Rightarrow$ III. Conversely, III$\Rightarrow$III$'$ (by implication III$\Rightarrow$I)$\Rightarrow$IV.

The equivalence I $\Leftrightarrow$ IV gives the following.

THEOREM 4.3. *Symmetry character of* $(j) \wedge (k)$: *If* $(e, e^*) \in (j) \wedge (k)$, *then also* $(e^*, e) \in (j) \wedge (k)$. *In other words, if* $(e, e^*) \in (j)$, *then* $(e, e^*) \in (k) \Leftrightarrow (e^*, e) \in (k)$.

THEOREM 4.4. *Symmetry of the k-property*[1]. *If* $(e, e^*) \in (k)$, *then* $(e^*, e) \in (k)$.

Proof. To begin with, if e is the null set, then for every e^*, $(e, e^*) \in {}'(k)$ and $(e^*, e) \in {}'(k)$. And if the null set is a member of e, then for every e^*, $(e, e^*) \in (k)$ and $(e^*, e) \in (k)$. It remains to consider cases where no sets involved are null. Suppose that $(e, e^*) \in {}'(k)$. Then there exists an x such that for every $e_1 \in e$, $e_1 \cap x \neq v$, and for every $e_1^* \in e^*$, $e_1^* \setminus x \neq v$. Let $y = \bigcup_{e_1^*} (e_1^* \setminus x)$ $(e_1^* \in e^*)$. Then for every $e_1^* \in e^*$, $y \cap e_1^* \neq v$; and if it can be proved that, for every $e_1 \in e$, $e_1 \setminus y \neq v$, it will follow that $(e^*, e) \in {}'(k)$. But for every $e_1 \in e$, $x_1 = e_1 \cap x \neq v$ and $x_1 \cap y = v$. Since $x_1 \neq v$, it follows that $x_1 \setminus y \neq v$ and therefore that $e_1 \setminus y \neq v$.

In what follows, the generality of Theorem 4.2 will be revealed. We will restrict ourselves to ordered sets. There we are naturally led to consider various operators which were the origin of the present investigations (cf. Kurepa [4], [6].)

5. Ordered sets, operators O, $\overline{O}$, O', $\overline{O}'$. Let S be any set ordered by $\leqq$. The operators O, $\overline{O}$, O', $\overline{O}'$ are defined in the following manner:

DEFINITION 5.1. OS designates the system of all *maximal chains* $\subseteqq S$.

[1] Theorem 4.4 and its proof are due to the referee.

DEFINITION $\overline{5.1}$. $\overline{O}S$ designates the system of all *maximal antichains* $\subseteqq S$.

DEFINITION 5.2. $O'S$ designates the system of all $X \in \overline{O}S$ such that

$$X \cap M \neq v \qquad (M \in OS) .$$

DEFINITION $\overline{5.2}$. $\overline{O'}S$ designates the system of all $X \in OS$ such that

$$X \cap A \neq v \qquad (A \in \overline{O}S) .$$

We shall be aware of a certain reciprocity between the notions chain and antichain, and in particular by passing from the system O, O' to the system $\overline{O}$, $\overline{O'}$.

To each ordered set S is associated the set consisting of

$$OS,\ \overline{O}S,\ O'S,\ \overline{O'}S \tag{5.1}$$

which are at most four elements of P^2S. The set (5.1) is of a great importance. Its elements form in a certain sense the spatial forms along which certain operations are to be taken. Each element e of (5.1) is as it were a system of paths for operations $(e, \perp, f)$, $(e, \perp', f)$, etc.

CONVENTION 5.1. The reciprocal of a statement s will be denoted $\bar{s}$. So the reciprocal of the Lemma 5.1 is denoted by Lemma $\overline{5.1}$. If X is a chain, then $\overline{X}$ is an antichain, etc, Here is an example.

LEMMA 5.1. *In order that $X \in O'S$, it is sufficient that X be an antichain of S such that $X \cap M \neq v$ $(M \in OS)$. In other words, if an antichain intersects each maximal chain of S it is necessarily a maximal antichain.*

The reciprocal result is as follows.

LEMMA $\overline{5.1}$. *In order that $X \in \overline{O'}S$, it is sufficient that X be a chain of S such that $X \cap A \neq v$ $(A \in \overline{O}S)$. In other words, if a chain X of S intersects each antichain of S, then X is necessarily a maximal one.*

Proof. Let X be an antichain satisfying $X \cap M \neq v$ $(M \in OS)$. To prove that $X \in O'S$, it is sufficient to prove that X is a *maximal* antichain, i.e., that each $b \in S$ is comparable to some point of X. Now, let $b \in B \in OS$. Then the point $B \cap X$ exists and is the required point of X which is comparable to b.

Reciprocally, let X be a chain such that $X \cap A \neq v$ $(A \in \overline{O}S)$. To prove that X is a maximal chain, suppose, on the contrary, that there is a chain $C \supset X$. Let $d \in C \setminus X$ and let $d \in D \in \overline{O}S$. Then necessarily $D \cap X = v$, because if $x \in D \cap X$, one would have two distinct comparable points d, x in the antichain D.

LEMMA 5.2. *$O'S \subseteqq \overline{O}S$, $\overline{O}'S \subseteqq OS$. (Each of the signs $\subseteqq$ here may be $=$ or $\subset$.) In particular there exists a non-void S such that*[2]

$$O'S = v, \ \overline{O}'S = v \tag{5.2}$$

EXAMPLE 5.1. Let σ_0 denote the system of all non-void bounded well ordered sets of rational numbers ordered by means of the relation $\subseteq$, where[3] (5.3) $x \subseteq y$ or $y \supseteq x$ means that x is an initial portion of y. In that case, $\overline{O}'\sigma_0 = v$, because, e,g., there is no chain in σ_0 intersecting each row of σ_0 (cf. [4, p. 95]). It is probable that $O'\sigma_0 = v$.

As an example of reciprocity considerations let us prove the following lemmas (5.3 and $\overline{5.3}$) which are mutually reciprocal and which will occur in distributive laws (cf. Theorem 9.1, Cases 2,$\overline{2}$).

LEMMA 5.3. *If the maximal chains of S are pairwise disjoint, then the comparability relation in S is transitive, and conversely. Also*

$$O'S = \overline{O}S = \prod_M M, \tag{5.4}$$

where $\prod$ denotes the combinatorial product of sets M, M running over OS; and $OS = \overline{O}'S$.

Peciprocally we have the following.

LEMMA $\overline{5.3}$. *If the maximal antichains of S are pairwise disjoint, then the incomparability relation in S is transitive, and conversely. Also*

$$OS = \overline{O}'S = \prod_A A$$

where $\prod$ denotes the combinatorial product of all the sets A, A running over $\overline{O}S$; and $O'S = \overline{O}S$.

Proof of Lemma 5.3. If OS is disjoint, then as it is easy to show, the comparability relation in S is a congruence relation, and vice versa. Each $A \in \overline{O}S$ intersects each $M \in OS$ (thus $\overline{O}S = O'S$) in a single point,

[2] According to W. Gustin, there exists a denumerable ramified set S satisfying (5.2) [c.f. Gustin, *Math. Rev.* **14**, 255 (1953) in connection with the review of Kurepa [**8**]].

[3] The relation $\subseteq$ is the very basis of the theory of ramified sets (cf. [**4**]).

since on the one hand OS is disjonint and on the other hand A is antichain; thus $A \in (5.4)_3 = \prod_M (M \in OS)$. Conversely, each $X \in (5.4)_3$ is an antichain because of the incomparability of each point of each $M \in OS$ to each point of each $M_0 \in OS$, $M_0 \neq M$. But X is also a maximal antichain. Analogously one proves the reciprocal of Lemma 5.3., that is Lemma $\overline{5.3}$.

REMARK 5.1. On Lemma 5.1 and Lemma $\overline{5.1}$ is based a very general distribution law (c.f. Theorem 9.1, Cases $2,\bar{2}$).

6. Operations $(v, \bigcap, f)$, $(v, \bigcup, f)$ and $(\Omega, \perp, f)$ for each $\Omega \in (5.1)$ and each $\perp \in \{\bigcap, \bigcup\}$.

Let Ω be any element of the set

$$\{OS,\ O'S,\ \overline{O}S,\ \overline{O}'S\}; \tag{6.1}$$

then $\Omega \in P^2S$; so that for each Ω and each $\perp \in \{\bigcap, \bigcup\}$, the operator

$$(\Omega,\ \perp,\ f) \tag{6.2}$$

is well defined. In the particular case that $\Omega = v$, we put

$$(v,\ \bigcup,\ f) = \text{universal set},\quad (v,\ \bigcap,\ f) = \text{void set}. \tag{6.3}$$

We shall consider ordered pairs (e, e^*) of elements of the set (6.1) and the corresponding sets (6.2) for $\Omega = e$ and $\Omega = e^*$, respectively.

EXAMPLE 6.1. Let

$$(T;\ \omega_0) \tag{6.4}$$

denote the system of all $<\omega_0$-complexes (finite complexes) of ordinals $<\omega_0$ ordered by means of the relation $\subseteq$ in (5.3). If f is a mapping of $(T; \omega_0)$ into the family of closed sets, then we can prove that $(O, \bigcap, f)$ and $(O', \bigcap, f')$, respectively, are the most general analytic set (A-set of Suslin) and the most general CA-set respectively (c.f. [10], [1], [2]; also [9]).[4]

Example 6.1 shows the importance of the operations (1.1) even in the particular cases (6.2) and $S = (T; \omega_0)$. (Cf. [6]).

7. Some simple lemmas.

LEMMA 7.1. *Either $O'S = v$ or each element of $O'S$ intersects each element of OS; and reciprocally, either $\overline{O}'S = v$ or $(\overline{O}S, \overline{O}'S)$ is a j-connected ordered*

[4] In our book [5] we defined A-sets just as sets $(OS, \bigcap, f)$ for the choice of S and f as in Example 6.1.

pair.

Lemma 7.1 and Theorem 2.1 yield the following.

THEOREM 7.1. $(OS, \cap, f) \subseteqq (O'S, \cap', f)$

and reciprocally,

$$(\bar{O}, \cap, f) \subseteqq (\bar{O}'S; \cap' f).$$

In general, we have here the sign $\subset$ *instead of* $\subseteqq$. *The duals of that relation hold also.*

THEOREM 7.2. *The two sets,*

$$(OS, \cap, f'), (\bar{O}S, \cap', f)$$

may be non-comparable if S is ramified.

To see this, let D denote the set of all integers ordered as in this diagram:

$$\cdots \rightarrow -4 \rightarrow -2 \rightarrow 0 \rightarrow 2 \rightarrow 4 \rightarrow 6 \rightarrow \cdots$$

$$\cdots \searrow -3 \searrow -1 \searrow 1 \searrow 3 \searrow 5 \searrow 7 \searrow \cdots$$

Obviously, the set D is ramified; for the sets $2D-1$ and $2D$ of all odd and, respectively, even integers one has $2D-1 \in \bar{O}D$, $2D \in OD$, $2D \cap (2D-1) = v$.

Let f be the characteristic function of $2D-1$; one proves then easily that

$$(OD, \cap, f') = \{1\}, (\bar{O}D, \cap', f) = \{0\},$$

and that proves Theorem 7.2.

8. **Ordered sets and k-condition.** If we consider the pair $(OS, O'S)$ or its reciprocal $(\bar{O}S, \bar{O}'S)$, then the j-condition is satisfied; therefore one obtains Theorem 7.1. On the other hand, in general one has neither

$$(OS, O'S) \in (k) \text{ nor reciprocally } (\bar{O}S, \bar{O}'S) \in (\bar{k}).$$

For the sake of simplicity, we present the following.

DEFINITION 8.1. The condition $(OS, O'S) \in (k)$ will be denoted $S \in (k)$ and reciprocally. Thus

$$(OS,\ O'S) \in (k) \Leftrightarrow S \in (k) \tag{8.1}$$

$$(\overline{O}S,\ \overline{O}'S) \in (\overline{k}) \Leftrightarrow S \in (\overline{k}) \tag{$\overline{8.1}$}$$

and we shall say that S *satisfies the* (k)*-condition and the* $(\overline{k})$*-condition* respectively.

In particular, $S \in (k)$ means the statement that each set $\subseteqq S$ which intersects each maximal chain of S contains a maximal antichain of S. Then Theorem 4.2. (implication I $\Rightarrow$ III) yields the following.

THEOREM 8.1. *For each ordered set S satisfying the (k)-condition, one has the following distribution law*[5]:

$$\underset{e_1}{\perp'}\ \underset{e_0}{\perp} f(e_0) = \underset{A}{\perp}\ \underset{a}{\perp'} f(a)\,, \quad (e_0 \in e_1 \in OS,\ a \in A \in O'S) \tag{8.2}$$

and reciprocally for $(\overline{8.2})$. ($\perp$ *designates* $\cap$ *or* $\cup$).

Usual distribution laws are special cases of (8.2). Thus if one takes the ordered set $S=\{1, 2, 3\}$ with diagram $\overset{2}{\underset{1}{\uparrow}}\ {}_3$ one has $OS=\{\{1, 2\}, \{3\}\}$, $O'S=\{\{1, 3\}, \{2, 3\}\}$ and the formula (8.2) yieds $f(3) \perp' (f(1) \perp f(2)) = (f(3) \perp' f(1)) \perp (f(3) \perp' f(2))$.

Analogously, considering the set $\overset{2}{\underset{1}{\uparrow}}\ \overset{4}{\underset{3}{\uparrow}}$ one has the binomial form of the distribution law. For other cases of distribution, cf. § 11.

9. Some classes of ordered sets satisfying (k) and $(\overline{k})$. We are going to prove that the conditions (k), $(\overline{k})$ are satisfied by ordered sets of some general classes—a fact which will give us a general distribution and duality law.

THEOREM 9.1. *The conditions (k) and $(\overline{k})$ are satisfied, provided S satisfies at least one of the following conditions*:

1) *S is a chain;*

$\overline{1}$) *S is an antichain;*

2) *OS is disjoint, i.e., the elements of OS are pairwise disjoint (this is equivalent to the statement that the comparability relation is transitive in S);*

$\overline{2}$) *The elements of $\overline{O}S$ are pairwise disjoint (this is equivalent to the transitivity of the incomparability relation in S).*

The cases 1), $\overline{1}$), 2), $\overline{2}$) are ranged according to relative importance. One sees that 1) and $\overline{1}$) as well as 2) and $\overline{2}$) are mutually reciprocal.

Let us prove, e.g., the case $\overline{2}$). At first, the elements of $\overline{O}S$ being

---

[5] The converse holds also.

pairwise disjoint, by Lemma 5.3, we have $OS = \prod_A A$ $(A \in \overline{O}S)$ and $\overline{O}S = O'S$. Now we prove $S \in (\bar{k})$. If no $A \in \overline{O}S$ were contained in an $X \leqq S$, where X intersects each $M \in OS$, there would be a point $x(A) \in A \setminus X$ for each $A \in \overline{O}S$. The set $\bigcup_A x(A)$ $(A \in \overline{O}S)$ would be a maximal chain of S which does not intersect X, contrary to the hypothesis on X.

REMARK 9.1. Later we shall see that the fact that each chain (antichain) satisfies (k) and $(\bar{k})$ is reflected in the fact that our duality theorem has as a special case the De Morgan duality theorem (cf. Theorem 13.1).

10. The case of ramified tables. At many opportunities we considered ramified tables, i.e., ordered sets satisfying the condition that for each $x \in T$, the set $(.,x)_T$ of all its predecessors in T is well-ordered. Let us recall that for a table T,

$$\gamma T \tag{10.1}$$

denotes the first ordinal number α such that there is no point $x \in T$ such that the order type of $(.,x)_T$ is α; γT is called rank or degree (order) of T.

THEOREM 10.1. *Each ramified table T satisfies (k); or explicitly and more precisely, let T be a set such that for each $x \in T$, the set $(.,x)_T$ is well-ordered. Let $X \leqq T$ and $M \cap X \neq v$ $(M \in OT)$. Then the set*

$$R_0 X \tag{10.2}$$

of all initial points of X is a maximal antichain of T; moreover, R_0X intersects each maximal chain of T. Thus, $R_0X \in O'T$.

THEOREM 10.2. *If $\gamma T < \omega_0$, then $T \in (\bar{k})$ and $OT = \overline{O}'T$, $\overline{O}T = O'T$. In particular, this holds for each finite table.*

Proof of Theorem 10.1. At first, $R_0X \in \overline{O}T$. As R_0X has no pair of distinct comparable points, it is sufficient to show that each $t \in T$ is comparable to a point $x_0(t) \in R_0X$. Now, by hypothesis, there exists at least one point $x(t) \in X$ comparable to t. Let $x_0(t)$ be the point in R_0X which is $\leqq x(t)$. In fact, it $x_0(t) = x(t)$, or if $x(t) \leqq t$, the comparability of t and $x_0(t)$ is obvious. On the other hand, if neither $x_0(t) = x(t)$ nor $x(t) \leqq t$, then $x_0(t) < x(t)$, $t < x(t)$. Thus, $x_0(t)$, t belong to the set $(.,x(t))_T$ which by the supposition on T is a chain.

It remains to prove that R_0X intersects each $M \in OT$. Again, by hypothesis, there exists a point $m \in X \cap M$; then the point $m' \in R_0X$

such that $m' \leqq m$ is a point of M. The set $(.,m]_T \cup M$ is a chain. By virtue of presupposed maximality of M, one has $(.,m]_T \subseteqq M$, thus $m' \in M$.

Proof of Theorem 10.2. At first we have the following.

LEMMA 10.1. *If* $\gamma T < \omega_0$, *then* $A \cap M \neq v$ $(A \in \overline{O}T,\ M \in OT)$, *thus* $OT = \overline{O'}T$, $O'T = \overline{O}T$ (cf. [8]).

Proof. Suppose, on the contrary, that T contains a maximal chain M and a maximal antichain A so that

$$A \cap M = v\ . \tag{10.3}$$

A being a maximal antichain of T, there exists for each $t \in T$ a point $\alpha(t) \in A$ such that $\{t,\ \alpha(t)\}$ is a chain; in particular, for each $m \in M$ the points m, $\alpha(m)$ are comparable. Now

$$m < \alpha(m)\ , \tag{10.4}$$

which is proved as follows. Since $M \in OT$, M is an initial portion of T. Consequently, if (10.4) did not hold, M would then contain also the point $\alpha(m)$ for at least a point $m_0 \in M$. Thus, $\alpha(m_0) \in A \cap M$ contrary to (10.3). Therefore (10.3)$\Rightarrow$(10.4). Now, since $\gamma T < \omega_0$, the chain M is finite.

Let l be the last point of M; l would be a last point of T also, contrary to the relation (10.4) for $m = l$. Thus the relation (10.3) is not possible, and Lemma 10.1 is proved.

To complete the proof of Theorem 10.2, we need to see that each $X \subseteqq T$ satisfying

$$X \cap A \neq v \qquad (A \in \overline{O}T) \tag{10.5}$$

contains a maximal chain of T. This holds for every T and we have the following statement which is reciprocal to Theorem 10.1.

THEOREM $\overline{10.1}$[6]. *Every ramified table* T *satisfies the* $\overline{k}$*-condition*: $T \in (\overline{k})$.

Proof. Suppose X satisfies (10.5). Since $R_0T \in \overline{O}T$, we have

$$X_0 \equiv X \cap R_0T \neq v\ . \tag{10.6}$$

The set (10.6) is an initial portion of X, that is,

$$x \in (10.6) \Rightarrow (.,x]_T \subseteqq (10.6)\ .$$

[6] Theorem $\overline{10.1}$ for $\gamma T \geqq \omega$ is due to the referee.

If X_0 contains a maximal chain of T, then Theorem 10.1 is proved. If $OX_0 \cap OT = v$, then

$$R_0(T \setminus X_0) \tag{10.7}$$

is a maximal antichain of T. As a matter of fact we have the following.

LEMMA 10.2. *If I is an initial portion of a ramified table T such that*

$$OI \cap OT = v\,, \tag{10.8}$$

then

$$R_0(T \setminus I) \in \bar{O}T\,. \tag{10.9}$$

To prove $(10.9)_1 \in \bar{O}T$, it suffices to show that each $i \in T$ is comparable to some point $i' \in (10.9)_1$. Obviously this holds for $i \in T \setminus I$. Suppose $i \in I$. Consider an M such that $i \in M \in OT$. By (10.8), $M \setminus I \neq v$. Let $P \in M \setminus I$, and let i' be the point such that $i' \in (10.9)_1$ and $i' \leqq P$. Since T is ramified and $i < P$, it follows that $i < i'$.

To prove Theorem 10.1, let us consider the sets

$$X_0,\ X_1, \cdots, X_\alpha, \cdots \tag{10.10}$$

defined as follows

$$\begin{aligned} &X_0 = X \cap R_0T,\ X_1 = X_0 \cup (X \cap R_0(T \setminus X_0)) \\ &X_\alpha = X_{\alpha-1} \cup (X \cap R_0(T \setminus X_{\alpha-1})) \\ &X_\alpha = \bigcup_{\alpha_0} X_{\alpha_0} \qquad (\alpha_0 < \alpha) \end{aligned} \tag{10.11}$$

depending upon whether α is isolated or a limit ordinal number.

Obviously, the sequence (10.10) is increasing and its terms $\subseteqq X$. Let δ be the first ordinal such that

$$X_\delta = X_{\delta+1}\,. \tag{10.12}$$

Of course, $\delta \leqq \gamma T$.

We say that

$$OX_\delta \cap OT \neq v \tag{10.13}$$

and hence

$$OX \cap OT \neq v \tag{10.14}$$

because $X_\delta \subseteqq X$.

First, each term of (10.10) is an initial portion of T-provable by an induction argument. Secondly, if the relation (10.13) were false, the set

$$R_0(T\setminus X) \tag{10.15}$$

by virtue of Lemma 10.2 would be a maximal antichain of T. By hypothesis on X (see (10.5)) there would be a point $z \in X \cap (10.15)$. Therefore

$$z \in X_{\delta+1},\ x \in {}'X_\delta .$$

Hence, $z \in X_{\delta+1}\setminus X_\delta$ and $X_\delta \subset X_{\delta+1}$, contrary to (10.12). Hence (10.14) holds and Theorem $\overline{10.1}$ is proved.

11. General distribution laws. To see how the previous investigations are linked with distribution questions, let us prove the following distribution theorem which is the most general distribution law expressible in usual terms.

THEOREM 11.1. *Let $\mathscr{F}$ be any non-void family of non-void sets $\subseteqq 1$, 1 being a standard set. Then for each mapping f of the set 1 we have*

$$\bigcap_{X\in\mathscr{F}}\ \bigcup_{x\in X} f(x) = \bigcup_{A}\ \bigcap_{a\in A} f(a) \qquad (A \in \prod_{X\in\mathscr{F}} X); \tag{11.1}$$

and dually.

Theorem 11.1 is a corollary to Theorem 4.2 (implication I$\Rightarrow$III). As a matter of fact, first the pair $(\mathscr{F}, \prod_{X\in\mathscr{F}} X)$ is j-connected; second it satisfies the k-conditions, as is easily probable.

A direct proof of Theorem 11.1 is as follows.

First, $(11.1)_1 \subseteqq (11.2)_2$, that is, if $\xi \in (11.1)_1$ then $\xi \in (11.1)_2$. In fact $\xi \in (11.1)_1$ means $\xi \in \bigcup_{x\in X} f(x)$ $(X \in \mathscr{F})$, that is, there exists an $X_e \in X$ such that $\xi \in f(X_e)$, $(X \in \mathscr{F})$. Putting $A_e = \bigcup_X X_e (X \in \mathscr{F})$, one has $\xi \in \bigcap_a f(a)$ $(a \in A_e)$ and $A_e \in \prod_X X$, thus $\xi \in (11.1)_2$.

Second, $(11.1)_2 \subseteqq (11.1)_1$: if $\xi \in (11.1)_2$, then $\xi \in (11.1)_1$. The relation $\xi \in (11.1)_2$ is equivalent to $\xi \in f(a)$ $(a \in A)$ for some $A \in \prod_x X$; since $A \cap X \neq v$, this implies $\xi \in \bigcup_{x\in X} f(x)$ for each $X \in \mathscr{F}$, hence $\xi \in (11.1)_1$.

From the proof of Theorem 11.1 we obtain the following interesting result.

THEOREM 11.2. (Cf. Theorem 2.1) *Let $(\mathscr{F}, \mathscr{F}_0)$ be any ordered pair of systems of sets $\subseteqq 1$ such that*

$$(11.2) \qquad X \cap X_0 \neq v \qquad (X \in \mathscr{F},\ X_0 \in \mathscr{F}_0);$$

then for each mapping f:

$$(11.3) \qquad \bigcup_{X \in \mathscr{F}} \bigcap_{x \in X} f(x) \subseteqq \bigcap_{X_0 \in \mathscr{F}_0} \bigcup_{x_0 \in X_0} f(x_0);$$

and dually,

$$\bigcup{}' \bigcap{}' f'(x) \subseteqq' \bigcap_{X^x \in \mathscr{F}^*}{}' \bigcup_{x^* \in X_0^*}{}' f'(x_0)\,.$$

In general, one reads here $\subset$ instead of $\subseteqq$. The case $\mathscr{F} = \mathscr{F}_0$ is not excluded. Therefore the relation (11.3) holds even if one or both sets $\mathscr{F}$, $\mathscr{F}_0$ are vacuous. In particular, (11.3) holds if $\mathscr{F} \in \{OS, O'S, \overline{O}S, \overline{O}'S\}$ and $\mathscr{F}_0 = \mathscr{F}'$ (obviously $(OS)'$ means $O'S$; $(O'S)' = OS$, $(\overline{O}'S)' = \overline{O}S$, even if $O'S = v = \overline{O}'S$. Consequently, we have the following.

THEOREM 11.3. *If* $\Omega \in \{O, \overline{O}, O', \overline{O}'\}$, *then*

$$\bigcup_X \bigcap_x f(x) \subseteqq \bigcap_{X'} \bigcup_{x'} f(x') \qquad (x \in X \in \Omega S, x' \in X' \in \Omega' S)$$

and dually.

Passing to complements in the relation and using the De Morgan formula, we have the following.

THEOREM 11.4. *For any* $\Omega \in \{O, \overline{O}, O', \overline{O}'\}$:

$$\bigcup_X \bigcap_x f(x)' \supseteqq \bigcup_{X_0} \bigcap_{x_0} f'(x_0) \quad (x \in X \in \Omega S,\ x_0 \in X_0 \in \Omega' S).$$

The question of whether sets forming $\mathscr{F}$ in Theorem 11.1 are pairwise disjoint or not disjoint is of no importance. However, without loss of generality, the system $\mathscr{F}$ may be supposed disjoint. In fact, let to each $X \in \mathscr{F}$ be associated the set X_d of all ordered pairs (X, x) $(x \in X)$; to each $x \in X$ we associate the pair (x, x). Instead of $\mathscr{F}$ we can consider the system $\mathscr{F}_d$ of all the X_d $(X \in \mathscr{F})$. Now, the family $\mathscr{F}_d$ is disjunctive and the system $\mathscr{F}_d$ can be interpreted either as OS or as $\bar{O}S$. If one orders totally each X_d and if one orders the set $S = \bigcup_X X_d$ $(X \in \mathscr{F})$ so that each element of X_d is incomparable to each element of each other element of $\mathscr{F}_d$ and if one leaves intact the ordering in each element of $\mathscr{F}_d$, then obviously

$$OS = \mathscr{F}_d\,,\ O'S = \prod_y y \qquad (y \in \mathscr{F}_d);$$

moreover

$$OS=\overline{O}'S,\ O'S=\overline{O}S\,,$$

the set satisfies the conditions (k) and $(\bar{k})$ and accordingly for the set S the distribution law (8.1) holds.

Combining Theorem 8.1 with Theorem 9.1, one has the following statements:

THEOREM 11.5. *If S is an ordered set of one of the cases* 1, $\bar{1}$, 2, $\bar{2}$, *in Theorem* 9.1, *then for each mapping f of S the following distribution law holds*:

(11.4) $$\underset{M}{\perp'}\ \underset{m}{\perp}\ f(m)=\underset{A}{\perp}\ \underset{a}{\perp'}\ f(a) \qquad (m\in M\in OS,\ a\in A\in O'S);$$

and reciprocally. In (11.4) $\perp$ *denotes either* $\bigcap$ *or* $\bigcup$.

Theorems 9.1 and 10.1, 10.2 yield the following.

THEOREM 11.6. *For each ramified table T and each mapping f of T one has*

(11.5) $$\underset{M}{\perp'}\ \underset{m}{\perp}(m)=\underset{A}{\perp}\ \underset{a}{\perp'}\ f(a) \qquad (m\in M\in OT,\ a\in A\in O'T);$$

and reciprocally.

12. A new duality law. We saw (Theorem 8.1) how the distribution law (11.4) is connected with the condition $(\bar{k})$. Now we will see the interconnection of the distribution law and of (k) or $(\bar{k})$ with some duality laws. Let us suppose that for each f one has

$$(OS,\ \perp,\ f)=(O'S,\ \perp'\ f)$$

(this happens if and only if $S\in(k)$ cf. Theorems 3.2, 4.1). In particular, since f is arbitrary, the same equality holds for the mapping f', f' being the complement of f; thus

$$(OS,\ \perp,\ f')=(O'S,\ \perp'\ f')\,.$$

From here, passing to complements, one has

$$(OS,\ \perp,\ f')'=(O'S,\ \perp'\ f')'=(O'S,\ \perp'',\ f'')=(O'S,\ \perp,\ f)\,,$$

(by De Morgan's formula). Thus we have the following.

THEOREM 12.1. GENERAL DUALITY LAW. *For each ordered set $S\in(k)$, one has*

(12.1) $$(\Omega,\ \perp,\ f)'=(\Omega',\ \perp,\ f');\ \textit{where, } \Omega=OS \textit{ or } O'S\,,$$

$$\perp=\bigcap \textit{ or } \bigcup\,.$$

Reciprocally, if $S \in (\bar{k})$, *then for each mapping* f *of* S:

$$(\Omega, \perp, f)' = (\Omega', \perp, f') . \tag{$\overline{12.1}$}$$

Where, Ω *denotes* $\bar{O}S$ *or* $\bar{O}'S$, $\perp = \bigcap$ *or* $\bigcup$.

It is interesting to observe that the converse of Theorem 12.1 holds also.

THEOREM 12.2. $\qquad V(f)\ (12.1) \Leftrightarrow S \in (k)$

$$V(f)\ (\overline{12.1}) \Leftrightarrow S \in (\bar{k}) ,$$

Let us express, e. g., the last equivalence directly.

THEOREM 12.3. *Given an ordered set* S: *in order that for each mapping* f *of* S, *one has*

$$\Big(\bigcup_A \bigcap_{a \in A} f(a)\Big)' = \bigcup_M \bigcap_{m \in M} f'(m) \qquad (A \in \bar{O}S,\ M \in \bar{O}'S), \tag{12.2}$$

it is necessary and sufficient that S *satisfies the* $(\bar{k})$*-condition* (cf. Theorem 3.1).

13. Some special cases of the duality theorem.

THEOREM 13.1. *If* S *is a chain or an antichain, then the duality Theorem* 12.1 *yields the theorem of De Morgan.*

Let us consider an antichain S; thus $\bar{O}S = \{S\}$; $\bar{O}'S$ is the system of all one-point sets $x \in S$. Then for each $M \in O'S$, one has $M = \{x\}$ with $x \in S$; thus $\bigcap_{m \in M} f'(m) = f'(x)$ where $\{x\} = M$ and one has

$$\bigcup_M \bigcap_m f'(m) = \bigcup_M f'(x) = \bigcup_{x \in S} f'(x) \tag{13.1}$$

On the other hand, as $\bar{O}S = \{S\}$,

$$\bigcap_{a \in A} f(a) = \bigcap_{s \in S} f(a)$$

and

$$\bigcup_{A \in \bar{O}S} \bigcap_{a \in A} f(a) = \bigcup_{A \in \{s\}} \bigcap_{s \in S} f(s) = \bigcap_{s \in S} f(s) . \tag{13.2}$$

By virtue of (13.1) and (13.2) the equality (12.2) yields

$$\Big(\bigcap_{s \in S} f(s)\Big)' = \bigcup_{s \in S} f'(s)$$

and this is just the equality of De Morgan. Since each family, or set, may be considered as an antichain, we see that Theorem 12.3 (its sufficient condition) for S an antichain gives the equality of De Morgan in its most general form.

THEOREM 13.2. *For each ramified table T and each mapping f of T one has*

$$(\perp'_M \perp_m f(m))' = \perp'_A \perp_a f'(a) \tag{13.3}$$

$$(m \in M \in OT, \ a \in A \in O'T, \ \perp \in \{\bigcap, \bigcup\})$$

In particular ($\perp = \bigcap$):

$$(\bigcup_M \bigcap_m f(m))' = \bigcup_A \bigcap_a f'(a) \quad (m \in M \in OT, \ a \in A \in O'T), \tag{13.4}$$

and reciprocally.

If one bears in mind the generality and importance of ramified tables (a tool for complete subdivisions or atomizations of sets), one is conscious of the importance and generality of Theorem 13.2.

REMARK 13.1. From a logical point of view it is very important that (13.4) as well as its reciprocal hold, especially for each table whose chains are finite.

Actually, we observe that such tables occur even in psychological processes, in subdivisions, evolution, etc. Thus is seems that the evolution processes follow a ramification scheme, as will be shown elsewhere.

REFERENCES

1. P. Alexandroff, *Sur l'invariances des ensembles complémentaires aux ensembles A*, Mat. Sb. **31** (1923), 310–318.

2. ———, *Sur les ensembles complémentaries aux ensembles A*, Fund. Math. **6** (1924), 160–165.

3. A. Fraenkel, *Abstract set theory*, Amsterdam, (1953).

4. G. Kurepa, *Ensembles ordonnés et ramifiés*, Thèse. Paris, 1935. Publ. Math., Belgrade, **4** (1935), 1–138.

5. ———, Teorija skupova, Zagreb (1951).

6. ———, *Ensembles partiellement ordonnés*, Proc. Int. Congr. Math. 1950 **1** (1952), 460–461.

7. ———, *Généralisation de l'opération de Suslin, de celle d'Alexandroff et de la formule de De Morgan*, Bull. Int. Acad. Sci., Zagreb, **5** (1955), 97–107.

8. ———, *On a characteristic property of finite sets*, Pacific J. Math. **2** (1952), 323–326.

9. W. Sierpinski, *Sur une généralisation des ensembles mesurables B*, Bull. Acad. Sci., Cracovie, (1918), 161–167.

10. M. Suslin, *Sur une définition des ensembles mesurables B sans nombres transfinis*, C.R. Acad. Sci., Paris **164** (1917), 88.

11. A. Tarski, *Zur Grundlegung der Boole'schen Algebra I*, Fund. Math. **24** (1935) 177–198.

LIE ALGEBRAS OF LOCALLY COMPACT GROUPS

RICHARD K. LASHOF

1. Introduction. We call an LP-group, a group which is the projective limit of Lie groups. Yamabe [8] has proved that every connected locally compact group is an LP-group. This permits the extension to locally compact groups of the notion of a Lie algebra. In §§ 2 and 3 we prove the existence and uniqueness of the Lie algebra of an LP-group and show the connection of the Lie algebra with the group by means of the exponential mapping.

In § 4, we extend the notion of a universal covering group for connected groups with the same Lie algebra. A covering group of a connected group g, in the extended sense used here, means a pair $(\bar{g}, w)$, where $\bar{g}$ is a connected LP-group and w is a continuous representation of $\bar{g}$ into g which induces an isomorphism of the Lie algebra of $\bar{g}$ onto the Lie algebra of g (see Definition 4.5). The universal covering group of a connected locally compact group is not necessarily locally compact and may not map onto the group. It turns out that the arc component of the identity in $\bar{g}$ is a covering space in the sense of Novosad [5] of the arc component of the identity of g (these components are dense subgroups, Lemma 3.7).

Finally, in § 5, we establish a one-to-one correspondence between " canonical LP-subgroups " of a group and subalgebras of its Lie algebra.

2. Projective limit of Lie algebras.

DEFINITION 2.1. By a *topological Lie algebra* (over the real numbers) we shall mean a (not necessarily finite dimensional) Lie algebra with an underlying topology such that the operations of addition, multiplication and scalar multiplication are continuous.

DEFINITION 2.2. Let J be an inductive set. Suppose given for each $a \in J$, a topological Lie algebra G_a such that if $a < b$ there exists a continuous representation $f_{ab}: G_b \to G_a$. Let $G = [\{X_a\} \in \prod_{a \in J} G_a$ such that $f_{ab}X_b = X_a$, all $a, b \in J$ with $a < b]$. Then G is a closed topological subalgebra of the direct product.

In analogy to A. Weil [7, p. 23], G will be called the *projective limit* of the G_a ($G = \lim G_a$) if the following hold:

Received September 27, 1955, and in revised form May 18, 1956. This paper is a portion of the author's doctoral dissertation, which was written at Columbia University. The author wishes to thank Professors Kadison and Taylor for their kind assistance.

LP I: If $a<b<c$, $f_{ac}X_c=f_{ab}(f_{bc}X_c)$;
LP II: f_{ab} is a continuous open homomorphism of G_b onto G_a;
LP III: f_a, the natural projection of G into G_a, is continuous and onto.
(Remark: f_a open is implied by f_{ab} open, (see [7, p. 24])

In particular, if the G_a are finite dimensional Lie algebras with the usual topology as a Euclidean space, we get the following.

THEOREM 2.3. *Let G_a, $a \in J$, be a system of finite dimensional Lie algebras satisfying* LP I *and*

LP II′: *f_{ab} is a representation onto.*

Then G as defined in Definition 2.2 *will necessarily satisfy* LP II *and* LP III *and hence $G=\lim G_a$.*

Proof. LP II′ implies LP II, since for finite dimensional vector spaces a representation onto is both continuous and open in the usual topology. That LP III is satisfied follows directly from the theory of linearly compact vector spaces [3 Ch. III, § 27]. In fact, this result holds for an inverse system of finite dimensional vector spaces.

DEFINITION 2.4. If $G=\lim G_a$, G_a *finite* dimensional Lie algebras, then G will be called an *LP algebra.*

LEMMA 2.5. *Let $G=\lim G_a$, where the G_a's are complete topological Lie algebras, with homomorphisms f_a and f_{ab} satisfying* LP I, II, *and* III. *Let N_a be the kernel of f_a, then*

A. *Every neighborhood of zero in G contains an N_a*
B. *For each N_a, N_b; there exists an $N_c \subset N_a \cap N_b$.*
C. *G is complete.*

Proof. It is easy to show (see [7]) that a fundamental system of neighborhoods of zero in G is given by $f_a^{-1}(V_a)$, $a \in J$, and V_a running through a fundamental system of neighborhoods of zero in G_a. Condition A then follows directly.

If $c>a$, b then N_c obviously satisfies B. Condition C is immediate from the definition of G.

LEMMA 2.6. *Given a topological Lie algebra G containing closed ideals N_a satisfying* A, B, C; *then* $\lim G/N_a$ *exists and is isomorphic to G (where we define $a<b$ if $N_a \supset N_b$ and let $f_{ab}:G/N_b \to G/N_a$ be the natural homomorphism).*

Proof. Since the Conditions A, B, C are identical to those for topological groups [7, p. 25] it follows that G is isomorphic to $\lim G/N_a$ as an additive topological group. The lemma now follows since the N_a are ideals.

THEOREM 2.7. *Suppose $G = \lim G_a$, G_a fiinite dimensional. If K is a closed subalgebra of G, then $K = \lim K_a$, where K_a is the image of K in G_a. If K is a closed ideal in G, then $G/K = \lim G_a/K_a$. In particular, G/K is complete.*

Proof. If $f_{ab}:G_b \to G_a$ $(a < b)$, then $f_{ab}:K_b \to K_a$ satisfies LP I, II′. Hence $\lim K_a$ exists. Since K maps onto K_a, K is dense in $\lim K_a$, [7]. Since K is closed, $K = \lim K_a$.

Likewise G_a/K_a, $a \in J$, satisfy LP I, II where $\bar{f}_{ab}:G_b/K_b \to G_a/K_a$ is induced by f_{ab}. Hence $\lim G_a/K_a$ exists. The natural maps $p_a:G_a \to G_a/K_a$ evidently induce a map $p:G \to \lim G_a/K_a$ defined by: $p\{X_a\} = \{p_a X_a\}$, $\{X_a\} \in G$. This definition is legitimate since:

$$\bar{f}_{ab}(p_b X_b) = p_a(f_{ab} X_b) = p_a(X_a)\,, \qquad a < b\,.$$

This in turn induces a map $i:G/K \to \lim G_a/K_a$. We have to show that i is an isomophism.

By its definition i is evidently continuous and one-to-one into. We show that it is an isomorphism into. Since the natural map of G onto G/K is open, it is sufficient to show that if W is a neighborhood in G then $p(W) = p(W+K)$ is open. Now if W is a neighborhood in G, take $V + V \subset W$. Then V contains an N_a, kernel of $f_a:G \to G_a$ (Lemma 2.5). Then $p(W)$ contains $p(V) + p(N_a)$. Now $(p(V))_a = p_a f_a(V)$, is an open neighborhood in G_a/K_a; and the preimage in $\lim G_a/K_a$ of a neighborhood in G_a/K_a is a neighborhood in lim G_a/K_a. But if $(p(X))_a \in p_a f_a(V)$, $X \in G$ then $X_a \in f_a(V) + K_a$ and hence $X \in V + N_a + K \subset W + K$. But this implies that i is open in $i(G/K)$ and hence is an isomorphism into.

It remains to show that i is onto. But this follows since as an abstract vector space G is linearly compact [3]. Hence

$$\bigcap_{a \in J} f_a^{-1} p_a^{-1}(Y_a) \neq \phi, \quad \{Y_a\} \in \lim G_a/K_a$$

since the intersection is nonempty for any finite subset of J.

LEMMA 2.8. *Let θ be a continuous representation of an* LP-*algebra G into an* LP-*algebra H. Then the image of G in H is a closed subalgebra of H.*

Proof. Suppose $G = \lim G_a$, $a \in J$; $H = \lim H_b$, $b \in K$; G_a and H_b finite dimensional. Consider the map $p_b: G \to H_b$, composed of θ and the projection of H onto H_b. This takes G onto a subalgebra H'_b of H_b. Obviously, the H'_b, $b \in K$, define a closed subalgebra $H' = \lim H'_b$ of H under the induced system of representations. Also $\theta(G) \subset H'$.

Let $Y = \{Y_b\}$ be in H'. The preimage $p_b^{-1}(Y_b)$ is a closed linear variety in G. Since G is linearly compact, $\bigcap p_b^{-1}(Y_b)$ is nonempty, as the intersection of a finite number is clearly nonempty. Hence there exists $X \in G$ such that $p_b(X) = Y_b$, all $b \in K$. Hence G maps onto H'. As H' is closed, this proves the Lemma.

LEMMA 2.9. *Let θ be a continuous representation of an* LP-*algebra G onto an* LP-*algebra H. Then θ is open.*

Proof. Let K be the kernel of θ. By Theorem 2.7, G/K is an LP-algebra. Hence it is clearly sufficient to prove the Lemma in the case that the map is also one-to-one.

Hence let θ be a continuous one-to-one representation of G onto H. Suppose $G = \lim G/K_a$; K_a, $a \in J$, closed ideals in G and G/K_a finite dimensional. By Lemma 2.8, $\theta(K_a)$ is a closed ideal in H. Further $H/\theta(K_a)$ is finite dimensional and hence is (topologically) isomorphic to G/K_a. Since θ is continuous every neighborhood of H contains some $\theta(K_a)$. It follows from Lemma 2.6 that

$$H = \lim H/\theta(K_a) = \lim G/K_a = G .$$

Clearly, the isomorphism so induced is the same map as θ.

THEOREM 2.10. *If G_1 and G_2 are two* LP-*algebras with the same underlying abstract algebra G, then G_1 and G_2 have the same topology.*

Proof. By Lemma 2.9 it is sufficient to construct an LP-algebra G_0 whose underlying abstract algebra is G and such that the identity maps of G_0 into G_1 and G_2 are both continuous. Let K_a, $a \in J$, be the set of all abstract ideals in G which have finite codimension. As is well known, if K_a and K_b have finite codimension, so does $K_a \cap K_b$. It follows that G/K_a, $a \in J$, satisfy the conditions of Theorem 2.3, where we define $b > a$ if $K_b \subset K_a$ and $f_{ab}: G/K_b \to G/K_a$ is the natural projection. Let $G_0 = \lim G/K_a$. We claim the underlying abstract algebra of G_0 is G.

Now G is linearly compact since it is the underlying algebra of G_1 and G_2. Let $p_a: G \to G/K_a$ be the natural map. If $Y = \{Y_a\} \in G_0$, then $\bigcap p_a^{-1}(Y_a)$ is nonempty, since the intersection of a finite number is nonempty. Hence there exists $X \in G$ such that $p_a(X) = Y_a$, all $a \in J$. Hence the map $p: X \to p_a(X)$ take G onto G_0. p is one-to-one, since for every $X \in G$ there is a K_a such that $p_a(X)$ is not zero, because the identity map of say G onto G_1 is one-to-one.

Now $G_1 = \lim G_1/K_b$, $b \in J'$, J' a subset of J. $p^{-1}: G_0 \to G_1$ is an isomorphism of this underlying abstract algebras and is continuous since J' is a subset of J. Hence the theorem follows by Lemma 2.9.

COROLLARY 2.11. *Let G be an abelian* LP*-algebra. Then G is the direct product of* 1*-dimensional algebras. In particular, the underlying topological vector space of an* LP*-algebra is algebraically and topologically the direct product of* 1*-dimensional vector spaces.*

3. Lie algebra of an LP-group.

DEFINITION 3.1. Let g_a, $a \in J$ be Lie groups. Suppose $g = \lim g_a$, the limit satisfying LP I, II, III of A. Weil [7]. Then we call g an *LP-group*. (Note that g is complete since the g_a are complete.)

DEFINITION 3.2. Suppose $g = \lim g_a$, g_a connected Lie groups. Let G_a be the Lie algebra of g_a. Then the homomorphisms $f_{ab}: g_b \to g_a$ $(a < b)$ induce homomorphisms $df_{ab}: G_b \to G_a$ satisfying LP I, II of Definition 2.1. Hence the G_a, $a \in J$, have a limit G. G is called the *Lie algebra* of g.

We show in Lemma 3.4 below that G is independent (in a natural sense) of the representation of g as a limit of Lie groups.

DEFINITION 3.3. Suppose $g = \lim g_a$, g_a connected Lie groups. Let G, $G = \lim G_a$, be the Lie algebra of g. Then we define a continuous map

$$\exp: G \to g, \quad \exp\{X_a\} = \{\exp X_a\}, \quad \{X_a\} \in G .$$

This mapping is legitimate, since if $f_{ab}: g_b \to g_a$ then

$$f_{ab}(\exp X_b) = \exp df_{ab} X_b = \exp X_a$$

and hence $\{\exp X_a\} \in g$.

LEMMA 3.4. *Suppose $g = \lim g_a$, g_a connected Lie groups, $a \in J$; $h = \lim h_b$, h_b connected Lie groups, $b \in K$. Let $G = \lim G_a$, $H = \lim H_b$ be the corresponding Lie algebras. If $\theta: g \to h$ is an isomorphism then we can define an isomorphism $d\theta: G \to H$ such that*

$$\theta(\exp X) = \exp d\theta(X), \quad X \in G$$

Proof. Let $f_a: g \to g_a$ and $\bar{f}_b: h \to h_b$ be the natural maps. Let n_a and $\bar{n}_b$ be the kernels of f_a and $\bar{f}_b$ respectively. Let $b \in K$. Since h_b is a Lie group there is a neighborhood V_b of h_b which contains no nontrivial subgroups (that is, h_b doesn't have arbitrarily small subgroups). Since $\bar{f}_b\theta: g \to h_b$ is continuous, there is a neighborhood W in g which maps into V_b. But W contains some n_a and this n_a must go into the unit element of h_b. This defines a homomorphism $\theta_{ba}: g_a \to h_b$ such that

$$\theta_{ba} f_a = \bar{f}_b \theta ,$$

this condition characterizing θ_{ba}.

If $a > a'$, a and a' in J, then $f_a = f_{aa'} f_{a'}$ and $\theta_{ba} f_{aa'} f_{a'} = \bar{f}_b \theta$. Hence $\theta_{ba} f_{aa'} = \theta_{ba'}$. Similarly, if $b' < b$, $\bar{f}_{b'b} \theta_{ba} = \theta_{b'a}$. The induced homomorphisms of the corresponding Lie algebras therefore satisfy

$$d\theta_{ba} df_{aa'} = d\theta_{ba'}, \quad d\bar{f}_{b'b} d\theta_{ba} = d\theta_{b'a} .$$

It follows that the maps $d\theta_{ba}$ define a continuous representation $d\theta$ of G into H, where $d\theta(X)$, $X = \{X_a\} \in G$, is defined by

$$(d\theta(X))_b = d\theta_{ba}(X_a) .$$

This map is well defined because of the conditions satisfied above, and is continuous because $d\theta_{ba}$ is continuous.

Similarly for each $a \in J$, we can find a $b \in K$ and a homomorphism $\psi_{ab}: h_b \to g_a$ such that $f_a \theta^{-1} = \psi_{ab} \bar{f}_b$. This defines a continuous representation $d\psi$ of H into G. Because of the conditions satisfied by the maps one sees easily that: $d\theta d\psi: H \to H$ and $d\psi d\theta: G \to G$ are the identities, and hence that $d\theta$ is an isomorphism.

Since $d\theta_{ba} df_a = d\bar{f}_b d\theta$ by definition, we have

$$\bar{f}_b \theta(\exp X) = \theta_{ba} f_a(\exp X) = \exp d\theta_{ba}\, df_a(X) = \exp d\bar{f}_a d\theta(X) .$$

By Definition 3.3, this implies that $\theta(\exp X) = \exp d\theta(X)$.

THEOREM 3.5. *Suppose $g = \lim g_a$, g_a connected Lie groups. Let $G = \lim G_a$ be the corresponding Lie algebra. Then g is the closed subgroup generated by the elements of the form $\exp X$, $X \in G$.*

Proof. Since G maps onto G_a, $\exp X_a$ for $X \in G$ generates g_a, and $\exp G$ generates a dense subgroup of g, proving the theorem.

LEMMA 3.6. *Suppose $G = \lim G_a$, G_a finite dimensional Lie algebras. Then the underlying space of G is arcwise connected.*

Proof. Since G is a topological vector space it is arcwise connected by straight lines.

LEMMA 3.7. *Suppose $g = \lim g_a$, g_a connected Lie groups. Then g is* *connected and the arcwise connected component of g is dense in g.*

Proof. The map $\exp: G \to g$ is continuous. Hence if A is the image of G, A is arcwise connected. Hence A^n is arcwise connected. Hence $\bigcup_1^\infty A^n$ is arcwise connected. By Theorem 3.5 this is dense in g. Hence g is connected.

THEOREM 3.8. *Let (g_a), $a \in J$ be a system of connected Lie groups satisfying* LP I, II *of A. Weil* [1]; *then* $g = (\{x_a\} \in \prod_{a \in J} G_a;\ f_{ab}x_b = x_a$, *all* $a, b \in J$ *with* $a < b)$ *satisfies* LP III, *and hence*:

$$g = \lim g_a$$

Proof. Let G_a be the Lie algebra of g_a. Then the (G_a), $a \in J$ satisfy LP I, II of Definition 1.1. Hence they have a limit G. Let $X \in G$, then if $X = \{X_a\}$ we have $\exp X_a \in g_a$ and $\{\exp X_a\} \in g$, since:

$$f_{ab} \exp X_b = \exp df_{ab} X_b = \exp X_a, \ a < b$$

But elements of the form $\exp X_a$ generate g_a since G maps onto G_a. Hence g maps onto g_a. Hence $g = \lim g_a$.

LEMMA 3.9. *Let $g = \lim g_a$, g_a arbitrary Lie groups. Let g_a^0 be the connected component of the identity of g_a. Then the (g_a^0), $a \in J$ form a system of groups satisfying* LP I, II *of A. Weil. Let $g^0 = \lim g_a^0$, then g^0 is the connected component of g.*

Proof. Since $f_{ab}: g_b \to g_a$ is continuous, open and g_b^0 is open in g_b, it takes g_b^0 onto g_a^0. Hence the (g_a^0), $a \in J$ satisfy LP I, II. By Theorem 3.8 they have a limit g^0. g^0 may obviously be considered as a subgroup of g, closed since complete.

By Lemma 3.7, g^0 is connected and hence contained in the connected component of g. On the other hand, if g_1 is the connected component of g, $f_a(g_1)$ is connected and hence contained in g_a^0. Hence g_1 is contained in the limit of the g_a^0. Hence $g_1 = g^0$.

DEFINITION 3.10. Let g be a topological group. If the connected component of the identity of g is an LP-group, we define the Lie algebra of g as the Lie algebra of its connected component.

REMARK. According to the result of Yamabe [8], every locally compact group is a generalized Lie group. This implies in particular that its connected component is an LP-group. Hence every locally compact group has a Lie algebra.

THEOREM 3.11. *Let g and h be topological groups for which Lie Algebras G and H are defined (Definition 3.10). Let f be a continuous representation of g into h, then f induces a unique continuous representation of G into H such that $f(\exp X)=\exp df(X)$, $X\in G$.*

Proof. Obviously f defines a continuous representation of the connected component of g into the connected component of h. Assume therefore that g and h are connected.

Suppose $g=\lim g_a$, $h=\lim h_b$ (g_a, h_b connected Lie groups). The map $g\rightarrow h\rightarrow h_b$ induces a map of $g_a\rightarrow h_b$ for some a, since h_b doesn't have arbitrarily small subgroups. This in turn induces a map of $G_a\rightarrow H_b$ and hence of G into H_b. As in the proof of Lemma 3.4, it is easy to see that the maps $G\rightarrow H_b$ induce a continuous representation of G into H, $df: G\rightarrow H$; such that $f(\exp X)=\exp df(X)$.

Suppose there are two such representations, say df and $\overline{df}$. Then $(dfX)_b\neq(\overline{df}X)_b$ some b and X. Since any neighborhood of zero generates G, and since df, $\overline{df}$ are linear; we can chose X such that dfX and $\overline{df}X$ are in any desired neighborhood of 0 in H_b. For a sufficiently small neighborhood exp is one-to-one on H_b. Hence $(\exp dfX)_b\neq(\exp\overline{df}X)_b$, a contradiction.

COROLLARY. *If g is connected and f, f_1 are two representations of g into h such that $df=df_1$, then $f=f_1$.*

Proof. Since $f(\exp X)=\exp dfX=\exp df_1X=f_1(\exp X)$ and since $\exp G$ generates a dense subgroup of g, we have $f=f_1$.

LEMMA 3.12. *Let g and h be locally compact topological group and g connected, G and H their Lie algebras. Let f be a continuous open homomorphism of g onto h, then df is a continuous open homomorphism of G onto H.*

Proof. According to A. Weil [7], if k is the kernel of f, we may take $g=\lim g_a$, $k=\lim k_a$, k_a the image of k in g_a, and $h=\lim g_a/k_a$, g_a and k_a Lie groups.

Then $G=\lim G_a$, G_a the Lie algebra of g_a. Let K be the Lie algebra of k; then $K=\lim K_a$, K_a the Lie algera of k_a. Then the Lie algebra of g_a/k_a is G_a/K_a [1]. Hence $H=\lim G_a/K_a$. By Theorem 2.5, $H=G/K$. It is easy to check that for $df: G\rightarrow G/K$ that $f(\exp X)=\exp dfX$. So that the natural map of G onto G/K is df. (Note: By a generalization of Pontrjagin's theorem on groups satisfying the second axiom of countability; if f is continuous and onto it is automatically open.)

4. Universal covering group.

DEFINITION 4.1. Suppose $g=\lim g_a$, (g_a), $a \in J$ connected Lie groups. Let $\bar{g}_a$ be the simply connected covering group of g_a. The map f_{ab} taking g_b onto g_a $(a < b)$ induces an open homomorphism $\bar{f}_{ab}$ of $\bar{g}_b$ onto $\bar{g}_a$. Hence the $(\bar{g}_a)$, $a \in J$ satisfy LP I, II and therefore have a limit $\bar{g}$ (Theorem 3.8). $\bar{g}$ is a complete, connected group. $\bar{g}$ is called the *universal covering group* of g.

PROPOSITION 4.2. *Let G be the Lie algebra of g. Then $\bar{g}$ has the Lie algebra G, and there exists a continuous representation w taking $\bar{g}$ onto a dense subgroup of g such that $dw: G \to G$ is the identity.*

Proof. The covering homomorphisms $w_a: \bar{g} \to g_a$ induce a continuous representation of $\bar{g}$ into g. Since $dw_a: G_a \to G_a$ is the identity, it follows that dw is the identity. Since w_a maps $\bar{g}_a$ onto g_a it follows that $w(\bar{g})$ is dense in g.

PROPOSITION 4.3. *The kernel of w is totally disconnected and is in the center.*

Proof. If k is the kernel of w, it follows from the definition of w that the image $f_a(k)$ of k in g_a belongs to the kernel of w_a. But this kernel is discrete and hence $f_a(k)$ is discrete, and therefore closed in g_a. It follows that the $f_a(k)$ satisfy LP I, II and since k is closed, $k=\lim f_a(k)$. Hence k is the projective limit of discret groups. It follows from Lemma 3.9 that k is totally disconnected. Further, a totally disconnected normal subgroup of a connected group belongs to the center.

LEMMA 4.4. *Let g be a connected* LP-*group with Lie algebra G. Let h be any other connected* LP-*group with Lie algebra H isomorphic to G. Then there exists an isomorphism of their universal covering groups $f: \bar{g} \to \bar{h}$ such that df is the given isomorphim of G onto H.*

Proof. Suppose $g=\lim g_a$, $h=\lim h_b$; then $\bar{h}=\lim \bar{h}_b$, $\bar{g}=\lim \bar{g}_a$ and $G=\lim G_a$, $H=\lim H_b$. The homomorphism $G \to H \to H_b$ induces a homomorphism $G_a \to H_b$ for some a, since H_b has no small subgroups when considered as an additive group. This in turn induces a homomorphism of $\bar{g}_a$ onto $\bar{h}_b$. Similary there exist homomorphisms of $\bar{h}_b$ onto $\bar{g}_a$, $a \in J$, some b. As in the proof of Lemma 3.4, it is easy to see that this implies that there exists an isomorphism $f:\bar{g} \to \bar{h}$. The induced homomorphism $df:G \to H$ such that $f(\exp X) = \exp df(X)$, is obviously the original isomorphism of G onto H.

DEFINITION 4.5. Let g and h be connected LP-groups, G and H their Lie algebras. A continuous representation w of g into h such that dw is an isomorphism of G onto H is called a *covering map* and g is called a *covering group* of h, (g, w) is called the *covering*.

PROPOSITION 3.6. *$w(g)$ is dense in h.*

Proof. In fact $w(\exp X) = \exp dw(X)$, $X \in G$. But $\exp dw(X)$, $X \in G$ generates a dense subgroup of h since dw is onto.

We now give a purely topological definition of covering space for arcwise connected spaces due to Novosad [5] and show that the arc component of the identity g^c of g in Definition 4.5 is actually a covering space in this sense of the arc component of the identity h^c of h. (Note that g^c is dense in g, Lemma 3.7) Similarly, we show the arc component of the identity of the universal covering group is a universal covering space.

DEFINITION 4.7. (Novosad) Let A be an arcwise connected space, $a \in A$. Let $f:(B, b) \to (A, a)$ be a continuous map of an arcwise connected space B into A taking b into a. Then (f, B, b) is called a *covering space* of (A, a), if given any contractible space C, and point $c \in C$ which is a deformation retract of C, and a map $\alpha:(C, c) \to (A, c)$, then there exists a map $\bar{\alpha}:(C, c) \to (B, b)$ which is unique with respect to the property $f\bar{\alpha} = \alpha$.

Let $(P_A, a\sharp)$ be the pair consisting of the space of paths starting from $a \in A$, with the compact open topology, of an arbitrary topological space A, and the constant path $a\sharp$ at $a \in A$. A continuous map $f:(B, b) \to (A, a)$ induces a continuous map $f\sharp:(P_B, b\sharp) \to (P_A, a\sharp)$ defined by:

$$f\sharp(p)(t) = f(p(t)), \quad p \in P_B, \quad t \in I \text{ (the unit interval)}$$

It is then easy to see that Definition 4.7 is equivalent to the following.

DEFINITION 4.7′. Let $f:(B, b) \to (A, a)$ be a continuous map of an arcwise connected space B into an arcwise connected space A taking $b \in B$ into $a \in A$. Then (f, B, b) is called a covering space of (A, a) if $f\sharp:(P_B, b\sharp) \to (P_A, a\sharp)$ is a homeomorphism onto.

PROPOSITION 4.8. *Definitions* 4.7 *and* 4.7′ *are equivalent.*

Proof.

(4.7) implies (4.7′) by Lemma 2.3 of [5].

(4.7′) implies (4.7); since, let $\lambda_A : P_A \to A$, $\lambda_A(p) = p(1)$, $p \in P_A$. Then λ_A defines a fiber space (in the sense of Serre) which obviously satisfies the covering homotopy theorem for arbitrary spaces. Hence if $\alpha:(C, c) \to (A, a)$ is homotopic to tne constant map the homotopy may be lifted to $(P_A, a\sharp)$ and hence to $(P_B, b\sharp)$ by the homeomorphism of (4.7′). λ_B maps the image into (B, b). The endpoint of the homotopy gives the desired covering of (4.7). The uniquenes follows since any point of C describes a path under the retraction and the image of this path in B is unique since covering paths are unique by (4.7′).

DEFINITION 4.9. An arcwise connected space A is called *simply connected* if every covering space (4.7) of (A, a) is trivial. This property is independent of the base point $a \in A$ (see [5]).

Let Ω_A be the (closed) subspace of P_A consisting of closed paths (that is, the loop space). Let Ω_A^c be the arc component of $a\sharp$ in Ω_A.

THEOREM 4.10. *Let A be an arcwise and locally arcwise connected space, $a \in A$; then if Ω_A is connected (not necessarily arcwise connected) and Ω_A^c is dense in Ω_A, A is simply connected (Definition* 4.9).

Proof. Let (f, B, b) be a covering space of (A, a). Then $f\sharp : P_B \to P_A$ is a homeomorphism, and hence $f\sharp$ maps Ω_B homeomorphically into Ω_A. But every loop in A contractible to a may be lifted to a unique loop in B (see proof of 4.8), hence $f\sharp(\Omega_B) \supset \Omega_A^c$. But $f\sharp(\Omega_B)$ is closed in P_A. Hence $f\sharp$ maps Ω_B onto Ω_A.

Now this means that $f: B \to A$ is one-to-one. For if $f\sharp(p)(1) = f\sharp(p') \cdot (1)$, $p, p' \in P_B$, then $p(1) = p'(1)$; since $f\sharp p$ and $f\sharp p'$ having the some endpoint form a loop in A, and hence must come from a loop in B by the above. Also $p(1) = p'(1)$ implies $(f\sharp p)(1) = (f\sharp p')(1)$. Hence since $\lambda_B : P_B \to B$, and $\lambda_A : P_A \to A$ (the endpoint maps) are onto, f is one-to-one.

Further λ_A is both continuous and open [4, Lemma 4] since A is locally arcwise connected, hence the continuous map $\lambda_B f^{-1} : P_A \to B$ induces a continuous map $f^{-1} : A \to B$. This proves the theorem.

We now apply the above to LP-groups. We remark that if g is a topological group; then P_g, the space of paths of g beginning at the identity, may be made into a topological group by pointwise multiplication of paths. Then Ω_g is a closed normal subgroup.

LEMMA 4.11. *Let $g = \lim g_\alpha$, g_α Lie groups; then $P_g = \lim P_{g_\alpha}$.*

Proof. First $f_a : g \to g_a$ has a local cross-section. In fact, it is obvious that $df_a : G \to G_a$ has a cross-section since these spaces are linear, further g_a has a neighborhood which is homeomorphic to a neighborhood of G_a and $\exp : G \to g$ is continuous.

Now P_{g_a} is arcwise connected, hence $f_a^\sharp : P_g \to P_{g_a}$ will be onto if the image covers a neighborhood of the identity. But this follows from the local cross-section of g_a in g. In fact, a fundamental system of neighborhoods of the identity in P_a is obtained from a fundamental system of neighborhoods of the identity in g by taking all those paths that are contained in a given neighborhood of the identity in g. Also it follows that $f_a^\sharp$ is open.

If V is a neighborhood of the identity in g that contains the kernel k_a of f_a, then the corresponding neighborhood in P_g contains P_{k_a}, and this last is clearly the kernel of $f_a^\sharp$. Finally, P_g is complete since g is complete. Hence all the conditions for a projective limit are satisfied and the lemma follows.

LEMMA 4.12. *If* $g = \lim g_a$, g_a *simply connected Lie groups, then* $\Omega_g = \lim \Omega_{g_a}$.

Pooof. The proof is the same as above, using the fact that since g_a is simply connected, Ω_{g_a} is arcwise connected.

LEMMA 4.13. *If* $g = \lim g_a$, g_a *simply connected, then* Ω_g^c *is dense in* Ω_g *and* Ω_g *is connected.*

Proof. Since $f_a : g \to g_a$ has local cross-sections (see 4.11), it defines a principal fiber bundle and hence satisfies the covering homotopy theorem of [6]. Hence since every loop in g_a is contractible each loop may be lifted to a contractible loop in g. Hence Ω_g^c maps onto Ω_{g_a}, all a. Hence Ω_g^c is dense in Ω_g. The last statement then follows.

LEMMA 4.14. *If* $g = \lim g_a$, g_a *simply connected, then* $P_g/\Omega \approx g^c$.

Proof. For Lie groups, $P_{g_a}/\Omega_{g_a} \approx g_a$, since g_a is locally arcwise connected and hence the map of P_{g_a} onto g_a is open (see [7]). Since P_g maps continuously onto P_{g_a}, P_g maps continuously onto $P_{g_a}/\Omega_{g_a} \approx g_a$. The induced map of P_g into $g = \lim g_a$ is obviously λ_g, and this induces a continuous one-to-one map of P_g/Ω_g onto $g^c \subset g$. Hence it is sufficient to show that P_g/Ω_g has the proper topology as a subgroup of g.

The neighborhoods of the identity in P_g/Ω_g are of the form $V\Omega_g$, where V is the preimage in P_g of a neighborhood V_a in P_{ga}. But this is the preimage in P_g/Ω_g of the neighborhood $V_a\Omega_{g_a}$ in $P_{g_a}/\Omega_{g_a} \approx g_a$. Hence the Lemma follows.

THEOREM 4.15. *If* $g = \lim g_\alpha$, g_α *simply connected Lie groups, then* g^c *(the arc-component of the identity) is arcwise connected, locally arcwise connected, and simply connected (Definition* 4.9).

Proof. g^c is locally arcwise connected since P_g is obviously so. Hence by Lemmas 4.12 and 4.13 and Theorem 4.10, g^c is simply connected.

COROLLARY 4.16. *If g is the universal covering group* (4.1) *of a metrisable* LP-*group or a connected locally compact group, then g is arcwise connected, locally arcwise connected and simply connected.*

Proof. For metrisable groups, P_g is metrisable and complete. Hence $P_g/\Omega_g \approx g^c$ is complete and thus $g^c = g$. The result for locally compact groups will follow from Theorem 4.25.

We write again w for the map of Definition 4.5 cut down to g^c.

LEMMA 4.17. (w, g^c, e) *is a covering space (Definition* 4.7) *of* (h^c, e).

Proof. First assume $g = \bar{h}$ the universal covering group of h. Then if $h = \lim h_\alpha$, $\bar{h} = \lim \bar{h}_\alpha$ and since $P_{\bar{h}_\alpha} \approx P_{h_\alpha}$ (this is obvious for Lie groups), $P_{\bar{h}} \approx \lim P_{\bar{h}_\alpha} \approx \lim P_{h_\alpha} \approx P_h$. Further, this isomorphism is clearly induced by the covering map w.

Now let (g, w) be any covering group of h. Then $\bar{h} \approx \bar{g}$ by (4.4) and hence $P_g \simeq P_{\bar{g}} \approx P_{\bar{h}} \approx P_h$. Since $\bar{g} \to g \to h$ is the same as $\bar{g} \to \bar{h} \to h$. The isomorphism $P_g \approx P_h$ is induced by w. Hence (w, g^c, e) satisfies (4.7)$'$.

THEOREM 4.18. *If g and h are* LP-*groups, G and H their Lie algebras, $\bar{g}$ and $\bar{h}$ their universal covering groups, P_g and P_h their group of paths, Ω_g^c and Ω_h^c the arc component of the identity in their group of loops, respectively; then the following are equivalent:*

(a) *G isomorphic to H* (b) *$\bar{g}$ isomorphic to $\bar{h}$*
(c) *P_g isomorphic to P_h such that the isomorphism takes Ω_g^c onto Ω_h^c.*

Proof.

(a) implies (b) follows from (4.4).
(b) implies (a) follows from (4.2).

(b) implies (c): $P_g \simeq P_{\bar{g}}$ by (4.17). Also $\Omega_{\bar{g}}^c \simeq \Omega_g^c$ under the same map since every contractible loop in g may be lifted to $\bar{g}$ (see proof of 4.8).

(c) implies (b): writing $\bar{\Omega}_g^c$ for the closure of Ω_g^c in P_g, we have $\bar{g}^c \simeq P_{\bar{g}}/\Omega_{\bar{g}} \simeq P_g/\bar{\Omega}_g^c \simeq P_h/\bar{\Omega}_h^c \simeq P_{\bar{h}}/\Omega_{\bar{h}} \simeq \bar{h}^c$, since Ω_g^c is dense in $\Omega_{\bar{g}}$. Hence $\bar{g} \simeq \bar{h}$.

LEMMA 4.19. *Every* LP-*algebra is the Lie algebra of an* LP-*group.*

Proof. By assumption, if G is an LP-algebra then $G = \lim G_a$, G_a finite dimensional. Let g_a be the simply connected groups corresponding to G_a. The homomorphisms of G_a onto G_b $(a < b)$, induce homomorphisms of g_a onto g_b which satisfy LP I, II. Hence they have a limit g (Theorem 3.8). But g obviously has Lie algebra G.

DEFINITION 4.20. The group g defined in the proof of (4.19) is called the *universal group corresponding to G.*

LEMMA 4.21. *Let g be a universal* LP-*group then every covering group (h, w) (Definition* 4.5) *is trivial, that is, w is an isomorphism of h onto g.*

Proof. h^c is a trivial covering of g^c. Since h and g are complete and h^c and g^c are dense, the lemma follows.

THEOREM 4.22. *Let h be an* LP-*group and let H be its Lie algebra. Let G be an* LP-*algebra and g the universal group corresponding to G. Let θ be a continuous representation of G into H. Then θ induces a continuous representation f of g into h such that $df = \theta$.*

Proof. If $h = \lim h_b$, h_b Lie groups; then $H = \lim H_b$, H_b finite dimensional Lie algebras. Suppose $G = \lim G_a$, G_a finite dimensional. Let g_a be the simply connected group corresponding to G_a, $g = \lim g_a$. The map $G \to H \to H_b$ induces a map $G_a \to H_b$, some a. But this in turn induces a map $g_a \to h_b$ and hence a map of $g \to h_b$ for every b. It is easy to see that this defines a map $f : g \to h$ such that $df = \theta$.

THEOREM 4.23. *The universal covering group $\bar{g}$ of a connected locally compact group g is the direct product of simply connected Lie groups. More explicitly $\bar{g} \simeq h \times a \times s$, where h is a simply connected Lie group, a is the (possibly infinite) direct product of the reals and s is the (possibly infinite) direct product of simple simply connected compact Lie groups.*

Proof. According to Yamabe [8] and Iwasawa (Theorem 11 of [2]), g is locally the direct product of a local Lie group h' and a compact normal subgroup k. Now $k=\lim k_a$, where $k_a=k/n_a$, n_a normal in k and hence in g (Theorem 4 of [2]). Hence $g=\lim g_a$, where $g_a=g/n_a$. Evidently g_a is locally isomorphic to $h'\times k_a$; and hence to $h'\times k_a^0$, k_a^0 connected component of k_a (since k_a a Lie group).

Since $f_{ab}:g_b\to g_a$ takes k_b onto k_a, it takes k_b^0 onto k_a^0. Hence f_{ab} induces a homomorphism of $h\times\bar{k}_b^0$ onto $h\times\bar{k}_a^0$, where h is the simply connected group associated with h' and $\bar{k}_a^0$ is the simply connected covering group of k_a^0. If $\bar{k}^0=\lim\bar{k}_a^0$, then $k\times\bar{k}^0=\lim h\times\bar{k}_a^0$; $h\times\bar{k}_a^0$ is the simply connected covering group of g_a and hence $h\times\bar{k}^0$ is the universal covering group of g.

Since $\bar{k}^0$ is the universal covering group of k^0, k^0 connected component of k; the problem is reduced to considering the universal covering group of a compact connected group.

According to A. Weil (p. 91 of [7]), k^0 is isomorphic to $(a'\times s)/d$, where a' is a compact abelian connected group, s is the (possibly infinite) direct product of simple simply connected compact Lie groups, and d is totally disconnected. It is evident that $\bar{k}^0=a\times s$, where a is the universal covering group of a'. Since a' in the projective limit of toroidal groups, a is the projective limit of vector groups, and hence is the direct product of the reals (2.11). This proves the theorem.

COROLLARY 4.24. *If g is a locally compact group, then its Lie algebra G has the form $G=H\times A\times S$, where H is a finite dimensional Lie algebra, A is the product (possibly infinite) of 1-dimensional Lie algebras, and S is the (possibly infinite) direct product of simple compact Lie algebras.*

EXAMPLE 1. Let $g=\Pi T_a$, $a\in J$, T_a isomorphic to the torus group, all a. Then g is compact. But $\bar{g}=\Pi R_a$, R_a isomorphic to the additive group of reals, $a\in J$. Hence for J infinite, $\bar{g}$ is not locally compact.

EXAMPLE 2. Let P be the p-adic solenoid (See for example: Eilenberg and Steenrod, *Foundations of algebraic topology*, p. 230). P is a compact connected group and is the projective limit of torus groups. If T is the multiplicative group of all complex numbers z with $|z|=1$, the projections $\varphi:T\to T$ are given by $\varphi(z)=z^p$, p an integer. φ induces the map $\bar{\varphi}:R\to R$, $\bar{\varphi}(x)=px$, which is an isomorphism of the additive group of the reals onto itself. Hence the universal covering group of P, which is the projective limit of the reals under these isomorphisms, is itself the additive group of reals. Hence the Lie algebra of P is 1-

dimensional. As is well known (see above reference) R maps continuously, one-to-one onto a dense subgroup of P, not the whole group. The map is not open and not onto P. More generally we have the following.

EXAMPLE 3. Let g be a connected, but not locally connected, locally compact group. Let $(\bar{g}, w)$ be its universal covering group. Then $w:\bar{g} \to g$ is not both open onto. Consequently, if $\bar{g}$ is locally compact, w is neither open (on the image) nor onto.

In fact $\bar{g}$ is locally connected. Hence if w is open and onto, g would be locally connected. If $\bar{g}$ is locally compact then if w is open on $w(\bar{g})$, $w(\bar{g})$ would be locally compact, hence closed, hence $w(\bar{g})=g$. On the other hand, if w is onto and $\bar{g}$ connected locally compact, w is open.

Hence, in particular we have the following. Let g be a connected, but not locally connected finite dimensional locally compact group. Then $\bar{g}$ is locally compact (a Lie group) and hence w is neither open nor onto.

EXAMPLE 4. Not every complete topological Lie algebra is an LP-algebra. In fact an infinite dimensional Banach space cannot contain arbitrarily small subspaces and cannot be an abelian LP-algebra.

5. Subgroups and subalgebras.

DEFINITION 5.1. Let g be an LP-group. An LP-group h is called an LP-*subgroup* of g if h is an abstract subgroup and the inclusion map $f:h \to g$ is a continuous representation such that $df:H \to G$ is an isomorphism into.

THEOREM 5.2. *Let* $g=\lim g_a$ *be a* LP-*group;* $G=\lim G_a$ *its Lie algebra. Let* H *be a closed subalgebra of* G, *then* $H=\lim H_a$ *where* H_a *is the image of* H *in* G_a. *Let* h_a *be the analytic subgroup of* g_a *corresponding to* H_a. *Then* $h=\lim h_a$ *exists and is a connected* LP-*subgroup of* g *with Lie algebra* H.

Proof. $H=\lim H_a$ follows from Theorem 2.5. Now $f_{ab}:g_b \to g_a$ induces $\bar{f}_{ab}:h_b \to h_a$, onto since the image of h_b in g_a is the analytic subgroup of g_a whose Lie algebra is H_a [1]. Therefore the h_a satisfy LP I, II and hence have a limit h (Theorem 3.8).

Obviously h is an abstract subgroup of g and the maps $h_a \to g_a$ induce a continuous one-to-one representation of h into g, namely the inclusion map. Obviously df is the inclusion map of H into G and hence an isomorphism into.

LEMMA 5.3. *Let h be the* LP-*subgroup of g defined in Theorem* 5.2. *Let h' be any other connected* LP-*subgroup with Lie algebra H' and such that the inclusion map $f': h' \to g$ induces an isomorphism df' of H' onto H. Then h' is a covering group of h and the covering map is abstractly an inclusion.*

Proof. Suppose $h' = \lim h'_b$, h'_b Lie groups. The map $h' \to g \to g_a$ induces a map $h'_b \to g_a$, for some b. But the image of h'_b in g_a is the analytic subgroup whose Lie algebra is H_a, the image of H'_b in G_a. Hence, $h'_b \to h_a$ is continuous and open, and induces $h' \to h_a$ continuous, open. This induces a continuous representation $\theta: h' \to h$ such that $d\theta$ is the isomorphism dg'. Since h' is contained abstractly in g, its elements are determined by their coordinates in g_a. Hence θ is an abstract inclusion of h' in h.

COROLLARY. *The subgroup h defined in Theorem* 5.2 *is uniquely characterized by Lemma* 5.3.

Proof. Suppose h and h' are two connected LP-subgroups such that h is a covering group of h' and h' is a covering group of h, and such that the covering maps are abstract inclusions. Then the maps $h \to h' \to h$ and $h' \to h \to h'$ are both the identity. Hence $h = h'$.

DEFINITION 5.4. The LP-subgroup h of g defined in Theorem 5.2 is called the *canonical* LP-*subgroup* corresponding to the subalgebra H of G.

We have proved the following.

THEOREM 5.5. *Let g be an* LP-*group, G its Lie algebra. There exists a one to one correspondence between canonical* PL-*subgroups of g and closed subalgebras of G.*

THEOREM 5.6. *Let g be the universal group corresponding to an* LP-*algebra G. Let k be a closed normal connected subgroup of g. Then k is the canonical* LP-*subgroup corresponding to an ideal in G. Conversely, the canonical* LP-*subgroup corresponding to an ideal in G is a closed normal connected topological subgroup of g.*

Proof. Let k be a closed, normal, connected subgroup of g. Suppose $g = \lim g_a$, g_a simply connected Lie groups. Let the image of k in g_a be k_a. The closure $\bar{k}_a$ of k_a in g_a is a closed connected normal subgroup of g_a. Let K_a be the ideal of G_a corresponding to $\bar{k}_a$.

The image by f_{ab} of $\bar{k}_b$ in g_a $(a < b)$, is the analytic subgroup of g_a corresponding to the image of K_b in G_a. But the image is an ideal in

G_a, hence the corresponding analytic subgroup is a closed normal connected subgroup of g_a. Hence $\bar{k}_a \subset f_{ab}(\bar{k}_b)$. On the other hand since f_{ab} is continuous, $f_{ab}(\bar{k}_b) \subset \bar{k}_a$. Hence $f_{ab}(\bar{k}_b) = \bar{k}_a$.

Hence $k' = \lim \bar{k}_a$ exists and is a closed normal subgroup of g, and is the canonical LP-subgroup corresponding to $K = \lim K_a$. On the other hand, $k < K'$; but k maps onto k_a, a dense subgroup of $\bar{k}_a$. Hence k is dense in k' and $k = k'$ since k is closed.

The converse is obvious since the analytic subgroups of g_a corresponding to ideals are closed topological subgroups.

EXAMPLE. Consider the p-adic solenoid of Example 2, § 4. The additive group of reals R may be considered as an LP-subgroup of the p-adic solenoid P. Then P itself is the canonical LP-subgroup corresponding to the Lie algebra of R.

REFERENCES

1. C. Chevalley, *Theory of Lie groups*, Vol. 1, Princeton, 1946.
2. K. Iwasawa, *On some types of topological groups*, Ann. of Math., **50** (1949), 507–558.
3. S. Lefschetz, *Algebric topology*, Amer. Math. Soc. Colloq. Pub., no. 27, 1942.
4. J. Moore, *Introduction to homotopy theory*, Mimeographed Notes, Princeton Univ., 1955.
5. R. S. Novosad, *Simply connected spaces*, Trans. Amer. Math. Soc., **79** (1955), 216–228.
6. N. Steenrod, *The topology of fibre bundles*, Princeton, 1951.
7. A. Weil, *L'Integration dans les groupes topologique*, 2nd ed., Hermann & Cie. Paris, 1953.
8. H. Yamabe, *A generalization of a theorem of Gleason*, Ann. of Math., **58** (1953), 351–365.

UNIVERSITY OF CHICAGO

NOTE ON NORMAL NUMBERS

CALVIN T. LONG

Introduction. Let α be a real number with fractional part $.a_1a_2a_3\cdots$ when written to base r. Let Y_n denote the block of the first n digits in this representation and let $N(d, Y_n)$ denote the number of occurrences of the digit d in Y_n. The number α is said to be *simply normal* to base r if

$$\lim_{n\to\infty}\frac{N(d, Y_n)}{n}=\frac{1}{r}$$

for each of the r distinct choices of d. α is said to be *normal* to base r if each of the numbers $\alpha, r\alpha, r^2\alpha, \cdots$ are simply normal to each of the bases $r, r^2, r^3, \cdots$. These definitions, due to Emile Borel [1], were introduced in 1909. In 1940 S. S. Pillai [3] showed that a necessary and sufficient condition that α be normal to base r is that it be simply normal to each of the bases $r, r^2, r^3, \cdots$, thus considerably reducing the number of conditions needed to imply normality. The purpose of the present note is to show that α is normal to base r if and only if there exists a set of positive integers $m_1 < m_2 < m_3 < \cdots$ such that α is simply normal to base r^{m_i} for each $i \geqq 1$, and also to show that no finite set of m's will suffice.

Notation. We make use of the following additional conventions.

If B_k is any block of k digits to base r, $N(B_k, Y_n)$ will denote the number of occurrences of B_k in Y_n and $N_i(B_k, Y_n)$ will denote the number of occurrences of B_k starting in positions congruent to i modulo k in Y_n.

The term "relative frequency" will denote the asymptotic frequency with which an event occurs. For example, B_k occurs in (α), the fractional part of α, with relative frequency r^{-k} if $\lim_{n\to\infty} N(B_k, Y_n)/n = r^{-k}$.

Proof of the theorems. The following lemmas are easily proved.

LEMMA 1. *If* $\lim_{n\to\infty}\sum_{i=1}^{m} f_i(n)=1$ *and if* $\liminf_{n\to\infty} f_i(n) \geqq 1/m$ *for* $i=1, 2, \cdots, m$; *then* $\lim_{n\to\infty} f_i(n)=1/m$ *for each* i.

LEMMA 2. *The real number* α *is simply normal to base* r^k *if and*

Received July 5, 1956. Results in this paper were included in a doctoral dissertation written under the direction of Professor Ivan Niven at the University of Oregon. 1955.

only if $\lim_{n\to\infty} N_1(B_k, Y_n)/n = 1/kr^k$ *for every block* B_k *of* k *digits to base* r.

THEOREM 1. *The real number* α *is normal to base* r *if and only if there exist positive integers* $m_1 < m_2 < m_3 < \cdots$ *such that* α *is simply normal to each of the bases* $r^{m_1}, r^{m_2}, r^{m_3}, \cdots$.

Proof. The necessity of the condition follows immediately from the definition of normality.

Now suppose the condition of the theorem prevails. Let ν be an arbitrary positive integer and let B_ν be an arbitrary block of ν digits to base r. Choose k so large that $m_k > \nu$. It follows from Lemma 2 that

$$\lim_{n\to\infty} \frac{N_1(A_{m_k}, Y_n)}{n} = \frac{1}{m_k r^{m_k}}$$

for each block A_{m_k} of m_k digits to base r. If B_ν occurs exactly $t = t(A_{m_k})$ times in each A_{m_k}, then it follows that

$$\liminf_{n\to\infty} \frac{N(B_\nu, Y_n)}{n} \geqq \frac{T}{m_k r^{m_k}}$$

where $T = \sum t(A_{m_k})$ with the sum running over all blocks of m_k digits to base r. Now there are $r^{m_k-\nu}$ distinct blocks A_{m_k} which contain B_ν starting in position i for $i = 1, 2, \cdots, m_k - \nu + 1$ so that $T = (m_k - \nu + 1) r^{m_k - \nu}$. Thus it follows that

$$\liminf_{n\to\infty} \frac{N(B_\nu, Y_n)}{n} \geqq \frac{(m_k - \nu + 1) r^{m_k-\nu}}{m_k r^{m_k}} = \frac{1}{r^\nu} - \frac{\nu - 1}{m_k r^\nu}.$$

But, since this argument can be made for arbitrarily large values of k and $m_k \geqq k$, this implies that

$$\liminf_{n\to\infty} \frac{N(B_\nu, Y_n)}{n} \geqq \frac{1}{r^\nu}.$$

With Lemma 1 this implies that

$$\lim_{n\to\infty} \frac{N(B_\nu, Y_n)}{n} = \frac{1}{r^\nu}$$

so that α is normal to base r by a result of Niven and Zuckerman [2].

The next theorem implies that no finite set of m's will suffice in Theorem 1.

THEOREM 2. *If* $m_1, m_2, \cdots, m_s$ *is an arbitrary collection of distinct* *positive integers, then there exists at least one real number α simply normal to each of the bases $r^{m_1}, r^{m_2}, \cdots, r^{m_s}$ but not normal to base r.*

Proof. Writing to base r^m form the periodic decimal

$$\alpha = .\dot{0}12\ldots(\dot{r^m}-1)$$

where m is the least common multiple of $m_1, m_2, \cdots, m_s$. It is clear that α is simply normal to base r^m and that it is not normal to base r. To show that it is simply normal to base r^{m_i} for $i=1, 2, \cdots, s$ we prove that if d divides m then α is simply normal to base r^d.

Let $m=qd$ and let B_d be an arbitrary but fixed block of d digits to base r. In view of Lemma 2 it suffices to show that

$$\lim_{n\to\infty} \frac{N_1(B_d, Y_n)}{n} = \frac{1}{dr^d}.$$

A simple counting process shows that there are precisely $\binom{q}{i}(r^d-1)^{q-i}$ distinct blocks A_m of m digits to base r which contain B_d exactly i times starting in a position congruent to one modulo d. Therefore, since

$$\lim_{n\to\infty} \frac{N_1(A_m, Y_n)}{n} = \frac{1}{mr^m}$$

for each A_m, it follows that

$$\lim_{n\to\infty} \frac{N_1(B_d, Y_n)}{n} = \frac{1}{mr^m} \sum_{i=1}^{q} i\binom{q}{i}(r^d-1)^{q-i} = \frac{1}{dr^d}$$

as required.

References

1. Émile Borel, *Les probabilités dénombrablés et leurs applications arithmétiques*, Rend. Circ. Mat. Palmero **27** (1909), 247-271.

2. Ivan Niven and H. S. Zuckerman, *On the definition of normal numbers*, Pacific J. Math., **1** (1951), 103-109.

3. S. S. Pillai, *On normal numbers*, Proc. Indian Acad. Sci., Sect. A, **12** (1940), 179-184.

University of Oregon and
National Security Agency

ON CERTAIN SUMS GENERATING THE DEDEKIND SUMS AND THEIR RECIPROCITY LAWS

M. MIKOLÁS

1. Introduction. Let $\{u\}=u-[u]$ denote the fractional part of u and let $((u))=\{u\}-\frac{1}{2}$. Dedekind sums are defined for example, by

$$s_1(h,k)=\sum_{\lambda=0}^{k-1}\left(\left(\frac{\lambda}{k}\right)\right)\left(\left(\frac{\lambda h}{k}\right)\right) \tag{1.1}$$

where h and k are relatively prime positive integers. These sums which were studied by Dedekind [7], and more recently by Rademacher and Whiteman [9], [12] in connection with the theory of the modular function $\eta(\tau)$, occur also in the theory of partitions and in a great number of special papers. (Cf. for example [1]-[13].) The most important property of $s_1(h,k)$ is the reciprocity law

$$s_1(h,k)+s_1(k,h)=(h^2+3hk+k^2+1)/(12hk)\ . \tag{1.2}$$

A few years ago, Apostol [1] (for $r=\nu$) and Carlitz [3] introduced and investigated the so-called generalized Dedekind sums

$$s_r^{(\nu)}(h,k)=\sum_{\lambda=0}^{k-1}P_{\nu+1-r}\left(\frac{\lambda}{k}\right)P_r\left(\frac{\lambda h}{k}\right) \qquad 0\leqq r\leqq \nu+1\ , \tag{1.3}$$

P_r denoting the well-known Bernoulli function defined by the expansion

$$ze^{uz}/(e^z-1)=\sum_{n=0}^{\infty}P_n(u)z^n/n! \qquad |z|<2\pi$$

for $0\leqq u<1$ and by $P_r(u)=P_r(\{u\})$ for u arbitrary real. They found the corresponding extensions of (1.2) too.

Now, we shall continue to develop these results in two directions. Next we give a systematic treatment of certain exponential sums (2.1), (2.3) generating

$$\mathfrak{S}_{m,n}\begin{pmatrix} a & b \\ & c \end{pmatrix}=\sum_{\nu=0}^{c-1}P_m\left(\frac{\lambda a}{c}\right)P_n\left(\frac{\lambda b}{c}\right) \qquad m,n=0,1,2,\cdots \tag{1.4}$$

with $(a,c)=(b,c)=1$, $c>0$. We obtain (among others) a three-term relation of new type (Theorem 1) which implies (in extended form) all the above reciprocity theorems (see (5.1)-(5.10)). Let us remark that the sum function (2.5) with other notations is also used in [6]. On the other hand, we get a functional equation for

Received July 31, 1956.

$$\mathfrak{D}_c^{a,b}(w, z) = \sum_{\lambda=1}^{c-1} \zeta\left(w, \left\{\frac{\lambda a}{c}\right\}\right) \zeta\left(z, \left\{\frac{\lambda b}{c}\right\}\right) \tag{1.5}$$

where $\zeta(s, u)$ is the Hurwitz zeta function (Theorem 2). By

$$\zeta(1-n, u) = -P_n(u)/n \qquad 0 < u \leqq 1;\ n=1, 2, \cdots,$$

(1.5) can be regarded substantially as a (transcendental) generalization of (1.4).

2. Preliminaries on $\mathfrak{S}_c^{a,b}(x, y)$, $\mathfrak{s}_{m,n}\begin{pmatrix} a & b \\ & c & \end{pmatrix}$. In what follows, x, y, w, z denote complex variables, a, b and c are integers and $c > 0$; for brevity we write, as usual, $e(z) = e^{2\pi i z}$.

Let us put

$$S_c^{a,b}(x, y) = \sum_{\lambda (\mathrm{mod}\, c)} e\left(\left\{\frac{\lambda a}{c}\right\} x + \left\{\frac{\lambda b}{c}\right\} y\right) \tag{2.1}$$

with $(a, c) = (b, c) = 1$, the summation extending over a complete residue system modulo c. It is obvious that (2.1) is independent of the choice of this residue system[1] and for $a = b$ or $c = 1, 2$ it is independent of a, b. The function $S_c^{a,b}(x, y)$ remains unaltered if we change a, b or x, y by multiplies of c. By this periodicity, it is no restriction to suppose for example, that $0 \leqq \Re(x) < c$, $-c < \Re(y) \leqq 0$.

We have $S_c^{a,b}(x, y) = S_c^{b,a}(y, x)$ and

$$S_c^{-a,b}(x, y) = e(x) S_c^{a,b}(-x, y) + 1 - e(x), \tag{2.2}$$

since $\{-u\} = 0$ or $1 - \{u\}$ according as u is an integer or not.

The function

$$\mathfrak{S}_c^{a,b}(x, y) = [e(x)-1]^{-1}[e(y)-1]^{-1} S_c^{a,b}(x, y) \qquad x, y \neq 0, \pm 1, \cdots \tag{2.3}$$

has corresponding trivial properties; in particular, (2.2) implies

$$\mathfrak{S}_c^{-a,b}(x, y) = -\mathfrak{S}_c^{a,b}(-x, y) - [e(y)-1]^{-1}. \tag{2.4}$$

By the definition of Bernoulli functions and (1.4) we obtain

$$xy\,\mathfrak{S}_c^{a,b}(x/2\pi i, y/2\pi i) = \sum_{m,n=0}^{\infty} \frac{x^m y^n}{m!\,n!} \mathfrak{s}_{m,n}\begin{pmatrix} a & b \\ & c & \end{pmatrix} \qquad |x|, |y| < 2\pi. \tag{2.5}$$

Here

[1] Hence we see that $S_c^{a,b}(x, y) = S_c^{1,b'}(x, y)$ for a suitable integer b'; however, the above symmetric notation seems the most convenient.

$$\mathfrak{S}_{0,n}\begin{pmatrix} a\ b \\ c \end{pmatrix}=\sum_{l=0}^{c-1} P_n\begin{pmatrix} l \\ c \end{pmatrix}=c^{1-n}B_n \qquad n=0,\, 1,\, \cdots , \tag{2.6}$$

$B_n=P_n(0)$ denoting the Bernoullian numbers.

Note that $\mathfrak{S}_{m,n}\begin{pmatrix} a\ b \\ c \end{pmatrix}=\mathfrak{S}_{n,m}\begin{pmatrix} b\ a \\ c \end{pmatrix}$ and $\mathfrak{S}_{m,n}\begin{pmatrix} a\ a \\ c \end{pmatrix}$ does not depend on a; especially we have $\mathfrak{S}_{m,n}\begin{pmatrix} 1\ b \\ c \end{pmatrix}=\mathfrak{S}_n^{(m+n-1)}(b,\, c)$, furthermore

$$\mathfrak{S}_{m,n}\begin{pmatrix} a\ b \\ 1 \end{pmatrix}=B_mB_n\ , \qquad \mathfrak{S}_{m,n}\begin{pmatrix} a\ b \\ 2 \end{pmatrix}=B_mB_n[1+(1-2^{1-m})(1-2^{1-n})] \tag{2.7}$$

$$m,\, n=0,\, 1,\, \cdots .$$

3. Representation by cotangents and Eulerian numbers respectively. Let $c>1$. The identity

$$\sum_{\mu=0}^{c-1} e\left(\frac{\mu x}{c}\right)e\left(\frac{\mu\nu}{c}\right)=[e(x)-1]\left[e\left(\frac{x+\nu}{c}\right)-1\right]^{-1} \tag{3.1}$$

yields after multiplication by $e\left(-\frac{\mu\nu}{c}\right)$ $(\nu=0,\, 1,\, \cdots,\, c-1)$ and summation

$$e\left(\frac{\mu x}{c}\right)=\frac{1}{c}[e(x)-1]\sum_{\nu=0}^{c-1}\left[e\left(\frac{x+\nu}{c}\right)-1\right]^{-1}e\left(-\frac{\mu\nu}{c}\right) \tag{3.2}$$

$$\mu=0,\, 1,\, \cdots,\, \nu-1\, ;$$

(3.1) and (3.2) hold clearly provided that $(x+\nu)/c$ is not an integer $(\nu=0,\, 1,\, \cdots,\, c-1)$. Hence by putting $\mu=c\{a\lambda/c\}$, a and c being coprime we get

$$e\left(x\left\{\frac{a\lambda}{c}\right\}\right)=\frac{1}{c}[e(x)-1]\sum_{\nu=0}^{c-1}\left[e\left(\frac{x+\nu}{c}\right)-1\right]^{-1}e\left(-\nu\frac{a\lambda}{c}\right) . \tag{3.3}$$

Furthermore, by using the corresponding expression for $e(y\{b\lambda/c\})$, $(b,\, c)=1$,

$$S_c^{a,b}(x,\, y)=\frac{1}{c^2}[e(x)-1][e(y)-1]\sum_{p,q (\operatorname{mod} c)}\left[e\left(\frac{x+p}{c}\right)-1\right]^{-1}\left[e\left(\frac{y+q}{c}\right)-1\right]^{-1}$$

$$\times\sum_{\lambda=0}^{c-1} e\left(-\frac{\lambda(ap+bq)}{c}\right) .$$

If we consider the complete residue systems (mod c): $p=-br$, $q=a\rho$ $(r,\, \rho=0,\, 1,\, \cdots,\, c-1)$ and take into account that $\sum_{\lambda=0}^{c-1} e\left(-\lambda\frac{ab(\rho-r)}{c}\right)$

vanishes except for $\rho=r$ when it has the value c, it follows simply that

$$(3.4)\qquad \mathfrak{S}_c^{a,b}(x,y)=\frac{1}{c}\sum_{r(\mathrm{mod}\,c)}\left[e\left(\frac{x-br}{c}\right)-1\right]^{-1}\left[e\left(\frac{y+ar}{c}\right)-1\right]^{-1},$$

holds for all $x, y\neq 0, \pm 1, \cdots$ and, because of the definition (2.3), in the case $c=1$ too. By $[1-e(z)]^{-1}=\frac{1}{2}(1+i\operatorname{ctg}\pi z)$ and

$$\sum_{\mu=0}^{c-1}\operatorname{ctg}\pi\left(z+\frac{\mu}{c}\right)=c\cdot\operatorname{ctg}c\pi z\ ,$$

we have the equivalent formula:

$$(3.5)\qquad \mathfrak{S}_c^{a,b}(x,y)=\frac{1}{4}[1+i(\operatorname{ctg}\pi x+\operatorname{ctg}\pi y)]$$
$$-\frac{1}{4c}\sum_{r(\mathrm{mod}\,c)}\operatorname{ctg}\pi\frac{x-br}{c}\operatorname{ctg}\pi\frac{y+ar}{c}\ ;$$

(3.4) or (3.5) expresses the sum (2.3) by means of periodic elementary functions, *without* using the arithmetical function $\{u\}$.

(3.4) leads immediately to corresponding representations of $\mathfrak{s}_{m,n}\binom{a\ b}{c}$ by means of the so-called Eulerian numbers $H_n(\eta^k)$, defined for a root of unity $\eta^k=e\left(\frac{k}{c}\right)$, $c>1$, $c\nmid k$ by

$$(3.6)\qquad (1-\eta^k)/(e^z-\eta^k)=\sum_{n=0}^{\infty}H_n(\eta^k)z^n/n!\qquad |z|<2\pi\{k/c\}\ .$$

In fact, after expanding the right-hand members of

$$xy\mathfrak{S}_c^{a,b}(x/2\pi i,\ y/2\pi i)=(xy/c)(e^{x/c}-1)^{-1}(e^{y/c}-1)^{-1}$$
$$+(xy/c)\sum_{r=1}^{c-1}(e^{x/c}\eta^{-br}-1)^{-1}(e^{y/c}\eta^{ar}-1)^{-1}\ ,$$

we find

$$(3.7)\qquad xy\mathfrak{S}_c^{a,b}(x/2\pi i,\ y/2\pi i)=c+\sum_{n=1}^{\infty}\frac{B_n}{n!c^{n-1}}(x^n+y^n)$$
$$+\sum_{m,n=1}^{\infty}\frac{x^my^n}{m!n!c^{m+n-1}}\left[B_mB_n+mn\sum_{r=1}^{c-1}\frac{H_{m-1}(\eta^{br})H_{n-1}(\eta^{-ar})}{(\eta^{ar}-1)(\eta^{-br}-1)}\right]\quad |x|,\ |y|<\frac{2\pi}{c}\ ,$$

so that comparison with (2.5) gives in addition to (2.6)

$$(3.8)\qquad \mathfrak{s}_{m,n}\binom{a\ b}{c}=\frac{1}{c^{m+n-1}}\left[B_mB_n+mn\sum_{r=1}^{c-1}\frac{H_{m-1}(\eta^{br})H_{n-1}(\eta^{-ar})}{(\eta^{ar}-1)(\eta^{-br}-1)}\right]$$
$$m,\ n=1, 2, \cdots ,$$

a formula implying a result of Carlitz [3, (6.5)]. In particular, for $m=n=1$ (3.8) becomes

$$\mathfrak{s}_{11}\begin{pmatrix} a & b \\ & c \end{pmatrix}=\frac{1}{4c}+\frac{1}{c}\sum_{r=1}^{c-1}(\eta^{ar}-1)^{-1}(\eta^{-br}-1)^{-1} \tag{3.9}$$

$$=\frac{1}{4}+\frac{1}{4c}\sum_{r=1}^{c-1}\operatorname{ctg}\frac{\pi ar}{c}\operatorname{ctg}\frac{\pi br}{c},$$

which contains two equivalent representations due to Rademacher and Rédei (for $a=1$; cf. for example, [4], (2.2) and [2], (5) respectively).

4. The main property of $\mathfrak{S}_c^{a,b}(x, y)$. Our next purpose is to deduce a peculiar symmetry relation relating to the sums in question, by applying the calculus of residues.

THEOREM 1. *We have for a, b, c positive, mutually coprime, and for $0\leq\Re(x)<1$, $-1<\Re(y)\leq 0$ the relation*

$$\mathfrak{S}_b^{c,a}(ax+by,\ -cx)+\mathfrak{S}_c^{a,b}(cx,\ cy)+\mathfrak{S}_a^{b,c}(-cy,\ ax+by) =[1-e(ax+by)]^{-1}, \tag{4.1}$$

provided that $ax+by$, cx and cy are not integers.

Proof. We consider the integral

$$\mathfrak{F}=\frac{1}{2\pi i}\int_Q [e(z)-1]^{-1}\left[e\left(x-\frac{b}{c}z\right)-1\right]^{-1}\left[e\left(y+\frac{a}{c}z\right)-1\right]^{-1}dz \tag{4.2}$$

the path of integration being a rectangle whose vertices are the points $-\varepsilon\pm ti$, $c-\varepsilon\pm ti$ with

$$t>\max\left\{\frac{c}{b}|\Im(x)|,\ \frac{c}{a}|\Im(y)|\right\}$$

and

$$0<\varepsilon<\min\left\{\frac{c}{b}(1-\Re(x)),\ \frac{c}{a}(1+\Re(y))\right\},$$

taken in positive direction. A straight-forward calculation shows that only singularities of the integrand inside Q are at the points:

$$z=\lambda \qquad \lambda=0,\ 1,\ \cdots,\ c-1;$$

$$z=\frac{c}{b}(\mu+x) \qquad \mu=0,\ 1,\ \cdots,\ b-1;$$

$$z=\frac{c}{a}(\nu-y) \qquad \nu=0,\ 1,\ \cdots,\ a-1;$$

by our assumptions, these are all distinct and poles of order 1 only of the first, second, and third factor respectively. Since

$$\operatorname*{res}_{z=\lambda}[e(z)-1]^{-1}=1/2\pi i$$

$$\operatorname*{res}_{z=(c/b)(\mu+x)}[e(x-bz/c)-1]^{-1}=-c/2\pi ib\ ,$$

$$\operatorname*{res}_{z=(c/a)(\nu-y)}[e(y+az/c)-1]^{-1}=c/2\pi ia\ ,$$

the residue theorem yields

$$\begin{aligned}2\pi i\cdot\mathfrak{F}=&\sum_{\lambda=0}^{c-1}\left[e\left(x-\frac{\lambda b}{c}\right)-1\right]^{-1}\left[e\left(y+\frac{\lambda a}{c}\right)-1\right]^{-1}\\&-\frac{c}{b}\sum_{\mu=0}^{b-1}\left[e\left(\frac{a}{b}x+y+\frac{\mu a}{b}\right)-1\right]^{-1}\left[e\left(\frac{c}{b}x+\frac{\mu c}{b}\right)-1\right]^{-1}\\&+\frac{c}{a}\sum_{\nu=0}^{a-1}\left[e\left(x+\frac{b}{a}y+\frac{\nu b}{a}\right)-1\right]^{-1}\left[e\left(-\frac{c}{a}y+\frac{\nu c}{a}\right)-1\right]^{-1}\end{aligned}$$

and therefore, by (3.4), we obtain

(4.3) $\quad \mathfrak{S}_c^{a,b}(cx,\ cy)-\mathfrak{S}_b^{c,-a}(ax+by,\ cx)+\mathfrak{S}_a^{c,b}(ax+by,\ -cy)=(2\pi i/c)\mathfrak{F}\ .$

Now, if we write

$$\int_Q=\int_{c-\varepsilon-ti}^{c-\varepsilon+ti}+\int_{c-\varepsilon+ti}^{-\varepsilon+ti}+\int_{-\varepsilon+ti}^{-\varepsilon-ti}+\int_{-\varepsilon-ti}^{c-\varepsilon-ti}$$

with the integrand of (4.2) and straight-line paths, the sum of the first and third member on the right vanishes because of the periodicity (with period c) of

$$[e(z)-1]^{-1}[e(x-bz/c)-1]^{-1}[e(y+az/c)-1]^{-1}\ .$$

On the other hand, using the estimate $|e(u+iv)-1|\geqq|e^{-2\pi v}-1|$ (u, v arbitrary real), we find at once that the integrals along the horizontal segments tend to zero as $t\to\infty$. Hence (4.3) implies for $t\to\infty$

(4.4) $\quad \mathfrak{S}_a^{c,b}(ax+by,\ -cy)-\mathfrak{S}_b^{c,-a}(ax+by,\ cx)+\mathfrak{S}_c^{a,b}(cx,\ cy)=0$

which is, by (2.4), equivalent to (4.1).

5. Applications; extension of the well-known reciprocity theorems. (1) If we write

(5.1) $$\mathfrak{T}_c^{a,b}(x,\ y)=\frac{1}{c}\sum_{r(\mathrm{mod}\ c)}\operatorname{ctg}\pi\frac{x-br}{c}\operatorname{ctg}\pi\frac{y+ar}{c}$$

and use (3.5), then (4.1) becomes

$$\mathfrak{T}_b^{c,a}(ax+by,\ -cx)+\mathfrak{T}_c^{a,b}(cx,\ cy)+\mathfrak{T}_a^{b,c}(-cy,\ ax+by)=1\ . \tag{5.2}$$

By (3.9), this may be regarded as a generalization of the reciprocity theorem of Dedekind sums. For, by putting $y=-x$ in (5.2) and making $x\to 0$, we obtain on the basis of the Laurent expansion $\operatorname{ctg} z=z^{-1}-\frac{1}{3}z-\cdots$

$$\mathfrak{S}_{11}\begin{pmatrix} b & c \\ & a \end{pmatrix}+\mathfrak{S}_{11}\begin{pmatrix} c & a \\ & b \end{pmatrix}+\mathfrak{S}_{11}\begin{pmatrix} a & b \\ & c \end{pmatrix}=\frac{1}{2}+\frac{1}{12}\left(\frac{a}{bc}+\frac{b}{ca}+\frac{c}{ab}\right), \tag{5.3}$$

a remarkably symmetric three-term relation which for $a=1$ reduces to (1.2) with $h=b$, $k=c$. (Cf. also a result of Rademacher in [11].)

(2) Let us replace in (4.1) x, y by $x/2\pi i$ and $y/2\pi i$ respectively, multiply both sides by $c^2xy(ax+by)$ and expand every member by applying (2.5), (2.6) and the power series of $z/(e^z-1)$. We obtain

$$\begin{aligned} cy\sum_{m,n=1}^{\infty}&\frac{(ax+by)^m(-cx)^n}{m!n!}\mathfrak{S}_{m,n}\begin{pmatrix} c & a \\ & b \end{pmatrix}-(ax+by)\sum_{m,n=1}^{\infty}\frac{(cx)^m(cy)^n}{m!n!}\mathfrak{S}_{m,n}\begin{pmatrix} a & b \\ & c \end{pmatrix} \\ &+cx\sum_{m\,n=1}^{\infty}\frac{(-cy)^m(ax+by)^n}{m!n!}\mathfrak{S}_{m,n}\begin{pmatrix} b & c \\ & a \end{pmatrix}=c^2xy\left[1+\sum_{\nu=1}^{\infty}\frac{B_\nu}{\nu!}(ax+by)^\nu\right] \\ &-cy\sum_{\nu=1}^{\infty}\frac{B_\nu}{\nu!b^{\nu-1}}[(ax+by)^\nu+(-cx)^\nu]+c(ax+by)\sum_{\nu=1}^{\infty}\frac{B_\nu}{\nu!}(x^\nu+y^\nu) \\ &-cx\sum_{\nu=1}^{\infty}\frac{B_\nu}{\nu!\,a^{\nu-1}}[(-cy)^\nu+(ax+by)^\nu]\ , \end{aligned}$$

this holding identically for $|x|$, $|y|<2\pi$. If one uses still the binomial theorem and arranges our absolutely convergent series in terms of x^ν, y^ν ($\nu=1, 2, \cdots$), then comparison of the corresponding coefficients leads without difficulty to the following system of relations:

$$\begin{aligned} a^\nu\cdot(\nu+1)b^\nu c\,\mathfrak{S}_{1,\nu}\begin{pmatrix} b & c \\ & a \end{pmatrix}&+b^\nu\sum_{\mu=1}^{\nu}(-1)^{\mu+1}\binom{\nu+1}{\mu}c^\mu a^{\nu+1-\mu}\mathfrak{S}_{\nu+1-\mu,\mu}\begin{pmatrix} c & a \\ & b \end{pmatrix} \\ +c^\nu\cdot(\nu+1)ab^\nu\,\mathfrak{S}_{\nu,1}\begin{pmatrix} a & b \\ & c \end{pmatrix}&=B_{\nu+1}(a^{\nu+1}+\nu b^{\nu+1}+(-c)^{\nu+1})-(\nu+1)B_\nu(ab)^\nu c \\ &\qquad\qquad \nu=1,\ 2,\ \cdots\ , \end{aligned} \tag{5.4}$$

furthermore, by $\binom{\alpha}{\beta}\binom{\gamma}{\alpha}=\binom{\gamma}{\beta}\binom{\gamma-\beta}{\gamma-\alpha}$,

$$\begin{aligned} a^\nu\cdot\binom{\nu+1}{p+1}&\sum_{\mu=1}^{p}(-1)^{\mu+1}\binom{p+1}{\mu}b^{\nu+1-\mu}c^\mu\mathfrak{S}_{\mu,\nu+1-\mu}\begin{pmatrix} b & c \\ & a \end{pmatrix} \\ &+b^\nu\cdot\binom{\nu+1}{p}\sum_{\mu=1}^{\nu+1-p}(-1)^{\mu+1}\binom{\nu+1-p}{\mu}c^\mu a^{\nu+1-\mu}\mathfrak{S}_{\nu+1-\mu,\mu}\begin{pmatrix} c & a \\ & b \end{pmatrix} \end{aligned} \tag{5.5}$$

$$+c^{\nu}\cdot\left[\binom{\nu+1}{p+1}a^{p+1}b^{\nu-p}\mathfrak{S}_{\nu-p,p+1}\binom{a\ b}{c}+\binom{\nu+1}{p}a^{p}b^{\nu+1-p}\mathfrak{S}_{\nu+1-p,p}\binom{a\ b}{c}\right.$$

$$=B_{\nu+1}\left[\binom{\nu+1}{p}a^{\nu+1}+\binom{\nu+1}{p+1}b^{\nu+1}\right]-(\nu+1)B_{\nu}\binom{\nu}{p}(ab)^{\nu}c$$

$$1\leqq p\leqq \nu-1\,.$$

The results can be written briefly in symbolic form as follows

$$(\nu+1)\left[ca^{\nu}\mathfrak{S}_{1,\nu}\binom{b\ c}{a}+c^{\nu}a\mathfrak{S}_{\nu,1}\binom{a\ b}{c}\right]-(a\mathfrak{S}-c\bar{\mathfrak{S}})^{\nu+1}\binom{c\ a}{b} \tag{5.6}$$

$$=\nu B_{\nu+1}b-(\nu+1)B_{\nu}a^{\nu}c \qquad \nu=1,\,2,\,\cdots\,,$$

$$a^{\nu}\cdot\binom{\nu+1}{p+1}(b\mathfrak{S}-c\bar{\mathfrak{S}})^{p+1}(b\mathfrak{S})^{\nu-p}\cdot\binom{c\ b}{a} \tag{5.7}$$

$$+b^{\nu}\cdot\binom{\nu+1}{p}(a\mathfrak{S}-c\bar{\mathfrak{S}})^{\nu+1-p}(a\mathfrak{S})^{p}\binom{c\ a}{b}$$

$$-c^{\nu}\cdot\left[\binom{\nu+1}{p+1}a\mathfrak{S}+\binom{\nu+1}{p}b\bar{\mathfrak{S}}\right](a\mathfrak{S})^{p}(b\bar{\mathfrak{S}})^{\nu+p}\binom{a\ b}{c}$$

$$=(p+1)\binom{\nu+1}{p+1}B_{\nu}a^{\nu}b^{\nu}c \qquad p=1,\,2,\,\cdots\,;\ \nu=p+1,\,p+2,\,\cdots\,,$$

where for example

$$(b\mathfrak{S}-c\bar{\mathfrak{S}})^{p+1}(b\mathfrak{S})^{\nu-p}\binom{c\ b}{a}$$

means that, after formal application of the binomial theorem to the first factor and formal multiplication by $b^{\nu-p}\cdot\mathfrak{S}^{\nu-p}\cdot\binom{c\ b}{a}$, every product $\mathfrak{S}^{m}\bar{\mathfrak{S}}^{n}\binom{c\ b}{a}$ is replaced by $\mathfrak{S}_{m,n}\binom{c\ b}{a}$.

(3) We remark at once that (5.4), (5.6) go over for $\nu=1$ to the reciprocity relation (5.3) and for $\nu>1$ odd, $b=1$ to the formula (cf. (1.3), (2.7))

$$(\nu+1)[ca^{\nu}\cdot s_{\nu}^{(\nu)}(c,\,a)+c^{\nu}a,\,s_{\nu}^{(\nu)}(a,\,c)]=(Bc-Ba)^{\nu+1}+\nu B_{\nu+1} \tag{5.8}$$

with [2]

$$(Bc-Ba)^{\nu+1}=\sum_{\mu=0}^{\nu+1}(-1)^{\mu}\binom{\nu+1}{\mu}c^{\mu}a^{\nu+1-\mu}B_{\mu}B_{\nu+1-\mu}\,;$$

[2] The factor $(-1)^{\mu}$ may plainly be suppressed in the last summand, that is,

$$(Bc-Ba)^{\nu+1}=(Bc+Ba)^{\nu+1}\,.$$

therefore (5.4), (5.6) generalize (5.3) and Apostol's reciprocity theorem [1, Theorem 1].

On the other hand, putting $\nu=3, 5, 7, \cdots$ in (5.7), we get for $c=1$

$$
\begin{aligned}
(5.9)\qquad & \binom{\nu+1}{p+1}a^{\nu-p}(s^{(\nu)}-b)^{p+1}(b, a)-\binom{\nu+1}{p}b^{p}(s^{(\nu)}-a)^{\nu+1-p}(a, b) \\
&=\binom{\nu+1}{p+1}aB_{\nu-p}B_{p+1}-\binom{\nu+1}{p}bB_{\nu+1-p}B_{p}\ ,
\end{aligned}
$$

while the case $b=1$ yields

$$
\begin{aligned}
(5.10)\qquad & c^{\nu}\left[\binom{\nu+1}{p+1}as^{(\nu)}_{\nu-p}(a, c)+\binom{\nu+1}{p}s^{(\nu)}_{\nu+1-p}(a, c)\right] \\
&=\binom{\nu+1}{p+1}(s^{(\nu)}-c)^{p+1}(as^{(\nu)})^{\nu-p}(c, a)+\binom{\nu+1}{p}(aB-c\overline{B})^{\nu+1-p}B^{p}\ ,
\end{aligned}
$$

the symbolic notations being understood in similar sense as above. (5.9) and (5.10) express the first and second reciprocity law of Carlitz respectively [3, Theorems 1, 2][3], so that we have in (5.5), (5.7) a common extension of them.

6. The sum $\mathfrak{D}_c^{a,b}(w, z)$. We now use the generalized zeta function, defined by

$$\zeta(z, u)=\sum_{n=0}^{\infty}(u+n)^{-z}$$

for $\Re(z)>1$ and by analytic continuation for other values $\neq 1$ of z, u denoting a fixed number with $0<u\leqq 1$. There holds the well-known formula of Hurwitz:

$$
\begin{aligned}
(6.1)\qquad & \zeta(z, u)=2(2\pi)^{z-1}\Gamma(1-z) \\
&\times\left(\sin\frac{\pi z}{2}\sum_{n=1}^{\infty}n^{z-1}\cos 2n\pi u+\cos\frac{\pi z}{2}\sum_{n=1}^{\infty}n^{z-1}\sin 2n\pi u\right)\qquad \Re(z)<0\ .
\end{aligned}
$$

Next we establish a functional equation for the sum

$$(6.2)\qquad \mathfrak{D}_c^{a,b}(w, z)=\sum_{\lambda=1}^{c-1}\zeta\left(w, \left\{\frac{\lambda a}{c}\right\}\right)\zeta\left(z, \left\{\frac{\lambda b}{c}\right\}\right)$$

with $(a, c)=(b, c)=1$, $c>1$, in observing that [cf. (1.4)]

$$(6.3)\qquad \mathfrak{D}_c^{a,b}(1-m, 1-n)=\frac{1}{mn}\left[\mathfrak{S}_{m,n}\binom{a\ \ b}{c}-B_mB_n\right]\qquad m, n=1, 2, \cdots$$

[3] In formula (3.2) of [**3**], the lack of the corresponding binomial coefficients before the Bernoullian numbers appears to be a typographical error.

and, by $\zeta(z, \frac{1}{2})=(2^z-1)\zeta(z)$ where $\zeta(z)=\zeta(z, 1)$ is Riemann's zeta function,

$$\mathfrak{D}_2^{a,b}(w, z)=(2^w-1)(2^z-1)\cdot\zeta(w)\zeta(z) . \tag{6.4}$$

THEOREM 2. *For $(a, c)=(b, c)=1$, $c>2$ and for any w, z distinct from 0 and 1 we have the relation*

$$\begin{aligned}\mathfrak{D}_c^{a,b}(w, z)=&(c^{w+z}-1)\zeta(w)\zeta(z)+\pi^{-1}(2c\pi)^{w+z-1}\Gamma(1-w)\Gamma(1-z)\\ &\times\left\{\cos\frac{\pi}{2}(w-z)\mathfrak{D}_c^{b,a}(1-w, 1-z)-\cos\frac{\pi}{2}(w+z)\mathfrak{D}_c^{b,-a}(1-w, 1-z)\right\} .\end{aligned} \tag{6.5}$$

Proof. 1° First let $\Re(w)<0$, $\Re(z)<0$. We transform

$$\overline{\mathfrak{D}}_c^{a,b}(w, z)=\sum_{\lambda=1}^{c}\zeta\left(w, \left\{\frac{\lambda a}{c}\right\}\right)\zeta\left(z, \left\{\frac{\lambda b}{c}\right\}\right) \tag{6.6}$$

by means of (6.1).

Since the series involved in Hurwitz's formula are absolutely convergent, one obtains after substitution into (6.6)

$$\begin{aligned}\overline{\mathfrak{D}}_c^{a,b}(w, z)=&4(2\pi)^{w+z-2}\Gamma(1-w)\Gamma(1-z)\\ &\times\sum_{m,n=1}^{\infty}m^{w-1}n^{z-1}\left(\phi_{m,n}\cdot\sin\frac{\pi w}{2}\sin\frac{\pi z}{2}+\psi_{m,n}\cdot\cos\frac{\pi w}{2}\cos\frac{\pi z}{2}\right) ,\end{aligned} \tag{6.7}$$

where

$$\phi_{m,n}=\sum_{\mu=1}^{c}\cos 2m\pi\frac{\mu a}{c}\cos 2n\pi\frac{\mu b}{c}=\begin{cases}c, \text{ if } & c\mid am\pm bn ,\\ 0 \text{ for } & c\nmid am\pm bn ,\\ c/2 \text{ otherwise} ,\end{cases} \tag{6.8}$$

$$\psi_{m,n}=\sum_{\mu=1}^{c}\sin 2m\pi\frac{\mu a}{c}\sin 2n\pi\frac{\mu b}{c}=\begin{cases}c/2, \text{ if } c\mid am-bn \text{ but } c\nmid am+bn ,\\ -c/2, \text{ if } c\mid am+bn \text{ and } c\nmid am-bn ,\\ 0 \text{ otherwise} .\end{cases} \tag{6.9}$$

Hence it follows easily that

$$\begin{aligned}\overline{\mathfrak{D}}_c^{a,b}(w, z)=&2c(2\pi)^{w+z-2}\Gamma(1-w)\Gamma(1-z)\cdot\left\{2\sin\frac{\pi w}{2}\sin\frac{\pi z}{2}\sum_{c\mid m, c\mid n}m^{w-1}n^{z-1}\right.\\ &\left.+\cos\frac{\pi}{2}(w-z)\sum_{\substack{am\equiv bn(\mathrm{mod}\ c)\\ c\nmid m, c\nmid n}}m^{w-1}n^{z-1}-\cos\frac{\pi}{2}(w+z)\sum_{\substack{am\equiv -bn(\mathrm{mod}\ c)\\ c\nmid m, b\nmid n}}m^{w-1}n^{z-1}\right\} .\end{aligned} \tag{6.10}$$

Now, by the functional equation of $\zeta(s)$ we have

$$(6.11)\qquad 4c(2\pi)^{w+z-2}\Gamma(1-w)\Gamma(1-z)\sin\frac{\pi w}{2}\sin\frac{\pi z}{2}\sum_{c\mid m,\, c\mid n} m^{w-1}n^{z-1} = c^{w+z-1}\zeta(w)\zeta(z)\,.$$

Furthermore, ar $(r=0, 1, \cdots, c-1)$ and br $(r=0, 1, \cdots, c-1)$ being complete systems of residues mod c, we can write

$$(6.12)\qquad \sum_{\substack{am\equiv bn \pmod c\\ c\nmid m,\, c\nmid n}} m^{w-1}n^{z-1} = \sum_{r=1}^{c-1}\Big(\sum_{m\equiv rb \pmod c} m^{w-1}\Big)\Big(\sum_{n\equiv ra \pmod c} n^{z-1}\Big)$$
$$= c^{w+z-2}\sum_{r=1}^{c-1}\Big[\sum_{M=0}^{\infty}\Big(\Big\{\frac{rb}{c}\Big\}+M\Big)^{w-1}\Big]\Big[\sum_{N=1}^{\infty}\Big(\Big\{\frac{ra}{c}\Big\}+N\Big)^{z-1}\Big]$$
$$= c^{w+z-2}\sum_{r=1}^{c-1}\zeta\Big(1-w, \Big\{\frac{rb}{c}\Big\}\Big)\zeta\Big(1-z, \Big\{\frac{ra}{c}\Big\}\Big)$$

and similarly

$$(6.13)\qquad \sum_{\substack{am\equiv -bn \pmod c\\ c\nmid m,\, c\nmid n}} m^{w-1}n^{z-1} = \sum_{r=1}^{c-1}\Big(\sum_{m\equiv rb \pmod c} m^{w-1}\Big)\Big(\sum_{n\equiv -ra \pmod c} n^{z-1}\Big)$$
$$= c^{w+z-2}\sum_{r=1}^{c-1}\zeta\Big(1-w, \Big\{\frac{rb}{c}\Big\}\Big)\zeta\Big(1-z, \Big\{\frac{ra}{c}\Big\}\Big)\,.$$

(6.10)$-$(6.13) yield together

$$(6.14)\qquad \overline{\mathfrak{D}}_c^{a,b}(w, z) = c^{w+z-1}\zeta(w)\zeta(z) + \pi^{-1}(2c\pi)^{w+z-1}\Gamma(1-w)\Gamma(1-z)$$
$$\times\Big\{\cos\frac{\pi}{2}(w-z)\mathfrak{D}_c^{b,a}(1-w, 1-z) - \cos\frac{\pi}{2}(w+z)\mathfrak{D}_c^{b,-a}(1-w, 1-z)\Big\}\,.$$

2° Finally, (6.5) follows immediately from (6.14), in view of

$$\mathfrak{D}_c^{a,b}(w, z) = \overline{\mathfrak{D}}_c^{a,b}(w, z) - \zeta(w)\zeta(z) \qquad \Re(w)<0\,,\ \Re(z)<0$$

and by analytic continuation.

7. Some remarks. In [2], Apostol finds certain finite sum representations for $s_\nu^{(\nu)}(h, k)$, involving cotangents, $\zeta(z, u)$, $\Gamma'(z)/\Gamma(z)$ and he uses these expressions to give a short analytic proof of (5.8) [Theorems 1, 2]. It may be noted that the above Theorem 2 implies the results in question, arising as limiting cases for $w\to 0$, and $z\to 0$, $z=-1, -2, \cdots$.

The form of $\mathfrak{S}_c^{a,b}(x, y)$, $\mathfrak{D}_c^{a,b}(w, z)$ suggests applications in connection with certain Lambert series, generalizing those investigated by Rademacher, Apostol and Carlitz. I hope to return on this problem in another paper.

References

1. T. M. Apostol, *Generalized Dedekind sums and transformation formulae of certain Lambert series*, Duke Math. J., **17** (1950), 147–157.
2. ———, *Theorems on generalized Dedekind sums*, Pacific J. Math., **2** (1952), 1–9.
3. L. Carlitz, *Some theorems on generalized Dedekind sums*, Pacific J. Math., **3** (1953), 513–522.
4. ———, *The reciprocity theorem for Dedekind sums*, Pacific J. Math., **3** (1953), 523–527.
5. ———, *A note on generalized Dedekind sums*, Duke Math. J., **21** (1954), 399–403.
6. ———, *Dedekind sums and Lambert series*, Proc. Amer. Math. Soc., **5** (1954), 580–584.
7. R. Dedekind, *Erläuterungen zu Riemann's Fragmenten über die Grenzfälle der elliptischen Funktionen*, Gesammelte mathematische Werke, vol. 1, Braunschweig, 1930, 159–173.
8. L. J. Mordell, *The reciprocity formula for Dedekind sums*, Amer. J. Math. **73** (1951), 593–598.
9. H. Rademacher, *Zur Theorie der Modulfunktionen*, J. Reine Angew. Math., **167** (1932), 312–336.
10. ———, *Die Reziprozitätsformel für Dedekindsche Summen*, Acta Sci. Math. Szeged, **12B** (1950), 57–60.
11. ———, *Generalization of the reciprocity formula for Dedekind sums*, Duke Math. J., **21** (1954), 391–397.
12. H. Rademacher and A. Whiteman, *Theorems on Dedekind sums*, Amer. J. Math., **63** (1941), 377–407.
13. L. Rédei, *Elementarer Beweis und Verallgemeinerung einer Reziprozitätsformel von Dedekind*, Acta Sci. Math. Szeged, **12B** (1950), 236–239.

Eötvös Loránd University, Budapest, Hungary.

INDUCED HOMOLOGY HOMOMORPHISMS FOR SET-VALUED MAPS

BARRETT O'NEILL

§ 1. **If X and Y are topological spaces,** a set-valued function F: $X \to Y$ assigns to each point x of X a closed nonempty subset $F(x)$ of Y. Let H denote Čech homology theory with coefficients in a field. If X and Y are compact metric spaces, we shall define for each such function F a vector space of homomorphisms from $H(X)$ to $H(Y)$ which deserve to be called the induced homomorphisms of F. Using this notion we prove two fixed point theorems of the Lefschetz type.

All spaces we deal with are assumed to be compact metric. Thus the group $H(X)$ can be based on a group $C(X)$ of projective chains [4]. Define the support of a coordinate c_i of $c \in C(X)$ to be the union of the closures of the kernels of the simplexes appearing in c_i. Then the intersection of the supports of the coordinates of c is defined to be the *support* $|c|$ of c.

If $F: X \to Y$ is a set-valued function, let $F^{-1}: Y \to X$ be the function such that $x \in F^{-1}(y)$ if and only if $y \in F(x)$. Then F is *upper* (*lower*) *semi-continuous* provided F^{-1} is *closed* (*open*). If both conditions hold, F is *continuous*. If $\varepsilon > 0$ is a real number, we shall also denote by $\varepsilon : X \to X$ the set-valued function such that $\varepsilon(x) = \{x' \mid d(x, x') \leqq \varepsilon\}$ for each $x \in X$.

Let A and B be chain groups with supports in X and Y respectively, and let $\varepsilon > 0$ be a number. A chain map $\varphi : A \to B$ is *accurate* with respect to a set-valued function $F: X \to Y$ provided $|\varphi(a)| \subset F(|a|)$ for each $a \in A$. Further, φ is *ε-accurate* with respect to F provided φ is accurate with respect to the composite function $\varepsilon F \varepsilon$.

(1) DEFINITION. A homomorphism $h: H(X) \to H(Y)$ is an *induced homomorphism* of a set-valued function $F: X \to Y$ provided that given $\varepsilon > 0$ there is a chain map $\varphi : C(X) \to C(Y)$ such that φ is ε-accurate with respect to F and $\varphi_* = h$.

We shall say that a homology homomorphism h is *nontrivial* provided the 0-dimensional component $h_0 : H_0(X) \to H_0(Y)$ is not the zero homomorphism. It will appear that a continuous set-valued function need not have a nontrivial induced homomorphism.

The set of all induced homomorphisms of an arbitrary set-valued function is, under the usual operations, a vector space. If h_F and h_G are induced homomorphisms of upper semi-continuous functions $F: X \to Y$ and $G: Y \to Z$, then $h_G h_F$ is an induced homomorphism of GF. If F:

Received July 9, 1956.

$X \to Y$ is a point-valued map of a connected (compact metric) space into a compact polyhedron, then the induced homology homomorphisms of F are exactly the scalar multiples of the Čech homology homomorphism F_*. Corresponding to each function $F: X \to Y$, let $F': X \to X \times Y$ be such that $F'(x) = \{x\} \times F(x)$, for $x \in X$. If $h: H(X) \to H(X \times Y)$ is a nontrivial induced homomorphism of F', then $q_* h$ is a nontrivial induced homomorphism of F, where q is the projection of the productspace on Y.

If T is a triangulation of a compact polyhedron X, let $C(X, T)$ be the group of oriented simplicial chains based on T, with the given field of coefficients. We may assume that the sequence of coverings used to define $C(X)$ consists of the star-coverings associated with the successive barycentric subdivisions of a fixed triangulation of X.

(2) LEMMA. *Let X be a compact polyhedron, $F: X \to Y$ a set-valued function. Then $h: H(X) \to H(Y)$ is an induced homomorphism of F if and only if given $\varepsilon > 0$ there is an arbitrarily fine triangulation T of X and an ε-accurate chain map $\psi: C(X, T) \to C(Y)$ such that $\psi_* = h$.*

Proof. Given $\varepsilon > 0$, let T be one of the selected triangulations of X with mesh at most $\varepsilon/2$. Denote by p the projection of $C(X)$ onto the chains of the nerve of the star-covering associated with T, followed by the natural isomorphism onto $C(X, T)$. Denote by s the chain map which assigns to $c \in C(X, T)$ the element of $C(X)$ whose coordinates correspond to the successive subdivisions $Sd^i(c)$ of c. If h is an induced homomorphism, let $\varphi: C(X) \to C(Y)$ be ε-accurate and such that $\varphi_* = h$.

Now s reduces supports, hence φs is the required chain map. Conversely, if $\psi: C(X, T) \to C(Y)$ is $\varepsilon/2$-accurate and such that $\psi_* = h$, then ψp is the required chain map, since p is $\varepsilon/2$-accurate with respect to the identity map of X.

(3) LEMMA. *Let X and Y be compact polyhedra. If, given $\varepsilon > 0$, there is an arbitrarily fine triangulation T of X and an ε-accurate ψ: $C(X, T) \to C(Y)$ such that ψ_* is nontrivial, then F has a nontrivial induced homomorphism.*

Proof. We may assume that X and Y are connected. Let L be the finite-dimensional vector space of homomorphisms from $H(X)$ to $H(Y)$. If $\varepsilon > 0$, let $A(\varepsilon)$ be the set of homomorphisms h in L such that h preserves Kronecker index and is induced by an ε-accurate chain map of $C(X, T)$, where T has mesh less than ε. By hypothesis $A(\varepsilon)$ is not empty. Furthermore one easily shows that $A(\varepsilon)$ is a coset modulo a subspace of L, and that if $\delta < \varepsilon$, then $A(\delta) \subset A(\varepsilon)$. Thus $\bigcap_{\varepsilon > 0} A(\varepsilon)$ is not empty, but an element of this intersection is, by the preceding lemma, a nontrivial induced homomorphism of F.

§ 2. **In order to establish the existence of** nontrivial induced homomorphisms in certain cases, we need some general properties of set-valued functions. Note first that if $F: X \to Y$ is continuous and K is a component of the graph $\Gamma = \{(x, y) | y \in F(x)\}$ of F, then K projects onto a component of X. In fact, the continuity of F implies that the projection $p: \Gamma \to X$ is open and closed.

(4) LEMMA. *Let X be an arcwise connected, simply connected space, $F: X \to Y$ a set-valued map. If, for each $x \in X$, $F(x)$ has exactly n components, then the graph Γ of F has exactly n components.*

Proof. Let $A = \{(x, \alpha) | x \in X$ and α is a component of $F(x)\}$. Topologize A as follows. If $(x, \alpha) \in A$, select mutually disjoint neighborhoods $V_1, \cdots, V_n$ in Y for the components $\alpha = \alpha_1, \cdots, \alpha_n$ of $F(x)$. Since F is continuous, there is a neighborhood of x such that if $x' \in U$, then $F(x') \subset V_1 \cup \cdots \cup V_n$ and $F(x')$ meets each V_i. Since$F(x')$ has n components, there is one in each V_i. Let $U_i = \{(x', \alpha') | x' \in U$ and $\alpha' = F(x') \cup V_i\}$. The collection of all such subsets of A generates a topology on A for which the projection $\pi: A \to X$ is a covering map, where $\pi(x, \alpha) = x$ if $(x, \alpha) \in A$. Consequently π is a homomorphism on each component of A—of which there are thus exactly n. If K' is a component of A, then $K = \{(x, y) | (x, \alpha) \in K'$ and $y \in \alpha\}$ is a component of Γ. In fact, since K is open and closed in Γ, it suffices to show that K is connected, but this follows from the fact that $p|K$ is strongly continuous and that $p^{-1}(x) \cap K = \alpha$ is connected for each $x \in X$.

Replacing the last step above by an application of the Vietoris mapping theorem [1] we obtain:

(5) LEMMA. *Let F be a set-valued map of a simplex σ into an arbitrary (compact metric) space. If, for each $x \in \sigma$, $F(x)$ consists of exactly n homologically trivial components, then the graph Γ of F consists of exactly n homologically trivial components.*

(6) THEOREM. *Let n be an integer, $F: X \to Y$ a set-valued map of compact polyhedra such that if $x \in X$ then $F(x)$ is either homologically trivial or consists of n homologically trivial components. Then F has a nontrivial induced homomorphism.*

Proof. Let I be the (closed) set of points for which $F(x)$ is homologically trivial. Replace F by the associated map $F': X \to X \times Y$. Since $X \times Y$ is a compact polyhedron and H is (weakly) continuous [3], if $x \in X$ and $\varepsilon > 0$ there is a δ_x such that the inclusion map $\delta(F'(x)) \subset$

$\varepsilon(F'(x))$ induces the zero homology homomorphism in positive dimensions. Using the compactness of I and the continuity of F one can show that a δ may be selected which is independent of x. Thus, if the dimension of X is $d \geqq 1$, there exist numbers $0 < \delta < \varepsilon_1 < \cdots \varepsilon_d = \varepsilon$ such that if $x \in X$ and $1 \leqq i \leqq d$ then :

(1) $\varepsilon_i + \varepsilon_1 < \varepsilon_{i+1}$,

(2) each positive-dimensional cycle of $(\varepsilon_i + \varepsilon_1)(F'(x))$ bounds in $\varepsilon_{i+1}(F'(x))$, and

(3) $F'(\delta(x)) \subset \varepsilon_1(F'(x))$. Let T be a triangulation of X with mesh at most δ, and for each simplex σ of T which meets I select a point $x_\sigma \in \sigma \cap I$. By Lemma 3 it suffices to find an ε-accurate chain map φ: $C(X, T) \to C(X \times Y)$ such that φ_* is nontrivial. We may associate with each point p of $X \times Y$ a 0-chain, $\bar{p} \in C(X \times Y)$ with support $\{p\}$ and such that $\bar{p} \not\sim 0$ and $\bar{p} \sim \bar{p}'$ if and only if p and p' are in the same component of $X \times Y$. Define a homomorphism φ_0: $C(X, T) \to C(X \times Y)$ as follows. If the vertex v is in I, let $\varphi_0(v) = n\bar{p}$, where $p \in F'(v)$. Otherwise, let $\varphi_0(v) = \bar{p}_1 + \cdots + \bar{p}_n$, where there is one point p_i in each component of $F'(v)$. Now φ_0 is extendable to a chain map φ_1 accurate with respect to F' provided that if vw is a 1-simplex of T then $\varphi_0(v) \sim \varphi_0(w)$ in $F'(vw)$. Using Lemma 4 and the preceding remark, one checks this homology in case vw does, or does not, meet I. Clearly φ_1 may be accurately extended to those chains in $C(X, T)$ whose supports avoid I. We complete the definition of φ by an induction on dimension, defining the chain map φ_q on chains of dimension at most q so that $|\varphi_q(c)| \subset \subset \varepsilon_q(F'(|c|)$. The homomorphism φ_1 is correctly defined. If φ_{q-1} has been defined ($q \geqq 1$), it suffices to define $\varphi_q(\sigma)$, where σ is an oriented q-simplex of T which meets I. If τ is an oriented $(q-1)$-face of σ, then

$$\varphi_{q-1}(\tau) \subset \varepsilon_{q-1}(F'(\tau)) \subset \varepsilon_{q-1}(F'(\partial(x_\sigma))) \subset (\varepsilon_{q-1} + \varepsilon_1)(F'(x_\sigma)) .$$

Thus $\varphi_{q-1}(\partial\sigma)$ bounds in $\varepsilon_q(F'(\sigma))$ and $\varphi_q(\sigma)$ may be correctly defined.

(7) THEOREM. *Let X be a compact 1-dimensional polyhedron with first Betti number $R_1 \leqq 1$. Then any set-valued map F of X into a compact polyhedron has a nontrivial induced homomorphism.*

Proof. Given $\varepsilon > 0$ and a triangulation T of X we must find a chain map ε-accurate with respect to F' and such that φ_* is nontrivial. If vw is a 1-simplex of T and φ is defined accurately on v, then φ can be accurately defined on w and vw so that $\partial(\varphi(vw) = \varphi(w) - \varphi(v)$, because each component of $F(v)$ is contained in a component of $F(vw)$ which meets $F(w)$. Thus if $R_1 = 0$, the required chain map exists. If $R_1 = 1$, then X may be expressed as the union of a circle C and a finite number of trees, each meeting C in at most one point. If $v^0, v^1, \cdots, v^n$ are the vertices of C, we may suppose that the 1-simplexes of C are $v^0v^1, v^1v^2, \cdots, v^nv^0$. Pick $y \in F(v^0)$ and let $\varphi^0(v^0)=\bar{y}$. There is a point $z \in F(v^1)$ such that if $\varphi^0(v^1)=\bar{z}$, then $\varphi^0(v^0v^1)$ can be correctly defined. Repeating this step we reach $\varphi^0(v^n)$, then $\varphi^1(v^0)$, and so on. Since Y is a compact polyhedron there exist integers $j<k$ such that $\varphi^k(v^0)-\varphi^j(v^0)=\partial c$, with $|c|\subset \varepsilon(F(v^0))$. Then on C let $\varphi=\sum_j^k \varphi^i$, except that

$$\varphi(v^nv^0)=\textstyle\sum_j^k \varphi^i(v^nv^0)+c\,.$$

So far φ is ε-accurate and accurate on vertices. But then as in the case $R_1=0$ this homomorphism may be extended correctly to $C(X, T)$.

We shall see in the next section that this theorem does not hold if the condition on either the dimension or the first Betti number is omitted.

§ 3. **The Lefschetz theorem holds for** set-valued functions in this form :

(8) LEMMA. *Let X be a compact polyhedron, $F: X \rightarrow X$ an upper semi-continuous set-valued function. If h is an induced homology homomorphism of F and the Lefschetz number $\Lambda(h)=\sum(-1)^q$ trace h_q is not zero, then F has a fixed point.*

The usefulness of this fact, of course, depends on our knowledge of the induced homomorphisms of a given set-valued map. From § 2 we get :

(9) THEOREM. *Let F be a set-valued self-map of a compact polyhedren X such that if $x \in X$, $F(x)$ is homologically trivial or consists of n homologically trivial components. Then F has a nontrivial homomorphism h such that if $\Lambda(h)\neq 0$, then F has a fixed point. If, further, X is homologically trivial, then F has a fixed point.*

The case $n=1$ is the polyhedral form of the Eilenberg-Montgomery theorem [2], except that the requirement that F be lower semi-continuous is then superfluous. However, if $n>1$ upper semi-continuity alone is not sufficient. For example, consider the self-map F of the Euclidean interval $[-1, 1]$ for which $F(0)=\{-1, 1\}$ and $F(x)$ is 1 for $x>0$, -1 for $x<0$. Also, if $n>1$ the space of induced homomorphisms need not be 1-dimensional as in the case $n=1$.

It does not appear that this result can be generalized by altering the number of components $F(x)$ is permitted to have. Mr. Richard Dunn has shown by a series of examples (unpublished) that if S is any finite set of positive integers—except certain sets of the form $\{2, n\}$ and, necessarily, $\{1, n\}$—there is a self-map F of the 2-cell without fixed points and such that for each point x the number of points in $F(x)$ occurs in S.

(10) THEOREM. *Let X be a compact* 1-*dimensional polyhedron with first Betti number* $R_1 \leqq 1$. *Every set-valued self-map F of X has a nontrivial induced homomorphism such that if $\Lambda(h) \neq 0$ then F has a fixed point.*

In particular, as is known, every set-valued self-map of a tree has a fixed point.

The last remark of § 2 may be justified by exhibiting suitable fixed-point-free maps. As for the restriction on the Betti number for example, let X be a compact connected 1-dimensional polyhedron without end points and such that $R_1 > 1$. If $\varepsilon > 0$ is sufficiently small, the function $F: X \rightarrow X$ for which $F(x) = \{x' \mid d(x, x') = \varepsilon\}$ will be continuous if d is a suitable metric and any induced homomorphism of F will be a scalar multiple of the identity homomorphism of $H(X)$. Thus a nontrivial induced homomorphism of such a function would have nonzero Lefschetz number, contradicting (8).

REFERENCES

1. E. G. Begle, *The Vietoris mapping theorem for bi-compact spaces*, Ann. of Math. **51** (1950), 534–543.
2. S. Eilenberg and D. Montgomery, *Fixed point theorems for multivalued transformations*, Amer. J. Math. **68** (1946), 214–222.
3. S. Eilenberg and N. Steenrod, *Foundations of algebraic topology*, Princeton, 1952.
4. S. Lefschetz, *Algebraic topology*, New York, 1942.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

A TOPOLOGICAL CHARACTERIZATION OF SETS OF REAL NUMBERS

MARY ELLEN RUDIN

We will say that a space E is of class L if E is a separable metric space which satisfies the following conditions:

(1) *Each component of E is a point or an arc (closed, open, or half-open), and no interior point of an arc-component A is a limit point of $E-A$.*

(2) *Each point of E has arbitrarily small neighborhoods whose boundaries are finite sets.*

The purpose of this note is to show that a necessary and sufficient condition that a space be homeomorphic to a set of real numbers is that it be of class L.

This gives an affirmative answer to a question raised by de Groot in [1].

In [2] L. W. Cohen proved that a separable metric space is homeomorphic to a set of real numbers if and only if it satisfies (1) above and (3) and (4) below:

(3) *E is zero-dimensional at each of its point-components.*

(4) *If p is an end point of an arc-component A, then the space $(E-A) \cup \{p\}$ is zero-dimensional at p.*

Any set of real numbers is clearly of class L. To prove the converse it is sufficient to show that every space of class L satisfies conditions (3) and (4). To this end it is clearly enough to show the following:

If X is a component of the space E of class L and ε is a positive number. there is a set $U(X, \varepsilon)$ which is both open and closed, contains X, and is contained in the union of X with the ε-neighborhoods of its endpoints (if any).

Suppose X is a component of a space E of class L and ε is a positive number. There exists an open set V which contains X but contains no point whose distance from X exceeds ε, such that the boundary B of V is finite; if X is a point, we can apply (2) directly to obtain V; if X is an arc, let V consist of X plus type (2) neighborhoods of the end points of X (if any).

Let G denote the sets of all points p of E such that E is the union of two mutually separated sets S_p and T_p, where S_p contains X and T_p contains p.

Received May 22, 1956. Research supported by the National Science Foundation.

Case I. $E-G=X$. Then G contains B. Let R be the union of all sets T_p for p in B. Since B is finite, R is both open and closed and $V-R$ is suitable for $U(X, \epsilon)$.

Case II. $E-G\neq X$. Since X is a component, $E-G$ is the union of two mutually separated sets Y and Z, where Y contains X and Z is not empty. It will be shown that there is a set K which is both open and closed and contains Z but does not intersect X, thus contradicting the fact that Z is not in G.

The definition of G, together with the fact that E has a countable base, implies that $G=\bigcup_{n=1}^{\infty} G_n$, where each G_n is both open and closed.

Let p be a point of Z. If q is a point of G, then T_q contains q and not p. The reasoning used in Case I shows that there is a neighborhood N_p of p which has no boundary point in G and whose diameter is less than half the distance from p to Y.

Let $\{H_n\}$ $(n=1, 2, 3, \cdots)$ be a countable base for E. If H_n is not a subset of N_p for any p in Z, put $K_n=0$. If, for some p in Z, H_n is a subset of N_p, let N be one such N_p and put $K_n=N-G_n$. Let $K=\bigcup_{n=1}^{\infty} K_n$. By the choice of N_p, K has no limit point in Y. No K_n has a boundary point in G and only finitely many sets K_n intersect any G_i. Consequently K has no boundary points in G and K is both open and closed. Since Z is a subset of K and X does not intersect K, the proof is complete.

References

1. J. de Groot, *On Cohen's topological characterization of sets of real numbers*, Nederl. Akad. Wetensch. Proc. Ser. A, **58** (1955), 33–35.

2. L. W. Cohen, *A characterization of those subsets of metric separable space which are homeomorphic with subsets of the linear continuum*, Fund. Math. **14** (1929), 281–303.

University of Rochester

THE FREDHOLM EIGEN VALUES OF PLANE DOMAINS

M. SCHIFFER

Introduction. The method of linear integral equations is an important tool in the theory of conformal mapping of plane simply-connected domains and in the boundary value problems of two-dimensional potential theory, in general. It yields a simple existence proof for solutions of such boundary value problems and leads to an effective construction of the required solution in terms of geometrically convergent Neumann-Liouville series. The convergence quality of these series is of considerable practical importance and has been discussed by various authors [4, 5, 6, 7]. It depends on the numerical value of the lowest nontrivial eigen value of the corresponding homogeneous integral equation which is an important functional of the boundary curve of the domain in question. Ahlfors [1, 10] gave an interesting estimate for this eigen value in terms of the extreme quasi-conformal mapping of the interior of this curve onto its exterior. Warschawski [15] gave a very useful estimate for it in terms of the corresponding eigen value of a nearby curve which allows often a good estimate of the desired eigen value in terms of a well-known one. This method is particularly valuable for special domains, for example, nearly-circular or convex ones.

It is the aim of the present paper to study the eigen functions and eigen values of the homogeneous Fredholm equation which is connected with the boundary value problem of two-dimensional potential theory. In particular, we want to obtain a sharp estimate for the lowest nontrivial eigen value in terms of function theoretic quantities connected with the curve considered. The steps of our investigation might become easier to understand by the following brief outline of our paper.

In § 1 we define the eigen values and eigen functions considered and transform the basic integral equation into a form which exhibits more clearly the interrelation with analytic function theory and extend the eigen functions as harmonic functions into the interior and the exterior of the curve. The boundary relations between these harmonic extensions are discussed and utilized to provide an example of a set of eigen functions and eigen values for the case of ellipses.

In § 2 we show the significance of the eigen value problem for the theory of the dielectric Green's function which depends on a positive parameter ε and is defined in the interior as well as the exterior of the curve. This Green's function has an immediate electrostatic interpretation and its theoretical value consists in the fact that it permits a

Received June 29, 1956. Prepared under contract Nonr-225 (11), (NR-041-086) for Office of Naval Research.

continuous transition from the Green's function of a domain to its Neumann's function. All dielectric Green's functions permit simple series developments in terms of the eigen functions and eigen values studied and the possible applications of these series developments to inequalities in potential theory are briefly indicated. Finally, it is shown that analytic completion of the dielectric Green's functions leads ultimately to univalent analytic functions in the interior as well as the exterior domain. This will lead, on the one hand, to interesting information on potential theoretical questions by use of the numerous distortion theorems of conformal mapping. On the other hand, we obtain in this way a one-parameter family of conformal maps of the domains which start with the identity and end up with the normalized mapping onto a circle. This parametrization is of importance in the theory of univalent functions; it is entirely different from the Löwner parametrization of univalent functions [8].

In §§ 3 and 4 we derive formulas for the variation of the various eigen values and dielectric Green's functions. We use at first interior variations and are thus able to derive precise variational formulas with uniform estimates for the error terms. By superposition of interior variations and simple transformations we can easily derive variational formulas of the Hadamard type. It is seen that the variational formula for the dielectric Green's function is surprisingly similar to that for the ordinary Green's function. It is seen that the circle is a curve for which all nontrivial eigen values are infinite. Thus, the circle leads to a homogeneous integral equation with an eigen value of infinite degeneracy and the usual perturbation theory cannot be applied. We show, therefore, by a special argument how eigen values for nearly-circular domains can be obtained.

Finally, we apply in § 5 the variational formula for the eigen values to a simple extremum problem for the lowest one which leads ultimately to the desired inequality. A characteristic difficulty, however, has to be overcome in this problem. It appears that the eigen values are only continuous functionals of the curve if the curve is deformed in such a way that normals in corresponding points are turned very little. Such a side condition is hard to preserve under general variations. We introduce, therefore, the concept of uniformly analytic curves which is closely related to the theory of univalent functions. Extremum problems within the class of uniformly analytic curves are easy to handle and the problem of the existence of extremum curves is likewise of very elementary nature. As the end result of our study an inequality then appears which estimates the lowest nontrivial Fredholm eigen value from below in terms of the modulus of uniform analyticity. This quantity is, however, easy to determine if a specific analytic curve is prescribed. It seems that the concept of uniform analyticity may play a useful role in many further extremum problems and variational investigations. As a side result of our study we obtain a new class of plane curves for which the Fredholm eigen functions and eigen values can be computed explicitly.

1. The Fredholm eigen values. Let C be a closed curve in the complex z-plane which is three times continuously differentiable; we denote the interior of C by D and its exterior by $\tilde{D}$. The kernel

$$k(z, \zeta)=\frac{\partial}{\partial n_\zeta}\log\frac{1}{|z-\zeta|} \tag{1}$$

is a continuous function of both argument points as these vary on the curve C only. We understand by $\dfrac{\partial}{\partial n_\zeta}$ the differentiation in direction of the normal at ζ on C pointing into D.

The first boundary value problem of potential theory with respect to the domain D can be reduced to the inhomogeneous integral equation

$$f(z)=\phi(z)+\frac{1}{\pi}\int_C k(z, \zeta)\phi(\zeta)ds_\zeta\,, \qquad z\in C\,, \tag{2}$$

while the second boundary value problem can be solved by reduction to an integral equation with transposed kernel

$$f(z)=\psi(z)-\frac{1}{\pi}\int_C k(\zeta, z)\psi(\zeta)ds_\zeta\,, \qquad z\in C\,. \tag{2'}$$

In view of the Fredholm alternative in the theory of integral equations one is then led naturally to discuss the eigen values and eigen functions of the corresponding homogeneous integral equation

$$\phi_\nu(z)=\frac{\lambda_\nu}{\pi}\int_C k(z, \zeta)\phi_\nu(\zeta)ds_\zeta\,. \tag{3}$$

These functionals of C play an important role in the potential theory of the domains D and $\tilde{D}$ as shall be seen in the following considerations. In this section we shall give a brief survey of their theory and various transformations of the integral equation (3) which will be used later.

We introduce the harmonic function

$$h_\nu(z)=\frac{\lambda_\nu}{\pi}\int_C k(z, \zeta)\phi_\nu(\zeta)ds_\zeta \tag{4}$$

which is defined in D and $\tilde{D}$; for the sake of clarity, we shall denote it by $\tilde{h}(z)$ if its argument point lies in $\tilde{D}$. By the well-known discontinuity behavior of the kernel (1), we have the limit relations, valid for an arbitrary point $z_0 \in C$:

$$(5) \qquad \lim_{z\to z_0} h_\nu(z)=(1+\lambda_\nu)\phi_\nu(z_0), \quad \lim_{z\to z_0} \tilde{h}_\nu(z)=(1-\lambda_\nu)\phi_\nu(z_0)\,.$$

On the other hand, the normal derivative of a double-layer potential goes continuously through the curve C which carries the charge and, hence,

$$(6) \qquad \frac{\partial}{\partial n} h_\nu(z)=-\frac{\partial}{\partial \tilde{n}}\tilde{h}_\nu(z)\,, \qquad \text{for } z\in C\,,$$

where $\frac{\partial}{\partial \tilde{n}}$ denotes normal differentiation into $\tilde{D}$.

By Green's identity, we have

$$(7) \qquad \frac{1}{2\pi}\int_C\left[h_\nu(\zeta)\frac{\partial}{\partial n_\zeta}\log\frac{1}{|\zeta-z|}-\log\frac{1}{|\zeta-z|}\frac{\partial}{\partial n}h_\nu(\zeta)\right]ds_\zeta=h_\nu(z)\delta(z)$$

and

$$(7') \qquad \frac{1}{2\pi}\int_C\left[\tilde{h}_\nu(\zeta)\frac{\partial}{\partial \tilde{n}_\zeta}\log\frac{1}{|\zeta-z|}-\log\frac{1}{|\zeta-z|}\frac{\partial}{\partial \tilde{n}}\tilde{h}_\nu(\zeta)\right]ds_\zeta=\tilde{h}_\nu(z)\tilde{\delta}(z)$$

where $\delta(z)$ and $\tilde{\delta}(z)$ are the characteristic functions of D and $\tilde{D}$, that is,

$$(8) \qquad \delta(z)=\begin{cases}1 & \text{if } z\in D\\ 0 & \text{if } z\in\tilde{D}\end{cases}\,, \qquad \tilde{\delta}(z)=1-\delta(z)\,.$$

Combining (7) with (7′) and observing the boundary relations (5) and (6) between $h_\nu(z)$ and $\tilde{h}_\nu(z)$, we obtain

$$(9) \qquad -\frac{\lambda_\nu}{\pi(\lambda_\nu-1)}\int_C\log\frac{1}{|\zeta-z|}\frac{\partial}{\partial n}h_\nu(\zeta)ds_\zeta=h_\nu(z)\,, \qquad z\in D$$

$$(10) \qquad -\frac{\lambda_\nu}{\pi(\lambda_\nu+1)}\int_C\log\frac{1}{|\zeta-z|}\frac{\partial}{\partial \tilde{n}}\tilde{h}_\nu(\zeta)ds_\zeta=\tilde{h}_\nu(z)\,, \qquad z\in\tilde{D}\,.$$

Define two analytic functions in D and $\tilde{D}$, respectively, by the formulas

$$(11) \qquad v_\nu(z)=\frac{\partial}{\partial z}h_\nu(z)\,, \quad \tilde{v}_\nu(z)=\frac{\partial}{\partial z}\tilde{h}_\nu(z)\,.$$

Differentiating (9) and (10) with respect to z, one obtains easily

$$v_\nu(z)=\frac{\lambda_\nu}{2\pi i}\int_C \frac{\overline{(v_\nu(\zeta)d\zeta)}}{\zeta-z}\,,\quad \tilde{v}_\nu(z)=\frac{\lambda_\nu}{2\pi i}\int_C \frac{\overline{(\tilde{v}_\nu(\zeta)d\zeta)}}{\zeta-z}\,. \tag{12}$$

These are elegant integral equations for the analytic functions $v_\nu(z)$ and $\tilde{v}_\nu(z)$ which are valid for $z \in D$ and $z \in \tilde{D}$, respectively.

We can also bring (12) into the form

$$v_\nu(z)=\frac{\lambda_\nu}{\pi}\iint_D \frac{\overline{v_\nu(\zeta)}}{(\zeta-z)^2}d\tau \qquad \text{for } z\in D \tag{13}$$

and

$$\tilde{v}_\nu(z)=\frac{\lambda_\nu}{\pi}\iint_{\tilde{D}} \frac{\overline{\tilde{v}_\nu(\zeta)}}{(\zeta-z)^2}d\tau \qquad \text{for } z\in \tilde{D} \tag{13'}$$

which expresses $v_\nu(z)$ and $\tilde{v}_\nu(z)$ as solutions of integral equations with improper kernels. The integrals involved have to be understood in the Cauchy principal value sense.

In the transition from the integral equations (3) defined on C to the integral equations (13), (13′) defined in D and $\tilde{D}$, we have lost one particular eigen function. Indeed, if $h(z)=\text{const.}$ were one of our eigen functions $h_\nu(z)$ it would have been cancelled out in the differentiation (11). But by (5), $h_\nu(z)=\text{const.}$ implies $\phi_\nu(z)=\text{const.}$ on C and from (3) and the identity

$$\frac{1}{\pi}\int_C k(z,\ \zeta)ds_\zeta=1\,, \qquad z\in C\,, \tag{14}$$

follows, in fact, that each constant is an eigen function of (3) with the eigen value $\lambda=+1$. The eigen value $\lambda=+1$ plays an exceptional role in the entire Poincaré-Fredholm theory of the boundary value problem; the fact that the equations (13), (13′) lead to all other eigen values and their corresponding eigen functions and eliminate $\lambda=+1$ represents, therefore, a strong argument in favor of this transition.

Let $z(s)$ be the parametric representation of C in terms of its length parameter s. Then

$$z'=\frac{dz}{ds} \tag{15}$$

will be the unit vector in tangential direction to C. The boundary relations (5) and (6) for h_ν and $\tilde{h}_\nu$ go over into the equations on C:

$$\Re\{v_\nu(z)z'\}=\frac{1+\lambda_\nu}{1-\lambda_\nu}\Re\{\tilde{v}_\nu(z)z'\} \tag{16}$$

$$\Im\{v_\nu(z)z'\} = \Im\{\tilde{v}_\nu(z)z'\} \tag{17}$$

which can be combined into the one complex equation

$$v_\nu(z)z' = \frac{1}{1-\lambda_\nu}\tilde{v}_\nu(z)z' + \frac{\lambda_\nu}{1-\lambda_\nu}\overline{\tilde{v}_\nu(z)z'}\,. \tag{18}$$

This relation combined with (12) throws an interesting light on the connection between $v_\nu(z)$ and $\tilde{v}_\nu(z)$. In fact, if we insert (18) into the first equation (12) and apply Cauchy's theorem, we find

$$v_\nu(z) = \frac{1}{2\pi i}\frac{\lambda_\nu}{1-\lambda_\nu}\int_C \frac{\overline{(\tilde{v}_\nu(\zeta)d\zeta)}}{\zeta - z}\,, \qquad z \in D\,. \tag{19}$$

Observe that the second formula (12) yields $\tilde{v}_\nu(z)$ for $z \in \tilde{D}$; now we see that the same expression yields $v_\nu(z)$ for $z \in D$, except for the factor $1-\lambda_\nu$. Similarly, one shows easily

$$\tilde{v}_\nu(z) = \frac{1}{2\pi i}\frac{\lambda_\nu}{1+\lambda_\nu}\int_C \frac{\overline{(v_\nu(\zeta)d\zeta)}}{\zeta - z}\,, \qquad z \in \tilde{D}\,. \tag{20}$$

If $f(z)$ is an arbitrary complex-valued function in the entire z-plane of the class $\mathscr{L}^2$, the equation

$$F(z) = \frac{1}{\pi}\iint \frac{\overline{f(\zeta)}}{(\zeta - z)^2}d\tau\,, \tag{21}$$

defines a new function in $\mathscr{L}^2$, its Hilbert transform. It is well-known [2] that the Hilbert transformation is norm-preserving, that is,

$$\iint |F|^2\,d\tau = \iint |f|^2\,d\tau\,. \tag{22}$$

Our formulas (13), (13′) and (19), (20) imply that the functions

$$f_\nu(z) = \begin{cases} v_\nu(z) & \text{in } D \\ 0 & \text{in } \tilde{D} \end{cases} \quad \text{and} \quad \tilde{f}_\nu(z) = \begin{cases} 0 & \text{in } D \\ \tilde{v}_\nu(z) & \text{in } \tilde{D} \end{cases} \tag{23}$$

have the Hilbert transforms

$$F_\nu(z) = \begin{cases} \dfrac{1}{\lambda_\nu}v_\nu(z) & \text{in } D \\ \left(\dfrac{1}{\lambda_\nu}+1\right)\tilde{v}_\nu(z) & \text{in } \tilde{D} \end{cases} \quad \text{and} \quad \tilde{F}_\nu(z) = \begin{cases} \left(\dfrac{1}{\lambda_\nu}-1\right)v_\nu(z) & \text{in } D \\ \dfrac{1}{\lambda_\nu}\tilde{v}_\nu(z) & \text{in } \tilde{D}\,. \end{cases} \tag{24}$$

Hence (22) yields

$$(25) \qquad (\lambda_\nu - 1)\iint_D |v_\nu|^2 \, d\tau = (\lambda_\nu + 1)\iint_{\tilde{D}} |\tilde{v}_\nu|^2 \, d\tau .$$

From (25) we conclude easily that

$$(26) \qquad |\lambda_\nu| > 1 .$$

For if, for example, $\lambda_\nu = 1$, we would have $\tilde{v}_\nu \equiv 0$ in $\tilde{D}$, $\tilde{h}_\nu(z) = \text{const.}$ and hence by (6) also $h_\nu(z) = \text{const.}$ But this would imply, in turn, $v_\nu(z) \equiv 0$ and no eigen functions would exist.

With each eigen value λ_ν of (4) the eigen value $-\lambda_\nu$ also occurs, except for $\lambda = 1$. In fact, if we denote the conjugate functions of $h_\nu(z)$ and $\tilde{h}_\nu(z)$ by $g_\nu(z)$ and $\tilde{g}_\nu(z)$, we have by the Cauchy-Riemann formulas the relations

$$(27) \qquad \frac{\partial}{\partial n} g_\nu(z) = -\frac{1+\lambda_\nu}{1-\lambda_\nu} \frac{\partial}{\partial \tilde{n}} \tilde{g}_\nu(z) , \quad \frac{\partial}{\partial s} g_\nu(z) = \frac{\partial}{\partial s} \tilde{g}_\nu(z) .$$

Hence, putting

$$(28) \qquad g_\nu^*(z) = (1-\lambda_\nu) g_\nu(z) , \quad \tilde{g}_\nu^*(z) = (1+\lambda_\nu) \tilde{g}_\nu(z)$$

and adding an appropriate constant we find for $z \in C$:

$$(29) \qquad g_\nu^*(z) = \frac{1-\lambda_\nu}{1+\lambda_\nu} \tilde{g}_\nu^*(z) , \quad \frac{\partial}{\partial n} g_\nu^*(z) = -\frac{\partial}{\partial \tilde{n}} \tilde{g}_\nu^*(z) .$$

These are the boundary relations between h_ν and $\tilde{h}_\nu$ but with $-\lambda_\nu$ instead of λ_ν. This proves our assertion.

If we start conversely from the complex integral equations (13) and (13′) and consider any eigen function $v_\nu(z)$ with the eigen value λ_ν, it will be observed that $e^{i\alpha} v_\nu(z)$ is also an eigen function to the eigen value $\lambda_\nu e^{-2i\alpha}$. Hence, if we focus our attention on the integral equations (12) or (13) we may assume without loss of generality that λ_ν is a real positive eigen value. Calculating backward, we can easily see that each such eigen value is also an eigen value of the Fredholm integral equation (4) and so is $-\lambda_\nu$.

It is readily verified that eigen functions $v_\nu(z)$ and $v_\mu(z)$ which belong to different eigen values λ_ν and λ_μ satisfy the orthogonality relation

$$(30) \qquad \iint_D v_\nu(z) \overline{v_\mu(z)} \, d\tau = 0 .$$

This condition can be extended to the case of any two linearly independent eigen functions. Similarly

$$(31) \qquad \iint_{\tilde{D}} \tilde{v}_\nu \overline{\tilde{v}}_\mu \, d\tau = 0$$

for any two different eigen functions $\tilde{v}_\nu$, $\tilde{v}_\mu$. In view of (25), we define

$$w_\nu(z)=\sqrt{\lambda_\nu-1}\ v_\nu(z)\,, \quad \tilde{w}_\nu(z)=i\sqrt{\lambda_\nu+1}\ \tilde{v}_\nu(z)\,. \tag{32}$$

Then we can assume the orthonormality relations

$$\iint_D w_\nu \bar{w}_\mu d\tau=\delta_{\nu\mu}\,, \quad \iint_{\tilde{D}} \tilde{w}_\nu \bar{\tilde{w}}_\mu d\tau=\delta_{\nu\mu}\,. \tag{33}$$

We have in view of (18) the boundary relations on C:

$$w_\nu(z)z'=\frac{i}{\sqrt{\lambda_\nu^2-1}}\tilde{w}_\nu(z)z'-\frac{\lambda_\nu i}{\sqrt{\lambda_\nu^2-1}}\overline{\tilde{w}_\nu(z)z'}\,; \tag{34}$$

from (19) and (20) follows

$$w_\nu(z)=-\frac{\lambda_\nu}{2\pi\sqrt{\lambda_\nu^2-1}}\int_C\frac{\overline{(\tilde{w}_\nu(\zeta)d\zeta)}}{\zeta-z}\,, \qquad z\in D \tag{35}$$

and

$$\tilde{w}_\nu(z)=\frac{\lambda_\nu}{2\pi\sqrt{\lambda_\nu^2-1}}\int_C\frac{\overline{(w_\nu(\zeta)d\zeta)}}{\zeta-z}\,, \qquad z\in \tilde{D}\,. \tag{36}$$

If we were able to guess two functions $w(z)$ and $\tilde{w}(z)$ which are analytic in D and $\tilde{D}$, respectively, and which satisfy on C the relation (34) for a properly chosen λ, we would have obtained a particular solution for the eigen value problems (13) and (13′). It is sometimes possible to construct such pairs of functions and to obtain thus eigen values and eigen functions for the Fredholm integral equation. One possibility of construction is the following: We refer the curve C by conformal mapping to the unit circumference. Let

$$z=f(\zeta) \tag{35}$$

be analytic on and near $|\zeta|=1$ and map it onto C. The condition (34) can now be referred to $|\zeta|=1$ and reads:

$$w[f(\zeta)]f'(\zeta)i\zeta=-\frac{1}{\sqrt{\lambda^2-1}}\tilde{w}[f(\zeta)]f'(\zeta)\zeta-\frac{\lambda}{\sqrt{\lambda^2-1}}\overline{\tilde{w}[f(\zeta)]f'(\zeta)\zeta}\,. \tag{38}$$

Since the conjugation $\bar{\zeta}$ means just $\dfrac{1}{\zeta}$ on $|\zeta|=1$ it is easier to guess solutions in this form.

Let, for example,

$$z=f(\zeta)=\zeta+\frac{\rho^2}{\zeta}\,, \qquad 0<\rho<1\,. \tag{39}$$

This means that C is an ellipse. Let us put

$$w[f(\zeta)]f'(\zeta)\zeta=a\zeta^n+b\zeta^{-n} \tag{40}$$

and

$$w[f(\zeta)]f'(\zeta)\zeta=i\zeta^{-n}. \tag{41}$$

Condition (38) will be fulfilled if we put

$$a=\frac{\lambda}{\sqrt{\lambda^2-1}}, \quad b=\frac{-1}{\sqrt{\lambda^2-1}}. \tag{42}$$

Define $W(f)=\int w\,df$ and $\tilde{W}(f)=\int \tilde{w}\,df$. Then (40) and (41) yield

$$W[f(\zeta)]=\frac{1}{n}(a\zeta^n-b\zeta^{-n})=\frac{\lambda}{n\sqrt{\lambda^2-1}}(\zeta^n+\lambda^{-1}\zeta^{-n}), \tag{43}$$

$$\tilde{W}[f(\zeta)]=-\frac{i}{n}\zeta^{-n}. \tag{44}$$

Now the function (39) is univalent outside of $|\zeta|=\rho$ and, hence, we may consider ζ as a regular analytic function of $z\in\tilde{D}$. Thus, $\tilde{W}(z)$ and $\tilde{w}(z)$ are regular analytic in $\tilde{D}$. In order that $W(z)$, $w(z)$ be analytic in D we must require that

$$\lambda=\rho^{-2n}. \tag{45}$$

In fact, $\zeta^n+\dfrac{\rho^{2n}}{\zeta^n}$ can be expressed as a Chebysheff polynomial of z. Thus we have guessed an infinity of eigen values and eigen functions for the case of the ellipse. It can be shown that $\lambda^n=\pm\rho^{-2n}$ gives all eigen values of the ellipse for $n=1, 2, \cdots$. Since $\rho=0$ describes a circle, we recognize, in particular, that all eigen values λ_ν for a circle have the the value infinity [3].

If we know the eigen values and eigen functions of a given domain D we can find immediately the eigen values and eigen functions of every domain D^* which is obtained from D by a linear transformation

$$z^*=\frac{az+b}{cz+d}=l(z). \tag{46}$$

In fact, let

$$w_\nu^*(z^*)l'(z)=w_\nu(z), \quad \tilde{w}_\nu^*(z^*)l'(z)=\tilde{w}_\nu(z), \quad \lambda_\nu^*=\lambda_\nu. \tag{47}$$

It follows from (33) that the w_ν^* and $\tilde{w}_\nu^*$ form an orthonormal set of analytic functions in D^* and $\tilde{D}^*$, respectively. Since we have on C^*

$$z^{*\prime}(z)=l'(z) \tag{48}$$

it is also obvious that (34) is fulfilled which shows that w_ν^* and $\tilde{w}_\nu^*$ are, in fact, the normalized eigen functions of the domains D^* and $\tilde{D}^*$. In particular, we note that the eigen values λ_ν of a domain are unchanged under linear transformation. Similar domains, for example, have the same set of eigen values.

2. Thc dielectric problem. The consideration of the electrostatic field of a point source at ζ in the presence of a dielectric medium in D with the dielectric constant ε leads to the following heuristic definition of a Green's function $G_\varepsilon(z, \zeta)$:

(a) $G_\varepsilon(z, \zeta)$ is harmonic in D and $\tilde{D}$, except for $z=\zeta$.

(b) $G_\varepsilon(z, \zeta)-\log \dfrac{1}{|z-\zeta|}$ is harmonic at ζ if $\zeta\in\tilde{D}$.

(b′) $G_\varepsilon(z, \zeta)-\varepsilon \log \dfrac{1}{|z-\zeta|}$ is harmonic at ζ if $\zeta\in D$.

(c) $G_\varepsilon(z, \zeta)$ is continuous through C.

(d) $\dfrac{\partial}{\partial n}G_\varepsilon(z, \zeta)+\varepsilon\dfrac{\partial}{\partial\tilde{n}}G_\varepsilon(z, \zeta)=0$ on C for $\zeta\in D$ or $\tilde{D}$.

(e) $\log|z|+G_\varepsilon(z, \zeta)\to 0$ if $z\to\infty$, for $\zeta\in D$ or $\zeta\in\tilde{D}$.

It is easily seen that $G_\varepsilon(z, \zeta)$ is uniquely determined by these conditions and that it satisfies the symmetry condition

$$G_\varepsilon(\zeta, \eta)=G_\varepsilon(\eta, \zeta)\,. \tag{1}$$

We may construct $G_\varepsilon(z, \zeta)$ by means of a line potential as follows. Let $\zeta\in\tilde{D}$ and put

$$G_\varepsilon(z, \zeta)=\log\frac{1}{|z-\zeta|}+\int_C \mu(\eta, \zeta)\log|\eta-z|ds_\eta\,. \tag{2}$$

This set-up satisfies automatically conditions (a), (b) and (c); we can fulfill condition (e) by the requirement

$$\int_C \mu(\eta, \zeta)ds_\eta=0 \tag{3}$$

and finally (d) by solving the integral equation

$$-\frac{1-\varepsilon}{1+\varepsilon}\cdot\frac{1}{\pi}\frac{\partial}{\partial n_z}\log\frac{1}{|z-\zeta|}=\mu(z, \zeta)-\frac{1-\varepsilon}{1+\varepsilon}\frac{1}{\pi}\int_C \mu(\eta, \zeta)\frac{\partial}{\partial n_z}\log\frac{1}{|\eta-z|}ds_\eta\,. \tag{4}$$

As long as $\varepsilon > 0$ this equation can be solved in a unique way since $\left|\frac{1-\varepsilon}{1+\varepsilon}\right| < 1$ and all eigen values of the corresponding homogeneous integral equation are larger or equal to one in absolute value. One verifies also from (4) that condition (3) is automatically fulfilled. In a similar way we proceed for $\zeta \in D$.

The integral equation (4) indicates already the close relation between the Green's function $G_\varepsilon(z, \zeta)$ and the Fredholm eigen functions. We obtain a further insight from the Dirichlet identities:

$$(5) \qquad \iint_D \nabla G_\varepsilon(z, \zeta)\nabla h_\nu(z)d\tau = -\int_C G_\varepsilon(z, \zeta)\frac{\partial}{\partial n}h_\nu(z)ds_z$$

$$= 2\pi\varepsilon h_\nu(\zeta)\delta(\zeta) - \int_C h_\nu(z)\frac{\partial}{\partial n}G_\varepsilon(z, \zeta)ds_z$$

and

$$(6) \qquad \iint_{\tilde{D}} \nabla G_\varepsilon(z, \zeta)\cdot\nabla \tilde{h}_\nu(z)\,d\tau = -\int_C G_\varepsilon(z, \zeta)\frac{\partial}{\partial \tilde{n}}\tilde{h}_\nu(z)ds_z$$

$$= 2\pi\tilde{h}_\nu(\zeta)\tilde{\delta}(\zeta) - \int_C \tilde{h}_\nu(z)\frac{\partial}{\partial \tilde{n}}G_\varepsilon(z, \zeta)ds_z\,.$$

Here we use $\delta(\zeta)$ and $\tilde{\delta}(\zeta)$ as defined in (1.8). Identity (6) is valid in spite of the logarithmic pole of G_ε at infinity since $\tilde{h}_\nu(z)$ vanishes there.

Adding (5) and (6) and using (1.6), we obtain

$$(7) \qquad \iint_{\tilde{D}} \nabla G_\varepsilon(z, \zeta)\nabla\tilde{h}_\nu(z)d\tau = -\iint_D \nabla G_\varepsilon(z, \zeta)\nabla h_\nu(z)\,d\tau\,.$$

Putting

$$(8) \qquad \rho_\nu = \frac{\lambda_\nu + 1}{\lambda_\nu - 1}$$

and using the boundary relations (1.5) and (d), we find:

$$(9) \qquad -\varepsilon\rho_\nu \iint_{\tilde{D}} \nabla G_\varepsilon(z, \zeta)\nabla\tilde{h}_\nu d\tau + \iint_D \nabla G_\varepsilon(z, \zeta)\nabla h_\nu d\tau$$

$$= 2\pi\varepsilon[h_\nu(\zeta)\delta(\zeta) - \rho_\nu\tilde{h}_\nu(\zeta)\tilde{\delta}(\zeta)]\,.$$

Thus, finally,

$$(10) \qquad \iint_D \nabla G_\varepsilon(z, \zeta)\nabla h_\nu(z)d\tau = \frac{2\pi\varepsilon}{1+\varepsilon\rho_\nu}[h_\nu(\zeta)\delta(\zeta) - \rho_\nu\tilde{h}_\nu(\zeta)\tilde{\delta}(\zeta)]$$

$$= -\iint_{\tilde{D}} \nabla G_\varepsilon(z, \zeta)\nabla\tilde{h}_\nu d\tau\,.$$

The eigen functions $h_\nu(z)$ connected with the Fredholm equation appear thus as the eigen functions of the integral equation in D:

$$\frac{1+\varepsilon\rho_\nu}{2\pi\varepsilon}\iint_D \nabla G_\varepsilon(z,\ \zeta)\cdot\nabla h_\nu(z)d\tau = h_\nu(\zeta)\ ,\quad \zeta\in D\ . \tag{10'}$$

Let $G(z,\ \zeta)$ be the ordinary Green's function of D; obviously

$$\iint_D \nabla G(z,\ \zeta)\cdot\nabla h_\nu(z)d\tau = 0\ . \tag{11}$$

Hence we obtain for $h_\nu(z)$ the integral equation

$$\frac{1+\varepsilon\rho_\nu}{2\pi\varepsilon}\iint_D \nabla K_\varepsilon(z,\ \zeta)\cdot\nabla h_\nu(z)d\tau = h_\nu(\zeta) \tag{12}$$

with the regular harmonic kernel

$$K_\varepsilon(z,\ \zeta) = G_\varepsilon(z,\ \zeta) - \varepsilon\, G(z,\ \zeta)\ . \tag{13}$$

Let

$$\tilde{G}(z,\ \infty) = \log|z| + \tilde{\gamma} + O\left(\frac{1}{|z|}\right) \tag{14}$$

represent the Green's function of $\tilde{D}$ with the source point at infinity. By (1.5) we have obviously

$$\int_C h_\nu(z)\frac{\partial\tilde{G}(z,\ \infty)}{\partial\tilde{n}}ds = \frac{1+\lambda_\nu}{1-\lambda_\nu}\int_C \tilde{h}_\nu(z)\,\frac{\partial\tilde{G}(z,\ \infty)}{\partial\tilde{n}} = 2\pi\frac{1+\lambda_\nu}{1-\lambda_\nu}\tilde{h}_\nu(\infty) = 0\ . \tag{15}$$

We now define the linear space Σ consisting of all functions $h(z)$ which are harmonic in D, have a finite Dirichlet integral there and satisfy the linear homogeneous condition

$$\int_C h(z)\,\frac{\partial\tilde{G}(z,\ \infty)}{\partial\tilde{n}}ds = 0\ . \tag{16}$$

Observe that the only constant element in Σ is the function $h\equiv 0$. All $h_\nu(z)$ lie in Σ; in view of (12) and the symmetry of $K_\varepsilon(z,\ \zeta)$ we may assume that they are orthonormalized by the conditions

$$\iint_D \nabla h_\nu\cdot\nabla h_\mu d\tau = \delta_{\nu\mu} \tag{17}$$

and it is easily seen that they form a complete orthonormal set in Σ [3].

If we use the conditions (c) and (e) in the definition of $G_\varepsilon(z,\ \zeta)$, we can show that the function

(18) $$h(z)=K_\varepsilon(z,\ \zeta)-\tilde{\gamma}$$

lies in Σ. Hence we have for it the following series development

(19) $$G_\varepsilon(z,\ \zeta)-\varepsilon G(z,\ \zeta)=\tilde{\gamma}+\sum_{\nu=1}^{\infty}\frac{h_\nu(z)h_\nu(\zeta)}{1+\varepsilon\rho_\nu}\cdot 2\pi\varepsilon\ , \qquad \zeta\in D\ .$$

The Fourier coefficients in this development have been calculated from (12); the series converges uniformly in each closed subdomain of D.

Suppose next $\zeta\in\tilde{D}$ and consider the harmonic function

(20) $$h(z)=G_\varepsilon(z,\ \zeta)+\tilde{G}(\zeta,\ \infty)-\tilde{\gamma}\ .$$

It is easily seen that $h(z)\in\Sigma$. Hence we may develop $h(z)$ into a series in the complete orthonormal system $h_\nu(z)$. Using (10), we find

(21) $$G_\varepsilon(z,\ \zeta)=\tilde{\gamma}-\tilde{G}(\zeta,\ \infty)-2\pi\varepsilon\sum_{\nu=1}^{\infty}\frac{\rho_\nu}{1+\varepsilon\rho_\nu}h_\nu(z)\tilde{h}_\nu(\zeta)\ .$$

This series converges for $\zeta\in\tilde{D}$ and z in a closed subdomain of D.

Observe that by definition of $G_\varepsilon(z,\ \zeta)$ we have for $\varepsilon=1$

(22) $$G_1(z,\ \zeta)=\log\frac{1}{|z-\zeta|}\ .$$

Hence (19) contains the following series representation for the ordinary Green's function of D:

(23) $$G(z,\ \zeta)=\log\frac{1}{|z-\zeta|}-\tilde{\gamma}-2\pi\sum_{\nu=1}^{\infty}\frac{h_\nu(z)h_\nu(\zeta)}{1+\rho_\nu}\ .$$

On the other hand, (21) reduces for $\varepsilon=1$ to

(24) $$\tilde{G}(\zeta,\ \infty)=\log|z-\zeta|+\tilde{\gamma}-2\pi\sum_{\nu=1}^{\infty}\frac{\rho_\nu}{1+\rho_\nu}h_\nu(z)\tilde{h}_\nu(\zeta)\ .$$

In a similar way we can derive series developments for $G_\varepsilon(z,\ \zeta)$ in the exterior $\tilde{D}$ of C. Observe that in view of the boundary conditions (1.5) and (1.6) the normalization (17) of the $h_\nu(z)$ implies

(25) $$\iint_{\tilde{D}}\nabla\tilde{h}_\nu\cdot\nabla\tilde{h}_\mu d\tau=\rho_\nu^{-1}\delta_{\mu\nu}\ .$$

Let $\tilde{\Sigma}$ be the linear function space consisting of all functions $\tilde{h}(z)$ which are harmonic in $\tilde{D}$, have a finite Dirichlet integral there and which vanish at infinity. Clearly the $\{\rho_\nu^{1/2}\tilde{h}_\nu(z)\}$ form a complete orthonormal set in $\tilde{\Sigma}$.

On the other hand, let $\tilde{G}(z, \zeta)$ be the Green's function of $\tilde{D}$. Then it is easily verified that by condition (e) and (14)

$$\tilde{h}(z)=G_\varepsilon(z, \zeta)-\tilde{G}(z, \zeta)+\tilde{G}(\zeta, \infty)+\tilde{G}(z, \infty)-\tilde{\gamma}, \qquad z, \zeta \in \tilde{D}, \tag{26}$$

lies in $\tilde{\Sigma}$. Again using the Dirichlet formula (10), we find

$$\begin{aligned} G_\varepsilon(z, \zeta)-\tilde{G}(z, \zeta)+\tilde{G}(\zeta, \infty)+\tilde{G}(z, \infty)-\tilde{\gamma} \\ =2\pi\varepsilon \sum_{\nu=1}^{\infty} \frac{\tilde{h}_\nu(z)\tilde{h}_\nu(\zeta)\rho_\nu^2}{1+\varepsilon\rho_\nu}. \end{aligned} \tag{27}$$

Putting, in particular, $\varepsilon=1$, we obtain by virtue of (22):

$$\tilde{G}(z, \zeta)-\tilde{G}(\zeta, \infty)-\tilde{G}(z, \infty)=\log\frac{1}{|z-\zeta|}-\tilde{\gamma}-2\pi\sum_{\nu=1}^{\infty}\frac{\tilde{h}_\nu(z)\tilde{h}_\nu(\zeta)\rho_\nu}{1+\rho_\nu}. \tag{28}$$

We have thus shown that all dielectric Green's functions can be constructed simultaneously and in D as well as in $\tilde{D}$, once the system of eigen functions $h_\nu(z)$ and the corresponding eigen values λ_ν are known.

Numerous inequalities can be drawn from these representations. We shall restrict ourselves to one single example. Denote, for $\zeta \in D$,

$$G_\varepsilon(z, \zeta)-\varepsilon\log\frac{1}{|z-\zeta|}=g_\varepsilon(z, \zeta), \quad G(z, \zeta)-\log\frac{1}{|z-\zeta|}=g(z, \zeta). \tag{29}$$

The functions $g_\varepsilon(z, \zeta)$ and $g(z, \zeta)$ represent the potentials induced by a unit pole at ζ in the presence of the dielectric in D, and in the presence of the grounded conductor C, respectively. We find from (19)

$$g_\varepsilon(\zeta, \zeta)-\varepsilon g(\zeta, \zeta) \geqq \tilde{\gamma}. \tag{30}$$

Since $e^{-\tilde{\gamma}}$ represents the electrostatic capacity of the conductor C, we obtain an interesting estimate for the dielectric reaction potential in terms of capacity constants connected with the conductor surface C. For $\varepsilon=1$, we have $g_1(\zeta, \zeta)=0$ and hence

$$\tilde{\gamma} \leqq -g(\zeta, \zeta). \tag{31}$$

This is an inequality connecting the inner and the outer Green's function of C; in the case that C is a circumference and ζ is its center, this inequality becomes an equality.

Up to this point we stressed the connection between the Green's function $G(z, \zeta)$ and the eigen functions $h_\nu(z)$. Since the Fredholm eigen functions appear also in the theory of the second boundary value problem, we should also expect some relations between the $h_\nu(z)$ and the Neumann's function of the domain D.

The Neumann's function is usually defined by its constant normal derivative on C

$$\frac{\partial N(z,\ \zeta)}{\partial n_z}=\frac{2\pi}{L}, \qquad z\in C,\ \zeta\in D,\ L=\text{length of } C, \tag{32}$$

and by the linear homogeneous side condition

$$\int_C N(z,\ \zeta)ds_z=0\,, \qquad \zeta\in D\,. \tag{33}$$

In order to operate within the class Σ, characterized by (16), we introduce the functions

$$\alpha(z)=\frac{1}{2\pi}\int_C N(t,\ z)\left[\frac{\partial\tilde{G}(t,\ \infty)}{\partial n}+\frac{2\pi}{L}\right]ds_t \tag{34}$$

and

$$\begin{aligned}h(z)=&N(z,\ \zeta)-G(z,\ \zeta)+\alpha(z)+\alpha(\zeta)\\&+\frac{1}{4\pi^2}\iint_{CC}N(t,\ \tau)\frac{\partial\tilde{G}(t,\ \infty)}{\partial n}\frac{\partial\tilde{G}(\tau,\ \infty)}{\partial n}ds_t\,ds_\tau.\end{aligned} \tag{35}$$

It is easily verified that $h(z)\in\Sigma$. Since obviously

$$\int_C\left[\frac{\partial\tilde{G}(t,\ \infty)}{\partial n}+\frac{2\pi}{L}\right]ds=0\,, \tag{34'}$$

the function $\alpha(z)$ is harmonic in D and has the normal derivative

$$\frac{\partial\alpha}{\partial n}=-\frac{\partial\tilde{G}(z,\ \infty)}{\partial n}-\frac{2\pi}{L}\,, \qquad z\in C\,. \tag{34''}$$

Hence, finally, we have for $z\in C$

$$\frac{\partial h}{\partial n}=-\frac{\partial G(z,\ \zeta)}{\partial n}-\frac{\partial\tilde{G}(z,\ \infty)}{\partial n} \tag{35'}$$

and consequently in view of (16) valid for each $h_\nu(z)$:

$$\iint_D \nabla h\cdot\nabla h_\nu d\tau=-\int_C h_\nu\frac{\partial h}{\partial n}ds=2\pi h_\nu(\zeta)\,. \tag{36}$$

Since $h(z)\in\Sigma$ and the $h_\nu(z)$ are a complete orthonormal system in Σ, we have the Fourier development

$$(37)\quad N(z,\ \zeta)-G(z,\ \zeta)+\frac{1}{2\pi}\int_C N(t,\ z)\frac{\partial\tilde{G}(t,\ \infty)}{\partial n}ds+\frac{1}{2\pi}\int_C N(t,\ \zeta)\frac{\partial\tilde{G}(t,\ \infty)}{\partial n}ds$$

$$+\frac{1}{4\pi^2}\iint_{CC} N(t,\ \tau)\frac{\partial\tilde{G}(t,\ \infty)}{\partial n}\frac{\partial\tilde{G}(\tau,\ \infty)}{\partial n}ds_t ds_\tau=\sum_{\nu=1}^{\infty} h_\nu(z)h_\nu(\zeta)\cdot 2\pi\,.$$

This formula is useful to establish the exact asymptotics of the function $G_\varepsilon(z,\ \zeta)$ as $\varepsilon\rightarrow 0$ as can be seen from formula (19).

The dielectric Green's functions $G_\varepsilon(z,\ \zeta)$ are closely related to a set of interesting univalent analytic functions. In order to show this connection we complete the harmonic functions $G_\varepsilon(z,\ \zeta)$ to analytic functions in z. We will obtain, of course, two entirely different functions when z lies in D or $\tilde{D}$. Let us denote the analytic completion of $G_\varepsilon(z,\ \zeta)$ by $P_\varepsilon(z,\ \zeta)$ if $z\in D$ and by $\tilde{P}_\varepsilon(z,\ \zeta)$ if $z\in\tilde{D}$. We want to show that for fixed $\zeta\in D$

$$(38)\qquad f_\varepsilon(z)=e^{-1/\varepsilon P_\varepsilon(z,\zeta)},\ \tilde{f}_\varepsilon(z)=e^{-\tilde{P}_\varepsilon(z,\zeta)}$$

represent univalent analytic functions in D and $\tilde{D}$, respectively.

For the sake of simplicity, we shall assume in the following consideration that C is an analytic curve. There exists, therefore, an analytic function $z=f(t)$ which maps a neighborhood of a segment of the real axis in the t-plane onto a neighborhood of a given arc of C. The function $G_\varepsilon(z,\ \zeta)$ becomes a harmonic function $g(t)$ to both sides of the segment. It goes continuously through the segment, but its normal derivatives satisfy the discontinuity law

$$(39)\qquad \frac{\partial g}{\partial n}+\varepsilon\frac{\partial g}{\partial\tilde{n}}=0\quad\text{for real } t.$$

Let

$$(40)\qquad p(t)=P_\varepsilon(f(t),\ \zeta),\ \tilde{p}(t)=\tilde{P}_\varepsilon(f(t),\ \zeta)\,.$$

We find easily for t in the segment and in view of the described discontinuity behavior of $g(t)$:

$$(41)\qquad \Re\{p'(t)\}=\Re\{\tilde{p}'(t)\},\ \Im\{p'(t)\}=\varepsilon\Im\{\tilde{p}'(t)\}\,.$$

We can combine the two relations (41) into the one equation:

$$(42)\qquad p'(t)=\frac{1+\varepsilon}{2}\tilde{p}'(t)+\frac{1-\varepsilon}{2}\overline{\tilde{p}'(\bar{t})}\,.$$

This formula allows an analytic continuation of $\tilde{p}'(t)$ into the upper halfplane and of $p'(t)$ into the lower. This proves that $p'(t)$ and $\tilde{p}'(t)$ are still analytic on the segment of the real axis in the t-plane. Returning to the z-plane we can infer that the functions

$$P'_\varepsilon(z,\ \zeta)=\frac{d}{dz}P_\varepsilon(z,\ \zeta),\ \ \tilde{P}'_\varepsilon(z,\ \zeta)=\frac{d}{dz}\tilde{P}_\varepsilon(z,\ \zeta) \tag{43}$$

are analytic beyond the curve C. Thus we proved that the two determinations of the Green's functions $G_\varepsilon(z,\ \zeta)$ are still regular harmonic on C if C is an analytic curve.

We derive from (40) and (42) that

$$\frac{P'_\varepsilon(z,\ \zeta)}{\tilde{P}'_\varepsilon(z,\ \zeta)}=\frac{1+\varepsilon}{2}+e^{-2i\alpha}\frac{1-\varepsilon}{2}\,, \qquad \alpha=\arg\tilde{p}'(t)\,. \tag{44}$$

Since we assume throughout $\varepsilon>0$, we see that the ratio (44) always lies in the right half of the complex plane. This implies

$$\varDelta\arg\tilde{P}'_\varepsilon(z,\ \zeta)=\varDelta\arg P'_\varepsilon(z,\ \zeta) \tag{45}$$

if z runs through the curve C in the positive sense with respect to D.

But by the argument principle we have

$$\varDelta\arg P'_\varepsilon(z,\ \zeta)=Z-P\,,\ \ \varDelta\arg\tilde{P}'_\varepsilon(z,\ \zeta)=\tilde{P}-\tilde{Z} \tag{46}$$

where P, Z are the numbers of zeros and poles of P'_ε in D and $\tilde{P}$, $\tilde{Z}$ have the same meaning with respect to $\tilde{P}'_\varepsilon$ and $\tilde{D}$. In case some zero of P'_ε should lie on C, we can deform the curve in such a way that it does not contain any zero and draw the same conclusion in view of the analyticity of P'_ε and $\tilde{P}'_\varepsilon$ on C.

We know by definition that if $\zeta\in D$ we have

$$P=1,\ Z\geqq 0;\ \tilde{P}=0,\ \tilde{Z}\geqq 1\,. \tag{47}$$

Hence, from (45), (46) and (47), we conclude

$$Z-1\leqq -1\,. \tag{48}$$

This is only possible if

$$\tilde{Z}=1,\ Z=0\,. \tag{49}$$

Hence we can state that $P'_\varepsilon(z,\ \zeta)$ and $\tilde{P}'_\varepsilon(z,\ \zeta)$ do not vanish at any finite point of the z-plane.

Consider now the system of differential equations $(z=x+iy)$

$$\frac{dx}{dt}=-\frac{\partial}{\partial x}G_\varepsilon(z,\ \zeta),\ \ \frac{dy}{dt}=-\frac{\partial}{\partial y}G_\varepsilon(z,\ \zeta)\,. \tag{50}$$

Along each solution curve $x(t)$, $y(t)$ of this system we have

$$\frac{d}{dt} G_\varepsilon(z(t),\ \zeta) = -\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2\right] < 0\ . \tag{51}$$

We have just shown that no critical point exists where $\nabla G_\varepsilon = 0$. Hence the net of solution curves covers the entire z-plane in a regular manner. All curves start out from the point $z=\zeta$ and run towards infinity. Each curve possesses the integral

$$\Im\{P(z,\ \zeta)\} = \text{const. or } \Im\{\tilde{P}_\varepsilon(z,\ \zeta)\} = \text{const.}\ , \tag{52}$$

according as it is considered in D or in $\tilde{D}$. From these facts it is evident that the functions (38) have the asserted univalency properties in D and $\tilde{D}$, respectively.

The importance of our result lies in the fact that the numerous distortion theorems of univalent function theory are now at our disposal in order to derive estimates of the various potential theoretical quantities connected with $G_\varepsilon(z,\ \zeta)$ in terms of the geometry of the curve C.

Let us observe, further, that for $\varepsilon = 1$ the function $\tilde{f}_1(z)$ represents the identity mapping while for $\varepsilon = 0$ we conclude from (21) that

$$\tilde{f}_0(z) = e^{-\tilde{\gamma}} \cdot e^{\tilde{P}(z,\infty)} = z + c_1 + \frac{c_1}{z} + \cdots \tag{53}$$

is the univalent function which maps $\tilde{D}$ onto the exterior of a circle of radius $e^{-\tilde{\gamma}}$ and which has at infinity the derivative one. Thus we can interpolate a continuous sequence of univalent mappings between the identity map of $\tilde{D}$ and its normalized mapping onto the exterior of a circle.

The preceding considerations show clearly the significance of the Fredholm eigen values and eigen functions for the dielectric problem and the general potential theory of the curve C. A generalization of most concepts to the physically more interesting case of three dimensions is easily done.

3. The variation of the eigen values. The variation of the eigen values λ_ν under a variation of the curve C can be determined by using the variational theory of the Green's function and of the various kernel functions connected with it [3]. In this paper we wish to give a straightforward and elementary derivation of the variational formulas.

Let z_0 be an arbitrary fixed point in $\tilde{D}$ and consider the mapping

$$z^* = z + \frac{\alpha}{z - z_0}\ . \tag{1}$$

For small enough α this will be a univalent mapping of C into a new smooth curve C^*. Let us denote its eigen values by λ_ν^* and its eigen functions by $w_\nu^*(z)$. We have used various eigen function definitions in the domain D; the $w_\nu^*(z)$ shall play the same role with respect to D^* (the domain bounded by C^*) as the $w_\nu(z)$ defined in Section 1 played with respect to D.

We have the integral equation

$$(2) \qquad w_\nu^*(z^*)=\frac{\lambda_\nu^*}{2\pi i}\int_{C^*}\frac{\overline{(w_\nu^*(\zeta^*)d\zeta^*)}}{\zeta^*-z^*}, \qquad z^*\in D^* .$$

Let us define

$$(3) \qquad m_\nu(z)=w_\nu^*\Big(z+\frac{\alpha}{z-z_0}\Big)\Big(1-\frac{\alpha}{(z-z_0)^2}\Big) .$$

This is a regular analytic function in D since (1) maps D univalently onto D^* where $w_\nu^*(z^*)$ is analytic. Using (3), we can rewrite (2) into the simpler form

$$(4) \qquad m_\nu(z)=\Big(1-\frac{\alpha}{(z-z_0)^2}\Big)\cdot\frac{\lambda_\nu^*}{2\pi i}\int_C\Big[1-\frac{\alpha}{(z-z_0)(\zeta-z_0)}\Big]^{-1}\frac{\overline{(m_\nu(\zeta)d\zeta)}}{\zeta-z} .$$

We have thus referred all variables back to our original domain D, but λ_ν^* and $m_\nu(z)$ appear now as the eigen values and eigen functions of an integral equation with slightly changed kernel.

We may transform the new integral equation (4) by easy calculations into

$$(5) \qquad m_\nu(z)=\frac{\lambda_\nu^*}{\pi}\iint_D\frac{\overline{m_\nu(\zeta)}}{(\zeta-z)^2}d\tau-\alpha\frac{\lambda_\nu^*}{\pi}\iint_D\frac{\overline{m_\nu(\zeta)d\tau}}{[(z-z_0)(\zeta-z_0)-\alpha]^2} .$$

Observe that by the definition (3) we have

$$(6) \qquad \iint_D|m_\nu(z)|^2d\tau=\iint_{D^*}|w_\nu^*(z^*)|^2d\tau^*=1 .$$

We have thus to determine the normalized eigen functions $m_\nu(z)$ to the integral equation (5) which differs from our original equation (1.13) by an α-term which can be estimated uniformly in z for $z_0\in\tilde{D}$ fixed.

Let us define the analytic function [3]

$$(7) \qquad L(z,\zeta)=-\frac{2}{\pi}\frac{\partial^2}{\partial z\,\partial\zeta}G(z,\zeta)=\frac{1}{\pi(z-\zeta)^2}-l(z,\zeta) .$$

It is well-known that for every function $f(z)$ which is analytic in D and for which $\iint_D|f|^2d\tau<\infty$ holds

$$\iint_D L(z, \zeta)\overline{f(\zeta)}d\tau = 0 . \tag{8}$$

Hence we have the identity, valid for each such $f(z)$,

$$\frac{1}{\pi}\iint_D \frac{\overline{f(\zeta)}}{(\zeta - z)^2}d\tau = \iint_D l(z, \zeta)\overline{f(\zeta)}d\tau . \tag{9}$$

Under our assumptions about the boundary curve C of D, it can be shown that $l(z, \zeta)$ is continuous in both variables in the closed region $D+C$. Thus (5) can be put into the form:

$$m_\nu(z) = \lambda_\nu^* \iint_D l(z, \zeta)\overline{m_\nu(\zeta)}d\tau - \alpha\frac{\lambda_\nu^*}{\pi}\iint_D \frac{\overline{m_\nu(\zeta)}}{[(z-z_0)(\zeta - z_0) - \alpha]^2}d\tau , \tag{10}$$

while $w_\nu(z)$ satisfies the unperturbed integral equation

$$w_\nu(z) = \lambda_\nu \iint_D l(z, \zeta)\overline{w_\nu(\zeta)}d\tau . \tag{11}$$

Now we can apply the general perturbation theory for regular kernels [9] and state that the eigen functions $m_\nu(z)$ and the eigen values λ_ν^* are analytic functions of the perturbation parameters α and $\bar{\alpha}$ and can be developed in power series in them. For $\alpha = 0$, λ_ν^* will coincide with λ_ν while $m_\nu(z)$ will then lie in the linear space spanned by the eigen functions of (11) which belong to the unperturbed eigen value λ_ν.

Let $w_\nu^{(j)}(z)$ $(j=1, \cdots, n)$ denote the eigen functions belonging to λ_ν. We have the developments

$$\lambda_\nu^* = \lambda_\nu + |\alpha|\kappa_\nu + O(|\alpha|^2) \tag{12}$$

and

$$m_\nu(z) = \sum_{j=1}^n A_j w_\nu^{(j)}(z) + |\alpha|\omega_\nu(z) + O(|\alpha|^2) . \tag{13}$$

Inserting (12) and (13) into (10) and making use of (11), we find

$$\begin{aligned} \sum_{j=1}^n A_j w_\nu^{(j)}(z) = \frac{\lambda_\nu^*}{\lambda_\nu}\sum_{j=1}^n \bar{A}_j w_\nu^{(j)}(z) + |\alpha|\lambda_\nu \iint_D l(z, \zeta)\overline{\omega_\nu(\zeta)}d\tau \\ - |\alpha|\omega_\nu(z) - \frac{\alpha\lambda_\nu}{(z-z_0)^2\pi}\sum_{j=1}^n \bar{A}_j \iint_D \frac{\overline{w_\nu^{(j)}(\zeta)}}{(\zeta - z_0)^2}d\tau + O(|\alpha|^2) . \end{aligned} \tag{14}$$

We multiply this identity with $w_\nu^{(k)}(z)$ and integrate over D. We use the orthonormality of the $w_\nu^{(k)}(z)$, the symmetry of $l(z, \zeta)$ and the integral equation (11). We also make use of the fact that by (1.36)

$$\frac{1}{\pi}\iint_D \frac{\overline{w_\nu(\zeta)}}{(\zeta-z_0)^2}\,d\tau = \frac{\sqrt{\lambda_\nu^2-1}}{i\lambda_\nu}\tilde{w}_\nu(z_0). \tag{15}$$

Hence we arrive at

$$A_k = \frac{\lambda_\nu^*}{\lambda_\nu}\overline{A}_k + 2i|\alpha|\Im\left\{\iint_D w_\nu^{(k)}(z)\overline{\omega_\nu(z)}\,d\tau\right\} \tag{16}$$

$$+\alpha\frac{\pi(\lambda_\nu^2-1)}{\lambda_\nu}\sum_{j=1}^{n}\overline{A}_k\tilde{w}_\nu^{(j)}(z_0)\tilde{w}_\nu^{(k)}(z_0)+O(|\alpha|^2)\ , \quad k=1, 2, \cdots, n\,.$$

Using the development (12) and comparing equal powers of $|\alpha|$ on both sides, we obtain

$$\Im\{A_k\}=0\ , \qquad A_k=\text{real},\ k=1, 2, \cdots, n\,. \tag{17}$$

Taking real parts in (16) and putting

$$\operatorname{sgn}\alpha = \frac{\alpha}{|\alpha|} = e^{i\beta}\,, \tag{18}$$

we find

$$\kappa_\nu A_k + \pi(\lambda_\nu^2-1)\sum_{j=1}^{n} A_j\Re\{e^{i\beta}\tilde{w}_\nu^{(j)}(z_0)\tilde{w}_\nu^{(k)}(z_0)\}=0\ , \quad k=1, 2, \cdots, n\,. \tag{19}$$

Thus the possible values of κ_ν in the development (12) of the perturbed eigen value λ_ν^* are the eigen values of the secular equation

$$\det\|\kappa_\nu\delta_{jk}+\pi(\lambda_\nu^2-1)\Re\{e^{i\beta}\tilde{w}_\nu^{(j)}(z_0)\tilde{w}_\nu^{(k)}(z_0)\}\|=0\,. \tag{20}$$

In particular, if λ_ν is a simple (nondegenerate) eigen value, we have the simple variational formula

$$\delta\lambda_\nu = |\alpha|\cdot\kappa_\nu = -\pi(\lambda_\nu^2-1)\Re\{\alpha\tilde{w}_\nu(z_0)^2\}\,. \tag{21}$$

Let us suppose next that we perform a variation (1) of the curve C but now with $z_0\in D$. Since the mapping (1) is regular and univalent in $\tilde{D}$, we can repeat the entire argument by interchanging the roles of D and $\tilde{D}$. We thus find

$$\det\|\kappa_\nu\delta_{jk}+\pi(\lambda_\nu^2-1)\Re\{e^{i\beta}w_\nu^{(j)}(z_0)w_\nu^{(k)}(z_0)\}\|=0 \tag{22}$$

as the secular equation for the κ_ν-terms and

$$\delta\lambda_\nu = |\alpha|\kappa_\nu = -\pi(\lambda_\nu^2-1)\Re\{\alpha w_\nu(z_0)^2\} \tag{23}$$

in the nondegenerate case. Formulas (21) and (23) exhibit the complete symmetry of our theory with respect to D and $\tilde{D}$.

We used the method of interior variations (1) in order to reduce the variational problem for the λ_ν explicitly to the theory of perturbation in classical integral equation theory. The formulas obtained are also very convenient in various extremum problems regarding the λ_ν as we shall show later. It seems, however, desirable to give also a variational formula for deformations of C which are described by the normal shift δn of each point on C. For this purpose we put

$$\Re\{\pi(\lambda_\nu^2-1)\alpha w_\nu^{(j)}(z_0)w_\nu^{(k)}(z_0)\} = \Re\left\{\frac{\alpha}{2i}(\lambda_\nu^2-1)\oint_C \frac{w_\nu^{(j)}(\zeta)w_\nu^{(k)}(\zeta)}{\zeta-z_0}d\zeta\right\}, \qquad z_0\in D. \tag{24}$$

Applying Cauchy's integral theorem with respect to $\tilde{D}$, we also find

$$0=\Re\left\{\frac{\alpha}{2i}(\lambda_\nu^2-1)\oint_C \frac{\tilde{w}_\nu^{(j)}(\zeta)\tilde{w}_\nu^{(k)}(\zeta)}{\zeta-z_0}d\zeta\right\}. \tag{25}$$

Finally, we derive from (1.34) that

$$(\lambda_\nu^2-1)\{w_\nu^{(j)}(\zeta)w_\nu^{(k)}(\zeta)-\tilde{w}_\nu^{(j)}(\zeta)\tilde{w}_\nu^{(k)}(\zeta)\}\zeta'^2 = 2\Re\{\lambda_\nu\tilde{w}_\nu^{(j)}(\zeta)\overline{\tilde{w}_\nu^{(k)}(\zeta)}-\lambda_\nu^2\tilde{w}_\nu^{(j)}(\zeta)\tilde{w}_\nu^{(k)}(\zeta)\zeta'^2\}. \tag{26}$$

Hence, if we subtract (25) from (24), we obtain

$$\Re\{\pi(\lambda_\nu^2-1)\alpha w_\nu^{(j)}(z_0)w_\nu^{(k)}(z_0)\} = \oint_C \Re\{\lambda_\nu\tilde{w}_\nu^{(j)}(\zeta)\overline{\tilde{w}_\nu^{(k)}(\zeta)}-\lambda_\nu^2\tilde{w}^{(j)}(\zeta)\tilde{w}_\nu^{(k)}(\zeta)\zeta'^2\}\,\delta n\, ds \tag{27}$$

where

$$\delta n=\Re\left\{\frac{1}{i\zeta'}\frac{\alpha}{\zeta-z_0}\right\} \tag{28}$$

represents the normal shift of C under the deformation (1). Thus the coefficients of the secular equation for $\delta\lambda$ have been expressed in terms of δn.

In particular, we have in the nondegenerate case in view of (23)

$$\delta\lambda_\nu=\int_C[\lambda_\nu^2\Re\{\tilde{w}_\nu(\zeta)^2\zeta'^2\}-\lambda_\nu|\tilde{w}_\nu(\zeta)|^2]\delta n\, ds. \tag{29}$$

It can easily be verified from (1.34) that on C

$$\lambda_\nu\Re\{\tilde{w}_\nu^{(j)}\overline{\tilde{w}_\nu^{(k)}}\}-\lambda_\nu^2\Re\{\tilde{w}_\nu^{(j)}\tilde{w}_\nu^{(k)}\zeta'^2\} = -\lambda_\nu\Re\{w_\nu^{(j)}\overline{w_\nu^{(k)}}\}+\lambda_\nu^2\Re\{w_\nu^{(j)}w_\nu^{(k)}\zeta'^2\}. \tag{30}$$

Thus we may replace $\tilde{w}$ by w in formulas (27) and (29); since transition from D to $\tilde{D}$ implies also a change of sign of the interior normal, the end result is unchanged. Thus the variational formula of the Hadamard type (29) is entirely symmetric with respect to the two complementary domains considered. If we had chosen $z_0 \in \tilde{D}$, we would have obtained the same end result (29).

We derived (29) in the case of a particular variation of the type (1). But since a variational formula depends linearly and additively on the variation, and since we can approximate general δn-variations by superposition of special variations of the type (1), we can extend (29) to the most general case of a δn-variation.

The value of the variational formula (29) is of heuristic nature; it shows the dependence of λ_ν on the geometry of C. For a precise study of extremum problems it is preferable to apply the variational formulas based on interior variations of the type (1).

We can derive, however, interesting monotonicity results by means of (29). Let, for example, $z=f(u)$ give the conformal mapping of the unit circle $|u|<1$ onto the domain D. Let C_r be the image under this map of the circumference $|u|=r<1$; and let $\lambda(r)$, $w(z, r)$ denote, say, the νth eigen value and eigen function of C_r. We assume, for the sake of simplicity, that $\lambda(r)$ is nondegenerate and then easily derive from (29):

$$\text{(29')} \qquad \frac{d}{dr}\lambda(r) = -\lambda(r)\oint_{|u|=r} |w(z, r)|^2 |f'(u)|^2\, ds_u$$

$$+\lambda(r)^2\, \Re\left\{\frac{i}{r}\oint_{|u|=r} w(z, r)^2 f'(u)^2 u\, du\right\}.$$

The function

$$F_r(u) = w[f(u), r]f'(u)$$

is regular analytic for $|u| \leq r$; hence the second integral in (29') vanishes by Cauchy's integral theorem and we obtain:

$$\text{(29'')} \qquad \frac{d}{dr}\log\lambda(r) = -\int_{|u|=r} |F_r(u)|^2\, ds_u < 0.$$

The eigen values $\lambda(r)$ of the level curves C_r are monotonically decreasing if r increases.

For every function $F(u)$ which is regular analytic for $|u| \leq r$ holds the obvious inequality

$$\text{29''')} \qquad \int_{|u|=r} |F(u)|^2\, ds_u \geqq \frac{2}{r}\iint_{|u|<r} |F(u)|^2\, d\tau.$$

Observe now that because of the normalization of $w(z, r)$ inside of C_r the function $F_r(u)$ is normalized in the circle $|u|<r$. Hence, combining (29″) with (29‴), we finally obtain

$$(29^{\text{iv}}) \qquad \frac{d}{dr}(\lambda r^2) \leqq 0 \,.$$

Since we have the trivial estimate $\lambda(1) \geqq 1$ for every curve C, we then derive from (29^{iv}) the useful estimate

$$(29^{\text{v}}) \qquad \lambda(r) \geqq \frac{1}{r^2} \qquad \text{for } r \leqq 1 \,.$$

In order to apply the usual perturbation method of integral equation theory we had to replace the integral equation (1.13) with singular kernel by the integral equation (11) which has the regular symmetric kernel $l(z, \zeta)$. The necessity for this transition becomes clear when we consider the exceptional case that C is a circumference. In this case (and only then), we have $l(z, \zeta)=0$. The original integral equation (1.13) has only the eigen value $\lambda=\infty$ and each function $f(z)$ which is analytic in D is an eigen function.

In fact, suppose for the sake of simplicity that C is the unit circumference $z \cdot \bar{z}=1$. We have

$$(31) \qquad \frac{1}{\pi}\iint_{|\zeta|<1} \frac{\overline{f(\zeta)}}{(\zeta-z)^2}d\tau = \frac{1}{2\pi i}\oint_{|\zeta|=1}\frac{\overline{f(\zeta)d\zeta}}{\zeta - z} = \overline{\frac{1}{2\pi i}\oint_{|\zeta|=1}\frac{\zeta f(\zeta) d\zeta}{\zeta\bar{z}-1}} \,.$$

By means of the residue theorem we conclude therefore

$$(32) \qquad \frac{1}{\pi}\iint_{|\zeta|<1}\frac{\overline{f(\zeta)}}{(\zeta-z)^2}d\tau = \begin{cases} 0 & \text{if } |z|<1 \\ \dfrac{1}{z^2}\bar{f}\left(\dfrac{1}{z}\right) & \text{if } |z|>1 \,. \end{cases}$$

This equation proves our statement that $\lambda=\infty$ is the only eigen value of (1.13) in this case and that it is of infinite degeneracy.

Our variational theory does not work in this exceptional case. However, let $|z_0|>1$ and C^* be the image of $|z|=1$ under the variation (1). We define its eigen function $w_\nu^*(z)$ and by (3) a function $m_\nu(z)$ which is regular analytic in D. It satisfies the integral equation (5) which, in view of (32), can be brought into the simple form

$$(33) \qquad m_\nu(z) = -\frac{\alpha\lambda_\nu^*}{(z-z_0)^2}\frac{1}{\eta^2}\overline{m}_\nu\left(\frac{1}{\eta}\right), \qquad \eta = z_0 + \frac{\alpha}{z-z_0} \,.$$

Let

$$m_\nu(z)=\frac{d}{dz}M_\nu(z)\ ; \tag{34}$$

if we choose the right constant of integration in the definition of $M_\nu(z)$, we can integrate (33) to the identity

$$M_\nu(z)=-\lambda_\nu^*\overline{M}_\nu(L(z))\ , \tag{35}$$

where

$$L(z)=\eta^{-1}=\frac{z-z_0}{z_0(z-z_0)+\alpha} \tag{36}$$

is a linear function of z. Thus we obtain a simple functional equation for the eigen functions $w_\nu^*(z)$ and the eigen values λ_ν^* of the varied curve C^*. α must be sufficiently small in order that the mapping (1) be univalent in D; but we have not made any neglection of higher powers of α and (35) will give the precise value of λ_ν^*.

If we iterate (35), we obtain

$$M_\nu(z)=\lambda_\nu^{*2}M_\nu(\Lambda(z))\ , \qquad \Lambda=\overline{L}(L(z))\ . \tag{37}$$

If z_1, z_2 are the fixed points of the linear transformation $Z=\Lambda(z)$, we can write

$$\frac{Z-z_1}{Z-z_2}=\tau^2\frac{z-z_1}{z-z_2} \tag{38}$$

where $|z_1|<1$, $|z_2|>1$. The eigen functions $M_\nu(z)$ are of the form

$$M_\nu(z)=A_\nu\left(\frac{z-z_1}{z-z_2}\right)^\nu, \qquad \nu=1, 2\cdots \tag{39}$$

and belong to the eigen values

$$\lambda_\nu^*=\pm\tau^{-\nu}\ . \tag{40}$$

Thus all eigen functions and eigen values of the curve C^* can be calculated explicitly. An easy computation shows that for small values of ε

$$\tau^{-1}=\frac{(|z_0|^2-1)^2}{|\alpha|}+O(1)\ . \tag{41}$$

An analogous calculation can be performed if the unit circle is transformed by a variation (1) with $|z_0|<1$. If we consider a superposition of variations (1), we can still derive an asymptotic formula for the eigen values λ_ν^* obtained. Thus we have shown that the eigen values for nearly circular domains can be obtained asymptotically in spite of the fact that the circle has an infinitely degenerate eigen value.

We showed at the end of § 1 that the eigen values of an ellipse can be calculated explicitly. This result is a particular case of our preceding investigation since the exterior of the ellipse is obtained from the exterior of the unit circle by a transformation (1) with $z_0=0$ and $|\alpha|<1$.

There are relatively few domains for which the eigen values and eigen functions of the Fredholm integral equation are known. It is, therefore, important to possess at least an asymptotic formula for the eigen values of nearly circular domains which admits many arbitrary parameters. Such formulas are particularly useful when one wishes to test hypotheses with respect to the eigen values of general domains.

4. The variation of the dielectric Green's function. In this section we want to derive the formula for the variation of the dielectric Green's function $G_\varepsilon(z, \zeta)$ defined in § 2. It will appear that it possesses a very simple variational formula which is quite similar to that for the ordinary Green's function of a plane domain. We shall again consider the interior variation

$$z^*=z+\frac{\alpha}{z-z_0} \tag{1}$$

which transforms the curve C into a curve C^* defining the two complementary domains D^* and $\tilde{D}^*$. Let $G_\varepsilon^*(z, \zeta)$ be the corresponding dielectric Green's function to the parameter ε.

If $z_0 \in \tilde{D}$, the mapping (1) will be univalent and regular in D for small enough α; hence the function

$$\Gamma_\varepsilon(z, \zeta)=G_\varepsilon^*\left(z+\frac{\alpha}{z-z_0}, \zeta+\frac{\alpha}{\zeta-z_0}\right) \tag{2}$$

will be harmonic in D. It will also be harmonic in $\tilde{D}$, except for the interior of a circle of radius $|\alpha|^{1/2}$ around the point z_0. The function $\Gamma_\varepsilon(z, \zeta)$ will have logarithmic poles at infinity and for $z=\zeta$ as follows from the definition of $G_\varepsilon(z, \zeta)$.

We consider now Green's identity:

$$\frac{1}{2\pi}\int_C\left[\Gamma_\varepsilon(t, z)\frac{\partial G_\varepsilon(t, \zeta)}{\partial n}-G_\varepsilon(t, \zeta)\frac{\partial \Gamma_\varepsilon(t, z)}{\partial n}\right]ds \tag{3}$$
$$=\varepsilon\Gamma_\varepsilon(\zeta, z)\delta(\zeta)-\varepsilon G_\varepsilon(z, \zeta)\delta(z) .$$

Observe that in view of the conformality of (1) on C the function $\Gamma_\varepsilon(z, \zeta)$ has the same continuity (and discontinuity) property on C as the original function $G_\varepsilon(z, \zeta)$. Hence we may transform (3) into

$$(4) \qquad \varepsilon[\Gamma_\varepsilon(\zeta, z)\delta(\zeta)-G_\varepsilon(z, \zeta)\delta(z)]=-\frac{\varepsilon}{2\pi}\int_c\left[\Gamma_\varepsilon(t, z)\frac{\partial G_\varepsilon(t, \zeta)}{\partial\tilde{n}} - G_\varepsilon(t, \zeta)\frac{\partial \Gamma_\varepsilon(t, z)}{\partial\tilde{n}}\right]ds .$$

Now we can apply Green's identity with respect to the domain $\tilde{D}$ after removing from it the interior of the circle $|z-z_0|=|\alpha|^{1/2}$ which we denote by c. Let us assume that neither z nor ζ lie inside c; then (4) yields

$$(5) \quad \Gamma_\varepsilon(z, \zeta)-G_\varepsilon(z, \zeta)=-\frac{1}{2\pi}\int_c\left[\Gamma_\varepsilon(t, z)\frac{\partial G_\varepsilon(t, \zeta)}{\partial n}-G_\varepsilon(t, \zeta)\frac{\partial \Gamma_\varepsilon(t, z)}{\partial n}\right]ds .$$

We have now fully utilized the boundary behavior of $G_\varepsilon(z, \zeta)$. The evaluation of the c-integral follows exactly the lines of the calculation for the ordinary Green's function. We put for $t \in c$

$$(6) \qquad t=z_0+|\alpha|^{1/2}e^{i\phi}$$

and evaluate the right-hand integral in (5) by power series development. We define again two analytic functions of z, namely $p_\varepsilon(z, \zeta)$ and $p_\varepsilon^*(z, \zeta)$, by

$$(7) \qquad \Re\{P_\varepsilon(z, \zeta)\}=G_\varepsilon(z, \zeta), \quad \Re\{P_\varepsilon^*(z, \zeta)\}=G_\varepsilon^*(z, \zeta) .$$

Further, let

$$(7') \qquad P_\varepsilon'(z, \zeta)=\frac{d}{dz}P_\varepsilon(z, \zeta), \quad P_\varepsilon^{*\prime}(z, \zeta)=\frac{d}{dz}P_\varepsilon^*(z, \zeta).$$

Then the usual calculations yield

$$(8) \qquad G_\varepsilon^*(z^*, \zeta^*)-G_\varepsilon(z, \zeta)=\Re\{\alpha P_\varepsilon^{*\prime}(z_0, z^*)P_\varepsilon'(z_0, \zeta)\}+O(|\alpha|^2) .$$

Further series development leads to the simple result

$$(9) \qquad G_\varepsilon^*(z, \zeta)=G_\varepsilon(z, \zeta)+\Re\left\{\alpha\left[P_\varepsilon'(z_0, z)P_\varepsilon'(z_0, \zeta) - \frac{P_\varepsilon'(z, \zeta)}{z-z_0}-\frac{P_\varepsilon'(\zeta, z)}{\zeta-z_0}\right]\right\}+O(|\alpha|^2) .$$

This is exactly the same variational formula as for the ordinary Green's function [12, 13]. It has been derived for $z_0 \in \tilde{D}$.

If we had chosen $z_0 \in D$ instead of $\tilde{D}$ analogous calculations would have been applicable. We could start with

(10)
$$\frac{1}{2\pi}\int_c\left[\Gamma_\varepsilon(t,\ z)\frac{\partial G_\varepsilon(t,\ \zeta)}{\partial \tilde{n}}-G_\varepsilon(t,\ \zeta)\frac{\partial \Gamma_\varepsilon(t,\ z)}{\partial \tilde{n}}\right]ds$$
$$=\Gamma_\varepsilon(\zeta,\ z)\tilde{\delta}(\zeta)-G_\varepsilon(z,\ \zeta)\tilde{\delta}(z)\ .$$

Using the discontinuity of $\dfrac{\partial G_\varepsilon}{\partial n}$ on C, we find

(11)
$$\varepsilon[\Gamma_\varepsilon(\zeta,\ z)\tilde{\delta}(\zeta)-G_\varepsilon(z,\ \zeta)\tilde{\delta}(z)]=-\frac{1}{2\pi}\int_c\left[\Gamma_\varepsilon(t,\ z)\frac{\partial G_\varepsilon(t,\ \zeta)}{\partial n}\right.$$
$$\left.-G_\varepsilon(t,\ \zeta)\frac{\partial \Gamma_\varepsilon(t,\ z)}{\partial n}\right]ds$$

and by means of Green's identity

(12)
$$\varepsilon[\Gamma_\varepsilon(z,\ \zeta)-G_\varepsilon(z,\ \zeta)]=-\frac{1}{2\pi}\int_c\left(\Gamma_\varepsilon\frac{\partial G_\varepsilon}{\partial n}-G_\varepsilon\frac{\partial \Gamma_\varepsilon}{\partial n}\right)ds$$

where c denotes again the circle $|z-z_0|=|\alpha|^{1/2}$. In this case the same procedure as before yields the result for $z_0\in D$:

(13)
$$G_\varepsilon^*(z,\ \zeta)-G_\varepsilon(z,\ \zeta)=\Re\left\{\alpha\left[\frac{1}{\varepsilon}P_\varepsilon'(z_0,\ z)P_\varepsilon'(z_0,\ \zeta)\right.\right.$$
$$\left.\left.-\frac{P_\varepsilon'(z,\ \zeta)}{z-z_0}-\frac{P_\varepsilon'(\zeta,\ z)}{\zeta-z_0}\right]\right\}+O(|\alpha|^2)\ .$$

Observe the factor $\dfrac{1}{\varepsilon}$ which is now introduced into (13) and causes a slight change in the variational formula.

We have thus derived a very elegant variational formula for the dielectric Green's function; its significance is seen from the numerous applications of its analogue in the case of the ordinary Green's function [12, 13, 14].

As mentioned in § 2, the function $P_\varepsilon(z,\ \zeta)$ consists in reality of two analytic functions, say, $P_\varepsilon(z,\ \zeta)$ if $z\in D$ and $\tilde{P}_\varepsilon(z,\ \zeta)$ if $z\in\tilde{D}$. The boundary behavior of $G_\varepsilon(z,\ \zeta)$ as described in § 2 implies for $z\in C$

(14)
$$\Re\{P_\varepsilon'(z,\ \zeta)z'\}=\Re\{\tilde{P}_\varepsilon'(z,\ \zeta)z'\},\quad \Im\{P_\varepsilon'(z,\ \zeta)z'\}=\varepsilon\Im\{\tilde{P}_\varepsilon'(z,\ \zeta)z'\}\ .$$

We can combine the variational formulas (9) and (13) into the integral form:

(15)
$$G_\varepsilon^*(z,\ \zeta)-G_\varepsilon(z,\ \zeta)$$
$$=\Re\left\{\frac{\alpha}{2\pi i}\oint_C\frac{\frac{1}{\varepsilon}P_\varepsilon'(t,\ z)P_\varepsilon'(t,\ \zeta)-\tilde{P}_\varepsilon(t,\ z)\tilde{P}_\varepsilon'(t,\ \zeta)}{t-z_0}dt\right\}+O(|\alpha|^2)\ .$$

By use of (14) this can be simplified to

$$(16)\qquad \delta G_\varepsilon(z,\ \zeta) = \frac{1}{2\pi}\left(\frac{1}{\varepsilon}-1\right)\int_C\left[\frac{\partial G_\varepsilon(t,\ z)}{\partial s}\frac{\partial G_\varepsilon(t,\ \zeta)}{\partial s} - \frac{\partial G_\varepsilon(t,\ z)}{\partial n}\frac{\partial G_\varepsilon(t,\ \zeta)}{\partial \tilde{n}}\right]\delta n\, ds$$

with

$$(17)\qquad \delta n = \Re\left\{\frac{1}{it'}\frac{\alpha}{(t-z_0)}\right\}.$$

This is the Hadamard type variational formula for the dielectric Green's function which has been proved in a precise manner through use of our interior variational method.

Since we can also write (16) in the form

$$(18)\qquad \delta G_\varepsilon(z,\ \zeta) = \frac{1}{2\pi}\left(\frac{1}{\varepsilon}-1\right)\int_C\left[\frac{\partial G_\varepsilon(t,\ z)}{\partial s}\frac{\partial G_\varepsilon(t,\ \zeta)}{\partial s} + \frac{1}{\varepsilon}\frac{\partial G_\varepsilon(t,\ z)}{\partial n}\frac{\partial G_\varepsilon(t,\ \zeta)}{\partial n}\right]\delta n\, ds$$

it is evident that if $\zeta\in D$ the expression $(G_\varepsilon(z,\ \zeta)+\varepsilon\log|z-\zeta|)_{z=\zeta}$ depends monotonically upon the domain D while for $\zeta\in\tilde{D}$ the same is true for $(G_\varepsilon(z,\ \zeta)+\log|z-\zeta|)_{z=\zeta}$. In a similar way many other expressions can be constructed which have a definite factor of $\delta n\, ds$ under the integral sign and which depend, therefore, monotonically upon D. The application of Hadamard's formula in order to obtain inequalities and comparison theorems for functionals connected with $G_\varepsilon(z,\ \zeta)$ is obvious.

For $\varepsilon=1$, we have $G_\varepsilon(z,\ \zeta)=-\log|z-\zeta|$ independently of the domain. For this reason the factor $\left(\frac{1}{\varepsilon}-1\right)$ must occur in the variational formulas (16) and (18).

We showed at the end of § 2 that the mapping of a domain onto a circle can be connected with the identical mapping by a one-parameter family of univalent functions which are closely related to the dielectric Green's functions. For this reason it is of interest to compute the derivative of $G_\varepsilon(z,\ \zeta)$ with respect to ε.

We start with Green's identity and with $\varepsilon>0$, $e>0$:

$$(19)\qquad e\delta(\zeta)G_\varepsilon(\zeta,\ \eta)-\varepsilon\delta(\eta)G_e(\zeta,\ \eta) = \frac{1}{2\pi}\int_C\left[G_\varepsilon(z,\ \eta)\frac{\partial G_e(z,\ \zeta)}{\partial n} - G_e(z,\ \zeta)\frac{\partial G_\varepsilon(z,\ \eta)}{\partial n}\right]ds.$$

Using the boundary relations of G_ε and G_e on C and Green's identity with respect to $\tilde{D}$, we find

$$(20)\qquad \frac{G_\varepsilon(\zeta,\ \eta)-G_e(\zeta,\ \eta)}{\varepsilon-e}=\frac{1}{\varepsilon}G_\varepsilon(\zeta,\ \eta)\delta(\zeta)+\frac{1}{2\pi\varepsilon}\int_C G_\varepsilon(z,\ \eta)\frac{\partial G_e(z,\ \zeta)}{\partial\tilde{n}}ds\,.$$

Passing to the limit $\varepsilon=e$, we then obtain

$$(21)\qquad \frac{\partial}{\partial\varepsilon}G_\varepsilon(\zeta,\ \eta)=\frac{1}{\varepsilon}G_\varepsilon(\zeta,\ \eta)\delta(\zeta)+\frac{1}{2\pi\varepsilon}\int_C G_\varepsilon(z,\ \eta)\frac{\partial G_\varepsilon(z,\ \zeta)}{\partial\tilde{n}}ds\,.$$

The symmetry of this expression is more clearly exhibited in the form

$$(22)\qquad \frac{\partial}{\partial\varepsilon}G_\varepsilon(\zeta,\ \eta)=\frac{1}{2\pi\varepsilon^2}\iint_D \nabla G_\varepsilon(z,\ \zeta)\cdot\nabla G_\varepsilon(z,\ \eta)d\tau\,.$$

This result could also have been obtained by straightforward calculation from (2.19) and its analogues.

It is obvious how numerous monotonicity results can be derived from expression (22) by considering combinations with positive derivative. This formula can also be used in order to develop G_ε in powers of ε. The formula is particularly useful in a more detailed discussion of the mapping functions $f_\varepsilon(z)$, defined in § 2; however, we do not enter into this subject in the present paper.

5. An extremum problem for the Fredholm eigen values. We shall now proceed to apply the variational formulas of § 3 to an important extremum problem for the lowest Fredholm eigen value of a given curve C. In order to explain the formulation of the problem considered we start with the following observation. Let C be a three times continuously differentiable curve as was supposed throughout; if λ_1 is its lowest eigen value we have shown that $\lambda_1>1$. Now let C^* be a continuum which consists of all points of C plus a segment which has one endpoint on C and the other in D; let λ_1^* be its lowest eigen value. It can be shown that $\lambda_1^*=1$ however small the additional segment of C^* is; thus two curves in an arbitrary Fréchet neighborhood can have very different lowest Fredholm eigen values.

The fact that λ_1 depends in this discontinuous way on its defining curve C makes it difficult to frame significant extremum problems for it. The side condition on C of three continuous derivatives is, on the one hand, somewhat unnatural and, on the other hand, hard to preserve under variation. We shall restrict ourselves, therefore, in this section to the consideration of analytic curves, but even in this case λ_1 can come as near as we wish to 1. In fact, formula (1.45) shows that we can find ellipses with λ_1 arbitrarily near 1. We have, therefore, to sharpen the concept of an analytic curve by introducing the concept of *uniform analyticity* of a curve. A curve C is called analytic if it is mapped by a regular univalent function $z=f(\zeta)$ from the unit circumference $|\zeta|=1$. $f(\zeta)$ must be regular and univalent in some circular ring $r<|\zeta|<R$ with $r<1<R$. The class of all curves C which are analytic and belong to functions $f(\zeta)$ which are regular and univalent in a fixed ring (r, R) shall be called the class of uniformly analytic curves with the modulus of analyticity (r, R).

Because of the normality of the family of univalent functions in a fixed region the concept of uniform analyticity lends itself easily to the construction of significant extremum problems. In particular, let us ask for the minimum value of λ_1 within the family of all uniformly analytic curves with modulus (r, R).

We may consider our problem as an extremum problem on univalent functions. Given the class of all functions $f(\zeta)$ which are regular and univalent in $r<|\zeta|<R$, to find one in the class which maps the unit circumference onto a curve C with minimum λ_1. The existence of such a function follows easily from the usual normality arguments and we proceed at once to characterize the extremum function by varying it and comparing it with nearby competing functions.

Since the curve C mapped by the extremum function is analytic and since its λ_1 is obviously finite, the lowest eigen value can have only a degeneracy of finite order. Let $w_1^{(1)}(z), \cdots, w_1^{(n)}(z)$ be a complete and linearly independent set of eigen functions belonging to λ_1 in D, while $\tilde{w}_1^{(1)}(z), \cdots, \tilde{w}_1{}^{(n)}(z)$ are the corresponding eigen functions in $\tilde{D}$. Suppose that the image of $|\zeta|=r$ forms a continuum Γ in D while the image of $|\zeta|=R$ forms the continuum $\tilde{\Gamma}$ in $\tilde{D}$. Let $z_0\in\tilde{\Gamma}$; there exists an infinity of analytic functions which are univalent outside of the continuum $\tilde{\Gamma}$ and which have a series development [11]

$$z^*=z+\sum_{\nu=1}^{\infty}\frac{a_\nu\rho^{\nu+1}}{(z-z_0)^\nu} \tag{1}$$

which converges for $|z-z_0|>\rho$. The coefficients a_ν of this development are uniformly bounded

$$|a_\nu|\leqq 4^{\nu+1} \tag{2}$$

and ρ is a positive parameter which can be chosen arbitrarily small.

Let us insert the extremum function $z=f(\zeta)$ into (1); we will thus obtain an infinity of competing functions regular and univalent in $r<|\zeta|<R$ of the form

$$f^*(\zeta)=f(\zeta)+\frac{a_1\rho^2}{f(\zeta)-z_0}+o(\rho^2)\,. \tag{3}$$

They define curves C^*, the images of $|\zeta|=1$ by $f^*(\zeta)$. If λ_1^* denotes the lowest eigen value of C^*, it defines a root of the secular equation derived in (3.20):

$$\det \|\delta\lambda_1 \cdot \delta_{jk} + \pi(\lambda_1^2 - 1)\Re\{a_1\rho^2 \tilde{w}_1^{(j)}(z_0)\tilde{w}_1^{(k)}(z_0)\}\| = 0 \tag{4}$$

with $\delta\lambda_1 = \lambda_1^* - \lambda_1 + o(\rho^2)$. $\delta\lambda_1$ is the lowest root of (4); on the other hand, we conclude from the minimum property of C that

$$\delta\lambda_1 \geqq o(\rho^2) \tag{5}$$

and this holds, *a fortiori*, for all other roots of (4). Hence we can assert that the quadratic form

$$Q_\rho(t) = \sum_{j,k=1}^{n} \Re\{a_1\rho^2 \tilde{w}_1^{(j)}(z_0)\tilde{w}_1^{(k)}(z_0)\}\, t_j t_k \tag{6}$$

satisfies the inequality

$$Q_\rho(t) \leqq o(\rho^2) \tag{7}$$

for every choice of the unit vector $t_1, \cdots, t_n$. Dividing by ρ^2 and passing to the limit $\rho = 0$, we obtain

$$\Re\left\{a_1 \sum_{j,k=1}^{n} \tilde{w}_1^{(j)}(z_0)\tilde{w}_1^{(k)}(z_0) t_j t_k\right\} \leqq 0\,. \tag{8}$$

In particular, we obtain

$$\Re\{a_1 \tilde{w}(z_0)^2\} \leqq 0\,, \quad \tilde{w}(z_0) = \tilde{w}_1^{(1)}(z_0)\,. \tag{8'}$$

This inequality holds for every choice of the univalent variation function (1). We now apply the following theorem [11, 14]:

If for every point $z_0 \in \tilde{\Gamma}$ *and every univalent function* (1) *holds*

$$\Re\{a_1 s(z_0)\} \leqq 0 \tag{9}$$

where $s(z_0)$ *is regular analytic on* $\tilde{\Gamma}$, *then* $\tilde{\Gamma}$ *itself is an analytic curve* $z(t)$ *which satisfies the differential equation*

$$\left(\frac{dz}{dt}\right)^2 s[z(t)] = 1\,. \tag{10}$$

Hence we can deduce from (8′) that $\tilde{\Gamma}$ satisfies the differential equation

$$\left(\frac{dz}{dt}\right)^2 \tilde{w}[z(t)]^2 = 1\,. \tag{11}$$

In exactly the same way we prove that the extremum function $f(\zeta)$ maps the circumference $|\zeta| = r$ onto an analytic arc Γ which satisfies the differential equation

$$\left(\frac{dz}{dt}\right)^2 w[z(t)]^2 = 1 . \tag{12}$$

Let us put

$$z(\phi) = f(re^{i\phi}) ; \tag{13}$$

if ϕ runs from 0 to 2π the image point $z(\phi)$ will vary over Γ. We deduce from (12) the inequality

$$\zeta^2 f'(\zeta)^2 w[f(\zeta)]^2 < 0 \qquad \text{for } |\zeta| = r . \tag{14}$$

Similarly, we derive from (11) the inequality

$$\zeta^2 f'(\zeta)^2 \tilde{w}[f(\zeta)]^2 < 0 \qquad \text{for } |\zeta| = R . \tag{15}$$

We introduce the analytic functions

$$A(\zeta) = \zeta f'(\zeta) w[f(\zeta)] ; \quad B(\zeta) = \zeta f'(\zeta) \tilde{w}[f(\zeta)] . \tag{16}$$

Clearly, $A(\zeta)$ is regular analytic in the ring domain $r < |\zeta| < 1$ while $B(\zeta)$ is regular analytic for $1 < |\zeta| < R$. (14) and (15) can be expressed as

$$A(\zeta) = \text{imaginary for } |\zeta| = r \tag{14'}$$

$$B(\zeta) = \text{imaginary for } |\zeta| = R \tag{15'}$$

while equation (1.34) leads to

$$-iA(\zeta) = \frac{1}{\sqrt{\lambda_1^2 - 1}} [B(\zeta) + \lambda_1 \overline{B(\zeta)}] \qquad \text{for } |\zeta| = 1 . \tag{17}$$

We have by the Schwarz' reflection principle in view of (14′) and (15′):

$$\overline{A(\zeta)} = -A\left(\frac{r^2}{\bar{\zeta}}\right) , \quad \overline{B(\zeta)} = -B\left(\frac{R^2}{\bar{\zeta}}\right) . \tag{18}$$

Now we can rewrite (17) into the form

$$-iA(\zeta) = (\lambda_1^2 - 1)^{-1/2} [B(\zeta) - \lambda_1 B(R^2 \zeta)] \qquad \text{for } |\zeta| = 1 , \tag{19}$$

since $\bar{\zeta} = \zeta^{-1}$ for $|\zeta| = 1$. By (18) we see that $A(\zeta)$ is analytic in the ring $r^2 < |\zeta| < 1$ while $B(\zeta)$ is analytic for $1 < |\zeta| < R^2$. From (19) we can continue $B(\zeta)$ into the ring $k < |\zeta| < 1$ where $k = \max(r^2, R^{-2})$. By (18) again $B(\zeta)$ is, therefore, analytic in the ring $k < |\zeta| < \dfrac{R^2}{k}$ and by (19) we may continue $A(\zeta)$ beyond the unit circumference. Thus $A(\zeta)$ and $B(\zeta)$ are certainly analytic for $|\zeta| = 1$. The interrelation between $A(\zeta)$ and $B(\zeta)$ is, however, best understood by the use of Laurent series development.

We put

$$A(\zeta)=i\sum_{n=-\infty}^{\infty} a_n\zeta^n\,,\quad B(\zeta)=i\sum_{n=-\infty}^{\infty} b_n\zeta^n \tag{20}$$

and are sure that both series have a ring of common convergence which contains the unit circumference. The functional equations (18) are reflected in the coefficient relations

$$a_{-n}=\overline{a}_n r^{2n},\ b_{-n}=\overline{b}_n R^{2n}\,. \tag{21}$$

On the other hand, a comparison of coefficients in (19) yields

$$-ia_n=(\lambda_1^2-1)^{-1/2}(1-\lambda_1 R^{2n})b_n\,. \tag{22}$$

If we replace n by $-n$ and apply (21), we also find

$$ia_n=(\lambda_1^2-1)^{-1/2}(R^{2n}-\lambda_1)r^{-2n}b_n\,. \tag{23}$$

But (22) and (23) lead obviously to the alternative

$$a_n=b_n=0 \quad \text{or} \quad \lambda_1=\frac{r^{2n}+R^{2n}}{1+(rR)^{2n}}\,. \tag{24}$$

Thus $A(\zeta)$ and $B(\zeta)$ are necessarily rational functions and the possible values of λ_1 are restricted to the various values in (24) for integer n. Observe that $n=0$ is excluded since λ_1 is surely greater than one. It is sufficient to consider only positive values of n since $-n$ yields the same λ_1-value as $+n$. We may put equation (24) into the form

$$\frac{\lambda_1-1}{\lambda_1+1}=\frac{R^{2n}-1}{R^{2n}+1}\,\frac{1-r^{2n}}{1+r^{2n}}\,. \tag{25}$$

This form makes it evident that the minimum value of λ_1 for fixed r and R is attained for $n=1$. Hence, for the lowest eigen value λ_1 which belongs to a uniformly analytic curve C with the modulus (r, R), we have established the inequality:

$$\lambda_1\geqq\frac{r^2+R^2}{1+(rR)^2}\,. \tag{26}$$

In order to conclude the investigation we have to show that there exists, in fact, a curve C within the class considered for which equality is attained in (26). This curve can be found by a careful analysis of the variational conditions (11) and (12). At first we shall state the nature of an extremum curve C and compute its λ_1-value from its definition. Later we shall show that C is uniquely determined up to linear transformations.

Let us consider the z-plane slit along the linear segment $-i\mu$, $+i\mu$ of the imaginary axis and along the segments $|x|>1$ of the real axis. Every circular ring $r \leqq |\zeta| \leqq R$ can be mapped on such a canonical domain; the real parameter μ depends on the ratio $\frac{R}{r}$. For reasons of symmetry we can obtain that the points $\zeta=R$ and $\zeta=-R$ are mapped into $z=1$ and $z=-1$, respectively, while the points $\zeta=ir$ and $\zeta=-ir$ go into $i\mu$ and $-i\mu$. The mapping function $f(\zeta)$ has the symmetry properties:

$$\overline{f(\zeta)}=f(\bar{\zeta}), \quad -\overline{f(\zeta)}=f(-\bar{\zeta}), \tag{27}$$

and is uniquely defined. Let C be the image of the unit circumference $|\zeta|=1$ under the mapping $z=f(\zeta)$. We want to prove that C is the required extremum curve.

We denote again the interior and exterior of C by D and $\tilde{D}$, respectively. Observe that the functions

$$W_n(z)=A_n\left(\zeta^n+\frac{(-r^2)^n}{\zeta^n}\right), \qquad \zeta=f^{-1}(z) \tag{28}$$

are regular analytic in the entire domain D while the functions

$$\tilde{W}_n(z)=B_n\left(\zeta^n+\frac{R^{2n}}{\zeta^n}\right), \qquad \zeta=f^{-1}(z) \tag{29}$$

are regular analytic in $\tilde{D}$. Let us define the eigen functions of D and $\tilde{D}$ by

$$w_n(z)=\frac{d}{dz}W_n(z), \quad \tilde{w}_n(z)=\frac{d}{dz}\tilde{W}_n(z). \tag{30}$$

Differentiating (28) and (29) with respect to ζ, we find

$$w_n[f(\zeta)]f'(\zeta)\zeta=nA_n\left(\zeta^n-\frac{(-r^2)^n}{\zeta^n}\right) \tag{31}$$

and

$$\tilde{w}_n[f(\zeta)]f'(\zeta)\zeta=nB_n\left(\zeta^n-\frac{R^{2n}}{\zeta^n}\right). \tag{32}$$

The boundary conditions (1.34) for the eigen functions of D and $\tilde{D}$ will lead to the requirement

$$-iA_n\left(\zeta^n-\frac{(-r^2)^n}{\zeta^n}\right)=(\lambda_n^2-1)^{-1/2}\left[B_n\left(\zeta^n-\frac{R^{2n}}{\zeta^n}\right)+\lambda_n\bar{B}_n\left(\frac{1}{\zeta^n}-R^{2n}\zeta^n\right)\right] \tag{33}$$

for $|\zeta|=1$. This can indeed be fulfilled by satisfying the conditions

$$-iA_n(\lambda_n^2-1)^{1/2}=B_n-\lambda_n R^{2n}\overline{B}_n \tag{34}$$
$$-iA_n(-r^2)^n(\lambda_n^2-1)^{1/2}=B_nR^{2n}-\lambda_n\overline{B}_n$$

which is always possible if and only if

$$\lambda_n=\frac{R^{2n}-(-r^2)^n}{|1-(-r^2R^2)^n|}\ . \tag{35}$$

Conversely, it is evident that the values λ_n determined by (35) for $n=1, 2, \cdots$ lead to actual eigen functions for the domains D and $\tilde{D}$. Observe, in particular, that

$$\lambda_1=\frac{R^2+r^2}{1+r^2R^2} \tag{36}$$

which verifies that C is indeed an extremum curve and that our estimate (26) is the best possible one.

There remains finally the uniqueness question relative to the extremum curve C. In order to answer it we return to the functions $A(\zeta)$ and $B(\zeta)$ connected with the extremum function $f(\zeta)$. Since we know now that in their Laurent development all coefficients vanish except for a_1, a_{-1} and b_1, b_{-1}, we have by (16), (21) and (22)

$$\zeta f'(\zeta)w[f(\zeta)]=ia_1\left(\zeta+\frac{r^2}{\zeta}\right) \tag{36'}$$

and

$$\zeta f'(\zeta)\tilde{w}[f(\zeta)]=ib_1\left(\zeta-\frac{R^2}{\zeta}\right) \tag{37}$$

with

$$ia_1(\lambda_1^2-1)^{1/2}=(\lambda_1R^2-1)b_1\ . \tag{38}$$

We made the unessential assumption that a_1 is real which leads to the consequence that b_1 is pure imaginary.

We integrate (36′) and (37) and find

$$W[f(\zeta)]=ia_1\left(\zeta-\frac{r^2}{\zeta}\right),\quad \tilde{W}[f(\zeta)]=ib_1\left(\zeta+\frac{R^2}{\zeta}\right) \tag{39}$$

where $W(z)$ and $\tilde{W}(z)$ are properly chosen integrals of $w(z)$ and $\tilde{w}(z)$. The function $W(z)$ is single-valued in D; $f(\zeta)$ is regular analytic on $|\zeta|=r$ and can be continued somewhat beyond this circumference. It will take values near the continuum Γ after this continuation; but these values in the z-plane were already attained for some values ζ in $|\zeta|>r$. Hence $W[f(\zeta)]$ must take the same values for $|\zeta|$ somewhat larger than r and for some $|\zeta|$ less than r. From (39) we recognize that these corresponding ζ-values must be connected by the equation

$$\zeta_1-\frac{r^2}{\zeta_1}=\zeta_2-\frac{r^2}{\zeta_2}\,. \tag{40}$$

Hence we proved the functional equation for $f(\zeta)$:

$$f(\zeta)=f\left(-\frac{r^2}{\zeta}\right). \tag{41}$$

In exactly the same manner we derive from the second formula (39) the functional equation

$$f(\zeta)=f\left(\frac{R^2}{\zeta}\right). \tag{42}$$

We know already that the extremum function $f(\zeta)$ will remain an extremum function after a linear transformation since we showed at the end of § 1 that λ_1 does not change under linear transformations. Hence we may assume without loss of generality that

$$f(r)=0\,,\quad f(R)=1\,,\quad f(iR)=\infty\,. \tag{43}$$

From (41) and (42) conclude then that

$$f(-r)=0\,,\quad f(-iR)=\infty \tag{44}$$

and in view of the univalent character of $f(\zeta)$ in $r<|\zeta|<R$ we conclude that $f(\zeta)$ has simple zeros and simple poles at these points. It is now easy to obtain for $f(\zeta)$ a product representation in terms of its known zeros and poles in the entire ζ-plane and to identify it with the function which maps the ring $r<|\zeta|<R$ on the above described slit domain. This completes the uniqueness argument.

Let us return to the inequality (26). An important special case deals with all uniformly analytic curves with the modulus (r,∞). This is the class of curves which are images of $|\zeta|=1$ mapped by functions which are regular and univalent for $|\zeta|>r$. We find the estimate

$$\lambda_1\geqq r^{-2} \tag{45}$$

and the extremum curve in this case is the ellipse C which is obtained from $|\zeta|=1$ by the mapping

$$z=\zeta+\frac{r^2}{\zeta}\,. \tag{46}$$

This follows directly from (1.45) as well as from our preceding characterization of the extremum domain. The inequality (45) can also be easily derived from the estimate (3.29^{v}); thus this particular result could have been proved by means of a Hadamard type variational formula.

As for the class of uniformly analytic functions with the modulus $(0, R)$, we have analogously the estimate

$$\lambda_1 \geqq R^2 . \tag{47}$$

The extremal curve C is obtained from the unit circumference by the mapping

$$z = \frac{2R\zeta}{R^2+\zeta^2} . \tag{48}$$

This mapping is best understood if we consider the intermediate step

$$\eta = R^{-2}\zeta + \frac{1}{\zeta} \tag{49}$$

which maps the unit circumference onto an ellipse with $\lambda_1 = R^2$ and the circumference $|\zeta| = R$ onto the linear segment $\left\langle -\frac{2}{R}, \frac{2}{R} \right\rangle$. The additional linear transformation $z = \frac{2}{R\eta}$ does not affect the eigen values and leads to a regular univalent function in $|\zeta| < R$. We could have obtained the mapping (48) also as a special case of the preceding characterization of the extremum curve C.

6. Concluding remarks. We have restricted ourselves in the present paper to the case of simply connected domains. It is possible to extend a considerable amount of the results to the case of multiply-connected domains [3, 10, 14]. The investigation becomes, however, more complicated for two reasons. First, we will have a larger number of complementary domains and, second, we will have additional eigen functions belonging to the eigen value one. In fact, let $C_1, C_2, \cdots, C_n$ denote the n components of the boundary C of the domain D; let $\omega_\nu(z)$ be that harmonic function in D which takes on C_μ the boundary value $\delta_{\nu\mu}$. Then it is easily seen that

$$w_\nu(z) = i\frac{\partial}{\partial z}\omega_\nu(z) \tag{1}$$

will satisfy the integral equation

$$w_\nu(z) = \frac{1}{\pi}\iint_D \frac{\overline{w_\nu(z)}}{(\zeta - z)^2} d\tau . \tag{2}$$

All other eigen functions of the integral equation (1.13) belong, however, to eigen values which are larger than one.

The concept of the dielectric Green's function carries over to the case of higher multiplicity and analogous series developments in terms of the eigen functions of the Fredholm integral equation are possible. Likewise, the different variational formulas can be extended to multiple connectivity. But, clearly, it will be much more difficult to draw simple conclusions from these formulas. One has only to consider the great use made in the preceding section of Laurent series developments in order to appreciate the great simplification introduced by the assumption of a simply connected domain.

REFERENCES

1. L. V. Ahlfors, *Remarks on the Neumann-Poincaré integral equation*, Pacific J. Math., **2** (1952), 271–280.

2. ———, *Conformality with respect to Riemann metrics*, Ann. Acad. Sci. Fenn., Series A 202 (1955).

3. S. Bergman and M. Schiffer, *Kernel functions and conformal mapping*, Compositio Math., **8** (1951), 205–249.

4. G. Birkhoff, D. M. Young and E. H. Zarantanello, *Effective conformal transformation of smooth simply-connected domains*, Proc. Nat. Acad. Sci., **37** (1951), 411–414.

5. ———, *Numerical methods in conformal mapping*, Proc. Symposia in Appl. Math., **IV** (1953), 117–140.

6. G. F. Carrier, *On a conformal mapping technique*, Quart. Appl. Math., **5** (1947), 101–104.

7. S. Gershgorin, *On conformal mapping of a simply-connected region onto a circle*, Math. Sb. **40** (1933), 48–58.

8. K. Löwner, *Untersuchungen über schlichte konforme Abbildungen des Einheitskreises*, Math. Ann., **89** (1923), 103–121.

9. F. Rellich, *Störungstheorie der Spektralzerlegung, 1. Mitteilung*, Math. Ann., **113** (1937), 600–619.

10. H. L. Royden, *A modification of the Neumann-Poincaré method for multiply-connected regions*, Pacific J. Math., **2** (1952), 385–394.

11. M. Schiffer, *A method of variation within the family of simple functions*, Proc. London Math. Soc., **44** (1938), 432–449.

12. ———, *Variation of the Green's function and theory of the p-valued functions*, Amer. J. Math., **65** (1943), 341–360.

13. ———, *Hadamard's formula and variation of domain functions*, Amer. J. Math., **68** (1946), 417–448.

13. M. Schiffer and D. C. Spencer, *Functionals of finite Riemann surfaces*, Princeton (1954).

14. S. E. Warschawski, *On the effective determination of conformal maps*, Contribution to the Theory of Riemann Surfaces, Princeton (1953).

STANFORD UNIVERSITY

A THREE POINT CONVEXITY PROPERTY

F. A. VALENTINE

There exist an interesting variety of set properties determined by placing restrictions on each triple of points of the set. It is the purpose here to study those closed sets in the n-dimensional Euclidean space E_n (in particular the plane E_2) which satisfy the following condition.

DEFINITION 1. A set S in E_n is said to possess the *three-point convexity property* P_3 if for each triple of points x, y, z in S at least one of the closed segments xy, yz, xz is in S.

The principal result obtained in this paper appears in Theorem 2. In order to achieve this result a series of lemmas and theorems is first established. Most of these are also of independent interest.

1. Closed connected sets in E_n, $n \geqq 1$.

In this section we assume that S is a closed connected set in E_n, $n \geqq 1$. The concept of *local convexity* is a useful one for our purpose, so we restate the well-known definition.

DEFINITION 2. A set S is said to be *locally convex at a point* $q \in S$ if there exists an open sphere N with center at q such that $S \cdot N$ is convex. If a set is locally convex at each of its points, it is said to be *locally convex*.

NOTATION 1. The open segment determined by points x and y is denoted by (xy), whereas xy denotes the closed segment. The line determined by x and y is denoted by $L(x, y)$. The boundary of a set S is $B(S)$, and $H(S)$ denotes the closed convex hull of S. The symbol $+$ stands for set union, and the symbol $\cdot$ stands for set product.

THEOREM 1. *Let S be a closed connected set in E_n $(n \geqq 1)$ which has property P_3. Then either S is convex or S is starlike with respect to each of its points of local nonconvexity (It may be starlike elsewhere).*

Proof. If S is locally convex, then by a theorem of Tietze [4, pp. 697–707], [2, pp. 448–449], the set S is convex, in which case it is starlike with respect to each of its points. Hence, suppose S is not locally convex, and let $q \in S$ be a point of local nonconvexity. This implies that in each spherical neighborhood N_i of q, there exist points

Received June 26, 1956.

x_i and y_i of S such that $(x_iy_i)\cdot S=0$ (see Notation 1). Choose *any* point $x\in S$. Property P_3 implies that either xy_i or xx_i is in S. If the radius of N_i is $1/i$, then as $i\rightarrow\infty$ the set xx_i+xy_i converges to qx, which then must belong to S. This completes the proof.

REMARK 1. *The set of all starlike points of a set S is called the convex kernel of S. The convex kernel of a set $S\subset E_n$ is convex. See Brunn* [1].

COROLLARY 1. *Each point of local nonconvexity of the set S in Theorem* 1 *is contained in the boundary of the convex kernel of S.*

COROLLARY 2. *For the set S above, let H be any r-dimensional plane section of S, where* $(1\leqq r\leqq n-1)$. *Then either $H\cdot S$ is starlike or $H\cdot S$ consists of two convex components.*

Proof of Corollary 2. If $H\cdot S$ is connected, then since $H\cdot S$ has property P_3, Theorem 1 implies $H\cdot S$ is starlike. If $H\cdot S$ is not connected, property P_3 implies trivially that $H\cdot S$ consists of two and only two components, each of which must be convex.

COROLLARY 3. *Each component of the complement of S is unbounded.* This is an immediate consequence of the starlikeness of S.

2. Closed connected sets in E_2. We restrict ourselves to closed connected sets in E_2 in this section, and the following definitions are useful.

DEFINITION 3. A component of the complement of a closed connected set S is called a *residual domain of* S. A *cross-cut xy of a residual domain K of S* is a closed segment such that $x\in S$, $y\in S$ and $(xy)\subset K$ (See Notation 1).

DEFINITION 4. An isolated point of local nonconvexity of S is called a p-point. A point of S which is a p-point or a limit point of p-points is called a *q-point.*

LEMMA 1. *Each open segment (uv) of the convex kernel of S contains no q-points of S.* (see Corollary 1).

Proof. Suppose w is a q-point contained in (uv). Clearly $S\not\subset L(u, v)$ (see Notation 1). Choose $z\in S-L(u, v)$. Since uv belongs to the convex kernel of S, we have triangle $uzv\subset S$. But this implies that each sufficiently small neighborhood of w contains no cross-cuts of the complement of S, since such a cross cut xy would have to have its interior (xy) in one of the open half-planes bounded by $L(u, v)$.

LEMMA 2. *Let S be a closed connected set in E_2 having property P_3. Then if S is not convex, it contains at least one isolated point of local nonconvexity.*

Proof. Let xy be a cross-cut of a residual domain K of S. Since S is closed and connected, the set $K-(xy)$ is the union of two mutually exclusive open sets, denoted by K_1 and K_2 [3, p. 118]. Since S is star-like, Corollary 3 implies that one and only one of these two sets is bounded. Let it be K_1, and denote the boundary $B(K_1)-(xy)\equiv C(K_1)$. The set $B(K_1)$ is a continuum [3, p. 124]. Since K_1 is a bounded domain, and since $C(K_1)\cdot(xy)=0$, it follows that $C(K_1)$ is a continuum. Define B_1 to be the set of points $z_1\in C(K_1)$ such that $xz_1\subset S$, and define B_2 to be the points $z_2\in C(K_1)$ such that $yz_2\subset S$. Since $xy\not\subset S$, property P_3 implies $C(K_1)=B_1+B_2$. Moreover, $x\in B_1$, $y\in B_2$, and moreover B_1 and B_2 are each closed since S and $C(K_1)$ are closed. Hence, since $C(K_1)$ is a continuum, it is well-known that $B_1\cdot B_2\neq 0$. Hence, let $p\in B_1\cdot B_2$, and we must have $xp\subset S$, $yp\subset S$, so that K_1 is interior to triangle xpy. Since $p\in B(K_1)$, it is clear that each neighborhood of p contains a cross-cut of K_1, so that p is a point of local nonconvexity of S.

To prove that p is an *isolated* point of local nonconvexity, observe that the lines $L(x, p)$ and $L(y, p)$ determine four V-shaped domains, each bounded by two rays. Order these V_1, V_2, V_3, V_4 so that $xyp\subset \overline{V}_1$, and so that the sets V_i are arranged consecutively in a clockwise direction about the point p. Suppose a p-point $p_1\in \overline{V}_1-p$ exists. Then since $p_1x+p_1y\subset S$, we would have $K_1\subset xyp_1$, which would violate the fact $p\in B(K_1)$. Suppose a p-point, say p_1 exists in V_2. But this implies that $xpp_1\subset S$, $ypp_1\subset S$. But this again would violate the fact $p\in B(K_1)$. In exactly the same way V_4 contains no p-point of S. Now consider V_3. If p_1 is a p-point of S in V_3, then $ypp_1+xpp_1\subset S$, which implies that p is an isolated p-point since V_1 contains no p-point of S. Finally, Lemma 1 implies that no sequence of p-points of S can exist on $L(x, p)\cdot \overline{V}_3$ or on $L(y, p)\cdot \overline{V}_3$ having p as a limit point. Thus we have shown that p is an isolated p-point of S.

REMARK 2. Let xy be the cross-cut in the above proof, and let p be the associated isolated p-point. Then the closed triangle xyp is such that the set $xyp\cdot S$ is the union of two convex sets having only the point p in common. One of these convex sets contains xp and is denoted by $C(xp)$, and the other denoted by $C(yp)$ contains py.

Proof. Let L_i and L be lines parallel to xy, such that L_i separates p and xy, and such that $p \in L$. Let H_i be the closed half-plane bounded by L_i and containing xy. Suppose $L_i \to L$ as $i \to \infty$ so that $H_{i+1} \supset H_i$. Since $S \cdot H_i \cdot xyp$ is locally convex, by Tietze's Theorem [4, loc. cit.] each of its components is convex. Property P_3 implies that there are at most two such components. The fact that $xy \not\subset S$, implies there are exactly two such components. Denote them by C_i and D_i. Clearly $C_{i+1} \supset C_i$, $D_{i+1} \supset D_i$, and hence C_i and D_i converge to convex sets having p in common. They have only p in common, otherwise p would not be a boundary point of K_1 as defined in the proof of Lemma 2. One of the convex sets contains xp and the other yp so that the notation in the remark is justified.

DEFINITION 5. Let Q be the set of q-points of S.

REMARK 3. Corollary 1 implies that Q is contained in the boundary of its own convex hull $H(Q)$, designated by $B(H)$.

LEMMA 3. *The boundary of $H(Q)$ is connected, and it can contain at most one ray.*

Proof. Since $H \equiv H(Q)$ is convex, if $B(H)$ were not connected, it would have to consist of two parallel lines (this is known). However, Lemma 1 would then imply that each of these parallel lines would contain at most two q-points. But this would imply that Q is bounded in which case $B(H)$ would be connected. If $B(H)$ contained two rays, then Lemma 1 would again imply that Q is bounded, which would again be contradictory.

DEFINITION 6. An *edge* of the boundary $B(H)$ is a closed segment xy or a closed ray $x\infty$ whose endpoints are q-points. An open half-plane whose boundary contains xy(*or* $x\infty$), and which does not intersect $H(Q)$ is called an *open half-plane of support*, and it is denoted by W.

LEMMA 4. *Let W be an open half-plane of support to $H(Q)$, which abuts on an edge xy(or $x\infty$). Then $H(Q)+W \cdot S$ is a convex subset of S.*

Proof. If $u \in H(Q)$ and if $v \in S \cdot W$, then $uv \subset S$, since S is starlike with respect to u. This, together with the facts $x \in B(S)$, $y \in B(S)$, and property P_3 imply that $uv \cdot xy \neq 0$ (or $uv \cdot x\infty \neq 0$), so that $uv \subset H(Q)+W \cdot S$. Suppose $u \in S \cdot W$, $v \in S \cdot W$. Let $z \in (xy)$ or $(x\infty)$. If $(uv) \not\subset S$, since $uz \subset S$, $vz \subset S$, then triangle uvz would contain a p-point of S (See the first paragraph of the proof of Lemma 2). But this is impossible, since W contains no p-points of S, and since by Lemma 1 the open segment (xy) or $(x\infty)$ contains no p-points of S. Hence $H(Q)+$

$W \cdot S$ is convex. It should be observed that if $H(Q) \equiv xy$, then $H(Q) + W \cdot S$ may or may not be closed.

LEMMA 5. *Let $x_i y_i$ be a countable number of pairwise disjoint edges in $B(H) \equiv B(H(Q))$. Assume that $B(H)$ contains at least three edges, and let W_i be the open half-plane of support to $H(Q)$ whose boundary contains $(x_i y_i)$ ($x_i y_i$ may be $x_i \infty$).*

Then the set $H(Q) + S \cdot \Sigma_i W_i$ is a closed convex set.

Proof. Without loss of generality establish an order on the boundary $B(H)$, and assume that in terms of this order, x_i is the beginning of the edge $x_i y_i$ and that y_i is the endpoint of $x_i y_i$. Select any two disjoint edges $x_i y_i$ and $x_j y_j$, and without loss of generality assume that x_i, y_i, x_j, y_j fall in an order so that an arc of $B(H)$ has x_i and y_j as its endpoints, and so that all four points lie on this arc in the order given above ($B(H)$ may be unbounded). Let the convex set which is bounded by the two lines $L(x_i, y_j)$ and $L(x_j, y_i)$, and which contains the quadrilateral $x_i y_i x_j y_j$ be denoted by V. The segments $x_i y_i$ and $x_j y_j$ divide V into three parts; one is the closed quadrilateral $x_i y_i x_j y_j$; the second is a three sided closed polygonal set adjacent to $x_i y_i$ and denoted by $\Delta(x_i, y_i)$; the third is a three sided closed polygonal set adjacent to $x_j y_j$ and denoted by $\Delta(x_j, y_j)$. The last two sets may or may not be bounded. If the edge $x_j y_j$ is a ray $x_j \infty$ instead, then the same type of division occurs, in which $L(x_i, \infty)$ is a line parallel to the ray $x_j \infty$ so that $\Delta(x_j, \infty)$ has two bounding sides instead of three. *We must have* $S \cdot W_i \subset \Delta(x_i, y_i)$, for if this were not so, it is easily seen that either x_i or y_i would be an interior point of a triangle which would belong to S. But this would contradict the fact that $x_i \in B(S)$, $y_i \in B(S)$. Similarly $S \cdot W_j \subset \Delta(x_j, y_j)$. This is true whether $x_j y_j$ is a finite segment or a ray $x_j \infty$.

Now, choose two points u and v in $U \equiv H(Q) + S \cdot \Sigma_i W_i$. If u and v are in $H(Q) + S \cdot W_i$, then Lemma 4 implies $uv \subset U$. If $u \in S \cdot W_i$ and $v \in S \cdot W_j$, then by the preceding paragraph $u \in \Delta(x_i, y_i)$, $v \in \Delta(x_j, y_j)$ (or $v \in \Delta(x_j, \infty)$). Since $V \equiv \Delta(x_i, y_i) + x_i y_i x_j y_j + \Delta(x_j, y_j)$ is convex, and since $\Delta(x_i, y_i) \cdot x_i y_i x_j y_j = x_i y_i$, we have $uv \cdot x_i y_i \neq 0$, whence $uv \subset U$.

To prove that U is closed, observe first that if there are a *finite* number of disjoint sets $x_i y_i$ (there are at least three edges) then U is closed, since $\overline{W}_i \cdot S \subset \Delta(x_i, y_i)$ implies $\overline{W_i \cdot S} \subset W_i \cdot S + B(H)$. If there are an infinite number of sets W_i, then let s be a limit point of an infinite sequence of sets $W_{i_n} \cdot S$. Since $W_{i_n} \subset \Delta(x_{i_n}, y_{i_n})$ by fixing (x_j, y_j) of the preceding paragraph, it follows that $(x_{i_n}, y_{i_n}) \to q$, a fixed point of $B(H)$, as $i_n \leftarrow \infty$. However, since in this situation, we must have $\Delta(x_{i_n}, y_{i_n}) \to q$ as $i_n \to \infty$, it follows that $s = q \in H(Q)$. Hence, it is clear that U is closed, since $H(Q)$ is closed.

THEOREM 2. *Suppose S is a closed connected set in E_2 such that for each triple of points x, y, z in S at least one of the segments xy, yz, xz is in S.*

Then S is expressible as the union of three or fewer closed convex sets having a nonempty intersection. The number three is best.

DEFINITION 7. Let N denote the cardinality of the set of p-points of S *in Theorem* 2.

THEOREM 3. *If N is not an odd integer greater than* 1, *then S can be expressed as the union of two or fewer closed convex sets having a nonempty intersection.*

Proofs of Theorems 2 *and* 3. We recall that Q is the closure of the set of p-points of S. The proof is divided into 5 cases, depending upon the value of N. The five cases are: $N=1$; $N=2$; $N=2m>1$; $N=2m+1>1$; $N=\infty$.

Case 1. $N=1$. Let $Q=p$, and let C be a circle with center at p and having radius r. The set $S\cdot C$ is a closed connected set having property P_3 and having p as its only p-point. If $S\cdot C$ satisfies the conclusions of either Theorem 2 or Theorem 3, it is quite clear that $S=\lim S\cdot C$ as $r\to\infty$ will satisfy the same conclusions. Let the boundary of the convex hull $H(S\cdot C)$ be $D(H)$, since $B(H)$ stands for the boundary of $H(Q)$. The rest of the proof will show incidentally that $D(H)\cdot S$ has one, two or four components.

Suppose $D(H)\cdot S$ has exactly three components and designate these by B_i $(i=1, 2, 3)$. It is easy to show that $B_i\neq\{p\}$ $(i=1, 2, 3)$. Choose points $x_i\in B_i$ with $x_i\neq p$ $(i=1, 2, 3)$. Property P_3 implies that at least one of the intervals x_1x_2, x_2x_3, x_3x_1 is in S. Suppose $x_1x_2\subset S$. Since $x_i\in D(H)$ $(i=1, 2)$, and since $B_1\cdot B_2=0$, we must have $L(x_1, x_2)\cdot S=x_1x_2$. If $p\notin(x_1x_2)$, then let H_{12} be the closed half-plane bounded by $L(x_1, x_2)$ and not containing p. Since $x_1\in B_1$, $x_2\in B_2$ with $B_1\cdot B_2=0$, there must exist a cross cut of the complement of S in H_{12}. However, by the proof of Lemma 2, there would exist a p-point in $H_{12}\cdot S$ which contradicts the fact that $Q=p$. Hence $p\in(x_1x_2)$. Since $p\in(x_1x_2)$, if $x_1x_3\subset S$, the proof of Lemma 2 would again imply the existence of a p-point in that closed half-space bounded by $L(x_1, x_3)$ which does not contain p. However this contradicts $Q=p$. Hence, $x_1x_3\not\subset S$. Similarly $x_2x_3\not\subset S$. Property P_3 and the closure of S implies that for points $x\in B_1$ sufficiently near x_1, we have $xx_2\subset S$, $xx_3\not\subset S$, $x_2x_3\not\subset S$. Applying the same reasoning to x, x_2, x_3 that we applied to x_1, x_2, x_3, we get $p\in(xx_2)$ for all x near

x_1. This can only be true if $B_1=x_1$. Similarly, $B_2=x_2$. However, since B_3 is contained in only one of the open half planes bounded by $L(x_1, x_2)$, the facts $B_i=x_i$ $(i=1, 2)$ simply that $x_1x_2 \subset D(H)\cdot S$, which contradicts the fact $B_1\cdot B_2=0$. Hence *$D(H)\cdot S$ cannot have exactly three components.* Suppose $D(H)\cdot S$ has at least four components, and designate four of these by B_i $(i=1, 2, 3, 4)$. The above argument implies that $B_i=x_i \in D(H)\cdot S$, and these can be renumbered so that $p\in(x_1x_2)$, $p\in(x_3x_4)$. Clearly any fifth component B_5 could not exist, since the above argument applied to x_1, x_2, x_3 and $x_5\in B_5$ would yield $p\in(x_1x_2)$, $p\in(x_3x_5)$, so that $x_5=x_4$, a contradiction. *Thus if $D(H)\cdot S$ has more than two components, then $S\cdot C$ is the union of two line segments having an interior point in common.*

Now, suppose $D(H)\cdot S$ has exactly two components denoted by B_1 and B_2. Let the end points of B_i be x_i and y_i ordered so that y_1x_2 and y_2x_1 are cross-cuts of the complement of S. The points x_i and y_i need not be distinct. We will prove that each of the sets $P_i\equiv H(B_i+p)+C(x_ip)+C(y_ip)$, $(i=1, 2)$ is convex. (See Remark 2 following Lemma 2 for the definitions of $C(x_ip)$ and $C(y_ip)$). Property P_3 and the fact that $D(H)\cdot S=B_1+B_2$ implies that $B_i+x_ip+y_ip$ is the boundary of $H(B_i+p)$. We will prove that P_1 is convex. Since each of the sets $H(B_1+p)$, $C(x_1p)$, $C(y_1p)$ is convex, to show that P_1 is convex, it suffices to select points $z\in H(B_1+p)$, $u\in C(x_1p)$, $v\in C(y_1p)$, and to show that $uv+uz+vz \subset P_1$. We must have $uv\cdot x_1p\neq 0$, $uv\cdot y_1p\neq 0$, for if this were not so, the fact $D(H)\cdot S=B_1+B_2$ would imply that $uv\cdot P_2\neq 0$. However, this would contradict property P_3. Hence, we have $uv\subset x_1y_1p+C(x_1p)+C(y_1p)$. Thus $uv\subset P_1$. In the same manner $uz\subset P_1$, $vz\subset P_1$, so that P_1 is convex. The same argument applies to P_2.

Finally if $D(H)\cdot S$ has exactly one component, and $Q=p$, it can be shown readily that there exists a line through p which divides $S\cdot C$ into two closed convex sets having p in common. This completes the proof for $N=1$, and oddly enough it appears to be the most difficult to prove.

Case 2. $N=2$. Let $Q=p_1+p_2$. The line $L(p_1, p_2)$ divides the plane into two open half-planes W_i $(i=1, 2)$. Lemma 4 implies that $W_i\cdot S$ is convex. If $W_1\cdot S=0$, then $S=\overline{W}_2\cdot S+S\cdot L(p_1, p_2)$ yields the desired conclusions of Theorem 2 and 3. Hence suppose $W_i\cdot S\neq 0$ $(i=1, 2)$. Let $U\equiv\overline{W_1\cdot S}+\overline{W_2\cdot S}$. If U is convex, then $S\equiv U+S\cdot L(p_1, p_2)$ yields the desired decomposition. Suppose U is not convex, then we can show that $S\cdot L(p_1, p_2)=U\cdot L(p_1, p_2)$, for suppose a point $u\in S\cdot L(p_1, p_2)-U\cdot L(p_1, p_2)$ exists. Since U is not convex, there exist points $x_i\in W_i\cdot S$ such that $x_1x_2\not\subset S$. Moreover $ux_i\not\subset S$, since $u\notin\overline{W_i\cdot S}$. However, this violates property P_3. Thus if U is not convex, $S=\overline{W_1\cdot S}+\overline{W_2\cdot S}$, and this is a desired decomposition of S into two convex sets.

Case 3. $N=2m>2$. In this case the hull $H(Q)$ is a convex polygon, each segment of which is an edge having p-points as endpoints (See definition 6). Order the edges x_ix_{i+1} of the boundary $B(H)$ counterclockwise so that $(i=1, 2, \cdots, 2m; x_1=x_{2m+1})$. The open half-plane of support to $H(Q)$ adjacent to x_ix_{i+1} is denoted by W_i. By Lemma 5 each of the sets

$$S_1 \equiv H(Q)+S\cdot\sum_{i=1}^{m} W_{2i-1}$$
(1)
$$S_2 \equiv H(Q)+S\cdot\sum_{i=1}^{m} W_{2i}$$

is a closed convex set. Moreover, since $S\subset H(Q)+\sum_{i=1}^{2m} W_i$ we have $S_1+S_2=S$.

Case 4. $N=2m+1>1$. As in Case 3, let $e_i\equiv x_ix_{i+1}$ $(i=1, \cdots, 2m+1; x_1=x_{2m+2})$ denote the ordered edges of $B(H)$, and define S_1 and S_2 as in (1).

Let

$$S_3\equiv H(Q)+S\cdot W_{2m+1}\ .$$

By Lemma 5, the sets S_1, S_2 and S_3 satisfy the conclusions of Theorem 2.

Case 5. $N=\infty$. In order to prove this case, the following definition is helpful.

DEFINITION 8. A connected closed subset I of the boundary $B(H)$ is called a *polygonal element* if the following conditions hold:

(a) It is the closure of the union of edges of $B(H)$ (see Definition 6).

(b) Its endpoints (one, two or none) are limit points of p-points of S.

(c) If $I=B(H)$, then I contains at most one limit point of p-points. If $I\neq B(H)$, then only its endpoints (one or two) are limit points of p-points.

Observe that these conditions imply that a polygonal element is maximal in the sense that it is not a proper subset of a larger polygonal element.

The number of polygonal elements of $B(H)$ is countable, hence we can well-order them easily. Let $I_1, I_2, \cdots, I_n \cdots$ designate such a well-ordering.

For each polygonal element I_n, divide the edges it contains (see Definition 6) into two classes M_n^1 and M_n^2 such that no two edges of M_n^i $(i=1, 2)$ are adjacent, that is, have an endpoint in common. It may happen that one of the M_n^i may be empty. For each edge $e \in M_n^i$ we let W_e^i denote the open half-plane of support to $B(H)$ whose boundary contains e. Define

$$F_n^i = \sum_{e \in M_n^i} W_e^i \cdot S \qquad (i=1, 2),$$

and let

$$S_i = H(Q) + \sum_n F_n^i \qquad (i=1, 2).$$

Since each edge in M_n^i is separated from each edge in M_m^i $(n \neq m)$, Lemma 5 implies that S_1 and S_2 are closed convex subsets of S. Moreover, since for each point $x \in S$, either $x \in H(Q)$, or x is contained in some $W_e^i \cdot S$, we have $S = S_1 + S_2$ and $S_1 \cdot S_2 \neq 0$.

To prove that the number "three" in Theorem 2 is best consider the familiar two cell formed by a five-pointed star. It is a simple matter to verify that this set has property P_3, and that it cannot be expressed as the union of two convex sets. The analogous $2m+1$ pointed star behaves the same way.

3. Concluding remarks.

(a) It should be noted that the converse of Theorem 2 is not true. For instance, the set consisting of three segments xx_i $(i=1, 2, 3)$, where each angle $\angle x_i x x_j = 120°$ $(i \neq j)$, is the union of three convex sets; yet it does not have property P_3.

(b) It would be of interest to characterize those sets in E_2 which are the union of two closed convex sets. It appears that such a characterization will follow from an investigation of the cardinality of the set $B(K) \cdot B(S)$, where K is the convex kernel of S.

(c) The theory in E_3 needs to be settled. In view of § 1, it is natural to ask the question. What are the closed connected sets in E_3 such that each of its plane sections is either starlike or the union of two disjoint convex sets?

References

1. H. Brunn, *Über Kerneigebiete*, Math. Ann. **73** (1913), 436–440.
2. V. Klee, *Convex sets in linear spaces*, Duke Math. J., **18**, No. 2 (1951), 443–466.
3. M. H. A. Newman, *Topology of plane sets of points*, second edition (reprinted 1954), Cambridge University Press.
4. H. Tietze, *Über Konvexheit im kleinen und im grossen und über gewisse den Punkten einer Menge zugeordnete Dimensionszahlen*, Math. Z. **28** (1929), 697–707.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

THE CENTER OF A COMPACT LATTICE IS TOTALLY DISCONNECTED

ALEXANDER DONIPHAN WALLACE

The purpose of this note is to prove the theorem of the title. A topological lattice is a Hausdorff space together with a pair of continuous functions $\wedge: L \times L \to L$, $\vee: L \times L \to L$ satisfying the usual conditions for lattice operations. As is customary we may write $x \wedge y$ in place of $\wedge(x, y)$. All references are to Chapter II of [1]. We assume the reader to be familiar with the elementary facts concerning topological algebras (groups, lattices, semigroups) and set-theoretic topology.

THEOREM. *The center of a compact lattice is totally disconnected.*

Proof. Let L be a compact lattice. As is wellknown L has a zero and a unit, 0 and 1. If A is the set of pairs $(x, y) \in L \times L$ such that $x \wedge y = 0$ and $x \vee y = 1$ then $A = \wedge^{-1}(0) \cap \vee^{-1}(1)$ so that A is closed. The projection $(x, y) \to x$ takes A onto the closed set B and B is the set of all $x \in L$ which admit a complement.

Now N, the set of neutral elements of L, is the intersection of the maximal distributive sublattices by Theorem 11. But if D is a distributive sublattice of L its closure is also a distributive sublattice. It follows that N is closed. By the corollary to Theorem 10 the center C of L is $N \cap B$ so that C is closed.

By the lemma on page 27 each element $x \in C$ has a *unique* complement $k(x) \in C$. We will show that $k: C \to C$ is continuous. If G is the subset of $C \times C$ consisting of all $(x, k(x))$ with $x \in C$ it is enough to show that G is closed since C is compact. But by the remarks above we have $G = (C \times C) \cap \wedge^{-1}(0) \cap \vee^{-1}(1)$.

Now C is a distributive lattice (Theorem 9 and Corollary p. 29) with unique complements. Thus C is a commutative *topological* group under the operations

$$x + y = (x \wedge k(y)) \vee (k(x) \wedge y), \quad -x = x$$

all of whose elements are of order 2, that is, $x + x = 0$ for all x. If Q is the component of C containing 0 and if $q \in Q$, $q \neq 0$, then there is a continuous homomorphism f taking Q into Z, the reals mod 1, such that $f(q) \neq f(0)$. Since $f(Q)$ is connected it contains an interval of Z and therefore contains an element not of finite order. Since the order of each element of Q is two this is a contradiction. Hence Q contains

Received August 6, 1956.

only 0 and therefore is totally disconnected. The proof of the Theorem is complete.

Reference

1. G. Birkhoff, *Lattice theory*, New York, 1948.

The Tulane University of Louisiana

TWO THEOREMS ON TOPOLOGICAL LATTICES

ALEXANDER DONIPHAN WALLACE

A *topological lattice* is a pair of continuous functions

$$\wedge: L\times L \rightarrow L, \quad \wedge: L\times L \rightarrow L$$

(L a Hausdorff space) satisfying the usual conditions for lattice operations. A set A is *convex* if $x, y \in A$ and $x \leqq a \leqq y$ implies $a \in A$. This is equivalent to $A=(A \wedge L) \cap (A \vee L)$.

After proving a separation theorem involving a convex set we show that a compact connected topological lattice is a cyclic chain in the sense of G. T. Whyburn and that each cyclic element is a convex sublattice. In doing so we rely on some results recently obtained by L. W. Anderson.

THEOREM 1. *Let L be a connected topological lattice and let A be a convex set such that $L\diagdown A$ is not connected. Then $L\diagdown A$ is the union of the connected separated sets $(A \wedge L)\diagdown A$ and $(A \vee L)\diagdown A$ which are open (closed) if A is closed (open). If L is also compact then A is connected if it is either open or closed.*

Proof. Let $L\diagdown A=U \cup V$ with $U^* \cap V=\phi=U \cap V^*$ and let $p \in U$, $q \in V$. The connected set $(p \wedge L) \cup (q \wedge L)$ meets both U and V; hence it meets A. Adjust the notation so that $(q \wedge L) \cap A \neq \phi$ and thus $q \in A \vee L$. If $(q \vee L) \cap A \neq \phi$ then $q \in A \wedge L$ and hence $q \in (A \wedge L) \cap (A \vee L)=A$. This being impossible we infer that $(q \vee L) \cap A=\phi$ and $q \in (A \vee L)\diagdown A=(A \vee L)\diagdown(A \wedge L)$. The connected set $(p \vee L) \cup (q \vee L)$ intersects U and V and so intersects A. But $(q \vee L) \cap A=\phi$ so that $(p \vee L) \cap A \neq \phi$ and hence $p \in A \wedge L$. Were $(p \wedge L) \cap A \neq \phi$ we would also have $p \in A \vee L$ and so $p \in A$, a contradiction. Thus $(p \wedge L) \cap A=\phi$ and $p \in (A \vee L)\diagdown A=(A \vee L)\diagdown(A \wedge L)$. Now take $y \in V$ and suppose that y is not in $A \vee L$ so that $(y \wedge L) \cap A=\phi$; then $(p \wedge L) \cap A \neq \phi$ since $(p \wedge L) \cup (y \wedge L)$ is a connected set meeting U and V. But this is contrary to the proven fact that $(p \wedge L) \cap A=\phi$. We conclude that $V \subset (A \vee L)\diagdown A$ and, dually, that $U \subset (A \wedge L)\diagdown A$. It follows that $L=(A \wedge L) \cup (A \vee L)$. Now $x \in (A \vee L)\diagdown A$ and $x \in L\diagdown V$ gives $x \in U \subset (A \wedge L)\diagdown A$ and this contradicts the convexity of A. Hence $U=(A \wedge L)\diagdown A$ and $V=(A \vee L)\diagdown A$. To see that $U \wedge L=U$ we need only note that $x \in U$ gives $(x \wedge L) \cap A=\phi$ and thus $(x \wedge L) \cap V=\phi$ (since $x \wedge L$ is connected and contains x) and hence $x \wedge L \subset (A \wedge L)\diagdown(A \vee L)=U$.

Received August 6, 1956. This work was supported by the National Science Foundation.

Dually, $V \wedge L = V$ and these equalities imply that U and V are connected. If A is closed (open) then U and V are open (closed). This completes the proof of the first sentence of the conclusion. If L is also compact then $H^1(L)=0$ [3] so that (as is well known) L is unicoherent. But L is locally connected, $L=(A \wedge L) \cup (A \vee L)$, and the sets $A \wedge L$ and $A \vee L$ are connected, and open (closed) [1] if A is open (closed). Hence by a known result [2] we see that $A=(A \wedge L) \cap (A \vee L)$ is connected.

We assume that the reader is familiar with the cyclic element theory of locally connected continua as given in [4]. We recall that a locally compact connected topological lattice is locally connected [1].

THEOREM 2. *Let L be a compact connected metrizable topological lattice. Then L is a cyclic chain, each cyclic element of which is a convex sublattice. If L is topologically contained in the plane then each true cyclic element of L is 2-cell and L has the fixed-point property.*

Proof. Let C be a true cyclic element of L, let $x, y \in C$ with $x \leqq y$ and let $p \in L$ such that $x \leqq p \leqq y$. If T is a maximal chain containing x, p, and y then T is an arc from 0 to 1, as is well known [1]. Hence the set $[x, y] = \{t \mid t \in T \text{ and } x \leqq t \leqq y\}$ is an arc from x to y [1]. Since C is an A-set [4] we know that $[x, y] \subset C$ and thus $p \in C$. Hence C is convex. Let D be the cyclic chain from 0 to 1, that is, D is the smallest A-set containing 0 and 1 [4]. Then, by definition, $T \subset D$ and if $x \in L \setminus D$ then the maximal chain T' containing 0, x, 1 is an arc from 0 to 1 and thus $T' \subset D$, a contradiction. Hence $D=L$ and L is the cyclic chain from 0 to 1. Let T_0 be 0, 1 and all points which separate 0 and 1. Then L is the union of T_0 and all true cyclic elements meeting T_0 in two points [4]. Suppose that the true cyclic element C meets T_0 in the cutpoints p and q. Note that neither 0 nor 1 is a cutpoint [3]. If z is a cutpoint then, since $\{z\}$ is convex, $L = (z \wedge L) \cup (z \vee L)$ and thus z is comparable with each $x \in L$, by Theorem 1. We may assume that $p < q$. We will show that $C = \{x \mid p \leqq x \leqq q\}$. The convexity of C proves the containment "$\supset$". If $x \in C$ and if, say, $x \leqq q$ is false then we have $q < x$. By Theorem 1, $L \setminus q = ((q \wedge L) \setminus q) \cup ((q \vee L) \setminus q)$ is a separation and C meets both members, contrary to the fact that C is a true cyclic element [4] . Dually, $x \leqq p$ cannot be false, proving the containment "$\subset$" of the desired equality. It follows that C is a convex sublattice. The cases $p=0$ or $q=1$ are treated similarly. The remaining results follow from the fact that $H^1(L)=0$ [3] so that L is a locally connected continuum [1] which does not cut the plane [4].

References

1. L. W. Anderson, *Topological lattices*, Tulane University dissertation (1956).

2. A. H. Stone, *Incidence relations in unicoherent spaces*, Trans. Amer. Math. Soc. **65** (1949) 427–447.

3. A. D. Wallace, *Cohomology, dimension and mobs*, Summa Bras. Math. **3** (1953) 43–54.

4. G. T. Whyburn, *Analytic topology*, New York, 1942.

The Tulane University of Louisiana

DIMENSION AND NON-DENSITY PRESERVATION OF MAPPINGS

G. T. WHYBURN

1. Introduction. In this paper consideration is given to conditions under which the property of being non-dense in a space in the sense of containing no open set in that space is invariant under certain types of mappings. In some spaces and for some mapping types the issue involved is essentially equivalent to the question of dimensionality preservation. These questions are of interest and importance in numerous mathematical fields. They are especially so in the study of topological aspects of the theory of functions and it is toward this connection that the results and methods in this note will be largely directed.

A single valued continuous transformation $f(X)=Y$ will be called a *mapping*. Such a mapping is *open* if open sets in X have open images in Y and is *light* provided $f^{-1}(y)$ is totally disconnected for each $y \in Y$. Also f has *scattered point inverses* provided that for each $y \in Y$, $f^{-1}(y)$ is a *scattered set* in the sense that no point of $f^{-1}(y)$ is a limit point of $f^{-1}(y)$.

As indicated above, a set K in a space X is *non-dense in* X provided K contains no open set in X. On the other hand that K is *dense in* X means that every point of X is either a point or a limit point of K. A mapping $f(X)=Y$ is said to *preserve non-density* for compact sets provided that $f(K)$ is non-dense in Y whenever K is compact and non-dense in X. For a mapping $f(X)=Y$, a subset X_0 of X is said to be *semi-dense* in X provided X_0 is dense in some open subset of every open set U in X whose image $f(U)$ is also open in Y. Thus the property of semi-density is a property of a subset of X *relative to a mapping* f *on* X and not an intrinsic property of X_0 alone.

For a mapping $f(X)=Y$, the set of all $x \in X$ such that x is a component of $f^{-1}f(x)$ will be designated by the symbol D_f. Also the symbol L_f will be used for the set of all $x \in X$ such that $f^{-1}f(x)$ is totally disconnected. Thus L_f is the maximum inverse set in X on which the mapping f is light, where by an *inverse set* I we mean a set which is the inverse of its transform under f, that is, one satisfying the relation

$$I=f^{-1}f(I) .$$

Accordingly L_f may be thought of as the *lightness kernel* or 0-*dimensional kernel* of the mapping f. Obviously we have $L_f \subset D_f$.

Received May 31, 1956.

2. General setting. We begin with a theorem which was suggested by the theorem of Alexandroff's [1] on invariance of dimension under countable-fold open mappings. Our proof closely parallels that of Alexandroff for his theorem.

(2.1) THEOREM. *Let $f(A)=B$ be open and have scattered point inverses, where A and B are locally compact separable and metric. Then A is the union $A=\sum A_n$ of a sequence of compact sets such that $f|A_n$ is topological for each n.*

Proof. Let (U_n) be a countable basis of open sets in A, so chosen that $\bar{U}_n$ is compact for each n. For each n, let F_n be the set of all $x \in U_n$ such that $g_n^{-1}g_n(x)=x$ where g_n denotes the mapping $f|U_n$. Then F_n is closed in U_n by openness of f. Accordingly each F_n is the union of a countable sequence of compact sets and thus we can write $\sum F_n = \sum A_n$ where each A_n is compact and lies in some F_m. Thus $f|A_n$ is topological for each n. Finally, $\sum A_n=A$, because if $x \in A$, there exists an m such that $x \in U_m$ and $U_m \cdot f^{-1}f(x)=x$ and hence so that $x \in F_m \subset \sum F_n = \sum A_n$.

(2.11) COROLLARY. *For any closed set K in A we have*

$$\dim f(K) = \dim K .$$

(2.12) COROLLARY. *If K is any closed set in A and V is any open subset of $f(K)$, then V contains an open subset U which is homeomorphic with a subset of K.*

For let K_n denote the set $K \cdot A_n$ for each n. Then since $V \subset \sum f(K_n)$ and each $f(K_n)$ is compact, some $f(K_n)$ contains an open subset U of V. Then $K_n \cdot f^{-1}(U)$ maps topologically onto U under f.

(2.13) COROLLARY. *If A and B are 2-manifolds, (or n-manifolds), then if K is non-dense in A, $f(K)$ is non-dense in B.*

(2.2) THEOREM. *Let A and B be locally compact separable metric spaces and let $f(A)=B$ be a mapping preserving compact non-dense sets. Then for some $y \in B$, $f^{-1}(y)$ is totally disconnected.*

Proof. For each integer $n>0$ and each $x \in A$, let U_x^n be an open set of diameter $<1/n$ containing x and having a compact boundary F_x^n. Let $U_{x_1}^n, U_{x_2}^n, \cdots$ be a countable collection of these sets U_x^n whose union covers A and set

$$F^n = \sum_i F_{x_i}^n , \qquad F = \sum_n F^n = \sum_n \sum_i F_{x_i}^n .$$

Now $f(F) \neq B$. For if $f(F)=B$, then for some n and i the set $f(F^n_{x_i})$ must contain an open set in B, as B is locally compact; and this is impossible by hypothesis because $F^n_{x_i}$ is compact and non-dense for each n and i. Accordingly there exists a $y \in B-f(F)$. Clearly $f^{-1}(y)$ is totally disconnected, because if it had a non-degenerate component C_y, then for $1/n <$ the diameter of C_y, we would have $C_y \cdot F^n_{x_i} \neq 0$ where $C_y \cdot U^n_{x_i} \neq 0$.

(2.21) Corollary. *Under the same hypothesis, the set Y of all $y \in B$ with $f^{-1}(y)$ totally disconnected is dense in B and $L_f=f^{-1}(Y)$ is semi-dense in A.*

For if B_0 is any open set in B, we have only to set $A_0=f^{-1}(B_0)$ and apply the theorem to the mapping $f|A_0$ to obtain the first conclusion that Y is dense in B. To prove the second conclusion suppose on the contrary that for an open set U which has an open image and which we first suppose conditionally compact, $L_f \cdot U$ is dense in no open subset of U. Then $\overline{L_f \cdot U}$ is compact and non-dense, whereas $f(\overline{L_f \cdot U})$ must contain $f(U)$ since Y is dense in $f(U)$ by openness of $f(U)$. This is a contradiction.

Finally, to see that U need not be conditionally compact, we need only show that *any open set V in A with an open image contains a conditionally compact open subset U with an open image.* To do this, set $V=\sum V_n$ where each V_n is open and conditionally compact and $\overline{V}_n \subset U$. Then since $f(V)=\sum f(\overline{V}_n)$, some $f(\overline{V}_n)$ contains an open set G. Since $f[\mathrm{Fr}(V_n)]$ is non-dense, $Q=G-f[\mathrm{Fr}(V_n)]$ is open and non-empty. Then $U=V_n \cdot f^{-1}(Q)$ meets our condition.

(2.3) Theorem. *Let A and B be locally compact separable metric spaces with* $\dim A=k<\infty$ *and let $f(A)=B$ be a mapping such that the image of every compact non-dense set K with* $\dim K<k$ *is non-dense. Then the set Y of all $y \in B$ with $f^{-1}(y)$ totally disconnected is dense in B.*

For, in the preceding proofs the sets F^n_x could now be taken of dimension $\leqq k-1$.

3. Quasi-open mappings. Region on a sphere. A mapping $f(X)=Y$ is *quasi-open* provided that if $y \in Y$ and K is a compact component of $f^{-1}(y)$, then for any open set U in X containing K, y is interior to $f(U)$ rel. Y, and is *strongly quasi-open* provided y is interior to $f(U)$ relative to a larger space $Y_0 \supset Y$. A mapping $f(X)=Y$ is *monotone* provided $f^{-1}(y)$ is a continuum (compact and connected set) for each $y \in Y$; and f is *compact* provided $f^{-1}(K)$ is compact for every compact set $K \subset Y$ or, equivalently, provided f is closed and has compact point inverses. For compact mappings, quasi-openness is equivalent to quasi-monotoneity as defined originally by Wallace [3].

(3.1) THEOREM. *Let $f(X)=Y$ be a compact and quasi-open mapping where X is a region on a sphere S, Y is a metric space and where no component of a point inverse separates X. In order that the image of every compact 1-dimensional set in X be of dimension ≤ 1 it is necessary and sufficient that the set D_f be semi-dense in X.*

Proof. Let $f=lm$, $m(X)=X'$, $l(X')=Y$ be the monotone-light factorization of f. Let the mapping m be extended to the whole sphere S by decomposing S into the sets $m^{-1}(x')$, $x' \in X'$ together with the components of $S-X$ so that we obtain a monotone mapping $\phi(S)=S'$ of S onto a sphere S' containing X' (ϕ is the natural mapping of the decomposition) which is identical with m on X. That S' is a topological sphere follows from the readily verified facts that the described decomposition of S is upper semi-continuous and no element of this decomposition separates S, together with the classical theorem of R. L. Moore [2] that the hyperspace of any such decomposition of a sphere into continua is itself a topological sphere. Then $l(X')=Y$ is a light open mapping which is compact; and since X' is a region on S', Y is a 2-manifold by the invariance of the 2-manifold property under such mappings [4].

Now to prove the sufficiency of the condition let K be a compact 1-dimensional set in X. Then $\dim m(K) \leq 1$. For, if not, then $m(K)$ contains an open set U in X'. Then $l(U)$ is open in Y and thus $m^{-1}(U)$ is an open set in X whose image under f is open in Y. Accordingly D_f is dense in an open subset Q of $m^{-1}(U)$. Since Q cannot lie wholly in K, $Q-Q\cdot K$ contains a point x of D_f. But then since $x=m^{-1}m(x)$, $m(x)$ cannot lie in $m(K)$, contrary to the supposition that $m(x) \in U \subset m(K)$. Thus $\dim m(K) \leq 1$.

It remains to show that $\dim lm(K) \leq 1$. Since l is compact, open and light and X' is a 2-manifold, l is finite to one [4]. Hence by (2.11) we have $\dim lm(K)=\dim m(K) \leq 1$.

To prove the necessity of the condition we note first that it follows from our hypothesis that f preserves non-density for compact sets. For if K is a compact, non-dense set in X we have $\dim K \leq 1$. Whence $\dim f(K) \leq 1$; and since as shown above Y also is a 2-manifold, it follows from this that $f(K)$ is non-dense. Accordingly, by (2.21) not only D_f but also L_f must be semi-dense in X.

Clearly we have the following alternative form of (3.1) which we state as

(3.2) THEOREM. *Let f, X and Y be as described in the first sentence of* (3.1). *In order that f preserve non-density for compact sets it is necessary and sufficient that L_f be semi-dense in X.*

4. Quasi-open mappings on the general 2-manifold.

We now show that the case of a mapping of this same type operating on an arbitrary 2-manifold can be reduced essentially to the case of a region on a sphere so that similar conclusions hold.

(4.1) LEMMA. *Let $f(X)=Y$ be quasi-open where X is a 2-manifold without edges and Y is a locally connected generalized continuum and suppose that $f(L_f)$ is dense in Y. If there exists in X a compact set K of dimension $\leqq 1$ whose image contains an open set in Y, then there exists a region R in X contained in a 2-cell of X such that $Q=f(R)$ is open in Y, the mapping $f(R)=Q$ is compact and quasi-open and for some compact subset K_1 of $K\cdot R$, $f(K_1)$ contains an open set.*

Proof. Let V be an open set in $f(K)$. Then there is a point $y\in V$ such that $f^{-1}(y)$ is totally disconnected. Now for each $x\in K\cdot f^{-1}(y)$ there exists a 2-cell E_x on X with edge C_x and interior I_x such that $f(E_x)\subset V$, $C_x\cdot f^{-1}(y)=0$. Thus if Q_x is the component of $Y-f(C_x)$ containing y and R_x is the component of $f^{-1}(Q_x)$ containing x we have $R_x\subset I_x$ because $R_x\cdot C_x=0$. Accordingly, R_x being conditionally compact [5], $f(R_x)=Q_x\subset V$ and the mapping $f(R_x)=Q_x$ is compact and quasi-open.

Now since $K\cdot f^{-1}(y)$ is covered by a finite union U of the sets R_x and $f(K\cdot U)$ contains an open set $V-f(K-K\cdot U)$ in V about y, some one of the sets R_x, say R, is such that $f(K\cdot R)$ contains an open set. Since $K\cdot R$ is closed in R, for some compact set $K_1\subset K\cdot R$, $f(K_1)$ must likewise contain an open set in $Q=f(R)$. Thus the lemma is proven.

Since a region on a 2-cell may be considered as a region on a sphere (by mapping the 2-cell topologically onto a 2-cell on a sphere), this lemma together with the theorems in § 3 yield at once

(4.2) THEOREM. *Given a quasi-open mapping $f(X)=Y$ where X is a 2-manifold without edges and Y is a locally connected generalized continuum such that no component of a point inverse lying inside a closed 2-cell on X separates X, in order that f preserve non-density for compact sets it is necessary and sufficient that L_f be semi-dense in X.*

Note. Most of the results in this paper were stated without proof, or with only brief indications of proof in some cases, by the author in his Presidential Address before the American Mathematical Society [6].

For further discussion of these results, in particular for cases in which alternative dimension preserving forms of (4.2) above are possible, see [6].

5. Differentiable functions. We now show that a mapping from a region of the z-plane Z into the w-plane W generated by a function $w=f(z)$ satisfying certain differentiability conditions will satisfy the requirements needed in the preceding sections to insure the preservation of non-density for compact sets.

(5.1) THEOREM. *Let $w=f(z)$ be continuous in a region X of Z and differentiable at all points of a dense set $f^{-1}(Y_0)$ in X which is the inverse of an open subset Y_0 of $Y=f(X)$. Then f is strongly quasi-open, no component of a point inverse lying inside a closed 2-cell on X separates X and the set $f(L_f)$ is dense in Y. Further, L_f is semi-dense in X.*

Proof. (Note. In the proof of all but the final statement use is made of only easily established topological properties of functions meeting minimum differentiability requirements. In proving the last one, however, we use the property, rather more difficult to establish topologically, that a non-constant function everywhere differentiable in a region R cannot be constant on any open set in R.)

To prove f strongly quasi-open it suffices (see § 7 of [6]) to show that for any elementary region R in X with boundary C in X,

$$(*)\quad f(R+C)=f(C)+\text{the union of bounded components of } W-f(C)\,,$$

where "elementary" means that R is bounded and C consists of a finite number of disjoint simple closed curves. To accomplish this, let S be a component of $W-f(C)$ such that the set $S_0=S\cdot f(R)$ is not empty. Since $R\cdot f^{-1}(S_0)$ is open and nonempty, it therefore intersects $f^{-1}(Y_0)$. Thus $S_0\cdot Y_0$ is not empty. Let Q be a component of $S_0\cdot Y_0$. Since $R\cdot f^{-1}(Q)$ is open and thus has only a countable number of components, there exists a component T of $R\cdot f^{-1}(Q)$ on which f is not constant. As f is differentiable on T by hypothesis [because $T\subset f^{-1}(Y_0)$] there exists a point $z_0\in T$ where $f'(z_0)\neq 0$. Now using properties of the circulation index, it readily follows that Q contains the interior of a square and thus contains a point q such that $f'(z)\neq 0$ for all $z\in f^{-1}(q)$. Since this makes the circulation index equal $2\pi i$ times a positive integer when taken around any sufficiently small circle enclosing a point of $f^{-1}(q)$, it results at once that the circulation index taken over all of C of f about q must be $\neq 0$. Further, since this latter index is constant throughout S, that is, it has the same value when any $p\in S$ is substituted for q, it follows that every point p of S must belong to $f(R)$. For details of the argument needed here using the circulation index the reader is referred to the last paragraph of § 5 of [7].

Hence we have $S \subset f(R)$. This gives (*), however, because $f(R+C)$ obviously cannot contain the whole unbounded component of $W-f(C)$. Thus any component of $W-f(C)$ intersecting $f(R)$ must be bounded and must lie wholly in $f(R)$. Accordingly f is strongly quasi-open.

Suppose, contrary to the second assertion, that some component K of $f^{-1}(w_0)$, for some $w_0 \in Y$, lies inside a closed 2-cell A on X and separates X. Then one component Q of $X-K$ must lie wholly inside A since only one component of $X-K$ intersects the edge of A. Let y be a point of $f(Q+K)$ such that $|y-w_0| = \max |f(z)-w_0|$ for $z \in Q+K$. Then Q contains a component H of $f^{-1}(y)$ and H is compact. Accordingly, by the strong quasi-openness of f, y must be interior to $f(Q)$ contrary to $|f(z)-w_0| \leqq |y-w_0|$ for all $z \in Q$.

That $f(L_f)$ is dense in Y is an immediate consequence of the fact that Y_0 is dense in Y and the quasi-openness of f already established. For any open set in Y thus contains the interior I of a square such that f is differentiable everywhere on $f^{-1}(I)$. Thus for some $q \in I$, $f'(z) \neq 0$ for all $z \in f^{-1}(q)$. This makes $f^{-1}(q)$ a scattered set which therefore surely lies in L_f.

Finally, to prove L_f semi-dense in X we note first that if T is any region in X on which f is non-constant and everywhere differentiable, then as shown above in the second paragraph of this proof, T contains a point z_0 where $f'(z_0) \neq 0$ and indeed $f(T)$ contains points q such that $f'(z)$ does not vanish on $f^{-1}(q)$, so that $f^{-1}(q) \subset L_f$. Accordingly any such region T intersects L_f. Now if U is any open set in X with an open image, $f(U) \cdot Y_0$ contains a region Q and if T is any component of $f^{-1}(Q) \cdot U$ on which f is not constant (and there are such components T because the collection of all components of $f^{-1}(Q) \cdot U$ is countable), then T intersects L_f as shown above. However, the same holds for an arbitrary subregion T_0 of T, because f is likewise non-constant and everywhere differentiable on T_0. Thus L_f is dense in T.

REFERENCES

1. P. S. Alexandroff, Dokl. (C. R.) Akad. Nauk SSSR. IV (XIII), No. 7 (111), 295–299.
2. R. L. Moore, *Concerning upper semi-continuous collections of continua*, Trans. Amer. Math. Soc., **27** (1925), 416–428.
3. A. D. Wallace, *Quasi-monotone transformations*, Duke Math. J., **7** (1940), 136–145.
4. G. T. Whyburn, *Interior transformations on surfaces*, Amer. J. Math., **60** (1938), 477–490.
5. ———, *An open mapping approach to Hurwitz's theorem*, Trans. Amer Math. Soc., **71** (1951), 114.
6. ———, *Topological analysis*, Bull. Amer. Math. Soc., **62** (1956), 204–218.
7. ———, *Lectures on functions of a complex variable*, Univ. of Michigan Press (1955), 1–14.

UNIVERSITY OF VIRGINIA

ON THE FUNCTIONAL REPRESENTATION OF CERTAIN ALGEBRAIC SYSTEMS

J. H. WILLIAMSON

1. Introduction, Definitions and Examples. In this paper an attempt is made to generalize the well-known representation theory of commutative Banach algebras by functions on the maximal ideals of the algebra [4]. The present paper is devoted almost exclusively to algebraic questions; topological aspects of the theory will be treated elsewhere.

In considering commutative algebras A over the complex field C, there are relatively few cases in which one can assert that the quotient A/M of the algebra by a maximal ideal is isomorphic to C. Apart from Banach algebras, there are the locally m-convex algebras of E.A. Michael [6] and R. Arens [1], and the 'algèbres à inverse continu' of L. Waelbroeck [8], [9] ($=Q$-algebras, in the terminology of Kaplansky [5], with continuous inversion). There are many interesting algebras which do not belong to either of these classes, and it would be desirable to have a theory to cover them as far as possible.

The basic idea is derived from the classical work of Carleman, von Neumann, and Stone on unbounded self-adjoint linear operators T in Hilbert space (see, for example, [7]). Here the analysis is carried out with the aid of the bounded transformations $(T-\lambda I)^{-1}$; the spectrum of T is the set of complex numbers λ such that $(T-\lambda I)^{-1}$ does not exist as a bounded transformation. This suggests that if we start with a commutative algebra A, and a suitable sub-algebra B (corresponding to the 'bounded' elements of A) we may be able to effect a useful analysis of A, and somehow represent an element $a \in A$ by a function whose values are those complex numbers λ such that $(a-\lambda e)^{-1}$ does not exists in B (e being the unit of A). It turns out that this is basically correct, although there are certain complications of detail. For instance, the representing functions may take infinite values; this is unavoidable. The space on which the functions are defined is that of the 'maximal B-ideals' or 'maximal ordinary B-ideals' of the algebra, not the space of maximal ideals in the ordinary sense.

Much of the theory of this paper applies to algebras over fields of fairly general type; for instance, many results are true for any algebraically closed field. It is no more difficult to develop the theory for the general case than for the case of the complex field. Let K be any

Received December 24, 1955. This paper was written while the author was a Commonwealth Fund Fellow.

(commutative) field, A a commutative linear algebra over K, with a unit e, and B a sub-algebra of A, containing e. A restriction will presently be put on B (immediately following Lemma 4), and after Theorem 1, K will be taken to be algebraically closed. Further special assumptions on A and K will be made in the later sections of the paper.

DEFINITION 1. A subset J of A is a *B-ideal* of A if

(i) $x-y \in J$ whenever $x \in J$, $y \in J$, and

(ii) $xb \in J$ whenever $x \in J$, $b \in B$.

The B-ideal J is *admissible* if $e \notin J$; it is *ordinary* if $xy \in J$ whenever $x \in J$, $y \in J$; otherwise it is *exceptional*. (B-ideal$=$$B$-submodule; ordinary B-ideal$=$$B$-submodule which is a sub-algebra).

It may be useful to remark that a B-ideal which is a proper subset of A is not necessarily an admissible B-ideal, by the above definition. For instance, B itself is clearly a B-ideal of A; it may be a proper subset of A but it is never an admissible B-ideal.

We give now one or two examples of the type of system under consideration.

(i) Let A be any algebra of the type specified above, and take $B=A$. The B-ideals of A are the ideals (in the usual sense) of A; all are ordinary.

(ii) Let A be as in (i), and take $B=Ke$ (which we shall sometimes write as K, if no danger of confusion exists). The B-ideals of A are the linear subspaces of A.

In particular, let A be the algebra of pairs of complex numbers (a_1, a_2), with pointwise addition and multiplication. The admissible B-ideals of A are the proper linear subspaces not containing $(1, 1)$. They are thus (a) the element $(0, 0)$, and (b) for each complex $\alpha \neq 1$, the subspace generated by $(1, \alpha)$, and the subspace generated by $(0, 1)$. There are precisely three ordinary admissible B-ideals, namely $(0, 0)$ and those generated by $(0, 1)$ and $(1, 0)$.

(iii) Let A be the algebra of polynomials, with complex coefficients, in the indeterminate t, and let B be the sub-algebra of constants. The sets $\{a: a(t_0)=0\}$ (t_0 a complex number) are clearly ordinary B-ideals of A. An elementary argument shows that they are maximal admissible ordinary B-ideals; it will appear later (after Theorem 2) that these are the only such B-ideals.

(iv) As for (iii), but with 'polynomial' replaced by 'rational function'. Here the maximal ordinary B-ideals are the sets $\{a: a(t_0)=0\}$ for each complex t_0 and the set $\{a: a(\infty)=0\}$.

(v) Let A be the algebra of (equivalence-classes of) complex almost everywhere finite Lebesgue measurable functions on $(0, 1)$, B the sub-algebra of essentially bounded functions. Among the B-ideals of A are (a) the set of all functions of A which are zero (almost everywhere) on E, for any fixed subset E of (0, 1) of positive measure (this is an ordinary B-ideal, and in fact an ideal); and (b) the set of functions $f(t)$ such that $|f(t)| \leqq kn^{-1}$ almost everywhere in E_n, for each n, where E_n is a decreasing sequence of measurable sets such that the measure of E_n tends to zero as n tends to infinity (k depends on f only). This is an ordinary B-ideal, but not an ideal.

(vi) Let A be an algebra of (possibly unbounded) self-adjoint or normal linear transformations of a Hilbert space into itself, and let B be the sub-algebra of bounded operators. This type of algebra will be considered in § 7.

In what follows it will be important to distinguish clearly between *ordinary maximal B-ideals*, that is, admissible B-ideals which are ordinary and which are not properly contained in any admissible B-ideal, and *maximal ordinary B-ideals*, that is, admissible B-ideals which are ordinary and which are not properly contained in any admissible ordinary B-ideal (maximal=maximal admissible). In example (ii) above, all the B-ideals (b) are clearly maximal. Of these only the ideals generated by (0, 1) and (1, 0) are ordinary; and these two are clearly also the only maximal ordinary B-ideals of A.

LEMMA 1. (i) *If J is a maximal B-ideal of A, then $B \cap J$ is a maximal ideal of B; if I is a maximal ideal of B, there is a maximal B-ideal of A containing I.*

(ii) *If J is a maximal ordinary B-ideal of A, then $B \cap J$ is a maximal ideal of B; if I is a maximal ideal of B, there is a maximal ordinary B-ideal of A containing I.*

Proof. (i) It is clear that $B \cap J$ is a proper ideal of B. Suppose that J' is a proper ideal of B which properly contains $B \cap J$; then $J + J'$ is a B-ideal of A which properly contains J and does not contain e. Since J was assumed to be maximal, this is a contradiction, and so $B \cap J$ is a maximal ideal of B.

The second assertion follows, by a simple application of Zorn's lemma, from the fact that any proper ideal of B is an admissible B-ideal of A, and the fact that the union of an ascending chain of admissible B-ideals is clearly an admissible B-ideal.

(ii) As for (i), with 'B-ideal' replaced by 'ordinary B-ideal'.

In general, a maximal ideal of B is contained in many maximal (or maximal ordinary) B-ideals of A; but in some cases it is possible to assert that the extension is unique; see § 4, Proposition 5 and § 7, Lemma 13.

2. A representation theorem. Let ∞ be a symbol such that $\infty+\lambda=\infty$ for all $\lambda \in K$, $\infty\cdot\infty=\infty$ and $\lambda\infty=\infty$ for all nonzero $\lambda\in K$. Denote the field K augmented by ∞ by the symbol K'. Now let J be any linear subspace of A, not containing e. Define a function with values in K' as follows:

DEFINITION 2.

$$f_J(a)\begin{cases}=\lambda \text{ if } a-\lambda e\in J\,,\\ =\infty \text{ if } a-\lambda e\notin J \text{ for all } \lambda\in K.\end{cases}$$

It is clear that the function is uniquely defined for all $a\in A$. There are one or two immediate consequences of the definition:

LEMMA 2. (i) *$f_J(\alpha a)=\alpha f_J(a)$ for all $\alpha\in K$, $a\in A$ ($0\cdot\infty=0$ here).*
(ii) *$f_J(a_1+a_2)=f_J(a_1)+f_J(a_2)$, (when the right-hand side is defined).*

Proof. (i) If $f_J(a)=\lambda\in K$ then $a-\lambda e\in J$, whence $\alpha a-\alpha\lambda e\in J$ and $f_J(\alpha a)=\alpha\lambda=\alpha f_J(a)$. If $f_J(a)=\infty$, then $a-\lambda e\notin J$ for all $\lambda\in K$; clearly if $\alpha\neq 0$ then $\alpha a-\mu e\notin J$ for all $\mu\in K$, and so $f_J(\alpha a)=\infty$. If $\alpha=0$ then $f_J(\alpha a)=f_J(0)=0$ for all $a\in A$.

(ii) If $f_J(a_1)=\lambda_1\in K$, $f_J(a_2)=\lambda_2\in K$, then $a_1+a_2-(\lambda_1+\lambda_2)e\in K$ and the result follows. If $f_J(a_1)=\lambda\in K$, and $f_J(a_2)=\infty$, then if $f_J(a_1+a_2)=\mu\in K$, we would have $f_J(a_2)=f_J(a_1+a_2-a_1)=\mu-\lambda\in K$, a contradiction.

Next we turn to the multiplicative properties of the function $f_J(a)$. It is clear that if we are to obtain any general results we must take J to be a B-ideal of A, and moreover an ordinary B-ideal; if J is not ordinary we could find $a_1\in J$, $a_2\in J$, with $a_1a_2\notin J$, that is,

$$f_J(a_1a_2)\neq f_J(a_1)f_J(a_2)=0\,.$$

The first result, however, is valid for any sub-algebra J:

LEMMA 3. *Let J be any sub-algebra of A not containing e. Then if neither of $f_J(a_1)$, $f_J(a_2)$ is ∞, we have $f_J(a_1a_2)=f_J(a_1)f_J(a_2)$.*

Proof. Let $a_1=f_J(a_1)e+j_1$, $a_2=f_J(a_2)e+j_2$, where $j_1\in J$, $j_2\in J$. Then $a_1a_2=f_J(a_1)f_J(a_2)e+f_J(a_2)j_1+f_J(a_1)j_2+j_1j_2$; the required result follows at once.

Difficulties arise when one or both of $f_J(a_1)$, $f_J(a_2)$ is ∞.

LEMMA 4. *If J is a sub-algebra of A and $f_J(a) \neq \infty$ then $aJ \subset J$.*

Proof. If $a = \lambda e + j$ $(j \in J)$ then for any $j' \in J$ we have $aj' = \lambda j' + jj' \in J$.

For the next lemma, and for all future developments, we require to make the following assumption.

Assumption. If M is any maximal ideal of B then $B/M \cong K$.

This assumption is satisfied in the cases in which we are interested.

LEMMA 5. *If J is a maximal B-ideal of A and $aJ \subset J$ then $f_J(a) \neq \infty$.*

Proof. The result is trivial if $a \in J$; we then have $f_J(a) = 0$. If $a \notin J$ then $J + aB$ is a B-ideal properly containing J; since J was maximal, $e = j + ab$ for some $j \in J$, $b \in B$. We have $b = \lambda e + j'$ for some $\lambda \in K$, $j' \in J$, by assumption; hence $e - \lambda a = j + aj' \in J$. We clearly cannot have $\lambda = 0$; hence $a - \lambda^{-1} e \in J$ and $f_J(a) = \lambda^{-1} \neq \infty$.

COROLLARY. *If J is maximal and $f_J(a) = \infty$, then $e = aj + j'$, where j, $j' \in J$.*

Proof. $aJ + J$ is a B-ideal of A properly containing J and hence containing e.

LEMMA 6. *If J is an ordinary maximal B-ideal of A, then $f_J(a_1 a_2) = f_J(a_1) f_J(a_2)$, whenever the right-hand side is defined.*

Proof. The case in which $f_J(a_1)$ and $f_J(a_2)$ are both finite has already been covered (Lemma 3). Suppose then that $f_J(a_1) = \infty$. By Lemma 5, Corollary, we have $e = a_1 j + j'$, where j, $j' \in J$. If $a_2 - \lambda e = j'' \in J$ $(\lambda \neq 0)$ we have $a_1 a_2 j = \lambda e + j'' - \lambda j' - j' j'' \notin J$, whence $f_J(a_1 a_2) = \infty$, by Lemma 4. If $e = a_2 j_1 + j_2$ $(j_1, j_2 \in J)$ then $a_1 a_2 j j_1 = e - j' - j_2 + j' j_2 \notin J$, whence $f_J(a_1 a_2) = \infty$ as before.

We can now collect the results obtained.

THEOREM 1. *Let $\mathscr{J}_0$ be the set of ordinary maximal B-ideals of A. Then there is a mapping of A into the set of K'-valued functions on $\mathscr{J}_0$: $a \rightarrow f_J(a)$, so that the structure of A is preserved as far as it can be, that is $f_J(\alpha a) = \alpha f_J(a)$, $f_J(a_1 + a_2) = f_J(a_1) + f_J(a_2)$, and $f_J(a_1 a_2) = f_J(a_1) f_J(a_2)$, whenever the right-hand sides of these equalities are defined.*

The above theorem has one serious flaw; given A and B, the set $\mathscr{J}_0$ may be empty. For example, let A be a field properly containing K, and take $B=Ke$. Then any maximal B-ideal is a maximal linear subspace of A not containing e; any $a \in A$ can be expressed uniquely as $a =\lambda e+j$, where $j \in J$. If J were ordinary we would have $aJ \subset J+JJ =J$, that is, J would be an ideal of A in the usual sense, which is impossible.

It is uncertain whether, given A, it is possible to choose B so that there is at least one ordinary maximal B-ideal. In any case, B will often be prescribed in advance, so that no choice is possible.

We are thus obliged to look at maximal ordinary B-ideals rather than ordinary maximal B-ideals. We have, by Lemma 1 (ii), the assurance that there always exist at least as many maximal ordinary B-ideals of A as there are maximal ideals of B, that is, always at least one.

3. A better representation theorem.

We now consider maximal ordinary B-ideals instead of maximal B-ideals. This introduces some technical difficulties (which can, however, be overcome), and also makes it necessary to confine attention to fields K which are algebraically closed. We shall make this assumption from now on. The sort of difficulty which arises if the field is not algebraically closed is adequately illustrated by considering the complex field C, as an algebra over the real field R. Here there is a unique maximal ordinary R-ideal $J =\{0\}$; if a is any complex number with a nonzero imaginary part, then $f_J(a)=\infty$. Clearly the multiplicative properties of f are quite unsatisfactory.

LEMMA 7. *If J is a maximal ordinary B-ideal and $aJ \subset J$ then $f_J(a) \neq \infty$.*

Proof. If $a \in J$ then $f_J(a)=0$; suppose then that $a \notin J$. The set $J+aB+a^2B+\cdots$ is an ordinary B-ideal of A, properly containing J; hence $e=j+ab_1+a^2b_2+\cdots+a^nb_n$ for some $j \in J$, $b_1, \cdots, b_n \in B$. We shall show that we can take $n=1$ here. First, it is to be noted that there is no loss of generality in supposing that $b_1, \cdots, b_n$ are all scalar multiples of e; if in the above representation we had $b_r=\lambda_r e+j_r$ $(1 \leqq r \leqq n)$ then we could also write

$$e=j'+\lambda_1 a+\lambda_2 a^2+\cdots+\lambda_n a^n, \text{ where } j'=j+j_1 a+\cdots+j_n a^n \in J .$$

Second, it is clearly permissible to assume that the representation of e in this way is of minimum degree. We do this. Let $\mu(a)$ be a polynomial in a, with coefficients in K, which is in J, and of minimum degree. Assume that the degree of μ is $n > 1$. Let $(a-\alpha e)$ be a factor of $\mu(a)$, and write $\mu(a)=(a-\alpha e)x$. By assumption $x \notin J$, and so the set $J+xB+x^2B+\cdots$ is an ordinary B-ideal of A, properly containing J. We thus have

$$e=j''+xb_1'+\cdots+x^m b_m', \text{ where } j'' \in J,\ b_1', \cdots, b_m' \in B .$$

This gives

$$a-\alpha e=(a-\alpha e)j''+\mu(a)b_1'+x\mu(a)b_2'+\cdots+x^{m-1}\mu(a)b_m' \in J ,$$

which contradicts the assumption that $\mu(a)$ was of minimum degree. Thus $\mu(a)=a-\alpha e$, and $f_J(a)=\alpha \neq \infty$.

COROLLARY. *If J is a maximal ordinary B-ideal and $f_J(a)=\infty$ then*

$$e=j_0+aj_1+a^2j_2+\cdots+a^n j_n \text{ for some } j,\ j_1, \cdots, j_n \in J.$$

Proof. The set $J+aJ+a^2J+\cdots$ is an ordinary B-ideal of A, properly containing J. Hence it contains e.

It will appear later (Lemma 9, Corollary) that we can always take $n=1$ in this representation. In the meantime it is convenient to formulate this as follows.

Property P. Let J be a maximal ordinary B-ideal of A, and a an element of A such that $f_J(a)=\infty$. We shall say that property P holds (for a and J) if $e=aj+j'$ for some $j,\ j' \in J$.

LEMMA 8. *If $f_J(a)=\infty$, and property P does not hold, then we can find $j^* \in J$ such that $f_J(aj^*)=\infty$.*

Proof. Clearly, by Lemma 7 we can find j^* such that $aj^* \notin J$. If $f_J(aj^*)=\alpha$, we would have $aj^*-\alpha e \in J$, whence $e=\alpha^{-1}aj^*+j'$, that is, property P would hold. Since we assume the contrary, $f_J(aj^*)=\infty$.

LEMMA 9. *If J is a maximal ordinary B-ideal of A, then $f_J(a_1a_2) =f_J(a_1)f_J(a_2)$, whenever the right-hand side is defined.*

Proof. (a) The case in which both factors on the right are finite has already been covered (Lemma 3).

(b) Suppose that $f_J(a_1)=\infty$, with property P, and $f_J(a_2)=\alpha$ ($\neq 0$). Then $e=a_1j+j'$, $a_2=\alpha e+j''$, so that

$$a_1a_2j=\alpha e-\alpha j'+j''-j'j'' \notin J .$$

Hence $f_J(a_1a_2)=\infty$, by Lemma 4.

(c) Suppose $f_J(a_1)=f_J(a_2)=\infty$, property P holding for both. Then $e=a_1j_1+j_1'$, $e=a_2j_2+j_2'$, and so

$$a_1a_2j_1j_2=e-j_1'-j_2'+j_1'j_2'\notin J\ ,$$

whence $f_J(a_1a_2)=\infty$, as before.

(d) Now suppose that $f_J(a_1)=\infty$, property P not being true, and $f_J(a_2)=\alpha\neq 0$. We shall show that $f_J(a_1a_2)=\beta\in K$ is impossible. Let

$$e=j_0+j_1a_1+\cdots+j_na_1^n\ ,$$

and $a_1a_2=\beta e+j$. Then

$$a_2^n=j_0a_2^n+j_1(\beta e+j)a_2^{n-1}+\cdots+j_n(\beta e+j)^n\ ,$$

that is,

$$(\alpha e+j')^n=j_0(\alpha e+j')^n+j_1(\beta e+j)(\alpha e+j')^{n-1}+\cdots+j_n(\beta e+j)^n\ .$$

This gives at once $\alpha^ne\in J$, which is impossible. Hence $f_J(a_1a_2)=\infty$.

(e) Finally suppose that $f_J(a_1)=\infty$, property P not holding, and $f_J(a_2)=\infty$ (property P possibly holding, possibly not). We note first that it will be enough to prove that under these hypotheses $f_J(a_1a_2)=0$ is impossible. For, if $f_J(a_1a_2)=\alpha\neq 0$, we could choose j^* as in Lemma 8, and replace a_1 by a_1j^*. Then property P fails to hold for a_1j^*, and we would have $f_J(a_1j^*)=f_J(a_2)=\infty$, $f_J(a_1j^*a_2)=0$. So, assume that $f_J(a_1a_2)=0$. Let

$$e=j_0+j_1a_1+\cdots+j_ma_1^m=j_0'+j_1'a_2+\cdots+j_n'a_2^n\ ,$$

where m and n are minimal. It is clearly no restriction to assume that $m\geqq n$. If $a_1a_2=j\in J$, then

$$a_1^n(e-j_0')=j_1'ja_1^{n-1}+\cdots+j_n'j^n\ .$$

Multiply this by $j_ma_1^{m-n}$, and we have

$$j_ma_1^m(e-j_0')=j_mj_1'ja_1^{m-1}+\cdots+j_mj_n'j^na_1^{m-n}\ .$$

But also

$$j_ma_1^m(e-j_0')=(e-j_0-j_1a_1-\cdots-j_{m-1}a_1^{m-1})(e-j_0')\ ,$$

so that, equating the right-hand sides of the last two equations, we have an expression for e as a polynomial in a_1, with coefficients in J, and of degree $\leqq m-1$, which contradicts the assumed minimality of m. Thus $f_J(a_1a_2)=\infty$ in this case also.

The above five cases exhaust all the possibilities, and so the lemma is proved.

COROLLARY. *Property P always holds; that is, if J is a maximal ordinary B-ideal and $f_J(a)=\infty$, then $e=aj+j'$ for some j, $j' \in J$.*

Proof. By Lemma 7, Corollary, we have $e=ah+j'$, where $h=j_1+\cdots+j_n a^{n-1}$. By Lemma 9, we must have $f_J(h)=0$, that is, $h \in J$.

As in the case of maximal B-ideals, we collect our results:

THEOREM 2. *Let $\mathscr{J}_0'$ be the set of maximal ordinary B-ideals of A. Then there is a mapping of A into the set of K'-valued functions on $\mathscr{J}_0'$: $a \rightarrow f_J(a)$, so that $f_J(\alpha a)=\alpha f_J(a)$, $f_J(a_1+a_2)=f_J(a_1)+f_J(a_2)$ and $f_J(a_1a_2)=f_J(a_1)f_J(a_2)$, whenever the right-hand sides of these equalities are defined.*

Since, as has been remarked, there always exists a maximal ordinary B-ideal of A, Theorem 2 always has content.

We can now show, as promised, that the B-ideals specified in Example (iii) of § 1 are the only maximal ordinary B-ideals. Suppose that there is a maximal ordinary B-ideal J such that $f_J(t)=\infty$. Then, by Lemma 9, $f_J(a)=\infty$ for every non-constant polynomial $a \in A$; that is, $J=\{0\}$, which is clearly not maximal. Thus $f_J(t)$ is always finite, from which it follows at once that J is one of the specified B-ideals.

It may be noted that if J is a maximal ordinary (or ordinary maximal) B-ideal of A, then the function $f_J(a)$ has the properties

(1) $f_J(e)=1$; $f_J(b) \in K$ for all $b \in B$, and

(2) $f_J(\alpha a)=\alpha f_J(a)$, $f_J(aa')=f_J(a)f_J(a')$, $f_J(a+a')=f_J(a)+f_J(a')$ whenever the right-hand sides are defined. Conversely, if we have a function f with these properties, the set $\{a\colon f(a)=0\}$ is clearly an ordinary B-ideal of A but not in general a maximal one (consider Example (iii) of § 1 and write $f(a)=\alpha$ if $a=\alpha$ (constant), $f(a)=\infty$ otherwise). This is in contrast to the situation in which J is an ideal and f a genuine homomorphism.

4. Further general results. The spectrum, etc. We shall for the most part be concerned with maximal ordinary B-ideals; in one or two cases we consider maximal B-ideals (which may or may not be ordinary).

DEFINITION 3. Denote by B_2 the set of elements of A such that $f_J(a)$ is finite for all maximal ordinary B-ideals J, and by B_2' the set such that $f_J(a)$ is finite for all maximal B-ideals J. If $B=B_2$, we say that B is *strongly saturated;* if $B=B_2'$, then B is said to be *weakly saturated.*

It is evident that $B_2 \supseteq B_2' \supseteq B$.

PROPOSITION 1. (i) *B_2 is a sub-algebra of A.*

(ii) *The maximal ordinary B_2-ideals of A are the same as the maximal ordinary B-ideals of A.*

(iii) *If M is any maximal ideal of B_2, then $B_2/M \cong K$.*

(iv) *$(B_2)_2 = B_2$, for any B.*

Proof. (i) This is an immediate consequence of Lemma 3. (Note that in general B_2' is not a sub-algebra of A).

(ii) Clearly every B_2-ideal of A is also a B-ideal, since $B_2 \supseteq B$. On the other hand, by Lemma 3 every ordinary B-ideal is also an ordinary B_2-ideal. Hence the result follows.

(iii) By Lemma 1 (ii) the maximal ideals of B_2 are the traces on B_2 of the maximal ordinary B_2-ideals of A, that is, of the maximal ordinary B-ideals. Hence, for any M and any $a \in B_2$, we have $a - \alpha e \in M$ for some $\alpha \in K$, that is, $B_2/M \cong K$.

(iv) This follows at once from (ii).

The last part of the above proposition shows that for any A it is always possible to choose a strongly saturated sub-algebra B; $(Ke)_2$ is of the required type.

THEOREM 3. (i) *The element $a \in A$ has an inverse $a^{-1} \in B$ if and only if it is in no maximal B-ideal of A.*

(ii) *The element $a \in A$ has an inverse $a^{-1} \in B_2$ if and only if it is in no maximal ordinary B-ideal of A. If such an inverse exists, it is expressible as a polynomial in a with coefficients in B.*

Proof. (i) If $aa^{-1} = e$, where $a^{-1} \in B$, then clearly a cannot be in any admissible B-ideal of A. If $ab \neq e$ for all $b \in B$, then aB is an admissible B-ideal of A, and hence is contained in some maximal B-ideal J. Then $a = ae \in J$.

(ii) If J is a maximal ordinary B-ideal, it is also a maximal ordinary B_2-ideal, by Proposition 1 (ii). Thus if $a \in J$, the relation $e = aa^{-1}$, with $a^{-1} \in B_2$, is impossible.

If a is such that e is not expressible as a polynomial in a, with coefficients in B and without constant term, then the set of all such polynomials clearly forms an admissible ordinary B-ideal of A. There is thus a maximal ordinary B-ideal containing a. So, if $a \notin J$ for all maximal ordinary J, it follows that $e = aa^{-1}$, where a^{-1} is expressed as a polynomial in a with coefficients in B. By Lemma 9, since $f_J(a)$ is never zero it follows that $f_J(a^{-1})$ is never infinite, that is, $a^{-1} \in B_2$.

COROLLARY. *If B is strongly saturated, the element $a \in A$ has an* *inverse in B if and only if it is in no maximal ordinary B-ideal.*

In general the expression for a^{-1} as a polynomial in a will necessarily be of degree $\geqq 1$. Consider, for example, the algebra of § 1, Example (ii). If $\alpha \neq \beta$, and neither α nor β is zero, the element $a = (\alpha, \beta)$ satisfies the equation

$$e = (\alpha\beta)^{-1}\{(\alpha+\beta)a - a^2\} ,$$

so that $a^{-1} = (\alpha\beta)^{-1}\{(\alpha+\beta)e - a\}$. It is clear that a^{-1} cannot be expressed as a polynomial of lower degree (a constant multiple of e in this case).

DEFINITION 4. The range of values of $f_J(a)$ as J varies over all maximal B-ideals of A is the *B-spectrum* of a, denoted $\sigma'_B(a)$. The range of values of $f_J(a)$ as J varies over all maximal ordinary B-ideals of A is the *B-spectroid* of a, denoted $\tau'_B(a)$. We write $\sigma_B(a) = \sigma'_B(a) \cap K$, and $\tau_B(a) = \tau'_B(a) \cap K$; these may be referred to as the *finite parts* of the respective sets.

The set $\sigma_B(a)$ consists of those scalars α such that $a - \alpha e$ has no inverse in B; the set $\tau_B(a)$ consists of those scalars α such that $a - \alpha e$ has no inverse in B_2. In general if D is any subset of A, we shall denote by $\sigma_D(a)$ the set of scalars $\lambda \in K$ such that $(a - \lambda e)^{-1}$ fails to exist in D. It is clear that neither $\sigma'_B(a)$ nor $\tau'_B(a)$ can be empty, although each set may consist of the element ∞ only; an example of this is easily found in the algebra A of formal power-series in an indeterminate, with B the sub-algebra of series with nonnegative powers only. Here there is a unique maximal B-ideal, which is ordinary, consisting of series with positive powers only; if J is this B-ideal, and $a \notin B$, then clearly $f_J(a) = \infty$.

Since every maximal ordinary B-ideal of A is contained in a maximal B-ideal, it follows that $\sigma'_B(a) \supseteqq \tau'_B(a)$ for all $a \in A$. The following lemma describes a case in which the two sets are equal:

PROPOSITION 2. *If B is strongly saturated then $\sigma'_B(a) = \tau'_B(a)$ for all $a \in A$.*

Proof. It is clear, in view of the remarks following Definition 4, that if $B = B_2$ then $\sigma_B(a) = \tau_B(a)$. If $B = B_2$ and $\infty \notin \tau'_B(a)$ then by definition $a \in B_2 = B$, and so $\infty \notin \sigma'_B(a)$, in view of the assumption on B made after Lemma 4. In view of the relation $\sigma'_B(a) \supseteqq \tau'_B(a)$, this completes the proof.

It is of course, not true that if $B = B_2$ then the maximal ordinary

B-ideals coincide with the maximal B-ideals — consider the algebra of § 1, Example (iii).

Suppose that

$$(a-\alpha_1 e)^{-1}, \cdots, (a-\alpha_r e)^{-1}$$

exist in A. Then if

$$q(a)=(a-\alpha_1 e)^{t_1}\cdots(a-\alpha_r e)^{t_r},$$

where $t_1, \cdots, t_r$ are positive integers, and $p(a)$ is any polynomial, the rational function $r(a)=p(a)/q(a)$ certainly exists as an element of A. If this is so we have the 'spectral mapping theorem':

THEOREM 4. *If the rational function $r(a)$ of a exists in A then the B-spectroid of $r(a)$ is the image under $r(.)$ of the B-spectroid of a; that is, $\alpha \in \tau'_B(a)$ if and only if $r(\alpha) \in \tau'_B(r(a))$, $(\alpha \in K')$.*

Proof. This follows at once from Theorem 2.

Notice that the spectroid, not the spectrum, is involved; the result is false in general if 'spectrum' is substituted for 'spectroid.'

COROLLARY. *A necessary and sufficient condition that the rational function $r(a)$ should exist as an element of B_2 is that $r(\tau'_B(a)) \subseteqq K$.*

Proof. If

$$r(a)=\Pi(a-\alpha_i e)^{\rho_i},$$

then

$$\Pi(f_J(a)-\alpha_i e)^{\rho_i} \in K$$

for all maximal ordinary J. Thus if $\rho_i < 0$ we cannot have $f_J(a)=\alpha_i$, and so $(a-\alpha_i e)^{\rho_i}$ exists (in B_2 and, a fortiori, in A) for all i with $\rho_i < 0$. Thus $r(a)$ exists in A and the result follows at once from the theorem.

THEOREM 5. *If a and a' are any elements of A, then*

$$\tau_B(aa') \subseteqq \tau_B(a).\tau_B(a') \text{ and } \tau_B(a+a') \subseteqq \tau_B(a)+\tau_B(a').$$

These relations are also true when τ is replaced by τ', provided that the sets which occur on the right-hand sides do not contain a product $0\cdot\infty$ or a sum $\infty+\infty$, respectively.

Proof. This also follows at once from Theorem 2.

Theorem 5 can of course be extended to several elements of A, and combined with Theorem 4 to give information about the spectroid of a rational function of several elements of A.

Next, a condition that the spectrum should consist of the whole of K:

PROPOSITION 3. *If* $\infty \notin \sigma'_B(a)$ *(that is,* $a \in B'_2$*) and* $a \notin B$*, then* $\sigma'_B(a) = K$.

Proof. Suppose that $\alpha \in K$ is not in $\sigma'_B(a)$. Then $f_J(a-\alpha e)$ is never zero, for any maximal B-ideal J; hence, by Theorem 3 (i), $(a-\alpha e)^{-1}$ exists in B. Since $f_J(a-\alpha e)$ is never ∞, it follows that $f_J((a-\alpha e)^{-1})$ is never zero. This implies that $((a-\alpha e)^{-1})^{-1}=(a-\alpha e)$ is in B, and hence that $a \in B$.

PROPOSITION 4. *Let* $\alpha \in K$ *be such that* $(a-\alpha e)^{-1} \in B$. *Then either* $a \in B$, *or* $\infty \in \tau'_B(a)$.

Proof. Suppose that $a \notin B$. The set $(a-\alpha e)^{-1}B$ is clearly an ideal of B; it is admissible, since $(a-\alpha e)^{-1}b=e$ would imply $a \in B$, which is not so. Hence, by Lemma 1 (ii) there is a maximal ordinary B-ideal, J say, containing this set. Then we must have $f_J(a)=\infty$; for $(a-\beta e) \in J$ would imply $(a-\beta e)(a-\alpha e)^{-1} \in J$, that is, $e+(\alpha-\beta)(a-\alpha e)^{-1} \in J$, that is, $e \in J$, which is impossible.

Note that it is possible to have $a \notin B$, $\infty \notin \tau'_B(a)$—consider Example (iii) of § 1. In this case, of course, if $a \notin B$ there is no $\alpha \in K$ such that $(a-\alpha e)^{-1} \in B$.

PROPOSITION 5. *If, for each* $a \in A$, *there exists* $\alpha \in K$ *such that* $(a-\alpha e)^{-1} \in B$, *then*

(i) *B is strongly saturated, and*

(ii) *each maximal ideal of B is contained in a unique maximal ordinary B-ideal of A.*

Proof. (i) This follows at once from Proposition 4.

(ii) Suppose that M is a maximal ideal of B, contained in two distinct maximal ordinary B-ideals of A, J and J'. Let $a \in A$ be such that $f_J(a) \neq f_{J'}(a)$, and $\alpha \in K$ such that $b=(a-\alpha e)^{-1}$ is in B. Then

$$f_M(b)=f_J(b)=(f_J(a)-\alpha)^{-1} \neq (f_{J'}(a)-\alpha)^{-1}=f_{J'}(b)=f_M(b)\,,$$

which is a contradiction.

5. The B-radical, semi-simplicity, etc. The theory given in this section is based on the definition of the B-radical of A as the intersection of all maximal ordinary B-ideals of A. There is, of course, a parallel theory based on the definition of the radical as the intersection of all maximal B-ideals; this set is a B-ideal but not in general an ordinary B-ideal. The two theories resemble each other so closely that there seems to be no point in writing out both sets of results explicitly.

DEFINITION 5. The intersection of all maximal ordinary B-ideals of A is the *B-radical* of A.

It is evident that the B-radical is an ordinary B-ideal.

PROPOSITION 6. *If $\tau'_B(a)=\{0\}$ implies $a \in B$ (in particular, if B is strongly saturated) then the B-radical of A consists of theose elements $b \in B$ such that $(e-\alpha b)$ has an inverse in B for all $\alpha \in K$.*

Proof. If a is in the B-radical then $\tau'_B(a)=\{0\}$, and $a \in B$, by assumption. If $e-\alpha a$ had no inverse in B, then $(e-\alpha a)B$ would be a proper ideal of B, and would be contained in a maximal ordinary B-ideal of A, by Lemma 1 (ii). If J is this B-ideal then $f_J(e-\alpha a)=0$, hence $f_J(a)=\alpha^{-1}\neq 0$, a contradiction. So $e-\alpha a$ has an inverse in B for each $\alpha \in K$.

On the other hand, if $a \in B$, and a is not in the B-radical, there will be a nonzero $\alpha \in K$ such that $f_J(a)=\alpha$ for some J. Then we cannot have $(e-\alpha^{-1}a)^{-1} \in B$; if this were so then $e=(e-\alpha^{-1}a).(e-\alpha^{-1}a)^{-1} \in JB \subseteqq J$, which is impossible.

DEFINITION 6. If the B-radical of A is $\{0\}$ then A is *B-semi-simple*. If, whenever $a\neq a'$ there is a maximal ordinary B-ideal J such that $f_J(a)\neq f_J(a')$, then A is *completely B-semi-simple*.

In the case of a Banach algebra, semi-simplicity implies complete semi-simplicity. Whether this is so in the present more general case remains an open question. We shall obtain partial results in this direction under restrictive hypotheses.

PROPOSITION 7. *If A is B-semi-simple, $a\neq a'$, and $\tau'_B(a)$ is not the whole of K', then there is a maximal ordinary B-ideal J such that $f_J(a)$ $\neq f_J(a')$.*

Proof. If $f_J(a)$ never takes the value ∞, then clearly $f_J(a)=f_J(a')$ for all maximal ordinary J implies $a=a'$, by the definition of B-semi-simplicity.

If $f_J(a)$ never takes the value α, then by Theorem 3, $(a-\alpha e)^{-1} \in B_2$. If $f_J(a)=f_J(a')$ for all J then $f_J((a-\alpha e)^{-1})=f_J((a'-\alpha e)^{-1})$ for all J, whence $(a-\alpha e)^{-1}=(a'-\alpha e)^{-1}$, by the assumed B-semi-simplicity. Hence $a-\alpha e=a'-\alpha e$, and $a=a'$.

LEMMA 10. *If A is B-semi-simple, and $\tau'_B(a)$ contains no nonzero elements of K, then $a=0$.*

Proof. Since K is assumed to be algebraically closed, there will be in K an element different from 0 and from 1; let α be any such element. Since $f_J(a)$ is never 1, $(a-e)^{-1}$ exists in B_2, by Theorem 3. Similarly, $(\alpha a-e)^{-1}$ exists in B_2. Clearly $f_J((a-e)^{-1})=f_J((\alpha a-e)^{-1})$ for all J ($=0$ if $f_J(a)=\infty$, $=-1$ if $f_J(a)=0$). Hence $(a-e)^{-1}=(\alpha a-e)^{-1}$, whence $a-e=\alpha a-e$, giving $a=0$.

LEMMA 11. *If $\tau'_B(a)$ contains a finite number of elements of K only, then it does not contain ∞, if A is B-semi-simple.*

Proof. Let $\alpha_1, \alpha_2, \cdots, \alpha_n$ be the elements of K in $\tau'_B(a)$. Then $f_J((a-\alpha_1 e)(a-\alpha_2 e)\cdots(a-\alpha_n e))=0$ or ∞ only. By Lemma 10 this implies that $(a-\alpha_1 e)(a-\alpha_2 e)\cdots(\alpha-\alpha_n e)=0$, and this is clearly inconsistent with $f_J(a)=\infty$ for any J.

COROLLARY. *If there are only a finite number of maximal ordinary B-ideals, then $f_J(a)$ is never ∞ for any $a \in A$, that is, $A=B_2$.*

It is clear that if we know that for each $a \in A$, $f_J(a)=\infty$ for a finite set of maximal ordinary B-ideals only, simplifications will result.

DEFINITION 7. The algebra A is of *finite type* (with respect to B) if for each $a \in A$ the function $f_J(a)$ is infinite on (at most) a finite set of maximal ordinary B-ideals J.

The algebra of rational functions of an indeterminate is evidently of finite type with respect to the sub-algebra of constants.

PROPOSITION 8. *If A is B-semi-simple and of finite type, it is completely B-semi-simple.*

Proof. Suppose that $f_J(a)=f_J(a')$ for all J. Then $f_J(a-a')$ takes a finite set of nonzero values at most. Hence, by Lemma 11, $f_J(a-a')$ is never ∞. If $a \neq a'$, then $a(a-a')$ would be such that (i) $f_J(a(a-a'))=\infty$ for some J; and (ii) $f_J(a(a-a'))$ takes a finite set of values in K

only—two contradictory properties. Hence $a=a'$.

There are two problems which are closely related to each other and to B-semi-simplicity. These are, broadly speaking, (i) on how large a set of maximal ordinary B-ideals can the function $f_J(a)$ be ∞? and (ii) on how small a set can $f_J(a)$ be nonzero, with $a\neq 0$? In the absence of B-semi-simplicity, of course, $f_J(a)$ may be ∞ for all maximal ordinary J, and $f_J(a')$ may be zero for all such J, with $a'\neq 0$ (consider the example of formal power-series discussed after Definition 4).

In the next two propositions we assume that A is B-semi-simple.

PROPOSITION 9. *Let $\mathscr{M}$ be a finite set of maximal ordinary B-ideals, and $a\in A$ an element such that $f_J(a)=\infty$ for $J\in\mathscr{M}$, and $f_J(a)\neq\infty$ for $J\notin\mathscr{M}$. Then, if $f_J(a')=0$ for all $J\notin\mathscr{M}$, it follows that $a'=0$.*

Proof. The function $f_J(aa')$ takes a finite set of values only, hence it is never ∞, by Lemma 11. This clearly implies that $f_J(a')=0$ for all J, and so $a'=0$, by the assumed B-semi-simplicity.

A somewhat similar result is the following.

PROPOSITION 10. *Let $\mathscr{M}$ be a set of maximal ordinary B-ideals such that there is an element $a\in A$ with $f_J(a)=\infty$ for $J\in\mathscr{M}$, and $f_J(a)\neq\infty$ for $J\notin\mathscr{M}$. Then there is no element $a'\in A$ with $f_J(a')\neq 0$ for $J\in\mathscr{M}$, and $f_J(a')=0$ for $J\notin\mathscr{M}$.*

Proof. If there were such an element a' then we would have $f_J(aa')=0$ or ∞ only, whence $aa'=0$, by Lemma 10. This contradicts $f_J(aa')=\infty$ for $J\in\mathscr{M}$.

If A is B-semi-simple, then Theorem 2 states that A is isomorphic' in a certain sense to an algebra of functions on the set of maximal ordinary B-ideals. In certain cases it is possible to assert that there is a genuine isomorphism between A and an algebra of equivalence-classes of functions. We introduce this as follows.

Let X be any set. We shall call a family $\mathscr{Q}$ of subsets of X a *Q-family* if (i) the union of two (and hence any finite number of) subsets of $\mathscr{Q}$ is in $\mathscr{Q}$; (ii) X is not in $\mathscr{Q}$. For example, if X is the real interval (0, 1), the subsets of measure zero form a Q-family. Take now the set S of functions defined on X, with values in K', which are finite outside a set of $\mathscr{Q}$. Let T be the set of functions which are zero outside a set of $\mathscr{Q}$. Let $(S;\ T)$ be the set of equivalence-classes of functions of S, modulo functions of T. Then, in the familiar way, $(S;\ T)$ can be made into an algebra by defining the sum of two classes to be the class determined by the sum of two functions, one from each class, etc.; it is easy to verify that the algebraic operations are well-defined. The object of condition (ii) is to ensure that the resulting algebra is nontrivial.

DEFINITION 8. If X is the set of maximal ordinary B-ideals of A, and there is a Q-family of subsets of X such that A is isomorphic to $(S;\ T)$, as defined above, then A has a *Q-representation.*

THEOREM 6. *If A is B-semi-simple, and of finite type, it has a Q-representation.*

Proof. If there are finitely many maximal ordinary B-ideals, then $f_J(a)$ is always finite, by Lemma 11, Corollary. Then result follows at once in this case, taking the Q-family consisting of the empty set only.

If there are infinitely many maximal ordinary B-ideals, then it is easy to verify that the family of sets on which $f_J(a)$ is infinite for some $a \in A$ forms a Q-family. The required result then follows from Proposition 9.

It would be of considerable interest to extend the above results, in particular, to remove the qualification 'finite' from the set $\mathscr{M}$ in Proposition 9. In § 7 we shall do this under additional hypotheses (Proposition 17). It is not evident that these restrictions are necessary for the validity of the result, and more information on the point would be welcome. There is one partial result in this direction, as follows:

PROPOSITION 11. *If A is completely B-semi-simple, then $f_J(a)=\infty$ for $J \in \mathscr{M}$, $f_J(a')=0$ for $J \notin \mathscr{M}$, $a' \in B$ together imply $a'=0$.*

Proof. Immediate.

6. Algebras over topological fields. We now consider the case of a field K with a topology. We are primarily interested in the complex case, but it is as easy to write out the results for much more general fields. We require very little of the topology; the essential feature is that it should provide a reasonable definition of 'bounded' subsets of K. We shall assume (until after Proposition 15) that K is a topological field in the sense of Bourbaki, that is, that the topology is Hausdorff and the algebraic operations are continuous.

We adopt the definition of boundedness given by Shafarevich; the subset H of K is bounded if, given any neighborhood N of 0, there is a neighborhood N' of 0 such that $HN' \subseteq N$. It is trivial that the union, sum and product of two bounded subsets of K are again bounded subsets. We shall further assume (again until after Proposition 15) that K is of type V, in the sense of Kaplansky; that is, if the set S is disjoint from some neighborhood of 0, then the set of inverses S^{-1} is bounded. We assume that K is not discrete; if K is discrete then every subset of K is closed and bounded, and the results reduce to those of § 4.

DEFINITION 9. Denote by B_1 the set of elements of A which have a bounded B-spectroid, and by B_1' the set of elements with a bounded B-spectrum. If $B_1 \supseteq B$ then B is *weakly bounded;* if $B_1' \supseteq B$ then B is *strongly bounded.* If $B_1 \subseteq B$, then B is *strongly boundedly saturated;* if $B_1' \subseteq B$ then B is *weakly boundedly saturated.*

It is evident that $B_1 \supseteq B_1'$; B_1 is clearly a sub-algebra of A, by Lemma 3, but B_1' is not a sub-algebra of A in general.

For the remainder of this section we shall assume that B is weakly bounded, unless the contrary is explicitly stated.

PROPOSITION 12. (i) *The maximal ordinary B_1-ideals of A are the same as the maximal ordinary B-ideals.*

(ii) *If M is any maximal ideal of B_1, then $B_1/M \cong K$.*

(iii) $(B_1)_1 = B_1$, *for any B.*

Proof. This is analogous to that of Proposition 1.

It is always possible, for any given A, to chose a strongly bounded sub-algebra B; take $B = Ke$. Also, it is always possible to choose a strongly boundedly saturated B; take $B = (Ke)_1$.

If B is not weakly bounded, there may be B_1-ideals of A which are not B-ideals. For example, let I be any infinite index-set, and A the algebra of complex-valued functions defined on I, $\{a_i\}_{i \in I}$, with pointwise addition and multiplication. Take $B = A$; then B_1 is the set of all bounded functions on I. Any function $a \in A$ with $a_i \neq 0$ for all $i \in I$, but $\inf_{i \in I} |a_i| = 0$, will be in an ordinary B_1-ideal but in no proper B-ideal of A.

LEMMA 12. *If 0 adheres to $\sigma_B'(a)$ then $0 \in \sigma_B'(a)$; if 0 adheres to*

$\tau'_B(a)$ then $0 \in \tau'_B(a)$.

Proof. We shall prove the second statement only; the proof of the first is similar and slightly simpler. The set $aB + a^2B + a^3B + \cdots$ is clearly an ordinary B-ideal of A. If e were in this B-ideal, then

$$e = b_1 a + b_2 a^2 + \cdots + b_n a^n$$

for some $b_1, b_2, \cdots, b_n \in B$. It is elementary to verify that if 0 adheres to $\tau'_B(a)$ then it adheres to $\tau'_B(b_1 a + \cdots + b_n a^n)$ also. Thus it would adhere to $\tau'_B(e) = \{1\}$. This is impossible, since the topology of K is Hausdorff. So the B-ideal specified above is admissible, and there is a maximal ordinary B-ideal, J say, containing it. Thus $a = ae \in J$, and $f_J(a) = 0$, so that $0 \in \tau'_B(a)$, as asserted.

THEOREM 7. *For each $a \in A$, $\sigma_B(a)$ and $\tau_B(a)$ are closed subsets of K.*

Proof. If $\alpha \in K$ adheres to $\sigma_B(a)$ then clearly 0 adheres to $\sigma_B(a - \alpha e)$; hence, by Lemma 12, 0 is in $\sigma_B(a - \alpha e)$ and so α is in $\sigma_B(a)$. Similarly for $\tau_B(a)$.

We may topologise K' by taking the basic neighborhoods of ∞ to be the complements in K' of the bounded subsets of K. In this topology, $\tau'_B(a)$ and $\sigma'_B(a)$ are not in general closed in K'. Example (iii) of § 1 shows that we may have ∞ adherent to $\tau'_B(a)$, but no maximal ordinary J such that $f_J(a) = \infty$.

THEOREM 8. *If a is in no maximal ordinary B-ideal of A then a has an inverse in B_1.*

Proof. (1) As in Theorem 3 (ii), a has an inverse a^{-1} in A. Since $f_J(a)$ is never zero, there is a neighborhood N of 0 such that $N \cap \tau'_B(a)$ is empty, by Theorem 7. Since we assume that K is of type V, this implies that the set of inverses of elements of $\tau'_B(a)$ is bounded; but this set of inverses is evidently the B-spectroid of a^{-1}; hence $a^{-1} \in B_1$.

COROLLARY. *If B is strongly boundedly saturated, then $a \in A$ has an inverse $a^{-1} \in B$ if and only if a is in no maximal ordinary B-ideal of A.*

PROPOSITION 13. *If $a \in B_2$, $a \notin B$, and B is strongly boundedly saturated, then $\tau'_B(a) = K$.*

Proof. This follows from Theorem 8, Corollary, just as Proposition 3 follows from Theorem 3 (i).

We note that if B is strongly boundedly saturated, the hypothesis of Proposition 6 is satisfied, and hence the conclusion of the proposition is valid.

PROPOSITION 14. *If $p(a) \in B_1$ for some polynomial p of degree $\geqq 1$, then $a \in B_1$.*

Proof. Since $\tau'_B(p(a))$ is bounded, we can choose $\alpha \in K$ so that $0 \notin \tau'_B(q(a))$, where $q(a) = p(a) - \alpha e$. Thus $q(a) \in B_1$ and $(q(a))^{-1}$ exists. Write $(q(a))^{-1}$ as a sum of terms of the type $\beta_{rs}(a - \alpha_r e)^{-s}$, it is a matter of routine to verify that if $\tau'_B(a)$ is unbounded, then 0 adheres to $\tau'_B((q(a))^{-1})$. This contradicts the fact that $q(a) \in B_1$, and $\tau'_B((q(a))^{-1}) = \{\tau'_B(q(a))\}^{-1}$.

The corresponding result, with B_2 in place of B_1, is true and trivial.

The following result is analogous to Theorem 4, Corollary, and is proved in exactly the same way:

PROPOSITION 15. *A necessary and sufficient condition that the rational function $r(a)$ exists as an element of B_1 is that $r(\tau'_B(a))$ is a bounded subset of K.*

For the remainder of this section we assume only that the field K has a Hausdorff topology, and that addition is continuous. We may topologise K' by taking the neighborhoods of ∞ to be the complements in K' of the bounded subsets of K. It is possible to introduce a topology on the maximal ordinary B-ideals of A in at least three obvious ways:

(i) Take as basic neighborhoods of the maximal ordinary B-ideal J_0 the sets $\{J: f_J(a_r) \in N,\ r = 1, 2, \cdots, n\}$, where N is any neighborhood of 0 and $a_1, a_2, \cdots, a_n$ are any elements of J_0. This clearly defines a Hausdorff topology in which each function $f_J(a)$ is continuous (as a function of J) wherever it is finite. In particular all functions representing elements of B_2 are continuous everywhere.

(ii) Take as basic neighborhoods of J_0 the sets

$$\{J:\ f_J(a_r) \in N_r(f_{J_0}(a_r)),\ r = 1, 2, \cdots, n\},$$

where $a_1, a_2, \cdots, a_n$ are any elements of A and $N_1, N_2, \cdots, N_n$ are any neighborhoods of $f_{J_0}(a_1), f_{J_0}(a_2), \cdots, f_{J_0}(a_n)$ respectively. This is the weakest topology in which all the functions $f_J(a)$ are continuous. It is evidently finer than (i).

(iii) Take as basic neighborhoods of J_0 the sets $\{J: f_J(b_r) \in N, r=1, 2, \cdots, n\}$, where N is any neighborhood of 0 and $b_1, b_2, \cdots, b_n$ are any elements of $J_0 \cap B$.

Other variations are possible; for instance, B may be replaced by B_1 or B_2 in (iii). We shall refer to these variations as (iii′), (iii″), respectively.

In general, topology (iii) will not be Hausdorff; a necessary and sufficient condition that it should be so is that each maximal ideal of B should be contained in precisely one maximal ordinary B-ideal of A. If the topology is Hausdorff, then the set $\mathscr{J}_0'$ of maximal ordinary B-ideals of A is compact; Gelfand's proof of the corresponding result for Banach algebras [4, Satz 9] applies to the present case. Similar remarks apply to (iii′) and (iii″).

In the case where A is a Banach algebra, and $B=A$, all the above topologies reduce to the customary Gelfand topology on the maximal ideals. In the context of § 8, topology (ii) seems the most appropriate.

Similar topologies could of course be imposed on the space of all maximal B-ideals of A.

7. Self-adjoint algebras. As in § 5, we use the maximal ordinary B-ideals; similar results could be obtained, starting from the maximal B-ideals. In this section the scalar field is taken to be the complex field C. The results could be formulated in a more general situation (in a field with a suitable 'conjugation'), but there seems to be no point in doing this. Asterisks applied to scalars denote complex conjugates, and $\infty^* = \infty$.

DEFINITION 10. The algebra A is *self-adjoint* (with respect to B) if, given $a \in A$, there exists $a^* \in A$ (not necessarily unique) such that $f_J(a^*) = f_J(a)^*$ for each maximal ordinary J.

From now on it is assumed that A is self-adjoint and B-semi-simple.

PROPOSITION 16. *The algebra A is completely B-semi-simple.*

Proof. Suppose that $f_J(a) = f_J(a')$ for all J. Then evidently $(e+aa^*)^{-1}$ and $(e+a'a'^*)^{-1}$ both exist (in B_1) and $f_J((e+aa^*)^{-1}) = f_J((e+a'a'^*)^{-1})$ for all J. Hence the two inverses are equal, by the assumed B-semi-simplicity, and this implies $aa^*=a'a'^*$. Next, it is easy to verify that $a(e+aa^*)^{-1}$ and $a'(e+a'a'^*)^{-1}$ are both in B_1, and $f_J(a(e+aa^*)^{-1})=f_J(a'(e+a'a'^*)^{-1})$ for all J. Hence the two elements are equal, and the conclusion $a=a'$ is immediate.

COROLLARY. *The element a^* is unique.*

It is clear that $a^*=0$ implies $a=0$, and $aa^*=0$ implies $a=0$.

The next result is, as promised in § 5, an improvement of Proposition 9 in the present special case:

PROPOSITION 17. *If $\mathscr{M}$ is a set of maximal ordinary B-ideals and $a\in A$ is such that $f_J(a)=\infty$ for $J\in\mathscr{M}$, $f_J(a)\neq\infty$ for $J\notin\mathscr{M}$, then $f_J(a')=0$ for $J\notin\mathscr{M}$ implies $a'=0$.*

Proof. Since $f_J(a'a'^*+e)$ is either real and $\geqq 1$, or is infinite, it is clear that $f_J(aa'a'^*+a)=f_J(a)$ for $J\in\mathscr{M}$; and since $f_J(a')=0$ for $J\notin\mathscr{M}$, the same equation holds for $J\notin\mathscr{M}$ also. Hence $aa'a'^*=0$, by Proposition 16. But this implies $a'=0$; if not, there would be a B-ideal $J\in\mathscr{M}$ with $f_J(a')\neq 0$, which would imply $f_J(aa'a'^*)=\infty$, which contradicts $aa'a'^*=0$.

THEOREM 9. *If A is B-semi-simple and self-adjoint, it has a Q-representation.*

Proof. If $f_J(a)=\infty$ for $J\in\mathscr{M}$, and $f_J(a')=\infty$ for $J\in\mathscr{M}'$, it follows that if $a''=(e+aa^*)(e+a'a'^*)$ then $f_J(a'')=\infty$ for $J\in\mathscr{M}\cup\mathscr{M}'$. Also, $f_J(a)$ cannot be infinite for all maximal ordinary B-ideals J, by Lemma 10. Hence the family of sets on which $f_J(a)$ is infinite for some $a\in A$ is a Q-family. The required result now follows from Proposition 17.

So far the topology of C has not been involved; it is essential for the results which follow. From now on we suppose $B=B_1$, that is, B is weakly bounded and strongly boundedly saturated (Definition 9). In the absence of this assumption the following results remain true, when suitably modified. But the statements then become more complicated, and the gain in generality is not significant.

PROPOSITION 18. $B=B_2$.

Proof. For any a, $f_J(e+aa^*)$ is never zero, and so $(e+aa^*)^{-1}$ exists. If $a \in B_2$, then $f_J((e+aa^*)^{-1})$ is never zero. By Lemma 12, it is therefore bounded away from zero, and so $f_J(a)$ is bounded away from infinity. Thus $a \in B_1 = B$. Since $B_2 \supseteqq B$, the theorem follows.

COROLLARY. *$\tau'_B(a)$ is closed in C' for any a.*

LEMMA 13. *Each maximal ideal of B is contained in exactly one maximal ordinary B-ideal of A.*

Proof. Suppose that the maximal ideal M of B is contained in the distinct maximal ordinary B-ideals, J and J', of A. Let $a \in J$, $a \notin J'$. Then

$$f_M((e+aa^*)^{-1}) = (1+|f_J(a)|^2)^{-1} \neq (1+|f_{J'}(a)|^2)^{-1} = f_M((e+aa^*)^{-1}),$$

a contradiction.

LEMMA 14. *All the topologies described in § 6 are equivalent.*

Proof. It is clearly sufficient to prove that (iii) is finer than (ii). Let N be the neighborhood

$$N = \{J\colon\ f_J(a_r) \in N_r(f_{J_0}(a_r)),\ r=1, 2, \cdots, n\} .$$

Write

$$b_r = (e+a_r a_r^*)^{-1} ;$$

it is easy to find neighborhoods N'_r, N''_r such that

$$f_J(b_r) \in N'_r(f_{J_0}(b_r)),\ f_J(a_r b_r) \in N''_r(f_{J_0}(a_r b_r))$$

together imply

$$f_J(a_r) \in N_r(f_{J_0}(a_r))$$

(if $f_{J_0}(a_r) = \infty$, then N''_r is superfluous). By translating the neighborhoods N'_r, N''_r to the origin if necessary, and taking their intersection, it is easily seen that there is a neighborhood in topology (iii) which is contained in N.

Combining the above results, we obtain at once:

THEOREM 10. *Let A be a self-adjoint, B-semi-simple algebra, with $B = B_1$. Then the maximal ordinary B-ideals of A can be topologised so as to become a compact Hausdorff space, and the mapping $a \to f_J(a)$*

sends elements of A into continuous C'-valued functions on this space. The sets on which $f_J(a)$ is infinite for some $a \in A$ form a Q-family of closed sets.

It is apparent that the structure space depends (set-wise and topologically) only on the 'bounded' sub-algebra B of A, provided that this satisfies reasonable conditions, which ensure that it is large enough. If we assume a little more, namely that all bounded continuous functions correspond to elements of B (for instance, if B is a Banach algebra under a suitable norm), then we can clearly assert that the set on which $f_J(a)$ is infinite is nowhere dense (since the set is closed, this is equivalent to its interior being empty). The conclusion of Theorem 10 is thus strengthened.

To conclude this section we turn to Example (vi) of § 1, and see to what extent the results of this section can be applied to it. First, it seems desirable to state precisely what we mean by an algebra of normal operators on a Hilbert space; we mean a collection A of normal operators such that any scalar multiple of an operator in A is in A, and the sum and product of any two operators in A have unique extensions in A. As always, we assume that A contains a unit (the identity operator, here) and is commutative (in the sense that the product of two operators, in a certain order, has the same extension in A as the product in the reverse order). We take B to consist of the bounded operators in A: we assume that if $a \in A$, and a^{-1} exists as a bounded operator, then $a^{-1} \in B$, and we also assume that B is uniformly closed. This implies that the maximal ideal condition $B/M \cong C$ is satisfied. If we denote by a^* the usual Hilbert space adjoint of a (we proceed immediately to show that this is in agreement with the previous use of a^*), and restrict attention to algebras A which are self-adjoint in the sense that $a \in A$ implies $a^* \in A$, then we have the following.

LEMMA 15. *The algebra A is self-adjoint in the sense of Definition* 10.

Proof. If a is bounded then it is clear that $f_J(a+a^*)$ is real, since $(a+a^*-\lambda e)$ has an inverse in B for nonreal λ. Similarly, $f_J(a-a^*)$ is imaginary, and so $f_J(a^*)=f_J(a)^*$ for bounded a. Next, for any $a \in A$, write $b=(e+aa^*)^{-1}$; it is well known that $b \in B$ and $ab \in B$; also b is self-adjoint ($b^*=b$), and $(ab)^*=a^*b$. If $f_J(a)$ and $f_J(a^*)$ are both finite, then from $f_J(a^*b)=f_J(ab)^*$ it follows that $f_J(a^*)=f_J(a)^*$, since $f_J(b)$ is real and nonzero. It remains to show that if one of $f_J(a)$, $f_J(a^*)$ is finite, then so is the other. Suppose the contrary; there is no loss of generality in supposing $f_J(a)=1$, $f_J(a^*)=\infty$. Then

$$1=f_J(e-b)=f_J(aa^*b)=f_J(a)f_J(a^*b)=f_J(a)f_J(ab)^*=0\,,$$

which is impossible.

The sub-algebra B is semi-simple, by the usual reasoning, and the B-semi-simplicity of A follows at once from this. The conclusions of Lemma 13 and 14 are true, independently of any assumption that $B=B_1$, since $b=(e+aa^*)^{-1}$ and ab are certainly in B.

The conclusion of Theorem 10 is thus valid for A. Moreover, the assumption that B is uniformly closed ensures that the functions $f_J(a)$ become infinite only on nowhere dense sets. The fact that $B=B_1$ follows from the same assumption; for each bounded continuous function on the maximal B-ideals of A (=maximal ideals of B) corresponds to an element of B.

In the paper of Fell and Kelley [3], the authors deal with algebras of operators from a somewhat different point of view. Starting from a strongly closed algebra of bounded of bounded operators, they select a class of unbounded functions on the structure space (the same class as we have obtained above, namely the continuous C'-valued functions infinite only on a nowhere dense set), and show that to each such function there corresponds a normal operator. Every normal operator can be obtained in this way, starting from a suitable algebra of bounded operators. The problem of the functional representation of an algebra of operators is not explicitly treated.

As a realization of the sort of algebra we have been considering, take the following trivial example. Let the Hilbert space be L_2 (0, 1), and consider continuous C'-valued functions on (0, 1) which are infinite only on a set with empty interior. To each such function a normal operator can be attached in an obvious way; the operator, applied to a function of L_2, yields the ordinary product of the two functions. If we assume that A is an algebra of such operators, containing all operators corresponding to bounded functions, then the above theory can be applied, and it is found that the operators are represented by the functions from which they have arisen.

8. Algebraic function fields. Although it is not our main objective, we give a few indications of the relation between the theory developed in the preceding sections, and the theory of fields of algebraic functions of one variable. All the relevant definitions, etc., will be found in Chevalley's book [2]. The first result is valid quite generally.

LEMMA 16. *If A is a field, and K is a proper sub-field of A, then for every maximal ordinary K-ideal J of A there is an element $a\in A$ with $f_J(a)=\infty$.*

Proof. If not, J would be a proper ideal of A, different from $\{0\}$, by Lemma 3; this is impossible.

The definition of a V-ring, as required in the next lemma, will be found in [2, p. 1].

THEOREM 11. *If A is a field of algebraic functions of one variable and K is an algebraically closed proper sub-field of A, then the maximal ordinary K-ideals of A are in one-to-one correspondence with the V-rings in A (over K).*

Proof. Let J be a maximal ordinary K-ideal of A, and write $Q = J+K$. Then clearly Q is a ring; further (i) Q contains K; (ii) $Q \neq A$, since, by Lemma 16 there is an element $a \in A$ with $f_J(a)=\infty$; and $f_J(q) \in K$ for all $q \in Q$; (iii) if $x \notin Q$ then $f_J(x)=\infty$; for if $f_J(x)=\alpha \in K$ then $x-\alpha e \in Q$ and so $x \in Q$. If $f_J(x)=\infty$ then $f_J(x^{-1})=0$, by Lemma 9, and so $x^{-1} \in Q$. Thus Q is a V-ring.

On the other hand, let Q be any V-ring, and let J be the ideal of non-units. Then $Q/J \cong K$ ([2], p. 10); every element of Q is of the form $\alpha e+j$, where $j \in J$. Clearly J is an ordinary B-ideal; we now show that it is maximal. Let a be any element of A, not in J. Then if $a \in Q$, $a-\alpha e \in J$ for some $\alpha \in K$ and so e is in the K-ideal generated by J and a. If a is not in Q then a^{-1} must be in J; for if a^{-1} were in Q but not in J then a^{-1}, and hence a, would be a unit in Q. So again e is in the K-ideal generated by J and a. That is, J is maximal. This establishes the required correspondence.

We may thus identify the maximal ordinary K-ideals in A with the places of A, where A is a field of algebraic functions of one variable over K. The value taken by $a \in A$ at the place J [2, p. 6] is the same as the value of the function $f_J(a)$ as defined in § 2.

The places of A may be topologised, if K is a topological field; the topology (ii) previously indicated (§ 6) reduces to that given by Chevalley for the complex case [2, p. 133].

REFERENCES

1. R. Arens, *A generalization of normed rings*, Pacific J. Math., **2** (1952), 455–471.
2. C. Chevalley, *Algebraic functions of one variable*, New York, 1951.
3. J. M. G. Fell and J. L. Kelley, *An algebra of unbounded operators*, Proc. Nat. Acad. Sci. **38** (1952), 592–598.
4. I. M. Gelfand, *Normierte Ringe*, Mat. Sbornik, N. S. **9** (51) (1941) 3–24.
5. I. Kaplansky, *Topological rings*, Amer. J. Math. **69** (1947), 153–183.
6. E. A. Michael, *Locally multiplicatively-convex topological algebras*, (Amer. Math. Soc. Memoir no. 11) 1952.

7. M. H. Stone, *Linear transformations in Hilbert space*, New York 1932.
8. L. Waelbroeck, *Le calcul symbolique dans les algèbres commutatives*, C. R. Acad. Sci. Paris **238** (1954), 556–558, 640–641, 762–764.
9. ———, *Le calcul symbolique dans les algèbres commutatives*, J. Math. Pures Appl. N. S. **33** (1954), 147–186.

THE INSTITUTE FOR ADVANCED STUDY, PRINCETON
AND QUEEN'S UNIVERSITY, BELFAST

Pacific Journal of Mathematics

Vol. 7, No. 2 February, 1957

William F. Donoghue, Jr., *The lattice of invariant subspaces of a completely continuous quasi-nilpotent transformation* 1031

Michael (Mihály) Fekete and J. L. Walsh, *Asymptotic behavior of restricted extremal polynomials and of their zeros* 1037

Shaul Foguel, *Biorthogonal systems in Banach spaces* 1065

David Gale, *A theorem on flows in networks* 1073

Ioan M. James, *On spaces with a multiplication* 1083

Richard Vincent Kadison and Isadore Manual Singer, *Three test problems in operator theory* 1101

Maurice Kennedy, *A convergence theorem for a certain class of Markoff processes* 1107

G. Kurepa, *On a new reciprocity, distribution and duality law* 1125

Richard Kenneth Lashof, *Lie algebras of locally compact groups* 1145

Calvin T. Long, *Note on normal numbers* 1163

M. Mikolás, *On certain sums generating the Dedekind sums and their reciprocity laws* 1167

Barrett O'Neill, *Induced homology homomorphisms for set-valued maps* 1179

Mary Ellen Rudin, *A topological characterization of sets of real numbers* 1185

M. Schiffer, *The Fredholm eigen values of plane domains* 1187

F. A. Valentine, *A three point convexity property* 1227

Alexander Doniphan Wallace, *The center of a compact lattice is totally disconnected* 1237

Alexander Doniphan Wallace, *Two theorems on topological lattices* 1239

G. T. Whyburn, *Dimension and non-density preservation of mappings* 1243

John Hunter Williamson, *On the functional representation of certain algebraic systems* 1251