Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 7, No. 1 **January, 1957**

Printed in Japan by Kokusai Bunken Insatsusha
(International Academic Printing Co., Ltd.), Tokyo, Japan

ZERO-DIMENSIONAL COMPACT GROUPS OF HOMEOMORPHISMS

R. D. ANDERSON

1. Introduction. All spaces and topological groups referred to in this paper will be compact and metric. All topological groups will additionally be zero-dimensional, that is, either finite or homeomorphic to a Cantor set. As general references we cite Zippin [6] and Montgomery and Zippin [4]. Several of our definitions are similar to those in [6].

A *topological transformation group* of a topological space is an association of a topological group G and a topological space E in the sense that each element g of G and point x of E determine a unique point of E. If this point be called x', we write $gx=x'$. The association is subject to the following conditions:

(1) if e denotes the identity of G, $ex=x$ for all $x \in E$,

(2) $g(g'x)=(gg')x$, $g, g' \in G$, $x \in E$, and

(3) gx is continuous simultaneously in g and x.

Each element of G may, under the association, be regarded as a homeomorphism of E onto itself.

The topological transformation group G is said to be *effective* if for each $g \in G$ not the identity, there is an $x_g \in E$ for which $gx_g \neq x_g$ and is said to be *strongly effective* (or *fixed-point-free*) if for each $g \in G$ not the identity and for each $x \in E$, $gx \neq x$. We shall use the symbol $Tg(G, E)$ to denote a particular association of G with E such that G is an effective topological transformation group of E. Thus by $Tg(G, E)$ we mean a particular group of homeomorphisms of E onto itself, the group being isomorphic to and identified with G. If $Tg(G, E)$ is strongly effective we write $TgS(G, E)$.

For $x \in E$, $G(x)$ will denote the set of all images of x under G and will be called the orbit of x under G. Similarly for $X \subset E$, $G(X)$ will denote the set of images of X under G. The individual orbits may be regarded as the "points" of a space, the orbit space, $O[Tg(G, E)]$ of $Tg(G, E)$. $O[Tg(G, E)]$ is a continuous decomposition of E.

The main purpose of this paper is to prove the following theorems:

THEOREM 1. *Let G be any compact zero-dimensional topological group. Let M be the universal curve.*[1] *Then there exists a $TgS(G, M)$*

Received May 11, 1956. Presented to the American Mathematical Society August 1956, and in part, December 1954. The research leading to this paper was supported in part by National Science Foundation Grant G 1013.

[1] The universal curve is a particular one-dimensional locally connected continuum. Its description and a characterization of it are given in § 3.

such that $O[TgS(G, M)]$ *is homeomorphic to* M.

THEOREM 2. *Let* G *be any infinite compact zero-dimensional topological group. Let* M *be the universal curve. Then there exists a* $TgS(G, M)$ *such that* $O[TgS(G, M)]$ *is a regular curve*[2].

Theorem 1 asserts that the universal curve is also universal in the sense that every compact zero-dimensional group can operate on it in a fixed-point-free fashion. It is well known and is easy to prove—see Example 1—that the Cantor set also has this property.

The following two theorems are corollaries of some of the methods used in the proofs of theorems 1 and 2. In particular, the argument of § 5 gives the essential structure of an argument for Theorem 3. Theorem 4 is a corollary of Theorem 3.

THEOREM 3. *Let* G *be any finite group. Then there exists in* E^3 *a 3-manifold* M *with connected boundary such that* $TgS(G, M)$ *exists.*

THEOREM 4. *Let* G *be any finite group. Then there exists in* E^3 *a 2-manifold* K *(without boundary) such that* $TgS(G, K)$ *exists.*

Any zero-dimensional compact group G can be expressed as the inverse (or projective) limit (simultaneously in both a topological and a group sense) of a sequence $\{G_i\}$ of finite groups under a sequence $\{\pi_i\}$ of homomorphisms with, for each i, π_i carrying G_{i+1} onto G_i (see §§ 2.5–2.7 of [4]). The group G is said to be p-adic if, for each i, G_i can be taken as a cyclic group with, for each i, π_i not an isomorphism. If G is a p-adic group and sequences $\{G_i\}$ and $\{\pi_i\}$ exist such that, for each i, π_i is two-to-one then G is called the dyadic group.

AGREEMENT 1. *We shall assume henceforth that* G *is a particular compact zero-dimensional topological group.*

AGREEMENT 2. *We shall assume that sequences* $\{G_i\}$ *and* $\{\pi_i\}$ *with respect to which* G *is an inverse limit are given and to avoid subdivision of the ensuing arguments into cases we shall further assume that* G *is infinite and that, for no* i, *is* π_i *an isomorphism.*

It will be clear that the argument we give for Theorem 1 actually includes the essentials of the argument for the case of G finite.

[2] A locally connected continuum is said to be a regular curve provided every point of it has arbitrarily small neighborhoods with finite boundaries or, equivalently, provided every pair of points of it can be separated by a finite point set,

NOTATION. Let e be the identity of G and, for each i, let e_i be the identity of G_i. For each i, let $n(G_i)$ be the number of elements in G_i.

REMARKS. At the heart of the theory of topological transformation groups is the open question as to whether any infinite compact zero-dimensional group can operate effectively on a Euclidean manifold E. In studying such a question it is natural to consider the "nice" spaces on which such a group can operate and to consider the characteristics of the group operation[3]. Zippin [6] has observed that the known examples of even the dyadic group D effective on locally connected continua involve a type of "branching" about subsets on which D is not strongly effective, and, in fact, usually a type of "branching" about points or sets which have periodic orbits under G (see Example 2). Thus our theorems and arguments contribute to the knowledge of the ways zero-dimensional infinite compact groups can operate on locally connected continua. In this connection, we also note in Example 3 that any p-adic group can be strongly effective on the infinite dimensional compact torus.

We mention the following questions: For E a continuum and G infinite, is it possible for $TgS(G, E)$ to be such that the dimension of $O[TgS(G, E)]$ exceeds the dimension of E? If such is possible, can E be one-dimensional?, locally connected?, the universal curve?, locally Euclidean? What are conditions on E for which $\dim(O[TgS(G, E)])$ must be $\leqq \dim E$?

In the classic example of Kolmogoroff [3], G (not made explicit by him) operated effectively but not strongly effectively on a one-dimensional locally connected continuum E, and $O[Tg(G, E)]$ was two-dimensional. The more recent example by Keldys [2] of a light open mapping of a one-dimensional continuum onto a square also involved a "branching" type operation.

2. Examples. In this section we wish to give three examples of topological transformation groups. Of these A and B, at least, are

[3] Smith, in [5], states "There exist, however, nearly periodic transformations which are not periodic. In all known examples the space M under transformation is of a highly irregular local structure which suggests the problem referred to above: Can there exist a non-periodic nearly periodic transformation T operating in M if M is fairly regular in its local structure, for example, locally Euclidean." If G is a p-adic group, if $TgS(G, M)$ exists, and if $g \in G$ with $g \neq e$, then g as a homeomorphism of M is a non-periodic nearly periodic transformation. As the universal curve is homogeneous, it is, in a sense, fairly regular in its local structure and thus our Theorems 1 and 2 contribute to this question of Smith.

well known.

A. The group G can operate on itself as follows: for each $g, h \in G$ with h thought of as a point of a space, $gh=h'$ where h' is the group-theoretic gh. With this definition G is transitive on itself. For each h, $h' \in G$ there is one (and only one) element $g \in G$ for which $gh=h'$.

If, contrary to our Agreement 2, G is finite then G can operate on itself in this same way and also G can operate on a Cantor set C as follows: let H be a collection of disjoint open and closed subsets of C such that[4] $H^*=C$ and H admits a one-to-one transformation φ onto G. For some $h \in H$ and any $g \in G$ let ρ_g be a homeomorphism of h onto $\varphi^{-1}(g\varphi(h))$ with ρ_e the identity on h. For any point $p \in C$, there exists a $g' \in G$ such that $\rho_{g'}^{-1}(p) \in h$. Define gp to be $\rho_{g''}(\rho_{g'}^{-1}(p))$ where $g''=gg'$. The technique which we use here is similar to one we shall use for Lemma 2 later in the argument for Theorems 1 and 2.

B. In this example we show that G can operate on a locally connected continuum in the plane, in fact, on a tree, the particular tree, however, depending on G. Let I be the unit interval $0 \leqq x \leqq 1$, $y=0$. Let K_1 be a collection on $n(G_1)$ disjoint subintervals of I formed by choosing every other element of a subdivision of I into $2n(G_1)-1$ equal subintervals. Inductively, for each $i>1$, let K_i be a collection of $n(G_i)$ disjoint subintervals of I formed by choosing every other one of a subdivision of each interval of K_{i-1} into $2\left(\frac{n(G_i)}{n(G_{i-1})}\right)-1$ equal subintervals. Then $\bigcap_i K_i^*$ is a Cantor set C which may, in the obvious way, be identified with G.

For each i, let Q_i be a set of $n(G_i)$ points on $y=2^{-i}$ such that for each element k of K_i, Q_i contains a point $q(k)$ whose x-coordinate is the x-coordinate of the midpoint of k. Let Q_0 be the point $(\frac{1}{2}, 1)$. Let t be $\bigcup_{i\geqq 0} Q_i+\bigcap_{i\geqq 1} K_i^*+$for each $i\geqq 0$, the sum of all intervals with endpoints one in Q_i and the other in Q_{i+1} which project parallel to the y-axis into K_i^*. Then G may be considered as operating effectively but not strongly effectively on t such that the "branchings" of the operation of G on t occur at the points of $\bigcup_{i\geqq 0} Q_i$ and such that each point p of $t-C$ has a finite orbit under G consisting of those points of t on the horizontal line through p. In developing G we may consider that, for each i, G_i permutes the elements of K_i consistent with π_{i-1} and G_{i-1} permuting the elements of K_{i-1}.

C. Let G be a p-adic group and hence let, for each i, G_i be cyclic.

[4] If H is a collection of point sets, H^* denotes the sum of the elements of H.

Let E be the infinite dimensional compact torus $J_1 \times J_2 \times \cdots$ where, for each i, J_i may be thought of as the circle of radius 2^{-i} and center at (0, 0). Then $TgS(G, E)$ exists. For each i, let φ_i be the group of order n_i of rotations of J_i and let $Tg(G_i, E)$ be the cyclic group of order n_i on E defined coordinatewise as φ_j for $j \leqq i$ and as the identity for $j > i$. Then $TgS(G, E)$ may be defined coordinatewise as φ_i on J_i, for each i.

3. Definitions and the universal curve. Let N be the set of points in E^3 for which $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$, $0 \leqq z \leqq 1$. For $w = x, y, z$ and $i = 1, 2, \cdots$ let $D_i(w)$ be the set of all open intervals on the w-axis of length 3^{-i} whose endpoints have w-coordinates which are positive rational numbers less than 1, the expression for each such rational number having 3^i as a denominator when in lowest terms. The length of $D_i^*(w)$, for any i, is $\frac{1}{3}$. Let M be the set of all points (x, y, z) of N for which, for no i, do two or more of the points $(x, 0, 0)$, $(0, y, 0)$, and $(0, 0, z)$ belong to the set $D_i^*(x) + D_i^*(y) + D_i^*(z)$. The set M is called the *universal curve*.

It is not hard to verify that M is a locally connected one-dimensional continuum with no local separating points. M is called "the universal curve" as every one-dimensional continuum can be imbedded in it.

We need several further definitions before characterizing the universal curve. We use a special case of the characterization given in [1] with resultant simpler definitions than those of [1].

If H and H' are collections of point sets, H is said to be a *refinement* of H$'$ if each element of H is a subset of an element of H' and each element of H' contains an element of H. A collection H of point sets is said to be *one-dimensional* provided no three elements of H intersect.

A collection H of point sets is said to be *simple* provided that (1) H is finite, and H^* is connected, (2) each element of H is a (closed) 3-cell, and (3) if two elements of H intersect their intersection is a 2-cell on the bounding 2-sphere of each such element.

Let H and H' be simple collections with H a refinement of H'. Let h be an element of H' and let Z be the collection of those elements of H in h which intersect elements of H not in h. Then H is said to *interlace* h provided that for any subdivision of Z into disjoint sets Z_1 and Z_2 with $Z_1 + Z_2 = Z$ there exist non-null connected sums of elements of H in h, namely X_1 and X_2 with $X_1 \supset Z_1^*$, $X_2 \supset Z_2^*$, and X_1 and X_2 having no element of H in common. H is said to *interlace* H' if H interlaces each element of H'.

A sequence $\{F_i\}$ is said to be a *λ-defining sequence* of a continuum

M provided

(1) for each i, F_i is a simple one-dimensional collection covering M,

(2) for each i, F_{i+1} is a refinement of F_i,

(3) $M=\bigcap_i F_i^*$

(4)[5] for any $\varepsilon > 0$ there exists a number n such that $m(F_n) < \varepsilon$,

(5) for each i, F_{i+1} is interlaced in F_i, and

(6) if two elements of F_i intersect then each contains two elements of F_{i+1} intersecting elements of F_{i+1} in the other but neither contains any element of F_{i+1} intersecting two elements of F_i distinct from the one containing it.

A non-degenerate continuum for which there exists a λ-defining sequence is called a *C-set*.

The following theorem is proved in [1]:

THEOREM. *Each C-set is homeomorphic to the universal curve.*

NOTATION. If E_i is a finite collection of closed point sets and $Tg(G_i, E_i^*)$ or $TgS(G_i, E_i^*)$ is such that for $h \in E_i$, and any $g \in G_i$, gh is an element of E_i then we will write $Tg(G, E_i^*, E_i)$ or $TgS(G_i, E_i^*, E_i)$ respectively. If $\{E_i\}$ is a λ-defining sequence and $TgS(G_i, E_i^*, E_i)$ and $TgS(G_{i+1}, E_{i+1}^*, E_{i+1})$ exist, then $TgS(G_{i+1}, E_{i+1}^*, E_{i+1})$ is said to *refine* $TgS(G_i, E_i^*, E_i)$ provided that for any $g \in G_{i+1}$ and any $x \in E_{i+1}$, if x' denotes the element of E_i containing x, $\pi_i(g)x'$ contains gx.

AGREEMENT 3. *In what follows we shall make many constructions in E^3 using 3-cells and homeomorphisms. Every 3-cell used is to be polyhedral and every homeomorphism defined over finite sums of 3-cells is to be piecewise-linear, that is, is to carry polyhedra into polyhedra. We interpret this understanding to apply also to appropriate subsets (2-cells) and homeomorphism over these subsets, such being used in the constructions and lemmas. All constructions are to be in E^3.*

4. Statements of lemmas and proof that the lemmas imply Theorems 1 and 2.

LEMMA 1. *Let n be any positive integer. Let K and K' be elements of a simple one-dimensional collection of 3-cells in E^3. Let D and D' be collections of n disjoint 2-cells on the boundaries of K and K' respectively. Let φ be a homeomorphism of D^* onto D'^* preserving orientation on the elements of D and D' relative respectively to K and K' as embed-*

[5] If H is a finite collection of point sets, $m(H)$ denotes the mesh of H, that is, the l.u.b. of the diameters of H.

ded in E^3. *Then there exists an orientation-preserving*[6] *homeomorphism* ψ *of* K *onto* K' *such that for each point* $p \in D^*$, $\psi(p)=\varphi(p)$.

Proof. This lemma is geometrically obvious and is well known.

LEMMA 2. *Let, for any* i, $X_1, \cdots, X_{n(G_i)}$ *be a set* X *of disjoint continua all homeomorphic to each other. For each* j, $1 \leqq j \leqq n(G_i)$, *let* η_j *be a homeomorphism of* X_1 *onto* X_j *with* η_1 *the identity on* X_1. *Let* ρ_i *be a one-to-one transformation of* X *onto* G_i *with* $\rho_i X_j = g_j$ *for each* j. *Then* $TgS(G_g, X^*, X)$ *exists with* gx *defined as follows*:

$$\textit{for} \quad x \in X_j, \; g_j x = \eta_{k'} \eta_k^{-1} x \quad \textit{where} \quad g_{k'} = g_j g_k \,.$$

Proof of Lemma 2. This lemma is almost obvious and is well known. We state it separately to simplify the argument for Lemmas 3, 3′ and 3″. To prove the lemma it is sufficient to note that

$$g_{j_1}(g_{j_2}x) = (g_{j_1}g_{j_2})x \quad \text{for} \quad x \in x_k \quad \text{and} \quad g_{j_1}, \; g_{j_2} \in G_i,$$

$$g_{j_1}(g_{j_2}x) = g_{j_1}(\eta_{k'}\eta_k^{-1}x) = \eta_{k''}\eta_{k'}^{-1}\eta_{k'}\eta_k^{-1}x = \eta_{k''}\eta_k^{-1}x$$

where $g_{k'}$ is $g_{j_2}g_k$ and $g_{k''}$ is $g_{j_1}g_{k'}$. Therefore $g_{k''}$ is $(g_{j_1}g_{j_2})g_k$ as was to be shown.

LEMMA 3. *There exists a continuum* M *and a* λ-*defining sequence* $\{F_i\}$ *of* M *such that for each* i, $TgS(G_i, F_i^*, F_i)$ *exists with* $TgS(G_{i+1}, F_{i+1}^*, F_{i+1})$ *refining* $TgS(G, F_i^*, F_i)$ *and for each element* f *of* F_i, $G_i(f)$ *consists of* $n(G_i)$ *disjoint elements of* F_i.

LEMMA 3′. *The same as Lemma* 3 *with the added condition that there exist a* λ-*defining sequence* $\{H_i\}$ *and a sequence* $\{\mu_i\}$ *such that*

(1) *for each* i, μ_i *is a mapping of* F_i^* *onto* H_i^* *with for* $f \in F_i$, $\mu_i(f) \in H_i$ *and* μ_i *a homeomorphism over* f,

(2) *for any* $g \in G_i$ *and* $x \in F_i^*$, $\mu_i(x) = \mu_i(gx)$, *and*

(3) *for each* i, $f \in F_i$ *and* $\tilde{f} \in F_{i+1}$, $\mu_i(f) \supset \mu_{i+1}(\tilde{f})$ *if and only if* $f \supset \tilde{f}$.

LEMMA 3″. *The same as Lemma* 3 *with the added condition that there exists a sequence* $\{H_i\}$ *of simple collections and a sequence* $\{\mu_i\}$ *such that*

[6] Orientation-preserving with respect to embedding in E^3.

(1) *for each* i, μ_i *is an* $n(G_i)$*-to-one mapping of* F_i^* *onto* H_i^* *with for* $f \in F_i$, $\mu_i(f) \in H_i$ *and* μ_i *a homeomorphism over* f,

(2) *for any* $g \in G_i$ *and* $x \in F_i^*$, $\mu_i(x) = \mu_i(gx)$

(3) *for each* i, $f \in F_i$ *and* $\tilde{f} \in F_{i+1}$, $\mu_i(f) \supset \mu_{i+1}(\tilde{f})$ *if and only if* $f \supset \tilde{f}$

(4) *for each* $h, h' \in H_i$ *for which* $h \cdot h'$ *exists,* H_{i+1} *contains exactly one element in* h *intersecting an element of* H_{i+1} *in* h, *and*

(5) *for any* $\varepsilon > 0$ *there exists an* n *such that* $m(H_n) < \varepsilon$.

Before proving Lemmas 3, 3′ and 3′′ in §§ 5 and 6 we wish to note that Lemma 3 implies a weaker form of Theorem 1 to the effect that $TgS(G, M)$ exists, that Lemma 3′ implies the full strength of Theorem 1, and that Lemma 3′′ implies Theorem 2.

Clearly, from the characterization of the universal curve cited in § 3, $\bigcap_i F_i^* = M$ is a universal curve. Let $g \in G$. Then g is defined by a unique sequence $\{g_i\}$ with, for each i, $g_i \in G_i$ and $\pi_i g_{i+1} = g_i$. For any point $p \in M$, gp is defined as $\bigcap_i g_i f_i$ where $\{f_i\}$ is a sequence such that for each i, $f_i \in F_i$, $f_i \supset f_{i+1}$, and $p \in f_i$. But gp must be unique for $m(F_i) \to 0$ and if $\{f_i'\}$ is another such sequence then, for each i, $g_i f_i'$ intersects $g_i f_i$.

That such definition of the association of G and M satisfies the conditions of the definition of topological transformation group is straigt-forward. First, $ex = x$ for all $x \in M$ as, for each i, e_i leaves all elements of F_i fixed. Second, as for each i, g, $g' \in G_i$ and $f \in F_i$, $g(g'f) = (gg')f$, it follows that $g(g'x) = (gg')x$ for g, $g' \in G$ and $x \in E$. Third gx is continuous simultaneously in g and x. Let $g^j \to g$ in G and let $x^j \to x$ in M. We wish to show that $g^j x^j \to vx$ in M. Let $\varepsilon > 0$. Let k be an integer such that (1) $m(F_k) < \varepsilon$, (2) for all $i > k$, x^i is in an element of F_k containing x, and (3) for all $j > k$, $\pi_k g_{k+1}^j = \pi_k g_{k+1}$ where g_{k+1}^j and g_{k+1} are the elements of G_{k+1} of the sequences $\{g_\lambda^j\}$ and $\{g_\lambda\}$ defining g^j and g respectively. Then for all $j > k$, $g^j x^j$ is at a distance of less than ε from gx as was to be shown.

We have now established that Lemma 3 imphes the weak form of Theorem 1 and it remains to show that Lemmas 3′ and 3′′ establish additionally that $O[TgS(G, M)]$ is, in the first case, a universal curve and, in the second, a regular curve.

We wish to show next that $H = \bigcap_i H_i^*$ is homeomorphic to $O[TgS(G, M)]$ with $\{H_i\}$ and $TgS(G, M)$ as in either Lemma 3′ or Lemma 3′′. For any $x \in H$, let $\{h_i\}$ be a sequence such that, for each i, $h_i \supset h_{i+1}$, $h_i \in H_i$, and $x \in h_i$. But then there exists a sequence $\{f_i\}$ such that, for each i, $f_i \supset f_{i+1}$, $f_i \in F_i$, and $\mu_i(f_i) = h_i$. For $x \in H$, let $\nu(x) = G(\bigcap_i f_i)$ for such a sequences $\{f_i\}$. For any other such sequence $\{f_i'\}$, $G(\bigcap_i f_i')$ is $G(\bigcap_i f_i)$. As $m(H_i) \to 0$, $m(F_i) \to 0$, and for h_i, $h_i' \in H_i$ h_i intersects

h'_i if and only if and only if for any $f_i \in F_i$ with $\mu_i(f_i)=h_i$ there exists an f'_i with $\mu_i(f'_i)=h'_i$ and f_i intersecting f'_i, then it follows that ν is one-to-one onto. A standard argument shows the continuity of ν. Hence ν is the desired homeomorphism of H onto $O[TgS(G, M)]$.

Finally for Theorem 1 we note that by the condition that $\{H_i\}$ is a λ-defining sequence in Lemma 3′ it follows that H is a universal curve.

For Theorem 2 by Condition (4) of Lemma 3″ we note that if $p \in H$ and k_i denotes the sum of all elements of H_i containing p then for any i, $H \cdot k_i$ has only a finite number of points on its boundary with respect to H. Hence H is a regular curve.

5. The first step of the proof of Lemmas 3, 3′ and 3″. The demonstration of the existence of suitable F_1 and $TgS(G, F_1^*, F_1)$ is applicable to each of the Lemmas 3, 3′ and 3″ and thus only one argument need be given.

DEFINITION. Let S denote a set of k disjoint 3-cells. A collection R is said to be an *n-developed collection about* S provided (1) R is a simple one-dimensional collection, (2) R contains S as a sub-collection, (3) $R-S$ contains $3n\binom{k}{2}$ elements, (4) for each pair of elements s_1 and s_2 of S there exist exactly n simple chains of elements of $R-S$ each consisting of 3 links and each having one end link intersecting s_1 and the other intersecting s_2, and (5) no link of any such 3-link chain intersects more than two elements of R distinct from itself.

Let S_1 be a set of $n(G_1)$ disjoint 3-cells and let R_1 be an $n(G_1)$-developed collection about S_1. Let R_1 be the desired set F_1.

For $s, s' \in S_1$ let $B(s, s')$ be the set of chains of $R_1 - S_1$ which join s and s'. Let λ be a one-to-one transformation of S_1 onto G_1 and for $s, s' \in S_1$ let $\mu_{s,s'}$ be a one-to-one transformation of $B(s, s')$ onto G_1.

In defining $Tg(G_1, F_1^*, F_1)$ which we shall show to be strongly effective and hence $TgS(G_1, F_1^*, F_1)$ we impose consecutively the following conditions:

(A) For any $s \in S_1$ and $g \in G_1$, $gs=\lambda^{-1}g\lambda s$.

(B) For any $g \in G_1$, $s, s' \in S_1$, and f a link of an element b_f of $B(s, s')$, gf is that link of $\mu_{gs,gs'}^{-1}(g[g_{s,s'}(b_f)])$ which intersects gs, intersects gs' or intersects neither gs nor gs' according as f intersects, s, intersects s' or intersects neither s nor s'.

With these conditions being satisfied, G_1 acts in a strongly effective way on the finite set F_1 as we show. (A) implies that G_1 thus acts on S_1 by permuting the elements of S_1 among themselves, for $s \in S_1$, and $s' \in S_1$ there is a unique $g \in G_1$ for which $gs=s'$ and if $s=s'$, $g=e_1$. For

$f \in F_1 - S_1$, $e_1 f = f$ by (B). For $g, g' \in G_1$ and $f \in b_f \in B(s, s')$, $g(g'f)$ must be $(gg')f$ for

$$g[\mu^{-1}_{g's, g's'}(g'[\mu_{s,s'}(b_f)] = gb'_f = \mu^{-1}_{g(g',s), g(g's')} g[\mu_{g's, g's'}(b'_f)]$$

$$= \mu^{-1}_{(gg')s, (gg')s'}([gg'][\mu_{s,s'}(b_f)]) = b''_f$$

and consistent with this, $g(g'f)$ and $(gg')f$ are each determined solely by the orders on b_f and b'_f and on b''_f respectively relative to s and s' and $g's$ and $g's'$ on the one hand and $(gg')s$ and $(gg')s'$ on the other. It is easy to see that such operation is not only strongly effective but if $f, f' \in F_1$ with for some $g \in G_1$, $gf = f'$ then f and f' do not intersect the same element of F_1.

Furthermore, it follows directly from the construction that if $f, f' \in F_1$ intersect then for any $g \in G_i$, gf and gf' intersect.

With this information in mind we proceed to define $Tg(G_1, F_1^*, F_1)$. Let C_1 be the set of all 2-cells which are the intersections of elements of F_1. Then we may think of G_1 acting on C_1 consistent with G_1 acting on F_1, that is, for $c \in C_1$, c is $f \cdot f'$ for some $f, f' \in F_1$, and for $g \in G_1$, gc is $gf \cdot gf'$. But G_1 structures C_1 into orbits. From Lemma 2 by considering these orbits one at a time we may define $TgS(G_1, C_1^*, C_1)$ such that gc is gc as defined above and such that g is a homeomorphism of c onto gc which is oriented to be consistent with some orientation preserving homeomorphism of $f+f'$ onto $gf+gf'$ carrying f onto gf and f' onto gf'. That the orientation property of this latter statement is true follows from a consideration like that of the proof of Lemma 2. The orientation property may be made valid directly for the homeomorphisms from an element c to the elements in its orbit but any other homeomorphism between elements of such orbit is composed from these and for any $f, f', f'' \in F_1$ with f' intersecting f'' there is at most one $g \in G_1$ for which $gf = f'$ or f''.

But now Lemma 1 and Lemma 2 applied to the various orbits of the elements of F_1 under G_1 assert the existence of $TgS(G_1, F_1^*, F_1)$ as we set out to show. Clearly there exists an H_1 as in Lemmas 3′ and 3″ such that we may map F_1^* onto H_1^* as in the Lemma.

6. The inductive step of the proofs of Lemmas 3, 3′, and 3″. To complete the proofs of Lemmas 3, 3′, and 3″ it now suffices to define and establish the existence of F_i and $TgS(G_i, F_i^*, F_i)$, $i > 1$, given F_1 and $TgS(G, F_1^*, F_1)$ defined as above and F_j and $TgS(G_j, F_j^*, F_j)$, $1 < j < i$, defined by the inductive procedure to be given. We seek to do this so that applicable parts of Lemma 3 are satisfied. Then we shall note variations on the argument to yield Lemmas 3′ and 3″.

The construction we give will be similar in many ways to that of the preceding section. We shall require that $m(F_i) < 2^{-i}$.

Let C_{i-1} denote the collection of intersections of the various elements of F_{i-1} with each other. Each element of C_{i-1} is a 2-cell. Let $c \in C_{i-1}$ and let $f(c)$ and $f'(c)$ be the two elements of F_{i-1} for which $c = f(c) \cdot f'(c)$. Let $S_i(c, f(c))$ and $S_i(c, f'(c))$ be collections of exactly $\dfrac{n(G_i)}{n(G_{i-1})}$ disjoint 3-cells in $f(c)$ and $f'(c)$ respectively such that

(1) each element of $S_i(c, f(c))$ intersects exactly one element of $S_i(c, f'(c))$ and that in a 2-cell in c,

(2) each element of $S_i(c, f(c))$ or $S_i(c, f'(c))$ intersects $B(f(c))$[7] or $B(f'(c))$ respectively in a 2-cell and such 2-cell is in $S_i^*(c, f'(c))$ or $S_i^*(c, f(c))$ respectively, and

(3) there exist $R_i(c, f(c))$ and $R_i(c, f'(c))$ which are $n(G_i)$-developed collections about $S_i(c, f(c))$ and $S_i(c, f'(c))$ respectively such that (a) $[R_i(c, f(c)) - S_i(c, f(c))]^* \subset f(c) - B(f(c))$ and $[R_i(c, f'(c)) - S_i(c, f'(c))]^* \subset f'(c) - B(f'(c))$ and (b) $m[R_i(c, f(c))] < \varepsilon$ and $m[R_i(c, f'(c))] < \varepsilon$.

As it is possible to define such sets $S_i(c, f(c))$, $R_i(c, f(c))$, $S_i(c, f'(c))$ and $R_i(c, f'(c))$ for all $c \in C_{i-1}$ such that for $c' \neq c$, $R_i^*(c, f(c)) + R_i^*(c, f'(c))$ does not intersect $R_i^*(c', f(c')) + R_i^*(c', f'(c'))$, we consider such a collection of sets to exist, each $c \in C_{i-1}$ being identified with just two elements $R_i(c, f(c))$ and $R_i(c, f'(c))$.

For $f \in F_{i-1}$ let $R_i(f)$ and $S_i(f)$ be the union of all such sets $R_i(c, f)$ and $S_i(c, f)$ respectively for $c \in C_{i-1}$ and $c \subset f$. Thus $S_i(f)$, for example, is a particular collection of disjoint 3-cells in f.

DEFINITION. Let S denote a set of n disjoint 3-cells. A collection R is said to be an *(n, m)-weakly developed collection about* S provided (1) R is a simple one-dimensional collection, (2) R contains S as a subcollection, (3) $R - S$ contains $m \cdot \binom{n}{2}$ elements, and (4) for each pair of elements s_1 and s_2 there is a simple chain of m elements of $R - S$ having one end link intersecting s_1 and the other intersecting s_2 such that no link of any such chain intersects more than two elements of R distinct from itself.

Let $n(S_i(f))$ be the number of elements of $S_i(f)$. For some fixed integer m and any $f \in F_{i-1}$ let $Q_i(f)$ be an $(n(S_i(f)), m)$-weakly developed collection about $S_i(f)$ such that (1) each element of $Q_i(f) - S_i(f) \subset f - B(f)$, (2) no element of $Q_i(f) - S_i(f)$ intersects any element of $R_i(f) - S_i(f)$, and (3) $m(Q_i(f)) < 2^{-i}$.

Let $L_i(f)$ be that subset of $Q_i(f)$ consisting of $S_i(f)$ and all links of all chains of the development of $Q_i(f)$ between elements of $S_i(f)$ not both in any one set $S_i(c, f)$ for $c \in C_{i-1}$ and $c \subset f$.

Let $$S_i = \bigcup_{f \in F_{i-1}} S_i(f), \quad R_i = \bigcup_{f \in F_{i-1}} R_i(f) \text{ and } L_i = \bigcup_{f \in F_{i-1}} L_i(f).$$

[7] By $B(f)$ is meant the boundary of f.

The set F_i is defined as the set of all elements in one or more of S_i, R_i, and L_i.

Next we shall define G_i acting on F_i in a strongly effective manner such that

(a) for $f \in F_i$, and $g \in G_i$, f and gf do not intersect the same element of F_i,

(b) if $f, f' \in F_i$ for which $f \cdot f'$ exists then for each $g \in G_i$, $gf \cdot gf'$ exists, and

(c) for $f \in F_i$, $\tilde{f} \in F_{i-1}$ with $\tilde{f} \supset f$ and for any $g \in G_i$, $gf \subset \pi_{i-1}(g)\tilde{f}$.

Let D_{i-1} be the collection of all sets $G_{i-1}(c)$ for $c \in C_{i-1}$. Each element of D_{i-1} consists of $n(G_{i-1})$ 2-cells. For $d \in D_{i-1}$, let $f(d)$ and $f'(d)$ be the two sets each of which is an element of F_{i-1} containing an element of d plus the sum of its images under G_{i-1}. Let $S(f(d))$ and $S(f'(d))$ be the collection of those elements of S_i which (1) intersect d^* and (2) lie in $f(d)$ and $f'(d)$ respectively. Then $S(f(d))$ and $S(f'(d))$ each consist of $n(G_i)$ disjoint 3-cells.

For $d \in D_{i-1}$ let $\lambda_{f(d)}$ and $\lambda_{f'(d)}$ be one-to-one transformations of $S(f(d))$ and $S(f'(d))$ respectively onto G_i such that

(1) for $s \in S(f(d))$ and $s' \in S(f'(d))$, s intersects s' if and only if $\lambda_{f(d)}(s)$ is $\lambda_{f'(d)}(s')$ and

(2) for $g \in G_i$, $s \in S(f(d))$ and $f \in F_{i-1}$ for which $s \subset f$, $\pi_{i-1}(g)f \supset \lambda^{-1}_{f(d)} g \lambda_{f(d)}(s)$.

Each element of S_i belongs to exactly one set $S(f(d))$ or $S(f'(d))$ and thus S_i is structured by these sets. We may now define $TgS(G_i, S_i)$ as follows: for $g \in G_i$ and $s \in S(f(d))$, gs is $\lambda^{-1}_{f(d)} g \lambda_{f(d)}(s)$.

Next, for any $s, s' \in S_i$ for which $s, s' \in S(f)d))$ for some $d \in D_{i-1}$ and for which for some $f \in F_{i-1}$, $s+s' \subset f$, let $B(s, s')$ denote the set of 3-element chains from s to s' of the definition of R_i and let $\mu_{s,s'}$ be a one-to-one transformation of $B(s, s')$ onto G_i.

Then we may define $TgS(G_1, R_i)$. For $s \in R_i$ and $s \in S_i$ and for any $g \in G_i$, gs is gs as defined in $TgS(G_i, S_i)$. For any $g \in G_i$ and $s, s' \in S_i$ for which $B(s, s')$ is defined as above and for any x a link of an element b of $B(s, s')$, gx is that link of $\mu^{-1}_{gs,gs'}(g[\mu_{s,s'}(b)])$ which intersects gs, intersects gs' or intersects neither gs nor gs' according as f intersects s, intersects s' or intersects neither s nor s'.

Next we define $TgS(G_i, L_i)$. For $s \in L_i$ and $s \in S_i$ and for any $g \in G_i$, gs is gs as defined in $TgS(G_i, S_i)$. For $s, s' \in S_i$, $f \in F_{i-1}$, with $s+s' \supset f$ and s and s' not both elements of any set $S(c, f(c))$, there is a simple chain $\beta(s, s')$ of exactly m elements of $L_i(f)-S_i(f)$ with $\beta(s, s')$ having one end element intersecting s and the other s'. For each link x of $\beta(s, s')$ let, for $g \in G_i$, gx be that link of $\beta(gs, gs')$ which is the same number of links removed from gs as is x from s.

The definition of $Tg(G_i, F_i)$ is now complete and it may easily be verified that conditions (a)-(c) above are satisfied.

Let C_i be the set of all intersections of pairs of elements of F_i. Let $TgS(G_i, C_i)$ be defined as follows: for $c \in C_i$, c is a 2-cell which is the intersection of some two elements $f, f' \in F_i$; for $g \in G_i$, gc is $gf \cdot gf'$. Then as in § 5 employing Lemma 2, we may define $TgS(G_i, C_i^*, C_i)$ so that gc is gc as defined immediately above and g preserves orientation on c and gc relative to the orientations on (f, f') and (gf, gf') respectively.

Finally employing Lemmas 1 and 2 we may define $TgS(G_i, F_i^*, F_i)$ consistent with $TgS(G_i, F_i)$ and $TgS(G_i, C_i^*, C_i)$ so that with this inductive definition, Lemma 3 is satisfied. In this connection we note that under $TgS(G_i, F_i)$, for $f \in F_i$, $G_i(f)$ consists of $n(G_i)$ disjoint 3-cells so that Lemma 2 is applicable.

To modify the argument given so as to prove Lemma 3′′ we must introduce some extra conditions. The sets H_j, $1 \leqq j \leqq i-1$ exist as in the Lemma. Then when we define S_i we also define a set $S_i(H)$ where for $h, h' \in H_{i-1}$ with h intersecting h' exactly one 3-cell is introduced in $S_i(H)$ in each of h and h' intersecting the other. In defining R_i we also define a set $R_i(H)$ where $R_i(H) - S_i(H)$ consists of exactly $3 \cdot n(G_i) \cdot N$ elements with N the number of elements in $S_i(H)$ and with for each element s of $S_i(H)$ there being $n(G_i)$ 3-link simple chains in $R_i(H) - S_i(H)$, both end links of each such chain intersecting s. We may additionally require that $m(R_i(H)) < 2^{-i}$. Then for each pair of elements of $S_i(H)$ in the same element of H_{i-1} we introduce a simple chain of 3-cells joining them, the simple chain having m links with m being so chosen that $m(H_i) < 2^{-i}$. This imposes an extra condition on the "m" of the preceding argument. It is now straightforward to see that the sequences of Lemma 3′′ can be asserted to exist.

Finally to prove Lemma 3′ we need one extra device. For each $c \in C_{i-1}$, we choose not one but two pairs of sets $[S_i(c, f(c)), S_i(c, f(c))]$ and $[S_i'(c, f(c)), S_i'(c, f(c))]$ such that we may introduce two pairs of sets $[R_i(c, f(c)), R_i(c, f'(c))]$ and $[R_i'(c, f(c)), R_i'(c, f(c))]$ similar to the one pair we introduced before with additionally $R_i^*(c, f(c)) + R_i^*(c, f'(c))$ and $R_i'^*(c, f(c)) + R_i'^*(c, f'(c))$ not intersecting each other. Finally for any $f \in F_{i-1}$ we may define $S_i(f)$ in the similar fashion to that used before but with $S_i(f)$ here containing twice as many elements as the corresponding set in the preceding argument. Then we may form the set $Q_i(f)$ as an $(n(S_i(f)), m)$-weakly developed collection about $S_i(f)$ and proceed as before using extra conditions analogous to those of the argument sketched for Lemma 3′′.

It is clear that under such conditions $\{H_i\}$ and $\{\mu_i\}$ can be defined so that $\{H_i\}$ will be a λ-defining sequence.

Thus Lemma 3′ is proved and our argument for Theorems 1 and 2 is completed.

REFERENCES

1. R. D. Anderson, *A characterization of the universal curve and a proof of its homogeneity*. Submitted to Ann. of Math.

2. L. Keldys, *Example of a one-dimensional continuum with a zero-dimensional and interior mapping onto the square*. Dokl. Akad. Nauk SSSR (N.S.) **97** (1954), 201–204 (Russian).

3. A. Kolmogoroff, *Über offene Abbildungen*, Ann. of Math., (2) **38** (1937), 36–38.

4. D. Montgomery and L. Zippin, *Topological transformation groups*, Interscience Publishers Inc., New York.

5. P. A. Smith, *Periodic and nearly periodic transformations*, Lectures in Topology, The University of Michigan Conference of 1940, pp. 159–190.

6. Leo Zippin, *Transformation groups*, Lectures in Topology, The University of Michigan Conference of 1940, pp. 191–221.

THE INSTITUTE FOR ADVANCED STUDY AND
UNIVERSITY OF PENNSYLVANIA

HOLOMORPHIC FUNCTIONALS AND COMPLEX CONVEXITY IN BANACH SPACES

H. J. BREMERMANN

1. Introduction. The present paper extends some basic theorems of the theory of several complex variables to Banach spaces. Results which are new even for finite dimension are also obtained. Considerable use is made of methods developed in "Complex Convexity" (Bremermann [8]), however, many modifications are necessary to adapt them to infinite dimension.

A complex valued functional is *Gâteaux holomorphic* (or in short *G-holomorphic*) in a domain D of a complex Banach space B_c if it is single valued and its restriction to an arbitrary analytic plane $\{z|z=z_0+\lambda a\}$ ($z_0 \in D$, $a \in B_c$, λ a complex parameter) is a holomorphic function of λ in the intersection of the plane with D. The space of n complex variables C^n can be considered as a Banach space, and for C^n the above definition is equivalent to the usual definition of a holomorphic function of several complex variables. In an infinite dimensional Banach space the Gâteaux holomorphic functions are not necessarily locally bounded, while in a finite dimensional space the local boundedness is a consequence of holomorphy. Therefore another notion of holomorphy, also coinciding with the notion of holomorphy in finite dimensional spaces, is possible: A function is *Fréchet holomorphic* in a domain D if it is Gâteaux holomorphic and locally bounded (compare Hille [11] and Soeder [17]). The theories of both types of holomorphic functions have been studied, the latter more than the former. Both theories are considerably less developed than the theory of finitely many variables. This may be partly due to the fact that the infinite dimensional spaces are not locally compact, in fact, if a space is locally compact, then it is finite dimensional (see Hille [11]).

In the present paper the theory of Gâteaux holomorphic functionals is studied exclusively. As a tool are used plurisubharmonic functionals (as defined by Oka [14] and [15], Lelong [12] and Thorin [19]) and a functional $d_D^{(N)}(z)$ which is the distance of the point z from the boundary of the domain D measured in the norm N. A notion of holomorphic continuation is defined and a "basic lemma" on the simultaneous continuation of G-holomorphic functionals is proved (3.1).[1] A consequence of

Received April 26, 1956. This research was supported by the United States Air Force, through the Office of Scientific Recearch of the Air Research and Development Command.

[1] The lemma in its present form is new also for finite dimension and permits to construct the envelope of an arbitrary domain in the C^n explicitely. This will be carried out in a further paper.

this lemma is the fact that there exist domains, as in finite dimension, such that all G-holomorphic functions can be continued G-holomorphically into a larger domain. Those domains for which a G-holomorphic function exists that is not continuable, are called *domains of holomorphy*. From the continuation lemma follows that the domains of holomorphy have the property that the functional $-\log d_D^{(N)}(z)$ is plurisubharmonic in D, and a theorem is proved which for finite dimension is known as "Kontinuitätssatz." The property of the functional $-\log d_D^{(N)}(z)$ to be plurisubharmonic is invariant with respect to all norms N that generate equivalent topologies. The domains for which $-\log d_D^{(N)}(z)$ is plurisubharmonic are called *pseudo-convex*, and some of their properties are studied. A domain D in a complex Banach space B_c is pseudo-convex if and only if its intersection with every (complex) two-dimensional linear submanifold of B is pseudo-convex.

The notion of pseudo-convexity bears some formal relationship to the ordinary convexity in real spaces, this is established by showing: A domain D in a real Banach B_r is convex if and only if $-\log d_D^{(N)}(x)$ is a convex functional in D. Finally tube domains are studied, that is domains of the form $\{z|x \in d_x,\ y \text{ arbitrary}\}$, where x is the real part and y the imaginary part of z, and d_x a domain in the real Banach space of the real parts. It is shown that for this particular class of domains the two notions coincide: A tube domain is pseudo-convex if and only if it is convex.

For simplicity's sake the present considerations are limited to complex valued functionals but can be extended without difficulty to vector valued functions. Also generalizations to spaces more general than Banach spaces (for instance locally convex spaces) are possible.

2. Holomorphic, plurisubharmonic and distance functionals.

2.1. We will consider in this paper Banach spaces where the field of scalars is either the field of real numbers or the field of complex numbers. Accordingly we speak of *real* and *complex Banach spaces* and write B_r and B_c respectively. By z we will denote exclusively elements of complex B-spaces and by x elements of real B-spaces.

2.2. The norm that is defined in a Banach space B provides it in a natural way with a topology (strong topology). As neighborhoods of a point $a \in B$ we define the pointsets $\{b|\,\|b-a\| < \varepsilon\}$. A *region* is an *open* set; a *domain* is an *open and connected set*.

2.3. DEFINITION. Let λ be a complex parameter. A complex valued functional $f(z)$, defined in a domain D of a complex Banach space B_c is

Gâteaux holomorphic in D (or in short G-holomorphic) if $f(z)$ is single-valued in D and if $f(z_0+\lambda a)$ is holomorphic in λ at the point $\lambda=0$ for all $z_0 \in D$ and $a \in B$. In other words $f(z)$ is required to be a holomorphic function of the one complex variable λ on the intersection of any two-dimensional analytic plane $\{z|z=z_0+\lambda a\}$ with the domain D. Obviously a function is G-holomorphic in D if and only if it is single-valued and G-holomorphic in a neighborhood of each point of D (locally holomorphic). This definition is equivalent to the requirement that the Gâteaux differential exists everywhere in D and that $f(z)$ is single-valued in D. We do *not* require that $f(z)$ be locally bounded or similar conditions. (Compare Hille [11], p. 71 and p. 81).

2.4. DEFINITION. A real-valued functional $V(z)$ defined in a domain D of a complex Banach space B_c is *quasi-plurisubharmonic* in D if $V(z_0+\lambda a)$ is quasi-subharmonic in λ at the point $\lambda=0$ for all $z_0 \in D$ and $a \in B_c$. $V(z)$ is *plurisubharmonic* if $V(z)$ is quasi-plurisubharmonic and upper-semicontinuous in D. (Cf. Thorin [19], p. 16) $V(z)$ is upper-semicontinuous at the point z_0 if for every $\varepsilon > 0$ there exists a δ, such that $V(z)-V(z_0) < \varepsilon$ for $\|z-z_0\| < \delta$. (For the definition of quasi-plurisubharmonic see T. Radó [16]. Cf. also P. Lelong [12]). What we call quasi-plurisubharmonic functions Lelong denotes as *functions of class M*. The plurisubharmonic functions have also been introduced by K. Oka [14] and [15] under the name *pseudo-convex functions*. Oka admits the constant, $-\infty$, Lelong excludes it. For our applications it is more convenient to admit $-\infty$ as a plurisubharmonic functional.)

2.5. We now have to define the notion of *holomorphic continuation*. In one and several variables this is being done by means of power series developments. However, power series are somewhat inconvenient here. Therefore we will define as holomorphic continuation a function that is holomorphic in a larger domain and coincides with the given function in the given domain. However, already in one variable the "larger domains" may be no longer schlicht but concrete complex manifolds with no branch points as interior points. We have to take care of this situation and therefore define the following.

2.6. D is a *domain over the space* B_c if D is a topological space carrying a mapping φ_D which maps D into B_c, such that φ_D is locally a homeomorphism.

We call φ_D the *projection mapping* of D and $\varphi_D(E)$, where E is a set in D, the *projection* of E.

Domains over a space B_c are special *complex analytic manifolds of infinite dimension*. (For general complex analytic manifolds of infinite dimension, cf. J. Eells [10].)

2.7. D is a *continuation* of a domain D_0 over B_c if there exists a subset $\tilde{D}_0$ of D and a homeomorphism h of $\tilde{D}_0$ onto D_0 such that $\varphi_{D_0}h(P)=\varphi_D(P)$ for every P in $\tilde{D}_0$.

We can then identify $\tilde{D}_0$ and D_0. In particular if D_0 is a domain *in* B_c—we will also say *schlicht domain*—then D is a continuation of D_0 if there exists a subset $\tilde{D}_0 \subset D$ such that $\varphi_D(\tilde{D}_0)$ is a homeomorphism onto D_0.

2.8. A functional f is *G-holomorphic in a domain D over a complex Banach space* B_c if f is G-holomorphic in a neighborhood of each point in D. And it is G-holomorphic in a neighborhood U_i if it is G-holomorphic in the homeomorphic image $\varphi_D U_i$ which is an open set in B_c where the notion of G-holomorphy is defined (2.3).

2.9. Let $f(z)$ be a G-holomorphic functional in a domain $D \subset B$. Then $g(z)$ is a *G-holomorphic continuation* of $f(z)$ if $g(z)$ is G-holomorphic in a continuation D_1 of D and coincides with $f(z)$ in D.

2.10. *Uniqueness of the G-holomorphic continuation.* Let D be a domain over B_c. Let D^* be a subdomain. Let $f(z)$ and $g(z)$ be G-holomorphic functionals in D, let $f(z) \equiv g(z)$ in D^*, then we have $f(z) \equiv g(z)$ throughout D.

Proof. Let S be the set of points such that $f(z)$ and $g(z)$ coincide. Then we have $D^* \subset S \subset D$. Let S^* be the largest open set contained in S. Suppose $S^* \neq D$. Then there exists a boundary point z_0 of S^* which is an interior point of D. Let U be the homeomorphic image of a neighborhood of z_0 in the B_c. In particular we can choose U as a sphere. In this sphere we have a point z_1 such that in a neighborhood of z_1 we have $f(z)=g(z)$ and a point z_2 such that $f(z_2) \neq g(z_2)$. Then we connect z_1 and z_2 by an analytic plane which cuts U in a circle. Restricting $f(z)$ and $g(z)$ to the analytic plane we obtain a contradiction to the identity theorem of holomorphic functions in one variable.

2.11. A domain H for which a functional $f(z)$ exists that is G-holomorphic in H and does not possess a G-holomorphic continuation into a proper continuation of H we call a *domain of holomorphy*.

2.12. *The distance function.* Let D be a domain *in* B, then we associate with every point of D the value

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

in other words $d_D^{(N)}(z)$ is the distance of the point z from the boundary of D measured in the norm N.

If D is different from the whole space B, then D has at least one finite boundary point, and then obviously $d_D^{(N)}(z)$ is finite in D. If D is the whole space B_c, then $d_D^{(N)} \equiv \infty$.

2.13. *If D is different from the whole space, then $d_D^{(N)}(z)$ is continuous with respect to the topology generated by the norm N.*

The poof is the same as in the finite case which is carried out in Bremermann [8].

2.14. Definition. Besides the distance function $d_D^{(N)}(z)$ we will consider the distance function

$$d_{a,D}^{(N)}(z)=\sup r \ni \{z' \mid z'=z+\lambda a,\ \|a\|_N=1,\ |\lambda|<r\} \subset D ,$$

in other words $d_{a,D}^{(N)}(z)$ is the radius of the largest circle with center at z on the analytic plane $\{z' \mid z'=z+\lambda a\}$ that is contained in D, that is the distance of z from the boundary of $\{z' \mid z'=z+\lambda a\} \cap D$.

From the definitions it follows immediately the relation

$$d_D^{(N)}(z)=\inf_a \{d_{a,D}^{(N)}(z)\} .$$

where a varies through all elements of B with norm 1.

2.15. *The function $d_{a,D}^{(N)}(z)$ is lower semicontinuous with respect to the topology generated by the norm N.*

If D is the whole space, then $d_{a,D}^{(N)}(z)$ will be $\equiv\infty$. However, even if D is not the whole space, $d_{a,D}^{(N)}(z)$ can be infinite for certain directions, though not for all directions.

(1) Let $d_{a,D}^{(N)}(z_1)=c_1$ be finite. Then for every $\varepsilon>0$ the point set

$$\{z \mid z=z_1+\lambda a,\ |\lambda| \leqq c_1-\varepsilon\}$$

is compact in D. Hence there exists for every $\varepsilon>0$ a δ such that for $\|z_1-z_2\|<\delta$ the point set

$$\{z \mid z=z_2+\lambda a,\ |\lambda| \leqq c_1-\varepsilon\}$$

is contained in D. Hence for $\|z_1-z_2\|<\delta$ we have

$$d_{a,D}^{(N)}(z_2)>c_1-\varepsilon , \quad \text{or}$$

$$d_{a,D}^{(N)}(z_1)-d_{a,D}^{(N)}(z_2)<\varepsilon .$$

Hence $d_{a,D}^{(N)}(z)$ is lower semicontinuous at z_1 with respect to the norm N.

(2) Let $d_{a,D}^{(N)}(z_1)=\infty$. Then for arbitrary large M the point set

$$\{z|z=z_1+\lambda a,\ |\lambda|\leqq M\}$$

is compact in D. Hence there exists a δ such that for $||z_1-z_2||<\delta$ the point set

$$\{z|z=z_2+\lambda a,\ |\lambda|\leqq M\}$$

is contained in D. Hence

$$d_{a,D}^{(N)}(z_2)>M \quad \text{for} \quad ||z_1-z_2||<\delta .$$

That means that also in this case $d_{a,D}^{(N)}(z)$ is lower semicontinuous at the point z_1 with respect to the norm N.

2.16. By a similar argument it follows that $d_{a,D}^{(N)}(z)$ is for fixed z lower semicontinuous with respect to variable direction a.

3. Simultaneous holomorphic continuation.

3.1. FUNDAMENTAL LEMMA. *Let D be a domain in a complex Banach space. Let S be a simply connected domain on an analytic plane $\{z|z=z_0+\lambda b\}$. Let T be the boundary of S and let $S\cup T\subset D$. Let $X(\lambda)$ be a function holomorphic in the image of S in the λ-parameter plane—in the following we will simply say holomorphic in S—and let $X(\lambda)\neq 0$ in $S\cup T$ and $|X(\lambda)|$ continuous in $S\cup T$.*

Let

$$|X(\lambda)|\, d_{a,D}^{(N)}(z_0+\lambda b)\geqq m>0 \qquad \text{for} \quad \lambda\in T .$$

Then any functional that is G-holomorphic in D can be continued G-holomorphically into all points

$$C=\{z|z=z_0+\lambda b+\tau a,\ \lambda\in S\cup T,\ |\tau|<m|X(\lambda)|^{-1}\} ,$$

(τ a complex parameter).

The idea of the proof is the following. We consider the subspace $\{z|z=z_0+\lambda b+\tau a\}$ and an arbitrary functional $f(z)$. The restriction of f to the intersection of D with this subspace is a holomorphic function in λ and τ. For fixed λ we can develop $f(z_0+\lambda b+\tau a)$ into an ordinary power series of powers of τ. From the maximum principle we derive that this series converges in the pointset C. Thus we have continued f into C. However it has to be checked that the continuation is not only a continuation of f as a holomorphic function of the one variable τ but as a G-holomorphic function in B_c.

This is not trivial. Functions of two complex variables are known which can be continued as functions of one variable beyond the domain where they are holomorphic in both variables. (Cf. Behnke-Thullen [3].)

In order to show that $f(z)$ is holomorphic in an arbitrary point P of C it is sufficient to show that $f(z)$ can be defined in a neighborhood of P such that the restriction of $f(z)$ to an arbitrary analytic plane $\{z|z=P+\sigma c\}$ through P is holomorphic. We do this by including in the proof an arbitrary direction c from the beginning.

Proof. Let $f(z)$ be an arbitrary functional G-holomorphic in D. We consider the subspace $\{z|z=z_0+\lambda b+\tau a+\sigma c\}$ where σ is a complex parameter and c an arbitrary direction with $||c||=1$. The restriction of $f(z)$ to the subspace is a holomorphic function of the three complex variables λ, τ, σ (no matter if a, b, c are linearly independent or not.).

For $\lambda \in T$ we have by assumption that $|X(\lambda)|d_{a,D}^{(N)}(z_0+\lambda b) \geqq m$, and because $X(\lambda) \neq 0$ on T we have

$$d_{a,D}^{(N)}(z_0+\lambda b) \geqq |X(\lambda)|^{-1} \qquad \text{for} \quad \lambda \in T\,.$$

Obviously there exists for every $\varepsilon > 0$ a sufficiently small $\delta > 0$ such that the set

$$C^* = \{z|z=z_0+\lambda b+\tau a+\sigma c,\ |\tau| \leqq (m-\varepsilon)|X(\lambda)|^{-1},\ \lambda \in T,\ |\sigma| \leqq \delta\}$$

is contained in D for arbitrary c with $||c||=1$. The set C^* is compact in the subspace $\{z|z=z_0+\lambda b+\tau a+\sigma c\}$, therefore the restriction of f to it is bounded (according to a well known theorem of n complex variables which was first proved by F. Hartogs (compare Carathéodory [9])). Let the bound be M. (M depends upon c, of course.)

We now develop the restriction of f in a power series in τ and σ.

$$f(z_0+\lambda b+\tau a+\sigma c) = \sum_{\mu,\nu=0}^{\infty} \frac{1}{\mu!\nu!} \frac{\partial^{\mu+\nu} f(z_0+\lambda b+\tau a+\sigma c)}{\partial\tau^\mu \partial\sigma^\nu}\Bigg/_{\tau=\sigma=0} \tau^\mu \sigma^\nu\,.$$

For $\lambda \in T$, $|\sigma| \leqq \delta$ and $|\tau| \leqq (m-\varepsilon)|X(\lambda)|^{-1}$ the point $z_0+\lambda b+\tau a+\sigma c$ belongs to C^* where f is holomorphic and its modulus smaller than M. Hence we obtain by Cauchy's formula for $\lambda \in T$ the inequality

$$\left| \frac{1}{\mu!\nu!} \frac{\partial^{\mu+\nu} f(z_0+\lambda b+\tau a+\sigma c)}{\partial\tau^\mu \partial\sigma^\nu}\Bigg/_{\tau=\sigma=0} \tau^\mu \sigma^\nu \right| \leqq \frac{M}{\delta^\nu[(m-\varepsilon)|X(\lambda)|^{-1}]^\mu}\,.$$

By multiplying with $|X(\lambda)|^{-\mu}$ we obtain for $\lambda \in T$:

$$\left| \frac{1}{\mu!\nu!} \frac{\partial^{\mu+\nu} f(z_0+\lambda b+\tau a+\sigma c)}{\partial\tau^\mu \partial\sigma^\nu}\Bigg/_{\tau=\sigma=0} \tau^\mu \sigma^\nu \right| |X(\lambda)|^{-\mu} \leqq \frac{M}{\delta^\nu (m-\varepsilon)^\mu}\,.$$

The left hand side is for $\sigma=\tau=0$ the modulus of a holomorphic function of λ and takes its maximum with respect to $S \cup T$ on T. *Therefore the* *inequality is valid not only for* $\lambda \in T$ *but for* $\lambda \in S \cup T$. Hence the series converges uniformly in every compact subset of the set

$$C^{**}=\{z|z=z_0+\lambda b+\tau a+\sigma c,\ \lambda \in S \cup T,\ |\tau|<(m-\varepsilon)|X(\lambda)|^{-1},\ |\sigma|<\delta\}\ .$$

The limit function of the series is a continuation of $f(z_0+\lambda b+\tau a+\sigma c)$ into the set C^{**}.

Letting c vary though all directions, that is, through all elements of B_c such that $\|c\|=1$ and letting ε tend to zero we define $f(z)$ in a full neighborhood (in the B_c) of each point of the set

$$C=\{z|z=z_0+\lambda b+\tau c,\ \lambda \in S \cup T,\ |\tau|<m|X(\lambda)|^{-1}\}\ .$$

We have to make sure that this definition of $f(z)$ is consistent, that at the same point not two different values are defined!

If a, b, c_1 and c_2 are linearly independent, then

$$z_1=z_0+\lambda_1 b+\tau_1 a+\sigma_1 c_1 \neq z_2=z_0+\lambda_2 b+\tau_2 a+\sigma_2 c_2$$

for all triples

$$(\lambda_1,\ \tau_1,\ \sigma_1) \neq (\lambda_2,\ \tau_2,\ \sigma_2)\ .$$

Therefore no contradictory values can be defined.

Suppose now that a, b, c_1 and c_2 are linearly dependent: $c_2=\alpha_1 a+\alpha_2 b+\alpha_3 c_1$. Then $\lambda_1=\lambda_2+\sigma_2\alpha_2$, $\tau_1=\tau_2+\sigma_2\alpha_1$, and $\sigma_1=\alpha_3\sigma_2$ imply $z_1=z_2$. Let

$$f(z_0+\lambda b+\tau a+\sigma c_1)=f_1(\lambda,\ \tau,\ \sigma) \quad \text{and} \quad f(z_0+\lambda b+\tau a+\sigma c_2)=f_2(\lambda,\ \tau,\ \sigma)\ .$$

Then in a neighborhood of (0, 0, 0) we have

$$f_1(\lambda+\sigma\alpha_2,\ \tau+\sigma\alpha_1,\ \alpha_3\sigma) \equiv f_2(\lambda,\ \tau,\ \sigma)\ .$$

This functional equation persists wherever both functions are holomorphic (in λ, τ, σ). Hence no contradictory values of $f(z)$ are defined at the same point.

Finally we observe that f is by construction G-holomorphic in a neighborhood (of the B_c) of each point of C. We observe further that C is simply connected. Hence f is single-valued in C. Hence the lemma is proved.

3.2. *Let the conditions of* 3.1 *be satisfied except that we replace* $d_{a,D}^{(N)}(z)$ *by* $d_D^{(N)}(z)$. *Let*

$$|X(\lambda)|d_D^{(N)}(z) \geqq m > 0 \qquad \text{for} \quad \lambda \in T\ .$$

Then any functional that is G-holomorphic in D can be continued G-holomorphically into all points

$$\tilde{C}=\{z|\,\|z_0+\lambda b-z\|<m|X(\lambda)|^{-1},\ \lambda\in S\cup T\}\ .$$

This follows immediately from the Fundamental Lemma 3.1. For if $|X(\lambda)|d_D^{(N)}(z)\geqq m$ for $\lambda\in T$, then in particular $|X(\lambda)|d_{a,D}^{(N)}(z)\geqq m$ for $\lambda\in T$ for every a. Hence by Lemma 3.1 $f(z)$ is G-holomorphically continued into all sets

$$\{z|z=z_0+\lambda b+\tau a,\ \lambda\in S\cup T,\ |\tau|<|X(\lambda)|^{-1}\}\ ,$$

and the union of all these sets is $\tilde{C}$. We observe further that $\tilde{C}$ is simply connected and therefore the continuation single valued.

3.3. How is the continuation of a functional f, G-holomorphic in D, into a set $\tilde{C}$ as described in 3.2 compatible with values already defined in D?

If the intersection of D and $\tilde{C}$ is connected, then f is single-valued. Now let $D\cap\tilde{C}$ not be connected. Then there is one component $\tilde{C}_0$ of $D\cap\tilde{C}$ containing $S\cup T$. f is a continuation from $\tilde{C}_0$ which can furnish function elements different from the ones already defined in the other components.

We therefore proceed as one does in one and finitely many variables. We pass to "domains *over* the space." We consider $\tilde{C}$ as the projection of a set $\tilde{C}^*$ under a mapping φ, being a homeomorphism of $\tilde{C}^*$ onto $\tilde{C}$. We identify $\tilde{C}_0$ and $\tilde{C}_0^*=\varphi^{-1}\tilde{C}_0$, while we consider the other components of $D\cap\tilde{C}$ and their images in C^* as different points.

In a further paper we will study the iteration of this process and we will show that in the limit we obtain the simultaneous continuation of all functionals that are holomorphic in D into "the pseudo-convex envelope" of D.

3.4. If D is a domain of holomorphy, then the sets C and $\tilde{C}$ belong to D. Therefore in this case no question of single-valuedness arises and we can admit S to be an arbitrary domain, not necessarily simply connected.

3.5. COROLLARY. *Let D be a (schlicht) domain of holomorphy. Let S be an arbitrary domain on an analytic surface $\{z|z=z_0+\lambda b\}$, T the boundary. Then*

$$\inf_{T}|X(\lambda)|d_{a,D}^{(N)}(z)=\inf_{S\cup T}|X(\lambda)|d_{a,D}^{(N)}(z)$$

and

$$\inf_{T} |X(\lambda)| d_{D}^{(N)}(z) = \inf_{S \cup T} |X(\lambda)| d_{D}^{(N)}(z) .$$

Let

$$m = \inf_{T} |X(\lambda)| d_{a, D}^{(N)}(z_0 + \lambda b) .$$

Now if D is a domain of holomorphy, then the set C defined in 3.1 belongs to D. Hence

$$d_{a, D}^{(N)}(z_0 + \lambda b) \geqq m |X(\lambda)|^{-1}$$

for $\lambda \in S \cup T$. Hence

$$|X(\lambda)| d_{a, D}^{(N)}(z_0 + \lambda b) \geqq m$$

for $\lambda \in S \cup T$ from which the first equality follows immediately. The second equality follows analogously from Lemma 3.2.

3.6. THEOREM. *Let D be a holomorphy, then the functionals $-\log d_{a, D}^{(N)}(z)$ and $-\log d_{D}^{(N)}(z)$ are plurisubharmonic in D.*

Proof. Suppose $-\log d_{D}^{(N)}(z)$ would not be plurisubharmonic in D. Now $-\log d_{D}^{(N)}(z)$ is continuous. Thus there would exist an analytic plane $\{z | z = z_0 + \lambda b\}$ on which it would not be subharmonic. That means there would exist a (small) circle and a harmonic function h being a majorant on the boundary of the circle but not inside.

We choose the representation of the analytic plane such that z_0 is the center of the circle. $h(\lambda)$ is harmonic in the open circle $|\lambda| < \rho$ and continuous in $|\lambda| \leqq \rho$ and for $|\lambda| = \rho$ we have

$$-\log d_{D}^{(N)}(z_0 + \lambda b) \leqq h(\lambda) .$$

On the other hand there exists a λ_0 with $|\lambda_0| < \rho$ such that

$$-\log d_{D}^{(N)}(z_0 + \lambda_0 b) > h(\lambda_0) .$$

Let $h^*(\lambda)$ be a conjugate harmonic function of $h(\lambda)$, then

$$|e^{h(\lambda) + ih^*(\lambda)}| d_{D}^{(N)}(z_0 + \lambda b) \geqq 1 \qquad \text{for} \quad |\lambda| = \rho$$

and

$$|e^{h(\lambda_0) + ih^*(\lambda_0)}| d_{D}^{(N)}(z_0 + \lambda_0 b) < 1 \qquad \text{for} \quad \lambda_0 .$$

This is a contradiction to 3.5 with $T = \{|\lambda| = \rho\}$, $S = \{|\lambda| < \rho\}$ and $X(\lambda) = e^{h(\lambda) + ih^*(\lambda)}$.

Hence $-\log d_{D}^{(N)}(z)$ is plurisubharmonic. The proof for $-\log d_{a, D}^{(N)}(z)$ is analogous. The only difference is that $-\log d_{a, D}^{(N)}(z)$ is upper-semi-continuous instead of continuous.

4. The Kontinuitätssatz. We will derive now a theorem which in the theory of finitely many variables is known as "Kontinuitätssatz." The term "Kontinuitätssatz" was introduced by Behnke-Thullen [3]; we will use this term because translating it as "theorem of continuity" might be misleading.

4.1. *Let D be a domain of holomorphy. Let $\{S_\nu\}$ be a family of bounded domains on one dimensional analytic planes and $\{T_\nu\}$ their boundaries. Let $S_0 = \lim S_\nu$ and $T_0 = \lim T_\nu$. Then $S_\nu, T_\nu \subset D$ for every ν and $T_0 \subset D$ imply $S_0 \subset D$.*

Proof. Applying 3.5 with $X(\lambda) \equiv 1$ we obtain

$$\inf_{S_\nu \cup T_\nu} d_D^{(N)}(z) = \inf_{T_\nu} d_D^{(N)}(z) .$$

Now $d_D^{(N)}(z)$ is a continuous functional in D. Then also $\inf_{S_\nu \cup T_\nu} d_D^{(N)}(z)$ and $\inf_{T_\nu} d_D^{(N)}(z)$ are continuous. Therefore the above equality holds also in the limit.

Hence

$$\inf_{S_0 \cup T_0} d_D^{(N)}(z) = \inf_{T_0} d_D^{(N)}(z) .$$

Now, because T_0 is compact and in D, we have

$$\inf_{T_0} d_D^{(N)}(z) > 0$$

and therefore $\inf_{S_0 \cup T_0} d_D^{(N)}(z) > 0$, which means

$$S_0 \subset D .$$

4.2. The Kontinuitätssatz can be expressed also in the following way:

Let $\{S_\nu\}$ be a family of bounded domains on one dimensional analytic planes. Let $S_0 = \lim S_\nu$ and $T_0 = \lim T_\nu$. *Let $f(z)$ be a functional holomorphic on T_0. If then $f(z)$ is singular at least at one point of S_0, then there exists a ν_0 such that for $\nu > \nu_0$ the domain S_ν contains at least one singularty of $f(z)$.*

5. Pseudo-convex domains. For finite dimensional domains the property of the functional $-\log d_D^{(N)}(z)$ to be plurisubharmonic is invariant with respect to the norm (Bremermann [8]). In this section we will extend this result to the infinite dimensional case. As in finite dimension we denote the domains for which the functionals $-\log d_D^{(N)}(z)$ are plurisubharmonic as *pseudo-convex*.

Thus *the domains of holomorphy are pseudo-convex*. For finite dimension the converse is true: The pseudo-convex domains are domains of holomorphy. This is very deep result due to K. Oka. (K. Oka [14] and [15]). Compare also F. Norgent [13] and H. J. Bremermann [7].) Most of the techniques applied to obtain this result cannot be generalized to infinite dimension (for instance the Weil-Bergmann integral formula [Weil [20], Bergmann [4], [5]] etc.). Nevertheless the pseudo-convexity may be characteristic for domains of holomorphy in the infinite dimensional case also.

5.1. *Let D be a domain such that for a certain norm N the functional $-\log d_{a,D}^{(N)}(z)$ is plurisubharmonic in D for every a.*

Then the intersection D^* with any finite dimensional linear submanifold L of B is a pseudo-convex region.

Proof. Let the linear submanifold L be $L=\{z|z=z_0+\tau_1 b_1+\cdots+\tau_n b_n\}$. Let $B^*=\{z|z=\tau_1 b_1+\cdots+\tau_n b_n\}$. The restriction of the norm N to the subspace B^* is a norm N^* in B^*. For every $a\in B^*$ the restriction of $d_{a,D}^{(N)}(z)$ to $D^*=L\cap D$ is equal to $d_{a,D^*}^{(N)}(z)$ by definition.

The restriction of any plurisubharmonic functional in D to D^* is a plurisubharmonic function in D^*, as one sees immediately from the definition of the plurisubharmonic functions.

Hence $-\log d_{a,D^*}^{(N^*)}(z)$ is a plurisubharmonic function for every $a\in B^*$. Hence (the finite dimensional) D^* is pseudo-convex according to a result by Bremermann [8].

REMARK. The intersection $D\cap L$ is not necessarily connected, but just an open set, that is, a region.

5.2. *Let D be a domain such that the intersection D^* of D with any two dimensional linear submanifold L of B is pseudo-convex, then $-\log d_{a,D}^{(N)}(z)$ is plurisubharmonic in D for every a and for any norm N, which generates a topology which is equivalent to the topology with respect to which D is defined.*

Proof. For finite dimensional Banach spaces it has been proved in Bremermann [8] that if D^* is a pseudo-convex region then $-\log d_{a,D^*}^{(N^*)}$ is plurisubharmonic in D^* for any norm N^*. Now if N is an arbitrary norm, then its restriction N^* to a finite dimensional subspace B^* is a norm in B^*. Hence $-\log d_{a,D^*}^{(N^*)}(z)$ is plurisubharmonic in D^* for $a\in B^*$ for every two dimensional subspace B^*, hence the restriction of $-\log d_{D}^{(N)}(z)$ to any D^* is plurisubharmonic, hence $-\log d_{a,D}^{(N)}(z)$ is plurisubharmonic in D.

5.1 and 5.2 combined yield:

5.3. COROLLARY. *The property of the functionals* $-\log d_{a,D}^{(N)}(z)$ *to be plurisubharmonic for a domain* D *is invariant with respect to all norms that generate equivalent topologies.*

5.4. DEFINITION. The domains which have the property that the functionals $-\log d_{a,D}^{(N)}(z)$ are plurisubharmonic for all a we call *pseudo-convex.*

5.5. COROLLARY. *The domains of holomorphy are pseudo-convex.*

5.6. With this definition we can express 5.1 and 5.2 also in the following way.

A domain D *is pseudo-convex if and only if the intersection* D *with any finite dimensional linear submanifold is a pseudo-convex region.*

The same is true if we replace "finite dimensional" by "two dimensional."

5.7. We now replace the functionals $-\log d_{a,D}^{(N)}(z)$ by $-\log d_D^{(N)}(z)$ and show: D *is pseudo-convex if and if* $-\log d_D^{(N)}(z)$ *is plurisubharmonic for arbitrary norms* N *with equivalent topology.*

Let D^* be as in 5.1. If $-\log d_D^{(N)}(z)$ is plurisubharmonic, then its restriction to D^* is plurisubharmonic. In general, however, we cannot say that this restriction is equal to $-\log d_{D^*}^{(N^*)}(z)$.

Let the finite dimensional submanifold be

$$\{z|z=z^0+\tau_1 b_1+\cdots+\tau_n b_n\} .$$

Then we take the upper envelope of

$$\{-\log d_D^{(N)}(z), \log|\tau_1|, \cdots \log|\tau_n|\} .$$

This is a plurisubharmonic function that becomes infinite everwhere at the boundary of D^*. Hence the finite dimensional region D^* is pseudo-convex according to Bremermann [8]. Hence D is pseudo-convex according to 5.6.

On the other hand if D is pseudo-convex, then

$$-\log d_{a,D}^{(N)}(z)$$

is plurisubharmonic for every a by definition. Now we have

$$-\log d_D^{(N)}(z) = \sup_a \{-\log d_{a,D}^{(N)}(z)\} \ .$$

In particular this relation holds also on every analytic plane. On analytic planes $-\log d_D^{(N)}(z)$ is thus the upper envelope of subharmonic functions, and thus according to Radó [16], subharmonic. Hence $-\log d_D^{(N)}(z)$ is plurisubharmonic in D.

5.8. *A domain D is pseudo-convex if there exists a plurisubharmonic functional $V(z)$ such that the closure of*

$$\{z \mid V(z) < M, \ z \in D\}$$

is contained in D for arbitrary large M. If D is bounded, then the converse is true.

Proof. If there exists such a $V(z)$, then we restrict $V(z)$ to finite dimensional subspaces and obtain by the analog theorem from the finite dimensional case (Bremermann [8] that all D^* are pseudo-convex, hence D is pseudo-convex.

On the other hand, if D is pseudo-convex, then $-\log {}_D^{(N)}(z)$ is plurisubharmonic and will tend to infinity at any finite boundary point of D.

5.9. *Let D be a pseudo-convex domain. Then the Kontinuitätssatz holds for D.* (Compare 4.1).

Proof. If D is pseudo-convex, then $-\log d_D^{(N)}(z)$ is plurisubharmonic (5.7). Then the restriction of $-\log d_D^{(N)}(z)$ to a one dimensional analytic plane is subharmonic. For subharmonic functions the maximum principle holds (Radó [16]). Hence we have

$$\inf_{S_\nu \cup T_\nu} d_D^{(N)}(z) = \inf_{T_\nu} d_D^{(N)}(z) \qquad \text{for every } \nu.$$

The rest of the proof follows as in 4.1.

5.10. Most theorems which hold for plurisubharmonic functions and pseudo-convex domains in the finite dimensional case also hold in the infinite dimensional case. (For instance: *The intersection of two pseudo-convex domains is a pseudo-convex region.*) We have listed here only some of the very basic facts. The reader will find it not difficult to extend most of the theorems listed in *Complex convexity* (Bremermann [8]) to the infinite dimensional case.

In [8] we have stressed the formal relationship between *complex convexity* (by which notion we denote the plurisubharmonic functions and the pseudo-convex domains jointly) to ordinary convexity.

In the following section we will show that the same relationship persists in the infinite dimensional case.

6. Elementary convexity and its relation to complex convexity.

6.1. A real valued function $U(t)$ of one real variable t is called convex in an interval D of the real t-axis if and only if the following condition holds for every closed subinterval $D' \subset D$:

If $l(t)$ is a linear function such that $l(t) \geqq U(t)$ on the boundary of D', then $l(t) \geqq U(t)$ holds also for $t \subset D'$.

6.2. A real valued functional $U(x)$ defined in a domain D of a real Banach space B_r is called *convex* if and only if its restriction to an arbitrary straight line $\{x|x=x_0+ta\}$ is a convex function of t in $\{x|x=x_0+ta\} \cap D$.

REMARK. Formally these definitions are similar to the definitions of subharmonic and plurisubharmonic functionals (compare Bremermann [8]).

6.3. A domain D in a real Banach space B_r is convex if with any two points x_1 and x_2 the connecting straight line segment $\{x|x=x_1+t(x_2-x_1),\ 0 \leqq t \leqq 1\}$ is contained in D.

REMARK. This definition bears no formal relationship to the definition of the pseudo-convex domains. We will establish this relationship—as in the finite dimensional case (compare Bremermann [8])—by proving that a domain D is convex if and only if $-\log d_D^{(N)}(x)$ is a convex functional in D. The proof which we have given in [8] for finite dimension does not apply for infinite dimension, therefore a different one is given in the following. (Convexity in several complex variables has also been studied from a different point of view by Behnke-Stein [2].)

6.4. *Let D be a convex domain. Let S be an interval on a straight line $\{x|x=x_0+tb\}$, T the boundary of S. Let $l(t)$ be a linear function of t. Then we have for every $a \in B$, $||a||=1$*

$$\inf_{T} d_{a,D}^{(N)}(x_0+tb)e^{l(t)} = \inf_{S \cup T} d_{a,D}^{(N)}(x_0+tb)e^{l(t)} .$$

Proof. In the following we consider the subspace generated by the vectors a and b. Let the parameter values belonging to the two points of T be t_1 and t_2 $(t_1 < t_2)$.

We observe that all the points

$$x=x_0+t_1b+\vartheta d_{a,D}^{(N)}(x_0+t_1b)\cdot a$$

and

$$x = x_0 + t_2 b + \vartheta d_{a,D}^{(N)}(x_0 + t_2 b) \cdot a$$

belong to D for $-1 < \vartheta < 1$ by definition of $d_{a,D}^{(N)}(x)$. Now D is convex by assumption. Hence all the points on the connecting straight line segment connecting any two of these points belong to D. Now let $\inf_T d(x_0 + tb) e^{l(t)} = m > 0$. Then

$$d(x_0 + t_1 b) e^{l(t)_1} \geqq m \quad \text{and} \quad d(x_0 + t_2 b) e^{l(t)_2} \geqq m \ .$$

Hence all the points on the straight line segment passing through

$$x_1(\vartheta) = x_0 + t_1 b + \vartheta m e^{-l(t_1)} a \quad \text{and} \quad x_2(\vartheta) = x_0 + t_2 b + \vartheta m e^{-l(t_2)} a$$

belong to D for $-1 < \vartheta < 1$.

Now the function $me^{-l(t)}$ is a convex function of t for any linear function $l(t)$. Hence the curve

$$x(t_1, \vartheta) = x_0 + tb + \vartheta m e^{-l(t)} a \ , \qquad t_1 \leqq t \leqq t_2$$

will lie for $-1 < \vartheta < 1$ within the parallelogram through the four points $x_1(1)$, $x_1(-1)$, $x_2(1)$, $x_2(-1)$. Hence

$$d_{a,D}^{(N)}(x_0 + tb) \geqq m e^{-l(t)} \qquad \text{for} \quad t_1 \leqq t \leqq t_2 \ .$$

And therefore

$$d_{a,D}^{(N)}(x_0 + tb) e^{l(t)} \geqq m \ ;$$

and

$$\inf_{S \cup T} d_{a,D}^{(N)}(x_0 + tb) e^{l(t)} \geqq m \ .$$

6.5. *Let D be a domain such that for an arbitrary linear function $l(t)$ and line segments s with boundary T we have*

$$\inf_T d_{a,D}^{(N)}(x) e^{l(t)} = \inf_{S \cup T} d_{a,D}^{(N)}(x) e^{l(t)} \ .$$

Then $-\log d_{a,D}^{(N)}(x)$ is a convex function in D.

Proof. Suppose $-\log d_{a,D}^{(N)}(x)$ would not be convex. Then there would exist a straight line $\{x | x = x_0 + tb\}$ and a segment S with boundary T and a linear function $l(t)$, such that

$$l(t) \geqq -\log d_{a,D}^{(N)}(x_0 + tb)$$

on T, but there would exist a $t_0 \in S$ such that

$$l(t_0) \leqq -\log d_{a,D}^{(N)}(x_0 + t_0 b) \ .$$

This is equivalent to

$$1 \leq d_{a,D}^{(N)}(x_0+tb)e^{l(t)}$$

on T, and

$$1 > d_{a,D}^{(N)}(x_0+t_0b)e^{l(t_0)} .$$

Thus the minimum principle would be violated in contradiction to the assumption, hence the functional $-\log d_{a,D}^{(N)}(x)$ is convex in D.

6.6. Summing up 6.4 and 6.5: *If D is a convex domain, then for an arbitrary norm N (generating an equivalent topology) and for every $a \in B_r$ with $\|a\|=1$ the functional $-\log d_{a,D}^{(N)}(x)$ is convex in D.*

6.7. *If D is convex, then $-\log d_D^{(N)}(x)$ is convex in D for an arbitrary norm N (with equivalent topology).*

This is an immediate consequence of 6.6 because $-\log d_D^{(N)}(x)$ is the upper envelope of the family $\{-\log d_{a,D}^{(N)}(x)\}$ and the upper envelope of a locally upper bounded family of convex functionals is convex.

6.8. We now proceed to prove the converse of 6.7, and for this purpose we show first:

Let D be a domain such that for one particular norm the functional $-\log d_D^{(N)}(x)$ is convex. Then the following "Kontinuitätssatz" holds for D:

Let $\{S_\nu\}$ be a family of straight line segment, T_ν their boundaries and $S_0=\lim_{\nu\to\nu_0} S_\nu$, $T_0=\lim_{\nu\to\nu_0} T_\nu$. Then $\{S_\nu\}$, $T_\nu\}$, $T_0 \subset D$ implies $S_0 \subset D$.

The proof is analogous to the proof of 5.9 and 4.1.

6.9. *If for a domain D the Kontinuitätssatz* 6.8 *holds, then D is convex.*

We have to show: Let x_1, x_2 be two arbitrary points in D. Then we can connect them by a straight line segment.

Since D is a domain, we can connect x_1 and x_2 by a continuous arc $x(t)$. Let $x(0)=x_1$ and $x(1)=x_2$. We connect x_1 with $x(t)$ by the straight line segment. For small t the point $x(t)$ is in a neighborhood of x_1 and therefore the connecting line segment in D. Now, there cannot be a first line segment $\{x_1, x(t_1)\}$ such that for $0 \leq t < t_1$ the line segments are in D, however $\{x_1, x(t_1)\}$ is not, because this would violate the Kontinuitätssatz. Hence the line segment connecting x_1 and x_2 is in D, hence D is convex.

6.10. Summing up the results of this section:

A domain D is convex if and only if

(a) $-\log d_D^{(N)}(x)$ *is convex in* D. This property is invariant with respect to all topologically equivalent norms.

(b) $-\log d_{a,D}^{(N)}(x)$ *is convex in* D *for all* $a \in B_r$ *with* $\|a\|=1$. This property, too, is invariant with respect to all topologically equivalent norms.

(c) *The Kontinuitätssatz* 6.8 *holds for* D.

6.11. We add: D is convex if and only if the intersection of D with any finite dimensional (two dimensional) linear submanifold of B_r is convex. This is obvious. We note further that most of the theorems and analogies to complex convexity which for the finite dimensional case are explicated in [8] are true for the infinite dimensional case also. The reader will find it very easy to carry out the proofs himself.

7. Tube domains.

7.1. Let B_r be a real Banach space. Then we can define a complex Banach space by considering pairs of elements of B_r:

$$B_c = \{x, y\}\ , \qquad x \in B_r\ , \qquad y \in B_r\ ,$$

and by defining for complex scalars $\lambda = \sigma + i\tau$ the multiplication

$$\lambda(x, y) = (\sigma x - \tau y,\ \sigma y + \tau x)\ .$$

As usual, we will write $(x, y) = x + iy$. If $\|\ \ \|_r$ is the norm defined in B_r, then one defines

$$\|x + iy\|_c = (\|x\|_r^2 + \|y\|_r^2)^{1/2}$$

and one easily checks that the axioms are satisfied.

7.2. DEFINITION. Let B_c be a complex vector space and B_x and B_y its "real" and "imaginary" components. Then a *tube domain* is a domain that has the form

$$T_X = \{z | z = x + iy,\ x \in X,\ y \in B_y\}\ ,$$

where X is a domain in B_x called the *basis* of T_X. (The notion of tube domain was introduced in the finite dimensional case by Bochner-Martin [6].)

7.3. *Any functional* $f(z)$ *holomorphic in a tube domain* T_X *is determined throughout* T_X *already by its values in the basis* X *of* T_X.

Proof. Let us consider the analytic plane passing through the two points

$$z_1 = x_0 \quad \text{and} \quad z_2 = x_0 + iy_0 \;, \qquad x_0 \in X \;, \qquad y_0 \text{ arbitrary.}$$

The parameter representation is

$$z = z_1 + \lambda'(z_2 - z_1) = x_0 + i\lambda' y_0 = x_0 + \lambda y_0 \;.$$

Its intersection with the B_x is

$$z = x_0 + (\mathscr{R}\lambda)y_0 \;.$$

This is a straight line passing through x_0. On the intersection of this straight line with X the values of $f(z)$ are prescribed. Then $f(z)$ is determined on the whole plane strip

$$\{z \mid z = x_0 + \lambda y_0,\ \mathscr{R}(x_0 + \lambda y_0) \in X\}$$

by a classical theorem on functions of one complex variable.

The union of the analytic planes considered contains the whole tube T_X, hence we conclude that $f(z)$ is determined throughout T_X.

7.4. *An upper semi-continuous functional $V(z)$ defined in a tube domain T_X that does not depend upon the imaginary part of z is a plurisubharmonic functional in T_X if and only if its restriction to the basis X is a convex functional in X.*

Proof. Let $V(z)$ be plurisubharmonic. Then $V(z)$ is subharmonic on every analytic plane $\{z \mid z = z_0 + \lambda a\}$. Then the Laplacian

$$\left.\frac{\partial^2 V(z_0 + \lambda a)}{\partial\lambda\partial\bar{\lambda}}\right|_{\lambda=0} \geqq 0$$

for every $a \in B_c$ and $z_0 \in T_x$. (Taken in the sense of L. Schwartz.)

Let $\lambda = \sigma + i\tau$ and $a = c + id$. Then

$$z_0 + \lambda a = x_0 + iy_0 + \sigma c - \tau d + i(\tau c + \sigma d) \;.$$

Now if $V(z)$ does not depend upon y, then

$$V(z_0 + \lambda a) = V(x_0 + \sigma c - \tau d) \;.$$

Hence

$$\left.\frac{\partial^2 V(z_0 + \lambda a)}{\partial\lambda\partial\bar{\lambda}}\right|_{\lambda=0} = \left(\frac{\partial^2}{\partial\sigma^2} + \frac{\partial^2}{\partial\tau^2}\right) V(x_0 + \sigma c - \tau d)\Big|_{\sigma=\tau=0} \;.$$

For $d = 0$ it follows that

$$\left.\frac{\partial^2 V(x_0 - \sigma c)}{\partial\sigma^2}\right|_{\sigma=0} \geqq 0$$

and letting $c=0$ we obtain

$$\left.\frac{\partial^2 V(x_0-\tau d)}{\partial\tau^2}\right|_{\tau=0} \geqq 0 .$$

This means however that $V(x)$ is a *convex* function in X.

On the other hand, if $V(x)$ is convex, then

$$\left.\frac{\partial^2 V(x_0+\sigma c)}{\partial\sigma^2}\right|_{\sigma=0} \geqq 0 \quad \text{and} \quad \left.\frac{\partial^2 V(x_0-\tau d)}{\partial\tau^2}\right|_{\tau=0} \geqq 0$$

for every c, $d \in B_r$, $x_0 \in X$, and therefore

$$\left.\frac{\partial^2 V(z_0+\lambda a)}{\partial\lambda\partial\bar{\lambda}}\right|_{\lambda=0} \geqq 0$$

for every $a \in B_c$ and $z_0 \in T_x$. Hence $V(z)$ is subharmonic on every analytic plane. $V(z)$ is by assumption upper semi-continuous. Hence $V(z)$ is plurisubharmonic.

7.5. *A tube domain T_X is pseudo-convex if and only if its basis X is convex.*

COROLLARY. *A tube domain is pseudo-convex if and only if it is convex.*

Let T_X be pseudo-convex. Then $-\log d^{(N)}_{T_X}(z)$ is plurisubharmonic and it does not depend upon the imaginary part y. Then norm N generates a norm N' in B_x, and the restriction of $-\log d^{(N)}_{T_X}(z)$ to X is equal to $-\log d^{(N')}_X(x)$. Hence $-\log d^{(N')}_X(x)$ is convex, hence the domain X is convex by Theorem 6.10. And passing through the conclusions in the reverse direction we conclude conversely if X is convex, then T_X is pseudo-convex.

Obviously T_X is convex if and only if X is convex. Hence follows the corollary.

7.6. It can be shown that all holomorphic functionals can be continued into the "pseudo-convex envelope." In the case of our tube domain T_X the pseudo-convex envelope is that tube that has the convex envelope of X as its base.

This, however, we will study in a further paper.

REFERENCES

1. H. Behnke and F. Sommer, *Über die Voraussetzungen des Kontinuitätssatzes*, Math. Ann., **121** (1950), 356–378.

2. H. Behnke and K. Stein, *Die Konvexität in der Funktionentheorie mehrerer komplexer Veränderlichen*, Mitt. Math. Ges. Hamburg, **8** (1940).

3. H. Behnke and P. Thullen, *Theorie der Funktionen mehrerer komplexer Veränderlichen*, Ergebn. d. Math., **3**, no. 3 (1934).

4. S. Bergman, *Über eine in gewissen Bereichen gültige Integraldarstellung von Funktionen zweier komplexer Variabler*, Math., **39** (1934), 76–94 and 605-608.

5. ———, *Über eine Integraldarstellung von Funktionen zweier komplexer Veränderlicher*, Rec. Math. (Mat.), N.S. **1** (43), (1936), 851-862.

6. ——— and W. T. Martin, *Several complex variables*, Princeton 1948.

7. H. J. Bremermann, *Über die Äquivalenz der pseudokonvexen Gebiete und der Holomorphiegebiete im Raum von n komplexen Veränderlichen* Math. Ann., **128** (1954), 63–91.

8. ———, *Complex convexity*, to appear in Trans. Amer. Math. Soc.

9. C. Carathéodory *Funktionentheorie*, vol. 2 Basel 1950.

10. J. Eells, *The differential geometry of mapping spaces.* To appear.

11. E. Hille, *Functional analysis and semi-groups*, New York 1948.

12. P. Lelong, *Les fonctions plurisosharmoniques*, Ann. Sci. Ecole. Norm. Sup., **62** (1945), 301–338.

13. F. Norguet, *Sur les domaines d'holomorphie des fonctions uniformes de plusieur variables complexes*, Bull. Soc. Math. France, **82** (1954), 137–159.

14. K. Oka, *VI, Domaines pseudoconvexes*, Tôhoku Math. J., **49** (1942), 15–52.

15. ———, *IX, Domaines finis sans point critique intérieur*, Jap. J. Math., **23** (1953), 97–155.

16. T. Radó, *Subharmonic functions*, Ergebn. d. Math., **5**, no. 1 (1937).

17. H. Soeder, *Funktionentheorie in Banachschen Räumen*, Staatsarbeit at the Math. Inst. Univ. Münster, 1954; Thesis Münster 1956.

18. A. E. Taylor, *Analysis in complex Banach spaces*, Bull. Amer. Math. Soc., **49** (1943), 652-669.

19. G. O. Thorin, *Convexity theorems*, Medd. Lunds Univ. Mat. Sem., **9** (1948).

20. A. Weil, *L'intégrale de Cauchy et les fonctions de plusieur variables*, Math. Ann., **111** (1935), 178-182.

THE INISTITUTE FOR ADVANCED STUDY

MINIMIZING INTEGRALS IN CERTAIN CLASSES OF MONOTONE FUNCTIONS

H. D. BRUNK, G. M. EWING AND W. R. UTZ

1. Introduction. This paper is concerned with the existence, uniqueness and representation of minimizing functions. It includes many results of [1] and [2]. Applications are discussed in [3].

The authors are indebted for various ideas to W. T. Reid with whom Brunk and Ewing collaborated in a study [2] of a particular integral (1.4) in the one-variable case. Also, the authors wish to acknowledge the helpful suggestions of the referee.

Extension to n variables and to more general integrands is of interest per se and is motivated by a variety of problems.

For example, let $\boldsymbol{x}$ ($\boldsymbol{y}$) be the random variable, maximum dilution (that is, unity minus concentration) of an insecticide I (J) which is lethal to an insect from a given population. Then

$$p(x, y) = \Pr\{\boldsymbol{x} > x \text{ or } \boldsymbol{y} > y\}$$

is the probability of death for an insect similtaneously dosed with respective dilutions x, y of I, J. Moreover

$$F(x, y) = 1 - p(x, y) = \Pr\{\boldsymbol{x} \leqq x \text{ and } \boldsymbol{y} \leqq y\} , \tag{1.1}$$

is the probability of survival and is a distribution function [5; pp. 78, 260]; hence $p(x, y)$ is nonincreasing in each variable and for each point-pair (x, y), (x', y'),

$$\Delta^2 p = p(x', y') - p(x', y) - p(x, y') + p(x, y) \leqq 0 . \tag{1.2}$$

For each of selected pairs (x_i, y_j) let $\Delta\mu_{ij}$ insects be dosed and let α_{ij} denote the fraction of the sample which is killed. The maximum likelihood estimate $P(x, y)$ of $p(x, y)$ is that function, subject to the restrictions stated above, which maximizes the product

$$\prod p_{ij}^{\alpha_{ij}\Delta\mu_{ij}}(1-p_{ij})^{(1-\alpha_{ij})\Delta\mu_{ij}} , \qquad p_{ij} = p(x_i, y_j) . \tag{1.3}$$

Equivalently, $P(x, y)$ minimizes the integral

$$-\int [\alpha \log p + (1-\alpha) \log (1-p)] d\mu , \tag{1.4}$$

in which μ describes the mass distribution consisting of masses $\Delta\mu_{ij}$ at

Received December 5, 1955, and in revised form March 22, 1956. This research was supported by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

the respective points (x_i, y_j) and no mass elsewhere.

Other problems, for example, [3, p. 610] require only that the function $P(x, y)$ minimizing (1.4) be monotone in each variable and not that it satisfy (1.2). As a further example of this type, suppose that $\alpha(t)$, $t=(t^1, \cdots, t^n)$, not necessarily monotone in any t^i, is a given approximation to $\theta(t)$ a function required to be monotone in each variable. The least squares determination $\Theta(t)$ of $\theta(t)$ minimizes the integral $\int[\theta(t)-\alpha(t)]^2 d\mu(t)$.

2. Formulation and preliminary lemmas. Given a fixed positive integer n and the space R_n with points $t=(t^1, \cdots, t^n)$, let μ be a measure defined on a Borel field $\mathscr{B}$ of subsets of R_n which is *totally finite*, that is, $R_n \in \mathscr{B}$, $\mu(R_n) < \infty$, and *complete*, that is, if $A \subset E \in \mathscr{B}$ and $\mu(E)=0$, then $A \in \mathscr{B}$ and $\mu(A)=0$. The term *measure* will mean μ-measure unless otherwise specified, *measurable set* will mean a set in $\mathscr{B}$, and *measurable function* a μ-measurable function. In particular μ can be a finite Lebesgue-Stieltjes measure.

Let I be a fixed nondegenerate interval of extended real numbers which includes its endpoints $a \geqq -\infty$, $b \leqq \infty$. Let $F(u, v)$ be an extended real-valued function for $u, v \in I$ subject to the following conditions.

(2.1) If $\alpha(t)$, $\theta(t)$ with ranges in I are both measurable then so is $F[\alpha(t), \theta(t)]$.

(2.2) For fixed u in I, either (i) $F(u, v)=\infty$ for $v \neq u$, with $F(u, u) < \infty$ or (ii) $F(u, v)$ is strictly decreasing (increasing) in v for $a \leqq v \leqq u (u \leqq v \leqq b)$ and right (left) continuous in v for $a < v < u (u < v < b)$. (See (5.3) to (5.6) for examples.)

$$\int F[\alpha(t), \alpha(t)]d\mu(t) > -\infty . \tag{2.3}$$

For fixed $\alpha(t)$ and arbitrary $\theta(t)$ with ranges in I and both measurable define

$$J[\theta]=\int F[\alpha(t), \theta(t]d\mu(t) . \tag{2.4}$$

Let M denote the class of all measurable functions $\theta(t)$ with ranges in I such that $J[\theta]$ exists finite or infinite and such that $\theta(t)$ is nondecreasing in each coordinate t^i of t. Define M^* as M if $n=1$ and, for $n>1$, let M^* consist of those $\theta(t)$ in M with the property that the difference $\Delta^2\theta$, defined as in (1.2) for each pair t^i, t^j, (with the other variables fixed for each choice of i, j) shall be nonnegative in the complement of the closure of the set on which $\theta(t)=\infty$ or $\theta(t)=-\infty$. The principal problems of this paper to minimize J in M and in M^*.

The methods apply, with suitable small changes, to problems like that of § 1 in which (1.2) is required with $\leqq$ instead of $\geqq$ and in which admissible functions are nonincreasing in the separate variables.

The relation $t_1 < t_2$ means that $t_1^i < t_2^i$, $i=1, \cdots, n$ while $t_1 \leqq t_2$ means that $t_1^i \leqq t_2^i$ for each i. Given a point v consider intervals of the types $(t : t < v)$ and $(t : t \leqq v)$. A measurable set L which is a union of intervals of the first and (or) second of these types is termed a *lower layer*. A measurable set L is then a lower layer if and only if $v \in L$ and $t \leqq v$ imply $t \in L$. An *upper layer* U is similarly defined. The complement $\tilde{L}$ of a lower layer L is an upper layer. If L is not void the common boundary of L and $\tilde{L}$ is called a *monotone graph*. Given a lower layer L and an upper layer U the measurable set $UL = L - \tilde{U}$ is termed a *layer*. For $n=1$, a layer is an interval of the reals which may be void, degenerate or of positive length and, in the latter case, may include either, neither, or both of its endpoints. The layer is the natural extension (for the purposes of this study) of the notion of interval. A monotone graph is connected and is a layer but, for $n > 1$, a layer need not be connected.

LEMMA 2.1. *Let $\mathscr{O}$ denote the union of all open sets of measure* 0. *Then $\mu(\mathscr{O})=0$ and given $t \notin \mathscr{O}$, every layer containing a neighborhood of t has positive measure.*

LEMMA 2.2. *If $\theta(t)$ is measurable and monotone nondecreasing in each variable t^i, then the set of points t, for which $\theta(t)$ is on a given finite or infinite interval of the reals, is a layer.*

The proofs of these lemmas are easy.

LEMMA 2.3. *If $\theta(t)$ is monotone nondecreasing in each variable t^i, the discontinuities of $\theta(t)$ lie on a countable set of monotone graphs.*

This result is Theorem 7 in [4].

LEMMA 2.4. *A monotone graph is of Lebesgue measure* 0.

Proof. The metric density is less than unity at each point of a monotone graph. Alternatively, observe that a line with direction numbers $(1, \cdots, 1)$ cuts a monotone graph in exactly one point and use Fubini's Theorem.

3. Existence theorems. Denote the respective infima of J in M, M^* by γ, γ^*.

LEMMA 3.1. *If* $\theta(t)$, $q=1, 2, \cdots$, *is a sequence in* M^*, *there exists a subsequence* $\theta_q^*(t)$ *of* $\theta_q(t)$ *and* $\theta(t) \in M^*$ *such that* $\lim \theta_q^*(t)=\theta(t)$ *except at most on two monotone graphs.*

Theorem 3 of [4] establishes this result. The exceptional sets, denoted by A and Ω, are respectively boundaries of layers on which $\theta(t)=-\infty, \infty$.

LEMMA 3.2. *If* $n=1$, *then any sequence in* $M=M^*$ *contains a subsequence converging everywhere on* R_1 *to a function* $\theta(t)$ *in* $M=M^*$.

Proof. If $n=1$ each of the sets A, Ω is either void or consists of a single point; hence the sequence $\theta_q^*(t)$ can be further refined to yield convergence (possibly to ∞ or $-\infty$) everywhere on R_1.

THEOREM 3.1. EXISTENCE THEOREM FOR M^*. *There exists a function* $\Theta(t)$ *in* M^* *such that* $J[\Theta]=\gamma^*$.

Proof. Attention is confined to the nontrivial case $\gamma^* < \infty$. Let $\theta_q(t)$ be a sequence in M^* such that $\lim J[\theta_q]=\gamma^*$. By Lemma 3.1 we may suppose that $\theta_q(t)$ converges, for $t \in R_n - A \cup \Omega$, to $\Theta(t) \in M^*$. Let $\theta_*(t)=\liminf \theta_q(t)$, $\theta^*(t)=\limsup \theta_q(t)$. Extend $\Theta(t)$ to $A \cup \Omega$ by the definition,

$$\Theta(t)\begin{cases} =\alpha(t)\,, & \text{if} \quad \theta_*(t) \leqq \alpha(t) \leqq \theta^*(t)\,, \\ =\theta_*(t)\,, & \text{if} \quad \alpha(t) < \theta_*(t)\,, \\ =\theta^*(t)\,, & \text{if} \quad \alpha(t) > \theta^*(t)\,. \end{cases} \tag{3.1}$$

Clearly $\Theta(t)$ is measurable. One varifies that $\Theta(t)$ is in M^*.

For fixed t, it follows from the definition of $\Theta(t)$ and property (2.2) of F that $F[\alpha(t), \Theta(t)] \leqq F[\alpha(t), v]$ for $\theta_*(t) \leqq v \leqq \theta^*(t)$. Since each point of accumulation of the sequence $\theta_q(t)$ lies in the interval $[\theta_*(t), \theta^*(t)]$ we have

$$F[\alpha(t), \Theta(t)] \leqq \liminf F[\alpha(t), \theta_q(t)]\,.$$

From Fatou's Lemma [6, p. 113; 7, p. 167] it then follows that

$$J[\Theta]=\int F[\alpha(t), \Theta(t)]d\mu(t) \leqq \liminf \int F[\alpha(t), \theta_q(t)]d\mu(t)\,;$$

hence $J[\Theta]=\gamma^*$.

THEOREM 3.2. EXISTENCE THEOREM FOR M. *If* $n=1$ *or* 2 *there exists a function* $\Theta(t)$ *in* M *such that* $J[\Theta]=\gamma$.

Proof. If $n=1$ the conclusion is contained in that of the preceding theorem. For $n=2$, let $\theta_q(t)$ be a sequence in M such that $\lim J[\theta_q]=\gamma$.

By Lemma 3.2 and the usual diagonalization process there exists a subsequence converging at all points with at least one rational coordinate.

Define $\theta_*(t)$, $\theta^*(t)$ as in the proof of Theorem 3.2 noting that $\theta_*(t)=\theta^*(t)$ if t^1 or t^2 is rational. From the density of these points in R_2, $\theta_*(t)=\theta^*(t)$ at any point t at which both functions are continuous; hence by Lemma 2.3 everywhere except on a countable set of monotone graphs. Define $\Theta(t)$ on the space R_2 as in (3.1).

If $t_1 \leqq t_2 \neq t_1$, the segment with endpoints t_1, t_2 is cut by at least one line on which t^1 is rational or on which t^2 is rational in a point t_0. One sees that

$$\Theta(t_1) \leqq \theta^*(t_1) \leqq \theta^*(t_0) = \theta_*(t_0) \leqq \theta_*(t_2) \leqq \Theta(t_2);$$

hence that Θ is nondecreasing in t^1 and in t^2 and is in M. The proof can be completed by following that of Theorem 3.1.

The point t_0 essential to the last proof need not exist for $n>2$.

A function $t(\tau)$, τ on a finite or infinite interval will be termed a *monotone nondecreasing vector-function* if $\tau_2 > \tau_1$ implies that $t(\tau_2) \geqq t(\tau_1)$. If $\theta(t)$ is nondecreasing in each variable t^i and $t(\tau)$ has the above property, then $\theta[t(\tau)]$ is nondecreasing in the real variable τ. *Monotone nonincreasing vector-functions* are similar. The graph of a monotone vector-function is a monotone graph in the sense of § 2 only for certain cases when $n=1$ or 2.

In the following theorem we suppose the class of measurable sets is contained in the class of Lebesgue measurable subsets of R_n. These is then a Lebesgue decomposition [6, p. 134] of μ; that is, μ is the sum of a measure α absolutely continuous with respect to Lebesgue measure λ and a measure σ singular with respect to λ. Thus if $\lambda(E)=0$, then $\alpha(E)=0$ and there is a decomposition of R_n into complementary sets A, $\tilde{A}$ such that $\lambda(A)=0$ and $\sigma(\tilde{A})=0$.

THEOREM 3.3. EXISTENCE THEOREM FOR M. *If $\mu=\alpha+\sigma$ is the Lebesgue decomposition of the given measure μ, and if the mass in R_n described by the singular part σ all lies on the graphs of a counterable set of monotone vector-functions, then there exists a function $\Theta(t)$ in M such that $J[\Theta]=\gamma$.*

This theorem applies in particular if σ describes a discrete mass distribution or if μ is Lebesgue measure. The proof, along lines similar to those followed in preceding theorems, is omitted.

4. Integrands generated by convex functions. The class of problems for which we are able to give more complete results in this section is more restricted than that of § 3. We are moreover primarily interested in cases in which the minimum of J in M is finite. It is convenient to introduce $\mathscr{M}$ to denote the subset of M consisting of those θ in M such that $J[\theta]<\infty$. Results of this section are for the minimum problem in $\mathscr{M}$ only.

Let I again denote a fixed nondegenerate interval of extended reals which includes its endpoints $a\geqq-\infty$ and $b\leqq\infty$. Let $T(z)$ be a continuous convex function of the real variable z on the interior I_0 of I. The derivative $T'(z)$ exists except on a countable subset of I_0 and it seems convenient to extend $T'(z)$ to I_0 by assigning it the value of the left derivative at each point of I_0, thereby making $T'(z)$ left-continuous on I_0. The extended real-valued function $F(u, v)$ is defined as follows:

$$F(u, v)=T(u)-T(v)-(u-v)T'(v)\ , \qquad u, v\in I_0\ . \tag{4.1}$$

The right member of (4.1) has an obvious interpretation in terms of the tangent to the graph of $T(z)$. $F(u, v)$ is extended to $I\times I$ by the additional definitions

$$\begin{aligned} F(a, v)&=\lim_{u\to a}F(u, v)\ , && v\in(a, b)\ ,\\ F(b, v)&=\lim_{u\to b}F(u, v)\ , && v\in(a, b)\ ,\\ F(u, a)&=\lim_{v\to a}F(u, v)\ , && u\in(a, b]\ ,\\ F(u, b)&=\lim_{v\to b}F(u, v)\ , && u\in[a, b)\ , \end{aligned} \tag{4.2}$$

One verifies that, for $u, v\in I_0$,

$$\begin{aligned} F(u, v)&=\int_{[u, v)}(z-u)dT'(z)\ , && \text{if} \quad u<v\ ,\\ &=-\int_{[v, u)}(z-u)dT'(z)\ , && \text{if} \quad u>v\ ,\\ &=0 && \text{if} \quad u=v\ . \end{aligned} \tag{4.3}$$

Essentially such functions F generalizing the particular integrand of [2] have been suggested independently by Reid.

Such functions F arise in connection with the applications (cf. examples in § 1, also [3], where $\exp\{-F[g(x), \theta]\}$ is the density function, with respect to a measure, of a random variable whose distribution belongs to the exponential family). F as defined above is nonnegative, and has properties (2.1) and (2.2) (except that F need not be right-continuous in v for $a<v<u$ and F is strictly monotone in v for $v\leqq u$ and for $v\geqq u$ only if T is strictly convex).

We again let t denote the generic point in R_n, let μ denote a totally finite complete measure on the given Borel field $\mathscr{B}$, and let $\alpha(t)$ denote a given integrable function with range in I such that $T[\alpha(t)]$ is integrable. It follows that $J[\theta]<\infty$ when $\theta(t)\equiv\theta_0$, θ_0 a constant in I_0, so that $\theta(t)\equiv\theta_0\in\mathscr{M}$.

For a measurable subset A or R_n we define

$$J[\theta\,;A]=\int_A F[\alpha(t),\,\theta(t)]d\mu(t)\,, \tag{4.4}$$

$$M(A)=\int_A \alpha(t)d\mu(t)/\mu(A)\,,\qquad \text{if}\quad \mu(A)>0\,. \tag{4.5}$$

THEOREM 4.1. *Let ν be a finite signed measure on the class of measurable subsets of R_n, absolutely continuous with respect to μ. Then there exist an upper layer P and a lower layer N such that*

(i) $\nu(PL)>0$ *for every lower layer L such that* $\mu(PL)>0$;

(ii) $\nu(U\tilde{P})\leqq 0$ *for every upper layer U;*

(iii) $\nu(UN)<0$ *for every upper layer U such that* $\mu(UN)>0$;

(iv) $\nu(\tilde{N}L)\geqq 0$ *for every lower layer L.*

Proof. The proof is an adaptation of that of the Hahn-Jordan decomposition theorem [6, p. 121] and will simply be sketched here in broad outline. Let $\mathscr{N}$ denote the class (a class of sets having a nonpositive property) of lower layers L such that $\nu(UL)\leqq 0$ for every upper layer U. Choose a sequence of lower layers in $\mathscr{N}$ whose measures approach $\beta=\sup_{L\in\mathscr{N}}\mu(L)$; one readily verifies that their union, $\tilde{P}$, is a maximal element of $\mathscr{N}$; that is, $\tilde{P}$ belongs to $\mathscr{N}$ and has measure β. Thus the lower layer $\tilde{P}$ has the nonpositive property (ii). It is possible that the void set is the only element of $\mathscr{N}$, in which event $\tilde{P}=\phi$. We shall now show that P, the complement of $\tilde{P}$, has also the positive property (i). Suppose the contrary. Then there is a lower layer $T\supset\tilde{P}$ such that $\nu(PT)\leqq 0$, while $\mu(PT)>0$, so that $T\notin\mathscr{N}$ (since $\tilde{P}$ is maximal). Hence there is an upper layer $U\subset P$, $U\supset\tilde{T}$, such that $\nu(UT)>0$. One may then determine an expanding sequence (as in the proof in [6], pp. 121–122, of the existence of a Hahn decomposition) U_i, $i=1,2,\cdots$, of upper layers, contained in P and containing $\tilde{T}$, whose limit, U^*, has a complement, $\tilde{U}^*$, belonging to $\mathscr{N}$, while

$$\nu(U_iU_{i-1})=\nu(U_i-U_{i-1})>0\,,\qquad i=1,2,\cdots;\quad U_0=\tilde{T}\,.$$

From the maximality of $\tilde{P}$ it follows that $\mu(P\tilde{U}^*)=0$, whence $\nu(PU^*)=0$. On the other hand,

$$\nu(U^*T)=\sum_{i=1}^{\infty}\nu(U_i-U_{i-1})>0\,,$$

so that $\nu(PT) > 0$, a contradiction. Thus P does indeed have the positive property (i). The determination of a lower layer N, possibly void, with the desired properties follows similarly on the introduction of a class $\mathscr{P}$ of upper layers U such that $\nu(UL) \geqq 0$ for every lower layer L. For each real x, define a lower layer N_x and an upper layer P_x as the lower and upper layers N and P given by Theorem 4.1 corresponding to the signed measure

$$\nu(A) = \int_A [\alpha(t) - x] d\mu(t) .$$

As a consequence of Theorem 4.1 applied to this signed measure, we have for upper and lower layers U and L,

$$
\begin{aligned}
&M(P_xL) > x \quad \text{if} \quad \mu(P_xL) > 0 , \\
&M(U\tilde{P}_x) \leqq x \quad \text{if} \quad \mu(U\tilde{P}_x) > 0 , \\
&M(UN_x) < x \quad \text{if} \quad \mu(UN_x) > 0 , \\
&M(\tilde{N}_xL) \geqq x \quad \text{if} \quad \mu(N_xL) > 0 ,
\end{aligned}
\tag{4.6}
$$

LEMMA 4.1. *If Λ is an index set and A_λ, $\lambda \in \Lambda$, is a family of measurable subsets of R_n such that $\mu(A_\lambda A_\sigma) = 0$ for $\Lambda \ni \sigma \neq \lambda \in \Lambda$, then $\mu(A_\lambda) = 0$ except for at most a countable subset of Λ.*

Proof. If the lemma is false then there is a positive number ϵ and a sequence of sets of the family $\{A_\lambda, \lambda \in \Lambda\}$, each having measure greater than ϵ. It follows from the hypothesis

$$\mu(A_\lambda A_\sigma) = 0 \quad \text{for} \quad \Lambda \ni \sigma \neq \lambda \in \Lambda$$

that the usual technique of replacing the sets of a sequence by mutually disjoint sets while preserving their union yields a sequence of disjoint sets each having measure greater than ϵ, so that their union has infinite measure, contradicting the property of μ of being totally finite.

COROLLARY 4.1. *$\mu(P_xN_x) = 0$ for every real x and $\mu(\tilde{N}_x\tilde{P}_x) = 0$ for all but a countable set of real numbers x.*

Proof. If $\mu(P_xN_x) > 0$, the first and third relations (4.6) yield the contradiction $x < x$. It can be seen as follows that the second conclusion is a consequence of Lemma 4.1. Since

$$\tilde{N}_x\tilde{P}_x \cap \tilde{N}_y\tilde{P}_y \subset \tilde{N}_y\tilde{P}_x ,$$

it follows that when $x < y$ and $\mu(\tilde{N}_y\tilde{P}_x) > 0$, then $y \leqq M(N_yP_x) \leqq x$, which is a contradiction. It follows that

$$\mu(\tilde{N}_x\tilde{P}_x \cap \tilde{N}_y\tilde{P}_y)=0$$

for $x \neq y$.

It is convenient to determine the upper and lower layers P_x and N_x so that

$$N_x \subset \tilde{P}_x \subset N_y \quad \text{for} \quad x<y \quad (\text{or } P_x \supset \tilde{N}_y \supset P_y \text{ for } x<y)\,, \tag{4.7}$$

$$N_x = \bigcup_{y<x} N_y\,, \qquad P_x = \bigcup_{y>x} P_y\,. \tag{4.8}$$

Let E denote the countable set consisting of reals r which are rational or for which $\mu(\tilde{N}_r\tilde{P}_r)>0$. It can be shown that

$$N_x^* = \bigcup_{E \ni r<x} (N_r \cup \tilde{P}_r)\,, \qquad P_x^* = \bigcup_{y>x} \bigcap_{E\ni r\leq y} (\tilde{N}_r \cap P_r)\,,$$

have properties (4.7) and (4.8) and that relations (4.6) hold with N_x^*, P_x^* in place of N_x, P_x. We shall understand from here on that this replacement has been made, but shall omit the asterisks.

Let us define $\Theta(t)$ as the infimum of those x such that $t \in \tilde{P}_x$.

LEMMA 4.2.

$$\Theta(t)>x \quad \textit{if and only if} \quad t \in P_x\,. \tag{4.9}$$

$$\Theta(t)<x \quad \textit{if and only if} \quad t \in N_x\,. \tag{4.10}$$

$$\Theta(t)=x \quad \textit{if and only if} \quad t \in (\tilde{N}_x\tilde{P}_x)\,. \tag{4.11}$$

$$\Theta(t)=\sup_{t \notin N_x} x\,. \tag{4.12}$$

Proof of (4.9). From its definition, $\Theta(t) \leq x$ if $t \notin P_x$. If $\Theta(t)=x_0>x$, then $t \in P_y$ for $y<x_0$; hence $t \in \bigcup_{y>x} P_y = P_x$.

Proof of (4.10). If $t \notin N_x$, then $t \in P_y$ for each $y<x$; hence

$$\Theta(t) = \inf_{t \notin P_y} y \geq x\,.$$

If $t \in N_x = \bigcup_{y<x} N_y$, there exists $y_0<x$ such that $t \in N_y$ for $y \geq y_0$; hence $t \in \tilde{P}_y$ for $y>y_0$; hence

$$\Theta(t) = \inf_{t \notin P_y} y \leq y_0 < x\,.$$

Relation (4.11) follows from (4.9) and (4.10).

Proof of (4.12). Set $\Theta_1(t) = \sup_{t \notin N_x} x$. Arguments similar to the above show that $\Theta_1(t)$ satisfies (4.9), (4.10), (4.11); hence that $\Theta(t) \equiv \Theta_1(t)$.

We remark that, for $\theta \in \mathscr{M}$,

$$M\{t:\ \Theta(t) \leqq z < \theta(t)\} \leqq z\ , \qquad M\{t:\ \Theta(t) \geqq z > \theta(t)\} \geqq z\ ,$$

provided that the measures of these sets are positive. Each strict inequality between z and $\theta(t)$ in these statements may be replaced by the corresponding weak inequality.

LEMMA 4.3. *If $\theta(t) \in \mathscr{M}$, if E is a measurable set, if a_n is a sequence of real numbers strictly decreasing to a (b_n a sequence strictly increasing to b), if $\theta_n(t) = \max[\theta(t), a_n]$ ($\min[\theta(t), b_n]$), then $\theta_n(t) \in \mathscr{M}$, $n=1, 2, \cdots$, and $\lim J[\theta_n; E] = J[\theta; E]$.*

Proof We recall that the function of t assuming the constant value a_n is in $\mathscr{M}$, and that, as a function of v, $F(u, v)$ is nondecreasing for $v > u$ and nonincreasing for $v < u$. Since

$$\theta(t) \leqq \theta_n(t) \leqq \theta_1(t) = \max[\theta(t), a_1]\ ,$$

we have

$$\begin{aligned} 0 \leqq F[\alpha(t), \theta_n(t)] &\leqq \max\{F[\alpha(t), \theta(t)],\ F[\alpha(t), \theta_1(t)]\} \\ &\leqq \max\{F[\alpha(t), \theta(t)],\ F[\alpha(t), a_1]\}\ . \end{aligned}$$

The functions $F[\alpha(t), \theta(t)]$, $F[\alpha(t), a_1]$ are integrable; so then is the function $\max\{F[\alpha(t), \theta(t)],\ F[\alpha(t), a_1]\}$. Also

$$\lim \theta_n(t) = \theta(t),\ \lim F[\alpha(t), \theta_n(t)] = F[\alpha(t), \theta(t)]\ ,$$

and by the dominated convergence theorem, $\lim J[\theta_n; E] = J[\theta; E]$.

LEMMA 4.4. *Given $\theta', \theta'' \in \mathscr{M}$, let*

$$E = \{t:\ \theta'(t) < \theta''(t)\} \quad \textit{and} \quad E(z) = \{t:\ \theta'(t) \leqq z < \theta''(t)\}\ .$$

Then

$$J[\theta''; E] - J[\theta'; E] = \int_{(a,b)} \{z - M[E(z)]\}\, \mu[E(z)]\, dT'(z)\ .$$

Proof. Let a_n and b_n, $n=1, 2, \cdots$, be sequences strictly decreasing and increasing to the endpoints a and b of I repectively. Set

$$\theta'_n(t) = \max[\theta'(t), a_n]\ , \quad \text{and} \quad \theta''_n(t) = \min[\theta''(t), b_n]\ ,$$

$n=1, 2, \cdots$. We have

$$J[\theta_n''\,;E]-J[\theta_n'\,;E]=\int_E \{F[\alpha(t)\,;\theta_n''(t)]-F[\alpha(t)\,;\theta_n'(t)]\}d\mu$$
$$=\int_E d\mu\int_{[\theta_n'(t),\theta_n''(t))}[z-\alpha(t)]dT'(z)\ .$$

For fixed n, set

$$A=\{(z,\,t):\ \theta'(t)\leqq z<\theta''(t)\ ,\quad a_n\leqq z<b_n\}\ .$$

Both z and $\alpha(t)$ are integrable over A with respect to the product measure $(d_\mu\times dT')$, so that Fubini's Theorem permits a change in the order of integration. We have that

$$J[\theta_n'';\,E]-J[\theta_n';\,E]=\int_{[a_n,b_n)}dT'(z)\int_{E(z)}[z-\alpha(t)]d\mu$$
$$=\int_{[a_n,b_n)}\{z-M[E(z)]\}\,\mu[E(z)]dT'(z)\ .$$

Applying Lemma 4.3 and taking limits as $n\to\infty$ we obtain the desired conclusion.

THEOREM 4.2. *$\Theta(t)$ minimizes J in $\mathscr{M}$.*

Proof. For $\theta(t)$ in $\mathscr{M}$, set

$$B_1=\{t:\ \Theta(t)<\theta(t)\}\ ,$$
$$B_2=\{t:\ \Theta(t)>\theta(t)\}\ ,$$
$$B_3=\{t:\ \Theta(t)=\theta(t)\}\ .$$

Then

$$J[\Theta]=\Sigma\, J[\Theta\,;\,B_i]$$

and similarly for $J[\theta]$. We have $J[\theta;\,B_3]-J[\Theta\,;\,B_3]=0$. In Lemma 4.4 set $\Theta=\theta'$, $\theta=\theta''$ so that E becomes B_1 and $E(z)$ becomes that set $\{t:\ \Theta(t)\leqq z<\theta(t)\}$. From Lemma 4.2 (see remark preceding Lemma 4.3) it follows that $M[E(z)]\leqq z$ if $\mu[E(z)]>0$; hence from Lemma 4.4 that

$$J[\theta;\,B_1]-J[\Theta;\,B_1]\geqq 0\ . \tag{4.13}$$

Now set $\Theta=\theta''$, $\theta=\theta'$ in Lemma 4.4 and then $E=B_2$ and

$$E(z)=\{t:\ \Theta(t)>z\geqq\theta(t)\}\ .$$

Again, from Lemma 4.2, $M[E(z)]\geqq z$ if $\mu[E(z)]>0$; hence, from Lemma 4.4,

$$J[\theta;\, B_2]-J[\Theta;\, B_2]\geqq 0\ . \tag{4.14}$$

Adding (4.13) and (4.14) we find that $J[\theta]-J[\Theta]\geqq 0$, completing the proof.

By (4.11), the minimizing function $\Theta(t)$ assumes a given value x on the layer $\tilde{N}_x\tilde{P}_x$. In calculating for specific examples it is useful to observe as a consequence of equations (4.6) that if $\mu(\tilde{N}_x\tilde{P}_x)>0$ then $\tilde{N}_x\tilde{P}_x$ is the maximal layer among layers $\tilde{N}_xL$ over which the mean is minimal:

$$M(\tilde{N}_x\tilde{P}_x)=x\leqq M(\tilde{N}_xL)$$

if $\mu(\tilde{N}_xL)>0$; while if $\tilde{N}_xL\supset\tilde{N}_x\tilde{P}_x$ and if $M(\tilde{N}_xL)=M(\tilde{N}_x\tilde{P}_x)$, then $\tilde{N}_xL$ and $\tilde{N}_x\tilde{P}_x$ differ by a set of measure 0. Similarly $\tilde{N}_x\tilde{P}_x$ is the maximal layer among layers $U\tilde{P}_x$ over which the integral mean of $\alpha(t)$ is maximal.

We term the subset of a neighborhood of a point t_0 consisting of points $t>t_0$, an *upper neighborhood* of t_0 and the subset consisting of points $t<t_0$, a *lower neighborhood* of t_0. Let $\mathscr{D}$ denote the set of points each of which has an upper or a lower neighborhood in $\mathscr{O}$ (defined in Lemma 2.1); $\mathscr{D}\supset\mathscr{O}$.

THEOREM 4.3. REPRESENTATION THEOREM. *If, given $\varepsilon>0$, we have $\mu(UN_{c+\varepsilon})>0$ for every upper layer U containing a given point t_0, and $\mu(P_{c-\varepsilon}L)>0$ for every lower layer L containing t_0, where $c=\Theta(t_0)$, then*

$$\Theta(t_0)=\sup_{U\ni t_0}\inf_{L\ni t_0} M(UL)\ , \tag{4.15}$$

$$\Theta(t_0)=\inf_{L\ni t_0}\sup_{U\ni t_0} M(UL)\ , \tag{4.16}$$

$$\Theta(t_0)=\sup_{U\ni t_0}\inf_{L} M(UL)\ , \tag{4.17}$$

$$\Theta(t_0)=\inf_{L\ni t_0}\sup_{U} M(UL)\ . \tag{4.18}$$

In particular (4.15), $\cdots$, (4.18) *hold if t_0 is a mass point of μ or if t_0 is a point of continuity of $\Theta(t)$ not in $\mathscr{D}$.*

We note that the measure of $\mathscr{O}$ is 0, and that the Lebesgue measure of $\mathscr{D}$-$\mathscr{O}$ is 0. Further, since $\Theta(z)\in\mathscr{M}$, its discontinuities lie on a countable set of monotone graphs (Lemma 2.3.). Theorem 4.3 thus gives almost everywhere representations of $\Theta(t)$, provided that μ is absolutely continuous with respect to Lebesgue measure, or provided that the Lebesgue singular part of μ concentrates its mass at a countable number of mass points. In general, these representations need not be valid almost everywhere.

Proof of (4.15) *and* (4.17). Given $U \ni t_0$ and given $\varepsilon > 0$, then $\mu(UN_{c+\varepsilon}) > 0$ by hypothesis, so that by (4.6), $M(UN_{c+\varepsilon}) < c+\varepsilon$. Hence

$$\inf_{L \ni t_0} M(UL) \leqq c\ , \quad \text{if} \quad U \ni t_0\ , \tag{4.19}$$

and also

$$\inf_{L} M(UL) \leqq c \quad \text{if} \quad U \ni t_0\ . \tag{4.20}$$

Further, if $\mu(\tilde{N}_c L) > 0$, then by (4.6), $M(\tilde{N}_c L) \geqq c$. But $\tilde{N}_c \ni t_0$, and hence relations (4.15) and (4.17) follow respectively from (4.19) and (4.20).

Relations (4.16) and (4.18) may be proved similarly.

We note that under the hypotheses of Theorem 4.3, if $\alpha(t)$ is monotone nonincreasing in each argument, then the constant function

$$\Theta(t) \equiv \int \alpha d\mu \Big/ \int d\mu$$

minimizes J in M and also in M^*. If $\alpha(t)$ is in the class $M(M^*)$ then clearly $\Theta(t) \equiv \alpha(t)$ minimizes J in $M(M^*)$, even under the less restrictive conditions of F in §§ 2 and 3.

5. Uniqueness theorems. By the relation $\theta(t) \simeq \Theta(t)$, we mean that equality holds almost everywhere.

THEOREM 5.1. *Under the conditions of* § 4, *if* $T(z)$ *is strictly convex* (*that is,* $T'(z)$ *is strictly increasing*) *on* I_0 *and if* $\theta(t)$ *and* $\Theta(t)$ *both minimize* J *in* $\mathscr{M}$, *then* $\theta(t) \simeq \Theta(t)$.

Proof. The set $\{t : \theta(t) \neq \Theta(t)\}$ is the union over all rationals r,

$$\bigcup \{t : \Theta(t) \leqq r < \theta(t)\} \bigcup \{t : \theta(t) \leqq r < \Theta(t)\}\ .$$

It suffices to prove that each of these sets has measure zero. Suppose there is an r_0 such that $\mu\{t : \Theta(t) \leqq r_0 < \theta(t)\} > 0$. Then there exists z_1 such that, for $r_0 < z < z_1$, $\mu\{t : \Theta(t) \leqq z < \theta(t)\} > 0$. As a consequence of Corollary 4.1, $\mu\{t : \Theta(t) = z\} = 0$ for all but a countable set of z, hence

$$M\{t : \Theta(t) \leqq z < \theta(t)\} = M\{t : \Theta(t) < z < \theta(t)\} < z$$

except for a countable set of z between r_0 and z_1. It follows from Lemma 4.4 that

$$J[\theta; B_1] > J[\Theta; B_1]\ , \qquad B_1 = \{t : \Theta(t) < \theta(t)\}\ .$$

Similarly, if

$$\mu\{t:\ \theta(t) \leqq r_0 < \Theta(t)\} > 0\ ,$$

then

$$J[\theta;\ B_2] > J[\Theta;\ B_2]\ , \qquad B_2 = \{t:\ \Theta(t) > \theta(t)\}\ ,$$

and hence $J[\theta] > J[\Theta]$, contradicting the hypothesis that $J[\theta] = J[\Theta]$.

If $T'(z)$ is not strictly increasing on I the above conclusion need not hold. For example if 0 is interior to I and $T(z)=0$ or z according as $z \leqq 0$ or > 0, then if $\Theta(t)$ minimizes J, any distinct admissible function $\theta(t)$, agreeing everywhere in sign with $\Theta(t)$, also minimizes J.

The next theorem applies either to problems covered by § 3 or to problems based on an integrand (4.1), and to both the minimum problems in M and in M^*.

If $\Theta(t)$ and $\theta(t)$ are both in M or both in M^*, then

$$\theta_z(t) = \Theta(t) + z[\theta(t) - \Theta(t)]$$

is in M (M^*) for $0 \leqq z \leqq 1$. Setting $\mathscr{J}(z) = J[\theta_z]$ we find that

$$\mathscr{J}''(z) = \int (\theta - \Theta)^2 F_{vv}(\alpha,\ \theta_z) d\mu\ , \tag{5.1}$$

provided the formal differentiation is valid. Moreover if Θ minimizes J in M (M^*), if $\mathscr{J}'(0)$ exists, and if Taylor's formula is applicable, then

$$J[\theta] - J[\Theta] = \mathscr{J}(1) - \mathscr{J}(0) \geqq \mathscr{J}(1) - \mathscr{J}(0) - \mathscr{J}'(0) = \mathscr{J}''(\zeta)/2\ , \quad 0 < \zeta < 1\ . \tag{5.2}$$

THEOREM 5.2. *If* (5.1) *and* (5.2) *are valid, if, for each z on the unit interval $F_{vv}[\alpha(t),\ \theta_z(t)]$ is positive for almost all t, and if $\theta(t)$ and $\Theta(t)$ both minimize J in M or both minimize J in M^*, then $\theta(t) \cong \Theta(t)$.*

The last two theorems apply in particular to integrands given by (4.1) and (4.2) in terms of any one of the convex functions

$$T(z) = z \log z + (1-z) \log (1-z), \qquad z \in I_0 = (0,\ 1)\ , \tag{5.3}$$

$$T(z) = z^2\ , \qquad z \in I_0 = (-\infty,\ \infty)\ , \tag{5.4}$$

$$T(z) = z - \log z\ , \qquad z \in I_0 = (0,\ \infty)\ , \tag{5.5}$$

$$T(z) = z \log z\ , \qquad z \in I_0 = (0,\ \infty)\ . \tag{5.6}$$

Applications of these examples in mathematical statistics are discussed in [3]. Each of these examples is covered by the hypotheses of § 3 and of § 4. It is easy to find suitable sufficient conditions for the validity of (5.1) and (5.2) in each case.

References

1. Miriam Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid and Edward Silverman, *An empirical distribution function for sampling with incomplete information*, Ann. Math. Stat. **26** (1955), 641–47.

2. H. D. Brunk, G. M. Ewing and W. T. Reid, *The minimum of a certain definite integral suggested by the maximum likelihood estimate of a distribution function*, Bull. Amer. Math. Soc., **60** (1954), 535.

3. H. D. Brunk, *Maximum likelihood estimates of monotone parameters*, Ann. Math. Stat. **26** (1955), 607–16.

4. H. D. Brunk, G. M. Ewing and W. R. Utz, *Some Helly Theorems for monotone functions*, Proc. Amer. Math. Soc. **7** (1956), 776–83.

5. H. Cramér, *Mathematical methods of statistics*, Princeton University Press, 1946.

6. P. R. Halmos, *Measure theory*, Van Nostrand, New York, 1950.

7. E. J. McShane, *Integration*, Princeton University Press, 1944.

8. M. E. Munroe, *Introduction to measure theory and integration*, Addison-Wesley, Cambridge, Mass., 1953.

UNIVERSITY OF MISSOURI

UNIQUENESS THEORY FOR ASYMPTOTIC EXPANSIONS IN GENERAL REGIONS

PHILIP DAVIS

1. Introduction. Let D be a simply connected region with an analytic boundary C. Assume that $z=0$ is an interior point while $z=1$ lies on the boundary. We assume further that the tangent to C at $z=1$ is not parallel to the real axis. In this case, we shall be able to fit into D small angles Γ placed symmetrically about the real axis and with vertex at $z=1$. These angles will be of the form $-\delta \leqq \theta \leqq \delta$ or $\pi-\delta \leqq \theta \leqq \pi+\delta$, $\delta>0$, depending upon the location of $z=1$. For a given $f(z)$ regular in D, we consider the following limits defined recursively

$$
\begin{aligned}
a_0 &= \lim_{z\to 1} f(z) \\
a_1 &= \lim_{z\to 1} (z-1)^{-1}[f(z)-a_0] \qquad (1)\\
a_2 &= \lim_{z\to 1} (z-1)^{-2}[f(z)-a_0-a_1(z-1)] \\
&\cdots\cdot
\end{aligned}
$$

If each limit in (1) exists and is independent of the manner in which $z \to 1$ through values in some angle Γ, then $f(z)$ is said to possess an asymptotic expansion at $z=1$ in the sense of Poincaré, and this is indicated by writing

$$f(z) \sim \sum_{n=0}^{\infty} a_n(z-1)^n . \qquad (2)$$

We shall designate by $A(=A(D))$ the linear class of functions which are regular in D and which possess asymptotic expansions at $z=1$ in the sense of Poincaré. The angle Γ in which (1) is valid may depend upon the particular $f \in A$ selected.

Uniqueness theory is concerned with distinguishing nontrivial subclasses of A within which the expansion $\sum_{n=0}^{\infty} a_n(z-1)^n$ determines the corresponding function uniquely. Write for the remainder

$$R_n(z)=f(z)-a_0-a_1(z-1)-\cdots-a_{n-1}(z-1)^{n-1}, \qquad (3)$$

and consider the ratios

Received January 6, 1956. The preparation of this paper was sponsored by the Office of Scientific Research and Development of the Air Research and Development Command, USAF.

(4) $$f_n(z)=(z-1)^{-n}R_n(z) \qquad (n=1,\,2,\,\cdots),\ f_0=f.$$

For $f\in A$, the functions $f_n(z)$ are regular in D and are bounded as $z\to 1$ in Γ. For a given sequence of positive quantities $\{m_n\}$, we consider the subset $A(m_n)$ of A consisting of those functions which satisfy in addition

(5) $$\|f_n\|^2 < Mk^n m_n^2 \qquad (n=0,\,1,\,2,\cdots)$$

for some $M>0$, $k>0$. Here $\|\ \|$ designates some conveniently chosen norm. The constants M and k may vary from function to function within the class. With the selection

(6) $$\|f\| = \max_{z\in D} |f(z)|\,,$$

it has been shown by Watson [1] and F. Nevanlinna [5] that when D is a sector, we may produce uniqueness classes by restricting the growth of the sequence $\{m_n\}$ sufficiently. When D is the unit circle, T. Carleman [2] has given necessary and sufficient conditions on $\{m_n\}$ in order that the resulting subclass $A(m_n)$ be a uniqueness class. At the same time Carleman raises the problem of giving necessary and sufficient conditions in the case of a more general region D. This problem (with the norm (6)) has been known in the literature at the generalized problem of Watson. It has been treated by Mandelbrojt and MacLane [3] using the theory of distortion in conformal mapping. See also Meili [4]. In the present paper, we adopt the norm

(7) $$\|f\|^2 = \int_C |f(z)|^2\,ds\,,$$

and show how it is possible to combine Carleman's idea of introducing an appropriate minimum problem with the techniques afforded by the theory of conformal kernel functions to arrive at a solution to this general problem. The class $A(m_n)$ will henceforth refer to the norm (7). Thus the question which we are treating may be worded as follows: *What are necessary and sufficient conditions on the sequence of constants $\{m_n\}$ in order that*

(8) $$\|f_n\|^2 = \int_C |f_n(z)|^2\,ds$$
$$= \int_C \left| \frac{f(z)-a_0-a_1(z-1)-\cdots-a_{n-1}(z-1)^{n-1}}{(z-1)^n} \right|^2 ds < Mk^n m_n^2$$

determine $f(z)$ uniquely from the asymptotic coefficients a_n.

2. Preliminary observations. We must first explain the sense in which we shall understand the expression

$$\int_C |f(z)|^2\, ds$$

when $f(z)$ is regular in D but not necessarily in its closure. Let $w=m(z)$ map D conformally onto the unit circle with $m(0)=0$ and $m(1)=1$. The images of $|w|=r$ will be designated by C_r, $0<r<1$. It is well known that the set of functions

$$\phi_n(z)=\frac{1}{\sqrt{2\pi}}\frac{[m'(z)]^{1/2}}{r^{n+1/2}}[m(z)]^n \qquad (n=0,\,1,\,2,\,\cdots) \tag{9}$$

is complete and orthonormal over each C_r, $0<r<1$, relative to the inner product

$$(f,\,g)=\int_{C_r} f\bar{g}\,ds\,.$$

Suppose then that we are given a function $f(z)$ which is regular in D. Then for any fixed $0<r<1$, $f(z)$ is continuous on C_r. Hence we can write

$$f(z)=\sum_{n=0}^{\infty} a_n\phi_n(z) \tag{10}$$

holding uniformly and absolutely in the interior of C_r. The coefficients a_n are given by

$$a_n=\int_{C_r} f(z)\,\overline{\phi_n(z)}ds \qquad (n=0,\,1,\,\cdots)\,. \tag{11}$$

Hence, for $r^*<r$, we have from (9) and (10),

$$\int_{C_{r^*}} |f(z)|^2\, ds=\sum_{n=0}^{\infty}|a_n|^2\frac{r^{*2n+1}}{r^{2n+1}}\,. \tag{12}$$

This equation tells us that

$$\int_{C_{r^*}} |f(z)|^2\, ds$$

is an increasing function of r^* and hence

$$\lim_{r^*\to 1^-}\int_{C_{r^*}} |f(z)|^2\, ds$$

exists (or equals $+\infty$). For $f(z)$ regular in D we shall therefore agree that

$$\int_C |f(z)|^2\, ds = \lim_{r\to 1-} \int_{C_r} |f(z)|^2\, ds\ .$$

LEMMA. *Given an arbitrary sequence of positive constants $\{m_n\}$; the class $A(m_n)$ is not a uniqueness class for asymptotic expansions at $z=1$ if and only if there exists an $f\not\equiv 0$ regular in D and constants $M>0$, $k>0$, for which*

$$\left\| \frac{f(z)}{(z-1)^n} \right\|^2 < Mk^n m_n^2 \qquad (n=0,\ 1,\ 2,\ \cdots). \tag{13}$$

Proof. If $A(m_n)$ is not a uniqueness class, there will exist two functions $g(z)$, $y(z) \in A(m_n)$, $g\not\equiv h$, possessing the same asymptotic expansion, say $\sum_{n=0}^{\infty} a_n(z-1)^n$, and satisfying

$$\int_C \left| \frac{g(z) - \sum_{k=0}^{n-1} a_k(z-1)^k}{(z-1)^n} \right|^2 ds < M_1 k_1^n m_n^2 \qquad (n=0,\ 1,\ \cdots) \tag{14}$$

$$\int_C \left| \frac{h(z) - \sum_{k=0}^{n-1} a_k(z-1)^k}{(z-1)^n} \right|^2 ds < M_2 k_2^n m_n^2$$

with $k_1 \leqq k_2$. Therefore, by Minkowski's inequality,

$$\begin{aligned} \int_C \left| \frac{g(z)-h(z)}{(z-1)^n} \right|^2 ds &< (M_1^{1/2} k_1^{n/2} + M_2^{1/2} k_2^{n/2})^2 m_n^2 \\ &= (M_1^{1/2}(k_1/k_2)^{n/2} + M_2^{1/2})^2 k_2^n m_n^2 \\ &< (M_1^{1/2} + M_2^{1/2})^2 k_2^n m_n^2 \end{aligned} \tag{15}$$

so that $g-h$ does not vanish identically and satisfies (13) with $M=(M_1^{1/2} + M_2^{1/2})^2$ and $k=k_2$.

Conversely, let $f\not\equiv 0$ satisfy (13). We shall show that (13) implies

$$\lim_{z\to 1} \frac{f(z)}{(z-1)^n} = 0 \qquad (n=0,\ 1,\ 2,\ \cdots) \tag{16}$$

as $z\to 1$ through values in some angle Γ. Assuming, for the moment, that this is so, (16) and (1) imply that

$$f(z) \sim 0 + 0\cdot(z-1) + 0\cdot(z-1)^2 + \cdots\ . \tag{17}$$

That is, $f(z)$ possesses an identically zero asymptotic expansion at $z=1$. Furthermore $f_n = f(z)(z-1)^{-n}$, so that (13) implies that $f\in A(m_n)$. Thus, $A(m_n)$ is not a uniqueness class for asymptotic expansions at $z=1$.

We show now that (13) implies (16). Given any $g(z)$ regular in D. Select any $0<r<1$. We have from (9), (10), (11), and the Schwarz inequality

$$|g(z)|^2 < K_{C_r}(z, \bar{z}) \int_{C_r} |g(z)|^2 \, ds, \tag{18}$$

for all z interior to C_r. K_{C_r} is the so-called Szegö kernel function for C_r whose explicit expression is (Szegö [6], Bergman [1])

$$K_{C_r}(z, \bar{z}) = \sum_{n=0}^{\infty} \phi_n(z)\overline{\phi_n(z)} = \frac{1}{2\pi} \frac{r|m'(z)|}{r^2 - |m(z)|^2}. \tag{19}$$

Writing $f(z)/(z-1)^n$ in place of $g(z)$ in (18), and using (13) and the monotonicity with r of

$$\int_{C_r} |f(z)|^2 \, ds,$$

we find for $j \leqq n$ and z interior to C_r,

$$\left|\frac{f(z)}{(z-1)^j}\right|^2 \leqq \frac{|(z-1)^{n-j}|^2 r|m'(z)|}{(2\pi)(r^2 - |m(z)|^2)} Mk^n m_n^2 \qquad (n=0, 1, 2, \cdots). \tag{20}$$

For each z in D we select an $r = r(z) = |m(z)| + \varepsilon(z) < 1$ where $\varepsilon(z)$ is defined by

$$\varepsilon(z) = \frac{1}{2}(1 - |m(z)|). \tag{21}$$

Thus,

$$\lim_{z \to 1} \varepsilon(z) = 0. \tag{22}$$

Here, $z \to 1$ through values in D. From (20), (21), and $r<1$,

$$\left|\frac{f(z)}{(z-1)^j}\right|^2 \leqq \frac{|(z-1)^{n-j}|^2}{2\pi} \cdot \frac{|m'(z)| Mk^n m_n^2}{2|m(z)|\varepsilon(z) + \varepsilon^2(z)} \tag{23}$$

$$< \frac{|(z-1)^{a-j}|^2 |m'(z)| Mk^n m_n^2}{4\pi |m(z)| \varepsilon(z)}.$$

We are now ready to consider the limit of (23) as $z \to 1$. First consider

$$\frac{\varepsilon(z)}{|z-1|} = \frac{1 - |m(z)|}{2|z-1|} = \frac{1}{2}(1 + |m(z)|)^{-1} \frac{(1 - |m(z)|^2)}{|z-1|}. \tag{24}$$

Since $m(z)$ is by assumption analytic at $z=1$, we have in a neighborhood of $z=1$,

(25)
$$m(z)=1+(z-1)R(z)\,,$$

where $R(z)$ is analytic there. Note that $R(1)=m'(1)\neq 0$, and write $R(z)=\sigma(z)e^{i\alpha(z)}$, $\sigma(z)>0$. We have $\sigma(1)\neq 0$ and $\alpha(1)\neq\pi/2,\ 3\pi/2$, inasmuch as the tangent to C at $z=1$ is assumed not parallel to the real axis. Furthermore, write $z=1+\rho e^{i\theta}$. Then, from (25),

(26)
$$\begin{aligned}\frac{1-|m(z)|^2}{|z-1|}&=\frac{-2\mathscr{R}\{(z-1)R(z)\}}{|z-1|}-\frac{|z-1|^2|R(z)|^2}{|z-1|}\\ &=-2\mathscr{R}\{e^{i\theta}\sigma(z)^{i\alpha(z)}\}-|z-1||R(z)|^2\\ &=-2\sigma(z)\cos(\theta+\alpha(z))-|z-1||R(z)|^2\,.\end{aligned}$$

If $z\to 1$ through some angle Γ: $-\delta\leqq\theta\leqq\delta$ or $\pi-\delta\leqq\theta\leqq\pi+\delta$, then, since $\alpha(1)\neq\pi/2,\ 3\pi/2$, it follows from the above that for δ sufficiently small, the expression (26) will be bounded away from 0. In view of (24) we will have

(27)
$$\frac{\varepsilon(z)}{|z-1|}\geqq\tau>0\ ;\ z\to 1$$

for z in some Γ. From (23), we have,

(28)
$$\left|\frac{f(z)}{(z-1)^j}\right|^2<|z-1|^{2n-2j-1}|m'(z)|Mk^n m_n^2\Big/\frac{4\pi|m(z)|\cdot\varepsilon(z)}{|z-1|}\,.$$

Thus, for $2n-2j-1>1$ it is now clear from (28) and (27) that

$$\lim_{z\to 1}\frac{f(z)}{(z-1)^j}=0\,.$$

For each j considered we need only use an $n>j+1$. This completes the proof of the lemma.

3. The uniqueness theorem.

THEOREM. *Given an arbitrary sequence of positive constants m_n. The class $A(m_n)$ is a uniqueness class for asymptotic expansions at $z=1$ if and only if for all $t>0$,*

(20)
$$\limsup_{n\to\infty}\int_C\log\left\{\sum_{k=0}^{n}\frac{t^k}{m_k^2}|(z-1)^{n-k}|^2\right\}\frac{\partial}{\partial n}\log|m(z)|ds=\infty\,.$$

Here $\partial/\partial n$ designates normal differentiation in the positive sense.

The above statement is equivalent to saying that $A(m_n)$ is not a uniqueness class if and only if there exists a $t>0$ and a $K>0$ such that

$$\int_C \log\left\{\sum_{k=0}^{n} \frac{t^k}{m_k^2}|(z-1)^{n-k/2}\right\}\frac{\partial}{\partial n}\log|m(z)|\,ds < K, \qquad n=0,1,2,\cdots \tag{30}$$

K may depend upon t, but is independent of n.

In view of the lemma of the preceding section, we shall prove that (30) is a necessary and sufficient condition for the existence of an $f(z) \neq 0$, and M, and a k which satisfy (13).

Consider the following sequence of integrals

$$I_n(f)=\sum_{k=0}^{n}\frac{t^k}{m_k^2}\int_C\left|\frac{f(z)}{(z-1)^k}\right|^2 ds; \tag{31}$$

$$=\sum_{k=0}^{n}\frac{t^k}{m_k^2}\|f\|_k^2 \qquad n=0,1,\cdots,$$

where we have written

$$\|f\|_k^2=\int_C\left|\frac{f(z)}{(z-1)^k}\right|^2 ds; \qquad k=0,1,\cdots. \tag{32}$$

We can also write (31) in the form

$$I_n(f)=\left\|\frac{\rho_n(z)f(z)}{(z-1)^n}\right\|^2 \tag{33}$$

where $\rho_n(z)$ is an analytic function which is regular in D, continuous on C and is such that

$$|\rho_n(z)|=\left\{\sum_{k=0}^{n}\frac{t^k}{m_k^2}|(z-1)^{n-k}|^2\right\}^{1/2}, \quad \text{for } z \text{ on } C. \tag{34}$$

We shall show below how a $\rho_n(z)$ may be constructed which has these properties and has, in addition, the property that

$$\rho_n(z)\neq 0 \quad \text{for} \quad z \text{ in } D. \tag{35}$$

Let n be fixed, and consider the following minimum problem P_n. *Determine that function $f(z)$ regular in D with $f(0)=1$ and such that*

$$I_n(f)=minimum. \tag{36}$$

This problem can be solved by passing to a related problem P_n'. *Determine that function $g(z)$ regular in D with $g(0)=1$ and such that*

$$\|g\|^2=minimum \tag{37}$$

The solution of the problem P_n' is given by the function (see, for example Szegö [6], Bergman [1])

$$g^*(z)=K_D(z, 0)/K_D(0, 0) \tag{38}$$

where $K_D(z, \overline{w})$ is the Szegö kernel function of the region D. The minimum value of the integral (37) is $1/K_D(0, 0)$. If we write

$$I_n(f)=|\rho_n(0)|^2\left\|\frac{\rho_n(z)f(z)}{\rho_n(0)(1-z)^n}\right\|^2, \tag{39}$$

we see, in view of (35) that the function $\rho_n(z)f(z)/\rho_n(0)(1-z)^n$ can play the role of $g(z)$ in the problem P_n'. The minimizing function f_n^* of the problem P_n is therefore

$$f_n^*(z)=\frac{K_D(z, 0)(1-z)^n\rho_n(0)}{\rho_n(z)K_D(0, 0)}, \tag{40}$$

and the minimum value of the integral is

$$I_n(f_n^*)=\frac{|\rho_n(0)|^2}{K_D(0, 0)}. \tag{41}$$

We now assert: a necessary and sufficient condition in order that there exist an $f(z)\neq 0$ and constants $M>0$, $k>0$ such that

$$\|f\|_n^2=\left\|\frac{f(z)}{(z-1)^n}\right\|^2<Mk^nm_n^2 \qquad (n=0, 1, \cdots) \tag{42}$$

is that there exists a $t>0$ and a $K>0$ such that

$$I_n(f_n^*)\leqq K \qquad n=0, 1, 2, \cdots. \tag{43}$$

Referring to (41), this is equivalent to asserting that there exist a $t>0$ and a K' such that

$$|\rho_n(0)|\leqq K' \qquad n=0, 1, 2, \cdots. \tag{44}$$

We can prove this as follows. Suppose first that $q(z)$ is such that (42) holds for it. This function $q(z)$ may have a zero of the pth order at $z=0$. The function $f(z)=q(z)/z^p$ is then regular in D and is such that $f(0)\neq 0$. Now,

$$\begin{aligned} I_n(f(z)/f(0)) &= \sum_{j=0}^n \frac{t^j}{m_j^2}\int_C\left|\frac{q(z)}{f(0)z^p(z-1)^j}\right|^2 ds \\ &\leqq \sum_{j=0}^n \frac{t^j}{m_j^2}\frac{1}{|f(0)|^2}\frac{1}{d^{2p}}M\cdot m_j^2k^j \\ &\leqq \frac{M}{d^{2p}|f(0)|^2}\sum_{j=0}^n t^jk^j \leqq \frac{M}{d^{2p}|f(0)|^2(1-tk)}, \end{aligned} \tag{45}$$

provided we select $0<t<1/k$. Here d designates the minimum distance from $z=0$ to C. Now since

$$I_n(f_n^*)\leqq I_n(f(z)/f(0))\leqq \frac{M}{d^{2p}|f(0)|^2(1-tk)}, \qquad (n=0,1,\cdots) \tag{46}$$

then (43) is satisfied with K equal to the right hand constant in (46).

Conversely, suppose that there exists a $t>0$ and $K>0$ such that (43) holds. Then from (31),

$$\sum_{k=0}^{n}\frac{t^k}{m_k^2}\|f_n^*\|_k^2\leqq K \qquad (n=0,1,2,\cdots). \tag{47}$$

In particular, taking the first term of (47) we obtain

$$\frac{1}{m_0^2}\|f_n^*\|_0^2<K \qquad n=0,1,2,\cdots. \tag{48}$$

Hence we have

$$\|f_n^*\|<\text{const.} \qquad (n=0,1,2,\cdots). \tag{49}$$

The inequalities (49) imply that the sequence of minimizing functions $\{f_n^*\}$ form a normal family and therefore there exist indices $n_1, n_2, \cdots$ such that $f_{n_k}^*\to F(z)$ uniformly in any closed region interior to D. Again, using (47) we have, for fixed j and for all $n\geqq j$

$$\frac{t^j}{m_j^2}\|f_n^*\|_j^2\leqq K. \tag{50}$$

Now for any $0<\rho<1$, we have

$$\|f_n^*\|_j^2=\int_C\left|\frac{f_n^*(z)}{(z-1)^j}\right|^2 ds\geqq\int_{C_\rho}\left|\frac{f_n^*(z)}{(z-1)^j}\right|^2 ds, \tag{51}$$

so that from (50) and (51),

$$\int_{C_\rho}\left|\frac{f_n^*(z)}{(z-1)^j}\right|^2 ds<Km_j^2t^{-j} \qquad (k=0,1,2,\cdots). \tag{52}$$

Let n take on the values n_i in (52) and let j be fixed. Then since $f_n^*(z)\to F(z)$ uniformly in and on C_ρ,

$$\int_{C_\rho}\left|\frac{F(z)}{(z-1)^j}\right|^2 ds\leqq Km_j^2t^{-j}. \tag{53}$$

This result is independent of ρ and hence we may allow $\rho\to 1$. Thus,

$$\int_C \left| \frac{F(z)}{(z-1)^j} \right|^2 ds < K m_j^2 t^{-j} \qquad (j=0, 1, 2, \cdots) . \tag{54}$$

Since obviously $F(0)=1$, we have exhibited in $F(z)$ a function regular in D, which does not vanish identically, a constant $M(=K)$ and a constant $k(=t^{-1})$ for which (42) holds.

It remains to construct $\rho_n(z)$, to show that it does not vanish, and to compute $\rho_n(0)$. Designate by $t_n(z)$ the positive function

$$t_n(z) = \left\{ \sum_{k=0}^{n} \frac{t^k}{m_k^2} |(z-1)^{n-k}|^2 \right\}^{1/2} \tag{55}$$

defined on C. Now $\log t_n(z)$ is continuous on C and hence

$$u_n(z) = \frac{1}{2\pi} \int_C \log t_n(w) \frac{\partial g(z, w)}{\partial n} ds \tag{56}$$

where $g(z, w)$ is the Green's function for D, is harmonic in D and assumes on C the boundary values $\log t_n(z)$. Designate by v_n the harmonic conjugate of u_n. Then $u_n(z) + iv_n(z)$ is regular and single valued in D, as is

$$p_n(z) = \exp\,[u_n(z) + iv_n(z)] . \tag{57}$$

Now, $|p_n(z)| = e^{u_n}$, so that on C, $|p_n(z)| = t_n(z)$. Furthermore $p_n(z) \neq 0$, as is clear from (57). Thus we may use $\rho_n(z) = p_n(z)$. The condition (44) then becomes: there exists a $t > 0$ and a $K' > 0$ such that

$$u_n(0) \leqq K' \qquad (n \to \infty) . \tag{58}$$

Finally, using the representation

$$g(z, w) = \log \left| \frac{m(z) - m(w)}{1 - m(z)\overline{m(w)}} \right| \tag{59}$$

with $z=0$ in (56), we obtain the stated condition (29).

4. Concluding remarks. Norms other than (6) might be contemplated. In particular, we might have used

$$\|f\|^2 = \iint_D |f(z)|^2 \, dA . \tag{60}$$

However (60) has the disadvantage that the solution of the corresponding minimum problem P_n can not be so elegantly expressed in terms of an analytic function $\rho_n(z)$ and so the role of the sequence $\{m_n\}$ is not immediately evident as with (29).

References

1. S. Bergman, *The kernel function and conformal mapping*, New York, 1950.
2. T. Carleman, *Les fonctions quasi-analytiques*, Paris, 1926.
3. S. Mandelbrojt and G. R. MacLane, *On functions holomorphic in a strip region and an extension of Watson's problem*, Trans. Amer. Math. Soc., **61** (1947), 454.
4. H. J. Meili, *Über das Eindeutigkeitsproblem in der Theorie der asymptotischen Reihen*, Comm. Math. Helv., **29** (1954), 93-96.
5. F. Nevanlinna, *Zur theorie der asymptotischen Potenzreihen*, Ann. Acad. Sci. Fenn., Sec. A, 1918.
6. G. Szegö, *Orthogonal polynomials*, New York, 1935.
7. G. N. Watson, *A theory of asymptotic series*, Trans. Royal Soc. Lond., **211**, 279-313.

NATIONAL BUREAU OF STANDARDS
WASHINGTON, D. C.

ON THE LEAST PRIMITIVE ROOT OF A PRIME

PAUL ERDÖS AND HAROLD N. SHAPIRO

1. Introduction. The problem of estimating the least positive primitive root $g(p)$ of a prime p seems to have been first considered by Vinogradov. His first result was [4, v. 2 part 7 chap. 14]

$$g(p) \leqq 2^m p^{1/2} \log p \,, \tag{1.1}$$

where m denotes the number of distinct prime factors of p-1. In 1930, [6], he improved this to

$$g(p) \leqq 2^m \frac{p-1}{\phi(p-1)} p^{1/2} \tag{1.2}$$

where $\phi(n)$ is the Euler ϕ-function. Next, in 1942, Hua [3] improved this to

$$g(p) < 2^{m+1} p^{1/2} \,, \tag{1.3}$$

and obtained also, for the primitive root of least absolute value, $h(p)$,

$$|h(p)| < 2^m p^{1/2} \,. \tag{1.4}$$

Lastly, Erdös [2] proved that for p sufficiently large

$$g(p) < p^{1/2}(\log p)^{17} \,. \tag{1.5}$$

This last result, of course, is not directly comparable with the others, giving better results for some primes and worse results for others.

In any event, all of the results are very weak (as is evidenced by a glance at tables of primitive roots [1]) in relationship to the conjecture that the true order of $g(p)$ is about $\log p$. In this connection, Pillai [5] has proved

$$g(p) > \log \log p \tag{1.6}$$

for infinitely many p.

In this note we shall give a very simple way of handling character sums, which not only yields (1.3) and (1.4) but allows a small improvement of these results; for example

$$g(p) = O(m^c p^{1/2}) \,, \quad (c \text{ a constant}) \,. \tag{1.7}$$

Received March 12, 1956.

2. A lemma concerning character sums. We consider first an inequality for certain character sums on which our later estimates will depend. Let S and T be any two sets of integers, such that modulo a given prime p, no two integers of S are congruent, and no two integers of T are congruent. Denote by $N(S)$, $N(T)$ the number of integers in S and T respectively. We have

LEMMA. For χ a non-principal character modulo p,

$$\left|\sum_{\substack{u\in S\\ v\in T}} \chi(u+v)\right| \leq p^{1/2}\sqrt{N(S)N(T)}\,. \tag{2.1}$$

Proof. Set

$$\tau(\chi)=\sum_{h=1}^{p}\chi(h)e^{2\pi ih/p}.$$

It is well known that $|\tau(\chi)|=p^{1/2}$, for χ a non-principal character. Also,

$$\tau(\bar{\chi})\chi(t)=\sum_{h=1}^{p}\bar{\chi}(h)e^{2\pi iht/p}\,.$$

From this we get

$$\tau(\bar{\chi})\sum_{\substack{u\in S\\ v\in T}}\chi(u+v)=\sum_{\substack{u\in S\\ v\in T}}\sum_{h=1}^{p}\bar{\chi}(h)e^{2\pi ih/p\cdot(u+v)}.$$

Then taking absolute values and using Schwarz's inequality

$$p^{1/2}\left|\sum_{\substack{u\in S\\ v\in T}}\chi(u+v)\right|\leq\sum_{h=1}^{p}\left|\sum_{u\in S}e^{2\pi ihu/p}\right|\left|\sum_{v\in T}e^{2\pi ihv/p}\right|$$

$$\leq\left\{\sum_{h=1}^{p}\left|\sum_{u\in S}e^{2\pi ihu/p}\right|^2\sum_{h=1}^{p}\left|\sum_{v\in T}e^{2\pi ihv/p}\right|^2\right\}^{1/2}.$$

But

$$\sum_{h=1}^{p}\left|\sum_{u\in S}e^{2\pi ihu/p}\right|^2=\sum_{h=1}^{p}\sum_{\substack{u_1\in S\\ u_2\in S}}e^{2\pi ih/p\cdot(u_1-u_2)}$$

$$=\sum_{\substack{u_1\in S\\ u_2\in S}}\sum_{h=1}^{p}e^{2\pi ih/p\cdot(u_1-u_2)}=pN(S)\,.$$

Similarly

$$\sum_{h=1}^{p}\left|\sum_{v\in T}e^{2\pi ihv/p}\right|^2=pN(T),$$

and the lemma follows immediately.

3. Another proof of Hua's result. By way of illustrating the manner in which the above lemma is to be applied we give here another proof of (1.3). It is well known that if t is not a primitive root modulo p then

$$P(t)=\sum_{d\mid p-1}\frac{\mu(d)}{\phi(d)}\sum_{o(\chi)=d}\chi(t)=0\,,$$

where $o(\chi)=d$ denotes that the inner summation is taken over all characters of order d.

Now if $x+1=g(p)$, the smallest positive primitive root mod p, we see that $P(t)=0$, $1\leqq t\leqq x$. Thus let $S=T$ denote the set of integers $1, 2, \cdots, [x/2]$; we have

$$\begin{aligned}0&=\sum_{d\mid p-1}\frac{\mu(d)}{\phi(d)}\sum_{o(\chi)=d}\sum_{\substack{u\in S\\ v\in T}}\chi(u+v)\\ &=[x/2]^2+\sum_{\substack{d\mid p-1\\ d>1}}\frac{\mu(d)}{\phi(d)}\sum_{o(\chi)=d}\sum_{\substack{u\in S\\ v\in T}}\chi(u+v)\,.\end{aligned}$$

Applying the lemma, this gives

$$(2^m-1)p^{1/2}[x/2]\geqq[x/2]^2$$

or

$$[x/2]\leqq(2^m-1)p^{1/2}\,.$$

Since $2[x/2]+2\geqq x+1=g(p)$ this yields

$$g(p)\leqq 2^{m+1}p^{1/2}-2p^{1/2}+2<2^{m+1}p^{1/2}$$

which is Hua's result (1.3).

Similarly, if in the above argument we use for $S=T$ the set of nonzero integers $-[x/2], \cdots, [x/2]$ where $x+1=|h(p)|$, we are led immediately to the result (1.4).

4. A small improvement in the estimate. The facility with which the lemma of § 2 enables us to handle the relevant character sums makes possible an improvement of the estimates for $g(p)$ and $h(p)$. We consider only the case of $g(p)$, since a similar estimate for $h(p)$ then follows automatically.

Let $F_x(d)$ denote the number of integers among

$$u+v,\ 1\leqq u\leqq[x/2],\ 1\leqq v\leqq[x/2]$$

such that $u+v$ is a dth power residue modulo p. Then, letting S denote the set of integers 1, 2, $\cdots$, $[x/2]$, we have

$$F_x(d)=\frac{1}{d}\sum_{\substack{u\in S\\ v\in S}}\sum_{o(\chi)\mid d}\chi(u+v)$$

$$=\frac{1}{d}[x/2]^2+\frac{1}{d}\sum_{\substack{o(\chi)\mid d\\ o(\chi)>1}}\sum_{\substack{u\in S\\ v\in S}}\chi(u+v)\,.$$

Applying the lemma of § 2 we obtain

$$F_x(d)=\frac{x^2}{4d}+O(xp^{1/2})\,. \tag{4.1}$$

If we let $N(x)$ denote the numbers among the

$$u+v,\quad u\in S,\quad v\in S$$

which are primitive roots modulo p, it is easily seen that

$$N(x)=\sum_{d\mid p-1}\mu(d)F_x(d)\,. \tag{4.2}$$

Applying Brun's method to (4.2), in conjunction with (4.1), in order to make a lower estimate for $N(x)$, one obtains

$$N(x)>\frac{x^2}{4}\sum_{d\mid p-1}\frac{\mu(d)}{d}+O(m^c p^{1/2}x)$$

or

$$N(x)>\frac{\phi(p-1)}{p-1}\,\frac{x^2}{4}+O(m^c p^{1/2}x)\,. \tag{4.3}$$

Thus if we take $x+1=g(p)$, $N(x)=0$ and (4.3) yields

$$x=O\left(\frac{p-1}{\phi(p-1)}\,m^c p^{1/2}\right)\,. \tag{4.4}$$

Finally since

$$\frac{p-1}{\phi(p-1)}=\prod_{q\mid p-1}\frac{1}{1-1/q}<\prod_{i=1}^{m}\frac{1}{1-1/p_i}=O(\log m)=O(m^\varepsilon)$$

(where p_i denotes the ith prime), (4.4) gives

$$x=O(m^c p^{1/2})\,,$$

and hence

$$g(p)=O(m^c p^{1/2})\,,$$

which is the desired result.

References

1. Allan Cunningham, H. J. Woodall and T. G. Creak, *On least primitive roots*, Proc. London Math. Soc. 2nd series, **21** (1922-23), 343-358.
2. P. Erdös, *Least primitive root of a primc*, Bull. Amer. Math. Soc., **55** (1945), 131-132.
3. L. K. Hua, *On the least primitive root of prime*, Bull. Amer. Math. Soc., **48** (1942), 726-730.
4. E. Landau, *Vorlesungen über Zahlentheorie*, Leipzig 1927, (New York 1947).
5. S. Pillai, *On the smallest primitive root of a prime*, J. Indian Math. Soc., **8** (1944), 14-17.
6. I. M. Vinogradov, *On the least primitive root of a prime*, Dokl. Akad. Nauk, S.S.S.R. (1930), 7-11.

Notre Dame University
New York University

REGULAR REGIONS FOR THE HEAT EQUATION

W. FULKS

1. Introduction. Let R be a region (open connected set) in the plane or in space ($x=[x_1, x_2]$ or $x=[x_1, x_2, x_3]$). We will say that R is a regular region for Laplace's equation

$$(1) \qquad \Delta u=0$$

if the Dirichlet problem for R always has a solution for continuous data. By this we mean: given a function $\psi(\xi)\in C$ (that is, continuous) for $\xi\in B$, the boundary of R, there is a unique function $u(x)\in C$ for $x\in\overline{R}=R\cup B$, for which

$$\Delta u=0 \qquad x\in R\,,$$

$$u(\xi)=\psi(\xi) \qquad \xi\in B\,.$$

We will further say that R is regular for the heat equation

$$(2) \qquad \Delta u=u_t$$

if the "Dirichlet problem" for the heat equation has a solution for continuous data, that is, if for each

$$\phi(x)\in C \qquad x\in\overline{R}$$

and

$$\psi(\xi,\ t)\in C \qquad \xi\in B,\ t\geqq 0$$

where

$$\phi(\xi)=\psi(\xi,\ 0)$$

there is a unique function $u(x, t)\in C$, for $x\in\overline{R}$, $t\geqq 0$ for which

$$\Delta u=u_t \qquad x\in R,\ t>0$$

$$u(x,\ 0)=\phi(x) \qquad x\in\overline{R}$$

$$u(\xi,\ t)=\psi(\xi,\ t) \qquad \xi\in B,\ t\geqq 0\,.$$

Tychonoff [4] has shown that if R is bounded and regular for

Received September 15, 1955. In revised form February 24, 1956. The work on this paper was performed under sponsorship of the Office of Naval Research, Contract Nonr-710 (16); (NR 043 041).

Laplace's equation, then it is regular for the heat equation and conversely. We give here a new proof that regularity for Laplace's equation implies regularity for the heat equation.

2. The work of Tychonoff. In the first half of the memoir cited above, Tychonoff proves the following three theorems.

A. *Each bounded region which is a regular region for the heat equation is also regular for Laplace's equation.*

B. *Each bounded regular region for the equation* $\Delta u=\bar{\lambda}u$ *for a certain* $\bar{\lambda}\geqq 0$ *is also regular for the equation* $\Delta u=\lambda u$ *for arbitrary* $\lambda\geqq 0$.

C. *Each bounded region which is regular for all the equations* $\Delta u=\lambda u$ *for* $\lambda\geqq\lambda_0$ *is also regular for the heat equation.*

This cycle of theorems shows the equivalence of regular regions for the equations $\Delta u=0$, $\Delta u=\lambda u$ $(\lambda\geqq 0)$, and $\Delta u=u_t$.

In the proof of B Tychonoff observes that the solution of the boundary value problems

$$\begin{aligned} \Delta u-\lambda u&=0 \qquad x\in R\\ u(\xi)&=\psi(\xi) \qquad \xi\in B \end{aligned}$$

is equivalent to the solution of the integral equations

$$u=(\lambda-\bar{\lambda})\int_R G(x,\ y)u(y)dy+w(x)$$

where G is Green's function for the region R for the equation $\Delta u=\bar{\lambda}u$, and $w(x)$ is the solution to the problem

$$\begin{aligned} \Delta w-\bar{\lambda}w&=0 \qquad x\in R\\ w(\xi)&=\psi(\xi) \qquad \xi\in B\ . \end{aligned}$$

The existence of both w and G are guaranteed by the assumption that R is regular for $\Delta u=\bar{\lambda}u$. He then deduces, via the Hilbert-Schmidt theory, that the desired solutions of the integral equations exist and hence these solve the boundary value problems.

However, in establishing C, he forsakes his integral equation methods and bases his argument on a refinement of a differential-difference method due to Rothe [2].

We may note that to complete the cycle of theorems it is sufficient to prove that if R is regular for $\Delta u=0$ it is regular for $\Delta u=u_t$, and we give here a proof of this result using a modification of the integral equation argument mentioned above.

In our argument we will use the following theorem which was indicated in a footnote in the paper by Tychonoff. For the sake of completeness we present the proof.

D. *Let R be a regular region for $\Delta u=0$, and let $\psi(\xi, t)$ be defined on B and be k times differentiable with respect to t, $0 \leqq t < T \leqq \infty$, and let ψ and each of its k derivatives respect to t be continuous for $\xi \in B$, $0 \leqq t < T$. Further, let $u(x, t)$ be the solution to the problem*

$$\Delta u(x, t)=0 \qquad x \in R$$

$$u(\xi, t)=\psi(\xi, t), \qquad \xi \in B, \ 0 \leqq t < T .$$

Then $u(x, t)$ has k continuous derivatives with respect to t and

$$v=\frac{\partial^j u}{\partial t^j}, \quad 0 \leqq j \leqq k ,$$

solves the problem

$$\Delta v(x, t)=0 \qquad x \in R$$

$$v(\xi, t)=\frac{\partial^j}{\partial t^j}\psi(\xi, t), \quad \xi \in B, \ 0 \leqq t < T .$$

Proof. Choose t_0, $0 \leqq t_0 < T$. By the maximum and minimum principles for harmonic functions

$$|u(x, t)-u(x, t_0)| \leqq \max_{\xi \in B} |\psi(\xi, t)-\psi(\xi, t_0)| .$$

But by the uniform continuity of $\psi(\xi, t)$ for $\xi \in B$, and t in a (sufficiently small) closed t interval about t_0, this maximum tends to zero as t tends toward t_0. So that $u(x, t)$ is continuous in t.

Since R is a regular region for $\Delta u=0$ there is a solution to the problem

$$\Delta v(x, t)=0 \qquad x \in R$$

$$v(\xi, t)=\frac{\partial}{\partial t}\psi(\xi, t), \quad \xi \in B, \ 0 \leqq t T .$$

Then

$$\left|\frac{u(x, t)-u(x, t_0)}{t-t_0}-v(x, t)\right| \leqq \max_{\xi \in R}\left|\frac{\psi(\xi, t)-\psi(\xi, t_0)}{t-t_0}-\frac{\partial}{\partial t}\psi(\xi, t_0)\right|$$

by the same argument used above. But

$$\frac{\psi(\xi, t)-\psi(\xi, t_0)}{t-t_0}=\frac{\partial \psi}{\partial t}(\xi, \bar{t}(\xi)) ,$$

where $\bar{t}(\xi)$ lies between t and t_0. Again by the uniform continuity of

$\frac{\partial\psi}{\partial t}(\xi,,t)$ this maximum vanishes as t tends toward t_0. Hence $u(x, t)$ is differentiable with respect to t and this derivative attains the continuous boundary data $\frac{\partial\psi}{\partial t}(\xi, t)$. Hence by the first part of the proof $\frac{\partial u}{\partial t}(x, t)$ is continuous in t. By iterating this argument k times the proof is completed.

We will need the following, also taken from Tychonoff.

E. *Let R be bounded and regular for $\Delta u=0$, and let $G(x, y)$ be the Green's function for this equation and this region:*

$$G(x, y)=\begin{cases} \frac{1}{2\pi}\log\frac{1}{r_{xy}}-g(x, y) & n=2 \\ \frac{1}{4\pi}\cdot\frac{1}{r_{xy}}-g(x, y) & n=3 \end{cases}$$

where $g(x, y)$ is the solution to the problem

$$\Delta_x g(x, y)=0 \qquad x\in R,\ y\in R$$

$$g(\xi, y)=\begin{cases} \frac{1}{2\pi}\log\frac{1}{r_{\xi y}} & \xi\in B,\ y\in R,\ n=2 \\ \frac{1}{4\pi}\cdot\frac{1}{r_{\xi y}} & \xi\in B,\ y\in R,\ n=3\,. \end{cases}$$

Then $G(x, y)=G(y, x)$, $x\in R$, $y\in R$.

Proof. Let R_j be a sequence of regions, $\overline{R}_j\subset R_{j+1}\subset R$, which tend to R with the property that the corresponding boundaries B_j are surfaces having continuous curvature and such that the distance from each point on B_j to B is not greater than δ_j where the sequence $\delta_j\to 0$ as $j\to\infty$. For such a construction see Kellog [1].

Let $G_j(x, y)$ be the Green's function for R_j. Under the hypotheses on R_j it is well known that $G_j(x, y)$ is symmetric (see Tamarkin and Feller [3]). It is therefore sufficient to prove that

$$\lim_{j\to\infty} G_j(x, y)=G(x, y)\,.$$

To this end we note that $G\geqq 0$: since it vanishes on B and is large and positive near the pole y it must be nonnegative by the minimum principle.

Let $\varepsilon>0$ be given, then if j is sufficiently large we have $0\leqq G(x, y)\leqq\varepsilon$ for each point $x\in\overline{R}-R_j$, and in particular on B_j. Hence

$$0 \leqq G(x, y) - G_j(x, y) = g(x, y) - g_j(x, y) < \varepsilon$$

everywhere in R_j since that inequality is true on B_j. This completes the argument.

3. Reduction of the data. We return now to the problem

$$\Delta u = u_t \qquad x \in R,\ t > 0$$

$$u(x, 0) = \phi(x) \qquad x \in \overline{R}$$

$$u(\xi, t) = \psi(\xi, t) \qquad \xi \in B,\ t > 0$$

under the assumption that R is regular for $\Delta u = 0$. We show that $\phi(x)$ may be assumed to be zero. Let R' be a sphere (or circle) containing $\overline{R}$ in its interior, and let $\phi'(x)$ be a continuous bounded extension of $\phi(x)$ into R'. Define

$$u_1(x, t) = \int_{R'} k(x-y, t)\phi'(y)dy ,$$

dy being the element of area or volume, and $k(x, t)$ being the fundamental solution

$$k(x, t) = (4\pi t)^{-n/2} \exp [\|x\|^2/4t]$$

where

$$\|x\|^2 = \sum_{j=1}^{n} x_j^2 \qquad n = 2, 3.$$

If $u(x, t)$ be the solution to our problem, the function

$$v(x, t) = u(x, t) - u_1(x, t)$$

solves the problem

$$\Delta v = v_t \qquad x \in R,\ t > 0$$

$$v(x, 0) = 0 \qquad x \in \overline{R}$$

$$v(\xi, t) = \psi(\xi, t) - u_1(\xi, t), \qquad \xi \in B,\ t \geqq 0$$

and

$$v(\xi, t)|_{t=0} = \psi(\xi, 0) - u_1(\xi, 0) = \psi(\xi, 0) - \phi(\xi) = 0.$$

4. The integral equations. We study now the problem

$$\Delta u = u_t \qquad x \in R,\ t > 0$$

$$u(x, 0) = 0 \qquad x \in \overline{R}$$

$$u(\xi,\ t)=\psi(\xi,\ t) \qquad \xi\in B,\ t\geqq 0$$

with

$$\psi(\xi,\ 0)=0, \qquad \xi\in B\,.$$

Since R is assumed regular for $\Delta u=0$, let $\bar{u}(x,\ t)$ be the solution to the problem

$$\Delta\bar{u}(x,\ t)=0 \qquad x\in R$$

$$\bar{u}(\xi,\ t)=\psi(\xi,\ t), \qquad x\in R$$

Also since R is regular for $\Delta u=0$, the Green's function $G(x,\ y)$ exists and is symmetric function by E, and if $f(x)$ is differentiable the function

$$g(x)=-\int_{R}G(x,\ y)f(y)dy$$

solves the problem

$$\Delta g=f(x) \qquad x\in R$$

$$g(\xi)=0 \qquad \xi\in B\,.$$

(See Tamarkin and Feller [3]). Hence if $u(x,\ t)$ be the solution to our problem it must also satisfy the integral equation

$$u(x,\ t)=\bar{u}(x,\ t)-\int_{R}G(x,\ y)\frac{\partial}{\partial t}u(y,\ t)dy\,. \tag{3}$$

Conversely any solution of our integral equation which is differentiable in x (and which attains the proper initial values) must also solve our problem.

We apply the Laplace transform: let

$$\mathscr{L}\{u(x,\ t)\}=w(x,\ s), \quad \mathscr{L}\{\bar{u}(x,\ t)\}=v(x,\ s),$$

so that (3) becomes

$$w(x,\ s)=v(x,\ s)-s\int_{R}G(x,\ y)w(y,\ s)ds \tag{4}$$

which is a Fredholm integral equation with a symmetrical kernel $-G(x,\ y)$.

5. Restricted solution of the problem. To facilitate the solution of our integral equations (3) and (4) we make additional restrictions which will be removed later. We assume

(i) there exists $T>0$ such that $\psi(\xi,\ t)=0$ for $t>T$.

(ii) $\psi(\xi, t)$ in addition to being continuous with respect to (ξ, t), has four derivatives with respect to t which are also continuous with respect to (ξ, t) and

$$\psi_t(\xi, 0)=\psi_{tt}(\xi, 0)=\psi_{ttt}(\xi, 0)=0\,, \quad \xi \in B\,.$$

From D it follows that $\bar{u}(x, t)$ has four continuous derivatives with respect to t; and

$$\bar{u}_t(x, 0)=\bar{u}_{tt}(x, 0)=\bar{u}_{ttt}(x, 0)=0$$

for $x \in R$, by the maximum principle. From (i) it follows that

$$\bar{u}(x, t)=0, \quad \text{for} \quad t>T, \quad x \in \bar{R}.$$

Since $-G(x, y)$ is symmetric in (x, y) it follows that the eigenvalues of our problem are all real and in fact it is well known that they are all negative. (See for example, Tamarkin and Feller [3]).

The solution of (4) is

$$w(x, s)=v(x, s)+\sum_n \frac{s v_n(s)}{\lambda_n - s}\phi_n(x)\,, \tag{5}$$

where $\phi_n(x)$ are the eigenfunctions for the kernel $-G(x, y)$ and where

$$v_n(s)=\int_R \phi_n(x) v(x, s) dx\,.$$

We must now invert the Laplace transform and show that $\mathscr{L}^{-1}\{w(x, s)\}$ is the solution to our restricted problem. To this end we examine some of the properties of $w(x, s)$. We begin with an examination of $v(x, s)$.

By its definition we have

$$v(x, s)=\int_0^\infty e^{-st}\bar{u}(x, t)dt\,,$$

the integral being uniformly and absolutely convergent for $x \in \bar{R}$, and $\mathscr{R}\, s \geqq 0$. In fact any of the x derivatives of v can be computed under the integral sign, since the resulting integral is uniformly and absolutely convergent for $\mathscr{R}\, s \geqq 0$ and x in any closed sub-domain of R. So that, in particular,

$$\Delta v(x, s)=\int_0^\infty e^{-st}\Delta\bar{u}(x, t)dt=0\,.$$

Furthermore $v(x, s)$ is analytic for $\mathscr{R}\, s>0$, and bounded for $\mathscr{R}\, s \geqq 0$, and by integrating by parts, under of course the restrictions (i) and (ii) we get

$$v(x, s)=\frac{1}{s^4}\int_0^\infty \bar{u}_{tttt}(x, t)e^{-st}dt\ .$$

From this we see that

$$|v(x, s)|\leqq K_1/|s^4|\ , \qquad \mathscr{R}\, s\geqq 0,\ x\in \bar{R}$$

which is of interest only for large $|s|$ since $v(x, s)$ is bounded.

Since

$$w(x, s)=v(x, s)-\sum_n \frac{sv_n(s)}{(s/\lambda_n)-1}\cdot\frac{\phi_n(x)}{\lambda_n}$$

we get

$$|w(x, s)|\leqq|v(x, s)|+\left[\sum_n \frac{|s|^2|v_n(s)|^2}{|(s/\lambda_n)-1|^2}\cdot\sum_n\frac{\phi_n^2(x)}{\lambda_n^2}\right]^{1/2}$$

Now $\lambda_n\leqq 0$ so that $|(s/\lambda_n)-1|\geqq 1$, and hence

$$|w(x, s)|\leqq v(x, s)+|s|\left[\int_R |v(x,\ s)|^2\,dx\cdot\int_R G^2(x, y)dy\right]^{1/2}.$$

But $\int_R G^2(x, y)dy$ is bounded since G is continuous except for a singularity at x like $\log\|x-y\|$ or $1/\|x-y\|$, as the case may be. Hence

$$|\,w(x, s)|\leqq \frac{K_1}{|s|^4}+\frac{K_2}{|s|^3}\leqq\frac{K_3}{|s|^3}\quad \text{for } |s|\geqq 1$$

uniformly for $x\in R$, $\mathscr{R}\, s\geqq 0$, and

$$|w(x, s)|\leqq K_4, \qquad |s|\leqq 1,\ \mathscr{R}\, s\geqq 0$$

since $v(x, s)$ is bounded there.

Hence $w(x, s)$ is also bounded for all $x\in\bar{R}$, $\mathscr{R}\, s\geqq 0$, and for large $|s|$,

$$w(x, s)=O(1/|s|^3)$$

uniformly for $x\in\bar{R}$.

The inverse transform

$$(6)\qquad u(x, t)=\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} w(x, s)e^{ts}ds \qquad \sigma>0$$

exists, and since e^{st} is bounded and $w(x, s)=O(1/|s|^3)$ converges uniformly. Also

$$\frac{\partial u}{\partial t}(x,\ t)=\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} sw(x,\ s)e^{ts}ds\ ,$$

since the integral converges uniformly.

Since $w(x, s)$ satisfies (4), by applying the inverse transform to each side we are led back to (3), the integration under the integral sign being permissible by the uniform convergence of the integrals involved. Hence $u(x, t)$ as given by (6) where $w(x, s)$ is given by (5) is the solution to the integral equation (3), and as such is a solution to the heat equation in R and attains the proper boundary conditions. Let us examine the initial values of $u(x, t)$:

$$u(x,\ 0)=\frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} w(x,\ s)ds\ , \qquad \sigma>0,\ x\in\overline{R}$$

$$|u(x,\ 0)|\leqq\frac{1}{2\pi}\int_{\sigma-i\infty}^{\sigma+i\infty}|w(x,\ s)|\cdot|ds|\ ,$$

$$\leqq\frac{K_3}{2\pi}\int_{-\infty}^{\infty}\frac{d\tau}{(\sigma+i\tau)^3}=\frac{K_1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\frac{dv}{|1+iv|^3}$$

which tends to zero as σ becomes infinite. Hence $u(x,\ 0)=0,\ x\in\overline{R}$. This completes the solution in the restricted case.

6. Removal of the restrictions. We first remove the restrictions (ii).

Let $\psi(\xi,\ t)$ be continuous, $\xi\in B,\ t\geqq 0$, with $\psi(\xi,\ 0)=0,\ \xi\in B$, and $\psi(\xi,\ t)=0,\ t>T$. By the Weierstrass approximation theorem there is a polynomial $p_n(\xi,\ t)$ such that

$$|\psi(\xi,\ t)-p_n(\xi,\ t)|<1/4n\ , \qquad \xi\in B,\ 0\leqq t\leqq T.$$

By the uniformity of the continuity of $\psi(\xi,\ t)$ there exists $t_n,\ t_n'$ such that

$$|\psi(\xi,\ t)|<1/4n,\ \text{for}\begin{cases}0\leqq t\leqq t_n, & \xi\in B\\ t_n'\leqq t\leqq T, & \xi\in B\end{cases}$$

and without loss of generality we may, assume $t_n<1/2n$ and $T-t_n'<1/2n$.

Let $q_n(t)\in C^5\ 0\leqq t$, increase from 0 to 1 as t increases from 0 to t_n and be identically 1 for $t_n\leqq t\leqq t_n'$ and decrease to zero again at $t=T$, and have four vanishing derivatives at $t=0$ and at $t=T$.

Now let $\psi_n(\xi,\ t)=q_n(t)p_n(\xi,\ t)$. This function is an admissible boundary function under the restricted proof, which we have already completed. Hence for each n there is a solution $u_n(x,\ t)$ of the heat equation assuming these boundary values and of course zero initial values. To show that this sequence converges to the solution to our present problem we consider first

$$|\psi(\xi, t)-\psi_n(\xi, t)|=|\psi(\xi, t)-p_n(\xi, t)|<\frac{1}{4n}$$

for $t_n<t<t'_n$. For $0\leqq t\leqq t_n$ and $t'_n\leqq t\leqq T$,

$$|\psi(\xi, t)-\psi_n(\xi, t)|\leqq|\psi(\xi, t)|+|\psi_n(\xi, t)|$$

$$\frac{1}{4n}+|q_n(t)|\cdot|p_n(\xi, t)|\leqq\frac{1}{4n}+|p_n(\xi, t)|,$$

but

$$|p_n(\xi, t)|\leqq|p_n(\xi, t)-\psi_n(\xi, t)|+|\psi_n(\xi, t)|\leqq\frac{1}{4n}+\frac{1}{4n}=\frac{1}{2n},$$

so that

$$|\psi(\xi, t)-\psi_n(\xi, t)|<\frac{1}{2n}, \qquad 0\leqq t\leqq T,$$

and consequently

$$|\psi_n(\xi, t)-\psi_m(\xi, t)|\leqq\frac{1}{\min(m, n)}, \qquad 0\leqq t\leqq T.$$

For $x\in\bar{R}$, $0\leqq t\leqq T$

$$u_n(x, t)-u_m(x, t)$$

is a solution of $\Delta u=u_t$ in R and continuous for $x\in\bar{R}$, $0\leqq t\leqq T$. Hence by the maximum and minimum principles for the heat equation this function attains its maximum and its minimum on the bottom or lateral parts of the space time cylinder defined by $x\in\bar{R}$, $0\leqq t\leqq T$

It follows that

$$|u_n(x, t)-u_m(x, t)|\leqq\max_{\xi\in B,\ 0\leqq t\leqq T}|\psi_n(\xi, t)-\psi_m(\xi, t)|\leqq\frac{1}{\min(m, n)}$$

from which the uniform convergence of the sequence $u_n(x, t)$ in the cylinder is clear. The limit function, $u(x, t)$, clearly attains the proper initial values, since each of the approximating functions does. And for $\xi\in B$,

$$u(\xi, t)=\lim_{n\to\infty}u_n(\xi, t)=\lim_{n\to\infty}\psi_n(\xi, t)=\psi(\xi, t),$$

so that $u(x, t)$ is the solution to our problem under the restriction (i).

Consider now any $\psi(\xi, t)$, continuous for $\xi \in B$, $t \geqq 0$, which vanishes for $t=0$. Then let

$$r_n(t)=\begin{cases}1 & 0\leqq t\leqq n\\ 1+(n-t) & n\leqq t\leqq n+1\\ 0 & n+1\leqq t\,.\end{cases}$$

and this time let

$$\psi_n(\xi, t)=\psi(\xi, t)r_n(t)\,.$$

If $u_n(x, t)$ be the solution to the problem with data ψ_n we will again show convergence. For let (x, t) be any point, $x \in \overline{R}$, $t \geqq 0$, and let n and m each be greater than, say $2t$. Then

$$|u_n(x, t)-u_m(x, t)| \leqq \max_{0\leqq\tau\leqq 2t} |\psi_n(\xi, \tau)-\psi_m(\xi, \tau)|$$

where the maximum is computed over all $\xi \in B$, $0 \leqq \tau \leqq 2t$. But this maximum vanishes, hence $u_n(x, t)=u_m(x, t)$ for n, m sufficiently large. So that $\lim_{n\to\infty} u_n(x, t)$ exists and is a solution of the heat equation and takes on the prescribed initial and boundary values.

References

1. O. D. Kellog, *Foundations of potential theory*, New York, 1929.
2. E. Rothe, *Wärmleitungsgleichung mit nicht-konstanten Koeffizienten 1*, Math. Ann. **104**, (1931).
3. J. D. Tamarkin and W. Feller, *Partial differential equations*, Brown University, 1941.
4. A. Tychonoff, *Sur l'equation de la chaleur de plusieurs variables*, Bull. Univ. Etal. Moscow, Ser. Int., Sect. A: Math. et Mecan. 1. Fasc. **9**, (1938).

The University of Minnesota

A REAL INVERSION FORMULA FOR A CLASS OF BILATERAL LAPLACE TRANSFORMS

WILLIAM R. GAFFEY

1. Introduction. The Post-Widder inversion formula for unilateral Laplace transformations [1] states that, under certain weak restrictions on $\phi(u)$,

$$\lim_{k\to\infty}\left(\frac{k}{c}\right)^{k+1}\frac{1}{k!}\int_0^\infty \phi(u)u^k \exp\left(-k\frac{u}{c}\right)du=\phi(c)\ ,$$

for any continuity point c of $\phi(u)$.

This formula applies when $\phi(u)$ is defined only for $u\geqq 0$. A similar formula may be deduced if $\phi(u)$ is defined for $u\geqq -a$, for some positive a. In such a case, we may let $\phi^*(u)=\phi(u-a)$, and we may then use the Post-Widder formula to determine $\phi^*(u)$ at the point $u=c+a$. The inversion formula then becomes

$$\lim_{k\to\infty}\left(\frac{k}{c+a}\right)^{k+1}\frac{1}{k!}\int_0^\infty \phi(u-a)u^k \exp\left(-k\frac{u}{c+a}\right)du=\phi(c)\ ,$$

or, if we make the transformation $z=u/(c+a)$,

$$(1)\qquad \lim_{k\to\infty}\frac{k^{k+1}}{k!}\int_0^\infty \phi[(c+a)z-a]z^k \exp(-kz)dz=\phi(c)\ .$$

This suggests that, if $\phi(u)$ is defined for $-\infty<u<\infty$, some sort of limiting form of (1) applies. We shall prove that under suitable restrictions on ε and on the behavior of $\phi(u)$,

$$(2)\qquad \lim_{k\to\infty}\frac{k^{k+1}}{k!}\int_{-\infty}^\infty \phi[(c+k^\varepsilon)z-k^\varepsilon]z^k \exp(-kz)dz=\phi(c)\ .$$

2. Remarks. In the following sections $\phi(u)$ will be assumed to be integrable over the interval from $-\infty$ to ∞, and c will be assumed to be a continuity point of $\phi(u)$. All limits should be understood to be for increasing values of k.

The expression $\delta/(c+k^\varepsilon)$, where δ and ε are positive numbers, occurs frequently. It will be denoted by $\delta(k,\varepsilon)$.

Finally, it may be noted that in terms of the Laplace transform of $\phi(u)$ for real t,

Received December 7, 1955, and in revised form April 13, 1956.

$$f(t)=\int_{-\infty}^{\infty}\phi(u)\exp(-tu)du ,$$

the inversion formula (2) may be written in the form

$$\lim \frac{(-1)^k}{k!}\left(\frac{k}{c+k^{\varepsilon}}\right)^{k+1}\frac{d^k}{dt^k}[f(t)\exp(-tk^{\varepsilon})]_{t=k/(c+k^{\varepsilon})}=\phi(c) .$$

3. Preliminary proofs. The results of the following four lemmas will be needed below. Proofs are given for the first two. The second two are proved in a similar way.

LEMMA 1. *If n is any fixed number and $1/3<\varepsilon<1/2$, then*

$$\lim k^n[1+\delta(k, \varepsilon)]^k\exp[-k\delta(k, \varepsilon)]=0 .$$

Proof. If the logarithm of the expression under the limit sign is expanded in powers of $\delta(k, \varepsilon)$, the sum of two of the terms in the expansion approaches $-\infty$ as $k\to\infty$, while the sum of the rest of the terms is bounded.

LEMMA 2. *If $1/3<\varepsilon<1/2$, then*

$$\lim \frac{k^{k+1}}{k!}\int_1^{1+\delta(k,\varepsilon)} z^k\exp(-kz)dz=\frac{1}{2} .$$

Proof. It is well known [1] that

$$\lim \frac{k^{k+1}}{k!}\int_1^{\infty} z^k\exp(-kz)dz=\frac{1}{2} .$$

Therefore, it is sufficient to show that

$$\lim \frac{k^{k+1}}{k!}\int_{1+\delta(k,\varepsilon)}^{\infty} z^k\exp(-kz)dz=0 .$$

Since $z\exp(-z)$ is a decreasing function of z for $z>1$, the above expression is, for fixed k, no larger than

$$\frac{k^{k+1}}{k!}[1+\delta(k, \varepsilon)]^{k-1}\exp[-(k-1)(1+\delta(k, \varepsilon))]\int_{1+\delta(k,\varepsilon)}^{\infty} z\exp(-z)dz .$$

By applying Stirling's formula and Lemma 1, we see that the upper bound approaches zero as k increases.

LEMMA 3. *If n is any fixed number and $0<\varepsilon<1/2$, then*

$$\lim k^n[1-\delta(k, \varepsilon)]^k \exp [k\delta(k, \varepsilon)]=0 ,$$

LEMMA 4. *If* $0<\varepsilon<1/2$, *then*

$$\lim \frac{k^{k+1}}{k!}\int_{1-\delta(k,\varepsilon)}^{1} z^k \exp(-kz)dz=\frac{1}{2} .$$

4. The inversion formula.

THEOREM. *If*

$$\text{(a)} \qquad \left|\int_{-\infty}^{-x} \phi(z)\,dz\right| \leqq A \exp(-dx^{2+\alpha})$$

for some positive quantities A, d, *and* α, *and if*

$$\text{(b)} \qquad \max(1/3,\ 1/(2+\alpha))<\varepsilon<1/2,$$

then

$$\lim I_k=\lim \frac{k^{k+1}}{k!}\int_{-\infty}^{\infty} \phi[(c+k^\varepsilon)z-k^\varepsilon]z^k \exp(-kz)dz=\phi(c) .$$

Proof. For any $\delta>0$, the infinite interval may be partitioned into the four subintervals $(-\infty,\ 1-\delta(k, \varepsilon))$, $(1-\delta(k, \varepsilon),\ 1)$, $(1,\ 1+\delta(k, z))$, and $(1+(k, \varepsilon),\ \infty)$. I_k may be considered as the sum of four integrals over these intervals, so that we may write

$$I_k=I_k^{(1)}+I_k^{(2)}+I_k^{(3)}+I_k^{(4)} .$$

$I_k^{(1)}$ is understood to represent the integral over $(-\infty,\ 1-\delta(k, \varepsilon))$ etc.

$$|I_k-\phi(c)| \leqq |I_k^{(1)}|+\left|I_k^{(2)}-\frac{\phi(c)}{2}\right|+\left|I_k^{(3)}-\frac{\phi(c)}{2}\right|+|I_k^{(4)}| .$$

We prove first that $I_k^{(1)}$ and $I_k^{(4)}$ approach zero as $k\to\infty$. For $I_k^{(1)}$, consider first the integral over the interval from 0 to $1-\delta(k, \varepsilon)$. The function $z\exp(-z)$ attains its maximum at the upper endpoint. Therefore an upper bound for the absolute value of this portion of the expression is

$$\frac{k^{k+1}}{k!}[1-\delta(k, \varepsilon)]^k \exp[-k+k\delta(k, \varepsilon)]\int_0^{1-\delta(k,\varepsilon)} |\phi[(c+k^\varepsilon)z-k^\varepsilon]|dz ,$$

which approaches zero by Stirling's formula and Lemma 3.

Consider now the integral over the interval from $-\infty$ to 0. Integrating by parts, we find that it is equal to

$$-\frac{1}{c+k^\varepsilon}\frac{k^{k+2}}{k!}\int_{-\infty}^{0} F[(c+k^\varepsilon)z-k^\varepsilon]z^{k+1}(1-z)\exp(-kz)dz\ ,$$

where $F(z)=\int_{-\infty}^{z}\phi(u)du$. Note that, by the assumption on $F(z)$,

$$|F[(c+k^\varepsilon)z-k^\varepsilon]|\leqq A\exp[-d\{-(c+k^\varepsilon)z+k^\varepsilon\}^{2+\alpha}]\ ,$$

which is in turn equal to or less than

$$A\exp[dz(c+k^\varepsilon)k^{\varepsilon(1+\alpha)}]\ .$$

The result of the integration by parts may be written as the difference between two integrals, the first containing z^{k-1} and the second containing z^k. The first integral is no greater in absolute value than

$$\frac{A}{(c+k^\varepsilon)}\frac{k^{k+2}}{k!}\int_{-\infty}^{0}|z^{k-1}|\exp[z\{d(c+k^\varepsilon)k^{\varepsilon(1+\alpha)}-k\}]dz\ .$$

Since $\varepsilon(2+\alpha)>1$, the coefficient of z in the exponent above is positive for sufficiently large k. Therefore, after some manipulation, this upper bound can be shown to be equal to

$$\frac{A}{(c+k^\varepsilon)}\frac{k^{k+2}}{k!}\cdot\frac{\Gamma(k)}{[d(c+k^\varepsilon)k^{\varepsilon(1+\alpha)}-k]^k}\ ,$$

which approaches zero as $k\to\infty$.

By the same argument, the second integral approaches zero, so that $\lim I_k^{(1)}=0$.

For $I_k^{(4)}$, observe that since $z\exp(-z)$ is a decreasing function of z for $z>1$, the expression has the following upper bound for its absolute value:

$$\frac{k^{k+1}}{k!}[1+\delta(k,\ \varepsilon)]^k\exp[-k-k\delta(k,\ \varepsilon)]\int_{1+\delta(k,\ \varepsilon)}^{\infty}|\phi[(c+k^\varepsilon)z-k^\varepsilon]|dz\ .$$

Since the integral is bounded, the whole upper bound approaches zero by virtue of Stirling's formula and Lemma 1.

We now prove that

$$\left|\lim I_k^{(3)}-\frac{1}{2}\phi(c)\right|<\frac{1}{2}\eta$$

for any $\eta>0$. By Lemma 2, it is sufficient to show that

$$\left|\lim\frac{k^{k+1}}{k!}\int_{1}^{1+\delta(k,\ \varepsilon)}\{\phi[(c+k^\varepsilon)z-k^\varepsilon]-\phi(c)\}z^k\exp(-kz)dz\right|<\frac{\eta}{2}\ .$$

Since c is a continuity point of $\phi(u)$, there is a $\delta > 0$ such that if $|(c+k^\varepsilon)z-k^\varepsilon-c| < \delta$, that is, if $|z-1| < \delta(k, \varepsilon)$, then

$$|\phi[(c+k^\varepsilon)z-k^\varepsilon]-\phi(c)| < \eta .$$

For such a δ, the absolute value of the expression above is equal to or less than

$$\eta \lim \frac{k^{k+1}}{k!}\int_1^{1+\delta(k,\varepsilon)} z^k \exp(-kz)dz = \frac{\eta}{2} .$$

By the use of Lemma 4, it may be shown in a similar way that

$$\left|\lim I_k^{(2)} - \frac{1}{2}\phi(c)\right| < \frac{1}{2}\eta .$$

Putting together these results, we have $|\lim I_k - \phi(c)| < \eta$ for any $\eta > 0$, which proves the theorem.

REFERENCE

1. C. V. Widder, *Inversion of the Laplace transform and the related moment problem*, Trans. Amer. Math. Soc. **36** (1934), 107–200.

UNIVERSITY OF CALIFORNIA, BERKELEY

ON CHARACTERISTIC FUNCTIONS OF BANACH SPACE VALUED RANDOM VARIABLES

R. K. GETOOR

1. Introduction. In recent years several authors have considered the notion of random variables with values in a Banach space, $\mathfrak{X}$. One of the basic problems is to characterize those positive definite functions on $\mathfrak{X}^*$ that are characteristic functions of such random variables. Mourier [4] has given a solution to this problem if $\mathfrak{X}$ is separable and reflexive. The purpose of this paper is to give another solution of this problem. Our results are valid if $\mathfrak{X}$ is reflexive. However the contribution of this paper is not so much the removal of the condition of separability, rather we feel that our method sheds new light on the problem and aids in understanding it. The basic tool that we use is the concept of a weak distribution as introduced by Segal [5], and this idea succeeds in unifying the theory.

Section 2 contains the basic definitions and preliminaries. The main results are contained in § 3 but in a form slightly more general than needed for the problem at hand. However we will need the results in this generality in a future paper. The contents of § 3 are clearly valid in any locally convex linear topological space. Finally in § 4 our solution to the problem stated above is given along with some examples and consequences.

The considerations of Bochner in chapters five and six of [1] are somewhat related to our problem.

2. Definitions. Let $(\Omega, \mathfrak{F}, P)$ be a probability space, that is, Ω is an abstract point set, $\mathfrak{F}$ a σ-algebra of subsets of Ω, and P is a measure on $(\Omega, \mathfrak{F})$ with $P(\Omega)=1$. Let $\mathfrak{X}$ be a real Banach space[1] and $\mathfrak{X}^*$ its conjugate space. Let $X: \Omega \to \mathfrak{X}$, we will call X an $\mathfrak{X}$ *valued random variable* if X is weakly measurable, that is, if $\langle x^*, X(\omega)\rangle$ is a real valued $\mathfrak{F}$-measurable function for each $x^* \in \mathfrak{X}^*$. Let $E(X)$ be the Pettis integral of X with respect to P, provided it exists. Thus $E(X)$ is the unique element of $\mathfrak{X}$ such that $\langle x^*, E(X)\rangle = E\{\langle x^*, X\rangle\} = \int \langle x^*, X(\omega)\rangle dP$ for each $x^* \in \mathfrak{X}^*$. The *characteristic function* of X is defined as follows,

$$\phi(x^*) = E\{e^{i\langle x^*, X\rangle}\} = \int e^{i\lambda} dF((x^*; \lambda) \tag{2.1}$$

Received January 16, 1956. This research supported in part by Office of Ordnance Research, U.S. Army, contract DA-36-034-ORD-1296 RD.

[1] The extension to a complex Banach space is essentially clear.

where $F(x^*; \lambda)$ is the distribution function of the real valued random variable $\langle x^*, X(w)\rangle$. It follows that $\phi(0)=1$, ϕ is positive definite, and ϕ is continuous. For a detailed discussion of the above concepts see [4].

If we put $L(x^*)=\langle x^*, X(\omega)\rangle$ then L is a linear map from $\mathfrak{X}^*$ to random variables. Segal [5] defines a "weak distribution", L, on $\mathfrak{X}$ to be a linear map from $\mathfrak{X}^*$ to random variables. However there are two interpretations of this statement. We may mean $L\left(\sum_{i=1}^{n} a_i x_i^*\right)=\sum_{i=1}^{n} a_i L(x_i^*)$ with probability one or the stronger statement that for almost all ω the function $L(\cdot, \omega)$ is linear[2]. Theorem 2 of the next section shows that these two possibilities are actually equivalent. Thus since there is a possible ambiguity and since we want to consider a weak distribution as the generalization of an ordinary n-dimensional distribution we make the following definition.

DEFINITION 2.1. A *weak distribution*, L, on $\mathfrak{X}$ is a map which assigns to each finite collection of elements $(x_1^*, \cdots, x_n^*)$ in $\mathfrak{X}^*$ an n-dimensional distribution function $F_n(x_1^*, \lambda_1; \cdots; x_n^*, \lambda_n)$ such that

(1) F_n is symmetric in the pairs (x_i^*, λ_i) .

(2) $F_n(x_1^*, \lambda_1; \cdots; x_n^*, \infty)=F_{n-1}(x_1^*, \lambda_1; \cdots; x_{n-1}, \lambda_{n-1})$.

(3) If $\sum_{i=1}^{n} a_1 x_i^*=0$ then

$$\int_{\sum_{i=1}^{n} a_i\lambda_i \leqq \lambda} dF_n(x_1^*, \lambda_1; \cdots; x_n^*, \lambda_n)=\varepsilon(\lambda)$$

where $\varepsilon(\lambda)$ is the unit distribution, $\varepsilon(\lambda)=\begin{cases}0 & \lambda<0\\ 1 & \lambda\geqq 0\end{cases}$.

Note. Condition (3) implies that if $x^*=\sum_{i=1}^{n} a_i x_i^*$ then

$$F(x^*, \lambda)=\int_{\sum_{i=1}^{n} a_i\lambda_i\leqq\lambda} dF_n(x_1^*, \lambda_1; \cdots; x_n^*, \lambda_n) .$$

EXAMPLE. If X is an $\mathfrak{X}$ valued random variable then there is associated with it in a natural way a weak distribution which assigns to $(x_1^*, \cdots, x_n^*)$ the joint distribution function of the random variables $\langle x_i^*, X(\omega)\rangle$, thus,

$$F(x_1^*, \lambda_1; \cdots; x_n^*, \lambda_n)=\Pr[\langle x_i^*, X(\omega)\rangle\leqq\lambda_i,\ i=1, \cdots, n] .$$

[2] Since "random variable" in [5] in treated as a residue class module null sets, it clearly seems that the definition of weak distribution given there refers to the first interpretation.

Given a weak distribution L we define its characteristic function $\phi(x^*)$ by

$$\phi(x^*)=\int e^{i\lambda}dF(x^*;\lambda) . \tag{2.2}$$

It is clear from (2.1) that if L comes from X (as in the example) then the characteristic function of L and X are the same.

We are going to give conditions that a weak distribution come from some $\mathfrak{X}$ valued random variable and hence that ϕ be the characteristic function of some $\mathfrak{X}$ valued random variable.

3. **The main theorems.** From the definitions in the preceding section we see that L and its characteristic fuunction ϕ are both defined relative to $\mathfrak{X}^*$. In other words in the study of the relations between ϕ and L the space $\mathfrak{X}$ plays no role. We are thus led to define a *q-weak distribution* on $\mathfrak{X}$ as a map, L, from finite sets of elements $(x_1, \cdots, x_n)$ in $\mathfrak{X}$ to distribution functions which satisfies the conditions of Definition 2.1. We can now state our first theorem.

THEOREM 1. *There is a unique one-to-one correspondence between q-weak distributions L defined on $\mathfrak{X}$ and positive definite functions ϕ defined on $\mathfrak{X}$ satisfying* (i) $\phi(0)=1$ (ii) *ϕ is continuous on each finite dimensional subspace of $\mathfrak{X}$. We say that ϕ is the Fourier transform of L and denote the correspondence by $\phi=\mathfrak{F}(L)$.*[3]

Proof. In the following we will need a formula for change of variables in Lebesgue-Stieltjes integrals that we give here for convenience. If

$$F(\lambda)=\int_{\sum_{i=1}^{n} a_i\lambda_i\leq\lambda} dF_n(\lambda_1, \cdots, \lambda_n)$$

then for any bounded Borel measurable function, f, we have

$$\int f(\lambda)dF(\lambda)=\int f\Big(\sum_{i=1}^{n} a_i\lambda_i\Big)dF_n(\lambda_1, \cdots, \lambda_n) . \tag{3.1}$$

Given L we define the corresponding ϕ by

$$\phi(x)=\int e^{i\lambda}dF(x;\lambda) . \tag{3.2}$$

[3] This result was essentially contained in a lecture of I. Segal given at the Institute for Advanced Study during the academic year 1954-55. See also [**1**]. Note that (ii) can be replaced by the equivalent condition that ϕ is continuous at 0 on each finite dimensional subspace.

Clearly $\phi(0)=1$ and since

$$\phi\Big(\sum_{k=1}^{n}\alpha_k x_k\Big)=\int e^{i\lambda}dF\Big(\sum_{k=1}^{n}\alpha_k x_k;\ \lambda\Big)=\int e^{i\sum_{k=1}^{n}\alpha_k\lambda_k}dF_n(x_1,\ \lambda_1;\ \cdots;\ x_n,\ \lambda_n)$$

(using the formula stated above) it is evident that ϕ is continuous on each finite dimensional subspace of $\mathfrak{X}$. Moreover

$$\begin{aligned}\sum_{j,k=1}^{n}\alpha_k\bar{\alpha}_j\phi(x_k-x_j)&=\sum_{j,k=1}^{n}\alpha_k\bar{\alpha}_j\int e^{i\lambda}dF(x_k-x_j,\ \lambda)\\&=\sum_{j,k=1}^{n}\alpha_k\bar{\alpha}_j\int e^{i(\lambda_k-\lambda_j)}dF_2(x_k,\ \lambda_k;\ x_j,\ \lambda_j)\\&=\sum_{j,k=1}^{n}\alpha_k\bar{\alpha}_j\int e^{i(\lambda_k-\lambda_j)}dF_n(x_1,\ \lambda_1;\ \cdots;\ x_n,\ \lambda_n)\\&=\int\Big|\sum_{j,k=1}^{n}\alpha_k e^{i\lambda_k}\Big|^2dF_n\geqq 0\ .\end{aligned}$$

Thus ϕ is positive definite and satisfies the conditions of Theorem 1.

Conversely suppose we are given ϕ satisfying the conditions of Theorem 1. For any finite set of elements $(x_1,\ \cdots,\ x_n)$ we consider the function $\psi(\alpha_1,\ \cdots,\ \alpha_n)=\phi\Big(\sum_{k=1}^{n}\alpha_k x_k\Big)$. It then follows that ψ is an n-dimensional characteristic function in the ordinary sense. Hence by the n-dimensional Bochner theorem there exists a distribution function $F_n(x_1,\ \lambda_1;\ \cdots;\ x_n,\ \lambda_n)$ such that

$$\phi\Big(\sum_{k=1}^{n}\alpha_k x_k\Big)=\psi(\alpha_1,\ \cdots,\ \alpha_n)=\int e^{i\sum_{k=1}^{n}\alpha_k\lambda_k}dF_n(x_1,\ \lambda_1;\ \cdots;\ x_n,\ \lambda_n)\ . \tag{3.3}$$

By using the uniqueness assertion of the n-dimensional Bochner theorem it is easy to show that the above construction actually defines a q-weak distribution on $\mathfrak{X}$. The fact that the correspondence established between the ϕ's and the L's is one-to-one (and unique) again follows from the uniqueness in the n-dimensional Bochner theorem.

COROLLARY 1. *A necessary and sufficient condition that ϕ be continuous on $\mathfrak{X}$ is that $F(x,\ \lambda)\rightarrow\varepsilon(\lambda)$ as $x\rightarrow 0$.*

Proof. This is an immediate consequence of the representation (3.2) and the properties of ordinary characteristic functions.

The following example shows that there actually exist positive definite functions continuous on each finite dimensional subspace without being continuous. Let $\mathfrak{X}$ be a separable Hilbert space and let $\{e_\sigma\}$ be a linear base, thus if $x\in\mathfrak{X}$ then $x=\sum_{1}^{n}\alpha_j e_{\sigma_j}$ and this expression is unique. It is no restriction to assume $\|e_\sigma\|=1$ for all σ. Let $\{\sigma_n\}$ be a given sequence from $\{\sigma\}$, and let Y_n be independent Gaussian random variables (real valued) such that $E(Y_n)=0$ and $E(Y_n^2)=n$. Put $Y_{\sigma_n}=Y_n$ and $Y_\sigma=0$ is $\sigma\neq\sigma_n$ for some n. If $x=\sum_1^k \alpha_j e_{\sigma_j}$ we define $L(x)=\sum_1^k \alpha_j Y_{\sigma_j}$. This then defines a q-weak distribution, L, on $\mathfrak{X}$ as described in the example of § 2, that is, $F_n(x_1, \lambda_1; \cdots; x_n, \lambda_n)$ is the joint distribution of $L(x_1), \cdots, L(x_n)$. Let $\phi=\mathfrak{F}(L)$ then, according to Theorem 2, ϕ is continuous on each finite dimensional subspace. However $\frac{1}{\sqrt{n}}e_{\sigma_n}\to 0$ while for each n $F\left(\frac{1}{\sqrt{n}}e_{\sigma_n}, \lambda\right)$ is the standard normal distribution with mean 0 and variance 1. Thus by Corollary 1 we see that ϕ is not continuous on $\mathfrak{X}$.

Since for any q-weak distribution on $\mathfrak{X}$ the family of associated distribution functions satisfies the Kolmogorov compatibility conditions we can construct a stochastic process in $R^{\mathfrak{X}}$ (R is the real number system) which induces the given distribution functions. If we put $\Omega=R^*$ then we can denote this stochastic process by $L(x, \omega)=\omega(x)$ and the joint distribution of $L(x_1, \omega), \cdots, L(x_n, \omega)$ is given by $F_n(x_1, \lambda_1; \cdots; x_n, \lambda_n)$. See [2]. Taking into account condition (3) of Definition 2.1 it is clear that one should expect the sample functions $L(\cdot, \omega)$ to be linear in some sense. The next theorem states that $L(\cdot, \omega)$ is a linear function for almost all ω.

THEOREM 2. *Given a q-weak distribution L on $\mathfrak{X}$ then the stochastic process $L(x, \omega)$ can be realized in the space of all linear functions from $\mathfrak{X}$ to R, that is, in the algebraic dual of $\mathfrak{X}$.*

Proof. Let Ω be the set of all linear functions from $\mathfrak{X}$ to R, let $\mathfrak{F}$ be the field of cylinder sets of Ω. $\mathfrak{A}\in\mathfrak{F}$ if and only if $\mathfrak{A}=\{\omega/(\omega(x_1), \cdots, \omega(x_n))\in A_n\}$ where A_n is a Borel set in R^n. Let P_n be the n-dimensional measure induced by $F_n(x_1, \lambda_1; \cdots; x_n, \lambda_n)$ and then we put $P(\mathfrak{A})=P_n(A_n)$. We will now show that P is a completely additive measure on $\mathfrak{F}$.

(1) If $\mathfrak{A}\in\mathfrak{F}$ then $P(\mathfrak{A})$ is uniquely determined. This is proved in exactly the same way as in Kolmogorov [2].

(2) $P(\Omega)=1$. Clear.

(3) If $\mathfrak{A}$ and $\mathfrak{B}$ are disjoint cylinder sets then $P(\mathfrak{A}\cup\mathfrak{B})=P(\mathfrak{A})+P(\mathfrak{B})$.

Let $\mathfrak{A}=\{\omega/(\omega(x_1), \cdots, \omega(x_k))\in A_k\}$ and $\mathfrak{B}=\{\omega/(\omega(x_1'), \cdots, \omega(x_j'))\in B_j\}$, then by assumption $\mathfrak{A}\cap\mathfrak{B}=0$. Let $(y_1, \cdots, y_n)$ contain all the x_i's and x_i''s in some order and let A be the cylinder set in R^n with base A_k in R^k and B be the cylinder set in R^n with base B_j in R^j. We claim that $P_n(A\cap B)=0$. Suppose not, that is, $P_n(A\cap B)>0$. We distinguish two cases. First suppose $(y_1, \cdots, y_n)$ are linearly independent. $A \cap B \neq 0$ since $P_n(A \cap B) > 0$. Let $(\lambda_1, \cdots, \lambda_n) \in A \cap B$, define $\omega(y_i) = \lambda_i$ and extend ω linearly to the linear extension, $\{y_1, \cdots, y_n\}_L$, of $(y_1, \cdots, y_n)$. Then we can extend ω to a linear function on all of $\mathfrak{X}$ (ω can even be taken to be continuous by the Hahn-Banach theorem). Thus $\omega \in \Omega$ and $\omega \in \mathfrak{A} \cap \mathfrak{B}$ which is a contradiction. Second suppose there is a linear relation, $\sum_1^n \alpha_i y_i = 0$, among the y_i's. Since

$$\int_{\sum \alpha_i \lambda_i \leqq \lambda} dF(y_1, \lambda_1; \cdots; y_n, \lambda_n) = \varepsilon(\lambda)$$

the measure P_n in R^n is concentrated on the subspace $\sum_{i=1}^n \alpha_i \lambda_i = 0$. Because $P_n(A \cap B) > 0$ there exists a point $(\lambda_1, \cdots, \lambda_n) \in A \cap B$ such that $\sum_{i=1}^n \alpha_i \lambda_i = 0$. If we define ω as before we obtain the same contradiction. Thus $P_n(A \cap B) = 0$. Now

$$P(\mathfrak{A} \cup \mathfrak{B}) = P_n(A \cup B) = P_n(A) + P_n(B) - P_n(A \cap B) = P(\mathfrak{A}) + P(\mathfrak{B}) .$$

(4) P is completely additive on $\mathfrak{F}$. This again can be proved exactly as in [2].

We can now extend P to a completely additive measure on the σ-algebra, $\mathfrak{F}'$, generated by $\mathfrak{F}$ and thus the proof of Theorem 2 is complete.

The next theorem gives conditions under which $L(\cdot, \omega)$ is continuous for almost all ω, that is, $L(\cdot, \omega) \in \mathfrak{X}^*$ for almost all ω. The proof is fashioned after a proof given by Mann [3] in the real valued case.

THEOREM 3. *A necessary and sufficient condition that $L(x, \omega)$ is realizable in the space, $\mathfrak{X}^*$, of all continuous linear functions from $\mathfrak{X}$ to R is that for any separable subspace $\mathfrak{X}'$ and any $\varepsilon, \eta > 0$ there exists $\delta = \delta(\varepsilon, \eta, \mathfrak{X}')$ such that for any finite collection $x_1, \cdots, x_n \in \mathfrak{X}'$ with $||x_i|| \leq \delta$ we have*

$$\int_{-\varepsilon}^{\varepsilon} \cdots \int_{-\varepsilon}^{\varepsilon} dF_n(x_1, \lambda_1; \cdots; x_n, \lambda_n) \geqq 1 - \eta \tag{3.4}$$

Proof of sufficiency. First note that if $\lambda_i = \varepsilon$ and $\lambda_i = -\varepsilon$ $(i = 1, 2, \cdots, n)$ are continuity points of F_n then the integral (3.4) is equal to $P[\max_{1 \leqq i \leqq n} |L(x_i)| \leq \varepsilon]$. For the purposes of this proof we denote $\mathfrak{X}^*$ by Ω and then as in the proof of Theorem 2 we can introduce a finitely additive measure, P, on the field, $\mathfrak{F}$, of cylinder sets in Ω. We will now show that P is completely additive. As is well known it is sufficient to show that if $\mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots$ is a decreasing sequence of cylinder sets such that $\bigcap_{n=1}^{\infty} \mathfrak{A}_n = 0$ then $P(\mathfrak{A}_n) \searrow 0$. Assume $P(\mathfrak{A}_n) \searrow \theta > 0$. It is no loss of generality to assume that $\mathfrak{A}_n$ is defined by $x_1, \cdots, x_n$ and a closed Borel set A_n in R^n

$$\mathfrak{A}_n = \{\omega | (\omega(x_1), \cdots, \omega(x_n)) \in A_n\} .$$

Let $\mathfrak{X}'$ be the separable subspace generated by $\{x_1, x_2, \cdots\}$, and let $\{y_1, \cdots\}$ be the set of all finite linear combinations of the x_i's with rational coefficients. Thus $\{y_i\}$ is a countable dense set in $\mathfrak{X}'$ and we arrange the notation so that $\mathfrak{A}_n$ depends on $y_{k_1}, \cdots, y_{k_n}$ where $k_i < k_j$ if $i < j$ and $k_n \to \infty$ as $n \to \infty$. By hypothesis we can choose a $\delta_2 > 0$, independent of n, such that

$$P\left[\omega | \max |\omega(y_i)| \leqq \frac{\theta}{4} \text{ for } i \leqq k_n \text{ and } ||y_i|| \leqq \delta_2\right] \geqq 1 - \frac{\theta}{4} .$$

Define $\mathfrak{A}_n^2 = \left\{\omega | \omega \in \mathfrak{A}_n \text{ and } \max |\omega(y_i)| \leqq \frac{\theta}{4};\ i \leqq k_n,\ ||y_i|| \leqq \delta_2\right\}$, then since $P(\mathfrak{A}_n) \geqq \theta$ we have $P(\mathfrak{A}_n^2) \geqq \frac{\theta}{2} + \frac{\theta}{4}$ and $\mathfrak{A}_n^2 \subset \mathfrak{A}_n$. Also $\mathfrak{A}_1^2 \supset \mathfrak{A}_2^2 \cdots$. Similarly we define inductively

$$\mathfrak{A}_n^p = \left\{\omega \in \mathfrak{A}_n^{p-1},\ \max |\omega(y_i)| \leqq \frac{\theta}{2^p} \text{ for } i \leqq k_n,\ ||y_i|| \leqq \delta_p\right\}$$

and $P(\mathfrak{A}_n^p) \geqq \frac{\theta}{2} + \frac{\theta}{2^p}$. More over $\mathfrak{A}_1^p \supset \mathfrak{A}_2^p \supset \cdots$, and $\mathfrak{A}_n^p \subset \mathfrak{A}_n^{p-1} \subset \cdots \subset \mathfrak{A}_n$. Consider the sequence $\mathfrak{A}_p^p$ and note that $\mathfrak{A}_1^1 \supset \mathfrak{A}_2^2 \supset \cdots$, also note that $\mathfrak{A}_p^p$ depends on $y_1, \cdots, y_{k_p}$. Moreover the above inequality shows that $P(\mathfrak{A}_p^p) \geqq \frac{\theta}{2}$. We can now replace the $\mathfrak{A}_p^p$ by sets $\mathfrak{B}_p$ depending on $y_1, \cdots, y_{k_p}$ such that $\mathfrak{B}_p \subset \mathfrak{A}_p^p$ and the corresponding Borel set, B_p, in R^p is closed and bounded and $P(\mathfrak{B}_p) \geqq \frac{\theta}{4}$. See [2].

Now choose $\omega_p \in \mathfrak{B}_p$ and by the diagonal process we can choose a subsequence (which we again denote by ω_p) such that $\omega_p(y_i) \to \lambda_i$ for each y_i. Since B_p is closed the point $(\lambda_1, \cdots, \lambda_{k_p})$ is in B_p and thus if we define $\omega(y_i) = \lambda_i$ for all y_i we see that if we can show $\omega \in \mathfrak{X}^*$ it will then follow that $\omega \in \mathfrak{B}_p$. Clearly ω is rational linear on $\{y_i\}$. We now show that ω is uniformly continuous on $\{y_i\}$. Given $\varepsilon > 0$ choose j such that $\frac{\theta}{2^j} \leqq \frac{\varepsilon}{2}$ and then choose $\delta = \delta_j$ (the δ_j used in the construction of $\mathfrak{A}_n^i$). For any y_i with $||y_i|| \leqq \delta$ we have

$$|\omega(y_i)| \leqq |\omega(y_i) - \omega_p(y_i)| + |\omega_p(y_i)| .$$

Since $\omega_p(y_i) \to \omega(y_i)$ we can choose a p_0 such that $p \geqq p_0$ implies $|\omega(y_i) - \omega_p(y_i)| \leqq \frac{\varepsilon}{2}$. We choose a p such that $p \geqq p_0$, $p \geqq j$, and $k_p \geqq i$, then $\omega_p \in \mathfrak{B}_p \subset \mathfrak{A}_p^p \subset \mathfrak{A}_p^j$ and hence $|\omega_p(y_i)| \leqq \frac{\theta}{2^j} \leqq \frac{\varepsilon}{2}$. Thus if $\|y_i\| \leqq \delta$ then $|\omega(y_i)| \leqq \varepsilon$. Since ω is rational linear on $\{y_i\}$ it follows that ω is uniformly continuous and hence can be extended by continuity to $\mathfrak{X}'$. Clearly the extension will be linear on $\mathfrak{X}'$, and hence by the Hahn-Banach theorem ω can be extended to be a continuous linear function on $\mathfrak{X}$. It now follows that $\omega \in \mathfrak{B}_p$ for all p and since $\mathfrak{B}_p \subset \mathfrak{A}_p$ we have that $\omega \in \bigcap_{k=1}^{\infty} \mathfrak{A}_k$. Hence P is completely additive on $\mathfrak{F}$.

Proof of necessity. Since $L(\cdot, \omega) \in \mathfrak{X}^*$ for almost all ω we can write $L(x, \omega) = \langle x, X^*(\omega) \rangle$. Let $\mathfrak{X}'$ be a separable subspace of $\mathfrak{X}$ and let $\|X^*(\omega)\|'$ be the norm of $X^*(\omega)$ when considered as a linear functional on $\mathfrak{X}'$, then $\|X^*(\omega)\|'$ is a measurable function. Given $\varepsilon, \eta > 0$ we can choose $\delta > 0$ such that $P[\|X^*\|' \leqq \varepsilon/\delta] \geqq 1 - \eta$ and this δ has the required properties.

4. Application to $\mathfrak{X}$ valued random variables. We can now give a solution to the problem stated in the introduction in case $\mathfrak{X}$ is a reflexive space.

THEOREM 4. *Let $\mathfrak{X}$ be a real reflexive Banach space and $\phi(x^*)$ be a positive definite function on $\mathfrak{X}^*$. A necessary and sufficient condition that ϕ is the characteristic function of an $\mathfrak{X}$ valued random variable is that:*

(i) *$\phi(0) = 1$ and ϕ be continuous on each finite dimensional subspace of $\mathfrak{X}^*$.*

(ii) *If $L = \mathfrak{F}(\phi)$ (which exists by* (i) *and Theorem* 1) *then for any separable subspace $\mathfrak{X}_1^*$ of $\mathfrak{X}^*$ and any $\varepsilon, \eta > 0$ there exists $\delta = \delta(\mathfrak{X}_1^*, \varepsilon, \eta)$ such that for any finite collection $x_1^*, \cdots, x_n^* \in \mathfrak{X}_1^*$ with $\|x_i^*\| \leqq \delta$ we have*

$$\int_{-\varepsilon}^{\varepsilon} \cdots \int_{-\varepsilon}^{\varepsilon} dF_n(x_1^*, \lambda_1; \cdots; x_n^*, \lambda_n) \geqq 1 - \eta .$$

Proof. L is a q-weak distribution on $\mathfrak{X}^*$ which satisfies the conditions of Theorem 3 relative to $\mathfrak{X}^*$. Hence L can be realized in $\mathfrak{X}^{**} = \mathfrak{X}$ since $\mathfrak{X}$ is assumed reflexive. Thus $L(x^*, \omega) = \langle x^*, X(\omega) \rangle$ and $X(\omega)$ is weakly measurable since $L(x^*, \cdot)$ is measurable for all x^*. But ϕ is the characteristic function of L and hence as remarked in § 2 it is the characteristic function of X. The necessity of the above conditions is obvious if we apply Theorem 3.

We conclude by giving two "continuity" theorems. Suppose $\phi_n = \mathfrak{F}(X_n)$ ($\mathfrak{F}(X_n)$ denotes the characteristic function of X_n) and $\phi_n(x^*) \to \phi(x^*)$. Clearly ϕ is positive definite and if ϕ is continuous at 0 on each finite dimensional subspace then there exists a weak distribution L on $\mathfrak{X}$ such that $\phi = \mathfrak{F}(L)$. The question naturally arises as to when there exists an X such that $\phi = \mathfrak{F}(X)$. We give two theorems which bear on this question and then two examples.

THEOREM 5. *Let $\mathfrak{X}$ be a real reflexive Banach space and let $\phi_n = \mathfrak{F}(X_n)$, if $\phi_n(x^*) \to \phi(x^*)$ then a necessary and sufficient condition that there exist an X such that $\phi = \mathfrak{F}(X)$ is that:*

(1) *ϕ restricted to any finite dimensional subspace of $\mathfrak{X}^*$ is continuous at 0.*

(2) *Given any separable subspace $\mathfrak{X}_1^*$ of $\mathfrak{X}^*$ and any $\varepsilon, \eta > 0$ there exists a δ such that for any finite collection $J = (x_1^*, \cdots, x_k^*) \in \mathfrak{X}_1^*$ with $||x_i^*|| \leqq \delta$ there exists $n(J, \delta)$ such that if $n > n(J, \delta)$ then*

$$P[\max_J |L_n(x_i^*)| \leqq \varepsilon] \geqq 1 - \eta \ . \tag{4.1}$$

Proof. Recall that $L_n(x^*) = \langle x^*, X_n(\omega) \rangle$. We now prove the sufficiency. Condition (1) implies that there exists a weak distribution L such that $\phi = \mathfrak{F}(L)$ and $L_n \to L$ in the sense that

$$F_k^{(n)}(x_1^*, \lambda_1; \cdots; x_k^*, \lambda_k) \to F_k(x_1^*, \lambda_1; \cdots; x_k^*, \lambda_k) \tag{4.2}$$

provided $(\lambda_1, \cdots, \lambda_k)$ is a continuity point of F_k. We show that L satisfies the conditions of Theorem 4. For convenience we put $\overline{F}(\varepsilon) = F(\varepsilon, \cdots, \varepsilon) - F(-\varepsilon, \cdots, -\varepsilon)$ for any distribution function F and we say ε is a continuity point provided $(\varepsilon, \cdots, \varepsilon)$ and $(-\varepsilon, \cdots, -\varepsilon)$ are continuity points of F.

Given $\mathfrak{X}_1^*$, ε, η, choose δ of Condition (2) corresponding to $\mathfrak{X}_1^*, \frac{\varepsilon}{2}, \frac{\eta}{2}$, Given any finite collection $J = (x_1^*, \cdots, x_k^*) \in \mathfrak{X}_1^*$ with $||x_i^*|| \leqq \delta$ we must show that

$$\int_{-\varepsilon}^{\varepsilon} \cdots \int_{-\varepsilon}^{\varepsilon} dF_k(x_1^*, \lambda_1; \cdots; x_k^*, \lambda_k) \geqq 1 - \eta \ .$$

Choose ε' such that ε' is a continuity point of F_k and $\frac{\varepsilon}{2} < \varepsilon' < \varepsilon$.

Choose n_0 such that $n_0 > n(J, \delta)$ and $|\overline{F}_k^{n_0}(\varepsilon') - \overline{F}_k(\varepsilon')| \leqq \frac{\eta}{4}$ then $\overline{F}_k(\varepsilon') \geqq \overline{F}_k^{n_0}(\varepsilon') - \frac{\eta}{4}$. Since $\varepsilon > \varepsilon'$ we have

$$\int_{-\varepsilon}^{\varepsilon}\cdots\int_{-\varepsilon}^{\varepsilon} dF_k(x_1^*, \lambda_1; \cdots; x_k^*, \lambda_k) \geqq \overline{F}_k(\varepsilon') \geqq \overline{F}_k^{n_0}(\varepsilon') - \frac{\eta}{4}$$

$$\geqq 1 - \frac{\eta}{2} - \frac{\eta}{4} \geqq 1 - \eta .$$

The necessity is proved by a similar computation.

THEOREM 6. *Let $\mathfrak{X}$ be a real separable reflexive Banach space and let $\phi_n = \mathfrak{F}(X_n)$, if $\phi_n(x^*) \to \phi(x^*)$ then a sufficient condition that there exist an X such that $\phi = \mathfrak{F}(X)$ is that:*

(1) *Condition* (1) *of Theorem* 5 *hold.*

(2) *If $G_n(a) = P[\|X_n\| \leqq a]$ then there exists a subsequence $G_{n_j}(a)$ converging to a distribution function $G(a)$.* ($\|X_n\|$ *is measurable since $\mathfrak{X}$ is separable.*)

Proof. In the same way as in the proof of Theorem 5 we have that $L_n \to L$ where $L = \mathfrak{F}(\phi)$ and $L_n(x^*) = \langle x^*, X_n\rangle$. Given $\varepsilon, \eta > 0$ choose $\delta > 0$ such that $\frac{\varepsilon}{2\delta}$ is a continuity point of G and $G\left(\frac{\varepsilon}{2\delta}\right) > 1 - \frac{\eta}{2}$. Choose N such that $n_j > N$ implies

$$\left| G_{n_j}\left(\frac{\varepsilon}{2\delta}\right) - G\left(\frac{\varepsilon}{2\delta}\right)\right| < \frac{\eta}{4} .$$

Now let $J = (x_1^*, \cdots, x_k^*)$ where $\|x_i^*\| \leqq \delta$, and let ε' be a continuity point of $F_k(x_1^*, \lambda_1; \cdots, x_k^*, \lambda_k)$ such that $\frac{\varepsilon}{2} < \varepsilon' < \varepsilon$. (We use same notation as in proof of Theorem 5.) Choose $n_j > N$ such that

$$|\overline{F}_k^{n_j}(\varepsilon') - \overline{F}_k(\varepsilon')| < \frac{\eta}{4} .$$

We now have

$$\overline{F}_k(\varepsilon) \geqq \overline{F}_k(\varepsilon') \geqq \overline{F}_k^{n_j}(\varepsilon') - \frac{\eta}{4} \geqq \overline{F}_k^{n_j}\left(\frac{\varepsilon}{2}\right) - \frac{\eta}{4} .$$

But

$$\overline{F}_k^{n_j}\left(\frac{\varepsilon}{2}\right) = P\left[\max_J |\langle x_i^*, X_{n_j}\rangle| \leqq \frac{\varepsilon}{2}\right] \geqq P\left[\delta \cdot \|X_{n_j}\| \leqq \frac{\varepsilon}{2}\right] = G_{n_j}\left(\frac{\varepsilon}{2\delta}\right) .$$

Therefore we finally obtain

$$\int_{-\varepsilon}^{\varepsilon}\cdots\int_{-\varepsilon}^{\varepsilon} dF_k(x_1^*, \lambda_1; \cdots; x_k^*, \lambda_k) = \overline{F}_k(\varepsilon) \geqq G\left(\frac{\varepsilon}{2\delta}\right) - \frac{\eta}{2}$$

$$\geqq 1 - \frac{\eta}{2} - \frac{\eta}{2} = 1 - \eta ,$$

and hence the proof of Theorem 6 is complete.

COROLLARY 1. *Condition* (2) *of Theorem* 6 *is implied by*

$$\lim_{a\to\infty} \inf_n G_n(a)=1 .$$

Proof. In this case every convergent subsequence (at least one exists by the Helly theorem) converges to a distribution function.

COROLLARY 2. (Mourier [4]). *The following condition implies* (2) *of Theorem* 6. *For some* $\alpha>0$ $E(\|X_n\|^\alpha)$ *exists for all* n *and* $E(\|X_n\|^\alpha)\leqq M$.

Proof. An immediate consequence of Corollary 1.

EXAMPLE. Let $\mathfrak{X}$ be a separable Hilbert space and let $\{e_n\}$ be a complete orthonormal system. Let Y_n be ordinary random variables mutually independent with normal distributions such that $E(Y_n)=0$, $E(Y_n^2)=\frac{1}{n}$. Define $X_n=\sum_1^n Y_k e_k$, clearly X_n is an $\mathfrak{X}$ valued random variable. Moreover (identifying $\mathfrak{X}^*$ with $\mathfrak{X}$), if $x=\sum_1^\infty \xi_k e_k$ then

$$\phi_n(x)=E\left\{e^{i\sum_{k=1}^n \xi_k Y_k}\right\}=e^{-1/2\sum_{k=1}^n \xi_k^2/k} .$$

But $\phi_n(x)\to\phi(x)=e^{-1/2\sum_{k=1}^\infty \xi_k^2/k}$ and the convergence is uniform if $\|x\|\leqq A$. Clearly $\phi(x)$ corresponds to the weak distribution $L(x)=\sum_1^\infty \xi_k Y_k$. However there is no $\mathfrak{X}$ valued random variable corresponding since $\sum_1^\infty Y_k e_k$ diverges with probability one. (This also follows from Theorem 4.) Thus uniform convergence of $\phi_n(x^*)\to\phi(x^*)$ on bounded sets is not sufficient to insure that ϕ corresponds to an $\mathfrak{X}$ valued random variable.

EXAMPLE. This example shows that condition (2) of Theorem 6 is not a necessary condition. Let $\mathfrak{X}$ and $\{e_n\}$ be as in Example 1. We define an ordinary random variable Y with the following distribution. $P[Y=\sqrt{n}]=P[Y^2=n]=\frac{1}{n!}(e^{-\lambda}\lambda^n)$ for $n=0,\cdots$. Clearly $E(Y)\leqq\lambda$. Let $Y_{n,k}$ be independent random variables each with the above distribution with parameter $\lambda_{n,k}$. Put

$$\lambda_{n,k}=\begin{cases}(1/n)^{3/2} & k\leqq n^2\\ 0 & k>n^2\end{cases} ; \quad \sum_{k=1}^\infty \lambda_{n,k}=\sum_1^{n^2}\lambda_{n,k}=\sqrt{n} .$$

Let $X_n = \sum_{k=1}^{\infty} Y_{n,k} e_k = \sum_{k=1}^{n^2} Y_{n,k} e_k$, then again X_n is an $\mathfrak{X}$ valued random variable. If $x = \sum_{k=1}^{\infty} \xi_k e_k$ then

$$E\{|(x, X_n)|\} \leq \sum_{k=1}^{n^2} |\xi_k| \lambda_{n,k} \leq \|x\| \cdot \left[\sum_{k=1}^{n^2} \lambda_{n,k}^2 \right]^{1/2} = \frac{1}{\sqrt{n}} \|x\| \to 0 .$$

Therefore $(x, X_n) \to 0$ in probability hence $(x, X_n) \to (x, X)$ in probability where $X \equiv 0$. Thus the weak distributions corresponding to X_n approach the weak distribution corresponding to X. However $\|x_n\|^2 = \sum_{k=1}^{n^2} Y_{n,k}^2$ where the $Y_{n,k}^2$ are independent Poisson variables with parameters $\lambda_{n,k}$. Thus the distributions of $\|x_n\|^2$ is Poisson with parameter $\sum_{k=1}^{n^2} \lambda_{n,k} = \sqrt{n}$ and clearly no subsequence converges to a distribution function.

References

1. S. Bochner, *Harmonic analysis and the theory of probability*, Berkeley and Los Angeles, 1955.

2. A. N. Kolmogorov, *Foundations of the theory of probability*, New York, 1950.

3. H. B. Mann, *On the realization of stochastic processes by probability distributions in function spaces*, Sankhyā, **11**, Part 1, (1951), 3-8.

4. E. Mourier, *Elements aléatoires dans un espace de Banach*, Ann. Inst. H. Poincaré **13** (1953), 161-244.

5. I. E. Segal, *Abstract probability spaces and theorem of Kolmogorov*, Amer. J. Math. **76** (1954), 721-732.

PRINCETON UNIVERSITY

SOME INEQUALITIES BETWEEN LATENT ROOTS AND MINIMAX (MAXIMIN) ELEMENTS OF REAL MATRICES

LOUIS GUTTMAN

1. Introduction. Because of the usual tediousness of computing latent roots, any quick information about them is often welcome and useful. We develop here some lower bounds to the absolute value of the major latent root (the one largest in absolute value) of any real symmetric matrix that depend only on a simple inspection of its elements. Also, lower bounds are developed for the largest latent root of a Gramian matrix of the form AA' that require inspection only of the elements of A. The latter case is especially important in linear regression theory of statistics, in factor analysis theories of psychology, and elsewhere.

The original motivation for our inequalities was to study the relationship between latent roots and the von Neumann value of a two-person zero-sum game matrix. We actually use the von Neumann theory to establish our bounds to latent roots, and in return we show how latent roots can be used to bound the game value of a matrix. The latter kind of bound will be useful whenever it is easier to get at the appropriate latent root than at the desired game value.

The bounds to latent roots are first exhibited in §§ 2–3, and then proved in § 4. How to reverse their emphasis to provide bounds for game values is shown in § 5.

2. Lower bounds to the major latent root. Let A be any real matrix of order $m \times n$. Let a_{ij} be the typical element of A ($i=1, 2, \cdots, m$; $j=1, 2, \cdots, n$), and let p_i and q_j be defined respectively as

$$(1) \qquad p_i = \min_j a_{ij}, \qquad q_j = \max_i a_{ij} \qquad \begin{pmatrix} i=1, 2, \cdots, m \\ j=1, 2, \cdots, n \end{pmatrix}.$$

Furthermore, let p and q be defined respectively as

$$(2) \qquad p = \max_i p_i, \qquad q = \min_j q_j.$$

From (1), it immediately follows that

$$(3) \qquad p_i \leqq a_{ij} \leqq q_j \qquad \begin{pmatrix} i=1, 2, \cdots, m \\ j=1, 2, \cdots, n \end{pmatrix},$$

Received February 20, 1956.

and in particular that $p \leqq q$.

Let λ^2 be the largest latent root of AA', where A' is the transpose of A. We shall prove in § 4 below that both of the following inequalities hold:

$$|\lambda| \geqq p\sqrt{n} \tag{4}$$

$$|\lambda| \geqq -q\sqrt{m}\ . \tag{5}$$

Inequality (4) is a useful lower bound to $|\lambda|$ if and only if $p > 0$, while (5) is useful if and only if $q < 0$. If $p \leqq 0 \leqq q$, we obtain no information about $|\lambda|$.

One interesting feature of (4) and (5) is that they show that λ^2 is generally at least of the order of m or of n, depending on whether $q < 0$ or $p > 0$.

Corresponding inequalities can be developed by considering A' in place of A. Let p'_j and q'_i be defined respectively as

$$p'_j = \min_i a_{ij}\ , \qquad q'_i = \max_j a_{ij} \qquad \begin{pmatrix} i=1, 2, \cdots, m \\ j=1, 2, \cdots, n \end{pmatrix}, \tag{6}$$

so that

$$p'_j \leqq a_{ij} \leqq q'_i \qquad \begin{pmatrix} i=1, 2, \cdots, m \\ j=1, 2, \cdots, n \end{pmatrix}. \tag{7}$$

Let p' and q' be defined by

$$p' = \max_j p'_j\ . \qquad q' = \min_i q'_i\ , \tag{8}$$

whence, from (7), $p' \leqq q'$.

Now, AA' and $A'A$ have the same nonzero latent roots, which are all positive. So if λ^2 is the largest latent root of AA', it is also the largest latent root of $A'A$. In addition to (4) and (5), we can write

$$|\lambda| \geqq p'\sqrt{m} \tag{9}$$

$$|\lambda| \geqq -q'\sqrt{n}\ . \tag{10}$$

Notice that the roles of m and n in (9) and (10) are reversed from those in (4) and (5). If $p' > 0$, λ^2 is at least of order m, while if $q' < 0$, λ^2 is at least of order n. If either of p or p' is positive, or if either of q or q' is negative, we get some information about $|\lambda|$.

Matrices of the form AA' or $A'A$ are called Gramian, or nonnegative definite symmetric. In statistics, any correlation matrix R is Gramian. A good deal of work in psychology, for example, is aimed at "factoring" an R into the form $R = AA'$. Given such a factoring, our inequalities immediately given lower bounds to the largest latent root of R from the minimax and maximin element of A. The latter are easily ascertainable by inspection.

3. **The case of symmetric matrices.** If $m \neq n$, A itself has no latent roots defined. However, if A is square, then it does have a characteristic equation and latent roots. A particularly important case is where A is symmetric, or $A = A'$. Then the latent roots of A are all real, and their squares are the latent roots of $AA' = A^2$. If λ^2 is the largest latent root of AA', then λ must be a root of A largest in absolute value, and conversely. In this symmetric case, we have not only $m = n$, but also $p = p'$, $q = q'$. So (9) and (4) are redundant, as are also (10) and (5). The inequalities can now be interpreted as referring to the major latent root of A itself, and not merely to a root of AA'.

When A is symmetric, we can usually improve on (4) and (5).

Let I be the unit matrix of order n, c be an arbitrary constant, and A^* be defined as

(11) $$A^* = A - cI .$$

If λ is a latent root of A, then $\lambda - c$ is a latent root of A^*, and conversely. Let p^* and q^* be the maximin and minimax of elements of A^* respectively, or, if δ_{ij} is Kronecker's delta,

(12) $$p^* = \max_i \min_j (a_{ij} - c\delta_{ij}) , \qquad q^* = \min_i \max_j (a_{ij} - c\delta_{ij}) .$$

Then in place of (4) and (5), we can write

(13) $$|\lambda - c| \geqq p^* \sqrt{n} , \qquad |\lambda - c| \geqq -q^* \sqrt{n} \qquad (A = A') ,$$

where $\lambda - c$ is the major latent root of A^*. In special cases, a judicious choice of c may be apparent that will make maximum $|\lambda - c|$ correspond to a λ which is either the most positive or the most negative latent root of A, and with a better bound than given by (4)-(5).

An especially important symmetric case is where A is a correlation matrix R, with all diagonal elements equal to unity. In such a case, the largest latent root of R cannot be less than 1, for the trace of R is n and all n latent roots are nonnegative. For this case, if $p > 0$, then choose $c = 1 - p$. This implies that the main diagonal elements of R^* are all equal to p. Then, clearly $p = p^*$; and since $\lambda \geqq 1$ for any R, $|\lambda - 1 + p| = \lambda - 1 + p$ when $p > 0$, and (13) becomes

(14) $$\lambda \geqq 1 + p(\sqrt{n} - 1) \qquad (p \geqq 0, \ A = R) .$$

Similarly, if $q < 0$, by choosing $c = 1 - q$ in (13) we get

(15) $$\lambda \geqq 1-q(\sqrt{n}-1) \qquad (q \leqq 0,\ A=R)\ .$$

4. Proof of the inequalities. Let P_k denote the space of all k-dimensional probability row vectors. That is $z \in P_k$ if and only if z is a row of k nonnegative numbers whose sum equals unity. Let z' denote the column vector that is the transpose of z. Then zz' is the sum of squares of the components of z, and it is easily established and well-known that

(16) $$\frac{1}{k} \leqq zz' \leqq 1 \qquad (z \in P_k)\ .$$

The equality on the left of (16) is always attained by letting $z=z_1$, where z_1 is a vector whose components all equal $1/k$ (and hence $z_1 \in P_k$).

von Neumann [1] has shown how each real matrix A has associated with it a unique real number v with certain important minimax properties. Since his theorem was developed in the context of his theory of games, we shall call v the *game value* of A. Our present interest of course is to regard von Neumann's theorem as a general theorem on real matrices, without necessary reference to the theory of games.

von Neumann's theorem is as follows. *If A is a real matrix of order $m \times n$, then there exist an x_0 and a y_0, where $x_0 \in P_m$ and $y_0 \in P_n$, and a unique real number v, such that*

(17) $$xAy_0' \leqq v \leqq x_0Ay' \qquad \textit{for all} \quad x \in P_m\ , \quad y \in P_n\ .$$

Furthermore,

(18) $$p \leqq v \leqq q\ ,$$

where p and q are as defined in (2).

To use this theorem for establishing our own inequalities, apply Schwarz's inequality to (17) to see that

(19) $$-\sqrt{(xx')(y_0A'Ay_0')} \leqq v \leqq \sqrt{(yy')(x_0AA'x_0')} \qquad (x \in P_m,\ y \in P_n)\ .$$

Let λ^2 be the largest latent root of AA' and $A'A$. Then

(20) $$x_0AA'x_0' \leqq \lambda^2 x_0x_0' \leqq \lambda^2\ , \qquad y_0A'Ay_0' \leqq \lambda^2 y_0y_0' \leqq \lambda^2\ ,$$

the second inequalities in each part of (20) following from the second inequality in (16). From the first inequality in (16),

(21) $$xx' \geqq \frac{1}{m}\ , \qquad yy' \geqq \frac{1}{n} \qquad (x \in P_m,\ y \in P_n)\ ,$$

and we have noted that the equalities in (21) are always attainable, by best possible x_1 and y_1 for this purpose. Using (20) and the equalities of (21) in (19) yield

$$\frac{-|\lambda|}{\sqrt{m}} \leqq v \leqq \frac{|\lambda|}{\sqrt{n}} . \tag{22}$$

Then (4) and (5) follow from (22) and (18). Inequalities (9) and (10) follow from the restatement of (22) for the game value v' of A':

$$\frac{-|\lambda|}{\sqrt{n}} \leqq v' \leqq \frac{|\lambda|}{\sqrt{m}} . \tag{23}$$

Inequalities (22) and (23) are of course sharper than those stated in § 2 above. If game values are known, they can be used in place of p, q, p', or q' in the latter inequalities. We have stated our inequalities in the form most practical to use, since p and q can be determined by inspection, whereas v usually cannot, except in the special case where $p=q=v$.

5. Application to game values. Let us now consider the converse problem of bounding game values. If an upper bound to $|\lambda|$ is known, this will serve to bound v and v' via (22) and (23). Thus, useful bounds to v can be set that may sometimes be better than (18) when $p \neq q$. Perhaps more important, (22) and (23) show how the magnitudes of v and v' compare with those of m and n in general, given some notion of the size of $|\lambda|$.

For the purpose of bounding v and v', (22) and (23) can be improved on. Let A_c be the $m \times n$ matrix whose typical element is $a_{ij}-c$, where c is an arbitrary constant. Thus A_c is obtained by subtracting c from *each* element of A (so $A_c \neq A^*$ if $c \neq 0$). It is easily verified that the game value of A_c is $v-c$, and optimal probability vectors x_0 and y_0 for A are optimal also for A_c. Let λ_c^2 be the largest latent root of $A_c A_c'$ (or of $A_c' A_c$). Then we can replace (22) and (23) by the more general inequalities

$$c - \frac{|\lambda_c|}{\sqrt{m}} \leqq v \leqq c + \frac{|\lambda_c|}{\sqrt{n}} \tag{24}$$

and

$$c - \frac{|\lambda_c|}{\sqrt{n}} \leqq v' \leqq c + \frac{|\lambda_c|}{\sqrt{m}} . \tag{25}$$

Evidently, the best choice of c is that which will minimize λ_c^2. A practical way to approximate this choice is to minimize instead the *sum of all the latent roots* of $A_c A_c'$, or the trace of $A_c A_c'$. This requires minimizing

$$\sum_{i=1}^{m}\sum_{j=1}^{n}(a_{ij}-c)^2 , \tag{26}$$

for which the minimizing value is $c=a$, where

$$\bar{a}=\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij} . \tag{27}$$

Reference

1. J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*, Princeton Univ. Press, 1954.

Center for Advanced Study in the Behavioral Sciences
(on leave from the Israel Institute of Applied Social Research)

THE NUMBER OF DISSIMILAR SUPERGRAPHS OF A LINEAR GRAPH

FRANK HARARY

1. Introduction. A (p, q) *graph* is one with p vertices and q lines. A formula is obtained for the number of dissimilar occurrences of a given (α, β) graph H as a subgraph of all (p, q) graphs G, $\alpha \leqq p$, $\beta \leqq q$, that is, for the number of dissimilar (p, q) supergraphs of H. The enumeration methods are those of Pólya [7]. This result is then applied to obtain formulas for the number of dissimilar complete subgraphs (cliques) and cycles among all (p, q) graphs. The formula for the number of rooted graphs in [2] is a special case of the number of dissimilar cliques. This note complements [3] in which the number of dissimilar (p, k) subgraphs of a given (p, q) graph is found. We conclude with a discussion of two unsolved problems.

A (*linear*) *graph* G (see [5] as a general reference) consists of a finite set V of *vertices* together with a prescribed subset W of the collection of all unordered pairs of distinct vertices. The members of W are called *lines* and two vertices v_1, v_2 are *adjacent* if $\{v_1, v_2\} \in W$, that is, if there is a line joining them. By the *complement* G' of a graph G, we mean the graph whose vertex-set coincides with that of G, in which two vertices are adjacent if and only if they are not adjacent in G.

Two graphs are *isomorphic* if there is a one-to-one adjacency-preserving correspondence between their vertex sets. An *automorphism* of G is an isomorphism of G with itself. The *group of a graph* G, written $\Gamma_0(G)$, is the group of all automorphisms of G. A *subgraph* G_1 of G is given by subsets $V_1 \subseteqq V$ and $W_1 \subseteqq W$ which in turn form a graph. If H is a subgraph of G, we also say G is a *supergraph* of H. Two subgraphs H_1, H_2 of G are *similar* if there is an automorphism of G which maps H_1 onto H_2. Obviously similarity is an equivalence relation and by the number of *dissimilar* vertices, lines, $\cdots$ of G, we mean the number of similarity classes (as in [3, 4, 6]).

Two supergraphs G_1 and G_2 of H are *H-similar* if there exists an isomorphism between G_1 and G_2 which leaves H invariant. It is clear that the number of dissimilar (p, q) supergraphs of H is equal to the number of dissimilar occurrences of H as a subgraph of all (p, q) graphs.

2. Supergraphs. Let H be an arbitrary (α, β) graph. We wish to enumerate the dissimilar (p, q) supergraphs of H where $p \geqq \alpha$, $q \geqq \beta$.

Received March 2, 1956. Presented to the American Mathematical Society August 31, 1955. Written while the author was a member of the Technical Staff of the Bell Telephone Laboratories, Murray Hill, New Jersey, Summer 1955.

Let $s^H_{p,q}$ be the number of dissimilar supergraphs of H with p vertices and q lines. For given p, let

$$(1) \qquad s^H_p(x) = \sum_{q=\beta}^{p(p-1)/2} s^H_{p,q} x^q$$

be the *counting polynomial* for the numbers $s^H_{p,q}$. We shall develop a formula for $s^H_p(x)$ using Pólya's enumeration theorem.

In precisely the form in which we require it, Pólya's Theorem is reviewed briefly in § 2 of [2]. Therefore, we shall not repeat here the definitions leading up to the statement of Pólya's Theorem, but shall only restate the theorem itself.

PÓLYA'S THEOREM. The configuration counting series $F(x)$ is obtained by substituting the figure counting series $\phi(x)$ into the cycle index $Z(\Gamma)$ of the configuration group Γ. Symbolically,

$$(2) \qquad F(x) = Z(\Gamma, \phi(x)) .$$

This theorem reduces the problem of finding the configuration counting series to the determination of the figure counting series and the cycle index of the configuration group.

The observations needed to make our problem amenable to Pólya's Theorem are as follows: A (p, q) supergraph G of the given (α, β) graph H is a configuration of length $p(p-1)/2-\beta$ whose figures are precisely those vertex-pairs of G not adjacent in H. The content of a figure is one if the vertices are adjacent and is zero otherwise, so that the figure counting series $\phi(x)=1+x$. Hence the content of the configuration G is $q-\beta$. The desired configuration series is $s^H_p(x)$.

In order to apply Pólya's Theorem, we still need to know the cycle index of the configuration group $\Gamma_{H,p}$. The degree of this group is $p(p-1)/2-\beta$ since the objects acted on by its permutations are the lines of the complement of H in the complete graph of p vertices containing H. All permutations of these lines which are compatible with $\Gamma_0(H)$ are in $\Gamma_{H,p}$. Before obtaining the cycle index of $\Gamma_{H,p}$ we state the form of the result by applying (2) to the present situation:

$$(3) \qquad s^H_p(x) = x^\beta Z(\Gamma_{H,p}, 1+x) .$$

We now turn to the development of the permutation group $\Gamma_{H,p}$ in a form which will yield its cycle index. Let F_p denote the complete graph of p vertices, that is, the graph with p vertices and all $p(p-1)/2$ possible lines. As in [3], let $\Gamma_1(G)$ be the *line-group* of the graph G, that is, the permutation group whose objects are the lines of G, and whose permutations are induced by those of $\Gamma_0(G)$, the group of automorphisms of G. If Γ is a permutation group of degree s, let $T(\Gamma)$

be the *pair-group* of Γ, that is, the permutation group of degree $s(s-1)/2$ which acts on the pairs of the object-set of Γ but is isomorphic to Γ as an abstract group. Then clearly $\Gamma_1(F_p)$ and $T(\Gamma_0(F_p))$ are isomorphic as permutation groups. Let $\Gamma_1 \cdot \Gamma_2$ denote the direct product of the permutation groups Γ_1 and Γ_2 whose object-sets are disjoint.

The lines of the object-set of the configuration group $\Gamma_{H,p}$ are of three possible kinds:

I. neither vertex is in H
II. both vertices are in H
III. one vertex is in H and the other is not.

For each of these three cases, we find the permutation group on the corresponding subset of lines and then form their direct product to get $\Gamma_{H,p}$. In case I, every rearrangement of the lines with neither vertex in H which is induced by a permutation of the vertices of $G-H$ is compatible with the group of H, so that we have the group $\Gamma_1(F_{p-\alpha})$. For case II, we obtain the line group of the complement of H, that is, $\Gamma_1(H')$. The third "mixed" case yields the group $M(H, F_{p-\alpha})$ of degree $\alpha(p-\alpha)$ on those lines of F_p joining a vertex of H with one of $F_{p-\alpha}$, consisting of those permutations of these lines which are compatible with $\Gamma_0(H)$. Then $\Gamma_{H,p}$ is the direct product:

$$\Gamma_{H,p} = \Gamma_1(F_{p-\alpha}) \cdot \Gamma_1(H') \cdot M(H, F_{p-\alpha}) \tag{4}$$

and by a remark of Pólya [7] to the effect that $Z(\Gamma_1 \cdot \Gamma_2) = Z(\Gamma_1) \cdot Z(\Gamma_2)$, we have

$$Z(\Gamma_{H,p}) = Z(\Gamma_1(F_{p-\alpha})) \cdot Z(\Gamma_1(H')) \cdot Z(M(H, F_{p-\alpha})) . \tag{5}$$

We note as a "dimensional check" that the degree of the groups of the right hand member of (4) are $(p-\alpha)(p-\alpha-1)/2$, $\alpha(\alpha-1)/2-\beta$, and $\alpha(p-\alpha)$ whose sum is $p(p-1)/2-\beta$, the degree of the configuration group.

Combining (5) and (3), we are now able to develop the counting polynomial for the dissimilar p vertex supergraphs of H. It is useful for this purpose to recall equation (10) of [2] which gives a formula for the first factor of (5). In this formula, which is equation (7) below, the letters g_i are employed for the indeterminates of the cycle index, S_p denotes the symmetric group of degree p, the sum is taken over all p-tuples (j) satisfying

$$1j_1 + 2j_2 + \cdots + pj_p = p , \tag{6}$$

and $d(q, r)$, $m(q, r)$ denote the greatest common divisor and least common multiple respectively.

$$Z(\Gamma_1(F_p))=Z(T(S_p))=\frac{1}{p!}\sum_{(j)}\frac{p}{\prod_{i=1}^{p} i^{j_i} j_i!} \tag{7}$$

$$\times \prod_{n=0}^{[(p-1)/2]} g_{2n+1}^{nj_{2n+1}+(2n+1)\binom{j_{2n+1}}{2}}$$

$$\times \prod_{n=1}^{[p/2]} (g_n g_{2n}^{n-1})^{j_{2n}}\, g_{2n}^{2n\binom{j_{2n}}{2}}$$

$$\times \prod_{1\leqq s<r\leqq p} g_{m(s,r)}^{j_s j_r d(s,r)}$$

Equation (7) gives the first factor of the right hand member of (5). The second factor depends on the particular graph H whose supergraphs are being enumerated. The third factor also depends on H, but can be readily computed as soon as $Z(\Gamma_0(H))$, the cycle index of the automorphism group of H, is found, by the following procedure. It is well known that for S_p, the symmetric group of degree p, one has

$$Z(S_p)=\frac{1}{p!}\sum_{(j)}\frac{1}{1^{j_1}j_1!\,2^{j_2}j_2!\cdots p^{j_p}j_p!}b_1^{j_1}b_2^{j_2}\cdots b_p^{j_p} \tag{8}$$

where the sum is taken over all partitions (j) of p satisfying (6) and the letters b_k are indeterminates. We write $Z(\Gamma_0(H))$ using the letters a_i as indeterminates, and then form the product $Z(\Gamma_0(H))\cdot Z(S_{p-\alpha})$. This will be a polynomial whose general term, aside from its numerical coefficient is of the form

$$(a_1^{h_1}a_2^{h_2}\cdots a_\alpha^{h_\alpha})(b_1^{j_1}b_2^{j_2}\cdots b_{p-\alpha}^{j_{p-\alpha}})=\prod_{s=1}^{\alpha} a_s^{h_s}\prod_{r=1}^{p-\alpha} b_r^{j_r} \tag{9}$$

If the letters c_k are the indeterminates of the third factor of (5), we then obtain $Z(M(H, F_{p-\alpha}))$ by substituting for (9) in $Z(\Gamma_0(H))\cdot Z(S_{p-\alpha})$ the expression:

$$\prod_{r,s} c_{m(r,s)}^{h_s j_r d(r,s)} \tag{10}$$

3. **Cliques.** We now specialize (5) to the case where H is a clique or complete graph, that is, to $H=F_\alpha$. For this to be meaningful, we define $Z(\Gamma_1(F_\alpha'))=1$, so that (5) becomes

$$Z(\Gamma_{F_\alpha,p})=Z(\Gamma_1(F_{p-\alpha}))\cdot Z(M(F_\alpha,\ F_{p-\alpha}))\,. \tag{11}$$

To illustrate (11), we take $p=4$, $\alpha=2$. Then the first factor is $Z(\Gamma_1(F_2))=c_1$ and the second factor is $Z(M(F_2,\ F_2))=\frac{1}{4}(c_1^4+3c_2^2)$. Therefore in thiscase, (3) yields the polynomial:

$$s_4^{F_2}(x)=x\cdot\frac{1}{4}(c_1^5+3c_1c_2^2,\ 1+x)$$

$$=x+2x^2+4x^3+4x^4+2x^5+x^6\,,$$

which can be readily verified pictorially by observing the number of dissimilar lines in all the graphs of 4 vertices: see Figure 1.

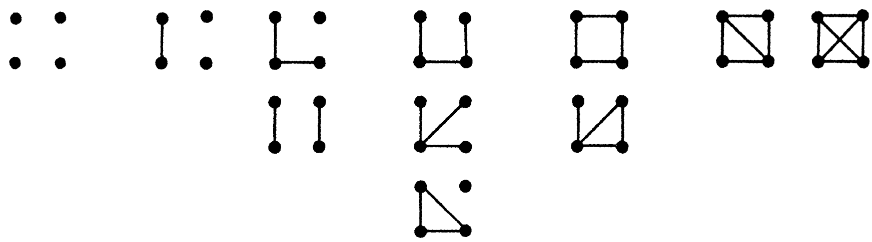

Figure 1

Equation (7) gives the first factor of (11) explicitly. One can also obtain an explicit formula for the second factor of (11) by applying (10) to two copies of (8) for the degrees α and $p-\alpha$. The result of this procedure is

$$(12)\qquad Z(M(F_\alpha,\ F_{p-\alpha}))=\frac{1}{\alpha!(p-\alpha)!}\sum_{(j)}\sum_{(h)}\frac{\alpha!}{\prod_{i=1}^{\alpha} i^{h_i}h_i!}\cdot\frac{(p-\alpha)!}{\prod_{i=1}^{p-\alpha} i^{j_i}j_i!}$$

$$\times\prod_{r=1}^{p-\alpha}\prod_{s=1}^{\alpha} c_{m(r,s)}^{h_s j_r d(r,s)}\,.$$

When (12) is specialized to $\alpha=1$, and then substituted into (3), the formula in [2] for the number of rooted graphs results.

4. **Cycles.** A *cycle of length n*, or an *n-cycle*, of a graph is a collection of n lines of the form A_1A_2, A_2A_3, $\cdots$, $A_{n-1}A_n$, A_nA_1 in which the vertices A_i are distinct. Let C_n be a graph consisting of an n-cycle. We now specialize (5) to the case $H=C_\alpha$. Since a 3-cycle is also a 3-clique, the particular case $\alpha=3$ for cycles has already been treated. In general, however, $\Gamma_0(C_n)\equiv D_n$, the dihedral group of degree n and order $2n$. From *Pólya* [7], we have

$$(13)\qquad Z(D_n)=\frac{1}{2n}\sum_{d\mid n}\phi(d)a_d^{n/d}+\begin{cases}\dfrac{1}{2}a_1a_2^{m-1}, & \text{when } n=2m-1\\[2ex] \dfrac{1}{4}(a_1^2a_2^{m-1}+a_2^m), & \text{when } n=2m\,.\end{cases}$$

When the cycle index of $Z(D_\alpha)$ is multiplied by $Z(S_{p-\alpha})$ from (8), and

(10) is applied, one obtains a formula for $Z(M(C_\alpha, F_{p-\alpha}))$ analogous to (12). Substituting $H=C_\alpha$ into (5), we see that

$$Z(\Gamma_{C_\alpha, p})=Z(\Gamma_1(F_{p-\alpha}))\cdot Z(\Gamma_1(C'_\alpha))\cdot Z(M(C_\alpha, F_{p-\alpha})) . \tag{14}$$

The only factor of the right-hand member of (14) for which we have not yet developed a formula is $Z(\Gamma_1(C'_\alpha))$.

To describe $Z(\Gamma_1(C'_\alpha))$, it is convenient to use a special case of the "Kranzgruppe" of Pólya [7]. Let Γ be any permutation group of degree d, and let E_n be the group of degree n and order 1. Then by $\Gamma[E_n]$, the *crown-group* of Γ around E_n, is meant the permutation group of degree nd obtained from Γ by replacing the d elements of the object-set acted on by the permutations belonging to Γ, by d disjoint sets of n elements each. Thus $Z(\Gamma[E_n])$ is obtained from $Z(\Gamma)$ when one replaces each factor $f_k^{j_k}$ occurring in each term of $Z(\Gamma)$ by $f_k^{nj_k}$.

For α odd, $\alpha=2n+1$, one sees that

$$\Gamma_1(C'_{2n+1})=D_{2n+1}[E_{n-1}] , \tag{15'}$$

from which $Z(\Gamma_1(C'_{2n+1}))$ is readily computed.

For α even, $\alpha=2n$ the group can be described using A. Cayley's[1] term "dimediation." For example the permutation group $\Gamma_1(C'_6)$ is generated by (123456)(789) and (12)(36)(45)(7)(89). Thus $\Gamma_1(C'_6)$ is isomorphic to D_6 as an abstract group, but as a permutation group it can be constructed from one copy of D_6 and two different copies of D_3. Abbreviating dimediation by "dim" following Cayley, we have in general

$$\Gamma_1(C'_{2n})=D_{2n}[E_{n-2}] \dim D_n . \tag{15''}$$

One can compute $Z(\Gamma_1(C'_{2n}))$ by multiplying each term of $Z(D_{2n}[E_{n-2}])$ by the appropriate term of $Z(D_n)$.

A *Hamilton cycle* of a graph is a cycle passing through all its vertices. Thus the number of dissimilar Hamilton cycles occurring in all (p, q) graphs is the number of dissimilar (p, q) supergraphs of C_p. In this situation, (14) becomes simplified to:

$$Z(\Gamma_{C_p, p})=Z(\Gamma_1(C'_p)) . \tag{16}$$

We illustrate (16) for $p=5$. Here (15') becomes $\Gamma_1(C'_5)=D_5[E_1]=D_5$, and by (13):

$$Z(D_5)=\frac{1}{10}(a_1^5+4a_5+5a_1a_2^2) ,$$

---

[1] See for example: A. Cayley, On the substitution groups for two, three, four, five, six, seven, and eight letters, Quart. J. Math. 25 (1890) especially p. 74.

so that applying (3), we get the counting polynomial for the number of dissimilar Hamilton cycles of length 5:

$$s_5^{C_5}(x) = x^5 + x^6 + 2x^7 + 2x^8 + x^9 + x^{10}\,.$$

This polynomial is verified by the graphs of Figure 2, in each of which the Hamilton cycle is drawn as the exterior cycle.

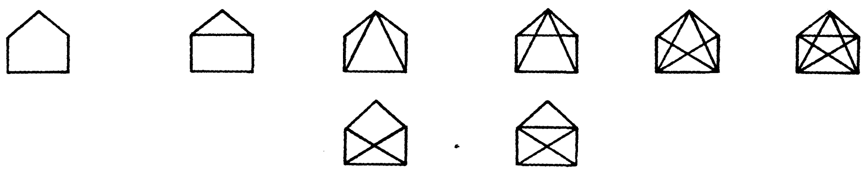

Figure 2.

For $p=5$, it turns out that each similarity type of Hamilton cycle occurs in a different graph; but this is not always so for larger p.

5. **Problems.** We discuss two unsolved problems implicit in [4] and [8] respectively.

I. It was shown in [4] that for any linear graph G; the dissimilarity characteristic equation:

$$v-(k-k_e)+(c-c_e)=1 \tag{17}$$

holds, where v, k, k_e denote the number of *dissimilar* vertices, lines, exceptional lines[2] respectively, and c, c_e denote the number of cycles, exceptional cycles respectively which appear in any dissimilarity cycle basis[3] of G. In the past, dissimilarity characteristic equations for trees and for Husimi trees [6] have proven useful in enumerating these kinds of graphs. The unsolved problem is to sum (17) over all (p, q) graphs, then multiply the resulting equation through by x^q and sum over $q=0$ to $\binom{p}{2}$. When this is done, the term 1 which is the right-hand member of (17) becomes $g_p(x)$, the counting polynomial for all p vertex graphs [2] and the term v clearly is manipulated into $G_p(x)$, the polynomial for p vertex rooted graphs [2]. By a result of Pólya [7], the enumeration of configurations in which all figures are distinct may be accomplished by using $Z(A_n)—Z(S_n)$, where A_n is the alternating group of degree n. But this is precisely the nature of the term $k—k_e$ of (17), which is the number of dissimilar lines of G whose vertices are not similar to each other. One sees by inspection from Figure 1 that for

[2] An *exceptional line* of a graph is one whose vertices are similar to each other.

[3] A *dissimilarity cycle basis* of a graph G is a minimal collection of cycles independent mod similarity on which all cycles of G depend mod similarity. Consult [**4**] for more details.

$p=4$, the counting polynomial induced by the term $k-k_e$ is

$$x^2+2x^3+2x^4+x^5\ .$$

To derive the general formula of which the preceding polynomial is the special case $p=4$, let us regard $\overline{F}_2$ as a line whose vertices are not similar. Then replacing F_α by $\overline{F}_2$ in (11), we get

$$Z(\Gamma_{F_2,\,p})=Z(\Gamma_1(F_{p-2}))\cdot Z(M(\overline{F}_2,\ F_{p-2})) \tag{18}$$

An explicit formula for the second factor is computed by noting that we may take $Z(\Gamma_0(\overline{F}_2))=Z(A_2)-Z(S_2)=\frac{1}{2}(a_1^2-a_2)$ by the above-mentioned result of Pólya, then multiplying this cycle index by (8) in which p is replaced by $p-2$, and applying (10).

The only term of (17) which we have been unable to sum is $c-c_e$. This appears to offer a nontrivial combinatorial problem, which if solved would provide a functional equation for $g_p(x)$ of the form

$$g_p(x)=G_p(x)-Z(\Gamma_{F_2,\,p},\ 1+x)+\text{the missing term.}$$

Using (14) for $\alpha=3, 4, \cdots, p$ one can enumerate *all* the dissimilar cycles among all (p, q) graphs, but this does not count just those in a dissimilarity cycle basis.

II. An *n-cube* can be described briefly as a graph whose vertices are the 2^n n-digit binary numbers in which 2 vertices are adjacent whenever they differ in exactly one place. An interesting unsolved problem with some potential applicability to switching theory is to determine the number h_n of dissimilar Hamilton cycles in an n-cube. It is well known that $h_2=h_3=1$ and it has been shown by E. N. Gilbert (unpublished) that $h_4=9$. From the formula of [3] one can find the number of dissimilar (p, p) subgraphs of any (p, q) graph, and of course, all the Hamilton cycles of the graph are included among these subgraphs. On the other hand, (16) gives a formula for the number of dissimilar Hamilton cycles occurring in all (p, q) graphs. However, each of these observations merely provides an upper bound for h_n and leaves the problem open. The more general problem of determining the number of dissimilar occurrences of a fixed graph H as a subgraph of a fixed graph G is also interesting.

One can give the results of this paper an interpretation in binary relations, following [1], and can also generalize them to directed graphs by employing the ordered-pair group of [2] instead of the pair group, but we shall not spell this out. We note finally that (3) implies that each such counting polynomial has end-symmetry with respect to its coefficients. This is explained geometrically by the one-to-one correspondence between the collection of all supergraphs G of H and the collection of their relative complements G'_H with respect to H defined by $G'_H = G' \cup H$.

REFERENCES

1. R. L. Davis, *The number of structures of finite relations*, Proc. Amer. Math. Soc., **4** (1953), 486–495.

2. F. Harary, *The number of linear, directed, rooted, and connected graphs*, Trans. Amer. Math. Soc., **78** (1955), 445–463.

3. ———, *The number of dissimilar line-subgraphs of a given graph*, Pacific, J. Math. **6** (1956), 57–64.

4. F. Harary and R. Z. Norman, *The dissimilarity characteristic of linear graphs*, Proc. Amer. Math. Soc., **5** (1954), 131–135.

5. D. König, *Theorie der endlichen und unendlichen Graphen*, Leipzig, 1936; reprinted New York, 1950.

6. R. Z. Norman, *On the number of linear graphs with given blocks*, Doctoral dissertation, University of Michigan, 1954 (in process of publication).

7. G. Pólya, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen, und chemische Verbindungen*, Acta Math. **68** (1937), 145–254.

8. D. Slepian, *On the number of symmetry types of Boolean functions of n variables*, Can. J. Math. **5** (1953), 185–193.

UNIVERSITY OF MICHIGAN

STRUCTURE THEORY FOR A CLASS OF CONVOLUTION ALGEBRAS

EDWIN HEWITT AND HERBERT S. ZUCKERMAN

Introduction. This paper is a chapter in the study of convolution algebras begun in [7]. The algebras studied here are algebras of Borel measures on certain compact semigroups, and we describe completely the structure of these algebras. The solution obtained seems remarkable in view of the extreme complexity of the corresponding measure algebras for compact Abelian groups (see [12]). Our success is explained by the simple algebraic structure of the semigroups we deal with.

In addition to the structure theory (§§ 2–6), we give an application to probability (§ 7), and some concrete examples and illustrations (§ 8).

Throughout this paper, we use the notation and terminology of [7]. In particular, the reader should be familiar with § 1 of [7]. The related papers [6] and [8] are not essential for understanding the present paper, but are referred to occasionally here at points of contact in subject-matter. For all measure-theoretic terms and techniques not explained here, see [4]. References are made throughout the present paper to [9] for topological matters, and to [10] for the elementary theory of Banach algebras. We use K to denote the complex number system. All other special symbols will be explained as they appear.

1. The semigroups to be studied.

1.1. We consider an arbitrary non-void set G, completely ordered by a transitive, irreflexive relation "$<$". That is, for all $x, y \in G$, exactly one of the relations $x<y$, $x=y$, $y<x$ obtains, and the relations $x<y$ and $y<z$ imply $x<z$. As usual, we write $y>x$, meaning $x<y$, and we write $x \leqq y$, meaning $x<y$ or $x=y$. For $u, v \in G$, we define

$$]u, v[=\{x: x \in G,\ u<x<v\} \qquad \text{(open interval)}\ ,$$

$$[u, v[=\{x: x \in G,\ u \leqq x<v\} \qquad \text{(half-open interval)}\ ,$$

$$]u, v]=\{x: x \in G,\ u<x \leqq v\} \qquad \text{(half-open interval)}\ ,$$

$$[u, v]=\{x: x \in G,\ u \leqq x \leqq v\} \qquad \text{(closed interval)}\ .$$

These sets may or may not be void, depending upon the relation between u and v.

Received January 17, 1956. The first-named author is a fellow of the John Simon Guggenheim Memorial Foundation. Research also supported in part by the National Science Foundation.

1.2. We make G into a semigroup by defining the product xy as $\max(x, y)$ for all $x, y \in G$. It is obvious that $x(yz)=(xy)z$ for all $x, y, z \in G$, that $xy=yx$ for all $x, y \in G$, and that $x^2=x$ for all $x \in G$.

1.3. Being a completely ordered set, G has a natural topology defined in terms of the ordering. For all $a \in G$ and all $u, v \in G$ such that $u < a < v$, the open interval $]u, v[$ is taken as an open neighborhood of the point a. If there is no u such that $u < a$ (*i.e.*, if a is the first element of G), then $[a, v[$ with $v > a$ is a neighborhood of a, and analogously if a is the last element of G. These are all of the open neighborhoods of a. It is obvious that Hausdorff's neighborhood axioms are satisfied and that Hausdorff's separation axiom is satisfied. A point a in G is isolated if and only if it has an immediate predecessor and an immediate successor. It has a complete neighborhood system consisting of intervals $[a, v[$ ($]u, a]$) if and only if it has an immediate predecessor (an immediate successor).

It is easy to verify that the semigroup operation $xy = \max(x, y)$ is continuous in both x and y for the topology described above. Hence G is a topological semigroup satisfying the Hausdorff separation axiom.

1.4. We impose the additional restriction on G that it be compact in the interval topology[1]. For this, it is both necessary and sufficient that every subset of G admit a least upper bound and a greatest lower bound. In particular, G has a least element, which we shall call α, and a greatest element, which we shall call ω (not to be confused with the ordinal number ω). For a sketch of the proof of this, see [9], p. 162, exercise C.

1.5. From now on, we shall suppose, save where the contrary is explicitly stated, that G is a completely ordered set that is compact in the interval topology, and made into a topological semigroup by the operation $\max(x, y)$.

1.6. Let $\mathfrak{C}(G)$ denote the linear space of all complex-valued continuous functions on G. We give $\mathfrak{C}(G)$ the usual norm:

$$\|f\| = \max_{x \in G} |f(x)|$$

for $f \in \mathfrak{C}(G)$. Let $\tilde{\mathfrak{C}}(G)$ denote the conjugate space of $\mathfrak{C}(G)$, that is, the linear space of all complex-valued linear functionals L on $\mathfrak{C}(G)$ such that the number

$$\|L\| = \sup\{|L(f)| : f \in \mathfrak{C}(G),\ \|f\| \leqq 1\}$$

is finite. It is well known that each $L \in \tilde{\mathfrak{C}}(G)$ has a unique representation

[1] See however 8.5.

as an integral with respect to a complex-valued, countably additive, regular measure λ defined on all Borel subsets of G (see [4], pp. 247–248). That is,

$$L(f)=\int_G f(x)d\lambda(x) \tag{1.6.1}$$

for all $f\in\mathfrak{C}(G)$. While many authors have contributed to this theorem, we call it for convenience the Riesz representation theorem. Elements of $\tilde{\mathfrak{C}}(G)$ will be denoted by capital Roman letters, $L, M, \cdots$, and the corresponding measures of the kind referred to will be denoted by the corresponding lower-case Greek letters $\lambda, \mu, \cdots$. Under our interpretation of the term "measure," the measures $\lambda, \mu, \cdots$ are set-functitns and not linear functionals (for a different point of view, consult [2], *passim*). However, we shall allow ourselves the abuse of notation $\lambda\in\tilde{\mathfrak{C}}(G)$, meaning that λ is connected with an element $L\in\tilde{\mathfrak{C}}(G)$ by the relation 1.6.1.

At various points in our discussion, it will be necessary to pass from an element $L\in\tilde{\mathfrak{C}}(G)$ to the corresponding measure λ. For nonnegative L (that is, $L(f)\geqq 0$ for f real and nonnegative), this process is simple. Let F be any closed subset of G. Then

$$\lambda(F)=\inf\{L(f):\ f\in\mathfrak{C}(G),\ f(x)\geqq 1 \text{ for } x\in F,\ f(x)\geqq 0 \text{ for } x\in G\}\ . \tag{1.6.2}$$

Let H be any open subset of G. Then

$$\lambda(H)=\sup\{\lambda(F):\ F \text{ is closed, } F\subset H\}\ . \tag{1.6.3}$$

Let X be any subset of G. Then

$$\lambda(X)=\inf\{\lambda(H):\ H \text{ is open, } H\supset X\}\ . \tag{1.6.4}$$

These three definitions of λ, on various families of sets, are all consistent, and λ is an outer measure on all subsets of G. Every Borel set is λ-measurable, λ is regular, and 1.6.1 holds.

For an arbitrary $L\in\tilde{\mathfrak{C}}(G)$, we obtain the corresponding measure λ by writing L as

$$L=L_1-L_2+i(L_3-L_4)\ , \tag{1.6.5}$$

where $L_1, \cdots, L_4$ are non-negative functionals on $\mathfrak{C}(G)$.

1.7. We recall that a semicharacter of a semigroup H is a bounded complex-valued function χ on H, not identically zero, satisfying the functional equation $\chi(xy)=\chi(x)\chi(y)$ for all $x, y\in H$ ([7], 3.1 and [8], 1.3). Semicharacters of our semigroup G play a vital rôle in the solution of the present problem, and we proceed to identify the semicharacters of G.

1.8. THEOREM. *Let G be as specified in* 1.5. *Then functions of the following two types are semicharacters of G:*

(1.8.1) *functions $\psi_{a]}$, where $a \in G$ and*

$$\psi_{a]}(x)=\begin{cases} 1 \text{ if } x \leqq a\ , \\ 0 \text{ if } x > a\ ; \end{cases}$$

(1.8.2) *functions $\psi_{a[}$, where $a \in]\alpha, \omega]$ and*

$$\psi_{a[}(x)=\begin{cases} 1 \text{ if } x < a\ , \\ 0 \text{ if } x \geqq a\ . \end{cases}$$

Furthermore, every semicharacter χ of G is one of these two types.

Proof. It is easy to see that all functions $\psi_{a]}(a \in G)$ and $\psi_{a[}(\alpha < a \leqq \omega)$ are semicharacters, and we omit the verification. To establish the converse, let χ be a semicharacter of G. Since $x^2=x$ for all $x \in G$, χ assumes no values other than 0 and 1. If χ is identically 1 (in this case we write $\chi=1$), then $\chi=\psi_{\omega]}$. If $\chi \neq 1$, then there exist a and b such that $\chi(a)=1$, $\chi(b)=0$. Let $A=\{x; x \in G, \chi(x)=1\}$, $B=\{x; x \in G, \chi(x)=0\}$. If $x \in A$ and $x' < x$, then we have $1=\chi(x)=\chi(x'x)=\chi(x')\chi(x)=\chi(x')$. If $x \in B$ and $x' > x$, then we have $\chi(x')=\chi(x'x)=\chi(x')\chi(x)=0$. The sets A and B are therefore non-void complementary sets forming a Dedekind cut in G. Since G is compact, A has a least upper bound a. If $a \in A$, we have $\chi=\psi_{a]}$; if $a \in B$, we have $\chi=\psi_{a[}$.

1.9. THEOREM. *Let G be as specified in* 1.5. *Suppose first that $\alpha \leqq a < \omega$. Then the function $\psi_{a]}$* (1.8.1) *is continuous if and only if a has an immediate successor. The function $\psi_{\omega]}$ is trivially continuous. Suppose next that $\alpha < a \leqq \omega$. Then the function $\psi_{a[}$ is continuous if and only if a has an immediate predecessor a_-, and in this case, $\psi_{a[}=\psi_{a_-]}$.*

We omit the proof of this theorem.

1.10. THEOREM. *The semigroup G admits a continuous semicharacter different from* 1 *if and only if G is disconnected.*

Proof. Since a semicharacter of G can assume only the values 0 and 1, the necessity of the condition is obvious. Conversely, suppose that G is disconnected, and that P and Q are non-void complementary open sets in G. Since $\sup P \in P$ and $\sup Q \in Q$ (P and Q being closed), we may suppose without loss of generality that $\sup P < \omega$. Let $B=\{x; x \in G, x > \sup P\}$. If $\sup P=\inf B$, then every open interval containing $\sup P$ contains points of B, and $B \subset Q$. Since P is open, this cannot occur. Hence $\sup P < \inf B$, and the function $\psi_{\sup P]}$ is a continuous nonconstant semicharacter.

1.11. REMARK. If G is *not* compact in its interval topology, then semicharacters of the types 1, $\psi_{a]}$, and $\psi_{a[}$ may or may not exhaust the class of all semicharacters. If G admits a Dedekind cut $\{A, B\}$ (where A is the lower class) and where A has no supremum and B no infimum, then the function ψ_A equal to 1 on A and 0 on B is a semicharacter different from 1, $\psi_{a]}$, and $\psi_{a[}$ for any a. The proof of 1.8 shows that the existence of such a Dedekind cut is also necessary for the existence of a semicharacter different from 1 and all $\psi_{a]}$ and $\psi_{a[}$.

1.12. THEOREM. *Let G be as in 1.5. Let $\hat{G}$ denote the set of all semicharacters of G. Then $\hat{G}$ is a semigroup under pointwise multiplication.*

Proof. If χ_1 and χ_2 are semicharacters, then the product $\chi_1\chi_2$ ($\chi_1\chi_2(x)=\chi_1(x)\chi_2(x)$ for $x \in G$) is obviously either 0 or a semicharacter. Since $\chi_1(\alpha)=\chi_2(\alpha)=1$, we cannot have $\chi_1\chi_2=0$.

2. The convolution algebra $\tilde{\mathfrak{C}}(G)$. In a previous paper, we have introduced the general notion of a convolution algebra ([7], p. 69, 1.3). We shall show here that $\tilde{\mathfrak{C}}(G)$ is a convolution algebra, where $\mathfrak{F}=\mathfrak{C}(G)$.

2.1. THEOREM. *Let $x \in G$ and let $f \in \mathfrak{C}(G)$. Then the function ${}_xf$ whose value at $y \in G$ is $f(xy)=f(\max(x, y))$ is continuous.*

Proof. This assertion follows immediately from the fact that

$${}_xf(y)=\begin{cases} f(x) \text{ for } y \leqq x \,, \\ f(y) \text{ for } y > x \,. \end{cases}$$

2.2. THEOREM. *Let $f \in \mathfrak{C}(G)$ and $L \in \tilde{\mathfrak{C}}(G)$. Then the function on G whose value at $x \in G$ is $L({}_xf)$ is continuous [we also write $L({}_xf)$ as $L_y(f(xy))$].*

Proof. Let $u, v \in G$ and suppose that $u \leqq v$. Then we have

$$(2.2.1) \qquad {}_vf(y)-{}_uf(y)=\begin{cases} f(v)-f(u) & \text{if} \quad \alpha \leqq y \leqq u \,, \\ f(v)-f(y) & \text{if} \quad u \leqq y \leqq v, \,, \\ 0 & \text{if} \quad v \leqq y \,. \end{cases}$$

Now let ε be a positive real number, and let x be an arbitrary element of G. Since f is continuous, there exist $a, b \in G$ such that $a < x < b$

(we omit the obvious changes needed when $x=\alpha$ or $x=\omega$) and such that $|f(s)-f(t)|<\varepsilon\cdot\|L\|^{-1}$ for all $s, t\in]a, b[$. It follows from 2.2.1 that $|_{x'}f(y)-{}_xf(y)|<\varepsilon\cdot\|L\|^{-1}$ for all $x'\in]a, b[$ and all $y\in G$. Hence we have

$$|L(_{x'}f)-L(_xf)|\leqq\|L\|\cdot\|_{x'}f-{}_xf\|<\varepsilon .$$

This completes the proof.

2.3. THEOREM. *Let L and M be elements of* $\tilde{\mathfrak{C}}(G)$. *For all* $f\in\mathfrak{C}(G)$, *let* $L*M(f)$ *be the value assumed by the functional L for the function whose value at x is* $M(_xf)$. *We write*

$$L*M(f)=L_x(M_y(f(xy))) . \tag{2.3.1}$$

Then $L*M\in\tilde{\mathfrak{C}}(G)$, *and*

$$\|L*M\|\leqq\|L\|\cdot\|M\| . \tag{2.3.2}$$

Proof. Theorem 2.2 shows that the right side of 2.3.1 has meaning. Now for all $x, y\in G$, we have $|f(xy)|\leqq\|f\|$, and hence $\|_xf\|\leqq\|f\|$. Therefore $|M_yf(xy)|\leqq\|M\|\cdot\|f\|$, and in turn $|L*M(f)|\leqq\|L\|\cdot\|M\|\cdot\|f\|$. This proves that $L*M$ is a bounded functional, and since $L*M$ is obviously linear, 2.3.2 and the present theorem follow.

2.4. REMARK. Theorems 2.1, 2.2, and 2.3 are verifications of [7] 1.3.1, 1.3.2, and 1.3.3, respectively. Therefore we have proved that $\tilde{\mathfrak{C}}(G)$ is a convolution algebra with the convolution $L*M$ of 2.3.

2.5. THEOREM. *Let L, M be elements of* $\tilde{\mathfrak{C}}(G)$ *and let* λ, μ *be the corresponding measures as in* 1.6. *Then we have*

$$L*M(f)=\int_G\int_G f(\max(x, y))d\mu(y)d\lambda(x) , \tag{2.5.1}$$

for all $f\in\mathfrak{C}(G)$.

Proof. The right side of 2.5.1 simply rewrites the right side of 2.3.1, making use of 1.6.1.

We shall write $\lambda*\mu$ to denote the measure associated with $L*M$ by 1.6.1.

2.6. THEOREM. *The algebra* $\tilde{\mathfrak{C}}(G)$ *is associative and commutative.*

Proof. Associativity is a property of all convolution algebras ([7], p. 73, Theorem 1.5). Commutativity follows immediately from Fubini's theorem (which applies since all measures under consideration are finite and countably additive) and 2.5.1 :

$$L*M(f)=\int_G\int_G f(\max(x, y))d\mu(y)d\lambda(x)=\int_G\int_G f(\max(x, y))d\lambda(x)d\mu(y)$$
$$=\int_G\int_G f(\max(y, x))d\lambda(x)d\mu(y)=M*L(f) .$$

2.7. To identify the unit in $\tilde{\mathfrak{C}}(G)$, and also for certain future purposes, we introduce a class of special linear functionals $E_a(a \in G)$:

(2.7.1) $$E_a(f)=f(a) \quad \text{for} \quad f \in \mathfrak{C}(G) .$$

It is clear that $\|t_1E_{a_1}+\cdots+t_sE_{a_s}\|=\sum_{j=1}^{s}|t_j|$ for all complex numbers $t_1, \cdots, t_s$ and distinct $a_1, \cdots, a_s$ in G. It is also clear that the measure ε_a corresponding to E_a is the unit mass at a:

(2.7.2) $$\varepsilon_a(X)=\begin{cases} 1 & \text{if } a \in X , \\ 0 & \text{if } a \notin X , \end{cases}$$

for all $X \subset G$.

2.8. For all $\lambda \in \tilde{\mathfrak{C}}(G)$ and every Borel set A in G, let λ^A be the measure such that $\lambda^A(X)=\lambda(A \cap X)$ for all Borel sets $X \subset G$.

2.9. THEOREM. *For all* $\lambda \in \tilde{\mathfrak{C}}(G)$ *and all* $a \in G$, *we have*

(2.9.1) $$\varepsilon_a*\lambda=\lambda([\alpha, a])\varepsilon_a+\lambda^{]a,\omega]}$$

and

(2.9.2) $$\varepsilon_a*\lambda=\lambda([\alpha, a[)\varepsilon_a+\lambda^{[a,\omega]} .$$

Proof. The set $[\alpha, a]$ being a closed subset of G, it is certainly a Borel set (although not necessarily a Baire set), and hence $\lambda([\alpha, a])$ is defined. Similarly, $]a, \omega]$ (which is void if $a=\omega$) is a Borel set, so that $\lambda^{]a,\omega]}$ is defined. Hence the right side of 2.9.1 is defined.

Consider the integral $I(x)=\int_G f(\max(x, y))d\varepsilon_a(y)$, where $f \in \mathfrak{C}(G)$. The integrand has the constant value $f(x)$ for $y \in [\alpha, x]$, and is equal to f in the interval $]x, \omega]$. Therefore if $x \leqq a$, then $I(x)=f(a)$. If $x > a$, then $I(x)=f(x)$. It follows that

(2.9.3) $$L*E_a(f)=\int_G I(x)d\lambda(x)=\int_{[\alpha,a]} f(a)\cdot 1 d\lambda(x) + \int_{]a,\omega]} f(x)d\lambda(x)$$
$$=\lambda([\alpha, a])E_a(f)+\int_G f(x)d\lambda^{]a,\omega]}(x) .$$

The relations 2.9.3 imply 2.9.1 immediately, and 2.9.2 is a trivial consequence of 2.9.1.

2.10. THEOREM. *For all* $\lambda \in \tilde{\mathfrak{C}}(G)$, *we have* $\varepsilon_\alpha * \lambda = \lambda$. *That is,* E_α *is the unit of* $\tilde{\mathfrak{C}}(G)$.

Proof. Putting $a = \alpha$ in 2.9.1, aud taking an arbitrary Borel set $X \subset G$, we have

$$\varepsilon_\alpha * \lambda(X) = \lambda(\{\alpha\}) \cdot \varepsilon_\alpha(X) + \lambda(]\alpha, \omega] \cap X) . \tag{2.10.1}$$

If $\alpha \notin X$, then $\varepsilon_\alpha(X) = 0$ and $]\alpha, \omega] \cap X = X$. Hence $\varepsilon_\alpha * \lambda(X) = \lambda(X)$ in this case. If $\alpha \in X$, the right side of 2.10.1 is equal to

$$\lambda(\{\alpha\} \cap X) + \lambda(]\alpha, \omega] \cap X) = \lambda((\{\alpha\} \cap X) \cup (]\alpha, \omega] \cap X)) = \lambda(X) .$$

Therefore $\varepsilon_\alpha * \lambda(X) = \lambda(X)$ in all cases, and $\varepsilon_\alpha * \lambda = \lambda$.

2.11. THEOREM. *For all* $L \in \tilde{\mathfrak{C}}(G)$, *we have* $E_\omega * L = L(1) E_\omega$. *In terms of measures, we have* $\varepsilon_\omega * \lambda = \lambda([\alpha, \omega]) \varepsilon_\omega$.

Proof. The set $]\omega, \omega]$ is void, and so, putting $a = \omega$ in 2.9.1, we get $\lambda^{]\omega, \omega]} = 0$ and $\varepsilon_\omega * \lambda = \lambda([\alpha, \omega]) \varepsilon_\omega$. The first statement is obviously equivalent to this.

2.12. THEOREM. *For* $a, b \in G$, *we have* $\varepsilon_a * \varepsilon_b = \varepsilon_{\max(a,b)}$.

Proof. This too follows at once from 2.9.1.

We summarize 2.3, 2.6, and 2.10 as follows.

2.13. THEOREM. *Under the convolution* 2.3.1, $\tilde{\mathfrak{C}}(G)$ *is a commutative Banach algebra with unit.*

3. The maximal ideals of $\tilde{\mathfrak{C}}(G)$.

In this section, we identify *all* of the maximal ideals in $\tilde{\mathfrak{C}}(G)$. Since $\tilde{\mathfrak{C}}(G)$ is a commutative Banach algebra with unit, every maximal ideal in $\tilde{\mathfrak{C}}(G)$ is closed and regular, and we may identify the class of maximal ideals in $\tilde{\mathfrak{C}}(G)$ with the class of all (algebra) homomorphisms of $\tilde{\mathfrak{C}}(G)$ onto K. For a discussion of Gel'fand's theory of commutative Banach algebras, see [10], pp. 66–81.

3.1. An obvious source of homomorphisms of $\tilde{\mathfrak{C}}(G)$ onto K is the set of all *continuous* semicharacters on G. If χ is such a semicharacter, then $L(\chi)$ is defined for all $L \in \tilde{\mathfrak{C}}(G)$ and the mapping

$$L \rightarrow L(\chi) \tag{3.1.1}$$

is obviously a linear functional on $\tilde{\mathfrak{C}}(G)$. If $L, M \in \tilde{\mathfrak{C}}(G)$, then

$$\begin{aligned} L*M(\chi) &= \int_G \int_G \chi(xy) d\mu(y) d\lambda(x) = \int_G \int_G \chi(x)\chi(y) d\mu(y) d\lambda(x) \\ &= \left[\int_G \chi(x) d\lambda(x)\right]\left[\int_G \chi(y)\, d\mu(y)\right] = L(\chi)M(\chi) . \end{aligned} \tag{3.1.2}$$

Hence the mapping 3.1.1 is multiplicative, that is, it is a homomorphism of $\tilde{\mathfrak{C}}(G)$ onto K.

However, as 1.8 and 1.10 show, G may have very few continuous semicharacters. Indeed, it can be shown that there exist mappings of the form 3.1.1 carrying an arbitrary $L \neq 0$ into a non-zero number if and only if G has Urysohn dimension zero. (We shall go no further into this minor point.) Therefore, if we have any hope of proving $\tilde{\mathfrak{C}}(G)$ semisimple, we must look further for homomorphisms of $\tilde{\mathfrak{C}}(G)$ onto K. Our construction hinges on the fact that while the functions $\psi_{a]}$ and $\psi_{a[}$ are often discontinuous, still they are Borel measurable and bounded. Therefore they are λ-integrable for all $\lambda \in \tilde{\mathfrak{C}}(G)$ even though $L(\psi_{a]})$ and $L(\psi_{a[})$ may be undefined *ab initio*. The Riesz representation theorem gives us a canonical method of extending L from $\mathfrak{C}(G)$ to the space of all bounded Borel measurable functions on G, and it is just this fact that we use.

3.2. THEOREM. *Let $a \in G$. Then the mapping*

$$L \rightarrow \lambda([\alpha, a]) = \int_G \psi_{a]}(x) d\lambda(x) \qquad (L \in \tilde{\mathfrak{C}}(G)) \tag{3.2.1}$$

is a homomorphism of $\tilde{\mathfrak{C}}(G)$ onto K. Let $a \in\,]\alpha, \omega]$. Then the mapping

$$L \rightarrow \lambda([\alpha, a[) = \int_G \psi_{a[}(x) d\lambda(x) \qquad (L \in \tilde{\mathfrak{C}}(G)) \tag{3.2.2}$$

is a homomorphism of $\tilde{\mathfrak{C}}(G)$ onto K.

Proof. First of all, it is clear that the mappings 3.2.1 and 3.2.2 are linear and not identically zero on $\tilde{\mathfrak{C}}(G)$. Our only task is to show that they are multiplicative. To this end, we consider first the mappings 3.2.1. If $a = \omega$, then we are dealing with the continuous semicharacter 1, and this case has already been treated in 3.1. We may therefore suppose that $a < \omega$. If a has an immediate successor a_+, then the interval $[\alpha, a]$ is open and closed, and the function $\psi_{a]}$ is a continuous semicharacter. Once again we can refer to 3.1. The remaining case is that in which $a < \omega$ and a has no immediate successor. In this case, the interval $]a, u[$ is non-void for every $u > a$, the semicharacter $\psi_{a]}$ is discontinuous, and a more detailed examination is needed.

It is convenient first to treat the case of non-negative, non-zero linear functionals L and M. It is obvious that if L and M are non-negative, then $L*M$ is non-negative. The set $[\alpha, a]$ being compact, we have $\lambda*\mu([\alpha, a]) = \inf L*M(f)$, the infimum being taken over all $f \in \mathfrak{C}(G)$ such that $f \geqq \psi_{a]}$ (see 1.6.2). Since the measure λ is regular, we have

$$\lambda([\alpha, a]) = \inf \{\lambda(T) : T \text{ is open, } T \supset [\alpha, a]\} .$$

Every such set T contains an interval $[\alpha, u[$, where $u > a$, and hence

$$\lambda([\alpha, a]) = \inf \{\lambda([\alpha, u[) : a < u \leqq \omega\} .$$

Similarly, we see that

$$\mu([\alpha, a]) = \inf \{\mu([\alpha, u[) : a < u \leqq \omega\} .$$

Now let ε be any positive real number. Since λ and μ are additive measures, the preceding two sentences show that there exists an element $u \in]a, \omega]$ for which the following inequalities hold:

$$\lambda(]a, u[) < \min\left(\frac{\varepsilon}{3M(1)}, \frac{\varepsilon}{3}\right), \tag{3.2.3}$$

$$\mu(]a, u[) < \min\left(\frac{\varepsilon}{3L(1)}, 1\right). \tag{3.2.4}$$

Since G is normal, there exists $f \in \mathfrak{C}(G)$ such that $f(x) = 1$ for $x \in [\alpha, a]$, $f(x) = 0$ for $x \in [u, \omega]$, and $0 \leqq f(x) \leqq 1$ for $x \in G$. (See [9], p. 141, Theorem 5.9.) We now consider the function $f(\max(x, y))$ on $G \times G$. The following facts are easily verified:

$$f(\max(x, y)) = \begin{cases} 1 & \text{for } (x, y) \in [\alpha, a] \times [\alpha, a] , \\ 0 & \text{for } (x, y) \in G \times [u, \omega] \cup [u, \omega] \times G , \\ f(x) & \text{for } (x, y) \in]a, u[\times [\alpha, a] , \\ f(y) & \text{for } (x, y) \in [\alpha, a] \times]a, u[. \end{cases} \tag{3.2.5}$$

We now have, applying 3.2.5, 3.2.3, and 3.2.4:

$$\begin{aligned} \lambda*\mu([\alpha, a]) \leqq L*M(f) &= \int_G \int_G f(\max(x, y)) d\mu(y) d\lambda(x) \\ &= \lambda([\alpha, a]) \cdot \mu([\alpha, a]) + \int_{]a,u[} f(x) d\lambda(x) \cdot \mu([\alpha, a]) \\ &\quad + \int_{]a,u[} f(y) d\mu(y) \cdot \lambda([\alpha, a]) + \int_{]a,u[} \int_{]a,u[} f(\max(x, y)) d\mu(y) d\lambda(x) \end{aligned}$$

$$\leqq \lambda([\alpha, a])\cdot\mu([\alpha, a]) + \lambda(]a, u[)\cdot\mu([\alpha, a]) + \mu(]a, u[)\cdot\lambda([\alpha, a]) + \lambda(]a, u[)\cdot\mu(]a, u[)$$

$$< \lambda([\alpha, a])\cdot\mu([\alpha, a]) + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \lambda([\alpha, a])\cdot\mu([\alpha, a]) + \varepsilon\ .$$

Since ε is arbitrary, we infer that $\lambda*\mu([\alpha, a]) \leqq \lambda([\alpha, a])\cdot\mu([\alpha, a])$.

To establish the reversed inequality, let ε again be an arbitrary positive real number, and let $f \in \mathfrak{C}(G)$ have the properties that $1 \geqq f(x) \geqq \varphi_{a]}(x)$ for $x \in G$ and $L*M(f) < \lambda*\mu([\alpha, a]) + \varepsilon$. The existence of such a function f follows at once from 1.6.2 and the non-negativity of $L*M$. It is obvious that $f(x)f(y) \leqq f(\max(x, y))$ for $(x, y) \in G \times G$. We now have

$$\lambda([\alpha, a])\cdot\mu([\alpha, a]) \leqq \int_G f(x)d\lambda(x)\cdot\int_G f(y)d\mu(y)$$

$$= \int_G\int_G f(x)f(y)d\mu(y)d\lambda(x) \leqq \int_G\int_G f(\max(x, y))d\mu(y)d\lambda(x)$$

$$= L*M(f) < \lambda*\mu([\alpha, a]) + \varepsilon\ .$$

Since ε is arbitrary, we have proved that

$$\lambda*\mu([\alpha, a]) = \lambda([\alpha, a])\cdot\mu([\alpha, a])\ . \tag{3.2.6}$$

We now prove that the mappings 3.2.2 are multiplicative for non-negative L and M. Since λ, μ, and $\lambda*\mu$ are regular measures, there exists, for every positive integer n, a compact subset C_n of $[\alpha, a[$ such that

$$\lambda([\alpha, a[) - \frac{1}{n} < \lambda(C_n)\ , \quad \mu([\alpha, a[) - \frac{1}{n} < \mu(C_n)\ ,$$

and

$$\lambda*\mu([\alpha, a[) - \frac{1}{n} < \lambda*\mu(C_n)\ .$$

We may evidently suppose that $C_n = [\alpha, b_n]$ for some $b_n < a$. (If a has an immediate predecessor a_-, so that $[\alpha, a[= [\alpha, a_-]$, we may refer to 3.2.6.) Then we have, applying 3.2.6:

$$\begin{aligned} \lambda*\mu([\alpha, a[) &= \lim_{n\to\infty} \lambda*\mu(C_n) = \lim_{n\to\infty} (\lambda(C_n)\cdot\mu(C_n)) \\ &= (\lim_{n\to\infty} \lambda(C_n))\cdot(\lim_{n\to\infty} \mu(C_n)) = \lambda([\alpha, a[)\cdot\mu([\alpha, a[)\ . \end{aligned} \tag{3.2.7}$$

To establish the present theorem for arbitrary $L, M \in \tilde{\mathfrak{C}}(G)$, we cite 1.6.5: $L = (L_1 - L_2) + i(L_3 - L_4)$, $M = (M_1 - M_2) + i(M_3 - M_4)$, where L_j and M_k are non-negative $(j, k = 1, \cdots, 4)$. The relations

$$\lambda*\mu([\alpha, a])=\lambda([\alpha, a])\cdot\mu([\alpha, a]) \qquad (a \in G) ,$$
$$\lambda*\mu([\alpha, a[)=\lambda([\alpha, a[)\cdot\mu([\alpha, a[) \qquad (a \in [\alpha, \omega]) ,$$

now follow from 3.2.6, 3.2.7, and the identity $(r\lambda+s\mu)(A)=r\lambda(A)+s\mu(A)$, valid for all $r, s \in K$; $L, M \in \tilde{\mathfrak{C}}(G)$, and Borel sets $A \subset G$. This completes the present proof.

3.3. THEOREM. *Let π be a homomorphism of $\tilde{\mathfrak{C}}(G)$ onto K. Then either there exists $b \in G$ such that $\pi(L)=\lambda([\alpha, b])$ for all $L \in \tilde{\mathfrak{C}}(G)$, or there exists $b \in]\alpha, \omega]$ such that $\pi(L)=\lambda([\alpha, b[)$ for all $L \in \tilde{\mathfrak{C}}(G)$.*

Proof. It follows from 2.10 that $\pi(\varepsilon_\alpha)=1$. Let x, y be elements of G. Then, using 2.12, we have

$$\pi(\varepsilon_x)\cdot\pi(\varepsilon_y)=\pi(\varepsilon_x*\varepsilon_y)=\pi(\varepsilon_{\max(x,y)}) .$$

The function p on G such that $p(x)=\pi(\varepsilon_x)$ for all $x \in G$ is therefore a semicharacter of G. Theorem 1.8 asserts that either there exists $b \in G$ such that

(3.3.1) $$\pi(\varepsilon_x)=\psi_{b]}(x) \qquad \text{for} \quad x \in G ,$$

or there exists $b \in]\alpha, \omega]$ such that

(3.3.2) $$\pi(\varepsilon_x)=\psi_{b[}(x) \qquad \text{for} \quad x \in G .$$

Suppose first that 3.3.1 holds. Applying π to the left side of 2.9.1, we have

(3.3.3) $$\pi(\varepsilon_a*\lambda)=\pi(\varepsilon_a)\cdot\pi(\lambda)=\psi_{b]}(a)\pi(\lambda) .$$

Applying π to the right side of 2.9.1, we have

(3.3.4) $$\pi(\lambda([\alpha, a])\varepsilon_a)+\lambda^{]a,\omega]})=\lambda([\alpha, a])\psi_{b]}(a)+\pi(\lambda^{]a,\omega]}) .$$

By 2.9.1, the last members of 3.3.3 and 3.3.4 are equal. We set $a=b$ in these expressions and equate them:

(3.3.5) $$\pi(\lambda)=\lambda([\alpha, b])+\pi(\lambda^{]b,\omega]}) .$$

We next show that $\pi(\lambda^{]b,\omega]})=0$. Here there are two cases. Suppose first that b has no immediate successor and that c is any element such that $c>b$. Then there is a d such that $b<d<c$. It follows at once from 2.9.1 that $\varepsilon_d*\lambda^{[c,\omega]}=\lambda^{[c,\omega]}$. Since $\pi(\varepsilon_d)=0$, we have

(3.3.6) $$\pi(\lambda^{[c,\omega]})=0 .$$

To infer from this that $\pi(\lambda^{]b,\omega]})=0$, we must use the continuity of π in the norm of $\tilde{\mathfrak{C}}(G)$. In fact,

(3.3.7) $$|\pi(L)| \leqq \|L\| \quad \text{for all} \quad L \in \tilde{\mathfrak{C}}(G)$$

([10], p. 69, Theorem). Let $\bar{\lambda}$ denote the total variation of λ ([5], p. 459, 1.2). It is easy to see that $\bar{\lambda}$ is regular. Thus, for every positive real number ε, there exists a compact subset of $]b, \omega]$ (which we may clearly take to be of the form $[c, \omega]$ with $c > b$) such that $\bar{\lambda}(]b, c[) < \varepsilon$. Then we have

(3.3.8) $$\|\lambda^{[c,\omega]} - \lambda^{]b,\omega]}\| = \sup\left\{\left|\int_{]b,c[} f(x)d\lambda(x)\right| : \|f\| \leqq 1\right\} \leqq \bar{\lambda}(]b, c[) < \varepsilon .$$

We infer from 3.3.6, 3.3.7, and 3.3.8 that $|\pi(\lambda^{]b,\omega]})| < \varepsilon$, and hence $\pi(\lambda^{]b,\omega]}) = 0$.

Suppose next that b has an immediate successor b_+. Then $]b, \omega] = [b_+, \omega]$. From 2.9.2, we have

$$\varepsilon_{b_+} * \lambda^{[b_+,\omega]} = \lambda^{]b_+,\omega]} ,$$

and since $\pi(\varepsilon_{b_+}) = \psi_{b]}(b_+) = 0$, we have

$$\pi(\lambda^{[b_+,\omega]}) = 0 = \pi(\lambda^{]b,\omega]}) .$$

Therefore $\pi(\lambda^{]b,\omega]}) = 0$ in both cases, and, returning to 3.3.5, we find

(3.3.9) $$\pi(\lambda) = \lambda([\alpha, b]) \qquad \text{for all} \quad \lambda \in \tilde{\mathfrak{C}}(G) .$$

This proves the present theorem in case 3.3.1 holds.

We have still to deal with the case in which 3.3.2 holds. If b has an immediate predecessor, we are actually in the case 3.3.1. We therefore may suppose that b has no immediate predecessor. Applying π to both sides of 2.9.1, we have as before

(3.3.10) $$\pi(\lambda) = \lambda([\alpha, a]) + \pi(\lambda^{]a,\omega]}) \qquad \text{for} \quad a \in [\alpha, b[.$$

Relations 2.9.2 and 3.3.2 imply that $\pi(\lambda^{[b,\omega]}) = 0$. Hence

$$\pi(\lambda^{]a,\omega]}) = \pi(\lambda^{]a,b[})$$

for all $a \in [\alpha, b[$. An argument based on $\|\lambda^{]a,b[}\|$, very like that used above, shows that for every positive real number ε, there exists $a_0 < b$ such that $|\pi(\lambda^{]a,\omega]})| < \varepsilon$ if $a_0 \leqq a < b$. Since λ is regular, there exists $a_1 < b$ such that

$$|\lambda([\alpha, a]) - \lambda([\alpha, b[)| < \varepsilon \quad \text{if} \quad a_1 \leqq a < b .$$

From these facts and 3.3.10, we obtain the present theorem in case 3.3.2 holds.

3.4. REMARK. It is interesting to compare 3.3 with the corresponding assertion for compact Abelian groups. Let H be a compact Abelian group with group operation xy. Then $\tilde{\mathfrak{C}}(H)$ is a convolution algebra, where

$$\lambda*\mu(f)=\int_H\int_H f(xy)d\mu(y)d\lambda(x) \qquad \text{for} \quad f\in\mathfrak{C}(H) \tag{3.4.1}$$

If H is infinite, then the homomorphisms of $\tilde{\mathfrak{C}}(H)$ onto K are enormously complicated, and in fact need not be described by characters of H (see [12] for a detailed discussion).

4. $\tilde{\mathfrak{C}}(G)$ is semisimple.

We establish first a preliminary result, which will also be of use in § 6.

4.1. THEOREM. *Let f be an element of $\mathfrak{C}(G)$ and let ε be a positive real number. Then there exists a finite subset $\{a_j\}_{j=0}^m$ of G, where*

$$\alpha=a_0<a_1<\cdots<a_j<a_{j+1}<\cdots<a_m=\omega\ ,$$

such that the oscillation of f is less than ε on each of the sets $]a_{j-1}, a_j]$ $(j=1, 2, \cdots, m)$.

Proof. The function f is continuous. Hence, for all $x\in G$, there exists an interval neighborhood $U(x)$ such that $|f(y)-f(y')|<\frac{\varepsilon}{2}$ for all $y, y'\in U(x)$. Since G is compact, a finite number of these neighborhoods cover G. Let $U_1, U_2, \cdots, U_p$ be such a collection of neighborhoods.

Each U_j has one of the following forms: $]u, v[$; $[u, v[$; $]u, v]$; $\{w\}(u<v)$. Whenever an interval U_j can be written in one of the last three forms, let the elements u, v or the element w be considered as the endpoints of U_j. Otherwise, call u, v the endpoints of U_j. There are at most $2p$ distinct endpoints of the sets U_j: we write them in increasing order as $a_0, a_1, \cdots, a_m$. Since α is in some U_j and since the only types of open intervals containing α are $[\alpha, v[$ $(\alpha<v)$ or $\{\alpha\}$ (if α is isolated), we must have $a_0=\alpha$. By the same token, we have $a_m=\omega$.

Now consider an arbitrary interval $]a_{k-1}, a_k]$ $(k=1, 2, \cdots, m)$. The point a_k lies in some interval U_s $(s=1, 2, \cdots, p)$. If U_s is of the form $]u_s, v_s[$ or $]u_s, v_s]$, it is obvious that $a_{k-1}\geqq u_s$ and hence $]a_{k-1}, a_k]\subset U_s$. If U_s has the form $[u_s, v_s[$ with $u_s<a_k$, then it is again obvious that $]a_{k-1}, a_k]\subset U_s$. In these cases, the oscillation of f on $]a_{k-1}, a_k]$ does not exceed $\varepsilon/2$. If U_s has the form $[a_k, v_s[$ or $\{a_k\}$, then since U_s

is open, a_k has an immediate predessor, say w. If $a_{k-1}=w$, then $]a_{k-1}, a_k]=\{a_k\}$. Consider an interval U_t that contains w. If U_t has the form $[w, v_t[$ or $]u_t, w]$ or $\{w\}$, then $w=a_{k-1}$. In these three cases, the oscillation of f on $]a_{k-1}, a_k]$ is 0. If U_t has the form $]u_t, v_t[$ and does not have any of the three preceding forms, then we have $a_k < v_t$, $u_t < w$, and necessarily $u_t \leqq a_{k-1}$. Again it follows that $]a_{k-1}, a_k] \subset U_t$, and the oscillation of f on $]a_{k-1}, a_k]$ does not exceed $\varepsilon/2$. Since $\varepsilon/2$ is less than ε, the lemma is proved.

Our next theorem shows that $\tilde{\mathfrak{C}}(G)$ is semisimple.

4.2. THEOREM. *Let L be an element of $\tilde{\mathfrak{C}}(G)$ such that $\lambda([\alpha, a])=0$ for all $a \in G$. Then $L=0$.*

Proof. Let f be any function in $\mathfrak{C}(G)$, let ε be a positive real number, and let $\{a_j\}_{j=0}^m$ be as in 4.1. Let p be the function on G such that

$$p(x)=\begin{cases} f(\alpha) & \text{for } x=\alpha ,\\ f(a_k) & \text{for } a_{k-1}<x\leqq a_k \quad (k=1, 2, \cdots, m) .\end{cases}$$

Then p is Borel measurable and bounded and hence is in $\mathfrak{L}_1(\lambda)$. Our hypothesis on λ implies that $\lambda(\{a\})=0$ and that $\lambda(]a_{k-1}, a_n])=0$ $(k=1, 2, \cdots, m)$. Consequently, $\int_G p(x)d\lambda(x)=0$. On the other hand, we have $|p(x)-f(x)|<\varepsilon$ for all $x \in G$. Therefore

$$\begin{aligned} |L(f)| &= \left|\int_G f(x)d\lambda(x)\right| = \left|\int_G f(x)d\lambda(x) - \int_G p(x)d\lambda(x)\right| \\ &\leqq \int_G |f(x)-p(x)|d\bar{\lambda}(x) \leqq \varepsilon\bar{\lambda}(G) . \end{aligned}$$

Since ε is an arbitrary positive real number, it follows that $L(f)=0$, and therefore $L=0$.

4.3. THEOREM. *Let L be an element of $\tilde{\mathfrak{C}}(G)$ such that $\lambda(\{\omega\})=0$ and $\lambda([\alpha, a[)=0$ for all $a \in]\alpha, \omega]$. Then $L=0$.*

The proof of this theorem differs only trivially from that of 4.2.

4.4. THEOREM. *The Banach algebra $\tilde{\mathfrak{C}}(G)$ is simisimple. If $L \in \tilde{\mathfrak{C}}(G)$ and $L \neq 0$, then the image of L under some homomorphism* 3.2.1 *is different from zero. If the image of L under every homomorphism* 3.2.2 *is zero, then $L=t\varepsilon_\omega$ for some $t \in K$.*

Proof. The second statement of this theorem merely repeats 4.2. The first statement follows from the second. To prove the third statement, let $t=\lambda(\{\omega\})$. Then $(\lambda-t\varepsilon_\omega)([\alpha, a[)=0$ for all $a\in]\alpha, \omega]$, and $(\lambda-t\varepsilon_\omega)(\{\omega\})=0$. We now appeal to 4.3.

5. $\hat{G}$ as the maximal ideal space of $\tilde{\mathfrak{S}}(G)$.

5.1. Theorems 3.2 and 3.3 identify completely the homomorphisms of $\tilde{\mathfrak{S}}(G)$ onto K. In order to study the space of all these homomorphisms, we introduce some new notation. For all $a\in G$, let $\boldsymbol{a}$ denote the homomorphism 3.2.1: $\boldsymbol{a}(L)=\lambda([\alpha, a])$. Let $\boldsymbol{G}$ denote the set of all homomorphisms $\boldsymbol{a}$. For $\lambda\in\tilde{\mathfrak{S}}(G)$, we define the function $\hat{\lambda}$ on $\boldsymbol{G}$ as usual: $\hat{\lambda}(\boldsymbol{a})=\boldsymbol{a}(L)$ for all $\boldsymbol{a}\in\boldsymbol{G}$. For $a\in]\alpha, \omega]$, let $\boldsymbol{a}'$ denote the homomorphism 3.2.2: $\boldsymbol{a}'(L)=\lambda[\alpha, a[)$. Let $\boldsymbol{G}'$ denote the set of all homomorphisms $\boldsymbol{a}'$. For $\lambda\in\tilde{\mathfrak{S}}(G)$, we define the function $\hat{\lambda}$ on $\boldsymbol{G}'$ as usual: $\hat{\lambda}(\boldsymbol{a}')=\boldsymbol{a}'(L)$ for all $\boldsymbol{a}'\in\boldsymbol{G}'$. By an abuse of notation, we identify $\boldsymbol{G}\cup\boldsymbol{G}'$ with the semigroup $\hat{G}$ of all semicharacters of G (1.8 and 1.12). Theorems 3.2 and 1.8 of course suggest this step. The function $\hat{\lambda}$ on $\hat{G}=\boldsymbol{G}\cup\boldsymbol{G}'$ is called the *Fourier transform* of L.

5.2. Before going further, we must agree on certain identifications that may have to be made between $\boldsymbol{G}$ and $\boldsymbol{G}'$. If $a\in G$ and a has an immediate successor a_+, then $[\alpha, a]=[\alpha, a_+[$, and hence $\boldsymbol{a}=\boldsymbol{a}'_+$. Equivalently, if $a\in]\alpha, \omega]$ and a has an immediate predecessor a_-, then $[\alpha, a[=[\alpha, a_-]$, and $\boldsymbol{a}'=\boldsymbol{a}_-$. For all such $a\in]\alpha, \omega]$, we agree to identify the point $\boldsymbol{a}'$ with the point $\boldsymbol{a}_-$.

5.3. For $u, v\in G$, we define $[\boldsymbol{u}, \boldsymbol{v}]$ as the set of all $\boldsymbol{c}\in\boldsymbol{G}$ such that $u\leqq c\leqq v$. The sets $[\boldsymbol{u}, \boldsymbol{v}[$, $[\boldsymbol{u}', \boldsymbol{v}']$, etc., are defined similarly.

5.4. The Gel'fand topology for $\hat{G}$ is the weakest topology (that is, the topology with the smallest family of open sets) that makes all of the functions $\hat{\lambda}$ continuous. It is well known that $\hat{G}$ is a compact Hausdorff space in this topology ([10], p. 52, Theorem 19B). We now describe the Gel'fand topology for $\hat{G}$.

5.5. THEOREM. *The point $\boldsymbol{\omega}$ is isolated in $\hat{G}$. If $b\in[\alpha, \omega[$ and b has no immediate successor, then a complete family of neighborhoods of $\boldsymbol{b}$ consists of all sets of the form*

$$(5.5.1)\qquad [\boldsymbol{b}, \boldsymbol{c}[\cup]\boldsymbol{b}', \boldsymbol{c}'] \qquad \textit{where} \quad c\in]b, \omega]\ .$$

If $b\in[\alpha, \omega[$ and b has an immediate successor b_+, then $\boldsymbol{b}=\boldsymbol{b}'_+$ is isolated in $\hat{G}$. If $b\in]\alpha, \omega]$ and b has no immediate predecessor, then a complete *family of neighborhoods of* $\boldsymbol{b}'$ *consists of all sets of the form*

$$]\boldsymbol{a}', \boldsymbol{b}'] \cup [\boldsymbol{a}, \boldsymbol{b}[\quad \textit{where} \quad a \in [\alpha, b[\ . \tag{5.5.2}$$

It b has an immediate predecessor b_-, *then* $\boldsymbol{b}' = \boldsymbol{b}_-$, *and is isolated.*

Proof. We use repeatedly the fact that all $\hat{\lambda}$ must be continuous on G. The function $\hat{\varepsilon}_\omega$ is 0 everywhere on $\hat{G}$ except at $\boldsymbol{\omega}$, and $\hat{\varepsilon}_\omega(\boldsymbol{\omega}) = 1$. Hence $\boldsymbol{\omega}$ is isolated.

Consider next any point $\boldsymbol{b}$ such that $\alpha \leqq b < \omega$. If b has no immediate successor, there exists, for every open set T containing $[\alpha, b]$, an element c such that $c > b$ and $[\alpha, b] \subset [\alpha, c[\subset T$. Every measure $\lambda \in \mathfrak{C}(\tilde{G})$ is regular, and hence we can find a $c_0 > b$ such that

$$|\lambda([\alpha, b]) - \lambda([\alpha, c)[| < \varepsilon \tag{5.5.3}$$

for all c such that $b < c \leqq c_0$, ε being an arbitrary positive real number. This means that

$$|\hat{\lambda}(\boldsymbol{b}) - \hat{\lambda}(\boldsymbol{c}')| < \varepsilon \qquad \text{if} \quad b < c \leqq c_0 \ . \tag{5.5.4}$$

If λ is non-negative, 5.5.3 clearly implies that $|\lambda([\alpha, b]) - \lambda([\alpha, c])| < \varepsilon$ for all c such that $b \leqq c \leqq c_0$. Since λ is a linear combination of four non-negative measures, we now have the following result.

5.5.5. Let $b \in [\alpha, \omega[$ and let b have no immediate successor. Let $\lambda \in \tilde{\mathfrak{C}}(G)$, and let ε be any positive real number. Then there is a $c_0 > b$ such that $|\hat{\lambda}(\boldsymbol{b}) - \hat{\lambda}(\boldsymbol{c})| < \varepsilon$ if $b \leqq c < c_0$ and $|\hat{\lambda}(\boldsymbol{b}) - \hat{\lambda}(\boldsymbol{c}')| < \varepsilon$ if $b < c \leqq c_0$.

If b has an immediate successor, b_+, then we have

$$\hat{\varepsilon}_b(\boldsymbol{x}) - \hat{\varepsilon}_{b_+}(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \boldsymbol{x} = \boldsymbol{b} = \boldsymbol{b}'_+ \ , \\ 0 & \text{elsewhere on } \hat{G} \ . \end{cases} \tag{5.5.6}$$

Since the function $\hat{\varepsilon}_b - \hat{\varepsilon}_{b_+}$ is continuous on $\hat{G}$, the point $\boldsymbol{b} = \boldsymbol{b}'_+$ is isolated.

We next consider a point $\boldsymbol{b}' \in \hat{G}$ such that b has no immediate predecessor. Then $[\alpha, b[$ is a nonclosed open subset of G, and for every closed subset F of $[\alpha, b[$, there exists $c < b$ such that $F \subset [\alpha, c] \subset [\alpha, b[$. If $\lambda \in \tilde{\mathfrak{C}}(G)$, then λ is regular, and we see just as in 5.5.5 that:

5.5.7. $\hat{\lambda}(\boldsymbol{c}')$ is arbitrarily close to $\hat{\lambda}(\boldsymbol{b}')$ if $c_0 < c \leqq b$ and $\hat{\lambda}(\boldsymbol{c})$ is arbitrarily close to $\hat{\lambda}(\boldsymbol{b}')$ if $c_0 \leqq c < b$ (here c_0 is an appropriately chosen element $< b$).

The case in which b has an immediate predecessor has already been dealt with.

The topology imposed on $\hat{G}$ by the neighborhood system 5.5.1 and 5.5.2 (and with isolated points as described) is obviously a Hausdorff topology. In 5.5.5, 5.5.6, and 5.5.7, we have shown that every function $\hat{\lambda}$ is continuous on $\hat{G}$ in this topology. From 5.4, we see that the Gel'fand topology is weaker than or equivalent to the topology just described.

To show that this topology is precisely the Gel'fand topology, consider any $b, c \in G$ such that $\alpha \leqq b < c \leqq \omega$ and such that b has no immediate successor. It is easy to see that

$$\hat{\varepsilon}_b(\boldsymbol{x}) - \hat{\varepsilon}_c(\boldsymbol{x}) = \begin{cases} 1 & \text{if} \quad \boldsymbol{x} \in [\boldsymbol{b}, \boldsymbol{c}[\cup]\boldsymbol{b}', \boldsymbol{c}'] \, , \\ 0 & \text{elsewhere on } G. \end{cases}$$

Hence all of the neighborhoods of $\boldsymbol{b}$ enumerated in 5.5.1 are necessarily open in the Gel'fand topology. Since c is the immediate successor of b if and only if b is the immediate predecessor of c, the same function $\hat{\varepsilon}_b - \hat{\varepsilon}_c$ shows that all of the neighborhoods of $\boldsymbol{b}'$ enumerated in 5.5.2 must be open in the Gel'fand topology. Points with immediate successors and the point $\boldsymbol{\omega}$ have already been dealt with: such points must be isolated in the Gel'fand topology for $\hat{G}$. This completes the present proof.

5.6. REMARK. Since $\tilde{\mathfrak{C}}(G)$ has the unit ε_α (2.10), $\hat{G}$ must be compact. Thus the topology of 5.5 is a compact Hausdorff topology. This fact could of course be established by a direct examination of $\hat{G}$.

5.7. The mapping $L \to \hat{\lambda}$ is a linear mapping of $\hat{\mathfrak{C}}(G)$ into the function space $\mathfrak{C}(\hat{G})$ that changes convolution into pointwise multiplication. That is, $L * M \to (\lambda * \mu)^\wedge = \hat{\lambda} \cdot \hat{\mu}$ for all $L, M \in \tilde{\mathfrak{C}}(G)$, where $\hat{\lambda} \cdot \hat{\mu}$ is the pointwise product of $\hat{\lambda}$ and $\hat{\mu}$ on $\hat{G}$. This follows at once from 3.2. Theorem 4.4 shows that this mapping is an algebraic isomorphism. The result of the present section is to describe the (unique) compact Hausdorff topology on $\hat{G}$ under which the functions $\hat{\lambda}$ are continuous. Thus in studying algebraic properties of $\tilde{\mathfrak{C}}(G)$, we may consider the subspace of $\hat{\mathfrak{C}}(G)$ consisting of all $\hat{\lambda}$. In 6.7 and 6.9, we will give a more precise description of these functions.

5.8. The Stone (or kernel-hull) topology for $\hat{G}$ ([10], p. 56) is identical with the Gel'fand topology. A neighborhood of $\boldsymbol{x} \in \hat{G}$ in the Stone topology consists of all $\boldsymbol{y}$ such that $\hat{\lambda}(\boldsymbol{y}) \neq 0$, where $\hat{\lambda}(\boldsymbol{x}) \neq 0$. It is clear that the Stone topology is weaker than or equal to the Gel'fand topology, and since the functions $\hat{\varepsilon}_b - \hat{\varepsilon}_c$ are different from 0 exactly on the neighborhoods 5.5.1 and 5.5.2, the two topologies coincide.

6. The Herglotz-Bochner theorem for $\tilde{\mathfrak{C}}(G)$.

6.1. Weil's generalization to locally compact Abelian groups of the Herglotz-Bochner theorem (see [10], pp. 141–142, Theorem 36A) gives an intrinsic characterization (positive definiteness and continuity) of all functions on the dual group that are Fourier-Stieltjes transforms of finite non-negative regular Borel measures. We here give two analogues of the Herglotz-Bochner theorem for the algebra $\tilde{\mathfrak{C}}(G)$.

6.2. Let λ be a non-negative measure in $\tilde{\mathfrak{C}}(G)$. Then the function $\hat{\lambda}$ is continuous, real-valued, and non-negative on $\hat{G}$. It is also nondecreasing in the sense that $\hat{\lambda}(\boldsymbol{a}) \leqq \hat{\lambda}(\boldsymbol{b})$ and $\hat{\lambda}(\boldsymbol{a}') \leqq \hat{\lambda}(\boldsymbol{b}')$ if $a \leqq b$. We shall show that these properties completely characterize Fourier transforms of non-negative measures. In fact if h is a continuous, real-valued, non-negative function on $\hat{G}$ such that $h(\boldsymbol{a}) \leqq h(\boldsymbol{b})$ for $a \leqq b$, and $h(\boldsymbol{\omega}) > 0$, then $h = \hat{\lambda}$ for some non-zero $\lambda \in \tilde{\mathfrak{C}}(G)$ such that $\lambda \geqq 0$. The proof requires a number of steps, which we state as separate theorems.

6.3. THEOREM. *Let h be a continuous function on $\hat{G}$ that is real-valued and non-decreasing on $\boldsymbol{G}$. Then h is also real-valued and non-decreasing on $\boldsymbol{G}'$.*

Proof. It is first clear that h is real-valued on $\boldsymbol{G}'$, since $\boldsymbol{G}$ is dense in $\hat{G}$. Let a, b be elements of G such that $a < b$, and let ε be a positive real number. There exists an element $c < a$ such that $|h(\boldsymbol{a}') - h(\boldsymbol{x})| < \varepsilon$ for all $\boldsymbol{x}$ such that $c \leqq x < a$ (see 5.5.2). This holds trivially if a has an immediate predecessor. Similarly, there exists an $e < b$ such that $|h(\boldsymbol{b}') - h(\boldsymbol{y})| < \varepsilon$ for all $\boldsymbol{y}$ such that $e \leqq y < b$. If we choose $e \geqq a$, then, as h is non-decreasing on $\boldsymbol{G}$, all of the numbers $h(\boldsymbol{x})$ are less than or equal to all of the numbers $h(\boldsymbol{y})$, and it follows that $h(\boldsymbol{a}') \leqq h(\boldsymbol{b}')$.

Given a function h as in 6.2, we must recapture the measure λ, or, equivalently, the linear functional L, whose Fourier transform is h. For this purpose, we introduce a Riemann integral with respect to h.

6.4. DEFINITION. Let h be any real-valued, non-decreasing function on $\boldsymbol{G}$. Let Δ denote a finite subset $\{a_0, a_1, \cdots, a_m\}$ of G, such that $a_0 = \alpha$, $a_m = \omega$, and $a_{j-1} < a_j$ $(j = 1, \cdots, m)$. For an arbitrary complex-valued function f on G, let

$$S(f, \Delta) = \sum_{j=1}^{m} f(a_j)(h(\boldsymbol{a}_j) - h(\boldsymbol{a}_{j-1})) .$$

6.5. THEOREM. *Let f be a continuous function on G. Then there exists a unique number $L(f)$ such that for every $\varepsilon > 0$ there exists a Δ_0 as in* 6.4 *with the property that $|L(f)-S(f, \Delta)| < \varepsilon$ for all $\Delta \supset \Delta_0$. We write this relation as $L(f) = \lim_{\Delta} S(f, \Delta)$.*

Proof. Let $\beta = h(\boldsymbol{\omega}) - h(\boldsymbol{\alpha})$. If $\beta = 0$, then $S(f, \Delta) = 0$ for all Δ and there is really nothing to prove. Otherwise, let θ be an arbitrary positive real number. Then, by 4.1, there exists a $\Delta = \{a_j\}_{j=0}^{m}$ such that the oscillation of f is less than $\beta^{-1}\theta$ in each of the sets $]a_{j-1}, a_j]$ $(j = 1, 2, \cdots, m)$. Suppose now that Γ is a finite subset of G such that $\Gamma \supset \Delta$. We shall prove that

$$|S(f, \Delta) - S(f, \Gamma)| < \theta . \tag{6.5.1}$$

Write $\Gamma = \{b_k\}_{k=0}^{r}$, $b_{k-1} < b_k$, and suppose that $b_s = a_1$, $b_k < a_1$ for $k < s$. Then we have

$$\begin{aligned} &\left| \sum_{k=1}^{s} f(b_k)(h(\boldsymbol{b}_k) - h(\boldsymbol{b}_{k-1})) - f(a_1)(h(\boldsymbol{a}_1) - h(\boldsymbol{a}_0)) \right| \\ &= \left| \sum_{k=1}^{s} (f(b_k) - f(a_1))(h(\boldsymbol{b}_k) - h(\boldsymbol{b}_{k-1})) \right| \\ &\leqq \beta^{-1}\theta \left(\sum_{k=1}^{s} (h(\boldsymbol{b}_k) - h(\boldsymbol{b}_{k-1})) \right) = \beta^{-1}\theta(h(\boldsymbol{a}_1) - h(\boldsymbol{a}_0)) . \end{aligned} \tag{6.5.2}$$

If $h(\boldsymbol{a}_1) - h(\boldsymbol{a}_0)$ is positive, it is clear that the inequality in 6.5.2 is strict. Estimates similar to 6.5.2 obviously hold for the b's lying in the intervals $]a_1, a_2], \cdots,]a_{m-1}, a_m]$. Adding these estimates together, we obtain the result that

$$|S(f, \Delta) - S(f, \Gamma)| < \left[\sum_{j=1}^{m} (h(\boldsymbol{a}_j) - h(\boldsymbol{a}_{j-1})) \right] \beta^{-1}\theta = \theta .$$

the strict inequality holding because some $h(\boldsymbol{a}_j) - h(\boldsymbol{a}_{j-1})$ is positive. This is just 6.5.1.

Let Δ_n be a subset of G as in 6.4 such that $|S(f, \Delta) - S(f, \Delta_n)| < n^{-1}$ for all $\Delta \supset \Delta_n$, and let $\Gamma_n = \bigcup_{j=1}^{n} \Delta_j$ $(n = 1, 2, 3, \cdots)$. Then $\{S(f, \Gamma_n)\}_{n=1}^{\infty}$ is a Cauchy sequence of complex numbers and hence has a limit, which we take as $L(f)$. If ε is a positive real number, then there exists an $n > 3/\varepsilon$ such that $|L(f) - S(f, \Gamma_n)| < \varepsilon/3$. If $\Delta \supset \Gamma_n$, then $\Delta \supset \Delta_n$, so that $|S(f, \Delta) - S(f, \Gamma_n)| < 2/n$. Thus $|L(f) - S(f, \Delta)| < \varepsilon$, as was to be proved. The uniqueness of $L(f)$ is proved by a standard argument, which we omit.

6.6. THEOREM. *The function L defined in* 6.5 *for all $f \in \mathfrak{C}(G)$ is a non-negative linear functional on $\mathfrak{C}(G)$.*

Proof. Since h is real and non-decreasing, it is clear that $S(\Delta, f)$ is real and non-negative for all real $f \in \mathfrak{C}(G)$ that are nonnegative and all Δ as in 6.4. Hence the limit $L(f)$ of these numbers is non-negative. The linearity of L follows at once from 6.5 and the obvious equality $S(\Delta, uf+vg) \equiv uS(\Delta, f)+vS(\Delta, g)$, valid for all complex numbers u, v, all $f, g \in \mathfrak{C}(G)$, and Δ as in 6.4.

We can now state and prove our main theorem.

6.7. THEOREM. *Let h be a continuous function on $\hat{G}$ that is real-valued, non-negative, and non-decreasing on $\boldsymbol{G}$. Let L be the non-negative linear functional associated with h as in* 6.5. *Let λ be the measure associated with L as in* 1.6.1. *Then h is the Fourier transform of $\lambda+h(\boldsymbol{\alpha})\varepsilon_{\alpha}$:*

$$h=\hat{\lambda}+h(\boldsymbol{\alpha})\hat{\varepsilon}_{\alpha}=\hat{\lambda}+h(\boldsymbol{\alpha}) . \tag{6.7.1}$$

Proof. Since $\hat{\lambda}$ and h are completely determined by their behavior on the dense subset $\boldsymbol{G}$ of $\hat{G}$, we have only to show that 6.7.1 holds on $\boldsymbol{G}$. That is, we must show that

$$\hat{\lambda}(\boldsymbol{a})=\lambda([\alpha, a])=h(\boldsymbol{a})-h(\boldsymbol{\alpha}) \quad \text{for all} \quad a \in G . \tag{6.7.2}$$

If $a \in G$ and a has an immediate successor, then the function $\psi_{a]}$ is continuous, and by the definition of λ given in 1.6.2, we have $\lambda([\alpha, a])= L(\psi_{a]})$. If Δ is any finite subset of G as in 6.4 that contains a, then it is plain that $S(\psi_{a]}, \Delta)=h(\boldsymbol{a})-h(\boldsymbol{\alpha})$. This implies that $L(\psi_{a]})=h(\boldsymbol{a})-h(\boldsymbol{\alpha})$, that is, that 6.7.2 holds for this value of a.

If a has no immediate successor, then, for every positive real number ε and every $b > a$, $b \in G$, there exists a non-negative real-valued function $f \in \mathfrak{C}(G)$ such that $f(x)=1$ for $x \leqq a$, $f(x)=0$ for $x \geqq b$, $0 \leqq f(x) \leqq 1$ for $x \in G$, and

$$|\lambda([\alpha, a])-L(f)| < \frac{\varepsilon}{3} . \tag{6.7.3}$$

This follows at once from 1.6.2 and the fact that G is a normal topological space. Now let Δ be any finite subset of G as in 6.4 that contains a and b. The inequalities

$$h(\boldsymbol{a})-h(\boldsymbol{\alpha}) \leqq S(f, \Delta) \leqq h(\boldsymbol{b})-h(\boldsymbol{\alpha}) \tag{6.7.4}$$

obviously hold. Since h is continuous on $\hat{G}$, we can choose the element $b > a$ such that $0 \leqq h(\boldsymbol{b})-h(\boldsymbol{a}) < \varepsilon/3$. By 6.5, there exists a finite subset Γ of G such that $\Gamma \supset \Delta$ and

$$|S(f, \Gamma)-L(f)| < \frac{\varepsilon}{3} . \tag{6.7.5}$$

Combining 6.7.3, 6.7.4, and 6.7.5, we have

(6.7.6) $$|\lambda([\alpha, a])-(h(\boldsymbol{a})-h(\boldsymbol{\alpha}))|<\varepsilon .$$

Since ε is arbitrary, we have proved 6.7.2.

6.8. REMARK. Theorem 6.7 is an analogue of the Herglotz-Bochner theorem, since it characterizes in a simple way those functions on $\hat{G}$ that are Fourier transforms of non-negative measures in $\tilde{\mathfrak{C}}(G)$. We can also obtain an exact analogue of the Herglotz-Bochner theorem in terms of positive definite functions. A function p on $\hat{G}$ is said to be positive definite if

(6.8.1) $$\sum_{j=1}^{m}\sum_{k=1}^{m}\xi_j\bar{\xi}_k p(\chi_j\bar{\chi}_k)\geqq 0$$

for all complex numbers $\xi_1, \cdots, \xi_m$ and all distinct $\chi_1, \cdots, \chi_m$ in $\hat{G}$. If λ is a non-negative measure in $\tilde{\mathfrak{C}}(G)$, then we have

$$\sum_{j=1}^{m}\sum_{k=1}^{m}\xi_j\bar{\xi}_k\hat{\lambda}(\chi_j\bar{\chi}_k)=\sum_{j=1}^{m}\sum_{k=1}^{m}\xi_j\bar{\xi}_k\int_G\chi_j(x)\bar{\chi}_k(x)d\lambda(x)$$
$$=\int_G\left|\sum_{j=1}^{m}\xi_j\chi_j(x)\right|^2 d\lambda(x)\geqq 0 .$$

Hence $\hat{\lambda}$ is continuous and positive definite in the sense of 6.8.1. Conversely, let p be a continuous function on $\hat{G}$ that satisfies 6.8.1. Let $a, b\in G$ and let $a<b$. For $m=2$, $\chi_1=\psi_{a]}$, $\chi_2=\psi_{b]}$, $\xi_1=1$, and $\xi_2=-1$, the inequality 6.8.1 obviously reduces to

(6.8.2) $$-p(\psi_{a]})+p(\psi_{b]})\geqq 0 .$$

Writing $p(\psi_{x]})=p(\boldsymbol{x})$ for $x\in G$, we have $p(\boldsymbol{a})\leqq p(\boldsymbol{b})$. From 6.8.1, we also see that p is non-negative. That is, p is continuous and non-decreasing on $\boldsymbol{G}$ and hence is the Fourier transform of a non-negative measure (6.7). Monotonicity is a much easier property to verify, in applications, than the inequality 6.8.1, so that the present characterization of Fourier transforms of nonnegative measures as continuous, positive definite functions is perhaps only a curiosity.

6.9. REMARK. Theorem 6.7 permits us to characterize general Fourier transforms $\hat{\lambda}$, where λ is an arbitrary complex-valued measure in $\tilde{\mathfrak{C}}(G)$, as being continuous functions on G that are linear combinations of continuous, real-valued, non-decreasing functions. However, there is another characterization of the functions $\hat{\lambda}$, more intrinsic in nature. Namely, let p be a function on $\hat{G}$ and let a, b be elements of G such that $a \leqq b$. We define the variation of p on the interval $[\boldsymbol{a}, \boldsymbol{b}]$ as the supremum of all numbers

$$\sum_{j=1}^{m} |p(\boldsymbol{a}_j) - p(\boldsymbol{a}_{j-1})| ,$$

taken over all finite sets $a = a_0 < a_1 < \cdots < a_m = b$ (if $a = b$, we take the variation as 0). We write this variation as $V(p\colon \boldsymbol{a}, \boldsymbol{b})$. One can then prove that a function q on $\hat{G}$ is the Fourier transform of some measure in $\tilde{\mathfrak{C}}(G)$ if and only if q is continuous and $V(q\colon \boldsymbol{\alpha}, \boldsymbol{\omega})$ is finite. The proof is suggested by standard arguments from the elementary theory of functions of a real variable (see for example [11], pp. 215–223). In the non-trivial direction, the proof is carried out by showing that every continuous real-valued function of finite variation on $\hat{G}$ is the difference of two continuous, real-valued, non-decreasing functions on $\hat{G}$. We omit the details.

7. An application to the theory of probability.

7.1. Theorem 6.7 has applications to the theory of probability. Let Φ be a random variable defined on a probability space (Y, π) with values in G. The function d on $\hat{G}$, defined by

$$\begin{cases} d(\boldsymbol{a}) = \pi\{y\colon y \in Y,\ \Phi(y) \leqq a\} & \text{for} \quad a \in G\ , \\ d(\boldsymbol{a}') = \pi\{y\colon y \in Y,\ \Phi(y) < a\} & \text{for} \quad a \in \,]\alpha, \omega]\ , \end{cases} \tag{7.1.1}$$

is obviously non-decreasing on $\boldsymbol{G}$ and $\boldsymbol{G}'$. Under some obvious hypotheses on π and Φ, this function d is continuous on $\hat{G}$ and hence is the Fourier transform of a probability measure λ in $\hat{\mathfrak{C}}(G)$ (6.7). It is clear that λ is non-negative, and since $d(\boldsymbol{\omega}) = 1$, we must have $\lambda(G) = 1$, that is, λ is a probability measure on the Borel sets of G. If Φ_j are *independent* random variables as above with corresponding probability measures $\lambda_j \in \tilde{\mathfrak{C}}(G)$ $(j = 1, \cdots, m)$, then the probability corresponding to the product $\Phi_1 \cdots \Phi_m$ is the convolution $\lambda_1 * \cdots * \lambda_m$. Thus the arithmetic of independent sets of random variables is just the arithmetic of the set of all continuous, nonnegative, non-decreasing functions p on $\hat{G}$ such that $p(\boldsymbol{\omega}) = 1$. The operation is of course pointwise multiplication on $\hat{G}$. If we denote the set of all probability measures in $\tilde{\mathfrak{C}}(G)$ by $\mathfrak{P}$, then the set of function on $\hat{G}$ that we are now considering is exactly $\hat{\mathfrak{P}}$. In the case of a finite semigroup G, the arithmetic arising in this way has been studied in detail in another place [6].

We proceed to a description of some of the properties of $\mathfrak{P}$ and $\hat{\mathfrak{P}}$.

7.2. It is clear that $\mathfrak{P}$ has a unit, namely ε_α (2.10). Since the measure ε_ω has the property that

$$\hat{\varepsilon}_\omega(\boldsymbol{x})=\begin{cases} 0 & \text{if } \boldsymbol{x}\neq\boldsymbol{\omega}\,, \\ 1 & \text{if } \boldsymbol{x}=\boldsymbol{\omega}\,, \end{cases}$$

it follows that $\varepsilon_\omega * \lambda=\varepsilon_\omega$ for all λ such that $\lambda(G)=1$. Hence ε_ω is a zero in the set $\mathfrak{P}$.

7.3. We next identify the idempotent elements of $\mathfrak{P}$. If $\lambda*\lambda=\lambda$, where $\lambda\in\tilde{\mathfrak{C}}(G)$, then $\hat{\lambda}^2=\hat{\lambda}$, and $\hat{\lambda}$ assumes only the values 0 and 1. If $\lambda\in\mathfrak{P}$, then $\hat{\lambda}$ is nondecreasing on $\boldsymbol{G}$ and $\lambda(\boldsymbol{\omega})=1$. The requirement of continuity makes it obvious that there exists an element $b\in G$ such that $\hat{\lambda}(\boldsymbol{a})=0$ for $a<b$, $\hat{\lambda}(\boldsymbol{a})=1$ for $a\geqq b$, $\hat{\lambda}(\boldsymbol{a}')=0$ for $a\leqq b$, and $\hat{\lambda}(\boldsymbol{a}')=1$ for $a>b$. This implies that $\hat{\lambda}=\hat{\varepsilon}_b$. Hence the only idempotent elements of $\mathfrak{P}$ are the measures ε_b.

7.4. Definition. Let $\{\lambda_n\}_{n=1}^{\infty}$ be a sequence of measures, where $\lambda_n\in\mathfrak{P}$ for all n. If there exists $\lambda\in\tilde{\mathfrak{C}}(G)$ such that $\lim\limits_{n\to\infty}\lambda_n(\boldsymbol{x})=\lambda(\boldsymbol{x})$ for all $\boldsymbol{x}\in\hat{G}$, then we say that λ is the limit of the sequence $\{\lambda_n\}_{n=1}^{\infty}$, and we write $\lambda=\lim\lambda_n$.

7.5. It is easy to show that $\lim\lambda_n$ is in $\mathfrak{P}$ whenever it exists. The notion of limit adopted here is very like that employed in the classical theory of probability (see for example [3], pp. 58–62, and esp. 102). There are obvious differences, as we insist on pointwise convergence throughout the *entire* space of homomorphisms $\hat{G}$, while the classical theory deals only with the homomorphisms defined by integrals $\int_{-\infty}^{\infty}e^{-ixy}\,d\lambda(x)$, which are not even dense in the space of all homomorphisms (see [12]).

7.6. Theorem. *Let* $\lambda\in\mathfrak{P}$ *and let* $\lambda^{[n]}-\lambda*\cdots*\lambda_{(n)}$ $(n=1,2,3,\cdots)$. *Then there is an element* $a\in G$ *such that* $\lim\lambda^{[n]}=\varepsilon_a$.

Proof. Consider the function $\hat{\lambda}^n=(\lambda^{[n]})^\wedge$ on the set $\boldsymbol{G}$. Let $A=\{x\colon x\in G,\ \hat{\lambda}(\boldsymbol{x})=1\}$. Since $\omega\in A$, A is non-void. Let $a=\inf A$. Since $\hat{\lambda}$ is non-decreasing, we have $\{x\colon x\in G,\ x>a\}\subset A$. Since $\hat{\lambda}$ is continuous on $\boldsymbol{G}$, we have $a\in A$. It follows that

$$\hat{\lambda}(\boldsymbol{x})\begin{cases} =1 & \text{if } x\geqq a\,, \\ <1 & \text{if } x<a\,. \end{cases} \tag{7.6.1}$$

This implies that

$$\lim_{n\to\infty} \hat{\lambda}(x) = \begin{cases} 0 & \text{if} \quad x < a \, , \\ 1 & \text{if} \quad x \geqq a \, . \end{cases} \tag{7.6.2}$$

Therefore $\lim_{n\to\infty} \hat{\lambda}^n(\boldsymbol{x}) = \hat{\varepsilon}_a(\boldsymbol{x})$ for all $\boldsymbol{x} \in \boldsymbol{G}$. The same relation holds if $\boldsymbol{x} \in \boldsymbol{G}'$, as is shown by the same argument. (Continuity shows that $a = \inf \{x : x \in G,\ \hat{\lambda}(\boldsymbol{x}') = 1\}$.) Therefore, by 7.4, $\lim \lambda^{[n]} = \varepsilon_a$.

7.7. Finally, we may look for the class of probability measures in $\mathfrak{P}$ that can be written in the form $\lim (\mu_1 * \cdots * \mu_n)$, where $\mu_n \in \mathfrak{P}$. For this purpose, it is convenient to go over to $\hat{\mathfrak{P}}$. Let λ be an arbitrary element of $\mathfrak{P}$. Write $\hat{\lambda} = p$. Then for every positive integer r, there is a unique nonnegative function $p^{1/r}$ on $\hat{G}$. It is easy to see that $p^{1/r}$ satisfies the conditions of 6.7 and has the property that $p^{1/r}(\boldsymbol{\omega}) = 1$. Hence $p^{1/r}$ is the Fourier transform of a probability μ_r such that $u_r^{[r]} = \lambda$. It is furthermore clear that

$$\lim_{n\to\infty} p^{1/2}(\boldsymbol{x}) \cdot p^{1/4}(\boldsymbol{x}) \cdot \ \cdots \ \cdot p^{1/2n}(\boldsymbol{x}) = p(\boldsymbol{x})$$

for all $\boldsymbol{x} \in \hat{G}$. Therefore, if we write $\lambda_n = \mu_2 * \mu_4 * \cdots * \mu_{2n}$, we have $\lim \lambda_n = \lambda$. Therefore every λ in $\mathfrak{P}$ is an "infinite product". If λ is not idempotent (that is, not of the form ε_a), then no μ_r is idempotent, and λ is an infinite product with "nondegenerate" factors.

If $\varepsilon_\alpha = \lim (\mu_1 * \cdots * \mu_n)$, then it is clear that all μ_j are equal to ε_α. For in the contrary case, we have $\hat{\mu}_j(\boldsymbol{a}) < 1$ for some $\boldsymbol{a} \in \boldsymbol{G}$ and some positive integer j. Hence $\lim_{n\to\infty} \hat{\mu}_1(\boldsymbol{a}) \cdots \hat{\mu}_n(\boldsymbol{a}) \leqq \hat{\mu}_j(\boldsymbol{a}) < 1$, and $\hat{\mu}_1 \cdot \ \cdots \ \cdot \hat{\mu}_n$ does not converge to $\hat{\varepsilon}_\alpha = 1$ everywhere on G.

On the other hand, if $b \in G$ and $b > \alpha$, choose any $u \in G$ such that $\alpha \leqq u < a$. A simple calculation shows that

$$\lim_{n\to\infty} (\tfrac{1}{2}\hat{\varepsilon}_u(\boldsymbol{x}) + \tfrac{1}{2}\hat{\varepsilon}_a(\boldsymbol{x}))^n = \hat{\varepsilon}_a(\boldsymbol{x})$$

uniformly on $\hat{G}$. Hence ε_a is an infinite product with all factors nondegenerate. (For the case of a finite G, see [**6**], 8.2.)

7.8. An intuitive interpretation of the results of 7.1–7.7 may be given. Consider a game whose possible outcomes are points of G, with the probability that the outcome lies in $A \subset G$ given by $\lambda(A)$, where $\lambda \in \mathfrak{P}$. We play the game repeatedly and keep score as follows. After the first game, we take its outcome, x_1, as our "score". After each subsequent game, we take as our score the maximum of its outcome and our previous score. That is, the score after n games is $\max(x_1,$

$\cdots, x_n)$. The probability that this score lies in $A \subset G$ is $\lambda^{[n]}(A)$. Hence, as $n \to \infty$, 7.6 shows that the outcome is almost certainly a, where $a = \inf \{x : x \in G,\ \lambda(]x, \omega]) = 0\}$. This is in accordance with what one intuitively expects. If there is a positive probability of obtaining x in some interval $[a, b]$, then, after sufficiently many repetitions, the probability is arbitrarily close to 1 that the maximum will be greater than or equal to a.

A similar interpretation, based on 7.7, can be given for games with *different* probabilities λ_n. Here an arbitrary $\lambda \in \mathfrak{P}$ can be obtained as the limiting probability as the number of games goes to ∞.

8. Examples and special results.

Our construction yields interesting results in certain classical cases. We here list a few of them.

8.1. Let G be the closed interval $[0, 1]$ on the real line, with the usual ordering. Then $\tilde{\mathfrak{S}}(G)$ consists of all complex, finite, countably additive Borel measures on $[0, 1]$. The space $\hat{G}$ is the union $I \cup I' \cup \{1\}$, where $I = [0, 1[$ and I' is a replica of $]0, 1]$ disjoint from I and $\{1\}$. The point 1 is isolated. Sets of the form $[t, t+\delta[\cup]t', t'+\delta']$, where $[t, t+\delta[\subset I$ and $]t', t'+\delta'] \subset I'$, are a basis for open sets in $I \cup I'$. This topology was described many years ago by Alexandroff and Urysohn for counter-example purposes [1], and it seems remarkable that it turns up here as the maximal ideal space of a certain Banach algebra.

As noted in 6.9, the Fourier transforms $\hat{\lambda}$ are just the continuous functions on G that have finite variation on $I \cup \{1\}$. Now let φ be any complex-valued function on $[0, 1]$ that has finite variation and is continuous on the right: $\varphi(t+0) = \varphi(t)$ for $0 \leqq t < 0$. It is well known ([3], p. 53) that φ determines and is determined by a $\lambda \in \tilde{\mathfrak{S}}(G)$: $\varphi(t) = \lambda([0, t])$ $(0 \leqq t \leqq 1)$. Hence $\hat{\lambda}(t) = \varphi(t)$ for all $t \in I \cup \{1\}$, and it is easy to see that $\hat{\lambda}(t') = \varphi(t-0)$ for $t' \in I'$. It follows that the algebra $\mathfrak{B}$ of all right-continuous functions of finite variation on $[0, 1]$ with pointwise operations is isomorphic to the algebra of Fourier transforms $\hat{\lambda}$ and hence to $\tilde{\mathfrak{S}}(G)$. Furthermore, the homomorphisms of $\mathfrak{B}$ onto K all have the form $\varphi \to \varphi(t)$ $(0 \leqq t \leqq 1)$ or $\varphi \to \varphi(t-0)$ $(0 < t \leqq 1)$. This answers a question put to the first-named author by Professor Einar Hille in 1946. Finally, if $\varphi_j \in \mathfrak{B}$, and φ_j corresponds to the measure $\lambda_j \in \tilde{\mathfrak{S}}(G)$ $(j = 1, \cdots, m)$, then the function $\varphi_1 \cdot \cdots \cdot \varphi_m$ corresponds to $\lambda_1 * \cdots * \lambda_m$.

8.2. Let G be any well-ordered set having a greatest element. It is obvious that G is compact and hence $\tilde{\mathfrak{C}}(G)$ is an algebra of the kind analyzed in the present paper. The measures in $\tilde{\mathfrak{C}}(G)$ are all uncomplicated. In fact, if $\lambda \in \tilde{\mathfrak{C}}(G)$, there exists a countable subset $\{a_n\}_{n=1}^{\infty}$ of G and a sequence $\{z_n\}_{n=1}^{\infty}$ of complex numbers such that $\sum_{n=1}^{\infty} |z_n| < \infty$ and such that

$$\lambda = \sum_{n=1}^{\infty} z_n \varepsilon_{a_n} . \tag{8.2.1}$$

The proof of this depends upon the following fact.

8.2.2. Let A be a well-ordered set with a greatest element and let δ be a finitely additive, real-valued, non-negative measure on the Borel sets of A such that $\delta(\{p\})=0$ for all $p \in A$ and δ is inner regular in the sense that $\delta(P) = \sup \{\delta(F) : F \text{ compact}, F \subset P\}$ for all intervals $P = [\alpha, u[\subset A$. Then $\delta = 0$.

Proof. We may suppose that A is infinite. Let α be the least element of A and let α_+ be the successor of α. Then $\delta([\alpha, \alpha_+[) = \delta(\{\alpha\}) = 0$. Suppose that $u \in A$ and that $\delta([\alpha, t[) = 0$ for all $t < u$. If u has an immediate predecessor u_-, then we have

$$\delta([\alpha, u[) = \delta([\alpha, u_-[\cup \{u_-\}) = \delta([\alpha, u_-[) + \delta(\{u_-\}) = 0 .$$

If u has no immediate predecessor, then for every compact set $F \subset [\alpha, u[$, there is a $t < u$ such that $[\alpha, t] \supset F$. There is also a t' such that $t < t' < u$, and we have $F \subset [\alpha, t] \subset [\alpha, t'[$. By our inductive hypothesis, we have $\delta([\alpha, t[) = 0$. By the regularity of δ, we infer $\delta([\alpha, u[) = \sup \{\delta(F) : F \text{ compact}, F \subset [\alpha, u[)\} = 0$. Hence $\delta([\alpha, u[) = 0$ for all $u \in A$. Since $\delta(\{\omega\}) = 0$, it follows that $\delta(A) = 0$.

In proving 8.2.1 from 8.2.2, we may clearly suppose that λ is non-negative (use 1.6.5). Let $\{a_n\}_{n=1}^{\infty}$ be the subset of G consisting of all points for which λ is positive, and let $z_n = \lambda(\{a_n\})$. Then $\delta = \lambda - \sum_{n=1}^{\infty} z_n \varepsilon_{a_n}$ is a measure satisfying the hypothesis of 8.2.2 (this δ is even countably additive).

It follows that the algebra $\tilde{\mathfrak{C}}(G)$ is isomorphic to the algebra $l_1(G)$ described in [8]. Since we have obtained all of the semicharacters of G in the present case, Theorems 1.8, 3.3, and 4.4 of the present paper are somewhat more precise that the corresponding Theorems 5.1, 2.7, and 5.8 of [8].

8.3. As another illustration of our techniques, we find *all* idempotent elements in $\tilde{\mathfrak{C}}(G)$, where G satisfies 1.5. If $\lambda * \lambda = \lambda$, then $\hat{\lambda}^2 = \hat{\lambda}$ and $\hat{\lambda}$ can assume only the values 0 and 1. According to 6.9, $\hat{\lambda}$ must have finite variation on $\boldsymbol{G}$ and be continuous on $\hat{G}$. Hence $\hat{\lambda}$ can have only a finite number of changes of sign on $\boldsymbol{G}$. A simple argument shows that there exists a finite subset $\{b_j\}_{j=1}^m$ of G such that $\alpha < b_1 < b_2 < \cdots < b_m < \omega$ (we write $\alpha = b_0$, $\omega = b_{m+1}$ in the following formulas) with the following properties. First, we may have

$$2\hat{\lambda}(\boldsymbol{x}) = \begin{cases} 1-(-1)^k & \text{for} \quad \boldsymbol{x} \in [\boldsymbol{b}_k, \boldsymbol{b}_{k+1}[\quad (k=0, \cdots, m)\,, \\ 1-(-1)^m & \text{for} \quad \boldsymbol{x} \in [\boldsymbol{b}_m, \boldsymbol{b}_{m+1}]\,. \end{cases} \tag{8.3.1}$$

Second, we may have

$$2\hat{\lambda}(\boldsymbol{x}) = \begin{cases} 1+(-1)^k & \text{for} \quad \boldsymbol{x} \in [\boldsymbol{b}_k, \boldsymbol{b}_{k+1}[\quad (k=0, \cdots, m)\,, \\ 1+(-1)^m & \text{for} \quad \boldsymbol{x} \in [\boldsymbol{b}_m, \boldsymbol{b}_{m+1}]\,. \end{cases} \tag{8.3.2}$$

These are the only possibilities. Translating this into a statement about the original measures, we see that λ must have the form

$$\lambda = \varepsilon_{c_0} - \varepsilon_{c_1} + \varepsilon_{c_2} + \cdots + (-1)^k \varepsilon_{c_k}\,, \tag{8.3.3}$$

where $\alpha \leqq c_0 < c_1 < \cdots < c_k \leqq \omega$. Since every measure 8.3.3 is obviously idempotent, we have found all idempotent measures in $\tilde{\mathfrak{C}}(G)$. This may be compared with Theorem 9.1 of [8], where we obtain a less precise result for a class of measure algebras related to but more complicated than those under study here.

8.4. Again let G satisfy 1.5. $\tilde{\mathfrak{C}}(G)$ admits an obvious involution. Let $L \in \hat{\mathfrak{C}}(G)$ and $L = M + iN$, where the functionals M and N are real-valued for real-valued $f \in \mathfrak{C}(G)$. Then the mapping $L \rightarrow \bar{\bar{L}} = M - iN$ is an involution of $\tilde{\mathfrak{C}}(G)$. Furthermore, $\tilde{\mathfrak{C}}(G)$ is obviously symmetric under this involution: $(\bar{\bar{\lambda}})^\wedge$ is the complex conjugate of $\hat{\lambda}$. However, $\tilde{\mathfrak{C}}(G)$ is never isomorphic to $\tilde{\tilde{\mathfrak{C}}}(G)$ (pointwise operations) if G is infinite. If G is infinite, we may suppose without loss of generality that G contains an infinite strictly increasing subset

$$a_1 < a_2 < a_3 < \cdots < a_n < \cdots\,. \tag{8.4.1}$$

Let b be the least upper bound of this set. It is easy to see that $T = \{\boldsymbol{a}_n\} \cup \{\boldsymbol{b}'\}$ is a closed subset of $\hat{G}$. The function γ on T such that $\gamma(\boldsymbol{a}_n) = \dfrac{1}{n}(1-(-1)^n)$ and $\gamma(\boldsymbol{b}') = 0$ is continuous on T. By Tietze's extension theorem, there is a continuous function γ_0 on $\hat{G}$ such that $\gamma_0(\boldsymbol{a}_n) = \gamma(\boldsymbol{a}_n)$ ([9], p. 242). Obviously γ_0 has infinite variation on $\boldsymbol{G}$ and hence is not a Fourier transform (6.9).

8.5. Following a suggestion of the referee, we note that if a semigroup G satisfies all of the hypotheses of 1.1–1.3 and if 1.4 is replaced by the hypothesis of *local* compactness, then it can be treated in much the same way as we have treated the compact case. Certain changes, however, are needed. The function space $\mathfrak{C}(G)$ of 1.6 is replaced by $\mathfrak{C}^*(G)$, the space of all bounded continuous functions on G. The conjugate space $\tilde{\mathfrak{C}}(G)$ is replaced by $\mathscr{M}(G)$, the space of all countably additive, complex-valued, finite Borel measures on G. (This is a realization of $\tilde{\mathfrak{C}}(G)$ for G compact but is ordinarily only a very small part of the conjugate space of $\mathfrak{C}^*(G)$ if G is non-compact.) The integral 1.6.1 exists for all $f \in \mathfrak{C}^*(G)$ and $\lambda \in \mathscr{M}(G)$ and defines a bounded linear functional on $\mathfrak{C}^*(G)$. Under this definition, $\mathscr{M}(G)$ is a convolution algebra. Every semicharacter of G is defined by a Dedekind cut, and it will be of the form 1.8.1, 1.8.2, or as in 1.11. $\mathscr{M}(G)$ has a unit if and only if G has a least element a and the unit in this case is ε_a. (See 2.10.) The results of §§ 3 and 4 can be carried over with obvious modifications. The maximal ideal space of $\mathscr{M}(G)$ is still $\hat{G}$ (see § 5), but the topological structure may be complicated. We omit the details. The changes necessary in §§ 6–7 are considerably greater, and the more general results to be obtained would not seem to justify carrying out all of the details.

References

1. P. S. Alexandroff and Paul Urysohn, *Mémoire sur les espaces compacts*, Verh. Nederl. Akad. Wetensch. Afd. Natuurk., **14**, No. 1.

2. N. Bourbaki, *Éléments de mathématique*, Première partie, Livre VI. Intégration, Ch. I–V. Act. Sci. et Ind. 1175. Hermann et Cie., Paris, 1952.

3. Harald Cramér, *Mathematical methods of statistics*, Princeton University Press, 1946.

4. Paul Halmos, *Measure theory*, D. Van Nostrand Co., Inc., New York, 1950.

5. Edwin Hewitt, *Remarks on the inversion of Fourier-Stieltjes transforms*, Ann. of Math. (2 Ser.) **57** (1953), 458–474.

6. Edwin Hewitt and Herbert S. Zuckerman, *Arithmetic and limit theorems for a class of random variables*, Duke Math. J., **22** (1955), 595–615.

7. ———, *Finite dimensional convolution algebras*, Acta Math. **93** (1955), 67–119.

8. ———, *The l_1 algebra of a commutative semigroup*, Trans. Amer. Math. Soc. **83** (1956), 70–97.

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

10. Lynn H. Loomis, *An introduction to abstract harmonic analysis*, D. Van Nostrand Co., Inc., New York, 1953.

11. I. P. Natanson, *Theory of functions of a real variable*, Frederick Ungar, New York, 1955.

12. Yu. A. Šreĭder, *Stroenie maksimal'nyh idealov v kol'cah mer so svertkoĭ*, Mat. Sb. N.S. **27** (69), (1950), 297–318.

THE INSTITUTE FOR ADVANCED STUDY AND
THE UNIVERSITY OF WASHINGTON

SOME TAUBERIAN THEOREMS

AMNON JAKIMOVSKI

1. Introduction. The following Taurberian theorem is well known.

THEOREM A. *If the sequence* $\{s_n\}$, $n=0, 1, 2, \cdots$, *is summable Abel*[1] *to* s *and the sequence* $\{n(s_n-s_{n-1})\}$ *is bounded on one side, then* $\{s_n\}$ *is convergent to* s.

Another Tauberian theorem, proved in [4], is

THEOREM B. *If the series* $\sum_{n=0}^{\infty} a_n$ *is summable Abel to* s *and the sequence* $\{n^2(a_{n-1}-a_n)\}$ *is bounded on one side, then* $\lim_{n\to\infty} na_n=0$.

An immediate consequence of Theorem B is the well known proposition that, for a convergent series $\sum_{n=0}^{\infty} a_n$ with monotonically decreasing terms, $\lim_{n\to\infty} na_n=0$.

By a well known theorem of Tauber, the series $\sum_{n=0}^{\infty} a_n$ of Theorem B is convergent and hence the sequence $\{s_n\}$ of partial sums of the series is summable $(H, -1)$, that is, $\{s_n\}$ is summable by the Hölder method of order -1, as defined in § 2. Thus Theorem B is equivalent to the following

THEOREM C. *If the sequence* $\{s_n\}$, $n=0, 1, 2, \cdots$, *is summable Abel to* s *and the sequence* $\left\{\binom{n}{2}(s_{n-2}-2s_{n-1}+s_n)\right\}$ *is bounded on one side, then* $\{s_n\}$ *is summable by the Hölder method of summability of order* -1.

As will be shown below both Theorem A and Theorem C are special cases of general results proved in § 5 of this paper.

The Tauberian conditions,

$$\binom{n}{1}(s_{n-1}-s_n)=O_L(1)$$

and

Received September 16, 1955 and in revised forms January 16, 1956, and May 28, 1956.

[1] Concepts and propositions mentioned or used in this paper without definition or proof are to be found in Hardy's book [3].

$$\binom{n}{2}(s_{n-2}-2s_{n-1}+s_n)=O_L(1)\ ,$$

belong to the general class of conditions of the form

$$\binom{n}{k}\Delta^k s_{n-k}=O_L(1)\ ,$$

where k is some fixed nonnegative integer and $\Delta^k s_r$ is defined by

$$\Delta^k s_r=\sum_{p=0}^{k}(-1)^p\binom{k}{p}s_{r+p}\ .$$

In this paper we prove for the Abel transformation Tauberian theorems in which the Tauberian conditions are of the form

$$\binom{n}{k}\Delta^k s_{n-k}=o(1)\ ,$$

or $O(1)$, or $O_L(1)$, as $n\to\infty$. For these theorems see specially § 5.

2. Some properties of Hausdorff and Hölder transforms. For all sequences appearing in this paper the index denoting the order of the terms will assume the values $0, 1, 2, \cdots$. If, in some formulae in this paper, a term appears with a negative value of the index denoting the order of the term, then we shall understand that this term assumes the value zero.

We say that a sequence $\{t_n\}$ is a *Hausdorff transform*, generated by the sequence $\{\mu_n\}$, of the sequence $\{s_n\}$, if

$$(1)\qquad t_n\equiv\sum_{m=0}^{n}\binom{n}{m}(\Delta^{n-m}\mu_m)s_m$$

for $n=0, 1, 2, \cdots$. A Hausdorff transform generated by a sequence $\{\mu_n\}$ will be called here, for shortness, a $(\mathfrak{H}, \mu_n)$ transform.

It is known that a necessary and sufficient condition for a sequence $\{t_n\}$ to be a $\{\mathfrak{H}, \mu_n\}$ transform of $\{s_n\}$ is the existence of

$$(2)\qquad \Delta^n t_0=\mu_n\cdot\Delta^n s_0$$

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

It is easy to see that, if $\{\lambda_n\}$ is defined by

$$(3)\qquad \lambda_n=\Delta^n\mu_0$$

for $n=0, 1, 2, \cdots$, where $\{\mu_n\}$ is an arbitrary sequence, then for each pair of nonnegative integers p and q

$$(4)\qquad \Delta^p\lambda_q=\Delta^q\mu_p\ .$$

If $\{\lambda_n\}$ is defined by (3) then (2) might be written in the form

$$\Delta^n t_0 = \Delta^n \lambda_0 \cdot \Delta^n s_0 \tag{5}$$

for $n=0, 1, 2, \cdots$. Equation (2) now shows that

$$t_n \equiv \sum_{m=0}^{n} \binom{n}{m} (\Delta^{n-m} \mu_m) s_m, \qquad n=0, 1, 2, \cdots,$$

is, by (4) and (5), equal to

$$\sum_{m=0}^{n} \binom{n}{m} (\Delta^m \lambda_{n-m}) s_m$$

which, by the symmetry of (5) in $\{\lambda_n\}$ and $\{s_n\}$, is equal to

$$\sum_{m=0}^{n} \binom{n}{m} (\Delta^m s_{n-m}) \lambda_m$$

$$= \sum_{m=0}^{n} \binom{n}{m} (\Delta^m \mu_0)(\Delta^m s_{n-m}),$$

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

Thus the $(\mathfrak{H}, \mu_n)$ transform of $\{s_n\}$ might be defined equivalently by

$$t_n \equiv \sum_{m=0}^{n} \binom{n}{m} (\Delta^m \mu_0)(\Delta^m s_{n-m}), \tag{6}$$

for $n=0, 1, 2, \cdots$; a fact which we use later.

We shall denote, in this paper, by $\{\mu_n^{(\alpha)}\}$, where α is an arbitrary fixed real number, the sequence $\{(n+1)^{-\alpha}\}$. The *Hölder transform of order* α, $\{h_n^{(\alpha)}\}$ (or, in short, the (H, α) transform) of a sequence, where α is a real number, is defined as the $(\mathfrak{H}, \mu_n^{(\alpha)})$ transform of the original sequence. We say that a sequence $\{s_n\}$ is *summable Hölder* to s if it is summable (H, α) to s for some real number α. We say that $\{s_n\}$ is *bounded Hölder* if it is bounded (H, α) for some real number α.

Let k be a fixed nonnegative integer. It is known that

$$\Delta^{k+1} \mu_n^{(-k)} = 0 \tag{7}$$

$$\Delta^k \mu_n^{(-k)} = (-1)^k \cdot k! \tag{8}$$

for $n=0, 1, 2, \cdots$; therefore, by (6),

$$h_n^{(-k)} = \sum_{m=0}^{k} (\Delta^m \mu_0^{(-k)}) \cdot \binom{n}{m} \cdot (\Delta^m s_{n-m}) \tag{9}$$

for $n=0, 1, 2, \cdots$. Equations (9) and (8) immediately yield the identity

$$(10)\quad \binom{n}{k}\cdot\Delta^k s_{n-k} = (-1)^{\binom{k+1}{2}}\cdot\left\{\prod_{p=0}^{k} p!\right\}^{-1}\cdot\begin{vmatrix} \mu_0^{(0)} & 0 & 0 & \cdots & 0 & h_n^{(0)} \\ \mu_0^{(-1)} & \Delta\mu_0^{(-1)} & 0 & \cdots & 0 & h_n^{(-1)} \\ \mu_0^{(-2)} & \Delta\mu_0^{(-2)} & \Delta^2\mu_0^{(-2)} & \cdots & 0 & h_n^{(-2)} \\ \cdot & \cdot & \cdot & \cdots & & \cdot \\ \cdot & \cdot & \cdot & \cdots & & \cdot \\ \cdot & \cdot & \cdot & \cdots & & \cdot \\ \mu_0^{(-k)} & \Delta\mu_0^{(-k)} & \Delta^2\mu_0^{(-k)} & \cdots & \Delta^{k-1}\mu_0^{(-k)} & h_n^{(-k)} \end{vmatrix}$$

for $n=0, 1, 2, \cdots$. If the determinant on the right side of (10) is expanded then we obtain

$$(11)\quad \binom{n}{k}\cdot\Delta^k s_{n-k} = \sum_{p=0}^{k} a_p^{(k)}\cdot h_n^{(-k+p)}$$

for $n=0, 1, 2, \cdots$; where, as is easy to see,

$$(12)\quad \sum_{p=0}^{k} a_p^{(k)} = 0\ ;\ a_0^{(k)} \neq 0\ ,$$

for $k=0, 1, 2, \cdots$. In the rest of this paper we shall denote by $a_0^{(k)}$, $\cdots$, $a_k^{(k)}$ the coefficients which appear in (11).

It is known that the Hölder transform of order α of the Hölder transform of order β of a sequence $\{s_n\}$ is identical with the Hölder transform of order $\alpha+\beta$ of $\{s_n\}$.

Let $\{\mu_n\}$ be defined by $\mu_n=\binom{n}{k}$, $n=0, 1, 2, \cdots$, where k is a fixed nonnegative integer. It is easy to see that

$$(13)\quad \Delta^p\mu_n = \begin{cases} (-1)^p\binom{n}{k-p} & \text{for } 0 \leqq p \leqq k \\ 0 & \text{for } p > k\ . \end{cases}$$

A consequence of (13) is that the sequence $\left\{\binom{n}{k}\Delta^k s_{n-k}\right\}$, $n=0, 1, 2, \cdots$, is a Hausdorff transform, generated by $\left\{(-1)^k\binom{n}{k}\right\}$, of the sequence $\{s_n\}$.

It is known that the product of two Hausdorff transformations is commutative; therefore, taking one the transformations to be that given by $\{h_n^{(\alpha)}\}$ and the other to be that given by $\left\{\binom{n}{k}\Delta^k s_{n-k}\right\}$ we obtain the following consequence of (11).

LEMMA 1. *Let α be a real number and k a nonnegative integer; then, for any sequence $\{s_n\}$,*

$$\binom{n}{k}\Delta^k h_{n-k}^{(\alpha)}=\sum_{p=0}^{k} a_p^{(k)}\cdot h_n^{(\alpha-k+p)}$$

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

3. A proposition concerning the product of two summability methods and three Tauberian theorems. We shall use later the following proposition (proved by O. Szász in [7]).

THEOREM D: *If* $\{s_n\}$ *is summable Abel to s and* $\{t_n\}$ *is a regular Hausdorff transform of* $\{s_n\}$; *then* $\{t_n\}$ *is summable Abel to s too,*

and the three theorems

THEOREM E. *If* $\{s_n\}$ *is summable Abel to s and* $\{s_n\}$ *is bounded, then* $\{s_n\}$ *is summable* (H, ε) *to s for each* $\varepsilon > 0$.

THEOREM F. *If* $\{s_n\}$ *is summable Abel to s and* $\{s_n\}$ *is bounded on one side, then* $\{s_n\}$ *is summable* $(H, 1)$ *to s.*

Theorem E may be deduced from Theorem 92 and Theorem 70 of [3], while Theorem F is Theorem 94 of the same book.

THEOREM G. *If* $f(x)$ *possesses a finite nth derivative,* $n \geqq 2$, *in the interval* $0 < x < 1$, *and if for some real number* α

$$f(x)=o((1-x)^{\alpha}), \qquad x \uparrow 1,$$

$$f^{(n)}(x)=O_L((1-x)^{\alpha-n}), \qquad x \uparrow 1,$$

then for all integers k satisfying $1 < k < n$,

$$f^{(k)}(x)=o((1-x)^{\alpha-k}), \qquad x \uparrow 1.$$

If, in Theorem G, we put $1-x=y^{-1}$, the theorem becomes a result first proved by N. Obrechkoff in [5] and subsequently generalized by M. Parthasarathy and C. T. Rajagopal in Theorems B and C of [6].

We shall now show the following proposition to be a consequence of Theorem G.

LFMMA 2. *Let the real sequence* $\{s_n\}$ *be summable Abel to s, that is*

$$\lim_{x\uparrow 1}(1-x)\sum_{n=0}^{\infty} s_n x^n = s . \tag{14}$$

If for some nonnegative integer k

$$(15)\qquad \binom{n}{k}\Delta^k s_{n-k} = O_L(1)\,, \qquad n\to\infty\,,$$

then for all integers p satisfying $0\leq p\leq k$,

$$(16)\qquad \lim_{x\uparrow 1}(1-x)^{-(k-p-1)}\cdot\sum_{n=p}^{\infty}\binom{n}{p}\cdot(\Delta^k s_{n-k})x^{n-p} = (-1)^{k-p}\binom{k-1}{p}s\,.$$

Proof. The identity

$$\sum_{n=0}^{\infty} s_n x^n = (-1)^r(1-x)^{-r}\cdot\sum_{n=0}^{\infty}(\Delta^r s_{n-r})x^n$$

for $r=0, 1, 2, \cdots$ combined with (14) yields (16) with $p=0$; that is

$$(17)\qquad \sum_{n=0}^{\infty}(\Delta^k s_{n-k})x^n \sim (-1)^k s(1-x)^{k-1}\,, \qquad x\uparrow 1\,.$$

Taking the kth derivative of the left side of (17) and using (15) we obtain

$$(18)\qquad \frac{d^k}{dx^k}\left\{\sum_{n=0}^{\infty}(\Delta^k s_{n-k})x^n\right. = k!\cdot\sum_{n=k}^{\infty}\binom{n}{k}(\Delta^k s_{n-k})x^{n-k}$$

$$= O_L\left(\sum_{n=k}^{\infty}x^{n-k}\right)\,, \qquad x\uparrow 1\,,$$

$$= O_L((1-x)^{-1})\,, \qquad x\uparrow 1\,.$$

The validity of (16), for all integers p satisfying $0<p<k$, follows now from (17) and (18) by an appeal to Theorem G with

$$f(x) = \sum_{n=0}^{\infty}\Delta^k s_{n-k}x^n - (-1)^k s(1-x)^{k-1}, \quad \alpha=k-1,\ n=k,\ k=p\,.$$

4. A Tauberian inequality for power series. In this section we prove one of the fundamental steps used in proving the main results of this paper. This step is the following.

THEOREM 1. *Let p be a fixed nonnegative integer. If for some real or complex sequence $\{s_n\}$,*

$$\overline{\lim_{n\to\infty}}\left|\binom{n}{p+1}\cdot\Delta^{p+1}s_{n-p-1}\right| \equiv S^{(p+1)} < +\infty\,,$$

then, for $x=1-(m+1)^{-1}$,

$$(19)\qquad \overline{\lim_{m\to\infty}}\left|-(1-x)^{-p}\cdot\Delta^p s_{m-p} - \sum_{r=0}^{p}(1-x)^{r-p}\cdot\sum_{n=0}^{\infty}\binom{n}{r}x^{n-r}\cdot\Delta^{p-1}s_{n-p-1}\right|$$

$$\leq \rho_p\cdot\overline{\lim_{n\to\infty}}\left|\binom{n}{p+1}\cdot\Delta^{p+1}s_{n-p-1}\right|\,,$$

where ρ_p is independent of $\{s_n\}$.

The case $p=0$ of Theorem 1 is well known. See for instance, inequality (15) of H. Hadwiger's paper [2].

The proof of Theorem 1 requires the following auxiliary proposition.

LEMMA 3. *For any pair m, n of integers satisfying $m \geqq 1$, $n \geqq 0$, and for $0 \leqq x \leqq 1$, we have*

$$0 \leqq 1-\sum_{p=0}^{m-1}\binom{n}{p}(1-x)^p x^{n-p} \leqq \binom{n}{m}(1-x)^m$$

where we suppose $\binom{n}{p}=0$ if $p>n$.

Proof. By the Taylor expansion

$$f(b)=f(a)+\frac{b-a}{1!}f'(a)+\cdots+\frac{(b-a)^{n-1}}{(n-1)!}f^{(n-1)}(a)+\frac{(b-a)^n}{n!}f^{(n)}(a+\theta(b-a)),$$

$$0<\theta<1,$$

we obtain, by choosing $b=1$, $a=1-x$ $(0 \leqq x \leqq 1)$ and $f(t)=t^n$,

$$1=\sum_{p=0}^{m-1}\binom{n}{p}(1-x)^p x^{n-p}+\binom{n}{m}(1-x)^m(x+\theta(1-x))^{n-m}, \quad 0<\theta<1.$$

Hence, for the stated values (in the theorem) of m, n and x,

$$0 \leqq 1-\sum_{p=0}^{m-1}\binom{n}{p}(1-x)^p x^{n-p} \leqq \binom{n}{m}(1-x)^m.$$

Proof of Theorem 1. We have

$$\begin{aligned}(20)\qquad & -(1-x)^{-p}\cdot\Delta^p s_{m-p}-\sum_{r=0}^{p}(1-x)^{r-p}\cdot\sum_{n=0}^{\infty}\binom{n}{r}x^{n-r}\cdot\Delta^{p+1}s_{n-p-1} \\ &=(1-x)^{-p}\sum_{n=0}^{m}\left\{1-\sum_{r=0}^{p}\binom{n}{r}(1-x)^r x^{n-r}\right\}\cdot\Delta^{p+1}s_{n-p-1} \\ &\quad -(1-x)^{-p}\sum_{n=m+1}^{\infty}\left\{\sum_{r=0}^{p}\binom{n}{r}(1-x)^r x^{n-r}\right\}\cdot\Delta^{p+1}s_{n-p-1} \\ &\equiv I_1+I_2.\end{aligned}$$

Lemma 3 yields

$$|I_1| \leqq (1-x)^{-p}\sum_{n=0}^{m}(1-x)^{p+1}\left|\binom{n}{p+1}\cdot\Delta^{p+1}s_{n-p-1}\right|.$$

Now, for $x=1-(m+1)^{-1}$,

$$\varlimsup_{m\to\infty} |I_1| \leqq \varlimsup_{n\to\infty} \left| \binom{n}{p+1} \cdot \Delta^{p+1} s_{n-p-1} \right| . \tag{21}$$

For each $\varepsilon > 0$ there exists an integer $m_0(\varepsilon)$ such that, for every $m > m_0(\varepsilon)$,

$$\left| \binom{m}{p+1} \cdot \Delta^{p+1} s_{m-p-1} \right| < S^{(p+1)} + \varepsilon .$$

We suppose now $m > m_0(\varepsilon)$; then

$$\begin{aligned} |I_2| &\leqq (S^{(p+1)} + \varepsilon) \cdot \sum_{n=m+1}^{\infty} \left\{ \sum_{r=0}^{p} \binom{n}{r} (1-x)^{-(p-r)} x^{n-r} \right\} \cdot \binom{n}{p+1}^{-1} \\ &= (p+1)(S^{(p+1)} + \varepsilon) \sum_{r=0}^{p} \binom{p}{r} (1-x)^{-(p-r)} \cdot \sum_{n=m+1}^{\infty} x^{n-r} \cdot \Delta^{p-r} (n-p)^{-1} . \end{aligned} \tag{22}$$

It is easy to show that for $0 \leqq r \leqq p$ we have

$$\begin{aligned} &(1-x)^{-(p-r)} \sum_{n=m+1}^{\infty} x^{n-r} \cdot \Delta^{p-r} (n-p)^{-1} \\ &= x^{m+1-r} \sum_{q=0}^{p-r-1} (-1)^q (1-x)^{-(p-r-q)} \cdot \Delta^{p-r-q-1} (m+1-q-p)^{-1} \\ &\quad + (-1)^{p-r} \sum_{n=m+1+p-r}^{\infty} (n-p)^{-1} x^{n-p} , \end{aligned} \tag{23}$$

and for $r = p$

$$(1-x)^{-(p-r)} \sum_{n=m+1}^{\infty} x^{n-r} \cdot \Delta^{p-r} (n-p)^{-1} = \sum_{n=m+1}^{\infty} (n-p)^{-1} x^{n-p} . \tag{24}$$

If we choose $x = 1 - (m+1)^{-1}$ and apply (23) and (24) to (22) we infer easily that, for $p \geqq 0$, there exists a positive constant λ_p which is independent of the sequence $\{s_n\}$ and such that

$$\varlimsup_{m\to\infty} |I_2| \leqq \lambda_p \cdot (S^{(p+1)} + \varepsilon) .$$

Since $\varepsilon > 0$ is chosen arbitrarily we infer that, for $x = 1 - (m+1)^{-1}$,

$$\varlimsup_{m\to\infty} |I_2| \leqq \lambda_p \cdot S^{(p+1)} . \tag{25}$$

Combining (20), (21) and (25) we see that our proposition is proved.

A consequence of Theorem 1 which will be used later is the following proposition.

LEMMA 4. *Let $\{s_n\}$ be summable Abel to s, and let there be a fixed positive integer k such that*

$$\binom{n}{k} \Delta^k s_{n-k} = O(1), \qquad n \to \infty . \tag{26}$$

Then (i) $\binom{n}{p}\Delta^p s_{n-p}=o(1)$, $n\to\infty$, *for* $1\leqq p\leqq k$, (ii) $\{s_n\}$ *is convergent to* s.

Proof. If $k=1$, we have to prove conclusion (ii) alone, and this follows from Theorem 1 with $p=0$. If $k\geqq 2$, then, by Theorem 1 and (26), for $x=1-(m+1)^{-1}$,

$$(27)\qquad \overline{\lim_{m\to\infty}}\left|-(1-x)^{-k+1}\cdot\Delta^{k-1}s_{m-k+1}-\sum_{r=0}^{k-1}(1-x)^{r-k+1}\sum_{n=0}^{\infty}\binom{n}{r}x^{n-r}\cdot\Delta^k s_{n-k}\right|=0\,.$$

The Abel summability of $\{s_n\}$, (26) and Lemma 3 show that

$$(28)\qquad \lim_{x\uparrow 1}\sum_{r=0}^{k-1}(1-x)^{r-k+1}\sum_{n=0}^{\infty}\binom{n}{r}x^{n-r}\cdot\Delta^k s_{n-k}=\sum_{r=0}^{k-1}(-1)^{k-r}\binom{k-1}{r}\cdot 0$$
$$=(-1)^k\cdot 0\cdot(1-1)^{k-1}$$
$$=0\,.$$

(28) and (27) show, for $x=1-(m+1)^{-1}$, that

$$\lim_{m\to\infty}|(1-x)^{-(k-1)}\cdot\Delta^{k-1}s_{m-(k-1)}|=0\,.$$

The last fact shows, immediately, that

$$\binom{n}{k-1}\cdot\Delta^{k-1}s_{n-(k-1)}=o(1)\,,\qquad n\to\infty\,.$$

Thus we reduced k in (26) by one, and by such a reduction (repeated if necessary) prove conclusion (i). Finnally we derive conclusion (ii) from conclusion (i) as already stated.

5. Some Tauberian theorems. The main result of this paper is the following.

THEOREM 2. *Be k some fixed positive integer. A necessary and sufficient condition for $\{s_n\}$ to be summable (H, k) is that $\{s_n\}$ should be summable Abel to s and* $\lim_{n\to\infty}\binom{n}{k}\cdot\Delta^k s_{n-k}=0$.

Proof. Proof of the sufficiency part. From the convergence of $\{s_n\}$ to s and the relations $\binom{n}{p}\Delta^p s_{n-p}=o(1)$, $n\to\infty$, for $p=1,\cdots,k$ (from Lemma 4) rewritten in the form (11), we get

$$\lim_{n\to\infty}h_n^{(-p)}=s$$

for $p=1, 2, \cdots, k$, successively; which proves the sufficiency part of the theorem. The proof of the necessity part of our proposition follows from (11) and the fact that the limits

$$\lim_{n\to\infty} h_n^{(-k)},\ \lim_{n\to\infty} h_n^{(-k+1)},\ \cdots,\ \lim_{n\to\infty} h_n^{(0)}$$

exist and are all equal to s.

Now we prove three interesting consequences of Theorem 2.

THEOREM 3. *A necessary and sufficient condition for a sequence $\{s_n\}$ to be summable (H, α), for some real value of α, is that $\{s_n\}$ should be summable Abel and that the sequence*

$$\left\{\binom{n}{k} \Delta^k s_{n-k}\right\}, \qquad n=0, 1, 2, \cdots$$

should be summable $(H, \alpha+k)$ to zero for some fixed positive integer k.

Proof. The necessity of the Abel summability of $\{s_n\}$ is obvious. The necessity of the $(H, \alpha+k)$ summability of

$$\left\{\binom{n}{k} \Delta^k s_{n-k}\right\}, \qquad n=0, 1, 2, \cdots,$$

to zero follows from Lemma 1 (if we replace α there by $\alpha+k$). Thus we have proved the necessity part of our theorem. The sufficiency part of our theorem is proved as follows. Suppose, first, that $\alpha \geqq -k$. Then, by Theorem D, the sequence

$$\{h_n^{(\alpha+k)}\}, \qquad n=0, 1, 2, \cdots,$$

is summable Abel to the same sum as the original sequence $\{s_n\}$, hence, using Theorem 2 with $\{h_n^{(\alpha+k)}\}$ instead of $\{s_n\}$, which is justified by Lemma 1 with α replaced by $\alpha+k$, $\{s_n\}$ is summable (H, α); which proves the sufficiency part of our theorem for $\alpha \geqq -k$. In the case $\alpha < -k$,

$$\left\{\binom{n}{k} \Delta^k s_{n-k}\right\}$$

being summable $(H, \alpha+k)$ to zero, is necessarily convergent to zero; and so, by Theorem 2, $\{s_n\}$ is summable $(H, -k)$, or $\{h_n^{(\alpha)}\}$ is summable $(H, -\alpha-k)$, and consequently summable Abel too. Thus, by Theorem D, $\{h_n^{(\alpha+k)}\}$ is also summable Abel and the proof can be completed as in the case $\alpha \geqq -k$.

The case $k=1$ is a special case of Theorem (9.4) of [1], with $\beta=\alpha+1$ there.

THEOREM 4. *Be k an arbitrary fixed nonnegative integer. If a* *sequence* $\{s_n\}$ *is summable Abel to s and the sequence*

$$\left\{\binom{n}{k}\Delta^k s_{n-k}\right\}$$

is bounded $(H, \alpha+k)$ *for some real number* α, *then* $\{s_n\}$ *is summable* $(H, \alpha+\varepsilon)$ *for each* $\varepsilon>0$.

The case $k=1$ of the last theorem is the special case $\beta=\alpha+1$ of Theorem (9.5) (for Abel summability) of [1].

Proof. First suppose $\alpha\geqq 0$. Then, by Theorem D, (11) and (12),

$$v_n \equiv \sum_{p=0}^{k} a_p^{(k)} h_n^{(\alpha+p)} = O(1), \qquad n\to\infty,$$

and $\{v_n\}$ is summable Abel to zero. Therefore, by Theorem E, $\{v_n\}$ is summable (H, ε), for each $\varepsilon>0$, to zero, or $\left\{\binom{n}{k}\Delta^k s_{n-k}\right\}$ is summable $(H, \alpha+k+\varepsilon)$ to zero, and the conclusion follows by Theorem 3. If $\alpha<0$, we apply the preceding argument to the $(H, -\alpha)$ transform of $\{v_n\}$ which is clearly $O(1)$, as $n\to\infty$, and summable Abel to zero. Thus we find that the $(H, -\alpha)$ transform of $\{v_n\}$ is summable (H, ε) to zero, for each $\varepsilon>0$, or that $\{v_n\}$ in summable $(H, -\alpha+\varepsilon)$ to zero and hence summable Abel to zero. Since $v_n=O(1)$, $\{v_n\}$ is, by Theorem E, summable (H, ε) to zero and the proof is completed exactly as in the case $\alpha\geqq 0$.

THEOREM 5. *Be k an arbitrary fixed positive integer. If a sequence* $\{s_n\}$ *is summable Abel to s and the sequence*

$$\left\{\binom{n}{k}\Delta^k s_{n-k}\right\}$$

is bounded $(H, \alpha+k)$ *on one side, then* $\{s_n\}$ *is summable* $(H, \alpha+1)$ *to s.*

The case $k=1$ is the special case $\beta=\alpha+1$ of Theorem (9.6) of [1].

The proof of Theorem 5 is exactly the same as that of Theorem 4. But now we have to use Theorem F in place of Theorem E.

In conclusion I wish to thank Professor C. T. Rajagopal for helpful suggestions.

REFERENCES

1. A. Amir (Jakimovski), *Some relations between the methods of summability of Abel, Borel, Cesàro, Hölder and Hausdorff*, J. Analyse Math. **3** (1953/4). 346-381.

2. H. Hadwiger, *Über ein Distanztheorem Bei der A Limitierung*, Comm. Math. Helv.,

16 (1943/4), 209–214.
3. G. H. Hardy, *Divergent Series*, Oxford University Press.
4. Kien Kwong Chen, *Some one-sided Tauberian theorems*, Anais Acad. Brasil, Ci **17** (1945), 249–259.
5. N. Obrechkoff, *Sur une formule pour be différences divisées et de leur dérivéees*, C. R. Acad. Bulgare Sci., **2** (1949), No. 1, 5–8.
6. M. Parthasaraty and C. T. Rajagopal, *A theorem on the Riemann-Liouville integral*, Math. Z., **55** (1951), 84–91.
7. O. Szász, *On the product of two summability methods*, Ann. Soc. Pol. Math., **25** (1952), 75–84 (1953).

THE UNIVERSITY OF TEL-AVIV,

SIMPLIFIED PROOFS OF "SOME TAUBERIAN THEOREMS" OF JAKIMOVSKI

C. T. RAJAGOPAL

1. Introduction. In this note, the preceding paper (mentioned in the title) will be referred to as [J], the papers or books numbered 1, 2, $\cdots$ in the bibliography concluding [J] will be referred to as [J1], [J2], $\cdots$, while those in the numbered list of references at the end will be referred to by their numbers in square brackets.

The notation in [J] is retained with a slight simplification as follows. As in Hardy's *Divergent series* [J3], a sequence $\{t_n\}$ is called a Hausdorff transform of another sequence $\{s_n\}$ when there is a sequence $\{\mu_n\}$ such that

$$(1) \qquad \Delta^n t_0 = \mu_n \Delta^n s_0 .$$

If α is a real number, the special case of $\{t_n\}$ defined by (1) with $\mu_n = (n+1)^{-\alpha}$, called the (H, α) transform, will be denoted by $H^\alpha s$ where s denotes the sequence $\{s_n\}$. Since two Hausdorff transformations are commutable, the operator H^α is such that $H^\alpha H^\beta = H^\beta H^\alpha = H^{\alpha+\beta}$ and H^0 is the identity operator.

From the Abel or (A) transform of $\{s_n\}$, defined as the left-hand member of

$$(2) \qquad (1-x)\sum_0^\infty s_n x^n = (-1)^p (1-x)^{-p+1} \sum_0^\infty \Delta^p s_{n-p} x^n ,$$

$$0 < x < 1,\ p=1, 2, 3, \cdots,$$

we deduce the equality (2) by induction on p. It is in the form of the right-hand member of (2) that the (A) transform is used in this note.

For any sequence $\{s_n\}$, summability (H, α) to a finite value l and summability (A) to l have their usual meanings as in [J].

2. The fundamental theorem in [J]. This theorem ([J], Theorem 2) may be restated as follows with its non-trivial parts separated, so that Tauber's first theorem ([J3], Theorem 85) emerges as the case $k=1$ of the first part, with the conclusion of the convergence of $\{s_n\}$ restated as that of the (H, -1) summability of $\{s_n\}$.

THEOREM 1. (a) *If* (i) $\{s_n\}$ *is summable* (A) *to* l, (ii) *for a positive integer* k, $n^k \Delta^k s_{n-k} = o(1)$, $n \to \infty$, *then* $\{s_n\}$ *is summable* (H, $-k$) *to* l.

Received June 18, 1956.

(b) *Conditions* (i) *and* (ii) *are also necessary for* $\{s_n\}$ *to be summable* (H, $-k$) *to* l.

For establishing this theorem, Jakimovski's tools are (1) the Tauberian technique embodied in Lemma 2 of this note with the additional complications necessary to bring in $n^p \Delta^p s_{n-p}$ to take the place of $n\Delta^p s_{n-1}$, (2) the technique of repeated differences (or differentiation) implicit in his appeal to one particular case of a theorem proved by Parthasarathy and Rajagopal ([J6], case $k=l+r$ of Theorem C). However, the second technique, while generally useful in proving Tauberian theorems of the Hardy-Littlewood class, is not required at all for proving the original Tauberian theorems; and it is perhaps not very satisfactory to use it to prove Theorem 1 which is essentially of the latter class of theorems. The present note supplies a new proof of Theorem 1 whose merit is that it depends only on Lemma 2 as it stands and on the interpretation, in Lemma 1, of $n(n-1)\cdots(n-p+1)\Delta^p s_{n-p}$, which is asympotically equal to $n^p \Delta^p s_{n-p}$, as a Hausdorff transform of s_n. Although the content of Lemma 1 is due to Jakimovski, the proof of Lemma 1 as it appears here is a simplification of his proof, resulting from the symbolic representation (5) of the Hausdorff transformation of s_n in question, suggested to me by Mr. M. R. Parameswaran.

LEMMA 1. *If k is a positive integer and*

$$t_n \equiv \binom{n}{k} \Delta^k s_{n-k}, \tag{3}$$

then t_n is related to s_n by (1) *with*

$$\mu_n = (-1)^k \binom{n}{k}, \tag{4}$$

that is, $\{t_n\}$ in (3) *is the Hausdorff transform of $\{s_n\}$ corresponding to the $\{\mu_n\}$ defined by* (4), *and further we have symbolically*

$$\{t_n\} \equiv \frac{(-1)^k}{k!} H^{-k}(H^0-H^1)(H^0-2H^1)\cdots(H^0-kH^1)s \tag{5}$$

$$\equiv \left(\sum_{r=0}^{k} a_r^{(k)} H^{-k+r}\right)s, \quad \sum_{r=0}^{k} a_r^{(k)} = 0,$$

the order of factors in (5) *being immaterial.*

Here I must record may indebtedness to Dr. Jakimovski who has pointed out an implication of the first part of (5), namely, that

$$\binom{n}{k}\Delta^k s_{n-k} = \frac{(-1)^k}{k!}\sum_{r=0}^{k} S_{k+1}^{k-r+1} H^{-k+r} s\,,$$

where S_n^m are Stirling's numbers of the first kind ([1], p. 142, (3)).

Proof. The relation between s_n and t_n is proved directly, starting from

$$\Delta^n t_0 = \sum_{r=0}^{n} (-1)^r \binom{n}{r} t_r\,,$$

and showing that substitution for t_r from (3) leads to (1) with the $\{\mu_n\}$ in (4).

Equation (5) follows from the fact that (4) can be written:

$$\mu_n = \frac{(-1)^k}{k!}(n+1)^k\left(1-\frac{1}{n+1}\right)\left(1-\frac{2}{n+1}\right)\cdots\left(1-\frac{k}{n+1}\right).$$

Now the factors $(n+1)^k$, $1-(n+1)^{-1}$, $1-2(n+1)^{-1}$, $\cdots$, $1-k(n+1)^{-1}$, taken successively instead of μ_n in (1), make the $\{t_n\}$ of $\{1\}$ the Hausdorff transforms of $\{s_n\}$ corresponding to the operators H^{-k}, H^0-H^1, H^0-2H^1, $\cdots$, H^0-kH^1 respectively. Hence the $\{t_n\}$ of (3) is the product of the several Hausdorff transforms last mentioned multiplied by $(-1)^k/k!$. We thus have the representation in (5) of the $\{t_n\}$ in (3), and we can take the factors in this representation in any order since Hausdorff transformations are commutable.

LEMMA 2. *If $\{s_n\}$ is such that $n\Delta^1 s_{n-1}=O(1)$, then, for $x=1-n^{-1}$,*

$$\limsup_{n\to\infty}\left|\sum_{r=0}^{n}\Delta^1 s_{r-1} - \sum_{r=0}^{\infty}\Delta^1 s_{r-1}x^r\right| \leqq \tau \limsup_{n\to\infty} |n\Delta^1 s_{n-1}|$$

where τ is the 'Tauberian constant':

$$\tau = C + 2\int_1^\infty e^{-x}x^{-1}\,dx\,, \qquad C=\textit{Euler's constant}.$$

This result, due to Hadwiger ([J2], inequality (15)), is a particular case of a more general result (e.g. [2], case $\alpha=1$, $\lambda_n=n$, $\phi^*(u)=e^{-u}$ of Theorem 2(b)).

Proof of Theorem 1. (a) We may suppose without loss of generality that $l=0$. For, we have only to consider, instead of $\{s_n\}$, the new sequence $\{s_n-l\}$ which is clearly subject to hypothesis (i) with $l=0$ and also hypothesis (ii).

First, we take $\Delta^k s_{n-k}$ instead of $\Delta^1 s_{n-1}$ in Lemma 1 and obtain, for $x=1-n^{-1}$,

$$(6)\ \limsup_{n\to\infty}(1-x)^{-k+1}\left|\sum_{r=0}^{n}\Delta^k s_{r-k}-\sum_{r=0}^{\infty}\Delta^k s_{r-k}x^r\right|\leqq\tau\limsup_{n\to\infty}(1-x)^{-k+1}|n\Delta^k s_{n-k}|$$
$$=\tau\limsup_{n\leftarrow\infty}|n^k\Delta^k s_{n-k}|\ .$$

Next, we take $p=k$ in (2) and get

$$(7)\qquad (-1)^k(1-x)^{-k+1}\sum_{r=0}^{\infty}\Delta^k s_{r-k}x^r=(1-x)\sum_{r=0}^{\infty}s_r x^r=o(1)\ ,\qquad x\to 1-0,$$

as a result of hypothesis (i) where $l=0$ according to our supposition. Using in (6) hypothesis (ii) and (7) with $x=1-n^{-1}$, we obtain

$$n^{k-1}\Delta^{k-1}s_{n-k+1}=-(1-x)^{-k+1}\sum_{r=0}^{n}\Delta^k s_{r-k}=o(1),\qquad n\to\infty.$$

If $k=1$, we infer at once that s_n converges to 0. If $k\geqq 2$, we repeat the foregoing argument with $k-1$, $k-2$, $\cdots$, 1 successively in place of k and find that $n^p\Delta^p s_{n-p}=o(1)$ for $p=k-2$, $k-3$, $\cdots$, 0, thus finally drawing the same inference as before. After this we use the fact, following from $n^p\Delta^p s_{n-p}=o(1)$, $1\leqq p\leqq k$, taken along with (5), that

$$(8)\qquad \frac{(-1)^p}{p!}H^{-p}(H^0-H^1)(H^0-2H^1)\cdots(H^0-pH^1)s\equiv\binom{n}{p}\Delta^p s_{n-p}=o(1)$$

as $n\to\infty$ for $p=1, 2, \cdots, k$, and prove successively that $H^{-1}s$, $H^{-2}s$, $\cdots$, $H^{-k}s$ all converge to $0=l$.

(b) If $\{s_n\}$ is summable (H, $-k$) to l, then $H^{-p}s$, $p=k, k-1,\cdots, 0$, are obviously each convergent to l and (8) necessarily holds for $p=k$; also $\{s_n\}$, being convergent to l, is necessarily summable (A) to l.

3. Remarks on other theorems in [J]. It may be pointed out how (5) in conjunction with the notation of this note simplifies the presentation of Jakimovski's main theorems ([J], Theorems 3,5) restated in this notation as Theorems 2,3. The simplified presentation, like the one given by Jakimovski, depends only on the results of the preceding section, O. Szász's theorem for the product of a regular Hausdorff method of summability and (A) summability ([J], Theorem D, generalized by Rajagopal in [3], Theorem I), and finally an idea whose simplest expression is the lemma which follows.

LEMMA 3. *If* $\{s_n\}$ *is summable* (A) *to* l *and the sequence denoted by* $H^\alpha s$, *where* α *is any real number, is bounded on one side, then* $H^{\alpha+1}s$ *is convergent to* l.

The case $\alpha=0$ of Lemma 4 is classical. The case $\alpha\neq 0$ is included in one of Jakimovski's theorems ([J4], Theorem (9.6)). However, it is best to deduce it from the case $\alpha=0$ by means of the following observation. If $\alpha>0$, then $H^{\alpha}s$ is summable (A) to l by Szász's product-theorem referred to above; while, if $\alpha<0$, $H^{\alpha}s$ is again summable (A) to l since it is summable (H, $-\alpha+1$) to l as a result of $s\equiv H^{-\alpha}(H^{\alpha}s)$ being bounded on one side and summable (A) to l.

In Lemma 3 we extend a Tauberian theorem for sequences s summable (A) by replacing s by $H^{\alpha}s$ in the Tauberian hypothesis and the conclusion. The method of extension shows that, in Theorem 1 (a), we may replace s by $H^{\alpha}s$, or, α being any real number, replace s by $H^{\alpha+k}s$, in hypothesis (ii) and the conclusion. The result of the replacement of s by $H^{\alpha+k}s$ is stated below.

THEOREM 2. (a) *If* (i) *the sequence* $\{s_n\}$ *is summable* (A) *to* l, (ii) *for a real number* α *and a positive integer* k, *the sequence* $H^{\alpha+k}t$ *is null, where* $t\equiv\{t_n\}$ *is defined by* (3) *or* (5), *then* $\{s_n\}$ *is summable* (H, α) *to* l.

(b) *Conditions* (i) *and* (ii) *are also clearly necessary for* $\{s_n\}$ *to be summable* (H, α) *to* l.

An immediate deduction from Theorem 2 is the next.

THEOREM 3. *If, in Theorem* 2(a), *condition* (ii) *is replaced by the condition that* $H^{\alpha+k}t$ *is bounded on one side, the conclusion will be that* $\{s_n\}$ *is summable* (H, $\alpha+1$) *to* l.

Proof. By Szász's product-theorem, $H^{k}t$ is summable (A) to 0. Hence, by Lemma 3 with $H^{k}t$ instead of s, $H^{\alpha+1+k}t$ is a null sequence, and the conclusion follows from Theorem 2(a) with $\alpha+1$ instead of α.

4. Addition. (November 23, 1956.) Szász ([4], p. 1019, Lemma 5) has proved the following theorem.

THEOREM X. *Let* $\{s_n\}$ *be a sequence which is* (i) *summable* (A) *to* l, (ii) *bounded below and quasi-monotonic-decreasing in the sense that there is a constant* $c>0$ *such that*

$$s_{n+1}\leqq(1+c/n)s_n\,, \qquad n>n_0(c).$$

Then $\{s_n\}$ *is convergent to* l.

Appealing to Lemma 3, we can replace $s\equiv\{s_n\}$ by $H^{\alpha}s$ in the hypothesis (ii) and the conclusion of Theorem X, and obtain the following theorem.

THEOREM Y. *Let* s *be a sequence such that* (i) *it is summable* (A) *to*

l, (ii) *its transform* $H^\alpha s$ *is bounded below and quasi-monotonic-decreasing according to the definition in Theorem* X. *Then* s *is summable* (H, α) *to* l.

The cases $\alpha=0$, $\alpha=-1$ of Theorem Y have applications to trigonometric series ([4]: p. 1020, Theorem 3 and p. 1031, Theorem 8).

REFERENCES

1. C. Jordan, *The calculus of finite differences* (Chelsea Reprint, 1950).
2. C. T. Rajagopal, *A generalization of Tauber's theorem and some Tauberian constants*, Math. Z., **57** (1953), 405-414.
3. ———, *Theorems on the product of two summability methods with applications*, J. Indian Math. Soc., **18** (1954), 89-105.
4. O. Szász, *Collected mathematical papers* (Cincinnati, 1955).

RAMANUJAN INSTITUTE OF MATHEMATICS
MADRAS, INDIA.

A CONGRUENCE THEOREM FOR TREES

PAUL J. KELLY

Let A and B be two trees with vertex sets $a_1, a_2, \cdots, a_n$ and $b_1, b_2, \cdots, b_n$ respectively. The trees are congruent, are isomorphic, or "are the same type", $(A \cong B)$, if there exists a one-to-one correspondence between their vertices which preserves the join-relationship between pairs of vertices. Let $c(a_i)$ denote the $(n-1)$-point subgraph of A obtained by deleting a_i and all joins (arcs, segments) at a_i from A. It is the purpose here to show that if there is a one-to-one correspondence in type, and frequency of type, between the sub-graphs of order $n-1$ in A and B, that is, if there exists a labeling such that $c(a_i) \cong c(b_i)$, $i=1, 2, \cdots, n$, then $A \cong B$. It is assumed throughout, therefore, that there is a labeling of the two trees A and B such that $c(a_i) \cong c(b_i)$, $i=1, 2, \cdots, n$, where $n \geqq 3$.

Some lemmas to the main theorem are established first. Let T denote a certain type of graph of order j, where $2 \leqq j < n$, which occurs as a subgraph α times in A and β times in B. If α_i is the number of T-type subgraphs which have a_i as a vertex, then,

$$\alpha = \left(\sum_1^n \alpha_i \right) \Big/ j.$$

Similarly,

$$\beta = \left(\sum_1^n \beta_i \right) \Big/ j,$$

where b_i is the number of T-type subgraphs having b_i as a vertex. Because $c(a_i) \cong c(b_i)$, the number of T-type subgraphs which do not have a_i as a vertex is the same as the number which do not have b_i as a vertex. Thus

$$\alpha - \alpha_i = \beta - \beta_i, \qquad i=1, 2, \cdots, n.$$

Therefore

$$\sum_1^n (\alpha - \beta) = \sum_1^n (\alpha_i - \beta_i),$$

so $n(\alpha-\beta) = j(\alpha-\beta)$, which implies $\alpha=\beta$. This, in turn, implies $\alpha_i=\beta_i$, $i=1, 2, \cdots, n$, and the lemma is established.

LEMMA 1. *Every type of proper subgraph which occurs in A or B*

Received December 16, 1955.

occurs the same number of times in both, and a_i and b_i are vertices in the same number of these subgraphs, $i=1, 2, \cdots, n$.

The case $j=2$ gives a special result.

LEMMA 2. *The vertices a_i and b_i have the same degree, $i=1, 2, \cdots, n$.*

Next it is clear that if either A or B consists of just a path between two end points then the other is also a path of the same length. If neither is just a path, then their maximal-length paths are proper subgraphs and have the same length because of Lemma 1.

This proves the third lemma.

LEMMA 3. *The trees A and B have the same radius r and both trees are central or both are bicentral.*

A correspondence between $c(a_i)$ and $c(b_i)$, under which $c(a_i) \cong c(b_i)$, will be called an a_i-mapping (or b_i-mapping), and the main theorem is obtained by using these submappings to define a congruence of A and B. Because such a congruence is more easily obtained when the trees are central, the proof will be carried through for bicentral trees only, with the simpler proof implied by analogy. It is supposed therefore that A has bicenters $\bar{a}_1$ and $\bar{a}_2$ and that B has bicenters $\overline{b_1}$ and $\overline{b_2}$ (where $\bar{a}_1$ is not necessarily a_1).

Let F be a component in the graph obtained by deleting from A the bicenters and all joins to them. There is a point of F joined in A to one bicenter, say $\bar{a}_1$, and no point of F is joined in A to $\bar{a}_2$. By $(\bar{a} \cup F)$ is meant the graph, which has for its vertices $\bar{a}$, and the vertices of F, and whose joins are the same as they are in A. The graph $(\bar{a}_1 \cup F)$ is a *limb at* $\bar{a}_1$. It is a *radial* or *nonradial* limb according as it does not possess an *r-point*, that is, a point whose distance in A from the nearest bicenter is r. An easy consequence of Lemma 1 is that a_i is an r-point if and only if b_i is an r-point.

Some special subgraphs of A and B are now defined. At $\bar{a}_i$ the radial limbs are

$$A_{i1},\ A_{i2},\ \cdots,\ A_{im_i}\,,$$

and the non-radial limbs are

$$C_{i1},\ C_{i2},\ \cdots,\ C_{is_i}\,,$$

while at $\overline{b_i}$ the radial limbs are

$$B_{i1},\ B_{i2},\ \cdots,\ B_{in_i}$$

and the non-radial limbs are

$$D_{i1},\ D_{i2},\ \cdots,\ D_{it_i}, \qquad i=1,\ 2\,.$$

Next,

$$A_i=(A_{i1}\cup A_{i2}\cup\cdots\cup A_{im_i})\,,\qquad B_i=(B_{i1}\cup B_{i2}\cup\cdots\cup B_{im_i})\,,$$
$$C_i=(C_{i1}\cup C_{i2}\cup\cdots\cup C_{is_i})\,,$$

and

$$D_i=(D_{i1}\cup D_{i2}\cup\cdots\cup D_{it_i})\,, \qquad i=1,\ 2.$$

Finally,

$$A_r=(A_1\cup A_2),\ B_r=(B_1\cup B_2),\ C=(C_1\cup C_2),$$

and

$$D=(D_1\cup D_2)\,.$$

In obtaining congruences for these special subgraphs, an important role is played by center preserving mappings, that is, those which pair $\bar{a}_1$ and $\bar{a}_2$ in some order with $\bar{b}_1$ and $\bar{b}_2$. It is useful, therefore, to define a vertex a_i to be a *nonessential point*, (n. e. point), if it is of degree one (is an end point) such that $c(a_i)$ is a bicentral tree of radius r. Every end point, which is not an r-point, is an n. e. point. An r-point is nonessential if it belongs to a limb with more than one r-point, or if the bicenter to which its limb belongs has more than one radial limb. If a_i is an n. e. point then b_i is an n. e. point and every a_i-mapping is center preserving. The following fact is also useful.

LEMMA 4. *If a_i is an* n. e. *point of A in A_r then b_i is an* n. e. *point of B in B_r.*

Proof. Assume $b_i\notin B_r$, that is, $b_i\in D$. Any a_i-mapping must pair the remainder of A_r (without a_i) with all of B_r, so the order of A_r is one greater than that of B_r. If A had a nonradial limb it would have an n. e. point in C, say a_j, and an a_j-mapping would have to pair A_r with all or part of B_r, which is impossible. Therefore A has no nonradial limb and b_i is the only point of B not in B_r. The sum of the degrees of $\bar{a}_1$ and $\bar{a}_2$ is therefore smaller than the sum of the degrees of $\bar{b}_1$ and $\bar{b}_2$. If B had an n. e. point b_l, distinct from b_i, the sum of the degrees of $\bar{b}_1$ and $\bar{b}_2$ would be the same in $c(b_l)$ as in B, and therefore a b_l-mapping could not be center preserving. From this it follows that a_i and b_i are the only n. e. points in A and B respectively. Thus

A consists of a $(2r+1)$-path and one extra point a_i joined to a point a_k, which is not a center, while B consists of a $(2r+1)$-path and one extra point b_i joined to a center. This center is b_k since a_k is the only point in A of degree three. But now it is clear that $c(a_k)$ has a component which is a path of greater length than any in $c(b_k)$, which contradicts $c(a_k) \cong c(b_k)$. The assumption that b_i is in D is therefore false and Lemma 4 is established.

THEOREM. *If A and B are trees with vertices $a_1, a_2, \cdots, a_n$ and $b_1, b_2, \cdots, b_n$, $n \geqq 3$, respectively, and $c(a_i) \cong c(b_i)$, $i=1, 2, \cdots, n$ then $A \cong B$.*

Proof. As previously indicated, the details will be given only for the case where A and B are bicentral.

Case 1. One of the trees, say A, has a nonradial limb. Then A has an n.e. point a_k in C and, from Lemma 4, b_k is in D. An a_k-mapping, therefore, pairs A_r with B_r, so

(1) $$A_r \cong B_r .$$

Next,

(2) There is a congruence of C and D which pairs $\bar{a}_1$, $\bar{a}_2$ in some order with $\bar{b}_1$ and $\bar{b}_2$.

Consider the n. e. points of A in A_r. First, suppose the limb to which one of these points a_i belongs is still of length r after a_i is deleted from it. Then an $a_i$mapping cannot take this sub-limb into D, and so must pair the remainder of A_r with the remainder of B_r. It therefore pairs C with D as stated in (2). Next, suppose every n.e. point which belongs to A_r or B_r is the end of an r-path limb. If a_i and b_i are such points, then deleting them from their limbs produces two $(r-1)$-path limbs. Since these sub-limbs are congruent, an a_i-mapping either pairs C and D as stated in (2) or else can be redefined to do so. The only remaining possibility is that no n.e. point occurs in either A_r or B_r, so each is a $(2r+1)$-path. Let a_i be the r-point in A_i and b_i be the r-point in B_i, $i=1, 2$. Since $C(a_1)$ is a tree of radius r and center $\bar{a}_2$ and $c(b_1)$ is a congruent tree with center $\bar{b}_2$, an a_1-mapping pairs $\bar{a}_2$ and $\bar{b}_2$. It must also pair the nonradial limbs of A at $\bar{a}_2$ with the nonradial limbs of B at $\bar{b}_2$, and hence $C_2 \cong D_2$. By the same reasoning, an a_2-mapping establishes a congruence of C_1 and D_1 which pairs $\bar{a}_1$ with $\bar{b}_1$, so there is clearly a congruence of C and D satisfying (2).

If a congruence of C and D, satisfying (2), and a congruence of

A_r and B_r both pair the bicenters in the same order, then clearly $A \cong B$. Assume, on the contrary, that every congruence of A_r with B_r pairs the bicenters in one order, say $\bar{a}_1$ with $\bar{b}_1$ and $\bar{a}_2$ with $\bar{b}_2$, while every congruence of C and D, satisfying (2), pairs the bicenters in the opposite order, namely $\bar{a}_1$ with $\bar{b}_2$ and $\bar{a}_2$ with $\bar{b}_1$. It will be shown that this leads to a contradiction and hence that $A \cong B$ under Case 1. First, from the assumption about the congruence of A_r and B_r, it follows that each is not just a $(2r+1)$-path. Therefore A has an n. e. point a_i in A_r, and it may be supposed that $a_i \in A_1$. Since an a_i-mapping implies a congruence of C and D satisfying (2) it must pair $\bar{a}_1$ and $\bar{a}_2$ with $\bar{b}_2$ and $\bar{b}_1$ in that order. By assumption, $A_1 \cong B_1$ and $A_2 \cong B_2$, so $b_i \in B_2$ would imply that an a_i-mapping pairs A_2 with B_1. But then $A_1 \cong A_2 \cong B_1 \cong B_2$ would contradict the unique mapping of bicenters in any congruence of A_r with B_r. Therefore $b_i \in B_1$. Let f_1 be the order of A_1 and B_1 and f_2 be that of A_2 and B_2. An a_i-mapping shows $f_2 = f_1 - 1$. Suppose A has an n. e. point a_j in A_2. Then an a_j-mapping pairs $\bar{a}_1$ and $\bar{b}_2$, so it pairs A_1 with all or part of B_2. But this is impossible because $f_1 > f_2$. Therefore there is no n. e. point of A in A_2, and, by the same reasoning, there is no n. e. point of B in B_2. Thus A_2 and B_2 are paths of length r, and A and B each have just two end points in A_1 and B_1 respectively.

Now consider nonradial limbs. At least one exists so, from Lemma 4, at least one each exists in each tree. Suppose there is a nonradial limb at a_1, and let a_j be an end point of A in this limb. Then $b_j \in D$. Because an a_j-mapping includes a congruence of A_r and B_r, it pairs a_1 with $\bar{b}_1$ and $\bar{a}_2$ with $\bar{b}_2$. If b_j were in D_1 such a mapping would imply $C_2 \cong D_2$, and this, with $C_1 \cong D_2$ and $C_2 \cong D_1$, would yield $C_1 \cong C_2 \cong D_1 \cong D_2$, contradicting the unique center pairings in a congruence of C and D. Therefore $b_j \in D_2$. Let f_3 be the order of C_1 and D_2 and f_4 by the order of C_2 and D_1. An a_j-mapping shows that $f_3 - 1 = f_4$. Therefore there is no n. e. point in C_2 and none in D_1. For if a_k in C_2 were an n. e. point, an a_k-mapping would pair $\bar{a}_1$ with $\bar{b}_1$ and therefore would pair C_1 with all or part of D_1. This is impossible because $f_3 > f_4$. There are, therefore, no nonradial limbs at $\bar{a}_2$ or $\bar{b}_1$, and there is just one nonradial limb at $\bar{a}_1$ and at b_2, each of length one. The center $\bar{a}_1$ and $\bar{b}_2$ are of degree three and $\bar{a}_2$ and $\bar{b}_1$ have degree two. Let a_1 be the end point of A in A_2. The tree $c(a_1)$ has only one center, namely $\bar{a}_1$ of degree three. If the r-point b_1 were in B_2, the center of $c(b_1)$ would have degree two, contradicting $c(a_1) \cong c(b_1)$. So $b_1 \in B_1$. Also b_1 is the only r-point in B_1, for otherwise $c(b_1)$ would be a bicentral tree. If a_2 and b_2 denote the other r-points of A and B respectively, it follows that $c(a_2)$ and $c(b_2)$ are central trees and that $\bar{a}_2$ and $\bar{b}_1$ are their respective centers. But in $c(a_2)$ the radial arm to a_1 is a path in $c(b_2)$ both radial arms branch, and this contradicts $c(a_2) \cong c(b_2)$. The supposition that a nonradial limb exists at $\bar{a}_2$ rather than $\bar{a}_1$ leads to the same kind of contradiction, hence $A \cong B$ under all the possibilities of Case 1.

Case 2. There are no nonradial limbs but one tree has at least three radial limbs. Suppose there are at least two radial limbs at $\bar{a}_2$, and now let $\bar{a}_1 = a_1$ and $\bar{b}_2 = b_2$. One and only one component of $c(a_1)$ is a central tree of radius r. Its center is $\bar{a}_2$ and all of its limbs are radial. Let b_1' be the center of the corresponding, congruent tree in $c(b_1)$. If b_1 is neither $\bar{b}_1$ or $\bar{b}_2$, then b_1' is a non-end point of B in some limb of B, say a limb at $\bar{b}_1$. There is a path P from $\bar{b}_1$, to b_1' and there also exists a path P' starting at b_1' and having no join in common with P. Since the length of P' must be less than r, all the limbs of the tree centered at b_1' cannot be radial. The supposition that b_1 is neither $\bar{b}_1$ or $\bar{b}_2$ is therefore false, and b_1 may be taken to be $\bar{b}_1$. Then $A_2 \cong B_2$ is implied by an a_1-mapping.

If there are at least two radial arms at either a_1 or b_1, the same reasoning shows that $A_1 \cong B_1$ and this, with $A_2 \cong B_2$, implies $A \cong B$. Suppose, then, that A_{11} and B_{11} are the only limbs at a_1 and b_1 respectively, and let the order of A_{11} be at least as great as that of B_{11}. There is an r-point a_j in A_{21} and it is an n. e. point. An a_j-mapping must pair a_1 with b_1 because these are of degree two while $\bar{a}_2$ and $\bar{b}_2$ are of degree at least three. The mapping therefore pairs A_{11} with all or part of B_{11}, and since the latter case is excluded by the orders of A_{11} and B_{11}, it follows that $A_{11} \cong B_{11}$. This, with $A_2 \cong B_2$, implies $A \cong B$ and completes Case 2.

Case 3. Each tree has exactly two limbs. Let n_i be the order of A_i and n_i' be the order of B_i, $i=1, 2$. Assume that the pair n_1, n_2 is not the pair n_1', n_2' in either order. Then, because $n_1+n_2=n_1'+n_2'$, one of the four numbers is a strict maximum. Suppose $n_2 > \max(n_1, n_1', n_2')$. Then A_2 is not congruent to B_1 or B_2 or any of their subgraphs, and therefore A_1 has no n. e. points. It is therefore a path with one r-point of A, say a_3. Then vertex b_3 is an r-point and is the only r-point of its limb because a_3 is not an n. e. point. The tree $c(a_3)$ is central, has radius r, and $\bar{a}_2$ is its center, so its two radial limbs have orders n_1 and n_2. The center of $c(b_3)$ is either $\bar{b}_1$ or $\bar{b}_2$, but in either case the two limbs have orders n_1' and n_2', so a congruence of $c(a_3)$ and $c(b_3)$ is impossible. From this contradiction it follows that n_1 and n_2 are in some order the numbers n_1' and n_2' and it may be supposed that $n_1 = n_1'$

and $n_2 = n'_2$.

Now consider the n. e. points. If none exist, then both trees are $(2r+1)$-paths and hence are congruent. If, on the other hand, a_i is an n. e. point of A, then b_i is an n. e. point of B and the following applies:

(3) If a_i and b_i are n. e. points, with a_i in A_1 and b_i in B_2 (or a_i in A_2 and b_i in B_1), then $A \cong B$.

For, suppose $a_i \in A_1$ and $b_i \in B_2$. Then because of the orders of the limbs, an a_i-mapping pairs A_2 with B_1, so $A_2 \cong B_1$ and $n_1 = n_2$. If there is no n. e. point of A in A_2 then A_2 is an r-path and so is B_1 because it is congruent to A_2. But then, because $n_1 = n_2$, both A_1 and B_2 are also r-paths, which is contradictory. Therefore, there exists an n. e. point a_j in A_2. Because $n_1 = n_2$, an a_j-mapping pairs A_1 either with B_2 or with B_1. The first case, together with $A_2 \cong B_1$, implies $A \cong B$ directly. The second case implies $A_1 \cong B_1 \cong A_2$, and from this it follows that there is an n. e. point, say b_k, in B_1. Then a b_k-mapping pairs B_2 with either A_1 or A_2. Therefore all the limbs are the limbs are the same type and $A \cong B$.

Because of (3), it is now only necessary to consider the case $a_i \in A_1$ and $b_i \in B_1$. There are two sub-cases.

Case 3.1. There is no n. e. point in either A_2 or B_2. Then $A_2 \cong B_2$ since they are both r-paths. Let the end point of A_2 be a_3. Then $c(a_3)$ is a central tree, of radius r, whose center is $\bar{a}_1$. From $c(a_3) \cong c(b_3)$, it follows that b_3 is the only r-point of some limb in B. Assume $b_3 \in B_1$. Let b_4 be the r-point of B in B_2. Then a_4 is the only r-point of A in A_1. An a_4-mapping pairs $\bar{a}_2$ with $\overline{b_1}$ and also pairs the limb of $c(a_4)$ which is not a path with the limit of $c(b_4)$ which is not a path. It therefore pairs $\bar{a}_1$ with b_{11}, the first point in the limb B_1. Because $\bar{a}_1$ is of degree two, the point b_{11} is of degree two and so is joined to a well defined second point in B_1, say b_{12}. An a_3-mapping pairs $\overline{b_2}$ and $\bar{a}_1$, and, by the same reasoning as before, pairs $\overline{b_1}$ with the first point, say a_{11}, in A_1. Then a_{11} is of degree two and so is joined to a well defined second point a_{12} in A_1. An a_4-mapping must, then, pairs a_{11} with b_{12}, so b_{21} is of degree two and joins the third point in B_1. Alternating this way between the a_3 and a_4 mappings, it follows that all points of A_1 and B_1 are of degree two, which is absurd. The assumption that b_3 is in B_1 is therefore false, so $b_3 \in B_2$. Now an a_3-mapping must pair $\bar{a}_1$ with $\overline{b_1}$ and must also pair the branching and non-branching limbs at $\bar{a}_1$ and $\overline{b_1}$. Therefore $A_1 \cong B_1$, and this, with $A_2 \cong B_2$, implies $A \cong B$.

Case 3.2. There is an n. e. point in A_2 or else there is one in B_2. Suppose $a_j \in A_2$ is nonessential. If b_j is in B_1, then, from (3), $A \cong B$, so suppose $b_j \in B_2$. If an a_i-mapping pairs A_2 with B_2 and an a_j-mapping pairs A_1 with B_1 then clearly $A \cong B$. So suppose an a_i-mapping pairs A_2 with the remainder of B_1 (without b_i). Then $n_1 = n_2 + 1$ and because of this an a_j-mapping pairs A_1 with B_1, hence $A_1 \cong B_1$. Let a_k be the point of A_1 paired with b_i in an a_j-mapping. Then A_1 minus a_k, that is the graph obtained from A_1 by deleting a_k and all joins to a_k, is congruent to B_1 minus b_i. But an a_i-mapping pairs A_2 with B_1 minus b_i. Therefore $c(a_k)$ is a bicentral tree both of whose limbs are congruent to A_2. From Lemma 1 there is a subgraph of the same type in B and hence $B_2 \cong A_2$. This, with $A_1 \cong B_1$, implies $A \cong B$ and completes the proof.

It is natural to wonder if any two graphs must be isomorphic when they have the same composition in terms of $(n-1)$-point subgraphs. The author has considered the question for graphs having at most one join for any pair of points, with no point joined to itself. Actual inspection shows that the theorem is valid for all such graphs up to order seven. It also holds for any two such graphs of general, finite order if either is disconnected or its transpose is disconnected. (The transpose is obtained by reversing the join relationship between every pair of vertices.) However, the author was unable to prove or disprove the general case. As a final comment, it is not true that the same composition in terms of $(n-2)$-point subgraphs implies isomorphism.

UNIVERSITY OF CALIFORNIA, SANTA BARBARA COLLEGE.
INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J.

ON THE MEASURE OF NORMAL FORMULAS

ROBERT MCNAUGHTON

1. Introduction. Quine has recently found (in [1], [2] and [3]) a reasonably practical method which yields the simplest normal equivalent of a given truth functional formula. The problem of this paper is to find a practical method which yields the simplest normal formula with a given measure. Roughly, the measure of a formula is the number of T's in the column under the formula in a truth table which has 2^d rows; these rows represent all possible assignments of T's and F's to d letters including all the letters of the formula and perhaps others. The problem, which is rather difficult, arises in the design of certain networks in digital computers (described at the end of § 2) as part of a more general problem which is all the more difficult. Networks, however, are not discussed at all in the remainder of the paper, where the main problem is attacked as a problem in pure logic. I have had no success in obtaining a method which is generally satisfactory, but have succeeded in proving a few theorems which will probably be indispensable in any future attack on the problem.

2. The problem and its origin. Most of the terminology which I shall use is Quine's. Where it conflicts with Quine's terminology of [1], [2] and [3] I shall explicitly say so; on the other hand, I shall not presuppose that the reader is familiar with any of these papers. An italicized word appearing in a sentence of this paper is defined in that sentence. In this section a sentence without an italicized word is often a theorem which is either well known or obvious.

A formula is made up in the usual manner from the letters $A_1, \cdots, A_d$ by means of negation, conjunction and disjunction (or alternation). For any formulas $\Phi_1, \cdots, \Phi_n$, $n \geq 2$, $\overline{\Phi}_1$ is the *negation* of Φ_1, $\Phi_1\Phi_2\cdots\Phi_n$ is the *conjunction* of $\Phi_1, \cdots, \Phi_n$ (these being *conjuncts*), and $\Phi_1 \vee \Phi_2 \vee \cdots \vee \Phi_n$ is the *disjunction* (called 'alternation' by Quine) of $\Phi_1, \cdots, \Phi_n$ (these being *disjuncts*). (I assume that the reader is familiar enough with the general literature to see how the circularity of definition in the last two sentences can be avoided.) A letter or its negation is a *literal*. If a formula is a disjunction, then the disjuncts are *clauses*; if it is not a disjunction, the formula itself is its only *clause*. A formula all of whose clauses are literals or conjunctions of literals is a *normal formula*. (For Quine a clause of a normal formula cannot have

Received January 30, 1956. This paper is an abridged version of report No. 1, contract DA-04-200-ORD-436 (Applied Mathematics and Statistics Laboratory, Stanford, California), sponsored by the Office of Ordnance Research.

repetitious letters.)

A formula is *consistent* if it comes out true under some interpretation of its letters. An inconsistent normal formula, then, must be one in which every clause is inconsistent; a clause of a normal formula is inconsistent if and only if there is a letter appearing both with and without a bar. An inconsistent clause can be omitted from any consistent normal formula and the resulting formula is equivalent to the original. A clause of a normal formula *subsumes* another if every literal of the second is a literal of the first. Any clause which subsumes another in a normal formula can be omitted and the resulting formula is equivalent to the original. A literal which has occurred previously in the same clause can be omitted and the resulting clause is equivalent to the original clause; hence, the resulting formula is equivalent to the original formula. A normal formula in which no clause subsumes another, no clause is inconsistent, and no clause contains a repeated literal, is an *apparently irredundant normal formula*. An *irredundant normal formula* is one in which no literal or clause can be omitted without sacrificing equivalence. Some apparently irredundant formulas are not irredundant as example 1 or example 2 of [1] is enough to show. An *interclausally consistent* formula is one in which the conjunction of any two clauses is consistent. A normal formula is interclausally consistent if and only if no letter appears in it at least once with a bar and at least once without a bar.

A normal formula is *developed* with respect to the letters $A_1, \cdots, A_d$ if every clause has one and only one occurrence of each of these letter. Every consistent formula Φ containing no letters other than $A_1, \cdots, A_d$ can be transformed into an irredundant normal formula which is developed with respect to the letters $A_1, \cdots, A_d$; the number of clauses in the letter is the *measure* of Φ, or $m(\Phi)$. If Φ is inconsistent, then the *measure* of Φ is 0. In general $m(\Phi)$ depends on d, but there is no need to make this dependence explicit in the notation in most of this paper. Where the notation $m(\Phi)$ is used, it is assumed that Φ contains only (perhaps not all) the letters $A_1, \cdots, A_d$. If a truth table is constructed for Φ with 2^d rows, representing the 2^d assignments of truth values to $A_1, \cdots, A_d$, then there will be $m(\Phi)$ T's in the column for Φ.

Two formulas Φ and Ψ are *isomorphic* if there is a one-to-one mapping f of the set of literals of Φ onto the set of literals of ψ such that, if both A_i and $\overline{A}_i$ occur in Φ and if $f(A_i)=A_j$ then $f(\overline{A}_i)=\overline{A}_j$, and if $f(A_i)=\overline{A}_j$ then $f(\overline{A}_i)=A_j$; and such that a formula Ψ' can be obtained from Φ by replacing each literal by its image under f, and Ψ can be obtained from Ψ' by changing the order of conjunts of zero or more conjunctions and changing the order of disjuncts of zero or more disjunctions. Thus $A_1A_2\vee\overline{A}_1A_3A_4\vee\overline{A}_3A_4A_5$ is isomorphic to $A_1A_5\overline{A}_6\vee A_4\overline{A}_5\overline{A}_6$

$\vee A_2\bar{A}_1$; here f is the mapping such that $f(A_1)=\bar{A}_1$, $f(A_2)=A_2$, $f(A_3)=A_5$, $f(A_4)=\bar{A}_6$, $f(A_5)=A_4$. If no letters except $A_1, \cdots, A_d$ occur in Φ and Ψ, and if Φ and Ψ are isomorphic, then $m(\Phi)=m(\Psi)$. (If this fact is not obvious enough to the reader, it is proved for the case in which Φ and Ψ are normal formulas as Theorem 2.4.)

For the purposes of this paper the word "simplicity" need not be, and is not, defined precisely. Let us understand merely that simplicity of a normal formula depends on the number of clauses and the number of literals in each clause.

A practical solution to the problem of finding the simplest normal formula with a given measure would have some application to the design of certain parts of digital computers. Dr. Montgomery Phister, Jr. of the Ramo Wooldridge Corporation has suggested the following problem which was the initial stimulus for the research for this paper.

Suppose that one is to devise a circuit with n outputs in such a way that in each of m given time intervals each output is to be in state 1 or state 0 as specified. The circuit engineer can select his inputs in any way he chooses, so that each input is either 0 or 1 in each interval. But he must do so in such a way that each output is a function of the inputs and the circuit is the most economical. If certain kinds of diode circuit are used, then the part of the circuit which relates any output to the inputs must be constructed as a normal formula.

The problem of finding the simplest normal formula with a given measure is relevant to this problem, even though a practical solution to the former would not necessarily mean a practical solution to the latter. If the number of intervals is between $2^{d-1}+1$ and 2^d inclusive and if each time interval itself is to be a unique function of the inputs, then there must be d inputs. With these assumptions, a practical way of choosing inputs so as to minimize the circuit for just one output is easily obtainable if there is a practical way of finding the simplest normal formula with a given measure. For example, if there are 16 time intervals and there is to be one output in state 1 in exactly 5 intervals, then it is necessary to find, for $d=4$, a simplest formula whose measure is 5; in this case, $A_1A_2\vee A_1A_3A_4$ seems to be a formula.

3. Calculation of the measure of a formula. There is a straightforward way of calculating the measure of a normal formula which is somewhat simpler than actually expanding it into a developed normal formula. At the basis of this method is an easily proved theorem relating the measures of two formulas, their conjunction and their disjunction.

THEOREM 3.1. *For any formulas Φ and Ψ*

$$m(\Phi \vee \psi) = m(\Phi) + m(\Psi) - m(\Phi\Psi)$$

Proof. Consider the developed normal formulas with respect to d variables equivalent to the four formulas concerned, Φ^*, Ψ^*, $(\Phi\vee\Psi)^*$, $(\Phi\Psi)^*$. The number of clauses in $(\Phi\vee\Psi)^*$ can be counted by counting the number of clauses in Φ^* (which is assumed to be 0 if Φ is inconsistent) and then counting the number of clauses in Ψ^*, remembering that any clauses which these have in common have been counted twice. But the number of clauses which Φ^* and Ψ^* have in common is precisely the number of clauses in $(\Phi\Psi)^*$, or 0 if $\Phi\Psi$ is inconsistent, which in either case is $m(\Phi\Psi)$. Hence Theorem 3.1 follows.

THEOREM 3.2. *If Φ is a conjunction of j distinct literals, no two of which are of the same letter, then $m(\Phi)=2^{d-j}$.*

This theorem follows readily from well known properties of truth tables or developed normal formulas.

For the remainder of this paper let Φ be a normal formula. Let $\varphi_1, \cdots, \varphi_k$ be the clauses of Φ in the order of their appearance in Φ. Let Φ_x, $1 \leq x \leq k$, be the normal formula $\varphi_1 \vee \varphi_2 \vee \cdots \vee \varphi_x$. Thus Φ_k is Φ. Let $C(\Phi)$ be the set of all clauses of Φ. For any $S \subseteq C(\Phi)$ let j_S be the total number of distinct letters appearing in the clauses of S. Let i_S be 0 if at least one letter appears in at least one clause with a bar and in at least one clause without a bar, that is, if the conjunction of all the clauses of S is inconsistent. If there is no such letter, let i_S be 1 if there are an odd number of clauses in S or -1 if there are an even number of clauses in S.

THEOREM 3.3.

$$m(\Phi) = \sum_{S \subseteq C(\Phi)} i_S 2^{d-j_S},$$

the summation being taken only over nonempty subsets S.

The proof is by induction on the number k of clauses in Φ. If $k=1$, then there is only one S to be considered, namely, the unit set of the one clause. If there is a letter which appears both with and without a bar, i_S is 0 and so is $m(\Phi)$. If there is no such letter, then $m(\Phi)=2^{d-j_S}$, by Theorem 3.2.

Suppose that $k>1$ and suppose that Theorem 3.3 holds for all formulas having fewer than k clauses. By Theorem 3.1,

$$(1) \qquad m(\Phi) = m(\Phi_{k-1}) + m(\varphi_k) - (\Phi_{k-1}\varphi_k)\,.$$

Now $\Phi_{k-1}\varphi_k$ is not a normal formula unless $k-1=1$. If $k-1>1$ let Ψ be the normal formula which is equivalent to $\Phi_{k-1}\varphi_k$ obtained by distributing the conjunct φ_k over the clauses of Φ_{k-1}, and if $k-1=1$, let Ψ be $\Phi_{k-1}\varphi_k$. Thus Ψ will have $k-1$ clauses and, for each $h \leqq k-1$, the h^{th} clause of Ψ will have the literals of the h^{th} clause of Φ_{k-1} and those of φ_k, but no others. Now since Ψ is equivalent to $\Phi_{k-1}\varphi_k$,

$$(2) \qquad m(\Psi)=m(\Phi_{k-1}\varphi_k)\,.$$

By the inductive hypothesis and Theorem 3.2,

$$(3) \qquad m(\Phi_{k-1})=\sum_{S\subseteqq C(\Phi_{k-1})} i_S 2^{d-j_S}$$

$$(4) \qquad m(\varphi_k)=i_{C(\varphi_k)}2^{d-j_{C(\varphi_k)}},$$

and

$$(5) \qquad m(\Psi)=\sum_{S\subseteqq C(\Psi)} i_S 2^{d-j_S},$$

where in each case only nonempty subsets S are considered.

Making substitutions in (1), justified by (2), (3), (4) and (5), we get

$$(6) \qquad m(\Phi)=\sum_{S\subseteqq C(\Phi_{k-1})} i_S 2^{d-j_S}+i_{C(\varphi_k)}2^{d-j_{C(\varphi_k)}}-\sum_{S\subseteqq C(\Psi)} i_S 2^{d-j_S}\,.$$

It remains only to show that we can equate the expression $\sum_{S\subseteqq C(\Phi)} i_S 2^{d-j_S}$ with the right side of (6). But these expressions are equal, term by term, as can be seen as follows. For every S in $C(\Phi)$, either S does not contain φ_k (case I), S contains φ_k and other clauses (case II), or S contains φ_k only (case III). In case III, the summand $i_S 2^{d-j_S}$ is the middle term of the right side of (6). In case I, $S \in C(\Phi_{k-1})$ and so the summand $i_S 2^{d-j_S}$ occurs as a summand in the first term on the right side of (6). In case II, finally, suppose S contains besides φ_k the $g_1^{\text{th}}, \cdots, g_n^{\text{th}}$ clauses of Φ. Then consider S', the set containing the $g_1^{\text{th}}, \cdots, g_n^{\text{th}}$ clause of Ψ. The literals appearing in these clauses are exactly the literals appearing in the clauses of S. A letter will appear both with a bar and without a bar in S if and only if it does in S'. Hence, $i_{S'}=0$ if and only if $i_S=0$. There an odd (even) number of clauses in S if and only if there are an even (odd) number in S' since S has just one more clause than S'. Hence the summand $i_S 2^{d-j_S}$ occurs negatively as a summand in the third term of the right side of (6). It is easy to see that this correspondence is one to one and that the equality of the two expressions is established.

THEOREM 3.4. *If Φ and Ψ are normal formulas, and if Φ is iso-*

morphic to Ψ, *then* $m(\Phi)=m(\Psi)$.

Proof. The *corresponding clause* in Ψ to a clause φ in Φ is the clause which contains those literals into which the literals of φ are mapped. Consider any set S of clauses of Φ and the corresponding set of clauses S' in Ψ. The total number of distinct letters of S equals the total number of distinct letters of S'. There is a letter appearing with a bar and without a bar in S if and only if there is such a letter appearing in S'. And, of course, the number of clauses of S is the same as the number of clauses of S'. Thus the expressions for $m(\Phi)$ and $m(\Psi)$ as given by Theorem 3.3 will be the same, term by term.

The formula Ψ *implies* Γ if every assignment of truth values which makes Ψ true also makes Γ true. As is well known, Ψ implies Γ if and only if every clause of a developed normal formula equivalent to Ψ is a clause of a developed normal formula with respect to the same variables equivalent to Γ. The formula Ψ is *equivalent* to Γ is Ψ implies Γ and Γ implies Ψ. The formulas Ψ and Γ are equivalent if and only if they are equivalent to a common developed normal formula. Theorems 3.5, 3.6 and 3.7 are direct consequences of these remarks.

THEOREM 3.5. *For any formulas* Ψ *and* Γ, $m(\Psi\vee\Gamma)\geqq m(\Psi)$. *The equality holds when, and only when,* Γ *implies* Ψ.

THEOREM 3.6. $m(\Psi\Gamma)\leqq m(\Psi)$. *The equality holds when, and only when,* Ψ *implies* Γ.

THEOREM 3.7. *If* Γ *implies* Ψ, *then* $m(\Gamma)\leqq m(\Psi)$. *The equality holds when, and only when,* Γ *and* Ψ *are equivalent.*

4. Bounds on the measure of a formula with a given structure. If $j_1\leqq j_2\leqq\cdots\leqq j_k$, then a formula has the *structure* $\langle j_1, j_2, \cdots, j_k\rangle$ if and only if it is an apparently irredundant normal formula with k clauses which have, in the order in which they appear in the formula, $j_1, \cdots, j_k$ literals. Note that a formula has some structure if and only if it is normal, it is apparently irredundant, and its clauses are in order of nondecreasing length. In finding a simplest normal formula we need only consider formulas which have some structure; for every normal formula which does not is equivalent to, and is no simpler than, a normal formula which does. In this section I shall give an upper bound and a lower bound on the measure of formula with a given structure. This result will be convenient in some cases where one is trying to determine a simplest formula with a given measure.

THEOREM 4.1. *If* Φ *has the structure* $\langle j_1, \cdots, j_k\rangle$, *then*

$$m(\Phi) \leqq 2^{d-j_1} + \cdots + 2^{d-j_k}.$$

The proof is by induction on k. For $k=1$, if Φ is of structure $<j_1>$, then since Φ is apparently irredundant, $m(\Phi)=2^{d-j_1}$ by Theorem 2.2. I shall assume that Theorem 4.1 is true about all structures whose formulas have less than k clauses, and show that it is true about the structure $<j_1, \cdots, j_k>$. If Φ has the structure $<j_1, \cdots, j_k>$, then, by inductive hypothesis, $m(\Phi_{k-1}) \leqq 2^{d-j_1} + \cdots + 2^{d-j_{k-1}}$, and $m(\varphi_k)=2^{d-j_k}$. But, by Theorem 3.1,

$$m(\Phi) \leqq m(\Phi_{k-1}) + m(\varphi_k) \leqq 2^{d-j_1} + \cdots + 2^{d-j_{k-1}} + 2^{d-j_k}.$$

(Note that, for any formula Φ with structure $<j_1, \cdots, j_k>$, $m(\Phi)=2^{d-j_1} + \cdots + 2^{d-j_k}$ if and only if the conjunction of every pair of clauses is inconsistent. It can be proved that such a formula exists if and only if $2^{d-j_1} + \cdots + 2^{d-j_k} \leqq 2^d$.)

Theorem 4.2, 4.3 and 4.4 are, in effect, lemmas to Theorem 4.5 which establishes a lower bound on formulas with a given structure.

THEOREM 4.2. *If Φ is not interclausally consistent, if the number of distinct literals of Φ does not exceed d, and if Φ has some structure, then there is an interclausally consistent normal formula with the same structure as Φ but with no greater measure.*

Proof. If Φ is not interclausally consistent, then there is at least one letter in Φ appearing both with and without a bar I shall prove Theorem 4.2 by proving that there is a formula Φ' with the same structure as Φ with exactly one less letter appearing both with and without a bar, such that $m(\Phi') \leqq m(\Phi)$. Suppose A_n appears both with and without a bar in Φ. Let Φ' be Φ with every occurrence of $\bar{A}_n$ replaced by a variable A_p with does not appear in Φ. Since Φ has some structure, it is apparently irredundant, and so A_n never appears both with and without a bar in any one clause of Φ. Hence Φ is equivalent to $A_n\psi \vee \bar{A}_n\Gamma \vee \Omega$ and Φ' is equivalent to $A_n\psi \vee A_p\Gamma \vee \Omega$, where ψ, Γ, and Ω are normal formulas in which A_n, $\bar{A}_n$, and A_p do not occur. By Theorem 2.1, then, the following hold.

$$m(\Phi)=m(\bar{A}_n\Gamma)+m(A_n\psi \vee \Omega)-m(\bar{A}_n\Gamma(A_n\psi \vee \Omega))$$

$$m(\Phi')=m(A_p\Gamma)+m(A_n\psi \vee \Omega)-m(A_p\Gamma(A_n\psi \vee \Omega)).$$

We have, $m(A_p\Gamma)=m(\bar{A}_n\Gamma)$ since $A_p\Gamma$ and $\bar{A}_n\Gamma$ are isomorphic and can easily be converted into isomorphic normal formulas. Therefore we can concentrate on the last term of each equation. $\bar{A}_n\Gamma(A_n\psi \vee \Omega)$ is equivalent to $\bar{A}_n\Gamma\Omega$, and $A_p\Gamma(A_n\psi \vee \Omega)$ is equivalent to $A_nA_p\Gamma\psi \vee A_p\Gamma\Omega$.

Also $m(\overline{A}_n\Gamma\Omega)=m(A_p\Gamma\Omega)$, the formulas being isomorphic. Finally $m(A_p\Gamma\Omega)\leqq m(A_nA_p\Gamma\psi\vee A_p\Gamma\Omega)$, by Theorem 3.5. Thus $m(\Phi')\leqq m(\Phi)$, Φ' has the same structure as Φ, and the number of literals of Φ' equals the number of literals of Φ.

Now if Φ' still has at least one letter with and without a bar, I construct Φ'', related to Φ' as Φ' is to Φ, and so forth. Eventually I shall obtain a formula $\Phi^{(q)}$ which has no letters appearing both with and without a bar, has the same structure as Φ and has demonstrably no greater measure. It is obvious enough that the number of variables in $\Phi^{(q)}$ will not exceed d if the hypothesis of the theorem is satisfied.

(Two things can be noted. First the formula $\Phi^{(q)}$ can easily be constructed from Φ as follows: supposing (without loss of generality) that $A_1, \cdots, A_q$ are the variables which appear both with and without bars and $A_{q+1}, \cdots, A_r$ are the other variables of Φ, replace $\overline{A}_1, \cdots, \overline{A}_t$ by $A_{r+1}, \cdots, A_{r+q}$ respectively. Second, if we prefer, we can delete all the bars from $\Phi^{(q)}$ and the resulting formula will have the same measure, being isomorphic to $\Phi^{(q)}$. In summary, then, given a formula Φ which satisfies the hypothesis of Theorem 4.2 it is an easy matter to write down another formula without bars, with the same structure and with the same structure and with no greater measure.)

THEOREM 4.3. *If Ψ, Γ have no letters in common, then*

$$m(\Psi\Gamma)=\frac{m(\Psi)\cdot m(\Gamma)}{2^d},$$

Proof. Suppose (without loss of generality) that $A_1, \cdots, A_n$ are the letters occurring in Ψ; then every letter appearing in Γ is one of the letters $A_{n+1}, \cdots, A_d$. From well known logical laws, the developed normal equivalent Ψ' of Ψ with respect only to the letters $A_1, \cdots, A_n$ has $m(\Psi)/2^{d-n}$ clauses. And the developed normal equivalent Γ' of Γ with respect to the letters $A_{n+1}, \cdots, A_d$ has $m(\Gamma)/2^n$ clauses. $\Psi\Gamma$ is equivalent to $\Psi'\Gamma'$; the developed normal equivalent of these can be obtained from the latter by the distributive law for disjunction over conjunction; the number of clauses will be the product of the number of clauses of Ψ' and Γ', which is $m(\Psi)m(\Gamma)/2^d$. Since this last formula is the developed normal equivalent of $\Psi\Gamma$ with respect to $A_1, \cdots, A_d$, this number is the measure of $\Psi\Gamma$. (For example, if $d=5$, Ψ is $A_1\vee A_2$ and Γ is A_4A_5, then $n=2$, Ψ' is $A_1A_2\vee A_1\overline{A}_2\vee\overline{A}_1A_2$ and Γ' is $A_3A_4A_5\vee\overline{A}_3A_4A_5$. The result of "multiplying out" Ψ' and Γ' yields the developed normal equivalent with respect to $A_1, \cdots, A_5$ of $\Psi\Gamma$.)

THEOREM 4.4. *If $k<d$, the formula $A_1\vee A_2\vee\cdots\vee A_k$ has the*

maximum measure $2^{d-1}+2^{d-2}+\cdots+2^{d-k}$ *of all interclausally consistent normal formulas having exactly* $k<d$ *clauses; the formula* $A_1\vee A_2\vee\cdots\vee A_d$ *has the maximum measure* $2^{d-1}+2^{d-2}+\cdots+2^0$ *of all interclausally consistent normal formulas having* d *or more clauses.*

Proof. The formula mentioned in the second part of the theorem is equivalent to $-(\bar{A}_1\bar{A}_2\cdots\bar{A}_d)$. Therefore its developed normal form has all the 2_d clauses except one, and therefore its measure is 2^d-1. This measure is a maximum for all interclausally consistent formulas since the one higer measure, 2^d, is that of a tautology, which is never interclausally consistent.

The formula of the first part of the theorem, when developed with respect to $A_1, \cdots, A_k$, has 2^k-1 clauses (by the first two sentences of the above paragraph with 'k' for 'd'). From this we obtain an equivalent formula developed with respect to all d letters by developing each clause into 2^{d-k} clauses. The measure of the formula, therefore, is

$$2^{d-k}(2^k-1)=2^d-2^{d-k}=2^{d-1}+2^{d-2}+\cdots+2^{d-k}\,.$$

Now every interclausally consistent formula with k clauses not isomorphic to $A_1\vee\cdots\vee A_k$ either has at least one clause with more than one literal or has one literal in every clause with some repetitions of clauses. In the latter case, the formula is equivalent to $A_1\vee\cdots\vee A_j$, for some $j<k$, whose measure is $2^d+2^{d-1}+\cdots+2^{d-j}<2^d+\cdots+2^{d-k}$. I dispose of the former case by showing that, in an interclausally consistent formula, if every clause containing more than one literal is replaced by just one of its literals then the measure of the formula is not decreased. Suppose the formula $\Psi\vee A_{i_1}\varphi_{i_1}\vee\cdots\vee A_{i_n}\varphi_{i_n}$ is thus replaced by $\Psi\vee A_{i_1}\vee\cdots\vee A_{i_n}$. The latter is implied by the former and hence, by Theorem 3.7, its measure is no smaller than that of the former.

(It is easy to extend this method of proof to prove that a formula with k clauses in which no letter appears both with and without a bar has this maximum measure if and only if each clause has one literal and no literals are repeated or, equivalently, no clauses are redundant.)

THEOREM 4.5. *If* Φ *has the structure* $\langle j_1, \cdots, j_k\rangle$, *then*

$$m(\Phi)\geqq 2^{d-j_1}+2^{d-j_2-1}+\cdots+2^{d-j_k-(k-1)}\,.$$

I prove first that Theorem 4.5 holds where Φ is interclausally consistent. The proof is by induction. If Φ is of structure $\langle j_1\rangle$, then equality holds. Assume, as an inductive hypothesis, that the measure of any formula Γ of structure $\langle j_1, \cdots, j_{h-1}\rangle$ satisfies

$$m(\Gamma) \geqq 2^{d-j_1} + 2^{d-j_2-1} + \cdots + 2^{d-j_{h-1}-(h-2)} .$$

Where Φ_h is of structure $<j_1, \cdots, j_h>$,

$$m(\Phi_h) = m(\Phi_{h-1}) + m(\varphi_h) - m(\varphi_h \Phi_{h-1}) .$$

The formula Φ_{h-1} has structure $<j_1, \cdots, j_{h-1}>$ and so, by inductive hypothesis,

$$m(\Phi_{h-1}) \geqq 2^{d-j_1} + 2^{d-j_2-1} + \cdots + 2^{d-j_{h-1}-(h-2)} .$$

Also $m(\varphi_h) = 2^{d-j_h}$. Therefore, it remains to prove that

$$m(\varphi_h \Phi_{h-1}) \leqq 2^{d-j_h-1} + 2^{d-j_h-2} + \cdots + 2^{d-j_h-(h-1)} .$$

The formula $\varphi_h \Phi_{h-1}$ is equivalent to $\varphi_h \Psi$, where Ψ is obtained from Φ_{h-1} by deleting literals which appear in φ_h; we know that there must be at least one literal in each clause of φ_{h-1} which does not occur in φ_h, for otherwise φ_h would subsume another clause of Φ contrary to the assumption that Φ has structure $<j_1, \cdots, j_h>$ and is therefore apparently irredundant. Therefore Ψ has $h-1$ clauses and has no literals of φ_h. Since Φ is interclausally consistent, it has no letters both with and without bars; it follows that, since Ψ has no literals of φ_h, it has no letters of φ_h. Thus,

$$m(\varphi_h \Psi) = \frac{1}{2^d} m(\varphi_h) m(\Psi) = \frac{2^{d-j_h}}{2^d} m(\Psi)$$

by Theorems 4.3 and 3.2. Since Ψ has $h-1$ clauses and since no letter appears both with and without bars in Ψ, it follows from Theorem 4.4 that, regardless of the value of h, $m(\Psi) \leqq 2^{d-1} + 2^{d-2} + \cdots + 2^{d-(h-1)}$. Therefore,

$$m(\varphi_h \Phi_{h-1}) = m(\varphi_h \Psi) \leqq \frac{2^{d-j_h}}{2^d} (2^{d-1} + 2^{d-2} + \cdots + 2^{d-(h-1)}$$

$$= 2^{d-j_h-1} + 2^{d-j_h-2} + \cdots + 2^{d-j_h-(h-1)} ,$$

which is what had to be proved.

To show that Theorem 4.5 still holds when Φ is not intercluasally consistent, I must discuss what happens to a formula when d varies. I shall use the notation $m_d(\Phi)$ here (and only here) to denote the measure of Φ for a given d. From the definition of measure, it can be seen that $m_d(\Phi) = m_{d'}(\Phi) \cdot 2^{d-d'}$, assuming that Φ has at most min (d, d') letters. Let Φ be a formula of structure $<j_1, \cdots, j_k>$ which is not interclausally consistent. I have to prove that, for any d not less than the number of distinct letters appearing in Φ,

$$m_d(\Phi) \geqq 2^{d-j_1} + 2^{d-j_2-1} + \cdots + 2^{d-j_k-(k-1)} .$$

Let e be the number of letters appearing both with and without bars in Φ; then $m_{d+e}(\Phi)=m_d(\Phi)2^e$. There are at most $d+e$ distinct literals in Φ, and so, by Theorem 4.2, there is an interclausally consistent Φ' of structure $<j_1, \cdots, j_k>$ (with at most $d+e$ letters) such that $m_{d+e}(\Phi') \leqq m_{d+e}(\Phi)$. But from what has been established it follows that

$$m_{d+e}(\Phi') \geqq 2^{d+e-j_1}+2^{d+e-j_2-1}+\cdots+2^{d+e-j_k-(k-1)},$$

which must now be true for $m_{d+e}(\Phi)$. Therefore,

$$m_d(\Phi) \geqq 2^{d-j_1}+2^{d-j_2-1}+\cdots+2^{d-j_k-(k-1)}.$$

This observation completes the proof of Theorem 4.5.

For $j_k+k \leqq d+1$, a formula with structure $<j_1, \cdots, j_k>$ which has the minimum measure $2^{d-j_1}+2^{d-j_2-1}+\cdots+2^{d-j_k-(k-1)}$ has been exhibited in the literature, namely in Quine's paper [2]. Quine does not discuss the measure of formulas, but proves, in his Theorem 2, that his formula has the value truth in just the first $2^{d-j_1}+\cdots+2^{d-j_k-(k-1)}$ rows of the conventional truth table. By the well known connection between truth tables and developed normal formulas it follows that the measure of Quine's formula is this number. The construction of this formula which has no bars can be described as follows: the first clause has j_1 distinct letters, and, in general, the h^{th} clause has all the letters of the $(h-1)^{\text{th}}$ clause except the last and enough letters which do not appear in any previous clause (at least one, since $j_{h-1} \leqq j_h$) to make a total of j_h distinct letters; the last letter of the h^{th} clause is a letter which has not appeared previously. It follows that the last letter of any clause of Quine's formula appears in that clause only. For example, if $d=10$, the Quine formula of structure $<1, 1, 3, 3, 6>$ whose measure is $2^9+2^8+2^5+2^4+2^0$ is

$$A_1 \vee A_2 \vee A_3A_4A_5 \vee A_3A_4A_6 \vee A_3A_4A_7A_8A_9A_{10}.$$

(It is possible to exploit the method used in proving Theorem 4.5 to prove that, for $j_k+k \leqq d+1$, the only formulas with structure $<j_1, \cdots, j_k>$ and measure $2^{d-j_1}+\cdots+2^{d-j_k-(k-1)}$ are those isomorphic to Quine's formula. The key property of Quine's formula is the fact that each clause φ contains a literal, say A_q, such that A_q is not in any other clause of the formula and all clauses followings φ contain all the literals of φ except A_q. In Quine's formula A_q is the last letter of the clause. This property is necessary, as well as sufficient, in order to insure that the formula Ψ in the proof of Theorem 4.5 has exactly $h-1$ clauses of one letter each, these letters being different from each other.)

One method of finding, for a given d and for a given measure $m \leqq 2^d$, a simplest normal formula whose measure is m is to construct some normal formula of measure m and then calculate the measure of all simpler normal formulas. This method is impractical (although effective) unless there is some way of limiting the number of formulas whose measure must be calculated. A method of constructing a formula due to Quine was described two paragraphs above; from what Quine shows, it follows that for any given measure such a formula can be constructed. Although Quine's formula is not always a simplest formula with that measure, it can serve to start the search for such a formula. Then the bounds on the measure of formulas with given structures established in this section serve to limit the number of formulas among which the search is to be made, (although not enough to make this method practicable). Only formulas having some structure need be examined; normal formulas without any structure are not apparently irredundant and have shorter equivalent formulas or can be converted to a formula with structure by changing the order of disjuncts. Needless to say, once a formula has been examined, formulas isomorphic to it need not be. The following theorem will be of some help in the search, although not enough to make it practicable in all examples.

THEOREM 4.6. *If a formula with some structure and with a least* $h+1$ *clauses, where* $h+1 \leqq k$, *has measure*

$$2^{d-j_1}+2^{d-j_2-1}+\cdots\cdot+2^{d-j_k-(k-1)}$$

and if the first h *clauses of it have exactly* $j_1, \cdots, j_h$, *respectively, letters, then the* $(h+1)^{\text{st}}$ *clause has at least* j_{h+1} *letters.*

Proof. A formula Φ of structure $\langle j_1, \cdots, j_h, j'_{h+1}, j'_{h+2}, \cdots \rangle$, where $j'_{h+1} < j_{h+1}$, has a measure which satisfies, by Theorem 4.5,

$$m(\Phi) \geqq 2^{d-j_1}+\cdots+2^{d-j_h-(h-1)}+2^{d-j'_{h+1}-h}+\cdots\cdot$$
$$> 2^{d-j_1}+\cdots+2^{d-j_h-(h-1)}+2^{d-j_{h+1}-h}+\cdots+2^{d-j_k-(k-1)} .$$

The last inequality is justified by the fact that the two expressions are each sums of powers of 2 with descending indices. As is well known about such expressions, since equality holds for the first h terms, the inequality of the $(h+1)^{\text{st}}$ term is decisive.

5. Conjectures and a counterexample. The results of the previous sections lead to no practical method of findining a simplest formula with a given measure. But there are two conjectures which, if they are true, would be of some significance. Another conjecture which suggests itself rather naturally turns out to be false, as a counterexample of mine will show. (I must admit that these conjectures may turn on the definition of "simplicity" which has not be given precisely in this paper.)

A one-clause formula with j literals, no two of which are of the same letter, has a measure 2^{d-j}. Any formula with at least two non-contradictory clauses having this measure cannot have a non-contradictory clause with less than $j+1$ literals (by Theorem 4.1). Thus a one-clause formula is simpler than any formula with more clauses but with the same measure. My first conjecture, in its strong form, is that any normal formula Φ is simpler than any formula with the same measure but with more clauses. The weaker form is that Φ is at least a simple any such formula. I have no conterexample to either of these propositions, nor do I have any good reason to believe that either of them is true in general.

Let r be the number of distinct letters of Φ. Then $r \leqq d$, and $m(\Phi)$ is divisible by 2^{d-r}. If. 2^{d-x} is the largest power of two which divides a given number m, then it is possible to find a formula with measure m with just x distinct letters. One example is Quine's formula with that measure (described near the end of § 4 of this paper). But for some Φ, $m(\Phi)$ is divisible by a power of two greater than 2^{d-r}. For example, for $d \geqq 3$, $m(A_1A_2 \vee A_2A_3 \vee A_1A_3) = 2^{d-1}$. In this example there is a simpler formula with the same measure, namely A_1. My second conjecture is that for any measure m, for any simplest formula Φ with measure m, m is divisible by 2^{d-r}. A weaker form of this conjecture is that, for any measure m, there exists a simplest formula Φ with measure m such that m is divisible by 2^{d-r}.

A formula with two clauses which are each consistent but which contradict each other (because a letter appears with a bar in one and without a bar in the other) has a measure $2^a + 2^b$, if there are d–a and d–b distinct literals in the respective clauses. If $a=b$ then a single clause formula with d–a–1 literals, no two of the same letter, has the same measure and is simpler. If $a > b$, then Quine's formula of measure $2^a + 2^b$ has two clauses with d–a and d–b–1, respectively, letters and is, therefore, simpler. A third conjecture that suggests itself is that, for any formula in which some letter appears both with and without a bar, there is another formula in which no letter appears both with and without a bar, which has the same measure and which is no less simple. However, for $d=6$ the formula

$$A_1A_2 \vee \bar{A}_1A_3A_4 \vee A_2A_5A_6$$

is simpler than any formula which has the same measure and which has no letter appearing both with and without a bar. (To verify this, the reader should note that Quine's formula with that measure has structure $\langle 2, 2, 3, 3 \rangle$. Using Theorems 4.6 and 4.1 the only possible structures for formulas which have the desired measure and are at least as simple as the displayed formula are $<2, 2, 3>$, $<2, 2, 4>$, $<2, 3, 3>$, and $<2, 2>$. Since the desired measure is not divisible by two, and since $d=6$, there must be exactly six distinct letters in any formula with that measure: for there are at most six, since $d=6$; and there are at least six, by an observation made in this section. Therefore, the structure $<2, 2>$ is excluded. Any formula with exactly six distinct letters in which no letter appears both with and without a bar is isomorphic to $A_1A_2 \vee A_1A_3 \vee A_4A_5A_6$ or $A_1A_2 \vee A_3A_4 \vee A_1A_5A_6$ if it has the structure $<2, 2, 3>$, is isomorphic to $A_1A_2 \vee A_3A_4 \vee A_1A_3A_5A_6$ or $A_1A_2 \vee A_1A_3 \vee A_3A_4A_5A_6$ or $A_1A_2 \vee A_1A_3 \vee A_3A_4A_5A_6$ if it has the structure $<2, 2, 4>$, and is isomorphic to $A_1A_2 \vee A_1A_3A_4 \vee A_1A_5A_6$ or $A_1A_2 \vee A_1A_3A_4 \vee A_2A_5A_6$ or $A_1A_2 \vee A_1A_3A_4 \vee A_3A_5A_6$ or $A_1A_2 \vee A_3A_4A_5 \vee A_1A_3A_6$ or $A_1A_2 \vee A_3A_4A_5 \vee A_3A_4A_6$ if is has the structure $<2, 3, 3>$. But none of these formulas has the desired measure.)

References

1. W. V. Quine, *The problem of simplifying truth functions*, Amer. Math. Monthly, **59** (1952), 521-531.

2. ———, *Two theorems about truth functions*, Bol. Soc. Mat. Mexicana, **10** (1953), 64-70.

3. ———, *A way to simplify truth functions*, Amer. Math. Monthly, **62** (1955), 627-631.

4. R. K. Richards, *Arithmetic operations in digital computers*, New York, 1955.

STANFORD UNIVERSITY

DISTRIBUTIVITY IN BOOLEAN ALGEBRAS

R. S. PIERCE

1. Introduction. Let α be an infinite cardinal number; suppose B is an α-complete Boolean algebra, that is, every subset of B which contains no more than α elements has a least upper bound in B.

DEFINITION 1.1. B is *α-distributive* if the following identity[1] holds in B whenever S and T are index sets of cardinality $\leqq \alpha$:

$$(1) \qquad \bigwedge_{\sigma \in S}(\bigvee_{\tau \in T} a_{\sigma\tau}) = \bigvee_{\varphi \in F}(\bigwedge_{\sigma \in S} a_{\sigma\varphi(\sigma)}) , \qquad \text{where} \quad F = T^S .$$

This paper studies α-distributive Boolean algebras, their Boolean spaces and the continuous functions on these Boolean spaces. A survey of the main results can be obtained by reading Theorems 6.5, 7.1, 8.1 and 8.2.

2. Notation. Throughout the paper, α denotes a fixed infinite cardinal number. The term α-B.A. is used to abbreviate α-complete Boolean algebra. Only α-complete algebras are considered, although some of the definitions apply to arbitrary Boolean algebras. We speak of α-subalgebras, α-ideals, α-homomorphisms, α-fields, etc., meaning that the relevant operations enjoy closure up to the power α. For example, an α-field is a field of sets, closed under α-unions, that is, unions of α or fewer element.

The lattice operations of join, meet and complement are designated by $\vee$, $\wedge$ and ($'$) respectively. The symbols 0 and u stand for the zero and unit elements of a Boolean algebra. Set operations are represented by rounded symbols: $\cup$, $\cap$ and $\subseteqq$ respectively denote union, intersection and inclusion. If A and B are sets, $B-A$ is the set of elements of B which are not in A; the complement (in a fixed set) of A is designated A^c. The empty set is denoted by 0. The symbol $\bar{\bar{A}}$ stands for the cardinality of the set A. Finally, for typographical reasons, we use the symbols 2^α and $\exp(\alpha)$ interchangeably.

Received February 17, 1956. The research in this paper was done, in part, while the author was a Jewett fellow of the Bell telephone laboratories.

[1] The notion of α-distributibity was introduced in [**1**]. It is assumed that the least upper bound an the right side of the equality (1) exists. However, by Corollary 3.4 below, it would suffice to make the equality in (1) contingent on the existence of this least upper bound.

3. Alternative characterizations of α-distributivity.

DEFINITION 3.1. A subset A of an α-B.A. B is called a *covering* of B if l.u.b. $A=u$. The covering A is called an *α-covering* if $\overline{\overline{A}}\leqq\alpha$. A *binary partition* of B is a pair consisting of an element of B and its complement. If A and $\tilde{A}$ are coverings of B, then $\tilde{A}$ is said to *refine* A if every $\tilde{a}\in\tilde{A}$ is $\leqq$ some $a\in A$.

PROPOSITION 3.2. *Let B be an α-B.A. Then the following are equivalent properties of B:*

(i) *B is α-distributive;*

(ii) *if $\{A_\sigma|\sigma\in S\}$ is a set of α-converings of B and $\overline{\overline{S}}\leqq\alpha$, then there exists a covering A which refines every A_σ;*

(iii) *if $\{A_\sigma|\sigma\in S\}$ is a set of binary partitions of B and $\overline{\overline{S}}\leqq\alpha$, then there exists a covering A which refines every A_σ.*

Proof.[2] (i) implies (ii). Indeed, if we index each A_σ by a set T of cardinality $\leqq\alpha$, say $A_\sigma=\{a_{\sigma\tau}|\tau\in T\}$ (allowing repetitions), then $\{\bigwedge_{\sigma\in S}a_{\sigma\varphi(\sigma)}|\varphi\in T^S\}$ is a covering which refines every A_σ.

(ii) implies (iii), obviously.

(iii) implies (i). Let $A_{\sigma\tau}=\{a_{\sigma\tau},(a_{\sigma\tau})'\}$ for all $\sigma\in S$, $\tau\in T$. Because the cardinality of $S\times T$ is $\leqq\alpha$, there exists a covering A which refines every $A_{\sigma\tau}$. Suppose $0\neq b\leqq\bigwedge_{\sigma\in S}(\bigvee_{\tau\in T}a_{\sigma\tau})$. Since l.u.b. $A=u$, there exists $a\in A$ such that $a\wedge b\neq 0$. Then for each $\sigma\in S$, we can find $\tau\in T$ such that $a\wedge a_{\sigma\tau}\neq 0$. Denoting this τ by $\varphi(\sigma)$ defines $\varphi\in F=T^S$. But A refines $A_{\sigma\varphi(\sigma)}$, so $a\leqq a_{\sigma\varphi(\sigma)}$ for all σ. Hence, $a\leqq\bigwedge_{\sigma\in S}a_{\sigma\varphi(\sigma)}$. It follows that $b\wedge(\bigwedge_{\sigma\in S}a_{\sigma\varphi(\sigma)})\neq 0$. Since b can be arbitrarily small, $\bigwedge_{\sigma\in S}(\bigvee_{\tau\in T}a_{\sigma\tau})$ is the least upper bound of the set $\{\bigwedge_{\sigma\in S}a_{\sigma\varphi(\sigma)}|\varphi\in F\}$, that is, (1) is satisfied.

COROLLARY 3.3. An α-B.A. *is α-distributive if and only if* (1) *is identically satisfied under the conditions $\overline{\overline{S}}\leqq\alpha$, $\overline{\overline{T}}=2$ and $\alpha_{\sigma 1}=(a_{\sigma 2})'$.*

Proof. By the argument that leads from (i) to (ii) in 3.2, the hypotheses of 3.3 imply (iii) of 3.2.

COROLLARY 3.4. *Let B be an α-B.A. which is not α-distributive.*

[2] The referee has pointed out that there is overlap between the first part of this paper and the independent work of Smith and Tarski [**5**]. In particular, 3.3 and 3.4 appear in [**5**] as Theorems 2.5 and 2.2, while our Corollaries 6.5 and 6.6 are special cases of Theorem 3.6 in [**5**]. It is a pleasure to acknowledge the contribution of a conscientious referee to the improvement of this paper.

Then there exists $b \neq 0$ *in* B *and a set of pairs* $\{[c_{\sigma 1}, c_{\sigma 2}] \subseteq B \mid \sigma \in \overline{\overline{S}}, \ S \leq \alpha\}$ *such that* $c_{\sigma 1} \wedge c_{\sigma 2} = 0$ *and* $c_{\sigma 1} \vee c_{\sigma 2} = b$ *for all* $\sigma \in S$ *and* $\bigwedge_{\sigma \in S} c_{\sigma \varphi(\sigma)} = 0$ *for all* $\varphi \in T^S$ $(T = [1, 2])$.

Proof. By 3.3, if B is not α-distributive, there exists a set of complementary pairs $\{[a_{\sigma 1}, a_{\sigma 2}] \mid \sigma \in S, \ \overline{\overline{S}} \leq \alpha\}$ such that the unit of B is not the least upper bound of the set of elements $\bigwedge_{\sigma \in S} a_{\sigma \varphi(\sigma)}$, $\varphi \in T^S$. Thus, there exists $b \neq 0$ in B such that $b \wedge (\bigwedge_{\sigma \in S} a_{\sigma \varphi(\sigma)}) = 0$ for all $\varphi \in T^S$. Then $c_{\sigma 1} = b \wedge a_{\sigma 1}$ and $c_{\sigma 2} = b \wedge a_{\sigma 2}$ have the required properties.

4. Examples of α-distributive Boolean algebras. Every α-field of sets is α-distributive. Moreover, from Definition 1.1.

(4.1) Every α-subalgebra of an α-distributive B.A. is α-distributive;

(4.2) Every 2^α-homomorph of an α-distributive B.A. is α-distributive.

Using (4.2), it is easy to construct α-distributive algebras which are not α-fields of sets (following Horn and Tarski [2, p. 492], or Sikorski [4, p. 253]): let B be the B.A. of all subsets of a set X with $\overline{\overline{X}} = \exp(\exp(\alpha))$. Let I be the α-ideal of all subsets M of X such that $\overline{\overline{M}} \leq \exp(\alpha)$. Then (see Tarski [8], or the remarks following 6.6 below), there exists no prime α-ideal of B which contains I, and consequently B/I has no prime α-ideals. Hence, B/I is not an α-field. On the other hand, by (4.2), B/I is α-distributive.

It is easy to see that (4.2) cannot be strengthened to assert that every α-homomorphic image of an α-field is α-distributive. In fact, by the theorem of Loomis (see [3]), every $\aleph_0$-B.A. is the $\aleph_0$-homomorph of an $\aleph_0$-field. But not every $\aleph_0$-B.A. is $\aleph_0$-distributive: an atomless measure algebra in which all nonzero elements have positive measure is not $\aleph_0$-distributive.

5. The representation of α-distributive algebras. In this section, we show that every α-distributive B.A. is the α-homomorph of an α-field. If $\alpha = 2^\beta$, then by (4.2) any α-homomorphic image of an α-field is β-distributive. This shows (as Sikorski observed in [4]) that the class of α-homomorphs of α-fields is rather elite when $\alpha \geq \exp(\aleph_0)$.

For any Boolean algebra B, let $X(B)$ denote the Boolean space of B. Then the points of $X(B)$ are the prime ideals of B and the topology is imposed by taking all the sets $X(a) = \{P \in X(B) \mid a \notin P\}$, with $a \in B$, as a neighborhood basis. As Stone [6] has shown, $X(B)$ is a totally disconnected, compact, Hausdorff space and the correspondence $a \to X(a)$ is an isomorphism of B onto the Boolean algebra of open-and-closed sets of $X(B)$.

DEFINITION 5.1. A set $T \subseteq X(B)$ is called *α-nowhere dense* if there is an α-covering A of B such that $T=(\bigcup_{a\in A}X(a))^c=\bigcap_{a\in A}X(a')$.

(5.2) A closed set $T\subseteqq X(B)$ is topologically nowhere dense in $X(B)$ (that is, T contains no open subset of $X(B)$) if and only if there is a covering A of B such that $T=(\bigcup_{a\in A}X(a))^c$. In particular, the α-nowhere dense sets are just the closed, nowhere dense sets which are α-intersections of open[3] sets.

LEMMA 5.3. *If B is an α-distributive* B.A., *and if $\{T_\sigma|\sigma\in S\}$ is a set of α-nowhere dense sets in $X(B)$ with $\overline{\overline{S}}\leqq\alpha$, then $\bigcup_{\sigma\in S}T_\sigma$ is nowhere dense in $X(B)$.*

Proof. By 5.1, $T_\sigma=(\bigcup_{a\in A_\sigma}X(a))^c$, where A_σ is an α-covering of B. By 3.2, there is a covering A which refines every A_σ. Then $T=(\bigcup_{a\in A}X(a))^c$ is a nowhere dense set (by (5.2)) which contains every T_σ.

THEOREM 5.4. *If B is an α-distributive* B.A., *then B is the α-homomorphic image of an α-field of sets.*

Proof.[4] Let F be the α-field generated by the open-and-closed subsets of $X(B)$. Let I be the α-ideal of F generated by the α-nowhere dense subset of $X(B)$. Consider the collection $\tilde{F}$ of sets in F which are congruent modulo I to some $X(a)$ with $a\in B$. The α-completeness of B implies that $\tilde{F}$ is an α-field; since $\tilde{F}$ contains every $X(a)$, $\tilde{F}=F$. By 5.3, $X(a)\in I$ only if $a=0$. Hence, F/I is isomorphic to B.

6. Quotients of α-distributive algebras.

We wish now to characterize the ideals I of an α-distributive B.A. for which B/I is α-distributive.

DEFINITION 6.1. Let S be an index set with $\overline{\overline{S}}\leqq\alpha$. For each $\sigma\in S$, suppose $A_\sigma=\{a_{\sigma\tau}|\tau\in T\}$ is a subset of the α-B.A. B. Denote

$$\prod_{\sigma\in S}A_\sigma=\{\bigwedge_{\sigma\in S}a_{\sigma\varphi(\sigma)}|\varphi\in T^S\}\cup\{0\}\ . \tag{2}$$

The sets $E\subseteqq B$ which are of the form $\prod_{\sigma\in S}A_\sigma$, with each A_σ a disjoint pair of elements of B, are called P_α subsets of B.

PROPOSITION 6.2. *Let B be an α-distributive* B.A. *and suppose I is*

[3] Note that since $X(B)$ is compact, every closed set which is an α-intersection of open sets is also an α-intersection of open-and-closed sets.

[4] This theorem is a special case of known results. (See [**1**] and the following abstracts from B.A.M.S. vol. **61** (1955): Smith 210, Chang 579, Scott 675 and Tarski 677.) We include the proof for the sake of completeness. The argument is the same as the topological proof of Loomis' theorem, given, for instance in Halmos' *Measure Theory*, p. 171.

an α-ideal of B. Then B/I is α-distributive if and only if every P_α subset of B which is contained in I has a l.u.b. *in I.*

Proof. Suppose B/I is α-distributive. Let $E=\prod_{\sigma\in S}A_\sigma$ be a P_α set with $E\subseteqq I$. Then

$$E=\{e_\varphi|\varphi\in F=T^S\}\cup\{0\}\ .$$

where $e_\varphi=\bigwedge_{\sigma\in S}a_{\sigma\varphi(\sigma)}$, $T=[1, 2]$. Let $a\to\bar{a}$ be the natural homomorphism of B onto B/I. By the α-distributivity of B/I,

$$\bigwedge_{\sigma\in S}(\bar{a}_{\sigma1}\vee\bar{a}_{\sigma2})=\bigvee_{\varphi\in F}\bar{e}_\varphi=\bar{0}$$

and hence

$$\bigvee_{\varphi\in F}e_\varphi\leqq\bigwedge_{\sigma\in S}(a_{\sigma1}\vee a_{\sigma2})\in I\ .$$

Conversely, suppose B/I is not α-distributive. Then by 3.4, there exists $\bar{b}\neq\bar{0}$ in B/I and

$$\{\bar{C}_\sigma\subseteqq B/I|\sigma\in S,\ \bar{\bar{S}}\leqq\alpha\}$$

such that

$$\bar{C}_\sigma=[\bar{c}_{\sigma1},\ \bar{c}_{\sigma2}]$$

with $\bar{c}_{\sigma1}\wedge\bar{c}_{\sigma2}=\bar{0}$, $\bar{c}_{\sigma1}\vee\bar{c}_{\sigma2}=\bar{b}$ and

$$\prod_{\sigma\in S}\bar{C}_\sigma=\{0\}\ ,$$

Choose an element $b\in B$ whose image in B/I is $\bar{b}$. Next, pick counter-images

$$[c_{\sigma1},\ c_{\sigma2}]\subseteqq B$$

of the pairs $\bar{C}_\sigma$ in such a way that $c_{\sigma1}\wedge c_{\sigma2}=0$ and $c_{\sigma1}\vee c_{\sigma2}=b$. Then $\prod_{\sigma\in S}C_\sigma$ is a P_α subset of B which is contained in I and whose least upper bound is b (since B is α-distributive), which is not in I.

PROPOSITION 6.3. *Let B an α-*B.A. *Then a subset E of B is a P_α subset if and, assuming B is α-distributive, only if*

(a) $0\in E$,

(b) *the elements of E are disjoint,*

(c) l.u.b. *E exists in B,*

(d) *there exists $E_\sigma\subseteqq E$ defined for each σ in an index set S with $\bar{\bar{S}}\leqq\alpha$, such that* l.u.b. *E_σ exists for all σ and the sets E_σ distinguish the nonzero elements of E, that is, if $e\neq\tilde{e}$ are nonzero elements of E,* *then there is an* E_σ *which contains* e *or* $\tilde{c}$, *but not both.*

Proof. The necessity of (a)–(c) is clear from 6.1. The subsets E_σ of (d) are obtained by letting $E_\sigma = [e \in E \mid e \leqq a_{\sigma 1}]$. Evidently, l.u.b. $E_\sigma = a_{\sigma 1} \wedge (\text{l.u.b. } E)$.

To show that (a)–(d) are sufficient, let $a = \text{l.u.b. } E$, $a_{\sigma 1} = \text{l.u.b. } E_\sigma$ and $a_{\sigma 2} = \text{l.u.b. } (E - E_\sigma)$. By (b), $a_{\sigma 2} = a \wedge (a_{\sigma 1})'$ and for $e \in E$, either $e \leqq a_{\sigma 1}$ and $e \wedge a_{\sigma 2} = 0$, or vice versa. We prove that $E = \prod_{\sigma \in S} A_\sigma$.

Suppose $\varphi \in F$ and $e \in E$ satisfy

$$e \wedge \bigwedge_{\sigma \in S} a_{\sigma \varphi(\sigma)} \neq 0 .$$

Then $e \wedge a_{\sigma \varphi(\sigma)} \neq 0$ for all $\sigma \in S$, so $e \leqq a_{\sigma \varphi(\sigma)}$. Consequently,

$$e = \bigwedge_{\sigma \in S} a_{\sigma \varphi(\sigma)}$$

(by (b) and (d)). Thus, $\prod_{\sigma \in S} A_\sigma \subseteqq E$. On the other hand, for $e \neq 0$ in E, define $\varphi \in F$ by $\varphi(\sigma) = 1$ if $e \in E_\sigma$, $\varphi(\sigma) = 2$ if $e \neq E - E_\sigma$. Then $0 \neq e = e \wedge \bigwedge_{\sigma \in S} a_{\sigma \varphi(\sigma)}$, and therefore $e = \bigwedge_{\sigma \in S} a_{\sigma \varphi(\sigma)}$. Hence, $E \subseteqq \prod_{\sigma \in S} A_\sigma$.

COROLLARY 6.4. *Let* B *be a* 2^α-B.A. *Then* $E \subseteqq B$ *is a* P_α *subset if and only if* $0 \in E$, *the elements of* E *are disjoint and* $\overline{\overline{E}} \leqq 2^\alpha$.

Proof. The necessity is clear. To prove the sufficiency, observe that since $\overline{\overline{E}} \leqq 2^\alpha$, it is possible to find a one-to-one map λ of E into the set of all two-valued functions on a set S of cardinality $\leqq \alpha$. For each $\sigma \in S$, let

$$E_\sigma = \{e \in E \mid [\lambda(e)](\sigma) = 1\} .$$

It is clear that the system $\{E_\sigma \mid \sigma \in S\}$ satisfies condition (d) of 6.3.

COROLLARY 6.5. *Let* B *be a* 2^α-B.A. *which is* α-*distributive. Let* I *be an* α-*ideal of* B. *Then* B/I *is* α-*distributive if and only if* I *is a* 2^α-*ideal.*

Proof.[5] If $C \subseteqq I$ satisfies $\overline{\overline{C}} \leqq 2^\alpha$, then using Zorn's lemma, it is possible to find a set E of disjoint elements such that $\overline{\overline{E}} \leqq \overline{\overline{C}}$, l.u.b. E = l.u.b. C and every $e \in E$ is contained in some $c \in C$ (so that $E \subseteqq I$). By 6.4, $E \cup [0]$ is a P_α subset of B. By 6.2, l.u.b. $(E \cup [0]) \in I$. Thus, l.u.b. $C \in I$.

[5] See footnote [2]. As noted in Smith and Tarski [**5**], the assumption that B is α-distributive in 6.6 is unnecessary. This condition was used only to prove the sufficiency in 6.2.

COROLLARY 6.6. *Let B be a 2^α-complete, α-distributivite* B.A. *Suppose α is weakly accessible from the infinite cardinal β. . Let I be a β-ideal of B such that B/I is α-distributive. Then I is a 2^α-ideal.*

Proof. First, observe that if ξ is a singular cardinal and I is an η-ideal for all $\eta < \xi$, then I is a ξ-ideal. Using this fact, 6.6 follows from 6.5 by transfinite induction on α.

It should be remarked that the methods and results of this section are related to the circle of ideas developed by Ulam and Tarski in [9] and [8]. For example, it follows directly from 6.6 that if B is a 2^α-field, where α is weakly accessible from β, then any prime β-ideal is also a 2^α-ideal (see [8], Theorem 3.19).

7. The lattice of continuous functions on $X(B)$. Stone has proved (see [7], p. 186) that a Boolean algebra B is α-complete if and only if the lattice of real valued, continuous functions on its Boolean space is conditionally α-complete. This result immediately suggests the

THEOREM 7.1. *Let B be a Boolean algebra. Then B is α-distributive if and only if the lattice $C(X(B))$ of real valued, continuous functions on the Boolean space of B is α-distributive*[6].

Proof. Assume first that $C(X(B))$ is conditionally α-complete. Then the set of all characteristic functions of open-and-closed subsets of $X(B)$ form an α-sublattice of $C(X(B))$ which is clearly lattice isomorphic to B (see the proof of Theorem 12 of [7]). Consequently, if $C(X(B))$ is α-distributive, so is B.

Conversely, suppose B is α-distributive (and α-complete). Then by Stone's result, cited above, $C(X(B))$ is conditionally α-complete and we have only to verify the relation (1) of 1.1.

First consider the special case where each function $a_{\sigma\tau}$ takes only finitely many real values. Let $A_{\sigma\tau} = \{b_{\sigma\tau n} | n = 1, 2, \cdots\}$ be a finite set of disjoint elements of B such that $\bigvee_n b_{\sigma\tau n} = u$ and $a_{\sigma\tau}$ is constant on each set $X(b_{\sigma\tau n})$. By 3.2, there is a covering A of B such that A refines every $A_{\sigma\tau}$. If $b \in A$, then every $a_{\sigma\tau}$ is constant on $X(b)$. Since $a \to (a|X(b), a|X(b'))$ is a direct decomposition of $C(X(B))$, the restriction homomorphism π_b: $a \to a|X(b)$ preserves arbitrary joins and meets. Moreover, π_b sends all $a_{\sigma\tau}$ into the conditionally complete sublattice of constant functions on $X(b)$. This sublattice, being isomorphic to the chain of real numbers, is evidently α-distributive. Hence,

[6] That is, $C(X(B))$ is a conditionally α-complete lattice which satisfies the identity (1) of Definition 1.1 when the elements $a_{\sigma\tau}$ are functions which have a common upper bound and a common lower bound.

$$\bigwedge_{\sigma\in S}\bigvee_{\tau\in T}\pi_b(a_{\sigma\tau})=\bigvee_{\varphi\in F}\bigwedge_{\sigma\in S}\pi_b(a_{\sigma\varphi(\sigma)}) .$$

Using this remark, we show that $\bigwedge_\sigma\bigvee_\tau a_{\sigma\tau}$ is the least upper bound in $C(X(B))$ of the set $\{\bigwedge_{\sigma\in S}a_{\sigma\varphi(\sigma)}|\varphi\in T^S\}$.

Suppose $f\geqq\bigwedge_\sigma a_{\sigma\varphi(\sigma)}$ for all φ. The if $b\in A$, $\pi_b(f)\geqq\bigwedge_\sigma\pi_b(a_{\sigma\varphi(\sigma)})$ for all φ, so

$$\pi_b(f)\geqq\bigvee_\varphi\bigwedge_\sigma\pi_b(a_{\sigma\varphi(\sigma)})=\bigwedge_\sigma\bigvee_\tau\pi_b(a_{\sigma\tau})=\pi_b(\bigwedge_\sigma\bigvee_\tau a_{\sigma\tau}) .$$

Thus $f(P)\geqq(\bigwedge_\sigma\bigvee_\tau a_{\sigma\tau})(P)$ pointwise on the dense set $\bigcup_{b\in A}X(b)$ and therefore, by continuity, $f\geqq\bigwedge_\sigma\bigvee_\tau a_{\sigma\tau}$. By definition of the least upper bound, $\bigwedge_\sigma\bigvee_\tau a_{\sigma\tau}=\bigvee_\varphi\bigwedge_\sigma a_{\sigma\varphi(\sigma)}$.

Now consider the general case of arbitrary functions $a_{\sigma\tau}$. Since $X(B)$ is compact and totally disconnected, the Stone-Weierstrass theorem guarantees the existence (for each $\sigma\in S$, $\tau\in T$ and integer n) of functions $f_{\sigma\tau}$, taking only finitely many real values, such that $|f_{\sigma\tau}-a_{\sigma\tau}|\leqq 1/n$. Suppose $f\in C(X(B))$ satisfies $f\geqq\bigwedge_\sigma a_{\sigma\varphi(\sigma)}$ for all $\varphi\in T^S$. Then

$$f\geqq\bigwedge_\sigma(f_{\sigma\varphi(\sigma)}-1/n)$$

for all φ. Hence, by the result of the special case,

$$\begin{aligned}f&\geqq\bigvee_\varphi\bigwedge_\sigma(f_{\sigma\varphi(\sigma)}-1/n)=(\bigvee_\varphi\bigwedge_\sigma f_{\sigma\varphi(\sigma)})-1/n=(\bigwedge_\sigma\bigvee_\tau f_{\sigma\tau})-1/n\\&\geqq(\bigwedge_\sigma\bigvee_\tau(a_{\sigma\tau}-1/n))-1/n=(\bigwedge_\tau\bigvee_\tau a_{\sigma\tau})-2/n .\end{aligned}$$

Since n can be arbitrarily large, $f\geqq\bigwedge_\sigma\bigvee_\tau a_{\sigma\tau}$. Thus, $\bigwedge_\sigma\bigvee_\tau a_{\sigma\tau}=\bigvee_\varphi\bigwedge_\sigma a_{\sigma\varphi(\sigma)}$.

8. The continuous functions on $X(B)$. In this section we consider the individual continuous functions on the Boolean space of an $\aleph_0$-distributive B.A.

LEMMA 8.1. *Let B be an $\aleph_0$-distributive* B.A. *Let $X(B)$ be the Boolean space of B. Let Y be a separable metric space. Then any continuous mapping f of $X(B)$ into Y is locally constant on a dense subset of $X(B)$, that is, the set of points P of $X(B)$ such that f is constant on some neighborhood of P is dense in $X(B)$.*

Proof. Let $\{N_1, N_2, \cdots, N_n, \cdots\}$ be a countable neighborhood basis of Y. Set $M_n=f^{-1}(N_n)$. Since Y is a metric space, N_n is an open F_σ (that is, a countable union of closed sets). By the continuity of f, so is M_n. But $X(B)$ is the Boolean space of an $\aleph_0$-B.A., so the closure of any open F_σ in $X(B)$ is open (see [5], Theorems 17 and 18). Hence, elements $b_n\in B$ exist so that $M_n^-=X(b_n)$.

Let $A_n=[b_n, b_n']$. Then there is a covering A of B which refines all A_n. By 5.2, $\bigcup_{a\in A}X(a)$ is dense in $X(B)$. It will be sufficient to prove that f is constant on $X(a)$ for each $a \in A$.

Suppose $f(P) \neq f(Q)$. Then there exists N_n such that $f(P) \in N_n$, $f(Q) \notin N_n^-$. Thus $P \in M_n \subseteqq X(b_n)$, but $Q \notin M_n^-$, since $f(M_n^-) \subseteqq N_n^-$. Hence, $Q \in X(b_n')$. Consequently, P and Q cannot lie in the same set $X(a)$ with $a \in A$. In other words, f is constant on each $X(a)$.

THEOREM 8.2. *Let B be an $\aleph_0$-B.A. and let $X(B)$ be the Boolean space of B. Then a necessary and sufficient condition that B be $\aleph_0$-distributive is that every real valued, continuous function on $X(B)$ be locally constant on a dense subset of $X(B)$.*

Proof. Necessity is a special case of 8.1. Suppose then that every real valued continuous function is locally constant on a dense set. Let $A_n = [a_n, a_n']$ be a countable set of binary partitions of B. Let $\varphi_n \in C(X(B))$ be defined by $\varphi_n(P) = 0$ if $P \in X(a_n)$, $\varphi_n(P) = 2$ if $P \in X(a_n')$. Set $f(P) = \sum_{n=1}^{\infty} \varphi_n(P)/3^n$. Then f is continuous on $X(B)$. Note that $f(P) = f(Q)$ if and only if $\varphi_n(P) = \varphi_n(Q)$ for all n (because the points of the Cantor set have unique representations in the form $\sum_{n=1}^{\infty} \delta_n/3^n$ with $\delta_n = 0, 2$). By assumption, f is locally constant on a dense set. Thus, there is a subset A of B such that $\bigcup_{a \in A} X(a)$ is dense in $X(B)$ and f is constant on each $X(a)$ with $a \in A$. This implies A is a covering of B and every φ_n is constant on each $X(a)$, so that A refines every A_n. By 3.2, B is $\aleph_0$-distributive.

9. Unsolved problems.

(9.1) What properties of the Boolean space of B characterize α-distributivity? One can deduce from 3.4 the following result, which, seemingly, is only slightly weaker than the converse of 5.3: if B is an α-B.A. which is not α-distributive, then there is a nonempty open subset of $X(B)$ which is contained in a 2^α-union of α-nowhere dense sets.

(9.2) Is the completion by Dedekind cuts of an α-field (or more generally, and α-distributive B.A.) itself α-distributive?

(9.3) Is every 2^α-complete, α-distributive B.A. the 2^α-homomorph of a 2^α-field? By 6.5, it would be enough to prove that every α-distributive, 2^α-B.A. is the α-homomorph of a 2^α-field.

REFERENCES

1. P. Erdös and A. Tarski, *On families of mutually exclusive sets*, Ann. of Math. **44** (1943), 315–329.

2. A. Horn and Tarski, *Measures in Boolean algebras*, Trans. Amer. Math. Soc. **64** (1948), 467-497.

3. L. H. Loomis, *On the representation of σ-complete Boolean algebras*, Bull. Amer. Math. Soc. **53** (1947), 757-760.

4. R. Sikorski, *On the representation of Boolean algebras as fields of sets*, Fund. Math. **35** (1948), 247-258.

5. E. C. Smith and A. Tarski, *Higher degrees of distributivity and completeness in Boolean algebras*, to appear in Trans. Amer. Math. Soc.

6. M. H. Stone, *Applications of the theory of Boolean rings to general topology*, Trans. Amer. Math. Soc. **41** (1937), 375-481.

7. ———, *Boundedness properties in function-lattices*, Canad. J. Math. **1** (1949), 176-186.

8. A. Tarki, *Über additive und multiplikative Mengenkörper und Mengenfunktionen*, Soc. Sci. Lett. Varsovie C.R. Cl. III, **30** (1937), 151-181.

9. S. Ulam, *Zur Masstheorie in der allgemeinen Mengenlehre*, Fund. Math. **16** (1930), 140-150.

UNIVERSITY OF WASHINGTON

CONTINUOUS SPECTRA AND UNITARY EQUIVALENCE

C. R. PUTNAM

1. Introduction. In the differential equation

$$(px')' + (\lambda + f(t))x = 0 \, , \tag{1}$$

let λ denote a real parameter and let $p(t)$ (>0) and $f(t)$ be continuous real-valued functions on $0 \leqq t < \infty$. Suppose that (1) is of the limit-point type, so that (1) and a linear homogeneous boundary condition

$$x(0) \cos \alpha + x'(0) \sin \alpha = 0 \, , \qquad 0 \leqq \alpha < \pi \, , \tag{2_α}$$

determine a boundary value problem with a spectrum $S = S_\alpha$ on the half-line $0 \leqq t < \infty$; cf. [7]. The continuous spectrum C_α (if it exists) is determined by a continuous monotone nondecreasing basis function $\rho_\alpha(\lambda)$. It is known that the set of cluster points, S', of S_α is independent of α, [7, p. 251]; the question as to whether the corresponding assertion for C_α is also true was raised by Weyl [7, 7. 252] but is still undecided.

Consider the self-adjoint operators $H_\alpha = \int \lambda dE_\alpha(\lambda)$ (all of which are extensions of the same symmetric operator) belonging to the various boundary value problems determined by (1) and (2_α); cf. for example, [2]. The object of this note is to shown that any two H_α possessing purely continuous (hence, in view of the above remark concerning S', necessarily identical) spectra are unitarily equivalent, at least if certain conditions concerning the nature of the sets C_α and the basis functions $\rho_\alpha(\lambda)$ are met. In fact there will be proved the following.

THEOREM (*). *Suppose that there exist two (distinct) values α_1 and α_2 $(0 \leqq \alpha_k < \pi)$ such that, for each of the two boundary value problems determined by* (1) *and* ($2_{\alpha k}$), *the following three conditions are satisfied:*

(i) $S_{\alpha k} \neq (-\infty, \infty)$,

(ii) *the point spectrum is empty, and*

(iii) $\rho_{\alpha k}(\lambda)$ *is absolutely continuous. Then $H_{\alpha 1}$ and $H_{\alpha 2}$ are unitarily equivalent.*

The condition (i) of (*) surely holds if, for instance, f is bounded or even bounded from below on $0 \leqq t < \infty$. It should be noted however that every (real) λ belongs to an S_α for some α (depending on λ); [1].

For other results on the continuous spectra of boundory value pro-

Received April, 1956. This work was supported by the National Science Foundation research grant NSF-G481.

blems with absolutely continuous basis functions (on certain intervals), see [4].

The proof of (*) in § 2 will depend upon the following result of M. Rosenblum [5] concerning perturbations of self-adjoint operators: Let the self-adjoint operators $A_k=\int \lambda dE(\lambda)$ (for $k=1, 2, 3$) satisfy $A_1-A_2=A_3$. Suppose, in addition, that A_3 is completely continuous and such that $\int|\lambda|dE_3(\lambda)$ has a finite trace while (E_1x, y) and (E_2x, y) are absolutely continuous functions of λ for arbitrary x and y in Hilbert space. Then A_1 and A_2 are unitarily equivalent.

2. Proof of (*). In the sequel, the index α_k will be replaced by k. It is clear from the assumptions that there exists some real value $\lambda=\lambda^*$ not belonging to S_k for $k=1, 2$. Consequently, since $f(t)$ can be replaced in (1) by $f(t)+\lambda^*$, it can be supposed without loss of generality that $\lambda=0$ is not in either of the sets S_k. Then the operators H_k^{-1}, where

$$H_k^{-1}=\int \lambda^{-1} dE_k(\lambda)=\int dF_k(\lambda) \qquad (F_k(\lambda)=E_k(\lambda^{-1}))$$

are bounded, self-adjoint integral operators with kernels $G_k(s, t)$ on $0\leqq s, t<\infty$; cf. for example, [2], [7]. Furthermore,

$$G_1(s, t)-G_2(s, t)=cg(s)g(t) ,$$

where c denotes a constant and $g(t)$ is a function of class $L^2[0, \infty)$; cf. [7, p. 251]. Thus $(H_1^{-1}-H_2^{-1})x$ is a multiple of g for every element x of class $L^2[0, \infty)$. Hence $H_1^{-1}-H_2^{-2}$ is a multiple of a one-dimensional projection operator; in particular, $H_1^{-}-H_2^{-1}$, corresponding to A_3, satisfies the trace condition on that operator mentioned at the end of § 1.

In view of the assumptions (ii) and (iii) of (*), it follows from the formulas relating the basis functions $\rho_k(\lambda)$ to the projections $E_k(\lambda)$ (cf., for example, [2]) that $||E(\lambda)x||$ is an absolutely continuous function of λ for every x in the Hilbert space; therefore $(E_k(\lambda)x, y)$, hence also $(F_k(\lambda)x, y)$, is absolutely continuous for every pair x, y of this space. According to the above mentioned theorem of Rosenblum, it now follows that the operators H_k^{-1} (hence also the H_k) are unitarily equivalent, and the proof of (*) is now complete.

3. Consider the special case of (1) in which $f\equiv 0$. It is readily seen that there are no eigenvalues for either of the boundary value problems determined by $x''+\lambda x=0$ and the respective boundary conditions $x(0)=0$ and $x'(0)=0$. (These boundary conditions correspond to $\alpha=0$,

$\pi/2$ in (2^α); in a somewhat more general connection, cf. [3, p. 792]). Thus, in each case, there is a purely continuous spectrum consisting of the half-line $0 \leqq \lambda < \infty$. Moreover, the basis functions, which, in this instance, are even known explicitly [6, p. 59] are absolutely continuous. Consequently, Theorem (*) is applicable and shows that the self-adjoint operators belonging to the above mentioned boundary value problems are unitarily equivalent.

REFERENCE

1. P. Hartman and A. Wintner, *An oscillation theorem for continuous spectra*, Proc. Nat. Acad. Sci., **33** (1947), 376–379.
2. K. Kodaira, *The eigenvalue problem for ordinary differential equations of the second order and Heisenberg's theory of S-matrices*, Amer. J. Math., **71** (1949), 921–945.
3. C. R. Putnam, *An application of spectral theory to a singular calculus of variations problem*, Amer. J. Math., **70** (1948), 780–803.
4. ———, *On the continuous spectra of singular boundary value problems*, Can. J. Math., **6** (1954), 420–426.
5. M. Rosenblum, *Perturbations of the continuous spectrum and unitary equivalence*, Technical Report No. 12, for Office of Ordnance Research, Department of Mathematics, University of California, December 1955.
6. E. C. Titchmarsh, *Eigenfunction expansions associated with second-order differential equations* (Oxford, 1946).
7. H. Weyl, *Ueber gewöhnliche Differentialgleichungen mit Singularitäten und die zugehörigen Entwicklungen willkürlicher Funktionen*, Math. Ann., **68** (1910), 222–269.

PURDUE UNIVERSITY

PERTURBATION OF THE CONTINUOUS SPECTRUM AND UNITARY EQUIVALENCE

MARVIN ROSENBLUM

1. Introduction. Suppose that A and B are self-adjoint operators in a Hilbert space H such that $B-A=P$ is a completely continuous operator. We shall concern ourselves with the problem of finding conditions sufficient to guarantee that B is unitarily equivalent to A. Clearly a necessary condition is that the spectrum of A (considered as a point set on the real line) is equal to the spectrum of B. However this condition is not sufficient; von Neumann [8] has proved the following result

1.1. *Let A and C be bounded self-adjoint operators in a separable Hilbert space, such that the spectra of A and C have the same limit points. Then there exists an operator B that is unitarily equivalent to C and such that $B-A$ is completely continuous.*

Thus we see that perturbation by a completely continuous operator can radically alter the multiplicity of the spectrum. Even if A and B have pure continuous spectra on the same interval, it does not follow that B is unitarily equivalent to A.

Our present investigation continues along lines begun by Friedrichs in [1] and [2]. He considered bounded operators A that have continuous spectrum of finite multiplicity, and worked in the representation space where A corresponds to a multiplication operator. One of Friedrichs' results is the following.

1.2. *Let $H=L^2(-1, 1)$ and let A be the operator that sends any function $f(x)$ of H into $xf(x)$.. Let P be the integral operator with the hermitian kernel $p(x, y)=\overline{p(y, x)}$, where p satisfies certain Lipschitz conditions. Then if ε is a sufficiently small real number, there exist unitary operators U_ε and V_ε such that*

(i) $e^{-i(A+\varepsilon P)t}e^{iAt}$ *converges strongly to* U_ε as $t \to \infty$;

(ii) $e^{-i(A+\varepsilon P)t}e^{iAt}$ *converges strongly to* V_ε as $t \to -\infty$;

Received April 23, 1956. Research performed in part under contract DA-04-200-ORD-171 Task Order 5 for Office of Ordnance Research, U.S. Army. This paper is a revised version of the author's doctoral dissertation submitted to the University of California, Berkeley. He wishes to thank Professor Frantisek Wolf for his advice and encouragement.

and

(iii) $U_\varepsilon^*(A+\varepsilon P)U_\varepsilon=A$ *and* $V_\varepsilon^*(A+\varepsilon P)V_\varepsilon=A$.

The operator $S=U_\varepsilon^* V_\varepsilon$ is the *scattering operator*, which is of intersect in quantum mechanics; see H. E. Moses [5] and Kay and Moses [4].

We shall make the following assumptions.

Assumption 1.3. $A=\int_{-\infty}^{\infty} x d_x E_x$ and $B=\int_{-\infty}^{\infty} x d_x F_x$ are (possibly unbounded) self-adjoint operators and $B-A=P$ is a completely continuous operator such that the trace of $|P|$ is finite[1].

Assumption 1.4. The spectral measure of A is weakly absolutely continuous, that is, $(E_x f, g)$ is an absolutely continuous function of x for all f, g in H.

We want to find conditions on B that will guarantee that B is unitarily equivalent to A, that is, that there exist a unitary operator U such that $BU=UA$, or equivalently, that $(F_x Uf, g)=(E_x f, U^*g)$ for all f, g in H. Thus a necessary condition is given in

Assumption 1.5. The spectral measure of B is weakly absolutely continuous.

We shall show in this paper that this condition is also sufficient. In fact, we shall prove the following.

THEOREM 1.6. *Suppose that* 1.3, 1.4, *and* 1.5 *hold. Then as* $t\to\infty$, *or* $t\to-\infty$, $e^{-iBt}e^{iAt}$ *converges strongly to unitary operators* U *and* V *respectively, such that* $U^*BU=A$ *and* $V^*BV=A$.

By von Neumann's theorem (see 1.1) Theorem 1.6 is no longer true if P is allowed to be an arbitrary completely continuous operator. It should be noticed that we have imposed no smallness condition on the norm of P, and that A may have continuous spectrum of any multiplicity.

1. Sketch of the proof. Actually, to prove Theorem 1.6, we have only to prove the following (seemingly) weaker result.

2.1. *Assume* 1.3–1.5 *hold. Then as* $t\to\infty$, $e^{-iBt}e^{iAt}$ *converges weak-*

[1] For any self-adjoint $A=\int_{-\infty}^{\infty}\lambda dE_\lambda$ we define $|A|=\int_{-\infty}^{\infty}|\lambda| dE_\lambda$.

ly to an operator U such that $BU=UA$ *and* $\|Uf\|=\|f\|$ *for all* f *in* H.

Proof. We shall deduce Theorem 1.6 from 2.1. We assume that the hypotheses and conclusions of 2.1 are valid. Recall that if a sequence $\{g_n\}$ of elements of H converges weakly to a limit element g, and if $\|g_n\|\to\|g\|$, then g_n converges strongly to g.

Let $f\in H$. $U_t=e^{-iBt}e^{iAt}$ is unitary, so

$$\lim_{t\to\infty}\|U_tf\|=\|f\|=\|Uf\|.$$

But, since U_tf also converges weakly to Uf, it follows from the preceding paragraph that as $t\to\infty$, U_tf converges strongly to Uf. Also $\|f\|=\|Uf\|$ implies that

$$(f,\ f)=\|f\|^2=\|Uf\|^2=(Uf,\ Uf)=(U^*Uf,\ f),\quad \text{so}\quad U^*U=I.$$

Now, 2.1 holds for all choices of A and B that satisfy the Assumptions 1.3–1.5. Since $B-A=P$ we see that $A-B=-P$. Thus we can substitute A for B, B for A, and $-P$ for P, and we can infer from 2.1 and the preceding paragraph that $e^{-iAt}e^{iBt}$ converges strongly to some operator W, and $W^*W=I$. Since $(e^{-iBt}e^{iAt})^*=e^{-iAt}e^{iBt}$ we deduce that $W=U^*$ and that $UU^*=U^*U=I$, so that U is unitary and $U^*BU=A$.

It is also true that $-B-(-A)=-P$, and if we substitute $-A$ for A and $-B$ for B in 2.1 we can repeat the above arguments to prove that as $t\to\infty$, $e^{iBt}e^{-iAt}$ converges strongly to a unitary operator V such that $V^*BV=A$.

In the remainder of this paper we prove 2.1. From now on we assume that assumptions 1.3–1.5 hold. We know that $P=B-A$ has a representation

$$P=\sum_{j=1}^{\infty}\lambda_j(\cdot,\ \phi_j)\phi_j\,,$$

where the ϕ_j are orthonormal and

$$\sum_{j=1}^{\infty}|\lambda_j|<\infty\,.$$

We put $U_t=e^{-iBt}e^{iAt}$. For any complex-valued Lebesgue-measurable function $K(x)$ that is almost everywhere finite we can define the normal operator $K(A)$ by specifying that

$$(K(A)f,\ g)=\int_{-\infty}^{\infty}K(x)\frac{d(E_xf,\ g)}{dx}dx$$

for suitable f in H. In particular, $(e^{-iAt}f,\ g)$ is the Fourier transform of $d(E_x f, g)/dx$. The letters f, g shall denote arbitrary elements in H.

The following sketch of our method of attack may prove instructive. We first derive the representation theorem

$$3.3.\ \ \text{(i)}\ \ (U_t f, g) = (f, g) + \frac{1}{i}\int_0^t (e^{-iBx}Pe^{iAx}f, g)dx .$$

We wish to take $t \to \infty$ and thus exhibit an operator U such that U_t converges weakly to U. But the integrand in 3.3 (i) is not necessarily integrable over $(0, \infty)$. However, we show in 4.4 (i) and 4.7 that there exists a function $w(x)$ that is finite a. e. and such that when f is in the domain of $w(A)$ and g is in the domain of $w(B)$, then the integrand in 3.3 (i) belongs to $L(0, \infty)$. Using this fact we prove in part 5 that U_t converges to an operator U such that $BU = UA$.

Now we have to show that $\| Uf \| = \| f \|$. We proceed in an indirect fashion. Rather than work with U_t we consider the operators $K_n(B)U_t$, where the $K_n(x)$ are a sequence of characteristic functions such that $K_n(x) \to 1$ and such that the integrand in the representation

$$(K_n(B)U_t f, g) = (K_n(B)f, g) + \frac{1}{i}\int_0^t (K_n(B)e^{-iBx}Pe^{iAx}f, g)dx$$

belongs to $L(0, \infty)$ for a dense set of f and *all* g in H. We show that, for each n, $K_n(B)U_t$ converges strongly to $K_n(B)U$, and thus

$$\lim_{n\to\infty}\lim_{t\to\infty}\| K_n(B)U_t f\| = \| Uf \| .$$

By means of representation Theorem 3.5 we show that this iterated limit is also equal to $\| f \|$, and thus $\| Uf \| = \| f \|$, and 2.1 is proved.

3. Derivation of the representation theorems.

LEMMA 3.1. *If s is a complex number with nonzero real part, then*

(i) $(s+iB)^{-1} = (s+iA)^{-1} - i(s+iB)^{-1}P(s+iA)^{-1}$,

(ii) $\qquad = (s+iA)^{-1} - i(s+iA)^{-1}P(s+iB)^{-1}$.

Proof. Since $B - A$ is bounded it follows that A and B have the same domain D. Then, for any $f \in H$, $(s+iB)^{-1}f \in D$, and

$$\begin{aligned} -i(s+iA)^{-1}P(s+iB)^{-1}f &= (s+iA)^{-1}(s+iA-s-iB)(s+iB)^{-1}f \\ &= (s+iB)^{-1}f - (s+iA)^{-1}f . \end{aligned}$$

(ii) follows similarly.

LEMMA 3.2. *If s is a complex number with nonzero real part, and if L is a bounded operator that commutes with $(s+iB)^{-1}$, then*

(i) $(s+iA)^{-1}L-L(s+iA)^{-1}=-i(s+iA)^{-1}[LP-PL](s+iA)^{-1}$.

Proof. By 3.1 (i),

$$L(s+iB)^{-1}=L(s+iA)^{-1}-iL(s+iB)^{-1}P(s+iA)^{-1}$$
$$=L(s+iA)^{-1}-i(s+iB)^{-1}LP(s+iA)^{-1} .$$

By 3.1 (ii), this last expression equals

$$L(s+iA)^{-1}-i(s+iA)^{-1}LP(s+iA)^{-1}-(s+iA)^{-1}P(s+iB)^{-1}LP(s+iA)^{-1} .$$

It can be similarly shown that

$$(s+iB)^{-1}L=(s+iA)^{-1}L-i(s+iA)^{-1}PL(s+iA)^{-1}$$
$$-(s+iA)^{-1}PL(s+iB)^{-1}P(s+iA)^{-1} .$$

Lemma 3.2 follows upon subtracting this last equation from the preceding equation and using the commutativity property of L.

In the following representation theorems all operator integrals are understood to be defined in the weak sense.

THEOREM 3.3. *For any real number t,*

(i) $U_t=I+\frac{1}{i}\int_0^t e^{-iBx}Pe^{iAx}\,dx$, and

(ii) $e^{-iBt}=e^{-iAt}+\frac{1}{i}\int_0^t e^{-iBx}Pe^{iA(x-t)}\,dx$.

Proof. Let s be a complex number with positive real part. Then 3.1 (i) holds and

$$\int_0^\infty e^{-st}e^{-iAt}\,dt=(s+iA)^{-1}$$

for any self-adjoint operator A. Hence by the Laplace transform, convolution and uniqueness theorems as found in Hille [3], chapter 10, we derive (ii).

(i) follows from (ii) by operating on the right with e^{iAt}.

THEOREM 3.4. *If L is a bounded operator that commutes with every bounded function of B, then*

(i) $e^{-iAt}Le^{iAt}=L+\frac{1}{i}\int_0^t e^{-iAx}(LP-PL)e^{iAx}\,dx$.

Proof. We start with 3.2, parallel the proof of Theorem 3.3, and derive the formula

$$e^{-iAt}L - Le^{-iAt} = \frac{1}{i}\int_0^t e^{-iAx}(LP - PL)e^{iA(x-t)}dx .$$

3.4 follows by operating on the right with e^{iAt}.

We shall use 3.4 (i) in the following form.

COROLLARY 3.5. *If K is a projection operator that commutes with every bounded function of B, then*

$$\text{(i)} \quad \|KU_tf\|^2 - \|f\|^2 = ([K-I]f, f) + \frac{1}{i}\int_0^t (e^{-iAx}(K-I)Pe^{iAx}f, f)dx$$

$$- \frac{1}{i}\int_0^t (e^{-iAx}P(K-I)e^{iAx}f, f)dx$$

Proof. We set $L = K - I$ in 3.4 (i) and take inner products. Then the right hand side of 3.4 (i) is equal to the right hand side of 3.5 (i). But,

$$\begin{aligned}(e^{-iAt}[K-I]e^{iAt}f, f) &= (e^{-iAt}Ke^{iAt}f, f) - (f, f)\\ &= \|Ke^{iAt}f\|^2 - \|f\|^2\\ &= \|KU_tf\|^2 - \|f\|^2 ,\end{aligned}$$

so the left sides are equal which proves 3.5.

4. Definition of the $K_n(x)$.

THEOREM 4.1.

(i) $0 \leq \dfrac{d(E_xf, f)}{dx} < \infty$ *for almost all* x;

(ii) $\displaystyle\int_{-\infty}^{\infty}\left|\frac{d(E_xf, g)}{dx}\right|dx \leq \|f\|\cdot\|g\|$

(iii) $\displaystyle\int_{-\infty}^{\infty}\frac{d(E_xf, g)}{dx}dx = (f, g)$; *and*

(iv) $\left|\dfrac{d(E_xf, g)}{dx}\right|^2 \leq \dfrac{d(E_xf, f)}{dx}\cdot\dfrac{d(E_xg, g)}{dx}$ *for almost all* x.

Proof. (i) follows from the fact that (E_xf, f) is a monotone increasing function of x;

(ii) is true because the total variation of $(E_x f,\ g)$ is $\leq \|f\|\cdot\|g\|$, (see Riesz and Nagy [6, p. 340]).

(iii) holds because $E_\infty = I$ and $E_{-\infty} = 0$.

We shall now derive (iv). If h is a nonzero real number, then

$$\left|\frac{([E_{x+h} - E_x]f,\ g)}{h}\right|^2 = \frac{1}{h^2}\,|([E_{x+h} - E_x]f,\ [E_{x+h} - E_x]g)|^2,$$

which by the Schwarz inequality is

$$\leq \left(\left[\frac{E_{x+h} - E_x}{h}\right]f,\ f\right)\cdot\left(\left[\frac{E_{x+h} - E_x}{h}\right]g,\ g\right).$$

Taking h to 0 completes the proof.

LEMMA 4.2.

(i) $\int_{-\infty}^{\infty} |(e^{iAx}f,\ g)|^2\, dx \leq 2\pi \int_{-\infty}^{\infty} \left|\frac{d(E_x f,\ g)}{dx}\right|^2 dx$

(ii) $\lim_{t\to\infty} (e^{iAt}f,\ g) = 0$.

Proof. If $d(E_x f,\ g)/dx$ is not square integrable, then the right hand side is infinite and there is nothing to prove. If $d(E_x f,\ g)/dx$ is square integrable, then its Fourier transform is also square integrable and (i) is an equality.

(ii) is a consequence of 4.1 (ii) and the Riemann-Lebesgue lemma.

LEMMA 4.3. *If Q is a bounded self-adjoint operator, then*

(i) $$\int_{-\infty}^{\infty} \|\ |P|^{\frac{1}{2}} Q e^{iAx} f\|^2 dx \leq 2\pi \int_{-\infty}^{\infty} \left[\sum_{j=1}^{\infty} |\lambda_j|\, d(E_x Q\phi\ ,\ Q\phi_j)/dx\right] d(E_x f, f)/dx\, dx$$
$$\leq 2\pi \text{ ess sup}\, d(E_x f,\ f)/dx \sum_{j=1}^{\infty} |\lambda_j|\ \|Q\phi_j\|^2$$

(ii) $$\int_{-\infty}^{\infty} \|\ |P|^{\frac{1}{2}} Q e^{iBx} f\|^2 dx \leq 2\pi \int_{-\infty}^{\infty} \left[\sum_{j=1}^{\infty} |\lambda_j|\, d(F_x Q\phi_j,\ Q\phi_j)/dx\right] d(F_x f, f)/dx\, dx$$

Proof. In this proof we have nonnegative integrands and thus we freely commute integration and summation.

$$\int_{-\infty}^{\infty} \|\ |P|^{\frac{1}{2}} Q e^{iAx} f\|^2\, dx = \int_{-\infty}^{\infty} \sum_{j=1}^{\infty} |\lambda_j|\ |(e^{iAx}f,\ Q\phi_j)|^2\, dx\ ,$$

which by 4.2 (i) and then 4.1 (iv) is

$$\leq 2\pi \int_{-\infty}^{\infty} \sum_{j=1}^{\infty} |\lambda_j|\, |d(E_x f,\ Q\phi_j)/dx|^2\, dx$$

$$\leq 2\pi\int_{-\infty}^{\infty}\left[\sum_{j=1}^{\infty}|\lambda_j|\, d(E_xQ\phi_j,\ Q\phi_j)/dx\right]d(E_xf,\ f)/dx\,dx\,.$$

The remaining inequality in (i) is readily derived from this and 4.1 (iii).

4.3. (ii) follows similarly.

LEMMA 4.4. *If Q_1 and Q_2 are bounded self-adjoint operators and $-\infty \leq s,\ t \leq \infty$, then*

(i) $$\left[\int_s^t|(e^{-iBx}Pe^{iAx}f,\ g)|dx\right]^2 \leq \int_s^t\|\ |P|^{\frac{1}{2}}e^{iAx}f\|^2\,dx\int_{-\infty}^{\infty}\|\ |P|^{\frac{1}{2}}e^{iBx}g\|^2\,dx\,.$$

(ii) $$\left[\int_{-\infty}^{\infty}|(e^{-iAx}Q_1PQ_2e^{iAx}f,\ f)|\,dx\right]^2$$
$$\leq [2\pi \text{ ess sup}\, d(E_xf,\ f)/dx]\left[\sum_{j=1}^{\infty}|\lambda_j|\ \|Q_1\phi_j\|^2\right]$$
$$\times\left[\sum_{j=1}^{\infty}|\lambda_j|\ \|Q_2\phi_j\|^2\right].$$

Proof. Let C be the operator A or the operator B. Since for any self-adjoint operator P we have the decomposition $P=|P|^{\frac{1}{2}}\text{ sgn } P\ |P|^{\frac{1}{2}}$, we have

$$|(e^{-iCx}Q_1PQ_2e^{iAx}f,\ g)|=|(e^{-iCx}Q_1|P|^{\frac{1}{2}}\text{ sgn } P\ |P|^{\frac{1}{2}}Q_2e^{iAx}f,\ g)|$$
$$=|(\text{sgn } P\ |P|^{\frac{1}{2}}Q_2e^{iAx}f,\ |P|^{\frac{1}{2}}Q_1e^{iCx}g)|\,,$$

which by the Schwarz inequality and the fact that $\|\text{sgn}\,P\|\leq 1$ is

$$\leq\|\ |P|^{\frac{1}{2}}Q_2e^{iAx}f\|\cdot\|\ |P|^{\frac{1}{2}}Q_1e^{iCx}g\,\|\,.$$

By the Schwarz inequality for integrals and the above calculation we see that

$$\left[\int_s^t|(e^{-iCx}Q_1PQ_2e^{iAx}f,\ g)|dx\right]^2$$
$$\leq\int_s^t\|\ |P|^{\frac{1}{2}}Q_2e^{iAx}f\|^2dx\int_s^t\|\ |P|^{\frac{1}{2}}Q_1e^{iCx}g\|^2\,dx$$
$$\leq\int_s^t\|\ |P|^{\frac{1}{2}}Q_2e^{iAx}f\|^2\,dx\int_{-\infty}^{\infty}\|\ |P|^{\frac{1}{2}}Q_1e^{iCx}g\|^2\,dx\,.$$

If we put $Q_1=Q_2=I$ and $C=B$ we see that we have derived (i). If we put $t=\infty$, $s=-\infty$, $C=A$, and employ 4.3 (i), we derive (ii).

DEFINITION 4.5.

$$w(x)=\sum_{j=1}^{\infty}|\lambda_j|[d(E_x\phi_j,\ \phi_j)/dx+d(F_x\phi_j,\ \phi_j)/dx].$$

THEOREM 4.6. *For almost all x, $w(x)$ is nonnegative and finite.*

Proof. From 4.1 (i) and the definition of $w(x)$ it follows that $w(x)$ is nonnegative a.e.

$$\int_{-\infty}^{\infty} w(x)dx=\sum_{j=1}^{\infty}|\lambda_j|\left[\int_{-\infty}^{\infty} d(E_x\phi_j,\ \phi_j)/dx\ dx+\int_{-\infty}^{\infty} d(F_x\phi_j,\ \phi_j)/dx\ dx\right]$$

which by 4.1 (iii), is $2\sum_{j=1}^{\infty}|\lambda_j|$. This last term is finite by assumption. Hence $w(x)$ is integrable, and thus is a. e. finite.

LEMMA 4.7. *If f is in the domain of $w(A)$, then*

(i) $\int_{-\infty}^{\infty}\|\,|P|^{\frac{1}{2}}e^{iAx}f\|^2\,dx\leq 2\pi(w(A)f,\ f)$.

If f is in the domain of $w(B)$, then

(ii) $\int_{-\infty}^{\infty}\|\,|P|^{\frac{1}{2}}e^{iBx}f\|^2\,dx\leq 2\pi(w(B)f,\ f)$.

Proof. By 4.3 (i)

$$\int_{-\infty}^{\infty}\|\,|P|^{\frac{1}{2}}e^{iAx}f\|^2\,dx$$

$$\leq 2\pi\int_{-\infty}^{\infty}\left[\sum_{j=1}^{\infty}|\lambda_j|\,d(E_x\phi_j,\ \phi_j)/dx\right]d(E_xf,\ f)/dx\ dx\,,$$

which is

$$\leq 2\pi\int_{-\infty}^{\infty} w(x)d(E_xf,\ f)/dx\ dx=2\pi(w(A)f,\ f)\,.$$

(ii) follows similarly.

DEFINITION 4.8. For every positive integer n, let $K_n(x)$ be the characteristic function of the set of real number x such that $w(x)=\infty$ or $w(x)\leq n$.

THEOREM 4.9. *For every positive integer n,*

(i) *$K_n(x)$ is a measurable function*

(ii) *$[K_n(x)]^2=K_n(x)$ and $K_n(x)$ is real;*

(iii) $\lim_{n\to\infty} K_n(x)=1$; *and*

(iv) $0 \leqq w(x)K_n(x) \leqq n$ *for almost all* x.

Proof. (i)-(iii) follow immediately from the definition of $K_n(x)$. (iv) is a consequence of 4.6.

THEOREM 4.10. *For every positive integer* n,

(i) $K_n(A)$ *is a projection operator such that*

(ii) $0 \leqq w(A)K_n(A) \leqq n$;

(iii) $\lim_{n\to\infty} (K_n(A)f,\ g)=(f,\ g)$; *and*

(iv) $\lim_{n\to\infty} \|K_n(A)f-f\|=0$.

(i)-(iv) *also hold when* A *is replaced everywhere by* B.

Proof. (i)-(iii) are direct consequences of 4.9. (iv) follows from (iii).

THEOREM 4.11. *Let* n *be any positive integer. Then as* $t \to \infty$, $K_n(B)U_t$ *converges strongly.*

Proof. For f in the domain of $w(A)$ and all g in H

$$|(K_n(B)[U_t-U_s]f,\ g)|^2=|([U_t-U_s]f,\ K_n(B)g)|^2,$$

which by 3.3 (i) and then 4.4 (i) and 4.7 (ii)

$$=\left|\frac{1}{i}\int_s^t (e^{-iBx}Pe^{iAx}f,\ K_n(B)g)dx\right|^2$$

$$\leqq 2\pi\int_s^t \|\,|P|^{\frac{1}{2}}e^{iAx}f\|^2\,dx\cdot(w(B)K_n(B)g,\ K_n(B)g)\ .$$

But by 4.10 (ii) this is

$$\leqq 2\pi\int_s^t \|\,|P|^{\frac{1}{2}}e^{iAx}f\|^2\,dx\cdot n\cdot\|g\|^2\ .$$

Now set $g=K_n(B)[U_t-U_s]f$ in preceding inequality. When then have

$$\|K_n(B)[U_t-U_s]f\|^4$$

$$\leqq 2\pi\int_s^t \|\,|P|^{\frac{1}{2}}e^{iAx}f\|^2\,dx\cdot n\cdot\|K_n(B)[U_t-U_s]f\|^2$$

$$\leqq 8\pi n\,\|f\|^2\int_s^t \|\,|P|^{\frac{1}{2}}e^{iAx}f\|^2\,dx\ ,$$

But, by 4.7 (i) the integrand in this last expression belongs to $L(-\infty, \infty)$. Thus $\lim_{s,t\to\infty} \|K_n(B)[U_t-U_s]f\|=0$.

We have proved that for all f in a dense set, $K_n(B)U_tf$ converges strongly. Since $\|K_n(B)U_t\|\leq 1$, it follows that 4.11 is true.

5. The Operator U.

THEOREM 5.1. *As $t\to\infty$, U_t converges weakly to an operator U. For any f, g in H,*

(i) $(Uf, g)=(f, g)-i\lim_{t\to\infty}\int_0^t(e^{-iBx}Pe^{iAx}f, g)dx$.

Proof. We know from 3.3 (i) that

$$(U_tf, g)=(f, g)-i\int_0^t(e^{-iBx}Pe^{iAx}f, g)dx .$$

The estimates 4.4 (i) and 4.7 assure us that the integrand in this expression belongs to $L(0, \infty)$ for all f in the domain D_1 of $w(A)$ and all g in the domain D_2 of $w(B)$. Thus the bilinear form

$$b(f, g)=b(f, g)=\lim_{t\to\infty}(U_tf, g)$$

is defined on $D_1\times D_2$. Since $\|U_t\|=1$ this form is bounded and it follows from the Frechet-Riesz representation theorem (see Stone [7], p. 63) that there exists a bounded, everywhere defined operator U such that $(Uf, g)=b(f, g)$ for all f, g in $D_1\times D_2$. In fact, since the U_t are uniformly bounded it is the case that $(Uf, g)=\lim_{t\to\infty}(U_tf, g)$ for all f, g in H. Thus 5.1 (i) holds.

LEMMA 5.2. *For all f in H,* $\lim_{t\to\infty}\|Pe^{iAt}f\|=0$.

Proof. Let $\varepsilon>0$. Since P is completely continuous there exists an integer n and an operator

$$P_n=\sum_{j=1}^n \lambda_j(\cdot, \phi_j)\phi_j$$

such that $\|P-P_n\|\leq\varepsilon$. Then

$$\|Pe^{iAt}f\|=\|(P_n+P-P_n)e^{iAt}f\|\leq\|P_ne^{iAt}f\|+\|(P-P_n)e^{iAt}f\|$$
$$\leq\|P_ne^{iAt}f\|+\varepsilon\|f\| .$$

But

$$\|P_n e^{iAt}f\|^2 = \sum_{j=1}^{n} \lambda_j^2 |(e^{iAt}f,\ \phi_j)|^2 ,$$

which, by 4.2 (ii), goes to 0 as $t \to \infty$. Since ε is arbitrary, the proof is complete.

LEMMA 5.3. *For all real* s, $e^{-iBs}U = Ue^{-iAs}$.

Proof.

$$e^{-iBs}U_t - U_t e^{-iAs} = e^{-iBs}e^{-iBt}e^{iAt} - e^{-iBt}e^{iAt}e^{-iAs}$$
$$= e^{-iBt}(e^{-iBs} - e^{-iAs})e^{iAt} ,$$

which by 3.3 (ii)

$$= \frac{1}{i}\int_0^s e^{-iB(x+t)}Pe^{iA(x+t-s)}dx .$$

Thus

$$|(e^{-iBs}U_t f,\ g) - (U_t e^{-iAs}f,\ g)|$$
$$\leq \int_0^s |(e^{-iB(x+t)}Pe^{iA(x+t-s)}f,\ g)|\, dx$$
$$\leq \int_0^s \|Pe^{iA(x+t-s)}f\| \cdot \|g\|\, dx .$$

By the preceding lemma and the bounded convergence theorem this last term goes to 0 as $t \to \infty$. Since U_t converges weakly to U we have $(e^{-iBs}Uf,\ g) - (Ue^{-iAs}f,\ g) = 0$, or 5.3.

THEOREM 5.4. $BU = UA$.

Proof. By 5.3, $(e^{-iBs}Uf,\ g) = (Ue^{-iAs}f,\ g)$, or

$$\int_{-\infty}^{\infty} e^{-ixs}d(F_x Uf,\ g)/dx\ dx = \int_{-\infty}^{\infty} e^{-ixs}d(UE_x f,\ g)/dx\ dx .$$

By the Fourier integral uniqueness theorem, $(F_x Uf,\ g) = (UE_x f,\ g)$, and thus $BU = UA$.

6. Conclusion of the proof.

6.1. We know that as $t \to \infty$, U_t converges weakly to U. Since $K_n(B)U_t$ converges strongly (Theorem 4.11) it follows that it converges strongly to $K_n(B)U$. From this we deduce that for all f in H,

$$\lim_{t\to\infty} \|K_n(B)U_tf\| = \|K_n(B)Uf\|,$$

and by 4.10 (iv),

$$\text{(i)}\quad \lim_{n\to\infty}\lim_{t\to\infty} \|K_n(B)U_tf\| = \|Uf\|.$$

We shall use 3.5 (i) in the form

$$\begin{aligned}\text{(ii)}\quad \|K_n(B)U_tf\|^2 - \|f\|^2 = ([K_n(B)-I]f, f) \\ + \frac{1}{i}\int_0^t (e^{-iAx}[K_n(B)-I]Pe^{iAx}f, f)dx \\ - \frac{1}{i}\int_0^t (e^{-iAx}P[K_n(B)-I]e^{iAx}f, f)dx,\end{aligned}$$

to prove that

$$\text{(iii)}\quad \lim_{n\to\infty}\lim_{t\to\infty} \|K_n(B)U_tf\| = \|f\|,$$

and thus show that $\|Uf\| = \|f\|$. When this has been done we will have proved 2.1, and Theorem 1.6 will follow from the argument after 2.1. It is clear that it is sufficient to show that (iii) is valid for all f is a dense set.

DEFINITION 6.2. Let D be the set of all f in H such that $d(E_xf, f)/dx$ is essentially bounded for all real x.

THEOREM 6.3. *D is dense in H.*

Proof. Let f be an arbitrary element in H. By 4.1 (i) $d(E_xf, f)/dx$ is almost everywhere finite. Let $M_n(x)$ be the characteristic function of the set of all real numbers x such that $d(E_xf, f)/dx \leqq n$ or $d(E_xf, f)/dx = \infty$. Then $M_n(A)f$ is a sequence of elements of D that converges strongly to f. Hence 6.3 is true.

THEOREM 6.4. *If $f \in D$, then* 6.1 (iii) *is true.*

Proof. We shall consider each of the terms on the right hand side of 6.1 (ii). By 4.10 (iii), $\lim_{n\to\infty}([K_n(B)-I]f, f) = 0$.

By 4.4 (ii)

$$\left|\frac{1}{i}\int_0^\infty (e^{-iAx}[K_n(B)-I]Pe^{iAx}f, f)dx\right|$$

$$\leqq 2\pi \text{ ess sup} \frac{d(E_xf, f)}{dx}\left[\sum_{j=1}^\infty |\lambda_j|\ \|[K_n(B)-I]\phi_j\|^2\right]^{\frac{1}{2}}\left[\sum_{j=1}^\infty |\lambda_j|\right]^{\frac{1}{2}},$$

which by 4.10 (iv) goes to 0 as $n \to \infty$. Thus it can be shown that all the terms on the right hand side of 6.1 (ii) go to 0 as $n \to \infty$, and thus 6.1 is true, and our proof of Theorem 1.6 is complete.

We conclude this paper with an interesting representation theorem for $F(B)-F(A)$.

THEOREM 6.5. *Assume* 1.3–1.5 *hold and that* $F(x)$ *is an essentially bounded function. Then*

(i) $\lim_{t\to\infty}(e^{-iAt}F(B)e^{iAt}f,\ g)=(F(A)f,\ g)$ *and*

(ii) $([F(B)-F(A)]f,\ g)=\lim_{t\to\infty} i\int_0^t (e^{-iAx}[F(B)P-PF(B)]e^{iAx}f,\ g)dx$.

Proof. $(e^{-iAt}F(B)e^{iAt}f,\ g)=(U_t^*F(B)U_tf,\ g)=(F(B)U_tf,\ U_tg)$. Since U_tf and U_tg converge strongly to Uf and Ug respectively and U is unitary we have

$$\lim_{t\to\infty}(e^{-iAt}F(B)e^{iAt}f,\ g)=(F(B)Uf,\ Ug)=(UF(A)f,\ Ug)=(F(A)f,\ g)\ .$$

(ii) is a consequence of (i) and Theorem 3.4.

REFERENCES

1. K. O. Friedrichs, *Über die Spektralzerlegung eines Integraloperators*, Math. Ann., **115** (1938), 249–272.

2. ———, *On the perturbation of the continuous spectra*, Communications on Applied Math., **1** (1948), 361–401.

3. E. Hille, *Functional analysis and semi-groups*, New York, 1948.

4. I. Kay and H. E. Moss, *The determination of the scattering potential from the spectral measure function*, N. Y. U. Research Report CX–18; January, 1955.

5. H. E. Moses, *The scattering operator and the adiabatic theorem*, Nuovo Cimento, **1** (1955), 103–131.

6. Riesz, F. and B. Sz.-Nagy, *Leçons d'analyse fonctionnelle*, Budapest, 1952.

7. M. H. Stone, *Linear transformations in Hilbert space*, New York, 1932.

8. J. von Neumann, *Charakterisierung des Spektrums eines Integraloperators*, Actualités Sci. Ind., **229**, 1935.

UNIVERSITY OF CALIFORNIA AND UNIVERSITY OF VIRGINIA

CERTAIN GENERALIZED HYPERGEOMETRIC IDENTITIES OF THE ROGERS-RAMANUJAN TYPE

V. N. SINGH

1. Introduction. In a recent paper H. L. Alder [1] has obtained a generalization of the well-known Rogers-Ramanujan identities. In this paper I have deduced the above generalizations as simple limiting cases of a general transformation in the theory of hypergeometric series given by Sears [5]. This method, besides being much simpler than that of Alder, also gives a simple form for the polynomials $G_{k,\mu}(x)$ given by him. In Alder's proof the polynomials $G_{k,\mu}(x)$ had to be calculated for every fixed k with the help of certain difference equations but in the present case we get directly the general form of these polynomials.

2. Notation. I have used the following notation throughout the paper. Assuming $|x|<1$, let

$$(a)_s \equiv (a;\,s)=(1-a)(1-ax)\cdots(1-ax^{s-1})\,, \qquad (a)_0=1$$

$$\prod^{s}(a_1,\,a_2,\,\cdots,\,a_r;\,b_1,\,b_2,\,\cdots,\,b_t)=\frac{(a_1;\,s)(a_2;\,s)\cdots(a_r;\,s)}{(b_1;\,s)(b_2;\,s)\cdots(b_t;\,s)}$$

$$\prod(a)=\prod_{n=0}^{\infty}(1-ax^n)$$

$$\prod(a_1,\,a_2,\,\cdots,\,a_r;\,b_1,\,b_2,\,\cdots,\,b_t)=\frac{\prod(a_1)\prod(a_2)\cdots\prod(a_r)}{\prod(b_1)\prod(b_2)\cdots\prod(b_t)}$$

$$K_s=\frac{(k;\,s)(x\sqrt{k}\,;\,s)(-x\sqrt{k}\,;\,s)}{(x;\,s)(\sqrt{k}\,;\,s)(-\sqrt{k}\,;\,s)}$$

$$K_{s,r}=K_s\frac{(x^{-r};\,s)}{(kx^{r+1};\,s)}\,x^{rs}$$

$$S_{n,n-1}=\sum_{r_n=0}^{r_{n-1}}\frac{k^{r_n}x^{r_n^2}(x^{r_{n-1}-r_n+1};\,r_n)}{(x;\,r_n)}\,, \qquad S_{1,0}=\sum_{r_1=0}^{r}\frac{k^{r_1}x^{r_1^2}(x^{r-r_1+1};\,r_1)}{(x;\,r_1)}$$

$$T_{n,M}=\sum_{t_n=0}^{\left[\frac{M-n-1}{M-n}-t_{n-1}\right]}\frac{(x^{t_{n-1}-2t_n+1};\,2t_n)x^{-2t_n(t_{n-1}-t_n)}}{(x;\,t_n)(x^{t_{n-2}-2t_{n-1}+1};\,t_n)}\,, \qquad (M=3,\,4,\,5,\cdots)$$

where $[a]$ denotes the integral part of a.

The numbers $s,\,r,\,r_1,\,r_2,\,\cdots,\,t,\,t_1,\,t_2,\,\cdots$ are either zero or positive

Received March 19, 1956.

integers. r_0 and t_0, wherever they occur, have been replaced simply by r and t respectively. Empty products are to mean unity.

3. **Sears** [5, § 4] has proved the following theorem:

$$(3.1)\quad \sum_{s=0}^{\infty} x^{\frac{1}{2}s(s-1)}(kx/a_1a_2)^s \prod^{s}(a_1, a_2; x, kx/a_1, kx/a_2)\theta_s$$
$$=\prod(kx, kx/a_1a_2; kx/a_1, kx/a_2)\sum_{r=0}^{\infty}(kx/a_1a_2)^r \prod^{r}(a_1, a_2; x, kx)$$
$$\times\sum_{t=0}^{r}\frac{(x^{-r}; t)(-1)^t x^{rt}}{(kx^{r+1}; t)(x; t)}\theta_t ,$$

wrere $|kx/a_1a_2|<1$, $|x|<1$ and θ_s is any sequence. The theorem holds provided only that the series on the left converges.

Take

$$\theta_s=\prod^{s}\begin{bmatrix} k, x\sqrt{k}, -x\sqrt{k}, a_3, a_4, \cdots, a_{2M+1}; \\ \sqrt{k}, -\sqrt{k}, kx/a_3, kx/a_4, \cdots, kx/a_{2M+1} \end{bmatrix}$$
$$\times\frac{(k^{M-1}x^{M-1})^s}{(a_3a_4\cdots a_{2M+1})^s}x^{\frac{1}{2}s(1-s)}, \qquad (M=1, 2, 3, \cdots)$$

Then

$$(3.2)\quad \sum_{s=0}^{\infty} K_s\frac{(a_1; s)(a_2; s)\cdots(a_{2M+1}; s)}{(kx/a_1; s)(kx/a_2; s)\cdots(kx/a_{2M+1}; s)}\frac{(k^M x^M)^s}{(a_1a_2\cdots a_{2M+1})^s}$$
$$=\prod(kx, kx/a_1a_2; kx/a_1, kx/a_2)\sum_{r=0}^{\infty}(kx/a_1a_2)^r\prod^{r}(a_1, a_2,; x, kx)$$
$$\times\sum_{t=0}^{r}K_{t,r}\frac{(a_3; t)(a_4; t)\cdots(a_{2M+1}; t)(-1)^t x^{\frac{1}{2}t(1-t)}(k^{M-1}x^{M-1})^t}{(kx/a_3; t)(kx/a_4; t)\cdots(kx/a_{2M+1}; t)(a_3a_4\cdots a_{2M+1})^t} .$$

Now let $a_1, a_2 \cdots, a_{2Mn1}\to\infty$ in (3.2). Then we get

$$(3.3)\quad \sum_{s=0}^{\infty}K_s(-1)^s k^{Ms}x^{\frac{1}{2}s\{(2M+1)s-1\}}$$
$$=\prod(kx)\sum_{r=0}^{\infty}\frac{k^r x^{r^2}}{(x; r)(kx; r)}\sum_{t=0}^{r}K_{t,r}k^{(M-1)t}x^{(M-1)t^2} .$$

And in (3.2) if we take $(M-1)$ for M, $a_1=x^{-r}$ and let $a_2, a_3, \cdots, a_{2M-1}$ tend to ∞, we have

$$(3.4)\quad \sum_{t=0}^{r}K_{t,r}k^{(M-1)t}x^{(M-1)t^2}$$
$$=(kx; r)\sum_{t=0}^{r}\frac{k^t x^{t^2}(x^{r-t+1}; t)}{(x; t)}\sum_{s=0}^{t}K_{s,t}k^{(M-2)s}x^{(M-2)s^2} .$$

On repeated application of (3.4) on the right-hand side of (3.3) it follows that

$$\{\prod (kx)\}^{-1}\sum_{s=0}^{\infty} K_s(-1)^s k^{Ms} x^{\frac{1}{2}s\{(2M+1)s-1\}} = \sum_{r=0}^{\infty}\frac{k^r x^{r^2}}{(x;\, r)}\prod_{n=1}^{M-2} S_{n,n-1}\,,$$

there being $(M-2)$ terminating series on the right since

$$\sum_{s=0}^{t} K_{s,t}=0 \tag{3.5}$$

by Watson's transformation [(2); § 8.5 (2)] of a terminating ${}_8\phi_7$ into a Saalschützian ${}_4\phi_3$.

Now it is easily verified that

$$\prod_{n=1}^{M-2} S_{n,n-1}$$

can, by suitable rearrangements, be simplified to

$$\sum_{t_1=0}^{(M-2)r}\frac{k^{t_1}x^{t_1^2}(x^{r-t_1+1};\, t_1)}{(x;\, t_1)}\sum_{t_2=0}^{\left[\frac{M-3}{M-2}t_1\right]}\frac{(x^{t_1-2t_2+1};\, 2t_2)x^{-2t_2(t_1-t_2)}}{(x;\, t_2)(x^{r-t_1+1};\, t_2)}\prod_{n=3}^{M-2} T_{n,M}\,,$$

where $t_h=r_h+r_{h+1}+\cdots+r_{M-2}$, $(h=1, 2, \cdots, M-2)$.

Thus on putting $r+t_1=t$, we finally have

$$\begin{aligned}\{\prod(kx)\}^{-1}\sum_{s=0}^{\infty} K_s(-1)^s k^{Ms}x^{\frac{1}{2}s\{(2M+1)s-1\}}\\ =\sum_{t=0}^{\infty}\frac{k^t x^{t^2}}{(x;\, t)}\sum_{t_1=0}^{\left[\frac{M-2}{M-1}t\right]}\frac{(x^{t-2t_1+1};\, 2t_1)x^{-2t_1(t-t_1)}}{(x;\, t_1)}\prod_{n=2}^{M-2} T_{n,M}\,.\end{aligned} \tag{3.6}$$

This is a k-cum-M generalization of the Rogers-Ramanujan identities. For any assigned values of M and t, the repeated terminating series can, by dividing out by the denominator factors, be evaluated as polynomials in x.

Let us now write

$$G_{M,t}(x)=x^{t^2}\sum_{t_1=0}^{\left[\frac{M-2}{M-1}t\right]}\frac{(x^{t-2t_1+1};\, 2t_1)x^{-2t_1(t-t_1)}}{(x;\, t_1)}\prod_{n=2}^{M-2} T_{n,M}\,. \tag{3.7}$$

Then, as usual, for $k=1$ and $k=x$ respectively, the left-hand side of (3.6) can be expressed as a product by means of Jacobi's classical identity

$$\sum_{n=-\infty}^{\infty}(-1)^n x^{n^2}z^n=\prod_{n=1}^{\infty}(1-x^{2n-1}z)(1-x^{2n-1}/z)(1-x^{2n}) \tag{3.8}$$

and we get Alder's generalization of the first and second Rogers-

Ramanujan identities in the form

$$(3.9)\qquad \prod_{n=0}^{\infty} \frac{(1-x^{(2M+1)n+M})(1-x^{(2M+1)n+M+1})}{(1-x^{(2M+1)n+1})(1-x^{(2M+1)n+2})\cdots(1-x^{(2M+1)n+2M})} = \sum_{t=0}^{\infty} \frac{G_{M,t}(x)}{(x;\,t)}$$

and

$$(3.10)\qquad \prod_{n=0}^{\infty} \frac{1}{(1-x^{(2M+1)n+2})(1-x^{(2M+1)n+3})\cdots(1-x^{(2M+1)n+2M-1})} = \sum_{t=0}^{\infty} \frac{x^t G_{M,t}(x)}{(x;\,t)}$$

where $G_{M,t}(x)$ is given by (3.7). The polynomials $G_{M,t}(x)$ can be seen by easy verification to be identical with $G_{k,\mu}(x)$ of Alder.

I am grateful to Dr. R. P. Agarwal for suggesting this problem and for his kind guidance in the preparation of this paper.

Added in Proof. If in (3.2) we take $a_1 = -\sqrt{kx}$, make $a_2, a_3, \cdots, a_{2M+1}$ tend to ∞, and proceed as in § 3, we get for $k=1$ and $k=x$ the respective identities

$$\prod_{n=1}^{\infty} \frac{(1-x^{2Mn-(M-\frac{1}{2})})(1-x^{2Mn-(M+\frac{1}{2})})(1-x^{2Mn})}{(1-x^n)}$$

$$= \{\Pi(-x^{\frac{1}{2}})\}^{-1} \sum_{t=0}^{\infty} \frac{x^{\frac{1}{2}t^2}(-x^{\frac{1}{2}})_t}{(x)_t} \sum_{t_1=0}^{\left[\frac{M-2}{M-1}t\right]} \frac{x^{-t_1(t-\frac{3}{2}t_1)}}{(x)_{t_1}} \times \frac{(x^{t-2t_1+1})_{2t_1}}{(-x^{\frac{1}{2}+t-t_1})_{t_1}} \prod_{n=2}^{M-2} T_{n,M}$$

and

$$\prod_{n=1}^{\infty} \frac{(1-x^{2Mn-1})(1-x^{2Mn-(2M-1)})(1-x^{2Mn})}{(1-x^n)}$$

$$= \{\pi(-x)\}^{-1} \sum_{t=0}^{\infty} \frac{x^{\frac{1}{2}t(t+1)}(-x)_t}{(x)_t} \sum_{t_1=0}^{\left[\frac{M-2}{M-1}\right]} \frac{x^{\frac{1}{2}t_1}x^{-t_1(t-\frac{3}{2}t_1)}}{(x)_{t_1}} \times \frac{(x^{t-2t_1+1})_{2t_1}}{(-x^{1+t-t_1})_{t_1}} \prod_{n=2}^{M-2} T_{n,M}\,.$$

REFERENCES

1. H. L. Alder, *Generalizations of the Rogers-Ramanujan identities*, Pacific J. Math., **4** (1954), 161-168.
2. W. N. Bailey, *Generalized Hypergeometric Series* (1935), Cambridge Tract No. 32.
3. ———, *Identities of the Rogers-Ramanujan type*, Proc. Lond. Math. Soc. (2), **50** (1949), 1-10.
4. ———, *Some identities in combinatory analysis*, ibid., **49** (1947), 421-435.
5. D. B. Sears, *Transformations of basic hypergeometric functions of special type*, ibid., **52** (1951), 467-483.

LUCKNOW UNIVERSITY

FAMILIES OF TRANSFORMATIONS IN THE FUNCTION SPACES H^p

P. SWERLING

I. Introduction

Let the interior of the unit circle be denoted by Δ; and let the set of functions single-valued and analytic in Δ be denoted by $\mathfrak{A}$.

It is well known that certain subsets of $\mathfrak{A}$ can be made into Banach spaces by the introduction of suitable norms. In particular, if $f \in \mathfrak{A}$, and if, for $1 \leq p \leq \infty$,

$$(I.1) \qquad \mathscr{M}_p(f, r) = \left\{ \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^p d\theta \right\}^{1/p}, \qquad p < \infty$$

$$\mathscr{M}_p(f; r) = \sup_{|z|<r} |f(z)|, \qquad p = \infty$$

and if $\sup_{r<1} \mathscr{M}_p(f; r) < \infty$, then f is said to be in the set H^p. Also, H^p is a Banach space with

$$(I.2) \qquad \|f\|_{H^p} = \sup_{r<1} \mathscr{M}_p(f; r)$$

A proof of these statements, together with a discussion of many properties of the spaces H^p, can be found in [8].

This paper is concerned with certain transformations in the spaces H^p [1].

Let $\omega(z)$ be a function of z which is analytic in Δ and such that $|\omega(z)| < 1$ for $z \in \Delta$. If $f \in \mathfrak{A}$, then so is the function defined by $f[\omega(z)]$. For $f \in \mathfrak{A}$, we define

$$(I.3) \qquad T_\omega f = g \underset{df}{\Longleftrightarrow} f[\omega(z)] = g(z) \text{ for } z \in \Delta.$$

T_ω is clearly an additive, homogeneous transformation.

It is well known [4] that if $f \in H^p$ and $\omega(0) = 0$, then $T_\omega f \in H^p$ and $\|T_\omega f\| \leq \|f\|$. In other words, if $\omega(0) = 0$, then $T_\omega \in [H^p]$ (the set of all linear bounded transformations on H^p to H^p), and $\|T_\omega\| \leq 1$. Our first task is to prove the following.

Received September 21, 1955, and revised form April 27, 1956. This paper forms a portion of Ph. D. thesis submitted at the University of California, Los Angeles. The author wishes to express his gratitude to Professor A. E. Taylor, under whose guidance the thesis was written.

[1] In the following, all statements about H^p refer to $1 \leq p \leq \infty$ unless further qualified.

THEOREM I.1. *If $\omega \in \mathfrak{A}$ and $|\omega(z)|<1$ for $z \in \Delta$, and if $|\omega(0)|=\alpha<1$, then $T_\omega \in [H^p]$ and $\|T_\omega\| \leqq \left(\frac{1+\alpha}{1-\alpha}\right)^{1/p}$ There is at least one such ω for which the equality holds.*

Proof. For $p=\infty$, the theorem is trivial. For $1 \leqq p < \infty$, a simple proof (for which the author is indebted to the referee) is as follows.

For $f \in H^p$, let u be the least harmonic majorant of $|f|^p$ in Δ (see [6]). Then $T_\omega u$ is a harmonic majorant of $|T_\omega f|^p$. Also,

$$\|f\| = \{u(0)\}^{1/p} \text{ and } \|T_\omega f\| \leqq \{(T_\omega u)(0)\}^{1/p} = \{u(\beta)\}^{1/p}$$

where $\beta=\omega(0)$. The Poisson integral for u shows that

$$u(\beta) \leqq u(0)\left(\frac{1+|\beta|}{1-|\beta|}\right)$$

Putting $\alpha=|\beta|$, it follows that

$$\|T_\omega f\| \leqq \|f\| \left(\frac{1+\alpha}{1-\alpha}\right)^{1/p}.$$

To complete the proof, we note that the following statement holds. Define the transformation L_α $(0 \leqq \alpha < 1)$ by

$$L_\alpha f(z) = f\left(\frac{z+\alpha}{1+\alpha z}\right).$$

Then the function

$$f(z) = \left(\frac{z+1}{z-1}\right)^\eta$$

is an eigenfunction of $L_\alpha : L_\alpha f = \lambda f$, belonging to the eigenvalue

$$\lambda = \left(\frac{1+\alpha}{1-\alpha}\right)^\eta,$$

provided $|\Re\eta| < 1/p$. This follows trivially from the fact that $f \in H^p$ provided $|\Re\eta| < 1/p$.

The result stated in Theorem I.1 can be sharpened as follows.

COROLLARY I.1. *For any ω ($\omega \in \mathfrak{A}$, mapping Δ into or onto itself),*

$$\text{(I.4)} \qquad \|T_\omega\| \leqq \inf_{\substack{\zeta\in\Delta \\ \eta\in\Delta}} \left\{\left(\frac{1+|\zeta|}{1-|\zeta|}\right)\left(\frac{1+|\eta|}{1-|\eta|}\right)\left(\frac{1+|\Gamma_\omega(\eta,\zeta)|}{1-|\Gamma_\omega(\eta,\zeta)|}\right)\right\}^{1/p}$$

where

$$\Gamma_\omega(\eta, \zeta)=\frac{\omega(\eta)+\zeta}{1+\bar{\zeta}\omega(\eta)}$$

Proof. For $\zeta \in \Delta$, define L_ζ by

$$L_\zeta f(z)=f\left(\frac{z+\zeta}{1+\bar{\zeta}z}\right)$$

Then

$$T_\omega=L_{-\eta}L_\eta T_\omega L_\zeta L_{-\zeta}$$

where

$$\eta \in \Delta, \ \zeta \in \Delta$$

so that

$$\| T_\omega \| \leqq \| L_{-\eta} \| \ \| L_{-\zeta} \| \ \| L_\eta T_\omega L_\zeta \|$$

Now, $\frac{z-\zeta}{1-\bar{\zeta}z}$ takes 0 into $-\zeta$; $\frac{z-\eta}{1-\bar{\eta}z}$ takes 0 into $-\eta$;

and $\omega\left(\frac{z+\eta}{1+\bar{\eta}z}\right)+\zeta \Big/ 1+\bar{\zeta}\omega\left(\frac{z+\eta}{1+\bar{\eta}z}\right)$ takes 0 into $\frac{\omega(\eta)+\zeta}{1+\bar{\zeta}\omega(\eta)}$

Applying Theorem I.1, we obtain (I.4).

We are thus assured that a transformation T_ω defined by $T_\omega f(z)=f[\omega(z)]$ is a member of $[H^p]$, $1 \leqq p \leqq \infty$. § II is devoted to a study of semigroups and groups of these transformations. Section III contains a discussion of two examples which illustrate the theorems of § II.

II. Families of Transformations in H^p

A. Definitions and preliminary results. Consider a family of functions $\{\omega(z; t)\}$ — also denoted by $\{\omega_t(z)\}$ — where $z \in \Delta$ and t belongs to a set $\mathscr{T}$ of complex numbers. The individual functions will be denoted by $\omega(z; t)$ or by $\omega_t(z)$, according to convenience.

Let the set $\mathscr{T}$ satisfy the following conditions.

(CII.1) (i) *If* $t_1, t_2 \in \mathscr{T}$, *then* $t_1+t_2 \in \mathscr{T}$.

(ii) $\mathscr{T}$ *contains the origin and some ray originating at the origin.*

(iii) *Every two points in* $\mathscr{T}$ *can be connected by a path*[2] *in* $\mathscr{T}$.

[2] Here a path is defined to mean a finite number of rectifiable Jordan arcs joined together; see [**3**, pp 13, 14].

Let the family $\{\omega(z;\ t)\}$ satisfy the following conditions:

(CII.2) (i) *For each* $t \in \mathcal{T}$, $\omega_t \in \mathfrak{A}$, *and* ω_t *maps* Δ *into* (*or onto*) *itself.*

(ii) *For* $t_1,\ t_2 \in \mathcal{T}$, *and* $z \in \Delta$,

$$\omega_{t_2}[\omega_{t_1}(z)] = \omega_{t_1}[\omega_{t_2}(z)] = \omega_{t_1+t_2}(z)$$

(iii) $\omega(z;\ 0) = z$ *for* $z \in \Delta$.

(iv) *For each* $z \in \Delta$, $\omega(z;\ t)$ *is differentiable*[3] *with respect to* t *for* $t \in \mathcal{T}$. *Also, if*

$$P(z) = \frac{\partial}{\partial t}\omega(z;\ t)|_{t=0}\ ,$$

then $P \in \mathfrak{A}$.

We can immediately state the following.

LEMMA II.1. *For fixed* $z \in \Delta$,

$$\text{(II.1)} \qquad \frac{\partial}{\partial t}[\omega(z;\ t)] = P[\omega(z;\ t)]$$

Proof. $\omega[\omega(z;\ t);\ h] = \omega(z;\ t+h)$ for $t,\ h \in \mathcal{T}$

Therefore

$$\frac{\omega(z;\ t+h) - \omega(z;\ t)}{h} = \frac{\omega[\omega(z;\ t);\ h] - \omega(z;\ t)}{h}$$

$$= \frac{\omega[\omega(z;\ t);\ h] - \omega[\omega(z;\ t);\ 0]}{h}$$

Letting $h \to 0$ (in $\mathcal{T}$), we obtain (II.1).

The family of transformations $\{T_{\omega_t}\}$ defined by (I.3) with $\omega = \omega_t$ will henceforth be denoted simply by $\{T_t\}$. This family forms a semi-group (possibly a group) of linear bounded transformations in the spaces H^p. (The boundedness is shown by Theorem I.1.)

We define the generator A of the family $\{T_t\}$ by

$$\text{(II.2)} \qquad Af = \lim_{t \to 0} \frac{T_t f - f}{t}\ , \qquad f \in H^p$$

the limit taken in the strong sense in H^p. The domain of A, denoted

[3] Here and in the following, "differentiability with respect to t for $t \in \mathcal{T}$" implies that the difference quotient approaches the *same* limit no matter how t is approached (as long as the approach is made entirely in $\mathcal{T}$).

by $\mathscr{D}(A)$, is defined to be the subset of H^p for which the limit in (II.2) exists as $t \to 0$, $t \in \mathscr{T}$ (the limit to be the same for all modes of approach within $\mathscr{T}$ to 0).

It follows from (II.2) that, for $f \in \mathscr{D}(A)$, and each $z \in \Delta$,

$$Af(z) = \lim_{t \to 0} \frac{T_t f(z) - f(z)}{t} \tag{II.3}$$

This is true since, for fixed $z \in \Delta$, $f(z)$ is a bounded linear functional of f, [7].

Now

$$\begin{aligned} Af(z) &= \lim_{t \to 0} \frac{f[\omega(z;\ t)] - f(z)}{t} \\ &= \lim_{t \to 0} \frac{f[\omega(z;\ t)] - f[\omega(z;\ 0)]}{t} \\ &= \frac{\partial}{\partial t} f[\omega(z;\ t)]|_{t=0} = f'[\omega(z;\ t)] \frac{\partial}{\partial t} \omega(z;\ t)|_{t=0} \end{aligned}$$

or

$$Af(z) = P(z) f'(z) \qquad z \in \Delta,\ f \in \mathscr{D}(A) \tag{II.4}$$

It is thus clear that $\mathscr{D}(A)$ is contained in the subset of H^p consisting of those elements f for which $f'(z)P(z)$ defines an element of H^p.

B. Differentiability properties of the family $\{T_t\}$

THEOREM II.1. *Let f be in H^p, and t_0 be in $\mathscr{T}$; let $g(z) = P(z) f'(z)$ and suppose that*

(i) *There exists a neighborhood $\mathscr{N}_{t_0}$ of t_0 and a positive constant M such that every point t of $\mathscr{N}_{t_0}$ can be connected to t_0 by a polygonal line in $\mathscr{N}_{t_0} \cap \mathscr{T}$ of length $\leq M|t_0 - t|$;*

(ii) *$T_t g \in H^p$ for $t \in \mathscr{N}_{t_0} \cap \mathscr{T}$;*

(iii) *$\| T_t g - T_{t_0} g \| \to 0$ as $t \to t_0$ ($t \in \mathscr{T}$).*

Then, $T_t f$ is strongly differentiable with respect to t at t_0 and

$$\frac{d}{dt} T_t f|_{t=t_0} = T_{t_0} g\ . \tag{II.5}$$

Before giving the proof, the following formal derivation might be of interest

$$\lim_{t\to t_0}\frac{T_tf-T_{t_0}f}{t-t_0}=\lim_{s\to 0}T_{t_0}\left\{\frac{T_sf-f}{s}\right\} \qquad (s=t-t_0)$$

$$=T_{t_0}Af=T_{t_0}g$$

This is however not a rigorous proof, even when $f\in\mathscr{D}(A)$, since $s=t-t_0$ may not be in $\mathscr{T}$ for all $t\in\mathscr{N}_{t_0}\cap\mathscr{T}$.

A rigorous proof is as follows.

Let $f[\omega(z;\ t)]=h(z;\ t)$ and let

$$D(z;\ t;\ t_0)=\frac{h(z;\ t)-h(z;\ t_0)}{t-t_0}-T_{t_0}g(z) \tag{II.6}$$

If $z=re^{i\theta}$, and if $\frac{\partial}{\partial t}h(z;\ t)$ is denoted by $h_t(z;\ t)$, then, from (II.1),

$$D(z;\ t;\ t_0)=\frac{h(re^{i\theta};\ t)-h(re^{i\theta};\ t_0)}{t-t_0}-h_t(re^{i\theta};\ t_0)$$

$$=\frac{1}{t-t_0}\int_{t_0}^{t}[h_t(re^{i\theta};\ \tau)-h_t(re^{i\theta};\ t_0)]d\tau$$

where t is chosen in $\mathscr{N}_{t_0}$ and the integral is taken along a polygonal line in $\mathscr{N}_{t_0}\cap\mathscr{T}$ connecting t and t_0 and of length $\leqq M|t-t_0|$.

First suppose that $1\leqq p<\infty$. Then

$$\mathscr{M}_p(D;\ r)=\left\{\frac{1}{2\pi}\int_0^{2\pi}|D(re^{i\theta};\ t;\ t_0)|^p\,d\theta\right\}^{1/p} \tag{II.7}$$

$$=\left\{\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{1}{t-t_0}\int_{t_0}^{t}[h_t(re^{i\theta};\ \tau)-h_t(re^{i\theta};\ t_0)]d\tau\right|^p d\theta\right\}^{1/p}$$

Let $\tau=\tau(s)$, $0\leqq s\leqq 1$, $\tau(0)=t_0$, $\tau(1)=t$. Here s is a constant times the arc length. Then [4], [1]

$$\mathscr{M}_p(D;\ r)=\left\{\frac{1}{2\pi}\int_0^{2\pi}\left|\frac{1}{t-t_0}\int_0^1[h_t(re^{i\theta};\ \tau)-h_t(re^{i\theta};\ t_0)]\tau'(s)ds\right|^p d\theta\right\}^{1/p}$$

$$\leqq\frac{1}{|t-t_0|}\int_0^1|\tau'(s)|\left\{\frac{1}{2\pi}\int_0^{2\pi}\left|h_t(re^{i\theta};\ \tau)-h_t(re^{i\theta};\ t_0)\right|^p d\theta\right\}^{1/p}$$

Hence,

$$\|D\|=\left\|\frac{T_tf-T_{t_0}f}{t-t_0}-T_{t_0}g\right\|=\sup_{r<1}\mathscr{M}_p(D;\ r)$$

$$\leqq \frac{1}{|t-t_0|}\int_0^1 |\tau'(s)|\left[\sup_{r<1}\left\{\frac{1}{2\pi}\int_0^{2\pi}\left|h_t(re^{i\theta};\tau)-h_t(re^{i\theta};t_0)\right|^p d\theta\right\}^{1/p}\right]ds$$

$$=\frac{1}{|t-t_0|}\int_0^1 |\tau'(s)|\ \| T_\tau g - T_{t_0} g\| \, ds \leqq M \sup_{0\leqq s\leqq 1} \| T_\tau g - T_{t_0} g\|$$

Now, by (iii), as $t \to t_0$, the quantity $\sup_{0\leqq s\leqq 1} \| T_\tau g - T_{t_0} g\|$ goes to zero. Thus $\| D\| \to 0$ as $t \to t_0$.

For $p=\infty$, the proof follows similar lines.

COROLLARY II.1–1. *Let f be in H^p, t_0 be in $\mathcal{T}$ and let $g(z)=P(z)f'(z)$. Suppose condition* (i) *of Theorem* II.1 *holds and in addition, suppose that*

(a) $|\omega(z;\ t_0)| < r < 1$ *for* $z \in \Delta$

(b) *$\omega(z;\ t)$ is continuous with respect to t at t_0, uniformly in z for $z \in \Delta$.*

Then, $T_t f$ is differentiable with respect to t at t_0 and (II.5) *holds.*

Proof. By (b), there exists a neighborhood $\mathcal{N}_{t_0}'$ of t_0 such that $|\omega(z;\ t)| < r' < 1$ for $z \in \Delta$, $t \in \mathcal{N}_{t_0}' \cap \mathcal{T}$.

Now, $g(z)$ is analytic in Δ. Therefore for $t \in \mathcal{N}_{t_0}' \cap \mathcal{T}$, $T_t g(z) = g[\omega(z;\ t)]$ is bounded in Δ and therefore $T_t g \in H^p$.

Also, $T_t g(z)$ is continuous with respect to t at t_0, uniformly in z for $z \in \Delta$. Hence $\sup_{z\in\Delta} |T_t g(z) - T_{t_0} g(z)| \to 0$ as $t \to t_0$.

THEOREM II.2. *Suppose*

(i) *Condition* (i) *of Theorem* II.1 *holds for* $t_0=0$;

(ii) $\| T_t f - f\| \to 0$ *as* $t \to 0$ ($t \in \mathcal{T}$) *for every* $f \in H^p$.

Then, $\mathcal{D}(A)$, the domain of the generator A (defined by II.2*), is the set of elements $f \in H^p$ for which $g(z)=f'(z)P(z)$ defines an element g of H^p.*

Proof. Let $\mathcal{G}$ denote the set of elements $f \in H^p$ such that $g(z) = f'(z)P(z)$ defines an element g of H^p. We already know (last paragraph of IIA) that $\mathcal{D}(A) \subset \mathcal{G}$. To show that $\mathcal{G} \subset \mathcal{D}(A)$, one must verify conditions (ii) and (iii) of Theorem II.1 for $f \in \mathcal{G}$, $t_0=0$.

Since $f \in \mathcal{G}$ implies $g \in H^p$, it follows from Theorem I.1 that $T_t g \in H^p$ for all $t \in \mathcal{T}$. Also, condition (iii) of Theorem II.1 is obtained for $t_0=0$ by applying condition (ii) of Theorem II.2 to the function g. Equation (II.5) becomes

$$\text{(II.8)} \qquad Af=g \quad \text{where} \quad g(z)=P(z)f'(z) .$$

THEOREM II.3. *Under conditions* (i) *and* (ii) *of Theorem* II.2, *A is a closed transformation. Also* $\mathscr{D}(A)$ *is dense in* H^p.

Proof. Let f_n be in $\mathscr{D}(A)$; $f_n \to f$ (in the norm of H^p) $Af_n \to g \in H^p$ (in the norm of H^p). Then [7]

$$\left.\begin{aligned} f_n(z) &\to f(z) \\ P(z)f_n'(z) &\to g(z) \end{aligned}\right\} \text{uniformly on compact subsets of } \Delta,$$

that is, $g(z)=P(z)f'(z)$ for $z \in \Delta$.

Therefore, since $g \in H^p$, then, by Theorem II.2, $f \in \mathscr{D}(A)$ and $Af=g$. See [2, Chap. 11] for the fact that $\mathscr{D}(A)$ is dense in H^p.

C. The family of transformations generated by a given operator of the form $Af(z)=P(z)f'(z)$. Suppose P is a given function in $\mathfrak{A}$. The following question arises: Is there a set $\mathscr{T}$ in the complex plane and a set of functions $\{\omega_t\}$ satisfying, respectively, conditions CII.1 and CII.2? If so, how, knowing just $P(z)$, can one determine the family $\{\omega_t\}$ and the maximum set $\mathscr{T}$?

To investigate these questions, additional conditions will be imposed on the given function $P(z)$. First,

(CII.3) $1/P(z)$ *is analytic in* Δ *except, possibly, for a single pole.*

Let the function $Q(z)$ be defined by

$$\text{(II.9)} \qquad Q(z)=\int_{z_0}^{z} \frac{d\zeta}{P(\zeta)} \qquad z_0,\ z \in \Delta$$

The path of integration is chosen in Δ so as not to pass through any singularity of $1/P(z)$; also, z_0 is chosen so as not to be a singularity of $1/P(z)$. $Q(z)$ may be a many-valued function.

$Q(z)$ depends on the choice of z_0; however, as will become clear below, it is not worthwhile to express this dependence in the notation. Clearly, all definitions of Q (corresponding to different choices of z_0) differ from each other by additive constants.

The following property of Q is worth noting.

Let z_1 and z_2 be in Δ, and not singularities of $1/P(z)$; let $Q^{(1)}(z_1)$, $Q^{(2)}(z_1)$ be two values of Q at $z=z_1$; and let $Q^{(1)}(z_1)-Q^{(2)}(z_1)=h$. Let $Q^{(1)}(z_2)$ be a value of Q at $z=z_2$. There exists a value of Q at $z=z_2$, which may be denoted by $Q^{(2)}(z_2)$, such that $Q^{(1)}(z_2)-Q^{(2)}(z_2)=h$. This is clear from the definition of Q and from (CII.3).

We shall further assume:

(CII.4) *If* z_1 *and* z_2 *are in* Δ, *are not singularities of* $1/P(z)$, *and* $z_1 \neq z_2$, *then* $Q(z_1) \neq Q(z_2)$.

This may, of course, be regarded as a condition on $P(z)$.

Now suppose $P \in \mathfrak{A}$ is given satisfying (CII.3) and (CII.4), and that a set $\mathscr{T}$ and a family $\{\omega_t\}$ exist satisfying (CII.1) and (CII.2). From (II.1) and (CII.2–iii), regarding z as fixed for the moment, one can write

$$\text{(II.10)} \quad \left.\begin{aligned} \frac{d}{dt}\,\omega(z;\ t) &= P[\omega(z;\ t)] \\ \omega(z;\ 0) &= z \end{aligned}\right\} \quad \begin{aligned} z &\in \Delta \\ t &\in \mathscr{T} \end{aligned}$$

Let z be fixed in Δ and not a singularity of $1/P(z)$. Then, from (II.10), $\omega(z;\ t)$ must satisfy

$$\text{(II.11)} \quad Q[\omega(z;\ t)] = Q(z) + t\,.$$

Now, for fixed $t \in \mathscr{T}$, $\omega(z;\ t)$ must be an analytic function of z in Δ, mapping Δ into itself.

Let I_Q be the image under Q of Δ (excluding the possible singularity of $1/P(z)$. The set I_Q includes *all* values of $Q(z)$ which can be obtained by integrating in (II.9) along paths which are entirely in Δ. If $\omega(z;\ t)$, for fixed $t \in \mathscr{T}$, is defined for all $z \in \Delta$, and such that $|\omega(z;\ t)| < 1$, then (II.11) implies that this t must translate I_Q into a subset of itself: $I_Q + t \subset I_Q$.

Let $\mathscr{T}_Q$ be the set of translations of I_Q into or onto itself. (Clearly $\mathscr{T}_Q$ does not depend on the choice of z_0 in defining Q.) Then $\mathscr{T} \subset \mathscr{T}_Q$.

On the other hand if P being given[4], $\mathscr{T}_Q$ contains a subset $\mathscr{T}^*$ satisfying conditions (CII.1), then a family $\{\omega_t\}$ satisfying (CII.2) exists (with $t \in \mathscr{T}^*$).

Define, for $t \in \mathscr{T}^*$, $z \in \Delta$,

$$\text{(II.12)} \quad \omega(z;\ t) = \begin{cases} Q^{-1}[Q(z)+t], & z \text{ not a singularity of } \dfrac{1}{P(z)} \\ z, & z \text{ a singularity of } \dfrac{1}{P(z)} \end{cases}$$

where Q^{-1} denotes the function inverse to Q.

This definition defines ω uniquely. If $Q(z)$ refers to a particular branch of Q, then ω is uniquely determined (in Δ) because of (CII.4); moreover, by the property of Q mentioned on p. it is seen that the same point ω is defined no matter what branch of Q is used in (II : 12).

[4] $P \in \mathfrak{A}$ and satisfying (CII. 3) and (CII. 4).

It is also clear that $\omega(z;\, t)$ does not depend on the choice of z_0.

The function $\omega(z;\, t)$ thus defined is analytic in z for each $t \in \mathscr{T}^*$. This is clear if z is not a singularity of $1/P(z)$. If z_1 is a singularity of $1/P(z)$ in $\varDelta$, it is necessary to show that $\omega(z;\, t)$ is (for fixed t) continuous at $z=z_1$; that is, (from II.12) $\omega_t(z) \to z_1$ as $z \to z_1$.

Since z_1 is a pole of $1/P(z)$, one can say, by the definition of Q, that there exist points $\omega_t(z)$ approaching z_1 as $z \to z_1$, such that (II.12) is satisfied. But, by (CII.4), these points are the only ones in $\varDelta$ for which (II.12) is satisfied.

The other conditions of (CII.2) are readily verified for the functions $\omega(z;\, t)$ as defined by (II.12).

The preceding results may be summed up as follows.

THEOREM 11.4. *Let $P(z)$ be in $\mathfrak{A}$, satisfying* (CII.3) *and* (CII.4). *Let $Q(z)$ be defined by* (II.9); *let I_Q be the image of $\varDelta$ under Q, let $\mathscr{T}_Q$ be the set of translations of I_Q into or onto itself.*

Then, there exists a set $\mathscr{T}$ and a family $\{\omega_t\}$ satisfying (CII.1) *and* (CII.2), *if and only if $\mathscr{T}_Q$ contains a subset $\mathscr{T}^*$ satisfying* (CII.1). *The maximum set $\mathscr{T}$ is the "direct sum" of all subsets of $\mathscr{T}_Q$ which satisfy* (CII.1). *Here "direct sum" is defined as follows: If $\{G_\alpha\}$ is a collection of subsets of the complex plane, each containing the origin, the direcct sum of the sets $\{G^\alpha\}$ is defined to be the set consisting of all elements of the form $t=t_1+\cdots+t_n$ where n is a finite* (*positive*) *integer and where $t_i \in \bigcup_\alpha G_\alpha$.*

The last statement follows from the fact that the direct sum of subsets of $\mathscr{T}_Q$ satisfying (C.II.1) is also a subset of $\mathscr{T}_Q$ which satisfies (C.II.1).

One result of the previous theorem is the following.

THEOREM II.5. *If $P(z) \in \mathfrak{A}$, satisfying* (CII.3) *and* (CII.4), *and if there exists a set $\mathscr{T}$ and a family $\{\omega_t\}$ satisfying* (CII.1) *and* (CII.2), *then $1/P(z)$ can have only a pole of first order in $\varDelta$.*

Proof. If $1/P(z)$ had a pole of order higher than the first, then I_Q would have a bounded (and non-null) complement; therefore $\mathscr{T}_Q$ would consist only of the point $t=0$.

Thus, if ζ_0 is the singularity of $1/P(z)$, then $Q(z)$ can be written

$$Q(z)=q_0 \ln (z-\zeta_0)+Q_1(z) \tag{II.13}$$

where $Q_1(z)$ is analytic in $\varDelta$.

Theorems II.6 and II.7 refer to families of transformations generated by $P(z)$ satisfying (CII.3) and (CII.4).

THEOREM II.6. *If* $\omega(z_1;\ t)=z_1,\ z_1\in\Delta$, for $t\neq 2\pi ikq_0,\ k=0,\ \pm1,\ \pm2,\ \cdots$, *then* $z_1=\zeta_0$.

Proof. $Q[\omega(z;\ t)]=Q(z)+t$ for $z\neq\zeta_0$.
Therefore $Q[z_1]=Q[z_1]+t$ if $z\neq\zeta_0$.
Therefore $t=2\pi ikq_0,\ k=0,\ \pm1,\cdots$.

THEOREM II.7. If $\omega(z_1;\ t)=\omega(z_2;\ t),\ t\in\mathscr{T}$, *then* $z_1=z_2$.

Proof. Suppose first that $z_1,\ z_2\neq\zeta_0$. Then $\omega(z_1;\ t)=\omega(z_2;\ t)$ would imply $Q(z_1)=Q(z_2)$ or, by (CII.4), $z_1=z_2$. On the other hand, if, say, $z_1=\zeta_0$, then $\omega(z_1,\ t)=z_1=\omega(z_2;\ t)$ and so $z_2=z_1$ by Theorem II.6.

Thus, conditions (CII.3) and (CII.4) when imposed on the function $P(z)$ imply that the family $\{\omega_t\}$ is a family of schlicht functions.

It is clear that the functions ω_t as well as the set $\mathscr{T}$ are unaltered if the definition of Q is altered by the addition of an arbitrary constant.

It is also easy to see that multiplying Q (that is, multiplying $1/P$) by a constant $c\neq0$ yields essentially the same family of transformations: Let $\mathscr{T}$, $\{\omega_t\}$ correspond to $P(z)$ and let $\mathscr{T}'$, $\{\omega'_{t'}\}$ correspond to $\frac{1}{c}P(z)$. (Here the primes do not, of course, imply differentiation.) Then clearly, $\mathscr{T}'=c\mathscr{T}$. Also, for $t'\in\mathscr{T}'$,

$$cQ[\omega'(z;\ t')]=cQ(z)+t',$$

or

$$Q[\omega'(z;\ t')]=Q(z)+\frac{t'}{c},$$

so that

$$\omega'(z;t')=\omega\left(z;\frac{t'}{c}\right);\quad t'\in\mathscr{T}',\quad \frac{t'}{c}\in\mathscr{T}. \tag{II.14}$$

In other words, there is a one-to-one correspondence between the transformations corresponding to $P(z)$ and those corresponding to $\frac{1}{c}P(z)$; the correspondence is given by (II.14).

Now consider, for $t\in\mathscr{T}\cap I_Q$, the parameter defined by

$$\beta=Q^{-1}(t)\qquad t\in\mathscr{T}\cap I_Q \tag{II.15}$$

Then $\beta\in\Delta$ and (II.12) becomes, writing $\omega[z;\ t(\beta)]$ simply as $\omega(z;\ \beta)$,

$$\omega(z;\ \beta)=Q^{-1}[Q(z)+Q(\beta)],\qquad z,\ \beta\in\Delta. \tag{II.16}$$

Here β is defined on $Q^{-1}[\mathscr{T} \cap I_Q]$.

It is always possible to define Q in such a way[5] that $\mathscr{T} \subset I_Q$ and therefore $\mathscr{T} \cap I_Q = \mathscr{T}$. In such a case, (II.15) and (II.16) hold for all $t \in \mathscr{T}$. For example, in defining Q by (II.9), it is clear that $Q(z_0)=0$ for $z_0 \in \Delta$. Thus, for Q defined as in (II.9) with $z_0 \in \Delta$, we have $\beta = Q^{-1}(t) = \omega(z_0;\ t)$.

It is, however, often possible and more convenient to define Q such that $\mathscr{T}$ is the closure of I_Q. It is also often possible to extend the definition of Q to the boundary of Δ in such a way that the boundary of Δ goes (under Q) into the boundary of I_Q. (An example of this is given by the family of transformations studied in the next section.) In such cases, (II.15) holds for all $t \in \mathscr{T}$ and, in (II.16), β may be a point on the boundary of Δ.

The law of composition of the transformations $T_{\omega_\beta} = T_\beta$ in terms of the parameter β is

$$\text{(II.17)} \qquad \begin{cases} T_{\beta_1} T_{\beta_2} = T_{\beta_3} \\ \beta_3 = \omega(\beta_1;\ \beta_2) \end{cases}$$

This can be shown as follows.

$$\omega[\omega(z;\ t_1);\ t_2] = \omega(z;\ t_1 + t_2),$$

so

$$\begin{aligned} \omega[\omega(z;\ \beta_1);\ \beta_2] &= \omega[z;\ t = Q(\beta_1) + Q(\beta_2)] \\ &= \omega[z;\ \beta = \omega(\beta_1;\ \beta_2)]. \end{aligned}$$

By simply looking at the set I_Q, one is usually able to determine many of the properties of the family $\{T_t\}$. For example, one may determine (a) whether or not such a family exists for the given $P(z)$; (b) what the maximum parameter domain $\mathscr{T}$ is; (c) whether $\{T_t\}$ is a group or a semigroup; (d) which of the functions ω_t transform Δ onto itself and which transform Δ into but not onto itself;

D. Possible applications. The above results provide the basis for obtaining a variety of theorems by rephrasing known results in the theory of transformations in Banach space in terms of transformations in the function spaces H^p of the kind studied above. Three possible categories of results are:

(a) Representations of the transformations T_t in terms of the generator A or the resolvent of A ([2] contains many such formulas).

(b) Application of results in the theory of analytic Banach-space-

[5] The addition of a constant to Q changes I_Q but leaves $\mathscr{T}$ unaltered.

valued functions of a complex variable ([2], [7], [9])

(c) Other theorems concerning properties of semigroups or groups of transformations in Banach space.

III. Two Special Cases

A. The family $\{T_w\}$ defined by $T_w f(z)=f(wz)$, $|w|\leq 1$.

Let

(III.1) $$P(z)=-z$$

and[6]

(III.2) $$Q(z)=\int_1^z \frac{-d\zeta}{\zeta}=-\ln z\,.$$

Then I_Q is the open right half plane: $\Re(z)>0$. $\bar{\mathscr{T}_Q}$ is the closed right half plane: $\Re(z)\geqq 0$. Clearly, $\bar{\mathscr{T}_Q}$ itself satisfies conditions (CII.1) and is therefore the maximum domain $\mathscr{T}$ of the parameter t. We have

(III.3) $$\omega(z;\ t)=ze^{-t} \qquad z\in\varDelta,\ t\in\bar{\mathscr{T}_Q}$$

or, if we let

(III.4) $$w=e^{-t}$$

then, writing $\omega[z;\ t(w)]$ simply as $\omega(z;\ w)$,

(III.5) $$\omega(z;\ w)=wz \qquad z\in\varDelta,\ |w|\leqq 1$$

The corresponding family of transformations $\{T_w\}$ is then given by

(III.6) $$T_w f=g$$

where $$g(z)=f(wz)$$

The generator A is defined for those $f\in H^p$ for which the limit

$$Af=\lim_{w\to 1}\frac{T_w f-f}{1-w} \qquad |w|\leqq 1$$

exists in the H^p norm. Thus,

(III.7) $$Af(z)=-zf'(z) \qquad \text{for } f\in\mathscr{D}(A).$$

For $1\leqq p<\infty$, $\mathscr{D}(A)$ is the set of functions $f\in H^p$ for which $f'(z)$ defines an element of H^p. This follows from Theorem II.2. The crucial point in applying Theorem II.2 is in verifying condition (ii) of

[6] Here $z_0=1$ is not in $\varDelta$, but in this case this is immaterial,

that theorem. This amounts to the following. Let h be in H^p ($1 \leqq p < \infty$), and let $T_w h(z) = h(wz)$ for $|w| \leqq 1$. Then $T_w h \to h$ in the norm of H^p as $w \to 1$ in the closure of Δ. It is not difficult to prove this.

Also, for $1 \leqq p < \infty$, A is a closed operator with domain dense in H^p.

For $p = \infty$, (III.7) still holds, but one cannot verify condition (ii) of Theorem II.2 and it is eassily seen that $\mathscr{D}(A)$ is not dense in H^∞.

B. The family $\{L_\alpha\}$ defined by $L_\alpha f(z) = f\left(\frac{z+\alpha}{1+\alpha z}\right)$, $-1 < \alpha < 1$.

Let

$$P(z) = (1 - z^2) \tag{III.8}$$

and[8]

$$Q(z) = \int_0^z \frac{d\zeta}{1-\zeta^2} = \tanh^{-1} z \tag{III.9}$$

Then I_Q is the strip $|\Im(z)| < \pi/4$. $\mathscr{T}_Q$ is the real axis. Clearly $\mathscr{T}_Q$ satisfies conditions (CII.1) and is therefore the maximum domain $\mathscr{T}$ of the parameter t. We have

$$\omega(z;\ t) = \frac{z + \tanh t}{1 + z \tanh t} \qquad t \in \mathscr{T}_Q,\ z \in \Delta. \tag{III.10}$$

If we let

$$\alpha = \tanh t, \qquad t \in \mathscr{T}_Q \tag{III.11}$$

then, writing $\omega[z; t(\alpha)]$ simply as $\omega(z; \alpha)$,

$$\omega(z; \alpha) = \frac{z+\alpha}{1+\alpha z}, \qquad z \in \Delta,\ -1 < \alpha < 1. \tag{III.12}$$

The family of transformations $\{L_\alpha\}$ is given by

$$L_\alpha f = g \tag{III.13}$$

where

$$g(z) = f\left(\frac{z+\alpha}{1+\alpha z}\right)$$

The norm of L_α is

$$\| L_\alpha \|_{H^p} = \left[\frac{1+|\alpha|}{1-|\alpha|}\right]^{1/p} \tag{III.14}$$

---

[8] The path of integration lying entirely in Δ.

The generator A is defined for those $f \in H^p$ for which the limit

$$Af = \lim_{\alpha \to 0} \frac{L_\alpha f - f}{\alpha}$$

exists in the H^p norm. Hence

(III.15) $$Af(z) = (1-z^2)f'(z) \qquad \text{for } f \in \mathscr{D}(A).$$

For $1 \leqq p < \infty$, $\mathscr{D}(A)$ is the set of functions $f \in P^p$ for which $(1-z^2)f'(z)$ defines an element of H^p; also, A is a closed operator with domain dense in H^p. As with the previous example, these statements do not hold for H^∞.

REFERENCES

1. G. H. Hardy J. E. Littlewood and G. Polya, *Inequalities*, Cambridge, 1934.
2. E. Hille, *Functional analysis and semigroups*, Amer. Math. Soc. Colloquium Publications, **31**, New York, 1948.
3. K. Knopp, *Theory of functions*, Part 1, New York, 1945.
4. J. E. Littlewood, *Lectures on the theory of functions*, Oxford, 1944.
5. F. Riesz, *Über die Randwerte einer analytischen Funktion*, Math. Z., **18** (1929), 87-95.
6. W. Rudin, *Analytic functions of class H_p*, Trans. Amer. Math. Soc., **78** (1955), 46-66.
7. A. E. Taylor, *Banach spaces of functions analytic in the unit circle, II*, Studia Math., **11** (1949), 145-170.
8. ———, *Banach spaces of functions analytic in the unit circle, II*, Studia Math., **12** (1951), 25-50.
9. ———, *New proofs of some theorems of Hardy by Banach space methods*, Math. Mag., **23** (1950), 115-124.

THE RAND CORP.

Pacific Journal of Mathematics

Vol. 7, No. 1 January, 1957