Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 11, No. 1 **November, 1961**

Printed in Japan by International Academic Printing Co., Ltd., Tokyo, Japan

Reprinted 1966 in the United States of America

GENERALIZED TWISTED FIELDS

A. A. ALBERT

1. Introduction. Consider a finite field $\mathfrak{K}$. If V is any automorphism of $\mathfrak{K}$ we define $\mathfrak{K}_V$ to be the *fixed field* of K under V. Let S and T be any automorphism of $\mathfrak{K}$ and define F to be the fixed field

$$\mathfrak{F} = \mathfrak{F}_q = (\mathfrak{K}_S)_T = (\mathfrak{K}_T)_S \ , \tag{1}$$

under both S and T. Then $\mathfrak{F}$ is the field of $q = p^\alpha$ elements, where p is the characteristic of $\mathfrak{K}$, and $\mathfrak{K}$ is a field of degree n over $\mathfrak{F}$. We shall assume that

$$n > 2 \ , \qquad q > 2 \ . \tag{2}$$

Then the period of a primitive element of $\mathfrak{K}$ is $q^n - 1$ and there always exist elements c in $\mathfrak{K}$ such that $c \neq k^{q-1}$ for any element k of $\mathfrak{K}$. Indeed we could always select c to be a primitive element of $\mathfrak{K}$.

Define a product (x, y) on the additive abelian group $\mathfrak{K}$, in terms of the product xy of the field $\mathfrak{K}$, by

$$(x, y) = xA_y = yB_x = xy - c(xT)(yS) \ , \tag{3}$$

for c in $\mathfrak{K}$. Then

$$A_y = R_y - TR_{c(yS)} \ , \qquad B_x = R_x - SR_{c(xT)} \ , \tag{4}$$

where the transformation $R_y = R[y]$ is defined for all y in $\mathfrak{K}$ *by the product* $xy = xR_y$ of $\mathfrak{K}$. Then the condition that $(x, y) \neq 0$ for all $xy \neq 0$ is equivalent to the property that

$$c \neq \frac{x}{xT} \frac{y}{yS} \ , \tag{5}$$

for any nonzero x and y of $\mathfrak{K}$. But the definition of a generating automorphism U of $\mathfrak{K}$ over $\mathfrak{F}$ by $xU = x^q$ implies that

$$S = U^\beta \ , \qquad T = U^\gamma \ . \tag{6}$$

We shall assume that $S \neq I$, $T \neq I$, so that

$$0 < \beta < n \ , \qquad 0 < \gamma < n \ . \tag{7}$$

Then $xy[(xS)(yT)]^{-1} = z^{q-1}$, where

$$1 - q^\beta = (q-1)\delta, \ 1 - q^\gamma = (q-1)\varepsilon, \ z = x^\delta y^\varepsilon \ . \tag{8}$$

Received April 25, 1960. This paper was supported in part by an Esso Educational Foundation Grant and by NSF Grant G-9504.

Thus the condition that $c \neq k^{q-1}$ is sufficient to insure the property that $(x, y) \neq 0$ whenever $xy \neq 0$.

For every c satisfying (5) we can define a division ring $\mathfrak{D} = \mathfrak{D}(\mathfrak{K}, S, T, c)$, with unity quantity $f = e - c$, where e is the unity quantity of $\mathfrak{K}$. It is the same additive group as K and we define the product $x \cdot y$ of D by

$$(9) \qquad xA_e \cdot yB_e = (x, y) .$$

These rings may be seen to generalize the twisted fields defined in an earlier paper.[1]

We shall show that $\mathfrak{D}$ is isomorphic to $\mathfrak{K}$ if and only if $S = T$. Indeed we shall derive the following result.

THEOREM 1. *Let $S \neq I$, $T \neq I$, $S \neq T$. Then the right nucleus of $\mathfrak{D}(\mathfrak{K}, S, T, c)$ is $f\mathfrak{K}_S$ and the left nucleus of $\mathfrak{D}(\mathfrak{K}, S, T, c)$ is $f\mathfrak{K}_T$. If $\mathfrak{L}$ is the set of all elements g of $\mathfrak{K}$ such that $gS = gT$ then $gA_e = gB_e$ and $\mathfrak{L}A_e = \mathfrak{L}B_e$ is the middle nucleus of $\mathfrak{D}$.*

The result above implies that $f\mathfrak{F}$ is the center of $\mathfrak{D}(\mathfrak{K}, S, T, c)$. Since it is known[2] that isotopic rings have isomorphic right (left and middle) nuclei, our results imply that the (generalized) twisted fields $\mathfrak{D}(\mathfrak{K}, S, T, c)$ are new whenever the group generated by either S or T is not the group generated by S and T. In this case our new twisted fields define new finite non-Desarguesian projective planes.[3]

2. The fundamental equation. Consider the equation

$$(9) \qquad A_x A_e^{-1} A_y = A_z ,$$

for x, y and z in $\mathfrak{K}$. Assume that the degree of $\mathfrak{K}$ over $\mathfrak{K}_T$ is m, where we shall now assume that

$$(10) \qquad m > 2 .$$

---

[1] For earlier definitions of twisted fields see the case $c = -1$ in *On nonassociative division algebras*, Trans. Amer. Math. Soc. **72** (1952), 296–309 and the general case in *Finite noncommutative division algebras*, Proc. Amer. Math. Soc. **9** (1958), 928–932. In those papers we defined a product $[x, y] = x(yT) - cy(xT)$ so that $(x, y) = [x, yT^{-1}] = xy - c(yS)(xT)$ is the product (3) with $S = T^{-1}$.

[2] This result was originally given for loops by R. H. Bruck. It is easy to show that, if $\mathfrak{D}$ and $\mathfrak{D}_0$ are isotopic rings with isotopy defined by the relation $QR_{xP} = R_x^{(c)} QR_z$, then the mapping $x \rightarrow (zx)P^{-1}$ induces an isomorphism of the right nucleus $\mathfrak{D}$ onto that of $\mathfrak{D}_0$, and the mapping $x \rightarrow (xz)P^{-1}$ induces an isomorphism of the middle nucleus of $\mathfrak{D}$ onto that of $\mathfrak{D}_0$.

[3] Two finite projective planes $\mathfrak{M}(\mathfrak{D})$ and $\mathfrak{M}(\mathfrak{D}_0)$ coordinatized by division rings $\mathfrak{D}$ and $\mathfrak{D}_0$ respectively are known to be isomorphic if and only if $\mathfrak{D}$ and $\mathfrak{D}_0$ are isotopic. See the author's *Finite division algebras and finite planes*, Proceedings of Symposia in Applied Mathematics; vol. 10, pp. 53–70.

Then the *norm* in $\mathfrak{K}$ over $\mathfrak{K}_T$ of any element k of $\mathfrak{K}$ is

$$\nu(k) = k(kT)\cdots(kT^{m-1}) , \tag{11}$$

and $\nu(k)$ is in $\mathfrak{K}_T$, that is,

$$\nu(k) = [\nu(k)]T \tag{12}$$

for every k of $\mathfrak{K}$. Thus

$$I - (TR_c)^m = I - R_{\nu(e)} = R_d , \tag{13}$$

where

$$d = e - \nu(c) = dT . \tag{14}$$

Now

$$A_e = I - TR_c , \qquad B_e = I - SR_c , \tag{15}$$

and we obtain

$$A_e[I + TR_c + (TR_c)^2 + \cdots + (TR_c)^{m-1}] = R_d , \tag{16}$$

so that

$$I + TR_c + (TR_c)^2 + \cdots + (TR_c)^{m-1} = A_e^{-1}R_d . \tag{17}$$

Our definition (4) implies that

$$R_aA_y = A_yR_a , \qquad R_bB_x = B_xR_b \tag{18}$$

for every x and y of K, providing that

$$a = aT , \qquad b = bS . \tag{19}$$

In particular, $R_dA_y = A_yR_d$, and so (9) is equivalent to

$$A_x[I + (TR_c) + (TR_c)^2 + \cdots + (TR_c)^{m-1}]A_y = A_zR_d . \tag{20}$$

It is well known that distinct automorphisms of any field $\mathfrak{K}$ are linearly independent in the field of right multiplications of $\mathfrak{K}$. Thus we can equate the coefficients of the distinct powers of T in the equation (20). The right member of (20) is $R_{zd} - TR_{cd(zS)}$ and so does not contain the term in T^{m-1} when $m > 2$. It follows that

$$\begin{aligned} R_x[(TR_c)^{m-1}R_y - (TR_c)^{m-2}(TR_c)R_{yS}] \\ - TR_{c(xS)}[(TR_c)^{m-2}R_y - (TR_c)^{m-3}(TR_c)R_{yS}] = 0 . \end{aligned} \tag{21}$$

This equation is equivalent to

$$xT^{m-1}(y - yS) = xST^{m-2}(y - yS) , \tag{22}$$

and so to the relation

$$[(x - xST^{-1})T^{m-1}](y - yS) = 0 . \tag{23}$$

By symmetry we have the following result.

LEMMA 1. *Let T have period $m > 2$. Then the equation $A_x A_e^{-1} A_y = A_z$ holds for some x, y, z in $\mathfrak{K}$ only if $y = yS$ or $x = xST^{-1}$. If S has period $m_0 > 2$ the equation $B_y B_e^{-1} B_x = B_z$ holds for some x, y, z in $\mathfrak{K}$ only if $x = xT$ or $y = yST^{-1}$.*

3. **The nuclei.** The ring $\mathfrak{D} = \mathfrak{D}(\mathfrak{K}, S, T, c)$ has its product defined by

$$x \cdot y = xR_y^{(c)} = yL_y^{(c)}, \tag{24}$$

where

$$R_{yB_e}^{(c)} = A_e^{-1} A_y , \qquad L_{xA_e}^{(c)} = B_e^{-1} B_x . \tag{25}$$

When $S = T$ our formula (3) becomes $(x, y) = xy - c[(xy)S] = xy(I - SR_c)$. But then the ring $\mathfrak{D}_0$, defined by the product (x, y), is isotopic to the field $\mathfrak{K}$. Since $\mathfrak{D} = \mathfrak{D}(\mathfrak{K}, S, S, c)$ is isotopic to $\mathfrak{D}_0$ it is isotopic to $\mathfrak{K}$, and it is well known that $\mathfrak{D}$ is then also isomorphic to $\mathfrak{K}$. Assume henceforth that

$$S \neq T . \tag{26}$$

The right nucleus of $\mathfrak{D}$ is the set $\mathfrak{N}_\rho$ of all elements z_ρ in $\mathfrak{K}$ such that

$$(x \cdot y) \cdot z_\rho = x \cdot (y \cdot z_\rho) , \tag{27}$$

for every x and y of $\mathfrak{K}$. Suppose that $b = bS$ so that

$$A_b = R_b - TR_{c(bS)} = (I - TR_c)R_b, \ A_e^{-1} A_b = R_b . \tag{28}$$

By (18) we know that $R_b B_x = B_x R_b$, and so $R_b(B_e^{-1} B_x) = (B_e^{-1} B_x) R_b$ for every x of $\mathfrak{K}$. By (25) this implies that the transformation

$$R_b = A_e^{-1} A_b = R_{bB_e}^{(c)} \tag{29}$$

commutes with every $L_x^{(c)}$. However, (27) is equivalent to

$$L_x^{(c)} R_{z_\rho}^{(c)} = R_{z_\rho}^{(c)} L_x^{(c)} . \tag{30}$$

Thus $bB_e = b(I - SR_c) = b(e - c) = bf$ is in $\mathfrak{N}_\rho$. We have proved that the right nucleus of $\mathfrak{D} = \mathfrak{D}(\mathfrak{K}, S, T, c)$ *contains* the field $f\mathfrak{K}_S$, a subring of $\mathfrak{D}$ isomorphic to $\mathfrak{K}_S$.

The left nucleus $\mathfrak{N}_\lambda$ of $\mathfrak{D}$ consists of all z_λ such that

$$(z_\lambda \cdot y) \cdot x = z_\lambda \cdot (y \cdot x) \tag{31}$$

for all x and y of $\mathfrak{K}$. This equation is equivalent to

(32) $$L_{z_\lambda}^{(c)} R_x^{(c)} = R_x^{(c)} L_{z_\lambda}^{(c)}$$

for every x of $\mathfrak{K}$. If $a = aT$ then $B_a = (I - SR_c)R_a$, $B^{-1}B_a = R_a = L_{aA_e}^{(c)}$ commutes with every A_y and every $R_x^{(c)}$, and we see that the left nucleus of $\mathfrak{D}(\mathfrak{K}, S, T, c)$ *contains* the field $f\mathfrak{K}_T$ isomorphic to $\mathfrak{K}_T$.

The *middle nucleus* of $\mathfrak{D} = \mathfrak{D}(\mathfrak{K}, S, T, c)$ is the set $\mathfrak{N}_\mu$ of all z_μ of $\mathfrak{K}$ such that

(33) $$(x \cdot z_\mu) \cdot y = x \cdot (z_\mu \cdot y)$$

for every x and y of $\mathfrak{K}$. This equation is equivalent to

(34) $$R_z^{(c)} R_y^{(c)} = R_{z \cdot y}^{(c)} ,$$

where $z = z_\mu$. However, we can observe that the assumption that

(35) $$R_z^{(c)} R_y^{(c)} = R_v^{(c)} ,$$

for some v in $\mathfrak{K}$, implies that $(f \cdot z) \cdot y = f \cdot v = v = z \cdot y$, Hence (34) holds for every y in $\mathfrak{K}$ if and only if

(36) $$A_g A_e^{-1} A_y = A_v ,$$

for every y of $\mathfrak{K}$, where v is in $\mathfrak{K}$ and

(37) $$gB_e = z = z_\mu .$$

If $gS = gT$ then $A_g = R_g - TR_{c(gS)} = R_g - TR_{c(gT)} = R_g - R_g TR_c = R_g A_e$. Then (36) becomes

(38) $$R_g A_y = R_g(R_g - TR_{c(yS)}) = R_{gy} - TR_{c(ySgT)} = A_{gy} .$$

Hence $gB_e = g(I - SR_c) = g - (gS)c = g - (gT)c = gA_e$, and $\mathfrak{N}_\mu$ contains the field of all elements gB_e for $gS = gT$.

We are now able to derive the converse of these results. We first observe that (27) is equivalent to

(39) $$R_y^{(c)} R_z^{(c)} = R_{y \cdot z}^{(c)} ,$$

for every y of $\mathfrak{K}$, where $z = z_\rho$. This equation is equivalent to

(40) $$A_y A_e^{-1} A_u = A_v ,$$

where $z = uB_e$. If the period of T is $m > 2$ we use Lemma 1 to see that, if we take $y \neq yST^{-1}$, then $u = uS$, $z = uB_e = fu$. The stated choice of y is always possible since we assuming that $S \neq T$ and so some element of $\mathfrak{K}$ is not left fixed by ST^{-1}. Thus $\mathfrak{N} = f\mathfrak{K}_S$. Similarly, is the period of S is not two then $\mathfrak{N}_\lambda = f\mathfrak{K}_T$. Assume that one of S and T has period two.

The automorphisms S and T cannot both have period two. For the group G of automorphisms of $\mathfrak{K}$ is a cyclic group and has a unique subgroup $\mathfrak{H}$ of order two. This group contains I and only one other automorphism. If S and T both had period two we would have $S = T$ and so $m = n = 2$, contrary to hypothesis. Thus we may assume that one of S and T has period two. There is clearly no loss of generality if we assume that T has period two, so that the period of S is at least three. By the argument already given we have $\mathfrak{N}_\lambda = f\mathfrak{K}_T$. We are then led to study (40) as holding for all elements y of $\mathfrak{K}$, where $z_\rho = uB_e$. Now

$$(41) \qquad A_e = I - TR_c, \; A_e(I + TR_c) = R_d, \; d = e - c(cT) = dT .$$

But then (40) becomes

$$(42) \qquad [R_y - TR_{c(yS)}](I + TR_c)[R_u - TR_{c(uS)}] = R_{vd} - TR_{cd(vS)} .$$

This yields the equations

$$(43) \qquad y[u - c(cT)(uS)] - (yST)[c(cT)](u - uS) = vd ,$$

$$(44) \qquad yT(u - uS) - yS[u - (uS)c(cT)] = - d(vS) .$$

Hence

$$\begin{aligned} d(yS)[uS - (cS)(cST)(uS^2)] - yS^2T(cS)(cST)(uS - uS^2)d &= vS(dS)d \\ &= (dS)yS[u - (uS)c(cT)] - yT(u - uS)(dS) . \end{aligned}$$

Since this holds for all y we have the transformation equation

$$(45) \qquad \begin{aligned} SR[d(uS) - d(cS)(cST)uS^2] - S^2TR[d(cS)(cST)(uS - uS^2)] \\ = SR[dSu - (dS)(uS)c(cT)] - TR[(u - uS)dS] . \end{aligned}$$

Since $S^2 \neq I$ and $T \neq S$, S^2T we know that the coefficient of S^2T is zero. Thus $(u - uS)dS = 0$ and $u = uS$ as desired. This shows that $\mathfrak{N}_\rho = f\mathfrak{K}_S$.

The *middle nucleus* condition (36) implies that $gS = gT$ if T does not have period two. When T does have period two but S does not have period two the analogous property

$$(46) \qquad L^{(c)}_{x \cdot z} = L^{(c)}_z L^{(c)}_x$$

is equivalent to

$$(47) \qquad B_g B_e^{-1} B_x = B_v ,$$

and we see again that $gS = gT$. This completes our proof of the theorem stated in the introduction.

4. Commutativity. It is known[4] that $\mathfrak{D} = (\mathfrak{K}, S, S^{-1}, c)$ is commutative if and only if $c = -1$. There remains the case where

$$S \neq I,\ T \neq I,\ ST \neq I,\ S \neq T\,. \tag{48}$$

Any $\mathfrak{D}(\mathfrak{K}, S, T, c)$ is commutative if and only if $R_x^{(e)} = L_x^{(c)}$ for every x of $\mathfrak{K}$. Assume first that $\mathfrak{K}_S \neq \mathfrak{K}_T$. There is clearly no loss of generality if we assume that there is an element b in $\mathfrak{K}_S$ and not in $\mathfrak{K}_T$, since the roles of S and T can be interchanged when $\mathfrak{D}(\mathfrak{K}, S, T, c)$ is commutative. Thus we have $b = bS \neq bT$. By (28) we know that $A_b = A_e R_b$ and so we have $R_{bf}^{(c)} = R_b$. Then $L_{bf}^{(c)} = B_e^{-1}B_y = R_b$, where $y = (bf)A_e^{-1}$. It follows that

$$B_g = R_y - SR_{c(yT)} = B_e R_b = (I - SR_c)R_b\,. \tag{49}$$

Then $R_y = R_b$, $y = b$, $c(yT) = c(bT) = cb$, and $b = bT$ contrary to hypothesis.

We have shown that if $\mathfrak{D}(\mathfrak{K}, S, T, c)$ is commutative the automorphisms S and T have the same fixed fields, that is, $b = bS$ if and only if $b = bT$, b is in $\mathfrak{F}$. Thus S and T both generate the cyclic automorphism group $\mathfrak{G}$ of order n of $\mathfrak{K}$ over $\mathfrak{F}$, and S is a power of T. Since $T^{-1} = T^{n-1} \neq S$ there exists an integer r such that

$$0 < r < n-1,\ S = T^r\,. \tag{50}$$

We now use the fact that $R_x^{(c)} = L_x^{(c)}$ for every x of K to see that $A_e^{-1}A_x = B_e^{-1}B_y$ for every x of $\mathfrak{K}$, where $y = xB_eA_e^{-1}$. Also $(TR_c)^n = (SR_c)^n = R_{\nu(c)}$, and our condition becomes

$$\begin{aligned}[I + TR_c + (TR_c)^2 + \cdots + (TR_c)^{n-1}][R_x - TR_{c(xS)}]\\ = [I + SR_c + (SR_c)^2 + \cdots + (SR_c)^{n-1}][R_y - SR_{c(yT)}]\,,\end{aligned} \tag{51}$$

where we have used the fact that $d = e - \nu(c) = dT = dS$. Compute the constant term to obtain the equation

$$R_x - (TR_c)^n R_{xS} = R_y - (SR_c)_u R_{yT}\,. \tag{52}$$

This is equivalent to the relation $x - [\nu(c)](xS) = y - [\nu(c)]yT$ for every x of K, where $y = xB_eA_e^{-1}$. Thus (52) is equivalent to

$$I - SR_{\nu(c)} = B_eA_e^{-1}[I - TR_{\nu(c)}]\,. \tag{53}$$

We also compute the term in T^r in (51). Since $r < n-1$ the left member of this term is $(TR_c)^r R_x - (TR_c)^r R_{xS}$, which is equal to $R^r R_{gc}(R_x - R_{xS})$, where $g = (cT)(cT)^2 \cdots (cT)^{r-1}$. The right member is the term in S, and this is $SR_c(R_y - R_{yT})$. Hence $(x - xS)g = y - yT$, a result equivalent to

[4] See footnote 1.

$$(I - S)R_g = B_e A_e^{-1}(I - T) . \tag{54}$$

Since the transformations $I - T$ and $I - TR_{\nu(c)}$ commute we may use (53) to obtain

$$(I - S)R_g[I - TR_{\nu(c)}] = [I - SR_{\nu(c)}](I - T) . \tag{55}$$

By (48) we may equate coefficients of I, S, T and ST, respectively. The constant term yields $g = e$. The term in S then yields $\nu(c) = e$ which is impossible when S and T generate the same group and $\mathfrak{D} = \mathfrak{D}(\mathfrak{K}, S, T, c)$ is a division algebra.

We have proved the following result.

THEOREM 2. *Let $\mathfrak{D} = \mathfrak{D}(\mathfrak{K}, S, T, c)$ be a division algebra defined for $S \neq I$, $T \neq I$, $S \neq T$. Then $\mathfrak{D}$ is commutative if and only if $ST = I$ and $c = -1$.*

THE UNIVERSITY OF CHICAGO

OPERATIONAL CALCULUS OF LINEAR RELATIONS

RICHARD ARENS

1. Introduction. Let X and Y be linear spaces, and T a linear subspace of $X \oplus Y$. We call T a *linear relation* to indicate our interest in those constructions with T which generalize those carried out when T is single-valued [4].

Properly many-valued linear relations arise naturally from operators T when T^{-1} or T^* is contemplated in cases where they are not single-valued. One advantage of not dismissing T^* when it is not single-valued is that $T^{**} = T$ if and only if T is closed (for the details, see 3.34, below.) A more superficial attraction is that linear relations, even self-adjoint linear relations in Hilbert space can exhibit phenomena (unbounded spectrum, domain $\neq X$) in finite-dimensional spaces which linear operators exhibit only in infinite-dimensional spaces.

We present an outline of the paper. In § 2 we define $p(T)$ where p is a polynomial with coefficients in the field Φ involved in X. We prove that $(pq)(T) = p(T)q(T)$, $(p \circ q)(T) = p(q(T))$, and point out that sometimes $(p + q)(T) \neq p(T) + q(T)$, etc.

In § 3 we turn to relations in dual pairs. In this situation, adjoints can be defined. We build an automorphism $\lambda \to \bar{\lambda}$ of Φ into the theory of dual pairs, so as not to *exclude* the Hilbert space situation, which dual pairs are intended to imitate. (Thus the transpose is a special kind of adjoint.) Closedness is defined algebraically, but in a way compatible with the topological concept. Closure of T^* and other algebraic properties of * are established. Finally, it is shown that if T is closed and its resolvent is not void then $p(T)$ is also closed.

Section 4 considers the self-dual case. We give a simple condition (4.3) always true in Hilbert space, that T^*T be self-adjoint, T being closed. In § 5 we give the spectral analysis of self-adjoint linear relations in Hilbert space. In a $1:1$ manner these correspond to the unitary *operators*, via the Cayley transform. However, it can be shown directly that X is the direct sum of orthogonal subspaces Y, Z which reduce T $(= T^*)$ giving in Z a self-adjoint operator and in Y the inverse of the zero-operator.

2. Linear relations. A *relation* T between members of a set X and members of a set Y is merely a subset of $X \times Y$. For $x \in X$, $T(x) = \{y : (x, y) \in T\}$. The *domain* of T consists of those x such that $T(x)$ is not void. T is called single-valued if $T(x)$ never contains more than one element. The *range* of T is the union of all $T(x)$.

Received April 13, 1960.

If T is as above and $S \subset Y \times Z$, then $S \circ T = \{(x, z) : (x, y) \in T, (y, z) \in S$ for some $y\}$. We shall write this ST. Finally, $T^{-1} = \{(y, x) : (x, y) \in T\}$. The *range* of T is the domain of T^{-1}.

If X and Y are linear spaces over a field Φ then $X \oplus Y$ is $X \times Y$ with the usual linear structure. A *linear relation* T between members of X and members of Y is a linear subspace of $X \oplus Y$. Linearity is characterized by

2.01 $$\alpha T(x_1) + \beta T(x_2) \subset T(\alpha x_1 + \beta x_2)\,, \qquad (\alpha, \beta \in \Phi; x_1, x_2 \in X)\,.$$

The null space of T is the class of x such that $(x, 0) \in T$. It is easy to see that

2.02 *if S and T are linear relations with the same null space, and the same range, then $S \subset T$ only if $S = T$.*

Let L be a linear subspace of X, and λ an element of Φ. Then λ_L denotes the single valued operator defined on L by $\lambda_L = \{(x, \lambda x) : x \in L\}$. The unit of Φ we denote by 1. Thus 1_L has a meaning according to the preceeding agreement. For T a linear relation with range L, we define λT as $\lambda_L T$. The zero of Φ we denote by 0. Thus OT is not O_X, but O_L where L is the domain of T.

Addition of linear relations S, T is defined as follows:

$$S + T = \{(x, y) : y = s + t \text{ for some } s, t \text{ such that } (x, s) \in S, (x, t) \in T\}\,.$$

The linear relations in $X \oplus X$ do not form a linear space, let alone a linear algebra. We list algebraic properties partly for use later, but mainly to call attention, as it were, to those that are lacking.

2.1 THEOREM. *The operations '$\circ$' and $(+)$ are associative, '$+$' is commutative. Let R, S, T be linear relations. Then*

2.11 *domain of* $R = X \Longleftrightarrow 1_X \subset R^{-1}R$;

2.12 *R is single-valued* $\Longleftrightarrow RR^{-1} \subset 1_L$, $L =$ *range of* R;

2.13 $\lambda \in \Phi \Rightarrow \lambda(ST) = (\lambda S)T = S(\lambda T) = ST\lambda_L$, $L =$ *domaiu of* T;

2.14 $R \subset S \Rightarrow R + T \subset S + T$, $RT \subset ST$, $TR \subset TS$, $R^{-1} \subset S^{-1}$;

2.15 $RS + RT \subset R(S + T)$, *with equality when the domain of R coincides with the whole space*;

2.16 $(S + T)R \subset SR + TR$, *with equality when R is single-valued*;

2.17 $(ST)^{-1} = T^{-1}S^{-1}$.

The proof of these may be left to the reader.

We say S and T commute is $ST = TS$. Suppose $SR = RS$, $TR = RT$. Then $(S + T)R \subset R(S + T)$. The equality may not hold, as the example $S = -T = 1_X$, domain of $R \neq X$, will show.

T^n is defined as $T^{n-1}T$, as usual. If T^n appears in a formula where $n = 0$ is allowed, then T^0 stands for 1_X.

These things can all be extended to the case of moduls over a ring Φ. However, we now turn to a lemma whose proof requires that Φ be a field.

For the remainder of § 2, T will denote a linear relation in $X \oplus X$, and for $\lambda \in \Phi$, we write just 'λ' for 'λ_X'.

It is clear that $\alpha_0 + \alpha_1 T + \cdots + \alpha_n T^n$ has for its domain, just the domain of T^n. *This is true even if* $\alpha_n = 0$! If a polynomial p has coefficients $\alpha_0, \alpha_1, \cdots, \alpha_n$, then by $p(T)$ we mean $\alpha_0 + \alpha_1 + \cdots + \alpha_n T^n$ provied $\alpha_n \neq 0$. Otherwise we omit α_n and consider whether $\alpha_{n-1} \neq 0$, *etc.* If $\alpha_n \neq 0$ and $\alpha_i = 0$ for some $i < n$, then it does not matter whether α_i is omitted or not (but we have already agreed to retain it) because, for example $T^3 + 0T = T^3$.

The next lemma settles a little difficulty that arises in the 'multiple-valued' situation. It enables us to include the multiple valued case in the succeeding theorem, whose substance is that the usual laws of algebra apply to the multiplication of linear polynomials in T. The importance of this theorem is based on the natural fear that even in the single valued case (see 2.15, 2.16), factoring might produce a proper extension of the "multiplied-out" polynomial.

2.2 LEMMA. *Let* $(x, y) \in \alpha_0 + \alpha_1 T + \cdots + \alpha_n T^n$, *where* $\alpha_n \neq 0$. *Then there exist* $y_0, y_1, \cdots, y_n$ *such that*

2.21 $$y_0 = x, \sum_{i=0}^{n} \alpha_i y_i = y$$

and

2.22 $$(y_{i-1}, y_i) \in T \qquad (i = 1, \cdots, n)\,.$$

Proof. Assume that for some j, we have $y_0, y_1, \cdots, y_n$ such that 2.21 holds, and (instead of 2.22)

(j) $$(y_{i-1}, y_i) \in T \qquad (1 \leqq i \leqq j)$$

and

$$(x, y_i) \in T^i \qquad (1 \leqq i \leqq n)\,.$$

Let k be the next integer greater than j such that $\alpha_k \neq 0$. We shall establish (k). This will suffice to prove the lemma.

Because $\alpha_k \neq 0$ we can find $\lambda_1, \cdots, \lambda_j$ such that, for $1 \leqq h \leqq j$,

$$\sum_{m=k-j+h}^{k} \alpha_m \lambda_{j-k+m+1-h} = \alpha_h\,.$$

We can find $z_1, z_2, \cdots, z_k$ where $z_k = y_k$ and $(x, z_1), (z_1, z_2), \cdots, (z_{k-1}, z_k) \in T$. This implies that $(0, y_1 - z_1) \in T$, and $(y_{i-1} - z_{i-1}, y_i - z_i) \in T$ for $i \leqq j$.

Now we define $w_0, w_1, \cdots, w_n$ as follows. $w_0 = x, w_1 = z_1$, for $1 \leqq m \leqq k$,

2.23 $$w_m = z_m + \sum_{i=1}^{j-k+m} \lambda_i (y_{j-k+m+1-i} - z_{j-k+m+1-i})$$

while $w_{k+1} = y_{k+1}, \cdots, w_n = y_n$. It is clear that $(w_{i-1}, w_i) \in T$ for $i \leqq k$, and $(x, w_i) \in T^i$ for all i. There remains only the question, does $\sum \alpha_i w_i = y$, or, equivalently, does

2.24 $$\sum_{m=1}^{k} \alpha_m (w_m - y_m) = 0?$$

The sum in 2.24 has the value

$$\sum_{m=1}^{k-1} \alpha_m (z_m - y_m) + \sum_{m=1}^{k} \sum_{i=1}^{j-k+m} \alpha_m \lambda_i (y_{j-k+m+1-i} - z_{j-k+m+1+i}) .$$

It is not hard to verify that for $0 \leqq h < k$ the coefficient of $y_h - z_h$ in this sum is

2.25 $$-\alpha_h + \sum_{m=k-j+h}^{k} \alpha_m \lambda_{j-k+m+1-h} ,$$

where the $\sum$-term is understood to be absent when $k - j + h > k$. These λ were chosen in order to make this vanish for $0 \leqq h \leqq j$. For $j < h < k, \alpha_h = 0$; since $k < k - j + h$, the $\sum$ term is absent. Thus the sum in 2.24 is 0, and this concludes the proof of the Lemma (2.2).

N.B. This lemma does not imply that T could be cut down to a linear operator U whose domain contains $c, Ux, \cdots$, and $U^{n-1}x$, where

$$\sum_{m=0}^{n} \alpha_m U^m(x) = y,$$

for x could be 0 and y be not 0.

2.3 THEOREM. *Let p and q be two polynomials with coefficients in Φ. Then*

2.31 $$(qp)(T) = q(T)p(T) .$$

Proof. Suppose the degrees of p and q are m and n respecively. Let $p(\xi) = \alpha_0 + \alpha_1\xi + \cdots + \alpha_m\xi^m$. *Mutatis mutandis*, let the coefficients of q and qp be β_j and γ_k.

Now suppose $(x, y) \in (pq)(T)$. By 2.2 there exist $x_1, \cdots, x_{m+n}$ such that $(x_{k-1}, x_k) \in T$ for $k = 1, \cdots, m + n$ where $x_0 = x$, and $\sum \gamma_k x_k = y$. Let $y_j = \sum_{i=1}^{m} d_i x_{i+j}$ for $j = 0, \cdots, n$. Then $(x, y_0) \in p(T)$ and $(y_{j-1}, y_j) \in T$. Let $z = \sum_{j=0}^{n} \beta_j y_j$, so that $(y_0, z) \in q(T)$. Then $(x, z) \in q(T)p(T)$. But obviously $z = \sum \gamma_k x_k = y$. This shows that $(qp)(T) \subset q(T)p(T)$.

Now suppose $(x, z) \in q(T)p(T)$. Then there must exist y such that $(x, y) \in p(T)$ and $(y, z) \in q(T)$. By 2.2 we can find $x_0, \cdots, x_m$ and $y_0, \cdots, y_n$ (where $x_0 = x$, and $y_0 = y$) such that $\sum \alpha_i x_i = y$ and $\sum \beta_j y_j = z$. We now turn to the free linear space Ξ (over Φ) generated by elements $\xi_0, \cdots, \xi_m, \eta_1, \cdots, \eta_n$. In Ξ we define a linear operator S, whose domain is spanned by $\xi_0, \cdots, \eta_{n-1}$, as follows:
$S(\xi_{i-1}) = \xi_i$ $(i = 1, \cdots, m)$, $S(\eta_0) = \eta_1$, where $\eta_0 = \sum \alpha_i \xi_i$, and $S(\eta_j) = \eta_{j+1}$

($j = 1, \cdots, n-1$). We can map Ξ linearly into X by a mapping f which sends ξ_i into X_i, and η_j into y_j. This mapping has the property that for ξ in the domain of S, $(f(\xi), f(S\xi)) \in T$. Derivable from this is that if r is a polynominal and $r(S)\xi$ is defined some ξ in Ξ then $(f(\xi), f(r(S)\xi)) \in r(T)$. We apply this to $\xi = \xi_0$ and $r = qp$. It is easy to see that $p(S)(\xi_0) = \eta_0$, whence $f(qp(S))(\xi_0) = f(\sum\beta_j\eta_j) = \sum\beta_j y_j = z$, and $(x, z) \in (qp)(T)$.

This completes the proof of 2.3.

[*Further remarks on polynomials of relations.* Inspection of the first argument in the proof of 2.3 yields the following result.

2.32 THEOREM. *Let p and q be as in* 2.3. *Then*

2.33 $$(p + q)(T) \subset p(T) + q(T)\ .$$

The '=' does not always hold. While

2.34 $$(\sum\alpha_i)T = \sum(\alpha_i T)$$

hold when $\sum\alpha_i \neq 0$, it does not hold when $\sum\alpha_i = 0$, some $\alpha_i \neq 0$, and T is not single-valued.

As the assertion connected with 2.34 implies, the reason that 2.33 cannot be strengthened to an inequality, is that $T - T$ is not 0 times some relation, if T is not single-valued. We close this little discourse on the peculiarities of many-valued relations by showing that the difficulty arises only with the terms of highest order.

2.35 THEOREM. *Let p, q be as above, and suppose the sum of their leading coefficients is not* 0. *Then $(p + q)(T) = p(T) + q(T)$.*

Proof. We combine the monomials of like degree on the right, and use 2.34 in each case. Eventually one may have to apply the following

2.36 LEMMA. *If $n \geqq k$ then $T^n = T^n + \lambda(T^k - T)$.*

Proof. Let (x, y) belong to the right side. Then $y = u + v$ where $(x, u) \in T^n + \lambda T^k$ and $(x, v) \varepsilon - \lambda T^k$. From 2.2 we obtain $u_0, \cdots, u_n$ which are successively T-related, $u_0 = x$, $u_n + \lambda_{u_k} = u$. Therefore $\lambda_{u_k} + v \in T^k(0)$, whence $u_n + \lambda_{u_k} + v \in T^k(u_{n-k}) \subset T^n(x)$. Thus $(x, y) \in T^n$.

2.37 THEOREM. *Let q and p be polynomials. Then $(q \circ p)(T) = q(p(T))$.*

Proof. The polynominal $q \circ p$ is the result of substituting p into q, by definition. The leading coefficients may be taken as not zero. We can multiply out the terms $\beta_j p(T)^j$ on the right side, without affecting that sum, by 2.3. (The associative law holds for addition.) We can arrange the sum as a polynominal, by virtue of 2.35 there being in fact at all times a unique term $\alpha_m\beta_n T^{m+n}$ of highest degree. The resulting polynomial is of course $(q\circ p)(T)$, for formal reasons.]

We now make some definitions which coincide with the usual ones for closed operators in F-spaces. We call a linear relation T *resolvable* if T^{-1} is single-valued with domain X (that is, by 2.11, if $T^{-1}T\subset 1_X\subset TT^{-1}$. If $T^{-1}T = 1 = TT^{-1}$ we call T *regular*.)

2.4 PROPOSITION. *The product of (finitely many pairwise) commuting linear relations is resolvable only if, and if, each factor is resolvable.*

Proof. It is inevitable and sufficient to consider the case of two factors. If these are resolvable, so is their product. The criterion $T^{-1}T \subset 1 \subset TT^{-1}$ can be used here.

If on the other hand, a linear relation S is not resolvable, then either $(x, 0) \in S$ for some $x \neq 0$, or the range $\neq X$. Accordingly, TS or ST shares the defect. (This sufficies for 2.4).

The *resolvent set* of a linear relation T is the class of λ in Φ for which $T - \lambda$ (by which we mean $T - \lambda 1_X$) is resolvable; and its complement is the spectrum $\sigma(T)$ of T.

2.5 (*Spectral polynomial theorem*). *Let Φ be algebraically closed, and let p be a polynomial over Φ. Then $\sigma(p(T)) = p(\sigma(T))$, where by the latter is meant the class of $p(\lambda)$, $\lambda \in \sigma(T)$.*

Proof. For $\mu \in \Phi$ we can write

$$p(T) - \mu = \alpha(T - \lambda_1)\cdots(T - \lambda_n),\ \mu = p(\lambda_1) = \cdots = p(\lambda_n)$$

where $T - \lambda_1, \cdots, T - \lambda_n$ commute.
If $\mu \in \sigma(p(T))$ then $p(T) - \mu$ is not resolvable, whence (by 2.4) some $\lambda_i \in \sigma(T)$, or $\mu \in p(T))$. If $\mu \in p(T))$ then $\mu = p(\lambda), \lambda \in \sigma(T)$, and so $\lambda = \lambda_i$ for some i. Then $p(T) - \mu$ has a non-resolvable factor, and so is not resolvable. Therefore $\mu \in \sigma(p(T))$. This proves 2.5.

We have defined the sum (and difference) of two linear subspaces U and V (say) of $X \oplus Y$, but occasionally one is concerned with the linear subspace of $X \oplus Y$ which they span. We will have to use some other symbol for this, and we choose

2.6 $$U \neq V\,.$$

Our purpose is to prove the following

2.61 THEOREM. *The range of $1 - V^{-1}U$ is the null-space of $U \neq V$, and the null-space of $1 - V^{-1}U$ is the domain of $U \cap V$.*

Proof. Let $(x, z) \in 1 - V^{-1}U$. Then $(x, z - x)\varepsilon - V^{-1}U$ whence $(x, y) \in U$ and $(y, x - z)\varepsilon - V^{-1}$, for some y. Therefore $(z - x, -y) \in V$ and so $(z, 0) \in U \neq V$. If moreover, $z = 0$ (so that x is in the null-space) then $(-x, -y)$ and thus (x, y) belongs to V and thus $x \in \operatorname{dom} U \cap V$. The reverse inclusions can be established by reversing the steps of this argument.

3. **Adjoints.** For the formalism of adjoints, it is good to suppose that the field Φ has an involutory automorphism

$$\lambda \to \bar{\lambda} ,$$

and we shall do so. Whether Φ admits a non-trivial involution or not, one *can* base the discussion on the identity. Thus the discussion includes the *transpose*.

Let X, A be two linear spaces over Φ. We shall say X, A *are a* $(\Phi, —)$ *dual pair* (or, more briefly, a dual pair) is there is a non-degenerate bi-additive, Φ-valued form $<, >$ defined on $X \oplus A$, linear in first argument, and semi-linear in the second:

$$\langle x, \lambda a\rangle = \bar{\lambda}\langle x, a\rangle .$$

Let Y, B be another $(\Phi, —)$ dual pair. Let T be a linear relation between elements of X and elements of Y, i.e., let T be a linear subspace of $X \oplus Y$. $X \oplus Y, A \oplus B$ form a $(\Phi, —)$ dual pair, in a natural way:

$$\langle (x, y), (a, b)\rangle = \langle x, a\rangle + \langle y, b\rangle .$$

The *adjoint* T^* is defined as follows:

3.11 $$T^* = \{(b, a) : \langle x, a\rangle = \langle y, b\rangle \text{ for all } (x, y) \in T\} .$$

T^* is (evidently) a linear subspace of $B \oplus A$.

For a linear subspace U of $B \oplus A$ we define

3.12 $$U^* = \{(x, y) : \langle x, a\rangle = \langle y, b\rangle \text{ for all } (b, a) \in U\} .$$

It is usually supposed that 3.12 need hardly be written down, once 3.11 is presented. We mention three obvious properties of this process (or, rather, these processes. See § 4)

3.2 $$T \subset T^{**}, S \subset T \Rightarrow T^* \subset S^*$$

3.21 $$(\lambda T)^* = \bar{\lambda} T^*$$

3.22 $$(T^{-1})^* = (T^*)^{-1} .$$

For a subset M of X, let

3.23 $$M^{\perp} = \{a : \langle x, a\rangle = 0 \text{ for all } x \in M\}$$

while if $M \subset A$ then

3.24 $$M^{\perp} = \{x : \langle x, a \rangle = 0 \text{ for all } a \in M\} .$$

In this sense (*cf.* [4])

3.3 $$T^* = (-T^{-1})^{\perp} .$$

In 3.3 we have in mind the natural pairing of $Y \oplus X$ and $B \oplus A$, of course.

Again, considering X, A as a typical pair, and M a linear subspace of X, we define $M^{\perp\perp}$ as the *closure* of M. This requires no topology in X, A, or Φ, and resembles the Stone topology [1, p. 466] in this respect—and in fact admits a natural, joint generalization.

M is *closed* if $M = M^{\perp\perp}$, and *dense* if $M^{\perp\perp} = X$.

PROPOSITION.

3.31 *The null-space of* $T^* =$ (*range of* $T)^{\perp}$

3.32 T^* *is single-valued only if and if the domain of* T *is dense*

3.33 T^* *is closed*

3.34 T^{**} *is the smallest closed linear relation containing* T.

Here 3.31 is easily established on the definitions, and 3.32 follows from it by considering the null space of T^{*-1}. 3.33 is obvious, because any $M^{\perp}$ is closed, while 3.34 follows from 3.33.

Turning to the adjoint of a sum, let S and T be two linear subspaces of $X \oplus Y$. It is quite elementary that

3.4 $$S^* + T^* \subset (S + T)^* .$$

The following gives an unsymmetric condition which insures the equality.

3.41 THEOREM. *If the domain of* $S^* = B$, *and the domain of* S *includes that of* T, *then*

$$(S + T) = S^* + T^* .$$

Proof. Let $(b, a) \in (S + T)^*$. Then there is an element a_1 such that $(b, a_1) \in S^*$. Let us show that $(b, a - a_1) \in T^*$. To this end, suppose $(x, t) \in T$. Then $(x, s) \in S$ for $s = S(x)$, and $(x, s + t) \in S + T$. Now

$$\begin{aligned}\langle x, a - a_1 \rangle - \langle t, b \rangle &= \langle x, a \rangle - \langle x, a_1 \rangle - \langle t, b \rangle \\ &= \langle x, a \rangle - \langle s, b \rangle - \langle t, b \rangle = \langle x, a \rangle - \langle s + t, b \rangle = 0 .\end{aligned}$$

Thus $(b, a - a_1) \in T^*$, which, with $(b, a_1) \in S^*$ gives $(b, a) \in S^* + T^*$ as was to be shown.

Although our T is not a *function*, we may adapt a symbolism usually used in a functional context, and write

$$X \overline{\quad\quad}_T Y \text{ , or } Y_T \overline{\quad\quad} X \text{ ,}$$

to convey that T is a linear subspace of $X \oplus Y$.

If we introduce S

$$Y \overline{\quad\quad}_S Z$$

where Z, C is another $(\Phi, —)$ dual pair, then

$$S \overline{\quad\quad}_{ST} Z \text{ , and } C \overline{\quad\quad}_{(ST)^*} A \text{ .}$$

Since $A\ {}_{T^*}\overline{\quad\quad} B\ {}_{S^*}\overline{\quad\quad} C$ we also have $C \overline{\quad\quad}_{T^*S^*} A$ and there arises the question of the relation of $(ST)^*$ and T^*S^*. In fact, it is quite elementary that $(ST)^* \supset T^*S^*$, but we wish to examine also the reverse inclusion, which is initiated by the following lemma. Here f_a (for example) is the linear functional on X defined by $f_a(x) = \langle x, a\rangle$, etc.

3.5 LEMMA. *Let $c \in C$, $a \in A$. Consider these linear functionals defined in Y*

3.51 $$f_c \circ S \text{ , } f_a \circ T^{-1} \text{ .}$$

Then $(c, a) \in (S \circ T)^*$ if and only if these functionals are single-valued and agree on the intersection of their domains; and $(c, a) \in T^* \circ S^*$ if and only if they have a common extension to some f_b, $b \in B$.

Proof. The second assertion is the easier to show. If $(c, a)\, T^* \circ S^*$ then $(c, b) \in S^*, (b, a) \in T^*$ for some $b \in B$. Let $y \in D(S) \cap D(T^{-1})$ ('D' means 'domain'). I say these functionals (3.51) agree with f_b for such y. Indeed, if $(y, z) \in S$ and $(y, x) \in T^{-1}$ then $f_c(z) = \langle z, c\rangle = \langle y, b\rangle = \langle x, a\rangle = f_a(x)$.

Conversely, if b having this property exists, then $(c, b) \in S^*$ and $(b, a) \in T^*$ or $(c, a) \in T^* \circ S^*$.

Now let $(c, a) \in (S \circ T)^*$, and let $y \in D(S) \cap D(T^{-1})$. Let $(y, z) \in S$, $(x, y) \in T$. Then $(x, z) \in S \circ T$ and $\langle x, a\rangle = \langle z, c\rangle$, and these are generic elements of $(f_a \circ T^{-1})(y)$, $(f_c \circ S^{-1})(y)$ respectively. Thus 3.51 are single-valued, and agree on $D(S) \cap D(T^{-1})$. The converse is obvious.

This establishes 3.5.

From this, a useful conclusion may be drawn.

3.52 PROPOSITION. *Suppose either that the domain of S^* is C, or that the range of T^* is A. Then*

$$(S \circ T)^* = T^* \circ S^* \text{ .}$$

Proof. Let $(c, a) \in (S \circ T)^*$. Consider the case in which the domain of S^* is c. Then $(c, b) \in S^*$ for some b. Let $(y, z) \in S$. Then $(f_c \circ S)(y) = \langle z, c\rangle = \langle y, b\rangle$, i.e., f_b is an extension of $f_c \circ S$. Hence it is also an extension of $f_a \circ T^{-1}$ (the latter confined, if need be, to the domain of $S + T^{-1}$.) We apply 3.5, and obtain $(c, a) \in T^* \circ S^*$.

If the range of T^* is A, the proof is similar. But it may be reduced to the case treated, by using 3.22, and the general fact $(U \circ V)^{-1} = V^{-1} \circ U^{-1}$.

We may now drop the '$\circ$' again, which was reintroduced to make 3.5 easier to present.

3.6 Proposition. *Let U be a linear subspace of $X \oplus Y$, and V, of $Y \oplus Z$. If either the domain of U^{**} is X, or the range of V^{**} is Z, then $(VU)^{**} \subset V^{**}U^{**}$.*

Proof. In any case $U^*V^* \subset (VU)^*$ and $(VU)^{**} \subset (U^*V^*)^*$. We think of U^* as S and V^* as T and apply 3.52, *mutatis mutandis*.

We recall (3.34) that T is closed precisely when $T \supset T^{**}$. The merit of our "many-valued" approach is that this criterion is available whether T^* is single-valued or not.

3.7 Theorem. *Let S and T be linear relations as above. Suppose they are closed, and that either the domain of T is X or the range of S is Z. Then ST is closed.*

Proof. By 3.6, we obtian $(ST)^{**} \subset S^{**}T^{**} = ST$ provided the domain of T is X or the range of S is Z, which suffices.

The relevance of the existence of resolvent values, to the question of closedness of polynomials in a (closed) operator, was noticed by Taylor [3] (see also [2, p. 56]).

3.8 Theorem. *Let T be a closed linear subspace of $X \oplus X$, for which there is at least one $\lambda \in \Phi$ such that $T - \lambda$ has range X. Then $p(T)$, for any polynomial p over Φ, is closed.*

Proof. By the algebraic Theorem 2.3 we have

$$[p - p(\lambda)](T) = (T - \lambda)q(T)$$

where q is a polynomial of degree less than that of p. By 3.7 and an obvious inductive approach, we see that $[p - p(\lambda)](T)$ is closed. Now $[p - p(\lambda)](T) = p(T) - p(\lambda)$ by 2.35, so the latter is closed. Note that $p(T) = U + V$ where $U = p(T) - p(\lambda)$, $V = p(\lambda)$.

Now $(U + V)^* \supset U^* + V^*$ and so $(U + V)^{**} \subset (U^* + V^*)^*$. Let V^* be the S of 3.41. Then its domain is the whole space, while $S^* = V$ and *its* domain is also the whole space. Thus $(U + V)^{**} \subset U^{**} + V^{**} = U + V$, so that $p(T)$ is closed. Of course, we also know that

$$p(T) = p(\lambda) + (T - \lambda)p(T)$$

which does not emerge from the proof given in [2].

4. Self-duality. When X, A is a $(\Phi, -)$ dual pair and $A = X$, we speak of a self-dual pair. This situation presents two definitions of $M^\perp$, that given by 3.23, and another, which we might call ${}^\perp M$, given by 3.24. These coincide if and only if

4.1 $$\langle x, y\rangle = 0 \text{ if and only if } \langle y, x\rangle = 0$$

which, in turn, is equivalent to

4.11 There exists a $p \in \Phi$ such that $p\bar{p} = 1$ and

$$\langle y, x\rangle = p\overline{\langle x, y\rangle} \text{ for all } x, y \in X\,.$$

(We leave the proof of this equivalence to the reader. One should note that 4.1 for X is transmitted, *via* 4.11, to $X \oplus X$, so that when $T \subset X \oplus X$, $T^\perp = {}^\perp T$ when 4.1 holds.)

The situation $M^\perp \neq {}^\perp M$ would not be awkward if one had ${}^\perp(M^\perp) = ({}^\perp M)^\perp$, but for all we know this condition might be equivalent to 4.1. In any case, it does not hold in general (see 5.41).

We therefore *assume* 4.1 in this section.

Let T be a linear subspace of $X \oplus X$. Then $W = T \mp T^\perp$ (see 2.6) is of interest, because for closed relations in Hilbert space, $W = X \oplus X$.

In general, the following relations hold:

4.2 $$\begin{array}{ccc} \multicolumn{2}{c}{W = X \oplus X} & \\ \Downarrow & \Downarrow & \\ \text{null-space of } W = X & W \text{ is dense} & \\ \Downarrow & \Downarrow & \Downarrow \\ \multicolumn{2}{c}{\text{null-space of } W \text{ is dense}} & T^\perp \cap T(0,0)\,. \end{array}$$

We proceed to generalize a proposition of von Neumann's [5].

4.3 THEOREM. *Let T be closed. Let $W = T \mp T^\perp$ and suppose that the null-space of W is all of X. Then the null-space of* $1 + T^*T$ *is* (0), *the range is* X, *and* $(T^*T)^* = T^*T$ (i.e., T^*T is self-adjoint.)

Proof. Let U (in 2.61) $= T$, and $V = T^\perp$. Then $-V^{-1} = T^*$. Therefore the range of $1 + T^*T$ is the null-space of W, that is, X. Moreover, the null-space of $(1 + T^*T)^*$ is (by 3.31) (range of $1 + T^*T)^\perp$, which is (0).

We know that $T^*S^* \subset (ST)^*$ in general, so if we set $S = T^*$, $S^* = T^{**} = T$, we get $T^*T \subset (T^*T)^*$, or $1 + T^*T \subset (1 + T^*T)^*$. Here we have used 3.41.

Considering 2.02, and what we know about the null-spaces and ranges, we conclude that $1 + T^*T = (1 + T^*T)^*$, $T^*T = (T^*T)^*$.

We have already defined T to be self-adjoint if $T = T^*$. We call

T unitary if $T^* = T^{-1}$. We say nothing about single-valuedness. In the Hilbert-space-situation, there are no unitary linear relations except those single-valued relations which are usually called unitary, as the following shows.

4.4 PROPOSITION. *$T^{-1} \subset T^*$ if and only if $\langle x, x\rangle = \langle y, y\rangle$ for all $(x, y) \in T$. If $T^* = T^{-1}$ and $T \mp T^{\perp} = X \oplus X$ then the domain and range of T both equal X.*

Proof. The statement about $\langle x, x\rangle$ and $\langle y, y\rangle$ is obviously true.

Now assume $T \mp T^{\perp} = X \oplus X$ and $T^* = T^{-1}$. Let $y \in X$. Then $(0, y) = (x, t) + (-x, y - t)$ where $(x, t) \in T$ and $(-x, y - t) \in T^{\perp} = (-T^*)^{-1} = -T$, or $(x, y - t) \in T$. Then $(2x, y) \in T$, or the given y is in the range of T. Now the things assumed about T are inherited by T^{-1} so that the range of T^{-1} is also X.

Returning briefly to the Hilbert-space-situation, if $T^* = T^{-1}$ then T is closed and so $T \mp T^{\perp}$ does equal $X \oplus X$, whence T *is* unitary in the usual sense.

To generalize the formal aspects of the Cayley transform [4] we *assume now that Φ contains an element i such that $i^2 = -1$ and $\bar{i} = -i$.*

Cayley's *map* sends $X \oplus X$ onto $X \oplus X$ thus

$$C(x, y) = (x - iy, x + iy) .$$

Its third iterate is scalar, and it preserves orthogonality, *etc.* If $T \subset X \oplus X$ then

$$C(T) = \{(s - it, s + it) : (s, t) \in T\}$$

is the *Cayley transform* of T.

We list several elementary properties.

4.51 $$S \subset T \Longleftrightarrow C(S) \subset C(T)$$

4.52 $$C(-T) = C(T)^{-1}$$

4.53 $$C(T^{-1}) = -C(T)^{-1}$$

4.54 $$C(T^{\perp}) = C(T)^{\perp}$$

4.55 $$C(T^*) = C(T)^{*^{-1}} .$$

4.6 THEOREM. *$T \subset T^*$ if and only if $C(T)^{-1} \subset C(T)^*$, $T = T^*$ if and only if $C(T)$ is unitary.*

If $C^2(T)$ were unitary, and we were in Hilbert space, then T would have a spectral resolution, but $C^2(T)$ is unitary if and only if $T^* = -T$.

The spectral mapping theorem holds for this Cayley transform:

4,7 $$\sigma(C(T)) = \{(1 + i\tau)(1 - i\tau)^{-1} : \tau \in (T)\}$$

with the following understanding: $\infty \in \sigma(S)$ means $0 \in \sigma(S^{-1})$, $2/0 = \infty$, $(1 + i\infty)(1 - i\infty)^{-1} = -1$. Moreover, *eigenvalues* correspond to *eigenvalues*.

The set consisting of the spectrum of T, plus the symbol ∞ if $0 \in \sigma(T^{-1})$ we call, following Taylor, the *augumented* spectrum. *The augmented spectrum thus contains ∞ whenever T is not single-valued.*

5. **Hilbert space.** In Hilbert space X, (Φ = complex numbers), self-adjoint linear relations T may be analyzed in just the same way as the single-valued ones are, by von Neumann, in [4]. The general theory is perfect in a way that the usual theory is not: every unitary operator is the Cayley transform of a unique self-adjoint linear relation, and conversely (4.6).

However, rather than repeat the application of the Cayley transform method, we prefer to analyze the general self-adjoint linear relation in term of self-adjoint operators.

If T is a closed linear subspace of $X \oplus X$, X being a Hilbert space (as shall be assumed in all of this section) then

5.1 $$T = T_\infty \pm T_1$$

where T_∞, T_1 are *orthogonal* closed linear subspaces (so we write '$\pm$' instead of '$\mp$') and $T_\infty = T \cap (\{0\} \oplus X)$. Thus T_∞ has only 0 in its domain, while its range is $T(0)$ (see § 2). $T(0)$ is closed, since $T_\infty = \{0\} \oplus T(0)$. The domain of T_1 is the domain of T, and T_1 is single-valued.

5.2 LEMMA. $T(0) = (\text{dom } T^*)^\perp$, dom T_1 *is dense in* $T^*(0)^\perp$, *and the range of* T_1 *lies in* $T(0)^\perp$.

Proof. 3.31 tells us that $T^{*^{-1}}(0) = (\text{dom } T^{-1})^\perp$. We can replace T here by T^{-1}, and then replace T^* by T since T is closed. Thus $T(0) = (\text{dom } T^*)^\perp$. From $T^*(0) = (\text{dom } T)^\perp$ we obtain $(\text{dom } T)^{-1} = T^*(0)^\perp$, and thus the second assertion. Finally, if $(x, y) \in T_1$, and $(0, z) \in T_\infty$ then $(x, y) \perp (0, z)$, because T_1 is the orthogonal complement of T_∞ relative to T. Hence $\langle y, z\rangle = 0$.

5.3 THEOREM. *Let T be a self-adjoint linear subspace of $X \oplus X$. Let $T = T_\infty \pm T_1$ as above. Then*

$$X = Y \pm Z$$

and T_∞ consists of all pairs $(0, y)$, $y \in Y$ while T_1 is a closed linear operator whose domain is dense in Z, and whose range is in Z. T_1, restricted to Z, coincides with a self-adjoint linear operator in Z.

Proof. Let $Y = T(0), Z = T(0)^{\perp}$. Then the domain of T_1 is dense in $T^*(0)^{\perp} = Y^{\perp} = Z$ and the range lines in $T(0)^{\perp} = Z$, all by 5.2.

Suppose that $(z, w) \in S^*$ where S is T_1 restricted to Z. Then $\langle x, w\rangle = \langle v, z\rangle$ for all $(x, v) \in T_1$. Each $(x, u) \in T$ is of the form $(x, y + v)$ where $y \in T(0)$ and $(x, v) \in T_1$. Now $\langle y, z\rangle = 0$, so $\langle x, w\rangle = \langle y + v, z\rangle$ for all $(x, y + v) \in T$. It follows that $(z, w) \in T^* = T$. But since $z, w \in Z$ we have $(z, w) \in T_1$. This proves 5.3.

We return here to the question raised in second paragraph of § 4, because a counterexample in a Hilbert space context is more desireable than any other. Let $X = L_2\,[0, 2]$, in which the inner product will be denoted by $\langle\, ,\rangle$, and orthogonality, by $\perp$. Select a bounded operator T, domain X, range dense, with single-valued inverse, and define a self-dual pairing by means of the formula

5.4 $$[f, g] = \langle Tf, g\rangle = \langle f, T^*g\rangle\,.$$

The associated orthogonality will be denoted by '$\circ$' to prevent confusion with '$\perp$' already present.

5.41 PROPOSITION. *It is possible to select T and M (a linear subspace of X) such that*

5.42 $$^\circ(M^\circ) = M \text{ but } (^\circ M)^\circ \neq M\,.$$

Before deciding on a specific T we shall establish

5.43 LEMMA. *$^\circ(M^\circ)$ is the closure of M in the norm $\|x\|_T = \|Tx\|$* [4, 298], *and $(^\circ M)^\circ$ is the closure of M in $\|\cdot\cdot\|_{T^*}$.*

Proof. $M^\circ = \{a : [M, a] = 0\} = {}^{\perp}(TM)$, and $^\circ M = {}^{\perp}(T^*M)$. Consequently $^\circ(M^\circ) = {}^{\perp}[T^{*\perp}(TM)]$, and so $g \in {}^\circ(M^\circ)$ precisely when $g \perp T^{*\perp}(TM)$ or $Tg \perp {}^{\perp}(TM)$, i.e.,

5.44 $$Tg \in (TM)^{\perp\perp} = \overline{TM}\,.$$

But this characterizes the closure of M in $\|\cdot\cdot\|_T$, and this observation suffices to establish 5.43.

Now we select $T = J$ where

$$(Jf)(t) = \int_0^t f(\tau)d\tau\,.$$

This J meets our requirement for T. We have

$$(J^*f)(t) = \int_t^2 f(\tau)d\tau\,,$$

whence $J^* = E - J$ where E is the projection on the constant functions in X.

Let N be the linear subspace of those functions that vanish on [1, 2]. Let

$$h(t) = \begin{cases} 1 & 0 \leqq t < 1 \\ 0 & 0 \leqq t \leqq 2 \, . \end{cases}$$

Then $h \in N$ and $M = N \cap \{h\}^{\perp} \neq N$. Thus $EM = (0)$. It is easy to establish, in the order given, the following: $JM \subset N, J^*N \subset N, \overline{JM} = N, \overline{J^*M} = N$.

Then one observes that $Jf \in N$ implies $f \in M$ while $J^*f \in N$ implies $f \in N$, (and each converse holds, because $JM \subset N, J^*N \subset N$.) Using 5.44 as a criterion for $Jg \in {}^\circ(M^\circ)$ we obtain ${}^\circ(M^\circ) = M, ({}^\circ M)^\circ = N$.

BIBLIOGRAPHY

1. Richard Arens, *A generalization of normed rings*, Pacific J. Math., **2** (1952), 455-471.

2. E. Hille and R. S. Phillips, *Functional Analysis and Semi-groups*, A.M.S. Collow. Publ., **31**.

3. A. E. Taylor, *Spectral Analysis of closed distributive operators*, Acta Math., **84** (1950), 189-224.

4. J. von Neumann, *Über adjungierte Funktional-operatoren*, Ann. Math., **33**, (1932), 294-310.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

DISCONJUGACY OF A SELF-ADJOINT DIFFERENTIAL EQUATION OF THE FOURTH ORDER[1]

JOHN H. BARRETT[2]

Introduction. In a recent paper [10] W. Leighton and Z. Nehari investigated oscillation properties of solutions of self-adjoint differential equations of the fourth order

$$(r(x)y'')'' + (q(x)y')' + p(x)y = 0$$

with particular attention to the cases where the middle term is missing, $r(x) > 0$ and $p(x)$ does not change sign. In the present paper one of these particular cases

$$(1) \qquad (r(x)y'')'' - p(x)y = 0$$

($r(x)$ and $p(x)$ positive and continuous on $[a, \infty)$) will be pursued further with the object of paralleling the known theory of second order equation

$$(2) \qquad (r(x)y')' + p(x)y = 0$$

with positive and continuous coefficients (e.g., see [2] and [12]). With only occasional minor modifications the terminology of [10], together with the fundamental properties of (1) established there, will be assumed throughout this paper. One point of departure is the distinction between "disconjugacy" and "non-oscillation" as the author has used them previously [2] for equation (2) in discussions which will be extended here to the fourth-order equation (1). It will be said that equation (1) is

(i) *disconjugate* if no nontrivial solution has more than 3 zeros on $[a, \infty)$ and, hence, no conjugate pairs exist on $[a, \infty)$ in the sense of Leighton and Nehari [10],

(ii) *oscillatory* if there is a nontrivial solution with infinitely many zeros on $[a, \infty)$.

(iii) *nonoscillatory* if every nontrivial solution has at most a finite number of zeros on $[a, \infty)$.

Recently, W. J. Coles [5] has developed Wirtinger-type inequalities in relation to the higher order equation

Received October 20, 1959.

[1] This work was sponsored by the Office of Ordnance Research, U. S. Army, Contract DA-04-495-ORD-1088 with the University of Utah. Presented to the Amer. Math. Soc., January, 1960.

[2] Member, Mathematics Research Center, U. S. Army, Madison, Wisconsin for the 1959-60 academic year.

$$(r(x)y^{(m)})^{(m)} + (-1)^{m+1}p(x)y = 0 \qquad (m=1, 2, 3, \cdots)$$

by use of his Riccati systems [4] and in this discussion are included various sets of two-point boundary conditions, one of which is analogous to the well-known focal-point conditions for the second order equation (1) (see [2], [12] and [13]).

Again following the second-order discussions [2], associate with (1) its "reciprocal" equation [10, p. 369]

$$(1^*) \qquad (y''/p(x))'' - (1/r(x))y = 0$$

as was done for (2) with

$$(2^*) \qquad (y'/p(x))' + (1/r(x))y = 0^3 .$$

Note that $y(x)$ is a solution of (1) if, and only if,

$$y_1(x) = r(x)y''(x)$$

is a solution of (1*). Throughout this paper, the subscript "1" on a solution will stand for the leading coefficient times the second derivative of the solution.

In the first section known second-order definitions and theorems will be listed, which will be shown to be true *almost verbatim* for the fourth order case in the second section. The third section contains results following from Wirtinger-type inequalities, which are the fourth-order special cases of the above-mentioned results of Coles, and an extension of the eigenvalue discussion of Leighton and Nehari. The last section contains Coles' general theorem with minor modifications, as utilized in the preceding sections.

1. The second-order case. Consider equation (2) with the stated conditions on its coefficients [2].

DEFINITION 1.1. If a nontrivial solution of (2) satisfies the two-point boundary conditions $y(a) = y(b) = 0$, $a < b$, then the smallest such number b is designated as $\eta_1(a)$ and is called the first (right) *conjugate point* of a. If no such solution and number b exist then equation (2) is said to be *disconjugate.*

DEFINITION 1.2. If a nontrivial solution of (2) satisfies $y(a)=y'(b)=0$, $a < b$, then the smallest such number b is designated by $\mu_1(a)$ and it is said that a is the first (left) *focal point* of $b = \mu_1(a)$. The first two theorems are almost trivial for (2) but their counterparts for (1) require some proof, as is seen in the next section.

[3] Previously utilized by Leighton for boundedness theorems in [8].

THEOREM 1.1. *If $\eta_1(a)$ exists then so does $\mu_1(a)$ and $a<\mu_1(a)<\eta_1(a)$. Furthermore, if (*) denotes the same notation for the reciprocal equation* (2*) *then the existence of $\eta_1(a)$ implies, also, the existence of $\mu_1^*(a)$. In other words, if either $\eta_1(a)$ or $\eta_1^*(a)$ exist then both of $\mu_1(a)$ and $\mu_1^*(a)$ exist.*

THEOREM 1.2. *If $\mu_1(a)$ does not exist then for every solution $y(x)$ of* (2), *for which $y(a)=0$ and $y'(a)>0$, it follows that $y(x)>0$ and $y'(x)>0$ on (a, ∞).*

THEOREM 1.3. [2, p. 554] *If $\mu_1(a)$ exists and $\int^\infty (1/r)=\infty$ then $\eta_1(a)$ exists. Furthermore, for any solution $y(x)$ of* (2), *for which $y'(b)=0$, $y(b)\neq 0$, $a\leqq b$, it follows that $y(x)$ has a zero on (b, ∞).*

This theorem is due to Hille for $r\equiv 1$ and was utilized by Nehari [12]. It is noted that disconjugacy of (2) implies the non-existence of $\mu_1(a)$, if $\int^\infty (1/r)=\infty$. Recall (e.g., [2]) that if $\int^\infty (1/r)<\infty$ then $\mu_1(a)$ can exist even though $\eta_1(a)$ does not—in particular, when $\int^\infty p=\infty$.

THEOREM 1.4. [7] *If $\int^\infty (1/r)=\infty$ and $\int^\infty p=\infty$ then $\eta_1(a)$ exists and, in fact, equation* (2) *is oscillatory* (for this result $p(x)$ may change sign).

The well-known relation of the focal-point problem to quadratic functionals[4] was reiterated recently by W. T. Reid [13], when he gave a concise, self-contained development with applications to new oscillation criteria of (2).

THEOREM 1.5. [2, 13]. *If the number $\mu_1(a)$ does not exist then the quadratic functional*

$$I_2[u;b]=\int_a^b (r(u')^2-pu^2) \tag{3}$$

is strictly positive for every $b>a$ and every function $u(x)$ such that $u(x)$ is absolutely continuous, $u'\in L_2(a,b)$ and u has a zero of at least order one at $x=a$. This conclusion can also be stated as a Wirtinger-type inequality

$$\int_a^b pu^2<\int_a^b r(u')^2\,. \tag{4}$$

2. The fourth-order case. Consider the equations

$$(r(x)y'')''-p(x)y=0\,, \tag{1}$$

[4] For the general classical theory see Morse [**11**] and for the theory for singular functionals see Leighton [**9**].

($r(x)$ and $p(x)$ positive and continuous on $[a, \infty)$), and

$$(y''/p(x)'' - y/r(x) = 0 \ . \tag{1'}$$

The following *conjugate point* definition is that of Leighton and Nehari [10].

DEFINITION 2.1. If a non-trivial solution of (1) satisfies the two-point boundary conditions

$$y(a) = y'(a) = y(b) = y'(b) = 0, a < b \ , \tag{5}$$

then the smallest such number b is designated by $\eta_1(a)$ and is called the first (right) *conjugate point* of a. Recall from [10] that such a number exists if (1) has a nontrivial solution which has double zeros at a and $\eta_1(a)$, is non-zero on $(a, \eta_1(a))$ and any essentially different (linearly independent) solution of (1) has at most 3 zeros on $[a, \eta_1(a)]$.

DEFINITION 2.2. If a nontrivial solution of (1) satisfies

$$y(a) = y'(a) = y_1(b) = y_1'(b) = 0, \ a < b, \tag{6}$$
$$(\text{recall } y_1 = ry'')$$

then the smallest such number b is designated by $\mu_1(a)$. The solutions of (1) which are particularly useful in the following analysis are those whose Wronskian at $x = a$ is

$$\begin{array}{lcccc} & y & y' & y_1 & y_1' \\ y = U(x): & 1 & 0 & 0 & 0 \\ y = V(x): & 0 & 1 & 0 & 0 \\ y = u(x): & 0 & 0 & 1 & 0 \\ y = v(x): & 0 & 0 & 0 & 1 \end{array} \tag{7}$$

By [10, Lemma 2.1] all of y, y', y_1 and y_1' for $y = U, V, u$ and v are positive on (a, ∞).

LEMMA 2.1. [10] *If $y(x)$ and $z(x)$ are solutions of* (1) *then*

$S[y; z] = yz_1' - zy_1' - y'z_1 + z'y_1 \equiv C$, a constant.

In [10] the non-self-adjoint form of the following is established and utilized in establishing conjugate point (oscillation) theorems.

LEMMA 2.2. *If $y(x)$ and $z(z)$ are solutions of* (1) *such that $S[y; z] = 0$ and $y(x) \neq 0$ on $I \subset [a, \infty)$ then $W(x) = yz' - zy'$ satisfies the second-order self-adjoint equation*

$$(rW'/y^2)' + (2y_1/y^3)W = 0 \ on \ I \ . \tag{8}$$

Note that if $y > 0$ and $y_1' > 0$ on I (such solutions exist for any such I) both coefficients in (8) are positive and the results of the first section apply. The following is an example of the importance of (8):

LEMMA 2.3. [10. Th. 3.11] *If $y(x)$ is a solution of* (1) *having at most a finite number of zeros then* (1) *is oscillatory if, and only if,* (8) *is oscillatory.*

The present discussion will be concerned with relations between the last two solutions, $u(x)$ and $v(x)$, defined by (7), since every solution of (1) having a double zero at $x = a$ can be expressed by

$$y = y_1(a)u(x) + y_1'(a)v(x) .$$

As in [10], note that in order for any nontrivial solution of (1) to have both a double zero at $x = a$ and at $x = b$ (i.e., satisfy boundary conditions (5)) it is necessary and sufficient that

$$\sigma(x) = u(x)v'(x) - u'(x)v(x) \tag{9}$$

vanish at $x = b$. Furthermore, $\eta_1(a)$ (of Definition 2.1) is the smallest such $b > a$. Observe further that in order for a nontrivial solution of (1) to satisfy the conditions (6) it is necessary and sufficient that

$$\rho(x) = u_1(x)v_1'(x) - u_1'(x)v_1(x) \tag{10}$$

vanish at $x = b$. The stage is now set for the verbatim fourth-order analog of Theorem 1.1.

THEOREM 2.1. *If $\eta_1(a)$ exists then so does $\mu_1(a)$ and $a<\mu_1(a)<\eta_1(a)$. Furthermore, if* (*) *denotes the same notation for the reciprocal equation* (1*) *then $\eta_1(a)$ implies, also, the existence of $\mu_1^*(a)$.*

Proof. Let $Z(x)$ be a solution of (1) having double zeros at $x = a$ and $x = \eta_1(a)$ and positive in $(a, \eta_1(a))$. There exist inflection points $x_1 < x_2$ of $y = Z(x)$ on $(a, \eta_1(a))$ such that $Z''(x_1) = Z''(x_2) = 0$ and $Z''<0$ on (x_1, x_2). Recall that $u(x)$ is the solution of (1) which satisfies

$$u(a) = u'(a) = u_1'(a) = 0, \ u_1(a) = 1 .$$

By use of a fundamental technique of [10], since $(Z''/u'')'$ must change sign on (x_1, x_2), say at $x = b$. There exists a number λ such that the solution $y = Z(x) - \lambda u(x)$ satisfies the boundary conditions (6) and $a<\mu_1(a) \leqq b < \eta_1(a)$[5]. For the second part consider the pair of solutions U, u. Since $S[u; v] = 0$ then $\sigma(x) = uv' - vu'$ is a solution of

[5] This result may also be established easily by means of an indirect argument using equation (8_1) following.

$$(8') \qquad \left(\frac{r\sigma'}{u^2}\right)' + \frac{2u_1}{u^3}\sigma = 0 \quad \text{and} \quad \left(\frac{r\sigma'}{v^2}\right)' + \frac{2v_1}{v^3}\sigma = 0 \quad \text{on} \quad (a, \infty)\,.$$

Note that $x = a$ is a singular point. Similarly, $S[U, u] = 0$ and $\sigma_1(x) = Uu' - uU'$ is a solution of (8′) and

$$(8'') \qquad \left(\frac{r\sigma_1'}{U^2}\right)' + \frac{2U_1}{U^3}\,\sigma_1 = 0 \quad \text{on} \quad [a, \infty)\,.$$

Finally, $S[U, V] = 0$ and $\sigma_2 = UV' - VU'$ is a solution of (8″). Since $\sigma(a) = 0 = \sigma(\eta_1) = \sigma_1(a)$ and $\sigma_2(a) > 0$, $\sigma_2'(a) = 0$, then simple comparison techniques give that σ_1 has a zero x_1 on (a, η_1) and that $\sigma_2(x)$ has a zero x_2 on $(a, x_1) \subset (a, \eta_1)$. Because of the relationship between (1) and (1*) it is easily seen that the smallest such number x_2 is actually $\mu_1^*(a)$ and the theorem is proved.

While Theorem 1.2 is not true for (1) a similar theorem does hold.

THEOREM 2.2. *If $\mu_1(a)$ does not exist then there exists a solution $y(x)$ such that $y(a) = y'(a) = 0$, $y(x) > 0$, $y'(x) > 0$ $y_1(x) > 0$ and $y_1'(x) < 0$, on (a, ∞).*

The proof will be accomplished by two lemmas concerning the ratios involving the particular solutions $u(x)$ and $v(x)$ of (7).

$$(11) \qquad \lambda_0 = u/v,\ \lambda_1 = u'/v',\ \lambda_2 = u_1/v_1,\ \lambda_3 = u_1'/v_1'\,.$$

LEMMA 2.4. *If $\mu_1(a)$ does not exist then*

$$\lambda_0 > \lambda_1 > \lambda_2 > \lambda_3 > 0 \quad \text{on} \quad (a, \infty)\,.$$

Proof of lemma. That all λ_i are positive is obvious. Also, if Lemma 2.2 is applied to (1*) and its solutions u_1 and v_1 then $\rho(x)$, defined by (10), satisfies

$$(8_1) \qquad \begin{aligned} &(\rho'/pu_1^2)' + 2u\rho/u_1^3 = 0 \quad \text{on} \quad [a, \infty) \\ &(\rho'/pv_1^2)' + 2v\rho/v_1^3 = 0 \quad \text{on} \quad (a, \infty)\,. \end{aligned}$$

Note that $x = a$ is not a singular point of the first equation of (8_1). The following useful relations are derived by routine calculations:

$$(12) \qquad \begin{cases} \lambda_0 - \lambda_1 = \sigma/vv' \\ \lambda_1 - \lambda_2 = \tau/v'v_1 \\ \lambda_2 - \lambda_3 = \rho/v_1v_1' \end{cases} \qquad\qquad (13) \quad \begin{cases} \rho'/p = -r\sigma' \\ (r\sigma')' = 2\tau \end{cases}$$

where

$$(14) \qquad \tau(x) = u'v_1 - v'u_1 \equiv uv_1' - vu_1'\,,$$

the latter identity following from the fact that $S[u; v] = 0$.

Since $\mu_1(a)$ does not exist and $\rho(a) = 1$ then $\rho(x) > 0$ on $[a, \infty)$. Also,

$\rho'(a) = 0$ and, hence, integration of the first equation of (8_1) gives $\rho'(x) < 0$ on (a, ∞). Therefore, $\sigma'(x) > 0$ and $\sigma(x) > 0$ on (a, ∞). In order to show that $\tau(x) > 0$, recall that $(\rho'/p)' = -2\tau$ and from the second equation of (8_1):

$$2\tau/v_1^2 = 2v\rho/v_1^3 - 2v_1'\rho'/pv_1^3 > 0 \quad \text{on} \quad (a, \infty) .$$

Therefore, all of the differences (12) are positive on (a, ∞).

LEMMA 2.5. *If $\mu_1(a)$ does not exist then λ_0, λ_1 and λ_2 are decreasing functions and λ_3 is an increasing function on (a, ∞).*

Proof of lemma. Simple calculations yield

$$\lambda_0' = -\sigma/v^2,\ \lambda_1' = -\tau/r(v')^2,\ \lambda_2' = -\rho/u_1^2,\ \lambda_3' = p\tau/(v_1')^2 \tag{14}$$

from which the result follows immediately.

To complete the proof of Theorem 2.2, let λ^* be a positive number such that

$$\lambda_0(x) > \lambda_1(x) > \lambda_2(x) > \lambda^* > \lambda_3(x) \quad \text{on} \quad (a, \infty) .$$

Then $y(x, \lambda^*) = u(x) - \lambda^* v(x)$ satisfies the required conditions.

An example of (1) for which $\mu_1(a)$ does not exist is $(e^{-2x}y'')'' - e^{-2x}y = 0$ for which direct calculations show that $\rho(x) > 0$ on (a, ∞). This example should be compared with the similar one for the second-order case [2].

THEOREM 2.3. *If $\mu_1(a)$ exists and $\int^\infty (1/r) = \infty$ then $\eta_1(a)$ exists.*

Proof. In addition to the identities (12), (13) and (14) note that

$$(15)\quad \begin{cases} \lambda_0 - \lambda_2 = r\sigma'/vv_1 \\ \lambda_0 - \lambda_3 = \tau/vv_1' \\ \lambda_1 - \lambda_3 = \tau'/v'v_1' \end{cases} \qquad (16)\quad \begin{cases} \tau' = (u'\tau - \sigma u_1')/u \\ \tau' = (u'\rho + \tau u_1')/u_1 \\ \tau'' = \dfrac{1}{r}\rho - p\sigma \quad \text{or} \\ (r\sigma')''' + 2p\sigma = 2\rho/r . \end{cases}$$

and that $\tau(x)$ satisfies the second order self-adjoint equation with positive coefficients:

$$\left(\frac{\tau'}{\sqrt{u'u_1'}}\right)' + \frac{1}{2\sqrt{u'u_1'}}\left(\frac{u_1'}{ru'} + \frac{pu'}{u_1'}\right)\tau = 0 \quad \text{on} \quad (a, \infty) . \tag{17}$$

Assume that the theorem is not true, i.e., (1) is disconjugate. Thus $\sigma(x) > 0$ on (a, ∞) and since $\int^\infty (u^2/r) = \infty$, Theorem 1.3 (second part) may be applied to

$$(8') \qquad (r\sigma'/u^2)' + 2u_1\sigma/u^3 = 0 \quad \text{on} \quad (a, \infty) .$$

to obtain that $\sigma'(x) > 0$ on (a, ∞). Hence $\rho'(x) < 0$ and $\rho(x)$ has only one zero, namely $\mu_1(a)$, on (a, ∞). Since

$$(\tau/u)' = -\sigma u_1'/u^2 < 0 \quad \text{on} \quad (a, \infty),\ \lim_{x\to a}(\tau/u) = \frac{1}{2} > 0 ,$$
$$\tau(a) = \tau'a) = 0 \quad \text{and} \quad \tau''(a) > 0 .$$

Then $\tau(x)$ can have at most one zero on (a, ∞). Suppose first that such a zero $x = t_1$ exists. Then $\tau(x) > 0$ on (a, t_1) and $\tau(x) < 0$ on (t_1, ∞). Note that $\int^\infty \sqrt{u'u_1'} = \infty$ and since (17) is disconjugate, Theorem 1.3 guarantees that its solution $\tau(x)$ has only one point $x = t_1'$ on (a, t_1) where $\tau' = 0$ and $\tau' < 0$ on (t_1', ∞). The first equation of (16) yields that $a < \mu_1 < t_1' < t_1$.

On the half-line (t_1, ∞) equations (12), (14) and (15) yield

$$\lambda_3(x) > \lambda_0(x) > \lambda_2(x) > \lambda_1(x)$$
$$\lambda_3' < 0,\ \lambda_0' < 0, \lambda_2' > 0, \lambda_1' > 0 .$$

Because of the above monotonicity there exists a positive constant λ^* such that on (t_1, ∞),

$$\lambda_3 > \lambda_0 > \lambda^* > \lambda_2 > \lambda_1 .$$

Let $y(x) = u(x) - \lambda^* v(x)$, a solution of (1). Then $y(x) > 0$, $y'(x) < 0$ and $y''(x) < 0$ (since $y_1 < 0$) on (t_1, ∞), which is contradictory information.

Therefore, $\tau(x) > 0$ on (a, ∞) and by Theorem 1.3, $\tau'(x) > 0$. Thus on the half-line (μ_1, ∞):

$$\lambda_0(x) > \lambda_1(x) > \lambda_3(x) > \lambda_2(x)$$
$$\lambda_0' < 0, \lambda_1' < 0,\ \lambda_3' > 0,\ \lambda_2' > 0 .$$

As above there exists a positive constant λ^* between λ_1 and λ_3 on (μ_1, ∞).

For $y(x) = u(x) - \lambda^* v(x)$ on the interval (μ_1, ∞):

$$y(x) > 0,\ y'(x) > 0,\ y_1(x) < 0 \quad \text{and} \quad y_1'(x) > 0$$

which is contradictory, using that $\int^\infty (1/r) = \infty$.

THEOREM 2.4. *If* $\int^\infty (1/r) = \infty$ *and* $\int^\infty pu_1^2 = \infty$ *then* $\eta_1(a)$ *exists and, in fact, equation* (1) *is oscillatory.*

The crucial point of the proof is the following which follows immediately by application of Theorem 1.3 to equation (8_1).

LEMMA 2.6. *If* $\int^\infty pu_1^2 = \infty$ *then* $\mu_1(a)$ *exists.*

Proof of theorem. $\int^\infty pu_1^2 = \infty$ implies that $\mu_1(a)$ exists which, together with $\int^\infty (1/r) = \infty$ and Theorem 2.3 gives that $\eta_1(a)$ exists. Since for any $a_1 > a$, $\eta_1(a_1)$ exists, then by [10] equation (1) is oscillatory.

Because of the monotonicity of $u(x)$ and $u_1(x)$ it follows that

COROLLARY 2.4.1. *If* $\int^\infty (1/r) = \infty$ *and* $\int^\infty p = \infty$ *then* (1) *is oscillatory.*

Further corollaries are obtained by a more careful examination of the properties of $u(x)$ and its derivatives. Integration of (1) and consideration of the initial conditions for $u(x)$ yield

$$u_1(x) = 1 + \overset{x}{\underset{a}{\mathrm{I}}}{}^2(pu) \quad \text{and} \quad u(x) = \overset{x}{\underset{a}{\mathrm{I}}}{}^2(u_1/r)\,, \tag{18}$$

where the Riemann-Liouville notation is used for the iterated integrals. For $a < x_0 \leqq x < \infty$ it follows that

$$\int_{x_0}^\infty (u^2/r) > u_1^2(x_0)\int_{x_0}^\infty (1/r(x))\left[\overset{x}{\underset{x_0}{\mathrm{I}}}{}^2(1/r)\right]^2 dx$$

and

$$\int_{x_0}^\infty pu_1^2 > \int_{x_0}^\infty p(x)\left[1 + u(x_0)\overset{x}{\underset{x_0}{\mathrm{I}}}{}^2 p\right]^2 dx$$

$$= \int_{x_0}^\infty p + 2u(x_0)\int_{x_0}^\infty (p(x)\left(\overset{x}{\underset{x_0}{\mathrm{I}}}{}^2 p\right)dx + u^2(x_0)\int_0^\infty p(x)\left(\overset{x}{\underset{x_0}{\mathrm{I}}}{}^2 p\right)^2 dx\,.$$

COROLLARY 2.4.2. *If* $\int^\infty (1/r(x))\,dx = \infty$ and $\int^\infty p(x)\left(\overset{x}{\underset{x_0}{\mathrm{I}}}{}^2 p\right)^2 dx = \infty$ *then* (1) *is oscillatory.*

Note that for $a < x_0 < x_1 \leqq x < \infty$, $\overset{x}{\underset{x_0}{\mathrm{I}}}{}^2 p \geqq (x - x_1)\int_{x_0}^{x_1} p \to \infty$ as $x \to \infty$ and the following result of Nehari and Leighton follows.

COROLLARY 2.4.3[6]. [10, Th. 6.8] *If* $r(x) \leqq m$ *and* (1) is *nonoscillatory then* $\int^\infty x^2 p(x)dx < \infty$.

In connection with the above oscillation theorems it is appropriate to list two known theorems insuring nonoscillation.

Theorem [10, Th. 6.12] If $P(x) = \overset{\infty}{\underset{x}{\mathrm{I}}}{}^3 p$ and $\int^\infty (P/r) < \infty$ then (1) is non-oscillatory.

Theorem [10, Th. 6.11] If $\int^\infty 1/r < \infty$ and $\int_x^\infty x^2 p(x)dx < \infty$ then (1) is non-oscillatory.

[6] Also, a special case of a more general result for fourth-order systems by Sternberg and Sternberg [**14**].

3. An eigenvalue problem and Wirtinger-type inequalities. Leighton and Nehari have shown [10, Th. 6.6 and 6.7] that equation (1) is disconjugate on $[a, \infty)$ if, and only if, the least eigenvalue $\lambda(b)$ of the "conjugate-point" problem

$$(ry'')'' - \lambda py = 0,\ y(a) = y'(a) = y(b) = y'(b) = 0 .$$

satisfies $\lambda(b) > 1$ for all $b > a$. Furthermore, if equation (1) is disconjugate on $[a, \infty)$ then $I_4[w; b] \geqq 0$ (see equation (25)) for all $w(x)$ of class $D''(a, b) \in L_2(a, b)$ which satisfy $w(a) = w'(a) = w(b) = w'(b) = 0$. Finally, in the spirit of [12] they obtained a number of nonoscillation theorems by taking special examples of such $w(x)$.

Consider here the "focal-point" problem as for the second-order case in [2, 12].

$$(ry'')'' - \lambda py = 0,\ y(a) = y'(a) = y_1(b) = y_1'(b) = 0 . \tag{19}$$

For each $b > a$ let $\lambda_1(b)$ be the least eigenvalue and $y = z(x)$ be a corresponding eigenfunction. Integration by parts gives

$$\int_a^b pz^2 = \lambda_1(b)\int_a^b r(z'')^2 . \tag{20}$$

If there does not exist a number $\mu_1(a)$ then $\lambda = 1$ is not an eigenvalue for any $b > a$ and by Theorem 2.2, there exists a solution $y(x)$ of (1) for which $y(a) = y'(a) = 0$ and $y(x) > 0,\ y'(x) > 0,\ y_1(x) > 0,\ y_1'(x) < 0$ on (a, ∞). This is but a special case of the general theorem of Coles [5], as will be seen in the last section, and it follows[7] that for every $b > a$

$$\int_a^b pu^2 < \int_a^b r(u'')^2, \text{ i.e., } I_4[u; b] > 0 \tag{21}$$

for every function $u(x)$ for which u' is absolutely continuous, $u'' \in L_2(a, b)$ and u has a double zero at $x = a$.

LEMMA 3.1. *The number $\mu_1(a)$ does not exist if, and only if, the eigenvalue $\lambda_1(b) > 1$ for all $b > a$.*

Proof. If $\mu_1(a)$ exists then for $b = \mu_1(a)$, $\lambda(\mu_1(a)) = 1$. If $\mu_1(a)$ does not exist then (20) and (21) yield that $\lambda_1(b) > 1$ for $b > a$ and the lemma is proved.

By combining the above lemma with Theorem 2.3 it follows that (recall the special monotone solution $u(x)$ of (1)):

THEOREM 3.1[8]. If $\int^\infty (1/r) = \infty$ *then equation* (1) *is disconjugate on*

[7] See concluding statement of section 4.

[8] This result has also been obtained by H. C. Howard by means of Rayleigh quotients in his dissertation at Carnegie Institute of Technology, June 1958; to appear in the *Transactions of the American Mathematical Society* **96** (1960), 296-311.

$[a, \infty)$ if, and only if, $\lambda_1(b) > 1$ for all $b > a$ and if (1) is disconjugate then for every $b > a$

$$\int_a^b pw^2 < \int_a^b r(w'')^2 \tag{21}$$

for all $w(x)$ such that $w'(x)$ is absolutely continuous, $w'' \in L_2(a, b)$ and $w(x)$ has a double zero at $x = a$.

Note that for such coefficients as $r \equiv 1$ or $\int^\infty (1/r) = \infty$ the Wirtinger-type inequality requires a double zero only at $x = a$ while the result stated at the beginning requires double zeros at both $x = a$ and $x = b$.

An application of Theorem 3.1 to the reciprocal equation (1*) yields

COROLLARY 3.1.1. *If $\int^\infty p = \infty$ and equation (1*) is disconjugate then for every $b > a$*

$$\int_a^b \frac{1}{r} w^2 < \int_a^b \frac{1}{p} (w'')^2 \tag{21'}$$

for the above class of functions w.

4. Higher order equations. The following is the theorem of Coles [5] which has been utilized several times in the preceding sections. It should be noted that his proof for the case $r \equiv 1$ carries over step-for-step for the following.

THEOREM C. *If m is a positive integer; $r(x) > 0$ and $p(x)$ are both continuous on $[a, b]$, $a < b$; and $y(x)$ is a solution of*

$$(r(x)y^{(m)})^{(m)} - p(x)y = 0; \;\; y_1(x) = r(x)y^{(m)}(x) \tag{22}$$

such that

$$\begin{cases} (-1)^m p(x)y(x) \geqq 0 \text{ but } \not\equiv 0 \text{ on } [a, b] \\ p_i y^{(m-i)}(c_i) \geqq 0 \quad (i = 1, 2, \cdots, m) \\ q_i y_1^{(i)}(d_i) \geqq 0 \quad (i = 0, 1, \cdots, m-1) \end{cases} \tag{23}$$

where

$k_i = 0$ or 1 $\quad (i = 0, 1, \cdots, m)$ such that $\sum_{i=0}^m k_i$ is even

$$c_i = \begin{cases} a, \; k_i = 0 \\ b, \; k_i = 0 \end{cases} \qquad d_i = \begin{cases} a, \; k_{i+1} = 0 \\ b, \; k_{i+1} = 0 \end{cases} \qquad \begin{cases} c_i^* = a + b - c_i \\ d_i^* = a + b - d_i \end{cases}$$

$$p_i = (-1)^{\Sigma_{j=0}^i k_j}, \; q_i = (-1)^i p_i \quad (i = 0, \cdots, m-1)$$

then

$$p_i y^{(m-i)}(x) > 0 \text{ on } (a, b) \text{ and at } c_i^* \qquad (i = 1, 2, \cdots, m) \tag{24}$$
$$q_i y_1^{(i)}(x) \geqq 0 \text{ on } [a, b] \text{ and } > 0 \text{ at } d_i^* \qquad (i = 0, 1, \cdots, m-1),$$

and the last inequality is strictly positive if $p(x)$ *is not identically zero on any subinterval of* $[a, b]$. *Furthermore, for every* $b > a$

$$I_{2m}[u; b] = \int_a^b [r(u^{(m)})^2 - (-1)^m p u^2] \tag{25}$$

is non-negative for all functions $u(x)$ *such that*

$$\begin{cases} u^{(m-1)}(x) \text{ is absolutely continuous} \\ u^{(m)} \in L_2[a, b] \text{ and } u^{(i)}(d^*_{m-i-1}) = 0 \end{cases} \tag{26}$$
(of at least order 1)

for $i = 0, 1, \cdots, m-1$; *with* $I_{2m} = 0$ *if, and only if,* $u(x)$ *is a constant times a solution of* (22) *which, in addition, satisfies*

$$q_i y^{(i)}(d_i) = 0 \qquad (i = 0, 1, \cdots, m-1). \tag{27}$$

Note that the special case in §3 *is that for* $m = 2$ *and* $k_1 = k_2 = k_3 = 0$.

For this case, Coles' method reduces to the following: If $y(x)$ is a solution of (1) such that $y > 0$ and $y' > 0$ on $[a, b]$ then by integrating by parts and completing squares

$$\int_a^b pu^2 = \left[\frac{y_1' u^2}{y} - \frac{y_1 (u')^2}{y'}\right]_a^b$$
$$+ \int_a^b \frac{y_1'}{y'}\left[u' - \frac{y'}{y}u\right]^2 - \int_a^b r\left[u'' - \frac{y''}{y'}u'\right]^2 + \int_a^b r(u'')^2 .$$

Using Theorem 2.2 the inequality (21) follows immediately.

BIBLIOGRAPHY

1. J. H. Barrett, *Behavior of solutions of second-order self-adjoint differential equations*, Proc. AMS **8** (1955), 510-518.

2. ———, *Disconjugacy of second-order linear differential equations with non-negative coefficients*, Proc. AMS **10** (1959), 552-561.

3. P. R. Beesack, *Integral inequalities of the Wirtinger-type*, Duke Math. J. **25** (1958), 477-498.

4. W. J. Coles, *Linear and Riccati Systems*, Duke Math. J. **22** (1955), 333-338.

5. ———, *A general Wirtinger-type inequality*, Duke. Math. J. **27** (1960), 133-138.

6. E. Hille, *Non-oscillation Theorems*, Trans. AMS **64** (1948), 234-252.

7. W. Leighton, *On self-adjoint differential equations of second-order*, J. London Math. Soc. **27** (1952), 37-47.

8. ———, *Bounds for the solutions of a second order linear differential equation*, Proc. Nat. Acad. Sci. **35** (1949), 190-191.

9. ———, *Principal quadratic functionals*, Trans. AMS **67** (1949), 253-274.

10. W. Leighton and Z. Nehari, *On the oscillation of solutions of self-adjoint linear differential equations of the fourth order* Trans. AMS. **89** (1958), 325-377.
11. M. Morse, *Calculus of Variations in the Large*, New York, 1934.
12. Z. Nehari, *Oscillation criteria for second-order linear differential equations*, Trans. AMS **85** (1958), 428-445.
13. W. T. Reid, *A comparison theorem for self-adjoint differential equations of second order*, Annals of Math. **65** (1957), 197-202.
14. H. M. and R. L. Sternberg, *A two-point boundary problem for ordinary self-adjoint differential equations of the fourth order*, Canadian J. Math. **6** (1954), 416-419.

THE UNIVERSITY OF UTAH

HARDY'S INEQUALITY AND ITS EXTENSIONS

PAUL R. BEESACK

1. Introduction. In this paper we are concerned with a systematic and uniform treatment of some analogues and extensions of Hardy's inequality for integrals. This result we state as

THEOREM 1. *If* $p > 1$, $f(x) \geqq 0$, *and* $F(x) = \int_0^x f(t)dt$, *then*

$$\int_0^\infty \left(\frac{F}{x}\right)^p dx < \left(\frac{p}{p-1}\right)^p \int_0^\infty f^p dx$$

unless $f \equiv 0$. *The constant is the best possible.*

This theorem was first proved by Hardy [1], and various alternative proofs have been given by other authors. (For reference to these, see [3, 240—243].) Theorem 1, together with the following generalization of this result (also due to Hardy, [2] and [3, Th. 330]) may be regarded as models of the class of inequalities with which this paper deals.

THEOREM 2. *If* $p > 1$, $r \neq 1$, $f(x) \geqq 0$, *and* $F(x)$ *is defined by*

$$F(x) = \begin{cases} \int_0^x f(t)dt & (r > 1)\,, \\ \int_x^\infty f(t)dt & (r < 1)\,, \end{cases}$$

then

$$\int_0^\infty x^{-r} F^p dx < \left(\frac{p}{|r-1|}\right)^p \int_0^\infty x^{-r}(xf)^p dx$$

unless $f \equiv 0$. *Again the constant is the best possible.*

Our integral inequalities will be of the form

$$\int_a^b s(x) F^p dx \leqq \int_a^b r(x) f^p dx \tag{1.1}$$

where $p > 1$ (or $p < 0$), and F is defined (as in Theorem 2) as a suitable integral of $f(x)$. For $0 < p < 1$, we obtain inequalities of the form (1.1), but with the inequality sign reversed. Our method of proof differs from those referred to above. We make use of the Euler-Lagrange differential equations

Received July 14, 1959.

$$\frac{d}{dx}\{r(x)y'^{p-1}\} + s(x)y^{p-1} = 0 \ , \tag{1.2}$$

$$\frac{d}{dx}\{r(x)(-y')^{p-1}\} - s(x)y^{p-1} = 0 \ , \tag{1.3}$$

associated with the minimum problem (1.1). Here, (1.2) applies if $F' = f$, while (1.3) applies if $F' = -f$. Nevertheless, the method is not a variational method, the difficulties involved in such an approach being considerable. (cf. [3, p. 181], where a variational proof of Theorem 1 is sketched.) Rather, we make use of certain Riccati-like equations associated with (1.2), (1.3) leading to integral identities. Aside from this, the main tools used are Hölder's inequality and two special, simple cases of the theorem of the arithmetic-geometric means.

In § 2 we begin by disposing of several lemmas on the "order of a zero" of a function. There will be needed in § 3, where we deal with the inequalities (1.1); this arrangement avoids interrupting the main thread of the argument. Finally, in § 4, we consider the case that p is a positive, even integer, so that (1.2), (1.3) are the same, and we may allow f (and F) to change sign.

2. Preliminary lemmas. Throughout this paper our integrals may be interpreted either in the Lebesgue sense, or as (absolutely convergent) improper Riemann integrals, with statements such as $f(x) \equiv g(x)$ to be interpreted accordingly. We always use the letters p and q to denote conjugate exponents, i.e., $p^{-1} + q^{-1} = 1$.

LEMMA 2.1. *Let $r(x)$ be positive and continuous on $a < x < b$, and suppose that $\int_a^b r\,|f|\,pdx < \infty$, where $p > 1$. Set*

$$F_1(x) = \int_a^x f(t)dt \ , \qquad (a \leqq x < b) \ ; \qquad F_2(x) = \int_x^b f(t)dt \ , \qquad (a < x \leqq b) \ .$$

If $r(x) = O[(x-a)^{p-1}]$, or if $r^{q/p}(x)\int_a^x r^{-q/p}(t)dt = O(x-a)$, then

$$r(x)\,|\,F_1(x)\,|^p = o[(x-a)^{p-1}] \ \text{as} \ x \to a+ \ . \tag{2.1}$$

If $r(x) = O[(b-x)^{p-1}]$, or if $r^{q/p}(x)\int_x^b r^{-q/p}(t)dt = O(b-x)$, then

$$r(x)\,|\,F_2(x)\,|^p = o[(b-x)^{p-1}] \ \text{as} \ x \to b- \ . \tag{2.2}$$

Either a or b may be infinite, the order conditions being modified appropriately.

Proof. We prove only (2.1), the proof of (2.2) being the same.

First, note that $F_1(x) = o(1)$, so that if $r(x) = O[(x-a)^{p-1}]$, then (2.1) follows. If the alternative hypothesis holds, then

$$|F_1(x)| \leqq \int_a^x |f|\,dt = \int_a^x (r^{1/p}|f|)(r^{-1/p})dt$$
$$\leqq \left(\int_a^x r\,|f|^p\,dt\right)^{1/p} \cdot \left(\int_a^x r^{-q/p}dt\right)^{1/q}$$

by Hölder's inequality. Hence

$$r(x)\,|F_1(x)|^p \leqq o(1)\left\{r^{q/p}(x)\int_a^x r^{-q/p}dt\right\}^{p/q} = o(1)\cdot O[(x-a)^{p/q}]$$
$$= o[(x-a)^{p-1}]\ ,$$

as asserted.

We remark that (2.1) is well-known in the case $r(x) \equiv 1$, where the assertion is simply that $f \in L_p(a, x)$ implies $|F_1(x)|^p = o[(x-a)^{p-1}]$. (cf. [3, Th. 222].) That $r(x)$ must satisfy some restriction in order to assure (2.1) for all F_1 such that $\int r\,|f|^p dx$ converges is easily seen by taking $a = 0$, $f \equiv 1$, $F \equiv x$. Finally, we note that the two hypotheses assuring (2.1) are mutually exclusive. For, $r(x) \leqq k(x-a)^{p-1}$ implies $r^{-q/p}(x) \geqq k(x-a)^{-1}$, so that $\int_a^x r^{-q/p}dt$ does not exist.

COROLLARY 2.1. *The hypotheses for* (2.1) *are satisfied if either*

(2.1.1) $\qquad 0 < c_1 \leqq r(x) \leqq c_2$ *on* $(a, x]$, *or*

(2.1.2) $\qquad r(x)$ *is nonincreasing on* $(a, x]$, *and* a *is finite.*

For (2.2), *the same result holds, with "nonincreasing" replaced by "nondecreasing,"* (*and* b *is finite*).

If $a = -\infty$, then (2.1.1) implies that $r(x)$ is bounded on $(-\infty, x]$, and hence that $r(x) = o(|x|^{p-1})$ as $x \to -\infty$. If a is finite, then

$$r^{p/q}(x)\int_a^x r^{-q/p}dt \leqq c_2^{q/p}c_1^{-q/p}\int_a^x dt = O(x-a)\ ,$$

or

$$r^{q/p}(x)\int_p^x r^{-q/p}dt \leqq \int_a^x dt = O(x-a)\ ,$$

according as (2.1.1) or (2.1.2) holds.

The next lemma, although not strictly required in the sequel, may shed some light on the question as to whether the inequality (1.1) can be "improved," in a given case. The notation is that of Lemma 2.1, and we again assume that $\int_a^b r\,|f|^p dx < \infty$, and $p > 1$.

LEMMA 2.2. *If either* $r^{q/p}(x)\int_a^x r^{-q/p}dt = o(b-x)$, *or if both* $r(x) = o[(b-x)^{p-1}]$ *and* $r^{q/p}(x)\int_k^x r^{-q/p}dt = O(b-x)$, *where* $k \geqq a$, *then*

$$r(x)\,|\,F_1(x)\,|^p = o[(b-x)^{p-1}] \text{ as } x \to b- \,. \tag{2.3}$$

Similarly, if $r^{q/p}(x)\int_x^b r^{-q/p}dt = o(x-a)$, *or if both* $r(x) = o[(x-a)^{p-1}]$ *and* $r^{q/p}(x)\int_x^k r^{-q/p}dt = O(x-a)$, *where* $k \leqq b$, *then*

$$r(x)\,|\,F_2(x)\,|^p = o[(x-a)^{p-1}] \text{ as } x \to a+ \,. \tag{2.4}$$

Proof. Again we shall prove only the first half of this lemma. If the first hypothesis is valid, then from the proof of (2.1) we have

$$r(x)\,|\,\mathrm{F}_1(x)\,|^p \leqq \int_a^b r\,|\,f\,|^p dx \cdot \left(r^{q/p}(x)\int_a^x r^{-q/p}dt\right)^{p/q} = o[(b-x)^{p-1}]\,.$$

Now, suppose the alternative hypotheses are valid. Then

$$r^{q/p}(x)\int_k^x r^{-q/p}dt \leqq K(b-x)$$

for x near b. Given $\varepsilon > 0$ there corresponds $X \geqq k$ such that

$$\left(\int_X^b r\,|\,f\,|^p dt\right)^{1/p} \cdot K^{1/q} < \varepsilon\,.$$

Proceeding now as in Lemma 2.1, we have

$$|\,F_1(x)\,| - |\,F_1(X)\,| \leqq \int_X^x |\,f\,|\,dt \leqq \left(\int_X^x r\,|\,f\,|^p dt\right)^{1/p} \cdot \left(\int_X^x r^{-q/p}dt\right)^{1/q},$$

$$r^{1/p}(x)\,|\,F_1(x)\,| \leqq r^{1/p}(x)\,|\,F_1(X)\,| + \left(\int_X^b r\,|\,f\,|^q dt\right)^{1/p} \cdot \left(r^{q/p}(x)\int_X^x r^{-q/p}dt\right)^{1/q}.$$

Hence

$$r^{1/p}(x)\,|\,F_1(x)\,| \leqq r^{1/p}(x)\,|\,F_1(X)\,| + \varepsilon(b-x)^{1/q}\,,$$

so

$$\frac{r^{1/p}(x)\,|\,F_1(x)\,|}{(b-x)^{(p-1)/p}} \leqq \frac{r^{1/p}(x)}{(b-x)^{1/q}} \cdot |\,F_1(X)\,| + \varepsilon\,.$$

Letting $x \to b-$, we obtain

$$\overline{\lim_{x\to b-}} \frac{r^{1/p}(x)\,|\,F_1(x)\,|}{(b-x)^{(p-1)/p}} \leqq \varepsilon\,.$$

Since ε is arbitrary, (2.3) is established.

We note, without proof, that if $b = \infty$ then (2.3) is valid if either $0 < c_1 \leqq r(x) \leqq c_2$ or if $r(x)$ is nonincreasing on $[x, \infty)$.

LEMMA 2.3. *Let $r(x)$ be positive and continuous on $a < x < b$, and suppose that $\int_a^b rf^p dx < \infty$, where $f > 0$ and $p < 0$. Set*

$$F_1(x) = \int_a^x f(t)dt\,, \qquad (a \leqq x < b)\,; \qquad F_2(x) = \int_x^b f(t)dt, \qquad (a < x \leqq b)\,.$$

If

$$\left\{r^{q/p}(x)\int_a^x r^{-q/p}dt\right\}^{-1} = O[(x-a)^{-1}]\,,$$

then

$$r(x)F_1^p(x) = o[(x-a)^{p-1}]$$

as $x \to a+$. Similarly, if

$$\left\{r^{q/p}(x)\int_x^b r^{-q/q}dt\right\}^{-1} = O[(b-x)^{-1}]\,,$$

then

$$r(x)F_2^p(x) = o[(b-x)^{p-1}]$$

as $x \to b-$. If a or b is infinite, the result is still valid, with $(x-a)$ or $(b-x)$ replaced by $|x|$.

Proof. This time we shall give the proof of the second half of the lemma. Proceeding as in Lemma 2.1, and noting that Hölder's inequality is reversed for $p < 0$, we have

$$F_2(x) = \int_x^b (r^{1/p}f)(r^{-1/p})dt \geqq \left(\int_x^b rf^p dt\right)^{1/p}\left(\int_x^b r^{-q/p}dt\right)^{1/q}\,,$$
$$F_2^p(x) \leqq \left(\int_x^b rf^p dt\right)\left(\int_x^b r^{-q/p}dt\right)^{p/q}\,.$$

Hence

$$r(x)F_2^p(x) \leqq o(1)\cdot\left\{r^{q/p}(x)\int_x^b r^{-q/p}dt\right\}^{p/q} = o\,[(b-x)^{p-1}]\,,$$

since $p/q = p - 1 < 0$.

We point out that the existence of the integrals $\int_x^b rf^p dt$, $\int_x^b f dt$ assures the existence of $\int_x^b r^{-q/p}dt$ in this case $(p < 0)$. Finally, we note that the appropriate hypothesis of the lemma is satisfied: if $a \neq -\infty$, and either $0 < c_1 \leqq r(x) \leqq c_2$ or $r(x)$ is nonincreasing on $(a, x]$; or if $b < \infty$, and either $0 < c_1 \leqq r(x) \leqq c_2$ or $r(x)$ is nondecreasing on $[x, b)$.

LEMMA 2.4. *With the same notation as in Lemma 2.3 ($p < 0$, $\int_a^b rf^p dx < \infty$). If either*

$$\left\{r^{q/p}(x)\int_a^x r^{p/q-}dt\right\}^{-1} = o[(b-x)^{-1}] ,$$

or

$$r(x) = o[(b-x)^{p-1}] ,$$

then

$$r(x)F_1^p(x) = o[(b-x)^{p-1}] \text{ as } x \to b- .$$

If either

$$\left\{r^{q/p}(x)\int_x^b r^{-q/p}dt\right\}^{-1} = o[(x-a)^{-1}] ,$$

or

$$r(x) = o[(x-a)^{p-1}] ,$$

then

$$r(x)F_2^p(x) = o[(x-a)^{p-1}] \text{ as } x \to a+ .$$

Again, we shall prove only the second assertion. Since F_2 increases as $x \to a+$, F_2^p decreases; hence if $r(x) = o[(x-a)^{p-1}]$, the conclusion follows.

If the alternative hypothesis holds, then from the proof of Lemma 2.3, we have

$$r(x)F_2^p(x) \leqq \int_a^b rf^p dx \cdot \left\{r^{q/p}(x)\int_x^b r^{-q/p}dt\right\}^{p/q} = o[(x-a)^{p-1}] .$$

Finally, we note that if $a \neq -\infty$, the second assertion of the lemma will be valid if $r(x)$ is bounded near $x = a$.

LEMMA 2.5. *Let $r(x)$, $s(x)$ be positive and continuous for $a < x < b$. Suppose $F_1(x)$ is nonnegative and nondecreasing on $a < x < b$, and that*

$$\int_a^b s(x)F_1^p(x)dx < \infty , \qquad 0 < p < 1 .$$

If $r(x) = O\left\{(x-a)^{p-1}\cdot\int_x^{(3x-a)/2} s(t)dt\right\}$, then $r(x)F_1^p(x) = o[(x-a)^{p-1}]$ as $x \to a+$.

If $r(x) = O\left\{(b-x)^{p-1}\int_x^{(b+x)/2} s(t)dt\right\}$, then $r(x)F_1^p(x) = o[(b-x)^{p-1}]$ as $x \to b-$.

If $a = -\infty$, or $b = \infty$, the assertions should be

$$r(x) = O\left\{|x|^{p-1}\int_x^{x/2} s(t)dt\right\} \text{ implies } r(x)F_1^p(x) = o(|x|^{p-1}) \text{ as } x \to -\infty ,$$

or

$$r(x) = O\left\{x^{p-1}\int_x^{2x} s(t)dt\right\} \textit{ implies } r(x)F_1^p(x) = o(|x|^{p-1}) \textit{ as } x \to \infty .$$

Proof. Since F_1^p is nondecreasing we have

$$\int_x^{(3x-a)/2} s(t)F_1^p(t)dt \geqq F_1^p(x)\int_x^{(3x-a)/2} s(t)dt \geqq Kr(x)F_1^p(x)(x-a)^{1-p} .$$

The result now follows from the fact that the left term of this inequality converges to zero as $x \to a+$. The second assertion of the lemma follows in the same way. Finally we note that if $a \neq -\infty$, and $r(x)$ is bounded near $x = a$, then $r(x)F_1^p(x) = o[(x-a)^{p-1}]$ is immediately valid.

LEMMA 2.6. *With the same hypotheses as in Lemma* 2.5, *except* $F_2(x)$ *is supposed nonnegative and nonincreasing on* $a < x < b$:

$$\textit{If } r(x) = O\left\{(x-a)^{p-1}\int_{(x+a)/2}^{x} s(t)dt\right\}, \textit{ then } r(x)F_2^p(x) = o[(x-a)^{p-1}] \textit{ as } x \to a+ .$$

$$\textit{If } r(x) = O\left\{(b-x)^{p-1}\int_{(3x-b)/2}^{x} s(t)dt\right\}, \textit{ then } r(x)F_2^p(x) = o[(b-x)^{p-1}] \textit{ as } x \to b- .$$

This is proved in precisely the same way as Lemma 2.5.

3. Integral inequalities with p real. Let p be a real parameter ($p \neq 0$, $p \neq 1$). Consider the pair of second-order, nonlinear differential equations

$$\frac{d}{dx}\{r(x)y'^{p-1}\} + s(x)y^{p-1} = 0 , \tag{3.1}$$

$$\frac{d}{dx}\{r(x)(-y')^{p-1}\} - s(x)y^{p-1} = 0 , \tag{3.2}$$

where $s(x)$, $r(x)$, $r'(x)$ *are assumed continuous on an interval* $a < x < b$, *and* $r(x) > 0$ *on this interval.* Here either a or b, or both, may be infinite. We note that these two equations are identical if p is an even integer. In particular, when $p = 2$, these equations reduce to the self-adjoint linear equation

$$(r(x)y')' + s(x)y = 0 .$$

Let $y(x)$ be a solution of (3.1) for which $y(x) > 0$, $y'(x) > 0$ on (a, b) and set $h(x) = [y'(x)/y(x)]^{p-1}$. Then $h(x)$ satisfies the Riccati-like equation

$$\frac{d}{dx}(rh) + (p-1)rh^q = -s(x) , \qquad q = p/(p-1) . \tag{3.1*}$$

Similarly, if $y(x)$ is a solution of (3.2) such that $y(x) > 0, y'(x) < 0$ on (a, b), and we set $h(x) = [-y'(x)/y(x)]^{p-1}$, then h satisfies

$$\frac{d}{dx}(rh) - (p-1)rh^q = s(x) \tag{3.2)*}$$

Now, suppose $f_1(x), f_2(x)$ are nonnegative, measurable functions on (a, b). With the pair of differential equations (3.1), (3.2), we shall associate the functions

$$F_1(x) = \int_a^x f_1(t)dt\,, \qquad a \leqq x < b\,, \tag{3.3}$$

$$F_2(x) = \int_x^b f_2(t)dt\,, \qquad a < x \leqq b\,, \tag{3.4}$$

Notice, in particular, that our notation implies $F_1(a) = F_2(b) = 0$, and that f_1, f_2 are integrable on any closed subinterval of (a, b). Since r, h and F_i are all continuous and $h \neq 0$ on such a closed subinterval, it follows that if $a < a' < b' < b$, then the integrals

$$I_i(a', b') = \int_{a'}^{b'} r\{f_i^p + (p-1)h^q F_i^p - phf_i F_i^{p-1}\}dx \tag{3.5}$$

exist, provided $f_i \in L_p(a', b')$. In the case $0 < p < 1$ this latter condition follows from the fact that $f_i \in L(a', b')$. If $p < 0$, we must also insist that f_i be strictly positive. Taking $i = 1$ and integrating by parts the last term of (3.5), we obtain

$$I_1(a', b') = \int_{a'}^{b'} rf_1^p dx + (p-1)\int_{a'}^{b'} rh^q\ F_1^p dx + \int_{a'}^{b'} (rh)' F_1^p dx - rhF_1^p \Big|_{b'}^{a'}\,,$$

or using (3.1)*,

$$I_1(a', b') = \int_{a'}^{b'} rf_1^p dx - \int_{a'}^{b'} sF_1^p dx + r(a')h(a')F_1^p(a') - r(b')h(b')F_1^p(b')\,. \tag{3.6}$$

Proceeding in the same way for I_2, and using (3.2)*, we obtain

$$I_2(a', b') = \int_{a'}^{b'} rf_2^p dx - \int_{a'}^{b'} sF_2^p dx + r(b')h(b')F_2^p(b') - r(a')h(a')F_2^p(a')\,. \tag{3.7}$$

We now use the fact that $I_i(a', b')$ is nonnegative if $p > 1$ or $p < 0$, and nonpositive if $0 < p < 1$. Indeed, this follows from the well-known inequalities [3, Th. 41]

$$x^p + (p-1)y^p - pxy^{p-1} \geqq 0\,, \qquad (p < 0 \text{ or } p > 1)\,, \tag{3.8}$$

$$x^p + (p-1)y^p - pxy^{p-1} \leqq 0\,, \qquad (0 < p < 1)\,. \tag{3.9}$$

Here, x and y are nonnegative (positive if $p < 0$), and in both cases strict inequality holds unless $y = x$. Setting $x = f_i$, $y = h^p F_i$ in (3.8), and recalling that $r(x) > 0$ on (a', b') ,we see from (3.5) that $I_i(a', b') \geqq 0$, with strict inequality unless $f_i/F_i \equiv \pm y'/y$, i.e., unless $f_i \equiv cy'$. Similarly, in the case $0 < p < 1$, we may apply (3.9) to prove $I_i(a', b') \leqq 0$. Hence from (3.6) and (3.7) we obtain

$$\int_{a'}^{b'} sF_1^p dx \lesseqgtr \int_{a'}^{b'} rf_1^p dx + r(x)h(x)F_1^p(x)\Big|_{b'}^{a'}, \tag{3.10}$$

$$\int_{x'}^{b'} sF_2^p dx \lesseqgtr \int_{a'}^{b'} rf_2^p dx + r(x)h(x)F_2^p(x)\Big|_{a'}^{b'}, \tag{3.11}$$

where, in both cases, the upper inequality sign holds if $p < 0$ or $p > 1$, and the lower sign holds if $0 < p < 1$. If $p < 0$ or $p > 1$ our hypothesis will be $\int_a^b rf_i^p dx < \infty$. This will incidentally assure $f_i \in L_p(a', b')$. Finally, we note that if inequality holds in (3.10) or (3.11) for any (a', b'), then (assuming the existence of the corresponding limits) inequality will also hold when $a' = a$, $b' = b$. This follows from the fact that $|I_i(a', b')|$ does not decrease as the interval (a', b') expands.

We must now consider separately the three cases $p > 1$, $p < 0$, $0 < p < 1$, as the details differ in the three cases.

3.1. *The case $p > 1$.* Here we have two theorems of which we prove only the first, the remaining theorem following by the same arguments.

THEOREM 3.1.1. *Suppose the differential equation* (3.1) (*where $p > 1$*) *has a solution $y(x)$ such that $y(x) > 0$, $y'(x) > 0$ on $a < x < b$, and that*

$$y'(x)/y(x) = O[(x - a)^{-1}] \text{ as } x \to a+ . \tag{3.1.1}$$

If $r(x) = O[(x - a)^{p-1}]$, or $r^{q/p}(x)\int_a^x r^{-q/p}dt = O(x - a)$, and $\int_a^b rf_1^p dx < \infty$, then

$$\int_a^b s(x)F_1^p dx \leqq \int_a^b r(x)f_1^p(x)dx . \tag{3.1.2}$$

Proof. Setting $h(x) = [y'(x)/y(x)]^{p-1}$, we have $h = O[(x - a)^{1-p}]$ by (3.1.1). Moreover, by Lemma 2.1, we have $rF_1^p = o[(x - a)^{p-1}]$. Hence, letting $a' \to a$ and $b' \to b$ in (3.10), we obtain

$$\int_a^b sF_1^p dx \leqq \int_a^b rf_1^p dx - \overline{\lim_{x \to b}}\, r(x)h(x)F_1^p(x) , \tag{3.1.3}$$

which certainly implies (3.1.2). By our previous remarks, equality can hold in (3.1.3) only if $F_1(x) = cy(x)$, so that equality can hold in (3.1.2) only if both this condition holds (so $y(a)$ must be zero), and

$$\lim_{x\to b} c^p r(x)y(x)y'^{p-1}(x) = 0 \ .$$

On the other hand, even if this holds for $c \neq 0$, equality may hold in (3.1.2) only for $f_1 \equiv 0$, since $\int ry'^p dx$ may diverge. Indeed, by Lemma 2.1, this integral will diverge unless

$$\lim_{x\to a} r(x)y(x)y'^{p-1}(x) = 0 \ . \tag{3.1.4}$$

In summary then, *equality holds in* (3.1.2) *only if* $f_1 \equiv cy'(x)$ *where* $c = 0$ *unless all of the conditions*

$$y(a) = 0 \ , \quad \lim_{x\to a} ryy'^{p-1} = 0 \ , \quad \int_a^b ry'^p dx < \infty \ , \tag{3.1.5}$$

hold. In particular, c must be zero unless (3.1.4) is satisfied.

The inequality (3.1.2) is certainly sharp (i.e, the unit constant on the right side cannot be reduced) if the conditions (3.1.5) all hold. Suppose that $\int_a^b ry'^p dx < \infty$, and at least one of the remaining conditions of (3.1.5) does not hold. Then, in general, (3.1.2) is not sharp. This is easily seen by taking $p = 2$, $r(x) = s(x) = 1$, $y(x) = \sin x$, with $0 < a < b \leqq \pi/2$; in this case the unit constant can be reduced to $4(b-a)^2/\pi^2$.

Finally, *if* $\int_a^b ry'^p dx = \infty$, *and* $s(x) \geqq 0$, *then* (3.1.2) *is sharp if*

$$\varliminf_{x\to a} r(x)y(x)y'^{p-1}(x) < \infty \ \textit{and} \ \varliminf_{x\to b} r(x)y(x)y'^{p-1}(x) < \infty \ . \tag{3.1.6}$$

In fact, if $\int_a^x ry'^p dx = \infty$, the first of conditions (3.1.6) is sufficient, and if $\int_x^b ry'^p dx = \infty$, the second of conditions (3.1.6) is sufficient to ensure the sharpness of (3.1.2). To prove this assertion, we take

$$f_1(x) = \begin{cases} 0 \ , & a \leqq x \leqq a' \ , \\ y'(x) \ , & a' < x < b' \ , \\ 0 \ , & b' \leqq x \leqq b \ , \end{cases}$$

where a', b' will be fixed later. Then $F_1(x) = y(x) - y(a')$ for $a' < x < b'$, and

$$F_1^p(x) = y^p(x)\left\{1 - \frac{y(a')}{y(x)}\right\}^p \geqq y^p(x)\left\{1 - p\frac{y(a')}{y(x)}\right\} , \ a' < x < b' \ . \tag{3.1.7}$$

This inequality is the special case of (3.8) obtained by taking $x = 1 - y(a')y^{-1}(x)$, $y = 1$. Using (3.1.7) and (3.6) we now have

$$\begin{aligned} \int_{a'}^{b'} sF_1^p dx &\geqq \int_{a'}^{b'} sy^p dx - py(a')\int_{a'}^{b'} sy^{p-1}dx \\ &= \int_{a'}^{b'} ry'^p dx + rhy^p \Big|_{b'}^{a'} - py(a')\int_{a'}^{b'} sy^{p-1}dx \ . \end{aligned}$$

From (3.1), we have

$$\int_{a'}^{b'} sy^{p-1}dx = ry'^{p-1}\Big|_{b'}^{a'} .$$

Hence

$$\begin{aligned}\int_a^b sF_1^p dx &> \int_{a'}^{b'} rf_1^p dx - r(b')y(b')y'^{p-1}(b') - py(a')r(a')y'^{p-1}(a') \\ &> (1-\delta)\int_{a'}^{b'} rf_1^p dx = (1-\delta)\int_a^b rf_1^p dx ,\end{aligned}$$

provided that

$$r(b')y(b')y'^{p-1}(b') + pr(a')y(a')y'^{p-1}(a') < \delta\int_{a'}^{b'} rf_1^p dx .$$

By (3.1.6) this inequality can be satisfied for any $\delta > 0$ by selecting a' or b', or both, appropriately close to a or b. Hence (3.1.2) is sharp.

It is of interest to note that the sharpness of (3.1.2) implies only that

$$\inf_{f_1}\left\{\overline{\lim_{x\to b}}\, r(x)h(x)F_1^p(x)\right\} = 0 ,$$

where the infimum is taken over all admissible $f_1 \not\equiv 0$. Hence, in general, (3.1.3) certainly states more than (3.1.2) even when (3.1.2) is sharp. On the other hand, if $r(x)$ satisfies the order conditions of Lemma 2.2 at $x = b$, and if $y'(x)y^{-1}(x) = O[(b-x)^{-1}]$ as $x \to b-$, then (3.1.2) and (3.1.3) are the same.

Finally, *we note that if* $p \geqq 2$, *then* (3.1.3) *can be improved to*

$$(3.1.8) \qquad \int_a^b sF_1^p dx \leqq \int_a^b rf_1^p dx - \overline{\lim_{x\to b}}\, r(x)h(x)F_1^p(x) - \int_a^b r\,|\,f_1 h^{1/(p-1)}F_1\,|^p dx .$$

This follows from the fact that the inequality (3.8) can be improved slightly to give

$$(3.1.9) \qquad x^p + (p-1)y^p - pxy^{p-1} \geqq |\,x-y\,|^p , \qquad p \geqq 2 .$$

THEOREM 3.1.2. *Suppose the differential equation* (3.2) (*with* $p > 1$) *has a solution* $y(x)$ *such that* $y(x) > 0$, $y'(x) < 0$ *on* $a < x < b$, *and that*

$$(3.1.10) \qquad y'(x)/y(x) = O[(b-x)^{-1}] \qquad \text{as } x \to b- .$$

If $r(x) = O[(b-x)^{p-1}]$, *or* $r^{q/p}(x)\int_x^b r^{-q/p}dt = O(b-x)$, *and* $\int_a^b rf_2^p dt < \infty$, *then*

$$(3.1.11) \qquad \int_a^b s(x)F_2^p(x)dx \leqq \int_a^b r(x)f_2^p(x)dt ,$$

where $f_2 \geqq 0$, $F_2(x) = \int_x^b f_2 dt$. *Equality holds in* (3.1.11) *only if* $f_2 \equiv cy'(x)$ *where* $c = 0$ *unless all of the conditions*

$$(3.1.12) \qquad y(b) = 0 , \qquad \lim_{x \to a} ry(-y')^{p-1} = 0 , \qquad \int_a^b r(-y')^p dx < \infty$$

hold. If $s(x) \geqq 0$, *and* $\int_a^b r(-y')^p dx = \infty$, *then* (3.1.11) *is sharp provided*

$$\varliminf_{x \to a} ry(-y')^{p-1} < \infty, \text{ and } \varliminf_{x \to b} ry(-y')^{p-1} < \infty .$$

Finally, if $p \geqq 2$, (3.1.11) *may be improved to*

$$(3.1.13) \quad \int_a^b sF_2^p dx \leqq \int_a^b rf_2^p dx - \varlimsup_{x \to a} r(x)h(x)F_2^p(x) - \int_a^b r\,|f_2 - h^{1/(p-1)}F_2|^p dx ,$$

where $h = [(-y')/y]^{p-1}$.

Theorem 1 is the special case of Theorem 3.1.1 obtained by setting $a = 0, b = \infty, y(x) = x^{(p-1)/p}$. More generally, Theorem 2 is obtained from Theorem 3.1.1. (for $r > 1$) and Theorem 3.1.2 (for $r < 1$) by taking $y(x) = x^{(r-1)/p}$. In this case, we have

$$r(x) = kx^{p-r} , \qquad r^{q/p}(x) = k_1 x^{(p-r)/(p-1)} , \qquad \left(k = \left(\frac{p}{|r-1|}\right)^p\right)$$

so that $r^{q/p}(x)\int_0^x r^{-q/p} dt = k_2 x$, or $r^{q/p}(x)\int_x^\infty r^{-q/p} dt = k_2 x$ according as $r > 1$ or $r < 1$. Since $\int_0^x r\,|y'|^p dx = \infty$, equality can hold in (3.1.2) only for $f_1 \equiv 0$, ond in (3.1.11) only for $f_2 \equiv 0$. On the other hand,

$$r(x)y(x)\,|y'(x)|^{p-1} \equiv K$$

so that the corresponding inequality is sharp. Finally, we note that $r(x) = o(x^{p-1})$ and $r^{q/p}(x)\int_0^x r^{-q/p} dt = O(x)$ as $x \to \infty$, for the case $r > 1$, and $r(x) = o(x^{p-1})$, $r^{q/p}(x)\int_x^\infty r^{-q/p} dt = O(x)$ as $x \to 0$, for the case $r < 1$. Hence, according to Lemma 2.2, we have $r(x)F_i^p(x) = o(x^{p-1})$ or $F_i^p(x) = o(x^{r-1})$ as $x \to 0$ (for $i = 2$, $r < 1$), or as $x \to \infty$ (for $i = 1$, $r > 1$). Since $y'/y = kx^{-1}$ in both cases, it follows that (3.1.3) reduces to (3.1.2), with a similar remark holding in case $r < 1$.

As another example for Theorem 3.1.1 we have the following inequality (cf. [3, Th. 256]): *If* $p > 1$, $y' \geqq 0$, $y(x) = \int_0^x y' dt$, *then*

$$(3.1.14) \qquad \int_0^{\pi/2} y^p dx \leqq \frac{1}{p-1}\left(\frac{p}{2} \sin \frac{\pi}{p}\right)^p \int_0^{\pi/2} y'^p dx$$

equality holding only if $y = cy(x)$ *where* $y(x)$ *is the unique solution of the equation*

$$x = \frac{p}{2} \sin \frac{\pi}{p} \int_0^y (1 - t^p)^{-1/p} dt , \qquad 0 \leqq y \leqq 1 .$$

We conclude this section with three examples similar to Theorem 2. *We suppose that* $p > 1$, *and* $i = 1$ *or* 2 *according as* $\alpha > 0$ *or* $\alpha < 0$ *in the first two inequalities, while* $i = 1$ *or* 2 *according as* α, β *are both positive, or both negative in the third inequality. Then*

(3.1.15) $$|\alpha|^p \int_0^\infty \frac{x^{-1-\alpha}}{(1 + x^\alpha)^{p-1}} F_i^p dx < \int_0^\infty x^{p(1-\alpha)-1} f_i^p dx \qquad \textit{unless } f_i \equiv 0 ;$$

(3.1.16) $$|\alpha|^p \left(\frac{p-1}{p}\right)^p \int_0^\infty \frac{x^{-1+\alpha}}{(1 + x^\alpha)^p} F_i^p dx < \int_0^\infty x^{(p-1)(1-\alpha)} f_i^p dx \quad \textit{unless } f_i \equiv 0 ;$$

(3.1.17) $$|\alpha|^{p-1}(|\alpha| + |\beta|)(p-1) \int_0^\infty \frac{x^{(p-1)(1-\alpha)-p+\beta}}{(1 + x^\beta)^p} F_i^p dx < \int_0^\infty x^{(p-1)(1-\alpha)} f_i^p dx$$

unless $F_i \equiv cx^\alpha (1 + x^\beta)^{-\alpha/\beta}$. *In all three cases, the constants are the best possible.* The (inadmissible) extremal functions for the inequalities (3.1.15), (3.1.16) are $y = 1 + x^\alpha$, $y = (1 + x^\alpha)^{1-(1/p)}$ respectively.

3.2. *The case* $p < 0$. The theorems corresponding to Theorems 3.1.1 and 3.1.2 are stated as Theorems 3.2.1 and 3.2.2, of which we prove only the second.

THEOREM 3.2.1. *Suppose the differential equation* (3.1) (*with* $p < 0$) *has a solution* $y(x)$ *such that* $y(x) > 0$, $y'(x) > 0$ *on* $a < x < b$, *and that*

(3.2.1) $$y(x)/y'(x) = O(x - a) \qquad \textit{as } x \to a+ ,$$

(3.2.2) $$\left\{ r^{q/p}(x) \int_a^x r^{-q/p} dt \right\}^{-1} = O[(x - a)^{-1}] \qquad \textit{as } x \to a+ .$$

If $f_1 > 0$, $F_1(x) = \int_a^x f_1 dt$, *and if* $\int_a^b r f_1^p dx < \infty$, *then*

(3.2.3) $$\int_a^b s(x) F_1^p dx \leqq \int_a^b r(x) f_1^p dx .$$

Equality holds in (3.2.3) *only if* $f_1 \equiv cy'(x)$, *and all of the conditions*

(3.2.4) $$y(a) = 0 , \qquad \lim_{x \to b} ryy'^{p-1} = 0 , \qquad \int_a^b ry'^p dx < \infty ,$$

hold. Finally, if $s(x) \geqq 0$, *and* $\int_a^b ry'^p dx = \infty$, *then* (3.2.3) *is sharp if*

(3.2.5) $$\varlimsup_{x \to b} ryy'^{p-1} < \infty, \textit{ and } \varlimsup_{x \to a} r(x)[y'(x)]^{p-1} \left(\int_a^x r^{-q/p} dt \right)^{1-(1/p)} < \infty .$$

THEOREM 3.2.2. *Suppose the differential equation* (3.2) (*with* $p < 0$)

has a solution $y(x)$ *such that* $y(x) > 0, y'(x) < 0$ *on* $a < x < b$, *and that*

$$y(x)/y'(x) = O(b - x) \qquad as\ x \to b- \ , \tag{3.2.6}$$

$$\left\{r^{q/p}(x)\int_x^b r^{-q/p}dt\right\}^{-1} = O[(b - x)^{-1}] \qquad as\ x \to b- \ . \tag{3.2.7}$$

If $f_2 > 0$, $F_2(x) = \int_x^b f_2 dt$, *and if* $\int_a^b r f_2^p dt < \infty$, *then*

$$\int_a^b s(x)F_2^p dx \leqq \int_a^b r(x)f_2^p dx \ . \tag{3.2.8}$$

Equality holds in (3.2.8) *only if* $f_2 \equiv cy'(x)$, *and all of the conditions*

$$y(b) = 0 \ , \lim_{x \to a} ry(-y')^{p-1} = 0 \ , \int_a^b r(-y')^p dx < \infty \ , \tag{3.2.9}$$

hold. Finally, if $s(x) \geqq 0$, *and* $\int_a^b r(-y')^p dx = \infty$, *then* (3.2.8) *is sharp if*

$$\varliminf_{x \to a} ry(-y')^{p-1} < \infty, \ and \ \varliminf_{x \to a} r(x)[-y'(x)]^{p-1}\left(\int_x^b r^{-q/p}dt\right)^{1-(1/p)} < \infty \ . \tag{3.2.10}$$

To prove Theorem 3.2.2, we set $h(x) = [-y'(x)/y(x)]^{p-1}$. Since $p < 1$, we have $h = O[(b - x)^{1-p}]$, and by Lemma 2.3 we also have $rF_2^p = o[(b - x)^{p-1}]$. Hence, from (3.11) we obtain

$$\int_a^b sF_2^p dx \leqq \int_a^b rf_2^p dx - \varlimsup_{x \to a} r(x)h(x)F_2^p(x) \ , \tag{3.2.11}$$

where equality can hold only if $F_2(x) \equiv cy(x)$. Comparing this with (3.2.8), we verify that conditions (3.2.9) are necessary and sufficient for equality (for an admissible function).

To prove the assertion concerning the sharpness of (3.2.8), we must modify the procedure used in Theorem 3.1.1 in view of our requirement $f_2 > 0$. Here, we set

$$f_2(x) = \begin{cases} g(x) \ , & a \leqq x \leqq a' \ , \\ -y'(x) \ , & a' < x < b' \ , \\ Mk(x) \ , & b' \leqq x \leqq b \ , \end{cases}$$

where a', b', M are to be assigned, $g(x)$ is any (fixed) admissible function, and $k(x)$ is an admissible function to be chosen later. For $a' < x < b'$, we have

$$F_2(x) = y(x)\left\{1 - \frac{y(b')}{y(x)} + \frac{M}{y(x)}\int_{b'}^b kdt\right\} .$$

Hence, as in Theorem 3.1.1,

$$F_2^p(x) \geqq y^p(x)\left\{1 - p\frac{y(b')}{y(x)} + p\frac{M}{y(x)}\int_{b'}^{b} kdt\right\},$$

and

$$\begin{aligned}\int_{a'}^{b'} sF_2^p dx &\geqq \int_{a'}^{b'} sy^p dx - py(b')\int_{a'}^{b'} sy^{p-1}dx + pM\int_{b'}^{b} kdt\int_{a'}^{b'} sy^{p-1}dx \\ &= \int_{a'}^{b'} r(-y')^p dx + rhy^p\Big|_{a'}^{b'} - p\left\{y(b') - M\int_{b'}^{b} kdt\right\}\cdot\left\{r(-y')^{p-1}\right\}\Big|_{a'}^{b'} \\ &> \int_{a'}^{b'} r(-y')^p dx - r(a')y(a')[-y'(a')]^{p-1} + py(b')r(a')[-y'(a')]^{p-1} \\ &\qquad + pMr(b')[-y'(b')]^{p-1}\int_{b'}^{b} kdt .\end{aligned}$$

Hence

$$\int_a^b sF_2^p dx > (1-\delta)\int_a^b fr_2^p dx ,$$

provided

$$\begin{aligned}&r(a')y(a')[-y'(a')]^{p-1} - py(b')r(a')[-y'(a')]^{p-1} - pMr(b')[-y'(b')]^{p-1} \\ &\qquad + (1-\delta)\left\{\int_a^{a'} rg^p dx + M^p\int_{b'}^{b} rk^p dx\right\} < \delta\int_{a'}^{b'} r(-y')^p dx .\end{aligned}$$

We first choose $M = M(b')$ so as to minimize the left side of this inequality. This is accomplished by choosing

$$M = r^{1/(p-1)}(b')\cdot[-y'(b')]\cdot\left(\int_{b'}^{b} kdx\right)^{1/(p-1)}\left(\int_{b'}^{b} rk^p dx\right)^{1/(1-p)} .$$

With this choice of M, we find (after some reduction) that we want to choose a', b', k so that

$$\begin{aligned}&r(a')y(a')[-y'(a')]^{p-1} - py(b')r(a')[-y'(a')]^{p-1} \\ &\qquad + (1-p)r^q(b')\cdot[-y'(b')]^p\left(\int_{b'}^{b} kdx\right)^q\left(\int_{b'}^{b} rk^p dx\right)^{1/(1-p)} < \delta\int_{a'}^{b'} r(-y')^p dx .\end{aligned}$$

An application of Hölder's inequality shows that the best possible choice for k is $k = cr^{-q/p}$, at last for x near b. Moreover, such a k is admissible since $K(x) = \int_x^b r^{-q/p}dt$ is well-defined, and since

$$\int_x^b rk^p dt = \int_x^b r^{-q/p}dt < \infty .$$

With this choice of k we want to choose a', b' so that

$$\begin{aligned}(3.2.12)\quad &r(a')y(a')[-y'(a')]^{p-1} - py(b')r(a')[-y'(a')]^{p-1} \\ &\qquad + (1-p)r^{p/(p-1)}(b')[-y'(b')]^p\int_{b'}^{b} r^{-q/p}dx < \delta\int_{a'}^{b'} r(-y')^p dx .\end{aligned}$$

Now, if $\int_a^x r(-y')^p dx = \infty$, we fix $b' < b$ arbitrarily. Using the first of conditions (3.2.10), together with the fact that $y(x)$ is a decreasing function, we see that (3.2.12) can be satisfied for a' appropriately close to a. Similarly, if $\int_x^b r(-y')^p dx = \infty$, we fix a' arbitrarily and, using the second of conditions (3.2.10), can choose b' so that (3.2.12) is satisfied. Hence (3.2.8) is sharp in either case.

We note that the second of conditions (3.2.10) implies (but does not seem to be equivalent to) the condition

$$\varliminf_{x\to b} ry(-y')^{p-1} < \infty .$$

By taking $a = 0$, $b = \infty$, $y(x) = x^{(r-1)/p}$ we obtain the following extension of Theorem 2 to the case $p < 0$: *if $f(x) > 0$, and $F(x)$ is defined by*

$$F(x) = \begin{cases} \int_0^x f(t)dt & (r < 1) , \\ \int_x^\infty f(t)dt & (r > 1) , \end{cases}$$

then

$$\int_0^\infty x^{-r}F^p dx < \left(\left|\frac{p}{|r-1|}\right|\right)^p \int_0^\infty x^{-r}(xf)^p dx . \tag{3.2.13}$$

The constant is the best possible. It may be of interest to note that in this case, as for the case $p > 1$, the hypotheses of Lemma 2.4 are satisfied by $r(x)$ so that (3.2.11) and (3.2.8) are identical.

As further examples of Theorem 3.2.2 we have the following:

If $\alpha > 0$, $1 < \beta \leqq 1 - (1/p)$, $f_2 > 0$, $F_2 = \int_x^1 f_2 dt$, then

$$(\alpha\beta)^{p-1}\alpha(1-p)(\beta-1)\int_0^1 \frac{x^{\alpha(1-p)(\beta-1)-1}}{(1-x^\alpha)^p}F_2^p dx \leqq \int_0^1 x^{(1-p)(\alpha\beta-1)}f_2^p dx , \tag{3.2.14}$$

where strict inequality always holds if $\beta = 1 - (1/p)$, and otherwise equality holds only if $F_2 \equiv c(1 - x^\alpha)^\beta$.

If $0 < \alpha < 1 - \gamma/(p-1)$, then

$$\alpha^{p-1}[\gamma + (\alpha-1)(p-1)]\int_0^1 \frac{x^{\gamma+(\alpha-1)(p-1)-1}}{(1-x^\alpha)^{p-1}}F_2^p dx < \int_0^1 x^\gamma f_2^p dx , \tag{3.2.15}$$

unless $F_2 \equiv c(1 - x^\alpha)$.

If $\alpha > 0$, $p < 0$, then

$$\alpha^p\left(\frac{p-1}{p}\right)^p \int_0^1 \frac{x^{\alpha-1}}{(1-x^\alpha)^p}F_2^p dx < \int_0^1 x^{(1-p)(\alpha-1)}f_2^p dx . \tag{3.2.16}$$

In each case, the constant is best possible. The inadmissible extremal function for the last inequality is $y = (1 - x^\alpha)^{1-(1/p)}$. Corresponding inequalities can also be obtained involving $F_1 = \int_0^x f_1 dt$, as well as for the case $p > 1$.

3.3. *The case* $0 < p < 1$. Here again we have two theorems corresponding to the two equations (3.1), (3.2).

THEOREM 3.3.1. *Suppose the differential equation* (3.1) (*with* $0 < p < 1$ *and* $s(x) > 0$) *has a solution* $y(x)$ *such that* $y(x) > 0$, $y'(x) > 0$ *on* $a < x < b$, *and that*

$$y(x)/y'(x) = O(b - x) \qquad \text{as } x \to b- \, , \tag{3.3.1}$$

$$(b - x)^{1-p} r(x) = O\left\{\int_x^{(b+x)/2} s(t)dt\right\}, \qquad \text{as } x \to b- \, . \tag{3.3.2}$$

If $f_1 \geqq 0$, $F_1(x) = \int_a^x f_1(t)dt$, *and* $\int_a^b sF_1^p dx < \infty$, *then*

$$\int_a^b sF_1^p dx \geqq \int_a^b rf_1^p dx \, . \tag{3.3.3}$$

Equality holds in (3.3.3) *only if* $f_1 \equiv cy'(x)$, *where* $c = 0$ *unless*

$$y(a) = 0 \, , \qquad \lim_{x\to a} ryy'^{p-1} = 0 \, , \qquad \int_a^b sy^p dx < \infty \, . \tag{3.3.4}$$

If $\int_a^b sy^p dx = \infty$, *then* (3.3.3) *is sharp if*

$$\varliminf_{x\to a} ryy'^{p-1} < \infty \, , \qquad \varliminf_{x\to b} y^p \int_x^b sdt < \infty \, . \tag{3.3.5}$$

As in Lemma 2.5, if $b = +\infty$, then $b - x$ is replaced by x, and $(b + x)/2$ by $2x$ in the order conditions (3.3.1), (3.3.2).

To prove (3.3.3) we need only apply (3.10) and Lemma 2.5 to obtain

$$\int_a^b sF_1^p dx \geqq \int_a^b rf_1^p dx + \varlimsup_{x\to a} r(x)h(x)F_1^p(x) \, , \tag{3.3.6}$$

where equality can hold only if $F_1 \equiv cy(x)$. This proves (3.3.3) as well as the assertion concerning (3.3.4).

If $f_1 = y'$ is *not* admissible, so $\int_a^b sy^p dx = \infty$, we set

$$f_1(x) = \begin{cases} 0 \, , & a \leqq x \leqq a' \, , \\ y'(x) \, , & a' < x < b' \, , \\ 0 \, , & b' \leqq x \leqq b \, . \end{cases}$$

Proceeding as in Theorem 3.1.1 (but using (3.9) rather than (3.8)), we obtain

$$\int_a^b sF_1^p dx < \{y(b') - y(a')\}^p \int_{b'}^b s dx + \int_a^b r f_1^p dx + (1-p) r(a') y(a') y'^{p-1}(a') + p y(a') r(b') y'^{p-1}(b')$$

$$< (1+\delta)\int_a^b r f_1^p dx ,$$

provided

$$y^p(b')\int_{b'}^b s dx + (1-p) r(a') y(a') y'^{p-1}(a') + p y(b') r(b') y'^{p-1}(b') < \delta \int_{a'}^{b'} r y'^p dx . \tag{3.3.7}$$

Now we note that since

$$\int_{a'}^{b'} r y'^p dx = \int_{a'}^{b'} s y^p dx + r y y'^{p-1}\Big|_{a'}^{b'} > \int_{a'}^{b'} s y^p dx - r(a') y(a') y'^{p-1}(a') ,$$

it follows that $\int_a^{b'} r y'^p dx = \infty$ if $\int_a^{b'} s y^p dx = \infty$ provided the first of conditions (3.3.5) is valid. Moreover, if $\int_{a'}^b s y^p dx = \infty$, then $\int_{a'}^b r y'^p dx = \infty$ in any case. Hence, if $\int_a^{b'} s y^p dx = \infty$, (3.3.7) can be satisfied for any $\delta > 0$ by fixing b' and letting $a' \to a$. On the other hand, if $\int_{a'}^b s y^p dx = \infty$, we fix a', and show that the left side of (3.3.7) remains finite for b' appropriately close to b. This is true of the first term by the second of conditions (3.3.5). For the other term of (3.3.7) involving b', we have

$$r y y'^{p-1} = r\left(\frac{y'}{y}\right)^{p-1} y^p \leqq K_1 (b-x)^{1-p} y^p r$$

$$\leqq K_2 (b-x)^{1-p} y^p (b-x)^{p-1} \int_x^{(b+x)/2} s dt ,$$

according to (3.3.1) and (3.3.2). It follows that

$$\varlimsup_{x \to b} r y y'^{p-1} < \infty ,$$

and that the left side of (3.3.7) remains bounded for b' appropriately close to b.

For completeness we state the theorem corresponding to equation (3.2), but omit the proof.

THEOREM 3.3.2. *Suppose the differential equation* (3.2) (*with* $0<p<1$, *and* $s(x) > 0$) *has a solution* $y(x)$ *such that* $y(x) > 0$, $y'(x) < 0$ *on* $a < x < b$, *and that*

$$(3.3.8) \qquad y(x)/y'(x) = O(x-a) \qquad \textit{as } x \to a+ \ ,$$

$$(3.3.9) \qquad (x-a)^{1-p} r(x) = O\left\{\int_{(x+a)/2}^{x} s(t)dt\right\} \qquad \textit{as } x \to a+ \ .$$

If $f_2 \geqq 0$, $F_2(x) = \int_x^b f_2(t)dt$, *and* $\int_a^b sF_2^p dx < \infty$, *then*

$$(3.3.10) \qquad \int_a^b sF_2^p dx \geqq \int_a^b rf_2^p dx \ ,$$

with equality only if $f_2 \equiv cy'(x)$, *where* $c = 0$ *unless*

$$(3.3.11) \qquad y(b) = 0, \lim_{x\to b} ryy'^{p-1} = 0, \int_a^b sy^p dx < \infty \ .$$

If $\int_a^b sy^p dx = \infty$, (3.3.10) *is sharp if*

$$(3.3.12) \qquad \varliminf_{x\to b} ryy'^{p-1} < \infty \ , \quad \varliminf_{x\to a} y^p \int_a^x sdt < \infty \ .$$

Taking $a = 0$, $b = \infty$, $y(x) = x^{(r-1)/p}$, the preceding theorems give the extension of Theorem 2 to the case $0 < p < 1$. This result is also due to Hardy ([2], and Theorem 347, [3]). By taking $y(x) = 1 + x^{\alpha}$, and $i = 1$ or 2 according as $\alpha > 0$ or $\alpha < 0$, we obtain the following analogue of (3.1.15);

$$(3.3.13) \qquad |\alpha|^p \int_0^\infty x^{-1-\alpha}(1+x^\alpha)^{1-p} F_i^p dx > \int_0^\infty x^{p(1-\alpha)-1} f_i^p dx \qquad \textit{unless } f_i \equiv 0 \ .$$

The corresponding analogues of (3.1.16) and (3.1.17) are not valid for $0 < p < 1$. The inequality (3.3.13) is sharp although only the second of conditions (3.3.5) (or (3.3.12)) is satisfied.

4. Integral inequalities with $p = 2k$. As noted previously, if $p = 2k$ the pair of differential equations (3.1), (3.2) reduce to the single equation

$$(4.1) \qquad \frac{d}{dx}\{r(x)y'^{2k-1}\} + s(x)y^{2k-1} = 0 \, .$$

If $y(x)$ is a solution of (4.1) for which $y(x) > 0$ on (a, b), and we set $h(x) = [y'(x)/y(x)]^{2k-1}$, then $h(x)$ satisfies the equation

$$(4.2) \qquad \frac{d}{dx}(rh) + (2k-1)rh^{2k/(2k-1)} = -s(x) \ .$$

We adopt a different notation from that used in (3.3), (3.4) by replacing f_i by u' and F_i by u, where *we assume throughout this section that*

$$(4.3) \qquad u(x) = \int_a^x u'(t)dt = -\int_x^b u'(t)dt \ , \qquad a \leqq x \leqq b \ ,$$

so that $u(a) = u(b) = 0$. Proceeding as in § 3 (and noting that (3.8) is valid for all real x, y when $p = 2k$) we obtain

$$\int_{a'}^{b'} su^{2k}dx \leqq \int_{a'}^{b'} ru'^{2k}dx + r(x)h(x)u^{2k}(x)\Big|_{b'}^{a'} \tag{4.4}$$

with strict inequality unless $u(x) \equiv cy(x)$. Note that

$$|u(x)| \leqq \int_a^x |u'|\,dt\,, \qquad |u(x)| \leqq \int_x^b |u'|\,dt\,.$$

It follows that Lemma 2.1 remains valid with f replaced by u' and F_i replaced by u.

We now want to weaken the hypotheses on (4.1); *in particular we want to allow y' and h to have a single discontinuity at a point c of (a, b), and to allow r to have a discontinuity or a zero at $x = c$.* Otherwise, we assume $r(x)$, $r'(x)$, $s(x)$ continuous, and $r(x) > 0$ on $a < x < b$, as in § 3. Under these hypotheses, by an *extended solution* of (4.1) we mean a function $y(x)$ positive and continuous on $a < x < b$ such that $y'(x)$ is continuous except perhaps at $x = c$, and such that rh is continuous on (a, b). Now, replacing $I_1(a', b')$ in (3.5) by $I_1(a', c - \varepsilon) + I_1(c + \varepsilon, b')$, carrying out the corresponding work following (3.5), and then letting $\varepsilon \to 0$, we again obtain (4.4), assuming the existence of $\int ru'^{2k}dx$. Finally, since $a < c < b$, Lemma 2.1 also holds.

THEOREM 4.1. *Suppose the differential equation* (4.1) *has an extended solution $y(x) > 0$ on $a < x < b$ and that*

$$\begin{aligned} y'(x)/y(x) &= O[(x-a)^{-1}] \ \text{ as } x \to a+\,, \\ y'(x)/y(x) &= O[(b-x)^{-1}] \ \text{ as } x \to b-\,, \end{aligned} \tag{4.5}$$

and both of the conditions

$$r(x) = O[(x-a)^{2k-1}]\,, \ \text{or}\ r^{q/p}(x)\int_a^x r^{-q/p}dt = O(x-a) \ \text{ as } x \to a+\,, \tag{4.6}$$

$$r(x) = O[(b-x)^{2k-1}]\,, \ \text{or}\ r^{q/p}(x)\int_x^b r^{-q/p}dt = O(b-x) \ \text{ as } x \to b-\,, \tag{4.7}$$

hold. If $u(x)$ satisfies (4.3), *and* $\int_a^b ru'^{2k}dx < \infty$, *then*

$$\int_a^b su^{2k}dx \leqq \int_a^b ru'^{2k}dx\,. \tag{4.8}$$

Equality holds only if $u \equiv cy(x)$, where $c = 0$ unless

$$y(a) = y(b) = 0\,, \qquad \int_a^b ry'^{2k}dx < \infty\,. \tag{4.9}$$

Moreover, if $\int_a^b ry'^{2k}dx = \infty$ *and* $s(x) \geqq 0$, *then* (4.8) *is sharp if* $y(a) = y(b)$ *and*

$$\overline{\lim_{x\to a}}\,|\,ryy'^{2k-1}\,| < \infty \ , \ \textit{and} \ \overline{\lim_{x\to b}}\,|\,ryy'^{2k-1}\,| < \infty \ . \tag{4.10}$$

The inequality (4.8), and the sufficiency of the conditions (4.9) for equality, follows from (4.4)—(4.7) together with Lemma 2.1. To prove the assertion concerning sharpness we assume that $y(a) = y(b)$, and that $\int_a^{b'} ry'^{2k}dx = \infty$, and define

$$u(x) = \begin{cases} 0\,, & a \leqq x \leqq a'\,, \\ y(x) - y(a')\,, & a' < x < b'\,, \\ 0\,, & b' \leqq x \leqq b\,. \end{cases}$$

Here a' and b' are to be chosen later, and in such a way that $y(b') = y(a')$. Thus $u(x)$ satisfies (4.3), and $\int_a^b ru^{2k}dx = \int_{a'}^{b'} ry'^{2k}dx < \infty$, so u is admissible. As in § 3.1 we find

$$\begin{aligned}\int_a^b su^{2k}dx &\geqq \int_a^b ru'^{2k}dx + (1-2k)r(a')y(a')y'^{2k-1}(a') - (1-2k)r(b')y(b')y'^{2k-1}(b') \\ &> (1-\delta)\int_a^b ru'^{2k}dx\end{aligned}$$

provided

$$(2k-1)r(a')y(a')y'^{2k-1}(a') - (2k-1)r(b')y(b')y'^{2k-1}(b') < \delta\int_{a'}^{b'} ry'^{2k}dx\,. \tag{4.11}$$

Since $(ry'^{2k-1})' = -sy^{2k-1} \leqq 0$, we see that ry'^{2k-1} is a nonincreasing function on $a < x < b$. It follows from this fact, together with $y(a) = y(b)$, that $y(x) \geqq y(a)$, $a \leqq x \leqq b$. Since $y(x) \not\equiv y(a)$ for x near a (otherwise $\int_a^{b'} ry'^{2k}dx < \infty$), $y(x)$ assumes a maximum value for $x = \alpha$, where $a < \alpha < b$. But then to each a', $a < a' < \alpha$, there corresponds at least one b', $\alpha < b' < b$, such that $y(b') = y(a')$. Choosing such a value of b' in (4.11), we see by (4.10) that for any $\delta > 0$, (4.11) can be satisfied for a' sufficiently close to a. The same proof holds if $\int_{a'}^b ry'^{2k}dx = \infty$.

Because of the symmetry of the extremal function, the inequality (3.1.14) can clearly be extended according to Theorem 4.1 to give: *If* $u(o) = u(\pi) = 0$, *then*

$$\int_0^\pi u^{2k}dx \leqq \frac{1}{2k-1}\left(k \sin \frac{\pi}{2k}\right)^{2k}\int_0^\pi u'^{2k}dx\,, \tag{4.12}$$

equality holding only if $u = cy(x)$, *where* $y((\pi/2) + x) = y((\pi/2) - x)$, *and for* $0 \leqq x \leqq (\pi/2)$, $y(x)$ *is the unique solution of the equation*

$$x = k\sin\frac{\pi}{2k}\int_0^y (1-t^{2k})^{-1/2k}dt\,, \qquad 0 \leqq y \leqq 1\,.$$

The next two inequalities are the extensions of (3.1.17) corresponding to the choices $\alpha = -(2k-1)^{-1}$, $\beta = -2k(2k-1)^{-1}$ and $\alpha = -(2k-1)^{-1}$, $\beta = -2n$ respectively.

$$(4.13) \qquad \frac{2k+1}{(2k-1)^{2k-1}}\int_{-\infty}^{\infty}\frac{(xu)^{2k}dx}{(1+x^{2k/2k-1})^{2k}} < \int_{-\infty}^{\infty}(xu')^{2k}dx$$

$$\textit{unless } u = c(1+x^{2k/(2k-1)})^{-1/2k}\,.$$

$$(4.14) \qquad \frac{2n(2k-1)+1}{(2k-1)^{2k-1}}\int_{-\infty}^{\infty}\frac{x^{2n(2k-1)}u^{2k}dx}{(1+x^{2n})^{2k}} < \int_{-\infty}^{\infty}(xu')^{2k}dx$$

$$\textit{unless } u = c(1+x^{2n})^{-1/(2n(2k-1))}\,.$$

The following examples are the extensions of the analogues (for $p > 1$) of the inequalities (3.2.14), (3.2.15).

$$(4.15) \qquad \int_{-1}^{1}\frac{u^{2k}dx}{(1-x^{2k/(2k-1)})^{2k}} < \int_{-1}^{1}u'^{2k}dx \qquad \textit{unless } u \equiv 0\,.$$

$$(4.16) \qquad [n(2k-1)]^{2k}\int_{-1}^{1}\frac{x^{n(2k-1)-1}u^{2k}dx}{(1-x^{2nk})^{2k}} < \int_{-1}^{1}\frac{u'^{2k}dx}{x^{(2k-1)[n(2k-1)-1]}}$$

unless $u \equiv 0$. In (4.16), n is an odd positive integer. The inadmissible extremal functions for these inequalities are

$$(1-x^{2k/(2k-1)})^{(2k-1)/2k},\ (1-x^{2nk})^{(2k-1)/2k}$$

respectively. The case $k = 1$ of (4.15) is due to Nehari [4].

$$(4.17) \quad \left(\frac{2k}{2k-1}\right)^{2k-1}(2n+1)\int_{-1}^{1}\frac{x^{2n}u^{2k}dx}{(1-x^{2k/(2k-1)})^{2k-1}} < \int_{-1}^{1}x^{2n}u'^{2k}dx \qquad (n \geqq 0)$$

unless $u \equiv c(1-x^{2k/(2k-1)})$.

$$(4.18) \qquad (2m)^{2k-1}[2n+(2m-1)(2k-1)]\int_{-1}^{1}\frac{x^{2n+(2m-1)(2k-1)-3}u^{2k}dx}{(1-x^{2m})^{2k-1}}$$

$$< \int_{-1}^{1}x^{2n}u'^{2k}dx \qquad \textit{unless } u \equiv c(1-x^{2m})\,.$$

In this inequality, we assume $m \geqq 1$, $n \geqq 1$.

$$(4.19) \left[\frac{2(n+k)}{2k-1}\right]^{2k-1}\int_{-1}^{1}\frac{u^{2k}dx}{(1-x^{2(n+k)/(2k-1)})^{2k-1}} < \int_{-1}^{1}\frac{u'^{2k}}{x^{2n}}dx \qquad (n \geqq 0)$$

unless $u \equiv c(1-x^{2(n+k)/(2k-1)})$.

$$(4.20) \qquad \left(\frac{2k}{2k+1}\right)^{2k-1}\int_{-1}^{1}\frac{u^{2k}dx}{(1-x^{2k/(2k+1)})^{2k-1}} < \int_{-1}^{1}x^{4k/(2k+1)}u'^{2k}dx$$

unless $u \equiv c(1 - x^{2k/(2k+1)})$. All of the preceding inequalities are sharp. The concept of an extended solution of (4.1) appears only in examples (4.16), (4.19) and (4.20); of these, y' has a discontinuity (at $x = 0$) only in (4.20). In examples (4.13), (4.14), $u(x)$ of course is to satisfy $u(\pm\infty) = 0$, while $u(\pm 1) = 0$ in examples (4.15)—(4.20).

A final example of Theorem 4.1 involving trigonometric functions is given by

$$(4.21) \qquad \left(\frac{2k}{2k-1}\right)^{2k-1}\int_0^\pi csc^2 x\left(\frac{u}{\sin x}\right)^{2k} dx < \int_0^\pi cot^2 x\left(\frac{u'}{\cos x}\right)^{2k} dx$$

when $u(0) = u(\pi) = 0$, *unless* $u \equiv c \sin^{2k/(2k-1)} x$.

We conclude by noting that in the case $p = 2k$, Theorems 3.1.1 and 3.1.2 (and their proofs) remain valid without the restriction $f_1 \geqq 0$, $f_2 \geqq 0$.

References

1. G. H. Hardy, *Note on a theorem of Hilbert*, Math. Zeit., **6** (1920), 314-317.
2. ———, *Notes on some points in the integral calculus* (LXIV), Messenger of Math., **57** (1928), 12-16.
3. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, second edition, Cambridge, 1952.
4. Z. Nehari, *The Schwarzian derivative and Schlicht functions*, Bull. Amer. Math., Soc., **55** (1949), 545-551.

McMaster University, Hamilton, Canada
and
Carleton University, Ottawa, Canada

ON INVARIANT PROBABILITY MEASURES II

J. R. BLUM AND D. L. HANSON

1. Summary. We continue the work begun in [1]. In this paper we investigate convergence properties of sequences of probability measures which are asympototically invariant.

2. Introduction. Let Ω be a set, $\mathscr{A}$ be a σ-algebra of subsets of Ω, and T be a mapping of Ω onto Ω which is one-to-one and bimeasurable. A set $A \in \mathscr{A}$ is said to be invariant if $A = TA$, a probability measure Q defined on $\mathscr{A}$ is invariant if $Q(A) = Q(TA)$ for all $A \in \mathscr{A}$, and an invariant probability measure P is said to be ergodic if every invariant set A is trivial for P, i.e., if $P(A) = 0$ or $P(A) = 1$. Alternately an invariant probability measure P is ergodic if whenever $P(A) > 0$ we have

$$P\Big(\bigcup_{n=-\infty}^{\infty} T^n A\Big) = 1 .$$

Let $\{Q_n\}$ be a sequence of probability measures defined on $\mathscr{A}$. We shall say that the sequence is asymptotically invariant if $\lim_n [Q_n(A) - Q_n(TA)] = 0$ for every $A \in \mathscr{A}$. In § 3 we give a simple condition which yields convergence of such a sequence to a given ergodic measure. In § 4 an example is given which shows that a reasonable conjecture is in fact false, and further conditions are given which insure uniform convergence of a sequence of asymptotically invariant measures. In the last section we investigate convergence properties of certain sequences of probability density functions.

Throughout the paper we shall have occasion to refer to the following theorem, proved in [1]. We state it here as:

THEOREM 1. *If P and Q are invariant measures which agree on the invariant sets then $P = Q$.*

3. A convergence theorem. Let P be an ergodic measure (we shall assume throughout that every measure considered is a probability measure) and let Q be a measure absolutely continuous with respect to P. Define the sequence $\{Q_n\}$ for $n = 1, 2, \cdots$ by the formula

$$Q_n(A) = \frac{1}{n}\sum_{i=0}^{n-1} Q(T^i A) , \quad A \in \mathscr{A}$$

Then it is an immediate consequence of the individual ergodic theorem that $\lim_n Q_n(A) = P(A)$ for every $A \in \mathscr{A}$. Clearly the sequence $\{Q_n\}$ is

Received March 2, 1960.

asymptotically invariant. It is equally clear that the sequence $\{Q_n\}$ is uniformly absolutely continuous with respect to P. It is the object of this section to show that in fact these properties alone are sufficient to insure convergence to P, and that the averaging is only incidental in this case.
More precisely we have

THEOREM 2. *Let P be an ergodic measure and $\{Q_n\}$ a sequence of measures satisfying*
(i) $\lim_n [Q_n(A) - Q_n(TA)] = 0$ *for every* $A \in \mathscr{A}$.
(ii) *For every* $\alpha > 0$ *there exists* $\delta > 0$ *and for every* $A \in \mathscr{A}$ *an integer* $N_{A,\alpha,\delta}$ such that if $P(A) \leqq \delta$ *and* $n \geqq N_{A,\alpha,\delta}$ *then* $Q_n(A) \leqq \alpha$.
Then $\lim_n Q(A) = P(A)$ *for every* $A \in \mathscr{A}$.

Proof. If the conclusion is false there exists $\alpha_0 > 0$, a set $A \in \mathscr{A}$ and a subsequence $\{Q_n\}$ (to avoid multiple subscripting we shall index subsequences in the same way as the original sequence) such that

$$|Q_n(A) - P(A)| \geqq \alpha_0, \text{ all } n . \tag{3.1}$$

Now let Σ be the class of sets $\{\phi, \Omega, T^nA, T^nA^c, n = 0, \pm 1, \cdots\}$, let $\mathscr{F}$ be the smallest field of sets containing Σ, and let $\mathscr{A}'$ be the smallest σ-algebra containing $\mathscr{F}$. We have $\Sigma \subset \mathscr{F} \subset \mathscr{A}' \subset \mathscr{A}$. Note that if $\beta \in \Sigma$ then $TB \in \Sigma$ and $T^{-1}B \in \Sigma$. Now $\mathscr{F}$ consists of finite intersections of finite unions of sets in Σ and it follows from the properties of T that $\mathscr{F}$ has the same property, i.e., T is bimeasurable with respect to $\mathscr{F}$. Let

$$\mathscr{B} = \{A \mid A \in \mathscr{A}', TA \in \mathscr{A}', T^{-1}A \in \mathscr{A}'\} ,$$

Then $\mathscr{F} \subset \mathscr{B} \subset \mathscr{A}'$. Suppose $A \in \mathscr{B}$. Then $TA^c = (TA)^c$ and $T^{-1}A^c = (T^{-1}A)^c$ and it follows that $A^c \in \mathscr{B}$. Similarly let $\{A_n\}$ be a sequence of elements of $\mathscr{B}$. Then $T\bigcup_n A_n = \bigcup_n TA_n$ and $T^{-1}\bigcup_n A_n = \bigcup_n T^{-1}A_n$. It follows that $\mathscr{B}$ is σ-algebra and consequently $\mathscr{B} = \mathscr{A}'$. Thus T is bimeasurable with respect to $\mathscr{A}'$.

Now $\mathscr{F}$ is generated by a denumerable collection of sets and is itself denumerable. By the usual diagonalization procedure we may extract a further subsequence $\{Q_n\}$ which converges on every set of $\mathscr{F}$. Define $Q(B) = \lim_n Q_n(B)$ for $B \in \mathscr{F}$. Since each Q_n is a measure on $\mathscr{F}$ it follows that Q is finitely additive and monotone on $\mathscr{F}$. Note that Q satisfies (3.1); i.e., $|Q(A) - P(A)| \geqq \alpha_0$. We proceed to show that Q is a probability measure on $\mathscr{F}$. Clearly $Q(\Omega) = 1$. Let $\{B_n\}$ be a sequence of sets in $\mathscr{F}$ which decrease to the null set. Then $\{Q(B_n)\}$ is a non-increasing sequence of numbers. Suppose $\lim_n Q(B_n) = \rho > 0$. Let $\alpha = \rho/2$ and choose an appropriate $\delta > 0$ according to (ii) of the hypothesis. Since $\lim_n P(B_n) = 0$ we may choose B_k so that $P(B_k) < \delta$. Then for n sufficiently large $Q_n(B_k) \leqq \rho/2$ and hence $Q(B_k) < \rho$ which is a contradiction. Thus $\rho = 0$ and Q is completely additive $\mathscr{F}$.

Since Q is a measure on $\mathscr{F}$ we may employ the usual Caratheodory technique to extend Q uniquely to $\mathscr{A}'$. From the hypothesis it follows that Q is invariant on $\mathscr{F}$ and the method used in extending Q to $\mathscr{A}'$ insures that Q is invariant on $\mathscr{A}'$.

Now let $B \in \mathscr{A}'$ and suppose B is invariant. Then $P(B) = 0$ or $P(B) = 1$. Suppose $P(B) = 0$. It is clear from the hypothesis that in that case $Q(B) = 0$ and similarly $Q(B) = 1$ if $P(B) = 1$. Thus Q agrees with P on the invariant elements of $\mathscr{A}'$, and it follows from Theorem 1 that $Q = P$ on $\mathscr{A}'$. In particular $Q(A) = P(A)$, which is a contradiction. The theorem is proved.

Theorem 2 has an interesting corollary. Consider the condition

$$\lim_n \frac{1}{n} \sum_{i=0}^{n-1} P(T_i A \cap B) = P(A)P(B) \text{ for all } A, B \in \mathscr{A}. \tag{3.2}$$

It is trivial to verify that if (3.2) holds then P is ergodic. Conversely if P is ergodic one may verify (3.2) by using the individual ergodic theorem. However (3.2) is also an immediate consequence of Theorem 2. It is clearly sufficient to consider the case when $P(B) > 0$. In that case define the sequence $\{Q_n\}$ by

$$Q_n(A) = \frac{1}{P(B)} \frac{1}{n} \sum_{i=0}^{n-1} P(T^i A \cap B) .$$

It follows at once that the hypotheses of Theorem 2 apply and (3.2) holds.

4. On uniform convergence. The converse of Theorem 2 evidently holds. If $\lim_n Q_n(A) = P(A)$ for every $A \in \mathscr{A}$ then (i) and (ii) of Theorem 2 are true. Furthermore if $\lim_n Q_n(A) = P(A)$ uniform for $A \in \mathscr{A}$ then $\lim_n [Q_n(A) - Q_n(TA)] = 0$ uniformly for $A \in \mathscr{A}$. It might therefore be reasonable to except that if hypothesis (i) of Theorem 2 is strengthened to $\lim_n [Q_n(A) - Q_n(TA)] = 0$ uniformly for $A \in \mathscr{A}$ we might obtain uniform convergence of Q_n to P. The following example, which is of some independent interest, shows that this is not the case. Let Ω be the unit interval closed on the left and open on the right, and $\mathscr{A}$ the Borel sets. Define T by $Tx = (x + c) \bmod 1$, where c is an irrational number. Then T is one-to-one, onto, and bimeasurable. Let P be Lebesgue measure. Clearly P is invariant and it can be shown that P is ergodic. For $n = 4, 5, \cdots$ let $A_n = [0, 1/n]$. Since $P(A_n) > 0$ we have

$$P\Big(\bigcup_{i=-\infty}^{\infty} T^i A_n\Big) = 1$$

and consequently for each n there is a unique first integer k_n such that

$$1/4 \leqq P\Big(\bigcup_{-k_n}^{k_n} T^i A_n\Big) \leqq 3/4 .$$

Let

$$B_n = \bigcup_{-k_n}^{k_n} T^i A_n$$

and let $b_n = P(B_n)$. Define the sequence $\{Q_n\}$ by $Q_n(A) = P(AB_n)/b_n$. Since $b_n \geqq 1/4$ it follows that the probability measures Q_n are uniformly absolutely continuous with respect to P. Furthermore

$$\begin{aligned} |\, Q_n(A) - Q_n(TA) \,| &= (1/b_n)\,|\, P(AB_n) - P(TAB_n) \,| \\ &\leqq 4\,|\, P(TATB_n) - P(TAB_n) \,| \\ &\leqq 4[P(TA(TB_n - B_n)) + P(TA(B_n - TB_n))] \\ &\leqq 4[P(TB_n - B_n) + P(B_n - TB_n)] . \end{aligned}$$

Now

$$TB_n - B_n \subset T^{k_n+1} A_n$$

and

$$B_n - TB_n \subset T^{-k_n} A_n .$$

Hence

$$|\, Q_n(A) - Q_n(TA) \,| \leqq 8P(A_n) = 8/n .$$

Thus

$$\lim_n \sup_{A \in \mathscr{A}} |\, Q_n(A) - Q_n(TA) \,| = 0 .$$

On the other hand $Q_n(B_n) - P(B_n) \geqq 1 - 3/4 = 1/4$ and we do not have uniform convergence. The remainder of this section remain is devoted to exhibiting conditions under which one does obtain uniform convergence of the sequence $\{Q_n\}$ to P. For this purpose we shall need several lemmas.

LEMMA 1. *Let P be an invariant measure and Q be an arbitrary measure. Then*

$$\sup_{\substack{i \\ A \in \mathscr{A}}} |\, Q(A) - Q(T^i A) \,| \leqq 2 \sup_{A \in \mathscr{A}} |Q(A) - P(A)\,| .$$

Proof.

$$\begin{aligned} |\, Q(A) - Q(T^i A) \,| \leqq \,&|\, Q(A) - P(A) + |\, P(A) - P(T^i A) \,| \\ &+ |\, P(T^i A) - Q(T^i A) \,| . \end{aligned}$$

Since P is invariant the middle term on the right vanishes and the lemma follows.

LEMMA 2. *Let P be an ergodic measure and f be a non-negative measurable function which is integrable with respect to P. Then for every $A \in \mathscr{A}$ and $\alpha > 0$ there exist infinitely many values of n such that*

$$\int_{T^n A} f(x)dP(x) < \left(\int_\Omega f(x)dP(x)\right) P(A) + \alpha .$$

Proof. Let $\beta = \int_\Omega fdP$. If $\beta = 0$ there is nothing to prove. Consequently assume $\beta > 0$. Define the measure Q by $Q(A) = \int_A fdP/\beta$ for $A \in \mathscr{A}$, and the sequence $\{Q_n\}$ by

$$Q_n(A) = \sum_{i=0}^{n-1} Q(T^i A)/n .$$

Since Q is absolutely continuous with respect to P it follows Theorem 2 that $\lim_n Q_n(A) = P(A)$ for $A \in \mathscr{A}$. If the conclusion of the lemma is false then for some $A \in \mathscr{A}$ and $\alpha > 0$ we have for sufficiently large n, $\int_{T^n A} fdP/\beta = Q(T^n A) \geqq P(A) + \alpha/2$. But then $\lim_n Q_n(A) > P(A)$ which is a contradiction.

LEMMA 3. *Let P be an ergodic measure and Q be a measure which is absolutely continuous with respect to P. Then*

$$\sup_{A \in \mathscr{A}} | Q(A) - P(A) | \leqq 2 \sup_{\substack{i \\ A \in \mathscr{A}}} | Q(A) - Q(T^i A) | .$$

Proof. Let f be the Radon-Nikodym derivative of Q with respect to P, and let $B = \{x \mid f(x) \geqq 1\}$. Then

$$\sup_{A \in \mathscr{A}} | Q(A) - P(A) | = \int_B [f(x) - 1]dP(x) .$$

Assume that $P(B) \leqq 1/2$; in the contrary case we can use B^c. Now if i is any integer we have

$$\begin{aligned}\sup_{\substack{i \\ A \in \mathscr{A}}} | Q(A) - Q(T^i A) | &\geqq Q(B) - Q(T^i B) \\ &= [Q(B) - P(B)] - [Q(T^i B) - P(T^i B)] \\ &= \int_B [f - 1]dP - \int_{T^i B} [f - 1]dP .\end{aligned}$$

Hence

$$\begin{aligned}\sup_{A \in \mathscr{A}} |Q(A) - P(A)| &\leqq \sup | Q(A) - Q(T^i A) | + \int_{T^i B} [f - 1]dP \\ &\leqq \sup_{\substack{i \\ A \in \mathscr{A}}} | Q(A) - Q(T^i A) | + \int_{B \cap T^i B} [f - 1]dP .\end{aligned}$$

Now let

$$g(x) = \begin{cases} f(x) - 1,\ x \in B \\ 0,\ x \in B^c \end{cases} .$$

Then

$$\int_{B \cap T^i B} [f - 1] dP = \int_{T^i B} g dP .$$

Let $\alpha > 0$. Then from Lemma 2 there exist integers i such that

$$\int_{T^i B} g dP < \left(\int_{\Omega} g dP\right) P(B) + \alpha .$$

But $P(B) \leqq 1/2$ and

$$\int_{\Omega} g dP = \int_{B} [f - 1]\, dP \leqq \sup_{A \in \mathscr{A}} | Q(A) - P(A) | .$$

Hence

$$\int_{T^i B} g dP < 1/2 \sup_{A \in \mathscr{A}} | Q(A) - P(A) | + \alpha$$

and we obtain

$$\sup_{A \in \mathscr{A}} | Q(A) - P(A) | \leqq 2 \sup_{\substack{i \\ A \in \mathscr{A}}} | Q(A) - Q(T^i A) | + \alpha$$

for abritrary $\alpha > 0$.

THEOREM 3. *Let P be an ergodic measure and let $\{Q_n\}$ be a sequence of measures each of which is absolutely continuous with respect to P. Then*

$$\lim_n \sup_{A \in \mathscr{A}} | Q_n(A) - P(A) | = 0$$

if and only if

$$\lim_n \sup_{\substack{i \\ A \in \mathscr{A}}} | Q_n(A) - Q_n(T^i A) | = 0 .$$

Proof. The theorem follows from Lemmas 1 and 3. Theorem 3 may also be formulated in terms of L_1 convergence. For if f_n is the Radon-Nikodym derivative of Q_n with respect to P, then

$$\sup_{A \in \mathscr{A}} | Q_n(A) - P(A) | = \int_{\{f_n > 1\}} [f_n - 1] dP = \int_{\{f_n < 1\}} [1 - f_n] dP .$$

Thus

$$\lim_n \sup_{A \in \mathscr{A}} | Q_n(A) - P(A) | = 0$$

if and only if

$$\lim_n \int_\Omega |f_n - 1| \, dP = 0 \, .$$

Similarly we have

$$\lim_n \sup_{\substack{i \\ A \in \mathscr{A}}} | Q_n(A) - Q_n(T^i A) | = 0$$

if and only if

$$\lim_n \sup_i \int_\Omega |f_n(x) - f_n(T^i x)| \, dP(x) = 0 \, .$$

Consequently we have the

COROLLARY. *Let P be an ergodic measure, let $\{Q_n\}$ be a sequence of measures each of which is absolutely continuous with respect to P, and let $\{f_n(x)\}$ be the corresponding sequence of Radon-Nikodym derivaties. Then*

$$\lim_n \int_\Omega |f_n(x) - 1| \, dP(x) = 0$$

if and only if

$$\lim_n \sup_i \int_\Omega |f_n(x) - f_n(T^i x)| \, dP(x) = 0 \, .$$

5. **Uniform convergence of densities.** In this section we shall be concerned with probability density functions with respect to an ergodic measure P, i.e., a function f is a probability if f is measurable, non-negative, and $\int_\Omega f dP = 1$. We begin with

LEMMA 4. *Let P be an ergodic measure and let f be a probability density with respect to P. Let $\alpha > 0$ and define the sets A and B by*

$$A = \left\{ x \,\middle|\, \sup_{i,j} | f(T^i x) - f(T^j x) | < \alpha \right\}$$

and

$$B = \{ x \mid | f(x) - 1 | > \alpha \} \, .$$

Then $P(AB) = 0$.

Proof. Let $B' = \{x \mid f(x) > 1 + \alpha\}$. Suppose $P(AB') > 0$. Let $C = \bigcup_{-\infty}^{\infty} T^i(AB')$. Since P is ergodic $P(C) = 1$. If $x \in C$ there exists an integer m such that $T^m x \in AB'$. Hence

$$\sup_{i,j} |f(T^i x) - f(T^j x)| \leqq \sup_{\substack{i,j \\ x \in A}} |f(T^i x) - f(T^j x)| \leqq \alpha .$$

In particular $|f(x) - f(T^m x)| \leqq \alpha$ or $f(x) \geqq f(T^m x) - \alpha$. But $T^m x \in B'$ which means $f(x) > 1$. Since an integer m can be found for each $x \in C$ we have $f(x) > 1$ for all $x \in C$. Then $\int_\Omega f dP = \int_C f dP > 1$, a contradiction to the fact that f is a probability density. A similar argument applies to the set $B'' = \{x \mid f(x) < 1 - \alpha\}$.

THEOREM 4. *Let P be an ergodic measure and let $\{f_n\}$ be a sequence of probability densities with respect to P. Then the following statements are equivalent*:

(i) $P\left(\lim_n \sup_{i,j} |f_n(T^i x) - f_n(T^j x)| = 0\right) > 0$.

(ii) $P\left(\lim_n \sup_{i,j} |f_n(T^i x) - f_n(T^j x)| = 0\right) = 1$.

(iii) $P\left(\lim_n \sup_x |f_n(x) - 1| = 0\right) = 1$.

Proof.

(a) (i) implies (ii). Suppose (i) is true. Let B be a set such that $P(B) > 0$ and such that

$$\lim_n \sup_{i,j} |f_n(T^i x) - f_n(T^j x)| = 0$$

for $x \in B$. But clearly this is also true for

$$x \in C = \bigcup_{i=-\infty}^{\infty} T^i B ,$$

and $P(C) = 1$. Thus (ii) holds.

(b) (ii) implies (iii). Let C be the set of measure one such that for $x \in C$ we have

$$\lim_n \sup_{i,j} |f_n(T^i x) - f_n(T^j)| = 0 .$$

Then for $x \in C$ and every positive integer k there exists a positive integer N_k such that

$$\sup_{i,j} |f_n(T^i x) - f_n(T^j x)| < 1/k$$

for $n \geqq N_k$. Let

$$A_k = \bigcup_{n \geqq N_k} \left\{x \middle| |f_n(x) - 1| > 1/k\right\} .$$

It follows from Lemma 4 that $P(A_k) = 0$ for $k = 1, 2, \cdots$. Let $A =$

$C - \bigcup_k A_k$. Then $P(A) = 1$, and for $x \in A$ we have $|f_n(x) - 1| \leqq 1/k$ for $n \geqq N_k$. Consequently

$$\lim_n \sup_{x \in A} |f_n(x) - 1| = 0$$

and (iii) follows.

(c) (iii) implies (i). Let A be the set of measure one such that

$$\lim_n \sup_{x \in A} |f_n(x) - 1| = 0 .$$

Let

$$A_0 = \bigcap_{i=-\infty}^{\infty} T^i A .$$

Then $P(A_0) = 1$ and for $x \in A_0$ we have

$$\sup_{i,j} |f_n(T^i x) - f_n(T^j x)| \leqq 2 \sup_i |f_n(T^i x) - 1|$$

and the last quantity approaches zero. Thus (i) holds and the theorem is proved.

Reference

1. J. R. Blum and D. L. Hanson, *On Invariant Probability Measures I*, Pacific J. Math., **10** (1960), 1125–1129.

Sandia Corporation and Indiana University

SYMMETRY IN GROUP ALGEBRAS OF DISCRETE GROUPS

ROBERT A. BONIC[1]

1. Introduction. The Banach algebras $\mathfrak{A}$ considered here are over the field of complex numbers, and have isometric involutions $*$. The involution is said to be *hermitian* if for any $x = x^* \in \mathfrak{A}$, the spectrum $Sp_{\mathfrak{A}}(x)$ of x contains only real numbers. The algebra $\mathfrak{A}$ is said to be *symmetric* if for any $y \in \mathfrak{A}$, $Sp_{\mathfrak{A}}(y^*y)$ contains only nonnegative real numbers.

A familiar example of a Banach algebra with an involution is the group algebra over the complex numbers of a locally compact group G. This is obtained by taking the Banach space $L^1(G)$ of all complex valued absolutely integrable functions with respect to the left invariant Haar measure dx on G. Multiplication is defined as convolution, and the involution by the formula $x^*(g) = \overline{x(g^{-1})}\rho(g)$, where $x \in L^1(G)$ and $\rho(\cdot)$ is the modular function relating the given measure to the right invariant measure by $dx^{-1} = \rho(x)dx$. This involution will be called the *natural involution* of the group algebra, and is the only involution on the group algebra we will consider.

It is known that when the group G is either compact or commutative, then its group algebra with respect to the natural involution is symmetric. On the other hand, in 1948 Neumark [6] showed that the natural involution in the group algebra of the homogeneous Lorentz group is not hermitian. (This implies that the algebra is not symmetric. See Theorem $A(a)$.) Later Gelfand and Neumark [3] extended this example to include all complex unimodular groups. Their proofs are quite difficult, entailing a knowledge of the irreducible unitary representations of the groups and considerable computation. Except for finite and commutative groups, the corresponding problems have not been studied for discrete groups. These problems will be our concern.

The main results will be summarized now. In § 2 several facts (some of which are well known) are collected to be used later. § 3 is concerned with the construction of group algebras that are symmetric, or at least have an hermitian involution. It is shown (Corollary 3.4) that the group algebra of the direct product of a commutative group and a group whose group algebra is symmetric, is a symmetric algebra. Theorem 3.7 shows that the natural involution is hermitian in the group algebra of a semidirect product of a commutative group by a finite group.

Received March 1, 1960.

[1] The work in this paper consists of a portion of the author's Yale doctoral dissertation (1960), written under the direction of Professor C. E. Rickart.

In § 4 examples are given of discrete groups for which the natural involution in the group algebra is not hermitian. The examples include free groups on two or more generators, and free groups on three or more generators of order two (Theorem 4.7). It is worth noting that these examples settle the following matrix problem negatively: suppose T is a bounded operator on l^1 (countable absolutely convergent sequences of complex numbers), and with respect to the usual basis, suppose that the matrix (t_{ij}) of T satisfies $t_{ij} = \overline{t_{ji}}$. Then, is the spectrum of the operator T a subset of the real axis? In this connection see Remark 4.8.

Finally in § 5 we show that various connections exist between the above problems and the question of the existence of an invariant mean on the group. The principal results are Theorem 5.6 and Theorem 5.8.

I wish to express my thanks to Professors C. E. Rickart and S. Kakutani for the advice, encouragement, and time they have generously given me.

2. Preliminary theorems.

THEOREM A. *Let $\mathfrak{A}$ be a Banach algebra over the complex numbers with an isometric involution * and identity e. Then*:

(a) *if $\mathfrak{A}$ is symmetric, then the involution is hermitian, and the converse holds whenever $\mathfrak{A}$ is commutative;*

(b) *the involution is hermitian whenever $ie + x$ is regular for any $x = x^*$.*

A proof of this theorem can be found in Rickart [7]. It is not known in general if an algebra with an hermitian involution is symmetric, and it is worth noting that this is exactly the problem in proving that a B^* algebra is a C^* algebra. The essential step in proving this is to show that the B^* algebra (whose involution is hermitian) is symmetric.

Let $\mathfrak{A}$ be a Banach algebra with an identity e of norm one, and let $\mathscr{B}(\mathfrak{A})$ denote the set of all bounded linear operators on $\mathfrak{A}$. For $x \in \mathfrak{A}$, *the left multiplication operator* L_x is defined by the formula $L_x y = xy$.

THEOREM B. (a) *The mapping $x \to L_x$ maps $\mathfrak{A}$ isometrically and isomorphically* into $\mathscr{B}(\mathfrak{A})$.

(b) *Let $\mathscr{L}(\mathfrak{A})$ denote the image of $\mathfrak{A}$ in $\mathscr{B}(\mathfrak{A})$. Then for $x \in \mathfrak{A}$, $Sp_{\mathfrak{A}}(x) = Sp_{\mathscr{L}(\mathfrak{A})}(L_x) = Sp(L_x)$, where $Sp(L_x) = \{\alpha : L_x - \alpha I$ is a singular operator on the Banach space$\}$.*

Proof. (a) and the first identity in (b) are immediate. If y is regular in $\mathfrak{A}$, then L_y is a regular operator on $\mathfrak{A}$, since it has as inverse the operator $L_{y^{-1}}$. This shows that $Sp_{\mathfrak{A}}(x) \supset Sp(L_x)$. Now if L_y is regular on $\mathscr{B}(\mathfrak{A})$, there exists an element $S \in \mathscr{B}(\mathfrak{A})$ such that $L_y S = SL_y = I$. It is then easily computed that $S = L_{Se}$ and Se is the inverse of y in $\mathfrak{A}$.

THEOREM C. *Let $\mathfrak{C}$ be a Banach algebra with an identity over the complex numbers and suppose C is a maximal left ideal* (hence closed) *in $\mathfrak{C}$. Then, with respect to the quotient norm, $\mathfrak{C}/C$ is a Banach space. For $x \in \mathfrak{C}$, $y + C \in \mathfrak{C}/C$ the mapping defined by $L_x^{\mathfrak{C}/C}(y + C) = xy + C$ gives a bounded algebraically irreducible representation of $\mathfrak{C}$ on $\mathfrak{C}/C$.*

The above representation $x \rightarrow L_x^{\mathfrak{C}/C}$ is called the left regular representation of $\mathfrak{C}$ on $\mathfrak{C}/C$. A proof of this theorem can also be found in Rickart [7].

3. Group algebras in which the natural involution is hermitian. It will be seen shortly that the symmetry problem for the group algebra of the direct product of two groups is a special case of a more general problem concerning tensor products of Banach algebras, so the latter will be taken up first. If $\mathfrak{A}$ and $\mathfrak{B}$ are Banach algebras, then the algebraic tensor product $\mathfrak{A} \otimes \mathfrak{B}$ can be normed with the so called *greatest cross norm* and then completed to give another Banach algebra called the *projective tensor product* $\mathfrak{A} \hat{\otimes} \mathfrak{B}$ of $\mathfrak{A}$ and $\mathfrak{B}$. The basic results concerning this can be found in Schatten [9]. We will summarize here only a few pertinent facts.

Let $\mathfrak{A}$ and $\mathfrak{B}$ be Banach algebras over the complex numbers having identities of norm one. It will be convenient for us not to distinguish notationally between the norms or the identities in the two algebras. Let $\mathfrak{A} \otimes \mathfrak{B}$ denote the usual algebraic tensor product of the vector spaces $\mathfrak{A}$ and $\mathfrak{B}$. An element $u \in \mathfrak{A} \otimes \mathfrak{B}$ can be represented in many ways in the form $\sum_{i=1}^n a_i \otimes b_i$ where $a_i \in \mathfrak{A}$, $b_i \in \mathfrak{B}$, $i = 1, 2, \cdots, n$. Whenever such a representation occurs, it will be denoted by $u \sim \sum_{i=1}^n a_i \otimes b_i$. The set $\mathfrak{A} \otimes \mathfrak{B}$ becomes an algebra by defining, for $u, v \in \mathfrak{A} \otimes \mathfrak{B}$, a representation of the product uv to be $\sum_{i=1}^n \sum_{j=1}^m a_i c_j \otimes b_i d_j$, where $u \sim \sum_{i=1}^n a_i \otimes b_i$, $v \sim \sum_{j=1}^m c_j \otimes d_j$. It becomes a normed algebra by defining $\|u\| = GLB \sum_{i=1}^n \|a_i\| \cdot \|b_i\|$ where the GLB is extended over all $\sum_{i=1}^n a_i \otimes b_i \sim u$. With this norm, any $u \in \mathfrak{A} \otimes \mathfrak{B}$ that satisfies $u \sim a \otimes b$ has a norm given by $\|u\| = \|a\| \cdot \|b\|$, and the identity $e \sim e \otimes e$ of $\mathfrak{A} \otimes \mathfrak{B}$ has norm one. The completion $\mathfrak{A} \hat{\otimes} \mathfrak{B}$ of $\mathfrak{A} \otimes \mathfrak{B}$ is hence a Banach algebra over the complex numbers with an identity of norm one. Finally we note that if $\mathfrak{A}$ and $\mathfrak{B}$ each have isometric involutions $*$, the definition of $u^* \sim \sum_{i=1}^n a_i^* \otimes b_i^*$ where $u \sim \sum_{i=1}^n a_i \otimes b_i$ gives a well-defined isometric involution on $\mathfrak{A} \otimes \mathfrak{B}$ which can hence be extended to $\mathfrak{A} \hat{\otimes} \mathfrak{B}$.

We now restrict ourselves to commutative $\mathfrak{A}$. Let $\Phi(\mathfrak{A})$ denote the space of maximal ideals of $\mathfrak{A}$, which will we be identify with the corresponding homomorphisms. For $h \in \Phi(\mathfrak{A})$, define T_h: $\mathfrak{A} \otimes \mathfrak{B} \rightarrow \mathfrak{B}$ by the formula $T_h(u) = \sum_{i=1}^n h(a_i) b_i$. Now, two formal sums represent the same element in $\mathfrak{A} \otimes \mathfrak{B}$ if and only if one can be transformed into the other by successive applications of the distributive law and the commutative law applied to scalars and $\otimes$. It is then clear that T_h is well-defined and a homomorphism. Also if $\mathfrak{A}$ and $\mathfrak{B}$ have involutions, and $\mathfrak{A}$ is symmetric, then for $u \sim \sum_{i=1}^n a_i \otimes b_i$, we have

$$T_h(u^*) = T_h\Big(\sum_{i=1}^n a_i^* \otimes b_i^*\Big) = \sum_{i=1}^n h(a_i^*)b_i^* = \sum_{i=1}^n \overline{h(a_i)}b_i^* = \Big(\sum_{i=1}^n h(a_i)b_i\Big)^* \\ = (T_h u)^* ,$$

so that T_h is a *-homomorphism. Finally since

$$\| T_h u \| = \Big\| \sum_{i=1}^n h(a_i)b_i \Big\| \leqq \sum_{i=1}^n \| a_i \| \cdot \| b_i \|$$

holds for any $\sum_{i=1}^n a_i \otimes b_i \sim u$, we have $\| T_h u \| \leqq \| u \|$ for all u, so that T_h can be extended to $\mathfrak{A} \hat{\otimes} \mathfrak{B}$. The extension will also be denoted by T_h.

Except for the notation, the following theorem is essentially the same as that of Bochner and Phillips [1: Theorem 3], which generalizes the Wiener-Gelfand theorem on the existence of an inverse.

THEOREM 3.1. *An element* $u \in \mathfrak{A} \hat{\otimes} \mathfrak{B}$ *has a left* (right) *inverse in* $\mathfrak{A} \hat{\otimes} \mathfrak{B}$ *if and only if* $T_h u$ *has a left* (right) *inverse in* $\mathfrak{B}$ *for every* $h \in \Phi(\mathfrak{A})$.

Proof. Only the case of left inverses will be shown. If $u \in \mathfrak{A} \hat{\otimes} \mathfrak{B}$ has a left inverse v, then for any $h \in \Phi(\mathfrak{A})$, $T_h v$ is a left inverse in $\mathfrak{B}$ for $T_h u$, since T_h is a homomorphism taking the identity of $\mathfrak{A} \hat{\otimes} \mathfrak{B}$ to the identity of $\mathfrak{B}$.

Conversely assume that $u_0 \in \mathfrak{A} \hat{\otimes} \mathfrak{B} = \mathfrak{C}$, that $T_h u_0$ has a left inverse in $\mathfrak{B}$ for every $h \in \Phi(\mathfrak{A})$, and that u_0 does not have a left inverse in $\mathfrak{C}$. Then $\mathfrak{C}u_0$ is a proper left ideal containing u_0 and can be extended to a maximal left ideal C. Now consider the left regular representation $u \to L_u^{\mathfrak{C}/C}$ of $\mathfrak{C}$ on $\mathfrak{C}/C$ (see Theorem C). Since this representation is algebraically irreducible, it follows from Theorem C that the set of all bounded operators on $\mathfrak{C}/C$ commuting with $\{L_u^{\mathfrak{C}/C} : u \in \mathfrak{C}\}$ consits of just scalar multiples of the identity operator. Clearly $L_{a\otimes e}^{\mathfrak{C}/C}$ commutes with all $L_u^{\mathfrak{C}/C}$ so that $L_{a\otimes e}^{\mathfrak{C}/C} = h(a)I$, and since

$$(L_{a_1\otimes e}^{\mathfrak{C}/C})(L_{a_2\otimes e}^{\mathfrak{C}/C}) = L_{a_1 a_2\otimes e}^{\mathfrak{C}/C} ,$$

it follows that h is an element of $\Phi(\mathfrak{A})$. Hence

$$L_{a\otimes b}^{\mathfrak{C}/C} = (L_{a\otimes e}^{\mathfrak{C}/C})(L_{e\otimes b}^{\mathfrak{C}/C}) = h(a)(L_{e\otimes b}^{\mathfrak{C}/C}) = L_{e\otimes h(a)b}^{\mathfrak{C}/C}$$

so that for $u \sim \sum_{i=1}^n a_i \otimes b_i$, we have

$$L_u^{\mathfrak{C}/C} = L_{\sum_{i=1}^n a_i\otimes b_i}^{\mathfrak{C}/C} = \sum_{i=1}^n L_{a_i\otimes b_i}^{\mathfrak{C}/C} = \sum_{i=1}^n L_{e\otimes h(a_i)b_i}^{\mathfrak{C}/C} L_{e\otimes \sum_{i=1}^n h(a_i)b_i}^{\mathfrak{C}/C} = L_{e\otimes T_h u}^{\mathfrak{C}/C} .$$

Moreover, since the representation is continuous we can extend this to $\mathfrak{C}$ so that we have $L_v^{\mathfrak{C}/C} = L_{e\otimes T_h v}^{\mathfrak{C}/C}$ for all $v \in \mathfrak{A} \hat{\otimes} \mathfrak{B}$.

Now by assumption $T_h u_0$ has a left inverse $b_0 \in \mathfrak{B}$. Hence

$$\begin{aligned}((L_{e\otimes b_0}^{\mathfrak{C}/C})(L_{u_0}^{\mathfrak{C}/C}))(e \otimes e + C) &= ((L_{e\otimes b_0}^{\mathfrak{C}/C})(L_{e\otimes T_h u_0}^{\mathfrak{C}/C}))(e \otimes e + C) \\ &= (L_{e\otimes b_0 T_h u_0}^{\mathfrak{C}/C})(e \otimes e + C) = (L_{e\otimes e}^{\mathfrak{C}/C})(e \otimes e + C) = e \otimes e + C .\end{aligned}$$

On the other hand, since $u_0 \in C$, we have

$$\begin{aligned}((L_{e\otimes b_0}^{\mathfrak{C}/C})(L_{u_0}^{\mathfrak{C}/C}))(e \otimes e + C) &= (L_{e\otimes b_0}^{\mathfrak{C}/C})(L_{u_0}^{\mathfrak{C}/C}(e \otimes e + C)) \\ &= (L_{e\otimes b_0}^{\mathfrak{C}/C})(u_0 + C) = (L_{e\otimes b_0}^{\mathfrak{C}/C})(C) = C\end{aligned}$$

and we have obtained a contradiction.

Since an element is regular if and only if it has a right inverse and a left inverse we have:

COROLLARY 3.2. *An element* $u \in \mathfrak{A} \hat{\otimes} \mathfrak{B}$ *is regular if and only if* $T_h u$ *is regular in* $\mathfrak{B}$ *for every* $h \in \Phi(\mathfrak{A})$. *More precisely*:

$$Sp_{\mathfrak{A}\hat{\otimes}\mathfrak{B}}(u) = \bigcup_{h\in\Phi(\mathfrak{A})} Sp_{\mathfrak{B}}(T_h u) .$$

COROLLARY 3.3. *If* $\mathfrak{A}$ *is symmetric and* $\mathfrak{B}$ *is symmetric* (has an hermitian involution), *then* $\mathfrak{A} \hat{\otimes} \mathfrak{B}$ *is symmetric* (has an hermitian involution).

Proof. If $\mathfrak{B}$ has an hermitian involution, then for $u = u^* \in \mathfrak{A} \hat{\otimes} \mathfrak{B}$ it follows that $(T_h u)^* = T_h u$ for all $h \in \Phi(\mathfrak{A})$. By the preceding corollary, $Sp_{\mathfrak{A}\hat{\otimes}\mathfrak{B}}(u)$ is a subset of the real axis. The "symmetry argument" is similar.

The following theorem is a special case of a theorem due to Grothendick [4: Théorème 2], and gives the connection between tensor products and group algebras.

THEOREM (Grothendieck). *If G and H are locally compact groups, then after a suitable normalization* $L^1(G) \hat{\otimes} L^1(H)$ *is isometrically * isomorphic to* $L^1(G \times H)$. ($G \times H$ denotes the direct product of the groups G and H).

The proof of this theorem is not easy. However our concern in the following corollary is with discrete groups, and for this special case the proof is quite direct. In any event, assuming this theorem, Corollary 3.3 gives:

COROLLARY 3.4. *If G is a discrete abelian group and H an arbitrary discrete group whose group algebra is symmetric* (has an hermitian involution), *then the group algebra of* $G \times H$ *is symmetric* (has an hermitian involution).

The case of semi-direct products will now be taken up.

DEFINITION 3.5. Let K and C be groups and suppose that for each $c \in C$ there is an automorphism φ_c of K such that the mapping $c \rightarrow \varphi_c$ is a homomorphism of C onto a group of automorphisms of K. The set of ordered pairs $\{\langle c, k\rangle : c \in C, k \in K\}$ with multiplication defined by $\langle c_1, k_1\rangle\langle c_2, k_2\rangle = \langle c_1c_2, k_1\varphi_{c_1}(k_2)\rangle$ then forms a group G called the *semi-direct product of* K *and* C *by* φ and denoted by $C \times_\varphi K$.

It is immediately verified that the set $\{\langle e, k\rangle : k \in K\}$ forms a normal subgroup of G isomorphic to K, and that $\langle c, k\rangle^{-1} = \langle c^{-1}, \varphi_{c-1}(k^{-1})\rangle$. The generality of semi-direct products is shown by the following theorem.

THEOREM 3.6. *If a group G contains subgroups K and C, where K is normal, $K \cap C = e$, and $G = KC$, then G is isomorphic to a semi-direct product of C and K.*

Proof. Since K is normal the mapping $c \rightarrow \varphi_c$, where $\varphi_c(k) = ckc^{-1}$, is a homomorphism of C onto a group of automorphisms of K. Since $G = KC$, any $g \in G$ can be written in the form $g = k_gc_g$ where $k_g \in K$, $c_g \in C$, and since $K \cap C = e$ this decomposition is unique. Then

$$gh = k_gc_gk_hc_h = k_gc_gk_hc_g^{-1}c_gc_h = k_g\varphi_{c_g}(k_g)c_gc_h$$

so that $c_{gh} = c_gc_h$, and $k_{gh} = k_g\varphi_{c_g}(k_h)$. It is now obvious that the correspondence $g \leftrightarrow \langle c_g, k_g\rangle$ is an isomorphism between G and the semi-direct product $C \times_\varphi K$.

Before stating the next theorem, it is convenient to establish some special conventions. The group algebra of a discrete group G will be denoted by $l^1(G)$, and elements of $l^1(G)$ will be written as sums rather than functions, i.e. if $x \in l^1(G)$, then $x = \sum_{g\in G} x(g)g$, where the $x(g)$ are complex numbers satisfying $\sum_{g\in G} |x(g)| < \infty$. Convolution in $l^1(G)$ is then the usual multiplication of these formal sums, and the involution is given by $x^* = \sum_{g\in G} \overline{x(g^{-1})}g$.

Let $G = C \times_\varphi K$ be a semi-direct product of C and K. We will abuse notation and consider C and K as subgroups of G. This is justified by Theorem 3.6. The elements of G can then be uniquely written in the form $g = kc$ and $gg' = kck'c' = k\varphi_c(k')cc'$ where $k, k' \in K$ and $c, c' \in C$. Finally, for $x \in l^1(G)$, we have

$$x = \sum_{k\in K}\sum_{c\in C} x(kc)kc = \sum_{c\in C}\left(\sum_{k\in K} x(kc)k\right)c = \sum_{c\in C} x'(c)c$$

where $x'(c) \in l^1(K)$. Dropping the primes, we will now write any element $x \in l^1(G)$ in the form $x = \sum_{c\in C} x(c)c$, $x(c) \in l^1(K)$.

THEOREM 3.7. *If C is a finite group and K is a discrete abelian group, then the natural involution is hermitian in the group algebra of any semi-direct product $G = C \times_\varphi K$.*

Proof. Let $G = C \times_\varphi K$ be a semi-direct product of C and K, and let $x = x^* \in l^1(G)$, $x = \sum_{c \in C} x(c)c$. Then

$$x^* = \sum_{c \in C} (x(c)c)^* = \sum_{c \in C} c^{-1}(x(c)^*) = \sum_{c \in C} c^{-1}(x(c)^*)cc^{-1}$$
$$= \sum_{c \in C} \Phi_{c^{-1}}(x(c)^*)c^{-1} = \sum_{c \in C} \Phi_c(x(c^{-1})^*)c \; .$$

(Φ_c denotes the extension of φ_c to $l^1(K)$ defined by $\Phi_c(\sum_{k \in K} x(k)k) = \sum_{k \in K} x(k)\varphi_c(k)$.) Since $x = x^*$ and the decomposition is unique we have $\Phi_c((x(c^{-1}))^*) = x(c)$ for all $c \in C$.

By Theorem A (b) the involution in $l^1(G)$ is hermitian if $ie + x$ is regular for all $x = x^*$. We will now construct a right inverse for $ie + x$. Indeed $ie + x$ will have a right inverse if and only if elements $y(c) \in l^1(K)$ can be found for each $c \in C$ such that $y = \sum_{c \in C} y(c)c$ satisfies $(ie + x)y = e$. Expressing this condition in terms of the coefficients we have:

$$e = \Big(ie + \sum_{c \in C} x(c)c\Big)\Big(\sum_{d \in C} y(d)d\Big) = i\sum_{d \in C} y(d)d + \sum_{c,d \in C} x(c)cy(d)d$$
$$= i\sum_{d \in C} y(d)d + \sum_{c,d \in C} x(c)\Phi_c(y(d))cd = i\sum_{b \in C} y(b)b + \sum_{b \in C}\Big(\sum_{c \in C} x(c)\Phi_c(y(c^{-1}b))\Big)b$$
$$= \sum_{b \in C} (iy(b) + \sum_{c \in C} x(c)\Phi_c(y(c^{-1})))b \; .$$

Hence our problem is to find $y(c)$'s satisfying the simultaneous set of equations:

$$iy(b) + \sum_{c \in C} x(c)\Phi_c(y(c^{-1}b)) = \begin{cases} e & \text{for} \quad b = e \\ 0 & \text{for} \quad b \neq e \; . \end{cases}$$

Write the elements for the finite group C as $\{e = c_0, c_1, \cdots, c_n\}$ so that we have:

$$iy(c_k) + \sum_{i=0}^{n} x(c_i)\Phi_{c_i}(y(c_i^{-1}c_k)) = e\delta_{k0} \, ,$$

or

$$(1) \qquad iy(c_k) + \sum_{i=0}^{n} x(c_k c_r^{-1})\Phi_{c_k c_r^{-1}}(y(c_r)) = e\delta_{k0} \; .$$

Since $\Phi_c(0) = 0$ and $\Phi_c(e) = e$ for any $c \in C$ the application of $\Phi_{c_k}^{-1}$ to the kth equation gives:

$$(2) \qquad i\Phi_{c_k^{-1}}(y(c_k)) + \sum_{r=0}^{n} \Phi_{c_r^{-1}}(x(c_k c_r^{-1}))\Phi_{c_r^{-1}}(y(c_r)) = e\delta_{k0}$$

for $k = 0, 1, \cdots, n$. The matrix of coefficients of these equations is:

$$\begin{bmatrix} ie + x(c_0) & x(c_1^{-1}) & \cdots & x(c_n^{-1}) \\ \Phi_{c_1^{-1}}(c_1 c_0^{-1})) & ie + \Phi_{c_1^{-1}}(x(c_1 c_1^{-1})) & \cdots & \Phi_{c_1^{-1}}(x(c_1 c_n^{-1})) \\ & \cdots & \cdots & \\ \Phi_{c_n^{-1}}(x(c_n c_0^{-1})) & \Phi_{c_n^{-1}}(x(c_n c_1^{-1})) & \cdots & ie + \Phi_{c_n^{-1}}(x(c_n c_n^{-1})) \end{bmatrix} .$$

Now the elements of this matrix are elements of the commutative algebra $l^1(K)$, and hence the determinant Δ of this matrix is a well defined element of $l^1(K)$. Moreover the usual "Cramer's rule" formula will furnish a solution of the set of equations (2) if it can be shown that Δ is a nonsingular element of $l^1(K)$. Let α_{ij} denote the element in the ith row and jth column of the above matrix so that $\Delta = \det(\alpha_{ij})$. Now Δ is nonsingular if and only if $h(\Delta)$ is non-zero for any $h \in \Phi(l^1(K))$. Since h is a homomorphism $h(\Delta) = \det(h(\alpha_{ij}))$, and $h(\Delta)$ will be non-zero if it can be shown that $\overline{h(\alpha_{ij})} = h(\alpha_{ji})$ for $i \neq j$ and $h(\alpha_{jj}) = i + \beta_{jj}$ where $\beta_{jj} = \overline{\beta_{jj}}$. Indeed in this case the matrix $(h(\alpha_{ij}))$ is the matrix corresponding to an operator on a finite dimensional Hilbert space of the form $iI + H$ where H is an hermitian operator. Hence the operator $iI + H$ is nonsingular so that the determinant of any matrix representations of it must be non-zero.

It remains to verify the above equations. Since $\Phi_c(x^*) = (\Phi_c(x))^*$ for any $c \in C$ we have

$$(\Phi_{c_k^{-1}}(x(c_0)))^* = \Phi_{c_k^{-1}}(x(c_0)^*) = \Phi_{c_k^{-1}}(x(c_0)) ,$$
$$(\Phi_{c_k^{-1}}(x(c_k)))^* = \Phi_{c_k^{-1}}(x(c_k)^*) = x(c_k^{-1}) ,$$

and

$$\begin{aligned}(\Phi_{c_i^{-1}}(x(c_i c_k^{-1})))^* &= \Phi_{c_i^{-1}}(x(c_i c_k^{-1})^*) = \Phi_{c_k^{-1}}(\Phi_{c_k c_i^{-1}}(x(c_i c_k^{-1})^*)) \\ &= \Phi_{c_k^{-1}}(\Phi_{c_k c_i^{-1}}(x((c_k c_i^{-1})^{-1})^*)) = \Phi_{c_k^{-1}}(x(c_k c_i^{-1})) .\end{aligned}$$

What we have shown is that the elements α_{ii} are of the form $\alpha_{ii} = ie + \delta_{ii}$ where $\delta_{ii}^* = \delta_{ii}$ and $(\alpha_{ij})^* = \alpha_{ji}$ for $i \neq j$. Finally any $h \in \Phi(l^2(K))$ satisfies $h(a^*) = \overline{h(a)}$ so that the matrix $(h(\alpha_{ij}))$ is in the desired form, and $h(\Delta)$ is non-zero. Hence we have a solution to the equations (2), and the application of Φ_{c_k} to the kth equation of (2) gives the solution to the equations (1) and therefore the desired right inverse y. A left inverse for $ie + x$ can be constructed in a similar way.

REMARK 3.8. We do not know in general if the group algebra of a semi-direct product of a finite group and a discrete abelian group is symmetric with respect to the natural involution, in spite of the fact that the above theorem shows the hermitianess of the involution. The following theorem describes a special case where this is true.

THEOREM 3.9. *If $C = \{e, a : a^2 = e\}$ and K is abelian, then any semi-direct product $G = C \times_\varphi K$ has a symmetric group algebra.*

Proof. $x \in l^1(G)$ has the form $x = x_1 + ax_2$, $x_1, x_2 \in l^1(K)$, and hence $e + x^*x = e + x_1^*x_1 + x_2^*x_2 + x_2^*ax_1 + x_1^*ax_2$. Let $\Phi_a = \Phi$, $z_1 = e + x_1^*x_1 + x_2^*x_2$, and $z_2 = \Phi(x_2^*)x_1 + \Phi(x_1^*)x_2$. Then $e + x^*x = z_1 + az_2$. Thus $e = (z_1 + az_2)(y_1 + ay_2)$ if and only if $z_1y_1 + \Phi(z_2)y_2 = e$ and $z_2y_1 + \Phi(z_1)y_2 = 0$ so that $y_2 = -\Phi(z_1^{-1})z_2y_1$ and $(z_1 - \Phi(z_2)\Phi(z_1^{-1})z_2)y_1 = e$. Assume $z_1 - \Phi(z_2)\Phi(z_1^{-1})z_2$ and hence $\Phi(z_1)z_1 - \Phi(z_2)z_2$ is singular. Then there is a homomorphism h such that $h(\Phi(z_1))h(z_1) = h(\Phi(z_2))h(z_2)$. But

$$\begin{aligned}
h(\Phi(z_2))h(z_2) &= (h(x_2^*)h(\Phi(x_1)) + h(x_1^*)h(\Phi(x_2)))(h(\Phi(x_2^*))h(x_1) + h(\Phi(x_1^*))h(x_2)) \\
&= \overline{h(x_2)}h(\Phi(x_1))\overline{h(\Phi(x_2))}h(x_1) + \overline{h(x_1)}h(\Phi(x_2))\overline{h(\Phi(x_2))}h(x_1) \\
&\quad + \overline{h(x_2)}h(\Phi(x_1))\overline{h(\Phi(x_1))}h(x_2) + \overline{h(x_1)}h(\Phi(x_2))\overline{h(\Phi(x_1))}h(x_2) \\
&= |h(x_1)h(\Phi(x_2))|^2 + |h(x_2)h(\Phi(x_1))|^2 + 2h(x_1)h(\Phi(x_1))\overline{h(x_2)h(\Phi(x_2))} \\
&\leqq |h(x_1)h(\Phi(x_2))|^2 + |h(x_2)h(\Phi(x_1))|^2 + |h(x_1)h(\Phi(x_2))|^2 \\
&\quad + |h(x_2)h(\Phi(x_1))|^2 = 2|h(x_1)h(\Phi(x_2))|^2 + 2|h(x_2)h(\Phi(x_1))|^2 \\
&\leqq |h(x_1)|^2 + |h(\Phi(x_2)|^2 + |h(x_2)|^2 + |h(\Phi(x_1)|^2 \\
&< (1 + |h(x_1)|^2 + |h(x_2)|^2)(1 + |h(\Phi(x_1))|^2 + |h(\Phi(x_2))|^2) \\
&= h(z_1)h(\Phi(z_1))
\end{aligned}$$

and we have obtained a contradiction.

It is known (see Rickart [7]) that the symmetry and hermitianess properties are preserved in passing from a Banach algebra to a norm closed * closed subalgebra. In the case of the group algebra of a discrete group, and the group algebra of a subgroup, an elementary proof of a more general result can be given. Specifically:

THEOREM 3.10. *Let G be a discrete group, and H a subgroup of G. Then the natural imbedding of H in G induces an isometric * isomorphic imbedding of $l^1(H)$ into $l^1(G)$. With respect to this imbedding, for $x \in l^1(H)$*

$$Sp_{l^1(H)}(x) = Sp_{l^1(G)}(x) .$$

In particular, if $l^1(G)$ is symmetric (has an hermitian involution), *then $l^1(H)$ is symmetric* (has an hermitian involution).

Proof. The only non-trivial part of the proof consists in showing that if $x \in l^1(H)$, and x is regular in $l^1(G)$, then x is already regular in $l^1(H)$.

Let $\{Hg_a : 0, a \in A, g_0 = e\}$ be a left coset decomposition of G with respect to H. Write $x = \sum_{k \in H} x_0(k)kg_0$, and its inverse $y \in l^1(G)$ as $y = \sum_{a \in A} (\sum_{h \in H} y_a(h)hg_a)$. Then $xy = e$ means

$$e = \sum_{a \in A} \Big(\sum_{k,h \in H} x_0(k)y_a(h)khg_a \Big) = \sum_{a \in A} \Big(\sum_{e \in H} \Big(\sum_{h \in H} x_0(lh^{-1})y_a(h)lg_a \Big) \Big) .$$

For any fixed $a \neq 0$ and $l \in H$ we then have $\sum_{h \in H} x_0(1h^{-1})y_a(h) = 0$. Define $y'_a \in l^1(H)$ by $y'_a = \sum_{h \in H} y_a(h)h$. Then the above equation gives that $xy'_a = 0$, so that x is a divisor of zero in $l^1(H)$. But then x is a divisor of zero in $l^1(G)$. Since x is assumed to be regular in $l^1(G)$ we must have that $y'_a = 0$, and hence that $y_a(h) = 0$, for all $h \in H$. As this is true for all $a \neq 0$ we have that the inverse $y = \sum_{h \in H} y_0(h)h$ is an element of $l^1(H)$.

REMARK 3.11. It is easily seen that if the group algebra of G is symmetric or hermitian, then so is the group algebra of any quotient group. However we do not know if the symmetry or hermitianess of the group algebras of both H and G/H imply that of G.

4. Group algebras where the natural involution is not hermitian. In this section G will be a countable discrete group. The notations following Theorem 3.5 will be used. The conjugate space of $l^1(G)$ will be denoted by $\mathscr{C}(G)$ (all bounded sequences of complex numbers). Let L_x be the left multiplication on $l^1(G)$ defined by x, i.e. $L_x y = xy$, the multiplication being convolution. For a given ordering $\{g_1, g_2, \cdots\}$ of all the elements of G, the matrix of L_x, $\operatorname{mat}(L_x)$, is then defined as (a_{ij}) where $x = \sum_{k=1}^{\infty} x(g_k)g_k$ and $a_{ij} = x(g_j g_i^{-1})$. Since

$$L_x g_n = \sum_{k=1}^{\infty} x(g_k)g_k g_n = \sum_{k=1}^{\infty} x(g_k g_n^{-1})g_k = \sum_{k=1}^{\infty} a_{nk} g_k ,$$

we may speak of the nth row of $\operatorname{mat}(L_x)$ as the image of g_n under L_x. An element $\varphi = (\varphi_1, \varphi_2, \cdots) \in \mathscr{C}(G)$ will be said to be *orthogonal* to a row $R_i = (a_{i1}, a_{i2}, \cdots)$ of $\operatorname{mat}(L_x)$ if $\sum_{j=1}^{\infty} a_{ij}\varphi_j = 0$.

LEMMA 4.1. (i) *If* $\operatorname{mat}(L_x) = (a_{ij})$, *then* $\operatorname{mat}(L_{x^*}) = (b_{ij})$ *where* $b_{ij} = \overline{a_{ji}}$. *In particular if* $x = x^*$, *then* $a_{ij} = \overline{a_{ji}}$.

(ii) *If there is a non-zero element* $\varphi \in \mathscr{C}(G)$ *orthogonal to all the rows of* $\operatorname{mat}(L_x)$, *then* L_x *is a singular operator.*

Proof. (i) Since $x^* = \sum_{g \in G} y(g)g$ where $y(g) = \overline{x(g^{-1})}$ we have that $\operatorname{mat}(L_{x^*}) = (b_{ij})$ with $b_{ij} = y(g_j g_i^{-1}) = \overline{x(g_i g_j^{-1})} = \overline{a_{ji}}$.

(ii) If such a φ exists, then $\varphi(L_x g_n) = 0$ for all n. Hence all finite linear combinations of the $L_x g_n$'s are in the nullspace of φ. From the continuity of φ and L_x, and the fact that linear combinations of the g_n's are dense in $l^1(G)$, it follows that L_x maps $l^1(G)$ into the nullspace of φ.

Since φ is non-zero, L_x is singular because it is not onto. Note that from Theorem B(b) we have that x is singular in $l^1(G)$.

LEMMA 4.2. *Let $a_1, a_2, \cdots, a_n$, $n \geqq 3$ be complex numbers of absolute value one, and $x_1, x_2, \cdots, x_r$ complex numbers of absolute value one or zero. Then for $2r \leqq n$ there are complex numbers $x_{r+1}, x_{r+2}, \cdots, x_n$ of absolute value one or zero, not all of which are zero, such that $a_1x_1 + a_2x_2 + \cdots + a_rx_r = a_{r+1}x_{r+1} + a_{r+2}x_{r+2} + \cdots + a_nx_n$. Moreover if $n \geqq 4$ and $2r < n$, there are at least two linearly independent solutions.*

Proof. Suppose first that not all the x_i's are zero. Let $x_{r+k} = a_kx_k/a_{r+k}$ for $1 \leqq k \leqq r$ and $x_{r+k} = 0$ for $k > r$. This gives a non-zero solution. If $2r < n$ then $x_n = 0$ and some $x_{r+i_0} \neq 0$. Let $x'_n = x_{r+i_0}a_{r+i_0}/a_n$, $x'_{i_0+r} = 0$, and $x'_j = x_j$ for $j \neq n, j \neq r + i_0$. Then the primed sequence is also a solution and is clearly not a scalar multiple of the unprimed sequence. (In the above case $n \geqq 3$ is all that is required).

Now assume that $x_1 = x_2 = \cdots = x_r = 0$. Since $n \geqq 3$ and $2r \leqq n$, a_{r+1} and a_{r+2} exist. Letting $x_{r+1} = 1$, $x_{r+2} = -a_{r+1}/a_{r+2}$, and the remaining x_i's zero we have a solution. Finally if $2r < n$ and $n \geqq 4$, a_{r+1}, a_{r+2}, and a_{r+3} exist. In this case pick $x'_{r+1} = 0$, $x'_{r+2} = 1$, $x'_{r+3} = -a_{r+2}/a_{r+3}$, and the remaining x'_i's zero. Again the primed sequence is a solution and clearly not a scalar multiple of the unprimed sequence.

Let $\{g_1, g_2, \cdots\}$ be an ordering $\mathscr{O}$ of all the elements of G. For a subset A of G, let $|A|$ denote the number of elements of A, and $[A]$ the subgroup generated by A. The following definition is pertinent to both the symmetry of $l^1(G)$, and the existence of an invariant mean on G.

DEFINITION 4.3. A finite set S of G will be said to be *singular with respect to the ordering* $\mathscr{O}$ if:

(i) $|S| \geqq 3$;

(ii) $[S]| = \infty$;

(iii) There is an integer n_0 such that

$$2\,|\,Sg_n \cap (Sg_1 \cup Sg_2 \cup \cdots \cup Sg_{n-1})\,| \leqq |\,S\,| \quad \text{for all} \quad n > n_0\,.$$

In the following theorem an element $\varphi = (\varphi_1, \varphi_2, \cdots) \in \mathscr{C}(G)$ is going to be constructed with respect to a given matrix. We will start out with the sequence consisting of all zeros, and then begin replacing the zeros by other entries. At any given stage in the construction, the kth column of the matrix will be termed an *old* column if φ_k has already replaced a zero (the φ_k may itself be zero), and a *new* column otherwise.

THEOREM 4.4. *Let S be a singular set in G with respect to the* *ordering* $\{g_1, g_2, \cdots\}$. *Then the element* $x = \sum_{s_i \in S} \alpha_i s_i$, $|\alpha_i| = 1$, *is singular in* $l^1(G)$.

Proof. By Theorem B (b) it is enough to show that L_x is a singular operator, and by Lemma 4.1 (ii) it suffices to find an element $\varphi \in \mathscr{C}(G)$ orthogonal to all the rows of $\mathrm{mat}\,(L_x)$.

Take $\mathrm{mat}\,(L_x)$. In the columns that contain a non-zero entry from one of the first n_0 rows of $\mathrm{mat}\,(L_x)$, replace the zeros in φ by zeros. In other words, these columns will now be called old columns. We have that $2\,|Sg_n \cap (Sg_1 \cup Sg_2 \cup \cdots \cup Sg_{n-1})| \leqq r$ for $n > n_0$ where $|S| = r$. The $(n_0 + 1)$ row of $\mathrm{mat}\,(L_x)$ contains non-zero entries $a_{c_1}, a_{c_2}, \cdots, a_{c_r}$ in columns $c_1, c_2, \cdots, c_r$ respectively, corresponding to the elements in the set Sg_{n_0+1}. Since S is singular, at least half of these columns are new. Denote the new columns by $c'_1, c'_2, \cdots, c'_s$ where $2s \geqq r$ and select, using Lemma 4.2, $\varphi_{c'_1}, \varphi_{c'_2}, \cdots, \varphi_{c'_s}$ of absolute value one or zero (but not all zero) such that $\sum_{i=1}^{s} \varphi_{c'_i} a_{c'_i} = 0$. At this stage the $\varphi \in \mathscr{C}(G)$ is orthogonal to the first $n_0 + 1$ rows of $\mathrm{mat}\,(L_x)$. Now take the $(n_0 + 2)$ row of $\mathrm{mat}\,(L_x)$. The non-zero entries $a_{p_1}, a_{p_2}, \cdots, a_{p_r}$ now occurs in columns $p_1, p_2, \cdots p_r$ respectively, and since S is singular at least half of the p_i's are new. Denote the old columns by $p'_1, p'_2, \cdots, p'_t$, and the new ones by $p''_1, p''_2, \cdots, p''_{r-t}$. Again by Lemma 4.2 there are complex numbers $\varphi_{p''_1}, \varphi_{p''_2}, \cdots, \varphi_{p''_{r-t}}$ of absolute value one or zero such that

$$\sum_{i=1}^{t} \varphi_{p'_i} a_{p'_i} = \sum_{j=1}^{r-t} \varphi_{p''_j} a_{p''_j} \,.$$

Replacing the zeros by these new φ_i's then gives an element of $\mathscr{C}(G)$ orthogonal to the first $n_0 + 2$ rows of $\mathrm{mat}\,(L_x)$.

The proof is completed by induction. For any $m \geqq n_0 + 2$ assume that scalars of absolute value one or zero have been selected in columns where a non-zero entry occurs in one of the first m rows of $\mathrm{mat}\,(L_x)$, and that the sequence constructed is orthogonal to these rows. By again using the definition of singularity and Lemma 4.2, new φ_i's of absolute value one or zero, in new columns corresponding to the non-zero entries of the $(m + 1)$ rows of $\mathrm{mat}\,(L_x)$ can be constructed so that the resulting sequence is orthogonal to the first $m + 1$ rows of $\mathrm{mat}\,(L_x)$.

COROLLARY 4.5. *If* $|S| \geqq 4$ *and* $2\,|Sg_n \cap (Sg_1 \cup Sg_2 \cup \cdots \cup Sg_{n-1})| < |S|$ *for* $n > n_0$, *the range of* L_x *is not of finite deficiency.*

Proof. The second part of Lemma 4.2 assures us that at each stage in the above construction, starting with the $(n_0 + 1)$ row, there are two linearly independent sets of new φ_i's to choose from. Therefore we can construct infinitely many linearly independent elements in $\mathscr{C}(G)$ orthogonal to all the rows of $\mathrm{mat}\,(L_x)$.

COROLLARY 4.6. *Suppose the singular set satisfies* $e \in S = S^{-1}$. *Let the coefficient of* e *be* i, *and the other coefficients be one. Then the hermitian element* $\sum_{s_i \in S - \{e\}} s_i$ *contains the element* $-i$ *in its spectrum.*

The following theorem gives examples of groups that contain singular sets S satisfying $e \in S = S^{-1}$. A definition is needed first. Let $n \geqq 2$, and let $F^{(n)}$ be the free group on generators $a_1, a_2, \cdots, a_n$. For any $f \in F^{(n)}$, the *length* of f is:

$$\min \left\{ \sum_{i=1}^{s} | n(i) | : f = a_{i(1)}^{n(1)} a_{i(2)}^{n(2)} \cdots a_{i(s)}^{n(s)} \right\} .$$

THEOREM 4.7. (i) *Let* $F^{(n)}$ *be the free group on generators* $a_1, a_2, \cdots, a_n$; $n \geqq 2$. *Then there is an ordering* $\mathcal{O}$ *of* $F^{(n)}$ *such that the set* $S = \{e = a_0, a_1, a_2, \cdots, a_n, a_1^{-1}, a_2^{-1}, \cdots, a_n^{-1}\}$ *is singular with respect to it.*

(ii) *Let* $G^{(n)}$ *be the free group on generators* $b_1, b_2, \cdots, b_n$; $n \geqq 3$, *each of order two. Then there is an ordering of* $G^{(n)}$ *such that the set* $S = \{e = b_0, b_1, b_2, \cdots, b_n\}$ *is singular with respect to it.*

Proof. (i) The ordering $\mathcal{O}$ is started with $g_i = a_i$, $i = 0, 1, 2, \cdots, n$; and $g_{n+j} = a_j^{-1}$, $j = 1, 2, \cdots, n$. Since the generators are free, each of the sets $Sg_1, Sg_2, \cdots, Sg_{2n}$ contains $2n - 1$ distinct elements of length 2. It is clear that no element of length 2 in Sg_i can equal an element of length 2 in Sg_j, $i \neq j$; and that included in the Sg_i's are all elements of length 2. Now successively adjoin to the set $\{g_0, g_1, \cdots, g_{2n}\}$ the elements of length 2 from $Sg_1, Sg_2, \cdots, Sg_n$ respectively. This gives $g_{2n+1}, g_{2n+2}, \cdots, g_{4n^2}$. Again since the generators are free, each of the sets $Sg_{2n+1}, Sg_{2n+2}, \cdots, Sg_{4n^2}$ contains $2n - 1$ distinct elements of length 3; no element of length 3 in Sg_i can equal an element of length 3 in Sg_j, $i \neq j$; and all the elements of length 3 are included in them. As before successively adjoin the elements of length 3 from $Sg_{2n+1}, Sg_{2n+2}, \cdots, Sg_{4n^2}$. The ordering $\mathcal{O}$, constructed in this manner by then adjoining elements of length 4.5 etc., satisfies the conditions of the theorem. Indeed, we have for any n, $| Sg_n \cap (Sg_0 \cup Sg_1 \cup \cdots \cup Sg_{n-1}) | = 2$, and since $n \geqq 2$, $2 \cdot 2 \leqq 2n + 1 = | S |$.

(ii) The proof is in the same spirit as that in (i). In this case start the ordering with S and successively adjoin the elements from

$$Sg_1 - (Sg_1 \cap S), Sg_2 - (Sg_2 \cap (Sg_1 \cup S)), \cdots, Sg_n - (Sg_n \cap (Sg_{n-1} \cup Sg_{n-2} \cup \cdots \cup S)) , \cdots$$

In this case $| Sg_n \cap (Sg_{n-1} \cup Sg_{n-2} \cup \cdots \cup S) | = 2$, and since $n \geqq 3$, $2 \cdot 2 \leqq n + 1$.

REMARK 4.8. It is not hard to see that for the case of $F^{(n)}$ an element of the form $\alpha e + a_1 + a_2 + \cdots + a_n + a_1^{-1} + a_2^{-1} + \cdots + a_n^{-1}$ where $|\alpha| < 2n - 2$ is singular in $l^1(F^{(n)})$. From this it follows that the hermitian element $x = a_1 + a_2 + \cdots + a_n + a_1^{-1} + a_2^{-1} + \cdots + a_n^{-1}$ contains in its spectrum the closed circle about the origin of radius $2n - 2$.

REMARK 4.9. For $G^{(2)}$ the theorem is false. One way to see this is to note that $G^{(2)}$ is the semi-direct product of the integers by a group of order two, where the automorphism sends an element to its inverse. Hence by Theorem 3.9, $l^1(G^{(2)})$ is symmetric with respect to the natural involution. Another way of seeing this will be given by Theorem 4.12 (see Remark 4.14).

REMARK 4.10. It is known that the group $F^{(n)}$, $n \geqq 2$ has a complete set of representations by finite groups, and it follows from this that $F^{(n)}$ can be algebraically imbedded in the complete direct sum of these finite groups. By Theorem 3.10 we then have that the natural involution is not hermitian in the group algebra of this complete direct sum. However, we do not know the answer to the involution question for the general case of the restricted direct sum (sequences reducing to the identity from some point on) of finite groups.

REMARK 4.11. Group algebras are A^* algebras in the sense introduced by Rickart [8]. Unfortunately, Hille and Phillips [5: pp 22] have defined an A^* algebra to be a Banach algebra with an hermitian involution. It follows from the above that these two definitions are not the same.

Perhaps the simplest example of an hermitian element with non-real spectrum can be found in the group algebra of the group $G = \{a, b : a^2 = e\}$. The element $x = a + b + b^{-1}$ is hermitian and with respect to an ordering of G constructed in the same fashion as above, the matrix of L_{ie+x} is:

$$\begin{bmatrix} i & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & i & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & i & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & i & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & i & 0 & 0 & 0 & 0 & 0 & 1 & 1 & \cdots \\ 0 & 1 & 0 & 0 & 0 & i & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & 0 & i & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & i & 0 & 0 & 0 & 0 & \cdots \\ & & & & & & \cdots & & & & & & \end{bmatrix}$$

The element $(i, 1, -1, -i, -i, \cdots) \in \mathscr{C}(G)$ will then be orthogonal to all the rows of this matrix, and hence $-i$ is in the spectrum of x.

Theorem 4.12 is quite special, however it does suffice to show that hermitian elements with finite support in $l^1(G^{(2)})$ have real spectrum.

Let $S = (s_{ij})$; $i, j = 1, 2, \cdots$ be any infinite matrix and define $S^{(n)} = (s'_{ij})$ by the equations $s'_{ij} = s_{ij}$ for $i, j \leqq n$ and $s'_{ij} = 0$ otherwise. $S^{(n)}$ will be called the *principal* $n \times n$ *section of* S. If there is an integer k so that $s_{mn} = 0$ whenever $|m - n| \geqq k$, S will be called a *corridor matrix of width* k.

THEOREM 4.12. *Let* $T = (a_{ij})$ *be an hermitian corridor matrix of width* k, *with* $\sup_{i,j}|a_{ij}| < \infty$. *Then for any real number* ρ, *the operator* $T + \rho iI$ *defined by this matrix maps* l^1 *onto a dense subset of* l^1.

Proof. Note first that since the norm of an operator on l^1 can be computed by taking the sup of the l^1 norms of the rows of its matrix with respect to the usual basis, T is a bounded operator (we will not distinguish between the matrix and the operator it represents) on l^1. Moreover since the matrix of T is hermitian, T can be extended to a bounded operator on l^2. Hence the spectrum of T as an operator on l^2 is real.

Assume there is a sequence $(\varphi_1, \varphi_2, \cdots)$ (not necessarily bounded), that is orthogonal to all the rows of $(T + \rho iI)$. Since $T + \rho iI$ is regular as an operator on l^2, it follows that $\sum_{i=1}^{\infty} |\varphi_i|^2 = \infty$. We are going to show that the sequence $(\varphi_1, \varphi_2, \cdots)$ is in fact unbounded.

Let $l^2_{(n)}$ denote n-dimensional Hilbert space and $\varphi^{(n)} = (\varphi_1, \varphi_2, \cdots, \varphi_n)$. Since T is hermitian, we have for any $y \in l^2_{(n)}$ that $\|(T + \rho iI)_1^{(n)} y\|_2 \geqq \rho \|y\|_2$, where $(T + \rho iI)_1^{(n)}$ denotes the $n \times n$ matrix in the upper left hand portion of the principal $n \times n$ section $(T + \rho iI)^{(n)}$. Let K be any large number, and pick $n_0 > K$ so that $\sum_{i=1}^{n_0} |\varphi_i|^2 > 4K^2M^2k^3\rho^{-2}$ where $M = \sup_{i,j} |t_{ij}|$ and $\operatorname{mat}(T + \rho iI) = (t_{ij})$. Let $(T + \rho iI)_1^{(n_0)} \varphi^{(n_0)} = (a_1, a_2, \cdots, a_{n_0})$, $a_k = \sum_{p=1}^{n_0} a_{kp}\varphi_p$. Since $\varphi = (\varphi_1, \varphi_2, \cdots)$ is orthogonal to all the rows of $T + \rho iI$ we have that $a_1 = a_2 = \cdots = a_{n_0 - k} = 0$. From the l^2 norm inequality above, we have that

$$\|(T + \rho iI)_1^{(n_0)} \varphi^{(n_0)}\|_2^2 \geqq \rho^2 \sum_{i=1}^{n_0} |\varphi_i|^2$$

or

$$\sum_{i=n_0-k+1}^{n_0} |a_i|^2 \geqq \rho^2(4K^2M^2k^3\rho^{-2}) = 4K^2M^2k^3 .$$

Hence some a_{i_0}, $n_0 - k < i_0 \leqq n_0$ is such that $|a_{i_0}|^2 > 4K^2M^2k^2$ or $|a_{i_0}| > 2KMk$. However $a_{i_0} = \sum_{p=1}^{n} t_{i_0 p}\varphi_p$, and since there are at most $2k$ non-zero terms, there is a p_0 with $|t_{i_0 p_0}\varphi_{p_0}| > KM$ and hence $|\varphi_{p_0}| > KM/|t_{i_0 p_0}| \geqq K$. In other words, the sequence $(\varphi_1, \varphi_2, \cdots)$ is unbounded, and it follows that the range of $T + \rho iI$ is dense in l^1.

For $x = \sum_{g \in G} x(g)g \in l^1(G)$, let G_0 denote the subgroup of G generated by $\{g \in G : x(g) \neq 0\}$. Since this set is countable, G_0 is countable. We have:

COROLLARY 4.13. *If $x = x^* \in l^1(G)$, and with respect to the basis in $l^1(G_0)$ defined by some ordering of G_0,* $\mathrm{mat}\,(L_x)$ *is a corridor matrix, then the spectrum of x is real.*

Proof. By Theorem 3.10 it sufficies to look at the spectrum of x as an element of $l^1(G_0)$. Now the theorem above gives that the ranges of L_{ie+x} and L_{-ie+x} are dense in $l^1(G_0)$. But these ranges are also ideals and since they are dense, they must be all of $l_1(G_0)$. This means that there are elements $y_1, y_2 \in l^1(G_0)$ such that $(ie + x)y_1 = e$ and $(-ie + x)y_2 = e$. Applying the involution to the latter equality gives $y_2^*(ie + x) = e$. Hence $ie + x$ has both a right and left inverse, and is hence regular.

REMARK 4.14. Take the ordering $(e, a, b, ab, ba, aba, \cdots)$ in the group $G^{(2)} = \{a, b : a^2 = b^2 = e\}$. Then it is easily seen that $\mathrm{mat}\,(L_x)$ is a corridor matrix whenever $x \in l^1(G^{(2)})$ has finite support. Hence Theorem 4.7 does not hold for $G^{(2)}$.

5. The involution and invariant means.

The main results in this section are Theorem 5.6 and Theorem 5.8. The first theorem gives us some information concerning the involution when the group has an invariant mean, and in the second theorem it is shown that a group containing a singular set cannot have an invariant mean.

A continuous linear functional λ on $\mathscr{C}(G)$ is said to be an *invariant mean*, if it satisfies:

(i) $\lambda(\varphi) \geqq 0, \quad \varphi \geqq 0, \quad \varphi \in \mathscr{C}(G)$;

(ii) $\lambda(\varphi_x) = \lambda(\varphi^x) = \lambda(\varphi)$ where $\varphi_x(y) = \varphi(x^{-1}y)$, and $\varphi^x(y) = \varphi(yx)$;

(iii) $\lambda(I) = 1$ where I is the function identically 1 on G.

Whenever the notation $\lambda(A)$, for A a subset of G, is used, it will mean the number $\lambda(\chi_A)$ where χ_A is the characteristic function of A.

For $\varphi, \psi \in \mathscr{C}(G)$ define a pseudo "inner product" $(\varphi, \psi) = \lambda(\varphi\bar{\psi})$. A few simple properties of this inner product are given in:

LEMMA 5.1. (i) $(\varphi, \psi_1 + \psi_2) = (\varphi, \psi_1) + (\varphi, \psi_2)$;

(ii) $(\varphi, \psi) = \overline{(\psi, \varphi)}$;

(iii) $(\alpha\varphi, \psi) = \alpha(\varphi, \psi)$;

(iv) $(\varphi, \varphi) \geqq 0$;

(v) $|(\varphi, \psi)| \leqq (\varphi, \varphi)^{1/2}(\psi, \psi)^{1/2}$;

where $\varphi, \psi, \psi_1, \psi_2 \in \mathscr{C}(G)$, and α is a complex number.

Proof. (v) will be proved, the other statements following immediately from the definitions.

$$0 \leqq (\varphi - \alpha\psi, \varphi - \alpha\psi) = (\varphi, \varphi) - \alpha(\psi, \varphi) - \bar{\alpha}(\varphi, \psi) + |\alpha|^2(\psi, \psi) .$$

If $(\varphi, \psi) = 0$, (v) is trivial, so assume that $(\varphi, \psi) \neq 0$, and let $\alpha = (\varphi, \varphi)/(\psi, \varphi)$; (v) then follows by direct calculation.

Let $\mathscr{K} = \{\varphi \in \mathscr{C}(G) : \lambda(|\varphi|) = 0\}$, and $\mathscr{L} = \{\varphi \in \mathscr{C}(G) : \lambda(|\varphi|^2) = 0\}$. We have:

LEMMA 5.2. *$\mathscr{K}$ is equal to $\mathscr{L}$, and is a closed subspace of $\mathscr{C}(G)$.*

Proof. By letting $\psi = I$ and replacing φ by its absolute value in (v) above, we have that, $|(|\varphi|, I)|^2 \leqq (|\varphi|, |\varphi|)(I, I)$ or $(\lambda(|\varphi|))^2 \leqq \lambda(|\varphi|^2)\lambda(I^2) = \lambda(|\varphi|^2)$. Hence $\lambda(|\varphi|^2) = 0$ implies $\lambda(|\varphi|) = 0$, and thus $\mathscr{L} \subset \mathscr{K}$.

Conversely if $\varphi \in \mathscr{K}$, then $|\varphi|^2 \leqq K|\varphi|$ where K is a bound for $|\varphi|$, and it follows that $\mathscr{K} \subset \mathscr{L}$.

Since $|\alpha\varphi + \beta\psi| \leqq |\alpha||\varphi| + |\beta||\psi|$, $\mathscr{K}$ is a subspace of $\mathscr{C}(G)$. Finally for $\varphi_n \in \mathscr{K}$ and $||\varphi_n - \varphi||_\infty \to 0$, it follows from the continuity of λ that $\lambda(|\varphi|) = 0$, and hence $\mathscr{K}$ is closed.

Let $\mathscr{C} - \mathscr{K}$ denote the space of cosets of $\mathscr{C}(G)$ with respect to $\mathscr{K}$. For $\dot{\varphi} \in \mathscr{C} - \mathscr{K}$ let $||\dot{\varphi}||_1 = \lambda(|\varphi|)$: $||\dot{\varphi}||_2 = (\lambda(|\varphi|^2))^{1/2}$, $\varphi \in \dot{\varphi}$; and $(\dot{\varphi}, \dot{\psi}) = \lambda(\varphi\bar{\psi})$, $\varphi \in \dot{\varphi}$, $\psi \in \dot{\psi}$. Then:

LEMMA 5.3. *For $\dot{\varphi}, \dot{\psi} \in \mathscr{C} - \mathscr{K}$.*

(i) *$||\dot{\varphi}||_1$ is well defined and a norm on $\mathscr{C} - \mathscr{K}$;*

(ii) *$(\dot{\varphi}, \dot{\psi})$ is well defined and makes $\mathscr{C} - \mathscr{K}$ into a pre-Hilbert space;*

(iii) $||\dot{\varphi}||_1 \leqq ||\dot{\psi}||_2$.

Proof. (i) Let $\varphi_1, \varphi_2 \in \dot{\varphi}$ so that $\varphi_1 = \varphi_2 + k$ where $k \in \mathscr{K}$. Then $|\varphi_1(g)| = |\varphi_2(g) + k(g)| \leqq |\varphi_2(g)| + |k(g)|$ so that $|\varphi_1| \leqq |\varphi_2| + |k|$. Hence $\lambda(|\varphi_1|) \leqq \lambda(|\varphi_2| + |k|) \leqq \lambda(|\varphi_2|) + \lambda(k|) = \lambda(|\varphi_2|)$. Now by reversing the roles of φ_1 and φ_2, it follows that $||\dot{\varphi}||_1$ is well defined. Also $||\dot{\varphi} + \dot{\psi}||_1 = \lambda(|\varphi + \psi|) \leqq \lambda(|\varphi| + |\psi|) = \lambda(|\varphi|) + \lambda(|\psi|) = ||\dot{\varphi}||_1 + ||\dot{\psi}||_1$, and $||\alpha\dot{\varphi}||_1 = ||(\dot{\alpha\varphi})||_1 = \lambda(|\alpha\varphi|) = |\alpha|\lambda(|\varphi|) = |\alpha|\,||\dot{\varphi}||$ for $\varphi \in \dot{\varphi}$, $\psi \in \dot{\psi}$, and α complex. Finally $||\dot{\varphi}||_1 = 0$ implies that $\lambda(|\varphi|) = 0$, and hence that $\dot{\varphi} = 0$. Thus $||\dot{\varphi}||_1$ is a norm on $\mathscr{C} - \mathscr{K}$.

(ii) If $\varphi_1, \varphi_2 \in \dot{\varphi}$ and $\psi_1, \psi_2 \in \dot{\psi}$, then $\varphi_1 = \varphi_2 + k$, $\psi_1 = \psi_2 + l$ where $k, l \in \mathscr{K}$. Then $(\varphi_1, \psi_1) = (\varphi_2 + k, \psi_2 + 1) = (\varphi_2, \psi_2) + (\varphi_2, 1) + (k, \psi_2) + (k, l)$. But $|(\varphi_2, l)|^2 \leqq (\varphi_2, \varphi_2)(l, l) = 0$, $|(k, \psi_2)|^2 \leqq (k, k)(\psi_2, \psi_2) = 0$, and $|(k, l)|^2 \leqq (k, k)(l, l) = 0$, so that $(\dot{\varphi}, \dot{\psi})$ is well defined. If $(\dot{\varphi}, \dot{\varphi}) = 0$, then $\lambda(|\varphi|^2) = 0$ for $\varphi \in \dot{\varphi}$, and by Lemma 5.2, $\varphi \in \mathscr{K}$ or $\dot{\varphi} = 0$. Hence with respect to $(\dot{\varphi}, \dot{\psi})$, $\mathscr{C} - \mathscr{K}$ becomes a pre-Hilbert space.

(iii) $\|\dot\varphi\|_1^2 = (\lambda(|\varphi|))^2 = |(|\varphi|, I)|^2 \leqq (|\varphi|, |\varphi|)(I, I) = \lambda(|\varphi|^2) = \|\dot\varphi\|_2^2$.

$L^1(G, \lambda)$ will denote the completion of $\mathscr{C} - \mathscr{K}$ with respect to $\|\dot\varphi\|_1$; and $L^2(G, \lambda)$ the completion with respect to $\|\dot\varphi\|_2$.

LEMMA 5.4.* *Let $g_1, g_2, \cdots, g_n$ be distinct elements of G. Then there is a subset A of G satisfying $\lambda(A) > 0$, and $Ag_i \cap Ag_j = \phi$, $i \neq j$.*

Proof. Let $\mathscr{A} = \{B \subset G : Bg_i \cap Bg_0 = \phi,\ i = j\}$. $\mathscr{A}$ is then non-empty since $\{e\} \in \mathscr{A}$, and is partially ordered by inclusion. An immediate application of Zorn's lemma gives a maximal element A. Let $C = Ag_1 \cup Ag_2 \cup \cdots \cup Ag_n \cup (\bigcup_{i\neq j} Ag_i g_j^{-1})$. It will be shown that $C = G$. Indeed if $h \in G - C$, let $A' = A \cup \{h\}$. Since A is a maximal element of $\mathscr{A}$, there are indices i_0 and j_0 such that $k \in A'g_{i_0} \cap A'g_{j_0}$ and $k \notin Ag_{i_0} \cap Ag_{j_0}$. Therefore either

(a) $k = hg_{i_0}$, $k \in Ag_{j_0}$;

(b) $k \in Ag_{i_0}$, $k = hg_{j_0}$; or

(c) $k = hg_{i_0}$, $k = hg_{j_0}$.

But (c) implies that $g_{i_0} = g_{j_0}$, a contradiction. (a) implies that $k = hg_{i_0}$, $k = ag_{j_0}$ where $a \in A$, and hence $hg_{i_0} = ag_{j_0}$ or $h = ag_{j_0}g_{i_0}^{-1}$ giving $h \in Ag_{j_0}g_{i_0}^{-1}$ which is also a contradiction. The proof that (b) is impossible, is similar to (a). Hence $C = G$, and $\lambda(A) > 0$, since G is then the finite union of sets, each of measure $\lambda(A)$.

Corresponding to an $x \in l^1(G)$, we are now going to define operators on $L^1(G, \lambda)$ and $L^2(G, \lambda)$.

For $\dot\varphi \in \mathscr{C} - \mathscr{K}$ and $g \in G$, $\dot\varphi_g$ will mean the coset in $\mathscr{C} - \mathscr{K}$ containing φ_g. This is well defined since $\varphi, \psi \in \dot\varphi$ imply $\varphi - \psi = k \in \mathscr{K}$. Since $\varphi_g - \psi_g = k_g$ is also in $\mathscr{K}$ it follows that $\dot\varphi_g = \dot\psi_g$. For $x = \sum_{g\in G} x(g)g \in l^1(G)$ define $T_x\dot\varphi = \sum_{g\in G} x(g)\dot\varphi_g$ for $\dot\varphi \in \mathscr{C} - \mathscr{K}$. For $\varphi \in \dot\varphi$ we have,

$$\| T_x\dot\varphi \|_1 = \lambda\Big(\Big| \sum_{g\in G} x(g)\varphi_g \Big|\Big) \leqq \lambda\Big(\sum_{g\in G} |x(g)|\,|\varphi_g|\Big)$$
$$= \lambda\Big(\sum_{g\in G_1} |x(g)|\,|\varphi_g| + \sum_{g\in G-G_1} |x(g)|\,|\varphi_g|\Big)$$

where G_1 is some subset of G satisfying $\sum_{g\in G-G_1} |x(g)| < \varepsilon \|\varphi\|_\infty$. Hence

$$\| T_x\dot\varphi \|_1 \leqq \lambda\Big(\sum_{g\in G_1} |x(g)|\,|\varphi_g|\Big) + \lambda\Big(\sum_{g\in G-G_1} |x(g)|\,|\varphi_g|\Big)$$
$$\leqq \lambda\Big(\sum_{g\in G_1} |x(g)|\,|\varphi_g|\Big) + \lambda\Big(\|\varphi_g\|_\infty \sum_{g\in G-G_1} |x(g)|\Big)$$
$$\leqq \lambda\Big(\sum_{g\in G_1} |x(g)|\,|\varphi_g| + (\varepsilon\|\varphi\|_\infty)\,\|\varphi_g\|_\infty$$
$$\leqq \sum_{g\in G_1} |x(g)|\,\lambda(|\varphi_g|) + \varepsilon \leqq \|\dot\varphi_g\|_1 \|x\| + \varepsilon\,.$$

* The author is thankful to Professor H. A. Dye who suggested this lemma.

Since ε was arbitrary we have that $\| T_x\dot{\varphi} \|_1 \leqq \| x \| \, \| \dot{\varphi}_g \|_1$.

Now let G_0 be the countable subgroup generated by $\{g : x(g) \neq 0\}$, and let $\{g_1, g_2, \cdots\}$ be an ordering of G_0. Let $x^{(n)} = \sum_{i=1}^n x(g_i)g_i$. By Lemma 5.4 there is a subset A of G with $\lambda(A) = d > 0$, and $Ag_i \cap Ag_j = \phi$ for $i \neq j$ and $1 \leqq i, j \leqq n$. Then

$$\begin{aligned} \| T_x(n)\dot{\chi}_A \|_1 &= \lambda\Big(\Big|\sum_{i=1}^n x(g_i)(\chi_A)_{g_i}\Big|\Big) = \lambda\Big(\sum_{i=1}^n | x(g_i) | \, | (\chi_A)_{g_i} |\Big) \\ &= \sum_{i=1}^n | x(g_i) | \, d = d \sum_{i=1}^n | x(g_i) | = \| \dot{\chi}_A \|_1 \, \| x^{(n)} \| \, . \end{aligned}$$

Since $\| x^{(n)} - x \| \to 0$ as $n \to \infty$ we have that $\| T_x \|_1 = \| x \|$. Finally since T_x is bounded on $\mathscr{C} - \mathscr{K}$ with respect to $\| \dot{\varphi} \|_1$, it can be extended to the completion $L^1(G, \lambda)$ without increasing its norm. The extension will be denoted by $T_x^{(1)}$.

The operator T_x on $\mathscr{C} - \mathscr{K}$ will now be extended to $L^2(G, \lambda)$. For $\dot{\varphi} \in \mathscr{C} - \mathscr{K}$, and $\varphi \in \dot{\varphi}$ we have

$$\| T_x\dot{\varphi} \|_2^2 = \lambda\Big(\Big| \sum_{g \in G} x(g)\varphi_g \Big|^2\Big) \leqq \lambda\Big(\Big(\sum_{g \in G} | x(g) | \, | \varphi_g |\Big)^2\Big) \, .$$

But $\{| x(g) |^{1/2} : g \in G\} \in l^2(G)$, and so for any $h \in G$ the sequence $\{| x(g) |^{1/2} | \varphi_g | (h) : g \in G\} \in l^2(G)$. Now

$$\lambda\Big(\Big(\sum_{g \in G} | x(g) |^2 | \varphi_g |\Big)^2\Big) = \lambda\Big(\Big(\sum_{g \in G} | x(g) |^{1/2} | x(g)^{1/2} | \varphi_g |\Big)^2\Big) \, ,$$

and

$$\Big(\sum_{g \in G} | x(g) |^{1/2} | x(g) |^{1/2} | \varphi_g | (h)\Big)^2 \leqq \Big(\sum_{g \in G} | x(g) |\Big)\Big(\sum_{g \in G} | x(g) | \, | \varphi_g |^2(h)\Big) \, ,$$

so that

$$\lambda\Big(\Big(\sum_{g \in G} | x(g) | \, | \varphi_g |\Big)^2\Big) \leqq \| x \| \, \lambda\Big(\sum_{g \in G} | x(g) | \, | \varphi_g |^2\Big)$$

and hence

$$\begin{aligned} \lambda\Big(\Big(\sum_{g \in G} | x(g) | \, | \varphi_g |^2\Big) &\leqq \| x \| \, \lambda\Big(\sum_{g \in G} | x(g) | \, | \varphi_g |^2\Big) = \| x \| \sum_{g \in G} | x(g) | \, \lambda(| \varphi_g |^2) \\ &= \| x \| \, \| x \| \, \| \dot{\varphi}_g \|_2^2 \, , \end{aligned}$$

so

$$\| T_x\dot{\varphi} \|_2^2 \leqq \| x \|^2 \| \dot{\varphi}_g \|_2^2 \quad \text{or} \quad \| T_x \|_2 \leqq \| x \| \, .$$

We can therefore extend T_x to a bounded operator on $L^2(G, \lambda)$ without increasing its norm. These results are summarized in the following theorem.

THEOREM 5.5. *The operator* T_x *on* $\mathscr{C}-\mathscr{K}$ *defined by* $T_x\dot{\varphi} = \sum_{g\in G} x(g)\dot{\varphi}_g$, *where* $x = \sum_{g\in G} x(g)g \in l^1(G)$, *can be uniquely extended to a bounded operator* $T_x^{(1)}(T_x^{(2)})$ *on* $L^1(G,\lambda)(L^2(G,\lambda))$, *and* $\| T_x^{(2)} \|_2 \leqq \| T_x^{(1)} \|_1 = \| x \|$.

THEOREM 5.6. *Let* G *have an invariant mean* λ, *and let* $x = x^* \in l^1(G)$. *If there is a* $\varphi \in \mathscr{C}(G)$ *whose nullspace contains the range of* $iI + L_x$, *then* $\lambda(|\varphi|) = 0$.

Proof. Since the nullspace of φ contains the range of $iI + L_x$, we have in particular that $\varphi((iI + L_x)(h)) = 0$ for all $h \in G$. Let $\varphi = \{\varphi(g)\}$ and $x = \sum_{g\in G} x(g)g$. Then

$$L_x h = \sum_{g\in G} x(g)gh = \sum_{g\in G} x(gh^{-1})g ,$$

and

$$(iI + L_x)(h) = ih + \sum_{g\in G} x(gh^{-1})g = (i + x(e))(h) + \sum_{g\neq h} x(gh^{-1})g .$$

Hence

$$0 = (i + x(e))\varphi(h) + \sum_{g\neq h} x(gh^{-1})\varphi(g) = (i + x(e))\varphi(h) + \sum_{g\neq e} x(g)\varphi(gh) .$$

Taking complex conjugates and letting $\psi(g) = \overline{\varphi(g)}$ we have,

$$0 = (-i + \overline{x(e)})\overline{\varphi(h)} + \sum_{g\neq e} \overline{x(g)}\,\overline{\varphi(gh)} = (-i + \overline{x(e)})\psi(h) + \sum_{g\neq e} \overline{x(g^{-1})}\psi(g^{-1}h)$$
$$= (-i + x(e))\psi(h) + \sum_{g\neq e} x(g)\psi_g(h)$$

for all $h \in G$, since $x = x^*$.

On the other hand,

$$\begin{aligned}((-iI + T_x^{(2)})\psi)(h) &= -i\psi(h) + \sum_{g\in G} x(g)\psi_g(h) \\ &= -i\psi(h) + x(e)\psi(h) + \sum_{g\neq e} x(g)\psi_g(h) \\ &= (-i + x(e))\psi(h) + \sum_{g\neq e} x(g)\psi_g(h) = 0\end{aligned}$$

for all $h \in G$. This means that $\overline{(-iI + T_x^{(2)})(\psi)} \in \mathscr{K}$, and since $T_x^{(2)}$ is an hermitian operator on $L^2(G,\lambda)$, we must have $\dot{\psi} = 0$. Therefore $\lambda(|\varphi|) = \lambda(|\psi|) = 0$.

We now show that the existence of a singular set (not necessarily inverse closed) in a countable group, implies that the group does not have an invariant mean. For this purpose we make essential use of a theorem due to Følner [2].

THEOREM (Følner). *A group* G *has an invariant mean if and only* *if for any finite set F and $\varepsilon > 0$, there exists a finite set A of G such that $|A \cap xA|/|A| > 1 - \varepsilon$ for all $x \in F$.*

LEMMA 5.7. *Let F be a finite subset of a group G such that $[F]$ is infinite. Then if there is a finite set A with $|A \cap xA|/|A| > 1 - \varepsilon$ for all $x \in F$, then $|A| \geqq 1/\varepsilon$.*

Proof. Let $F = \{f_1, f_2, \cdots, f_s\}$, $|A| = r$, and assume that $r < 1/\varepsilon$. Then $\varepsilon < 1/r$, and for any $f_i \in F$, $|A \cap f_iA| > (1-\varepsilon)r > (1 - 1/r)r = r - 1$, and hence $A \cap f_iA = A$ or $f_iA = A$ ($i = 1, 2, \cdots, s$). It follows that $gA = A$ for any $g \in [F]$. Now since A is finite and $[F]$ infinite, there must exist elements $a_0 \in A$, $g_1, g_2 \in F$, $g_1 \neq g_2$ such that $g_1a_0 = g_2a_0$. But this gives $g_1 = g_2$, and we have contradicted the assumption that $|A| < 1/\varepsilon$.

THEOREM 5.8. *If G contains a singular set F with respect to the ordering $\{g_1, g_2, \cdots\}$, then G does not possess an invariant mean.*

Proof. Since F is singular there exists an integer t_0 such that $2|Fg_t \cap (Fg_1 \cup Fg_2 \cup \cdots \cup Fg_{t-1})| \leqq |F| = s$ for $t > t_0$. Assume G does have an invariant mean, so that Fϕlner's condition is satisfied. Let $\varepsilon = 1/72t_0(s-1)$. Then there exists a finite set A with $|A| = r$, and $|A \cap f_iA| > (1-\varepsilon)|A|$ for any $f_i \in F$. From Lemma 5.7 $|A| \geqq 1/\varepsilon = 72t_0(s-1) > 6t_0$. Let $A = \{g_{n_1}, g_{n_2}, \cdots, g_{n_r} : n_1 < n_2 < \cdots < n_r\}$. Then $2|Fg_{n_t} \cap (Fg_{n_1} \cup Fg_{n_2} \cup \cdots \cup Fg_{n_{t-1}})| \leqq s$ for $t > t_0$.
Consider the matrix

$$\begin{bmatrix} f_1g_{n_1} & f_1g_{n_2} & \cdots & f_1g_{n_r} \\ f_2g_{n_1} & f_2g_{n_2} & \cdots & f_2g_{n_r} \\ \cdots & \cdots & & \cdots \\ f_sg_{n_1} & f_sg_{n_2} & \cdots & f_sg_{n_r} \end{bmatrix}$$

and let B denote the set of distinct elements of this matrix. We are first going to get an upper bound for $|B|$ by counting the elements of the matrix row by row, and then a lower bound for $|B|$ by counting them column by column. It will turn out that these bounds are incompatible and the proof completed.

The kth row of the matrix is simply f_kA, and $|f_jA - (f_iA \cap f_jA)| < 3r\varepsilon$. Indeed $|A \cap f_iA)| > (1-\varepsilon)r$ implies $|A - (A \cap f_iA)| < r\varepsilon$ so that

$$\begin{aligned} A = &((A - (A \cap f_iA)) \cap (A - (A \cap f_jA))) \cup ((A - (A \cap f_iA)) \cap (A \cap f_jA)) \\ &\cup ((A \cap f_iA) \cap (A - (A \cap f_jA))) \cup ((A \cap f_iA) \cap (A \cap f_jA)) = A_1 \cup A_2 \cup A_3 \cup A_4 , \end{aligned}$$

where the A_i's are disjoint. Therefore $r = |A| = |A_1| + |A_2| + |A_3| + |A_4| < r\varepsilon + r\varepsilon + r\varepsilon + |A_4|$, and $|(f_iA \cap f_jA)| \geqq |A \cap f_iA \cap f_jA| = |A_4| > r - 3r\varepsilon$ or $|f_jA - (f_iA \cap f_jA)| \, 3r\varepsilon$. Now the first row of the matrix has

r elements and, as has just been shown, each additional row adds less than $3r\varepsilon$ additional distinct elements. Adding, we have $|B| < r + (s-1)3r\varepsilon$.

The first t_0 columns obviously contain at least s distinct elements, and from the singularity condition it follows that each additional column from $t_0 + 1$ through r adds at least $s/2$ distinct elements to B. Hence $|B| \geqq s + (r - t_0)s/2$.

Therefore $s + (r - t_0)s/2 < r + (s-1)3r\varepsilon = r + (s-1)3r/72t_0(s-1) = r + r/24t_0$. Since $r = |A| \geqq 1/\varepsilon > 6t_0$ we have $r - t_0 > 5r/6$. Hence $s + (5r/6)(s/2) < s + (r - t_0)s/2 < r + r/24t_0$. Since $s \geqq 3$, $s(1 + 5r/12) \geqq 3 + 5r/4$ so that $3 + 5r/4 < r + r/24t_0$ or $12 + 5r < 4r + r/6t_0 = r(4 + 1/6t_0) < 5r$, and we have obtained the desired contradiction.

BIBLIOGRAPHY

1. S. Bochner and R. S. Phillips, *Absolutely convergent Fourier expansions for non-commutative normed rings*, Ann. of Math., **43** (1942), 409-418.

2. E. Følner, *On groups with full Banach mean value*, Math. Scand., **3** (1955), 243-254.

3. I. M. Gelfand and M. A. Neumark, *Unitare Darstellungen der Klassischen Gruppen*, Akademie-Verlog, Berlin (1957).

4. A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires*, Memoirs of American Mathematical Society, no. **16** (1955).

5. E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc. Colloq. publications, **31** (1957).

6. M. A. Neumark, *Rings with involution*, Uspehi Mat. Nauk., **3** (1948), 52-145; Translation no. 25, Amer. Math. Soc., New York (1950).

7. C. E. Rickart, *General theory of Banach algebras*, D. Van Nostrand Co. (1960).

8. ———, *The uniqueness of norm problem in Banach algebras*, Ann. of Math., **51** (1950), 615-628.

9. R. Schatten, *A theory of cross spaces*, Ann. of Math. Studies, no. 26, Princeton (1950).

UNIVERSITY OF SOUTHERN CALIFORNIA,
UNIVERSITY OF CALIFORNIA AT LOS ANGELES, AND INSTITUTE FOR ADVANCED STUDY.

MULTIPLICATION OPERATORS

R. C. BUCK

1. Introduction. The prototype for partially ordered linear spaces is $C[X]$, the space of all real valued continuous functions on a topological space X, with the natural ordering defined by: $f \geqq 0$ if and only if $f(x) \geqq 0$ for all $x \in X$. If V is a real linear space with a partial order defined by a suitable positive cone P, then V has a canonical embedding in a function space $C[X]$.

The containing space $C[X]$ has a more elaborate structure than did the original space V; in particular, $C[X]$ is an algebra. If we take any aspect of $C[X]$, we may ask how it appears when transferred back to V. This paper deals with one aspect of this.

Among the linear operators on $C[X]$, an interesting class that arises in many contexts is the class of multiplication operators. These are defined by:

$$T(f) = g \quad \text{where} \quad g(x) = \phi(x)f(x) \qquad x \in X ,$$

and where ϕ is a specific member of $C[X]$.

The central result in this paper is a simple characterization, in terms of order, of the linear operators on V which become multiplication operators when V is represented in a function space $C[X]$. This in turn yields a new and more transparent proof of the Stone-Krein theorem on ordered algebras.

2. A simpler case. Let V be a real linear space. We assume that there is a convex cone P with vertex at 0 which defines an order relation $\leqq$ in V by $x \leqq y$ if and only if $y - x \in P$. On P, we impose three conditions:

(1) $\qquad P \cap -P = \{0\}$

(2) $\qquad P$ is generating

(3) $\qquad P$ is linearly closed in V.

The second condition implies that every element $x \in V$ is the difference of positive elements; the third condition requires that every line meet P in a (possibly unbounded) closed interval. Note that we do not impose any further lattice properties on V, nor do we assume that there is an order unit. If V' denotes the dual space of V, consisting of all linear functionals on V, then V' has a natural partial ordering derived from that of V. A functional L is said to be positive if $L(x) \geqq 0$ for

Received January 11, 1960. Some of this work was done while the author was a Guggenheim Fellow.

all $x \geqq 0$; the positive cone in V' is P'. The space V' will not in general obey all the properties (1), (2), (3).

Let $\mathscr{L}(V)$ denote the algebra of all linear transformations on V. We single out a subclass $\mathfrak{A} \subset \mathscr{L}(V)$ consisting of the order-bounded transformations:

DEFINITION 1. An operator $T \in \mathscr{L}(V)$ is order bounded if there is a constant r such that

$$-rx \leqq Tx \leqq rx \quad \text{for all} \quad x \geqq 0 \quad \text{in } V. \tag{4}$$

We observe that $\mathfrak{A}$ is a subalgebra of $\mathscr{L}(V)$ containing the identity operator I; for, if T_1 and T_2 are in $\mathfrak{A}$, with associated constants r_1 and r_2, then it follows readily from (4) that T_1T_2 obeys (4) with $r = 3r_1r_2$. We wish to show that V has function space representations in which the algebra $\mathfrak{A}$ becomes multiplication operators. We will prove this first under the strong restriction that V has an "order unit", and then remove this restriction.

Let us suppose that there is an element $e \in V$ such that $e \geqq 0$ and

$$\text{for every } x \geqq 0, \text{ there is } \lambda > 0 \text{ such that } x \leqq \lambda e. \tag{5}$$

This restriction can be described geometrically: the point e is a radially interior point of P, so that every line thru e meets P in a line segment containing e as interior point.

THEOREM 1. *Let V be a partially ordered linear space obeying* (1), (2), (3) *and* (5). *Let $\mathfrak{A}$ be the order bounded operators on V. Then there is a compact set Γ and an order preserving representation $\theta: x \rightarrow \hat{x}$ of V onto a subspace of $C[\Gamma]$, and an isomorphism $\bar{\theta}: T \rightarrow \hat{T}$ of $\mathfrak{A}$ into the multiplication operators on $C[\Gamma]$ such that*

$$\theta(Tx) = \hat{T}\hat{x}$$

for all $x \in V$, $T \in \mathfrak{A}$.

Otherwise described, the diagram

$$\begin{array}{ccc} V & \xrightarrow{\theta} & C[\Gamma] \\ \downarrow{\scriptstyle T} & & \downarrow{\scriptstyle \hat{T}} \\ V & \xrightarrow{\theta} & C[\Gamma] \end{array}$$

commutes. Corresponding to T, there is a function $\phi \in C[\Gamma]$ such that if $Tx = y$, then $\hat{y}(p) = \phi(p)\hat{x}(p)$, for all $p \in \Gamma$.

COROLLARY 1. *$\mathfrak{A}$ is a commutative subalgebra of $\mathscr{L}(V)$.*

The method we use will be to construct certain appropriate real homomorphisms of $\mathfrak{A}$. Recall first the important notion of a minimal positive element (See Brelot [3] for background.)

DEFINITION 2. An element $u \geqq 0$ in V is said to be minimal if $0 \leqq x \leqq u$ implies that $x = \lambda u$ for some real λ.

This can be described geometrically: u is minimal if the ray ρ generated by u is extremal in P, and this is so if u cannot be expressed as the midpoint of two points in P that are not on ρ. In contrast with the situation for finite dimensional spaces, a cone P in a general linear space will usually have no extremal rays (or minimal elements). This is the case for $C[X]$ when X is the line, but is not the case if X is discrete. The dual cone P' of positive linear functionals on V can be better behaved; however, if V is the space $L^1[0, 1]$, neither P nor P' have extremal rays.

LEMMA 1. *If P is the positive cone in a space V and P contains a radially interior point, then P' has a separating family of extremal rays.*

This is more or less familiar. (See Bonsall [2], Kadison [8], Kelly [9].) One defines a norm in V by

$$\|x\| = \inf \{\text{all } r \text{ with } -re \leqq x \leqq re\} .$$

Let D be the functionals L on V such that $\|L\| \leqq 1$ and $L(e) = 1$. This is then a w^* compact convex set in the dual space of $\langle V, \|\ \|\rangle$. Invoking the Krein-Milman theorem, D has extreme points L_0 whose convex hull is dense in D. These are in fact minimal positive elements in V', generating extremal rays in P'. Moreover, if $L_0(x) = 0$ for all L_0, then $x = 0$.

The key to the proof of Theorem 1 is the observation that minimal elements of P will yield homomorphism of $\mathfrak{A}$ onto the reals. If $T \in \mathfrak{A}$, then by (4) there is a number r such that

$$0 \leqq rx + Tx \leqq 2rx \qquad \text{all } x \geqq 0 . \tag{6}$$

Let $x = u$, a minimal element of P. Then, we see at once that u is an eigenvector for T. Denoting the corresponding eigenvalue by $\lambda(T)$, we have $Tu = \lambda(T)u$, holding for all $T \in \mathfrak{A}$. But, it then follows that $T \rightarrow \lambda(T)$ is a homomorphism of $\mathfrak{A}$ onto the real field k; for, given T_1 and T_2, we have

$$\begin{aligned} \lambda(T_1T_2)u &= T_1T_2(u) \\ &= T_1(\lambda(T_2)u) \\ &= \lambda(T_1)\lambda(T_2)u . \end{aligned}$$

Unfortunately, except in unusual cases, P will not have any minimal elements. Let us go over to the adjoint algebra $\mathfrak{A}^* \subset \mathscr{L}(V')$ consisting of all operators T^* for $T \in \mathfrak{A}$. T^* is defined on V', the dual space of V, by:

$$T^*(L)(x) = L(Tx) \qquad \begin{array}{l}\text{all } L \in V' \\ \text{all } x \in V,\end{array} \tag{7}$$

and the mapping $T \to T^*$ is an anti-isomorphism of $\mathfrak{A}$ onto $\mathfrak{A}^*$. From (7) and (5), we see that if T obeys (4), then

$$-rL \leqq T^*(L) \leqq rL \qquad \text{all } L \geqq 0 . \tag{8}$$

Thus, $\mathfrak{A}^*$ is an algebra of order-bounded operators on the partially ordered space V'. By Lemma 1, since P was assumed to have an order unit e, there are many minimal elements L_0 in P'.

Let D be the convex cross-section of P' consisting of all $L \geqq 0$ with $L(e) = 1$. Each extreme point of D is a minimal positive element in P' and generates an extremal ray; let Γ be the closure of the set of extreme points in D, in the w^* topology arising from the natural norm topology on V. By the simple argument given above, each $L_0 \in \Gamma$ yields a real homomorphism λ_{L_0} of $\mathfrak{A}^*$, defined by the equation

$$T^*(L_0) = \lambda_{L_0}(T^*)L_0 .$$

Since $\mathfrak{A}^*$ is (anti) isomorphic to $\mathfrak{A}^*$, λ_{L_0} in turn defines a real homomorphism h_{L_0} of $\mathfrak{A}$; using (7), this takes the explicit form:

$$L_0(Tx) = h_{L_0}(T)L_0(x) \qquad \begin{array}{l}\text{all } x \in V \\ \text{all } T \in \mathfrak{A} . \\ \text{all } L_0 \in \Gamma\end{array} \tag{10}$$

By Lemma 1, the functionals L_0 separate V so that the collection of homomorphisms h_{L_0} separate $\mathfrak{A}$. We may conclude that $\mathfrak{A}$ is isomorphic to a product of fields k, and is therefore commutative; this proves the corollary.

To complete the proof of Theorem 1, we examine (10). We first represent V in $C[\Gamma]$, mapping x onto $\theta(x) = \hat{x}$ where $\hat{x}(L_0) = L_0(x)$ for all $L_0 \in \Gamma$. Since $L_0(e) = 1$ for all L_0, $\hat{e}$ is the constant function 1; in fact, the mapping θ is one-to-one and order preserving. For fixed $T \in \mathfrak{A}$, define a function ϕ on Γ by

$$\phi(L_0) = h_{L_0}(T) . \tag{11}$$

Let $Tx = y$; then, (10) can be rewritten as:

$$\hat{y}(L_0) = \phi(L_0)\hat{x}(L_0) . \tag{12}$$

The representation θ is such that every order-bounded operator T is carried into a multiplication operator on $C[\Gamma]$, and the correspondence is an isomorphism of $\mathfrak{A}$ with a subalgebra of $\mathscr{L}(C[\Gamma])$, and in fact, with a subalgebra of $C[\Gamma]$ itself.

3. The Krein-Stone theorem. Before removing the assumption that V possesses an order unit e, we insert an immediate application of our results. (See Stone [14], Krein [10], Kadison [8]).

THEOREM 2. *Let A be a real algebra with unit e and having a partial order such that if $x \geqq 0$, $y \geqq 0$, then $x + y \geqq 0$ and $xy \geqq 0$. Assume further that, as a linear space, A obeys restrictions* (1), (2), (3) *and* (5). *Then, A is commutative and can be represented as a subalgebra of a function algebra $C[X]$.*

Proof. Consider the left regular representation of A. This sends $a \in A$ into the operator $U_a \in \mathscr{L}(A)$ where $U_a(x) = ax$ for all $x \in A$. Since A has a unit, this is an isomorphism of A onto a subalgebra $\bar{A} \subset \mathscr{L}(A)$. By virture of (5), we can choose r depending upon a so that $-re \leqq a \leqq re$. If $x \geqq 0$, then $-rx \leqq ax \leqq rx$ so that U_a is an order bounded operator on the linear space $\langle A, +\rangle$. Hence, $\bar{A} \subset \mathfrak{A}$, and since this is a commutative algebra, so is A.

As a matter of fact, it is not necessary in this proof to assume that A is even associative, since this too can be deduced from the representation. Since $U_a U_b = U_b U_a$, it follows that $a(bx) = b(ax)$ for all $x \in A$; with $x = e$, we find that A is commutative. Then, $a(bc) = a(cb)$ while $b(ac) = (ac)b$ and A is associative.

Conversely, we note that Corollary 1 follows from Theorem 2, since $\mathfrak{A}$ itself is an ordered algebra, with I as unit.

Other proofs which have been given for this result rely upon the construction of appropriate real homomorphisms h of A. These are linear functionals on $\langle A, +\rangle$ which are multiplicative and obey $h(e) = 1$. It is natural to look for these among the extreme points of an appropriate convex set D in the dual space of $\langle A, +\rangle$. Since any finite set of distinct real homomorphisms of A are linearly independent, the collection of h are precisely the extreme points of the convex set D_0 which they generate. Unfortunately, we cannot obtain D_0 directly. Instead, one selects a $D \supset D_0$, easily described, and then proves $D = D_0$. For example, the method adopted in Tate [15], Kadison [8] and Kelley [9] is to select D as all functionals L on $\langle A, +\rangle$ such that $L(e) = 1$ and $L(x^2) \geqq 0$ for all $x \in A$. We note that the proof of $D = D_0$ depends strongly upon the hypotheses on A; one can construct a finite dimensional algebra B for which D is a closed disc, having a circle for its extreme points, but such that B has no proper real homomorphisms.

4. Reduction of the general case. Suppose now that V is not assumed to satisfy (5). This is true for example, of the space $C_0[R]$ of functions with compact support, continuous on the real line R. We reduce this case to the previous one. Let e be an element in P and form

$$V(e) = \{\text{all } x \in V \text{ such that for some } \lambda,\ -\lambda e \leqq x \leqq \lambda e\} . \tag{13}$$

This is a linear subspace of V; it inherits a partial order from V, and in its positive cone $P \cap V(e)$, the element e is an order unit. Suppose that $T \in \mathfrak{A}$. Then, from (4), if $x \in V(e)$, then for the appropriate λ, we have

$$-3\lambda re \leqq Tx \leqq 3\lambda re .$$

Thus, $V(e)$ is left invariant under all operators $T \in \mathfrak{A}$. Accordingly, if we restrict $\mathfrak{A}$ to $V(e)$, we obtain a representation of $\mathfrak{A}$ in $\mathscr{L}(V(e))$. Applying Theorem 1 to the resulting algebra, we find that $\mathfrak{A}$ is commutative in its action on $V(e)$, and also obtain a representation (homomorphic) of $\mathfrak{A}$ as multiplication operators on an appropriate function space $C[\Gamma_e]$. Finally, as e ranges over P, the subspaces $V(e)$ cover V, and we have proved the following result:

THEOREM 3. *Let V be a partially ordered linear space obeying* (1), (2) *and* (3), *but not necessarily* (5). *Let $\mathfrak{A}$ be its algebra of order bounded operators. Then, $\mathfrak{A}$ is commutative, and corresponding to any positive element e in V, there is a compact set Γ_e, an order preserving linear representation θ of $V(e)$ into $C[\Gamma_e]$ and a homomorphism $\bar{\theta}$ of $\mathfrak{A}$ into the multiplication operators on $C[\Gamma_e]$ such that $\theta(Tx) = \bar{\theta}(T)\theta(x)$ for all $x \in V(e)$ and $T \in \mathfrak{A}$.*

A footnote to this is in order. Although we have shown that the algebra $\mathfrak{A}$ is commutative, we have not shown that it need contain more than the multiples of the identity operator I. This can in fact, happen, although it does not in most of the interesting cases discussed in the next section. A glance at the finite dimensional case will be helpful. Let P be a polyhedral cone in n-space, and let $u_1, u_2, \cdots u_N$ generate its extremal rays. Each u_j is an eigenvector for all the order bounded operators $T \in \mathfrak{A}$, and in turn generates real homomorphisms h_j of $\mathfrak{A}$, with

$$T(u_j) = h_j(T)u_j .$$

Suppose that the $\{u_j\}_1^N$ are such that $N > n$ and every set of n is independent. Then, it follows that all the h_j coincide on $\mathfrak{A}$. Since together they define a faithful representation of $\mathfrak{A}$, we conclude that $\mathfrak{A}$ consists exactly of the scalar multiples of I. In contrast, if $N = n$, and the u_j form a basis, then $\mathfrak{A}$ becomes the algebra of diagonal matrices; these, of course, are the multiplication operators in this representation.

5. Examples. In this section, we give a number of interesting illuastrations of Theorem 3, together with a counterexample to show the necessity of the assumption that P is a linearly closed cone.

First, choose V as the space $C_0[X]$ of all real valued continuous functions on the locally compact space X which vanish at infinity. With the usual ordering ($f \geqq 0$ means $f(p) \geqq 0$ for all $p \in X$) this is a partially ordered linear space satisfying the hypotheses of Theorem 3. Note in particular that $C_0[X]$ does not have an order unit. What are the order bounded operators on $C_0[X]$? Applying Theorem 3, we choose any $e \geqq 0$ in $C_0[X]$ and form the subspace $V(e)$. By (13), $f \in V(e)$ if and only if f/e is a bounded function on X. Thus, $V(e)$ is isomorphic to the space of bounded continuous functions on the open support O_e of e. The set Γ_e is the Čech compactification of O_e, which contains O_e densely. Any point $p \in O_e$ defines a minimal functional L_p on $V(e)$ so that by (10) and (12),

$$L_p(Tf) = (Tf)(p) = \phi(p)f(p) \tag{14}$$

for all $p \in 0_e$ and any $T \in \mathfrak{A}$. If X is σ-compact, we can take e so that $O_e = X$, and we find that the only order bounded transformations on $C_0[X]$ are those defined as point-wise multiplication by *bounded* continuous functions ϕ on X. If X is not σ-compact, we arrive at the same conclusion by varying e.

We note that if V is $C[X]$ itself, a simple and direct characterization of the order bounded operators is available. Using the fact that if $f(p_0) = 0$, then we may write $f = f_1 - f_2$ where $f_i \geqq 0$ and $f_i(p_0) = 0$, it readily follows from the characteristic property of T that $(Tf)(p_0) = 0$. Applying this to $f = g - g(p_0)$, we have $Tg = \phi g$ where $\phi = T(1)$.

Another interesting special case is obtained by taking V as the space H of all bounded harmonic functions on an open domain Ω. The constant function is an order unit for H so that we do not need the full machinery of Theorem 3. The extremal rays in P are generated by the R. S. Martin minimal functions (see Brelot [3]) and H is represented as a subspace of the space of continuous functions on the ideal boundary Γ of Ω. The order bounded transformations are represented in turn as $C[\Gamma]$ itself; for any $T \in \mathfrak{A}$, Tf is the harmonic function $g \in H$ which is described by the (abstract) Dirichlet problem $g\,|_\Gamma = \phi f\,|_\Gamma$ where ϕ is the function in $C[\Gamma]$ corresponding to T. Note that T is not a multiplication on Ω itself. With Ω chosen as the unit disc and $\phi(x, y) = x$, we have $T(1) = x$, $T(y) = xy$, but $T(x) = (1/2)\{x^2 - y^2 + 1\}$, and $T(xy) = (1/4)\{3x^2y - y^3 + y\}$.

A somewhat more complicated illustration is provided by the space $C[X : E]$ of all bounded functions f on a locally compact space X with values in a fixed partially ordered linear space E. We order this by saying $f \geqq g$ when $f(p) \geqq g(p)$ for all $p \in X$. We shall also assume that E has an order unit e and require that each f be continuous when E is given the norm topology associated with e. If $v \in E$, denote by $\bar{v}$ the constant function on X with value v. Note that $\bar{e}$ is then an order unit for $C[X : E]$. To apply Theorem 3, we must determine minimal functionals in the dual space of V. We can find one associated with each point

$p_0 \in X$ and any minimal functional θ on E; define L_0 on $C[X: E]$ by $L_0(f) = \theta(f(p_0))$. The following argument proves that L_0 is indeed minimal. Suppose $0 \leqq L \leqq L_0$. Then, for any $v \geqq 0$ in E, $0 \leqq L(\bar{v}) = \theta(v)$. Thus, $v \rightarrow L(\bar{v})$ is a positive linear functional on E which is dominated by θ. Since θ is minimal on E, there is a constant ρ such that $L(\bar{v}) = \rho\theta(\bar{v}) = \rho L_0(\bar{v})$ for all $v \geqq 0$ in E (and thus for all $v \in E$). Suppose now that $f \in C[X: E]$ with $f(p) \leqq f(p_0)$ for all $p \in X$; we shall say that such a function f takes a maximum value at p_0 and that $f \in \mathscr{F}_{p_0}$. Setting $v = f(p_0)$, we have $\bar{v} - f \geqq 0$ so that $0 \leqq L(\bar{v} - f) \leqq L_0(\bar{v} - f)$. But, $L_0(\bar{v} - f) = \theta(v - f(p_0)) = 0$ so that $L(f) = L(\bar{v}) = \rho L_0(\bar{v}) = \rho L_0(f)$. Thus, $L = \rho L_0$ on the linear span of the special class $\mathscr{F}_{p_0}$. Consider now a general function $F \in C[X: E]$; since F is bounded, $\| F(p) \| \leqq M$ for all $p \in X$. Define g, g_1, and g_2 on X by:

$$\left.\begin{aligned} g(p) &= F(p) - F(p_0) \\ g_1(p) &= \frac{1}{2}\{2 \| g(p) \| e + g(p)\} \\ g_2(p) &= \frac{1}{2}\{2 \| g(p) \| e - g(p)\}\} \end{aligned}\right. \qquad p \in X .$$

One sees that $g_i \geqq 0$ and $g_i(p_0) = 0$, with $\| g_i(p) \| \leqq 3M$ for all $p \in X$. Moreover,

$$g(p) = \{4M - g_2(p)\} - \{4M - g_1(p)\}$$

for all $p \in X$, so that $g \in \mathscr{F}_{p_0} - \mathscr{F}_{p_0}$. We conclude that $L(F) = \rho L_0(F)$, so that L_0 is indeed a minimal positive functional on $C[X: E]$.

Let Γ be the set of extreme points in the set D of functionals α on E with $\alpha \geqq 0$ and $\alpha(e) = 1$. Applying Theorem 3, we find that any order bounded operator T has the property that

$$\alpha(T(f)(p_0)) = \alpha(T(\bar{e})(p_0))\alpha(f(p_0)) \tag{15}$$

for all $f \in C[X: E]$, $p_0 \in X$ and $\alpha \in \Gamma$. If we represent the functions f in $C[X : E]$ as functions f on $X \times \Gamma$, then

$$\bar{\theta}(Tf)(p, \alpha) = \phi(p, \alpha)f(p, \alpha)$$

for all (p, α).

The original space $C[X: E]$ is not an algebra, but is a module over the algebra $C[X]$. Formula (9) shows immediately that any order bounded transformation on $C[X: E]$ is in fact *algebraic*. If $\psi \in C[X]$ and $f \in C[X: E]$, then $T(\psi f) = \psi T(f)$. For,

$$\begin{aligned} \alpha(T(\psi f)(p)) &= \phi(p, \alpha)\alpha(\psi(p)f(p)) \\ &= \psi(p)\phi(p, \alpha)\alpha(f(p)) \\ &= \psi(p)\alpha(T(f)(p)) \\ &= \alpha(\psi(p)T(f)(p)) \end{aligned}$$

for each $p \in X$ and $\alpha \in \Gamma$.

Finally, we use a familiar example to show that the most crucial hypothesis on the partially ordered linear space V in Theorem 1 and 3 is that P be *linearly closed*. Take for V the space of all polynomials, with the ordering: $a_0 + a_1x + \cdots + a_mx^m > 0$ if $a_m > 0$. P satisfies the first and second requirements, but is not linearly closed; in fact

$$\lambda(x^2) + (1 - \lambda)(-x) \in P \quad \text{only if } \lambda > 0 .$$

There is no order unit. We can still introduce the algebra $\mathfrak{A}$ of order bounded transformations on V. It is easy to see, however, that $\mathfrak{A}$ is *not* commutative. Let T be defined on V by $T(x^n) = q_n$ where q_n is a polynominal of degree less than n. Then, $I \pm T \geqq 0$ so that $T \in \mathfrak{A}$. In particular, $T_1 = x(d^2/dx^2)$ and $T_2 = d/dx$ are in $\mathfrak{A}$; however, $T_1T_2 \neq T_2T_1$. In this example, the reason for this can be traced to the fact that P is so large that there are too many positive linear operators on V, (and no non-degenerate positive linear functionals).

REFERENCES

1. G. Birkhoff, *Lattice Theory*, Amer. Math. Soc. Colloq. No. 25 (Rev. edition) 1948.
2. F. F. Bonsall, *Sublinear functionals and ideals in partially ordered vector spaces*, Proc. London Math. Soc. **4** (1954), 402–418.
3. M. Brelot, *Topology of R. S. Martin and Green lines, Lectures on Functions of a complex variable*, University of Michigan Press, (1955), 105–121.
4. R. C. Buck, *A factoring theorem for homomorphisms*, Proc. Amer. Math. Soc., **2** (1951), 135–137.
5. H. Cartan, *Un théorème sur les groupes ordonnés*, Bull. Sci. Math., **63** (1939), 201–205.
6. P. C. Curtis, *Order and commutativity in Banach algebras*, Proc. Amer. Math. Soc., **9** (1958), 643–645.
7. N. Jacobson, *Structure theory for algebraic algebras of bounded degree*, Annals of Math., **46** (1945), 695–707.
8. R. V. Kadison, *A representation theory for commutative topological algebras*, Memoirs Amer. Math. Soc. No. 7, (1951).
9. J. L. Kelley and R. L. Vaught, *The positive cone in Banach algebras*, Trans. Amer. Math. Soc., **74** (1953), 44–55.
10. S. Krein, I. Vernikov and A. Tovbin, *Sur les anneaux semiordonnés*, C. R. (Doklady) Akad. Soc. USSR, **30** (1941), 785–787.
11. N. McCoy, *Rings and ideals*, Carus Monograph No. 8, (1948).
12. N. McCoy and A. Forsythe, *On the commutativity of certain rings*, Bull. Amer. Math. Soc., **52** (1946), 523–526.
13. S. Sherman, *Order in operator algebras*, Amer. J. Math., **73** (1951), 227–232.
14. M. H. Stone, *A general theory of spectra*, I, Proc. Nat. Acad. Sci., **26** (1940), 280–283.
15. J. T. Tate, *On the relation between extremal points of convex sets and homomorphisms of algebras*, Commun. Pure and Applied Math., **4** (1951), 31–32.
16. F. B. Wright, *Hölder groups*, Duke Math. J., **24** (1957), 567–571.

INSTITUTE FOR DEFENSE ANALYSES
PRINCETON, NEW JERSEY

SOME GENERALIZATIONS OF METRIC SPACES

JACK G. CEDER

1. Introduction. This paper consists of a study of certain classes of topological spaces (called M_1-, M_2-, and M_3-spaces) which include metric spaces and CW-complexes and are included in the class of all paracompact and perfectly normal spaces. It is shown, for example, that like the case in metric spaces, a subset of an M_2- (or M_3-) space is an M_2- (or M_3-) space; a countable product of M_i-spaces ($i = 1, 2, 3$) is again an M_i-space; and separable is equivalent to Lindelöf in an M_i-space. Moreover, unlike the case in metric spaces, the quotient space obtained by identifying the points of a closed subset of an M_2- (or M_3-) space is again an M_2- (or M_3-) space (for metric spaces such a quotient space need not be first countable). Also, we have $M_1 \rightarrow M_2 \rightarrow M_3$, but whether $M_3 \rightarrow M_2$ or $M_2 \rightarrow M_1$ is unknown.[1]

These classes of spaces are derived from generalizations of the following well-known characterization of metrizability in terms of specific properties of the base:

THEOREM 1.1. (Smirnov [14] or Nagata [12]). *A regular space is metrizable if and only if it has a σ-locally finite base.*

Recall that a σ-locally finite family is a union of countably many locally finite families. It is easily checked that a locally finite family $\boldsymbol{U}$ of sets has the property, called *closure preserving*, that for any

$$\boldsymbol{V} \subset \boldsymbol{U}, \quad (\cup \{V \in \boldsymbol{V}\})^- = \cup \{V^- : V \in \boldsymbol{V}\} .$$

This, then, suggests we consider spaces having a *σ-closure preserving base* (that is, a base which is the union of countably many closure preserving families).

DEFINITION 1.1. An M_1-space is a regular space having a σ-closure preserving base.

Although conceptually simple, M_1-spaces prove unsatisfactory in some respects, so we weaken the condition of having a σ-closure preserving base. We begin by calling a collection $\boldsymbol{B}$ of (not necessarily open!) subsets of X a *quasi-base* if, whenever $x \in X$ and U is a neighborhood of

Received June 12, 1959, resubmitted November 13, 1959. This paper represents part of the authors doctoral dissertation at the University of Washington, prepared under the guidance of Professor E. A. Michael, to whom the author wishes to express his gratitude for his advice and encouragement. The author is also indebted to Professor Jun-iti Nagata for some helpful correspondence.

[1] Nearly all topological terminology appearing in this paper is consistent with that used in Kelley [4]. Exceptions are that our regular, and normal spaces are assumed to be T_1-spaces.

x, then there exists a $B \in \boldsymbol{B}$ such that $x \in B^0 \subset B \subset U$ where B^0 denotes the interior of B).

DEFINITION 1.2. An M_2-space is a regular space with a σ-closure preserving quasi-base.

Now we proceed to weaken the condition of having a σ-closure preserving quasi-base. Let $\boldsymbol{P}$ be a collection of ordered pairs $P = (P_1, P_2)$ of subsets of X, with $P_1 \subset P_2$ for all $P \in \boldsymbol{P}$. Then $\boldsymbol{P}$ is called a *pair-base* for X if P_1 is open for all $P \in \boldsymbol{P}$ and if, for any $x \in X$ and neighborhood U of x, there exists a $P \in \boldsymbol{P}$ such that $x \in P_1 \subset P_2 \subset U$. Moreovor, $\boldsymbol{P}$ is called *cushioned* if for every $\boldsymbol{P}' \subset \boldsymbol{P}$,

$$(\bigcup\{P_1 : P \in \boldsymbol{P}'\})^- \subset \bigcup\{P_2 : P \in \boldsymbol{P}'\} .$$

$\boldsymbol{P}$ is called *σ-cushioned* if it is the union of countably many cushioned subcollections.

DEFINITION 1.3. An M_3-space is a T_1-space with a σ-cushioned pair-base.

2. Properties of M_i-spaces.

THEOREM 2.1. (Michael [6]). *A T_1-space is paracompact if and only if every open cover $\boldsymbol{U}$ has a σ-cushioned open refinement $\boldsymbol{V}$ (that is, $\boldsymbol{V} = \bigcup_{n=1}^{\infty} \boldsymbol{V}_n$, where for each n, and $V \in \boldsymbol{V}_n$ one can assign a $U_{V,n} \in \boldsymbol{U}$ such that $\{(V, U_{V,n}) : V \in \boldsymbol{V}_n\}$ is cushioned).*

THEOREM 2.2. *The following implications hold: Metrizable $\rightarrow M_1 \rightarrow M_2 \rightarrow M_3 \rightarrow$ paracompact and perfectly normal.*

Proof. Metrizable $\rightarrow M_1$ and $M_1 \rightarrow M_2$ are obvious.

To show $M_2 \rightarrow M_3$, let $\bigcup_{n=1}^{\infty} \boldsymbol{B}_n$ be a σ-closure preserving quasi-base. For each n, put $\boldsymbol{P}_n = \{(B^0, \bar{B}) : B \in \boldsymbol{B}_n\}$. Then clearly $\bigcup_{n=1}^{\infty} \boldsymbol{P}_n$ becomes a σ-cushioned pair-base.

To show $M_3 \rightarrow$ paracompactness, let $\bigcup_{n=1}^{\infty} \boldsymbol{P}_n$ be a σ-cushioned pair-base. Let $\boldsymbol{U}$ be an open cover and for each n, let $\boldsymbol{W} = \{P_1 \subset P_2 \subset U_{W,n}$ for some $U \in \boldsymbol{U}$, $U \in \boldsymbol{P}_n\}$. For $W \in \boldsymbol{W}_n$, pick $U_{W,n} \in \boldsymbol{U}$ such that for some $P \in \boldsymbol{P}_n$, $W = P_1 \subset P_2\, U_{W,n}$. Then $\boldsymbol{W} = \bigcup_{n=1}^{\infty} \boldsymbol{W}_n$ becomes a σ-cushioned open refinement of $\boldsymbol{U}$ and hence, by Theorem 2.1, X is paracompact.

To show $M_3 \rightarrow$ perfectly normal, let G be an open set in X. For each n, put $F_n = (\cup\{P_1 : P_2 \subset G, P \in \boldsymbol{P}_n\})^-$. Then $G = \bigcup_{n=1}^{\infty} F_n$, so every open set is an F_σ, whence X is perfectly normal since X is normal by paracompactness, thus completing the proof of Theorem 2.2.

Example 9.2 furnishes us with a separable and first countable M_1-space which is non-metrizable. The "half-open interval" space R (the real line R with base the family $\{[x, y) : x, y \in R\}$ is paracompact and perfectly normal and $R \times R$ is not paracompact (Sorgenfrey [16] or Kelley [4]). Hence, by Theorem 2.2, $R \times R$ is not M_3, and by Theorem 2.4 it follows that R is not M_3. The questions of whether $M_2 \to M_1$ or $M_3 \to M_2$ remain unsolved. However, see Proposition 7.7 for a partial result.

The following three theorems exhibit properties which metric spaces have in common with M_i-spaces.

THEOREM 2.3. *If A is a subset of an M_2- (or M_3-) space X, then A is M_2 (or M_3).*

Proof. We prove it only for the M_2-case. Let $\bigcup_{n=1}^{\infty}\boldsymbol{B}_n$ be a σ-closure preserving quasi-base for X. For each n, put $\boldsymbol{B}_n' = \{A \cap \bar{B} : B \in \boldsymbol{B}_n\}$. To show $\boldsymbol{B}_n'$ is closure preserving in A it suffices to show for $x \in A$ and $\boldsymbol{A} \subset \boldsymbol{B}_n$, that $x \notin \cup \{(A \cap \bar{B})^- : B \in \boldsymbol{A}\}$ implies $x \notin (\cup\{A \cap \bar{B} : B \in \boldsymbol{A}\})^-$. But for any $B \in \boldsymbol{A}, x \notin (A \cap \bar{B})^-$ implies $x \notin A \cap \bar{B}$ and $x \notin \bar{B}$. So $x \notin \cap \{\bar{B} : B \in \boldsymbol{A}\} = (\cup\{\bar{B} : B \in \boldsymbol{A})^-$ and hence, $x \notin (\cup\{A \cap \bar{B} : B \in \boldsymbol{A}\})^-$ and $\boldsymbol{B}_n'$ is closure preserving. Let U be open about x in A. Then for some U' open in X we have $U = U' \cap A$, so there exists B in some $\boldsymbol{B}_n$ so that $x \in B^0 \subset B \subset \bar{B} \subset U'$. Then with $A \cap \bar{B} \in \boldsymbol{B}_n'$, we have $x \in (\boldsymbol{B}^0 \cap A) \subset (A \cap \bar{B})^0 \subset (A \cap \bar{B}) \subset (U' \cap A) = U$. Hence A is M_2, which completes the proof.

The foregoing proof breaks down in the case of an M_1-space (since in general $(B^0 \cap A)^- \neq (A \cap \bar{B})$), and it is unsolved whether a subspace, or even a closed subspace, of an M_1-space is M_1.

THEOREM 2.4. *A countable product of M_i-spaces is M_i.*

Proof. We prove it only for the M_1 case; the other cases follow similarly. For each n, let X_n be an M_1-space with a σ-closure preserving base $\bigcup_{m=1}^{\infty}\boldsymbol{B}_n^m$. Without loss of generality we can assume that, for all m, n, $X_n \in \boldsymbol{B}_n^m$ and $\boldsymbol{B}_n^m \subset \boldsymbol{B}_n^{m+1}$. Now put $X = \prod_{n=1}^{\infty}X_n$ and, for each n, let

$$\boldsymbol{B}_n = \prod_{i=1}^{n}\left\{B_i : B_i \in \boldsymbol{B}_i^n\right\},$$

where

$$\prod_{i=1}^{n}B_i = \{x \in X : x_i \in B_i \text{ for } i \leqq n\}.$$

Then $\bigcup_{n=1}^{\infty}\boldsymbol{B}_n$ becomes a σ-closure preserving base for X, making X an M_1-space.

We can also prove the following result:

THEOREM 2.5. *Let X be an M_i-space. Then the following are equivalent*:

(1) *X is separable,*

(2) *X is Lindelöf,*

(3) *X is satisfies the countable chain condition* (that is, every disjoint family of open sets is countable).

A separable M_1-space need not have a countable base; for example, see Example 9.2.

Smirnov [15] has shown that any locally metrizable paracompact space is metrizable. And Nagata [13] has obtained the stronger result that a space which is the union of a locally finite family of closed metrizable subsets in metrizable. We can obtain analogous results as follows:

THEOREM 2.6. *If X is paracompact and locally M_i, then X is M_i.*

Proof. We prove it only for the M_1 case, and note that the others follow analogously. For each $x \in X$, there exists an open neighborhood $W(x)$ of x such that $W(x)$ is M_1. By paracompactness, let $\{U_\alpha : \alpha \in A\}$ be an open locally finite refinement of $\{W(x) : x \in X\}$. Then, since an open subset of an M_1-space is clearly M_1, each U_α is M_1. Let $\boldsymbol{B}^\alpha = \bigcup_{n=1}^\infty \boldsymbol{B}_n^\alpha$ be a σ-closure preserving base for U_α such that, for each $B \in \boldsymbol{B}^\alpha$, $\bar{B} \subset U_\alpha$. For each n, put $\boldsymbol{C}_n = \cup\{\boldsymbol{B}_n^\alpha : \alpha \in A\}$. Then it easily follows that each $\boldsymbol{C}_n$ is closure preserving and $\bigcup_{n=1}^\infty \boldsymbol{C}_n$ is a base for X.

LEMMA 2.7. *If $X = A_1 \cup A_2$, where A_1 and A_2 are closed M_2- (or M_3-) subspaces, then X is M_2 (or M_3).*

Proof. First we get X to be regular (Nagata [12]). For the M_2 case, let $\bigcup_{n=1}^\infty \boldsymbol{B}_n^1$ and $\bigcup_{n=1}^\infty \boldsymbol{B}_n^2$ be σ-closure preserving quasi-bases for A_1 and A_2 respectively, with $\phi \in \boldsymbol{B}_n^1 \cap \boldsymbol{B}_n^2$ for all n. Now for each n, m, we put $\boldsymbol{B}_{n,m} = \{B_1 \cup B_2 : B_1 \in \boldsymbol{B}_n^1, B_2 \in \boldsymbol{B}_m^2\}$. Then it is easily checked that $\bigcup_{n,m=1}^\infty \boldsymbol{B}_{n,m}$ is a σ-closure preserving quasi-base for X. Hence X is M_2. The M_3 case is similar.

THEOREM 2.8. *If X is a locally finite union of closed M_2- (or M_3-) spaces, then X is M_2 (or M_3).*

Proof. First we apply a theorem of Michael [7, pp. 379–380] and Morita [10] (see Theorem 8.1 of this paper) to get X paracompact. Let X be the union of a locally finite family $\boldsymbol{A}$ of closed M_2- (or M_3-) spaces. Then, for each $x \in X$, there exists an open U_x containing x which intersects only finitely many members of $\boldsymbol{A}$, say $F_1, \cdots, F_n$. Then $x \in U_x \subset \bigcup_{i=1}^n F_i$. But by Lemma 2.7 $\bigcup_{i=1}^n F_i$ is M_2 (or M_3), and then by Theorem 2.3 we see that U_x is M_2 (or M_3). Now, since X is paracompact and locally M_2 (or M_3), we get X to be M_2 (or M_3) by Theorem 2.6, which completes the proof.

Whether Theorem 2.9 is true for M_1-space is unknown.

3. Nagata spaces.

DEFINITION 3.1. A *Nagata* space X is a T_1-space such that for each $x \in X$ there exist sequences of neighborhoods of x, $\{U_n(x)\}_{n=1}^\infty$ and $\{S_n(x)\}_{n=1}^\infty$, such that:

(1) for each $x \in X$, $\{U_n(x)\}_{n=1}^\infty$ is a local base of neighborhoods of x,

(2) for all $x, y \in X$, $S_n(x) \cap S_n(y) \neq \phi$ implies $x \in U_n(y)$.

The order pair $\langle\{U_n(x)\}_{n=1}^\infty, \{S_n(x)\}_{n=1}^\infty\rangle$ is said to be a *Nagata structure* for X if and only if, for each x, $\{U_n(x)\}_{n=1}^\infty$ and $\{S_n(x)\}_{n=1}^\infty$ are sequences of neighborhoods of x satisfying the above two conditions.

Now having defined Nagata spaces, we get the following relation between a Nagata space and an M_3-space:

THEOREM 3.1. *A topological space is a Nagata space if and only if it is first countable and M_3.*

Proof. Let X be a Nagata space with a Nagata structure $\langle\{U_n(x)\}_{n=1}^\infty, \{S_n(x)\}_{n=1}^\infty\rangle$. Define $\boldsymbol{P}_n = \{(S_n(x)^0, U_n(x)) : x \in X\}$ for each n. Then obviously $\bigcup_{n=1}^\infty \boldsymbol{P}_n$ is a pair-base. To show that each $\boldsymbol{P}_n$ is cushioned, we must show, for any index set A, that $(\cup\{S_n(x_\alpha)^0 : \alpha \in A\})^- \subset \cup\{U_n(x_\alpha) : \alpha \in A\}$. Suppose $y \notin \cup\{U_n(x_\alpha) : \alpha \in A\}$. Then $S_n(y)^0 \cap S_n(x_\alpha)^0 = \phi$ for all α in A. Hence, $S_n(y)^0 \cap (\cup\{S_n(x_\alpha)^0 : \alpha \in A\}) = \phi$ and $y \notin (\cup\{S_n(x_\alpha)^0 : \alpha \in A\})^-$. Thus X is M_3 and first countable.

Now let X be M_3 and first countable. For each $x \in X$, let $\{W_n(x)\}_{n=1}^\infty$ be a local base at x. Suppose $\bigcup_{n=1}^\infty \boldsymbol{P}_n$ is a σ-cushioned pair-base for X. We can assume that for all n, $(X, X) \in \boldsymbol{P}_n$. For m, n and $x \in X$ define

$$U_{m,n}(x) = \cap\{\bar{P}_2 : W_m(x) \subset P_1, P \in \boldsymbol{P}_n\}$$

and

$$S_{m,n}(x) = \cap\{P_1 : W_m(x) \subset P_1, P \in \boldsymbol{P}_n\} - \cup\{\bar{P}_1 : x \notin P_2, P \in \boldsymbol{P}_n\}.$$

We wish to show that $\langle\{U_{m,n}(x)\}_{m,n=1}^\infty, \{S_{m,n}(x)\}_{m,n=1}^\infty\rangle$ is a Nagata structure for X. Obviously $\{U_{m,n}(x)\}_{m,n=1}^\infty$ and $\{S_{m,n}(x)\}_{m,n=1}^\infty$ are sequences of neighborhoods of x satisfying condition (1) in Definitition 3.1. To show (2), suppose $y \notin U_{m,n}(x)$. Then there exists a $P \in \boldsymbol{P}_n$ such that $W_m(x) \subset P_1$ and $y \notin \bar{P}_2$. Then, by definition of $S_{m,n}(x)$, we have $S_{m,n}(y) \cap \bar{P}_1 = \phi$. But $S_{m,n}(x) \subset P_1$, so $S_{m,n}(x) \cap S_{m,n}(y) = \phi$, which completes the proof.

Now by virture of Theorem 3.1 and the fact subsets and countable products of first countable spaces are first countable, we obtain the results that: any subspace of a Nagata space is a Nagata space; a countable product of Nagata spaces is Nagata; and in a Nagata space, separable $\longleftrightarrow$ Lindelöf $\longleftrightarrow$ the countable chain condition.

We can also get the following generalization (from X being metric to X being Nagata) of a well known extension theorem of Dugundji [3]:

THEOREM 3.2. *Let A be a closed subset of a Nagata space X and let f be a continuous map from A into a convex subset K of a locally convex topological linear space Y. Then f can be extended to a continuous g from X to K.*

Proof. Let $\langle\{U_n(x)\}_{n=1}^{\infty}, \{S_n(x)\}_{n=1}^{\infty}\rangle$ be a Nagata structure for X. Without loss of generality we can suppose that, for $n < m$ and $y \in X$, we have $S_m(y) \subset S_n(y)$, $U_m(y) \subset U_n(y)$, and $S_1(y) = U_1(y) = X$. Now for $x \in X - A$, put $n_x = \max\{n : \text{for some } y \in A, x \in S_n(y)\}$ and $m_x = \min\{n : U_n(x) \cap A = \phi\}$. By the paracompactness of $X - A$, let $\boldsymbol{V}$ be an open locally finite refinement of $\{S_{m_x}(x) : x \in X - A\}$. For each $V \in \boldsymbol{V}$ pick x_V such that $V \subset S_{m_{x_V}}(x_V)$, and pick a_V such that $x_V \in S_{n_{x_V}}(a_v)$. Now let $\{p_V : V \in \boldsymbol{V}\}$ be a partition of unity subordinate to $\boldsymbol{V}$, and define $g: X \rightarrow Y$ by

$$g(x) = f(x) \quad \text{for } x \in A$$

and

$$g(x) = \sum_{V \in \boldsymbol{V}} p_V(x) f(a_V) \quad \text{for } x \notin A .$$

Then it can be shown without difficulty that g is the desired extension of f.

4. Some metrization theorems. The following is a recent characterization of metrizability by Nagata [13], which has the dual virture of being obviously satisfied by a metric space and of easily implying many other known metrization theorems. (The concept of a Nagata space was actually abstracted from this characterization.)

THEOREM 4.1. (Nagata [13]). *A T_1-space X is metrizable if and only if X is a Nagata space with a Nagata structure $\langle\{U_n(x)\}_{n=1}^{\infty}, \{S_n(x)\}_{n=1}^{\infty}\rangle$ with the property that $x \in S_n(y)$ implies $S_n(x) \subset U_n(y)$ for all $x, y \in X$.*

The following theorems are consequences of this result:

THEOREM 4.2. *A regular space X is metrizable if and only if X is an M_1-space with σ-closure preserving base $\boldsymbol{B} = \bigcup_{n=1}^{\infty} \boldsymbol{B}_n$ such that, for each $x \in X$ and each n, $\bigcap\{B : x \in \boldsymbol{B}_n\}$ is neighborhood of x.*

Proof. The sufficiency follows easily from Theorem 1.1. For the necessity, we put, for $x \in X$ and m,

$$U_m(x) = \bigcap\{\bar{B} : x \in B \in \boldsymbol{B}_m\} ,$$

and

$$S_m(x) = \bigcap\{B : x \in B \in \boldsymbol{B}_m\} - \bigcup\{\bar{B} : x \notin \bar{B} \text{ and } B \in \boldsymbol{B}_m\} .$$

Then it is easily checked that $\langle\{U_m(x)\}_{m=1}^\infty, \{S_m(x)\}_{m=1}^\infty\rangle$ is a Nagata structure for X with the property that $x \in S_n(y)$ implies $S_n(x) \subset U_n(y)$ for all $x, y \in X$. Hence, according to Theorem 4.1, X is metrizable.

COROLLARY 4.3. *A regular space X is metrizable if and only if X has a σ-closure preserving base $\boldsymbol{B} = \bigcup_{n=1}^\infty \boldsymbol{B}_n$ where each $\boldsymbol{B}_n$ is point finite.*

Proof. The sufficiency follows from Theorem 1.1 and the necessity from Theorem 4.2.

The above theorem and corollary have analogues for the case of M_2- and M_3-spaces.

An interesting but unsolved problem poses itself here, namely: is an M_1-space with a σ-closure preserving base $\boldsymbol{B} = \bigcup_{n=1}^\infty \boldsymbol{B}_n$, where each $\boldsymbol{B}_n$ is point countable, necessarily metrizable?

We also have the following metrization theorem on M_1-spaces:

THEOREM 4.4. (Bing [1]). *A T_1-space X is metrizable if and only if X is an M_1-space with a σ-closure preserving base $\bigcup_{n=1}^\infty \boldsymbol{B}_n$ such that, for any $x \in X$ and open set U containing x, there exists an n such that $\phi \neq \bigcup\{B : x \in B \in \boldsymbol{B}_n\} \subset U$.*

We can easily generalize this result to the following:

THEOREM 4.5. *A T_1-space X is metrizable if and only if X is an M_3-space with a σ-cushioned pair-base $\bigcup_{n=1}^\infty \boldsymbol{P}_n$ with the property that for each $x \in X$ and open set U containing x, there exists an n such that $\phi \neq \bigcup\{P_1 . x \in P_1, P \in \boldsymbol{P}_n\} \subset U$.*

5. Completeness. According to Čech [2], a Hausdorff space is *topologically complete* if it is a G_δ in some compact Hausdorff space, and a Hausdorff space is *completely metrizable* if it has a compatible complete metric. Čech then proves that a metrizable space is completely metrizable if and only if it is topologically complete. In this section we investigate topologically complete M_i-spaces.

THEOREM 5.1. (Nagata [13]). *A topologically complete Nagata space is completely metrizable.*

Actually Nagata's proof of Theorem 5.1 establishes the following result.

THEOREM 5.2. *Let X be a paracompact topologically complete space, and suppose there exists a sequence of open converings $\{S_n\}_{n=1}^{\infty}$ such that, for every $x, y \in X, x \neq y$ implies there exists an m such that $y \notin (\bigcup\{S : x \in S \in S_m\})^{-}$. Then X is completely metrizable.*

We can generalize this result by virture of the following lemmas:

LEMMA 5.3. *Let X be a paracompact space. Then, if there exists a sequence of open coverings $\{V_n\}_{n=1}^{\infty}$ such that $x \neq y$ implies there exists an m such that $y \notin \bigcup\{V : x \in V \in V_m\}$, then there exists a sequence of open coverings $\{S_n\}_{n=1}^{\infty}$ such that $x \neq y$ implies there exists an m such that $y \notin (\bigcup\{S : x \in S \in S_m\})^{-}$.*

Proof. Let W_m be an open locally finite refinement of V_m such that, if $W \in W_m$, then $\bar{W} \subset$ some $V \in V_m$. For $V \in V_m$, define $S_V = \bigcup\{W \in W_m : \bar{W} \subset V\}$. Let $S_m = \{S_V : V \in V_m\}$. Then S_m is cushioned in V_m and in particular, if $x \notin \bigcup\{V \in V_m : y \in V\}$, then $x \notin (\bigcup\{S_V \in S_m : y \in V\})^{-}$, and the conclusion of the lemma follows.

LEMMA 5.4. *The diagonal is a G_δ in $X \times X$ if and only if there exists a sequence of open coverings $\{S_n\}_{n=1}^{\infty}$ of X such that for each $x, y \in X$ $x \neq y$ implies there exists an m such that $y \notin \bigcup\{S : x \in S \in S_m\}$.*

Proof. Let Δ be the diagonal in $X \times X$. Suppose $\Delta = \bigcap_{n=1}^{\infty} G_n$ where each G_n is open in $X \times X$. For each n, put $S_n = \{S : S$ open in $X, S \times S \subset G_n\}$. Then if $x \neq y$, there exists an m such that $(x, y) \notin G_m$ and hence $y \notin \bigcup\{S : x \in S \in S_m\}$.

Now assume we have such a sequence of open coverings $\{S_n\}_{n=1}^{\infty}$. For each n, put $G_n = \bigcup\{S \times S : S \in S_n\}$. Then clearly $\Delta = \bigcap_{n=1}^{\infty} G_n$, which completes the proof.

Then obviously we can strengthen Theorem 5.2 to:

THEOREM 5.5. *A paracompact topologically complete space whose diagonal is a G_δ in $X \times X$ is completely metrizable.*

Now we generalize Theorem 5.1. to:

THEOREM 5.6. *A topologically complete M_i-space is completely metrizable.*

Proof. Let X be an M_i-space. Then $X \times X$ is an M_i-space and thus perfectly normal; so the diagonal is a G_δ. Now applying the previous theorem we complete the proof.

COROLLARY 5.7. *A locally compact M_i-space is completely metrizable.*

Proof. It is well known that a locally compact space is open in any Hausdorff space in which it is densely embedded (Kelly [4], p. 163). Hence X is open in $\beta(X)$, the Stone-Čech compactification of X, and, by Theorem 5.6, X is completely metrizable.

Now we proceed to establish a "completeness-like" condition that will make a Nagata space topologically complete.

DEFINITION 5.1. Let X be a Nagata space. Then the Nagata structure $\langle\{U_n(x)\}_{n=1}^{\infty}, \{S_n(x)\}_{n=1}^{\infty}\rangle$ is *complete* if, whenever $\{A_n\}_{n=1}^{\infty}$ is a decreasing sequence of nonempty closed sets such that for every n there exists x_n and k_n such that $A_{k_n} \subset S_n(x_n)$, we have $\bigcap_{n=1}^{\infty} A_n \neq \phi$.

First we note without proof that:

THEOREM 5.8. *Let X be a Nagata space with Nagata structure $\langle\{U_n(x)\}_{n=1}^{\infty}, \{S_n(x)\}_{n=1}^{\infty}\rangle$. Then the following are equivalent:*

(1) $\langle\{U_n(x)\}_{n=1}^{\infty}, \{S_n(x)]_{n=1}^{\infty}\rangle$ *is complete.*

(2) *Whenever $\boldsymbol{A}$ is a family of closed sets having the finite intersection property such that for every n, there exists $A_n \in \boldsymbol{A}$ and $x_n \in X$ so that $A_n \subset S_n(x_n)$, then $\bigcap \boldsymbol{A} \neq \phi$.*

(3) *If $\{x_m\}_{m=1}^{\infty}$ is a sequence such that for any n there exists k_n, y_n such that $k_n \leqq m$ implies $x_m \subset S_n(y_n)$, then $\{x_m\}_{m=1}^{\infty}$ has a cluster point.*

THEOREM 5.9. *A Nagata space with a complete Nagata structure is completely metrizable.*

Proof. For the proof, we need the concept of the Wallman compactification of a normal space (Wallman [18], Kelly [4, pp. 167–168]). Let X be normal and let $\boldsymbol{F}$ be the family of all closed subsets of X. Define $w(X)$ to be the collection of all subfamilies of $\boldsymbol{F}$ which have the finite intersection property and are maximal with respect to this property. For U open in X, we put $U^+ = \{\boldsymbol{A} \in w(X)$: for some $A \in \boldsymbol{A}, A \subset U\}$. Then $\{U^+ : U \text{ open in } X\}$ is a base for some topology τ. Then $\langle w(X), \tau\rangle$ is called the *Wallman compactification* of X. Then $w(X)$ is compact Hausdorff and X is densely embedded in $w(X)$ by the homeomorphism $\phi(x) = \{A \in \boldsymbol{F} : x \in A\}$.

To show that X is completely metrizable we need only show that X is a G_δ in $w(X)$. Let $\langle\{U_n(x)\}_{n=1}^{\infty}, \{S_n(x)\}_{n=1}^{\infty}\rangle$ be the complete Nagata structure for X. For each n, put $G_n = \cup\{S_n(x)^+ : x \in X\}$. Then G_n is open and obviously $\phi(X) \subset \bigcap_{n=1}^{\infty} G_n$. Now suppose $\boldsymbol{A} \in \bigcap_{n=1}^{\infty} G_n$. Then for each n there exists an $x_n \in X$ such that $\boldsymbol{A} \in S_n(x_n)^+$, which means that for each n there exists $x_n \in X$ and $A_n \in \boldsymbol{A}$ so that $A_n \subset S_n(x_n)$. Hence by completeness $\bigcap \boldsymbol{A} \neq \phi$. So let $x \in \bigcap \boldsymbol{A}$, then since $\boldsymbol{A}$ is maximal with respect to the finite intersection property we must have $\boldsymbol{A} = \phi(x) \in \phi(X)$. Hence, $\phi(X) = \bigcap_{n=1}^{\infty} G_n$, showing that X is a G_δ in $w(X)$.

6. Semi-metric spaces.

DEFINITION 6.1. Let d be a real-valued nonnegative function defined on $X \times X$. Then d is a *semi-metric* for X provided:

$$(1) \qquad d(x, y) = 0 \text{ if and only if } x = y\,,$$

$$(2) \qquad d(x, y) = d(y, x) \text{ for all } x, y \in X\,.$$

If d is a semi-metric for X, the semi-metric topology is that determined by: p is a limit point of $A \subset X$ if and only if $\inf\{d(p, x) : x \in A\} = 0$. A topological space $\langle X, \tau \rangle$ is *semi-metrizable* if and only if there is a semi-metric d such that the semi-metric topology agrees with τ.

We can characterize semi-metric spaces as follows:

THEOREM 6.1. *A Hausdorff space X is semi-metrizable if and only if for all $x \in X$, there exists sequences of neighborhoods $\{U_n(x)\}_{n=1}^\infty$ and $\{S_n(x)\}_{n=1}^\infty$ such that $\{U_n(x)\}_{n=1}^\infty$ is a nested local base of neighborhoods of x, and for each n and $x, y \in X$, $S_n(x) \subset U_n(x)$ and $y \in S_n(x)$ implies $x \in U_n(y)$.*

Proof. For the sufficiency, put $S_n(x) = U_n(x) = \{y : d(x, y) \leqq 1/n\}$. For the necessity, define $d(x, y) = \inf\{1/n : x \in U_n(y) \text{ and } y \in U_n(x)\}$ where we assume $U_1(x) = X$ for all $x \in X$.

Now by virtue of the preceding characterization of semi-metrizability, we obviously have:

THEOREM 6.2. *A Nagata space is semi-metrizable.*

McAuley [5] has given an example of a regular separable semi-metric space X which is not hereditarily sparable (that is, subsets are not necessarily separable). It follows by Theorems 2.3 and 2.5 that X is not a Nagata space. In fact, it can be shown that X is not even paracompact. An interesting unsolved problem is whether a paracompact (or even a regular Lindelöf) semi-metric space must be a Nagata space.

McAuley [5] has defined a semi-metric space to be *strongly-complete* if, whenever $\{A_n\}_{n=1}^\infty$ is a decreasing sequence of nonempty closed sets such that for every n there exists k_n and x_n such that $A_{k_n} \subset \{y : d(x_n, y) \leqq 1/n\}$, then we have $\bigcap_{n=1}^\infty A_{k_n} \neq \phi$. (Theorem 5.8 has an analogue for semi-metric spaces). McAuley has proved the following result concerning strongly complete semi-metric spaces:

THEOREM 6.3. (McAuley [5]). *A regular, hereditarily separable, strongly complete semi-metric space is metrizable.*

The following two theorems, taken together, clarify and improve the above theorem of McAuley.

THEOREM 6.4. *A regular, hereditarily separable, semi-metric space is hereditarily Lindelöf (hence paracompact).*

Proof. Let $\boldsymbol{U}$ be an open cover of X. For each $x \in X$, there exists n_x and $U_x \in \boldsymbol{U}$ such that $S_{n_x}(x) = \{y : d(x, y) < 1/n_x\} \subset U_x$. Put $A_n = \{x \in X : n_x = n\}$. Then A_n has a separable subset $\{d_n^m\}_{m=1}^{\infty}$ and it follows that $A_n \subset \bigcup_{m=1}^{\infty} S_n(d_n^m)$. Now choose $U_n^m \in \boldsymbol{U}$ such that $S_n(d_n^m) \subset U_n^m$. Then

$$X = \bigcup_{n=1}^{\infty} A_n \subset \bigcup_{n,m=1}^{\infty} S_n(d_n^m) \subset \bigcup_{n,m=1}^{\infty} U_n^m .$$

So $\{U_n^m\}_{n,m=1}^{\infty}$ is a countable subcover of $\boldsymbol{U}$. So X is Lindelöf and hence normal, but a normal semi-metric space is easily seen to be perfectly normal, and a perfectly normal Lindelöf space is easily seen to be hereditarily Lindelöf. So we conclude that X is hereditarily Lindelöf and hence paracompact, which completes the proof.

THEOREM 6.5. *A paracompact, strongly complete semi-metric space is completely metrizable.*

Proof. Exactly analogously to the proof of Theorem 5.9 we show that X is a G_δ in $w(X)$. Then we apply Lemma 5.4 and Theorem 5.5, where we take $\mathcal{S}_m = \{S_m(x))^0 : x \in X\}$ and $S_m(x) = \{y : d(x, y) < 1/m\}$, which completes the proof.

7. Closed continuous images.

We have the following theorem about closed continuous images of metric spaces:

THEOREM 7.1. (Stone [17], Morita and Hanai [11]). *Let f be a closed continuous map of a metric space X onto a topological space Y. Then the following are equivalent:*

(1) *Y is first countable,*
(2) *for each $y \in Y$, the boundary of $f^{-1}(y)$, $\partial f^{-1}(y)$, is compact,*
(3) *Y is metrizable.*

A special case of a closed continuous image of a space X is X/A, the quotient space of X formed by identifying the points of a closed subset A. Here, the natural map is clearly closed and continuous. Then, according to Theorem 7.1, if X is a metric space and A is a closed subset of X with a non-compact boundary, then X/A is not metrizable.

We have the following partial analogue to Theorem 7.1:

THEOREM 7.2. *Let X be an M_2- (or M_3-) space and f a closed continuous function from X onto any space Y. Then*

(1) *if Y is first countable, then for all $y \in Y$, $\partial f^{-1}(y)$ is compact,*
(2) *if for all $y \in Y$, $\partial f^{-1}(y)$ is compact, then Y is M_2 (or M_3).*

Proof. The proof of (1) is somewhat similar to Stone's proof of (1) $\rightarrow$ (2) in Theorem 7.1. To prove (2) for the M_2-case let $\bigcup_{n=1}^{\infty}\boldsymbol{B}_n$ be a σ-closure preserving quasi-base for X. Then $\bigcup_{n=1}^{\infty}A_n$ becomes a σ-closure preserving quasi-base for Y, where $\boldsymbol{A}_n = \{f(\bigcup_{i=1}^{k}A_i) : A_1, \cdots, A_k \in \boldsymbol{B}_n\}$. The M_3-case is similar.

The converse of (1) is easily seen to be false by taking the identity map from a non-first countable M_2- (or M_3-) space onto itself. Also, Example 9.2 shows that the converse of (2) is false. It is unknown whether Theorem 7.2 is true for M_1-spaces.

It is also unsolved whether an arbitrary closed continuous image of an M_i-space is again M_i. However we can obtain the partial result that the quotient space of an M_2- (or M_3-) space with respect to a closed subset is again M_2 (or M_3).

For the M_2 case this result would follow if every closed subset A of X had a "local σ-closure preserving quasi-base" in the sense that there exists a σ-closure preserving family $\boldsymbol{V}$ such that for every open U containing A, $A \subset V^0 \subset V \subset U$ for some $V \in \boldsymbol{V}$. For then, if $\boldsymbol{B}$ were a σ-closure preserving quasi-base for X, the image under the natural map of the family $\boldsymbol{V} \cup \{B \in \boldsymbol{B} : \bar{B} \cap A = \phi\}$ would be a σ-closure preserving quasi-base for X/A. As it turns out, we have the stronger result that every closed subset has a "local closure preserving quasi-base" as follows:

LEMMA 7.3. *Let A be a closed subset of an M_2-space X. Then there exists a closure preserving family $\boldsymbol{V}$ of neighborhoods of A such that for every open U containing A, $A \subset V^0 \subset V \subset U$ for some $V \in \boldsymbol{V}$.*

Proof. Let $\boldsymbol{B} = \bigcup_{n=1}^{\infty}\boldsymbol{B}_n$ be a σ-closure preserving quasi-base for X. Without loss of generality we can assume that the members of B are closed and $\boldsymbol{B}_n \subset \boldsymbol{B}_m$ for $n < m$. For each $B \in \boldsymbol{B}_n$ we put

$$R(B, n) = B - \bigcup\{W^0 : A \cap W = \phi,\ W \in \boldsymbol{B}_n\} .$$

Now let $\{\boldsymbol{S}_\alpha : \alpha \in E\}$ be the family of all subcollections of $\boldsymbol{B}$. For each $\alpha \in E$ and n, we put

$$V_{\alpha,n} = \bigcup\{R(B, n) : B \in (\boldsymbol{S}_\alpha \cap \boldsymbol{B}_n)\}$$
$$V_\alpha = \bigcup_{n=1}^{\infty} V_{\alpha,n} \text{ and } D = \{\alpha \in E : A \subset V_\alpha\} .$$

To show $\boldsymbol{V} = \{V_\alpha : \alpha \in D\}$ is closure preserving, let $C \subset D$ and suppose $x \notin \bigcup\{\bar{V}_\alpha : \alpha \in C\}$. Then clearly $x \notin A$; so let k be the least integer for which there exists a $B \in \boldsymbol{B}_{k+1}$ such that $x \in B^0$ and $B \cap A \neq \phi$. Then we have $V_{\alpha n} \cap B^0 = \phi$ for every $n > k$ and $\alpha \in C$. Hence

$x \notin (\bigcup\{V_{\alpha,n} : n > k, \alpha \in C\})^-$. If $k \geqq 1$ (otherwise we are finished), then we also have $x \notin \cup\{W^0 : A \cap W = \phi, W \in \boldsymbol{B}_k\}$. From the facts that $x \notin \bigcup\{W^0 : A \cap W = \phi, W \in \boldsymbol{B}_k\}$ and $x \notin \bigcup\{R(B, k) : B \in (\boldsymbol{S}_\alpha \cap \boldsymbol{B}_k)\}$ it follows that $x \notin \bigcup(\boldsymbol{S}_\alpha \cap \boldsymbol{B}_k)$. Since

$$(\bigcup\{V_{\alpha m} : m \leqq k, \alpha \in C\})^- \subset (\bigcup(\boldsymbol{S}_\alpha \cap \boldsymbol{B}_k))^- = \bigcup(\boldsymbol{S}_\alpha \cap \boldsymbol{B}_k)$$

(because $\boldsymbol{B}_k$ is closure preserving), we have that $x \notin (\bigcup\{V_{\alpha,n} : n \leqq k, \alpha \in C\})^-$. Hence $x \notin (\bigcup\{V_\alpha : \alpha \in C\})^-$.

Finally, suppose U is an open set containing A. For each $x \in A$ there exists n_x and $B_x \in \boldsymbol{B}_{n_x}$ such that $x \in B_x^0 \subset B_x \subset U$. Then x is in the open set $B_x^0 - \bigcup\{W : x \notin W, W \in \boldsymbol{B}_{n_x}\}$ which is included in $R(B_x, n_x)^0$. Hence $x \in R(B_x, n_x)^0 \subset R(B_x, n_x) \subset U$. So putting $\boldsymbol{S}_\alpha = \{B_x : x \in A\}$ we clearly get $A \subset V_\alpha^0 \subset V_\alpha \subset U$ with $\alpha \in D$, which completes the proof.

Lemma 7.4 has an analogue for M_3-spaces. Now by virtue of the remarks preceding Lemma 7.3 we clearly obtain:

THEOREM 7.4. *Let X be an M_2- (or M_3-) space and A a closed subset of X. Then X/A is M_2 (or M_3).*

It is unknown whether the above theorem is true for M_1-spaces. However, we can get X/A to be M_1 if X is metrizable, as follows:

LEMMA 7.5. *Let A be a closed subset of the metric space X. Then there exists a closure preserving family $\boldsymbol{V}$ of open sets such that for every open U containing A, $A \subset V \subset U$ for some $V \in \boldsymbol{V}$.*

Proof. Let $\boldsymbol{B} = \bigcup_{n=1}^\infty \boldsymbol{B}_n$ be a σ-locally finite base for X such that $\boldsymbol{B}_n \subset \boldsymbol{B}_m$ for $n < m$. For each n put

$$A_n = \{y \in X : \operatorname{dist}(y,A) < 1/n\} \text{ and } \boldsymbol{A}_n = \{B \cap A_n : B \in \boldsymbol{B}_n\} .$$

Then each $\boldsymbol{A}_n$ is locally finite. Let $\{\boldsymbol{W}_\alpha : \alpha \in D\}$ be the family of all subcollections of $\bigcup_{n=1}^\infty \boldsymbol{A}_n$ which cover A, and put $\boldsymbol{V} = \{V_\alpha : V_\alpha = \bigcup \boldsymbol{W}_\alpha, \alpha \in D\}$. Then obviously for every open U containing A there exists $\alpha \in D$ such that $A \subset V_\alpha \subset U$. Now consider any $C \subset D$ and suppose $x \notin \bigcup\{\bar{V}_\alpha : \alpha \in C\}$. Then $x \notin A$ and there exists a k such that $x \notin \bar{A}_k$; hence $(X - \bar{A}_k) \cap W = \phi$ for $W \in \boldsymbol{A}_m \cap \boldsymbol{W}_\alpha$ with $k \leqq m$ and $\alpha \in C$. Since $\bigcup_{i=1}^{k-1} \boldsymbol{A}_i$ is closure preserving, it follows that $x \in (\bigcup\{W \in \boldsymbol{A}_m \cap \boldsymbol{W}_\alpha : m < k, \alpha \in C\})^-$. Then we get $x \notin (\bigcup\{V_\alpha : \alpha \in C\})^-$, which completes the proof.

Now we obviously obtain the following:

THEOREM 7.6. *Let X be a metric space and A a closed subset of X. Then X/A is M_1.*

According to Lemma 7.3 every point in an M_2-space has a "local closure preserving quasi-base." It is unsolved, however, if every point in an M_1-space has a "local closure preserving base" (that is, an open local base which is closure preserving). Nevertheless, we can establish the following negative result:

PROPOSITION 7.7. Suppose there exists an M_1-space X with some point p at which there does not exist a closure preserving open local base. Then

(1) there exists an M_2-space which is not M_1,

(2) there exists an M_1-space Y with a closed subset A such that Y/A is not M_1.

Proof. Let $Y = \bigcup_{n=1}^{\infty} X_n$ where $n \neq m$ implies $X_n \cap X_m = \phi$ and each X_n is homeomorphic to X by a map i_n. Topologize Y by: O is open $\longleftrightarrow O \cap X_n$ is open in X_n for all n. Let $p_n = i_n(p)$ and $A = \{y \in Y : y = p_n$ for some $n\}$. Let i be the natural map from Y onto Y/A. Then clearly A is closed and Y is M_2; hence Y/A is M_2. Now suppose Y/A has a σ-closure preserving base $\boldsymbol{B} = \bigcup_{n=1}^{\infty} \boldsymbol{B}_n$. Then for each n, $\{i^{-1}(B) \cap X_n : A \in B \in \boldsymbol{B}_n\}$ is closure preserving in X_n. Hence, there exists an open V_n in X_n so that $p_n \in V_n$ and $A \in B \in \boldsymbol{B}_n$ implies $(i^{-1}(B) \cap X_n) \not\subset V_n$. Now put $V = \bigcup_{n=1}^{\infty} V_n$. Since $\boldsymbol{B}$ is a base there exists some B in some $\boldsymbol{B}_k$ such that $A \in B \subset i(V)$, whence $(i^{-1}(B) \cap X_k) \subset V_k$, which is a contradiction. Hence, Y/A is M_2 but not M_1.

8. The Topology of chunk-complexes. A *chunk-complex* is a topological space $\langle K, \tau \rangle$ having a family $\boldsymbol{K}$ of closed subsets, called *chunks*, such that

(1) $\bigcup \boldsymbol{K} = K$,

(2) for $S, T \in \boldsymbol{K}$, either $S \cap T = \phi$ or $S \cap T \in \boldsymbol{K}$,

(3) for $S \in \boldsymbol{K}$, $\{T \in \boldsymbol{K} : T \subset S\}$ is finite,

(4) each $S \in \boldsymbol{K}$ is a compact metric space $\langle S, \rho_S \rangle$,

(5) $U \in \tau$ if and only if for every $S \in \boldsymbol{K}$, $S \cap U$ is open in $\langle S, \rho_S \rangle$.

If $\boldsymbol{B}$ is a collection of closed subsets of a space X, then $\boldsymbol{B}$ *dominates* X provided that, for every subfamily $\boldsymbol{A}$ of $\boldsymbol{B}$, if $C \subset \bigcup \boldsymbol{A}$ and $A \cap C$ is closed in A for all $A \in \boldsymbol{A}$, then C is closed in X.

THEOREM 8.1. (Michael [7, pp. 379–380], Morita [10]). *If X is dominated by a collection of paracompact (or perfectly normal) subsets, then X is paracompact (or perfectly normal).*

Using Theorem 8.1, it is easy to show that.

LEMMA 8.2. *A chunk-complex is dominated by the set of its chunks, and hence is paracompact and perfectly normal.*

In this section we establish the stronger result that each chunk-complex is an M_1-space.

For the proof we establish the following notation: For $S \in \boldsymbol{K}$ define $\Delta(S) = \{T \in \boldsymbol{K} : T \subset S, T \neq S\}$. Define $\boldsymbol{K}_0 = \{S \in \boldsymbol{K} : \Delta(S) = \phi\}$ and, assuming $\boldsymbol{K}_m$ has been defined for $0 \leqq m < n$, we define

$$\boldsymbol{K}_n = \left\{S \in \boldsymbol{K} : \Delta(S) \subset \bigcup_{i=0}^{n-1} \boldsymbol{K}_i\right\} - \bigcup_{i=0}^{n-1} \boldsymbol{K}_i \,.$$

Then $\bigcup_{n=1}^{\infty} \boldsymbol{K}_n = \boldsymbol{K}$, by induction on the number of subchunks. For $S \in \boldsymbol{K}$ put $\partial S = \bigcup(\Delta(S))$, $S^0 = S - \partial S$, and $\boldsymbol{A}_S = \{T \in \boldsymbol{K} : S \subset T\}$. Then obviously $\bigcup\{S^0 : S \in \boldsymbol{K}\} = K$. Let N be the set of nonnegative integers and $M = \{1/n : n \in N - \{0\}\}$.

THEOREM 8.3. *A chunk-complex is an M_1-space.*

Proof. Let $\langle K, \tau \rangle$ be a chunk-complex with a set of chunks $\boldsymbol{K}$. First we observe that for each $n \in N$, $P \in \boldsymbol{K}_n$, there exists a countable family $\boldsymbol{B}(P) = \{P_m : m \in N\}$ of open sets in P^0 forming a base for points in P^0 so that $\bar{P}_m \in P^0$ for all $m \in N$. Fix $n \in N$, $P \in \boldsymbol{K}_n$ and $B \in \boldsymbol{B}(P)$. Let $g \colon \boldsymbol{A}_P \to M$. Then we define a candidate B_g for our base as follows: By normality, let W be an open set containing $\bar{B}$ and such that $\bar{W} \cap (\bigcup\{T \in \boldsymbol{K} : T \cap P^0 = \phi\}) = \phi$. Now, by induction, for any $T \in \boldsymbol{K}_n \cap \boldsymbol{A}_P$ we necessarily have $T = P$ and we define $B_g^P = B$ and $\dot{B}_g^P = \phi$. Now assume we have defined B_g^S for all $S \in \boldsymbol{K}_{n+k} \cap \boldsymbol{A}_P$ with $k < m$. Then for any $T \in \boldsymbol{K}_{n+m} \cap \boldsymbol{A}_P$ we put

$$\dot{B}_g^T = \bigcup\{B_g^S : S \in \Delta(T) \cap \boldsymbol{A}_P\}$$

and

$$B_g^T = W \cap \{y \in T : \rho_T(\dot{B}_g^T, y) < \min\,[g(T), \rho_T(y, \partial T \dot{-} \dot{B}_g^T)]\} \,.$$

Finally we put

$$B_g = \bigcup\{B_g^T : T \in \boldsymbol{A}_P\} \,.$$

We note that for all $T \in \boldsymbol{A}_P$ we have $(B_g^T \cap \partial T) \subset \dot{B}_g^T$, $((B_g^T)^- \cap \partial T) \subset (\dot{B}_g^T)^-$, and if $S \notin \boldsymbol{A}_P$, $(B_g^T)^- \cap S = \phi$.

Now we need to establish the following lemma:

LEMMA 8.4. *For all $P \in \boldsymbol{K}_n$ and $S, T \in \bigcup_{k=0}^{m} \boldsymbol{K}_{n+k} \cap \boldsymbol{A}_P$,*

(a) *$\dot{B}_g^S$ is open in ∂S,*

(b) *$\dot{B}_g^S \subset B_g^S$,*

(c) *$(B_g^S \cap T) \subset B_g^T$,*

(d) *$((B_g^S)^- \cap T) \subset (B_g^T)^-$.*

Proof. By induction on m: if $m = 0$, then $S = T = P$ and all conditions are obviously satisfied. Now assume that (a), (b), (c) and (d) hold for all $k < m$, and let us prove this for m.

(a) Applying the induction hypothesis on (c) we get for all $R, Q \in \Delta(S) \cap \boldsymbol{A}_P$ that $(B_g^R \cap Q) \subset B_g^Q$. Hence

$$\partial S - \dot{B}_g^S = \partial S - \bigcup\{B_g^T \in \Delta(S)\} = \bigcup\{R - B_g^R : R \in \Delta(S)\} .$$

But each $R - B_g^R$ is closed in R which is in turn closed in ∂S. Hence $\partial S - \dot{B}_g^S$ is closed in ∂S for $S \in \boldsymbol{K}_{n+m}$.

(b) Then if $y \in \dot{B}_g^S$, $\rho_S(y, \dot{B}_g^S) = 0$ and $\rho_S(y, S - \dot{B}_g^S) > 0$, so $y \in B_g^S$. Hence we have $\dot{B}_g^S \subset B_g^S$ for all $S \in \boldsymbol{K}_{n+m}$.

(c) If $S \not\subset T$, then $(B_g^S \cap T) \subset (B_g^S \cap (T \cap S)) \subset (B_g^S \cap \partial S) \subset \dot{B}_g^S$. So if $x \in B_g^S \cap T$, then $x \in$ some B_g^R with $R \in \Delta(S)$, and then $x \in (B_g^R \cap (T \cap S)) \subset B^{T\cap S}$ by the induction hypothesis on (c). If $S \cap T = T$, then $B_g^{S\cap T} = B_g^T$. If $S \cap T \neq T$, then $S \cap T \in \Delta(T)$, and by (b) we have $B_g^{S\cap T} \subset B_g^T$. Hence if $S \not\subset T$, $(B_g^S \cap T) \subset B_g^T$. If $S \subset T$, then $B_g^S \subset B_g^T$ by (b). Hence $(B_g^S \cap T) \subset B_g^T$ for all $S, T \in \boldsymbol{K}_{n+m} \cap \boldsymbol{A}_P$.

(d) The proof of (d) is exactly similar to (c) above; but here we use the fact that $((B_g^S)^- \cap S) \subset (\dot{B}_g^S)^-$.

This completes the proof of Lemma 8.4.

For $m, n \in N$, $P \in \boldsymbol{K}_n$, define $\boldsymbol{V}_P^m = \{(P_m)_g : g : \boldsymbol{A}_P \to M\}$ and $\boldsymbol{U}_n^m = \bigcup\{\boldsymbol{V}_P^m : P \in \boldsymbol{K}_n\}$. Now we will show that

(a) each $(P_m)_g$ is open,

(b) $\bigcup\{\boldsymbol{V}_P^m : m \in N\}$ is a base for points in P^0,

(c) each $\boldsymbol{V}_P^m$ is closure preserving,

(d) each $\boldsymbol{U}_n^m$ is closure preserving.

Then, since $\bigcup\{P^0 : P \in \bigcup_{n=1}^\infty \boldsymbol{K}_n\} = K$, $\boldsymbol{B} = \bigcup\{\boldsymbol{U}_n^m : n, m \in N\}$ will be the desired σ-closure preserving base for K.

(a) each $(P_m)_g$ is open. Let $P_m = B$. It then suffices to show that for every $S \in \boldsymbol{A}_p$, $S \cap B_g$ is open in S. But by Lemma 8.4, $S \cap B_g = \bigcup\{S \cap B_g^T : T \in \boldsymbol{A}_P\} = S \cap B_g^S$, which is open in S by construction.

(b) $\bigcup\{\boldsymbol{V}_P^m : m \in N\}$ is a base for points in P^0. Let $P \in \boldsymbol{K}_n$, $x \in P^0$, and U be on open set containing x. Choose $B \in \boldsymbol{B}(P)$ such that $x \in B \subset \bar{B} \subset (U \cap P^0)$. We want to find $g : \boldsymbol{A}_P \to M$ so that $x \in B_g \subset U$. By induction on m, we define $g(T)$ for $T \in \boldsymbol{K}_{n+m} \cap \boldsymbol{A}_P$ so that $(B_g^T)^- \subset U$. For $m = 0$ we have $T = P$ and $(B_g^T)^- = \bar{B} \subset (P \cap U)$ for any $g : \boldsymbol{A}_P \to M$, so put $g(P) = 1$. Now assume we have defined $g(S)$ for every $S \in \boldsymbol{K}_{n+k} \cap \boldsymbol{A}_P$ with $k < m$ so that $(B_g^S)^- \subset U$. Let $T \in \boldsymbol{K}_{n+m} \cap \boldsymbol{A}_P$. Then, by the induction hypothesis, $(\dot{B}_g^T)^- = \bigcup\{(B_g^S)^- : S \in \Delta(T)\} \subset (U \cap T)$. So by the compactness of T there exists $\beta \in M$ so that $\{y \in T : \rho_T(y, \dot{B}_g^T) \leqq \beta\} \subset (T \cap U)$. Then put $g(T) = \beta$. Then we have

$$\begin{aligned}(B_g^T)^- &= (W \cap \{y \in T : \rho_T(y, \dot{B}_g^T) < \min[g(T), \rho_T(y, \partial T - \dot{B}_g^T)]\})^- \\ &\subset \{y \in T : \rho_T(y, \dot{B}_g^T) \leqq g(T)\} \subset (T \cap U) .\end{aligned}$$

Hence $x \in B_g = \bigcup\{B_g^T : T \in \boldsymbol{A}_P\} \subset U$, with $B_g \in \boldsymbol{V}_P^m$ and $B = P_m$.

(c) each V_P^m is closure-preserving. First we need the following result:

LEMMA 8.5. (Michael [8]). *Let* $D = \prod_{i=1}^{j} M_i$, *where* $M_i = M$ *for all* i. *For all* $x, y \in D$, *define* $x \leqq y$ *if and only if* $x_i \leqq y_i$ *for all* i. *Then* $\langle D, \leqq \rangle$ *is a partially ordered set with the property that, for each* $C \subset D$, *there exist* $c_1, \cdots c_m \in C$ *so that, for all* $c \in C$, *there exists* c_k $(1 \leqq k \leqq m)$ *such that* $c \leqq c_k$.

Now let $\{B_g : g \in G\}$ be a subfamily of V_P^m with $P_m = B$. For every $T \in A_P$ we must show $T \cap (\bigcup\{\bar{B}_g : g \in G\})$ is closed. First we show that $\bar{B}_g = \bigcup\{(B_g^S)^- : S \in A_P\}$. For this it suffices to show, for every $T \in A_P$, that $T \cap (\bigcup\{B_g^S)^- : S \in A_P\})$ is closed. But by part (d) of Lemma 8.4, $T \cap (\bigcup\{(B_g^S)^- : S \in A_P\}) = (B_g^T)^-$. Then

$$T \cap (\bigcup\{\bar{B}_g : g \in G\}) = T \cap (\bigcup\{(B_g^S)^- : g \in G, S \in A_P\}) = \bigcup\{(B_g^T)^- : g \in G\} .$$

Now we apply Lemma 8.5 above to the subset $A = \{(g(S_1), \cdots, g(S_k)) : g \in G\}$ of the partially ordered set $\prod_{i=1}^{k} M_i$, where $\{S_1, \cdots, S_k\} = \Delta(T) \cap A_P$. Notice that, if $g(S_i) \leqq h(S_i)$ for all i with $g, h \in G$, then we have $(B_g^T)^- \subset (B_h^T)^-$. Hence by Lemma 8.5 we get $g_1, \cdots, g_n \in G$ such that

$$T \cap (\bigcup\{\bar{B}_g : g \in G\}) = \bigcup\{(B_g^T)^- : g \in G\} = \bigcup_{i=1}^{n}\{(B_{g_i}^T)^-\} ,$$

which is closed.

(d) each U_n^m is closure preserving. Let U be a subfamily of U_n^m. Then we can express U as $\{(P_m)_g : g \in G_P, P \in P\}$ for some $P \subset K_n$ and $G_P \subset \{g : g : A_P \to M\}$. Let $T \in K$. If $P \not\subset T$, then $T \notin A_P$ and $((P_m)_g)^- \cap T = \phi$. But there are only finitely many $P \in P$ contained in T. Hence there exist $P^1, \cdots, P^k \in P$ so that

$$T \cap (\bigcup\{\bar{B}_g : B_g \in U\}) = T \cap (\bigcup\{((P_m^i)g)^- : 1 \leqq i \leqq k, g \in G_{Pi}\})$$

which is closed by part (c) above.

This completes the proof of the theorem.

COROLLARY 8.6. *A CW-complex* (Whitehead [19]) *is an* M_1*-space.*

Proof. Let $\langle K, \tau \rangle$ be a CW-complex. Then the family of finite subcomplexes is a family of chunks, whence the CW-complex $\langle K, \tau \rangle$ is M_1. (See Whitehead [19] for terminology).

COROLLARY 8.7. *A countable product of CW-complexes is an* M_1*-space; hence; both paracompact and perfectly normal.*

Proof. Apply Theorems 2.2 and 2.4 and Corollary 8.6.

9. Some examples. In the sequel, R will denote the real numbers and N the natural numbers. We will also use the notation $\langle x, y\rangle$ for the point $(x, y) \in R \times R$ to distinguish it from (s, t) which will mean the open interval $\{x \in R : s < x < t\}$ and $[s, t]$ which will be the closed interval $\{x \in R : s \leqq x \leqq t\}$.

EXAMPLE 9.1. A non-metrizable first countable M_1-space.

Let R' be the rational numbers. For $x \in R$, put $L_x = \{\langle x, y\rangle : \langle x, y\rangle \in R \times R, 0 < y\}$ and $X = R \cup (\bigcup\{L_x : x \in R\})$. Then we will define a base for X as follows: For $s, t \in R'$ and $z = \langle x, w\rangle \in L_x$ such that $0 < s < w < t$ we put $\bigcup^x_{s.t}(z) = \{\langle x, y\rangle : S < y < t\}$ and let $\boldsymbol{U}$ be the set of all such $U^x_{s,t}(z)$. For $r, s, t \in R'$ and $z \in R$ such that $s < z < t$ and $r > 0$, we put

$$V_{r,s,t}(z) = (s, t) \cup (\bigcup\{\langle w, y\rangle : 0 < y < r, w \in (s, t) - \{z\}\}) ,$$

and let $\boldsymbol{V}$ be the set of all such $V_{r,s,t}(z)$. Now put $\boldsymbol{B} = \boldsymbol{U} \cup \boldsymbol{V}$. Then it can be easily shown that $\boldsymbol{B}$ is a σ-closure preserving base making X into a non-metrizable first countable M_1-space.

The following example is more powerful than Example 9.1. But here the proof of M_1-ness, which is due to Jun-iti Nagata, is far from being straightforward. (The space of the example seems to have first appeared in McAuley [5]; Nagata [13] gives it without proof of M_1-ness as an example of a non-metrizable, separable Nagata space.)

EXAMPLE 9.2. [Nagata]. A non-metrizable, separable, first countable M_1-space.

Let $X = \{\langle x, y\rangle : \langle x, y\rangle \in R \times R, 0 < x < 1, 0 \leqq y\}$. Clearly $X - (0, 1)$, as a subset of $R \times R$, has a σ-closure preserving base $\boldsymbol{B}$. For $n \in N$ and $\langle p, 0\rangle \in X$, we define

$$U_n(p) = \{p\} \cup \{\langle x, y\rangle \in X : y < n - (n^2 - (x - p)^2)^{1/2}, |x - p| < 1/n\} .$$

Then $\boldsymbol{B} \cup \{U_n(p) : n \in N, \langle p, 0\rangle \in X\}$ is a base which clearly generates a regular topology. Obviously X is separable, first countable, and not second countable; hence X is not metrizable.

To show the existence of a σ-closure preserving base for X, it suffices to show one for points in $(0, 1)$. For $m, q \in N, m < q$, and $0 \leqq k \leqq 2^{m+1} - 2$, we define

$$W_{q,m,k} = \{\langle x, y\rangle : (k)2^{-m-1} < x < (k + 2)2^{-m-1}, 0 < y \leqq 2^{-q}\} .$$

Now consider any $U_n(p)$. Then we can choose $m, k \in N$ so that

$$(k)2^{-m-1} < n^{-1} + p \quad \text{and} \quad (k - 4)2^{-m-1} \leqq p < (k - 3)2^{-m-1} .$$

For this m, k, we put

$$q = \min\{j : W_{j,m,k-2} \subset U_n(p)\} ,$$

$$
\begin{aligned}
I_1 &= W_{q,m,k-2}\ , \\
a_1 &= (k)2^{-m-1}\ , \\
a_2 &= (k-2)2^{-m-1}\ , \\
b_1 &= 2^{-q}\ .
\end{aligned}
$$

Now for each $i \in N$, we choose k_i so that

$$(k_i - 4)2^{-m-i-1} \leqq p < (k_i - 3)2^{-m-i-1}\ .$$

Then we put

$$
\begin{aligned}
q_i &= \min\{j : W_{i,m+i,k_i-2} \subset U_n(p)\}\ , \\
I_{i+1} &= W_{q_i,m+i,k_i-2}\ , \\
a_{i+2} &= (k_i - 2)2^{-m-i-1}\ , \\
b_{i+1} &= 2^{-q_i}\ .
\end{aligned}
$$

Now it follows that for each $i, j \in N$, $i < j$ implies $a_j < a_i$ and $b_j < b_i$, and obviously $b_i \to 0$ and $a_i \to p$.

We also choose $m', k' \in N$ such that

$$p - n^{-1} < (k')2^{-m'-1} \quad \text{and} \quad (k'+3)2^{-m'-1} < p \leqq (k'+4)2^{-m'-1}\ .$$

Then we put

$$
\begin{aligned}
q' &= \min\{j : W_{j,m',k'} \subset U_n(p)\}\ , \\
I'_1 &= W_{q',m',k'}\ , \\
a'_1 &= (k')2^{-m'-1}\ , \\
a'_2 &= (k'+2)2^{-m'-1}\ , \\
b'_1 &= 2^{-q'}\ .
\end{aligned}
$$

Now for $i \in N$, we choose k'_i so that

$$(k'_i + 3)2^{-m'-i-1} < p \leqq (k'_i + 4)2^{-m'-i-1}\ .$$

Then put

$$
\begin{aligned}
q'_i &= \min\{j : W_{j,m'+i,k'_i} \subset U_n(p)\}\ , \\
I'_{i+1} &= W_{q'_i,m+i,k'_i}\ , \\
a'_{i+2} &= (k'_i + 2)2^{-m'-i-1}\ , \\
b'_{i+1} &= 2^{-q'_i}\ .
\end{aligned}
$$

Then for each $i, j \in N$, $i < j$ implies $a'_i < a'_j$ and $b'_i < b'_j$, and obviously $b'_i \to 0$ and $a'_i \to p$.

Now putting

$$N_n(p) = \left(\left(\left(\bigcup_{j=1}^{\infty} I_j\right) \cup \left(\bigcup_{j=1}^{\infty} I'_j\right)\right)^{-}\right)^{0}\ ,$$

it can be shown that $p \in N_n(p) \subset U_n(p)$.

Now consider the countable set

$$T = \{\langle (k')2^{-m'}, (k)2^{-m}\rangle : k, k', m, m' \in N, (k')2^{-m'} < (k)2^{-m}\} .$$

For $t = \langle (k')2^{-m'}, (k)2^{-m}\rangle \in T$, put

$$\boldsymbol{B}_t = \{N_n(p) : a_1' = (k')2^{-m'}, a_1 = (k)2^{-m}\} .$$

Then obviously $\bigcup\{\boldsymbol{B}_t : t \in T) = \{N_n(p) : n \in N, p \in (0, 1)\}$, which is a base for points in $(0, 1)$. Finally, it can be shown that each $\boldsymbol{B}_t$ is closure preserving. Hence $\bigcup\{\boldsymbol{B}_t : t \in T\}$ is a σ-closure preserving base and X is an M_1-space.

If X is the space in Example 9.2, then it can be shown without difficulty that $X/(0, 1)$ is an M_1-space with $(0, 1)$ having a closure preserving local base.

EXAMPLE 9.3. There exists a non-metrizable M_1-space X with $p \in X$ such that p has an uncountable closure preserving local base and $X - \{p\}$ is homeomorphic to R.

Let $p \notin R$ and put $X = R \cup \{p\}$. Let $\{r_n\}_{n=1}^{\infty}$ be an enumeration of the integers and put $B = \{1/n : n \in N - \{1\}\} \cup \{0\}$. Let F be the set of all functions from the integers I to B such that either there exists $r \in I$ such that if $s < r$, then $f(s) = 0$ and if $r \leqq s$, then $f(s) \neq 0$; or for all $r \in I$, $f(r) \neq 0$. For $f \in F$, put $U_f = \bigcup_{n=1}^{\infty}(r_n - f(r_n), r_n + f(r_n))$ where if $f(r_n) = 0$, $(r_n, r_n) = \phi$. Let $\boldsymbol{U} = \{\{p\} \cup U_f : f \in F\}$ and $\boldsymbol{B}$ be a countable base for R. Then it is obvious that $\boldsymbol{U} \cup \boldsymbol{B}$ is a σ-closure preserving base for X. Moreover, it is easy to see that X is not first countable at p and R is homeomorphic to $X - \{p\}$.

It is clear that this construction can be carried out for any noncompact metric space without isolated points. In particular, carrying it out for the rational numbers we get a countable non-metrizable M_1-space.

EXAMPLE 9.4. (Michael [9]). We can get another countable nonmetrizable M_1-space by taking the subspace $I \cup \{p\}$ of $\beta(I)$, where I is the integers and $\beta(I)$ is the Stone-Čech compactification of I and $p \in \beta(I) - I$. Here the family of all open sets containing p is closure preserving.

BIBLIOGRAPHY

1. R. H. Bing, *Metrization of topological spaces*, Canadian J. Math., **3** (1951), 175-186.
2. E. Čech, *On bicompact spaces*, Ann. Math., **38** (1937), 823-844.
3. J. Dugundji, *An extension of Tietze's theorem*, Pacific J. Math., **1** (1951), 353-367.
4. J. L. Kelley, *General Topology*, New York, Van Nostrand Co., Inc., 1955.
5. L. F. McAuley, *A relation between perfect separability, completeness and normality in semi-metric spaces*, Pacific J. Math., **6** (1956), 315-326.
6. E. A. Michael, *Yet another note on paracompact spaces*, Proc. Amer. Math. Soc., **10** (1959), 309-314.

7. ———, *Continuous selection I*, Ann. Math., **63** (1956), 361-382.
8. ———, *On a class of partially ordered sets*, Amer. Math. Monthly, **67** (1960), 448-449.
9. ———, *Another note on paracompact spaces*, Proc. Amer. Math. Soc., **8** (1957), 822-828.
10. K. Morita, *On spaces having the weak topology with respect to closed coverings*, II, Proc. Japan Acad., **30** (1954), 711-717.
11. ——— and S. Hanai, *Closed mappings and metric spaces*, Proc. Japan Acad., **32** (1956), 10-14.
12. J. Nagata, *On a necessary and sufficient condition of metrizability*, J. Inst. Polytech., Osaka City Univ., **1** (1950), 93-100.
13. ———, *A contribution to the theory of metrization*, J. Inst. Polytech., Osaka City Univ., **8** (1957), 185-192.
14. Yu. M. Smirnov, *A necessary and sufficient condition for metrizability of a topological space*, Doklady Akad. Nauk S. S. S. R., (N. S.), **77** (1951), 197-200.
15. ———, *On metrization of topological spaces*, Uspehi Matem. Nauk, **6** (1951), 100-111.
16. R. H. Sorgenfrey, *On the topological product of paracompact spaces*, Bull. Amer. Math. Soc., **53** (1947), 631-632.
17. A. H. Stone, *Metrizability of decomposition spaces*, Proc. Amer. Math. Soc., **7** (1956), 690-700.
18. H. Wallman, *Lattices and topological spaces*, Ann. Math., **39** (1938), 112-126.
19. J. H. C. Whitehead, *Combinatorial homotopy*, I, Bull. Amer. Math. Soc., **55** (1949), 213-245.

University of California,
Santa Barbara

RANDOM CROSSINGS OF CUMULATIVE DISTRIBUTION FUNCTIONS

MEYER DWASS

1. Introduction. Let $X_1, \cdots, X_n$ be n independent and identically distributed random variables, each with continuous c.d.f. (cumulative distribution function), $F(x)$. Let $F_n(x)$ be the empirical c.d.f. of the n random variables and let $N_1(n)$ be the number of times F_n equals F. There is no loss of generality in supposing that the X_i's are distributed uniformly over the interval $(0, 1)$, and to be specific, $N_1(n)$ is defined by

$$N_1(n) = \text{number of indices, } i, \text{ for which } F_n(i/n) = i/n, \quad i = 1, \cdots, n.$$

Similary, let $X_1, \cdots, X_n, \cdots, Y_1, \cdots, Y_n$ be $2n$ independent random variables, each with the same continuous c.d.f., $F(x)$, and let F_n, G_n denote the empirical c.d.f.'s of the X_i's and Y_i's respectively. Let $N_2(n)$ be the number of times F_n equals G_n. That is.

$$\begin{aligned} N_2(n) = {} & \text{number of indices } i \text{ for which } F_n(X_i) = G_n(X_i), \\ & \text{plus} \\ & \text{number of indices } i \text{ for which } F_n(Y_i) = G_n(Y_i), \quad i = 1, \cdots, n. \end{aligned}$$

The purpose of this paper is to show that

$$\lim_{n\to\infty} P\left(\frac{N_1(n)}{\sqrt{2n}} < t\right) = \lim_{n\to\infty} P\left(\frac{N_2(n)}{\sqrt{4n}} < t\right) = 1 - e^{-t^2} .$$

The methods for obtaining these results are practically the same for N_1 and N_2, so the first case is treated with somewhat greater detail. In both cases, the random variables are related to other random variables on appropriate stochastic processes with independent increments, to obtain generating functions for the moments of N_i. The Karamata Tauberian theorem is then applied to describe the asymptotic behavior of these moments.

2. Some preliminaries on the Poisson process. Let $Y(t)$ be the Poisson process with stationary independent increments, $t \geqq 0$, $Y(0) = 0$, $E[Y(1)] = 1$. Consider γt, the straight line coming out of the origin with slope $\gamma > 1$. The random function $Y(t)$ can equal γt at times $1/\gamma$, $2/\gamma$, etc. The event that $Y(t) = \gamma t$ is a recurrent event in the sense of Feller [4]. Because γ is greater than 1, this recurrent event is an *uncertain* one. It was shown by Baxter and Donsker [1] that

Received February 23, 1960. Research done under contract with the U. S. Office of Naval Research. Contract Nonr-1228(10), project NR 047-021.

$$P[Y(t) < \gamma t, \text{ all positive } t] = 1 - 1/\gamma \ .$$

A completely elementary proof of this fact was given by Dwass [3]. In other words, the probability that the uncertain recurrent event under discussion *never* takes place is $1 - 1/\gamma$. To introduce some specific notation, let $N =$ number of times that $Y(t)$ equals γt. That is,

$$N = \text{number of indices } i \text{ for which } Y(i/\gamma) = \gamma(i/\gamma) = i, \quad i = 1, 2, \cdots.$$

The random variable, N is geometrically distributed, specificially,

$$P(N = k) = (1/\gamma)^k(1 - 1/\gamma) \ ,$$

and for the rth factorial moment we have,

$$EN^{(r)} = EN(N-1)(N-2) \cdots (N-r+1) = r!/(r-1)^r \ . \tag{2.1}$$

3. A generating function for $E[N_1^{(r)}(n)]$. The link between the random variables N and $N_1(n)$ lies in the following lemma.

LEMMA 3.1. *The conditional distribution of N given that $Y(t) = \gamma t$ for the last time at time $t = n/\gamma$ is exactly the same as the distribution of $N_1(n)$.*

Proof of Lemma 3.1. This follows directly from the well-known fact that the places where the jumps of $Y(t)$ occur in the interval $(0, a)$ are distributed as n randomly chosen points in $(0, a)$ under the condition that $Y(a) = n$.

Making use of this lemma, we can compute the rth factorical moment of N in the following iterative way. Let A_k denote the event that the last crossing of γt by $Y(t)$ takes place at time k/γ. Then

$$E(N^{(r)}) = \sum_{k=0}^{\infty} E(N^{(r)} \mid A_k)P(A_k) \ .$$

Since

$$P(A_k) = (k/\gamma)^k e^{-k/\gamma}(1 - 1/\gamma)/k!$$

and

$$E(N^{(r)} \mid A_k) = E[N_1^{(r)}(k)], \qquad (k, 0, 1, 2, \cdots) \ ,$$

we have, making use of (2.1), the following theorem.

THEOREM.

$$\sum_{k=0}^{\infty} \frac{e^{-k}k^k}{k!} EN_1^{(r)}(k) \left(\frac{e^{1-1/\gamma}}{\gamma}\right)^k = \frac{r!\,\gamma}{(\gamma - 1)^{r+1}} \ . \tag{3.1}$$

REMARKS.

(a) In (3.1), $e^{-k}k^k/k!$ should be understood to be 1 when $k = 0$.

(b) $u = e^{1-1/\gamma}$ is a strictly decreasing function of $1/\gamma$, for $\gamma \geqq 1$, and maps $(1, \infty)$ onto $(0, 1)$. Let $1/\gamma = P(u)$ denote the inverse function. Then (2.2) can be rewritten,

$$\sum_{k=0}^{\infty} \frac{e^{-k}k^k}{k!} EN_1^{(r)}(k)u^k = \frac{r!}{P(u)[P^{-1}(u) - 1]^{r+1}} = h(u) , \tag{3.2}$$

$$0 \leqq u < 1 .$$

Since
$$\lim_{x \to 1} \frac{1 - xe^{1-x}}{(x-1)^2} = 1/2 ,$$

or equivalently,

$$\lim_{u \to 1} \frac{1-u}{(P^{-1}(u) - 1)^2} = 1/2 ,$$

it follows that

$$\lim_{u \to 1} (1-u)^{(r+1)/2} h(u) = \frac{r!}{2^{(r+1)/2}} . \tag{3.3}$$

If the coefficients of u^k in $h(u)$ form an increasing sequence, then Karamata's Tauberian theorem is applicable and we could conclude that the sum of the first k coefficients of powers of u in $(1-u)h(u)$ is asymptotically equal to

$$k^{(r-1)/2} \frac{r!}{2^{(r+1)/2}\Gamma\left(\frac{r+1}{2}\right)} ,$$

or equivalently,

$$\lim_{k \to \infty} \frac{\frac{e^{-k}k^k}{k!} EN_1^{(r)}(k)}{k^{(r-1)/2}} = \frac{r!}{2^{(r+1)/2}\Gamma\left(\frac{r+1}{2}\right)}$$

$$\lim_{k \to \infty} \frac{EN_1^{(r)}(k)}{k^{r/2}} = \frac{r!\sqrt{\pi}}{2^{r/2}\Gamma\left(\frac{r+1}{2}\right)} = 2^{r/2}\Gamma\left(\frac{r}{2} + 1\right)$$

by the "duplication formula" for the gamma function (p. 240 [6]).

Since the asymptotic behavior of the rth factorial moment is the same as that of the rth ordinary moment, we would have finally,

$$\lim_{k \to \infty} E\left(\frac{N_1(k)}{\sqrt{2k}}\right)^r = \Gamma\left(\frac{r}{2} + 1\right) = \int_0^{\infty} x^r f(x)dx$$

where $f(x)$ is the probability density function

$$f(x) = \begin{cases} 2xe^{-x^2}, & x \geqq 0 \\ 0, & x < 0 \end{cases},$$

However, it is not at all clear that the usual conditions for Karamata's theorem to hold are applicable, and a slightly more delicate argument is required.

4. The limiting distribution of $N_1(n)$. Following the discussion in the last section, the main effort which remains is to justify the applicability of Karamata's theorem.

LEMMA 4.1. *Let $a_i(u)$, $(i = 1, \cdots, r)$ be power series having positive, non-decreasing coefficients. Then $a(u) = \prod_i a_i(u)$ has the same property.*

Proof of Lemma 4.1. $a_i(u)$ has positive, non-decreasing coefficients means that the coefficients of $(1-u)a_i(u)$ are non-negative.

$$(1-u)\prod_i \mathrm{a}_i(u) = \prod_i [(1-u)a_i(u)](1-u)^{-(r-1)}$$

is a product of power series all with non-negative coefficients, which completes the proof.

LEMMA 4.2.

(a) $$\sum_{k=0}^{\infty} \frac{k^k e^{-k}}{k!} u^k = f(u) = 1/(\gamma - 1)$$

(b) $$c(1-u)^{-1/2} - f(u) = g(u)$$

is a power series with positive, non-decreasing coefficients if c is sufficiently large.

Proof of *Lemma* 4.2. Part (a) follows from (3.1) for $r = 0$. The coefficients of $(1-u)^{-1/2}$ are of the order of $1/\sqrt{k}$ and strictly positive. The coefficients of $-f(u)$ are strictly increasing and also of the order of $1/\sqrt{k}$. Hence choosing c sufficiently large will guarantee the result.

Finally, we want to state the following form of Karamata's theorem.

LEMMA 4.3. *Let $a(u) = \sum_{k=0}^{\infty} a_k u^k$ where $\{a_k\}$ is a non-decreasing sequence, and suppose $(1-u)^{\gamma} a(u) \to A$ as $u \to 1$, for $\gamma \geqq 0$. Then*

$$\frac{a^k}{k^{\gamma-1}} \to \frac{A}{\Gamma(\gamma)}$$

as $k \to \infty$.

Proof of Lemma 4.3. For $\gamma > 1$ the result follows from the conventional form of Karamata's theorem (for example see Theorem 4.3, p. 192, [5]) by considering that

$$(1-u)^{\gamma-1}(1-u)a(u) \to A$$

and that the partial sums of coefficients in $(1-u)a(u)$ are a_k.

For $0 \leqq \gamma < 1$ we have that

$$a_1 + \cdots + a_k \sim \frac{A}{\Gamma(\gamma+1)} k^\gamma$$

and we can apply Hilfassatz 3, p. 517, Doetsch, [2], to conclude that

$$a_k \sim \frac{A\gamma}{\Gamma(\gamma+1)} k^\gamma = \frac{A}{\Gamma(\gamma)} k^\gamma .$$

We can now prove the following.

THEOREM.

$$\lim_{n\to\infty} P\left(\frac{N_1(n)}{\sqrt{2n}} < t\right) = 2\int_0^t xe^{-x^2}dx = 1 - e^{-t^2}$$

Proof. The limiting distribution is determined by its moments, hence it is sufficient to show that

$$\lim_{n\to\infty} E\left(\frac{N_1(n)}{\sqrt{2n}}\right)^r = 2\int_0^\infty x^{r+1}e^{-x^2}dx = \Gamma(r/2+1) , \qquad r = 1, 2, \cdots .$$

Referring to (3.2) and to Lemma 4.2, we can write

$$\begin{aligned} \text{(4.1)} \qquad h(u) &= r!\,[1+f(u)][f(u)]^r \\ &= [c(1-u)^{-1/2} - g(u) + 1][c(1-u)^{-1/2} - g(u)]^r . \end{aligned}$$

Since $g(u)$ has positive and increasing coefficients then by Lemma 4.1 so does $(1-u)^{-m/2}[g(u)]^n$ for m, n positive integers, because

$$(1-u)[(1-u)^{-m/2}(g(u))^n] = (1-u)^{-m/2}(1-u)[g(u)]^n$$

has positive coefficients. Hence by Karamata's theorem, since

$$(1-u)^{(m+n)/2}(1-u)^{-m/2}[g(u)]^n \to \left(c - \frac{1}{\sqrt{2}}\right)^n ,$$

the coefficients of $(1-u)^{-m/2}[g(u)]^n$ are asymptotically equivalent to

$$\frac{(c-1/\sqrt{2})^n}{\Gamma\left(\frac{m+n}{2}\right)} k^{(m+n)/2} .$$

On expanding the right side of (4.1), an elementary computation yields the result that the coefficients of $h(u)$ are asymptotically equivalent to

$$\frac{r!}{2^{(r+1)/2}\Gamma\left(\frac{r+1}{2}\right)} k^{(r-1)/2} .$$

According to the discussion in § 4, we conclude then that

$$\lim_{k\to\infty} E\left(\frac{N_1(k)}{\sqrt{2k}}\right)^r = \Gamma\left(\frac{r}{2}+1\right) ,$$

which completes the proof of the theorem.

5. The limiting distribution of $N_2(n)$. In this section we prove the following.

THEOREM.

$$\lim_{n\to\infty} P\left(\frac{N_2(n)}{\sqrt{4n}} < t\right) = 1 - e^{-t^2}$$

The main points of the proof are essentially the same as in the preceding theorem, so we offer an outline of the method only.

Let $X_1, X_2, \cdots$ be a sequence of independent, identically distributed random variables such that

$$X_i = \begin{cases} 1 \text{ with probability } p, \\ 0 \text{ with probability } 1-p. \end{cases}$$

and let S_n denote the sum of the first n random variables.

The event that for a positive integer n, $S_{2n} = n$, is a well-known recurrent event, representing return to the origin, in a discrete random walk on the line. Suppose $p < 1/2$. Then the probability that the recurrent event never takes place is $1 - 2p$. (See Feller, p. 288, [4].) Using N exactly as above, let $N =$ number of indices i for which $S_{2i} = i$, $i = 1, 2, \cdots$. As before, N is a geometric random variable, such that

$$P(N = k) = (1 - 2p)(2p)^k ,$$

and hence

$$EN^{(r)} = \frac{r!}{\left(\frac{1}{2p} - t\right)^r} .$$

Analogous to Lemma 3.1 is the following combinatorial lemma.

LEMMA 5.1. *The conditional distribution N given that $S_{2i} = i$ for the last time when $i = n$ is exactly the same as the distribution of*

$N_2(n)$. We omit the proof which is elementary.

Let A_k denote the event that $S_{2i} = i$ for the last time when $i = k$. Then

$$\begin{aligned} EN^{(r)} &= \sum_{k=0}^{\infty} E(N^{(r)} \mid A_k)P(A_k) \\ &= \sum_{k=0}^{\infty} EN_2^{(r)}(k)\binom{2k}{k}p^k(1-p)^k(1-2p) \,. \end{aligned}$$

Hence

$$f(u) = \sum_{k=0}^{\infty} EN_2^{(r)}(k)\binom{2k}{k}4^k u^k = \frac{r!}{2h(u)\left(\dfrac{1}{2h(u)} - 1\right)^{r+1}} \tag{5.1}$$

where $4p(1-p) = u$, $0 \leqq p \leqq 1/2$, is an increasing function of p which maps $(0, 1/2)$ onto $(0, 1)$, and where $p = h(u)$ is the inverse function.

We next notice that

$$\lim_{u \to 1} (1-u)^{(r+1)/2} f(u) = r! \,,$$

This follows from the fact that

$$\lim_{p \to 1} \frac{1 - 4p(1-p)}{\left(\dfrac{1}{2p} - 1\right)^2} = 1 \,.$$

The application of the Karamata theorem can now be justified exactly as before. In fact if $g(u)$ is defined in terms of $f(u)$ as in § 4, then the details go through exactly word for word. Hence we conclude that

$$\lim_{k \to \infty} E \frac{N_2^{(r)}(k)\binom{2k}{k}4^{-k}}{k^{(r-1)/2}} = \frac{r!}{\Gamma\left(\dfrac{r+1}{2}\right)} \,,$$

hence

$$\lim_{k \to \infty} E \frac{N_2^{(r)}(k)}{(4k)^{r/2}} = \frac{r! \sqrt{\pi}}{2^r \Gamma\left(\dfrac{r+1}{2}\right)} = \Gamma\left(\frac{r}{2} + 1\right)$$

which completes the proof.

6. Final remarks. The asymptotic distribution of $N_1(n)$ has been studied by N. V. Smirnov in "Sur les écarts de la courbe de distribution empirique", Mat. Sbornik, 6 (48), pp. 3–26 (1939), (Russian, French summary). His methods are not based on the Karamata Tauberian theorem and seem considerably more complicated than those of this paper, though he actually dealt with a more general situation. Also, the referee has kindly pointed out that the random variable $N_2(n)$ is related to a random variable studied by W. Feller in "The number of zeros and of changes of sign in a symmetric random walk", L'Enseignement Mathématique, III, 3, (1957), 229–235.

REFERENCES

1. G. Baxter and M. D. Donsker, *On the distribution of the supremum functional for processes with stationary independent increments*, Trans. Amer. Math. Soc., **85** (1957), 73-87.
2. G. Doetsch, *Handbuch der Laplace-Transformation*, Band I, Basel: Birkhäuser Verlag, 1950.
3. Meyer Dwass, *The distribution of a generalized D_n^+ statistic*, Ann. Math. Stat. 30, December, (1958), 1024-1028.
4. William Feller, *An introduction to probability theory and its applications*, 2nd edition, New York: John Wiley and Sons, 1957.
5. D. V. Widder, *The Laplace Transform*, Princeton: Princeton University Press, 1946.
6. E. T. Whittaker, and G. N. Watson, *A Course of Modern Analysis*, 4th ed. Cambridge: Cambridge Press, 1950.

NORTHWESTERN UNIVERSITY

RADIAL DISTRIBUTION AND DEFICIENCIES OF THE VALUES OF A MEROMORPHIC FUNCTION

ALBERT EDREI[1], WOLFGANG H. J. FUCHS[2]
AND SIMON HELLERSTEIN

Introduction. Let $f(z)$ be a meromorphic function. Throughout this note we make the following conventions.

I. $f(0) = 1$; this simplifies the exposition without affecting the generality of the results.

II. We denote by

$$a_1, a_2, a_3, \cdots$$

the sequence of the zeros of $f(z)$ and by

$$b_1, b_2, b_3, \cdots$$

the sequence of its poles.

The moduli of the terms of these two sequences are taken to be nondecreasing and each zero or pole appears as often as indicated by its multiplicity.

III. The standard symbols of Nevanlinna's theory:

$$\log^+, \; m(r, f), \; \log M(r, f), \; n(r, f), \; N(r, f), \; T(r, f), \; \delta(\tau, f)$$

are used systematically; familiarity with their meaning is assumed.

We investigate here the following problem, a special case of which has already been mentioned by two of the authors [**1**; p. 295]:

To find sequences $\{a_\mu\}$, $\{b_\nu\}$ *such that if* $f(z)$ *is a meromorphic function with zeros* $\{a_\mu\}$ *and poles* $\{b_\nu\}$ *(and no other zeros or poles), then*

$$(1) \qquad \delta(0, f) > 0 \; , \qquad \delta(\infty, f) > 0 \; .$$

The results of the present note show that a simple behavior of the arguments of the zeros and poles is almost sufficient to induce the inequalities (1). We prove

THEOREM 1. *Let* $f(z)$ *be a meromorphic function with positive zeros and negative poles.*

Received April 25, 1960.

[1] The research of this author was supported by the United States Air Force, Contract No. AF49(638)-571, monitored by the Office of Scientific Research.

[2] The research of this author was supported by a grant from the National Science Foundation (G 5253).

Assume that

$$\sum_\mu \frac{1}{a_\mu} + \sum_\nu \frac{1}{|b_\nu|} = +\infty , \tag{2}$$

and that

$$\sum_\mu \frac{1}{a_\mu^\xi} + \sum_\nu \frac{1}{|b_\nu|^\xi} < +\infty , \tag{3}$$

for some finite positive value of ξ.
Then

$$\limsup_{r\to\infty} \frac{N\left(r, \dfrac{1}{f}\right) + N(r,f)}{T(r,f)} \leqq \frac{1}{1+A} , \tag{4}$$

where $A(>0)$ *is an absolute constant.*
If the condition (3) *is omitted, we still have*

$$\limsup_{r\to\infty} \frac{N\left(r, \dfrac{1}{f}\right) + N(r,f)}{T(r,f)} \leqq 1 . \tag{5}$$

COROLLARY 1.1. *The assumptions of Theorem* 1 *imply*

$$\delta(0,f) \geqq \frac{A}{1+A} , \qquad \delta(\infty,f) \geqq \frac{A}{1+A} .$$

If the condition (3) *is omitted, but*

$$0 < \alpha \leqq \liminf_{r\to\infty} \frac{N(r,f)}{N\left(r, \dfrac{1}{f}\right)} \leqq \limsup_{r\to\infty} \frac{N(r,f)}{N\left(r, \dfrac{1}{f}\right)} \leqq \frac{1}{\beta} < +\infty ,$$

we still have

$$\delta(0,f) \geqq \frac{\alpha}{1+\alpha} , \qquad \delta(\infty,f) \geqq \frac{\beta}{1+\beta} .$$

COROLLARY 1.2. *Let* $f(z)$ *be an entire function with real zeros. If*

$$\sum_\mu \frac{1}{|a_\mu|^2} = +\infty , \tag{6}$$

and if

$$\sum_\mu \frac{1}{|a_\mu|^\xi} < +\infty , \tag{7}$$

for some finite positive value of ξ, *then*

$$\delta(0, f) \geqq \frac{A}{1+A}, \tag{8}$$

where A *is the absolute constant in* (4).

The condition (2) of Theorem 1 cannot be omitted; we shall see that the theorem does not hold for certain meromorphic functions of finite order, with positive poles and such that

$$\sum_{\mu} \frac{1}{a_{\mu}^{\kappa}} + \sum_{\nu} \frac{1}{|b_{\nu}|^{\kappa}} = +\infty,$$

for every κ less than one.

Similarly, Corollary 1.2 does not hold for certain entire functions of finite order, with real zeros and such that

$$\sum_{\mu} \frac{1}{|a_{\mu}|^{\kappa}} = +\infty,$$

for every κ less than two.

The conditions (3) and (7) are used essentially in our proofs, but it is possible that our results hold without such restrictions. This conjecture is plausible if we observe that the assertions (4) and (8) do not contain the parameter ξ.

Our method gives a little more than has been stated. In the special case of entire functions it yields

THEOREM 2. *Let* $f(z)$ *be entire. Assume that all its zeros* a_{μ} *lie on the radii defined by*

$$re^{i\omega_0}, re^{i\omega_1}, \cdots re^{i\omega_m} \quad (r > 0),$$

where the ω*'s are real.*

Then, there exists a positive constant K, *depending only on the* ω*'s, and such that the condition*

$$\sum_{\mu} \frac{1}{|a_{\mu}|^{K}} = +\infty, \tag{9}$$

and the condition

$$\sum_{\mu} \frac{1}{|a_{\mu}|^{\xi}} < +\infty, \tag{10}$$

for some finite value of ξ, *imply*

$$\delta(0, f) \geqq \frac{A}{1+A}, \tag{11}$$

where A is the absolute constant in (4).

All the previous theorems and corollaries assert that 0 and ∞ are among the deficient values of certain functions $f(z)$.

Hence, by Theorem 4 of [1], the lower order μ, of $f(z)$ is positive[3].

Assume now that $h(z)$ denotes a meromorphic function which does not vanish identically, is of order less than μ, but is otherwise arbitrary. Then, by elementary inequalities of Nevanlinna's theory,

$$T(r, hf) \backsim T(r, f) ,$$

$$\frac{m(r, fh)}{T(r, fh)} = \frac{m(r, f)}{T(r, f)} + o(1), \frac{m\left(r, \frac{1}{fh}\right)}{T(r, fh)} = \frac{m\left(r, \frac{1}{f}\right)}{T(r, f)} + o(1) ,$$

and hence

$$\delta(0, fh) = \delta(0, f) , \qquad \delta(\infty, fh) = \delta(\infty, f) .$$

This shows that our theorems remain true even if infinitely many zeros and poles have unknown arguments but are sufficiently rare.

It will be shown in [2] that a radial distribution of zeros and poles makes it, in general, impossible for the function to have other deficient values than 0 and ∞. Combining the results of [2] with those of the present investigation, it is possible to obtain information concerning *all* the deficient values of certain interesting classes of functions. The following result is one of the simplest which may be obtained in this way.

Let $f(z)$ be an entire function of finite order λ. Assume that all the zeros of $f(z)$ are real and that $\lambda > 2$.

Then (11) *holds and*

$$\delta(\tau, f) = 0 ,$$

for $\tau \neq 0$, $\tau \neq \infty$.

1. Consequences of an identity of Nevanlinna.

LEMMA 1. *Let $f(z)$ be meromorphic with zeros $\{a_\mu\}$ and poles $\{b_\nu\}$. Assume*

$$|\arg a_\mu| \leqq \gamma < \frac{\pi}{2} \qquad (\mu = 1, 2, 3, \cdots); \tag{1.1}$$

$$|\arg b_\nu - \pi| \leqq \gamma < \frac{\pi}{2} \qquad (\nu = 1, 2, 3, \cdots); \tag{1.2}$$

[3] A direct study of the lower order of our functions will be found in [2]. For the functions in Theorem 1 and its Corollaries, this study yields "best possible" bounds for μ.

$$\sum_{\mu} \frac{1}{|a_\mu|} + \sum_{\nu} \frac{1}{|b_\nu|} = +\infty . \tag{1.3}$$

Then, for r large enough,

$$\frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\cos\theta d\theta \geqq \cos\gamma\left\{N\left(r,\frac{1}{f}\right) + N(r,f)\right\} . \tag{1.4}$$

Proof. Put $q = 0$, $z = 0$ in a well-known identity of *R.* Nevanlinna [3; p. 222]. Adapting the formula to our notation, we obtain

$$\frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\, e^{-i\theta}d\theta = \frac{r}{2}\left\{\sum_{|a_\mu|\leqq r}\left(\frac{1}{a_\mu} - \frac{\overline{a_\mu}}{r^2}\right)\right. \tag{1.5}$$
$$\left. - \sum_{|b_\nu|\leqq r}\left(\frac{1}{b_\nu} - \frac{\overline{b_\nu}}{r^2}\right)\right\} + \frac{f'(0)}{f(0)}\frac{r}{2} ,$$

and hence, in view of the assumptions (1.1) and (1.2)

$$\frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\cos\theta d\theta \geqq \frac{r}{2}\cos\gamma\left\{\sum_{|a_\mu|\leqq r}\left(\frac{1}{|a_\mu|} - \frac{|a_\mu|}{r^2}\right)\right. \tag{1.6}$$
$$\left. + \sum_{|b_\nu|\leqq r}\left(\frac{1}{|b_\nu|} - \frac{|b_\nu|}{r^2}\right)\right\} - \left|\frac{f'(0)}{f(0)}\right|\frac{r}{2} .$$

An elementary evaluation yields

$$\frac{r}{2}\sum_{|a_\mu|\leqq r}\left(\frac{1}{|a_\mu|} - \frac{|a_\mu|}{r^2}\right) = N\left(r,\frac{1}{f}\right) + \frac{r}{2}\int_0^r N\left(x,\frac{1}{f}\right)\left(\frac{1}{x^2} - \frac{1}{r^2}\right)dx . \tag{1.7}$$

Using (1.7) (and the analogous formula for poles) in (1.6), we obtain

$$\frac{1}{2\pi}\int_0^{2\pi} \log|f(re^{i\theta})|\cos\theta d\theta \geqq \cos\gamma\left\{N\left(r,\frac{1}{f}\right) + N(r,f)\right\} \tag{1.8}$$
$$+ \frac{r}{2}\left\{\cos\gamma\int_0^r\left\{N\left(x,\frac{1}{f}\right) + N(x,f)\right\}\left(\frac{1}{x^2} - \frac{1}{r^2}\right)dx - \left|\frac{f'(0)}{f(0)}\right|\right\} .$$

If r is large enough, this implies (1.4) since, by our assumption (1.3), the integral in the right-hand side of (1.8) tends to $+\infty$ as $r \to +\infty$.

2. Lower bounds for $m(r, f)$.

LEMMA 2. *Let $g(z)$ be an absolutely convergent product of primary factors of genus 2.*

Assume that the zeros of $g(z)$ lie in the sector $\Delta(\varepsilon)$ defined by

$$|\arg z| \leqq \frac{\pi}{6} - \varepsilon \qquad \left(0 < \varepsilon \leqq \frac{\pi}{6}\right) . \tag{2.1}$$

Then

$$\int_{(\pi/3)-(\varepsilon/2)}^{(\pi/3)+(\varepsilon/2)} \log^{+}\left|\frac{g(re^{i\theta})}{g(-re^{i\theta})}\right| d\theta \geqq 2\varepsilon \sin\frac{\varepsilon}{2} r^{3}\int_{0}^{\infty}\frac{n\left(t, \frac{1}{g}\right)}{t^{2}(t^{2}+r^{2})}dt . \tag{2.2}$$

Proof. Let

$$E(u, q) = (1-u)\exp\left(u + \frac{u^{2}}{2} + \cdots + \frac{u^{q}}{q}\right),$$

denote the primary factor of genus q; we write $E(u)$ instead of $E(u, 2)$.
It follows from the definitions that

$$\log\frac{E(u)}{E(-u)} = \text{lon}\left(\frac{1-u}{1+u}\right) + 2u = 2\int_{0}^{u}\frac{t^{2}}{t^{2}-1}dt ,$$

and hence, if

$$u = re^{i\phi} , \qquad \phi \not\equiv 0 \ (\text{mod } \pi) ,$$

$$\log\left|\frac{E(re^{i\phi})}{E(-re^{i\phi})}\right| = 2\int_{0}^{r}\frac{x^{4}\cos\phi - x^{2}\cos 3\phi}{x^{4}-2x^{2}\cos 2\phi + 1}dx . \tag{2.3}$$

Let $\{c_{\nu}\}$ be the sequence of the zeros of $g(z)$; putting

$$\theta_{\nu} = \arg c_{\nu} ,$$

we have, by assumption

$$|\theta_{\nu}| \leqq \frac{\pi}{6} - \varepsilon . \tag{2.4}$$

If $z(= re^{i\theta})$ is confined to the sector

$$\left|\theta - \frac{\pi}{3}\right| \leqq \frac{\varepsilon}{2} , \tag{2.5}$$

(2.3), (2.4) and (2.5) yield

$$\log\left|\frac{E\left(\frac{z}{c_{\nu}}\right)}{E\left(\frac{-z}{c_{\nu}}\right)}\right| = 2\int_{0}^{r/|c_{\nu}|}\frac{x^{4}\cos(\theta-\theta_{\nu}) - x^{2}\cos 3(\theta-\theta_{\nu})}{x^{4}-2x^{2}\cos 2(\theta-\theta_{\nu})+1}dx$$

$$\geqq 2\sin\frac{\varepsilon}{2}\int_{0}^{r/|c_{\nu}|}\frac{x^{2}}{1+x^{2}}dx .$$

Hence, in the region defined by (2.5)

$$\log^+\left|\frac{g(re^{i\theta})}{g(-re^{i\theta})}\right| \geqq 2\sin\frac{\varepsilon}{2}\sum_{\nu=1}^{\infty}\int_0^{r/|c_\nu|}\frac{x^2}{1+x^2}\,dx \tag{2.6}$$

$$= 2\sin\frac{\varepsilon}{2}\,r^3\int_0^{\infty}\frac{n\left(t,\dfrac{1}{g}\right)}{t^2(t^2+r^2)}\,dt$$

and this clearly implies (2.2).

Lemma 3. *Let $f(z)$ be a meromorphic function of genus not greater than* 2.

Assume

(i) *that its zeros $\{a_\mu\}$ lie in the region $\Delta(\varepsilon)$ defined by* (2.1);

(ii) *that its poles $\{b_\nu\}$ lie in the region $\Delta^*(\varepsilon)$ defined by*

$$|\arg z - \pi| \leqq \frac{\pi}{6} - \varepsilon \qquad \left(0 < \varepsilon \leqq \frac{\pi}{6}\right);$$

(iii) $\displaystyle\sum_\mu \frac{1}{|a_\mu|} + \sum_\nu \frac{1}{|b_\nu|} = +\infty$.

Then

$$\frac{1}{2\pi}\int_{(\pi/3)-(\varepsilon/2)}^{(\pi/3)+(\varepsilon/2)} \log^+\left|\frac{f(re^{i\theta})}{f(-re^{i\theta})}\right| d\theta \tag{2.7}$$

$$\geqq \frac{(1-\eta(r))}{2\pi}\,\varepsilon\sin\frac{\varepsilon}{2}\left\{N\left(r,\frac{1}{f}\right) + N(r,f)\right\},$$

where $\eta(r) \to 0$ as $r \to +\infty$.

Proof. Since the genus of $f(z)$ does not exceed 2, it is possible to represent the function by

$$f(z) = e^{P(z)}\frac{\Pi E\left(\dfrac{z}{a_\mu}, 2\right)}{\Pi E\left(\dfrac{z}{b_\nu}, 2\right)}, \tag{2.8}$$

where the polynomial $P(z)$ is of degree not greater than 2 [it is obvious that the infinite products in (2.8) are not necessarily canonical].

Clearly

$$\frac{f(z)}{f(-z)} = e^{2P'(0)z}\frac{g(z)}{g(-z)}, \tag{2.9}$$

where

$$(2.10) \qquad g(z) = \Pi E\left(\frac{z}{a_\mu}, 2\right) \Pi E\left(-\frac{z}{b_\nu}, 2\right).$$

By (2.9)

$$\log^+ \left|\frac{f(re^{i\theta})}{f(-re^{i\theta})}\right| \geqq \log^+ \left|\frac{g(re^{i\theta})}{g(-re^{i\theta})}\right| - 2\,|\,P'(0)\,|\,r\,,$$

and the assumptions (i) and (ii) of Lemma 3 enable us to apply Lemma 2 to the function defined by (2.10). We thus obtain

$$(2.11) \qquad \int_{(\pi/3)-(\varepsilon/2)}^{(\pi/3)+(\varepsilon/2)} \log^+ \left|\frac{f(re^{i\theta})}{f(-re^{i\theta})}\right| d\theta$$

$$\geqq 2\varepsilon \sin\frac{\varepsilon}{2}\, r^3 \int_0^\infty \frac{n\left(t, \frac{1}{f}\right) + n(t, f)}{t^2(t^2 + r^2)}\, dt - 2\varepsilon\,|\,P'(0)\,|\,r\,.$$

Now

$$r^3 \int_0^\infty \frac{n\left(t, \frac{1}{f}\right) + n(t, f)}{t^2(t^2 + r^2)}\, dt > \frac{r}{2} \int_0^r \frac{n\left(t, \frac{1}{f}\right) + n(t, f)}{t^2}\, dt\,,$$

and by assumption (iii) the latter integral tends to $+\infty$ as $r \to +\infty$. Hence (2.11) yields

$$(2.12) \qquad \frac{1}{2\pi} \int_{(\pi/3)-(\varepsilon/2)}^{(\pi/3)+(\varepsilon/2)} \log^+ \left|\frac{f(re^{i\theta})}{f(-re^{i\theta})}\right| d\theta$$

$$\geqq (1 - \eta(r))\, \frac{\varepsilon}{\pi} \sin\frac{\varepsilon}{2}\, r^3 \int_0^\infty \frac{n(t)}{t^2(t^2 + r^2)}\, dt\,,$$

where

$$n(t) = n\left(t, \frac{1}{f}\right) + n(t, f)\,,$$

and $\eta(r) \to 0$ as $r \to +\infty$.

Putting

$$N(t) = \int_0^t \frac{n(x)}{x}\, dx\,,$$

an integration by parts and obvious estimates yield

$$\int_0^\infty \frac{n(t)}{t^2(t^2 + r^2)}\, dt = \int_0^\infty N(t)\, d\left\{-\frac{1}{t(t^2 + r^2)}\right\}$$

$$\geqq N(r) \int_r^\infty d\left\{-\frac{1}{t(t^2 + r^2)}\right\} = \frac{N(r)}{2r^3}\,.$$

Using the latter estimate in (2.12), we obtain (2.7).

LEMMA 4. *If, in Lemma 3, we restrict the value of the parameter* ε *by the inequalities*

$$\frac{9}{10}\frac{\pi}{6} \leqq \varepsilon \leqq \frac{\pi}{6}\,, \tag{2.13}$$

then, for all sufficiently large values of r,

$$T(r,f) \geqq (1+A)\Big\{N\Big(r,\frac{1}{f}\Big)+N(r,f)\Big\}\,, \tag{2.14}$$

where $A(>0)$ *is an absolute constant.*

The inequality (2.14) *still holds if* $f(z)$ *is replaced by* $F(z)$:

$$F(z) = e^{S(z)}f(z) \tag{2.15}$$

where $S(z)$ *is an entire function (which may reduce to a polynomial).*

Proof. We apply Lemma 1 to the function $f(z)/f(-z)$ (instead of $f(z)$). By (2.13) and the definition of $\Delta(\varepsilon)$ and $\Delta^*(\varepsilon)$, we obtain, for large values of r,

$$\frac{1}{2\pi}\int_0^{2\pi}\log\left|\frac{f(re^{i\theta})}{f(-re^{i\theta})}\right|\cos\theta d\theta \geqq \cos\Big(\frac{\pi}{60}\Big)\Big\{2N\Big(r,\frac{1}{f}\Big)+2N(r,f)\Big\}\,.$$

Hence, in view of the trivial relation

$$\frac{1}{2\pi}\int_0^{2\pi}\log\left|\frac{f(-re^{i\theta})}{f(re^{i\theta})}\right|d\theta = 0$$

we find, for r large enough,

$$\frac{1}{2\pi}\int_0^{2\pi}\log\left|\frac{f(-re^{i\theta})}{f(re^{i\theta})}\right|(1-\cos\theta)d\theta \geqq 2\cos\frac{\pi}{60}\Big\{N\Big(r,\frac{1}{f}\Big)+N(r,f)\Big\}\,,$$

$$2m\Big(r,\frac{f(-z)}{f(z)}\Big) \geqq \frac{1}{2\pi}\int_0^{2\pi}\log^+\left|\frac{f(re^{i\theta})}{f(-re^{i\theta})}\right|(1-\cos\theta)d\theta$$
$$+\,2\cos\frac{\pi}{60}\Big\{N\Big(r,\frac{1}{f}\Big)+N(r,f)\Big\}\,,$$

$$2m\Big(r,\frac{f(-z)}{f(z)}\Big) \geqq \Big(1-\cos\frac{\pi}{4}\Big)\frac{1}{2\pi}\int_{(\pi/3)-(\varepsilon/2)}^{(\pi/3)+(\varepsilon/2)}\log^+\left|\frac{f(re^{i\theta})}{f(-re^{i\theta})}\right|d\theta \tag{2.16}$$
$$+\,2\cos\frac{\pi}{60}\Big\{N\Big(r,\frac{1}{f}\Big)+N(r,f)\Big\}\,.$$

Using (2.7) and inequalities for the means of Nevanlinna, (2.16) yields

$$m(r, f(z)) + m\left(r, \frac{1}{f(z)}\right) \geqq \left\{(1 - \eta(r))\frac{\left(1 - \cos\frac{\pi}{4}\right)}{4\pi} \varepsilon \sin\frac{\varepsilon}{2} + \cos\frac{\pi}{60}\right\}$$
$$\times \left\{N\left(r, \frac{1}{f}\right) + N(r, f)\right\},$$

and hence, by Jensen's formula,

$$2T(r, f) \geqq \left(1 + \cos\frac{\pi}{60} + \frac{(1 - \eta(r))\left(1 - \cos\frac{\pi}{4}\right) \varepsilon \sin\frac{\varepsilon}{2}}{4\pi}\right) \tag{2.17}$$
$$\times \left\{N\left(r, \frac{1}{f}\right) + N(r, f)\right\} \; (f(0) = 1).$$

Using (2.13), it is easy to obtain an explicit numerical bound for the coefficient of $N(r, 1/f) + N(r, f)$ in (2.17). Since this bound exceeds 2, we obtain (2.14).

In order to see that (2.14) holds if $f(z)$ is replaced by $F(z)$, we observe that

$$m(r, e^{S(z)}) \leqq T(r, F(z)) + T(r, f(z)) \qquad (f(0) = 1). \tag{2.18}$$

Now

$$T(r, f) = o(r^3) \qquad (r \to +\infty), \tag{2.19}$$

because, by assumption, $f(z)$ is of genus not greater than 2 [3; p. 235].

If $S(z)$ is a polynomial of degree not greater than 2, there is nothing to prove since $F(z)$ is still of genus not greater than 2. In all other cases

$$Xr^3 \leqq m(r, e^{S(z)}), \tag{2.20}$$

for some $X(> 0)$ and r sufficiently large. Hence we obtain the last assertion of the lemma by combining (2.14), (2.18), (2.19), and (2.20).

3. Proof of Theorem 1.

Inequality (5) of Theorem 1 follows readily from Lemma 1 and Jensen's theorem: with $\gamma = 0$, (1.4) yields

$$m(r, f) + m\left(r, \frac{1}{f}\right) \geqq N\left(r, \frac{1}{f}\right) + N(r, f),$$

$$2T(r, f) \geqq 2\left\{N\left(r, \frac{1}{f}\right) + N(r, f)\right\} \qquad (f(0) = 1),$$

which obviously implies (5).

The first part of Theorem 1 is contained in the following Lemma 5 which we now state and prove.

LEMMA 5. *Let $f(z)$ be meromorphic. Assume that there exists an integer $q(\geqq 1)$ such that*

$$\sum_{\mu} \frac{1}{|a_{\mu}|^{q}} + \sum_{\nu} \frac{1}{|b_{\nu}|^{q}} = +\infty . \tag{3.1}$$

$$\sum_{\mu} \frac{1}{|a_{\mu}|^{q+1}} + \sum_{\nu} \frac{1}{|b_{\nu}|^{q+1}} < +\infty . \tag{3.2}$$

Let p be an odd integer

$$1 \leqq p \leqq q . \tag{3.3}$$

Consider the sectors Δ_k defined by

$$\left| \arg z - \frac{2\pi k}{p} \right| \leqq \frac{\pi}{60q} \qquad (k = 0, 1, 2, \cdots, p-1). \tag{3.4}$$

and the sectors Δ_k^* defined by

$$\left| \arg z - \pi - \frac{2\pi k'}{p} \right| \leqq \frac{\pi}{60q} \quad (k' = 0, 1, 2, \cdots, p-1). \tag{3.5}$$

If every zero of $f(z)$ lies in one of the sectors Δ_k and every pole in one of the sectors $\Delta_{k'}^*$, then

$$\limsup_{r \to \infty} \frac{N\left(r, \dfrac{1}{f}\right) + N(r, f)}{T(r, f)} \leqq \frac{1}{1 + A}, \tag{3.6}$$

where A is the absolute constant in Lemma 4.

Proof. Consider the odd integer s defined by

$$s \leqq \frac{q}{p} < s + 2 ; \tag{3.7}$$

in view of (3.3)

$$1 \leqq s .$$

Put

$$l = ps , \qquad \omega = \exp\left(\frac{2\pi i}{l}\right) . \tag{3.8}$$

Clearly l is a positive odd integer and, by (3.7)

$$l \leqq q < l + 2p \leqq 3l \,. \tag{3.9}$$

In view of (3.1) and (3.2), the function $f(z)$ is of the form

$$f(z) = e^{S(z)} \frac{\Pi E\left(\dfrac{z}{a_\mu}, q\right)}{\Pi E\left(\dfrac{z}{b_\nu}, q\right)},$$

where $S(z)$ is entire.

Consider now the auxiliary function

$$G(z) = f(z)f(\omega z) \cdots f(\omega^{l-1}z) = e^{R(z^l)} \frac{\Pi E\left(\dfrac{z^l}{a_\mu^l}, \left[\dfrac{q}{l}\right]\right)}{\Pi E\left(\dfrac{z^l}{b_\nu^l}, \left[\dfrac{q}{l}\right]\right)}, \tag{3.10}$$

where $R(z)$ is entire and the genus $[q/l]$ of the primary factors is, by (3.9), either 1 or 2.

Putting

$$\phi_\mu = \arg a_\mu \,, \qquad \psi_\nu = \arg b_\nu \,,$$

our inequalities (3.4), (3.5), and (3.7) imply

$$|\, \phi_\mu l - 2\pi ks \,| \leqq \frac{\pi}{60}\,, \qquad |\, \psi_\nu l - \pi l - 2\pi k's \,| \leqq \frac{\pi}{60}\,. \tag{3.11}$$

We also notice that our assumptions prevent the possibility of cancellation between the zeros of one of the functions $f(\omega^j z)$ $(j = 0, 1, \cdots, l-1)$ and the poles of another of these functions. Hence

$$N(r, G(z)) = lN(r, f)\,, \qquad N\left(r, \frac{1}{G(z)}\right) = lN\left(r, \frac{1}{f}\right). \tag{3.12}$$

Put

$$H(u) = e^{R(u)} \frac{\Pi E\left(\dfrac{u}{a_\mu^l}, \left[\dfrac{q}{l}\right]\right)}{\Pi E\left(\dfrac{u}{b_\nu^l}, \left[\dfrac{q}{l}\right]\right)},$$

and rewrite (3.10) as

$$G(z) = H(z^l)\,. \tag{3.13}$$

The inequalities (3.11), the assumption (3.1), and the first of the inequalities (3.9) show that it is possible to apply Lemma 4 to $H(u)$ (instead of $f(z)$). Hence

$$T(r, H(u)) \geqq (1 + A)\left\{N\left(r, \frac{1}{H(u)}\right) + N(r, H(u))\right\} \quad (r \geqq r_0) . \tag{3.14}$$

On the other hand, the fundamental definitions of Nevanlinna's theory show that, for any meromorphic function $w(z)$:

$$N(r, w(z^l)) = N(r^l, w(z)) , \qquad T(r, w(z^l)) = T(r^l, w(z)) ,$$

so that (3.13) and (3.14) yield

$$T(r, G(z)) \geqq (1 + A)\left\{N\left(r, \frac{1}{G(z)}\right) + N(r, G(z))\right\} \quad (r^l \geqq r_0) . \tag{3.15}$$

Since

$$l\,T(r, f) \geqq T(r, G(z)) ,$$

we see that (3.6) follows from (3.12) and (3.15).

We obtain the first part of Theorem 1 by taking $p = 1$ in Lemma 5.

4. Proof of the Corollaries. Corollary 1.1 follows trivially from the inequalities (4) and (5) and the definition of deficiency.

Corollary 1.2 is contained in the following.

LEMMA 6. *Let $f(z)$ be entire. Modify the assumptions of Lemma 5 by:*

(i) *omitting all reference to poles;*

(ii) *omitting the restriction that p be odd (p may be any integer satisfying the inequality* (3.3)).

Then (3.6) *still holds.*

The proof of Lemma 5 also yields Lemma 6 provided the integer s (even or odd) is defined by

$$s \leqq \frac{q}{p} < s + 1 ,$$

instead of (3.7). The definitions (3.8) remain unchanged and (3.9) takes the sharper form

$$l \leqq q < 2l .$$

The other changes in the proof are obvious and need not be mentioned here.

We obtain Corollary 1.2 by taking $p = 2$, in Lemma 6.

5. Best possible character of the conditions (2) and (6).
Let

$$s_1, s_2, s_3, \cdots \tag{5.1}$$

be a sequence of integers such that

$$s_1 \geqq 2, \quad s_{\lambda+1} > 2s_\lambda \qquad (\lambda = 1, 2, 3, \cdots). \tag{5.2}$$

Consider the entire function

$$f(z) = \prod_{\lambda=1}^{\infty} \prod_{m=s_\lambda}^{2s_\lambda} \left(1 - \frac{z}{m(\log m)^2}\right).$$

Denoting by $\{a_\mu\}$ the sequence of the zeros of $f(z)$, elementary estimates yield

$$\sum_\mu \frac{1}{a_\mu} < +\infty, \quad \sum_\mu \frac{1}{a_\mu^\kappa} = +\infty \qquad (\kappa < 1). \tag{5.3}$$

These relations hold for every choice of the sequence (5.1). Hence we may take the ratios $s_{\lambda+1}/s_\lambda$ to be rapidly increasing with λ and, using the well-known formula [4; p. 271]:

$$\log M(r, f) = r\int_0^\infty \frac{n\left(t, \dfrac{1}{f}\right)}{t(t+r)}\, dt,$$

choose (5.1) so that

$$\liminf_{r\to\infty} \frac{\log T(r, f)}{\log r} \leqq \liminf_{r\to\infty} \frac{\log\log M(r)}{\log r} = 0. \tag{5.4}$$

It is sufficient to choose the sequence (5.1) in such a way that, for some arbitrarily large u, $n(t, 1/f)$ is constant in $u \leqq t \leqq e^u$.
Hence, putting

$$F(z) = \frac{f(z)}{f(-z)},$$

$$\liminf_{r\to\infty} \frac{\log T(r, F(z))}{\log r} = 0. \tag{5.5}$$

It has been shown elsewhere [1; p. 297, Theorem 4] that the condition (5.5) implies

$$\delta(\tau, F(z)) = 0,$$

except possibly for a single value of τ, finite or infinite.

Hence the inequalities

$$\delta(0, F(z)) > 0 , \qquad \delta(\infty, F(z)) > 0 ,$$

are both impossible since one of them would imply the other one. We thus have

$$1 = \limsup_{r\to\infty} \frac{N(r, F)}{T(r, F)} = \frac{1}{2} \limsup_{r\to\infty} \frac{N\left(r, \frac{1}{F}\right) + N(r, F)}{T(r, F)} ,$$

although $F(z)$ satisfies all the conditions of Theorem 1 except (2) which is replaced by the weaker condition (5.3).

Similarly, (5.4) and Theorem 4 of [1] yield

$$\delta(\tau, f(z)) = 0 \qquad (\tau \neq \infty)$$

and hence, putting

$$F^*(z) = f(z^2)$$

we have

$$\delta(\tau, F^*(z)) = 0 \qquad (\tau \neq \infty).$$

In particular $\delta(0, F^*(z)) = 0$, although $F^*(z)$ satisfies all the conditions of Corollary 1.2 except (6) which is replaced by a weaker condition analogous to (5.3).

6. Proof of Theorem 2. Our proof is a straightforward consequence of Lemma 6 and of a classical theorem of H. Weyl [5; p. 335, Satz 16].

We consider the arguments ω_j of the radii carrying the zeros of $f(z)$ and assume

$$\omega_0 = 0 ;$$

this is clearly no restriction.

Let $k + 1$ $(0 \leqq k \leqq m)$ be the maximum number of linearly independent elements among

$$2\pi, \omega_1, \omega_2, \cdots \omega_m . \tag{6.1}$$

Renumbering, if necessary, the ω's we may assume:

(i) that a relation such as

$$\mu_0 2\pi + \sum_{j=1}^{k} \mu_j \omega_j = 0 , \tag{6.2}$$

is impossible for integral values of the μ's, not all zero;

(ii) if $k < m$, there exist integers n_{ij} and $\sigma(> 0)$ such that

$$\sigma\omega_l = 2\pi n_{l0} + \sum_{j=1}^{k} n_{l_j}\omega_j \qquad (l = k+1, \cdots, m). \tag{6.3}$$

Put

$$M_l = \sum_{j=1}^{k} | n_{l_j} |$$

and

$$M = \sup \{\sigma, M_{k+1}, M_{k+2}, \cdots, M_m\} . \tag{6.4}$$

Since no relation such as (6.2) is possible, Weyl's theorem asserts the existence of a sequence

$$\lambda_1, \lambda_2, \lambda_3, \cdots \tag{6.5}$$

of increasing integers such that

$$| \lambda_s\omega_j - L_{sj}2\pi | \leqq \frac{\pi}{120M} \qquad (j = 1, 2, \cdots k; s = 1, 2, 3, \cdots), \tag{6.6}$$

where the L_{sj} are integers. Weyl's theorem also asserts that the sequence (6.5) has a positive density. The latter property is unnecessarily precise for our purposes; we only need the obvious implication

$$\lambda_{s+1} < 2\lambda_s \qquad (s \geqq s_0). \tag{6.7}$$

We set

$$K = \sigma\lambda_{s_0}$$

and observe that the integer K depends only on the ω's.

By the assumptions (9) and (10), there exists an integer q such that

$$q \geqq K , \qquad \sum_{\mu} \frac{1}{| a_\mu |^q} = +\infty , \qquad \sum_{\mu} \frac{1}{| a_\mu |^{q+1}} < +\infty .$$

Define h by the inequalities

$$\sigma\lambda_h \leqq q < \sigma\lambda_{h+1} . \tag{6.8}$$

In view of the definition of K and (6.7)

$$q < 2\sigma\lambda_h . \tag{6.9}$$

We now obtain Theorem 2 by verifying that Lemma 6 may be applied with the value of q chosen above and

$$p = \sigma\lambda_h . \tag{6.10}$$

It is clear that we only have to ascertain that the zeros of $f(z)$ lie in regions such as (3.4) with p defined by (6.10).

Using (6.6) and (6.4) in (6.3), we obtain

$$|\sigma\lambda_h\omega_l - \Lambda_{hl}2\pi| \leqq \frac{\pi}{120} \qquad (l = k+1, k+2, \cdots m), \tag{6.11}$$

where the Λ's are integers.

By (6.6) and (6.4), it is clear that (6.11) holds also for $l = 1, 2, \cdots k$, with

$$\Lambda_{hl} = \sigma L_{hl} \qquad (l = 1, 2, \cdots k).$$

Hence, by (6.9), (6.10) and (6.11)

$$\left|\omega_l - \frac{\Lambda_{hl}2\pi}{p}\right| \leqq \frac{\pi}{60q} \qquad (l = 1, 2, \cdots m).$$

This shows that the location of zeros allows the application of Lemma 6. Theorem 2 is an immediate consequence.

References

1. A. Edrei and W. H. J. Fuchs, *On the growth of meromorphic functions with several deficient values,* Trans. Amer. Math. Soc. **93** (1959), 292-328.
2. ——— and ———, *On the maximum number of deficient values of certain classes of functions,* Air Force Technical Note, AFOSR TN 60-402 (April 1960).
3. R. Nevanlinna, *Eindeutige Analytische Funktionen,* 2nd ed., Berlin, 1953.
4. E. C. Titchmarsh, *The theory of functions 2d ed.,* Oxford, 1950.
5. H. Weyl, *Über die Gleichverteilung von Zahlen mod. Eins,* Math. Annalen **77** (1916), 313-352.

Syracuse University and Cornell University

HARMONIC FUNCTIONS WITH ARBITRARY LOCAL SINGULARITIES

WILLIAM C. FOX

1. Introduction. This paper is concerned with a new and more informative solution of an old existance problem, that of determining what conditions must be imposed upon the nature of a local harmonic singularity in order to imply the global existence of a harmonic function which "has" the given singularity. In 1870, [17, vol. II, p. 133–143 and p. 144–177], H. A. Schwarz solved the problem for closed surfaces giving, as sufficient, the condition that the harmonic singularity function must have vanishing flux across the curve (bounday of a disk) on which it is given, and he solved the problem for open surfaces which are interiors of compact manifolds-with-boundaries, with no restriction on the singularity function. In 1909 [4, vol. 3, p. 73–80] Hilbert announced that the problem for open surfaces, with singularities having flux, can be solved by a special extremal method. Hilbert worked out an illustrative example and left the general account to be presented in the thesis of his student Richard Courant. A few months later Koebe [6], in the last of his series of four papers on the uniformization of analytic curves, gave the first full account of the existence of harmonic functions with a prescribed local singularity on open surfaces. Koebe based his proofs on exhaustion and the results of Schwarz; he did not use Hilbert's special extremal method. Moreover, his convergence arguments still used the assumption that the singularity's flux is zero. In 1910 and 1912 [2, 3] Courant published accounts of special cases taken from his thesis; not again did Hilbert's special extremal method appear in print. In 1913, Weyl [20] re-proved Koebe's theorem using an extremal method, namely that of minimizing the Dirichlet integral of what he called the "concurrence functions." (In all these works the singularity function was specified in concrete terms, e.g., as the real part of $1/z^n$ near the origin. However, the proofs remain valid for any singularity with vanishing flux. Accordingly, I have described them in those terms.)

Not until 1953 were any further advances made with respect to this existence problem. At that time, Sario [13] published a modern account (based on preliminary notes dated 1949 and 1950) of the alternating series method of Schwarz which went far beyond the work of Schwarz both in method and in generality. When Sario's results are

Received March 15, 1960. Research which led to this paper was supported in part by the National Science Foundation. A preliminary report was made at the 1957 summer meeting of the American Mathematical Society.

restricted to the case of a *local* singularity, they duplicate those of Schwarz for closed surfaces and in the case of open surfaces they relieve Koebe's theorem of the vanishing flux restriction on the singularity. Sario states further that when the flux *does* vanish one may conclude that the function asserted to exist is bounded and has finite Dirichlet integral on any domain on whose closure it is harmonic.

In this paper, Sario's results are sharpened in various ways. Among others: a necessary and sufficient condition for the flux to vanish is that a potential function exists which "has" the prescribed singularity and whose normal derivatives vanish (in a certain strong sense) on the ideal boundary. On open surfaces, there always exists a potential function which "has" the prescribed singularity and which vanishes on the ideal boundary. These conditions (on the functions' behavior at the ideal boundary) determine the two potential functions uniquely (up to additive constants) as solutions to certain extremal problems. Concerning either of these two potential functions, one may *always* state that it is bounded and of finite Dirichlet integral on any domain on whose closure it is harmonic, even when the singularity's flux is not zero. Moreover, the extremal properties shed some light on the role of Sario's assumption that the given singularity function, harmonic on certain Jordan curves, vanishes on these curves.

An alternative existence proof is given here also. Its preliminary part (Theorem 1) on uniform boundedness is of some intrinsic interest. Although it parallels Sario's Lemma 3 [13, p. 636], it was suggested by similar arguments used by Koebe in 1909 [6] and in his 1910 recapitulations [7, 8]. Mainly, however, the alternative existence proof is given here because it also yields the other results described above.

2. Uniform boundedness. The existence theorem of the next section makes use of the Ascoli theorem. For this purpose some information is needed on the existence of a uniform bound for certain families of potential functions. In what follows, the term "local coordinate" refers to any homeomorphism (*from* a domain in the sphere *onto* a domain in the Riemann surface in question) which is also analytic.

THEOREM 1. *In the Riemann surface X, let $B_I \subset B_0$ be the images under a local coordinate of concentric open disks, and let S be harmonic on the closed "annulus" $(B_0 - B_I)^-$. If $\mathscr{U}$ is a set of functions, u, with each of which is associated a domain $D(u)$ in X containing $(B_0)^-$ such that*

(i) *each function u is harmonic on $D(u) - B_I$ and is bounded there by its extreme values on the boundary of B_I,*

(ii) *each function $u - S$ determines a harmonic function U on $(B_0)^-$ which agrees with $u - S$ on $B_0 - B_I$, and*

(iii) *for some point* Q *of* B_I *the set of values* $\{U(Q) : u \in \mathscr{U}\}$ *is bounded,*

then there exists a constant K *such that* $|u| \leqq K$ *on* $D(u) - B_I$ *for every* u *in* $\mathscr{U}$.

Proof. Let $M(u)$ denote the maximum value of $|u|$ on the boundary of the inner disk B_I, by hypotheses, the relation $|u| \leqq M(u)$ holds on $D(u) - B_I$ and there then must exist a point $p(u)$ in the boundary of B_I at which $|u(p) = M(u)$. If the set $\{M(u) : u \in \mathscr{U}$ is bounded above then the theorem is proved. Otherwise, there exists a sequence $\{u_n\}$ in $\mathscr{U}$ such that (2.1) $\lim_n M(u_n) = +\infty$ and (2.2) $\lim_n p(u_n) - p$. On the outer disk B_0, the function U_n (which agrees with $u_n - S$ on $B_0 - B_I$) is harmonic and so is bounded there by its extreme values on the boundary of B_0, so that

$$|U_n| \leqq \operatorname{lub}\{|u_n(x)| : x \in \partial B_0\} + \operatorname{lub}\{|S(x)| : x \in \partial B_0\}. \tag{2.3}$$

If the right hand term in this inequality is abbreviated by the symbol M, then relation (2.3), by hypotheses, may be written in the form

$$|U_n| \leqq M(u_n) + M \text{ on } B_0. \tag{2.4}$$

Thus

$$\frac{|U_n - U_n(Q)|}{M(u_n)} \leqq 2\left(1 + \frac{M}{M(u_n)}\right) \text{ on } B_0. \tag{2.5}$$

By (2.1), this implies that the sequence $(U_n - U_n(Q))/M(u_n)$ is uniformly bounded on B_0. Moreover, the sequence $u_n/M(u_u)$ is uniformly bounded on the ring $(B_0 - B_I)^-$. The Ascoli theorem guarantees the existence of a subsequence of indeces for which the following limits exist uniformly on the domains indicated:

$$\lim_n \frac{U_n - U_n(Q)}{M(u_n)} = H_0, \text{ harmonic on } B_0, \tag{2.6}$$

$$\lim_n \frac{u_n}{M(u_n)} = H, \text{ harmonic on } B_0 - (B_I^-), \text{ continuous on } (B_0 - B_I)^-. \tag{2.7}$$

By hypotheses the sequence of numbers $\{U_n(Q)\}$ is bounded. Thus by (2.1), one may conclude that

$$\lim_n \frac{S}{M(u_n)} = 0 \text{ and } \lim_n \frac{U_n(Q)}{M(u_n)} = 0, \tag{2.8}$$

whence

$$H_0 = \lim_n \frac{u_n}{M(u_n)} - \lim_n \frac{S}{M(u_n)} - \lim_n \frac{U_n(Q)}{M(u_n)} = H \text{ on } B_0 - B_I. \tag{2.9}$$

It is now necessary to show that $H_0(p) = 1$ when p is given by (2.2). Let p_n denote $p(u_n)$ and let $a_{nj} = |u_n(p_j)/M(u_n)|$. In view of the facts.

$$(2.10) \qquad \begin{cases} \lim_n a_{nj} = |H(p_j)| \text{ uniformly in } j, \text{ and} \\ \lim_j a_{nj} = |u_n(p)/M(u_n)| \text{ for each } n , \end{cases}$$

one may apply the Moore-Smith theorem on iterated limits. It follows that the double limit $\lim_{n,j} a_{nj}$ exists and equals the common value of the two iterated limits:

$$(2.11) \qquad \lim_n (\lim_j a_{nj}) = \lim_j (\lim_n a_{nj}) = H_0(p) .$$

(The last equality follows from (2.8) and (2.9), since the points p_n are all in $B_0 - B_I$, for that set contains the boundary of B_I.) Since $a_{nn} = 1$ for all n (by definition of p_n), it *is* the case that $|H_0(p)| = 1$. Since $H_0(Q) = 0$, this means that H_0 is not constant on B_0.

On the other hand, H_0 must be constant, in virtue of the following considerations. Restricted to B_I, H_0 certainly is bounded by its extreme values on the boundary of B_I. Restricted to the ring $B_0 - B_I$, it is still true that H_0 is bounded by its extreme values on the boundary of B_I because H_0 inherits this property from the sequence $u_n/M(u_n)$ which by (2.8) and (2.9) converges to H_0 on $B_0 - B_I$. This means that H_0 must have a local maximum at some point of the boundary of B_I, necessarily an interior point of B_0, whence H_0 is constant. This contradiction completes the proof.

3. The existence theorem. The method of exhaustion requires that the existence problem be solved on subdomains with compact closure and smooth boundaries. For the present purposes this was started by H. A. Schwarz in 1870 and was completed by Koebe in 1910, [8]. A portion of the proof is sketched here to indicate how the hypotheses enter the arguments.

If r and s are C' on the interior of Y except for a closed set of measure zero then the Dirichlet integral of r relative to s over Y will be denoted by $\mathfrak{D}(r, s; Y)$. When $r = s$, the Dirichlet integral will be written simply $\mathfrak{D}(r; Y)$. When Y is a plane domain, one has

$$\mathfrak{D}(r, s; Y) = \int_Y \left(\frac{\partial r}{\partial x} \frac{\partial s}{\partial x} + \frac{\partial r}{\partial y} \frac{\partial s}{\partial y} \right) d\mu$$

where μ is Lebesgue measure in Y.

LEMMA. *Let X^- be a compact two-manifold-with-boundary and let its interior, $X \neq X^-$, be a Riemann surface. If $B_I \subset B_0$ are images* *under a local coordinate in X of concentric open disks and if S is harmonic on the closed "annulus" $(B_0 - B_I)^- \subset X$, then there exists a function u, and when S has a single-valued conjugate, there exists another function v, such that*

(i) *they are harmonic on $X - B_I$,*

(ii) *their differences with S have harmonic extensions from $B_0 - B_I$ to B_0,*

(iii) *they are bounded on $X - B_0$ by their extreme values on the boundary of B_0,*

(iv) $\mathfrak{D}(v, T; X - B_0) = +\int_{\partial B_0} T dv^*$ *for every function T, C' on $X - B_0$; and*

(v) $\mathfrak{D}(u, H; X - B_0) = +\int_{\partial B_0} u dH^*$ *for every function H harmonic on $X - B_0$.*

Proof. By definition of a two-manifold-with-boundary, every boundary point of X^- is contained in a Jordan curve one of whose complementary domains in X is simply connected. By the Osgood-Caratheodory extension of the Riemann mapping function, there exists a homeomorphism of this domain's closure onto the closed unit disk, a homeomorphism which is analytic on the domain itself and which sends an arc of the boundary of X onto an arc of the unit circle. Thus, it is no restriction to suppose that each component of the boundary of X is an *analytic* Jordan curve, which in turn makes it possible to form the "double" of X^-. (Of course, when X^- is contained in some larger Riemann surface Y this argument does not imply that the boundary of X is *analytic in* Y, but only analytic relative to the coordinate system of X itself.)

Let X^* denote the double of X^-; there is then an analytic homeomorphism, a reflection, R of X^* onto itself which leaves the boundary of X pointwise fixed and which sends X onto $X^* - X^-$. Let $B_0^* = R(B_0)$, $B_I^* = R(B_I)$ and $S^* = S(R^{-1})$. Both S and S^*, by hypothesis, have single-valued harmonic conjugates on the annuluses $B_0 - B_I$ and $B_0^* - B_I$ respectively. Since X^* is a closed surface, this condition is needed to warrant the conclusion that a function v^*, harmonic on $X^* - (B_I \cup B_I^*)$, exists such that $v^* - S$ and $v^* - S^*$ have harmonic extensions to B_0 and B_0^* respectively. Moreover these conditions make v^* unique up to an additive constant since the only functions harmonic on a closed surface are the constants. Therefore $v^*(R) - v^*$ is not only constant but is zero because R leaves the (non-empty) boundary of X pointwise fixed. This implies that v^* is bounded on $X^* - (B_0 \cup B_0^*)$ by its extreme values on the boundary of B_0 alone, for they are related by the reflection R to those on the boundary of B_0^*. Of course the normal derivatives of

v^* vanish on the boundary of X, also because of the relation $v^*(R)-v^*=0$ and the fact that, on neighborhoods of points of the boundary of X, R is literally a reflection. Thus, if v denotes the restriction of v^* to $X^- - B_I$ then (i), (ii) and (iii) have been proved, and (iv) is a consequence of the Green's formula. This use of the "double" of X is due to Koebe [8].

To construct u, one applies the existence theorem of Schwarz (for open surfaces) with boundary values being the constant zero. Properties (i) and (ii) are then immediate; (v) is a consequence of the Green's formula and the boundary values of u, and the latter implies (iii) also.

It is now possible to prove the main theorem.

THEOREM 2, PART I: *Existence. Let X be a Riemann surface, let B and β be images under a local coordinate in X of an open disk and its boundary respectively, and let S be harmonic on β. A necessary and sufficient condition that $\int_\beta dS^* = 0$ is that there exists a real function v*

(i) *which is a potential function, on X, whose singularity is S, (i.e., which is harmonic on $X-B$ and whose difference with S has a harmonic extension to B^-).*

(ii) *which is bounded on $X-B$ by its extreme values on β,*

and

(iii) *for which $dv^*=0$ on ∂X, i.e.,*

$$\int_{\partial X} wdv^* = \mathfrak{D}(w, v; X-B) - \int_\beta wdv^* = 0$$

for every function w, C' on $X-B^-$ and continuous on the closure.

If S is an arbitrary harmonic function on β and if X is open, there always exists a potential function, u, on X, whose singularity is S, which has property (ii) *and*

(iv) *for which $u=0$ on ∂X, i.e.,*

$$\int_{\partial X} udw^* = \mathfrak{D}(u, w; X-B) - \int_\beta udw^* = 0$$

for every function w harmonic on $X-B$.

THEOREM 2, PART II: *Uniqueness: When v exists there then also exists a function r, harmonic on B^-, for which $d(S-r)^*=0$ on β; whether or not v exists, there always exists a function s, harmonic on B^-, for which $S-s=0$ on β. The functions v and u are determined uniquely up to an additive constant, among all potential functions on*

X whose singularities are S, by their respective properties (iii) *and* (iv) *as the functions for which*:

(v) *The quantity* $\int_\beta (S - r)dw^* + \int_{\partial X} dw^*$ *is minimized by* $w = v$ *among all functions* w, *harmonic on* $X - B$.

(vi) *The quantity* $\int_\beta wd(S - s)^* + \int_{\partial X} wdw^*$ *is minimized by* $w=u$ *among all functions* w *harmonic on* $X - B$.

In each case the minimum values are $\int_\beta (S - r)dv^*$ *and* $\int_\beta ud(S - s)^*$ *respectively.*

Proof. Let $B = B_0$ and let B_I be concentric with B_0 such that S is harmonic on $(B_0 - B_I)^-$. Let $\{D_n\}$ be an exhaustion of X such that $(B_0)^- \subset D_I$, i.e., $\{D_n\}$ is a sequence of domains, the union of which is X, each with compact closure and with boundary consisting of sectionally analytic Jordan curves. By the lemma, there exists a potential function u_n, on D_n whose singularity is S, which vanishes continuously on the boundary of D_n, and is bounded, by its extreme values on β, on $D_n - B_0$. Let Q is a point in B_I and let a_n be the value taken at Q by the harmonic extension to B_0 of $u_n - S$. If $\{a_n\}$ is an unbounded sequence let $\{u_n\}$ be replaced by $\{u_n - a_n\}$. By Theorem 1 the sequence $\{u_n\}$ is uniformly bounded on each set $D_n - B_I$ and so contains a subsequence which converges uniformly on each set $D_n - B_I$ to a function u which then inherits properties (i) and (ii).

NOTE. When each u_n is the Green's function of D_n with pole at Q then a_n is called the principle part of u_n and the sequence $\{a_n\}$ is necessarily monotone increasing. By use of Harnack's theorem one sees that $\{a_n\}$ is bounded above if and only if u is the Green's function for X. This characterization of the Green's function's existence (the convergence of a sequence of principle parts) was first discovered by Koebe in his proof of the so-called uniformization theorem [9], when X is simply connected.

By (v) in the lemma, one may establish the following relations, once u_n has been extended continuously to $X - D_n$ to be constant there:

$$
(3.1) \qquad \begin{cases} \mathfrak{D}(u, H; X - B_0) = \lim_n \mathfrak{D}(u_n, H; X - B_0) \\ \qquad = \lim_n \mathfrak{D}(u_n, H; D - B_0) \\ \qquad = \lim_n \int_\beta u_n dH^* = \int_\beta udH^*. \end{cases}
$$

This establishes (iv) here. Note that the additive constants a_n, should they be present, do not have any effect, for they disappear in the integrand of the Dirichlet integral.

If S does have a single-valued harmonic conjugate, i.e., if

$$\int dS^* = 0$$

then the corresponding functions v_n certainly exist, by the lemma, and so therefore does v by Theorem 1. Moreover

$$(3.2)\qquad \begin{cases} \mathfrak{D}(v, T; X - B_0) = \lim_n \mathfrak{D}(v_n, T; X - B_0) \\ \qquad\qquad = \lim_n \mathfrak{D}(v_n, T; D - B_0) \\ \qquad\qquad = \lim_n \int_\beta T dv_n{}^* = \int_\beta T dv^*, \end{cases}$$

as required by (iii). In (3.1) and (3.2), one needs to know that the partial derivatives of u_n and v_n will converge to those of u and v. Conversely, given (iii), one may choose $w = 1$ and obtain

$$\int_\beta dv^* = 0.$$

Since $v - S$ has a harmonic extension to $(B_0)^-$ and therefore has a single-valued harmonic conjugate there, it follows that

$$\int_\beta d(v - S)^* = 0,$$

whence

$$\int_\beta dS^* = 0$$

It is worth observing that the choices $T = v$ and $H = u$ lead to the conclusions

$$(3.3)\qquad \begin{cases} \mathfrak{D}(u; X - B_I) = \int_\beta u du^*, \text{ and} \\ \mathfrak{D}(v; X - B_I) = \int_\beta v dv^*. \end{cases}$$

Thus both u and v have finite Dirichlet integrals over $X - B_0$. Moreover, according to (iii) v has the property that $\mathfrak{D}(v, T; X - B_0)$ is finite even for functions T such that $\mathfrak{D}(T; X - B_0)$ is *not* finite. A similar remark holds for u.

Using the notation

$$(3.4)\qquad \mathfrak{D}(w; X - B_0) = \int_\beta w dw^* + \int_{\partial X} w dw^*$$

for harmonic w, with the above observations, one may verify that

$$(3.5)\qquad \int_{\partial X} u du^* = 0 \text{ and } \int_{\partial X} v dv^* = 0,$$

facts which will be used below.

Let t be a function harmonic on B^-. To prove part II one must establish that the quantities to be minimized in (v) and (vi) exhibit quadratic-form properties. For this purpose it is convenient to introduce the notion of an $(S-t)$-concurrence function as a function W, C' on $X-\beta$ for which

$$W+(S-t)\chi_{X-B^-}$$

is continuous on a neighborhood of β. (Of course, χ_Z denotes the characteristic function of Z.) If w belongs to the class $\mathscr{C}$ of all functions C' on $X-B^-$ and continuous on the closure, then w determines an $(S-t)$-concurrence function, denoted by W_{S-t}, by the rule that $W_{S-t}=w$ on $X-B^-$ and W_{S-t} is, on B^-, the harmonic function determined by the boundary values $w-(S-t)$. Thus, every W_{S-t} is harmonic on B^-.

If y is a potential function on X whose singularity is S and w is an arbitrary member of $\mathscr{C}$ then

$$(3.6)\quad \begin{cases} \mathfrak{D}(Y_{S-t};\, Y_{S-t}-W_{S-t};\, B) = -\int_\beta (y-w)d(y-(S-t))^*, \\ \mathfrak{D}(Y_{S-t};\, Y_{S-t}-W_{S-t};\, X-B) = \int_\beta (y-w)dy^* + \int_{\partial X}(y-w)dy^*, \text{ and} \\ \mathfrak{D}(Y_{S-t},\, Y_{X-t}-W_{S-t};\, X) = \int_\beta (y-w)d(S-t)^* + \int_{\partial X}(y-w)dy^*. \end{cases}$$

If w is a member of the class $\mathscr{H}$ of all functions harmonic on $X-B$, then

$$(3.7)\quad \begin{cases} \mathfrak{D}(Y_{S-t},\, Y_{S-t}-W_{X-t};\, B) = -\int_\beta (y-(S-t))d(y-w)^*, \\ \mathfrak{D}(Y_{S-t},\, Y_{S-t}-W_{S-t};\, X-B) = \int_\beta yd(y-w)^* + \int_{\partial X} yd(y-w)^*, \text{ and} \\ \mathfrak{D}(Y_{S-t},\, Y_{X-t}-W_{S-t};\, X) = \int_\beta (S-t)d(y-w)^* + \int_{\partial X} yd(y-w)^*. \end{cases}$$

These relations are consequences of the Green's formula and the facts that every W_{S-t} takes, on β, the values $w-(S-t)$ and w according as one approaches β from B or from $X-B^-$ and the difference of two $(S-t)$-concurrence functions is C' on all of X. Therefore

(3.8) The Dirichlet-variation $\mathfrak{D}(Y_{S-t},\, Y_{S-t}-W_{S-t};\, X)$ vanishes in each of the following cases:
(a) when $y=v$ and $w-y=0$ on β and on ∂X,
(b) when $y=v$ and $t=r$
(c) when $y=u$ and $d(w-y)^*=0$ on β and on ∂X, and
(d) when $y=u$ and $t=s$.

Cases (a) and (c) are immediate, whereas (b) and (d) are consequence of (iii) and (iv), and the properties of r and of s.

The quadratic form character of the Dirichlet integral makes it easy to verify that

$$\begin{cases} \mathfrak{D}(a, a-b; Z) = 0 \text{ if and only if} \\ \mathfrak{D}(b; Z) - \mathfrak{D}(a; Z) = \mathfrak{D}(b-a; Z). \end{cases} \tag{3.9}$$

Since $\mathfrak{D}(b - a; Z)$ is non-negative the vanishing of the Dirichlet-variations in (3.8) are equivalent, respectively, with the following:

$$\begin{cases} \text{(a)}' & \mathfrak{D}(V_{S-t}; X) \leqq \mathfrak{D}(W_{S-t}; X) \text{ for all } w \text{ in } \mathscr{C} \text{ for which } w-v=0 \\ & \text{on } \beta \text{ and on } \partial X, \\ \text{(b)}' & \mathfrak{D}(V_{S-r}; X) \leqq \mathfrak{D}(W_{S-r}; X) \text{ for all } w \text{ in } \mathscr{C} \\ \text{(c)}' & \mathfrak{D}(U_{S-t}; X) \leqq \mathfrak{D}(W_{S-t}; X) \text{ for all } w \text{ in } \mathscr{H} \text{ for which} \\ & d(w-u)^* = 0 \text{ on } \beta \text{ and on } \partial X, \text{ and} \\ \text{(d)}' & \mathfrak{D}(U_{S-t}; X) \leqq \mathfrak{D}(W_{S-s}; X) \text{ for all } w \text{ in } \mathscr{H}. \end{cases} \tag{3.10}$$

In each of these cases, $\mathfrak{D}(b-a; Z) = 0$ implies that $b - a$ is constant, which establishes the uniqueness claims. The extremal properties stated in (v) and (vi) are derived from the inequalities (b)′, and (d)′ by expanding both sides. In general, when w is in $\mathscr{H}$,

$$\begin{cases} \mathfrak{D}(W_{S-t}; X) = \mathfrak{D}(W_{S-t}; B) + (W_{S-t}, X - B) \\ \qquad = -\int_\beta (w-(S-t))d(w-(S-t))^* + \int_\beta w dw^* + \int_{\partial X} w dw^* \\ \qquad = +\int_\beta w d(S-t) + \int_\beta (S-t) dw^* - \int_\beta (S-t) d(S-t)^* \\ \qquad\quad + \int_{\partial X} w dw^*. \end{cases} \tag{3.11}$$

In (a)′, the first and third terms, above, are common to both sides; in (b)′ those terms both vanish since $d(S - r)^* = 0$ on β. In (c)′ the second and third terms, above, are common to both sides, and in (d)′ these become zero because $S - s = 0$ on β.

The extremal properties of (a)′ and (c)′ are of no interest because any two of the harmonic functions involved must differ by a constant. Whereas v and u were shown to solve (b)′ and (d)′ by use of (iii) and (iv), they were shown to solve (a)′ and (c)′ *automatically*. Hence any other of the competing functions also solves (a)′ or (c)′, whence the competing functions all differ from u (or v) by constants.

It remains only to verify the existence of r and s. By the Poisson formula there is a function s, harmonic on B_0, given by the boundary values S. Since $S - s = 0$ on the boundary, β, of B_0, it may be continued across the the boundary by reflection. Therefore, s is harmonic on $(B_0)^-$ because S is harmonic on $(B_0 - B_I)^-$. When S is the real part of a complex function $S + iT$ analytic on $(B_0)^-$ then in a like manner one may construct a function t, harmonic on $(B_0)^-$, which agrees with T on the boundary. Since B_0 is a disk, t is the imaginary part of a complex function $r+it$ analytic on $(B_0)^-$. Since $T - t = 0$ on the boundary of B_0 and is conjugate to $S-r$, it follows that $d(S-r)^*=0$ on the boundary of B_0 also, as required to complete the proof.

The evaluation of $\mathfrak{D}(W_{S-t}; X)$ in (3.11) relies upon the assumption that w is harmonic on $X - B$ so that the Green's formula may be used there. This restriction makes it possible to phrase (v) in a compact form, though it does cause that statement to be incomplete. However, the supply of functions; harmonic on $X - B$ is sufficiently ample to make it possible for the existence of v to be proved by a direct approach to the extremal problem (v) rather than by exhaustion. (Indeed, except for the restriction to harmonic functions, this is exactly how Weyl's existence proof was accomplished, for he minimized the Dirichlet integral of all $(S - r)$-concurrence functions.) This fact makes one suspect that it may be possible to prove the existence of u by a direct approach to the extremal problem (vi). I will discuss this possibility in another paper.

From (3.9) and the expansion in (3.6) it is clear that

$$\mathfrak{D}(y, X - B) \geqq \mathfrak{D}(w; X - B) = \int_\beta w dw^* + \int_{\partial X} w dw^* \tag{3.12}$$

for every w harmonic on $X-B$ for which $w=y$ on β and on ∂x. Sario [14, p. 354] has discovered that, when $y = v$, the requirement "$w = y$ on ∂X" may be replaced by the requirement "$\int_\beta dw^* = 0$" which is of course equivalent with "$\int_{\partial X} dw^* = 0$", and (3.12) continues to hold. He uses this fact to show that the existence of v implies $\int_\beta dS^* = 0$, whereas in the present discussion property (iii) is used for this purpose as well as to characterize v uniquely up to an additive constant. The function v is Sario's "principal function P^0_{-1}." By (iii), when $y=v$, the relation "$w = y$ on ∂X" holds for *every* C' function w, so that Sario's condition "$\int_\beta dw^* = 0$" is not necessary, for the extremal property (3.12) itself. The fact that (3.12) holds for all w which agree with v on β regardless of their behavior "at infinity" was known to Hilbert in 1909 [4, vol. 3, p. 78].

The existence of v was announced first by Hilbert [4, vol. 3, pp. 73–80] and fully proved first by Koebe [6]. Its extremal property (v) was discovered and proved by Weyl [20]. The existence of u with property (i) was proved first by Sario [13]. Properties (ii), (iv) and (v) for u are new, as well as their unique determination of u.

The results given here may be generalized along the lines of Sario's linear operators [14, 15] by consideration of more general domains for the singularity function S and of extremal properties involving other combinations of the quantities appearing in (v) and (vi). Alternatively, once the existence of potential functions with an arbitrary local singularity has been settled (as in the present theorem 2) one may build a sequence of potential functions each with local singularity in one of a sequence of "localities" and then combine them with coefficients which force convergence of the resulting series. Such a technique was first proposed by Koebe [10] in his proof that every open surface is conformally equivalent with a continuation manifold (needed to fill a gap in his first proof of the so-called uniformization theorem); a more detailed version was given by Stoilow [18, p. 59–60].

BIBLIOGRAPHY

1. Constantin Caratheodory, *Gesammelte Mathematische Schriften*, C. H. Beck, 1955.
2. Richard Courant, *Zur Begrundung des Dirichletschen Prinzipes*, Göttingen Nachrichten, (1910), 154-160.
3. ———, *Uber die Methode des Dirichletschen Prinzipes*, Mathematische Annalen, **72** (1912), 517-550.
4. David Hilbert, *Gesammelte Abhandlungen*, Springer, 1935.
5. O. D. Kellogg, *Foundations of Potential Theory*, Frederick Ungar, 1929.
6. R. Koebe, *Uber die Uniformisierung beliebieger analytischer Kurven*, Göttingen Nachrichten, (1909), 324-361.
7. ———, *Uber die Uniformisierung beliebieger analytischer Kurven*, Journal fur die reine und angewandte Mathematik, **138** (1910), 192-253.
8. ———, *Uber die Hilbertsche Uniformisierungsmethode*. Göttingen Nachrichten, 1910, 59-74.
9. ——— *Uber die Uniformisierung beliebieger analytischer Kurven*, Göttingen Nachrichten, (1907), 191-210.
10. ———, *Fonction potentialle et fonction analytique ayant un domaine d'existence donné à un nombre quelconque (fini ou infini) de feuillets*, Comptes Rendue, **148** (1909), 1446.
11. C. Neumann, Berichten der K. Sachs. Gesellschaft der Wissenschaft, April 21 and October 31, 1870.
12. ———, *Vorlesungen uber Riemanns Theorie*, 2nd Edition, Leipzig, 1884.
13. Leo Sario, *Alternating method on arbitrary Riemann surfaces*, Pacific J. Math., **3** 1953), 631-645.
14. ———, *Minimizing Operators on Subregions*, Proc. Amer. Math. Soc., **4** (1953), 350-355.
15. ———, *A linear operator method on arbitrary Riemann surfaces*, Trans. Amer. Math. Soc. **72**, (1952), 281-295.
16. ———, *Construction of functions with prescribed properties on Riemann surfaces* Ann. Math. Studies, **30** (1953), 63-76.
17. H. A. Schwarz, *Gesammelte Mathematische Abhandlungen*, Springer, 1890.
18. S. Stoilow, *Leçons sur les Principes topologiques de la Théorie des Fonctions Analytiques*, Gauthier-Villars, 1938.
19. Karl Weierstrass, *Mathematische Werke von Karl Weierstrass*, Berlin, 1894.
20. Hermann Weyl, *Die Idee der Riemannschen Flache*, Teubner, 1913.

MANIFOLDS WITH POSITIVE CURVATURE

THEODORE FRANKEL

0. Introduction and a conjecture. In 1936 J. L. Synge [10] proved that an even dimensional orientable compact manifold M_n with positive sectional curvature is simply connected. His proof was an application of a formula for the second variation of arc length derived by him in an earlier article.[1] In the present paper we continue the study of positively curved manifolds again using the ideas of Synge and applying them to an only slightly different situation, namely to the "position" of certain submanifolds of M.

Theorem 1 states that two compact totally geodesic (see § 2 for definitions) submanifolds V_r and W_s of M_n must necessarily intersect if their dimension sum is at least that of M, i.e. if $r + s \geqq n$. As remarked above the proof is a straightforward continuation of Synge's method. Unfortunately totally geodesic submanifolds are not a too common occurrence.

If M_n is a Kähler manifold[2] the situation is much more satisfactory. There, instead of requiring V and W to be totally geodesic, we need only ask that they be complex analytic submanifolds (Theorem 2).

Examples of compact Riemannian manifolds of positive sectional curvature are the spheres, the real, complex and quaternionic projective spaces and the Cayley plane. Rauch [8] has shown that if the sectional curvatures do not differ too much from that of the sphere and if the space is simply connected, then it is itself topologically a sphere (see also the recent improvements by W. Klingenberg, Über kompakte Riemannsche Mannigfaltigkeiten, Math. Ann., **137** (1959), pp. 351–61). Berger [2] has shown that if M_n is an even dimensional, simply connected manifold and if the sectional curvature K satisfies $1/4 \leqq K \leqq 1$, then the manifold is one of the spaces listed above.

In the list the only Kähler manifolds are the complex projective n-spaces $P_n(\boldsymbol{C})$ with the usual Fubini metric. If (e_1, e_2) is a pair of orthonormal tangent vectors to $P_n(\boldsymbol{C})$, then the sectional curvature $K(e_1, e_1)$ satisfies $1/4 \leqq K(e_1, e_2) \leqq 1$ with $K = 1$ if and only if the plane $e_1 \wedge e_2$ is a "complex direction." It may very well be that

CONJECTURE. *The positively curved Kähler manifolds of complex dimension n are analytically homeomorphic to* $P_n(\boldsymbol{C})$. The Gauss Bonnet

Received January 11, 1960. Work supported in part by the National Science Foundation.

1 For completeness we include in § 1 a derivation of the second variation formula.

2 Since the Ricci curvature of a positively curved manifold is positive, the Kähler manifold is a "Hodge manifold" and Kodaira's theorem [6] states that the manifold is algebraic.

theorem shows that this is true for $n = 1$. Using Theorem 2, A. Andreotti has shown that the conjecture is true for $n = 2$ and his proof is presented in Theorem 3. It relies heavily on the known classification of algebraic surfaces.[2]

Difficulties in attempting to construct counter examples stem from the fact that the product of two positively curved manifolds has only *nonnegative* curvature (in the product metric). If $e_1 \wedge e_2$ is a product plane (e.g., if e_1 is "horizontal" and e_2 is "vertical"), then $K(e_1, e_2) = 0$ and this is the only time 0 curvature can occur. Our results in general do not apply to such spaces.

The last section is devoted to proving the existence of fixed points for certain maps, thus showing further similarities with $P_n(C)$.

I should like to thank A. Andreotti, E. Calabi and N. Hawley for discussions of the results.

1. Second variation of arc length. Our notation is as follows. M_n is a complete n dimensional Riemannian manifold and V_r and W_s are submanifolds of dimension r and s respectively. $\mathscr{C}(t)$ is a geodesic going from $\mathscr{C}(0) = P \in V$ to $\mathscr{C}(l) = Q \in W$ striking V and W orthogonally; t represents arc length along $\mathscr{C}$. X_t is a unit vector field that is displaced parallel along $\mathscr{C}$ and is tangent to V and W at P and Q respectively; X_t (if it exists) is thus orthogonal to $\mathscr{C}$ for all t. Finally T_t is the unit tangent vector to $\mathscr{C}$.

We construct a "variation" of the geodesic $\mathscr{C}$ as follows. We pass a small "ribbon" of surface through $\mathscr{C}$ that is tangent to X_t at $\mathscr{C}(t)$ for all t such that $0 \leqq t \leqq l$. This ribbon cuts V and W in two curves. We now pass curve segments on the ribbon tangent to X_t at $\mathscr{C}(t)$, the curves varying smoothly from V to W. The ribbon is chosen so "thin" that no two segments intersect. On each segment we use the directed arc length α from $\mathscr{C}$ as parameter and we may suppose that $-\varepsilon \leqq \alpha \leqq +\varepsilon$. Each point on the ribbon carries two coordinates (t, α) and we have two systems of coordinate curves $t =$ constant and $\alpha =$ constant (the original geodesic is of course $\alpha = 0$). We have two coordinate vector fields $T = \partial/\partial t$ and $X = \partial/\partial\alpha$ defined on the ribbon with $T = T_t$ at $(t, 0)$ and $X = X_t$ at this same point. The problem is to investigate the lengths of the curves $\alpha =$ constant.

We recall some facts and notation of Riemannian geometry (our notation follows [7]). We let $g(Y, Z)$ denote the Riemannian scalar product of two vectors Y and Z; if $(x_1, \cdots, x_n)$ are local coordinates for M, then $g(Y, Z) = \sum_{ij} g_{ij} Y^i Z^j$. If Y is a vector at a point and if f is a function, then the covariant derivative of f with respect to Y, written $\nabla_Y(f)$, is the directional derivative of f in the direction Y. If Z is a vector field, the covariant derivative of Z with respect to Y is again a vector, written $\nabla_Y Z$. If Y is also a vector field, the Lie or commutator bracket of Y and Z is given by $[Y, Z] = YZ - ZY = \nabla_Y Z - \nabla_Z Y$. In particular, if Y and Z are coordinate vectors $\nabla_Y Z - \nabla_Z Y = [Y, Z] = 0$. Hence in the case of our particular vectors we have

$$\nabla_X T = \nabla_T X . \tag{1}$$

Next we have the Ricci operator identity

$$\nabla_Y \nabla_Z - \nabla_Z \nabla_Y = R(Y, Z) + \nabla_{[Y,Z]}$$

where $R(Y, Z)$ is, for each pair (Y, Z), a linear transformation on tangent vectors. $R(Y, Z)$ is constructed from the Riemann curvature tensor and in terms of coordinates the transformation of vectors $U \to R(Y, Z)U$ is given by

$$\sum_i U^i \frac{\partial}{\partial x^i} \to \sum_i \Big(\sum_{jkl} - R^i_{jkl} Y^k Z^l U^j \Big) \frac{\partial}{\partial x^i} .$$

$R(Y, Z)$ is skew symmetric; $R(Y, Z) = -R(Z, Y)$. In our case the Ricci identity becomes

$$\nabla_X \nabla_T - \nabla_T \nabla_X = R(X, T) . \tag{2}$$

The Riemannian sectional curvature corresponding to the 2-plane $T \wedge X$ is given by

$$K(T, X) = g(R(X, T)T, X) = -g(R(X, T)X, T) . \tag{3}$$

Finally we recall that the scalar product is "covariant constant," i.e.

$$\frac{\partial}{\partial \alpha} g(Y, Z) = \nabla_X g(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) .$$

The length of the curve $\alpha =$ constant is given by

$$L(\alpha) = \int_0^l g(T, T)^{1/2} dt .$$

LEMMA ([9]). *The first and second variations of arc length are*

$$\begin{cases} L'_X(0) = \left.\dfrac{dL}{d\alpha}\right|_0 = 0 \\ L''_X(0) = \left.\dfrac{d^2L}{d\alpha^2}\right|_0 = g(\nabla_X X, T)_Q - g(\nabla_X X, T)_P - \displaystyle\int_0^l K(T, X) dt . \end{cases}$$

Proof.

$$L'(\alpha) = \int_0^l \frac{\partial}{\partial \alpha} g(T, T)^{1/2} dt = \int_0^l \nabla_X g(T, T)^{1/2} dt ,$$

thus

$$L'(\alpha) = \int_0^l \frac{g(\nabla_X T, T)}{g(T, T)^{1/2}} dt . \tag{4}$$

But $g(T, T) \equiv 1$ along $\alpha = 0$ (T is unit tangent to $\mathscr{C}(t)$) and so from (1) we get

$$L'(0) = \int_0^l g(\nabla_X T, T)dt = \int_0^l g(\nabla_T X, T)dt = 0$$

since $\nabla_T X = 0$ for a parallel displaced X.

For the second variation we continue from (4)

$$L''(\alpha) = \int_0^l \nabla_X \left\{ \frac{g(\nabla_T X, T)}{g(T, T)^{1/2}} \right\} dt$$

which expands to

$$L''(0) = \int_0^l \nabla_X g(\nabla_T X, T)dt - \int_0^l g(\nabla_T X, T)^2 dt .$$

But X is displaced parallel along $\mathscr{C}$; $\nabla_T X = 0$ and so the second integral vanishes. Thus

$$L''(0) = \int_0^l g(\nabla_X \nabla_T X, T)dt + \int_0^l g(\nabla_T X, \nabla_X T)dt$$

but again the second integral vanishes. Using (2) the first term becomes

$$L''(0) = \int_0^l g(\nabla_T \nabla_X X, T)dt + \int_0^l g(R(X, T)X, T)dt .$$

The first integral transforms by means of

$$g(\nabla_T \nabla_X X, T) = \nabla_T g(\nabla_X X, T) - g(\nabla_X X, \nabla_T T) = \frac{\partial}{\partial t} g(\nabla_X X, T)$$

and using (3) we get the desired second variation.

The end terms in the second variation are interpreted as follows. $B_T(X)_P \equiv g(\nabla_X X, T)$ is the second fundamental form for V at P corresponding to the normal vector T, evaluated at the tangent vector X.

2. Real manifolds with positive curvature. A submanifold V of a Riemannian M_n is *totally geodesic* if any geodesic of M that is tangent to V at a point lies wholly in V. This implies that every geodesic of V (in the naturally induced metric from M) is at the same time a geodesic of M.

THEOREM 1. *Let M_n be a complete*[3] *connected manifold with positive Riemannian sectional curvature and let V_r and W_s be compact totally geodesic submanifolds. If $r + s \geqq n$ then V_r and W_s have a non-empty intersection.*

Proof. At first we assume that V_r and W_s are *any* compact submanifolds. We suppose they do not intersect. Then there is a shortest geodesic $\mathscr{C}(t)$, say of length $l > 0$, from V to W and let P and Q be the points $\mathscr{C}(0)$ and $\mathscr{C}(l)$ respectively. Since $\mathscr{C}$ is the shortest geodesic from V to W it strikes V and W orthogonally. We will arrive at a contradiction by exhibiting a variation X for which $L''_X(0) < 0$, thus showing that $\mathscr{C}$ cannot be minimizing.

Let $\mathscr{V}_0$ be the tangent space to V_r at P. By parallel translation along $\mathscr{C}$ we get a subspace $\mathscr{V}_l$ of $\mathscr{M}$, the tangent space to M_n at Q. Since $\mathscr{V}_0$ is orthogonal to $\mathscr{C}$ at P, $\mathscr{V}_l$ is also orthogonal to $\mathscr{C}$ at Q. Let $\mathscr{W}$ be the tangent space to W_s at Q. Then $\mathscr{V}_l$ and $\mathscr{W}$ are two subspaces of the linear space $\mathscr{M}$; moreover, both $\mathscr{V}_l$ and $\mathscr{W}$ are orthogonal to $\mathscr{C}$ at Q. Thus the dimension of their intersection is

$$\dim(\mathscr{V}_l \cap \mathscr{W}) \geqq r + s - (n - 1) \geqq 1 \tag{5}$$

and thus $\mathscr{V}_l$ and $\mathscr{W}$ have at least a one dimensional subspace in common. But this simply means that there is a unit vector X_0 tangent to V at P whose parallel translate is tangent to W at Q. Let X_t be the parallel translate of X_0 along $\mathscr{C}$. The term $-\int_0^l K(T, X)dt$ of the second variation formula is strictly negative by the curvature assumption.

So far V and W were arbitrary. To evaluate the end terms in the second variation we use the fact that V and W are totally geodesic. The variation vector X_t is given. For the construction of the "ribbon" we can choose geodesics of M through each X_t; since X_0 is tangent to V at P and since V is totally geodesic, the geodesic through X_0 will lie entirely in V. Likewise the geodesic through X_l will lie entirely in W. Thus the curves $\alpha =$ constant will have their endpoints on V and W as required for the variation. But since X_0 and X_l are tangent vectors to geodesics of M we have $\nabla_X X = 0$ at P and Q. Hence the end terms of the second variation formula vanish and we have

$$L''_X(0) = -\int_0^l K(T, X)dt < 0 \qquad \text{Q.E.D.}$$

as desired.

We note that $g(\nabla_X X, T)_P = g(\nabla_X X, T)_Q = 0$ is merely the statement that all second fundamental forms for a totally geodesic submanifold vanish identically.

[3] If the curvature is bounded away from 0, $K \geqq \varepsilon > 0$, the classical result of Bonnet-Myers states that M_n is necessarily compact.

There is at least one situation when totally geodesic submanifolds arise "naturally." If $f\colon M_n \to M_n$ is an *isometric* map of a Riemannian manifold into itself, then the set of fixed points $F = \{P \in M \mid f(P) = P\}$ has as components totally geodesic submanifolds (see [4]). Hence

COROLLARY. *If $f\colon M_n \to M_n$ is an isometry of a compact connected Riemannian manifold with positive curvature, then no two fixed set components can have dimension sum $\geqq n$.*

3. Kähler manifolds with positive curvature. A Kähler manifold M is a special type of Riemannian manifold whose underlying space is a complex manifold. There is a linear transformation J on each tangent space that sends any vector Y into a vector JY orthogonal to Y (J represents multiplication by $(-1)^{1/2}$). J has the properties $J^2 = -I$ and $g(JY, JZ) = g(Y, Z)$ for all vectors Y and Z (this last property states that g is a "Hermitian" metric). From J we construct the Kähler exterior 2-form ω, defined by

$$\omega(Y, Z) = g(JY, Z) \, .$$

ω is exterior because $\omega(Y, Z) = -\omega(Z, Y)$. All that has been said so far holds for any Hermitian manifold. The further condition defining a Kähler manifold can be stated as requiring that ω be covariant constant, $\nabla_U \omega = 0$ for all vectors U; i.e., for any vector fields Y and Z we have

$$\nabla_U \omega(Y, Z) = \omega(\nabla_U Y, Z) + \omega(Y, \nabla_U Z) \, .$$

Since g is also covariant constant we conclude that J is also, i.e., we have the operator equation

$$\nabla_U \circ J = J \circ \nabla_U \tag{6}$$

for any vector U.

A linear subspace $\mathscr{V}$ of the tangent space to a complex manifold at a point is said to be *complex* if it is invariant under J, $J\colon \mathscr{V} \to \mathscr{V}$. A submanifold is *complex analytic* if its tangent space at each point is complex.

When dealing with complex manifolds dimension subscripts will denote complex dimension.

The following result is easily true for $P_n(C)$ since it holds for the linear subspaces.

THEOREM 2. *Let M_n be a complete, connected Kähler manifold with positive sectional curvature and let V_r and W_s be compact complex analytic submanifolds. If $r + s \geqq n$, then V_r and W_s we have a non-empty intersection.*

Proof. The proof is again by contradiction, starting exactly as in Theorem 1. We again arrive at a variation vector X_t, parallel displaced along $\mathscr{C}$ and tangent to V and W at P and Q respectively. Now, however, we have additional information. Since V and W are complex analytic the vector field $J(X_t)$ is tangent to V and W at P and Q respectively. Further, from (6) we have $\nabla_T J(X_t) = J\nabla_T X_t = 0$ since X_t is parallel displaced. Thus $J(X_t)$ is also parallel displaced and gives us the same type of variation vector as X_t. We claim

$$\begin{cases}\text{the second variation corresponding to at least one of} \\ \text{the fields } X_t \text{ or } JX_t \text{ is strictly negative}\end{cases}$$

again giving a contradiction.

To prove our claim we suppose

$$L''_X(0) = g(\nabla_X X, T)_Q - g(\nabla_X X, T)_P - \int_0^l K(T, X)dt \geqq 0 . \tag{7}$$

By the hypothesis of positive curvature we conclude that

$$g(\nabla_X X, T)_Q - g(\nabla_X X, T)_P > 0 .$$

We will be finished if we can show $g(\nabla_{JX} JX, T)_Q - g(\nabla_{JX} JX, T)_P < 0$. But this is actually the case as follows from the fact that every second fundamental form of a complex analytic submanifold of a Kähler manifold is skew-hermitian,[4] i.e.

$$\begin{cases} g(\nabla_{JX} JX, T)_P = -g(\nabla_X X, T)_P & \text{for} \quad V \\ g(\nabla_{JX} JX, T)_Q = -g(\nabla_X X, T)_Q & \text{for} \quad W . \end{cases} \tag{8}$$

The proof of this is simple and we include it here for completeness.

Let $\mathscr{R}$ be a complex analytic curve (real dimension 2) on V tangent to X_0 and JX_0 at P. Then X_0 can be extended to a tangent vector field X on $\mathscr{R}$ and of course JX is an extension of JX_0. Since X and JX are tangent vector fields to $\mathscr{R}$ the commutator bracket $[JX, X]$ is again a vector field tangent to $\mathscr{R}$, and thus orthogonal to T at P. Using $[JX, X] = \nabla_{JX} X - \nabla_X JX$ and (6) and $J^2 = -I$ we get at P

$$\begin{aligned} g(\nabla_{JX} JX, T) &= g(J\nabla_{JX} X, T) = g(J[JX, X] + J\nabla_X JX, T) \\ &= g(J[JX, X], T) - g(\nabla_X X, T) . \end{aligned}$$

Since $[JX, X]$ is tangent to $\mathscr{R}$, so is $J[JX, X]$ and so the first term vanishes and the result follows. Q.E.D.

[4] This is a reflection of the fact that Kähler submanifolds of a Kähler manifold are minimal submanifolds in the sense of the calculus of variations. Thus their mean curvatures vanish for all normal directions.

4. Kähler surfaces with positive curvature. We now consider the case of Kähler surfaces M_2 (real dimension 4). We noticed previously[2] that by Kodaira's theorem such a surface is necessarily algebraic.

We recall that an *exceptional curve* (of the first kind) arises in the following fashion. There is a surface N_2 and a point $P \in N_2$ such that M_2 is a quadratic transform [3] of N_2 and the exceptional curve is the quadratic transform of p. Thus exceptional curves result from blowing up a point p of a surface by means of the Hopf σ-process; i.e., the point p is replaced by the complex projective line $P_1(\boldsymbol{C})$ of complex directions at p. Since there clearly are curves that do not intersect the exceptional curve (hyperplane section of N_2 for example) we conclude from Theorem 2 that a positively curved compact Kähler surface has no exceptional curves (of the first kind).

THEOREM 3. *A compact Kähler surface M_2 with positive sectional curvature is complex analytically homeomorphic to $P_2(\boldsymbol{C})$.*

Proof (Andreotti). As mentioned before[2] the Ricci curvature of a positively curved Kähler M_n is positive. The negative of the exterior Ricci form represents the characteristic class of the canonical bundle K over M. By Kodaira's "vanishing theorem" [5] we conclude $H^p(M_n; \Omega^0(K^1)) = 0$, $p \neq n$, where K^i is the line bundle $K \otimes \cdots \otimes K$, i factors and where $\Omega^0(K^i)$ is the sheaf of germs of holomorphic sections of K^i. Thus the plurigenera $P_i = \dim H^0(M_n; \Omega^0(K^i))$ all vanish and since M_2 is simply connected the arithmetic genus $p_a = P_1 - h^{1,0} = 0$ also. We now apply results in the classification theory of *surfaces*, i.e., $n = 2$. By a theorem of Castelnuovo-Enriques (for references see, for example, Zariski's book, *Introduction to the problem of minimal models in the theory of algebraic surfaces*, Math. Soc. of Japan, 1958, p. 84) we conclude that M_2 is rational. As we have just seen M_2 can have no exceptional curves (of the first kind). By a result of Andreotti [1] M_2 is either birationally equivalent, without exceptions, to $P_2(\boldsymbol{C})$ or else it is a ruled surface. Since the rulings would be compact curves that do not intersect, Theorem 2 eliminates this last possibility. Q.E.D.

5. Correspondences. A (holomorphic) correspondence of a complex manifold N_n with itself is a complex analytic n dimensional submanifold of $N_n \times N_n$.

A holomorphic map $f: N_n \to N_n$ gives rise to a correspondence, the graph $G(f)$ of f; $G(f) = \{(p, fp) \mid p \in N_n\}$. $G(f)$ is of course a special type of correspondence since f is single valued. Let $\Delta = \{(p, p) \mid p \in N_n\}$ be the diagonal of $N_n \times N_n$. It is clear that a map f will have a fixed point whenever $G(f)$ intersects the diagonal Δ. A correspondence will be said to have a fixed point if it intersects the diagonal.

THEOREM 4. *Every (holomorphic) correspondence of a connected compact Kähler manifold N_n with positive curvature has a fixed point.*

Proof. Again this is a simple known property of $P_n(\boldsymbol{C})$.

The correspondence is a complex analytic submanifold V_n of $N_n \times N_n$. The same is true for the diagonal Δ. We need only show that V_n and Δ intersect, and this almost follows from Theorem 2. However, as pointed out in the introduction, $N_n \times N_n$ has only nonnegative curvature; product planes give 0 sectional curvature. This, however, is easily mended as follows.

In our previous notation $V_n = V$, $\Delta = W$ and $N_n \times N_n = M$. In the proof of Theorem 2 positive curvature occurs only in the statement $\int_0^l K(T, X)dt > 0$. Now we can only say.

$$\begin{cases} L_X''(0) = (\nabla_X X, T)_Q - (\nabla_X X, T)_P - \int_0^l K(T, X)dt \\ \qquad\qquad\qquad \int K(T, X)dt \geqq 0 \end{cases}$$

Again we suppose $L_X''(0) \geqq 0$.

Case 1. $(\nabla_X X, T)_Q - (\nabla_X X, T)_P > 0$. Then from (8) we $L_{JX}''(0) < 0$ and we are finished.

Case 2. $(\nabla_X X, T)_Q = (\nabla_X X, T)_P$ and $\int_0^l K(T, X)dt = 0$. We will then be finished if we can show $\int_0^l K(T, JX)dt > 0$. Now $\int_0^l K(T, X)dt = 0$ means $T \wedge X$ is a product plane along $\mathscr{C}$, in particular at $Q \in W = \Delta$. Choose a real basis for the tangent space to $N_n \times N_n$ at Q consisting of the $2n$ "horizontal" orthonormal vectors $e_1, Je_1, \cdots, e_n, Je_n$ and the $2n$ "vertical" orthonormal vectors $f_1, Jf_1, \cdots, f_n, Jf_n$. Since $T \wedge X$ is a product plane the basis can be so chosen that

$$\begin{aligned} X &= (\cos\theta)e_1 + (\sin\theta)f_1 \\ T &= -(\sin\theta)e_1 + (\cos\theta)f_1 \,. \end{aligned}$$

Thus

$$JX = (\cos\theta)Je_1 + (\sin\theta)Jf_1 \,.$$

This means that the only possibilities for $T \wedge JX$ to be a product plane are either $\cos\theta = 0$ or $\sin\theta = 0$, i.e., either $T = \pm e_1$ or $T = \pm f_1$. But e_1 and f_1 being respectively horizontal and vertical cannot be orthogonal to the diagonal $W = \Delta$ while the geodesic tangent T must be. We thus conclude that if $T \wedge X$ is a product plane then $T \wedge JX$ *cannot* be. Hence $\int_0^l K(T, JX)dt > 0$. Q.E.D.

The isometries (rotations) of the 3-sphere without fixed points show that there is no real analogue of Theorem 4.

References

1. A. Andreotti, *On the complex structures of a class of simply connected manifolds*, in the Lefschetz symposium volume *Algebraic geometry and topology*, Princeton, 1957.
2. M. Berger, *Les variétés riemanniennes* (1/4)-*pincées*. C. R. Acad. Sci. Paris, **250** (1960), 442–44.
3. H. Hopf, *Zur Topologie der komplexen Mannigfaltigkeiten*, Studies and Essays presented to R. Courant, New York, (1948), 167–85.
4. S. Kobayashi, *Fixed points of isometries*, Nagoya Math. J., **13** (1958) 63–68.
5. K. Kodaira, *On a differential-geometric method in the theory of analytic stacks*, Proc. Nat. Acad. Sci. U. S. A., **39** (1953), 1268–73.
6. ———, *On Kähler varieties of restricted type*, Ann. of Math., **60** (1954), 28–48.
7. K. Nomizu, *Lie groups and differential geometry*, Math, Soc. Japan (1956).
8. H. E. Rauch, *A contribution to differential geometry in the large*, Ann. of Math., **54** (1951), 38–55.
9. J. L. Synge, *The first and second variations of length in Riemannian space*, Proc. London Math. Soc., **25** (1926).
10. ———, *On the connectivity of spaces of positive curvature*, Quart. J. Math. (Oxford series) **7** (1936), 316–20.

Stanford University

A STRONG MAXIMUM PRINCIPLE FOR WEAKLY SUBPARABOLIC FUNCTIONS

AVNER FRIEDMAN

Introduction. It has been proved by E. Hopf [3], over thirty years ago, that solutions of second order elliptic equations satisfy the maximum principle. A similar principle, well known for solutions of the heat equation, has been, relatively recently, extended to second order parabolic equations by Nirenberg [5]. In various problems, such as in solving the Dirichlet problem by the methods of Poincaré and Perron, subsolutions have been introduced and the maximum principle has been extended to such functions. In the elliptic case (see [6]) the subsolutions used are continuous, whereas in the parabolic case, they may have certain discontinuities (see [2]). In the elliptic case, they are called L-subharmonic or subelliptic functions. Likewise, in the parabolic case, we call them L-subcaloric or subparabolic functions; L is the elliptic or the parabolic operator.

Recently, Walter Littman [4] has generalized the concept of L-subharmonic functions to include measurable integrable functions. This generalization is obtained by expressing the condition $Lu \geqq 0$ in an integrated form, namely, $\int uL^*vdx \geqq 0$ for any twice differentiable $v \geqq 0$ with compact support, L^* being the adjoint of L. He then established the maximum principle in the following sense: If an L-subharmonic function assumes its essential supremum at a point of continuity, then it is equal to a constant almost everywhere.

The purpose of this paper is to prove a similar result for measurable L-subcaloric functions. The general outline of the proof is similar to that of Littman's method. However, the crucial step in the proof is the construction of two kernal functions with certain required properties. Our construction is entirely different from that of Littman.

In § 1 we state some definitions and the results of the paper. In § 2 we prove Lemma 2. In § 3 we recall some properties of fundamental solutions. These are used in § 4 to prove Lemma 1. Lemmas 1, 2 immediately yield the maximum principle.

1. Statement of the results.

Consider the differential operators

$$Lu \equiv \sum_{i,j=1}^{n} a_{ij}(x, t)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{n} a_i(x, t)\frac{\partial u}{\partial x_i} + a(x, t)u - \frac{\partial u}{\partial t}$$

Received March 21, 1960. Prepared under Contract Nonr 710(16) (NR 044 004) between the Office of Naval Research and the University of Minnesota.

$$L^*u \equiv \sum_{i,j=1}^{n} b_{ij}(x, t)\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x, t)\frac{\partial u}{\partial x_i} + b(x, t)u + \frac{\partial u}{\partial t}$$

where L^* is the adjoint of L (thus, $b_{ij} = a_{ij}$, etc.). Throughout this paper it will always be assumed that:

$$a_{ij}, \frac{\partial}{\partial x_k} a_{ij}, \frac{\partial^2}{\partial x_k \partial x_m} a_{ij}, a_j, \frac{\partial}{\partial x_k} a_i, a$$

are Hölder continuous (exponent α) in (x, t) which varies in a bounded domain D, and that

$$a \leqq 0 \text{ in } D, \ \Sigma a_{ij}\xi_i\xi_j \geqq A_0 \Sigma \xi_i^2 \text{ in } D (A_0 > 0)$$

for any real vector ξ.

DEFINITION. A bounded measurable function $u(x, t)$ in D is called *weakly L-subcaloric* (or simply, *weakly subparabolic* when there is no confusion about the L) if for any compact subdomain E of D with piecewise smooth boundary (so that Green's formula holds)

$$\iint_E u(x, t)L^*v(x, t)dxdt \geqq 0 \tag{1}$$

for any function $v(x, t)$ satisfying the following properties:

(i) $v \geqq 0$ in E,

(ii) $v, \partial v/\partial v_i, \partial^2 v/\partial x_i \partial x_j, \partial v/\partial t$ are continuous in E and vanish on the boundary ∂E of E.

We note that, for the establishment of the maximum principle below, it is enough that (1) holds only for some special types of domains, namely, for cylindrical domains and for certain sections of paraboloids.

DEFINITIONS. For any point $P(x^0, t^0)$ in D, we denote by $C(P)$ the set of all points (x^1, t^1) in D such that there exists a differentiable curve connecting (x^0, t^0) to (x^1, t^1) and along which the t-coordinate is non-increasing. A function $u(x, t)$ is said to be *continuous from below* at a point $P = (x^0, t^0)$ if u, as a function in $C(P)$, is continuous at P in the usual sense. By a *neighborhood-from-below* of a point P we mean the intersection of a neighborhood of P with $C(P)$.

Our purpose is to prove the following theorem.

THEOREM. *Let u be a weakly L-subcaloric in D. If u assumes its essential supremum M (in D) at a point $P = (x^0, t^0)$ at which u is continuous from below, and if $M \geqq 0$, then $u = M$ almost everywhere in $C(P)$.*

As in [4], the proof follows immediately once we have established the following lemmas.

LEMMA 1. *Under the assumptions of the theorem, there exists a neighborhood-from-below N of P such that $u = M$ almost everywhere in N.*

LEMMA 2. *Let u be a weakly L-subcaloric function in D. If $u = M$ almost everywhere in a neighborhood-from-below of some point P of D, and $M \geqq 0$, then $u = M$ almost everywhere in $C(P)$.*

2. Proof of Lemma 2. We shall prove that, given a compact subset E of D, we can construct, for each point $Q = (y, \tau)$ in E a domain

$$\Omega = \Omega_{\delta\varepsilon} : -\delta < t - \tau < 0, \varepsilon \,|\, x - y \,|^2 < |\, t - \tau \,| \, (\varepsilon > 0, \delta > 0)$$

and a function $w(x, t) = w^{y,\tau}(x, t)$ having the following properties:

(a) $w > 0$ in Ω.

(b) $w, \partial w/\partial x_i, \partial^2 w/\partial x_i \partial x_j, \partial w/\partial t$ are continuous in $\bar{\Omega} - \{(y, \tau)\}$ and vanish on the boundary $\partial\Omega - \{(y, \tau)\}$.

(c) $L^* w > 0$ in Ω.

Furthermore, ε may be any number between 0 and 1 and δ may be taken to be dependent *only* on L, ε and E, but not on the particular point $Q = (y, \tau)$. Finally, as $\varepsilon \to 0$, the radius of the base (or δ/ε) can be taken to be bounded away from zero.

Once w has been constructed, a simple argument of [4] can easily be extended to complete the proof of the lemma. For the sake of completeness we reproduce it here.

Let S be the set of points (x, t) in $C(P)$ having the property that $u = M$ almost everywhere in an open-from-below set containing (x, t). By assumption S is nonempty. Clearly S is open from below. If we show that S is also closed, then S coincides with $C(P)$. To prove it, we take any sequence $Q_m \to R$, Q_m in S, R in D, and use the above construction with $E = \{R, Q_1, Q_2, \cdots\}$. If we show that $u = M$ almost everywhere in each domain Ω_i corresponding to Q_i, then it would follow that R also belongs to S. (Note that in the construction of the Ω below, the radius of the base of Ω can be made bounded away from zero as $\varepsilon \to 0$.)

For simplicity we denote Ω_i by Ω and the corresponding w_i by w. We now modify the definition of $w(x, t)$ in the intersection of Ω with a neighborhood-from-below N of Q_i where $u = M$ almost everywhere. The modified function is denoted by W, and is taken to satisfy the conditions imposed on the function v in the definition of subcaloricity (in § 1) with E replaced by Ω. Denote $A = N \cap \Omega$, $B = \Omega - A$. Using the definition of weakly L-subcaloric functions, we get

$$\iint_\Omega u L^* W dx dt \geqq 0 \, , \tag{2}$$

Now,

$$\iint_\Omega L^* W dxdt = \iint_\Omega W L 1 dxdt \leqq 0 \ ;$$

hence

$$\iint_B L^* W dxdt \leqq -\iint_A L^* W dxdt \ . \tag{3}$$

On the other hand, by (2),

$$\iint_B u L^* W dxdt \geqq \iint_A u L^* W dxdt = -M \iint_A L^* W dxdt \ .$$

Using (3) we obtain

$$\iint_B (u - M) L^* W dxdt \geqq 0 \ .$$

Since $L^* W = L^* w > 0$ in B, $u - M$ must vanish in B almost everywhere.

To complete the proof of Lemma 2 we have to construct a function w which the required properties (a) — (c). For simplicity we shall do it in the special case is $\tau = 0$, $y = 0$; the general case is immediately obtained by translation.

DEFINITION OF w:

$$w = (\delta + t)^2(-t - \varepsilon r^2)^3 \tilde{r}^{-k} \tag{4}$$

where

$$r = |x|, \ \tilde{r}^2 = r^2 - k^{1/2} t,$$

where k is a positive integer to be determined later. Clearly, w satisfies (a), (b). It remains to prove that $L^* w > 0$ in Ω. We have

$$\frac{\partial w}{\partial x_i} = -6\varepsilon x_i(\delta + t)^2(-t - \varepsilon r^2)^2 \tilde{r}^{-k} - k x_i (\delta + t)^2 (-t - \varepsilon r^2)^3 \tilde{r}^{-k-2} \ ,$$

$$\begin{aligned}\frac{\partial^2 w}{\partial x_i \partial x_j} &= 24\varepsilon^2 x_i x_j (\delta + t)^2(-t - \varepsilon r^2)\tilde{r}^{-k} - 6\varepsilon\delta_{ij}(\delta + t)^2(-t - \varepsilon r^2)^2 \tilde{r}^{-k} \\ &\quad + 12k\varepsilon x_i x_j (\delta + t)^2(-t - \varepsilon r^2)^2 \tilde{r}^{-k-2} - k\delta_{ij}(\delta + t)^2(-t - \varepsilon r^2)^3 \tilde{r}^{-k-2} \\ &\quad + k(k+2) x_i x_j (\delta + t)^2(-t - \varepsilon r^2)^3 \tilde{r}^{-k-4} \ ,\end{aligned}$$

$$\begin{aligned}\frac{\partial w}{\partial t} &= 2(\delta + t)(-t - \varepsilon r^2)^3 \tilde{r}^{-k} - 3(\delta + t)^2(-t - \varepsilon r^2)^2 \tilde{r}^{-k} \\ &\quad + \frac{1}{2} k^{3/2} (\delta + t)^2 (-t - \varepsilon r^2)^3 \tilde{r}^{-k-2} \ .\end{aligned}$$

We now form $L^* w$, and restrict δ to be sufficiently small and restrict $|x|$ to be sufficiently small (depending only on L^*), say $|x| \leqq \rho$. Then, the contribution to $L^* w$ made by the terms of $\Sigma b_i \partial w / \partial x_i + bw$ is small compared with the corresponding last two terms in $\partial w / \partial t$. Also, the negative contribution in $\Sigma b_{ij} \partial^2 w / \partial x_i \partial x_j$ corresponding to the fourth term in $\partial^2 w/\partial x_i \partial x_j$ (calculated above) can be neglected as compared to the third term in $\partial w/\partial t$ (provided k is sufficiently large, depending on b_{ij}). Discarding (as we may) the positive contribution corresponding to the first and the last terms in $\partial^2 w/\partial x_i \partial x_j$, we conclude that in order to prove that $L^* w > 0$, it is sufficient to prove that

$$k\varepsilon \frac{r^2}{\tilde{r}^2} + k^{3/2}(-t - \varepsilon r^2)\frac{1}{\tilde{r}^2} \geqq \lambda > 0 \tag{5}$$

where λ is a constant depending only on L and ρ ($|x| \leqq \rho$ in Ω).

To prove (5) we take $k > 1/\varepsilon^2$, which imples that, in Ω (where $\varepsilon r^2 < |t|$),

$$k^{1/2}|t| \leqq \tilde{r}^2 = r^2 + k^{1/2}|t| \leqq 2k^{1/2}|t| .$$

Hence (5) is a consequence of

$$k^{1/2}\varepsilon r^2 + k(-t - \varepsilon r^2) \geqq 2\lambda |t|$$

which is clearly true if $k^{1/2} \geqq 2\lambda$, $k \geqq 1$.

3. **Properties of fundamental solutions.** Assume that the closure of a cylinder $C : |x|^2 < \beta$, $-\delta < t < 0$ with base B is contained in D. By our assumptions on L, there exists (by Pogorzelski [7]) in C a fundamental solution $\Gamma(x, t; \xi, \tau)(t < \tau)$ of L^* with pole (ξ, τ); $L^*\Gamma = 0$ as a function of (x, t), and Γ can be constructed as follows:

Let (B_{ij}) be the matrix inverse to (b_{ij}) and define

$$\sigma(x, t ; \xi, \tau) = \Sigma B_{ij}(\xi, \tau)(x_i - \xi_i)(x_j - \xi_j)$$
$$Z(x, t ; \xi, \tau) = (\tau - t)^{-n/2} \exp\left\{-\frac{\sigma(x, t; \xi, \tau)}{4(\tau - t)}\right\}$$

$$\Gamma(x, t; \xi, \tau) = Z(x, t; \xi, \tau) + \int_t^\tau \int_B Z(x, x; \eta, s)\Phi(\eta, s; \xi, \tau)d\eta ds$$

where Φ is the solution of the integral equation

$$L^*_{(x,t)}Z(x, t; \xi, \tau) - \rho(x, t)\Phi(x, t; \xi, \tau)$$
$$+ \int_\tau^t \int_B [L^*_{(x,t)}Z(x, t; \eta, s)]\Phi(\eta, s; \xi, \tau)d\eta ds = 0 .$$

Here,

$$\rho(x, t) = (4\pi)^{n/2}/(\det (B_{ij}(x, t))^{1/2} .$$

Note that

$$0 < \text{const.} \leqq \frac{\sigma(x, t; \xi, \tau)}{|x - \xi|^2} \leqq \text{const.} < \infty .$$

In the following we shall be interested in the special case $(\xi, \tau) = 0$. We define

$$g(x, t) = \Gamma(x, t; 0, 0)$$
$$\sigma(x, t) = \sigma(x, t; 0, 0)$$
$$Z(x, t) = Z(x, t; 0, 0) .$$

By simple calculation we get

$$g(x, t) = Z(x, t)(1 + o(1)) \tag{6}$$

where $o(1) \to 0$ as $t \to 0$. Hence, in particular, $g(x, t) > 0$ if the height δ of C is sufficiently small, as we shall assume. We also mention, although this is not used later on, that for any bounded measurable function $\varphi(x, t)$ in C, which is continuous at $(0, 0)$ we have (see [8])

$$\lim_{t\to 0}\int_B g(x, t)\varphi(x, t)dx = \rho(0, 0)\varphi(0, 0) . \tag{7}$$

We conclude this section with estimating the following expression (which will appear in the next section)

$$I \equiv - \sum_{i,j} b_{ij}(x, t)x_i\frac{\partial g(x, t)}{\partial x_j} . \tag{8}$$

Since

$$- \frac{\partial}{\partial x_j}Z(x, t) = \frac{1}{2t}(\sum_k B_{jk}(0, 0)x_k)Z(x, t) ,$$

and since

$$\sum_{i,k}\sum_{j=1}^{n} b_{ij}(x, t)B_{jk}(0, 0)x_ix_k$$
$$= | x |^2 + \sum_{i,j,k}[b_{ij}(x, t) - b_{ij}(0, 0)]B_{jk}(0, 0)x_ix_k = | x |^2(1 + o(1))$$

where $o(1) \to 0$ as $|x| \to 0$, we conclude that

$$I_1 \equiv -\sum_{i,j} b_{ij}(x, t)x_i\frac{\partial Z(x,t)}{\partial x_j} \geqq \frac{| x |^2}{3 | t |}Z(x, t) \tag{9}$$

provided $|x|$ is sufficiently small.

To evaluate $I - I_1$, we use the definitions of g and Γ, and proceed to estimate the x_j-derivatives of the integral which appears in the definition of Γ. Noting that

$$\left|\frac{\partial}{\partial x_j}Z(x, t; \eta, s)\right| \leqq \text{const.}\ (s - t)^{-1/2}Z(x, t; \eta, s)$$

and using the estimate of [7] for Φ and Dressel [1; Lemma 2] we find that

$$|I - I_1| \leqq \lambda_0 |x|^\gamma Z(x, t) \qquad (\lambda_0 > 0, 0 < \gamma \leqq 1) ,$$

where λ_0, γ depend only on L. In what follows we shall only need the weaker inequality

$$I \geqq -\lambda_0 |x|^\gamma Z(x, t) . \tag{10}$$

4. **Proof of Lemma 1.** We may assume, without loss of generality, that the essential supremum M is assumed at the origin. Following the procedure of Littman [4], we claim that it is enough to construct a function $G(x, t)$ in a cylinder C: $|x|^2 < \beta$, $-\delta < t < 0$, with base B, which satisfies the following conditions:

(a) G, $\partial G/\partial x_i$, $\partial^2 G/\partial x_i \partial x_j$, $\partial G/\partial t$ are continuous in $\bar{C} - \{(0, 0)\}$ and vanish on the boundary $\partial C - \{(0, 0)\}$.

(b) $L^*G > 0$ in C.

(c) If $f(x, t)$ is L-subcaloric in a domain which contains C, and if f is continuous from below at the origin and $f(0, 0) = 0$, then

$$0 \leqq \iint_C f L^* G dx dt . \tag{11}$$

Once G is constructed, the proof of Lemma 1 follows very easily. Indeed, $u - M$ is L-subcaloric, and using (c) we get

$$\iint_C (u - M) L^* G dx dt \geqq 0 .$$

Since, by (b), $L^*G > 0$, we conclude that $u = M$ almost everywhere in C.

DEFINITION OF $G(x, t)$:

$$G(x, t) = (t + \delta)^2 (\beta - r^2)^3 g(x, t) \tag{12}$$

where $g(x, t)$ is defined in § 3. Clearly (a) is satisfied. We proceed to establish (b), (c).

Proof of (b).

$$\frac{\partial G}{\partial x_i} = -6x_i(t + \delta)^2(\beta - r^2)^2 g + (t + \delta)^2(\beta - r^2)^3 \frac{\partial g}{\partial x_i} ,$$

$$\begin{aligned} \frac{\partial^2 G}{\partial x_i \partial x_j} &= -6\delta_{ij}(t + \delta)^2(\beta - r^2)^2 g + 24 x_i x_j (t + \delta)^2 (\beta - r^2) g \\ &\quad - 6x_i(t + \delta)^2(\beta - r^2)^2 \frac{\partial g}{\partial x_j} - 6x_j(t + \delta)^2(\beta - r^2)^2 \frac{\partial g}{\partial x_i} \\ &\quad + (t + \delta)^2(\beta - r^2)^3 \frac{\partial^2 g}{\partial x_i \partial x_j} , \end{aligned}$$

$$\frac{\partial G}{\partial t} = 2(t+\delta)(\beta - r^2)^3 g + (t+\delta)^2(\beta - r^2)^3 \frac{\partial g}{\partial t} .$$

Recalling that $L^*g = 0$ we obtain

$$\begin{aligned} L^*G = & -6(t+\delta)^2(\beta - r^2)^2 g \Sigma x_i b_i - 6(\Sigma b_{ii})(t+\delta)^2(\beta - r^2)^2 g \\ & + 24(\Sigma b_{ij} x_i x_j)(t+\delta)^2(\beta - r^2) g - 12(t+\delta)^2(\beta - r^2)^2 \Sigma b_{ij} x_i \frac{\partial g}{\partial x_j} \\ & + 2(t+\delta)(\beta - r^2)^3 g . \end{aligned}$$

Now the first term in L^*G is small compared with the second one, if $|x|$ (or β) is small. Using (8), (10), (6) to estimate the fourth term, we conclude that, if it is negative, then its absolute value is smaller than that of the second term. Hence, if we prove that

$$(t+\delta)r^2 + (\beta - r^2)^2 > \mu(t+\delta)(\beta - r^2) \tag{13}$$

for sufficiently large μ depending only on L (provided β is smaller than an appropriate constant), then $L^*G > 0$.

To prove (13) we note that if $\mu(\beta - r^2) < r^2$ then (13) clearly holds. Hence it remains to consider the case where

$$\mu(\beta - r^2) \geqq r^2 .$$

However, in this case

$$(\beta - r^2)^2 \geqq \frac{\beta^2}{(1+\mu)^2} > \mu(t+\delta)\beta$$

for sufficiently small δ (i.e., if $(\mu+1)^2\mu\delta < \beta$), from which (13) follows.

Proof of (c). We modify G as follows: Let

$$\sigma_\varepsilon(x,t) = \begin{cases} \sigma(x,t) \text{ if } -\delta < t < -\varepsilon \\ \sigma(x,t) + (t+\varepsilon) \text{ if } -\varepsilon \leqq t \leqq 0 . \end{cases}$$

Clearly $\sigma_\varepsilon(x,t)$ has second continuous x-derivatives and a first continuous t-derivative in $\bar{C}$. We next define

$$\begin{aligned} Z_\varepsilon(x,t) &= \frac{1}{(-t)^{n/2}} \exp\left\{\frac{\sigma_\varepsilon(x,t)}{4t}\right\} . \\ g_\varepsilon(x,t) &= Z_\varepsilon(x,t) + \int_t^0 \int_B Z(x,t;\eta,s)\Phi(\eta,s;0,0)d\eta ds , \\ G_\varepsilon(x,t) &= (t+\delta)^2(\beta - r^2)^3 g_\varepsilon(x,t) . \end{aligned}$$

G_ε is differentiable also at the origin where it vanishes. We now proceed to prove (c).

By the definition of L-subcaloricity (see (1)) we have,

$$\iint_C f(x, t) L^* G_\varepsilon(x, t) dx dt \geqq 0 . \tag{14}$$

If we prove that

$$\lim_{\varepsilon \to 0} \iint_{C_\varepsilon} f x, t) L^* G(x, t) dx dt = 0 , \tag{15}$$

$$\lim_{\varepsilon \to 0} \iint_{C_\varepsilon} f(x, t) L^* G_\varepsilon(x, t) dx dt = 0 , \tag{16}$$

where $C_\varepsilon = C \cap \{-\varepsilon < t < 0\}$, then (c) follows from (14).

In what follows we denote any positive constant (independent of ε) by the same symbol A. To prove (15) we write

$$\iint_{C_\varepsilon} f L^* G dx dt = \int_{-\varepsilon}^{0} \int_{|x|<\eta} f L^* G dx dt + \int_{-\varepsilon}^{0} \int_{\eta<|x|<\beta} f L^* G dx dt , \tag{17}$$

where η is any positive number smaller than β. Since f is continuous from below at (0, 0) and $f(0, 0) = 0$, the first integral on the right side of (17) tends to zero as $\eta \to 0$, independently of ε.

Here we have made use of (see [7])

$$| L^* G(x, t) | \frac{A}{| t |^{(n+\nu+1)/2}} \exp\left\{\frac{Ar^2}{t}\right\} \text{ for some } 0 \leqq \nu < 1 . \tag{18}$$

The second integral on the right side of (17), for any fixed η, also tends to zero as follows by using (18).

Proof of (16). Proceeding similarly to the proof of (15), we find that all that remains to be proved is that

$$\iint_{C_\varepsilon} | L^* G_\varepsilon | dx dt \leqq A < \infty \tag{19}$$

for all $\varepsilon > 0$ (A is independent of ε). Now,

$$-L^* g_\varepsilon = L^*(g - g_\varepsilon) = L^*(Z - Z_\varepsilon) = L^*\left[Z(x, t)\left(1 - \exp\left\{\frac{(t+\varepsilon)^2}{4t}\right\}\right)\right]$$

$$= L^* Z\left(1 - \exp\left\{\frac{(t+\varepsilon)^2}{4t}\right\}\right) - Z\left[\exp\left\{\frac{(t+\varepsilon)^2}{4t}\right\}\right]\left[\frac{t+\varepsilon}{2t} - \frac{(t+\varepsilon)^2}{4t^2}\right] .$$

Since

$$| L^* Z | \leqq \frac{A}{| t |^{(n+1+\nu)/2}} \exp\left\{\frac{Ar^2}{t}\right\}$$

for some $0 \leqq \nu < 1$, we find, denoting $\left|\frac{t+\varepsilon}{2t} - \frac{(t+\varepsilon)^2}{4t^2}\right|$ shortly by $[\cdots]$,

$$| L^* G_\varepsilon | \leqq \frac{A}{| t |^{(n+1+\nu)/2}} \exp\left\{\frac{Ar^2}{t}\right\} \tag{20}$$

$$+ \frac{A}{| t |^{n/2}} [\cdots] \exp\left\{\frac{Ar^2}{t}\right\} \exp\left\{\frac{(t+\varepsilon)^2}{4t}\right\} \equiv K_1 + K_2 .$$

The integral of K_1 is easily seen to be bounded. Hence it remains to evaluate

$$J \equiv \iint_{C_\varepsilon} K_2 dxdt \ .$$

We split J in the following way:

$$J = \int_{-\varepsilon}^{-\varepsilon/2}\int_B K_2 dxdt + \int_{-\varepsilon/2}^{0}\int_B K_2 dxdt \equiv J_1 + J_2 \ .$$

As for J_1, $[\cdots] \leqq 1$ and hence $J_1 \leqq A$. As for J_2, $[\cdots] \leqq A\varepsilon^2/t^2$ and hence

$$J_2 \leqq A\varepsilon^2 \int_{-\varepsilon/2}^{0}\left(\int_B \frac{1}{|t|^{n/2}} \exp\left\{\frac{Ar^2}{4t}\right\} dr\right)\frac{1}{t^2}\exp\left\{\frac{A\varepsilon^2}{t}\right\} dt \ .$$

The inner integral is bounded. Substituting $z = \varepsilon^2/|t|$ we get

$$J_2 \leqq A\int_{2\varepsilon}^{\infty} e^{-Az} dz \ .$$

We have thus proved that $J = J_1 + J_2 \leqq A$, which completes the proof of (19). Hence, the proof of (16) is completed.

REMARK. The maximum principle for subelliptic functions [4] follows from the maximum principle for subparabolic functions proved in this paper. Indeed, as is easily seen, a weak subelliptic function is necessarily a weak subparabolic function.

REFERENCES

1. F. G. Dressel, *The fundamental solution of the parabolic equation II*, Duke Math. J., **13** (1946), 61-70.

2. A. Friedman, *Parabolic equations of the second order*, Trans. Amer. Math. Soc., **93** (1959), 509-530.

3. E. Hopf, *Elementare Bemerkungen über die Lösungen partieller Differentialgleichungen zweiter Ordnung von elliptischen Typus*, Sitzungsberichte, Berlin Akad. Wiss., **19** (1927), 147-152.

4. W. Littman, *A strong maximum principle for weakly L-subharmonic functions*, J. Math. and Mech., **8** (1959), 761-770.

5. L. Nirenberg, *A strong maximum principle for parabolic equations*, Comm. Pure App. Math., **6** (1953), 167-177.

6. ———, *Existence Theorems in Partial Differential Equations*, New York University Notes.

7. W. Pogorzelski, *Étude de la solution fondamentale de l'équation parabolique*, Ricerche di Mat., **5** (1956), 25-57.

8. ———, *Propriétés des intégrales de l'équation parabolique normale*, Ann. Polon, Mat., **4** (1957), 61-91.

UNIVERSITY OF MINNESOTA, INSTITUTE OF TECHNOLOGY, MINNEAPOLIS, MINNESOTA

ASYMPTOTICS II: LAPLACE'S METHOD FOR MULTIPLE INTEGRALS

W. FULKS AND J. O. SATHER

Laplace's method is a well known and important tool for studying the rate of growth of an integral of the form

$$I(h) = \int_a^b e^{-hf} g dx$$

as $h \to \infty$, where f has a single minimum in $[a, b]$. It's extension to multiple integrals has been studied by L. C. Hsu in a series of papers starting in 1948, and by P. G. Rooney (see bibliography). These authors establish what amount to a first term of an asymptotic expansion. All but one (see [7]) of these results are under fairly heavy smoothness conditions.

In this paper we examine multiple integrals of the form

$$I(h) = \int_R e^{-hf} g dx$$

where f and g are measurable functions defined on a set R in E_p. Without making any smoothness assumptions on f and g, and using only the existence of $I(h)$ and, of course, asymptotic expansions of f and g near the minimum point of f we obtain an asymptotic expansion of I. The special features of our procedure are the lack of smoothness assumptions and the fact that we get a complete expansion.

Without loss of generality we may assume that the essential infimum of f occurs at the origin, and that this minimal value is zero. We introduce polar coordinates: $x = (\rho, \Omega)$ where

$$\rho = |x| = \sqrt{x_1^2 + x_2^2 + \cdots + x_p^2} ,$$

and where $\Omega = x/|x|$ is a point on the surface, S_{p-1}, of the unit sphere.

Our hypothesis are the following:

(1) The origin is an interior point of R.

(2) For each $\rho_0 > 0$ there is an $A > 0$ such that $f(\rho, \Omega) \geqq A$ if $\rho \geqq \rho_0$. (This says that f can be close to zero only at the origin.)

(3) There is an $n \geqq 0$ and $n + 1$ continuous functions $f_k(\Omega)$, $k = 0, 1, 2, \cdots, n$, defined on S_{p-1} with $f_0 > 0$ for which

$$f(\rho, \Omega) = \rho^\nu \sum_{k=0}^{n} f_k(\Omega)\rho^k + o(\rho^{n+\nu}) \text{ as } \rho \to 0$$

Received April 29, 1960. The work on this paper was performed under sponsorship of the Office of Naval Research, Contract Nonr 710 (16), at the University of Minnesota.

where $\nu > 0$. (This is meant in the following sense: for each $\varepsilon > 0$ there is a $\rho_0 > 0$ for which

$$|f(\rho, \Omega) - \rho^{\nu} \sum_{k=0}^{n} f_k(\Omega)\rho^k| < \varepsilon\rho^{n+\nu}$$

whenever $\rho \leqq \rho_0$. Besides giving the asymptotic behavior of f near the origin (3) implies that the infimum of f in R is indeed zero.)

(4) There are $n + 1$ functions $g_k(\Omega), k = 0, 1, \cdots n$, for which

$$g = \rho^{\lambda-p} \sum_{k=0}^{n} g_k(\Omega)\rho^k + o(\rho^{n+\lambda-k}) \text{ as } \rho \to 0$$

where $\lambda > 0$. (Thus g is permitted a mild singularity at the origin. The expansion is meant in the same sense as the one in (3).)

Under these conditions we will prove that if there is a h_0 for which $I(h)$ exists then it exists for all $h \geqq h_0$ and

$$I(h) = \sum_{k=0}^{n} c_k h^{-(k+\lambda)/\nu} + o(h^{-(n+\lambda)/\nu})$$

where the c_k's are constants depending only on the f_j's and g_j's for $j \leqq k$. Their evaluation will be described in the proof of this result. In particular

$$C_0 = \frac{\Gamma((\lambda + 1)/\nu)}{\lambda} \int_{S_{p-1}} g_0(\Omega)/[f_0(\Omega)]^{\lambda/\nu} d\Omega$$

where $d\Omega$ is the element of $(p - 1)$-dimensional measure on S_{p-1}.

In the course of the proof we will use the following lemmas, which are given now so as to not interrupt the main thread of the argument.

LEMMA 1. *Let f be a measurable function on a set R in E_p, and let $g \in L_1(R)$. Then the function $G(z)$ defined by*

$$G(z) = \int_{\{f \leqq z\}} g dx$$

has bounded variation on $\{-\infty < z < \infty\}$.

Proof. Let $g = g_1 - g_2$, where

$$g_1(x) = \begin{cases} g(x), & g(x) \geqq 0 \\ 0, & g(x) < 0 \end{cases}; \qquad g_2(x) = \begin{cases} 0, & g(x) \geqq 0 \\ -g(x), & g(x) < 0, \end{cases}$$

and define G_1 and G_2 by

$$G_1(z) = \int_{\{f \leqq z\}} g_1 dx , \qquad G_2(z) = \int_{\{f \leqq z\}} g_2 dx .$$

Clearly G_1 and G_2 are increasing and bounded on $\{-\infty < z < \infty\}$, and $G = G_1 - G_2$.

LEMMA 2. *Let $F(t)$ be a continuous function defined on a possibly infinite interval $\{a < t < b\}$, and let f be a measurable function on a set R in E_p taking values in the interval $\{a < t < b\}$. If $g \in L_1(R)$, and $F(f)g \in L_1(R)$ and G is defined as in Lemma 1, then*

$$\int_R F(f)gdx = \int_a^b F(t)dG(t) .$$

Proof. Suppose first that a and b are finite, and that $g \geqq 0$. Form a partition: $a = t_0 < t_1 < \cdots < t_n = b$, and set

$$E_j = \{x \mid t_{j-1} < f \leqq t_j\} ,$$

and let $M_j = \sup_{\{t_{j-1} \leqq t \leqq t_j\}} F(t)$ and $m_j = \inf_{\{t_{j-1} \leqq t \leqq t_j\}} F(t)$.

Then

$$\begin{aligned} \int_R F(f)gdx &= \sum_{j=1}^n \int_{E_j} F(f)gdx \leqq \sum_{j=1}^n M_j \int_{E_j} gdx \\ &= \sum_{j=1}^n M_j[G(t_j) - G(t_{j-1})] . \end{aligned}$$

Similarly

$$\int_R F(f)gdx \geqq \sum_{j=1}^n m_j[G(t_j) - G(t_{j-1})] .$$

If we let $n \to \infty$ so that $\max_{1 \leqq j \leqq n}(t_j - t_{j-1}) \to 0$ then both

$$\sum_{j=1}^n M_j[G(t_j) - G(t_{j-1})] \text{ and } \sum_{j=1}^n m_j[G(t_j) - G(t_{j-1})]$$

converge to $\int_a^b F(t)dG(t)$, since F is continuous and G monotone.

If g is not positive we can write $g = g_1 - g_2$ as in Lemma 1, apply the proof just completed to each of g_1 and g_2, and combine the results to complete the proof for the case where a and b are finite.

Suppose for example b is infinite. Then for any finite b',

$$\begin{aligned} \int_R F(f)gdx &= \lim_{b' \to \infty} \int_{\{f \leqq b'\}} F(f)gdx = \lim_{b' \to \infty} \int_a^{b'} F(t)dG(t) \\ &= \int_a^\infty F(t)dG(t) . \end{aligned}$$

A similar argument applies if $a = -\infty$.

We now return to the proof of the main theorem. First we note that if $h \geqq h_0$ then $e^{-h_0 f}g$ forms a dominating function for $e^{-hf}g$, so that

$I(h)$ exists.

For each $\varepsilon > 0$ we define the two functions $f_+(\rho, \Omega)$ and $f_-(\rho, \Omega)$ by

$$f_\pm(\rho, \Omega) = \rho^\nu \sum_{k=0}^{n} f_k(\Omega)\rho^k \pm \varepsilon\rho^{n+\nu} .$$

These functions are defined in all of E_p. Now given an $\varepsilon > 0$ there is a ρ_0 so that

(i) $|f(\rho, \Omega) - \rho^\nu \sum_{k=0}^{n} f_k(\Omega)\rho^k| < \varepsilon\rho^{n+\nu}$

(ii) $|g(\rho, \Omega) - \rho^{\lambda-p} \sum_{k=0}^{n} g_k(\Omega)\rho^k| < \varepsilon\rho^{n+\lambda-p}$ for $\rho < \rho_0$,

and so that

(iii) both the functions $f_\pm(\rho, \Omega)$ are increasing in ρ for $\{0 \leqq \rho \leqq \rho_0\}$ for each $\Omega \in S_{p-1}$. This can easily be achieved since f_0 is positive (and therefore bounded away from zero) and the other f_k's are bounded.

(iv) the sphere $\{\rho \leqq \rho_0\}$ is in R.

We denote $\{\rho \leqq \rho_0\}$ by R_0 and write $I(h)$ in the form

$$I(h) = \int_{R_0} e^{-hf} g dx + \int_{R-R_0} e^{-hf} g dx \equiv I_1(h) + I_2(h)$$

respectively. We proceed to estimate I_2: by hypothesis (2) there is an $A > 0$ so that $f \geqq A$ if $\rho \geqq \rho_0$. Thus

$$\begin{aligned} |I_2(h)| &\leqq \int_{R-R_0} e^{-hf} |g| dx \leqq e^{-(h-h_0)A} \int_{R-R_0} e^{-h_0 f} |g| dx \\ &= Ce^{-hA} \text{ where } C \text{ is a constant.} \end{aligned}$$

That is,

$$I_2(h) = O(e^{-hA}) \text{ as } h \to \infty ,$$

so it is clear that the dominant part of $I(h)$ must arise from $I_1(h)$. The remainder of the proof is largely concerned with estimating I_1.

In R_0 we define $r(\rho, \Omega)$ by

$$g(\rho, \Omega) = \rho^{\lambda-p} \sum_{0}^{n} g_k(\Omega)\rho^k + r(\rho, \Omega)\rho^{n+\lambda-p} .$$

Let

$$g_k^+(\Omega) = \begin{cases} g_k(\Omega), & g_k(\Omega) \geqq 0 \\ 0, & g_k(\Omega) < 0 \end{cases}, \qquad g_k^-(\Omega) = \begin{cases} 0, & g_k(\Omega) \geqq 0 \\ -g(\Omega), & g_k(\Omega) > 0 \end{cases}$$

and

$$r^+(\rho, \Omega) = \begin{cases} r(\rho, \Omega), & r(\rho, \Omega) \geqq 0 \\ 0, & r(\rho, \Omega) < 0 \end{cases}; \quad r^-(\rho, \Omega) = \begin{cases} 0, & r(\rho, \Omega) \geqq 0 \\ -r(\rho, \Omega), & r(\rho, \Omega) < 0 . \end{cases}$$

In R_0 we now define $g^+(\rho, \Omega)$ and $g^-(\rho, \Omega)$ by

$$g^+(\rho, \Omega) = \rho^{\lambda-p} \sum_{k=0}^{n} g_k^+(\Omega)\rho^k + r^+(\rho, \Omega)\rho^{n+\lambda-p}$$

and

$$g^-(\rho, \Omega) = \rho^{\lambda-p} \sum_{k=0}^{n} g^-(\Omega)\rho^k + r^-(\rho, \Omega)\rho^{n+\lambda-p} .$$

Then $g = g^+ - g^-$ and

$$I_1 = \int_{R_0} e^{-hf} g^+ dx - \int_{R_0} e^{-hf} g^- dx .$$

Thus we may assume that $g \geqq 0$ in R_0.

We recall the definition of f_+ and f_- and define $I_+(h)$ and $I_-(h)$ by

$$I_+(h) = \int_{R_0} e^{-hf_+} g dx,\ I_-(h) = \int_{R_0} e^{-hf_-} g dx .$$

Since $g \geqq 0$ we conclude

$$I_+(h) \leqq I_1(h) \leqq I_-(h) .$$

Next we turn our attention to I_+: Let $R_t = \{x \mid f_+ \leqq t\}$ and choose a so small that $R_a \subset R_0$. Then

$$I_+(h) = \int_{R_a} e^{-hf_+} g dx + \int_{R_0 - R_a} e^{-hf_+} g dx = I'_+ + I''_+ ,$$

respectively. Now f_+ is bounded away from zero in R_0 outside any neighborhood of the origin. Thus by the same argument used on I_2 we get

$$I''_+ = O(e^{-hA'}) .$$

Furthermore e^{-hf_+} is bounded away from zero in R_a, since f_+ is bounded there. Thus $e^{-hf_+} g \in L_1(R_a)$ and by Lemma 2,

$$I'_+ = \int_0^a e^{-ht} dG(t) ,$$

where $G(t) = \int_{R_t} g dx$. Integrating by parts we get

$$\begin{aligned} I'_+ &= e^{-ha} G(a) + h \int_0^a e^{-ht} G(t) dt \\ &= h \int_0^a e^{-ht} G(t) dt + O(e^{-ha}) . \end{aligned}$$

We next do some preliminary calculations, preparatory to estimating $G(t)$. For each $t, 0 \leqq t \leqq a$, the equation $t = f_+(\rho, \Omega)$ has a unique solution for ρ which is continuous in Ω, since f_+ is increasing in ρ.

Thus the solution defines a star-shaped curve (or surface) given by $\rho = \rho(t, \Omega)$. We proceed to estimate $\rho(t, \Omega)$. Set $t = U^\nu$ then $t = f_+(\rho, \Omega)$ can be written in the form

$$U^\nu = \rho^\nu\Big[\sum_0^n f_k(\Omega)\rho^k + \varepsilon\rho^n\Big]$$

or

$$U = \rho[f_0(\Omega) + f_1(\Omega)\rho + \cdots (f_n(\Omega) + \varepsilon)\rho^n]^{1/\nu} .$$

From here on we assume $n > 0$, for if $n = 0$, we can solve directly for ρ and the estimates are considerably simpler than those which follow.

Now the right hand side of the last equation is a monotone function of $\rho, 0 \leqq \rho \leqq a$, hence an inverse function exists. Since, for each fixed Ω, U is an $(n + 2)$-times differentiable (it's even analytic!) function of ρ, $0 \leqq \rho \leqq a$, then ρ is an $(n + 2)$-times differentiable function of U, and it can therefore be expanded in a Taylor series with remainder. Thus since $f_0(\Omega) > 0$ we get

$$\rho = \psi_1(\Omega)U + \psi_2(\Omega)U^2 + \cdots + \psi_{n+1}(\Omega, \varepsilon)U^{n+1} + \psi_{n+2}(\Omega, \varepsilon, U)U^{n+2}$$

where $\psi_1(\Omega) = 1/[f_0(\Omega)]^{1/\nu}$. Since the ψ_k's are expressible in terms of the f_k's it is easy to check that ψ_k depends only on f_j's for $j \leqq k$, that ψ_k is independent of ε for $k \leqq n$, that ψ_{n+1} depends only linearly on ε and finally that ψ_{n+2} is uniformly bounded for $\Omega \in S_{p-1}$, $0 \leqq \varepsilon \leqq 1$, and $0 \leqq U \leqq a^{1/\nu}$.

Since $U = t^{1/\nu}$ we express ρ in terms of t, Ω, and ε by

$$\begin{aligned}\rho(t, \Omega) = \psi_1(\Omega)t^{1/\nu} + \psi_2(\Omega)t^{2/\nu} + \cdots + \psi_{n+1}(\Omega, \varepsilon)t^{(n+1)/\nu} \\ + \psi_{n+2}(\Omega, \varepsilon, U)t^{(n+2)/\nu}\end{aligned}$$

By definition $G(t) = \int_{R_t} g dx$, which we can write as

$$G(t) = \int_{S_{p-1}}\int_0^{\rho(t,\Omega)} g(\rho, \Omega)\rho^{p-1}d\rho d\Omega ,$$

where $d\Omega$ represents the element of measure on the sphere $S_{p-1} : \{\rho = 1\}$. We proceed to compute:

$$\begin{aligned}G(t) &= \int_{S_{p-1}}\int_0^{\rho(t,\Omega)}\Big(\sum_0^n g_k(\Omega)\rho^{k+\lambda-1} + o(\rho^{n+\lambda-1})\Big)d\rho d\Omega \\ &= \int_{S_{p-1}}\Big[\rho^\lambda(t, \Omega)\Big(\sum_0^n \frac{g_k(\Omega)}{k + \lambda}\rho^k(t, \Omega)\Big) + o(\rho^{n+\lambda}(t, \Omega))\Big]d\Omega .\end{aligned}$$

If we substitute for $\rho(t, \Omega)$ the expression previously computed for it, the preceding integral can be written in the form

$$G(t) = \int_{S_{p-1}} \left[t^{\lambda/\nu} \sum_0^{n-1} \gamma_k(\Omega) t^{k/\nu} + \gamma_n(\Omega, \varepsilon) t^{(n+\lambda)/\nu} + o(t^{(n+\lambda)/\nu}) \right] d\Omega$$

where γ_k is independent of ε for $k = 0, 1, 2, \cdots, n-1$, and γ_n is linear in ε. We may also note that each of the g_j's enter the γ_k's linearly. In particular

$$\gamma_0 = g_0(\Omega)/[f_0(\Omega)]^{\lambda/\nu} .$$

Now if we write $\gamma_n(\Omega, \varepsilon) = \gamma_n(\Omega) - \varepsilon\gamma'_n(\Omega)$ we have

$$\begin{aligned} G(t) &= \int_{S_{p-1}} \left(\sum_0^n \gamma_k(\Omega) t^{(k+\lambda)/\nu} - \varepsilon\gamma'_n(\Omega) t^{(n+\lambda)/\nu} \right) d\Omega + o(t^{(n+\lambda)/\nu}) , \\ &= \sum_0^n \eta_k t^{(k+\lambda)/\nu} - \varepsilon\eta'_n{}^{(n+\lambda)/\nu} + o(t^{(n+\lambda)/\nu}) \end{aligned}$$

where $\eta_k = \int_{S_{p-1}} \gamma_k(\Omega) d\Omega$. In particular $\eta_0 = (1/\lambda)\int_{S_{p-1}} [g_0(\Omega)/[f_0(\Omega)]^{\lambda/\nu}] d\Omega$.

Now by Watson's lemma we can multiply this asymptotic formula for G by e^{-ht} and integrate termwise to get

$$I'_+ = \sum_0^n c_k h^{-(k+\lambda)/\nu} - \varepsilon c'_n h^{(n+\lambda)/\nu} + o(h^{-(n+\lambda)/\nu})$$

where $c_k = \eta_k \Gamma((k + \lambda + 1)/\nu)$. In particular $c_0 = \eta_0 \Gamma((\lambda + 1)/\nu)$. Since $I_+ = I'_+ + I''_+ = I'_+ + o(e^{-hA'})$, we have also

$$I_+ = \sum_0^n c_k h^{-(k+\lambda)/\nu} - \varepsilon c'_n h^{-(n+\lambda)\nu} + o(h^{-(n+\lambda)/\nu}) .$$

By the same argument, since I_- differs from I_+ only in the sign of ε, we get

$$I_- = \sum_0^n c_k h^{-(k+\lambda)/\nu} + \varepsilon c'_n h^{-(n+\lambda)/\nu} + o(h^{-(n+\lambda)/\nu}) .$$

Now as we have shown before

$$I_+(h) \leqq I_1(h) \leqq I_-(h) .$$

Thus

$$I_+ - \sum_0^n c_k h^{-(k+\lambda)/\nu} \leqq I_1(h) - \sum_0^n c_k h^{-(k+\lambda)/\nu} \leqq I_- - \sum_0^n c_k h^{-(k+\lambda)/\nu} .$$

If we multiply through by $h^{(n+\lambda)/\nu}$ and let $h \to \infty$ we get

$$-\varepsilon c'_n \leqq \underline{\overline{\lim}} \left[(I_1(h) - \sum_0^n c_k h^{-(k+\lambda)/\nu}) h^{(n+\lambda)/\nu} \right] \leqq \varepsilon c'_n .$$

But $I(h) = I_1(h) + o(e^{-hA})$ so that we have also

$$-\varepsilon c'_n \leqq \underline{\overline{\lim}} \left[\left(I(h) - \sum_0^n c_k h^{-(k+\lambda)/\nu} \right) h^{(n+\lambda)/\nu} \right] \leqq \varepsilon c'_n ,$$

for every $\varepsilon > 0$. Let $\varepsilon \to 0$ to complete the proof for $g \geqq 0$.

If g may change sign near the origin we can decompose g with g^+ and g^- as described earlier. The proof just completed applies to each of these. We can then subtract the results to obtain the result for g. Also since g_j''s enter into the c_k''s linearly, the same formula for the c's applies whether g is one signed or has a variable sign near the origin.

BIBLIOGRAPHY

1. L. C. Hsu, *Approximations to a class of double integrals of functions of large numbers*, Amer. J. of Math., **70** (1948).
2. ———, *A theorem on the asymptotic behavior of a double integral*, Duke Math. J., **15** (1948).
3. ———, *An asymptotic expression for an integral involving a parameter*, Acad. Sinica Sci. Record, **2** (1949).
4. ———, *The asymptotic behavior of an integral involving a parameter*, Sci. Rep. Nat. Tsing Hua Univ., **5** (1949).
5. ———, *On the asymptotic behavior of a class of multiple integrals involving a parameter*, Amer. J. Math., **73** (1951)
6. ———, *The asymptotic behavior of a kind of multiple integrals involving a parameter*, Quart. J. Math. Ser., **2** (1951).
7. ———, *A theorem concerning an asymptotic integration*, Chung Kuo L'o Hsueh (Chinese Science), **2** (1951).
8. ———, *One kind of asymptotic integrals having absolute maximum at boundary points*, Acta Math. Sinica, **4** (1954).
9. ———, *On an asymptotic integral*, Proc. Edinburgh Math. Soc. (2), **10** (1956).
10. P. G. Rooney, *Some remarks on Laplace's method*. Trans. Roy. Soc. Canada. III, **47** (1953).

OREGON STATE COLLEGE
AND
INSTITUTE OF TECHNOLOGY
UNIVERSITY OF MINNESOTA

AN EMBEDDING OF RIEMANN SURFACES OF GENUS ONE

A. M. GARSIA AND E. RODEMICH

The C^k embedding of a Riemann surface S will mean here the construction of a C^k surface S' in 3-space which is conformally equivalent to S, if angles on the surface S' are measured in the natural way.[1] The result to be obtained is:

THEOREM. *Any compact Riemann surface of genus one can be C^∞ embedded in 3-space.*

As is well known, any Riemann surface of genus one is conformally equivalent to a parallelogram in the plane with opposite sides identified. The method used here utilizes surfaces which are approximately isometric to the canonical surfaces determined by parallelograms. The parallelogram for a given conformal class may be picked in a standard way. We may take the vertices at the points $0, 2\pi, \omega, \omega + 2\pi$ in the complex plane. Then the parallelogram is determined by a single complex number ω. For any surface S conformally equivalent to this parallelogram with opposite sides identified, ω will be called a modulus of S, and the parallelogram a fundamental parallelogram of S. ω is not completely determined. A complete set of inequivalent canonical surfaces corresponds to the values of $\omega = \theta + i\lambda$ in the region

$$(1) \qquad -\pi < \theta \leqq \pi, \quad \theta^2 + \lambda^2 > 4\pi^2$$

or

$$0 \leqq \theta \leqq \pi, \quad \theta^2 + \lambda^2 = 4\pi^2 .$$

For each value of ω in this region a surface is needed.

A torus has a pure imaginary modulus which is easily computed. More generally, any surface with a plane of symmetry has pure imaginary modulus. Thus there are many ways in which one can construct a family of surfaces whose moduli fill the line $\theta = 0, \lambda \geqq 2\pi$.

For finding surfaces with $\theta \neq 0$, we may note first that under a reflection of space a surface with modulus $\theta + i\lambda$ is transformed into one with modulus $-\theta + i\lambda$. This means that if surfaces whose moduli represent all points of the region

---

Received February 23, 1960. This research was performed while the first author was supported by the United States Air Force under Contract No. AF 49(638)-42, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command.

[1] In our considerations compact surfaces of 3-space will be considered oriented by the outward pointing normal.

$$(1') \qquad 0 < \theta \leqq \pi, \quad \theta + \lambda^2 \geqq 4\pi^2$$

are available, then every point of (1) with $\theta < 0$ will be found among the moduli of the reflected surfaces.

One type of surface whose modulus can be computed is a canal surface, the envelope of a one-parameter family of spheres. In many cases it becomes difficult, however, to determine when the surface enveloped by a given family of spheres is really a good surface of genus one, with no undesired self-intersections. One two parameter family of canal surfaces has been given [2] which yields all values of ω in (1) with $|\theta| < \theta_0$, where $\theta_0 > 0$. More complicated families yield these moduli and also those for which $\lambda > \lambda_0$. In all the families which have been investigated by the authors, the surfaces for which ω is close to the vertex $\pi + (\pi + i\pi\sqrt{3})$ of (1) have self-intersections. Perhaps there is a region of values of ω near this point which cannot be realized by canal surfaces. However, by using the methods of Nash [4] as extended by Kuiper [3] it should not be hard to show that there exists an analytic surface whose modulus is arbitrarily close to any given modulus.

The method used here to prove the existence of embeddings for all moduli is to construct C^∞ surfaces which are approximations to singular surfaces. The singular surfaces used are composed of polygonal faces joined along edges and at vertices, and have the property that although points on different faces are distinguished, the faces all lie in the same plane in space, and may partly or wholly overlap. For each value of ω in a region including (1′) a singular surface is constructed. It is isometric be the canonical surface of modulus ω.

The singular surfaces are idealizations of figures which may be physically constructed by folding paper parallelograms and joining the edges. The physical model approaches the ideal surfaces as the thickness of the paper approaches zero. From this it follows that the singular surface can be approximated by a true surface. Such an approximation is given by the central plane of the paper.

1. The deformation lemma.[2] Suppose that we have a C^1 mapping of a surface S onto a surface S' such that all first derivatives of the mapping do not simultaneously vanish at any point of S. Then the dilation quotient D of the mapping is a function of position on S defined as follows: The image of an infinitesimal circle about the point P of S is an ellipse of definite eccentricity. $D(P)$ is the ratio of major to minor diameters in this ellipse. If S and S' are Riemann surfaces and the mapping is given in a neighborhood of P in terms of local uniformizers z, z' by $z' = f(z)$, then

[2] The methods used in this section are inspired by the work of Teichmuller [5] and Ahlfors [1].

$$D(P) = \frac{|f_z| + |f_{\bar z}|}{|f_z| - |f_{\bar z}|}\Bigg|_{z=z(P)} \tag{2}$$

if the mapping preserves orientation. For a conformal mapping $D = 1$. In general, $1 \leqq D \leqq \infty$.

Now let S and S' be surfaces of genus one, and let the mapping be one-to-one. If the moduli of S, S' are ω, ω' respectively, then the induced mapping of the fundamental parallelograms of S and S' has the form

$$z' \equiv f(z) \quad (\text{mod } 2\pi, \omega') ,$$

where f is a differentiable function of z and $\bar z$. The boundary values of $f(z)$ are related by the condition that equivalent values of z go into equivalent values of z'. We can choose the fundamental parallelogram of S' so that its sides are homotopic to the images under $S \to S'$ of the corresponding sides of the fundamental parallelogram of S. Then

$$\begin{aligned} f(z + 2\pi) &= f(z) + 2\pi , \\ f(z + \omega) &= f(z) + \omega' . \end{aligned} \tag{3}$$

After these preliminaries we may prove the following

LEMMA. *Let S, S' be Riemann surfaces of genus* 1. *Suppose there is a one-to-one mapping of S onto S' which is piecewise C^1, is conformal except on a region R of the fundamental parallelogram of S of area α, and has $D \leqq D_0 < \infty$. Then if $\omega = \theta + i\lambda$ is the modulus of S, S' has a modulus ω' such that*

$$|\omega - \omega'| < \sqrt{\eta(\eta + 2\lambda)} \tag{4}$$

where

$$\eta = \frac{\alpha}{2\pi D_0}(D_0 - 1)^2 .$$

Proof. Let $z = x + iy, z' = x' + iy'$. If P is the fundamental parallelogram of S and C its boundary, we have, by using (3),

$$\int_C f(z)dz = 2\pi(\omega - \omega') .$$

By Green's theorem,

$$\int_C f(z)dz = 2i\iint_P f_{\bar z} dxdy .$$

Since $f_{\bar z} = 0$ outside R,

$$-\pi i(\omega - \omega') = \iint_R f_{\bar z} dxdy \ .$$

Now apply Schwarz' inequality:

$$(5) \qquad \pi^2|\omega - \omega'|^2 < \iint_R dxdx \iint_R |f_{\bar z}|^2 dxdy \ .$$

To estimate the last integral, we use (2) to get

$$|f_{\bar z}|^2 = \frac{(D-1)^2}{4D}\{|f_z|^2 - |f_{\bar z}|^2\} = \frac{(D-1)^2}{4D}\frac{\partial(x', y')}{\partial(x, y)} \ .$$

Hence

$$\iint_R |f_{\bar z}|^2 dxdy < \iint_P \frac{(D_0-1)^2}{4D_0}\frac{\partial(x', y')}{\partial(x, y)} dxdy = \frac{(D_0-1)^2}{4D_0} \cdot 2\pi\lambda' \ ,$$

where $\lambda' = \operatorname{Im}\omega'$. Inserting this in (5),

$$|\omega - \omega'|^2 \leqq \frac{\alpha}{\pi^2}\iint_R |f_{\bar z}|^2 dxdy \leqq \eta\lambda' \ .$$

To get an inequality without λ' on the right, note that

$$(\lambda' - \lambda)^2 \leqq |\omega - \omega'|^2 \leqq \eta\lambda'$$

implies $\lambda' < \eta + 2\lambda$.

Thus

$$|\omega - \omega'| \leqq \sqrt{\eta\lambda'} < \sqrt{\eta(\eta + 2\lambda)} \ .$$

Observe that $|\omega - \omega'|$ is small, according to the lemma, in two cases: (1) if D_0 is close to 1, and (2) for fixed D_0, if α is small. Both cases will occur in the applications of the lemma.

2. The singular surfaces. The construction of a singular surface will be described in terms of operations on a paper parallelogram of modulus $\omega = \theta + i\lambda$. If ideal paper of zero thickness is used, the resulting surface is isometric to the canonical surface of modulus ω.

The first operation is to bend the parallelogram into a cylinder of radius 1 which fits together along the sides of length $|\omega|$. Glueing the cylinder together along the line where these sides meet gives the proper identification of the vertical sides of the fundamental parallelogram. The cylinder must be folded up so that the ends come into coincidence in the proper way.

If $\theta = 0$, the points at the ends of each generator must be identified. To do this, first flatten the cylinder by folding along two opposite generators. Then fold on the center line perpendicular to the generators, so that the two ends of the cylinder come together. Along the line where the ends meet, there are the edges of four layers of paper. If these are numbered in order of position, the proper identification of points is accomplished by joining sheet 1 to sheet 4 and sheet 2 to sheet 3.

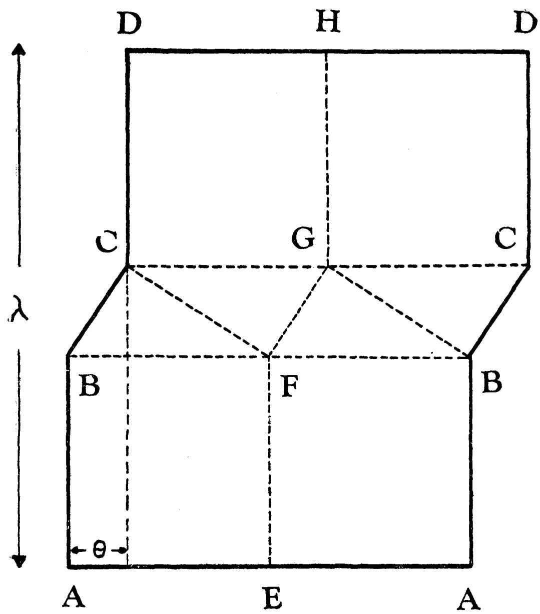

Figure IA

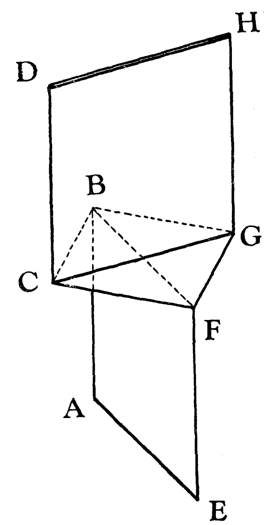

Figure IB

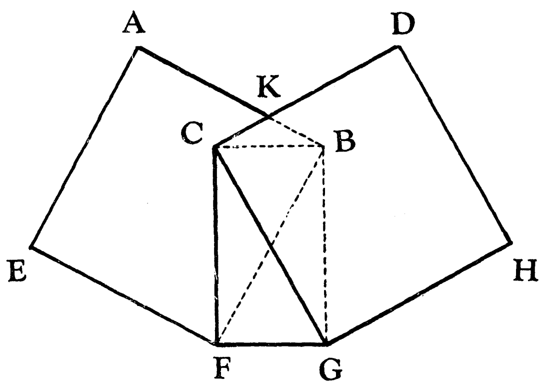

Figure IC

If $\theta \neq 0$, the point at the bottom of a generator must be identified with the point at the top of another generator, separated from the first by the angle θ. The cylinder will be folded flat in such a way that at each end it is folded along two opposite generators, but the generators used at the top are displaced from those used at the bottom by the angle θ. Then the two ends can be connected as they were for $\theta = 0$.

One such method of folding the cylinder is illustrated in figure 1. In 1(a) the surface of the cylinder is shown, developed on the plane. AB, EF and CD, GH are pairs of opposite generators, with CD to the right of AB by the distance θ. CGC and BFB are perpendicular to the generators. $AB = CD$, and the angles BCF, CFG, etc. are right angles. This determines the construction of 1(a). The cylinder may now be folded into a polygonal surface as shown in 1(b). Each line in 1(a) becomes one or more edges in 1(b). The part of the cylinder outside the lines CGC and BFB are flattened into the rectangles $DHGC$, $BFAE$. The part between these lines is flattened into the rectangle $BGFC$.

Next 1(b) is flattened out to give 1(c) by bending 1(b) along the lines CG, BF until all faces lie in the same plane. In figure 1(c) there

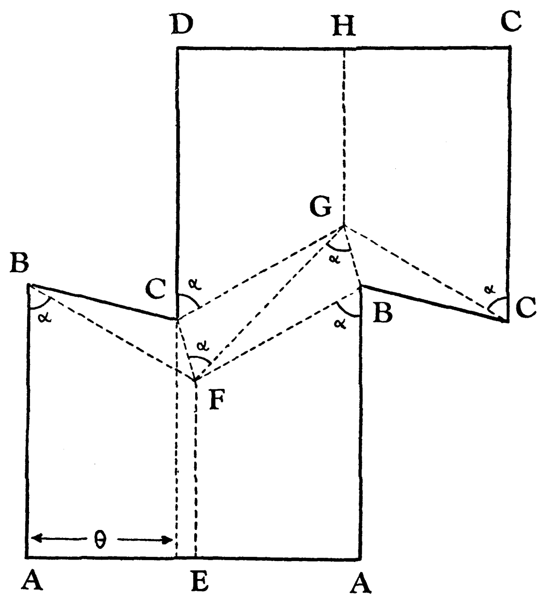

Figure 2 A

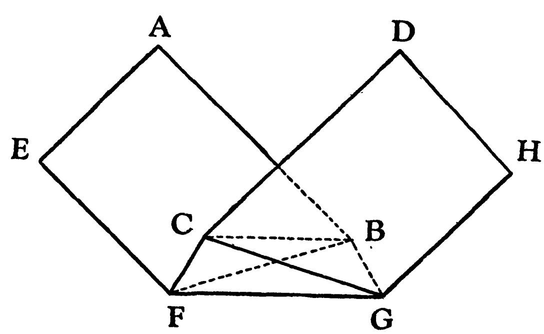

Figure 2 B

are six sheets, joined in appropriate ways along their edges. The flattened circles DHD and AEA are now in a suitable position and may be identified as when $\theta = 0$, after folding 1(c) along its vertical center line. This gives a singular surface consisting of twelve polygonal faces lying in the same plane. It is a Riemann surface isometric to the original parallelogram with its opposite sides identified.

A restriction must be placed on ω for this construction to work. λ must be sufficiently large that the lines AE, DH lie outside the rectangle $CBGF$ in 1(c). Also the lines CD and BA of 1(c) must extend at least as far as their intersection K. The first condition implies the second if $\theta \leqq \pi/2$, which is the only case in which this construction will be used. All the dimensions of 1(c) may be determined by elementary methods. The first condition on λ is

$$\lambda \geqq 3\sqrt{\theta(\pi - \theta)}\,. \tag{6}$$

For $0 < \theta \leqq \pi/2$, the model described may be used when λ satisfies this inequality.

Another model may be constructed as indicated in figure 2. The configuration $BFBCGC$ in 2(a) is determined by the angle

$$\alpha = ctn^{-1}[1 - \sqrt{2(1 - \theta/\pi)}]$$

which is marked in six locations. The condition $AB = CD$ determines the rest of the construction of 2(a). Figure 2(b) is analogous to 1(c). The central quadrilateral is now a trapezoid instead of a rectangle, and the lines AB, CD in 2(b) are perpendicular. The condition that AE and DH do not enter $FGBC$ in this diagram allows the construction of the singular surface to be completed as before. This gives the following condition on λ:

$$\lambda \geqq 2\pi - \theta \tag{7}$$

This model will be used for $\theta > \pi/2$. Note that the two models are the same for $\theta = \pi/2$.

By the two constructions just described, we have a family of singular surfaces associated one-to-one with the values of ω in the region

$$\mathrm{R} : \begin{cases} 0 < \theta < \pi, \\ \lambda \geqq \begin{cases} 3\sqrt{\theta(\pi - \theta)}\,, \ \theta \leqq \pi/2 \\ 2\pi - \theta, \ \theta \geqq \pi/2 \end{cases} \end{cases} \tag{8}$$

The surface $S_0(\omega)$ corresponding to ω has modulus ω. The curve which forms the lower boundary of this region is below the circular arc $\lambda = \sqrt{4\pi^2 - \theta^2}$ which forms the lower boundary of (1′). Thus all the points of (1′) for which $\theta \neq \pi$ are in R. Another construction to be used later for the case $\theta = \pi$ depends on the fact that R is not bounded away from 0, and the lower boundaries of (1′) and R do not meet even at $\theta = \pi$. Note that $S_0(\omega)$ depends continuously on ω.

3. The C^∞ surfaces. The singular surface $S_0(\omega)$ will be transformed into a family of C^∞ surfaces $S(\omega, \vartheta)$, $0 < \vartheta < \vartheta_1(\omega)$ with modulus $\Omega(\omega, \vartheta)$ which approaches ω as $\vartheta \to 0$. These surfaces will also depend on a large positive constant K.

For convenience, suppose that $S_0(\omega)$ is situated in a horizontal plane. In this plane, around each vertex V_k of $S_0(\omega)$ construct a circular disc γ_k^0 of radius $K\vartheta$. For each edge of $S_0(\omega)$, consider the strip of the plane extending the distance ϑ to each side of the edge. Construct the regions ζ_e^0 interior to these strips and exterior to the circles γ_k^0. These are well defined if $K > 1$. For fixed K, the γ_k^0's and ζ_e^0's will be disjoint for sufficiently small $\vartheta : \vartheta < \vartheta_1(\omega)$, if none of the angles in the figure are too small. This will be true if θ is not too close to 0 or $\pi : \theta_1(K) \leqq \theta \leqq \theta_2(K)$, where $\theta_1(K) \to 0$ and $\theta_2(K) \to \pi$ as $K \to \infty$.

Form ζ_e from ζ_e^0 by replacing the bounding circular arcs by chords. Form γ_k from γ_k^0 by removing the segments which have been added to the ζ_e's.

Let C_k be the cylindrical region of space with base γ_k, and let R_e be the cylindrical region with base ζ_e. The C^∞ surface $S(\omega, \vartheta)$ will be constructed above the plane so that in the mapping from $S_0(\omega)$ to $S(\omega, \vartheta)$, (1) the parts of the faces of $S_0(\omega)$ which lie outside the γ_k's and the ζ_e's are each translated upwards by a certain amount, (2) the pieces of $S_0(\omega)$ in ζ_e are mapped isometrically into R_e, and (3) the pieces in γ_k are mapped into C_k.

Let the twelve faces of $S_0(\omega)$ be numbered in order from bottom to top. The jth face is cut into a number of sections by the boundaries of the γ_k's and ζ_e's. Raise each section bounded entirely by these boundaries to the height $j/12$ above the plane of $S_0(\omega)$. This includes the vertical translation referred to above. The remainder of $S_0(\omega)$ consists of strips of surface containing edges, and a finite number of regions in each γ_k.

Each edge strip S_m has width 2ϑ. In extending the mapping to S_m, the sides of the strip must go into parallel horizontal lines at a distance less than ϑ, one above the other. Opposing points on the two sides are to go into points of the parallel lines which lie on the same vertical line, and the mapping is to be isometric along each side.

Such a mapping can be constructed by mapping S_m isometrically onto a cylindrical surface of width 2ϑ bounded by the two parallel lines as generators. For $S_m \subset \zeta_e$, its image shall lie in R_e. This surface can be so chosen that the image of S_m and the adjacent sections of faces of $S_0(\omega)$ lies on a C^∞ cylindrical surface. If several of the S'_m's lie along the same edge of $S_0(\omega)$, use the same cylindrical surface for each of their images. Then the cylindrical mapping may be extended through the intervening regions which lie in the γ_k's.

The image of the region around each edge of $S_0(\omega)$ may be constructed in turn, so as not to meet any of the parts of $S(\omega, \vartheta)$ previously constructed, if we take the edges in the right order, for example first all the edges of the top face, than all the unconnected edges of the next face, and so on. It remains to construct the images of the regions about vertices of $S_0(\omega)$.

Let v_j be the region about a vertex of $S_0(\omega)$, bounded by the boundary of γ_k. It is to be mapped into a piece of C^∞ surface in C_k which joins to previously constructed parts of $S(\omega, \vartheta)$ at the boundary of C_k, so as to make a C^∞ surface. Let V_j be any piece of surface with these properties. Then a mapping from v_j to V_j can be made which agrees at the boundary with the mapping already constructed, and has a bounded dilation quotient. $S(\omega, \vartheta)$ and the mapping from $S_0(\omega)$ will be completed when this has been done for each v_k.

The resulting mapping is isometric, hence conformal, except in the regions of v_k. The dilation quotient has an upper bound $D_0(\omega, \vartheta)$. If $\Omega = \Omega(\omega, \vartheta)$ is the modulus of $S(\omega, \vartheta)$, then by the lemma

$$|\Omega - \omega| < \sqrt{\eta(\eta + 2\lambda)}\,, \tag{9}$$

where

$$\eta = \frac{(D_0 - 1)^2}{2\pi D_0} \sum_k \text{Area}\,(v_k) < \frac{(D_0 - 1)^2}{2\pi D_0} \cdot 12\pi K^2\vartheta^2\,,$$

$$< 6K^2\vartheta^2 \frac{(D_0 - 1)^2}{D_0}\,, \tag{10}$$

since there are twelve vertices, and each v_k is a subregion of a disc of radius $K\vartheta$.

Suppose that this construction is made only for particular values of ω and ϑ, e.g. $\omega = \omega_1 = \theta_1 + i\lambda_1$, $\vartheta = \vartheta_1$. Then from this we can derive $S(\omega, \vartheta)$ for other values of ω and ϑ in a useful way.

First, let ϑ vary. For all sufficiently small ϑ, we may take the vertex pieces V_k to be similar to those for ϑ_1, and take the cylindrical strips which contain the images of the edges to have a cross-section which is similar to the cross-section for $\vartheta = \vartheta_1$. If we take the mapping from v_k to V_k to be that which is induced by this similarity, then D_0 in (10) is independent of ϑ.

Next let λ vary. In $S_0(\omega)$, certain sides change in length, but all angles at the vertices are unchanged. Looking at the skeleton of $S(\omega, \vartheta)$, consisting of the V_k's and the cylindrical strips which contain the images of edges, to get a skeleton for $S(\theta_1 + i\lambda, \vartheta)$ it is sufficient to change appropriately the lengths of some of the strips. If $S(\theta_1 + i\lambda, \vartheta)$ is formed in this way, D_0 can be made independent λ.

To go to a value of ω with $\theta \neq \theta_1$, the angles at the vertices must change as well as lengths. Looking again at the skeleton of $S(\omega_1, \vartheta)$, $V_j \subset C_k$ must be transformed so that the adjacent cylindrical strips are rotated into different positions about the axis of C_k. We may impose on V_j the condition that it does not intersect the axis of C_k. Then the desired result may be obtained by performing a transformation of C_k which rotates points about the axis through a variable angle, is isometric in the sectors adjacent to the cylindrical strips, and is C^∞ with non-vanishing Jacobian at all points off the axis.

Let t_ϑ be the transformation which takes the vertex sections $V_k(\theta_1)$ into $V_k(\theta)$. Since all the angles involved are continuous functions of θ, we may choose t_θ so that $t_{\theta'}\, t_\theta^{-1}$ is a mapping whose dilation quotient has a bound $D(\theta, \theta')$ such that

$$\lim_{\theta' \to \theta} D(\theta, \theta') = 1$$

and $t_{\theta'}\, t_\theta^{-1}$ approaches the identity as $\theta' \to \theta$.

This transformation of the V_k's may be extended to a transformation of $S(\omega_1, \vartheta)$ into $S(\omega, \vartheta)$, by transforming the cylindrical strips of $S(\omega_1, \vartheta)$ linearly into the strips of $S(\omega, \vartheta)$, and then extending the mapping over the plane faces. Since the lengths of the edges of $S_0(\omega)$ are continuous functions of θ, if the extension to the faces is done properly the induced mapping $T_{\omega,\omega'}$ of $S(\omega, \vartheta)$ into $S(\omega', \vartheta)$ will have a dilation with a bound $D(\omega, \omega', \vartheta)$ such that

$$\lim_{\omega' \to \omega} D(\omega, \omega', \vartheta) = 1 \tag{11}$$

The mapping from $S_0(\omega)$ to $S(\omega, \vartheta)$ may be chosen so that $D_0(\omega, \vartheta)$ is a continuous function of ω. For this, it is necessary to map v_j on V_j in the right way.

A mapping of $S_0(\omega_1)$ onto $S_0(\omega)$ is associated with the mappings already described:

$$S_0(\omega_1) \longrightarrow S(\omega_1, \vartheta) \xrightarrow{T_{\omega_1\omega}} S(\omega, \vartheta) \longrightarrow S_0(\omega)\,.$$

Thus the map of $v_j(\omega)$ on $V_j(\omega)$ will be determined by that for $\omega = \omega_1$ and a mapping of $v_j(\omega_1)$ on $v_j(\omega)$. Since $T_{\omega,\omega'}$ has the property (11), $D_0(\omega, \vartheta)$ will be a continuous function of ω if the transformation between the v_k's also has this property.

To transform a v_j it is convenient to look at this region unfolded. It is a disc with several segments cut off, and is to be transformed into another such disc. The mapping is given on the boundary, and is to be extended to the interior so that the bound $D'(\omega, \omega')$ for the dilation quotient of the mapping from $v_j(\omega)$ to $v_j(\omega')$ approaches 1 as $\omega' \to \omega$. One way to get such an extension is to take the mapping given in rectangular coordinates by harmonic functions with the proper boundary values.

4. The existence theorem for $0 < \theta < \pi$. Any modulus $\omega_0 = \theta_0 + i\lambda_0$ in (1′) with $0 < \theta < \pi$ lies in the interior of R. Pick K and Λ so large that $\theta_1(K) < \theta_0 < \theta_2(K)$ and $\Lambda > \lambda_0$. Then ω_0 lies in the interior of the closed subregion R_1 of R for which

$$\theta_1(K) \leqq \theta \leqq \theta_2(K) ,$$
$$\lambda \leqq \Lambda$$

Let the distance from ω_0 to the boundary of R_1 be greater than ε ($\varepsilon > 0$)

In §3, we have constructed surfaces $S(\omega, \vartheta)$ for each $\omega \in R_1$ and $\vartheta < \bar{\vartheta} = \min \vartheta_1(\omega)$, $\omega \in R_1$. By (11) and the lemma, $\Omega(\omega, \vartheta)$ is a continuous function of ω, so $|\Omega(\omega, \vartheta) - \omega|$ is bounded on R_1, for each ϑ. An explicit bound is given by (9):

$$|\Omega(\omega, \vartheta) - \omega| < C\vartheta ,$$

where, setting $D_1 = \max D_0(\omega, \vartheta)$, $\omega \in R_1$,

$$C = \sqrt{6K^2 \frac{(D_1 - 1)^2}{D_1} [6K^2\bar{\vartheta}^2 \frac{(D^1 - 1)^2}{D_1} + \Lambda]} .$$

Take $\vartheta < \varepsilon / C$. Then in the mapping of R_1 by $\omega \to \Omega(\omega, \vartheta)$ each point is moved by less than ε. Hence the image of the boundary is a curve which winds around ω_0 once. It follows that $\Omega(\omega, \vartheta) = \omega_0$ for some $\omega \in R_1$.

5. The existence theorem for $\theta = \pi$. If a family of singular surfaces is available which varies continuously over a region of moduli containing the right-hand boundary of (1′) in its interior, then the procedure of §§ 3 and 4 may be applied to prove the existence theorem for $\theta = \pi$.

As observed in § 2, the lower boundary of R is at a positive distance from the lower boundary of (1′). Let this distance be d. Also, R contains

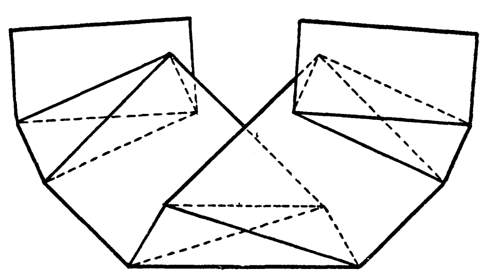

Figure 3

values of ω for which $|\omega| < d/2$. Let $\omega_2 = \theta_2 + i\lambda_2$ be such a value. Now take the singular surface of § 2 for modulus ω, constructed up to the point of figure 1(c) or 2(b). Join to each of its open ends the corresponding figure for $\omega = \omega_2$, as illustrated in figure 3. This gives a folded cylinder of length $\lambda + 2\lambda_2$. When it is folded over the vertical center line of figure 3 and has its ends joined properly, it is a singular surface with modulus $\omega + 2\omega_2$. If this is done for each $\omega \in R$, we get for a region of moduli the region R shifted up and to the right by a positive distance less than d, with singular surfaces varying continuously over this region. The line $\theta = \pi$, $\lambda \geqq \sqrt{3}\,\pi$ lies in the interior, as required.

References

1. L. Ahlfors, *On Quasiconformal mappings*, Journal d'Analyse Mathematique (Jerusalem), **III** (1954), 1-58.

2. A. Garsia, *The calculation of conformal parameters for some imbedded Riemann surfaces*, Pacific J. Math., **10** (1960), 121-165.

3. N. H. Kuiper, *On C^1 isometric imbeddings*, Proc. Kon. Ac. Wet. Amsterdam A 58 (Indagationes Mathematicae) no. **4** (1955), 545-556.

4. J. Nash, *C^1 isometric imbeddings*, Ann. of Math., **60** (1954), 383-396.

5. Oswald Teichmüller, *Beweis der Analytischen Abhängigkeit des Konformen Moduls Einer Analytischen Ringflächenschar Von der Parametern*, Deutsche Math., **7** (1944), 309-336.

UNIVERSITY OF MINNESOTA
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

WEAK COMPACTNESS AND SEPARATE CONTINUITY

IRVING GLICKSBERG

1. For a locally compact space X let $C(X)$ denote the Banach space of all bounded continuous complex valued functions on X, $C_0(X)$ the subspace of functions vanishing at infinity, so that the adjoint $C_0(X)^*$ consists of all finite complex regular Borel measures on X. In a natural fashion, we may view $C(X)$ as a subspace of $C_0(X)^{**}$.

When X is compact Grothendieck [6; Th. 5] has shown that a bounded set $K \subset C(X)$ is compact in the weak topology if (and of course only if) K is compact in the topology of pointwise convergence on X, and then both topologies, being comparable, coincide on K. In some recent work the writer was led to a simple corollary of Grothendieck's result which yields the significance, when X is only locally compact, of compactness in $C(X)$ under pointwise convergence:

1.1. *Let K be a bounded subset of $C(X)$, X locally compact. Then K is compact in the topology of pointwise convergence on X (if and) only if K is compact in the weak* topology of $C_0(X)^{**}$* [4, 5.1].

Again both topologies coincide on K. A direct corollary of 1.1 is

1.2. *Let X and Y be locally compact spaces, and f a bounded complex function on $X \times Y$ which is separately continuous, i.e., for which all the maps*

$$x \to f(x, y) \text{ and } y \to f(x, y)$$

are continuous. Then for $\mu \in C_0(X)^*$,

$$y \to \int f(x, y)\mu(dx)$$

is continuous [4, 5.2].

The continuity obtained in 1.2 allows one to form the iterated integral

$$\iint f(x, y)\mu(dx)\nu(dy), \qquad \mu \in C_0(X)^*, \ \nu \in C_0(Y)^*, \tag{1.21}$$

and thus one can extend the notion of convolution of a pair of finite measures to a locally compact semigroup S in which the operation is only separately continuous. Moreover 1.2 shows (1.21) is identical with

Received March 24, 1960.

$$\iint f(x, y)\nu(dy)\mu(dx) \tag{1.22}$$

so that convolution is commutative if S is. Consequently we show in § 4 how some results of [3] extend to the separately continuous situation; these in turn yield an analogue of the Weyl equidistribution theorem which applies to weakly almost periodic functions on locally compact abelian groups (4.6 below).

Although the fact will not be needed in what follows, note that 1.1 is actually a weak compactness criterion for the complete locally convex space $C(X)_\beta$ formed from $C(X)$ by endowing it with the strict topology (cf. [0]). For since the dual $C(X)_\beta^*$ consists precisely of the measures in $C_0(X)^*$, the weak topology of $C(X)_\beta$ is just the weak* topology in 1.1 and the bounded sets of $C(X)$ and $C(X)_\beta$ coincide. But as a consequence the topology of pointwise convergence on $C(X)$, when restricted to bounded sets, shares some properties of weak topologies of complete locally convex spaces: conditionally countably compact sets are conditionally compact, and have compact convex hulls.

Notation. For a function f, $f \mid E$ will denote its restriction to E, while for a set K of functions, $K \mid E$ will denote the corresponding set of restrictions. $C(X)_p$ and $C(X)_{w^*}$ will denote $C(X)$ in the topology of pointwise convergence on X, and in the weak* topology of $C_0(X)^{**}$, respectively. In general X and Y will denote locally compact (Hausdorff) spaces, and, for a function f on $X \times Y$, $f(\cdot, y)$ will be its section $x \to f(x, y)$ (with $f(x, \cdot)$ defined analogously). As we have indicated f is separately continuous only if all of sections are continuous. Other notation is standard.

2. Since the proofs of 1.1 and 1.2 (given in [4]), are quite short, we shall include them for completeness.

Consider 1.1, and let $\mathscr{F}$ be an ultrafilter on K. $\mathscr{F}$ converges to some f_0 in K in $C(X)_p$, and we need only show $\mathscr{F}$ converges to f_0 in $C(X)_{w^*}$. On the bounded set K the weak* topology is defined by the dense set of measures μ with compact carriers C_μ, so we need only show $\int f_0 d\mu = \lim_{\mathscr{F}} \int f d\mu$ for such μ. But $K \mid C_\mu$ is compact in $C(C_\mu)_p$ and thus, by Grothendieck's theorem, compact in the weak topology, and both topologies coincide on $K \mid C_\mu$. Clearly then $\int f_0 d\mu = \lim_{\mathscr{F}} \int f d\mu$ as desired.

In order to prove 1.2 we have to show the map $y \to f(\cdot, y)$ of Y into $C(X)_{w^*}$ is continuous. But it is a continuous map into $C(X)_p$, so that any compact neighborhood V of $y_0 \in Y$ has an image which is compact in the weak* topology by 1.1. And since the weak* topology coincides on this image with that of pointwise convergence, the desired continuity is immediate.

As a first application of 1.2 we note the following simple proof of the well known fact (due to Krein and Smulian) that if K is a weakly compact subset of a complete locally convex linear space E, then the closed convex hull $\mathscr{C}(K)$ is weakly compact. Take, as our X and Y of 1.2, K in the weak topology, and the polar $V^0 \subset E^*$ of a neighborhood V of 0 in E, in the weak* topology. Since $x \to \langle x, x^* \rangle$ and $x^* \to \langle x, x^* \rangle$ are each continuous in the appropriate topologies, by 1.2 we have, for $\mu \in C(K)^*$,

$$x^* \to \int \langle x, x^* \rangle \mu(dx) \tag{2.11}$$

continuous on V^0. Since V is an arbitrary neighborhood of 0, and E is complete, a well known result of Grothendieck [5] shows (2.11) represents a weak* continuous functional on E^*, and thus there is an x_μ in E satisfying

$$\langle x_\mu, x^* \rangle = \int \langle x, x^* \rangle \mu(dx) , \qquad x^* \in E^* . \tag{2.12}$$

Let $N = \{\mu : \mu \in C(K)^*, \mu \geqq 0, \mu(K) = 1\}$, a weak* compact convex subset of $C(K)^*$, and endow N with the weak* topology. Since

$$\mu \to \int \langle x, x^* \rangle \mu(dx)$$

is clearly continuous on N, (2.12) implies $\mu \to x_\mu$ is a continuous map from N into E under the weak topology; thus the range of this map is a convex weakly compact subset of E, which clearly contains K. Since $\mathscr{C}(K)$ is weakly closed by Mazur's theorem, this is all we need to show.

3. As was noted in the introduction, 1.2 allows one to form the iterated integral

$$\iint f(x, y) \mu(dx) \nu(dy), \qquad \mu \in C_0(X)^*, \nu \in C_0(Y)^* ,$$

for any bounded separately continuous f. The desirable interchangability of the order of integration would of course be immediate once f is, say, locally Borel measurable; however the writer is not aware of any general answer to the question of measurability of separately continuous functions (a special case is covered in [7, § 39]). Nevertheless the independence of order is easily obtained from 1.2.

THEOREM 3.1. *Let f be a bounded separately continuous complex function on $X \times Y$. Then*

$$\iint f(x, y) \mu(dx) \nu(dy) = \iint f(x, y) \nu(dy) \mu(dx), \quad \mu \in C_0(X)^*, \nu \in C_0(Y)^* . \tag{3.11}$$

Proof. Let μ be fixed. For K a compact subset of Y let $E_K = \{\nu: \|\nu\| \leqq 1,\ \nu \text{ vanishes on subsets of } K'\}$. Clearly (3.11) holds when ν is a finite linear combination of point masses; since these are weak* dense in E_K we can prove (3.11) holds for all ν in E_K by showing both sides are continuous functions on E_K, taken in the weak* topology of $C_0(Y)^*$. By Urysohn's lemma this topology coincides on E_K with the weak* topology of $C(K)^*$, and thus the left side of (3.11) is continuous since the inner integral represents an element of $C(K)$. On the other hand E_K is compact in the weak* topology of $C(K)^*$ and

$$(x, \nu) \to \int f(x, y)\nu(dy)$$

defines a bounded separately continuous function on $X \times E_K$ (by 1.2 and the definition of the weak* topology). Thus 1.2 implies

$$\nu \to \iint f(x, y)\nu(dy)\mu(dx)$$

is continuous on E_K.

Consequently (3.11) holds for any given μ, and any ν with compact carrier. Since such ν are strongly dense in $C_0(X)^*$, (3.11) follows.

4. Let S be a compact space which is also a semigroup (group), and suppose the operation is separately continuous:

$$x \to xy \text{ and } y \to xy$$

are continuous; then we shall call S a compact separately continuous semigroup (group). For μ and ν in $C(S)^*$ we can form the convolution of μ and ν, an element $\mu\nu$ of $C(S)^*$, by virtue of the Riesz representation theorem and 1.2:

$$\int f(x)\mu\nu(dy) = \iint f(xy)\mu(dx)\nu(dy), \qquad f \in C(S) .$$

Convolution is easily seen to be associative, and endowing $C(S)^*$ with its weak* topology, separately continuous (by 3.1). Moreover 3.1 shows convolution is commutative when S is.

Let $\tilde{S} = \{\mu: \mu \in C(S)^*, \mu \geqq 0, \mu(S) = 1\}$; $\tilde{S}$ forms a compact separately continuous semigroup under convolution and the weak* topology. In [3] the writer determined the subgroups of $\tilde{S}$ when S is also jointly continuous; in the present section we shall see how some of the results of [3] extend to the separately continuous situation. (We might remark that compact separately continuous semigroups arise naturally in the study of weakly almost periodic functions on, for example, the real line (cf. [2])).

That most of these results carry over to the separately continuous situation is due to the consequences of Grothendieck's theorem given above. We shall also make mild use[1] of a fact due to Ellis [1] which can be obtained, interestingly enough, from Grothendieck's result [2, Appendix]: *a compact separately continuous group is a compact topological group.* In particular any *closed* algebraic subgroup of S is a compact topological group. (However an algebraic subgroup need not have its closure an algebraic subgroup, as in the jointly continuous case.)

To begin, let us note some distinctions between the present, separately continuous, situation, and that of [3], preserving, insofar as possible, the notation of [3]. When S is separately continuous, only the same is true of $\tilde{S}$ in general. But all of the ideal structure used in [3] continues to hold (with one exception: (1.11) of [3] fails); in particular every abelian separately continuous compact semi-group S contains a least ideal ($\bigcap_{x \in S} xS$) which is closed, a group, and thus a compact topological group. (In [3] we allowed S to be abelian, or a group; by virtue of the result cited above nothing new is obtained by allowing S to be a group here, and we shall insist that S be abelian in all but our first result.) The following is, in modified form, the key lemma of [3].

LEMMA 4.1. *Let S be a compact separately continuous semi-group, and let $\mu, \nu \in \tilde{S}$. Then*

$$\text{carrier } \mu\nu = [(\text{carrier } \mu)(\text{carrier } \nu)]^- . \tag{4.11}$$

Proof. The proof given in [3, Lemma 2.1] with $A \cdot B$ replaced by the right side of (4.11) shows the right side has $\mu\nu$-measure 1. To see that any open set W which meets the right side of (4.11) has $\mu\nu(W) > 0$, we argue as follows.

Let $x_0 y_0 \in W$, $x_0 \in$ carrier μ, $y_0 \in$ carrier ν. Then if $f \in C(S)$ vanishes off W while $f(x_0 y_0) = 1$, $0 \leqq f \leqq 1$, we have $\int f(xy_0)\mu(dx) > 0$ since $x \to f(xy_0)$ is positive near $x = x_0$. Since $y \to \int f(xy)\mu(dx)$ is continuous by 1.2, and positive at $y = y_0$,

$$0 < \iint f(xy)\mu(dx)\nu(dy) = \int f(z)\mu\nu(dz) \leqq \mu\nu(W) .$$

Consequently the right side of (4.11) is indeed carrier $\mu\nu$.

In the remainder of this section *we assume that S is an abelian compact separately continuous semigroup.*

[1] Essentially we use this to assert that μ in 4.2, when shown to be an invariant normalized measure on a separately continuous compact group, is the Haar measure; of course this could easily be avoided.

THEOREM 4.2. *Let $\mu^2 = \mu \in \tilde{S}$. Then carrier μ is a compact subgroup of S, and μ its Haar measure.*

If $H =$ carrier μ, then 4.1 shows $H^{2-} = H$, and scrutiny of the proof of [3, Th. 2.2] shows this is an adequate replacement for $H^2 = H$. (Note that 1.2 must be used to obtain the continuity of f'.)

THEOREM 4.3. *Let Γ be an algebraic subgroup of $\tilde{S}$. Then $G = \bigcup_{\mu \in \Gamma}$ carrier μ is an algebraic subgroup of S. If η is the identity of Γ, $g =$ carrier η is a compact topological group, η its Haar measure, and Γ the set of G-translates of η. Furthermore if Γ is closed, G is closed.*

Proof. G is algebraically a subsemigroup of S by 4.1, while g is a compact group and η its Haar measure by 4.2. Let e be the identity of g. Then for $\mu \in \Gamma$, $x \in$ carrier $\mu = [g \text{ carrier } \mu]^-$ implies $ex = x$ since this holds for x in g carrier μ. Consequently e acts as an identity on G.

Again let $\mu \in \Gamma$, $x \in$ carrier μ, $z \in$ carrier μ^{-1}; then $zg \subset$ carrier μ^{-1} by 4.1, so $xzg \subset$ (carrier μ)(carrier μ^{-1}) $\subset g$, and thus $g = (xzg)g = xzg$. Consequently there is a y in zg for which $xy = e$ and G is a group. Moreover $x^{-1} = y \in zg$ so $z \in x^{-1}g$; since z was any element of carrier μ^{-1}, carrier $\mu^{-1} \subset x^{-1}g = yg \subset zg \subset$ (carrier μ^{-1})$g \subset$ carrier μ^{-1}. Thus carrier μ^{-1} $= zg$ for any $z \in$ carrier μ^{-1}, or carrier $\mu = xg$, for any x in carrier μ, and carrier μ is a coset of g in G. Now

$$\int f(z)\,\mu(dz) = \iint f(xy)\eta(dx)\mu(dy), \qquad f \in C(S)\,,$$

since $\mu = \eta\mu$. Since $y \to \int f(xy)\eta(dx)$ is constant on carrier μ,

$$\int f(z)\,\mu(dz) = \int f(xy)\eta(dx)$$

for any y in carrier μ. Thus μ is exactly the translate to yg of η.

Finally suppose Γ is closed. If $x \in G^-$ we can find nets $\{x_\delta\}$ and $\{\mu_\delta\}$ for which $x_\delta \to x$, $x_\delta \in$ carrier μ_δ, $\mu_\delta \in \Gamma$ and $\mu_\delta \to \mu \in \Gamma$. If $x \notin$ carrier μ then $xg \cap$ carrier $\mu = \phi$ and there is an f in $C(S)$, $0 \leqq f \leqq 1$, which is 1 on xg and 0 on carrier μ. Since

$$y \to \int f(yz)\eta(dz)$$

is continuous by 1.2, and assumes the value 1 for y in xg, we have

$$\frac{1}{2} \leqq \int f(x_\delta z)\eta(dz) = \int f(z)\mu_\delta(dz) \qquad \text{for } \delta \geqq \delta_0,$$

despite the fact that

$$\int f(z)\mu_\delta(dz) \to \int f(z)\mu(dz) = 0 .$$

Thus $x \in \text{carrier } \mu \subset G$, and G is closed, completing our proof.

Actually we can obtain all of the analogous result (Th. 2.3) of [3]; it is easy to see that if Γ is closed (as [3] required) then the weak* closed convex hull $\mathscr{C}(\Gamma)$ of Γ is the image of $(G/g)^\sim$, using exactly the map T_η of [3, 2.3] (alternatively we could note that our measures all lie on a compact topological group G, and apply 2.3 of [3]).

THEOREM 4.4. *Let Σ be a closed subsemigroup of $\tilde{S}$ with least ideal $\mathscr{I}$; let*[2] $S_1 = (\bigcup_{\mu \in \Sigma} \text{carrier } \mu)^-$, *with least ideal I. Then* $I = \bigcup_{\mu \in \mathscr{I}} \text{carrier } \mu$.

Proof. Since $\mathscr{I}$ is a closed subsemigroup of Σ, and thus of $\tilde{S}$, by 4.3, $G = \bigcup_{\mu \in \mathscr{I}}$ carrier μ is a closed subgroup of S, and thus of S_1. Let $S_0 = \bigcup_{\mu \in \Sigma}$ carrier μ, and algebraic subsemigroup of S with $S_0^- = S_1$.

Suppose xS_1 does not contain G for some x in S_1. Then since $y \in xS_1 \cap G$ implies $G = yG \subset xS_1G \subset xS_1$, $xS_1 \cap G = \phi$. Consequently there is an f in $C(S)$ which vanishes on xS_1 and is 1 on G. Since $x \in S_1 = S_0^-$, there is a net $x_\delta \to x$, $x_\delta \in \text{carrier } \mu_\delta$, $\mu_\delta \in \Sigma$. For ν in $\mathscr{I}$, x carrier $\nu \subset xS_1$, so that $\int f(xy)\nu(dy) = 0$, and therefore $\int f(x_\delta y)\nu(dy) \to 0$ by 1.2. On the other hand $\mu_\delta \nu \in \mathscr{I}$ so that x_δ carrier $\nu \subset$ carrier $\mu_\delta \nu \subset G$, and $\int f(x_\delta y)\nu(dy) = 1$, a contradiction, whence we conclude that $G \subset xS_1$ for all x in S_1. Thus $G \subset I = \bigcap_{x \in S_1} xS_1$.

Now for x in S_0 and ν in $\mathscr{I}$, the fact that x carrier $\nu \subset G$ shows $xG \subset G$; for y in G then $xy \in G$ for all x in S_1 since G is closed and $x \to xy$ continuous. Consequently $xG \subset G$, all x in S_1, and G is an ideal in S_1; of course G must then contain the least ideal I, whence $G = I$ and our proof is complete.

By virtue of 4.4 and the remark immediately preceding it we obtain, by exactly the proof of [3, 3.2],

THEOREM 4.5. *Let $\mu \in \tilde{S}$. Then $(1/N)\sum_{n=1}^N \mu^n \to$ Haar measure on the least ideal of the closed subsemigroup of S generated by carrier μ.*

For the proofs of some of our next remarks (and for definitions of the basic entities involved) the reader is referred to [2]. Let G be a locally compact abelian group. Then the weakly almost periodic functions on G form a closed translation invariant subalgebra $W(G)$ of $C(G)$ containing $C_0(G)$. Moreover $W(G)$ is isometrically isomorphic to $C(G^w)$, where G^w is a compact abelian separately continuous semigroup, the

[2] Separate continuity (applied twice) is sufficient to guarantee that the closure of an algebraic subsemigroup is a subsemigroup.

weakly almost periodic compactification of G, in which G forms (topologically[3] and algebraically) a dense open subgroup; the elements in $W(G)$ are just the restrictions, to G, of the elements of $C(G^w)$. (G^w is *not* jointly continuous, or a group, unless G is compact.) Naturally each finite measure μ on G induces an element μ' of $C(G^w)^*$, and $\mu \to \mu'$ is easily seen to preserve convolution, norm and order; in particular $\mu \geqq 0$, $\|\mu\| = 1$ imply $\mu' \in \widetilde{G}^w$. If we define the carrier, in G, of such a non-negative μ to be the closed complement of the union of all open sets of μ-measure zero, then carrier μ' in G^w contains the carrier of μ (since open sets in G remain open in G^w, and $C_0(G) \subset W(G)$). Finally let the translate $R_g f$ of f be defined by $R_g f(g') = f(g'g)$, g, g' in G, f in $W(G)$. We need only apply 4.5 to $S = G^w$ and μ' to obtain

THEOREM 4.6. *Let G be a locally compact abelian group, and let $\mu \geqq 0$ be an element of $C_0(G)^*$ of norm 1. Then there is a non-negative functional F of norm 1 on $W(G)$ for which*

$$\frac{1}{N}\sum_{n=1}^{N}\int_G f(g)\mu^n(dg) \to F(f)\,, \qquad f \text{ in } W(G)\,,$$

and $F(R_g f) = F(f)$ for all g in the carrier of μ.

Here μ^n is, of course, the ordinary n-fold convolve of μ. As the reader will observe, a related result can be obtained when G is merely an abelian topological semigroup, as in [2].

Familiar results from ergodic theory suggest an alternative approach to 4.6, but yield a result of a different nature. Indeed if we define $\mu^n * f$, for f in $W(G)$, by $\mu^n * f(g) = \int f(gg')\mu^n(dg')$ then $\mu^n * f$ lies in the weakly compact closed convex hull K of the set of translates of f, and ergodic theory shows $(1/N)\sum_{n=1}^{N}\mu^n * f$ converges strongly to an f_1 in K with $\mu * f_1 = f_1$. From this alone it is not all apparent that f_1 should have the stronger invariance property that $R_g f_1 = f_1$ for g in the carrier of μ. But since $\mu^n * f(g) = \int R_g f(g')\mu^n(dg')$, 4.6 shows

$$\frac{1}{N}\sum_{n=1}^{N}\mu^n * f(g) \to F(R_g f)$$

and $f_1(g) = F(R_g f)$, so f_1 does indeed have the invariance property. Consequently we have proved

COROLLARY 4.7. *Let G be a locally compact abelian group, μ a non-negative measure of norm 1 on G. Then the operators*

[3] In the more general context of [2] G is only imbedded continuously in G^w; here $C_0(G) \subset W(G)$ guarantees the imbedding is open as well.

$$f \to \frac{1}{N}\sum_{n=1}^{N}\mu^n * f$$

on $W(G)$ *converge in the strong operator topology to a projection onto the manifold of functions left fixed by* $\{R_g$: g in the carrier of $\mu\}$.

4.8 REMARK. The remaining result of § 3 of [3], 3.5, extends to the present context with no change in proof; beyond this point, however, there are difficulties in obtaining extensions. In particular § 4 makes strong use of the now lacking property that the closure of an algebraic subgroup of $\tilde{S}$ be a group.

5. For $E \subset C(X)$ let $\sigma(C_0(X)^*, E)$ denote the least fine topology for which the maps

$$\mu \to \int f(x)\mu(dx), \qquad f \in E,$$

are continuous. When X is taken to be a locally compact abelian group G, 1.1 can be applied to some topologies on $C_0(G)^*$ by virtue of the Fourier-Stieltjes transformation. Let $G^\wedge$ denote the character group of G, $\hat{\mu}$ the Fourier-Stieltjes transform of $\mu \in C_0(G)^*$, $C_0(G)^{*\wedge}$ the set of all such transforms.

THEOREM 5.1. *Let* $K \subset C_0(G)^*$ *have a uniformly bounded set of Fourier-Stieltjes transforms. Then* K *is* $\sigma(C_0(G)^*, C_0(G^\wedge)^{*\wedge})$ *compact if (and of course only if)* K *is* $\sigma(C_0(G)^*, G^\wedge) = \sigma(C_0(G)^*, P(G^\wedge)^\wedge)$ *compact, where* $P(G^\wedge)$ *is the set of point masses on* $G^\wedge$. *Moreover* K *is then weak* compact if bounded.*

We need only note that by virtue of the identity

$$\int_G \hat{\nu}(g^{-1})\mu(dg) = \int_{G^\wedge}\hat{\mu}(\hat{g})\nu(d\hat{g})$$

(for $\mu \in C_0(G)^*$, $\nu \in C_0(G^\wedge)^*$), $\sigma(C_0(G)^*, C_0(G^\wedge)^{*\wedge})$ is the topology $\sigma(C_0(G)^{*\wedge}, C_0(G^\wedge)^*)$ (or the weak* topology of $C_0(G^\wedge)^{**}$) transported to $C_0(G)^*$, while $\sigma(C_0(G)^*, P(G^\wedge)^\wedge)$ corresponds in the same way to $\sigma(C_0(G)^{*\wedge}, P(G^\wedge))$ (or the topology of pointwise convergence). Thus 1.1 can be applied. For the final statement, note that $C_0(G^\wedge)^{*\wedge}$ contains $L_1(G^\wedge)^\wedge$, which defines the weak* topology on bounded subsets of $C_0(G)^*$.

REFERENCES

O. R. C. Buck, *Bounded continuous functions on a locally compact space*, Michigan J., **5** (1958), 95-104.

1. R. Ellis, *Locally compact transformation groups*, Duke Math. J., **24** (1957), 119-126.
2. K. de Leeuw and I. Glicksberg, *Applications of almost periodic compactifications*, to appear.

3. I. Glicksberg, *Convolution semigroups of measures,* Pacific. J. Math., **9** (1959), 51–67.
4. I. Glicksberg, *Uniform boundedness for groups,* to appear.
5. A. Grothendieck, *Sur la complétion du dual d'un espace localement convexe,* Comptes Rendues, **230** (1950), 605–606.
6. A. Grothendieck, *Critères de compacité dans les espaces fonctionnels généraux,* Amer. J. Math., **74** (1952), 168–186.
7. H. Hahn, *Reele Functionen I,* Leipzig, 1932.

UNIVERSITY OF NOTRE DAME

ON A CONJECTURE OF H. HADWIGER

BRANKO GRÜNBAUM

1. For any convex body (i.e., compact convex set with interior points) K in the Euclidean plane E^2 let $i(K)$ denote the greatest integer with the following property:

There exist translates K_n, $1 \leqq n \leqq i(K)$, of K such that

$$(1) \qquad \begin{aligned} &K \cap K_n \neq \phi && \text{for all } n; \\ &\operatorname{Int} K_n \cap \operatorname{Int} K_m = \phi && \text{for } n \neq m. \end{aligned}$$

It is well known (see e.g., Hadwiger [3]) that $7 \leqq i(K) \leqq 9$ for any $K \subset E^2$,[1] and that the bounds are attained (e.g., $i(K) = 7$ if K is a circle, $i(K) = 9$ if K is a parallelogram). Hadwiger conjectured,[2] moreover, that if K is not a parallelogram, then $i(K) = 7$.

We shall establish Hadwiger's conjecture in the following theorem:

If K is not a parallelogram, then $i(K) = 7$. Moreover, if 7 *translates o K satisfy conditions* (1) *then one of them coincides with K.*

In the proof we shall use some results on centrally symmetric convex sets; they are collected in § 2. The proof of the theorem follows in § 3. In § 4 we make some remarks on related problems in higher-dimensional spaces. § 5 contains some results on the related problem on the number of translates of a convex set needed to "enclose" the set.

2. Let K be any centrally symmetric plane convex body with the origin 0 as center. Then a Minkowski geometry, with norm $\| \;\; \|$, is defined in the plane, for which K is the unit cell.

We note the following propositions:

(i) *For any point x with $\|x\| = 1$ there exist points y, z satisfying* $\|y\| = \|z\| = \|x - y\| = \|y - z\| = \|x + z\| = 1$. (In other words, any $x \in \operatorname{Front} K$ is a vertex of at least one affine-regular hexagon whose vertices belong to Front K).

(ii) *Let x, y, z be different points belonging to Front K, such that the origin* 0 *does not belong to that open half-plane determined by x and y which contains z. Then $\|x - y\| \geqq \|x - z\|$, with equality taking place only in case y, z, and $(y - x)/\|y - x\|$ belong to a straight-line segment contained in Front K.*

Received November 3, 1958. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under contract No. AF 49(637)-253. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[1] Related results, pertaining to more general sets, are given in [4].

[2] Oral communication from Dr. H. Debrunner.

(iii) *Let* x, y, z, u *be different points belonging to* Front K, *such that* z *and* u *belong to an open half-plane determined by* x *and* y, *while* 0 *belongs to its complement. Then either* $\|x - y\| = \|z - u\| = 2$ *or* $\|x - y\| > \|z - u\|$.

Proofs of (i) have been given in [5], [6], [9]; (ii) and (iii) are proved in [2].

(iv) *Let* $y_i, 1 \leqq i \leqq 8$, *be such that* $\|y_i\| = 1$, $\|y_i - y_j\| \geqq 1$ *for* $i \neq j$. *Then* K *is a parallelogram.*

Proof. Since in Minkowski geometry a straight-line segment is a path of minimal length between two points, the above hypotheses imply that the perimeter of K is $\geqq 8$ (in the Minkowski metric). But it is well known (see, e.g., [6], [9]) that the unit cell of any Minkowski plane has a perimeter $\leqq 8$; moreover, the same proofs easily yield also the fact that the perimeter equals 8 only if K is a parallelogram, which ends the proof of (iv).

(v) *If there exists a set* $Y = \{y_i, 1 \leqq i \leqq 7\} \subset$ Front K *such that* $\|y_i - y_j\| \geqq 1$ *for* $i \neq j$, *then* K *is a parallelogram.*

Proof. Let $\pm x_i$, $i - 1, 2, 3$, be the vertices of any affine-regular hexagon H inscribed in K (such hexagons exist by (i)). We note that:

(a) If two points of Y are opposite vertices of H, then $Y \cup (-Y)$ contains 8 points satisfying the assumptions of (iv), and therefore K is a parallelogram;

(b) No pair of points $y_i, y_j \in Y$ can belong to the interior of a small arc of Front K determined by two neighboring vertices of H, since in such a case (iii) would imply that $\|y_i - y_j\| < 1$.

Now, if (a) does not hold, it is clear that we may find H such that, after suitably changing the indices if necessary, the following relations hold ($\prec$ denotes equality, or precedence according to a fixed orientation of Front K):

$$x_1 = y_1 \prec y_2 \prec x_2 \prec y_3 \prec x_3 \prec y_4 \prec -x_1,\ y_2 \neq x_2,\ y_4 \neq -x_1.$$

Then (ii) implies that $\|y_1 - y_2\| = 1$, and that y_2, x_2, y_3, x_3 belong to a maximal straight-line segment $[a, b] \subset$ Front K, with $x_1 \prec a \prec x_2$. Now, if $y_4 \in [a, b]$ we have $\|a - b\| \geqq 2$ which establishes K as a parallelogram. Let us therefore assume $y_4 \notin [a, b]$. Jointly with $y_4 \neq -x_1$ this implies that $y_2 = a$, $y_3 = a - x_1$, and $\|y_3 - y_2\| = 1$, since otherwise the affine-regular hexagon with vertices $\pm x_1$, $\pm a$, $\pm(a - x_1)$ would yield the situation described in (b). Now $\|y_3 - (-x_1)\| = \|a\| = 1$ and $\|y_3 - y_4\| \geqq 1$ imply, by (ii), that $\|y_3 - y_4\| = 1$ and that y_4, $-x_1$, and $-a$ are points of a segment $[-a, c] \subset$ Front K, which is obviously adjacent to the segment $[-a, -b]$.

Using (ii) repeatedly we see that $-x_1 \prec y_5$ and $y_5 \neq -x_1$; therefore $-b-x_1 \prec y_6$ and $y_6 \neq -b-x_1$, so $-b \prec y_7$ with $y_7 \neq -b$. But this is impossible since it would imply $\|y_1 - y_7\| < \|x_1 + b\| = 1$. Accordingly, y_4 must belong to $[a, b]$, and (v) is proved.

(vi) *If $P = -P$ is a parallelogram, if C is a convex set, and if $P = (1/2)[C + (-C)]$, then $C = P + x$ for a suitable point x.*

Proof. Considering the supporting lines of P it is immediate that C must be a parallelogram with sides parallel to those of P; therefore $P = (1/2)[C + (-C)]$ implies that C is a translate of P.

REMARK. The author is indebted to Professor E. G. Straus for the remark that (vi) has to be used in order to complete the original proof of the theorem. Professor Straus also observed that if K is a centrally symmetric plane convex body different from a parallelogram, then $K=(1/2)[C + (-C)]$ for some C which is not a translate of K. The following particularly simple proof of this fact was given by Dr. E. Asplund:

Inscribe an affine regular hexagon H in P (see (i)) and construct a curve $(1/2)C$ consisting of translates of the arcs of the boundary of P which are determined by alternate sides of H. It is easy to see that $(1/2)C$ is not homothetic to the boundary of P unless P is a parallelogram. On the other hand $(1/2)C$ has constant width 1 in the Minkowski metric whose unit sphere is P (it is in fact a Reuleux triangle for that metric) and thus $-(1/2)C + (1/2)C$ is the sphere P as the only centrally symmetric body of constant width.

The related question of non-trivial decomposability of centrally symmetric convex bodies in three-dimensional space seems to be much more complicated. Using results of Gale [1] it is easily established that parallelepipeds, octahedra and other centrally-symmetric anti-prisms, as well as other sets, are only trivially decomposable in the form $(1/2)[C + (-C)]$.

3. We now turn to the proof of our theorem. First of all we remark that without loss of generality we may assume K to be centrally symmetric. Indeed, if K is any convex set, $(1/2)[K + (-K)]$ is centrally symmetric; but, as has been noted by Minkowski [8] and used also by Hadwiger [3], $(x + K) \cap (y + K)$ and $(x + (1/2)[K + (-K)]) \cap (y + (1/2)[K + (-K)])$ are simultaneously empty, non-empty, or have interior points. Therefore, (vi) implies that the general case follows from the symmetric one.

Assuming now that K is centrally symmetric and that the translates $K_n = z_n + K$ satisfy conditions (1), we construct a new family of translates $\{K_n^*\}$ as follows: If $z_n = 0$ we put $K_n^* = K_n$; if $z_n \neq 0$, we define $K_n^* = (2z_n/\|z_n\|) + K$. The family K_n^* then satisfies the conditions

(1). Indeed, $K_n^* \cap K$ obviously contains $y_n = z_n/\| z_n \|$ (resp. $y_n = 0$ if $z_n = 0$), and for $n \neq m$ we have

$$\text{Int}\, K_n^* \cap \text{Int}\, K_m^* = \phi \tag{2}$$

since (1), assumed to hold for the family $\{K_n\}$, implies

$$\text{Int}\,(\lambda z_n + K) \cap \text{Int}(\mu z_m + K) = \phi \text{ for any } \lambda, \mu \geqq 1.$$

Now, (2) implies that $\| 2y_n - 2y_m \| \geqq 2$, i.e., $\| y_n - y_m \| \geqq 1$, and therefore the theorem follows from (v).

4. The number $i(K)$ may be defined in the same way for convex bodies in any Euclidean space. Hadwiger proved that $i(K) \leqq 3^k$ for $K \subset E^k$, the bound being attained for k-dimensional parallelotopes. On the other hand we have:

If $K \subset E^k$ then $i(K) \geqq k^2 + k + 1$.

Proof. As above, we may without loss of generality assume that K is centrally symmetric with center 0. Let the points x_i, $0 \leqq i \leqq k$, satisfy $\| x_i - x_j \| = 2$ for $i \neq j$, where the norm is taken in the Minkowski metric determined by K. (The existence of such a family $\{x_i\}$ may be established by obvious continuity arguments.) Then the $k^2 + k + 1$ sets $x_i - x_j + K$, for $0 \leqq i$, $j \leqq k$, satisfy conditions (1). Thus our assertion is established.

The above estimate $i(K) \geqq k^2 + k + 1$ is the best possible; it is attained if K is, e.g., a simplex. This is obvious for $k \leqq 3$, and may be established also in the general case.

As a generalization of the result of §1, we conjecture that $i(K)$ is odd for any K and that any odd value between $k^2 + k + 1$ and 3^k is assumed. The last part of the conjecture is easily verified for $k = 3$.

5. We end with a related result. Following [4], we shall say that a set A encloses a set B if every unbounded connected set which intersects B also intersects A. For any convex body K in the Euclidean plane let $e(K)$ denote the smallest natural number with the property:

There exist translates K_n, $1 \leqq n \leqq e(K)$, such that

$$\text{Int}\, K \cap \text{Int}\, K_n = \phi \text{ for all } n$$

$$\bigcup_{n=1}^{e(K)} K_n \text{ encloses } K.$$

With this terminology we have

If K is not a parallelogram, then $e(K) = 6$. For a parallelogram P, $e(P) = 4$.

This result may be established by the same methods we used in

§§ 2 and 3. Using the conventions of § 2, the main step of the proof (which is used instead of (iv) and (v)) may be formulated as follows:

(vii) *If* $Y = \{y_i; 1 \leqq i \leqq 5\} \subset \text{Front } K$ *with* $\| y_i - y_{i+1} \| \leqq 1$ *and* $y_i \prec y_{i+1}$ *for all* i $(y_6 = y_1)$, *and if the origin belongs to the convex hull of* Y, *then* K *is a parallelogram.*

We may also mention another theorem of a similar kind, established by Levi [7]: If K is a convex body in the plane, different from a parallelogram, then there exist three translates of Int K such that their union covers K (and therefore encloses it). For centrally symmetric sets a stronger theorem of the same type is given in [2].

References

1. D. Gale, *Irreducible convex sets*, Proc. Congr. Amsterdam, 2, (1954), 217-218.
2. B. Grünbaum, *Borsuk's partition conjecture in Minkowski planes*, Bull. Research Council of Israel, **7**F (1957), 25-30.
3. H. Hadwiger, *Ueber Treffanzahlen bei translationsgeichen Eikörpern*, Archiv der Mathematik, **8** (1957), 212-213.
4. C. J. A. Halberg, Jr., E. Levin, and E. G. Straus, *On contiguous congruent sets in Euclidean space*, Proc. Amer. Math. Soc., **10** (1959), 335-344.
5. P. J. Kelly, *A property of Minkowskian circles*, Amer. Math. Monthly, **57** (1950), 677-678.
6. D. Laugwitz, *Konvexe Mittelpunktsbereiche und normierte Räume*, Math. Zeitschr., **61** (1954), 235-244.
7. F. W. Levi, *Ueberdeckung eines Eibereiches durch Parallelverschiebungen seines offenen Kerns*, Archiv der Math., **6** (1955), 369-370.
8. H. Minkowski, *Dichteste gitterformige Lagerung kongruenter Körper*, Nachr. der K. Ges. der Wiss. zu Göttingen, Math.-Phys. Kl., 1904, 311-355 (=Gesam. Abh., Vol. 2, Leipzig und Berlin, 1911, pp. 3-42).
9. Yu. G. Rešetnyak, *An extremal problem from the theory of convex curves*, Uspehi Matem. Nauk (N. S.) **8** No. 6 (58) (1953), 125-126. (Russian)

THE INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY

ON THE ACTION OF A LOCALLY COMPACT GROUP ON E_n

F. J. HAHN

It is known [2, p. 208] that if a locally compact group acts effectively and differentiably on E_n then it is a Lie group. The object of this note is to show that if the differentiability requirements are replaced by some weaker restrictions, given later on, the theorem is still true. Let G be a locally compact group acting on E_n and let the coordinate functions of the action be given by $f_i(g, x_1, \cdots, x_n)$, $1 \leqq i \leqq n$. For economy we introduce the following notation

$$Q_{ij}(g, t, x) = \frac{f_i(g, x_1, \cdots, x_j + t, \cdots, x_n) - f_i(g, x_1, \cdots, x_j, \cdots, x_n)}{t}.$$

We denote by $\sigma(Q_{ij}(e, 0, x))$ the oscillation of $Q_{ij}(g, t, x)$ at the point $(e, 0, x)$.

Before proceeding there is one simple remark to be made on matrices. If $A = (a_{ij})$ is an $n \times n$ matrix such that $| a_{ij} - \delta_{ij} | < (1/n)$ then A is non-singular. If A were singular there would be a vector x such that $\sum_i x_i^2 = 1$ and $Ax = 0$. From the Schwarz inequality it follows that $x_i^2 = (\sum_j (a_{ij} - \delta_{ij})x_j)^2 < (1/n)$ and consequently $1 = \sum x_i^2 < 1$ which is impossible. If $| a_{ij} - \delta_{ij} | \leqq (\alpha/n)$, where $0 < \alpha < 1$, then the determinant of A is bounded away from zero since the determinant is a continuous function and the set $\{a_{ij}: | a_{ij} - \delta_{ij} | \leqq (\alpha/n)\}$ is compact in E_{n^2}.

THEOREM 1. *If T is a pointwise periodic homeomorphism of E_n then T is periodic.*

Proof. [2, p. 224.]

THEOREM 2. *If G is a compact, zero dimensional, monothetic group acting effectively on E_n and satisfying*

$$(*) \qquad \sigma(Q_{ij}(e, 0, x)) < \frac{\varepsilon}{n}, \quad 0 < \varepsilon < 1, \quad \textit{for each } x \textit{ in } E_n;$$

then G is a finite cyclic group.

Proof. Since G is monothetic, let a be an element whose powers are dense in G. It is enough to show that there is a power of a which leaves E_n pointwise fixed since the action of G is effective.

Received April 12, 1960. The author is a National Science Foundation Fellow.

If q is a positive integer we let

$$T_i^q(g, x) = x_i + f_i(g, x) + \cdots + f_i(g^{q-1}, x) .$$

If $y = (y_i)$ and $x = (x_i)$ let

$$T_{ij}^q(g, x, y) = \frac{T_i^q(g, x_1, \cdots, x_{j-1}, y_j, \cdots, y_n) - T_i^q(g, x_1, \cdots, x_j, y_{j+1}, \cdots, y_n)}{y_j - x_j}$$

for $y_j \neq x_j$ and zero otherwise. If we let $y = f(g, x)$ then we obtain

$$\begin{aligned} f_i(g^q, x) - x_i &= T_i^q(g, y) - T_i^q(g, x) \\ &= \sum_{j=1}^{n} T_{ij}^q(g, x, y)(y_j - x_j) \\ &= q \cdot \sum_{i=1}^{n} \frac{1}{q} T_{ij}^q(g, x, y)(y_j - x_j) . \end{aligned}$$

Because of the fact that $f_i(e, x) = x_i$ and because of $(*)$ it follows that there is a compact neighborhood $U(x)$ of the identity of G such that if $g, \cdots, g^q \in U(x)$ then $|(1/q)T_{ij}^q(g, x, y) - \delta_{ij}| \leqq (\alpha/n)$, $0 < \varepsilon < \alpha < 1$. It follows that if T is the matrix with entries $(1/q)T_{ij}^q(g, x, y)$ then T is non-singular and its determinant is bounded away from zero uniformly in q, so the determinant of the inverse is bounded uniformly in q; thus

$$(f(g, x) - x) = (y - x) = \left(\delta_{ij}\frac{1}{q}\right) \cdot T^{-1} \cdot (f(g^q, x) - x) .$$

Since G is monothetic and zero dimensional there is a power of a such that if $g = a^p$ then all the powers of g lie in $U(x)$. Since $U(x)$ is compact it follows that the vectors $f(g^q, x) - x$ are bounded uniformly in q and thus $f(g, x) - x = f(a^p, x) - x = 0$. Hence a is pointwise periodic on E_n and it follows from Theorem 1 that it is periodic and consequently has a power leaving E_n pointwise fixed.

From this it follows quickly that if G is a locally compact group acting effectively on E_n and satisfying $(*)$ then it is a Lie group. This follows from the fact that since G is effective it must be finite dimensional [1] and then if G is not a Lie group it must contain a compact, non-finite zero dimensional subgroup H [2, p. 237] which acts effectively. H has small subgroups which act effectively and it follows from Newman's theorem [3, 4] that H cannot have arbitrarily small elements of finite order. Thus H has an element a of infinite order such that the compact subgroup generated by a acts effectively on E_n and satisfies $(*)$ but by Theorem 2 this is impossible.

BIBLIOGRAPHY

1. D. Montgomery, *Finite dimensionality of certain transformation groups*, Ill. J. Math. 1, No. **1** (1957), 28–35.

2. D. Montgomery and L. Zippin, *Topological Transformation Groups,* Interscience Publishers, New York (1955).
3. M. H. A. Newman, *A theorem on periodic transformations of spaces,* Quart. J. Math. **2** (1931), 1-8.
4. P. A. Smith, *Transformations of finite period, III, Newman's theorem,* Ann. of Math. **42** (1941), 446-458.

INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY

RELATIVE HERMITIAN MATRICES

MAGNUS R. HESTENES

1. Introduction. The purpose of the present paper is to develop a spectral theory for an arbitrary $m \times n$ dimensional matrix A, which is analogous to that given in the hermitian case and which reduces to the usual spectral theory when A is hermitian. The theory is centered around the triple product AB^*C of matrices of the same dimension. Here B^* is the transpose of B in the field of real numbers and the conjugate transpose of B in the field of complex numbers. The matrix T will be said to be elementary in case $T = TT^*T$. Elementary matrices play the role of units and in case of vectors are unit vectors. Given an elementary matrix T and a matrix A of the same dimension the matrix TA^*T can be considered to be the conjugate transpose of A relative to T. If $A = TA^*T$, then A is hermitian relative to T. The polar decomposition theorem for matrices implies that to each matrix A there is a unique elementary matrix R such that A is hermitian relative to R, AR^* is non-negative hermitian in the usual sense, and R has the same null space as A. Every elementary matrix T relative to which A is hermitian is of the form $T = T_0 + R_1 - R_2$, where $R_1 + R_2 = R$ and T_0, R_1, R_2 are mutually $*$-orthogonal. Two matrices A and B are $*$-orthogonal in case $AB^* = 0$ and $A^*B = 0$. A matrix B will be called a section of A if B and $A - B$ are $*$-orthogonal.

If A is hermitian relative to an elementary matrix T, it is shown below that A and T can be written as sums of sections

$$A = A_1 + \cdots + A_q , \qquad T = T_1 + \cdots + T_q$$

such that $A_i = \lambda_i T_i$, where λ_i is a real number. Moreover these sections can be chosen so that $\lambda_i \neq \lambda_j$, $(i \neq j)$. If in this event the decomposition is unique. If $AT^* \geqq 0$, then $\lambda_i \geqq 0$. If in addition A and T have the same null space then $\lambda_i > 0$. In the event T is the identity, this result gives the usual spectral representation of hermitian matrices.

A matrix A will be said to be normal relative to an elementary matrix T in case $A = AT^*T = TT^*A$, $AA^*T = TA^*A$. In this event the spectral decomposition theorem described above holds, the coefficients λ_i being complex instead of real.

In the development of the theory the concept of $*$-commutativity of two matrices plays a significant role. The matrices A and B will be said to $*$-commute (see § 4 below) in case $AB^* = BA^*$ and $A^*B = B^*A$. If A and B $*$-commute, there is an elementary matrix T relative to

Received April 26, 1960. The preparation of this paper was sponsored by the Office of Naval Research and the Office of Ordnance Research, U. S. Army. Reproduction in whole or in part is permitted for any purpose of the United States Government.

which they are hermitian. Moreover $A, B\ T$ can be written as sums of section A_i, B_i, T^i such that $A_i = \lambda_i T_i$ and $B_i = \mu_i T_i$, where λ_i and μ_i are real. If $\mathfrak{A}$ is a linear class of $m \times n$ dimensional matrices that are hermitian relative to an elementary matrix T in $\mathfrak{A}$ (including T itself) and have the property that the product AB^*C is invariant under permutations A, B and C, then this class forms an algebra with AT^*B as the product of A and B. The elements A of $\mathfrak{A}$ are all matrices expressible in the form $A = \lambda_i T_i + \cdots + \lambda_q T_q$, where $T_1, \cdots, T_q$ are suitably chosen sections of T. Throughout the paper the reciprocal A^{*-1} of A^* plays a role analogous to that of A itself.

The main results obtained in the present paper can be extended to a closed operator A from a Hilbert space $\mathfrak{R}$ to a second Hilbert space $\mathfrak{R}'$. Whenever it is convenient to do so, the theorems are stated so as to be valid for operators in Hilbert space. The terminology used has been chosen so as to make this transition as simple as possible. The extension to Hilbert spaces yields a common spectral theory for the gradient of a function and the divergence of a vector field.

2. Terminology and notations. Throughout the following pages matrices will be denoted by capital letters $A, B, C, P, Q, R, \cdots$. The elements can be considered to be real or complex. The conjugate transpose of A will be denoted by A^*. It will be convenient to consider all matrices to be square, since this can be obtained by the addition of zero-elements. However, this is not essential. The paper is written so as to be valid for rectangular matrices, the equality of the dimensions of two or more matrices being implied by the condition that the operations used should be well defined.

Occasionally we shall use column vectors and row vectors. A column vector will be denoted by $x, y, z, \cdots$. Row vectors are conjugate transposes of column vectors. If $x^*x = 1$ then x is a unit vector. Given two vectors x and y then $A = yx^*$ is a matrix of rank 1. Every matrix of rank 1 is represented in the form $A = \lambda yx^*$, where λ is a real number and x, y, are unit vectors. In fact λ can be taken to be positive. The greek letters, $\alpha, \beta, \gamma, \mu, \cdots$ appearing in the text normally denote real numbers.

A matrix is *hermitian* if $A^* = A$. A hermitian matrix A is *non-negative*, written $A \geqq 0$, if $x^*Ax \geqq 0$ for every column vector x. If $A \geqq 0$, there is a unique matrix $B \geqq 0$ such that $B^2 = A$. The matrix B will be called the *square root* of A. A matrix E will be called a *pro-*

[1] E. H. Moore, *General Analysis*, Part I, Mem. Philos. Soc. (1935), p. 197. See also Penrose, "A generalized inverse of matrices", *Proc.* Cambridge Philos. Soc. **51** (1953), 406–413; M. R. Hestenes, "Inversion of matrices by biorthogonalization and related results," *J. Soc. Ind. Appl. Math.* **6** (1958), p. 84; J. von Neumann, On regular rings, Proc. Nat. Acad. Sci. U. S. A. **22** (1936), 707–713.

jection if $E = E^* = E^2$. The identity matrix will be denoted by I. The null space of the matrix A will be denoted by $\mathfrak{N}_A$.

To each matrix A there is a unique matrix B such that

$$A = AA^*B = AB^*A = BA^*A\ , \qquad B = BB^*A = BA^*B = AB^*B\ .$$

The matrix B is the reciprocal of A^* in the sense of E. H. Moore[1] and will be called the $*$-reciprocal of A. It is also the conjugate transpose of the *reciprocal* A^{-1} of A. If A is nonsingular, A^{-1} is the inverse of A. We shall accordingly use the symbols A^{*-1}, A^{-1*} for the $*$-reciprocal of A. The matrices

$$E = A^{-1}A = A^*A^{*-1}\ , \qquad E' = AA^{-1} = A^{*-1}A^* \tag{2.1}$$

are projections and satisfy the relations

$$E'A = AE = A\ , \qquad E'A^{*-1} = A^{*-1}E = A^{*-1}\ . \tag{2.2}$$

They will be called the *projections associated with* A. It should be noted that the reciprocal of A^*A is $A^{-1}A^{*-1}$ and that the reciprocal of AA^* is $A^{*-1}A^{-1}$. If A is hermitian then $A^{*-1} = A^{-1}$. If A is nonsingular then A^{-1} is the inverse of A.

A matrix R will be said to be an *elementary matrix* in case $RR^*R = R$. It is easily seen that R is elementary if and only if $R = R^{*-1}$ or equivalently if and only if $R^* = R^{-1}$. If R is elementary so also is R^*. A projection is an elementary matrix. If R is a hermitian elemetary matrix, then

$$E_+ = \frac{1}{2}(R^2 + R)\ , \qquad E_- = \frac{1}{2}(R^2 - R)$$

are projections such that

$$R = E_+ - E_-\ , \quad E_+E_- = E_-E_+ = 0\ , \quad R^2 = E_+ + E_-\ .$$

Conversely, the difference of two projections that are orthogonal is a hermitian elementary matrix.

A matrix A will be said to be *hermitan relative to an elementary matrix* R if $A = RA^*R$. The following result is fundamental.

THEOREM 2.1. *Suppose that A is hermitian relative to an elementary matrix R. Then A^{*-1} is hermitian relative to R and A^*, A^{-1} are hermitian relative to R^*. Moreover, $\mathfrak{N}_R \subset \mathfrak{N}_A$ and $\mathfrak{N}_{R^*} \subset \mathfrak{N}_{A^*}$. The matrices A and R satisfy the further relations*

$$A = RR^*A = AR^*R = RA^*R\ , \quad RA^*A = AA^*R = AR^*A \tag{2.3a}$$

$$A^*R = R^*A\ , \quad AR^* = RA^*\ . \tag{2.3b}$$

$$(A^*R)^2 = A^*A\ , \quad (AR^*)^2 = AA^*\ . \tag{2.3c}$$

(2.3d) $A^{*-1}A^*R = RA^*A^{*-1} = AR^*A^{*-1} = AA^{-1}R = RA^{-1}A = A^{*-1}R^*A$.

Since $A = RA^*R$ we have $A^* = R^*AR^*$ and $\mathfrak{N}_R \subset \mathfrak{N}_A$, $\mathfrak{N}_{R^*} \subset \mathfrak{N}_{A^*}$. Moreover

$$RR^*A = RR^*RA^*R = RA^*R = A = AR^*R$$
$$AR^*A = RA^*RR^*RA^*R = RA^*RA^*R = AA^*R = RA^*A .$$

It follows that (2.3a) holds. The relations (2.3b) and (2.3c) follow from the computations

$$A^*R = R^*AR^*R = R^*A , \quad RA^* = RR^*AR^* = AR^*$$
$$(A^*R)^2 = A^*RR^*A = A^*A , \quad (AR^*)^2 = AR^*RA^* = AA^* .$$

It is easily verified that $RA^{-1}R$ is the reciprocal $R^*AR^* = A^*$. Consequently $A^{*-1} = RA^{-1}R$, that is, A^{*-1} is hermitian relative to R. Similarly A^{-1} is hermitian relative to R^*. The relations (2.3d) follow from (2.3b) and the relations $AA^{-1} = A^{*-1}A^*$, $A^{-1}A = A^*A^{*-1}$.

COROLLARY. *Suppose $A = RA^*R$, $R = RR^*R$ and set $P = A^*R$, $Q = RA^*$. Then*

(2.4) $$A = RP = QR , \quad P = R^*QR , \quad Q = RPR^* .$$

The matrix P is nonnegative if and only if Q is nonnegative. Moreover $\mathfrak{N}_R = \mathfrak{N}_A$ if and only if

(2.5) $$R = A^{*-1}R^*A$$

and hence if and only if $\mathfrak{N}_{R^*} = \mathfrak{N}_{A^*}$.

In view of this result we define a matrix A to be *nonnegative hermitian relative to R* in case $A = RA^*R$ and $A^*R \geqq 0$.

THEOREM 2.2. *Given a matrix A there is a unique elementary matrix R such that A is nonnegatively hermitian relative to R and such that $R = A^{*-1}R^*A$. Moreover*

(2.6) $$R^*R = A^{-1}A = A^*A^{*-1} , \quad RR^* = AA^{-1} = A^{*-1}A^* .$$

Let P be the square root of A^*A. The matrix $R = A^{*-1}P$ has the properties described in the theorem. Clearly $A^*R = A^*A^{*-1}P = P \geqq 0$. Moreover

$$RR^*R = A^{*-1}P^2A^{-1}R = A^{*-1}(A^*AA^{-1}A^{*-1})P = A^{*-1}P = R ,$$
$$RA^*R = RP = A^{*-1}P^2 = A^{*-1}A^*A = A ,$$
$$A^{*-1}R^*A = A^{*-1}PA^{-1}A = A^{*-1}P = R ,$$
$$R^*R = R^*A^{*-1}R^*A = A^{-1}A = A^*A^{*-1} ,$$
$$RR^* = A^{*-1}R^*AR^* = A^{*-1}A^* = AA^{-1} .$$

The uniqueness of R follows from the uniqueness of P as the square root of A^*A. This proves the theorem.

If R is chosen as described in Theorem 2.2 then the formula (2.4) for A in terms of $P = A^*R$ and R is called the polar decomposition[2] of A.

The matrix R described in Theorem 2.2 will be called the *elementary matrix associated with* A.

COROLLARY. *Let R be the elementary matrix associated with A and let S and T be elementary matrices such that*

$$A = SS^*A = AT^*T$$

Then $V = TR^*S$ is the elementary operator associated with $B = TA^*S$

This follows because

$$\begin{aligned} VB^*V = TR^*SS^*AT^*TR^*S = TR^*AR^*S = TA^*S = B \\ B^*V = S^*AT^*TR^*S = S^*(AR^*)S \geqq 0 \ . \\ B^{*-1}V^*B = B^{*-1}S^*RT^*TA^*S = B^{*-1}S^*RA^*S \\ = TA^{-1}SS^*AR^*S = TA^{-1}AR^*S = TR^*S = V \ . \end{aligned}$$

As a further result we have

THEOREM 2.3. *Let R be the elementary matrix associated with A. If A is normal so also is R. If A is nonnegative hermitian, then R is a projection. If A is hermitian, then R is hermitian and is the difference of two orthogonal projections.*

If A is normal, its associated projections E, E' coincide. By virtue of (2.6) we have $RR^* = R^*R$ and R is normal. If A is nonnegative hermitian, then $R = E$. If A is hermitian, let P be the square root of A^2. Then $A = RP = PR = PR^* = R^*P$. Consequently $R = R^*$, as was to be proved.

The following result is of interest.

THEOREM 2.4. *Let R be the elementary matrix associated with A. There exists a unique pair of matrices B, C such that*

$$B + C = R \ , \qquad A = BR^*C^{*-1} = C^{*-1}R^*B$$

and having R as their associated elementary matrix. The matrices B and C are defined by the formulas

$$B^{-1} = A^{-1} + R^* \ , \qquad C^{-1} = A^* + R^*$$

and satisfy the relations

$$BB^*C = BC^*B = CB^*B \ , \qquad CC^*B = CB^*C = BC^*C$$

[2] See MacDuffee, C. B., "Theory of Matrices", Ergebnisse der Mathematik und ihrer Grenzgebiete (1933), pp. 77.

$$B^{-1}RC^{-1} = C^{-1}RB^{-1} = B^{-1} + C^{-1}$$
$$A^{*-1} = B^{*-1}R^*C = CR^*B^{*-1} , \qquad A^* = B^*RC^{-1} = C^{-1}RB^* ,$$
$$A^{-1} = B^{-1}RC^* = C^*RB^{-1} .$$

Since no direct use of this result will be made, its proof will be omitted.

3. $*$-orthogonality. Two matricies A and B will be said to be *$*$-orthogonal* in case

$$A^*B = B^*A = 0 , \qquad AB^* = BA^* = 0 . \tag{3.1}$$

Consider now two matrices A and B and let

$$E = A^{-1}A , \quad E' = AA^{-1} , \quad F = B^{-1}B , \quad F' = BB^{-1} \tag{3.2}$$

be the associated projections. We have the following

LEMMA 3.1. *Two matrices A and B are $*$-orthogonal if and only if*

$$EF = FE = 0 \qquad E'F' = F'E' = 0 . \tag{3.3}$$

Moreover two matrices A and B are $*$-orthogonal if and only if their associated elementary matrices R and S are $*$-orthogonal. Finally A is $*$-orthogonal to B if and only if A^{-1*} is $*$-orthogonal to B.

If (3.1) holds, then

$$EF = (A^{-1}A)(B^*B^{-1*}) = A^{-1}(AB^*)B^{-1*} = 0$$
$$E'F' = (A^{-1*}A^*)(BB^{-1}) = 0 .$$

Hence (3.3) holds. The converse follows from the relations $A = AE = E'A$, $B = BF = F'B$. The last two statements in the lemma follow from the first.

LEMMA 3.2. *Let A and B be $*$-orthogonal matrices and set $C = A + B$. Then*

$$C^* = A^* + B^* , \quad C^{-1} = A^{-1} + B^{-1} , \quad C^{*-1} = A^{*-1} + B^{*-1} . \tag{3.4}$$

The rank of C is the sum of the ranks of A and B. The elementary matrix associated with C is the sum $T = R + S$ of the elementary matrices R, S associated with A, B, respectively. The matrix C is elementary if and only if A and B are elementary.

By the use of Lemma 3.1 it is seen that

$$A^{-1}B = B^{-1}A = 0 , \quad AB^{-1} = BA^{-1} = 0 .$$

It follows that

$$(A^{-1} + B^{-1})C = A^{-1}A + B^{-1}B = E + F$$

$$C(A^{-1} + B^{-1}) = AA^{-1} + BB^{-1} = E' + F' .$$

In view of (3.3) the matrices $G = E + F$ and $G' = E' + F'$ are projections. Moreover, setting $C^{-1} = A^{-1} + B^{-1}$ we have

$$G'C = CG = C , \quad C^{-1}G' = GC^{-1} = C^{-1} .$$

The matrix C^{-1} is therefore reciprocal of C and the relations (3.4) hold. To show that $T = R + S$ is the elementary matrix associated with C observe that $R^*B = S^*A = 0, BR^* = AS^* = 0$, by Theorem 3.1. Hence

$$T^*C = R^*A + S^*B = A^*R + B^*S = C^*T \geqq 0 , \quad CT^* = TC^* \geqq 0 .$$
$$T^*T = R^*R + S^*S = E + F = G = C^{-1}C , \qquad TT^* = G' = CC^{-1} ,$$

as was to be proved. The remaining statements in the lemma are easily established.

A matrix A will be said to be a *section* of a matrix C if there is a matrix B $*$-orthogonal to A such that $C = A + B$. By virtue of the last lemma the elementary matrix R associated with a section A of C is a section of the elementary matrix T belonging to C. A section of an elementary matrix is elementary.

LEMMA 3.3. *Let E, E' be the projections associated with a matrix A and let F and F' be projections such that $F'A = AF$. Then $EF = FE$, $E'F' = F'E'$. Moreover $A_1 = AF$ is a section of A.*

Since $AE = A$ it follows that $AFE = F'AE = F'A = AF$. Consequently

$$EFE = A^{-1}AFE = A^{-1}AF = EF .$$

This implies that $EF = FE$. Similarly $E'F' = F'E'$. Observe that

$$FA^{-1} = FEA^{-1} = EFA^{-1} = A^{-1}AFA^{-1} = A^{-1}F'AA^{-1} = A^{-1}F'E' = A^{-1}F'$$
$$FA^{-1}AF = FEF = EF , \qquad F'AA^{-1}F' = F'E'F' = E'F' .$$

Consequently $A_1^{-1} = FA^{-1} = A^{-1}F'$ is the reciprocal of $A_1 = AF = F'A$. The projections $E_1 = EF, E_0 = E - E_1$ are orthogonal as are $E_1' = E'F'$, $E_0' = E' - E_1'$. Consequently $A_0 = AE_0 = A - A_1 = E_0'A$ is $*$-orthogonal to A_1. Since $A = A_0 + A_1$ it follows that A_1 is a section of A, as was to be proved.

LEMMA 3.4. *A matrix B is a section of A if and only if $A^*B = B^*B, BA^* = BB^*$.*

Let

$$F = B^{-1}B , \qquad F' = BB^{-1} .$$

If $A^*B = B^*B, BA^* = BB^*$

$$A^*F' = A^*BB^{-1} = B^*BB^{-1} = B^*$$

$$FA^* = B^{-1}BA^* = B^{-1}BB^* = B^* .$$

Consequently $B = AF = F'A$ and B is a section of A, by Lemma 3.3. The converse is immediate.

LEMMA 3.5. *Let R be an elementary matrix and set $E = R^*R$. Let $S = RF$, where F is a projection. Then S is a section of R if and only if $EF = FE$. If r is the rank of R then R is expressible as the sum $R = R_1 + \cdots + R_r$ of r $*$-orthogonal sections of rank* 1.

If S is a section of R then $R^*S = S^*S = EF = EFE$. Consequently $EF = FE$. Conversely if $EF = FE$ then

$$S^*S = FR^*RF = FEF = EF = R^*RF = R^*S$$
$$SS^* = RF^2R^* = RFR^* = SR^* .$$

It follows from Lemma 3.4 that S is a section of R.

In order to prove the last statement in the theorem suppose that $R \neq 0$ and choose a unit vector x such that $Ex = x$. Then $E_1 = xx^*$ is a projection that commutes with E. Hence $R_1 = RE_1$ is a section of R of rank 1. Moreover $R - R_1$ is a section of R of rank $r - 1$ and is $*$-orthogonal to R_1. If $R - R_1 \neq 0$ it has a section R_2 of rank 1. Clearly R_2 is $*$-orthogonal to R_1 and $R - R_1 - R_2$ is a section of R of rank $r - 2$. By a repetition of this argument it is seen that R is expressible as the sum of r $*$-orthogonal sections, as was to be proved.

4. $*$-commutativity. Given two matrices A and B the products A^*B and AB^* can be considered to be two types of $*$-products of A and B. If these $*$-products are unaltered upon interchanging A and B, that is, if

$$A^*B = B^*A , \qquad AB^* = BA^* , \tag{4.1}$$

then A and B will be said to *$*$-commute*. It should be noted that A and B $*$-commute if and only if A^*B and AB^* are hermitian in the usual sense. As a first result we have

LEMMA 4.1. *If A and B $*$-commute and*

$$E = A^{-1}A , \quad E' = AA^{-1} , \quad F = B^{-1}B , \quad F' = BB^{-1} , \tag{4.2}$$

then $F'A = AF$, $E'B = BE$ and $EF = FE$, $E'F' = F'E'$.

For if (4.1) holds then, since $F'B = BF = B$, we have

$$F'AF = B^{*-1}B^*AF = B^{*-1}A^*BF = B^{*-1}A^*B = B^{*-1}B^*A = F'A$$
$$F'AF = F'AB^*B^{*-1} = F'B^*AB^{*-1} = B^*AB^{*-1} = AB^*B^{*-1} = AF .$$

Consequently $AF = F'A$. Similarly $E'B = BE$. In view of Lemma 3.3 the relations $EF = FE$, $E'F' = F'E'$ hold.

THEOREM 4.1. *Two matrices A and B $*$-commute if and only if they are expressible in the form $A = A_0 + A_1$, $B = B_0 + B_1$, where A_0 is $*$-orthogonal to A_1 and B, B_0 is $*$-orthogonal to B_1 and A, A_1 and B_1 $*$-commute and have the same associated projections*

$$E_1 = A_1^{-1}A_1 = B_1^{-1}B_1 \,, \qquad E_1' = A_1A_1^{-1} = B_1B_1^{-1} \,. \tag{4.3}$$

If A and B $*$-commute, then, by Lemmas 4.1 and 3.3, the matrices $A_1 = AF$, $B_1 = BE$ are sections of A and B respectively and have $E_1 = EF$, $E_1' = E'F'$ as their associated projections. Moreover $A_0 = A - A_1$ has $E - EF$, $E' - E'F'$ as its projections and hence is $*$-orthogonal to B and A_1 and hence also to B_1 and $B_0 = B - B_1$. Similarly B_0 is $*$-orthogonal to B_1, A, A_0 and A_1. The converse is immediate and the lemma is proved.

COROLLARY. *Suppose that A and B $*$-commute. Then A and A^{*-1} $*$-commute with B and B^{*-1}. Moreover A^*, A^{-1} $*$-commute with B^* and B^{-1}.*

We shall see later that their associated matrices R, S $*$-commute with $A, B, R,$ and S.

THEOREM 4.2. *Let R be the elementary matrix associated with a matrix A and let S be an elementary matrix that $*$-commutes with A. Then S $*$-commutes with R. Moreover A, R, S are expressible uniquely as sums and differences*

$$A = A_0 + A_+ + A_- \,, \quad R = R_0 + R_+ + R_- \,, \quad S = S_0 + R_+ - R_- \tag{4.4}$$

of $*$-orthogonal matrices such that the matrices R_0, R_+, R_- are the elementary matrices associated with A_0, A_+, A_- respectively and such that S_0 is $*$-orthogonal to A_0 and R_0. Conversely if A, R, S can be decomposed in this manner the A and R $*$-commute with S.

By virtue of the last theorem A and S can be expressible uniquely as the sum of $*$-orthogonal sections $A = A_0 + A_1$, $S = S_0 + S_1$ such that A_1 and S_1 have the same associated projections, S_0 being $*$-orthogonal to A and A_0 being $*$-orthogonal to S. The elementary matrix R associated with A is expressible in the form $R = R_0 + R_1$, where R_0 and R_1 are the elementary matrices associated with A_0 and A_1 respectively. In view of these remarks we can restrict ourselves to the case in which $A_0 = 0$, $S_0 = 0$ and $R_0 = 0$. Then

$$E = A^{-1}A = R^*R = S^*S \,, \qquad E' = AA^{-1} = RR^* = SS^* \,.$$

Since A^*S is self-adjoint, its associated elementary matrix T is the difference $T = E_+ - E_-$ of two orthogonal projections E_+ and E_- whose sum is E. The matrix A^*ST is nonnegative and self-adjoint. It follows from Theorem 2.1 that $R = ST$. The matrices $R_+ = RE_+$, $R_- = RE_-$ are $*$-orthogonal elementary matrices such that

$$R = RE = R_+ + R_- , \qquad S = RT = R_+ - R_- .$$

Since AR^* and AS^* are hermitian it follows that the matrices

$$AR_+^* = \frac{1}{2}A(R^* + S^*) , \qquad AR_-^* = \frac{1}{2}A(R^* - S^*)$$

are hermitian and orthogonal. Moreover they are nonnegative because of the relations

$$0 \leqq AR^* = AR_+^* + AR_-^* .$$

The elementary matrices R_+ and R_- are accordingly the elementary matrices associated with $A_+ = AE_+$ and $A_- = AE_-$ respectively. It is clear that A_+ and A_- are $*$-orthogonal and that $A = A_+ + A_-$. The matrices A, R, S are therefore expressible in the form (4.4). The converse is immediate and the theorem is established.

COROLLARY 1. *Two elementary matrices R and S $*$-commute if and only if there exist mutually matrices R_0, R_+, R_-, S_0 such that $R = R_0 + R_+ + R_-$, $S = S_0 + R_+ - R_-$. Moreover this decomposition is unique.*

COROLLARY 2. *If the matrix S appearing in Theorem 4.2 is of rank 1 then the decomposition (4.4) of A takes the simpler form*

$$A = \mu S + A_0 \tag{4.5}$$

where μ is a real number and A_0 is $*$-orthogonal to S.

For in this case two of the matrices S_0, R_+, R_- are zero since S has rank 1. If $S = S_0$, then (4.5) holds with $\mu = 0$. If $S = R_+$, then $R_- = A_- = 0$ and A_+ is of rank 1. Since S^*A_+ is a nonegative hermitian matrix of rank 1 it follows that A_+ is of the form $A_+ = \mu S$, where $\mu > 0$. If $S = -R_-$, then A_- is of the form $A_- = \mu S$ with $\mu < 0$.

COROLLARY 3. *If $S_1, \cdots, S_r$ are r mutually $*$-orthogonal elementary matrices of rank 1 that $*$-commute with A, then A is expressible in the form*

$$A = \mu_1 S_1 + \cdots + \mu_r S_r + A_0 \tag{4.6}$$

where $\mu_1, \cdots, \mu_r$ are real numbers and A_0 is $*$-orthogonal to each S_i $(i = 1, \cdots, r)$ and hence to $S_1 + \cdots + S_r$.

This result follows from Corollary 2 by induction. At the kth step one applies Corollary 2 with A replaced by $A - \mu_1 S_1 - \cdots - \mu_{k-1} S_{k-1}$ and with $S = S_k$.

THEOREM 4.3. *If a matrix A $*$-commutes with every section of an elementary matrix S than A is expressible in the form*

$$A = \mu S + A_0$$

where μ is a real number and A_0 is $*$-orthogonal to S.

If S has rank 1, the theorem holds by virtue of Corollary 2 to Theorem 4.2. If S is of rank $r > 1$, then, by Lemma 3.5, S is expressible in the form $S = S_1 + \cdots + S_r$ where $S_1, \cdots, S_r$ are mutually $*$-orthogonal elementary matrices. Consequently A is expressible in the form (4.6). It remains to show that $\mu_1 = \mu_2 = \cdots = \mu_r$. To this end choose unit vectors x_i and y_i such that $S_i = x_i y_i^*$. Then for $i \neq j$ the vector x_i is orthogonal to x_j and y_i is orthogonal to y_j. Let α and β be two nonnull real numbers such that $\alpha^2 + \beta^2 = 1$ and set

$$x = \alpha x_i + \beta x_j \,, \qquad y = \alpha y_i + \beta y_j \,.$$

Then $T = xy^*$ is a section of S and is $*$-orthogonal to S_k if $k \neq i, k \neq j$. The matrix

$$A^* T = (\mu_i S_i^* + \mu_j S_j^*) T = (\mu_i \alpha y_i + \mu_j \beta y_j) y^*$$

is hermitian if and only if $\mu_i = \mu_j$, that is T $*$-commutes with A if and only if $\mu_i = \mu_j$. This completes the proof of the theorem.

THEOREM 4.4. *A matrix A $*$-commutes with an elementary matrix S and has no nonnull section $*$-orthogonal to S if and only if $A = SA^*S$.*

This result is easily established. The condition that $A = SA^*S$, when $S = I$ is the condition that A be hermitian. Accordingly one can consider the condition $A = SA^*S$ to be an extension of the concept of a matrix being hermitian.

In the complex domain we have the following:

COROLLARY. *If A is a matrix and S is an elementary matrix such that $SS^*A = AS^*S = A$, then $B = (1/2)(A + SA^*S)$ and $C = (1/2i)(A - SA^*S)$ $*$-commute with S. Moreover, $A = B + iC$.*

5. Principal values and principal sections of matrices. Here and elsewhere the symbol $\|y\|$ denotes the length or norm of the vector y. By the norm $\|A\|$ of a matrix A will be meant the least upper bound of the quantity $\|Ax\|$ for all unit vectors x. If R is the elementary matrix associated with A, then $\|A\|$ is equal to the least upper bound of $\|Ax\|$ subject to the condition $\|Rx\| = 1$. As is well known there is a unit vector x such that $\|Ax\| = \|A\|$. For such a vector x we have $\|Rx\| = 1$ also. It is well known that $\|A\| = \|A^*\|$. If $A \neq 0$ then $\|A^{-1}\| = \|A^{*-1}\|$ is equal to the least number m such that $\|Ax\| \geqq (1/m)\|Rx\|$.

THEOREM 5.1. *Let R be the elementary matrix associated with A. Given a positive number λ there exists a unique decomposition*

$$A = A_+ + A_0 + A_- \,, \qquad R = R_+ + R_0 + R_- \tag{5.1}$$

on A and R into mutually $*$-orthogonal sections such that R_+, R_0, R_-, $R_+ - R_-$ are the elementary matrices associated with $A_+, A_0, A_-, A - \lambda R$ respectively. Moreover $A_0 = \lambda R_0$ and

$$(5.2) \qquad \begin{aligned} \| A_+x \| &> \lambda \| R_+x \| \qquad \textit{whenever } R_+x \neq 0 \\ \| A_-x \| &< \lambda \| R_-x \| \qquad \textit{whenever } R_-x \neq 0 \,. \end{aligned}$$

If $\lambda \geqq \| A \|$, then $A_+ = R_+ = 0$. If $1/\lambda \leqq \| A^{*-1} \|$ then $A_- = R_- = 0$.

In order to prove this result let $B = A - \lambda R$ and let S be the associated elementary matrix. Since R $*$-commutes with A and R, it $*$-commutes with B and hence also with S. Since S $*$-commutes with B and R it follows that S $*$-commutes with A. Applying Theorem 4.2 to A, R, S and to B, R, S it is seen that they are expressible as sums

$$\begin{aligned} A &= A_+ + A_0 + A_- \,, \qquad & R &= R_+ + R_0 + R_- \\ B &= B_+ + B_0 - B_- \,, \qquad & S &= R_+ + S_0 - R_- \,. \end{aligned}$$

of $*$-orthogonal matrices such that R_+ is the elementary matrix associated with A_+ and B_+, R_- is the elementary matrix associated with A_- and B_-, R_0 is the elementary matrix associated with A_0. It is clear that $B_0 = S_0 = 0$ since every matrix that is $*$-orthogonal to A and R is also $*$-orthogonal to B and S. From the relation $A = B + \lambda R$ it follows that

$$A_0 = \lambda R_0 \,, \quad A_+ = B_+ + \lambda R_+ \,, \quad A_- = B_- + \lambda R_- \,.$$

Consequently

$$R_+^* A_+ = R_+^* B_+ + \lambda E_+ \,, \qquad R_-^* A_- + R_-^* B_- = \lambda E_- \,.$$

Since these matrices are nonnegative and hermitian, it is seen that (5.2) holds. The last statement in the theorem follows from the relations (5.2).

THEOREM 5.2. *A nonnull matrix A and its associated elementary matrix R have unique decompositions of the form*

$$(5.3) \qquad A = \lambda_1 R_1 + \cdots + \lambda_k R_k \,, \qquad R = R_1 + \cdots + R_k$$

into $*$-orthogonal sections, where $\lambda_1, \cdots, \lambda_k$ are distinct positive numbers.

In order to prove this result let $\lambda_1 = \| A \|$. By virtue of the last theorem the matrices A and R are expressible as sums

$$(5.4) \qquad A = \lambda_1 R_1 + B \,, \qquad R = R_1 + S$$

of $*$-orthogonal sections with $\| B \| < \lambda_1$. If $B \neq 0$, choose $\lambda_2 = \| B \|$ and, by Theorem 5.1, again, B and S are sums

$$B = \lambda_2 R_2 + C \,, \qquad S = R_2 + T$$

of sections. Proceeding in this manner one obtains the representation described in the theorem.

The numbers $\lambda_1, \cdots, \lambda_k$ appearing in the last theorem will be called the *principal values* of A and the matrices $A_1 = \lambda_1 R_1, \cdots, A_k = \lambda_k R_k$ will be called the *principal sections* of A. The rank of A_i will be called the *multiplicity* of λ_i as a principal value of A. It is easily seen that a number $\lambda > 0$ is a principal value of A of multiplicity m if and only if it is an eigenvalue of R^*A (or AR^*) of multiplicity m. Similarly a number $\lambda > 0$ is a principal value of A if and only if λ^2 is an eigenvalue of A^*A (or of AA^*) and the multiplicities are the same. It is easily seen that a number λ is a principal value if and only if the equation $Ax = \lambda Rx$ has a solution x with $Rx \neq 0$.

It follows from the last theorem that the principal values of A are the norms of the nonnull sections of A and in particular the norms of the principal sections of A. A section B of rank 1 of the principal section $A_i = \lambda_i R_i$ of A is expressible in the form $B = \lambda_i yx^*$, where x and y are unit vectors. A vector of the form αx $(\alpha \neq 0)$ will be called a *principal vector* of A and a vector of the form $\beta y\, (\beta \neq 0)$ will be called a *reciprocal principal vector* of A corresponding to the principal value λ_i. It is easily seen that x is an eigenvector of A^*A and that y is an eigenvector of AA^*.

THEOREM 5.3. *A matrix A is normal if and only if its principal sections are normal. A matrix A is hermitian if and only if its principal sections are hermitian. A matrix A is hermitian and nonnegative if and only if its principal sections are hermitian and nonnegative.*

In order to prove this result let A and R be represented in the form (5.3) with $\lambda_1 > \lambda_2 > \cdots > \lambda_k$. Let

$$B = A - \lambda_k R = (\lambda_1 - \lambda_k)R_1 + \cdots + (\lambda_{k-1} - \lambda_k)R_{k-1} .$$

If A is normal so also are R and B. It follows that $S = R_1 + R_2 + \cdots + R_{k-1}$, the elementary matrix belonging to B, is also normal. This implies that $R_k = R - S$ and $A_k = \lambda_k R_k$ are normal. The same argument applied to B shows that R_{k-1} and $A_{k-1} = \lambda_{k-1} R_{k-1}$ are normal. It follows that each princiqal section $A_i = \lambda_i R_i$ $(1 \leqq i \leqq k)$ of A is normal whenever A is normal. Conversely if $A_1, \cdots, A_k$ are normal so also is A. This proves the first statement in the theorem. The second statement can be proved similarly. The third statement is an easy consequence of the second and the concept of nonnegativeness.

THEOREM 5.4. *Let X and Y be elementary matrices such that the relation*

$$YY^*A = AXX^* = A$$

holds for a given matrix A. Then $B = Y^*AX$ has the same principal *values as those of A and have the same multiplicities. If R is the associated elementary matrix for A, then $S = Y^*RX$ is the associated elementary matrix for B. There exist elementary matrices X and Y such that B is a nonnegative diagonal matrix and $R = YX^*$.*

Setting $S_i = Y^*R_iX$, $S = S_1 + \cdots + S_k$ it is found that $B = \lambda_1 S_1 + \cdots + \lambda_k S_k$. The ranks of S_i and R_i coincide and S_i $*$-commutes with S_j $(i \neq j)$. The last statement can be obtained by selecting a maximal set of mutually orthogonal principal vectors $x_1, \cdots, x_r$ of A of unit length and setting $y_h = Rx_h$ $(h = 1, \cdots, r)$. Let X be the matrix whose first r column vectors are $x_1, \cdots, x_r$ and the remaining vectors are null vectors. The matrix $Y = RX$ has $y_1, \cdots, y_r$ as its first r column vectors. It is easily seen that X and Y are elementary matrices of rank r having the properties described in the theorem. In fact the nonzero elements of $B = Y^*AX$ are the principal values of A. One could restrict X to have only r columns if one so desires. One could modify X and Y so as to be nonsingular. In this event we would have $R = YEX^*$, where $E = R^*R$. In either event the column vectors of $C = AX$ are mutually orthogonal and the lengths of the nonnull column vectors of C are the principal values of A. This fact can be used to devise a modified Jacobi method for finding the principal values of A. A discussion of a method of this type will be given by the author in a forthcoming paper.

6. Further properties of $*$-commutativity. Throughout the present section let A denote a given matrix and let R be its associated elementary matrix. Let

$$A = \lambda_1 R_1 + \cdots + \lambda_k R_k\,, \qquad R = R_1 + \cdots + R_k \tag{6.1}$$

be its decomposition into principal sections, given in Theorem 5.2. As before we set

$$A_i = \lambda_i R_i\,, \quad E_i = R_i^* R_i\,, \quad E_i' = R_i R_i^*\,, \quad E = R^*R\,, \quad E' = RR^*\,. \tag{6.2}$$

The first result to be established is given in the following.

THEOREM 6.1. *If a matrix B $*$-commutes with A, then it $*$-commutes with every matrix of the form*

$$C = \nu_1 R_1 + \cdots + \nu_k R_k \tag{6.3}$$

where $\nu_1, \cdots, \nu_k$ are real numbers. In particular B $*$-commutes with R and with each principal section A_j $(j = 1, \cdots, k)$. The matrix B is expressible uniquely as the sum

$$B = B_0 + B_1 + \cdots + B_k \tag{6.4}$$

of $*$-orthogonal sections such that B_i is $*$-orthogonal to A_j $(j \neq i)$ and

$B_i\,(i > 0)$ *$*$-commutes with A_i.*

In order to prove this result we may suppose that the principal sections of A have been ordered so that $\lambda_1 < \lambda_2 < \cdots < \lambda_k$. Recall that, by virtue of the corollary to Theorem 4.1, the matrix B not only $*$-commutes with A but also with

$$A^{*-1} = (1/\lambda_1)R_1 + \cdots + (1/\lambda_k)R_k \ .$$

It follows that B $*$-commutes with

$$C_2 = A - \lambda_1^2 A^{*-1} = \lambda_{22}R_2 + \cdots + \lambda_{k2}R_k$$

where $\lambda_{i2} = \lambda_i - \lambda_1(\lambda_1/\lambda_i) > 0\ (i = 2, \cdots, k)$. Moreover $\lambda_{i2} < \lambda_{j2}\,(i < j)$, as one readily verifies. Using the recursion formula

$$C_{j+1} = C_j - \lambda_{jj}^2 C_j^{*-1}\ , \qquad \lambda_{j+1,i} = \lambda_{ji} - \frac{\lambda_{jj}^2}{\lambda_{ji}} \qquad (i = j+1, \cdots, k)$$

one obtains matrices of the form

$$C_j = \lambda_{jj}R_j + \cdots + \lambda_{kj}R_k \qquad (j = 2, \cdots, k)$$

that $*$-commute with B. Moreover each R_i is a linear combination of $C_1 = A, C_2, \cdots, C_k$. It follows that B $*$-commutes with each of the sections $R_1, \cdots, R_k$ of R. Consequently B $*$-commutes with any matrix C of the form (6.3).

In order to prove that B is of the form (6.4) it follows from Theorem 4.1 with A replaced A_i that B is expressible in the form $B = C_i + B_i$ where C_i is $*$-orthogonal to A_i and $B_0 = B - (B_1 + \cdots + B_k)$ is $*$-orthogonal to each B_j and A_j $(j = 1, \cdots, k)$ and hence also to A. This completes the proof of the theorem.

THEOREM 6.2. *If a matrix B $*$-commutes with every section of A, then B is expressible in the form*

$$B = B_0 + \mu_1 R_1 + \cdots + \mu_k R_k \tag{6.5}$$

where $\mu_1, \cdots, \mu_k$ are real numbers and B_0 is $*$-orthogonal to A.

Let $B_0, B_1, \cdots, B_k$ be the sections of B given in (6.4). Since every section of A_i and hence every section of R_i $*$-commutes with B and hence with B_i it follows from Theorem 4.3 that B_i is of the form $B_i = \mu_i R_i + B_{i0}$, where μ_i is a real number and B_{i0} is $*$-orthogonal to R_i. It is clear from the definition of B_i that B_{i0} must be zero. This proves the theorem.

THEOREM 6.3. *A matrix B $*$-commutes with A if and only if it $*$-commutes with R and $AR^*B = BR^*A$.*

If B $*$-commutes with A, then B $*$-commutes with R and

$$AR^*B = AB^*R = BA^*R = BR^*A \ .$$

Conversely, suppose that B $*$-commutes with R and that $AR^*B = BR^*A$. Then $A^*RB^* = B^*RA^*$ and

$$A^*B = R^*RA^*B = R^*AR^*B = R^*BR^*A = B^*RR^*A = B^*A$$
$$AB^* = RR^*AB^* = RA^*RB^* = RB^*RA^* = BR^*RA^* = BA^* ,$$

as was to be proved.

COROLLARY. *If A is a positive definite hermitian matrix, then B $*$-commutes with A if and only if B is hermitian and $AB = BA$.*

This result is immediate since $R = I$ for a positive definite matrix.

It should be observed that if A is hermitian but not definite, then there are nonhermitian matrices that $*$-commute with A. For example the matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

$*$-commute even though B is not hermitian. However A is elementary and $B = AB^*A$, that is, B is hermitian relative to A.

THEOREM 6.4. *Let A be a matrix and let T be an elementary matrix such that $TA^*T = A$. Let B be a matrix that $*$-commutes with T. Then B $*$-commutes with A if and only if $AT^*B = BT^*A$.*

The proof is similar to that of the last theorem and will be omitted.

THEOREM 6.5. *Given a matrix B that $*$-commutes with A there exists a set of mutually $*$-orthogonal elementary matrices $T_1, \cdots, T_q$ with the property that A and B are expressible in the form*

$$A = \alpha_1 T_1 + \cdots + \alpha_q T_q , \qquad B = \beta_1 T_1 + \cdots + \beta_q T_q \tag{6.6}$$

where $\alpha_1, \cdots, \alpha_q$ are equal numbers, $\beta_1, \cdots, \beta_q$ are equal numbers and $\alpha_i = \alpha_j$, $\beta_i = \beta_j$ holds only in case $i = j$. If $\alpha_i \neq 0$, then $|\alpha_i|$ is a principal value of A. Similarly if $\beta_i \neq 0$ then $|\beta_i|$ is a principal value.

It is clear that T_i may be replaced by $-T_i$ in the theorem. The matrices $T_1, \cdots, T_q$ will be uniquely determined if, for example, one requires that $\beta_i \geqq 0$ and that $\alpha_i > 0$ if $\beta_i = 0$.

Let

$$A = \lambda_1 R_1 + \cdots + \lambda_k R_k , \qquad B = \mu_1 S_1 + \cdots + \mu_m S_m \tag{6.7}$$

be the decompositions of A and B respectively into principal sections. Recall that by virtue of Corollary 1 to Theorem 4.2 the matrices R_i and S_j are expressible as sums

$$R_i = R_{ij0} + R_{ij+} + R_{ij-} , \qquad S_j = S_{ij0} + R_{ij+} - R_{ij-} .$$

Let $T_1, \cdots, T_p$ be all non-null elementary matrices R_{ij+} and R_{ij-} obtained in this manner. Adjoin to these the maximal nonnull section of each R_i that is $*$-orthogonal to S and the maximal nonnull section of each S_j that is $*$-orthogonal to R. The elementary matrices $T_1, \cdots, T_q$ obtained in this manner are $*$-orthogonal and have the property that each R_i and S_j is expressible uniquely in the form

$$R_i = \rho_{i1}T_1 + \cdots + \rho_{iq}T_q\ , \qquad S_j = \sigma_{j1}T_1 + \cdots + \sigma_{jq}T_q$$

where $\rho_{ik} = 0$ if T_k is $*$-orthogonal to R_i, $\rho_{ih} = 1$ if T_h is a section of R_i, $\rho_{ih} = -1$ if $-T_h$ is a section of R_i, $\sigma_{jh} = 0$ if T_h is $*$-orthogonal to S_j, $\sigma_{jh} = 1$ if T_h is a section of S_j and $\sigma_{jh} = -1$ if $-T_h$ if a section of S_j. Combining this result with (6.7) one obtains (6.6). The last statement of the theorem follows from the construction just made.

As a consequence of this result we have

THEOREM 6.6. *Let T be an elementary matrix and let A be a matrix satisfying the condition $A = TA^*T$. Then A and T can be represented uniquely as the sum*

$$A = \alpha_1T_1 + \cdots + \alpha_qT_q\ , \qquad T = T_1 + \cdots + T_q \tag{6.7}$$

of mutually $*$-orthogonal matrices such that $\alpha_i \neq \alpha_j$ $(i \neq j)$.

This result follows from the last theorem with $B = T$ and the condition that $\beta_i \geqq 0$. Since no nonnull section of A is $*$-orthogonal to T we have $\beta_i = 1$, and the theorem follows.

If T is the identity then A is hermitian and $\alpha_1, \cdots, \alpha_q$ are the eigenvalues of A. The rank of T_i is the multiplicity of α_i as an eigenvalue of A. This result suggests that we call $\alpha_1, \cdots, \alpha_q$ the principal values or eigenvalues of A relative to T, the rank of T_i being the multiplicity of α_i.

As an extension of the last theorem we have

THEOREM 6.7. *Let T be an elementary matrix and let A and B be $*$-commutative matrices such that $A = TA^*T$ and $B = TB^*T$. There is a unique decomposition*

$$T = T_1 + \cdots + T_q$$

into sections such that A and B are representable in the form

$$A = \alpha_1T_1 + \cdots + \alpha_qT_q\ , \qquad B = \beta_1T_1 + \cdots + \beta_qT_q$$

where $\alpha_i = \alpha_j$, $\beta_i = \beta_j$ holds only in case $i = j$.

The proof of this result can be made by a simple modification of the proof of the last two theorems and will be omitted.

In the complex domain we have the following

COROLLARY 1. *Let T be an elementary matrix and let C be a* *matrix such that $TT^*C = CT^*T = C$ and $TC^*C = CC^*T$. Then C and T have unique decompositions*

$$C = \gamma_1 T_1 + \cdots + \gamma_q T_q , \qquad T = T_1 + \cdots + T_q$$

in sections, where $\gamma_1, \cdots, \gamma_q$ are distinct complex numbers.

For, by the corollary to Theorem 4.4, the matrix C is expressible in the form $C = A + iB$, where A and B $*$-commute with T. From the relation $TC^*C = CC^*T$ it is found that $AT^*B = BT^*A$ and hence that A and B $*$-commute. The corollary follows from the last theorem with $\gamma_j = \alpha_j + i\beta_j$ $(j = 1, \cdots, q)$.

If $T = I$, the result described in the corollary yields the spectral decomposition for normal matrices.

By the use of an argument like that given in the proof of Theorem 5.4 one obtains the further result described in the following

COROLLARY 2. *Let A and B be $*$-commutative matrices and let T be an elementary matrix such that $A = TA^*T$ and $B = TB^*T$. There exist elementary matrices X and Y such that*

$$YY^*T = TXX^* = T$$

and such that Y^*AX, Y^*BX are diagonal matrices. If $C = A + iB$, then Y^*CX is also a diagonal matrix.

7. Certain classes of matrices. Let $\mathscr{S}(A)$ be the class of matrices B that $*$-commute with A and have no non-null section that is $*$-orthogonal to A. Let $\mathscr{C}(A)$ be all matrices B such that $\mathscr{S}(A)$ is a subclass of $\mathscr{S}(B)$. It is clear that A is in $\mathscr{C}(A)$. We have the following

THEOREM 7.1. *Let R be the elementary matrix associated with A and let*

$$A = \lambda_1 R_1 + \cdots + \lambda_k R_k , \qquad R = R_1 + \cdots + R_k$$

be the decomposition of A into principal sections. The class $\mathscr{C}(A)$ consists of all matrices B that are expressible in the form

$$B = \mu_1 R_1 + \cdots + \mu_k R_k \tag{7.1}$$

where $\mu_1, \cdots, \mu_k$ are real numbers. If B is in $\mathscr{C}(A)$ so also is B^{*-1} and its associated elementary matrix S.

This result follows from Theorems 6.1 and 6.2.

COROLLARY. *If B is in $\mathscr{C}(A)$ then $\mathscr{C}(B) \subset \mathscr{C}(A)$. Moreover $\mathscr{C}(B) = \mathscr{C}(A)$ if and only if B has the same number of distinct principal values as A.*

As a further result we have

THEOREM 7.2. *If B, C, D are matrices in $\mathscr{C}(A)$ so also is $M = BC^*D$.*

In fact

$$BC^*D = BD^*C = CB^*D = CD^*B = DB^*C = DC^*B\,. \tag{7.2}$$

The relations (6.2) follow from the fact that B, C, D $*$-commute with each other. Observe that $M^* = B^*CD^*$. If N is a matrix in $\mathscr{C}(A)$ then

$$N^*M = N^*BC^*D = B^*NC^*D = B^*CN^*D = B^*CD^*N = M^*N\,.$$

Similarly $NM^* = MN^*$. This proves the theorem.

In view of the formula (7.1) one obtains the following

THEOREM 7.3. *Let*

$$B = \mu_1 R_1 + \cdots + \mu_k R_k\,, \qquad C = \nu_1 R_1 + \cdots + \nu_k R_k \tag{7.3}$$

be two matrices in $\mathscr{C}(A)$. Then

$$\alpha B + \beta C = (\alpha\mu_1 + \beta\nu_1)R_1 + \cdots + (\alpha\mu_k + \beta\nu_k)R_k$$

is in $\mathscr{C}(A)$ for every pair of real numbers α and β. If we define the product of B and C by the formula $B\cdot C = BR^*C$, then

$$B\cdot C = \mu_1\nu_1 R_1 + \cdots + \mu_k\nu_k R_k$$

is in $\mathscr{C}(A)$ and the usual laws of algebra hold. In particular $B\cdot R = R\cdot B = B$. Given a polynomial

$$p_m(\lambda) = \alpha_0 + \alpha_1\lambda + \cdots + \alpha_m\lambda^m$$

with real coefficients set

$$p_m(A, R) = \alpha_0 A^{(0)} + \alpha_1 A^{(1)} + \cdots + \alpha_{\hat{n}} A^{(n)}\,,$$

where $A^{(0)} = R$, $A^{(1)} = A$, $A^{(h)} = A\cdot A^{(h-1)}$. Then the polynomial

$$p_m(A, R) = p_m(\lambda_1)R_1 + \cdots + p_m(\lambda_k)R_k$$

in A relative to R is in $\mathscr{C}(A)$. Conversely every matrix B in $\mathscr{C}(C)$ is expressible as a real polynominal in A relative to R of degree $\leqq k-1$. There is a unique polynomial $p_k(\lambda)$ with leading coefficient $\alpha_k = 1$ such that $p_k(A, R) = 0$.

The first three statements in the lemma are immediate. The matrix B is given by the relative polynomial $p_{k-1}(A, R)$ whose coefficients $\alpha_0, \alpha_1, \cdots, \alpha_{k-1}$ are given by the solutions of the equations

$$\alpha_0\lambda_1^h + \alpha_1\lambda_2^h + \cdots + \alpha_{k-1}\lambda_k^h = \mu_h \qquad (h = 0, 1, \cdots, k-1)\,.$$

Finally the polynomial $p_k(\lambda)$ described in the last statement in the theorem is the polynomial of degree k whose roots are $\lambda_1, \lambda_2, \cdots, \lambda_k$.

COROLLARY. *On the class $\mathscr{C}(A)$ the norm $\|B\|$ satisfies the rela-*

tion $\|B\cdot C\| \leqq \|B\|\,\|C\|$.

As a final property of the classes $\mathscr{C}(A)$ we have

THEOREM 7.4. *If B is in $\mathscr{S}(A)$ there is a matrix C such that A and B are in $\mathscr{C}(C)$.*

This result follows from Theorem 6.5 with

$$C = T_1 + 2T_2 + 3T_3 + \cdots qT_q\,.$$

Let T be a given elementary matrix and let A be a matrix in $\mathscr{S}(T)$. Let $\mathscr{S}(A, T)$ be all matrices in $\mathscr{S}(T)$ that $*$-commute with A. Let $\mathscr{C}(A, T)$ be all matrices B in $\mathscr{S}(A, T)$ such that $\mathscr{S}(A, T) \subset \mathscr{S}(B, T)$. If T is the elementary matrix R associated with A then $\mathscr{C}(A, T) = \mathscr{C}(A)$. Let

$$A = \alpha_1 T_1 + \cdots + \alpha_q T_q\,, \qquad T = T_1 + \cdots T_q$$

be the decomposition of A and T given in Theorem 6.6. Then, as is easily seen, the class $\mathscr{C}(A, T)$ consists of all matrices of the form

$$B = \beta_1 T_1 + \cdots + \beta_q T_q$$

where $\beta_1, \cdots, \beta_q$ are real numbers. If we set

$$C = T_1 + 2T_2 + \cdots + qT_q$$

then the class $\mathscr{C}(A, T)$ coincides with the class $\mathscr{C}(C)$. Consequently the results stated above are applicable to the class $\mathscr{C}(A)$. If T is the identity, then A is hermitian and $\mathscr{C}(A, T)$ consists of all hermitian matrices that commute with every hermitian matrix that commutes with A.

Consider now an elementary matrix T and let $\mathfrak{N}(T)$ be the class of all matrices A such that $TT^*A = AT^*T = A$ and $AA^*T = TA^*A$. Given a matrix A in $\mathfrak{N}(T)$ let $\mathfrak{M}(A, T)$ be the class of all matrices B in $\mathfrak{N}(T)$ such that $AT^*B = BT^*A$ and $AB^*T = TB^*A$. If B is in $\mathfrak{M}(A, T)$ then $BA^*T = TA^*B$ also. Moreover, TB^*T is in $\mathfrak{M}(A, T)$. Let $\mathscr{B}(A, T)$ be the class of all matrices B such that $\mathfrak{M}(B, T) \supset \mathfrak{M}(A, T)$. In view of Corollary 1 to Theorem 6.7 the matrices A and T are expressible uniquely in the form

$$A = \alpha_1 T_1 + \cdots + \alpha_q T_q\,, \qquad T = T + \cdots + T_q\,,$$

where $\alpha_1, \alpha_2, \cdots, \alpha_q$ are distinct complex numbers. It is not difficult to show a matrix B is in $\mathscr{B}(A, T)$ if and only if it is expressible in the form

$$B = \beta_1 T_1 + \cdots + \beta_q T_q\,,$$

where $\beta_1, \cdots, \beta_q$ are complex numbers. If B and C are in $\mathscr{B}(A, T)$ so also are $\alpha B + \beta C$, where α and β are complex numbers. Moreover, the product $B \cdot C = BT^*C$ is in $\mathscr{B}(A, T)$. If polynomials of A relative to T are defined as before, but with complex coefficients, it is seen that the class $\mathscr{B}(A, T)$ is made up of all polynomials of A relative to T of degree $\leqq q - 1$. Again we have the relation $\| B \cdot C \| \leqq \| B \| \| C \|$. These results generalize the corresponding theory for normal matrices.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

ON SIMILARITY INVARIANTS OF CERTAIN OPERATORS IN L_p

G. K. KALISCH

The purpose of this paper is to extend the result of Corollary, Theorem 2 of the author's paper on Volterra operators (Annals of Math., 66, 1957, pp. 481–494 quoted as A; we shall use the definitions and notations of that paper) to the most general situation applicable: We are dealing with operators T_F where $F(x, y) = (y - x)^{m-1}\, aG(x, y)$ is a function defined on the triangle $0 \leqq x \leqq y \leqq 1$, where m is a positive integer, a a complex number of absolute value 1, G is a complex valued function which is continuously differentiable and $G(x, x)$ is positive real. We recall that if $f \in L_p\,[0, 1]$, then $(T_F)(f)(x) = \int_x^1 F(x, y) f(y) dy$ is again in $L_p\,[0, 1]$. The only difference from A is the presence of the constant a which affects none of results except Theorem 2 and its Corollary. Theorems 1 and 2 of the present paper fill the gap. Theorem 3 shows that differentiability conditions imposed on F cannot be abandoned entirely—and also that the integral equation (1) of A cannot be solved unless K (which corresponds to our F) has at least first derivatives near $y = x$.

If c is constant and E is the function identically equal to 1, we define T_E^c as T_H which $H(x, y) = (y - x)^{c-1}/\Gamma(c)$ (fractional integration of order c).

THEOREM 1. *Let c_1 and c_2 be complex numbers and let r_1 and r_2 be real numbers such that $r_i \geqq 1$, then $c_1 T_E^{r_1}$ is similar to $c_2 T_E^{r_2}$ if and only if $c_1 = c_2$ and $r_1 = r_2$.*

Proof. The first part of the Proof of Theorem 2 of A applies and implies that $r_1 = r_2$ $(= r)$ and $|c_1| = |c_2|$. Thus suppose that $c_1 T_E^r$ is similar to $c_2 T_E^r$ or that $c T_E^r$ is similar to

$$T_E^r = P c T_E^r P^{-1} \text{ for } |c| = 1 \tag{1}$$

where P is a bounded linear transformation of $L_p\,[0, 1]$ onto itself with the bounded linear inverse P^{-1}. If T is similar to $S = PTP^{-1}$, then $f(T)$ is similar to

$$f(S) = P f(T) P^{-1} \tag{2}$$

for polynomials and even analytic functions f. Let

Received February 29, 1960. The author was supported in part by the United States Air Force under Contract No. AF 49(638)-64 and by a grant from the National Science Foundation.

$$f(z) = \sum_{i=0}^{\infty} a_i z^{i+1}$$

Then

$$f(cT_E^r) = \sum_{i=0}^{\infty} a_i c^{i+1} T_E^{r(i+1)} = T_{g_1(y-x)}$$

where $g_1(t) = ct^{r-1}g(ct^r)$ where we have written t for $y - x$ and where

$$g(z) = \sum_{i=0}^{\infty} b_i z^i$$

with $b_i = a_i/\Gamma(r(i+1))$. Equations (1) and (2) imply that $\| f(T_E^r) \| \leqq \| P \| \| P^{-1} \| \| f(cT_E^r) \|$. The definition of the norm of a linear transformation in a Banach space implies the following inequality:

$$\| f(T_E^r) \| = \| T_{t^{r-1}g(t^r)} \| \geqq \left\| \int_x^1 (y - x)^{r-1} g((y - x)^r) k(y) dy \right\|_p$$

for all $k \in L_p\,[0, 1]$ such that $\| k \|_p = 1$. On the other hand, Lemma 2 of A implies that

$$\| T_{ct^{r-1}g(t^r)} \| \leqq \| ct^{r-1}g(ct^r) \|_1 = \| t^{r-1}g(ct^r) \|_1 \,.$$

Thus if $k(y) = 1$, we obtain

$$\begin{aligned} L = \left\| \int_x^1 (y - x)^{r-1} g((y - x)^r) dy \right\|_p &\leqq \| f(T_E^r) \| \\ &\leqq \| P \| \| P^{-1} \| \| f(cT_E^r) \| \\ &\leqq \| P \| \| P^{-1} \| \| t^{r-1} g(ct^r) \|_1 = R \,. \end{aligned} \tag{3}$$

We shall find a family of functions g_v (and correspondingly f_v) depending on a positive parameter v such that if we use the notations L_v and R_v for the corresponding left and right hand sides of (3), $L_v \to \infty$ and $R_v \to 0$ as $v \to \infty$ contradicting the inequality (3): this contradiction then proves our theorem.

Let us first consider the case where the real part of c, $Re(c)$, is less than 0. Let $g_v(t) = \exp(vt)$. Since T_E^r is generalized nilpotent for $r \geqq 1$, the corresponding function $f_v(T_E^r)$ exists and (1) indeed implies (2) for $S = T_E^r$ and $T = cT_E^r$. Then

$$R_v = \| t^{r-1} g_v(ct^r) \|_1 = \int_0^1 | t^{r-1} \exp(vct^r) | \, dt$$

and $R_v \to 0$ as $v \to \infty$. On the other hand

$$L_v = (1/r^p) \int_0^1 (\exp(v(1 - x)) - 1/v)^p dx \to \infty$$

as $v \to \infty$. If finally $Re(c) \geqq 0$ and $c \neq 1$, then there exist a positive integer n such that $Re(c^n) < 0$. But then (1) implies that $c^n T_E^{nr}$ is similar to $T_E^{nr} = Pc^n T_E^{nr} P^{-1}$ which contradicts the preceding result and the proof of the theorem is complete.

THEOREM 2. *Let $F(x, y) = (y - x)^{m-1} aG(x, y)$ satisfy, in addition to the general hypotheses stated above, one of the following*:

(1) *G is analytic in a suitable region and m is arbitrary*;

(2) *$G(x, y) = G(y - x)$, $G(0) \neq 0$, $G \in C^2$ and m is arbitrary*;

(3) *$G \in C^2$ and $m = 1$. Let A be a complex number. Then $AI + T_F$ and $AI + T_F^*$ are similar to the unique operator $AI + caT_E^m$ and $AI + c\bar{a}T_E^m$ respectively where $c = \left(\int_0^1 (G(u, u))^{1/m} du\right)^m$.*

Here I is the identity operator and T_K^*, the adjoint of T_K, is defined by

$$(T_K^*)(f)(x) = \int_0^x \overline{K(y, x)} f(y) dy \, .$$

Proof. Note first that A implies that $AI + T_F$ is similar to $AI + caT_E^m$ and that $AI + T_F^*$ is similar to $AI + c\bar{a}T_E^{*m}$ (see Cor. Theorem 2 of A). Observe next that $T_E^* f(x) = \int_0^x f(y) dy$ and

$$T_E^{*m} f(x) = (1/\Gamma(m)) \int_0^x (x - y)^{m-1} f(y) \, dy$$

and that if $(S_{1-x} f)(x) = f(1 - x)$ then S_{1-x} is an isometry of $L_p[0, 1]$ onto itself and $S_{1-x} T_E^m S_{1-x}^{-1} = T_E^{*m}$. It remains to show uniqueness. Suppose that $A_1 I + c_1 a_1 T_E^{m_1}$ is similar to $A_2 I + c_2 a_2 T_E^{m_2}$. Then $A_1 = A_2$ (because of the complete continuity of T_E) and $c_1 a_1 T_E^{m_1}$ is similar to $c_2 a_2 T_E^{m_2}$ which by Theorom 1 implies that $c_1 = c_2$, $a_1 = a_2$, $m_1 = m_2$.

THEOREM 3. *The linear transformation $T_E + T_E^{1+a}$ where $0 < a < 1$ of $L_p[0, 1]$ into itself is not similar to any linear transformation cT_E^r for complex c and real $r \geqq 1$.*

Proof. Preliminaries. 1. If two linear transformations S and T are similar, i.e., if there exists P such that $S = PTP^{-1}$, then there exists a constant K such that

$$1/K \leqq \| T^n \| / \| S^n \| \leqq K \, , \tag{4}$$

for all positive integers n. It suffices to take $K = \| P \| \| P^{-1} \|$.

2. The following inequality is a consequence of the fact that if $0 \leqq F_1(x, y) \leqq F_2(x, y)$ then $\| T_{F_1} \| \leqq \| T_{F_2} \|$:

$$\| (T_E + T_E^{1+a})^n \| \geqq n \| T_E^{n+a} \| \tag{5}$$

for all positive integers n.

3. Our next task is to find estimates for $\| T_E^n \|$. An estimate from above is the following:

$$\| T_E^n \| \leqq 1/(n\Gamma(n)p^{1/p}) \tag{6}$$

for all positive integers n. An estimate from below is furnished by the following Proposition:

Given the real positive number e there exists a positive number $K = K(e)$ and a positive integer $N = N(e)$ such that for all integers $n \geqq N$,

$$\| T_E^n \| \geqq K/(n^{1+e}\Gamma(n)) . \tag{7}$$

Proof of (6). If $f \in L_p[0, 1]$,

$$T_E^n f(x) = \int_x^1 [(y-x)^{n-1}/\Gamma(n)] f(y) dy .$$

If $(1/p) + (1/q) = 1$, Hölder's inequality yields

$$\begin{aligned}\int_x^1 (y-x)^{n-1} f(y) dy &\leqq \left(\int_x^1 (y-x)^{(n-1)q} dy\right)^{1/q} \| f \|_p \\ &= (1-x)^{((n-1)q+1)/q} \| f \|_p / (((n-1)q+1)^{1/q})\end{aligned}$$

so that

$$\begin{aligned}&\| T_E^n f \|_p^p \\ &= \int_0^1 | (T_E^n f)(x) |^p dx \\ &= (1/\Gamma(n))^p \int_0^1 \left| \int_x^1 (y-x)^{n-1} f(y) dy \right|^p dx \\ &\leqq (1/\Gamma(n))^p (1/((n-1)q+1)^{p/q}) \int_0^1 (1-x)^{((n-1)p+(p/q))} dx \| f \|_p^p \\ &= (1/\Gamma(n))^p (1/((n-1)q+1)^{p/q}) (1/((n-1)p+(p/q)+1)) \| f \|_p^p\end{aligned}$$

which implies that

$$\| T_E^n \| \leqq (1/\Gamma(n))(1/((n-1)q+1)^{1/q})(1/((n-1)p+(p/q)+1)^{1/p})$$

which in turn implies (6).

Proof of (7). We first observe that elementary considerations concerning the gamma function imply that given c such that $0 < c < 1$ and given a positive real number d there exists an integer N depending on c and d such that for all integers $n \geqq N$

$$(8) \qquad \Gamma(n + c) < (n + c)^{c+d}\Gamma(n) .$$

Consider next the function $f(x) = r(1 - x)^{-s} \in L_p\,[0, 1]$ such that $\|f\|_p = 1$, i.e., $r^p = 1 - sp$ and $0 < s < 1/p$. Then

$$T_E^n f(x) = r\Gamma(1 - s)(1 - x)^{n-s}/\Gamma(n + 1 - s)$$

and

$$\| T_E^n \| \geqq r\Gamma(1 - s)/\Gamma(n + 1 - s)(p(n - s) + 1)^{1/p} .$$

We now choose s (and hence r) such that for the positive real number e of (7), $0 < (1/p) - s < e$ and then we choose d such that $0 < d < e + s - (1/p)$ and finally by virtue of (8) we obtain N as a function of e such that for all integers $n \geqq N$, $\Gamma(n + 1 - s) < (n + 1 - s)^{1-s+d}\Gamma(n)$ whence

$$\| T_E^n \| \geqq r\Gamma(1 - s)/(n + 1 - s)^{1-s+d}\Gamma(n)(p(n - s) + 1)^{1/p}$$

which upon choosing $K = K(e)$ properly implies (7).

After these preliminaries, we turn to the proof of the theorem. We distinguish several cases. Let $T = T_E + T_E^{1+a}$.

Case 1. $|c| \leqq 1$. Consider

$$h_n = \| (cT_E^r)^n \| / \| T^n \| \leqq \| T_E^n \| / (n \| T_E^{n+a} \|)$$

where we have used (5) and the fact that $r \geqq 1$. Take now positive real numbers e and d such that $a + e + d < 1$. Then there exists by (7) a positive constant K and an integer N such that for all integers $n \geqq N$

$$(9) \qquad \begin{aligned} h_n &\leqq (n + a)^{1+e}\Gamma(n + a)/(n^2\Gamma(n)p^{1/p}K) \\ &\leqq (n + a)^{1+e+a+d}\Gamma(n)/(n^2\Gamma(n)p^{1/p}K) \end{aligned}$$

where we have made use of (8) and (6). The last inequality implies that $h_n \to 0$ which in conjunction with (4) implies the truth of our theorem in the case under consideration.

Case 2. $r < 1$. Using the notations and making similar choices as under Case 1, (9) becomes

$$h_n \leqq |c|^n(n + a)^{1+e+a+d}\Gamma(n)/(n^2 r\Gamma(rn)p^{1/p}K)$$

which, since $|c|^n\Gamma(n)/\Gamma(rn)$ is bounded (in fact converges to 0) for $r > 1$ as $n \to \infty$, again proves the truth of the theorem in the present case.

Case 3. $r = 1$, $|c| > 1$. This time we consider the quotient

$$
\begin{aligned}
k_n &= \| T^n \| / \| (cT_E)^n \| \\
&\leqq \sum_{i=0}^{n} \binom{n}{i} \| T_E^{n+a(n-i)} \| / (| c |^n \| T_E^n \|) \qquad (10) \\
&\leqq ((n^{1+e}\Gamma(n)/(| c |^n K p^{1/p})) \sum_{i=0}^{n} \binom{n}{i} / (\Gamma(n + a(n-i) + 1)) ,
\end{aligned}
$$

which is valid for sufficiently large n; again we used (6) and (7).

In order to complete the proof of our theorem, we need the following fact:

Given any positive real number e and given the positive real number $a < 1$, there exists an integer $N = N(e; a)$ such that for all integers i and n such that $0 \leqq i \leqq n \leqq N$

$$\Gamma(n)/\Gamma(n + a(n-i) + 1) \leqq 2e^{n-i} . \tag{11}$$

Proof. The case $i = 0$ results from elementary considerations about the gamma function. If $i = 1$, we find N_1 so that (11) is valid for $i = 0$ and $n \geqq N_1$. We then find N_2 so that (8) is true for some arbitrary but fixed d, for $c = a$ and for $n \geqq N_2$. Then $\Gamma(n)/\Gamma(n+(n-1)a+1) \leqq (\Gamma(n)/\Gamma(n + na + 1))/(n + na + 1)^{a+d}$ which for $n \geqq \max(N_1, N_2, e^{-1/a}) = N_3$ implies (11) for $i = 2$ and $n \geqq N_3$. The remaining cases are settled by induction (except $i = n$ which is obvious); note that we never have to go above N_3 at any point. This completes the proof of (11).

The proof is now completed by substituting (11) into (10):

$$k_n \leqq 2n^{1+c}(1 + e_1)^n / | c |^n K p^{1/p}$$

where e_1 is the constant e of (11). Thus $k_n \to 0$ upon proper choice of e_1 and our theorem is again true in view of (4). This completes the proof of Theorem 3.

UNIVERSITY OF MINNESOTA

THE STONE-WEIERSTRASS PROPERTY IN BANACH ALGEBRAS

YITZHAK KATZNELSON AND WALTER RUDIN*

Introduction. Let A be a semi-simple commutative Banach algebra with maximal ideal space Δ. Regarding the elements of A as functions on Δ, we call a subalgebra B of A *self-adjoint* if corresponding to every $f \in B$ the function $\bar{f}$ defined on Δ by $\bar{f}(x) = \overline{f(x)}$ is also in B; we call B *separating* if to every pair of distinct points $x_0, x_1 \in \Delta$ there is an $f \in B$ such that $f(x_0) = 0$, $f(x_1) = 1$.

If every separating self-adjoint subalgebra of A is dense in A, we say that *A has the Stone-Weierstrass property.*

The Stone-Weierstrass property is related, to some extent at least, to the ideal structure of A. For instance, it is obvious that if A has a unit and a closed primary ideal I which is not maximal, then the algebra generated by I and the constants is not dense in A. More generally, suppose A is self-adjoint, I is a closed self-adjoint ideal in A which is not the intersection of the regular maximal ideals containing it, and A/I is the direct sum of its radical and a subalgebra B_0. If h is the canonical homomorphism of A onto A/I, then $I + h^{-1}(B_0)$ is a separating self-adjoint subalgebra of A which is not dense in A, so that A does not have the $S - W$ property.

Also, it was pointed out by Herz that the Schwartz counterexample [9] to spectral synthesis in $L^1(R^3)$ yields immediately an example of a closed, separating, self-adjoint, *proper* subalgebra of $L^1(R^3)$. After Malliavin's solution of the spectral synthesis problem for $L^1(\Gamma)$, where Γ is any locally compact abelian group, it was natural to investigate the $S - W$ property for these algebras.

In Part I (whose contents were announced in [5]) this is done for $\Gamma = Z$, the additive group of the integers. The general case is settled in Part II; the solution shows that the relation between the $S - W$ property and the ideal structure is, after all, not a very close one. Part III deals with the relation between the self-adjointness of A and the total disconnectedness of Δ.

For convenience of notation, we shall phrase our results on group algebras in $A(G)$ rather than in $L^1(\Gamma)$. Here G and Γ are dual groups of each other, and $A(G)$ is the algebra of all Fourier transforms of functions in $L^1(\Gamma)$. The circle group (the dual of Z) will be denoted by T, so that $A(T)$ is the algebra of all absolutely convergent Fourier series.

Since every locally compact abelian group is locally isomorphic to a

Received April 4, 1960. * Research Fellow of the Alfred P. Sloan Foundation.

compact group, nothing of interest is lost by restricting our attention to algebras $A(G)$ with G compact.

PART I

LEMMA 1.1. *If g and its derivative g' are in $A(T)$, and if $\varepsilon > 0$, $0 < \delta < \pi$, there exists a function φ on T with the following properties:*

(i) $0 \leqq \varphi \leqq 1$;

(ii) $\varphi \equiv 1$ *in some neighborhood of* 0, $\varphi \equiv 0$ *outside* $(-\delta, \delta)$;

(iii) *if* $h = \varphi g$, *then* $|n\hat{h}(n)| < \varepsilon \, (n = 0, \pm 1, \pm 2, \cdots)$.

Here $\hat{h}(n)$ denotes the nth Fourier coefficient of h:

$$\hat{h}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} h(x)e^{-inx}dx \,. \tag{1}$$

Proof. Let u be a continuous odd function on the real line, vanishing outside the segments $(-2, -1)$ and $(1, 2)$, such that $\int_1^2 u(s)ds = -1$, and put

$$\hat{u}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} u(s)e^{-ist}ds \qquad (t \text{ real}). \tag{2}$$

Note that

(a) $\hat{u}(0) = 0$,

(b) $\hat{u}$ is continuous,

(c) $\|\hat{u}\|_\infty < 1$, and

(d) $\hat{u}(t) \to 0$ as $|t| \to \infty$.

For $r = 1, 2, 3, \cdots$, put $u_r(x) = ru(rx)$. Then $\hat{u}_r(t) = \hat{u}(t/r)$, and the above mentioned properties of $\hat{u}$ show that there exists a sequence of positive integers r_i (which must increase sufficiently rapidly), so that

$$|\hat{u}_{r_1}(t) + \cdots + \hat{u}_{r_k}(t)| < 1 \qquad (k = 1, 2, 3, \cdots; t \text{ real}). \tag{3}$$

Take r_1 and k so large that $r_1 > 2/\delta$ and

$$\left(\frac{1}{k} + \frac{2}{\pi r_1}\right)(\|g\|_A + \|g'\|_A) < \varepsilon \,. \tag{4}$$

(The subscripts A indicate that the norms are taken in $A(T)$.) Define

$$v = \frac{1}{k}(u_{r_1} + \cdots + u_{r_k}) \,, \tag{5}$$

$$\varphi(x) = \int_{-\pi}^{x} v(t)dt \,. \tag{6}$$

Our construction shows immediately that φ has properties (i) and (ii) of the lemma. If $h = \varphi g$, then $h' = vg + \varphi g'$. Note that $|\hat{\varphi}(0)| \leqq 2/(\pi r_1)$

and that

$$|\hat{\varphi}(n)| = |\hat{v}(n)/n| \leqq |\hat{v}(n)| \leqq 1/k$$

for $n \neq 0$, by (3) and (5). Thus

$$\sup_n |n\hat{h}(n)| = \| \hat{h}' \|_\infty \leqq \| \hat{v} \|_\infty \| g \|_A + \| \hat{\varphi} \|_\infty \| g' \|_A$$
$$\leqq \frac{1}{k} \| g \|_A + \left(\frac{1}{k} + \frac{2}{\pi r_1}\right) \| g' \|_A < \varepsilon ,$$

by (4), and the lemma is proved.

THEOREM 1.2. *$A(T)$ does not have the Stone-Weierstrass property.*

Proof. We shall construct a totally disconnected perfect set P on T and a function f, not equivalent to 0, which vanishes outside P, such that $|n\hat{f}(n)|$ is bounded.

Once this is done, we let $B[P]$ be the algebra of all twice continuously differentiable functions g on T such that $g'(x) = 0$ for every $x \in P$. Since P is totally disconnected, $B[P]$ is a separating (and evidently self-adjoint) subalgebra of $A(T)$. The bounded sequence $\{n\hat{f}(n)\}$ defines a non-zero bounded linear functional U on $L^1(Z)$, hence on $A(T)$, and U annihilates $B[P]$: for $g \in B[P]$, the Fourier series of g' converges uniformly, and we have

$$Ug = \sum_{-\infty}^{\infty} \hat{g}(-n) n\hat{f}(n) = \frac{i}{2\pi} \int_{-\pi}^{\pi} f(x) g'(x) dx = 0 . \tag{7}$$

Hence $B[P]$ is not dense in $A(T)$, and the theorem follows.

Now to the construction of f and P. Put $f_0 = 1$, and suppose that f_i is constructed, so that

$$|n\hat{f}_i(n)| \leqq 1 - 2^{-i} \qquad (n = 0, \pm 1, \pm 2, \cdots) . \tag{8}$$

Let I_i be the largest interval on which f_i is identically 1, let x_i be the midpoint of I_i, and choose φ_i (by Lemma 1.1) such that

(i) $0 \leqq \varphi_i \leqq 1$;

(ii) $\varphi_i \equiv 1$ in a neighborhood of x_i, $\varphi_i \equiv 0$ outside $(x_i - \delta_i, x_i + \delta_i)$, where $\delta_i = 2^{-i-1}$;

(iii) if $h_i = f_i \varphi_i$, then $|n\hat{h}_i(n)| < 2^{-i-1}$ for $n = 0, \pm 1, \pm 2, \cdots$.

Define $f_{i+1} = f_i(1 - \varphi_i)$. Then

$$|n\hat{f}_{i+1}(n)| \leqq |n\hat{f}_i(n)| + |n\hat{h}_i(n)| < 1 - 2^{-i-1} ,$$

so that our induction hypothesis (8) is satisfied with $i + 1$ in place of i.

The sequence $\{f_i\}$ converges monotonically to a function f. Applying the Lebesgue convergence theorem to the computation of $\hat{f}(n)$, (8) shows that $|n\hat{f}(n)| \leqq 1$. Our choice of the points x_i shows that $f \equiv 0$ on a dense open set V. Let P be the complement of V. Finally, observe that $f = 1$ at those points at which every φ_i is 0, and this happens on the complement of a set whose measure does not exceed $2\sum_0^\infty \delta_i < 2\pi$. Hence $f = 1$ on a set of positive measure. This completes the proof.

1.3. It was essential in our preceding construction to have P of positive measure. For suppose $m(P) = 0$ and $B[P]$ is defined as in the proof of Theorem 1.2. If $\{c_n\}$ is any bounded sequence such that $\sum_{-\infty}^{\infty} c_n \hat{g}(-n) = 0$ for every $g \in B[P]$, and if $f(x) \sim \sum_{n\neq 0}(c_n/n)e^{inx}$, a computation analogous to (7) shows that $\int fg' = 0$ for every $g \in B[P]$. It follows that f must vanish outside P, and since $m(P) = 0$, $c_n = 0$ for $n \neq 0$. Then c_0 much also be 0, and we conclude that $B[P]$ is dense in $A(T)$.

1.4. However, measure theoretic conditions on P are not enough. To show this, we shall now construct a totally disconnected perfect set P of positive measure, such that $B[P]$ is dense in $A(T)$. Our construction will also show that for every function $g \in A(T)$ there exist differentiable g_n such that $\|g - g_n\| \to 0$ and $g'_n \to 0$ a.e.

Put $n_k = 2^{k+1}$, $N_k = n_1 n_2 \cdots n_k$, let $k = 1, 2, 3, \cdots$,

$$L_{k,j} = \left(\frac{2\pi(j-1)}{N_k}\,,\quad \frac{2\pi j}{N_k}\right),$$

and let I_k be the union of those $L_{k,j}$ which have $j \equiv 1 \pmod{n_k}$. The desired set P is the intersection of the complements of the I_k.

Since $m(I_k) = 2\pi/n_k$, we see that $m(P) \geqq \pi$.

Let g_k be the characteristic function of I_k, and define

$$f_k(x) = n_k \int_0^x g_k(t)dt\ .$$

Then $f_k(2\pi) = 2\pi$, hence $e^{if_k} \in A(T)$, and since f_k is constant on each interval of the complement of I_k, e^{if_k} belongs to the closure $\bar{B}$ of $B[P]$.

Next, $f_k(x) - x = h_k(N_{k-1}x)$, where $h_k(0) = h_k(2\pi) = 0$, h_k is linear on $[0, 2\pi/n_k]$ and on $[2\pi/n_k, 2\pi]$, h_k has period 2π, and

$$h_k\left(\frac{2\pi}{n_k}\right) = \frac{2\pi}{N_{k-1}} - \frac{2\pi}{N_k}\ .$$

Then $\|f_k(x) - x\| = \|h_k(N_{k-1}x)\| = \|h_k\|$. Computation of the Fourier coefficients of h_k (see 1.5 below) shows that $\|h_k\| \to 0$ as $k \to \infty$. It follows that

$$\|e^{ix} - e^{if_k(x)}\| = \|1 - e^{i(f_k(x)-x)}\| \to 0$$

on $k \to \infty$, so that $\bar{B}$ contains the function e^{ix}.

Similarly, $e^{-ix} \in \bar{B}$, hence $\bar{B} = A(T)$, and $B[P]$ is dense in $A(T)$.

1.5. Suppose $0 < a < 1/2$, $b > 0$. Let h be linear on $[0, a]$ and on $[a, 2\pi]$, such that $h(0) = h(2\pi) = 0$, and $h(a) = b$. Then

$$h(x) = \frac{b}{2} - \frac{b}{2\pi - a} \sum_{n \neq 0} \frac{1 - e^{-ina}}{n^2 a} e^{inx} .$$

Hence $\|h\|$ (the norm being taken in $A(T)$) is bounded by $Cb \log 1/a$, where C is an absolute constant.

1.6. In our proof of Theorem 1.2 we could have chosen the intervals on which $\varphi_i \equiv 1$ so that the resulting set P satisfies a certain arithmetic condition which assures that P is a set of spectral synthesis. The condition we have in mind is due Herz [3; Theorem 6.5]: there should be an increasing sequence of integers n_k such that a point $2\pi j/n_k$ either lies in P or its distance from P is at least $2\pi/n_k$.

If P is so constructed, let I be the ideal in $A(T)$ of all functions vanishing on P. Since P is a set of spectral synthesis, I lies in the closure $\bar{B}$ of $B[P]$, $\bar{B}/I$ is a proper subalgebra of $A(T)/I$, and the latter algebra is semi-simple. Also, $A(T)$ has no closed primary ideals. We conclude:

$A(T)/I$ is a semi-simple Banach algebra, without closed primary ideals, which is not spanned by its set of idempotents, although its maximal ideal space, P, is totally disconnected.

PART II

THEOREM 2.1. *If A is a semi-simple commutative Banach algebra which is spanned by its set of idempotents, then A has the Stone-Weierstrass property.*

Proof. Let B be the closure of a separating self-adjoint subalgebra B_0 of A. (Note that we do not assume that B is self-adjoint; see 3.3.) Let $\Delta(A)$ and $\Delta(B)$ be the maximal ideal spaces of A and B. Since B is separating, $\Delta(A) \subset \Delta(B)$.

For any $f \in B$, the norm of f is the same whether we consider f as an element of A or of B. Hence the two spectral norms of f (in A and in B) are the same, so that

$$(9) \qquad \sup_{x \in \Delta(B)} |f(x)| = \sup_{x \in \Delta(A)} |f(x)| \, (f \in B) .$$

In other words, the Silov boundary S of B [6; p. 80] lies in $\Delta(A)$.

The equation $|e^f| = e^{Re f}$ shows that (9) holds with the real or imaginary part of f in place of $|f|$. Since B_0 is self-adjoint on $\Delta(A)$, this maximum modulus property shows that B_0 is also self-adjoint as an algebra of functions on $\Delta(B)$. Since B_0 is dense in B, B_0 separates points on $\Delta(B)$, and the Stone-Weierstrass theorem implies that every continuous function on $\Delta(B)$ can be uniformly approximated by elements of B_0. Thus $S = \Delta(B)$, and we conclude: $\Delta(B) = \Delta(A)$.

Since A is spanned by its idempotents, $\Delta(A)$ is totally disconnected. Silov's theorem on idempotents [**10**] thus applies to B and shows that B contains every idempotent of A. Hence $B = A$, and the theorem is proved.

THEOREM 2.2. *Let G be a compact abelian group. Then $A(G)$ has the Stone-Weierstrass property if and only if G is totally disconnected.*

Proof. One half of the theorem is an immediate corollary of Theorem 2.1.

To prove the other half, suppose G is not totally disconnected. Its dual group Γ then contains an infinite cyclic group Λ which can be mapped isomorphically onto Z. Regarding Z as a subgroup of R_a (the additive group of the real numbers, with the discrete topology), the divisibility of R_a [**4**] implies that our isomorphism of Λ onto Z can be extended to a homomorphism of Γ into R_a. The duality theory for compact and discrete abelian groups now shows that G contains a compact subgroup K whose dual group $\hat{K}$ is a subgroup of R_a, and that therefore K is a homomorphic image of the Bohr compactification of R. It follows that K contains a dense one-parameter subgroup J.

We now use Theorem 1.2 to prove that $A(K)$ does not have the $S - W$ property.

Note that a continuous function f on K belongs to $A(K)$ if and only if its restriction to J is of the form

$$f(\varphi(t)) = \sum_{s \in \hat{K}} a(s) e^{ist} \qquad (t \text{ real}), \tag{10}$$

with $\sum | a(s) | < \infty$; here φ is a fixed continuous isomorphism of R onto J. Conversely, every function of the form (10) has a continuous extension to K.

A lemma of Wiener [**11**; p. 80] implies that the functions $f(\varphi)$ of the form (10) are locally the same as the members of $A(T)$. Let $\alpha = [-\pi + \varepsilon, \pi - \varepsilon]$, for some fixed $\varepsilon > 0$. Choose $P \subset \alpha$, as in the proof of Theorem 1.2, so that $B[P]$ is not dense in $A(T)$, and let B_1 be the algebra of all $f \in A(K)$ such that $f(\varphi(t))$ coincides with a function in $B[P]$ on α. Then B_1 is a separating self-adjoint subalgebra of $A(K)$ which is not dense in $A(K)$, and $A(K)$ does not have the $S - W$ property.

Finally, we take all $f \in A(G)$ whose restriction to K lies in B_1, and we obtain a separating self-adjoint subalgebra of $A(G)$ which is not dense in $A(G)$.

This completes the proof.

2.3. Suppose G is a totally disconnected infinite compact abelian group, E is a compact subset of G which is not of spectral synthesis (such sets exist [7]), I is the ideal in $A(G)$ consisting of all f which vanish on E, and I_0 is the closure of the ideal consisting of all f which vanish on a neighborhood of E.

Define $B = A(G)/I_0$. If the idempotents in B spanned a proper closed subalgebra B' of B, the inverse image of B' under the canonical homomorphism of $A(G)$ onto B would be a proper, closed, separating, self-adjoint subalgebra of $A(G)$, in contradiction to Theorem 2.2. The radical of B is I/I_0, which by a theorem of Helson [2], is infinite dimensional. We conclude:

B is a commutative Banach algebra which is spanned by its idempotents and whose radical R is infinite dimensional.

We shall show, furthermore, *that B has no subalgebra C, algebraically isomorphic to B/R, such that B is the direct sum $C + R$.*

Thus the Wedderburn principal theorem does not hold for B.

Feldman has constructed an algebra, spanned by its idempotents, with one dimensional radical R, which is not the direct sum of R and any *closed* subalgebra; however, his algebra is the direct sum of R and a *non-closed* subalgebra [1; Theorem 6.1].

Suppose C is a subalgebra of B, and $B = C + R$. Let h be the natural isomorphism of the semi-simple Banach algebra B/R onto C (i.e., into B). Let e', e'' be the characteristic functions of disjoint compact sets E', E'' whose union is E, and define $B' = e'B$, $C' = e'C$, $R' = e'R$. A result of Bade and Curtis [1; Theorems 3.7, 3.9], combined with the fact that $A(G)$ has no primary ideals, shows that E' can be chosen so that

(α) the restriction of h to $e' \cdot (B/R) = B'/R'$ is continuous,

(β) B' is not semi-simple.

Since h^{-1} is continuous (by the definition of the quotient norm), (α) implies that $C' = h(B'/R')$ is a closed subalgebra of B'. Since

(a) C' contains all idempotents of B',

(b) these idempotents span B' (by the same reasoning that applied to B), and

(c) C' is closed, we conclude that $C' = B'$, so that B' is semi-simple, in contradiction to (β).

The preceding argument yields a more general result:

THEOREM 2.4. *Suppose A is a regular commutative Banach algebra without primary ideals, and suppose E is a compact subset of the* *maximal ideal space of A. If $B = A/I_0(E)$ is spanned by its idempotents and if B is not semi-simple, then the Wedderburn theorem does not hold for B.*

Bade and Curtis, in an as yet unpublished paper "The Wedderburn decomposition of commutative Banach algebras", have constructed other examples of commutative Banach algebras in which the radical cannot be split off algebraically.

PART III

3.1. The standard examples of non-self-adjoint Banach algebras involve analytic functions of one or more complex variables, and their maximal ideal spaces are at least two-dimensional. Before turning to the construction of examples with totally disconnected maximal ideal space, we insert two remarks.

(a) *If A is semi-simple and self-adjoint, then there is a constant M such that $\|\bar{f}\| \leqq M\|f\|$ for every $f \in A$.*

Indeed, considering A as a Banach space over the real field, the map $f \to \bar{f}$ is linear; the closed graph theorem applies (since A is semi-simple) and shows that this map is continuous.

(b) *If a semi-simple Banach algebra A has a dense self-adjoint subalgebra B, and if the map $f \to \bar{f}$ is bounded on B, then A is self-adjoint.*

This is obvious. (It was tacitly used in 2.3, in the assertion that B' is self-adjoint.) We mention (b) mainly because of the two examples which follow. In 3.2 we construct an algebra which is not self-adjoint, although the map $f \to \bar{f}$ is bounded on a separating subalgebra; the algebra constructed in 3.3 contains a dense self-adjoint subalgebra although it is not itself self-adjoint. In both examples, the maximal ideal spaces are totally disconnected.

3.2. Let $\{z_n\}$ be a sequence of complex numbers, with $0 < |z_n| < 1$, such that $z_n \to 0$ as $n \to \infty$, and such that the sequence $\{z_n/|z_n|\}$ is dense on the unit circle.

Let A be the algebra of all sequences $a = \{a_n\}$, $n = 1, 2, 3, \cdots$, with termwise addition and multiplication, for which the limit

$$L(a) = \lim_{n\to\infty} \frac{a_n}{z_n}$$

exists as a finite number; norm A by

$$\|a\| = \sup_n \left|\frac{a_n}{z_n}\right|.$$

Then A is a Banach algebra. Let B be the set of all $a \in A$ for which

$L(a) = 0$. Since $L(ab) = 0$ whenever $a \in A$ and $b \in B$, B is a closed ideal in A. Furthermore, if $e_k \in A$ is the sequence whose K^{th} term is 1, while all others are 0, it is easily verified that B is the closure of the set of all finite linear combinations of the e_k; in other words, B is spanned by the idempotents of A.

Let h be a homomorphism of A onto the complex field. If $h(B) = 0$, then $[h(a)]^2 = h(a^2) = 0$ for every $a \in A$ (since $a^2 \in B$), a contradiction. Thus $h(e_k) \neq 0$ for some k, and $h(e_k)h(a) = h(e_k a) = a_k h(e_k)$, so that $h(a) = a_k$ for all $a \in A$.

It follows that A is semi-simple, and that its maximal ideal space is discrete and countable. By 3.1 (b), B is self-adjoint.

Suppose next that $a \in A$ and a is real. Then

$$L(a) = \lim_{n\to\infty} \frac{a_n}{|z_n|} \cdot \frac{|z_n|}{z_n} .$$

Since $\{|z_n|/z_n\}$ is dense on the unit circle, and since a_n is real, this can exist only when it is 0. Hence $a \in B$.

Thus if $a \in A$ and $\bar{a} \in A$, then $a + \bar{a} \in B$ and $i(a - \bar{a}) \in B$, so that $a \in B$. We summarize:

A is a commutative semi-simple Banach algebra whose maximal ideal space is discrete and countable; A contains a proper, closed, separating ideal B which consists precisely of the self-adjoint elements of A.

The non-self-adjoint algebra A is, in turn, a closed ideal in the self-adjoint algebra A_1 which consists of all sequences a such that $\|a\| = \sup_n |a_n/z_n| < \infty$.

The last assertion follows from the inclusions $A \cdot A_1 \subset B \subset A$.

3.3. Our next example is *a regular semi-simple commutative Banach algebra A which is not self-adjoint, although it is spanned by its idempotents.**

Since the algebra of all finite linear combinations of idempotents is always self-adjoint, we see that A contains a dense self-adjoint subalgebra.

Define $w_n = 1$ if $n \leqq 0$, $w_n = 1 + \log(n + 1)$ if $n \geqq 1$, and let A_0 be the algebra of all functions f of the form

$$f(x) = \sum_{-\infty}^{\infty} a_n e^{inx}$$

for which the norm

$$\|f\|_0 = \sum_{-\infty}^{\infty} |a_n| w_n$$

is finite.

* *Added in proof.* Other algebras with this property were found by Coddington (Proc. Amer. Math. Soc. **8** (1957), 258–261).

The inequality $w_{n+m} \leqq w_n w_m$ shows that A_0 is a Banach algebra and it is easily verified that its maximal ideal space is unit circle T. Since A_0 contains every f with two continuous derivatives, A_0 is regular. i.e., given any two disjoint compact sets C_0 and C_1 on T, there exists $f \in A_0$ such that $f = 0$ on C_0, $f = 1$ on C_1. Wermer has pointed out that algebras like A_0 furnish simple examples of non-self-adjoint algebras with one-dimensional maximal ideal space.

Given a rapidly increasing sequence of positive integers p_k ($k = 1, 2, 3, \cdots$), let

$$L_{j,k} = \left(\frac{2\pi(j-1)}{p_k^2}, \quad \frac{2\pi j}{p_k^2}\right) \qquad (1 \leqq j \leqq p_k^2),$$

let I_k be the union of those $L_{j,k}$ with $j \equiv 1 \pmod{p_k}$, and define P to be the intersection of the complements of the sets I_k.

Our desired algebra A is the restriction of A_0 to P.

We first prove that A is spanned by its idempotents. As in 1.4, let g_k be the characteristic function of I_k, and put $f_k(t) = p_k \int_0^x g_k(t)dt$. Then $e^{if_k} \in A_0$, and $f_k(x) - x = h_k(p_k x)$, where $h_k(0) = h_k(2\pi) = 0$, h_k is linear on $[0, 2\pi/p_k]$ and on $[2\pi/p_k, 2\pi]$, and

$$h_k\left(\frac{2\pi}{p_k}\right) = \frac{2\pi}{p_k} - \frac{2\pi}{p_k^2}.$$

We have

$$\| h_k(p_k x) \|_0 = \sum_{-\infty}^{\infty} | \hat{h}_k(n) | w_{np_k}.$$

The Fourier series of h_k is exhibited in 1.5, with $a = 2\pi/p_k$, $b = 2\pi(p_k - 1)/p_k^2$, and a simple computation now shows that

$$\| h_k(p_k x) \|_0 < C \frac{\log^2 p_k}{p_k}$$

which tends to 0 as $k \to \infty$. As in 1.4, it follows that $\| e^{ix} - e^{if_k(x)} \|_0 \to 0$ as $k \to \infty$.

Since e^{if_k} is constant on each arc contiguous to I_k, the restriction of e^{if_k} to P is a finite linear combination of idempotents of A. It follows that the restriction of e^{ix} to P is in the span of the idempotents; the same is true of e^{-ix}, and hence A is spanned by its idempotents.

It remains to be shown that $\{p_k\}$ can be so chosen that A will not be self-adjoint. We do this inductively.

Let P_k be the complement of $I_1 \cup \cdots \cup I_k$, and let A_k be the restriction of A_0 to P_k. We claim that A_k is not self-adjoint. To prove this, note that A_k contains the restriction to P_k of all $f(x) = \sum_{-\infty}^{0} a_n e^{inx}$ with $\sum_{-\infty}^{0} | a_n | < \infty$. If A_k were self-adjoint, it would follow that A_k consists of all restrictions of functions in $A(T)$ to P_k, and since P_k

contains an arc, this would imply that $A(T)$ and A_0 coincide locally. Being regular, these algebras would therefore have to be the same, a contradiction.

Let $c(n, k)$ be the A_k-norm of the restriction of e^{inx} to P_k, i.e.,

$$c(n, k) = \inf \{\|f\|_0 : f(x) = e^{inx} \text{ on } P_k\} \qquad (n = 0, \pm 1, \pm 2, \cdots) . \tag{11}$$

Since $c(n, k) = 1$ for $n \leqq 0$ and A_k is not self-adjoint, 3.1 (b) shows that $c(n, k)$ is unbounded as $n \to +\infty$. In particular, there exists n_k such that

$$c(n_k, k) > k . \tag{12}$$

We now claim that there exists $\delta_k > 0$ with the following property: *If* $0 \leqq n \leqq k$, *if* V *is an open set with* $m(V) < \delta_k$, *if* $g \in A_0$ *and* $g(x) = e^{inx}$ *on* $P_k - V$, *then* $\|g\|_0 > c(n, k) - 2^{-k}$.

Suppose this is false. Then there exists

(i) an integer n, $0 \leqq n \leqq n_k$,

(ii) open sets V_r with $m(V_r) < 2^{-r}$ $(r = 1, 2, 3, \cdots)$,

(iii) $g_r \in A_0$ satisfying $g_r(x) = e^{inx}$ on $P_k - V_r$, such that

(iv) $\|g_r\|_0 \leqq c(n, k) - 2^{-k}$.

By (iv), the diagonal process yields a sequence $\{r_i\}$ such that $\hat{g}_{r_i}(m)$ converges, say to a_m, for $m = 0, \pm 1, \pm 2, \cdots$. Put

$$g(x) = \sum_{-\infty}^{\infty} a_m e^{imx} . \tag{13}$$

Then $g \in A_0$, and

$$\|g\|_0 \leqq c(n, k) - 2^{-k} . \tag{14}$$

For every $f \in L(T)$, $\int g_{r_i} f \to \int gf$ as $i \to \infty$. Combined with (ii) and (iii), this shows that $g(x) = e^{inx}$ a.e. on P_k, and since g is continuous, this equality holds everywhere on P_k. But then (14) contradicts (11), the definition of $c(n, k)$.

Having determined δ_k, we choose P_{k+1} so that $2\pi - m(P_{k+1}) < (1/2)\delta_k$, i.e., so that $m(I_{k+1}) < (1/2)\delta_k$, and we furthermore subject the sequence $\{P_k\}$ to the requirement

$$m(I_{k+1}) + m(I_{k+2}) + \cdots < \delta_k .$$

Then $m(P_k - P) < \delta_k$ for every k, and it follows that the A-norm of the restriction of $e^{in_k x}$ to P is not less than $c(n_k, k) - 1$, i.e., not less than $k - 1$, by (12). Since the restrictions of the trigonometric polynomials to P are dense in A, 3.1 (b) implies that A is not self-adjoint. This completes the proof.

3.4. We conclude with a theorem which shows that under certain conditions the hypothesis of self-adjointness can be dropped from the Stone-Weierstrass theorem (a special case of this appeared in [8]):

THEOREM. *If A is a semi-simple commutative Banach algebra which is spanned by its idempotents and whose maximal ideal space contains no perfect subset, then every separating subalgebra of A is dense in A.*

Proof. Let B be a closed separating subalgebra of A, and denote the maximal ideal spaces of A and B by $\Delta(A)$ and $\Delta(B)$.

Fix $x_0, x_1 \in \Delta(A)$, $x_0 \neq x_1$. There exists $f \in B$ such that $f(x_0) = 0$, $f(x_1) = 1$. Since $\Delta(A)$ has no perfect subsets, $f(\Delta(A)) \cup \{0\}$ is a compact countable subset of the complex plane [8].

Suppose there is a point $y \in \Delta(B)$ such that $f(y) \neq 0$ and $f(y)$ is not in $f(\Delta(A))$. Then there is a polynomial $P(z) = \sum_1^N a_n z^n$ such that

$$|P(f(y))| > \sup \{|f(z)| : z \in f(\Delta(A))\} ,$$

and the function $P(f)$, which belongs to B, does not attain its maximum modulus (relative to $\Delta(B)$) on $\Delta(A)$. But the Silov boundary of B is in $\Delta(A)$, as in the proof of Theorem 2.1, and we have a contradiction.

Thus $f(\Delta(B)) = f(\Delta(A))$.

We can therefore find disjoint open sets V, W in the plane, such that $0 \in V$, $1 \in W$, $f(\Delta(B)) \subset V \cup W$. Define $g = 0$ on V, $g = 1$ on W. The theorem on analytic functions in Banach algebras [**6**; p. 78] shows that $g(f) \in B$. That is to say, B contains the characteristic function of a compact open set $E \subset \Delta(A)$ such that $x_1 \in E$ but $x_0 \notin E$.

Since x_0, x_1 were arbitrary, the sets E so obtained form a basis for the topology of $\Delta(A)$. This implies that B contains the characteristic function of every compact open set in $\Delta(A)$, and so $B = A$.

This proof, unlike our proof of Theorem 2.1, does not use Silov's theorem on idempotents. In fact, the preceding proof establishes Silov's theorem in the special case in which $\Delta(A)$ contains no perfect set.

REFERENCES

1. W. G. Bade and P. C. Curtis, Jr., *Homomorphisms of commutative Banach algebras*, Amer. Math. J., **82** (1960), 589-608.

2. Henry Helson, *On the ideal structure of group algebras*, Ark. Mat. **2** (1952), 83-86.

3. C. S. Herz, *The spectral theory of bounded functions*, Trans. Amer. Math. Soc., **94** (1960), 181-232.

4. Irving Kaplansky, *Infinite Abelian groups*, Ann Arbor, 1954.

5. Yitzhak Katznelson, *Some examples in the theory of absolutely convergent Fourier series*, Abstract 564-148, Notices of Amer. Math. Soc., Dec. 1959.

6. L. H. Loomis, *An Introduction to Abstract Harmonic Analysis,* New York, 1953.
7. Paul Malliavin, *Sur l'impossibilité de la synthèse spectrale sur les groupes abéliens non compacts,* Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques Paris, (1959), 61-68.
8. Walter Rudin, *Continuous functions on compact spaces without perfect subsets,* Proc. Amer. Math. Soc. **8** (1957), 39-42.
9. Laurent Schwartz, *Sur une propriété de synthèse spectrale dans les groupes non compacts,* C. R. Acad. Sci. Paris **227** (1948), 424-426.
10. G. E. Silov, *On decomposition of a commutative normed ring in a direct sum of ideals,* Mat. Sbornik N. S. **32** (74) (1953), 353-364.
11. Norbert Wiener, *The Fourier Integral and Certain of its Applications,* Cambridge University Press, 1933.

UNIVERSITY OF CALIFORNIA
AND
UNIVERSITY OF WISCONSIN

THE SUBGROUPS OF A DIVISIBLE GROUP G WHICH CAN BE REPRESENTED AS INTERSECTIONS OF DIVISIBLE SUBGROUPS OF G

SAMIR A. KHABBAZ

Introduction. In [1], page 70, L. Fuchs asks the following question: Which are those subgroups of a divisible group G that can be represented as intersections of divisible subgroups of G?

The main purpose of this paper is to give an answer to this question.

NOTATION.

N1: If H is a primary p-group, let $S(H)$ denote the subgroup of elements of H whose orders are 1 or p.

N2: If G is Abelian, let $T(G)$ be the torsion subgroup of G; let G_p denote the primary p-component of $T(G)$; and, in case G is divisible, let $F(G)$ denote a maximal torsion free subgroup of G.

N3: Let the symbol $\oplus$ denote a direct sum. Let the symbol $<$ denote "properly contained in." Let $\subset$ denote "contained in." Let $N\backslash M$ denote "the set of elements in N and not in M." Let $\cong$ denote "is isomorphic to." Let $\exists$ denote "there exists (exist)." Let $\ni$ denote "such that." Let $(N_\alpha)_{\alpha\in A}$ denote a family of sets N_α indexed by members of the set A. Finally if Q is a subset of a group, let $\{Q\}$ denote the subgroup of that group generated by the elements of Q.

N4: Let R denote the additive group of rationals. Let P denote the set of primes. Let the group $C(p^\infty)$ be the indecomposable divisible primary p-group.

N5: Let $C = C(2^\infty) \oplus C(3^\infty) \oplus C(5^\infty) \oplus \cdots$; and if $S \subset P$, let $C_S = \bigoplus_{p\in S} C(p^\infty)$.

N6: If G is a group, let $P(G)$ be the set of $p \in P$, such that $\exists x \in G$ with order $x = p$.

N7: Finally, we recall the following convenient and succinct classification of the subgroups of R [see Kurosh I, page 208]. Let $p_1, p_2, p_3, \cdots$ be the sequence of primes in natural order. A characteristic is a sequence $a = (a_1, a_2, a_3, \cdots)$, where $a_i =$ a non-negative integer or ∞. A type is a class of equivalent characteristics, two characteristic $a = (a_1, a_2, a_3, \cdots)$ and $b = (b_1, b_2, b_3, \cdots)$ being equivalent if and only if $\sum_{i=1}^{\infty} |a_i - b_i| < \infty$, where $\infty - \infty = 0$.

$A \subset R$ has type a if and only if it is isomorphic to the subgroup

Received April 15, 1960. This paper constitutes the first part of the author's doctoral dissertation submitted to the University of Kansas, and was presented before the Amer. Math. Soc. An abstract of this paper was received by the Society on July 17, 1959.

of R consisting of those rationals whose denominators in the reduced form are divisible by no higher power of the prime p_i than the a_ith if $a_i < \infty$, and by every power of p_i if $a_i = \infty$.

Define $a \geqq b$ if and only if $a_i \geqq b_i$, for $i = 1, 2, 3, \cdots$.

N8: Let $S \subset P$. We shall say that A above has type T_S is and only if $a_i = 0$ for $p_i \in S$ and $a_i = \infty$ otherwise. Then it is well known that $R/B \subset C_S$ if and only if B contains a subgroup A of type T_S, and that the intersection of two subgroups of R containing subgroups of type T_S again contains a subgroup of type T_S.

N9: Let the symbol $\bigcap_G$ stand for the phrase "an intersection of divisible subgroups of G."

LEMMA 1. (Kulikov):

a. *A divisible group M is a minimal divisible group containing the subgroup L if and only if $H \subset M$ and $H \cap L = 0$ imply $H = 0$, H* being a subgroup.

b. *If M is a minimal divisible group containing L, then M/L is torsion and divisible.*

LEMMA 2. *Let G be divisible and L a subgroup of G. Let M be a minimal divisible subgroup of G containing L. Thus, using the notational in N2, we may write, $G = M \oplus E = M \oplus T(E) \oplus F(E)$. We have:*

a. *If M is minimal divisible containing L, then $S(M_p) = S(L_p)$ for each $p \in P$, and $T(M)$ is minimal divisible containing $T(L)$.*

b. Kulikov: *If L is torsion, then M is minimal divisible containing L if and only if $S(L_p) = S(M_p)$ for all $p \in P$.*

c. *If L is $\bigcap_G$ then $T(L)$ is $\bigcap_{T(G)}$ and hence $\bigcap_G$.*

Proof.

a. Let $x \in S(M_p)\backslash S(L_p)$. Then $\{x\} \cap S(L_p) = 0$, and therefore $\{x\} \cap L = 0$. By Lemma 1a, $x = 0$. Next, $T(M)$ is divisible, and contains $T(L)$. If $N \subset T(M)$ is divisible and contains $T(L)$, $T(M)$ can be written as $T(M) = N \oplus K$, where $K \cap T(L) = 0$; and hence $K \cap L = 0$, so that by Lemma 1a, $K = 0$; and hence, $N = T(M)$.

b. The "only if" part is contained in part a. For the "if" part, assume $N \subset M$ is divisible and contains L. Then we may write $M = N \oplus K$. Then, by hypothesis, K cannot have elements of prime order, and must therefore be 0.

c. Assume $L = \bigcap_{a \in A} M_a$, where each M_a is divisible and contains L. Then each $T(M_a)$ is divisible and contains $T(L)$. Moreover, we have $\bigcap_{a \in A} T(M_a) = T(\bigcap_{a \in A} M_a) = T(L)$. Hence $T(L)$ is $\bigcap_{T(G)}$ and hence $\bigcap_G$.

LEMMA 3. *Let G be a minimal divisible group containing the* *subgroup* L, *and having a representation of the form* $G = \bigoplus_{a \in A} G_a$. *Then* G/L (which by Lemma 1b is divisible and torsion) *contains a subgroup isomorphic to* $C(p^\infty)$ *if and only if for some* $a \in A$, $G_a/G_a \cap L$ *contains a subgroup of the same kind. In other words*: $P(G/L) = \bigcup_{a \in A} P(G_a/G_a \cap L) = \bigcup_{a \in A} P(G_a + L/L)$.

Proof. Because of the divisibility of all the groups concerned, it suffices to check the existence of elements of order p. Suppose $x \in G_a/G_a \cap L$ has order p. Then $G/L \supset G_a + L/L \cong G_a/(G_a \cap L)$. Hence, G/L has an element of order p. Conversely, suppose that for no $a \in A$ does $G_a/(G_a \cap L)$ contain an element of order p. Then no $G_a + L/L$ contains an element of order p. Hence, the subgroup of G/L generated by the $(G_a + L)/L$ contains no elements of order p. But since the G_a's generate G, the $G_a + L/L$'s generate all of G/L.

THEOREM 1. *Let G be a divisible group; let L be a subgroup of G; let M be a minimal divisible subgroup of G containing L. Suppose G has a representation of the form $G = M \oplus E$, then $L = M \cap (\bigcap_{\omega \in \Omega} M_\omega)$, where M_ω is a divisible subgroup of G containing L for each $\omega \in \Omega$, if and only if there exists homomorphisms $\underline{h_\omega\colon M \to E}$ for each $\omega \in \Omega$ such that $\bigcap_{\omega \in \Omega} \ker h_\omega = L$.*

Proof. To prove the "if" part, let I be the identity map of M; and for each $\omega \in \Omega$ let $g_\omega\colon M \to G$ be defined by $g_\omega = I + h_\omega$. Let $M_\omega = g_\omega(M)$. Then $L \subset M_\omega$ since $h_\omega(L) = 0$, and therefore M_ω is divisible since it is a homomorphic image of M. Finally, $x \in M_\omega \cap M$ implies $x \in \ker h_\omega$ (since $x = y + h_\omega(y)$ implies $h_\omega(y) = x - y \in M \cap E = 0$); and hence, $L \subset \bigcap_{\omega \in \Omega} M_\omega \cap M = \bigcap_{\omega \in \Omega} \ker h_\omega = L$.

To prove the "only if" part, suppose $L = M \cap (\bigcap_\omega M_\omega)$. It can be assumed that each M_ω is minimal divisible containing L and, therefore, also minimal divisible containing $M \cap M_\omega$. Also, M is minimal divisible containing $M \cap M_\omega$. Also, M is minimal divisible containing $M \cap M_\omega$; so there is an isomorphism $i_\omega\colon M \to M_\omega$ which is the identity on $M \cap M_\omega$. Note that $i_\omega(x) \in M \Rightarrow i_\omega(i_\omega(x)) = i_\omega(x) \Rightarrow i_\omega(x) = x \Rightarrow x \in M \cap M_\omega$. Let p: $G \to E$ be the projection determined by the decomposition $G = M \oplus E$. Let h_ω be defined by $h_\omega = pi_\omega$. Then $h_\omega(x) = 0$, $i_\omega(x) \in M$, and $x \in M \cap M_\omega$ are equivalent. Thus $\bigcap_{\omega \in \Omega} \ker h_\omega = \bigcap_{\omega \in \Omega} M \cap M_\omega = L$.

REMARK. The underlined portion of Theorem 1 may be replaced by $h_\omega\colon M \to T(E)$.

COROLLARY 1. *Let G be a divisible group; let L be a subgroup of G; and let M be a subgroup of G which is minimal with respect to being divisible and containing L. Thus, we may write $G = M \oplus E =$*

$M \oplus T(E) \oplus F(E)$, and $P(E) = P(T(E))$. Then L is $\bigcap_G$ if and only if $P(M/L) \subset P(E)$.

Proof. The condition $P(M/L) \subset P(E)$ is easily seen to be equivalent to the existence of the family $\{h_\omega\}_{\omega \in \Omega}$ of homomorphisms in Theorem 1.

REMARK. Let G be divisible and torsion free, then $L \subset G$ is $\bigcap_G$ if and only if L is divisible or, equivalently, is a direct summand of G.

COROLLARY 2. *Let G be divisible, and let L be a torsion subgroup of G. Then L is $\bigcap_G$ if and only if for each $p \in P$, $S(L_p) < S(G_p)$ whenever $S(L_p) \neq 0$, and L_p is not divisible.*

Proof. If L is $\bigcap_G$ then by Lemma 2c for each $p \in P$ obviously L_p is $\bigcap_{G_p}$; and, hence, to prove that our condition is necessary, we may assume that G is primary, and L is not divisible, in which case the necessity becomes obvious in view of the fact that otherwise G would be the only minimal divisible subgroup of itself containing L; and, consequently, $L = G$, since L is $\bigcap_G$, contrary to L being not divisible.

To Prove the "only if" part, note that $p \in P(M/L)$ implies $M_p/L_p = (M/L)_p \neq 0$, since L is torsion. Thus, by hypothesis, $p \in P(M/L) \Rightarrow S(L_p) < S(G_p) \Rightarrow S(M_p) < S(G_p) = S(M_p) \oplus S(E_p) \Rightarrow S(E_p) \neq 0 \Rightarrow p \in P(E)$.

COROLLARY 3. *Let G be divisible and $L \subset G$ be torsion, reduced and $\bigcap_G$. Then every subgroup of L is $\bigcap_G$.*

COROLLARY 4. *Let G be divisible and L be $\bigcap_G$. Let M be a minimal divisible subgroup of G containing L. Let K be a subgroup of G such that $L \subset K \subset M$. Then K is $\bigcap_G$.*

Proof. If $G = M \oplus E$, then $P(M/K) \subset P(M/L) \subset P(E)$.

COROLLARY 5. *Let K be any Abelian group of arbitrary cardinal number A. Then K can be embedded in a divisible group G of power $A\aleph_0$ in such a way that any subgroup of K can be represented as an intersection of two divisible subgroups of G.*

Proof. Let M be a minimal divisible group containing K and let E be a group isomorphic to M/K. Let G be the direct sum of M and E. The cardinality of G is clearly $\aleph_0 A$ and the isomorphism of M/K into E induces a homomorphism h: $M \rightarrow E$ with $\ker h = K$. Thus, Theorem 1 gives the required conclusion.

REMARK. Let $L \subset G$, and let $L = L_1 \oplus L_2$, where L_2 is divisible and

L_1 is reduced. Then, also, $G = L_2 \oplus K$, where K may be chosen to contain L_1. It is easy to see that L is $\bigcap_G$ if and only if L_1 is $\bigcap_K$. Thus, in order to avoid excessive wording, we may in the following theorems assume without loss of generality that L is reduced.

THEOREM 2. *Assume $L \subset G$ is reduced, then L is $\bigcap_G$ if and only if $T(L)$ is $\bigcap_G$ and $P(G/L) \subset P(G)$, equality holding if L is $\bigcap_G$.*

Proof. Let $G = M \oplus E$, where M and E are as in Theorem 1. Then $P(G) = P(T(M)) \cup P(E)$, and $P(G/L) = P(M/L) \cup P(E)$, because $G/L \cong (M/L) \oplus E$. Note that if $T(L)$ is $\bigcap_G$, then $T(L)$ is $\bigcap_{T(G)}$ and hence $P(T(M)/T(L)) \subset P(T(E)) = P(E)$, by Corollary 1. But since $T(L)$ is reduced, $P(T(M)/T(L)) = P(T(M))$. Thus the assumption that $T(L)$ is $\bigcap_G$ implies $P(G) = P(E)$ and, therefore, that the conditions $P(G/L) = P(G)$, $P(G/L) \subset P(G)$, and $P(M/L) \subset P(E)$ are equivalent. This observation, together with Lemma 2c and Corollary 1, proves Theorem 2.

COROLLARY 6. *Let G be any divisible group. Let C be as defined in N5, and let $\bar{G} = C \oplus G$. Then any subgroup $K \subset G$ is $\bigcap_{\bar{G}}$.*

REMARK. In Corollary 6, C may be replaced by any Abelian group containing it.

COROLLARY 7. *Any torsion free subgroup T of $\bar{G}$ above is $\bigcap_{\bar{G}}$.*

Proof. T is contained in a direct summand of $\bar{G}$ whose complimentary direct summand contains a subgroup isomorphic to C.

REMARK. The following example shows that if $L \subset G$ is $\bigcap_G$ and if $\bar{L} \subset G$ is isomorphic to L, $\bar{L}$ need not be $\bigcap_G$. Let $G = \bigoplus_{i=1}^{\infty} C_i$ where $C_i \cong C(p^\infty)$ and where p is fixed. Then,

$$S(\bigoplus_{i=1}^{\infty} C_i) \cong S(\bigoplus_{i=2}^{\infty} C_i) ;$$

however, $S(\bigoplus_{i=1}^{\infty} C_i)$ is not $\bigcap_G$, while $S(\bigoplus_{i=2}^{\infty} C_i)$ is $\bigcap_G$.

In this connection we have:

COROLLARY 8. *Let $L \subset G$ be $\bigcap_G$, and let $\bar{L} \subset G$ be isomorphic to L. Then if $T(\bar{L})$ is $\bigcap_G$—this is in particular the case if $\bar{L} \subset L$—$\bar{L}$ is also $\bigcap_G$. Thus, $\bar{L}$ is $\bigcap_G$ if and only if $T(\bar{L})$ is $\bigcap_G$.*

Proof. For the proof we may assume L is reduced. By Theorem 2, it suffices to show that $P(G/\bar{L}) \subset P(G)$. Let M and $\bar{M}$ be minimal divisible subgroup of G containing L and $\bar{L}$, respectively, so that $G =$

$M \oplus N = \bar{M} \oplus \bar{N}$. Then we have $M/L \cong \bar{M}/\bar{L}$ [see Kurosh I, page 168]. Thus, $P(G/\bar{L}) = P(\bar{M}/\bar{L}) \cup P(\bar{N})$
$= P(M/L) \cup P(\bar{N})$
$\subset P(G/L) \cup P(G)$
$\subset P(G)$, since L is $\bigcap_G$.

THEOREM 3. *Let G be a divisible group and let $L \subset G$ be reduced. Then the following statements are equivalent:*

(a) *L is $\bigcap_G$.*

(b) *$T(L)$ is $\bigcap_G$ and for any subgroup $\bar{R}$ of G isomorphic to R, either $L \cap \bar{R}$ is zero or of type $\geqq T_{P(G)}$.*

(c) *$S(L_p) < S(G_p)$ if $p \in P(G)$ and for any subgroup $\bar{R}$ of G isomorphic to R, either $L \cap \bar{R}$ is zero or of type $\geqq T_{P(G)}$.*

(d) *$T(L)$ is $\bigcap_G$ and $L \subset (\bigoplus_{a \in A} R_a) \oplus T(G) \subset G$, where $R_a \cong R$ and each $L \cap R_a$ is either zero or of $\geqq T_{P(G)}$.*

Proof. By Theorem 2, (a) is equivalent to the conditions $T(L)$ is $\bigcap_G$ and $P(G/L) \subset P(G)$. These conditions imply (b) since G/L contains a subgroup isomorphic to $\bar{R}/\bar{R} \cap L$, so that $P(\bar{R}/\bar{R} \cap L) \subset P(G/L) \subset P(G)$ and therefore $\bar{R} \cap L$ is zero or of type $\geqq T_{P(G)}$ by N8. Properties (b) and (c) are equivalent by Lemma 2c and Corollary 2. Also (b) implies (d). Finally, suppose (d) holds. Then, $P(T(G)/(T(G) \cap L)) = P(T(G)/T(L)) \subset P(T(G)) = P(G)$ by Theorem 2. Let $G = (\bigoplus_{b \in B} R_b) \oplus (\bigoplus_{a \in A} R_a) \oplus T(G)$. Then $R_b \cap L$ is 0 for all $b \in B$, and for each $a \in A$, $R_a \cap L$ is 0 or has type $\geqq T_{P(G)}$ by hypothesis. Thus, by Lemma 3, $P(G/L) = \bigcap_{b \in B} P(R_b/(R_b \cap L)) \cup \bigcup_{a \in A} P(R_a/(R_a \cap L)) \cup P(T(G)/(T(G) \cap L)) \subset P(G)$. By Theorem 2, this implies (a).

DEFINITION. Define a subset $(x_a)_{a \in A}$ of elements of an Abelian group H to be independent if and only if

$$n_1 x_{a_1} + n_2 x_{a_2} + \cdots + n_m x_{a_m} = 0$$

implies $n_1 = n_2 = \cdots = n_m = 0$ where each $a_i \in A$ and the n_i's are integers.

COROLLARY 10. Assume $L \subset G$ is reduced, then L is $\bigcap_G$ if and only if $T(L)$ is $\bigcap_G$ and L contains a subgroup H which contains a maximal independent subset of L and which has the form $\bigoplus_{a \in A} S_a$, where each S_a is isomorphic to a subgroup of R having type $T_{P(G)}$.

Proof. Assume L is $\bigcap_G$, then by Lemma 2c $T(L)$ is $\bigcap_G$. Let $M \subset G$ be minimal divisible containing L, and assume that $M = T(M) \oplus F(M) = T(M) \oplus (\bigoplus_{a \in A} R_a)$, where $R_a \cong R$ for all $a \in A$. Then, by Theorem 3c and Lemma 1a, each $L \cap R_a$ contains a subgroup S_a of type $T_{P(G)}$. Then it is easy to see that $\bigoplus_{a\in A}S_a$ exists; and that, if $x_a \in S_a$, then $(x_a)_{a\in A}$ is a maximal independent subset of M and therefore, of L.

Next assume the condition holds and let M be as usual. For each S_a, let $R_a \supset S_a$ be a subgroup of M of type R. Since any two non-zero subgroups of R have a non-zero intersection, and since $\bigoplus_{a\in A}S_a$ exists, also $\bigoplus_{a\in A}R_a$ exists. Let $x_a \in S_a$; then $(x_a)_{a\in A}$ is a maximal independent subset of H; and since H contains a maximal independent subset of L, $(x_a)_{a\in A}$ is also a maximal independent subset of L. Thus, we must have $M = T(M) \oplus (\bigoplus_{a\in A}R_a)$. Theorem 3d and the fact that $T(L)$ is $\bigcap_G$ imply that L is $\bigcap_G$.

REMARK. Concerning the last definition given above, it is well known that H contains a maximal independent subset and that if $(x_a)_{a\in A}$ is independent, then $H/\{(x_a)_{a\in A}\}$ is torsion if and only if $(x_a)_{a\in A}$ is also maximal independent. Thus, Corollary 10 may be worded as follows: Assume L is reduced; then L is $\bigcap_G$ if and only if $T(L)$ is $\bigcap_G$ and L contains a subgroup H which has the form $\bigoplus_{a\in A}S_a$ where S_a is isomorphic to a subgroup of R of type $T_{P(G)}$ and such that L/H is torsion.

REMARK. The author wishes to thank the referee, R. S. Pierce, for the present arrangement of the material in this paper, as well as for many changes in the proofs. The author also owes thanks to W. R. Scott.

BIBLIOGRAPHY

1. L. Fuchs, *Abelian Groups*, Publishing House of the Hungarian Academy of Sciences, Budapest, 1958.

2. Kaplansky, Irving, *Infinite Abelian Groups*, University of Michigan Press, Ann Arbor, 1954.

3. A. G. Kurosh, *The Theory of Groups*, Chelsea Publishing Company, New York, N. Y., 1955.

CONSTRUCTION OF A CLASS OF MODULAR FUNCTIONS AND FORMS

MARVIN ISADORE KNOPP

1. Introduction. Let $G(j)$ be the principal congruence subgroup, of level j, of the modular group. In this paper we construct functions which are invariant under $G(j)$, for each integer $j \geqq 2$.

We begin by defining certain functions $\lambda_\nu(j;\tau)$ which, although not in general invariant under $G(j)$, do possess the transformation properties

$$(1.01)\quad \lambda_\nu(j;T\tau) = \lambda_\nu(j;\tau) + \text{constant}, \text{ for all } T \text{ in } G(j).$$

This is the content of the main theorem, Theorem (4.02). Once this result has been established it is a simple matter to construct invariants for $G(j)$ by forming certain linear combinations of the $\lambda_\nu(j;\tau)$. This is done in § 5.

These functions $\lambda_\nu(j;\tau)$ are defined as Fourier series which generalize the Fourier series expansion of $\lambda(\tau)$, given by Simons [6]. To derive the transformation equations (1.01), we proceed directly from the Fourier series, extending a method introduced by Rademacher [4], and since generalized by Lehner [2] and the author [1]. Although in [4] only the invariant $J(\tau)$ for the modular group is treated, the method of [4] has much wider applicability. Thus, in [2] it is used in the case of the modular group to overcome the usual convergence difficulties encountered in constructing forms of dimension -2 by means of Poincaré series, while in [1] it is used to construct forms of nonnegative even integral dimension (in which case we, of course, do not have the method of the Poincaré series) for the modular group and several other closely related groups.

We will indicate in section 6 how the method of this paper can be used to construct automorphic forms of all positive even integral dimensions for the groups $G(j)$. In a future publication these same methods will be applied to construct automorphic functions and forms for certain other congruence subgroups of the modular group and for congruence subgroups of several other groups.

I would like to thank the referee of this paper for his helpful remarks.

2. Several lemmas. In [4] the principal analytic tool is a rather delicate lemma in which the terms of a certain conditionally convergent double series are rearranged. Several variations of this lemma can be

Received January 20, 1960 The author is a National Science Fellow.

found in [1] and [2]. In this section we derive two generalizations of the lemma that will be needed in § 4.

LEMMA (2.01). *Let $a < 0, b < 0, d > c > 0$. Let $y > 0, r \geqq 0$, and ν and j be positive integers. Let $t = (c - 1/2b)d^{-1}$. Then*

$$\sum_{k=1}^{\infty}{}^{\#} \lim_{N\to\infty} \sum_{|m|\leqq N}{}^{*} \frac{e^{-2\pi i\nu m'/k}}{k^{1+r}(kiy - m)} \tag{2.02}$$

$$= \lim_{K\to\infty}\left\{ \sum_{k=1}^{[K(dt-c)]}{}^{\#} \sum_{|m-bk/d|\leqq K/d}{}^{*} \frac{e^{-2\pi i\nu m'/k}}{k^{1+r}(kiy - m)} + \sum_{k=[K(dt-c)]+1}^{[K(dt+c)]}{}^{\#} \sum_{\frac{ak-Kt}{c}\leqq m\leqq\frac{bk+K}{d}}{}^{*} \frac{e^{-2\pi i\nu m'/k}}{k^{1+r}(kiy - m)}\right\}.$$

where the asterisk (*) indicates that the inner sum is taken on those m such that $(m, k) = 1$ and $m \equiv 1 \pmod j$, the sharp (♯) indicates that the outer sum is taken on those k such that $k \equiv j \pmod{j^2}$, and m' is is defined by $mm' \equiv -1 \pmod k$.

LEMMA (2.03). *Let y, r, ν, and j be as above. Let ρ be any positive number. Then*

$$\sum_{k=1}^{\infty}{}^{\#} \lim_{N\to\infty} \sum_{|m|\leqq N}{}^{*} \frac{e^{-2\pi i\nu m'/k}}{k^{1+r}(kiy - m)} = \lim_{K\to\infty} \sum_{k=1}^{[\rho K]}{}^{\#} \sum_{|m|\leqq K}{}^{*} \frac{e^{-2\pi i\nu m'/k}}{k^{1+r}(kiy - m)}. \tag{2.04}$$

REMARK. With care, (2.03) could have been included as a special case of (2.01). However, it is simpler and somewhat more germane to our purpose to state them as separate lemmas. It should be noted that Lemma (2.03) is the same as a lemma in [1], except for the congruence conditions on m and k.

A geometric interpretation may be helpful. By a "lattice point" we will mean a pair of relatively prime integers k, m such that $k \equiv j \pmod{j^2}$ and $m \equiv 1 \pmod j$. Rademacher's lemma [4] shows that the sum can be taken by first summing over the lattice points of the half square in the $k - m$ plane defined by $1 \leqq k \leqq K, |m| \leqq K$, and then letting $K \to \infty$. Lemma (2.03) allows us to first sum over the lattice points of the rectangle $1 \leqq k \leqq [\rho K]$, $|m| \leqq K$, while Lemma (2.01) shows that the sum can be taken first over the lattice points of the trapezoid bounded by the lines $k = 0, m = (ak - Kt)/c, m = (bk - K)/d, m = (bk + K)/d$.

The lemma can actually be proved for other trapezoids, but the form in which we have stated it will suffice for our application.

Proof of (2.01). We prove the lemma in the case $r = 0$, the proof for $r > 0$ being virtually the same. We first show the convergence of the left hand side of (2.02).

$$\sum_{|m|\leqq N}^{*} \frac{e^{-2\pi i m'\nu/k}}{k(kiy - m)} = k^{-1} \sum_{0\leqq h<k}^{*} e^{-2\pi i h'\nu/k} \sum_{\substack{n \\ |nk+h|\leqq N}} \frac{1}{kiy - h - nk},$$

where we have put $m = h + nk$. Therefore,

$$\lim_{N\to\infty} \sum_{|m|\leqq N}^{*} \frac{e^{-2\pi i m'\nu/k}}{k(kiy - m)} = k^{-2} \sum_{0\leqq h<k}^{*} e^{-2\pi i h'\nu/k} \lim_{N\to\infty} \sum_{\substack{n \\ |nk+h|\leqq N}} \frac{1}{iy - h/k - n}$$

$$= k^{-2} \sum_{0\leqq h<k}^{*} e^{-2\pi i h'\nu/k} \cdot 2\pi i(1/2 - \{1 - e^{2\pi i(iy - h/k)}\}^{-1})$$

$$= \pi i k^{-2} \sum_{0\leqq h<k}^{*} e^{-2\pi i h'\nu/k} - 2\pi i k^{-2} \sum_{p=0}^{\infty} e^{-2\pi y p} \sum_{0\leqq h<k}^{*} \exp\left[-\frac{2\pi i}{k}(\nu h' + ph)\right].$$

Now, the inner sum of the second term is a Kloosterman sum, for which we have the estimate (see [5])

$$\sum_{0\leqq h<k}^{*} \exp\left[-\frac{2\pi i}{k}(\nu h' + ph)\right] = O(k^{2/3+\varepsilon}). \tag{2.05}$$

Also, the sum in the first term can be written

$$\sum_{0\leqq h<k}^{*} \exp\left[-\frac{2\pi i}{k}(\nu h' + kh)\right] = O(k^{2/3+\varepsilon}).$$

We conclude that

$$\lim_{N\to\infty} \sum_{|m|\leqq N}^{*} \frac{e^{-2\pi i m'\nu/k}}{k(kiy - m)} = O(k^{-4/3+\varepsilon}\{1 - e^{-2\pi y}\}^{-1}),$$

and the left hand side of (2.01) converges.

Let Z denote the set of integers. Let $z_1(K) = [K(dt - c)]$ and $z_2(K) = [K(dt + c)]$. We let $\mathscr{A}(K, N) = \{m \in Z \mid -N \leqq m < (bk - K)/d$ or $(bk + K)/d < m \leqq N\}$ and $\mathscr{B}(K, N) = \{m \in Z \mid (bk + K)/d < m \leqq N$ or $-N \leqq m < (ak - Kt)/c\}$.

We can now state the lemma in the following form

$$\lim_{K\to\infty} \left\{ \sum_{k=1}^{z_1(K)}{}^{\#} \lim_{N\to\infty} \sum_{m\in\mathscr{A}(K,N)}^{*} \frac{e^{-2\pi i m'\nu/k}}{k(kiy - m)} + \sum_{k=z_1(K)+1}^{z_2(K)}{}^{\#} \lim_{N\to\infty} \sum_{m\in\mathscr{B}(K,N)}^{*} \frac{e^{-2\pi i m'\nu/k}}{k(kiy - m)} \right\} = 0. \tag{2.06}$$

The function defined by

$$g(m) = \begin{cases} e^{-2\pi i m'\nu/k}, & \text{if } (m, k) = 1 \text{ and } m \equiv 1 \pmod{j} \\ 0, & \text{otherwise} \end{cases}$$

is periodic modulo k. This is easily seen if we recall that $k \equiv j \pmod{j^2}$

and therefore $j \mid k$. It follows that

$$g(m) = \sum_{l=1}^{k} B_l e^{2\pi i l m/k} ,$$

where

$$B_l = k^{-1} \sum_{0 \leqq m < k}^{*} \exp\left[-\frac{2\pi i}{k}(\nu m' + lm)\right] .$$

Using (2.05) we see that

$$B_l = O(k^{-1/3+\varepsilon}) . \tag{2.07}$$

In the first double sum of (2.06) put

$$\begin{aligned} T_k(K, N) &= \sum_{m \in \mathscr{A}(K,N)}^{*} \frac{e^{-2\pi i \nu m'/k}}{k(kiy - m)} = \sum_{m \in \mathscr{A}(K,N)} \sum_{l=1}^{k} B_l \frac{e^{2\pi i l m/k}}{k(kiy - m)} \\ &= \sum_{l=1}^{k-1} B_l \sum_{m \in \mathscr{A}(K,N)} \frac{e^{2\pi i l m/k}}{k(kiy - m)} + B_k \sum_{m \in \mathscr{A}(K,N)} \frac{1}{k(kiy - m)} . \end{aligned} \tag{2.08}$$

Let $T_k(K) = \lim_{N\to\infty} T_k(K, N)$, $z_3(K) = [(K + bk)/d]$, and $z_4(K) = [(K - bk)/d]$. Recalling the definition of $\mathscr{A}(K, N)$ and making use of (2.08), we may write

$$\begin{aligned} T_k(K) &= k^{-1} \sum_{l=1}^{k-1} B_l \sum_{m=z_3(K)+1}^{\infty} \frac{e^{2\pi i l m/k}}{kiy - m} \\ &\quad + k^{-1} \sum_{l=1}^{k-1} B_l \sum_{m=z_4(K)+1}^{\infty} \frac{e^{-2\pi i l m/k}}{kiy + m} \\ &\quad + B_k k^{-1} \sum_{m=z_4(K)+1}^{\infty} \left(\frac{1}{kiy - m} + \frac{1}{kiy + m}\right) \\ &\quad + B_k k^{-1} \sum_{m=z_3(K)+1}^{z_4(K)} \frac{1}{kiy - m} \\ &= S_1 + S_2 + S_3 + S_4 . \end{aligned} \tag{2.09}$$

To handle S_1, put

$$E_m = \sum_{p=z_3(K)+1}^{m} e^{2\pi i l p/k} = \frac{e^{\pi i l(2m+1)/k} - e^{\pi i l(2z_3(K)+1)/k}}{e^{\pi i l/k} - e^{-\pi i l/k}} .$$

Therefore,

$$| E_m | \leqq (\sin \pi l/k)^{-1} \leqq (\min \{2l/k, 2(k - l)/k\})^{-1} \leqq \frac{k}{2}(1/l + 1/(k - l)) .$$

Now,

$$\begin{aligned} \sum_{m=z_3(K)+1}^{\infty} \frac{e^{2\pi i l m/k}}{kiy - m} &= \sum_{m=z_3(K)+1}^{\infty} \frac{E_m - E_{m-1}}{kiy - m} \\ &= \sum_{m=z_3(K)+1}^{\infty} E_m\left(\frac{1}{kiy - m} - \frac{1}{kiy - m - 1}\right) . \end{aligned}$$

Hence,

$$\begin{aligned}
&\left|\sum_{m=z_3(K)+1}^{\infty}\frac{e^{2\pi i l m/k}}{kiy-m}\right| \\
&\quad\leqq \frac{k}{2}(1/l+1/(k-l))\sum_{m=z_3(K)+1}^{\infty}\{k^2y^2+m^2\}^{-1/2}\{k^2y^2+(m+1)^2\}^{-1/2} \\
&\quad< \frac{k}{2}(1/l+1/(k-l))\int_{z_3(K)}^{\infty}\frac{dx}{x^2}=\frac{k}{2}(1/l+1/(k-l))[(K+bk)/d]^{-1}\,.
\end{aligned}$$

Now, we are here considering only those k in the range $1 \leqq k \leqq z_1(K) = [K(dt-c)]$. Since $b<0$, $d>0$, $(K+bk)/d \geqq \{K+Kb(dt-c)\}/d = K/2d$. Making use of (2.07), we conclude that

$$\begin{aligned}
S_1 &= k^{-1}\sum_{l=1}^{k-1}B_l\sum_{m=z_3(K)+1}^{\infty}\frac{e^{2\pi i l m/k}}{kiy-m} \\
&= O\Big(k^{-1}\sum_{l=1}^{k-1}k^{-1/3+\varepsilon}\frac{k}{2}\{1/l+1/(k-l)\}K^{-1}\Big)\,.
\end{aligned}$$

Therefore,

$$S_1 = O(k^{-1/3+\varepsilon}K^{-1}\log k)\,. \tag{2.10}$$

We can estimate S_2 in exactly the same way simply by noticing that $(K-bk)/d \geqq K/d$. We obtain

$$S_2 = O(k^{-1/3+\varepsilon}K^{-1}\log k)\,. \tag{2.11}$$

The estimation of S_3 is simpler. We notice that

$$S_3 = B_k\cdot k^{-1}\sum_{m=z_4(K)+1}^{\infty}\frac{2iyk}{-y^2k^2-m^2}$$

and hence

$$|S_3| < |B_k|\sum_{m=z_4(K)+1}^{\infty}\frac{2y}{m^2} < |B_k|\int_{z_4(K)}^{\infty}\frac{2y\,dx}{x^2}\,.$$

Therefore,

$$S_3 = O(k^{-1/3+\varepsilon}[(K-bk)/d]^{-1}) = O(k^{-1/3+\varepsilon}K^{-1})\,. \tag{2.12}$$

We consider S_4. Recalling that $z_3(K)+1 > (K+bk)/d \geqq K/2d$, we find that

$$\begin{aligned}
&\left|\sum_{m=z_3(K)+1}^{z_4(K)}\frac{1}{kiy-m}\right| \leqq \sum_{m=z_3(K)+1}^{z_4(K)}(k^2y^2+m^2)^{-1/2} \leqq \sum_{m=z_3(K)+1}^{z_4(K)}(k^2y^2+K^2/4d^2)^{-1/2} \\
&\qquad\leqq 2dK^{-1}\{(K-bk)/d-(K+bk)/d\} = -4bkK^{-1}\,.
\end{aligned}$$

Therefore, using (2.07),

$$S_4 = B_k \cdot k^{-1} \sum_{m=z_3(K)+1}^{z_4(K)} \frac{1}{kiy - m} = O(k^{-1/3+\varepsilon}K^{-1}) . \tag{2.13}$$

Collecting our results (2.10), (2.11), and (2.12), we have $T_k(K) = O(k^{-1/3+\varepsilon}K^{-1} \log k)$. Hence,

$$\begin{aligned} \sum_{k=1}^{z_1(K)}{}^{\#} \lim_{N\to\infty} \sum_{m\in\mathscr{A}(K,N)}{}^{*} \frac{e^{-2\pi i\nu m'/k}}{k(kiy - m)} &= \sum_{k=1}^{z_1(K)}{}^{\#} T_k(K) \\ &= O\Big(K^{-1}\sum_{k=1}^{z_1(K)} k^{-1/3+\varepsilon} \log k\Big) \\ &= O(K^{-1/3+\varepsilon} \log K) \end{aligned} \tag{2.14}$$

In the second double sum of (2.06) put

$$\begin{aligned} U_k(K, N) &= \sum_{m\in\mathscr{B}(K,N)}{}^{*} \frac{e^{-2\pi i\nu m'/k}}{k(kiy - m)} \\ &= \sum_{l=1}^{k-1} B_l \sum_{m\in\mathscr{B}(K,N)} \frac{e^{-2\pi i\nu m'/k}}{k(kiy - m)} + B_k \sum_{m\in\mathscr{B}(K,N)} \frac{1}{k(kiy - m)} . \end{aligned} \tag{2.15}$$

Let $U_k(K) = \lim_{N\to\infty} U_k(K, N)$ and $z_5(K) = [(Kt - ak)/c]$. Then using (2.15) and the definition of $\mathscr{B}(K, N)$ we find

$$\begin{aligned} U_k(K) &= k^{-1}\sum_{l=1}^{k-1} B_l \sum_{m=z_5(K)+1}^{\infty} \frac{e^{2\pi i l m/k}}{kiy - m} + k^{-1}\sum_{l=1}^{k-1} B_l \sum_{m=z_5(K)+1}^{\infty} \frac{e^{-2\pi i l m/k}}{kiy + m} \\ &\quad + B_k \cdot k^{-1} \sum_{m=z_5(K)+1}^{\infty} \Big(\frac{1}{kiy - m} + \frac{1}{kiy + m}\Big) \\ &\quad + k^{-1} \sum_{l=1}^{k-1} B_l \sum_{m=z_3(K)+1}^{z_5(K)} \frac{e^{-2\pi i l m/k}}{kiy - m} + B_k \cdot k^{-1} \sum_{m=z_3(K)+1}^{z_5(K)} \frac{1}{kiy - m} \\ &= S_5 + S_6 + S_7 + S_8 + S_9 . \end{aligned} \tag{2.16}$$

Since $(Kt - ak)/c > Kt/c$, we can estimate S_5 and S_6 in the same way as S_1, and S_7 in the same way as S_3 We obtain

$$S_5 + S_6 + S_7 = O(k^{-1/3+\varepsilon}K^{-1} \log k) . \tag{2.17}$$

To handle S_8 define E_m as before. Then

$$\begin{aligned} &\sum_{m=z_3(K)+1}^{z_5(K)} \frac{e^{2\pi i l m/k}}{kiy - m} \\ &= \sum_{m=z_3(K)+1}^{z_5(K)} E_m\Big(\frac{1}{kiy - m} - \frac{1}{kiy - m - 1}\Big) + E_{z_5(K)}/(kiy - z_5(K) - 1) . \end{aligned}$$

Recalling that $|E_m| \leqq (k/2)\{1/l + 1/(k - l)\}$, we have

$$\begin{aligned}
&\left|\sum_{m=z_3(K)+1}^{z_5(K)} \frac{e^{2\pi i l m/k}}{kiy - m}\right| \\
&\leqq \frac{k}{2}\{1/l + 1/(k-l)\}\Big(\sum_{m=z_3(K)+1}^{z_5(K)} \{k^2y^2 + m^2\}^{-1/2}\{k^2y^2 + (m+1)^2\}^{-1/2} \\
&\qquad\qquad + \{k^2y^2 + (Kt - ak)^2/c^2\}\Big) \\
&< \frac{k}{2}\{1/l + 1/(k-l)\}\Big(\sum_{m=z_3(K)+1}^{z_5(K)} (k^{-2}y^{-2}) + c/tK\Big) \\
&\leqq \frac{k}{2}\{1/l + 1/(k-l)\}\{y^{-2}k^{-2}((Kt - ak)/c - (K + bk)/d)) + c/tK\} \\
&\leqq \frac{k}{2}\{1/l + 1/(k-l)\}(Ky^{-2}k^{-2}c^{-1}d^{-1}\{(dt-c) + (-ad-bc)(dt+c)\} + c/tK)\ ,
\end{aligned}$$

since $-ad - bc > 0$ and k is in the range $K(dt - c) < z_1(K) + 1 \leqq k \leqq z_2(K) \leqq K(dt + c)$. Therefore,

$$\begin{aligned}
S_8 &= k^{-1}\sum_{l=1}^{k-1} B_l \sum_{m=z_3(K)+1}^{z_5(K)} \frac{e^{2\pi i l m/k}}{kiy - m} \\
&= O\Big(k^{-1}\sum_{l=1}^{k-1} k^{-1/3+\varepsilon}\cdot\frac{k}{2}\{1/l + 1/(k-l)\}\{Kk^{-2} + K^{-1}\}\Big) \qquad (2.18)\\
&= O(k^{-1/3+\varepsilon}\log k\{Kk^{-2} + K^{-1}\})\ .
\end{aligned}$$

Finally, we estimate S_9.

$$\begin{aligned}
\left|\sum_{m=z_3(K)+1}^{z_5(K)} \frac{1}{kiy - m}\right| &\leqq \sum_{m=z_3(K)+1}^{z_5(K)} (k^2y^2 + m^2)^{-1/2} \leqq y^{-1}k^{-1}\{(Kt - ak)/c - (K - bk)/d\} \\
&\leqq K(cdyk)^{-1}\{(dt - c) - (ab + bc)(dt + c)\}\ .
\end{aligned}$$

Therefore,

$$S_9 = B_k\cdot k^{-1}\sum_{m=z_3(K)+1}^{z_5(K)} \frac{1}{kiy - m} = O(k^{-7/3+\varepsilon}K)\ . \qquad (2.19)$$

Using (2.17), (2.18), and (2.19), we find that

$$U_k(K) = O(k^{-1/3+\varepsilon}\log k\{K^{-1} + K\cdot k^{-2}\})\ .$$

Hence,

$$\begin{aligned}
&\sum_{k=z_1(K)+1}^{z_2(K)}{}^{\#} \lim_{N\to\infty} \sum_{m\in\mathscr{B}(K,N)}^{*} \frac{e^{-2\pi i\nu m'/k}}{k(kiy - m)} \\
&= \sum_{k=z_1(K)+1}^{z_2(K)} U_k(K) = O\Big(\sum_{k=z_1(K)+1}^{z_2(K)} k^{-1/3+\varepsilon}\log k\{K^{-1} + K\cdot k^{-2}\}\Big) \qquad (2.20)\\
&= O\Big(K^{-1}\log K\sum_{k=z_1(K)+1}^{z_2(K)} k^{-1/3+\varepsilon}\Big) = O(K^{-1}\log K\cdot K^{-1/3+\varepsilon}\cdot 2cK) \\
&= O(K^{-1/3+\varepsilon}\log K)\ .
\end{aligned}$$

Now (2.06) follows from (2.14) and (2.20) and the lemma is proved.

Proof of (2.03). We outline the proof for the case $r = 0$. The left hand side of (2.04) is the same as the left hand side of (2.02) and its convergence has already been demonstrated.

The lemma may be stated as follows

$$\lim_{K\to\infty} \sum_{k=1}^{[\rho K]}{}^{\#} \lim_{N\to\infty} \sum_{K<|m|\leqq N}{}^{*} \frac{e^{-2\pi i\nu m'/k}}{k(kiy - m)} = 0 \,. \tag{2.21}$$

Let

$$\begin{aligned} V_k(K, N) &= \sum_{K<|m|\leqq N}{}^{*} \frac{e^{-2\pi i\nu m'/k}}{k(kiy - m)} \\ &= \sum_{l=1}^{k-1} B_l \sum_{K<|m|\leqq N}{}^{*} \frac{e^{2\pi i l m/k}}{k(kiy - m)} + B_k \sum_{K<|m|\leqq N} \frac{1}{k(kiy - m)} \,. \end{aligned}$$

Then,

$$\begin{aligned} V_k(K) = \lim_{N\to\infty} V_k(K, N) &= k^{-1}\sum_{l=1}^{k-1} B_l \sum_{m=K+1}^{\infty} \frac{e^{2\pi i l m/k}}{kiy - m} \\ &+ k^{-1}\sum_{l=1}^{k-1} B_l \sum_{m=K+1}^{\infty} \frac{e^{-2\pi i l m/k}}{kiy + m} + B_k \cdot k^{-1} \sum_{m=K+1}^{\infty}\left(\frac{1}{kiy - m} + \frac{1}{kiy + m}\right) \\ &= S_1' + S_2' + S_3' \,. \end{aligned}$$

Now S_1', and S_2' can be estimated in the same way as S_1 and S_3' in the same way as S_3. Once we have these estimates the proof of (2.21) proceeds exactly as the proof of (2.14) of the previous lemma.

3. The functions $\lambda_\nu(j; \tau)$. Let j be an integer $\geqq 2$ and let ν be a positive integer. We define the function

$$\begin{aligned} \lambda_\nu(j; \tau) &= \sum_{n=1}^{\infty} a_n(\nu, j) e^{2\pi i n\tau/j} \\ a_n(\nu, j) &= (\pi/8) \sum_{k=1}^{\infty}{}^{\#} k^{-1} A_{k,\nu}(n) \cdot (\nu/n)^{1/2} I_1(4\pi(n\nu)^{1/2}/k) \,, \end{aligned} \tag{3.01}$$

where

$$A_{k,\nu}(n) = \sum_{0\leqq h<k}{}^{*} \exp\left[\frac{-2\pi i}{k}(\nu h' + nh)\right] ,$$

a Kloosterman sum, and I_1 is the modified Bessel function of the first kind. Recall that the sharp ($^{\#}$) means that we allow only those k such that $k \equiv j \pmod{j^2}$ and the asterisk ($*$) indicates that we allow only those h such that $h \equiv 1 \pmod{j}$ and $(h, k) = 1$.

We need the following

LEMMA (3.02). (a) *If $a_n(\nu, j)$ is defined as in* (3.01) *then*

$$a_n(\nu, j) \sim \{\nu^{1/4} n^{-3/4} (2j)^{-1/2}/16\} e^{-2\pi i(n-\nu)/j} \exp\left(4\pi(n\nu)^{1/2}/j\right) .$$

(b) If $|z| < 1$, *then*

$$\sum_{n=1}^{\infty} z^n \sum_{p=0}^{\infty} (4\pi^2 n\nu k^{-2})^p/p!\,(p+1)!$$

is absolutely convergent.

Proof. (a) The first term that occurs in the sum defining $a_n(\nu, j)$ is for $k = j$. This term is equal to

$$(\pi/8)j^{-1}A_{j,\nu}(n)(\nu/n)^{1/2}I_1(4\pi(n\nu)^{1/2}/j)\ .$$

But

$$A_{j,\nu}(n) = \exp\left[-2\pi i\{n + (j-1)\nu\}/j\right] = e^{-2\pi i(n-\nu)/j}\ .$$

Therefore the first term is

$$(\pi/8)j^{-1}e^{-2\pi i(n-\nu)/j}\cdot(\nu/n)^{1/2}I_1(4\pi(n\nu)^{1/2}/j)\ .$$

It follows that

$$\begin{aligned}
&|\, a_n(\nu, j) - (\pi/8j)e^{-2\pi i(n-\nu/j}(\nu/n)^{1/2}I_1(4\pi(n\nu)^{1/2}/j)\,| \\
&\quad = |\,(\pi/8)\sum_{k=2j}^{\infty}{}^{\#} k^{-1}A_{k,\nu}(n)(\nu/n)^{1/2}I_1(4\pi(n\nu)^{1/2}/k)\,| \\
&\quad \leqq C_1(\nu/n)^{1/2}\sum_{k=2j}^{\infty}{}^{\#} k^{-1}k^{2/3+\varepsilon}I_1(4\pi(n\nu)^{1/2}/k)\ ,
\end{aligned}$$

where we have made use of (2.05)

It is a simple consequence of the power series definition of I_1

$$I_1(\eta) = \sum_{p=0}^{\infty}(\eta/2)^{2p+1}/p!\,(p+1)! \tag{3.03}$$

that $I_1(\eta) \leqq \sinh\eta$. We also need the fact that $\sinh\eta \leqq (\eta/B)\sinh B$, for $0 \leqq \eta \leqq B$. We find that

$$\begin{aligned}
&|\, a_n(\nu, j) - (\pi/8j)e^{-2\pi i(n-\nu)/j}(\nu/n)^{1/2}I_1(4\pi(n\nu)^{1/2}/j)\,| \\
&\quad \leqq C_1(\nu/n)^{1/2}\sum_{k=2j}^{\infty}{}^{\#} k^{-1/3+\varepsilon}\{(4\pi(n\nu)^{1/2}/k)/(4\pi(n\nu)^{1/2}/2j)\}\sinh(4\pi(n\nu)^{1/2}/2j) \\
&\quad \leqq C_2\exp(2\pi(n\nu)^{1/2}/j)\cdot n^{-1/2}\ .
\end{aligned}$$

Now in ([7], p. 203, formula 2), it is shown that $I_1(\eta) \sim e^{\eta}/(2\pi\eta)^{1/2}$. Therefore,

$$I_1(4\pi(n\nu)^{1/2}/j) \sim \exp(4\pi(n\nu)^{1/2}/j)/2\pi(2j^{-1})^{1/2}(n\nu)^{1/4}\ .$$

and the result follows.

(b) $$\sum_{p=0}^{\infty}(4\pi^2 n\nu/k^2)^p/p!\,(p+1)! = \{k/2\pi(n\nu)^{1/2}\}I_1(4\pi(n\nu)^{1/2}/k)$$

$$\leqq \{k/2\pi(n\nu)^{1/2}\} \sinh (4\pi(n\nu)^{1/2}/k)$$
$$< \{k/2\pi(n\nu)^{1/2}\} \exp (4\pi(n\nu)^{1/2}/k) \,.$$

The result follows.

Lemma (302a) shows that the series defining $\lambda_\nu(j;\tau)$ converges absolutely for $\mathscr{I}(\tau) > 0$. Therefore, $\lambda_\nu(j;\tau)$ is analytic in the upper half τ-plane.

In order to derive the transformation properties of $\lambda_\nu(j;\tau)$ we transform (3.01) into a certain double series. The computations involved are a repetition of those found in [4, pp. 244–5] and in [1] and [2] and we omit them. Briefly, the series definition of $a_n(\nu, j)$ is inserted into the series for $\lambda_\nu(j;\tau)$, I_1 is replaced by the power series (3.03), Lemma (3.02) is used to justify several interchanges of summation, and use is made of the Lipschitz formula [3]

$$\sum_{n=1}^{\infty} n^p\{\exp[2\pi i(\tau/j - h/k)]\}^n$$
$$= \begin{cases} (p!/(2\pi)^{p+1}) \cdot \sum\limits_{l=\infty}^{\infty} (-i\tau/j + ih/k + li)^{-p-1}\,, & \text{for } p > 0 \\ -1/2 + (1/2\pi) \lim\limits_{N\to\infty} \sum\limits_{l=-N}^{N} (-i\tau/j + ih/k + li)^{-1}\,, & \text{for } p = 0\,. \end{cases}$$

We obtain the double series

$$(3.04) \quad \lambda_\nu(j;\tau) = \text{constant} + \frac{1}{16}\sum_{k=1}^{\infty}{}^{\#} \sum_{0\leqq h<k}{}^{*} e^{-2\pi i\nu h'/k}$$
$$\cdot \lim_{N\to\infty} \sum_{l=N}^{N} \left\{\exp\left[\frac{2\pi i\nu}{k(k\tau/j - h - kl)}\right] - 1\right\}.$$

4. **Transformation properties of the $\lambda_\nu(j;\tau)$.** In (3.04) put $m = h + kl$. Since $j \mid k$ it follows that $m \equiv h \pmod j$. Hence $m \equiv 1 \pmod j$ is a consequence of $h \equiv 1 \pmod j$. Also $(h, k) = 1$ implies $(m, k) = 1$. It is easy to see that as l runs through all the integers and h through a residue class modulo k with the restrictions $(h, k) = 1$ and $h \equiv 1 \pmod j$, then $h + kl$ takes on, exactly once, each integer value m such that $(m, k) = 1$ and $m \equiv 1 \pmod j$. Then (3.04) becomes

$$(4.01) \quad \lambda_\nu(j;\tau) = A + \frac{1}{16}\sum_{k=1}^{\infty}{}^{\#} \lim_{N\to\infty} \sum_{|m|\leqq N}{}^{*} e^{-2\pi i\nu m'/k}\left\{\exp\left[\frac{2\pi i\nu}{k(k\tau/j - m)}\right] - 1\right\}.$$

Let a, b, c, d be integers such that $ad - bc = 1$, $a \equiv d \equiv 1 \pmod j$, and $b \equiv c \equiv 0 \pmod j$. Denote by $T_{a,b,c,d}$ the element of $G(j)$ defined by

$$T_{a,b,c,d}(\tau) = \frac{a\tau + b}{c\tau + d}\,.$$

We wish to prove

THEOREM (4.02)[1]. *The function $\lambda_\nu(j;\tau)$ satisfies the transformation*

[1] See correction at end of paper.

equations

$$\lambda_\nu(j;\, T_{a,b,c,d}(\tau)) \equiv \lambda_\nu\left(j;\, \frac{a\tau + b}{c\tau + d}\right) = \lambda_\nu(j;\, \tau) + \omega_\nu(j;\, c, d)\,, \tag{4.03}$$

for all $T_{a,b,c,d}$ *in* $G(j)$ *and* $\mathscr{I}(\tau) > 0$. *Here* $\omega_\nu(j; c, d)$ *does not depend on* $\tau, a,$ *or* b.

Proof. Let us suppose we have already shown that

$$\lambda_\nu\left(j;\, \frac{a\tau + b}{c\tau + d}\right) = \lambda_\nu(j;\, \tau) + \omega\,,$$

where ω does not depend on τ. Under this assumption we prove that ω is independent of a and b.

Let $T_{a',b',c,d}$ be in $G(j)$. Then, since $a - a' \equiv b - b' \equiv 0 \pmod j$ and $ad - bc = a'd - b'c = 1$, we have that $a' = a + q'j, b' = b + r'j$, with q', r' integers and $q'd = r'c$. Since $(c, d) = 1$ it follows that $q' = qc$, $r' = qd$ with q an integer, and therefore $a' = a + qcj, b' = b + qdj$. Hence $T_{a',b',c,d} = T_{1,qj,0,1} \cdot T_{a,b,c,d}$, and clearly

$$\lambda_\nu\left(j;\, \frac{a'\tau + b'}{c\tau + d}\right) = \lambda_\nu\left(j;\, \frac{a\tau + b}{c\tau + d}\right) = \lambda_\nu(j;\, \tau) + \omega\,.$$

Therefore, ω does not depend on a or b.

It suffices to prove (4.03) subject to the restrictions $d > jc > 0, a < 0$, $b < 0$. First we may assume $c > 0$, changing the signs of a, b, c, d if necessary. It is then simple to compute that $T_{a,b,c,d} = T_{1,sj,0,1} \cdot T_{\alpha,\beta,\gamma,\delta} \cdot T_{1,-rj,0,1}$, with $\alpha = a - sjc, \beta = rj(a - sjc) + b - sjd, \gamma = c, \delta = d + rjc$, and we can choose integers r and s so large that $\alpha < 0, \beta < 0, \delta > jc$. But $\lambda_\nu(j; \tau)$ is clearly invariant under $T_{1,sj,0,1}$ and $T_{1,-rj,0,1}$ since these are translations by sj and $-rj$ respectively. Hence, if $\lambda_\nu(j;\, T_{\alpha,\beta,\gamma,\delta}(\tau)) = \lambda_\nu(j;\tau) + \omega$, then $\lambda_\nu(j;\, T_{a,b,c,d}(\tau)) = \lambda_\nu(j; \tau) + \omega$.

Now, in order to apply Lemmas (2.01) and (2.03) we assume that τ is a pure imaginary number. Expanding the expression in the braces in (4.01) into a power series, we get

$$\begin{aligned} \lambda_\nu(j;\, \tau) &= A + \frac{1}{16}\sum_{k=1}^{\infty}{}^{\#} \lim_{N\to\infty} \sum_{|m|\leqq N}{}^{*} e^{-2\pi i\nu m'/k} \sum_{p=1}^{\infty} \frac{1}{p!}\left(\frac{2\pi i\nu}{k(k\tau/j - m)}\right)^p \\ &= A + \frac{1}{16}\sum_{k=1}^{\infty}{}^{\#} \lim_{N\to\infty} \sum_{|m|\leqq N}{}^{*} e^{-2\pi i\nu m'/k} \frac{2\pi i\nu}{k(k\tau/j - m)} \\ &\quad + \frac{1}{16}\sum_{k=1}^{\infty}{}^{\#} \lim_{N\to\infty} \sum_{|m|\leqq N}{}^{*} e^{-2\pi i\nu m'/k} \sum_{p=2}^{\infty} \frac{1}{p!}\left(\frac{2\pi i\nu}{k(k\tau/j - m)}\right)^p . \end{aligned} \tag{4.04}$$

The separation into two sums is justified since the first is convergent by Lemma (2.01) and the second is an absolutely convergent triple sum.

It follows that the second sum can be rearranged in any fashion. Making use of this fact and noting that the restrictions $a < 0, b < 0, d > jc > 0$ make it possible to apply Lemma (2.01), with $r = 0$ and a, b, c, d replaced by $a, b/j, jc, d$ to the first sum, we obtain

$$\begin{aligned}\lambda_\nu(j;\tau) = A &+ \frac{1}{16}\lim_{K\to\infty}\Big\{\sum_{k=1}^{[K(dt-jc)]}{}^{\#}\sum_{|m-bk/jd|\leq K/d}^{*}\frac{e^{-2\pi i\nu m'/k}}{k(k\tau/j-m)}\cdot 2\pi i\nu \\ &+\sum_{k=[K(dt-jc)]+1}^{[K(dt+jc)]}{}^{\#}\sum_{(ak-Kt)/jc\leq m\leq bk/jd+K/d}^{*}\frac{e^{-2\pi i\nu m'/k}}{k(k\tau/j-m)}\cdot 2\pi i\nu\Big\} \\ &+\frac{1}{16}\lim_{K\to\infty}\Big\{\sum_{k=1}^{[K(dt-jc)]}{}^{\#}\sum_{|m-bk/jd|\leq K/d}^{*}e^{-2\pi i\nu m'/k}\sum_{p=2}^{\infty}\frac{1}{p!}\Big(\frac{2\pi i\nu}{k(k\tau/j-m)}\Big)^p \\ &+\sum_{k=[K(dt-jc)]+1}^{[K(dt+jc)]}{}^{\#}\sum_{(ak-Kt)/jc\leq m\leq bk/jd+K/d}^{*}e^{-2\pi i\nu m'/k}\sum_{p=2}^{\infty}\frac{1}{p!}\Big(\frac{2\pi i\nu}{k(k\tau/j-m)}\Big)^p\Big\}.\end{aligned}$$

Therefore,

(4.05)

$$\begin{aligned}\lambda_\nu(j;\tau) = A &+ \frac{1}{16}\lim_{K\to\infty}\Big\{\sum_{k=1}^{[K(dt-jc)]}{}^{\#}\sum_{|m-bk/jd|\leq K/d}^{*}e^{-2\pi i\nu m'/k}\Big(\exp\Big[\frac{2\pi i\nu}{k(k\tau/j-m)}\Big]-1\Big) \\ &+\sum_{k=[K(dt-jc)]+1}^{[K(dt+jc)]}{}^{\#}\sum_{(ak-Kt)/jc\leq m\leq bk/jd+K/d}^{*}e^{-2\pi i\nu m'/k}\Big(\exp\Big[\frac{2\pi i\nu}{k(k\tau/j-m)}\Big]-1\Big)\Big\}.\end{aligned}$$

Now, let

$$\begin{aligned}S_K(\tau) = &\sum_{k=1}^{[K(dt-jc)]}{}^{\#}\sum_{|m-bk/jd|\leq K/d}^{*}e^{-2\pi i\nu m'/k}\exp\Big[\frac{2\pi i\nu}{k(k\tau/j-m)}\Big] \\ &+\sum_{k=[K(dt-jc)]+1}^{[K(dt+jc)]}{}^{\#}\sum_{(ak-Kt)/jc\leq m\leq bk/jd+K/d}^{*}e^{-2\pi i\nu m'/k}\exp\Big[\frac{2\pi i\nu}{k(k\tau/j-m)}\Big].\end{aligned}$$

A little computing shows that

$$\begin{aligned}S_K(\tau) = &\sum_{k=1}^{[K(dt-jc)]}{}^{\#}\sum_{|m-bk/jd|\leq K/d}^{*}\exp\Big[2\pi i\nu\frac{-k'-m'\tau/j}{k\tau/j-m}\Big] \\ &+\sum_{k=[K(dt-jc)]+1}^{[K(dt+jc)]}{}^{\#}\sum_{(ak-Kt)/jc\leq m\leq bk/jd+K/d}^{*}\exp\Big[2\pi i\nu\frac{-k'-m'\tau/j}{k\tau/j-m}\Big],\end{aligned}$$

where $-k' = (mm'+1)/k$. We see that $kk' + mm' + 1 = 0$, so $kk' \equiv -1 \pmod{m}$. Now given the relatively prime pair k, m, the pair k', m' is not uniquely determined. In fact, m' can be replaced by $m' + qk$, where q is any integer. Then k must be replaced by $k' - qm$. The corresponding term in $S_K(\tau)$ is replaced by

$$\begin{aligned}\exp\Big[2\pi i\nu\frac{-k'+qm-(m'+qk)\tau/j}{k\tau/j-m}\Big] &= \exp\Big[2\pi i\nu\Big(\frac{-k'-m'\tau/j}{k\tau/j-m}-q\Big)\Big] \\ &= \exp\Big[2\pi i\nu\Big(\frac{-k'-m'\tau/j}{k\tau/j-m}\Big)\Big],\end{aligned}$$

so that $S_K(\tau)$ is unaffected by the ambiguity in the choice of k' and m'.

Recall that in $S_K(\tau)$ we are summing over the lattice points of the trapezoid bounded by the lines $k = 0$, $m = bk/jd - K/d$, $m = bk/jd + K/d$, $m = (ak - Kt)/jc$. Now, if the pair k, m is replaced by $-k, -m$, the pair k', m' is replaced by $-k', -m'$, and the corresponding term in $S_K(\tau)$ is unchanged. Therefore, if we extend our region of summation in $S_K(\tau)$ by reflecting the trapezoid through the origin, $S_K(\tau)$ is multiplied by 2. The new region of summation is the parallelogram, $\mathscr{P}(K)$, bounded by the four lines $m = bk/jd \pm K/d$, $m = (ak \pm Kt)/jc$. Therefore,

$$S_K(\tau) = \frac{1}{2} \sum^{\#}\sum^{*}_{(k,m)\in\mathscr{P}(K)} \exp\left[2\pi i\nu \frac{-k' - m'\tau/j}{k\tau/j - m}\right]. \tag{4.06}$$

It follows from this that

$$S_K\left(\frac{a\tau + b}{c\tau + d}\right) = \frac{1}{2} \sum^{\#}\sum^{*}_{(k,m)\in\mathscr{P}(K)} \exp\left[2\pi i\nu \frac{-(dk' + bm'/j) - (\tau/j)(jck' + am')}{(\tau/j)(ak - jcm) - (md - bk/j)}\right].$$

If the transformation $l = ak - jcm$, $n = -bk/j + md$ is performed, the parallelogram $\mathscr{P}(K)$ in the $k - m$ plane is mapped onto the rectangle defined by $|l| \leqq tK$, $|n| \leqq K$ in the $l - n$ plane. Furthermore, since $a \equiv d \equiv 1 \pmod j$, $b \equiv c \equiv 0 \pmod j$, and $ad - bc = 1$, there is a one-to-one correspondence set up between the set of all lattice points (k, m) in $\mathscr{P}(K)$ and the set of all lattice points (l, n) of the rectangle $|l| \leqq tK$, $|n| \leqq K$. Also, a little computing shows that $(ak - jcm)(dk' + bm'/j) + (md - bk/j)(jck' + am') + 1 = kk' + mm' + 1 = 0$. Therefore we can put $l' = dk' + bm'/j$, $n' = jck' + am'$, and we finally obtain

$$\begin{aligned} S_K\left(\frac{a\tau + b}{c\tau + d}\right) &= \frac{1}{2} \sum^{\#}_{|l|\leqq tK} \sum^{*}_{|n|\leqq K} \exp\left[2\pi i\nu \frac{-l' - n'\tau/j}{l\tau/j - n}\right] \\ &= \sum^{\#\,[tK]}_{l=1} \sum^{*}_{|n|\leqq K} \exp\left[2\pi i\nu \frac{-l' - n'\tau/j}{l\tau/j - n}\right]. \end{aligned} \tag{4.07}$$

Therefore, it follows from (4.05) that

$$\begin{aligned} &\lambda_\nu\left(j; \frac{a\tau + b}{c\tau + d}\right) \\ &= A + \frac{1}{16} \lim_{K\to\infty} \left\{ S_K\left(\frac{a\tau + b}{c\tau + d}\right) - \sum^{\#\,[K(dt-jc)]}_{k=1} \sum^{*}_{|m-bk/jd|\leqq K/d} e^{-2\pi i\nu m'/k} \right. \\ &\qquad \left. - \sum^{\#\,[K(dt+jc)]}_{k=[K(dt-jc)]+1} \sum^{*}_{(ak-Kt)/jc\leqq m\leqq bk/jd+K/d} e^{-2\pi i\nu m'/k} \right\} \\ &= A + \frac{1}{16} \lim_{K\to\infty} \left\{ \sum^{\#\,[tK]}_{l=1} \sum^{*}_{|n|\leqq K} \exp\left[2\pi i\nu \frac{-l' - n'\tau/j}{l\tau/j - n}\right] \right. \\ &\qquad - \sum^{\#\,[K(dt-jc)]}_{k=1} \sum^{*}_{|m-bk/jd|\leqq K/d} e^{-2\pi i\nu m'/k} \end{aligned} \tag{4.08}$$

$$- \sum_{k=[K(dt-jc)]+1}^{[K(dt+jc)]}{}^{\#} \sum_{(ak-Kt)/jc \leq m \leq bk/jd+K/d}{}^{*} e^{-2\pi i\nu m'/k} \Bigg\} .$$

We now return to (4.01) and apply Lemma (2.03) with $r = 0, \rho = t$. Proceeding in the same way as in the proof of (4.05), we find that

$$\begin{aligned}(4.09)\quad & \lambda_\nu(j;\tau) \\ &= A + \frac{1}{16}\lim_{K\to\infty}\sum_{k=1}^{[tK]}{}^{\#}\sum_{|m|\leq K}{}^{*} e^{-2\pi i\nu m'/k}\left(\exp\left[\frac{2\pi i\nu}{k(k\tau/j-m)}\right]-1\right) \\ &= A + \frac{1}{16}\lim_{K\to\infty}\left\{\sum_{k=1}^{[tK]}{}^{\#}\sum_{|m|\leq K}{}^{*}\exp\left[2\pi i\nu\frac{-k'-m'\tau/j}{k\tau/j-m}\right] - \sum_{k=1}^{[tK]}{}^{\#}\sum_{|m|\leq K}{}^{*} e^{-2\pi i\nu m'/k}\right\} .\end{aligned}$$

Upon comparing (4.08) and (4.09), we conclude that

$$\begin{aligned}& \lambda_\nu\left(j;\frac{a\tau+b}{c\tau+d}\right) - \lambda_\nu(j;\tau) \\ (4.10)\qquad & = \frac{1}{16}\lim_{k\to\infty}\left\{\sum_{k=1}^{[tK]}{}^{\#}\sum_{|m|\leq K}{}^{*} e^{-2\pi i\nu m'/k} - \sum_{k=1}^{[K(dt-jc)]}{}^{\#}\sum_{|m-bk/jd|\leq K/d}{}^{*} e^{-2\pi i\nu m'/k}\right. \\ & \left. - \sum_{k=[K(dt-jc)]+1}^{[K(dt+jc)]}{}^{\#}\sum_{(ak-Kt)/jc\leq m\leq bk/jd+K/d}{}^{*} e^{-2\pi i\nu m'/k}\right\} \equiv \omega_\nu(j;c,d) .\end{aligned}$$

We have proved the required transformation properties when τ is a pure imaginary number. But $\lambda_\nu(j;\tau)$ is regular for $\mathscr{I}(\tau) > 0$, and therefore, by analytic continuation, (4.10) holds for $\mathscr{I}(\tau) > 0$, and the proof of the theorem is complete.

There are other transformation properties of the $\lambda_\nu(j;\tau)$ for special values of ν. These can be summarized in the following.

THEOREM (4.11). (a) *If ν is a multiple of j, then for $\mathscr{I}(\tau) > 0$,*

$$(4.12)\qquad \lambda_\nu(j;-1/\tau) = \lambda_\nu(j;\tau) .$$

(b) *If j is even and ν is an odd multiple of $j/2$, then for $\mathscr{I}(\tau) > 0$,*

$$(4.13)\qquad \lambda_\nu(j;-1/\tau) = \sigma_\nu(j) - \lambda_\nu(j;\tau) ,$$

where $\sigma_\nu(j)$ does not depend on τ.

Proof. We again begin by assuming that τ is a pure imaginary number. Returning to (4.01), applying Lemma (2.03) with $r = 0, \rho = j$, and proceeding as in the proof of (4.05), we obtain

$$(4.14)\quad \lambda_\nu(j;\tau) = A + \frac{1}{16}\lim_{K\to\infty}\sum_{k=1}^{jK}{}^{\#}\sum_{|m|\leq K}{}^{*} e^{-2\pi i\nu m'/k}\left(\exp\left[\frac{2\pi i\nu}{k(k\tau/j-m)}\right]-1\right) .$$

This time, put

$$S_K(\tau) = \sum_{k=1}^{jk}{}^{\#}\sum_{|m|\leq K}{}^{*} e^{-2\pi i\nu m'/k}\exp\left[\frac{2\pi i\nu}{k(k\tau/j-m)}\right]$$

$$\text{(4.15)} \qquad = \sum_{k=1}^{jK}{}^{\#} \sum_{m=1}^{K}{}^{*} \exp\left[2\pi i\nu \frac{-k' - m'\tau/j}{k\tau/j - m}\right] + \sum_{k=1}^{jK}{}^{\#} \sum_{m=1}^{K}{}^{*} \exp\left[2\pi i\nu \frac{-k' + m'\tau/j}{k\tau/j + m}\right],$$

where we have separated the terms for $m < 0$ and $m > 0$.

It follows that

$$\text{(4.16)} \qquad S_K(-1/\tau) = \sum_{k=1}^{jK}{}^{\#} \sum_{m=1}^{K}{}^{*} \exp\left[2\pi i\nu \frac{-k'\tau + m'/j}{-k/j - m\tau}\right] + \sum_{k=1}^{jK}{}^{\#} \sum_{m=1}^{K}{}^{*} \exp\left[2\pi i\nu \frac{-k'\tau - m'/j}{-k/j + m\tau}\right].$$

Put $l = k/j$ and $n = jm$; it follows from $(k, m) = 1$, $k \equiv j \pmod{j^2}$, and $m \equiv 1 \pmod j$ that $(l, n) = 1$, $l \equiv 1 \pmod j$, and $n \equiv j \pmod{j^2}$. Also, we may put $l' = jk' - m$, $n' = (k/j + m')/j$. For $mm' \equiv -1 \pmod k$, $m \equiv 1 \pmod j$, and $j \mid k$ together imply that $m' \equiv -1 \pmod j$. Using the fact that $k/j \equiv 1 \pmod j$, we find that $k/j + m' \equiv 0 \pmod j$ and n', as defined above, is an integer. Furthermore, $ll' + nn' + 1 = kk' + mm' + 1 = 0$. With the above definition of l' and n', we have $k' = (l' + n/j)/j$ and $m' = jn' - l$. Now, (4.16) becomes

$$\text{(4.17)} \qquad \begin{aligned} S_K(-1/\tau) &= \sum_{n=1}^{jK}{}^{\#} \sum_{l=1}^{K}{}^{*} \exp\left[2\pi i\nu \frac{-(l' + n/j)\tau/j + (jn' - l)/j}{-l - n\tau/j}\right] \\ &\quad + \sum_{n=1}^{jK}{}^{\#} \sum_{l=1}^{K}{}^{*} \exp\left[2\pi i\nu \frac{-(l' + n/j)\tau/j - (jn' - l)/j}{-l + n\tau/j}\right] \\ &= \sum_{n=1}^{jK}{}^{\#} \sum_{l=1}^{K}{}^{*} \exp\left[2\pi i\nu \left(\frac{-n' + l'\tau/j}{n\tau/j + l} + 1/j\right)\right] \\ &\quad + \sum_{n=1}^{jK}{}^{\#} \sum_{l=1}^{K}{}^{*} \exp\left[2\pi i\nu \left(\frac{-n' - l'\tau/j}{n\tau/j - l} - 1/j\right)\right]. \end{aligned}$$

We see from (4.14) and the definition of $S_K(\tau)$ that

$$\lambda_\nu(j; \tau) = A + \frac{1}{16} \lim_{K\to\infty} \left\{S_K(\tau) - \sum_{k=1}^{jK}{}^{\#} \sum_{|m| \leqq K}{}^{*} e^{-2\pi i\nu m'/k}\right\}.$$

Now, if ν is a multiple of j, a comparison of (4.15) and (4.17) shows that $S_K(-1/\tau) = S_K(\tau)$ and therefore (4.12) follows. This is part (a) of the theorem. In part (b), $S_K(-1/\tau) = -S_K(\tau)$, and therefore,

$$\lambda_\nu(j; -1/\tau) + \lambda_\nu(j; \tau) = 2A - \frac{1}{8} \lim_{K\to\infty} \sum_{k=1}^{jK}{}^{\#} \sum_{|m|\leqq K}{}^{*} e^{-2\pi i\nu m'/k} \equiv \sigma_\nu(j).$$

This is part (b) of the theorem. Here again the theorem has been proved for τ a pure imaginary number, but as before we extend our results by analytic continuation to all τ such that $\mathscr{I}(\tau) > 0$.

5. Construction of functions for $G(j)$. In order to construct functions which are invariant under the group $G(j)$, we make use of Theorem (4.02) and the fact that $G(j)$ is finitely generated. Let T_l, $l = 1, \cdots, q(j)$, be a set of generators for $G(j)$. Then by Theorem (4.02), we have

$$\lambda_\nu(j; T_l(\tau)) - \lambda_\nu(j; \tau) = c_{l,\nu}(j)\ , \qquad l = 1, \cdots, q(j)\ , \tag{5.01}$$

for any integer $\nu \geqq 1$.

Let $1 \leqq \nu_1 < \nu_2 < \cdots < \nu_m$ be integers and consider the function defined by

$$F(\tau) = b_1\lambda_{\nu_1}(j; \tau) + \cdots + b_m\lambda_{\nu_m}(j; \tau)\ . \tag{5.02}$$

Then $F(\tau)$ satisfies the functional equations

$$F(T_l(\tau)) - F(\tau) = b_1 c_{l,\nu_1}(j) + \cdots + b_m c_{l,\nu_m}(j)\ , \quad l = 1, \cdots, q(j)\ . \tag{5.03}$$

Let $m \geqq q(j) + 1$ and consider the homogeneous linear system in the m unknowns $b_1, \cdots, b_m$

$$b_1 c_{l,\nu_1}(j) + \cdots + b_m c_{l,\nu_m}(j) = 0\ , \qquad l = 1, \cdots, q(j)\ . \tag{5.04}$$

This has $m - q(j)$ linearly independent solutions $(b_1, \cdots, b_m)$. With $b_1, \cdots, b_m$ chosen to satisfy (5.04), put

$$\mathscr{L}(j; b_1, \cdots, b_m; \nu_1, \cdots, \nu_m; \tau) = b_1\lambda_{\nu_1}(j; \tau) + \cdots + b_m\lambda_{\nu_m}(j; \tau)\ . \tag{5.05}$$

It follows from (5.03) and (5.04) that $\mathscr{L}(j; b_1, \cdots, b_m; \nu_1, \cdots, \nu_m; T_l(\tau)) = \mathscr{L}(j; b_1, \cdots, b_m; \nu_1, \cdots, \nu_m; \tau)$ for $l = 1, \cdots, q(j)$ and therefore, since the T_l generate $G(j)$, we have

$$\mathscr{L}(j; b_1, \cdots, b_m; \nu_1, \cdots, \nu_m; T(\tau)) = \mathscr{L}(j; b_1, \cdots, b_m; \nu_1, \cdots, \nu_m; \tau)\ , \tag{5.06}$$

for all T in $G(j)$.

In order to show that the function $\mathscr{L}$ defined by (5.05) cannot reduce to a constant we prove

LEMMA (5.07). *Let d_n be the* nth *Fourier coefficient of the function $\mathscr{L}$. Then*

$$d_n \sim (b_m/16)\nu_m^{1/4} n^{-3/4}(2j)^{-1/2} e^{-2\pi i(n-\nu_m)/j} \exp\,[4\pi(n\nu_m)^{1/2}/j]\ . \tag{5.08}$$

Proof. We see immediately from (5.05) that $d_n = \sum_{i=1}^m b_i a_n(\nu_i, j)$, with $a_n(\nu_i, j)$ defined as in (3.01). The lemma now is direct consequence of Lemma (3.02a)

In particular, (5.08) implies that $\mathscr{L}$ is not a constant.

6. Construction of forms for $G(j)$. Let r be any positive even integer. We define the function

$$(6.01)\qquad \begin{aligned} \lambda_\nu(j;\tau,r) &= \sum_{n=1}^{\infty} a_n(\nu,j,r)e^{2\pi i n\tau/j} \\ a_n(\nu,j,r) &= \{(-1)^{r/2}\pi/8\}\sum_{k=1}^{\infty}{}^{\#} k^{-1}A_{k,\nu}(n)\cdot(\nu/n)^{(r+1)/2}I_{r+1}(4\pi(n\nu)^{1/2}/k)\ , \end{aligned}$$

where $A_{k,\nu}(n)$ is defined as in (3.01) and I_{r+1} is again a Bessel function of the first kind. It should be noted that if we put $r = 0$ in (6.01) we obtain the function $\lambda_\nu(j;\tau)$ defined by (3.01).

The computations of §§ 3 and 4, using Lemmas (2.01) and (2.03), with $r > 0$, yield the following two theorems.

THEOREM (6.02)[2]. *The function $\lambda_\nu(j;\tau,r)$ satisfies the transformation equations*

$$(6.03)\qquad \begin{aligned} (c\tau+d)^r\lambda_\nu(j;T_{a,b,c,d}(\tau),r) &\equiv (c\tau+d)^r\lambda_\nu\Big(j;\frac{a\tau+b}{c\tau+d},r\Big) \\ &= \lambda_\nu(j;\tau,r) + p_\nu(j;\tau,r;c,d)\ , \end{aligned}$$

for all $T_{a,b,c,d}$ in $G(j)$ and $\mathscr{I}(\tau) > 0$, where $p_\nu(j;\tau,r;c,d)$ is a polynomial in τ of degree at most r.

THEOREM (6.04). (a) *If ν is a multiple of j, then for $\mathscr{I}(\tau) > 0$,*

$$(6.05)\qquad \tau^r\lambda_\nu(j;-1/\tau,r) = \lambda_\nu(j;\tau,r) + p_{\nu,1}(j;\tau,r)\ ,$$

where $p_{\nu,1}(j;\tau,r)$ is a polynomial in τ of degree at most r.
(b) *If j is even and ν is an odd multiple of $j/2$, then for $\mathscr{I}(\tau) > 0$,*

$$(6.06)\qquad \tau^r\lambda_\nu(j;-1/\tau,r) = p_{\nu,2}(j;\tau,r) - \lambda_\nu(j;\tau,r)\ ,$$

where $p_{\nu,2}(j;\tau,r)$ is a polynomial in τ of degree at most r.

Now, in order to construct forms of dimension r for $G(j)$, we make use of Theorem (6.02) and proceed as in § 5. We take a linear combination of the $\lambda_\nu(j;\tau,r)$ in such a way that the resulting linear combination of polynominals occurring in the transformation equation connected with T_l, $l = 1,\cdots,q(j)$, vanishes identically. In this case m, the number of $\lambda_\nu(j;\tau,r)$ in the linear combination, must be such that $m \geqq (r+1)\cdot q(j) + 1$.

A simple generalization of Lemma (5.07), to cover the present case, shows that the forms constructed in this way are not identically zero.

7. Conclusion. Other functions of the type dealt with in this paper can be constructed by generalizing the congruence conditions on k and h in (3.01) and (6.01). Let n_1 and n_2 be any integers. If, in

[2] See correction at end of paper.

(3.01), we impose the new congruence conditions $k \equiv n, j \pmod{j^2}$, $h \equiv n_2 \pmod{j}$, we obtain new functions which satisfy (4.03), and which, therefore, can be used to construct functions which are invariant under $G(j)$.

If $(n_2, j) > 1$, the sum defining $A_{k,\nu}(n)$ is empty and so each Fourier coefficient is zero. Also the case $n_1 \equiv 0 \pmod{j}$, $n_2 \equiv 1 \pmod{j}$ is unique and will receive separate treatment in another publication. The distinctive feature here is the fact that, in order to construct functions satisfying (4.03), we must introduce a pole term at $i\infty$. This situation occurs, for example, in the Fourier expansion of $\mu(\tau)$, the reciprocal of $\lambda(\tau)$ (see [6]).

Making the additional assumptions $n_1 = n_2$, $n_1^2 \equiv 1 \pmod{j}$ in (3.01), we obtain functions for which we can prove Theorem (4.11).

Correspondingly, if we impose the conditions $k \equiv n_1, j \pmod{j^2}$, $h \equiv n_2 \pmod{j}$ in (6.01), we obtain functions satisfying (6.03), and making the further assumptions, $n_1 = n_2$, $n_1^2 \equiv 1 \pmod{j}$, we obtain functions for which Theorem (6.04) holds.

It should be noted that all of our functions vanish at the parabolic cusp at infinity. As the referee has pointed out, it is of interest to consider the behavior of these functions at the other parabolic cusps of $G(j)$. This question will be treated at a later time.

Correction to "Construction of a Class of Modular Functions and Forms". As it stands the proof of Theorem (4.02) is incorrect. The difficulty arises in the paragraph immediately preceding (4.06), where we extend the region of summation in $S_K(\tau)$. In the original expression for $S_K(\tau)$ we are summing over the points (k, m) of a certain trapezoid subject to the additional restrictions $(m, k) = 1$, $k \equiv j \pmod{j^2}$, $m = 1 \pmod{j}$. In order to extend the region of summation to the parallelogram $\mathscr{P}(K)$, we reflect this trapezoid through the origin. That is, when (k, m) appears in the summation, we also include the point $(-k, -m)$. The trouble is, that when $j \geqq 3$, $(-k, -m)$ does not satisfy the proper congruence conditions, but rather the new conditions $-k \equiv -j \pmod{j^2}$, $-m \equiv -1 \pmod{j}$, or equivalently, $-k \equiv j^2 - j \pmod{j^2}$, $-m \equiv j - 1 \pmod{j}$. Hence the expression (4.06) for $S_K(\tau)$ is incorrect, when $j \geqq 3$. For $j = 2$, of course, this difficulty does not arise.

The situation can be readily rectified if we go back to (3.01) and modify the definition of the function $\lambda_\nu(j; \tau)$. Put $b_n^+(\nu, j) = a_n(\nu, j)$, with $a_n(\nu, j)$ as in (3.01) and define $b_n^-(\nu, j)$ to be the same as $a_n(\nu, j)$, except that the congruence condition on k is changed to $k \equiv j^2 - j \pmod{j^2}$ and the congruence condition on h is changed to $h \equiv j - 1 \pmod{j}$. We now define $\lambda_\nu(j; \tau)$ by

$$\lambda_\nu(j; \tau) = \sum_{n=1}^{\infty} b_n(\nu, j) e^{2\pi i n\tau/j} ,$$

where

$$b_n(\nu, j) = \tfrac{1}{2}[b_n^+(\nu, j) + b_n^-(\nu, j)] .$$

when $j = 2$, $b_n^+(\nu, j) = b_n^-(\nu, j) = b_n(\nu, j) = a_n(\nu, j)$, and no change has been made.

With this new definition of $\lambda_\nu(j; \tau)$ (4.06) becomes

$$S_K(\tau) = \tfrac{1}{2} \sum \sum \exp \left[2\pi i \nu \frac{-k' - m'\tau/j}{k\tau/j - m} \right],$$

where the summation is over all points of $\mathscr{P}(K)$ such that $(m, k) = 1$. and *either* $k \equiv j \pmod{j^2}$, $m \equiv 1 \pmod{j}$ *or* $k \equiv j^2 - j \pmod{j^2}$, $m \equiv j-1 \pmod{j}$ The remainder of the proof now carries through.

The same remark is necessary in connection with Theorem (6.02). That is, Theorem (6.02) is incorrect as it stands, but if we modify the function $\lambda_\nu(j; \tau, r)$ in the same way as we modified $\lambda_\nu(j; \tau)$, the proof goes through.

We should point out that Theorems (4.11) and (6.04) are correct as they are, but in addition Theorem (4.11) is true for the modified $\lambda_\nu(j; \tau)$ and Theorem (6.04) is true for the modified $\lambda_\nu(j; \tau, r)$.

Similar modifications have to be made in the definition of the functions mentioned in § 7.

Bibliography

1. M. I. Knopp, *Fourier series of automorphic forms of nonnegative dimension*, Illinois J. Math. to appear.
2. J. Lehner, *On modular forms of negative dimension*, Michigan Math. J. **6** (1959), 71-88.
3. R. Lipschitz, *Untersuchung der eigenschaften einer gattung von unendlichen Reihen*, Journal für Math. **105** (1889), 127-156.
4. H.Rademacher, *The Fourier series and the functional equation of the absolute modular invariant* $J(\tau)$, Amer. J. Math. **61** (1939), 237-248.
5. H. Salié, *Zur abschätzung der Fourierkoeffizienten ganzer modulformen*, Math. Zeit. **36** (1933), 263-278.
6. W. H. Simons, *The Fourier coefficients of the modular function* $\lambda(\tau)$, Canadian J. Math. **4** (1952), 67-80.
7. G. N. Watson, *Theory of Bessel Functions*, Cambridge, Cambridge University Press, 1945.

THE INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY

COMMUTING BOOLEAN ALGEBRAS OF PROJECTIONS

C. A. McCarthy

0. Introduction. One of the more important problems in the theory of spectral operators is to decide when the sum and product of two bounded commuting spectral operators is again spectral. J. Wermer [7] has shown that the sum and product of two bounded commuting spectral operators on Hilbert space is again spectral. N. Dunford [4, Theorem 19] and S. R. Foguel [5, Theorem 7] have shown that if the Boolean algebra of projections generated by the resolutions of the identity of two bounded commuting spectral operators on a weakly complete Banach space is bounded, then the sum and product of these operators are spectral. We therefore wish to determine conditions that insure the boundedness of the Boolean algebra of projections generated by two bounded commuting algebras of projections on a Banach space. We shall show that it suffices that one of the original algebras be strongly complete, countably decomposable, and contains no projection of infinite multiplicity. The example of S. Kakutani [6] shows that the Boolean algebra of projections generated by two commuting, strongly complete, algebras of bound 1, but both of infinite multiplicity on a non weakly complete space, need not be bounded. By slightly reworking his example, we shall show that the order of magnitude of our estimates is sharp, even for spaces of finite dimension. By taking a suitable direct sum of these examples, we obtain a separable reflexive Banach space on which we have two commuting, strongly complete, Boolean algebras of projections, both of bound 1, neither having a projection of infinite uniform multiplicity, but such that the algebra of projections they generate is unbounded. On this same Banach space we also show that the sum and product of two bounded commuting spectral operators need not be spectral.

This paper is divided into four sections: the first is devoted to the proof of a combinatorial inequality, the second contains our main theorem on the boundedness of projections, the third section consists of examples. The last section is an appendix to section two.

1. A combinatorial inequality. The required inequality is the

Received August 18, 1959, and in revised form January 4, 1961.

Part of this work appeared in the author's doctoral dissertation presented to Yale University, written under Professor Einar Hille, supported by the National Science Foundation. The remainder of the work was supported by the United States Air Force under contract AF-49(638)-42, monitored by the Air Force Office of Scientific Research of the Air Research and Development Command. The author would like to express particular thanks to W. G. Bade who provided access to unpublished material.

assertion of the following theorem.

THEOREM 1.1. *Let $\alpha_1, \cdots, \alpha_N$ be any N complex numbers, and let $\mathscr{S}$ be the collection of all subsets S of the set $1, \cdots, N$ of indices. Then for any S_0 in $\mathscr{S}$,*

$$\left|\sum_{s\in S_0} \alpha_s\right| \leqq 2\sqrt{N\pi}\cdot 2^{-N} \sum_{S\in\mathscr{S}} \left|\sum_{s\in S} \alpha_s\right| . \tag{1.1}$$

That is, the sum of any particular subset of the α's cannot exceed in absolute value the average of the absolute values of sums taken over all subsets by more than a factor which has order of magnitude $N^{1/2}$.

It suffices to prove the slightly stronger

THEOREM 1.1. a. *Let $\beta_1, \cdots, \beta_{2N}$ be any $2N$ complex numbers, and $\mathscr{R}$ the collection of all subsets R of $\{1, \cdots, 2N\}$. Then*

$$\left|\sum_{r=1}^{n} \beta_r\right| \leqq 2\sqrt{N\pi}\cdot 4^{-N} \sum_{R\in\mathscr{R}} \left|\sum_{r\in R} \beta_r\right| . \tag{1.2}$$

This implies Theorem 1.1, for suppose that N, S_0, and the α's of that theorem are given, with $S_0 = \{s_1, \cdots, s_n\}$. Define

$$\beta_r = \alpha_{s_r}, \quad 1 \leqq r \leqq n; \quad \beta_{N+r} = 0, 1 \leqq r \leqq N;$$
$$\beta_r = 0, \quad n+1 \leqq r \leqq N; \quad \beta_{N+r} = \alpha_{s_r}, \quad n+1 \leqq r \leqq N;$$

where $s_{n+1}, \cdots, s_N$ are those integers between 1 and N which are not in S_0. Then we have

$$\left|\sum_{s\in S_0} \alpha_s\right| = \left|\sum_{r=1}^{n} \beta_r\right| .$$

Also, every S in $\mathscr{S}$ determines 2^N R's in $\mathscr{R}$: namely

$$\{r \mid 1 \leqq r \leqq n \text{ and } s_r \in S\} \cup \{r \mid n+1 \leqq r \leqq N \text{ and } s_r \in S\}$$

together with any of the 2^N subsets of $\{n+1, \cdots, N+n\}$, such that

$$\left|\sum_{s\in S} \alpha_s\right| = \left|\sum_{r\in R} \beta_r\right| ,$$

so that

$$2^N \sum_{S\in\mathscr{S}} \left|\sum_{s\in S} \alpha_s\right| = \sum_{R\in\mathscr{R}} \left|\sum_{r\in R} \beta_r\right| .$$

Now if (1.2) holds, then we have

$$\left|\sum_{s\in S_0} \alpha_s\right| \leqq 2\sqrt{N\pi}\cdot 2^{-N} \sum_{S\in\mathscr{S}} \left|\sum_{s\in S} \alpha_s\right|$$

which is (1.1).

We will now show that it suffices to prove Theorem 1.1.a in the special case

$$\beta_1 = \cdots = \beta_N = 1 \ , \quad \beta_{N+1} = \cdots = \beta_{2N} = -1 \ .$$

We will first show that if we replace both β_i and β_j, $1 \leqq i, j \leqq N$, by their common average $\frac{1}{2}(\beta_i + \beta_j)$ and we have (1.2) for this new set of β's, then we necessarily had (1.2) for our original β's (Lemma 1.2 below). We then show that we can perform these two-at-a-time averagings in such a way as to eventually make the resulting β_i's, $1 \leqq i \leqq N$, all arbitrarily close to their common average (Lemma 1.3 below). By the continuity of both sides of (1.2) in the β_i's, it then suffices to prove (1.2) in the case $\beta_1 = \cdots = \beta_N$. Similarly, we may assume $\beta_{N+1} = \cdots = \beta_{2N}$. By re-indexing the β's if necessary, we may suppose

$$\left| \sum_{r=1}^{N} \beta_r \right| \geqq \left| \sum_{r=N+1}^{2N} \beta_r \right| \ ;$$

and by the homogenity of both sides of (1.2), it suffices to prove Theorem 1.2 in the case $\beta_1 = \cdots = \beta_N = 1$, $\beta_{N+1} = \cdots = \beta_{2N} = \gamma$ where γ is some complex number, $|\gamma| \leqq 1$. We will then show that we need only consider $\gamma = -1$ (Lemma 1.4 below).

LEMMA 1.2. *Suppose we set* $\beta'_1 = \beta'_2 = \frac{1}{2}(\beta_1 + \beta_2)$, $\beta'_r = \beta_r$, $3 \leqq r \leqq N$. *Then if* (1.2) *holds for the* β''s, *then it holds for the* β's.

Proof. Partition $\mathscr{R}$ into four disjoint classes:

$$\mathscr{R}_1 = \{R \mid 1 \in R, 2 \in R\} \ , \quad \mathscr{R}_3 = \{R \mid 1 \notin R, 2 \in R\} \ ,$$
$$\mathscr{R}_2 = \{R \mid 1 \in R, 2 \notin R\} \ , \quad \mathscr{R}_4 = \{R \mid 1 \notin R, 2 \notin R\} \ .$$

If R is in $\mathscr{R}_1$ or $\mathscr{R}_4$, then $\sum_{r\in R} \beta_r = \sum_{r \in R} \beta'_r$. Now note that there is a one-to-one correspondence between $\mathscr{R}_2$ and $\mathscr{R}_3$: R is in $\mathscr{R}_2$ if and only if $R' = R \cup \{2\} - \{1\}$ is in $\mathscr{R}_3$. Then we have

$$\begin{aligned} \left| \sum_{r\in R} \beta'_r \right| + \left| \sum_{r \in R'} \beta'_r \right| &= \left| \beta_1 + \beta_2 + 2 \sum_{r \in R \cap R'} \beta_r \right| \\ &= \left| \sum_{r \in R} \beta_r + \sum_{r \in R'} \beta_r \right| \\ &\leqq \left| \sum_{r\in R} \beta_r \right| + \left| \sum_{r \in R'} \beta_r \right| \ . \end{aligned}$$

Summing over all R in $\mathscr{R}_2$, we have

$$\sum_{R \in \mathscr{R}_2 \cup \mathscr{R}_3} \left| \sum_{r \in R} \beta'_r \right| \leqq \sum_{R \in \mathscr{R}_2 \cup \mathscr{R}_3} \left| \sum_{r \in R} \beta_r \right| \ ;$$

together with equality for R in $\mathscr{R}_1$ and $\mathscr{R}_4$, this proves the lemma. Note that the use of the particular indices 1 and 2 is irrelevant for our purposes; we only need that both indices are no greater than N or that both exceed N, so that $\sum_{r=1}^{N} \beta'_r = \sum_{r=1}^{N} \beta_r$.

LEMMA 1.3. *Let $\beta_1, \cdots, \beta_N$ be any N complex numbers. Then by a finite sequence of two-at-a-time averagings, we may obtain new numbers $\beta'_1, \cdots, \beta'_N$ such that $\max_{i,j} |\beta'_i - \beta'_j|$ is arbitrarily small.*

Proof. Suppose that all the β's are real and let β be their average. Let $\theta = \max_r |\beta - \beta_r|$. Partition $\{1, \cdots, N\}$ into three disjoint classes:

$$R_1 = \{r \mid \beta - \theta \leqq \beta_r < \beta - \theta/3\} ,$$
$$R_2 = \{r \mid \beta - \theta/3 \leqq \beta_r \leqq \beta + \theta/3\} ,$$
$$R_3 = \{r \mid \beta + \theta/3 < \beta_r \leqq \beta + \theta\} .$$

By averaging a β_i, i in R_1 with a β_j, j in R_3, we obtain numbers between $\beta - \theta/3$ and $\beta + \theta/3$; by doing this, we may exhaust either R_1 or R_3, so that we may initially assume that one of these, say R_3, is empty. In this case the cardinality of R_2 must exceed that of R_1, for otherwise the sum of the β's would be less than $N\beta$. Now we may average each β_i, i in R_1, with a distinct β_j, j in R_2, and obtain numbers between $\beta - 2\theta/3$ and β. Then if β'_r are the resultant set of numbers, $\max_r |\beta - \beta'_r| \leqq 2\theta/3$. By repeating this process, we may arrive at numbers differing arbitrarily little from β. For complex β's, we first perform two-at-a-time averagings to make the real parts of the β's as nearly equal as desired, and then do the same for the imaginary parts. Notice that when we perform any averagings, neither the maximum difference of the real parts nor of the imaginary parts can increase, so that when we average to make the imaginary parts nearly equal, we do not increase the maximum difference of the real parts.

We therefore assume $\beta_1 = \cdots = \beta_N = 1$ and $\beta_{N+1} = \cdots = \beta_{2N} = \gamma$, $|\gamma| \leqq 1$. Now each set R of $\mathscr{R}$ determines two integers k and p which are respectively the numbers of indices of R which do not, resp. do, exceed N. For such an R, $|\sum_{r \in R} \beta_r| = |k + p\gamma|$. Since there are $\binom{N}{k}$ subsets of $\{1, \cdots, N\}$ of cardinality k, and $\binom{N}{p}$ subsets of $\{N+1, \cdots, 2N\}$ of cardinality p, the number of R's for which $|\sum_{r \in R} \beta_r| = |k + p\gamma|$ is $\binom{N}{k}\binom{N}{p}$. Thus in this case (1.2) becomes

$$N \leqq A_N(\gamma) = 2 \cdot \sqrt{N\pi} 4^{-N} \sum_{k=0}^{N} \sum_{p-0}^{N} |k + p\gamma| \binom{N}{k}\binom{N}{p} .$$

Since $|k + p\gamma| \geqq |k - p|\gamma||$, it suffices to prove that $A_N(-1) \leqq A_N(\gamma)$, $-1 \leqq \gamma \leqq 0$, and then that $A_N(-1) \geqq N$.

Lemma 1.4. *$A_N(-1) \leqq A_N(\gamma)$ for all $|\gamma| \leqq 1$*

Proof. We have just seen that it suffices to consider real negative γ; to see that it suffices to consider $\gamma = -1$, note that for fixed N,

$$\frac{1}{2\sqrt{N\pi}}\cdot 4^N A_N = (\gamma)\sum_{k=1}^{N}\sum_{p=0}^{N}|k+p\gamma|\binom{N}{k}\binom{N}{p} = G_N(\gamma)$$

is a piecewise linear continuous function of γ. Where it exists, its derivative with respect to γ is

$$\begin{aligned}
&\sum_{k=0}^{N}\sum_{p=0}^{[k/|\gamma|]} p\binom{N}{k}\binom{N}{p} - \sum_{k=0}^{N}\sum_{p=[k/|\gamma|]+1}^{N} p\binom{N}{k}\binom{N}{p} \\
&\geqq \sum_{k=0}^{N}\sum_{p=0}^{k} p\binom{N}{k}\binom{N}{p} - \sum_{k=0}^{N}\sum_{p=k+1}^{N} p\binom{N}{k}\binom{N}{p} \\
&= \sum_{k=0}^{N}\sum_{p=0}^{k} p\binom{N}{k}\binom{N}{p} - \sum_{N-k=0}^{N}\sum_{N-p+1=1}^{N-k} p\binom{N}{k}\binom{N}{p} \\
&= \sum_{k=0}^{N}\sum_{p=0}^{k}\left[p\binom{N}{k}\binom{N}{p} - (N-p+1)\binom{N}{N-k}\binom{N}{N-p+1}\right] \\
&= 0\,.
\end{aligned}$$

Thus $G_N(\gamma)$ is a non-decreasing function of γ and so obtains its minimum at $\gamma = -1$.

Finally, we compute $G_N = G_N(-1)$. We have

$$\begin{aligned}
G_{N+1} &= \sum_{k=0}^{N+1}\sum_{p=0}^{N+1}|k-p|\binom{N+1}{p}\binom{N+1}{k} \\
&= \sum_{k=0}^{N+1}\sum_{p=0}^{N+1}|k-p|\left[\binom{N}{p}\binom{N}{k} + \binom{N}{p}\binom{N}{k-1} + \binom{N}{p-1}\binom{N}{k} + \binom{N}{p-}\binom{N}{k-1}\right] \\
&= \sum_{k=0}^{N}\sum_{p=0}^{N}|k-p|\binom{N}{p}\binom{N}{k} + \sum_{k=0}^{N}\sum_{p=0}^{N}|k+1-p|\binom{N}{p}\binom{N}{k} \\
&\quad + \sum_{k=0}^{N}\sum_{p=0}^{N}|k-p-1|\binom{N}{p}\binom{N}{k} + \sum_{k=0}^{N}\sum_{p=0}^{N}|k-p|\binom{N}{p}\binom{N}{k} \\
&= 4G_N + 2\sum_{k=0}^{N}\binom{N}{k}^2 = 4G_N + 2\binom{2N}{N}\,.
\end{aligned}$$

We have used the convention $\binom{N}{n} = 0$ if $n < 0$ or $n > N$. The third equality is a simple change of index of summation. The next-to-last equality comes from noting that

$$|k-p-1| + |k-p+1| - 2|k-p| = \begin{cases} 0 & \text{if } k \neq p \\ 2 & \text{if } k = p\,. \end{cases}$$

We then have by an easy induction

$$G_N = 4^N\frac{\Gamma(N+1/2)}{\sqrt{\pi}\,\Gamma(N)}\,,$$

whence by Stirling's formula, and the crudest sort of estimates,

$$G_N \geqq 4^N \cdot \frac{1}{2}\sqrt{\frac{N}{\pi}},$$

so that $A_N \geqq N$.

2. The boundedness theorem. Let X be a Banach space, X^* its adjoint, $\mathscr{E}$ and $\mathscr{F}$ bounded Boolean algebras of projections on X, with bounds M_1 and M_2 respectively, such that $EF = FE$ for all E in $\mathscr{E}$ and F in $\mathscr{F}$; E will be assumed to be strongly complete [1, Definition 2.1]. I is the identity operator on X and will be assumed to belong to both $\mathscr{E}$ and $\mathscr{F}$; we denote $I - E$ $(I - F)$ by E' (F'). The operator $\sum_\iota a_\iota E_\iota$, where the E_ι are mutually disjoint projections from $\mathscr{E}$ and $\sup |a_\iota| < \infty$, is a bounded operator on X with norm at most $4M_1 \cdot \sup |a_\iota|$ [4, p. 341]. We use the usual lattice supremum, infimum, and comparison signs for our projections as well as for closed subspaces of X: $E_1 \vee E_2 = E_1 + E_2 - E_1E_2$, $E_1 \wedge E_2 = E_1E_2$, $E_1 \leqq E_2$ if and only if $E_1E_2 = E_1$; $\mathfrak{M}_1 \vee \mathfrak{M}_2$ is the smallest closed manifold in X containing both of the closed manifolds $\mathfrak{M}_1$ and $\mathfrak{M}_2$, $\mathfrak{M}_1 \wedge \mathfrak{M}_2$ is the intersection of $\mathfrak{M}_1$ and $\mathfrak{M}_2$, and $\mathfrak{M}_1 \leqq \mathfrak{M}_2$ means that $\mathfrak{M}_1$ is contained in $\mathfrak{M}_2$. $\mathfrak{M}(x)$ denotes the least closed manifold of X containing Ex for all E in $\mathscr{E}$. If x is in X, we call the projection in $\mathscr{E}$, $C(x) = \wedge\{Ex \mid Ex = x\}$ the *carrier projection* of x; x is *full over* E if $C(x) \geqq E$.

We assume that there is an integer N for which the following condition $(*_N)$ holds:

$(*_N)$ *Let x be in X, and suppose that $\mathfrak{M}(F_ix) \wedge \bigwedge_{j\neq i} \mathfrak{M}(F_jx) = 0$ for all i, $1 \leqq i \leqq n$, for some choice of $F_1, \cdots, F_n$. Then either $\bigwedge_{i=1}^n C(F_ix) = 0$, or else $n \leqq N$.*

This condition holds, for example, if $\mathscr{E}$ is countably decomposable and has no projection of infinite multiplicity. The proof requires rather extensive background material which we will have no other occasion to use, and so is deferred to an appendix.

We wish to obtain a bound for the norm of $\bigvee_{m=1}^M E_mF_m$ which is independent of M and the particular E_m's in $\mathscr{E}$ and F_m's in $\mathscr{F}$. Accordingly, fix $E_m \in \mathscr{E}$, and $F_m \in \mathscr{F}$, $m = 1, \cdots, M$; $x \in X$ and $x^* \in X^*$ with $|x| \leqq 1$, $|x^*| \leqq 1$. We will estimate $x^*\sum_{m=1}^M E_mF_mx$.

First notice that, without loss of generality, we may assume that the F_m's are all disjoint: let L be an index running over all subsets of $\{1, \cdots, M\}$, and define

$$E_L = \bigvee_{l\in L} E_l\,, \qquad F_L = \bigwedge_{l\in L} F_l \wedge \bigwedge_{l\notin L} F_l'\,.$$

It is well known that the non-zero F_L are the atoms of the Boolean algebra of projections generated by the F_m's, and are mutually disjoint with sum I. Now we have

$$\begin{aligned}\bigvee_\iota E_\iota F_{\iota'} &\leqq \bigvee_\iota E_\iota(\bigvee_{\{L|\iota\in L\}} F_L) \leqq \bigvee_\iota(\bigvee_{\{L|\iota\in L\}} E_L F_L) \\ &\leqq \bigvee_L E_L F_L \leqq \bigvee_L(\bigvee_{\iota\in L} E_\iota)F_L \leqq \bigvee_L(\bigvee_{\iota\in L} E_\iota F_\iota) \leqq \bigvee_\iota E_\iota F_\iota ;\end{aligned}$$

thus we have found a way of expressing $\bigvee_{m=1}^M E_m F_m$ with the F's disjoint.

Now let J and K be two indices running over all subsets of $\{1, \cdots, M\}$, and define

$$E_J = \bigwedge_{j\in J} E_j \wedge \bigwedge_{j\notin J} E_j' , \qquad G_K = \bigwedge_{k\in K} C(F_k x) \wedge \bigwedge_{k\notin K} C(F_k x)' .$$

$\{E_J\}$ and $\{G_K)$ are both disjoint families of projections with sum I.

LEMMA 2.1. 1. $C(F_k x) = \bigvee_{\{K|k\in K\}} G_K$,
2. $G_K F_k x = 0$ *if* $k \notin K$,
3. *If* $k \in K$ *and* $G_K \neq 0$, *then* $G_K F_k x \neq 0$,
4. $F_k x = \sum_J \sum_K E_J G_K F_k x$,
5. $\sum_{m=1}^M E_m F_m x = \sum_J \sum_K \sum_{\{m\in J\cap K\}} E_J G_K F_m x$,
6. *For a fixed* K, *there are most* N *integers* m *for which* $G_k F_m x \neq 0$.

Proof. $1-4$ are clear. 5 follows from the fact that the E_J's and G_K's have sum I, and if $m \notin J$, then $E_J E_m = 0$; if $m \notin K$, then $G_K F_m z = 0$; while if $m \in J \cap K$, then $E_J G_K E_m F_m x = E_J G_K F_m x$.

6. Suppose that $G_K F_m x \neq 0$ for $m = m_1, \cdots, m_{N+1}$. Then by 2, $\{m_1, \cdots, m_{N+1}\} \subseteqq K$, and by 1, each $F_{m_n} x$ is full over G_K. Since $F_m z = z$ for every z in $\mathfrak{M}(F_m x)$, the disjointness of the F_m's gives

$$\mathfrak{M}(F_{m_i} x) \wedge \bigvee_{j\neq 1} \mathfrak{M}(F_{m_j} x) = 0$$

for $1 \leqq i \leqq N+1$, which contradicts $(*_N)$.

Now define

$$\alpha(m, J, K) = x^* E_J G_K F_m x .$$

As a corollary to Lemma 2.1, parts 5 and 6, we have

5a. $x^* \sum_{m=1}^M E_m F_m x = \sum_J \sum_K \sum_{\{m\in J\cap K\}} \alpha(m, J, K)$,

6a. *For a fixed* K, *there are at most* N *integers* m *for which* $\alpha(m, J, K) \neq 0$.

Let P be any subset of $\{1, \cdots, M\}$ and define

$$\beta(P, J, K) = \sum_{p\in P} \alpha(p, J, K) = x^*\Big(\sum_{p\in P} F_p\Big) E_J G_K .$$

Let T_P be the operator $\sum_J \sum_K \overline{\operatorname{sgn}} \beta(P, J, K) E_J G_K$, where $\overline{\operatorname{sgn}}\, re^{i\theta} = e^{-i\theta}$ if $r \neq 0$, and 0 if $r = 0$. T_P is an operator on X of norm at most $4M_1$.

Thus we have

$$\left| x^*\Big(\sum_{p\in P} F_p\Big) T_p x \right| \leqq | x^* | \left| \sum_{p\in P} F_p \right| | T_p | | x | \leqq 4M_1M_2 ;$$

but on the other hand

$$\begin{aligned} (2.1) \qquad x^*\Big(\sum_{p\in P} F_p\Big) T_P x &= \sum_J \sum_K \Big[\overline{\operatorname{sgn}}\, \beta(P, J, K) \cdot x^*\Big(\sum_{p\in P} F_p\Big) E_J G_K x \Big] \\ &= \sum_J \sum_K | \beta(P, J, K) | \leqq 4M_1M_2 . \end{aligned}$$

We are now in a position to prove the principal theorem of this paper.

THEOREM 2.2. *Let $\mathscr{E}$ and $\mathscr{F}$ be commuting bounded Boolean algebras of projections on a Banach space with bounds M_1 and M_2 respectively, $\mathscr{E}$ strongly complete. Suppose condition $(*N)$ is satisfied for some N. Then the Boolean algebra of projections generated by $\mathscr{E}$ and $\mathscr{F}$ is bounded, with bound $8\sqrt{N\pi}M_1M_2$.*

Proof. For each J, K, there are at most N integers $m_1, \cdots, m_N$ for which $\alpha(m, J, K) \neq 0$. Let

$$\alpha_s = \alpha(m_s, J, K) , \quad 1 \leqq s \leqq N ,$$
$$S_0 = \{ s \mid m_s \in J \cap K \} ,$$

and apply Theorem 1.1. We obtain

$$\left| \sum_{m\in J\cap K} \alpha(m, J, K) \right| \leqq 2\sqrt{N\pi} \cdot 2^{-N} \sum_{S\in\mathscr{S}} \left| \sum_{s\in S} \alpha(m_s, J, K) \right| .$$

Now for any S, there are 2^{M-N} distinct sets P of $\{1, \cdots, M\}$ for which $\sum_{s\in S}\alpha(m_s, J, K) = \sum_{p\in P}\alpha(p, J, K)$; namely, $\{m_s \mid s \in S\}$ together with any of the 2^{M-N} subsets of integers between 1 and M which are not one of $m_1, \cdots, m_N$. Thus

$$2^{M-N} \sum_{S\in\mathscr{S}} \left| \sum_{s\in S} \alpha(m_s, J, K) \right| = \sum_P \left| \sum_{p\in P} \alpha(p, J, K) \right| ,$$

and

$$\left| \sum_{m\in J\cap K} \alpha(m, J, K) \right| \leqq 2\sqrt{N\pi} \cdot 2^{-M} \sum_P \left| \sum_{p\in P} \alpha(p, J, K) \right| ,$$

Summing over all J, K, we have for arbitrary x, x^* of norm 1, E_m's and F_m's,

$$\begin{aligned}\left|x^* \sum_{m=1}^{M} E_m F_m x\right| &\leqq \sum_J \sum_K \left|\sum_{m \in J \cap K} \alpha(m, J, K)\right| \\ &\leqq 2\sqrt{N\pi}\cdot 2^{-M} \sum_P \sum_J \sum_K \left|\sum_{p \in P} \alpha(p, J, K)\right| \\ &\leqq 2\sqrt{N\pi}\cdot 2^{-M} \sum_P (4M_1M_2) = 2\sqrt{N\pi}\cdot 4M_1M_2 ,\end{aligned}$$

which is exactly our theorem.

3. Examples. Inspired by the example of S. Kakutani [6], we construct an example in a finite dimensional space to show that the order of magnitude of our bound is sharp. We imitate his paper in the construction of algebras of projections as much as possible and omit proofs which essentially appear in his paper.

Let N be a power of 2, $N = 2^n$, and let S and S' be the set of integers $\{1, \cdots, N\}$; $C(S)$, the continuous functions on S with the sup norm, is simply the N dimensional vector space of N-tuples. Let $S^* = S \times S'$, and let our Banach space X be $C(S^*)$, but with the minimal cross product norm induced from $C(S)$ and $C(S')$. Our X corresponds to the space $C(S) \circledast C(S')$ of Kakutani, and has dimension N^2. The elements of X may be thought of in a natural way as $N \times N$ matrices $x(s, s')$. Let $\mathscr{E}_N$ and $\mathscr{F}_N$ be the commuting Boolean algebras of projections of bound 1 generated respectively by E_i and F_i, $1 \leqq i \leqq N$, both of multiplicity N:

$$E_i x(s, s') = \begin{cases} x(s, s') \text{ if } s = i , \\ 0 \text{ if } s \neq i , \end{cases} \qquad E_i x(s, x') = \begin{cases} x(s, s') \text{ if } s' = i , \\ 0 \text{ if } s' \neq i . \end{cases}$$

Then there is a projection G in the Boolean algebra of projections generated by $\mathscr{E}_N$ and $\mathscr{F}_N$ such that $2G - I$ takes the element of X defined by $x(s, s') \equiv 1$ into the element $\rho(s, s')$ defined by

$$\rho(s, s') = (-1) \sum_{i=1}^{n} \varepsilon_i(s)\varepsilon_i(s')$$

where s has the unique representation

$$s = \varepsilon_1(s)2^{n-1} + \delta_2(s)2^{n-2} + \cdots + \varepsilon_{n-1}(s)2 + \varepsilon_n(s) + 1, \varepsilon_i(s) = 0 \quad \text{or} \quad 1 .$$

If we put a measure μ on S which assigns to each point the measure $1/N$, then the N functions on S, $\rho(s, i)$, $1 \leqq i \leqq N$, form an orthonormal base for $L^2(S, \mu)$, and the computations on pp. 368 and 369 of [6] carry over exactly to show that the norm of $\rho(s, s')$ in X is no less than $\sqrt{N}$. Since the element of X, $x(s, s')$, has norm 1, this says that the norm of $2G - I$ is at least $\sqrt{N}$, or that the norm of G is at least $\frac{1}{2}(\sqrt{N} - 1)$.

Let us now take one copy X_N of the above example for each N, and form the l_2 direct sum of the X_N, which we call X. Elements of X are sequences $\{x_N\}$ where $x_N \in X_N$ and

$$\|\{x_N\}\| = \left[\sum_{N=1}^{\infty} \|x_N\|_{X_N}^2\right]^{1/2} < \infty .$$

The algebras $\mathscr{E}_N$ and $\mathscr{F}_N$ on X_N have a natural extension to all of X by defining $\mathscr{E}_N(X_M) = \mathscr{F}_N(X_M) = 0$, $M \neq N$. Let $\mathscr{E}$ and $\mathscr{F}$ be respectively the commuting Boolean algebras of bound 1 of projections on X generated by all the $\mathscr{E}_N$, resp. $\mathscr{F}_N$, and note that the generated algebra contains a projection of norm at least $\frac{1}{2}(\sqrt{N} - 1)$ on the subspace X_N; we thus see that the algebra generated by $\mathscr{E}$ and $\mathscr{F}$ is not bounded. Since X is an l_2 direct sum of finite dimensional (hence reflexive) spaces, X must be itself reflexive and also separable.

Now let T and T' be operators on X, defined by

$$T\left(\sum_{N=1}^{\infty} \circledast\, x_N(s, s')\right) = \sum_{N=1}^{\infty} \circledast\, 2^{-N}3^{-s}x(s, s')$$

$$T'\left(\sum_{N=1}^{\infty} \circledast\, x_N(s, s')\right) = \sum_{N=1}^{\infty} \circledast\, 5^{-s'}x_N(s, s') .$$

Then T and T' are bounded commuting scalar-type spectral operators on X. The operator TT' has simple eigenvalues at the distinct points $2^{-M}3^{-i}5^{-j}$, $1 \leqq i, j \leqq M < \infty$. The projection $E_{M,i,j}$ corresponding to the eigenvalue $2^{-M}3^{-i}5^{-j}$ satisfies

$$E_{M,i,j}\left(\sum_{N=1}^{\infty} \circledast\, x_N(s, s')\right) = \sum_{N=1}^{\infty} \circledast\, \delta_{MN}\delta_{is}\delta_{js'}x_N(s, s') ,$$

where δ_{ij} is the Kronecker delta. Thus the Boolean algebra of projections generated by the $E_{M,i,j}$ contains both $\mathscr{E}$ and $\mathscr{F}$, and therefore is unbounded. TT' cannot be spectral. Also the sum of two spectral operators on X need not always be spectral. For if this were so, $T + T'$ would be spectral, hence $(T + T')^2$; also $(T + T')^2 - T'^2 = 2TT'$.

4. **Appendix.** We show that $(*_N)$ is satisfied if the Boolean algebra $\mathscr{E}$ is countably decomposable and has no projection of infinite multiplicity. We will make use of the representation theory of such algebras of projections originally given by J. Dieudonné [3] but used here in the form due to W. G. Bade [2]:

There is a compact Hausdorff space Ω, the Stone space for $\mathscr{E}$, and a natural correspondence between $\mathscr{E}$ and the Boolean algebra of Borel sets of Ω. We will allow ourselves to confuse the set $\sigma \subset \Omega$ with the corresponding projection $E(\sigma)$ in $\mathscr{E}$. A projection E *has multiplicity* N if there exist N elements $x_1, \cdots, x_N$ of X such that $EX = \bigvee_{n=1}^{N}\mathfrak{M}(x_n)$, and if for every $N-1$ elements $y_1, \cdots, y_{N-1}$ of X, $EX \neq \bigvee_{n=1}^{N-1}\mathfrak{M}(y_n)$. E *has uniform multiplicity* N if E has multiplity N, and $0 < E_1 \leqq E$ implies that E_1 has multiplicity N. By using theorem of Bade [2, Theorem 3.4], and assuming that $\mathscr{E}$ contains no projection of infinite multiplicity, we can decompose Ω into a finite union of disjoint sets, $\Omega = e_1 \cup \cdots \cup e_N$ for some N, where e_n has uniform multiplicity n. It will suffice to consider the case $\Omega = e_N$. In this case, we can find an $\mathscr{E}$-basis $x_1, \cdots, x_N$ for X and a dual basis $x_1^*, \cdots, x_N^*$ such that $X = \bigvee_{n=1}^N \mathfrak{M}(x_n)$ and $x_m^* E(\sigma)x_n = 0$ if $m \neq n$ and is > 0 if $m = n$ and $E(\sigma)x_n \neq 0$. Let us write $\mu(x^*, x)$ for the measure $x^*E(\cdot)x$. Then each x in X determines, essentially uniquely, N scalar functions $f_n(\omega)$ on Ω, $f_n(\omega)$ being the Radon-Nikodým derivative of $\mu(x_n^*, x)$ with respect to $\mu(x_n^*, x_n)$. Also each x^* in X^* determines, essentially uniquely, N scalar functions $g_n(\omega)$ on Ω, $g_n(\omega)$ being the Radon-Nikodým derivative of $\mu(x^*, x_n)$ with respect to $\mu(x_n^*, x_n)$. The product $f_n g_n$ is in $L^1(\Omega, \mu(x_n^*, x_n))$ for each n, and $x^*x = \sum_{n=1}^N \int f_n(\omega)g_n(\omega)d\mu(x_n^*, x_n)$.

Note that the measures $\mu(x_n^*, x_n)$ are all absolutely continuous with respect to one another, and every measure $\mu(x^*, x)$ is absolutely continuous with respect to all of the $\mu(x_n^*, x_n)$. When we say measurable, we mean with respect to any, hence all, $\mu(x_n^*, x)$.

Now suppose that $F_1, \cdots, F_{N+1}$ are disjoint projections, commuting with each $E \in \mathscr{E}$, and such that for some x and some $\sigma \subset \Omega$, $\sigma \neq 0$, each $F_n x$ is full over σ. We can assume for simplicity that $\sigma = \Omega$. The fact that each F is a bounded projection commuting with every E in $\mathscr{E}$, insures that $F_z = z$ for every z in $\mathfrak{M}(Fx)$. The disjointness of the F_n's then gives us $\mathfrak{M}(F_n x) \wedge \bigvee_{i \neq n} \mathfrak{M}(F_i x)$ for $n = 1, \cdots, N+1$.

The following two lemmas will allow us to reach a contradiction.

LEMMA 4.1. *Let $A(\omega)$ be a matrix of measurable functions on Ω. Then if $M(\omega)$ is a fixed minor of $A(\omega)$, det $M(\omega)$ is a measurable function. If $r(A, \omega)$ denotes the rank of $A(\omega)$, then $r(A, \omega)$ is a measurable function.*

Proof. If $M(\omega)$ is a fixed minor of $A(\omega)$, $\det M(\omega)$ is a sum of products of measurable functions, hence is measurable. Also the set on which $\det M(\omega) \neq 0$ is measureable, and so the Boolean algebra of sets generated by the supports of $M(\omega)$ for all minors M of A, is an algebra of measurable sets. $r(A, \omega)$ is a simple function on this algebra, and so is measurable.

$\sigma(r_0, A)$ will denote the set of ω for which $r(A, \omega) = r_0$. $\sigma(r_0, A, M)$ will denote the subset of $\sigma(r_0, A)$ for which the r_0-rowed minor M has non-zero determinant. The $\sigma(r_0, A, M)$ mutually exhaust $\sigma(r_0, A)$. Let $\{\sigma\}$ be a finite collection of mutually disjoint Borel sets such that each σ is contained in some $\sigma(r, A, M)$, and mutually exhaust $\sigma(r, A)$ and hence exhaust Ω.

For the moment, fix σ. Let M be a r-rowed minor of $A(\omega)$ for which $\sigma \subset \sigma(r, A, M)$. Let $p_1, \cdots, p_r$ be the row indices of M and $q_1, \cdots, q_r$ the column indices.

LEMMA 4.2. *Let $g_1(\omega), \cdots, g_N(\omega)$ be N measurable functions such that on σ, the column N-tuple $(g_1(\omega), \cdots, g_N(\omega)$ is pointwise linearly dependent upon the r columns $(a_{1,q_j}(\omega), \cdots, a_{N,q_j}(\omega))$ of $A(\omega)$. Then there exist r measurable functions $u_j(\omega)$ such that on σ,*

$$g_n(\omega) = \sum_{j=1}^{r} u_j(\omega) a_{n,q_j}(\omega) \quad \text{for } n = 1, \cdots, N .$$

Proof. The minor $M(\omega)$ has non-zero determinant on. Let $M^{-1}(\omega) = (w_{p_i,q_j}(\omega))$, the w's being measurable functions on Ω. We have

$$\sum_{j=1}^{r} a_{p_i,q_j}(\omega) w_{p_j,q_k}(\omega) \equiv \delta_{ik} .$$

Define

$$u_j(\omega) = \sum_{i=1}^{r} w_{p_j,q_i}(\omega) \cdot g_{p_i}(\omega)$$

Then, if n is one of the p_i, we have

$$\begin{aligned}\sum_{j=1}^{r} u_j(\omega) a_{n,q_j}(\omega) &= \sum_{j=1}^{r}\sum_{i=1}^{r} w_{p_j,q_i}(\omega) a_{n,q_j}(\omega) g_{p_i}(\omega) \\ &= \sum_{i=1}^{r} \delta_{n,p_i} g_{p_i}(\omega) = g_n(\omega) .\end{aligned}$$

And if for some ω_0 and some n_0 not a p_i,

$$\sum_{j=1}^{r} u_j(\omega_0) a_{n_0,q_j}(\omega_0) \neq g_{n_0}(\omega_0) ,$$

then the matrix, evaluated at ω_0,

$$\begin{pmatrix} a_{p_1,q_1} \cdots a_{p_1,q_r} & g_{p_1} \\ \cdots \qquad \cdots & \cdots \\ a_{p_r,q_1} \cdots a_{p_r,q_r} & g_{p_r} \\ a_{n_0,q_1} \cdots a_{n_0,q_r} & g_{n_0} \end{pmatrix} ,$$

has rank $r+1$, contrary to the assumption that the g_n are linearly dependent upon the r columns of A with indices q_j.

Now let the matrix A have its entires defined by

$$a_{ij}(\omega) = \frac{d\mu(x_i * F_j x)}{d\mu(x_i * x_i)} , \quad 1 \leq i \leq N, \quad 1 \leq j \leq N+1 .$$

Then the $N+1$st column is pointwise linearly dependent upon the first N columns. Selecting one of the non-zero sets σ and applying Lemma 6.2, we have the existence of N measurable functions $u_j(\omega)$ on σ for which we have

$$a_{i,N+1}(\omega) = \sum_{j=1}^{N} u_j(\omega) a_{ij}(\omega) , \quad 1 \leqq i \leqq N .$$

Let now $\tau \neq 0$ be a subset of σ on which each of the functions $u_j(\omega)$ is bounded. We then have

$$x_i^* E(\tau) F_{N+1} x = x_i^* \sum_{j=1}^{N} \int_\tau u_j(\omega) E(d\omega) F_j x$$

which implies

$$E(\sigma) F_{N+1} x = \sum_{j=1}^{N} \left(\int_\tau u_j(\omega) E(d\omega) \right) F_j x$$

(this makes sense all the u_j's are bounded on τ); that is,

$$E(\tau) F_{N+1} x \in \bigvee_{j=1}^{N} \mathfrak{M}(F_j x) ,$$

which is the desired contradiction.

References

1. W. G. Bade, *On Boolean algebras of projections and algebras of opertors*, Trans. AMS, **80** (1955).

2. ———, *A multiplicity theory for Boolean algebras of projections on Banach spaces*, Trans. AMS, **92** (1959).

3. J. Dieudonné, *Sur la théorie spectrale*, J. Math. Pures Appl. (9), **35** (1956).

4. N. Dunford, *Spectral operators*, Pacific J. Math, **4** (1954).

5. S. R .Foguel, *Sums and products of commuting spectral operator*, Arkiv för Mat., **3** (1957).

6. S. Kakutani, *An example concerning uniform boundedness of spectral measures*, Pacific J. Math., **4** (1954).

7. J. Wermer, *Commuting spectral measures on Hilbert space*, Pacific J. Math., **4** (1954).

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

TRANSFORMATIONS OF SERIES OF E-FUNCTIONS

T. M. MACROBERT

1. Introductory. The transformation [1, p. 25, 2, p. 369]

$$(1)\qquad F\begin{pmatrix}\alpha, & \beta, & \gamma, & \delta, & -l; & 1\\ \alpha-\beta+1, & \alpha-\gamma+1, & \alpha-\delta+1, & \alpha+l+1 & & \end{pmatrix}$$
$$=\frac{(\alpha+1;l)\left(\frac{1}{2}\alpha-\beta+1;l\right)}{\left(\frac{1}{2}\alpha+1;l\right)(\alpha-\beta+1;l)}F\begin{pmatrix}\alpha-\gamma-\delta+1, \frac{1}{2}\alpha, \beta, -l; 1\\ \alpha-\gamma+1, \alpha-\delta+1, \beta-\frac{1}{2}\alpha-l\end{pmatrix},$$

where l is a positive integer, is a special case of a formula of Whipple's. It, and other transformations of the same kind, can be employed to obtain transformations of series of E-functions. Two such transformations are:

$$\sum_{n=0}^{l}\frac{(\alpha;n)(\beta;n)(-l;n)}{n!(\alpha-\beta+l;n)(\alpha+l+1;n)}E\left\{\begin{matrix}p;\alpha_r\\ \hline \Delta(m;\rho-n),\Delta(m;\sigma-n),\Delta(m;\alpha+\rho+n)\\ \Delta(m;\alpha+\sigma+n),\rho_1,\rho_2,\cdots,\rho_q\end{matrix}:z\right\}$$
$$(2)\qquad =\frac{(\alpha+1;l)\left(\frac{1}{2}\alpha-\beta+1;l\right)}{\left(\frac{1}{2}\alpha+1;l\right)(\alpha-\beta+1;l)}\sum_{n=0}^{l}\frac{\left(\frac{1}{2}\alpha;n\right)(\beta;n)(-l;n)}{n!\left(\beta-\frac{1}{2}\alpha-l;n\right)}\left(\frac{2}{m}\right)^n$$
$$\times E\left\{\begin{matrix}\Delta(2m;\alpha+\rho+\sigma+n-1),\alpha_1,\alpha_2,\cdots,\alpha_p\\ \hline \Delta(2m;\alpha+\rho+\sigma-1),\Delta(m;\rho),\Delta(m;\sigma),\\ \Delta(m;\alpha+\rho+n),\Delta(m;\alpha+\sigma+n),\rho_1,\rho_2,\cdots,\rho_q\end{matrix}:z\right\};$$

$$\sum_{n=0}^{l}\frac{(\alpha;n)(\beta;n)(-l;n)}{n!(\alpha-\beta+1;n)(\alpha+l+1;n)}E\left\{\begin{matrix}\Delta(m;\gamma+n),\Delta(m;\gamma-\alpha-n),\alpha_1,\cdots,\alpha_p\\ \Delta(m;\sigma+n),\Delta(m;\sigma-\alpha-n),\rho_1,\cdots,\rho_q\end{matrix}:z\right\}$$
$$(3)\qquad =\frac{(\alpha+1;l)\left(\frac{1}{2}\alpha-\beta+1;l\right)}{\left(\frac{1}{2}\alpha+1;l\right)(a-\beta+1;l)}\sum_{n=0}^{l}\frac{(\sigma-\gamma;n)\left(\frac{1}{2}\alpha;n\right)(\beta;n)(-l;n)}{n!\left(\beta-\frac{1}{2}\alpha-l;n\right)(-m^2)^n}$$
$$\times E\left\{\begin{matrix}\Delta(m;\gamma),\Delta(m;\gamma-\alpha-n),\alpha_1,\cdots,\alpha_p\\ \Delta(m;\sigma-\alpha),\Delta(m;\sigma+n),\rho_1,\cdots,\rho_q\end{matrix}:z\right\}.$$

In these formulae m is a positive integer,

$$(4)\qquad (\alpha;0)=1,\ (\alpha;m)=\alpha(\alpha+1)\cdots(\alpha+m-1),$$

and $\Delta(m;\alpha)$ denotes the set of parameters

Received February 23, 1960.

$$\frac{\alpha}{m}, \; \frac{\alpha+1}{m}, \; \cdots, \; \frac{\alpha+m-1}{m}.$$

The proofs of (2) and (3) are given in § 2. The following formulae are required.

If m is a positive integer

$$(5) \quad \Gamma(mz)=(2\pi)^{[(1/2)-(1/2)m]}m^{[mz-(1/2)]}\Gamma(z)\Gamma\left(z+\frac{1}{m}\right)\cdots\Gamma\left(z+\frac{m-1}{m}\right).$$

From this it follows that, if m and n are positive integers,

$$(6) \quad \Gamma\left(\frac{\alpha+n}{m}\right)\Gamma\left(\frac{\alpha+1+n}{m}\right)\cdots\Gamma\left(\frac{\alpha+m-1+n}{m}\right)$$
$$=\Gamma\left(\frac{\alpha}{m}\right)\Gamma\left(\frac{\alpha+1}{m}\right)\cdots\Gamma\left(\frac{\alpha+m-1}{m}\right)m^{-n}(\alpha;n),$$

and

$$(7) \quad \Gamma\left(\frac{\alpha-n}{m}\right)\Gamma\left(\frac{\alpha+1-n}{m}\right)\cdots\Gamma\left(\frac{\alpha+m-1-n}{m}\right)$$
$$=\Gamma\left(\frac{\alpha}{m}\right)\Gamma\left(\frac{\alpha+1}{m}\right)\cdots\Gamma\left(\frac{\alpha+m-1}{m}\right)(-m)^{n}/(1-\alpha;n).$$

The Barnes' integral for the E-function is [2, p. 374]

$$(8) \quad E(p;\alpha_r;q;\rho_s:z)=\frac{1}{2\pi i}\int\frac{\Gamma(\zeta)\Pi\Gamma(\alpha_r-\zeta)}{\Pi\Gamma(\rho_s-\zeta)}z^{\zeta}d\zeta$$

where $|\operatorname{amp} z|<\pi$ and the integral is taken up the η-axis with loops, if necessary, to ensure that the origin lies to the left of the contour and the points $\alpha_1, \alpha_2, \cdots, \alpha_p$ to the right of the contour. Zero and negative integral values of the parameters are excluded, and the α's must not differ by integral values. When $p<q+1$ the contour is bent to the left at both ends. When $p>q+1$ the formula is valid for $|\operatorname{amp} z|<\frac{1}{2}(p-q+1)\pi$.

2. Proofs of the transformations. Using (8), (6) and (7), the left-hand side of (2), with $p=q=0$, can be written

$$\sum_{n=0}^{l}\frac{(\alpha;n)(\beta;n)(-l;n)}{n!(\alpha-\beta+1;n)(\alpha+l+1;n)}$$
$$\times\frac{1}{2\pi i}\int\frac{\Gamma(\zeta)z^{\zeta}(1-\rho+m\zeta;n)(1-\sigma+m\zeta;n)d\zeta}{\prod_{\rho,\sigma}\left[\prod_{u=0}^{m-1}\left\{\Gamma\left(\frac{\rho+u}{m}-\zeta\right)\Gamma\left(\frac{\alpha+\rho+u}{m}-\zeta\right)\right\}(\alpha+\rho-m\zeta;n)\right]}$$
$$=\frac{1}{2\pi i}\int\frac{\Gamma(\zeta)z^{\zeta}}{\prod_{\rho,\sigma}\left[\prod_{u=0}^{m-1}\left\{\Gamma\left(\frac{\rho+u}{m}-\zeta\right)\Gamma\left(\frac{\alpha+\rho+u}{m}-\zeta\right)\right\}\right]}Id\zeta\,,$$

where

$$I=F\begin{pmatrix}\alpha, & \beta, & 1-\rho+m\zeta, & 1-\sigma+m\zeta, & -l; & 1\\ \alpha-\beta+1, & \alpha+\rho-m\zeta, & \alpha+\sigma-m\zeta, & \alpha+l+1\end{pmatrix},$$

From (1) it follows that this can be put in the form

$$\frac{(\alpha+1;l)\left(\frac{1}{2}\alpha-\beta+1;l\right)}{\left(\frac{1}{2}\alpha+1;l\right)(\alpha-\beta+1;1)}$$

$$\times\frac{1}{2\pi i}\int\frac{\Gamma(\zeta)z^{\zeta}}{\prod_{\rho,\sigma}\left[\prod_{u=0}^{m-1}\left\{\Gamma\left(\frac{\rho+u}{m}-\zeta\right)\Gamma\left(\frac{\alpha+\rho+u}{m}-\zeta\right)\right\}\right]}J d\zeta,$$

where

$$J=F\begin{pmatrix}\alpha+\rho+\sigma-1-2m\zeta, \frac{1}{2}\alpha, \beta, -l; 1\\ \alpha+\rho-m\zeta, \alpha+\sigma-m\zeta, \beta-\frac{1}{2}\alpha-l\end{pmatrix}.$$

Now the integral is equal to

$$\sum_{n=0}^{l}\frac{\left(\frac{1}{2}\alpha;n\right)(\beta;n)(-l;n)}{n!\left(\beta-\frac{1}{2}\alpha-l;n\right)2\pi i}$$

$$\times\int\frac{\Gamma(\zeta)z^{\zeta}\Gamma(\alpha+\rho+\sigma-1+n-2m\zeta)}{\prod_{\rho,\sigma}\left[\sum_{u=0}^{m-1}\left\{\Gamma\left(\frac{\rho+u}{m}-\zeta\right)\Gamma\left(\frac{\alpha+\rho+u}{m}-\zeta\right)\right.\right.}\Bigg/$$

$$\Bigg/\frac{\Gamma(\alpha+\rho-m\zeta)\Gamma(\alpha+\sigma-m\zeta)d\zeta}{\left.\left.\Gamma(\alpha+\rho+n-m\zeta)\right\}\right]\Gamma(\alpha+\rho+\sigma-1-2m\zeta)},$$

and, on applying (5) to the gamma functions whose arguments contain $-2m\zeta$ or $-m\zeta$, the right-hand side of (2) with $p=q=0$ is obtained. Formula (2) can then be derived by generalising.

Formula (3) can be proved in the same way. It should be noted that

$$(\alpha-\gamma+1+m\zeta;n)=(-1)^n(\gamma-\alpha-n-m\zeta;n).$$

The restrictions on amp z and on the parameters can be removed by analytical continuation, provided that the functions exist.

References

1. W. N. Bailey, *Generalized Hypergeometric Series,* Cambridge 1935.
2. T. M. MacRobert, *Functions of a Complex Variable,* London 1954.

AN INEQUALITY FOR LOGARITHMIC CAPACITIES

HEINZ RENGGLI

1. Introduction. In his work on capacities, G. Choquet proved that for many capacities the inequality of strong subadditivity holds [1]. It is the purpose of this note to show that a similar inequality holds for logarithmic capacities. More precisely we shall prove the

THEOREM. *Let A and B be compact sets in the complex z-plane E. By $C(S)$ we denote the logarithmic capacity* [2] *of a given compact set $S, S \subset E$, where we agree to put $C(S) = 0$ whenever $S = \phi$. Then*

$$C(A \cup B) \cdot C(A \cap B) \leqq C(A) \cdot C(B) .$$

2. Proof of the theorem. Let $S, S \subset E$, be a compact set whose boundary consists of a finite number of analytic arcs. By S^* we denote that component of $E - S$ which is unbounded. Then Green's function of S^* is defined by the properties: it is harmonic in S^*, vanishes at the finite boundary points of S^* and has a logarithmic singularity at infinity. We will denote this function by $g_S(z, \infty)$.

First we shall deal with the case when the respective boundaries of A, B and $A \cap B$ consist of a finite number of non-degenerate analytic arcs. We remark that the difference $g_{A \cap B}(z, \infty) - g_A(z, \infty)$ is harmonic in $A^*, A^* \subset (A \cap B)^*$, and at infinity. It is furthermore non-negative on the boundary of A^* and hence non-negative in A^* by the maximum principle. Similarly $g_{A \cup B}(z, \infty) \geqq g_B(z, \infty)$ holds in $B^*, B^* \subset (A \cap B)^*$.

The function

$$h(z) = g_{A \cup B}(z, \infty) + g_{A \cap B}(z, \infty) - g_A(z, \infty) - g_B(z, \infty)$$

is harmonic in $(A \cup B)^*$ and at infinity. From $(A \cup B)^* = A^* \cap B^*$ it follows that the boundary points of $(A \cup B)^*$ belong either to the boundary of A^* or to the boundary of B^*. Therefore $g_{A \cup B}(z, \infty)$ and either $g_A(z, \infty)$ or $g_B(z, \infty)$ vanish at these boundary points. With the aid of the remark made above we get the result that $h(z)$ is non-negative in $(A \cup B)^*$.

Therefore

$$g_A(z, \infty) + g_B(z, \infty) \leqq g_{A \cup B}(z, \infty) + g_{A \cap B}(z, \infty)$$

holds in $(A \cup B)^*$. From this general inequality and using the fact that

Received December 16, 1959. Presented to the Amer. Math. Soc. in Jan. 1958, where a less elementary proof based on an inequality for extremal lengths was given.

$$\lim_{z\to\infty} \{g_S(z, \infty) - \log |z|\}$$

is the constant $\gamma(S)$ of Robin [2] we deduce

$$\gamma(A) + \gamma(B) \leqq \gamma(A \cup B) + \gamma(A \cap B) .$$

But

$$C(S) = \exp \{-\gamma(S)\}$$

by definition. Hence our theorem is proven for the special case.

The general case follows by the usual approximation techniques [2].

REFERENCES

1. G. Choquet, *Theory of capacities,* Ann. Inst. Fourier, Grenoble **5** (1953–54), 131–295.
2. R. Nevanlinna, *Eindeutige analytische Funktionen,* Grundlehren 46, 2. Aufl., Springer, 1953.

RUTGERS UNIVERSITY

APPLICATIONS OF THE SUBORDINATION PRINCIPLE TO UNIVALENT FUNCTIONS

M. S. ROBERTSON

1. Introduction. Let

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots \tag{1.1}$$

be regular and univalent in $|z| < 1$ and map $|z| < 1$ onto a simply-connected domain D. Let

$$\phi(z) = b_1 z + b_2 z^2 + \cdots + b_n z^n + \cdots \tag{1.2}$$

also be regular in $|z| < 1$. $\phi(z)$ is said to be subordinate to $f(z)$ if for each z of the unit circle $|z| < 1$ the corresponding point $w = \phi(z)$ lies in the domain D. In this case [2] there exists an analytic function $\omega(z)$ regular in $|z| < 1$ for which $\omega(0) = 0$, $|\omega(z)| \leqq |z| < 1$ and $\phi(z) \equiv f\{\omega(z)\}$.

It is the purpose of this paper to establish the following basic Theorems A and B which concern analytic functions $F(z, t)$ and $\omega(z, t)$, depending upon a real parameter t, and then to use them to obtain results in the theory of univalent functions. Some of the results are well known and others are new, but the method of attack seems to be novel, simple and of sufficient generality to be of interest in itself. The functions $F(z, t)$ and $\omega(z, t)$ will be related to the univalent function $f(z)$ of (1.1) by means of the subordination concept.

An interesting biproduct of Theorem B is the following statement. A sufficient condition that $f(z)$, regular and univalent in $|z| < 1$, be convex in $|z| < 1$ is that the de la Vallée Poussin means $V_n(z)$ of (1.1) be subordinate to $f(z)$ in $|z| < 1$ for $n = 1, 2, \cdots$. Recently [3] G. Pólya and I. J. Schoenberg showed that this condition for convexity is also necessary.

THEOREM A. *Let*

$$\omega(z, t) = \sum_1^\infty b_n(t) z^n \tag{1.3}$$

be regular in $|z| < 1$ *for* $0 \leqq t \leqq 1$. *Let*

$$|\omega(z, t)| < 1 \text{ for } |z| < 1,\ 0 \leqq t \leqq 1,\ \omega(z, 0) \equiv z\,.$$

Let ρ *be a positive real number for which*

$$\omega(z) = \lim_{t \to 0+} \left\{ \frac{\omega(z, t) - z}{z t^\rho} \right\} \tag{1.4}$$

Received October 2, 1959.

exists. Then

$$\mathscr{R}\omega(z) \leqq 0 \text{ for } |z| < 1 . \tag{1.5}$$

If $\omega(z)$ is also analytic in $|z| < 1$ and $\mathscr{R}\omega(0) \neq 0$, then

$$\mathscr{R}\omega(z) < 0 \text{ for } |z| < 1 .$$

Proof. By Schwarz' lemma we have for $|z| < 1$ $|\omega(z, t)| \leqq |z|$ with equality only if $\omega(z, t) = z \exp i\theta(t)$, then the function

$$\mu(z, t) = \frac{\omega(z, t) - z}{\omega(z, t) + z} \tag{1.6}$$

is regular and $\mathscr{R}\mu(z, t) < 0$ for $|z| < 1$. But when $\omega(z, t) = z \exp i\theta(t)$, $\mu(z, t) = i \tan(1/2\theta(t))$ is purely imaginary. Thus $\mu(z, t)$ is regular and $\mathscr{R}\mu(z, t) \leqq 0$ in $|z| < 1$ with equality occurring only if $\omega(z, t) = z \exp i\theta(t)$.

For $t > 0$, $|z| < 1$ we may write

$$\mathscr{R}\left\{\frac{\omega(z, t) - z}{zt^\rho} \cdot \frac{2z}{\omega(z, t) + z}\right\} = \mathscr{R}\left\{\frac{2\mu(z, t)}{t^\rho}\right\} \leqq 0 . \tag{1.7}$$

(1.4) implies that $\lim_{t\to 0+}\omega(z, t) = z = \omega(z, 0)$. Therefore, on letting $t \to 0$ in (1.7) we obtain $\mathscr{R}\omega(z) \leqq 0$ for $|z| < 1$. When $\omega(z)$ is also analytic in $|z| < 1$ and $\mathscr{R}\omega(0) \neq 0$ we have further that $\mathscr{R}\omega(z) < 0$ in $|z| < 1$. This follows since the maximum, in this case zero, of a non-constant harmonic function cannot occur at an interior point.

As an illustration of Theorem A, the following example is useful. Let

$$\omega(z, t) = \frac{(1 - 2t)z + z^2}{1 + (1 - 2t)z} , \qquad 0 \leqq t \leqq 1 . \tag{1.8}$$

Then $\omega(z, 0) \equiv z$, $|\omega(z, t)| \leqq 1$ in $|z| < 1, 0 \leqq t \leqq 1$.

$$\omega(z) = \lim_{t\to 0}\left\{\frac{\omega(z, t) - \omega(z, 0)}{zt}\right\} = \frac{1}{z}\left[\frac{\partial}{\partial t}\omega(z, t)\right]_{t=0} = 2\frac{z - 1}{z + 1} , \tag{1.9}$$

$$\mathscr{R}\omega(z) = 2\mathscr{R}\left(\frac{z - 1}{z + 1}\right) < 0 , \quad |z| < 1 . \tag{1.10}$$

Theorem A is a special case of Theorem B to follow. However, the proof of Theorem B depends upon Theorem A.

THEOREM B. *Let*

$$f(z) = z + a_2z^2 + \cdots + a_nz^n + \cdots \tag{1.11}$$

be regular and univalent in $|z| < 1$. For $0 \leqq t \leqq 1$ let $F(z, t)$ be regular in $|z| < 1$. Let $F(z, 0) \equiv f(z)$ and $F(0, t) \equiv 0$: Let ρ be a positive *real number for which*

$$F(z) = \lim_{t \to 0+} \left\{ \frac{F(z, t) - F(z, 0)}{zt^\rho} \right\} \tag{1.12}$$

exists. Let $F(z, t)$ be subordinate to $f(z)$ in $|z| < 1$ for $0 \leqq t \leqq 1$. Then

$$\mathscr{R}\left\{ \frac{F(z)}{f'(z)} \right\} \leqq 0 \, , \quad |z| < 1 \, . \tag{1.13}$$

If in addition $F(z)$ is also analytic in $|z| < 1$ and $\mathscr{R}F(0) \neq 0$, then

$$\mathscr{R}\left\{ \frac{f'(z)}{F(z)} \right\} < 0 \, , \quad |z| < 1 \, .$$

Proof. Since $F(z, t)$ is subordinate to $f(z)$ in $|z| < 1$ we have

$$F(z, t) = f\{\omega(z, t)\} \, , \quad |z| < 1 \, , \quad 0 \leqq t \leqq 1 \, ,$$

where $\omega(z, t)$ is regular and bounded $|\omega(z, t)| \leqq 1$ in $|z| < 1$, $0 \leqq t \leqq 1$. Since $F(z, 0) \equiv f(z)$ and since $f(z)$ is univalent in $|z| < 1$ we have $\omega(z, 0) \equiv z$. Also since $f(0) = 0$, $F(0, t) = 0$ and since $f(z)$ is univalent we have $\omega(0, t) = 0$. We now write

$$\frac{F(z, t) - F(z, 0)}{zt^\rho} = \left[\frac{f(\omega(z, t)) - f(\omega(z, 0))}{\omega(z, t) - \omega(z, 0)} \right] \left[\frac{\omega(z, t) - \omega(z, 0)}{zt^\rho} \right] . \tag{1.14}$$

(1.12) implies that $F(z, t)$ is continuous from the right at $t = 0$ and a similar statement holds for $\omega(z, t)$ because of the subordination. Let $t \to 0+$ in (1.14). The left side of equation (1.14) has for a limit $F(z)$ by (1.12). On the right side of (1.14) the square bracket has a limit $f'(z) \neq 0$. Thus

$$\omega(z) = \lim_{t \to 0+} \left[\frac{\omega(z, t) - \omega(z, 0)}{zt^\rho} \right] \tag{1.15}$$

exists and equals $F(z)/f'(z)$. Furthermore $\mathscr{R}\omega(0) = \mathscr{R}F(0)$. If $F(z)$ is analytic so is $\omega(z)$. Since the conditions of Theorem A are fulfilled by $\omega(z, t)$ we have

$$\mathscr{R}\frac{F(z)}{f'(z)} = \mathscr{R}\omega(z) \leqq 0 \, , \quad |z| < 1 \, . \tag{1.16}$$

When $F(z)$ is analytic in $|z| < 1$ and $\mathscr{R}F(0) \neq 0$ we also have

$$\mathscr{R}\left\{ \frac{f'(z)}{F(z)} \right\} < 0 \, , \quad |z| < 1 \, . \tag{1.17}$$

2. Applications to univalent functions. The properties of univalent functions $W = f(z)$, given by (1.1), which are also star-like with respect to the origin in $|z| < 1$ are well-known [2]. If $W = f(z)$ maps $|z| < 1$ onto a star-like domain D of the W-plane, then by definition the line segment joining the origin to the point $W = f(z)$ lines entirely within D for each z in $|z| < 1$. One then shows that it is necessary that

$$(2.1) \qquad \mathscr{R}\left\{\frac{zf'(z)}{f(z)}\right\} > 0 \text{ in } |z| < 1 .$$

In establishing (2.1) one is obliged to show first that if $f(z)$ is star-like with respect to the unit circle it is also star-like with respect to each smaller circle $|z| = r < 1$. At this stage one then appeals to an alternative definition of a star-like domain. This requires that the radius vector, joining the origin to the point $f(z)$, turns always in one direction as the argument of z advances.

A much simpler proof of the necessity of (2.1) follows immediately from Theorem B. Since $(1 - t)f(z)$ is subordinate to $f(z)$ for $0 \leqq t \leqq 1$, we have

$$(2.2) \qquad (1 - t)f(z) = f\{\omega(z, t)\}$$

where $\omega(z, t)$ satisfies the conditions of Theorem A. Taking $\rho = 1$ and letting

$$(2.3) \qquad F(z, t) = (1 - t)f(z)$$

in Theorem B we obtain at once $F(z) = -f(z)/z \neq 0$, so that (2.1) follows from (1.17) very simply.

More generally we have the following theorem.

THEOREM 1. *Let.*

$$f(z) = z + a_2z^2 + \cdots + a_nz^n + \cdots$$

be regular and univalent in $|z| < 1$ *and such that* $(1 - te^{i\alpha})f(z)$ *is subordinate to* $f(z)$ *in* $|z| < 1$ *for an interval* $0 \leqq t \leqq t_0$, α *a real constant* $|\alpha| < \pi/2$, *then*

$$(2.4) \qquad \mathscr{R}\left\{e^{-i\alpha}\frac{zf'(z)}{f(z)}\right\} > 0 , \quad |z| < 1 .$$

For the proof of Theorem 1 we take

$$(2.5) \qquad F(z, t) = (1 - te^{i\alpha})f(z)$$

in Theorem B and (1.13) becomes (2.4) in this case. The condition (2.4) is the one given for spiral-like functions by L. Špaček [7].

The following theorem from an intuitive point of view appears to be almost self-evident. Our new technique, however, furnishes an easy and precise proof.

THEOREM 2. *Let $f(z)$ of (1.1) be regular and univalent in $|z| < 1$. For an interval $0 \leqq t \leqq t_0$ let the function*

$$\frac{1}{2}[f(e^{it}z) + f(e^{-it}z)] \tag{2.6}$$

be subordinate to $f(z)$ in $|z| < 1$. Then.

$$\mathscr{R}\left\{1 + \frac{zf''(z)}{f'(z)}\right\} > 0\ , \quad |z| < 1\ , \tag{2.7}$$

and $f(z)$ is convex in $|z| < 1$.

Proof. In Theorem B we choose $\rho = 2$ and $F(z, t)$ to be the function (2.6). Then

$$F(z) = \lim_{t\to 0} \frac{F(z, t) - F(z, 0)}{zt^2} = \lim_{t\to 0} \frac{1}{2zt}\frac{\partial F(z, t)}{\partial t} \tag{2.8}$$

$$F(z) = \lim_{t\to 0} \frac{1}{2z}\frac{\partial^2 F(z, t)}{\partial t^2} = -zf''(z) - f'(z)\ .$$

Since $f'(0) = 1$, it follows that $F(0) = -1$ so that $\mathscr{R}F(0) \neq 0$. Thus (1.17) of Theorem B is equivalent to

$$\mathscr{R}\left\{1 + z\frac{f''(z)}{f'(z)}\right\} > 0\ , \quad |z| < 1\ . \tag{2.9}$$

It is well known [2] that (2.9) implies that $f(z)$ is convex in $|z| < 1$.

For odd functions and an appropriate choice of $F(z, t)$ we obtain a result perhaps not so intuitively obvious as Theorem 2. It is the following theorem.

THEOREM 3. *Let*

$$f(z) = z + \sum_2^\infty a_{2n-1}z^{2n-1}\ , \tag{2.10}$$

be an odd function, regular and univalent in $|z| < 1$. For all real α and for an interval $0 \leqq t \leqq t_0$ let the function

$$\frac{1}{2}\left[f\left(\frac{z + x}{1 + \bar{x}z}\right) + f\left(\frac{z - x}{1 - \bar{x}z}\right)\right], \quad x = te^{i\alpha}\ , \tag{2.11}$$

be subordinate to $f(z)$ in $|z| < 1$. Then $f(z)$ is convex in $|z| < 1$.

For the proof of Theorem 3 we take $F(z, t)$ of Theorem B to be the function (2.11) and select $\rho = 2$. A calculation of $F(z)$ in (1.12), together with (1.16), leads to the inequality

$$(2.12) \qquad \mathscr{R}\left[(1 - e^{-2i\alpha}z^2)^2 e^{2i\alpha}\frac{f''(z)}{zf'(z)} - 2(1 - e^{-2i\alpha}z^2)\right] \leqq 0, \quad |z| < 1 .$$

Choose $\alpha = \text{amp}\, z$. Let $|z| = r < 1$. Then (2.12) becomes

$$(2.13) \qquad \mathscr{R}\left[(1 - r^2)^2 \cdot \frac{zf''(z)}{r^2 f'(z)} - 2(1 - r^2)\right] \leqq 0 ,$$

$$(2.14) \qquad \mathscr{R}\left[1 + \frac{zf''(z)}{f'(z)}\right] \leqq \frac{1 + r^2}{1 - r^2} .$$

Similarly, for $\alpha = \pi/2 + \text{amp}\, z$, we obtain

$$(2.15) \qquad \mathscr{R}\left[1 + \frac{zf''(z)}{f'(z)}\right] \geqq \frac{1 - r^2}{1 + r^2} > 0 .$$

It follows from (2.15) that $f(z)$ is convex for $|z| < 1$. It is to be noticed that equality occurs in (2.13) for the convex function

$$f(z) = \frac{1}{2}\nu \cdot \log\left(\frac{1 + z}{1 - z}\right) .$$

when $\alpha = 0$. In this case $F(z) \equiv 0$.

For another application of Theorem B we turn now to a class of function which need not be convex but which form a subclass of the class of close-to-convex functions introduced by W. Kaplan [1].

It is well known that if

$$(2.16) \qquad f(z) = z + \sum_2^\infty a_n z^n$$

is univalent and convex in $|z| < 1$, then $|a_n| \leqq 1$ [2]. The author [5, 6] has shown that if the coefficients are all real and if $f(z)$ is univalent and convex only in the direction of the imaginary axis for $|z| < 1$, then again $|a_n| \leqq 1$, but that if the coefficients are complex the results $|a_n| \leqq n$ is sharp. For the class of functions $f(z)$ which are close-to-convex in $|z| < 1$, the inequalities $|a_n| \leqq n$ again hold [4]. We now consider another class of functions, which are also close-to-convex in $|z| < 1$, but not necessarily convex, for which $|a_n| \leqq 1$. This class contains the odd star-like functions as a sub-class. The result is stated in the following theorem.

THEOREM 4. *Let the function*

$$(2.17) \qquad (1 - t)f(z) + tf(-z)$$

be subordinate to the univalent, regular function

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots \tag{2.18}$$

in $|z| < 1$ *for an interval* $0 \leqq t \leqq t_0$. *Then*

$$\mathscr{R}\left\{\frac{zf'(z)}{f(z) - f(-z)}\right\} > 0\,, \quad |z| < 1\,, \tag{2.19}$$

and the vector $\{f(z) - f(-z)\}$ *turns continuously in one direction as* z *traverses each circle* $|z| = r < 1$. $f(z)$ *is close-to-convex in* $|z| < 1$.

Proof. Let $\rho = 1$ and let $F(z, t)$ be the function in (2.17). Then $F(z)$ of (1.12) reduces to $(1/z)[f(-z) - f(z)]$ and $F(0) \neq 0$. (1.13) then leads to (2.19).

Now let

$$\arg\,[f(z) - f(-z)] = \phi, \arg z = \theta\,.$$

$$\begin{aligned}\frac{d\phi}{d\theta} &= \mathscr{R}z\left\{\frac{f'(z) + f'(-z)}{f(z) - f(-z)}\right\} \\ &= \mathscr{R}\left\{\frac{zf'(z)}{f(z) - f(-z)}\right\} + \mathscr{R}\left\{\frac{(-z)f'(-z)}{f(-z) - f(z)}\right\} > 0\,, \quad |z| < 1\,,\end{aligned} \tag{2.20}$$

by (2.19).

Since by (2.20) $\{f(z) - f(-z)\}$ is univalent and star-like in $|z| < 1$, it follows that

$$\psi(z) = \int_0^z \frac{f(t) - f(-t)}{t}\, dt\,, \quad |z| < 1\,, \tag{2.21}$$

is convex in $|z| < 1$. Thus (2.19) may be cast in the form

$$\mathscr{R}\left\{\frac{f'(z)}{\psi'(z)}\right\} > 0\,, \quad |z| < 1\,, \quad \varphi(z) \text{ convex,} \tag{2.22}$$

which implies that $f(z)$ is close-to-convex [1] in $|z| < 1$. This completes the proof of Theorem 4.

In a recent paper [3] G. Pólya and I. J. Schoenberg have shown that if $f(z)$ of (1.1) is univalent and convex in $|z| < 1$ then so are the de la Vallée Poussin means $V_n(z)$ of the power series (1.1),

$$\begin{aligned}V_n(z) = \frac{n}{n+1} z &+ \frac{n(n-1)}{(n+1)(n+2)} a_2 z^2 + \cdots \\ &+ \frac{n(n-1)\cdots 1}{(n+1)(n+2)\cdots(2n)} a_n z^n\,,\end{aligned} \tag{2.23}$$

and if D and D_n denote the convex domains into which the unit circle is mapped by $f(z)$ and $V_n(z)$, respectively, then $D_n \subset D$. In other words, $V_n(z)$ is necessarily subordinate to $f(z)$ for $n = 1, 2, \cdots$ when $f(z)$ is univalently convex. By means of Theorem B we can now prove that the condition $D_n \subset D$ for infinitely many values of n is also a sufficient condition that $f(z)$ be convex when $f(z)$ is univalent. The theorem of Pólya and Schoenberg in its extended form is now stated as Theorem 5.

THEOREM 5. *A necessary and sufficient condition that the function*

$$f(z) = z + a_2 z^2 + \cdots + a_n z^n + \cdots ,$$

regular and univalent in $|z| < 1$, *be convex in* $|z| < 1$ *is that the de la Vallée Poussin means* $V_n(z)$ *in* (2.23) *be subordinate to* $f(z)$ *in* $|z| < 1$ *for* $n = 1, 2, \cdots$.

Proof of sufficiency. In Theorem B we choose $\rho = 1$ and $F(z, t) = V_n(z)$ where $t = (n+1)^{-1}$. We define $F(z, 0) = \lim_{t \to 0+} F(z, t) = \lim_{n \to \infty} V_n(z) = f(z)$, uniformly in $|z| \leqq r < 1$. For $\rho = 1$ we shall show that the limit defining $F(z)$ in (1.12) exists uniformly and is precisely the analytic function $-\{zf''(z) + f'(z)\}$, $F(0) = -1$. When this is done (1.17) will give

$$\mathscr{R}\left\{1 + \frac{zf''(z)}{f'(z)}\right\} > 0 , \quad |z| < 1 ,$$

and the convexity of $f(z)$ follows. We need the following lemma.

LEMMA. *If n and k are positive integers, $k \leqq n$, then*

$$(n+1)\left[1 - \frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)}\right] \leqq k^2 . \tag{2.24}$$

We establish the lemma by mathematical induction. Let n be an assigned positive integer. It is readily seen that (2.24) holds for $k = 1$. Assuming that (2.24) is true for a value $k < n$ we prove that (2.24) also holds when k is replaced by $(k+1)$. Indeed, we have

$$\begin{aligned}(2.25)\quad &(n+1)\left[1 - \frac{n(n-1)\cdots(n-k+1)(n-k)}{(n+1)(n+2)\cdots(n+k)(n+k+1)}\right] \\ &= (n+1)\left[1 - \frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)}\left(1 - \frac{2k+1}{n+k+1}\right)\right] \\ &= (n+1)\left[1 - \frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)}\right] \\ &\quad + (2k+1)\frac{(n+1)}{n+k+1}\left[\frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)}\right] \\ &\leqq k^2 + (2k+1) = (k+1)^2 .\end{aligned}$$

Turning to the calculation of $F(z)$ we have

(2.26)

$$F(z) = \lim_{t\to 0+}\left[\frac{F(z, t) - F(z, 0)}{tz}\right] = \lim_{n\to\infty}\frac{n+1}{z}(V_n(z) - f(z))$$

$$= -\lim_{n\to\infty}(n+1)\sum_{k=1}^{n}\left\{1 - \frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)}\right\}a_k z^{k-1},\ a_1 = 1,$$

$$-\lim_{n\to\infty}(n+1)z^n\sum_{\nu=0}^{\infty}a_{\nu+n+1}z^{\nu}\ .$$

Let $|z| \leqq r < 1$. Since $f(z)$ is univalent we have $|a_{\nu+n+1}| < e(\nu + n + 1)$. Consequently for large n

$$(2.27)\qquad (n+1)z^n\sum_{\nu=0}^{\infty}a_{\nu+n+1}z^{\nu} = 0\left\{\frac{n^2r^n}{(1-r)^2}\right\} = p_n\ ,$$

$$(2.28)\qquad \lim p_n = 0$$

uniformly in $|z| \leqq r$.

Let N be a positive integer. Then

$$(2.29)\qquad \lim_{n\to\infty}(n+1)\sum_{k=1}^{N}\left\{1 - \frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)}\right\}a_k z^{k-1}$$

$$= \sum_{k=1}^{N}k^2 a_k z^{k-1}\ .$$

For $n > N$, $|z| \leqq r$, by the lemma we have

$$(2.30)\qquad \left|(n+1)\sum_{k=N+1}^{n}\left\{1 - \frac{n(n-1)\cdots(n-k+1)}{(n+1)(n+2)\cdots(n+k)}\right\}a_k z^{k-1}\right|$$

$$\leqq \sum_{k=N+1}^{n}k^2|a_k|r^{k-1} < e\sum_{N+1}^{\infty}k^3r^{k-1}\ .$$

Given $\varepsilon > 0$, we now choose $N_0(\varepsilon, r)$ so that for $N > N_0$

$$(2.31)\qquad e\sum_{N+1}^{\infty}k^3r^{k-1} < \varepsilon\ .$$

From (2.26), (2.28), (2.29), (2.30) and (2.31) it follows that the limit in (2.26) exists uniformly in $|z| \leqq r < 1$ and is the analytic function

$$(2.32)\qquad F(z) = -\sum_{k=1}^{\infty}k^2a_kz^{k-1},\ a_1 = 1,\ |z| < 1,$$

$$= -zf''(z) - f'(z)\ .$$

This completes the proof of the sufficiency part of Theorem 5. The necessity part was shown in [3]. In (2.26) since n is a positive integer we have let $t \to 0$ through a discrete set of values of t. This, however, in no way affects the validity of Theorem B.

REFERENCES

1. W. Kaplan, *Close-to-convex Schlicht functions*, Michigan Math. Journal, **1** (1952), 169-185.
2. Z. Nehari, *Conformal Mapping*, New York, 1952.
3. G. Pólya and I. J. Schoenberg, *Remarks on de la Vallée Poussin means and convex conformal maps of the circle*, Pacific J. Math., **8** (1958), 295-334.
4. M. O. Reade, *Sur une classe de fonctions univalentes*, C. R. Acad. Sci. Paris, **239** (1954), 1758-1759.
5. M. S. Robertson, *On the theory of univalent functions*, Annals of Math., **37** (1936), 374-408.
6. ———, *Analytic functions star-like in one direction*, Amer. J. Math., **58** (1936), 465-472.
7. L. Špaček, *Contribution à la theorie des fonctions univalentes*, Časopsis pro Pěstováni Matematiky a Fysiki, **62** (1932), 12-19.

RUTGERS * THE STATE UNIVERSITY.

PARTITION AND MODULATED LATTICES

DAVID SACHS

Introduction. The lattice of equivalence relations on a set S, or equivalently the lattice of partitions on a set S, is perhaps one of the most interesting lattices from the point of view of abstract algebra. Partition lattices were studied rather thoroughly by O. Ore [6], who also gave a characterization of them in geometric terms. Later, another characterization of partition lattices was given by U. Sasaki and S. Fujiwara [8]. Their characterization makes specific use of the notions of lines and planes and is somewhat combinatorial in point of view. In this paper we introduce the notion of a *modulated* lattice and give a characterization of partition lattices (Theorem 14) which is remarkably similar to the lattice-theoretic characterization of the classical projective geometries. Moreover, our study suggests that there may be continuous analogues of partition lattices much in the same way as the continuous geometries of J. Von Neumann are analogues of the classical projective geometries. After developing some preliminary on modulated lattices, we focus our attention on irreducible modulated matroid lattices. A simple property which may or may not be present in such lattices enables us to give our characterization of partition lattices. Curiously enough, we are able to give a characterization of partition lattices on an infinite set which is simpler in appearance than our more general result. We devote some attention to metric lattices and show that certain continuous modulated lattices must be continuous geometries. Finally, we mention some problems and extensions suggested by this paper.

Preliminaries. Let L be a lattice with operations $+, \cdot$, partially ordered by the relation $\leqq$. The zero (unit) element is written as 0 (I), and we shall usually assume that these elements are present. We write $(a, b)M$ and say that the pair (a, b) is modular if and only if $(c + a)b = c + ab$ for every $c \leqq b$. If the modular relation is symmetric, then the lattice is said to be *semi-modular*. If $(a, b)M$ and $ab = 0$, then we write $(a, b) \perp$ and say that the pair (a, b) is independent. We say that b covers c if $b > c$ and there is no x for which $b > x > c$. A *point* is an element which covers 0. An interval $[a, b]$ is the set of elements x such that $a \leqq x \leqq b$. For the convenience of the reader we include here some properties of semi-modular lattices, the proofs of which can be found in [1] and [3]. All maximal chains between two elements $b, c, b < c$, are finite and have the same length if there exists one finite maximal chain between the two elements [3, p. 88]. By the length of

Received April 25, 1960. This paper was presented to the American Mathematical Society under the title "Modulated and Partition Lattices" on January 29, 1960.

an interval we mean the number of elements in a finite maximal chain in the interval if there is such a chain. If $(x, y)M$, then the intervals $[xy, x]$, $[y, x + y]$ have the same length if one of the intervals has a finite maximal chain; moreover, if the intervals $[xy, x]$, $[y, x + y]$ have the same length, then $(x, y)M$. An element c is said to have *dimension* n if the interval $[0, c]$ has length $n + 1$; it has *codimension* m if the interval $[c, I]$ has length $m + 1$. By an *independent complement* of c we mean an element b such that $c + b = I, (c, b) \perp$. By a *line* we mean an element which covers a point, and by a *plane* we mean an element which covers a line. A *hyperplane* is an element which is covered by I.

Let L be a lattice with 0. L is *left-complemented* if for every $a, b \in L$, there exists $b' \leqq b$ such that $a + b' = a + b, (b', a) \perp$. A left-complemented lattice is semi-modular [10]. Later, we shall show that matroid lattices are left-complemented.

Now let L be a semi-modular lattice. We say that b is a *modular* element (bM) if $(b, c)M$ for every $c \in L$. The reader can easily verify that OM, IM and pM if p is a point. In the case of affine geometry these are the only modular elements. Evidently a necessary and sufficient condition that every element be modular is that L be a modular lattice.

The meet of two modular elements is modular. For if aM, bM, $c \in L$ and $d \leqq ab$, then $(d + c)ab = [(d + c)a]b = (d + ca)b = d + cab$ since $(c, a)M$ and $(ca, b)M$.

THEOREM 1. *Let aM. If $b \leqq a$ and $(b, e)M$ relative to $[0, a]$ for every $e \in [0, a]$, then bM.*

Proof. Notice that $(g, h)M$ in a lattice if and only if $(g, h)M$ relative to the interval $[gh, g + h]$. Let $c \in L$, $d \leqq b$. Then $(d + c)b = (d + c)ba = [(d + c)a]b = (d + ca)b = d + cab$ since aM and $ca \leqq a$. But $d + cab = d + cb$, and this completes the proof.

If bM, then the intervals $[bc, b], [c, b + c]$ are isomorphic, and the mappings $x \to x + c$ and $y \to yb, x \in [bc, b], y \in [c, b + c]$, are inverse isomorphisms between the two intervals.

Modulated lattices.

DEFINITION 1. A left-complemented lattice L with unit I is said to be *modulated* if for every yM and for every $z \geqq y$, there exists xM such that $x + z = I, xz = y$.

Since the zero element is modular, the above conditions cannot be satisfied vacuously. It is easily seen that every complemented modular lattice is a modulated lattice. We shall now give an example of a modulated lattice which is not a complemented modular lattice but is describable in terms of such a lattice. Let Λ be a complemented modular lattice with operations $+, \cdot$, of length $\geqq 3$ for which every interval sublattice is irreducible and which contains a point p. An example of such a lattice is a projective geometry of not necessarily finite length or its dual. We define

$$L' = \Lambda - [p] .$$

If L' is partially ordered in the natural manner, then it is easily seen that L' is a lattice. Moreover, if the join and meet operations are denoted by $\cup$ and $\cap$ respectively, then the following properties hold:

$$a \cup b = a + b ,$$
$$a \cap b = ab \text{ if } ab \neq p ,$$
$$a \cap b = 0 \text{ if } ab = p .$$

We observe that

(1) $0 \in L', I \in L'$;

(2) $x \in L', x \neq 0, y \geqq x$ implies $y \in L'$;

(3) $x \in L', y \leqq x$ implies the existence of $z \in L'$ such that $z + y = x$, $zy = 0$.

We obtain (3) by observing that $[0, x]$ is irreducible, and hence y has at least two complements in $[0, x]$. By a result due to Wilcox [10], L' is left-complemented and

$$(a, b)M \text{ if and only if } ab \neq p.$$

From this we deduce that aM in L' if and only if $a = I$ or $a \not> p$. Suppose now that $a \neq I$ and aM. Let $b > a$. If $b \geqq a + p \equiv c$, then a complement z of b within $[a, I]$ cannot contain $a + p$ so that $z \not> p$, and thus zM. If $b \not> c$, then $b + c$ covers b and therefore $bc = a$. By the irreducibility assumption applied to $[a, b + c]$, there exists $p' \neq c$ such that $p' + b = b + c, p'b = a$. If $x + b + c = I, x(b + c) = p'$, then $xb = x(b + c)b = p'b = a, x + b = x + p' + b = I$. Furthermore, $xc = x(b + c)c = p'c \neq c$. Therefore $x \not> c$ so that $x \not> p$, and xM. Since such an x exists, L' is a modulated lattice.

THEOREM 2. *Let L be a modulated lattice, and let a be a modular element. Then if $c \leqq a$, $(c, b)M$ relative to $[0, a]$ for all $b \in [0, a]$, and $c \leqq x \leqq a$; there exists an element y such that $yM, x + y = a, xy = c$.*

Proof. It follows from Theorem 1 that cM. There exists $z \in L$ such that $z + x = I, zx = c, zM$. Define $y \equiv za$. Then $yM, x + y = x + za = (x + z)a = a; xy = xza = xz = c$.

Since the meet of two modular elements in a semi-modular lattice is a modular element, it follows by induction that the meet of a finite number of modular elements is a modular element. If L has finite length, then the modular elements form a lattice; however, this lattice is not usually a sublattice of L. It will be shown later that the modular elements of a matroid lattice form a lattice, and our example given above shows there are other examples as well. We shall call the partially ordered system of modular elements $\mathfrak{M}$, and the dual of this system will be denoted by $\overline{\mathfrak{M}}$.

THEOREM 3. *If L is a modulated lattice and $\mathfrak{M}$ is a lattice, then $\overline{\mathfrak{M}}$ is a left-complemented lattice.*

Proof. If $\mathfrak{M}$ is a lattice, then the meets of elements in $\mathfrak{M}$ and L are the same. The join of two elements x, y in $\mathfrak{M}$ will be denoted by $x \cup y$. Notice that $x \cup y \geqq x + y$ in L. Let $a, b \in \mathfrak{M}$. To prove the theorem we need to show the existence of an element $b' \geqq b$ such that

$$ab' = ab \, ,$$
$$a \cup b' = I$$

and $c(b' \cup a) = cb' \cup a$ for every $c \in \mathfrak{M}, c \geqq a$. (This is the dual of left-complementation.) Since L is modulated, there exists an element $b'M$ such that $(a + b)b' = b, a + b + b' = I$. Thus $ab = ab', a + b' = I$, and so $a \cup b' = I$. If $c \geqq a$, then $c(b' \cup a) = c$; also $c(b' + a) = cb' + a$, or $c = cb' + a$. Thus $c \leqq cb' \cup a$, and since the reverse inequality is obvious, we have $c = cb' \cup a$. This proves the theorem.

THEOREM 4. *If L is modulated and $\mathfrak{M}$ is a lattice, then $(a, b)M$ in $\overline{\mathfrak{M}}$ if and only if $a + b = a \cup b$.*

Proof. Suppose $a + b = a \cup b$. Then if $c \geqq b$, $c(a \cup b) = c(a + b) = ca + b$, and so $c(a \cup b) \leqq ca \cup b$. The reverse inequality is obvious, and so $c(a \cup b) = ca \cup b$. Thus $(a, b)M$ in $\overline{\mathfrak{M}}$. Conversely, suppose $a \cup b \neq a + b$. Then $a \cup b > a + b$. If we consider the b' in Theorem 3, then we see that $b'a \cup b = b$. But $b'(a \cup b) > b$, for if $b'(a \cup b) = b$, then $a \cup b = (a \cup b)(b' + (a + b)) = (a \cup b)b' + (a + b) = a + b$. Hence $(a, b)M'$ in $\overline{\mathfrak{M}}$, and the proof is complete.

If, in our example, Δ is the dual of a projective geometry, then $\overline{\mathfrak{M}}$ is an affine geometry. What we want to show next is that L and $\overline{\mathfrak{M}}$ have the same center if every element in L is a join of points. We first state a few lemmas about left-complemented lattices. The following lemma is easily proved:

LEMMA 1. *If L is a left-complemented lattice, $(b, b') \perp$, and $c = cb + cb'$ for every $c \in L$, then L is isomorphic to the cardinal product of $[0, b]$ and $[0, b']$.*

LEMMA 2. *If L is left-complemented, then an element is in the* *center if and only if it has a unique complement.*

Proof. Suppose e has a unique complement e'. Then $(e, e')M$. Let $ue = 0$. There exists an element x such that $(u + e) + x = I, (u + e, x) \perp$, Now $e + (u + x) = I$ and $e(u + x) = e(u + e)(u + x) = e[(u + e)x + u] = eu = 0$. Since e' is the unique complement of $e, u + x = e'$. Thus if $ue = 0$, then $u \leqq e'$. For every x there exists an element b such that $ex + e'x + b = x, (ex + e'x, b) \perp$. Now $ex(e'x + b) = (ex + e'x)(e'x + b)ex = e'xex = 0$. Hence $e(e'x + b) = 0$ since $e'x + b \leqq x$. It then follows that $e'x + b \leqq e'$, and so $b \leqq e'x$ because $b \leqq x$. Thus $x = ex + e'x$ for every $x \in L$. We now use Lemma 1 to see that e is in the center. The converse is trivial.

LEMMA 3. *Let L be a left-complemented lattice. If e has a unique complement e', then e has a unique complement.*

Proof. Suppose that $e + e' = I$, $(e, e') \perp$, and that e' is the only element with these properties. If our lemma is false, then there exists an element b such that $e + b = I, eb = 0$ and $(e, b)M'$. Since there exists $b' \leqq b$ such that $e + b' = e + b = I, (e, b') \perp$, it follows that $b > e'$. Then there exists $x \leqq b$ such that $e' + x = b$ and $(e', x) \perp$. Also, there exists $x' \leqq x$ such that $e + x' = e + x$ and $(e, x') \perp$. Moreover, there exists an element y which is an independent complement of $e + x'$. Therefore $e + (x' + y) = I$ and $e(x' + y) = e(e + x')(x' + y) = ex' = 0$. If $c \leqq e$, then $(c + x' + y)e = [((c + x') + y)(e + x')]e = (c + x')e = c$. Thus $(e, x'+ y) \perp$. From the uniqueness of e' we obtain $y + x' = e'$, and this implies $x' \leqq e'$. Therefore $x' \leqq e'x = 0$, and then we have $e + x = e$. Thus $x \leqq eb = 0, e' = b$, and the proof is complete.

THEOREM 5. *Let L be a modulated lattice such that every element is a join of points. If $\mathfrak{M}$ is a lattice, then L and $\overline{\mathfrak{M}}$ have the same center.*

Proof. If an element e lies in the center of L, then it is modular and so it lies in $\overline{\mathfrak{M}}$. Since e has a unique complement e' in L, e' must be the unique independent complement of e in $\overline{\mathfrak{M}}$; thus e lies in the center of $\overline{\mathfrak{M}}$. Conversely, suppose e lies in the center of $\overline{\mathfrak{M}}$. Then e also lies in the center of $\mathfrak{M}$. Let e' be its unique complement in $\mathfrak{M}$. Suppose $c \in L$. Obviously $c \geqq ce + ce'$. If p is a point within c, then since $p \in \mathfrak{M}, p = pe \cup pe'$. Thus $p \leqq e$ or $p \leqq e'$. In any case $p = pe + pe'$. Hence $p \leqq ce + ce'$, and since c is a join of points, $c \leqq ce + ce'$. Consequently, $c = ce + ce'$ for every $c \in L$. Since $ee' = 0, e \cup e' = I$ and $(e, e')M$ in $\overline{\mathfrak{M}}$; $e + e' = I$, and now the result is obvious.

COROLLARY. *Let L be a modulated lattice such that every element* *is a join of points. If $\mathfrak{M}$ is a lattice, then L is irreducible if and only if $\overline{\mathfrak{M}}$ is irreducible.*

Matroid lattices.

DEFINITION 2. A lattice L is a *matroid* lattice if it has the following properties:
(4) L is a complete lattice;
(5) Every element in L is the join of points;
(6) If p is a point and $p \not\leqq b$, then $p + b$ covers b;
(7) If $p \leqq \sum_{\alpha \in A} p_\alpha$ where p and p_α are points, then there exists a finite subset B of A such that $p \leqq \sum_{\beta \in B} p_\beta$.[1]

A set Y of elements in L is said to be increasingly directed if for every $x, y \in Y$, there exists $z \in Y$ such that $z \geqq x, y$. We define a decreasingly directed set in an analogous manner. If L is a lattice satisfying (4) and (5), then condition (7) is equivalent to the following one:

$$x \sum_{y \in Y} y = \sum_{y \in Y} (xy) \tag{8}$$

where Y is an increasingly directed set. We call this property *meet continuity*.[2] The proof of the equivalence of (7) and (8) can be found in [3]. We shall now show that a matroid lattice is left-complemented.

LEMMA 4. *If p is a point, then $(b + p)a = b + pa$ if $b \leqq a$, i.e., $(p, a)M$.*

Proof. If $p \leqq b$ or $p \leqq a$, this is obvious. If not, then $b + p$ covers b so that since $b + p \not\leqq a$, $(b + p)a = b$. But obviously then $(b + p)a = b + pa$.

LEMMA 5. *Let $a, b \in L$. There exists a maximal element b' such that $b' \leqq b$, $(b', a) \perp$.*

Proof. Define

$$S = [x \in S; x \leqq b, (x, a) \perp] .$$

S is partially ordered by the relation $\leqq$ in L. If C is a chain in S and $q = \sum_{c \in C} c$, then $q \leqq b$; moreover, a direct application of meet continuity shows that $qa = 0$. If $m \leqq a$, then $(m + q)a = (m + \sum_{c \in C} c)a = (\sum_{c \in C}(m + c))a$. Since the set of elements of the form $b + c$ is an

[1] A rather through discussion of matroid lattices can be found in [3] where they are called "*treillis geometriques*". We use the name matroid because it seems to be the more common term.

[2] This property is sometimes called *upper continuity*.

increasingly directed set,

$$(m + q)a = \sum_{c \in C}(m + c)a = \sum_{c \in C}(m + ca) = m .$$

Thus $q \in S$, and the existence of b' follows from Zorn's lemma.

LEMMA 6. *If $a, b \in L$, there exists $b' \leqq b$ such that $a + b' = a + b$, $(b', a) \perp$.*

Proof. There exists a maximal element $b' \leqq b$ such that $(b', a) \perp$. If $a + b' \neq a + b$, then there exists a point p such that $p \leqq b$, $p \nleqq a + b'$. Since p is a point, $(p, b' + a) \perp$. This implies that $(p + b', a) \perp$ which contradicts the definition of b'.

THEOREM 6. *A matroid lattice is left-complemented and hence semi-modular.*

We state without proof the following structure theorem for matroid lattices:

Structure Theorem. Every matroid lattice is the cardinal product of irreducible matroid lattices; moreover, irreducible matroid lattices are characterized by the fact that any two points have a common complement.[3]

This is the main result of [7], but the theorem was proved in the finite length case in [2] and [4]. Using this theorem, we can easily prove a theorem important for our investigation.

THEOREM 7. *If L is an irreducible matroid lattice and aM, then $[0, a]$ is an irreducible matroid lattice.*

Proof. It is obvious that $[0, a]$ is a matroid lattice. Let b, c be points in $[0, a]$. Since L is irreducible, there exists x such that $b + x = c + x = I$, $bx = cx = 0$. Thus $b + ax = (b + x)a = a = (c + x)a = c + ax$ since aM, and $bax = cax = 0$. Therefore b and c have a common complement in $[0, a]$ from which it follows that $[0, a]$ is irreducible.

COROLLARY. *If L is an irreducible matroid lattice, l is a line, and lM, then l contains at least three distinct points.*

THEOREM 8. *Let L be a matroid lattice. If H is a set of elements each of which is modular, then $\prod_{h \in H} h \equiv h'$ is a modular element.*

Proof. The set of all finite meets of elements in H has the same meets as H, consists of modular elements, and is decreasingly directed.

[3] By cardinal product we mean cardinal product in the unrestricted sense, i.e., we do not require almost all of the entries to be 0.

Thus without loss of generality we assume that H is decreasingly directed, and that its elements are indexed. Let b be an element of finite dimension, and let $c \leqq h'$. Thus $c \leqq h_\beta$ for every β. If for every h_γ there is an h_α such that $h_\alpha < h_\gamma$ and $bh_\alpha < bh_\gamma$, then there exists an infinite chain between b and 0. But this is impossible since b has finite dimension. Therefore for some h_γ, $bh_\alpha = bh_\gamma$ for every $h_\alpha < h_\gamma$. Thus $bh' = bh_\gamma$, and $c + bh' = c + bh_\gamma = (c+b)h_\gamma \geqq (c+b)h'$. Hence h' is modular with every finite-dimensional element. Let d' be any element in L. The set D of finite-dimensional elements contained in d' is increasingly directed, and its join is d'. If $c \leqq h'$, then we have

$$\begin{aligned}(c + d')h' &= \Big(c + \sum_{d\in D} d\Big)h' = \Big(\sum_{d\in D}(c+d)\Big)h' = \sum_{d\in D}(c+d)h' \\ &= \sum_{d\in D}(c + dh') = c + \Big(\sum_{d\in D} dh'\Big) = c + \Big(\sum_{d\in D} d\Big)h' = c + d'h' ,\end{aligned}$$

where we have used meet continuity twice. This proves the theorem.

Theorem 8 shows that $\mathfrak{M}$ is always a lattice, in fact a complete lattice, when L is a matroid lattice. If L is also modulated, then every element in $\mathfrak{M}$ is a join of points since $\mathfrak{M}$ is left-complemented and every modular element in $L \neq I$ is contained in a modular hyperplane.

THEOREM 9. *If X is an increasingly directed set of modular elements in a matroid lattice L, then $\sum X \equiv x'$ is a modular element.*

Proof. Let $b \in L$, with $c \leqq b$. Then

$$\begin{aligned}(c + x')b &= \Big(c + \sum_{x\in X} x\Big)b = \Big(\sum_{x\in X}(c+x)\Big)b = \sum_{x\in X}(c+x)b \\ &= \sum_{x\in X}(c + xb) = c + \Big(\sum_{x\in X} xb\Big) = c + \Big(\sum_{x\in X} x\Big)b = c + x'b .\end{aligned}$$

We shall now restrict our attention to modulated matroid lattices. It is easily shown that a cardinal product of lattices is modulated if and only if each of the factors is modulated. In view of the Structure Theorem we shall therefore concentrate on the irreducible case.

LEMMA 7. *Let L be a matroid modulated lattice which is irreducible. Then any two hyperplanes have a common independent complement.*

Proof. Let h', h'' be any two hyperplanes in L. We choose a point p such that $ph' = 0$. If $ph'' = 0$, then p is the common independent complement. If $p \leqq h''$, then we choose a line lM such that $l + h'' = I$, $lh'' = p$. Now l must contain at least two more points r, s. Neither r nor s can be contained in h'', for then $l \leqq h''$. If both r and s are contained in h', then $h' \geqq p$ which is false. Hence r or s is a common independent complement.

THEOREM 10. *If L is an irreducible modulated matroid lattice, then any two elements of the same dimension (or codimension) have a common independent complement.*

Proof. Let a and b have the same dimension or codimension. From the meet continuity condition, there exists a maximal element x such that $(a, x) \perp, (b, x) \perp$. Because a and b have the same dimension or codimension, $a + x = I$ if and only if $b + x = I$. Suppose then that $a + x \neq I, b + x \neq I$. Then $a + x$ and $b + x$ are contained in hyperplanes, and there exists a point p such that $(a + x, p) \perp, (b + x, p) \perp$. One easily sees that $(a, x + p) \perp, (b, x + p) \perp$ contrary to the definition of x.

COROLLARY. *If L is an irreducible modulated matroid lattice, and if aM, bM and a, b have the same dimension or codimension, then $[0, a]$ is isomorphic to $[0, b]$.*

Proof. If x is a common independent complement of a and b, then $[0, a]$ and $[0, b]$ are both isomorphic to $[x, I]$.

We shall now restrict our attention to irreducible modulated matroid lattices of length $\geqq 5$. Let us consider the following property:

(γ) L contains a point p which lies in a plane tM such that three M-lines contain p and are contained in t.

LEMMA 8. *Suppose that L is an irreducible matroid modulated lattice of length $\geqq 5$ satisfying condition (γ). Let dM be a plane and let p' be a point contained in d. Then p' is contained in three M-lines l, m, n where $l, m, n \leqq d$.*

Proof. Suppose that $p' = p$ and $d \neq t$. Let d' be a common complement of t and d. The perspective mapping from $[0, t]$ to $[0, d]$ with d' as a common complement leaves p fixed for $(p + d')d = p + d'd = p$ since $p \leqq d$. The images of the three M-lines in t containing p are the required lines. (The images are obviously M-lines in $[0, d]$ and since dM, they are M-lines.) Suppose now that p' is any point in t distinct from p. If $(p + p')M$, then $p + p'$ contains a third point p'' because L is irreducible. From the length condition, t is contained in a 3-space S which is modular. From Theorem 2 we see that there exist M-planes $t', t'' \leqq S$ such that $t' > p, t'' > p'$ and such that p'' is a common complement of t' and t'' within $[0, S]$. Using our previous results, we conclude that t'' contains three M-lines which contain p', and a repeated application shows that t contains three M-lines which contain p'. If p' is a point for which $(p + p')M'$, then we can find a point $q \leqq t$ such that $(p + q)M$ and $(q + p')M$. Since q has three M-lines in

t containing it, we conclude that p' has the same property. Hence every point in t is contained in three M-lines which lie in t. We now deduce the conclusion of the theorem by noting that every two M-planes are perspectively isomorphic, and that in the isomorphisms M-lines map onto M-lines.

THEOREM 11. *If L is an irreducible matroid modulated lattice of length $\geqq 5$ which satisfies* (γ), *then every line in an M-plane d contains at least three points.*

Proof. Let l be such a line. We choose a point $p \leqq d$ such that $p \nleqq l$ and then apply Lemma 8.

If L is an irreducible modulated matroid lattice of length $\geqq 5$, then we say it satisfies (δ) if every line in an M-plane contains at least three points. Thus Theorem 11 says that (γ) implies (δ).

LEMMA 9. *Let L satisfy* (δ). *If sM, s covers z, z covers l, z covers m, mM and $l \neq m$, then there exists a point p such that $p + l = p + m = z$.*

Proof. Since z covers l, m and $(l, m)M$; l, m cover lm. There exists an element xM such that $x + lm = s, xlm = 0$. Clearly the interval $[0, x]$ is isomorphic to the interval $[lm, s]$. Since every line in $[0, x]$ contains three points, there exists an element r such that $l + r = m + r = z, lr = mr = lm$. There exists a point p such that $p + lm = r$. Then $l + p = l + lm + p = l + r = z$, and similarly $m + p = z$.

If s is a modular element, then in view of Theorem 1 $[0, s]$ is a modular lattice if and only if every element in $[0, s]$ is a modular element.

LEMMA 10. *Let L be a matroid lattice, and let C be a chain of elements each of which is modular and has the property that every element contained in it is modular. Then $\sum C$ is a modular element, and every element contained in it is modular.*

Proof. That $\sum C$ is modular follows from Theorem 9. Let $a \leqq \sum C$. The set of elements of the form $ac, c \in C$, is an increasingly directed set of elements with join $\sum_{c \in C} ac = a \sum_{c \in C} c = a$. Each of the elements ac is modular, and therefore a is modular.

We now consider the set of all modular elements c such $[0, c]$ is a modular lattice. According to the preceding lemma and Zorn's lemma, there must exist a maximal such element. The next lemma tells us the character of such maximal elements if L satisfies (δ).

LEMMA 11. *Let L be a lattice satisfying* (δ). *If l' is a maximal element with the property that $l'M$ and $[0, l']$ is a modular lattice, then l' is a hyperplane or $l' = I$.*

Proof. Suppose that l' is not a hyperplane or I. Then there exists an element $Q \neq I$ which is modular and covers l'. Since $Q \neq I$, there exists a modular element s which covers Q. Suppose that l is covered by Q, and lM'. Define $ll' \equiv q$. Since $(l, l')M$, l covers q; moreover, q is modular because $q \leqq l'$. There exists mM such that $Qm = q$, $Q + m = s$, Evidently m covers q. Define $z \equiv m + l$; then z covers l, m and is not contained in Q. By Lemma 9 there exists a point p such that $p + l = p + m = z$. Since $z \nleqq Q$, it follows that $p \nleqq Q$. There exists an M-element R such that $R + z = s, Rz = m$. Thus $Q + p = R + p = s$, $Qp = Rp = 0$. The mapping $x \rightarrow (x + p)Q, x \leqq R$, is an isomorphic mapping from $[0, R]$ onto $[0, Q]$. Evidently $(m + p)Q = (m + l)Q = l + mQ = l + q = l$. Since m is modular with every element in R, l is modular with every element in Q. Thus every maximal element in Q is a modular element, and therefore by Theorem 8 every element in Q is a modular element. Thus $[0, Q]$ is a modular lattice with QM, and this contradicts the fact that l' was a maximal element with this property.

THEOREM 12. *Let L be an irreducible matroid modulated lattice with length $\geqq 5$ satisfying* (γ) *or* (δ). *Then if $b \neq I$ and bM, then $[0, b]$ is modular lattice.*

Proof. If L satisfies (γ), then L satisfies (δ). According to the previous lemma, there exists a hyperplane hM such that $[0, h]$ is a modular lattice. Now if tM and t is a hyperplane, then $[0, t]$ is a modular lattice because $[0, t]$ is isomorphic to $[0, h]$. If $b \neq I$ and bM, then b is contained in a modular hyperplane. The result is now evident.

COROLLARY. *If L is an irreducible matroid modulated lattice with length $\geqq 5$, then L satisfies* (γ) *if and only if it satisfies* (δ). *Moreover, L satisfies* (γ) *if and only if $ab \neq 0$ implies $(a, b)M$.*

Proof. We have already shown that (γ) implies (δ). Suppose that L satisfies (δ) and $ab \neq 0$. There exists an element mM such that $ab + m = I, abm = 0$. The interval $[ab, I]$ is isomorphic to $[0, m]$. Since $m \neq I$, $[0, m]$ is a modular lattice, and therefore $(a, b)M$ within $[ab, I]$. Thus $(a, b)M$.

If $ab \neq 0$ implies $(a, b)M$, then $[p, I]$ is a modular lattice if p is a point. If hM is a complement of p, then $[0, h]$ is a modular lattice. Since h must have at least dimension 3, (γ) follows immediately.

Partition lattices. The corollary to Theorem 12 tells us a great deal about the irreducible modulated matroid lattices of length $\geqq 5$ which satisfy (γ) and also gives us a condition equivalent to (γ) which is free of the notion of lines and planes. We digress from our abstract theory to discuss partition lattices.

It is well known [3; p. 265] that partition lattices are matroid lattices. Following Ore [6], we shall call a set of a partition a block and a partition *singular* if it contains at most one block with more than one element. It is implicit in a result due to Ore [6; p. 583] that the singular partitions are precisely the modular elements. We give here a proof in line with our ideas. If a partition is not singular, then one can easily construct a line which is not modular with it. It is also easily seen that a singular partition which is a hyperplane is a modular element. Since every singular partition is a meet of hyperplane singular partitions, we conclude the result if we use Theorem 8.

Let A be a singular partition. If B is a partition containing A, we can construct a singular partition B' as a complement of B within $[A, I]$ as follows:

$$A = [a_1, \cdots, a_\alpha]\,[a]\,[b]\,[c]\cdots$$
$$B = [a_1, \cdots, a_\alpha, \cdots]\,[b_1, \cdots]\,[b_2, \cdots]\,[b_\omega, \cdots]\cdots$$
$$B' = [a_1, \cdots, a_\alpha, b_1, b_2, b_\omega, \cdots]\,[e]\,[f]\cdots .$$

That is, we select one element from each of the blocks of B that does not contain the main block of A and combine these elements with the main block of A into one block. If we let $A = 0$, i.e., we let A be the partition with all blocks of one element, then we see immediately that B' is not unique if $B \neq I, 0$. This shows that a partition lattice is irreducible. Thus partition lattices are irreducible, matroid and modulated.

We now consider the case when L is an irreducible modulated matroid lattice of length $\geqq 5$ and does not satisfy (γ). Since L does not satisfy (γ), then in a modular plane P some line l contains only two points. The line l is not an M-line, for an M-line must contain at least three points since L is irreducible. If t_1 and t_3 are the points on l, then each must be contained in at least two M-lines in P because P is an M-element. Since l is not an M-line, P must contain at least four M-lines. If there is a fifth M-line, then it cannot meet l at t_1 or t_3 because L would then satisfy condition (γ). Thus it must meet l at a third point which is impossible. Hence there are exactly four M-lines in P. If these lines are $l_1, l_2 \geqq t_1$ and $l_3, l_4 \geqq t_3$, then $t_1, t_3, l_1l_3, l_1l_4, l_2l_3, l_2l_4$ are points distinct from each other. There no other points in P because every point must be the meet of two M-lines. The plane contains two more lines $l_1l_3 + l_2l_4$, $l_1l_4 + l_2l_3$ and no others because there are no points remaining to make lines. It is easily verified that P is isomorphic to the lattice of partitions on a set with four elements. Our next aim is to show that the lattice $\mathfrak{M}$ associated with L is isomorphic to the lattice of singular partitions of some partition lattice.

LEMMA 12. *If $h \in L$ is an M-element covered by a hyperplane, then h is contained in exactly two M-hyperplanes.*

Proof. Since L is modulated, h is contained in at least two M-hyperplanes. Suppose it is contained in three M-hyperplanes h_1, h_2, h_3. Let p be a point such that $p \leqq h$. There exists an M-element m such that $m + h = I, mh = p$. Obviously m is a plane. If $mh_1 = mh_2$, then $mh_1 + h = mh_2 + h$ or $h_1(m + h) = h_2(m + h)$ and therefore $h_1 = h_2$. Thus mh_1, mh_2, mh_3 are three distinct M-lines containing p in m. But then L satisfies (γ) which is false.

LEMMA 13. *Every point in L has exactly two M-hyperplanes as complements.*

Proof. Since L is irreducible, $\overline{\mathfrak{M}}$ is irreducible, and thus every element in $\overline{\mathfrak{M}}$ must have at least two independent complements. This simply means that every point in L has at least two M-hyperplanes as complements. Suppose then that the point p has three M-complements h_1, h_2, h_3. Now $h_1h_2 \neq h_2h_3$ because equality would contradict Lemma 12. Thus h_1h_2, h_2h_3, h_1h_3 are distinct elements that cover their intersection $h_1h_2h_3$. If m is an M-complement of $p + h_1h_2h_3$ within $[p, I]$, then $m \geqq p$, $m + h_1h_2h_3 = I$ and $mh_1h_2h_3 = 0$. Evidently m is a plane, and the intervals $[0, m]$ and $[h_1h_2h_3, I]$ are isomorphic. Thus mh_1, mh_2, mh_3, are three M-lines in m which do not contain p. But this is impossible because in the lattice of partitions on four elements, any three M-lines contain all the points.

LEMMA 14. *In a matroid modulated lattice if two elements b, c have the same M-complements, then $b = c$.*

Proof. Suppose $b \neq c$. Without loss of generality we can assume that c contains a point p which b does not contain. If m is an M-complement of $p + b$ within $[p, I]$, then $m \geqq p, m + b = I, mb = 0$. But $mc \geqq p$ and the contradiction is apparent.

LEMMA 15. *Given any two distinct M-hyperplanes h', h'' in L, there exists a point p which is a complement of both h' and h''.*

Proof. Let l be an M-element which is a complement of $h'h''$. A point on l distinct from lh' and lh'' satisfies the condition.

Let G be the set of M-hyperplanes in L.

THEOREM 13. *The* $\mathfrak{M}$ *lattice of* L *is isomorphic to the lattice of singular partitions of* G.

Proof. With each point in L we associate the set of elements in G that are complementary to it. The previous lemmas show that each point p is associated with a two element set, different points are associated with different sets, and that every two element subset of G is associated with a point in L. Moreover, an M-hyperplane contains a point p if and only if it is not a member of the set associated with p. Consider the lattice of partitions of the set G. If α is an M-hyperplane of L, then we map it onto the maximal singular partition $[\alpha][\cdots]$. This mapping is obviously one-to-one onto the maximal singular partitions of G. If p is a point in L and is associated with the subset $[\beta, \gamma]$ of G, then we map it onto the partition $[\beta, \gamma][a][b]\cdots$ where all blocks but the first have one element. This mapping is also one-to-one from the points of L onto the points of the partition lattice of G. Now a maximal singular partition $[\alpha][\cdots]$ contains a point partition $[\beta, \gamma][a][b]\cdots$ if and only if $\alpha \notin [\beta, \gamma]$. From this we immediately see that we have defined an order preserving mapping in both directions between the M-hyperplanes and points of L and the M-hyperplanes and points of the partition lattice on G. According to [5; p. 200], two complete lattices in which every element is the join of points and the meet of hyperplanes are isomorphic if the partially ordered sets of hyperplanes and points are isomorphic. It thus follows that the $\mathfrak{M}$ lattice of L is isomorphic to the lattice of singular partitions of G.

DEFINITION 3. Let L' be a modulated matroid lattice. A set H of elements in $\mathfrak{M}'$ is said to be a *quasi-ideal* if

(8) $x \in H$ and $y \leqq x$ imply $y \in H$;

(9) $x, y \in H$ and $(x, y)M$ in $\overline{\mathfrak{M}}'$ (i.e., $x + y = x \cup y$) imply $x \cup y \in H$;

(10) the join of an increasingly directed set of elements in H is also in H (note that join in the sense of $\mathfrak{M}'$ and L' are synonymous for increasingly directed sets).

H is a *maximal* quasi-ideal if in addition it satisfies the following property:

(11) $I \notin H$; if K is a quasi-ideal and $H \subset K$, then $K = H$ or $K = \mathfrak{M}'$.

LEMMA 16. *If* H *is the set of elements in* $\mathfrak{M}'$ *contained in a hyperplane* h *in* L', *then* H *is a maximal quasi-ideal.*

Proof. The first three properties are immediately evident even if h is not a hyperplane. Suppose $H \subset K$ and $H \neq K$. Property (10) implies that there exists a maximal element m in K not in H and therefore not contained in h. If $m \neq I$, there exists $m'M$ such that m' covers m. Thus $m'h + m = m'$; consequently there is a point $p \leqq h$ such that $p + m = m'$. Thus $m' \in K$ by (9) which is impossible.

LEMMA 17. *The maximal quasi-ideals of a lattice of singular partitions of a partition lattice T are the set of singular partitions contained in a partition of two blocks.*

Proof. It is obvious that the set of singular partitions $\leqq$ a singular partition of two blocks is a maximal quasi-ideal. If we note that two singular partitions form a modular pair in $\overline{\mathfrak{M}}_T$ (the $\overline{\mathfrak{M}}$ lattice of T) if and only if their main blocks overlap, then we readily see that the set of singular partitions $\leqq$ a partition P of two non-trivial blocks is a quasi-ideal. The quasi-ideal determined by P has two maximal singular partitions m', m'' whose main blocks do not overlap. If a point p is not in P, then its main block must overlap the main blocks of m' and m''. Then $(p, m')M$ in $\overline{\mathfrak{M}}_T$, and also $(p \cup m', m'')M$ in $\overline{\mathfrak{M}}_T$. Thus I is in any quasi-ideal containing the quasi-ideal determined by P since $p \cup m' \cup m'' = I$, and this proves the maximality of the quasi-ideal determined by P. If Q is any quasi-ideal, then in view of (10) it must contain maximal elements. The main blocks of these maximal elements cannot overlap, for otherwise they would be modular in $\overline{\mathfrak{M}}_T$ and then they could not be maximal. This observation immediately shows that any quasi-ideal Q must consist of the singular partitions $\leqq$ some partition. But obviously such a quasi-ideal cannot be maximal unless the partition has two blocks.

LEMMA 18. *Every maximal quasi-ideal in $\mathfrak{M}$ of L is determined by a hyperplane in L.*

Proof. In view of Lemma 17, every maximal quasi-ideal in $\mathfrak{M}$ of L has one or two maximal elements since $\mathfrak{M}$ is isomorphic to the lattice of the singular partitions of G. If a maximal quasi-ideal has one maximal element, then this element must be a hyperplane in L (hyperplanes in $\mathfrak{M}$ are hyperplanes in L), and the lemma is true in this case. Let Q be a maximal quasi-ideal with two maximal elements m' and m''. In the lattice $\mathfrak{M}$, $m' \cup m'' = I$. Since $(m', m'')M'$ in $\overline{\mathfrak{M}}$, $m' \cup m'' > m' + m''$ in L. Evidently the set of M elements contained in $m' + m''$ is a quasi-ideal containing Q, hence equal to Q since Q is maximal. If $m' + m''$ is not a hyperplane, then there is a hyperplane $h > m' + m''$. Since h determines a maximal quasi-ideal, h and $m' + m''$ must determine the same quasi-ideal. But this is impossible, for h contains a point which is not $\leqq m' + m''$. The proof is complete.

THEOREM 14. *A necessary and sufficient condition that a lattice L*

of length $\geqq 5$ *be isomorphic to a lattice of partitions of a set* G *is that*

(12) L *be an irreducible modulated matroid lattice;*

(13) *there exist a pair of elements* (a, b) *such that* $(a, b)M', ab \neq 0$.

Proof. Let L satisfy conditions (12) and (13). Our previous results have shown that the $\mathfrak{M}$ lattice of L is isomorphic to the lattice of singular partitions of the set G of modular hyperplanes in L. Furthermore, the maximal quasi-ideals of $\mathfrak{M}$ are in a one-to-one correspondence with the hyperplanes of L and the partitions of two blocks in the lattice of partitions of G. Thus there is a one-to-one order preserving correspondence in both directions between the hyperplanes and points of L and the hyperplanes and points the lattice of partitions of G. From this we conclude that the two lattices are isomorphic.

Conversely, if L is a partition lattice, then it evidently satisfies (12). It is easily shown that a partition lattice of length $\geqq 5$ has two hyperplanes which do not form a modular pair and do not meet in 0 since every interval $[a, I]$ of a partition lattice is itself a partition lattice. The proof is complete.

REMARK. It is impossible for (13) to be satisfied in a matroid lattice of length $\leqq 4$. Hence our condition in Theorem 14 is neither necessary nor sufficient if L has length $\leqq 4$ although (12) is necessary.

A lattice is said to be *simple* if it has only trivial congruence relations. Obviously a simple lattice is irreducible although the converse is not necessarily true. Ore [6] has shown that a partition lattice is simple. Thus Theorem 14 is still correct if we replace the word "irreducible" by the word "simple". What we intend to show is that if L is of infinite length, then condition (13) may be deleted if simplicity replaces irreducibility in (12).

By a *neutral* ideal of a relatively complemented lattice L with 0, we mean an ideal which is the kernel of a homomorphic mapping. Ore [6] has given the following intrinsic characterization of a neutral ideal: an ideal N is neutral if and only if for every $x, y \in L, a \in N$, there exists $b \in N$ such that $xy + b = (x + a)y + b$.

LEMMA 19. *Let* L *be a semi-modular lattice. If* b *covers* a *and* $(a, y)M$, *then* $by = ay$ *or* by *covers* ay.

Proof. If $(a, y)M$, then $(c + a)yb = (c + ay)b = c + ay = c + ayb$ if $c \leqq yb$. Thus $(a, by)M$. Since b covers a, $a = a + by$ or $a + by$ covers a. If $a = a + by$, then $ay \geqq by$ and therefore $ay = by$ since b covers a. If $a + by$ covers a, then by covers $aby = ay$ since if $(u, v)M$ and $u + v$ covers v, then u covers uv.

LEMMA 20. *Let L be an irreducible modulated lattice of infinite length satisfying (γ). Then the set F of all finite-dimensional elements is a neutral ideal.*

Proof. It is obvious that F is an ideal. Suppose that $x, y \in L$ and $a \in F$. Let h be an M-hyperplane. Since a is finite-dimensional, there exists a finite maximal chain between x and $x + a$. Using the results of Lemma 19 and the fact that yh is an M-element (cf. Theorem 12), we see that there is a maximal chain between xyh and $(x + a)yh$ of length no greater than the maximal chain between x and $x + a$. Since $(x + a)yh = (x + a)y$ or $(x + a)yh$ is covered by $(x + a)y$, there exists a finite maximal chain between xy and $(x + a)y$. Let b be an independent complement of xy within $[0, (x + a)y]$. Then $b \in F$ and $(x + a)y + b = xy + b$. This completes the proof.

THEOREM 15. *A lattice L of infinite length is isomorphic to a partition lattice if and only if it is a simple modulated matroid lattice.*

Proof. The necessity is evident. If L satisfies (γ) and is irreducible, then it cannot be simple in view of Lemma 20.

REMARK. Every projective geometry of finite length is a simple modulated matroid lattice, so that our condition in Theorem 15 is not sufficient if L has finite length.

Metric lattices. By a valuation on a lattice L we mean a real-valued function $| \ |$ defined for each element in L such that

(14) $x < y$ implies $|x| < |y|$;

(15) $|x + y| + |xy| \leqq |x| + |y|$;

(16) $(x, y)M$ if and only if $|x + y| + |xy| = |x| + |y|$.

As is well known, every semi-modular lattice of finite length has such a valuation. If we define $d(x, y) = 2|x + y| - |x| - |y|$, then L becomes a metric space [9] in which $|(|a| - |b|)| \leqq d(a, b)$. Moreover, $d(a + e, b + f) \leqq d(a, b) + d(e, f)$ [9], so that the join operation is a uniformly continuous function. We shall refer to L as a *metric* lattice and say that L is metrically complete if it is complete in the metric defined above. A metrically complete lattice L is complete as a lattice if it contains 0 and I [9].

LEMMA 21. *If $(a, x)M$ and $(a, y)M$, then $d(ax, ay) \leqq d(x, y)$.*

Proof. We have

$$
\begin{aligned}
&|a(x+y)| - |ay| \\
&\qquad \leqq |a| + |x+y| - |a+x+y| - |a| - |y| + |a+y| \\
&\qquad \leqq |x+y| - |y| .
\end{aligned}
$$

Similarly, $|a(x+y)| - |ax| \leqq |x+y| - |x|$. Thus $2|a(x+y)| - |ax| - |ay| \leqq 2|x+y| - |x| - |y|$. But since $a(x+y) \geqq ax + ay$, $|a(x+y)| \geqq |ax+ay|$. Hence $2|ax+ay| - |ay| - |ax| \leqq 2|x+y| - |x| - |y|$, and this proves the lemma.

THEOREM 16. *Let L be a metrically complete lattice. If the sequence (a_i) has limit a and $(a_i, x)M$ for every i, then $(a, x)M$ and* lim $(a_i x) = ax$.

Proof. For each i we have $|a_i + x| + |a_i x| = |a_i| + |x|$. Thus $|a_i x| = |a_i| + |x| - |a_i + x|$. If $\lim (y_i) = y$, then $\lim (|y_i|) = |y|$, for $|(|y| - |y_i|)| \leqq d(y, y_i)$. Since $\lim (a_i + x) = a + x$, $\lim (|a_i + x|) = |a + x|$. Therefore $\lim (|a_i x|) = |a| + |x| - |a + x|$. The sequence $(a_i x)$ is a Cauchy sequence, for $d(a_i x, a_j x) \leqq d(a_i, a_j)$ since $(a_i, x)M$, $(a_j, x)M$. Since L is metrically complete, there exists a' such that $\lim (a_i x) = a'$, and therefore $\lim (|a_i x|) = |a'|$. Now $\lim (a_i + a_i x) = \lim (a_i) = a$. But since the join operation is continuous, we have $\lim (a_i + a_i x) = \lim (a_i) + \lim (a_i x) = a + a'$. Therefore $a' \leqq a$. We also have that $x = \lim (x + a_i x) = \lim (x) + \lim (a_i x) = x + a'$. Thus $a' \leqq x$ and consequently $a' \leqq ax$. Since $|ax| \leqq |x| + |a| - |a + x|$ and $|a'| = |x| + |a| - |a + x|$, it follows that $|ax| \leqq |a'|$. This implies that $a' = ax$ and the result follows.

It is to be noted that our proof requires the metric completeness of L, and the theorem is false if one does not assume metric completeness as can be shown in an example in [9]. The reason this is so is that when one metrically completes a lattice, the join operation is preserved but the meet operation need not be.

We say that a lattice with more than one element is *dense* if $x > y$ implies the existence of an element z such that $x > z > y$. A left-complemented lattice of length > 1 is dense if and only if it contains no points. It is easily shown that a maximal chain C in $[a, b]$ of a metrically complete dense lattice is isomorphic and isometric to a closed interval of real numbers.

Let L be a left-complemented lattice. If a, b are contained in $[c, d]$, and have a common independent complement relative to $[c, d]$, then they have a common independent complement relative to any interval containing $[c, d]$.

LEMMA 22. *Let L be a left-complemented lattice such that every interval sublattice is irreducible. If a and b are incomparable, then there exists an element $p > 0$ such that $(a, p) \perp$, $(b, p) \perp$.*

Proof. There exist a', b' such that $a' + b = a + b$, $(a', b) \perp$, $a' \leqq a$, $b' + a = b + a$, $(b', a) \perp$, $b' \leqq b$. Therefore $a' \neq 0$, $b' \neq 0$ and $(a', b') \perp$. Since the interval $[0, a' + b']$ is irreducible and contains more than two elements, there exists $x < a' + b'$ such that $x(a' + b') \neq xa' + xb'$. Therefore $x > xa' + xb'$. Now there exists $u \neq 0$ such that $xa' + xb' + u = x$, $(xa' + xb', u) \perp$. Since $u \leqq x$, $ua' \leqq xa'$; therefore $ua' = ua'xa' = uxa' \leqq u(xa' + xb') = 0$. Therefore $ua' = 0$ and also $ub' = 0$. But $ua = u(a' + b')a = ua' = 0$, and in a similar way $ub = 0$. We choose u' such that $u' + a = u + a$, $(u', a) \perp$, $u' \leqq u$. Consequently $u' \neq 0$. There exists p such that $p + b = u' + b$, $(p, b) \perp$, $p \leqq u'$. Quite obviously $p \neq 0$ and $(p, a) \perp$. This completes the proof.

THEOREM 17. *Let L be a left-complemented metrically complete lattice in which every interval sublattice is irreducible. If $|a| = |b|$, there exists $c \in L$ such that $a + b = a + c = b + c$, $ac = bc = ab$, $(a, c)M$, $(b, c)M$.*

Proof. Without loss of generality we assume that $ab = 0$. The case $a = b$ is trivial. Suppose $a \neq b$. Since a and b are incomparable, there exists $p > 0$, $p \leqq a + b$, such that $(a, p) \perp$, $(b, p) \perp$. Let c be a maximal element with this property.[4] If $a + c = a + b$, then $b + c = a + b$; for $|a| = |b|$, $|a + c| = |c| + |a| - |0|$, $|b + c| = |c| + |b| - |0|$, and therefore $|a + c| = |b + c|$. Suppose then that $a + b \neq a + c$, $a + b \neq b + c$. If $a + c = b + c$, then $a + c = a + b$. Thus $a + c$ and $b + c$ are incomparable, and therefore there exists $m > 0$, $m \leqq a + b$, such that $(a + c, m) \perp$, $(b + c, m) \perp$. It is easily seen that $(a, c + m) \perp$, $(b, c + m) \perp$, and this contradicts the definition of c.

Let M be an irreducible continuous geometry valuated in such a way that $|I| = 2$, $|0| = 0$. If we define L to be the set of all elements x with $|x| < 1$ plus I and valuate L so that $|I| = 1$ and all other elements have the same valuation as they do in M, then L is a lattice without points which satisfies the hypothesis of Theorem 17 and is not a continuous geometry.

We have already studied metric lattices which are modulated because every semi-modular lattice of finite length has a valuation. To facilitate our study of the irreducible modulated lattices of finite length, we introduced the condition (γ) which makes no sense if our lattice has no points. By the use of Theorem 12 one can show that all the intervals except possibly the intervals $[0, l]$ where l is a line are irreducible. We have shown that irreducible lattices which satisfy (γ) have the additional property that $ab \neq 0$ implies $(a, b)M$. It is not too difficult to show

[4] Let S be a set of elements in L where L is a metrically complete lattice with 0 and I. Theorem 16 and the fact that every increasing transfinite sequence is countable imply that there exists a maximal element independent of each element of S.

that metrically complete lattices which are dense and have this latter property must be modular lattices if we use Theorem 16. This leads one to suspect that modulated lattices which are dense, metrically complete, and which have no reducible interval sublattices are actually continuous geometries. This is the case as is seen below.

THEOREM 18. *Let L be a modulated lattice without points which is metrically complete and contains at least two elements. If every interval sublattice of L is irreducible, then L is an irreducible continuous geometry.*

Proof. We assume without loss of generality that our valuation is normalized, i.e., $|0| = 0, |I| = 1$. Since L is dense and metrically complete, there are elements of any valuation between 0 and 1; thus there are modular elements of any value between 0 and 1. We define a set S of real numbers as follows:

$S = [a \in S; a \in [0, 1]$; for every yM with $|y| \leqq a$, the interval $[0, y]$ is a modular lattice].

S is non-empty because $0 \in S$. Let ω be the least upper bound of S. Then $\omega \in S$. To prove this it suffices to consider those modular elements y for which $|y| = \omega$. There exists an increasing sequence (y_i) with limit y such that $y_i M$ for each i. Let $c \in [0, y]$. Since $cy_i M$ and $(c, y_i)M$ for every i, $\lim (cy_i) = cy = c$ and cM. Thus $[0, y]$ is a modular lattice. Suppose then that $1 - \omega \neq 0$. There exists δ with $1 > \delta > \omega$ such that $1 - \delta \geqq \delta - \omega$. To prove the theorem it suffices to show that $\delta \in S$. Let a be an M-element with $\omega < |a| \leqq \delta$. Let $c \leqq a$. There exists $q \leqq a$ such that $|q| = \omega$ and qM. Evidently qcM since $[0, q]$ is a modular lattice. Since L is a modulated lattice, there exists d' such that $a + d' = I, ad' = qc, d'M$. There exists a real number α such that $\alpha - |a| = |c| - |qc|$. Then $\alpha = |a| + |c + q| - |q|$. Thus $\alpha \leqq \delta + \delta - \omega \leqq 1$. But there exists $r \geqq a$ with rM such that $|r| = \alpha$. Since $|r| - |a| = |c| - |qc|$ and $|r| + |d'| = 1 + |rd'|$, it follows that $|d'| + |a| = 1 - |c| + |rd'| + |qc|$. Using the fact that $|d'| + |a| = 1 + |qc|$, we find that $|c| = |rd'|$. If we define $rd' \equiv d$, then $a + d = r$, $ad = qc = cd$, and dM since rM and $d'M$. From the irreducibility of the interval $[cd, c + d]$ there exists $z \in L$ such that $d + z = c + z = c + d, dz = cz = cd, (c, z)M$ and $(d, z)M$. Now $az = a(c + d)z = cz = cd$. Also $a + z = a + c + z = a + c + d = a + d = r$. Since rM and dM, there exists $b \in L$ such that $c + d + b = r, (c + d)b = d, bM$. Therefore $bz = b(c + d)z = dz = cd, b + z = b + z + d = b + c + d = r$. This shows that $|b| = |a|$. There exists $x \in L$ such that $x + cd = z$ and $(x, cd) \perp$. Then $a + x = a + cd + x = r = b + cd + x = b + x, ax = azx = cdx = 0 = bzx = bx$. The mapping $y \rightarrow (y + x)a, y \in [0, b]$, is an isomorphism from $[0, b]$ onto $[0, a]$. Now $d \rightarrow (d + x)a = (d + cd + x)a = (d + z)a =$

$(c+z)a = c$. Since c is the image of d under the isomorphism, c is modular with every element in $[0, a]$ and so cM. Thus $[0, a]$ is a modular lattice for every aM with $\omega < |a| \leqq \delta$. Hence δ is in S which is a contradiction. Thus $\omega = 1$, and the proof is complete.

Conclusion. It might be of some interest to determine the existence or non-existence of an irreducible metrically complete modulated lattice which is dense but is not a continuous geometry. Such a lattice would be a natural generalization of finite partition lattices in view of our Theorems 14 and 17. In a subsequent paper we shall show how to represent lattices satisfying (γ) in terms of projective geometries. Using this representation, we can show that every interval sublattice of a matroid modulated lattice is a modulated lattice.

BIBLIOGRAPHY

1. Garrett Birkhoff, *Lattice Theory*, Rev. Ed. New York: Amer. Math. Soc., 1948.
2. R. P. Dilworth, *The structure of relatively complemented lattices*, Ann. of Math., **51** (1950), 348-59.
3. M. L. Dubreil-Jacotin, L. Lesieur and R. Croisot, *Leçons sur la Théorie Des Treillis Des Structures Algebriques Ordonnées et Des Treillis Géométriques*. Paris: Gauthier-Villars, 1953.
4. J. E. McLaughlin, *Projectivities in relatively complemented lattices*, Duke Math. J., **18** (1951), 73-84.
5. ———, *Structure theorems for relatively complemented lattices*, Pacific J. Math., **3** (1953), 197-208.
6. O. Ore, *Theory of equivalence relations*, Duke Math. J., **9** (1942), 573-627.
7. U. Sasaki and S. Fujiwara, *The decomposition of matroid lattices*, J. Sci. Hiroshima Univ. Ser. A., **15** (1952), 183-88.
8. ———, *The characterization of partition lattices* J. Sci. Hiroshima Univ. Ser. A., **15** (1952), 189-201.
9. L. R. Wilcox and M. F. Smiley, *Metric lattices*, Ann. of Math., **40** (1939), 309-27, corrected *ibid.*, **47** (1946), 831.
10. ———, *A Note on complementation in lattices*, Bull. Amer. Math. Soc., **48** (1942), 453-58.

ILLINOIS INSTITUTE OF TECHNOLOGY
UNIVERSITY OF ILLINOIS

ABSTRACT MARTINGALE CONVERGENCE THEOREMS

FRANK S. SCALORA

Introduction. The study of probability theory in abstract spaces became possible with the introduction of integration theories in such spaces. Thus the idea of the expectation of a random variable which takes its value in a Banach space was studied by Frechet [6] with what amounted to the Bochner integral, and by Mourier [13] with the Pettis integral. Doss [2] studied the problem in a metric space. Kolmogorov [10] generalized the notion of characteristic function. Generalizations of the laws of large numbers and the ergodic theorem appear in Mourier [13] and Fortet-Mourier [5]. In this paper we generalize the concept of martingale and prove various convergence theorems.

Chapter I is devoted to listing various definitions and theorems which we shall have to refer to later. In Chapter II we introduce the idea of the conditional expectation of a Banach space valued random variable. We also prove the existence of the strong conditional expectation for strongly measurable random variables. This part of our work was also done by Moy [14] independently, and without the knowledge of the author. Chapter III is devoted to the definition and study of weak and strong $\mathfrak{X}$-martingales, with emphasis on the latter.

In Chapter IV we prove a series of convergence theorems for $\mathfrak{X}$-Martingales with the help of theorems of Doob [1]. The main theorem says that if $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-Martingale where $\mathfrak{X}$ is a reflexive Banach space, and if $\{\| x_n \|, n \geqq 1\}$ is a uniformly integrable class of functions, then there is a strongly measurable $\mathfrak{X}$-valued function x_∞ such that $\| x_n(\omega) - x_\infty(\omega) \| \to 0$ as $n \to \infty$ with probability 1 and $\{x_n, \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ is an $\mathfrak{X}$-martingale. We close by discussing examples where $\mathfrak{X}$ is one of the standard Banach spaces, l^p, $L^p(I)$, and $C(I)$.

CHAPTER I.

PRELIMINARY DEFINITIONS

1. Measurability concepts. A. Let $(\Omega, P, \mathscr{M})$ be a probability space. Thus Ω is an abstract set of points ω, $\mathscr{M}$ is a Borel field of subsets of Ω, and P is a probability measure defined on $\mathscr{M}$. We recall that a Borel field of sets is a class of sets which is closed under countable unions and intersections, and complementation. A probability

Received August 18, 1959, and in revised form April 19, 1960. Presented to the American Mathematical Society August 29, 1958. This paper is essentially the author's doctoral dissertation submitted to the University of Illinois. He wishes to express his gratitude to Professor J. L. Doob under whose guidance the thesis was written.

measure P is a completely additive non negative set function defined on a Borel field of sets, such that $P\{\Omega\} = 1$. We will be concerned with functions $x(\cdot)$ defined on Ω, and taking their values in a Banach space $\mathfrak{X}$. The sets of $\mathscr{M}$ will be referred to as the measurable sets.

DEFINITION 1.1. x is a *weak random variable* if it is a weakly measurable function from Ω to $\mathfrak{X}$.

DEFINITION 1.2. x is a *finitely* (*countably*) *valued random variable* if it is constant on each of a finite (countable) number of disjunct measurable sets Δ_j; with $\Omega = \bigcup_j \Delta_j$.

DEFINITION 1.3. x is a *strong random variable* if it is a strongly measurable function from Ω to $\mathfrak{X}$.

DEFINITION 1.4. x is *almost separably valued* if there is a set Δ in $\mathscr{M}$ such that $P\{\Delta\} = 0$ and $x(\Omega - \Delta)$ is separable.

Note. x is strongly measurable if and only if it is weakly measurable and almost separably valued. (Pettis [15] and Hille-Phillips [9] Theorem 3.5.3, p. 72).

B. The measure induced in $\mathfrak{X}$. Suppose x is a function from Ω to $\mathfrak{X}$. We define a class of subsets of $\mathfrak{X}$ in the following way: Let $\mathscr{F}$ be a Borel field of measurable subsets of Ω, $\mathscr{F} \subseteqq \mathscr{M}$. Let $\mathscr{F}_x$ be the class of subsets of $\mathfrak{X}$ with the property that $\mathscr{A} \in \mathscr{F}_x$ if $\mathscr{A} \subseteqq X$ and $\{\omega: x(\omega) \in \mathscr{A}\}$ is an $\mathscr{F}$ set. $\mathscr{F}_x$ is a Borel field.
If $\mathscr{A} \in \mathscr{F}_x$ define $P^x\{\mathscr{A}\} = P\{\omega: x(\omega) \in \mathscr{A}\}$. Clearly P^x is a probability measure on $\mathscr{F}_x$. This gives us a probability triple on $\mathfrak{X}$, $(\mathfrak{X}, P^x, \mathscr{F}_x)$. Now, let $\mathscr{F} = \mathscr{M}$, the class of measurable sets of Ω. In order to assure that $\mathscr{M}_x$ will contain some interesting subsets of $\mathfrak{X}$ we shall have to assume some measurability properties for x, which we now proceed to do.

a. Suppose that x is weakly measurable. Then $f(x)$ is a real measurable function for all $f \in \mathfrak{X}^*$, the real first conjugate space of $\mathfrak{X}$. Thus for every real Borel set B, $\{\omega: f(x(\omega)) \in B\}$ is an $\mathscr{M}$ set. Next $\{\omega: f(x(\omega)) \in B\} = \{\omega: x(\omega) \in f^{-1}(B)\}$. Hence $f^{-1}(B)$ is in $\mathscr{M}_x$ for every f in $\mathfrak{X}^*$ and real Borel set B. Since f is continuous, $f^{-1}(B)$ is open (closed) if B is open (closed).

Further, $\mathscr{M}_x$ contains all the weak neighborhoods of $\mathfrak{X}$ if x is weakly measurable. In fact, let $N(\xi_0; f_1, \cdots, f_n; \varepsilon)$ be a weak neighborhood of $\mathfrak{X}$. Then

$$N(\xi_0; f_1, \cdots, f_n; \varepsilon) = \{\xi: |f_j(\xi) - f_j(\xi_0)| < \varepsilon, \quad j = 1, \cdots, n\}$$
$$= \bigcap_{j=1}^{n} \{\xi: |f_j(\xi) - f_j(\xi_0)| < \varepsilon\} .$$

But the inverse image of each of the sets in the intersection by x is clearly an $\mathscr{M}$ set since $f(x)$ is a real valued measurable function for every linear functional f. Thus $\mathscr{M}_x$ contains all of the weak neighborhoods of $\mathfrak{X}$, and hence the smallest Borel field containing the weak neighborhoods.

Conversely, if $\mathscr{M}_x$ contains all the weak neighborhoods of $\mathfrak{X}$ then x is weakly measurable. To prove this, we must show that $f(x)$ is a real valued measurable function on Ω for every f in $\mathfrak{X}^*$. If f is the zero functional then $f(x(\omega)) = 0$ for all ω, and thus $f(x)$ is clearly measurable. Otherwise f takes on all real values. In this case we show that $\{\omega: f(x(\omega)) \in B\}$ is an $\mathscr{M}$ set for every real Borel set B and linear functional f. If B is the open interval $(a - \varepsilon, a + \varepsilon)$, then $\{\omega: f(x(\omega)) \in B\} = \{\omega: |f(x(\omega)) - a| < \varepsilon\}$. Since f takes on all real values there is an element ξ_0 in $\mathfrak{X}$ such that $f(\xi_0) = a$. Hence $\{\omega: f(x(\omega)) \in B\} = \{\omega: x(\omega) \in N(\xi_0; f; \varepsilon)\}$ which is an $\mathscr{M}$ set by hypothesis for $\mathscr{M}_x$ contains all the weak neighborhoods of $\mathfrak{X}$. Next, every open set in the reals, in fact, in any separable metric space, is a countable union of open spheres. Thus, if B is an open set in the reals $B = \bigcup_n V_n$ where V_n is an open interval for every n. Since $\mathscr{M}$ is closed under countable unions $\{\omega: f(x(\omega)) \in B\} = \bigcup_n \{\omega: f(x(\omega)) \in V_n\}$ is an $\mathscr{M}$ set. Finally, the class of real sets B for which $\{\omega: f(x(\omega)) \in B\}$ is an $\mathscr{M}$ set is a Borel field which contains the open sets, thus it must contain all the real Borel sets, and so x is weakly measurable. Thus the definition of weak measurability may be rephrased as follows:

DEFINITION 1.1.* x is weakly measurable if $\mathscr{M}_x$ contains all the weak neighborhoods of $\mathfrak{X}$, that is, if $\{\omega: x(\omega) \in N\}$ is an $\mathscr{M}$ set for every weak neighborhood N.

b. Suppose that x is strongly measurable. Then there is a sequence x_n of finitely valued functions, and a set Λ in $\mathscr{M}$ such that $P\{\Lambda\} = 0$, $\|x_n(\omega) - x(\omega)\| \to 0$ as $n \to \infty$ for $\omega \in \Omega - \Lambda$. Let g be a real valued continuous function. Then $g(x)$ is a real valued measurable function on Ω. Consequently, $\{\omega: g(x(\omega)) \in B\}$ is an $\mathscr{M}$ set and $g^{-1}(B)$ is an $\mathscr{M}_x$ set for every real Borel set B and real continuous function g. Next let $\mathscr{C}$ be the class of real valued functions g defined on $\mathfrak{X}$ such that $g(x)$ is a real valued measurable function on Ω. Then $\mathscr{C}$ contains the continuous functions and is closed under the limit operation, thus it contains all the Baire functions on $\mathfrak{X}$ to the reals. Now let A be a Borel set in $\mathfrak{X}$. Then there is a real number a and a Baire function g

such that $A = \{\xi: g(\xi) > a\}$. Now $A = g^{-1}(B)$ where $B = (a, \infty)$. Thus A is an $\mathscr{M}_x$ set since $\{\omega: g(x(\omega)) \in B\}$ is an $\mathscr{M}$ set by the measurability of $g(x)$. Therefore if x is strongly measurable, then $\mathscr{M}_x$ contains all the Borel sets of $\mathfrak{X}$, or $\{\omega: x(\omega) \in B\}$ is an $\mathscr{M}$ set for every Borel set B of $\mathfrak{X}$.

C. Independence. Let x and y be (weakly or strongly) measurable random variables on Ω to $\mathfrak{X}$. We can then define a Borel field $\mathscr{M}_{x,y}$ of subsets $\mathfrak{X} \times \mathfrak{X}$ in an analogous way. Consider $\mathscr{M}_x \times \mathscr{M}_y = \{A \times B: A \in \mathscr{M}_x, B \in \mathscr{M}_y\}$. Let $P^{x,y}(A \times B) = P\{\omega: x(\omega) \in A, y(\omega) \in B\}$. This probability is well defined for the set on the right is the intersection of two $\mathscr{M}$ sets and hence is itself an $\mathscr{M}$ set. Let $R_{x,y}$ be the field of finite unions of sets of $\mathscr{M}_x \times \mathscr{M}_y$. Then $P^{x,y}$ can be defined on $R_{x,y}$ to be a probability measure in the obvious way in a unique fashion. Next $P^{x,y}$ can be extended uniquely to $\mathscr{M}_{x,y}$, the smallest Borel field of measurable subsets of $\mathfrak{X} \times \mathfrak{X}$ containing $R_{x,y}$ (Doob [1] Theorem 2.2, p. 605).

DEFINITION 1.5. x and y are said to be independent if $P\{\omega: x(\omega) \in A, y(\omega) \in B\} = P\{\omega: x(\omega) \in A\}P\{\omega: y(\omega) \in B\}$ for A, B subsets of $\mathfrak{X}$ whenever all of the probabilities in the equality are defined; i.e., whenever the above sets are in $\mathscr{M}$. The equality may be rewritten as $P^{x,y}(A \times B) = P^x(A)P^y(B)$.

Notice that this definition can be rephrased to say that the product relationship holds whenever A is in $\mathscr{M}_x$ and B is in $\mathscr{M}_y$, for only then will all of the probabilities in the product be defined. This is the type of definition that has been given by Kolmogorov; e.g., Gnedenko-Kolmogorov ([7], p. 26). The definition used by Doob [1] differs in that it says that the product relationship holds whenever A and B belong to a possibly smaller class of sets, namely the Borel sets. For a full discussion of the connection between the two types of definition the reader is referred to Doob's appendix to the above mentioned book by Gnedenko and Kolmogorov.

THEOREM 1.1. *If x and y are independent, then $f_1(x), \cdots, f_n(x)$ are independent of $g_1(y), \cdots, g_m(y)$ in the sense of Kolmogorov for every finite set of real valued linear functionals $f_1, \cdots, f_n, g_1, \cdots, g_m$ on $\mathfrak{X}$.*

Proof. Let $A_1, \cdots, A_n, B_1, \cdots, B_m$ be real sets such that $\{\omega: f_j(x(\omega)) \in A_j\}$ and $\{\omega: g_k(y(\omega)) \in B_k\}$ are $\mathscr{M}$ sets for $j = 1, \cdots, n$ and $k = 1, \cdots, m$. Then $f_j^{-1}(A_j)$ is in $\mathscr{M}_x$ and $g_k^{-1}(B_k)$ is in $\mathscr{M}_y$. Next, $\bigcap_{j=1}^n f_j^{-1}(A_j) \in \mathscr{M}_x$ and $\bigcap_{k=1}^m g_k^{-1}(B_k) \in \mathscr{M}_y$. Thus

$$\begin{aligned} P\{\omega: f_1(x(\omega)) \in A_1, \cdots, f_n(x(\omega)) \in A_n, g_1(y(\omega)) \in B_1, \cdots, g_m(y(\omega)) \in B_m\} \\ = P\{\omega: x(\omega) \in \bigcap_{j=1}^{n} f_j^{-1}(A_j), y(\omega) \in \bigcap_{k=1}^{m} g_k^{-1}(B_k)\} \\ = P\{\omega: x(\omega) \in \bigcap_{j=1}^{n} f_j^{-1}(A_j)\} P\{\omega: y(\omega) \in \bigcap_{k=1}^{m} g_k^{-1}(B_k)\} \end{aligned}$$

by the independence of x and y

$$= P\{\omega: f_1(x(\omega)) \in A_1, \cdots, f_n(x(\omega)) \in A_n\} P\{\omega: g_1(y(\omega)) \in B_1, \cdots, g_m(y(\omega)) \in B_m\} \quad \text{Q.E.D.}$$

THEOREM 1.2. *If x and y are weakly measurable and independent, then $f_1(x), \cdots, f_n(x)$ are independent of $g_1(y), \cdots, g_m(y)$ in the sense of Doob for every finite set of real valued linear functionals $f_1, \cdots, f_n, g_1, \cdots, g_m$ on $\mathfrak{X}$.*

Proof. Let A_j and B_k in the above proof be real Borel sets; then $\{\omega: f_j(x(\omega)) \in A_j\}$ and $\{\omega: g_k(y(\omega)) \in B_k\}$ are $\mathscr{M}$ sets for $f_j(x)$ and $g_k(y)$ are real valued measurable functions by the weak measurability of x and y. The rest of the proof goes as above.

THEOREM 1.3. *If x and y are weakly measurable, and such that $f_1(x), \cdots, f_n(x)$ are independent of $g_1(y), \cdots, g_m(y)$ for every finite set of real valued linear functionals $f_1, \cdots, f_n, g_1, \cdots, g_m$ on $\mathfrak{X}$, then x and y are independent relative to the smallest Borel field of $\mathfrak{X}$ sets containing the weak neighborhoods;* i.e.,

$$P\{\omega: x(\omega) \in A, y(\omega) \in B\} = P\{\omega: x(\omega) \in A\} P\{\omega: y(\omega) \in B\}$$

for all A and B in the smallest Borel field containing the weak neighborhoods of $\mathfrak{X}$.

Proof. Let $A = N(\xi_0; f_1, \cdots, f_n; \varepsilon)$ and $B = N(\eta_0; g_1, \cdots, g_m; \delta)$: then

$$\begin{aligned} P\{\omega: x(\omega) \in A, y(\omega) \in B\} \\ = P\{\omega: |f_i(x(\omega)) - f_i(\xi_0)| < \varepsilon, \quad i = 1, \cdots, n; \\ |g_j(y(\omega)) - g_j(\eta_0)| < \delta, \quad j = 1, \cdots, m\} \\ = P\{\omega: |f_i(x(\omega)) - f_i(\xi_0)| < \varepsilon, \quad i = 1, \cdots, n\} \\ P\{\omega: |g_j(y(\omega)) - g_j(\eta_0)| < \delta, \quad j = 1, \cdots, m\} \end{aligned}$$

by the hypothesis, and so

$$P\{\omega: x(\omega) \in A, y(\omega) \in B\} = P\{\omega: x(\omega) \in A\} \cdot P\{\omega: y(\omega) \in B\}$$

when A and B are weak neighborhoods of $\mathfrak{X}$. Now the class of weak neighborhoods is closed under finite intersections and thus the independence multiplicative relationship is preserved if we extend this class to the smallest Borel field containing it (Loève [12] p. 225).

The notion of independence is easily generalized to aggregates of random variables. For a fuller discussion of the measurability concepts mentioned in this section, see Pettis [15] and Hille and Phillips [9].

Note. Let $(\xi, \eta) \in \mathfrak{X} \times \mathfrak{X}$. Define $\|(\xi, \eta)\| = \sqrt{\|\xi\|^2 + \|\eta\|^2}$. By this definition, $\mathfrak{X} \times \mathfrak{X}$ becomes a Banach space. Let f be a real linear functional on $\mathfrak{X} \times \mathfrak{X}$. If $f_1(\xi) = f[(\xi, \theta)]$ and $f_2(\eta) = f[(\theta, \eta)]$, then f_1 and f_2 are real linear functionals on $\mathfrak{X}$, and $f[(\xi, \eta)] = f_1(\xi) + f_2(\eta)$. If x and y are weakly measurable $\mathfrak{X}$-valued functions on Ω, then $f_1(x)$ and $f_2(y)$ are real valued measurable functions on Ω. Thus the weak measurability of x and y implies the weak measurability of (x, y) on Ω to $\mathfrak{X} \times \mathfrak{X}$. Similarly, if x and y are strongly measurable, there exist sequences x_n and y_n of finitely-valued measurable $\mathfrak{X}$-valued functions such that $\|x_n - x\| \to 0$ and $\|y_n - y\| \to 0$ as $n \to \infty$ with probability 1. But (x_n, y_n) gives a sequence of $\mathfrak{X} \times \mathfrak{X}$ finitely-valued functions, and $\|(x_n, y_n) - (x, y)\| = \sqrt{\|x_n - x\|^2 + \|y_n - y\|^2} \to 0$ with probability 1 as $n \to \infty$. Thus, if x and y are strongly measurable, then so is (x, y).

2. Integrability concepts. Let x be a countably valued function taking the value ξ_j on the measurable set Δ_j. Then x is said to be Bochner integrable if and only if $\|x(\cdot)\|$ is integrable, and by definition

$$(B) \int_\Omega x(\omega) dP = \sum_{j=1}^{\infty} \xi_j P(\Delta_j) .$$

DEFINITION 2.1. $x(\cdot)$ is integrable in the sense of Bochner if there is a sequence $x_n(\cdot)$ of countably valued random variables converging with probability 1 to $x(\cdot)$, and such that

$$\lim_{m,n\to\infty} \int_\Omega \|x_m(\omega) - x_n(\omega)\| dP = 0 .$$

Then the limit of $(B)\int_\Omega x_n(\omega)dP$ exists and by definition

$$(B)\int_\Omega x(\omega)dP = \lim_{n\to\infty} (B) \int_\Omega x_n(\omega)dP .$$

Since $P\{\Omega\} = 1$, we may again replace the word countably by finitely.

We will later need the following result apparently proved first by Pettis ([15] Theorem 5.2, p. 293), and later by Moy ([14] Theorem 1, pp. 3, 4.)

THEOREM 2.1. *If x is strongly measurable relative to the Borel field*

$\mathscr{F}$ of measurable sets and Bochner integrable and such that $\int_A x(\omega)dP = \theta$ for every set A in $\mathscr{F}$ then $x(\omega) = \theta$ almost everywhere.

CHAPTER II

GENERALIZATIONS OF THE RADON-NIKODYM THEOREM AND ABSTRACT CONDITIONAL EXPECTATIONS

1. It is well known that a real or complex valued completely additive set function which is absolutely continuous on a σ-finite measure space is actually the integral in the usual sense of a finite measurable point function (unique almost everywhere). The existence of this point function is assured by the classical Radon-Nikodym theorem (Halmos [8] p. 128).

Using a theorem due to Dunford and Pettis ([4], p. 339) it is possible to get a definition of conditional expectations for more general random variables such as Dunford and Pettis integrable functions. Since it is too weak for our purposes, we will no longer refer to it in this paper.

2. Strong conditional expectations. If we restrict ourselves to Bochner integrable random variables it is possible to get a sharper version of the conditional expectation.

With this end in mind, let $x(\cdot): \Omega \to \mathfrak{X}$ be finitely valued; in fact, let $x(\omega) = \xi_j$ on A_j; $j = 1, \cdots, k$. Then $x(\omega) = \sum_{j=1}^{k} \xi_j \cdot \chi_{A_j}(\omega)$ where χ_{A_j} is the characteristic function of A_j.

DEFINITION 2.1. $\mathscr{E}^s\{x \mid \mathscr{F}\}(\omega) = \sum_{j=1}^{k} \xi_j \cdot E\{\chi_{A_j}/\mathscr{F}\}(\omega)$, where $E\{\chi_{A_j} \mid \mathscr{F}\}$ is the ordinary conditional expectation (Doob [1]) of χ_{A_j} relative to $\mathscr{F}$. $\mathscr{E}^s\{x \mid \mathscr{F}\}$ will be referred to as the strong conditional expectation of x relative to $\mathscr{F}$.

In this section all integrals will be in the sense of Bochner, so we will remove the letter B preceding the integral sign.

LEMMA 2.1. *If x is a measurable finitely valued function on Ω to $\mathfrak{X}$, then $\int_A x(\omega)dP = \int_A \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega)dP$ for every $A \in \mathscr{F}$.*

Proof.

$$\int_A \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega)dP = \int_A \Big(\sum_{j=1}^{k} \xi_j E\{\chi_{A_j} \mid \mathscr{F}\}(\omega)\Big)dP$$
$$= \sum_{j=1}^{k} \xi_j \int_A E\{\chi_{A_j} \mid \mathscr{F}\}(\omega)dP$$

where the integral is in the ordinary sense

$$= \sum_{j=1}^{k} \xi_j P\{A_j \cap A\}$$
$$= \int_A x(\omega) dP \,. \qquad \text{Q.E.D.}$$

LEMMA 2.2. *If x is a measurable finitely valued function on Ω to $\mathfrak{X}$, then $\| \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega) \| \leqq E\{\|x\| \mid \mathscr{F}\}\,(\omega)$ with probability* 1.

Proof.

$$\| \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega) \| = \left\| \sum_{j=1}^{k} \xi_j E\{\chi_{A_j} \mid \mathscr{F}\}(\omega) \right\|$$
$$\leqq \sum_{j=1}^{k} \| \xi_j \| \cdot E\{\chi_{A_j} \mid \mathscr{F}\}(\omega)$$

for $\chi_{A_j} = 1$ or 0.

$$= E\{\|x\| \mid \mathscr{F}\}(\omega) \,. \qquad \text{a.e.} \qquad \text{Q.E.D.}$$

LEMMA 2.3. *If $x_1, \cdots, x_k$ are finitely valued measurable functions, and $a_1 \cdots, a_k$ are scalars, then*

$$\mathscr{E}^s\{a_1 x_1 + \cdots + a_k x_k \mid \mathscr{F}\}(\omega) = \sum_{j=1}^{k} a_j \mathscr{E}^s\{x_j \mid \mathscr{F}\}(\omega)$$

with probability 1.

Proof. Let $\{A_m\}: m = 1, \cdots, p$ be a decomposition of Ω such that each x_j takes on only one value on each A_m; in fact, let $x_j(\omega) = \varphi_j(A_m)$ for $\omega \in A_m$. Then since $\mathscr{E}^s\{x \mid \mathscr{F}\}$ depends on x and $\mathscr{F}$ and not on the decomposition of Ω, the same representation holds for all the $\mathscr{E}^s\{x_j \mid \mathscr{F}\}$. Hence

$$\mathscr{E}^s\{x_j \mid \mathscr{F}\}(\omega) = \sum_{m=1}^{p} \varphi_j(A_m) E\{\chi_{A_m} \mid \mathscr{F}\}(\omega) \,.$$

Thus

$$\mathscr{E}^s\{a_1 x_1 + \cdots + a_k x_k \mid \mathscr{F}\}(\omega)$$
$$= \sum_{m=1}^{p} [a_1 \varphi_1(A_m) + \cdots + a_k \varphi_k(A_m)] E\{\chi_{A_m} \mid \mathscr{F}\}(\omega)$$
$$= a_1 \sum_{m=1}^{p} \varphi_1(A_m) E\{\chi_{A_m} \mid \mathscr{F}\}(\omega) + \cdots + a_k \sum_{m=1}^{p} \varphi_k(A_m) E\{\chi_{A_m} \mid \mathscr{F}\}(\omega)$$
$$= \sum_{j=1}^{k} a_j \mathscr{E}^s\{x_j \mid \mathscr{F}\}(\omega) \qquad \text{with probability 1.} \qquad \text{Q.E.D.}$$

THEOREM 2.1. *Let $x(\cdot): \Omega \to \mathfrak{X}$ be integrable in the sense of Bochner and $\mathscr{F}$ a Borel field of measurable Ω sets. Then there exists a func-*

tion $\mathscr{E}^s\{x \mid \mathscr{F}\}(\cdot)\colon \Omega \to \mathfrak{X}$ *which is Bochner integrable, strongly measurable relative to* $\mathscr{F}$, *unique* a.e., *and*

$$\int_\Lambda x(\omega)dP = \int_\Lambda \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega)dP \textit{ for all } \Lambda \in \mathscr{F} .$$

Proof. Let x be strongly measurable and integrable in the sense of Bochner. Then there exists a sequence x_n of finitely valued measurable functions such that $x_n(\omega) \to x(\omega)$ with probability 1 as $n \to \infty$; $\int_\Omega \| x_n(\omega) - x_m(\omega) \| dP \to 0$ as $n, m \to \infty$; and $\int_\Omega x_n(\omega)dP \to \int_\Omega x(\omega)dP$. Now $\mathscr{E}^s\{x_n \mid \mathscr{F}\}$ is defined for all x_n by Definition 2.1. Also

$$\begin{aligned}
&\int_\Omega \| \mathscr{E}^s\{x_n \mid \mathscr{F}\}(\omega) - \mathscr{E}^s\{x_m \mid \mathscr{F}\}(\omega) \| \, dP \\
&= \int_\Omega \| \mathscr{E}^s\{x_n - x_m \mid \mathscr{F}\}(\omega) \| \, dP && \text{by Lemma 2.3.} \\
&\leqq \int_\Omega E\{\| x_n - x_m \| \mid \mathscr{F}\}(\omega)dP && \text{by Lemma 2.2.} \\
&= \int_\Omega \| x^n(\omega) - x_m(\omega) \| \, dP && \text{by the definition of ordinary conditional expectations}
\end{aligned}$$

$\to 0$ by the defining property of the x_n's as $n, m \to \infty$.
Then according to Hille and Phillips ([9] p. 82, Theorem 3.7.7), there exists a function, y, which is Bochner integrable, strongly measurable relative to $\mathscr{F}$, unique a.e., and such that

$$\int_\Omega \| \mathscr{E}^s\{x_n \mid \mathscr{F}\}(\omega) - y(\omega) \| \, dP \to 0 \qquad \text{as } n \to \infty . \tag{1}$$

Next,

$$\begin{aligned}
&\left\| \int_\Lambda y(\omega)dP - \int_\Lambda x(\omega)dP \right\| \\
&\quad = \left\| \int_\Lambda y(\omega)dP - \int_\Lambda \mathscr{E}^s\{x_n \mid \mathscr{F}\}(\omega)dP \right. \\
&\qquad\qquad \left. + \int_\Lambda \mathscr{E}^s\{x_n \mid \mathscr{F}\}(\omega)dP - \int_\Lambda x(\omega)dP \right\| \\
&\quad \leqq \int_\Lambda \| y(\omega) - \mathscr{E}^s\{x_n \mid \mathscr{F}\}(\omega) \| \, dP \\
&\qquad\qquad + \left\| \int_\Lambda x_n(\omega)dP - \int_\Lambda x(\omega)dP \right\| \qquad \text{by Lemma 2.1.}
\end{aligned}$$

$\to 0$ as $n \to \infty$ by (1) above and by the definition of $\int_\Lambda x(\omega)dP$. Thus $\int_\Lambda y(\omega)dP = \int_\Lambda x(\omega)dP$ for all $\Lambda \in \mathscr{F}$. We are now justified in calling $y(\cdot)$ the strong conditional expectation of x relative to $\mathscr{F}$ and we use the notation $\mathscr{E}^s\{x \mid \mathscr{F}\}(\cdot)$. Q.E.D.

DEFINITION 2.2. $\mathscr{E}^s\{x \mid \mathscr{F}\}$ is called the strong conditional expectation of x relative to $\mathscr{F}$.

We shall now examine the properties of the strong conditional expectation. In what follows we will be concerned mainly with the strong rather than the weak conditional expectation.

THEOREM 2.2.

1. *If* $x(\omega) = \xi$ *on* Ω *then* $\mathscr{E}^s\{x \mid \mathscr{F}\}(\omega) = \xi$ *with probability* 1.
2. $\mathscr{E}^s\left\{\sum_{j=1}^n c_j x_j \mid \mathscr{F}\right\} = \sum_{j=1}^n c_j \mathscr{E}^s\{x_j \mid \mathscr{F}\}$ *with probability* 1.
3. $\| \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega) \| \leqq E\{\| x \| \mid \mathscr{F}\}$ *with probability* 1.
4. *If* $\| x_n(\omega) - x(\omega) \| \to 0$ *as* $n \to \infty$ *with probability* 1, *and there is a real random variable* $a(\omega) \geqq 0$ *such that* $\| x_n(\omega) \| \leqq a(\omega)$ *with probability* 1 *and* $E\{a\} < \infty$, *then* $\lim_{n\to\infty} \mathscr{E}^s\{x_n \mid \mathscr{F}\} = \mathscr{E}^s\{x \mid \mathscr{F}\}$ *with probability* 1.

Proof.

(1) The function $x(\omega) = \xi$ has the defining property of $\mathscr{E}^s\{x \mid \mathscr{F}\}$ and is measurable relative to any Borel field $\mathscr{F}$.

(2) $$\int_\Lambda \mathscr{E}^s\left\{\sum_{j=1}^n c_j x_j \mid \mathscr{F}\right\}(\omega) dP = \int_\Lambda \left(\sum_{j=1}^n c_j x_j(\omega)\right) dP \qquad \text{by Theorem 2.1.}$$
$$= \int_\Lambda \left(\sum_{j=1}^n c_j \mathscr{E}^s\{x_j \mid \mathscr{F}\}(\omega)\right) dP \text{ for all } \Lambda \in \mathscr{F}.$$

Thus

$$\mathscr{E}^s\left\{\sum_{j=1}^n c_j x_j \mid \mathscr{F}\right\} = \sum_{j=1}^n c_j \mathscr{E}^s\{x_j \mid \mathscr{F}\} \text{ with probability 1.}$$

(3) Let x_n be as in the proof of Theorem 2.1. and let $\Lambda \in \mathscr{F}$. Now $\| \mathscr{E}^s\{x_n \mid \mathscr{F}\}(\omega) \| \leqq E\{\| x_n \| \mid \mathscr{F}\}(\omega)$ with probability 1 by Lemma 2.2. Thus

$$\int_\Lambda \| \mathscr{E}^s\{x_n \mid \mathscr{F}\}(\omega) \| dP \leqq \int_\Lambda E\{\| x_n \| \mid \mathscr{F}\}(\omega) dP .$$

But

$$\int_\Lambda \| \mathscr{E}^s\{x_n \mid \mathscr{F}\}(\omega) \| dP \to \int_\Lambda \| \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega) \| dP \qquad \text{as } n \to \infty$$

by Theorem 2.1., and

$$\int_\Lambda E\{\| x_n \| \mid \mathscr{F}\}(\omega) dP = \int_\Lambda \| x_n(\omega) \| dP \to \int_\Lambda \| x(\omega) \| dP$$
$$= \int_\Lambda E\{\| x \| \mid \mathscr{F}\}(\omega) dP .$$

Hence

$$\int_\Lambda \| \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega) \| dP \leqq \int_\Lambda E\{\| x \| \mid \mathscr{F}\}(\omega) dP \text{ for } \Lambda \in \mathscr{F} ,$$

and thus $\| \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega) \| \leqq E\{\| x \| \mid \mathscr{F}\}(\omega)$ with probability 1.

(4) $\| \mathscr{E}^s\{x_n \mid \mathscr{F}\}(\omega) - \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega) \|$
$= \| \mathscr{E}^s\{x_n - x \mid \mathscr{F}\}(\omega) \|$ by (2) with probabilility 1.
$\leqq E\{\| x_n - x \| \mid \mathscr{F}\}(\omega)$ by (3) with probability 1.
$\rightarrow 0$ as $n \rightarrow \infty$ by Doob ([1] p. 23). Q.E.D.

Next it will be convenient to show that every linear transformation distributes over $\mathscr{E}^s$.

THEOREM 2.3. *Let x be Bochner integrable, $\mathscr{F}$ a Borel field of measurable sets, f a linear (bounded) transformation from $\mathfrak{X}$ to another Banach space $\mathfrak{Y}$. Then*

$$f[\mathscr{E}^s\{x \mid \mathscr{F}\}(\omega)] = \mathscr{E}^s\{f(x) \mid \mathscr{F}\}(\omega) \textit{ with probability } 1.$$

Proof. Since f is a linear (bounded) transformation, $f(x)$ and $f[\mathscr{E}^s\{x \mid \mathscr{F}\}]$ are Bochner integrable (Hille-Phillips [9] p. 84). Let $\Lambda \in \mathscr{F}$. Then

$$(B)\int_\Lambda f[\mathscr{E}^s\{x \mid \mathscr{F}\}(\omega)]dP = f[(B)\int_\Lambda \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega)dP]$$

(Hille-Phillips [9] Theorem 3.7.12, p. 83)

$$= f[(B)\int_\Lambda x(\omega)dP]$$

$$= (B)\int_\Lambda f(x(\omega))dP \qquad \text{by the preceding reference}$$

$$= (B)\int_\Lambda \mathscr{E}^s\{f(x) \mid \mathscr{F}\}(\omega)dP .$$

Thus $f[\mathscr{E}^s\{x \mid \mathscr{F}\}(\omega)] = \mathscr{E}^s\{f(x) \mid \mathscr{F}\}(\omega)$ with probability 1 by Theorem 2.1. of Chapter I. Q.E.D.

COROLLARY. *Let x be Bochner integrable, $\mathscr{F}$ a Borel field of measurable Ω sets, $f \in \mathfrak{X}^*$, then*

$$f[\mathscr{E}^s\{x \mid \mathscr{F}\}(\omega)] = E\{f(x) \mid \mathscr{F}\}(\omega) \textit{ with probability } 1.$$

A final remark. If $\mathscr{F} \subseteqq \mathscr{S}$, *then*

$$\mathscr{E}^s\{\mathscr{E}^s\{x \mid \mathscr{F}\} \mid \mathscr{S}\} = \mathscr{E}^s\{\mathscr{E}^s\{x \mid \mathscr{S}\} \mid \mathscr{F}\} = \mathscr{E}^s\{x \mid \mathscr{F}\}$$

with probability 1. *For*

$$\int_A \mathscr{E}^s\{\mathscr{E}^s\{x \mid \mathscr{F}\} \mid \mathscr{S}\}(\omega)dP = \int_A \mathscr{E}^s\{x \mid \mathscr{F}\}(\omega)dP \text{ for } A \in \mathscr{S}$$

$$\begin{aligned}\int_A \mathscr{E}^s\{\mathscr{E}^s\{x \mid \mathscr{S}\} \mid \mathscr{F}\}(\omega)dP &= \int_A \mathscr{E}^s\{x \mid \mathscr{S}\}(\omega)dP \text{ for } A \in \mathscr{F} \\ &= \int_A x(\omega)dP \text{ for } A \in \mathscr{S};\ \therefore \text{ also for } A \in \mathscr{F} \\ &= \int_A E\{x \mid \mathscr{F}\}(\omega)dP \text{ for } A \in \mathscr{F}. \qquad \text{Q.E.D.}\end{aligned}$$

CHAPTER III.

ABSTRACT MARTINGALES

1. Preliminary definitions.

DEFINITION 1.1. Let T be a linear index set. Let $x_\tau(\cdot)\colon \Omega \to \mathfrak{X}$ be integrable in the sense of Bochner for $\tau \in T$ and $\mathscr{F}_\tau$ be a Borel field of measurable subsets of Ω for $\tau \in T$. Let $\mathscr{F}_\sigma \subset \mathscr{F}_\tau$ if $\sigma < \tau$. Suppose x_τ is strongly measurable relative to $\mathscr{F}_\tau$ or equal almost everywhere to such a function. If $\mathscr{E}^s\{x_\tau \mid \mathscr{F}_\sigma\} = x_\sigma$ with probability 1 when $\sigma < \tau$ then $\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ is a *strong $\mathfrak{X}$-martingale.*

In most of our work we will be concerned with the case in which T is the set of positive integers, and in this case the martingale will be denoted by $\{x_n, \mathscr{F}_n, n \geqq 1\}$ and the martingale equality becomes $\mathscr{E}^s\{x_n \mid \mathscr{F}_m\} = x_m$ with probability 1 for $n > m$.

By using the Dunford-Pettis Theorem alluded to in Chapter II, it is possible to get a definition of weak $\mathfrak{X}$-martingales, but because of a separability assumption in the theorem, they turn out to be strong X-martingales.

2. General properties of strong $\mathfrak{X}$-martingales.

From this point we will denote $(B)\int_A x(\omega)dP$ by $\int_A x(\omega)dP$, $(B)\int_\Omega x(\omega)dP$ by $\mathscr{E}\{x\}$, and $\mathscr{E}^s\{x \mid \mathscr{F}\}$ by $\mathscr{E}\{x \mid \mathscr{F}\}$, and omit the word strong when discussing strong martingales.

THEOREM 2.1. *$\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ is an $\mathfrak{X}$-martingale if and only if $\int_A x_\tau(\omega)dP = \int_A x_\sigma(\omega)dP$ for $\sigma < \tau$ and A in $\mathscr{F}_\sigma$.*

Proof. If $\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ is an $\mathfrak{X}$-martingale, then $\mathscr{E}\{x_\tau \mid \mathscr{F}_\sigma\} = x_\sigma$ with probability 1. Thus for every A in $\mathscr{F}_\sigma$ we have the equality

$$\int_A x_\sigma(\omega)dP = \int_A \mathscr{E}\{x_\tau \mid \mathscr{F}_\sigma\}(\omega)dP = \int_A x_\tau(\omega)dP ,$$

the last equality following from the definition of conditional expectations. Conversely, if $\int_A x_\tau(\omega)dP = \int_A x_\sigma(\omega)dP$, for A in $\mathscr{F}_\sigma, \sigma < \tau$, then $\int_A \mathscr{E}\{x_\tau \mid \mathscr{F}_\sigma\}(\omega)dP = \int_A x_\sigma(\omega)dP$. Therefore, $\mathscr{E}\{x_\tau \mid \mathscr{F}_\sigma\} = x_\sigma$ with probability 1 by Theorem 2.1 of Chapter I, and hence the process in question is an $\mathfrak{X}$-martingale.

THEOREM 2.2. *If $\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ is an $\mathfrak{X}$-martingale, and f is a linear* (continuous) *transformation from $\mathfrak{X}$ to another Banach space $\mathfrak{Y}$, then $\{f(x_\tau), \mathscr{F}_\tau, \tau \in T\}$ is a $\mathfrak{Y}$-martingale. Thus, in particular, the conclusion is true for every f in $\mathfrak{X}^*$. On the other hand, if $\{f(x_\tau), \mathscr{F}_\tau, \tau \in T\}$ is a real martingale for every f in $\mathfrak{X}^*$, and the x_τ are Bochner integrable, then $\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ is an $\mathfrak{X}$-martingale.*

Proof.

(1) x_τ is strongly measurable relative to $\mathscr{F}_\tau$; thus $f(x_\tau)$ is also strongly measurable relative to $\mathscr{F}_\tau$ by the continuity of f. Next, $\mathscr{E}\{f(x_\tau) \mid \mathscr{F}_\sigma\}(\omega) = f[\mathscr{E}\{x_\tau \mid \mathscr{F}_\sigma\}(\omega)]$ with probability 1 by Theorem 2.3 of Chapter II, where both sides of the equality are in $\mathfrak{Y}$. The expression on the right is equal to $f(x_\sigma(\omega))$ with probability 1 by the definition of $\mathfrak{X}$-martingale. Hence, $\mathscr{E}\{f(x_\tau) \mid \mathscr{F}_\sigma\}(\omega) = f(x_\sigma(\omega))$ with probability 1; thus, $\{f(x_\tau), \mathscr{F}_\tau, \tau \in T\}$ is a $\mathfrak{Y}$-martingale. In particular, this is true for all real linear functionals f, and in this case, the resulting martingale is a real one.

(2) On the other hand, if x_τ is Bochner integrable and strongly measurable relative to $\mathscr{F}_\tau$, then by hypothesis $\mathscr{E}\{f(x_\tau) \mid \mathscr{F}_\sigma\} = f(x_\sigma)$ with probability 1 for every f in $\mathfrak{X}^*$. Then we can write

$$\begin{aligned} f\Big(\int_A x_\tau(\omega)dP\Big) &= \int_A f(x_\tau(\omega))dP = \int_A E\{f(x_\tau) \mid \mathscr{F}_\sigma\}(\omega)dP \\ &= \int_A f(x_\sigma(\omega))dP = f\Big(\int_A x_\sigma(\omega)dP\Big) \end{aligned}$$

for every f in $\mathfrak{X}^*$ and A in $\mathscr{F}_\sigma$. Therefore, $\int_A x_\tau(\omega)dP = \int_A x_\sigma(\omega)dP$ for every A in $\mathscr{F}_\sigma$. Hence $\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ is an $\mathfrak{X}$-martingale by Theorem 2.1. Q.E.D.

Note. By virtue of Hille-Phillips ([9] Theorem 3.7.12, p. 83), the theorem is true for f, a closed additive transformation from $\mathfrak{X}$ to $\mathfrak{Y}$, if we assume that $f(x_\tau)$ is Bochner integrable for every τ in T.

DEFINITION 2.1. Let $\mathfrak{Y}$ be a Banach space. A subset $\mathfrak{K}$ of $\mathfrak{Y}$ is called a positive cone if

(1) $\theta \in \mathfrak{K}$,

(2) $\xi \in \mathfrak{K}$ and a nonnegative imply $a\xi \in \mathfrak{K}$,

(3) if $\xi \in \mathfrak{K}$ and $-\xi \in \mathfrak{K}$, then $\xi = \theta$,

(4) if $\xi \in \mathfrak{K}$ and $\eta \in \mathfrak{K}$, then $\xi + \eta \in \mathfrak{K}$,

(5) $\mathfrak{K}$ is closed. By definition $\xi \geqq \eta$ if and only if $\xi - \eta \in \mathfrak{K}$. The order thus induced is a partial order (Hille-Phillips [9] Theorem 1.11.1, p. 15).

DEFINITION 2.2. Let $\mathfrak{Y}$ be a Banach space with a positive cone. Let T and $\mathscr{F}_\tau$ for $\tau \in T$ be as in Definition 1.1 of this chapter. Let x_τ be a Bochner integrable $\mathfrak{Y}$-valued strongly measurable (relative to $\mathscr{F}_\tau$) function on Ω for $\tau \in T$. Then $\{y_\tau, \mathscr{F}_\tau, \tau \in T\}$ is a $\mathfrak{Y}$-semi-martingale if $\mathscr{E}\{y_\tau \mid \mathscr{F}_\sigma\}(\omega) \geqq y_\sigma(\omega)$ with probability 1 for $\sigma < \tau$.

DEFINITION 2.3. A function g defined on $\mathfrak{X}$ with values in $\mathfrak{Y}$, a Banach space equipped with a positive cone, is said to be sub-additive if $g(\xi + \eta) \leqq g(\xi) + g(\eta)$, positive-homogeneous if $g(a\xi) = ag(\xi)$ for $a \geqq 0$.

THEOREM 2.3. *If x is a Bochner integrable $\mathfrak{X}$-valued function on Ω, $\mathscr{F}$ a Borel field of measurable subsets of Ω, and g a continuous subadditive positive-homogeneous function on $\mathfrak{X}$ to $\mathfrak{Y}$, a Banach space with a positive cone, such that $g(x)$ is Bochner integrable, then $g\left(\int_\Omega x(\omega)dP\right) \leqq \int_\Omega g(x(\omega))dP$ and $g(\mathscr{E}\{x \mid \mathscr{F}\}(\omega)) \leqq \mathscr{E}\{g(x) \mid \mathscr{F}\}(\omega)$ with probability 1. In particular, the conclusion follows for real valued g without the assumption of integrability on $g(x)$.*

Proof. If x and $g(x)$ are Bochner integrable, then by the methods of Hille-Phillips ([9] Corollary, p. 81, and Theorem 3.7.17, p. 83) there exists a sequence of countably valued integrable random variables x_n such that $\| x_n(\omega) - x(\omega) \| \to 0$, $\| g(x_n(\omega)) - g(x(\omega)) \| \to 0$ uniformly with probability 1 as $n \to \infty$, and also $\int_A \| x_n(\omega) - x(\omega) \| \, dP \to 0$ and $\int_A \| g(x_n(\omega)) - g(x(\omega)) \| \, dP \to 0$ as $n \to \infty$ for every measurable set A. Thus $\int_A x_n(\omega)dP \to \int_A x(\omega)dP$ and $\int_A g(x_n(\omega))dP \to \int_A g(x(\omega))dP$ as $n \to \infty$. Furthermore, $\mathscr{E}\{x_n \mid \mathscr{F}\} \to \mathscr{E}\{x \mid \mathscr{F}\}$, $\mathscr{E}\{g(x_n) \mid \mathscr{F}\} \to \mathscr{E}\{g(x) \mid \mathscr{F}\}$ uniformly with probability 1 as $n \to \infty$, and $\int_A \| \mathscr{E}\{x_n \mid \mathscr{F}\} - \mathscr{E}\{x \mid \mathscr{F}\} \| \, dP \to 0$, (Moy [14] p. 7) $\int_A \| \mathscr{E}\{g(x_n) \mid \mathscr{F}\} - \mathscr{E}\{g(x) \mid \mathscr{F}\} \| \, dP \to 0$ as $n \to \infty$ for every measurable set A. Let $x_n(\omega) = \xi_n^j$ for ω in A_n^j, where the A_n^j are disjunct measurable sets such that

$$\sum_{j=1}^{\infty} P\{A_n^j\} = 1 .$$

Then

$$\int_\Omega x_n(\omega)dP = \sum_{j=1}^\infty \xi_n^j P\{A_n^j\} = \lim_{N\to\infty}\sum_{j=1}^N \xi_n^j P\{A_n^j\}.$$

Now

$$g\Big(\sum_{j=1}^N \xi_n^j P\{A_n^j\}\Big) \leqq \sum_{j=1}^N g(\xi_n^j)P\{A_n^j\}$$

by the subadditivity and positive-homogeneity of g. Further,

$$\int_\Omega g(x_n(\omega))dP = \sum_{j=1}^\infty g(\xi_n^j)P\{A_n^j\} = \lim_{N\to\infty}\sum_{j=1}^N g(\xi_n^j)P\{A_n^j\} .$$

Hence,

$$\begin{aligned} g\Big(\int_\Omega x_n(\omega)dP\Big) &= g\Big(\lim_{N\to\infty}\sum_{j=1}^N \xi_n^j P\{A_n^j\}\Big) = \lim_{N\to\infty} g\Big(\sum_{j=1}^N \xi_n^j P\{A_n^j\}\Big) \\ &\leqq \lim_{N\to\infty}\sum_{j=1}^N g(\xi_n^j)P\{A_n^j\} = \int_\Omega g(x_n(\omega))dP , \end{aligned}$$

since g is continuous and the positive cone in $\mathfrak{Y}$ is closed. Similarly, $\mathscr{E}\{x_n \mid \mathscr{F}\} = \sum_{j=1}^\infty \xi_n^j E\{\chi_{A_n^j} \mid \mathscr{F}\}$ almost everywhere and thus,

$$\begin{aligned} g(\mathscr{E}\{x_n \mid \mathscr{F}\}(\omega)) &= g\Big(\lim_{N\to\infty}\sum_{j=1}^N \xi_n^j E\{\chi_{A_n^j} \mid \mathscr{F}\}\Big) \\ &\leqq \lim_{N\to\infty}\sum_{j=1}^N g(\xi_n^j)E\{\chi_{A_n^j} \mid \mathscr{F}\} = \mathscr{E}\{g(x_n) \mid \mathscr{F}\}(\omega) \text{ a.e.} \end{aligned}$$

Finally, $g\Big(\int_\Omega x_n(\omega)dP\Big) \to g\Big(\int_\Omega x(\omega)dP\Big)$ and $g(\mathscr{E}\{x_n \mid \mathscr{F}\}) \to g(\mathscr{E}\{x \mid \mathscr{F}\})$ a.e. by the continuity of g and the known convergence of the integrals and conditional expectations in question. Thus,

$$\begin{aligned} g\Big(\int_\Omega x(\omega)dP\Big) &= g\Big(\lim_{n\to\infty}\int_\Omega x_n(\omega)dP\Big) = \lim_{n\to\infty} g\Big(\int_\Omega x_n(\omega)dP\Big) \\ &\leqq \lim_{n\to\infty}\int_\Omega g(x_n(\omega))dP = \int_\Omega g(x(\omega)dP \end{aligned}$$

and

$$\begin{aligned} g(\mathscr{E}\{x \mid \mathscr{F}\}) &= g(\lim_{n\to\infty}\mathscr{E}\{x_n \mid \mathscr{F}\}) \quad \text{a.e.} \\ &= \lim_{n\to\infty} g(\mathscr{E}\{x_n \mid \mathscr{F}\}) \text{ a.e.} \leqq \lim_{n\to\infty}\mathscr{E}\{g(x_n) \mid \mathscr{F}\} \text{ a.e.} \\ &= \mathscr{E}\{g(x) \mid \mathscr{F}\} \text{ a.e.} \end{aligned}$$

If, in particular, g is a real valued subadditive positive-homogeneous continuous function, then there exists a finite nonnegative number M_g, $M_g = \sup\,[g(\xi); \|\xi\| \leqq 1]$, such that $|g(\xi)| \leqq M_g(\|\xi\| + 1)$ (Hille-

Phillips [9] Theorem 2.5.2, p. 25). Thus, $|g(x(\omega))| \leqq M_g(\|x(\omega)\| + 1)$, and, since the function on the right is integrable on Ω, it being a finite measure space, $g(x)$ is Lebesgue integrable, and the conclusion of the theorem follows. Q.E.D.

THEOREM 2.4. *Let $\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ be an $\mathfrak{X}$-martingale, and let g be a continuous subadditive positive-homogeneous function on $\mathfrak{X}$ to $\mathfrak{Y}$, a Banach space with a positive cone such that $g(x_\tau)$ is Bochner integrable for every τ in T. Then $\{g(x_\tau), \mathscr{F}_\tau, \tau \in T\}$ is a $\mathfrak{Y}$-semi-martingale. In particular, if g is a continuous subadditive positive-homogeneous functional the conclusion is that the resulting process is a real semi-martingale without assuming that $g(x_\tau)$ is integrable. Finally $\{\|x_\tau\|, \mathscr{F}_\tau, \tau \in T\}$ is a real semi-martingale.*

Proof. By Theorem 2.3, $\mathscr{E}\{g(x_\tau) \mid \mathscr{F}_\sigma\}(\omega) \geqq g(\mathscr{E}\{x_\tau \mid \mathscr{F}_\sigma\}(\omega))$ a.e. But the righthand side is equal almost everywhere to $g(x_\sigma(\omega))$ since $\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ is an $\mathfrak{X}$-martingale. Thus, $\mathscr{E}\{g(x_\tau) \mid \mathscr{F}_\sigma\}(\omega) \geqq g(x_\sigma(\omega))$ a.e. for $\sigma < \tau$. Since $g(x_\tau)$ is clearly strongly measurable relative to $\mathscr{F}_\tau$, $\{g(x_\tau), \mathscr{F}_\tau, \tau \in T\}$ is a $\mathfrak{Y}$-semi-martingale. Q.E.D.

Next we consider some examples.

EXAMPLE 2.1. Let z be Bochner integrable and $\{\mathscr{F}_\tau\}$ as before. Let $x_\tau = \mathscr{E}\{z \mid \mathscr{F}_\tau\}$. Then $\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ is an $\mathfrak{X}$-martingale. For let $\Lambda \in \mathscr{F}_\sigma, \sigma < \tau$,

$$\int_\Lambda x_\sigma(\omega)dP = \int_\Lambda \mathscr{E}\{z \mid \mathscr{F}_\sigma\}(\omega)dP = \int_\Lambda z(\omega)dP$$

as a consequence of the definition of $\mathscr{E}\{z \mid \mathscr{F}_\sigma\}$, and

$$\int_\Lambda x_\tau(\omega)dP = \int_\Lambda \mathscr{E}\{z \mid \mathscr{F}_\tau\}(\omega)dP = \int_\Lambda z(\omega)dP\,,$$

the last equality being true for all $\Lambda \in \mathscr{F}_\tau$ and hence for all $\Lambda \in \mathscr{F}_\sigma \subseteqq \mathscr{F}_\tau$. Thus $\int_\Lambda x_\sigma(\omega)dP = \int_\Lambda x_\tau(\omega)dP$ for $\Lambda \in \mathscr{F}_\sigma$. Hence, by Theorem 2.1, $\{x_\tau, \mathscr{F}_\tau, \tau \in T\}$ is an $\mathfrak{X}$-martingale.

Before proceeding to the next example we shall have to prove the following lemma.

LEMMA 2.1. *Let x and y be strongly measurable independent random variables. Let $\mathscr{F}$ be the Borel field of measurable sets generated by x;* i.e., *the smallest Borel field of measurable sets with respect to which x is strongly measurable. Suppose $\mathscr{E}\{y \mid \mathscr{F}\}$ exists, and define $\mathscr{E}\{y \mid x\} = \mathscr{E}\{y \mid \mathscr{F}\}$. Then $\mathscr{E}\{y \mid x\} = \mathscr{E}\{y\}$ with probability* 1.

Proof. If x and y are independent, then $f(x)$ and $f(y)$ are real valued independent random variables by Theorem 1.1 of Chapter I for every f in $\mathfrak{X}^*$. Thus $E\{f(y) \mid \mathscr{F}\} = E\{f(y)\}$ with probability 1. Next, let A be an $\mathscr{F}$ set. Then

$$f\Big(\int_A \mathscr{E}\{y \mid x\}(\omega)dP\Big) = f\Big(\int_A \mathscr{E}\{y \mid \mathscr{F}\}(\omega)dP\Big) = \int_A f(\mathscr{E}\{y \mid \mathscr{F}\}(\omega))dP$$
$$= \int_A E\{f(y) \mid \mathscr{F}\}(\omega)dP = \int_A E\{f(y)\}dP = f\Big(\int_A \mathscr{E}\{y\}dP\Big)$$

by Theorem 3.3 of Chapter II. Thus

$$\int_A \mathscr{E}\{y \mid x\}(\omega)dP = \int_A \mathscr{E}\{y\}dP ,$$

for every A in $\mathscr{F}$. Hence $\mathscr{E}\{y \mid x\} = \mathscr{E}\{y\}$ with probability 1 by Theorem 2.1 of Chapter I. Q.E.D.

In like manner, it can be shown that if $\{y_n\}$ are mutually independent, then $\mathscr{E}\{y_n \mid \mathscr{F}\} = \mathscr{E}\{y_n\}$ with probability 1 if $\mathscr{F}$ is the smallest Borel field relative to which $y_1, \cdots, y_{n-1}$ are strongly measurable.

EXAMPLE 2.2. Let $\{y_j, j \geqq 1\}$ be mutually independent, $\mathscr{E}\{y_j\} = \theta$ for $j > 1$, $\mathscr{F}_j$ be the smallest Borel field relative to which $y_1, \cdots, y_j$ are all strongly measurable, and $x_n = \sum_{j=1}^n y_j$. Then $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-martingale.

We show that $\mathscr{E}\{x_n \mid \mathscr{F}_{n-1}\} = x_{n-1}$ with probability 1.

Note. $\mathscr{E}\{x_n \mid \mathscr{F}_{n-1}\} = \mathscr{E}\{x_n \mid y_1, \cdots, y_{n-1}\} = \mathscr{E}\{x_n \mid x_1, \cdots, x_{n-1}\}$.

Clearly

$$x_n = \sum_{j=1}^{n} y_j = \sum_{j=1}^{n-1} y_j + y_n = x_{n-1} + y_n .$$

Then

$$\begin{aligned}\mathscr{E}\{x_n \mid \mathscr{F}_{n-1}\} &= \mathscr{E}\{x_{n-1} + y_n \mid \mathscr{F}_{n-1}\} \\ &= \mathscr{E}\{x_{n-1} \mid \mathscr{F}_{n-1}\} + \mathscr{E}\{y_n \mid \mathscr{F}_{n-1}\} \quad \text{by Theorem 2.2 of Chapter II.} \\ &= x_{n-1} + \mathscr{E}\{y_n \mid \mathscr{F}_{n-1}\} \quad \text{with probability 1 for } x_{n-1} \text{ is measurable relative to } \mathscr{F}_{n-1}. \\ &= x_{n-1} + \mathscr{E}\{y_n\} \quad \text{with probability 1 by Lemma 2.1} \\ &= x_{n-1} \text{ for } \mathscr{E}\{y_n\} = \theta \text{ for } n > 1 .\end{aligned}$$

Thus $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-martingale.

CHAPTER IV

MARTINGALE CONVERGENCE THEOREMS IN A BANACH SPACE

Let $\mathfrak{X}$ be a Banach space. We will prove various convergence theorems for $\mathfrak{X}$-martingales. Thus we will show that if $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-martingale, then under certain conditions there will exist an $\mathfrak{X}$-valued random variable x such that $x_n \to x$ with probability 1 in various senses.

THEOREM 1. *Let $\{x_n, \mathscr{F}_n, n \geqq 1\}$ be an $\mathfrak{X}$-martingale, and let $\mathscr{F}_\infty$ be the smallest Borel field of Ω sets such that $\mathscr{F}_\infty \supseteqq \bigcup_{n=1}^{\infty} \mathscr{F}_n$. Let $y_n(\omega) = \| x_n(\omega) \|$. Then*

$$E\{\| x_1 \|\} \leqq E\{\| x_2 \|\} \leqq \cdots \leqq E\{\| x_n \|\} \leqq \cdots$$

(1) If $\text{l.u.b.}_n E\{\| x_n \|\} < \infty$ then $\lim_{n\to\infty} \| x_n \| = y_\infty$ exists with probability 1, and $E\{y_\infty\} < \infty$. In fact, the boundedness condition reduces to $\lim_{n\to\infty} E\{\| x_n \|\} = K < \infty$, and then $E\{y_\infty\} \leqq K$.

(2) a. If the $\| x_n \|$'s are uniformly integrable then

$$\underset{n}{\text{l.u.b.}}\, E\{\| x_n \|\} < \infty, \lim_{n\to\infty} E\{| y_\infty - \| x_n \| |\} = 0 ,$$

and the process $\{y_n, \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ is a real semi-martingale dominated by a semi-martingale relative to the same fields. (Doob [1] p. 297)

b. If $\text{l.u.b.}_n E\{\| x_n \|\} < \infty$ so that y_∞ exists, and if the process $\{y_n, \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ is a real semi-martingale, then $\lim_{n\to\infty} E\{\| x_n \|\} = E\{y_\infty\}$ and the $\| x_n \|$'s are uniformly integrable.

Proof. If $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-martingale, then $\{\| x_n \|, \mathscr{F}_n, n \geqq 1\}$ is a real semi-martingale by Theorem 2.4 of Chapter III, and then $E\{\| x_1 \|\} \leqq \cdots E\{\| x_n \|\} \leqq \cdots$ according to Doob ([1] Theorem 2.1 (ii) p. 311). The other conclusions follow from Theorem 4.1 s of Doob ([1] p. 324–325). Q.E.D.

THEOREM 2. *Let $\{x_n, \mathscr{F}_n, n \geqq 1\}$ be an $\mathfrak{X}$-martingale. Let $\mathfrak{X}$ be reflexive. Suppose $\lim_{n\to\infty} E\{\| x_n \|\} = K < \infty$. Then there is an $\mathfrak{X}$-valued strongly measurable random variable x_∞ such that $x_n \to x_\infty$ weakly as $n \to \infty$ with probability* 1.

Proof. Since x_n is strongly measurable, there is a measurable set A_n such that $P\{A_n\} = 0$ and $x_n(\Omega - A_n)$ is separable, for strongly measurable functions are almost separably valued (Hille-Phillips [9] Theorem 3.5.3, p. 72). Let $\mathfrak{Y}_n = x_n(\Omega - A_n)$ and let $\mathfrak{Y}$ be the closed linear manifold spanned by $\bigcup_{n=1}^{\infty} \mathfrak{Y}_n$. Then $\mathfrak{Y}$ is a separable subspace of $\mathfrak{X}$ and $x_n(\omega) \in \mathfrak{Y}$ for almost all ω, for each n. Now $\mathfrak{Y}$ is reflexive since $\mathfrak{X}$ is. (Hille-Phillips [9] Corollary 1 to Theorem 2.10.3, p. 38). Further, since $\mathfrak{Y}$ is separable, then so is $\mathfrak{Y}^{**}$ for $\mathfrak{Y} \cong \mathfrak{Y}^{**}$. But then $\mathfrak{Y}^*$ is separable by Theorem 2.8.4 of Hille-Phillips ([9] p. 34). Now if $f \in \mathfrak{Y}^*$ then $\{f(x_n), \mathscr{F}_n, n \geqq 1\}$ is a real martingale by Theorem 2.2 of Chapter III. Also

$$E\{|f(x_n)|\} \leqq E\{||f||\,||x_n||\} = ||f||\,E\{||x_n||\} \leqq ||f||\,K$$

because $E\{||x_1||\} \leqq \cdots E\{||x_n||\} \leqq \cdots \leqq K$ by Theorem 1. By virtue of Doob ([1] Theorem 4.1, p. 319) for every $f \in \mathfrak{Y}^*$ there exists a real measurable function z_f, and a measurable set Λ_f such that $P\{\Lambda_f\} = 0$ and $|f(x_n(\omega)) - z_f(\omega)| \to 0$ as $n \to \infty$ for $\omega \in \Omega - \Lambda_f$. By the separability of $\mathfrak{Y}^*$ there is a countable dense subset $\{f_j\}$ of $\mathfrak{Y}^*$. Thus for every f_j there is a Λ_{f_j} and z_{f_j} as we have seen, such that $P\{\Lambda_{f_j}\} = 0$ and $|f_j(x_n(\omega)) - z_{f_j}(\omega)| \to 0$ as $n \to \infty$ for $\omega \in \Omega - \Lambda_{f_j}$. Let $\Lambda_1 = \bigcup_{j=1}^{\infty} \Lambda_{f_j}$. Then

$$P\{\Lambda_1\} = P\left\{\bigcup_{j=1}^{\infty} \Lambda_{f_j}\right\} \leqq \sum_{j=1}^{\infty} P\{\Lambda_{f_j}\} = 0\ .$$

By Theorem 1 there is a measurable set M such that $P\{M\} = 0$ and such that $||x_n(\omega)||$ is a convergent sequence for $\omega \in \Omega - M$. Let $\Lambda = \Lambda_1 \cup M$. Then $P\{\Lambda\} = 0$. Next, let $\omega \in \Omega - \Lambda$. Then $\omega \in \Omega - M$ so that $||x_n(\omega)||$ is a convergent sequence. Thus there is a constant C such that $||x_n(\omega)|| \leqq C$ for all n.

Define $Q_n(f) = f(x_n(\omega))$ for $f \in \mathfrak{Y}^*$. The Q_n's form an equi-continuous sequence of functions on $\mathfrak{Y}^*$, for, given $\epsilon > 0$, $\exists \delta = \varepsilon/C$ such that for every n, $||f - g|| < \delta$ implies

$$|Q_n(f) - Q_n(g)| = |f(x_n(\omega)) - g(x_n(\omega))| \leqq ||f - g||\,||x_n(\omega)|| < \varepsilon/C \cdot C = \varepsilon\ .$$

Furthermore, since $\omega \in \Omega - \Lambda_{f_j}$ for every j,

$$|Q_n(f_j) - Q_m(f_j)| = |f_j(x_n(\omega)) - f_j(x_m(\omega))| \to 0 \text{ as } n, m \to \infty\ .$$

But, an equicontinuous sequence of functions converging on a dense set of a metric space converges on the whole space. Thus for every $f \in \mathfrak{Y}^*$, $|Q_n(f) - Q_m(f)| \to 0$ as $n, m \to \infty$; i.e., $|f(x_n(\omega)) - f(x_m(\omega))| \to 0$ as $n, m \to \infty$ for every $\omega \in \Omega - \Lambda$.

Therefore $f(x_n(\omega))$ is a convergent sequence for all $\omega \in \Omega - \Lambda$ and $f \in \mathfrak{Y}^*$. The reflexiveness of $\mathfrak{X}$ and $\mathfrak{Y}$ implies that $\mathfrak{X}$ and $\mathfrak{Y}$ are weakly complete. Thus there is an x_∞ (strongly measurable) such that for every $f \in \mathfrak{Y}^*$ and $\omega \in \Omega - \Lambda$ we have $|f(x_n(\omega)) - f(x_\infty(\omega))| \to 0$ as $n \to \infty$; i.e. x_n converges to x_∞ weakly with probability 1. Q.E.D.

Note. x_∞ is strongly measurable since it is the weak limit of strongly measurable functions (Hille-Phillips [9] Theorem 3.5.4, p. 74). Theorem 2 may be restated as follows:

THEOREM 2*. *Let* $\{x_n, \mathscr{F}_n, n \geqq 1\}$ *be an* $\mathfrak{X}$*-martingale. Let* $\mathfrak{X}$ *be weakly complete and suppose that* $\mathfrak{X}^*$ *is separable, and* $\lim_{n\to\infty} E\{\|x_n\|\} = K < \infty$. *Then there is an* $\mathfrak{X}$*-valued strongly measurable random variable* x_∞ *such that* x_n *converges to* x_∞ *weakly with probability* 1.

COROLLARY 1. *Let* $\{x_n, \mathscr{F}_n, n \geqq 1\}$ *be an* $\mathfrak{X}$*-martingale. Suppose* $\mathfrak{X}$ *is a Hilbert space, and that* $\lim_{n\to\infty} E\{\|x_n\|\} = K < \infty$. *Then there exists a strongly measurable* $\mathfrak{X}$*-valued random variable* x_∞ *such that* $x_n \to x_\infty$ *weakly with probability* 1.

Proof. Since $\mathfrak{X}$ is a Hilbert space, it is reflexive and weakly complete. Hence all of the hypotheses of Theorem 2 are satisfied, and so the above conclusion follows. Q.E.D.

By making a stronger assumption on the $\|x_n\|$'s we will show that the last result may be sharpened to give strong convergence with probability 1.

THEOREM 3. *Let* $\{x_n, \mathscr{F}_n, n \geqq 1\}$ *be an* $\mathfrak{X}$*-martingale; let* $\mathfrak{X}$ *be reflexive. If the* $\|x_n\|$*'s are uniformly integrable, then there is a strongly measurable* $\mathfrak{X}$*-valued random variable* x_∞ *such that* $\|x_n(\omega) - x_\infty(\omega)\| \to 0$ *as* $n \to \infty$ *with probability* 1, *and in fact* $\{x_n, \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ *is an* $\mathfrak{X}$*-martingale.*

Proof. As in the proof of Theorem 2, there is a separable sub-space $\mathfrak{Y}$ of $\mathfrak{X}$, and for each n, $x_n(\omega) \in \mathfrak{Y}$ for almost all ω. Also $\mathfrak{Y}$ is reflexive, so therefore $\mathfrak{Y}^{**}$ is separable, which implies that $\mathfrak{Y}^*$ is separable. Now $E\{\|x_1\|\} \leqq E\{\|x_2\|\} \leqq \cdots \leqq E\{\|x_n\|\} \leqq \cdots$ since $\{\|x_n\|, \mathscr{F}_n, n \geqq 1\}$ is a semi-martingale. Therefore $\lim_{n\to\infty} E\{\|x_n\|\} = K \leqq \infty$, while $\lim_{n\to\infty} E\{|f(x_n)|\} \leqq \lim_{n\to\infty} \|f\| E\{\|x_n\|\} = \|f\| K$. But the uniform integrability of the $\|x_n\|$'s makes $K < \infty$ (Doob [1] Theorem 4.1, p. 319). Theorem 1 tells us that there is a y_∞ such that $|\|x_n\| - y_\infty| \to 0$ as $n \to \infty$ with probability 1, and such that $\{y_n, \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ is a real semi-martingale, where $y_n(\omega) = \|x_n(\omega)\|$ and $y_\infty(\omega) = \lim_{n\to\infty} \|x_n(\omega)\|$. In fact, $E\{|y_\infty - \|x_n\||\} \to 0$ as $n \to \infty$. By Theorem 2, there is a strongly measurable $\mathfrak{X}$-valued random variable x_∞ such that $|f(x_n(\omega)) - f(x_\infty(\omega))| \to 0$ as $n \to \infty$ with probability 1 for every $f \in \mathfrak{Y}^*$. Furthermore, if the $\|x_n\|$'s are uniformly integrable, then so are the $f(x_n)$'s for every $f \in \mathfrak{Y}^*$ because, first of all,

$$\{\omega: |f(x_n(\omega))| > M\} \subseteqq \left\{\omega: \|x_n(\omega)\| > \frac{M}{\|f\|}\right\}$$

if $||f|| < 0$. (If $||f|| = 0$, then trivially the $f(x_n)$'s are uniformly integrable.) Thus

$$\int_{\{\omega: |f(x_n(\omega))| > M\}} |f(x_n(\omega))|\, dP \leqq \int_{\{\omega: ||x_n(\omega)|| > M/||f||\}} |f(x_n(\omega))|\, dP$$
$$\leqq ||f|| \int_{\{\omega: ||x_n(\omega))|| > M/||f||\}} ||x_n(\omega)||\, dP \to 0$$
uniformly in n as $M \to \infty$.

By the uniform integrability of the $||x_n||$'s, thus proving the uniform integrability of the $f(x_n)$'s for every $f \in \mathfrak{Y}^*$. Hence $\{f(x_n), \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ is a real martingale for every $f \in \mathfrak{Y}^*$ by Doob ([1] Theorem 4.1, p. 319).

Next, x_∞ is strongly measurable (in fact, relative to $\mathscr{F}_\infty$) by Theorem 2. Furthermore, $E\{||x_\infty||\} < \infty$, for, $x_n \to x_\infty$ weakly with probability 1. Hence $||x_\infty(\omega)|| \leqq \lim_{n\to\infty} \inf ||x_n(\omega)||$ for almost all ω. But the right hand side equals $y_\infty(\omega)$ with probability 1 by Theorem 1. Thus $||x_\infty(\omega)|| \leqq y_\infty(\omega)$ a.e. Since y_∞ is integrable, so is $||x_\infty||$; hence, by Theorem 3.7.4 of Hille-Phillips ([9] p. 80), x_∞ is Bochner integrable. Thus, by Theorem 2.2 of Chapter III, $\{x_n, \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ is an $\mathfrak{X}$-martingale. Then $\{||x_n||, \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ is a semi-martingale by Theorem 3.4 of Chapter III. But so is $\{||x_1||, \cdots, ||x_n||, \cdots, y_\infty\}$ relative to $\mathscr{F}_1, \cdots, \mathscr{F}_n, \cdots, \mathscr{F}_\infty$.

We now show that $||x_\infty|| = y_\infty$ with probability 1. We have already shown $E\{||x_\infty||\} \leqq E\{y_\infty\}$. But $E\{||x_n||\} \leqq E\{||x_\infty||\}$ since $\{||x_n||, \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ is a semi-martingale, and since $E\{||x_n||\} \to E\{y_\infty\}$ by Theorem 1, we have $E\{y_\infty\} \leqq E\{||x_\infty||\}$. Hence, $E\{||x_\infty||\} = E\{y_\infty\}$. But $||x_\infty(\omega)|| \leqq y_\infty(\omega)$ for almost all ω. Therefore by Theorem B of Halmos ([8] p. 104), $||x_\infty(\omega)|| = y_\infty(\omega)$ for almost all ω, and $||x_n(\omega)|| \to ||x_\infty(\omega)||$ with probability 1, even as $x_n \to x_\infty$ weakly with probability 1. Next, let $\xi \in \mathfrak{Y}$. Then $\{x_n - \xi, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-martingale, for

$$\mathscr{E}\{x_n - \xi \mid \mathscr{F}_m\} = \mathscr{E}\{x_n \mid \mathscr{F}_m\} - \mathscr{E}\{\xi \mid \mathscr{F}_m\}$$ with probability 1 by Theorem 2.2 of Chapter II

$$= x_m - \xi$$ with probability 1, since $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-martingale, and by Theorem 2.2 of Chapter II.

Now by what we have already proved in this theorem, since the $||x_n - \xi||$'s are clearly uniformly integrable, there is a u_∞ such that $f(x_n - \xi) \to f(u_\infty)$ with probability 1 for every $f \in \mathfrak{Y}^*$ and $||x_n(\omega) - \xi|| \to u_\infty(\omega)$ with probability 1. But $f[x_n(\omega) - \xi] = f(x_n(\omega)) - f(\xi) \to f(x_\infty(\omega)) - f(\xi) = f[x_\infty(\omega) - \xi]$ as $n \to \infty$ with probability 1. Thus $u_\infty(\omega) = x_\infty(\omega) - \xi$ with probability 1. Hence $||x_n(\omega) - \xi|| \to ||x_\infty(\omega) - \xi||$ with probability 1. Let $\{\xi_j\}$ be a denumerable dense set in $\mathfrak{Y}$. Then there is a Λ_j such that $P\{\Lambda_j\} = 0$ and $||x_n(\omega) - \xi_j|| \to ||x_\infty(\omega) - \xi_j||$ for

$\omega \in \Omega - \Lambda_j$. Let $\Lambda = \bigcup_{j=1}^{\infty} \Lambda_j$. Then $P\{\Lambda\} = 0$. Let $\omega \in \Omega - \Lambda$, and define $R_n(\xi) = \| x_n(\omega) - \xi \|$ for $\xi \in \mathfrak{Y}$. The R_n's form an equicontinuous sequence of functions on $\mathfrak{Y}$, for given $\varepsilon > 0$, $\exists \delta = \varepsilon$, such that for every n, $\| \xi - \eta \| < \delta = \varepsilon$ implies $| R_n(\xi) - R_n(\eta) | = | \| x_n(\omega) - \xi \| - \| x_n(\omega) - \eta \| | \leqq \| \xi - \eta \| < \varepsilon$. Furthermore, since $\omega \in \Omega - \Lambda_j$ for every j,

$$| R_n(\xi_j) - \| x_\infty(\omega) - \xi_j \| | = | \| x_n(\omega) - \xi_j \| - \| x_\infty(\omega) - \xi_j \| | \to 0 \text{ as } n \to \infty .$$

But, as an equicontinuous sequence of functions converging on a dense set of a metric space converges on the whole space, thus for every $\xi \in \mathfrak{Y}$, $| R_n(\xi) - \| x_\infty(\omega) - \xi \| | = | \| x_n(\omega) - \xi \| - \| x_\infty(\omega) - \xi \| | \to 0$ as $n \to \infty$ for every $\omega \in \Omega - \Lambda$. Now, for $\omega \notin \Lambda$, let $\xi = x_\infty(\omega)$. Then $\| x_n(\omega) - x_\infty(\omega) \| \to \| x_\infty(\omega) - x_\infty(\omega) \| = 0$. Thus there is a measurable set Λ such that $P\{\Lambda\} = 0$ and such that for $\omega \in \Omega - \Lambda$, $\| x_n(\omega) - x_\infty(\omega) \| \to 0$ as $n \to \infty$. Q.E.D.

COROLLARY 2. *If $\mathfrak{X}$ is a Hilbert space, or l^p, or L^p, $1 < p < \infty$, and $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-martingale in which the $\| x_n \|$'s are uniformly integrable, then there is an x_∞ such that $\{x_n, \mathscr{F}_n, 1 \leqq n \leqq \infty\}$ is an $\mathfrak{X}$-martingale, and $\| x_n(\omega) - x_\infty(\omega) \| \to 0$ as $n \to \infty$ with probability* 1.

Proof. All of the above named Banach spaces are reflexive, and thus the result follows from Theorem 3.

REMARK. Let $\mathfrak{X}$ be a Banach space with a partial order induced by a positive cone. Suppose $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-semi-martingale. Then, as in Doob ([1] p. 297), x_n can be represented in the form

$$x_n = x'_n + \sum_{j=1}^{n} \Delta_j ,$$

where $\Delta_1 = \theta$; $\Delta_j = \mathscr{E}\{x_j \mid \mathscr{F}_{j-1}\} - x_{j-1} \geqq \theta, j > 1$; and $\{x'_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-martingale. Thus convergence problems for $\mathfrak{X}$-semi-martingales can be reduced to convergence of $\mathfrak{X}$-martingales if reasonable conditions can be found for the convergence of the monotone sequence $y_n = \sum_{j=1}^{n} \Delta_j$.

THEOREM 4. *Let $\{x_n, \mathscr{F}_n, n \leqq -1\}$ be an $\mathfrak{X}$-martingale in which $\mathfrak{X}$ is reflexive, and let $\mathscr{F}_{-\infty} = \bigcap_{-\infty}^{-1} \mathscr{F}_n$. Then $x_{-\infty}$ exists, such that $\| x_n(\omega) - x_{-\infty}(\omega) \| \to 0$ as $n \to -\infty$ with probability* 1, *and $\{x_n, \mathscr{F}_n, -\infty \leqq n \leqq -1\}$ is an $\mathfrak{X}$-martingale.*

Proof. $\{\| x_n \|, \mathscr{F}_n, n \leqq -1\}$ is a real semi-martingale; thus by Doob ([1] Theorem 4.25, p. 329) $\lim_{n \to -\infty} \| x_n(\omega) \| = y_{-\infty}$ exists with probability 1, and $-\infty \leqq y_{-\infty} < \infty$, while $\{\| x_n \|, \mathscr{F}_n, -\infty \leqq n \leqq -1\}$ is a semi-martingale. By Theorem 4.2 of Doob ([1] p. 328) $\lim_{n \to -\infty} f(x_n)$ exists for almost all ω and every $f \in \mathfrak{X}^*$. Using the methods of Theorem 2, we can show that there is an $x_{-\infty}$ such that $f(x_n(\omega)) \to f(x_{-\infty}(\omega))$ as $n \to -\infty$ for almost all ω and all f. Using the methods of Theorem 3, we show that $\{x_n, \mathscr{F}_n, -\infty \leqq n \leqq -1\}$ is an $\mathfrak{X}$-martingale, and that $\| x_{-\infty} \| = y_{-\infty}$ and $\| x_n(\omega) - x_{-\infty}(\omega) \| \to 0$ as $n \to -\infty$ with probability 1. Q.E.D.

THEOREM 5. *Let z be a strongly measurable random variable, $\mathfrak{X}$ reflexive, with $E\{\| z \|\} < \infty$; let $\cdots \mathscr{F}_{-n} \subseteqq \cdots \subseteqq \mathscr{F}_0 \subseteqq \cdots \subseteqq \cdots \subseteqq \mathscr{F}_n \subseteqq \cdots$ be Borel fields of measurable Ω sets. Let $\mathscr{F}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \mathscr{F}_n$, be the smallest Borel field of Ω sets with $\mathscr{F}_\infty \supseteqq \bigcup_{n=-\infty}^{\infty} \mathscr{F}_n$. Then $\lim_{n\to-\infty} \mathscr{E}\{z \mid \mathscr{F}_n\} = \mathscr{E}\{z \mid \mathscr{F}_{-\infty}\}$, and $\lim_{n\to\infty} \mathscr{E}\{z \mid \mathscr{F}_n\} = \mathscr{E}\{z \mid \mathscr{F}_\infty\}$ with probability* 1.

Proof. Let $x_n = \mathscr{E}\{z \mid \mathscr{F}_n\}$, $-\infty \leqq n \leqq \infty$. Then $\{x_n, \mathscr{F}_n, -\infty \leqq n \leqq \infty\}$ is an $\mathfrak{X}$-martingale by Example 2.1 of Chapter III. Thus by Theorem 4, $\lim_{n\to-\infty} \mathscr{E}\{z \mid \mathscr{F}_n\} = \mathscr{E}\{z \mid \mathscr{F}_{-\infty}\}$. Next, $\{\| x_n \|, \mathscr{F}_n, -\infty \leqq n \leqq \infty\}$ is a real semi-martingale, with a last term in which all the random variables are nonnegative. Thus by Theorem 3.1 of Doob ([1] p. 311) the $\| x_n \|$'s are uniformly integrable. Hence by Theorem 3, there is a y such that $\| x_n(\omega) - y(\omega) \| \to 0$ as $n \to \infty$ for almost all ω and $\{x_n, 1 \leqq n < \infty, y\}$ is an $\mathfrak{X}$-martingale. We finally must show that $x_\infty(\omega) = y(\omega)$ with probability 1. But this is true for both x_∞ and y are equal almost everywhere to functions measurable relative to $\mathscr{F}_\infty$. Also $\int_\Lambda x_\infty(\omega) dP = \int_\Lambda \mathscr{E}\{z \mid \mathscr{F}_\infty\}(\omega) dP = \int_\Lambda z(\omega) dP$ for $\Lambda \in \mathscr{F}_\infty$ and $\int_\Lambda y(\omega) dP = \int_\Lambda x_n(\omega) dP = \int_\Lambda \mathscr{E}\{z \mid \mathscr{F}_n\}(\omega)\, dP = \int_\Lambda z(\omega) dP$ for every $\Lambda \in \mathscr{F}_n$ and thus for every $\Lambda \in \bigcup_n \mathscr{F}_n$. Hence $\int_\Lambda y(\omega) dP = \int_\Lambda x_\infty(\omega) dP$ for every $\Lambda \in \bigcup_n \mathscr{F}_n$; thus, $\int_\Lambda f(y(\omega)) dP = \int_\Lambda f(x_\infty(\omega)) dP$ for every $\Lambda \in \bigcup_n \mathscr{F}_n$ and $f \in \mathfrak{X}^*$. But these integrals define completely additive set functions of $\mathscr{F}_\infty$ sets which are identical on the fields $\bigcup_n \mathscr{F}_n$ and therefore identical on $\mathscr{F}_\infty$ (Doob [1] Theorem 2.1, p. 605). Thus $\int_\Lambda y(\omega) dP = \int_\Lambda x_\infty dP$ for every $\Lambda \in \mathscr{F}_\infty$. Hence $y(\omega) = x_\infty(\omega)$ with probability 1 and $\lim_{n\to\infty} \mathscr{E}\{z \mid \mathscr{F}_n\} = \mathscr{E}\{z \mid \mathscr{F}_\infty\}$ with probability 1. Q.E.D.

COROLLARY 3. *Let z be a strongly measurable random variable, with $E\{\| z \|\} < \infty$ and let $y_1, y_2, \cdots$ be strongly measurable. Let $\mathscr{S}_n$ be the smallest Borel field with respect to which $y_n, y_{n+1} \cdots$ are strongly measurable. Then $\lim_{n\to\infty} \mathscr{E}\{z \mid \mathscr{S}_n\} = \mathscr{E}\{z \mid \bigcap_{n=1}^{\infty} \mathscr{S}_n\}$, $\lim_{n\to\infty} \mathscr{E}\{z \mid \mathscr{H}_n\} = \mathscr{E}\{z \mid \mathscr{H}_\infty\}$ where $\mathscr{H}_n$ is the smallest Borel field relative to which $y_1, y_2, \cdots, y_n$ are strongly measurable, $\mathscr{H}_\infty$ the smallest Borel field containing $\bigcup_{n=1}^{\infty} \mathscr{H}_n$.*

Proof. In Theorem 5, let $\mathscr{S}_n = \mathscr{F}_{-n}$ and $\mathscr{H}_n = \mathscr{F}_n$. Q.E.D.

Using this corollary it is possible to get a proof of the Banach space version of the strong law of large numbers. In fact, such a proof is virtually along the lines outlined in Doob ([1] p. 341). Mourier [13] has proved an ergodic theorem, more general than this one, by a more direct approach.

EXAMPLE 1. Let $\mathfrak{X} = l^p, 1 < p < \infty$ (real l^p). Then

$$x_n(\omega) = (\xi_1^{(n)}(\omega), \cdots, \xi_j^{(n)}(\omega), \cdots) \text{ where } \sum_{j=1}^{\infty} |\xi_j^{(n)}(\omega)|^p < \infty ,$$

and

$$|| x_n(\omega) || = \left\{\sum_{j=1}^{\infty} |\xi_j^{(n)}(\omega)|^p\right\}^{1/p} .$$

If x_n is Bochner integrable, then its integral satisfies the equation

$$\int_\Omega x_n(\omega)dP = \left\{\int_\Omega \xi_1^{(n)}(\omega)dP, \cdots, \int_\Omega \xi_j^{(n)}(\omega)dP, \cdots\right\}$$

where the components are ordinary Lebesgue integrals; thus the components of x_n are real-valued Lebesgue integrable functions.

The martingale equality becomes

$$\left\{\int_\Lambda \xi_1^{(n)}(\omega)dP, \cdots, \int_\Lambda \xi_j^{(n)}(\omega)dP, \cdots\right\}$$
$$=\left\{\int_\Lambda \xi_1^{(m)}(\omega)dP, \cdots, \int_\Lambda \xi_j^{(m)}(\omega)dP, \cdots\right\}, m < n, \Lambda \in \mathscr{F}_m \subset \mathscr{F}_n$$

or, alternatively,

$$\int_\Lambda \xi_j^{(n)}(\omega)dP = \int_\Lambda \xi_j^{(m)}(\omega)dP, m < n, \Lambda \in \mathscr{F}_m \subset \mathscr{F}_n \text{ for } j = 1, 2, \cdots .$$

Thus for every j, $\{\xi_j^{(n)}, \mathscr{F}_n, n \geqq 1\}$ is a real martingale, which can also be seen by noticing that the mapping from an l^p vector to any of its components is a linear functional. Then if

$$E\{|| x_n ||\} = \int_\Omega \left\{\sum_{j=1}^{\infty} |\xi_j^{(n)}(\omega)|^p\right\}^{1/p} dP \leqq K < \infty ,$$

by Theorem 2 there is an $x(\omega) = \{\xi_1(\omega), \cdots, \xi_j(\omega), \cdots\} \in l^p$ such that for every $\eta = (\eta_1, \cdots, \eta_j, \cdots) \in l^q$, $1/p + 1/q = 1$, $\sum_{j=1}^\infty \eta_j \xi_j^{(n)}(\omega) \to \sum_{j=1}^\infty \eta_j \xi_j(\omega)$ as $n \to \infty$ for almost all ω. Note that the boundedness assumption on the $E\{|| x ||\}$'s implies boundedness for $E\{| \xi_j^{(n)} |\}$'s for every j; thus we could get convergence in each component by the ordinary martingale convergence theorems.

Finally, if the $\|x_n\|$'s are uniformly integrable, that is, if

$$\int_{\Lambda_K}\left[\sum_{j=1}^{\infty}|\xi_j^n(\omega)|^p\right]^{1/p}dP\to 0$$

uniformly in n as $K\to\infty$, $\Lambda_K=\{\omega\colon[\sum_{j=1}^{\infty}|\xi_j^n(\omega)|^p]^{1/p}>K\}$. We can get by the ordinary martingale convergence theorem that $\int_\Lambda \xi_j(\omega)dP=\int_\Lambda \xi_j^n(\omega)dP$ for $\Lambda\in\mathscr{F}_n$, $n\geqq 1$ for every j.

However, we get more by Theorem 3, namely, $\sum_{j=1}^{\infty}|\xi_j^n(\omega)-\xi_j(\omega)|^p\to 0$ with probability 1 as $n\to\infty$, and also, of course $\sum_{j=1}^{\infty}|\xi_j^n(\omega)|^p\to\sum_{j=1}^{\infty}|\xi_j(\omega)|^p$ with probability 1 as $n\to\infty$.

EXAMPLE 2. Let $\mathfrak{X}=L^p(I)$, where I is the closed unit interval with Lebesgue measure, $p>1$. Then $x_n(\omega)=g_n(t,\omega)$ where $\int_\Omega |g_n(t,\omega)|^p dt<\infty$. Now if $x_n(\omega)$ is strongly measurable relative to $\mathscr{F}_n$, there is a representation $g_n(t,\omega)$ which is measurable over $\Omega\times I$ such that $g_n(\cdot,\omega)=x_n(\omega)$ in $L^p(I)$ a.e. in Ω, and any two representations of $x_n(\cdot)$ differ over $\Omega\times I$ on at most a set of measure zero. (Dunford-Pettis [4] Theorem 1.3.2, p. 336).

If $x_n(\cdot)$ is Bochner integrable, then besides being strongly measurable, $\int_\Omega \|x_n(\omega)\|dP<\infty$.

Thus

$$\int_\Omega\left\{\int_I |g_n(t,\omega)|^p dt\right\}^{1/p}dP=\int_\Omega\|x_n(\omega)\|\,dP<\infty\ .$$

Hence

$$\int_\Omega\left\{\int_I |g_n(t,\omega)|\,dt\right\}dP\leqq\int_\Omega\left\{\int_I |g_n(t,\omega)|^p dt\right\}^{1/p}dP<\infty$$

by the Hölder Inequality. Therefore, by the Fubini Theorem,

$$\int_\Omega\int_I g_n(t,\omega)dt\,dP=\int_I\int_\Omega g_n(t,\omega)dP\,dt\ ,$$

and

$$\begin{aligned}\int_I\left\{\int_\Omega x_n(\omega)dP\right\}(t)dt&=\int_\Omega\int_I x_n(\omega)(t)dt\,dP\\&=\int_\Omega\int_I g_n(t,\omega)dt\,dP=\int_I\int_\Omega g_n(t,\omega)dP\,dt\ .\end{aligned}$$

Hence

$$\left\{\int_\Omega x_n(\omega)dP\right\}(t)=\int_\Omega g_n(t,\omega)dP$$

for almost all t.

If $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an L^p-martingale, then $\int_A x_n(\omega)dP = \int_A x_m(\omega)dP$ for $A \in \mathscr{F}_m$, $m < n$, i.e., $\int_A g_n(t, \omega)dP = \int_A g_m(t, \omega)dP$ for almost all t, and $A \in \mathscr{F}_m$, $m < n$. Hence, for almost all $t \in I$ (Lebesgue measure) if $\mathscr{F}_n$ is generated by countably many sets, $\{g_n(t, \cdot), \mathscr{F}_n, n \geqq 1\}$ is a real martingale.

Next, if

$$E\{\| x_n \|\} = \int_\Omega \left[\int_I | g_n(t, \omega)|^p dt\right]^{1/p} dP \leqq K < \infty ,$$

there is an $x(\omega) = g(t, \omega) \in L^p(I)$, $\int_I | g(t, \omega) |^p dt < \infty$ by Theorem 2 such that $\int_I h(t)g_n(t, \omega)dt \to \int_I h(t)g(t, \omega)dt$ as $n \to \infty$ with probability 1 for every $h \in L^q(I)$, $1/p + 1/q = 1$

Furthermore, by Theorem 3, if the $\| x_n \|$'s are uniformly integrable, then $\int_I | g_n(t, \omega) |^p dt \to \int_I | g(t, \omega) |^p dt$ as $n \to \infty$ with probability 1, and even better, $\int_I | g_n(t, \omega) - g(t, \omega) |^p dt \to 0$ as $n \to \infty$ with probability 1.

The uniform integrability condition says that

$$\int_{A_N} \left[\int_I | g_n(t, \omega) |^p dt\right]^{1/p} dP \to 0$$

uniformly in n as $N \to \infty$,

$$\left\{\omega : \left[\int_I | g_n(t, \omega) |^p dt\right]^{1/p} > N\right\} = A_N .$$

This implies uniform integrability of the random variables in the real martingales $\{g_n(t, \cdot), \mathscr{F}_n, n \geqq 1\}$. Thus for almost all t, we can apply the ordinary Doob martingale theorems, and thus get convergence theorems in each coordinate.

The functions $g_n(t, \omega)$ as functions of t might, as a further illustration, be sample functions of a sequence of measurable stochastic processes (Doob [1] p. 60) with the property of being absolutely integrable over $\Omega \times I$.

EXAMPLE 3. We have seen in Example 2.2 of Chapter III that if $\{y_j, j \geqq 1\}$ are mutually independent, as $\mathfrak{X}$-valued random variables, with $\mathscr{E}\{y_j\} = \theta$ for $j > 1$, and $\mathscr{F}_j$ is the smallest Borel field relative to which $y_1, \cdots, y_j$ are all strongly measurable, and if $x_n = \sum_{j=1}^n y_j$, then $\{x_n, \mathscr{F}_n, n \geqq 1\}$ is an $\mathfrak{X}$-martingale. Theorem 2 tells us that if $\lim_{n\to\infty} E\{\|x_n\|\} = K < \infty$, then $\sum_{j=1}^\infty f(y_j(\omega))$ converges with probability 1. If, further, the $\| x_n \|$'s are uniformly integrable, then by Theorem 3, $\sum_{j=1}^\infty y_j(\omega)$ converges with probability 1.

Examples 1 and 2 above illustrate an important point. It is clear from them that an l^p-martingale is really a countable collection of one-dimensional martingales, while an L^p-martingale is a non-denumerable collection of ordinary real martingales. Thus, it is possible to prove convergence theorems for l^p or L^p by first proving convergence in each coordinate, using the Doob theorems on convergence of ordinary martingales. One could prove the convergence theorem for abstract Hilbert space by first proving the theorem for l^2 in each coordinate and then using the fact that there is a one-to-one linear norm preserving transformation between l^2 and abstract Hilbert space. In fact, one could prove convergence theorems for any $\mathfrak{X}$-martingale in which $\mathfrak{X}$ is a function space or a coordinate space by first proving martingale convergence theorems in each coordinate.

Let $\{\xi_t, t \in I = [0, 1]\}$ be a separable Brownian motion process (Doob [1] p. 52, p. 392). Then there is a measurable set $\Omega_0 \subset \Omega$, such that $P\{\Omega - \Omega_0\} = 0$, and such that for $\omega \in \Omega_0$, $\xi_t(\omega)$ is a continuous function of $t \in I$. Let $x(\omega) = \xi_t(\omega) = g(t, \omega)$. Then $x(\cdot)\colon \Omega \to C(I)$, the continuous function space on the unit interval, and $\| x(\omega) \| = \sup_{t \in I} | g(t, \omega) |$.

We next show that $x(\cdot)$ is strongly measurable. Let $f \in \mathfrak{X}^* = C(I)^*$. Then there is a function of bounded variation F such that $f(x(\omega)) = \int_I g(t, \omega) dF(t)$

$$= \lim_{\max|t_j - t_{j-1}| \to 0} \sum_{j=1}^{n} g(u_j, \omega)[F(t_j) - F(t_{j-1})]$$

where $0 = t_0 < t_1 < \cdots < t_n = 1$ and $t_{j-1} < u_j < t_j$. But each sum is clearly measurable in ω, so the limit must be too. Thus $x(\cdot)$ is weakly measurable, but since $C(I)$ is separable, this is equivalent to strong measurability of x.

To show that $x(\cdot)$ is Bochner integrable, we need only show that $E\{\| x \|\} < \infty$, for $x(\cdot)$ is strongly measurable. To this end, let $\xi_0 = 0$ with probability 1, and let $h(\omega) = \| x(\omega) \| = \sup_{t \in I} | g(t, \omega) |$.

Then

$$P\{\omega\colon h(\omega) \geqq n\} \leqq \frac{\sigma}{n}\sqrt{\frac{2}{\pi}} e^{-n^2/2\sigma^2} .$$

(Doob [1] p. 392) Thus

$$\sum_{n=1}^{\infty} P\{\omega\colon h(\omega) \geqq n\} \leqq \sigma\sqrt{\frac{2}{\pi}} \sum_{n=1}^{\infty} \frac{1}{n} e^{-n^2/2\sigma^2} < \infty .$$

Hence, $E\{\| x \|\} < \infty$, and $x(\cdot)$ the sample function of a separable Brownian motion process is Bochner integrable.

Let $\mathscr{F}_1$ be the Borel field of Ω sets generated by $\xi_0, \xi_{1/2}, \xi_1$; $\mathscr{F}_2$ the

Borel field generated by $\xi_0, \xi_{1/4}, \xi_{1/2}, \xi_{3/4}, \xi_1$, and in general $\mathscr{F}_n$ the Borel field generated by $\xi_0, \xi_{1/2^n}, \cdots, \xi_{2^n-1/2^n}, \xi_1$. Then $\mathscr{F}_1 \subset \mathscr{F}_2 \subset \cdots \subset \mathscr{F}_n \subset \cdots$ Let $f_n(t)(\omega) = E\{\xi_t \mid \mathscr{F}_n\}(\omega)$. Lévy ([11]) p. 18) has shown that $f_n(t)(\omega)$ is a polygonal line function of t for almost all ω, and that $|f_n(t)(\omega) - \xi_t(\omega)| \to 0$ as $n \to \infty$ uniformly in t for almost all ω. If we let $y_n(\omega) = f_n(t)(\omega) \in C(I)$ for $\omega \in \Omega$, then $\{y_n, \mathscr{F}_n, n \geqq 1\}$ is a $C(I)$-martingale. Lévy's result does not as yet come out of our work because $C(I)$ is not reflexive.

The validity of the Martingale Convergence Theorem for non-reflexive spaces is not known to the author. In fact, various, attempts in proving it have failed. If it were established, then further interesting examples like the last one for important non-reflexive spaces, e.g., L^1 or l^1, could be given.

BIBLIOGRAPHY

1. J. L. Doob, *Stochastic processes*, John Wiley and Sons, New York, 1953.

2. S. Doss, *Sur la moyenne d'un élément aléatoire dans un espace distancié*, Bull, Soc. Math., LXXIII (1949) p. 1.

3. N. Dunford, *Uniformity in linear spaces*, Trans. Amer. Math. Soc., **44** (1938), 305-356.

4. N. Dunford and B. J. Pettis, *Linear operations on summable functions*, Trans. Amer. Math. Soc., **47** (1938), 323-392.

5. R. Fortet and E. Mourier, *Loi des grands nombres et théorie ergodique*, Comptes Rendus, Acad. Sci., Paris, **234** (1952), 699.

6. M. Frechet, *Généralités sur les probabilités, éléments aléatoires*, Gauthier-Villars, Paris, 1950.

7. B. Gnedenko and A. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley, Boston, 1954.

8. P. Halmos, *Measure theory*, D. Van Nostrand Company, Inc., New York, 1950.

9. E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc. Coll. Publ., **31** (1957), revised edition.

10. A. Kolmogorov, *La transformation de La Place dans les espaces linéaires*, Comptes Rendus, Acad. Sci., Paris, **200** (1935), 1717-1718.

11. P. Lévy, *Processus stochastiques et mouvement Brownien*, Gauthier-Villars, Paris, 1948.

12. M. Loève, *Probability theory*, D. Van Nostrand Company, Inc., New York, 1955.

13. E. Mourier, *Eléments aléatoires dans un espace de Banach*, Gauthier-Villars, Paris, 1954.

14. S. C. Moy, *Conditional expectations of random variables with values in a Banach space and their properties*, Wayne University, Office of Ordnance Research Project, 1956.

15. B. J. Pettis, *On integration in vector spaces*, Trans. Amer. Math. Soc., **44** (1938). 277-304.

UNIVERSITY OF ILLINOIS, URBANA, ILLINOIS
I. B. M. CORPORATION, NEW YORK, N. Y.

TORSION ENDOMORPHIC IMAGES OF MIXED ABELIAN GROUPS

ELBERT A. WALKER

In this paper we will answer Fuchs' PROBLEM 32 (a), and the corresponding part of his PROBLEM 33. (See [1], pg. 203.) The statements of these PROBLEMS are the following.

I. "Which are the torsion groups T that are endomorphic images of all groups containing them as maximal torsion subgroups?"

II. "Which are the torsion groups T such that a basic subgroup of T is an endomorphic image of any group G containing T as its maximal torsion subgroup?"

Actually, we will answer question II and the following question which is more general than I.

III. What groups H are endomorphic images of all groups G containing H such that G/H is torsion free?

The solutions will be effected by using some homological results of Harrison [2]. All groups considered here will be Abelian. The definitions and results stated in the remainder of this paragraph are due to Harrison, and may be found in [2]. A reduced group G is *cotorsion* if $\text{Ext}(A, G) = 0$ for all torsion free groups A. If H is a reduced group, then $\text{Ext}(Q/Z, H) = H'$ is cotorsion, where Q and Z denote the additive group of rationals and integers, respectively. Furthermore, H is a subgroup of H', (that is, there is a natural isomorphism of H into H') and H'/H is divisible torsion free. This implies, of course, that if T is a torsion reduced group, then T is the torsion subgroup of $T'=\text{Ext}(Q/Z, T)$.

Now it is easy to see that *if G is a group such that* $\text{Ext}(A, G) = 0$ *for all torsion free groups A, then any homomorphic image of G is the direct sum of a cotorsion group and a divisible group.* In fact, let H be a homomorphic image of G. This gives us an exact sequence

$$0 \to K \to G \to H \to 0$$

which yields the exact sequence

$$\begin{aligned} 0 \to \text{Hom}(A, K) \to \text{Hom}(A, G) \to \text{Hom}(A, H) \to \\ \text{Ext}(A, K) \to \text{Ext}(A, G) \to \text{Ext}(A, H) \to 0 \, . \end{aligned}$$

If A is any torsion free group, then $\text{Ext}(A, G) = 0$, and so $\text{Ext}(A, H) = 0$. Write $H = D \oplus L$, where D is the divisible part of H. Then L is reduced, and $0 = \text{Ext}(A, D \oplus L) \cong \text{Ext}(A, D) \oplus \text{Ext}(A, L) = \text{Ext}(A, L)$, so that L is cotorsion. Our assertion is proved.

Received April 28, 1960.

Now we are ready to give the solutions promised earlier. The following theorem settles III.

THEOREM. *The group H is an endomorphic image of every group G containing it such that G/H is torsion free if and only if $H = D \oplus C$, where D is divisible and C is cotorsion. This is equivalent to the assertion that H is a direct summand of every such G.*

Proof. Suppose H is an endomorphic image of every group G containing it such that G/H is torsion free. Let $H = D \oplus C$, where D is divisible and C is reduced. Then C is a subgroup of the cotorsion group $\text{Ext}(Q/Z, C) = C'$ such that C'/C is torsion free, so that H is a subgroup of $D \oplus C' = H'$ such that H'/H is torsion free. Therefore H is an endomorphic image of H'. $\text{Ext}(A, D \oplus C') = 0$ for all torsion free groups A, and as we have just proved, any homomorphic image of $D \oplus C'$ is the direct sum of a cotorsion and a divisible group. It follows that C must be cotorsion.

If $H = D \oplus C$, with D divisible and C cotorsion, then $\text{Ext}(A, H) = 0$ for all torsion free groups A, and hence H is a direct summand of any group G containing it such that G/H is torsion free. If H is a direct summand of any such G, then clearly H is an endomorphic image of any such G. Thus our theorem is proved.

The torsion group T is a direct summand of every group containing it as its maximal torsion subgroup if and only if $T = D \oplus B$, with D divisible and B of bounded order. (See [1], pg. 187.) Thus, by our theorem, we see that *the torsion group T is an endomorphic image of every group containing it as its maximal torsion subgroup if and only if $T = D \oplus B$, with D divisible and B of bounded order.*

The solution of II goes as follows. Suppose a basic subgroup of T is an endomorphic image of every group G in which T is the maximal torsion subgroup. Let $T = D \oplus B$, with D divisible and B reduced. Then a basic subgroup of T must be an endomorphic image of $D \oplus B' = D \oplus \text{Ext}(Q/Z, B)$. Therefore a basic subgroup of T must be cotorsion, since it is reduced, and since it is torsion, it is of bounded order. (See [1], pg. 187. The remark by Harrison in [2], pg. 371 is incorrectly worded.) Writing T as $D \oplus B$, we see that a basic subgroup of B is a basic subgroup of T. But any two basic subgroups of T are isomorphic, and if B has a basic subgroup of bounded order, then B must be of bounded order. In fact, the only basic subgroup of B is B itself. Thus $T = D \oplus B$, with D divisible and B of bounded order. If $T = D \oplus B$, with D divisible and B of bounded order, then B is a basic subgroup of T. Now $D \oplus B$, and hence B, is a direct summand of any G in which T is the maximal torsion subgroup. Therefore B is an endomorphic image of any such G, and hence any basic subgroup of T

is such an endomorphic image. Thus we see that *the answers to questions* I *and* II *are the same.*

References

1. L. Fuchs, *Abelian Groups*, Budapest, 1958.
2. D. K. Harrison, *Infinite Abelian groups and homological methods*, Annals of Math., **69** (1959), 366-391.

NEW MEXICO STATE UNIVERSITY
UNIVERSITY PARK, NEW MEXICO

THE PRIME DIVISORS OF FIBONACCI NUMBERS

MORGAN WARD

1. Introduction. Let

$$(U) : U_0, U_1, U_2, \cdots, U_n, \cdots$$

be a linear integral recurrence of order two; that is,

$$U_{n+2} = PU_{n+1} - QU_n (n = 0, 1, \cdots) .$$

P, Q integers, $Q \neq 0$; U_0, U_1, integers. It is an important arithmetical problem to decide whether or not a given number m is a divisor of (U); that is, to find out whether the diophantine equation

$$U_x = my , \qquad m \geqq 2 \tag{1.1}$$

has a solution in integers x and y. Our information about this problem is scanty except in the cases when it is trivial; that is when the characteristic polynomial of the recursion has repeated roots, or when some term of (U) is known to vanish.

If we exclude these trivial cases, there is no loss in generality in assuming that m in (1.1) is a prime power. It may further be shown by p-adic methods [7] that we may assume that m is a prime. Thus the problem reduces to characterizing the set $\mathfrak{P}$ of all the prime divisors of (U). $\mathfrak{P}$ is known to be infinite [6], and there is also a criterion to decide a priori whether or not a given prime is a member of $\mathfrak{P}$, [2], [6], [7]. But this criterion is local in character and tells little about $\mathfrak{P}$ itself.

I propose in this paper to study in detail a special case of the problem in the hope of throwing light on what happens in general. I shall discuss the prime divisors of the Fibonacci numbers of the second kind:

$$(G) : 2, 1, 3, 4, 7, \cdots, G_n, \cdots$$

These and the Fibonacci numbers of the first kind

$$(F) : 0, 1, 1, 2, 3, 5, \cdots, F_n, \cdots$$

are probably the most familiar of all second order integral recurrences; (F) and (G) have been tabulated out to one hundred and twenty terms by C. A. Laisant [3].

2. Preliminary classification of primes. Let R denote the rational field and $\mathscr{R} = R(\sqrt{5})$ the root field of the characteristic polynomial

Received April 14, 1960.

$$f(x) = x^2 - x - 1 \tag{2.1}$$

of (F) and (G). Then if α and β are the roots of $f(x)$ in $\mathscr{R}$,

$$F_n = \frac{\alpha^n - \beta^n}{\alpha - \beta}, \; G_n = \alpha^n + \beta^n , \qquad (n = 0, 1, 2, \cdots).$$

If p is any rational prime, by its rank of apparition in (F) or rank, we mean the smallest positive index x such that p divides F_x. We denote the rank of p by ρ_p or ρ. Its most important properties are: $F_n \equiv o \pmod p$ if and only if $n \equiv o \pmod \rho$; $p - (5/p) \equiv o \pmod \rho$. Here $(5/p)$ is the usual Legendre symbol.

The following consequence of (2.1) and the formula $F_{2n} = F_n G_n$ is well known.

LEMMA 2.1. *p is a divisor of (G) if and only if the rank of apparition of p in (F) is even.*

The formula

$$G_n^2 - 5F_n^2 = (-1)^n 4 \tag{2.2}$$

gives more information. For if $p \equiv 1 \pmod 4$, and p divides (G), (2.2) implies that $(5/p) = 1$. On the other hand if $p \equiv 3 \pmod 4$, p must divide (G). For otherwise Lemma 2.1 and formula (2.2) with $n = \rho_p$ imply $(-1/p) = 1$.

On classifying the primes according to the quadratic characters of 5 and -1 modulo p, they are distributed into eight arithmetical progressions $20n + 1$, $20n + 3$, $20n + 7$, $20n + 9$, $20n + 11$, $20n + 13$, $20n + 17$, $20n + 19$. By the remarks above, only primes of the form $20n + 1$ and $20n + 9$ for which both -1 and 5 are quadratic residues need be considered; the following lemma disposes of all others.

LEMMA 2.2. *p is a divisor of (G) if $p \equiv 3 \pmod 4$; that is if $p \equiv 3, 7, 11, 19 \pmod{20}$. p is a non-divisor of (G) if $p = 1 \pmod 4$ and $p \equiv 2$ or $3 \pmod 5$; that is if $p \equiv 13, 17 \pmod{20}$.*

3. Further classification criteria. Let $\mathfrak{Q}$ denote the set of all primes having both 5 and -1 as quadratic residues; that is primes of the $20n+1$ or $20n+9$. For the remainder of the paper all primes considered belong to $\mathfrak{Q}$. Let $\mathfrak{P}$ denote the subset of divisors of (G) and $\mathfrak{P}^* = \mathfrak{Q} - \mathfrak{P}$ the complementary set of non-divisors of (G). We shall derive criteria to decide whether p belongs to $\mathfrak{P}$ or to $\mathfrak{P}^*$.

If p is any element of $\mathfrak{Q}$, we may write

$$p \equiv 2^k + 1 \pmod{2^{k+1}}, \; p - 1 = 2^k q, \, q \text{ odd}; \, k \geqq 2 . \tag{3.1}$$

We shall call k the (dyadic) order of p. Thus primes of order two are of the forms $40n + 21$ and $40n + 29$, primes of order three, of the form $80n + 9$ and $80n + 41$ and so on. The difficulty of classifying p as a divisor or non-divisor of (G) increases rapidly with its order.

Let R_p denote the finite field or p elements. For every $p \varepsilon \mathfrak{Q}$, the characteristic polynomial (2.2) splits in R_p:

$$(3.2) \qquad x^2 - x - 1 = (x = a)(x - b),\, a,\, b \varepsilon R_p\ .$$

If we represent the elements of R_p by the least positive residues of p, then by a classical theorem of Dedekind's, the factorization of p in the root-field $\mathscr{R}$ of $f(x)$ is given by

$$(3.3) \qquad p = \mathfrak{q}\mathfrak{q}',\ \mathfrak{q} = (p, \alpha - a),\ \mathfrak{q}' = (p, \alpha - b)\ .$$

Here $\mathfrak{q}$ and $\mathfrak{q}'$ are conjugate prime ideals of $\mathscr{R}$ of norm p.

Now assume $p \varepsilon \mathfrak{P}^*$; then rank ρ of p divides q in (3.1). Consequently $F_q \equiv o \pmod p$, so that $\alpha^q \equiv \beta^q (\bmod\ \mathfrak{q})$ in $\mathscr{R}$. But then $\alpha^{2q} \equiv \alpha^q\beta^q \equiv (-1)^q \equiv -1 \pmod{\mathfrak{q}}$ so that $a^{2q} \equiv -1 \pmod{\mathfrak{q}}$. But then $a^{2q} \equiv -1 \pmod p$ in R. Conversely, assume that $a^{2q} \equiv -1 \pmod p$. Then in $\mathscr{R}$, $\alpha^{2q} \equiv -1 \pmod{\mathfrak{q}}$ or $\alpha^{2q} \equiv (\alpha\beta)^q (\bmod\ \mathfrak{q})$, $(\alpha - \beta)\alpha^q F_q \equiv O \pmod{\mathfrak{q}}$. But $(\alpha - \beta, \mathfrak{q}) = (\alpha, \mathfrak{q}) = (1)$ in $\mathscr{R}$. Hence $F_q \equiv O \pmod{\mathfrak{q}}$ so that $F_q \equiv O \pmod p$ in R. Thus the rank of p in (F) must divide q and is consequently odd. Hence $p \varepsilon \mathfrak{P}^*$.

It follows that $p \varepsilon \mathfrak{P}^*$ if and only if $a^{2q} = -1$ in R_p. Since $(ab)^{2q} = (-1)^{2q} = +1$ in R_p, it is irrelevant which root of $f(x) = 0$ in R_p we choose for a. An equivalent way of stating this result is that $p \varepsilon \mathfrak{P}^*$ if and only if $a^{4q} \equiv 1 \pmod p$ but $a^{2q} \not\equiv 1 \pmod p$.

For ease of printing, let

$$[u/p]_n = (u/k)_{2^n}$$

denote the $2^n ic$ character of u modulo p. Thus $[u/p]_1$ is an ordinary quadratic character, $[u/p]_2$ or $(u/p)_4$ a biquadratic character and so on. The result we have obtained may be stated as follows:

THEOREM 3.1. *Let p be any prime of order $k \geqq 2$. Then if a is a root of $x^2 - x - 1$ in the finite field R_p, a necessary and sufficient condition that p belong to $\mathfrak{P}^*$ is*

$$(3.3) \qquad [a/p]_{k-1} = -1\ .$$

There is another useful way of stating this result. Let

$$(3.4) \qquad g(x) = f(x^{2^{k-2}}) = x^{2^{k-1}} - x^{2^{k-2}} - 1\ .$$

Assume that $p \varepsilon \mathfrak{P}$. Then each of the equations

$$x^{2^{k-2}} = a,\ x^{2^{k-2}} = b$$

where a, b are the roots of $f(x)$ in R_p, has 2^{k-2} roots in R_p. If c is any one of these roots, it follows from (3.4) that c is a root of $g(x)$. Hence the polynomial $g(x)$ splits completely in R_p. On the other hand since neither of the equations

$$x^{2^{k-1}} = a,\ x^{2^{k-1}} = b$$

has a root in R_p, $g(x^2)$ has no roots in R_p. Evidently, by Theorem 3.1, these splitting conditions imply conversely that $p \varepsilon \mathfrak{P}^*$. Hence

THEOREM 3.2. *Necessary and sufficient conditions that p belong to $\mathfrak{P}^*$ are that the polynomial $g(x)$ defined by* (3.4) *splits completely into linear factors modulo p, but the polynomial $g(x^2)$ has no linear factor modulo p.*

For example, assume that $p \equiv 5 \pmod 8$ so that $k = 2$. Then $g(x) = f(x)$ so the first condition of Theorem 3.2 is always satisfied. Since $g(x^2) = x^4 - x^2 - 1$ we may state the following corollary.

COROLLARY 3.1. *If p is of order two, $p \varepsilon \mathfrak{P}$ if and only if the polynomial $x^4 - x^2 - 1$ is completely reducible modulo p.*

In like manner if $p \equiv 1 \pmod 8$ so that $k \geqq 2$, we may state the following corollary

COROLLARY 3.2. *If p is of order three or more, a sufficient condition that $p \varepsilon \mathfrak{P}$ is that the polynomial $x^4 - x^2 - 1$ is not completely reducible modulo p.*

Now let

$$p = u^2 + 4v^2 \tag{3.5}$$

be the representation of p as a sum of two squares. Either u or v is divisible by 5.

LEMMA. *The polynomial $z^4 - z^2 - 1$ splits completely in R_p if and only if in the representation* (3.5) *either* $u \equiv \pm 1 \pmod 5$ *or* $v \equiv \pm 1 \pmod 5$.

Proof. Since $z^4 - z^2 - 1 = ((2z^2 - 1)^2 - 5)/4$, $z^4 - z^2 - 1$ always splits into quadratic factors in R_p. But if i denotes an element of R_p whose square is $p - 1$, then $z^4 - z^2 - 1 = (z^2 + i)^2 - (1 + 2i)z^2$. Hence a necessary and sufficient condition that $z^4 - z^2 - 1$ split completely in R_p is that $1 + 2i = ((-1)(-1 - 2i))$ be a square in R_p.

Now let $\mathfrak{T}$ denote the ring of the Gaussian integers, and let $p = (u + 2iv)(u - 2iv)$ be the decomposition of p into primary factors in $\mathfrak{T}$.

(Bachmann [1]). Then $u - 2iv$ is a prime ideal of norm p so that the residue class ring $\mathfrak{T}/(u - 2iv)$ is isomorphic to R_p. Now $-1 - 2i$ is primary in $\mathfrak{T}$. Also since $p \equiv 1 \pmod 4$, -1 is a quadratic residue of $u - 2iv$. Hence $1 + 2i$ is a square in R_p if and only if $-1 - 2i$ is a quadratic residue of $u - 2iv$ in $\mathfrak{T}$. By the quadratic reciprocity law in $\mathfrak{T}$, (Bachmann [1])

$$\left(\frac{-1-2i}{u-2iv}\right) = \left(\frac{u-2iv}{-2-2i}\right) = \left(\frac{u+v}{-1-2i}\right) .$$

Now either u or v must be divisible by $-1 - 2i$. But $(-1 - 2i)$ is a prime ideal in $\mathfrak{T}$ of norm five. Therefore $-1 - 2i$ is a quadratic residue of $u - 2iv$ if and only if $u \equiv 0$, $v \equiv 1, 4 \pmod 5$ or $v \equiv 0$, $u \equiv 1, 4 \pmod 5$. This completes the proof of the lemma.

On combining the results of Corollaries 3.1 and 3.2 into the lemma, we obtain

THEOREM 3.3. *Let p be congruent to* 5 *modulo* 8. *Then a necessary and sufficient condition that $p \varepsilon \mathfrak{P}$ is that in the representation* (3.5) *of p as a sum of two squares, either $u \equiv \pm 1 \pmod 5$ or $v \equiv \pm 1 \bmod 5$. If p is congruent to* 1 *modulo* 8, *a sufficient condition that $p \varepsilon \mathfrak{P}$ is that $u \equiv \pm 2 \pmod 5$ or $v \equiv \pm 2 \bmod 5$.*

4. Applications of the criteria. The theorems of § 3 classify unambiguously all primes of $\mathfrak{Q}$ either into $\mathfrak{P}$ or into $\mathfrak{P}^*$. But in the absence of workable reciprocity laws beyond the biquadratic case, they tell us little more than Lemma 2.1 for primes of order greater than three; that is, primes of the forms $160n + 9$ or $160n + 81$. However the theorems may be extended so as to give useful information about primes of any order by utilizing the following elementary properties of the character symbol $[u/p]_k$:

$$(4.1)\qquad \begin{aligned} &[uv/p]_k = [u/p]_k[v/p]_k \\ &[u^2/p]_k = [u/p]_k^2 = [u/p]_{k-1} \\ &[u/p]_k = 1 \text{ implies } [u/p]_n = 1 \text{ for } 1 \leqq n \leqq k-1 . \end{aligned}$$

From (4.1) (iii) and Theorem 3.1 we immediately obtain.

THEOREM 4.1. *If p is of order $k \geqq 3$, then a necessary condition that p belong to $\mathfrak{P}^*$ is that*

$$(4.2)\qquad [a/p]_n = 1 \qquad (n = 1, 2, \cdots, k-2) .$$

COROLLARY 4.1. *A sufficient condition that p belong to $\mathfrak{P}$ is that* (4.2) *be false for some $n \leqq k - 2$.*

Now suppose that a solution $x = c$ of the congruence $c^2 \equiv a \pmod{p}$ is known, p of order four or more. Then by (4.1) (ii) and the theorem just proved we obtain.

THEOREM 4.2. *If p is of order $k \geqq 4$, then a necessary condition that p belong to $\mathfrak{P}^*$ is that*

$$[c/p]_n = 1\ , \qquad (n - 1, 2, \cdots, k - 3). \tag{4.4}$$

A necessary and sufficient condition that p belong to $\mathfrak{P}^*$ is that

$$[c/p]_{k-2} = -1\ . \tag{4.5}$$

There is a method for obtaining a, the root of (2.1) modulo p, which leads to another useful criterion for primes of low order. For every prime p of $\mathfrak{O}$ there exists a unique representation in the form

$$p = r^2 - 5s^2,\ 0 < r,\ 0 < s < \sqrt{4p/5}\ . \tag{4.6}$$

(Uspensky [5]). If this representation is known, a is easily shown to be the least positive solution of the congruence

$$2sa \equiv (r + s) \quad (\mathrm{mod}\ p.)\ . \tag{4.7}$$

By using property (4.1) (i) of the character symbol and Theorem 3.1, we see that

$$[2s/p]_{k-1} = -[(r + s)/p]_{k-1}$$

is a necessary and sufficient condition that p belong to $\mathfrak{P}^*$.

If $k = 2$, the criterion becomes $(2s/p) = -((r + s)/p)$. But since $p \equiv 5 \pmod{8}$ and $p = r^2 - 5s^2$, r is odd and $s = 2s'$ where s' is odd. Hence by the reciprocity law for the Jacobi symbol, $(2s/p) = (s'/p) = (p/s') = (r^2/s') = +1$. Hence $p \varepsilon \mathfrak{P}^*$ if and only $((r + s)/p) = -1$. But $((r + s)/p) = ((r^2 - 5s^2)/(r + s)) = (-4s^2/(r + s)) = (-1/(r + s)) = (-1)^{(r+1)/2}$ since $s \equiv 2 \pmod{4}$. We have thus proved

THEOREM 4.3. *If p is of order two, so that p is of the form $40n + 21$ or $40n + 29$, then p belongs to $\mathfrak{P}$ or to $\mathfrak{P}^*$ according as r in the representation* (4.6) *is congruent to three or one modulo* 4.

Now if $k > 2$, $p \equiv 1 \pmod{8}$ so that r in the representation (4.6) is odd. Hence using the corollary to Theorem 4.1 with $n = 1$ and the results established in the proof of Theorem 4.3, we obtain

THEOREM 4.4. *If p is of order greater than two, p belongs to $\mathfrak{P}$ if r in the representation* (4.6) *is congruent to one modulo* 4.

To illustrate, suppose that $p = 101$. Then $p \equiv 5 \pmod{8}$ so that

Theorem 3.3 is applicable. Since $101 = 1^2 + 4 \cdot 5^2$, $101 \varepsilon \mathfrak{P}$. Also $101 = 11^2 - 5 \cdot 2^2$ and $11 \equiv 3 \pmod 4$. Hence $101 \varepsilon \mathfrak{P}$ by Theorem 4.3. In fact we find from Laisant's table that $G_{50} = 12586269025 = 101 \times 124616525$.

Again, there are seven primes in $\mathfrak{Q}$ less than one thousand of order greater than three; namely 241, 401, 449, 641, 769, 881 and 929. But only two of these need be discussed; Theorem 3.3 assigns 241, 449, 641, 881 and 929 to $\mathfrak{P}$. For $241 = 15^2 + 4.2^2$, $449 = 7^2 + 4.10^2$, $641 = 25^2 + 4.2^2$, $881 = 25^2 + 4.8^2$ and $929 = 23^2 + 4.10^2$. There remain 401 and 729. Now $401 \equiv 17 \pmod{32}$. Hence $k = 4$. Since $112^2 - 112 - 1 = 31 \times 401$, $a = 112$. Hence by Theorem 3.1, $401 \varepsilon \mathfrak{P}^*$ if and only if $[112/401]_3 = -1$. Now using the idea in Theorem 4.2, $112 = 2^4 \times 7$ and $85^2 \equiv 7 \pmod{401}$. Hence $[112/401]_3 = [85/401]_2$. But $(85/401) = -1$. Hence $401 \varepsilon \mathfrak{P}$. This conclusion is easily checked. For $401 - 1 = 25.16$ and by Laisant's table, $F_{25} = 75025 \not\equiv 0 \pmod{401}$. Hence $401 \varepsilon \mathfrak{P}$ by Lemma 2.1.

Finally $769 \equiv 257 \pmod{512}$ so that $k = 8$. Using Jacobi's Canon, $a = 43$, ind $a = 500 \not\equiv 0 \pmod{64}$ so that $769 \varepsilon \mathfrak{P}$. Indeed $769 - 1 = 3 \cdot 256$ and $F_3 = 2$. Hence $769 \varepsilon \mathfrak{P}$ by Lemma 2.1.

We have shown incidentally that every prime $p < 1000$ in $\mathfrak{Q}$ of order greater that three is a divisor of (G).

5. Conclusion. The methods of this paper may be easily extended to obtain information about the prime divisors of the Lucas or Lehmer [4] numbers of the second kind $\alpha^n + \beta^n$ where α and β now are the roots of any quadratic polynomial $x^2 - \sqrt{P}x + Q$ with P, Q integers, $Q(P - 4Q) \neq 0$. It is worth noting that just as in the special case $P = 1$ $Q = -1$ investigated here, there will be arithmetical progressions whose primes cannot be characterized as divisors or non-divisors by their quadratic or biquadratic characters alone.

In the absence of any criterion like Lemma 2.1 for a prime divisor of an arbitrarily selected recurrence (U), it seems difficult to characterize the divisor of (U) in any general way. It would be interesting to make a numerical study of several recurrences (U) to endeavor to find out whether the two Lucas sequences $0, 1, P, \cdots$ and $2, P, P^2 - 2Q, \cdots$ and their translates are essentially the only ones for which a global characterization of the divisors is possible.

References

1. Paul Bachmann, *Kreistheilung*, Leipzig (1921), 150-185.
2. Marshall Hall, *Divisors of second order sequences*, Bull. Amer. Math. Soc., **43** (1937), 78-80.
3. C. A. Laisant, *Les deux suites Fibonacciennes fondamentales*, Enseignement Math., **21** (1920), 52-56.
4. D. H. Lehmer, *An extended theory of Lucas functions*, Annals of Math., **31** (1930), 419-448.

5. J. V. Uspensky and M. A. Heaslet, *Elementary number theory,* New York (1939), 358–359.
6. Morgan Ward, *Prime divisiors of second order recurrences,* Duke Math. Journal **21** (1954), 607–614.
7. ———, *The linear p-adic recurrence of order two,* Unpublished.

CALIFORNIA INSTITUTE, PASADENA.

ON THE NILPOTENCY CLASS OF A GROUP OF EXPONENT FOUR

C. R. B. WRIGHT

Introduction. If G is a multiplicative group with elements $x, y, \cdots$, we define the commutator (x, y) by $(x, y) = x^{-1}y^{-1}xy$ and, inductively for length n, $(x_1, \cdots, x_{n-1}, x_n) = ((x_1, \cdots, x_{n-1}), x_n)$. We also use the notation $(x, \cdots, y; \cdots; z, \cdots, w)$ for the commutator $((x, \cdots, y), \cdots, (z, \cdots, w))$. For each positive integer n, let G_n be the subgroup of G generated by all commutators of length n.

A group, G, is of exponent 4 in case $x^4 = 1$ for every x in G but $y^2 \neq 1$ for some y in G. Let F be a free group of rank k, and let F^4 be the subgroup generated by fourth powers of elements of F. Let $B(k) = F/F^4$. Then $B(k)$ is clearly a group of exponent 4 on k generators. Moreover, every group of exponent 4 on k generators is a homomorphic image of $B(k)$.

I. N. Sanov has shown that $B(k)$ is finite. (See [2], pp. 324–325, or [3]). Unfortunately, his proof gives very little additional information about $B(k)$. The present paper is devoted to the study of relations between commutators in the group $B(k)$, a consequence of the relations obtained being that $B(k)_{3k} = 1$.

Preliminaries. Let G be a group of exponent 4, and let $a, b, \cdots$ be elements of G. Then

$$(a, b)^2 \equiv (a, b, b, b)(a, b, b, a)(a, b, a, a) \bmod G_4 \tag{1}$$

$$(a, b, a)^2 = (a, b, a, a, a) = (a, b, a; a, b) \tag{2}$$

$$(a, b, c) \equiv (b, c, a)(c, a, b) \bmod G_4 \tag{3}$$

$$(a, b; c, d) \equiv (a, c; b, d)(a, d; b, c) \bmod G_5 \tag{4}$$

$$(\mathbf{a}, b; \mathbf{c}, d; f) \equiv (\mathbf{a}, d; \mathbf{c}, f; b)(\mathbf{a}, f; \mathbf{c}, b; d) \bmod G_6 \tag{5}$$

where the bold-face type in (5) has no significance other than to point out which entries are left fixed while the others are cyclicly permuted—whenever bold-face type appears in a computation an application of (5) is about to be made. The relations (1) and (2) can be shown to hold in $B(2)$; hence they certainly hold in any group, G, of exponent 4. Relation (3) is simply the Jacobi identity (which holds in any group) adapted to exponent 4. Relations (4) and (5) were proved in [4] for the case in which the entries are of order 2, but the proofs clearly go through without this restriction, since in proving the relations we are simply

Received April 25, 1960. This work was supported by the Office of Naval Research under contract MR 044-213.

looking at the first significant terms of $(abcd)^4$ and $(abcdf)^4$ as collected by P. Hall's process. It should be noted that these relations are "identical" in the sense that they hold for every choice of a, b, c, d and f in G. This property gives us the freedom of substitution which we shall use later.

The following result, which appeared in a slightly different form as the Corollary to Lemma 3.2 in [4], is easily proved using (1) and (3).

(A). *Let G be a group of exponent* 4. *Let*

$$C = (x_1, \cdots, x_i, a, x_{i+1}, \cdots, x_{n-1})$$

where $x_1, \cdots, x_{n-1}$ and a are in G. Then, modulo G_{n+1}, C is a product of commutators of the form $(a, y_1, \cdots, y_i, x_{i+1}, \cdots, x_{n-1})$, where $y_1, \cdots, y_i$ are $x_1, \cdots, x_i$ in some order.

Finally, we need to know that if a and b are the generators of $B(2)$, then $B(2)_5$ is generated by $(b, a, a; b, a)$ and $(b, a, b; b, a)$, and $B(2)_6 = 1$. These results may be verified directly or deduced from Burnside's original work in [1].

Throughout this paper we shall be concerned with the relations between commutators in $B(k)$. Our first lemma gives us a method of reducing our problems to a few relatively tractable cases.

LEMMA 1. *Suppose $(x_1, \cdots, x_n)$ is a commutator of length n in a group, G, of exponent 4. If one of $x_3, \cdots, x_n$ is a and one b, then, modulo G_{n+1}, $(x_1, \cdots, x_n)$ is a product of commutators of length n of the following four types*:

(i) $(x, y, \cdots, a, b, \cdots)$

(ii) $(x, y, \cdots, b, a, \cdots)$

(iii) $(x, y, \cdots, a, z, b)$

(iv) $(x, y, \cdots, b, z, a)$.

Loosely stated, Lemma 1 says that we may bring a and b more or less together and keep them out of the first two positions.

Proof of Lemma 1. Observe first that we can rewrite (3) as

$$(a, b, c) \equiv (a, c, b)(a; b, c) \bmod G_4 .$$

Using this form and working modulo G_7 we have

$$\begin{aligned}(x, y, a, z, b, w) &\equiv (x, y, a, b, z, w)(\boldsymbol{x}, \boldsymbol{y}, a; \boldsymbol{b}, z; w)\\ &\equiv (x, y, a, b, z, w)(x, y, z; b, w; a)(x, y, w; b, a; z)\\ &\equiv (x, y, a, b, z, w)(x, y, z, b, w, a)(x, y, z, w, b, a)\\ &\quad \cdot(x, y, w, b, a, z)(x, y, w, a, b, z) .\end{aligned}$$

Let $G(n, a, b)$ be the (normal) subgroup of G generated by G_{n+1} and all commutators of length n of types (i) and (ii). Let $G^*(n, a, b)$ be the (normal) subgroup of G generated by $G(n, a, b)$ and all commutators of length n of types (iii) and (iv). Then certainly if w is in $G(n, a, b)$ and g is in G, (w, g) is in $G(n+1, a, b)$, and by the relation just proved, if z is in $G^*(n, a, b)$, then (z, g) is in $G^*(n+1, a, b)$. Thus it will be sufficient to prove the lemma under the assumption that x_n is either a or b (say b).

We have reduced the problem to showing that if C has length n and if $C = (x_1, x_2, \cdots, x_i, a, \cdots, b)$, then C is in $G^*(n, a, b)$. If $2 \leqq n - i \leqq 3$, then C is in $G^*(n, a, b)$. We proceed by induction on $n - i$. Suppose for induction that for some $j \geqq 4$ and all $n \geqq j + 2$, C is in $G^*(n, a, b)$ whenever $n - i < j$. We shall show that if $n - i = j$, then C is in $G^*(n, a, b)$, so that by finite induction we shall have C in $G^*(n, a, b)$ for all i such that $2 \leqq n - i \leqq n - 2$, i.e., such that $2 \leqq i \leqq n - 2$. Thus the lemma will be proved.

Let $i = n - j$. By the inductive assumption and (3) we have, modulo $G^*(n, a, b)$,

$$(x_1, x_2, \cdots, x_i, a, \cdots, x_{n-3}, x_{n-2}, b) \equiv (X, x_i; A, x_{n-3}; x_{n-2}; b) ,$$

where $X = (x_1, \cdots, x_i)$, and where $A = (a, \cdots, x_{n-4})$ if $n - 4 > i$ but $A = a$ if $n - 4 = i$. Now, modulo G_{n+1}, using (4), (3) and (5),

$$\begin{aligned}(X, x_i; A, x_{n-3}; x_{n-2}; b) &\equiv (X, x_{n-3}; A, x_i; x_{n-2}; b)(x_i, x_{n-3}; A, X; x_{n-2}; b)\\ &\equiv (X, x_{n-3}, x_{n-2}; A, x_i; b)(\boldsymbol{A}, \boldsymbol{x}_i, x_{n-2}; X, \boldsymbol{x}_{n-3}; b)\\ &\qquad \cdot(\boldsymbol{A}, \boldsymbol{X}, x_{n-2}; x_i, \boldsymbol{x}_{n-3}; b)(\boldsymbol{A}, X; \boldsymbol{x}_i, \boldsymbol{x}_{n-3}, x_{n-2}; b)\\ &\equiv (X, x_{n-3}, x_{n-2}; A, x_i; b)(A, x_i, X; b, x_{n-3}; x_{n-2})(A, x_i, b; x_{n-2}, x_{n-3}; X)\\ &\qquad \cdot(A, X, x_i; b, x_{n-3}; x_{n-2})(A, X, b; x_{n-2}, x_{n-3}; x_i)\\ &\qquad \cdot(A, x_{n-2}; x_i, x_{n-3}, b; X)(A, b; x_i, x_{n-3}, X; x_{n-2}) .\end{aligned}$$

But by the inductive assumption $(X, x_{n-3}, x_{n-2}; A, x_i; b)$, $(A, x_i, b; x_{n-2}, x_{n-3}; X)$, $(x_i, x_{n-3}, b; A, x_{n-2}; X)$ and $(A, b; x_{n-3}, x_i, X; x_{n-2})$ are all in $G^*(n, a, b)$. Further,

$$\begin{aligned}&(A, x_i, X; b, x_{n-3}; x_{n-2})(A, X, x_i; b, x_{n-3}; x_{n-2})\\ &\qquad\equiv (X, x_i, A; b, x_{n-3}; x_{n-2}) \bmod G_{n+1} .\end{aligned}$$

Thus, modulo $G^*(n, a, b)$,

$$\begin{aligned}&(X, x_i; A, x_{n-3}; x_{n-2}; b)\\ &\qquad\equiv (\boldsymbol{X}, \boldsymbol{x}_i, A; \boldsymbol{b}, x_{n-3}; x_{n-2})(A, X, b; x_{n-2}, x_{n-3}; x_i)\\ &\qquad\equiv (X, x_i, x_{n-3}; b, x_{n-2}; A)(X, x_i, x_{n-2}; b, A; x_{n-3})\\ &\qquad\quad \cdot(A, X; x_{n-2}, x_{n-3}, b; x_i)(x_{n-2}, x_{n-3}; A, X; b; x_i)\end{aligned}$$

$$\begin{aligned}
&\equiv (\boldsymbol{A}, X; \boldsymbol{x}_{n-2}, \boldsymbol{x}_{n-3}, b; x_i)(\boldsymbol{x}_{n-2}, x_{n-3}; \boldsymbol{A}, X; b; x_i) \\
&\equiv (A, b; x_{n-2}, x_{n-3}, x_i; X)(A, x_i; x_{n-2}, x_{n-3}, X; b) \\
&\qquad \cdot(x_{n-2}, X; A, b; x_{n-3}; x_i)(x_{n-2}, b; A, x_{n-3}; X; x_i) \\
&\equiv (x_{n-2}, b; A, x_{n-3}; X; x_i) \\
&\equiv (\boldsymbol{x}_{n-2}, b; \boldsymbol{A}, x_{n-3}; x_i; X)(\boldsymbol{x}_{n-2}, b; A, \boldsymbol{x}_{n-3}; X, x_i) \\
&\equiv (x_{n-2}, x_{n-3}; A, x_i; b; X)(x_{n-2}, x_i; A, b; x_{n-3}; X) \\
&\qquad \cdot(x_{n-2}, A; X, x_i, x_{n-3}; b)(X, x_i, x_{n-2}; b, x_{n-3}; A) \\
&\equiv 1\,.
\end{aligned}$$

Hence, $(x_1, x_2, \cdots, x_i, a, \cdots, x_{n-3}, x_{n-2}, b)$ is in $G^*(n, a, b)$, as desired. Thus the lemma is proved.

An immediate consequence of Lemma 1 is the following.

COROLLARY. *If $C = (x_1, \cdots, x_n)$ and if two of $x_3, \cdots, x_n$ are a, then modulo G_{n+1}, C is a product of commutators of length n of the forms*:

(i) $(x, y, \cdots, a, a, \cdots)$

(ii) $(x, y, \cdots, a, z, a)$.

We next observe that, using (1),

$$\begin{aligned}
(x_1, \cdots, x_m, a^2) &= (x_1, \cdots, x_m, a)^2(x_1, \cdots, x_m, a, a) \\
&\equiv (x_1, \cdots, x_m, a, a) \bmod G_{m+3}\,.
\end{aligned}$$

Hence,

$$(x_1, \cdots, x_i, a, a, x_{i+1}, \cdots, x_n) \equiv (x_1, \cdots, x_i, a^2, x_{i+1}, \cdots, x_n) \tag{6}$$

modulo G_{n+3}.

We may now prove the following useful result.

LEMMA 2. *Let G be a group of exponent 4, and let $(x_1, \cdots, x_n)$ be a commutator of length n in elements of G. If some three of $x_3, \cdots, x_n$ are a, then modulo G_{n+1}, $(x_1, \cdots, x_n)$ is a product of commutators of the forms*:

(i) $(y_1, y_2, \cdots, y_{n-3}, a, a, a)$

(ii) $(y_1, y_2, \cdots, y_{n-4}, a, a, y_{n-3}, a)$.

Proof. We first derive two shifting relations. Using (1) and (3) we obtain modulo G_7,

$$\begin{aligned}
(x, y, a, a, a, z) &\equiv ((x, y, a)^2, z) \equiv (x, y, a, z)^2 \equiv (x, y; a, z)^2(x, y, z, a)^2 \\
&\equiv (x, y, z, a)^2 \equiv (x, y, z, a, a, a)\,.
\end{aligned}$$

Hence,

$$(7) \qquad (x, y, a, a, a, z) \equiv (x, y, z, a, a, a) \bmod G_7 .$$

Thus, modulo longer commutators, a string of three a's can be shifted to the right.

We also have, modulo G_7,

$$(x, y, a, a, z, a) \equiv (x, y, a, z, a, a) \cdot (x, y, \boldsymbol{a}; z, \boldsymbol{a}; a) \equiv (x, y, a, z, a, a) .$$

Thus

$$(8) \qquad (x, y, a, z, a, a) \equiv (x, y, a, a, z, a) \bmod G_7 .$$

Further, modulo G_8,

$$\begin{aligned}(x, y, a, a, z, a, w) &\equiv (x, y, a, a, a, z, w)(x, y, a, a; a, z; w) \\ &\equiv (x, y, a, a, a, z, w)(\boldsymbol{x}, \boldsymbol{y}, a^2; \boldsymbol{a}, z; w) \\ &\equiv (x, y, a, a, a, z, w)(x, y, z; a, w; a^2) \\ &\equiv (x, y, a, a, a, z, w)(x, y, z, a, w, a, a)(x, y, z, w, a, a, a) .\end{aligned}$$

Applying (7) and (8) we get

$$(9) \qquad (x, y, a, a, z, a, w) \equiv (x, y, z, a, a, w, a) \bmod G_8 .$$

Thus, modulo longer commutators, a trio of a's with one gap may be shifted to the right.

It is clear from (7) and (9) that it is sufficient to prove the lemma under the assumption that $x_n = a$. Considering $(x_1, \cdots, x_{n-1})$ now, it is clear from the Corollary of Lemma 1 that we may restrict ourselves to the consideration of commutators of the following two types:

I $(x_1, x_2, \cdots, a, a, \cdots, x_{n-1}, a)$

II $(x_1, x_2, \cdots, a, x_{n-1}, a, a)$.

By (8), commutators of type II are already of type (ii), Further,

$$(x_1, x_2, \cdots, a, a, \cdots, x_{n-1}, a) \equiv (x_1, x_2, \cdots, a^2, \cdots, x_{n-1}, a) \bmod G_{n+1} .$$

Now applying Lemma 1 with b replaced by a^2 we find that modulo G_{n+1}, $(x_1, x_2, \cdots, a^2, \cdots, x_{n-1}, a)$ is a product of commutators of form $(y_1, y_2, \cdots, a, a, a, \cdots)$ and commutators of form $(y_1, y_2, \cdots, a, a, y_{n-1}, a)$. Thus, by (7), any commutator of type I is a product to commutators of types (i) and (ii) modulo G_{n+1}. The lemma follows.

The main theorems.

In this section we derive more consequences of Lemma 1 and find an upper bound on the nilpotency class of $B(k)$. The first theorem is much like Lemma 2.

THEOREM 1. *Let G be a group of exponent* 4, *and suppose* $(x_1, \cdots, x_n)$

is a commutator of length n with entries from G such that some four (or more) of $x_1, \cdots, x_n$ are a. If $n \geqq 6$, then $(x_1, \cdots, x_n)$ is in G_{n+1}.

Proof. If $(x_1, \cdots, x_n) \neq 1$, then since four entries of $(x_1, \cdots, x_n)$ are a, it follows that at least three of $x_3, \cdots, x_n$ are a. By Lemma 2 and (A) we may restrict attention to commutators of the following types:

(i) $(a, x_2, \cdots, x_{n-3}, a, a, a)$

(ii) $(a, x_2, \cdots, a, a, x_{n-3}, a)$.

Applying (7) and (9), we may confine our study to commutators of the following types:

(i′) $(a, x_2, a, a, a, x_3, \cdots, x_{n-3})$

(ii′) $(a, x_2, a, a, x_3, a, \cdots)$.

But now, modulo G_7, using (2) and (5),

$$(a, x, a, a, a, y) \equiv (a, x, a; a, x; y) \equiv (a^2, \boldsymbol{x}; \boldsymbol{a}, x; y) \equiv 1 ,$$

and

$$\begin{aligned}(a, x, a, a, y, a) &\equiv (a, x, a^2, y, a) \equiv (a, x, y, a^2, a)(\boldsymbol{a}, x; \boldsymbol{a}^2, y; a)\\ &\equiv (a, x, y, a, a, a) \equiv (a, x, a, a, a, y) = 1 .\end{aligned}$$

Thus a commutator of type (i′) or (ii′) is in G_{n+1}. The theorem follows.

Let $x_1, \cdots, x_k$ be generators of $B(k)$. Then it is easy to show that $x_1, \cdots, x_{k-1}$ generate a group isomorphic to $B(k-1)$. We may thus consider $B(k-1)$ as imbedded in $B(k)$.

If A and B are subgroups of a group, G, we define (A, B) as the subgroup of G generated by all commutators (a, b) with a in A and b in B.

THEOREM 2. *For each positive integer k,*

$$(B(k)_{3k-1}, B(k+1)) \subseteqq B(k+1)_{3k+1} .$$

Proof. We first show that the theorem holds for $k = 2$, then we proceed by induction on k. Thus suppose first that $k = 2$. Now as noted above, $B(2)_5$ is generated by $(x_1, x_2, x_1; x_1, x_2)$ and $(x_2, x_1, x_2; x_2, x_1)$. But if y is in $B(3)$, then modulo $B(3)_7$,

$$(x_1, x_2, x_1; x_1, x_2; y) = (x_1^2, \boldsymbol{x}_2; \boldsymbol{x}_1, x_2; y) \equiv 1 .$$

Similarly, $(x_2, x_1, x_2; x_1, x_2; y) \equiv 1$ modulo $B(3)_7$. Thus the theorem is true if $k = 2$.

Now suppose inductively that for some n the theorem is true for all k such that $2 \leqq k < n$. We shall show that

$$(B(n)_{3n-1}, B(n+1)) \subseteqq B(n+1)_{3n+1} .$$

It will be sufficient to show that if $y_1, \cdots, y_{3n-1}$ are chosen in any way from $x_1, \cdots, x_n$ and if z is in $B(n+1)$, then $(y_1, \cdots, y_{3n-1}, z)$ is in $B(n+1)_{3n+1}$. Now if four of $y_1, \cdots, y_{3n-1}$ are equal, then by Theorem 2 $(y_1, \cdots, y_{3n-1}, z)$ is in $B(n+1)_{3n+1}$. Thus suppose the contrary, i.e., suppose that each of (say) $x_2, \cdots, x_n$ appears three times among $y_1, \cdots, y_{3n-1}$ and that x_1 appears twice. By (A) we may restrict attention to the case in which $y_1 = x_1$. But in this case, since $n \geqq 3$, we must have at least one (say x_n) of $x_2, \cdots, x_n$ appearing three times among $y_3, \cdots, y_n$, so that by Lemma 2 we may restrict ourselves to consideration of commutators of the following types:

(i) $(y_1, y_2, \cdots, y_{3n-4}, x_n, x_n, x_n, z)$

(ii) $(y_1, y_2, \cdots, x_n, x_n, y_{3n-4}, x_n, z)$,

where x_1 appears twice among $y_1, \cdots, y_{3n-4}$ and each of $x_2, \cdots, x_{n-1}$ appears three times. Now by (9),

$$(y_1, y_2, \cdots, x_n, x_n, y_{3n-4}, x_n, z) \equiv (y_1, \cdots, y_{3n-4}, x_n, x_n, z, x_n)$$

modulo $B(n+1)_{3n+1}$. But $(y_1, \cdots, y_{3n-4})$ is in $B(n-1)_{3(n-1)-1}$ so that, by the inductive assumption, a commutator of type (i) or type (ii) is in $B(n+1)_{3n+1}$. The theorem follows.

Finally, we have the principal goal of this paper.

THEOREM 3. *For each positive integer k, $B(k)_{3k} = 1$.*

Proof. It follows immediately from Theorem 2 that $B(k)_{3k} = B(k)_{3k+1}$ so that, since $B(k)$ is nilpotent, $B(k)_{3k} = 1$.

One may apply the foregoing results to obtain numerical estimates of the derived length and order of $B(k)$. It follows immediately from Theorem 3 that if $B(k)^{(m)} \neq 1$, then $2^m < 3k$, so that the derived length of $B(k)$ is at most $\log_2(3k-1)$. By means of the Witt formulae (see, for example, [2], p. 169) one can also obtain an upper bound on the order of $B(k)$ using Theorems 2 and 3. Such estimates, both of derived length and order, are easily seen to be imprecise. For example, the Witt formula calculations give the order of $B(3)$ as at most 2^{484}, whereas a little direct computation shows that the order is at most 2^{70}. Also, $\log_2(3 \cdot 3 - 1) = 3$, but one can show that $B(3)''' = 1$.

Finally we would like to point out that it can be shown that $B(k)_k \neq 1$, so that perhaps the upper bound on the class given here is not too far from the true class. Indeed, the bound is precise for $k = 2$, and preliminary work suggests that it may be precise for $k = 3$.

BIBLIOGRAPHY

1. W. Burnside, *On an unsettled question in the theory of discontinuous groups*, Quart. J. Pure and Appl. Math., **33** (1902), 230-238.
2. M. Hall, Jr., *The Theory of Groups*, the Macmillan Co., 1959.
3. I. N. Sanov, *Solution of Burnside's problem for exponent* 4, Leningrad State Univ. Ann., **10** (1940), 166-170.
4. C. R. B. Wright, *On groups of exponent four with generators of orders two*, Pacific, J. Math., **10**, (1960), 1097-1105.

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pacific Journal of Mathematics

Vol. 11, No. 1 November, 1961