Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 10, No. 1 **September, 1960**

PACIFIC JOURNAL OF MATHEMATICS

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

EXTENSIONS OF BANACH ALGEBRAS

RICHARD ARENS

1. Introduction. We are concerned with propositions of four types (1.1–1.4) about a commutative Banach algebra A and its various commutative Banach algebra extensions B.

1.1 TPr. *If $\{B_i : i \in I\}$ is a family of extensions of A, then there is an extension B of A and topological isomorphism $\{f_i : i \in I\}$ where $f_i(B_i) \subset B$ and $f_i(a) = a$ for $a \in A$.*

Let us call [*normally*] *solvable over A* a system Σ of polynomials over A (or more generally, multiple power series elements) such that there is an extension B of A in which there is a system of elements [whose norms do not exceed 1 and] whose substitution into Σ reduces each member equal to 0.

1.2 Sol. *Let $\{\Sigma_i : i \in I\}$ be a family of solvable systems such that no indeterminate occurs in more than one system. Then $\Sigma = \bigcup \Sigma_i$ is solvable.*

A system $\mathscr{J}$ of ideals is *removable* if in some extension, each ideal J of $\mathscr{J}$ generates the ideal (1).

1.3 RId. *Let $\{J_i : i \in I\}$ be a family of removable ideals. Then it is a removable system.*

An element $c \in A$ is called [*normally*] *subregular* if it has an inverse [of norm $\leqq 1$] in some extension.

1.4 Inv. *Let $\{c_i : i \in I\}$ be a family of subregular elements. Then, in some extension, each c_i has an inverse.*

Our findings on such propositions is that **TPr** is false, and that **Inv** is true if I is finite, but false if a natural norm restriction is brought in. By the *finite* form of 1.1–1.4 we mean that in which I is finite. By the *normal* form we mean the statements obtained if in (1.1) the f_i are required to be isometries, if 'solvable' in (1.2) is replaced by 'normally solvable', and 'subregular' in (1.4) by 'normally subregular'.

This gives four forms of propostions of each type:

(1.5)	*normal*	(*no qualification*)
	finite normal	*finite* .

For each type (1.1–1.4), there are rather obvious implications in (1.5), namely to the right, and downward. (To see this, one need only observe that c is subregular if and only if λc is normally subregular for some $\lambda \in \boldsymbol{C}$, etc.). For each form (1.5) there are implications among the

Received May 4, 1959.

types:

$$
\begin{tikzcd}
\mathbf{TPr} \arrow[r, Rightarrow] & \mathbf{Sol} \arrow[dl, Rightarrow] \arrow[d, Rightarrow] \\
\mathbf{RId} & \mathbf{Inv}\,.
\end{tikzcd}
\tag{1.6}
$$

(For example, the diagonal rests on this observation: if J is removable then $1 - j_1x_1 - \cdots - j_nx_n$ is solvable for some $j_1, \cdots, j_n \in J$; and solving the latter removes the former.) We present our *results on these sixteen conceivable propositions* in this diagrammatic way. In each quadrant of (1.5) imagine a cluster of four symbols as in (1.6). Affix a dagger if the proposition is false, a star if true, and a reference to the crucial theorem. Unsettled cases have a question mark.

	†	†	†	?
	?	†	?	?
(1.7)				
	†	†	† (5.2)	?
	?	† (3.2)	?	* [1, 3.8]

Besides this there is a small positive result (7.1) which is a special case of **RId.**

Further results not included in the scheme (1.7) are as follows.

The *cortex* (class of non-removable maximal ideals) is sometimes greater than the Shilov boundary. This is based an a class of algebras of Shilov, whose theory we have felt obliged to sketch (sec. 4)

For completeness we have considered also the case where A has the *sup*-norm (that is, $||a|| = \sup |\xi(a)|$, ξ ranging over all complex valued homomorphisms of A.) There **Sol** holds (6.1): There is one extension which normally solves all normally solvable systems. Necessary and sufficient conditions for **TPr** are given (5.3)

For some subalgebras A of the l_1-algebra B of a discrete abelian group, B provides inverses of norm 1 for all normally subregular elements (3.5, 3.6).

Section 2 provides more careful definitions of extension, and shows that when **Sol** can be proved, then the solving algebra can always be taken as a quotient-algebra of a power-series algebra.

2. Analytic extension. In order to save space we shall list here properties of a Banach algebra which we shall usually, if not always, require.

(2.11) *It is a Banach space.*

(2.12) *It is a linear algebra over the complex numbers* $\boldsymbol{C}$, *with unit* 1.

(2.13) $||ab|| < ||a|| \, ||b||, \quad ||1|| = 1$.

(2.14) *It is commutative.*

Let A be such an algebra. Let I be a set (to be used as indices). We want to define the commutative Banach algebra $A(X)$ *generated by the family* $X = \{x_i : i \in I\}$ and A. Because a norm has to be defined, we need some details. First we define the free commutative semi-group $S(X)$ with unit generated by X. $S(X)$ is the set of all functions from I to $\{0, 1, 2, \cdots\}$ which vanish at all but finitely many places. The operation is addition. We write it multiplicatively, and use the notation

$$x_{i_1}^{k_1} x_{i_2}^{k_2} \cdots x_{i_n}^{k_n} \tag{2.2}$$

for the element which has the value k_j at i_j $(j = 1, \cdots, n)$ and is 0 otherwise. The function which is 0 everywhere is written as 1. A change in the order of the factors in (2.2) does not produce a different element, of course. Now $A(X)$ is the set of functions f from $S(X)$ to A such that

$$||f|| = S(X) \, ||f(\quad)|| < \infty \ . \tag{2.3}$$

We may let

$$a x_{i_1}^{k_1} \cdots x_{i_n}^{k_n}$$

stand for the element of $A(X)$ which has the value $a \in A$ at the element 2.2 of $S(X)$, and the value 0 elsewhere. We write a for $a1$ $(1 \in S(X))$. Then each f has the form

$$f = \sum_{j=1}^{\infty} a_j \xi_j \tag{2.4}$$

where each ξ_j has the form (2.2), and

$$||f|| = \sum_{j=1}^{\infty} ||a_j|| \ . \tag{2.5}$$

Clearly the element of $A(X)$ can be added and multiplied, being functions with values in A. The algebra $A(X)$ is easily seen to satisfy the conditions 2.11–2.14. It clearly "contains" the algebra $A[X]$ of *polynomials* in the indeterminates with coefficients in A.

If I_0 is a subset of I, and X_0 is the corresponding system of indeterminates, then $A(X_0)$ can be canonically embedded in $A(X)$. The algebra $A(X)$ is not very interesting in itself. For example, its space of multiplicative linear functionals of the form

$$\Delta(A) \times D^I$$

where $\Delta(A)$ is the corresponding space for A (compare [2, 4.1])

A *Banach algebra extension* of A is an isometric isomorphism of A onto a subalgebra A_1 of a Banach algebra B where the unit of A_1 is that of B. When possible we abbreviate this by saying that B is an extension of A, and pretend that $A \subset B$.

A system $\Sigma = \{\gamma_k : k \in K\}$ of elements of $A(X)$ is called *normally solvable over* A if there is a Banach algebra extension B satisfying 2.11–2.14 and if for each $i \in I$ there is an element $b_i \in B$ with $||b_i|| \leqq 1$ such that if b_i be substituted for x_i in γ_k, then 0 results for each k. (If X contains any x_i not appearing in any γ_k, the corresponding b_i need not be expressly exhibited. It may be chosen as $0 \in B$.)

For an example, see (2.9) below.

A natural attempt to "solve Σ normally" is to form the closed ideal J generated by Σ in $A(X)$, and form

$$A_\Sigma = A(X) \bmod J \tag{2.6}$$

The norm in A_Σ is the canonical one for residue-class algebras [5, p. 14]. The main theorem of this section (2.8) is that this construction is always successful when Σ *is* normally solvable. (Obviously, if the construction is successful, Σ must be normally solvable.)

The only possible obstacle to this approach is that, whereas $A(X)$ is a Banach algebra extension of A, A_Σ might not be, because norms of elements in $A \subset A(X)$ might be diminished when A_Σ is formed (compare [1, pp. 537–8; 2, p. 204.])

2.7 PROPOSITION. *A_Σ is a Banach algebra extension of A and is normally solvable if, for each finite collection of polynomials $p_i, \cdots, p_n \in A(X)$ and indices $j_1, \cdots, j_n$, and each element $a \in A$, the inequality*

$$||a|| \leqq ||a - p_1\gamma_{j_1} - \cdots - p_n\gamma_{j_n}|| \tag{2.71}$$

holds.

The norm on the right is the one mentioned in (2.5). The proof of (2.7) may be omitted. It suffices to deal only with polynomials p_k in (2.7) because there are dense in $A(X)$.

It would be a waste of effort to have X contain any elements not involved in Σ, in essaying to verify (2.71).

The converse (2.7) is valid and we thus arrive at the following.

2.8 THEOREM. *Σ is normally solvable if and only if* (2.71) *holds for all $a, p_1, \cdots, p_n$ as specified in* (2.7).

To see the "only if", suppose B normally solves the system Σ, containing elements $\{b_i\}_{i \in I_0}$ where I_0 are the indices of the elements actually appearing in Σ, such that $\gamma_j(b) = 0$ for all j. Then we can set up a homomorphism

$$h : A(X) \to B$$

wherein $h(x_i) = b_i$, $i \in I_0$, $h(x_i) = 0$, $i \notin I_0$, and $h(a) = a$ for $a \in A$ (regarded as a subalgebra of $A(X)$ as well as of B.) Clearly $||h(f)|| \leqq ||f||$. The ideal J is contained in the kernel of h, because $h(\gamma_j) = 0$ for all $\gamma_j \in \Sigma$. We thus arrive at a homomorphism h^* of bound at most 1 [5, 7D] of A_Σ into B. Therefore the natural image $a + J$ of an element a from A has a norm (in A_Σ) not less than the norm of its image in B. The latter is a itself, so that $||a|| \leqq ||a + J||$. This implies (2.71), so that (2.8) is shown.

The necessary and sufficient condition given by (2.8) can in special cases be replaced by a simpler one.

2.9 THEOREM. *Let c, $d \in A$, and let n be a positive integer. Then*

$$c = dx^n \qquad\qquad ||x|| \leqq 1$$

can be solved in some extension algebra if and only if, for every $a \in A$,

$$||ca|| \leqq ||da|| .$$

The proof, which resembles [1, sec. 3], is simple and may be omitted.

An illustration of the two-way utilization of (2.8) is the following.

2.91 THEOREM. *Let $c \in A$, and let $\mu > 0$. Then*

$$c = e^z \qquad\qquad ||z|| \leqq \mu$$

can be solved in some extension algebra if and only if for each $\nu > \mu$, and positive integer N, there is an extension in which for some n

$$c = \left(1 + \frac{y}{n}\right)^n \qquad\qquad n \geqq N,\ ||y|| < \nu ,$$

Proof. It is evidently a matter of showing that $c - e^{\mu x}$ is normally solvable precisely when $c - \{1 + (\mu x\delta/n)\}^n$ is normally solvable for infinitely many n, whenever $\delta > 1$. The former can be solved in the latter circumstances because the class of normally solvable *elements* of $A(x)$ is closed, by (2.8). Conversely, if the latter is normally solved with x in some extension algebra B, and $n > \mu\delta$, then (letting $\lambda = \mu/n$) take $y = \lambda^{-1} \log(1 + \lambda x\delta)$, and obtain $e^{\mu y} = c$, with $||y|| \leqq -\lambda^{-1} \log(1 - \lambda\delta)$.

3. The union of normally solvable systems. In (1.2) we included the condition that the solvable families whose union is to be formed involve distinct collections of indeterminates. This is natural, because while each of the one-member families

$$\{1 - x\}, \ \{1 - 2x\} \tag{3.1}$$

is normally solvable over any algebra, the union is never solvable. As indicated in § 1, we do not know if this condition is enough to make even **Sol** (finite) hold, but we shall now show that **Sol** (normal, finite) is not generally true. Our example has the special merit of dealing with systems whose solution consists in constructing inverses, so that it destroys **Inv** (normal, finite) as well, as promised by (1.7).

3.2 THEOREM. *There exists a Banach algebra A* (2.11–2.14) *with elements p, q over which*

$$1 - qx \tag{3.21}$$

and

$$1 - py \tag{3.22}$$

are normally solvable, but

$$\{1 - qx, \ 1 - py\} \tag{3.23}$$

is not normally solvable.

Proof. The algebra A is isomorphic as a topological algebra, to the algebra of absolutely convergent power series on the unit disc. In order to reserve letters such as z for possible use as indeterminates, we use p for the "complex variable". Select a real number $\alpha, \alpha > 1$. For $a \in A$, say,

$$a = \lambda_0 + \lambda_1 p + \lambda_2 p^2 + \cdots \tag{3.24}$$

we define

$$||a|| = |\lambda_0| + \alpha(|\lambda_1| + |\lambda_2| + \cdots) . \tag{3.25}$$

The operations are the usual addition and multiplication of series. (2.11–2.14) are easily verified. (In fact, this algebra is a simple specimen of a ring of Shilov's type $K_{<\alpha n>}$ (§ 4) where $\{\alpha_n\} = \{1, \alpha, \alpha, \cdots\}$).

It is clear that

$$||pa|| \geqq ||a|| \qquad (a \in A) . \tag{3.26}$$

Moreover, for $0 \leqq \delta < 1$ we also have

$$||qa|| \geqq ||a|| \tag{3.27}$$

where

$$q = (1 - \delta)^{-1}(1 - \delta p)p . \tag{3.28}$$

To see (3.27), consider that

$$\|(1-\delta p)pa\| \geqq \|pa\| - \delta\|p^2a\| .$$

Now

$$\|p^2a\| = \|pa\|, \text{ so } \|(1-\delta p)pa\| \geqq \|pa\|(1-\delta) \geqq (1-\delta)\|a\| ,$$

by (3.26).

By [1, 3.5] each of the one-element systems

$$\{\gamma_1\} = \{1 - qx\} , \tag{3.29}$$

$$\{\gamma_2\} = \{1 - py\} , \tag{3.30}$$

is normally solvable (and, the combined system is solvable [1, 3.8]).

We submit that the following is an identity in x, y:

$$\delta(1-\delta)(1-\delta p)^{-1} = x - (1-\delta)y + \gamma_1[\delta(1-\delta)(1-\delta p)^{-1} + (1-\delta)y] + \gamma_2[-x(1-\delta p)] . \tag{3.31}$$

This is readily verified by substituting (3.29), (3.30), and (3.28) into (3.31).

Let us now suppose $\{\gamma_1, \gamma_2\}$ is normally solvable. Let $a = \delta(1-\delta)(1-\delta p)^{-1}$. Then, from (3.31)

$$a - \gamma_1[\cdots] - \gamma_2[\cdots] = x - (1-\delta)y .$$

Comparing this with (2.71), we see that

$$\|a\| \leqq \|x - (1-\delta)y\| = 1 + (1-\delta) ,$$

where we have used (2.5) for the norm in $A(x, y)$. Now $a = \delta(1-\delta)(1 + \delta p + \delta^2 p^2 + \cdots)$ and the norm of this is given by (3.25):

$$\|a\| = \delta(1-\delta)(1 + \delta\alpha + \delta^2\alpha + \cdots) = \delta - \delta^2 + \delta^2\alpha .$$

It thus appears that

$$\alpha \leqq 1 + 2(1-\delta)\delta^{-2} . \tag{3.32}$$

Thus the desired counterexample is possible. In fact, if $\alpha > 1$ then some δ will make (3.32) false.

The next proposition shows how plentiful these counterexamples really are.

3.4 THEOREM. *Let A be any algebra satisfying* (2.11–2.14), *containing an element c which is not regular but is no topological zero-divisor. Then A contains p, q and can be given an equivalent norm such that $\{1-qx\}$, $\{1-py\}$ are each normally solvable, but $\{1-qx, 1-py\}$ is not.*

Proof. Select a complex-valued homomorphism ξ of A such that $\xi(c) = 0$. We may assume that $||ac|| \geqq ||a||$ for all $a \in A$. Select any δ such that $0 < \delta < 1$. *I* now present the p and q: $p = 3c$, $p = 3\beta c$ $(1 - \delta c)$ where $\beta = ||(1 - \delta c)^{-1}||$; and the new norm

$$|a| = |\xi(a)| + \alpha||a - \xi(a)|| \qquad (\alpha > 1)\,. \tag{3.41}$$

Here α is a parameter to be fixed later. It is not hard to see that (3.41) satisfies (2.11–2.14). Furthermore,

$$||a|| < |a| < 3\alpha\, ||a||\,. \tag{3.42}$$

Since $\xi(c) = 0$ we have $|3ca| = \alpha||3ca|| \geqq 3\alpha||a|| > |a|$. It follows that $1 - 3cy$ is normally solvable. It is similarly established that $1 - 3\beta c$ $(1 - \delta c)x$ is normally solvable.

Now suppose some extension B of A (A with the $|\cdot\cdot|$ norm, be it understood) had elements x, y of norm not exceeding 1 such that

Then
$$3\beta c(1 - \delta c)x = 1\,, \qquad 3cy = 1\,.$$
$$\beta(1 - \delta c)x = y\,, \qquad \beta x - y = \beta\delta c x\,.$$

Now $3\beta c x = (1 - \delta c)^{-1}$ so we have

$$\delta(1 - \delta c)^{-1} = 3\beta x - 3y$$

Whence

$$\delta|(1 - c)^{-1}| \leqq 3(\beta + 1)\,. \tag{3.43}$$

But

$$|(1 - \delta c)^{-1}| = 1 + \alpha||(1 - c)^{-1} - 1||$$

where the coefficient of α is not 0 because $\delta \neq 0$. Hence α can be chosen so that (3.43) is impossible.

The Banach algebra A used in (3.2) cannot yield a counterexample if the parameter α is taken as 1. This follows from the following.

3.5 THEOREM. *Let B satisfy* (2.11–214) *and let A be a subalgebra with unit. Let $\Delta(B)$ be the space of complex-valued homomorphisms of B, and $\Delta(A)$, the corresponding set for A. Suppose that every $\xi \in \Delta(B)$ when restricted to A falls into the Shilov boundary* [4, 5] $\partial_A\Delta(A)$. *Suppose moreover that there is a collection U of elements in B such that*

$$u \in U \text{ and } a \in A \text{ implies } ||ua|| = ||a|| \tag{3.51}$$

and

(3.52) $\{ua : u \in U,\ a \in A\}$ *is dense in* B.

Then each element c *of* A *which has an inverse of norm not exceeding* 1 *in some extension of* A, *has such an inverse in* B.

Proof. If $c \in A$ has an inverse in some superalgebra, then it cannot vanish on $\partial_A \Delta(A)$ as each of these homeomorphisms can be extended to any superalgebra of A. Thus c has an inverse b in B, and what remains to be proved is that $||b|| \leqq 1$. This will result from the fact that necessarily, $||ca|| \geqq ||a||$ for all $a \in A$ [1, 3.5].

By (3.52), there exist $\{u_n\}$, $\{a_n\}$ such that $u_n a_n \to b$. Therefore $||cu_n a_n|| \to ||cb|| = 1$. However (by 3.51)

$$||cu_n a_n|| = ||ca_n|| \geqq ||a_n|| = ||u_n a_n|| \to ||b|| .$$

This completes the proof of (3.5).

From this general proposition we now consider another which shows that for the A of (3.2) with $\alpha = 1$ there is a B to which A bears the relation described in (3.5). In fact, $B = L^1(\boldsymbol{Z})$ where $\boldsymbol{Z}$ is the discrete group of integers, and A can be identified with those elements of B which are supported by the semi-group $\boldsymbol{Z}_+$ of non-negative integers. This pair is discussed in [5. 23C and 24E].

3.6 THEOREM. *Let* G *be a discrete abelian group and* S *a subsemigroup containing* $e \in G$. *Let* $B = L^1(G)$, *and* A *be the subalgebra of* B *consisting of those functions whose support lies in* S. *Let* U *be the group* G *as naturally imbedded in* $B : x \to \delta_x$, $\delta_x(y) = \delta(y - x)$, *where* $\delta(x) = 0$ *or* 1 *according to whether* $x \neq e$, *or* $x = e$. *Then* U, A, B *satisfy the conditions of* (3.5).

The specific properties of U are obvious, and the relation of $\Delta(B)$ and $\partial_A \Delta(A)$ is easily established, either by analogy with [5, 24E], or by [7, 4.6].

4. The cortex. Let A satisfy (2.11–2.14) as always. By $\Delta(A)$ we mean the space of complex linear homomorphisms of the algebra A onto the complex numbers $\boldsymbol{C}$, with the weak topology. By the *cortex* $\Gamma(A)$ of $\Delta(A)$ (or, more briefly, the *cortex* of A) we mean the set of those homomorphisms which can be extended to every extension B of A.

Now those $\xi \in \Delta(A)$ which can be extended to B form a compact set E_B which is the continuous image (under the restriction map) of $\Delta(B)$, and the cortex is evidently the intersection of these E_B. Moreover, each E_B contains the Shilov boundary $\partial_A \Delta(A)$ [4, 5] which is never void. Thus we have the following.

4.1 THEOREM. *The cortex* $\Gamma(A)$ *is compact, and contains the Shilov* *boundary.*

When A has the *sup* norm, i.e, when $||a|| = \sup \{|\xi(a)| : \xi \in \Delta(A)\}$, then $\Gamma(A) = \partial_A \Delta(A)$ since the extension $B = \mathscr{C}(\partial_A(\Delta(A))$ admits only homomorphisms which are on the Shilov boundary. There are algebras in which the norm is not equivalent to the *sup-norm* in which $\Gamma(A)$ and $\partial_A(\Delta(A))$ coincide, for example the A of (3.6) above.

However, the work of Shilov [6] makes it possible to exhibit algebras with one generator in which $\Gamma(A) \neq \partial_A(\Delta(A))$. Because of the rarity of this paper in these parts it may be permissible to sketch proofs of some of Shilov's theorems.

Let $\{\alpha_n\} = \{\alpha_0, \alpha_1, \cdots\}$ be a sequence of real numbers where, for $m, n > 0$,

$$\alpha_0 = 1 \leqq \alpha_{m+n} \leqq \alpha_m \cdot \alpha_n \; . \tag{4.21}$$

Let $K(\boldsymbol{\alpha})$ be the space of these formal power series (which notation makes the algebraic operations more evident)

$$f(z) = a_0 + a_1 z + a_2 z^2 + \cdots \tag{4.22}$$

for which

$$||f|| = \Sigma \, |a_i| \alpha_i \; \infty \; . \tag{4.23}$$

$K(\boldsymbol{\alpha})$ satisfies (2.11–2.14). It follows from (4.21) that $r = \lim \; (\alpha_n)^{1/n}$ exists. Thus the spectrum of z (see 4.22) is the disc $\{|\lambda| \leqq r\}$, and this is a homeomorphic image of $\Delta(K(\boldsymbol{\alpha}))$ under the map $\xi \to \xi(z)$.

We now consider the possibility of enlarging the algebra by defining α_n also for $n > 0$, and norming formal Laurent series as in (4.23). It is easy to see that if such α_n $(n < 0)$ can be defined with (4.21) holding, then for each $r = 1, 2, \cdots$,

$$\beta_r = \sup \; \alpha_n (\alpha_{n+r})^{-1} : n = 1, \; 2, \; \cdots \tag{4.24}$$

must be finite. Moreover

4.25 PROPOSITION (Shilov). *Letting $\alpha_{-n} = \beta_n$ for $n = 1, 2, \cdots$ provides an extension of the $K(\boldsymbol{\alpha})$ when β_1 is finite.*

The proof lies in verifying (4.21), and then observing that $\beta_r < (\beta_1)^r$. Let us call this extension algebra $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$. The spectral radius of $1/z$ is evidently

$$s = \lim \; (\beta_n)^{1/n} \tag{4.26}$$

and $\Delta(K(\boldsymbol{\alpha}, \boldsymbol{\beta}))$ is homeomorphic to the spectrum of z, which is

$$\{\lambda : s^{-1} \leqq |\lambda| \leqq r\} \; . \tag{4.27}$$

When $\beta_1 = \infty$ we set $s^{-1} = 0$.

4.3 THEOREM (Shilov). *The element* $\lambda - z$ *is a generalized zero-divisor in* $K(\boldsymbol{\alpha})$ *if and only if* $s^{-1} < |\lambda| < r$.

Now $\lambda - z$ is *not* a generalized (or topological) zero-divisor if

$$\inf \{||(\lambda - z)f(z)|| : ||f(z)|| = 1\} \tag{4.31}$$

is positive. The evaluation of (4.31) is facilitated by

4.32 LEMMA. *Let* T *be a convex compact subset of a topological (real) linear space* L, *and let* $\varphi_0, \cdots, \varphi$ *be* $N_N + 1$ *real valued linear functionals on* L. *Let* $S(x) = |\varphi_0(x)| + \cdots + |\varphi_N(x)|$, $\mu = \min \{S(x) : x \in T\}$. *Then there exist* $i_1, i_2, \cdots, i_n$ *and a point* $x_1 \in T$ *such that*

$$S(x_1) = \mu \tag{4.33}$$

and

(4.34) x_1 *is an extreme point of* T *relative to* $Z(i_1, \cdots, i_n)$, *where the latter is the linear subspace defined by* $\varphi_{i_1} = \varphi_{i_2} = \cdots = \varphi_{i_n} = 0$.

Proof. Selection an x_0 such that $S(x_0) = \mu$ and such that the number of φ_i that vanish at x_0 is a maximum. Let

$$\varphi(x) = \textstyle\sum'(\operatorname{sgn} \varphi_i(x_0))\, \varphi_i(x) \tag{4.35}$$

where $'$ is to remind the reader that sgn $0 = 0$. Let $\{i_1, \cdots, i_n\}$ be those indices for which $\varphi_i(x_0) = 0$, and define Z as in (4.34). By the maximum-property of n, each sgn φ_i is constant on $T \cap Z$. Therefore $\varphi(x) = S(x)$ on $T \cap Z$. Now φ is linear and so there is an extreme point x_1 of the convex set $T \cap Z$ such that $\varphi(x_1) = \mu$.

Having established the Lemma we apply it as follows to the space L_N of polynomials of degree $< N$. Let T_N be the collection of those members of L_N whose $K(\boldsymbol{\alpha})$ norm is $\leqq 1$. For $f \in L_N$ let $\varphi_i(f)$ be α_i times the ith coefficient of $(\lambda - y) f(z)$. It is clear that the inf in (4.31) has the value

$$\lim_{N \to \infty} \inf_{f \in T_N} S_N(f)\,.$$

The set $Z(i_1, \cdots, i_n)$ of those $f \in L_N$ such that $\varphi_{i_1}(f) = \cdots = \varphi_{i_n}(f) = 0$ is clearly the set of those f such that $\lambda a_i = a_{i-1}$ for $i = i_1, i_2, \cdots, i_n$ (here a_{-1} and a_N are interpreted as 0.) Let the indices be arranged so that $i_1 < i_2 < \cdots < i_n$. This sequence decomposes into maximal blocks without gaps. If $\{m + 1, \cdots, m + p\} = \sigma$ is such a block then

$$f_\sigma = z^m(\lambda^p + \lambda^{p-1}z + \cdots + \lambda z^{p-1} z^p) \tag{4.36}$$

lies in $Z(i_1, \cdots, i_n)$. If $\sigma = \{0, \cdots\}$ let $f_\sigma = 0$. If $0 \neq i_1$ then $1 \in Z(i_1, \cdots, i_n)$. If i, $i + 1$ do not occur $(0 < i,\ i + 1 \leqq \mathrm{N})$ then

$z^i \in Z(i_1, \cdots, i_n)$. *Conversely,* every function in $Z(i_1, \cdots, i_n)$ is uniquely expressible as a linear combination of these functions just associated with $\{i_1, \cdots, i_n\}$. Considering how the norm is taken, the extreme points of $Z(i_1, \cdots, i_n) \cap T_N$ are obviously just functions of the type

$$cz^m(\lambda^p - z^p)(\lambda - z)^{-1} \tag{4.37}$$

where c is any number that makes the norm 1. (Here we allow $p = 1$ and also $m = 0$ to take care of those monomials mentioned above which are not due to gap-less blocks.)

4.4 LEMMA. *The inf* (4.31) *can be evaluated by letting f run through the system* (4.37).

Shilov does not seem to examine the inf (4.31) to the extent we do here, but the functions (4.37) occur in his considerations.

We now pass to a proof of (4.3). First of all, if $|\lambda| < s^{-1}$ then z and $z - \lambda$ have an inverse in $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$, whence $z - \lambda$ cannot be a topological zero-divisor. If $|\lambda| > r$ then $z - \lambda$ has an inverse in $K(\boldsymbol{\alpha})$.

Now suppose $z - \lambda$ is not a topological zero-divisor. We wish to show that $|\lambda| < s^{-1}$ if $|\lambda| \leqq r$. We confine ourselves to $\lambda \geqq 0$, and assume $\lambda \leqq r$. λ cannot be r or s^{-1} for these yield topological zero-divisors by Shilov's earlier theorem [6]. If $\lambda < r$ then for N, some we have $\lambda^p < \alpha_p$ for $p > N$. From (4.31) we obtain an $M < \infty$ such that $||(x - z)f(z)||\, M > ||f(z)||$ for all f. Inspired by (4.4) we select $f(z) = \lambda^p + \lambda^{p-1}z + \cdots + z^p$ and obtain

$$\lambda^p + \lambda^{p-1}\alpha_1 + \cdots + \alpha_p < M(\lambda^{p+1} + \alpha_{p+1}) < 2M\alpha_{p+1} .$$

For $q > N$ and $p = k + q - 1$ we obtain

$$\lambda^k\alpha_q < 2M\lambda\alpha_{k+q} ,$$

whence (by 4.24) $\lambda^k\beta_k < 1$, and (by 4.26) $\lambda s < 1$.

Using the notation of (4.3), we deduce the following [see **6**. Th. 7].

4.4 COROLLARY. *For $K(\boldsymbol{\alpha})$ the Shilov boundary is $\{\lambda : |\lambda| = r\}$, and the cortex is $\{\lambda : s^{-1} \leqq |\lambda| \leqq r\}$. For $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$ the Shilov boundary is $\{\lambda : s^{-1} = |\lambda|$ or $|\lambda| = r\}$ while the cortex is the same as for $K(\boldsymbol{\alpha})$. Thus if $s^{-1} < r$, then in each case the cortex is greater than the Shilov boundary.*

Shilov remarks that if $t < \rho$ then examples can be constructed such that $s^{-1} = t$, $r = \rho$. He gives no example, so we may just give one producing the interesting case $s^{-1} = 0$, $r = 1$. We have, of course, to define $\{\alpha_0, \alpha_1, \cdots\}$. Let $\alpha_m = \exp(\lambda(m))$ where $\lambda(m)$ is defined as follows. Set

$$\begin{aligned}\lambda_n(m) &= n^{-2}m \quad &\text{when} \quad & m \leqq n^3 \\ &= 0 \quad &\text{when} \quad & m \geqq n^3 .\end{aligned}$$

Let $\lambda(m) = \lambda_1(m) + \lambda_2(m) + \cdots$. It is not hard to see that $\lambda(m) - \lambda(m+1) > n - \pi^2/6$. It follows that $\beta_1 = +\infty$ (see 4.24) and because $\beta_r \alpha_{r-1} \geqq \beta_1$, that $\beta_r = \infty$ whence $s^{-1} = 0$ (see 4.26).

We close this section by comparing the extension $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$ which Shilov has provided for $K(\boldsymbol{\alpha})$ when $\beta_1 < \infty$ with that provided by [1, 3.1]

The least norm of z^{-1} in any extension of $K(\boldsymbol{\alpha})$ is easily seen to be β_1, for this least norm is by [1. 3.1] the reciprocal of the *inf* (4.31). Thus it is reasonable to compare Shilov's extension $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$ with that provided by § 2, for normally solving $1 - \beta_1 zx$.

4.5 THEOREM. *Let $\beta_1 < \infty$. In $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$ the norm of z^{-n} is β_n. In the "canonical" extension of $K(\boldsymbol{\alpha})$ normally solving $1 - \beta_1 zx$, the norm* of z^{-n} *is* $(\beta_1)^n$.

Proof. The statement about $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is obvious.

Denote β_1 by β. In the extension (2.6), the norm of x^n is inf $S[x^n - (1 - xy)g(x, y)]$ extended over all polynomials g, S meaning sum of all norms of coefficients of powers of x, where$\|y^n\| = \|\beta^n z^n\| = \beta^n \alpha_n$. Let $p + q$ be a maximum for the term $\gamma x^p y^q$ in some particular polynomial $g(x, y)$; and suppose $p + q > n - 1$. Then $x^n - (1 - xy)g(x, y)$ has $\gamma x^{p+1} y^{q+1}$ as its highest degree term. If we modify $g(x, y)$ by omitting the term $\gamma x^p y^q$, then the S-contribution from $x^p y^q$-terms in $x^n - (1 - xy)g(x, y)$ might become $|\gamma|\ \beta^p \alpha_q$ larger, but the contribution $|\gamma|\ \beta^{p+1} \alpha_{q+1}$ will disappear. Hence this modification changes the $S[\cdots]$ by at most a negative increment. Therefore, we may confine ourselves to g's whose terms have $p + q < n - 1$. Of these, $g = 0$ gives the minimum possible value for $S[\cdots]$, and it is β^n.

This theorem (4.5) shows that $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$ gives the smallest possible inverse not only to z, but to all its powers (whereas the "canonical" one may not do justice, so to speak, to the inverses of $z^2, z^3, \cdots$). This being so, one wonders if $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$ might not provide the best (i.e., least-in-norm) inverses to $z - \lambda$ for $|\lambda| < s^{-1}$. Our result (4.4), whose full force has not really been employed above, shows that this is not generally true.

4.6 THEOREM. *Consider $K(\boldsymbol{\alpha})$ with $\{\alpha_0, \alpha_1, \alpha_2, \cdots\} = \{1, 2, 1, 1, \cdots\}$. Then $s^{-1} = r = 1$. For $|\lambda_1| = \lambda < 1$ the norm of $(z - \lambda_1)^{-1}$ in $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is $2(1 - \lambda)^{-1}$. For each λ_1 such that $|\lambda_1| = \lambda < 1$ there is an extension of $K(\boldsymbol{\alpha})$ in which the norm of $(z - \lambda_1)^{-1}$ is*

(4.61) $\quad (2 + \lambda)(1 + \lambda^2)^{-1}$ *if* $\lambda \leqq 1/2$, *and* $(1 - \lambda)^{-1}$ *if* $\lambda \geqq 1/2$.

Proof. The details are tedious and should be accepted or verified by the reader. The system (4.24) comes out $\{\beta_1, \beta_2, \cdots\} = \{2, 2, 2, \cdots\}$ whence $\|(z-\lambda_1)^{-1}\| = 2(1-\lambda)^{-1}$ in $K(\boldsymbol{\alpha}, \boldsymbol{\beta})$. Taking however a fixed λ (one might as well suppose $\lambda = \lambda_1 \geqq 0$), the best that can be done by the canonical (and thus by any) method is a norm for $(z-\lambda)^{-1}$ equal to the reciprocal of the *inf* (4.31). A page of calculation, based on (4.4), yields the result stated. Curiously, the formula (4.61) gives a function which is not monotonely increasing, but has a minimum at $\lambda = 1/2$.

5. The tensor-product problem TPr. The proposition **TPr** (*norm, finite*) is false in general because it would conflict with (3.2). However, there is a simpler argument, which also destroys **TPr** (*top, finite*). It rests upon the following.

5.1 THEOREM. *Let B satisfy* (2.11–2.14), *and let $\{B_i : i \in I\}$ be a family of closed subalgebras; and let A be a closed subalgebra with the unit of B included in each of the B_i. Let Δ_i be the space of complex-valued homomorphisms of B_i, and let Γ_i be the cortex, $i \in I$. Let T_i^* be the restriction map*

$$T_i^* : \Delta_i \to \Delta\ ;\ T_i^*(\zeta) = \zeta|_A\ . \tag{5.11}$$

Then for each $j \in I$ we must have

$$T_j^*(\Gamma_j) \subset \bigcap T_i^*(\Delta_i)\ . \tag{5.12}$$

Proof. Let ζ belong to the right hand side of (5.12). Then ζ extends to B, and thus ζ is for each i a restriction of some $\zeta_i \in \Delta_i$, to A.

For a pair of real numbers $r, \rho\,(0 \leqq \rho \leqq r)$ let $A(r, \rho)$ be the algebra of functions continuous for $\rho \leqq |\lambda| \leqq r$ and holomorphic for $\rho < |\lambda| < r$, with the *sup-norm.* The cortex is the set $\{|\lambda| = \rho\} \cup \{|\lambda| = r\}$ (deleting the former when $\rho = 0$). Let $A = A(0, 1)$, $B_1 = A(1/3, 1)$, $B_2 = A(2/3, 1)$. Then $A \subset B_1 \subset B_2$ and in some sense B_2 is the desired tensor product of B_1 and B_2 (over the ring A) but not in the sense **TPr** for the injection of $B_1 \to B_2$ is not *bi*-continuous. In fact, because $T^*(1/3) \notin T^*(\Delta_2)$, we have:

5.2 THEOREM. **TPr** *does not hold for* $A\{B_1\, B_2\}$.

Returing briefly to (5.12), we show that it is *sufficient* for **TPr** when all norms are *sup*-norms.

5.3 THEOREM. *If $\{B_i : i \in I\}$ is a family of Banach algebra extensions of A* (*all satisfying* 2.11–2.14) *and each B_i has the sup-morm then an algebra B as in* **TPr** (1.1) *exists such that the mappings f_i are isometries, provided condition* (5.12) *holds for each j.*

We sketch a demonstration with close reference to [3, Appendice I].

One forms $B_0 = \bigotimes_{(I)} B_i$. Let ζ be any element of $\times_{i \in I} \Delta_i$ such that $\zeta_i|_A$ is independent of i. Then $\bigotimes_{i \in I} \zeta_i = \zeta^*$ is a C–homomorphism of B_0. We define the norm of an element $b \in B_0$ as $||b|| = \sup |\zeta^*(b)|$, and complete B_0 in that norm. Now let $b_j \in B_j$ have norm 1. Then $|\zeta_j(b_j)| = 1$ 1 for some $\zeta_j \in \Gamma_j$. By (5.12), this ζ_j is part of a collection $\{\zeta_i\}$ of the type used in forming the homomorphisms ζ^*, and surely $|\zeta^*(f_j(b_j))| = 1$. Thus $||f_j(b_j)|| \geqq ||b_j||$. On the other hand, if $|\zeta^*(f_j(b_j)| > 1$ then $|\zeta_i(b_j)| > 1$ for some $\zeta_i \in \Delta_i$ which cannot be if $||b_i|| = 1$.

Thus (5.3) is proved.

6. The Sol problem for sup-normed algebras. The non-equivalence of **Sol** and **TPr** is brought out by the fact that **Sol** (norm, arb) is true when A has the *sup*-norm. For then, the algebra $M(\partial_A \Delta(A))$ of bounded functions on the Shilov boundary solves normally all normally solvable systems.

6.1 THEOREM. *Let B, satisfying* (2.11–2.14), *normally solve a system Σ over a subalgebra A having the* sup-*norm. Then $M(\partial_A \Delta(A))$ also normally solves Σ.*

Proof. Well-order the class $\Delta(B)$ of C-homomorphisms of B. For each $\zeta \in \partial_A(\Delta(A))$, let ζ' be the first element of $\Delta(B)$ which is an extension of ζ. Define $T(b)(\zeta) = \zeta'(b)$. This mapping is isometric on A, and of bound 1 on B. These two properties insure that the homomorphism T preserves "normal solution". Thus (6.1) is proved.

It is worth noting that $M(\partial_A \Delta(A))$ not only solves all systems solvable over A but also all systems solvable over itself.

7. A fragmentary result on joint removal of ideals. Let A satisfy (2.11–2.14), and let Δ be its space of C-homomorphisms. Let J be an ideal of A. The hull H of J is $\{\zeta : \zeta \in \Delta, \zeta = 0 \text{ on } J\}$. This is compact.

7.1 THEOREM. *Let $\{J_i : i \in I\}$ be a family of removable ideals* (*see* § 1) *and let J be a removable ideals. Let each J_i, and J be a principal ideal. Let the hulls H_i converge to the hull H of J in this sense: every neighborhood W of H contains all but finitely many of the H_i. Then the system $\{J\} \cup \{J_i : i \in I\}$ is removable.*

Proof. If $J = cA$ then c *is* subregular (see § 1); and we can select c so that $||ca|| \geqq ||a||$ for all $a \in A$. Elements c_i can be selected so that each $J_i = c_i A$ with $||c_i a|| \geqq ||a||$ for all $a \in A$. Let $W = \{\zeta : |\zeta(c)| < 1/2\}$. Let $H_{i_1}, \cdots, H_{i_n}$ include all those hulls which are not included in W. Let $d = c_{i_1} \cdots c_{i_n}$. One can find an integer p such that $|\zeta(dc^p)| < 1$ for all $H \cup \bigcup H_i$. In some extension algebra B, dc^p has an inverse b, $||b|| \leqq 1$. If the ideals J, J_i $(i \in I)$ are not all removed by B then B has a C-homomorphism ζ_0 which is an extension of some ζ in $H \cup \bigcup H_i$. Now $\zeta_0(bdc^p) = 1$, $|\zeta_0(b)| \leqq 1$, but $|(dc^p)| < 1$. This is a contradiction.

A question. Suppose $c_1, \cdots, c_n$ are elements of A which generate a removable ideal. Then there are numbers $\mu_1, \cdots, \mu_n$ such that $||a|| \leqq ||c_1 a|| \mu_1 + \cdots + ||c_n a|| \mu_n$. (Indeed, if $1 = c_1 x_1 + \cdots + c_n x_n$ in some superalgebra, then one can take $\mu_k = ||x_k||$.) Is the converse true? If so, we would have that every finite collection of removable ideals is a removable family of ideals, that is, **RId** (finite). The method would be, using the given systems

$$c_{1,i}, \cdots, c_{n,i} \qquad (i \in I),$$

to construct a new system

$$c_k = \prod_{i \in I} c_{k_i, i}$$

($k_1, \cdots, k_n$ a permutation of $1, \cdots, n$) and apply that converse.

Bibliography

1. Richard Arens, *Inverse-producing extensions of normed algebras*, Trans. Amer. Math. Soc., **88** (1958), 536–548.

2. ——— and Kenneth Hoffman, *Algebraic extensions of normed algebras*, Proc. Amer. Math. Soc., **7** (1956),, 203–210.

3. N. Bourbaki, *Eléments de mthématique*, Livre II, Algebra, Chap. III. A. S. I. 1044, 1948.

4. I. M. Gelfand, D. Raikov, and G. E. Shilov, *Commutative normed rings* (in Russian,) Uspehi Mat. Nauk N. S., **2** (1946), 44–146,

5. Lynn H. Loomis, *Abstract Harmonic Analysis*, van Nostrand, New York, 1953.

6. G. E. Shilov, *On normed rings possessing one generator* (in Russian) Mat. Sb. N. S., **21** (1947), 35–47. See also C. E. Rickart's review, Math. Rev. 9, 455.

7. I. M. Singer and Richard Arens, *Generalized analytic functions*, Trans. Amer. Math. Soc., **81** (1956),, 379–393.

University of California at Los Angeles

SPECTRAL THEORY FOR LINEAR SYSTEMS OF DIFFERENTIAL EQUATIONS

FRED BRAUER

Introduction. The study of boundary value problems for systems of first order differential equations was begun by Bliss in 1926 [1]. Such problems are of interest not only because they include boundary value problems for single equations of arbitrary order, but also because they arise in the calculus of variations and relativistic quantum mechanics. Until now, attention has been concentrated on boundary value problems on a finite interval [1, 2, 8], but an application to a particular boundary value problem on an infinite interval has also been considered [6]. It seems reasonable to expect that the theory of boundary value problems and eigenfunction expansions on an infinite interval for a single differential equation of arbitrary order can be extended to first order systems. In this paper, the extension will be carried out along lines similar to those used by the author in [3]. It will be shown that all the results obtained in [3] can be formulated so as to be valid for systems. Vector and matrix notation will be used extensively, and as a result, formulae will take a simpler and more natural form than in [3].

The elements of a matrix A will be denoted by A_{ij}, and the components of a row or column vector f will be denoted by f_i in the usual manner. The adjoint of a matrix A, written A^*, will be the matrix with $\bar{A}_{ji}$ in the ith row, jth column, the bar indicating the complex conjugate. The adjoint f^* of a row or column vector f will be the column or row vector respectively with components $\bar{f}_i$. It is easily seen that $(AB)^* = B^*A^*$, whether A and B are vectors or matrices such that AB is defined. If A is a matrix and α is a scalar, then $(\alpha A)^* = \bar{\alpha}A^*$. Also, if A is a Hermitian matrix ($A = A^*$), and f and g are column vectors, then it is easy to see that $(f^*Ag)^* = g^*Af = \overline{(f^*Ag)}$. The matrix dA/dt, or A', will be the matrix with elements A'_{ij}, and the vector df/dt, or f', will be the vector with components f'_i. Any analytic properties, such as continuity or differentiability, postulated for a vector or matrix will be understood to be assumed for each element separately.

1. The expansion theorem. Let A_0, A, B be $n \times n$ continuous complex-valued matrix functions of t defined on an interval $I = (a, b)$, not necessarily a bounded interval, with not all elements of B vanishing identically on I and with A_0 non-singular at every point of I. We are

Received April 20, 1959.

interested in boundary value problems for the linear system of differential equations

$$(1)\qquad A_0x' + Ax = \lambda Bx ,$$

where x is an n-dimensional column vector. The adjoint system is defined to be

$$(2)\qquad -(A_0^*y)' + A^*y = \lambda B^*y .$$

The system (1) is called symmetric if there exists a transformation $y = C(t)x$, with C a non-singular continuously differentiable matrix on I, which transforms (1) into (2) for all values of λ. It can easily be shown (cf. [8]) that (1) is symmetric if and only if

$$(3)\qquad (A_0^*C)' - A_0^*CA_0^{-1}A - A^*C = 0, \qquad B^*C + A_0^*CA_0^{-1}B = 0 .$$

If (3) is satisfied, it may easily be verified that

$$(4)\qquad -(A_0^*Cx)' + A^*Cx - \lambda B^*Cx = -A_0^*CA_0^{-1}(A_0x' + Ax - \lambda Bx) .$$

It may be shown by integration by parts that if f and g are two differentiable vector functions vanishing at the ends of the interval I, then

$$\int_I (Cg)^*(A_0f' + Af - \lambda Bf)dt = \int_I [-(A_0^*Cg)' + A^*Cg - \bar{\lambda}B^*Cg]^*f dt .$$

If the system (1) is symmetric, (4) yields

$$(5)\qquad \int_I (Cg)^*(A_0f' + Af - \lambda Bf)dt$$
$$= -\int_I (A_0g' + Ag - \bar{\lambda}Bg)^*A_0^{*-1}C^*A_0f dt .$$

Let $C_0^1(I)$ denote the set of continuously differentiable vector functions which vanish identically outside some compact subinterval of I. A symmetric linear system (1) is called definite if

(i) the matrix $S = C^*B$ is Hermitian, so that $C^*B = B^*C$,

(ii) $\int_I f^*Sf dt \geqq 0$ for any $f \in C_0^1(1)$, and

(iii) $A_0u' + Au = 0, Bu = 0$ on any subinterval J of I implies that u vanishes identically on J.

In view of these conditions,

$$(6)\qquad [f, g] = \int_I g^*Sf dt$$

may be regarded as an inner product on $C_0^1(I)$. Let H be the Hilbert space completion of $C_0^1(I)$ in the inner product (6). Then H is the set of equivalence classes of vector functions f such that $\int_I f^*Sf\,dt < \infty$. The norm in H will be denoted by $\|f\|$.

Let D denote the set of functions f in $C_0^1(I)$ such that

$$A_0f' + Af = Bp \tag{7}$$

for some p in H. Although p may not be uniquely determined as a function by (7), the function Bp is uniquely determined. If p_1 and p_2 are elements of H with $Bp_1 \equiv Bp_2$, then

$$\|p_1 - p_2\|^2 = \int_I (p_1 - p_2)^*C^*B(p_1 - p_2)dt = 0 ,$$

and p_1 and p_2 define the same element of H. Thus the equation (7) determines a unique element p of H. We define an operator L in H with domain D, by defining $Lf = p$ for $f \in D$, with p determined by (7).

LEMMA 1. *If the system* (1) *is symmetric and definite, then the operator* L *is symmetric on* D.

Proof. Let $f, g \in D$, with p as in (7) and

$$A_0g' + Ag = Bq . \tag{8}$$

Then,

$$\begin{aligned}
[Lf, g] &= \int_I g^*Sp\,dt = \int_I g^*B^*Cp\,dt = -\int_I g^*A_0^*CA_0^{-1}Bp\,dt \\
&= -\int_I g^*A_0^*CA_0^{-1}(A_0f' + Af)dt \\
&= -[g^*A_0^*Cf]_a^b + \int_I (g^*A_0^*C)'f\,dt - \int_I g^*A_0^*CA_0^{-1}f\,dt \\
&\quad -[g^*A_0^*Cf]_a^b + \int_I g^{*\prime}A_0^*Cf\,dt + \int_I g^*[(A_0^*C)' - A_0^*CA_0^{-1}A]f\,dt ,
\end{aligned}$$

using (3), (7), and integration by parts. Also,

$$\begin{aligned}
[f, Lg] &= \int_I q^*Sf\,dt = \int_I q^*B^*Cf\,dt = \int_I (A_0g' + Ag)^*Cf\,dt \\
&= \int_I (g^{*\prime}A_0^*Cf + g^*A^*Cf)dt ,
\end{aligned}$$

using (8). Thus

$$\begin{aligned}
[Lf, g] - [f, Lg] = &-[g^*A_0^*Cf]_a^b \\
&+ \int_I g^*[(A_0^*C)' - A_0^*CA_0^{-1}A - A^*C]f\,dt .
\end{aligned}$$

The integral vanishes because of (3), and the first term on the right side vanishes because f and g vanish outside a compact subinterval of I. Therefore $[Lf, g] = [f, Lg]$, and L is symmetric on D.

Throughout this paper, we shall assume that (1) is symmetric and definite, and that the symmetric operator L has a self-adjoint extension T, considered as an operator in H. If A_0, A, and B have real coefficients, then L is a real operator and always has at least one self-adjoint extension ([9], p. 329).

LEMMA 2. *There exists a matrix $k(t, s, \lambda)$ with the following properties:*

(i) *k is continuous on $I \times I$ for fixed λ except on $t = s$, and analytc in λ for fixed t, s,*

(ii) *$k(s + 0, s, \lambda) - k(s - 0, s, \lambda)$ is the identity matrix E for $s \in I$ and any λ,*

(iii) *the columns of k satisfy* (1) *as functions of t for $t \neq s$,*

(iv) *if J is any compact subinterval of I and f is any function in $C_0^1(J)$, then*

$$f(t) = \int_J k(t, s, \lambda)[A_0(s)f'(s) + A(s)f(s) - \lambda B(s)f(s)]ds \,, \tag{9}$$

for $t \in J$.

Proof. Let $\Phi(t, \lambda)$ be a fundamental matrix solution of (1), that is, a matrix whose columns are linearly independent solutions of (1). This matrix is non-singular for all $t \in I$, and can be chosen so that all its elements are analytic in λ for each fixed t. For $t < s$, define $k(t, s, \lambda) = 0$, and for $t \geqq s$, define $k(t, s, \lambda) = \Phi(t, \lambda)\Phi^{-1}(s, \lambda)$. The properties (i)-(iii) are immediate consequences of this definition, and the property (iv) follows from the variation of constants formula ([5], p. 74).

The function $k(t, s, 0)$ will be denoted by $k(t, s)$. In this section, we will use only $k(t, s)$, but the more general $k(t, s, \lambda)$ will be required later. An expression such as $k(t, \ )$ will stand for $k(t, s)$, considered as a function of s for any fixed t. Let J be any compact subinterval of I and let θ_J be a real continuously differentiable scalar functions which is 1 on J and which vanishes identically outside some compact subinterval of I. Let $z(t, s) = C^{-1}(s)k^*(t, s)\theta_J(s)$, an $n \times n$ matrix. It is clear that the columns $z_i(t, \ )$ of $z(t, \ )$ are continuously differentiable vectors which vanish outside a compact subinterval of I, and that each $z_i(t, \ )$ is an element of H. If f belongs to D and vanishes identically outside J, then we can write

$$f(t) = \int_J \theta_J(s)k(t, s)B(s)p(s)ds = \int_J z^*(t, s)C^*(s)B(s)p(s)ds$$

$$= \int_I z^*(t, s)S(s)p(s)ds ,$$

using (7), (9), and $S = C^*B$. This means that each component f_i of f $(i = 1, \cdots, n)$ can be written

$$f_i(t) = \int_J z_i^*(t, s)S(s)p(s)ds = [p, z_i(t,\ )] = [Lf, z_i(t,\ )] . \tag{10}$$

We will make use of the theory of direct integrals and the spectral theorem as given in [7]. The notation will be similar, but not identical, to that used in [3]. The elements of the direct integral $L^2(\sigma, \nu)$ are $\nu(\lambda)$-dimensional vectors $F(\lambda)$, and the inner product

$$(F, G) = \int_R \sum_{k=1}^{\nu(\lambda)} F_k(\lambda)\overline{G}_k(\lambda)d\sigma(\lambda)$$

of two elements F, G of $L^2(\sigma, \nu)$ will be denoted by $\int_R G^*(\lambda)F(\lambda)d\sigma(\lambda)$, in analogy to our other notation. R will always mean the real line.

We can now state the result of this section.

THEOREM 1. *Let T be a self-adjoint extension with domain D_T of the operator L defined for a symmetric definite system* (1). *The spectral theorem furnishes a direct integral $L^2(\sigma, \nu)$ and a unitary transformation U from H to $L^2(\sigma, \nu)$ which diagonalizes T. This transformation is given by*

$$(Uf)(\lambda) = \int_I E^*(t, \lambda)S(t)f(t)dt , \tag{11}$$

and its inverse by

$$(U^{-1}F)(t) = \int_R E(t, \lambda)F(\lambda)d\sigma(\lambda) , \tag{12}$$

with the integrals converging to the functions in the norms of the Hilbert spaces $L^2(\sigma, \nu)$ and H respectively. Here, $E(t, \lambda)$ is a matrix function with n rows and $\nu(\lambda)$ columns, whose elements have locally square-integrable derivatives with respect to t. The columns of $E(t, \lambda)$ are improper eigenfunctions (not necessarily belonging to H) *of the differential equation* (1) *for almost all λ. If λ_0 is an eigenvalue of T, then the columns of $E(t, \lambda_0)$ are proper eigenfunctions.*

Proof. Let $L^2(\sigma, \nu)$ be a suitable direct integral and let U be the unitary mapping of H to $L^2(\sigma, \nu)$ which diagonalizes the self-adjoint extension T of L. The fact that U is unitary is expressed by the Parseval formula

$$[f, g] = (Uf, Ug) = \int_R (Ug)^*(\lambda)(Uf)(\lambda)d\sigma(\lambda) . \tag{13}$$

Let f belong to D_T, the domain of T, and let g be any function in H such that Sg vanishes identically outside some compact subinterval J of I. Let $F = Uf$, $G = Ug$, $Z_i = Uz_i$, $E^i(t, \lambda) = \lambda Z_i^*(t, \lambda)$, where z_i is as in (10). Then

$$\begin{aligned} f_i(t) &= [Tf, z_i(t, \ )] = (UTf, Z_i(t, \ )) = (\lambda Uf, Z_i(t, \ )) \\ &= (F, E^{i*}(t, \ )) = \int_R E^i(t, \lambda)F(\lambda)d\sigma(\lambda) , \end{aligned}$$

using (10), (13), and the spectral theorem. In addition,

$$\begin{aligned} [f, g] &= \int_J g^*Sf dt = \int_J \sum_{i=1}^n [g^*S]_i f_i dt \\ &= \int_J \sum_{i=1}^n [g^*(t)S(t)]_i \left\{ \int_R E^i(t, \lambda)F(\lambda)d\sigma(\lambda) \right\} dt \\ &= \int_R \left\{ \int_J \sum_{i=1}^n [g^*(t)S(t)]_i E^i(t, \lambda)dt \right\} F(\lambda)d\sigma(\lambda) , \end{aligned}$$

the interchange in the order of integration being justified by the absolute convergence of the integral. We define the $n \times \nu(\lambda)$ matrix $E(t, \lambda)$ with rows $E^i(t, \lambda)$. Then we can write

$$[f, g] = \int_R \left\{ \int_J g^*(t)S(t)E(t, \lambda)dt \right\} F(\lambda)d\sigma(\lambda) .$$

On the other hand,

$$[f, g] = \int_R G^*(\lambda)F(\lambda)d\sigma(\lambda) ,$$

and thus

$$G^*(\lambda) = \int_J g^*(t)S(t)E(t, \lambda)dt ,$$

or,

$$G(\lambda) = \int_J E^*(t, \lambda)S(t)g(t)dt$$

for almost all λ.

For $g \in D$, g vanishing identically outside J, we have seen that $(Ug)(\lambda) = \int_J E^*(t, \lambda)S(t)g(t)dt$. If $A_0g' + Ag = Bp$, then $Bp = BTg = 0$ outside J, $STg = 0$ outside J, and we can apply the above relation to Tg, obtaining

$$(UTg)(\lambda) = \int_J E^*(t, \lambda)S(t)Tg(t)dt .$$

Since

$$(UTg)(\lambda) = \lambda(Ug)(\lambda) = \int_J E^*(t, \lambda)S(t)g(t)dt ,$$

we obtain

$$\int_J E^*(t, \lambda)S(t)[Tg(t) - \lambda g(t)]dt = 0 , \tag{14}$$

when λ does not belong to a set N_g of measure zero, with N_g dependent on g. The same is true for a sequence g_j of functions when λ does not belong to the null set $N = \bigcup_{i=1}^{\infty} N_{g_j}$. We choose the sequence g_j dense in $D \cap C_0^1(J)$, and then (14) holds for all $g \in D \cap C_0^1(J)$ if $\lambda \notin N$. We let $E(t, \lambda) = 0$ for $\lambda \in N$, and then (14) holds for all λ. Since $S = C^*B$, (14) yields

$$\int_J E^*(t, \lambda)C^*(t)[B(t)Tg(t) - \lambda B(t)g(t)]dt = 0 ,$$

or

$$\int_J E^*(t, \lambda)C^*(t)[A_0(t)g'(t) + A(t)g(t) - \lambda B(t)g(t)]dt = 0 .$$

Thus the columns of $C(t)E(t, \lambda)$ are weak solutions of (1) on J. It follows from a well-known theorem on weak solutions of partial differential equations that the columns of $C(t)E(t, \lambda)$ have locally square-integrable derivatives with respect to t which are continuous after correction on a null set for each λ, and that each column is a solution of (1). This theorem is easily proved using the properties of $k(t, s, \lambda)$. Since $C(t)$ is non-singular, the columns of $E(t, \lambda)$ are improper eigenfunctions of (1).

The matrix E depends on the compact subinterval J. Let E' be another matrix with the same properties, corresponding to an interval $J' \supseteqq J$. Then

$$\int_J [E^*(t, \lambda) - E'^*(t, \lambda)]S(t)g(t)dt = 0$$

for almost all λ, independent of $g \in C_0^1(J)$. It follows that $S(t)E(t, \lambda) - S(t)E'(t, \lambda) = 0$ for λ outside some null set. For λ in this null set we redefine $E(t, \lambda) = E'(t, \lambda) = 0$. The columns of $E(t, \lambda) - E'(t, \lambda)$ satisfy $Bu = 0$. At the same time, since E and E' are eigenfunctions of (1), they satisfy $A_0u' + Au = \lambda Bu = 0$. By hypothesis (iii) in the definiteness of (1), $E(t, \lambda) = E'(t, \lambda)$ on J for all λ. By taking a sequence of compact subintervals J tending to I, we can extend E uniquely to a matrix function defined for $t \in I$ and all λ.

If λ_0 is an eigenvalue of T, then σ has a jump, which we may assume to be a jump of 1, at λ_0. We choose $F = 0$ except at λ_0, and $F_j(\lambda_0) = \delta_{jk}$ for any fixed index $k \leqq \nu(\lambda_0)$. Then $F \in L^2(\sigma, \nu)$ and

$$(U^{-1}F)(t) = \int_R E(t, \lambda)F(\lambda)d\sigma(\lambda) = E_k(t, \lambda_0) ,$$

the kth column of $E(t, \lambda_0)$, an element of H. Thus the columns of $E(t,\lambda_0)$ are proper eigenfunctions of T if λ_0 is an eigenvalue of T.

The inversion formulae (11), (12), obtained for functions f in D_T which vanish identically outside a compact subinterval J, can be extended to all functions in D_T by a standard density argument. They are valid with the integrals converging to the functions in the norms of the appropriate Hilbert spaces. These formulae give the expansion of an arbitrary function $f \in D_T$ in eigenfunctions of the system of differential equations (1). The proof of Theorem 1 is now complete.

To prepare for the next section, we write the expansion formulae in a different form. Let $\Phi(t, \lambda)$ be a fundamental matrix solution of (1), with each element analytic in λ for fixed t. The matrix $E(t, \lambda)$ can be expressed in terms of $\Phi(t, \lambda)$ by

$$E(t, \lambda) = \Phi(t, \lambda)R(\lambda) , \tag{15}$$

where $R(\lambda)$ is a matrix with n rows and $\nu(\lambda)$ columns whose elements are functions of λ only. With the use of (15), the Parseval equality (13) takes the form

$$\begin{aligned}
||f||^2 &= \int_R F^*(\lambda)F(\lambda)d\sigma(\lambda) \\
&= \int_R \left[\int_I f^*(t)S(t)E(t, \lambda)dt\right]\left[\int_I E^*(s, \lambda)S(s)f(s)ds\right]d\sigma(\lambda) \\
&= \int_R \left[\int_I f^*(t)S(t)\Phi(t, \lambda)R(\lambda)dt\right]\left[\int_I R^*(\lambda)\Phi^*(s, \lambda)S(s)f(s)ds\right]d\sigma(\lambda) \\
&= \int_R (Vf)^*(\lambda)R(\lambda)R^*(\lambda)(Vf)(\lambda)d\sigma(\lambda) ,
\end{aligned}$$

where

$$(Vf)(\lambda) = \int_I \Phi^*(t, \lambda)S(t)f(t)dt . \tag{16}$$

The formula

$$d\rho(\lambda) = R(\lambda)R^*(\lambda)d\sigma(\lambda) \tag{17}$$

defines a Hermitian positive semi-definite $n \times n$ matrix, called a spectral matrix. Let H^* be the Hilbert space of all complex-valued n-dimensional vector functions $F(\lambda)$ such that

$$\int_R \sum_{k=1}^{n} \bar{F}_j(\lambda)F_k(\lambda)d\rho_{jk}(\lambda) = \int_R F^*(\lambda)d\rho(\lambda)F(\lambda) < \infty ,$$

with inner product

$$(F, G) = \int_R G^*(\lambda) d\rho(\lambda) F(\lambda) .$$

Then (16) defines a unitary mapping of H onto H^* which diagonalizes T. A straightforward computation gives

$$(V^{-1}F)(t) = \int_R \Phi(t, \lambda) d\rho(\lambda) F(\lambda) . \tag{18}$$

2. Green's function and the spectral matrix. Let T be a self-adjoint extension of L as in §1, and let $R_\lambda = (T - \lambda)^{-1}$, for $Im\lambda \neq 0$, be the resolvent of T, a bounded operator in H.

THEOREM 2. *There exists an $n \times n$ matrix $G(t, s, \lambda)$ defined for $t, s \in I$, $Im\lambda \neq 0$, such that*

$$S(t) R_\lambda f(t) = \int_J G(t, s, \lambda) S(s) f(s) ds , \tag{19}$$

where J is a compact subinterval of I, $t \in J$, and $f \in C_0^1(J)$. This matrix G, called the Green's matrix of the operator T, has the following properties:

(i) *G is analytic in λ for fixed t, s and $Im\lambda \neq 0$, is continuous in (t, s) on $I \times I$ for fixed λ except on the diagonal $t = s$*

(ii) *$G(s + 0, s, \lambda) - G(s - 0, s, \lambda) = E$ for $s \in I$, $Im\lambda \neq 0$*

(iii) *$G(t, s, \lambda) S(s) = S(t) G^*(s, t, \bar{\lambda})$*

(iv) *considered as functions of t, the columns of G satisfy (1) if $t \neq s$*

(v) *G is uniquely determined by T*

(vi) *if $f \in C_0^1(I)$, then $S(t) f(t) = \int_I G(t, s, \lambda) S(s) (T - \lambda) f(s) ds$.*

Proof. (cf. [7], p. 14). If $f \in C_0^1(J)$, $g \in C_0^1(I)$, then

$$\begin{aligned} [f, g] &= \int_I g^*(t) S(t) f(t) dt = [R_\lambda f, (T - \lambda) g] \\ &= \int_J [(T - \lambda) g(t)]^* S(t) R_\lambda f(t) dt , \end{aligned} \tag{20}$$

by (6) and the definition of the resolvent. We make use of a matrix $k(t, s, \lambda)$ as in Lemma 2. Let s_0 be any point of J, V a neighbourhood of s_0 whose closure is contained in J, and θ_V a real scalar function in $C_0^1(J)$ which is equal to 1 on V. For $t \in I$, $s \in V$, define

$$p(t, s) = (T_t - \lambda)[k(t, s, \lambda)(1 - \theta_V(t))] , \tag{21}$$

the subscript t indicating that the operator is applied to $k(t, s, \lambda)(1-\theta_V(t))$ considered as a function of t for fixed s. The result of application of an operator to a matrix will be understood as the matrix whose columns are obtained by applying the operator to the columns of the original matrix. For fixed $s \in V$, $p(\ ,s)$ vanishes except on a "ring" contained in $J - V$, and the matrix function $p(t, s)$ is continuous on $I \times V$. Consider $v(t) = \int_J p^*(s, t)S(s)R_\lambda f(s)ds$. If $g \in C_0^1(V)$, then

$$\int_V g^*(t)v(t)dt = \int_V g^*(t)\left[\int_J p^*(s, t)S(s)R_\lambda f(s)dsdt\right. \tag{22}$$

$$= \int_J\left[\int_V g^*(t)p^*(s, t)dt\right]S(s)R_\lambda f(s)ds ,$$

However, if $u(s, g) = \int_J k(s, t, \lambda)g(t)dt$, then

$$\int_V p(s, t)g(t)dt = \int_V (T_s - \lambda)k(s, t, \lambda)g(t)dt$$

$$- \int_V (T_s - \lambda)k(s, t, \lambda)\theta_V(s)g(t)dt$$

$$= g(s) - (T_s - \lambda)\theta_V(s)u(s, g) ,$$

using the properties of k and (21). Substituting in (22),

$$\int_V g^*(t)v(t)dt = \int_J g^*(s)S(s)R_\lambda f(s)ds - \int_J [(T_s - \lambda)\theta_V(s)u(s, g)]^*S(s)R_\lambda f(s)ds$$

$$= \int_J g^*(s)S(s)R_\lambda f(s)ds - \int_J \theta_V(s)u^*(s, g)S(s)f(s)ds$$

$$= \int_J g^*(s)S(s)R_\lambda f(s)ds - \int_J\int_J \theta_V(s)g^*(t)k^*(s, t, \lambda)S(s)f(s)ds ,$$

using (20) and the definition of $u(s, g)$. Since this holds for all $g \in C_0^1(V)$, we obtain $S(t)R_\lambda f(t) = v(t) + \int_J \theta_V(s)k^*(s, t, \lambda)S(s)f(s)ds$ for almost all $t \in V$. If $k_1(s, t, \lambda) = R_\lambda^* p(s, t)$, then

$$v(t) = \int_J p^*(s, t)S(s)R_\lambda f(s)ds = [R_\lambda f, p(\ , t)] = [f, R_\lambda^* p(\ , t)]$$

$$= [f, k_1(\ , t, \lambda)] = \int_J k_1^*(s, t, \lambda)S(s)f(s)ds ,$$

using the definition of the adjoint operator R_λ^*. It is clear that $k_1(s, t, \lambda)$ is continuous on $V \times I$ for fixed λ, $Im\lambda \neq 0$, since R_λ^* is the inverse of a differential operator. Now

$$S(t)R_\lambda f(t) = \int_J [k_1^*(s, t, \lambda) + \theta_V(s)k^*(s, t, \lambda)]S(s)f(s)ds .$$

and the definition

$$G(t, s, \lambda) = k_1^*(s, t, \lambda) + \theta_V(s)k^*(s, t, \lambda) \tag{23}$$

yields (19). As this can be done for any $s_0 \in J$, (19) holds for all $t, s \in J$. The analogue of this result in [3] is proved incorrectly, as has been pointed out to the author by Professor M. H. Stone. A correct proof can be given essentially following the argument used here. The matrix G depends on the interval J, but is uniquely determined by J. If J' is another compact subinterval of I which contains J, and G' is the corresponding matrix, it is easy to see that $G(t, s, \lambda) = G'(t, s, \lambda)$ for t, $s \in J$, $Im\lambda \neq 0$. Thus, by taking a sequence of compact subintervals J tending to I, we can extend G uniquely to a matrix function defined for $t, s \in I$.

The remainder of the proof consists of the verification of the properties of the Green's matrix. The property (vi) follows immediately from the definition of the resolvent and (19). Since $R_\lambda^* = R_{\bar\lambda}$, $[R_\lambda f, g] = [f, R_{\bar\lambda} g]$ for any $f, g \in C_0^1(I)$. Then

$$\int_I g^*(t)S(t)R_\lambda f(t)dt = \int_I [R_{\bar\lambda} g(s)]^* S(s)f(s)ds = \int_I [S(s)R_{\bar\lambda} g(s)]^* f(s)ds ,$$

and, using (19), this yields

$$\int_I \int_I g^*(t)[G(t, s, \lambda)S(s) - S(t)G^*(s, t, \bar\lambda)]f(s)dsdt = 0 .$$

Since this holds for all $f, g \in C_0^1(I)$, we obtain

$$G(t, s, \lambda)S(s) = S(t)G^*(s, t, \bar\lambda) , \tag{24}$$

which is property (iii), for almost all $s, t \in I$. As $k_1(s, t, \lambda)$ is continuous, (23) shows that $G(t, s, \lambda)$ has the same analytic behaviour as $k^*(s, t, \lambda)$, in particular the same discontinuity at $s = t$, and the properties (i) and (ii) follow from Lemma 2 of §1. In view of the continuity of the matrices involved, (24) must actually be true for all $s, t \in I$. To prove (iv), we begin with (vi), written as

$$\begin{aligned} S(t)f(t) &= \int_I G(t, s, \bar\lambda)C^*(s)\left[A_0(s)\frac{d}{ds} + A(s) - \bar\lambda B(s)\right]f(s)ds \\ &= \int_I [C(s)G^*(t, s, \bar\lambda)]^*\left[A_0(s)\frac{d}{ds} + A(s) - \bar\lambda B(s)\right]f(s)ds . \end{aligned}$$

using the definition $S = C^*B$. Application of (5) yields

$$\begin{aligned} S(t)f(t) = &- \int_I \Big[\Big(A_0(s)\frac{d}{ds} + A(s) - \lambda B(s)\Big) \\ &\times G^*(t, s, \bar\lambda)\Big]^* A_0^{*-1}(s)C^*(s)A_0(s)f(s)ds . \end{aligned} \tag{25}$$

Since (25) is true for all $f \in C_0^1(I)$, the columns of $G^*(t, s, \bar\lambda)$, considered as functions of s, satisfy (1) for $t \neq s$. This, together with (24), proves (iv).

If there were two Green's matrices for $Im\lambda \neq 0$, their difference would be continuous everywhere and would be an eigenfunction of the operator T. As the spectrum of the self-adjoint operator T is real, this is impossible, and the Green's matrix is therefore unique. This completes the proof of Theorem 2.

Now we express the Green's matrix in terms of the fundamental matrix solution $\Phi(t, \lambda)$ of (1) introduced at the end of §1. From the properties of the Green's matrix, it is easy to deduce that G may be written

$$G(t, s, \lambda) = S(t)\Phi(t, \lambda)P^{+}(\lambda)\Phi^{*}(s, \bar{\lambda}) \qquad (s \geqq t) \tag{26}$$
$$G(t, s, \lambda) = S(t)\Phi(t, \lambda)P^{-}(\lambda)\Phi^{*}(s, \bar{\lambda}) \qquad (s \leqq t) .$$

The matrices P^{+} and P^{-} are analytic in λ except possibly on the real axis, and $P^{-*} = P^{+}$. We define the matrix $P = \frac{1}{2}(P^{+} + P^{-})$, and then P is analytic for $Im\lambda \neq 0$ and Hermitian.

THEOREM 3 (Titchmarsh-Kodaira formula). *The Green's matrix G of T is related to the spectral matrix ρ associated with the fundamental matrix solution Φ of* (1) *by the formula*

$$P(\mu) = \int_{-\infty}^{\infty} d\rho(\lambda)/(\lambda - \mu) , \tag{27}$$

where P is as defined above, and (27) *is to be taken in the sense that $P(\mu) - \int_{-N}^{N} d\rho(\lambda)/(\lambda - \mu)$ is analytic across the real axis on the interval* $(-N, N)$.

Proof. Let $f \in D_T$, $F = Vf$. Then, by (18),

$$f(t) = \int_R \Phi(t, \lambda)d\rho(\lambda)F(\lambda) .$$

Let

$$u(t) = \int_R \Phi(t, \lambda)d\rho(\lambda)F(\lambda)/(\lambda - \mu) .$$

Then

$$A_0u' + Au - \mu Bu = \int_R \lambda B(t)\Phi(t, \lambda)d\rho(\lambda)F(\lambda)/(\lambda - \mu)$$
$$- \int_R \mu B(t)\Phi(t, \lambda)d\rho(\lambda)F(\lambda)/(\lambda - \mu) = B(t)f(t) ,$$

or $u = R_{\mu}f$. Thus

$$\mu(Vu)(\lambda) = \mu\int_I \Phi^{*}(t, \lambda)S(t)u(t)dt = \int_I \Phi^{*}(t, \lambda)C^{*}(t)\mu B(t)u(t)dt$$

$$= \int_I \Phi^*(t, \lambda)C^*(t)[A_0(t)u'(t) + A(t)u(t) - \mu B(t)f(t)]dt$$
$$= \int_I \Phi^*(t, \lambda)C^*(t)[A_0(t)u'(t) + A(t)u(t)]dt - (Vf)(\lambda)$$
$$= V(Tu)(\lambda) - (Vf)(\lambda) = \lambda(Vu)(\lambda) - (Vf)(\lambda) ,$$

using (16), $u = R_\mu f$, and the fact that V diagonalizes T. Thus $(\lambda - \mu)(Vu)(\lambda) = (Vf)(\lambda)$. Applying the Parseval equality to u and f, $[u, f] = \int_R (Vf)^*(\lambda)d\rho(\lambda)(Vu)(\lambda) = \int_R F^*(\lambda)d\rho(\lambda)F(\lambda)/(\lambda - \mu)$, which is $F^*(\mu)\left[\int_{-N}^{N} d\rho(\lambda)/(\lambda - \mu)\right]F(\mu)$ plus a matrix which is analytic unless μ is real and $|\mu| \geqq N$. On the other hand, $S(t)u(t) = \int_I G(t, s, \mu)S(s)f(s)ds$, and $[u, f] = \int_I\int_I f^*(t)G(t, s, \mu)S(s)f(s)dsdt$, which, using (26), is equal to $F^*(\mu)P(\mu)F(\mu)$ plus an analytic function. Letting f run through a dense subset of H, which means, that F runs through a dense subset of H^*, we conclude that $P(\mu) - \int_{-N}^{N} d\rho(\lambda)/(\lambda - \mu)$ is analytic unless μ is real and $|\mu| \geqq N$.

Another form of the Titchmarsh-Kodaira formula is

$$\rho(\lambda) = \lim_{\delta \to 0+} \lim_{\varepsilon \to 0+} \frac{1}{2\pi i}\int_{\delta}^{\lambda+\delta} [P(\mu + i\varepsilon) - P(\mu - i\varepsilon)]d\mu ,$$

with ρ normalized to be continuous from the right and $\rho(0) = 0$, and with the formula interpreted in the same way as (27). The proof is exactly the same as the corresponding proof in [3], a straightforward inversion.

3. **Boundary conditions.** Let D_0 be the set of functions f in H such that $A_0 f' + Af$ exists almost everywhere on I and such that (7) is satisfied for some p in H. Let T_0 be the operator in H with domain D_0 defined by $T_0 f = p$ for $f \in D_0$, p as in (7). We assume that T_0 has at least one self-adjoint restriction. Let R_λ be the resolvent of some self-adjoint restriction of T_0, so that

$$S(t)R_\lambda f(t) = \int_I G(t, s, \lambda)S(s)f(s)ds ,$$

for $f \in H$, $Im\lambda \neq 0$. Then R_λ is a bounded operator for $Im\lambda \neq 0$, mapping H into D_0, whose adjoint is $R_{\bar{\lambda}}$. Let $\varepsilon(\lambda)$ be the eigenspace of T_0 corresponding to the value λ, the set of all solutions in D_0 of the differential system (1).

LEMMA 3. *T_0 is a closed operator whose domain consists of all $f \in H$ of the form $f = R_\lambda h + w$, where $h \in H$, $w \in \varepsilon(\lambda)$, for any λ with $Im\lambda \neq 0$.*

Proof. Since R_λ maps H into D_0 and $\varepsilon(\lambda)$ is containined in D_0, it is clear that every f of this form belongs to D_0. Conversely, suppose $f \in D_0$ is given. Let $h = T_0 f - \lambda f$, $w = f - R_\lambda h$. Then

$$T_0 w = T_0 f - T_0 R_\lambda h = T_0 f - \lambda R_\lambda h - h = T_0 f - h - \lambda(h - w) = \lambda w ,$$

and thus $w \in \varepsilon(\lambda)$, while $f = R_\lambda h + w$. If f is written in this way, $T_0 f - \lambda f = h$. If f_k is a sequence in D_0 such that $f = \lim f_k$ and $f^* = \lim T_0 f_k$ exist, we can write $f_k = R_\lambda(T_0 f_k - \lambda f_k) + w_k$, and deduce that $w = \lim w_k$ exists and belongs to $\varepsilon(\lambda)$. Letting $k \to \infty$, we obtain $f = R_\lambda(f^* - \lambda f) + w$, which implies $f \in D_0$ and $T_0 f = f^*$. This proves that T_0 is closed.

Since T_0 is closed and its domain D_0 is dense in H, T_0 has a closed adjoint T_0^* whose domain D_0^* is dense in H. Also, $T_0 = T_0^{**} = (T_0^*)^*$. For any subspace M of H, we let $H - M$ denote the orthogonal complement of M in H.

LEMMA 4. *D_0^* consists of all $g \in D_0$ of the form $g = R_\lambda z$, where $z \in H - \varepsilon(\bar{\lambda})$. The operator T_0^* is a restriction of T_0 and is closed and symmetric.*

Proof. $g^* = T_0^* g$ means

$$[T_0 f, g] = [f, g^*] \tag{28}$$

for every $f \in D_0$. By Lemma 3, any $f \in D_0$ may be written $f = R_{\bar{\lambda}} h + w$, with $h \in H$, $w \in \varepsilon(\bar{\lambda})$, and then $T_0 f = \bar{\lambda} f + h$. Substitution in (28) gives

$$[R_{\bar{\lambda}} h + w, g^*] = [\bar{\lambda} f + h, g] = [\bar{\lambda} R_{\bar{\lambda}} h + \bar{\lambda} w + h, g] ,$$

or

$$[h, \lambda \overset{*}{R_{\bar{\lambda}}} g + g - \overset{*}{R_{\bar{\lambda}}} g^*] + [w, \lambda g - g^*] = 0$$

for all $h \in H$, $w \in \varepsilon(\bar{\lambda})$. Then $g^* - \lambda g = z$ is orthogonal to $\varepsilon(\bar{\lambda})$, or $z \in H - \varepsilon(\bar{\lambda})$, and $g = \overset{*}{R_{\bar{\lambda}}}(g^* - \lambda g) = R_\lambda z$. Since R_λ maps H into D_0, g belongs to D_0. Thus $D_0^* \subseteqq D_0$. As it is assumed that there exists a self-adjoint restriction T of T_0 with domain D_T, $D_0 \supseteqq D_T \supseteqq D_0^*$, and since T is symmetric, its restriction T_0^* is also symmetric.

As we have seen in Lemma 1,

$$[T_0 f, g] - [f, T_0 g] = g^*(a) A_0^*(a) C(a) f(a) - g^*(b) A_0^*(b) C(b) f(b)$$

for $f, g \in D_0$. Here, $g^*(t) A_0^*(t) C(t) f(t)$ is a bilinear form in f, g which is non-degenerate for all $t \in I$ and skew-Hermitian. We define

$$\langle fg \rangle = g^*(a) A_0^*(a) C(a) f(a) - g^*(b) A_0^*(b) C(b) f(b) .$$

A homogeneous boundary condition is a condition on $f \in D_0$ of the form $\langle f\alpha\rangle = 0$, where α is a fixed function in D_0 The conditions

$$\langle f\alpha_j\rangle = 0, \qquad (j = 1, \cdots, p) \tag{29}$$

are said to be linearly independent if the only set of complex numbers $\gamma_1, \cdots, \gamma_p$ for which $\sum_{j=1}^{p} \gamma_j \langle f\alpha_j\rangle = 0$ identically in $f \in D_0$ is $\gamma_1 = \cdots = \gamma_p = 0$. Since $[T_0 f, g] - [f, T_0^* g] = \langle fg\rangle$ for $f \in D_0, g \in D_0^*$, it is easily seen that these boundary conditions are linearly independent if and only if the functions $\alpha_1, \cdots, \alpha_p$ are linearly independent (mod D_0^*). A set of p linearly independent boundary conditions (29) is said to be self-adjoint if $\langle \alpha_j \alpha_k\rangle = 0$ for $j, k = 1, \cdots, p$. Two sets of boundary conditions are said to be equivalent if the sets of functions satisfying the two sets of conditions are identical.

The assumption that T_0^* has a self-adjoint extension is equivalent to the assumption that the linear spaces $\varepsilon(i)$ and $\varepsilon(-i)$ have the same dimension τ, the defect index of T_0^*. By exactly the same proof as that used in [3], originally used in [4], we can obtain the following relation between self-adjoint extensions of T_0^* and boundary conditions.

THEOREM 4. *If T is a self-adjoint extension of T_0^** (or, equivalently, restriction of T_0) *with domain D_T, then there exists a self-adjoint set of τ linearly independent boundary conditions such that D_T is the set of all $f \in D_0$ satisfying these conditions. Conversely, corresponding to a self-adjoint set of τ linearly independent boundary conditions, there exists a self-adjoint extension T of T_0^* whose domain D_T is the set of all $f \in D_0$ satisfying these boundary conditions.*

4. **Examples.** The results of this paper include as a special case the corresponding results for a single differential equation of arbitrary order as obtained in [3]. For simplicity, we consider only equations of even order with real coefficients. Let L and M be formally self-adjoint linear differential operators of orders $2r$ and $2s$ respectively $(r > s)$. Then L and M can be written

$$Lu = \sum_{i=0}^{r} [p_{r-i} u^{(i)}]^{(i)}, \qquad Mu = \sum_{i=0}^{s} [q_{s-i} u^{(i)}]^{(i)} ,$$

where p_{r-i}, q_{s-i} are real functions having continuous derivatives up to order i on I. We assume $p_0 \neq 0$ on I. It is not difficult to verify, as suggested in ([5], p. 206, problem 19), that the differential equation $Lu = \lambda Mu$ is equivalent to a system (1). If we let x be the vector with components $(x_1, \cdots, x_{2r})$, with

$$x_j = u^{(j-1)} [j = 1, \cdots, r],\ x_{r+j} = (-1)^j [p_{r-j} u^{(j)} + (p_{r-j-1} u^{(j+1)})'$$

$$+ \cdots + (p_0 u^{(r)})^{(r-j)}] + (-1)^{j+1}[q_{s-j}u^{(j)} + (q_{s-j-1}u^{(j+1)})' \\ + \cdots + (q_0 u^{(s)})^{(s-j)}],\ [j = 1, \cdots, r] ,$$

understanding zero for any expression q_{-k}, $k > 0$, we obtain the system

$$\begin{aligned}
&- x'_{r+1} + p_r x_1 = \lambda q_s x_1 \\
&- x'_{r+2} - p_{r-1}x_2 - x_{r+1} = - \lambda q_{s-1} x_2 \\
&\qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \\
&- x'_{r+s} - (-1)^s p_{r+1-s} x_s + x_{r+s-1} = (-1)^s \lambda q_0 x_s \\
&\qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \\
&- x'_{2r} - (-1)^r p_1 x_r + x_{2r-1} = 0 \\
&x'_1 - x_2 = 0 \\
&x'_2 - x_3 = 0 \\
&\quad \cdot \quad \cdot \quad \cdot \\
&x'_{r-1} - x_r = 0 \\
&x'_r - (-1)^r x_{2r}/p_0 = 0 ,
\end{aligned} \tag{30}$$

which is of the form (1), where

$$A = \begin{pmatrix} 0_r & -E_r \\ E_r & 0_r \end{pmatrix}, \qquad A = \begin{pmatrix} P & Q \\ Q^* & R \end{pmatrix},$$

$$P = \begin{pmatrix} p_r & & & & \\ & -p_{r-1} & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & (-1)^{r-1}p_1 \end{pmatrix}, \quad Q = \begin{pmatrix} 0 \cdots 0 \\ 1 \cdots 0 \\ \cdot \quad\ \cdot \\ 0 \cdot\cdot\, 10 \\ 0 \cdots 1 \end{pmatrix},$$

$$R = \begin{pmatrix} 0 \cdots 0 & \\ \cdot \cdots\cdot & \\ 0 \cdots & (-1)^r/p_0 \end{pmatrix}, \qquad B = \begin{pmatrix} q_s & & & & & \\ & -q_{s-1} & & & & \\ & & \cdot & & & \\ & & & (-1)^s q_0 & & \\ & & & & \cdot & \\ & & & & & 0 \end{pmatrix}$$

E_r denoting the r-dimensional unit matrix, 0_r the r-dimensional zero matrix, and all elements not shown being zero. It is an immediate consequence of (31) that the system (30) is its own adjoint. The set of functions D may be regarded as the set of scalar functions with $2r$ continuous derivatives on I which vanish identically outside some compact subinterval of I, the condition (7) being no restriction. The norm is given by

$$||f||^2 = \sum_{i=0}^{s} \int_I (-1)^i q_{s-i}(t) |f^{(i)}|^2 dt ,$$

and to make the problem definite in the sense of §1, we must assume $(-1)^i q_{s-i}(t) \geqq 0\,(i = 0, 1, \cdots, s)$. With this restriction, we obtain the eigenfunction expansion theorem, the existence of the Green's function, the Titchmarsh-Kodaira formula, and the nature of the boundary conditions as in [3] from the results of this paper.

A problem which has arisen in relativistic quantum mechanics (cf. [6]) involves the pair of differential equations

$$(32) \qquad x_1' = q_1(t)x_2 + \lambda x_2, \qquad x_2' = -\, q_2(t)x_1 - \lambda x_1 ,$$

where q_1 and q_2 are real and continuous on $0 \leqq t < \infty$. This is of the form (1) with $A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $A = \begin{pmatrix} 0 & -q_1 \\ q_2 & 0 \end{pmatrix}$, $B = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. The adjoint system is

$$(33) \qquad y_1' = q_2(t)y_2 + \lambda y_2, \qquad y_1' = -\, q_1(t)y_1 - \lambda y_1 ,$$

and (32) may be transformed into (33) by $y = Cx$ with $C = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, so that $S = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Then $||f||^2 = \int_I (|f_1|^2 + |f_2|^2)dt$, where $f = (f_1, f_2)$. It can be determined that $\rho(\lambda) = \begin{pmatrix} \lambda/\pi & 0 \\ 0 & 0 \end{pmatrix}$. If $(u(x, \lambda), v(x, \lambda))$ is a solution of (32), the expansion formulae are

$$f_1(t) = \frac{1}{\pi}\int_R F_1(\lambda)u(t, \lambda)d\lambda, \, f_2(t) = \frac{1}{\pi}\int_R F_1(\lambda)v(t, \lambda)d\lambda,$$

$$F_1(\lambda) = \int_I [f_1(t)\bar{u}(t, \lambda) + f_2(t)\bar{v}(t, \lambda)]dt ,$$

with F_2 not appearing because ρ has rank 1. Possibly this approach can be used to prove the existence of eigenfunction expansions in more general applications, but its usefulness will be limited by the difficulty in computing the spectral matrix.

REFERENCES

1. G. A. Bliss, *A boundary value problem for a system of ordinary differential equations of the first order*, Trans. Amer. Math. Soc. **28** (1926), 561–584.

2. ———, *Definitely self-adjoint boundary value problems*, Trans. Amer. Math. Soc. **44** (1938), 413–428.

3. F. Brauer, *Spectral theory for the differential equation $Lu = \lambda Mu$*, Can. J. Math. **10** (1958), 431–446.

4. E. A. Coddington, *The spectral representation of ordinary differential operators*, Ann. of Math. **60** (1954), 192–211.

5. E. A. Coddington and N. Levinson, *Theory of ordinary differential equations*, (New

York, 1955),
6. S. D. Conte and W. C. Sangren, *An expansion theorem for a pair of singular first order equations*, Can. J. Math. **6** (1954), 554–560.
7. L. Garding, *Applications of the theory of direct integrals of Hilbert spaces to some integral and differential operators*, Institute for Fluid Dynamics and Applied Mathematics, University of Maryland, 1954.
8. W. T. Reid, *A class of two-point boundary problems*, Ill. J. Math. **2**. (1958), 434–453.
9. F. Riesz and B. Sz. Nagy, *Functional analysis*, (New York, 1955).

UNIVERSITY OF BRITISH COLUMBIA

AREA AND NORMALITY

H. BUSEMANN AND E. G. STRAUS

1. Introduction. The simplest non-Riemannian a-dimensional area (concisely: a-area) is a translation invariant positive continuous measure (or area) defined on the a-dimensional linear subspaces, called a-flats, of an n-dimensional affine space A_n $(1 \leq a \leq n)$. Such areas have been studied by Wagner [15] and they are the subject of the present investigation which is in part related to Wagner's, but has no connection with the differential geometry of general area metrics persued principally in Japan by Kawaguchi, Iwamoto and others.

The simplest case, $a = 1$, is well known. In that case a segment with endpoints x, y has a translation invariant length $d(x, y)$. If the sphere $d(z, x) = 1$ (z fixed) has at x_0 a supporting $(n-1)$-flat (hyperplane) H_0 then H_0 is transversal to the 1-flat (line) L_0 through z and x_0, and L_0 is normal to H_0.

Therefore the existence of an $(n-1)$-flat transversal to a given line is equivalent to the convexity of the sphere $d(z, x) = 1$; which, in turn, is equivalent to the triangle inequality for $d(a, b)$, in other words, to the space being Minkowskian (normed linear).

If L_0 is normal to H_0 at x_0 then it is normal to every line L through x_0 in H_0 in the two-flat spanned by L_0 and L. A well-known theorem of Blaschke [2] states that for $n \geq 3$ normality between lines is symmetric only in euclidean space. However, as shown by Radon [13], this is not the case for $n = 2$.

Here we treat the analogous problems for arbitrary a, and then study the special case of Minkowski area.

We cannot give more than this vague hint without some definitions. Let $(x^1, \cdots, x^n)$ be affine coordinates of a point x in A^n with origin $z = (0, \cdots, 0)$. The *a-box* $[x_0, x_1, \cdots, x_a]$ consists of all points of the form $(1-\theta_i)x_0 + \sum_{i=1}^{a} \theta_i x_i$ where $0 \leq \theta_i, \leq 1$; and hence is a (possibly degenerate) parallelepiped.

An *a-area* assigns to every Borel[1] set M in an a-flat a measure $\alpha(M)$ which is invariant under the translations of A^n, and continuous; that is, $\alpha([x_0, \cdots, x_a])$ depends continuously on $x_0, \cdots, x_a$. The invariance under translation applied to sets in the same a-flat A yields at once that the measure in A is determined up to a factor depending on A. If we introduce an auxiliary euclidean metric

Received, June 19, 1958. The first named author was supported by a grant from the National Science Foundation.

[1]All sets considered will be Borel sets.

$$e(x, y) = \left[\sum_{i,k=1}^{n} g_{ik}(x^i - y^i)(x^k - y^k)\right]^{1/2}$$

where the form $\sum g_{ik}x^i x^k$ is positive definite, then the a-dimensional Lebesgue measure, $|M|_a^e$, in A which results from this euclidean metric is invariant under translations so that

$$(1) \qquad \alpha(M) = f(A)|M|_a^e\,, \quad f(A) > 0\,.^2$$

Translation invariance implies that $f(A) = f(A')$ if A and A' are parallel a-flats, and the continuity of α implies continuity of $f(A)$. Because of the invariance under translation we may also write.

$$\alpha([x_0, \cdots, x_a]) = F(x_1 - x_0, \cdots, x_a - x_0)$$

where the function $F(x_1, \cdots, x_a)$ satisfies some simple conditions $F_1, \cdots F_4$ listed at the end of § 2.

We call the area α convex if

$$(2) \qquad F(x_1' + x_1'', x_2, \cdots, x_a) \le F(x_1', x_2, \cdots, x_a) + F(x_1'', x_2, \cdots, x_a)$$

and *strictly convex* if the strict inequality holds for independent x_1', x_1'', $x_2, \cdots, x_a$.

If an a-flat A and a b-flat B intersect in a d-flat D, where $0 \le d < \min(a, b)$, then they span a q-flat Q with $q = a + b - d$. We call B *totally transversal to* A, or A *totally normal to* B (at D in Q, where ambiguities are possible) if $\alpha(M) \le \alpha(M')$ for a projection[3] M parallel to B on A of any set M' which lies in an a-flat A' through D in Q. For $d = 0$, $b = n - a$ this is Caratheodory's concept of transversality[4]. If A is totally normal to B at D, $d > b + 1$, then A is totally normal to every b'-flat, $d < b' < b$ through D in B. We call A *normal to* B *at* D and B *transversal to* A, if A is totally normal to every $(d + 1)$-flat in B through D. For $d = 0$, $b = n - a$ this is Wagner's concept of transversality. Only for $d = \min(a, b) - 1$ does normality of A to B at D imply total normality. This is the only case with $d > 0$ which was studied previously in the literature, namely in [7] for Minkowski area.

We call α *totally convex* if an $(n - a)$-flat totally transversal to a given a-flat at a point exists. For totally convex α the a-flats minimize area in the sense that the α-area of the union of all but one face of a closed a-dimensional polyhedron is not less than the area of that face.

---

[2] Therefore the case $a = n$ is uninteresting as long as only areas for one definite A^n are considered. Hence we assume $1 \le a \le n - 1$ except in the last three sections.

[3] This concept needs clarification when $d > 0$. The precise form is found in § 2.

[4] Caratheodory treats more general a-dimensional variational problems. His ideas on transversality are easiest understood by consulting volume 1 of his Gesammelte Mathematische Schriften, München 1954; see in particular p. 364 and paper XX pp. 404–426.

However, the a-flats may minimize α-area for α which are not totally covex. On the other hand for $1 < a < n - 1$ the a-flats need not minimize area when α is merely convex. They will minimize a-area if α is *extendably convex* which means the following; α assigns an area $\phi(\mathfrak{a})$ to every simple a-vector, $\mathfrak{a}$, in the space V_a^n of all a-vectors, if $\phi(\mathfrak{a})$ can be extended to a convex function in all of V_a^n then α is extendably convex. The difference between extendable and total convexity has a very palpable interpretation in V_a^n.

If $F^2(x_1, \cdots, x_a)$ is a quadratic form in each set of variables $x_i^1, \cdots, x_i^n; i = 1, \cdots, a$; then we call $\alpha(M)$ *quadratic*. If $\alpha(M)$ is euclidean, that is if $\alpha(M) = |M|_a^e$ for a suitable choice of $e(x, y)$, then it is quadratic, but a quadratic area is not necessarily euclidean when $1 < a < n - 1$. The quadratic areas enter naturally as follows.

Let $0 \leq d < a \leq b < n$ and let a convex a-area α and a convex b-area β be defined in A^n. If normality (with respect to α) of an a-flat A to a b-flat B at a d-flat D is equivalent to normality (with respect to β) of B to A at D then both areas are quadratic unless $a + b = n, d = 0$. Whether the latter cases are really exceptional is not known except for $a = 1, b = n - 1$ (see below). If, in particular, $a = b$ and $\alpha \equiv \beta$, then equivalence of normality means that normality of two a-flats at a d-flat is a symmetric relation. Hence symmetry of normality implies—except for $a = n/2$, $d = 0$—that the area is quadratic. It will be euclidean only in special cases, for instance when $a < n/2$ and $d = 0$ or $a > n/2$ and $d = 2a - n$. For $a = b = 1, n > 2$ this becomes the above mentioned result of Blaschke [2].

All the results on symmetry and equivalence of normality also hold for total normality.

The a-dimensional Minkowski area (or measure), $2 \leq a \leq n$, in an n-dimensional Minkowski space with distance $F(x - y)$ is the area of the above type for which an a-dimensional unit ball in any a-flat A, that is the set $\{x \mid F(x - x_0) \leq 1; x, x_0 \in A\}$, has the euclidean volume $\pi^{a/2}/\Gamma(a/2 + 1)$. It is shown in [7] that these areas are convex and are strictly convex or differentiable if $F(x) = 1$ is strictly convex or differentiable.

We do not know whether Minkowski area is totally or extendably convex for $1 < a < n - 1$.

If the a-dimensional area $1 \leq a \leq n - 1$ of a Minkowski space is quadratic then the space is euclidean. Hence if normality of an a-flat A to a b-flat B at a d-flat D with respect to the a-area of one Minkowski space is equivalent to normality of B to A at D with respect to the b-area of another, then both Minkowski spaces are euclidean, unless $a + b = n, d = 0$. However only the case $a = 1, b = n - 1, d = 0$ is really known to be exceptional when the two spaces are different. When they are identical then already this case leads for $n > 2$ to an unsolved problem on convex bodies [10, Problem 5].

There are many interesting and difficult problems involving two areas in a Minkowski space of which we settle only a few. In the last section we obtain from the method and result of [8] a result of a different nature. If $b > a$ and $f_b(B), f_a(A)$ are the functions of (1) for a-and b-dimensional area of the same Minkowski space, we give an estimate from above for $f_b(B)$ in terms of $f_a(A)$ with $A \subset B$.

2. Normality. Our first objects are the relations between the various concepts of normality arising from different choices of d and b. In all that follows let $0 \leq d < \min(a, b); q = a + b - d \leq n$. Moreover, A, B, D, Q with or without subscripts denote a-, b-, d-, q-flats respectively with $D \subset B \subset Q, A \subset Q, A \cap B = D$.

Choose in B a c-flat $C, c = b - d$, which intersects D in exactly one point and hence intersects A in this point only. The association of the points of A and A_0 which lie in the same c-flat parallel to C is a projection of A_0 on A, which depends on the choice of C. The restriction of this mapping to a subset M_0 of A_0 gives the projection of M_0 on a set M in A.

If C' is a second c-flat in B which intersects D in a point, and B^* is any b-flat in Q parallel to B, then the projection of $B^* \cap A_0$ on A with the use of C' is the product of the projection of $B^* \cap A_0$ on A with the use of C and of a translation parallel to D (which depends continuously on B^*).

This and (1) imply.

(2.1) LEMMA. *If M_0 is a set in A_0 and M, M' are its projections on A with the use of C and C' respectively, then $\alpha(M) = \alpha(M')$.*

Thus the arbitrariness of C does not influence the measures of the projections. Moreover, if $0 < \alpha(M_0) < \infty$, then $\alpha(M)/\alpha(M_0)$ is according to (1) independent of the choice of M_0 in A_0.

We now define: A is totally *normal to B at D in Q*, or *B totally transversal to A at D in Q*, if for a fixed $M_0 \subset A_0$ with $0 < \alpha(M_0) < \infty$ and a fixed C the area $\alpha(M)$ of the projection of M_0 on A is minimal.

The preceding discussion shows that this definition is independent of the choice of A_0, M_0 and C; and hence depends only on D, B and Q.

The existence of an A normal to B at D in Q follows from two observations.

(i) The function $f(A)$ is continuous and has the same value for parallel A. Hence $f(A)$ attains its positive minimum f_1 and its finite maximum f_2 on the compact set of a-flats through z, so that

$$f_1|M|_a^e \leq \alpha(M) \leq f_2|M|_a^e .$$

(ii) $|M|_a^e \to \infty$ and hence $\alpha(M) \to \infty$ when A approaches a position for which $A \cap B$ is greater than D.

As previously observed, a B totally transversal to a given A at D in Q will in general fail to exist.

We now consider some properties of normality. In many of the following statements "totally" appears in parentheses, because they remain valid for the weaker concept of normality defined in the Introduction.

(2.2) *If A is* (totally) *normal to B at D in Q and the b'-flat B' lies in Q and contains B but does not contain A, then A is* (totally) *normal to B' at $D' = B' \cap A$ in Q.*

This is nearly obvious. A $(b-d)$-flat C in B which intersects D in exactly one point also intersects A and hence D' in this point only. Therefore the same C can be used for projection in both cases of normality.

(2.3) *If A is* (totally) *normal to B at D, $d<b'<b$, then A is* (totally) *normal to any b'-flat B' through D in B.*

Take a $(b'-d)$-flat C' in B' that intersects D in a point and choose a $(b-d)$-flat C in B which contains C' and intersects D in this point only. For any A' through D in the space spanned by A and B' the projection of A' on A parallel to B and B' respectively coincide if we use C and C'.

Proposition (2.3) implies in particular that A is totally normal to every $(d+1)$-flat through D in B. We shall see in § 5 that the converse is in general not true. It does hold in an important special case.

(2.4) THEOREM. *If $A \cap B = D$, $a = d+1$, $b-d \geq 2$ and A is normal to every $(d+1)$-flat in B through D, then A is totally normal to B at D.*

For an indirect proof, assume that A is not totally normal to B and let $\bar{A} \neq A$ be totally normal to B at D in the space Q spanned by A and B. A suitable b-flat B' in Q parallel to B intersects A and $\bar{A}$ in two distinct d-flats D' and $\bar{D}'$ parallel to D. These lie therefore in a $(d+1)$-flat $D_+ \subset B'$. In D^+ take a line L which intersects D' in a point.

Consider a set M in A with $0 < \alpha(M) < \infty$. Since A is normal to D^+, the projection $\bar{M}$ of M on $\bar{A}$ parallel to L satisfies $\alpha(\bar{M}) \geq \alpha(M)$.

On the other hand, let C' be a $(b-d)$-flat in B' which contains L and intersects D' in $L \cap D'$ only. Projection of M on $\bar{A}$ parallel to B with the use of C' again yields the set $\bar{M}$. Since $\bar{A}$ is totally normal to B and A is not, we would have $\alpha(M) > \alpha(\bar{M})$, a contradiction.

Defining normality of A to B at D as in the Introduction we conclude from (2.4) that normality and total normality coincide for $d = \min(a, b) - 1$. Obviously (2.3) remains valid for normalily instead of total normality. To prove (2.2) in this case we observe that a $(d'+1)$-flat E through D' in B' intersects B in a $(d+1)$-flat $F \supset D$. For $b' - b = d' - d$ and $E \cup B$ spans B' so that

$$\dim E \cap B + b' = \dim E + \dim B = d' + 1 + b = b' + d + 1 .$$

By hypothesis A is totally normal to F at D, by (2.2) it is also totally normal to E at D' and hence normal to B'.

Moreover (2.2) and (2.3) also show that the case $b = n - a$, $q = n$ is decisive in the following sense.

(2.5) *If an $(n-a)$-flat (totally) transversal to A exists, then for given $D \subset A \subset Q$, $q = a + b - d$, a b-flat (totally) transversal to A at D in Q exists.*

By hypothesis there is an $(n-a)$-flat N transversal to A through a point $p \in D$. By (2.2) A is normal to the $(n-a+d)$-flat B' spanned by D and N. This settles the case $q = n$. If $q < n$ then according to (2.3) A is normal to the b-flat $B = Q \cap B'$.

For later purposes we note the following consequence of (2.4) and (2.5).

(2.6) LEMMA. *A b-flat B transversal to A at D in Q for any given $D \subset A \subset Q$ will exist if and only if*

(i) *For $p \in A$, every $(a+1)$-flat through A contains a line transversal to A at p.*

(ii) *The set formed by the totality of all transversals to A at p in the different $(a+1)$-flats though A contains an $(n-a)$-flat N.*

The flat N is then transversal to A.

Also for later application we notice as a consequence of the continuity of $f(A)$ the following.

(2.7) LEMMA. *If $A_\nu \to A$, $D_\nu \to D$, $B_\nu \to B$ and A_ν is (totally) normal to B_ν at D_ν then A is (totally) normal to B at D.*

We follow these considerations up analytically using Barthel [1]. The invariance of $\alpha(M)$ under translation implies that the area of the box $[x_0, x_1, \cdots, x_a]$ has the form $F(x_1 - x_0, \cdots, x_a - x_0)$ and

$$(2.8) \qquad \begin{aligned} F(x_1, \cdots, x_a) &= F(x_1', \cdots, x_1^n, \cdots, x_a', \cdots, x_a^n) \\ &= f(A_x)|[z, x_1, \cdots, x_a]|_a^e , \end{aligned}$$

where A_x is the flat spanned by $x_1, \cdots, x_a$, if $x_1, \cdots, x_a$ are linearly independent and $F(x_1, \cdots, x_a) = 0$ otherwise. Thus $F(x_1, \cdots, x_a)$ has the following properties.

F_1: *$F(x_1, \cdots, x_a)$ is continuous in the $a \cdot n$ variables and symmetric in $x_1, \cdots, x_a$.*
F_2: *$F(x_1, \cdots, x_a) > 0$ if $x_1 \wedge \cdots \wedge x_a \neq 0$.*
F_3: *$F(\lambda x_1, x_2, \cdots, x_a) = |\lambda|\ F(x_1, \cdots, x_a)$.*
F_4: *$F(x_1 + \lambda x_j, x_2, \cdots, x_a) = F(x_1, \cdots, x_a)$ for $j > 1$.*

Conversely, if a function $F(x_1, \cdots, x_a)$ has the properties $F_1, \cdots, F_4$ then a well known argument (see e.g. [14, pp. 118, 124]) shows that $F(x_1, \cdots, x_a)$ has the form (2.8) with continuous $f(A_x)$ and vanishes for $x_1 \wedge \cdots \wedge x_a = 0$. Hence it defines an area function.

We now take definite independent vectors $u_1, \cdots, u_a$ and assume that $F(x_1, \cdots, x_a)$ possesses a differential as function of $x_1^1, \cdots, x_1^n$ at $x_i = u_i$. Then F_3 and F_4 yield for small $\lambda > 0$ and $j = 1, \cdots, a$

$$(2.9) \qquad \begin{aligned} \delta_j^1 \lambda F(u_1, \cdots, u_a) &= F(u_1 + \lambda u_j, u_2, \cdots, u_a) - F(u_1, \cdots, u_a) \\ &= \sum_i \frac{\partial F(u_1, \cdots, u_a)}{\partial x_1^i} \lambda u_j^i + o(\lambda) . \end{aligned}$$

For $\lambda \to 0$, using the symmetry of $F(x_1, \cdots, x_a)$ we obtain the following.

(2.10) *If $F(x_1, \cdots, x_a)$ possesses a differential as function of $x_k^1, \cdots, x_k^n$ at $u_1, \cdots, u_a$; $u_1 \wedge \cdots \wedge u_a \neq 0$, then*

$$(2.11) \qquad \sum_i \frac{\partial F(u_1, \cdots, u_a)}{\partial x_k^i} u_j^i = \delta_j^k F(u_1, \cdots, u_a) .$$

Let A be normal to B at D in Q, $z \in D$. Choose a non-degenerate q-box $[z, y_1, \cdots, y_b, u_{d+1}, \cdots, u_a]$ such that $y_1, \cdots, y_d$ lie in D; $y_{d+1}, \cdots, y_b$ in B and $u_{d+1}, \cdots, u_a$ in A. For any $\lambda_{d+1}, \cdots, \lambda_b$ the box $[z, y_1, \cdots, y_d, u_{d+1}, \cdots, u_a]$ originates from the box

$$\left[z, y_1, \cdots, y_d, u_{d+1} + \sum_{i=d+1}^{b} \lambda_i y_i, u_{d+2}, \cdots, u_a\right]$$

by projection parallel to B. If F possesses a differential at $y_1, \cdots, y_d$, $u_{d+1}, \cdots, u_a$ as function of $x_{d+1}', \cdots, x_{d+1}^n$, then normality of A to B implies that $F(y_1, \cdots, y_d, u_{d+1} + \sum \lambda_i y_i, u_{d+2}, \cdots, u_a)$ has a minimum for $\lambda_i = 0$. Hence

$$\sum \frac{\partial F(y_1, \cdots, y_d, u_{d+1}, \cdots, u_a)}{\partial x^i_{d+1}} y^i_j = 0 \qquad j = d+1, \cdots, b\,.$$

Thus we have found the following.

(2.12) *If the a-flat through z spanned by* $y_1, \cdots, y_d, u_{d+1}, \cdots, u_a$ *is normal to the b-flat B through z spanned by* $y_1, \cdots, y_b$ *and F is differentiable at* $y_1, \cdots, y_d, u_{d+1}, \cdots, u_a$ *as function of* $x^1_k, \cdots, x^n_k$ *for* $k = d+1, \cdots, a$ *then*

$$\sum_i \frac{\partial F(y_1, \cdots, y_d, u_{d+1}, \cdots, u_a)}{\partial x^i_k} y^i_j = 0\,, \qquad \begin{array}{l} k = d+1, \cdots, a \\ j = d+1, \cdots b\,. \end{array} \tag{2.13}$$

We conclude from (2.11) that the matrix $\partial F(y, u)/\partial x^i_k$ has rank $a - d$. Therefore, if D, A, Q are given there can be at most one b-flat transversal to A at D in Q. For brevity we say that $F(x_1, \cdots, x_a)$ is *individually differentiable* at $u_1, \cdots, u_a$ if it possesses a differential at $u_1, \cdots, u_a$ with respect to each of the sets of variables $x^1_k, \cdots, x^n_k$; $k = 1, \cdots, a$.

With property (ii) of (2.6) in mind we state explicitly the following consequence of our discussion.

(2.14) LEMMA. *If* $F(x_1, \cdots, x_a)$ *is individually differentiable at* $u_1, \cdots, u_a$ *with* $u_1 \wedge \cdots \wedge u_a \neq 0$, *and if in each* $(a+1)$*-flat containing the a-flat* A_u *spanned by* $u_1, \cdots, u_a$ *there exists a transversal to* A_u; *then this transversal is unique and the y corresponding to the different* $(a+1)$*-flats through* A_u *form the* $(n-a)$*-flat*

$$\sum_i \frac{\partial F(u_1, \cdots, u_a)}{\partial x^i_k} y^i = 0 \qquad k = 1, \cdots, a\,.$$

3. **Convexity.** Convexity, strict convexity and differentiability for the area α were determined in terms of the function $F(x_1, x_2, \cdots, x_a)$ in the introduction as follows.

(3.1) DEFINITION. Writing $F(y, x) = F(y, x_2, \cdots, x_a)$ we say that α is *convex*, *strictly convex*, or *differentiable* according as the curve $F(\lambda_1 y_1 + \lambda_2 y_2, x) = 1$ has those properties in the plane spanned by y_1, y_2 for any linearly independent $y_1, y_2, x_2, \cdots, x_a$.

Thus for convex α we have

$$F(y_1 + y_2, x) \leq F(y_1, x) + F(y_2, x) \text{ for } y_1 \wedge y_2 \wedge \cdots \wedge x_a \neq 0$$

with strict inequality for strict convexity. If we do not exclude linear dependence of y_1, y_2, then setting $y_1 = \mu y_2$ we have

$$F(y_1 + y_2, x) = |1 + \mu| F(y_1, x) \begin{cases} = F(y_1, x) + F(y_2, x) \text{ if } \mu \geq 0 \\ < F(y_1, x) + F(y_2, x) \text{ if } \mu < 0 \,. \end{cases}$$

Thus we find the following.

(3.2) LEMMA. *The area function α is convex if and only if*

$$F(y_1 + y_2, x_2, \cdots, x_a) \leq F(y_1, x_2, \cdots, x_a) + F(y_2, x_2, \cdots, x_a)$$

for $y_i \wedge x_2 \wedge \cdots \wedge x_a \neq 0$; and is strictly convex if and only if equality implies $y_1 = \mu y_2, \mu > 0$.

Let α be convex and $u_1 \wedge \cdots \wedge u_a \neq 0$. The function F has a differential with respect to $x_1^1, \cdots, x_1^n$ at $u_1, \cdots, u_a$ if and only if the curve

$$F(\lambda u_1 + \mu v, u_2, \cdots, u_a) = 1$$

is differentiable at $\lambda = 1, \mu = 0$ for all v with $v \wedge u_1 \wedge \cdots \wedge u_a \neq 0$. We have thus proved the following.

(3.3) *A convex area function α is differentiable if and only if the corresponding function $F(x_1, \cdots, x_a)$ is individually differentiable for $x_1 \wedge \cdots \wedge x_a \neq 0$.*

The differentiability properties of convex functions imply that for every convex α the corresponding F has strong differentiability properties, of which we need only the following.

(3.4) LEMMA. *If α is convex and $u_1, \wedge \cdots \wedge u_a \neq 0$, then there exist sequences $\{u_{i\nu}\}$ such that $u_{i\nu} \to u_i$ $(i = 1, \cdots, a)$ and such that $F(x_1, \cdots, x_a)$ is individually differentiable at $u_{i\nu}, \cdots, u_{a\nu}$.*

Reformulation of these properties in terms of the function $f(A)$ will prove useful. Since $f(A)$ is defined relative to a definite euclidean metric $e(x, y)$ we may use euclidean concepts. In particular we will speak of "perpendicularity" when we mean normality with respect to $e(x, y)$.

Consider a plane P perpendicular at z to the $(a - 1)$-flat L_{a-1} and choose in L_{a-1} an $(a - 1)$-box $[z, x_2, \cdots, x_a]$ with euclidean $(a - 1)$-volume 1. On each ray R in P with origin z choose y_R such that $F(y_R, x_2, \cdots, x_a) = 1$. The euclidean a-volume of this box is $e(z, y_R)$. Hence, if A_R is the a-flat containing R and L_{a-1} then

$$F(y_R, x_2, \cdots, x_a) = f(A_R) e(z, y_R) = 1 \,.$$

If the t-flat L, $2 \leq t \leq n - a + 1$ is perpendicular to L_{a-1} at z we denote by $S(L_{a-1}, L)$ the locus in L_t obtained by taking the point y_R

with $e(z, y_R) = f^{-1}(A_R)$ on a variable ray R in L_t with origin z. Then we can express our result as follows.

(3.5) LEMMA. *The area function α is convex, strictly convex, convex and differentiable if for any L_2, L_{a-1} and only if for all L_t, L_{a-1}, $2 \leq t \leq n - a + 1$ the surface $S(L_{a-1}, L_t)$ is convex, strictly convex, convex and differentiable.*

Following the arguments of [7] we now settle the case $d = \min(a, b) - 1$. The emphasis is not only on the result, but also on the *method* of constructing normal and transversal flats which the proof provides.

(3.6) THEOREM. *Let $d = \min(a,b) - 1, q = a + b - d \leq n$. For given d-, a-, q-flats $D \subset A \subset Q$, there exists a b-flat B transversal to A at D in Q if and only if the area function α is convex. B is unique when α is differentiable. The normal to B at D in Q is unique for all given $D \subset B \subset Q$ if and only if α is strictly convex.*

Proof. There are two cases.

CASE I: $a = d + 1, b = q - 1$. If $z \in D \subset Q$ are given we take the $(q - d)$-flat L_{q-d} perpendicular to D in Q at z and construct the surface $S = S(D, L_{q-d})$ of (3.5). An a-flat A through D in Q intersects L_{q-d} in two rays, each containing a point of S. Let y_A be one of these points. We claim that *B is transversal to A at D in Q if and only if it is spanned by D and a $(q - d - 1)$-flat through z parallel to a supporting flat H of S in L_{q-d} at y_A.*

The additional remarks on strict convexity and differentiability are then obvious. For if $H \cap S$ contains more points than y_A then the normal A to B at D in Q is not unique, and if S has two different supporting flats at y_A then B is not unique.

To prove our assertion we take A_1 perpendicular to B through D in Q, and in A_1 we take a set M_1 with $0 < \alpha(M_1) < \infty$. If we use $C = L_{q-d} \cap B$ to define projection parallel to B, then we have for the projection M of M_1 on any A

$$\alpha(M) = |M|_a^e f(A) = |M_1|_a^e |\sec(y_A z y_{A_1})| f(A) . \tag{3.7}$$

Therefore B is transversal to A if and only if $|\cos(y_A z y_{A_1})| f^{-1}(A)$ is maximal; or if and only if S has a supporting plane at y_A which is perpendicular to the ray from z through y_{A_1}, in other words is parallel to B.

The construction is easily freed from the intervening metric $e(x, y)$. Let $1 \leq a = d + 1 < q \leq n$ and let $z \in D \subset Q$ be given. Take a non-degenerate d-box $[z, x_1, \cdots, x_d]$ in D and a $(q - d)$-flat L_{q-d} in Q which intersects D at z only. In L_{q-d} construct the locus

$$S = \{x \mid F(x, x_1, \cdots, x_d) = 1\} .$$

Then the a-flat spanned by $x, x_1, \cdots, x_d$ with $x \in S$ is normal to the b-flat B in Q through D if and only if $B \cap L_{q-d}$ is parallel to a supporting $(q - d - 1)$-flat of S at x.

CASE II: $b = d + 1, a = q - 1$. As in Case I take L_{q-d} perpendicular to D at z in Q. Instead of using S we now take the line perpendicular to a variable a-flat A through D in Q. The two points y_A with $e(z, y_A) = f^{-1}(A)$ generate a locus T. When α is convex, strictly convex, convex and differentiable then T has the corresponding property.

This time we claim that B is transversal to A at D in Q if and only if it is spanned by D and the perpendicular to a supporting $(q - d - 1)$-flat of T in L_{q-d} at y_A. We define A_1 and M_1 as in Case I and use the line C perpendicular to A_1 at z for projection parallel to B. Then the projection M of M_1 on any A again satisfies (3.7) and $f^{-1}(A)|\cos(y_A z y_{A_1})|$ is maximal if and only if y_A lies on a supporting flat of T which is perpendicular to C. Since C is perpendicular to A_1 it lies in B. The additioned remarks follow as in Case I.

The definition of T cannot be entirely freed from extraneous concepts, but their role can be reduced.

If T is convex, let T' be the polar reciprocal in L_{q-d} of T with respect to the metric $e(x, y)$ (see [5, p. 28]). If T is strictly convex (differentiable) then T' is differentiable (strictly convex). In terms of T' we can interpret the normality relation in a manner similar to that of Case I; only the roles of normality and transversality are interchanged.

If $x \in T'$ then the $(d + 1)$-flat spanned by x and D is transversal to the a-flat A through D in Q if and only if A is spanned by D and a $(q - d - 1)$-flat parallel to a supporting flat of T' at x.

In the most interesting case, $d = 0$, the surface T' has a very interesting meaning. In $(Q = L_{q-d})$ take any $(q = a + 1)$-measure invariant under translation. The only arbitrariness is then the unit of measure. Then T' is a *solution of the isoperimetric problem* to minimize the α-area among all closed convex hyper-surfaces in Q which bound a set of given $(a + 1)$-measure. For details see [6]. Of course T' remains a solution even if we change the unit of $(a + 1)$-measure.

Assume that α is convex and consider an a-flat A_u through z spanned by $u_1, \cdots, u_a$ and such that F is individually differentiable at $u_1, \cdots, u_a$. Then (3.6) (more particularly Case II) guarantees that in every $(a + 1)$-flat containing A_u there exists a transversal to A_u at z. We conclude from (2.14) that the transversals at z to A_u in the different

$(a+1)$-flats form an $(n-a)$-flat N_A and from Theorem (2.6) that this N_A is transversal to A.

If F is not individually differentiable at $u_1, \cdots, u_a$ then we can find sequences $\{u_{i\nu}\}$ with $u_{i\nu} \to u_i$ $(i = 1, \cdots, a)$ such that F is individually differentiable at $u_{i\nu}, \cdots, u_{a\nu}$. Hence if A_ν contains $z, u_{i\nu}, \cdots, u_{a\nu}$ then there exists an $(n-a)$-flat N_ν transversal to A_ν at z. By the continuity of the area function every limit $(n-a)$-flat of a subsequence of N_ν is transversal to A. Thus if α is convex there exists an $(n-a)$-flat transversal to A. Using (2.14) and (2.5) we have proved

(3.8) THEOREM. *If the area function α is convex then, given an a-flat A, a d-flat $D \subset A$ and a q-flat $Q \supset A$ with $0 \leq d < a < q \leq n$; there exists a b-flat, $b = q - a + d$, transversal to A at D in Q, which is unique when α is differentiable.* (*Wagner* [15], *for* $d = 0$).

The conditions in (3.8) are also necessary, but we conclude from (2.5) and (3.6) that we need consider only fixed d and q.

(3.9) THEOREM. *With the notation of* (3.8); *if for fixed d, q and all A, D, Q a b-flat transversal to A at D in Q exists* (*and is unique*) *then α is convex* (*and differentiable*).

A normal to B at D in Q is in general not unique even for strictly and extendably convex α (as we shall see in (5.14)) when $d < \min(a, b) - 1$. For in that case normality is not equivalent to total normality. However, because total normals exist and are normal we have

(3.10) *If the a-flat A normal to B at D in Q is unique, then A is totally normal to B.*

Even the total normal is not necessarily unique for strictly and extendably convex α, see (5.14).

4. Area minimizing a-flats. Total and extendable convexity. The area $\alpha(\varDelta)$ of an a-dimensional polyhedron $\varDelta$ is defined as the sum of the a-areas of its a-faces. In the following we reserve $\varDelta$ for the union of all a-faces but one, $\varDelta_0$, of an a-dimensional polyhedron in A^n which is abstractly a closed orientable a-dimensional manifold but may have self interersections in A^n. By $A_\varDelta$ we denote the a-flat containing the face $\varDelta_0$ and hence the boundary of $\varDelta$.

We say that the a-flat A (strictly) minimizes α-area in the q-flat $Q \supset A$, $q > a$, if $\alpha(\varDelta) \geq \alpha(\varDelta_0)$ $(\alpha(\varDelta) > \alpha(\varDelta_0))$ for all choices of $\varDelta \neq \varDelta_0$ in Q for which $A_\varDelta = A$. If this is true for all a-flats A in Q we say that the a-flats (strictly) minimize area in Q.

The case $a = 1$ is familiar; with the help of (3.6) we may formulate these results as follows.

The line L minimizes α-length in the q-flat Q if and only if a $(q-1)$-flat B transversal to L in Q at a point z exists. The line L strictly minimizes length in Q if and only if L is the only line normal to B at z.

The lines (strictly) minimize α-length in A^n if and only if α is (strictly) convex.

A few of these facts extend to the general case.

(4.1) *The a-flat A minimizes α-area in Q if a $(q-a)$-flat B totally transversal to A at a point z exists. Let B exist. Then A strictly minimizes α-area when A is the only a-flat totally normal to B at z or when a is strictly convex.*

Project Δ on A_Δ parallel to B. For topological reasons this projection covers Δ_0. Let σ be an a-dimensional face of Δ which lies in the a-flat A and let σ_0 be its projection on A_Δ. If $\dim(B \cap A) > 0$ then obviously $0 = \alpha(\sigma_0) < \alpha(\sigma)$. If $\dim(B \cap A) = 0$ then the transversality of B to A_Δ implies $\alpha(\sigma_0) \leq \alpha(\sigma)$. This proves $\alpha(\Delta) \geq \alpha(\Delta_0)$.

If α is strictly convex and $\Delta \neq \Delta_0$ then $\alpha(\Delta) > \alpha(\Delta_0)$ is obvious when $\dim(B \cap A) > 0$ for some A containing an a-face of Δ. Assume therefore $\dim(B \cap A) = 0$ for all such A. There is at least one pair of a-faces σ^1, σ^2 of Δ which have a common $(a-1)$-face and at least one of which is not parallel to A_Δ. If A^i is the a-flat containing σ^i then not both A^1, A^2 can be normal to B. For, if A_i is the a-flat parallel to A^i through $A_\Delta \cap B$ then $\dim(A_1 \cap A_2) = a - 1$ and hence $A_1 \cup A_2$ spans an $(a+1)$-flat Q which intersects B in a line L through $A_\Delta \cap B$. Since α is strictly convex at least one of the two a-flats, say A_1, is by (2.3) and (3.6) not normal to B. Hence A' is not normal to B and $\alpha(\sigma') > \alpha(\sigma_0')$. Hence $\alpha(\Delta) > \alpha(\Delta_0)$.

If A is the only total normal to B at z then at least one a-face σ' of Δ is not totally normal to B and again $\alpha(\sigma') > \alpha(\sigma_0')$.

The case $q = a+1$ is completely known essentially through Minkowski (Theorie der konvexen Körper, § 27, Ges. Abh. 2, Leipzing 1911, 131–229). His terminology is so different that we give the argument here.

For each $(a+1)$-flat Q through z we construct the surface T_Q, analogous to T in the discussion of Case II in the proof of (3.6), as the locus T_Q of the points y_A with $e(z, y_A) = f^{-1}(A)$ on the perpendiculars to the a-flats A through z in Q.

(4.2) *The a-flat A minimizes α-area in the $(a+1)$-flat Q if and only if a line transversal to A in Q exists.*

A strictly minimizes area in Q if and only if a line transversal to in Q exists and y_A is not an interior point of an a-flat region on T_Q.

The sufficiency of the first part of (4.2) follows from (4.1) and the fact that a line transversal to A is totally transversal to A. We next prove the necessity statements in both parts of (4.2).

We choose rectangular coordinates such that Q is the flat $x^{a+2} = \cdots = x^n = 0$ and define, as usual,

$$H(0) = 0 \text{ for } x = 0, H(x) = |x| f(A_x) \text{ for } x \neq 0 ,$$

where A_x is the a-flat through z in Q with normal x and $|x| = (\sum x_i^2)^{1/2}$, so that T_Q has the equation $H(x) = 1$. The function $H(x)$ is convex with α.

If no transversal to A exists then, according to Case II in (3.6), T_Q does not possess a supporting a-flat at y_A; so that y_A is an interior point of the convex closure of T_Q. Hence independent points $x_1, \cdots, x_{a+1}$ on T_Q exist such that

$$H(y_A) > \sum_{i=1}^{a+1} \lambda_i H(x_j) , \qquad y_B = \sum_{i=1}^{a+1} \lambda_i x_i , \qquad \lambda_i > 0 . \tag{4.3}$$

If y_A is an interior point of an a-flat set on T_Q then independent $x_1, \cdots, x_{a+1}$ on T_Q exist with

$$H(y_A) = \sum_{i=1}^{a+1} \lambda_i H(x_i) , \qquad y_A = \sum_{i=1}^{a+1} \lambda_i x_i , \qquad \lambda_i > 0 . \tag{4.4}$$

Setting $\xi = y_A/|y_A|$, $\xi_i = x_i/|x_i|$ we have $-|y_A|\xi + \sum \lambda_i |x_i| \xi_i = 0$.

Therefore (see Bonnesen-Fenchel [5, p. 118]),[5] an $(a + 1)$-simplex in Q exists whose faces have exterior normals, $-\xi, \xi_1, \cdots, \xi_a$ and area $|x|, \lambda_1|x_1|, \cdots, \lambda_a|x|$. The total area of the faces with normals $\xi_1, \cdots, \xi_a$ is

$$\sum \lambda_i |x_i| f(A_{x_i}) = \sum \lambda_i H(x_i)$$

and $|x| f(A_x) = H(x)$ is the area of the face with normal $-\xi$.

The relations (4.3), (4.4) prove the necessity statements in (4.2).

To establish sufficiency in the second part of (4.2) we resume the notation used in the last part of the proof of (4.1). We assume that Δ lies in Q and replace B by a line L transversal to $A = A_\Delta$.

For $\alpha(\Delta) = \alpha(\Delta_0)$ it is necessary that the mapping of Δ on Δ_0 by projection parallel to L be one-to-one and that all a-flats carrying a-faces of Δ be normal to L.

Now there are two supporting flats A', A'' of T_Q perpendicular to L. On the other hand the construction of the transversal in the discussion of Case II in (3.6) shows that at the points y_A which corresponds to an A normal to L the surface T_Q has supporting planes perpendicular to

[5] The proof there is involved but becomes very simple in the present case where the number of faces is $a + 1$.

L. Therefore A' and A'' each contain one of the two points y_A and one of the two points y_A for each A which carries an a-face of Δ.

Since projection of Δ on Δ_0 is one-to-one and $\Delta \neq \Delta_0$ it follows that among the points y_A, y_A in A' there are $a+1$ which do not lie in an $(a-1)$-flat. These points span an a-simplex which lies on T_Q.

(4.5) COROLLARY. *The a-flats minimize α-area in all $(a+1)$-flats if and only if α is convex. They strictly minimize area if and only if in addition the surface T_Q contains no a-flat piece for any $(a+1)$-flat Q.*

Our results are not as complete for $q > a+1$, $a \neq 1$. Consider the vector space V_a^n of all contravariant a-vectors $\mathfrak{A}$ in A^n. A simple a-vector $\mathfrak{A} \neq 0$ determines an oriented a-flat in A^n through the origin. For the α-area determined by $\mathfrak{A}$ we obtain a function $\Phi(\mathfrak{A})$ defined on all simple $\mathfrak{A}$-vectors whose relation to F is given by

$$a!\, \Phi(x_1 \wedge \cdots \wedge x_a) = F(x_1, \cdots, x_a) .$$

Obviously Φ satisfies the conditions

$$\Phi_1 \qquad \Phi(\mathfrak{A}) > 0 \quad \text{for} \quad \mathfrak{A} \neq 0$$

$$\Phi_2 \qquad \Phi(\lambda\mathfrak{A}) = |\lambda|\, \Phi(\mathfrak{A}) \quad \text{for all real } \lambda .$$

All a-vectors are simple only when $a = 1$ and $a = n-1$. (If we exclude the trivial cases $a = 0, n$). We shall prove at the end of this section that for $1 < a < n-1$ and convex α it is in general impossible to extend $\Phi(\mathfrak{A})$ to a convex function defined for all a-vectors. An obviously necessary condition for extendability is

$$\Phi(\mathfrak{A}) \leq \sum_{i=1}^{r} \Phi(\mathfrak{A}_i) \quad \text{for simple } \mathfrak{A}, \mathfrak{A}_i, \text{ with } \quad \mathfrak{A} = \sum_{i=1}^{r} \mathfrak{A}_i . \tag{4.6}$$

Condition (4.6) is also sufficient. The simple a-vectors form a basis of V_a^n. Hence if $\mathfrak{A}$ is any a-vector then simple a-vectors $\mathfrak{A}_i$ exist so that

$$\mathfrak{A} = \sum_{i=1}^{r} \mathfrak{A}_i , \tag{4.7}$$

since any scalar multiple of a simple vector is simple. We can now extend $\Phi(\mathfrak{A})$ to all of V_a^n by defining

$$\Phi(\mathfrak{A}) = \inf \sum_{i=1}^{r} \Phi(\mathfrak{A}_i)$$

where the $\{\mathfrak{A}_i\}$ traverse all sets of simple vectors whose sum is $\mathfrak{A}$. Because of (4.6) $\Phi(\mathfrak{A})$ is not changed by this definition for simple $\mathfrak{A}$, and the extended function obviously is convex and satisfies Φ_1 and Φ_2.

We call α *extendably convex* if it satisfies (4.6). As before consider a polyhedron $\Delta \cup \Delta_0$. Orient it and let $\mathfrak{A}_1, \cdots, \mathfrak{A}_r$ be the simple a-vectors corresponding to the a-faces in Δ. Let $\mathfrak{A}_0$ correspond to Δ_0. Then

$$\sum_{i=0}^{r} \mathfrak{A}_i = 0 \quad \text{or} \quad \mathfrak{A} = -\mathfrak{A}_0 = \sum_{i=1}^{r} \mathfrak{A}_i$$

so that $\alpha(\Delta) \geq \alpha(\Delta_0)$ is equivalent to condition (4.6). In general the relation $\mathfrak{A} = \sum_{i=1}^{r}\mathfrak{A}_i$ for simple $\mathfrak{A}, \mathfrak{A}_i$ does not imply that $-\mathfrak{A}, \mathfrak{A}_1, \cdots, \mathfrak{A}_r$ correspond to the faces of a closed polyhedron. For example, the a-flats corresponding to $\mathfrak{A}, \mathfrak{A}_i$ through the origin z may intersect at z alone. However it is not unlikely that the validity of (4.6) for $\mathfrak{A}, \mathfrak{A}_i$ deriving from polyhedra implies its general validity. We have not been able to prove this. Thus we can only state:

(4.7) *If α is extendably convex then the a-flats minimize area.*

We call α *totally convex* if an $(n - a)$-flat totally transversal to a given a-flat at a point exists. If the condition in (4.7) is necessary then (4.1) shows that total convexity entails extendable convexity. We shall prove this directly, obtaining at the same time a very interesting geometric interpretation for the two types of convexity. The arguments are closely related to those of Wagner [15].

Denote by W_a the affine space associated with the vector space V_a^n, so that we may speak of hyperplanes etc. which do not pass through 0. The simple vectors in V_a^n form the *Grassmann cone* and the equation $\Phi(\mathfrak{A}) = 1$ defines on that cone the indicatrix I of the area α.

Extendable convexity of α means that I lies on the boundary of its convex closure in W_a; that is, that I possesses at every point a supporting hyperplane in W_a.

In order to interpret total convexity we provide A^n with the euclidean metric $g_{ik} = \delta_{ik}$. This metric induces a scalar product $\mathfrak{A} \cdot \mathfrak{B}$ for the simple a-vectors in A^n whose geomentric meaning, apart from sign, is the product of the (euclidean) area of one vector and the area of the orthogonal projection of the other on the a-flat of the first.

This scalar product for the vectors on the Grassmann cone can be extended to an inner product in V_a^n and hence induces a euclidean metric in W_a. To the projection of an a-flat A_1 on an a-flat A parallel to the $(n - a)$-flat B perpendicular to the a-flat B^* at a point there corresponds in V_a^n the projection of the line A_1 on the line A parallel to the hyperplane H_B perpendicular to the line B^*.

Assume now $\mathfrak{A} \in I$ and that I possesses at $\mathfrak{A}$ a simple supporting hyperplane H_B; that is a hyperplane H_B perpendicular to a line B^* on the Grassmann cone. If $\mathfrak{A}_1$ is a simple vector lying on H_B

(that is, interpreted in A^n, if $|\mathfrak{A}| = |\mathfrak{A}_1'|$ for the projection $\mathfrak{A}_1'$ of $\mathfrak{A}_1$ on the a-flat of $\mathfrak{A}$ parallel to the $(n-a)$-flat B which is perpendicular to B^*), then $\Phi(\mathfrak{A}_1) \geq \Phi(\mathfrak{A})$ since H_B is a supporting plane of I. Therefore B is totally transversal to A.

Conversely, if B is totally transversal to A at a point, then any simple $\mathfrak{A}_1$ whose projection parallel to H_B is $\mathfrak{A}$ satisfies $\Phi(\mathfrak{A}_1) \leq \Phi(\mathfrak{A}) = 1$, so that H_B is a supporting hyperplane of I. This could, of course, be formulated without the use of an auxiliary metric:

(4.10) *The area α is totally convex if and only if the indicatrix I posseses at every point $\mathfrak{A} = (a^\lambda)$ a simple supporting hyperplane $\sum a^\lambda b_\lambda = 1$, where $\mathfrak{B} = (b_\lambda)$ satisfies the conditions of a simple vector.*

If I is differentiable at $\mathfrak{A}$, so that the $a(n-a)$-flat, T, tangent to I at $\mathfrak{A}$ exists, then any supporting hyperplane of I at $\mathfrak{A}$ must pass through T. Through a given $a(n-a)$-flat there is exactly one simple hyperplane (see [15]). Since extendable convexity means only the existence of some supporting hyperplane of I at a given point we deduce from (4.10):

(4.11) *Total convexity implies extendable convexity but not conversely.*

That the converse is not valid does not follow from the preceding arguments, but in (5.13) we give an example of an extendably but not totally convex area.

We now show that convexity of α does not imply extendable convexity (Wagner [15] states this fact for $\min(a, n-a) > 2$; but, as it seems to us, he only proves that a certain definite extension of convex area is in general not convex). For this purpose we prove a lemma which seems to be of some independent interest.

(4.12) LEMMA. *Let S_a be a simple closed $(a-1)$-surface in an a-flat A so that at every point of S_a there is both an interior and an exterior supporting $(a-1)$-sphere of radius c in A. Let $z \in A$ be in the interior of S_a so that at the line zx from z to any $x \in S_a$ makes an angle no less than $\alpha > 0$ with the tangent $(a-1)$-flat of S_a at x.*

Then for every $\varepsilon > 0$ there exists a hypersurface $S \supset S_a$ such that every L_2 through z which contains a line that makes an angle greater than ε with A intersects S in convex curve.

Proof. For sufficiently small $\delta > 0$ the interior parallel surface S_a', which is the locus in the interior of S_a of points whose distance from S_a is δ, obviously satisfies the hypotheses of the lemma provided the constants c and α are replaced by suitable constants c' and α'. Let T_a' be the a-body bounded by S_a'.

Let S be the locus of points whose distance from T'_a is δ. Clearly $S_a \subset S$. Every $L_2 \ni z$ intersects S in a curve C. Assume that C is not convex; then there is an $x \in C$ at which C does not have a line of support in L_2 and therefore S does not have a plane of support at x. Thus the point x' nearest to x on T'_a must lie on S'_a and the line zx makes an angle less than $\tan^{-1}[d(x, x')/d(z, x')] \leq \tan^{-1}(\delta/d)$ with A, where d is the distance from z to S_a.

Now let L be the tangent line to C at x. Since L intersects the interior of C, the cylinder L_δ, which is the locus of points whose distance from L is δ, must intersect the interior of T'_a. Since the quadric $Q_\delta = L_\delta \cap A$ is tangent to S'_a at x' it follows that the minimal curvature of Q_δ at x' is less than $1/c'$. Let L' be the tangent line to Q_δ at x' in the direction of minimal curvature then the tangent of the angle between L and L' is less than $\sqrt{\delta/c'}$.

Thus for sufficiently small δ the two lines L and zx make arbitrarily small angles with the lines L' and zx' in A. Since the last named lines make an angle with each other which exceeds α' it follows that every line in L_2 makes an arbitrarily small angle with A.

Now, for example, in the space V_2^4 of 2-vectors in A^4 we can find a three-plane generated by simple vectors which contains no two-plane of simple vectors. Such a three-plans is L_3 generated by $e_1 \wedge e_2$, $e_3 \wedge e_4$ and $(e_1 + e_3) \wedge (e_2 + e_4)$. The simple vectors which it contains are all of the form $\lambda(e_1 + \mu e_3) \wedge (e_3 + \mu e_4)$. We can now define the area function F so that the indicatrix I does not lie on the boundary of its convex hull in L_3, for instance by $F(e_1, e_2) = F(e_3, e_4) = F(e_1 + e_3, e_2 + e_4) = 1$ and $F(e_1 + 2e_3, e_2 + e_2 + 2e_4) > 6$ in violation of (4.6); but so that $I \cap L_3$ satisfies all the conditions of Lemma (4.12) where z is the zero element of V_2^4. By Lemma (4.12) we can now extend I in such a way that its intersection with every two-plane of simple vectors is convex, in other words, so that F is convex. However, since I does not lie on the boundary of its convex hull, the area is not extendably convex.

5. Equivalence of normality. Example. Quadratic area. The normality relations determine the area up to a constant factor in the following sense.

(5.1) THEOREM. *Let α and α' be two a-dimensional convex area functions, $a + b - d \leq n$ and $d \leq \min(a, b) - 1$. For any d-flat D and any b-flat B through D let A be normal to B at D with respect to α' whenever this is the case with respect to α. Then $\alpha'(M)$ and $\alpha(M)$ differ only by a constant factor.*

The same holds for total normality if there exists a b-flat totally transversal with respect to α for any given a-flat at any given d-flat *in any given* $(a + b - d)$*-flat* (*in particular, when* α *is totally convex*).

Proof. Let $a = d + 1$. With the notation of Case I in (3.6) we construct the surfaces S, S' belonging to α and α' respectively. The hypothesis of (5.1) means in terms of S, S': If H and H' are parallel supporting $(q - d - 1)$-flats of S and S' then a line through z containing a point x of $S \cap H$ also contains a point of $S' \cap H'$. It folllows that S and S' are homothetic, and this conclusion remains valid when this condition on the line zx is assumed only for those $x \in S$ at which S is differentiable, that is H is unique.

This weakening of the hypothesis amounts to requiring that A be normal to B at D with respect to α' only when B is the unique transversal to A at D in $A \oplus B$ with respect to α.

The fact that S and S' are homothetic means that $\alpha'(M)/\alpha(M)$ is constant for all M lying in a-flats through a fixed $(a - 1)$-flat in an $(a + b - d)$-flat. This yields the general answer, because two arbitrary a-flats A', A'' can be joined by a finite number of a-flats $A_1 = A', A_2, \cdots, A_r = A''$ such that $\dim A_i \cap A_{i+1} = a - 1$ for $i = 1, \cdots, r - 1$,

Application of the result just obtained to the pencils determined by A_i and A_{i+1} proves the theorem.

The case $d < a - 1$ is reduced to $d = a - 1$ as follows. Let B^+, $\dim B^+ = b + a - d - 1$ be the unique transversal to A at an $(a - 1)$-flat D^+ in $A \oplus B^+$. In D^+ chose a d-flat D and an $(a - d - 1)$-flat E such that $D^+ = D \oplus E$. Then $D^+ = D \oplus E$ where B is a b-flat and $A \oplus B = A \oplus B^+$ because $E \subset A$.

For normality we know, and for total normality we assume, that a b-flat B' totally transversal to A at D in $A \oplus B^+$ exists. By (2.2) $B' \oplus E$ is transversal to A at D^+ in $A \oplus B^+$ and $B' \oplus E = B^+$ because B^+ is unique. By hypothesis B' is transversal to A at D with respect to α', and again by (2.2) B^+ is transversal to A at D^+ with respect to α'.

This means that the hypothesis of the theorem is satisfied for $d = a - 1$ and $b = a + b - d - 1$, so that the assertion follows from the first part of the proof.

Let $0 \leq d < a < n$. For a given a-area α we say that normality at d-flats is *symmetric*, if normality of an a-flat A to an a-flat A' at a d-flat D implies that A' is normal to A at D.

If $0 \leq d < a < n$ and an a-area α and a b-area β are given, we say that α-normality and β-normality at d-flats are *equivalent*, if normality of an a-flat A to a b-flat B at a d-flat D with respect to α implies that B is normal to A at D with respect to β and conversely, normality of B to A at D implies that A is normal to B at D.

This formulation admits the possibility that $a = b$. If at the same time $\alpha = \beta$ then equivalence means symmetry. If $a = b$ but $\alpha \neq \beta$ then equivalence means that normality in one norm is equivalent to transversality in the other.

Symmetry and equivalence of total normality are defined in the same way by replacing everywhere normality and transversality by total normality or transversality.

In the next section we discuss the implications of symmetry or equivalence of normality. Here we give some examples where these phenomena occur and the area is not euclidean.

(5.2) *For* $d = 0, a = 1, n = 2$ *symmetry of normality does not imply that the length,* i.e. *the corresponding two-dimensional Minkowski metric, is euclidean.* All these metrics have been determined by Radon [13], (see also [9, p. 104]).

(5.3) *For any* $(n-1)$*-dimensional convex area function* β *there is a convex one-dimensional area,* i.e. *a Minkowski metric* $F(x-y)$, *such that normality of a hyperplane to a line for* β *is equivalent to normality of the line to the hyperplane for* $F(x-y)$.

To see this we construct the surface T' of Case II of (3.6) for β and $d = 0$. That is, on the perpendicular to a variable hyperplane $B \ni z$ at z we take the two points y_B with $e(y_B, z) = f^{-1}(B)$. These points y_B traverse a convex hypersurface T and T' is the polar reciprocal of T. As Minkowski metric $F(x-y)$ we take the metric with T' as unit sphere $F(x) = 1$. Then the discussion under (3.6) shows that the hyperplanes normal (for β) to a line zw at $w \in T'$ are the supporting planes of T' at w and these are exactly the planes transversal to zw at w for $F(x-y)$.

The a-area α, $1 \leq a \leq n-1$ is euclidean if $\alpha(M) = |M|_a^e$ for a suitable choice of $e(x, y)$. With the summation convention $\sum_k g_{ik}x^k = g_{ik}x^k$ this means for F that

$$(5.4) \qquad F^2(x_1, \cdots, x_a) = \sum_{i_1 < \cdots < i_a} \begin{vmatrix} x_1^{i_1} & \cdots & x_1^{i_a} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_a^{i_1} & \cdots & x_a^{i_a} \end{vmatrix} \begin{vmatrix} g_{i_1k}x_1^k & \cdots & g_{i_ak}x_1^k \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ g_{i_1k}x_a^k & \cdots & g_{i_ak}x_a^k \end{vmatrix} .$$

We shall call α *quadratic* if F^2 is a quadratic form in each set of variables $x_i^1, \cdots, x_i^n$ $(i = 1, \cdots, a)$. A quadratic F^2 is a quadratic form in the Plücker coordinates.

$$P^{i_1 \cdots i_a} = \begin{vmatrix} x_1^{i_1} & \cdots & x_1^{i_a} \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ x_a^{i_1} & \cdots & x_a^{i_a} \end{vmatrix}, \qquad 1 \leq i_1 < \cdots < i_a \leq n ,$$

of the a-flat through z spanned by $x_1, \cdots, x_a$, since $F(x_1, \cdots, x_a) = f(A)|[z, x_1, \cdots, x_a]|_a^e$ where the terms on the right depend only on the Plücker coordinates.

If F is quadratic then for any L_{a-1} and L_2 perpendicular to L_{a-1} at z the curve $S(L_{a-1}, L_2)$ of (3.5) is an ellipse and conversely. If Q is any $(a+1)$-flat through z we construct in Q the surface T of Case II of (3.6) for $D = z$. The section of T with any plane $L_2 \ni z$ is obtained from $S(_{a-1}, L_2)$, where L_{a-1} is perpendicular at z to L_2 in Q, by a rotation through $\pi/2$. Hence T is an ellipsoid. This implies that the area restricted to Q is euclidean Thus we have the following.

(5.5) THEOREM. *An a-area is quadratic if and only if it is euclidean in every $(a+1)$-flat; that is to say, if and only if normality of a-flats at $(a-1)$-flats in $(a+1)$-flats is symmetric.*

We now wish to determine under what conditions a quadratic area is euclidean.

(5.6) *A quadratic a-area is euclidean if $a = 1$ or $a = n - 1$, and in general is not euclidean if $1 < a < n - 1$.*

The first part of the statement is obvious since a quadratic length is euclidean by definition and a quadratic $(n-1)$-area is euclidean in n-space by (5.5).

A simple counting argument convinces us of the truth of the second part since a euclidean quadratic area is determined by the metric (g_{ij}) so that the manifold of euclidean quadratic areas is $n(n+1)/2$-dimensional, while the manifold of Plücker coordinates is of dimension $1 + a(n-a)$; or, in other words, there are $\binom{n}{a} - a(n-a) - 1$ independent (quadratic) identities satisfied by the $P^{i_1 \cdots i_a}$ (see e.g. [2]). The distinct quadratic form in the Plücker coordinates therefore have dimension

$$\frac{1}{2}\binom{n}{a}\left[\binom{n}{a} + 1\right] - \binom{n}{a} + a(n-a) + 1$$

which exceeds $\binom{n+1}{2}$ whenever $1 < a < n - 1$.

If, for example, we restrict our attention to a-areas for which

$$F^2(x_1, \cdots, x_a) = \sum_{i_1 < \cdots < i_a} A_{i_1} \cdots {}_{i_a}(P^{i_1 \cdots i_a})^2$$

then no two different forms can be identical. Thus the dimension of this set is $\binom{n}{a}$ while the dimension of each equivalence class is no greater than $\binom{n+1}{2}$.

In particular, the Cartesian form

$$F^2(x_1, \cdots, x_n) = \sum (P^{i_1 \cdots i_a})^2 \tag{5.7}$$

is preserved only under orthogonal transformations. For, if we assume the existence of a matrix $g_{ij} \neq \delta_{ij}$ which preserves the form (5.7) then, for a suitable choice of Cartesian coordinates, we have $g_{ij} = A_i \delta_{ij}$ and (5.7) becomes

$$F^2(x_1, \cdots, x_n) = \sum A_{i_1} \cdots A_{i_a} (P^{i_1 \cdots i_a})^2 \tag{5.8}$$

Now, if (5.8) is Cartesian in one Cartesian coordinate system then it is Cartesian in all. Thus $A_{i_1} \cdots A_{i_a} = 1$ for all $i_1 < \cdots < i_a$. Since $n > a$ this implies $A_i = 1$ and $g_{ij} = \delta_{ij}$.

Since every euclidean area can be brought to Cartesian form we have also proved the following (which also follows from Theorem 9.1).

(5.9) *If two metrics g_{ij} and g'_{ij} give rise to identical a-areas, $a < n$, then $g_{ij} = g'_{ij}$.*

We can now determine the relations which suffice to make a quadratic a-area euclidean:

(5.10) THEOREM. *A quadratic a-area is euclidean if it is euclidean in every $(a + 2)$-flat.*

Proof. We proceed by induction. Assuming the area is euclidean is every m-flat, $m \geq a + 2$, we wish to prove it euclidean in every $(m + 1)$-flat. Let the $(m + 1)$-flat L_{m+1} have the equations $x^{m+2} = \cdots = x^n = 0$. Since the area is euclidean in every sub-flat $x^i = 0$ $(i = 1, \cdots, m + 1)$, there exists a matrix $g^{(i)}_{pq}$ $(1 \leq p, q \leq m + 1; p, q \neq i)$ so that the area function has the form (5.4) in this sub-flat. By (5.9) we have $g^{(i)}_{pq} = g^{(j)}_{pq}$ if $p, q \neq i, j$ since that is the unique metric in the common sub-flat $x^i = x^j = 0$. Thus there exists a matrix $g_{pq} = g^{(i)}_{pq}$ $(i \neq p, q)$ that defines a euclidean a-area in L_{m+1} which coincides with the given a-area in every coordinate sub-flat.

Without loss of generality we may assume the coordinates in L_{m+1} chosen so that $g_{pq} = \delta_{pq}$. Then on L_{m+1} we have

$$F^2(x_1, \cdots, x_a) = \sum (P^{i_1 \cdots i_a})^2 + R \tag{5.11}$$

where R involves the products of distinct Plücker coordinates so that every index $1, \cdots, m + 1$ appears in every product (if $m + 1 > 2a$ then there are no such terms and the proof is complete).

Consider the sub-flat $x^{m+1} = \lambda x^m$ of L_{m+1} and introduce the coordinates $y^i = x^i$ $(i = 1, \cdots, m - 1)$, $y^m = (1 + \lambda^2)^{-1} x^m$. In terms of these coordinates (5.11) becomes

(5.12) $$F^2(x_1, \cdots, x_a) = \sum (P^{i_1\cdots i_a})^2 + R'$$

where R' involves only products in which there appears every index $1, \cdots, m$. Now (5.12) is euclidean by hypothesis and the matrix g'_{ij} which represents it in the form (5.4) reduces to the identity matrix in every coordinate sub-flat. Hence $g'_{ij} = \delta_{ij}$ and $R' \equiv 0$. This means $R = 0$ in every sub-flat $x^{m+1} = \lambda x^m$, that is, $R \equiv 0$ so that (5.11) is euclidean.

The simplest case of an a-area with $1 < a < n - 1$, namely quadratic 2-area in A^4, already povides examples to show that:

(5.13) *For $1 < a < n - 1$ an extendably convex a-area need not be totally convex.*

For, denote the euclidean area in A^4 which belongs to $g_{ik} = \delta_{ik}$ by $E(x_1, x_2)$ and put $e_i = (\delta_{i1}, \cdots, \delta_{i4})$. For any $\varepsilon > 0$

$$F^2(x_1, x_2) = \varepsilon E^2(x_1, x_2) + \left(\begin{vmatrix} x_1^1 & x_1^2 \\ x_2^1 & x_2^2 \end{vmatrix} + \begin{vmatrix} x_1^3 & x_1^4 \\ x_2^3 & x_2^4 \end{vmatrix}\right)^2$$

defines a quadratic 2-area which obviously is extendably convex. The (x^1, x^2)-plane P^{12} is normal to every line in the (x^3, x^4)-plane P^{34}, because for arbitrary λ, μ, ρ we have

$$\begin{aligned} F^2(e_1 + \lambda e_3 + \mu e_4, e_2 + \rho\lambda e_3 + \rho\mu e_4) &\geq \varepsilon E^2(e_1, e_2) + \left(\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + \begin{vmatrix} \lambda & \mu \\ \rho\lambda & \rho\mu \end{vmatrix}\right)^2 \\ &= \varepsilon + 1 = F^2(e_1, e_2) . \end{aligned}$$

Thus P^{12} is normal to P^{34}. However, for small ε, P^{12} is not totally normal to P^{34}, since then

$$F^2(e_1 + e_3 + e_4/2, e_2 + e_3 - e_4/2) = \varepsilon E^2(e_1 + e_3 + e_4/2, e_2 + e_3 - e_4/2) < 1 + \varepsilon.$$

According to (3.10) the plane normal to P^{34} at z cannot be unique. Actually there is a one-parameter family of planes totally normal to P^{34} at z. To see this we observe that

$$\begin{aligned} &F^2(e_1 + \lambda e_3 + \mu_4, e_2 + \rho e_3 + \sigma e_4) \\ &\qquad = \varepsilon(1 + \lambda^2 + \mu^2 + \rho^2 + \sigma^2 + (\lambda\sigma - \mu\rho)^2) + (1 + \lambda\sigma - \mu\rho)^2. \end{aligned}$$

For a given ε with $0 < \varepsilon < 1$ this expression attains the minimal value $4\varepsilon/(1+\varepsilon)$ for $\lambda = -\sigma = \delta \cos\theta$, $\mu = \rho = \delta \sin\theta$ where $\delta = (1-\varepsilon)^{1/2}(1+\varepsilon)^{-1/2}$ and θ is arbitrary. Hence

(5.14) *If $1<a<n-1$ then extendable strict convexity of an a-area does not imply that the a-flat totally normal to an $(n-a)$-flat at a point is unique. More generally, the a-flat totally normal to a b-flat at a d-flat is not necessarily unique when $d < \min(a, b) - 1$.*

6. Equivalence of normality. Implications. Equivalence of normality for two convex areas implies for most combinations of the dimensions a, b, d that both areas are quadratic.

(6.1) THEOREM. *Let $0 \leq d < a \leq b < n$ $a+b-d \leq n$, but not $a+b=n$ and $d=0$. If a convex a-area α and a convex b-area β have the property that (total) α-normality and (total) β-normality at d-flats are equivalent, then both α and β are quadratic.*

(6.2) COROLLARY. *If for a convex a-area (total) normality at d-flats is symmetric then the area is quadratic unless $n = 2a$ and $d = 0$.*

We know from (5.3) that $a = 1, b = n - 1$ is actually exceptional but no examples are known for $a > 1$. (See note at end of paper).

The following proof is arranged so that only the existence of normals and not of transversals is used. Since the total normals exist, the proofs remain valid when normality is replaced everywhere by total normality.

Since normals and total normals do exist for non-convex areas, it is possible that (6.1) also holds without the assumption that α and β be convex. However the present proof uses convexity.

The hypothesis on the dimensions means that either (1) $a + b < n + d$ or (2) $a + b = n + d$ and $d > 0$. We consider the two cases separately.

In case (1) we show first (denoting an i-flat by L_i):

(A) Given[6] $L_{a-1} \subset L_{a+1} \subset L_{a+2}$ there exists an $L_a \subset L_{a+2}$ with $L_a \cap L_{a+1} = L_{a-1}$ such that the a-flats through L_{a-1} in L_{a+1} are normal to L_a.

(A′) The same as (A) with b replacing a.[6]

The proofs are entirely analogous with a slight simplification for (A) which we shall point out.

To prove (A) take $L_{n-b+a} \supset L_{a+1}$ with $L_{n-b+a} \cap L_{a+2} = L_{a+1}$, then take B normal to L_{n-b+a} at $D \subset L_{a-1}$. If $d + 1 < a$ choose the $(a-1-d)$-flat C such that $D \oplus C = L_{a-1}$. Since $B \oplus L_{n-b-a} = A^n$ we can find L_a with $L_{a-1} \subset L_a \subset B \oplus C$ and $L_a \oplus L_{a+1} = L_{a+2}$. (Here we can take $L_a \subset B$, but in the proof of (A′) there would exist no $L_b \subset A$ for $b > a$, whereas $L_b \subset A \oplus C$ exists because C is a $(b-1-d)$-flat and hence dim $A \oplus C = b-1-d+a \geq b$.) B is normal, hence by hypothesis transversal at D to any a-flat A' in L_{n-b+a} through D. If $L_{n-b+a} \supset A' \supset L_{a-1}$ then A' is normal to $B \oplus C$ at L_{a-1} and hence is normal to L_a at L_{a-1}.

We now show that (A) implies that α is quadratic. Let $z \in L_a \subset L_{a+2}$ and take L_3 through a perpendicular to L_{a-1} in L_{a+2}. Construct the surface $S = S(L_3, L_{a-1})$ of (3.5). It follows from the discussion of (3.6) Case I that for two lines G, H through z in L_3 the a-flat $G \oplus L_{a-1}$ is normal to $H \oplus L_{a-1}$ if and only if H is parallel to a supporting line of S at one of the two points $G \cap S$.

[6] $a + 2 \leq b + 2 \leq n$ since $b < n + d - a \leq n - 1$.

Now it follows from (A): Given L_2 through z in L_3 there exists in L_3 a $G \ni z$ such that for $z \in H \subset L_2$ the a-flat $H \oplus L_{a-1}$ is normal to $G \oplus L_{a-1}$. In terms of S this means that every intersection of S with a plane through z lies in some circumscribed cylinder of S.

A well known theorem of Blaschke [3] (see also [4, p. 157]) states that a closed convex surface S' in A^3 is an ellipsoid if every cylinder touches S' in a plane curve. Blaschke assumes that S' is differentiable but not that S' has a center. The differentiability hypothesis is very easily removed (see e.g. [9, p. 93]).

Under the hypothesis that S' has a center z the hypothesis may be relaxed in two ways.

(B_1) *S' is an ellipsoid when every plane section of S' through z lies on a circumscribed cylinder.*

(B_2) *S' is an ellipsoid when every circumscribed cylinder contains a plane section of S' through z.*

(B_1) is proved by a trivial modification of the proof of Blaschke's theorem and is also well known from the theory of Banach spaces.

The proof of (B_2) requires a less obvious but far from difficult modification of Blaschke's proof. (B_1) and (A) show that S is an ellipsoid. It follows that $S(L_{a-1}, L_{n-a+1})$ is also an ellipsoid (compare for example [9, p. 91]).

In the same way we deduce from (A′) and (B_1) that the surface $S(L_{b-1}, L_3)$ constructed with the area β is an ellipsoid so that β is also quadratic.

We now turn to the case $a + b = n + d$, $d > 0$ and prove:

(C) *Given $z \in L_{a-2} \subset L_{a-1} \subset L_{a+1}$ there is an a-flat A in L_{a+1} with $A \cap L_{a-1} = L_{a-2}$ such that the a-flats A_G in L_{a+1} through L_{a-1} are normal to A. The same holds with b replacing a.*

Take B normal to L_{a-1} at an $L_{d-1} \subset L_{a-2}$. Such a B exists because $a - 1 + b = n + d - 1$, moreover $L_{a-1} \oplus B = A^n$.

For any line G through z in B the a-flat $A_G = L_{a-1} \oplus G$ is transversal to B at $D_G = L_{d-1} \oplus G$. Hence A_G is, by hypothesis, normal to B at D_G. If $a > d + 1$ choose an $(a - d - 1)$-flat C through z in L_{a-1} such that $L_{d-1} \oplus C = L_{a-2} \subset L_{a-1}$. Then A_G is normal to $B \oplus C$ at $L^G_{a-1} = D_G \oplus C$. Let $z \in L_2 \subset B$, $L_2 \cap L_{d-1} = z, L_2 \oplus L_{d-1} = L_{d-1}$. This L_2 exists because $L_{a-1} \cap B = L_{d-1}$ and $L_{a-1} \oplus B = A^n$. Then $A = L_2 \oplus L_{a-2} \subset B \oplus C$ and for $z \in G \subset L_2$ the a-flat A_G is normal to A at L^G_{a-1}.

We now construct a surface T as in Case II of (3.6). On the line perpendicular to a given a-flat A' through z in L_{a+1} we take $y_{A'}$ with $e(z, y_{A'}) = f^{-1}(A')$. The points $y_{A'}$ traverse T.

Also, for a given L_{a-2} with $z \in L_{a-2} \subset L_{a+1}$ we take the L_3

perpendicular to L_{a-2} through z ($L_3 = L_{a+1}$ if $a = 2$). If $A' \supset L_{a-2}$ then the perpendicular to A' at z lies in L_3. The perpendiculars to the $A' \supset L_{a-2}$ therefore intersect T in a surface T_0 and it suffices to prove that T_0, or its polar reciprocal T_0' in L_3, is an ellipsoid.

According to the discussion of Case II the a-flat spanned by $x \in T_0'$ and L_{a-2} is transversal to the a-flat $A' \supset L_{a-2}$ if and only if A' is spanned by L_{a-2} and a plane L_2 through z parallel to a supporting plane of T_0' at x. Then A' is normal also to every a-flat in L_{a+1} through L_{a-2} and x.

Statement (C) means in terms of T_0', that given a line H through z ($H \oplus L_{a-2}$ is the L_{a-1} in the hypothesis of (C)) the cylinder parallel to H circumscribed to T_0' touches T_0' in a set containing a section of T_0' by a plane L_2 through z ($H \oplus L_{a-2}$ is the L_a in the assertion of (C)). It now follows from (B_2) that T_0' is an ellipsoid.

The proof that β is quadratic for $a + b = n + d, d > 0$ is again entirely analogous.

The Corollary (6.2) can be improved in special cases as follows:

(6.3) THEOREM. *If $a < n/2$ and $d = 0$ or $a > n/2$ and $d = 2a - n$ and for a (totally) convex a-area α (total) normality at d-flats is symmetric, then α is euclidean.*

The area function is differentiable because, according to (6.2), it is quadratic (in other respects the present proof is independent of (6.2)).

Let $a < n/2, A \ni z$ and let B_A be the $(n - a)$-flat transversal to A at z. Then each a-flat $A' \ni z$ in B_A is transversal to A. Hence by hypothesis A' is normal to A so that $B_{A'} \supset A$. Thus $A' \supset B_A$ implies $B_{A'} \supset A$. The mapping $A \to B_A$ can therefore be extended to a correlation φ on itself of the bundle consisting of all i-flats $(1 \leq i \leq n - 1)$ through z (see [2, pp. 51–53]). Moreover φ is a polarity because $A\varphi^2 = A$, and if $L_1 \ni z$ then $L_1\varphi$ does not contain L_1. Thus φ coincides with the mapping which belongs to a suitable ellipsoid E with center z which associates $L_1 \ni z$ with its diametral hyperplane $L_1\varphi$. This nearly obvious fact may be seen as follows.

We extend A^n to a projective space P^n and the correlation φ to a correlation of P^n by first associating $z = (0, \cdots, 0,1)$ with the hyperplane at infinity $H = (0, \cdots, 0,1)$. With the intersection $L_1 \cap H = (x_1, \cdots, x_n, 0)$ we associate the hyperplane $L_1\varphi = (\xi_1, \cdots, \xi_n, 0)$. If T is the (symmetric) matrix of $(x_1, \cdots, x_n, 0) \to (\xi_1, \cdots, \xi_n, 0)$ then $\begin{pmatrix} T & 0 \\ 0 & -1 \end{pmatrix}$ is the matrix of a polarity in P^n which defines the ellipsoid E with the above property.

This ellipsoid taken as unit sphere defines a euclidean metric in A^n and also a euclidean a-area. By construction normality of a-flats at z for this area coincides with α-normality of a-flats at z. According to

(5.1) this shows that the two areas differ only by a constant factor so that α is also euclidean.

The case $a > n/2$, $d = 2a - n$ is very similar. If $A \ni z$ we take the $(n-a)$-flat B_A transversal to A at z. This time if $A' \supset B_A$ then A' is transversal to A at $A' \cap A$ where $\dim A' \cap A = 2a - n$. By hypothesis A' is normal to A so that $B_{A'} \subset A$. Since $A' \supset B_A$ implies $B_{A'} \subset A$, the mapping $A \to B_A$ can again be extended to a correlation of the bundle of all i-flats $(1 \leq i \leq n-1)$ through z on itself. From here on the proof proceeds exactly as in the first case.

7. Minkowski area. We now apply our results to the special cases from which the general theory originated.

Consider a symmetric Minkowski metric (or a 1-dimension convex area) $F(x)$ in A^n. We denote its unit ball $F(x) \leq 1$ by U and let $U(A)$ denote the intersection of U with the a-flat A through z. For any a-flat $\bar{A}$ parallel to A the intersection $(F(x - \bar{z}) \leq 1) \cap \bar{A}$, $\bar{z} \in \bar{A}$ originates from $U(A)$ by translation and is a unit ball in $\bar{A}$ for the metric induced by $F(x)$ in $\bar{A}$. Following [7] we define an a-dimensional area $1 \leq a \leq n$ in A^n by stipulating that the measure of $U(A)$ have the euclidean volume

$$\pi_a = \pi^{a/2} / \Gamma\left(\frac{a}{2} + 1\right),$$

(in particular $\pi_1 = 2$, $\pi_2 = \pi$) so that for a definite euclidean metric e we have

$$f_a(\bar{A}) = f_a(A) = \pi_a / |U(A)|_a^e .$$

The functions corresponding to our previous $\alpha(M)$ and $F(x_1, \cdots, x_a)$ will be denoted by $|M|_a$ and $F_a(x_1, \cdots, x_a)$ so that

$$|M|_a = f_a(A)\,|M|_a^e ,$$

$$F_a(x_1, \cdots, x_a) = f_a(A)\,|[z, x_1, \cdots, x_a]|_a^e$$

and $F_1(x) = F(x)$. Since we admitted $a = n$ we also have an n-dimensional measure

$$|M|_n = f_n |M|_n^e = \pi_n |M|_n^e / |U|_n^e .$$

For $a < n$ let $L_{a-1} \ni z$ be an $(a-1)$-flat and L_2 the plane perpendicular to L_{a-1} at z. On a variable ray R with origin z in L_2 take the point y_R with

$$e(z, y_R) = f_a^{-1}(A_R) = |U(A_R)|_a^e \pi_a^{-1} , \qquad A_R = L_{a-1} \oplus R .$$

That is the curve $S(L_{a-1}, L_2)$ for f_a as constructed in (3.5). It is a fundamental and non-trivial fact (see [7, p. 164]) that $S(L_{a-1}, L_2)$ for f_a is always convex and is strictly convex or differentiable when the unit sphere $F(x) = 1$ of the space is strictly convex or differentiable respectively. Thus we have the following.

(7.1) THEOREM. *The Minkowski areas* $|M|_a, (1 \leq a \leq n-1)$ *are all convex. They are strictly convex or differentiable if the unit sphere* $F(x) = 1$ *is strictly convex or differentiable.*

The question whether Minkowski areas are totally convex for $1 < a < n - 1$ is equivalent to a difficult problem on convex bodies. Even extendable convexity is not known (see Problem 10 in [10]).

We mention the following further property of Minkowski area which is important for differential geometric investigations and was proved by Barthel [1].

(7.2) *If* $F(x)$ *is of class* C^r *for* $x \neq 0$ *then* $F_a(x_1, \cdots, x_a)$ *is of class* C^r *for* $x_1 \wedge \cdots \wedge x_a \neq 0$.

We also note:

(7.3) *If the a-area,* $1 \leq a \leq n-1$, *of a Minkowski space is quadratic then the space is euclidean.*

For, if $a > 1$ then we conclude from (6.3) that the area in any $(a + 1)$-dimensional subspace is euclidean. It is easily seen and contained in Theorem (9.1) that therefore the metric in this subspace is euclidean. It is well known (see e.g. [9, (16.12) p. 91]) that then the metric of the whole space is euclidean. Therefore (6.1) and (6.2) yield the following.

(7.4) THEOREM. *Let* $0 \leq d < a \leq b < n$ *but not* $a + b = n$ *and* $d = 0$. *If* α, β *are Minkowski a-and b-areas respectively (not necessarily relative to the same Minkowski metric) and* α*-normality and* β*-normality at d-flats are equivalent then both Minkowski metrics are euclidean.*

If normality of a-flats at d-flats in a Minkowski space is symmetric then the space is euclidean unless $a = n/2, d = 0$.

We note in particular that for all $n > 2$ symmetry of normality of a-flats at $(a - 1)$-flats suffices to make the Minkowski space euclidean. From (5.2) we know that the case $a = 1, b = n - 1, d = 0$ is exceptional for two distinct Minkowski metrics. Whether this case is exceptional when α and β belong to the same Minkowski space amounts (unless $a = b = 1$) to an interesting open problem on convex bodies (see [10, Problem 5]).

Finally we see from Theorem (7.4) and the example at the end of § 6 that convex area functions—even quadratic area functions—are in general not Minkowskian. The problem of characterization of Minkowski areas among the convex area functions remains open, (see § 9).

The fact that area functions are now defined for all a leads to new concepts, in particular to a sine function. If $A \cap B = D \ni z$, where $0 \leq d \leq \min(a, b) - 1$ and $A \oplus B = Q$, take a non-degenerate q-box, $[z, y_1, \cdots, y_b, x_{d+1}, \cdots, x_a]$, such that $y_1, \cdots, y_d \in D$; $y_{d+1}, \cdots, y_b \in B - D$ and $x_{d+1}, \cdots, x_a \in A - D$. Now put

$$(7.5) \qquad \operatorname{sm}(A, B) = \frac{F_d(y_1, \cdots, y_d)F_q(y_1, \cdots, y_b, x_{d+1}, \cdots, x_a)}{F_a(y_1, \cdots, y_d, x_{d+1}, \cdots, x_a)F_b(y_1, \cdots, y_b)} ,$$

where $F_0 = 1$. The number $\operatorname{sm}(A, B)$ is called the *Minkowski sine* of the flats A, B because it depends only on the latter and not on the choice of the q-box. For example, if $d > 0$ then replacing $y_1, \cdots, y_d$ by other independent $\bar{y}_1, \cdots, \bar{y}_d \in D$ amounts to multiplying all four terms F_d, F_q, F_a, F_b in (7.5) by

$$|[z, \bar{y}_1, \cdots, \bar{y}_d]|_a^e / |[z, y_1, \cdots, y_d]|_a^e .$$

If D does not contain z, but $\bar{z} \in D$ then the vectors y_i, x_j in (7.5) must be replaced by $y_i - \bar{z}, x_j - \bar{z}$.

The sine function is *not* the function of a number, "the angle between A and B". Even in the euclidean case this angle is defined *only* for $d = \min(a, b)$-1. Hence the restriction to this case in [7] and [1]. The sine function for the euclidean metric will be denoted by se. Then obviously, with $f(L_0) = 1$, we have

$$(7.6) \qquad \operatorname{sm}(A, B) = \operatorname{se}(A, B) f_d(D) f_q(0) f_a^{-1}(A) f_b^{-1}(B) .$$

For any λ_j^k, $k = d + 1, \cdots, b$; $j = d + 1, \cdots, a$ put

$$y_k(\lambda) = \sum_{j=d+1}^{b} \lambda_j^k y_j .$$

Then the boxes of the form $[z, y_1, \cdots, y_d, x_{d+1} + y_{d+1}(\lambda), \cdots, x_a + y_a(\lambda)]$ have $[z, y_1, \cdots, y_d, x_{d+1}, \cdots, x_a]$ as projection in Q parallel to B on A. Since

$$F_a(y_1, \cdots, y_b, x_{d+1} + y_{d+1}(\lambda), \cdots, x_a + y_a(\lambda))$$

does not depend on the λ_j^k, the a-flat A is totally normal to B at D in Q if and only if

$$\operatorname{sm}(A, B) \geq \operatorname{sm}(A^*, B) \text{ for } A^* \cap B = D, A^* \subset Q .$$

We denote this maximal value of $\operatorname{sm}(A^*, B)$ for given B, D, Q by $\alpha(B, D, Q)$. If $q = n$ then $Q = A^n$ is unique and we write simply $\alpha(B, D)$.

(7.7) *If*

$$\operatorname{sm}(A_1, B_1) = \max_A \alpha(A, D, Q),$$

then

$$\operatorname{sm}(A_1, B_1) = \max_B \alpha(B, D, Q),$$

and conversely, hence A_2 is totally normal to B_2 and B_2 is totally normal to A_2.

Proof. If A is normal to B then

$$\alpha(B, D, Q) = \operatorname{sm}(A, B) \leq \alpha(A, D, Q) \tag{7.8}$$

hence

$$\max_B \alpha(B, D, Q) \leq \max_A \alpha(A, D, Q) \tag{7.9}$$

Similarly, if B' is totally normal to A' then

$$\alpha(A', D, Q) = \operatorname{sm}(A', B') \leq \alpha(B', D, Q)\,. \tag{7.10}$$

Whence together with (7.9) we have

$$\max_B \alpha(B, D, Q) = \max_A \alpha(A, D, Q)\,. \tag{7.11}$$

If

$$\operatorname{sm}(A_1, B_1) = \max_A \alpha(A, D, Q) = \alpha(A_1, D, Q)$$

then B_1 is totally normal to A_1. Hence (7.11) and (7.10) imply

$$\operatorname{sm}(A_1, B_1) = \alpha(B_1, D, Q) = \max(B, D, Q)$$

so that A_1 is totally normal to B_1.

(7.12) *If for given A (B) in Q through D there exists a b-flat (a-flat) totally transversal to B (A) at D in Q (which is always the case for* $\min(a, b) = d + 1$) *and*

$$\operatorname{sm}(A_2, B_2) = \min_A \alpha(B, D, Q)$$

then

$$\operatorname{sm}(A_2, B_2) = \min_B \alpha(B, D, Q)$$

and A_2, B_2 are normal to each other.

For, if A is totally transversal to a given B, then

$$\alpha(A_2, D, Q) = \operatorname{sm}(A, B) \leq \alpha(B, D, Q) .$$

Hence

$$\min_A \alpha(B, D, Q) \leq \min_B \alpha(B, D, Q) .$$

The proof is analogous to that of (7.7).

As a consequence of (7.11) and (7.12) we have the following.

(7.13) COROLLARY. *If the function $\alpha(A, D, Q)$ is constant for fixed D, Q then $\alpha(B, D, Q)$ is constant and conversely. Moreover the constants have the same value. If $\alpha(A, D, Q)$ or $\alpha(B, D, Q)$ is constant then total normality of A to B and total normality of B to A are equivalent.*

The equivalence of total normality follows from the fact that for any A totally normal to B we have

$$\operatorname{sm}(A, B) = \max_A \alpha(A, D, Q) .$$

The equivalence of normality implies that B (A) totally transversal to A (B) at D in Q exist. Therefore both (7.11) and (7.12) apply.

Whether the converse of the second statement in (7.13) always holds is not known. However the proof of (3.6) yields the following special case.

(7.14) *If $d = \min(a, b) - 1$ and normality of A to B at D in Q is equivalent to normality of B to A, then $\alpha(A, D, Q)$ and $\alpha(B, D, Q)$ are constant.*

Proof. If $z \in D$ we take as in Case I of (3.6) the $(q-d)$-flat L_{q-d} perpendicular to D at z and construct, if $a \leq b$ say, the surface S by taking on each ray R in L_{q-d} with origin z the point y_R with $e(z, y_R) = f_a^{-1}(A_R)$ where $A_R = D \oplus R$.

For the b-area we construct T as in Case II by taking on the perpendicular in Q to a b-flat B through D in Q the two points y'_R with $e(z, y'_R) = f_b^{-1}(B)$, and denote by T' the polar reciprocal of T in L_{q-d} with respect to the metric $e(x, y)$.

If w_R is the point $R \cap T'$ then the supporting $(q-d-1)$-flat of T' at w_R spans together with the d-flat parallel to D through w_R a b-flat B normal to A_R. The reciprocity of T and T' implies that B has distance $f_b(B)$ from z. Hence by (7.6) we have[7]

[7] Because $d = \min(a, b) - 1$ the function se is the ordinary sine of the angle between A_R and B in the metric $e(x, y)$.

$$\alpha(A_R, D, Q) = \operatorname{sm}(A_R, B) = \operatorname{se}(A_R, B) f_d(D) f_q(Q) f_a^{-1}(A_R) f_b^{-1}(B) .$$

But

$$f_b(B) = \operatorname{se}(A_R, B) e(z, w_R), f_a(A_R) = e^{-1}(z, y_R) ,$$

so that we have the following nice *interpretation for* $\alpha(A_R, D, Q)$:

$$\alpha(A_R, D, Q) = f_d^{-1}(D) f_q^{-1}(Q) e(z, y_R)/e(z, w_R) .$$

If normality of A to B at D in Q is equivalent to that of B to A then S and T' are homothetic. Hence $e(z, y_R)/e(z, w_R)$ is constant, which proves (7.14).

8. The range of the sine functions. Problems regarding the ranges of $\alpha(B, D, Q)$ are important for Minkowskian geometry and are geometrically very attractive, but unfortunately often quite difficult—only in the simplest case $n = 2$ hence $a = b = 1, d = 0$ do we have complete answers owing to Petty [12] who found the following.

For any line L_1 *in* A^2 *through* z *we put* $\alpha(L_1, z, A^2) = \alpha(L)$ *and denote by* C_F *the unit circle* $F(x) = 1$. *Then*

$$\min_{L_1, F} \alpha(L_1) = \pi/4, \quad \max_{L_1, F} \alpha(L_1) = \pi/2 ,$$

and $\alpha(L_1) = \pi/4$ *or* $\alpha(L_1) = \pi/2$ *imply that* C_F *is a parallelogram and* L_1 *a suitable line (different in the two cases).*

Also

$$\max_F \min_{L_1} \alpha(L_1) = \pi/3 ,$$

where the maximum is attained only when C_F *is a hexagon which is regular for a suitable* $e(x, y)$.

Finally

$$\min_F \max_{L_1} \alpha(L_1) = 1 ,$$

where the minimum is attained only when C_F *is an ellipse, that is when the metric is euclidean.*

By (7.13) and (7.14) we have $\alpha(L_1) = k_F$, that is $\alpha(L_1)$ is independent of L_1, if and only if normality of lines in the plane is symmetric. This means that C_F is one of the curves discovered by Radon [13] which we encountered already several times implicitly and which we shall call Radon curves. Their construction is also found in Petty [12] and in [9, p. 104]. Since the regular hexagon is a Radon curve we find $1 \leq k_F \leq \pi/3$ with $k_F = 1$ only for the euclidean metric and $k_F = \pi/3$ only when C_F is a regular hexagon.

Under the hypothesis of (7.14), if $a = b$ and hence $d = a - 1$ then S and T' are Radon curves and we can derive the range of $\alpha(L_a, L_{a-1}, L_{a+1})$ (when constant) from Petty's results. Otherwise the ranges for $\alpha(A, D, Q)$ with D, Q fixed are not known. For variable D, Q we deduce from (7.13) and (7.14) the following.

(8.1) THEOREM. *If $0 \leq d < a \leq b < n$ but not $a + b = n$ and $d = 0$ then $\alpha(L_a, L_d, L_{a+b-d})$ is independent of L_a, L_d, L_{a+b-d} only in the euclidean geometry (where all α-functions are equal to* 1).

Beyond this result only very few facts on the ranges of the sine functions are known for $n > 2$, which we shall now discuss.

$$\min_{F, L_1} \alpha(L_1, L_0) = \min_{F, L_{n-1}} \alpha(L_{n-1}, L_0) = \pi_n / 2\pi_{n-1} \tag{8.2}$$

$$\max_{F, L_1} \alpha(L_1, L_0) = \max_{F, L_{n-1}} \alpha(L_{n-1}, L_0) = n\pi_n / 2\pi_{n-1} \,. \tag{8.3}$$

In the first of these relations equality is obtained only when the unit sphere S, that is $F(x) = 1$, is a *cylinder* and in the second only when S is a *double cone.*

The proof is very simple. The equality of the first two members in (8.2) or (8.3) follows from (7.12) and (7.7). Let H be a hyperplane through z and L_1 normal to H at z. If p, p' are the points $L_1 \cap S$ and $U_H = U \cap H$ then the hyperplanes parallel to H through p and p' are supporting planes of U. Moreover U_H has maximal $(n-1)$-dimensional volume among all sections U by hyperplanes parallel to H. Therefore

$$\pi_n = |U|_n \leq F(p - p')|U_H|_{n-1} \operatorname{sm}(L_1, H) = 2\pi_{n-1}\alpha(H, z)$$

with equality only for cylinders.

On the other hand U contains the double cone formed by the cones with apexes p, p' and bases U_H so that

$$\pi_n \geq n^{-1} 2\pi_{n-1}\alpha(H, z)$$

with equality only for double cones.

These relations successively provide bounds for all $\alpha(L_a, L_d)$, but these bounds are not sharp. We exemplify the procedure with $\alpha(L_{n-2}, L_0)$. If L_{n-2} is normal to L_2 at z then we consider in L_2 lines L_1' and L_1 through z such that L_1' is normal to L_1. Since L_{n-2} is normal to L_1 and L_1' we have, with $L_{n-1} = L_{n-2} \oplus L_1$,

$$\operatorname{sm}(L_{n-2}, L_2) \operatorname{sm}(L_1', L_1) = \operatorname{sm}(L_{n-2}, L_1) \operatorname{sm}(L_1', L_{n-1})$$

or

$$\begin{aligned} \operatorname{sm}(L_{n-2}, L_2) &= \alpha(L_1, z, L_{n-1}) \operatorname{sm}(L_1', L_{n-1})\alpha^{-1}(L_1, z, L_2) \\ &\leq \frac{n-1}{2} \frac{\pi_{n-1}}{\pi_{n-2}} \frac{n}{2} \frac{\pi_n}{\pi_{n-1}} \frac{4}{\pi} = \frac{n(n-1)}{\pi} \frac{\pi_n}{\pi_{n-2}} \,. \end{aligned}$$

It is easily seen that with a proper choice of L_1, L_1' in L_2 the line L_1' is normal to L_{n-1}. Hence

$$\operatorname{sm}(L_{n-2}, L_2) \geq \frac{1}{2}\frac{\pi_{n-1}}{\pi_{n-2}}\frac{1}{2}\frac{\pi_n}{\pi_{n-1}}\frac{2}{\pi} = \frac{1}{2\pi}\frac{\pi_n}{\pi_{n-2}},$$

so that

$$\frac{1}{2\pi}\frac{\pi_n}{\pi_{n-2}} \leq \alpha(L_{n-2}, L_0) \leq \frac{n(n-1)}{\pi}\frac{\pi_n}{\pi_{n-2}}.$$

The only exact bound other than (8.2) and (8.3) which has been determined is the following.

$$\max_{F, L_{n-1}} \alpha(L_{n-1}, L_{n-2}) = 2\pi_{n-2}\pi_n/\pi_{n-1}^2. \tag{8.4}$$

This equality holds only for a cylindrical unit sphere with $(n-2)$-dimensional generators and a parallelogram as 2-dimensional crosssection whose exact definition will emerge from the proof.

If an L_{n-2} is given we choose coordinates so that its equations are $x_{n-1} = x_n = 0$ and put $x_{n-1} = \rho\cos\varphi$, $x_n = \rho\sin\varphi$ so that $x_1, \cdots, x_{n-2}, \rho, \varphi$ are our coordinates. Set $U(L_{n-2}) = V$. For given x, φ with $x \in V$ let $(x, r(x, \varphi), \varphi)$ lie on the unit sphere S. Then, with $e^2(x, y) = \sum(x^i - y^i)^2$,

$$|U|_n^e = \frac{1}{2}\int_0^{2\pi}\int_V r^2(x, \varphi)dx\,d\varphi \geq \frac{1}{2|V|_{n-2}^e}\int_0^{2\pi}\left(\int_V r(x, \varphi)dx\right)^2 d\varphi$$

with equality only when $r(x, \varphi)$ is independent of x.

Now $\int_V r(x, \varphi_0)dx$ is the euclidean volume $A(\varphi_0)$ of the intersection of U with the half-hyperplane $\varphi = \varphi_0$. Hence if P_{φ_0} is the hyperplane containing $\varphi = \varphi_0$ we have

$$\begin{aligned}\operatorname{sm}(P_{\varphi_1}, P_{\varphi_2}) &= \frac{\sin|\varphi_1 - \varphi_2|2A(\varphi_1)2A(\varphi_2)}{|V|_{n-2}^e|U|_n^e}\cdot\frac{\pi_{n-2}\pi_n}{\pi_{n-1}^2}\\ &\leq \frac{4\sin|\varphi_1 - \varphi_2|A(\varphi_1)A(\varphi_2)}{(1/2)\int_0^{2\pi}A^2(\varphi)d\varphi}\cdot\frac{\pi_{n-2}\pi_n}{\pi_{n-1}^2}\end{aligned}$$

Considering the convex curve $\rho = A(\varphi)$ in $x_1 = \cdots = x_{n-2} = 0$ we see that the first factor on the right attains its maximum 2 when the curve is a parallelogram and φ_1, φ_2 fall in the diagonals. There will be equality in (8.4) if and only if in addition $r(x, \varphi)$ is independent of x. For $n = 3$ we have equality only for a parallelepiped.

The most important questions regarding the ranges of the sine functions concern

$$\min_F \max_{L_a} \alpha(L_a, L_d) = \min_F \max_{L_{n-a+d}} \alpha(L_{n-a+d}, L_d),$$

in particular whether, or for which a, d this number equals 1; and whether the value 1 characterizes euclidean geometry. The case $a = 1, d = 0$ is Problem 6 in [10].

9. Relations between the functions f_a. The Minkowski areas are derived from—and hence determined by—the Minkowski length. The question arises whether in a Minkowski geometry any of the areas $(1 < a < n)$ determine the remaining ones.

(9.1) THEOREM. *An a-dimensional area function $F_a(x_1, \cdots, x_a), 1 \leq a \leq n-1$, is an a-dimensional Minkowski area for at most one Minkowski geometry. In other words, if $F_a(x_1, \cdots, x_a)$ is known then $F(x)$ and hence the remaining $F_b(x_1, \cdots, x_b)$ are determined.*

This follows from a theorem of P. Funk [11]:

Let S_e be the sphere $e(z, x) = 1$ in B and let $S(A)$ be its intersection with $A \ni z$. Let $g_i(x), i = 1, 2$ be an even continuous function on S_e and denote by $S(A, g_i)$ the integral of $g_i(x)$ over $S(A)$ with respect to $(a-1)$-dimensional area. If $S(A, g_1) = S(A, g_2)$ for each A with $z \in A \subset B$ then $g_1(x) \equiv g_2(x)$.

Induction reduces this statement to $a = b - 1$.

A proof for $b = 3$ is found in [5, p. 138]. A proof for general b is obtained by using expansion in terms of spherical harmonics. If $x \in S_e$ then $xF^{-1}(x)$ lies on $F(x) = 1$. Hence $|U(A)|_a^e = S(A, a^{-1}F^{-a}(x))$ so that by Funk's theorem this relation determines $F(x)$.

An explicit expression of $F(x)$ in terms of $f_a(A)$ can be found in [4, pp. 154, 155], and this yields, in principle, the value $f_b(B)$ for given B. Actually the expression thus obtained is much too involved to deduce pertinent information from it. There is however an inequality of a very simple form, although its proof is involved, which relates f_b and f_a and which we are now going to derive from the results of [8].

If $n \geq b > a > 1, B \ni z$ then

$$D(b, a) f_b^{-a}(B) \geq \int_{B \supset A \ni z} f_a^{-b}(A) dA \tag{9.2}$$

with equality only for the ellipsoid. In this formula dA is the kinematic density for a-flats in B, the quantity $D(b, a)$ is the measure of all a-flats through z in B and hence is a constant which depends only on a and b.

Since in (9.2) B acts as the whole space we may take $b = n$. The inequality is a special case of a relation between the functions

$$f_{i,a}(A) = \pi_a / |U_i(A)|_a^e, f_{i,n} = \pi_n / |U_i|_n^e \qquad i = 1, \cdots, a$$

belonging to different Minkowski metrics with unit spheres $U_1, \cdots, {}_aU$

with common center z:

$$D(n, a)\prod_{i=1}^{a} f_{i,n}^{-1} \geq \int_{A\ni z}\prod_{i=1}^{a} f_{i,a}^{-n/a}(A)dA\ , \tag{9.3}$$

with equality only when the U_i are homothetic ellipsoids, i.e. when the corresponding Minkowski metrics are proportional euclidean metrics.

The inequality (9.3) is in turn a consequence of a still more general inequality.

Let $M_1, \cdots, M_a$ be convex bodies in the n-dimensional euclidean space $E^n, n \geq 3, 2 \leq a \leq n-1$ then

$$\begin{aligned} |M_1|_n \cdots |M_a|_n &\geq \pi_n^a \pi_a^{-n} D^{-1}(n, a) \\ &\times \int_{A\ni z} |M_1 \cap A|_a^{n/a} \cdots |M_a \cap A|_a^{n/a} dA\ , \end{aligned} \tag{9.4}$$

with equality for $|M_i|_n > 0$ only when the M_i are homothetic ellipsoids with center z. The measure $|M|_i$ is of course, the i-dimensional Lebesgue measure in E^n.

We deduce (9.4) from the following relation for any closed bounded sets $M_1, \cdots, M_a$.

$$\begin{aligned} |M_1|_n \cdots |M_a|_n &= C^1(n, a) \\ &\times \int_{A\ni z}\int_{M_1\cap A}\cdots\int_{M_a\cap A} T^{n-a}(P_1, \cdots, P_a, z)dV_{P_1}^a \cdots dV_{P_a}^a dA \end{aligned} \tag{9.5}$$

where $T(P_1, \cdots, P_a, z)$ is the a-dimensional measure of the (possibly degenerate) simplex with vertices $P_1, \cdots, P_a, z$ and $dV_{P_i}^a$ is the area element of A at $P_i \in M_i \cap A$. The symbol $C^i(n, a)$ denotes a constant which depends only on n and a.

For $a = n - 1$ and $a = n - 2$ (9.5) is proved in [8, (2), (17)], hence we prove (9.5) by induction for decreasing a. Assume (9.5) to hold for some $a + 1 \leq n - 1$. As M_{a+1} we take the euclidean unit ball U with center z. Then if B denotes an $(a + 1)$-flat we have

$$\begin{aligned} |M_1|_n \cdots |M_a|_n &= \pi_n^{-1} C^1(n, a+1) \\ &\times \int_{B\ni z}\int_{M_1\cap B}\cdots\int_{M_{a+1}\cap B} T^{n-a-1}(P_1, \cdots, P_{a+1}, z)dV_{P_1}^{a+1} \cdots dV_{P_{a+1}}^{a+1} dB\ . \end{aligned}$$

Now $M_{a+1} \cap B$ is an $(a + 1)$ – dimensional unit ball $\overline{U}$, and if φ is the angle between the a-flats spanned by $P_1, \cdots, P_a$ and the line through z and P_{a+1}, then

$$T(P_1, \cdots, P_{a+1}, z) = (a + 1)^{-1}e(z, P_{a+1})|\sin\varphi|\, T(P_1, \cdots, P_a, z)\ .$$

Since

$$\int_{\bar{U}} e^{n-a-1}(z, P_{a+1})|\sin^{n-a-1}\varphi|\,dV_{P+1}^{a+1}$$

depends only on n and a we obtain, after carrying out the integration ever U,

$$|M_1|_n \cdots |M_a|_n = C^2(n, a)$$
$$\times \int_{B\ni z}\int_{M_1\cap B}\cdots\int_{M_a\cap B} T^{n-a-1}(P_1, \cdots, P_a, z)\,dV_{P_1}^{a+1}\cdots dV_{P_{a+1}}^{a+1}dB\ .$$

For a variable a-flat A through z in B we have (see [8, (12)])

$$dV_{P_1}^{a+1}\cdots dV_{P_a}^{a+1} = a!\ T(P_1, \cdots, P_a, z)\,dV_{P_1}^{a}\cdots dV_{P_a}^{a}dA\ .$$

Integration first over all A through z in B, and then over all B through z can, according to the properties of kinematic measure, be interpreted as an integration over all A through z (except for a factor which depends only on a), and this proves (9.5).

Steiner's symmetrization leads from (9.5) to (9.4). Consider a fixed a-flat A through z and let $M_1, \cdots, M_a$ be convex bodies. It is shown in [8, pp. 8-10] that under simultaneous symmetrization of the sets $M_i \cap A$ in any $(a-1)$-flat C through z in A the integral

$$\int_{M_1\cap A}\cdots\int_{M_a\cap A} T^{n-a}(P_1, \cdots, P_a, z)\,dV_{P_1}^{a}\cdots dV_{P_a}^{a}$$

decreases unless the centers of all chords of all $M_i \cap A$ perpendicular to C are coplanar with z. Hence the $M_1 \cap A$ are homothetic ellipsoids with center z if the last integral is to be minimized. The minimum is actually attained for such ellipsoids [8, pp. 10, 11] and the integral has then the value

$$C^3(n, a)|M_1 \cap A|_a^{n/a} \cdots |M_a \cap A|_a^{a/a}\ .$$

This proves (9.4).

We note two consequences of these results.

(9.6) *The ellipsoids with center z maximize* $\int_{A\in z} |M \cap A|_a^n dA$ *among all convex bodies with a given volume.*

Application of (9.4) to the case $M_2 = \cdots = M_a = U$ yields

$$|M|_n \geq \pi_n \pi_a^{-n/a} D^{-1}(n, a)\int_{A\in z}|M \cap A|_a^{n/a}dA\ , \tag{9.7}$$

with equality only for the sphere. Hence the sphere gives the maximum of $\min_A |M \cap A|_a^n |M|_n^{a-n}$ for given volume $|M| > 0$.

References

1. W. Barthel, *Zum Inhaltsbegriff in der Minkowskischen Geometric*, Math. Z. **58** (1953), 358-375.

2. E. Bertini, *Einführung in die projektive Geometrice mehrdimensionaler Räume*, Wien 1924.

3. W. Blaschke, *Räumliche Variationsprobleme mit symmetrischer Transversalitätsbedingung*, Ber. Sächs. Akad. Wiss. Leipzig **68** (1916), 50-55.

4. ———, *Kreis and Kugel*, 2nd edition, Berlin 1956.

5. T. Bonnesen and W. Fenchel, *Theorie der konvexen Körper*, Ergebnisse Math. III 1, Berlin 1934 and New York 1958.

6. H. Busemann, *The isoperimetric problem for Minkowski area*, Amer. J. Math. **71** (1949), 743-762.

7. ———, *The foundations of Minkowskian geometry*, Comm. Math. Helv. **24** (1950), 156-186.

8. ———, *Volume in terms of concurrent cross-sections*, Pacific J. Math. **3** (1953), 1-12.

9. ———, *The geometry of geodesies*, New York 1955.

10. H. Busemann and C. M. Petty, *Problems on convex bodies*, Math. Scand. **4** (1956), 88-94.

11. P. Funk, *Über Flächen mit lauter geschlossenen geodätischen Linien*, Math. Ann. **74** (1913), 278-300.

12. C. M. Petty, *On the geometry of the Minkowski plane*, Riv. Mat. Univ. Parma **6** (1955), 269-292.

13. J. Radon, *Über eine besondere Art ebener konvexer Kurven*, Ber. Sächs. Akad. Wiss Leipzig **68** (1916), 131-134.

14. E. Sperner, *Einführung in die analytische Geometrie and Algebra I*, Göttingen 1948.

15. V. Wagner, *Geometry of a space with areal metric and its application to the calculus of variations*, (Russian), Mat. Sb. N. S. **19 (41)** (1946), 341-404.

UNIVERSITY OF SOUTHERN CALIFORNIA AND UNIVERSITY OF CALIFORNIA, LOS ANGELES

Note: While this paper was in print it was shown in H. Busemann: *Areas in affine space* II (to appear in the Rend. Circ. Mat. Palermo) that the case $a+b=n, d=1$ in (6.1) is exceptional for all a and also that $a=n/2,\ d=0$ in (6.2) is always exceptional.

CHARACTERIZATIONS OF TREE-LIKE CONTINUA

J. H. CASE AND R. E. CHAMBERLIN

1. Introduction. It has been conjectured by J. R. Isbell that every one dimensional continuum with trivial Čech homology (arbitrary coefficient group) is tree-like. In this note we give an example showing the conjecture is false. Moreover, the example has the Čech homology groups, the Čech cohomology groups, and the Čech fundamental group (see [3]) of a point. Also, the example cannot be mapped essentially onto a circle, but can be mapped essentially onto a "figure 8". We precede the example with two characterizations of tree-like continua.

2. Preliminaries. Throughout this note by a continuum we will mean a compact connected metric space and unless otherwise specified by a complex we will mean a finite complex. Also, by a linear graph we will mean a one dimensional connected complex.

For this section let X be any one dimensional continuum, K be any linear graph, and $\mathscr{C}(X)$ be the collection of all essential finite open covers of order two of X. For $U \in \mathscr{U} \in \mathscr{C}(X)$ let $\mathscr{N}(\mathscr{U})$ denote the nerve (see page 68 of [5]) of $\mathscr{U}$ and $\sigma(U)$ denote that vertex in $\mathscr{N}(\mathscr{U})$ corresponding to U. Note that for any $\mathscr{U} \in \mathscr{C}(X)$, $\mathscr{N}(\mathscr{U})$ is a linear graph. Where $\mathscr{U} \in \mathscr{C}(X)$ and $x \in X$ let $\Delta(\mathscr{U}, x)$ be the simplex in $\mathscr{N}(\mathscr{U})$ which has as vertices the collection of all $\sigma(U)$ such that $x \in U \in \mathscr{U}$. Where $\mathscr{U} \in \mathscr{C}(X)$, a continuous function f from X to $\mathscr{N}(\mathscr{U})$ is said to be a $\mathscr{U}$-canonical mapping provided that $f(x) \in \Delta(\mathscr{U}, x)$ for all $x \in X$. Where f is a continuous function from X to K, let $\mathscr{L}(f)$ be the collection of all non-empty inverse images under f of open stars of vertices in K. Note that $\mathscr{L}(f) \in \mathscr{C}(X)$. Where f is a continuous function from X to K let f' be that simplicial mapping from $\mathscr{N}(\mathscr{L}(f))$ to K which satisfies the condition.

$$f'(\sigma(f^{-1}\,[\text{open star of } v])) = v$$

for all vertices v in K such that

$$f^{-1}\,[\text{open star of } v]$$

is non-empty. Where $\mathscr{V}$ and $\mathscr{U}$ are two elements of $\mathscr{C}(X)$ such that $\mathscr{V}$ refines $\mathscr{U}$, a simplicial mapping p from $\mathscr{N}(\mathscr{V})$ to $\mathscr{N}(\mathscr{U})$ is said to be a projection if

$$p(\sigma(V)) = \sigma(U)$$

Received March 16, 1959.

implies that $V \subset U$ for all $V \in \mathscr{V}$ and $U \in \mathscr{U}$. If $\mathscr{V}$ refines $\mathscr{U}$ then there is always a projection from $\mathscr{N}(\mathscr{V})$ to $\mathscr{N}(\mathscr{U})$ (see page 135 of [5]).

P1. For any continuous function f from X to K there exists another, say g, which is homotopic to f and is such that $g[X]$ is a subcomplex of K.

This is proved by deforming the map f so as to uncover the interior of any simplex whose interior is not completely covered by f and keeping f fixed on the rest of the complex.

P2. For any $\mathscr{U} \in \mathscr{C}(X)$ there exists a $\mathscr{U}$-canonical mapping g from X to $\mathscr{N}(\mathscr{U})$ such that $g[X]$ is a subcomplex of $\mathscr{N}(\mathscr{U})$.

The existence of a $\mathscr{U}$-canonical mapping f from X to $\mathscr{N}(\mathscr{U})$ is established on page 286 of [2]. We use the method described for proving P1 to deform f to a mapping g such that $g[X]$ is a subcomplex of $\mathscr{N}(\mathscr{U})$. Under this natural deformation, g is $\mathscr{U}$-canonical.

P3. If $\mathscr{U} \in \mathscr{C}(X)$ and c is a $\mathscr{U}$-canonical mapping from X to $\mathscr{N}(\mathscr{U})$ then $\mathscr{L}(c)$ refines $\mathscr{U}$.

The proof is immediate from the fact that

$$c^{-1}\,[\text{open star of } \sigma(U)] \subset U$$

for any $U \in \mathscr{U}$.

P4. If f is a continuous function from X onto K then f' is a simplicial isomorphism from $\mathscr{N}(\mathscr{L}(f))$ onto K.

This is a special case of proposition D on page 69 of [5].

P5. If f is a continuous function from X onto K, $\mathscr{U} \in \mathscr{C}(X)$, $\mathscr{U}$ refines $\mathscr{L}(f)$, p is a projection from $\mathscr{N}(\mathscr{U})$ to $\mathscr{N}(\mathscr{L}(f))$, and c is a canonical mapping from X to $\mathscr{N}(\mathscr{U})$ then $f'\,pc$ is homotopic to f.

For any $x \in X$, let $S(x)$ be the smallest simplex in K which contains $f(x)$. It follows immediately from the definitions that $f'\,pc(x)$ and $f(x)$ are both in $S(x)$ for any $x \in X$. Therefore $f'\,pc$ is homotopic to f.

3. Two characterizations of tree-like continua. A one dimensional continuum X is said to be tree-like provided that every open cover of X can be refined by a finite open cover having nerve a tree, that is, having nerve a simply connected linear graph. A continuous mapping f from X to K is said to be inessential if it is homotopic to a constant map — otherwise it is said to be essential.

We shall prove the following two theorems simultaneously:

THEOREM 1. *A given one dimensional continuum X is tree-like if* *and only if every continuous mapping of X into any linear graph is inessential.*

THEOREM 2. *A given one dimensional continuum X is tree-like if and only if for every $\mathscr{U} \in \mathscr{C}(X)$ there exists an element $\mathscr{V}$ of $\mathscr{C}(X)$ which refines $\mathscr{U}$ and is such that any projection from $\mathscr{N}(\mathscr{V})$ to $\mathscr{N}(\mathscr{U})$ is inessential.*

Proof.

Part A. Suppose that X is tree-like and $\mathscr{U}$ is any element of $\mathscr{C}(X)$. Since X is tree-like we may take $\mathscr{V} \in \mathscr{C}(X)$ such that $\mathscr{V}$ refines $\mathscr{U}$ and $\mathscr{N}(\mathscr{V})$ is a tree. Since $\mathscr{N}(\mathscr{V})$ is a tree this nerve is contractible and any mapping (in particular any projection) from $\mathscr{N}(\mathscr{V})$ to $\mathscr{N}(\mathscr{U})$ is inessential.

Part B. Suppose that for any $\mathscr{U} \in \mathscr{C}(X)$ there exists $\mathscr{V} \in \mathscr{C}(X)$ such that $\mathscr{V}$ refines $\mathscr{U}$ and any projection from $\mathscr{N}(\mathscr{V})$ to $\mathscr{N}(\mathscr{U})$ is inessential. Let f be any continuous mapping of X into any linear graph K. In view of P1 we may assume that f is onto. Now we have that $\mathscr{L}(f) \in \mathscr{C}(X)$. Take $\mathscr{U} \in \mathscr{C}(X)$ such that $\mathscr{U}$ refines $\mathscr{L}(f)$ and any projection from $\mathscr{N}(\mathscr{U})$ to $\mathscr{N}(\mathscr{L}(f))$ is inessential. Let p be such a projection and let c be a canonical mapping from X to $\mathscr{N}(\mathscr{U})$.

By P5, the composite mapping $f'\,pc$ is homotopic to f. Since p is inessential so are $f'\,pc$ and f.

Part C. Suppose that every continuous mapping of X into any linear graph is inessential. Let $\mathscr{O}$ be any open cover of X. Since X is a one dimensional continuum we may take $\mathscr{U} \in \mathscr{C}(X)$ such that $\mathscr{U}$ refines $\mathscr{O}$.

Let c be a canonical mapping of X into $\mathscr{N}(\mathscr{U})$ such that $c[X]$ is a subcomplex of $\mathscr{N}(\mathscr{U})$. Let K be the universal covering space of $\mathscr{N}(\mathscr{U})$ with projection π. The space K is a complex (in general infinite) and π is simplicial. Since by our hypothesis c is inessential there exists a continuous mapping c^* from X to K such that $\pi c^* = c$. Let $T = c^*[X]$. Then T is a tree. By P4 the nerve $\mathscr{N}(\mathscr{L}(c^*))$ is isomorphic to T and hence is a tree. Also $\mathscr{L}(c^*)$ refines $\mathscr{L}(c)$ which refines $\mathscr{U}$ which refines $\mathscr{O}$ and therefore $\mathscr{L}(c^*)$ refines $\mathscr{O}$.

4. A group theoretic Lemma. The group theoretic situation discussed in this section is fundamental to the construction of the example in the following section.

Let G be a free non-Abelian group on two generators a and b. Let h be that endomorphism of G characterized by

$$h(a) = aba^{-1}b^{-1}$$

and

$$h(b) = a^2b^2a^{-2}b^{-2} .$$

Let Q be the set of all ordered pairs (α, n) such that n is a integer and α is either a or b. For any $(\alpha, n) \in Q$ let $e(\alpha, n) = \alpha^n$. Let S be the collection of all finite sequences $\{(\alpha_i, n_i)\}_{i=1}^{r}$ in Q such that

$$n_i \neq 0 \qquad \text{for } i = 1, \cdots, r$$

and

$$a_i \neq a_{i+1} \qquad \text{for } i = 1, \cdots, r-1$$

For each element g of G other than the identity there exists a unique $\{(\alpha_i, n_i)\}_{i=1}^{r}$ in S such that

$$g = \prod_{i=1}^{r} e(\alpha_i, n_i) .$$

In this case $\prod_{i=1}^{r} e(\alpha_i, n_i)$ is called the *preferred* representation of g and r is called the *length* of g.

LEMMA. *If g is an element of positive length r in G then the length of $h(g)$ is greater than or equal to $3r$.*

In order to prove the Lemma we will prove by induction the following somewhat stronger proposition:

($*$) If g is any element of G with preferred representation $\prod_{i=1}^{r} e(\alpha_i, n_i)$ and $h(g)$ has preferred representation $\prod_{i=1}^{s} e(\beta_i, m_i)$ then

(i) $s \geqq 3r$,

(ii) $\alpha_r = a$ and $n_r > 0$ imply $\beta_s = b$ and $m_s = -1$,

(iii) $\alpha_r = a$ and $n_r < 0$ imply $\beta_s = a$ and $m_s = -1$,

(iv) $\alpha_r = b$ and $n_r > 0$ imply $\beta_s = b$ and $m_s = -2$,

and

(v) $\alpha_r = b$ and $n_r < 0$ imply $\beta_s = a$ and $m_s = -2$.

Observing that

$$h(a^{-1}) = bab^{-1}a^{-1}$$

and

$$h(b^{-1}) = b^2a^2b^{-2}a^{-2}$$

proposition ($*$) is obviously true for $r = 1$.

Suppose that proposition ($*$) is true for the positive integer r. Let g be any element of G with length $r + 1$.

Say

$$g = \prod_{i=1}^{r} e(\alpha_i, n_i)\, e(\alpha, n)$$

and this is the preferred representation of g. Let $f = \prod_{i=1}^{r} e(\alpha_i, n_i)$. This is the preferred representation of f. Let

$$h(f) = \prod_{i=1}^{s} e(\beta_i, m_i)$$

be the preferred representation of $h(f)$. Note that

$$h(g) = h[f \cdot e(\alpha, n)] = h(f) \cdot h(\alpha, n)$$

$$= \prod_{i=1}^{s} e(\beta_i, m_i)\ h(\alpha, n)\ .$$

In order to conclude (i) – (v) of (*) we break the situation down into the following cases:

Case I $\alpha_r = a, n_r > 0,\ \alpha = b,$ and $n > 0$.
Case II. $\alpha_r = a, n_r > 0,\ \alpha = b,$ and $n < 0$.
Case III. $\alpha_r = a, n_r < 0,\ \alpha = b,$ and $n > 0$.
Case IV. $\alpha_r = a, n_r < 0,\ \alpha = b,$ and $n < 0$.
Case V. $\alpha_r = b, n_r > 0,\ \alpha = a,$ and $n > 0$.
Case VI. $\alpha_r = b, n_r\quad 0,\ \alpha = a,$ and $n < 0$.
Case VII. $\alpha_r = b, n_r < 0,\ \alpha = a,$ and $n > 0$.
Case VIII. $\alpha_r = b, n_r < 0,\ \alpha = a,$ and $n < 0$.

For convenience let k be the absolute value of n.
Case I. Define $\{\gamma_j, q_j\}_{j=1}^{4k}$ by

$$\begin{aligned} \gamma_{4i+1} &= a,\ q_{4i+1} = 2 \\ \gamma_{4i+2} &= b,\ q_{4i+2} = 2 \\ \gamma_{4i+3} &= a,\ q_{4i+3} = -2 \\ \gamma_{4i+4} &= b,\ q_{4i+4} = -2 \end{aligned}$$

for $i = 0, 1, 2, \cdots, k-1$. Then

$$\prod_{i=1}^{s} e(\beta_i, m_i) \prod_{j=1}^{4k} e(\gamma_j, q_j)$$

is the preferred representation of $h(g)$ and $h(g)$ has length $s + 4k \geqq 3r + 4k \geqq 3r + 3 \geqq 3(r+1)$.
Also

$$\gamma_{4k} = b \text{ and } q_{4k} = -2\ .$$

Case II. Define $\{\gamma_j, q_j\}_{j=2}^{4k}$ by

$$\begin{aligned} \gamma_{4i+1} &= b,\ q_{4i+1} = 2 \\ \gamma_{4i+2} &= a,\ q_{4i+2} = 2 \\ \gamma_{4i+3} &= b,\ q_{4i+3} = -2 \\ \gamma_{4i+4} &= a,\ q_{4i+4} = -2 \end{aligned}$$

for $i = 0, 1, 2, \cdots, k-1$ and $4i + 1 \neq 1$.

Then

$$\prod_{i=1}^{s-1} e(B_i, m_i)\ e(b, 1) \prod_{j=2}^{4k} e(\gamma_j, q_j)$$

is the prefered representation of $h(g)$ and $h(g)$ has lengh

$$s + 4k - 1 \geqq 3r + 4k - 1 \geqq 3r + 3 \geqq 3(r + 1)\ .$$

Also

$$\gamma_{4k} = a \text{ and } q_{4k} = -2\ .$$

These two cases are representative of all of them. In every case the length of $h(g)$ is either $s + 4k$ or $s + 4k - 1$ and hence is greater than or equal to $3(r + 1)$. Cancellation can occur in at most one place and that is where the terminal factor of $h(f)$ lies next to the initial factor of $h(\alpha, n)$. Conditions (ii) — (v) of (*) follow immediately.

5. An example. Let C be the collection of all complex numbers having modulus 1. Let

$$B = [C \times \{1\}] \cup [\{1\}] \times C]$$

and let $b_0 = (1, 1)$. In geometrical terms B is the union of two tangent circles and b_0 is the point of tangency. Define the function f from B to itself by the following formulas:

$$f(u, 1) = \left\{\begin{array}{ll} (u^4, 1) & \text{for } 0 \leqq \arg(u) \leqq \pi/2 \\ (1, u^4) & \text{for } \pi/2 \leqq \arg(u) \leqq \pi \\ (u^{-4}, 1) & \text{for } \pi \leqq \arg(u) \leqq 3\pi/2 \\ (1, u^{-4}) & \text{for } 3\pi/2 \leqq \arg(u) \leqq 2\pi \end{array}\right\}$$

and

$$f(1, v) = \left\{\begin{array}{ll} (v^8, 1) & \text{for } 0 \leqq \arg(v) \leqq \pi/2 \\ (1, v^8) & \text{for } \pi/2 \leqq \arg(v) \leqq \pi \\ (v^{-8}, 1) & \text{for } \pi \leqq \arg(v) \leqq 3\pi/2 \\ (1, v^{-8}) & \text{for } 3\pi/2 \leqq \arg(v) \leqq 2\pi \end{array}\right\}$$

for all $u, v \in C$ where u^{-1} and v^{-1} are the complex conjugates of u and v respectively. Note that f is continuous, onto, and takes b_0 into b_0.

We will represent the fundamental group $\pi(\mathrm{B}, b_0)$ of B with base point b_0 as the homotopy classes of continuous mappings σ from C to B such that $\sigma(1) = b_0$. Let a be that element of $\pi(B, b_0)$ represented by the continuous mapping α from C to B defined by

$$\alpha(u) = (u, 1)$$

for all $u \in C$. Let b be that element of $\pi(B, b_0)$ represented by the continuous mapping β from C to B defined by

$$\beta(u) = (1, u)$$

for all $u \in C$. It is well known that $\pi(B, b_0)$ is a free non-Abelian group on the two generators a and b. It follows immediately from the definition of the group operation on $\pi(B, b_0)$ that the natural induced endomorphism f^* of $\pi(B, b_0)$ satisfies the conditions

$$f^*(a) = aba^{-1}b^{-1}$$

and

$$f^*(b) = a^2b^2a^{-2}b^{-2} .$$

Therefore the group $\pi(B, b_0)$ and the endomorphism f^* of this group satisfy the hypothesis of the group theoretic lemma in the preceding section. It also follows that the induced endomorphisms $f_\#$ and $f^\#$ on the one dimensional homology group $H_1(B)$ and on the one dimensional cohomology group $H^1(B)$ respectively are the zero endomorphisms — no matter what coefficient group is used.

We define the space M to be the limit of the inverse system

$$B \xleftarrow{f} B \xleftarrow{f} B \xleftarrow{f} \cdots .$$

M may be described in a more elementary but more tedious way as the intersection of certain nest of closed tubular neighborhoods of "figure 8's" in three dimensional Euclidean space.

Q1. M is a continuum

This follows immediately from results in Chapter VIII of [2].

In establishing some of the other properties it will be convenient to give a more explicit definition of M. The set M is the collection of all sequences

$$x = \{x_i\}_{i=1}^{\infty}$$

of points in B such that

$$f(x_{i+1}) = x_i$$

for all i. For each i we define the projection π_i from M to B by the formula

$$\pi_i(x) = x_i$$

for all $x \in M$. The collection of all subsets of M of the form $\pi_i^{-1}[U]$ where i is any positive integer and U is any open subset of B from a basis for the topology of M. Moreover, the intersection of any two of these basic open sets is another. For all $i < j$ we let

$$\pi_i^j = f^{j-i}$$

be the mapping from B onto itself obtained by $j-i$ iterations of f. Let m_0 be that point in M defined by

$$\pi i(m_0) = b_0$$

for all i.

Q2. For each i, π_i is a continuous mapping of M onto B.
This follows from Corollary 3.9 on page 218 of [2].

Q3. For any open cover $\mathscr{U}$ of M there exists a positive integer j and a finite open cover $\mathscr{V}$ of B such that the collection of all $\pi_j^{-1}[V]$ for $V \in \mathscr{V}$ refines $\mathscr{U}$.

Proof. Let $\mathscr{U}$ be any open cover of M. Since M is compact $\mathscr{U}$ may be refined by a finite cover $\mathscr{B}$ of basic open sets. Let F be a finite set of pairs having first coordinate a positive integer and second coordinate an open subset of B such that $\mathscr{B}$ is the collection of all $\pi_i^{-1}[N]$ for $(i, N) \in F$. Let j be greater than any of the first coordinates of members of F. Let $\mathscr{V}$ be the collection of all $[\pi_i^j]^{-1}[N]$ for $(i, N) \in F$. Then for each $(i, N) \in F$ we have

$$[\pi_j^{-1}][\pi_i^j]^{-1}[N] = [\pi_i^j\pi_j]^{-1}[N] = \pi_i^{-1}[N] .$$

Therefore, $\mathscr{B}$ is the set of all $\pi_j^{-1}[V]$ for $V \in \mathscr{V}$ and $\mathscr{B}$ refines $\mathscr{U}$.

Q4. M is one dimensional.

Proof. Let $\mathscr{U}$ be any open cover of M. By Q3 we may take a positive integer j and a finite open cover $\mathscr{V}$ of B such that the collection of all $\pi_j^{-1}[V]$ for $V \in \mathscr{V}$ refines $\mathscr{U}$. Since B is a one dimensional continuum we may take $\mathscr{W}$ to be a finite open cover of B which refines $\mathscr{U}$ and is of order 2. Now the collection of all $\pi_j^{-1}[W]$ for $W \in \mathscr{W}$ is of order 2. The collection of all $\pi_j^{-1}[W]$ for $W \in \mathscr{W}$ is a finite open cover of M which refines $\mathscr{U}$ and is of order 2. Therefore, the continuum M has dimension less than or equal to one. We need only observe that it contains more than one point to see that it is one dimensional.

Since M is one dimensional all of the higher groups of M are trivial and we do not mention them further.

Q5. $H_1(M)$ and $H^1(M)$ are zero for an arbitrary coefficient group.

Proof. The Čech homology and Čech cohomology satisfy the continuity axiom (see Chapter X of [2]). Therefore, $H_1(M)$ is isomorphic to the limit of the inverse system

$$H_1(B) \xleftarrow{f^\#} H_1(B) \xleftarrow{f^\#} H_1(B) \xleftarrow{f^\#} \cdots\cdots$$

and $H^1(B)$ is isomorphic to the limit of the direct system

$$H^1(B) \xrightarrow{f^\#} H^1(B) \xrightarrow{f^\#} H^1(B) \xrightarrow{f^\#} \cdots$$

We have already observed that both f and f are the zero homomorphisms. Therefore both $H_1(M)$ and $H^1(M)$ are the zero groups.

Q6. M cannot be mapped essentially onto the circle.

Proof. Making use of the fact that M is a one dimensional compact space and $H^1(M)$ with integer coefficients is zero, we see that Q6 follows immediately from the corollary on page 150 in [5] to Hopf's extension theorem.

Q7. The Čech fundamental group $\pi(M, m_0)$ of M with base point m_0 is zero.

Proof. The Čech fundamental group also satisfies the continuity axiom (see [3]) and agrees with the usual fundamental group on complexes. Therefore $\pi(M, m_0)$ is isomorphic to the limit of the inverse system.

$$\pi(B, b_0) \xleftarrow{f^*} \pi(B,b_0) \xleftarrow{f^*} \pi(B, b_0) \xleftarrow{f^*} \cdots$$

of non-Abelian groups and homomorphisms. We now apply the group theoretic lemma of the previous section. Suppose there is an element g other than the identity in this inverse limit. Then $g = \{g_i\}_{i=1}^{\infty}$ where for each i

$$g_i \in \pi(B, b_0)$$

and $f^*(g_{i+1}) = g_i$. Moreover we may take n such that g_i is not equal to the identity for all $i \geqq n$. Therefore for any $i > n$, g_i has positive length,

$$(f^*)^{i-n}\,(g_i) = g_n,$$

and g_n has length greater than or equal to $3(i-n)$. This says that g_n has infinite length, a contradiction.

We have yet to establish that M is not tree-like. For this purpose we make use of Theorem 1 although we could just as well use Theorem 2.

We construct an essential mapping q of M onto B. For each i, let

f^i be the mapping from B to itself obtained by i iterations of f. Note that for all $i < j$

$$f^i \pi_i^j = f^j .$$

Define the mapping q from M to B by

$$q(x) = f(x_1) .$$

Actually $q = f\pi_1$, but for clarity we use this different notation. Note that for any i

$$q(x) = f^i(x_i) .$$

Q8. For every positive integer i the mapping f^i from B into itself is essential.

Proof. According to the algebraic lemma of the preceding section the endomorphism f^* of $\pi(B, b_0)$ is an automorphism. Therefore, the endomorphism $(f^i)^*$ which equals $(f^*)^i$ is also an automorphism. Now since the group $\pi(B, b_0)$ is not zero, the automorphism $(f^i)^*$ is not zero and the mapping f^i is essential.

Q9. The continuous mapping q from M onto the "figure 8" B is essential and therefore M is not tree-like.

Proof. By way of contradiction, suppose that q is inessential. Let $\tilde{B}$ be the universal covering space of B with projection p. Since the mapping q from M to B is inessential there exists a continuous mapping q^* from M to $\tilde{B}$ such that $pq^* = q$.

Let $\mathscr{E}$ be an open cover of $q^*[M]$ by open sets of $\tilde{B}$ such that for any $E \in \mathscr{E}$ the mapping p_E obtained by restricting p to E is a homeomorphism of E onto the open set $p[E]$ in B. Let $\mathscr{U}$ be the collection of all inverse images under q^* of elements of $\mathscr{E}$. According to Q3 we may take a positive integer i and an open cover $\mathscr{V}$ of B such that the collection of all $\pi_i^{-1}[V]$ for $V \in \mathscr{V}$ refines $\mathscr{U}$.

Define the function c from B to $\tilde{B}$ as follows:

$$c(x) = q^*(z)$$

for any $x \in B$ and $z \in \pi_i^{-1}[x]$.
Clearly

$$pc(x) = pq^*(z) = q(z) = f^i(z_i) = f^i(x)$$

for any $x \in B$ and $z \in \pi_i^{-1}[x]$. Therefore, if c is a continuous function it will cover f^i.

To show that c is single valued, let x be any point in B and let $z, z' \in \pi_i^{-1}[x]$.
Then

$$q(z) = q(z') = f^i(x) .$$

Therefore $pq^*(z) = pq^*(z')$. By the choice of i and $\mathscr{V}$ we know that there exists $E \in \mathscr{E}$ such that

$$q^*[\pi_i^{-1}[x]] \subset E .$$

Therefore $q^*(z), q^*(z') \in E$. We also know that p restricted to E is a homeomorphism. Therefore $q^*(z) = q^*(z')$ and we have that c as defined above is single valued.

It is immediate from the fact that $\pi_i[M] = B$ that the domain of c is B.

In order to show continuity note that for any $x \in V \in \mathscr{V}$,

$$c(x) = (p_E)^{-1} f^i(x)$$

where E is an element of $\mathscr{E}$ such that

$$q^*[\pi_i^{-1}[V]] \subset E .$$

Therefore c is continuous on each member of $\mathscr{V}$, an open cover of B, and therefore c is continuous on all of B.

Now we have lifted the map f^i from B to B to the map c from B to the universal covering space of B. Since B is a linear graph the sub-continuum $c[B]$ of $\tilde{B}$ is contractible and hence c is inessential. Since $pc = f^i$ the map f^i is inessential, a contradiction of Q8.

Further remarks. Theorems 1 and 2 give us two conditions each of which is equivalent to saying that a given one dimensional continuum X is tree-like. We list without proof some other likely characterizations:
(1) X has no non-trivial connected generalized covering space.
(2) X cannot be mapped essentially into the Universal one dimensional curve.
(3) X is an inverse limit of 2-cells.
Condition (3) leads us to stating another question. First let us say that a continuum X is disk-like if it is an inverse limit of 2-cells or equivalently if every open cover of X can be refined by one which has nerve a disk. The tree-like continua are precisely the one dimensional disk-like continua. Also, those disk-like continua that can be imbedded in the plane are precisely the continua which can be imbedded in the plane and do not separate the plane.

Question. Does every disk-like continuum have the fixed point property?

Obviously an affirmative answer to this problem would give affirmative answers to the fixed point problem for tree-like continua and for those sub-continua of the plane which do not separate it.

The continuum M described in this paper gives further insight into the difficulties of generalizing the definition of the fundamental group. We may conclude that any generalization of the fundamental group, which agrees with the usual fundamental group on complexes, and which also satisfies the continuity axiom cannot distinguish the tree-like continua from the other one-dimensional continua. This difficulty will be described more explicitly in another paper which will include the verification of condition (1) of this section.

The referee pointed out our lack of reference to the known results on the fixed point problem for continua which are inverse limits of n-cells with $n \neq 2$. We remark that snake-like continua (those which are inverse limits of 1-cells) have the fixed point property (see [4]) but there exist cube-like continua (those which are inverse limits of 3-cells) which do not have the fixed point property (see [1]).

REFERENCES

1. K. Borsuk, *Sur un continu acyclique qui se laisse transformer topologiquement en lui meme sans points invariants*, Fund. Math., **24** (1935), 51-58.

2. S. Eilenberg and N. Steenrod, *Fundations of Algebraic Topology*, Princeton University Press, 1952.

3. A. Goldman, *A Čech fundamental group*, Proceedings of the Summer Institute on Set Theoretic Topology, Madison, Wisconsin, 1955.

4. O. H. Hamilton, *A fixed point theorem for pseudo-arcs and certain other metric continua*, Proc. Amer. Math. Soc., **2** (1951), 173-174.

5. W. Hurewicz and H. Wallman, *Dimension Theory*, Princeton University Press, 1948.

UNIVERSITY OF UTAH

CHARACTERISTIC SUBGROUPS OF MONOMIAL GROUPS

R. B. CROUCH

1. Introduction. Let U be a set, $o(U) = B = \aleph'_u$, $u \geqq 0$, where $o(U)$ means the number of elements of U. Let H be a fixed group. A monomial substitution y is a transformation that maps every x of U in a one-to-one fashion into an x of U multiplied on the left by an element h_x of H. Multiplication of substitutions means successive applications. The set of all monomial substitutions forms the monomial group Σ. Ore [5] has studied this group for finite U, and some of his results have been generalized to general U in [2], [3], and [4].

This paper determines the structure of the characteristic subgroups for the case when U is infinite in the cases where normal subgroups and automorphisms are known. The method used makes clear how corresponding theorems for the case where U is finite might be proved but does not list these results.

2. Definitions and preliminaries. Let d be the cardinal of the integers. Let B be an infinite cardinal; B^+, the successor of B; U, a set such that $o(U) = B$; and C such that $d \leqq C \leqq B^+$. Let H be a fixed group and e the identity of H. Denote by $\Sigma = \Sigma(H; B, d, C)$ the monomial group of U over H whose elements are of the form

$$y = \begin{pmatrix} \cdots, & x_\varepsilon, & \cdots \\ \cdots, & h_\varepsilon x_{i_\varepsilon}, & \cdots \end{pmatrix} \tag{1}$$

where only a finite number of the h_ε are not e and the number of x not mapped into themselves is less than C. Any element of Σ may be written in the form

$$y = \begin{pmatrix} \cdots, & x_\varepsilon, & \cdots \\ \cdots, & h_\varepsilon x_\varepsilon, & \cdots \end{pmatrix} \begin{pmatrix} \cdots, & x_\varepsilon, & \cdots \\ \cdots, & e x_{i_\varepsilon}, & \cdots \end{pmatrix}$$

or $y = vs$ where v sends every x into itself and every h of s is e. Elements of the form of

$$v = \begin{pmatrix} \cdots, & x_\varepsilon, & \cdots \\ \cdots, & h_\varepsilon x_\varepsilon, & \cdots \end{pmatrix} = [\cdots, h_\varepsilon, \cdots]$$

are *multiplications* and all such elements form a normal subgroup, *the basis groups* $V(B, d) = V$ of Σ. The h_ε of y are called the *factors* of y. Elements of the form of s are *permutations* and all such elements form a subgroup, the *permutation group*, $S(B, C) = S$ of $\Sigma(H; B, d, C)$. Cycles

Received March 6, 1959.

of s will also be written as $(x_1, \cdots, x_n)$ and $(\cdots, x_{-1}, x_0, x_1, \cdots)$. Baer [1] has shown that the normal subgroups of $S(B, C)$ are the alternating group, $A = A(B, d)$, and $S(B, D)$ where $d \leqq D \leqq C$. Let E be the identity of Σ, I the identity of S.

3. Characteristic subgroups of $\Sigma(H; B, d, C), d \leqq C < B^+$. The normal subgroups of $\Sigma(H; B, d, C)$ are known [2], [3]. They are classified first as to whether or not they are contained in the basis group V.

If N is normal in Σ and $N \subset V$ its elements are multiplications with only a finite number of non-identity factors which are contained in a normal subgroup G of H. The set of all possible products of factors of all elements of N form a normal subgroup G_1 of H. The group G/G_1 is Abelian and G/G_1 is in the center of H/G_1.

If M is normal in Σ and $M \not\subset V$ then $M \cap S = P \neq E$ is a normal subgroup of S. The group $N = M \cap V$ is as above except that $G = H$. It becomes necessary to consider the cases where $P = S(B, D)$ with $d \leqq D \leqq C$ and $P = A(B, d)$. When $P = S(B, D)$ then $M = N \cup P$.

If M is normal in Σ, $M \not\subset V$, $P = A(B, d)$, $M \cap V = N$, $M/N \cong A(B, d)$ then $M = N \cup A(B, d)$.

If M is normal in Σ, $M \not\subset V$, $P = A(B, d)$, $M \cap V = N$, $M/N \not\cong A(B, d)$ then $M = N \cup A(B, d) \cup L$ where L is the cyclic group generated by $[e, a](1, 2)$ with $a^2 \in G_1$, $a \notin G_1$.

The converses of these theorems are true. That is, if one starts with the proper subgroups of H and constructs N or M as above the resulting group is normal in Σ.

The automorphisms of $\Sigma(H; B, d, C)$ are known [4]. A mapping θ is an automorphism of $\Sigma(H; B, d, C)$ if and only if $\theta = T^+ I_{(s^+)} I_{(v^+)}$ where $T^+, I_{(s^+)}, I_{(v^+)}$ are automorphisms of Σ defined as follows. Let T be any automorphism of H. Then

$$yT^+ = vst^+ = [h_1, \cdots, h_\varepsilon, \cdots]sT^+ = [h_1^T, \cdots, h_\varepsilon^T, \cdots]s \, .$$

Let $s^+ \in S(B, B^+)$. Then $I_{(s^+)}$ is defined by $yI_{(s^+)} = s^+ y (s^+)^{-1}$. Let $v^+ \in V(B, B^+)$ if $C = d$, $v^+ \in V(B, d)$ if $d < C$ then $I_{(v^+)}$ is defined by $yI_{(v^+)} = v^+ y (v^+)^{-1}$.

THEOREM 1. *If N is a subgroup of $\Sigma(H; B, d, C)$ contained in the basis group then N is characteristic in Σ if and only if N is normal in Σ, (hence is as described above) and G, G_1 are characteristic in H.*

Proof. Assume N is characteristic in Σ. Then N is normal in Σ and its structure is known. Choose $\theta = T^+$ with T arbitrary in the automorphism group of H and v arbitrary in N. Then

$$\begin{aligned} v\theta &= [e, \cdots, e, e, g_{i_1}, e, \cdots, e, g_{i_n}, e, \cdots]T^+ \\ &= [e, \cdots, g^T_{i_1}, e, \cdots, e, g^T_{i_n}, e, \cdots] . \end{aligned}$$

The elements $g^T_{i_1}$ must be in G. This shows G is characteristic in H. Furthermore $g^T_{i_1}g^T_{i_2}\cdots g^T_{i_n} = (g_{i_1}\cdots g_{i_n})^T$ must be in G_1 and since $g_{i_1}\cdots g_{i_n}$ is arbitrary in G_1, G_1 is characteristic in H.

Conversely, if $N \subset V(B, d)$, N is normal in Σ, G, G_1 are characteristic in H then N is characteristic in Σ. To see this let v_1 be arbitrary in N. Then $v_1\theta = v_1TI_{(s^+)}I_{(v^+)} = v_2I_{(s^+)}I_{(v^+)}$. The non-identity factors of v_2 are in G and their product in G_1 by G, G_1 characteristic in H. Now $v_2I_{(s^+)}\,I_{(v^+)} = (v^+)(s^+)v_2(s^+)^{-1}(v^+)^{-1}$. The effect of $I_{(s^+)}$ on v_2 is to permute the non-identity factors so $(v^+)(v_3)(v^+)^{-1}$ is now to be considered with v_3 in N. Since G is normal in H in G/G_1 is in the center of H/G_1, $(v^+)v_3(v^+)^{-1}$ will be in N.

THEOREM 2. *Let $M = N \cup P$ be a normal subgroup of $\Sigma(H; B, d, C)$, $d \leqq C < B^+$, where N is as described above, $P = S(B, D)$. Then M is characteristic in Σ if and only if G_1 is characteristic in H.*

Proof. By an argument similar to that used in Theorem 1, G_1 is characteristic in H.

Conversely, if $y = v_1s_1$ is arbitrary in M then

$$v_1s_1\theta = v_1s_1T^+I_{(s^+)}I_{(v^+)} = v_2s_1I_{(s^+)}I_{(v^+)} .$$

Since G_1 is characteristic in H, v_2 belongs to N. Now consider

$$(v^+)(s^+)v_2s_1(s^+)^{-1}(v^+)^{-1} = (v^+)v_3(s^+)s_1(s^+)^{-1}(v^+)^{-1} = (v^+)v_3s_2(v^+)^{-1}.$$

The multiplication v_3 is in N since the factors are still in H, and the product of the factors is still in G_1 since H/G_1 is Abelian. The permutation s_2 is in P since P is normal in $S(B, B^+)$. It is now convenient to consider two cases. If $C = d$ the permutation s_2 is finite and $(v^+)v_3s_2(v^+)^{-1} = (v^+)v_3v_4s_2$ where the factors of v_4 differ from the inverse of those in(v^+) in only a finite number of places. Therefore $(v^+)v_3v_4$ will have a finite number of factors of the form $k_\xi h_\xi k^{-1}_{i_\xi}$. If $k_\xi \neq k_{i_\xi}$ then $k_{i_\xi}h_{i_\xi}k_\alpha$, $k_{i_\xi} \neq k_\alpha$, will be a factor of $(v)v_3v_4$. Since H/G_1 is Abelian the product of the factors is in G_1. Therefore, $(v^+)v_3v_4s_2 = v_5s_2$ belongs to M. If $C > d$ then (v^+), v_4 have only a finite number of non-identity factors and the same argument holds. Therefore $(v^+)v_3v_4s_2$ belongs to M.

THEOREM 3. *Let $M = N \cup A(B, d)$ be a normal subgroup of $\Sigma(H; B, d, C)$, $d \leqq C < B^+$. Then M is characteristic in Σ if and only if G_1 is characteristic in H.*

Proof. The argument used in the proof of Theorem 1 may be used to show that G_1 is characteristic in H if M is characteristic in Σ.

Conversely, if $y = v_1 s_1$ is arbitrary in M then

$$y\theta = v_1 s_1 \theta = v_1 s_1 T^+ I_{(s^+)} I_{(v^+)} = v_2 s_1 I_{(s^+)} I_{(v^+)} = (v^+)(s^+) v_2 s_1 (s^+)^{-1} (v^+)^{-1}$$
$$= (v^+) v_3 (s^+) s_1 (s^+)^{-1} (v^+)^{-1} = (v^+) v_3 s_2 (v^+)^{-1} = (v^+) v_3 v_4 s_2 .$$

Now $v_2 \in N$ by G_1 characteristic in H and v_3 will be in N by H/G_1 Abelian. Since $A(B, d)$ is normal in $S(B, B^+)$, s_2 belongs to $A(B, d)$. The factors of v_4 differ from the inverse of those in v in only a finite number of places since s_2 moves only a finite number of x. Therefore, $(v^+) v_3 v_4 \in N$, $s_2 \in A(B, d)$ and M is characteristic in Σ.

THEOREM 4. *Let $M_1 = N \cup A \cup L$ be a normal subgroup of $\Sigma(H; B, d, C)$, $d \leqq C < B^+$. Let L be generated by $y = [e, a](1, 2)$ with $a^2 \in G_1$, $a \notin G_1$. Then M_1 is characteristic in Σ if and only if G_1 is characteristic in H, and a^T belongs to the coset aG_1 for all automorphisms T of H.*

Proof. By considering $v \in N$ and $\theta = T^+$ we see that G_1 is characteristic in H. By considering $y = [e, a]\,(1, 2)$ of M_1 and $\theta = T^+$ we see that $[e, a^T](1, 2)$ must belong to M_1. This means a^T belongs to aG.

Conversely, if $v_1 s_1 \in M_1$ then

$$v_1 s_1 \theta = v_1 s_1 T^+ I_{(s^+)} I_{(v^+)} = v_2 s_1 I_{(s^+)} I_{(v^+)} = (v^+)(s^+) v_2 s_1 (s^+)^{-1} (v^+)^{-1}$$
$$= (v^+) v_3 (s^+) s_1 (s^+)^{-1} (v^+)^{-1} = (v^+) v_3 s_2 (v^+)^{-1} = (v^+) v_3 v_4 s_2 .$$

Now $v_2 s_1$ is in M_1 by G_1 characteristic if the product of the factors of v_1 is in G_1 and by a^T in aG_1 if the product of the factors is in aG_1. The multiplication v_3 has only a finite number of non-identity factors because v_2 has only a finite number of non-identity factors. Since s_1 is finite, s_2 is a finite permutation and is even or odd as s_1 is even or odd. Therefore, v_4 has only a finite number of factors different from the inverse of the factors of (v^+). The factors of $(v^+) v_3 v_4$ have their product in G_1 or aG_1 according as v_3 has its product in G_1 or aG_1. Therefore, if s_1 was even s_2 is even, v_1 had the product of its factors in G_1 and so does $(v^+) v_3 v_4$. If s_1 was odd so is s_2 and v_1 had the product of its factors in aG_1 and so does $(v^+) v_3 v_4$. That is, M_1 is characteristic.

4. Characteristic subgroups of $\Sigma_A(H; B, d, d)$. The normal subgroups of $\Sigma_A(H; B, d, d)$ are precisely those of $\Sigma(H; B, d, d)$ that are contained in $\Sigma_A(H; B, d, d)$ [2, p. 210]. The automorphism of $\Sigma_A(H; B, d, d)$ are those of $\Sigma(H; B, d, d)$ restricted to $\Sigma(H; B, d, d)$ [4, p. 84].

THEOREM 5. *Let N be a subgroup of $\Sigma_A(H; B, d, d)$ contained in the basis group. Then N is characteristic in Σ_A if and only if N is normal in Σ_A and G, G_1 are characteristic in H.*

THEOREM 6. *Let M be a subgroup of $\Sigma_A(H; B, d, d)$, $M \not\subset V(B, d)$. Then M is characteristic in Σ_A if and only if M is normal,* i.e. $M = N \cup A$, *and G_1 is characteristic in H.*

BIBLIOGRAPHY

1. R. Baer, *Die Kompositionsreihe der Gruppe aller einendeutigen Abbildungen einer unendlichen Menge auf sich*, Studia Mathematica, **5** (1934), 15–17.
2. R. B. Crouch, *Monomial Groups*, Trans. Amer. Math. Soc., **80** (1955), 187–215.
3. R. B. Crouch and W. R. Scott, *Normal subgroups of monomial groups*, Proc. Amer. Math. Soc., **8**, No. 5 (1957), 931–936.
4. C. V. Holmes, *Contributions to the Theory of Groups*, Research Grant NSF-G 1126, Report No. 5, Feb., (1956), 23–93.
5. O. Ore, *Theory of monomial groups*, Trans. Amer. Math. Soc., **51** (1942), 15–64.

NEW MEXICO STATE UNIVERSITY
UNIVERSITY PARK, NEW MEXICO

EXISTENCE THEOREMS FOR CERTAIN CLASSES OF TWO-POINT BOUNDARY PROBLEMS BY VARIATIONAL METHODS

RICHARD J. DRISCOLL

Prefatory remarks. The principal results of this paper are existence theorems for solutions of two classes of vector differential systems; in each case the existence theorem is established by variational methods. In particular, the second system considered is a generalization of a scalar system, including as a special case the so-called Fermi-Thomas equation, studied by Sansone [8; pp. 445-450]. In spite of similarities occurring in the discussion of the two systems considered, the two problems are sufficiently distinct to warrant separate treatment. Accordingly, we shall divide the remaining sections of this paper into two parts, in which the numbering of sections and of displayed material will be independent; the bibliography, however, will apply to both parts.

Matrix notation will be used throughout and all matrices will have real elements; in particular, a vector $u = (u_j)$, $(j = 1, 2, \cdots, n)$, will be regarded as an $n \times 1$ matrix. If M is a matrix, M^* will denote the transpose of M, while for a vector $u = (u_j)$, $(j = 1, 2, \cdots, n)$, we define $|u| = (u_1^2 + \cdots + u_n^2)^{1/2}$. For $F(u)$ a scalar function of the vector u, the symbol $F_u(u)$ will denote the vector function $(F_{u_j}(u))$; if $G(u)$ is a vector function $(G_i(u))$, $(i = 1, 2, \cdots, m)$, of the vector u, then $G_u(u)$ will denote the $m \times n$ matrix $\|\partial G_i / \partial u_j\|$, $(i = 1, \cdots, m; j = 1, \cdots, n)$. If M and N are matrices, the notation $M \geq N$ is used to signify that M and N are real symmetric matrices of the same dimensions and $M - N$ is non-negative. As usual, the symbol $C^{(n)}$ represents the class of finite dimensional matrix functions which are continuous and have continuous derivatives of the first n orders on some given set.

PART I

1. Introduction. This part of the paper will be concerned with vector differential systems of the form

$$(1.1)\qquad \begin{aligned} &y'' = f(x, y, y'), \qquad a \leq x \leq b, \\ &y(a) = y_1, \quad y(b) = y_2, \end{aligned}$$

Received March 2, 1959. [1] This research was supported in part by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract Nos. AF-(603)-86 and AF-49(638)-400. Reproduction in whole or in part is permitted for any purpose of the United States Government This paper was part of a thesis submitted to Northwestern University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author wishes to express his gratitude to Professor W. T. Reid for his encouragement and guidance.

where $f(x, y, z)$ is an n-dimensional real vector function of the real scalar x and the real n-dimensional vectors y and z. It will be shown that the system (1.1) has a solution, under the hypotheses H_1, H_2, H_3, H_5, of § 3 and H_4^* of § 4. For reasons of convenience, we shall work primarily with the system

$$\begin{aligned} y'' &= f(x, y, y') , \qquad a \leq x \leq b , \\ y(a) &= 0 = y(b) , \end{aligned} \tag{1.2}$$

and show in § 4 how a system (1.1) may be reduced to such a system.

The existence proof will use variational methods applied to the functional

$$I(y, z) \equiv \int_a^b (|y' - z|^2 + |z' - f(x, y, z)|^2) dx , \tag{1.3}$$

with (y, z) in the class K of function pairs defined below. In § 2 there are listed some lemmas to be used later. In § 3 an existence theorem for a solution of (1.2) is established by showing, in effect, that $I(y, z)$ has a minimum for (y, z) in K, and that this minimum is zero. The relation between systems of the forms (1.1) and (1.2) is considered in § 4, while § 5 contains a comment on a modification of hypotheses. Finally, § 6 is devoted to an example of a class of problems to which the existence theorem proved here is applicable.

In what follows, A_2 will denote the class of vector functions $y(x)$ which are absolutely continuous and for which $|y'(x)|^2$ is integrable on $a \leq x \leq b$, while K is the class of vector function pairs (y, z) with $y(x)$ and $z(x)$ in A_2 and with $y(a) = 0 = y(b)$.

2. Some useful lemmas. For future reference we collect here certain auxiliary results.

LEMMA 2.1. *Suppose that the matrix $f_z(x, y, z)$ exists and is continuous for $a \leq x \leq b$, all y, and all z. If for each $\rho > 0$ the elements of f_z are bounded for $a \leq x \leq b$, $|y| < \rho$ and z arbitrary, then there are values $K_1 = K_1(\rho)$ and $K_2 = K_2(\rho)$ such that*

$$|f(x, y, z)| \leq K_1 |z| + K_2 , \text{ for } a \leq x \leq b, \ |y| < \rho, \ z \text{ arbitrary}.$$

LEMMA 2.2. *If $\{w_m(x)\}$, $(m = 1, 2, \cdots)$, is a sequence of vector functions of class A_2 such that the two sequences $\left\{\int_a^b |w_m|^2 dx\right\}$ and $\left\{\int_a^b |w'_m|^2 dx\right\}$ are bounded, then the $w_m(x)$ are uniformly bounded on $a \leq x \leq b$, and there exists a $w(x)$ in A_2 and a subsequence $\{w_{m_j}(x)\}$ such that $w_{m_j}(x) \rightarrow w(x)$ uniformly and $w'_{m_j}(x) \rightarrow w'(x)$ weakly in the class of integrable square functions on $a \leq x \leq b$.*

This lemma is a ready consequence of well-known results for the Hilbert space of real-valued measurable functions whose squares are Lebesgue integrable on $a \leq x \leq b$, see, for example, [7; §§ 32, 99].

LEMMA 2.3. *If $y(x)$ is in A_2 and $y(a) = 0$, then*

$$\int_a^b |y'|^2 dx \geq \frac{\pi^2}{4(b-a)^2} \int_a^b |y|^2 dx .$$

This is a well-known condition on the smallest proper value of the differential system $y'' + \lambda y = 0$, $y(a) = 0 = y'(b)$. For an independent proof see [2; p. 184]; the present inequality follows from (7.7.1) of [2] by a simple change of variable.

We will also need some results related to non-oscillation of the scalar differential equation

$$(2.1) \qquad (\psi_1(x)u'(x))' - \psi_2(x)u(x) = 0 , \quad a \leq x \leq b ,$$

where ψ_1 is of class C' and ψ_2 continuous on $a \leq x \leq b$. The equation (2.1) is termed non-oscillatory on $a \leq x \leq b$ if for two arbitrary points x_1, x_2 satisfying $a \leq x_1 < x_2 \leq b$, any solution $u(x)$ of (2.1) vanishing at x_1 and at x_2 vanishes identically on $a \leq x \leq b$. It is well-known that if $\psi_1(x) > 0$ on $a \leq x \leq b$, then (2.1) is non-oscillatory on $a \leq x \leq b$ if and only if

$$(2.2) \qquad J(u) \equiv \int_a^b (\psi_1(x)u'^2(x) + \psi_2(x)u^2(x))dx > 0$$

holds for all non-identically vanishing $u(x)$ belonging to A_2 and satisfying $u(a) = 0 = u(b)$. For a proof of this statement see, for example, [5; Theorem 2.1], where a more general result is proved. Moreover, if (2.1) is non-oscillatory on $a \leq x \leq b$, the infimum of $J(u)$ for $u(x)$ in A_2 and satisfying $u(a) = 0 = u(b)$, $\int_a^b u^2 dx = 1$ is greater than zero, as can be seen from an indirect argument. Indeed, if the infimum were equal to zero, then there would be a sequence of functions u_j in A_2 with $u_j(a) = 0 = u_j(b)$, $\int_a^b u_j^2 dx = 1$, $j = 1, 2, \cdots$, and with $J(u_j) \to 0$. One readily verifies that the sequence $\left\{\int_a^b u_j'^2 dx\right\}$ would be bounded, so that, by Lemma 2.2, there would be a $u(x)$ in A_2 and a subsequence of $\{u_j\}$, denoted again by $\{u_j\}$, such that $u_j(x) \to u(x)$ uniformly on $a \leq x \leq b$, and $u_j'(x) \to u'(x)$ weakly on this interval. The identity

$$J(u_j) = J(u) + \int_a^b \psi_1(x)[2u'(u_j' - u') + (u_j'^2 - u'^2)]dx + \int_a^b \psi_2(x)(u_j^2 - u^2)dx$$

would then imply that $0 = \lim J(u_j) \geq J(u)$, contrary to (2.2), since $u(a) = 0 = u(b)$ and $\int_a^b u^2 dx = 1$. With these comments one readily establishes the following result.

LEMMA 2.4. *If* (2.1) *is non-oscillatory on* $a \leq x \leq b$, *and* $\psi_1(x) > 0$ *on this interval, then there exists an* $\varepsilon > 0$ *such that if* $h(x)$ *is any function continuous and satisfying* $|h(x)| < \varepsilon$ *on* $a \leq x \leq b$, *then the equation* $(\psi_1(x)u')' - (\psi_2(x) + h(x))u = 0$ *is non-oscillatory on* $a \leq x \leq b$.

3. Existence theorem for a solution of (1.2). In the future sections we will make reference to the following hypotheses on the real-valued vector function $f(x, y, z)$:

H_1. $f(x, y, z)$ is continuous for (x, y, z) in $\Omega : a \leq x \leq b$, $|y| < \infty$, $|z| < \infty$.

H_2. *The matrices* f_y *and* f_z *exist and are continuous for* (x, y, z) *in* Ω.

H_3. *For any* $\rho > 0$, *there exists a* $K = K_\rho$ *such that* $|\partial f_i/\partial z_j| \leq K$ *for* $|y| < \rho$, $a \leq x \leq b$, $|z| < \infty$, $(i, j = 1, \cdots, n)$.

H_4. *For arbitrary* $\rho > 0$ *there exist scalar functions* $\psi_1(x) = \psi_1(x; \rho) \in C$, $\psi_2(x) = \psi_2(x; \rho) \in C'$ *with* $\psi_2(x) > 0$ *on* $a \leq x \leq b$, *and a constant* $N = N(\rho)$ *such that:*

(a) *the scalar differential equation* $(\psi_2(x)w')' - \psi_1(x)w = 0$ *is non-oscillatory on* $a \leq x \leq b$;

(b) *the integral inequality*

$$2\int_a^b y^* f(x, y, z)dx \geq \int_a^b [(\psi_2 - 1)|y'|^2 + \psi_1 |y|^2]dx - N$$

holds for all $y(x)$, $z(x)$ *in* A_2 *satisfying* $y(a) = 0 = y(b)$ *and*

$$\int_a^b |y' - z|^2 dx \leq \rho .$$

H_5. *For arbitrary* $y(x)$, $z(x)$ *in* A_2, *the vector differential system*

$$\begin{aligned} w'' - f_z(x, y(x), z(x))w' - f_y(x, y(x), z(x))w &= 0 , \quad a \leq x \leq b \\ w(a) = 0 &= w(b) \end{aligned} \tag{3.1}$$

has only the solution $w(x) \equiv 0$, $a \leq x \leq b$.

We now prove the following theorem.

THEOREM 3.1. *Under the hypotheses* H_1–H_5 *there exists a solution of the system* (1.2).

Let $\{y_m(x), z_m(x)\}$, $m = 1, 2, \cdots$, be a sequence of function pairs of class K such that $I(y_m, z_m) \to I_0$, where I_0 denotes the infimum of $I(y, z)$ on K. Since $\{I(y_m, z_m)\}$ is a convergent sequence, there exists a constant M_0 such that $I(y_m, z_m) \leq M_0$, $m = 1, 2, \cdots$. It will be shown first that the inequality

$$\int_a^b (|y_m'(x)|^2 + 2y_m^*(x)f(x, y_m(x), z_m(x)))dx \leq M , \quad m = 1, 2, \cdots \tag{3.2}$$

holds for

$$M = 2M_0/k\,, \quad \text{where} \quad k = \text{Min}\,(1, \pi^2/[4(b-a)^2])\,. \tag{3.3}$$

Let $v_m(x) = \int_a^x f_m(s)ds + z_m(a)$, where $f_m(x) = f(x, y_m(x), z_m(x))$. Then $u_m(x) = z_m(x) - v_m(x)$ is in A_2, and $u_m(a) = 0$, so that by Lemma 2.3,

$$\int_a^b |z'_m - f_m|^2 dx = \int_a^b |u'_m|^2 dx \geq \frac{\pi^2}{4(b-a)^2}\int_a^b |u_m|^2 dx\,.$$

This inequality yields

$$M_0 \geq k\int_a^b (|y'_m - z_m|^2 + |z_m - v_m|^2)dx \geq (k/2)\int_a^b |y'_m - v_m|^2 dx\,, \tag{3.4}$$

where k is as in (3.3). Since

$$\int_a^b y'^*_m v_m dx = y^*_m v_m \Big|_a^b - \int_a^b y^*_m v'_m dx = -\int_a^b y^*_m f_m dx\,,$$

relation (3.2), with M given by (3.3), results from (3.4) and the obvious inequality

$$\int_a^b |y'_m - v_m|^2 dx \geq \int_a^b (|y'_m|^2 - 2y^{*\prime}_m v_m)dx\,.$$

Since the sequence $\left\{\int_a^b |y'_m - z_m|^2 dx\right\}$ is bounded, we may use H_4 to write

$$2\int_a^b y^*_m f(x, y_m, z_m)dx \geq \int_a^b [(\psi_2 - 1)|y'_m|^2 + \psi_1|y_m|^2]dx - N\,,$$

where $\psi_2(x)$, $\psi_1(x)$, and N have the properties stated in H_4. This relation with (3.2) yields

$$\int_a^b (\psi_2|y'_m|^2 + \psi_1|y_m|^2)dx \leq M + N\,.$$

Since $(\psi_2(x)u')' - \psi_1(x)u = 0$ is non-oscillatory on $a \leq x \leq b$, Lemma 2.4 implies that there is an r with $0 < r < 1$ such that $(\psi_2 u')' - (1/r)\psi_1 u = 0$ is non-oscillatory on $a \leq x \leq b$. As $y_m(a) = 0 = y_m(b)$, we then have

$$\int_a^b (r\psi_2|y'_m|^2 + \psi_1|y_m|^2)dx \geq 0$$

and

$$\int_a^b (\psi_2|y'_m|^2 + \psi_1|y_m|^2)dx \geq (1-r)\int_a^b \psi_2|y'_m|^2 dx \geq r_0\int_a^b |y'_m|^2 dx\,,$$

where $r_0 = (1-r)\,\text{Min}_{a\leq x\leq b}\psi_2(x)$. Thus, the sequence $\left\{\int_a^b |y'_m|^2 dx\right\}$ is bounded, and since each $y_m(x)$ vanishes at a and b, the vector functions $y_m(x)$ are uniformly bounded on $a \leq x \leq b$. Moreover,

$$\begin{aligned}\int_a^b |z_m|^2 dx &\leq \int_a^b |y'_m + (z_m - y'_m)|^2 dx \\ &\leq 2\int_a^b |y'_m|^2 dx + 2\int_a^b |z_m - y'_m|^2 dx \\ &\leq 2\int_a^b |y'_m|^2 dx + 2M_0\end{aligned}$$

so that the sequence $\left\{\int_a^b |z_m|^2 dx\right\}$ is bounded. Finally, with $f_m(x)$ continuing to denote $f(x, y_m(x), z_m(x))$, we have

$$\begin{aligned}\int_a^b |z'_m|^2 dx &= \int_a^b |(z'_m - f_m) + f_m|^2 dx \\ &\leq 2\int_a^b |z'_m - f_m|^2 dx + 2\int_a^b |f_m|^2 dx \\ &\leq 2M_0 + 2\int_a^b |f_m|^2 dx\ .\end{aligned}$$

As the vector functions $y_m(x)$, $(m = 1, 2, \cdots)$, are bounded uniformly on $a \leq x \leq b$, in view of hypothesis H_3 and the result of Lemma 2.1, this latter inequality implies $\int_a^b |z'_m|^2 dx \leq K' + K''\int_a^b |z_m|^2 dx + 2M_0$, for suitable constants K', K''. Hence, the two sequence $\{y_m(x)\}$, $\{z_m(x)\}$ satisfy the hypotheses of Lemma 2.2, and we conclude that there exist subsequences, which will be denoted simply by $\{y_m(x)\}$ and $\{z_m(x)\}$, and a pair of functions $y(x)$, $z(x)$ in A_2, such that $y_m(x) \to y(x)$ and $z_m(x) \to z(x)$ uniformly on $a \leq x \leq b$, while $y'_m(x) \to y'(x)$ and $z'_m(x) \to z'(x)$ weakly on the same interval.

With $f_m(x)$ as above and $f(x) = f(x, y(x), z(x))$ we have

$$I(y_m, z_m) = I(y, z) + I_{1,m} + I_{2,m}\ ,$$

where

$$I_{1,m} = \int_a^b [|(z_m - z) - (y'_m - y')|^2 + |(f_m - f) - (z'_m - z')|^2] dx$$

and

$$\begin{aligned}I_{2,m} = 2\int_a^b \{&(y' - z)^*[(y'_m - y') - (z_m - z)] \\ &+ (z' - f)^*[(z'_m - z') - (f_m - f)]\} dx\ .\end{aligned}$$

Since $y_m(x) \to y(x)$, $z_m(x) \to z(x)$ uniformly, we also have $f_m(x) \to f(x)$ uniformly on $a \leq x \leq b$. This, and the fact that $y'_m \to y'$, $z'_m \to z'$ weakly on the same interval, implies that $I_{2,m} \to 0$ as $m \to \infty$. As $I_{1,m} \geq 0$, it follows that

$$I_0 = \lim_{m\to\infty} I(y_m, z_m) \geq I(y, z) .$$

On the other hand, the definition of I_0 requires $I_0 \leq I(y, z)$, so that $I_0 = I(y, z)$; that is, (y, z) renders $I(y, z)$ a minimum in the class of function pairs K.

It follows that for arbitrary $\eta(x)$, $\zeta(x)$ in A_2 with $\eta(a) = 0 = \eta(b)$, and θ a real parameter, the functional $I(y + \theta\eta, z + \theta\zeta)$ has a minimum at $\theta = 0$, and therefore $(d/d\theta)I(y + \theta\eta, z + \theta\zeta) = 0$ for $\theta = 0$; that is,

$$\int_a^b [(y^{*\prime} - z^*)(\eta' - \zeta) + (z^{*\prime} - f^*)(\zeta' - f_y\eta - f_z\zeta)]dx = 0 , \tag{3.5}$$

where the arguments of f, f_y, f_z are x, $y(x)$, $z(x)$.

In view of H_5, (see [4; pp. 213–214]), for an arbitrary continuous function $g(x)$, $a \leq x \leq b$, there exists a solution $(\eta(x), \zeta(x))$ of the differential system

$$\begin{aligned} &\eta' - \zeta = 0 , \\ &\zeta' - f_y(x, y(x), z(x))\eta - f_z(x, y(x), z(x))\zeta = g(x) , \quad a \leq x \leq b , \\ &\eta(a) = 0 = \eta(b) . \end{aligned}$$

Therefore, $\int_a^b [z^{*\prime} - f^*(x, y, z)]g(x)dx = 0$ for arbitrary $g(x)$ continuous on $a \leq x \leq b$, and consequently $z'(x) - f(x, y(x), z(x)) = 0$ a.e. on the same interval. Relation (3.5), with $\eta(x)$ chosen identically zero on $a \leq x \leq b$, then yields $\int_a^b \zeta^*(y' - z)dx = 0$ for arbitrary ζ in A_2, and hence $y'(x) = z(x)$ a.e. on $a \leq x \leq b$. From the relations $z(x) = z(a) + \int_a^x f(s, y(s), z(s))ds$, $y(x) = \int_a^x z(s)ds$, it then follows that $y(x)$ and $z(x)$ are of class C', and that $y'(x) = z(x)$, $z'(x) = f(x, y(x), z(x))$ for $a \leq x \leq b$, so that $y(x)$ is of class C'' and satisfies (1.2).

4. Existence of a solution of (1.1).

For the system

$$\begin{aligned} &y'' = f(x, y, y') , \qquad a \leq x \leq b , \\ &y(a) = y_1 , \qquad y(b) = y_2 , \end{aligned} \tag{1.1}$$

let $F(x, y, z) \equiv f(x, y + \lambda(x), z + \lambda'(x)) - \lambda''(x)$, where $\lambda(x)$ is any vector function of class C'' on $a \leq x \leq b$ satisfying $\lambda(a) = y_1$, $\lambda(b) = y_2$. Then (1.1) is equivalent, with $u = y - \lambda$, to

$$\begin{aligned} &u'' = F(x, u, u') , \qquad a \leq x \leq b , \\ &u(a) = 0 = u(b) . \end{aligned} \tag{4.1}$$

This leads us to formulate the following hypothesis.

H_4^*. *There exists* $\lambda(x)$ *of class* C'' *on* $a \leq x \leq b$ *with* $\lambda(a) = y_1$, $\lambda(b) = y_2$, *and such that for arbitrary* $\rho > 0$ *there exist scalar functions*

$\psi_1(x) = \psi_1(x;\rho)$, continuous on $a \le x \le b$, $\psi_2(x) = \psi_2(x;\rho)$ of class C' on $a \le x \le b$ with $\psi_2(x) > 0$, and a constant $N = N(\rho)$ such that:

(a) *the scalar differential system $(\psi_2(x)w')' - \psi_1(x)w = 0$ is non-oscillatory on $a \le x \le b$;*

(b) *the integral inequality*

$$2\int_a^b y^* f(x, y+\lambda, z+\lambda')dx \ge \int_a^b [(\psi_2 - 1)|y'|^2 + \psi_1|y|^2]dx - N$$

holds for all $y(x)$, $z(x)$ in A_2 satisfying $y(a) = 0 = y(b)$ and

$$\int_a^b |y' - z|^2 dx < \rho\,.$$

THEOREM 4.1. *Under hypotheses $H_1, H_2, H_3, H_4^*, H_5$, the system* (1.1) *has a solution.*

Let $F(x, y, z) = f(x, y+\lambda(x), z+\lambda'(x)) - \lambda''(x)$, where $\lambda(x)$ is the function described in H_4^*. Clearly, $F(x, y, z)$ satisfies H_1, H_2, H_3. Since f satisfies H_4^*, we have

$$2\int_a^b y^* f(x, y+\lambda, z+\lambda')dx \ge \int_a^b [(\psi_2 - 1)|y'|^2 + \psi_1|y|^2]dx - N$$

for arbitrary $y(x)$, $z(x)$ satisfying $y(a) = 0 = y(b)$ and $\int_a^b |y' - z|^2 dx \le \rho$. Hence, for such $y(x)$, $z(x)$ we have

$$\begin{aligned} 2\int_a^b y^* F(x, y, z)dx &\ge \int_a^b [(\psi_2 - 1)|y'|^2 + \psi_1|y|^2]dx - N - 2\int_a^b y^*\lambda''(x)dx \\ &\ge \int_a^b [(\psi_2 - 1)|y'|^2 + (\psi_1 - \varepsilon)|y|^2]dx \\ &\quad - \left(N + \frac{1}{\varepsilon}\int_a^b |\lambda''|^2 dx\right), \end{aligned}$$

for any $\varepsilon > 0$. But by Lemma 2.4, ε can be chosen so small that $(\psi_2 w')' - (\psi_1 - \varepsilon)w = 0$ is still non-oscillatory on $a \le x \le b$. Thus, $F(x, y, z)$ satisfies H_4. Finally, one easily verifies that if $f(x, y, z)$ satisfies H_5 then $F(x, y, z)$ satisfies H_5. Thus, whenever $f(x, y, z)$ satisfies the hypotheses of Theorem 4.1, the corresponding function $F(x, y, z)$ of (4.1) satisfies the hypotheses of Theorem 3.1, so that the result of Theorem 4.1 is a direct corollary of Theorem 3.1.

5. A comment on altered hypotheses. We note here that hypothesis H_4 is implied by the more restrictive but simpler hypotheses H_4' and H_4''.

H_4'. *There exists a constant C such that*

$$|y^*(f(x, y, z_2) - f(x, y, z_1))| \le C|y||z_2 - z_1|\,, \textit{ for } (x, y, z_1), (x, y, z_2) \textit{ in } \Omega.$$

H_4''. *There exist scalar functions $\psi_1(x)$, continuous on $a \le x \le b$,* *and* $\psi_2(x) > 0$ *of class* C' *on* $a \leq x \leq b$, *and a constant* N *such that:*

(a) *the scalar differential system* $(\psi_2(x)w')' - \psi_1(x)w = 0$ *is non-oscillatory on* $a \leq x \leq b$;

(b) *the integral inequality*

$$2\int_a^b y^* f(x, y, y')dx \geq \int_a^b [(\psi_2 - 1)|\, y' \,|^2 + \psi_1 |\, y \,|^2]dx - N$$

holds for all $y(x)$ *in* A_2 *satisfying* $y(a) = 0 = y(b)$.

To see that H_4 is implied by H_4' and H_4'' (assuming, of course, H_1, H_2), for $y(x)$, $z(x)$ in A_2 and $\varepsilon > 0$ we write,

$$\begin{aligned}
2\int_a^b y^* f(x, y, z)dx &= 2\int_a^b y^* f(x, y, y')dx + 2\int_a^b y^*[f(x, y, z) - f(x, y, y')]dx \\
&\geq 2\int_a^b y^* f(x, y, y')dx - 2C\int_a^b |\, y \,|\,|\, y' - z \,|dx \\
&\geq 2\int_a^b y^* f(x, y, y')dx - C\varepsilon\int_a^b |\, y \,|^2 dx - (C/\varepsilon)\int_a^b |\, y' - z \,|^2 dx \\
&\geq \int_a^b [(\psi_2 - 1)|\, y' \,|^2 + (\psi_1 - C\varepsilon)|\, y \,|^2]dx - [(C\rho)/\varepsilon + N]
\end{aligned}$$

for all $y(x)$, $z(x)$ in A_2 with $y(a) = 0 = y(b)$ and $\int_a^b |\, y' - z \,|^2 dx < \rho$. Since ε can be chosen so small that $(\psi_2 w')' - (\psi_1 - \varepsilon C)w = 0$ is still non-oscillatory on $a \leq x \leq b$, we see that H_4' and H_4'' imply H_4.

It is to be noted that if the elements of $f_z(x, y, z)$ are bounded for (x, y, z) in Ω, then $f(x, y, z)$ satisfies both H_3 and H_4'.

6. An example. Let $f(x, y, z) = g(x, y)(1 + z^2)^{1/2}$, where z is a scalar and $g(x, y)$ is a scalar function of the scalars x and y satisfying the conditions:

(6.1)
(a) $g(x, y)$ *and* $g_y(x, y)$ *are continuous for* $a \leq x \leq b$, $-\infty < y < \infty$;
(b) $g_y(x, y) \geq 0$ *for* $a \leq x \leq b$, $-\infty < y < \infty$;
(c) *there exists a constant* $A > 0$ *such that if* $|\, y \,| \geq A$ *then*

$$yg(x, y) \geq 0\,, \qquad a \leq x \leq b\,.$$

One may verify that $f(x, y, z)$ satisfies hypotheses H_1, H_2, H_3, H_4^*, and H_5.

PART II

1. Introduction. Sansone [8; pp. 445–450] has proved the existence and uniqueness of a solution of the scalar differential system

$$\begin{aligned}
&y'' = \psi(x)\phi(x, y)\,, \qquad 0 < x < \infty\,, \\
(1.1)\qquad &y(0) = y_0\,, \qquad y(+\infty) = 0\,, \\
&y \in C' \quad \text{on} \quad 0 \leq x < \infty\,,
\end{aligned}$$

under assumptions which are related to hypotheses H_1–H_6 (see §§ 2, 7) of this paper. The product $\psi(x)\phi(x, y)$ appears in (1.1) to facilitate stating the hypotheses in such a way as to include the Fermi-Thomas system (see [8; p. 445]),

$$
\begin{aligned}
&y'' = x^{-1/2}y^{3/2}\ , \\
&y(0) = 1\ , \qquad \lim_{x\to\infty} y(x) = 0\ .
\end{aligned} \tag{1.2}
$$

In this paper we consider solutions of a vector differential system, for which we prove an existence and uniqueness theorem which includes the results of Sansone.

The proof given in [8] may be considered in two parts. In the first part the author proves, in effect, that under his hypotheses the system

$$
\begin{aligned}
&y'' = \psi(x)\phi(x, y)\ , \qquad 0 < x < \infty\ , \\
&y(0) = y_0\ , \quad y(x)\ \textit{bounded on}\ 0 \le x < \infty, \\
&y \in C' \quad \textit{on} \quad 0 \le x < \infty\ ,
\end{aligned} \tag{1.3}
$$

has a unique solution. Essential to Sansone's proof of this result is the fact that his hypotheses guarantee a local uniqueness property for solutions of

$$
y'' = \psi(x)\phi(x, y)\ ; \tag{1.4}
$$

that is, under his hypotheses, (1.4) has for $0 \le x_0 < \infty$ exactly one solution satisfying $y(x_0) = y_0$, $y'(x_0) = y'_0$. The hypotheses of the present paper, however, are not strong enough to imply such local uniqueness, as will be shown by an example in § 2. In the second phase of his proof, Sansone appeals to hypotheses which are designed to guarantee that the bounded solutions of (1.3) actually satisfy (1.1). In this paper we make a similar step, but again our hypothesis is weaker than the corresponding ones in [8], as will be made clear in § 7.

Sections 2–5 of this paper present an existence and uniqueness theorem for a solution of the vector generalization of Sansone's system mentioned above. This proof is primarily by variational methods, and the solutions are shown to be characterized by an extremal property. In § 6 there is given a different characterization of these solutions, while § 7 contains several theorems relating to the asymptotic behavior of solutions. Finally, § 8 is devoted to properties of solutions of (2.1) as functions of initial values.

2. Formulation of the problem. Let $g(x, y)$ be a real-valued scalar function of the scalar x and the n-dimensional vector $y = (y_j)$. We will denote by $g_y(x, y)$ the vector $(g_{y_j}(x, y))$, and consider the problem of solving the vector differential system

$$(2.1) \qquad \begin{aligned} &y''(x) = g_y(x, y(x)) \;, \qquad 0 < x < \infty \;, \\ &y(0) = y_0 \;, \qquad y(x) \textit{ bounded on } 0 \leq x < \infty \;, \end{aligned}$$

where $y(x) \in C'$ on $0 \leq x < \infty$ and $y(x) \in C''$ on $0 < x < \infty$. We will suppose that $g(x, y)$ has the form $g(x, y) = \psi(x)G(x, y)$, where $\psi(x)$ and $G(x, y)$ are real-valued functions which satisfy the following hypotheses:

H_1. *$G(x, y)$ is continuous in (x, y) on Ω: $0 \leq x < \infty$, $|y| < \infty$, and $G(x, 0) \equiv 0$ for $0 \leq x < \infty$.*

H_2. *$G_y(x, y)$ exists and is continuous in (x, y) on Ω.*

H_3. *$y^*G_y(x, y) \geq 0$ for (x, y) on Ω.*

H_4. *$\eta^*[G_y(x, y + \eta) - G_y(x, y)] \geq 0$ for (x, y), (x, η) on Ω.*

H_5. *$\psi(x)$ is continuous and positive for $x > 0$ and integrable on any finite closed interval $0 \leq x \leq A$.*

It is to be noted that $g(x, y)$ may satisfy H_1–H_5 without the equation $y'' = g_y(x, y)$ having the local uniqueness property mentioned in § 1. Indeed, if we take

$$g(x, y) = \begin{cases} 8y^{3/2} \;, & y \geq 0 \;, \\ 0 \;, & y \leq 0 \;, \end{cases}$$

so that

$$g_y(x, y) = \begin{cases} 12y^{1/2} \;, & y \geq 0 \;, \\ 0 \;, & y \leq 0 \;, \end{cases}$$

it is easily verified that $g(x, y)$ satisfies H_1–H_5, with $\psi(x) \equiv 1$. However, the function $y_1(x) = (x - x_0)^4$, $x_0 > 0$, satisfies the equation $y''(x) = g_y(x, y(x))$, as does the function $y_2(x) \equiv 0$. Since we have $y_1(x_0) = y_2(x_0)$, $y_1'(x_0) = y_2'(x_0)$, it follows that the local uniqueness property does not obtain here.

A few consequences of the above hypotheses are worthy of comment. First, observe that H_2 and H_3 imply $G_y(x, 0) \equiv 0$ for $0 \leq x < \infty$. Also, since $G(x, 0) \equiv 0$ by H_1, and

$$G(x, y) = \int_0^1 \left[\frac{d}{ds} G(x, sy) ds \right] = \int_0^1 y^* G_y(x, sy) ds = \int_0^1 s^{-1}(sy^* G_y(x, sy)) ds \;,$$

we have by H_3 that $G(x, y) \geq 0$ on Ω. Moreover, if $y(x)$ is continuous on $0 \leq x \leq A$, $A > 0$, and $y''(x)$ exists and satisfies $y''(x) = g_y(x, y(x))$ for $0 \leq x \leq A$, then $y \in C'$ on $0 \leq x \leq A$. To see this we write

$$y'(x) = y'(A) - \int_x^A y''(t) dt = y'(A) - \int_x^A \psi(t) G_y(t, y(t)) dt \;, \quad 0 < x \leq A \;.$$

Hence, $\lim_{x \to 0} y'(x)$ exists. This, with the fact that $y(x)$ is continuous for $0 \leq x \leq A$, implies $y'(0)$ exists and that $\lim_{x \to 0} y'(x) = y'(0)$.

Next we note that if $G(x, y)$ satisfies H_1 and H_2, then H_4 is equivalent to the statement that $G(x, y)$ is convex in y; that is,

$$G(x, y_2) - G(x, y_1) - (y_2 - y_1)^* G_y(x, y_1) \geq 0$$

for arbitrary (x, y_1), (x, y_2) in Ω. Finally, we note that the condition $G(x, 0) \equiv 0$ of H_1 is no essential restriction, since if $G(x, y)$ satisfies H_1–H_5 with the exception of this condition, then the function $G_1(x, y) \equiv G(x, y) - G(x, 0)$ satisfies H_1–H_5 and presents the same differential system (2.1).

3. Some properties of solutions. In addition to the system (2.1), we will consider also the system

$$\begin{aligned} y''(x) &= g_y(x, y(x)) \,, \qquad 0 \leq a \leq x \leq b \,, \\ y(a) &= y_a \,, \qquad y(b) = y_b \,, \end{aligned} \tag{3.1}$$

where y is of class C'' on $a \leq x \leq b$, with the obvious modification in case $a = 0$. For these systems we prove the following theorem.

THEOREM 3.1. *Under hypotheses H_1–H_5, the systems* (2.1) *and* (3.1) *have at most one solution.*

We will give the proof for (2.1); the proof for (3.1) is similar. If $y_1(x)$ and $y_2(x)$ are two solutions of (2.1), let $\eta(x) = y_1(x) - y_2(x)$. By H_4 and H_5 we have for $0 < x < \infty$,

$$0 \leq \eta^*[g_y(x, y_2 + \eta) - g_y(x, y_2)] = \eta^*[g_y(x, y_1) - g_y(x, y_2)] = \eta^*\eta'' \,,$$

and hence,

$$\int_0^x \eta^*(t)\eta''(t)dt \geq 0 \,, \qquad 0 < x < \infty \,.$$

Consequently, upon integration by parts, we get

$$\eta^*(x)\eta'(x) \geq \int_0^x |\eta'|^2 dt \geq 0 \,.$$

Since $(|\eta|^2)' = 2\eta^*\eta'$ and $(|\eta^2|)'' = 2|\eta'|^2 + 2\eta^*\eta''$, it then follows that

$$(|\eta(x)|^2)' \geq 0 \,, \quad \text{and} \quad (|\eta(x)|^2)'' \geq 0 \,, \quad 0 \leq x < \infty \,.$$

Consequently, either $\eta(x) \equiv 0$, $0 \leq x < \infty$, or else $|\eta(x)| \to \infty$ as $x \to \infty$. Since the latter is impossible, (2.1) has at most one solution.

The following result will be of use later.

LEMMA 3.1. *If $g(x, y)$ satisfies H_1–H_5, and $y(x)$ is a solution of $y''(x) = g_y(x, y(x))$ on $0 < x < \infty$ with $\int_0^\infty |y'|^2 dx < \infty$, then $y(x)$ is bounded on $0 \leq x < \infty$.*

If $y(x)$ satisfies $y'' = g_y(x, y)$, then, since $(|y|^2)'' = 2|y'|^2 + 2y^*y'' = 2|y'|^2 + 2y^*g_y(x, y) \geq 0$, we know that either there is an x_1 such that

$|y| \equiv 0$ for $x \geq x_1$, or else there is an x_2 such that $|y| \neq 0$ for $x \geq x_2$. In the latter case we have

$$|y|\,|y|' = y^*y' , \qquad x \geq x_2 ,$$

and

$$|y|^3|y|'' = |y|^2(y^*y'') + (|y|^2|y'|^2 - (y^*y')^2) \geq 0 , \qquad x \geq x_2 ,$$

since $y^*y'' = y^*g_y(x, y) \geq 0$. Hence, either $|y|' \leq 0$ for $x \geq x_2$, in which case $\lim_{x\to\infty}|y(x)| \leq |y(x_2)|$, or there is an $\alpha > 0$ and an $x_3 \geq x_2$ such that $|y|' \geq \alpha > 0$ for $x \geq x_3$. In this latter case, for $x \geq x_3$ we have $|y|\,|y'| \geq y^*y' = |y|\,|y|' \geq \alpha|y|$, so that $|y'| \geq \alpha > 0$ and consequently $\int_0^\infty |y'|^2dx = \infty$. Since this is the only case in which $y(x)$ would be unbounded, we conclude that if $y(x)$ is unbounded then $\int_0^\infty |y'|^2dx = \infty$.

4. A preliminary existence theorem. In what follows $I(y; a, b)$ will denote the functional

$$I(y; a, b) = \int_a^b [|y'|^2 + 2g(x, y)]dx , \qquad y(x) \text{ in } K(a, b) ,$$

where $K(a, b)$ is the class of absolutely continuous vector functions $y(x)$ with $|y'(x)|^2$ integrable on $a \leq x \leq b$, and satisfying $y(a) = y_a$, $y(b) = y_b$. We prove the following result.

THEOREM 4.1. *If $g(x, y)$ satisfies hypotheses H_1–H_5, then for any a, b satisfying $0 \leq a < b$, the system (3.1) has a unique solution. Moreover, this solution is a unique minimizing function for $I(y; a, b)$ in the class $K(a, b)$.*

By H_5 and the fact that $g(x, y) \geq 0$, we see that $I(y; a, b) \geq 0$ for y in $K(a, b)$. Let $\bar{I}(a, b)$ denote the infimum of $I(y; a, b)$ for y in $K(a, b)$, and let $\{y_m(x)\}$ be a sequence of elements of $K(a, b)$ such the $I(y_m; a, b) \to \bar{I}(a, b)$. As $g(x, y_m(x)) \geq 0$ on $a < x \leq b$, we have

$$\int_a^b |y_m'|^2dx = I(y_m; a, b) - 2\int_a^b g(x, y_m)dx \leq I(y_m; a, b) ,$$

so that there exists an N such that $\int_a^b |y_m'|^2dx < N$ for $m = 1, 2, \cdots$. Moreover, for $a \leq x \leq b$,

$$|y_m(x) - y_a|^2 = \left|\int_a^x y_m'(t)dt\right|^2 \leq (x - a)\int_0^x |y_m'|^2dt \leq (b - a)\int_a^b |y_m'|^2dt ,$$

so that $|y_m(x) - y_a| \leq [(b - a)N]^{1/2}$, and hence $|y_m(x)| \leq |y_a| + [(b - a)N]^{1/2}$. Consequently, we may use Lemma 2.2 of Part I to conclude that there is a subsequence, which we will denote again by $\{y_m(x)\}$, and a function

$y(x)$ in $K(a, b)$, such that $y_m(x) \to y(x)$ uniformly on $a \leq x \leq b$, and $y'_m(x) \to y'(x)$ weakly on this interval.

From the identity

$$I(y_m; a, b) - I(y; a, b) = \int_a^b [|y'_m - y'|^2 + 2(g(x, y_m) - g(x, y)) \\ + 2(y'_m - y')^* y']dx ,$$

and the fact that $y_m(x) \to y(x)$ uniformly on $a \leq x \leq b$ while $y'_m(x) \to y'(x)$ weakly on this interval, one obtains the lower semi-continuity relation

$$\bar{I}(a, b) = \lim_{m \to \infty} I(y_m; a, b) \geq I(y; a, b) .$$

Since the definition of $\bar{I}(a, b)$ requires that $\bar{I}(a, b) \leq I(y; a, b)$, we see that $\bar{I}(a, b) = I(y; a, b)$; that is, $y(x)$ minimizes $I(y; a, b)$ in the class $K(a, b)$.

It follows that if $\eta(x)$ is absolutely continuous with $\eta(a) = 0 = \eta(b)$ and $|\eta'(x)|^2$ is integrable on $a \leq x \leq b$, and θ is a real parameter, then $I(y + \theta\eta; a, b)$ has a minimum at $\theta = 0$. From this it follows that $(d/d\theta)I(y + \theta\eta; a, b) = 0$ at $\theta = 0$; that is,

$$\int_a^b [\eta'^* y' + \eta^* g_y(x, y)]dx = 0 .$$

In particular, this last equality holds for arbitrary η of class C'' on $a \leq x \leq b$ with $\eta(a) = 0 = \eta(b) = \eta'(a) = \eta'(b)$, and for such an η integration by parts leads to

$$\int_a^b \eta''^* \left[y(x) - \int_a^x ds \int_a^s g_y(t, y(t))dt \right] dx = 0 . \tag{4.1}$$

By the fundamental lemma of the calculus of variations, there exist constant vectors ξ_1 and ξ_2 such that

$$y(x) = \int_a^x ds \int_a^s g_y(t, y(t))\, dt + \xi_1 x + \xi_2 , \qquad a \leq x \leq b .$$

Therefore, $y''(x)$ exists and satisfies

$$y''(x) = g_y(x, y(x)) , \qquad a \leq x \leq b ,$$

with the understanding that if $a = 0$, then $y''(x)$ may fail to exist at $x = 0$. Since $y(a) = y_a$, $y(b) = y_b$, it follows that $y(x)$ satisfies (3.1). The uniqueness of this solution follows from Theorem 3.1. Moreover, since the above argument shows that any function of class $K(a, b)$ that minimizes $I(y; a, b)$ is a solution of (3.1), it follows that the above determined $y(x)$ is the unique minimizing function for $I(y; a, b)$ in $K(a, b)$.

5. An existence theorem for a solution of (2.1). In what follows, K will denote the class of absolutely continuous vector functions $y(x)$ with $|y'|^2$ integrable on $0 \leq x < \infty$ and satisfying $y(0) = y_0$, $I(y; 0, \infty) < \infty$, where

$$I(y; 0, \infty) = \int_0^\infty [|y'|^2 + 2g(x, y)]dx .$$

We now prove the following result. .

THEOREM 5.1. *Under hypotheses H_1–H_5 the system* (2.1) *has a unique solution; moreover, this solution is a unique minimizing function for $I(y; 0, \infty)$ in the class K.*

Let $\{y_m(x)\}$, $m = 1, 2, \cdots$, be a sequence of functions in K such that $I(y_m; 0, \infty) \to \bar{I}$, where $\bar{I}$ denotes the infimum of $I(y; 0, \infty)$ for y in K. Then the non-negativeness of $g(x, y)$ implies that the sequence $\left\{\int_0^\infty |y_m'|^2 dx\right\}$ is bounded, and since $y_m(0) = y_0$ for every m, as in the proof of Theorem 4.1, the $y_m(x)$ are uniformly bounded on each finite interval. Hence, by Lemma 2.2 of Part I, there is a subsequence, say $\{y_m(x)\}$ again, and an absolutely continuous function $y(x)$, such that on each finite interval $y_m(x) \to y(x)$ uniformly, and $y_m'(x) \to y'(x)$ weakly. Now for any $A > 0$ we have $I(y_m; 0, \infty) \geq I(y_m; 0, A)$; moreover, as in § 4 we have

$$I(y_m; 0, A) - I(y; 0, A) \geq 2\int_0^A [(g(x, y_m) - g(x, y)) + (y_m' - y')^* y']dx$$

and consequently $\liminf_{m\to\infty} I(y_m; 0, A) \geq I(y; 0, A)$. Hence

$$\bar{I} = \lim_{m\to\infty} I(y_m; 0, \infty) \geq I(y; 0, A) , \qquad A > 0 ,$$

and finally,

$$\bar{I} \geq I(y; 0, \infty) = \lim_{A\to\infty} I(y; 0, A) .$$

In particular, this latter relation implies that $y(x)$ is in K, and in view of the definition of $\bar{I}$ we have $I(y; 0, \infty) \geq \bar{I}$, so that $I(y; 0, \infty) = \bar{I}$. That is, $y(x)$ minimizes $I(y; 0, \infty)$ in the class K.

Now on any finite interval $0 \leq x \leq A$, the thus determined $y(x)$ must coincide with the unique vector function which minimizes $I(y; 0, A)$ in the class $K(0, A)$ of curves joining $(0, y_0)$ and $(A, y(A))$, for otherwise one could piece together a curve which would give $I(y; 0, \infty)$ a smaller value than does $y(x)$. By Theorem 4.1 it then follows that $y(x)$ satisfies $y''(x) = g_y(x, y(x))$ on $0 < x \leq A$, where A is arbitrary, and consequently $y''(x) = g_y(x, y(x))$ on $0 < x < \infty$. Since $I(y; 0, \infty)$ is finite, Lemma 3.1 implies that $y(x)$ is bounded on $0 \leq x < \infty$ and therefore is a solution of (2.1). The uniqueness of this solution follows from Theorem 3.1, Inasmuch as we have actually shown that any $y(x)$ that minimizes $I(y; 0, \infty)$ in K is a solution of (2.1), the uniqueness of $y(x)$ as a minimizing function follows from its uniqueness as a solution of (2.1).

6. A further characterization of solutions of (2.1).

THEOREM 6.1. *Suppose that hypotheses H_1–H_5 are satisfied, and $y_\infty(x)$ is the unique solution of* (2.1) *guaranteed by Theorem* 5.1. *If, for a given vector, ξ, $y = y_N(x, \xi)$, $0 \le x \le N$, is the solution of*

$$(6.1a) \qquad y'' = g_y(x, y(x)) \,,$$

$$(6.1b) \qquad y(0) = y_0 \,, \qquad y(N) = \xi \,, \qquad N = 1, 2, \cdots,$$

then $y_N(x, \xi) \to y_\infty(x)$ and $y'_N(x, \xi) \to y'_\infty(x)$ uniformly on each subinterval $0 \le x \le A$.

We will suppose in what follows that the definition of $y_N(x, \xi)$ has been extended so that $y_N(x, \xi) = \xi$ for $x \ge N$. The inequality $(|y_N(x, \xi)|^2)'' \ge 0$ and the end conditions (6.1b) then imply that

$$(6.2) \qquad |y_N(x, \xi)| \le \operatorname{Max}(|y_0|, |\xi|) \,, \quad 0 \le x < \infty \,, \quad N = 1, 2, \cdots.$$

Moreover, the identity

$$(6.3) \qquad y'_N(x, \xi) = \frac{1}{A}\left[y_N(A, \xi) - y_0 - \int_0^A ds \int_x^s g_y(t, y_N(t, \xi))dt\right] \,,$$
$$0 \le x \le A \,, \quad N > A \,,$$

shows that the sequence $\{|y'_N(x, \xi)|\}$ is uniformly bounded on $0 \le x \le A$. Consequently, the sequence $\{y_N(x, \xi)\}$ is uniformly bounded and equicontinuous on any finite interval, so that we may select a subsequence $\{y_{N_j}(x, \xi)\}$ which converges uniformly on any finite interval to a continuous function $y(x)$. From (4.1) it follows that if $\eta(x)$ is of class C'' on $0 \le x < \infty$, and $\eta(0) = 0 = \eta'(0) = \eta'(A)$, $\eta(x) \equiv 0$ for $x \ge A$, then

$$\int_0^A \eta''^*\left[y_{N_j}(x, \xi) - \int_0^x ds \int_0^s g_y(t, y_{N_j}(t, \xi))dt\right]dx = 0 \,, \quad N > A \,.$$

Since $y_{N_j}(x, \xi) \to y(x)$ uniformly on $0 \le x \le A$, we then have

$$\int_0^A \eta''^*\left[y(x) - \int_0^x ds \int_0^s g_y(t, y(t))dt\right]dx = 0 \,.$$

As before, application of the fundamental lemma of the calculus of variation yields the result that $y''(x)$ exists and $y'' = g_y(x, y)$ for $0 < x \le A$. Since A is arbitrary, it follows that $y''(x) = g_y(x, y(x))$ on $0 < x < \infty$. Moreover, $y(0) = y_0$, while the relation (6.2) shows that $y(x)$ is bounded on $0 \le x < \infty$, so that in view of Theorem 5.1 we have $y(x) = y_\infty(x)$.

Now for $0 \le x_0 < \infty$, let η be any accumulation point of the bounded sequence $\{y_N(x_0, \xi)\}$, and let a subsequence $\{y_{N_i}(x_0, \xi)\}$ be chosen such that $y_{N_i}(x_0, \xi) \to \eta$. Then, as before, the sequence $\{y_{N_i}(x, \xi)\}$ is uniformly bounded and equicontinuous on any finite interval, so that we may select a subsequence which approaches $y_\infty(x)$ on $0 \leq x < \infty$. Consequently, the sequence $\{Y_N(x_0, \xi)\}$ has only one accumulation point, namely $\eta = y_\infty(x_0)$, from which it follows that $y_N(x, \xi) \to y_\infty(x)$ for $0 \leq x < \infty$.

With $\zeta_N(x) = y_N(x, \xi) - y_\infty(x)$, as in the proof of Theorem 3.1 we have that $(|\zeta_N(x)|^2)' \geq 0$, $0 \leq x \leq N$. This implies that for any $A > 0$ and $N > A$ we have $|\zeta_N(x)| \leq |\zeta_N(A)|$ on $0 \leq x \leq A$, and thus $y_N(x, \xi) \to y_\infty(x)$ uniformly on $0 \leq x \leq A$.

The fact that $y'_N(x, \xi) \to y'_\infty(x)$ uniformly on $0 \leq x \leq A$ now follows from (6.3), and the corresponding identity obtained by replacing $y_N(x, \xi)$ by $y_\infty(x)$.

7. Asymptotic behavior of solutions of (2.1). At this point we introduce the following hypotheses:

H_6. *For each $c > 0$ there is an $x_c \geq 0$ and a $\Psi(x, c) \geq 0$ with $x\Psi(x, c)$ integrable on every finite subinterval of $x_c \leq x < \infty$, $\int_{x_c}^{\infty} x\Psi(x, c)dx = \infty$, and such that for $x \geq x_c$, $|y| \geq c$ we have $y^* g_y(x, y) \geq \Psi(x, c)$.*

H_7. *If $y(x)$ is in C' and $|y(x)| \geq c > 0$ for $0 \leq x < \infty$, then $I(y(x); 0, \infty) = \infty$.*

We have the following result.

THEOREM 7.1. *If in addition to H_1–H_5, either H_6 or H_7 is also satisfied, then any solution of* (2.1) *approaches zero as $x \to \infty$.*

If $y(x)$ is a solution of (2.1), then $(|y|^2)'' = 2y^*y'' + 2|y'|^2 \geq 0$. Since $y(x)$ is bounded on $0 \leq x < \infty$, it follows that $(|y|^2)' \leq 0$, so that either $|y(x)|$ is bounded away from zero or else $y(x) \to 0$. If H_7 is satisfied then, in view of the fact that $I(y(x); 0, \infty)$ is finite for $y(x)$ a solution of (2.1), it follows that $|y(x)|$ cannot be bounded from zero; that is, $y(x) \to 0$.

Suppose now that H_6 is satisfied. As was noted in the preceding paragraph, $(|y|^2)'$ is non-decreasing and non-positive, so that $\lim_{x\to\infty}(|y|^2)'$ exists. This limit is zero, since $|y|^2$ is non-negative, and hence $\lim_{x\to\infty} y^*y' = 0$. This fact leads to the following relations,

$$-2y^*(x)y'(x) = \int_x^\infty (|y|^2)''dt = 2\int_x^\infty (y^*y'' + |y'|^2)dt ,$$

$$-2y^*(x)y'(x) = 2\int_x^\infty (y^*(t)g_y(t, y(t)) + |y'(t)|^2)dt ,$$

$$-y^*(x)y'(x) \geq \int_x^\infty y^*(t)g_y(t, y(t))dt .$$

Integration now yields

$$\frac{1}{2}|y(x)|^2 - \frac{1}{2}|y(A)|^2 \geq \int_x^A ds \int_s^\infty y^* g_y dt ,$$

and hence

$$\frac{1}{2}|y(x)|^2 \geq \int_x^A ds \int_s^\infty y^* g_y dt .$$

Finally, upon integration by parts we obtain

$$\frac{1}{2}|y(x)|^2 \geq A\int_A^\infty y^* g_y dt - x\int_x^\infty y^* g_y dt + \int_x^A s y^*(s) g_y(s, y(s)) ds .$$

If there is a $c > 0$ such that $|y(x)| \geq c$ on $x_c \leq x < \infty$, then by H_6 it follows that for all x and A satisfying $x_c \leq x < A < \infty$

$$\frac{1}{2}|y(x)|^2 \geq \int_x^A s\Psi(s, c) ds - x\int_x^\infty y^* g_y(t, y(t)) dt .$$

But this implies that $\int_x^\infty s\Psi(s, c)ds < \infty$, contrary to assumption. Thus, there is no $c > 0$ such that $|y(x)| \geq c$ on an interval of the form $x_c \leq x < \infty$, and since $|y(x)|$ is non-increasing it follows that $|y(x)| \to 0$ as $x \to \infty$.

In connection with the comments in §1 of this paper, it is to be noted that the hypotheses used in [8] to establish the analogue of our Theorem 7.1 correspond to the assumption that the $\Psi(x)$ of H_6 satisfies $\int_0^\infty \Psi(x)dx = \infty$, instead of the weaker requirement made here.

For the next two theorems we will make use of the following hypothesis.

H_8. *There exists a function $\phi(x)$ such that*

$$|g_y(x, y_1) - g_y(x, y_2)| \leq \phi(x)|y_2 - y_1| ,$$
$$\textit{for } 0 \leq x < \infty, \ |y_1| < \infty , \ |y_2| < \infty ,$$

where $\phi(x) \geq 0$, $x\phi(x)$ is integrable on any finite subinterval of $0 \leq x < \infty$, and $\int_0^\infty x\phi(x)dx < \infty$.

We prove the following theorem:

THEOREM 7.2. *If $g(x, y)$ satisfies H_1, H_2, H_5, H_8, and $g_y(x, 0) \equiv 0$, and if α is any constant vector, then there is a unique solution $y(x)$ of $y'' = g_y(x, y)$ for which $y(x) \to \alpha$ as $x \to \infty$.*

Let $G(x) = \int_x^\infty t\phi(t)dt$. Imitating Hille [3; p. 238], we consider the following successive approximations corresponding to a given vector α,

$$y_0(x) \equiv \alpha , \qquad 0 \leq x < \infty .$$
$$y_k(x) = \alpha + \int_x^\infty (t - x) g_y(t, y_{k-1}(t)) dt , \qquad 0 \leq x < \infty .$$

We will show by induction that for $0 \leq x < \infty$,

$$
(7.1)\quad \begin{aligned} &\text{(a)}\quad y_k(x) \textit{ is defined};\\ &\text{(b)}\quad |y_k(x) - y_{k-1}(x)| \leq |\alpha| \frac{[G(x)]^k}{k!} \leq \frac{|\alpha|[G(0)]^k}{k!}, \qquad k = 1, 2, \cdots. \end{aligned}
$$

We have $|y_1(x) - y_0(x)| = \left|\int_0^\infty (t - x)g_y(t, \alpha)dt\right|$. The integral here exists since on $x \leq t < \infty$ we have $|t - x||g_y(t, \alpha)| \leq t|g_y(t, \alpha)| \leq t\phi(t)|\alpha|$, which is integrable on $x \leq t < \infty$. Moreover,

$$
|y_1(x) - y_0(x)| \leq |\alpha| \int_x^\infty t\phi(t)dt = |\alpha| G(x),
$$

so that (7.1) is satisfied for $k = 1$.

Suppose (7.1) is true for $k = 1, 2, \cdots, N - 1$. Then $y_N(x)$ is defined, since $|g_y(t, y_{N-1}(t)| \leq \phi(t)|y_{N-1}(t)|$, where $y_{N-1}(t)$ is bounded on $0 \leq t < \infty$. Moreover,

$$
\begin{aligned}
|y_N(x) - y_{N-1}(x)| &= \left|\int_x^\infty (t - x)(g_y(t, y_{N-1}(t)) - g_y(t, y_{N-2}(t)))dt\right| \\
&\leq \int_k^\infty t\phi(t)|y_{N-1}(t) - y_{N-2}(t)|dt \\
&\leq \frac{|\alpha|}{(N-1)!}\int_x^\infty t\phi(t)G^{N-1}(t)dt .
\end{aligned}
$$

Since $G^{N-1}(t)$ is bounded, all the integrals above exist. Hence,

$$
|y_N(x) - y_{N-1}(x)| \leq \frac{-|\alpha|}{(N-1)!}\int_x^\infty [G(t)]^{N-1}G'(t)dt = \frac{|\alpha|[G(x)]^N}{N!} .
$$

Now $y_N(x) - \alpha = (y_1 - y_0) + (y_2 - y_1) + \cdots + (y_N - y_{N-1})$, and the series $\sum_{k=1}^\infty |y_k(x) - y_{k-1}(x)|$ converges uniformly on $0 \leq x < \infty$ by (7.1b). Hence $y(x) = \lim_{N\to\infty} y_N(x)$ exists; moreover $y(x)$ is continuous on $0 \leq x < \infty$ and satisfies $|y(x)| \leq |\alpha| \exp\{G(x)\}$. Therefore $|y(x)|$ is bounded on $0 \leq x < \infty$, and H_8 with the uniform convergence of $\{y_N(x)\}$ on $0 \leq x < \infty$ implies $y(x) = \alpha + \int_x^\infty (t - x)g_y(t, y(t))dt$, so that

$$
\begin{gathered}
y''(x) = g_y(x, y(x)), \qquad 0 < x < \infty, \\
\lim_{x\to\infty} y(x) = \alpha .
\end{gathered}
$$

If $Y(x)$ satisfies $Y''(x) = g_y(x, Y(x))$ on $0 < x < \infty$ and $Y(x) \to \beta$ as $x \to \infty$, then the integral $\int_x^\infty (t - x)g_y(t, Y(t))dt$, $0 \leq x < \infty$, exists and $\eta(x) \equiv Y(x) - \beta - \int_x^\infty (t - x)g_y(t, Y(t))dt$ is such that $\eta''(x) = 0$, $0 < x < \infty$, and $\eta(x) \to 0$ as $x \to \infty$. Hence, $\eta(x) \equiv 0$ and $Y(x) = \beta + \int_x^\infty (t - x)g_y(t, Y(t))dt$. With $y(x)$ as above we then have

$$\begin{aligned} |y(x) - Y(x)| &= \left|\alpha - \beta + \int_x^\infty (t - x)[g_y(t, y(t)) - g_y(t, Y(t))]dt\right| , \\ &\leq |\alpha - \beta| + \int_x^\infty t\phi(t)\,|y(t) - Y(t)|\,dt , \end{aligned}$$

so that by a simple modification of Gronwall's lemma, (see [1; p. 35]), it follows that

$$|y(x) - Y(x)| \leq |\alpha - \beta| \exp\{G(x)\} .$$

If $\beta = \alpha$ then $Y(x) \equiv y(x)$, which proves the uniqueness of solutions of $y'' = g_y(x, y)$ with given limit as $x \to \infty$. Moreover, $|y(x) - Y(x)| \leq |\alpha - \beta| \exp\{G(0)\}$, so that we have the following corollary.

COROLLARY 7.1. *The solution $y(x)$ described in Theorem 7.2 is a continuous function of $\alpha = y(\infty)$.*

We now prove the following theorem on the order of growth of solutions.

THEOREM 7.3. *If $g(x, y)$ satisfies H_1, H_2, H_5, H_8, and $g_y(x, 0) \equiv 0$, and if $y(x)$ satisfies $y'' = g_y(x, y)$ on $0 < x < \infty$, then $\eta = lim_{x\to\infty} y'(x)$ exists and is finite, and $y(x) = x[\eta + o(1)]$.*

Note the H_8 implies $|g_y(x, y)| \leq \phi(x)\,|y|$, which is all that is needed here. If $y(x)$ satisfies $y'' = g_y(x, y)$, then following Bellman [1; p. 114] we write

$$y(x) = y(0) + xy'(0) + \int_0^x (x - t)g_y(t, y(t))dt .$$

Hence, for $x \geq 1$,

$$|y(x)| \leq x(|y(0)| + |y'(0)|) + x\int_0^x \phi(t)\,|y(t)|\,dt$$

or

$$\frac{|y(x)|}{x} \leq (|y(0)| + |y'(0)|) + \int_0^x t\phi(t)\frac{|y(t)|}{t}dt .$$

Therefore, by Gronwall's lemma, (see [1; p. 35]),

$$\frac{|y(x)|}{x} \leq (|y(0)| + |y'(0)|) \exp\left\{\int_0^x t\phi(t)dt\right\} ,$$

and hence there is a constant M such that

$$|y(x)| \leq Mx , \qquad x \geq 1 .$$

Now for $x \geq 1$ we have

$$\int_1^x |g_y(t, y(t))|\,dt \le \int_1^x \phi(t)\,|y(t)|\,dt \le M\int_1^x t\phi(t)dt\,,$$

so that $\int_0^\infty |g_y(t, y(t))|\,dt$ exists. Since

$$y'(x) = y'(0) + \int_0^x g_y(t, y(t))dt\,,$$

we have that $y'(x) \to \eta$ as $x \to \infty$, where

$$\eta = y'(0) + \int_0^\infty g_y(t, y(t))dt\,.$$

The final equality in the theorem is a ready consequence of this finite limit of $y'(x)$.

8. Behavior of solutions of (2.1) with respect to initial values. We continue to suppose here that H_1–H_5 are satisfied, but not necessarily any other hypotheses. Let $y_1(x)$, $y_2(x)$ be two bounded solutions of $y'' = g_y(x, y)$ on $0 \le x < \infty$, and set $\eta(x) = y_1(x) - y_2(x)$. Then by H_4, we have $\eta^*\eta'' \ge 0$, so that $(|\eta|^2)'' = 2|\eta'|^2 + 2\eta^*\eta'' \ge 0$. Since $\eta(x)$ is bounded, we must have $|\eta(x)|$ non-increasing; in particular, $|\eta(x)| \le |\eta(0)|$ on $0 \le x < \infty$. Suppose now we denote by $y(x;\alpha)$ the unique bounded solution of $y'' = g_y(x, y)$ which satisfies $y(0;\alpha) = \alpha$. Then $y(x;\alpha)$ is continuous in x and α jointly, as may be seen from the inequality

$$\begin{aligned} |y(\bar{x};\bar{\alpha}) - y(x, \alpha)| &\le |y(\bar{x};\bar{\alpha}) - y(\bar{x};\alpha)| + |y(\bar{x};\alpha) - y(x;\alpha)|\,, \\ &\le |\bar{\alpha} - \alpha| + |y(\bar{x};\alpha) - y(x;\alpha)|\,. \end{aligned}$$

Moreover, for any $A > 0$,

$$(8.1)\qquad y'(x;\alpha) = \frac{1}{A}\left[y(A;\alpha) - y(0;\alpha) - \int_0^A ds\int_x^s g_y(t, y(t;\alpha))dt\right],$$

so that $y'(x;\alpha)$ is also continuous in x and α.

We turn now to the question of differentiability of solutions with respect to initial values. The derivation of our results will involve the use of a lemma, the proof of which is based on certain theorems due to W. T. Reid. In [6], Reid has discussed a class of non-oscillatory linear matrix differential equations which includes as a special case the matrix equation

$$(8.2)\qquad U'' = P(x)U\,, \qquad a \le x < \infty$$

where $P(x)$ is a non-negative definite symmetric matrix with continuous real-valued elements. As shown in Theorem 6.1 of [6], if $U(x)$ is a solution of (8.2) which is non-singular on a subinterval $b \le x < \infty$, and the necessarily constant matrix $U^*(x)U'(x) - U^{*\prime}(x)U(x)$ is the zero matrix, then

$$M(b; U) = \lim_{x\to\infty}\left(\int_b^x U^{-1}(t)U^{*-1}(t)dt\right)^{-1}$$

exists and is finite. Moreover, such a $U(x)$ is a *principal solution* of (8.2) in the sense of Reid [6] if and only if $M(b; U) = 0$. In addition, a principal solution $U(x)$ is characterized by $U(x) = U_{b,\infty}(x)C$, where C is a non-singular constant matrix and $U_{b,\infty} = \lim_{t\to\infty} U_{b,t}(x)$, where $U_{b,t}(x)$, $t > b$, is the unique solution of (8.2) satisfying $U_{b,t}(b) = E$, $U_{b,t}(t) = 0$.

It follows as a special case of Theorem 5.1 of this paper that the vector system

$$(8.3)\qquad \begin{gathered} u'' = A(x)u\ , \qquad 0 \le x < \infty \\ u(0) = u_0\ , \quad |u(x)| \text{ bounded on } 0 \le x < \infty\ , \end{gathered}$$

where $A(x)$ is a real symmetric non-negative matrix of functions continuous on $0 \le x < \infty$, has a unique solution. Moreover, Theorem 6.1 shows that the solution $u(x)$ of (8.3) is the limit, as $N \to \infty$, of a function $u_N(x)$ satisfying $u_N'' = A(x)u_N$, $u_N(0) = u_0$, $\mu_N(N) = 0$, $N = 1, 2, \cdots$. In view of the similar characterization of this solution and of the principal solutions described above, it follows that the column vectors of $U(x)$, where $U(x)$ is a principal solution of $U'' = A(x)U$, are particular bounded solutions of $u'' = A(x)u$. This fact will be used in the proof of the following lemma.

LEMMA 8.1. *Suppose $A(x; h)$ is an $n \times n$ non-negative definite symmetric matrix, continuous jointly in the scalar x and the vector h for $0 \le x < \infty$ and h in some open set H. Let $W_h(x)$ be the unique principal solution of*

$$W_h''(x) = A(x; h)W_h(x)\ ,$$

satisfying

$$W_h(0) = E\ .$$

Then, if h_0 is in H we have $\lim_{h\to h_0} W_h(x) = W_{h_0}(x)$, uniformly for x on any interval $0 \le x \le X$.

We consider the solutions $U = U_h(x)$ of the system

$$(8.4)\qquad \begin{gathered} U'' = A(x; h)U \\ U(0) = E\ , \qquad U'(0) = E \end{gathered}$$

or, equivalently,

$$(8.5)\qquad \begin{gathered} U' = V \\ V' = A(x; h)U \\ U(0) = E\ , \qquad V(0) = E\ . \end{gathered}$$

The latter is of the form (2.4′) of [6] with $A = 0$, $B = E$, $C = A(x; h)$. The solution of (8.4) is non-singular on $0 \leq x < \infty$, since if ξ is a constant vector such that $u = U(x)\xi$ satisfies $u(x_0) = 0$ with $x_0 > 0$, then

$$\begin{aligned} 0 &= \int_0^{x_0} u^*(u'' - Au)dx \ , \\ &= u^*u'\Big|_0^{x_0} - \int_0^{x_0}(|\,u'\,|^2 + u^*Au)dx \ , \\ &= -|\,\xi\,|^2 - \int_0^{x_0}(|\,u'\,|^2 + u^*Au)dx \ , \end{aligned}$$

so that $\xi = 0$.

Continuing to use the notation of [6; § 3], we compute the value of the constant matrix $\{U, U\} = U^*(x)V(x) - V^*(x)U(x)$ to be $U^*(0)V(0) - V^*(0)U(0) = 0$, and we find that $T = E$. By Theorem 3.1 of [6] we know that any solution $Y(x)$ of $Y'' = A(x)Y$ with $Y(0) = E$ has the form

$$Y(x) = U(x)\Big[E + \Big(\int_0^x U^{-1}(t)U^{*-1}(t)dt\Big)K_0\Big] \ ,$$

for some constant matrix K_0.

Now by Theorems 5.1 and 6.1 of [6] we have $W_h(x) = \lim_{N\to\infty} Y_{0N}(x)$, where $Y''_{0N} = A(x, h)Y_{0N}$, $Y_{0N}(0) = E$, $Y_{0N}(N) = 0$. But in view of the boundary conditions satisfied by $Y_{0N}(x)$ we have

$$Y_{0N}(x) = U(x)\Big[E + \Big(\int_0^x U^{-1}(t)U^{*-1}(t)dt\Big)K_0\Big] \ ,$$

with

$$K_0 = -\Big(\int_0^N U^{-1}(t)U^{*-1}(t)dt\Big)^{-1} \ .$$

Hence,

$$Y_{0N}(x) = U(x)\Big[E - \Big(\int_0^x U^{-1}U^{*-1}dt\Big)\Big(\int_0^N U^{-1}U^{*-1}dt\Big)^{-1}\Big] \ ,$$

and finally,

$$W_h(x) = U_h(x)\Big[E - \Big(\int_0^x U_h^{-1}U_h^{*-1}dt\Big)M(0; U_h)\Big] \ . \tag{8.6}$$

We now need an estimate of $U_h^{-1}(x)U_h^{*-1}(x)$ for large x. To this end put $Z_h(x) = (1 + x)^{-1}U_h(x)$. In view of (8.4), one readily verifies that

$$((1 + x)^2Z_h')' - (1 + x)^2A(x; h)Z_h = 0 \ , \quad Z_h(0) = E \ , \quad Z_h'(0) = 0 \ .$$

From this fact it follows that

$$0 = \int_0^x Z_h^*[((1+t)^2 Z_h')' - (1+t)^2 A(t;h) Z_h]dt$$
$$= (1+t)^2 Z_h^* Z_h' \Big|_0^x - \int_0^x (1+t)^2 [Z_h^{*\prime} Z_h' + Z_h^* A(t;h) Z_h]dt\ ,$$

and therefore

$$(1+x)^2 Z_h^*(x) Z_h'(x) = \int_0^x (1+t)^2 [Z_h^{*\prime} Z_h' + Z_h^* A Z_h]dt\ .$$

Consequently, $(Z_h^* Z_h)' = Z_h^* Z_h' + Z_h^{*\prime} Z_h = 2Z_h^* Z_h' \geq 0$ on $0 \leq x < \infty$, and $Z_h^*(x) Z_h(x) \geq Z_h^*(0) Z_h(0) = E$ for $x \geq 0$; that is, $U_h^*(x) U_h(x) \geq (1+x)^2 E$ and hence $U_h^{-1}(x) U_h^{*-1}(x) \leq (1+x)^{-2} E$ on $0 \leq x < \infty$ for h in H.

Since as $h \to h_0$ we have $U_h(x) \to U_{h_0}(x)$ uniformly on each finite interval $0 \leq x \leq X$, it follows that

$$\lim_{h \to h_0} M(0; U_h) = M(0; U_{h_0})\ .$$

This result, with (8.6), proves the lemma.

We can now prove the following theorem:

THEOREM 8.1. *If* $g_{yy}(x, y) = \| g_{y_i y_j} \|$ *exists and is continuous for* (x, y) *in* Ω: $0 \leq x < \infty$, $|y| < \infty$, *and if* $g(x, y)$ *satisfies* H_1–H_5, *then with* $y(x; \alpha)$ *as in the beginning of this section, we have that* $\partial y(x; \alpha)/\partial \alpha_j$ *and* $\partial y'(x; \alpha)/\partial \alpha_j$ *exist and are continuous in* x, α *for* $0 \leq x < \infty$, $|\alpha| < \infty$, $j = 1, 2, \cdots, n$.

Note that if the hypotheses of this theorem are satisfied, then $g_{yy}(x, y) \geq 0$ for (x, y) in Ω. We denote by $e^{(j)}$ the unit vector having all components zero but the jth, and we let $\Delta\alpha = e^{(j)} h$, $\Delta y = y(x; \alpha + \Delta\alpha) - y(x; \alpha)$, where h is a real scalar. Then

$$(\Delta y)'' = g_y(x, y(x; \alpha + \Delta\alpha)) - g_y(x, y(x; \alpha))$$
$$= \left(\int_0^1 g_{yy}(x, y(x; \alpha) + \theta \Delta y) d\theta\right) \Delta y\ ,$$

so that

$$\left(\frac{\Delta y}{h}\right)'' = \left(\int_0^1 g_{yy}(x, y(x; \alpha) + \theta \Delta y) d\theta\right)\left(\frac{\Delta y}{h}\right), \quad h \neq 0\ .$$

In Lemma 8.1 we identity $A(x; h)$ as $\int_0^1 g_{yy}(x, y(x; \alpha) + \theta \Delta y) d\theta$, where α is fixed, and we identify h_0 as zero. We note that

$$\left(\frac{\Delta y}{h}\right)_{x=0} = e^{(j)}\ , \quad \text{and} \quad \left|\frac{\Delta y(x)}{h}\right| = \frac{|\Delta y(x)|}{|h|} \leq \frac{|\Delta y(0)|}{|h|} \leq 1\ ,$$

$0 \leq x < \infty$. Hence, $(1/h)\Delta y$ is the unique bounded solution of $z'' = A(x; h) z$ satisfying $z(0) = e^{(j)}$. As explained above, the unique principal solution of

$$(8.7') \qquad Z_h'' = A(x; h)Z_h ,$$

satisfying

$$(8.7'') \qquad Z_h(0) = E ,$$

is the same as the bounded solution of (8.7'), (8.7'') guaranteed by Theorem 6.1 of this paper, of which $(1/h)\Delta y$ is the jth column vector, for $h \neq 0$. Lemma 8.1 then implies that $\lim_{h\to 0}(1/h)\Delta y(x)$ exists and is equal to the jth column vector of $Z_0(x)$; that is, for all α, the vector function $y_{\alpha_j}(x; \alpha) = (\partial/\partial\alpha_j)y(x; \alpha)$ exists and satisfies

$$(8.8) \qquad (y_{\alpha_j}(x; \alpha))'' = g_{yy}(x, y(x; \alpha))y_{\alpha_j}(x; \alpha) ; \qquad 0 \leq x < \infty .$$

Since $|y_{\alpha_j}(x; \alpha)| \leq 1$, we may use Lemma 8.1 with $h = \alpha$ in conjunction with the inequality

$$|y_{\alpha_j}(\bar{x}, \bar{\alpha}) - y_{\alpha_j}(x, \alpha)| \leq |y_{\alpha_j}(\bar{x}; \bar{\alpha}) - y_{\alpha_j}(\bar{x}; \alpha)| + |y_{\alpha_j}(\bar{x}; \alpha) - y_{\alpha_j}(x; \alpha)|$$

to show that $y_{\alpha_j}(x; \alpha)$ is continuous in x and α. Differentiation of the right hand member of (8.1) with respect to α_j shows the existence of $(\partial/\partial\alpha_j)y'(x; \alpha)$ and its continuity with respect to x and α.

BIBLIOGRAPHY

1. R. Bellman, *Stability theory of differential equations*, McGraw-Hill, New York, 1953.
2. G. H. Hardy, J. E. Littlewood and G. Pólya, *Inequalities*, 2nd ed., Cambridge University Press, London, 1952.
3. E. Hille, *Non-oscillation theorems*, A.M.S. Trans., **64** (1948), 234-252.
4. E. L. Ince, *Ordinary differential equations*, Longmans, Green and Co. Ltd., London, 1927.
5. W. T. Reid, *Oscillation criteria for linear differential systems with complex coefficients*, Pacific J. Math., **6** (1956), 733-751.
6. ———, *Principal solutions of non-oscillatory self-adjoint linear differential systems*, Pacific J. Math., **8** (1958), 147-169.
7. F. Riesz and B. Sz.-Nagy, *Leçons d'Analyse fonctionnelle*, 2d ed., Budapest, 1953.
8. G. Sansone, *Equazioni differenziali nel Campo reale*, part 2, 2d ed., Bolonga, 1949.

LOYOLA UNIVERSITY
CHICAGO

A CLASS OF HYPER-FC-GROUPS

A. M. DUGUID

1. Introduction. An element g of an arbitrary group G is called an FC element if it has a finite number of conjugates in G. The set of all FC elements of G forms a characteristic subgroup H of G (see Baer [1]). The *upper FC-series* of G, introduced by Haimo [4] as the *FC-chain,* may be defined by

$$H_0 = \{1\} ,$$
$$H_{i+1}/H_i = H(G/H_i) ,$$

the subgroup of all FC elements of G/H_i. The upper FC-series is continued transfinitely in the usual way, by defining

$$H_\alpha = \bigcup_{\beta<\alpha} H_\beta ,$$

when α is a limit ordinal. If $H_\gamma = G$, but $H_\delta \neq G$, for all $\delta < \gamma$, we say that the group G is *hyper-FC of FC-class* γ, following McLain [7].

A group G in which the transfinite upper central series

$$\{1\} = Z_0 \leq Z_1 \leq \cdots \leq Z_\alpha \leq \cdots$$

reaches the whole group is called a ZA-group (Kurosh [6]), and we may say that G has class α if $Z_\alpha = G$, but $Z_\beta \neq G$, for all $\beta < \alpha$. Glushkov [3] and McLain [7] have given constructions for a ZA-group of any given class. The main object of this note is to construct groups of given FC-class.

2. Constructions and proofs.

DEFINITION. We say that a group G is of type Q_α if

1. G has FC-class α, with upper FC-series
$\{1\} = H_0 \leq H_1 \leq \cdots \leq H_\alpha = G$,
2. $H_{\gamma+1}/H_\gamma$ is infinite, for all $\gamma < \alpha$, and
3. $H_{\gamma+1}/H_\gamma$ has the unit subgroup for its centre, for all $\gamma < \alpha$.

Thus the group with only one element is of type Q_0, and, in constructing a group G of type Q_α, we may assume the existence of a group G_β of type Q_β, for each $\beta < \alpha$. If α is a limit ordinal, we define G to be the ordinary (restricted) direct product of the groups G_β, for all $\beta < \alpha$. Then G has the properties $1 - 3$, and thus has type Q_α. For the case $\alpha = \beta + 1$ we shall construct G by 'wreathing' the regular

Received March 5, 1959.

representation of G_β with a certain kind of infinite centreless FC-group of permutations of the positive integers. (For convenience, we say that a group is *centreless* if its centre consists of the unit element alone.)

DEFINITION. A faithful representation of a group G by permutations of the positive integers will be called a *special* representation of G if

(i) the stabiliser of each integer has finite index in G and
(ii) the intersection of the stabilisers of the elements of any set of all but a finite number of these integers is the unit subgroup.

DEFINITION. An infinite centreless FC-group possessing a special representation will be called a group of type F.

To construct an example of a group of type F, let $D = B_1 \times B_2 \times \cdots$ be the ordinary direct product of an infinite sequence of finite centreless groups B_i, $i = 1,2, \cdots$. Let $D_n = B_{n+1} \times B_{n+2} \times \cdots$, let k_n be the order of D/D_n and let the elements of D/D_n, in an arbitrary order, be

$$X_1^n\,,\ X_2^n\,,\ \cdots,\ X_{k_n}^n\,.$$

For each element $g \in D$ and each $n = 1, 2, \cdots$, define the permutation π_{gn} of the integers $1, 2, \cdots, k_n$ by the rule

$$\pi_{gn}(i) = j \text{ when } gX_i^n = X_j^n\,. \tag{1}$$

Now, for each $g \in G$, define the permutation π_g of the positive integers by the rule

$$\pi_g\Big(i + \sum_{j=1}^{n-1} k_j\Big) = \pi_{gn}(i) + \sum_{j=1}^{n-1} k_j\,, \tag{2}$$

for all $i = 1, 2, \cdots, k_n$, and $n = 1, 2, \cdots$. The systems of transitivity in this permutation representation of D are the sets T_n of integers m such that $\sum_{i=1}^{n-1} k_i < m \le \sum_{i=1}^{n} k_i$, for $n = 1, 2, \cdots$. If $m \in T_n$, then the subgroup D_n of D is contained in the stabiliser of m. Hence the stabiliser in D of each positive integer has finite index in D. On the other hand, suppose g is in the stabiliser in D of all but a finite number of the positive integers. Then there is a number n_0 such that g is in the stabiliser of each integer of each system T_n with $n \ge n_0$. So if i is any integer in the range $1 \le i \le k_n$, $n \ge n_0$, we know that g is in the stabiliser of $i + \sum_{j=1}^{n-1} k_j$, and this means that $gX_i^n = X_i^n$. Thus $g \in D_n$. But the subgroups D_n, with $n \ge n_0$, intersect in the unit subgroup of D. So $g = 1$. We observe also that the permutation representation of D defined by (1) and (2) is faithful. Thus we have a special representation of the infinite centreless FC-group D, which is therefore a group of type F.

LEMMA. *If G_β is a group of type Q_β and J is a group of type F,* *then a group G formed by wreathing the regular representation of G_β with a special representation R of J is a group of type $Q_{\beta+1}$.*

Proof. The wreath group G may be regarded as a semi-direct product

$$G = KE\,, \quad K \cap E = 1\,,$$

where $K = \prod_{i=1}^{\infty} A_i$ is the direct product of a sequence of groups, each isomorphic to G_β, and E is isomorphic to J. The automorphisms of K induced by elements of E permute the subgroups A_i, $i = 1, 2, \cdots$, realizing the special representation R of $J \simeq E$. Associated with G is a set of isomorphisms θ_{ij}, $i, j = 1, 2, \cdots$ such that $\theta_{ij}(A_i) = A_j$, and if $a \in A_i$, $g \in E$ and $g^{-1}A_i g = A_j$, then $g^{-1}ag = \theta_{ij}(a)$. θ_{ii} is the identity automorphism, for all i. (A brief general description of wreath groups, and further references, may be found in Hall [5].)

Let C_i be the set of all elements g in E such that $g^{-1}A_i g = A_i$. Then C_i is the centraliser in E of each element of A_i. Since the representation R is special, the subgroup C_i of E has finite index in E, for each i, and the unit element is the only element of E common to all the subgroups of any set of all but a finite number of the C's.

For all $\gamma \leq \beta$, put $H_\gamma = H_\gamma(K)$, the γth term of the upper FC-series of K. If possible, let $\tau + 1$ be the least such ordinal for which $H_{\tau+1}(G) \neq H_{\tau+1}$. Now any element k of K can be written as the product of a finite number of elements $a_{i_\nu} \in A_{i_\nu}$, $\nu = 1, 2, \cdots, n$, and the subgroup $C(k) = \bigcap_{\nu=1}^{n} C_{i_\nu}$ has finite index in E. But $C(k)$ is contained in the centraliser of k in E, so $g^{-1}kg$, with $g \in E$, is finite valued. Hence

$$H_{\tau+1}(G) \cap K = H_{\tau+1}\,.$$

Suppose $kg \in H_{\tau+1}(G)$, where $k \in K$ and $g \in E$, $g \neq 1$. Let $\sigma + 1$ be the least ordinal in the range $\tau + 1 \leq \sigma + 1 \leq \beta$ such that $k \in H_{\sigma+1}$. Now H_σ is a characteristic subgroup of K, and hence is normal in G, and both kH_σ and kgH_σ are FC elements of G/H_σ. Hence gH_σ is FC in G/H_σ.

We can choose an infinite sequence of distinct positive integers, $\mu_1, \mu_2, \cdots$, such that $g^{-1}A_{\mu_i}g \neq A_{\mu_i}$, for all $i = 1, 2, \cdots$, for otherwise g would belong to all but a finite number of the C's. Moreover, since C_i has finite index in E, for each i, we can choose the sequence μ_1, $\mu_2, \cdots$ so that distinct terms belong to distinct systems of transitivity in the representation R of E. By relabelling the subgroups A_i, $i = 1, 2, \cdots$, we may arrange that the sequence $\mu_1, \mu_2, \cdots$ is just the sequence of odd positive integers. So if n is any odd positive integer, and $g^{-1}A_n g = A_{\dot{n}}$, then $\dot{n}$ is even. Since $\sigma < \beta$, we can choose

$a_n \in A_n - H_\sigma(A_n)$, for $n = 1, 3, \cdots$. Let $a_{\dot{n}} = g^{-1} a_n g$, and define

$$c_n = g^{-1} g^{a_n} = a_{\dot{n}}^{-1} a_n , \qquad n = 1, 3, \cdots .$$

Then

$$c_n^{-1} c_m = (g^{a_n})^{-1} g^{a_m} = a_n^{-1} a_{\dot{n}} a_{\dot{m}}^{-1} a_m .$$

If $n \neq m$, the four integers $n, \dot{n}, m$ and $\dot{m}$ are all distinct and thus $(g^{a_n})^{-1} g^{a_m} \notin H_\sigma$. Thus gH_σ is not FC in G/H_σ, contrary to what we have already proved.

It follows that the upper FC-series of G is

$$\{1\} = H_0 \leq H_1 \leq \cdots \leq H_\beta = K < G ,$$

for $G/K \simeq E \simeq J$, and J is an FC-group. Moreover J is infinite and centreless, and the factors $H_{\gamma+1}/H_\gamma$ are infinite and centreless, for all $\gamma < \beta$, since G_β is a group of type Q_β, and K is a direct product of groups isomorphic with G_β. Thus G is a group of type $Q_{\beta+1}$, as required.

We have now shown how to construct a group of type Q_α, given groups of type Q_β for all $\beta < \alpha$, whether α is a limit ordinal or not. So, by transfinite induction, we have:

THEOREM. *There exist groups of type Q_α, for any ordinal α.*

I should like to express my thanks to Prof. P. Hall of Kings College, Cambridge, who suggested the topic of this paper to me while I was studying under his direction.

REFERENCES

1. R. Baer, *Finiteness properties of groups*, Duke Math. J. **15** (1948), 1021–1032.
2. A. M. Duguid and D.H. McLain, *FC-nilpotent and FC-soluble groups*, Proc. Cambridge Philos. Soc. **52** (1956), 391-398.
3. V. M. Glushkov, *On the central series of an infinite group*, Mat. Sbornik N. S. **31** (1952), 491-495 (Russian).
4. F. Haimo, *The FC-chain of a group*, Canad. J. Math. **5** (1953), 498–511.
5. P. Hall, *Finiteness conditions in soluble groups*, Proc. London Math. Soc. (3) **4** (1954), 419-436.
6. A. G. Kurosh, *The theory of groups*, New York, 1955.
7. D. H. McLain, *Remarks on the upper central series of a group*, Proc. Glasgow Math. Soc. **3** (1956), 38-44.

BROWN UNIVERSITY

THE CALCULATION OF CONFORMAL PARAMETERS FOR SOME IMBEDDED RIEMANN SURFACES

A. M. GARSIA

Introduction. Riemann surfaces were originally introduced as a tool for the study of multiple valued analytic functions. In Riemann's work they appear as covering surfaces of the complex plane with given branch points. Since then Riemann surfaces have been considered from several different aspects.

Here we shall follow the point of view assumed by Beltrami and Klein, who visualized these surfaces as two-dimensional submanifolds of Euclidean space whose conformal structure is defined by the surrounding metric.

Recent results of J. Nash[1] on isometric imbeddings of Riemannian manifolds assure that all models of Riemann surfaces with the natural Poincaré metric can be C^∞ isometrically imbedded in a sufficiently high (51) dimensional Euclidean space. However, the question still remains open whether or not every Riemann surface has a conformally equivalent representative in the ordinary three-dimmensional space.

Although the dimension requirement seems restrictive, there is reason to believe that, since only conformality is required, at least the compact surfaces can be conformally imbedded. We shall not be directly concerned here with this existence problem; instead, we shall present a family of elementary surfaces which may contain all conformal types and whose conformal structure can be easily characterized.

In the genus one case, the conformal structure is usually described by a complex parameter ν which gives the ratio of two principal periods of an abelian differential of the surface. It is always possible to choose these periods so that their ratio ν lies in the region $\mathfrak{M}$ of the Gauss plane defined by the inequalities:

$$\mathfrak{Im}\nu < 0, \ -\tfrac{1}{2} < \mathfrak{Re}\nu \leqq \tfrac{1}{2}; \ |\nu| > 1 \text{ for } \mathfrak{Re}\nu < 0, \ |\nu| \geqq 1 \text{ for } \mathfrak{Re}\nu \geqq 0 .$$

It is well known that every Riemann surface of genus one has in $\mathfrak{M}$ one and only one representative point.

It is easy to verify that the representative points ν of the tori of revolution lie in the imaginary axis and cover it completely. Thus it seems plausible that the affine images of the tori of revolution should cover all conformal types in the genus one case; however, we have

Received February 12, 1959. This research was supported in whole by the United States Air Force No. AF49 (638)-42, monitored by the AF Office of Scientific Research of the Air Research and Development Command.

[1] "The imbedding problem for Riemannian manifolds". *Annals of Mathematics*, 63 (1956), pp. 20-63.

found no proof of this fact. Indeed the characterization of the parameter ν for an imbedded surface leads in general to rather difficult problems.

For this reason, for quite some time there have been no known examples of surfaces whose representative point in $\mathfrak{M}$ lies off the imaginary axis. In 1944, O. Teichmüller[2] proved the existence of these surfaces by showing that there are small deformations of the tori of revolution for which the variation of ν is not purely imaginary.

Led by these observations we have tried to develop a method of uniformizing a given Riemann surface that could be of practical application for some wide enough family of surfaces. To make our considerations applicable to surfaces of higher genus we needed to introduce some parameters to take the role that ν plays in the genus one case. To this end we have adopted as a canonical form of a Riemann surface the result of the Schottky uniformization. In fact, some imbedded surfaces can be considered topologically "marked" in a natural way, and the Schottky uniformization associates with every marked surface of genus g (>1) a complete set of geometrical invariants which can be expressed by means of $3g-3$ independent complex parameters.

In view of the importance of these parameters we deemed necessary to include in the first section of this paper a description of the Schottky uniformization and some general facts associated with it. In the second section we present a definition of "M-surfaces". These are imbedded surfaces which may have edge type singularities along curves but can be made into Riemann surfaces in a natural way. To generate these surfaces we adopt a process which uses surfaces of genus zero as building blocks to construct surfaces of genus one and surfaces of genus one to construct surface of higher genus.

In the third section we present a method of constructing the Schottky uniformization of a given M-surface. This method is more general than it appears in the context since from the existence of the Schottky uniformization, every marked surface can be considered an M-surface (dropping the condition that the building surfaces of genus zero should be globally imbedded.) As will be shown in the fourth section, this method assumes practical importance when the building blocks of M-surfaces are ordinary spheres. These special M-surfaces we have called "natural".

To present our results in this case we made use of anallagmatic coordinates of spheres as introduced by E. Cartain in [2]; for the sake of completeness a brief introduction to these coordinate is also included.

In the last section a few properties of natural M-surfaces of genus

[2] "Beweis der analytischen Abhängigkeit des konformen Moduls einer analytischen Ringflächenschar von den Parametern", Deutsche Math. **7** (1944), 309–336.

one are studied, and some of the results are used to construct the Teichmüller models. At the end a process is given by means of which all natural M-surfaces can be made into C^∞ smooth canal surfaces.

ACKNOWLEDGEMENT

We wish to express here our gratitude towards Professor H. Royden for introducting us to the subject and suggesting these problems and to Professor L. Ahlfors and S. S. Chern for their friendly encouragement and advice.

1. A choice of conformal parameters for compact Riemann surfaces.

1.1 Here and in the following Σ shall denote a given 2-sphere; "a coordinate in Σ" shall mean an extended valued complex coordinate introduced by a stereographic projection of Σ upon the Gauss-plane. Let z be such a coordinate. Since z is defined up to a Moebius transformation of Σ onto itself, we can assume that the points $0, 1, \infty$ are situated wherever we may wish. Whenever it does not lead to ambiguities, we shall make use of the same symbol for a point of Σ and its complex coordinate.

If Λ is a Jordan curve and α a point of Σ not lying in Λ, we shall denote by $\Lambda(\alpha)$ the connected component of $\Sigma - \Lambda$ which contains α. $\Lambda(\alpha)$ will be called the interior of Λ with respect to α. If Λ separates α from another point β of Σ we have of course

$$\Sigma = \Lambda(\alpha) + \Lambda + \Lambda(\beta) .$$

Let now α_i, β_i $(i = 1, 2, \cdots, g)$ be $2g$ distinct points of Σ and ω_i $(i = 1, 2, \cdots, g)$ given complex numbers of absolute value greater than one. Let τ_i be the Moebius transformation of Σ onto itself defined by the equation

$$\frac{\tau_i z - \alpha_i}{\tau_i z - \beta_i} = \omega_i \frac{z - \alpha_i}{z - \beta_i} \tag{1}$$

We assume for a moment that $\alpha_1 = 0$ and $\beta_1 = \infty$. Under this coordinate system we have

$$\tau_1 z = \omega_1 z .$$

Let ρ_1 and ρ_2 be the smallest and the largest of the absolute values

$$|\alpha_i| , \ |\beta_i| \qquad\qquad i = 2, 3, \cdots, g .$$

If $|\omega_1| > (1/\eta)(\rho_2/\rho_1)$ for some $0 < \eta < 1$, a circle with center at 0 and radius $r = \eta\rho_1$ is transformed by τ_1 onto a concentrical circle of radius

$$r' = |\omega_1| r > \rho_2 .$$

Thus if $|\omega_1| > \rho_2/\rho_1$ there are infinitely many circles Λ such that the points $\alpha_2, \beta_2; \cdots; \alpha_g, \beta_g$ are all interior to the anulus

$$\Lambda(\infty) \cap \tau_1\Lambda(0) .$$

Before expressing this fact in an invariant way we shall introduce a notation. If α and β are two distinct points of Σ by $P(\alpha, \beta)$ we shall denote the pencil of circles which admit α, β as a couple of inverse points.

We have thus shown that:

I. *Provided* $|\omega_1|$ *is sufficiently large we can choose a circle* Λ *in an infinite number of ways so that*

(a) $\Lambda \in P(\alpha_1, \beta_1)$

(b) *the points* $\alpha_2, \beta_2; \cdots: \alpha_g, \beta_g$ *are contained in the domain* $\Lambda(\beta_1) \cap \tau_1\Lambda(\alpha_1)$.

Let Λ_1 be one of these circles.

We shall show now that:

II. *Provided the* $|\omega_i|$'s *are sufficiently large the circles* Λ_i *can be chosen in an infinite number of ways so that*

(a) $\Lambda_i \in P(\alpha_i, \beta_i)$

(b) *the closed disks*

$$\overline{\Lambda_1(\alpha_1)},\ \overline{\tau_1\Lambda_1(\beta_1)},\ \cdots,\ \overline{\Lambda_g(\alpha_g)},\ \overline{\tau_g\Lambda_g(\beta_g)}$$

are exterior to each other.

Because of I we can prove II inductively.

Suppose that the circles $\Lambda_1, \Lambda_2, \cdots, \Lambda_{i-1}$ have been chosen in such a way that

(a) $\Lambda_j \in P(\alpha_j, \beta_j)$ $\qquad (j = 1, 2, \cdots, i-1)$,

(b) the closed disks $\overline{\Lambda_1(\alpha_1)}, \overline{\tau_1\Lambda_1(\beta_1)}; \cdots; \overline{\Lambda_{i-1}(\alpha_{i-1})}, \overline{\tau_{i-1}\Lambda_{i-1}(\beta_{i-1})}$ are exterior to each other,

(c) the remaining points α_j, β_j $(j = i, i+i, \cdots, g)$ are contained in the domain

$$\bigcap_{j=1,i-1} \{\Lambda_j(\beta_j) \cap \tau_j\Lambda_j(\alpha_j)\} .$$

We temporarily assume that $\alpha_i = 0$ and $\beta_i = \infty$. We let S be the set consisting of the closed disks

$$\overline{\Lambda_1(\alpha_1)},\ \overline{\tau_1\Lambda_1(\beta_1)},\ \cdots,\ \overline{\Lambda_{i-1}(\alpha_{i-1})},\ \overline{\tau_{i-1}\Lambda_{i-1}(\beta_{i-1})}$$

and (if $i < g$) the points

$$\alpha_{i+1}, \beta_{i+1}, \cdots, \alpha_g, \beta_g .$$

Under this coordinate system let ρ_1 and ρ_2 be the minimum and the maximum value assumed by $|z|$ as z varies in S. Clearly the argument can be completed since, for the same reasons as before, if $|\omega_i| > \rho_2/\rho_1$, the circle Λ can be chosen in an infinite number of ways so that

(a) $\Lambda \in P(\alpha_i, \beta_i)$

(b) the set S is exterior to $\overline{\Lambda(\alpha_i)}$ and $\overline{\tau_i\Lambda(\beta_i)}$. Let Λ_i be one of these circles.

A further investigation on the nature of the inequalites to which the $|\omega_i|$'s are to be subjected, for such a construction to be possible would be of some interest, but for our immediate purposes it is not needed.

We would like to point out, however, that if for a given set of complex numbers $\{\alpha_1, \beta_1, \omega_1; \cdots; \alpha_g, \beta_g, \omega_g\}$ the construction in II is possible, then it is also possible for any other set $\{\alpha_1, \beta_1, \omega_i'; \cdots; \alpha_g, \beta_g, \omega_g'\}$ such that

$$|\omega_i'| \geqq |\omega_i| \qquad\qquad i = 1, 2, \cdots, g\ .$$

1.2. Let $\mathfrak{M}_g$ be the subset of the $3g$-dimensional complex cartesian space composed of those points

$$m \sim \{\alpha_1, \beta_1, \omega_1; \cdots; \alpha_g, \beta_g, \omega_g\}$$

for which it is possible to choose g Jordan curves $\Lambda_1, \Lambda_2, \cdots, \Lambda_g$ of Σ such that

(a) each Λ_i separates α_i from β_i,

(b) the closed sets $\overline{\Lambda_1(\alpha_1)}, \overline{\tau_1\Lambda_1(\beta_1)}, \cdots, \overline{\Lambda_g(\alpha_g)}, \overline{\tau_g\Lambda_g(\beta_g)}$[3] are exterior to each other.

III. *The points of $\mathfrak{M}_g$ give rise to compact Riemann surfaces of genus g.*[4]

If $m \sim \{\alpha_1, \beta_1, \omega_1; \cdots; \alpha_g, \beta_g, \omega_g\}$ and $\Lambda_1, \Lambda_2, \cdots, \Lambda_g$ are chosen to satisfy (a) and (b), we set

$$R = \bigcap_{i=1,g} \{\overline{\Lambda_i(\beta_i)} \cap \overline{\tau_i\Lambda_i(\alpha_i)}\}\ .$$

We then identify the points of the boundaries Λ_i and $\tau_i\Lambda_i$ of R by means of the transformation τ_i. In other words we set $Q \sim \tau_i Q$ for each $Q \in \Lambda_i$. We do this for $i = 1, 2, \cdots, g$. Let X denote the resulting space.

We shall make X into a Riemann surface introducing local uniformizers.

[3] Here the τ_i's are again given by (1).

[4] The construction presented here is to some extent contained in a paper of Schottky published in Crelle's Journal (1887, cfr. [**8**]). See also Hurwitz-Courant [**5**], p. 462.

If P is a point of X which is interior to R and N is a neighborhood of P contained in R we take as a local uniformizer any coordinate in Σ which does not attain the value ∞ within N.

If P is a point of X which lies on one of the Λ's, say Λ_i, we have to proceed in a different way.

First we take a neighborhood N of P *in* Σ which is so small that it is contained in the set

$$R \cup \tau_i^{-1} R .$$

Then we define a corresponding neighborhood N^* of P *in* X by setting

$$N^* = \{\overline{\Lambda_i(\beta_i)} \cap N\} + \tau_i\{\overline{\Lambda_i(\alpha_i)} \cap N\} = R \cap (N + \tau_i N) .$$

If $z(p)$ is a coordinate in Σ which does not attain the value ∞ in N, we introduce as a local uniformizer in N^* the function on X which takes the value $z(p)$ for $p \in R \cap N$ and the value $z(\tau_i^{-1}p)$ for a point p of $R \cap \tau_i N$.

We proceed in a similar way for each of the curves Λ_i. The resulting manifold is a Riemann surface of genus g; it will be denoted by $\Gamma(m;\ \Lambda_1,\ \Lambda_2, \cdots, \Lambda_g)$ and referred to as a "Schottky model".

1.3. We shall give statement III a more precise meaning by showing that

IV. *Any two surfaces* $\Gamma(m;\ \Lambda_1,\ \Lambda_2, \cdots, \Lambda_g)$ *and* $\Gamma'(m;\ \Lambda'_1,\ \Lambda'_2, \cdots, \Lambda'_g)$ (*same* m), *are conformally equivalent.*

Let G be the group of Moebius transformations generated by the τ_i's. G constitutes what is usually called a "Schottky group".

We shall denote by $\hat{\Gamma}(m)$ the set obtained from Σ by deleting the limit points of G.

The following properties of G are well known (cfr. for instance [4] pages 37 to 66), and can be easily established:

(a) The group G is free.

(b) The sets $D = \bigcap_{i=1,g} \{\overline{\Lambda_i(\beta_i)} \cap \tau_i \Lambda_i(\alpha_i)\}$ and $D' = \bigcap_{i=1,g} \{\overline{\Lambda'_i(\beta_i)} \cap \tau_i \Lambda'_i(\alpha_i)\}$ are fundamental regions of G.

(c) The images of D (as well as those of D') decompose and cover completely the set $\hat{\Gamma}(m)$, i.e. $\hat{\Gamma}(m) = \sum_{\tau \in G} \tau D = \sum_{\tau \in G} \tau D'$[5].

These relations yield

$$D = \sum_{\tau \in G} D \cap \tau D' \tag{3}$$

$$D' = \sum_{\tau \in G} D' \cap \tau D\ ; \tag{4}$$

[5] We should emphasize that $\hat{\Gamma}(m)$ is a *disjoint* union of the images of D and D'.

since D and D' are bounded away from the limit points of G[6] both these sums, after a finite number of terms, terminate with a string of empty sets. The equality in (3) is also equivalent to

$$D = \sum_{\tau \in G} D \cap \tau^{-1} D' \tag{5}$$

and (4) can be written in the form

$$D' = \sum_{\tau \in G} \tau(D \cap \tau^{-1} D') . \tag{6}$$

We define a mapping[7] φ: $D \leftrightarrow D'$ by setting

$$\varphi p = \tau p \text{ for } p \in D \cap \tau^{-1} D' .$$

Since the unions on the right hand sides of (5) and (6) are disjoint φ is well defined. Clearly φ preserves the identification of points in Γ and Γ' and thus defines a topological mapping of Γ onto Γ', in addition it maps every sufficiently small neighborhood of Γ conformally onto neighborhood of Γ'.

From this the assertion follows.

1.4. The abstract Riemann surface represented by any one of the surfaces $\Gamma(m; A_1, A_2, \cdots, A_g)$ shall be denoted by $\Gamma(m)$; it shall be referred to as "the Schottky model corresponding to m."

Suppose now that there exists a Moebius transformation of Σ onto itself which sends the points $\alpha_1, \beta_1; \cdots; \alpha_g, \beta_g$ respectively onto the points $\alpha_1', \beta_1'; \cdots; \alpha_g', \beta_g'$ and assume that the parameters $\omega_1, \omega_2, \cdots, \omega_g$ have been chosen in such a way that both $m \sim (\alpha_1, \beta_1, \omega_1; \cdots; \alpha_g, \beta_g, \omega_g)$ and $m' \sim (\alpha_1', \beta_1', \omega_1; \cdots; \alpha_g', \beta_g', \omega_g)$ lie in $\mathfrak{M}_g$. Then the corresponding models $\Gamma(m)$ and $\Gamma(m')$ are conformally equivalent. Under these circumstances, it is natural to identify any two points m and m' of $\mathfrak{M}_g$ for which we have

$$\begin{aligned} &\omega_i = \omega_i' , \\ &\text{if } g \geqq 2 \ (\beta_i, \alpha_1, \alpha_2, \beta_1) = (\beta_i', \alpha_1', \alpha_2', \beta_1')^{8} \qquad i = 2, \cdots, g , \\ &\text{if } g \geqq 3 \ (\alpha_i, \alpha_1, \alpha_2, \beta_1) = (\alpha_i', \alpha_1', \alpha_2', \beta_1') \qquad i = 3, \cdots, g . \end{aligned} \tag{7}$$

If Γ is a Riemann surface of genus g, the Jordan curves $A_1, A_2, \cdots, A_g$ will be said to form a "canonical semi-basis" if they can be completed to a canonical basis for the cycles of Γ.

The Riemann surface Γ will be said "marked" if a canonical semi-

[6] The limit points of G are contained in the sets $\tau_i^{-1} A_j(\alpha_j)$, $\tau_i^{-1}\tau_j A_j(\beta_j)$ and $\tau_i A_j(\alpha_j)$ $(i, j = 1, 2, \cdots, g)$.

[7] Here and in the following a "mapping" shall mean a "one-to-one mapping".

[8] By the symbol (x, y, z, w) where x, y, z, w are given distinct complex numbers we mean the cross-ratio $(x-y)(z-w)/(x-w)(z-y)$.

basis has been chosen in Γ. The surface Γ marked by $\Lambda_1, \Lambda_2, \cdots, \Lambda_g$ shall be denoted by the symbol $\Gamma(\Lambda_1, \Lambda_2, \cdots, \Lambda_g)$.

We shall consider two marked surfaces $\Gamma(\Lambda_1, \Lambda_2, \cdots, \Lambda_g)$ and $\Gamma'(\Lambda'_1, \Lambda'_2, \cdots, \Lambda'_g)$ as the same object whenever $\Gamma \sim \Gamma'$ (conformally) and Λ_i is homotopic to Λ'_i (for $i = 1, 2, \cdots, g$). With these identifications the following theorem holds:

V. *The points of $\mathfrak{M}_g$ are in a one-to-one correspondence with the marked Riemann surfaces of genus g.*

Proof. Clearly, every Schottky model $\Gamma(m; \Lambda_1, \Lambda_2, \cdots, \Lambda_g)$ can be considered a marked surface by the choice of $\Lambda_1, \Lambda_2, \cdots, \Lambda_g$ as a canonical semi-basis.

But the converse is also true: namely, to each marked surface $\Gamma(\Lambda_1, \Lambda_2, \cdots, \Lambda_g)$ there corresponds a Schottky model, uniquely defined up to a Moebius transformation, and thereby a point of $\mathfrak{M}_g$. This correspondence is easily established after constructing the so-called "Schottky covering surface" of each marked surface. This concept is well known (see for instance [4], pp. 256–257), but for the sake of completeness, we shall sketch its definition.

Let $\Gamma(\Lambda_1, \Lambda_2, \cdots, \Lambda_g)$ be a given marked surface.

Let $M_1, M_2, \cdots, M_g$ be a completion of $\Lambda_1, \Lambda_2, \cdots, \Lambda_g$ to a canonical basis, and $\mathscr{M}$ denote the free group generated by the cycles M_1, $M_2, \cdots, M_g$.

We imagine the surface $\Gamma(\Lambda_1, \Lambda_2, \cdots, \Lambda_g)$ cut along the curves Λ_i to yield a planar region X bounded by the $2g$ Jordan curves $\Lambda_1, \Lambda_2, \cdots, \Lambda_g$; $\Lambda_1^{-1}, \Lambda_2^{-1}, \cdots, \Lambda_g^{-1}$[9] of Γ. We then reproduce an infinite number of exact replicas X_M of X, one for each $M \in \mathscr{M}$. The closed sets $\overline{X}_M$ are then glued together according to the following rules:

(i) If $M = M_i M^*$ (and the first factor of M^* is not M_i^{-1}) then the points of the curve Λ_i^{-1} of $\overline{X}_{M^*}$ are identified with the corresponding ones in the curve Λ_i of $\overline{X}_M$.

(ii) If $M = M_i^{-1} M^*$ (and the first factor of M^* is not M_i) then the points of the curve Λ_i of $\overline{X}_{M^*}$ are identified with the corresponding ones in the curve Λ_i^{-1} of $\overline{X}_M$.

With these identifications the set $\sum_{M \in \mathscr{M}} \overline{X}_M$ becomes a covering surface of Γ. We shall denote it $\hat{\Gamma}_\Lambda$ and call it the "Schottky covering surface" of $\Gamma(\Lambda_1, \Lambda_2, \cdots, \Lambda_g)$.

What then remains to be proved is a consequence of the following well known properties of the surface $\hat{\Gamma}_\Lambda$. (cfr. for instance [5] pp. 483–484 or [4] Chapter X).

[9] We tacitly assume, without restriction, that the curves Λ_i do not intersect each other.

(a) $\hat{\Gamma}_A$ is of planar character, it can be conformally mapped into the sphere Σ.

(b) The mapping μ_i of $\hat{\Gamma}_A \leftrightarrow \hat{\Gamma}_A$ which sends each region X_M of $\hat{\Gamma}_A$ onto the adjacent region $X_{M_i M}$ is a cover transformation of $\hat{\Gamma}_A$.

(c) The group of cover transformations of $\hat{\Gamma}_A$ is free and admits the mappings $\mu_1, \mu_2, \cdots, \mu_g$ as generators.

(d) If φ is any conformal mapping of $\hat{\Gamma}_A$ into Σ, the cover transformations of $\hat{\Gamma}_A$ induce in Σ, through the mapping φ, a set G of Moebius transformations which is a Schottky group. The generators of G are given by the Moebius transformations

$$\tau_1 = \varphi\mu_1\varphi^{-1},\ \tau_2 = \varphi\mu_2\varphi^{-1},\ \cdots,\ \tau_g = \varphi\mu_g\varphi^{-1}\ .$$

(e) The image φX_E of X_E (where by E we mean the identity in $\mathscr{M}$) constitutes a fundamental region for G; its boundary consists of the curves $\varphi A_1, \varphi A_2, \cdots, \varphi A_g$; $\varphi A_1^{-1}, \varphi A_2^{-1}, \cdots, \varphi A_g^{-1}$, and φA_i^{-1} is the image of φA_i under the transformation τ_i for each i.

Thereby φX_E and $\tau_1\tau_2, \cdots, \tau_g$ originate a Schottky model which is conformally equivalent to $\Gamma(A_1, A_2, \cdots, A_g)$.

(f) If φ' is any other conformal mapping of $\hat{\Gamma}_A$ into Σ, $\varphi'\varphi^{-1}$ induces a Moebius transformation of Σ; thus, if we set

$$\tau_i = \varphi\mu_i\varphi^{-1},\ \tau_i' = \varphi'\mu_i\varphi'^{-1}$$

and

$$\frac{\tau_i z - \alpha_i}{\tau_i z - \beta_i} = \omega_i \frac{z - \alpha_i}{z - \beta_i},\quad \frac{\tau_i' z - \alpha_i'}{\tau_i' z - \beta_i'} = \omega_i' \frac{z - \alpha_i'}{z - \beta_i'}$$

(under some coordinate system in Σ), the corresponding points

$$m \sim (\alpha_1, \beta_1, \omega_1; \cdots; \alpha_g, \beta_g, \omega_g)$$
$$m' \sim (\alpha_1', \beta_1', \omega_i'; \cdots; \alpha_g', B_g', \omega_g')$$

of $\mathfrak{M}_g$ are to be considered the same since the equalities in (7) will necessarily be satisfied.

1.5. After Statement V it is natural to adopt the following:

Definition. If $\Gamma(A_1, A_2, \cdots, A_g)$ is a given marked Riemann surface and $m \sim \{\alpha_1, \beta_1, \omega_1; \cdots; \alpha_g, \beta_g, \omega_g\}$ is the point of $\mathfrak{M}_g$ corresponding to it, the complex numbers

$$(8)\qquad \begin{aligned} &\omega_1, \omega_2, \cdots, \omega_g \\ &\omega_{i+g-1} = (\beta_i, \alpha_1, \alpha_2, \beta_1) \quad (i = 2, \cdots, g \text{ if } g \geqq 2) \\ &\omega_{i+2g-3} = (\alpha_i, \alpha_1, \alpha_2, \beta_1) \quad (i = 3, \cdots, g \text{ if } g \geqq 3) \end{aligned}$$

will be called "*the conformal parameters*" of $\Gamma(\Lambda_1, \Lambda_2, \cdots, \Lambda_g)$.

In the following we shall say that a marked Riemann surface $\Gamma(\Lambda_1, \Lambda_2, \cdots, \Lambda_g)$ has been "uniformized" if the mapping of $\hat{\Gamma}_\Lambda$ into Σ and the conformal parameters of $\Gamma(\Lambda_1, \Lambda_2, \cdots, \Lambda_g)$ have been characterized.

It is interesting to note that Schottky in [8] expressed the abelian differentials and their periods as analytic functions of the parameters $\alpha_1, \beta_1, \omega_1; \cdots; \alpha_g, \beta_g, \omega_g$; unfortunately, there are some restrictive hypotheses in his proofs, and the results, although explicit, assume formidable expressions.

2. Some special models of compact Riemann surfaces.

2.1. The three-dimensional Euclidean space shall be denoted by E_3. Any smooth (four times continuously differentiable), non self-intersecting surface of E_3, homeomorphic to a sphere, shall be called a p-sphere.

A p-sphere shall always be assumed to have been assigned a specific orientation.

Let Λ be a Jordan curve of a p-sphere Γ. If α is a point of Γ not lying in Λ, as before, we shall denote by $\Lambda(\alpha)$ the connected component of $\Gamma - \Lambda$ which contains α.

We can define an orientation of Λ by specifying which of the two connected components of $\Gamma - \Lambda$ is to be the interior or the exterior of Λ; conversely if Λ has been oriented, we can accordingly speak of the interior and the exterior of Λ in Γ. To this end we shall adopt the following convention:

If Q is a point of Λ, $\boldsymbol{t}$ and $\boldsymbol{b}$ are unit vectors having respectively the direction of the positive tangent to Λ and the positive normal to Γ at Q, and if the unit vector $\boldsymbol{n}$, normal to Λ and tangent to Γ at Q, points towards the interior of Λ, then the ordered triplet $\boldsymbol{t}, \boldsymbol{n}, \boldsymbol{b}$ should form a left handed frame.

Any oriented surface of E_3 can be made into a Riemann surface in a natural way by means of the conformal structure induced by the surrounding metric. In this fashion every p-sphere can be considered a compact Riemann surface of genus zero, and therefore it can be mapped conformally onto a sphere.

2.2. Let Σ be a sphere, and z a complex coordinate in Σ. If Γ is a p-sphere, let $z = \varphi p$ be a conformal mapping of Γ onto Σ. By means of φ we can transfer to Γ several conformally invariant properties of Σ. We shall define the cross-ratio of any four points $\alpha, \beta, \gamma, \delta$ of Γ by setting

$$(1) \qquad (\alpha, \beta, \gamma, \delta) = (\varphi\alpha, \varphi\beta, \varphi\gamma, \varphi\delta) .$$

The right hand side of (1) is independent of the mapping φ. In fact, if ψ is any other conformal mapping of Γ onto Σ, the mapping $\tau = \psi\varphi^{-1}$ of Σ onto itself is conformal and necessarily a Moebius transformation.

A Jordan curve Λ of Γ will be called a p-circle if the cross ratio of any four points of Λ is real; i.e., if the curve $\varphi\Lambda$ is a circle in Σ.

If Λ is a p-circle of Γ and α, β, γ are distinct points of Λ by an "inversion with respect to Λ" we shall mean the transformation σ defined by the equation

$$(\sigma p,\ \alpha,\ \beta,\ \gamma) = \overline{(p,\ \alpha,\ \beta,\ \gamma)} \tag{2}$$

the bar meaning complex conjugation. Clearly $\varphi\sigma\varphi^{-1}$ is in Σ an inversion with respect to the circle $\varphi\Lambda$.

The most general conformal mapping τ of Γ onto itself is determined by the images α', β', γ' of any three distinct points α, β, γ of Γ, and its equation can be written in the form

$$(\tau p,\ \alpha',\ \beta',\ \gamma') = (p,\ \alpha,\ \beta,\ \gamma)\ .$$

Such a mapping will be referred to as "a Moebius transformation of the p-sphere Γ".

We will find it convenient, in order to avoid having to refer back to the sphere Σ, to consider Schottky models imbedded in a p-sphere. Indeed, the construction of these models can be carried out for p-spheres in exactly the same way it was done in the last section for ordinary spheres; thus we shall not repeat it.

2.3. Let Γ_1 and Γ_2 be two p-spheres which intersect along a Jordan curve Λ. Suppose that there exists a conformal mapping φ of Γ_1 onto Γ_2 which leaves fixed the points of the intersection Λ.

The mapping φ is unique.

In fact, if ψ is another conformal mapping of Γ_1 onto Γ_2 which leaves the points of Λ fixed, then the mapping $\psi\varphi^{-1}$: $\Gamma_2 \leftrightarrow \Gamma_2$ leaves more than three points fixed and must necessarily be the identity.

This shows that φ is completely determined by the conditions imposed on it by three distinct points of the curve Λ, hence φ may not exist if the intersection of Γ_1 and Γ_2 is arbitrary.

A class of examples of couples of intersecting p-spheres for which such a mapping exists can be obtained by constructing surfaces which have a common axis of revolution and intersect along a common parallel, then taking their images under arbitrary Moebius transformations of space.

Suppose now that the finite ordered set of p-spheres $\Gamma_0, \Gamma_1, \cdots, \Gamma_{n-1}$ is such that for each $i = 1, 2, \cdots, n$:

(a) The surface Γ_{i-1} intersects the successive one Γ_i along a Jordan curve A_i which we shall suppose sufficiently well behaved. (We set $A_n = A_0$, $\Gamma_n = \Gamma_0$).

(b) There exists a conformal mapping Δ_i of Γ_{i-1} onto Γ_i which leaves fixed the points of the curve A_i.

(c) A_{i-1} has on points in common with A_i.

Let each A_i be oriented in such a way that the interior of A_i in Γ_{i-1} contains the curve A_{i-1}. Let A_i^- and A_i^+ denote respectively the interior of A_i in Γ_{i-1} and the exterior of A_i in Γ_i. With this notation we have

$$\Delta_i A_i^- + A_i + A_i^+ = \Gamma_i \ .$$

The ordered set of p-spheres $\Gamma_0, \Gamma_1, \cdots, \Gamma_{n-1}$ will be said to generate "a link of M-surface", if in addition to (a), (b), (c) it satisfies the following conditons:

(d) The exterior A_{i-1}^+ of A_{i-1} in Γ_{i-1} contains the curve A_i.

(e) No two of the sets $A_{i-1}+A_{i-1}^+ \cap A_i^-$ have any points in common.

These conditions being satisfied, the set

$$L = A_0 + A_0^+ \cap A_1^- + A_1 + A_1^+ \cap A_2^- + \cdots + A_{n-1} + A_{n-1}^+ \cap A_0^-$$

constitutes a compact, piece wise smooth, surface of genus one. We shall make L into a Riemann surface.

For each $i = 0, 1, \cdots, n-1$[10] let φ_i be a conformal mapping of Γ_i onto a given sphere Σ.

Let $\varphi_n = \varphi_0$, $\Delta_n = \Delta_0$, $\Gamma_{-1} = \Gamma_{n-1}$, $A_{-1} = A_{n-1}$, etc...

If p_0 is a point of $A_i^+ \cap A_{i+1}^-$ and N a neighborhood of p_0 in Γ_i, small enough to be contained in $A_i^+ \cap A_{i+1}^-$, we take as local uniformizer in N the function $z = \varphi_i p$, where z is any coordinate in Σ which does not assume the value ∞ in $\varphi_i N$.

If p_0 is a point of A_i, let N be a neighborhood of p_0 in Γ_i small enough to be contained in the domain $\{\Delta_i A_{i-1}^+\} \cap A_{i+1}^-$. We take as a neighborhood of p_0 in L the set

$$N^* = \{\Delta_i^{-1} N\} \cap A_i^- + N \cap A_i + N \cap A_i^+ \ .$$

We introduce as local uniformizer in N^* the function defined by setting

$$z = \varphi_i \Delta_i p \text{ for } p \in \{\Delta_i^{-1} N\} \cap A_i^-$$

and

$$z = \varphi_i p \quad \text{for } p \in N \cap \{A_i + A_i^+\}$$

Again, z is any coordinate in Σ which does not assume the value infinity in $\varphi_i N$.

[10] Here and in the following we shall assume a link to consist of at least 3 p-spheres.

The conformal structure thus introduced in L agrees in a natural way with that induced by the surrounding metric of E_3. Of course, in general along the curves Λ_i there will be discrepancies between angles measured in E_3 and angles measured in L.

The surface L will be referred to as a "link of M-surface" or briefly a "link". It will be denoted by $L(\Gamma_0, \Gamma_1, \cdots, \Gamma_{n-1})$.

2.4. We shall now construct surfaces of higher genus by putting together several links. There are several ways to achieve this. For our purposes it will be sufficient to construct only surfaces which consist of a p-sphere Γ_0 with many handles, each handle being part of a link containing Γ_0.

Let $L_1, L_2, \cdots, L_g$ be the links

$$\begin{gathered} L_1(\Gamma_{1,0}, \Gamma_{1,1}, \cdots, \Gamma_{1,n_1-1}) \\ L_2(\Gamma_{2,0}, \Gamma_{2,1}, \cdots, \Gamma_{2,n_2-1}) \\ \cdots \quad \cdots \quad \cdots \\ L_g(\Gamma_{g,0}, \Gamma_{g,2}, \cdots, \Gamma_{g,n_g-1}) \,. \end{gathered}$$

With the same notations as before we shall use the symbols $\Lambda_{i,j}$, $\Delta_{i,j}$, $\varphi_{i,j}$ where the first index will denote which link the object represented belongs to, and the second index, which position it occupies in the link itself.

Suppose that $L_1, L_2, \cdots, L_g$ satisfy the following conditions:

(f) The initial surfaces $\Gamma_{1,0}, \cdots, \Gamma_{g,0}$ are all the same p-sphere Γ_0.

(g) No two of the sets $L_i - \Gamma_0$ have any point in common.

(h) The closed sets $\Gamma_0 - \Lambda^+_{i,0}$, $\Gamma_0 - \Lambda^-_{j,1}$ $(i, j = 1, 2, \cdots, g)$ are all exterior to each other.

Then the set Ξ defined by

$$\Xi = L_1 \cap L_2 \cap \cdots \cap L_g + \sum_{1,g} {}_i (L_i - \Gamma_0) \,,$$

or, which is the same, by

$$\Xi = \sum_{1,g} {}_i (\Lambda_{i,0} + \Lambda_{i,1}) + \bigcap_{i=1,g} \{\Lambda^+_{i,0} \cap \Lambda^-_{i,1}\} + \sum_{1,g} {}_i (L_i - \Gamma_0)$$

shall be called an "M-surface".

Ξ can be made into a Riemann surface using the same local uniformizers which were introduced for the L_i's themselves.

However, some care has to be applied in the choice of permissible neighborhoods, and this is solely for points of the surface Γ_0.

We shall illustrate the situation with representative cases:

Suppose that P is a point of Ξ that is in Γ_0.

If $P \in \bigcap_{i=1,g} \{\Lambda^+_{i,0} \cap \Lambda^-_{i,1}\}$, then we can take as a neighborhood of P in

Ξ any neighborhood of P in Γ_0 which is small enough to be contained in $\bigcap_{i=1,g} \{\Lambda_{i,0}^+ \cap \Lambda_{i,1}^-\}$.

If $P \in \Lambda_{j,0}$, we choose first a neighborhood N of P in Γ_0 which is small enough to be contained in the domain

$$\Delta_{j,n_j}\{\Lambda_{j,n_j-1}^+ \cap \Lambda_{j,0}^-\} + \Lambda_{j,0} + \bigcap_{i=1,g} \{\Lambda_{i,0}^+ \cap \Lambda_{i,1}^-\} ,$$

then we take as a neighborhood of P in Ξ the set

$$N^* = \{\Delta_{j,n_j}^{-1} N\} \cap \Lambda_{j,0}^- + N \cap \Lambda_{j,0} + N \cap \Lambda_{j,0}^+ .$$

If $P \in \Lambda_{j,1}$, we choose a neighborhood N of P in $\Gamma_{j,1}$ so small that

$$N \subset \Delta_{j,1}\{\bigcap_{i=1,g} (\Lambda_{i,0}^+ \cap \Lambda_{i,1}^-)\} + \Lambda_{j,1} + \Lambda_{j,1}^+ \cap \Lambda_{j,2}^- .$$

We then take as a neighorhood of P in Ξ the set

$$N^* = \{\Delta_{j,1}^{-1} N\} \cap \Lambda_{j,1}^- + N \cap \Lambda_{j,1} + N \cap \Lambda_{j,1}^+$$

3. Characterization of the conformal parameters.

3.1. Let $\Xi \sim (L_1, L_2, \cdots, L_g)$ be a given M-surface, and $\Xi(\Lambda_{1,1}, \Lambda_{2,1}, \cdots, \Lambda_{g,1})$ denote the surface Ξ marked by the set of curves

$$\Lambda_{1,1}, \Lambda_{2,1}, \cdots, \Lambda_{g,1} .$$

We shall now present a construction of the Schottky model corresponding to $\Xi(\Lambda_{1,1}, \Lambda_{2,1}, \cdots, \Lambda_{g,1})$.

Let us first take under consideration the case that Ξ consists of a single link $L(\Gamma_0, \Gamma_1, \cdots, \Gamma_{n-1})$.

We imagine to have cut L along the curve Λ_1

Using the mapping Δ_2 we can collapse the portion $\Lambda_1 + \Lambda_1^+ \cap \Lambda_2^-$ of L into the p-sphere Γ_2. The new set

$$X_1 = \Delta_2\{\Lambda_1 + \Lambda_1^+ \cap \Lambda_2^-\} + \Lambda_2 + \Lambda_2^+ \cap \Lambda_3^- + \cdots + \Lambda_{n-1} + \Lambda_{n-1}^+ \cap \Lambda_0 + \Lambda_0 + \Lambda_0^+ \cap \Lambda_1^-$$

with the points of its boundaries Λ_1 and $\Delta_2\Lambda_1$ identified by the transformation Δ_2, can also be considered a Riemann surface.

We shall briefly describe the neighborhoods and the local uniformizers at the points of the set $\Delta_2\{\Lambda_1 + \Lambda_1^+ \cap \Lambda_2^-\} + \Lambda_2$.

If $p \in \Delta_2\Lambda_1$, we choose $N \ni p$ in Γ_2 so that

$$\Delta_2^{-1} N \subset \Delta_1(\Lambda_0^+ \cap \Lambda_1^-) + \Lambda_1 + \Lambda_1^+ \cap \Lambda_2^- ,$$

we then take

$$N^* = \{\Delta_1^{-1}\Delta_2^{-1} N\} \cap \Lambda_1^- + N \cap \Delta_2\{\Lambda_1 + \Lambda_1^+\} .$$

As a uniformizer in N^* we take the function

$$z = \varphi_2 p \qquad \text{for } p \in N \cap \Delta_2\{A_1 + A_1^+\}$$
$$z = \varphi_2 \Delta_2 \Delta_1 p \text{ for } p \in \{\Delta_1^{-1}\Delta_2^{-1}N\} \cap A_1^-$$

(provided that $z \neq \infty$ in N).

If $p \in \Delta_2\{A_1^+ \cap A_2^-\}$ we choose $N \ni p$ so that

$$N \subset \Delta_2\{A_1^+ \cap A_2^-\}$$

then set $N^* = N$ and $z = \varphi_2 p$ (assuming $z \neq \infty$ in N).

If $p \in A_2$ we choose $N \ni p$ so that

$$N \subset \Delta_2\{A_1^+ \cap A_2^-\} + A_2 + A_2^+ \cap A_3^- \ ,$$

then set $N^* = N$ and $z = \varphi_2 p$ (assuming $z \neq \infty$ in N).

L and X_1 are conformally equivalent.

In fact, the function ψ_1 defined by

$$\psi_1 p = p \quad \text{ for } p \in A_2 + A_2^+ \cap A_3^- + \cdots + A_n + A_n^+ \cap A_1^-$$
$$\psi_1 p = \Delta_2 p \text{ for } p \in A_1 + A_1^+ \cap A_2^-$$

induces a conformal mapping of L onto X_1.

We proceed in a similar way, and collapse the subset

$$\Delta_2\{A_1 + A_1^+ \cap A_2^-\} + A_2 + A_2^+ \cap A_3^-$$

of Γ_2 into Γ_3 by means of the mapping Δ_3, the subset

$$\Delta_3\Delta_2\{A_1 + A_1^+ \cap A_2^-\} + \Delta_3\{A_2 + A_2^+ \cap A_3^-\} + A_3 + A_3^+ \cap A_4^-$$

of Γ_3 into Γ_4 by means of the mapping Δ_4, etc..., the subset

$$\Delta_{k-1} \cdots \Delta_2\{A_1 + A_1^+ \cap A_2^-\} + \cdots + \Delta_{k-1}\{A_{k-2} + A_{k-2}^+ \cap A_{k-1}^-\} + A_{k-1} + A_{k-1}^+ \cap A_k^-$$

of Γ_{k-1} into Γ_k by means of the mapping Δ_k, and set

$$X_{k-1} = \Delta_k \cdots \Delta_2\{A_1 + A_1^+ \cap A_2^-\} + \cdots + \Delta_k\{A_{k-1} + A_{k-1}^+ \cap A_k^-\}$$
$$+ A_k + A_k^+ \cap A_{k+1}^- + \cdots + A_n + A_n^+ \cap A_1^- + A_1 \ .$$

Again, X_{k-1} is made into a Riemann surface, by introducing local uniformizers in such a way that the function ψ_{k-1} defined by

$$\psi_{k-1} p = p \quad \text{ for } p \in A_k + A_k^+ \cap A_{k+1}^- + \cdots + A_n + A_n^+ \cap A_1^- \ ,$$
$$\psi_{k-1} p = \Delta_k p \text{ for } p \in A_{k-1} + A_{k-1}^+ \cap A_k^- \ ,$$
$$\cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots \quad \cdots$$
$$\psi_{k-1} p = \Delta_k \Delta_{k-1} \cdots \Delta_2 p \text{ for } p \in A_1 + A_1^+ \cap A_2^-$$

induces a conformal mapping between L and X_{k-1}.

In this fashion, at each step of the process L and X_{k-1} are kept conformally equivalent, in particular for $k = n$ we obtain that L is conformally equivalent to the subset

$$X_{n-1} = \Delta_n\Delta_{n-1}\cdots\Delta_2\{\Lambda_1 + \Lambda_1^+ \cap \Lambda_2^-\} + \cdots$$
$$+ \Delta_n\{\Lambda_{n-1} + \Lambda_{n-1}^+ \cap \Lambda_n^-\} + \Lambda_n + \Lambda_n^+ \cap \Lambda_1^- + \Lambda,$$

of the p-sphere Γ_0. Of course the points of the boundaries Λ_1 and $\Delta_n\Delta_{n-1}\cdots\Delta_2\Lambda_1$, of X_{n-1} are to be considered identified by the mapping $\Delta_n\Delta_{n-1}\cdots\Delta_2$ or, which is the same[11], by the transformation $\tau = \Delta_n\Delta_{n-1}\cdots\Delta_2\Delta_1$.

3.2. We shall now prove that

I. *X_{n-1} is a Schottky model in Γ_0.*

Since τ is necessarily a Moebius transformation of Γ_0, all we have to show, to justify our assertion, is that τ is hyperbolic or loxodromic, that it has two fixed points $\alpha \in \Gamma_0 - \Lambda_1^-$ and $\beta \in \tau\Lambda_1^-$, and that

$$\tau\Lambda_1(\alpha) \supset \overline{\Lambda_1(\alpha)}\ .$$

Now for each k we have

$$\Delta_k\{\Gamma_{k-1} - \Lambda_k^-\} = \Lambda_k + \Lambda_k^+$$

and since

$$\Lambda_{k-1}^+ \supset \Gamma_{k-1} - \Lambda_k^-\ ,$$

we have

$$(1) \qquad \Delta_k\Lambda_{k-1}^+ \supset \Lambda_k^+\ .$$

Thus if

$$\Delta_{k-1}\cdots\Delta_1\{\Gamma_0 - \Lambda_1^-\} \supset \Lambda_{k-1}^+\ ,$$

because of (1) it will follow that

$$(2) \qquad \Delta_k\cdots\Delta_1\{\Gamma_0 - \Lambda_1^-\} \supset \Lambda_k^+\ .$$

However, we have $\Delta_1\{\Gamma_0 - \Lambda_1^-\} = \Lambda_1 + \Lambda_1^+ \supset \Lambda_1^+$; hence (2) is true and for $k = n$ we have

$$(3) \qquad \tau\{\Gamma_0 - \Lambda_1^-\} \supset \Lambda_0^+ \supset \Gamma_0 - \Lambda_1^-\ .$$

Since $\Gamma_0 - \Lambda_1^-$ is closed and Λ_0^+ is open, the boundaries Λ_1 and $\tau\Lambda_1$ of $\Gamma_0 - \Lambda_1^-$ and $\tau\{\Gamma_0 - \Lambda_1^-\}$ cannot have any point in common. Therefore, if α^* and β^* are two points of Γ_0 such that $\alpha^* \in \Gamma_0 - \bar{\Lambda}_1^-$ and $\beta^* \in \Lambda_1^-$, otherwise arbitrary, from (3) follows:

[11] $\Delta_n\Delta_{n-1}\cdots\Delta_2$ and $\Delta_n\Delta_{n-1}\cdots\Delta_2\Delta_1$ agree along Λ_1.

$$\tau^{-1}\overline{\Lambda_1(\alpha^*)} \subset \Lambda_1(\alpha^*)$$

and

$$\tau\overline{\Lambda_1(\beta^*)} \subset \Lambda_1(\beta^*) .$$

From these inclusions we can deduce that τ is neither parabolic nor elliptic:

In fact, if τ were parabolic with γ as a fixed point, then

$$\gamma = \lim_{n\to\infty} \tau^{-n}\alpha^* = \lim_{n\to\infty} \tau^n\beta^* .$$

But this would imply that

$$\gamma \in \Lambda_1(\alpha^*) \cap \Lambda_1(\beta^*)$$

which is absurd.

If τ were elliptic and $p \in \overline{\Lambda_1(\alpha^*)}$, then $\tau^{-1}p \in \Lambda_1(\alpha^*)$ and thus $\tau^{-1}p$ would be contained in an open set $D \subset \Lambda_1(\alpha^*)$; consequently $\tau^{-n}D\subset\Lambda_1(\alpha^*)$ for all $n \geqq 1$; but for a suitable value of n $\tau^{-n}D$ would cover p. This would imply that every point of $\overline{\Lambda_1(\alpha^*)}$ is interior to $\Lambda_1(\alpha^*)$ which is absurd.

Thus τ is hyperbolic or loxodromic and its fixed points are determined by the limits

$$\alpha = \lim_{n\to\infty} \tau^{-n}\{\Gamma_0 - \Lambda_1^-\}$$

$$\beta = \lim_{n\to\infty} \tau^n\Lambda_1^- .$$

With this notation under any coordinate system in Γ_0 the equation of τ takes the form

$$\frac{\tau z - \alpha}{\tau z - \beta} = \omega\frac{z-\alpha}{z-\beta}$$

with $|\omega| > 1$. Finally, since $\alpha \in \Gamma_0 - \overline{\Lambda_1^-}$, from (3) we obtain

$$\tau\Lambda_1(\alpha) \supset \overline{\Lambda_1(\alpha)} .$$

3.3. We shall now consider the general case.

Let $\Xi \sim (L_1, L_2, \cdots, L_g)$, imagine $\Xi(\Lambda_{1,1}, \Lambda_{2,1}, \cdots, \Lambda_{g,1})$ cut along the curves

$$\Lambda_{1,1}, \Lambda_{2,1}, \cdots, \Lambda_{g,1} .$$

We then apply to each link L_i the previous construction. Each handle

$$\Lambda_{i,1} + \Lambda_{i,1}^+ \cap \Lambda_{i,2}^- + \cdots + \Lambda_{i,n_i-1} + \Lambda_{i,n_i-1}^+ \cap \Lambda_{i,0}^- \quad (i = 1, 2, \cdots, g)$$

of Ξ, is flattened into Γ_0 by means of the mapping ψ_i defined by the equalities:

$$
(4)\quad
\begin{array}{lll}
\psi_i p = \Delta_{i,n_i} \cdots \Delta_{i,3}\Delta_{i,2} p & \text{for } p \in \Lambda_{i,1} + \Lambda^+_{i,1} \cap \Lambda^-_{i,2} \\
\psi_i p = \Delta_{i,n_i} \cdots \Delta_{i,3} p & \text{for } p \in \Lambda_{i,2} + \Lambda^+_{i,2} \cap \Lambda^-_{i,3} \\
\cdots \quad \cdots & \cdots \quad \cdots \quad \cdots \\
\psi_i p = \Delta_{i,n_i} p & \text{for } p \in \Lambda_{i,n_i-1} + \Lambda^+_{i,n_i-1} \cap \Lambda^-_{i,0} \,.
\end{array}
$$

The resulting subregion X of the p-sphere Γ_0 can be considered to be the intersection

$$X = X_{n_1-1} \cap X_{n_2-1} \cap \cdots \cap X_{n_g-1}$$

of the Schottky models X_{n_i-1} corresponding to each link of Ξ.

The pairs of boundaries $\Lambda_{i,1}$ and $\Delta_{i,n_i} \cdots \Delta_{i,3}\Delta_{i,2}\Lambda_{i,1}$ of X should be considered identified by the mapping $\Delta_{i,n_i} \cdots \Delta_{i,3}\Delta_{i,2}$ or, which is the same thing, by the mapping $\tau_i = \Delta_{i,n_i} \cdots \Delta_{i,2}\Delta_{i,1}$. Furthermore:

II. *X is a Schottky model conformally equivalent to Ξ.*

Proof. As a by-product of the proof of Statement I we obtain that

(a) Each mapping τ_i ($i = 1, 2, \cdots, g$) is a hyperbolic or loxodromic Moebius transformation of Γ_0.

(b) The fixed points α_i, β_i of τ_i are respectively contained in $\Gamma_0 - \overline{\Lambda^-_{i,1}}$ and $\Lambda^-_{i,1}$.

(c) In any coordinate system in Γ_0 the equation of τ_i writes

$$(5)\qquad \frac{\tau_i z - \alpha_i}{\tau_i z - \beta_i} = \omega_i \frac{z - \alpha_i}{z - \beta_i}$$

with $|\omega_i| > 1$.

(d) Each τ_i satisfies the inclusions (see (3))

$$\tau_i\{\Gamma_0 - \Lambda^-_{i,1}\} \supset \Lambda^+_{i,0} \supset \Gamma_0 - \Lambda^-_{i,1}$$

or, changing notation:

$$(6)\qquad \tau_i\overline{\Lambda_{i,1}(\alpha_i)} \supset \Lambda_{i,0}(\alpha_i) \supset \overline{\Lambda_{i,1}(\alpha_i)} \,.$$

Since $\Lambda_{i,0}(\alpha_i)$ is open we can safely conclude that (6) implies

$$\tau_i \Lambda_{i,1}(\alpha_i) \supset \overline{\Lambda_{i,1}(\alpha_i)} \,.$$

Condition (c) in the definition of a M-surface requires the closed sets

$$\Gamma_0 - \Lambda^+_{i,0} = \overline{\Lambda_{i,0}(\beta_i)},\ \Gamma_0 - \Lambda^-_{j\,1} = \overline{\Lambda_{j,1}(\alpha_j)} \quad (i, j = 1, 2, \cdots, g)$$

to be disjoint. However, the inclusions

$$\tau_i\overline{A_{i,1}(\alpha_i)} \supset A_{i,0}(\alpha_i)$$

imply

$$\tau_i A_{i,1}(\beta_i) \subset \overline{A_{i,0}(\beta_i)}$$

hence we must have

$$\tau_{i,}\overline{A_{i,1}(\beta_i)} \subset \overline{A_{i,0}(\beta_i)}\ .$$

Therefore also the closed sets

$$\tau_i\overline{A_{i,1}(\beta_i)},\ \overline{A_{j,1}(\alpha_j)} \qquad (i, j = 1, 2, \cdots, g)$$

are disjoint. With this, the conditions for X to be a Schottky model are all satisfied.

The conformal equivalence of X to Ξ is a consequence of the fact that the function ψ defined by the equalities

$$\psi p = p \quad \text{for } p \in L_1 \cap L_2 \cap \cdots \cap L_g - \sum_{i=1,g} A_{i,1}$$

and (see (4))

$$\psi p = \psi_i p \quad \text{for } p \in L_i - \Gamma_0 + A_{i,1} \qquad (i = 1, 2, \cdots, g)$$

induces a conformal mapping of Ξ onto X.

3.4. The mapping ψ, or rather its analytic continuation in Ξ, uniformizes the marked surface $\Xi(A_{1,1}, A_{2,1}, \cdots, A_{g,1})$.

Let $\hat{\Xi}_A$ represent the Schottky covering surface of $\Xi(A_{1,1}, A_{2,1}, \cdots, A_{g,1})$ and X_E the region obtained by cutting Ξ along the curves $A_{1,1}, A_{2,1}, \cdots, A_{g,1}$.

Let the cycles $M_1, M_2, \cdots, M_g$ of a completion of $A_{1,1}, A_{2,1}, \cdots, A_{g,1}$ to a canonical basis of Ξ be chosen in such a way that each M_i intersects the curves $A_{i,j}$ $(j = 1, 2, \cdots, n_i)$ in the order

$$A_{i,n_i},\ A_{i,n_i-1},\ \cdots,\ A_{i,2},\ A_{i,1}\ .$$

As before, let $\mathscr{M}$ be the free group generated by the M_i's and X_M for each $M \in \mathscr{M}$ an exact replica of X_E.

Then we have

$$\hat{\Xi}_A = \sum_{M \in \mathscr{M}} \overline{X}_M$$

where again the boundaries of the $\overline{X}_M$'s are identified according to the rules (i), (ii) stated in § 1.4.

For each $M \in \mathscr{M}$ let $\tau_M \in G$[12] be the Moebius transformation corresponding to M under the isomorphism of $\mathscr{M}$ onto G defined by setting

[12] As before G denotes the group generated by the τ_i's.

$$M_i \longleftrightarrow \tau_i \qquad (i = 1, 2, \cdots, g) .$$

The mapping $\hat{\psi}$ of $\hat{\Xi}_\Lambda$ into Γ_0 is then obtained taking

$$\hat{\psi} p = \tau_M \psi p \quad \text{for } p \in \overline{X}_M - \sum_{i=1,g} \Lambda_{i,1}^{-1} ,$$

and the region of Γ_0 onto which $\hat{\Xi}_\Lambda$ is mapped is given by the union

$$\hat{\psi}\hat{\Xi}_\Lambda = \sum_{\tau \in G} \tau\psi\overline{X}$$

This shows that X is the Schottky model corresponding to $\Xi(\Lambda_{1,1}, \Lambda_{2,1}, \cdots, \Lambda_{g,1})$ and therefore that the conformal parameters of $\Xi(\Lambda_{1,1}, \Lambda_{2,1}, \cdots, \Lambda_{g,1})$ are characterized by the invariants ω_i and the fixed points α_i, β_i of the transformations τ_i.

4. Links of spheres.

4.1. Given two oriented spheres Γ_1 and Γ_2 intersecting along a circle Λ, there always exists a conformal mapping Δ of Γ_1 onto Γ_2 which leaves unchanged the points of Λ.

The mapping Δ can be constructed in the following way:

Let τ be a Moebius transformation of E_3 which sends a point of Λ onto the point at infinity. The circle Λ is taken by τ onto a straight line $\tau\Lambda$ and the spheres Γ_1 and Γ_2 onto two planes $\tau\Gamma_1$, $\tau\Gamma_2$ intersecting along $\tau\Lambda$. If π_1 and π_2 denote the two planes through $\tau\Lambda$ which bisect the dihedral angle formed by $\tau\Gamma_1$ and $\tau\Gamma_2$, the two transformations τ_{π_1} and τ_{π_2} obtained by reflection across π_1 and π_2 respectively, map $\tau\Gamma_1$ onto $\tau\Gamma_2$ with preservation of angles and leave unchanged the points of $\tau\Lambda$.

The corresponding spheres $\tau^{-1}\pi_1$ and $\tau^{-1}\pi_2$ generate the inversions $\tau_1 = \tau^{-1}\tau_{\pi_1}\tau$, $\tau_2 = \tau^{-1}\tau_{\pi_2}\tau$ which map Γ_1 onto Γ_2 with preservation of angles and leave unchanged the points of Λ. These two spheres are called the spheres of antisimilitude of Γ_1 and Γ_2 (see also [3] page 230).

To see which of τ_1 and τ_2 defines the conformal mapping Δ, suppose that we transfer the orientation of Γ_1 and Γ_2 onto $\tau\Gamma_1$ and $\tau\Gamma_2$ by means of τ. The product $R = \tau_{\pi_1}\tau_{\pi_2}$ is a rotation of π radians around $\tau\Lambda$, therefore whatever may be the orientations of $\tau\Gamma_1$ and $\tau\Gamma_2$, R generates a sense reversing transformation of $\tau\Gamma_1$ and $\tau\Gamma_2$ onto themselves. The same will also be true for the product

$$R' = \tau^{-1}R\tau$$

with respect to Γ_1 and Γ_2. Since $\tau_1 = \{\tau^{-1}\tau_{\pi_1}\tau_{\pi_2}\tau\}\{\tau^{-1}\tau_{\pi_2}\tau\} = R'\tau_2$, either τ_1 or τ_2 is orientation preserving (as a transformation of Γ_1 onto Γ_2). But each of them is a sense reversing transformation of E_3, therefore the transformation Δ is given by that one of τ_1 and τ_2 which sends the interior of Γ_1 onto the exterior of Γ_2. The one of $\tau^{-1}\pi_1$ and $\tau^{-1}\pi_2$

which generates Δ will be called the "direct" sphere of antisimilitude of Γ_1 and Γ_2.

We can thus construct M-surfaces by means of collections of intersecting oriented spheres. Such M-surfaces will be called "natural".

Natural M-surfaces form a wide family for which the canal surfaces[13] are limit elements. It seems reasonable to conjecture that every Riemann surface can be realized as a natural M-surface. We shall later show that every natural M-surface can be deformed into a C^∞ canal surface without altering its conformal structure. For these reasons we found it of some interest to present a brief study of the conformal parameters of natural M-surfaces. This will lead to a few results concerning the conformal imbedding of Riemann surfaces of genus one.

Before presenting these results we need to introduce a few tools.

4.2. The conformal geometry of the 3 dimensional space is simplified by the use of "anallagmatic coordinates". An introduction to these coordinates can be found in a paper by E. Cartan [2] or in a book by R. Lagrange [6]. Here we will give only a brief description of them.

The collection of all planes, properly or improperly real spheres, and points of E_3 shall be called the "3 dimensional anallagmatic space"; we shall denote it by $\mathscr{A}_3$.

A one-to-one correspondence between the points of a 4-dimensional real projective space $\mathscr{P}_4 \sim (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)$ and the elements of $\mathscr{A}_3$ can be generated in the following way:

To each point $\boldsymbol{\alpha} \sim (\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4)$ of $\mathscr{P}_4$, if x_1, x_2, x_3 denote the cartesian coordinates of a point of E_3, we can associate the equation

$$(1) \qquad \alpha_0(x_1^2 + x_2^2 + x_3^2) - 2\alpha_1 x_1 - 2\alpha_2 x_2 - 2\alpha_3 x_3 + \alpha_4 = 0 .$$

If $\alpha_0 = 0$ this equation defines a plane of E_3.

If $\alpha_0 \neq 0$ (1) is equivalent to the equation

$$(2) \quad \left(x_1 - \frac{\alpha_1}{\alpha_0}\right)^2 + \left(x_2 - \frac{\alpha_2}{\alpha_0}\right)^2 + \left(x_3 - \frac{\alpha_3}{\alpha_0}\right)^2 = \frac{\alpha_1^2 + \alpha_2^2 + \alpha_3^2 - \alpha_0\alpha_4}{\alpha_0^2} ,$$

which defines a real sphere, a point or an improperly real sphere according as the quadratic form

$$(3) \qquad (\boldsymbol{\alpha}, \boldsymbol{\alpha}) = \alpha_1^2 + \alpha_2^2 + \alpha_3^2 - \alpha_0\alpha_4$$

is greater, equal or less than zero.

This correspondence between $\mathscr{P}_4$ and $\mathscr{A}_3$ is clearly invertible. The five real numbers $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4$ (determined up to a common factor of proportionality) thus associated to each element of $\mathscr{A}_3$, are called the "anallagmatic coordinates" of that element. When expressed in anal-

[13] Surfaces which are envelopes of spheres (see [**1**]).

lagmatic coordinates, the Moebius transformations of E_3 become the homographies of $\mathscr{P}_4$ which leave invariant the binary form

$$(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \alpha_1\beta_1 + \alpha_2\beta_2 + \alpha_3\beta_3 - \tfrac{1}{2}(\alpha_0\beta_4 + \alpha_4\beta_0) . \tag{4}$$

This form is assumed as a scalar product in $\mathscr{P}_4$. We have to distinguish it from the Euclidean scalar product

$$\boldsymbol{x} \cdot \boldsymbol{y} = x_1y_1 + x_2y_2 + x_3y_3 , \tag{5}$$

which will also figure in our subsequent formulas. To this end vectors with 5 components will be denoted by means of Greek characters and vectors with 3 components by means of Latin characters. We shall always denote (4) by $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and (5) by $\boldsymbol{x} \cdot \boldsymbol{y}$, $\boldsymbol{x} \cdot \boldsymbol{x}$ often by $\boldsymbol{x}^2$, a point $\boldsymbol{\alpha}$ of $\mathscr{P}_4$ briefly

$$\boldsymbol{\alpha} \sim (\alpha_0, \boldsymbol{a}, \alpha_4) ,$$

and the binary form (4)

$$(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \boldsymbol{a} \cdot \boldsymbol{b} - \tfrac{1}{2}(\alpha_0\beta_4 + \alpha_4\beta_0) . \tag{6}$$

To represent oriented spheres of E_3 it is convenient to normalize the anallagmatic coordinates by making use of the factor of proportionality so as to express orientations in an invariant way (see [2]). This is achieved by requiring that:

(1) If $\boldsymbol{\alpha} \sim (\alpha_0, \boldsymbol{a}, \alpha_4)$ corresponds to a point of E_3 we should have

$$\alpha_0 + \alpha_4 > 0$$

(2) If $\boldsymbol{\alpha}$ corresponds to a real oriented sphere Γ of E_3 and $\xi \sim (x_0, \boldsymbol{x}, x_4)$ corresponds to an interior point of Γ we should have

$$(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 1$$
$$(\boldsymbol{\alpha}, \boldsymbol{\xi}) > 0 .$$

(3) If $\boldsymbol{\alpha}$ corresponds to an oriented plane π and $\boldsymbol{\xi}$ to a point of the half-space towards which the positive normal of π is directed, we should have

$$(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = 1$$
$$(\boldsymbol{\alpha}, \boldsymbol{\xi}) > 0 .$$

(4) If $\boldsymbol{\alpha}$ corresponds to an improperly real sphere, we should have

$$(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = -1$$
$$\alpha_0 + \alpha_4 > 0 .$$

The transition from Euclidean to normalized anallagmatic coordinates can be carried out according to the following rules:

(a) If $\boldsymbol{p} \sim \boldsymbol{\xi}$ is a point of E_3 and $\lambda > 0$ then

$$\boldsymbol{\xi} = \lambda(1,\ \boldsymbol{p},\ \boldsymbol{p}^2)\ .$$

(b) If $\Gamma \sim \boldsymbol{\alpha}$ is a sphere or radius R and center in $\boldsymbol{c}$, oriented so that $\boldsymbol{c}$ is an interior point

$$\boldsymbol{\alpha} = \frac{1}{R}(1,\ \boldsymbol{c},\ \boldsymbol{c}^2 - R^2)\ .$$

(c) If $\Gamma \sim \boldsymbol{\alpha}$ has the same center but imaginary radius

$$\boldsymbol{\alpha} = \frac{1}{R}(1,\ \boldsymbol{c},\ \boldsymbol{c}^2 + R^2)\ .$$

(d) If $\pi \sim \boldsymbol{\alpha}$ is a plane which contains the point Q and has the unit vector $\boldsymbol{n}$ as positive normal

$$\boldsymbol{\alpha} = (0,\ \boldsymbol{n},\ 2\boldsymbol{n} \cdot \boldsymbol{Q})\ .$$

By means of these formulas it can be easily verified that:

(i) The cosine of the Euclidean angle formed by two oriented spheres $\Gamma_1 \sim \boldsymbol{\alpha}$ and $\Gamma_2 \sim \boldsymbol{\beta}$ is given by the binary form (6).

(ii) A point $p \sim \boldsymbol{\xi}$ belongs to a sphere $\Gamma \sim \boldsymbol{\alpha}$ if and only if $(\boldsymbol{\alpha},\ \boldsymbol{\xi}) = 0$.

(iii) The equation of the inversion Δ generated by a real sphere $\Gamma \sim \boldsymbol{\delta}$ when expressed in normalized anallagmatic coordinates takes the form[14]

$$\Delta\boldsymbol{\xi} = \boldsymbol{\xi} - 2(\boldsymbol{\xi},\ \boldsymbol{\delta})\boldsymbol{\delta}\ , \tag{7}$$

where $\boldsymbol{\xi}$ denotes a variable element of $\mathscr{P}_4$.

The normalization (1) for anallagmatic coordinates of points of E_3 is invariant under products of inversions generated by real spheres. In fact, from (7) follows that if $\boldsymbol{\delta} = 1/R(1,\ \boldsymbol{c},\ \boldsymbol{c}^2 - R^2)$ and $\boldsymbol{\xi} = \lambda(1,\ \boldsymbol{p},\ \boldsymbol{p}^2)$ then

$$\Delta\boldsymbol{\xi} = \lambda\frac{|\boldsymbol{p} - \boldsymbol{c}|^2}{R^2}(1,\ \boldsymbol{p}',\ \boldsymbol{p}'^2) \tag{7*}$$

with

$$\boldsymbol{p}' = \boldsymbol{c} + \frac{R^2}{|\boldsymbol{p} - \boldsymbol{c}|^2}(\boldsymbol{p} - \boldsymbol{c})\ .$$

Thus $\Delta\boldsymbol{\xi}$ satisfies condition (1) whenever $\boldsymbol{\xi}$ does.

---

[14] See also **[6]** pages 25–26.

Using (7) we can readily obtain the anallagmatic coordinates of the direct sphere of antisimilitude of two given intersecting oriented spheres $\Gamma_1 \sim \boldsymbol{\alpha}_1$ and $\Gamma_2 \sim \boldsymbol{\alpha}_2$. According to the considerations in § 4.1, the sphere $\Gamma \sim \boldsymbol{\delta}$ is the direct sphere of antisimilitude of Γ_1 and Γ_2 if and only if the inversion Δ which it generates, transforms the oriented sphere Γ_1 onto the sphere Γ_2 oriented in the opposite way; thus in anallagmatic coordinates we should have

$$\Delta\boldsymbol{\alpha}_1 = -\boldsymbol{\alpha}_2 ,$$

and by (7)

$$\boldsymbol{\alpha}_1 - 2(\boldsymbol{\alpha}_1, \boldsymbol{\delta})\boldsymbol{\delta} = -\boldsymbol{\alpha}_2 , \tag{8}$$

or

$$(\boldsymbol{\alpha}_1, \boldsymbol{\delta})\boldsymbol{\delta} = \frac{\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2}{2} . \tag{9}$$

To find $(\boldsymbol{\alpha}_1, \boldsymbol{\delta})$ we multiply both sides of (9) scalarly by $\boldsymbol{\alpha}_1$ obtaining

$$(\boldsymbol{\alpha}_1, \boldsymbol{\delta}) = \pm \sqrt{\frac{1 + \cos\varphi}{2}} \tag{10}$$

where by φ we indicate the Euclidean angle formed by Γ_1 and Γ_2. Now, $(\boldsymbol{\alpha}_1, \boldsymbol{\delta})$ does not vanish, for otherwise (8) yields $\boldsymbol{\alpha}_1 = -\boldsymbol{\alpha}_2$; and since the orientation of $\boldsymbol{\delta}$ does not affect the outcome of (7) we can choose the positive sign in (10) so that we obtain

$$\boldsymbol{\delta} = \frac{\boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2}{2\cos\varphi/2} . \tag{11}$$

4.3. The conformal parameters of natural M-surfaces admit a purely algebraic characterization in terms of the anallagmatic coordinates of the generating spheres.

Suppose first that $L \sim (\Gamma_0, \Gamma_1, \cdots, \Gamma_{n-1})$ is a given natural link, and that $\Gamma_i \sim \boldsymbol{\alpha}_i$ $(i = 0, 1, \cdots, n-1)$. Set $\Gamma_n = \Gamma_0$, $\boldsymbol{\alpha}_n = \boldsymbol{\alpha}_0$ and φ_i equal to the angle formed by Γ_{i-1} and Γ_i $(i = 1, 2, \cdots, n)$.

Let $\Gamma_i' \sim \boldsymbol{\delta}_i$ be the direct sphere of antisimilitude of Γ_{i-1} and Γ_i and Δ_i be the Moebius inversion generated by $\boldsymbol{\delta}_i$. In other words

$$\boldsymbol{\delta}_i = \frac{\boldsymbol{\alpha}_{i-1} + \boldsymbol{\alpha}_i}{2\cos\varphi_i/2} ,$$

$$\Delta_i\boldsymbol{\xi} = \boldsymbol{\xi} - 2(\boldsymbol{\xi}, \boldsymbol{\delta}_i)\boldsymbol{\delta}_i .$$

The results of § 3.2 imply that the Moebius transformation which defines in Γ_0 the Schottky model corresponding to L is given by the product of inversions

$$\tau = \Delta_n \Delta_{n-1} \cdots \Delta_1 .$$

The conformal parameter of L is related in a simple way to the eigenvalues of τ.

The study of this transformation can be simplified if we introduce a complex coordinate in Γ_0 and make use of the results established in § 3.2.

To construct a stereographic projection $p = \varphi z$ of the complex plane π onto the sphere Γ_0 we can proceed in the following way:

We first choose a basis in $\mathscr{P}_4$ which consists of $\boldsymbol{\alpha}_0$ and four other normalized vectors $\boldsymbol{\gamma}_0$, $\boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_2$, $\boldsymbol{\gamma}_1$ representing respectively

(a) $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$: two distinct real points of Γ_0.

(b) $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$: two real spheres containing the points represented by $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$, orthogonal to each other and to the sphere Γ_0.

We then normalize $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ so that

$$(\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1) = -1/2 \text{ ,}$$ [15]

and set for each $z = x + iy$ of π:

$$\varphi z = \lambda_0(\boldsymbol{\gamma}_0 + x\boldsymbol{\varepsilon}_1 + y\boldsymbol{\varepsilon}_2 + \{x^2 + y^2\}\boldsymbol{\gamma}_1) ,$$

where the indeterminate λ_0 is only restricted to be a positive real number.

Introducing the two complex points

$$\bar{\boldsymbol{\gamma}} = (\boldsymbol{\varepsilon}_1 - i\boldsymbol{\varepsilon}_2)/2, \quad \boldsymbol{\gamma} = (\boldsymbol{\varepsilon}_1 + i\boldsymbol{\varepsilon}_2)/2 ,$$

the equation of φ assumes the more suggestive form

$$\varphi z = \lambda_0(\boldsymbol{\gamma}_0 + z\bar{\boldsymbol{\gamma}} + \bar{z}\gamma + z\bar{z}\boldsymbol{\gamma}_1) . \tag{12}$$

To find the inverse of φ, we observe that if

$$\boldsymbol{\xi} = \lambda_0\boldsymbol{\gamma}_0 + \lambda\bar{\boldsymbol{\gamma}} + \lambda'\boldsymbol{\gamma} + \lambda_1\boldsymbol{\gamma}_1$$

represents a real point of Γ_0 we must have $\bar{\boldsymbol{\xi}} = \boldsymbol{\xi}$ and $(\boldsymbol{\xi}, \boldsymbol{\xi}) = 0$; this yields

$$\lambda' = \bar{\lambda}$$

and

$$\lambda_0\lambda_1 = \lambda\bar{\lambda} .$$

This means that such a $\boldsymbol{\xi}$ can always be written in the form

[15] The scalar product of two normalized vectors of $\mathscr{P}_4$ which represent real points of E_3 is always negative.

$$\boldsymbol{\xi} = \lambda_0\Big(\boldsymbol{\gamma}_0 + \frac{\lambda}{\lambda_0}\bar{\boldsymbol{\gamma}} + \frac{\bar{\lambda}}{\lambda_0}\boldsymbol{\gamma} + \frac{\lambda}{\lambda_0}\frac{\bar{\lambda}}{\lambda_0}\boldsymbol{\gamma}_1\Big)\,.$$ [16]

Thus we can set

$$\varphi^{-1}\boldsymbol{\xi} = \frac{\lambda}{\lambda_0}\,. \tag{13}$$

In view of the results of § 3.2, the mapping $\tau = \Delta_n \cdots \Delta_i$, restricted to Γ_0, is a loxodromic (in particular hyperbolic) Moebius transformation. Let us denote then by $\boldsymbol{A}$ and $\boldsymbol{B}$ its two fixed points in Γ_0 and assume that $\boldsymbol{A}$ is the source and $\boldsymbol{B}$ is the sink.

If we set

$$\boldsymbol{\gamma}_0 = \frac{1}{\overline{AB}}(1,\ \boldsymbol{A},\ \boldsymbol{A}^2),\ \boldsymbol{\gamma}_1 = \frac{1}{\overline{AB}}(1,\ \boldsymbol{B},\ \boldsymbol{B}^2)\,,$$

and take for $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ any two spheres satisfying condition (b), since φ^{-1} maps $\boldsymbol{A}$ onto the origin and $\boldsymbol{B}$ onto the point at infinity of π, the equation of the Moebius transformation $\tau^* = \varphi^{-1}\tau\varphi$ of π will assume the simple form

$$\tau^* z = \rho e^{i\theta} z\,, \tag{14}$$

with $\rho > 1$ and $-\pi < \theta \leqq \pi$.

Thus for each point

$$\boldsymbol{\xi} = \lambda_0\boldsymbol{\gamma}_0 + \lambda\bar{\boldsymbol{\gamma}} + \bar{\lambda}\boldsymbol{\gamma} + \frac{\lambda\bar{\lambda}}{\lambda_0}\boldsymbol{\gamma}_1\,,$$

we have (using (13), (14) and (12)):

$$\begin{aligned}\tau\boldsymbol{\xi} &= \varphi\tau^*\varphi^{-1}\boldsymbol{\xi} = \varphi\tau^*\frac{\lambda}{\lambda_0} = \varphi\rho e^{i\theta}\,\frac{\lambda}{\lambda_0}\\ &= \lambda_0'\Big(\boldsymbol{\gamma}_0 + \frac{\rho e^{i\theta}\lambda}{\lambda_0}\,\bar{\boldsymbol{\gamma}} + \frac{\overline{\rho e^{i\theta}\lambda}}{\lambda_0}\,\boldsymbol{\gamma} + \rho^2\,\frac{\lambda\bar{\lambda}}{\lambda_0^2}\boldsymbol{\gamma}_1\Big)\end{aligned}$$

or

$$\tau\boldsymbol{\xi} = \frac{\lambda_0'}{\lambda_0}\Big(\lambda_0\boldsymbol{\gamma}_0 + \rho e^{i\theta}\lambda\bar{\boldsymbol{\gamma}} + \rho\bar{e}^{i\theta}\bar{\lambda}\boldsymbol{\gamma} + \rho^2\frac{\lambda\bar{\lambda}}{\lambda_0}\boldsymbol{\gamma}_1\Big)\,.$$

A priori the indeterminate λ_0' is only restricted to be a positive real number. However, the ratio λ_0'/λ_0 depends solely upon the transformation τ.

[16] By λ/λ_0 we mean an extended valued complex number.

In fact, since τ preserves the scalar product of $\mathscr{P}_4$ we must have

$$(\tau\boldsymbol{\xi},\ \boldsymbol{\gamma}_0) = (\boldsymbol{\xi},\ \tau^{-1}\boldsymbol{\gamma}_0)\ ,$$
$$(\tau\boldsymbol{\xi},\ \boldsymbol{\gamma}_1) = (\boldsymbol{\xi},\ \tau^{-1}\boldsymbol{\gamma}_1)\ , \tag{15}$$
$$(\tau_0\boldsymbol{\gamma},\ \tau\boldsymbol{\gamma}_1) = (\boldsymbol{\gamma}_0,\ \boldsymbol{\gamma}_1)\ .$$

Since $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ represent fixed points of τ

$$\tau\boldsymbol{\gamma}_0 = \mu_0\boldsymbol{\gamma}_0\ ,$$
$$\tau\boldsymbol{\gamma}_1 = \mu_1\boldsymbol{\gamma}_1\ ,$$

for some positive real numbers μ_0 and μ_1. Substituting in the equations (15) we obtain

$$\frac{\lambda_0'}{\lambda_0} = \frac{1}{\rho}, \qquad \mu_0 = \frac{1}{\rho}, \qquad \mu_1 = \rho\ .$$

This gives

$$\tau\boldsymbol{\xi} = \frac{\lambda_0}{\rho}\boldsymbol{\gamma}_0 + \lambda e^{i\theta}\bar{\boldsymbol{\gamma}} + \bar{\lambda}\bar{e}^{i\theta}\boldsymbol{\gamma} + \frac{\lambda\bar{\lambda}}{\lambda_0}\rho\boldsymbol{\gamma}_1\ ,$$

and since λ is arbitrary

$$\tau\bar{\boldsymbol{\gamma}} = e^{i\theta}\bar{\boldsymbol{\gamma}}\ ,$$
$$\tau\boldsymbol{\gamma} = e^{-i\theta}\boldsymbol{\gamma}\ .$$

Finally, the relation $\Delta_i\boldsymbol{\alpha}_{i-1} = -\boldsymbol{\alpha}_i$ for $i = 1, 2, \cdots, n$ implies

$$\tau\boldsymbol{\alpha}_0 = (-1)^n\boldsymbol{\alpha}_0\ .$$

With this we have shown that the eigenvalues of τ are $(-1)^n$, $1/\rho$, ρ, $e^{i\theta}$, $e^{-i\theta}$. Thereby the relation between these eigenvalues and the conformal parameter of L is established.

Only little has to be added concerning the general case.

If $\Xi \sim (L_1, L_2, \cdots, L_g)$ is a given natural M-surface and Γ_0 is the common initial sphere of the L_i's, we operate separately on each link L_i and determine the transformation τ_i generated by the spheres of L_i.

These transformations alone carry complete information regarding the conformal parameters of Ξ.

However, unlike the case of a single link, the eigenvalues of the τ_i's are not sufficient by themselves to characterize the conformal parameters of Ξ, since they yield only the first g of them. The real eigenvectors of these transformations have to be determined also, and among them those representing the fixed points $\boldsymbol{A}_i$, $\boldsymbol{B}_i$ of each τ_i have to be selected. Then, according to the definition (formulas (8) of § 1.5), the remaining parameters are given by the coordinates of the points

B_2; A_3, B_3; $\cdots$; A_g, B_g in a coordinate system in Γ_0 for which A_1B_1 and A_2 have coordinates 0, ∞ and 1 respectively.

4.4. With the same notation as in § 2.3, let $L \sim (\Gamma_0, \Gamma_1, \cdots, \Gamma_{n-1})$ be a natural link and Λ_i be the intersection of the sphere Γ_{i-1} with the sphere Γ_i. If $\omega = \rho e^{i\theta}$ is the conformal parameter of L, we shall say that ρ is the *thinness* and θ the *torsion* of L.

The thinness of the link L can be estimated in terms of the capacities of the annular domains $\Lambda_{i-1}^+ \cap \Lambda_i^-$. In fact, we have the following:

THEOREM. *Suppose that each annulus $\Lambda_{i-1}^+ \cap \Lambda_i^-$ has a capacity c_i satisfying the inequality*

$$c_i \leqq \frac{1}{\log \rho_i} \tag{16}$$

for some $\rho_i > 1$. Then the thinness ρ of L satisfies the inequality

$$\rho \geqq \rho_1 \rho_2 \cdots \rho_n , \tag{17}$$

and the equal sign holds if and only if (16) *are equalities and the spheres Γ_0, Γ_1, $\cdots$, Γ_{n-1} are all orthogonal to the spheres of a hyperbolic pencil.*

To prove this theorem, we need a few preliminary considerations.

If τ is a loxodromic transformation of a sphere Γ; i.e. if for some coordinate system in Γ

$$\frac{\tau z - \alpha}{\tau z - \beta} = \omega \frac{z - \alpha}{z - \beta}$$

the number $|\omega|$ (which can always be supposed greater than one) will be called the "stretching factor" of τ.

Let Λ and Λ' be two circles of Γ having no points in common and suppose that α_0 and β_0 are the two points of Γ which belong to the elliptic pencil generated by Λ and Λ'. Let α_0 and β_0 be ordered in such a way that the disks $\Lambda(\alpha_0)$ and $\Lambda'(\beta_0)$ are exterior to each other.

LEMMA I. *Among all Moebius transformations of Γ which map $\Lambda(\alpha_0)$ onto $\Lambda'(\alpha_0)$ only those which admit α_0 and β_0 as fixed points have the smallest stretching factor.*

Proof. Let us choose a complex coordinate in Γ which is such that $\alpha_0 = 0$, $\beta_0 = \infty$ and Λ has the equation $|z| = 1$. The equation of Λ' will then be

$$|z| = \rho$$

for a suitable $\rho > 1$.

If τ is a Moebius transformation of Γ which sends $\Lambda(\alpha_0)$ onto $\Lambda'(\alpha_0)$ its equation can be written in the form

$$\frac{\tau z - \alpha}{\tau z - \beta} = \omega \frac{z - \alpha}{z - \beta}$$

with $\alpha \in \Lambda(\alpha_0)$, $\beta \in \Lambda'(\beta_0)$ and $|\omega| > 1$[17]. Now, τ must send the points $1/\bar{\alpha}$ and $1/\bar{\beta}$ respectively onto the points $\rho^2/\bar{\alpha}$ and $\rho^2/\bar{\beta}$. In other words, we must have

$$\text{(18)a, b} \qquad \frac{\rho^2/\bar{\alpha} - \alpha}{\rho^2/\bar{\alpha} - \beta} = \omega \frac{1/\bar{\alpha} - \alpha}{1/\bar{\alpha} - \beta}, \quad \frac{\rho^2/\bar{\beta} - \alpha}{\rho^2/\bar{\beta} - \beta} = \omega \frac{1/\bar{\beta} - \alpha}{1/\bar{\beta} - \beta}$$

and

$$\text{(19)} \qquad (\alpha, \beta, 1/\bar{\alpha}, 1/\bar{\beta}) = (\alpha, \beta, \rho^2/\bar{\alpha}, \rho^2/\bar{\beta}) .$$

Equation (18)a gives

$$\omega = \frac{\rho^2 - \alpha\bar{\alpha}}{1 - \alpha\bar{\alpha}} \cdot \frac{1 - \bar{\alpha}\beta}{\rho^2 - \bar{\alpha}\beta} ,$$

equation (19), after a few eliminations, yields

$$\frac{\rho^2 - \bar{\alpha}\beta}{1 - \bar{\alpha}\beta} \cdot \frac{\rho^2 - \alpha\bar{\beta}}{1 - \alpha\bar{\beta}} = \rho^2 .$$

Therefore we have

$$|\omega| = \frac{1}{\rho} \left| \frac{\rho^2 - \alpha\bar{\alpha}}{1 - \alpha\bar{\alpha}} \right| .$$

But $\alpha\bar{\alpha} < 1$ (since $\alpha \in \Lambda(\alpha_0)$), thus

$$|\omega| \geqq \rho$$

and the equality sign holds if and only if $\alpha\bar{\alpha} = 0$. However, when $\alpha = 0$ equations (18)a,b give $\beta = \infty$. This proves the assertion.

Let the Moebius transformation $(\tau z - \alpha)/(\tau z - \beta) = \omega(z - \alpha)/(z - \beta)$ define in Σ a Schottky model $M(\tau)$. Any circle Λ such that the closed disks $\overline{\Lambda}(\alpha)$ and $\overline{\tau\Lambda(\beta)}$ are mutually exclusive cuts $M(\tau)$ into a region $\Lambda(\beta) \cap \tau\Lambda(\alpha)$ which is an annulus. As a consequence of the previous lemma we can show that:

[17] This follows from an argument similar to that presented in § 3.2.

LEMMA II. *Among all circles Λ for which $\overline{\Lambda(\alpha)}$ and $\overline{\tau\Lambda(\beta)}$ are disjoint, only those belonging to the pencil $P(\alpha, \beta)$ cut $M(\tau)$ into an anulus of minimum capacity.*

Proof. Let α_0 and β_0 be the two points belonging to the elliptic pencil generated by Λ and $\tau\Lambda$, and assume that $\alpha_0 \in \Lambda(\alpha)$ and $\beta_0 \in \tau\Lambda(\beta)$.[18] If c denotes the capacity of the anulus $\Lambda(\beta) \cap \tau\Lambda(\alpha)$ the stretching factor of every Moebius transformation which sends $\Lambda(\alpha)$ onto $\tau\Lambda(\alpha)$ and admits α_0 and β_0 as fixed points is given by $\rho = e^{1/c}$.

By Lemma I we must have

$$|\omega| \geqq e^{1/c}$$

or, which is the same (since $|\omega| > 1$)

$$c \geqq 1/\log|\omega|$$

with equality possible if and only if $\alpha = \alpha_0$ and $\beta = \beta_0$. Q.E.D.

We can now give a proof of the theorem.

If c denotes the capacity of the annulus

$$\begin{aligned} \Lambda_1^- - \tau\Lambda_1^- = \Delta_n \cdots \Delta_2\{\Lambda_1 + \Lambda_2^+ \cap \Lambda_3^-\} + \cdots \\ + \Delta_n\{\Lambda_{n-1} + \Lambda_{n-1}^+ \cap \Lambda_n^-\} + \Lambda_0 + \Lambda_0^+ \cap \Lambda_1^- , \end{aligned}$$

from a well known inequality of potential theory (cfr. [7]) we obtain

$$\frac{1}{c} \geqq \frac{1}{c_1} + \frac{1}{c_2} + \cdots + \frac{1}{c_n}, \tag{20}$$

and the equality sign holds if and only if the circles $\Delta_n \cdots \Delta_2\Lambda_1$, $\Delta_n \cdots \Delta_3\Lambda_2, \cdots, \Delta_n\Lambda_{n-1}, \Lambda_0, \Lambda_1$ belong to the same pencil. Since the thinness ρ of the link L is equal to the stretching factor of the transformation $\Delta_n\Delta_{n-1} \cdots \Delta_1$, from Lemma II we get

$$\rho \geqq e^{1/c}, \tag{21}$$

thus from (20) and (16) the desired inequality follows.

To prove the last statement of the theorem, we observe that the equal sign will occur in (17) if and only if, (16) being equalities, equality holds simultaneously in (20) and (21). However, this happens if and only if all the circles $\Delta_n \cdots \Delta_2\Lambda_1$, $\Delta_n \cdots \Delta_3\Lambda_2, \cdots, \Delta_n\Lambda_{n-1}, \Lambda_0, \Lambda_1$ belong to the pencil generated by the fixed points α, β of the transformation

$$\tau = \Delta_n\Delta_{n-1} \cdots \Delta_1 .$$

Let then Γ be any sphere orthogonal to Γ_0 and containing α and

[18] This is always the case after a suitable labeling of α_0 and β_0.

β. Since Γ is orthogonal to Λ_0, Γ will be orthogonal to Γ_{n-1} and Γ'_n (the direct sphere of antisimilitude of Γ_{n-1} and Γ_0.)

Therefore $\Delta_n\Gamma = \Gamma$ and consequently Γ is orthogonal to $\Delta_n(\Delta_n\Lambda_{n-1}) = \Lambda_{n-1}$. Γ will then be orthogonal to Γ_{n-2} and to Γ'_{n-1} (the direct sphere of antisimilitude of Γ_{n-2} and Γ_{n-1}). But this implies that $\Delta_{n-1}\Delta_n\Gamma = \Gamma$ and consequently Γ is orthogonal to $\Delta_{n-1}\Delta_n(\Delta_n\Delta_{n-1}\Lambda_{n-2}) = \Lambda_{n-2}$, etc. Proceeding in this fashion we obtain that Γ is also orthogonal to $\Gamma_{n-3}, \Gamma_{n-4}, \cdots, \Gamma_2, \Gamma_1$. The spheres orthogonal to Γ_0 and containing α and β form a hyperbolic pencil.

Conversely if the spheres $\Gamma_0, \Gamma_1, \cdots, \Gamma_{n-1}$ are orthogonal to the spheres of a hyperbolic pencil P, so will also be the spheres of antisimilitude $\Gamma'_2, \Gamma'_2, \cdots, \Gamma'_n$; consequently each sphere of P will be invariant under any of the transformations $\Delta_1, \Delta_2, \cdots, \Delta_n$.

We can then easily deduce that the circles $\Delta_n \cdots \Delta_2\Lambda_1, \cdots, \Delta_n\Lambda_{n-1}, \Lambda_0$ are orthogonal to the spheres of P and thus they all belong to the pencil generated by the two points α and β intersection of Γ_0 and the spheres of P. But α and β are the fixed points of the transformation $\Delta_n\Delta_{n-1} \cdots \Delta_1$.

Our proof is thus complete.

Although it will not be needed in the following we would like to point out that the inequality (17) holds also for general links. In fact, Lemma II is valid in the stronger form:

"*Among all smooth Jordan curves Λ for which $\overline{\Lambda(\alpha)}$ and $\overline{\tau\Lambda(\beta)}$ are disjoint, only the circles of the pencil $P(\alpha, \beta)$ cut $M(\tau)$ into an annulus of minimum capacity.*"

This statement follows from standard potential theoretical considerations.

5. Some special links.

5.1. Let π_1 denote the w-plane and w_1, w_2 two complex numbers for which

$$\mathfrak{Im}\, w_1/w_2 < 0 .$$

Let G denote the group generated by the translations

$$\begin{aligned} \tau_1 w &= w + w_1 \\ \tau_2 w &= w + w_2 . \end{aligned} \tag{1}$$

If we identify the points of π which are images of each other under the transformations of G, we obtain a Riemann surface of genus one $\Gamma(w_1, w_2)$.

The surface $\Gamma(w_1, w_2)$ can also be thought of as the parallelogram

$$\mathscr{P} = \{w : w = \lambda w_1 + \mu w_2;\ 0 \leqq \lambda \leqq 1,\ 0 \leqq \mu \leqq 1\} .$$

with opposite sides identified by the transformations (1).

This standard construction generates every Riemann surface of genus one: as a matter of fact, as w_1 and w_2 vary, $\Gamma(w_1, w_2)$ assumes every conformal type and each an infinite number of times.

It is clear that two Riemann surfaces $\Gamma(w_1, w_2)$ and $\Gamma(w_1', w_2')$ are conformally equivalent if and only if the lattices

$$\begin{matrix} L \sim \{m_1 w_1 + m_2 w_2\} \\ L' \sim \{m_1 w_1' + m_2 w_2'\} \end{matrix} \qquad m_1,\ m_2 = 0,\ \pm 1,\ \pm 2, \cdots$$

can be superimposed by a similarity. Now it is well known that this is possible if and only if the two ratios

$$\nu = \frac{w_1}{w_2} \text{ and } \nu' = \frac{w_1'}{w_2'}$$

are images of each other under a transformation of the restricted unimodular group; in other words if and only if there exist integers a, b, c, d for which $ad - bc = 1$ and

$$\nu' = \frac{a\nu + b}{c\nu + d} .$$

The set

$$\begin{aligned} \mathfrak{M} = \{\nu : \mathfrak{Im}\, \nu < 0;\ -1/2 < \mathfrak{Re}\, \nu \leqq 1/2;\ |\nu| > 1 \\ \text{for } \mathfrak{Re}\, \nu < 0;\ |\nu| \geqq 1 \text{ for } \mathfrak{Re}\, \nu \geqq 0\} \end{aligned}$$

is a fundamental region of the restricted unimodular group; thus two Riemann surfaces $\Gamma(w_1, w_2)$ and $\Gamma'(w_1, w_2)$ will be conformally equivalent if and only if the complex numbers w_1/w_2 and w_1'/w_2' have the same image point in $\mathfrak{M}$.

If we have a Schottky model $M(\tau)$ defined by a Moebius transformation

$$\frac{\tau z - \alpha}{\tau z - \beta} = \rho e^{i\theta} \frac{z - \alpha}{z - \beta} \qquad (\rho > 1;\ -\pi < \theta \leqq \pi)$$

of some sphere Σ, a conformally equivalent model is given by the surface $\Gamma(\log \rho + i\theta,\ 2\pi i)$. In fact, the function $w = \log (z - \alpha)/(z - \beta)$ defines a conformal mapping of $M(\tau)$ onto $\Gamma(\log \rho + i\theta,\ 2\pi i)$.

The point

$$\nu = \frac{\theta}{2\pi} - i \frac{\log \rho}{2\pi}$$

belongs to $\mathfrak{M}$ if

$$(2)\qquad \begin{aligned} &\left(\frac{\theta}{2\pi}\right)^2+\left(\frac{\log\rho}{2\pi}\right)^2 \geqq 1 \text{ when } \theta \geqq 0 \\ &\left(\frac{\theta}{2\pi}\right)^2+\left(\frac{\log\rho}{2\pi}\right)^2 > 1 \text{ when } \theta < 0\,. \end{aligned}$$

Thus can we conclude that two distinct Schottky models $M(\tau)$ and $M(\tau')$ whose conformal parameters $\rho e^{i\theta}$ and $\rho' e^{i\theta'}$ satisfy the inequalities (2) are never conformally equivalent.

We shall proceed to show that there exist natural links which are not conformally equivalent to any of the models $\Gamma(\log\rho,\ 2\pi i)$.

5.2. Let $\boldsymbol{\alpha} = 1/R(1,\ \boldsymbol{c},\ \boldsymbol{c}^2 - R^2)$, $\alpha_1 = 1/R(1,\ \boldsymbol{c}_1,\ \boldsymbol{c}_1^2 - R^2)$ and $\boldsymbol{\alpha}_2 = 1/R(1,\ \boldsymbol{c}_2,\ \boldsymbol{c}_2^2 - R^2)$ be three given spheres[19] of equal radius and suppose that $\overline{\boldsymbol{c}\boldsymbol{c}_1} = \overline{\boldsymbol{c}\boldsymbol{c}_2} = 2\delta,\ \delta < R < \overline{\boldsymbol{c}_1\boldsymbol{c}_2}/2$.

Let Λ_1 and Λ_2 be the circles of intersection of $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}_2$ respectively, π_1 and π_2 be the planes containing Λ_1 and Λ_2, d the intersection of π_1, and π_2 (proper or improper), $\boldsymbol{p}$ the intersection of d with the plane through $\boldsymbol{c}$ perpendicular to d, and $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ represent the points of contact of the two planes through d which are tangent to α.

We would like to compute the capacity of the annulus

$$D = \Lambda_1(\boldsymbol{p}_2) \cap \Lambda_2(\boldsymbol{p}_1)\,.$$

To do this it is sufficient to compute the stretching factor of a Moebius transformation of $\boldsymbol{\alpha}$ which admits $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ as fixed points and sends $\Lambda_1(\boldsymbol{p}_1)$ onto $\Lambda_2(\boldsymbol{p}_1)$.

Let π denote the plane through $\boldsymbol{c}$ and d, and $\boldsymbol{\sigma}$ the sphere through Λ_2 which is orthogonal to $\boldsymbol{\alpha}$. Clearly the product

$$\tau = \tau_\sigma \tau_\pi$$

of the inversions τ_π and τ_σ with respect to π and $\boldsymbol{\sigma}$ generates a transformation of $\boldsymbol{\alpha}$ which is of the type requested. We shall compute its equation.

We indicate by $\boldsymbol{a}$ and $\boldsymbol{b}$ two unit vectors with the directions of $\boldsymbol{c}_1\boldsymbol{c}_2$ and $\boldsymbol{c}\boldsymbol{p}$ respectively. Let us assume for simplicity that the origin of the coordinate system of E_3 is at $\boldsymbol{c}$. We then have

$$\boldsymbol{\alpha} = \frac{1}{R}(1,\ 0,\ -R^2)$$

$$\pi = (0,\ \boldsymbol{a},\ 0)\,.$$

[19] Occasionally we shall make use of the same symbol to denote a geometric object and its representative in $\mathscr{P}_4$.

Setting $\varphi = \widehat{pcc_1} = \widehat{pcc_2}$ and $\psi = \widehat{pcp_1} = \widehat{pcp_2}$:

$$\boldsymbol{\pi}_1 = (0,\ -\sin\varphi \boldsymbol{a} + \cos\varphi \boldsymbol{b},\ 2\delta)\ ,$$

$$\boldsymbol{\pi}_2 = (0,\ \sin\varphi \boldsymbol{a} + \cos\varphi \boldsymbol{b},\ 2\delta)\ ,$$

$$\boldsymbol{\gamma}_1 = \frac{1}{\overline{\boldsymbol{p}_1\boldsymbol{p}_2}}(1,\ \boldsymbol{p}_1^2,\ \boldsymbol{p}_1^2) = \frac{1}{2R\sin\psi}(1,\ -R\sin\psi\boldsymbol{a} + R\cos\psi\boldsymbol{b},\ R^2)\ ,$$

$$\boldsymbol{\gamma}_2 = \frac{1}{\overline{\boldsymbol{p}_1\boldsymbol{p}_2}}(1,\ \boldsymbol{p}_2,\ \boldsymbol{p}_2^2) = \frac{1}{2R\sin\psi}(1,\ R\sin\psi\boldsymbol{a} + R\cos\psi\boldsymbol{b},\ R^2)\ .$$

By its definition $\boldsymbol{\sigma}$ belongs to the pencil generated by $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$, as well as to the pencil generated by $\boldsymbol{\alpha}$ and $\boldsymbol{\pi}_2$.

Thus for suitable values of μ, λ, μ', λ'

$$\boldsymbol{\sigma} = \mu\boldsymbol{\gamma}_1 + \lambda\boldsymbol{\gamma}_2 = \mu'\boldsymbol{\alpha} + \lambda'\boldsymbol{\pi}_2\ . \tag{3}$$

Observing that since $(\boldsymbol{\sigma}, \boldsymbol{\sigma}) = 1$ and $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = -1/2$ we must have $\lambda = -1/\mu$, equating the middle components of (3) we obtain

$$\frac{\lambda^2 + 1}{\lambda^2 - 1} = \frac{\tan\psi}{\tan\varphi}\ .$$

Now

$$\tau_\pi\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_2,\ \tau_\pi\boldsymbol{\gamma}_2 = \boldsymbol{\gamma}_1\ ,$$

$\tau_\sigma\boldsymbol{\gamma}_1 = \boldsymbol{\gamma}_1 - 2(\boldsymbol{\gamma}_1, \boldsymbol{\sigma})\boldsymbol{\sigma} = \lambda^2\boldsymbol{\gamma}_2$ and analogously $\tau_\sigma\boldsymbol{\gamma}_2 = (1/\lambda^2)\boldsymbol{\gamma}_1$.

Thus for the stretching factor ρ of the product $\tau_\sigma\tau_\pi$ we get

$$\rho = \lambda^2 = \frac{\tan\varphi + \tan\psi}{\tan\varphi - \tan\psi} \tag{4}$$

this determines the capacity of D[20].

5.3. It is easy to show that every point of $\mathfrak{M}$ which lies in the imaginary axis can be obtained as an image of an imbedded surface.

In fact, the image of a torus in $\mathfrak{M}$ is always pure imaginary, and as we vary the radius of the generating circle, keeping the center fixed, we can describe the whole imaginary axis.

We shall exhibit a family of natural links with the same property, and at the same time illustrate our way of computing the conformal parameters of natural links.

Let $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ be unit vectors forming a left handed orthogonal triplet and set

---

[20] The obvious argument based on the fact that the stereographic projection is a cross-ratio-preserving transformation would lead to the same result with more or less the same effort.

$$\boldsymbol{\alpha}_i = \frac{1}{R}\left(1,\ \cos i\frac{2\pi}{n}\boldsymbol{a} + \sin i\frac{2\pi}{n}\boldsymbol{b},\ 1 - R^2\right)$$

(assume $n \geqq 3$). It can be readily verified that $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{n-1}$ define a natural link for every value of R greater than $\sin \pi/n$ and less than one.

Let Λ_i denote the intersection of $\boldsymbol{\alpha}_{i-1}$ with $\boldsymbol{\alpha}_i$, and the sets Λ_i^-, Λ_i^+ have the same meaning as in § 2.3. To compute the conformal parameters of the link

$$L(n,\ R) = \Lambda_0 + \Lambda_0^+ \cap \Lambda_1^- + \cdots + \Lambda_{n-1} + \Lambda_{n-1}^+ \cap \Lambda_n^-\ ,$$

according to the results of § 4.3 we should study the transformation τ product of successive inversions with respect to the spheres

$$\boldsymbol{\delta}_i = \frac{R}{2}\frac{\boldsymbol{\alpha}_{i-1} + \boldsymbol{\alpha}_i}{\sqrt{R^2 - \sin^2 \pi/n}} \qquad (i = 1, 2, \cdots, n).$$

This does not present any difficulty. In fact, we observe that each of the $\boldsymbol{\alpha}_i$'s is orthogonal to the plane

$$\boldsymbol{\varepsilon}_1 = (0,\ \boldsymbol{c},\ 0)$$

and the sphere

$$\boldsymbol{\varepsilon}_2 = \left(\frac{1}{\sqrt{1 - R^2}},\ 0,\ -\sqrt{1 - R^2}\right).$$

Thus all the spheres of $P(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2)$ (the pencil generated by $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$) are orthogonal to each of the $\boldsymbol{\alpha}_i$'s and therefore also to each of the $\boldsymbol{\delta}_i$'s. This implies that the spheres of $P(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2)$ are all invariant under the transformation τ. Consequently also the points $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ which $\boldsymbol{\alpha}_0$ has in common with the spheres of the pencil $P(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2)$ are invariant under τ. We can then conclude that τ admits the decomposition

$$\begin{aligned}
\tau\boldsymbol{\gamma}_0 &= \frac{1}{\rho}\boldsymbol{\gamma}_0 \\
\tau\boldsymbol{\varepsilon}_1 &= \boldsymbol{\varepsilon}_1 \\
\tau\boldsymbol{\varepsilon}_2 &= \boldsymbol{\varepsilon}_2 \\
\tau\boldsymbol{\alpha}_0 &= (-1)^n\boldsymbol{\alpha}_0 \\
\tau\boldsymbol{\gamma}_1 &= \rho\boldsymbol{\gamma}_1\ ,
\end{aligned}$$

with a suitable $\rho > 0$ (if $\boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}_1$ are properly labeled ρ will result greater than one).

Thus the torsion of $L(n,\ R)$ vanishes independently of n and R. To determine the thinness ρ we use the formula (4) of last section and obtain for the capacity c_i of each anulus $\Lambda_{i-1} + \Lambda_{i-1}^{+} \cap \Lambda_i^{-}$

$$c_i = \left\{\log \frac{R\cos\pi/n + \sin\pi/n\sqrt{1-R^2}}{R\cos\pi/n - \sin\pi/n\sqrt{1-R^2}}\right\}^{-1} .$$

Applying the theorem of § 4.4 we obtain

$$\rho = \left(\frac{R\cos\pi/n + \sin\pi/n\sqrt{1-R^2}}{R\cos\pi/n - \sin\pi/n\sqrt{1-R^2}}\right)^{n} . \tag{5}$$

Clearly for any given $n > 3$ this function increases from 1 to ∞ as R decreases from 1 to $\sin\pi/n$.

It is interesting to note that if R is kept fixed in (5) and we let n tend to infinity we obtain

$$\lim_{n\to\infty} \rho = e^{2\pi\frac{\sqrt{1-R^2}}{R}} .$$

This result is not surprising since the link $L(n, R)$ then approaches the torus enveloped by a sphere of radius R as its center describes a circle of radius one,

5.4. The fact that each link $L(n, R)$ has torsion zero could have been predicted. We can show that if a natural link admits a plane of symmetry or a sphere of inversion (which amounts to the same thing) then its torsion must vanish.

We shall consider two representative cases.

Case 1. *All the spheres of the link are orthogonal to the sphere of inversion.*

Let $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{n-1}$ be the generating spheres and $\boldsymbol{\varepsilon}$ be a real sphere such that

$$(\boldsymbol{\alpha}_i, \boldsymbol{\varepsilon}) = 0 \qquad (i = 0, 1, \cdots, n-1) .$$

From this follows that the spheres of antisimilitude $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \cdots, \boldsymbol{\delta}_n$ will also be orthogonal to $\boldsymbol{\varepsilon}$ and therefore

$$\tau\boldsymbol{\varepsilon} = \Delta_n\Delta_{n-1}\cdots\Delta_1\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon} . \tag{6}$$

We suppose that $\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}, \bar{\boldsymbol{\gamma}}$ decompose τ, and set (as in § 4.3)

$$\tau\boldsymbol{\gamma}_0 = \frac{1}{\rho}\boldsymbol{\gamma}_0,\ \tau\boldsymbol{\gamma}_1 = \rho\boldsymbol{\gamma}_1,\ \tau\bar{\boldsymbol{\gamma}} = e^{i\theta}\bar{\boldsymbol{\gamma}},\ \tau\boldsymbol{\gamma} = e^{-i\theta}\boldsymbol{\gamma} ,$$

Since $\boldsymbol{\varepsilon}$ is orthogonal to $\boldsymbol{\alpha}_0$ it must be of the form

$$\boldsymbol{\varepsilon} = \lambda_0\boldsymbol{\gamma}_0 + \lambda_1\boldsymbol{\gamma}_1 + \lambda\bar{\boldsymbol{\gamma}} + \bar{\lambda}\boldsymbol{\gamma} ; \tag{7}$$

however, for a natural link $\rho > 1$ (cfr. theorem of § 4.4), and thus the hypothesis $e^{i\theta} \neq 1$ is incompatible with (6) and (7).

Case 2. *The spheres of the link are interchanged by the sphere of inversion.* By means of two or more additional spheres we can reduce (without altering the conformal structure of the link) every possible situation to the following one:

The spheres $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{n-1}$ are an even number $n = 2p$ and furthermore the sphere of inversion $\boldsymbol{\varepsilon}$ is such that

$$\tau_\varepsilon \boldsymbol{\alpha}_0 = \boldsymbol{\alpha}_0,\ \tau_\varepsilon \boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_{n-1},\ \tau_\varepsilon \boldsymbol{\alpha}_2 = \boldsymbol{\alpha}_{n-2},\ \cdots,\ \tau_\varepsilon \boldsymbol{\alpha}_p = \boldsymbol{\alpha}_p.$$ [21]

The spheres of antisimilitude will then be related in the following way

$$\boldsymbol{\delta}_n = \tau_\varepsilon \boldsymbol{\delta}_1,\ \boldsymbol{\delta}_{n-1} = \tau_\varepsilon \boldsymbol{\delta}_2,\ \cdots,\ \boldsymbol{\delta}_{p+1} = \tau_\varepsilon \boldsymbol{\delta}_p\ .$$

This implies that the transformation $\tau = \Delta_n \Delta_{n-1} \cdots \Delta_1$ can be written in the form

$$\tau = \tau_\varepsilon \Delta_1 \Delta_2 \cdots \Delta_p \tau_\varepsilon \Delta_p \Delta_{p-1} \cdots \Delta_1$$

or, setting $\sigma = \Delta_p \Delta_{p-1} \cdots \Delta_1$:

$$\tau = \tau_\varepsilon \sigma^{-1} \tau_\varepsilon \sigma\ .$$

Assuming that $\boldsymbol{\gamma}_0, \boldsymbol{\gamma}_1$ are the source and the sink of the transformation τ, for a suitable $\rho > 1$ we have

$$\tau(\tau_\varepsilon \boldsymbol{\gamma}_0) = \tau_\varepsilon \sigma^{-1} \tau_\varepsilon \sigma \tau_\varepsilon \boldsymbol{\gamma}_0 = \tau_\varepsilon \tau^{-1} \boldsymbol{\gamma}_0 = \rho \tau_\varepsilon \boldsymbol{\gamma}_0\ .$$

In view of the unicity of $\boldsymbol{\gamma}_1$ (since $\rho \neq 1$) we must have

$$\tau_\varepsilon \boldsymbol{\gamma}_0 = \boldsymbol{\gamma}_0 - 2(\boldsymbol{\gamma}_0, \boldsymbol{\varepsilon})\boldsymbol{\varepsilon} = \lambda \boldsymbol{\gamma}_1 \tag{8}$$

for some $\lambda > 0$ (cfr. the properties of the normalization in § 4.2). Scalar multiplication of (8) by $\boldsymbol{\gamma}_0$ yields $2(\boldsymbol{\gamma}_0, \boldsymbol{\varepsilon}) = \pm \sqrt{\lambda}$ so that choosing the positive sign (the orientation of $\boldsymbol{\varepsilon}$ is irrelevant) we obtain

$$\boldsymbol{\varepsilon} = \frac{1}{\sqrt{\lambda}} \boldsymbol{\gamma}_0 - \sqrt{\lambda}\, \boldsymbol{\gamma}_1\ . \tag{9}$$

Considering the spheres $\boldsymbol{\alpha}_i$ in the different order

$$\boldsymbol{\alpha}_p, \boldsymbol{\alpha}_{p+1}, \cdots, \boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_{p-1}\ ,$$

we obtain again the same link; the source and the sink of the corresponding Moebius transformation $\tau^* = \sigma \tau_\varepsilon \sigma^{-1} \tau_\varepsilon$ will then be the points $\boldsymbol{\gamma}'_0 = \sigma \boldsymbol{\gamma}_0$ and $\boldsymbol{\gamma}'_1 = \sigma \boldsymbol{\gamma}_1$. Therefore we must also have

[21] By τ_ε we mean the inversion generated by $\boldsymbol{\varepsilon}$.

$$\varepsilon = \frac{1}{\sqrt{\mu}}\gamma_0' - \sqrt{\mu}\,\gamma_1' \tag{10}$$

for some $\mu > 0$.

We set $\bar{\gamma} = (\varepsilon_1 - i\varepsilon_2)/2$, $\gamma = (\varepsilon_1 + i\varepsilon_2)/2$ where ε_1 and ε_2 are any real spheres containing γ_0 and γ_1 orthogonal to each other and to the sphere α_0; from (9) follows that

$$\tau_\varepsilon\bar{\gamma} = \bar{\gamma},\ \tau_\varepsilon\gamma = \gamma\ .$$

Now $(\sigma\bar{\gamma}, \sigma\gamma_0) = (\bar{\gamma}, \gamma_0) = 0$ and similarly $(\sigma\bar{\gamma}, \sigma\gamma_1) = (\sigma\gamma, \sigma\gamma_0) = (\sigma\gamma, \sigma\gamma_1) = 0$, therefore in view of (10) we deduce

$$\tau_\varepsilon\sigma\tau_\varepsilon\bar{\gamma} = \tau_\varepsilon\sigma\bar{\gamma} = \sigma\bar{\gamma}\ ,\ \ \tau_\varepsilon\sigma\tau_\varepsilon\gamma = \tau_\varepsilon\sigma\gamma = \sigma\gamma\ ,$$

and

$$\tau\bar{\gamma} = \sigma^{-1}\sigma\bar{\gamma} = \bar{\gamma}\ ,\ \ \tau\gamma = \sigma^{-1}\sigma\gamma = \gamma\ .$$ [22]

which is what we wanted to show.

Case 2 illustrates the intuitive fact that if a link L admits a plane of symmetry then whatever torsion L might inherit from one of its symmetric parts is taken away by the other. This property is not peculiar to natural links but it holds for all Riemann surfaces of genus one imbedded in E_3.

We shall give only a sketch of the proof for the general case.

If a surface admits a plane of symmetry then it admits an anticonformal (sense-reversing angle-preserving) mapping onto itself. This fact by itself is sufficient to exclude that the corresponding parallelgramm lattice could be a general one, it must have rectangular or rhomboidal generators.[23]

However, the case of rhomboidal generators can be excluded also. The anticonformal mapping generated by a plane of symmetry in E_3 will always leave invariant two distinct closed curves of the surface as loci of fixed points. On the other hand, if a rhomboidal lattice is a general one, the reflections which preserve the identification of points admit also two distinct invariant curves, but only one of them as a locus of fixed points.

5.5. In contrast with the results of the previous section, it is not difficult to construct natural links whose torsion does not vanish. The simplest models of such links can be obtained using five linearly independent spheres.

[22] A shorter but less illustrative proof could be derived from the fact that the equation $\tau_\varepsilon\gamma = \nu\bar{\gamma}$ together with (8) leads to an absurdity.

[23] We owe this observation to Professor H. Royden.

In fact, we can show that
If a link L is generated by five spheres $\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_4$ then its torsion vanishes if and only if the vectors $\boldsymbol{\alpha}_i$ are linearly dependent.

The torsion of L vanishes if and only if there exist vectors which are invariant under the product of inversions $\tau = \Delta_5\Delta_4 \cdots \Delta_1$ generated by the spheres $\boldsymbol{\delta}_i$. Now, the transform of a vector $\boldsymbol{\xi}$ by τ (after a repeated application of formula (7) of § 4.2) can be written in the form

$$\tau\boldsymbol{\xi} = \boldsymbol{\xi} - 2(\boldsymbol{\xi}, \boldsymbol{\delta}_1)\boldsymbol{\delta}_1 - 2(\Delta_1\boldsymbol{\xi}, \boldsymbol{\delta}_2)\boldsymbol{\delta}_2 - \cdots - 2(\Delta_4 \cdots \Delta_1\boldsymbol{\xi}, \boldsymbol{\delta}_5)\boldsymbol{\delta}_5$$

and the equation

$$(\boldsymbol{\xi}, \boldsymbol{\delta}_1)\boldsymbol{\delta}_1 + (\Delta_1\boldsymbol{\xi}, \boldsymbol{\delta}_2)\boldsymbol{\delta}_2 + \cdots + (\Delta_4 \cdots \Delta_1\boldsymbol{\xi}, \boldsymbol{\delta}_5)\boldsymbol{\delta}_5 = 0$$

can be satisfied when and only when the $\boldsymbol{\delta}_i$'s are dependent. On the other hand if we let α denote the matrix whose columns are the vectors $\boldsymbol{\alpha}_i$, δ denote the matrix whose columns are the vectors $\boldsymbol{\delta}_i$, and set $\mu_i = \sqrt{1 + (\boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_i)/2}$ [24] we have

$$\delta = \alpha \left|\begin{array}{lllll} 1/2\mu_1 & 0 & 0 & 0 & 1/2\mu_5 \\ 1/2\mu_1 & 1/2\mu_2 & 0 & 0 & 0 \\ 0 & 1/2\mu_2 & 1/2\mu_3 & 0 & 0 \\ 0 & 0 & 1/2\mu_3 & 1/2\mu_4 & 0 \\ 0 & 0 & 0 & 1/2\mu_4 & 1/2\mu_5 \end{array}\right|$$

and

$$\det \delta = \frac{\det \alpha}{2^4\mu_1\mu_2 \cdots \mu_5}. \tag{11}$$

Thus the $\boldsymbol{\delta}_i$'s are dependent or independent together with the $\boldsymbol{\alpha}_i$'s. This proves the assertion.

This result does not quite solve our original problem of constructing models whose representative point in $\mathfrak{M}$ is off the imaginary axis, at least as long as we do not know when the point $\theta/2\pi - i(\log \rho)/2\pi$ is contained in $\mathfrak{M}$. We shall get around this difficulty by showing that our models can be made sufficiently thin (cfr. the inequalities (2) of § 5.1). To this end we shall exhibit a family of links within which this deformation is possible.

Let $\boldsymbol{C}_0, \boldsymbol{C}_1, \cdots, \boldsymbol{C}_4$ be points of E_3 and P denote the closed polygonal line $\boldsymbol{C}_0\boldsymbol{C}_1\cdots\boldsymbol{C}_4\boldsymbol{C}_0$. Suppose that each segment $\overline{\boldsymbol{C}_i\boldsymbol{C}_{i+1}}$ ($i = 0, \cdots, 4$; $\boldsymbol{C}_5 = \boldsymbol{C}_0$) has length equal to twice that of the unit of measure, and set $2\varphi_i =$ angle $\boldsymbol{C}_{i-1}\widehat{\boldsymbol{C}_i\boldsymbol{C}_{i+1}}$. Let α_i be a sphere of radius R and center $\boldsymbol{C}_i$, i. e.,

[24] cfr. (10), (11) of § 4.2.

$$\alpha_i = \frac{1}{R}(1, C_i, C_i^2 - R^2)\,. \tag{12}$$

In order that the spheres $\boldsymbol{\alpha}_i$ fulfill the conditions (a), (b), (c), (d), (e) of § 2.3, so that they can be used to define a link, it is sufficient to require that for each $i = 1, \cdots, 5$ $\boldsymbol{\alpha}_{i-1}$ intersects $\boldsymbol{\alpha}_i$ and does not intersect $\boldsymbol{\alpha}_{i+1}$ (Set $\boldsymbol{\alpha}_6 = \boldsymbol{\alpha}_1$). We shall thus assume that P is such that

$$\varphi_i > \pi/6 + \sigma, \text{ or } \overline{C_{i-1}C_{i+1}} > 2(1+\varepsilon)\,. \tag{13}$$

for some $0 < \sigma < \pi/3, 0 < \varepsilon < 1$, and restrict R to satisfy

$$1 < R < 1 + \varepsilon\,. \tag{14}$$

Let $L(P, R)$ denote the link defined by such a choice of P and R. From (12) follows that

$$\det \alpha = \frac{1}{R^5} \det \begin{vmatrix} 1 & C_1 & C_1^2 \\ 1 & C_2 & C_2^2 \\ \cdots & \cdots & \cdots \\ 1 & C_5 & C_5^2 \end{vmatrix}, \tag{15}$$

therefore the torsion of $L(P, R)$ vanishes if and only if the vertices of P lie on the same sphere. Now, it is geometrically evident that if we keep P fixed and let R decrease to 1 the capacities of the anuli $\Lambda_{i-1}^+ \cap \Lambda_i^-$ will decrease to zero (see also formula (4) of § 5.2) and thus by the theorem of § 4.4 we can predict that the thinness of $L(P, R)$ will tend to infinity.

This proves the existence of links whose torsion does not vanish and whose representative point is in $\mathfrak{M}$.

More accurate results about the links $L(P, R)$ could be obtained by a direct calculation of the eigenvalues of the corresponding Moebius transformations. However, without going into tedious computations we can show that: *the portion of $\mathfrak{M}$ covered by the images of the links $L(P, R)$ contains a strip of constant width around the imaginary axis.*

It can be shown (see [6] pp. 26–28 and 154–155) that the characteristic polynomial of the Moebius transformation generated by a set of linearly independent spheres $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \cdots, \boldsymbol{\delta}_5$ is given by the expression

$$x(\lambda) = \det \begin{vmatrix} 1+\lambda & 2(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) & \cdots\, 2(\boldsymbol{\delta}_1, \boldsymbol{\delta}_5) \\ 2\lambda(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2) & 1+\lambda & \cdots\, 2(\boldsymbol{\delta}_2, \boldsymbol{\delta}_5) \\ \cdots & \cdots & \cdots \\ 2\lambda(\boldsymbol{\delta}_1, \boldsymbol{\delta}_5) & 2\lambda(\boldsymbol{\delta}_2, \boldsymbol{\delta}_5) & \cdots\, 1+\lambda \end{vmatrix}. \tag{16}$$

On the other hand, from the results of § 4.3 we have

$$x(\lambda) = (\lambda^2 - 2\cos\theta\lambda + 1)(\lambda^2 - 2\cos h\sigma\lambda + 1)(\lambda + 1)^{25} . \tag{17}$$

(we have set $\cos h\sigma = 1/2(\rho + 1/\rho)$). Evaluating (16) and (17) for $\lambda = 1$ and equating the results we obtain

$$\sin h^2\sigma/2 \sin^2\theta/2 = - \det \| (\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) \| . \tag{18}$$

If we recall the definition of the scalar product ((4) of § 4.2) we see that it is

$$\| (\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) \| = \delta^T \begin{pmatrix} 0 & 0 & 0 & 0 & -1/2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ -1/2 & 0 & 0 & 0 & 0 \end{pmatrix} \delta$$

this means that

$$\det \| (\boldsymbol{\delta}_i, \boldsymbol{\delta}_j) \| = - 1/4 \{\det \delta\}^2 .$$

Substituting in (18) we obtain

$$| \sin\theta/2 | = 1/2 \frac{| \det \delta |}{\sin h\sigma/2} . \tag{19}$$

We now observe that for a link $L(P, R)$ we have $(\boldsymbol{\alpha}_{i-1}, \boldsymbol{\alpha}_i) = 1 - 2/R^2$ and setting $r = \sqrt{R^2 - 1}$, (11) gives

$$\det \delta = \frac{R^5 \det \alpha}{2^4 r^5} ,$$

so that, using (15), (19) yields

$$| \sin\theta/2 | = \frac{\left| \det \begin{vmatrix} 1 & \boldsymbol{C}_0 & \boldsymbol{C}_0^2 \\ \cdots & \cdots & \cdots \\ 1 & \boldsymbol{C}_4 & \boldsymbol{C}_4^2 \end{vmatrix} \right|}{2^5 r^5 \sin h\sigma/2} . \tag{20}$$

We shall get upper and lower bounds for $\sin h\sigma/2$.

Let $\boldsymbol{\gamma}^0$ be the sink of the Moebius transformation corresponding to $L(P, R)$, and set

$$\boldsymbol{\gamma}^0 = (1, \boldsymbol{G}_0, \boldsymbol{G}_0^2),\ \Delta_1\boldsymbol{\gamma}^0 = \lambda_1\boldsymbol{\gamma}^1 = \lambda_1(1, \boldsymbol{G}_1, \boldsymbol{G}_1^2),\ \cdots,\ \Delta_5\boldsymbol{\gamma}^4 = \lambda_5\boldsymbol{\gamma}^5 = \lambda_5(1, \boldsymbol{G}_5, \boldsymbol{G}_5^2) = \rho\boldsymbol{\gamma}^0 .$$

Since $\boldsymbol{\delta}_i = 1/r(1, \boldsymbol{A}_i, \boldsymbol{A}_i^2)$ with $\boldsymbol{A}_i = (\boldsymbol{C}_{i-1} + \boldsymbol{C}_i)/2$, recalling formula (7)* of § 4.2 we obtain

$$\lambda_1 = \frac{\overline{G_0A_1}^2}{r^2} ,\ \lambda_2 = \frac{\overline{G_1A_2}^2}{r^2} ,\ \cdots,\ \lambda_5 = \frac{\overline{G_4A_5}^2}{r^2} ;$$

[25] $(\lambda + 1)$, since the number of spheres is odd.

this gives

$$\rho = \frac{\overline{G_0A_1}^2 \cdot \overline{G_1A_2}^2 \cdots \overline{G_4A_5}^2}{r^{10}} .$$

But each G_i is a point of the corresponding sphere α_i; thus we get

$$\rho \leqq \frac{(1+R)^{10}}{r^{10}} . \tag{21}$$

The theorem of § 4.4 gives a bound from below. Let C_i denote the capacity of the anulus $\Lambda_{i-1}^+ \cap \Lambda_i^-$, using (4) of § 5.2 and some geometrical considerations we obtain

$$C_i = \left\{ \log \frac{R \sin \varphi_i + \sqrt{1 - R^2 \cos^2 \varphi_i}}{R \sin \varphi_i - \sqrt{1 - R^2 \cos^2 \varphi_i}} \right\}^{-1} ,$$

thus

$$\rho \geqq \frac{(R \sin \varphi_1 + \sqrt{1 - R^2 \cos^2 \varphi_1})^2 \cdots (R \sin \varphi_5 + \sqrt{1 - R^2 \cos^2 \varphi_5})^2}{r^{10}} ;$$

since we keep $R < 2 \sin \varphi_i$ each of the factors in the numerator of the right hand side is greater than one therefore

$$\rho > \frac{1}{r^{10}} . \tag{22}$$

Finally (21) and (22) used in (20) yield (assuming $r \leqq 1$):

$$\frac{\left| \det \begin{vmatrix} 1 & C_0 & C_0^2 \\ \cdots & \cdots & \cdots \\ 1 & C_4 & C_4^2 \end{vmatrix} \right|}{2^4\{(R+1)^5 - (R-1)^5\}} \leqq |\sin \theta/2| \leqq \frac{\left| \det \begin{vmatrix} 1 & C_0 & C_0^2 \\ \cdots & \cdots & \cdots \\ 1 & C_4 & C_4^2 \end{vmatrix} \right|}{2^4(1 - r^{10})} . \tag{23}$$

These inequalities imply our assertion:

For each polygon $P \sim C_0C_1 \cdots C_4C_0$ let $D(P)$ denote the value of

$$\left| \det \begin{vmatrix} 1 & C_0 & C_4^2 \\ \cdots & \cdots & \cdots \\ 1 & C_4 & C_4^2 \end{vmatrix} \right| .$$

If P_0 is a regular pentagon of side 2 then the link $L(P_0, R)$ is certainly well defined when $1 < R\sqrt{2}$. Simple geometrical considerations together with formula (5) of § 5.3 show that the link $L(P_0, \sqrt{2})$has a thinness ρ_0 for which $\log \rho_0 < 2\pi$. Let then P vary among the polygonals which satisfy the following conditions.

(1) $D(P) \neq 0$.

(2) The link $L(P, \sqrt{2})$ is well defined.

(3) The point $\nu(P) = (\theta(P)/2\pi) - i(\log \rho(P))/2\pi$ corresponding to $L(P, \sqrt{2})$ is contained in the region $|Re\nu| < 1/2, |\nu| \leqq 1$.

Assume $1 < R < \sqrt{2}$ and set $\nu(P, R) = (\theta(P, R)/2\pi) - i(\log \rho(P, R))/2\pi$ where $\theta(P, R)$ and $\rho(P, R)$ represent the thinness and the torsion of $L(P, R)$.

For every fixed P, as R decreases from $\sqrt{2}$ to 1, the point $\nu(P, R)$ describes a curve $M(P)$ which starts from a point outside $\mathfrak{M}$, enters $\mathfrak{M}$ for a suitably small value of R and tends to infinity from within $\mathfrak{M}$ as $R \to 1$.

The first inequality in (23) shows that each curve $M(P)$ is bounded away from the imaginary axis. Then, if we let P approach P_0, because of the second inequality in (23), $M(P)$ will tend to the imaginary axis and sweep a neighborhood of the type asserted.

A family of polygons satisfying the conditions (1), (2), (3) can be obtained from the following model. Let (x, y, z) be a cartesian coordinate system in E_3. Let

$$\boldsymbol{C}_0 = (1/\sin \pi/5, 0, 0),\ \boldsymbol{C}_1 = (x(\psi), y(\psi), z(\psi)),\ \boldsymbol{C}_2 = (-\cot \pi/5, 1, 0)$$
$$\boldsymbol{C}_3 = (-\cot \pi/5, -1, 0),\ \boldsymbol{C}_4 = (x(\psi), -y(\psi), -z(\psi))$$

with

$$\begin{aligned} x(\psi) &= 1/2 \frac{1}{\sin 2\pi/5} + 2 \sin \pi/5 \sin \pi/10 \cos \psi \\ y(\psi) &= 1/2 + 2 \sin \pi/5 \cos \pi/10 \cos \psi \\ z(\psi) &= 2 \sin \pi/5 \sin \psi\ , \end{aligned}$$

and set $P(\psi) \sim \boldsymbol{C}_0\boldsymbol{C}_1(\psi)\boldsymbol{C}_2\boldsymbol{C}_3\boldsymbol{C}_4(\psi)\boldsymbol{C}_0$. The points $\boldsymbol{C}_i$ have been chosen so that $P(0)$ is the regular pentagon of side 2 which lies in the plane x, y, has its center at the origin and a vertex in the positive real axis. When ψ varies $\boldsymbol{C}_1(\psi)$, $\boldsymbol{C}_4(\psi)$ describe the circles H, K loci of points whose distances from $\boldsymbol{C}_0, \boldsymbol{C}_2$ and $\boldsymbol{C}_0, \boldsymbol{C}_3$ respectively are equal to 2. A short calculation gives (for $\psi < \pi/2$)

$$D(P) = 2^5 \sin \pi/5 \sin \pi/10 \sin \psi(1 - \cos \psi)\ . \tag{24}$$

It can be easily seen that the links $L(P(\psi), \sqrt{2})$ are well defined when defined when $|\psi| < \pi/4$ (the only critical distance in this range is $\overline{\boldsymbol{C}_1\boldsymbol{C}_4}$ and it is well above $2\sqrt{2}$).

Numerical estimates of the width of the strip covered are poor, since (21) is rather crude. Nevertheless using (23) and (24) with $R = 1.2$ and $\rho \geqq 11$ we obtain $|\theta| > 2$ degrees.

5.6. We shall conclude by showing that each natural M-surface can be deformed into a conformally equivalent C^∞ canal surface. Our construction is based on the following observation.

Let Γ be a Riemann surface, N a subregion of Γ and Λ the boundary of N. Let N^* be a Riemann surface with a boundary Λ^* and suppose there exists a conformal mapping Δ of N^* onto N which is defined and continuous up to Λ^*. Then we can make the set

$$\Gamma^* = (\Gamma - N) + N^*$$

into a Riemann surface conformally equivalent to Γ. The proof is immediate. We introduce local uniformizers in Γ^* so that the mapping $\varphi(P)$ of Γ^* onto Γ defined by

$$\varphi(P) = P \quad \text{for} \quad P \in \Gamma^* - N^*$$
$$\phi(P) = \Delta P \quad \text{for} \quad P \in N^*$$

is conformal.[26]

We shall illustrate the use of this observation in a simple case. Suppose Γ is imbedded in E_3. Assume that N is a simply connected piece of a surface of revolution whose boundary is a parallel. Let N^* be any other simply connected piece of surface of revolution which has the same boundary and the same axis as N. The existence of the mapping Δ in this case is trivial. The observation can thus be applied, and we can deduce that Γ and $\Gamma^* = (\Gamma - N) + N^*$ must inherit the same conformal structure from E_3.

If Γ is C^∞ across Λ and we want Γ^* to possess the same property, then we have to restrict N^* to osculate N along Λ to an infinite degree.

Our next application will be the smoothing of natural M-surfaces. Let L be a given natural link and suppose that we want to render smooth the edge formed by the spheres Γ_1 and Γ_2 of L. Let Λ be the circle of intersection of Γ_1 and Γ_2. For simplicity we shall assume that the whole space has been subjected to a Moebius transformation so that Γ_1 and Γ_2 have become spheres of equal radius, their centers being interior points. Let Λ^-, Λ^+ be the portions of Γ_1 and Γ_2 which are exterior to Γ_2 and Γ_1 respectively, π the plane of Λ; π_1 and π_2 two planes parallel to π at a small distance ε from π. Assume that π_1 and π_2 intersect Λ^- and Λ^+ respectively and set

$$\Lambda_1 = \pi_1 \cap \Lambda^- \ , \ \Lambda_2 = \pi_2 \cap \Lambda^+ \ .$$

Let a be the straight line which contains the centers of Γ_1 and Γ_2, ν a half plane bounded by a; k_1 and k_2 the semicircles $\Gamma_1 \cap \nu, \Gamma_2 \cap \nu$ respectively. Let

[26] In § 2.3 we have proceeded in a similar way.

$$A_1 = \nu \cap \Delta_1 \,, \; A = \nu \cap \Delta \,, \; A_2 = \nu \cap \Delta_2 \,.$$

Let N be the portion of L generated by the rotation of the arcs A_1k_1A and Ak_2A_2 around a.[27]

We shall choose k to be a curve of ν which joins A_1 to A_2 and fits with k_1 and k_2 at its end points in a C^∞ fashion. Let $N^*(k)$ be the surface of revolution generated by rotation of the arc A_1kA_2 around a. It is easy to see that when the non-Euclidean length of the arc A_1kA_2 in the half-plane ν is equal to the sum of the non-Euclidean lengths of the arcs A_1k_1A and Ak_2A_2 there exists a conformal mapping Δ of $N^*(k)$ onto N which leaves invariants the points of Δ_1 and Δ_2. And then, in view of our observation, $N^*(k)$ can be used to replace N in L. It remains to be shown that such a k can be found.

Let us first choose k to be the semicircle of ν which joins A_1 with A_2 and is orthogonal to a. Since k is then a geodesic, using the triangle inequality, we obtain

$$n.\mathscr{E}.l.A_1kA_2 < n.\mathscr{E}.l.A_1k_1A + n.\mathscr{E}.l.Ak_2A_2 \,. \tag{25}$$

Now, k can be deformed at its end points to fit with k_1, and k_2 as smoothly as we please, increasing its length as little as we wish. Thereafter, if necessary, we can increase the length of k to change (25) into an equality.

To complete our argument we must show that L can be rendered smooth without introducing self-intersections. However, it is clear that k can be chosen to be a simple curve contained in the circle of center A and radius the (Euclidean) length of the segment $\overline{AA_1}$, for any given ε.

REFERENCES

1. W. Blaschke, *Vorlesungen über Differentialgeometrie*, Berlin: Julius Springer, 1929, vol. III.
2. E. Cartan, *OEuvres completes*, Paris: Gauthier-Villars, 1955; vol. 2, part III, pp. 1701–1726.
3. J. L. Coolidge, *A Treatise on the Sphere and the Circle*, London: Oxford U. Press, 1916.
4. L. Ford, *Automorphic Functions*, New York: Chelsea Publ. Co., 1951.
5. A. Hurwitz and R. Courant, *Funktionentheorie*, Berlin: Julius Springer, 1925.
6. R. Lagrange, *Produits d'inversions et metrique conforme*, Cahiers Scientifiques, Fasc. XXIII, Paris: Gauthier-Villars, 1957.
7. G. Pólya and G. Szegö, *Isoperimetric inequalities in mathematical physics*, Annals of of Math. Studies, No. 27, Princeton U. Press, (1951), 51–54.
8. F. Schottky, *Ueber eine specielle Function, welche bei einer bestimmten linearen Transformation ihres Arguments unverändert bleibt*, Crelle's J., **101** (1887), 227–272.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

[27] This sentence is meaningful when ε is sufficiently small.

HOMOMORPHISMS OF CERTAIN ALGEBRAS OF MEASURES

IRVING GLICKSBERG

The problem of determining all isomorphisms between the L_1 algebras of a pair of locally compact groups G and H has been considered by J. G. Wendel [16, 17] and H. Helson [7] (in the abelian case); these authors showed in particular that all norm-decreasing isomorphisms arise essentially from isomorphisms between the groups (and are isometries). In the abelian case a device suggested by Helson leads to much more, and we shall determine all norm-decreasing *homomorphisms* of certain algebras of measures (similar to L_1) on G into the algebra of measures on H(cf. 2.1 below.).

Let $M(G)$ denote the Banach algebra of all finite, complex, regular Borel measures on G, with convolution as multiplication. $L_1(G)$ forms a subalgebra of $M(G)$, in fact an ideal. Because of this, knowledge of the norm-decreasing homomorphisms of L_1 algebras into algebras of measures on another group leads to the determination of all norm-decreasing isomorphisms between $M(G)$ and $M(H)$; indeed when G and H are abelian we shall show that for each norm-decreasing isomorphism of a (not necessarily closed) subalgebra of $M(G)$ which contains $L_1(G)$ with a similar subalgebra of $M(H)$ there is an isomorphism γ of G onto H and a fixed character $\hat{g}$ of G for which $T\mu$ is just the measure $\hat{g}\mu$ transported to H via γ (whence $TL_1(G) = L_1(H)$ and T is an isometry). This is exactly the abelian Helson-Wendel result extended to superalgebras of L_1; in the non-commutative situation we can only obtain the analogous result for compact groups.

Aside from familiar facts about harmonic analysis (as given in [10, 15]) our main tools will be the following results obtained in [6] for a compact group G:

(1) *each multiplicative subgroup of non-negative elements of the unit ball of $M(G)$, other than the trivial subgroup $\{0\}$, consists of translates of Haar measure of a fixed normal subgroup of G* [6, 2.4];

(2) *each non-zero idempotent in the unit ball of $M(G)$ is Haar measure of a subgroup multiplied by a multiplicative character of this subgroup* [6, 4.3].

It is a pleasure to record the author's indebtedness to K. de Leeuw for many stimulating comments and suggestions.

NOTATION. As usual $C_0(G)$ will denote the continuous complex

Received February 16, 1959. Presented to the A.M.S. January 22, 1959. Work supported in part by the United States Air Force under Contract AF 49 (638)-294

functions on G vanishing at infinity; $M(G)$ is of course $C_0(G)^*$. The space of all continuous bounded complex functions on G will be denoted by $C(G)$.

When G is abelian, $G^\wedge$ will denote its character group with generic element $\hat{g}$; the respective identities of G and $G^\wedge$ will be g_0 and $\hat{g}_0$. In general measures on G will be denote by the letter μ and those on H by ν with $\mu_g(\nu_h)$ the mass 1 at g (h). It will be convenient to use μ for the measure and also for the corresponding integral, writing $\mu(f) = \int f(g)\mu(dg)$ where integration is always over the entire group. For notational ease we shall take the Fourier-Stieltjes transform $\hat{\mu}$ of μ to be defined by $\hat{\mu}(\hat{g}) = \int (g, \hat{g})\mu(dg)(=\mu(\hat{g}))$; in particular for absolutely continuous measures, inversion will involve the familiar conjugation.

On occasion we shall need to multiply a measure μ by a function f: $f\mu$ will denote the measure we might define by $f\mu(dx) = f(x)\mu(dx)$. Finally it should perhaps be stated explicitly that the term "subalgebra" should only be taken in the algebraic sense, and all references to norms on subalgebras of $M(G)$ are to the norm of $M(G)$.

1. Preliminaries. If T is an isomorphism of $L_1(G)$ onto $L_1(H)$, and G and H are abelian then one has a dual homeomorphism τ of $H^\wedge$ onto $G^\wedge$ for which $(T\mu)^\wedge = \hat{\mu} \circ \tau$. This fact from the Gelfand theory formed the starting point of Helson's investigation [7], which proceeded to show τ had algebraic properties as well when T is norm-decreasing. Helson observed [7, §2] that τ could be extended to map almost periodic functions in a linear norm-decreasing fashion, but found no application for his observation, which will be fundamental for our abelian results.

Our first result yields the algebraic content of the norm-decreasing character of somewhat more general maps. Here and elsewhere $\|\ \|_\infty$ will denote the usual supremum norm for functions, and 0 the function identically zero.

THEOREM 1.1. *Let G and H be a pair of abelian topological groups, with $G^\wedge$ and $H^\wedge$ their (algebraic) groups of continuous characters. Let τ be any map of $H^\wedge$ into $G^\wedge \cup \{0\}$ with $\tau\hat{h}_0 = \hat{g}_0$. If*

$$\left\|\sum_{i=1}^{n} a_i \tau\hat{h}_i\right\|_\infty \leq \left\|\sum_{i=1}^{n} a_i \hat{h}_i\right\|_\infty \tag{1.11}$$

for any trigonometric polynomial $\sum_{i=1}^{n} a_i\hat{h}_i$ on H, then $\tau^{-1} G^\wedge$ is a subgroup of $H^\wedge$ and the restriction of τ to this subgroup an algebraic homomorphism.[1]

[1] Actually we could take $G^\wedge$ and $H^\wedge$ to be any groups of (multiplicative) characters on a pair of (not necessarily abelian) groups G and H. One need only replace G^* and H^* (below) by the duals of the (discrete) groups $G^\wedge$, $H^\wedge$ (into which G and H map onto dense subsets).

COROLLARY 1.2. *If* $\tau\colon H^\wedge \to G^\wedge \cup \{0\}$ *satisfies* (1.11) *and* $\tau\hat{h}_1 \in G^\wedge$ *then* $\sigma : \hat{h} \to \tau(\hat{h}_1)^{-1}\tau(\hat{h}\hat{h}_1)$ *is multiplicative on the subgroup* $\hat{h}_1^{-1}\tau^{-1}(G^\wedge)$ *of* $H^\wedge$, *and otherwise vanishes.*

COROLLARY 1.3. *If* $\tau : H^\wedge \to G^\wedge$ *satisfies* (1.11) *then* $\sigma : \hat{h} \to (\tau\hat{h}_0)^{-1}\tau\hat{h}$ *is a homomorphism of* $H^\wedge$ *into* $G^\wedge$. *Conversely if* σ *is a homomorphism* (1.11) *holds. Finally identical equality obtains in* (1.11) *iff* τ *is one-to-one as well.*

Proofs. In (1.11) we are of course demanding that the obvious linear extension of τ mapping trigonometric polynomials on H into those on G be norm-decreasing, and thus we have a norm-decreasing extension of this map taking $\mathfrak{A}(H)$, the almost periodic functions on H, into $\mathfrak{A}(G)$. Letting H^* and G^* be the almost periodic compactifications[2] of H and G we then have a norm-decreasing map T of $C(H^*)$ into $C(G^*)$ with $TH^{*\wedge} \subset G^{*\wedge} \cup \{0\}$. As a consequence the norm-decreasing adjoint map T^* of $C(G^*)^* = M(G)$ into $C(H^*)^* = M(H^*)$ is multiplicative, for $T^*(\mu_1 * \mu_2)(\hat{h}) = \mu_1 * \mu_2(T\hat{h}) = \mu_1(T\hat{h})\mu_2(T\hat{h}) = T^*\mu_1(\hat{h})T^*\mu_2(\hat{h})$ since $T\hat{h}$ is either 0 or a character. Hence $(T^*(\mu_1 * \mu_2))^\wedge = (T^*\mu_1)^\wedge(T^*\mu_2)^\wedge$ and from the one-to-one nature of $^\wedge$ we obtain $T^*(\mu_1 * \mu_2) = T^*\mu_1 * T^*\mu_2$.

Moreover from $T\hat{h}_0 = \tau\hat{h}_0 = \hat{g}_0$ we see that T^* preserves non-negativity; for $\mu \geq 0$ and $\|\mu\| = 1$ imply $1 = \mu(\hat{g}_0) = \mu(T\hat{h}_0) = T^*\mu(\hat{h}_0) \leq \|T^*\mu\| \leq \|\mu\| = 1$ so that $T^*\mu(1) = 1 = \|T^*\mu\|$, and therefore $T^*\mu \geq 0$.

Consequently T^* maps the multiplicative subgroup $\{\mu_g : g \in G^*\}$ of the unit ball of $M(G^*)$ into a subgroup of the unit ball of $M(H^*)$ which consists of non-negative measures. Thus by [**6**, 2.4] (cf. introduction (1)) the image consists of translates of Haar measure ν of some subgroup K of H^*, and we can write $T^*\mu_g = \nu^{\gamma(g)}$ where $\nu^{\gamma(g)}$ is the translate of ν to the coset $\gamma(g) \in H^*/K$. For $\hat{h}$ in $K^\perp$(=the subgroup of $H^{*\wedge} = H^\wedge$ of all characters identically 1 on K, hence constant on cosets mod K) we have $(\gamma(g), \hat{h}) = \nu^{\gamma(g)}(\hat{h}) = T^*\mu_g(\hat{h}) = \mu_g(T\hat{h}) = (g, \tau\hat{h})$ for all g in $G \subset G^*$. But as usual this implies τ is multiplicative on the subgroup $K^\perp$ of $H^\wedge$ since, for $\hat{h}_1, \hat{h}_2$ in $K^\perp$, $(g, \tau(\hat{h}_1\hat{h}_2)) = (\gamma(g), \hat{h}_1\hat{h}_2) = (\gamma(g), \hat{h}_1)(\gamma(g), \hat{h}_2) = (g, \tau\hat{h}_1)(g, \tau\hat{h}_2) = (g, \tau\hat{h}_1\tau\hat{h}_2)$ for all g in G. On the other hand for $\hat{h} \notin K^\perp$ we have $0 = \nu^{\gamma(g)}(\hat{h}) = T^*\mu_g(\hat{h}) = \mu_g(T\hat{h}) = (g, \tau\hat{h})$ for all g in G, and thus $\tau\hat{h} = 0$; consequently $\tau^{-1}G^\wedge$ is precisely the subgroup $K^\perp$ of $H^\wedge$, and our proof of Theorem 1.1 is complete.

We might remark that the converse of 1.1 can be obtained in somewhat the fashion of the corresponding assertion of 1.3 (below), and

---

[2] It will be convenient to view G as a dense subset of G^* and $\mathfrak{A}(G)$ as the restrictions to G of the elements of $C(G^*)$ [**10**, **15**]. Similarly we consider the elements of $G^\wedge$ as the restrictions to G of the elements of $G^{*\wedge}$.

equality obtains identically in (1.11) iff $\tau H^\wedge \subset G^\wedge$ and, as in 1.3, τ is one-to-one. Since we shall have no use for these facts proofs will be omitted.

The proof of Corollary 1.2 follows immediately from noting that

$$\left\|(\tau\hat{h}_1)^{-1}\sum_{i=1}^n a_i\tau\hat{h}_i\right\|_\infty = \left\|\sum_{i=1}^n a_i\tau\hat{h}_i\right\|_\infty$$

(so that (1.11) holds for σ) while $\sigma(\hat{h}_0) = \hat{g}_0$. Evidently σ is independent of the particular choice of $\hat{h}_1$.

The direct portion of Corollary 1.3 is a consequence of 1.2, taking $\hat{h}_1 = \hat{h}_0$. For the converse part we note that if σ is a homomorphism then interpreting it as a map of $H^{*\wedge}$ into $G^{*\wedge}$ we have a dual homomorphism γ of G^* into H^*, and

$$\left(\sum_{i=1}^n a_i\sigma\hat{h}_i\right)(g) = \left(\sum_{i=1}^n a_i\hat{h}_i\right)(\gamma(g)) .$$

Consequently (since we may consider G and H as dense subsets of G^* and H^*) we have

$$\begin{aligned}(1.12)\qquad \left\|\sum_{i=1}^n a_i\sigma\hat{h}_i\right\|_\infty &= \sup_{G^*}\left|\sum_{i=1}^n a_i\sigma\hat{h}_i(g)\right| = \sup_{G^*}\left|\sum_{i=1}^n a_i\hat{h}_i(\gamma(g))\right| \\ &= \sup_{\gamma(G^*)}\left|\sum_{i=1}^n a_i\hat{h}\,(h)\right| \le \sup_{H^*}\left|\sum_{i=1}^n a_i\hat{h}_i(h)\right| = \left\|\sum_{i=1}^n a_i\hat{h}_i\right\|_\infty ,\end{aligned}$$

and (1.11) holds. Clearly identical equality obtains if $\gamma(G^*) = H^*$. On the other hand since γ is continuous and G^* compact, $\gamma(G^*)$ is a compact subgroup of H^*, and if $\gamma(G^*) \neq H^*$ some non-zero f in $C(H^*)$ vanishes on $\gamma(G^*)$; since f can be approximated uniformly by trigonometric polynomials equality in (1.12) cannot always obtain. Thus identical equality is equivalent to $\gamma(G^*) = H^*$, or dually, to the one-to-oneness of σ, hence of τ.

We shall return to some reformulations and analogues of these results in §6.

2. Homomorphisms. In order to utilize the device suggested by Helson we need not restrict our attention to Banach algebras. We need only insist that our subalgebra A of $M(G)$ have $G^\wedge \cup \{0\}$ as its space of multiplicative functionals and be large enough to determine the norm of each trigonometric polynomial on G. Unless something to the contrary is stated G and H will represent locally compact abelian groups throughout this section.

It will be convenient to extend the definition of the Fourier-Stieltjes transform $\hat{\mu}$ of μ in $M(G)$ by setting $\hat{\mu}(0) = \mu(0) = 0$, and regard

$G^\wedge \cup \{0\}$ as the one point compactification of $G^\wedge$. Consider the following conditions on a subalgebra A of $M(G)$:

(2.01) *For each trigonometric polynomial* $\sum_{i=1}^n a_i \hat{g}_i$ *on* G

$$\left\|\sum_{i=1}^n a_i \hat{g}_i\right\|_\infty = \sup\left\{\left|\mu\left(\sum_{i=1}^n a_i \hat{g}_i\right)\right| : \mu \in A, \|\mu\| \leq 1\right\};$$

(2.02) *The set of maps* $\mu \to \hat{\mu}(\hat{g}), \hat{g} \in G^\wedge \cup \{0\}$, *corresponds in a one-to-one fashion to the set of all multiplicative linear functionals on* A, *and* $A^\wedge \subset C(G^\wedge \cup \{0\})$.

When both conditions hold $A^\wedge$ *contains*[3] *sufficiently many functions to determine the topology of* $G^\wedge \cup \{0\}$; for (2.01) implies $A^\wedge$ separates any pair of elements of the compact space $G^\wedge \cup \{0\}$, and thus each $\hat{g}_1$ in $G^\wedge \cup \{0\}$ has a base of neighborhoods of the form $\{\hat{g} : |\hat{\mu}_i(\hat{g}) - \hat{\mu}_i(\hat{g}_1)| < \varepsilon, i = 1, 2, \cdots, n\}$, where $\mu_i \in A$. $A = L_1(G)$ clearly satisfies these condition, and will of course be the most important example.

THEOREM 2.1 *Let A satisfy* (2.01) *and* (2.02). *Then if T is a non-zero norm-decreasing homomorphism of A into $M(H)$ there is a compact subgroup H_0 of H, a continuous* (*not necessarily open*) *homomorphism γ of G into H/H_0, and characters $\hat{g}$ of G and $\hat{h}$ of H for which*

$$T\mu(f) = \mu(\hat{g}[S(\hat{h}f) \circ \gamma]), \qquad f \in C_0(H), \tag{2.11}$$

where S denotes the map of $C_0(H)$ onto $C_0(H/H_0)$ defined by $Sf(hH_0) = \int_{H_0} f(hh')\nu(dh')$ (where ν is Haar measure on H_0); alternatively

$$T\mu = \hat{h}S^*\Gamma\hat{g}\mu \tag{2.12}$$

where Γ is the homomorphism of $M(G)$ into $M(H/H_0)$ defined by setting $\Gamma\mu(f) = \mu(f \circ \gamma), f \in C_0(H/H_0)$, and S^* is the adjoint of S mapping $M(H/H_0)$ into $M(H)$. Conversely each such quadruple $H_0, \gamma, \hat{g}, \hat{h}$ defines a non-zero norm-decreasing T via (2.11) *or* (2.12).

Proof. For each $\hat{h}$ in $H^\wedge, \mu \to (T\mu)^\wedge(\hat{h})$ defines a multiplicative functional on A, and thus we obtain a unique $\tau\hat{h}$ in $G^\wedge \cup \{0\}$ for which $(T\mu)^\wedge(\hat{h}) = \hat{\mu}(\tau\hat{h})$. Since the elements of $A^\wedge$ suffice to define the topology of $G^\wedge \cup \{0\}$ and the functions $(T\mu)^\wedge$ are continuous on $H^\wedge$ one clearly has τ: $H^\wedge \to G^\wedge \cup \{0\}$ continuous.

On the other hand τ satisfies (1.11) as a consequence of (2.01):

$$\left\|\sum_{i=1}^n a_i \tau\hat{h}_i\right\|_\infty = \sup_{\|\mu\| \leq 1}\left|\mu\left(\sum_{i=1}^n a_i \tau\hat{h}_i\right)\right| \tag{2.13}$$

[3] Of course this holds when $A^\wedge$ is only a separating subalgebra of $C(G^\wedge \cup \{0\})$; but then we can only assert that $G^\wedge$ forms a subspace of the space of multiplicative functionals on A (taken in the w^* topology).

$$= \sup_{\|\mu\| \le 1} \left| T\mu\left(\sum_{i=1}^{n} a_i \hat{h}_i\right)\right|$$

$$\le \sup_{\|T\mu\| \le 1} \left| T\mu\left(\sum_{i=1}^{n} a_i \hat{h}_i\right)\right| \le \left\|\sum_{i=1}^{n} a_i \hat{h}_i\right\|_\infty .$$

Thus in order to apply Corollary 1.2 we need only verify that $\tau\hat{h}_1 \in G^\wedge$ for some $\hat{h}_1$ in $H^\wedge$; but such an $\hat{h}_1$ exists since otherwise $\tau H^\wedge = 0$, $(TA)^\wedge(H^\wedge) = 0$ and thus $TA = 0$ by the one-to-oneness of the Fourier-Stieltjes transformation. Consequently $\sigma : \hat{h} \to (\tau\hat{h}_1)^{-1}\tau(\hat{h}\hat{h}_1)$ is multiplicative on the subgroup $K = \sigma^{-1}(G^\wedge) = \hat{h}_1^{-1}\tau^{-1}(G^\wedge)$ of $H^\wedge$, and of course vanishes elsewhere. As we have seen τ, and thus σ, is continuous on $H^\wedge$ so that $\sigma^{-1}\{0\}$ is closed and $K = \sigma^{-1}(G^\wedge)$ is open. Therefore K is an open and closed subgroup of $H^\wedge$, whence $H^\wedge/K$ is discrete, and the dual $H_0 = K^\perp$ of $H^\wedge/K$ is a compact subgroup of H.

Dual to the continuous homomorphism $\sigma|K : K \to G^\wedge$ we have a continuous homomorphism γ of G into $K^\wedge = H/K^\perp = H/H_0$, and thus for $\hat{h}$ in K and g in G, $(g, \sigma\hat{h}) = (\gamma(g), \hat{h}) = \nu^{\gamma(g)}(\hat{h})$, where $\nu^{\gamma(g)}$ is again the translate to the coset $\gamma(g)$ of Haar measure ν on H_0. Moreover the formula

$$(g, \sigma\hat{h}) = \nu^{\gamma(g)}(\hat{h}) \tag{2.25}$$

clearly also holds when $\hat{h} \notin K = H_0^\perp$, since both sides are then zero. Combining (2.25) with $\sigma(\hat{h})\tau(\hat{h}_1) = \tau(\hat{h}\hat{h}_1)$, or $\sigma(\hat{h}\hat{h}_1^{-1})\tau(\hat{h}_1) = \tau(\hat{h})$, we make the following computation, with $F \in L_1(H^\wedge)$:

$$\begin{aligned}
T\mu(\bar{\hat{F}}) &= \int \bar{\hat{F}}(h) T\mu(dh) = \int \overline{F(\hat{h})}(T\mu)^\wedge(\hat{h})d\hat{h} \\
&= \int \overline{F(\hat{h})}\hat{\mu}(\tau\hat{h})d\hat{h} = \iint \overline{F(\hat{h})}(g, \tau\hat{h})\mu(dg)d\hat{h} \\
&= \iint \overline{F(\hat{h})}(g, \sigma(\hat{h}\hat{h}_1^{-1}))(g, \tau\hat{h}_1)d\hat{h}\mu(dg) \\
&= \iint \overline{F(\hat{h})}\nu^{\gamma(g)}(\hat{h}\hat{h}_1^{-1})(g, \tau\hat{h}_1)d\hat{h}\mu(dg) \\
&= \iiint \overline{F(\hat{h})}(h, \hat{h}\hat{h}_1^{-1})(g, \tau\hat{h}_1)\nu^{\gamma(g)}(dh)d\hat{h}\mu(dg) \\
&= \iint \left(\int \overline{F(\hat{h})}(h, \hat{h})d\hat{h}\right)(h, \hat{h}_1^{-1})\nu^{\gamma(g)}(dh)(g, \tau\hat{h}_1)\mu(dg) \\
&= \iint \bar{\hat{F}}(h)(h, \hat{h}_1^{-1})\nu^{\gamma(g)}(dh)(g, \tau\hat{h}_1)\mu(dg) \\
&= \int S(\hat{h}_1^{-1}\bar{\hat{F}})(\gamma(g))(g, \tau\hat{h}_1)\mu(dg) ,
\end{aligned}$$

or, setting $\hat{h} = \hat{h}_1^{-1}$ and $\hat{g} = \tau\hat{h}_1$

$$T\mu(\bar{F}) = \mu(\hat{g} \cdot [S(\hat{h}\bar{F}) \circ \gamma]) ; \tag{2.26}$$

since $L_1(H^\wedge)^\wedge$ is dense in $C_0(H)$ and both sides of (2.11) are continuous in f, (2.11) follows. The alternative form (2.12) follows when we make the obvious notational transfer.

Conversely given $H_0, \gamma, \hat{g}$ and $\hat{h}$, and thus S^* and Γ, the right side of (2.12) clearly defines a norm-decreasing homomorphism of $M(G) \to M(H)$, as the composition of four norm-decreasing homomorphisms. To see that $T \neq 0$ we need only verify that (2.11) remain valid for $f = \hat{h}' \in H^\wedge$; for then

$$T\mu(\hat{h}') = \mu(\hat{g}[S(\hat{h}\hat{h}') \circ \gamma])$$

while $S(\hat{h}\hat{h}') \in (H/H_0)^\wedge$ if $\hat{h}\hat{h}' \in H_0^\perp$, and then $S(\hat{h}\hat{h}') \circ \gamma = \hat{g}' \in G^\wedge$. Consequently $TA(\hat{h}') = A(\hat{g}\hat{g}') = A^\wedge(\hat{g}\hat{g}') \neq \{0\}$ for an appropriate $\hat{h}'$, by (2.02).

But that (2.11) remains valid for $f = \hat{h}' \in H^\wedge$ follows from the same sort of computation as the preceding; with $F \in L_1(H^\wedge)$ one obtains

$$\begin{aligned}
\int \overline{F(\hat{h}')}(T\mu)(\hat{h}')d\hat{h}' = T\mu(\bar{F}) &= \int (g, \hat{g})S(\hat{h}\bar{F})(\gamma(g))\mu(dg) \\
&= \iint (g, \hat{g})(h, \hat{h})\overline{\hat{F}(h)}\nu^{\gamma(g)}(dh)\mu(dg) \\
&= \int \overline{F(\hat{h}')}\Big(\iint (g, \hat{g})(h, \hat{h})(h, \hat{h}')\nu^{\gamma(g)}(dh)\mu(dg)\Big)d\hat{h}' \\
&= \int \overline{F(\hat{h}')}\Big(\int (g, \hat{g})[S(\hat{h}\hat{h}')(\gamma(g))]\mu(dg)\Big)d\hat{h}' \\
&= \int \overline{F(\hat{h}')}\mu(\hat{g}[S(\hat{h}\hat{h}') \circ \gamma])d\hat{h}
\end{aligned}$$

whence $(T\mu)(\hat{h}') = \mu(\hat{g}[S(\hat{h}\hat{h}') \circ \gamma])$ for almost all $\hat{h}'$. But the second expression vanishes for $\hat{h}'$ in the open complement of $\hat{h}^{-1}H_0^\perp$ so that (by continuity) the first also vanishes there. On the other hand for $\hat{h}'$ in $\hat{h}^{-1}H_0^\perp$ (also open) we have $\mu(\hat{g}[S(\hat{h}\hat{h}') \circ \gamma])$ continuous as a function of $\hat{h}'$ since $S(\hat{h}\hat{h}') \circ \gamma = \sigma(\hat{h}\hat{h}')$ where σ is the continuous homomorphism of $H_0^\perp = (H/H_0)^\wedge \to G^\wedge$ dual to γ. Consequently both expressions are continuous functions of $\hat{h}'$ on $\hat{h}^{-1}H_0^\perp$ as well, and thus coincide on this open set.

2.2 REMARK. When the subgroup H_0 is trivial (i.e. $= \{h_0\}$) one may write 2.12 in the more concise form $T\mu = \Gamma\hat{g}\mu$; for clearly we have $T\mu = \hat{h}\Gamma\hat{g}\mu$ so that $T\mu(f) = \mu(\hat{g}[(\hat{h}f) \circ \gamma]) = \mu(\hat{g}(\hat{h} \circ \gamma)(f \circ \gamma))$ and we may replace $\hat{g}$ by $\hat{g}(\hat{h} \circ \gamma) \in G^{\wedge}$. This situation will of course occur if each $\hat{h}$ in $H^{\wedge}$ produces a non-zero functional on TA, i.e. when $\tau H^{\wedge} \subset G^{\wedge}$; for then $H_0^{\perp} = K = H^{\wedge}$.

2.3 REMARK. If A is an ideal of a larger subalgebra A_0 of $M(G)$ and A satisfies (2.01) and (2.02) there is the possibility of applying Theorem 2.1 to certain norm-decreasing homomorphisms T on A_0. For provided $TA \neq \{0\}$, we may apply the result to the pair A and $T|A$ to obtain $T|A = T_1|A$ where T_1 represents the homomorphism (given by the right side of (2.12)) of all of $M(G)$ into $M(H)$; consequently (since A is an ideal in A_0) for $\mu \in A$, $\mu' \in A_0$,

$$T\mu' * T\mu = T(\mu' * \mu) = T_1(\mu' * \mu) = T_1\mu' * T_1\mu = T_1\mu' * T\mu$$

and $(T\mu' - T_1\mu') * T\mu = 0$. Hence $T\mu' - T_1\mu'$ annihilates TA, and we need only know that TA has no non-zero annihilators in $M(H)$ (not TA_0) to conclude that $T\mu' = T_1\mu'$ for all μ' in A_0. As a particular case

COROLLARY 2.31. *Let A satisfy* (2.01) *and* (2.02) *and let A_0 be a larger subalgebra of $M(G)$ in which A forms an ideal. If T is a norm-decreasing isomorphism of A_0 onto $M(H)$, then T is determined as in Theorem* 2.1, *indeed as in* 2.2 *since $H_0 = \{h_0\}$.*

Since $\mu * A = 0$ implies $\hat{\mu} A^{\wedge} = 0$ while $A^{\wedge}(\hat{g}) \neq \{0\}$ for each $\hat{g}$ in $\hat{G}$ by (2.02), A has no non-zero annihilators in A_0. Thus since T is an isomorphism, TA has no non-zero annihilators in $TA_0 = M(H)$, and $T\mu = \hat{h} S^* \Gamma \hat{g} \mu$. But if ν denotes Haar measure of H_0 one clearly has $(\hat{h}\nu) * T\mu = T\mu$ so that we must have $\hat{h}\nu$ the identity of $M(H)$, hence $H_0 = \{h_0\}$.

2.4. The following example shows how completely wrong Theorem 2.1 is for arbitrary large subalgebras of $M(G)$ in general; it was suggested to the author by K. de Leeuw. Let G be any non-discrete locally compact abelian group and, for μ in $M(G)$, let $\mu = \mu^p + \mu^c$ be the Lebesgue decomposition of μ into discrete and continuous parts, i.e., μ^p is a countable linear combination of point masses (converging in norm) and μ^c vanishes on all one point sets. Since the continuous measures form an ideal and $\mu_1^p * \mu_2^p$ is still discrete, $\mu \to \mu^p$ is a norm-decreasing homomorphism of $M(G) \to M(G)$, or indeed of $M(G)$ onto $M(G^d)(G^d = G$ in the discrete topology); clearly the map is not induced by any continuous $\gamma : G \to G^d$.

2.5. The restriction that T be norm-decreasing in Theorem 2.1 can be replaced by apparently weaker conditions in certain cases. The following result has a much simpler proof when $A = L_1(G)$.

THEOREM 2.5. *Let A be a subalgebra of $M(G)$ satisfying* (2.02) *which is spanned by its non-negative elements and has sufficiently many of these to determine the non-negative almost periodic functions, i.e.,*

(2.51) $f \in \mathfrak{A}(G)$ and $\mu(f) \geq 0$ for all $\mu \geq 0$ in A imply $f \geq 0$. *If T is any non-zero homomorphism of A into $M(H)$ which preserves order ($\mu \geq 0 \Rightarrow T\mu \geq 0$) then T is norm-decreasing. If A also satisfies* (2.01) *then $T\mu = S^*\Gamma\mu$, $\mu \in A$.*

Proof. As in Theorem 2.1 we obtain $\tau : H^\wedge \to G^\wedge \cup \{0\}$ with $T\mu(\hat{h}) = \mu(\tau\hat{h})$. The functional $\mu \to T\mu(\hat{h}_0)$ cannot be zero for then $0 = T\mu(1) = \|T\mu\|$ for $\mu \geq 0$, whence $TA = 0$ since the non-negative elements span A. For $\mu \geq 0$ in A, $\mu(\tau\hat{h}_0) = T\mu(\hat{h}_0) \geq 0$ so that $\tau\hat{h}_0 \geq 0$ by (2.51), and thus $\tau\hat{h}_0 = \hat{g}_0$. Consequently for $\mu \geq 0$, $\|\mu\| = \mu(\hat{g}_0) = \mu(\tau\hat{h}_0) = T\mu(\hat{h}_0) = \|T\mu\|$[4].

Let τ_0 denote the linear extension of τ mapping trigonometric polynomials. If p is a non-negative trigonometric polynomial on H and $\mu \geq 0$ is in A then $\mu(\tau_0 p) = T\mu(p) \geq 0$, so that $\tau_0 p \geq 0$ by (2.51). Thus τ_0 preserves the order of real valued trigonometric polynomials, and since $\tau\hat{h}_0 = \hat{g}_0$, $-1 \leqq p \leqq 1$ implies $-1 \leqq \tau_0 p \leqq 1$. But for any trigonometric polynomial $p = \sum_{i=1}^n a_i\hat{h}_i$, if $p^* = \sum_{i=1}^n \bar{a}_i\hat{h}_i^{-1}$ then $(p + p^*)/2$ and $(p - p^*)/2i$ are real valued, with values bounded by $-\|p\|_\infty$, $\|p\|_\infty$. Hence $\|\tau_0(p + p^*)/2\|_\infty \leqq \|p\|_\infty$, $\|\tau_0(p - p^*)/2\|_\infty = \|\tau_0(p - p^*)/2i\|_\infty \leqq \|p\|_\infty$, and therefore $\|\tau_0 p\| \leqq 2\|p\|_\infty$.

Consequently τ_0 extends to a bounded map of $\mathfrak{A}(H)$ into $\mathfrak{A}(G)$, which we may view as a map of $C(H^*)$ into $C(G^*)$; calling the extension τ_0 we have $T\mu(f) = \mu(\tau_0 f)$, $\mu \in A, f \in C(H^*)$, since this held for trigonometric polynomials. Moreover this identity implies τ_0 (as extended) preserves order by (2.51), so the adjoint $\tau_0^* : M(G^*) \to M(H^*)$ must also preserve order. As before we conclude from $\tau_0\hat{h}_0 = \hat{g}_0$ that $\|\tau_0^*\mu\| = \|\mu\|$ for $\mu \geq 0$ in $M(G^*)$. Therefore τ_0^* maps the point masses on G^* into the unit ball of $M(H^*)$, and thus their w^* closed convex circled hull into the same set. Since the hull coincides with the unit ball of $M(G^*)$, $\|\tau_0^*\| = \|\tau_0\| \leqq 1$, and, for μ in A,

$$\sup_{\|f\|_\infty \leqq 1} |T\mu(f)| = \sup_{\|f\|_\infty \leqq 1} |\mu(\tau_0 f)| \leqq \sup_{\|\tau_0 f\|_\infty \leqq 1} |\mu(\tau_0 f)| \leqq \|\mu\|$$

where f varies in $\mathfrak{A}(H)$. But the norm of $T\mu$ as a functional on $\mathfrak{A}(H)$ coincides with its norm as a measure ([4], [6, §5]), whence $\|T\mu\| \leqq \|\mu\|$, and T is norm-decreasing.

For the final statement in 2.5 we need only note that since $\tau\hat{h}_0 = \hat{g}_0$,

[4] At this point the proof for $A = L_1$ is essentially complete; for T is clearly norm-decreasing on simple functions (rather, on the corresponding measures) and these are dense.

and since our present τ coincides with that obtained in the proof of Theorem 2.1, we may take $\hat{h}_1 = \hat{h}_0$ in deriving (2.26), so that $\hat{h} = \hat{h}_0$, $\hat{g} = \hat{g}_0$ in (2.12), completing our proof.

It should perhaps be noted that portions of the above proof can be used to obtain an analogue of Theorem 1.1 in which (1.11) is replaced by "$\sum_{i=1}^n a_i \tau\hat{h}_i \geq 0$ if $\sum_{i=1}^n a_i\hat{h}_i \geq 0$"; for clearly our argument shows this condition implies (1.11).

If the group H has a connected dual we can replace "norm-decreasing" in Theorem 2.1 by "bounded".

THEOREM 2.6. *Let A be a subalgebra of $M(G)$ satisfying* (2.01) *and* (2.02), *and suppose $H^\wedge$ is connected. If T is any bounded non-zero homomorphism of A into $M(H)$, then T is norm-decreasing; consequently there is a homomorphism $\gamma : G \to H$ and a $\hat{g}$ in $G^\wedge$ for which $T\mu = \Gamma\hat{g}\mu$, $\mu \in A$. In particular if A is a closed subalgebra, all non-zero homomorphisms of A into $M(H)$ arise in this fashion.*[5]

Proof. As in the proof of Theorem 2.1 we obtain a continuous map $\tau : H^\wedge \to G^\wedge \cup \{0\}$ with $\tau^{-1}G^\wedge \neq \phi$; further, the linear extension of τ mapping trigonometric polynomials is bounded by a computation analogous to (2.13), and we may view this as extending to a bounded map $\tau_0 : C(H^*) \to C(G^*)$. Again $\tau_0^* : M(G^*) \to M(H^*)$ is multiplicative (as in 1.1), for $\tau_0^*\mu(\hat{h}) = \mu(\tau_0\hat{h}) = \mu(\tau\hat{h})$, or $(\tau_0^*\mu)^\wedge = \hat{\mu} \circ \tau$, $\hat{\mu} \in M(G^*)$.

Now (for any locally compact abelian G) if we define $\tilde{\mu} \in M(G^*)$ corresponding to $\mu \in M(G)$ by setting $\tilde{\mu}(f) = \int_G f(g)\mu(dg), f \in C(G^*)$ (so that $\tilde{\mu}$ represents the restriction of the integral corresponding to μ to almost periodic functions) then $\mu \to \tilde{\mu}$ is an isometric isomorphism of $M(G)$ into $M(G^*)$ ([4], or [6, § 5]), and, as functions on the *set* $G^\wedge$, $\hat{\tilde{\mu}} = \hat{\mu}$. Moreover as a consequence of a theorem of Bochner-Schoenberg-Eberlein [4], $M(G)^\sim$ consists of just those μ in $M(G^*)$ with $\hat{\mu}$ continuous on the *space* $G^\wedge$. Thus, for μ in $M(G)$, since τ is continuous and $(\tau_0^*\tilde{\mu})^\wedge = \hat{\tilde{\mu}} \circ \tau = \hat{\mu} \circ \tau$, we have $(\tau_0^*\tilde{\mu})^\wedge$ the transform of some measure $\sigma\mu$ in $M(H)$, i.e., $\tau_0^*\tilde{\mu} = (\sigma\mu)^\sim$. Clearly σ is a multiplicative map of $M(G)$ into $M(H)$. Since $\tau^{-1}G^\wedge \neq \phi$ for any $\hat{h}$ therein we have $|\sigma\mu_g(\hat{h})| = |\tau_0^*\tilde{\mu}_g(\hat{h})| = |\tilde{\mu}_g(\tau\hat{h})| = |\mu_g(\tau\hat{h})| = 1$ for all g in G, whence $\sigma\mu_g \neq 0$. Consequently if E denotes the set of all point masses on $G, \sigma E$ forms a bounded non-zero subgroup of $M(H)$ so that ($H^\wedge$ being connected) by a theorem of Beurling and Helson [3, § 5] σE consists of unimodular multiples of point masses on H. Thus E maps into the unit ball of $M(H)$ under σ,

[5] Such homomorphisms being automatically bounded since A is a Banach algebra and $M(H)$ is semisimple.

or equivalently $E^{\sim}$ maps into the unit ball of $M(H^*)$ under τ_0^*. But $E^{\sim}$ is w^* dense in the set of point masses on G^*, and thus τ_0^* carries all point masses on G^* into the unit ball of $M(H^*)$. As in the proof of Theorem 2.5 this implies $\|T\mu\| \leq \|\mu\|$, $\mu \in A$. The final assertions of 2.6 now follow from 2.1 and 2.2, since the connectedness of $H^{\wedge}$ precludes the existence of any non-trivial compact subgroup H_0 of H. A consequence of our proof is

COROLLARY 2.61. *Let $H^{\wedge}$ be connected, and let $\tau : H^{\wedge} \to G^{\wedge} \cup \{0\}$ be any non-zero continuous map for which*

$$\left\|\sum_{i=1}^{n} a_i \tau \hat{h}_i\right\|_{\infty} \leq M \left\|\sum_{i=1}^{n} a_i \hat{h}_i\right\|_{\infty}$$

for all trigonometric polynomials $\sum_{i=1}^{n} a_i \hat{h}_i$ on H. Then M can be replaced by 1, $\tau H^{\wedge} \subset G^{\wedge}$, and $\hat{h} \to (\tau\hat{h}_0)^{-1}\tau\hat{h}$ is a homomorphism.[6]

For the map again extends to a bounded map τ_0 of $C(H^*)$ into $C(G^*)$ with $\|\tau_0\| = \|\tau_0^*\| \leq 1$ so that M can be replaced by 1. Since a translate of $\tau^{-1}G^{\wedge}$ provides us with an open subgroup of $H^{\wedge}$ by 1.1, $\tau^{-1}G^{\wedge} = H^{\wedge}$ and we need only apply 1.3.

2.7. A result of Leibenson [9], improved by Kahane [8], can be stated as follows: the only maps τ of the circle group T^1 into itself for which $f \circ \tau$ has an absolutely convergent series whenever f does are of the form $\tau(t) = t_1 \cdot t^n$, where $t_1 \in T^1$ and n is an integer. The following corollary of 2.6 yields a stronger assertion as a special case ($G^{\wedge} = H^{\wedge} = T^1$, $A = L_1(G)$); the result is of course essentially a dual formulation of 2.6.

COROLLARY 2.71. *Let A be a closed subalgebra of $M(G)$ satisfying* (2.01) *and* (2.02), *and let $H^{\wedge}$ be connected.*[7] *Then any map τ of $H^{\wedge}$ into $G^{\wedge}$ for which*

$$f \in A^{\wedge} \text{ implies } f \circ \tau \in M(H)^{\wedge}$$

must be of the form

$$\tau(\hat{h}) = \hat{g} \cdot \sigma(\hat{h})$$

where $\hat{g} \in G^{\wedge}$ and σ is a continuous homomorphism of $H^{\wedge}$ into $G^{\wedge}$.

---

[6] Note that continuity cannot be dropped from our hypothesis: for a map of $R^{\wedge}$ which merely interchanges two elements produces a bounded map of trigonometric polynomials on R.

[7] In this and our subsequent results involving a connected dual (viz: parts of 3.5, 4.2, 4.3, and 5.1) (2.01) can always be replaced by the requirement that $\|p\|_\infty \leq K \sup\{|\mu(p)| : \mu \in A, \|\mu\| \leqq 1\}$ for all trigonometric polynomials p.

Proof. Let $T\mu$ be that element of $M(H)$ for which $(T\mu)^\wedge = \hat{\mu} \circ \tau$, $\mu \in A$. Clearly T is an algebraic homomorphism of A into $M(H)$, which must be bounded since A is a Banach algebra and $M(H)$ is semisimple. Moreover T is non-zero, since otherwise $A^\wedge(\tau H^\wedge) = 0$, contradicting (2.02). Thus 2.6 applies to yield a continuous homomorphism $\gamma: G \to H$ and a $\hat{g}$ in $G^\wedge$ with $T\mu = \Gamma\hat{g}\mu$, $\mu \in A$, whence as before

$$\hat{\mu}(\tau\hat{h}) = (T\mu)^\wedge(\hat{h}) = T\mu(\hat{h}) = \Gamma\hat{g}\mu(\hat{h}) = \mu(\hat{g}(\hat{h} \circ \gamma)) = \hat{\mu}(\hat{g}(\hat{h} \circ \gamma))$$

for all μ in $\hat{A}, h$ in $H^\wedge$. Consequently $\tau(\hat{h}) = \hat{g}(\hat{h} \circ \gamma) = \hat{g}\sigma(\hat{h})$ where $\sigma: H^\wedge \to G^\wedge$ is the continuous homomorphism dual to γ.

It should be noted that we cannot obtain the type of boundedness required in 2.6 by simply assuming A is a Banach algebra under *some* norm.

An analogous result, in which connectedness is replaced by more stringent requirements on τ, is a consequence of 2.5 and Bochner's theorem. We shall omit its most general statement, taking our algebra A to be $L_1(G)$ so that no specific hypotheses concerning the algebra appear.

COROLLARY. 2.72. *Let τ be a map of $H^\wedge$ into $G^\wedge$ for which $\varphi \circ \tau$ is positive definite on $H^\wedge$ whenever φ is a positive definite element of $C_0(G^\wedge)$. Then τ is a continuous (but not necessarily open) homomorphism.*

Proof. Since the Fourier-Stieltjes transform of a measure is a linear combination of four positive definite functions we may define $T\mu$ as before for μ in $A = L_1(G)$ to obtain a non-zero homomorphism of $L_1(G)$ into $M(H)$. Moreover, $\mu \geq 0, \mu \in L_1(G)$ imply $\hat{\mu}$ is a positive definite element of $C_0(G^\wedge)$, and thus $(T\mu)^\wedge = \hat{\mu} \circ \tau$ is positive definite. Thus (by Bochner's theorem again) $T\mu \geq 0$, and we may apply 2.5 to obtain $T\mu = S^*\Gamma\mu$, $\mu \in L_1(G)$, in the notation of Theorem 2.1. But for $\hat{h} \notin H_0^\perp$ we have $L_1(G)^\wedge(\tau\hat{h}) = (TL_1(G))^\wedge(\hat{h}) = (S^*\Gamma L_1(G))^\wedge(\hat{h}) = 0$; hence from $\tau H^\wedge \subset G^\wedge$ we conclude that $H_0^\perp = H^\wedge$ and H_0 is trivial, $T\mu = \Gamma\mu$ and therefore $T\mu(\hat{h}) = \mu(\tau\hat{h}) = \mu(\hat{h} \circ \gamma)$, so that τ appears as the dual to γ, completing our proof.

The same proof (except for the final step) applies if one takes τ only to be a non-trivial map of $H^\wedge$ into $G^\wedge \cup \{0\}$ (i.e., with $\tau^{-1}G^\wedge \neq \phi$); one obtains the fact that $T\mu = S^*\Gamma\mu$, $\mu \in L_1(G)$, and concludes that τ is a continuous homomorphism on the open subgroup $\tau^{-1}G^\wedge$ of $H^\wedge$ (in order to consider $\varphi \circ \tau$ as defined on all of $H^\wedge$ one should include 0 in the domain of φ, with $\varphi(0) = 0$).

2.8. It is tempting to try the same approach in the non-commutative situation, replacing characters by finite dimensional matricial representations; apparently only in case H is compact can we obtain any consequences without a deeper investigation.

For any map σ of functions and matrix $U = (u_{ij})$ of functions let σU represent the matrix $(\sigma(u_{ij}))$. Then if U is any bounded continuous finite dimensional matricial representation of H, $\nu \to \nu(U)$ is a bounded representation of $M(H)$. Moreover if $T : L_1(G) \to M(H)$ is any bounded homomorphism, then $\mu \to T\mu(U)$ is a bounded representation of $L_1(G)$ and, as is well known, must be of the form $\mu \to \mu(\tilde{U})$, where $\tilde{U}$ is a continuous bounded matricial representation[8] of G. Viewing $C(H)$ as a subspace of $M(H)^*$, the adjoint T^* maps $C(H)$ into $L_1(G)^* = L_\infty(G)$, and we may clearly identify $\tilde{U}$ and $T^*U = (T^*u_{ij})$ as identical matrices of elements of $L_\infty(G)$. Consequently we can take T^*u_{ij} as a continuous function, indeed an almost periodic function, on G.

Now if H is compact the Peter-Weyl theorem assures us that we can view T^* as mapping $C(H)$ into $\mathfrak{A}(G)$; moreover this map τ is clearly norm-decreasing if T is. Each μ in $M(G)$ provides us with a functional $\tilde{\mu}$ on $\mathfrak{A}(G)$, and since $\tau^* : \mathfrak{A}(G)^* \to M(H)$ is norm-decreasing, $\|\tau^*\tilde{\mu}\| \leq \|\tilde{\mu}\| \leq \|\mu\|$ so that $\sigma : \mu \to \tau^*\tilde{\mu}$ is a norm-decreasing map of $M(G)$ into $M(H)$. But σ is automatically multiplicative: for

$$\begin{aligned}\sigma(\mu * \mu')(U) = (\mu * \mu')^{\sim}(\tau U) = \mu * \mu'(\tilde{U}) &= \mu(\tilde{U})\mu'(\tilde{U}) \\ &= \sigma\mu(U)\sigma\mu'(U) = (\sigma\mu * \sigma\mu')(U)\end{aligned}$$

for all U, so that $\sigma(\mu * \mu') = \sigma(\mu) * \sigma(\mu')$ by the Peter-Weyl theorem. Thus $E = \{\sigma\mu_g : g \in G\}$ forms a multiplicative group in the unit ball of $M(H)$.

Unfortunately the results of [6] do not determine all groups in the ball of $M(H)$ in the non-abelian case, but only those consisting of non-negative measures. E will be such a group if T (and therefore T^*, τ, τ^* and σ) preserves order; moreover we then have $\mu \to T\mu(1)$ a non-zero representation of $L_1(G)$ if $T \neq 0$ (otherwise $0 = T\mu(1) = \|T\mu\|$ for all $\mu \geq 0$, hence for all μ). Since $\mu \to T\mu(1)$ also preserves order, $T^*1 = 1$. As a consequence T is automatically norm-decreasing (cf. footnote 4), and $E \neq \{0\}$ since $\sigma\mu_g(1) = \mu_g(\tau 1) = \mu_g(T^*1) = 1$. We thus have E a set of translates of Haar measure of a normal subgroup H_0 of H, and can write as before $\sigma\mu_g = \nu^{\gamma(g)}$, $\gamma(g) \in H/H_0$.

But the map $g \to \tilde{\mu}_g$ of G into $\mathfrak{A}(G)^*$ (taken in the w^* topology) is continuous, so that $g \to \tau^*\tilde{\mu}_g = \nu^{\gamma(g)}$ is w^* continuous, and one can easily conclude that γ is a continuous homomorphism of G into H/H_0. Moreover since $g \to \tilde{\mu}_g$ is w^* continuous we can represent $\tilde{\mu}$ as the w^* convergent vector valued integral $\int \tilde{\mu}_g \mu(dg)$, $\mu \in M(G)$. Applying τ^* we

[8] That is, a homomorphism into a group of (possibly singular) matrices.

obtain $\tau^*\tilde{\mu} = \int \tau^*\tilde{\mu}_g \mu(dg) = \int \nu^{\gamma(g)}\mu(dg)$ so that $\tau^*\tilde{\mu}(f) = \int \nu^{\gamma(g)}(f)\mu(dg) = \mu(Sf \circ \gamma), f \in C(H)$, in our earlier notation. Finally we have $\tau^*\tilde{\mu} = T\mu$, $\mu \in L_1(G)$: for $\tau^*\tilde{\mu}(U) = \mu(\tau U) = \mu(T^*U) = T\mu(U)$, all U. Hence we may write $T = S^*\Gamma$.

Actually if T is any non-zero norm-decreasing homomorphism what we really need to know is that some one-dimensional representation of H induces a non-zero representation of $L_1(G)$. For then we have multiplicative characters χ' and χ of H and G respectively for which $T\mu(\chi') = \mu(\chi)$; consequently $\chi' T\chi^{-1}\mu(1) = T\chi^{-1}\mu(\chi') = \chi^{-1}\mu(\chi) = \mu(1)$ and the norm-decreasing map $T_0 : \mu \rightarrow \chi' T\chi^{-1}\mu$ has $T_0^*1 = 1$, whence it is easily seen to preserve order (as in 1.1). Thus $T\mu = (\chi')^{-1}S^*\Gamma\chi\mu$.

THEOREM 2.9. *Let G be any locally compact group, H any compact group. Then any non-zero order-preserving homomorphism $T : L_1(G) \rightarrow M(H)$ is of the form $S^*\Gamma$. If T is merely norm-decreasing and $T^*\chi'$ is a non-zero element of $L_\infty(G)$ for some multiplicative character χ' of H, then $T\mu = \chi'' S^*\Gamma\chi\mu$, where χ'', χ are multiplicative characters of H and G respectively; indeed $\chi'' = (\chi')^{-1}, \chi = T^*\chi'$.*

3. **Isomorphisms.** An almost immediate consequence of Corollary 2.31 is the fact that isometric isomorphisms between $M(G)$ and $M(H)$ arise in the same simple fashion as in the case of L_1 algebras. Actually we have a stronger result.

THEOREM 3.1. *Let G and H be locally compact abelian groups, and let A be a subalgebra of $M(G)$ containing $L_1(G)$, B a similar subalgebra of $M(H)$. Then for any isomorphism T of A onto B which is norm-decreasing on $L_1(G)$ there is an isomorphism γ of G onto H and a character $\hat{g}$ of G for which*

$$T\mu(f) = \mu(\hat{g}(f \circ \gamma)), \quad f \in C_0(H), \mu \in A .$$

Thus T is an isometry and $T_1(G) = L_1(H)$.

Before proceeding to the proof of Theorem 3.1 we might note that $L_1(G)$ can be replaced in our hypothesis by any subalgebra of $M(G)$ satisfying (2.01) and (2.02) which is an ideal in A.

Proof of Theorem 3.1. Applying Theorem 2.1 to the restriction of T to $L_1(G)$ we obtain characters $\hat{g}_1$ and $\hat{h}_1$, and operators S^* and Γ for which $T\mu = \hat{h}_1 S^*\Gamma\hat{g}_1\mu$, $\mu \in L_1(G)$. Consider the norm-decreasing isomorphism $T_0 = \hat{h}_1^{-1}T\hat{g}_1^{-1}$ of $A_0 = \hat{g}_1 A$ onto $B_0 = \hat{h}_1^{-1}B$. A_0 contains $L_1(G)$, and B_0 contains $L_1(H)$, while $T_0\mu = S^*\Gamma\mu$ for μ in $L_1(G)$. Evidently $\tilde{\nu} * T_0\mu = T_0\mu$, $\mu \in L_1(G)$, where $\tilde{\nu}$ is Haar measure on H_0. Since $\tilde{\nu}$ is an idempotent, $\mu \rightarrow \tilde{\nu} * T_0\mu$ is a homomorphism of A_0 into $M(H)$ which is one-to-one on $L_1(G)$. Consequently it is one-to-one on all of A_0: for $\tilde{\nu} * T_0\mu = 0$ implies $\tilde{\nu} * T_0(\mu * \mu') = 0$, $\mu' \in L_1(G)$, whence $\mu * \mu' = 0$ by the one-to-oneness on $L_1(G)$, and $\mu = 0$. But if $H_0 \neq \{h_0\}$ we have $H_0^\perp$ a proper open and closed subgroup of $H^\wedge$ so that we can find a ν in $L_1(H)$, $\nu \neq 0$, with $\hat{\nu}(H_0^\perp) = 0$, by the regularity of $L_1(H)$. Since $\hat{\tilde{\nu}}$ is the characteristic function of $H_0^\perp$, $(\tilde{\nu} * \nu)^\wedge = \hat{\tilde{\nu}}\hat{\nu} = 0$, and $\tilde{\nu} * \nu = 0$; on the other hand $\nu = T_0\mu$, $\mu \in A_0$, $\mu \neq 0$, so that $\tilde{\nu} * \nu \neq 0$ by the one-to-oneness of $\mu \to \tilde{\nu} * T_0\mu$, and we conclude that $H_0 = \{h_0\}$. Thus γ appears as a continuous homomorphism of G into H, and we may now write $T_0\mu = \Gamma\mu$, $\mu \in L_1(G)$.

As a consequence[9] we have $(T_0\mu)^\wedge(\hat{h}) = (\Gamma\mu)^\wedge(\hat{h}) = \mu(\hat{h} \circ \gamma) = \hat{\mu}(\hat{h} \circ \gamma)$, $\mu \in L_1(G)$, with $\hat{h} \circ \gamma \in G^\wedge$, so $T_0\mu \to (T_0\mu)^\wedge(\hat{h})$ is a non-zero functional on $T_0L_1(G)$. Repeating a previous computation, we have, for μ in A_0 and μ' in $L_1(G)$

$$T_0\mu * T_0\mu' = T_0(\mu * \mu') = \Gamma(\mu * \mu') = \Gamma\mu * \Gamma\mu' = \Gamma\mu * T_0\mu' ,$$

$L_1(G)$ being an ideal, so that $(T_0\mu - \Gamma\mu) * T_0L_1(G) = 0$. Thus for each $\hat{h}$, $(T_0\mu - \Gamma\mu)^\wedge(\hat{h}) = 0$ whence $T_0\mu = \Gamma\mu$, $\mu \in A_0$. Consequently $T\mu(f) = \hat{h}\Gamma\hat{g}\mu(f) = \Gamma\hat{g}\mu(\hat{h}f) = \hat{g}\mu(\hat{h} \circ \gamma \cdot f \circ \gamma) = \mu(\hat{g}_1(f \circ \gamma))$ for μ in A, and it remains to show γ is an isomorphism of G onto H.

First $\gamma(G)$ is dense in H; for otherwise we have a non-zero f in $C_0(H)$ with $f \circ \gamma = 0$, while $\nu(f) \neq 0$ for some ν in $L_1(H)$, $\nu = T_0\mu$, whence $0 \neq \nu(f) = T_0\mu(f) = \mu(f \circ \gamma) = 0$. Moreover γ is one-to-one since if $\gamma(g_1) = \gamma(g_2)$ then $\mu = \mu_{g_1} - \mu_{g_2}$ has $\Gamma\mu = 0$ (for $\mu(f \circ \gamma) = f(\gamma(g_1)) - f(\gamma(g_2)) = 0$). But then for μ' in $L_1(G)$ we have $\mu * \mu' \in L_1(G)$ and $T_0(\mu * \mu') = \Gamma(\mu * \mu') = \Gamma\mu * \Gamma\mu' = 0$ whence $\mu * \mu' = 0$ for all μ' in $L_1(G)$, and clearly $\mu = 0$, $g_1 = g_2$. Indeed the argument shows Γ is one-to-one on $M(G)$.

Consequently it is sufficient to show γ^{-1} is continuous on $\gamma(G)$; for then γ is a homeomorphism, $\gamma(G)$ is therefore locally compact and, being dense in H, must coincide with H as is well known. Suppose then that the net $h_\delta = \gamma(g_\delta) \to h_0 = \gamma(g_0)$. Clearly $\Gamma\mu_{g_\delta} = \nu_{h_\delta}$. For μ in A_0 with $T_0\mu$ in $L_1(H)$ we have $\nu_{h_\delta} * T_0\mu \in L_1(H) \subset T_0A_0 = \Gamma A_0$; clearly $\Gamma(\mu_{g_\delta} * \mu) = \nu_{h_\delta} * \Gamma\mu = \nu_{h_\delta} * T_0\mu$ so that $\mu_{g_\delta} * \mu \in A_0$ since Γ is one-to-one on $M(G)$, and further $T_0(\mu_{g_\delta} * \mu) = \nu_{h_\delta} * T_0\mu$. But $\|T_0(\mu_{g_\delta} * \mu) - T_0\mu\| = \|\nu_{h_\delta} * T_0\mu - T_0\mu\| \to 0$, $T_0\mu$ being in $L_1(H)$, and since $T_0^{-1}|L_1(H)$ is automatically continuous, $\|\mu_{g_\delta} * \mu - \mu\| \to 0$. As a consequence $\{g_\delta : \delta \geq \delta_0\}$ is contained in some compact $K \subset G$ for some δ_0; otherwise a cofinal subnet tends to infinity and $\overline{\lim}\, \|\mu_{g_\delta} * \mu - \mu\| = 2\|\mu\|$ for each such μ. If g is any cluster point of $\{g_\delta\}$ in K then, for each $\hat{g}$, $\mu_g * \mu(\hat{g})$ is a cluster

[9] This follows as in the final part of the proof of 2.1.

point of $\{\mu_{g_\delta} * \mu(\hat{g})\}$, which of course converges to $\mu(\hat{g})$ since $\|\mu_{g_\delta} * \mu - \mu\| \to 0$. Thus $\mu_g * \mu = \mu$ and $T_0\mu = \Gamma\mu = \Gamma(\mu_g * \mu) = \nu_{\gamma(g)} * \Gamma\mu = \nu_{\gamma(g)} * T_0\mu$; since $T_0\mu$ is an arbitrary element of $L_1(H)$ we clearly have $\gamma(g) = h_0$ and $g = g_0$. Consequently $\{g_\delta\}$ converges to g_0 by the compactness of K, and γ^{-1} is continuous.

Finally we have $\Gamma L_1(G) = L_1(H)$ since strong continuity of the map $g \to \mu_g * \mu$ is equivalent to strong continuity of $h \to \nu_h * \Gamma\mu$, and L_1 consists of just those measures for which strong continuity holds, by a theorem of Plessner. Consequently $TL_1(G) = L_1(H)$ and our proof is complete.

Applying 2.5 and 2.6 to $T|L_1(G)$, we obtain

COROLLARY 3.11. *Let T be any isomorphism of A onto B for which $\mu \geq 0$, $\mu \in L_1(G)$ imply $T\mu \geq 0$. Then T is an isometry Γ induced by an isomorphism γ of G onto H.*

COROLLARY 3.12. *If $H^\wedge$ is connected any isomorphism T of A onto B is on isometry determined as in* 3.1.

THEOREM 3.2. *When G and H are arbitrary compact groups, the conclusions drawn in Theorem* 3.1 *and Corollary* 3.11 *continue to hold.*

Proof. Consider first the situation indicated by 3.1, and let μ^0, ν^0 be the Haar measures on G and H. Then $T\mu^0$ is a non-zero idempotent in the unit ball of $M(H)$, and thus, by the result (2) cited in the introduction, of the form $\chi_1\nu$ where ν is Haar measure of a subgroup of H, and χ_1 is a multiplicative character of this subgroup.

But $A * \mu^0 = K\mu^0$, K the complex field, so $B * (\chi_1\nu) = K(\chi_1\nu)$. Taking $M(H) = C(H)^*$ in the w^* topology, the linear map $\nu' \to \nu' * (\chi_1\nu)$ of $M(H)$ into itself is of course continuous, and clearly is of norm $\leqq 1$. In particular the unit ball of B maps into $D \cdot (\chi_1\nu)$, where D is the unit disc $\{z : |z| \leqq 1\}$ in K. Since each ν_h is w^* adherent to the unit ball of $L_1(H) \subset B$, we obtain $\nu_h * (\chi_1\nu) \in D(\chi_1\nu)$ for each h in H, and the carrier of ν must be translation invariant. Consequently $\nu = \nu^0$ and χ_1 appears as a character of the full group H.

Thus $\mu \to T\mu(\chi_1^{-1})$ is a non-trivial one-dimensional representation of $L_1(G)$: for $T\mu^0(\chi_1^{-1}) = \chi_1\nu^0(\chi_1^{-1}) = 1$. As in 2.8 we obtain a multiplicative character χ of G for which $T\mu(\chi_1^{-1}) = \mu(\chi)$; since $\mu^0(\chi) = 1$, $\chi = 1$, and by 2.9 we have $T\mu = \chi_1 S^*\Gamma\mu$, $\mu \in L_1(G)$. Setting $T_0\mu = \chi_1^{-1}T\mu$ we obtain an isomorphism of A onto $B_0 = \chi_1^{-1}B$, with $T_0\mu = S^*\Gamma\mu$, $\mu \in L_1(G)$, and, in particular, $T_0\mu^0 = \chi_1^{-1}T\mu^0 = \nu^0$.

As in 3.1, γ must be one-to-one; otherwise we have a $\mu \neq 0$ in $M(G)$ with $\Gamma\mu = 0$ so that $T_0(\mu * \mu') = S^*\Gamma(\mu * \mu') = S^*\Gamma\mu * S^*\Gamma\mu' = 0$

for all μ' in $L_1(G)$, and[10] $\mu * L_1(G) = 0$, $\mu = 0$. Moreover if the compact image $\gamma(G)$ of G in H/H_0 were not all of H/H_0 we should have an f in $C(H/H_0)$ with $f \neq 0, f \geq 0, f \circ \gamma = 0$; thus if ρ denotes the canonical map of H onto H/H_0,

$$0 < \nu^0(f \circ \rho) = T_0\mu^0(f \circ \rho) = S^*\Gamma\mu^0(f \circ \rho) = \Gamma\mu^0(f) = \mu^0(f \circ \gamma) = 0 ;$$

consequently γ maps G onto H/H_0, and therefore is a homeomorphism and isomorphism between these groups. But now Γ appears as an isometry mapping $M(G)$ onto $M(H/H_0)$, and since S^* is easily seen to be an isometry, $T_1 = T_0 | L_1(G) = S^*\Gamma | L_1(G)$ is isometric. This combines with $T_0\mu^0 = \nu^0$ to show T_1 and T_1^{-1} preserve order: for

$$\mu \geq 0 \Longleftrightarrow \mu * \mu^0 = ||\mu||\mu^0 \Longleftrightarrow T_1\mu * \nu^0 = ||\mu||\nu^0 = ||T_1\mu||\nu^0$$
$$\Longleftrightarrow T_1\mu \geq 0 .$$

Consequently T_1 maps $\{\mu : 0 \leq \mu \leq \mu^0\}$ onto $\{\nu : 0 \leqq \nu \leqq \nu^0\}$, or, more generally, the algebra $L_\infty(G)\mu^0 = \{f \cdot \mu^0 : f \in L_\infty(G)\}$ onto $L_\infty(H) \cdot \nu^0$. As an isometry T_1 thus maps closure onto closure, or $L_1(G)$ onto $L_1(H)$, and we are forced to conclude that H_0 is trivial since its Haar measure acts as an identity on $T_1L_1(G) = L_1(H)$. Hence γ is an isomorphism of G onto H, and $T_1\mu = \Gamma\mu$. As before we conclude that $T_0\mu = \Gamma\mu$, $\mu \in A$: for with $\mu' \in L_1(G)$, $T_0\mu * T_0\mu' = T_0(\mu * \mu') = \Gamma(\mu * \mu') = \Gamma\mu * \Gamma\mu' = \Gamma\mu * T_0\mu'$ and $(T_0\mu - \Gamma\mu) * L_1(H) = 0$. Thus we have $T\mu = \chi_1\Gamma\mu = \Gamma\chi\mu$, as in 2.2, or $T\mu(f) = \mu(\chi \cdot (f \circ \gamma))$, proving the analogue of 3.1. The analogue of 3.11 follows since our T must then be norm-decreasing on $L_1(G)$ as in 2.8.

3.3. Returning to the abelian case, the results of Šreider [14] for $G = R$ indicate that $G^\wedge$ forms a smaller part of the maximal ideal space of $M(G)$ than one might initially presume. As one would suspect from the one-to-one nature of the Fourier-Stieltjes transformation however, $G^\wedge$ would seem still to occupy a rather dominant rôle in the Gelfand representation of $M(G)$; this view is certainly reinforced by 3.1 since it shows the norm-decreasing automorphisms of $M(G)$ can only induce self-homeomorphisms of the maximal ideal space which leave $G^\wedge$ invariant, and indeed preserve its algebraic structure.

3.4. A variant of the proof of Theorem 3.1 yields the form of all norm-decreasing isomorphisms of $L_1(G)$ onto a closed subalgebra of $L_1(H)$; when $G^\wedge$ is connected this yields the answer to the question: what (proper, closed) subalgebras isomorphic to $L_1(G)$ can $L_1(G)$ contain? Clearly if G_1 is a proper open subgroup isomorphic to G then $L_1(G)$

[10] If $\mu * L_1(G) = 0$ then for $f, F \in C(G)$ we have $0 = \int\int f(g_1g_2)\mu(dg_1)F(g_2^{-1})\mu^0(dg_2) = \int f * F(g_1)\mu(dg_1)$, whence $\mu = 0$ since such convolutions $f * F$ are dense in $C(G)$.

provides such a subalgebra; when $G^\frown$ is connected these are the only candidates.[11]

THEOREM 3.5. *Let A be a closed ideal in $M(G)$ satisfying* (2.01) *and* (2.02) *and let T be an isomorphism of A onto a closed subalgebra B of $L_1(H)$. If T is norm-decreasing then $T\mu = \hat{h}S^*\Gamma\hat{g}\mu$ (as in* 2.1) *where γ is an isomorphism of G onto an open subgroup of H/H_0. In particular if $H^\frown$ is connected any isomorphism of A onto B is of the form $\mu \to \Gamma\hat{g}\mu$, where γ maps G isomorphically onto an open subgroup of H.*

Proof. By 2.6 the second assertion follows from the first. By 2.1 we have $T\mu = \hat{h}S^*\Gamma\hat{g}\mu$; as before we can eliminate $\hat{g}, \hat{h}$, and may as well assume $T\mu = S^*\Gamma\mu$. Since A is an ideal in $M(G)$ and $\mu * A = 0$ implies $\mu = 0$ by (2.02), we conclude exactly as in 3.1 that γ is one-to-one.

Moreover if γ^{-1} isn't continuous on $\gamma(G) \subset H/H_0$, for some neighborhood U of g_0 we have $\gamma^{-1}(VH_0) \cap U' \neq \phi$ for each neighborhood V of h_0; let $g_V \in \gamma^{-1}(VH_0) \cap U'$ and let $\gamma(g_V) = h_V H_0$ where $h_V \in V$. Since

$$T(\mu * \mu_{g_V}) = S^*\Gamma(\mu * \mu_{g_V}) = S^*\Gamma\mu * \nu^{\gamma(g_V)} = S^*\Gamma\mu * \nu_{h_V} = T\mu * \nu_{h_V}$$

(where $\nu^{\gamma(g)}$ is the translate to $\gamma(g)$ of Haar measure on H_0 as before) we conclude from the strong convergence of $T\mu * \nu_{h_V}$ to $T\mu$ and the automatic continuity of T^{-1} (B being closed) that $\|\mu * \mu_{g_V} - \mu\| \to 0$. Noting that $\mu * \mu_g = \mu$ for all μ in A implies $g = g_0$ by (2.02), our previous argument yields the fact that $g_V \to g_0$, contradicting $g_V \in U'$. Thus γ is a (topological) isomorphism, $\gamma(G)$ is locally compact and therefore closed, and we need only show $\gamma(G)$ open to complete our proof.

Let H_1 be the inverse image of $\gamma(G)$ under the canonical homomorphism of H onto H/H_0, a closed subgroup of H. If $f \in C_0(H)$ vanishes on H_1 we clearly have $S^*\Gamma\mu(f) = 0$ so that the regular Borel measure $S^*\Gamma\mu$ vanishes on all Borel subsets of the complement of H_1. Since $S^*\Gamma\mu$ is a non-zero element of $L_1(H)$ for some μ, H_1 clearly contains some compact set C of positive Haar measure, and thus must be open[12]; hence $\gamma(G)$ is open and our proof complete.

---

[11] It is of course not the answer otherwise. For example let A_n be the algebra of integrable f on the circle T^1 with $f(e^{i\theta}) = f(e^{i(\theta+2\pi/n)})$; then setting $g(t) = f(t^{1/n})$ yields a well defined element g of $L_1(T_1)$, and $f \to g$ is easily seen to be an isomorphism of A_n with $L_1(T^1)$.

[12] For we can find a Baire subset E of H containing C with $E \backslash C$ of (Haar) measure zero; then the Borel measurable function $h \to \varphi_E(h)\varphi_E{}^{-1}(h^{-1}h') - \varphi_C(h)\varphi_C{}^{-1}(h^{-1}h')$ (for h' fixed, φ_E the characteristic function) differs from zero only on a subset of $(E\backslash C) \cup (h'E\backslash h'C)$ so that $\varphi_E * \varphi_E{}^{-1}(h') = \int \varphi_C(h)\varphi_C{}^{-1}(h^{-1}h')dh$. As usual the fact that $\varphi_E * \varphi_E{}^{-1} \neq 0$ on a neighborhood U of h_0 yields for $h' \in U$ a h in C with $h' \in hC^{-1} \subset H_1$ whence $U \subset H_2$ and H_1 is open.

4. Some other isomorphisms. The rôle of condition (2.02) in §2 was confined to providing us with a map τ of character groups dual to a given homomorphism of our algebra A. In certain situations such a τ arises naturally in the absence of (2.02) and provided (2.01) holds, our approach may again be applicable. For example suppose A is a closed subalgebra of $M(G)$ satisfying (2.01) for which $G^{\wedge}$ forms a subspace of the maximal ideal space $\mathfrak{M}$ of A; further suppose $G^{\wedge}$ is connected. Then any endomorphism T of A for which the dual map $\tau : \mathfrak{M} \to \mathfrak{M}$ sends $G^{\wedge}$ into itself necessarily has $\tau(\hat{g}) = \hat{g}_1 \cdot \sigma(\hat{g}), \hat{g} \in G^{\wedge}$, where σ is an endomorphism of $G^{\wedge}$. For T is necessarily bounded so that τ induces a bounded map of $\mathfrak{A}(G)$ into itself (by an analogue of (2.13), using (2.01)), and Corollary 2.61 applies. Consequently T is itself determined as before; similarly if $G^{\wedge}$ is not connected but T is also norm-decreasing, or order-preserving while (2.51) obtains, we can apply Corollary 1.3 or the remark following 2.5 to the same end.

Exactly such a situation arises in connection with the Arens-Singer theory of generalized analytic functions [1], in particular in Arens' subsequent generalization of the conformal mappings of the disc [2]. There (among other things) Arens is interested in the automorphisms[13] of a certain closed subalgebra A_1 of $L_1(G), G$ locally compact abelian; one has a fixed closed subset G_+ of G satisfying [1, §2]

(4.01) *G_+ is a subsemigroup of G, i.e., $x, y \in G_+$ imply $xy \in G_+$,*

(4.02) *the interior of G_+ is dense in G_+ and generates G;*

$A_1 = L_1(G_+)$ is then the set of all elements of $L_1(G)$ vanishing off G_+. As Arens and Singer showed, $L_1(G_+)$ has $G^{\wedge}$ as the Šilov boundary of its maximal ideal space; consequently (by a well known property of the Šilov boundary) any automorphism T of $L_1(G_+)$ induces a self-homeomorphism τ of its maximal ideal space which maps $G^{\wedge}$ onto itself. Moreover the fact that G_+ generates G shows (2.01) and (2.51) hold for $L_1(G_+)$. For the closure G_+^- of G_+ in G^* is a generating subsemigroup of G^*, while any closed subsemigroup of a compact group is a subgroup [5, 11]. Thus $G_+^- = G^*$, and G_+, as well as its interior, is dense in G^*; since point masses concentrated at interior points can clearly be approximated by elements of the unit ball of $L_1(G_+)$ in the weak topology defined by almost periodic functions, we obtain (2.01) and (2.51).

Consequently if $G^{\wedge}$ is connected we have $\tau(\hat{g}) = \hat{g}_1 \cdot \sigma(\hat{g}), \hat{g} \in G^{\wedge}$, by 2.61, where σ is an automorphism of $G^{\wedge}$. Writing elements of $L_1(G)$ as functions rather than measures, we thus have $(Tf)^{\wedge}(\hat{g}) = \hat{f}(\hat{g}_1\sigma(\hat{g}))$ $=(\hat{g}_1 f)^{\wedge}(\sigma(\hat{g})) = k[(\hat{g}_1 f) \circ \gamma]^{\wedge}(\hat{g})$, where γ^{-1} is the automorphism of G dual to σ, and $k > 0$ compensates for the change in Haar measure produced by γ (of course $k = 1$ if G is discrete). Clearly $Tf = k(\hat{g}_1 f) \circ \gamma$ says $\gamma G_+ = G_+$

[13] These are not Arens' full set of generalized conformal mappings, which correspond to the automorphisms of his algebra A_0.

and we have additional information about τ. Thus in the classical case of the Arens-Singer theory (where G is the group Z of integers, G_+ the non-negative integers, and $L_1(G_+)^\wedge$ may be viewed[14] as the algebra of analytic functions with absolutely convergent Taylor series on the disc $|z| \leqq 1$ (the maximal ideal space of $L_1(G_+)$)) γ must be the identity, so that τ reduces to a rotation on $|z| = 1$, hence[15] is a rotation of $|z| \leq 1$. In other words *the only self-homeomorphisms τ of the disc which* (via $F \to F \circ \tau$) *map the set of analytic functions with absolutely convergent series on the disc onto itself are rotations.*

Again in the case $G = Z \times Z, G_+ = \{(m, n) : m, n \geq 0\}$, where $L_1(G_+)$ can be viewed as the algebra of analytic functions of two complex variables with power series absolutely convergent on $|z| \leq 1, |w| \leqq 1$, there are clearly only two candidates for γ ($(m, n) \to (n, m)$ and the identity), and thus the general automorphism is of the form

$$\sum a_{mn} z^m w^n \to \sum a_{mn} c^m z^m d^n w^n$$

or

$$\sum a_{mn} z^m w^n \to \sum a_{mn} c^m w^m d^n z^n$$

where c and d are fixed unimodular constants; in other words each automorphism is induced by separate rotations of each disc $|z| \leq 1$, $|w| \leq 1$, plus a possible interchange of variables. Clearly this extends to n complex variables.

Generalizing our setting slightly we have

THEOREM 4.1. *Let G and H be locally compact abelian groups with closed subsemigroups G_+ and H_+ satisfying* (4.02), *and let $L_1(G_+)$, $L_1(H_+)$ be defined as above. Then if either group has a connected dual an isomorphism T of $L_1(G_+)$ onto $L_1(H_+)$ is an isometry of the form $Tf = k(\hat{g}f) \circ \gamma$, where k is a positive constant and γ an isomorphism of H onto G with $\gamma H_+ = G_+$. Without connectedness the same applies to order-preserving or norm-decreasing isomorphisms.*

4.2. Clearly most of what we have said applies equally well to any closed algebra satisfying (2.01) for which $G^\wedge$ yields the Šilov boundary. And any closed subalgebra A of $L_1(G)$, with $A^\wedge$ a translation invariant

[14] More precisely $L_1(G_+)^\wedge$ is the set of restrictions to $|z| = 1$ of these functions (since $^\wedge$ still is the Fourier transformation and not the full Gelfand representation, cf. [**10**, p. 72]).

[15] For τ is analytic as the function representing the characteristic function of $\{1\}$ under the full Gelfand representation. Alternatively we could note that knowledge of τ on the Šilov boundary determines τ among all automorphism-inducing self-homeomorphisms of $\mathfrak{M}$; since here rotation of the full disc is clearly such a homeomorphism it coincides with τ on the full disc.

set of functions on $G^\frown$ which separate the elements of $G^\frown \cup \{0\}$, has $G^\frown$ the Šilov boundary ∂ of its maximal ideal space. For $G^\frown$ forms a subspace of the maximal ideal space (cf. footnote 3), while if $\mu \to \mu^0$ is the Gelfand representation of A, $\|\mu^0\|_\infty = \lim \|\mu^{(n)}\|^{1/n} = \|\hat{\mu}\|_\infty = |\hat{\mu}(\hat{g})|$ for some $\hat{g}$ in $G^\frown$, for each μ in A, and $\partial \subset G^\frown$. But since $A^\frown$ is translation invariant we clearly have ∂ a translation invariant subset of $G^\frown$, and $\partial = G^\frown$ (this is precisely the argument of [1]).

Consequently we obtain as before

THEOREM 4.3. *Let A be a closed subalgebra of $L_1(G)$ which is closed under multiplication by elements of $G^\frown$, B a similar subalgebra of $L_1(H)$, and suppose A satisfies* (2.01) *while B merely has $B^\frown$ a separating set of functions on $H^\frown \cup \{0\}$. Then if $H^\frown$ is connected any isomorphism T of A onto B is an isometry of the form $T\mu = \Gamma\hat{g}\mu$ (notation as in* 2.1), *where γ is an isomorphism of G onto H. Without connectedness the same applies to norm-decreasing* (*or, if A satisfies* 2.51, *order-preserving*) *isomorphisms.*

Here γ is the isomorphism dual to the isomorphism σ we obtain from 2.61, etc., rather than its inverse, which is the γ of 4.1.

5. When G is discrete a general theorem of Šilov [13] shows that $L_1(G)$ is the direct sum of a pair of ideals if and only if $G^\frown$ is disconnected. When $G^\frown$ is connected $L_1(G)$ may still be the vector space direct sum of a closed ideal and a closed subalgebra, and Theorem 2.6 then reveals the exact situation.

THEOREM 5.1. *Let $G^\frown$ be connected, and $L_1(G) = A \oplus I$ where A is a* (*non-zero*) *closed subalgebra and I a* (*non-zero*) *closed ideal. Then G is the direct product of a discrete subgroup G_1 and an open subgroup G_2 for which $A = L_1(G_2)$ and $I = \{\mu : \mu \in L_1(G), \hat{\mu}(\hat{g}_1 G_1^\perp) = 0\}$, where $\hat{g}_1 \in G_2^\perp$. Conversely any such decomposition of G and character $\hat{g}_1$ orthogonal to G_2 yields a decomposition of $L_1(G)$ of the type described.*

Proof. Let T be the projection of $L_1(G)$ onto A, a nonzero homomorphism. By 2.6, $T\mu = \Gamma\hat{g}_1\mu$ where Γ is induced by a continuous endomorphism γ of G. Let σ be the endomorphism of $G^\frown$ dual to γ, so that $\hat{g} \circ \gamma = \sigma(\hat{g})$ and $T\hat{\mu}(\hat{g}) = \hat{\mu}(\hat{g}_1(\hat{g} \circ \gamma)) = \hat{\mu}(\hat{g}_1\sigma(\hat{g}))$. Since $T^2 = T$, $\hat{\mu}(\hat{g}_1\sigma(\hat{g})) = T^2\hat{\mu}(\hat{g}) = T\hat{\mu}(\hat{g}_1\sigma(\hat{g})) = \hat{\mu}(\hat{g}_1\sigma(\hat{g}_1\sigma(\hat{g})))$. Consequently $\sigma(\hat{g}) = \sigma(\hat{g}_1\sigma(\hat{g})) = \sigma(\hat{g}_1)\sigma(\sigma(\hat{g}))$ whence (setting $\hat{g} = \hat{g}_0$) $\hat{g}_0 = \sigma(\hat{g}_1) = \hat{g}_1 \circ \gamma$ and $\sigma \circ \sigma = \sigma$. Dually $\gamma \circ \gamma = \gamma$, and thus the algebraic subgroup $G_2 = \gamma(G)$ of G, on which γ acts as an identity map, is closed (for $\gamma(g_\delta) \to g$ implies $\gamma(\gamma(g_\delta)) = \gamma(g_\delta) \to \gamma(g)$ and $\to g$ whence $g = \gamma(g) \in G_2$). Moreover the fact that $\hat{g}_0 = \hat{g}_1 \circ \gamma$ says $\hat{g}_1 \in G_2^\perp$.

But G_2 is open as well. For $\Gamma\mu$ is a non-zero element of $L_1(G)$ for some μ in $L_1(G)$, while $\Gamma\mu(f) = \mu(f \circ \gamma) = 0$ for $f \in C_0(G)$ vanishing on $G_2 = \gamma(G)$, so that the regular Borel measure $\Gamma\mu$ vanishes on all Borel sets in the complement of G_2; thus G_2 contains some compact subset C of positive Haar measure, and must be open (cf. footnote 12).

Set $G_1 = \{g\gamma(g)^{-1} \colon g \in G\}$, clearly an algebraic subgroup of G. Then $g = (g\gamma(g)^{-1}) \cdot \gamma(g)$ yields a direct product decomposition of G, $G = G_1 \otimes G_2$: for $g \in G_1 \cap G_2$ implies $g = g'\gamma(g')^{-1} = \gamma(g) = \gamma(g')\gamma(g')^{-1} = g_0$. Since G_2 is open, G_1 is clearly discrete, and evidently γ is the projection of G onto G_2 corresponding to our decomposition.

Let $\mu^{g_1G_2}$ be the restriction of the measure μ in $L_1(G)$ to g_1G_2, so that $\mu = \sum_{g_1 \in G_1} \mu^{g_1G_2}$ and

$$\Gamma\mu(f) = \mu(f \circ \gamma) = \sum_{g_1 \in G_1} \int_{g_1G_2} f(\gamma(g))\mu(dg), \quad f \in C_0(G) .$$

Since

$$\mu_{g_1^{-1}} * \mu^{g_1G_2}(f) = \int f(g_1^{-1}g)\mu^{g_1G_2}(dg) = \int_{g_1G_2} f(g_1^{-1}g)\mu(dg)$$

and $g_1^{-1}g = \gamma(g)$ for $g \in g_1G_2$ we have $\Gamma\mu = \sum_{g_1 \in G_1} \mu_{g_1^{-1}} * \mu^{g_1G_2}$. But this clearly implies Γ, and therefore T, maps $L_1(G)$ into $L_1(G_2)$; indeed it shows Γ and (since $\hat{g}_1 \in G_2^{\perp}$) T leave elements of $L_1(G_2)$ fixed so that $A = TL_1(G) = L_1(G_2)$. On the other hand I, being the kernel of T, consists of just those μ in $L_1(G)$ with $\Gamma\hat{g}_1\mu = 0$, i.e. with $\hat{\mu}(\hat{g}_1\sigma(\hat{g})) = 0$, $\hat{g} \in G^{\wedge}$. Thus $\mu \in I$ if and only if $\hat{\mu}(\hat{g}_1\sigma(G^{\wedge})) = 0$ or $\hat{\mu}(\hat{g}_1G_1^{\perp}) = 0$ since σ, as the dual to the projection γ of $G_1 \otimes G_2$ onto G_2, is the projection of $G^{\wedge} = G_2^{\perp} \otimes G_1^{\perp}$ onto $G_1^{\perp}$.

Conversely given $G = G_1 \otimes G_2$, $\hat{g}_1 \in G_2^{\perp}$ one need only set $T\mu = \sum_{g_1 \in G_1} \mu_{g_1^{-1}} * (\hat{g}_1\mu^{g_1G_2})$ to obtain a projection of $L_1(G)$ onto $L_1(G_2)$; writing $\hat{g}_1'\hat{g}_2'$ (with $\hat{g}_1' \in G_2^{\perp}$, $\hat{g}_2' \in G_1^{\perp}$) as the generic element of $G^{\wedge}$ an easy computation shows $T\hat{\mu}(\hat{g}_1'\hat{g}_2') = \hat{\mu}(\hat{g}_1\hat{g}_2')$ so that T is clearly multiplicative and I, as described, is its kernel.

If $G^{\wedge}$ is disconnected our present tools can only be applied to those decompositions for which $\|T\| = 1$ (that other cases occur can be seen from the results of [12] for G the circle group); one can then obtain an analogous result, somewhat complicated by the fact that γ appears as a homomorphism of G into G/G_0, G_0 compact, and indeed the decomposition of L_1 arises from a decomposition of G/G_0, $G/G_0 = G_1/G_0 \otimes G_2/G_0$, and A appears as $S^*L_1(G_2/G_0)$.

6. Some reformulations. When G and H are compact abelian groups Corollary 1.3 has an interesting reformulation; our final section will be devoted to this result and some analogues.

THEOREM 6.1. *Let G and H be compact abelian groups and let T*

be any norm-decreasing linear map of the Banach space $C(H)$ into $C(G)$ for which $TH^\wedge \subset G^\wedge$. Then there is a homomorphism γ of G into H for which $Tf = (T\hat{h}_0) \cdot f \circ \gamma, f \in C(H)$. In particular if $T\hat{h}_0 = \hat{g}_0$ then T is a Banach algebra homomorphism when $C(G)$ and $C(H)$ are equipped with ordinary multiplication.

Further the range of T is dense iff γ is one-to-one, and then $TH^\wedge = G^\wedge$ and T is onto, while T is an isometry iff $\gamma(G) = H$.

Although we could obtain a proof by noting that T is merely the linear extension of $\tau = T|H^\wedge$ we obtain in the proof of Theorem 1.1, an appeal to Corollary 1.3 is more direct. Clearly τ satisfies the hypothesis of 1.3, and thus $\sigma : \hat{h} \to (\tau\hat{h}_0)^{-1}\tau\hat{h}$ is a homomorphism of $H^\wedge$ into $G^\wedge$. Since $H^\wedge$ and $G^\wedge$ are discrete, and σ thus continuous, we have a continuous dual homomorphism $\gamma : (g, \sigma(\hat{h})) = (\gamma(g), \hat{h})$. Thus

$$\left(\sum_{i=1}^{n} a_i\hat{h}_i\right)(\gamma(g)) = \left(\sum_{i=1}^{n} a_i\sigma(\hat{h}_i)\right)(g) = (g, (\tau\hat{h}_0)^{-1})\left(T\sum_{i=1}^{n} a_i\hat{h}_i\right)(g),$$

or

$$T\sum_{i=1}^{n} a_i\hat{h}_i = \tau\hat{h}_0\left[\left(\sum_{i=1}^{n} a_i\hat{h}_i\right)\circ\gamma\right].$$

Since trigonometric polynomials are dense $Tf = \tau\hat{h}_0 \cdot (f \circ \gamma), f \in C(H)$.

For the final statements, we clearly need only consider the case $T\hat{h}_0 = \hat{g}_0$. Note that if γ is not one-to-one then $Tf = f \circ \gamma$ says the range of T consists of functions constant on the cosets of the non-trivial kernel of γ, and thus the range cannot be dense in $C(G)$. On the other hand if γ is one-to-one, then (by compactness) it is an isomorphism of G with a subgroup $\gamma(G)$ of H. Thus for any character χ of $\gamma(G)$ we have a character $\hat{g}$ of G for which $\chi \circ \gamma = \hat{g}$, and since $\chi = \hat{h}|\gamma(G)$ for some $\hat{h}$ in $H^\wedge$ we obtain $\hat{h} \circ \gamma = \chi \circ \gamma = \hat{g}$, whence $G^\wedge = TH^\wedge$. Further if $F \in C(G)$ then any continuous extension f of $F \circ \gamma^{-1} \in C(\gamma(G))$ to all of H (available by Urysohn's lemma) yields $f \circ \gamma = F$, and T is onto. Lastly, if $\gamma(G)$ is proper we have an non-zero $f \in C(H)$ vanishing on $\gamma(G)$ so that $Tf = f \circ \gamma = 0$, and T is not even one-to-one, while if $\gamma(G) = H$ then T is clearly an isometry.

In one case specific mention of characters as such can be eliminated, yielding the weaker result: if T is a linear norm-decreasing one-to-one map of $C(H)$ into $C(G)$ taking the positive definite functions in the ball of $C(H)$, $P_0(H)$, onto $P_0(G)$, then $f \to (T\hat{h}_0)^{-1}Tf$ is multiplicative. For with one-to-oneness the set of extreme points $H^\wedge \cup \{0\}$ of $P_0(H)$ maps onto those of $P_0(G)$, $G^\wedge \cup \{0\}$.

In this form we have an indication that a similar result can be obtained for the L_1 algebras of locally compact abelian groups.

THEOREM 6.2. *Let G and H be locally compact abelian groups and $P_1(G), P_1(H)$ be the integrable positive definite functions. If T is a* *linear isometry of the Banach space* $L_1(G)$ *onto* $L_1(H)$ *with* $TP_1(G) = P_1(H)$ *then* T *is an algebra isomorphism.*

Before proceeding to a proof of 6.2 we should perhaps note an abstract version. Recall that an extreme positive (extendable) functional on a commutative Banach * algebra is a * preserving multiplicative functional. Then

THEOREM 6.3. *Let* A *and* B *be commutative Banach * algebras with (without) identities, and suppose* B *is semisimple and symmetric. Let* T *be a linear isometry of the Banach space* A *into* B *for which the adjoint map* T^* *takes the positive (extendable) functionals on* B *onto those on* A. *Then* T *is a * isomorphism of the algebras* A *into* B.

Proof. Let $P(A)$, $P(B)$ be the set of positive (extendable) functionals of norm 1 on A, B. We know T^*, being an isometry, maps $P(A)$ onto $P(B)$. Since it is one-to-one T^* must map the set $P(A)^e$ of extreme points of $P(A)$ onto $P(B)^e$. But these sets consist of * preserving multiplicative functionals, and since each multiplicative functional on B is * preserving by hypothesis, and thus an extreme positive (extendable) functional, T^* provides us with a map of $\mathfrak{M}_B$, the maximal ideal space of B, into $\mathfrak{M}_A$. Consequently (with $\hat{}$ now the Gelfand representation), $(Taa')\hat{}(M) = (aa')\hat{}(T^*M) = \hat{a}(T^*M)\hat{a}'(T^*M) = (Ta)\hat{}(M) \cdot (Ta')\hat{}(M)$. Since B is semisimple, $Taa' = Ta \cdot Ta'$, and we need only verify $Ta^* = (Ta)^*$. But since M and T^*M are * preserving for M in $\mathfrak{M}_B$, $(Ta^*)\hat{}(M) = \hat{a}^*(T^*M) = \bar{\hat{a}}(T^*M) = \overline{(Ta)\hat{}(M)} = (Ta)^{*\wedge}(M)$, so $Ta^* = (Ta)^*$ also follows from the semisimplicity of B.

The proof of Theorem 6.2 now follows quite simply, for, as is well known, the positive (extendable) functionals on $L_1(G)$ form the polar cone of $P_1(G)$. Thus the adjoint of T satisfies the requirements of 6.3 when $A = L_1(G)$, $B = L_1(H)$, and T is an algebra isomorphism.

REFERENCES

1. R. Arens and I. M. Singer, *Generalized analytic functions*, Trans. Amer. Math. Soc. **81** (1956), 379-393
2. R. Arens, *A Banach algebra generalization of conformal mappings on the unit disc*, Trans. Amer. Math. Soc., **81** (1956), 501-513.
3. A. Beurling and H. Helson, *Fourier-Stieltjes transforms with bounded powers*, Math. Scand. **1** (1953), 120-126.
4. W. F. Eberlein, *Characterizations of Fourier-Stieltjes transforms*, Duke Math. J. **22** (1955), 465-468.
5. B. Gelbaum, G. Kalish, and J.M.H. Olmstead, *On the imbedding of topological semigroups and integral domains*, Proc. Amer. Math. Soc. **2** (1951), 807-821,
6. I. Glicksberg, *Convolution semigroups of measures*, Pacific J. Math. **9** (1959), 51-67.
7. H. Helson, *Isomorphisms of abelian group algebras*, Arkiv for Mat. **2** (1953), 475-487.
8. J. P. Kahane, *Sur less fonctions sommes des séries trigonométriques absolument convergentes*, Comptes Rendues **240** (1955), 36-37.

9. Z. L. Leibenson, *On the ring of functions with absolutely convergent Fourier series*, Uspehi Matem. Nauk N.S. **9** (1954), 157-162.

10. L. Loomis, *An Introduction to Abstract Harmonic Analysis*, New York, 1953.

11. K. Numakura, *On bicompact semigroups*, Math. J. of Okayama University, **1** (1952), 99-108.

12. W. Rudin, *The automorphisms and endomorphisms of the group algebra of the unit circle*, Acta Math. **95** (1956), 39-55.

13. G. E. Šilov, *On the decomposition of a commutative normed ring into a direct sum of ideals*, Translations of the A.M.S., series 2, vol. 1, pp. 37-48.

14. Yu. A. Sreider, *The structure of maximal ideals in rings of measures with convolution*, Mat. Sbornik N. S. **27** (69) (1950), 279-318; Amer. Math. Soc. Translation No. 81.

15. A. Weil, *L'Integration dans les Groupes Topologiques et ses Applications*, Actualites Scientifiques et Industrielles, no. 369, Paris, 1940.

16. J. G. Wendel, *On isometric isomorphisms of group algebras*. Pacific J. Math. **1** (1951), 305-311.

17. J. G. Wendel, *Left centralizers and isomorphisms of group algebras*, Pacific J. Math. **2** (1952) 251-261.

UNIVERSITY OF NOTRE DAME
STANFORD UNIVERSITY

SOME APPLICATIONS OF EXPANSION CONSTANTS

B. GRÜNBAUM

1. For any metric space X (with distance function d) the *expansion constant* $E(X)$ of X is the greatest lower bound of real numbers μ which possess the following property ($S(x;\rho) = S_X(x;\rho) = \{y \in X; d(x,y) \leq \rho\}$ denotes the closed cell with center x and radius ρ):

For any family $\{S(x_\alpha;\rho_\alpha); \alpha \in A\}$ of pairwise intersecting cells in X,

$$\bigcap_{\alpha \in A} S(x_\alpha; \mu\rho_\alpha) \neq \phi .$$

If for every such family $\bigcap_{\alpha \in A} S(x_\alpha; E(X)\rho_\alpha) \neq \phi$, $E(X)$ is called *exact*.

The expansion constants of Minkowski spaces have been studied in [5]. In the present paper we deal (in § 2) with an application of the expansion constants to a problem on projections in Banach spaces; as corollaries we obtain Nachbin's [10] geometric characterization of Banach spaces with the Hahn-Banach extension property (§ 2) and Bohnenblust's [3] result on projections in Minkowski spaces, as well as some results which we believe to be new (§ 4). In § 3 we discuss the relation of expansion constants to a property of retractions in metric spaces, especially those convex in Menger's sense; as a corollary we obtain Aronszajn-Panitchpakdi's [2] characterization of spaces with the unlimited uniform extension property. Section 4 contains additional remarks and examples.

2. In order to apply expansion constants to projections in Banach spaces, it is convenient to introduce the notion of projection constants.

DEFINITION 1. For any normed space X the *projection constant* $p(X)$ is the greatest lower bound of real numbers μ which possess the following property: For any normed space Y which contains X as a subspace of deficiency 1, there exists a projection P of Y onto X such that $\|P\| \leq \mu$. If for any such Y there exists a projection of norm less than or equal to $p(X)$, the projection constant $p(X)$ is called *exact*.

(The projection constant $p(X)$ should not be confused with the projection constant $\mathscr{P}(X)$ studied in [6].)

We show now that if X is a normed space then $E(X)$ actually coincides with $p(X)$.

Received April 6, 1959. Part of the results of this paper are taken from § 6 of the author's Ph. D. thesis, "On some properties of Minkowski spaces," (in Hebrew), prepared under the guidance of Prof. A. Dvoretzky at The Hebrew University in Jerusalem. The remaining results have been obtained under contract AF 61 (052)-04.

THEOREM 1. *For any normed space X we have $p(X) = E(X)$; moreover, if one of the constants is exact, so is the other.*

Proof. If X is not complete, then $p(X) = \infty = E(X)$. The first part follows immediately from the remark that, if Y is any subspace of the completion of X containing X as a subspace of deficiency 1, there exists no projection of Y onto X. On the other hand, $E(X) = \infty$ for any metric space X which is not complete. Indeed, if $\{x_n; n = 1, 2, \cdots\}$ is a Cauchy sequence in X which is not convergent, let $\rho_n = 2\lim_{k\to\infty} d(x_n, x_k)$, for $n = 1, 2, \cdots$. Then the cells $\{S(x_n; \rho_n); n = 1, 2, \cdots\}$ are mutually intersecting, but $\bigcap_{n=1}^{\infty} S(x_n; \mu\rho_n) = \phi$ for any μ, which implies $E(X) = \infty$. (We shall see in § 4 that if X is a complete metric space then $E(X) \leq 2$.)

Thus we may assume that X is complete. We shall first prove that $E(X) \leq p(X)$. To that effect, let Y be the linear sum of X (considered as a vector space) and a point $y_0 \notin X$. For any given family $\{S(x_\alpha; \rho_\alpha); \alpha \in A\}$ of mutually intersecting cells (of X) we shall define a norm in Y, such that Y becomes a Banach space containing X as a subspace, and that for any projection P of Y onto X we have:

$$\bigcap_{\alpha \in A} S(x_\alpha; \| P \| \cdot \rho_\alpha) \neq \phi\,. \tag{2.1}$$

If $\inf \rho_\alpha = 0$ any norm on Y establishes (2.1) since $\bigcap_{\alpha\in A} S(x_\alpha; \rho_\alpha) \neq \phi$. We prove this relation in the following way: If for some $\beta \in A$ we have $\rho_\beta = 0$, then $x_\beta \in S(x_\alpha; \rho_\alpha)$ for all $\alpha \in A$. On the other hand, if for a sequence of indices $\alpha_n \in A$ we have $\lim \rho_{\alpha_n} = 0$, then (since $d(x_\alpha, x_\beta) \leq \rho_\alpha + \rho_\beta$) $\{x_{\alpha_n}\}$ is a Cauchy sequence. Since X is complete there exists $x_0 = \lim_n x_{\alpha_n}$. We claim that $x_0 \in \bigcap_{\alpha\in A} S(x_\alpha, \rho_\alpha)$. Indeed, for any $\alpha \in A$ and any $\varepsilon > 0$, let n be such that $\rho_{\alpha_n} < 1/2\varepsilon$ and $d(x_{\alpha_n}, x_0) < 1/2\,\varepsilon$; then $d(x_\alpha, x_0) \leq d(x_\alpha, x_{\alpha_n}) + d(x_{\alpha_n}, x_0) < \rho_\alpha + \varepsilon$. Since ε is arbitrary, it follows that $x_0 \in S(x_\alpha; \rho_\alpha)$ for any $\alpha \in A$, as required.

Thus we are left with the case $\inf \rho_\alpha > 0$. Let $y_\alpha = (x_\alpha + y_0)/\rho_\alpha$ for each $\alpha \in A$ and $K = \{y_\alpha; \alpha \in A\}$. If S is the unit cell of X, we denote by T the closure (in the product topology of $Y = X \times Ry_0$) of the convex hull of the set $S \cup K \cup (-K) \subset Y$. Since T is a centrally symmetric convex body in Y it defines a norm (according to which T is the unit cell). We claim that $T \cap X = S$, i.e. that X (as a Banach space) is a subspace of Y. Obviously, this will be established if we show that the intersection of X with a segment connecting a point of K with a point of $-K$ belongs to S. Now, an elementary computation shows that $[y_\alpha, -y_\beta] \cap X$ is the point $(x_\alpha - x_\beta)/(\rho_\alpha + \rho_\beta)$ whose norm is less than or equal to 1, since $\| x_\alpha - x_\beta \| \leq \rho_\alpha + \rho_\beta$ is a consequence of the assumption that the members of $\{S(x_\alpha; \rho_\alpha)\}$ are pairwise intersecting. Now let P be any projection of Y onto X and let $x_0 = -P(y_0)$. Then

$$P(y_\alpha) = P\left(\frac{x_\alpha + y_0}{\rho_\alpha}\right) = \frac{x_\alpha - x_0}{\rho_\alpha} \text{ and therefore}$$

$$\left|\left|\frac{x_\alpha - x_0}{\rho_\alpha}\right|\right| \leq ||P|| \cdot ||y_\alpha|| \leq ||P|| \text{ for each } \alpha \in A .$$

In other words, $x_0 \in S(x_\alpha; ||P|| \rho_\alpha)$ for each $\alpha \in A$ which implies (2.1) and thus establishes $E(X) \leq p(X)$.

In order to derive the converse inequality $p(X) \leq E(X)$ let Y be any Banach space containing X as a maximal subspace, and let $y_0 \in Y, y_0 \notin X$. The triangle inequality implies that $S_X(x, ||x - y_0||) \cap S_X(x', ||x' - y_0||) \neq \phi$ for any $x, x' \in X$. Let μ be such that $\bigcap_{x \in X} S_X(x; \mu ||x - y_0|| \neq \phi$, and denote by x_0 any point of that intersection. Thus $||x - x_0|| \leq \mu ||x - y_0||$ for any $x \in X$. We define a projection P of Y onto X by $P(x + \lambda y_0) = x + \lambda x_0$, and we shall show that $||P(x + \lambda y_0)|| \leq \mu ||x + \lambda y_0||$, i.e., that $||P|| \leq \mu$. Obviously, we may assume $\lambda \neq 0$ and then, by the definition of x_0, we have $||P(x + \lambda y_0)|| = ||x + \lambda x_0|| = |\lambda| \cdot ||-x/\lambda - x_0|| \leq \mu |\lambda| \cdot ||-x/\lambda - y_0|| = \mu ||x + \lambda y_0||$.

This completes the proof of our last assertion, and thus also the proof of Theorem 1.

The connection between projection constants and extensions of linear transformations may be found using the following lemma:

If X and Y are any normed spaces, if Z contains Y as a subspace of deficiency 1, *and if f is any linear transformation from Y to X, then there exist a normed space W and a linear transformation F from Z to W such that:*

(i) *W contains X as a subspace of deficiency* 1;

(ii) *F coincides with f on Y;*

(iii) $||F|| = ||f||$.

We omit the simple proof of this lemma since a more general extension theorem of this type has been proved by Sobczyk [13, Theorem 4.1].

Using the above lemma, the following corollary results immediately from Theorem 1:

For any Banach spaces X, Y and Z, with Y a maximal subspace of Z, any linear transformation f from Y to X, and any $\mu > p(X) = E(X)$, there exists a linear transformation F from Z to X, coinciding with f on Y, such that $||F|| \leq ||\mu|| \, ||f||$; if $p(X)$ is exact, there exists such an F even for $\mu = p(X)$.

A standard application of Zorn's lemma or transfinite induction yields therefore:

The following two properties of a normed space X are equivalent:

(i) *$E(X) = 1$ and is exact;*

(ii) *For any normed spaces Y and Z, with $Y \subset Z$, and any linear* *transformation f from Y to X, there exists a linear transformation F from Z to X such that F coincides with f on Y and $\|F\| = \|f\|$.*

Since the "binary intersection property" of Nachbin [10] is equivalent to "$E(X)=1$ and is exact," the last statement is precisely Nachbin's characterization of spaces with the Hahn-Banach extension property [10, Theorem 1].

3. In the case of metric spaces, expansion constants may be used to obtain information on retraction properties, in close analogy to the procedure applied in § 2 to projections in normed spaces.

A *retraction* r of a metric space Y onto a metric space $X \subset Y$ is a (continuous) mapping of Y onto X such that $r(x) = x$ for each $x \in X$.

DEFINITION 2. The norm $\|r\|$ of a retraction r of Y onto $X \subset Y$ is the greatest lower bound of numbers μ such that $d(r(y_1), r(y_2)) \leq \mu d(y_1, y_2)$ for all $y_1, y_2 \in Y$. The *retraction constant* $r(X)$ of a metric space X is the greatest lower bound of numbers μ with the property: For any metrix space Y which contains X any only one point not in X, there exists a retraction of Y onto X with norm less than or equal to μ. If $r(X) = \min \mu$, the retraction constant $r(X)$ is called *exact*.

Obviously $r(X) = \infty$ if X is not complete, and it is easily shown that for complete spaces $r(X) \leq 2$.

Since metrically convex spaces have special properties with respect to retractions, we recall their definition (essentially that of Menger [9]):

A metric space X is called *metrically convex* if for any pair of points $x', x'' \in X$ and any λ, $0 < \lambda < 1$, there exists a point $y \in X$ such that $d(x', y) = \lambda d(x', x'')$ and $d(x'', y) = (1 - \lambda)d(x', x'')$.

In analogy to Theorem 1 we have:

THEOREM 2. (i) *For any metric space X, $E(X) \leq r(X)$.*

(ii) *For any metrically convex metric space X, $E(X) = r(X)$; moreover, if one of the constants is exact, so is the other.*

Proof. (i) Since for uncomplete spaces both constants are infinite, we will assume that X is complete. Let $\{S(x_\alpha; \rho_\alpha); \alpha \in A\}$ be any family of mutually intersecting cells in X. We shall define a space $Y = X \cup \{y_0\}$ (with distance function D) such that for any retraction r of Y onto X we have

$$(3.1) \qquad \bigcap_{\alpha \in A} S(x_\alpha; \|r\| \cdot \rho_\alpha) \neq \phi \; .$$

This will prove part (i) of the Theorem. Now, if $\inf \rho_\alpha = 0$ the reasoning used in the proof of Theorem 1 shows that any metrization of Y is appropriate. Thus there remains the case $\inf \rho_\alpha > 0$. Then, let

$D(x', x'') = d(x', x'')$ for all $x', x'' \in X$, and let $D(x, y_0)$ be the greatest lower bound of those numbers μ for which $S(x; \mu)$ contains $S(x_\alpha; \rho_\alpha)$ for some $\alpha \in A$. (This metric on Y was used also in [2]). Since $D(x, y_0) > 0$ follows obviously from $\inf \rho_\alpha > 0$, in order to establish that D is indeed a distance function on Y we have only to prove the triangle inequality for triples of points containing y_0, i.e. the relations

$$D(x', x'') \leq D(x', y_0) + D(y_0, x'') \tag{3.2}$$

and

$$D(x', y_0) \leq D(x', x'') + D(x'', y_0) \tag{3.3}$$

for all $x', x'' \in X$. To that effect let $\varepsilon > 0$ be given; then $S(x'; D(x', y_0) + \varepsilon) \supset S(x_\alpha; \rho_{\alpha'})$ and $S(x''; D(x'', y_0) + \varepsilon) \supset S(x_{\alpha''}; \rho_{\alpha''})$ for suitable $\alpha', \alpha'' \in A$. Since any two of the cells $S(x_\alpha; \rho_\alpha)$ have common points, there exists a $z \in X$ such that $d(x', z) \leq D(x', y_0) + \varepsilon$ and $d(x'', z) \leq D(x'', y_0) + \varepsilon$. Then $D(x', x'') = d(x', x'') \leq d(x', z) + d(x'', z) \leq D(x', y_0) + D(x'', y_0) + 2\varepsilon$. Since ε was arbitrary, (3.2) results. On the other hand, since $S(x'; d(x', x'') + D(x'', y_0) + \varepsilon) \supset S(x''; D(x'', y_0) + \varepsilon) \supset S(x_{\alpha''}; \rho_{\alpha''})$, we have $D(x', y_0) \leq d(x', x'') + D(x'', y_0) + \varepsilon$ for any $\varepsilon > 0$, which establishes (3.3).

Now, let r be a retraction of Y onto X, and let $x_0 = r(y_0)$. Then, for any $\alpha \in A$ we have $d(x_\alpha, x_0) \leq \| r \| D(x_\alpha, y_0) \leq \| r \| \cdot \rho_\alpha$, which is equivalent to $x_0 \in \bigcap_{\alpha \in A} S(x_\alpha; \| r \| \rho_\alpha)$. Thus (3.1) holds, and the proof of (i) is complete.

The proof of (ii) is now easy. Let $Y = X \cup \{y_0\}$ and let D be the distance function of Y. We consider the family of cells in X defined by $\{S(x; D(x, y_0)); x \in X\}$. The triangle inequality which is satisfied by D, and the metric convexity of X imply that these cells are mutually intersecting. Let μ be a number such that $\bigcap_{x \in X} S(x; \mu D(x, y_0)) \neq \phi$, and let x_0 be any point of this intersection. Then the retraction r of Y onto X defined by $r(y_0) = x_0$ obviously satisfies $\| r \| \leq \mu$. This completes the proof of Theorem 2.

REMARKS. (i) If X is not metrically convex, $E(X) < r(X)$ is possible. The simplest example to this effect is that of a space X consisting of precisely two points. Then $E(X) = 1$ and $r(X) = 2$.

(ii) Let X, Y and $Z = Y \cup \{z_0\}$ be any metric spaces, and f a uniformly continuous transformation from Y to X, with subadditive modulus of continuity $\delta(\varepsilon)$ (see, e.g., [2]). It is easily established that there exists a metric space $X^* = X \cup \{x^*\}$, whose distance function coincides on X with the distance function of X, such that there exists an extension F of f, with domain Z and range in X^*, which is uniformly continuous with the modulus $\delta(\varepsilon)$. Therefore, using transfinite induction or Zorn's lemma, we obtain the following corollary of Theorem 2, which is equivalent to Theorem 2, § 3 of [2]:

For any metrically convex metric space X the following properties are equivalent:

(a) $r(X) = 1$ *and is exact;*

(b) *For any metric spaces Y and Z, with $Z \supset Y$, and any uniformly continuous transformation f from Y to X with subadditive modulus of continuity $\delta(\varepsilon)$, there exists a uniformly continuous transformation F from Z to X, coinciding with f on Y and having $\delta(\varepsilon)$ as modulus of continuity.*

4. Some properties of expansion constants $E(X)$ for finite dimensional Banach spaces X have been established in [5]. As a consequence of Theorem 1 of the present paper, these results yield the following information on the projection constants $p(X)$:

(i) If X is an n-dimensional Minkowski space then $1 \leq p(X) \leq 2n/(n+1)$. (This was first established by Bohnenblust [3].)

(ii) If E^n denotes the n-dimensional Euclidean space then $p(E^n) = \sqrt{2n/(n+1)}$.

The characterization of those n-dimensional Minkowski spaces X for which $E(X) = 2n/(n+1)$, given in Theorem 2 of [5], yields immediately a characterization of spaces X for which the upper bound is attained in (i).

As observed by Bohnenblust [3], $p(X) \leq 2$ for any Banach space X. By Theorem 1 this is a corollary of the following more general proposition:

$E(X) \leq 2$ *for any complete metric space X.*

Proof. Let $\{S(x_\alpha; \rho_\alpha); \alpha \in A\}$ be any family of mutually intersecting cells in X. The reasoning used in the proof of Theorem 1 shows that if $\inf \rho_\alpha = 0$ then $\bigcap_{\alpha \in A} S(x_\alpha; \rho_\alpha) \neq \phi$. Thus we may assume $\inf \rho_\alpha = \rho_0 > 0$. Given any $\varepsilon > 0$ let $\beta \in A$ be such that $\rho_\beta \leq (1+\varepsilon)\rho_0$. Since then $d(x_\alpha, x_\beta) \leq \rho_\alpha + \rho_\beta \leq \rho_\alpha + (1+\varepsilon)\rho_0 \leq (2+\varepsilon)\rho_\alpha$, we have $x_\beta \in \bigcap_{\alpha \in A} S(x_\alpha; (2+\varepsilon)\rho_\alpha)$, which proves our assertion.

The notion of expansion constants gives us a convenient method of obtaining information on the exactness of projection and retraction constants.

Definition 3. A metric space X is said to have the *finite intersection property* if each family of cells $\{S(x_\alpha; \rho_\alpha); \alpha \in A\}$ of X satisfies the condition: Whenever every finite subfamily has a non-void intersection, then $\bigcap_{\alpha \in A} S(x_\alpha; \rho_\alpha) \neq \phi$.

Obviously, compact spaces and spaces with compact cells have the finite intersection property. As a consequence of the w^*-compactness of the unit cell of any adjoint Banach space ([1], [4]), adjoint Banach spaces (and thus especially reflexive, unitary, or finite dimensional Banach spaces) also have the finite intersection property.

For spaces with the finite intersection property we have:

THEOREM 3. *If X has the finite intersection property then $E(X)$, and therefore $r(X)$ (and $p(X)$ if X is a Banach space), are exact.*

Proof. Let $\{S(x_\alpha; \rho_\alpha); \alpha \in A\}$ be any family of mutually intersecting cells in X. By the definition of the expansion constant, any finite subfamily of the family $\{S_{\alpha,\varepsilon} = S(x_\alpha; (E(X) + \varepsilon)\rho_\alpha); \alpha \in A, \varepsilon > 0\}$ has a non-void intersection. Since X has the finite intersection property this implies that $\bigcap_{\alpha \in A, \varepsilon > 0} S_{\alpha,\varepsilon} \neq \phi$. But $S(x_\alpha; E(X) \cdot \rho_\alpha) = \bigcap_{\varepsilon > 0} S_{\alpha,\varepsilon}$ for each $\alpha \in A$ and therefore $\bigcap_{\alpha \in A} S(x_\alpha; E(X)\rho_\alpha) \neq \phi$ as claimed.

REMARK. We know of no Banach space which has the finite intersection property but is not an adjoint space; indeed, it seems reasonable to conjecture that only adjoint spaces have this property. On the other hand, a wider class of Banach spaces has exact projection and expansion constants. E.g., it is well known (Sobczyk [12]) that $p(c_0) = 2$ and is exact (it is not difficult to show directly that $E(c_0) = 2$ and is exact) but it is easily seen that c_0 does not have the finite intersection property (not even for families of cells having the same radius).

Another question, raised by Bohnenblust [3], is whether there exists a projection of norm ≤ 2 from any Banach space onto each of its maximal closed subspaces. A negative answer to Bohnenblust's problem follows immediately from the following example.

EXAMPLE. *Let X be the subspace of l defined by*

$$X = \left\{x = (x_1, x_2, \cdots) \in l; \sum_{n=1}^{\infty} \frac{n}{n+1} x_n = 0\right\}.$$

Then $E(X) = 2$ but $E(X)$ is not exact.

For reasons of convenience we shall, instead of X, consider its translate $H = \{x; \sum_{n=1}^{\infty} n/(n+1)x_n = 1\} \subset l$. Since X and H are isometric metric spaces, this is permissible. Now let $\{e_n; n = 1, 2, \cdots\}$ denote the usual basis of l, and S its unit cell $S = \{x \in l; \|x\| = \sum_{n=1}^{\infty} |x_n| \leq 1\}$. Obviously $n + 1/(n)e_n \in H$ for $n = 1, 2, \cdots$, but $S \cap H = \phi$ although $\operatorname{dist}(S, H) = 0$. (This last property was Klee's reason for introducing H in [8].) We consider in l the family of cells

$$\left\{S_n = S\left(\frac{n+1}{n} e_n; \frac{n+1}{n}\right) = \left\{x \in l; \left\| x - \frac{n+1}{n} e_n \right\| \leq \frac{n+1}{n}\right\}; n = 1, 2, \cdots .\right\}.$$

Then $S_n^* = S_n \cap H$ is a family of cells in H which are mutually intersecting since

$$\frac{(n+1)(k+1)}{n(k+1)+k(n+1)}(e_n + e_k) \in S_n^* \cap S_k^* .$$

We shall show that

$$H \cap \left[\bigcap_{n=1}^{\infty} S\left(\frac{n+1}{n} e_n ; 2\frac{n+1}{2}\right)\right] = \phi ,$$

which will obviously prove our assertion. Let $K = \bigcap_{n=1}^{\infty} S(1 + n^{-1})e_n ; 2(n+1)/n)$; then, since $H \cap S = \phi$, it is sufficient to prove that $K \subset S$. Assuming that there exists an $x \in K$ such that $\|x\| = 1 + \varepsilon > 1$, we have by the definition of K:

$$0 \leq \left\| x - \frac{n+1}{n} e_n \right\| = \left| \frac{n+1}{n} - x_n \right| + \sum_{i \neq n} |x_i| \leq 2\frac{n+1}{n} .$$

for each n. Now, if for some n we have $x_n \geq 0$, it follows that either $x_n > (n+1)/n > 1$ or $-x_n + \sum_{i \neq n} |x_i| \leq (n+1)/n$. Both cases are possible only for a finite number of indices n; in the first case this is obvious, while in the second it follows from the fact that it implies $\|x\| - (n+1)/n \leq 2x_n$, i.e. $\varepsilon \leq (1/n) + 2x_n$. On the other hand, for those n for which $x_n \leq 0$ we have

$$\frac{n+1}{n} - x_n + \sum_{i \neq n} |x_i| \leq 2\frac{n+1}{n} ,$$

or

$$\|x\| \leq \frac{n+1}{n}, \text{ i.e. } \varepsilon \leq \frac{1}{n} ,$$

which is again possible only for a finite number of indices n. Thus $K \subset S$, which completes the proof.

REMARKS. (i) Since adjoint Banach spaces have exact expansion constants, the space X of the above example is not an adjoint space, although it is a maximal closed subspace of an adjoint space. It would be interesting to know whether every non-reflexive Banach space has a closed maximal subspace which is not an adjoint space.

(ii) Jung's constant $J(X)$ has been defined [5] in the same way as $E(X)$, with the additional condition that all the radii ρ_α be equal. The space X of the last example shows that it is possible to have $J(X) = E(X)$ with $J(X)$ exact and $E(X)$ not exact.

(iii) Theorem 4 of § 3 of [2] implies:

If X is a bounded, metrically convex metric space and $E(X) = 1$, then $E(X)$ is exact.

Although the condition of boundedness is perhaps redundant, the following example shows that it is impossible to drop the condition that X be metrically convex.

EXAMPLE. *Let $X = \{x_n; n = 1, 2, \cdots\}$ with $d(x_n, x_x) = 1 + 1/n + 1/k$ for $n \neq k$. Then $E(X) = 1$ but $E(X)$ is not exact.*

Indeed, it is easily verified that $E(X) = 1$. On the other hand, the cells $S_n = S(x_n, 1 + 2/n) = \{x_k; k \geq n\}$ for $n = 1, 2, \cdots$, are not only mutually intersecting, but we even have $S_k \supset S_n$ for $k \leq n$. But obviously $\bigcap_{n=1}^{\infty} S_n = \phi$. (Complete metric spaces containing descending sequences of cells with empty intersections have been considered by Sierpiński [11]; see also Harrop-Weston [7].)

REFERENCES

1. L. Alaoglu, *Weak topologies of normed spaces*, Ann. of Math., (2) **41** (1940), 252-267.
2. N. Aronszajn, and P. Panitchpakdi, *Extension of uniformly continuous transformations and hyperconvex metric spaces*, Pacific J. Math., **6** (1956), 405-439.
3. F. Bohnenblust, *Convex regions and projections in Minkowski spaces*, Ann. of Math., (2) **39** (1938), 301-308.
4. N. Bourbaki, *Sur les espaces de Banach*, C. R. Paris., **206** (1938), 1701-1704.
5. B. Grünbaum, *On some covering and intersection properties in Minkowski spaces*, Pacific J. Math., **9** (1959), 487-494.
6. ———, *Projection constants* (submitted to Trans. Amer. Math. Soc.).
7. R. Harrop, and J. D. Weston, *An intersection property in locally convex spaces*, Proc. Amer. Math. Soc., **7** (1956), 535-538.
8. V. L., Klee, Jr. *The support property of a convex set in a linear normed space*, Duke Math. J., **15** (1948), 767-772.
9. K. Menger, *Untersuchungen über allgemeine Metrik*, Math. Ann., **100** (1928), 75-163.
10. L. Nachbin, *A theorem of the Hahn-Banach type for linear transformations*, Trans. Amer. Math. Soc., **68** (1950), 28-46.
11. W. Sierpiński, *Sur une définition des espaces complets*, Ganita, **1** (1948), 13-16.
12. A. Sobczyk, *Projections of the space (m) on its subspace (c_0)*, Bull. Amer. Math. Soc., **47** (1941), 938-947.
13. ———, *On the extension of linear transformations*, Trans. Amer. Math. Soc., **55** (1944), 153-169.

THE INSTITUTE FOR ADVANCED STUDY
PRINCETON, N. J.
AND
THE HEBREW UNIVERSITY
JERUSALEM

ERROR BOUNDS FOR AN APPROXIMATE SOLUTION TO THE VOLTERRA INTEGRAL EQUATION

JOHN HILZMAN

In 1945 Michal [2] obtained several results which he asserted were useful for approximating the solution to the Volterra integral equation. These results were concerned with certain equations in Fréchet differentials having as their unique solutions the resolvent kernel and the exact solution to the Volterra integral equation of the second kind. Michal treated the resolvent kernel $S[K \mid x, t]$ and the solution $y[K \mid x]$ as functions[1] of the given kernel $K(x, t)$, the setting being the Banach spaces

$$T = \{G(x, t) \mid G(x, t) \text{ is real and continuous on } a \leq t \leq x \leq b\}$$

and

$$I = \{g(x) \mid g(x) \text{ is real and continuous on } a \leq x \leq b\}$$

with the norms

$$\begin{aligned} &\|G(x, t)\| = \max |G(x, t)| \quad (a \leq t \leq x \leq b), \\ &\|g(x)\| \equiv \max |g(x)| \qquad (a \leq x \leq b), \end{aligned} \tag{1}$$

respectively. In another work [3, pp. 16-17] Michal showed that the solution $y[K \mid x]$ can be expressed by a Taylor-type expansion in Fréchet differentials of $y[K \mid x]$ about an arbitrary $K_0(x, t)$ from T. In this paper we shall use Michal's results to obtain approximations to the solution of the Volterra integral equation with error bounds.

I wish to thank Professor A. T. Lonseth for suggesting this course of investgation and the Referee for recommendations which have improved this paper.

Consider the integral equation

$$y(x) + \int_a^x K(x, t)y(t)dt = f(x) \tag{2}$$

where $K(x, t)$ is in T and $f(x)$ is in I. It is known that the exact solution to (2) is given by

$$y(x) = f(x) + \int_a^x S(x, t)f(t)dt \tag{3}$$

Received January 15, 1959, and in revised form March 30, 1959. This work was partially supported by the Office of Ordnance Research, U.S. Army, under contract with Oregon State College.

[1] The symbols $S[K \mid x, t]$ and $y[K \mid x]$ were used to indicate the functional dependence of $S(x, t)$ and $y(x)$ on $K(x, t)$.

where the resolvent kernel $S(x, t)$ is in T. Let $K_0(x, t)$ from T be another kernel such that $S_0(x, t)$, the resolvent of $K_0(x, t)$, is known and that $\|h(x, t)\| = \|K(x, t) - K_0(x, t)\|$ is small in the sense of (1). Then by (3) the solution to (2) with kernel $K_0(x, t)$ is

$$y_0(x) = f(x) + \int_a^x S_0(x, t)f(t)dt \ . \tag{4}$$

Now treat $y(x)$ as a function of the kernel $K(x, t)$. The first Fréchet differential $dy(x)$ of $y(x)$ with increment $h(x, t)$ (applied to $K(x, t)$) is

$$dy(x) = -\int_a^x \left[h(x, t) + \int_t^x S(x, z)h(z, t)dz \right] y(t)dt$$

[2, p. 253]. In particular, the Fréchet differential of $y(x)$ evaluated at $K_0(x, t)$ with increment $h(x, t) = K(x, t) - K_0(x, t)$ will be

$$dy_0(x) = -\int_a^x \left[h(x, t) + \int_t^x S_0(x, z)h(z, t)dz \right] y_0(t)dt \ . \tag{5}$$

Furthermore, by Theorem 2 of [2] the differential system

$$\begin{cases} dy_0(x) = -\int_a^x \left[h(x, t) + \int_t^x S_0(x, z)h(z, t)dz \right] y_0(t)dt \\ y_0(x) = f(x) \qquad (K_0(x, t) = 0) \end{cases}$$

has a unique solution which is given by (4). Thus a first order approximation to the solution $y(x)$ of (2) will be

$$y_0(x) + dy_0(x) \ .$$

The exact solution to (2) is given by the Taylor expansion [3; 1. p. 112]

$$y(x) = y_0(x) + \sum_{j=1}^{\infty} (j!)^{-1} d^j y_0(x) \tag{6}$$

where, in terms of composition powers[2],

$$d^j y_0(x) = (-1)^j j! \, [h + S_0 h]^j * y_0 \ . \tag{7}$$

Thus knowledge of the higher order differentials will allow closer approximations to $y(x)$.

We now take up the problem of establishing error bounds for any order of approximation to $y(x)$ from (6). If A_j $(j = 1, 2, \cdots, n)$ is in T and g is in I, and

---

[2] $VW = \int_t^x V(x, z)W(z, t)dz$, $W^2 = \int_t^x W(x, z)W(z, t)dz$, $W^n = \int_t^x W(x, z)W^{n-1}(z, t)dz$, and $W^n * g = \int_a^x W^n(x, t)g(t)dt$

$$A = A_1A_2 \cdots A_n = \int_t^x \int_t^{z_1} \cdots \int_t^{z_{n-2}} A_1(x, z_1)A_2(z_1, z_2) \cdots A_n(z_{n-1}, t)dz_{n-1} \cdots dz_1 \,,$$

it is seen that

$$(8) \qquad \| A \| \leq \frac{|b-a|^{n-1}}{(n-1)!} \prod_{j=1}^{n} \| A_j \|$$

and

$$(9) \qquad \| A * g \| \leq \frac{\| g \| \, |b-a|^n}{n!} \prod_{j=1}^{n} \| A_j \| \,.$$

Let $P_{n-i,i}[h(S_0h)]$ denote the sum of terms obtained from the composition $h^{n-i}(S_0h)^i$ by a permutation on the n places occupied by

$$\underbrace{hh \cdots h}_{n-i}\underbrace{(S_0h)(S_0h) \cdots (S_0h)}_{i} = h^{n-i}(S_0h)^i \,.$$

For example, by setting

$$P_{2,1}[h(S_0h)] = h^2(S_0h) + h(S_0h)h + (S_0h)h^2$$

and

$$P_{1,2}[h(S_0h)] = h(S_0h)^2 + (S_0h)h(S_0h) + (S_0h)^2h$$

we can write with brevity

$$[h + S_0h]^3 = h_3 + P_{2,1}[h(S_0h)] + P_{1,2}[h(S_0h)] + (S_0h)^3 \,.$$

Now let

$$c = \| h(x,t) \|, \; m = \| y_0(x) \|, \; B = \| S_0(x,t) \|, \text{ and } u = |b-a| \,.$$

Then from (7), (8), (9), and the mechanics of composition we obtain

$$\begin{aligned}
&\| (n!)^{-1}d^ny_0(x) \| = \| (-1)^n[h + S_0h]^n * y_0 \| \\
&= \| h^n * y_0 + P_{n-1,1}[h(S_0h)] * y_0 + \cdots + P_{1,n-1}[h(S_0h)] * y_0 + (S_0h)^n * y_0 \| \\
&\leq \| h^n * y_0 \| + \| P_{n-1,1}[h(S_0h)] * y_0 \| + \cdots + \| (S_0h)^n * y_0 \| \\
(10) \; &\leq \frac{mc^nu^n}{n!} + \binom{n}{1} \frac{mc^nu^{n+1}B}{(n+1)!} + \cdots \binom{n}{n} \frac{mc^nu^{2n}B^n}{(2n)!} \\
&\leq mc^nu^n \sum_{j=0}^{n} \binom{n}{j} \frac{(uB)^j}{(n+j)!} \\
&\leq \frac{m[cu(1+uB)]^n}{n!} \,.
\end{aligned}$$

Thus transposing the desired nth order approximation to $y(x)$ from the right side of (6) to the left side and applying (10) we get

$$\begin{aligned}\left\|y(x) - y_0(x) - \sum_{j=1}^{n} \frac{1}{j!} d^j y_0(x)\right\| &= \left\|\sum_{j=n+1}^{\infty} \frac{1}{j!} d^j y_0(x)\right\| \\ &\leq \sum_{j=n+1}^{\infty} m(j!)^{-1}\theta^j \\ &\leq m\left[e^\theta - \sum_{j=0}^{n} (j!)^{-1}\theta^j\right]\end{aligned} \tag{11}$$

where $\theta = cu[1 + uB]$, For small values of θ we readily discern the asymptotic relation

$$\left\|y(x) - y_0(x) - \sum_{j=1}^{n-1} \frac{1}{j!} d^j y_0(x)\right\| = 0(\theta^n)\ . \tag{12}$$

A simple numerical example will be given next.

Consider the Volterra equation

$$y(x) + \frac{1}{3}\int_0^x xt[3 + x^3 - t^3]y(t)dt = x \exp[1/3x^3] \tag{13}$$

where $K(x, t) = 1/3\, xt[3 + x^3 - t^3]$ is in T, $f(x) = x \exp[1/3x^3]$ is in I and $a = 0, b = 1$. Take $K_0(x, t) = xt \exp[1/3(x^3 - t^3)]$. The resolvent kernel for $K_0(x, t)$ is $S_0(x, t) = -xt$. By (4) the solution to (13) with kernel $K_0(x, t)$ is

$$y_0(x) = x \exp[1/3x^3] + \int_0^x - xt^2 \exp[1/3t^3]dt = x\ . \tag{14}$$

By virtue of (5), the Fréchet differential of $y(x)$ evaluated at $K_0(x, t)$ with increment

$$h(x, t) = K(x, t) - K_0(x, t) = \frac{1}{3} xt[3 + x^3 - t^3] - xt \exp[1/3(x^3 - t^3)]$$

is

$$\begin{aligned} dy_0(x) = &- \int_0^x \left\{\frac{1}{3} xt(3 + x^3 - t^3 - 3) \exp[1/3(x^3 - t^3)]\right. \\ &\left. + \int_t^x - xz\left(\frac{1}{3} zt(3 + z^3 - t^3 - 3) \exp[1/3(z^3 - t^3)]\right)dz\right\}tdt \\ = &\ \frac{x^{10}}{162}\ . \end{aligned} \tag{15}$$

Thus a first order approximation to $y(x)$ will be

$$y(x) \approx x + \frac{x^{10}}{162} \, . \tag{16}$$

It is easily established that

$$\| h(x, t) \| < 0.04, \; \| S_0(x, t) \| = 1, \; \| y_0(x) \| = 1 \, .$$

Hence, with $\theta = 0.08$, it follows from (11) that

$$\| y(x) - y_0(x) - dy_0(x) \| < 0.0033 \, . \tag{17}$$

References

1. E. Hille and R. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc. Coll. Publ. **31** (1957).

2. A. D. Michal, *Differential equations in Frechet differentials occurring in integral equations*, Proc. Nat. Acad. Sci., **31** (1945), 252-258.

3.———, *On a non-linear total differential equation in normed linear spaces* Acta Mathematica, **80** (1948), 1-21.

HARPUR COLLEGE

THE FRATTINI SUBGROUP OF A p-GROUP

CHARLES HOBBY

The Frattini subgroup $\Phi(G)$ of a group G is defined as the intersection of all maximal subgroups of G. It is well known that some groups cannot be the Frattini subgroup of any group. Gaschütz [3, Satz 11] has given a necessary condition for a group H to be the Frattini subgroup of a group G in terms of the automorphism group of H. We shall show that two theorems of Burnside [2] limiting the groups which can be the derived group of a p-group have analogues that limit the groups which can be Frattini subgroups of p-groups.

We first state the two theorems of Burnside.

THEOREM A. *A non-abelian group whose center is cyclic cannot be the derived group of a p-group.*

THEOREM B. *A non-abelian group, the index of whose derived group is p^2, cannot be the derived group of a p-group.*

We shall prove the following analogues of the theorems of Burnside.

THEOREM 1. *If H is a non-abelian group whose center is cyclic, then H cannot be the Frattini subgroup $\Phi(G)$ of any p-group G.*

THEOREM 2. *A non-abelian group H, the index of whose derived group is p^2, cannot be the Frattini subgroup $\Phi(G)$ of any p-group G.*

We shall require four lemmas, the first two of which are due to Blackburn and Gaschütz, respectively.

LEMMA 1. [1, Lemma 1] *If N is a normal subgroup of the p-group G such that the order of N is p^2, then the centralizer of N in G has index at most p in G.*

LEMMA 2. [3, Satz 2] *If $H = \Phi(G)$ for a p-group G and N is a subgroup of H that is normal in G, then $\Phi(G/N) = \Phi(G)/N$.*

LEMMA 3. *If $N = \{a\} \times \{b\}$ is a subgroup of order p^3 normal in the p-group G such that N is contained in $\Phi(G)$, and if $\{a\}$ is a group of order p^2 in the center of $\Phi(G)$, then N is in the center of $\Phi(G)$.*

Proof. N normal in G implies that N contains a group C of order p which is in the center of G. If C is not contained in $\{a\}$ the proof

Received March 18, 1959.

is trivial, hence we may assume $C = \{a^p\}$. Since an element of order p in a p-group cannot be conjugate to a power of itself the possible conjugates of b under G are

$$b, ba^p, \cdots, ba^{(p-1)p} .$$

The index of the centralizer of b in G is equal to the number of conjugates of b under G, hence is at most p. Thus b is in the center of $\Phi(G)$, and the lemma follows.

LEMMA 4. *If H is a non-abelian group of order p^3 then there is no p-group G such that $\Phi(G) = H$.*

Proof. If $H = \Phi(G)$ for a p-group G, then H is normal in G and must contain a group N of index p which is also normal in G. Then N is a group of order p^2, hence (Lemma 1) the centralizer C of N in G has index at most p in G. Therefore C contains H, and N is in the center of H. Since the center of H has order p this is a contradiction, and the lemma follows.

We can now prove Theorems 1 and 2.

Proof of Theorem 1. We proceed by induction on the order of H. The theorem is true if H has order p^3 (Lemma 4). Suppose H is group of minimal order for which the theorem is false, and let C of a subgroup of H of order p which is contained in the center of G. Then (Lemma 2)

$$\Phi(G/C) = \Phi(G)/C = H/C .$$

Thus the induction hypothesis implies that H/C cannot be a non-abelian group with cyclic center. We consider two cases: H/C is abelian; or, the center of H/C is non-cyclic.

Case 1. Suppose H/C is abelian. Since H is not abelian, and C has order p, we conclude that C is the derived group of H. Thus H/C, which coincides with its center, is not cyclic, and we are in Case 2.

Case 2. Suppose that the center Z of H/C is non-cyclic. The elements of order p in Z form a characteristic subgroup P of Z. Since Z is not cyclic, P is also not cyclic and hence has order at least p^2. Thus we can find subgroups $\bar{M}$ and $\bar{N}$ of P which are normal in G/C and have orders p and p^2, respectively, where $\bar{M}$ is contained in $\bar{N}$. Let M and N be the subgroups of G which map on $\bar{M}$ and $\bar{N}$. Then M and N are subgroups of H which contain C and are normal in G; M and N have orders p^2 and p^3, respectively, and M is contained in N.

We see from Lemma 1 that the centralizer of M in G has index at most p in G, hence M is in the center of H, which is cyclic. Also, N is abelian since N is contained in H and the index of M in N is p. Now $\bar{N}$ is contained in P, hence is not cyclic. Therefore N is a non-cyclic group which (Lemma 3) is in the center of H. Since the center of H is cyclic this is a contradiction, and the proof is complete.

Proof of Theorem 2. We denote the derived group of a group K by K'. Suppose G is a p-group such that $\Phi(G) = H$ where $H' \neq \{1\}$ and $(H: H') = p^2$. Let N be a normal subgroup of G which has index p in H'. Then G/N is a p-group such that (Lemma 2)

$$\Phi(G/N) = \Phi(G)/N = H/N .$$

But $(H/N)' = H'/N \neq \{1\}$, and the order of H/N is

$$(H: N) = (H: H')(H': N) = p^3 .$$

Thus H/N is a non-abelian group of order p^3 which is the Frattini subgroup of the p-group G/N. This is impossible (Lemma 4) and the theorem follows.

REMARK 1. The only properties of the Frattini subgroup used in the proof of Theorems 1 and 2 are the following: $\Phi(G)$ is a characteristic subgroup of G which is contained in every subgroup of index p in G; and, $\Phi(G/N) = \Phi(G)/N$ whenever N is normal in G and contained in $\Phi(G)$. Thus if we have a rule ψ which assigns a unique subgroup $\psi(G)$ to every p-group G, then Theorems 1 and 2 will hold after replacing "the Frattini subgroup $\Phi(G)$" by "the subgroup $\psi(G)$" if $\psi(G)$ satisfies the following conditions.

(1) $\psi(G)$ is a characteristic subgroup of G.

(2) $\psi(G)$ is contained in $\Phi(G)$.

(3) $\psi(G/N) = \psi(G)/N$ if N is normal in G and N is contained in $\psi(G)$.

In particular, if $\psi(G) = G'$, the derived group of G, we have the theorems of Burnside. The proofs are unchanged.

REMARK 2. Blackburn [1] has used Theorem A to characterize the groups having two generators which are the derived group of a p-group. Using Theorem 1 it is easy to see that Blackburn's proof establishes the following

THEOREM 3. *If $H = \Phi(G)$ for a p-group G and if H has at most two generators, then H contains a cyclic normal subgroup N such that H/N is cyclic.*

References

1. N. Blackburn, *On prime-power groups*, Proc. Camb. Phil. Soc. **53** (1957), 19-27.
2. W. Burnside, *On some properties of groups whose orders are powers of primes* Proc. Lond. Math. Soc. (2) **11** (1912), 225-45.
3. W. Gaschütz, *Über die Φ-Untergruppe endlicher Gruppen.* Math. Z. **58** (1953) 160–170.

California Institute of Technology

VON NEUMANN DIFFERENCE APPROXIMATION TO HYPERBOLIC EQUATIONS

MILTON LEES

1. Introduction.

Consider the finite difference equation

$$v_{t\bar{t}} = v_{x\bar{x}} + \alpha k^2 v_{x\bar{x}t\bar{t}} \tag{1.1}$$

(1.[1])

devised by von Neumann for the numerical solution of the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}\,. \tag{1.2}$$

An essential property of any finite difference approximation to a differential equation is the convergence of solutions of the difference equation to the solution of the corresponding differential equation as $h, k \to 0$. It turns out that a sufficient condition for a difference approximation to be convergent is that it be consistent and stable (cf. Lax and Richtmyer [3]). Roughly speaking, a difference approximation to a differential equation is *consistent* when the difference equation converges to the differential equation, and it is *stable* when the solutions of the difference equation can be estimated (in a suitable norm) in terms of the prescribed data. A finite difference approximation to a hyperbolic differential equation is said to be *unconditionally stable* when it is stable for *all* positive values of the mesh ratio $R = k/h$; it is said to be *conditionally stable* when it is stable for some values of R, but not unconditionally stable.

O'Brien, Hyman and Kaplan [5] determined completely the stability properties of the difference equation (1.1). They showed that (1.1) is unconditionally stable when $\alpha \geq 1/4$, and conditionally stable when $\alpha < 1/4$ (the mesh ratio limitation being $R \leq (1 - 4\alpha)^{-\frac{1}{2}}$). This reduces to a classical result of Courant, Friedrichs and Lewy [1] when $\alpha = 0$.

It is well known that the *implicit backward* difference approximation

$$v_{\bar{t}\bar{t}} = v_{x\bar{x}} \tag{1.3}$$

to the wave equation (1.2) is unconditionally stable (cf. [5]). However, the unconditionally stable approximation (1.3) involves an error of approximation which is of order $k + k^2$, while the conditionally stable approximation

Received October 20, 1958, and in revised form March 18, 1959. Work on this paper was supported by National Science Foundation Grant G-5863.

[1] See section 2 for notation.

$$v_{t\bar{t}} = v_{x\bar{x}} \tag{1.4}$$

leads to an error of order $k^2 + h^2$. The error incurred when the wave equation (1.2) is approximated by the von Neumann difference equation (1.1) is of order $k^2 + h^2$ for all values of the parameter α.

The difference approximation (1.4) was extended to a difference analogue of the hyperbolic differential equation

$$\frac{\partial^2 u}{\partial t^2} = a(x, t)\frac{\partial^2 u}{\partial x^2} + b(x, t)\frac{\partial u}{\partial x} + c(x, t)\frac{\partial u}{\partial t} + d(x, t)u + e(x, t) \tag{1.5}$$

by Courant, Friedrichs and Lewy, and shown to be conditionally stable, the mesh ratio limitation being $1 - aR^2 > 0$. The implicit backward difference approximation (1.3) has been extended to difference analogues of equations of the form (1.5) by Lees [4], and shown to be unconditionally stable. Both extensions preserve the order of magnitude of the error of approximation.

There are two natural extensions of the von Neumann difference equation (1.1) to a difference analogue of the hyperbolic differential equation (1.5). We consider both of these extensions of (1.1), and give sufficient conditions that they be unconditionally stable. Both of these extensions lead to an error of approximation of order $k^2 + h^2$ for all values of the parameter α. Unfortunately, our method of proving stability gives no information about the conditional stability of the von Neumann difference approximations to (1.5).

We establish the stability theorems by showing that the solutions of the von Neumann difference approximations to (1.5) satisfy an *energy inequality* similar to the classical energy inequality of Friedrichs and Lewy [2]. It is the energy inequality which allows us to handle differential equations with variable coefficients; the case of constant coefficients can be treated by Fourier analysis (cf. [3]).

2. Preliminary remarks. Let $\bar{\Omega}_h$ be a rectangular lattice with mesh widths h and $k = Rh$ fitted to the region

$$\bar{\Omega}: 0 \leq x \leq 1, 0 \leq t \leq T .$$

More precisely, $\bar{\Omega}_h$ is the set of all points of intersection of the coordinate lines

$$x = nh, n = 0, 1, \cdots, N ,$$
$$t = mk, m = 0, 1, \cdots, M ,$$

where $Nh = 1$ and $Mk = T$. The quantity R is called the *mesh ratio* of the lattice.

Let

$$\Delta^i(m, l) = \{(x, t) \in \bar{\Omega}_h \,|\, x = ih \text{ and } mk \le t \le lk\}$$

and put

$$\Omega_h = \bigcup_{i=1}^{N-1} \Delta^i(k, M-k) .$$

Denote by $T_{\pm nh}$ and $T_{\pm mk}$ the translation operators defined as follows:

$$T_{\pm nh}v(x, t) = v(x \pm nh, t) ,$$
$$T_{\pm mk}v(x, t) = v(x, t \pm mk) .$$

For the first order partial difference quotients of functions $v(x, t)$ we employ the following notation:

$$v_x = \frac{1}{h}(T_h - 1)v, \qquad v_t = \frac{1}{k}(T_k - 1)v ,$$

$$v_{\bar{x}} = \frac{1}{h}(1 - T_{-h})v, \qquad v_{\bar{t}} = \frac{1}{k}(1 - T_{-k})v ,$$

$$v_{\hat{x}} = \frac{1}{2h}(T_h - T_{-h})v, \qquad v_{\hat{t}} = \frac{1}{2k}(T_k - T_{-k})v .$$

Difference quotients of order higher than the first are fo
peated application of the above formulas, for example,

$$v_{x\bar{x}} = \frac{1}{h^2}(T_h - 2 + T_{-h})v .$$

We shall use von Neumann's finite difference method to approximate the sufficiently smooth solutions of the following mixed initial-boundary value problem

$$\frac{\partial^2 u}{\partial t^2} = a(x, t)\frac{\partial^2 u}{\partial x^2} + b(x, t)\frac{\partial u}{\partial x} + c(x, t)\frac{\partial u}{\partial t} + d(x, t)u + e(x, t), \tag{2.1}$$
$$(0 < x < 1, 0 < t \le T) ,$$

$$\begin{aligned} u(x, 0) &= f(x) \\ \frac{\partial u}{\partial t}(x, 0) &= g(x) \\ u(0, t) &= h_0(t) \\ u(1, t) &= h_1(t) . \end{aligned} \tag{2.2}$$

We assume that the functions b, c, d and e are continuous in $\bar{\Omega}$, and that there exist constants c_i, $(i = 0, 1, 2, 3)$ such that

$$0 < c_0 \le a(x, t) \le c_1 \text{ in } \bar{\Omega} , \tag{2.3}$$

$$(2.4) \qquad |a(x, t) - a(\bar{x}, \bar{t})| \le c_2|x - \bar{x}| + c_3|t - \bar{t}| \; .$$

We also assume that the functions g, h_0 and h_1 are continuous and that f is twice continuously differentiable.

The first finite difference approximation of (2.1) is

$$(2.5) \quad v_{t\bar{t}} = a(x, t)v_{x\bar{x}} + \alpha k^2 v_{x\bar{x}t\bar{t}} + b(x, t)v_{\hat{x}} + c(x, t)v_{\hat{t}} + d(x, t)v + e(x, t) \; ,$$

where α is a parameter to be specified later.

The initial and boundary conditions (2.2) are approximated as follows:

$$(2.6) \quad \begin{aligned} v(x, 0) &= f(x) \\ v(x, k) &= f(x) + kg(x) + \frac{k^2}{2}[a(x, 0)f''(x) + b(x, 0)f'(x) \\ &\qquad + c(x, 0)g(x) + d(x, 0)f(x) + e(x, 0)]\text{[2]} \\ v(0, t) &= h_0(t) \\ v(1, t) &= h_1(t) \; . \end{aligned}$$

As a second approximation to (2.1), we take

$$(2.7) \quad \begin{aligned} v_{t\bar{t}} = a(x, t)v_{x\bar{x}} &+ a(x, t)\alpha k^2 v_{x\bar{x}t\bar{t}} + b(x, t)v_{\hat{x}} \\ &+ c(x, t)v_{\hat{t}} + d(x, t)v + e(x, t) \; . \end{aligned}$$

Both of these difference equations are of the implicit type when $\alpha > 0$, i.e., their solution, subject to the auxilliary conditions (2.6), requires the inversion of a linear system of $(N - 1)$ algebraic equations in the same number of unknowns at each time step.

3. Energy inequalities. In this section, we derive sufficient conditions for the solutions of the difference equations (2.5) and (2.7) to satisfy an energy inequality. As a corollary of the energy inequalities, we prove the existence of a unique solution to the systems (2.5), (2.6), and (2.7), (2.6).

Before giving the energy inequalities, we prove several preliminary results.

LEMMA 1. *The function* $E(t) = (1 + \beta k)^{-t/k}$, $(\beta > 0)$ *satisfies the following conditions*:

(i) $E_{\bar{t}} + \beta E = 0$,

(ii) $T_k E = (1 + \beta k)^{-1} E \le E$,

(iii) $E^{-1}(t) \le e^{\beta t}$.

[2] This approximation is taken to insure that the approximation error is of uniform order of magnitude over the region $\bar{\Omega}_h$.

Proof. Properties (i) and (ii) are readily verified, and property (iii) follows by exponentiating the inequality

$$t/k \log(1 + \beta k) \leq \beta t \ .$$

The next three lemmas give finite difference analogues of certain differential identities employed in the proof of the classical energy inequality of Friedrichs and Lewy.

LEMMA 2. *If $E(t) = (1 + \beta k)^{-t/k}$, $(\beta > 0)$, and $v(x, t)$ is any function defined on $\bar{\Omega}_h$, then the following identity holds:*

$$Ev_{\hat{t}}v_{t\bar{t}} = \tfrac{1}{2}[(T_k E)v_t^2]_{\bar{t}} + \tfrac{1}{2}\beta(T_k E)v_t^2 \tag{3.1}$$

Proof. We have that

$$[v_t^2]_{\bar{t}} = v_t v_{t\bar{t}} + (T_{-k}v_t)v_{t\bar{t}} = 2v_t v_{t\bar{t}} - (v_t - T_{-k}v_t)v_{t\bar{t}} = 2v_t v_{t\bar{t}} - k(v_{t\bar{t}})^2 \ .$$

Similarly,

$$[v_t^2]_{\bar{t}} = 2v_{\bar{t}}v_{t\bar{t}} + k(v_{t\bar{t}})^2 \ .$$

Hence,

$$v_{\hat{t}}v_{t\bar{t}} = \tfrac{1}{2}[v_t^2]_{\bar{t}} \ .$$

Therefore

$$Ev_{\hat{t}}v_{t\bar{t}} = \tfrac{1}{2}[(T_k E)v_t^2] - \tfrac{1}{2}(T_k E)_{\bar{t}}v_t^2$$

which reduces to (3.1) in view of property (i) of Lemma 1.

LEMMA 3. *Let $E(t) = (1 + \beta k)^{-t/k}$, and let $v(x, t)$ be any function defined on $\bar{\Omega}_h$. Then the following identity holds:*

$$\begin{aligned} Ev_{\hat{t}}av_{x\bar{x}} = {} & (v_{\hat{t}}Eav_x)_{\bar{x}} - v_{\bar{x}}Ea_{\bar{x}}(T_{-h}v_{\hat{t}}) \\ & - \frac{1}{4}[(T_{-k}Ea)v_{\bar{x}}^2]_t - \frac{1}{4}[(T_k Ea)v_{\bar{x}}^2]_{\bar{t}} \\ & + \frac{1}{2}(Ea)_{\hat{t}}v_{\bar{x}}^2 + \frac{k^2}{4}[(T_k Ea)v_{\bar{x}t}^2]_{\bar{t}} - \frac{k^2}{4}(Ea)_t v_{\bar{x}t}^2 \ . \end{aligned} \tag{3.2}$$

Proof. We have

$$\begin{aligned} Ev_t av_{x\bar{x}} &= (v_t Eav_x)_{\bar{x}} - (v_t Ea)_{\bar{x}}v_{\bar{x}} \\ &= (v_t Eav_x)_{\bar{x}} - Eav_{\bar{x}}v_{\bar{x}t} - (T_{-k}v_t)v_{\bar{x}}Ea_{\bar{x}} \\ &= (v_t Eav_x)_{\bar{x}} - (T_{-h}v_t)v_{\bar{x}}Ea_{\bar{x}} - \tfrac{1}{2}[(T_{-k}Ea)v_{\bar{x}}^2]_t \\ &\quad + \frac{1}{2}(Ea)_{\bar{t}}v_{\bar{x}}^2 + \frac{k}{2}Eav_{\bar{x}t}^2 \ . \end{aligned}$$

Combining this identity with a similar representation for $Ev_{\bar{t}}av_{x\bar{x}}$ we obtain (3.2).

LEMMA 4. *Let* $E(t) = (1 + \beta k)^{-t/k}$, *and let* $v(x, t)$ *be any function defined on* $\bar{\Omega}_h$. *Then the following identity holds:*

$$(3.3) \quad Ev_{\bar{t}}av_{x\bar{x}t\bar{t}} = (Ev_{\hat{t}}av_{xt\bar{t}})_{\bar{x}} - Ea_{\bar{x}}(T_{-h}v_{\hat{t}})v_{\bar{x}t\bar{t}} - \tfrac{1}{2}[(T_kEa)v^2_{\bar{x}t}]_{\bar{t}} + \tfrac{1}{2}(Ea)_tv^2_{\bar{x}t} .$$

The proof of this lemma is similar to the proof of Lemma 3, and will be omitted.

In order to present the energy inequalities in a convenient form, we introduce several norms. If $v(x, t)$ is defined on $\bar{\Omega}_h$, then

$$||v||^2_{0,t} = h\sum_{n=1}^{N} v^2(nh, t)$$

$$^*||v||^2_{0,t} = h\sum_{n=1}^{N-1} v^2(nh, t)$$

$$||v||^2_{1,t} = ||v_{\bar{t}}||^2_{0,t} + \tfrac{1}{2}\{||v_{\bar{x}}||^2_{0,t} + ||v_{\bar{x}}||^2_{0,t-k}\} .$$

For functions defined on $\bar{\Omega}$ we introduce the maximum norm

$$|v|_{\bar{\Omega}} = \max_{\bar{\Omega}} |v(x, t)| .$$

THEOREM 1. (Energy Inequality) *Let* $v(x, t)$ *be a solution of the difference equation* (2.5) *in* Ω_h. *Assume that* $v(x, t)$ *vanishes* $\Delta^0(0, M)$ *and* $\Delta^N(0, M)$. *If the quantity*

$$(3.5) \qquad 4\alpha - a(x, t)$$

is bounded away from zero in $\bar{\Omega}$, *then there exists a constant* c *depending only on* α, T, c_i, $(i = 0, 1, 2, 3)$ *and the coefficients of* (2.1) *such that for all sufficiently small* k

$$(3.6) \qquad ||v||^2_{1,t} \leq c[||v||^2_{1,k} + k\sum_{\tau=k}^{t-k} {}^*||e||^2_{0,\tau}] .$$

Proof.[3] We have

$$hk\sum_{\Omega_h} Ev_{\hat{t}}[v_{t\bar{t}} - av_{x\bar{x}} - \alpha k^2v_{x\bar{x}t\bar{t}} - bv_{\hat{x}} - cv_{\hat{t}} - dv - e] = 0 .$$

Let $\Delta_h = \Omega_h \cup \Delta^N(k, M - k)$. Since v vanishes on $\Delta^N(0, M)$, we can write the preceding equation in the form

$$(3.7) \qquad hk\sum_{\Lambda_h} Ev_{\hat{t}}[v_{t\bar{t}} - av_{x\bar{x}} - \alpha k^2v_{x\bar{x}t\bar{t}}] = \tfrac{1}{2}B(v) ,$$

where

[3] The basic idea in the proof is made more transparent by taking $b = c = d = e = 0$.

$$(3.8) \qquad B(v) = 2hk\sum_{\Omega_h} Ev_{\hat{t}}[bv_{\hat{x}} + cv_{\hat{t}} + dv + e] \,.$$

Using the identities of Lemmas 2, 3 and 4[4], we can write (3.7) as

$$(3.9) \qquad \begin{aligned} hk\sum_{\Lambda_h} [(T_kE)v_{\hat{t}}^2]_{\bar{t}} &+ \tfrac{1}{2}[(T_{-k}Ea)v_{\bar{x}}^2]_t + \tfrac{1}{2}[(T_kEa)v_{\bar{x}}^2]_{\bar{t}} \\ &- \tfrac{1}{2}k^2[(T_kEa)v_{\bar{x}t}^2] + \alpha k^2[(T_kEa)v_{\bar{x}t}^2]_{\bar{t}} = R(v) + B(v) \,, \end{aligned}$$

where

$$(3.10) \qquad \begin{aligned} R(v) = hk\sum_{\Lambda_h} \{&-\beta(T_kE)v_{\hat{t}}^2 - 2Ea_{\bar{x}}v_{\bar{x}}(T_{-h}v_{\hat{t}}) \\ &+ (Ea)_{\hat{t}}v_{\bar{x}}^2 - \tfrac{1}{2}Eak^2v_{\bar{x}t}^2 + \alpha k^2E_tv_{\bar{x}t}^2\} \,. \end{aligned}$$

In deriving (3.9), we have used the fact that

$$\sum_{\Lambda_h} (Ev_{\hat{t}}av_x)_{\bar{x}} = \sum (Ev_{\hat{t}}v_{xt\bar{t}})_{\bar{x}} = 0$$

which follows from our assumption that $v(x, t)$ vanishes on $\Delta^0(0, M)$ and $\Delta^N(0, M)$.

Summation with respect to t yields the following formulas.

$$(3.11) \qquad hk\sum_{\Lambda_h} [(T_kE)v_{\hat{t}}^2]_{\bar{t}} = E(T)\|v_{\bar{t}}\|_{0,T}^2 - \|v_{\bar{t}}\|_{0,k}^2$$

$$(3.12) \qquad \begin{aligned} &hk\sum_{\Lambda_h} [(T_{-k}Ea)v_{\bar{x}}^2 \\ &\quad = h\sum_{x=h}^{Nh} (T_{-k}E(T)a(x, T)v_{\bar{x}}^2(x, T) - h\sum_{x=h}^{Nh} a(x, 0)v_{\bar{x}}^2(x, k) \end{aligned}$$

$$(3.13) \qquad \begin{aligned} &hk\sum_{\Lambda_h} [(T_kEa)v_{\bar{x}}^2]_{\bar{t}} \\ &\quad = h\sum_{x=h}^{Nh} E(T)a(x, T)v_{\bar{x}}^2(x, T - k) - h\sum_{x=h}^{Nh} E(k)a(x, k)v_{\bar{x}}^2(x, 0) \end{aligned}$$

$$(3.14) \qquad \begin{aligned} &hk\sum_{\Lambda_h} k^2[(T_kEa)v_{\bar{x}t}^2]_{\bar{t}} \\ &= h\sum_{x=h}^{Nh} E(T)a(x, T)[v_{\bar{x}}^2(x, T) - 2v_{\bar{x}}(x, T)v_{\bar{x}}(x, T - k) + v_{\bar{x}}^2(x, T - k)] \\ &\quad - h\sum_{x=h}^{Nh} E(k)a(x, k)[v_{\bar{x}}^2(x, k) - 2v_{\bar{x}}(x, k)v_{\bar{x}}(x, 0) + v_{\bar{x}}^2(x, 0)] \end{aligned}$$

$$(3.15) \qquad \begin{aligned} &hk\sum_{\Lambda_h} \alpha k^2[(T_kE)v_{\bar{x}t}^2]_{\bar{t}} \\ &= \alpha h\sum_{x=h}^{Nh} E(T)[v_{\bar{x}}^2(x, T) - 2v_{\bar{x}}(x, T)v_{\bar{x}}(x, T - k) + v_{\bar{x}}^2(x, T - k)] \\ &\quad - \alpha h\sum_{x=h}^{Nh} E(k)[v_{\bar{x}}^2(x, k) - 2v_{\bar{x}}(x, k)v_{\bar{x}}(x, 0) + v_{\bar{x}}^2(x, 0)] \,. \end{aligned}$$

It follows now from (3.9), (3.11)–(3.15) that

[4] In Lemma 4, we take $a \equiv 1$.

$$
\begin{aligned}
E(T)\|v_{\bar t}\|^2_{0,T} &+ \frac{h}{2}\sum_{x=h}^{Nh}[2\alpha E(T)v_{\bar x}^2(x,T-k) \\
&+ \{2\alpha E(T) - k(E(T)a(x,T))_{\bar t}\}v_{\bar x}^2(x,T) \\
&+ \{2E(T)a(x,T) - 4\alpha E(T)\}v_{\bar x}(x,T)v_{\bar x}(x,T-k)] \\
&= \|v_{\bar t}\|^2_{0,k} + \frac{h}{2}\sum_{x=h}^{Nh}[2\alpha E(k)a(x,k)v_{\bar x}^2(x,0) \\
&+ \{2\alpha E(k) - k(E(k)a(x,k))_{\bar t}\}v_{\bar x}^2(x,k) \\
&+ \{2E(k)a(x,k) - 4\alpha E(k)\}v_{\bar x}(x,k)v_{\bar x}(x,0)] + R(v) + B(v)\,. \qquad (3.16)
\end{aligned}
$$

Consider the real quadratic form

$$Q(\xi, \lambda) = 2\alpha E\xi^2 + (2Ea - 4\alpha E)\xi\eta + (2\alpha E - k(Ea)_{\bar t})\eta^2\,. \qquad (3.17)$$

Now,

$$(Ea)_{\bar t} = E_{\bar t}(T_{-k}a) + Ea_{\bar t} = E(-\beta(T_{-k}a) + a_{\bar t})$$

by property (i) of Lemma 1. Therefore

$$2\alpha E - k(Ea)_{\bar t} \geq E2\alpha \qquad (3.18)$$

if $\beta a \geq a_{\bar t}$.

It follows from (3.17) and (3.18) that by choosing β large enough

$$Q(\xi, \eta) \geq E(2\alpha\xi^2 + 2\{a - 2\alpha\}\xi\eta + 2\alpha\eta^2)\,. \qquad (3.19)$$

In view of our assumption concerning the expression (3.5), we see that the right side of (3.19) is a positive definite quadratic form, and

$$Q(\xi, \eta) \geq E\mu_0(\xi^2 + \eta^2)\,, \qquad (3.20)$$

where

$$\mu_0 = \min\,[4\alpha - a, a]\,.$$

Also, there is a constant μ_1 such that

$$Q(\xi, \eta) \leq E\mu_1(\xi^2 + \eta^2)\,. \qquad (3.21)$$

Hence, combining (3.16), (3.20) and (3.21) we fined that

$$
\begin{aligned}
E(T)\|v_{\bar t}\|^2_{0,T} &+ E(T)\mu_0\tfrac{1}{2}\{\|v_{\bar x}\|^2_{0,T} + \|v_{\bar x}\|^2_{0,T-k}\} \\
&\leq \|v_{\bar t}\|^2_{0,k} + \mu_1\tfrac{1}{2}\{\|v_{\bar x}\|^2_{0,0} + v_{\bar x}\|^2_{0,k}\} + R(v) + B(v)\,. \qquad (3.22)
\end{aligned}
$$

If we shown that there is a constant B such that

$$R(v) + B(v) \leq kB\|v_{\bar t}\|^2_{0,k} + k\sum_{\tau=k}^{T-k}{}^{*}\|e\|^2_{0,\tau}E(\tau) \qquad (3.23)$$

for all sufficiently small k, then (3.6) will follow.

It is readily verified that

$$B(v) \leq hk\sum_{\Omega_h} E(|b|_{\bar{\Omega}} + |c|_{\bar{\Omega}} + |d|_{\bar{\Omega}} + 1)v_t^2$$
$$+ hk\sum_{\Omega_h} E(|b|_{\bar{\Omega}} + |d|_{\bar{\Omega}})v_x^2 + k\sum_{\tau-k}^{T-k} {}^*\|e\|_{0,\tau}^2 E(\tau) .$$

Since

$$\sum_{\Omega_h} Ev_x^2 \leq \sum Ev_{\bar{x}}^2$$

and

$$\sum_{\Omega_h} Ev_t^2 \leq \sum_{\Omega_h} Ev_{\bar{t}}^2 + k\frac{h}{2}\sum_{x=h}^{Nh-h} E(k)v_t^2(x, k)$$

it follows that there are constants B_1 and B_2 such that

$$B(v) \leq B_1 hk\sum_{\Lambda_h} v_{\bar{x}}^2 E + B_2 hk\sum_{\Omega_h} v_{\bar{t}}^2 E + B_2 E(k)\frac{k}{2}\|v_{\bar{t}}\|_{0,k}^2 \tag{3.24}$$
$$+ k\sum_{\tau=k}^{T-k} {}^*\|e\|_{0,\tau}^2 E(\tau) .$$

Using (3.10), it is not difficult to show that

$$R(v) \leq \sum_{\Lambda_h} E\left\{\frac{-\beta}{1+\beta k}v_t^2 + \left|\frac{\partial a}{\partial x}\right|_{\bar{\Omega}} v_t^2 + \left|\frac{\partial a}{\partial x}\right|_{\bar{\Omega}} v_{\bar{x}}^2\right. \tag{3.25}$$
$$+ \frac{1}{2}|a|_{\bar{\Omega}} k^2 v_{\bar{x}t}^2 - \alpha k^2\frac{2}{1+\beta k}v_{\bar{x}t}^2$$
$$\left. + \left(-\frac{\beta}{2}(T_{-k}a) + a_t - \frac{\beta}{2(1+\beta k)}(T_k a)\right)v_{\bar{x}}^2\right\} + kB_3\|v_{\bar{t}}\|_{0,k}^2$$

for a suitable constant B_3. It follows now from (3.24) and (3.25) that we can choose β such that for all sufficiently small k (3.23) holds. This completes the proof of Theorem 1.

Restating Theorem 1, we have

COROLLARY 1. *If $4\alpha - a$ is bounded away from zero in $\bar{\Omega}$, then the von Neumann difference equation* (2.5) *is unconditionally stable for all sufficiently small k.*

THEOREM 2. *If $4\alpha - a$ is bounded away from zero in $\bar{\Omega}$, then for all sufficiently small k, the finite difference equation* (2.5) *with the auxiliary conditions* (2.6) *possesses a unique solution.*

Proof. At the end of §2, we remarked that the system (2.5), (2.6) is equivalent to a system of $(N-1)(M-2)$ linear equations in the same number of unknowns. It is sufficient to prove that the associated homogeneous system of equations has only the trivial solution. The homogeneous system is obtained by putting e, $v(x, 0)$, $v(x, k)$, $v(0, t)$ and $v(1, t)$ equal to zero. From Theorem 1 we conclude that $||v||_{1,t} = 0$, which implies that $v = 0$ in Ω_h. Hence, the associated homogeneous system has only the trivial solution.

We now state without proof the following approximation theorem.

THEOREM 3. *Let $u(x, t)$ be of class C^4 in $\bar{\Omega}$ and satisfy the mixed initial boundary value problem* (2.1), (2.2). *Let $4\alpha - a$ be bounded away from zero in $\bar{\Omega}$ and let $v(x, t)$ denote the solution of the von Neumann approximation* (2.5), (2.6). *Then for all sufficiently small k there exists a constant B_4 independent of h and k such that*

$$\max_{\bar{\Omega}_h} |u(x, t) - v(x, t)| \leq B_4(h^2 + k^2) .$$

We now consider the finite difference equation (2.7). The preceding theorems can all be extended to this difference equation provided that we modify the range of the parameter α.

THEOREM 4. (Energy Inequality) *Let $v(x, t)$ be a solution of the difference equation* (2.7) *in Ω_h. Assume that $v(x, t)$ vanishes on $\Delta^0(0, M)$ and $\Delta^N(0, M)$. If*

$$4\alpha - 1 > 0 , \tag{3.26}$$

then there exists a constant c independent of h and k such that

$$||v||_{1,t}^2 \leq c\left[||v||_{1,k}^2 + k\sum_{\tau=k}^{t-k}{}^* ||e||_{0,\tau}^2\right] . \tag{3.27}$$

The proof of Theorem 4 is quite similar to the proof of Theorem 1. Only slight changes in the proof are required due to the fact that we must use the full form of Lemma 4.

BIBLIOGRAPHY

1. R. Courant, K. O. Friedrichs, and H. Lewy, *Über die Partiellen Differentialgleichungen der Mathematischen Physik*, Math. Ann., **100** (1928), 32–74.
2. K. O. Friedrichs and H. Lewy, *Über die Eindeutigkeit und das abhängigkutsgebeit der Läsungen beim Aufanswert problem linearer Hyperbolischer Differentialgleichungen*, Math. Ann. **98** (1927), 192-204.
3. P. D. Lax and R. D. Richtmyer, *Survey of the stability of linear finite difference equations*, Comm. Pure Appl. Math., **9** (1956), 267–293.
4. M. Lees, *Energy inequalities for the solution of differential equations*, Trans. Amer. Math. Soc., to appear.
5. G. O'Brien, M. A. Hyman, and S. Kaplan, *A study of the Numerical solution of partial differential equations*, Jour. Math. Physics, **29** (1951), 223–251.

THE INSTITUTE FOR ADVANCED STUDY

AXIOM SCHEMATA OF STRONG INFINITY IN AXIOMATIC SET THEORY

AZRIEL LÉVY

1. Introduction. There are, in general, two main approaches to the introduction of strong infinity assertions to the Zermelo-Fraenkel set theory. The arithmetical approach starts with the regular ordinal numbers, continues with the weakly inaccessible numbers and goes on to the ρ_0-numbers of Mahlo [4], etc. The model-theoretic approach, with which we shall be concerned, introduces the strongly inaccessible numbers and leads to Tarski's axioms of [14] and [15]. As we shall see, even in the model-theoretic approach we can use methods for expressing strong assertions of infinity which are mainly arithmetical. Therefore we shall introduce strong axiom schemata of infinity by following Mahlo [4,5,6,]. Using the ideas of Montague in [7] we shall give those axiom schemata a purely model-theoretic form. Also the axiom schemata of replacement in conjunction with the axiom of infinity will be given a similar form, and thus the new axiom schemata will be seen to be natural continuations of the axiom schema of replacement and infinity.

A provisional notion of a standard model, introduced in § 2, will be basic for our discussion. However, in § 5 it is shown that this definition cannot serve as a general definition for the notion of a standard model.

2. Standard models of set theories. For the forthcoming discussion we need the notion of a standard model of a set theory. A general principle which distinguishes between standard and non-standard models of set theory is not yet known. Nevertheless, a notion of a standard model for various set theories will be given here, but this will serve only as an ad-hoc principle and we shall see later that its general application is not justified.

The Zermelo-Fraenkel set theory is generally formalized in the simple applied first-order functional calculus, since this is the most natural language for a set theory. In that formulation the Zermelo-Fraenkel set theory has an infinite number of axioms. From that formulation one passes directly to a formulation of the Zermelo-Fraenkel set theory by a finite number of axioms in the non-simple applied first-order functional calculus (we shall denote functional variables with $p, p_1, p_2, \cdots$). The axioms of extensionality, pairing, sum-set, power-set and infinity are as in [2]. The changed axioms are

The axiom of subsets $(x)(\exists y)(z)(z \in y \equiv : z \in x \,.\, p(z))$

Received February 20, 1959. This paper was written while the author was a Sloan Fellow of the School for Advanced Study at M.I.T.

The axiom of replacement
$(x, y, z,) (p(x, y) \,.\, p(x, z) : \supset y=z) \supset (x) (\exists\, y) (z) (z \in y \equiv (\exists\, u) (u \in x \,.\, p(u, z)))$.

The axiom of foundation $(\exists\, x)\, p(x) \supset (\exists\, x) (p(x) \,.\, (y) (y \in x \supset \sim p(y)))$.

If we regard as mathematical theorems of a theory Q formulated in the non-simple applied first order functional calulus only those theorems of Q which do not contain functional variables then it can be shown, by the method of *Rückverlegung der Einsetzungen* (compare [3], pp. 248-249) that the set of all the mathematical theorems of Q coincides with the set of all the theorems of the corresponding theory Q' formulated in the simple applied first order functional calculus (whose axioms are the axioms and the axiom schemata corresponding to the axioms of Q). Therefore Q and Q' could be regarded, from the mathematical point of view as the same theory. Nevertheless, we shall see that Q' is not obtained uniquely from Q if we disregard the actual axiomatic representation of Q.

We are interested in passing to set theories based on a finite set of axioms in the non-simple applied first-order functional calculus, since in this case we can define the notion of standard models for these theories in the sense of Henkin. A standard model of such a theory Q will be a model where the functional variables range over all the subsets of the universe set of the model. The statement that the universe u and the membership relation e (which are both taken to be sets) determine a standard model of Q can be easily formulated in set theory. This is done as follows: We take the conjunction of the axioms of Q and effect the following replacements[1]

$(x)($ by $(x)(x \in u \supset$ $\qquad$ $(\exists\, x)($ by $(\exists\, x)(x \in u.$

$x \in y$ by $\langle x\, y \rangle \in e$ $\quad$ $p_i(x_1, \cdots, x_{n_i})$ by $\langle x_1, \cdots, x_{n_i} \rangle \in f_i$

and

then we close the resulting formula with respect to the variables f_i by the prefix $(f_1, \cdots, f_j)(f_1 \subseteq u \,.\, \cdots \,.\, f_j \subseteq u : \supset$. Thus we obtain a formula which we shall denote with $Sm^Q(u, e)$.

Standard models for set theories for which $\langle x\, y \rangle \in e \equiv : y \in u \,.\, x \in y$ and $y \in u \,.\, x \in y : \supset x \in u$ are called standard complete models: $Scm^Q(u) \equiv : (y)(y \in u \supset y \subseteq u) \,.\, (e) (\langle x\, y \rangle \in e \equiv : y \in u \,.\, x \in y : .\, Sm^Q(u, e))$.

We denote by S the set theory which consists of the axioms of extensionality, pairing, sum-set, power-set, subsets and foundation. SF will denote the theory obtained from S by adding to it the axiom of replacement. Z (resp. ZF) will denote the theory obtained from S (resp. SF) by the addition of the axiom of infinity (axiom VII* of [2]). We shall assume that these theories are formulated in the simple first-order

[1] Alphabetic change of bound variables may also be needed.

functional calculus unless we are dealing with standard models of these theories, in which case we shall assume that we have passed to corresponding formalizations in the non-simple first-order functional calculus.

By the methods of Shepherdson [12] 1.5 and Mostowski [9] it is easy to prove (in SF) that each standard model of a set theory Q which includes the axioms of extensionality and foundation is isomorphic to some standard complete model of Q.

The function $R(\alpha)$ is defined by $R(\alpha) = \sum_{\beta<\alpha} \mathfrak{P}(R(\beta))$ ($\mathfrak{P}(x)$ is the power-set of x). The rank of an element x of $R(\alpha)$ is defined to be the first β such that $x \in R(\beta)$. We shall assume in the following that the properties of these functions are known.[2]

We can prove, in the same way as Shepherdson [12] 3.14 and 3.3 [3] that if Q contains the axioms and the axiom schemata of SF then each standard complete model of Q is of the form $R(\alpha)$, where α is some limit number. Thus we can conclude that each standard model of a theory Q which contains the axioms and axiom schemata of SF is isomorphic to some standard complete model of Q of the form $R(\alpha)$. If we regard as assertions of infinity those statements which assert the existence of standard models for strong set theories, we see now why all assertions of infinity reduce to statements about the existence of ordinal numbers with appropriate properties.

The (strongly) inaccessible numbers α are usually defined as regular initial numbers greater than ω which satisfy $(\lambda)(\lambda < \alpha \supset \overline{\overline{2^{\lambda}}} < \alpha)$. This definition leads to the expected consequence only if the axiom of choice is assumed, since, for example, if the cardinal of the continuum is not an aleph then according to this definition no ordinal is inaccesible. Shepherdson [12] established the close connection between the inaccessible numbers and what we call the standard complete models of ZF. These results of Shepherdson can serve to give a new definition of inaccessible numbers which will have a satisfactory meaning even if the axiom of choice is not assumed.

DEFINITION 1. α is called inaccessible if $R(\alpha)$ is a standard complete model of ZF.

$In(\alpha) \equiv Scm^{ZF}(R(\alpha))$

Shepherdson [12] proves, in effect, that this definition is equivalent to the usual definition if the axiom of choice is assumed. Without using the axiom of choice it can be proved that α is inaccessible if and only if

(1) $\alpha > \omega$

(2) α is regular

(3) $(z)(z \in R(\alpha) \supset \sim \overline{\overline{z}} \geq \overline{\overline{\alpha}})$ [4]

---

[2] See, for example, Shepherdson [**12**] 3.2– he denotes the function R by G.

[3] Shepherdson's super-complete models are our standard complete models of ZF.

We shall widely use in the following the fact that every inaccessible number is regular (this is proved by Shepherdson [12] 3.42).

Definition 1 shows clearly why such a number is called inaccessible, i. e., unobtainable from the smaller ordinal numbers by means of the set theory ZF. Following Specker [13] we can generalize this definition as follows:

DEFINITION 2. Let Q be a set theory formulated by a finite number of axioms in the non-simple applied first-order functional calculus. An ordinal number α is called inaccessible with respect to Q if $R(\alpha)$ is a standard complete model of Q.

$In^Q(\alpha) = Scm^Q(R(\alpha))$.

3. A strong axiom schema of infinity. The numbers inaccessible with respect to ZF are the inaccessible numbers. The numbers inaccessible with respect to the theory obtained from ZF by addition of the axiom $(\exists\, \sigma)\, In(\sigma)$ are all the inaccessible numbers except the first one. Thus we can go on and observe numbers inaccessible with respect to systems which require the existence of more and more inaccessible numbers. We can also observe the numbers inaccessible with respect to the extension of ZF which is obtained by adding $(\mu)\,(\exists\, \sigma)\,(\sigma > \mu\, .\, In(\sigma))$ to its axioms, etc. But if we want to have a really fast trip into the realm of infinity we shall use the means provided by the arithmetical approach to assertions of infinity.

Mahlo [4] defined a function $\pi_{\alpha,\beta}$ such that $\pi_{\alpha,0}$ counts the regular ordinal numbers, $\pi_{\alpha,1}$ counts the weakly inaccessible number and for increasing β $\pi_{\alpha,\beta}$, regarded as a function of α, counts ordinals which satisfy higher and higher requirements of weak inaccessibility. The whole hierarchy of Mahlo [4] is based on the class [5] of the regular numbers — the range of $\pi_{\alpha,0}$. If we replace $\pi_{\alpha,0}$ by a function $\pi'_{\alpha,0}$ whose range is a subclass of the class of the regular numbers we can define analogously functions $\pi'_{\alpha,\eta}$ and $\pi'_{\alpha,\eta,\xi}$ and prove theorems corresponding to Mahlo's theorems in [4, 5, 6,]. We shall take for the range of $\pi'_{\alpha,0}$ the class of the inaccessible numbers.

Our exposition will differ from Mahlo's also in a technical point: Whereas Mahlo uses any strictly increasing functions to count the members of given classes of ordinal numbers we shall use for this purpose normal functions (*Normalfunktionen*) [6] which are much easier to handle. A normal function at limit-number arguments may take values

---

[4] Since we do not assume that the cardinal numbers are formally defined $\sim \bar{\bar{z}} \geqq \bar{\bar{\alpha}}$ is an abbreviation of a statement about equivalence of sets.

[5] We shall use the word '*class*' instead of the word '*property*', e.g., instead of '*the property of being a regular number*' we shall say '*the class of the regular numbers*'.

outside the class whose members it counts, since the normal function counts the members of the closure (in the order topology) of the given class.

DEFINITION 3. The functions $P_\eta(\alpha)$[7] are defined by transfinite induction as follows: $P_0(0)$ is the first inaccessible number; $P_0(\beta+1)$ is the first inaccessible number greater than $P_0(\beta)$; for limit-number α $P_0(\alpha) = \lim_{\beta<\alpha} P_0(\beta)$. $P_\eta(\beta+1)$ (resp. $P_\eta(0)$) is the first inaccessible number σ greater than $P_\eta(\beta)$ (resp. the first inaccessible number) such that for each $\eta' < \eta$ $\sigma = P_{\eta'}(\gamma)$ for some limit number γ.

The functions $P_\eta(\alpha)$ are not assumed to be defined for evey η and α.

DEFINITION 4. $Q(\beta+1)$ (resp. $Q(0)$) is the first inaccessible number α greater than $Q(\beta)$ (resp. the first inaccessible number) such that $P_\alpha(0) = \alpha$. For a limit-number α $Q(\alpha) = \lim_{\beta<\alpha} Q(\beta)$.

We can also define functions $Q_\eta(\alpha)$ such that $Q_0(\alpha) = Q(\alpha)$, $Q_{\beta+1}$ is related to Q_β as Q_0 is related to P_0 and for limit-ordinal η Q_η counts the inaccessible numbers which are in the intersection of the ranges of all the functions $Q_{\eta'}$, $\eta' < \eta$. The numbers α for which $Q_\alpha(0) = \alpha$ we call Q^*-numbers.

We shall now consider the following axiom schema

M *Every normal function defined for all ordinals* (*d.f.a.o.*) *has at least one inaccessible number in its range*[8]

THEOREM 1. *M is equivalent to each of the following schemata*

M' *Every normal function d.f.a.o. has at least one fixed point which is inaccessible*

M'' *Every normal function d.f.a.o. has arbitrarily great fixed points which are inaccessible.*

Proof. Obviously M'' implies M' and M' implies M. We shall prove that M implies M''.

Let F be a normal function *d.f.a.o.* Let G be the derivative of F, i.e., the normal function which counts the fixed points of F. Since F is *d.f.a.o.* then by [1] § 8 G is also *d.f.a.o.* For any given γ let $H_\gamma(\xi) =$

---

[6] A function $F(\alpha)$ on the ordinal numbers into the ordinal numbers is called normal if:
(1) It is strictly increasing: $\alpha < \beta \supset F(\alpha) < F(\beta)$
(2) It is continous: For limit-number α $F(\alpha) = \lim_{\beta<\alpha} F(\beta)$.

[7] These are the functions analogous to the functions $\pi_{\alpha,\eta}$ of Mohlo [4].

[8] This schema is written formally as
$(\alpha,\beta,\gamma)(\varphi(\alpha,\beta) . \varphi(\alpha,\gamma) : \supset \beta = \gamma) . (\alpha)(\exists \beta)\varphi(\alpha,\beta) . (\alpha,\beta,\gamma,\delta)(\alpha < \gamma . \varphi(\alpha,\beta) . \varphi(\gamma,\delta) : \supset \beta < \delta) . (\alpha, \beta)(\sim(\exists \sigma)(\sigma+1=\alpha) . \alpha \neq 0 . \varphi(\alpha,\beta) : \subset (\gamma)(\gamma < \beta \supset (\exists \delta,\eta)(\delta < \alpha . \varphi(\delta,\eta) . \eta > \gamma))) : \supset (\exists \alpha,\beta)(\varphi(\alpha,\beta) . \mathrm{In}(\beta))$ where φ is a formula of set theory such that there is no confusion of variables in the corresponding instance of the schema.

$G(\gamma+\xi)$. H_γ is a normal function *d.f.a.o.* and hence by M there is an ordinal ξ such that $\beta = H_\gamma(\xi)$ is inaccessible. Since $\beta = G(\gamma + \xi)$, $F(\beta) = \beta$. By a well-known theorem the value of a normal function is not less than the argument and hence $\beta \geq \gamma + \xi \geq \gamma$.

In order to see how near M is to a purely arithmetical assertion it is interesting to note that M is equivalent to the conjunction of

(1) *There exist arbitrarily great inaccessible numbers*

(2) *Every normal function d.f.a.o. has at least one regular number in its range*

The proof of this makes use of the fact that every regular ordinal which is the limit of a set of inaccessible numbers is inaccessible (since an ordinal is inaccessible if and only if it is regular, greater than ω and $(z)(z \in R(\alpha) \supset \sim \bar{\bar{z}} \geq \bar{\alpha}))$. Let F be any normal function *d.f.a.o.* If there exist arbitrarily great inaccessible numbers then the function $P_0(\alpha))$ is *d.f.a.o.* and also the normal function $F(P_0(\alpha))$ is *d.f.a.o.* By (2), using the same reasoning as in Theorem 1, there is a regular ordinal β such that $F(P_0(\beta)) = \beta$, i.e., $P_0(\beta) = \beta$ and $F(\beta) = \beta$. Since β is a limit number and $P_0(\beta) = \beta$, β is the limit of a sequence of inaccessible numbers and since β is regular it is inaccessible.

ZM will denote the set theory obtained from ZF by the addition of M.

We shall now introduce a principle of reflection over ZF. This will be an axiom schema which will assert the existence of standard complete models of ZF which reflect in some sense the situation of the universe.

Let φ be a formula of set theory. We denote by $\mathrm{Rel}(u, \varphi)$ the formula obtained from φ by relativizing all the quantifiers in it to u, i.e., by replacing each occurrence $(z)\chi$ or $(\exists z)\chi$ by $(z)(z \in u \supset \chi)$ or $(\exists z)(z \in u \,.\, \chi)$, respectively.[9]

The principle of complete reflection over ZF

N $(\exists u)(Scm^{ZF}(u) \,.\, (x_1, \cdots, x_n)(x_1, \cdots x_n \in u \supset .\, \varphi \equiv \mathrm{Rel}(u, \varphi)))$ *where* φ *is any formula which has no free variables except* $x_1 \cdots, x_n$.

As seen from the formulation of N, it is closely connected with the notion of an arithmetical extension of Tarski and Vaught [17]. In the proofs of Theorems 2, 3, 5 and 6 we shall use the methods used by Montague and Vaught [8] for arithmetical extensions.

We shall see now that another principle of reflection, which seems at first sight to be stronger than N is equivalent to N.

THEOREM 2. *N is equivalent in S to the following schema*

[9] If φ contains u bounded then u is replaced in φ before the relativization by the first variable, in alphabetic order, which does not occur in φ.

N' $(\exists u)(z \in u \,.\, Scm^{ZF}(u). \quad (x_1, \cdots, x_n)(x_1, \cdots, x_n \in u \supset :$
$\varphi_1 \equiv \mathrm{Rel}(u, \varphi_1). \cdots\cdots. \varphi_m \equiv \mathrm{Rel}(u, \varphi_m)))$
where m is any natural number and φ_i, $1 \leq i \leq m$, is a formula which has no free variables except $x_1, \cdots, x_n$.

Proof. Obviously N' implies N. Now we assume N and we shall prove first the schema N'' which is like N', only that it does not contain the part $z \in u$. Let $\varnothing$ be the formula $\bigvee_{i=1}^m t = i \,.\, \varphi_i$. Since the natural numbers $1, 2, \cdots, m$ are absolute with respect to standard complete models (see, e.g., [12] 2.320) we have $Scm(u) \supset : . \mathrm{Rel}(u, \varnothing) \equiv : \bigvee_{i=1}^m t = i \,.\, \mathrm{Rel}(u, \varphi_i)$. We use now N with respect to $\varnothing$ and we obtain the existence of a set u such that $Scm^{ZF}(u)$ and

$$(t)(x_1, \cdots, x_n)(t, x_1, \cdots, x_n \in u \supset .\!:\! . \bigvee_{i=1}^m t = i \,.\, \varphi_i \; : \equiv : \bigvee_{i=1}^m t = i \,.\, \mathrm{Rel}(u, \varphi_i))$$

From $Scm^{ZF}(u)$ we can prove easily by induction that $\omega \subseteq u$, and therefore, substituting j for t in the above formula, $1 \leq j \leq m$, we get $x_1, \cdots, x_n \in u \supset . \varphi_j \equiv \mathrm{Rel}(u, \varphi_j)$, and thus we have proved N''. Now we shall prove N' from N''.

Given $\varphi_1, \cdots, \varphi_m$ we denote

$$\varphi_{m+1} = (\exists u)\Big(z \in u \,.\, Scm^{ZF}(u) \,.\, (x_1, \cdots, x_n)\Big(x_1, \cdots, x_n \in u \supset . \bigwedge_{i=1}^m \varphi_i \equiv \mathrm{Rel}(u, \varphi_i)\Big)\Big)$$

$$\varphi_{m+2} = (z)\, \varphi_{m+1}$$

We use now N'' for $\varphi_1, \cdots, \varphi_{m+2}$. Thus we have the existence of u such that $Scm^{ZF}(u)$ and

$$x_1, \cdots, x_n \in u \supset . \varphi_i \equiv \mathrm{Rel}(u, \varphi_i) \qquad 1 \leq i \leq m \tag{3}$$
$$z \in u \supset . \varphi_{m+1} \equiv \mathrm{Rel}(u, \varphi_{m+1}) \tag{4}$$
$$\varphi_{m+2} \equiv \mathrm{Rel}(u, \varphi_{m+2}) . \tag{5}$$

By $Scm^{ZF}(u)$ and (3) we have $(z)(z \in u \supset \varphi_{m+1})$, and hence, by (4), also $(z)(z \in u \supset \mathrm{Rel}(u, \varphi_{m+1}))$; but the latter formula is $\mathrm{Rel}(u, \varphi_{m+2})$ and hence, by (5), we have φ_{m+2}, which is the instance of N' corresponding to $\varphi_1, \cdots, \varphi_m$.

We note that Theorem 2 will remain valid if ZF is replaced in both N and N' by S or by any extension of S.

THEOREM 3. *In ZF the schema M is equivalent to the schema N and to the following schema*

N''' $(\exists\, \alpha)\,(In(\alpha)\,.\,(x_1, \cdots, x_n)\,(x_1, \cdots, x_n \in R(\alpha) \supset .\, \varphi \equiv \mathrm{Rel}\,(R(\alpha),\ \varphi)))$ *where* φ *is any formula which has no free variables except* $x_1, \cdots, x_n$.

Proof. As we have already mentioned in § 2 all the standard complete models of ZF are of the form $R(\alpha)$. Hence, by Definition 1, N and N''' are equivalent.

We shall now prove M from N'''. Let $\varphi(x, y\,; x_1, \cdots, x_n)$ be any formula. Let $\chi(x_1, \cdots, x_n)$ be the formula asserting that if $\varphi(\xi, \eta : x_1, \cdots, x_n)$ gives a function $\eta = F(\xi)$ which is normal and *d.f.a.o.* then $F(\xi)$ has at least one inaccessible number in its range. From N''' we shall pass, as in Theorem 2, to a schema which is like N''' only that $\varphi \equiv \mathrm{Rel}\,(R(\alpha), \varphi)$ is replaced by $\bigwedge_{i=1}^{4} \varphi_i \equiv \mathrm{Rel}\,(R(\alpha), \varphi_i)$. We shall take $\varphi_1 \equiv \varphi$, $\varphi_2 \equiv (\xi)\,(\exists\, \eta)\, \varphi\,(\xi, \eta)$, $\varphi_3 \equiv \chi$, $\varphi_4 \equiv (x_1, \cdots, x_n)\, \chi$. By the corresponding instance of that schema there exists an inaccessible number α such that

(6) $x_1, \cdots, x_n, x, y \in R\,(\alpha) \supset .\, \varphi \equiv \mathrm{Rel}\,(R(\alpha),\ \varphi)$
(7) $x_1, \cdots, x_n \in R\,(\alpha) \supset .\, (\xi)\,(\exists\, \eta)\, \varphi\,(\xi, \eta) \equiv \mathrm{Rel}\,(R(\alpha),\, (\xi)\,(\exists\, \eta)\, \varphi\,(\xi, \eta))$
(8) $x, \cdots, x_n \in R(\alpha) \supset .\, \chi \equiv \mathrm{Rel}\,(R(\alpha), \chi)$
(9) $(x_1, \cdots, x_n\,)\, \chi \equiv \mathrm{Rel}\,(R(\alpha),\ (x_1, \cdots, x_n)\chi)$.

We shall now assume that for certain $x_1, \cdots, x_n \in R(\alpha)$ $\varphi(\xi, \eta)$ gives a function $\eta = F(\xi)$ which is normal and *d.f.a.o.* The relativization of an ordinal-number-variable μ to the set $R(\alpha)$ is $\mu < \alpha$ (see Shepherdson [12] 2.316) and thus, since we assume the left-hand side of (7), we get

$$(\xi)\,(\xi < \alpha \supset (\exists\, \eta)\,(\eta < \alpha\,.\, \mathrm{Rel}\,(R(\alpha)\,,\ \varphi(\xi\,,\eta))))$$

and by (6) we have $(\xi)\,(\xi < \alpha \supset (\exists\, \eta)\,(\eta < \alpha\,.\, \varphi\,(\xi, \eta)))$. Since F is normal and α is a limit number we have $F(\alpha) = \alpha$, thus proving $x_1, \cdots, x_n \in R\,(\alpha) \supset \chi\,(x_1, \cdots, x_n)$. By (8) we have $x_1 \cdots, x_n \in R\,(\alpha) \supset \mathrm{Rel}\,(R(\alpha), \chi(x_1, \cdots, x_n))$ which is $\mathrm{Rel}\,(R(\alpha),(x_1, \cdots, x_n)\, \chi)$ and hence, by (9), we have $(x_1, \cdots, x_n)\, \chi$, thus proving M.

Now we shall prove N from M. In this we shall make use of ideas of Montague in [7]. Let φ be any formula of set theory. We write φ in prenex normal form. Let φ be of the form $(y)\,(\exists\, z)\,(u)\,(\exists\, t)\, \varphi^*$ where φ^* does not contain any quantifiers, and let φ have the two free variables x_1, x_2. For formulae φ of any other structure the treatment is analogous to the treatment of this case.

Given any x_1, x_2, y let $F_1(x_1, x_2, y)$ be the set of all the sets z which satisfy $(u)\,(\exists\, t)\varphi^*$ and which are of minimal rank among the sets satisfying this requirement. If there are z's satisfying $(u)\,(\exists\, t)\, \varphi^*$ then by the axiom of foundation they have certain ranks and hence $F_1(x_1, x_2, y) \neq 0$, otherwise $F_1(x_1, x_2, y) = 0$. $F_1(x_1, x_2, y)$ is a set since it is the subset of some set $R(\alpha)$ or it is the void set. Given any x_1, x_2, y, z, u we denote by $F_2(x_1, x_2, y, z, u)$ the set of all the sets t which satisfy φ^* and which are of the least rank among the sets t satisfying φ^*. $\sim \varphi = (\exists\, y)\,(z)\,(\exists\, u)\,(t) \sim \varphi^*$. We define for this formula corresponding functions $F_3(x_1, x_2)$ and $F_4(x_1, x_2, y, z)$.

$$\begin{aligned} H(x) = x + \sum_{x_1, x_2, y \in x} F_1(x_1, x_2, y) + \sum_{x_1, x_2 y, z, u \in x} F_2(x_1, x_2, y, z, u) \\ + \sum_{x_1, x_2 \in x} F_3(x_1, x_2) + \sum_{x_1, x_2, y, z \in x} F_4(x_1, x_2, y, z)\,. \end{aligned}$$

Let ξ be the rank of the set x, then $x \subseteq R(\xi - 1)$ (ξ cannot be a limit-number). Let us define

$$J(x) = R(\xi - 1), \quad K(x) = J(H(x)), \quad P(x) = \sum_{n \in \omega} K^n(x)\,.$$

It follows immediately from the definition of $P(x)$ that

$$\begin{aligned} x_1, x_2, y, z, u \in P(x) \supset : F_1(x_1, x_2, y) \subseteq P(x) \,.\, F_2(x_1, x_2, y, z, \mathrm{u}) \subseteq P(x) \\ .\, F_3(x_1, x_2) \subseteq P(x) \,.\, F_4(x_1, x_2, y, z) \subseteq P(x)\,. \end{aligned}$$

Denote

$$\begin{aligned} \varnothing(s) \equiv (x_1, x_2, y, z, u)\,(x_1, x_2, y, z, u \in s \supset : F_1(x_1, x_2, y) \subseteq s \\ .\, F_2(x_1, x_2, y, z, u) \subseteq s \,.\, F_3(x_1, x_2) \subseteq s \,.\, F_4(x_1, x_2, y, z) \subseteq s). \end{aligned}$$

Assume $\varnothing(s)$. We shall see that $x_1, x_2 \in s \supset .\, \varphi \equiv \mathrm{Rel}(s, \varphi)$. We have

$$\mathrm{Rel}\,(s, \varphi) \equiv (y)\,(y \in s \supset (\exists\, z)(z \in s \,.\, (u)\,(u \in s \supset (\exists\, t)\,(t \in s \,.\, \varphi^*))))$$

and by definition of $F_1 - F_4$

$$(10) \qquad \begin{aligned} (x_1, x_2, y, z, u, t)\,(x_1, x_2, y \in s \,.\, z \in F_1(x_1, x_2, y) \,.\, u \in s \\ .\, t \in F_2(x_1, x_2, y, z, u) : \supset \varphi^*) \end{aligned}$$

$$(11) \qquad \begin{aligned} (x_2, x_2, y, z, u, t)\,(x_1, x_2 \in s \,.\, y \in F_3\,(x_1, x_2) \,.\, z \in s \\ .\, u \in F_4(x_1, x_2, y, z) \,.\, t \in s : \supset \sim \varphi^*) \end{aligned}$$

If φ holds for $x_1, x_2 \in s$ then $F_1(x_1, x_2, y) \neq 0$ and $F_2(x_1, x_2, y, z, u) \neq 0$ for $y \in s$, $z \in F_1(x_1, x_2, y)$ and $u \in s$ and hence by (10) $\mathrm{Rel}\,(s, \varphi)$ holds for x_1, x_2. If $\sim \varphi$ holds for x_1, x_2 then we have by (11), in the same way, that $\mathrm{Rel}\,(s, \sim \varphi)$ holds, i.e., $\sim \mathrm{Rel}\,(s, \varphi)$ holds.

Since we have always $\sum_{\nu<\mu} R(\alpha_\nu) = R(\sup_{\nu<\mu} \alpha_\nu)$ and by the definition of the function K $K^n(x)$ is of the form $R(\beta)$ also $P(x) = \sum_{n\in\omega} K^n(x)$ is equal to $R(\alpha)$ for some α. Since $\varnothing(P(x))$ we have $\varnothing(R(\alpha))$. If we want α to be greater than μ it is enough to take $x = \{\mu\}$ and by $x \subseteq P(x)$ we have $\mu \in P(x) = R(\alpha)$, i.e., $\mu < \alpha$. Now let F be the normal function counting, in the order of their magnitude, the ordinals α which satisfy $\varnothing(R(\alpha))$. Since we have arbitrarily great ordinals α satisfying $\varnothing(R(\alpha))$ F is *d.f.a.o.* For ξ which is not a limit-number we have $\varnothing(R(F(\xi)))$. Let η be a limit-number, and let $x_1, x_2, y, z, u \in R(F(\eta))$. Let γ be the maximum of the ranks of x_1, x_2, y, z, u. Since η is a limit-number $F(\eta)$ is also a limit-number and therefore $\gamma < F(\eta)$. Since $F(\eta) = \lim_{\xi<\eta} F(\xi)$ there is an ordinal ξ, $\xi + 1 < \eta$, such that $\gamma \leq F(\xi + 1) < F(\eta)$, and hence $x_1, x_2, y, z, u \in R(F(\xi + 1))$. But, as we have already mentioned, $\varnothing(R(F(\xi + 1)))$ holds and therefore $F_1(x_1, x_2, y) \subseteq R(F(\xi + 1)) \subseteq R(F(\eta))$ and the same holds for $F_2 - F_4$. Thus we have proved $\varnothing(R(F(\eta)))$ also for limit-number η, hence $(\eta) \varnothing (R(F(\eta)))$.

By M the function $F(\eta)$ has in its range an inaccessible number α. Therefore we have $\varnothing(R(\alpha))$ and hence

$$(x_1, x_2)(x_1, x_2 \in R(\alpha) \supset . \varphi \equiv \mathrm{Rel}(R(\alpha), \varphi)).$$

N follows from Definition 1.

THEOREM 4. *In ZM it is provable that all the functions P_η are d.f.a.o. as well as the function Q.*

Proof. Let η be the least ordinal such that P_η is not *d.f.a.o.* and let α be the least ordinal for which $P_\eta(\alpha)$ is not defined. α cannot be a limit-number, since in that case $P_\eta(\alpha) = \lim_{\beta<\alpha} P_\eta(\beta)$. Let us "define" $P_\eta(\alpha)$ to be the class of all the ordinal numbers. By exactly the same arguments as those in the proof of Theorem 2 of Mahlo [4] (for the case $\alpha = \pi_{\mu,\nu}$ $\mu, \nu < \alpha$) we can define a normal function "converging to $P_\eta(\alpha)$" which does not have inaccessible values at limit-number arguments, i.e., we have a normal function *d.f.a.o.* which does not satisfy M'. Now that we proved that for each η P_η is *d.f.a.o.* Let $Q(0)$ be undefined. As in the former case we "define" $Q(0)$ to be the class of all ordinals and use the arguments in the proof of Theorem 2 of Mahlo [4] (for the case of the least ξ such that $\xi = \pi_{1,\xi}$) to construct a normal function *d.f.a.o.* which does not satisfy M'. In the same way we prove, by transfinite induction, the existence of $Q(\alpha)$ for each α.

Arguments which are very similar to those of Theorem 4 can be used in order to prove in ZM that all the functions Q_η are *d.f.a.o.* as well as the normal function counting the Q^*-numbers, and so on.

4. **An hierarchy of set theories.** In analogy with Mahlo [4] we can give axioms of infinity stronger than M.

DEFINITION 5. α is call a hyper-inaccessible number of type 1 if it is inaccessible with respect to ZM, i.e., if it is inaccessible and each normal function whose domain is α and whose range is included in α

has at least one inaccessible number in its range. α is hyper-inaccessible of type $\mu+1$ if it is inaccessible and each normal function whose domain is α and whose range is included in α has at least one hyper-inaccessible number of type μ in its range. For a limit-number μ α is hyper-inaccessible of type μ if it is hyper-inaccessible of type λ for every $\lambda < \mu$.[10]

It follows immediately from Definition 5 that if α is hyper-inaccessible of type μ it is also hyper-inaccessible of type λ for every $\lambda < \mu$.

Let Λ be a definite ordinal number. To avoid going into details we assume that existence and uniqueness of Λ are provable in ZF and also that it is provable in ZF that the definition of Λ is absolute with respect to standard complete models of ZF. Observe the following axiom schema:

M_Λ (for $\Lambda \geq 2$) *Every normal function d.f.a.o. has for every $\mu < \Lambda$ at least one hyper-inaccessible number of type μ in its range.*

Obviously we have that if $ZF \vdash \Lambda < M$ then M_M implies M_Λ. Let ZM_Λ denote the theory obtained from ZF by addition of M_Λ. By Definition 5 α is a hyper-inaccessible number of type Λ if and only if $R(\alpha)$ is a standard complete model of M_Λ (here we use the absoluteness of Λ with respect to standard complete models of ZF).

In complete analogy to Theorem 3 we have:

THEOREM 5. *M_Λ is equivalent in ZF to the schemata*

N_Λ''' $(\mu)(\mu < \Lambda \supset (\exists\alpha)(\alpha$ *is hyper-inaccessible of type* $\mu\,.\,(x_1, \cdots, x_n)(x_1, \cdots, x_n \in R(\alpha) \supset .\, \varphi \equiv \mathrm{Rel}(R(\alpha), \varphi))))$
where φ is any formula which has no free variables except $x_1, \cdots, x_n$.

and

N_Λ $(\mu)(\mu < \Lambda \supset (\exists u)(Scm^{ZM_\mu}(u)\,.$
$(x_1, \cdots, x_n)(x_1, \cdots, x_n \in u \supset .\, \varphi \equiv \mathrm{Rel}(u, \varphi))))$ *where φ is any formula which has no free variables except $x_1, \cdots, x_n$.*

By $Scm^{ZM_\mu}(u)$ we mean that u is a standard complete model of an axiom system like ZM_Λ only that in ZM_μ μ is taken as a parameter. Thus $Scm^{ZM_\mu}(u)$ is a formula with the two free variables μ and u.

By replacing Λ by $\Lambda+1$ in Theorem 5 we obtain easily that $M_{\Lambda+1}$ is equivalent to the schemata

$$(\exists\alpha)(\alpha \textit{ is hyper-inaccessible of type } \Lambda\,.$$
$$(x_1, \cdots, x_n)(x_1, \cdots, x_n \in R(\alpha) \supset .\, \varphi \equiv \mathrm{Rel}(\alpha), \varphi))$$

[10] The hyper-inaccessible numbers of type 1 correspond to the ρ_0-numbers of Mahlo [**4**]. The hyper-inaccessible numbers of type λ correspond to the members of the range of $\pi_{\alpha,0,\lambda}$ of Mahlo [**4**].

and

$$(\exists\, u)\,(Scm^{ZM_A}(u)\,.\,(x_1, \cdots, x_n)\,(x_1, \cdots, x_n \in u \supset .\, \varphi \equiv \mathrm{Rel}\,(R(\alpha), \varphi))).$$

Now we shall see that the same relation which holds between ZF and ZM, and between ZM_A and ZM_{A+1} holds also between S and ZF.

THEOREM 6. *In S the axiom schema of replacement in conjunction with the axiom of infinity is equivalent to the schema*

N_0 $\quad (\exists\, u)\,(Scm^S(u)\,.\,(x_1, \cdots, x_n)\,(x_1, \cdots, x_n \in u \supset .\, \varphi \equiv \mathrm{Rel}\,(u, \varphi)))$ *where φ is a formula which does not contain free variables except $x_1, \cdots, x_n$.*

Proof. That N_0 is provable in ZF is Montague's theorem proved in [7] and it is proved by the same method as the corresponding part of Theorem 3.

Now we assume N_0 and prove the axioms of infinity and replacement. By N_0, taking any φ, we obtain $(\exists\, u)\, Scm^S(u)$. This u obviously satisfies the requirements of the axiom of infinity. Now, given $\varphi(v, w)$ with the only free variables $v, w, x_1, \cdots, x_n$ let χ denote the formula

$$(r, s, t)\,(\varphi(r, s)\,.\,\varphi(r, t) : \supset s = t) \supset (\exists\, y)\,(w)\,(w \in y \equiv (\exists\, v)\,(v \in x\,.\,\varphi\,(v, w)))$$

By N_0 we have, as in Theorem 2, that there exists a set u such that $Scm^S(u)$ and

(12) $\quad x_1, \cdots, x_n, v, w \in u \supset .\, \varphi \equiv \mathrm{Rel}\,(u, \varphi)$
(13) $\quad x_1, \cdots, x_n, v \in u \supset .\, (\exists\, w)\, \varphi \equiv \mathrm{Rel}\,(u, (\exists\, w)\, \varphi)$
(14) $\quad x_1, \cdots, x_n, x \in u \supset .\, \chi \equiv \mathrm{Rel}\,(u, \chi)$
(15) $\quad (x_1, \cdots, x_n)\,(x)\, \chi \equiv \mathrm{Rel}\,(u, (x_1, \cdots, x_n)\,(x)\, \chi)$

Since $\mathrm{Rel}\,(u,(\exists\, w)\, \varphi)$ is $(\exists\, w)\,(w \in u\,.\,\mathrm{Rel}\,(u, \varphi))$ we have by (12) and (13) $x_1, \cdots, x_n, v \in u \supset .\, (\exists\, w)\,(w \in u\,.\, \varphi) \equiv (\exists\, w)\, \varphi$; hence if $(r, s, t,)$ $(\varphi(r, s)\,.\,\varphi(r, t) : \supset s = t)$ then for $x \in u$, since $Scm^S(u)$ implies that then $x \subseteq u$, the function represented by $\varphi(v, w)$ maps the members of x on members of u, and therefore, by the axiom of subsets, that function maps x on some set y. Thus we have $x_1, \cdots, x_n, x \in u \supset \chi$ and by (14) $x_1, \cdots, x_n, x \in u \supset \mathrm{Rel}\,(u, \chi)$; but the closure of the latter formula is $\mathrm{Rel}\,(u, (x_1, \cdots, x_n)\,(x)\chi)$ and hence, by (15), we have χ.

By Theorem 6 we can view the axiom schemata M and M_A as natural continuations of the axioms of infinity and replacement. Therefore, although the consistency of ZF does not imply, even in ZM (if ZM is consistent), the consistency of ZM, it seems likely that if in the sequence $S, ZF, ZM, ZM_2, \cdots$ no inconsistency is introduced in the first step, from S to ZF, also no inconsistency is introduced in the further steps.

In the following definitions and statements we essentially follow Montague in [7].

Let the theory Q be an extension of the theory P. Let φ be any sentence of Q. $P + \{\varphi\}$ denotes the theory obtained from P by adding to it φ as a new axiom. $\mathrm{Con}\,(P + \{\varphi\})$ is the arithmetic sentence which asserts the consistency of $P + \{\varphi\}$. Q is called essentially reflexive over P if for every sentence φ of Q $\varphi \supset \mathrm{Con}\,(P + \{\varphi\})$ is a theorem of Q. Q is called an essentially infinite extension of P if no consistent extension of Q without new symbols is obtained from P by adding to it a finite number of axioms. If Q is essentially reflexive over P then Q is an essentially infinite extension of P. By the same argument as that of Montague in [7] each of the theories $S.\, ZF, ZM, \cdots$ is essentially reflexive over the preceding ones.

Let $E_{R(\alpha)} = \{\langle xy \rangle\, ;\, x \in y\, .\, x, y \in R(\alpha)\}$, $A_\alpha = \langle R(\alpha), E_{R(\alpha)} \rangle$. Montague and Vaught proved in [8] that if $\beta < \alpha$ and $R(\alpha)$ is an arithmetical extension of $R(\beta)$ (i.e., for any formula φ with no free variables except $x_1, \cdots, x_n$

$$(x_1, \cdots, x_n)\,(x_1, \cdots, x_n \in R(\beta) \supset .\ \mathrm{Rel}\,(R(\alpha), \varphi) \equiv \mathrm{Rel}\,(R(\beta), \varphi)))$$

then both A_α and A_β are models of ZF (in the sense of models of the type S_4 of Tarski [16]). [11]

THEOREM 7. *If A_α and A_β are as mentioned above and β is inaccessible then both A_α and A_β are models of ZM. If β is hyper-inaccessible of type Λ then both A_α and A_β are models of $ZM_{\Lambda+1}$.*

The proof that A_α is a model as required is exactly like the second part of the proof of Theorem 3. A_β is also a model as required since if φ holds in A_α it holds in A_β.

Another aspect of the phenomena discovered by Montague and Vaught in [7] and [8] is the following theorem:

THEOREM 8. *Let Sb be a theory with the same language and axioms as S with the additional set-constant b and the additional axioms*

(16) $Scm^S(b)$

(17) $(x_1, \cdots, x_n)\,(x_1, \cdots, x_n \in b \supset .\ \varphi \equiv \mathrm{Rel}\,(b, \varphi))$ *where φ is any formula of S without free variables except $x_1, \cdots, x_n$.*

The theorems of Sb which do not contain the constant b are exactly the

[11] This and the following Theorem 7 can be read in two different ways. Either we take the theorems and proofs informally, in which case all the notions retain their verbal meaning; or that the theorems are taken to be formal theorems of S and then the notions of model and arithmetical extension are formal notions defined by means of the formal notion of satisfaction, which is given, for example, in Mostowski [**11**].

theorems of ZF*; the theorems of* $Sb + \{Scm^{ZF}(b)\}$ *which do not contain* b *are exactly the theorems of* ZM*; and the theorems of* $Sb + \{(\mu)(\mu < \Lambda \supset Scm^{ZM_\mu}(b))\}$ *which do not contain* b *are exactly the theorems of* ZM_Λ *(the theorems of* $Sb + \{Scm^{ZM_\Lambda}(b)\}$ which do not contain b are exactly the theorems of $ZM_{\Lambda+1}$).

Proof. Every theorem of ZF is provable in Sb since Sb contains the axioms of S and all the instances of N_0 are obviously provable in Sb. Now let the sentence χ be a theorem of Sb which does not contain b. Let $\varnothing(b)$ be the conjunction of all the instances of (16) and (17) used in the proof of χ. By the deduction theorem $\varnothing(b) \supset \chi$ is provable from the axioms of S, hence $(\exists\, u)\, \varnothing\, (u) \supset \chi$ is provable in S. But Montague's theorem (Theorem 6) $(\exists\, u)\, \varnothing\, (u)$ is a theorem of ZF, hence χ is provable in ZF.

The other statements of Theorem 8 follow in the same way from Theorems 3 and 5.

We see, by Theorem 8, that even though in the sequence ZF, ZM, $ZM_2, \cdots$ each theory is an essentially infinite extension of all the preceding ones we can get a corresponding sequence Sb, $Sb + \{Scm^{ZF}(b)\}$, $Sb + \{Scm^{ZM}(b)\}, \cdots$ in which the theories which are "almost the same" as the respective theories in the former sequence, and in which all the theories are obtained from the first one by the addition of respective single axioms.

5. **Peculiar behavior of models.** We shall now see examples illustrating the inadequacy for general use of the notion of standard model introduced in § 2. In our examples we shall use a formal satisfaction definition. The idea of using the formalized notion of satisfaction in these problems and the special way in which that notion is given here are due to Mostowski.[12] Our notations will be those of Mostowski [10].

Our first example will be an axiomatic representation ZF^* of ZF which has no standard model.

Let Φ_n be the nth formula in a given Gödelization of ZF. Given the functional variable $p(i, f)$ we shall construct a formula $\Psi(p)$ which asserts that $p(i, f)$ is a satisfaction definition.

$p(i, f)$ is a satisfaction definition if the following holds for every finite number i and every finite sequence of sets f:

(a) If Φ_i is the formula $x_k = x_j$ or $x_k \in x_j$ then $p(i, f)$ if and only if $D(f) = \{k, j\}$ and $f(k) = f(j)$ or $f(k) \in f(j)$, respectively.

(b) If $\Phi_i = \Phi_j \mid \Phi_h$ then
$p(i, f) \equiv : D(f) = s_i \,.\, \sim p(j, f/s_j) \vee \sim p(h, f/s_h)$.

(c) If $\Phi_i = (\exists\, x_m)\, \Phi_j$ and x_m is free in Φ_j then $p(i, f) \equiv : D(f) = s_i$ $.(\exists\, a)\, p(j, f + \{\langle ma \rangle\})$. If x_m is not free in Φ_j then $p(i, f) \equiv p(j, f)$.

[12] By oral communication.

This inductive definition can be replaced by an explicit one in the usual method and thus we get the required formula $\Psi(p)$ which asserts that p is a definition of satisfaction.

Now substitute for p in $\Psi(p)$ any formula φ of ZF. Assume $\Psi(\varphi)$, then by the usual methods, e,g., those of Mostowski [10] pp. 114-115, we obtain a truth definition for ZF in ZF and we arrive at the Tarski contradiction. Thus we have proved in $ZF \sim \Psi(\varphi)$ for any φ of ZF. Therefore we can add the axiom schema

$$\sim \Psi(\varphi) \quad \textit{for any } \varphi$$

to ZF without changing the theory and we call the new array of axioms ZF^*. The sets u and e form a standard model of ZF^* if $Sm^{ZF}(u, e)$ and there exists no subset v of u of ordered pairs $\langle if \rangle$ such that the formula obtained from the relativization of $\Psi(p)$ to the model by substituting $\langle if \rangle \in v$ for $p(i, f)$ holds. But form $Sm^{ZF}(u, e)$ it is easy to prove (in S) the existence of such a subset v of u, e.g., by the methods of Mostowski [11]. Hence ZF^* has no standard model. In other words, $\sim \Psi(p)$ is a true statement of set theory if p varies over the relations expressible in the set theory itself, but $\sim \Psi(p)$ is not a true statement if p varies over all the relations.

We shall now sketch briefly a second example. This will be a theory T which contains all the theorems of ZF, but has more standard complete models than ZF.

Mostowski defines in [10] when a class F of ordered pairs $\langle if \rangle$ is called an S-sequence for the formula Φ_j. This definition can be formulated without class variables, except F. Therefore, using the analogy between classes and functional variables, we can define, using only set variables beside p, when the functional variable $p(i, f)$ is an S-sequence for Φ_j. Let $Stf(u, i, f)$ be a formula which asserts that the finite sequence of sets f satisfies Φ_i in the complete model u (for the existence of such a formula cf. Mostowski [11]). We consider the following formula $\Omega(p)$ *p is an S-sequence for $\Phi_i \supset (\exists u)(Scm^S(u)\,.\,(f)(f$ is a finite sequence of sets whose range is in $u \supset .\, p(i, f) \equiv Stf(u, i, f)))$.* If we add to S the schema $\Omega(\varphi)$ where φ is any formula of S then we get a theory T which is an extension of ZF since all the instance of N_0 are provable in T (to prove the instance of N_0 corresponding to the formula φ with Gödel-number i we write down an S-sequence χ for Φ_i — this can be done by Mostowski [10] Σ_4 — and $\Omega(\chi)$ implies $x_1, \cdots, x_n \in u \supset . \varphi \equiv \mathrm{Rel}\,(u, \varphi))$. We shall now see that every standard complete model of ZF is a standard complete model of T but there are standard complete models of T with universes of smaller cardinality than that of and standard complete model of ZF. That every standard complete model of ZF is a standard complete model of T is the formal counterpart of Montague's theorem (that the axioms of infinity and replacement imply N_0). Now let τ be the first inaccessible number. By Montague and Vaught [8] there exists an ordinal number $\alpha < \tau$ such that $R(\alpha)$ is the union of the sets definable in the model A_τ and in conseqence A_τ is an arithmetical extension of A_α. In exactly the same way we can prove that there exists an ordinal β $\alpha < \beta < \tau$ such that $R(\beta)$ is the union of all the sets definable in the model A_τ by means of the new constant α, and in consequence A_τ is also an arithmetical extension of A_β. Hence, by Theorem 1.8 of [17], A_β is an arithmetical extension of A_α. It is easily seen that A_β is a standard model of T, where u required in the schema $\Omega(\varphi)$ is always taken to be $R(\alpha)$.

BIBLIOGRAPHY

1. H. Bachmann, *Transfinite Zahlen*, Berlin 1955.
2. A. A. Fraenkel and J. Bar-Hillel, *Foundations of set theory*, Amsterdam 1958.
3. D. Hilbert and P. Bernays, *Grundlagen der Mathematik*, Part I, Berlin 1934.
4. P. Mahlo, *Uber lineare transfinite Mengen*, Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematish-Physische Klasse, **63** (1911), 187-225.
5. ———, *Zur Theorie und Anwendung der ρ_0-Zahlen*, ibid, **64** (1912), 108-112.
6. ———, *Zur Theorie und Anwendung der ρ_0-Zahlen* II, ibid, **65** (1913), 268-282.
7. R. Montague and S. Feferman, *The method of arithmetization and some of its applications*, North Holland Pub. Co., to appear. See abstracts by R. Montague, Bull. of the A.M.S., **61** (1955), p. 172, **62** (1956), p. 260, *Summaries of talks at the Summer Institute of Symbolic Logic in* 1957 *at Cornell University*, pp. 256-259.
8. R. Montague and R. L. Vaught, *Natural models of set theories*, to appear in Fundamenta Mathematicae.
9. A. Mostowski, *An undecidable arithmetical statement*, Fundamenta Mathematicae, **36** (1949), 143-164.
10. ———, *Some impredicative definitions in the axiomatic set theory*, ibid, **37** (1950), 111–124.
11. ———, *On models of axiomatic systems*, ibid, **39** (1952), 133-158.
12. J. C. Shepherdson, *Inner models for set theory*, Journal of Symbolic Logic, Part I - **16** (1951), 161-190; Part II - **17** (1952), 225-237; Part III - **18** (1953), 145-167.
13. E. Specker, *Zur Axiomatik der Mengenlehre* (*Fundierugs und Auswahlaxiom*), Zeitschrift für mathematische Logik und Grundlagen der Mathematik, **3** (1957), 173-210.
14. A. Tarski, *Uber unerreichbare Kardinalzahlen*, Fundamenta Mathemticae, **30** (1938), 68–89.
15. ———, *On well-ordered subsets of any set*, ibid, **32** (1939), 176-183.
16. ———, *Notions of proper models for set theoris* (abstract), Bulletin of the A.M.S., **62** (1956), p. 601.
17. A. Tarski and R. L. Vaught, *Arithmctical extensions of relational systems*, Compositio Mathematica, **13** (1957), 81-102.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
AND
HEBREW UNIVERSITY, JERUSALEM

ON CERTAIN SINGULAR INTEGRALS

BENJAMIN MUCKENHOUPT

1. Introduction. The purpose of this paper is to consider a modification of the Hilbert transform and the singular integrals treated by Calderon and Zygmund in [1] and [3], and to use the results to generalize some standard results on fractional integration. In the one dimensional case the Hilbert transform of a function $f(x)$ is essentially the integral $\int_{-\infty}^{\infty} \frac{f(x-t)}{t} dt$. In the one dimensional case the transform to be considered will be a convolution with $\frac{1}{|t|^{1+i\gamma}}$ instead of with $\frac{1}{t}$. Throughout this paper γ will denote a real number not zero. As in the Hilbert transform case there is trouble with the definition; for the Hilbert transform this is solved by taking a Cauchy value at the origin. The obvious extension of this method was used by Thorin [6] when he considered a transform of the type

$$\lim_{\varepsilon \to 0} \int_{\varepsilon}^{\infty} \frac{f(x-t) - f(x+t)}{t^{1+i\gamma}} dt .$$

Here and subsequently ε will always be greater than 0 and the limits in ε will be one sided. In this case, however, obtaining cancellation by taking a Cauchy value is unnecessary; the kernel already has sufficient oscillations to accomplish this. The integral $\lim_{\varepsilon \to 0} \int_{\varepsilon}^{\infty} \frac{f(x-t)}{t^{1+i\gamma}} dt$ will not, in general, exist, but by using some suitable summation procedure, it may be given meaning. Starting with two such methods, it is shown that this transform has the usual singular integral properties. Specifically, for functions in a Lebesgue L^p class $1 < p < \infty$, it is shown that the summation procedure converges in L^p and that the resulting transformation is bounded in L^p. For $p = 1$ substitute results are obtained. Furthermore, for functions in $L^p, 1 \leq p < \infty$, the summation procedure is shown to converge pointwise almost everywhere.

Carried along simultaneously with the preceding is the n dimensional extension of the sort considered by Calderón and Zygmund for the Hilbert transform. In Euclidean n space, E^n, let $x = (x_1, x_2 \cdots x_n)$, $|x| = (x_1^2 + \cdots x_n^2)^{\frac{1}{2}}$ and $dx = dx_1 \cdots dx_n$. The transforms to be considered are of the form

$$\int_{E^n} \frac{f(x-t)}{|t|^{n+i\gamma}} \Omega(t) \, dt .$$

Received April 27, 1959.

where $\Omega(t) = \Omega\left(\frac{t}{|t|}\right)$ is integrable on the unit sphere, and the integral in the neighborhood of the origin is again defined by a suitable summation method. In this case, unlike the Calderón and Zygmund results, the integral of $\Omega(t)$ on the unit sphere need not be zero. Again for functions in L^p, $1 < p < \infty$, the summation procedure converges in L^p, pointwise almost everywhere, and the resulting transformation is bounded in L^p. Substitute results for L^1 including pointwise convergence are also proved although for some it must be assumed that $\Omega(t)$ satisfies a continuity condition. The method used to obtain all these results is first to reduce the summation definition to one more closely resembling the Cauchy value definition of ordinary singular integrals. After this, lemmas similar to some lemmas in [1] make the methods of [1] and [3] applicable to these transformations.

In the last section the preceding results and an interpolation theorem of Stein [4] are used to prove the following theorem.

Let p, q, and λ be positive numbers such that $1 < p < q < \infty$ and $\frac{1}{p} = \frac{1}{q} + \lambda$. Let $f(x)$ be in L^p in E^n and let $\Omega(t) = \Omega\left(\frac{t}{|t|}\right)$ be in L^s, $s = \frac{1}{1-\lambda}$, on the unit sphere. Then the integral

$$D_\lambda(f) = \int_{E^n} \frac{\Omega(t) f(x-t)}{|t|^{n(1-\lambda)}}\, dt$$

exists for almost all x and $\|D_\lambda(f)\|_q \leq A\|f\|_p$ where A is independent of f.

For $\Omega(t) = 1$ this is a well known theorem on fractional integrals. See for example [5]. Substitute results are also obtained for $p = 1$ and $q = \infty$ using the proof for the weaker results in [8].

2. Summation.

A summation method for the integral $\lim_{\varepsilon \to 0} \int_\varepsilon^1 f(x)\, dx$ of the form $\lim_{\varepsilon \to 0} \int_0^1 \varphi_\varepsilon(\alpha) d\alpha \int_\alpha^1 f(x)\, dx$ is a *regular* method if

$$\lim_{\varepsilon \to 0} \int_a^1 |\varphi_\varepsilon(\alpha)| d\alpha = 0 \text{ for } a > 0,\ \lim_{\varepsilon \to 0} \int_0^1 \varphi_\varepsilon(\alpha)\, d\alpha = 1 \text{ and } \int_0^1 |\varphi_\varepsilon(\alpha)| d\alpha \leq B\,.$$

LEMMA 1. *If $\lim_{\varepsilon \to 0} \int_\varepsilon^1 f(x) dx$ exists, then any regular method of summation will give the same limit.*

This is a standard fact about these summation methods.

LEMMA 2. *If $\int_\varepsilon^1 f(x, y)\, dy$ converges in L^p norm to $g(x)$ as $\varepsilon \to 0$*

and has a uniformly bounded L^p norm, then any regular summation method will also converge to $g(x)$ in L^p norm.

Let B be a bound for $\int_0^1 |\varphi_\varepsilon(\alpha)| d\alpha$ and C a bound for $\left\|\int_\varepsilon^1 f(x,y)\,dy\right\|_p$. Then given $\eta > 0$, choose β so that

$$\left\|g(x) - \int_\delta^1 f(x,y)\,dy\right\|_p \le \frac{\eta}{3B}$$

for $\delta \le \beta$ and γ so that

$$\int_\beta^1 |\varphi_\varepsilon(\alpha)| d\alpha \le \frac{\eta}{6C}$$

and

$$\left|\int_0^1 \varphi_\varepsilon(\alpha)\,d\alpha - 1\right| \le \frac{\eta}{3C}$$

provided that $\varepsilon \le \gamma$. The existence of β and γ follows from the hypotheses of the lemma.

$$\begin{aligned}
&\text{If } \varepsilon \le \gamma, \text{ then } \left\|g(x) - \int_0^1 \varphi_\varepsilon(\alpha)d\alpha \int_\alpha^1 f(x,y)dy\right\|_p \\
&= \left\|g(x)\Big(1 - \int_0^1 \varphi_\varepsilon(\alpha)\,d\alpha\Big) + \int_0^\beta \varphi_\varepsilon(\alpha)\Big(g(x) - \int_\alpha^1 f(x,y)\,dy\Big)\,d\alpha \right. \\
&\qquad \left. + \int_\beta^1 \varphi_\varepsilon(\alpha)\Big(g(x) - \int_\alpha^1 f(x,y)\,dy\Big)d\alpha\right\|_p \\
&\leqq \|g(x)\|_p \frac{\eta}{3C} + \int_0^\beta |\varphi_\varepsilon(\alpha)| \left\|g(x) - \int_\alpha^1 f(x,y)\,dy\right\|_p d\alpha \\
&\qquad + \int_\beta^1 |\varphi_\varepsilon(\alpha)| \left\|g(x) - \int_\alpha^1 f(x,y)\,dy\right\|_p d\alpha
\end{aligned}$$

by use of Minkowski's inequality and Minkowski's integral inequality. Observing that $\|g(x)\|_p$ is also less than or equal to C, this last expression is clearly less than or equal to $\frac{C\eta}{3C} + \frac{B\eta}{3B} + \frac{\eta}{6C}2C = \eta$. Since η was arbitrary, the lemma follows.

3. Definitions of the transform. To give meaning to the integral

$$\hat{f}(x) = \int_0^\infty \frac{f(x-t)}{t^{1+i\gamma}}\,dt$$

it may be written as

$$(S)\int_0^1 \frac{f(x-t)}{t^{1+i\gamma}}\,dt + \int_1^\infty \frac{f(x-t)}{t^{1+i\gamma}}\,dt$$

where the first integral must be obtained by using a suitable method of summation. For this purpose logarithmic Abel summation defined by

$$(S)\int_0^1 g(t)\,dt = \lim_{\varepsilon\to 0}\int_0^1 t^{\varepsilon} g(t)\,dt$$

or logarithmic Cesaro summation defined by

$$(S)\int_0^1 g(t)dt = \lim_{\varepsilon\to 0}\int_\varepsilon^1 g(t)\Big(1 - \frac{\log t}{\log \varepsilon}\Big)dt$$

may be used. Both are regular methods for they may be written as

$$\lim_{\varepsilon\to 0}\int_0^1 \varphi_\varepsilon(\alpha)\,d\alpha\int_\alpha^1 g(t)\,dt$$

where $\varphi_\varepsilon(\alpha) = \varepsilon\alpha^{\varepsilon-1}$ in the case of logarithmic Abel summation and

$$\varphi_\varepsilon(\alpha) = \begin{cases} \dfrac{-1}{\alpha\log\varepsilon} & \varepsilon \le \alpha \le 1 \\ 0 & 0 \le \alpha < \varepsilon \end{cases}$$

for logarithmic Cesaro summation. That these satisfy the necessary conditions is clear from their forms.

In either case $\tilde{f}(x)$ may be written as

$$\begin{aligned}(S)&\int_0^1 \frac{f(x-t)}{t^{1+i\gamma}}dt + \int_1^\infty \frac{f(x-t)}{t^{1+i\gamma}}dt \\ &= (S)\int_0^1 \frac{f(x-t)-f(x)}{t^{1+i\gamma}}\,dt - \frac{f(x)}{i\gamma} + \int_1^\infty \frac{f(x-t)}{t^{1+i\gamma}}\,dt\,.\end{aligned}$$

By the first lemma the existence of this expression can be shown by proving the existence of

$$\lim_{\varepsilon\to 0}\int_\varepsilon^1 \frac{f(x-t)-f(x)}{t^{1+i\gamma}}\,dt\,.$$

Therefore, showing the convergence almost everywhere of the expression

$$\begin{aligned}&\lim_{\varepsilon\to 0}\int_\varepsilon^1 \frac{f(x-t)-f(x)}{t^{1+i\gamma}}dt - \frac{f(x)}{i\gamma} + \int_1^\infty \frac{f(x-t)}{t^{1+i\gamma}}\,dt \\ &\quad = \lim_{\varepsilon\to 0}\left[\int_\varepsilon^\infty \frac{f(x-t)}{t^{1+i\gamma}}dt - \frac{f(x)\varepsilon^{-i\gamma}}{i\gamma}\right]\end{aligned} \tag{3.1}$$

will imply convergence almost everywhere for the original definition of $\tilde{f}(x)$. Furthermore, by Lemma 2 the convergence in L^p norm of (3.1) will imply the convergence in L^p norm of the original definition of $\tilde{f}(x)$.

4. Convergence in L^2 norm. Define

$$K_{N,\varepsilon}(t) = \begin{cases} \dfrac{1}{t^{1+i\gamma}} & \varepsilon \leq t \leq N \\ \\ 0 & \text{elsewhere} \end{cases} \text{, and let}$$

$$\tilde{f}_{N,\varepsilon}(x) = \int_{-\infty}^{\infty} f(x-t) K_{N,\varepsilon}(t)\, dt - \frac{f(x)\varepsilon^{-i\gamma}}{i\gamma}\,.$$

Then the transform of $f(x)$ defined before, $\tilde{f}(x) = \lim_{\varepsilon\to 0} \lim_{N\to\infty} \tilde{f}_{N,\varepsilon}(x)$ if this last limit exists. Now if $f(x)\varepsilon L^2$, it is possible to take Fourier transforms and obtain

$$\hat{\tilde{f}}_{N,\varepsilon}(x) = \hat{f}(x)\left(\hat{K}_{N,\varepsilon}(x) - \frac{\varepsilon^{-i\gamma}}{i\gamma}\right)$$

where $\hat{g}(x)$ denotes the Fourier transform of $g(x)$.

LEMMA 3. *The expression $\hat{K}_{N,\varepsilon}(x) - \dfrac{1}{i\gamma\varepsilon^{i\gamma}}$ is in absolute value less than $c(\gamma) = C\dfrac{(|\gamma|+1)^2}{|\gamma|}$ where C is an absolute constant. As $M \to \infty$ the expression converges to a function $\hat{K}_\varepsilon(x)$ except for $x = 0$. Furthermore, as $\varepsilon \to 0$, $\hat{K}_\varepsilon(x)$ converges to a function $\hat{K}(x)$ except for $x = 0$.*

From its definition $\hat{K}_{N,\varepsilon}(x) - \dfrac{1}{i\gamma\varepsilon^{i\gamma}}$ is equal to

$$\int_\varepsilon^N \frac{e^{ixt}}{t^{1+i\gamma}}\, dt - \frac{1}{i\gamma\varepsilon^{i\gamma}} = |x|^{i\gamma}\int_{\varepsilon|x|}^{N|x|} \frac{e^{it \operatorname{sgn} x}}{t^{1+i\gamma}}\, dt - \frac{\varepsilon^{-i\gamma}}{i\gamma}\,.$$

Now

$$\int_a^b \frac{e^{it \operatorname{sgn} x}}{t^{1+i\gamma}}\, dt = \frac{e^{it \operatorname{sgn} x}}{i \operatorname{sgn} x t^{1+i\gamma}}\bigg|_a^b + \frac{1+i\gamma}{i \operatorname{sgn} x}\int_a^b \frac{e^{it \operatorname{sgn} x}}{t^{2+i\gamma}}\, dt \tag{4.1}$$

and

$$\int_a^b \frac{e^{it \operatorname{sgn} x}}{t^{1+i\gamma}}\, dt = \frac{e^{it \operatorname{sgn} x}}{-i\gamma t^{i\gamma}}\bigg|_a^b + \frac{\operatorname{sgn} x}{\gamma}\int_a^b \frac{e^{it \operatorname{sgn} x}}{t^{i\gamma}}\, dt\,. \tag{4.2}$$

If necessary, split the integral

$$\int_{\varepsilon|x|}^{N|x|} \frac{e^{it \operatorname{sgn} x}}{t^{1+i\gamma}}\, dt$$

into two parts, the first with limits less than or equal to one, and the second with limits greater than or equal to one. Then applying (4.2) to the first part and (4.1) to the second part, it is clear that the whole integral is in absolute value less than $C\frac{(|\gamma|+1)^2}{|\gamma|}$ for some absolute constant C.

Using (4.1), it is clear that

$$\lim_{N\to\infty}\hat{K}_{N,\varepsilon}(x)-\frac{1}{i\gamma\varepsilon^{i\gamma}}=|x|^{i\gamma}\Big(\int_{\varepsilon|x|}^{1}\frac{e^{it\operatorname{sgn}x}}{t^{1+i\gamma}}dt+\lim_{N\to\infty}\frac{e^{it\operatorname{sgn}x}}{it^{1+i\gamma}\operatorname{sgn}x}\Big|_1^{N|x|}$$
$$+\lim_{N\to\infty}\frac{1+i\gamma}{i\operatorname{sgn}x}\int_1^{N|x|}\frac{e^{it\operatorname{sgn}x}}{t^{2+i\gamma}}\Big)-\frac{1}{i\gamma\varepsilon^{i\gamma}}$$

and the two limits certainly exist. From this and (4.2) it follows that

$$\lim_{\varepsilon\to 0}\hat{K}_\varepsilon(x)=\lim_{\varepsilon\to 0}\Big[|x|^{i\gamma}\Big(\frac{e^{it\operatorname{sgn}x}}{-i\gamma t^{i\gamma}}\Big|_{\varepsilon|x|}^{1}+\int_{\varepsilon|x|}^{1}\frac{e^{it\operatorname{sgn}x}}{\gamma t^{i\gamma}\operatorname{sgn}x}dt-\frac{e^{i\operatorname{sgn}x}}{i\operatorname{sgn}x}$$
$$+\frac{1+i\gamma}{i\operatorname{sgn}x}\int_1^{\infty}\frac{e^{it\operatorname{sgn}x}}{t^{2+i\gamma}}dt\Big)-\frac{1}{i\gamma\varepsilon^{i\gamma}}\Big].$$

The limit of the integral clearly exists. The lower limit on the first integrated part and the last term combined give

$$\lim_{\varepsilon\to 0}|x|^{i\gamma}\Big(-\frac{|x|^{-i\gamma}e^{i\varepsilon x}}{-i\gamma\varepsilon^{i\gamma}}\Big)-\frac{1}{i\gamma\varepsilon^{i\gamma}}=\lim_{\varepsilon\to 0}\frac{\varepsilon^{-i\gamma}}{i\gamma}(e^{i\varepsilon x}-1)=0\,.$$

It follows that

$$\begin{aligned}\hat{K}(x)=\lim_{\varepsilon\to 0}\hat{K}_\varepsilon(x)=|x|^{i\gamma}\Big(&\frac{ie^{i\operatorname{sgn}x}}{\gamma}+\frac{ie^{i\operatorname{sgn}x}}{\operatorname{sgn}x}+\int_0^1\frac{e^{it\operatorname{sgn}x}}{\gamma t^{i\gamma}\operatorname{sgn}x}dt\\ &+\frac{1+i\gamma}{i\operatorname{sgn}x}\int_1^\infty\frac{e^{it\operatorname{sgn}x}}{t^{2+i\gamma}}dt\Big)\end{aligned}\tag{4.3}$$

COROLLARY 1. *If $f(x)$ belongs to L^2, then the transformation $\tilde{f}_\varepsilon(x)=\int_\varepsilon^\infty\frac{f(x-t)}{t^{1+i\gamma}}dt-\frac{f(x)}{i\gamma\varepsilon^{i\gamma}}$ satisfies $\|\tilde{f}_\varepsilon(x)\|_2\le c(\gamma)\|f(x)\|_2$. As $\varepsilon\to 0$, $\tilde{f}_\varepsilon(x)$ converges in L^2 norm to a function $\tilde{f}(x)$ which also satisfies $\|\tilde{f}(x)\|_2\le c(\gamma)\|f(x)\|_2$*

The expression $\Big(\hat{K}_{N,\varepsilon}(x)-\frac{1}{i\gamma\varepsilon^{i\gamma}}\Big)\hat{f}(x)$ converges in L^2 norm to $\hat{K}_\varepsilon(x)\hat{f}(x)$ because the first part of the product converges boundedly. Consequently, taking Fourier transforms, $\tilde{f}_{N,\varepsilon}(x)$ converges in L^2 norm to $\tilde{f}_\varepsilon(x)$. Similarly, since $\hat{K}_\varepsilon(x)\ \hat{f}(x)$ converges in L^2 norm, the Fourier transform, $\tilde{f}_\varepsilon(x)$, converges in L^2 norm to a function $\tilde{f}(x)$. The statements concerning the norms follow immediately from the estimate in Lemma 3.

For later proofs there is a more convenient form for $\hat{K}(x)$. Adding the identity

$$0 = -\frac{|x|^{i\gamma}}{i\gamma} + \frac{|x|^{i\gamma}e^{i\,\mathrm{sgn}\,x}}{i\gamma} - \frac{|x|^{i\gamma}\,\mathrm{sgn}\,x}{\gamma}\int_0^1 e^{it\,\mathrm{sgn}\,x}\,dt$$

to (4.3) gives

$$\hat{K}(x) = \frac{-|x|^{i\gamma}}{i\gamma} + |x|^{i\gamma}\Big[\mathrm{sgn}\,x\int_0^1 \frac{t^{-i\gamma}-1}{\gamma}e^{it\,\mathrm{sgn}\,x}\,dt \tag{4.4}$$
$$+\, i e^{i\,\mathrm{sgn}\,x}\,\mathrm{sgn}\,x + \frac{1+i\gamma}{i\,\mathrm{sgn}\,x}\int_1^\infty \frac{e^{it\,\mathrm{sgn}\,x}}{t^{2+i\gamma}}\,dt\Big].$$

Now for $|\gamma| \le 1$ the expression in brackets is uniformly bounded. This is obvious for the last two terms. Furthermore,

$$\left|\frac{t^{-i\gamma}-1}{\gamma}\right| = \left|\frac{\log t\int_0^\gamma t^{-iu}\,du}{\gamma}\right| \le |\log t|$$

so that the first integral is also uniformly bounded. This leads to the following.

COROLLARY 2. *The transform* $\tilde{F}_\gamma(x) = \lim_{\varepsilon\to 0}\frac{1}{\pi}\int_{|t|>\varepsilon}\frac{f(x-t)\,\mathrm{sgn}\,t}{|t|^{1+i\gamma}}dt$, $|\gamma| \le 1$, *satisfies* $\|\tilde{F}_\gamma(x)\|_2 \le A\|f(x)\|_2$ *where* A *is independent of* γ *and* f. *As* $\gamma \to 0$, $\tilde{F}_\gamma(x)$ *converges in* L^2 *to the ordinary Hilbert transform of* $f(x)$.

$\tilde{F}_\gamma(x)$ may be written in the form

$$\lim_{\varepsilon\to 0}\frac{1}{\pi}\int_\varepsilon^\infty \frac{f(x-t)}{t^{1+i\gamma}}dt - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}} - \int_{-\infty}^{-\varepsilon}\frac{f(x-t)\,dt}{(-t)^{1+i\gamma}} + \frac{f(x)}{i\gamma\varepsilon^{i\gamma}}.$$

Now observing that

$$\int_{-N}^{-\varepsilon}\frac{e^{ixt}}{(-t)^{1+i\gamma}}dt = \int_\varepsilon^N \frac{e^{i(-x)t}}{t^{1+i\gamma}}\,dt,$$

it is clear that

$$\hat{\tilde{F}}_\gamma(x) = \frac{1}{\pi}\lim_{\varepsilon\to 0}\lim_{N\to\infty}\Big(\hat{K}_{N,\varepsilon}(x) - \hat{K}_{N,\varepsilon}(-x)\Big)\hat{f}(x) = \frac{1}{\pi}\Big(\hat{K}(x) - \hat{K}(-x)\Big)\hat{f}(x).$$

From (4.4) it is clear that $\hat{K}(x) - \hat{K}(-x)$ is bounded uniformly in γ since the unbounded terms cancel. Letting $\gamma \to 0$ in (4.4) then gives $\lim_{\gamma\to 0}(\hat{K}(x) - \hat{K}(-x))$

$$= -2i\,\mathrm{sgn}\,x\int_0^1 \log t\cos t\,dt + 2i\,\mathrm{sgn}\,x\cos 1 + \frac{2}{i\,\mathrm{sgn}\,x}\int_1^\infty \frac{\cos t}{t^2}\,dt$$

$$= 2i \operatorname{sgn} x\left(\lim_{N\to\infty}\int_0^N \frac{\sin t}{t}dt\right) = \pi i \operatorname{sgn} x \ .$$

Therefore,

$$\lim_{\gamma\to 0} \hat{\hat{F}}_\gamma(x) = i\hat{f}(x) \operatorname{sgn} x \ .$$

The Fourier transform of the Hilbert transform of $f(x)$ may be written as

$$\frac{1}{\pi}\lim_{\varepsilon\to 0}\lim_{N\to 0}\left(\int_\varepsilon^N \frac{e^{ixt} - e^{-ixt}}{t}dt\right)\hat{f}(x) = i\hat{f}(x) \operatorname{sgn} x \ .$$

Thus, the two transforms are the same.

5. The N dimensional case. Most of the important results for the n dimensional case can be obtained from one dimensional results quite simply by the method of rotation which is treated in §8. Rotation methods, however, fail in certain cases, and for these a direct approach must be used. This will be similar to the one dimensional methods and is actually just a generalization of them.

In n dimensions the transforms will be of the form

$$\hat{f}(x) = \int_{E^n} \frac{f(x-t)\Omega(t)}{|t|^{n+i\gamma}}dt \text{ where } \Omega(t) = \Omega\left(\frac{t}{|t|}\right)$$

is a function only of angle and is integrable on the unit sphere, Σ. The part of the integral for which $0 \le |t| \le 1$ is obtained by using the same summation methods as before. The same reasoning shows that the existence of

$$\lim_{\varepsilon\to 0}\int_{|t|\ge\varepsilon} \frac{f(x-t)\Omega(t)}{|t|^{n+i\gamma}}dt - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}}\int_\Sigma \Omega(t)\,d\sigma \tag{5.1}$$

where $d\sigma$ is the element of "area" of the unit sphere, implies the existence of the original definition. The convergence in norm implies the convergence in norm of the original definition.

In n dimensions define

$$K_{N,\varepsilon}(t) = \begin{cases} \dfrac{\Omega(t)}{|t|^{n+i\gamma}} & \varepsilon \le |t| \le N \\ 0 & \text{elsewhere} \ . \end{cases}$$

LEMMA 4. *The expression* $\hat{K}_{N,\varepsilon}(x) - \dfrac{1}{i\gamma\varepsilon^{i\gamma}}\displaystyle\int_\Sigma \Omega(t)\,d\sigma$ *is in absolute vvlue less than* $c(\gamma) = C\dfrac{(|\gamma|+1)^2}{|\gamma|}\displaystyle\int_\Sigma |\Omega(t)|\,d\sigma$ *where* C *is an absolute* *constant. As $N \to \infty$ the expression converges to a function $\hat{K}_\varepsilon(x)$ except for $x = 0$. Furthermore, as $\varepsilon \to 0$, $\hat{K}_\varepsilon(x)$ converges to a function $\hat{K}(x)$ except for $x = 0$.*

Let θ be the angle between x and t. Then using polar coordinates

$$\begin{aligned}&\hat{K}_{N,\varepsilon}(x) - \frac{1}{i\gamma\varepsilon^{i\gamma}}\int_\Sigma \Omega(t)\,d\sigma \\ &= \int_\Sigma \Omega(t)\,d\sigma\Big(\int_\varepsilon^N \frac{e^{ir|x|\cos\theta}}{r^{1+i\gamma}}dr - \frac{1}{i\gamma\varepsilon^{i\gamma}}\Big)\,.\end{aligned} \tag{5.2}$$

The inner expression is the same at the one dimensional Fourier transform except that x has been replaced by $|x|\cos\theta$. Hence by Lemma 3 it is in absolute value less than $C\dfrac{(|\gamma|+1)^2}{|\gamma|}$. The convergence as $N \to \infty$, and $\varepsilon \to 0$ follow from this. Applying Holder's inequality then shows these conclusions hold for the whole expression.

COROLLARY 3. *If $f(x)$ belongs to L^2, the transform*

$$\hat{f}_\varepsilon(x) = \int_{|t|\geq\varepsilon} \frac{f(x-t)\Omega(t)}{|t|^{n+i\gamma}}\,dt - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}}\int_\Sigma \Omega(t)\,d\sigma$$

satisfies $||\hat{f}_\varepsilon(x)||_2 \leq c(\gamma)||f(x)||_2$. As $\varepsilon \to 0, \hat{f}_\varepsilon(x)$ converges in L^2 norm to a function $\hat{f}(x)$ which also satisfies $||\hat{f}(x)||_2 \leq c(\gamma)||f(x)||_2$.

The existence almost everywhere of $\hat{f}_\varepsilon(x)$ follows from the reasoning of [3] p. 292. The result then follows from Lemma 4 in the same way that Corollary 1 followed from Lemma 3.

COROLLARY 4. *If $\int_\Sigma \Omega(t)d\sigma = 0$ and $\Omega(t)$ belongs to $L\log^+ L$ on Σ, then the transform $\tilde{F}_\gamma(x) = \lim_{\varepsilon\to 0}\int_{|t|\geq\varepsilon} \dfrac{f(x-t)\Omega(t)}{|t|^{n+i\gamma}}\,dt$ for $|\gamma| \leq 1$ satisfies $||\tilde{K}_\gamma(x)||_2 \leq A||f(x)||_2$ where A is independent of γ and f. As $\gamma \to 0$, $F(x)$ converges in L^2 to the ordinary Calderon and Zygmund singular integral $\lim_{\varepsilon\to 0}\int_{|t|\geq\varepsilon} \dfrac{f(x-t)\Omega(t)}{|t|^n}\,dt$.*

Using the one dimensional formula (4.4) in (5.2) shows that

$$\hat{K}(x) = \int_\Sigma \Omega(t)\Big(\frac{-|x\cos\theta|^{i\gamma}}{i\gamma}\Big)d\sigma + \int_\Sigma \Omega(t)H(|x|\cos\theta,\gamma)\,d\sigma \tag{5.3}$$

where $H(|x|\cos\theta,\gamma)$ is uniformly bounded in both arguments. The first term may be written as

$$-\frac{|x|^{i\gamma}}{i\gamma}\int_\Sigma \Omega(t)\Big(\frac{|\cos\theta|^{i\gamma}-1}{\gamma}\Big)\,d\sigma$$

since $\int_\Sigma \Omega(t)d\sigma = 0$. Now

$$\left|\frac{|\cos\theta|^{i\gamma}-1}{\gamma}\right| = \left|\frac{\log|\cos\theta|\int_0^\gamma i|\cos\theta|^{iu}du}{\gamma}\right| \leq \log\frac{1}{|\cos\theta|},$$

and since $\Omega(t)$ belongs to $L\log^+ L$ on Σ, an application of Young's inequality[1] shows that the first part of (5.3) is also uniformly bounded. Convergence follows from the pointwise convergence of the expressions in the integral signs and the bounded convergence theorem. The first part converges to 0 and the second part as in Corollary 2 converges to $\int_\Sigma \pi i \operatorname{sgn}(\cos\theta)\Omega(t)\,d\sigma$. That the Fourier transform in the case of the ordinary singular integral converges to the same value follows by expressing the transform in polar coordinates and again applying the reasoning of Corollary 2.

6. Convergence in norm. Let $\beta^f(y) = \sup|S|$ where all sets S such that $\int_S f(x)\,dx \geq |S|y$ are considered. Further, given a function $\Omega(x)$ of the type considered in the last section, let $\omega(r)$ be its modulus of continuity; that is $\omega(r) = \sup|\Omega(x) - \Omega(y)|$ where x and y both lie on the unit sphere and $|x-y| \leq r$.

LEMMA 5. *Let $f(x)$ be non negative and belong to $L^p, 1 \leq p \leq 2$, in E^n. Let $\Omega(t) = \Omega\left(\frac{t}{|t|}\right)$ be such that its modulus of continuity satisfies $\int_0^1 \frac{\omega(r)}{r}dr < \infty$. Let E_y be the set where*

$$\tilde{f}_\varepsilon(x) = \int_{|t|\geq\varepsilon}\frac{\Omega(t)}{|t|^{n+i\gamma}}f(x-t)\,dt - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}}\int_\Sigma \Omega(t)\,d\sigma \tag{6.1}$$

exceeds y in absolute value. Then $|E_y| \leq \frac{c(\gamma)}{y^2}\int_{E^n}[f(x)]_y^2dx + c(\gamma)\beta^f(y)$, where $[f(x)]_y = \min(f(x), y)$ and $c(\gamma) = \frac{C(|\gamma|+1)^2}{|\gamma|}$ where C depends only on Ω.

Note. The primary use of this lemma will be for the one dimensional case where the continuity condition is automatically satisfied and the constant C is an absolute constant.

This lemma is the same as Lemma 2, Chapter I of [1] except that the transform

$$\int_{|t|\geq 1/\lambda}\frac{\Omega(t)}{|t|^n}f(x-t)\,dt$$

has been replaced by (6.1) and λ by $1/\varepsilon$. The proof is almost identical, and therefore will not be repeated. The few minor differences will be

[1] See [7] Vol. I, p. 116.

mentioned.

When $f(x)$ is split into the two parts $g(x)$ and $h(x)$, the proof for the one in L^2, $h(x)$, is a consequence of Corollary 3. The proof that

$$k(x) = \int_{|t|\geq\varepsilon} \frac{\Omega(t)}{|t|^{n+i\gamma}} g(x-t)\,dt$$

satisfies

$$\int_{\bar{D}'_y} |k(x)|\,dx \leq c\int_{D_y} |g(x)|\,dx$$

is the same except where the expression for the difference of the kernels is obtained. The principal difference there is that the expression

$$\frac{1}{|t|^{n+i\gamma}} - \frac{1}{|t_k|^{n+i\gamma}}$$

arises instead of

$$\left|\frac{1}{|t|^n} - \frac{1}{|t_k|^n}\right| .$$

However, using the fact that

$$\frac{1}{a^{n+i\gamma}} - \frac{1}{b^{n+i\gamma}} = -\int_a^b \frac{n+i\gamma}{u^{n+1+i\gamma}}\,du ,$$

the same inequality can be obtained. Now

$$\tilde{g}_\varepsilon(x) = k(x) - \frac{g(x)}{i\gamma\varepsilon^{i\gamma}}\int_\Sigma \Omega(t)\,d\sigma$$

so that

$$\int_{\bar{D}'_y} |\tilde{g}_\varepsilon(x)|\,dx = \int_{\bar{D}'_y} |k(x)|\,dx \leq c\int_{D_y} |g(x)|\,dx .$$

From this point the proofs are again identical. Following the details closely also shows that the constants are of the desired form.

From this result Theorems 1 through 7 of Chapter I of [1] follow immediately, either with the same proofs or with minor modifications. In some cases where only norms are concerned it is more convenient to carry through the proof for

$$\int_{|t|\geq\varepsilon} \frac{\Omega(t)}{|t|^{n+i\gamma}} f(x-t)\,dt$$

and then to add in the other term for which the theorems are obviously true. Lemma 5 is also obviously valid for just this term of the transform. Thus, for example, the following are true.

THEOREM 1. *Let $f(x)$ belong to $L^p, 1 < p < \infty$, in E^n. Then with the continuity condition on Ω of Lemma 5, the function $\hat{f}_\varepsilon(x)$ of* (6.1) *also belongs to L^p. Furthermore, $\|\hat{f}_\varepsilon(x)\|_p \le c(\gamma, p)\|f(x)\|_p$ where $c(\gamma, p) = C\dfrac{(|\gamma| + 1)^2 p^2}{|\gamma|(p-1)}$ and C depends only on Ω.*

The form of $c(\gamma, p)$ can be obtained by using the reasoning of the remark on page 99 of [1], following the constants through the proof, and using the fact that for

$$p > 1, \left(\frac{(|\gamma| + 1)^2}{|\gamma|}\right)^{\frac{1}{p}} \le \frac{(|\gamma| + 1)^2}{|\gamma|} .$$

THEOREM 2. *Let $f(x)$ be a function such that*

$$\int_{E^n} |f(x)|(1 + \log^+|f(x)|)\, dx < \infty .$$

Then with the continuity condition of Ω of Lemma 5 $\tilde{f}_\varepsilon(x)$ is integrable over any set S of finite measure and

$$\int_S |\tilde{f}_\varepsilon(x)|dx \le c(\gamma)\int_{E^n} |f(x)|dx + c(\gamma)\int_{E^n} f(x) \log^+ \left(|S|^{1+\frac{1}{n}} f(x)\right) dx + c(\gamma) S^{-\frac{1}{n}}$$

where $c(\gamma) = C\dfrac{(|\gamma| + 1)^2}{|\gamma|}$ and C depends only on Ω.

THEOREM 3. *Let f be integrable in E^n and Ω satisfy the continuity condition of Lemma 5. Then if S is a set of finite measure,*

$$\int_S |\tilde{f}_\varepsilon(x)|^{1-\alpha} dx \le \frac{c}{\alpha} |S|^\alpha \left(\int_{E^n} |f(x)|dx\right)^{1-\alpha}$$

where c is a constant independent of α, S, ε and f.

THEOREM 4. *Let $\mu(x)$ be a mass-distribution, that is a completely additive function of Borel set in E^n, and suppose that the total variation V of μ in E^n is finite. Let $\mu'(x)$ denote the derivative of $\mu(x)$ which exists almost everywhere. Then if Ω satisfies the continuity condition of Lemma 5 and if*

$$\tilde{f}_\varepsilon(x) = \int_{|t|>\varepsilon} \frac{\Omega(t)}{|t|^{n+i\gamma}}\, d\mu(x - t) - \frac{\mu'(x)}{i\gamma\varepsilon^{i\gamma}} \int_\Sigma \Omega(t)\, d\sigma ,$$

over every set S of finite measure

$$\int_S |\tilde{f}_\varepsilon(x)|^{1-\alpha} dx \leq \frac{c}{\alpha} |S|^\alpha V^{1-\alpha} .$$

THEOREM 5. *Let $f(x)$ belong to $L^p, 1 < p < \infty$, and let Ω satisfy the continuity condition of Lemma 5. Then $\tilde{f}_\varepsilon(x)$ converges in the mean of order p as $\varepsilon \to 0$ to a function $\tilde{f}(x)$.*

From this last theorem it follows by use of Lemma 2 that the original summation definition of $\tilde{f}(x)$ also converges in L^p norm if f is in L^p and $1 < p < \infty$.

7. Pointwise convergence.

THEOREM 6. *If $f(x)$ belongs to $L^p, 1 < p < \infty$, then $\tilde{f}_\varepsilon(x)$ converges almost everywhere to a function $\tilde{f}(x)$ as $\varepsilon \to 0$. Moreover, the function $\sup_\varepsilon |\tilde{f}_\varepsilon(x)|$ belongs to L^p and $\| \sup_\varepsilon |\tilde{f}_\varepsilon(x)| \|_p \leq c \| f(x) \|_p$, c being a constant which depends on p, γ, and Ω only.*

The proof is similar to that of Theorem 1, Chapter II of [1]. Define

$$K_\varepsilon(x) = \begin{cases} \dfrac{\Omega(x)}{|x|^{n+i\gamma}} & |x| \geq \varepsilon \\ 0 & |x| < \varepsilon . \end{cases}$$

Let $H(x)$ be non negative, zero outside the unit sphere, have continuous first derivatives, and have $\int_{E^n} H(x) dx = 1$. Denote by $\tilde{f}(x)$ the limit in norm of $\tilde{f}_\varepsilon(x)$ and define

$$\hat{f}_\varepsilon(x) = \frac{1}{\varepsilon^n} \int_{E^n} H\left(\frac{x-t}{\varepsilon}\right) \tilde{f}(t)\, dt .$$

By the lemmas in Chapter II of [1], $\hat{f}_\varepsilon(x)$ converges almost everywhere to $\tilde{f}(x)$ and $\| \sup_\varepsilon \hat{f}_\varepsilon(x) \|_p \leq c \| \tilde{f}(x) \|_p \leq c \| f(x) \|_p$. As in [1] every constant not depending on f will be denoted by c simply.

Using the fact that $\tilde{f}_\varepsilon(x)$ converges in norm to $\tilde{f}(x)$,

$$\hat{f}_\varepsilon(x) = \lim_{\lambda \to 0} \int_{E^n} \frac{1}{\varepsilon^n} H\left(\frac{x-t}{\varepsilon}\right) \left[\int_{E^n} f(t-v) K_\lambda(v)\, dv \right.$$
$$\left. - \frac{f(t)}{i\gamma\lambda^{i\gamma}} \int_\Sigma \Omega(w) dw \right] dt .$$

This may be considered as the difference of two integrals and written

$$\hat{f}_\varepsilon(x) = \lim_{\lambda \to 0} \left[\int_{E^n} \int_{E^n} \frac{1}{\varepsilon^n} H\left(\frac{x-t}{\varepsilon}\right) f(t-v) K_\lambda(v)\, dt\, dv \right.$$
$$\left. - \int_{E^n} \frac{1}{\varepsilon^n} H\left(\frac{x-t}{\varepsilon}\right) \frac{f(t)}{i\gamma\lambda^{i\gamma}}\, dt \int_\Sigma \Omega \right] .$$

Making the substitutions $t=x-u+v$ in the first integral and $t=x-u$ in the second gives

$$\hat{f}_\varepsilon(x) = \lim_{\lambda\to 0}\int_{E^n} f(x-u)\left[\int_{E^n}\frac{K_\lambda(v)}{\varepsilon^n}H\left(\frac{u-v}{\varepsilon}\right)dv - \frac{1}{\varepsilon^n}H\left(\frac{u}{\varepsilon}\right)\frac{\lambda^{-i\gamma}}{i\gamma}\int_\Sigma \Omega\right]du$$

$$= \lim_{\lambda\to 0}\int_{E^n} f(x-u)\left[\int_{|v|\geq\varepsilon}\frac{K_\lambda(v)}{\varepsilon^n}H\left(\frac{u-v}{\varepsilon}\right)dv\right.$$

$$\left.+\int_{|v|<\varepsilon}\frac{K_\lambda(v)}{\varepsilon^n}\left[H\left(\frac{u-v}{\varepsilon}\right)-H\left(\frac{u}{\varepsilon}\right)\right]dv - \frac{H\left(\frac{u}{\varepsilon}\right)\int_\Sigma \Omega}{i\gamma\varepsilon^{n+i\gamma}}\right]du\,.$$

Since $H(x)$ is differentiable, the limit may be taken inside the integral signs to give

$$\hat{f}_\varepsilon(x) = \int_{E^n} f(x-u)\left[\int_{|v|\geq\varepsilon}\frac{K(v)}{\varepsilon^n}H\left(\frac{u-v}{\varepsilon}\right)dv\right.$$

$$\left.+\int_{|v|<\varepsilon}\frac{K(v)}{\varepsilon^n}\left[H\left(\frac{u-v}{\varepsilon}\right)-H\left(\frac{u}{\varepsilon}\right)\right]dv - \frac{H\left(\frac{u}{\varepsilon}\right)\int_\Sigma \Omega}{i\gamma\varepsilon^{n+i\gamma}}\right]du$$

where $K(v) = K_0(v)$.

Now it is also true that

$$\tilde{f}_\varepsilon(x) = \int_{E^n} f(x-u)\left[\int_{E^n}\frac{K_\varepsilon(u)}{\varepsilon^n}H\left(\frac{u-v}{\varepsilon}\right)dv\right]du - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}}\int_\Sigma \Omega$$

since the integral $\int_{E^n} H(x)\,dx = 1$. For $|u|\geq 3\varepsilon$ it is clear that

$$\left|\int_{E^n}\frac{K_\varepsilon(v)}{\varepsilon^n}H\left(\frac{u-v}{\varepsilon}\right)dv - \int_{E^n}\frac{K_\varepsilon(u)}{\varepsilon^n}H\left(\frac{u-v}{\varepsilon}\right)dv\right|$$

$$= \left|\int_{E^n}\varepsilon^{-n}H\left(\frac{u-v}{\varepsilon}\right)(K_\varepsilon(v)-K_\varepsilon(u))\,dv\right|$$

$$\leq \int_{E^n}\varepsilon^{-n}H\left(\frac{u-v}{\varepsilon}\right)\frac{c\omega\left(\frac{c\varepsilon}{|u|}\right)}{|u|^n}\,dv = \frac{c\omega\left(\frac{c\varepsilon}{|u|}\right)}{|u|^n}\,.$$

As before ω is the modulus of continuity of Ω and c is independent of ε. The last inequality for $|K_\varepsilon(v)-K_\varepsilon(u)|$ is the one used in the proof of Lemma 5; it is valid here because $|u-v|<\varepsilon$ when the integrand is not zero.

For $|u|\leq 3\varepsilon$ it is clear that both

$$\left|\int_{E^n}\frac{K_\varepsilon(v)}{\varepsilon^n}H\left(\frac{u-v}{\varepsilon}\right)dv\right| \text{ and } \left|\int_{E^n}\frac{K_\varepsilon(u)}{\varepsilon^n}H\left(\frac{u-v}{\varepsilon}\right)dv\right|$$

are less than or equal to

$$\int_{|v|\leq 4\varepsilon} \frac{c}{\varepsilon^{2n}} H\left(\frac{u-v}{\varepsilon}\right) dv \leq \frac{c}{\varepsilon^n}\chi_{(0,5)}\left(\frac{|u|}{\varepsilon}\right).$$

Here $\chi_{(0,5)}$ is the characteristic function of the interval $(0, 5)$.

Similarly

$$\left|\int_{|v|<\varepsilon} \frac{K(v)}{\varepsilon^n}\left[H\left(\frac{u-v}{\varepsilon}\right) - H\left(\frac{u}{\varepsilon}\right)\right] dv\right| \leq c\left|\int_{|v|<\varepsilon} \frac{K(v)}{\varepsilon^n}\frac{|v|}{\varepsilon}\chi_{(0,2)}\left(\frac{|u|}{\varepsilon}\right) dv\right|$$

$$\leq \frac{c}{\varepsilon^n}\chi_{(0,2)}\left(\frac{|u|}{\varepsilon}\right).$$

Combining all of these results

$$|\hat{f}_\varepsilon(x) - \tilde{f}_\varepsilon(x)| \leq \int_{E^n} |f(x-u)| \left[\frac{c\omega\left(\frac{c\varepsilon}{|u|}\right)}{|u|^n}\chi_{(3,\infty)}\left(\frac{|u|}{\varepsilon}\right)\right.$$

$$\left. + \frac{c}{\varepsilon^n}\chi_{(0,5)}\left(\frac{|u|}{\varepsilon}\right) + \frac{c}{\varepsilon^n}\chi_{(0,2)}\left(\frac{|u|}{\varepsilon}\right)\right] du$$

$$+ \left|\frac{\varepsilon^{-i\gamma}}{i\gamma}\left[f(x) - \frac{1}{\varepsilon^n}\int_{E^n} f(x-u)H\left(\frac{u}{\varepsilon}\right) du\right]\int_\Sigma \Omega\right|.$$

From this the lemmas of the second chapter of [1] give

$$\|\sup_\varepsilon |\hat{f}_\varepsilon(x) - \tilde{f}_\varepsilon(x)|\|_p \leq c\|f(x)\|_p .$$

Then since $\lim_{\varepsilon\to 0}\hat{f}_\varepsilon(x) = \tilde{f}(x)$ almost everywhere and $\|\sup_\varepsilon|\hat{f}_\varepsilon(x)|\|_p \leq c\|f(x)\|_p$ and $\tilde{f}_\varepsilon \to \tilde{f}$ in mean of order p, the theorem follows.

THEOREM 7. *Let $\mu(x)$ be a mass distribution, that is, a completely additive function of Borel set in E^n and suppose that the total variation V of $\mu(x)$ in E^n is finite. Then the expression*

$$\tilde{f}_\varepsilon(x) = \int_{|t|>\varepsilon} \frac{\Omega(t)}{|t|^{n+i\gamma}} d\mu(x-t) - \frac{\mu'(x)}{i\gamma\varepsilon^{i\gamma}}\int_\Sigma \Omega ,$$

where $\mu'(x)$ is the derivative of $\mu(x)$ where this exists, has a limit $\tilde{f}$ almost everywhere as ε tends to zero, and over every set S of finite measure $\int_S |\tilde{f}(x)|^{1-\alpha} dx \leq \frac{c}{\alpha}|S|^\alpha V^{1-\alpha}$.

This corresponds to Theorem 2, Chapter II of [2]. The proof is the same except that Theorem 6 is used to obtain the convergence of the integral involving $g(x)$.

8. Other theorems. With this basis all the basic theorems in [2] and [3] can easily be shown to have their analogues for transforms of the type considered here. The periodic and discrete cases can be done simply; in the discrete case the subtracted term even disappears.

The rotation method as presented in [3] can also be applied in this case but the proof is much simpler. The method that applies only to odd kernels in the ordinary case applies to all kernels in this case. To illustrate this the following important theorem is given.

THEOREM 8. *Let $f(x)$ belong to $L^p, 1 < p < \infty$, in E^n. Let $\Omega(t) = \Omega\left(\frac{t}{|t|}\right)$ be merely integrable on the unit sphere Σ. Then if*

$$\tilde{f}_\varepsilon(x) = \int_{|t|>\varepsilon} \frac{\Omega(t)}{|t|^{n+i\gamma}} f(x-t)\,dt - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}} \int_\Sigma \Omega(t)\,d\sigma\,,$$

it satisfies

$$||\tilde{f}_\varepsilon(x)||_p \le \frac{C(|\gamma|+1)^2 p^2}{|\gamma|(1-p)}\,||b(x)||_p$$

where C depends only on Ω. As $\varepsilon \to 0$, $\tilde{f}_\varepsilon(x)$ converges in L^p norm to a function $\tilde{f}(x)$. Furthermore, $||\sup_\varepsilon |\tilde{f}_\varepsilon(x)|\,||_p \le c||f(x)||_p$ where c is independent of f, and $\tilde{f}_\varepsilon(x)$ converges almost everywhere to $\tilde{f}(x)$ as $\varepsilon \to 0$.

That $\tilde{f}_\varepsilon(x)$ exists almost everywhere is shown on page 292 of [3].

Let the norm symbol $||\ \ ||_p$ apply to the variable x. To write the integrals in polar coordinates let $t = rt'$, t' on the unit sphere. Then

$$\begin{aligned}
||\sup_\varepsilon |\tilde{f}_\varepsilon(x)|\,||_p &= \left|\left|\sup_\varepsilon \left|\int_{|t|\ge\varepsilon} \frac{\Omega(t)}{|t|^{n+i\gamma}} f(x-t)\,dt - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}}\int_\Sigma \Omega\right|\right|\right|_p \\
&= \left|\left|\sup_\varepsilon \left|\int_\Sigma \Omega(t')\,d\sigma\left(\int_\varepsilon^\infty \frac{f(x-rt')}{r^{1+i\gamma}}\,dr - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}}\right)\right|\right|\right|_p \\
&\le \left|\left|\int_\Sigma |\Omega(t')|\,d\sigma\left(\sup_\varepsilon \left|\int_\varepsilon^\infty \frac{f(x-rt')}{r^{1+i\gamma}}\,dr - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}}\right|\right)\right|\right|_p .
\end{aligned}$$

Using Minkowski's integral inequality this is less than or equal to

$$\int_\Sigma |\Omega(t')|\,d\sigma \left|\left|\sup_\varepsilon\left|\int_\varepsilon^\infty \frac{f(x-rt')}{r^{1+i\gamma}}\,dr - \frac{f(x)}{i\gamma\varepsilon^{i\gamma}}\right|\right|\right|_p .$$

Using the one dimensional version of theorem 6 on the inner integral by first integrating x parallel to t' and then over the space of such lines gives

$$||\sup_\varepsilon |\tilde{f}_\varepsilon(x)|\,||_p \le \left(\int_\Sigma |\Omega(t')|\,d\sigma\right) c||f(x)||_p \le c||f(x)||_p .$$

The inequality for $||\tilde{f}_\varepsilon(x)||_p$ follows using the same method and the one dimensional version of Theorem 1. The rest of the proof is the same as that of Theorem 3 of [3] once convergence for continuously differentiable f vanishing outside a bounded set is shown. Writing $\tilde{f}_\varepsilon(x)$ as

$$\tilde{f}_\varepsilon(x) = \int_{\varepsilon \le |t| \le 1} \frac{f(x-t) - f(x)}{|t|^{n+i\gamma}} \Omega(t)\, dt + \int_{|t| \ge 1} \frac{f(x-t)\Omega(t)}{|t|^{n+i\gamma}}\, dt - \frac{f(x)}{i\gamma}\int_\Sigma \Omega$$

shows clearly that it converges pointwise in this case.

9. Transforms of fractional integral type.[2]

DEFINITION. $T_z(f) = c_z \int_{E^n} s(t)\left(\frac{\theta(t)}{|t|^n}\right)^z f(x-t)\,dt$ for $0 \le R(z) < 1$,

$$T_z(f) = c_z \lim_{\varepsilon \to 0} \int_{|t| \ge \varepsilon} s(t)\left(\frac{\theta(t)}{|t|^n}\right)^z f(x-t)\,dt - \frac{f(x)\int_\Sigma s(t)(\theta(t))^z\, d\sigma}{(nz-n)\varepsilon^{nz-n}}$$

for $R(z) = 1$ and $z \ne 1$, and $T_1(f) = -\frac{1}{n} f(x)\int_\Sigma s(t)\theta(t)\,d\sigma$, where $c_z = \frac{z-1}{(z-2)^2}$, o^0 is taken as 0, $\theta(t) = \theta\left(\frac{t}{|t|}\right) \ge 0$ is integrable on the unit sphere Σ, $s(t) = s\left(\frac{t}{|t|}\right)$ has absolute value one, and $R(z)$ denotes the real part of z.

To obtain the principal theorem of this section a theorem of Stein [4] p. 483 will be used. For this purpose it will be necessary to show that the operators T_z as defined above satisfy the conditions of this theorem. Using the terminology of [4], the following lemma may be proved.

LEMMA 6. *Consider the set T_z as a family of operators from functions in E^n that are zero off the sphere $|x| \le D$ to functions in E^n. The set T_z is then an analytic family of operators of admissible growth in the strip $0 \le R(z) \le 1$. For a simple function φ in the given set, the inequalities $\|T_{1+iy}\varphi\|_p \le \frac{Cp^2}{p-1}\|\varphi\|_p$ for $1 < p < \infty$, and $\|T_{iy}\varphi\|_\infty \le \|\varphi\|_1$ hold where C depends only on $\theta(t)$ and not on D.*

Throughout the proof φ and ψ will be simple non negative functions and M the maximum of φ. Since any simple function can be written as the difference of two such functions, it will be sufficient to prove the assertions for these. The lemma will be proved in parts as indicated.

a. Simple functions in the given set are transformed into measurable functions for $0 \le R(x) \le 1$. For $R(z) = 1$ this follows from the preced-

[2] The method of this section was suggested by A. P. Calderon.

ing sections. To consider the case $0 \leq R(z) < 1$, let $r = |t|$ and $t' = t/|t|$. Then changing to polar coordinates

$$T_z(\varphi) = c_z \int_\Sigma s(t')(\theta(t'))^z d\sigma \int_0^\infty \frac{f(x - rt')}{r^{1-n+nz}} dr . \tag{9.1}$$

Using this,

$$|T_z(\varphi)| \leq M\left(\int_\Sigma (1 + \theta(t'))d\sigma\right)\left(\int_A^B \frac{dr}{r^{1-n+nR(z)}}\right)$$

where A is the greater of $|x| - D$ and 0 and $B = |x| + D$. Both integrals are obviously finite so that $T_z\varphi$ exists. The measurability follows from the Fubini theorem.

b. If $R(z) = 1$ and $1 \geq \varepsilon > 1/3n$, then $|T_{z-\varepsilon}(\varphi)|$ is bounded by a constant that is independent of z and ε. For $\varepsilon = 1$

$$|T_{z-1}(\varphi)| \leq \int_{E^n} |\varphi(x - t)| dt$$

and is obviously bounded. For $1 > \varepsilon > 1/3n$

$$\begin{aligned} |T_{z-\varepsilon}(\varphi)| &\leq |c_{z-\varepsilon}| \int_{E^n} \left(\frac{\theta(t)}{|t|^n}\right)^{1-\varepsilon} \varphi(x - t) dt \\ &\leq \left[\int \left(\frac{\theta(t)}{|t|^n}\right)^{\frac{3n-1}{3n}} \varphi(x - t) dt\right]^{\frac{3n-3n\varepsilon}{3n-1}} \left[\int \varphi(x - t) dt\right]^{\frac{3n\varepsilon-1}{3n-1}} \end{aligned}$$

by use of Holder's inequality. The second integral is certainly bounded. Writing the first integral in polar coordinates shows that it is in absolute value less than

$$\int_\Sigma (1 + \theta(t'))d\sigma \int_0^{2D} \frac{Mdr}{r^{2/3}}$$

so that it too is bounded. Since the exponents are between 0 and 1 the whole expression is bounded.

c. If $R(z) = 1$ and $1/3n \geq \varepsilon > |I(z)|$, where $I(z)$ denotes the imaginary part of z, then $|T_{z-\varepsilon}(\varphi)|$ is bounded by a constant that is independent of z and ε. Using polar coordinates,

$$\begin{aligned} |T_{z-\varepsilon}(\varphi)| &\leq |c_{z-\varepsilon}| \int_\Sigma [\theta(t')]^{1-\varepsilon} d\sigma \int_0^\infty \frac{f(x - rt')}{r^{1-n\varepsilon}} dr \\ &\leq 2\varepsilon \int_\Sigma (1 + \theta(t'))d\sigma \int_0^{2D} \frac{M}{r^{1-n\varepsilon}} dr \\ &\leq 2\varepsilon \int_\Sigma (1 + \theta(t'))d\sigma \frac{M}{n\varepsilon}(2D)^{n\varepsilon} \\ &\leq \frac{4DM}{n} \int_\Sigma (1 + \theta(t'))d\sigma . \end{aligned}$$

d. If $R(z) = 1$, $\varepsilon < |I(z)|$ and $\varepsilon < 1/2n$, then the integral

$$c_{z-\varepsilon}\int_{|t|\geq 1} s(t)\left[\left(\frac{\theta(t)}{|t|^n}\right)^{z-\varepsilon} - \frac{(\theta(t))^{z-\varepsilon}}{|t|^{nz}}\right]\varphi(x-t)\,dt \tag{9.2}$$

is uniformly bounded. For $z \neq 1$ it converges to 0 as ε approaches 0. The integral of (9.2) is clearly dominated by $\int_{|t|\geq 1} \frac{2(1+\theta(t))}{|t|^{n-\frac{1}{2}}}\varphi(x-t)\,dt$ which is finite. Since $c_{z-\varepsilon}$ is bounded, the expression (9.2) is bounded; convergence follows from the dominated convergence theorem.

e. If $R(z) = 1$, $\varepsilon < |I(z)|$ and $\varepsilon < 1/2n$, then the integral

$$c_{z-\varepsilon}\int_0^1 n\varepsilon\alpha^{n\varepsilon-1}d\alpha\int_{|t|\geq\alpha} \frac{s(t)(\theta(t))^{z-\varepsilon}\varphi(x-t)}{|t|^{nz}}\,dt\,. \tag{9.3}$$

has uniformly bounded L^2 norm. For $z \neq 1$ it converges in L^2 to $T_z(\varphi)$ as ε approaches 0.

As before, let the norm symbol $\|\ \|_2$ apply to the variable x. Then changing to polar coordinates the L^2 norm of (9.3) is

$$\left\|c_{z-\varepsilon}\int_\Sigma s(t')(\theta(t'))^{z-\varepsilon}d\sigma\int_0^1 n\varepsilon\alpha^{n\varepsilon-1}d\alpha\int_\alpha^\infty \frac{\varphi(x-rt')dr}{r^{nz-n+1}}\right\|_2 .$$

Then applying Minkowski's integral inequality twice shows that this is less than or equal to

$$|c_{z-\varepsilon}|\int_\Sigma (1+\theta(t'))\,d\sigma\int_0^1 n\varepsilon\alpha^{n\varepsilon-1}d\alpha\left\|\int_\alpha^\infty \frac{\varphi(x-rt')\,dr}{r^{nz-n+1}}\right\|_2 .$$

Using Corollary 1 and performing the integration of x first over lines parallel to t' and then over the space of such lines shows that the whole expression is bounded by

$$\frac{2|I(z)|}{1+|I(z)|^2}\int_\Sigma (1+\theta(t'))\,d\sigma\frac{C(1+|I(z)|^2)}{|I(z)|}\|\varphi\|_2 = 2C\|\varphi\|_2\int_\Sigma (1+\theta(t'))\,d\sigma\,.$$

To prove the convergence consider the expression

$$c_{z-\varepsilon}\int_\Sigma s(t')(\theta(t'))^z d\sigma\int_0^1 n\varepsilon\alpha^{n\varepsilon-1}d\sigma\int_\alpha^\infty \frac{\varphi(x-rt')}{r^{nz-n+1}}\,dr\,. \tag{9.4}$$

This converges in L^2 norm to $T_z(\varphi)$ by Corollary 3 and Lemma 2 since its limit is the Abel summation definition of $T_z(\varphi)$ written in polar coordinates. The reasoning used above to show that (9.3) had bounded L^2 norm can be applied to the difference of (9.3) and (9.4). This shows that the L^2 norm of the difference is less than or equal to

$$2C\int_\Sigma |(\theta(t'))^{z-\varepsilon} - (\theta(t))^z|\,d\sigma\,\|\varphi\|_2\,,$$

and this converges to 0 as ε approaches 0. Consequently, (9.3) converges to $T_z(\varphi)$.

f. $F(z) = \int \psi T_z(\varphi)\,dx$ is analytic in $0 \le R(z) < 1$. For $1 - R(z) \ge |I(z)|$ or $1 - R(z) \ge 1/3n$ this follows immediately from the majorizing expressions for $T_z(\varphi)$ in parts b and c. Since $T_z(\varphi)$ is a uniformly convergent integral of an analytic function in these cases, $T_z(\varphi)$ and hence $F(z)$ are analytic. For $1 - R(z) < |I(z)|$ and $1 - R(z) < 1/3n$ observe that $T_z(\varphi)$ is the sum of (9.2) and (9.3). By the same reasoning as in the other case, the integral of the product of ψ with either (9.2) or (9.3) is analytic. Therefore, the sum of these parts, $F(z)$, is analytic.

g. $F(z) = \int \psi T_z(\varphi)\,dx$ is continuous on $R(z) = 1$. By its definition $T_z(\varphi)$ is the product of c_z and the transformation of the previous sections where $s(t)(\theta(t))^z$ has replaced $\Omega(t)$ and $(nz - n)/i$ has replaced γ. Using Fourier transforms then gives $\hat{T}_z(\varphi) = c_z\hat{K}_z\hat{\varphi}$ where $\hat{K}_z$ is the function $\hat{K}$ of Lemma 4 with $\gamma = (nz-n)/i$, provided that $z \ne 1$. Using the expression (4.4) in (5.2) gives an expression for $c_z\hat{K}_z$. Its form shows that $c_z\hat{K}_z$ is uniformly bounded in x and z. Furthermore, for $0 < a \le x \le b < \infty$, it is also clear that $c_z\hat{K}_z$ is continuous in z, uniformly in x. Both statements remain valid if $-\frac{1}{n}\int_\Sigma s(t)\theta(t)\,d\sigma$ is used for $c_1\hat{K}_1$. Using this, it is also clear that $\hat{T}_1(\varphi) = c_1\hat{K}_1\hat{\varphi}$.

Now let z be a complex number with $R(z) = 1$, and let $\varepsilon > 0$ be arbitrary. Choose real numbers a and b so that if S consists of points in E^n whose distance from the origin lies between a and b, and S' is the complement of S in E^n, then

$$\left(\int_{S'} (\varphi(x))^2\,dx\right)^{\frac{1}{2}} \le \frac{\varepsilon}{4\|\psi\|_2 \sup\limits_{x,z} |c_z\hat{K}_z|} .$$

Let w be another complex number with $R(w) = 1$. Then

$$\begin{aligned} |F(z) - F(w)| &\le \|\psi\|_2\|T_z\varphi - T_w\varphi\|_2 \le \|\psi\|_2\|\hat{T}_z\hat{\varphi} - \hat{T}_w\hat{\varphi}\|_2 \\ &\le \|\psi\|_2\left(\int_{S'} \hat{\varphi}^2\,dx\right)^{\frac{1}{2}} \sup_{x\in S'} |c_z\hat{K}_z - c_w\hat{K}_w| \\ &\quad + \|\psi\|_2\left(\int_{S} \hat{\varphi}^2\,dx\right)^{\frac{1}{2}} \sup_{x\in S} |c_z\hat{K}_z - c_w\hat{K}_w| . \end{aligned}$$

The first part is less than $\varepsilon/2$ and the second part approaches 0 as w approaches z. This shows the desired continuity.

h. $F(z)$ is continuous and bounded on $0 \le R(z) \le 1$. From parts b through e it is clear that $F(z)$ is uniformly bounded in $0 \le R(z) < 1$ and $\lim\limits_{\varepsilon\to 0} F(z - \varepsilon) = F(z)$ for $R(z) = 1$ and $z \ne 1$. These facts, together with the analyticity and continuity on $R(z) = 1$, give the desired continuity and boundedness.

i. $\|T_{iy}\varphi\|_\infty \leq \|\varphi\|_1$ and $\|T_{1+iy}\varphi\|_p \leq A\dfrac{p^2}{p-1}\|\varphi\|_p, 1 < p < \infty$,

where y is real and A depends only on $\theta(t)$ and not on D or φ. The first is trivial. The second follows from Theorem 8 since $(1 + |y|^2)|c_{1+iy}|/|y|$ is bounded.

This completes the proof of the lemma.

THEOREM 9. *Let p, q, and λ be positive numbers such that $1 < p < q < \infty$ and $\dfrac{1}{q} = \dfrac{1}{p} - \lambda$. Let f be in L^p on E^n and $\Omega(t) = \Omega\left(\dfrac{t}{|t|}\right)$ be in L^s, $s = \dfrac{1}{1-\lambda}$, on the unit sphere. Then the integral*

$$D_\lambda(f) = \int_{E^n} \frac{\Omega(t)}{|t|^{n(1-\lambda)}} f(x-t)dt$$

exists for almost all x and

$$\|D_\lambda(f)\|_q \leq C\frac{1}{\lambda}\left(\frac{pq}{p-1}\right)^{1-\lambda}\|f\|_p$$

where C depends only on Ω.

Applying the theorem of Stein [4] p. 483 to the T_z with $p_1 = 1$, $q_1 = \infty$, $p_2 = q_2 = q(1-\lambda)$, $z = 1-\lambda$ gives for simple φ,

$$\|T_z\varphi\|_q \leq A\left(\frac{p_2^2}{p_2-1}\right)^{R(z)}\|\varphi\|_p \,.$$

Now let $\theta(t) = |\Omega(t)|^{\frac{1}{1-\lambda}}$, and $s(t) = \operatorname{sgn}\Omega(t)$. Then dividing the above inequality by $c_{1-\lambda}$ gives

$$\begin{aligned}\|D_\lambda(\varphi)\|_q &\leq A\left(\frac{q^2(1-\lambda)^2}{q(1-\lambda)-1}\right)^{1-\lambda}\frac{1+\lambda^2}{\lambda}\|\varphi\|_p \\ &= A\left(\frac{pq(1-\lambda)^2}{p-1}\right)^{1-\lambda}\frac{1+\lambda^2}{\lambda}\|\varphi\|_p \\ &\leqq \frac{2A}{\lambda}\left(\frac{pq}{p-1}\right)^{1-\lambda}\|\varphi\|_p\end{aligned}$$

Now if $\Omega \geq 0$ all the integrands are positive. Given an arbitrary positive function f in L^p, take a sequence of simple functions φ_n that vanish off bounded sets and converge in L^p norm to f. Then taking the limit in the inequality above gives

$$|D_\lambda(f)\|_q \leq \frac{2A}{\lambda}\left(\frac{pq}{p-1}\right)^{1-\lambda}\|f\|_p \,.$$

From this $D_\lambda(f)$ exists almost everywhere. In the case where f and Ω are not positive the integrand of $D_\lambda(f)$ is majorized by a positive function that does satisfy the desired inequality. This completes the proof.

It is known that the usual fractional integration theorem and, as a result, Theorem 9 fail for the cases $p = 1$ and $q = \infty$. Zygmund [8] p. 605-6 proved substitute results for the usual fractional integral case, and these results can be extended to the present case. The proof of Theorem 10 is an adaptation of the corresponding proof in [8].

THEOREM 10. *Let $p = 1/\lambda$ be a positive number greater than 1. Let f be in L^p on E^n, vanish off a bounded set R and $\|f\|_p \le 1$. Let $\Omega(t) = \Omega\left(\frac{t}{|t|}\right)$ be in L^s, $s = \frac{1}{1-\lambda}$ on the unit sphere. Then the expression $D_\lambda(f)$ exists for almost all x. Furthermore, if $\Phi(x) = e^{x^s} - x^s - 1$, there exist constants a and A, independent of f and R, such that*

$$\int_{E^n} \Phi(a|D_\lambda(f)|) \le A|R| \, .$$

Using Theorem 9

$$\int_{E^n} \Phi(a|D_\lambda(f)|) \le \sum_2^\infty \frac{a^{ns}}{n!}\int_R |D_\lambda(f)|^{ns}$$
$$\le \sum_2^\infty \frac{1}{n!}\left(\frac{n^2}{(n-1)(1-\lambda)^2}\right)^n \left(\frac{aC}{\lambda}\right)^{ns} \|f\|_{p_n}^{ns}$$

where $p_n = \frac{n}{1-\lambda+\lambda n}$. Now using the fact that $\left(\frac{1}{|R|}\int_R |f|^p\right)^{\frac{1}{p}}$ increases with p shows that the preceding sum is less than or equal to $\sum_2^\infty \frac{(a^s Dn)^n}{n!}|R|\|f\|_p^{ns}$ where D is a constant independent of $n, f,$ and R. Then using the fact that $\|f\|_p \le 1$ and Stirling's formula shows that for $a^s = 1/(2eD)$ the series converges to a constant A.

THEOREM 11. *Let $q = 1/(1-\lambda)$ be a positive number, $1 < q < \infty$. Let $\Psi(x) = (1 + x)[\log(1 + x)]^{1-\lambda}$ and f be a function in E^n such that $\int_{E^n} \Psi(|f|)$ is finite. Let $\Omega(t) = \Omega\left(\frac{t}{|t|}\right)$ be in L^s, $s = 1/(1-\lambda)$ on the unit sphere. Then the expression $D_\lambda(f)$ exists for almost all x, and over any set R of finite measure*

$$\left(\int_R |D_\lambda(f)|^q\right)^{\frac{1}{q}} \le A\left(|R| + \int_R \Psi(|f|)\right)$$

where A is independent of f and R.

By differentiating it is clear that $\Psi(x)$ is greater than the function conjugate to $\Phi(x) = e^{x^s} - x^s - 1$ in the sense of Young.[3] Consequently, for real positive numbers b and d, $bd \leq \Phi(b) + \Psi(d)$ by Young's inequality. Now consider a function g in L^p, $p = 1/\lambda$, vanishing outside R and with $||g||_p \leq 1$. Then using Theorem 10

$$\left|\int_{E^n} D_\lambda(g)f\right| \leq \frac{1}{a}\int_{E^n} a|D_\lambda(g)||f| \leq \frac{1}{a}\left(\int_{E^n} \Phi(a|D_\lambda(g)|)\right) + \int_{E^n} \Psi(|f|)$$
$$\leq \frac{1}{a}\left(A|R| + \int_{E^n} \Psi(|f|)\right).$$

However, by interchanging the order of integration

$$\left|\int_{E^n} D_\lambda(g)f\right| = \left|\int_R gD_\lambda(f)\right|.$$

Since g is an arbitrary function in L^p on R, the least upper bound for this integral is $\left(\int_R |D_\lambda(f)|^q\right)^{1/q}$ by the converse of Holder's inequality. Therefore $\left(\int_R |D_\lambda(f)|^q\right)^{1/q} \leq \dfrac{A|R|}{a} + \dfrac{1}{a}\int_{E^n} \Psi(|f|)$.

References

1. A. P. Calderon and A. Zygmund, *On the existence of certain singular integrals*, Acta Mathematica, **88** (1952), 85–139.
2. ———, *Singular integrals and periodic functions*, Studia Mathematica, **14** (1954), 249-271.
3. ———, *On singular integrals*, Amer. J. Math. **88** (1956), 289–309.
4. E. M. Stein, *Interpolation of linear operators*, Trans. Amer. Math. Soc. **83** (1956), 482–492.
5. E. M. Stein and G. Weiss, *Fractional integrals on n dimensional space*, J. Math. and Mech. **7** (1958), 503–514.
6. G. O. Thorin, *Convexity theorems*, Meddelanden fran Lunds Universitets Matematiska Seminarium, **9** (1948).
7. A. Zygmund, *Trigonometrical Series*, Second Ed., Vol. I, p. 1–383, Vol. II, p. 1–343, Cambidge University Press, 1959.
8. ———, *Some points in the theory of trigonometric and power series*, Trans. Amer. Math. Soc. **36** (1934), 586–617.

DE PAUL UNIVERSITY,
CHICAGO, ILLINOIS

[3] See [7] Vol. I, p. 16

ON THE STABILITY OF BOUNDARY COMPONENTS

KÔTARO OIKAWA

I. PRESENTATION OF THE PROBLEM

1. Definitions.

1. A *boundary component* of a plane region $D \subset (|z| \leqq \infty)$ is a component of the boundary ∂D of D, i.e., a connected subset of ∂D which is not a proper subset of any connected subset of ∂D.

There is an alternate definition. Let $\{\Omega_n\}_{n=1}^{\infty}$ be a sequence of subregions of D such that

(i) $\Omega_1 \supset \Omega_2 \supset \cdots$,

(ii) the relative boundary $\partial\Omega_n \cap D$ consists of one closed analytic curve in D,

(iii) $\bigcap_{n=1}^{\infty} \Omega_n = \phi$. Two sequences $\{\Omega_n\}$ and $\{\Omega'_n\}$ are said to be equivalent if, for any n, there exists m such that $\Omega_m \subset \Omega'_n$ and $\Omega'_m \subset \Omega_n$. A *boundary component* of D is an equivalence class of $\{\Omega_n\}$.

These two definitions are equivalent in the following sense:

(i) Given a sequence $\{\Omega_n\}$, the set $\bigcap_{n=1}^{\infty} \bar{\Omega}_n$ is a component of ∂D and, for two sequences, these sets coincide if and only if the sequences are equivalent.

(ii) Given a component Γ of ∂D, there exists a sequence such that $\Gamma = \bigcap_{n=1}^{\infty} \bar{\Omega}_n$.

For a boundary component Γ, the sequence $\{\Omega_n\}$ such that $\Gamma = \bigcap_{n=1}^{\infty} \bar{\Omega}_n$ is called a *defining sequence of* Γ.

Let $w = f(z)$ be a topological mapping of D onto a plane region D'. Then we can immediately see from the second definition that f gives a one-to-one correspondence between the boundary components of D and D'. We shall speak of the *image of a boundary component* Γ *under* f in this sense and denote it by $f(\Gamma)$.

2. Let D^c denote the complement of D with respect to the extended plane $|z| \leqq \infty$. For a boundary component Γ, there exists a uniquely determined component of D^c whose boundary coincides with Γ. We call it the *component of* D^c *corresponding to* Γ and denote it by Γ^*.

If D does not contain the point $z = \infty$, the boundary component Γ

Received January 7, 1959. The present paper is a part of the author's doctoral dissertation submitted to the University of California, Los Angeles. The author wishes to express his heartiest gratitude to Professor Leo Sario for his guidance and encouragement during the preparation of this paper.

such that $\infty \in \Gamma^*$ is called the *outer boundary of D.*

3. We call a region D a *circular* (or *radial*) *slit disk* if $0 \in D$, $D \subset (|z| < R < \infty)$, the outer boundary is $|z| = R$, and every other boundary component is either a point or an arc on $|z| = \text{const.}$ (or a line segment on $\arg z = \text{const.}$).

2. The stability problem of boundary components.

4. Let D be a plane region and let Γ be a boundary component. Sario [16, 17] gave the following classification:

(a) If $f(\Gamma)$ is a point for every univalent function $w = f(z)$ on D, then Γ is said to be *weak*.

(b) If $f(\Gamma)$ is a continuum, i.e., a connected closed set containing more than one point, for every f, then Γ is said to be *strong*.

(c) If Γ is neither weak nor strong, it is said to be *unstable*.

Weak boundary components were first investigated by Grötzsch in connection with the so-called "Kreisnormierungsproblem" (Grötzsch [7]; see also Denneberg [5] and Strebel [21]). He called them vollkommen punktförmig. Regions of class $O_{SB} = O_{SD}$ introduced by Ahlfors and Beurling [2] coincide with those possessing merely weak boundary components. Sario [16] has generalized the concept weak boundary components for open Riemann surfaces. It has been discussed also by Savage [19] and Jurchescu [10].

We are now lead to the following natural problems:

PROBLEM A. *Given a boundary component consisting of a single point, determine whether it is weak or unstable.*

PROBLEM B. *Given a boundary component consisting of a continuum, determine whether it is strong or unstable.*

We shall attempt to obtain concrete tests with practical applicability.

3. Related extremal problems.

5. Let D be a region containing the point $z = 0$. Let $\mathfrak{B}$ be the family consisting of all functions $w = \varphi(z)$ which are regular and univalent in $D - \{0\}$, and have the expansion $1/z + cz + \cdots$ near $z = 0$.

Consider, with Grötzsch [6], the diameter of the image $\varphi(\Gamma)$ of the boundary component Γ. It is quite easy to see that *Γ is weak if and only if* $\sup_{\varphi \in \mathfrak{B}} \operatorname{diam} \varphi(\Gamma) = 0$, *and Γ is* strong *if* $\inf_{\varphi \in \mathfrak{B}} \operatorname{diam} \varphi(\Gamma) > 0$.

6. Let $\mathfrak{F}_\Gamma$ be the family consisting of functions $w = f(z)$ such that

(i) regular and univalent in D,

(ii) $f(0) = 0$ and $f'(0) = 1$,

(iii) $f(\Gamma)$ is the outer boundary of $f(D)$.

Rengel [14] introduced the following functionals on $\mathfrak{F}_\Gamma$:

$$M(f) = \max_{w \in f(\Gamma)} |w| = \sup_{z \in D} |f(z)| ,$$

$$m(f) = \min_{w \in f(\Gamma)} |w| ,$$

and considered the quantities

$$R(\Gamma) = R(\Gamma;\ D) = \sup_{f \in \mathfrak{F}_\Gamma} m(f)$$

and

$$r(\Gamma) = r(\Gamma;\ D) = \inf_{f \in \mathfrak{F}_\Gamma} M(f) .$$

From the definition we have immediately the basic

THEOREM 1. *Γ is strong if $R(\Gamma) < \infty$. Γ is weak if and only if $r(\Gamma) = \infty$.*

These criteria are equivalent to those in No. 5, since

$$R(\Gamma) = 2/\inf_{\varphi \in \mathfrak{B}} \operatorname{diam} \varphi(\Gamma) ,$$

$$r(\Gamma) = 4/\sup_{\varphi \in \mathfrak{B}} \operatorname{diam} \varphi(\Gamma) .$$

In fact, for an arbitrary function $f(z) \ni \mathfrak{F}_\Gamma$, the functions

$$\varphi_f(z) = \frac{1}{f(z)} + \frac{f''(0)}{2}$$

and

$$\psi_f(z) = \varphi_f(z) + \frac{1}{M(f)^2} \cdot \frac{1}{\varphi_f(z)}$$

belong to $\mathfrak{B}$, and

$$m(f) \leqq 2/\operatorname{diam} \varphi_f(\Gamma)$$

$$M(f) \geqq 4/\operatorname{diam} \varphi_f(\Gamma) .$$

On the other hand, for $\varphi(z) \in \mathfrak{B}$, let $F(w)$ be the function which maps $(\varphi(\Gamma)^*)^c$ conformally onto the exterior of a disk with the center at the origin. Assume further that $F(w) = w + c + c'/w + \cdots$ near $w = \infty$. Then $f_\varphi(z) = 1/F \circ \varphi(z) \in \mathfrak{F}_\Gamma$ and

$$2/\operatorname{diam} \varphi(\Gamma) \leqq M(f_\varphi) = m(f_\varphi) \leqq 4/\operatorname{diam} \varphi(\Gamma) .$$

The proof of the above equalities is hereby complete.

7. Whether or not $R(\Gamma)<\infty$ is necessary for strength is still an open problem. We shall discuss this problem in No. 24.

We shall see in No. 17 that $1/r(\Gamma)$ equals the "capacity" of the boundary component Γ introduced by Sario [16] (it is not necessarily equal to the logarithmic capacity of the closed set Γ), and, therefore, that the latter half of Theorem 1 is equivalent to Sario's result ([17], Theorem 6). Jurchescu [10] showed that the "capacity" coincides with the "perimeter" introduced by Ahlfors and Beurling [2].

It will be shown in No. 22 that $R(\Gamma)$ coincides with the quantity which Strebel [22] called "extremal Durchmesser". Finally, Theorem 4 in No. 21 shows that the first half of the above theorem coincides with Sario's result ([17], Theorem 4).

II. Preliminaries

In this chapter, we collect a number of known results which will be needed later.

4. Extremal length.

8. A curve γ considered here is either a closed rectifiable curve or a curve of the form $z=z(t)$ $(0<t<1)$ every subarc of which is rectifiable. If $\lim_{t\to 0} z(t)$ or $\lim_{t\to 1} z(t)$ exists, it is called an end point.

Let D be a reginon and let $\{\gamma\}$ be a family of curves $\gamma\subset D$. Let $\{\rho\}$ be the collection of functions ρ which are $\geqq 0$ and lower semi-continuous in D. With the understanding that $0/0=\infty/\infty=0$, take

$$\lambda\{\gamma\}=\sup_{\rho}\frac{\left(\inf\limits_{\gamma}\int_{\gamma}\rho\, ds\right)^2}{\iint_D \rho^2\, dxdy}.$$

It is called the *extremal length of* $\{\gamma\}$ (Ahlfors and Beurling [2], Ahlfors and Sario [3]).

9. The following properties (I)–(V) are well known; for the proofs the reader is referred to, e.g., Hersch [8][1]:

(I) $\lambda\{\gamma\}$ is independent of the choice of D.

(II) $\lambda\{\gamma\}$ is conformally invariant.

(III) $\lambda\{\gamma'\}\leqq\lambda\{\gamma\}$ if every γ contains a γ'.

(IV) For $\{\gamma_1\}$ and $\{\gamma_2\}$, assume the existence of disjoint regions D_1 and D_2 such that $\gamma_\nu\subset D_\nu$ $(\nu=1, 2)$. If, for any γ of the third family

[1] His definition is different from ours, but his proofs remain valid.

$\{\gamma\}$, there exist γ_1 and γ_2 such that $\gamma_1 \cup \gamma_2 \subset \gamma$, then

$$\lambda\{\gamma_1\} + \lambda\{\gamma_2\} \leqq \lambda\{\gamma\} .$$

(V) Let $\{\gamma_1\}$ and $\{\gamma_2\}$ be the same as above. If $\{\gamma_1\} \cup \{\gamma_2\} \subset \{\gamma\}$, then

$$\frac{1}{\lambda\{\gamma_1\}} + \frac{1}{\lambda\{\gamma_2\}} \leqq \frac{1}{\lambda\{\gamma\}} .$$

(VI) (Hersch [8][1]). For three families with $\{\gamma\} = \{\gamma_1\} \cup \{\gamma_2\}$,

$$\frac{1}{\lambda\{\gamma\}} \leqq \frac{1}{\lambda\{\gamma_1\}} + \frac{1}{\lambda\{\gamma_2\}} .$$

(VIII) Let $\{\gamma_1\}$ be the subfamily of $\{\gamma\}$ consisting of γ having both end points and such that $z(t)$ $(0 \leqq t \leqq 1)$ is rectifiable. Then $\lambda\{\gamma\} = \lambda\{\gamma_1\}$.

In fact, since the extremal length of $\{\gamma_2\} = \{\gamma\} - \{\gamma_1\}$ is infinite, (VI) shows that $\lambda\{\gamma_1\} \leqq \lambda\{\gamma\}$, and $\lambda\{\gamma\} \leqq \lambda\{\gamma_1\}$ by (III).

(VIII) For a curve $\gamma : z = z(t)$ $(0 < t < 1)$, let $\bar{\gamma}$ be the curve $z = \overline{z(t)}$ $(0 < t < 1)$. If $z(0) = \lim_{t\to 0} z(t)$ exists and is real, put $\hat{\gamma} = \gamma \cup \bar{\gamma} \cup \{z(0)\}$. Let $\{\gamma_0\}$ be a family of curves which are contained in the upper half-plane and have the end points $z(0)$ on the real axis. Let $\{\gamma\}$ be a family which contains all $\hat{\gamma}_0$ and $\bar{\gamma}$. Furthermore it is assumed that, for any γ, there exist γ_0 and γ_0' in $\{\gamma_0\}$ such that $\bar{\gamma}_0 \cup \gamma_0' \subset \gamma$. Then

$$\lambda\{\gamma\} = 2\lambda\{\gamma_0\} .$$

In fact, to define $\lambda\{\gamma\}$, we may restrict $\{\rho\}$ to the subfamily consisting of functions symmetric about the real axis. Since $2\inf_{\gamma_0}\int_{\gamma_0}\rho\, ds = \inf_{\gamma}\int_{\gamma}\rho\, ds$ for such ρ, we conclude that $\lambda\{\gamma\} = 2\lambda\{\gamma_0\}$.

(IX) Let A be the annulus $1 < |z| < q$ or a region obtained by deleting a finite number of circular slits from this annulus. Let $\{\gamma\}$ be the family of all closed rectifiable curves in A separating $|z| = 1$ from $|z| = q$. Then $\lambda\{\gamma\} = 2\pi/\log q$. This is true even if each γ is restricted to a concentric circle in A.

The proof is found, e.g., in Hersch [8][1].

10. Let D be a region, and let E_0 and E_1 be compact sets such that $E_\nu \cap \bar{D} \neq \phi$ $(\nu = 0, 1)$. Let $\{\gamma\}$ be the family consisting of γ: $z = z(t)$ $(0 < t < 1)$ such that $\gamma \subset D$, $\bigcap_{\varepsilon>0}\overline{\{z(t);\ 0 < t < \varepsilon\}} \subset E_0$, and $\bigcap_{\varepsilon>0}\overline{\{z(t);\ 1 - \varepsilon < t < 1\}} \subset E_1$. Then $\lambda\{\gamma\}$ is called the *extremal distance* $\delta_D(E_0, E_1)$ *between* E_0 *and* E_1 *with respect to* D.

By (VII), $\delta_D(E_0, E_1)$ coincides with the extremal length of the family of rectifiable curves in D whose end points are on E_0 and E_1 respectively. Under a certain restriction of the configuration, it is also equal to that of a subfamily consisting of analytic curves (Wolontis [25]).

From this consideration, we get

(X) If no point of E_1 is accessible from D by a rectifiable curve, then $\delta_D(E_0, E_1) = \infty$.

(XI) (Pfluger [12][1]). If cap $E_1 = 0$, then $\delta_D(E_0, E_1) = \infty$. For $D = (|z| = 1)$, $E_0 = (|z| = \varepsilon < 1)$, and $E_1 \subset (|z| = 1)$, $\delta_D(E_0, E_1) = \infty$ if and only if cap $E_1 = 0$.

Combining (VI), (X), and (XI), we get

(X′) If no point on E_1, except for a set of capacity zero, is accessible from D by a rectifiable curve, then $\delta_D(E_0, E_1) = \infty$.

(XII) Let D, E_0, and E_1 be contained in the closed upper half-plane. Let $\hat{D}$ be the region which is the union of D, the reflection of D across the real axis, and the part of ∂D on the real axis. Let $\hat{E}_0$ and $\hat{E}_1$ have analogous meanings. If $\delta_{\hat{D}}(\hat{E}_0, \hat{E}_1)$ is expressed in terms of the extremal length of a family consisting of analytic curves[2], then

$$\delta_{\hat{D}}(\hat{E}_0, \hat{E}_1) = \frac{1}{2}\delta_D(E_0, E_1) .$$

Proof. Let $\delta_{\hat{D}}(\hat{E}_0, \hat{E}_1) = \lambda\{\gamma\}$ where γ is an analytic curve and let $\delta_D(E_0, E_1) = \lambda\{\gamma'\}$. Using the notation in (VII), we see immediately that $\{\gamma'\}$ and $\{\bar{\gamma}'\}$ are contained in $\{\gamma\}$. Since $\lambda\{\gamma'\} = \lambda\{\bar{\gamma}'\}$, we find, on applying (V), that $\lambda\{\gamma\} \leqq \lambda\{\gamma'\}/2$.

In order to prove the inequality in the opposite direction, we first remark that, to define $\lambda\{\gamma\}$, we may restrict ρ to a function symmetric about the real axis. For a curve $\gamma : z = z(t)$ $(0 < t < 1)$, let γ^* be

$$z = \begin{cases} z(t) & \text{if } \Im z(t) \geqq 0 \\ \overline{z(t)} & \text{if } \Im z(t) \leqq 0 . \end{cases}$$

Evidently $\int_\gamma \rho\, ds = \int_{\gamma^*} \rho\, ds$ for a symmetric ρ.

Since it is assumed that γ is an analytic curve, γ^* intersects the real axis at only a finite number of points $z_1, z_2, \cdots, z_k$. Let Δ_ν be the punctured disk $0 < |z - z_\nu| < r$ $(\nu = 1, 2, \cdots, k)$, where r is taken so small that the Δ_ν are mutually disjoint. The extremal length of the family of curves in Δ_ν separating z_ν from $|z - z_\nu| = r$ is, by (IX), equal to infinite. Therefore, for arbitrary $\varepsilon > 0$ and ρ, there exists a closed curve $\gamma_\nu \subset \Delta_\nu$ encircling z_ν and such that $\int_{\gamma_\nu} \rho\, ds < \varepsilon/k$. On replacing a part of $\gamma^* \cap \Delta_\nu$ by a part of $\gamma_\nu (\nu = 1, 2, \cdots, k)$, we obtain from γ^* a

[2] This restriction is satisfied in our subsequent applications. It is perhaps superfluous. However, the author has not succeeded in furnishing the proof without it.

curve γ' belonging to the family $\{\gamma'\}$ and such that $\int_{\gamma'}\rho\, ds - \varepsilon < \int_{\gamma}\rho\, ds$. Since γ and ε are arbitrary, we get $\inf_{\gamma'}\int_{\gamma'}\rho\, ds \leqq \inf_{\gamma}\int_{\gamma}\rho\, ds$ for every symmetric ρ. Since $\iint_{\hat{D}}\rho^2\, dxdy = 2\iint_{D}\rho^2\, dxdy$, we conclude that $\lambda\{\gamma'\} \leqq 2\lambda\{\gamma\}$.

(XIII) Let A be the annulus $1 < |z| < q$ or a region obtained by deleting a finite number of radial slits from this annulus. Let $E_0 = (|z| = 1)$ and $E_1 = (|z| = q)$. Then $\delta_A(E_0, E_1) = (\log q)/2\pi$, and it is also equal to the extremal length of the family of all radials from E_0 to E_1 in A.

For the proof, the reader is referred to, e.g., Strebel [20].

5. Teichmüller's extremal region.

11. Let D be a doubly connected region and let $\{\gamma\}$ be the family of all closed rectifiable curves in D separating the boundary components. The quantity $2\pi/\lambda\{\gamma\}$ is called the *modulus of* D and is denoted by mod D. As is well known, D can be mapped conformally onto an annulus $1 < |z| < q$ where $\log q = \text{mod } D$.

For $P > 0$, the doubly connected region

$$D_P = \{[-1,\ 0] \cup [P,\ \infty]\}^c$$

where the brackets express a closed interval on the real axis, is called *Teichmüller's extremal region.* It has the following extremal property (Teichmüller [23]): *Let D be a doubly connected region such that one component of D^c contains the point $z = 0$ as well as a point on $|z| = 1$ and the other contains the point $z = \infty$ as well as a point on $|z| = P$. Then $\text{mod } D \leqq \text{mod } D_P$ and the equality holds if and only if D is a region obtained by rotating D_P about the origin.*

12. It was proved by Teichmüller [23] that $\Psi(P) = \exp(\text{mod } D_P)$ is a continuous function of P such that

$$\lim_{P\to\infty}\frac{\Psi(P)}{P} = 16\ . \tag{1}$$

It is easy to see that

$$\log\Psi\left(\frac{1}{P}\right) = \frac{\pi^2}{\log\Psi(P)}\ . \tag{2}$$

On combining (1) and (2), we have

$$\log\Psi(P) \sim \frac{\pi^2}{\log\dfrac{1}{P}} \qquad \text{for } P\to 0\ . \tag{3}$$

13. The following result will be used later:

LEMMA 1. *Let*

$$A = (1 < |z| < q),$$
$$\Gamma = (|z| = 1),$$

and

$$E_\theta = \{z;\ |z| = q,\ |\arg z| \leqq \theta\}.$$

Then

$$\delta_A(\Gamma,\ E_\theta) \sim \frac{1}{\pi}\log\frac{1}{\theta} \qquad \textit{for } \theta \to 0.$$

Proof. $\delta_A(\Gamma,\ E)$ is equal to the extremal length $\lambda\{\gamma\}$ where $\{\gamma\}$ is the family of all analytic curves in A connecting Γ with E_θ (cf. Wolontis [25]). By (VIII) and (XIII), it is equal to $\delta_Q(E'_\theta,\ E''_\theta)/4$ where

$$Q = (1/q < |z| < q) \cap (\Im z > 0),$$
$$E'_\theta = \{z;\ |z| = 1/q,\ 0 \leqq \arg z \leqq \theta\},$$

and

$$E''_\theta = \{z;\ |z| = q,\ 0 \leqq \arg z \leqq \theta\}.$$

Map Q onto the upper half-plane in such a way that $1/q$ and q correspond to 0 and 1, respectively. Let $-\alpha$ and $1 + \beta$ ($\alpha,\ \beta > 0$) be the images of $e^{i\theta}/q$ and $qe^{i\theta}$, respectively. It is not difficult to see that

$$\begin{cases} \alpha \sim c\dfrac{\theta^2}{q} \\ \beta \sim c'q\theta^2 \end{cases} \qquad \text{for } \theta \to 0$$

where c and c' are constants independent of θ. The region obtained by deleting the intervals $[-\infty,\ -\alpha]$, $[0,\ 1]$, and $[1 + \beta,\ \infty]$ from the extended plane is conformally equivalent to Teichmüller's extremal region with

$$P = \frac{\alpha\beta}{1 + \alpha + \beta} \sim c''\theta^4 \qquad (\theta \to 0).$$

Therefore, on applying (VIII) again, we get $\delta_A(\Gamma,\ E_\theta) = \pi/(4\log\Psi(P))$ and, by (3),

$$\delta_A(\Gamma,\ E_\theta) \sim \frac{1}{4\pi}\log\frac{1}{P} \sim \frac{1}{\pi}\log\frac{1}{\theta} \qquad \text{for } \theta \to 0.$$

6. Koebe's distortion theorem.

14. The following is a slight modification of the original form of Koebe's well-known distortion theorem, which will be used frequently:

Let $\varphi(z)$ be a function which is univalent and regular in $|z| < \varepsilon_0$ with $\varphi(0) = 0$ and $\varphi'(0) = 1$. Then there are numbers $a(\varepsilon)$ and $b(\varepsilon)$ which are independent of φ and have the properties that

$$a(\varepsilon) \leqq |\varphi(z)| \leqq b(\varepsilon) \qquad \text{on } |z| = \varepsilon < \varepsilon_0$$

and

$$\lim_{\varepsilon \to 0} \frac{a(\varepsilon)}{\varepsilon} = \lim_{\varepsilon \to 0} \frac{b(\varepsilon)}{\varepsilon} = 1 .$$

In fact, we may take $a(\varepsilon) = \varepsilon\varepsilon_0^2/(\varepsilon + \varepsilon_0)^2$ and $b(\varepsilon) = \varepsilon\varepsilon_0^2/(\varepsilon - \varepsilon_0)^2$.

7. Quasi-conformal mappings.

15. In Chapters IV and V, we shall make use of quasi-conformal mappings to illustrate our results by examples. As in the type problem of Riemann surfaces, they are utilized to replace a given region by a simpler one.

A sense-preserving topological mapping $w = T(z)$ of a region D onto another is said to be *quasi-conformal* if there exists a finite number K such that mod $T(Q) \leqq K$ mod Q for any quadrilateral $Q \subset D$ (Ahlfors [1]). Here, mod Q of a quadrilateral Q means the extremal distance between two opposite sides of Q. The minimum value of K is called the *maximal dilatation of* T.

For the proofs of the following properties (I)–(III), the reader is referred to Ahlfors [1]:

(I) If T is quasi-conformal of maximal dilatation K, then mod $T(A) \leqq K$ mod A for any doubly connected region $A \subset D$.

(II) Let E be a set which is contained in a finite number of analytic arcs. Let D be a region containing E, and let T be a topological mapping of D which is quasi-conformal in $D - E$. Then it is quasi-conformal in D with the same maximal dilatation.

(III) If T is a topological mapping of class C^1, then the maximal dilatation is given by $K = \sup_{z \in D} (|T_z| + |T_{\bar{z}}|)/(|T_z| - |T_{\bar{z}}|)$ where T_z and $T_{\bar{z}}$ are complex derivatives.

(IV) Let $\{\gamma\}$ be a family of curves in D. Let T be a quasi-conformal mapping of class C^1 with the maximal dilatation K. Then

$$\lambda\{T(\gamma)\} \leqq K\lambda\{\gamma\} .$$

The proof is found in Hersch [9][1].

REMARK. Even if T is not of class C^1 throughout D, this inequality holds under, e.g., the following restriction: T is of C^1 in D except for a countable number of analytic arcs clustering nowhere in D, i.e., every point of D has a neighborhood intersecting at most a finite number of the arcs, and every γ is the union of a countable number of analytic arcs clustering nowhere in D. This generalization will be needed in No. 35.

III. CIRCULAR AND RADIAL SLIT DISKS

8. Circular slit disks.

16. Let D be a plane region containing the point $z = 0$, and let Γ be a boundary component. The problem of minimizing $M(f)$ in $\mathfrak{F}_\Gamma$ for a region of finite connectivity has been discussed by Rengel [14]. To consider it for a region of arbitrary connectivity, in particular to show the uniqueness of the minimizing function, Sario [16] introduced the functional

$$J(f) = \int_{\partial D} \log |f| \cdot d \arg f \qquad (f \in \mathfrak{F}_\Gamma) .$$

Here the line integral means $\lim_{n\to\infty} \int_{\partial D_n} \log |f| \cdot d \arg f$ for an exhaustion $D_n \uparrow D$; the limiting value exists and is independent of the exhaustion. He proved the existence of a function g_0 such that

$$M(g_0) = m(g_0)$$

and

$$2\pi \log M(g_0) = J(f) - D(\log |f| - \log |g_0|)$$

for all $f \in \mathfrak{F}_\Gamma$, where the second term means the Dirichlet integral over D. Evidently g_0 is the unique function which minimizes $J(f)$.

From these relations we can derive the following facts (Sario [16]):

(I) *There exists a function $g_0 \in \mathfrak{F}_\Gamma$ such that $M(g_0) = \min_{f\in\mathfrak{F}_\Gamma} M(f) = r(\Gamma)$. If $r(\Gamma) < \infty$, the minimizing function is determined uniquely. It maps D onto a circular slit disk $|w| < r(\Gamma)$, where the area of slits, i.e., $g_0(\partial D - \Gamma)^*$, vanishes,*

(II) *Let $0 \in D_n \uparrow D$ be an exhaustion and let Γ_n be the component of ∂D_n separating D_n from Γ. Then*

$$r(\Gamma) = \lim_{n\to\infty} r(\Gamma_n) .$$

If $r(\Gamma) < \infty$, the sequence $\{g_n\}$ of the minimizing functions on D_n converges to g_0 uniformly on each compact set in D.

17. By making use of this result, we can express $r(\Gamma)$ in terms of extremal length. Let ε_0 be a small number such that $|z| \leqq \varepsilon_0$ is contained in D. For $0 < \varepsilon < \varepsilon_0$, the numbers $a(\varepsilon)$ and $b(\varepsilon)$ were defined in No. 14. The following theorem has been proved, in essence, by Jurchescu [10]:

THEOREM 2. *Let $\{\gamma\}_\varepsilon$ be the family of all closed curves in $D_\varepsilon = D - (|z| \leqq \varepsilon)$ which separate Γ from the point $z = 0$. Then*

$$\log \frac{r(\Gamma)}{b(\varepsilon)} \leqq \frac{2\pi}{\lambda\{\gamma\}_\varepsilon} \leqq \log \frac{r(\Gamma)}{a(\varepsilon)}$$

and, therefore,

$$\log r(\Gamma) = \lim_{\varepsilon \to 0} \left(\log \varepsilon + \frac{2\pi}{\lambda\{\gamma\}_\varepsilon} \right).$$

The result remains valid if the γ are restricted to analytic curves.

Proof. Consider the metric given by $\rho = |g_0'| / |g_0|$. Since the area of the circular slits is zero, $\iint_{D_\varepsilon} \rho^2 \, dxdy \leqq 2\pi \log (r(\Gamma)/a(\varepsilon))$. Therefore,

$$\lambda\{\gamma\}_\varepsilon \geqq (2\pi)^2/2\pi \log (r(\Gamma)/a(\varepsilon)) .$$

To prove the left inequality, take an exhaustion $D_n \uparrow D$ and consider the family $\{\gamma_n\}_\varepsilon$ of all closed curves γ_n in $D_n - (|z| \leqq \varepsilon)$ separating Γ_n from $z = 0$. Since D_n is of finite connectivity, the proposition (IX), No. 9, shows that $2\pi/\lambda\{\gamma_n\}_\varepsilon \geqq \log (r(\Gamma_n)/b(\varepsilon))$. When we take the limit for $n \to \infty$, we have by virtue of the relation $\lambda\{\gamma\}_\varepsilon \leqq \lambda\{\gamma_n\}_\varepsilon$ that

$$2\pi/\lambda\{\gamma\}_\varepsilon \geqq \log (r(\Gamma)/b(\varepsilon)) .$$

18. The following criterion for weakness due to Grötzsch [7] will be useful in the next chapter:

THEOREM 3. *In order that Γ be weak, it is necessary and sufficient that, for an arbitrary positive number l, there exist a finite number of doubly connected regions $A_1, A_2, \cdots A_k$ in $D - (|z| \leqq \varepsilon)$ satisfying the following conditions:*

(i) *The A_ν are mutually disjoint,*

(ii) *A_ν separates Γ from $(|z| \leqq \varepsilon)$ $(\nu = 1, 2, \cdots, k)$ and separates $A_{\nu-1}$ from $A_{\nu+1}$ $(\nu = 2, 3, \cdots, k-1)$,*

(iii)

$$\sum_{\nu=1}^{k} \operatorname{mod} A_\nu \geqq l .$$

Proof. Sufficiency: By (V), No. 9, and by Theorem 2, $l \leqq \sum_{\nu=1}^{k} \bmod A_\nu \leqq 2\pi/\lambda\{\gamma\}_\varepsilon \leqq \log(r(\Gamma)/(\varepsilon))$. Therefore, $r(\Gamma)=\infty$ and, by Theorem 1, Γ is weak.

Necessity: Take an exhaustion $(|z| \leqq \varepsilon) \subset D_1 \subset D_2 \subset \cdots \subset D_n \subset \cdots \uparrow D$ and consider the extremal function g_n on D_n. By Koebe's distortion theorem, No. 14, the image of $|z|=\varepsilon$ is contained in $a(\varepsilon) \leqq |w| \leqq b(\varepsilon)$, so that the set $b(\varepsilon) < |w| < r(\Gamma_n)$ minus the circular slits is contained in the image of $D_n - (|z| \leqq \varepsilon)$. From the annulus $b(\varepsilon) < |w| < r(\Gamma_n)$, delete all the concentric circles containing the circular slits. Then we get a finite number of concentric annuli $A_1', A_2', \cdots, A_k'$ such that $\sum_{\nu=1}^{k} \bmod A_\nu' = \log\ (r(\Gamma_n)/b(\varepsilon))$. Since $r(\Gamma)=\lim_{n\to\infty} r(\Gamma_n)=\infty$, we can take n so large that the right hand side is greater than the given l. The inverse images $A_1, A_2, \cdots, A_k$ of $A_1', A_2', \cdots, A_k'$ are what we desired.

REMARK. We see from this theorem that the weakness of Γ depends merely on the configuration of ∂D near l. Furthermore, by (I), No. 15, the weakness is invariant under quasi-conformal mappings.

9. Radial slit disks for special regions.

19. Unlike the case of the functional $M(f)$, the function maximizing $m(f)$ does not exist in general; by slightly modifying the example given by Strebel [20], we get a region on which $m(f) < R(\Gamma) = \sup_{f\in\mathfrak{F}_\Gamma} m(f)$ for all $f \in \mathfrak{F}_\Gamma$.

Under a restriction, however, we get a result analogous to that of No. 15. Let G be a region containing the point $z=0$ and such that a component Γ of ∂G consists of a closed analytic curve which is isolated, i.e., $\overline{\partial G - \Gamma} \cap \Gamma = \phi$. Let $\mathfrak{A}_\Gamma$ be the subfamily of $\mathfrak{F}_\Gamma$ consisting of all functions with $M(f)=m(f)$. On this family Sario [17, 18] introduced the functional

$$I(f) = 2\pi \log m(f) - \int_{\partial D - \Gamma} \log |f| \cdot d \arg f$$

and proved the existence of a function $f_0 \in \mathfrak{A}_\Gamma$ such that

$$(4) \qquad 2\pi \log m(f_0) = I(f) + D(\log |f| - \log |f_0|)$$

for all $f \in \mathfrak{A}_\Gamma$. Evidently f_0 is the unique maximizing function of $I(f)$ in $\mathfrak{A}_\Gamma$.

We can derive from this relation the following facts (Sario [18]), which have been obtained by Rengel [14] for a region G of finite connectivity:

(I) *$R(\Gamma)$ is finite. f_0 is the unique function maximizing $m(f)$ in $\mathfrak{A}_\Gamma$. It maps G onto a radial slit disc $|w| < R(\Gamma)$, where the area of slits, i.e., $f_0(\partial G - \Gamma)^*$, vanishes.*

(II) *Let $\{G_n\}$ be a sequence of regions such that $0 \in G_n \uparrow G$ and ∂G_n consists of Γ and a finite number of closed analytic curves. Then*

$$R(\Gamma;\ G) = \lim_{n\to\infty} R(\Gamma_n;\ G_n)$$

and the sequence $\{f_n\}$ of the maximizing functions on G_n converges to f_0 uniformly on each compact set in $G \cup \Gamma$.

20. Let $\{\gamma\}_\varepsilon$ be the family of rectifiable curves which connect $|z| = \varepsilon$ with Γ in $G - (|z| \leqq \varepsilon)$. In a method similar to the proof of Theorem 2 we can obtain the following relations:

$$\frac{\left(\log \dfrac{R(\Gamma)}{b(\varepsilon)}\right)^2}{\log \dfrac{R(\Gamma)}{a(\varepsilon)}} \leqq 2\pi\lambda\{\gamma\}_\varepsilon \leqq \log \frac{R(\Gamma)}{a(\varepsilon)}\,, \tag{5}$$

$$\log R(\Gamma) = \lim_{\varepsilon\to 0} (\log \varepsilon + 2\pi\lambda\{\gamma\}_\varepsilon)\,. \tag{6}$$

Here $\{\gamma\}_\varepsilon$ can be replaced by the subfamily of analytic curves.

10. Characterizations of $R(\Gamma)$.

21. Let D be an arbitrary region containing the point $z = 0$. Let $\{\Omega_n\}_{n=1}^\infty$ be a defining sequence of Γ such that $0 \notin \Omega_n$ $(n = 1, 2, \cdots)$. Then $G_n = D - \Omega_n$ is a region and its boundary component $\Gamma_n - \partial G_n \cap \partial\Omega_n$ satisfies the condition of No. 19.

THEOREM 4. *$\{R(\Gamma_n, G_n)\}_{n=1}^\infty$ is an increasing sequence and $R(\Gamma) = \lim_{n\to\infty} R(\Gamma_n;\ G_n)$.*

Proof. $\{R(\Gamma_n;\ G_n)\}$ is an increasing sequence by (6).

For an arbitrary $\varepsilon > 0$, there exists an $f(z) \in \mathfrak{F}_\Gamma$ such that $m(f) > R(\Gamma) - \varepsilon/2$. Then there exists an n_0 such that the m of this $f(z)$ on G_n (we denote it by $m_n(f)$) has the property that $m_n(f) > m(f) - \varepsilon/2$ whenever $n \geqq n_0$. Therefore, $R(\Gamma_n;\ G_n) \geqq m_n(f) > R(\Gamma) - \varepsilon$ and $\underline{\lim}_{n\to\infty} R(\Gamma_n;\ G_n) \geqq R(\Gamma)$.

Next, let A_n be the doubly connected region bounded by Γ_n and Γ. Then Γ is an isolated boundary component of the region $\tilde{G}_n = G_n \cup A_n \cup \Gamma_n$. Γ is not necessarily a closed analytic curve, but from the result of No. 19 we can see the existence of the function $\tilde{f}_n(z)$ in $\mathfrak{F}_\Gamma$ of $\tilde{G}_n$ such that $m(\tilde{f}_n) = R(\Gamma;\ \tilde{G}_n)$. Evidently $\tilde{f}_n(z)$ belongs to $\mathfrak{F}_\Gamma$ of D. By (6),

$R(\Gamma_n;\ G_n) \leqq R(\Gamma;\ \tilde{G}_n)$. Consequently, $R(\Gamma_n;\ G_n) \leqq R(\Gamma;\ \tilde{G}_n) = m(\tilde{f}_n) \leqq R(\Gamma)$ and $\overline{\lim}_{n\to\infty} R(\Gamma;\ G_n) \leqq R(\Gamma)$.

This reasoning remains valid for the case where $R(\Gamma) = \infty$.

REMARK. Combining Theorem 4 with Theorem 1, we see that $\lim_{n\to\infty} R(\Gamma_n;\ G_n) < \infty$ implies the strength of Γ. This fact was proved by Sario [17].

22. Let $\{\gamma\}_\varepsilon$ be the family of curves $\gamma : z = z(t)$ $(0 < t < 1)$ in $D - (|z| \leqq \varepsilon)$ such that $\bigcap_{\varepsilon>0} \overline{\{z(t);\ 0 < t < \varepsilon\}} \subset (|z| = \varepsilon)$ and $\bigcap_{\varepsilon>0} \overline{\{z(t);\ 1 - \varepsilon < t < 1\}} \subset \Gamma$. Let $\{\gamma_n\}_\varepsilon$ be the corresponding family in G_n. Strebel [22] has proved the relation $\lambda\{\gamma\}_\varepsilon = \lim_{n\to\infty} \lambda\{\gamma_n\}_\varepsilon$. On combining this with (5), (6), and Theorem 4, we have

THEOREM 5.

$$\frac{\left(\log \dfrac{R(\Gamma)}{b(\varepsilon)}\right)^2}{\log \dfrac{R(\Gamma)}{a(\varepsilon)}} \leqq 2\pi\lambda\{\gamma\}_\varepsilon \leqq \log \frac{R(\Gamma)}{a(\varepsilon)} ,$$

$$\log R(\Gamma) = \lim_{\varepsilon\to 0} (\log \varepsilon + 2\pi\lambda\{\gamma\}_\varepsilon) .$$

Here γ can be restricted to the curve which is the union of a countable number of analytic arcs which cluster nowhere in D (cf. No. 15, Remark).

REMARK. The exponential of the right hand side of the second relation was called "extremal Durchmesser" by Strebel [22]. On combining Theorem 5 with Theorem 1, or directly from (XI), No. 10, we see that $\lambda\{\gamma\}_\varepsilon < \infty$ implies the strength of Γ. This result was generalized for open Riemann surfaces by Constantinescu [4].

23. For an exhaustion $D_n \uparrow D$ in the ordinary sense, it has not been proved whether $\lim_{n\to\infty} R(\Gamma_n;\ D_n)$ exists or not. We obtain merely the following

THEOREM 6. *Let Δ be a region such that $0 \in \Delta$, $\bar{\Delta} \subset D$, and bounded by a finite number of closed analytic curves. Denote by Γ_Δ the component of $\partial\Delta$ which separates Δ from Γ. Then*

$$R(\Gamma) = \underline{\lim}_{\Delta\to D} R(\Gamma_\Delta;\ \Delta) ,$$

where the right hand side is a directed limit.

Proof. For $\varepsilon > 0$, there exists by Theorem 4 an n such that

$R(\Gamma) - \varepsilon < R(\Gamma_n;\ G_n)$. By Theorem 5 $R(\Gamma_n;\ G_n) \leqq R(\Gamma_\Delta;\ \Delta)$ for any $\Delta \supset \Gamma_n \cup \{0\}$. Therefore, $R(\Gamma) \leqq \underline{\lim}_{\Delta\to D} R(\Gamma_\Delta;\ \Delta)$. On the other hand, for $\varepsilon > 0$ and a compact set $K \subset D$, take an n_0 such that $K \subset G_{n_0}$. There exists, by (II), No. 19, a $\Delta \subset G_{n_0}$ such that $R(\Gamma_\Delta;\ \Delta) \subset R(\Gamma_{n_0};\ G_{n_0}) + \varepsilon$, and, therefore, $R(\Gamma_\Delta;\ \Delta) < R(\Gamma) + \varepsilon$. Consequently $\underline{\lim}_{\Delta\to D} R(\Gamma_\Delta;\ \Delta) \leqq R(\Gamma)$.

REMARK. On combining Theorem 6 with Theorem 1 we see that $\underline{\lim}_{\Delta\to D} R(\Gamma_\Delta;\ \Delta) < \infty$ implies the strength of Γ. Sario [18] has shown that Γ is strong if $\overline{\lim}_{\Delta\to D} R(\Gamma_\Delta;\ \Delta) < \infty$.

11. Unsolved problems.

24. As we pointed out in No. 7, the following problem has not been solved:

(1) Is $R(\Gamma) < \infty$ necessary for the strength of Γ ?

Since the maximizing function of $m(f)$ in $\mathfrak{F}_\Gamma$, or equivalently the minimizing function of diam $\varphi(\Gamma)$ in $\mathfrak{B}$, does not exist in general, the case is different from that of a weak boundary component. The example of Strebel [20] stated in No. 19 is for $R(\Gamma) > \infty$, and it does not answer this question.

Let $\{G_n\}_{n=1}^{\infty}$ be the sequence introduced in No. 21 and let $f_n(z)$ be the extremal function on G_n. Since $\{f_n\}_{n=1}^{\infty}$ is a normal family, we may assume that f_n converges to a univalent function $f(z)$. One can imagine that, if $R(\Gamma) = \infty$, then $f(\Gamma)$ would be a point. However, we can only prove that $f(\Gamma)$ consists of the point $w = \infty$ and possibly of radial segments emanating from it whose arguments form a set of measure zero (Strebel [22]). Such line segments appear in our Example 10, Nos. 39, 40. Nevertheless the boundary component of this example is unstable, because we can map it onto a region such that $f(\Gamma)$ is a point and $f(\partial D - \Gamma)$ consists of circles (No. 39).

We have several other unsolved problems as follows:

(2) Is strength a boundary property?

(3) Is $\overline{\lim}_{\Delta\to D} R(\Gamma_\Delta;\ \Delta)$ equal to $\underline{\lim}_{\Delta\to D} R(\Gamma_\Delta;\ \Delta)$?

(4) Is strength preserved under quasi-conformal mappings?

IV. CRITERIA FOR WEAKNESS AND INSTABILITY

In this chapter we consider Problem A presented in No. 4. Several sufficient conditions for weakness have been obtained by Savage [19]. Here we shall consider some special regions and attempt to get more concrete necessary or sufficient conditions.

12. Boundary on the positive real axis.

25. Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ be sequences of positive numbers such that

$$1 < b_{n-1} \leqq a_n < b_n \qquad (n = 1, 2, \cdots),$$

$$\lim_{n\to\infty} a_n = \infty .$$

Denote by $[a, b]$ the closed interval on the real axis. Then

$$D = (|z| < \infty) - \bigcup_{n=1}^{\infty} [b_{n-1}, a_n]$$

is a region and $\Gamma = \{\infty\}$ is its boundary component. The present section is devoted to discussing the following problem: *When is Γ weak and when is it unstable?*

26. THEOREM 7. (i) *If*

$$\sum_{n=1}^{\infty} \left(\frac{b_n}{a_n} - 1\right) = \infty , \tag{7}$$

then Γ is weak.

(ii) *If*

$$\lim_{n\to\infty} \frac{b_n}{a_n} = 1^3 \tag{8}$$

and

$$\sum_{n=1}^{\infty} \frac{1}{\log \dfrac{1}{(b_n/a_n) - 1}} < \infty \tag{9}$$

then Γ is unstable.

Proof. (i) Consider the annuli $A_n = (a_n < |z| < b_n)$ $(n = 1, 2, \cdots)$. Since $\sum \mathrm{mod}\, A_n = \sum \log (b_n/a_n) = \infty$, Theorem 3 shows that Γ is weak.

(ii) Let $A_1, A_2, \cdots, A_k$ be doubly connected regions satisfying the conditions (i) and (ii) of Theorem 3. For any A_ν, there exists an n such that A_ν passes through the open interval (a_n, b_n) and a component of A_ν contains 0 as well as a_n. The region

$$D^{(n)} = \{[0, a_n] \cup [b_n, \infty]\}^c$$

is conformally equivalent to Teichmüller's extremal region with $P = (b_n/a_n) - 1$. By the extremal property of $D^{(n)}$, No. 11, the sum of the

[3] If $\overline{\lim}_{n\to\infty} b_n/a_n > 1$, then Γ is weak by (i), Theorem 7

moduli of all such A_ν does not exceed mod $D^{(n)} = \log \Psi((b_n/a_n) - 1)$.

$$\sum_{\nu=1}^{k} \operatorname{mod} A_\nu \leqq \sum_{n=1}^{\infty} \log \Psi\left(\frac{b_n}{a_n} - 1\right). \tag{10}$$

By (3), No. 12,

$$\log \Psi\left(\frac{b_n}{a_n} - 1\right) \sim \frac{\pi^2}{\log \dfrac{1}{(b_n/a_n) - 1}}.$$

Therefore, the right hand side of (10) converges and, by Theorem 3, Γ is unstable.

EXAMPLE 1. $a_n = 2n + 1$, $b_n = 2n + 2$. Evidently (7) diverges so that *Γ is weak.*

EXAMPLE 2. $a_n = n^k$, $b_n = n^k + 1$ $(k > 1)$. Since (7) converges and (9) diverges, we cannot decide by Theorem 7 (see also No. 27).

EXAMPLE 3. $a_n = e^n$, $b_n = e^n + 1$. Similarly, we cannot decide (see also No. 27).

EXAMPLE 4. $a_n = e^{n^\alpha}$, $b_n = e^{n^\alpha} + 1$ $(\alpha > 1)$. *Γ is unstable* by (ii).

27. We derive another criterion applicable to Examples 2 and 3. To this end, we first prove

LEMMA 2. *For the doubly connected region*

$$A_h = (1 < |z| < q) - [1 + h,\ q)$$

where $h > 0$ and q is fixed,

$$\operatorname{mod} A_h \sim \frac{\pi^2}{2 \log \dfrac{1}{h}} \qquad \textit{for } h \to 0.$$

Proof. By (VIII), No. 9, mod $A_h = 4\pi/\lambda\{\gamma\}$ where $\{\gamma\}$ is the family of rectifiable curves in $Q = A_h \cap (\Im z > 0)$ joining $[-q,\ -1]$ with $[1,\ 1 + h]$. Map Q conformally onto the upper half-plane in such a manner that $-q$, -1, 1 correspond to $-\infty$, -1, 0, respectively. The image P of $1 + h$ has the property that

$$P \sim ch^2 \qquad \text{for } h \to 0$$

where c is a constant independent of h. From (VIII), No. 9, we conclude that

$$\operatorname{mod} A_h = \log \Psi(P) \sim \frac{\pi^2}{\log \dfrac{1}{P}} \sim \frac{\pi^2}{2 \log \dfrac{1}{h}} \qquad (h \to 0) .$$

THEOREM 8. *Suppose that* $\lim_{n\to\infty} b_n/a_n = 1$. *If* a_{n+1}/a_n *is bounded away from* 1, *then* Γ *is weak if and only if*

$$\sum_{n=1}^{\infty} \frac{1}{\log \dfrac{1}{(b_n/a_n) - 1}} = \infty .$$

Proof. If the series converges, Γ is unstable by (ii) of Theorem 7.

Conversely, suppose that the series diverges. The doubly connected region $A_n = (a_n < |z| < a_{n+1}) - [b_n, a_{n+1})$ is conformally equivalent to the region $A'_n = (1 < |z| < a_{n+1}/a_n) - [b_n/a_n, a_{n+1}/a_n)$. By the assumption $1 < 1 + \delta < a_{n+1}/a_n$ and, therefore, $A''_n = (1 < |z| < 1 + \delta) - [b_n/a_n, 1+\delta) \subset A'_n$ so that $\operatorname{mod} A''_n \leqq \operatorname{mod} A_n$. By Lemma 2

$$\operatorname{mod} A''_n \sim \frac{\pi^2}{2 \log \dfrac{1}{(b_n/a_n) - 1}} \qquad (n \to \infty) .$$

Consequently, the assumption implies that $\sum \operatorname{mod} A_n = \infty$, and we infer from Theorem 3 that Γ is weak.

EXAMPLE 3 (No. 26). $a_n = e^n$, $b_n = e^n + 1$. By Theorem 8, Γ *is weak.*

EXAMPLE 2 (No. 26). $a_n = n^k$, $b_n = n^k + 1$ $(k > 1)$. Since $a_{n+1}/a_n = (n + 1)^k/n^k$ is not bounded away from 1, the above theorem is not applicable. However, we can see as follows that Γ *is weak*. For simplicity, we consider the case $k = 2$; the general case can be treated in a similar fashion. Consider the region $A_n = (a_{2^n} < |z| < a_{2^{n+1}}) - [b_{2^n}, a_{2^{n+1}})$, which is conformally equivalent to $(1 < |z| < 4) - [1 + 2^{-2n}, 4)$. By Lemma 2, $\operatorname{mod} A_n \sim \pi^2/(4n \log 2)$ for $n \leftarrow \infty$ and $\sum \operatorname{mod} A_n = \infty$. It follows from Theorem 3 that Γ is weak.

More generally, this result can be stated as follows:

THEOREM 8′. *Suppose that* $\lim_{n\to\infty} b_n/a_n = 1$ *and that there exists a subsequence* $\{n_i\} \subset \{n\}$ *such that* $a_{n_{i+1}}/a_{n_i}$ *is bounded away from* 1 *and*

$$\sum_{i=1}^{\infty} \frac{1}{\log \dfrac{1}{(b_{n_i}/a_{n_i}) - 1}} = \infty . \tag{12}$$

Then Γ *is weak.*

28. When a_{n+1}/a_n is not bounded away from 1, we may also apply the following criterion:

THEOREM 9. *Suppose* $\lim_{n\to\infty} b_n/a_n = 1$ *and* $\lim_{n\to\infty} a_{n+1}/a_n = 1$. *If*

$$\lim_{n\to\infty} \frac{\log (b_n/a_n)}{\log (a_{n+1}/a_n)} \tag{13}$$

exists, then

$$\sum_{n=1}^{\infty} \frac{\log (a_{n+1}/a_n)}{\log \dfrac{1}{\left(\dfrac{b_n}{a_n}\right)^{1/\log(a_{n+1}/a_n)} - 1}} = \infty \tag{14}$$

implies that Γ *is weak.*

Proof. Consider the doubly connected region $A'_n = (1 < |z| < q_n) - [1 + h_n, q_n)$ $(n = 1, 2, \cdots)$, where $0 < h_n < q_n - 1$ and $\lim_{n\to\infty} q_n = 1$. Map the annulus $1 < |z| < q_n$ onto $1 < |w| < e$ by the quasi-conformal mapping

$$w = T_n(z) = r^{1/\log q_n} e^{i\theta} \qquad (z = re^{i\vartheta}) .$$

Its dilatation equals $1/\log q_n$ provided n is so large that $q_n < e$. The image of A'_n is $A''_n = (1 < |w| < e) - [(1 + h_n)^{1/\log q_n}, e)$. From (I), No. 15, we have

$$\log q_n \cdot \operatorname{mod} A''_n \leqq \operatorname{mod} A'_n . \tag{15}$$

Now suppose that $\lim_{n\to\infty} (\log (1 + h_n))/\log q_n$ exists. If

$$\lim_{n\to\infty} (1 + h_n)^{1/\log q_n} > 1 ,$$

then $\operatorname{mod} A''_n$ and $\log \{1/[(1 + h_n)^{1/\log q_n} - 1]\}$ are bounded and bounded away from zero. Hence the divergence of

$$\sum_{n=1}^{\infty} \frac{\log q_n}{\log \dfrac{1}{(1 + h_n)^{1/\log q_n} - 1}} \tag{16}$$

implies that $\sum_{n=1}^{\infty} \log q_n \cdot \operatorname{mod} A''_n = \infty$ and, by (14), that $\sum_{n=1}^{\infty} \operatorname{mod} A'_n = \infty$. If $\lim_{n\to\infty} (1 + h_n)^{1/\log q_n} = 1$ we obtain by Lemma 2

$$\log A''_n \sim \frac{\pi^2}{2 \log \dfrac{1}{(1 + h_n)^{1/\log q_n} - 1}} \qquad (n \to \infty) .$$

Therefore, the divergence of (16) again implies that of $\sum_{n=1}^{\infty} \operatorname{mod} A'_n$.

In the given region, consider $A_n = (a_n < |z| < a_{n+1}) - [b_n, a_{n+1})$. It is conformally equivalent to the above A'_n for $1 + h_n = b_n/a_n$ and $q_n = a_{n+1}/a_n$. Therefore, $\sum_{n=1}^{\infty} \mathrm{mod}\, A_n = \infty$ and Γ is weak.

This criterion is applicable to Example 2.

EXAMPLE 5. $a_n = n$, $b_n = n + e^{-n}$. In this case (7) converges and (9) diverges, so that we cannot decide by Theorem 7. Since a_{n+1}/a_n is not bounded away from zero, we cannot apply Theorem 8.[4] For every subsequence such that $\underline{\lim}_{i\to\infty} a_{n_{i+1}}/a_{n_i} > 1$, (12) converges, and we cannot use Theorem 8′. (14) also converges and, therefore 9 is inapplicable. We have not been able to decide whether Γ is weak or unstable. In general, for $a_n = n$, $b_n = n + e^{-n^\alpha}$ $(\alpha > 0)$, Γ is unstable for $\alpha > 1$ but it is unknown if it remains true for $0 < \alpha \leqq 1$.

13. A generalization.

29. Consider the case where the intervals are distributed on the whole real axis. We treat again the simplest case.

PROBLEM. *Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=0}^{\infty}$ be the sequence of positive numbers such that*

$$0 < b_{n-1} \leqq a_n < b_n \qquad (n = 1, 2, \cdots)$$

$$\lim_{n\to\infty} a_n = \infty\ .$$

Consider the region

$$\tilde{D} = (|z| < \infty) - \bigcup_{n=1}^{\infty} [b_{n-1}, a_n] - \bigcup_{n=1}^{\infty} [-a_n, -b_{n-1}]\ .$$

Under what condition is $\tilde{\Gamma} = \{\infty\}$ a weak boundary component of $\tilde{D}$?

This problem can be reduced to the case which we discussed in the previous section. More precisely, let $\Gamma = \{\infty\}$ be a boundary component of

$$D = (|z| < \infty) - \bigcup_{n=1}^{\infty} [b_{n-1}, a_n]\ ;$$

then we have

THEOREM 10. *$\tilde{\Gamma}$ is weak if and only if Γ weak.*

Proof. If Γ is unstable, then, since $\tilde{D} \subset D$, $\tilde{\Gamma}$ is unstable by the definition.

[4] The author is indebted to Professor R. Redheffer for the argument that follows in this example.

Suppose that $\tilde{\Gamma}$ is unstable. Since weakness is a boundary property (No. 18), we may assume without loss of generality that $b_0 > 1$. By Theorem 2, $\lambda\{\gamma\} > 0$ where $\{\gamma\}$ is the family of curves in $\tilde{D} - (|z| \leqq 1)$ separating $\tilde{\Gamma}$ from $|z| = 1$. Let $\{\gamma_1\}$ be the family consisting of curves in the upper half of $\tilde{D} - (|z| \leqq 1)$ connecting $(1, \infty) - \bigcup_{n=1}^{\infty} [b_{n-1}, a_n]$ with $(-\infty, -1) - \bigcup_{n=1}^{\infty} [-a_n, -b_{n-1}]$. Let $\{\gamma_1'\}$ be its subfamily consisting of curves whose end points are symmetric with respect to the origin. Then, by (VIII), No. 9,

$$\lambda\{\gamma_1'\} \geqq \lambda(\gamma_1\} = \lambda\{\gamma\}/2 > 0 .$$

Consider the region $\Delta = (|\zeta| < \infty) - \bigcup_{n=1}^{\infty} [b_{n=1}^2, a_n^2]$ and its boundary component $(\zeta = \infty)$. Let $\{\gamma^*\}$ be the family of curves in $\Delta - (|\zeta| \leqq 1)$ separating ∞ from $|\zeta| \leqq 1$. By making use of the mapping $\zeta = z^2$, we can immediately see that $\lambda\{\gamma^*\} = \lambda\{\gamma_1'\}$ and, therefore, $(\zeta = \infty)$ is an unstable boundary component of Δ.

The mapping

$$\zeta = T(z) = r^2 e^{i\theta} \qquad (z = re^{i\theta})$$

is quasi-conformal and maps D onto Δ, $z = \infty$ onto $\zeta = \infty$. Since weakness is preserved under quasi-conformal mappings (No. 18), Γ is unstable.

REMARK. Using the same method, we can also prove Theorem 10 when the intervals are distributed on k half-lines $re^{i2\pi\nu/k}$ $(0 \leqq r < \infty)$, $\nu = 0, 1, \cdots, k$.

14. Criteria for arbitrary regions.

30. Let D be a plane region such that $\Gamma = \{\infty\}$ is a boundary component. If D is contained in another region discussed in preceding sections and $\{\infty\}$ is its unstable boundary component, then, by the definition of instability, Γ is an unstable boundary component of D.

If such a condition is not satisfied, the following criterion may be applicable. It is a simple generalization of (ii) of Theorem 7, and we omit the proof.

THEOREM 11. *Let D be a region such that $0 \in D$ and $\Gamma = \{\infty\}$ is a boundary component. Γ is unstable if there exists a sequence $\{C_n\}_{n=1}^{\infty}$ of components of $\partial D - \Gamma$ satisfying the following conditions:*

(i) *For a doubly connected region $A \subset D$ separating 0 from ∞, there exists a number n such that A separates C_n from C_{n+1}.*

(ii) *For every n, there exist points $a_n \in C_n$ and $b_n \in C_{n+1}$ such that* $|a_n - b_n| = \operatorname{dist}(C_n, C_{n+1})$,

$$\lim_{n\to\infty}\frac{b_n}{a_n}=1$$

and

$$\sum_{n=1}^{\infty}\frac{1}{\log\dfrac{1}{|(b_n/a_n)-1|}}<\infty\ .$$

31. This criterion is not a necessary condition for instability. This is apparent from the following

EXAMPLE 6. Consider the closed sets

$$E_n=\{z;\ n^2+1\leqq|z|\leqq(n+1)^2,\ |\arg z|\leqq\pi-\varepsilon_n\}\ ,$$
$$0<\varepsilon_n<\pi\ ,\qquad\qquad n=1,\ 2,\cdots\ .$$

If $\varepsilon_n(n=1,\ 2,\cdots)$ are taken sufficiently small, then $\Gamma=\{\infty\}$ is an unstable boundary component of $D=(|z|<\infty)-\bigcup_{n=1}^{\infty}E_n$. It does not satisfy the assumption of Theorem 11.

Proof. For an arbitrary subsequence $\{C_n\}_{n=1}^{\infty}$ of $\{E_n\}_{n=1}^{\infty}$ and every choice of a_n and b_n,

$$\sum_{n=1}^{\infty}\frac{1}{\log\dfrac{1}{|(b_n/a_n)-1|}}\geqq\frac{1}{2}\sum_{n=1}^{\infty}\frac{1}{\log n}=\infty\ .$$

Therefore, the assumption of Theorem 11 is not satisfied.

In order to show the instability of Γ, consider the following cross cuts of D:

$$\alpha_n:\ \Re z=0,\ (n+1)^2\leqq\Im z\leqq(n+1)^2+1\ ,$$
$$\beta_n:\ |z|=(n+1)^2,\ |\arg z|\leqq\pi-\varepsilon_n\ ,$$
$$\beta'_n:\ |z|=(n+1)^2+1,\ |\arg z|\leqq\pi-\varepsilon_{n+1}\ ,$$
$$(n=1,\ 2,\cdots)\ .$$

Let δ_n be the extremal distance between α_n and $\beta_n\cup\beta'_n$ with respect to the region $(n+1)^2<|z|<(n+1)^2+1$. It is possible to take ε_n and ε_{n+1} so small that $\delta_n>n^2$ $(n=1,\ 2,\cdots)$. Let $\{\gamma\}_n$ be the family consisting of closed curves in $D-(|z|\leqq1)$ separating Γ from $|z|\leqq1$ and passing through α_n. Let $\{\gamma_1\}_n\subset\{\gamma\}_n$ be the subfamily of curves contained in $(n+1)^2<|z|<(n+1)^2+1$ and put $\{\gamma_2\}_n=\{\gamma\}_n-\{\gamma_1\}_n$. By (VI), No. 9,

$$\frac{1}{\lambda\{\gamma\}_n} \leqq \frac{1}{\lambda\{\gamma_1\}_n} + \frac{1}{\lambda\{\gamma_2\}_n}\,.$$

Since $n^2 < \delta_n \leqq \lambda\{\gamma_2\}_n$ and $2\pi/\lambda\{\gamma_1\}_n = \log(1 + 1/(n+1)^2)$, we get

$$\frac{1}{\lambda\{\gamma\}_n} \leqq \frac{1}{2\pi}\log\left(1 + \frac{1}{(n+1)^2}\right) + \frac{1}{n^2}$$

if n is sufficiently large, and, therefore, $\sum_{n=1}^{\infty} 1/\lambda\{\gamma\}_n$ converges.

To apply Theorem 3, take $A_1, A_2, \cdots, A_k$. Then evidently

$$\sum_{\nu=1}^{k} \operatorname{mod} A_\nu \leqq \sum_{n=1}^{\infty} 1/\lambda\{\gamma\}_n < \infty$$

and we conclude that Γ is unstable.

32. Finally, for the sake of completeness, we shall present a well-known sufficient condition for weakness. For a bounded doubly connected region A, we have that $\operatorname{mod} A \geqq \log(1 + (\pi d/4l))$. Here d is the distance between the components of ∂A and l is the infimum of the lengths of closed curves which separate the components of ∂A and whose distance from ∂A is $\geqq d/2$ (Sario [15], Meschkowsky [11]). Therefore we get immediately from Theorem 3 the following result (Meschkowsky [11], Savage [19]):

THEOREM 12. *Let D be a plane region containing the point $z = 0$ and such that $\Gamma = \{\infty\}$ is a boundary component. Suppose there exists a sequence of doubly connected regions $A_n \subset D - (|z| \leqq \varepsilon)$ $(n = 1, 2, \cdots)$ with the following properties:*

(i) *The A_n are mutually disjoint,*

(ii) *A_n separates Γ from $|z| \leqq \varepsilon$ $(n = 1, 2, \cdots)$ and also separates A_{n-1} from A_{n+1} $(n = 2, 3, \cdots)$,*

(iii)

$$\sum_{n=1}^{\infty} d_n/l_n = \infty\,.$$

Then Γ is a weak boundary component of D.

On applying this theorem, we obtain

EXAMPLE 7 (Denneberg [5]). Let D be a region such that $\Gamma = \{\infty\}$ is the only accumulating boundary component. If there exist numbers $\alpha > 0$ and $\beta < \infty$ such that the distance between every pair components of $\partial D - \Gamma$ is $\geqq \alpha$ and the diameter of every component of $\partial D - \Gamma$ is $\leqq \beta$, then Γ is weak.

EXAMPLE 8 (Cf. Wagner [24]). Let $\mathfrak{G}$ be the group of transformations $z' = z + m\omega + n\omega'$ $(m,\ n = 0,\ \pm 1,\ \pm 2, \cdots)$ and let E_0 be a closed set contained in the interior of the fundamental parallelogram of $\mathfrak{G}$. Then $\Gamma = \{\infty\}$ is a weak boundary component of the region $D = (|z| < \infty) - \bigcup_{T \in \mathfrak{G}} T(E_0)$.

V. Criteria for Strength and Instability

In this chapter we shall discuss Problem B, No. 4. For simplicity we mean by a *boundary continuum* a boundary component of a region which is a continuum containing more than one point.

15. Strong boundary components.

33. If Γ is an isolated boundary continuum of D, i.e., if there exists an open set U such that $\Gamma \subset U$ and $U \cap (\partial D - \Gamma) = \phi$, then Γ is evidently strong. More generally,

THEOREM 13. *A boundary continuum Γ of a region D is strong if there exists a disk U such that $U \cap \Gamma \neq \phi$ and $U \cap (\partial D - \Gamma) = \phi$.*

This theorem is also almost trivial. To prove it rigorously, we shall use the following

LEMMA 3. *Let Δ be a simply connected region which is a proper subset of $(|\zeta| < 1)$. Map Δ conformally onto the upper half-plane. Then the image E of $\overline{\partial\Delta \cap (|\zeta| < 1)}$ is a set which does not belong to the class N_D.*[5]

The proof is easy and we omit it. It may appear plausible that E contains an interval. That this is however not so has been remarked by Koebe (see Radó [13], p. 2, Bemerkung). We can even see that the condition of Lemma 3 is necessary and sufficient.

Proof of Theorem 13. Map a component Δ of $U \cap D$ onto the upper half-plane by φ and let E be the image of $\Gamma \cap \bar{\Delta}$. By Lemma 3 $E \notin N_D$ and, therefore, E is of positive measure (Ahlfors and Beurling [2]). If Γ is unstable, a univalent function $f(z)$ transforms Γ to a point. Therefore, the univalent function $f \circ \varphi$ on the upper half-plane takes a constant boundary value on E, contrary to the well-known theorem of F. and M. Riesz.

REMARK 1. In this case, $R(\Gamma) < \infty$ and we can also use Theorem 1 to conclude that Γ is strong. To prove the finiteness of $R(\Gamma)$, we apply Theorem 5. Take a component V of $U \cap D$. It is easy to find

[5] A compact set E is said to belong to the class N_D if E^c does not admit a function with a finite Dirichlet integral.

a simply connected region Δ such that $\Delta \subset D$, $V \subset \Delta$ and $(|z| \leqq \varepsilon) \subset \Delta$. Since the set $E \notin N_D$ is of positive capacity (Ahlfors and Beurling [2]), $\lambda\{\gamma\}_\varepsilon < \infty$ by Lemma 3 and (XI), No. 10.

REMARK 2. Because of this theorem, we may consider from now on only the case where every point of Γ is an accumulation point of $\partial D - \Gamma$.

34. We shall now give two other kinds of examples of strong boundary components which do not satisfy the condition of Theorem 13.

EXAMPLE 7. Let D be a radial slit disc $|z| < a$ in the sense of No. 3 and let $\Gamma = (|z| = a)$. If the arguments of the slits form a set of measure μ less than 2π, then $R(\Gamma) < \infty$ and, consequently, Γ is strong.

In fact, we can easily obtain the estimate

$$\lambda\{\gamma\}_\varepsilon \leqq \{\log (a/\varepsilon)\}/(2\pi - \mu) < \infty .$$

35. EXAMPLE 8. Let $\{c_n\}_{n=1}^\infty$ be a sequence of numbers such that $0 < c_n \leqq \pi/2^{n+1}$. Put $r_n = 1 - 1/(n+1)$ and let

$$s_n^k = \left\{z;\ |z| = r_n,\ \frac{\pi(k-1)}{2^n} + c_n \leqq \arg z \leqq \frac{\pi k}{2^n} - c_n\right\}$$
$$(k = 1, 2, \cdots, 2^{n+1};\ n = 1, 2, \cdots) .$$

$\Gamma = (|z| = 1)$ is a boundary continuum of the circular slit disc $D = (|z| < 1) - \bigcup_{n,k} s_n^k$. If $\varliminf_{n\to\infty} c_n 2^n > 0$, then $R(\Gamma) < \infty$ and therefore, Γ is strong.

Proof. Clearly it is sufficient to give the proof for $c_n 2^n = \delta > 0$. For simplicity, we choose $\delta = \pi/4$, i.e., $c_n = \pi/2^{n+2}$. In order to show the finiteness of $R(\Gamma)$, we map D quasi-conformally onto the radial slit disc $\Delta = (|w| < 1) - \bigcup_{n,k} \sigma_n^k$, where

$$\sigma_n^k = \left\{w;\ r_n e^{-c_n/2} \leqq |w| \leqq r_n e^{c_n/2},\ \arg w = \frac{\pi(2k-1)}{2^{n+1}}\right\}$$
$$(k = 1, 2, \cdots, 2^{n+1};\ n = 1, 2, \cdots) .$$

Consider the doubly connected regions

$$A_z = \{z;\ -1 < \Re z < 1,\ -\tfrac{1}{2} < \Im z < \tfrac{1}{2}\}$$
$$- \{z;\ -\tfrac{1}{2} \leqq \Re z \leqq \tfrac{1}{2},\ \Im z = 0\}$$

and

$$A_w = \{w;\ -1 < \Re w < 1,\ -\tfrac{1}{2} < \Im w < \tfrac{1}{2}\}$$
$$- \{w;\ \Re w = 0,\ -\tfrac{1}{4} \leqq \Im w \leqq \tfrac{1}{4}\} .$$

It is not difficult to map A_z quasi-conformally onto A_w by a function which is of class C^1 in A_z and is the identity mapping on the outer periphery of A_z.

In our region D, consider the quadrilaterals

$$Q_n^k = \left\{z;\ r_n e^{-c_n} < |z| < r_n e^{c_n},\ \frac{\pi(k-1)}{2^n} < \arg z < \frac{\pi k}{2^n}\right\}$$

$$(k = 1, 2, \cdots, 2^{n+1};\ n = 1, 2, \cdots) .$$

They are mutually disjoint and all $Q_n^k - s_n^k$ and $Q_n^k - \sigma_n^k$ are conformally equivalent to A_z and A_w, respectively. Therefore, we can contruct the mapping $w = T_n^k(z)$ of $Q_n^k - s_n^k$ onto $Q_n^k - \sigma_n^k$ which is the identity mapping on ∂Q_n^k and whose maximal dilatation K depends neither on k nor on n. Then

$$w = T(z) = \begin{cases} T_n^k(z) & \text{in } Q_n^k - s_n^k \ (k = 1, 2, \cdots, 2^{n+1};\ n = 1, 2, \cdots) \\ z & \text{in } D - \bigcup_{n,k} Q_n^k \end{cases}$$

is a qussi-conformal mapping of D onto Δ such that $T(T) = (|w| = 1) = \Gamma'$.

Since Δ belongs to the case of Example 7, $R(\Gamma';\ \Delta) < \infty$, and, by Theorem 5, $\lambda\{\gamma'\}_\varepsilon < \infty$. Here γ' is a rectifiable curve in $\Delta - (|w| \leqq \varepsilon)$ connecting $|w| = \varepsilon$ with Γ'. It is furthermore assumed that γ' is a union of a countable number of analytic arcs clustering nowhere in Δ (cf. Remark, No. 15). On D, we have the corresponding family $\{\gamma\}_\varepsilon$ and, by (IV), No. 15, $\lambda\{\gamma\}_\varepsilon \leqq K\lambda\{\gamma'\}_\varepsilon < \infty$. Therefore, by Theorem 5, $R(\Gamma) < \infty$ and Γ is strong.

35. We continue to consider Example 8. If c_n decreases sufficiently fast, then $R(\Gamma) = \infty$. In fact, let $\{\gamma_n\}_\varepsilon$ be the subfamily of $\{\gamma\}_\varepsilon$ which consists of curves passing through the arc $\{z;\ z = r_n,\ |\arg z| \leqq c_n\}$. By (VI), No. 9, $\lambda\{\gamma\}_\varepsilon \geqq \lambda\{\gamma_n\}_\varepsilon/2^{n+1}$ and, By Lemma 1, No. 13,

$$\lambda\{\gamma_n\}_\varepsilon \sim \frac{1}{2\pi}\log\frac{1}{c_n} \qquad (n \to \infty) .$$

For this reason $R(\Gamma) = \infty$ if, for instance, $c_n = \exp(-2^{2n})$. However, it is unknown in this case whether Γ is strong or unstable.

16. Unstable boundary continua.

37. As in No. 21, let $\{\Omega_n\}_{n=1}^{\infty}$ be a defining sequence of Γ and let $0 \in G_n = D - \Omega_n \uparrow D$. Consider the function $w = f_n(z)$ maximizing the functional $m(f)$ in $\mathfrak{F}_{\Gamma_n}$ on G_n (No. 19). We may assume that $\{f_n(z)\}_{n=1}^{\infty}$ converges to a univalent function $w = f(z)$.

In the following case, $R(\Gamma) = \infty$ implies that $f(\Gamma) = \{\infty\}$:

THEOREM 14. *Let D be a region containing $z=0$ and let Γ be a boundary continuum. Suppose that*

(i) *D is symmetric with respect to the lines*

$$l_\nu:\ re^{\nu\pi/2k}\ (-\infty<r<\infty),\ \nu=1,\ 2,\cdots,\ 2^k$$

for some integer $k\geqq 0$, and

(ii) *every component of $\partial D-\Gamma$ intersects at least one l_ν.*

Then Γ is strong if and only if $R(\Gamma)<\infty$.

Proof. We may assume that each G_n is symmetric with respect to all the l_ν. By the uniqueness of $f_n(z)$ (No. 19), we can immediately see that $f_n(z)$ and, *a fortiori*, $f(D)$ are symmetric about these lines. As has been shown by Strebel [22], $f(\partial D-\Gamma)$ consists of radial segments. By the assumption $f(\partial D-\Gamma)$ is contained in $\bigcup_{\nu=1}^{2k} l_\nu$.

Now assume that $f(\Gamma)\neq\{\infty\}$. If $f(\Gamma)\subset\bigcup_{\nu=1}^{2k} l_\nu\cup\{\infty\}$, then $f(\Gamma)\cap l_\nu$ is a line segment which does not meet $\overline{f(\partial D-\Gamma)}$, so that $R(\Gamma)<\infty$ by Remark 1, No. 33. If $f(\Gamma)\not\subset\bigcup_{\nu=1}^{2k} l_\nu\cup\{\infty\}$ there exists a sector S bounded by two neighboring l_ν's such that $S\cap f(\Gamma)$ does not intersect $f(\partial D-\Gamma)$ and we have $R(\Gamma)<\infty$. Consequently, the strength of Γ implies that $R(\Gamma)<\infty$.

38. We can find many examples of unstable boundary continua belonging to this category, e.g., as follows:

EXAMPLE 9. Consider the region

$$D=(|z|\leqq\infty)-\Gamma-\bigcup_{k=1}^{\infty}(s_k^+\cup s_k^-\cup\sigma_k^+\cup\sigma_k^-)\ ,$$

where

$$\Gamma=\{z;\ -1\leqq\Re z\leqq 1,\ \Im z=0\}\ ,$$

$$s_k^+=\left\{z;\ 1+\frac{1}{2k+1}\leqq\Re z\leqq 1+\frac{1}{2k},\ \Im z=0\right\}\ ,$$

$$s_k^-=\left\{z;\ -1-\frac{1}{2k}\leqq\Re z\leqq -1-\frac{1}{2k+1},\ \Im z=0\right\}\ ,$$

$$\sigma_k^\pm=\left\{z;\ -1\leqq\Re z\leqq 1,\ \Im z=\frac{\pm 1}{k}\right\}\ .$$

Since every point on Γ, except ± 1, is inaccessible, $R(\Gamma)=\infty$ by (X'), No. 10. From this and from Theorem 14, we infer that Γ is an unstable boundary continuum of D.

39. Meschkowsky [11] has proved that a region satisfying certain metric conditions can be mapped conformally onto a region bounded by circles or points in such a way that the image of a preassigned boundary continuum is a point. This case is also an example of an unstable boundary continuum.

40. The following example belongs to this category but does not necessarily satisfy Meschkowsky's conditions. Moreover, the function $f(z) = \lim_{n\to\infty} f_n(z)$ of No. 37 does not transform Γ to a point.

EXAMPLE 10. Let $I = \{z;\ -1 \leqq \Re z \leqq 1,\ \Im z = 0\}$ and let

$$I' = \{z;\ \Re z = 0,\ -1 \leqq \Im z \leqq 1\}\ .$$

Choose a sequence $\{c_k;\ k = \pm 1,\ \pm 2, \cdots\}$ such that

$$c_{-k} = -c_k,\ c_1 > c_2 > \cdots \downarrow 0\ ,$$

and let

$$s_k^0 : z = re^{ic_k} \qquad (1/|k|! \leqq r \leqq 1)\ ,$$

$$s_k^{\pi/2} : z = re^{i(c_k+\pi/2)} \qquad (1/|k|! \leqq r \leqq 1)\ ,$$

$$s_k^{\pi} : z = re^{i(c_k+\pi)} \qquad (1/|k|! \leqq r \leqq 1)\ ,$$

$$s_k^{-\pi/2} : z = re^{i(c_k-\pi/2)} \qquad (1/|k|! \leqq r \leqq 1)\ ,$$

where $k = \pm 1,\ \pm 2, \cdots$. Then $\Gamma = I \cup I'$ is an unstable boundary continuum of the region

$$D = (|z| \leqq \infty) - \Gamma - \bigcup_{\substack{k=-\infty \\ k\neq 0}}^{\infty} (s_k^0 \cup s_k^{\pi/2} \cup s_k^{\pi} \cup s_k^{-\pi/2})\ .$$

In fact, D can be mapped onto a region such that $f(\Gamma)$ is a point and every component of $f(\partial D - \Gamma)$ is a circle. For the proof, map the region

$$(|z|) \leqq \infty) - \bigcup_{\substack{k=-n \\ k\neq 0}}^{\infty} (s_k^0 \cup s_k^{\pi/2} \cup s_k^{\pi} \cup s_k^{-\pi/2})$$

conformally onto a region bounded by $8n$ circles; we may require that the mapping function $w = f^{(n)}(z)$ has the expansion $z + b_n/z + \cdots$ near $z = \infty$ $(n = 1, 2, \cdots)$. The existence and the uniquess of such a mapping are well known. A suitable subsequence of $\{f^{(n)}(z)\}_{n=1}^{\infty}$ converges to a univalent function $w = f(z)$. We can easily prove that every component of $f(\partial D - \Gamma)$ is a circle (see, e.g., Meschkowsky [11]). In what follows we shall show that $f(\Gamma) = \{0\}$.

First we remark that $R(\Gamma) = \infty$, because every point on Γ, except $0,\ \pm 1,\ \pm i$, is inaccessible (cf. (X'), No. 10). Second, D and, therefore,

$f(D)$ are symmetric with respect to the following four lines: $l_0 =$ (real axis), $l_{\pi/4} = (\Re z = \Im z)$, $l_{\pi/2} =$ (imaginary axis), and $l_{-\pi/4} = (\Re z = -\Im z)$.

The component $f(\Gamma)^*$ of $f(D)^c$ corresponding to $f(\Gamma)$ is a compact connected set which contains the point $w = 0$ and is symmetric about these four lines.

The component $f(s_k^\beta)^*$ of D^c ($\beta = 0, \pm\pi/2, \pi$; $k = \pm 1, \pm 2, \cdots$) is a disk, which we denote by

$$\Delta_k^\beta : \ | w - a_k^\beta | \leqq \rho_k \ .$$

The radius ρ_k does not depend on β because of the symmetry. Furthermore,

$$\lim_{k\to\infty} \rho_k = 0 ; \tag{17}$$

in fact, all the Δ_k^β cluster to $f(\Gamma)^*$, so that the sum $8\pi \sum_{k=1}^\infty \rho_k^2$ of their areas converges.

Consider a quadrilateral

$$Q_k = \left\{ z ; \ \frac{1}{k!} < | z | < \frac{1}{(k-1)!}, \ c_k < \arg z < \frac{\pi}{2} - c_k \right\} ,$$

which connects s_k^0 with $s_{-k}^{\pi/2}$ ($k = 1, 2, \cdots$). The extremal distance between s_k^0 and $s_{-k}^{\pi/2}$ with respect to D does not exceed

$$\operatorname{mod} Q_k = \frac{(\pi/2) - 2c_k}{\log k} \ .$$

Let L_k be the infimum of lengths of curves in $f(D)$ connecting Δ_k^0 with $\Delta_{-k}^{\pi/2}$. Then

$$\frac{L_k^2}{\mu U} \leqq \frac{(\pi/2) - 2c_k}{\log k} \longrightarrow 0 \qquad (k \to \infty) \tag{18}$$

where μU expresses the area of a bounded open set U containing $f(\Gamma)^*$. For this reason and by virtue of (17) and (18), we have

$$\lim_{k\to\infty} | a_k^0 - a_{-k}^{\pi/2} | \leqq \lim_{k\to\infty} (L_k + 2\rho_k) = 0 \ .$$

It follows, by symmetry, that $\{a_k^0\}_{k=1}^\infty$ and $\{a_{-k}^{\pi/2}\}_{k=1}^\infty$ cluster to $l_{\pi/4}$ in the first quadrant. From this and again from the symmetry, we see that the set H of all accumulation points of a_k^β ($\beta = 0, \pm\pi/2, \pi$; $k = \pm 1, \pm 2, \cdots$) is contained in $l_{\pi/4} \cup l_{-\pi/4}$. Evidently it is symmetric about l_0 and $l_{\pi/2}$, and $H \subset f(\Gamma)^*$.

Next we shall show that $H = \{0\}$. Suppose that H contains a point $w_0 = pe^{i\pi/4}$ ($p > 0$). Then there must exist a point $qe^{i\pi/4} \in H$ ($0 \leqq q < p$). For otherwise H would consist of four points: $H = \{pe^{i\theta} ; \theta = \pm\pi/4, \pm 3\pi/4\}$.

Then all but a finite number of components of $f(\partial D - \Gamma)$ in the first quadrant would be contained in $|w - pe^{i\pi/4}| < p/4$. Since w_0 and 0 are contained in $f(\Gamma)^*$ and $f(\Gamma)^*$ is a continuum, $f(\Gamma)$ would have a "free" subset as in Theorem 13. But the reasoning of Remark 1, No. 33, shows that this property of $f(\Gamma)$ contradicts the fact that $R(\Gamma) = \infty$ and, therefore, $qe^{i\pi/4} \in H$ exists. Take a subsequence $\{k_j\} \subset \{k\}$ such that

$$\lim_{j\to\infty} a^0_{k_j} = \lim_{j\to\infty} a^{\pi/2}_{-k_j} = qe^{i\pi/4}\,.$$

Then

$$L_{k_j} + 2\rho_{k_j} \geqq \frac{p-q}{2} > 0$$

for sufficiently great j, contrary to (17) and (18). Consequently, w_0 does not exist and $H = \{0\}$.

Finally, if $f(\Gamma)^* \supsetneqq H$, then $f(\Gamma)$ would again have a "free" subset, contrary to the fact that $R(\Gamma) = \infty$. We conclude that $f(\Gamma)^* = \{0\}$.

41. Transform the region D by $\zeta = 1/z$ and, for simplicity, denote the image again by D. For the sequence $G_n \uparrow D$ of No. 37, we take

$$\begin{aligned} G_n = (|z| < n! + c_{n+1}) \cap D \\ - \bigcup_{h=1}^{3} \Big\{z;\ 1 - c_{n+1} \leqq |z|,\ \frac{h\pi}{2} - \frac{c_n + c_{n+1}}{2} \leqq \arg z \\ \leqq \frac{h\pi}{2} + \frac{c_n + c_{n+1}}{2}\Big\}\,, \end{aligned}$$

$n = 1, 2, \cdots$, and consider the extremal function $f_n(z)$. We shall show:

If $c_k = -c_{-k}$ decreases sufficiently fast (e. g., $c_k = e^{-k!}$), *then $\lim_{n\to\infty} f_n(z) = z$ uniformly on every compact set in D.*

In order to prove this, we estimate the Dirichlet integral of $\log|f_n(z)/z|$ over $\varDelta = (|z| \leqq 1/2)$:

$$\begin{aligned} &D_\varDelta(\log|f_n(z)| - \log|z|) \leqq D_{G_n}(\log|f_n(z)| - \log|z|) \\ &= \int_{\partial G_n} (\log|f_n| \cdot d\arg f_n - \log|z| \cdot d\arg f_n \\ &\quad - \log|f_n| \cdot d\arg z + \log|z| \cdot d\arg z) \\ &= \int_{\partial G_n} (\log|f_n| \cdot d\arg f_n - 2\log|f_n| \cdot d\arg z \\ &\quad + \log|z| \cdot d\arg z) \\ &= 2\pi \log R(\Gamma_n;\ G_n) - 2\log R(\Gamma_n;\ G_n)\int_{\Gamma_n} d\arg z \\ &\quad + \int_{\Gamma_n} \log|z|\, d\arg z \leqq 2\pi\{\log n! - \log R(\Gamma_n;\ G_n)\}\,. \end{aligned}$$

To estimate the last term, we shall use the relation $\log R(\Gamma_n; G_n) = \lim_{\varepsilon\to 0}(\log\varepsilon + 2\pi\lambda\{\gamma\}_\varepsilon^{(n)})$, where the sequence is increasing (No. 22). Here $\{\gamma\}_\varepsilon^{(n)}$ is the family of curves in $G_n - (|z| \leqq \varepsilon)$ connecting Γ_n with $|z| = \varepsilon$. We take the closed disks

$$\Delta_n^h : |z - e^{i\pi h/2}| \leqq c_n ,$$

$$\Delta_n'^h : |z - n!\, e^{i\pi h/2}| \leqq n!\, c_n ,$$

$h = 0, 1, 2, 3$; $n = 1, 2, \cdots$. Let $\{\gamma_1\}_\varepsilon^{(n)} \subset \{\gamma\}_\varepsilon^{(n)}$ be the family of curves connecting $|z| = \varepsilon$ with $\bigcup_{h,n} \Delta_n^h \cup \Delta_n'^h$ and put $\{\gamma_2\}_\varepsilon^{(n)} = \{\gamma\}_\varepsilon^{(n)} - \{\gamma_1\}_\varepsilon^{(n)}$. By (VI), No. 9,

$$\frac{1}{\lambda\{\gamma\}_\varepsilon^{(n)}} \leqq \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \qquad (\lambda_\nu = \lambda\{\gamma_\nu\}_\varepsilon^{(n)},\ \nu = 1, 2) ,$$

or

$$\lambda\{\gamma\}_\varepsilon^{(n)} \geqq \lambda_2 - \frac{\lambda_2^2}{\lambda_1} .$$

It is evident that

$$\frac{1}{2\pi - 8c_n}\log\frac{n! + c_n}{\varepsilon} \geqq \lambda_2 \geqq \frac{1}{2\pi}\log\frac{n!}{\varepsilon} .$$

Therefore,

$$\log R(\Gamma_n;\ G_n) \geqq \log\varepsilon + 2\pi\lambda\{\gamma\}_\varepsilon^{(n)} \geqq \log n! - 2\frac{\lambda_2^2}{\lambda_1} ,$$

whence

$$D_\Delta(\log|f_n(z)| - \log|z|) \leqq 4\pi^2\frac{\lambda_2^2}{\lambda_1} .$$

If c_n is taken sufficiently small, then $\lim_{n\to\infty}\lambda_2^2/\lambda_1 = 0$. For instance, if $c_n = e^{-n!}$, we have $\lambda_1 \sim (8\cdot n!)/\pi$ $(n \to \infty)$ by Lemma 1, No. 13, and $\lambda_2^2/\lambda_1 \to 0$. In such a case, $\lim_{n\to\infty} D_\Delta(\log|f_n(z)| - \log|z|) = 0$ and we conclude that $\lim_{n\to\infty} f_n(z) = z$ uniformly on each compact set in D.

Consequently $R(\Gamma) = \infty$ for our region, but $\lim_{n\to\infty} f_n(z)$ does not transform Γ to a point.

References

1. L. V. Ahlfors, *On quasiconformal mappings*, J. Analyse Math. **3** (1953/54), 1–58, 207–208.
2. L. V. Ahlfors, and A. Beurling, *Conformal invariants and function-theoretic null-sets*, Acta Math. **83** (1950), 101–129.
3. L. V. Ahlfors, and L. Sario, *Riemann surfaces*, Princeton University Press, to appear.
4. C. Constantinescu, *Sur la comportement d'une fonction analytique à la frontière d'une surface de Riemann*, C. R. Acad. Sci. Paris, **245** (1957), 1995–1997.

5. H. Denneberg, *Konforme Abbildung einer Klasse unendlich-vielfach zusammenhängender schlichter Bereiche auf Kreisbereiche*, Ber. Verh. Sächs. Akad. Wiss. Leipzig. Math-Nat. Kl. **84** (1932), 331-352.

6. H. Grötzsch, *Über die Verzerrung bei schlichter konformen Abbildung mehrfach zusammenhängender schlichter Bedeiche*, I. Ibid. **81** (1929), 38-47. II. 217-221.

7. ———, *Eine Bemerkung zum Koebeschen Kreisnormierungsprinzip*, Ibid. **87** (1935), 319-324.

8. J. Hersch, *Longueurs extrémales et théorie des fonctions*, Comment. Math. Helv. **29** (1955), 301-337.

9. ———, *Contribution à la théorie des fonctions pseudo-analytiques*, Ibid. **30** (1956), 1-19.

10. M. Jurchescu, *Modulus of a boundary component*, Pacific J. Math. **8** (1958), 791-809.

11. H. Meschkowsky, *Über die konforme Abbildung gewisser Bereiche von unendlich hohem Zusammenhang auf Vollkreisbereiche*, I. Math. Ann. **123** (1951), 392-405. II. Ibid. **124** (1952), 178-181.

12. A. Pfluger, *Extremallängen und Kapazität*, Comment. Math. Helv. **29** (1955), 120-131.

13. T. Radó, *Über eine nicht fortsetzbare Riemannsche Mannigfaltigkeit*, Math. Z. **20** (1924), 1-6.

14. E. Rengel, *Existenzbeweise für schlichte Abbildungen mehrfach zusammenhängender Bereiche auf gewisse Normalbereiche*, Jber. Deutsch Math. Verein. **45** (1935), 83-87.

15. L. Sario, *Über Riemannsche Flächen mit hebbarem Rånd*, Ann. Acad. Sci. Fenn. Ser. A. I. **50** (1948), 79 pp.

16. ———, *Capacity of the boundary and of a boundary component*, Ann. Math. **59** (1954), 135-144.

17. ———, *Stability problems on boundary components*, Proc. Conference Anal. Func., Princeton, 1957. To appear.

18. ———, *Strong and weak boundary components*, J. Analyse Math. **5** (1958), 389-398.

19. N. Savage, *Weak boundary components of an open Riemann surface*, Duke Math. J. **24** (1957), 79-96.

20. K. Strebel, *Eine Ungleichung für extremale Längen*, Ann. Acad. Sci. Fenn. Ser. A. I. **90** (1951), 8 pp.

21. ———, *Über das Kreisnormierungsproblem der konformen Abbildung*, Ibid. **101** (1951), 22 pp.

22. ———, *Die extremale Distanz zweier Enden einer Riemannschen Fläche*, Ibid. **179** (1955), 22 pp.

23. O. Teichmüller, *Untersuchungen über konforme und quasikonforme Abbildung*, Deutsche Math. **3** (1938), 621-678.

24. R. Wagner, *Ein Kontaktproblem der konformen Abbildung*, J. Reine Angew. Math. **196** (1956), 99-132.

25. V. Wolontis, *Properties of conformal invariants*, Amer. J. Math. **74** (1952), 587-606.

UNIVERSITY OF CALIFORNIA, LOS ANGELES, CALIFORNIA
AND
TOKYO INSTITUTE OF TECHNOLOGY, JAPAN

LOOPS WITH THE WEAK INVERSE PROPERTY

J. MARSHALL OSBORN

Let the left and right inverses of an element x of a loop G be denoted by x^λ and x^ρ respectively, then G is said to have the inverse property if the two identities $x^\lambda(xy) = y$ and $(yx)x^\rho = y$ are satisfied by all elements x, y of G. Perhaps the two most basic properties of inverse property loops are that (i) the left, middle and right nuclei coincide, and that (ii) if every loop isotopic to G has the inverse property, then G is a Moufang loop[1]. More recently, R. Artzy has defined cross inverse property loops (G has the cross inverse property if any two elements x and y of G satisfy either of the two equivalent identities $x^\lambda(yx)=y$ and $(xy)x^\rho = y$), and has shown that the same two properties hold for these loops[2]. In the present paper, we shall consider (i) and (ii) for a class of loops which includes both of the classes already mentioned. In § 1 we introduce the weak inverse property and prove (i) for loops with this property. In § 2 and § 3 we discuss loops all of whose isotopes have the weak inverse property, and show that those loops are not necessarily Moufang loops but come very close (see Theorems 2 and 3). An interesting by-product of this investigation is the construction in § 3 of a class of loops, each of which is isomorphic to all its isotopes. The only previously known examples of such loops have been Moufang loops[3].

In dealing with isotopy and cross inverse property loops, Artzy does not discuss the question of whether a cross inverse property loop can arise as an isotope of an inverse property loop. In § 4 we answer this question in the negative.

1. Let G be a loop with identity element 1, then G will be said to satisfy the *weak inverse property*[4] if whenever three elements x, y, z of G satisfy the relation $xy \cdot z = 1$, they also satisfy the relation $x \cdot yz = 1$. Using the right inverse operator ρ, we may transform this definition into more usable form by observing that the relation $xy \cdot z = 1$ is equivalent to $z = (xy)^\rho$, and by substituting this into $x \cdot yz = 1$ to yield

$$y \cdot (xy)^\rho = x^\rho \,. \tag{1}$$

Received April 6, 1959.

1. These concepts will all be defined in the body of the paper. A proof of these properties and those used in § 4 for inverse property loops and Moufang loops will be found in [**3**]. Note that Bruck uses the term "associator" instead of "nucleus" in his earlier papers.

2. See [**1**].

3. See [**5**].

4. Loops with this property have previously been considered in connection with nets. For a brief discussion of this and references, see [**2**].

Iterating this relation, we readily obtain

$$(xy)^{\rho}\cdot x^{\rho^2} = y^{\rho}\ , \tag{2}$$

and iterating again gives $x^{\rho^2}\cdot y^{\rho^2} = (xy)^{\rho^2}$, showing that ρ^2 is an automorphism of G. Since λ is the inverse of the operator ρ, λ^2 is also an automorphism, and applying it to (2) gives

$$(xy)^{\lambda}\cdot x = y^{\lambda}\ , \tag{3}$$

which is the dual of (1). From (3) it is easy to see that $x\cdot yz = 1$ implies $xy\cdot z = 1$, so that we could have equivalently defined G to have the weak inverse property if $xy\cdot z = 1$ whenever $x\cdot yz = 1$. It might also be remarked that if ρ is an anti-automorphism or automophism in a weak inverse property loop G, then equations (1) and (3) tell us that G has the inverse property or the cross inverse property respectively. Conversely, either of the latter two properties imply the weak inverse property.

Letting $R(y)$ and $L(y)$ denote right and left multiplication by the element y, we may rewrite (1) in the form $R(y)\rho L(y) = \rho$, which yields the two useful relations

$$R^{-1}(y) = \rho L(y)\lambda\ , \text{ and } L^{-1}(y) = \lambda R(y)\rho\ . \tag{4}$$

To develop the properties of weak inverse property loops further, we shall need to introduce the concepts of isotopism and autotopism. Let G_0 be a loop consisting of the elements of G under a new binary operation "$\circ$" (the old operation shall be denoted by "$\cdot$"), and let U, V, W be three permutations on the elements of G satisfying the relation $xU\cdot yV = (x\circ y)W$ for all x, y of G. Then we shall say that G_0 is *isotopic* to G (or, equivalently, that it is an *isotope* of G) by means of the *isotopism* (U, V, W). In case "$\circ$" is just the original binary operation "$\cdot$", we shall call (U, V, W) an *autotopism*. Observe that if T is an automorphism of G, then it gives rise to the autotopism (T, T, T), and conversely. It is well known[5] that the set of isotopisms of G form a group under the operation $(U_1, V_1, W_1)(U_2, V_2, W_2) = (U_1 U_2, V_1 V_2, W_1 W_2)$, and that the autotopisms form a subgroup.

LEMMA 1. *If (U, V, W) is an autotopism of a weak inverse property loop, then so are $(V, \lambda W\rho, \lambda U\rho)$ and $(\rho W\lambda, U, \rho V\lambda)$.*

Using (1) on the relation $xU\cdot yV = (xy)W$, we obtain $yV\cdot[(xy)W]^{\rho} = [xU]^{\rho}$. And making the substitution $x = (yz)^{\lambda}$ in this equation yields $yV\cdot[(z^{\lambda})W]_{\rho} = [(yz)^{\lambda}U]^{\rho}$, which tells us that $(V, \lambda W\rho, \lambda U\rho)$ is an autotopism. The other autotopism of the lemma arises in the same way using (3).

Next, we define the *left nucleus* of G to be the set of all elements

5. See [**3**], for example.

a of G satisfying the relation $ax \cdot y = a \cdot xy$ for every pair of elements x, y of G. We may equivalently characterize the left nucleus as the set of all a such that $(L(a), 1, L(a))$ is an autotopism of G. Similarly the dual concept of *right nucleus* may be characterized as the set of all a such that $(1, R(a), R(a))$ is an autotopism. If we now assume that a is in the right nucleus, then Lemma 1 tells us that $(R(a), \lambda R(a)\rho, 1)$ and $(\rho R(a)\lambda, 1, \rho R(a)\lambda)$ are autotopisms. From the latter it is clear that $(\lambda^2, \lambda^2, \lambda^2)(\rho R(a)\lambda, 1, \rho R(a)\lambda)^{-1}(\rho^2, \rho^2, \rho^2) = (L(a), 1, L(a))$ is an autotopism, so that a is in the left nucleus of G. On the other hand, from the former we get the equation $xa \cdot [(z^\lambda)a]^\rho = xz$, or $xa \cdot zL_a^{-1} = xz$ for all x, z of G. Setting $z = ay$ gives $xa \cdot y = x \cdot ay$ for all x, y of G, which is the definition of the element a being in the *middle nucleus*. Since all these steps are reversible, we have proved:

THEOREM 1. *The left, middle, and right nuclei of a weak inverse property loop coincide.*

2. We now turn to the question of when an isotope of G also has the weak inverse property. First of all, if G_0 is an isotope of an arbitrary loop G, then it is well known[5] that, up to isomorphism, "$\circ$" is given in terms of "$\cdot$" by the relation $x \circ y = xR^{-1}(g) \cdot yL^{-1}(f)$, for some fixed pair of elements f and g of G. If ρ_0 is the right inverse operator of G_0, then the weak inverse property in G_0 is equivalent to the identity $y \circ (x \circ y)^{\rho_0} = x^{\rho_0}$, or $yR^{-1}(g) \cdot [xR^{-1}(g) \cdot yL^{-1}(f)]^{\rho_0} L^{-1}(f) = x^{\rho_0}$. Setting $x = ug$ and $y = fv$, this becomes $(fv)R^{-1}(g) \cdot (uv)^{\rho_0} L^{-1}(f) = (ug)^{\rho_0}$, and using the weak inverse property in G yields

$$(5) \qquad (ug)^{\rho_0 \lambda} \cdot (fv)R^{-1}(g) = [(uv)^{\rho_0} L^{-1}(f)]^\lambda .$$

Since fg is the unit of G_0 (as may be verified from its definition), the mapping ρ_0 is defined by the relation $fg = x \circ x^{\rho_0} = xR^{-1}(g) \cdot (x^{\rho_0})L^{-1}(f)$. Using (3), we may put this in the form $[(x^{\rho_0})L^{-1}(f)]^\lambda = (fg)^\lambda \cdot xR^{-1}(g)$, which leads to the formula $\rho_0 = R^{-1}(g)L((fg)^\lambda)\rho L(f)$. But (5) says precisely that $(R(g)\rho_0\lambda, L(f)R^{-1}(g), \rho_0 L^{-1}(f)\lambda)$ is an autotopism, which may now be rewritten as

$$(6) \qquad (L([fg]^\lambda)R^{-1}(f), L(f)R^{-1}(g), R^{-1}(g)L([fg]^\lambda))$$

after substituting for ρ_0 and using the first relation of (4).

Next, if (U, V, W) is an autotopism of G, and if f and g are the images of the identity under U and V respectively, then we may obtain the relations $UR(g) = W$, $VL(f) = W$ and $fg = 1W$ as special cases of the relation $xU \cdot yV = (xy)W$. Our autotopism may then be written in the form $(WR^{-1}(g), WL^{-1}(f), W)$. But then the isotope G_0 given by $x \circ y = xR^{-1}(g) \cdot yL^{-1}(f) = [xW^{-1} \cdot yW^{-1}]W$ is isomorphic to G, and hence has the weak inverse property. Conversely, if G_0 is isomorphic to G by the mapping W^{-1}, then $(WR^{-1}(g), WL^{-1}(f), W)$ will be an autotopism. We have shown:

LEMMA 2: *If f and g are two elements of a weak inverse property loop G, then the isotope G_0 given by $x \circ y = xR^{-1}(g) \cdot yL^{-1}(f)$ has the weak inverse property if and only if the expression in* (6) *is an autotopism. Furthermore, G_0 is isomorphic to G if and only if f and g are the images of the identity under the first two permutations of some autotopism.*

Consider now the special case of Lemma 2 with $g = 1$. The autotopism (6) is then $(L(f^\lambda)R^{-1}(f), L(f), L(f^\lambda))$, which may be transformed into $(L(f), \lambda L(f^\lambda)\rho, \lambda L(f^\lambda)R^{-1}(f)\rho) = (L(f), R^{-1}(f^\rho)R^{-1}(f^\rho)L(f))$ using Lemma 1. Applying this autotopism to the pair x, 1 gives the relation $L(f)R(f) = R^{-1}(f^\rho)L(f)$, which allows us to write our autotopism in the form

$$(7) \qquad (L(f), R^{-1}(f^\rho), L(f)R(f)) .$$

We shall also need to use this in the following equivalent form:

$$(8) \qquad L(fx) = R(f^\rho)L(x)L(f)R(f) = R(f^\rho)L(x)R^{-1}(f^\rho)L(f) .$$

From (7) and (8), it is clear that f is a weakened type of Moufang element of G[6]. Similar to the case of inverse and cross inverse property loops, one may say something about the structure of the set of elements which give isotopes with the weak inverse property, or which are the images of the identity under a permutation from some autotopism. However, since neither of these sets need form a subloop, this structure does not seem sufficiently interesting to be discussed except in the case where all isotopes have the weak inverse property, to which we turn next.

If all isotopes of G have the weak inverse property, then we may use (6), (7) and (8) for any elements f, g, x of G. In particular, if we take the inverse of (6) and set $f = 1$, we get $(L^{-1}(g^\lambda), R(g), L(g^\lambda)R(g))$. Applying this to $1, x$ gives $R(g)L(g) = L(g^\lambda)R(g)$, allowing us to write $(L^{-1}(g^\lambda), R(g), R(g)L(g))$, which is the dual of (7). Replacing f by g in (7) and multiplying by the inverse of its dual, we get $(L(g)L(g^\lambda), R^{-1}(g^\rho)R^{-1}(g), L(g)R(g)L^{-1}(g)R^{-1}(g))$. But each of these three permutations preserves the identity element of G, and hence, by an easy argument, they are all equal. Defining θ_g by

$$(9) \qquad \theta_g = L(g)L(g^\lambda) = R^{-1}(g^\rho)R^{-1}(g) = L(g)R(g)L^{-1}(g)R^{-1}(g) ,$$

we have shown that θ_g is an automorphism. From (9) we get the relation $R^{-1}(g^\rho) = \theta_g R(g)$, which allows us to put (7) in the form $(L(g), \theta_g R(g), L(g)R(g))$, which says that

6. In case the isotope G_0 defined by this element f is isomorphic to G, then f is a "companion" in the terminology of [4].

(10) $$gx\cdot(z\theta_g\cdot g) = (g\cdot xz)\cdot g\,, \quad \text{for all} \quad g, x, z \quad \text{of} \quad G\,.$$

But if θ_g were the identity automorphism for all g, then (10) would be just one of the Moufang identities. For example, if G has the inverse property, then $L(g^\lambda) = L^{-1}(g)$, and θ_g is the identity. Similarly, if G has the cross inverse property, then $L(g^\lambda)=R^{-1}(g)$, and (9) yields $L(g)R^{-1}(g)= L(g)R(g)L^{-1}(g)R^{-1}(g)$, or $L(g) = R(g)$. Hence G is commutative and θ_g is again the identity. We have proved:

THEOREM 2. *If G is an inverse property, cross inverse property or commutative loop such that every isotope of G has the weak inverse property, then G is a Moufang loop.*

Now let α be the autotopism (6), and let β and γ be the special cases of this with $g = 1$ and $f = 1$ respectively, then $\beta\gamma\alpha^{-1}=(L(f^\lambda)R^{-1}(f), L(f), L(f^\lambda))\cdot(L(g^\lambda), R^{-1}(g), R^{-1}(g)L(g^\lambda))\cdot(R(f)L^{-1}([fg]^\lambda), R(g)L^{-1}(f), L^{-1}([fg]^\lambda)R(g)) = (L(f^\lambda)R^{-1}(f)L(g^\lambda)R(f)L^{-1}([fg]^\lambda), 1, L(f^\lambda)R^{-\lambda}(g)L(g^\lambda) L^{-1}([fg]^\lambda)R(g))$. Applying this autotopism to the pair x, g gives $L(f^\lambda)R^{-1}(f)L(g^\lambda)R(f) = R(g)L(f^\lambda)R^{-1}(g)L(g^\lambda)$. But from (8) this is just $L(g^\lambda f^\lambda)$, so that the first permutation of $\beta\gamma\alpha^{-1}$ is $L(g^\lambda f^\lambda)L^{-1}([fg]^\lambda)$. Denoting this permutation by U, we thus have an autotopism of the form $(U, 1, W)$, or $xU\cdot y = (xy)W$ for all x, y of G. But setting $y=1$ in this equation gives $U = W$, and setting $x = 1$ shows that $U = L(u)$ where u is the image of the identity under U. Hence $u=(g^\lambda f^\lambda)L^{-1}([fg]^\lambda)$ is in the nucleus of G. Furthermore, in case $u = 1$ for all f and g of G, then we may conclude that $[fg]^\lambda = g^\lambda f^\lambda$ for all f, g of G, so that G has the inverse property. Since the nucleus of G is normal by a theorem of Bruck[7], we have proved:

THEOREM 3: *Let G be a loop all of whose isotopes have the weak inverse property, and let N be its nucleus. Then N is normal, and G/N is a Moufang loop.*

We turn next to a closer examination of the automorphism θ_g. First of all, if x is an arbitrary element of G and if b is an element of the nucleus, then $x\theta_b = b^\lambda\cdot bx = b^\lambda b\cdot x = x$, $b\theta_x=x^\lambda\cdot xb = x^\lambda x\cdot b = b$, and $(bx)^\lambda= [(bx)^\lambda\cdot b]\cdot b^{-1} = x^\lambda b^{-1}$. Also, $\theta_{bx} = \theta_x$, since $y\theta_{bx} = (bx)^\lambda\cdot(bx\cdot y) = x^\lambda b^{-1}\cdot (b\cdot xy) = x^\lambda\cdot xy = y\theta_x$ for any element y of G. Again, setting $x = g^\rho$ in (10) gives the relation $\theta_g = L(g^\rho)L(g)$, or $\theta_g = \theta_h$ for $h = g^\rho$ using (9). Using the iterates of this relation, we compute $x^{\lambda^i}\theta_x=x^{\lambda^i}L(x^{\lambda^{i+1}})L(x^{\lambda^{i+2}}) = x^{\lambda^{i+2}}$. As a special case of this we have $x\theta_x = x^\lambda\cdot xx = x^{\lambda^2}$, which may

7. See Theorem 4.1 of [**4**]. Although the statement of Bruck's theorem assumes that G is a Moufang loop, it is clear from the proof (and his Lemma 2.1) that he only uses the fact that every permutation of the form $L(x)$ or $R(x)$ occurs as the first permutation in some autotopism of G. This is true in our situation from (7), and from the special case $g = f_\rho$ of (6).

be transformed using the weak inverse property into $x^{\lambda^3}\cdot x^{\lambda} = (xx)^{\lambda}$, or $x^{\lambda}\cdot x^{\rho} = (xx)^{\rho}$. Similarly, $x\theta_x^{-1} = xx\cdot x^{\rho} = x^{\rho^2}$ leads to $x^{\lambda}\cdot x^{\rho} = (xx)^{\lambda}$. Hence $(xx)^{\lambda} = (xx)^{\rho}$, and so squares have unique inverses in G.

We now define a by the equation $xa = x\theta_x$, and observe that a will be in the nucleus. Since $1 = x^{\lambda^2}\cdot x^{\lambda} = xa\cdot x^{\lambda} = x\cdot ax^{\lambda}$, or $ax^{\lambda} = x^{\rho}$, we have $x^{\lambda}x^{\lambda} = (x^{\lambda}x^{\lambda})^{\rho^2} = x^{\rho}x^{\rho} = x^{\rho}\cdot ax^{\lambda} = x^{\rho}a\cdot x^{\lambda}$, and hence $x^{\rho}a = x^{\lambda}$. But then $x^{\rho}a = x^{\rho}\theta_x$, and we can conclude from the properties developed in the last paragraph that $x^{\lambda^i}\cdot a = x^{\lambda^{i+2}}$ and $a\cdot x^{\lambda^i} = x^{\lambda^{i-2}}$. If K is the subloop of G generated by x, and A the cyclic subgroup generated by a, then A is contained in the nucleus of K and $xA = Ax$ from the relations just derived. But then A is normal, K/A is cyclic, and every element of K can be expressed in the form $x^i a^j$ for some pair of integers i, j (note that x^i may be defined to be any element of K that maps into the ith power of the image of x in K/A). It is possible to determine how the elements of K multiply, and hence the structure of K, by an inductive argument. However, this can be avoided by exhibiting a loop, and proving that it is the free loop on one generator with the property that every isotope has the weak inverse property. These one-generator loops are of interest to us, on the one hand as proof that the class of loops we are studying is strictly larger than the class of Moufang loops, and on the other hand as examples of loops which are isomorphic to all their isotopes.

3. Let H be the set of all ordered pairs of integers $[i, k]$ under the binary operation defined by the following four equations

$$\begin{aligned}[2i, k][2j, m] &= [2i + 2j, k + m]\\ [2i + 1, k][2j, m] &= [2i + 2j + 1, k + m + j]\\ [2i, k][2j + 1, m] &= [2i + 2j + 1, m - k]\\ [2i + 1, k][2j + 1, m] &= [2i + 2j + 2, m - k - j]\,,\end{aligned}\tag{11}$$

where i, j, k, m are arbitrary integers. It is easy to verify that each product is uniquely defined by these equations and that H is a loop. As suggested by (11) it will be convenient hereafter to call an element of H odd or even if its first component is odd or even respectively. By checking each of the eight possible cases, the following result may easily be verified:

LEMMA 3. *A triple of elements u, v, w of H associate if and only if at least one of them is even. If all three are odd, then $u\cdot vw = (uv\cdot w)a$, where $a = [0, 1]$.*

COROLLARY 1. *The nucleus of H is the set of all even elements.*

Since three elements whose product is the identity cannot all be odd, we also have:

COROLLARY 2. *H has the weak inverse property.*

The fact that all isotopes of H also have the weak inverse property will follow from the following stronger result:

THEOREM 4. *Every isotope of H is isomorphic to it.*

Let H_0 be the isotope of H defined by $y \circ z = yR^{-1}(g) \cdot zL^{-1}(f)$, and let u and v be defined as follows: $u = [0, 0]$ if f is even, $u = [-1, 0]$ if f is odd, $v = [0, 0]$ if g is even, and $v = [1, 0]$ if g is odd. Then defining s and t by the relations $f = us$ and $sg = vt$, we observe that s and t are even, and hence in the nucleus. Thus,

$$L^{-1}(f) = L^{-1}(us) = [L(s)L(u)]^{-1} = L^{-1}(u)L^{-1}(s) ,$$

and

$$\begin{aligned} R^{-1}(g)R^{-1}(s) &= [R(s)R(g)]^{-1} = R^{-1}(sg) = R^{-1}(vt) \\ &= [R(v)R(t)]^{-1} = R^{-1}(t)R^{-1}(v) . \end{aligned}$$

Using these relations we have

$$\begin{aligned} y \circ z &= yR^{-1}(g) \cdot zL^{-1}(u)L^{-1}(s) = yR^{-1}(g) \cdot s^{-1} \cdot zL^{-1}(u) \\ &= yR^{-1}(g)R^{-1}(s) \cdot zL^{-1}(u) = yR^{-1}(t)R^{-1}(v) \cdot zL^{-1}(u) \\ &= yR^{-1}(t)R^{-1}(v) \cdot zR^{-1}(t)L^{-1}(u)R(t) \\ &= [yR^{-1}(t)R^{-1}(v) \cdot zR^{-1}(t)L^{-1}(u)]R(t) . \end{aligned}$$

But then, defining the isotope $y \otimes z = [yR(t) \circ zR(t)]R^{-1}(t) = yR^{-1}(v) \cdot zR^{-1}(u)$, we see that, up to isomorphism, we need only consider the four cases where $f = u$ and $g = v$.

Now, if $y \circ z = yR^{-1}(v) \cdot z$ where $v = [1, 0]$, define $G_\times$ by $y \times z = [yR(v) \cdot zR(v)]R^{-1}(v) = [y \cdot zR(v)]R^{-1}(v)$, and up to isomorphism we may consider the isotope $G_\times$ instead of G_0. From Lemma 3, we observe that $y \times z = yz$ if either y or z is even, and $y \times z = yz \cdot a^{-1}$ if both are odd. Similarly, if $y \circ z = y \cdot zL^{-1}(u)$ where $u = [-1, 0]$, we define $G_\otimes$ by $y \otimes z = [yL(u) \circ zL(u)]L^{-1}(u) = [yL(u) \cdot z]L^{-1}(u)$, and computing $y \otimes z$ from Lemma 3, we find that $y \otimes z = y \times z$. Finally, if $y \circ z = yR^{-1}(v) \cdot zL^{-1}(u)$ where $u = [-1, 0]$ and $v = [1, 0]$, we would like to show that $y \circ z = y \times z$. But this is equivalent to $yR^{-1}(v) \cdot zL^{-1}(u) = [y \cdot zR(v)]R^{-1}(v)$, or $pq \cdot v = pv \cdot (uq \cdot v)$, where we have right-multiplied by v and set $y = pv$ and $z = uq$. Using Lemma 3, this identity may be easily checked for all four cases of p and q odd and even.

It now only remains to show that the isotope $G_\times$ is isomorphic to G. Letting T be the permutation sending $[i, k]$ into $[-i, -k]$, we shall verify that $y \times z = (yT \cdot zT)T^{-1}$. If either y or z is even, it is easy to check visually from (11) that $yT \cdot zT = (yz)T$. And if both are odd, then we have

$$\begin{aligned}([2i+1, k]T \cdot [2j+1, m]T)T^{-1} &= ([-2i-1, -k][-2j-1, -m])T \\ &= [-2i-2j-2, -m+k+j+1]T \\ &= [2i+2j+2, m-k-j-1] = [2i+1, k][2j+1, m] \cdot a^{-1},\end{aligned}$$

to complete the proof.

Now let K be the free loop on one generator with the property that every isotope has the weak inverse property. Then we may induce a homomorphism φ of K onto H by sending the generator x of K onto the element $[1, 0]$, which generates H. Under this homomorphism, the element a, defined at the end of § 3, can be seen to go onto $[0, 1]$ (by mapping the relation $x^\rho a = x^\lambda$, for example). If A is the cyclic subgroup of K generated by a, then no element of A is in the kernel of φ since $[0, 1]$ has infinite order in H. But $K/A \rightarrow H/\varphi(A)$ also has no kernel, since both are infinite cyclic, and hence, φ is an isomorphism.

THEOREM 5. *The loop H defined by the relations* (11) *is the free loop on one generator with the property that every isotope has the weak inverse property.*

It might be pointed out that every homomorph of H also has the property that it is isomorphic to all its isotopes. By imposing the relations $x^{4m} = a^n = 1$ for integers $m \geq 1$ and $n \geq 2$, we get a loop of order $4mn$ with this property, which is not a group.

4. In this section we shall prove that a cross inverse property loop can only be isotopic to an inverse property loop if it is commutative (and hence already satisfies the inverse property itself). In addition to clarifying the relation between two well known classes of weak inverse property loops, this result is of interest to us here because the method of proof is identical with those used in the rest of the paper.

Let G be an inverse property loop and let G_0 be the isotope given by $a \circ b = ag^{-1} \cdot f^{-1}b$. If G_0 has the cross inverse property, then $(a \circ b) \circ a^{\rho_0} = b$, or $(ag^{-1} \cdot f^{-1}b)g^{-1} \cdot f^{-1}a^{\rho_0} = b$, where ρ_0 is the right inverse operator in G_0. Setting $a = x^{-1}g$ and $b = y^{-1}$ gives

$$(x^{-1} \cdot f^{-1}y^{-1})g^{-1} \cdot f^{-1}(x^{-1}g)^{\rho_0} = y^{-1},$$

or

$$(x^{-1} \cdot f^{-1}y^{-1})g^{-1} = y^{-1} \cdot [f^{-1}(x^{-1}g)^{\rho_0}]^{-1},$$

and taking the inverse of both sides yields $g(yf \cdot x) = [f^{-1}(x^{-1}g)^{\rho_0}]y$. Using the special case obtained by setting $y = 1$, we may rewrite this equation as $g(yf \cdot x) = (g \cdot fx)y$. Finally, replacing y by yf^{-1} gives $(g \cdot fx) \cdot yf^{-1} = g(yx)$. We are motivated by this relation to define an *anti-autotopism* $[U, V, W]$ to be an ordered triple of permutations on G satisfying $xU \cdot yV = (yx)W$ for all pairs of elements x and y of G. We may then express our last relation by saying that

$$[L(f)L(g), R^{-1}(f), L(g)] \tag{12}$$

is an anti-autotopism. By adapting Lemma 1 to the case of anti-autotopisms and of loops with the inverse property, it is easy to verify that if $[U, V, W]$ is an anti-autotopism, then so is $[W, \rho V \rho, U]$. Hence (12) may also be put in the form

$$[L(g), L(f), L(f)L(g)] \, . \tag{13}$$

But if $[U_1, V_1, W_1]$ and $[U_2, V_2, W_2]$ are anti-autotopisms, then $(xy) W_1 W_2 = (yU_1 \cdot x V_1) W_2 = x V_1 U_2 \cdot y U_1 V_2$, so that $(V_1 U_2, U_1 V_2, W_1 W_2)$ is an autotopism. Hence the anti-autotopism $[U, V, W]$ has an inverse anti-autotopism given by $[V^{-1}, U^{-1}, W^{-1}]$. Using this information, we may combine (12) and (13) to get the autotopism

$$\begin{aligned}
&[L(g), L(f), L(f)L(g)]\,[L(f)L(g), R^{-1}(f), L(g)]^{-1} \\
&\quad = [L(g), L(f), L(f)L(g)]\,[R(f), L^{-1}(g)L^{-1}(f), L^{-1}(g)] \\
&\quad = (L(f)R(f), L^{-1}(f), L(f)).
\end{aligned}$$

Then the analogue of Lemma 1 for inverse property loops allows us to conclude that $(L(f), R(f), L(f)R(f))$ is also an autotopism, which shows that f is in the Moufang nucleus of G. Since the inverse property and cross inverse property are both symmetric, we may conclude by duality that g is also in the Moufang nucleus. But then the isotope G_0 given by $x \circ y = xg^{-1} \cdot f^{-1}y$ will also have the inverse property, as was to be proved.

It might be remarked that a non-commutative cross inverse property loop may not be obtained from an inverse property loop even by allowing anti-isotopisms. This is because every anti-isotopism is the product of an ordinary isotopism and the canonical anti-isotopism given by $x \circ y = y \cdot x$, which clearly preserves both the inverse property and the cross inverse property.

BIBLIOGRAPHY

1. R. Artzy, *Cross inverse and related loops,* Trans. Amer. Math. Soc., **91** (1959), 480–492.
2. R. Baer, *Net and groups II.* Trans. Amer. Math. Soc., **47** (1940), 435–439.

3. R. H. Bruck, *Contributions to the theory of loops,* Trans. Amer. Math. Soc., **55** (1944), 245–354.
4. ———, *Pseudo-automorphisms and Moufang loops,* Proc. Amer. Math. Soc., **3** (1952), 66–72.
5. ———, *Some theorems on Moufang loops,* Submitted to Illinois Jour. Math.).

UNIVERSITY OF WISCONSIN

UN THÉORÈME DE RÉALISATION DE GROUPES RÉTICULÉS

P. Ribenboim

En 1939, Lorenzen [4] a demontré que tout groupe réticulé est isomorphe à un sous-groupe sous-réticulé d'un produit direct de groupes totalement ordonnés. Pour le faire, il a utilisé la théorie des systèmes d'idéaux, introduite auparavant par Krull [2] dans l'arithmétique des anneaux d'intégrité commutatifs.

Dans cette note, nous démontrons ce même théorème par une methode distincte, qui utilise la notion de filet de Jaffard [1]. Notre démonstration semble plus transparente et met en relief certains aspects d'intérêt qui ne sont pas du tout apparents dans le travail de Lorenzen: (1) la réalisation qui nous obtenons est complètement régulière, de Hausdorff et fidèle (ces termes sont définis ci-dessous); (2) il y a une relation entre les ultrafiltres de l'ensemble des filets du groupe donné et les pré-ordres totaux plus fins, laquelle est utile pour exprimer l'ordre donné comme la conjonction d'ordres totaux plus fins (cf. Krull [3], Ribenboim [5, 6]).

1. Rappelons d'abord les définitions et résultats qui seront utilisés. Soit G un groupe (abélien additif) réticulé (selon l'ordre $\leqq$) et notons P l'ensemble des éléments positifs de G; G est un réticulé distributif. Si $f \in P$ soit $E(f) = \{g \in P \mid g \wedge f = 0\}$ l'ensemble des éléments de P étrangers à f. Posons $f \equiv g$ si et seulement si $E(f) = E(g)$. La classe d'équivalence $\bar{f}$ contenant l'élément $f \in P$ s'appelle le *filet* de f. Soit $\mathscr{F}$ l'ensemble des filets du groupe G, determinés par les éléments $f \in P$. $\mathscr{F}$ est ordonné en posant $\bar{f} \leqq \bar{g}$ si et seulement si $E(f) \supseteq E(g)$; $\mathscr{F}$ possède un premier élément $\bar{0} = \{0\}$; si $f, g \in P, f \leqq g$ alors $\bar{f} \leqq \bar{g}$; $\mathscr{F}$ est un réticulé distributif:

$$\bar{f} \wedge \bar{g} = \overline{f \wedge g}, \bar{f} \vee \bar{g} = \overline{f \vee g} = \overline{f + g};$$

on a $\bar{f} \wedge \bar{g} = \bar{0}$ si et seulement si $f \wedge g = 0$; $\mathscr{F}$ est disjonctif: si $\bar{f}, \bar{g} \in \mathscr{F}$ $\bar{f}$ ne suit pas $\bar{g}$ dans $\mathscr{F}$, il existe $\bar{h} \in \mathscr{F}$ tel que $\bar{0} \neq \bar{h} \leqq \bar{g}, \bar{h} \wedge \bar{f} = \bar{0}$.

Si $f \in G$ (mais non nécessairement $f \in P$) posons par définition: $\bar{f} = \bar{f}_+ \vee \bar{f}_-$, donc $\bar{f} = \overline{|f|}$, où $|f| = f_+ + f_- \in P$.

Soit $(G_\iota)_{\iota \in I}$ une famille de groupes totalement ordonnés, $\prod_{\iota \in I} G_\iota$ leur produit direct ordonné; un isomorphisme θ d'un groupe ordonné G dans $\prod_{\iota \in I} G_\iota$ s'appelle une *réalisation* lorsque $pr_\iota \theta(G) = G_\iota$ quelque soit $\iota \in I$. Par un théoréme de Lorenzen-Dieudonné, un groupe ordonné G admet une réalisation si et seulement s'il vérifie la condition suivante: si $f \in G$

Received March 16, 1959.

et $nf \geqq 0$, où n est un entier strictement positif, alors $f \geqq 0$. En particulier, tout groupe réticulé satisfait cette condition et admet alors une réalisation.

La réalisation $\theta : G \to \theta(G) \subseteq \prod_{\iota \in I} G_\iota$ est dite *concordante* (ou propre) lorsque $\theta(G)$ est un sous-réticulé de $\prod_{\iota \in I} G_\iota$ (la réalisation obtenue dans la démonstration du théorème de Lorenzen-Dieudonné n'est pas concordante).

Si $f \in G$ notons $\sigma(f) = \{\iota \in I \mid pr_\iota \theta(f) \neq 0\}$. Alors $\sigma(f \vee g) = \sigma(f) \cup \sigma(g)$, $\sigma(f \wedge g) = \sigma(f) \cap \sigma(g)$ lorsque θ est une réalisation concordante.

La réalisation $\theta : G \to \theta(G) \subseteq \prod_{\iota \in I} G_\iota$ est dite *complètement régulière* lorsque: si $\iota \in I$, $f \in P$, $\iota \notin \sigma(f)$ alors il existe $g \in P$ tel que $\iota \in \sigma(g)$, $\sigma(f) \cap \sigma(g) = \phi$.

La réalisation $\theta : G \to \theta(G) \subseteq \prod_{\iota \in I} G_\iota$ est dite *de Hausdorff* lorsque: si $\iota, \kappa \in I$, $\iota \neq \kappa$, il existe $f, g \in P$ tels que $\iota \in \sigma(f)$, $\kappa \in \sigma(g)$, $\sigma(f) \cap \sigma(g) = \phi$.

La réalisation indiquée dans le théorème de Lorenzen-Dieudonné n'est pas complètement régulière, ni de Hausdorff.

LEMME 1. *Si θ est une réalisation concordante et complètement régulière alors θ est fidèle, c'est-à-dire: si $f, g \in P$ alors $\bar{f} = \bar{g}$ si et seulement si $\sigma(f) = \sigma(g)$; ainsi σ définit un isomorphisme du réticulé $\mathscr{F}$ des filets de G sur un sous-réticulé de celui de parties de l'ensemble I, en posant $\bar{\sigma}(f) = \sigma(\bar{f})$.*

En effet, soit $\bar{f} = \bar{g}$ et $\iota \in \sigma(g)$, $\iota \notin \sigma(f)$; alors il existe $h \in P$ tel que $\iota \in \sigma(h)$, $\sigma(h) \cap \sigma(f) = \phi$; donc $\sigma(h \wedge f) = \phi$ et $h \wedge f = 0$; or, θ étant concordante, $\iota \in \sigma(h) \cap \sigma(g) = \sigma(h \wedge g)$ donc $h \wedge g \neq 0$ et alors $\bar{f} \neq \bar{g}$, absurde! Ainsi, on a bien $\sigma(f) = \sigma(g)$.

Réciproquement, si $\sigma(f) = \sigma(g)$, si $h \in P$ est tel que $h \wedge f = 0$, alors

$$\phi = \sigma(h \wedge f) = \sigma(h) \cap \sigma(f) = \sigma(h) \cap \sigma(g) = \sigma(h \wedge g)$$

donc $h \wedge g = 0$, et vice-versa, donc $\bar{f} = \bar{g}$.

Par conséquent, si on pose $\bar{\sigma}(\bar{f}) = \sigma(f)$ alors $\bar{\sigma}(\bar{0}) = \phi$, $\bar{\sigma}(\bar{f} \vee \bar{g}) = \bar{\sigma}(\bar{f}) \cup \bar{\sigma}(\bar{g})$, $\bar{\sigma}(\bar{f} \wedge \bar{g}) = \bar{\sigma}(\bar{f}) \cap \bar{\sigma}(\bar{g})$(où $\bar{f}, g \in P$) et $\bar{\sigma}$ est bien un isomorphisme du réticulé $\mathscr{F}$ des filets de G dans celui des parties de I.

2. Théoréme de réalisation. Tout groupe réticulé admet une réalisation concordante et complètement régulière.

Démonstration.

1°. Soit G le groupe réticulé donné, $\mathscr{F}$ le réticulé des filets de G, Ω l'ensemble des ultrafiltres de $\mathscr{F}$. Pour tout $U \in \Omega$ soit $P_U = \{f \in G \mid$ il existe $\alpha \in U$ tel que $\bar{f_-} \wedge \alpha = \bar{0}\}$. Alors $P_U + P_U \subseteq P_U$. En effet, si $f, g \in G$ et $\beta, \gamma \in U$ sont tels que $\bar{f_-} \wedge \beta = \bar{0}$, $\bar{g_-} \wedge \gamma = \bar{0}$ alors

$$\overline{(f+g)_-} \wedge (\beta \wedge \gamma) \leqq \overline{f_- + g_-} \wedge (\beta \wedge \gamma) = (\bar{f_-} \vee \bar{g_-}) \wedge (\beta \wedge \gamma)$$
$$= (\bar{f_-} \wedge \beta \wedge \gamma) \vee (\bar{g_-} \wedge \beta \wedge \gamma) = \bar{0}$$

avec $\beta \wedge \gamma \in U$, donc $f + g \in P_U$. De même, $P_U \cap (-P_U) = \{f \in G \mid \text{il existe } \alpha \in U \text{ tel que } \bar{f} \wedge \alpha = \bar{0}\}$. En effet, si $\bar{f} \wedge \alpha = \bar{0}$ alors de $\bar{f} = \bar{f}_+ \vee \bar{f}_-$ vient $\bar{0} = \bar{f} \wedge \alpha = (\bar{f}_+ \vee \bar{f}_-) \wedge \alpha = (\bar{f}_+ \wedge \alpha) \vee (\bar{f}_- \wedge \alpha)$ donc $\bar{f}_- \wedge \alpha = \bar{0}$, $\overline{(-f)_-} \wedge \alpha = \bar{f}_+ \wedge \alpha = \bar{0}$; réciproquement. Enfin, $G = P_U \cup (-P_U)$. En effet, si $f \notin P_U$ alors $\bar{f}_- \wedge \alpha \neq \bar{0}$ quelque soit $\alpha \in U$, donc $\bar{f}_+ \notin U$ (car $\bar{f}_- \wedge \bar{f}_+ = \bar{0}$) et alors il existe $\beta \in U$ tel que $\bar{0} = f_+ \wedge \beta = \overline{(-f)_-} \wedge \beta$, c'est-à-dire $-f \in P_U$.

2°. Nous venons de voir que P_U définit un pré-ordre total compatible sur G. Soit $G'_U = G/(P_U \cap (-P_U))$ et considérons $P'_U = P_U/(P_U \cap (-P_U))$ donc P'_U définit un ordre total compatible sur G'_U. Pour tout $f \in G$ soit f'_U son image canonique en G'_U. Si $f \in G$ posons $\theta(f) = (f'_U)_{U \in \Omega} \in \prod_{U \in \Omega} G'_U$ et montrons que θ est un isomorphisme de G dans le produit direct ordonné $\prod_{U \in \Omega} G'_U$. D'abord $\theta(f + g) = \theta(f) + \theta(g)$ car $(f + g)'_U = f'_U + g'_U$ quelque soit $U \in \Omega$. Si $f \neq 0$ il existe $U \in \Omega$ tel que $\bar{f} \in U$ donc $\bar{f} \wedge \alpha \neq \bar{0}$ quelque soit $\alpha \in U$ et $f \notin P_U \cap (-P_U)$, c'est-à-dire $f'_U \neq 0$ et alors $\theta(f) \neq 0$. Si $f \in P$ alors $f_- = 0$ donc $f \in P_U$ quelque soit $U \in \Omega$ et alors $\theta(f) = (f'_U)_{U \in \Omega}$ est positif dans $\prod_{U \in \Omega} G'_U$. Réciproquement, si $f \notin P$ alors $f_- \neq 0$ donc il existe $U \in \Omega$ tel que $\overline{f_-} \in U$ et alors $f \notin P_U$ car $\bar{f}_- \wedge \alpha \neq \bar{0}$ quelque soit $\alpha \in U$, donc $\theta(f)$ n'est pas positif, car $f'_U \notin P'_U$.

3°. θ est une réalisation, car $pr_U \theta(G) = G'_U$ quelque soit $U \in \Omega$; en effet, si $f'_U \in G'_U$ avec $f \in G$ alors $\mathrm{pr}_U \theta(f) = f'_U$. L'isomorphisme θ est concordant, c'est-à-dire, $\theta(f \wedge g) = \inf \{\theta(f), \theta(g)\}$, $\theta(f \vee g) = \sup \{\theta(f), \theta(g)\}$ (inf et sup pris dans $\prod_{u \in \Omega} G'_U$); pour cela, on doit montrer que $pr_U \theta(f \wedge g) = \inf \{pr_U \theta(f), pr_U \theta(g)\}$ (inf pris dans G'_U) pour tout $U \in \Omega$ (analoguement pour le sup). Puisque G'_U est totalement ordonné, alors par exemple $f'_U \leqq g'_U$ dans $g'_U - f'_U = (g - f)'_U \in P'_U$ et alors $g - f \in P_U$, donc il existe $\beta \in U$ tel que $\overline{(g - f)_-} \wedge \beta = \bar{0}$; or, alors $f \wedge g$ et f sont quasi-égaux selon le pré-ordre défini par P_U, c'est-à-dire, $f \wedge g - f \in P_U \cap (-P_U)$; en effet, $\overline{f \wedge g - f} \wedge \beta = \overline{0 \wedge (g - f)} \wedge \beta = \overline{-(g - f)_-} \wedge \beta = \bar{0}$. Donc $pr_U \theta(f \wedge g) = (f \wedge g)'_U = f'_U = f'_U \wedge g'_U = \inf \{pr_U \theta(f), pr_U \theta(g)\}$.

4°. Montrons maintenant que la réalisation est complètement régulière. Soit $U \in \Omega, f \in P$ tel que $U \notin \sigma(f) = \{V \in \Omega \mid pr_V \theta(f) \neq 0\}$, donc $pr_U \theta(f) = 0$ et alors $f \in P_U \cap (P_U)$; ainsi, il existe $g \in P, \bar{g} \in U$, tel que $\bar{f} \wedge \bar{g} = \bar{0}$, donc $f \wedge g = 0$ et par conséquent $\sigma(f) \cap \sigma(g) = \phi$; enfin $pr_U \theta(g) \neq 0$, c'est-à-dire $g \notin P_U \cap (-P_U)$, sinon il existe $\beta \in U$ tel que $\bar{g} \wedge \beta = \bar{0}$, absurde!

Remarque. On a $\sigma(f) = \{U \in \Omega \mid \bar{f} \in U\}$ quelque soit $f \in G$, et θ est une réalisation de Hausdorff. En effet, si $f \in G, \bar{f} \in U$, alors $pr_U \theta(f) \neq 0$, car sinon $f \in P_U \cap (-P_U)$ et il existe $\alpha \in U$ tel que $\bar{f} \wedge \alpha = \bar{0}$, absurde!

Réciproquement, si $\bar{f} \notin U$ il existe $\beta \in U$ tel que $\bar{f} \wedge \beta = \bar{0}$ donc $f \in P_U \cap (-P_U)$ et alors $pr_U \theta(f) = 0$. Il résulte que si $U, V \in \Omega$, $U \neq V$, il existe $f, g \in P$ tel que $\bar{f} \in U, \bar{f} \notin V, \bar{g} \in V, \bar{f} \wedge \bar{g} = \bar{0}$, donc $\bar{g} \notin U$ et alors $U \in \sigma(f)$, $V \in \sigma(g), f \wedge g = 0$, donc $\sigma(f) \cap \sigma(g) = \phi$.

References

1. P. Jaffard, *Contribution à l'Etude des Groupes Ordonnés*, J. Math. Pures et Appl., **32** (1953), 203–280.
2. W. Krull, *Beiträge zur Arithmetik kommutativer Integritätsbereiche* I, Math. Z., **41** (1936), 545–577.
3. W. Krull, *Uber geordnete Gruppen von reellen Funktionen*, Math. Z., **64** (1956), 10–40.
4. P. Lorenzen, *Abstrakte Begründung der multiplikativen Idealtheorie*, Math. Z., **45** (1939), 533–553.
5. P. Ribenboim, *Conjonction d'Ordres dans les Groupes Abéliens Ordonnés* Anais Acad. Bras. Ci., **29** (1957), 201–224.
6. ———, *Sur quelques Construction de Groupes Réticulés et l'Equivalence Logique entre l'Affinement de Filtres et d'Ordres*, Summa Bras. Math., **4** (1958) 65-89.

Instituto de Matemática Pura e Aplicada,
Rio de Janeiro, Brasil

AN INVERSION THEOREM FOR LAPLACE-STIELTJES TRANSFORMS

DANIEL SALTZ

E. Phragmén [2; p. 360] showed that under certain assumptions of boundedness for $F(x)$,

$$\lim_{s\to+\infty}\int_0^t F(\tau)[1-\exp(-e^{(t-\tau)s})]\,d\tau=\int_0^t F(\tau)d\tau\ .$$

If we write $1-\exp(-e^{s(t-\tau)})=\sum_1^\infty(-1)^{n+1}e^{nx(t-\tau)}/n!$ in the above formula, and interchange sum and integral, we formally obtain

$$\lim_{s\to\infty}\sum_1^\infty\frac{(-1)^{n+1}}{n!}e^{nst}\int_0^t e^{-ns\tau}F(\tau)d\tau=\int_0^t F(\tau)d\tau\ .$$

G. Doetsch [1; pp. 286–288] showed that for reals s, if $f(s)=\int_0^\infty e^{-s\tau}F(\tau)d\tau$ converges absolutely in some half-plane, then

$$\int_0^t F(\tau)d\tau=\lim_{s\to+\infty}\sum_1^\infty\frac{(-1)^{n+1}}{n!}f(ns)e^{nst}\ \text{for}\ t>0\ .$$

This paper will generalize this result to Laplace-Stieltjes transforms

$$\text{(I)}\qquad f(s)=\int_0^\infty e^{-st}\,d\alpha(t)$$

and will eliminate the assumption of absolute convergence. Unless specifically written otherwise, all integrals will be evaluated from 0 to $+\infty$ and all summations from 1 to ∞. We shall need the following two propositions [3; pp 39,41]:

LEMMA 1. *If the integral*

$$f(s_0)=\int e^{-s_0t}\,d\alpha(t)$$

converges with $Rs_0>0$, *then*

$$f(s_0)=s_0\int e^{-s_0t}\,\alpha(t)dt-\alpha(0)$$

and $\int e^{-s_0t}\,\alpha(t)dt$ *converges absolutely if* s_0 *is replaced by any number with larger real part.*

Received December 27, 1957, and in revised form April 20, 1959.

Lemma I remains valid for $Rs_0 < 0$ if we insist that $\alpha(\infty) = 0$. In this paper we shall make the following

Assumption. In (I), s is real and positive, and $\alpha(t)$ is of bounded variation in $(0, R)$, for every $R > 0$.

LEMMA 2. *If the integral*

$$\int_0^\infty e^{-s\tau}\, d\alpha(\tau)$$

converges for $s = s_0$ and if the real part γ of s_0 is positive, then $\alpha(\tau) = 0(e^{\gamma\tau})$ as $\tau \to \infty$.

We shall now prove some useful lemmas.

LEMMA 3. *If* (I) *converges in some half plane Γ, then*

(a) $$\lim_{s\to\infty}\left|\int_\sigma^\infty [1 - \exp(-e^{-s\tau})]d\alpha(\tau)\right| = 0 \textit{ for fixed } \sigma > 0\,,$$

(b) $$\lim_{\sigma\to\infty}\left|\int_\sigma^\infty [1 - \exp(-e^{-s\tau})]d\alpha(\tau)\right| = 0 \textit{ for fixed } s > 0\,.$$

Proof. Since $1 - \exp(-e^{-s\tau}) = O(e^{-s\tau})$ for $s,\tau \geq 0$, a standard argument involving integration by parts shows that

$$\int_\sigma^8 [1 - \exp(-e^{-s\tau})]d\alpha(\tau) = O\{e^{-s\sigma}[|\alpha(\sigma)| + s]\}$$

for $s \in \Gamma$ and $\sigma \geq 0$. The desired result now follows from Lemmas 1 and 2.

LEMMA 4. *If* (I) *converges in some half-plane Γ, then for $s \in \Gamma'$ where Γ' is a half-plane properly contained in Γ,*

$$\sum \frac{(-1)^{n+1}}{n!}\int e^{ns(t+\tau)}\, d\alpha(\tau) = \int d\alpha(\tau) \sum \frac{(-1)^{n+1}}{n!} e^{ns}(t-\tau)\,.$$

Proof. Upon integration by parts, application of Lemma 2, and some algebra, the desired equality takes the form

$$\sum_1^\infty \frac{(-1)^{n+1}}{(n-1)!}\int e^{ns(t-\tau)}\,\alpha(\tau)d\tau = \int \sum_1^\infty \frac{(-1)^{n+1}}{(n-1)!} e^{ns(t-\tau)}\,\alpha(\tau)d\tau\,.$$

To verify this latter equality, it suffices to show that

$$\sum_0^\infty \frac{1}{(n-1)!}\int e^{ns(t-\tau)}\,|\alpha(\tau)|d\tau < \infty\,,$$

but this follows from Lemma 1.

LEMMA 5. *If* (I) *converges in some half-plane Γ, then*

$$\alpha(t)-\alpha(0)=\lim_{s\to+\infty}\sum\frac{(-1)^{n+1}}{n!}f(ns)e^{nst}$$

for all non-negative t which are points of continuity of $\alpha(t)$.

Proof. We have

$$\sum\frac{(-1)^{n+1}}{n!}f(ns)e^{nst}=\int d\alpha(\tau)\sum\frac{(-1)^{n+1}}{n!}e^{ns(t-\tau)}$$

$$=\int[1-\exp(-e^{s(t-\tau)})]d\alpha(\tau)\,,$$

the interchange in the order of summation and integration being justified by Lemma 4. For $t=0$ ($t>0$) and a point of continuity of $\alpha(t)$, write the integral on the right as

$$\int_0^\delta+\int_\delta^\infty\left(\text{or, for } t>0,\int_0^{t-\delta}+\int_{t-\delta}^{t-\delta}+\int_{t+\delta}^\infty\right),$$

with $0<\delta<t$ chosen that the total variation of α on $[0,\delta]$ (respectively. $[t-\delta,\ t+\delta]$) is less than ε, and apply Lemma 3 to $\int_\delta^\infty\left(\text{resp.,}\int_{t+\delta}^\infty\right)$. We see that $\int_0^\delta\left(\text{resp.,}\int_{t-\delta}^{t+\delta}\right)$ is less than ε for all $s\geq 0$. (For $t>0$, $\int_0^{t-\delta}=\alpha(t-\delta)-\alpha(0)-\int_0^{t-\delta}\exp[-e^{s(t-\tau)}]d\alpha(\tau)$, and this clearly tends to $\alpha(t-\delta)-\alpha(0)$ as $s\to\infty$. Thus the integral $\int_0^\infty$ is $\alpha(t)-\alpha(0)+o(1)$ as $s\to\infty$).

We can now prove our main result.

THEOREM *If* $\alpha(t)=[\alpha(t^+)+\alpha(t^-)]/2$ *for* $t>0$ *and* (I) *converges for some* $s>0$, *then*

$$\lim_{s\to\infty}\sum\frac{(-1)^{n+1}}{n!}f(ns)e^{nst}=\begin{cases}[\alpha(0^+)-\alpha(0)]\,(1-e^{-1}), & t=0\\ \alpha(t)-\alpha(0)-[\alpha(t^+)-\alpha(t^-)]\,(e^{-1}-1/2)\,, & t>0\end{cases}$$

Proof. Define

$$\beta(\tau)=\begin{cases}\alpha(\tau)-[\alpha(0^+)-\alpha(0)]\ \text{sign}\ \tau, & \tau>0\\ \alpha(0) & ,\ \tau=0\end{cases}$$

for $t=0$, and

$$\beta(\tau)=\begin{cases}\alpha(\tau)-[\alpha(t^+)-\alpha(t^-)]\ \text{sign}\ (\tau-t), & \tau>0\\ \alpha(t) & ,\ \tau=0\end{cases}$$

for $t>0$, β is then continuous at t, and

$$F(s) = \int_0^\infty e^{-s\tau}\, d\beta(\tau) = \begin{cases} f(s) - \alpha(0^+) + \alpha(0) & , t = 0 \\ f(s) - [\alpha(t^+) - \alpha(t^-)]e^{-st}, & t > 0 \, . \end{cases}$$

Now apply Lemma 5 with β and F substituted for α and f, respectively.

BIBLIOGRAPHY

1. G. Doetsch, *Handbuch der Laplace Trasformation* vol. **1**, Birkhäuser Basel, 1950.
2. E. Phragmén. *Sur une extension d'un théorème classique de la théorie des functions*, Acta Math., **28** (1904), 351-368.
3. D. V. Widder, *The Laplace Transform*, Princeton University Press, 1946.

NORTHWESTERN UNIVERSITY

STATISTICAL METRIC SPACES

B. SCHWEIZER AND A. SKLAR

Introduction. The concept of an abstract metric space, introduced by M. Fréchet in 1906 [2], furnishes the common idealization of a large number of mathematical, physical and other scientific constructs in which the notion of a "distance" appears. The objects under consideration may be most varied. They may be points, functions, sets, and even the subjective experiences of sensations. What matters is the possibility of associating a non-negative real number with each ordered pair of elements of a certain set, and that the numbers associated with pairs and triples of such elements satisfy certain conditions. However, in numerous instances in which the theory of metric spaces is applied, this very association of a *single* number with a pair of elements is, realistically speaking, an over-idealization. This is so even in the measurement of an ordinary length, where the number given as the distance between two points is often not the result of a single measurement, but the average of a series of measurements. Indeed, in this and many similar situations, it is appropriate to look upon the distance concept as a statistical rather than a determinate one. More precisely, instead of associating a number—the distance $d(p, q)$—with every pair of elements p, q, one should associate a distribution function F_{pq} and, for any positive number x, interpret $F_{pq}(x)$ as the probability that the distance from p to q be less than x. When this is done one obtains a generalization of the concept of a metric space—a generalization which was first introduced by K. Menger in 1942 [5] and, following him, is called a statistical metric space.

The history of statistical metric spaces is brief. In the original paper, Menger gave postulates for the distribution functions F_{pq}. These included a generalized triangle inequality. In addition, he constructed a theory of betweeness and indicated possible fields of application.

In 1943, shortly after the appearance of Menger's paper, A. Wald published a paper [14] in which he criticized Menger's generalized triangle inequality and proposed an alternative one. On the basis of this new inequality Wald constructed a theory of betweeness having certain advantages over Menger's theory [15].

In 1951 Menger continued his study of statistical metric spaces in a paper [7] devoted to a resume of the earlier work, the construction of several specific examples and further considerations of the possible applications of the theory. In this paper Menger adopted Wald's version

Received May 25, 1959.

of the triangle inequality.[1]

Statistical metric spaces are also considered by Menger in the last chapter of his book *Géométrie Générale* [9]; and references to these spaces are scattered throughout his other works, e.g., [6], [8].[2]

In the present paper we continue the study of statistical metric spaces. Our paper is divided into three parts, each devoted to one main topic. They are:

(I) The axiomatics of statistical metric spaces, with particular emphasis on the triangle inequality;

(II) The construction and study of particular spaces;

(III) A consideration of topological notions in statistical metric spaces and a study of the continuity properties of the distance function.[3]

In concluding this introduction, we wish to express our thanks to Professor K. Menger for his never-failing interest and encouragement, and to our colleagues, Professors T. Erber and M. McKiernan, for their many valuable comments and suggestions.

I. Definitions and Preliminaries

1. Statistical metric spaces. As is customary, we call a real-valued function defined on the entire real line a *distribution function* if it is non-decreasing, left-continuous and has inf 0 and sup 1. We shall use various symbols for distribution functions. However, in the sequel, H will always denote the specific distribution function defined by

$$H(x) = \begin{cases} 0, & x \leqq 0\,, \\ 1, & x > 0\,. \end{cases}$$

We shall also, for convenience, adhere to the convention that, for any distribution function F, and any $x > 0$, $F(x/0) = 1$, while $F(0/0) = 0$.

For purposes of reference and comparison, we list here the postulates, due originally to Fréchet, for an ordinary metric space. A metric space (briefly, an *M-space*) is an ordered pair (S, d), where S is an abstract set and d a mapping of $S \times S$ into the real numbers—i.e., d associates

[1] However, as Prof. Menger informs us, even before the paper was written, both he and Wald, in a number of conversations, had come to feel that the Wald inequality was in some respects too stringent a requirement to impose on all statistical metric spaces. Some support for this is furnished in the present paper (Theorems 5.4 and 6.4).

[2] In addition, in a note on Menger's paper [**7**], A. Špaček [**13**] has considered the question of determining the probability that a random function defined on every pair of points in a set is an ordinary metric on that set. In particular, he has established necessary and sufficient conditions for such a random function to be a metric with probability one. The connection between the concepts of Špaček and that of a statistical metric is considered in [**10**].

[3] Some of the results of this paper have been presented in [**12**].

a real number $d(p, q)$ with every pair (p, q) of elements of S. The mapping d is assumed to satisfy the following conditions:

(i) $d(p, q) = 0$ if, and only if, $p = q$. (Identity)

(ii) $d(p, q) \geqq 0$. (Positivity)

(iii) $d(p, q) = d(q, p)$. (Symmetry)

(iv) $d(p, r) \leqq d(p, q) + d(q, r)$. (Triangle Inequality)

DEFINITION 1.1. *A statistical metric space* (briefly, an *SM-space*) is an ordered pair $(S, \mathscr{F})$ where S is an abstract set (whose elements will be called *points*) and $\mathscr{F}$ is a mapping of $S \times S$ into the set of distribution functions—i.e., $\mathscr{F}$ associates a distribution function $\mathscr{F}(p, q)$ with every pair (p, q) of points in S. We shall denote the distribution function $\mathscr{F}(p, q)$ by F_{pq}, whence the symbol $F_{pq}(x)$ will denote the value of F_{pq} for the real argument x. The functions F_{pq} are assumed to satisfy the following conditions:

I. $F_{pq}(x) = 1$ for all $x > 0$ if, and only if, $p = q$.

II. $F_{pq}(0) = 0$.

III. $F_{pq} = F_{qp}$.

IV. If $F_{pq}(x) = 1$ and $F_{qr}(y) = 1$, then $F_{pr}(x + y) = 1$.

In view of Condition II, which evidently implies that $F_{pq}(x) = 0$ for all $x \leqq 0$, Condition I is equivalent to the statement: $p = q$ if, and only if, $F_{pq} = H$.

Every M-space may be regarded as an SM-space of a special kind. One has only to set $F_{pq}(x) = H(x - d(p, q))$ for every pair of points (p, q) in the M-space. Furthermore, with the interpretation of $F_{pq}(x)$ as the probability that the distance from p to q is less than x, one sees that Conditions I, II, and III are straightforward generalizations of the corresponding conditions i, ii, iii. Condition IV is a 'minimal' generalization of the triangle inequality iv which may be interpreted as follows: If it is certain that the distance of p and q is less than x, and likewise certain that the distance of q and r is less than y, then it is certain that the distance of p and r is less than $x + y$.

Condition IV is always satisfied in M-spaces, where it reduces to the ordinary triangle inequality. However, in those SM-spaces in which the equality $F_{pq}(x) = 1$ does not hold (for $p \neq q$) for any finite x, IV will be satisfied only vacuously. It is therefore of interest to have 'stronger' versions of the generalized triangle inequality. We shall consider two such versions in detail. Before doing so, it is convenient to make the following:

DEFINITION 1.2. A triangle inequality will be said to hold *universally* in an *SM*-space if and only if it holds for all triples of points, distinct or not, in that space.

2. Menger spaces. In his original formulation [5], Menger gave as a generalized triangle inequality the following:

IVm. $F_{pr}(x + y) \geqq T(F_{pq}(x), F_{qr}(y))$ for all $x, y \geqq 0$,

where T is a 2-place function on the unit square satisfying:

(a) $0 \leqq T(a, b) \leqq 1$,
(b) $T(c, d) \geqq T(a, b)$ for $c \geqq a, d \geqq b$,
(c) $T(a, b) = T(b, a)$,
(d) $T(1, 1) = 1$,
(e) $T(a, 1) > 0$ for $a > 0$.

In view of condition (d), it follows that IVm contains IV as a special case. Because of the rather general nature of the function T, about the most that can be given as an interpretation of IVm is a statement such as: Our knowledge of the third side of a triangle depends in a symmetric manner on our knowledge of the other two sides and increases, or at least does not decrease, as our knowledge of these other two sides increases. The interpretation can, however, be made precise by choosing T to be a specific function. There are numerous possible choices for T. We list here six of the simplest:

T_1: $T(a, b) = \text{Max}(a + b - 1, 0)$, i.e., $T = \text{Max}(\text{Sum}-1, 0)$;
T_2: $T(a, b) = ab$, ,, $T = \text{Product}$;
T_3: $T(a, b) = \text{Min}(a, b)$, ,, $T = \text{Min}$;
T_4: $T(a, b) = \text{Max}(a, b)$, ,, $T = \text{Max}$;
T_5: $T(a, b) = a + b - ab$, ,, $T = \text{Sum-Product}$;
T_6: $T(a, b) = \text{Min}(a + b, 1)$, ,, $T = \text{Min}(\text{Sum}, 1)$.

The six functions are listed in order of increasing 'strength', where T'' is said to be *stronger* than T' (and T' *weaker* than T'') if $T''(a, b) \geqq T'(a, b)$ for all (a, b) on the unit square with strict inequality for at least one pair (a, b). Evidently, if IVm holds for any given T, it will hold a fortiori for all weaker T's. For $T = \text{Product}$, IVm may be interpreted as follows: The probability that the distance of p and r is less than $x + y$ is not less than the joint probability that, independently, the distance of p and q is less than x, and the distance of q and r is less than y. For $T = \text{Min}$ (Max), the interpretation is: The probability that the distance of p and r is less than $x + y$ is not less than the smaller (larger) of the probabilities that the distance of p and q is less than x and the distance of q and r is less than y. Similar interpretations may be given to the other choices of T. However as the following lemmas indicate, the three functions T_4, T_5, T_6 are actually too strong for most purposes.

LEMMA 2.1. *If an SM-space contains two distinct points, then* IVm *cannot hold universally in the space under the choice* $T = \text{Max}$.

Proof. Let p and q be two distinct points of the space and let x and y satisfy $0 < y < x$. Suppose IVm holds universally with $T = \text{Max}$. Then,

$$F_{pq}(x) \geqq \text{Max}\,(F_{pq}(x - y), F_{qq}(y)) = 1 \,.$$

But x can be *any* positive number, which by Condition I means $p = q$ and contradicts the assumption $p \neq q$.

LEMMA 2.2. *If an SM-space is not an M-space, and if* IVm *holds universally in the space for some choice of T satisfying* (a)—(e), *then the function T has the property that there exists a number a, $0 < a < 1$, such that* $T(a, 1) \leqq a$.

Proof. If an SM-space is not an M-space, then there is at least one pair p, q of (necessarily distinct) points for which F_{pq} assumes values other than 0 or 1. By the left-continuity and monotonicity of F_{pq}, this means that there is, not merely one point, but an open interval (x, y) on which we have $0 < F_{pq} < 1$. Now assume that $T(a, 1) = a + \phi(a)$, where $\phi(a) > 0$ for $0 < a < 1$. Let z be any point in (x, y) and take $t > 0$. Then

$$\begin{aligned} F_{pq}(z + t) &\geqq T(F_{pq}(z), F_{qq}(t)) \\ &= T(F_{pq}(z), 1) \\ &= F_{pq}(z) + \phi(F_{pq}(z)) \,. \end{aligned}$$

Letting $t \to 0+$, we have:

$$F_{pq}(z+) \geqq F_{pq}(z) + \phi(F_{pq}(z)) > F_{pq}(z) \,.$$

Thus F_{pq} is discontinuous at z, and therefore at every point of (x, y). But this is a contradiction, since a non-decreasing function can be discontinuous at only denumerably many points.

LEMMA 2.3. *If* IVm *holds universally in an SM-space and if T is continuous, then, for any $x > 0$,* $T(F_{pq}(x), 1) \leqq F_{pq}(x)$.

Proof. Let p, q and $x > 0$ be given and choose y such that $0 < y < x$. Then,

$$F_{pq}(x) \geqq T(F_{pq}(x-y), F_{qq}(y)) = T(F_{pq}(x-y), 1) .$$

Letting $y \to 0+$, we obtain

$$F_{pq}(x) \geqq \lim_{y\to 0+} T(F_{pq}(x-y), 1) .$$

But, by the assumed continuity of T,

$$\lim_{y\to 0+} T(F_{pq}(x-y), 1) = T(\lim_{y\to 0+} F_{pq}(x-y), 1) ,$$

while, by the left-continuity of F_{pq},

$$\lim_{y\to 0+} F_{pq}(x-y) = F_{pq}(x) .$$

This completes the proof.

Motivated by these lemmas, and noticing that the three weaker functions in our list of T's satisfy $T(a, 1) = a$, we are led to replace conditions (a), (d) and (e) by the condition,

(a′) $T(a, 1) = a$, $T(0, 0) = 0$.

This new condition implies that $T \leqq \mathrm{Min}$, for we have the inequalities

$$T(a, b) \leqq T(a, 1) = a ,$$
$$T(a, b) = T(b, a) \leqq T(b, 1) = b ,$$

whence $T(a, b) \leqq \mathrm{Min}\,(a, b)$. Thus, under (a′), Min becomes the strongest possible universal T. Similarly, the weakest possible T satisfying (a′), (b) and (c) is the function, henceforth denoted by T_w, which is given by,

$$T_w(x, y) = \begin{cases} a , & x = a, y = 1 \text{ or } y = a, x = 1 , \\ 0 , & \text{otherwise.} \end{cases}$$

It must not be construed, however, that functions stronger than Min or weaker than T_w thereby lose all interest; in fact, on numerous occasions, we shall find it of value to determine under what conditions —i.e., for which points p, q, r and for which numbers x, y—IVm holds for a function stronger than Min or weaker than T_w.

To the conditions on T considered thus far we also add the associativity condition,

(d′) $T[T(a, b), c] = T[a, T(b, c)]$,

which permits the extension of IVm to a polygonal inequality. Accordingly, we make the following:

DEFINITION 2.1. A *Menger space* is an *SM*-space in which IVm holds universally for some choice of T satisfying conditions (a′), (b), (c) and (d′).

The following lemma shows that, in determining whether or not an SM-space is a Menger space, only triples of distinct points need be considered.

LEMMA 2.4. If the points p, q, r are not all distinct, then IVm holds for the triple p, q, r under any choice of T satisfying (a′), (b), (c) and (d′).

Proof. We need only consider the choice $T = \text{Min}$. If $p = r$, then $F_{pr} = H$ and the conclusion is immediate. If $p = q \neq r$, then for $x, y \geqq 0$,

$$\text{Min}\,(F_{pq}(x), F_{qr}(y)) = \text{Min}\,(H(x), F_{qr}(y)) \leqq F_{qr}(y) \leqq F_{qr}(x+y) = F_{pr}(x+y)\ .$$

3. **Wald spaces.** The other generalized triangle inequality that we consider is the one due to A. Wald [14, 15]. It is:

IVw. $\qquad F_{pr}(x) \geqq [F_{pq} * F_{qr}](x)\ , \quad$ for all $x \geqq 0$,

where $*$ denotes convolution, i.e.,

$$[F_{pq} * F_{qr}](x) = \int_{-\infty}^{\infty} F_{pq}(x - y)dF_{qr}(y)\ .$$

Since $F_{pq}(x - y) = 0$ for $y \geqq x$, and $F_{qr}(y) = 0$ for $y \leqq 0$, we may evidently write

$$[F_{pq} * F_{qr}](x) = \int_{0}^{x} F_{pq}(x - y)dF_{qr}(y)\ .$$

Since the convolution of the distribution functions of two independent random variables gives the distribution function of their sum, the interpretation of IVw is: The probability that the distance of p and r is less than x is not less than the probability that the sum of the distance of p and q and the distance of q and r (regarded as independent) is less than x.

DEFINITION 3.1. A *Wald space* is an SM-space in which IVw holds universally.

THEOREM 3.1. *A Wald space is a Menger space under the choice $T = \text{Product}$.*

Proof. In a Wald space, for any $x, y \geqq 0$, we have

$$\begin{aligned} F_{pr}(x + y) &\geqq \int_0^{x+y} F_{pq}(x + y - z)dF_{qr}(z) \\ &= \int_0^{x+y}\left[\int_0^{x+y-z} dF_{pq}(t)\right]dF_{qr}(z) \\ &= \iint\limits_{\substack{t,z\geqq 0\\ t+z\leqq x+y}} dF_{pq}(t)dF_{qr}(z)\ . \end{aligned}$$

Now,

$$\iint\limits_{\substack{t,z\geqq 0\\ t+z\leqq x+y}} dF_{pq}(t)dF_{qr}(z) \geqq \iint\limits_{\substack{0\leqq t\leqq x\\ 0\leqq z\leqq y}} dF_{pq}(t)dF_{qr}(z)$$

since the rectangle $\{(t, z)\,;\, 0 \leqq t \leqq x,\, 0 \leqq z \leqq y\}$ is contained in the triangle $\{(t, z)\,;\, t, z \geqq 0,\, t + z \leqq x + y\}$ and the F's are non-decreasing. But,

$$\iint\limits_{\substack{0\leqq t\leqq x\\ 0\leqq z\leqq y}} dF_{pq}(t)dF_{qr}(z) = \int_0^x\int_0^y dF_{pq}(t)dF_{qr}(z)$$

$$= \int_0^x dF_{pq}(t)\int_0^y dF_{qr}(z) = F_{pq}(x)F_{qr}(y)\ .$$

Combining the various inequalities we obtain

$$F_{pr}(x + y) \geqq F_{pq}(x)F_{qr}(y) \tag{1}$$

which is IVm with $T =$ Product, and completes the proof of the theorem.

COROLLARY. *If the Wald inequality holds, then so does the inequality* IV.

Proof. By (1), if $F_{pq}(x) = 1$ and $F_{qr}(y) = 1$ then $F_{pr}(x + y) = 1$.

The following lemma is a counterpart to Lemma 2.4:

LEMMA 3.1. *If the points p, q, r are not all distinct, then* IVw *holds for the triple p, q, r.*

Proof. If $p = r$, this is immediate, since in this case, $F_{pr} = H$.

Otherwise, if $p = q \neq r$, then, for $x \geqq 0$,

$$F_{pr}(x) = F_{qr}(x) = \int_0^x dF_{qr}(y) = \int_0^x H(x - y)dF_{qr}(y)$$

$$= \int_0^x F_{pq}(x - y)dF_{qr}(y)\ .$$

The case $p \neq q = r$ follows on interchanging r and p.

THEOREM 3.2. *If in an SM-space,* IVm *holds under* $T =$ Max *for all triples of distinct points, then the space is a Wald space.*

Proof. Let p, q, r be distinct. Then for any $x \geqq 0$,

$$F_{pr}(x) \geqq \text{Max}\,(F_{pq}(0), F_{qr}(x)) = F_{qr}(x) = \int_0^x dF_{qr}(y)$$

$$\geqq \int_0^x F_{pq}(x - y)dF_{qr}(y)\ ,$$

since $0 \leqq F_{pq}(x - y) \leqq 1$. Therefore IVw holds for all triples of distinct points in the space. But, by the preceding lemma, IVw holds automatically for triples of non-distinct points. Consequently, IVw holds for all triples of points in the space.

This theorem is, in a sense, a partial converse to Theorem 3.1. As will be seen later, $T = \text{Max}$ in the theorem cannot be weakened to $T = \text{Min}$, let alone $T = \text{Product}$. Thus the true converse of Theorem 3.1 is false.

II. Particular Spaces

4. Equilateral spaces. The simplest metric spaces are the equilateral spaces in which

$$d(p, q) = \begin{cases} a\,, & \text{if } p \neq q \\ 0\,, & \text{if } p = q\,, \end{cases}$$

where a is positive.

Accordingly, we call an SM-space *equilateral* if, for some distribution function G satisfying $G(0) = 0$,

$$F_{pq}(x) = \begin{cases} G(x)\,, & \text{if } p \neq q\,, \\ H(x)\,, & \text{if } p = q\,, \end{cases} \tag{1}$$

where H is the distribution function defined in § 1. From (1) it follows that the Conditions I–IV defining an SM-space are satisfied.

THEOREM 4.1. *The means, medians, etc., of the statistical distances in an equilateral SM-space form an equilateral M-space.*

Proof. Any one of these quantities is zero when $p = q$ and a fixed positive number for any p, q when $p \neq q$.

THEOREM 4.2. *In an equilateral SM-space, the Menger triangle inequality* IVm *holds for any triple of distinct points under* $T = \text{Max}$, *and universally under* $T = \text{Min}$.

Proof. Since G is non-decreasing,

$$G(x + y) \geqq \text{Max}\,(G(x), G(y)) \geqq \text{Min}\,(G(x), G(y))\,,$$

and

$$G(x + y) \geqq \text{Min}\,(G(x), 1)\,.$$

COROLLARY. *An equilateral SM-space is a Wald space.*

Proof. This is a direct consequence of Theorem 3.2.

There are also equilateral SM-spaces in which IVm holds under a stronger choice of T.

EXAMPLE 1.

$$G(x) = \begin{cases} 0\,, & x \leqq 0\,, \\ x\,, & 0 \leqq x \leqq 1\,, \\ 1\,, & 1 \leqq x\,. \end{cases}$$

For any triple of distinct points in this space, IVm holds under $T = \mathrm{Min}\,(\mathrm{Sum}, 1)$, since in all cases we have $G(x + y) \geqq \mathrm{Min}\,(G(x) + G(y), 1)$.

EXAMPLE 2.

$$G(x) = \begin{cases} 0\,, & x \leqq 0\,, \\ 1 - e^{-x}\,, & x \geqq 0\,. \end{cases}$$

For any triple of distinct points in this space, IVm holds under $T = \mathrm{Sum}$—Product. This follows from the fact that $e^{-x}e^{-y} = e^{-(x+y)}$.

However, even through, as Examples 1 and 2 show, there are equilateral SM-spaces in which the generalized triangle inequality IVm holds under a stronger T than Max, the result of Theorem 4.2 is the best possible. This is shown by:

EXAMPLE 3.

$$G(x) = \begin{cases} 0\,, & x \leqq 0\,, \\ a\,, & 0 < x \leqq k\,, \\ b\,, & k < x \leqq 3k\,, \\ 1\,, & 3k < x\,, \end{cases}$$

where $0 < a \leqq b < 1$ and k is any positive number. Then for $0 < x \leqq k$, $k < y \leqq 2k$, we have $G(x + y) = b = \mathrm{Max}\,(a, b)$, whence IVm cannot hold under any choice of T which is stronger than Max.

5. Simple spaces. A class of SM-spaces, more interesting than the equilateral, may be obtained as follows:

Let (S, d) be an M-space and G a distribution function, different from H, satisfying $G(0) = 0$. For every pair of points p, q in S, define the distribution function F_{pq} as follows:

$$F_{pq}(x) = \begin{cases} G[x/d(p, q)]\,, & p \neq q\,, \\ H(x)\,, & p = q\,. \end{cases} \tag{2}$$

DEFINITION 5.1. An SM-space $(S, \mathscr{F})$ is said to be a *simple space* if and only if there exists a metric d on S and a distribution function G satisfying $G(0) = 0$, such that, for every pair of points p, q in S, $\mathscr{F}(p, q) = F_{pq}$ is given by (2). Furthermore, we say that $(S, \mathscr{F})$ is the simple space *generated* by the M-space (S, d) and the distribution function G.

THEOREM 5.1. *A simple space is a Menger space under any choice of T satisfying* (a′), (b), (c) *and* (d′).

Proof. It is sufficient to show that IVm holds universally under $T = \text{Min}$, since this is the strongest choice of T possible. Thus, in view of Lemma 2.4, we have only to show that for p, q, r distinct,

$$G\left(\frac{x+y}{d(p,r)}\right) \geqq \text{Min}\,[G(x/d(p,q)), G(y/d(q,r))] \,. \tag{3}$$

Now, since d is an ordinary metric,

$$d(p, r) \leqq d(p, q) + d(q, r) \,.$$

Thus,

$$\frac{x+y}{d(p,r)} \geqq \frac{x+y}{d(p,q)+d(q,r)} \,. \tag{4}$$

Furthermore, since $d(p, q)$ and $d(q, r)$ are positive,

$$\begin{aligned} \text{Max}\,[x/d(p,q), y/d(q,r)] &\geqq \frac{x+y}{d(p,q)+d(q,r)} \\ &\geqq \text{Min}\,[x/d(p,q), y/d(q,r)] \end{aligned} \tag{5}$$

with equality on either side if and only if $x/d(p, q) = y/d(q, r)$. Consequently, on combining (4) and the right-hand inequality in (5) we have,

$$\frac{x+y}{d(p,r)} \geqq \text{Min}\,[x/d(p,q), y/d(q,r)] \,,$$

which, since G is non-decreasing, implies (3) and completes the proof.

COROLLARY 1. *An equilateral M-space generates an equilateral SM-space.*

COROLLARY 2. *If $G(x) = H(x-1)$, the generated SM-space reduces to the generating M-space.*

Proof.

$$F_{pq}(x) = H\left(\frac{x}{d(p,q)} - 1\right) = H(x - d(p, q)) \,.$$

In most simple spaces $T = \text{Max}$ will be too strong since the left-hand inequality in (5) shows that, for a triple of distinct points p, q, r

such that $d(p, r) = d(p, q) + d(q, r)$, IVm under $T = \text{Max}$ fails. Indeed, in simple spaces having sufficient structure the choice $T = \text{Min}$ implicit in Theorem 5.1 is the best possible.

THEOREM 5.2. *If* (S, d) *is a finite-dimensional Euclidean space,* G *a continuous distribution function such that* $G(0) = 0$ *and* $0 < G(x) < 1$ *for all* $x > 0$, *then* Min *is the strongest* T *under which* IVm *holds for all triples of distinct points.*

Proof. Suppose T is stronger than Min. Then there exists at least one pair of numbers, $a, b\,(0 < a, b < 1)$, such that $T(a, b) > \text{Min}\,(a, b)$. We distinguish two cases:

(i) If $a = b$, choose $x = y$ such that $G(x) = a$, and choose $d(p, q) = d(q, r) = 1$, $d(p, r) = 2$. Then, since equality is attained in (3), we cannot have $T(a, a) > \text{Min}\,(a, a) = a$.

(ii) If $a \neq b$, we may suppose that $a < b$. Let $\varepsilon = T(a, b) - \text{Min}\,(a, b)$ and choose u, v so that $a = G(u)$ and $b = G(v)$. Such numbers u, v clearly exist since G is a continuous distribution function; moreover $u < v$. Also, since G is continuous, there exists an $h > 0$ such that

$$G(u + h) < G(u) + \varepsilon = a + \varepsilon\,.$$

Now let $d(q, r) = t$ be fixed and choose

$$d(p, q) = s \text{ such that } \frac{t}{s + t} < \frac{h}{v - u}\,,$$

$$d(p, r) = d(p, q) + d(q, r) = s + t\,,$$

$$x = ud(p, q) \text{ and } y = vd(q, r)\,.$$

Then,

$$\text{Min}\,[G(x/d(p, q)), G(y/d(q, r))] = \text{Min}\,[G(u), G(v)] = \text{Min}\,(a, b) = a\,.$$

Furthermore,

$$G\Big(\frac{x + y}{d(p, r)}\Big) = G\Big(\frac{us + vt}{s + t}\Big) = G\Big(u + \frac{(v - u)t}{s + t}\Big) \leqq G(u + h) < a + \varepsilon\,,$$

which contradicts the hypothesis

$$G\Big(\frac{x + y}{d(p, r)}\Big) \geqq T(a, b) = \text{Min}\,(a, b) + \varepsilon = a + \varepsilon\,.$$

THEOREM 5.3. *In a simple space, the means (if they exist), medians, modes (if unique) each form an M-space homothetic*[4] *to the original M-space.*

[4] Two M-spaces, M_1 and M_2, with distance functions d_1 and d_2, respectively are said to be homothetic if there exists a number $a > 0$ and a one-to-one mapping, f, from M_1 to M_2 such that, for every $p, q \in M_1$, $d_1(p, q) = ad_2(f(p), f(q))$.

Proof. Let $E[G] = \int_0^\infty x dG(x) = \mu$.

If $p = q$, then

$$E[F_{pq}] = \int_0^\infty x dH(x) = 0 \ .$$

If $p \neq q$, then

$$E[F_{pq}] = \int_0^\infty x dG(x/d(p, q))$$

which on substituting $t = x/d(p, q)$ becomes

$$E[F_{pq}] = d(p, q)\int_0^\infty t dG(t) = \mu d(p, q) \ .$$

The other cases are similar.

In Theorem 3.1 it was shown that every Wald space is a Menger space under $T =$ Product and in the corollary to Theorem 4.2 that every equilateral space is a Wald space. However, as the following examples show, there exist simple spaces which are not Wald spaces. Thus the converse of Theorem 3.1 is false. For, if in a Menger space IVm holds under $T =$ Min, it holds a fortiori under $T =$ Product.

THEOREM 5.4. *There exist simple spaces which are not Wald spaces.*

Proof. We give two counter-examples:

EXAMPLE 4.

$$F_{pq}(x) = 1 - e^{-x/d(p,q)} \ .$$

With $d(p, q) = R$, $d(q, r) = S$ and $d(p, r) = T$, one obtains (taking into account the fact that the lower limit of the convolution integral is 0 and the upper limit is x),

$$(6) \qquad [F_{pq} * F_{qr}](x) = \begin{cases} 1 - \dfrac{1}{R - S}(Re^{-x/R} - Se^{-x/S}) \ , & R \neq S \ , \\ 1 - \left(1 + \dfrac{x}{R}\right)e^{-x/R} \ , & R = S \ . \end{cases}$$

In order that the Wald inequality IVw be satisfied, we must have

$$(7) \qquad F_{pr}(x) \geqq [F_{pq} * F_{qr}](x) \ , \text{ for every } x \geqq 0 \ ,$$

Suppose $R \neq S$, say $R > S$. Then, keeping x fixed and applying the mean value theorem to the second term on the upper right-hand side of (6), we have

$$(8) \qquad [F_{pq} * F_{qr}](x) = 1 - \left(1 + \frac{x}{t}\right)e^{-x/t}\ , \text{ where } S < t < R\ .$$

Furthermore, if $R = S$, we observe, on comparing (8) with (6), that (8) holds with $t = R$. Thus, in both cases, in order that (7) hold, it is necessary that

$$\left(1 + \frac{x}{t}\right)e^{-x/t} \geqq e^{-x/T}\ , \text{ for all } x \geqq 0\ ;$$

that is,

$$1 + \frac{x}{t} \geqq e^{x(1/t - 1/T)}\ , \text{ for all } x \geqq 0\ .$$

This will be true if and only if $(1/t - 1/T) \leqq 0$, i.e., $T \leqq t$. In particular, therefore, it is necessary that $T \leqq R$. But this means that the side of the triangle pqr whose length is T certainly cannot be the longest side of that triangle. Thus we conclude: If $d(p, r) \geqq \text{Max}\,(d(p, q), d(q, r))$, then the Wald inequality will fail for sufficiently large x.

EXAMPLE 5.

$$F_{pq}(x) = \begin{cases} 0\ , & x \leqq 0\ , \\ x/d(p, q)\ , & 0 \leqq x \leqq d(p, q)\ , \\ 1\ , & x \geqq d(p, q)\ . \end{cases}$$

In this simple space, one can construct an example in which IVw holds for x sufficiently small and again for x sufficiently large, but fails in an intermediate range. For instance, if, $d(p, q) = 1$, $d(q, r) = 3$, and $d(p, r) = 3.75$, then for $2.50 < x < 3.68$, we have $[F_{pq} * F_{qr}](x) > F_{pr}(x)$.

6. Normal spaces. Statistical metric spaces also arise very naturally in the following manner: Let $p, q, \cdots$, be random variables on a common M-space E with distance function d. Then $d(p, q)$ is a random variable on the Cartesian product $E \times E$. Let F_{pq} be the distribution function of $d(p, q)$. Then the ordered pair $(S, \mathscr{F})$, where S is the set $\{p, q, \cdots\}$ and $\mathscr{F}$ the class of ordered pairs $\{(p, q), F_{pq}\}$ is an SM-space.

Particularly interesting SM-spaces of this type result when S is taken to be a set of mutually independent spherically-symmetric Gaussian random variables on an n-dimensional Euclidean space. We have investigated these at some length. However, reproducing the details here would take us too far afield. We shall therefore restrict ourselves to a brief presentation, without proof, of the main results.[5]

Let S_n^* be the set of all mutually independent, spherically-symmetric, n-dimensional Gaussian random variables, $p, q, \cdots$, etc., on a Euclidean n-space E^n. Let the n-dimensional mean of p (which is a point in E^n) be $\boldsymbol{m}_p$ and the common standard deviation of the one-dimensional marginal distributions of p be σ_p; and let the corresponding objects for q be $\boldsymbol{m}_q$ and σ_q, respectively. Then, upon setting

$$r = r(p, q) = d(\boldsymbol{m}_p, \boldsymbol{m}_q) ,$$

where d is the metric in E^n, and

$$\sigma = \sigma(p, q) = (\sigma_p^2 + \sigma_p^2)^{1/2} ,$$

the distribution function, F_{pq}, of $d(p, q)$ is given by:

$$F_{pq}(x) = R_n(\sigma, r; x) = \begin{cases} 0 , & x \leqq 0 , \\ \left(\dfrac{r}{\sigma}\right)^{1-n/2} e^{-r^2/2\sigma^2} \displaystyle\int_0^{x/\sigma} y^{n/2} I_{(n/2)-1}\left(\frac{r}{\sigma} y\right) e^{-y^2/2} dy , & x \geqq 0 , \end{cases} \tag{9}$$

if $\sigma > 0$, where I_ν is the modified Bessel function of order ν; and by

$$F_{pq}(x) = R_n(0, r; x) = H(x - r) ,$$

if $\sigma = 0$. The distribution function R_n has been found in different contexts, first by Bose [1][6], and more recently, by K. S. Miller, R. I. Bernstein and L. E. Blumenson [11].

An SM-space $(S_n, \mathscr{F})$ will be called an (n-dimensional) *normal space* if: (a) S_n is a subset of S_n^* having the property that, for every point $\boldsymbol{m}$ in E^n, there is at least one p in S_n whose expectation is $\boldsymbol{m}$; (b) For p, q in S_n, $\mathscr{F}(p, q) = F_{pq}$ is given by (9). A normal space is called *homogeneous* if $\sigma_p = \sigma_q$ for all p, q in S_n; otherwise, *inhomogeneous*.

Normal spaces have the following properties:

1. For all n, the means[7] of the F_{pq}'s form a metric space. This metric space is 'asymptotic in the large' to E^n in the following sense: If the distance between the means of p and q is r, then (for σ fixed) the mean of F_{pq} as a function of r is asymptotic to r for large r. In the 'small', the metric is definitely non-Euclidean. Furthermore, for $\sigma > 0$, there is a *positive minimum distance* between distinct points. That is to say, if $p \neq q$, and $r = d(\boldsymbol{m}_p, \boldsymbol{m}_q) > 0$, then the mean of F_{pq}, which is the expected value of the random variable $d(p, q)$, is greater than $\sqrt{2}\,\sigma[\Gamma(n/2 + 1/2)/\Gamma(n/2)]$. When $r = 0$ equality is attained.

[5] The detailed discussion will be the subject of another paper.

[6] See also Mahalanobis [**3**].

[7] These always exist; indeed, the F_{pq}'s in a normal space have moments of all orders.

2. All normal spaces are Menger spaces under the choice $T = T_w$.

3. A homogeneous normal space is a Menger space under the choice $T = \text{Max}(\text{Sum}-1, 0)$. For some triples of points, p_1, p_2, p_3, and for some numbers, x, y—subject to certain restrictions—IVm will hold under a stronger choice of T. However, other than T_w and $\text{Max}(\text{Sum}-1, 0)$, none of the T's listed in § 2 will hold universally. In particular, there are triples of points for which IVm does not hold for all x, y under $T = \text{Product}$. Consequently,

4. No normal space is a Wald space.

Whether IVm will hold in a normal space under a T that is weaker than Product, yet stronger than $\text{Max}(\text{Sum}-1, 0)$ is not known.

III. Topology, Convergence, Continuity

7. In the theory of metric spaces, the concept of a neighborhood can be introduced and defined with the aid of the distance function. A similar procedure applies in the theory of statistical metric spaces. In fact, neighborhoods in SM-spaces may be defined in several non-equivalent ways. Here we shall consider only one of these—the one which seems to be the strongest, in that its consequences most nearly resemble the classical results on M-spaces.

Definition 7.1. Let p be a point in the SM-space $(S, \mathscr{F})$. By an ε, λ-neighborhood of p, $\varepsilon > 0$, $\lambda > 0$, we mean the set of all points q in S for which $F_{pq}(\varepsilon) > 1 - \lambda$. We write:

$$N_p(\varepsilon, \lambda) = \{q; F_{pq}(\varepsilon) > 1 - \lambda\} .$$

The interpretation is: $N_p(\varepsilon, \lambda)$ is the set of all points q in S for which the probability of the distance from p to q being less than ε is greater than $1 - \lambda$. Observe that this neighborhood of a point in an SM-space depends on two parameters.

Theorem 7.1. *In a simple space, $N_p(\varepsilon, \lambda)$ is an ordinary spherical neighborhood of p in the generating M-space.*

Proof. For any p, q, we have

$$F_{pq}(\varepsilon) = G(\varepsilon/d(p, q)) ,$$

which will be greater than $1 - \lambda$ provided only that $d(p, q)$ is sufficiently small.

Lemma 7.1. *If $\varepsilon_1 \leqq \varepsilon_2$ and $\lambda_1 \leqq \lambda_2$, then $N_p(\varepsilon_1, \lambda_1) \subset N_p(\varepsilon_2, \lambda_2)$.*

Proof. Suppose $q \in N_p(\varepsilon_1, \lambda_1)$ so that $F_{pq}(\varepsilon_1) > 1 - \lambda_1$. Then, $F_{pq}(\varepsilon_2) \geqq F_{pq}(\varepsilon_1) > 1 - \lambda_1 \geqq 1 - \lambda_2$, whence, by definition, $q \in N_p(\varepsilon_2, \lambda_2)$.

THEOREM 7.2. *If $(S, \mathscr{F})$ is a Menger space and T is continuous then $(S, \mathscr{F})$ is a Hausdorff space in the topology induced by the family of ε, λ-neighborhoods $\{N_p\}$.*

Proof. We have to show that the following four properties are satisfied:

(A) For every p in S, there exists at least one neighborhood, N_p, of p; every neighborhood of p contains p.

(B) If N_p^1 and N_p^2 are neighborhoods of p, then there exists a neighborhood of p, N_p^3, such that $N_p^3 \subset N_p^1 \cap N_p^2$.

(C) If N_p is a neighborhood of p, and $q \in N_p$, then there exists a neighborhood of q, N_q, such that $N_q \subset N_p$.

(D) If $p \neq q$, then there exist disjoint neighborhoods, N_p and N_q, such that $p \in N_p$ and $q \in N_q$.

Proof of (*A*). For every $\varepsilon > 0$ and every $\lambda > 0$, $p \in N_p(\varepsilon, \lambda)$ since $F_{pp}(\varepsilon) = 1$ for any $\varepsilon > 0$.

Proof of (*B*). Let

$$N_p^1(\varepsilon_1, \lambda_1) = \{q;\, F_{pq}(\varepsilon_1) > 1 - \lambda_1\}$$

and

$$N_p^2(\varepsilon_2, \lambda_2) = \{q;\, F_{pq}(\varepsilon_2) > 1 - \lambda_2\}$$

be the given neighborhoods of p, and consider

$$N_p^3 = \{q;\, F_{pq}(\mathrm{Min}\,(\varepsilon_1, \varepsilon_2)) > 1 - \mathrm{Min}\,(\lambda_1, \lambda_2)\}\ .$$

Clearly $p \in N_p^3$; and since $\mathrm{Min}\,(\varepsilon_1, \varepsilon_2) \leqq \varepsilon_1$ and $\mathrm{Min}\,(\lambda_1, \lambda_2) \leqq \lambda_1$, by Lemma 7.1, $N_p^3 \subset N_p^1$. Similarly, $N_p^3 \subset N_p^2$, whence $N_p^3 \subset N_p^1 \cap N_p^2$.

Proof of (*C*). Let $N_p = \{r;\, F_{pr}(\varepsilon_1) > 1 - \lambda_1\}$ be the given neighborhood of p. Since $q \in N_p$,

$$F_{pq}(\varepsilon_1) > 1 - \lambda_1\ .$$

Now, F_{pq} is left-continuous at ε_1. Hence, there exists an $\varepsilon_0 < \varepsilon_1$ and a $\lambda_0 < \lambda_1$, such that

$$F_{pq}(\varepsilon_0) > 1 - \lambda_0 > 1 - \lambda_1\ .$$

Let $N_q = \{r;\, F_{qr}(\varepsilon_2) > 1 - \lambda_2\}$, where $0 < \varepsilon_2 < \varepsilon_1 - \varepsilon_0$, and λ_2 is chosen such that

$$T(1-\lambda_0, 1-\lambda_2) > 1-\lambda_1 .$$

Such a λ_2 exists since, by hypothesis, T is continuous, $T(a, 1) = a$, and $1-\lambda_0 > 1-\lambda_1$. Now suppose $s \in N_q$, so that

$$F_{qs}(\varepsilon_2) > 1-\lambda_2 .$$

Then

$$F_{ps}(\varepsilon_1) \geqq T(F_{pq}(\varepsilon_0), F_{qs}(\varepsilon_1-\varepsilon_0)) \geqq T(F_{pq}(\varepsilon_0), F_{qs}(\varepsilon_2))$$
$$\geqq T(1-\lambda_0, 1-\lambda_2) > 1-\lambda_1 .$$

But this means $s \in N_p$, whence $N_q \subset N_p$.

Proof of (D). Let $p \neq q$. Then there exists an $x > 0$ and an a, $0 \leqq a < 1$, such that, $F_{pq}(x) = a$. Let

$$N_p = \{r;\, F_{pr}(x/2) > b\} \text{ and } N_q = \{r;\, F_{qr}(x/2) > b\} ,$$

where b is chosen so that $0 < b < 1$ and $T(b, b) > a$. Such a number b exists, since T is continuous and $T(1, 1) = 1$. Now suppose there is a point s in $N_p \cap N_q$, so that $F_{ps}(x/2) > b$ and $F_{qs}(x/2) > b$. Then

$$a = F_{pq}(x) \geqq T(F_{ps}(x/2), F_{qs}(x/2)) \geqq T(b, b) > a ,$$

which is a contradiction. Thus N_p and N_q are disjoint.

It should be noted that the function T appeared only in the proofs of (C) and (D). Also, the ε, λ-neighborhoods defined at various stages in the proof may consist of only a single point. The situation here is analogous to the one that arises in connection with isolated points in M-spaces.

8. As is well known, in an M-space the notion of convergence of a sequence of points $\{p_n\}$ to a point p may be introduced with the aid of the neighborhood concept. There is also a well known theorem which states that the distance function d is continuous on S: that is to say, if $p_n \to p$ and $q_n \to q$, then $d(p_n, q_n) \to d(p, q)$. The proof of this theorem depends strongly on the triangle inequality.[8] Having defined neighborhoods in SM-spaces, it is natural to consider the above questions in this more general setting. In so doing, a significant difference arises, for there are now two distinct types of questions regarding convergence and continuity that must be considered: (a) Those relating to the distance function, $\mathscr{F}$, considered as a function on $S \times S$—either for a fixed value of x, or for a range of values; (*b*) Those relating to the individual distribution functions F_{pq}—either for a fixed pair of points (p, q) or for

[8] In fact, as Menger has shown [**4**, pp. 142-145], the triangle inequality is, in a certain sense, the simplest condition implying continuity that can be imposed on a semi-metric.

a set of pairs of points. As is to be expected, these questions are not independent.

DEFINITION 8.1. A sequence of points $\{p_n\}$ in an SM-space is said to converge to a point p in S (and we write $p_n \to p$) if and only if, for every $\varepsilon > 0$ and every $\lambda > 0$, there exists an integer $M_{\varepsilon,\lambda}$, such that $p_n \in N_p(\varepsilon, \lambda)$, i.e., $F_{pp_n}(\varepsilon) > 1 - \lambda$, whenever $n > M_{\varepsilon,\lambda}$.

LEMMA 8.1. *If $p_n \to p$, then $F_{pp_n} \to F_{pp} = H$, i.e., for every x, $F_{pp_n}(x) \to F_{pp}(x) = H(x)$, and conversely.*

Proof. (a) If $x > 0$, then for every $\lambda > 0$, there exists an integer $M_{x,\lambda}$ such that $F_{pp_n}(x) > 1 - \lambda$ whenever $n > M_{x,\lambda}$. But this means that $\lim_{n\to\infty} F_{pp_n}(x) = 1 = F_{pp}(x)$.

(b) If $x = 0$, then for every n, $F_{pp_n}(0) = 0$ and hence $\lim_{n\to\infty} F_{pp_n}(0) = 0 = F_{pp}(0)$.

The converse is immediate.

COROLLARY. *The convergence is uniform on any closed interval $[a, b]$ such that $a > 0$, i.e., $M_{x,\lambda}$ is independent of x for $a \leqq x \leqq b$.*

Proof. For any x, $a \leqq x \leqq b$, $F_{pp_n}(x) \geqq F_{pp_n}(a)$.

THEOREM 8.1. *If $(S, \mathscr{F})$ is a Menger space and T is continuous, then the statistical distance function, $\mathscr{F}$, is a lower semi-continuous function of points, i.e., for every fixed x, if $q_n \to q$ and $p_n \to p$, then,*

$$\liminf_{n\to\infty} F_{p_nq_n}(x) = F_{pq}(x) .$$

Proof. If $x = 0$, this is immediate, since for every n, $F_{p_nq_n}(0) = 0 = F_{pq}(0)$. Suppose then that $x > 0$, and let $\varepsilon > 0$ be given. Since F_{pq} is left-continuous at x, there is an h, $0 < 2h < x$, such that

$$F_{pq}(x) - F_{pq}(x - 2h) < \varepsilon/3 .$$

Set $F_{pq}(x - 2h) = a$. Since T is continuous, and $T(a, 1) = a$, there is a number t, $0 < t < 1$, such that

$$T(a, t) > a - \varepsilon/3$$

and

$$T(a - \varepsilon/3, t) > a - 2\varepsilon/3 .$$

Since $q_n \to q$ and $p_n \to p$, by Lemma 8.1 there exists an integer $M_{h,t}$ such that $F_{qq_n}(h) > t$ and $F_{pp_n}(h) > t$ whenever $n > M_{h,t}$. Now,

$$F_{p_nq_n}(x) \geqq T(F_{p_nq}(x-h), F_{qq_n}(h))$$

and

$$F_{p_nq}(x-h) \geqq T(F_{pq}(x-2h), F_{pp_n}(h)) .$$

Thus, on combining the various inequalities, we obtain

$$F_{p_nq}(x-h) \geqq T(a, t) > a - \varepsilon/3 ,$$

whence

$$F_{p_nq_n}(x) \geqq T(a-\varepsilon/3, t) > a - 2\varepsilon/3 > F_{pq}(x) - \varepsilon .$$

COROLLARY 1. *Let p be a fixed point and suppose $q_n \to q$. Then*

$$\liminf_{n\to\infty} F_{pq_n}(x) = F_{pq}(x) .$$

COROLLARY 2. *If $(S, \mathscr{F})$ is a Wald space, then $\mathscr{F}$ is a lower semi-continuous function of points.*

Proof. By Theorem 3.1, in a Wald space, the Menger inequality holds under $T =$ Product, which is continuous.

THEOREM 8.2. *Let $(S, \mathscr{F})$ be a Menger space. Suppose that T is continuous*[9] *and at least as strong as* Max (Sum-1, 0). *Suppose further that $p_n \to p$, $q_n \to q$, and that F_{pq} is continuous at x. Then $F_{p_nq_n}(x) \to F_{pq}(x)$, i.e., the distance function $\mathscr{F}$ is a continuous function of points at (p, q, x); or, expressed in another way, the sequence of functions $\{F_{p_nq_n}\}$ converges weakly to F_{pq}.*

Proof. In view of Theorem 8.1, it suffices to prove upper semi-continuity, i.e., that for $\varepsilon > 0$ and n sufficiently large,

$$F_{p_nq_n}(x) < F_{pq}(x) + \varepsilon . \tag{1}$$

Suppose then that $\varepsilon > 0$ is given. Since F_{pq} is continuous, and in particular therefore right-continuous at x, there exists an $h > 0$, such that

$$F_{pq}(x+2h) - F_{pq}(x) < \varepsilon/3 . \tag{2}$$

By Lemma 8.1, there is an integer M such that the conditions,

$$F_{pp_n}(h) > 1 - \varepsilon/3 , \tag{3}$$

$$F_{qq_n}(h) > 1 - \varepsilon/3 , \tag{4}$$

are simultaneously satisfied for all $n > M$. And from IVm, we have

[9] Here, as well as in Theorems 7.2 and 8.1, the continuity of T may be replaced by the weaker condition, $\lim_{x\to 1} T(a, x) = a$.

$$(5) \qquad F_{pq}(x+2h) \geqq T(F_{pq_n}(x+h),\ F_{qq_n}(h))$$

and

$$(6) \qquad F_{pq_n}(x+h) \geqq T(F_{p_nq_n}(x),\ F_{pp_n}(h)) .$$

Now, by hypothesis, T is at least as strong as Max (Sum$-1, 0$), so that on combining (3) and (6) we obtain

$$(7) \qquad F_{pq_n}(x+h) \geqq F_{p_nq_n}(x) + F_{pp_n}(h) - 1 > F_{p_nq_n}(x) - \varepsilon/3 ;$$

and, on combining (7) with (4) and (5), we obtain

$$F_{pq}(x+2h) \geqq F_{pq_n}(x+h) + F_{qq_n}(h) - 1 > F_{p_nq_n}(x) - 2\varepsilon/3 .$$

Finally, combining (8) with (2) yields (1) and completes the proof.

COROLLARY 1. *Under the hypotheses of Theorem* 8.2, *if* $q_n \to q$, *then* $F_{pq_n}(x) \to F_{pq}(x)$.

COROLLARY 2. *If the functions* F_{pq} *are each continuous functions for all* p, q *in* S, *then* $\mathscr{F}$ *is a continuous function of points.*

COROLLARY 3. *If* $(S, \mathscr{F})$ *is a Wald space and if the functions* F_{pq} *are each continuous, then* $\mathscr{F}$ *is a continuous function of points.*

BIBLIOGRAPHY

1. R. Bose, *On the exact distribution and moment-coefficients of the D^2-statistics*, Science and Culture, **1** (1935), p. 205.
2. M. Fréchet, *Sur quelques points du calcul fonctionnel*, Rendiconti de Circolo Matematico di Palermo, **22** (1906), 1-74.
3. P. C. Mahalanobis, *On the generalized distance in statistics*, Proc. Nat. Inst. of Sci. India, **2** (1936), 49-55.
4. K. Menger, *Untersuchungen über allgemeine Metrik*, Mathematische Annalen, **100** (1928), 75-163.
5. ———, *Statistical metrics*, Proc. Nat. Acad. of Sci., U. S. A., **28** (1942), 535-537.
6. ———, *Probabilistic theories of relations*, Ibid., **37** (1951), 178-180.
7. ———, *Probabilistic geometry*, Ibid., **37** (1951), 226-229.
8. ———, *Ensembles flous et fonctions aléatoires*, Comptes Rendus Acad. Sci. Paris, **232** (1951), 2001-2003.
9. ———, *Géométrie générale*, Mémorial des Sciences Mathématiques, no. 124, Paris 1954.
10. K. Menger, B. Schweizer and A. Sklar, *On probabilistic metrics and numerical metrics with probability 1*, Czechoslovak Math. Journal, **9** (84), (1959), 459-466.
11. K. S. Miller, R. I. Bernstein and L. E. Blumenson, *Generalized Rayleigh processes*, Quarterly of Applied Math., **16** (1958), 137-145.
12. B. Schweizer and A. Sklar, *Espaces métriques aléatoires*, Comptes Rendus Acad. Sci. Paris, **247** (1958), 2092-2094.
13. A. Špaček, *Note on K. Menger's probabilistic geometry*, Czechoslovak Math. Journal **6** (81), (1956), 72-74.

14. A. Wald, *On a statistical generalization of metric spaces*, Proc. Nat. Acad. Sci., U. S. A., **29** (1943), 196-197.

15. ———, *On a statistical generalization of metric spaces*, Reports of a Mathematical Colloquium, Second Series, 5-6, (1944), 76-79. Reprinted in: Abraham Wald: Selected Papers in Statistics and Probability, New York 1955, pp. 413-416.

UNIVERSITY OF CALIFORNIA, LOS ANGELES
ILLINOIS INSTITUTE OF TECHNOLOGY

ON DERIVATIONS IN DIVISION RINGS

MORRIS WEISFELD

We are concerned with studying division rings in which Lie rings of derivations are acting. The results include the determination of dimension over the constant subring, an outer Galois theory, and miscellaneous results on inner automorphisms and powers of derivations.

Let A be a ring with an identity 1 and B be a subring of A containing 1.

1. The mappings $R_x : y \rightarrow yx$, $L_x : y \rightarrow xy$, $I_x : y \rightarrow x^{-1}yx$, $x \in A$ are called right multiplications, left multiplications and inner automorphisms respectively. For any subset N of A, $R_N = \{R_x \mid x \in N\}$, $L_N = \{L_x \mid x \in N\}$, $I_x = \{I_x \mid x \in N\}$. D is a derivation of B into A if and only if $(x + y)D = xD + yD$ and $(xy)D = xDy + xyD$ for all $x, y \in A$. The set of all such mappings is denoted by $\mathrm{Der}(B, A)$. If $B = A$, D is called a derivation in A and the set of these denoted by $\mathrm{Der}(A)$. If $D_1, \cdots, D_s \in \mathrm{Der}(A)$, we have for all $x \in A$

$$(1) \qquad R_x D_1^{k_1} \cdots D_s^{k_s} = \sum \left\{ \binom{k_1}{i_1} \cdots \binom{k_s}{s_s} D_1^{k_1 - i_1} \cdots D_s^{k_s - i_s} R_{xD_1^{i} \cdots D_s^{i_s}} \mid 0 \leq i_j \leq k_j,\ j = 1, \cdots, s \right\}.$$

For all $D, D' \in \mathrm{Der}(A)$, $[DD'] = DD' - D'D \in \mathrm{Der}(A)$ and, if A has prime characteristic p, $D^p \in \mathrm{Der}(A)$. $\{x \mid xy - yx = 0 \text{ for all } y \in B; x \in A\}$ is called the centralizer of B in A. If c belongs to the centralizer of B in A, $DR_c \in \mathrm{Der}(B, A)$. The centralizer of A in A is called the center of A. Let C be the center of A. $\mathscr{D}$ is a Lie ring (Lie ring over C) of derivations in A if and only if $\mathscr{D} \subseteq \mathrm{Der}(A)$ and for all $D, D' \in \mathscr{D}$, $D - D' \in \mathscr{D}$, $[DD'] \in \mathscr{D}$ $(DR_c \in \mathscr{D})$. If A has prime characteristic p, $\mathscr{D}$ is restricted if, in addition, $D^p \in \mathscr{D}$.

For $x \in A$, the mapping $I'_x : y \rightarrow yx - xy$ is a derivation called an inner derivation. For $N \subseteq A$, $I'_N = \{I'_x \mid x \in N\}$. The elements of $\mathrm{Der}(A)$ not in I'_A are called outer derivations. Lie ideals are defined in the usual way for Lie rings, restricted or not, over C or not. The inner derivations in $\mathscr{D}$ form a Lie ideal in $\mathscr{D}$.

Let T be a subset of $\mathrm{Der}(B, A)$. The set of $x \in B$ such that $xD = 0$ for all $D \in T$ is a subring of B which we call the subring of T-constants and which we denote by $B(T)$. If $x \in B(T)$ and x has a multiplicative inverse x^{-1} in B, then $x^{-1} \in B(T)$. The set of derivations D in A such

Received April 15, 1959. This is a revised version of part of a dissertation submitted in candidacy for a Ph. D. degree to Yale University. The author wishes to thank Professor N. Jacobson for his advice and encouragement. An abstract was presented to the American Mathematical Society (Bull. A. M. S., Vol. 60, 1954, p. 142).

that $B \subseteq A(D)$ is a (restricted) Lie subring over C of $\mathrm{Der}(A)$ which we denote by $\mathscr{D}(B)$.

If $T_1 \subseteq T_2 \subseteq \mathrm{Der}(A)$, then $A(T_1) \supseteq A(T_2)$; and, if B_1 and B_2 are subrings of A containing 1, and $B_1 \subseteq B_2$, then $\mathscr{D}(B_1) \supseteq \mathscr{D}(B_2)$. We have the following relations:

$$\mathscr{D}' \subseteq \mathscr{D}(A(\mathscr{D}')) \tag{2}$$

for all (restricted) Lie subrings over C of $\mathrm{Der}(A)$

$$B \subseteq A(\mathscr{D}(B)) \tag{3}$$

(for all subrings B of A containing 1. These give

$$A(\mathscr{D}') = A(\mathscr{D}(A(\mathscr{D}'))) \tag{4}$$

for all (restricted) Lie subrings over C, $\mathscr{D}'$, of $\mathrm{Der}(A)$

$$\mathscr{D}(B) = \mathscr{D}(A(\mathscr{D}(B))) \tag{5}$$

for all subrings of A containing 1. Thus, $B = A(\mathscr{D}(B))$ if and only if $B = A(\mathscr{D})$ for some (restricted) Lie subring over C of $\mathrm{Der}(A)$, and $\mathscr{D}' = \mathscr{D}(A(\mathscr{D}'))$ if and only if $\mathscr{D}' = \mathscr{D}(B)$ for some subring B of A containing 1.

Let $\mathscr{D}$ be a (restricted) Lie subring over C of $\mathrm{Der}(A)$. Define

$$\sum(\mathscr{D}) = \{x \mid x \in A \text{ and } I'_x \in \mathscr{D}\} .$$

Regard the ring A as a (restricted) Lie algebra over C under the compositions $(x, y) \to xy - yx$ ($x \to x^p$ if A has prime characteristic p). $\Sigma(\mathscr{D})$ is then a (restricted) Lie subalgebra of A. $\Sigma(\mathscr{D})$ is invariant under $\mathscr{D}$; that is $xD \in \Sigma(\mathscr{D})$ for $x \in \Sigma(\mathscr{D})$ and $D \in \mathscr{D}$. If $\mathscr{D} = \mathscr{D}(B)$ where B is a subring of A containing 1, then $\Sigma(\mathscr{D})$ is closed with respect to ordinary multiplication and taking multiplicative inverses. This leads us to make the following definition: We say $\mathscr{D}$ is a (restricted) N-Lie subring over C or $\mathrm{Der}(A)$ if and only if $\Sigma(\mathscr{D})$ is a subring of A closed with respect to taking multiplicative inverses and invariant under $\mathscr{D}$. In this case $\Sigma(\mathscr{D})$ over C is called the associated algebra of $\mathscr{D}$. If A is a division ring, $\Sigma(\mathscr{D})$ is a division algebra over C.

Let Δ be a division ring, Φ be its center and E a division subring of Δ. The additive group of homomorphisms of $(\mathrm{E}, +)$ into $(\Delta, +)$, $\mathrm{Hom}(\mathrm{E}, +; \Delta, +)$ is an $(R_{\mathrm{E}}, R_\Delta)$-space; that is, $\mathrm{Hom}(\mathrm{E}, +; \Delta; +)$ is a vector space over R_{E} and a right vector space over R_Δ such that

$$(R_xT)R_y = R_x(TR_y) \tag{7}$$

for all $x \in \mathrm{E}, y \in \Delta$ and $T \in \mathrm{Hom}(\mathrm{E}, +; \Delta, +)$. $(R_{\mathrm{E}}, R_\Delta)$-subspaces are defined in the usual way. If $\mathrm{E} = \Delta$, we write $\mathrm{End}(\Delta, +)$. If S is a set of endomorphisms of Δ, End $\{S\}$ is the ring of endomorphisms generated by S.

Let Γ be a division subring of Δ. $L_\Gamma(E, \Delta)$ denotes the subgroup of $\mathrm{Hom}(E, +; \Delta, +)$ of homomorphisms of the vector space E over Γ into the vector space Δ over Γ. If $E = \Delta$, we write $L_\Gamma(\Delta)$.

Topologize $\mathrm{Hom}(E, + \Delta, +)$ as follows: The sets

$$\{g \mid xg = xf;\, g \in \mathrm{Hom}(E, +; \Delta, +)\}$$

where $x \in E$ is a subbase of neighborhoods of $f \in \mathrm{Hom}(E, +; \Delta, +)$. One verifies that $L_\Gamma(\Delta)$ is a closed subring containing R_Δ. That the foregoing properties characterize $L_\Gamma(\Delta)$ is a consequence of the Jacobson-Bourbaki theorem: if we associate with a closed subring B of $\mathrm{End}(\Delta, +)$ which contains R_Δ, the set Γ of $x \in \Delta$ such that L_x commutes with the elements of B, then $B = L_\Gamma(\Delta)$, and, in fact, the mapping $\Gamma \to L_\Gamma(\Delta)$ is a lattice anti-isomorphism of the set of division subrings of Δ onto the set of closed subrings of $\mathrm{End}(\Delta, +)$ containing R_Δ.

If A is a vector space and right vector space over P, $[A:P]_L$ denotes its dimension and $[A:P]_R$ its right a dimensional over p. We note the following $[E:\Gamma]_L$ is finite if and only if $[L_\Gamma(E, \Delta): R_\Delta]_R$ is finite and when both are finite, they are equal. If B is a subring of $\mathrm{End}(\Delta, +)$ containing R_Δ and $[B:R_\Delta]_R$ is finite, then B is a closed subring.

Again Δ is a division ring, Φ is its center, E is a division subring and M is the centralizer of E in Δ. If $T \in \mathrm{End}(\Delta, +)$, then T^* denotes the restriction of T to E. If $D_1, \cdots, D_s$ are elements of $L_\Gamma(E, \Delta)$ not in the algebra generated by L_Δ and R_Δ and σ, μ are not zero, the degree of the endomorphism

$$D_1^{k_1} \cdots D_s^{k_s} L_\sigma R_\mu, \quad k_j \geq 0, \quad D_j^0 = I_1, \quad j = 1, \cdots, s, \tag{8}$$

is $k_1 + \cdots + k_s$. The weight of a sum of endomorphisms of the form (8) is the largest degree for which a term with that degree appears non-trivially. If all the terms appearing non-trivially have equal degree h, we say the endomorphism is homogeneous (of weight h). Any endomorphism is a sum of homogeneous endomorphisms. Suppose $D_1, \cdots, D_s \in L_\Gamma(\Delta)$ and are derivations of E into Δ and $\sigma \in \Delta$. $D_1^{k_1} \cdots D_s^{k_s} L_\sigma$ is called an admissible endomorphism if and only if (1), restricted to E and multiplied by L_σ, holds, any term appearing in (1) is admissible, and, if Δ has prime characteristic p, $k_j < p$.

LEMMA 1. *Let $\mathfrak{M}$ be an (R_E, R_Δ)-subspace of* $\mathrm{Hom}(E, +; \Delta, +)$, $\sigma_1, \cdots, \sigma_t$ *be elements of Δ, and $D_1, \cdots, D_s$ be derivations of* E *into Δ belonging $L_\Gamma(\Delta)$. Suppose no right linear combination of $L_{\sigma_1}^*, \cdots, L_{\sigma_t}^*$ with coefficients in R_M is zero, and no right linear combination of $L_{\sigma_{u+1}}^*, \cdots, L_{\sigma_t}^*$ with coefficients in R_M belongs to $\mathfrak{M}$. Suppose $D_1^*, \cdots, D_s^*$ are right* *linearly independent over R_Δ modulo $I_\Delta'^*$ and no linear combination of $D_{r+1}^*, \cdots, D_s^*$ with coefficients $R_M + L_{\sigma_1}R_M + \cdots + L_{\sigma_u}R_M$ belongs to $L_\Delta^* + \mathfrak{M}$. Then the set of non-zero admissible endomorphisms by $D_1, \cdots, D_s$ and $\sigma_1, \cdots, \sigma_t$ such that some k_j with $j > r$ is not zero or L_{σ_j} appears with $j > u$ is right linearly independent over R_Δ modulo $\mathfrak{M}$.*

Proof. Suppose we had a non-trivial linear relation. Let F be a linear relation of lowest weight q and shortest length in q. Suppose $q = 0$. Then $F = L_{\sigma_{j_1}}^* R_{\mu_1} + \cdots + L_{\sigma_{j_n}}^* R_{\mu_n} \in \mathfrak{M}$, $j_1, \cdots, j_n > u$, $\mu_i \in \Delta$, $\mu_1 = 1, n > 1$. For all $x \in \mathrm{E}$

$$R_x F - F R_x = L_{\sigma_{j_2}}^* R_{(x\mu_2 - \mu_2 x)} + \cdots + L_{\sigma_{j_n}}^* R_{(x\mu_n - \mu_n x)} \,.$$

This gives a shorter non-trivial relation, or $\mu_1, \cdots, \mu_n \in M$, contradicting our hypothesis. Note that for $\mu_1, \cdots, \mu_n \in M$

$$\begin{aligned} L_{\sigma_{j_1}}^* R_{\mu_1} + \cdots + L_{\sigma_{j_n}}^* R_{\mu_n} &= (R_{\mu_1} L_{\sigma_{j_1}})^* + \cdots + (R_{\mu_n} L_{\sigma_{j_n}})^* \\ &= (L_{\mu_1} L_{\sigma_{j_1}} + \cdots + L_{\mu_n} L_{\sigma_{j_n}})^* = L^*{}_{\mu_1\sigma_{j_1} + \cdots + \mu_n\sigma_{j_n}} \,. \end{aligned}$$

Suppose $q \geq 1$. Write $F = F_q + F_{q-1} + \cdots + F_0$ where the F_j are homogeneous. Let s_1 be the largest element among $1, \cdots, s$ such that a term in F_q has $k_{s_1} > 0$. Make its coefficient R_μ equal to 1. Form

$$\mathfrak{M} \ni R_x F - F R_x = G_q(x) + G_{q-1}(x) + \cdots + G_0(x) \,.$$

This will have lower weight or shorter length in q. The coefficients of terms in $G_q(x)$ have the form $R_{x\mu - \mu x}$. If these coefficients are zero, then the μ belong to M. Since we have shorter length in q, the coefficients in $G_q(x)$ must be 0. The coefficients in $G_{q-1}(x)$ have the form $R_{xD_{n_1}R_{\rho_1} + \cdots + xD_{n_k}R_{\rho_k} + xI_\lambda'}$, where the ρ_j's belong to M. These coefficients being zero for all $x \in \mathrm{E}$ would contradict our hypothesis. If $q > 1$, a term in F_q has the factor $D_j D_{s_1}$ and so G_{q-1} has a term with factor D_{s_1} and, if x is chosen to make its coefficient non-zero, we would contradict the choice of F. Hence the only possibility left is $q = 1$. Hence F has the form

$$\begin{aligned} D_{s_1}^* L_{\rho_{j_0}} + D_{n_1}^* L_{\rho_{j_1}} R_{\mu_1} + \cdots + D_{n_k}^* L_{\sigma_{j_k}} R_{\mu_k} + D_{n_{k+1}}^* R_{\mu_{k+1}} + \cdots \\ + D_{n_m}^* R_{\mu_m} + L_{\sigma_{u+1}}^* R_{\rho_{u+1}} + \cdots + L_{\sigma_t}^* R_{\rho_t} \,. \end{aligned}$$

Forming $R_x F - F R_x$ would yield non-trivial relations unless all right multiplications appearing belong to R_M and $j_1, \cdots j_k \leq u$. Hence F could be written as a linear combination of $D_{r+1}^*, \cdots, D_s^*$ with coefficients in $R_M + R_M L_{\sigma_1} + \cdots + R_M L_{\sigma_u}$ plus an element of $L_\Delta^* + \mathfrak{M}$ and this contradicts our hypothesis. Hence our assertion is true.

Suppose Δ has characteristic 0 and D is an outer derivation in Δ. Then the powers $D^k, k \geq 1$ of D are right linearly independent over R_Δ

as just shown. If $\Gamma = \Delta(D)$, $[L_\Gamma(\Delta) : R_\Delta] = \infty$ and hence $[\Delta : \Gamma]_L = \infty$. Thus if Δ has an outer derivation D and $[\Delta : \Gamma]_L < \infty$ where $\Gamma = \Delta(D)$, then Δ must have prime characteristic p.

Suppose Δ has prime characteristic p and Γ is a division subring such that $\mathscr{D}' = \mathscr{D}(\Gamma)$. $\mathscr{D}'$ is a restricted N-Lie ring over Φ of derivations in Δ. Suppose $\mathscr{D}'$ is infinite dimensional over Φ. Note that $I'_{\sigma_1}, \cdots, I'_{\sigma_t}$ are right linearly independent over R_Φ if and only if $1, \sigma_1, \cdots, \sigma_t$ are linearly independent over Φ. For if $\Sigma\varphi_i\sigma_i + \varphi_0 = 0$, then $\Sigma I'_{\sigma_i}R_{\varphi_i} = I'_{\Sigma\varphi_i\sigma_i+\varphi_0} = 0$. If $\Sigma I'_{\sigma_i}R_{\varphi_i} = 0$, then $0 = \Sigma(R_{\sigma_i} - L_{\sigma_i})R_{\varphi_i} = R_{\Sigma\varphi_i\sigma_i} - \Sigma L_{\sigma_i}R_{\varphi_i}$. Applying Lemma 1 yields the result. Thus $\mathscr{D}'$ has either infinitely many outer derivations right linearly independent over Φ modulo I'_Δ or $[\Sigma(\mathscr{D}') : \Phi] = \infty$. In either case $[L_\Gamma(\Delta) : R_\Delta]_R$ will, by Lemma 1, be infinite dimensional and so will $[\Delta : \Gamma]_L$.

THEOREM 1. *Let Δ be a division ring having prime characteristic p, Φ be its center, $\mathscr{D}$ be a finite dimensional restricted N-Lie ring over Φ of derivations in Δ and $\Gamma = \Delta(\mathscr{D})$. Then if $D_1, \cdots, D_m$ is a complete set of representatives of a basis for the right vector space $\mathscr{D} - I'_{\Sigma(\mathscr{D})}$ over Φ, and $\sigma_1, \cdots, \sigma_q$ is a basis for $\Sigma(\mathscr{D})$ over Φ, then*

$$(9) \qquad \{D_1^{k_1} \cdots D_m^{k_m} L_{\sigma_j} \mid k_i = 0, 1, \cdots, p-1, \quad D_i^0 = I_1, \quad i = 1, \cdots, m, \quad j = 1, \cdots, q\}$$

is a basis for the right vector space End $\{\mathscr{D}, R_\Delta\}$ *over R_Δ. Moreover,* End $\{\mathscr{D}, R_\Delta\} = L_\Gamma(\Delta)$, $[L_\Gamma(\Delta) : R_\Delta]_R = p^m q = [\Delta : \Gamma]_L$ *and* $\mathscr{D} = \mathscr{D}(\Gamma)$.

Proof. Consider the set A of right linear combinations of elements of (9) with coefficients in R_Δ. Then clearly End $\{\mathscr{D}, R_\Delta\} \supseteq A$. Because $1 = \Sigma\sigma_j\lambda_j$, $\lambda_j \in \Phi$, $I_1 = L_{\sigma_j}R_{\lambda_j} \in A$ and, hence, $A \supseteq R_\Delta$ Since any inner derivation belonging to $\mathscr{D}$ can be written as $\Sigma(R_{\sigma_j\varphi_j} - L_{\sigma_j}R_{\varphi_j})$, where $\varphi_j \in \Phi$, $A \supseteq I'_{\Sigma(\mathscr{D})}$. Since any $D \in \mathscr{D}$ can be written as $\Sigma D_iR_{\lambda_i} + I'_\sigma$, where $\lambda_L \in \Phi$ and $\sigma \in \Sigma(\mathscr{D})$, $A \supseteq \mathscr{D}$. We have $R_xD_i = D_iR_x + R_{xD_i}$ for $x \in \Delta$, $D_i^p = \Sigma D_iR_{\lambda_i} + I'_\sigma$ with $\lambda_i \in \Phi$ and $\sigma \in \Sigma(\mathscr{D})$, $D_iD_j = D_jD_i + D$ with $D \in \mathscr{D}$, and $L_\sigma D_i = D_iL_\sigma + L_{\sigma D_i}$, for $\sigma \in \Sigma(\mathscr{D})$ and, hence, $\sigma D_i \in \Sigma(\mathscr{D})$. Also $L_\sigma L_\tau = L_{\sigma\tau} \in A$, for $\sigma, \tau \in \Sigma(\mathscr{D})$ and, hence $\sigma\tau \in \Sigma(\mathscr{D})$. Thus A is a ring and so $A =$ End $\{\mathscr{D}, R_\Delta\}$. The elements of (9) generate A and are right linearly independent by Lemma 1 and so they constitute a basis for A over R_Δ. Thus [End $\{\mathscr{D}, R_\Delta\} : R_\Delta]_R = p^m q$ and so is a closed subring of End$(\Delta, +)$ containing R_Δ. By the Jacobson-Bourbaki theorem we have End $\{\mathscr{D}, R_\Delta\} = L_\Gamma(\Delta)$ and $[\Delta : \Gamma]_L = [L_\Gamma(\Delta) : R_\Delta]_R = p^m q$. Suppose $D \in \mathscr{D}(\Gamma)$. Then $D \in L_\Gamma(\Delta)$. Hence, by Lemma 1, $D = D_1\varphi_1 + \cdots + D_m\varphi_m + I'_\tau$, $\varphi_1, \cdots, \varphi_m \in \Phi$. $I'_\tau = R_\tau - L_\tau \in \mathscr{D}(\Gamma)$ and so another application of Lemma 1 yields $\tau = \sigma_1\lambda_1 + \cdots + \sigma_q\lambda_q$, $\lambda_1, \cdots, \lambda_q \in \Phi$, so that $D \in \mathscr{D}$.

We remark that because of the symmetry in the above situation, we also have $[\Delta : \Gamma]_R = [\Delta : \Gamma]_L$.

LEMMA 2. *Let $\mathscr{D}$ be a* (restricted if Δ has prime characteristic p) *Lie ring over Φ of derivations in Δ and $\Gamma = \Delta(\mathscr{D})$. Suppose $\mathscr{D}$ contains all inner derivations belonging to $\mathscr{D}(\Gamma)$. Let* E *be a division subring of Δ containing Γ and* $[\mathrm{E} : \Gamma]_L < \infty$. *Let* M *be the centralizer of* E *in Δ. If D^* is a derivation of* E *into Δ and $\Gamma \subseteq \Delta(D^*)$, then D^* can be extended to a derivation in the centralizer of* M.

Proof. $D^* \in L_\Gamma(\mathrm{E}, \Delta)$; hence D^* can be extended to an element of $L_\Gamma(\Delta)$. The proof of Theorem 1 shows that $L_\Gamma(\Delta)$ is the closure of $\mathrm{End}\{\mathscr{D}, R_\Delta\}$ in $\mathrm{End}(\Delta, +)$. Since $[\mathrm{E}:\Gamma]_L < \infty$, there is an $F \in \mathrm{End}\{\mathscr{D}, R_\Delta\}$ such that $D^* = F^*$, the asterick denoting restriction to E. We have

$$D^* = F^* = \Sigma(D_i^{k_1} \cdots D_s^{k_s} L_{\sigma_j} R \mu_{k_1 \cdots k_s j})^*$$

where $D_1, \cdots, D_s, L_{\sigma_1}, \cdots, L_{\sigma_t}$ satisfy the hypotheses of Lemma 1 with $\mathfrak{M} = \{0\}$. Hence

$$D^* = D_{j_1}^* R_{\mu_1} + \cdots + D_{j_k}^* R_{\mu_k} + I_\tau'^*,\ \mu_1, \cdots, \mu_k \in M$$

I_τ' leaves the elements of Γ fixed so that the right hand side is a derivation in the centralizer of M and belongs to $\mathscr{D}$.

In particular, if the centralizer of Γ is Φ; that is, every non-zero derivation in $\mathscr{D}(\Gamma)$ is outer, then every derivation of E into Δ can be extended to a derivation in Δ.

Henceforth, we restrict ourselves to Δ having prime characteristic p.

LEMMA 3. *Let $\mathscr{D}$ be a finite-dimensional restricted N-Lie ring over Φ of derivations in Δ and $\Gamma = \Delta(\mathscr{D})$. If B is a subring of $L_\Gamma(\Delta)$ containing R_Δ and $\mathscr{D}' = B \cap \mathscr{D}$, then $\mathscr{D}'$ is a restricted N-Lie subring over Φ of $\mathscr{D}$. If $\mathscr{D}$ consists only of outer derivations, then* $B = \mathrm{End}\{\mathscr{D}', R_\Delta\}$.

Proof. Clearly $\mathscr{D}'$ is a finite-dimensional restricted Lie ring over Φ of derivations in Δ. Now $\Sigma(\mathscr{D}')$ is contained in $\Sigma(\mathscr{D})$ and $[\Sigma(\mathscr{D}) : \Phi] < \infty$. Hence $[\Sigma(\mathscr{D}') : \Phi] < \infty$. Since $1 \in \Phi \subseteq \Sigma(\mathscr{D}')$ and $\Sigma(\mathscr{D}')$ has no zero divisors, it is a division ring provided that it is a ring. Let $\sigma_1, \sigma_2 \in \Sigma(\mathscr{D}')$; that is, $I_{\sigma_1}', I_{\sigma_2}' \in \mathscr{D}'$. Since B contains R_Δ and I_{σ_1}', it contains L_{σ_1}. Since B contains R_Δ and I_{σ_2}', it contains $I_{\sigma_1\sigma_2}' = I_{\sigma_1}' R_{\sigma_2} + L_{\sigma_1} I_{\sigma_2}'$. Hence, $I_{\sigma_1\sigma_2}' \in \mathscr{D}' = \mathscr{D} \cap B$ which implies $\sigma_1\sigma_2 \in \Sigma(\mathscr{D}')$. If $\sigma \in \mathscr{D}'$ and $D \in \mathscr{D}'$, $[I_\sigma', D] = I_{\sigma D}' \in \mathscr{D}'$ and $\Sigma(\mathscr{D}')$ is invariant under $\mathscr{D}'$. Now let $D_1, \cdots, D_s$ be a basis for $\mathscr{D}$ over Φ such that $D_1, \cdots, D_r$ is a basis for $\mathscr{D}'$ over Φ, $r \leq s$. By Theorem 1, $B \supseteq \mathrm{End}\{\mathscr{D}, R_\Delta\}$ and we can write

$$b = \Sigma D_1^{k_1} \cdots D_s^{k_s} R_{\mu_{k_1} \cdots k_s}$$

Since we have assumed $\Sigma(\mathscr{D}) = \Phi$. Applying Lemma 1 with $B = \mathfrak{M}$, we note that no D_j with $j > r$ can appear in this expression. Hence, $B \subseteq \operatorname{End}\{\mathscr{D}', R_\Delta\}$. Clearly, $\operatorname{End}\{\mathscr{D}', R_\Delta\} \subseteq B$ and these facts give the desired conclusion.

THEOREM 2. *Let $\mathscr{D}$ be a finite dimensional restricted Lie ring over Φ of outer derivations in Δ, $\Gamma = \Delta(\mathscr{D})$, and $\mathscr{D} = \mathscr{D}(\Gamma)$. To each restricted Lie subring over Φ, $\mathscr{D}'$, of $\mathscr{D}$ assign the division ring $\Delta(\mathscr{D}')$. To each division subring* E *of Δ containing Γ assign the restricted Lie ring $\mathscr{D}$*(E) *over Φ of derivations in Δ. Then the correspondences $\mathscr{D}' \to \Delta(\mathscr{D}')$ and* $\mathrm{E} \to \mathscr{D}(\mathrm{E})$ *are inverses of each other; that is, $\mathscr{D}' = \mathscr{D}(\Delta(\mathscr{D}'))$ and* $\mathrm{E} = \Delta(\mathscr{D}(\mathrm{E}))$. *Moreover, $\mathscr{D}'$ is a Lie ideal over Φ of $\mathscr{D}$ if and only if* $\mathrm{E} = \Delta(\mathscr{D}')$ *is invariant under $\mathscr{D}$, and, in this case, the restricted Lie ring over Φ of derivations in* E *leaving the elements of Γ fixed is isomorphic to $\mathscr{D}/\mathscr{D}'$.*

Proof. Let $\mathscr{D}'$ be a restricted Lie subring over Φ of $\mathscr{D}$. $\mathscr{D}'$ satisfies the conditions of Theorem 1, for, in this case, $\Sigma(\mathscr{D}') = \Phi$, and thus $\mathscr{D}(\Delta(\mathscr{D}')) = \mathscr{D}'$. Next let E be a division subring of Δ containing Γ and $B = L_{\mathrm{E}}(\Delta)$. Then B is a subring of $L_\Gamma(\Delta)$ containing R_Δ. By Lemma 3, $B = \operatorname{End}\{\mathscr{D}', R_\Delta\}$ where $\mathscr{D}' = B \cap \mathscr{D}$. Clearly $\mathscr{D}' = \mathscr{D}(\mathrm{E})$. On the other hand, since $B = \operatorname{End}\{\mathscr{D}', R_\Delta\} = L_{\mathrm{E}}(\Delta)$, $\mathrm{E} = \Delta(\mathscr{D}')$. Thus, $\mathrm{E} = \Delta(\mathscr{D})\mathrm{E}))$.

If E is invariant under $\mathscr{D}$, then $\mathscr{D}(\mathrm{E})$ is a Lie ideal over Φ in $\mathscr{D}$. For if $D \in \mathscr{D}$ and $D' \in \mathscr{D}(\mathrm{E})$, then $x(DD' - D'D) = (xD)D' - (xD')D = 0$ for all $x \in \mathrm{E}$. Conversely, if $\mathscr{D}'$ is a Lie ideal over Φ in $\mathscr{D}$, then $\mathrm{E} = \Delta(\mathscr{D}')$ is invariant under $\mathscr{D}$. For if $D \in \mathscr{D}$, $(xD)D' = (xD')D = 0$ for all $x \in \mathrm{E}$ and $D' \in \mathscr{D}'$. Hence $xD \in \Delta(\mathscr{D}') = \mathrm{E}$ for all $x \in \mathrm{E}$. Consider the mapping $D \to D^*$ the restriction of D to D^*. This is clearly a homomorphism of $\mathscr{D}$ into the Lie ring over Φ of derivations in E leaving the elements of fixed. Using Lemma 2, one finds that the mapping is onto. Its kernel is $\mathscr{D}' = \mathscr{D}(\mathrm{E})$.

THEOREM 3. *Let $\mathscr{D}$ be a* (restricted if Δ has prime characteristic p) *Lie ring over Φ of derivations in Δ and $\Gamma = \Delta(\mathscr{D})$. Suppose $\mathscr{D}$ contains all inner derivations belonging to $\mathscr{D}(\Gamma)$. Let* E *be a division subring of Δ containing Γ and* $[\mathrm{E}:\Gamma]_L < \infty$. *If a^* is an isomorphism of* E *into Δ leaving the elements of Γ fixed, then a^* can be extended to an inner automorphism in Δ.*

Proof. a^* belongs to $L_\Gamma(\mathrm{E}, \Delta)$; hence a^* can be extended to $a \in L_\Gamma(\Delta)$. $L_\Gamma(\Delta)$ is the closure of $\operatorname{End}\{\mathscr{D}, R_\Delta\}$ in $\operatorname{End}\{\Delta, +)$. Since $[\mathrm{E}:\Gamma]_L < \infty$, there is an $F \in \operatorname{End}\{\mathscr{D}, R_\Delta\}$ such that $a^* = F^*$, where the asterick denotes restriction to E. We can write

$$F = \Sigma D_1^{k_1} \cdots D_s^{k_s} L_{\sigma_j} R_{\mu_{k_1 \cdots k_s j}}$$

where $D_1, \cdots, D_s, \sigma_1, \cdots, \sigma_t$, etc. are as in Lemma 1 with $\mathfrak{M} = \{0\}$. Since $R_x a^* - a^* R_{xa^*} = 0$ for all $x \in$ E and $s^* = F^*$, we have

$$\Sigma(D_1^{k_1} \cdots D_s^{k_s} L_{\sigma_j} R_{x\mu_{k_1 \cdots k_s j} - \mu_{k_1 \cdots k_s j}(xa^*)})^* + \text{terms} = 0$$

By Lemma 1, we obtain from a term of highest weight

$$x\lambda - \lambda(xa^*) = 0 \text{ for all } x \in \text{E and for some } 0 \neq \lambda \in \Delta .$$

Clearly s^* can be extended to the inner automorphism determined by λ.

The following theorem is a special case of one due to Amitsur: Let Δ be a division ring, D a derivation in Δ and $\Gamma = \Delta(D)$. Let S be the set of $x \in \Delta$ such that

$$x(D^n R_{\mu_n} + D^{n-1} R_{\mu_{n-1}} + \cdots + R_{\mu_0}) = 0$$

where $\mu_n \neq 0, \mu_{n-1}, \cdots, \mu_0 \in \Delta$. Then S is a vector space over Γ of dimension $\leq n$. This result is applied in the following theorem.

THEOREM 4. *Let Δ be a division ring and D an outer derivation in Δ. Let k be the least integer greater than* 1 *such that D^k is a derivation. Then Δ has prime characteristic k. If k doesn't exist, then Δ has characteristic zero.*

Proof. By hypothesis and Leibniz's rule

$$0 = R_x D^k - D^k R_x - R_{xD^k} = \binom{k}{1} D^{k-1} R_{xD} + \cdots + \binom{k}{k-1} D R_{xD^{k-1}} .$$

Choose x so that $R_{xD} \neq 0$: If Δ has characteristic 0, $\binom{k}{1} = k \neq 0$. If k exists by Lemma 1 and the Jacobson-Bourbaki Theorem, $[\Delta : \Gamma]_L = \infty > k$, and by Amitsur's Theorem, $[\Delta : \Gamma]_L \leq k - 1 < k$. Hence, in this case, k, can't exist. If Δ has prime characteristic p, then since D^p is a derivation, $k \leq p$. Assume in this case $k < p$. Then, $\binom{k}{1} = k \neq 0$. Applying Amitsur's Theorem yields the fact that $[\Delta : \Gamma]_L \leq k - 1 < p$. By Lemma 1 and the Jacobson-Bourbaki Theorem, $[\Delta : \Gamma]_L \geq p$. This is a contradiction. Hence, if some power $k > 1$ of an outer derivation in Δ is a derivation, then Δ must have prime characteristic p and the least power of D greater than 1 which is a derivation is D^p.

BIBLIOGRAPHY

1. F. Artin, *Galois Theory*, Notre Dame, Indiana, 1944.
2. A. A. Albert, *Structure of Algebras*, Amer. Math. Soc., New York, 1939.
3. S. Amitsur, *A generalization of a theorem on linear differential equations*, Bull. Amer. Math. Soc., **54** (1948), 937–941.
4. J. Dieudonné, *Les semi-derivations dans les extensions radicielles* C. R. Acad. Sci. Paris, **227** (1948) 1319–1320.
5. ———, *Theorie de Galois des extensions radicielles d'exposent quelconque*. C. R. Acad. Sci. Paris, **228** (1949) 148–150.
6. N. Jacobson, *Abstract derivation and Lie algebras*, Trans. Amer. Math. Soc., **42** (1937) 206–224.
7. ———, *Galois theory of purely inseparable fields of exponent one*, Amer. J. of Math., **66** (1944), 645–649.
8. ———, *Structure of Rings*, Amer. Math. Soc., New York, 1956.

YALE UNIVERSITY

FAITHFUL*-REPRESENTATIONS OF NORMED ALGEBRAS

Bertram Yood

1. Introduction. Let B be a complex Banach algebra with an involution $x \to x^*$ in which, for some $k > 0$, $\|xx^*\| \geqq k\|x\|\|x^*\|$ for all x in B. Kaplansky [8, p. 403] explicitly made note of the conjecture that all such B are symmetric. An equivalent formulation is the conjecture that all such B are B^*-algebras in an equivalent norm. In 1947 an affirmative answer had already been provided by Arens [1] for the commutative case. We consider in § 2 the general (non-commutative) case. It is shown that the answer is affirmative if k exceeds the sole real root of the equation $4t^3 - 2t^2 + t - 1 = 0$. This root lies between .676 and .677. In any case these algebras are characterized spectrally as those Banach algebras with involution for which self-adjoint elements have real spectrum and there exists $c > 0$ such that $\rho(h) \geqq c\|h\|$, h self-adjoint (where $\rho(h)$ is the spectral radius of h).

A basic question concerning a given complex Banach algebra B with an involution is whether or not it has a faithful*-representation as operators on a Hilbert space. In § 3 we give a necessary and sufficient condition entirely in terms of algebraic and linear space notions in B. This is that $\rho(h) = 0$ implies $h = 0$ for h self-adjoint and that $R \cap (-R) = (0)$. Here R is the set of all self-adjoint elements linearly accessible [11, p. 448] from the set of all finite sums of elements of the form x^*x. This is related to a previous criterion of Kelley and Vaught [10] which however involves topological notions (in particular, the assumption that the involution is continuous).

If B is semi-simple with minimal one-sided ideals a simpler discussion of *-representations (§ 5) is possible even if B is incomplete. For example if B is primitive then B has a faithful*-representation if and only if $xx^* = 0$ implies $x^*x = 0$. The incomplete case has features not present in the Banach algebra case. In the former case, unlike the latter, a*-representation may be discontinuous. A class of examples is provided in § 5.

2. Arens*-algebras. Let B be a complex normed algebra with an involution $x \to x^*$. An *involution* is a conjugate linear anti-automorphism of period two. Elements for which $x = x^*$ are called *self-adjoint* (s. a.) and the set of s. a. elements is denoted by H. Let $\mathfrak{H}$ be a Hilbert space and

Received May 4, 1959. This research was supported by the National Science Foundation, research grant NSF G 5865, and by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command under Contract No. SAR/AF – 49(636) – 153.

$\mathfrak{C}(\mathfrak{H})$ be the algebra of all bounded linear operators on $\mathfrak{H}$. By a*-representation of B we mean a homomorphism $x \to T_x$ of B into some $\mathfrak{C}(\mathfrak{H})$ where T_{x^*} is the adjoint of T_x. A*-representation which is one-to-one is called *faithful*.

We shall be mainly, but not exclusively, interested in the case where B is complete (a Banach algebra). In § 2 we shall assume throughout that B is a Banach algebra with an involution $x \to x^*$.

As in [5, p. 8] we set $x \circ y = x + y - xy$ and say that x is quasi-regular with quasi-inverse y if $x \circ y = y \circ x = 0$. The quasi-inverse of x is unique, if it exists, and is denoted by x'. As, for example, in [16, p. 617] we define the *spectrum* of x, $sp(x)$, to be the set consisting of all complex numbers $\lambda \neq 0$ such that $\lambda^{-1}x$ is not quasi-regular, plus $\lambda = 0$ provided there does not exist a subalgebra of B with an identity element and containing x as an invertible element. (The treatment of zero as a spectral value plays no role below.) The *spectral radius* $\rho(x)$ if x is defined to be $\sup |\lambda|$ for $\lambda \in sp(x)$.

We say that B is an *Arens*-algebra* [1] if there exists $k > 0$ such that $\| x^*x \| \geqq k \| x \| \, \| x^* \|$, $x \in B$. As usual, we say that B is a *B*-algebra* if $\| x^*x \| = \| x \|^2$, $x \in B$.

2.1. LEMMA. *Let B an Arens*-algebra with $\| xx^* \| \geqq k \| x \| \, \| x^* \|$, $x \in B$. Then for each* s. a. *element h, $\rho(h) \geqq k \| h \|$ and $sp(h)$ is real.*

That the spectrum of a s. a. element h is real is shown in [1, p. 273]. By use of the inequality $\| h^{2^n} \| \geqq k \| h^{2^{n-1}} \|^2$ as in [16, p. 626] it follows that $\rho(h) \geqq k \| h \|$. We shall show (Theorem 2.4) that the spectral conditions of Lemma 2.1 imply that B is an Arens*-algebra.

2.2. LEMMA. *Suppose that for each* s. a. *element h, $\rho(h) \geqq c \| h \|$ and $sp(h)$ is real, where $c > 0$. Let h be* s. a., *$sp(h) \subset [-a, b]$ where $a \geqq 0$, $b \geqq 0$ and let $r > 0$. Then*

(1) $\| (-t^{-1}h)' \| < r$ *if* $t > (1 - cr)b/cr$ *and* $t > (1 + cr)a/cr$,

(2) $\| (t^{-1}h)' \| < r$ *if* $t > (1 - cr)a/cr$ *and* $t > (1 + cr)b/cr$.

Note that (2) follows from (1) as applied to the element-h. By [18, Theorem 3.4] the involution is continuous on B. Therefore h generates a closed*-subalgebra B_0. Let $\mathfrak{M}$ be the space of regular maximal ideals of B_0. For $t > a$ set $u = (-t^{-1}h)'$. By [8, Theorem 4.2], $u \in B_0$. It is readily seen that u is s. a. Since $-t^{-1}h + u + t^{-1}hu = 0$ we have, for each $M \in \mathfrak{M}$, $u(M) = h(M)/(t + h(M))$. By, [8, p. 402] the spectrum of h is the same whether computed in B or in B_0 so that $-a \leqq h(M) \leqq b$. Since $\lambda/(t + \lambda)$ is an increasing function of λ we see that $-a/(t - a) \leqq u(M) \leqq b/(t + b)$. Now $\rho(u) = \sup | u(M) |$, $M \in \mathfrak{M}$. Therefore, since u is s.a.,

$$c \| u \| \leqq \rho(u) \leqq \max [a/(t - a), b/(t + b)] \, . \tag{2.1}$$

From formula (2.1), $\| u \| < r$ if $a/(t-a) < cr$ and $b/(t+b) < cr$. This yields (1).

Note that, under the given hypotheses, $c \leqq 1$.

2.3. LEMMA. *Let x and y be quasi-regular. Then $x + y$ is quasi-regular if and only if $x'y'$ is quasi-regular.*

The formulas $x' \circ (x + y) \circ y' = x'y'$ and $x + y = x \circ (x'y') \circ y$ yield the desired result. Let $r > 0$. If $\| x' \| < r$ and $\| y' \| < r^{-1}$ it follows from Lemma 2.3 and [12, p. 66] that $(x + y)'$ exists.

Consider the situation of Lemma 2.2 and let h_k be s. a., $k = 1, 2$ where $N = \max(\rho(h_1), \rho(h_2))$. By Lemma 2.2, $\| (t^{-1}h_k))' \| < 1$ and $\| (-t^{-1}h_k)' \| < 1$ if $t > (1 + c)N/c$. Then, by Lemma 2.3,

$$sp(h_1 + h_2) \subset [-(1 + c)N/c, (1 + c)N/c] . \tag{2.2}$$

Suppose next that $sp(h_k) \subset [0, \infty)$, $k = 1, 2$. Then $\| (t^{-1}h_k)' \| < 1$ if $t > (1 + c)N/c$ and $\| (-t^{-1}h_k)' \| < 1$ if $t > (1 - c)N/c$. Then by Lemma 2.3,

$$sp(h_1 + h_2) \subset [-(1 - c)N/c, (1 + c)N/c] . \tag{2.3}$$

2.4. THEOREM. *Suppose that for each* s. a. *element h, $\rho(h) \geqq c \| h \|$ and $sp(h)$ is real, where $c > 0$. Then B is an Arens*-algebra with $\|xx^*\| \geqq k \| x \| \| x^* \|$, $x \in B$, where k can be chosen to be $c^5/(1 + c)(1 + 2c^2)$.*

Let $x = u + iv$ where u and v are s. a. Then $x^*x = u^2 + v^2 + i(uv - vu)$, $xx^* = u^2 + v^2 + i(vu\text{-}uv)$ and $xx^* + x^*x = 2u^2 + 2v^2$. We next compare $\rho(u^2) = [\rho(u)]^2$ and $\rho(v^2)$ with $\rho(xx^*)$. For this purpose we may suppose that $\rho(u) \geqq \rho(v)$ for otherwise we can replace x by $ix = -v + iu$. If $\lambda \neq 0$ then $\lambda \in sp(xx^*)$ if and only if $\lambda \in sp(x^*x)$. Thus $\rho(xx^*) = \rho(x^*x)$. By (2.2), $sp(xx^* + x^*x) \subset [-(1 + c)\rho(xx^*)/c, (1 + c)\rho(xx^*)/c]$. Now $2u^2 = xx^* + x^*x - 2v^2$. Let $r > 0, t > 0$. By Lemma 2.2,

$$\| [t^{-1}(xx^* + x^*x)]' \| < r, t > (1 + cr)(1 + c)\rho(xx^*)/c^2r . \tag{2.4}$$

Since $sp(-2v^2) \subset (-\infty, 0]$ and $\rho(2v^2), \leqq \rho(2u^2)$, by Lemma 2.2 we have, for $t > 0$,

$$\| [t^{-1}(-2v^2)]' \| < r^{-1}, t > (r - c)\rho(2u^2)/c . \tag{2.5}$$

we select $c < r < 2c$. For such r, Lemma 2.3 and formulas (2.4) and (2.5) show that $[t^{-1}(2u^2)]'$ exists if $t > \max\{(1+cr)(1+c)\rho(xx^*)/c^2r, (r-c)\rho(2u^2)/c\}$. Now $(r - c)/c < 1$ and $sp(2u^2) \subset [0, \infty)$. Therefore, letting $r \to 2c$, we have

$$\rho(2u^2) \leqq (1 + 2c^2)(1 + c)\rho(xx^*)/(2c^3) . \tag{2.6}$$

On the other hand $\| x \| \leqq \| u \| + \| v \| \leqq [\rho(u) + \rho(v)]/c \leqq 2\rho(u)/c$ and $\| x^* \| \leqq 2\rho(u)/c$. Therefore, by (2.6),

$$(2.7)\qquad \|x\|\,\|x^*\| \leqq 4\rho(u^2)/c^2 \leqq (1+2c)(1+c)\rho(xx^*)/c^5 .$$

But $\rho(xx^*) \leqq \|xx^*\|$. This together with (2.7) completes the proof.

2.5. COROLLARY. *Under the hypotheses of Theorem* 2.4, *the norm of the involution as an operator on B does not exceed* $(1+c)(1+2c^2)/c^5$.

In (2.7) we may replace $\|x\|\,\|x^*\|$ by $\|x^*\|^2$ and $\rho(xx^*)$ by $\|x\|\,\|x^*\|$. This gives $\|x^*\| \leqq (1+c)(1+2c^2)\,\|x\|/c^5$.

We denote by $P(N)$ the set of $x \in B$ such that $sp(x^*x) \subset [0, \infty)$ $(sp(x^*x) \subset (-\infty, 0])$.

2.6. LEMMA. *For an Arens*-algebra B the following are equivalent.*

(a) *B is a B*-algebra in an equivalent norm.*

(b) $N = (0)$.

(c) $P = B$.

Suppose that $N = (0)$. Let $y \in B$. Since the involution on B is continuous, the element y^*y generates a closed*-subalgebra B_0. Let $\mathfrak{M}$ be the space of regular maximal ideals of B_0. By [1, p. 279] the commutative algebra B_0 is *-isomorphic to $C(\mathfrak{M})$. Also $sp(y^*y)$ is real. Hence there exist $u, v \in B_0$ such that $u(M) = \sup(y^*y(M), 0)$ and $v(M) = -\inf(y^*y(M), 0)$, $M \in \mathfrak{M}$. Then u and v are s. a., $y^*y = u - v$ and $uv = 0$. As in [14, p. 281], $(yv)^*(yv) = -v^3$ so that $yv = 0$ by hypothesis. Then $v = 0$ and $sp(y^*y) \subset [0, \infty)$.

A theorem of Gelfand and Neumark [13] asserts that if B is semi-simple, has a continuous involution, is symmetric $(B = P)$ and has an identity then there exists a faithful*-representation $x \to T_x$ of B. This theorem is also valid when B has no identity [4, Theorem 2.16]. In our situation, B is semi-simple [18, Lemma 3.5] and the involution is continuous. Thus a faithful*-representation exists. This representation is bi-continuous by [18, Corollary 4.4].

That (a) implies (b) follows from the well-known fact that any B^*-algebra is symmetric [14, p. 207 and p. 281].

The equation $4t^3 - 2t^2 + t - 1 = 0$ has exactly one real root a. This root a lies between .676 and .677.

2.7. THEOREM. *Suppose that for each* s. a. *element h,* $\rho(h) \geqq c\,\|h\|$ *and* $sp(h)$ *is real, where* $c > 0$. *Then there is an equivalent norm for B in which B is a B*-algebra if* $c > a$.

Suppose that $sp(x^*x) \subset (-\infty, 0]$. By Lemma 2.6 it is sufficient to show that $x = 0$. Suppose that $x \neq 0$. By Theorem 2.4 it is clear that $x^*x \neq 0$ and $\rho(x^*x) \neq 0$. Set $x = u + iv$ where u and v are s. a. As in the proof of Theorem 2.4, $xx^* + x^*x = 2u^2 + 2v^2$ and we may assume that $\rho(u) \geqq \rho(v)$. Since $sp(u^2) \subset [0, \infty)$, $sp(v^2) \subset [0, \infty)$ formula 2.3 shows that $sp(2u^2+2v^2) \subset [-(1-c)\rho(2u^2)/c, (1+c)\rho(2u^2)/c]$. Let $r > 0, t > 0$. From Lemma 2.2,

$\| [-t^{-1}(2u^2 + 2v^2)]' \| < r$ if $t > (1 - cr)(1 + c)\rho(2u^2)/(c^2r)$ and $t > (1 + cr)(1 - c)\rho(2u^2)/(c^2r)$.

We write $x^*x = 2u^2+2v^2+(-xx^*)$. By Lemma 2.2, $\| [-t^{-1}(-xx^*)]'\| < r^{-1}$ if $t > 0$ and $t > (r - c)\rho(x^*x)/c$ since $sp(-xx^*) \subset [0, \rho(x^*x)]$. By Lemma 2.3, $(-t^{-1}x^*x)'$ exists if $t > \max \{(1+cr)(1-c)\rho(2u^2)/c^2r, (1-cr)(1+c)\rho(2u^2)/c^2r, (r - c)\rho(x^*x)/c\}$. Since $sp(x^*x) \subset (-\infty, 0]$, $\rho(x^*x)$ cannot exceed this maximum. Now select r, $1 \leqq r < 2c$ which is possible since $c > a$. Then $(r - c)/c < 1$ and $(1 + cr)(1 - c) \geqq (1 - cr)(1 + c)$. Therefore $\rho(x^*x) \leqq (1 + cr)(1 - c)\rho(2u^2)/c^2r$. Letting $r \to 2c$ we obtain

$$\rho(x^*x) \leqq (1 + 2c^2)(1 - c)\,\rho(2u^2)/2c^3 . \tag{2.8}$$

Next we express $-2u^2 = 2v^2 + (-xx^* - x^*x)$. By formula (2.3), $sp(-xx^* - x^*x) \subset [-(1 - c)\rho(x^*x)/c, (1 + c)\rho(x^*x)/c]$. Recall that $\rho(2v^2) \leqq \rho(2u^2)$. Repeating the above reasoning we see that for $r > 0, t > 0$, $(-t^{-1}(-2u^2))'$ exists for $t > \max \{1-cr)(1+c)\rho(x^*x)/c^2r, (1+cr)(1-c)\rho(x^*x)/c^2r, (r - c)\rho(2u^2)/c\}$. But $sp(-2u^2) \subset (-\infty, 0]$. Then by the argument above we obtain

$$\rho(2u^2) \leqq (1 + 2c^2)(1 - c)\rho(x^*x)/2c^3 . \tag{2.9}$$

From formulas (2.8) and (2.9) we see that $(1 + 2c^2)(1 - c) \geqq 2c^3$ or $4c^3 - 2c^2 + c - 1 \leqq 0$. This gives $c \leqq a$ which is impossible by hypothesis.

Thus if $c > a$ we have $N=(0)$. We subsequently show (Corollary 2.11) that, in any case, N and P are closed in an Arens*-algebra B.

Following Rickart [16, p. 625] we say that B is an *A^*-algebra* if there exists in B an auxiliary normed-algebra norm $| x |$ (B need not be complete it this norm) such that, for some $c > 0$, $| x^*x | \geqq c \, | x |^2$. He raises the question of whether every A^*-algebra has a faithful*-representation.

2.8. COROLLARY. *An A^*-algebra B where $| x^*x | \geqq c \, | x |^2$, $x \in B$, in the auxiliary norm has a faithful*-representation if $c > a$.*

Observe that $| x^* || x | \geqq c \, | x |^2$ so that $| x^* | \leqq c^{-1} | x |$, $x \in B$. Thus the involution on B is continuous in the topology provided by the norm $| x |$. Let B_0 be the completion of B in the norm $| x |$. We extend the function $| x |$ from B to B_0 by continuity. Likewise the involution $x \to x^*$ can be extended by continuity to provide a continuous involution $y \to y^*$ on B_0. We then have $| y^*y | \geqq c \, | y |^2$, $y \in B_0$. As in [16, p. 626] we obtain $\rho(h) \geqq c \, | h |$ for h s. a. in B_0 where $\rho(h)$ is the spectral radius computed for h as an element of the Banach algebra B_0, $\rho(h) = \lim | h^n |^{1/n}$. Also $| y^*y | \geqq c^2 \, | y^* | \, |y|$, $y \in B_0$, so that B_0 is an Arens*-algebra. Hence, by Lemma 2.1, the spectrum of each s. a. element of B_0 is real. By Theorem 2.7, B_0 is a B^*-algebra in an equivalent norm. Therefore B has the desired faithful*-representation.

We have no information on the truth or falsity of Theorem 2.7 for $c \leqq a$.

To prove Theorem 2.7 without restriction on the size of c one can assume without loss of generality that B has an identity. For suppose that B has no identity, $\| x^*x \| \geqq k \| x^* \| \| x \|$, $x \in B$. Adjoin an identity e to B to form the algebra B_1 with the norm defined in B_1 by the rule

$$\| \lambda e + x \| = \sup_{\substack{\|y\|=1 \\ y \in B}} \| \lambda y + xy \| .$$

Then B_1 is a Banach algebra with the involution $(\lambda e + x)^* = \bar{\lambda} e + x^*$[1, p. 275]. By changing in minor ways arguments in [14, p. 207] we see that B_1 is an Arens*-algebra. There is a constant K such that $\| x^* \| \leqq K \| x \|$, $x \in B$. Choose $0 < r < 1$. Given $\lambda e + x \in B_1$ there exists $y \in B$, $\| y \| = 1$, such that

$$\begin{aligned} r^2 \| \lambda e + x \|^2 < \| \lambda y + xy \|^2 &\leqq K \| (\lambda y + xy)^* \| \| \lambda y + xy \| \\ &\leqq K k^{-1} \| y^*(\lambda e + x)^*(\lambda e + x) y \| \\ &\leqq K^2 k^{-1} \| (\lambda e + x)^*(\lambda e + x) \| . \end{aligned}$$

Then

$$\| (\lambda e + x)^*(\lambda e + x) \| \geqq k K^{-2} \| \lambda e + x \|^2 \geqq (kK^{-2})^2 \| \lambda e + x \| \| (\lambda e + x)^* \| .$$

We use this fact later.

Some results on spectral theory in Arens*-algebras were obtained by Newburgh [15]. In a B^*-algebra $\rho(x)$ is a continuous function on the set H of s.a. elements since $\rho(h) = \| h \|$, $h \in H$. This property holds for all Arens*-algebras.

2.9. THEOREM. *In any Arens*-algebra, $\rho(x)$ is a continuous function on H.*

We assume that $\rho(h) \geqq c \| h \|$ and $sp(h)$ is real, $h \in H$. We shall use the following principle [12, p. 67]. If y' exists and $\| z \| < (1 + \| y' \|)^{-1}$ then $(y + z)'$ exists.

Let $h \in H$, $h \neq 0$. Select $t > \rho(h)$ and set $u = (t^{-1}h)'$. We proceed as in the proof of Lemma 2.2. Let B_0 be the closed*-subalgebra generated by h and let $\mathfrak{M}$ be its space of regular maximal ideals. Then $u \in B_0$. Since $t^{-1}h \circ u = 0$ we obtain, for each $M \in \mathfrak{M}$, $u(M) = h(M)/(h(M) - t)$. Since $\lambda/(\lambda - t)$ is a decreasing function of λ, $\sup | u(M) |$ can be majorized by $\rho(h)/(t - \rho(h))$. Then $(1 + \|u\|)^{-1} \geqq (1 + c^{-1}\rho(u))^{-1} \geqq c(t - \rho(h))/(ct + (1 - c)\rho(h)) = a(t)$, say.

Therefore $t^{-1}h + t^{-1}h_1$ is quasi-regular if $\| t^{-1}h_1 \| < a(t)$ or if

$$ct^2 - c[\rho(h) + \| h_1 \|]t - (1 - c)\rho(h) \| h_1 \| > 0 . \tag{2.10}$$

We apply this to $h_1 \in H$, $\| h_1 \| < \rho(h)$. The larger zero d of the left hand side of (2.10) is given by

$$(2.11) \qquad 2d = \rho(h) + \| h_1 \| + [(\rho(h) - \| h_1 \|)^2 + 4c^{-1}\rho(h) \| h_1 \|]^{1/2} .$$

The radical term of (2.11) is majorized by $\rho(h) - \|h_1\| + 2(c^{-1}\rho(h)\|h_1\|)^{1/2}$. Hence $d \leqq \rho(h) + (c^{-1}\rho(h) \|h_1\|)^{1/2}$. Thus $t \notin sp(h + h_1)$ if $t > \rho(h) + (c^{-1}\rho(h)\|h_1\|)^{1/2}$. Likewise $t \notin sp(-h - h_1)$ under the same condition. This shows that

$$(2.12) \qquad \rho(h + h_1) \leqq \rho(h) + (c^{-1}\rho(h) \| h_1 \|)^{1/2} .$$

provided $h_1 \in H$ and $\| h_1 \| < \rho(h)$.

Note that $\rho(h + h_1) \geqq c \| h + h_1 \| \geqq c (\| h \| - \| h_1 \|) \geqq c(\rho(h) - \| h_1 \|)$. Therefore if $\| h_1 \| < c(\rho(h) - \| h_1 \|)$ or equivalently if $\| h_1 \| < c\rho(h)/(1+c)$ we have $\| h_1 \| < \rho(h + h_1)$. We may then apply the above analysis to the pair of s. a. elements $(h + h_1)$, $-h_1$, to obtain (if $\| h_1 \| < c\rho(h)/(1 + c)$)

$$(2.13) \qquad \rho(h) \leqq \rho(h_1 + h_2) + (c^{-1}\rho(h + h_1) \| h_1 \|^{1/2} .$$

From (2.12), $\rho(h + h_1) \leqq [c^{-1/2} + 1]\rho(h)$. Inserting this estimate in the radical term of (2.13) we obtain

$$(2.14) \qquad \rho(h) \leqq \rho(h + h_1) + (c^{-1} + c^{-3/2})^{1/2}(\rho(h) \| h_1 \|)^{1/2}$$

Combining (2.12) and (2.14) we obtain

$$| \rho(h + h_1) - \rho(h) | \leqq [(c^{-1} + c^{-3/2})\rho(h) \| h_1 \|]^{1/2}$$

provided $\| h_1 \| < c\rho(h)/(1 + c)$.

This show that $\rho(x)$ is continuous on H at $x = h$. Clearly we have continuity on H at $x = 0$.

For x s.a. in an Arens*-algebra let $[a(x), b(x)]$ be the smallest closed interval containing $sp(x)$.

2.10. COROLLARY. *For an Arens*-algebra B, $a(x)$ and $b(x)$ are continuous functions of x on H.*

As remarks above indicate, there is no loss of generality in supposing that B has an identity e. Let h be s.a. Choose $\lambda > 0$ such that $sp(\lambda e + h) \subset [1, \infty)$. Let $h_n \rightarrow h$, where each h_n is s.a., and choose $0 < \varepsilon < 1$. We have $\rho(\lambda e + h) = b(\lambda e + h) = \lambda + b(h)$. By the "spectral continuity theorem" (see e. g. [15, Theorem 1]) for all n sufficiently large $sp(\lambda e + h_n) \subset (1 - \varepsilon, b(\lambda e + h) + \varepsilon)$. Also for all n sufficiently large $|\rho(\lambda e + h_n) - \rho(\lambda e + h)| < \varepsilon$ by Theorem 2.9. Since, for such n, $sp(\lambda e + h_n) \subset (0, \infty)$, then $\lambda + b(h_n) = \rho(\lambda e + h_n) \rightarrow \lambda + b(h)$. Therefore $b(h_n) \rightarrow b(h)$. A similar argument shows that $a(h_n) \rightarrow a(h)$.

2.11. COROLLARY. *For an Arens*-algebra B, N and P are closed sets.*

This follows directly from the continuity of the involution on B and Corollary 2.10. Likewise the set H^+ of all s.a. elements whose spectrum is non-negative is closed.

3. Faithful*-representations. Let B be a Banach algebra with an involution $x \to x^*$. Our aim here is to give necessary and sufficient conditions for B to possess a faithful*-representation. Our criterion (Theorem 3.4) is in terms of algebraic and linear space properties of B. A criterion of Kelley and Vaught [10] is largely topological in nature. To discuss this we first prove a simple lemma. We adopt the following notation. Let R_0 be the collection of all finite sums of elements of B of the form x^*x. Let $R = \{x \in H \mid \text{there exists } y \in R_0 \text{ such that } ty + (1-t)x \in R_0, 0 < t \leqq 1\}$. In the notation of Klee [11, p. 448], $R = \text{lin } R_0$ (computed in the real linear space H, the union in H of R_0 and the points of H linearly accessible from R_0). Let P be the closure in B of R_0. If B has an identity e and the involution is continuous then H is closed, e is an interior point of R_0 [10] and $R = P$ [11, p. 448]. If B has no identity or if the involution is not assumed continuous we see no relation, in general, between R and P other than $R \subset P$.

3.1. LEMMA. *Suppose that B has a continuous involution $x \to x^*$ and an identity e. Then there is an equivalent Banach algebra norm $\| x \|_1$ where $\| x^* \|_1 = \| x \|_1$, $x \in B$, and $\| e \|_1 = 1$.*

We first introduce an equivalent norm $\| x \|_0$ in which $\| x^* \|_0 = \| x \|_0$, $x \in B$, by setting $\| x \|_0 = \max(\| x \|, \| x^* \|)$. Let $L_x(R_x)$ be the operator on B defined by left (right) multiplicaton by x; $L_x(y) = xy$ and $R_x(y) = yx$. Let $\| L_x \|$ be the norm of L_x as an operator on B where the norm $\| y \|_0$ is used for B. $\| R_x \|$ is defined in the same way. We set $\| x \|_1 = \max(\| L_x \|, \| R_x \|)$. Then $\| x + y \|_1 \leqq \| x \|_1 + \| y \|_1$ and $\| xy \|_1 \leqq \| x \|_1 \| y \|_1$. Clearly $\| x \|_1 \leqq \| x \|_0$. Moreover $\| L_x \| \geqq \| x \|_0 / \| e \|_0$ and the norms $\| x \|_0$ and $\| x \|_1$ are equivalent. Trivially $\| e \|_1 = 1$. Also

$$\| L_{x^*} \| = \sup_{\| y \|_0 = 1} \| x^* y \|_0 = \sup_{\| y \|_0 = 1} \| y^* x \|_0 = \| R_x \| .$$

Then $\| x^* \|_1 = \max(\| L_{x^*} \|, \| R_{x^*} \|) = \max(\| L_x \|, \| R_x \|) = \| x \|_1$.

In view of Lemma 3.1 the result [10, p. 51] of Kelley and Vaught in question may be expressed as follows.

3.2. THEOREM. *Let B be a Banach algebra with an identity and an involution $x \to x^*$. Then B has a faithful*-representation if and only if * is continuous and $P \cap (-P) = (0)$.*

As it stands this criterion breaks down if B has no identity. For let $B = C([0, 1])$ with the usual involution $x \to x^*$ and norm. Let B_0 be the algebra obtained from B by keeping the norm and involution but defining all products to be zero. Then* is still continuous and $P \cap (-P) = (0)$. But B_0 has no faithful*-representation, for otherwise B_0 would be semi-simple [16, p. 626].

As in [4] we call the involution $x \to x^*$ in B *regular* if, for h s.a., $\rho(h) = 0$ implies $h = 0$. By [4, Lemma 2.15]. * is regular if and only if every maximal commutative *-subalgebra of B is semi-simple. Also every maximal commutative*-subalgebra of B is closed [4, Lemma 2.13].

By a *positive linear functional* f on B we mean a linear functional such that $f(x^*x) \geqq 0$, $x \in B$. The functional f is not assumed to be continuous. If B has an identity then [13, p. 115], $f(h)$ is real for h s.a. and $f(x^*) = \overline{f(x^*)}$. Trivial examples show this to be false, in general. However, from the positivity of f, $f(x^*y)$ and $f(y^*x)$ are complex conjugates which is the fact really needed for the introduction of the inner product in Theorem 3.4.

3.3 LEMMA. *Let the involution on B be regular. Then*
(1) *a positive linear f satisfies the inequalities*

$$f(y^*hy) \leqq f(y^*y) \| h \|, \; y \in B, h \in H \, , \tag{3.1}$$

$$f(y^*x^*xy) \leqq f(y^*y) \| x^*x \|, \; x, y \in B \, , \tag{3.2}$$

(2) *if B has an identity e, any $h \in H$, $\| e-h \| \leqq 1$ has a* s.a. *square root and, moreover, any positive linear functional is continuous on H.*

Suppose first that B has an identity e, $\| e - h \| \leqq 1$, h s.a. In the course of the proof of [4, Theorem 2.16] it was shown that h has a s.a. square root. Next do not assume that B has an identity. Let B_1 be the Banach algebra obtained by adjoining an identity e to B. Consider the power series $(1 - t)^{1/2} = 1 - t/2 - t^2/8 \cdots$. Let $h \in B$, h s.a. and $\| h \| \leqq 1$. Then the expansion $-h/2 - h^2/8 - \cdots$ converges to an element $z \in B$. Let B_0 be a maximal abelian*-subalgebra of B containing h. As noted above, B_0 is a semi-simple Banach algebra. The involution is continuous on B_0 ([16, Corollary 6.3]). Therefore z is s.a. Also $(e + z)^2 = e - h$. Let $y \in B$ and set $k = y + zy$. Then $k^*k = (y^* + y^*z)(y + zy) = y^*(e + z)^2y = y^*y - y^*hy$. For any positive linear functional f on B, $f(k^*k) \geqq 0$ which yields (3.1). Formula (3.2) is a special case.

Suppose that B has an identity e. If we set $y = e$ in (3.1) we obtain $| f(h) | \leqq f(e) \| h \|$ which shows that f is continuous on H.

3.4. THEOREM. *B has a faithful*-representation if and only if* is regular and $R \cap (-R) = (0)$.*

Suppose that B has a faithful*-representation $x \rightarrow T_x$ as operators on a Hilbert space $\mathfrak{H}$. Let h be s.a. and $\rho(h) = 0$. Then $\rho(T_h) = 0$. As T_h is a s.a. operator on a Hilbert space, $T_h = 0$ and therefore $h = 0$. Thus the involution is regular. Let $x \in R \cap (-R)$ and let f be a positive linear functional on B. Then clearly $f(y) \geqq 0$, $y \in R_0$. From the definition of R there exists $y \in R_0$ such that $tf(y) + (1 - t)f(x) \geqq 0$, $0 < t \leqq 1$. It follows that $f(x) \geqq 0$ and hence $f(x) = 0$. Let $\xi \in \mathfrak{H}$ and set $f(x) = (T_x\xi, \xi)$. Then $(T_x\xi, \xi) = 0$ for all $\xi \in \mathfrak{H}$. Since T_x is a s.a. operator we see that $T_x=0$ and $x = 0$.

Suppose now that* is regular and $R \cap (-R) = (0)$. We show first that the regularity of the involution makes available a general representation procedure of Gelfand and Neumark [13].

Let f be a positive linear functional on B. Let $I_f = \{x | f(x^*x) = 0\}$. I_f is a left ideal of B. Let π be the natural homomorphism of B onto B/I_f. Since $f(x^*y) = f(y^*x)$, $\mathfrak{H}'_f = B/I_f$ is a pre-Hilbert space if we define $(\pi(x), \pi(y)) = f(y^*x)$. As in [13, p. 120] we associate with $y \in B$ an operator A^f_y on $\mathfrak{H}'_f$ defined by $A^f_y[\pi(x)] = \pi(yx)$. Formula (3.2) yields

$$(3.3) \qquad || A^f_y[\pi(x)] ||^2 = f(x^*y^*yx) \leqq || y^*y || \, || \pi(x) ||^2 .$$

Thus A^f_y is a bounded operator with norm not exceeding $|| y^*y ||^{1/2}$. It may then be extended to T^f_y, a bounded operator on the completion $\mathfrak{H}_f$ of $\mathfrak{H}'_f$. The mapping $x \to T^f_x$ is a *-representation of B with kernel $\{y \in B \mid yx \in I_f$, for all $x \in B\} = K$. Note that $K^* = K$.

Now take the direct sum $\mathfrak{H}$ of the Hilbert spaces $\mathfrak{H}^f$ as f ranges over all positive linear functionals on B ([13, p. 95]). Since $|| T^f_y || \leqq || y^*y ||^{1/2}$ by (3.3) and this estimate is independent of f, the direct sum ([13, p. 113]) $x \to T_x$ of the representations $x \to T^f_x$ yields a*-representation of B as bounded operators on $\mathfrak{H}$ with kernel $\{y \in B \mid yx \in \cap I_f$, for all $x \in B\}$. If B has an identity, the kernel is the reducing ideal of B ([13, p. 130]), namely $\cap I_f$.

Supppose first that B has an identity e. The set R_0 has the property that $x, y \in R_0, \lambda, \mu \geqq 0$ imply $\lambda x + \mu y \in R_0$. By Lemma 3.3, $R_0 \supset \{x \in H \mid || e - x || \leqq 1\}$. Thus e is an interior point of R_0. By the theory of convex sets in normed linear spaces, R is the closure in H of R_0 and R is a closed cone in H ([11, p. 448]).

Let f be a positive linear functional on B. By Lemma 3.2, f is continuous on H. Also $f(w) \geqq 0$, $w \in R$. Let H' be the conjugate space of H and $G = \{g \in H' \mid g(w) \geqq 0, w \in R\}$. It is easy to see ([10, p. 48]) that G, the dual cone of R, is the set of linear functionals on H which are the restrictions to H of positive linear functional on B. There is no loss generality in assuming that $|| e || = 1$. Let $x \in H$. By [10, Lemma 1.3], dist $(-x, R) = \sup \{g(x) \mid g \in G, g(e) \leqq 1\}$.

We show that $R \cap (-R) = H \cap (\cap I_f)$. Let $y \in H$, $y \in \cap I_f$. For any fixed f, $T^f_y = 0$ and $(T^f_y\xi, \xi) = 0$, $\xi \in \mathfrak{H}_f$. Then $(\pi(yx), \pi(x)) = 0$ for all $x \in B$ in the notation used above. Therefore $f(x^*yx) = 0$, $x \in B$. Setting $x = e$ we see that $f(y) = 0$. Then by the distance formula, $-y \in R$. Likewise $y \in R$. Suppose conversely that $y \in R \cap (-R)$. It is easy to see that for each $z \in B$, $z^*R_0z \subset R_0$. Therefore $z^*Rz \subset R$. Hence $z^*yz \in R \cap (-R)$, $z \in B$. From the distance formula, $\sup \{f(z^*yz) \mid f \text{ positive}, f(e) \leqq 1\} = 0 = \sup \{f(-z^*yz) \mid f \text{ positive}, f(e) \leqq 1\}$. Hence $f(z^*yz) = 0$ for each positive linear functional. Then $(T^f_y\pi(z), \pi(z)) = 0$ for all z whence $T^f_y = 0$. Therefore $T_y = 0$ and $y \in H \cap (\cap I_f)$.

This proves the theorem in case B has an identity. Suppose that B has no identity. Let B_1 be the algebra obtained by adjoining an identity e to B. We extend the involution to B_1 by setting $(\lambda e + x)^* = \overline{\lambda e} + x^*$. The involution on B_1 is regular [4, Lemma 2.14]. Let R_0' and R' be the sets R_0 and R respectively computed for the algebra B_1. By the above it is sufficient to show that $R \cap (-R) = (0)$ implies $R' \cap (-R') = (0)$. Suppose that $R \cap (-R) = (0)$.

Let $x, y \in B$. Then $y^*(\lambda e + x)^*(\lambda e + x)y = (\lambda y + xy)^*(\lambda y + xy)$. This shows that $y^*R_0'y \subset R_0$ which implies $y^*R'y \subset R$. Note also that B is semi-simple [18, Lemma 3.5] which implies that $zB = (0)$, or $Bz = (0)$, $z \in B$, can hold only for $z = 0$.

Suppose that $\lambda e + x \in R' \cap (-R')$ where $x \in B$ and λ is a scalar. We derive a contradiction from $\lambda \neq 0$. For every $y \in B$, $y^*(\lambda e + x)y \in R \cap (-R)$. Setting $u = -x/\lambda$ we have $y^*(e-u)y = 0$ or $y^*y = y^*uy$ for all $y \in B$. Then

$$h^2 = huh,\ h \text{ s.a.} \tag{3.4}$$

Let h_1 and h_2 be s.a. Then $(h_1 + h_2)^2 = (h_1 + h_2)u(h_1 + h_2)$. From (3.4) we obtain

$$h_1h_2 + h_2h_1 = h_1uh_2 + h_2uh_1\ . \tag{3.5}$$

Also $(h_1 - ih_2)(h_1 + ih_2) = (h_1 - ih_2)u(h_1 + ih_2)$ From (3.4) we get

$$h_2h_1 - h_1h_2 = h_2uh_1 - h_1uh_2 \tag{3.6}$$

From (3.5) and (3.6) we see that $h_1h_2 = h_1uh_2$. Consequently for h_k s.a., $k = 1, 2, 3, 4$, we see that $(h_1 + ih_2)(h_3 + ih_4) = (h_1 + ih_2)u(h_3 + ih_4)$. In other words

$$zw = zuw,\ z, w \in B\ . \tag{3.7}$$

From (3.7) $(z - zu)w = 0$ for all $w \in B$ so that $z = zu$ for each z. Hence u is a right identity for B. Likewise from $z(w - uw) = 0$ for all $z \in B$ we see that u is an identity for B. But this is impossible since we are considering the case where B has no identity.

We now have $x \in R' \cap (-R')$. Then $y^*xy = 0$ for all $y \in B$. Therefore $hxh = 0$, h s.a. Also for h_k s.a., $k = 1, 2$, $(h_1 + h_2)x(h_1 + h_2) = 0$ so that $h_2xh_1 + h_1xh_2 = 0$. Also $(h_1 - ih_2)x(h_1 + ih_2) = 0$ so that $h_1xh_2 - h_2xh_1 = 0$. Therefore $h_1xh_2 = 0$. It follows that $zxw = 0$ for all $z, w \in B$. This implies that $x = 0$ and completes the proof.

4. Preliminary ring theory.

Let R be a semi-simple ring with minimal one-sided ideals. For a subset A of R let $\mathfrak{L}(A) = \{x \in R \mid xA = (0)\}$ and $\mathfrak{R}(A) = \{x \in R \mid Ax = (0)\}$. Consider a two-sided I of R. If $x \in R(I)$, $y \in R$, $z \in I$ then $zy \in I$, $z(yx) = 0$ so that $\mathfrak{R}(I)$ is a two-sided ideal of R. Therefore $\mathfrak{R}(I)I$ is an ideal. But $[\mathfrak{R}(I)I]^2 = (0)$. Thus, by semi-simplicity, $\mathfrak{R}(I)I = (0)$

and $\mathfrak{R}(I) \subset \mathfrak{L}(I)$. Likewise we have $\mathfrak{L}(I) \subset \mathfrak{R}(I)$ and thus $\mathfrak{R}(I) = \mathfrak{L}(I)$. Let S be the socle [5, p. 64] of R. This is the algebraic sum of the minimal left (right) ideals of R. S is a two-sided ideal. Therefore $\mathfrak{L}(S) = \mathfrak{R}(S)$. This set we denote by $S^\perp$. Note that $S \cap S^\perp = (0)$.

We call an idempotent e of R a *minimal idempotent* if eR is a minimal right ideal.

4.1. LEMMA. (a) *Let I be a left (right) ideal of R, $I \neq (0)$. Then I contains no minimal left (right) ideal of R if and only if $I \subset S^\perp$.*

(b) *$R/S^\perp$ is semi-simple. If S_0 is the socle of $R/S^\perp$ then $S_0^\perp = (0)$.*

Let $I \neq (0)$ be a left ideal of R. Suppose that $I \subset S^\perp$. Then I cannot contain a minamal left ideal J of R for any such J would be contained in $S \cap S^\perp$. Next suppose that $I \not\subset S^\perp$. We must show that I contains a minimal left ideal of R. There exists a minimal idempotent e such that $e I \neq (0)$. Choose $u \in I$ such that $eu \neq 0$. By semi-simplicity and the minimality of eR, $eR = euR$. Thus there exists $z \in R$ such that $euz = e$. Since $(euz)^2 = e$, we have $j \neq 0$ where $j = zeu$. Note that $j^2 = j$. As $u \in I$ we have $Rj \subset I$. To see that Rj is the desired minimal ideal it is sufficient to see that jRj is a division ring [5, p. 65].

Note that $jz = zeuz = ze \neq 0$. Then $Rze = Re$ so that there exists $v \in R$ where $vze = e$. Then $vj = vzeu = eu$ and $vjz = e$.

We assert that $jx_1j = jx_2j$ if and only if $eux_1ze = eux_2ze$. For if $jx_1j = jx_2j$, multiply on the left by v and on the right by z and use the relations $vj = eu$ and $jz = ze$. If $eux_1ze = eux_2ze$ multiply on the left by z and on the right by u and use $zeu = j$.

Therefore the mapping $\tau : \tau(jxj) = euxze$ is a well-defined one-to-one mapping of jRj into eRe. The mapping is onto. For let $ewe \in eRe$. Then $ewe = euzwvze = \tau(jzwvj)$. τ is clearly additive. But also $\tau[(jxj)(jyj)] = \tau(jxjyj) = euxjyze = (euxze)(euyze) = \tau(jxj)\tau(jyj)$. Therefore τ is a ring isomorphism of jRj onto eRe. Since eRe is a division ring so is jRj.

Let J be the radical of $R/S^\perp$ and π be the natural homomorphism of R onto $R/S^\perp$. Suppose that $J \neq 0$. Then $\pi^{-1}(J) \supset S^\perp$ and $\pi^{-1}(J) \neq S^\perp$. By (a), $\pi^{-1}(J)$ contains a minimal idempotent e of R. We then have $\pi(e) \in J$, $\pi(e) \neq 0$. This is impossible since the radical of a ring contains no non-zero idempotents.

Let S_0 be the socle of $R/S^\perp$ and e be a minimal idempotent of R. Clearly $\pi(e) \neq 0$ and π is one-to-one on eRe. Then $\pi(e)\pi(R)\pi(e)$ is a division ring so that, since $R/S^\perp$ is semi-simple, $\pi(e) \in S_0$. Let $\pi(x) \in S_0^\perp$. Then $\pi(ex) = 0$ so that $ex \in S^\perp \cap S = (0)$. Hence $x \in S^\perp$ and $\pi(x) = 0$.

The following result is due to Rickart [17, Lemma 2.1.]:

4.2. LEMMA. *Let A be any ring. Let $x \rightarrow x^*$ be a mapping of A onto A such that $x^{**} = x$, $(xy)^* = y^*x^*$ and $xx^* = 0$ implies $x = 0$. Then any* *minimal right (left) ideal I of A can be written in the form $I=eA(I=Ae)$ where $e^2 = e \neq 0$, $e^* = e$.*

We improve this result by relaxing the conditions on $x \to x^*$ but at the expense of assuming the ring to be semi-simple.

4.3. LEMMA. *Let R be semi-simple with minimal one-sided ideals. Let $x \to x^*$ be a mapping of R onto R satisfying $x^{**} = x$ and $(xy)^* = y^*x^*$. Then the following statements are equivalent.*

(1) *Every minimal right ideal is generated by a* s.a. *idempotent.*

(2) *Every minimal left ideal is generated by a* s.a. *idempotent.*

(3) *$jj^* \neq 0$ for each minimal idempotent j of R.*

(4) *$xx^* = 0$ implies $x \in S^\perp$*

We say that the idempotent e is s.a. if $e^* = e$. Note that $x \to x^*$ is one-to-one and $0^*=0$. As a preliminary we show that j^* is a minimal idempotent if j^* is a minimal idempotent. The ideal $I = jR$ is a minimal right ideal. Then $I^* = Rj^*$ is a left ideal $\neq (0)$. Suppose $I^* \supset K \neq (0)$, $I^* \neq K$ where K is a left ideal of R. By semi-simplicity there exists $x \in K$ such that $x^2 \neq 0$. Then $I^* \supset Rx \neq (0)$, $I^* \neq Rx$. This implies that $I \supset x^*R \neq (0)$, $I \neq x^*R$. This is impossible. Therefore I^* is a minimal left ideal and j^* is a minimal idempotent. It is clear from this argument that (1) and (2) imply each other.

Assume (1). Let j be a minimal idempotent, $I = Rj$ a minimal left ideal. We can write $I = Re$ where e is a s.a. idempotent. Then for some $v \in R$, $vj = e$. But $e = ee^* = vjj^*v$. Therefore $jj^* \neq 0$. Thus (1) implies (3).

Assume (3). Suppose that $xx^* = 0$, $x \neq 0$. Let $I = Rx$. Then $I \neq (0)$. Suppose that I contains a minimal left ideal Rj of R where j is a minimal idempotent. We can write $j = yx$, $y \in R$. Then $0 \neq jj^* = yxx^*y^* = 0$. This shows that I contains no minimal left ideal of R. By Lemma 4.1, $I \subset S^\perp$. Then for any minimal idempotent e, $0 = e(ex)$ and $x \in S^\perp$. Thus (3) implies (4).

Assume (4). If j is a minimal idempotent and $jj^* = 0$ then $j \in S^\perp$. But $j \in S$ and $S \cap S^\perp = (0)$. This shows that (4) implies (3).

Assume (3). Let j be a minimal idempotent, $I = jR$. Since $jj^* \neq 0$, $jj^*R = I$. There exists $u \in R$, $jj^*u = j$. As noted above j^* is a minimal idempotent. By (3), $0 \neq j^*j$. Then $0 \neq (u^*jj^*)(jj^*u) = u^*(jj^*)^2u$. Therefore $(jj^*)^2 \neq 0$. Set $h = jj^*$. Since I is minimal, $I = hI$. As in the proof of [17, Lemma 2.1] there exists $u \in I$ such that $h = hu$. Set $e = uu^*$. As in that proof, e is a s.a. idempotent and it remains only to check that $e \neq 0$ to obtain (2) from (3). If $e = 0$ then $0 = uu^* = huu^*h = h^2$ which is impossible.

5. Normed algebras with minimal ideals.

We are concerned here with*-representations of semi-simple normed algebras B with an involution where B has minimal one-sided ideals. B may be incomplete.

5.1. LEMMA. *Let B be a complex semi-simple normed algebra with minimal one-sided ideals. Let e_1, e_2 be minimal idempotents of B. Then the following statements are equivalent.*

(1) $e_1Be_2 \neq (0)$.

(2) $e_2Be_1 \neq (0)$,

(3) e_1Be_2 *is one-dimensional.*

(4) e_2Be_1 *is one-dimensional.*

Suppose (1). There exists $u \in B$, $e_1ue_2 \neq 0$. Since $e_1ue_2B = e_1B$, there exists $v \in B$ where $e_1ue_2v = e_1$. Then $e_2ve_1 \neq 0$ and (1) implies (2). Let $E = \{\lambda e_2ve_1 \mid \lambda \text{ complex}\}$. Clearly $e_2Be_1 \supset E$. Let $e_2xe_1 \in e_2Be_1$. Then $e_2xe_1 = e_2x(e_1ue_2ve_1) = (e_2xe_1ue_2)e_2ve_2$, a scalar multiple of e_2 by the Gelfand-Mazur Theorem. Thus (1) implies (4). The remainder of the argument is trivial.

For the remainder of § 5, B denotes a semi-simple complex normed algebra with an involution and with minimal one-sided ideals.

5.2. THEOREM. *The following statements concerning B are equivalent.*

(1) *Every minimal one-sided ideal is generated by a* s.a. *idempotent.*

(2) *There exists a*-representation with kernel $S^{\perp}$.*

(3) *There exists a*representation with kernel contained in $S^{\perp}$.*

(4) $j - j^*$ *is quasi-regular for every minimal idempotent j.*

(5) $jBj^* \neq (0)$ *for every minimal idempotent j and $xx^* = 0$ implies $x^*x \in S^{\perp}$, $x \in B$.*

Suppose that (1) holds. Let Q be the set of all s.a. minimal idempotents of B and let $j \in Q$. By the Gelfand-Mazur Theorem, $jBj = \{\lambda j \mid \lambda \text{ complex}\}$. Suppose $jx^*xj = \lambda j$. Taking adjoints, $\lambda = \bar{\lambda}$ so λ is real. We show that $jx^*xj = -j$ is impossible. For suppose $jx^*xj = -j$. Now $jxj = \alpha j$ for some scalar $\alpha = a + bi$, where a, b are real. Set $c = a + (a^2 + 1)^{1/2}$. By the use of $jx^*xj = -j$ one obtains $(jx^* - cj)(jx^* - cj)^* = 0$. From Lemma 4.3 we have $jx^* - cj = 0$. Then $(a - bi)j = jx^*j = cj$. It follows that $c = a$ and $b = 0$. This is impossible.

For $j \in Q$ we define the functional $f_j(x)$ on B by the rule $f_j(x)j = jxj$. By the above, $f_j(x^*x) \geqq 0$, $x \in B$, $x \in B$ and $f_j(x^*) = \overline{f_j(x)}$. The functional f_j is a positive linear functional on B and is continuous on B.

The following inequality of Kaplansky [9, p. 55] is then available.

$$f_j(y^*x^*xy) \leqq \nu(x^*x)f_j(y^*y),\ x, y \in B\ , \tag{5.1}$$

where $\nu(x^*x) = \lim \|(x^*x)^n\|^{1/n}$. Let $I_j = \{x \mid f_j(x^*x) = 0\}$. Let π be the natural homomorphism of B onto B/I_j. The definition $(\pi(x), \pi(y)) = f_j(y^*x)$ makes B/I_j a pre-Hilbert space. Let $\mathfrak{H}_j$ be its completion. See the discussion of the Gelfand-Neumark procedure in § 3. To each $y \in B$ we correspond the operator A_y^j defined by $A_y^j[\pi(x)] = \pi(yx)$. Then

$$\| A_y^j[\pi(x)] \|^2 = f_j(x^*y^*yx) \leqq \nu(y^*y) \| \pi(x) \|^2$$

by (5.1). Thus A_y^j can be extended to a bounded linear operator T_y^j on $\mathfrak{H}_j$, and the mapping $y \to T_y^j$ is a*-representation of B.

Since $\| T_y^j \| \leqq \nu(y^*y)^{1/2}$ and the estimate is independent of $j \in Q$ we can take the direct sum $\mathfrak{H}$ of the Hilbert spaces $\mathfrak{H}_j$, $j \in Q$ and the direct sum $x \to T_x$ of the representations $x \to T_x^j$. This gives a*-representation of B with kernel K where

$$K = \{x \in B \mid xy \in \bigcap_{j \in Q} I_j, \text{ for all } y \in B\} .$$

We show that $K = S^\perp$.

It is clear that $S^* = S$ and therefore $(S^\perp)^* = S^\perp$. Using this and Lemma 4.3 we obtain the following chain of equivalences: $x \varepsilon \cap I_j \leftrightarrow jx^*xj = 0$, all $j \in Q \leftrightarrow jx^* \in S^\perp$, all $j \in Q \leftrightarrow jx^* = 0$, all $j \in Q \leftrightarrow x^* \in S^\perp \leftrightarrow x \in S^\perp$. Therefore $\cap I_j = S^\perp$. Thus $K = \{x \mid xy \in S^\perp, \text{ all } y \in B\}$. If $x \in K$ then $xj \in S^\perp \cap S = (0)$ for all $j \in Q$ and $x \in S^\perp$. Clearly $S^\perp \subset K$. Therefore $K = S^\perp$. Hence (1) implies (2). Clearly (2) implies (3).

Assume (3) and let φ be a*-representation whose kernel $\subset S^\perp$. Let j be a minimal idempotent of B. Let A be the subalgebra of B generated by j and j^*. By the Gelfand-Mazur Theorem, $jj^*j = \lambda j$ for some scalar λ. Thus A is the linear space spanned by j, j^*, jj^* and j^*j. A is finite-dimensional and $A \subset S$. Since $S \cap S^\perp = (0)$, φ is one-to-one on A. Note that $A = A^*$. Let E be the B^*-algebra obtained by taking the closure in the operator algebra on the appropriate Hilbert space of $\varphi(B)$. Clearly $\varphi(A)$ is a closed*-subalgebra of E. The element $\varphi(j - j^*)$ is a skew element of E and therefore quasi-regular in E. By [8, Theorem 4.2] its quasi-inverse in E already lies in $\varphi(A)$. As φ is one-to-one on A, $j - j^*$ has a quasi-inverse in A. Thus (3) implies (4).

Assume (4). Let j be a minimal idempotent of B. There exists $u \in B$ such that $j - j^* + u - (j - j^*)u = 0$. If $jj^* = 0$ then left multiplication by j gives $j = 0$ which is impossible. Therefore $jj^* \neq 0$. By Lemma 4.3, we see that (4) implies (1). Clearly (1) implies (5) by Lemma 4.3. Assume (5). Let j be a minimal idempotent of B. If $j^*j = 0$ then $0 = x^*j^*jx = (jx)^*(jx)$. Also $jxx^*j^* \in S^\perp \cap S = (0)$ for all $x \in B$. Since $jBj^* \neq (0)$, jBj^* is one-dimensional by Lemma 5.1. Hence there exists $u \neq 0$ in B and a linear functional $f(x)$ on B such that $jxj^* = f(x)u$. Then $f(xx^*) = 0$ for all $x \in B$. Expanding $0 = f[(x + y)(x + y)^*] = f[(x + iy)(x + iy)^*]$ we see that $f(xy^*) = 0$ for all $x, y \in B$. Hence f vanishes on B^2. Take any $z \in B$. We have $f(jz) = 0$ or $jzj^* = 0$. Thus $jBj^* = (0)$ which is impossible. Therefore $j^*j \neq 0$. By Lemma 4.3, (5) implies (1).

Algebras to which Theorem 5.2 can be applied most easily are those for which $S^\perp = (0)$. Examples are semi-simple annihilator algebras studied by Bonsall and Goldie [3] and primitive algebras (Corollary 5.4).

5.3. COROLLARY. *If B is an Arens*-algebra with non-zero socle then $N \subset S^\perp$.*

Let $x_0 \in N$, $sp\,(x_0x_0^*) \subset (-\infty, 0]$. Then we can write $x_0x_0^* = -h^2$ where h is s.a. The ideal $S^\perp$ is closed and self-adjoint. Let π be the natural homomorphism of B onto $B/S^\perp$. An involution can be defined in $B/S^\perp$ by the rule $[\pi(x)]^* = \pi(x^*)$. Since B is semi-simple, $B/S^\perp$ has non-zero socle. Let $\pi(x)$ be a minimal idempotent of $B/S^\perp$. Then $[\pi(x)]^* - \pi(x) = \pi(x^* - x)$ is quasi-regular in $B/S^\perp$ since $x^* - x$ is quasi-regular in B. By Theorem 5.2 and Lemma 4.1, $B/S^\perp$ has a faithful*-representation. Then, by Theorem 3.4, $\pi(x_0x_0^*) = 0 = \pi(h^2)$. Therefore $x_0x_0^* \in S^\perp$ and $(jx_0)(jx_0)^* = 0$ for each minimal idempotent j of B. Therefore $jx_0 = 0$ for all such j and $x_0 \in S^\perp$.

We call the involution $x \to x^*$ *proper* if $xx^* = 0$ implies $x = 0$. We call the involution *quasi-proper* if $xx^* = 0$ implies $x^*x = 0$. Not every involution is quasi-proper. For example let B be all 2×2 matrices with the involution defined by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^* = \begin{pmatrix} \bar{a} & -\bar{c} \\ -\bar{b} & \bar{d} \end{pmatrix}.$$

To see that this is not quasi-proper choose x as

$$\begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix}.$$

Every proper involution is quasi-proper but the converse is false. Consider, for example $B = C([0, 1])$ and set $x^*(t) = \overline{x(1-t)}$.

5.4. COROLLARY. *Let B be primitive with non-zero socle. Then the following statements are equivalent.*

(1) *The involution* is proper.*

(2) *The involution* is quasi-proper.*

(3) *There exists a faithful*-representation of B.*

Suppose that $S^\perp \neq (0)$. Then by [5, p. 75], $S \subset S^\perp$. Since $S \cap S^\perp = (0)$ this is impossible. Therefore $S^\perp = (0)$. Assume (2). Let j be a minimal idempotent of B. Then $jBj^* \neq (0)$ (see the prooof of [16, Theorem 4.4]) and, consequently (5) of Theorem 5.2 is satisfied. Then by Theorem 5.2, (2) implies (3); the remainder of the proof is obvious.

The equivalence of (1) and (3) was noted by Rickart [17, Theorem 3.5]. By Lemma 4.3 and Theorem 5.2 this equivalence of (1) and (3) holds for any B for which $S^\perp = (0)$.

If B is complete the following statements hold. (1) Any*-representation of B is continuous [16, Theorem 6.2]. (2) If B has a faithful*-representation then the involution is continuous [16, Lemma 5.3]. We show that both these statements can be false for B incomplete. Our discussion is based on work of Kakutani and Mackey [6, p. 56] (see also [7] for the complex case). Let $\mathfrak{X}$ be an infinite-dimensional complex Hilbert space, $(x, x)^{1/2} = \| x \|$. Let $||| x |||$ be any other norm on $\mathfrak{X}$ such that $||| x ||| \leqq \| x \|$, $x \in \mathfrak{X}$. Let $\mathfrak{X}_1 = \{y \in \mathfrak{X} \mid (x, y)$ is continuous on $\mathfrak{X}$ in the norm $||| x |||\}$ and endow $\mathfrak{X}_1$ with the norm $||| x |||$. Then [6, p. 56] a linear functional $f(x)$ on $\mathfrak{X}_1$ has the form $f(x) = (x, y)$. Moreover $\mathfrak{X}_1$ is dense in $\mathfrak{X}$ in both norms. If there exists $c > 0$ such that $\| x \| \leqq c ||| x |||$, $x \in \mathfrak{X}_1$ then $\mathfrak{X} = \mathfrak{X}_1$ and $\mathfrak{X}_1$ is complete.

Let $\mathfrak{E}(\mathfrak{X}_1)$ be the normed algebra of all bounded linear operators on $\mathfrak{X}_1$. As shown in [6, p. 56], $\mathfrak{E}(\mathfrak{X}_1)$ has an involution $T \to T^*$ where $(T(x), y) = (x, T^*(y))$, $x, y \in \mathfrak{X}_1$. In these terms we show the following.

5.5. THEOREM. *The following statements are equivalent.*

(1) $\mathfrak{X}_1$ *is complete.*

(2) *The involution in* $\mathfrak{E}(\mathfrak{X}_1)$ *is continuous.*

(3) *The faithful*-representation of Theorem* 5.2 *for* $\mathfrak{E}(\mathfrak{X}_1)$ *is continuous.*

As already noted (1) implies (2) and (3). Assume (2) and let M be the norm of the involution. By [2] any minimal idempotent of $\mathfrak{E}(\mathfrak{X}_1)$ is one-dimensional and the operator J defined by the rule $J(x) = (x, u)u$ where $(u, u) = 1$ is a minimal idempotent. Since $(J(x), y) = (x, u)(u, y) = (x, J(y))$ we have $J = J^*$. The functional f defined by $f(U)J = JUJ$ is a continuous positive linear functional on $\mathfrak{E}(\mathfrak{X}_1)$. For $z \in \mathfrak{X}_1$ define the operator W_z by the rule $W_z(x) = (x, u)z$. Then we can write the norm of W_z as $C ||| z |||$ where C is independent of z. A simple computation gives $JW_z^* W_z J = (z, z)J$. By formula (5.1), where $\| U \|$ denotes the norm in $\mathfrak{E}(\mathfrak{X}_1)$,

$$\| z \|^2 = (z, z) \leqq \nu(W_z^* W_z) \leqq \| W_z^* W_z \| \leqq C^2 M ||| z |||^2 .$$

This shows that $\mathfrak{X}_1$ is complete.

Assume (3) and let N be the norm of the faithful*-representation. Let $I_f = \{U \in \mathfrak{E}(\mathfrak{X}_1) \mid f(U^*U) = 0\}$, π be the natural homomorphism of $\mathfrak{E}(\mathfrak{X}_1)$ onto $\mathfrak{E}(\mathfrak{X}_1)/I_f$ and $(\xi, \eta)_f$ be the inner product for the pre-Hilbert space $\mathfrak{E}(\mathfrak{X}_1)/I_f$. Let $V \to T_V^f$ be the partial*-representation induced by f. Its norm cannot exceed N. Now $(\pi(J), \pi(J))_f = 1$ and

$$N^2 \| U \|^2 \geqq \| T_U^f[\pi(J)] \|^2 = (UJ, UJ)_f = f(JU^*UJ) = f(U^*U) .$$

Applying this formula to $U = W_z$ we obtain $N^2 C^2 ||| z |||^2 \geqq (z, z)$ and again $\mathfrak{X}_1$ is complete.

A specific example is suggested in [6, p. 57]. Let $\mathfrak{X} = l^2$, $||| \{x_n\} ||| = \sup | x_n |$. An easy computation gives $\mathfrak{X}_1 = l^2 \cap l^1$ in the sup norm. Here the involution and*-representation are therefore not continuous.

6. Involutions on $\mathfrak{E}(\mathfrak{H})$.

Let $\mathfrak{H}$ be a Hilbert space and $\mathfrak{E}(\mathfrak{H})$ the B^*-

algebra of all bounded linear operators on $\mathfrak{H}$. We determine in Theorem 6.2 all the involutions on $\mathfrak{E}(\mathfrak{H})$ for which there are faithful adjoint-preserving representations.

6.1. LEMMA. *Let$^\sharp$ be any involution on $\mathfrak{E}(\mathfrak{H})$. Then there exists an invertible* s.a. *element U in $\mathfrak{E}(\mathfrak{H})$ such that $T^\sharp = U^{-1}T^*U$ for all $T \in \mathfrak{E}(\mathfrak{H})$. Conversely any such mapping is an involution.*

The mapping $T \to T^{\sharp *}$, $T \in \mathfrak{E}(\mathfrak{H})$, is an automorphism of $\mathfrak{E}(\mathfrak{H})$. Thus there exists $V \in \mathfrak{E}(\mathfrak{H})$ where $T^{\sharp *} = VTV^{-1}$, $T \in \mathfrak{E}(\mathfrak{H})$. Set $U = V^*$. Then $T^\sharp = U^{-1}T^*U$. Since $T^{\sharp\sharp} = T$, $T = (U^{-1}T^*U)^\sharp = U^{-1}U^*T(U^*)^{-1}U$. Thus $U^{-1}U^*$ lies in the center of $\mathfrak{E}(\mathfrak{H})$. Consequently $U = \lambda U^*$ for some scalar λ. Since $U^*U = |\lambda|^2U^*U$ we see that $|\lambda| = 1$. Set $\lambda = \exp(i\theta)$ and $W = \exp(-i\theta/2)U$. Then $W^* = W$ and $T^\sharp = W^{-1}T^*W$, $T \in \mathfrak{E}(\mathfrak{H})$. The remaining statement is easily verified.

6.2. THEOREM. *An involution $T \to T^\sharp$ on $\mathfrak{E}(\mathfrak{H})$ is proper if and only if it can be expressed in the form $T^\sharp = U^{-1}T^*U$, $U \in \mathfrak{E}(\mathfrak{H})$* where U is s.a. and $sp(U) \subset (0, \infty)$.

If $T \to T^\sharp$ is a proper involution then (see [7]) an inner product can be defined in $\mathfrak{H}$ in terms of which $T^\sharp$ is the adjoint of T. Hence the proper involutions are those for which there is an adjoint preserving faithful representation.

Let W be a one-dimensional operator, $W(x) = (x, z)w$ with $w \neq 0, z \neq 0$. Then $W^*(x) = (x, w)z$. By Lemma 6.1 we can write $T^\sharp = U^{-1}T^*U$, $T \in \mathfrak{E}(\mathfrak{H})$, where U is s.a. Then $0 \neq W^\sharp W = U^{-1}W^*UW$. Hence $0 \neq W^*UW$. But $W^*UW(x) = (x, z)W^*U(w) = (x, z)(U(w), w)z$. Therefore $(U(w), w) \neq 0$ for an arbitrary non-zero $w \in \mathfrak{H}$. Hence $(U(w), w) \neq 0$ for an arbitrary non-zero $w \in H$. Hence $(U(w), w)$ has a constant sign and, by changing to $-U$ if necessary, we may suppose that $(U, w), w) \geqq 0$, $w \in \mathfrak{H}$. Then we can write $U = V^2$ where V is s.a. in $\mathfrak{E}(\mathfrak{H})$.

Suppose conversely that $T^\sharp = V^{-2}T^*V^2$, $T \in \mathfrak{E}(\mathfrak{H})$ where V is s.a. Then $TT^\sharp = (TV^{-1})(TV^{-1})^*V^2$. Thus $TT^\sharp = 0$ implies that $TV^{-1} = 0$ and that $T = 0$.

REFERENCES

1. R. Arens, *Representations of *-algebras*, Duke Math. J. **14** (1947), 269–282.
2. B. H. Arnold, *Rings of operators on vector spaces*, Ann. of Math. **45** (1944), 24–49.
3. F. F. Bonsall and A. W. Goldie, *Annihilator algebras*, Proc. London Math. Soc. **14** (1954), 154–167.
4. P. Civin and B. Yood, *Involutions on Banach algebras*, Pacific J. Math., **9** (1959), 415–436.
5. N. Jacobson, *Structure of rings*, Amer. Math. Soc. Colloq. Publ. vol. **37**, Providence, 1956.
6. S. Kakutani and G. W. Mackey, *Two characterizations of real Hilbert space*, Ann. of Math. **45** (1944), 50–58.

7. S. Kakutani and G. W. Mackey, *Ring and lattice characterization of complex Hilbert space*, Bull. Amer. Math. Soc. **52** (1946), 727-733.

8. I. Kaplansky, *Normed algebras*, Duke Math. J. **16** (1949), 399-418.

9. ———, *Topological algebra*, Department of Mathematics, University of Chicago, 1952, mimeographed notes.

10. J. L. Kelley and R. L. Vaught, *The positive cone in Banach algebras*, Trans. Amer. Math. Soc. **74** (1953), 44-55.

11. V. L. Klee, Jr., *Convex sets in linear spaces*, Duke Math. J. **18** (1951), 443-466.

12. L. H. Loomis, *An introduction to abstract harmonic analysis*, D. Van Nostrand Co., New York, 1953.

13. M. A. Naimark, *Involutive Algebren*, Sowjetische Arbeiten zur Funktionalysis, Berlin (1954), 89-196.

14. M. A. Naimark, *Normed rings*, Moscow, 1956. (Russian).

15. J. D. Newburgh, *The variation of spectra*, Duke. Math. J. **18** (1951), 165-177.

16. C. E. Rickart, *The uniqueness of norm problem in Banach algebras*, Ann. of Math. **51** (1950), 615-628.

17. ———, *Representations of certain Banach algebras on Hilbert space*, Duke Math. J. **18** (1951), 27-39.

18. B. Yood, *Topological properties of homomorphisms between Banach algebras*, Amer. J. Math. **76** (1954), 155-167.

YALE UNIVERSITY AND UNIVERSITY OF OREGON

Pacific Journal of Mathematics

Vol. 10, No. 1 September, 1960