Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 8, No. 1 **March, 1958**

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

COMMUTATIVE LINEAR DIFFERENTIAL OPERATORS

S. A. AMITSUR

1 Introduction. Let $D=d/dx$ be the operator of differentiation with respect to a variable x. Let $f(D)=a_0D^n+\cdots+a_n$, $a_0\neq 0$ be a differential operator of degree n. The problem we intend to study in this paper is to determine the set $C[f]$ of all linear operators which commute with f. This problem, is old and for a complete discussion of old and new results see the report of H. Flanders [2]. The most pronounced result in this subject is the fact that $C[f]$ is a commutative ring and that it is finitely generated over the algebra of all polynomials in $f(D)$ with constant coefficients.

In his report [2], Flanders obtains this theorem by algebraic methods with the aid of a deep theorem of Tsen on division algebras over the field of all rational functions in one variable. The first part of the present paper contains a simple algebraic proof of this result which uses only elementary facts of linear algebra.

In the second part of this paper we obtain necessary and sufficient conditions for the existence of non-trivial differential operators which commute with $f(D)$. This is obtained by adjoining a parameter λ to the domain of definition of the coefficient of $f(D)$ and by considering the invariant ring [1] of the operator $f(D)-\lambda$. It is shown that the structure of the $C[f]$ is closely related with the factorization of $f(D)-\lambda$. In this part, use is made of the theory of abstract differential polynomials as developed in [4], [3] and [1]. All proofs are purely algebraic.

2. The centralizer of $f(D)$. To be more precise we make the following assumptions: Let K be a field of characteristic zero with a derivation $D: a\to a'$. Let F denote the field of constants of K. That is: $F=\{a\,;\ a\in K.\ \ a'=0\}$.

Let $K[D]$ be the ring of all *formal* differential polynomials $p(D)=p_0D^n+\cdots+p_m$, $p_i\in K$ with multiplication defined in $K[D]$ by the relation

$$Da=aD+a'\,, \qquad a\in K\,.$$

Clearly $K[D]$ can be considered also the ring of linear differential operators on K.

Let $f(D)=a_0D^n+a_1D^{n-1}+\cdots+a_n$, $n\geqq 1$, $a_0\neq 0$ be a polynomial of degree n in $K[D]$. We shall denote by $C[f]$ the centralizer of f in $K[D]$. That is, $C[f]=\{g(D)\,;\ g(D)\in K[D],\ gf=fg\}$. Clearly $C[f]$ is a subring of $K[D]$ and it contains the ring $F[f]$ of all polynomials in $f(D)$ with constant coefficients.

Received November 14, 1957.

The main object of the first section is to prove the following.

THEOREM 1. (1) *$C[f]$ is a free $F[f]$-module of dimension ρ, where ρ is a divisior of n (=the degree of $f(D)$)*

(2) *$C[f]$ is a commutative ring.*

We shall need the following two known lemmas.

LEMMA 1. ([2] Lemma 10.1) *If p_0, q_0 are respectively the leading coefficients of two polynomials $p(D)$, $q(D)$ of the same degree which commute with $f(D)$ then $p_0=cq_0$ for some constant $c \in F$.*

LEMMA 2. ([2] Lemma 10.2) *The set of all polynomials of $C[f]$ of degree $\leqq m$ is a finite—dimensional vector space over the field of constants F.*

For completeness we include the proof of these lemmas in the abstract case we are dealing with.

Indeed, if $p(D)f(D)=f(D)p(D)$ and $p(D)=p_0D^m+p_1D^{m-1}+\cdots+p_m$, then by comparing the coefficient of D^{n+m-1} on both sides we obtain:

$$ma_0'p_0+a_0p_1+p_0a_1=np_0'a_0+p_0a_1+p_1a_0$$

Thus, the leading coefficient p_0 satisfies the homogeneous linear differential equation: $na_0p_0'-ma_0'p_0=0$. Hence if $q(D)=q_0D^m+\cdots+q_m$ also commutes with $f(D)$, q_0 satisfies the same differential equation and, therefore $q_0=cp_0$ for some constant c, which proves Lemma 1.

The proof of Lemma 2 follows immediately from Lemma 1, by induction on the degree m.

We proceed now with the proof of Theorem 1:

Let Z_f be the set of all integers which are the degrees of the polynomials of $C[f]$. Since $C[f]$ is a ring, and since $\deg(p(D)q(D))=\deg p(D)+\deg q(D)$, it follows that Z_f is closed under addition. Let $\overline{m}$ denote the residue class modulo $n(=\deg f(D))$ of the integer m, and let $\overline{Z}_f=\{\overline{m}\,;\, m\in Z_f\}$. Then clearly $\overline{Z}_f$ is a subgroup of the additive cyclic group of all residue classes mod n. Hence $\overline{Z}_f$ is cyclic of order ρ and ρ is a divisor of n.

Let $\overline{o}=\overline{m}_1, \cdots, \overline{m}_\rho$ be the ρ classes mod n of $\overline{Z}_f$ and let m_i be the minimal integer of its class $\overline{m}_i$. Let $g_i(D)\in C[f]$ be a polynomial of degree m_i and we can clearly choose $g_1=1$. We shall show that these polynomials g_i are free generators of $C[f]$ over $F[f]$.

Indeed, if $g_1\varphi_1(f)+\cdots+g_\rho\varphi_\rho(f)=0$ for some polynomials $\varphi_i(f)\in F(f)$, then evidently: if $\varphi_h(f)\neq 0$, for some h, then

$$\deg\,[g_i\varphi_i(f)]=\deg\,[g_j\varphi_j(f) \text{ for some } i\neq j.$$

But, since deg $[g_i\varphi_i(f)] \equiv$ deg $g_i \equiv m_i (\text{mod } n)$ and deg $[g_j\varphi_j(f)] \equiv m_j$ (mod n), and $m_i \not\equiv m_j$ (mod n), we are led to a contradiction. Consequently $\varphi_i(f)=0$ for all i.

It remains now to show that if $g \in C[f]$ then $g=g_1\varphi_1(f)+ \cdots +g_\rho\varphi_\rho(f)$ for some $\varphi_i(f) \in F[f]$. This is obtained by induction on deg g. It deg $g=0$, then it follows by Lemma 1 that $g=c \in F$, and hence $g=cg_1$. Now, let deg $g=k$. Since $k \in Z_f$, it follows that $k \equiv m_i (\text{mod } n)$ for some i. By the minimality of m_i, it follows that $k \geqq m_i$. Hence $k=m_i+nq$, which implies that deg $g=$deg g_if^q. It follows, therefore, by Lemma 1 that $g'=g-cg_if^q \in C[f]$ for some $c \in F$ and deg $g'<$deg g. Hence, by induction we obtain $g-cg_if^q=g_1\varphi_1(f)+ \cdots +g_\rho\varphi_\rho(f)$, and the proof is readily completed. This proves the validity of (1) of Theorem 1.

We turn now to the proof of (2). Let $g(D) \in C[f]$ be a polynomial whose residue class of deg $g(D)$ mod n generates the cyclic group $\overline{Z}_f$. One readily verifies that in this case, the set of all degrees of the polynomials of the form

$$H(g,f)=\varphi_0(f)+g\varphi_1(f)+ \cdots +g^{\rho-1}\varphi_{\rho-1}(f),\ \varphi_i(f) \in C[f]\ ,$$

contains all integers of Z_f with at most an exception of a finite number of integers. Hence, we may assume that this set contain all integers $m \in Z_f$ for which $m \geqq t$, for some fixed t. One then proves as in the preceding part that every polynomial $h(D) \in C[f]$ can be written in the form $h(D)=H_0(g,f)+h_0(D)$, where $h_0 \in C[f]$ and deg $h_0 \leqq t$. From Lemma 2 we know that the set of all polynomials h_0 is an F-space of dimension T, for some T. Let $f^\nu h=H_\nu(g,f)+h_\nu$, $\nu=0,1,\cdots,T$, and deg $h_\nu \leqq t$; thus the polynomials h_ν are F-dependent and, therefore, $\sum c_\nu h_\nu=0$, for $c_\nu \in F$ and not all $c_\nu=0$. This yield that $(\sum c_\nu f^\nu)h=\sum c_\nu H_\nu(g,f)$, which proves that for every $h \in C[f]$ there exists polynomials $H(g,f)$ and $F(f)$ with constant coefficients such that: $F(f)h=H(g,f)$.

Clearly the set of all polynomials $H(g,f)$ commute with each other, and the polynomials of $C[f]$ commute with the polynomial of $F[f]$; hence, if $F_i(f)h_i=H_i(g,f)$ for $h_i \in C[f]$ $i=1,2$, then

$$F_1(f)F_2(f)h_1h_2=(F_1h_1)(F_2h_2)=H_1H_2=H_2H_1=(F_2h_2)(F_1h_1)=F_2F_1h_2h_1\ .$$

Now $K[D]$ is a ring without zero divisors, hence $h_1h_2=h_2h_1$ and the proof of Theorem 1 is completed.

It was thus shown that $C[f]$ is an integral domain, let $C(f)$ denote the quotient field of $C[f]$. If $F(f)$ denotes the field of all rational functions in f over F, that is the quotient field of $F[f]$, then clearly $F(f) \subseteqq C(f)$. Actually, the preceding proof shows that the chosen polynomial g is algebraic of degree ρ over $C(f)$, since $F(f)g^\rho=H(g,f)$ and moreover $C(f)$ is an algebraic extension of $F(f)$ generated by g. Thus we have shown :

COROLLARY 1. *$C(f)$ is an algebraic extension of degree ρ of $F(f)$ and if the residue class of a polynomial $g \in C[f]$ generates the group of residue classes $\overline{Z}_F$, then g is of degree ρ over $F(f)$ and $C(f)=F(f)$ $[g]$.*

This clearly implies the following.

COROLLARY 2. *If $h \in C[f]$ then h is algebraic over $F(f)$ and its degree is a divisor of ρ, that is, there exists a polynomial $H(h,f)\equiv 0$ with constant coefficient and where degree in h is a divisor of ρ.*

This follows from the fact that $F(f)\subseteqq F(f)[h]\subseteqq F(f)[g]$.

REMARK. The fact that h is algebraic is well known, but here we obtained some additional information on its degree. In fact, one can prove by the previous methods that the degree of the minimal polynomial $H(h,f)$ in h is equal to the order of the subgroup of the additive group of all residue classes mod n generated by the degree of h.

Additional information on the degree ρ of $C(f)$ over $F(f)$ will be obtained in the following section.

3. **The field $C(f)$.** Let λ be a commutative indeterminate over the field K. We extend the derivation of K to a derivation of the field of all rational functions $K(\lambda)$ so that $F(\lambda)$ will be the field of constants of the extended derivation. Consider the ring $K(\lambda)[D]$ of all differential polynomials in D with coefficients in $K(\lambda)$.

LEMMA 3. *Every polynomial $g(D) \in K(\lambda)[D]$ can be expressed in the form $g=a(\lambda)^{-1}$ $G[\lambda,D]$, where $G[\lambda,D]=\sum g_\nu(\lambda)D^\nu$ is of the same degree as $g(D)$, and $g_\nu(\lambda)$, $a(\lambda)$ are relatively prime polynomials in λ. Similarly $g=G_1[\lambda,D]b(\lambda)^{-1}$ with similar restrictions for G_1 and $b(\lambda)$.*

The proof is evident.

Let $\overline{F}_\lambda$ be the algebraic closure of $F(\lambda)$, and let $\overline{K}_\lambda=K(\overline{F}_\lambda)$ be the field obtained by adjoining $\overline{F}_\lambda$ to K. that is, $\overline{K}_\lambda$ is the composition field of K and $\overline{F}_\lambda$ over $F(\lambda)$. One extends the deviation of K to $\overline{K}_\lambda$ so that $\overline{F}_\lambda$ is the new field of constants. These extended derivations yield the following sequence of rings of differential polynomials

$$K[D]\subset K(\lambda)[D]\subset \overline{K}_\lambda[D]\ .$$

If $f(D) \in K[D]$ then $f(D)-\lambda \in K(\lambda)[D]$ and first we show the following.

LEMMA 4. *$f(D)-\lambda$ is an irreducible polynomial in $K(\lambda)$ $[D]$.*

PROOF. Suppose $f(D)-\lambda=f_1(D)f_2(D)$ and $\deg f_i<\deg f$, $f_i(D) \in$

$K(\lambda)[D]$. In view of Lemma 3 we set $f_1(D)=g_1[D, \lambda]a^{-1}(\lambda)$ and $a^{-1}(\lambda)f_2(D)=g_2[D, \lambda]b^{-1}(\lambda)$. Thus $(f(D)-\lambda)b(\lambda)=g_1[D, \lambda]g_2[D, \lambda]$.

We consider now $g,[D, \lambda]$ as a polynomial in λ with coefficients in the ring $K[D]$, and we obtain by the remainder theorem[1]

$$g_1[D, \lambda]=(f(D)-\lambda)H[D, \lambda]+R[D]$$

where $R[D]\in K[D]$. Hence, it follows readily that

$$R[D]g_2[D, \lambda]=(f(D)-\lambda)G[D, \lambda] .$$

Let $g_2[D, \lambda]=\sum\lambda^\nu h_\nu(D)$. Then the fact that $f(D)-\lambda$ is a left divisor of $R[D]g_2[D, \lambda]$ implies by the remainder theorem that $\sum f^\nu Rh_\nu=0$. If $R\neq 0$, then we must have that deg $(f^\nu Rh_\nu)=$deg $(f^\mu Rh_\mu)\neq 0$ for some $\nu\neq\mu$. But suppose $\nu>\mu$ then we have

$$\deg h_\mu=\deg h_\nu+(\nu-\mu) \deg f\geqq\deg f.$$

On the other hand, clearly,

$$\deg h_\mu\leqq\deg g_2=\deg f_2<\deg f$$

which is impossible. Hence $R=0$, which means that $g_1[D, \lambda]=(f(D)-\lambda)H[D, \lambda]$. But this leads to a contradiction since deg $F_1<$deg f, and proof of the lemma is completed.

The polynomial $f(D)-\lambda$, when considered as a polynomial in $\overline{K}_\lambda[D]$, may be reducible, and indeed its factorization in the extended field $\overline{K}_\lambda$ is closely connected with the field $C(f)$. This we propose to show in what follows, and we begin with some preliminary lemmas.

Let $K[D, \lambda]$ be the ring of all polynomials in λ and D, and let $K[\lambda]$ be the polynomial ring in λ.

LEMMA 5. *Let* $0\neq p\in K(\lambda)$, $A, G, H\in K[\lambda, D]$ *such that* $AG=Hp$, *and* $A=f_0D^n+\cdots+f_n$ *where* $f_i\in K[\lambda]$ *and* $(f_0, p)=1$, *then* $G=G_1p$ *for some* $G_1\in K[\lambda, D]$.

PROOF. If the lemma is not valid, then let G be a polynomial of minimum degree in D which is not a left multiple of p and which satisfies the conditions of the lemma. Let[2] $G=D^mg_0+\cdots+g_m$. Since $AG=Hp$. It follows by comparison of the leading coefficients of both sides that $f_0g_0=h_0p$. Now $(f_0p)=1$, hence p divides g_0 and we have $g_0=qp$ for some polynomial $q\in K(\lambda)$. But then $G-(D^mq)p$ is of degree$<$

[1] See, e.g. A. A. Albert, Modern Higher Algebra, Chicago 1937 p. 25.

[2] Note that the polynominal $G(D)$ may be written with coefficients either on the right or on the left of the power of D.

deg G; it is not a left multiple of p, since G is not, but nevertheless $A(G-D^m qp)=(H-FD^m q)p$, which contradicts the minimality of G.

LEMMA 6. *Every polynomial $p(\lambda)\in K[\lambda]$ can be written as $p(\lambda)=c(\lambda)q(\lambda)$ where $c(\lambda)$ is a monic polynomial in λ and $c'(\lambda)=0$, and in the factorization of $q(\lambda)=aq_1^{\nu_1}\cdots q_k^{\nu_k}$, $a\in K$, into prime factors, the polynomials $q_i(\lambda)$ are relatively prime to their derivatives $q_i'(\lambda)$.*

Let $p(\lambda)=ap_1^{\nu_1}(\lambda)p_2^{\nu_2}(\lambda)\cdots p_n^{\nu_n}(\lambda)$ be the factorization of $p(\lambda)$ into prime factors. We may assume that each p_i is a monic polynomial, i.e. its leading coefficient is 1. For each p_i, the polynomial $p_i'(\lambda)$ is of lower degree in λ than p_i; hence, since p_i is prime, it follows that either $(p_i, p_i')=1$ or p_i divides p_i', which in the latter case must yield that $p_i'=0$. Thus, $c(\lambda)$ is the product of all p_i for which $p_i'=0$ and $q(\lambda)$ is the product of the rest.

LEMMA 7. *Let $p\in K[\lambda]$, $G,H\in K[D,\lambda]$ and $(p,p')=1$, then $pG=Hp$ implies that $G=G_0p$ for some $G_0\in K[D,\lambda]$.*

PROOF. If the lemma is not true then let G be the polynomial of minimum degree in D which do not satisfy our lemma.

Let $G=D^n g_0+\cdots+g_n$, $H=D^n h_0+\cdots+h_n$, g_i and $h_i\in K[\lambda]$, and $g_0h_0\neq 0$. Compare the coefficient of D^n of both sides of the equation $pG=Hp$. This yields $pg_0=h_0p$ which gives $g_0=h_0$. The coefficient of D^{n-1} yields

$$-np'g_0+pg_1=h_1p.$$

Hence, $-np'g_0=p(h_1-g_1)$. Since $(p,p')=1$ it follows that $g_0=kp$ for some $k\in K[\lambda]$. But then the polynomial $G-D^nkp$ is of lower degree then G; it is not a left multiple of p, but nevertheless $p(G-D^nkp)=(H-pD^nk)p$. This contradicts the minimality of G.

We can now turn to the main object of this section, and we recall the notion of the invariant ring of a differential polynomial. ([1 §5 p.260] and [3 §10 p.502]).

Let $h(D)$ be a polynomial in $K[D]$. The invariant ring $\mathscr{R}(h)$ of h is the ring of all classes $g(D)+h(D)K[D]$ which have a representative $g(D)$ satisfying $g(D)h(D)=h(D)g_1(D)$. It is known [1, Theorem 9] that $\mathscr{R}(h)$ is a finite dimensional algebra over the constant field, and if h is irreducible, then $\mathscr{R}(h)$ is a division ring.

We shall consider the invariant ring $\mathscr{R}(f(D)-\lambda)$ in the ring $K(\lambda)[D]$. Since it was shown in Lemma 4 that $f(D)-\lambda$ is irreducible, it follows that $\mathscr{R}(f-\lambda)$ is a division ring (e.g. [1, Theorem 10]). First we show the following.

THEOREM 2. *The field $C(f)$ is isomorphic with $\mathscr{R}(f-\lambda)$.*

PROOF. This elements of $\mathscr{R}(f-\lambda)$ are classes of the form $g(D)+(f-\lambda)K(\lambda)[D]$, $g\in K(\lambda)[D]$, and the first part of the proof is to show that we may choose a representative of this class of the form $q(D)c(\lambda)^{-1}$, where $q(D)\in C[f]$ and $c(\lambda)$ is a polynomial in λ with constant coefficients. The converse, that is: that every class which has a representative of the form $q(D)c(\lambda)^{-1}$ belongs to $\mathscr{R}(f-\lambda)$, follows easily since $qc^{-1}(f-\lambda)=q(f-\lambda)c^{-1}=(f-\lambda)\ qc^{-1}$.

So let $g(D)$ be a representative of a class in $\mathscr{R}(f-\lambda)$, then $g(D)(f-\lambda)=(f-\lambda)h(D)$. We set $g(D)=a^{-1}(\lambda)G[D,\lambda]$ and $h(D)=H[D,\lambda]b(\lambda)^{-1}$, in accordance with Lemma 3, we may assume and that a, b and monic polynomials. Then we have

$$G[D,\lambda](f-\lambda)b(\lambda)=a(\lambda)(f-\lambda)H[D,\lambda]\ . \tag{3.1}$$

Suppose $b(\lambda)\neq 1$. Let $b(\lambda)=p_1p_2\cdots p_n$ be the factorisation of b into prime factors, then we may assume that $(p_1', p_i)=1$. Since if $(p_1', p_i)\neq 1$ we have seen that $p_1'=0$, and we may deal with $p_ig(D)$, which also belongs to $\mathscr{R}(f-\lambda)$, instead of $g(D)$.

It follows by Lemma 3 that H was so chosen that it is not a left multiple of any prime factor of $b(\lambda)$ say p_1; furthermore, clearly $a(\lambda)(f-\lambda)=D^na(\lambda)a_0+\cdots$, where $f-\lambda=D^na_{0+}\cdots$ and $a_0\neq 0$. Hence, it follows by Lemma 5 that p_1 divides $a(\lambda)$. So let $a(\lambda)=p_1q_1$. Hence (3.1) yields

$$p_1q_1(f-\lambda)H[D,\lambda]=G_1p_1,\ \text{where } G_1=G(f-\lambda)p_2\cdots \mathrm{p}_r\ .$$

Since $(p_1, p_1')=1$, it follows by Lemma 7 that $q_1(f-\lambda)H=G_2p_1$. By similar reasons it follows from Lemma 5 that $q_1=p_1q_2$, and thus $p_1q_2(f-\lambda)H=G_2p_1$. Again Lemma 7 will yield that $q_2(f-\lambda)H=G_3p_1$. This cannot proceed indefinitely since the degrees of $a(\lambda), q_1, q_2, \cdots$ in λ are reduced in each step. Thus we are led to a contradiction, which leads us to the result that $b(\lambda)=1$. Thus (3.1) states that $G(f-\lambda)=a(f-\lambda)H$. The leading coefficient of $f-\lambda$ is an element of K, hence if we assume that $a\neq 1$, we must have by a result parallel to Lemma 5, that $G=aG_1$ which contradicts the way we have chosen G and a by Lemma 3.

We thus have shown that by multiplying $g(D)$ by polynomials $c(\lambda)$ (that is the product of the p_i for which $p_i'=0$) we obtained a representative $G[D,\lambda]$ which is a polynomial both in λ and D. If $G[D,\lambda]=\sum\lambda^\nu g_\nu(D)$, then the remainder theorem yields that

$$G[D,\lambda]=\sum f^\nu g_\nu+(f-\lambda)H_1[D,\lambda]\ .$$

Thus $q(D)=\sum f^\nu g_\nu$ is a representative of the same class $\mathrm{mod}(f-\lambda)$ as $G[D,\lambda]$.

Since $q\in\mathscr{R}(f-\lambda)$, we have $q(D)(f-\lambda)=(f-\lambda)Q(D)$. Let, in view of Lemma 3, $Q(D)=P[D,\lambda]d(\lambda)^{-1}$. Then $q(D)(f-\lambda)d(\lambda)=(f-\lambda)P[D,\lambda]$. We must have $d(\lambda)=1$. For if the degree of $d(\lambda)$ in λ is >1, then since the leading coefficient of $f-\lambda \in K$, it is relatively prime with $d(\lambda)$, and hence Lemma 5 implies that $P[D, \lambda]=P_1 d(\lambda)$ which contradicts Lemma 3. Thus $q(D)(f-\lambda)=(f-\lambda)Q(D)$ and $Q(D) \in K[D, \lambda]$. By comparing the coefficients of the powers of λ of both sides, one readily obtains that $Q=q$, and $qf=fq$, that is $q \in C[f]$. Consequently, we obtained that class of $c(\lambda)g(D)$ has a representative $q \in C[f]$. Hence

$$g(D)+(f-\lambda)K(\lambda)[D]=c(\lambda)^{-1}q(D)+(f-\lambda)K(\lambda)[D]$$

which proves our assertion.

We prove now Theorem 2. Clearly every element of $C(f)$ has the form $q(D)c(f)^{-1}$ where $q \in C[f]$ and $c(f) \in F[f]$, and we map $C(f)$ onto $\mathscr{R}(f-\lambda)$ by the correspondence

$$q(D)c(f)^{-1} \to q(D)c(\lambda)^{-1}+(f-\lambda)K(\lambda)[D] .$$

From the previous part of the proof it follows that this mapping is onto, and one readily verifies that this is an isomorphism. We shall show here only that it is a one-to-one correspondence; namely, that $q_1(D)c_1(f)^{-1} =q_2(D)c_2(f)^{-1}$ if and only if

$$q_1(D)c_1(\lambda)^{-1}+(f-\lambda)K(\lambda)[D]=q_2(D)c_2(\lambda)^{-1}+(f-\lambda)K(\lambda)[D] .$$

Indeed, the first hold if and only if $q_1(D)c_2(f)=q_2(D)c_1(f)$, and one readily verifies by the remainder theorem, in view of the fact that $c_i(f)$ commute with q_i, that the latter is eqivalent to the fact that $q_1(D)c_2(\lambda)-q_2(D)c_1(\lambda)=(f-\lambda)H[\lambda, D]$, and the rest is evident.
We now apply Theorem 2 to show the following.

THEOREM 3. *The polynomial $f(D)-\lambda$ is completely reducible* [3, p. 489] *in $\overline{K}_\lambda[D]$; if $g(D)$ is an irreducible polynomial which is right* (*or left*) *divisor of $f[D]-\lambda$ in $\overline{K}_\lambda[D]$ then deg $f=\mu$deg g and $\mu=\rho=(C(f):F(f))$.*

PROOF. Let θ be any automorphism of $\overline{F}_\lambda$ over $F(\lambda)$ This automorphism is readily extended to $\overline{K}_\lambda$ over $K(\lambda)$, and to $\overline{K}_\lambda[D]$ over $K(\lambda)[D]$. Since $f-\lambda=hg$, and $f-\lambda \in K(\lambda)[D]$, one readily verifies that $f-\lambda=(f-\lambda)^\theta=h^\theta g^\theta$. This means that $f-\lambda$ is a left common multiple of all g^θ, where θ ranges over all automorphisms of $\overline{K}_\lambda$ over $K(\lambda)$.

Let $G(D)$ be the least common left multiple of all g^θ whose leading coefficient is 1. Then, clearly $G^\varphi(D)$ will also be a least common multiple, whence one readily obtains that $G^\varphi=G$ for all automorphisms φ. This will yield that $G \in K(\lambda)[D]$. Now $f-\lambda$ is also a common left multiple, hence (Ore [4]) $f-\lambda=G_1G$, $G_1 \in \overline{K}_\lambda[D]$. Clearly, one obtains that also $G_1 \in K(\lambda)[D]$, but Lemma 4 states that $f-\lambda$ is irreducible. Consequently

$f-\lambda$ is the least left common multiple of all g^θ. From which one obtains (Ore [4]) that $f-\lambda$ is completely irreducible, and moreover, $f-\lambda=[g_1, \cdots, g_\mu]$, the least common multiple of all $g_i=g^{\theta_i}$ for some θ_i. In particular, this yields that all g_i have the same degree as g.

Thus (Ore [4]) μ deg $g=\deg(f-\lambda)=n$, or deg $g=\deg g_i=n/\mu$

To prove the second part of the theorem we need the following lemmas.

LEMMA 8. *Let $\overline{\mathscr{R}}(f-\lambda)$ be the invariant ring of $f-\lambda$ in $\overline{K}_\lambda[D]$, then $\overline{\mathscr{R}}(f-\lambda)=\mathscr{R}(f-\lambda)\otimes\overline{F}_\lambda$, where the tensor product is taken with respect to $F(\lambda)$.*

For let (c_α) be a $F(\lambda)$-base of $\overline{F}_\lambda$, then clearly, this set is also a $K(\lambda)$-base of K as well as a $K(\lambda)[D]$-base of $\overline{K}_\lambda[D]$. Let $g(D)\in\overline{K}_\lambda[D]$ belong to $\overline{\mathscr{R}}(f-\lambda)$, and let $g=\sum g_\alpha c_\alpha$ where $g_\alpha\in K(\lambda)[D]$. One readily observes that since $f-\lambda\in K(\lambda)[D]$, the relation $g(f-\lambda)=(f-\lambda)h$ implies that $h=\sum h_\alpha c_\alpha$, $h_\alpha\in K(\lambda)[D]$ and $g_\alpha(f-\lambda)=(f-\lambda)h_\alpha$ for all α. That is $g_\alpha\in\mathscr{R}(f-\lambda)$. Conversely, if $g_\alpha\in\mathscr{R}(f-\lambda)$ and only a finite number of g_α is different from 0, then clearly $\sum c_\alpha g_\alpha\in\overline{\mathscr{R}}(f-\lambda)$; from which one readily deduces the lemma.

LEMMA 9. *If $g(D)\in\overline{K}_\lambda[D]$ is irreducible, then $\overline{\mathscr{R}}(g)$ is the field of constants $\overline{F}_\lambda$ that is, if $hg=gh_1$, then $h=c+g\overline{K}_\lambda[D]$ for some constant $c\in\overline{F}_\lambda$.*

Indeed, if g is irreducible, then it follows by [1] Theorem 10 that $\overline{\mathscr{R}}(g)$ is a finite dimensional division algebra over the field of constant $\overline{F}_\lambda$. But $\overline{F}_\lambda$ is algebraically closed, hence the only division algebra over $\overline{F}_\lambda$ is $\overline{F}_\lambda$ itself. Thus $\overline{\mathscr{R}}(g)=\overline{F}_\lambda$.

We return now to the proof of Theorem 3.

It follows by [3, Theorem 19] and by the relation between the invariant ring of differential polynomials and differential linear transformaions ([3, §10 p.503] and [4]) that the invariant ring of a completely reducible polynomial is a direct sum of complete matrix rings over division algebra, and each division algebra is isomorphic to the invariant ring of one of the prime factors of the polynomial considered. In our case, since $\overline{\mathscr{R}}(f-\lambda)$ is commutative (by Lemma 8 and Theorem 1), and the invariant rings of irreducible polynomials are isomorphic with $\overline{F}_\lambda$, it follows that $\overline{\mathscr{R}}(f-\lambda)=F_1\oplus\cdots\oplus F_\mu$, where each F_i is a field isomorphic with $\overline{F}_\lambda$ (compare with [3, Theorem 19]), and μ is the number of prime divisors given in the first part of Theorem 3.

On the other hand $\mathscr{R}(f-\lambda)$ is an algebraic extension of $F(\lambda)$; and from Theorem 2 and Theorem 1 it follows that $(\mathscr{R}(f-\lambda)\colon F(\lambda))=\rho$.

Hence it is well known that $\mathscr{R}(f-\lambda)\otimes\bar{F}_\lambda$ is a direct sum of ρ fields isomorphic with the algebraic closed field $\bar{F}_\lambda$. Consequently Lemma 8 implies that $\overline{\mathscr{R}}(f-\lambda)$ has the same decomposition. Comparing the two results, we obtain that $\rho=\mu$.
From Theorem 3 we can conclude the following known result.

COROLLARY 3. *If $f(D)$ is a polynomial with constant coefficients then $C[f]$ is the ring $F[D]$ of all polynomials with constant coefficients.*

PROOF. The factorization of $f(D)-\lambda$ in $\bar{K}_\lambda[D]$ is readily obtained. Indeed, let $\mu_1, \cdots \mu_n$ be the roots of $f(x)-\lambda$ in $\bar{F}_\lambda$ where x is a commutative variable, then it is easily seen that the factorization of $f(D)-\lambda$ in $\bar{K}_\lambda[D]$ is $f(D)-\lambda=\Pi_{i=1}^n(D-\mu_i)$. It follows therefore by Theorem 3 that $(C(f)\colon F(f))=n$. But the field of all rational functions in D, that is $F(D)$, is of degree n over $F(f)$ and clearly $F(D)\subset C(f)$. Hence $C(f)=F(D)$. The rest is readily obtained.

We can also determine the dimension $(C(f)\colon F(f)=\rho$ by the methods of [1, §5]. In [1] we have introduced the notion of a resultant of two differential polynomials which was denoted by $f(D)\times g(D)$, and the notion of the nullity of a polynomial $f(D)$ in a field K. We recall here that the nullity of f in K was the number of independent solutions of the differential equation $f(D)z=O$ in K.

From Theorem 2 and [1, Theorem 2] we now obtain the following

THEOREM 4. *The dimension $\rho=(C(f)\colon F(f))=(\mathscr{R}(f-\lambda)\colon F(\lambda))$ is equal to the nullity of the polynomial $(f(D)-\lambda)\times(f^*(D)-\lambda)$ in $K(\lambda)$ where $f^*(D)$ is the adjoint polynomial of $f(D)$.*

REFERENCES

1. S. A. Amitsur, *Differential polynomials and division algebras*, Ann. of Math. **59** (1954), 245-278.

2. H. Flanders, *Commutative linear differential operators*, Report No. 1, University of California, Berkely 1955.

3. N. Jacobson, *Pseudo-linear transformations*, Ann. of Math., **38** (1937), 484-507.

4. O. Ore, *Theory of non commutative polynomials*, Ann. of Math. **34** (1933), 480-508.

THE HEBREW UNIVERSITY

UNIFORM CONTINUITY OF CONTINUOUS FUNCTIONS OF METRIC SPACES

MASAHIKO ATSUJI

In this paper we intend to find equivalent conditions under which continuous functions of a metric space are always uniformly continuous. Isiwata has attempted to prove a theorem in a recently published paper [3] by a method that has a close relation with ours. Unfortunately he does not accomplish his purpose, so we shall give a correct theorem (Theorem 3) in the last part of this paper and, for this purpose, give a condition for the existence of a uniformly continuous unbounded function in a metric space (Theorem 2).

In this paper the space S, unless otherwise specified, is the metric space with a distance function $d(x, y)$, and, for a positive number α, the α-sphere about a subset A $\{x\,;\, d(A, x)<\alpha\}$ is denoted by $S(A, \alpha)$; the function is the real valued continuous mapping.

DEFINITION 1. Let us consider a family of neighborhoods U_n of x_n such that $\{x_n\}$ is a sequence of distinct points and $U_m \cap U_n=\phi$ (=empty) for $m\neq n$. Let $f_n(x)$ be a function such that $f_n(x_n)=n$ and $f_n(x)=0$ for $x \notin U_n$. Then a *mapping constructed from the family* is a mapping $f(x)$ defined by $f(x)=f_n(x)$ for x belonging to some U_n and $f(x)=0$ for the other x.

LEMMA. *Consider a family of neighborhoods U_n of x_n satisfying the following conditions*:

(1) *$\{x_n\}$, which consists of distinct points, has no accumulation point,*

(2) *$\overline{U}_m \cap \overline{U}_n=\phi$, $m\neq n$ ($\overline{U}$ a closure of U), and $U_n \subset S(x_n, 1/n)$,*

(3) *there is a sequence of points y_n such that distances of x_n and y_n converge to* 0 *and y_n does not belong to any U_m; then the mapping constructed from the family is continuous and not uniformly continuous. When $\{x_n\}$ is a sequence containing infinitely many distinct points and has no accumulation point, there is a family of neighborhoods of x_n satisfying* (2); *if $\{x_n\}$ further contains infinitely many distinct accumulation points, then the family besides satisfies* (3).

Proof. The continuity of the mapping constructed from the family follows from $\overline{\cup U_{n_i}} = \cup \overline{U}_{n_i}$ for any subsequence $\{n_i\}$ of indices; the mapping is not uniformly continuous by (3). Suppose $\{x_n\}$ consists of distinct accumulation points and has no accumulation point, then, by an inductive process, we have neighborhood V_n of x_n such that $V_n \subset S(x_n, 1/n)$ and $\overline{V}_m \cap \overline{V}_n=\phi$, and have y_n and a neighborhood U_n of x_n

Received May 24, 1957. In revised form September 6, 1957.

such that $U_n \not\ni y_n \in V_n$, $U_n \subset V_n$.

DEFINITION 2. Let x be isolated in a metric space, then we write $I(x)$ for a supremum of positive numbers α such that $S(x, \alpha)$ consists of x alone.

THEOREM 1. *The following conditions on a metric space S are equivalent*

(1) *If $\{x_n\}$ is a sequence of points without accumulation point, then all but finitely many members of x_n are isolated and* inf $I(x_n)$ *for the isolated points is positive.*

(2) *If a subset A of S has no accumulation point then all but finitely many points of A are isolated and* inf $I(x)$ *for all the isolated points of A is positive.*

(3) *The set A of all accumulation points in S is compact and* inf $I(x_n)$ *is positive for any sequence $\{x_n\}$ in S-A which has no accumulation point* (Isiwata [2], Theorem 2).

(4) $\bar{A} \cap \bar{B} = \phi$ *implies* $S(A, \alpha) \cap S(B, \alpha) = \phi$ *for some* α (Nagata [4], Lemma 1).

(5) $\bigcap_{n=1}^{\infty} \bar{A}_n = \phi$ *implies* $\bigcap_{n=1}^{\infty} S(A_n, \alpha) = \phi$ *for some* α.

(6) *For any function $f(x)$, there is a positive integer n such that every point of $A = \{x ; |f(x)| \geqq n\}$ is isolated and* $\inf_{x \in A}$ $I(x)$ *is positive.*

(7) *All functions of S are uniformly continuous.*

(8) *All continuous mappings of S into an arbitrary uniform space S' are uniformly continuous.*

Proof. Since the equivalence of (1) and (3) is simple, we shall show (1)→(8)→(7)→(6)→(5)→(4)→(2)→(1).

(1)→(8): If a continuous mapping $f(x)$ of S is not uniformly continuous, there is an "entourage" V (in the sense of Bourbaki) of S' such that $d(x_n, y_n) < 1/n$ and $(f(x_n), f(y_n)) \notin V$ for any positive integer n and for some x_n and y_n. $\{x_n\}$ contains infinitely many distinct points. If $\{x_n\}$ has an accumulation point x, there are subsequences $\{x_{n_i}\}$ and $\{y_{n_i}\}$ of $\{x_n\}$ and $\{y_n\}$ converging to x, and, since $f(x)$ is continuous, $(f(x), f(x_{n_i})) \in W$ and $(f(x), f(y_{n_i})) \in W$ for W satisfying $W \cdot W \subset V$ (we may assume $W^{-1} = W$) and for all sufficiently large i. Hence we have $(f(x_{n_i}), f(y_{n_i})) \in V$, which is excluded. Consequently $\{x_n\}$ hes no accmulation point and inf $I(x_n) = r > 0$ for all sufficiently large n, which contradicts the first inequality of f for n satisfying $r > 1/n$.

(8)→(7) is obvious.

(7)→(6): If, for some function $f(x)$ and every n, there is an accumulation point x_n such that $|f(x_n)| \geqq n$, $\{x_n\}$ contains infinitely many distinct elements and has no accumulation point, then, by the Lemma, we have a function which is not uniformly continuous. Suppose that every point of $A=\{x;\ |f(x)|\geqq n\}$ is isolated and inf $I(x)=0$. Then there is a sequence $\{x_n\}$ in A such that inf $I_n=0$, $I_n=I(x_n)$. $\{x_n\}$ has no accumulation point, and we may assume $I_n<1/n$. If distances of distinct points of $\{x_n\}$ are greater than a positive number e, then, for all n satisfying $e>4I_n$, x_n and y_n $(\neq x_n,\ \in S(x_n,\ 2I_n))$ satisfy the conditions of the Lemma. In the other case, there are arbitrarily large m and n satisfying $d(x_m,\ x_n)<e$ for any positive number e, and we have, by an inductive process, a subsequence $\{y_i\}$ of $\{x_n\}$ satisfying $d(y_{2i-1},\ y_{2i})<1/i$. Then y_{2i-1} and y_{2i} satisfy the conditions of the Lemma.

(6)→(5): Let $\bigcap_n S(A_n,\ 1/m)\neq\phi$ for every m in spite of $\bigcap_n \bar{A}_n=\phi$. We have a point x_1 contained in $\bigcap_n S(A_n,\ 1)$ and a point y_1 distinct from x_1 satisfying $d(x_1,\ y_1)<1$. Suppose $B_i=\{x_1,\cdots,\ x_i\}$ consists of distinct points such that $x_j\in \bigcap_n S(A_n,\ 1/j)$, x_j and y_j are distinct and $d(x_j,\ y_j)<1/j$, $j=1,\cdots,\ i$. Since, for any point x, $\bigcap_n S(A_n,\ 1/m)$ does not contain x for a sufficiently large m, $\bigcap_n S(A_n,\ 1/(i+1))$ contains a point x_{i+1} being not contained in B_i, and some A_n contains y_{i+1} distinct from x_{i+1} satisfying $d(x_{i+1},\ y_{i+1})<1/(i+1)$. Thus we have a sequence $\{x_n\}$ of distinct points and $\{y_n\}$ such that $x_m\ \in \bigcap_n S(A_n,\ 1/m)$, x_n and y_n are distinct, and $d(x_n,\ y_n)<1/n$. $\{x_n\}$ has no accumulation point because of $\bigcap \bar{A}_n=\phi$. The function obtained from the Lemma does not satisfy the condition (6) whether all but finitely many members of x_n are isolated or not.

(5)→(4) is obvious.

(4)→(2): Suppose A has infinitely many accumulation points x_n, $n=1,\ 2,\cdots$. Since $B=\{x_n\}$ has no accumulation point, there is a sequence $C=\{y_n\}$ having no accumulation point such that $d(x_n,\ y_n)<1/n$, $B\cap C=\phi$. $\bar{B}\cap\bar{C}=B\cap C=\phi$, and $S(B,\ \alpha)\cap S(C,\ \alpha)=\phi$ for no α. If every point of A is isolated and inf $I(x)=0$, we have a sequence $\{x_n\}$ such that lim $I(x_n)=0$, and have a sequence $\{y_n\}$ with the same properties as the above.

(2)→(1) is obvious.

Recently Isiwata has stated a theorem ([3], Theorem 4) which is related to our Theorem 1. However the first step in his proof is wrong. We shall give a correct form of the theorem in Theorem 3. Let us first give a counterexample for the statement "In a connected metric space which is not totally bounded, there exists a sequence $\{x_n\}$ and a uniformly continuous function f such that $f(x_n)=n$".

EXAMPLE. Denoting the points of the plane by polar-coordinate, we consider the following subsets of the plane:

$$A_m=\{(r,\ \theta);\ 0\leqq r\leqq 1,\ \theta=\pi/m\},$$
$$S=\bigcup_{m=1}^{\infty} A_m\ .$$

We define the distance of the points of S by

$$\begin{aligned} d((r,\ \theta),\ (r',\ \theta'))&=|r-r'| \qquad \text{as } \theta=\theta' \text{ or } rr'=0\ ,\\ &=r+r' \qquad \text{as } \theta\neq\theta'\ ,\end{aligned}$$

then S is obviously a connected metric space which is not totally bounded. When $f(x)$, $x\in S$, is a uniformly continuous function of S, there is a positive integer n such that $d(x,\ y)<1/n$ implies $|f(x)-f(y)|<1$. If x is contained in A_m, there are points $y_0=0=$pole, $y_1, \cdots, y_r=x$, $r\leqq n+1$, of A_m such that $d(y_{i-1},\ y_i)<1/n$, $i=1, \cdots, r$.

$$|f(0)-f(x)|\leqq|f(0)-f(y_1)|+\cdots+|f(y_{r-1})-f(x)|\leqq n+1\ ;$$

namely $f(x)$ is bounded.

DEFINITION 3. Let e be a positive number, then the finite sequence of points $x_0, x_1, \cdots, x_m$ satisfying $d(x_{i-1},\ x_i)<e$, $i=1, \cdots, m$, is said to be an *e-chain with length m*. If, for any positive number e, there are finitely many points $p_1, \cdots, p_i$ and a positive integer m such that any point of the space can be bound with some p_j, $1\leqq j\leqq i$, by an e-chain with length m, then the space is said to be *finitely chainable*.

THEOREM 2. *A metric space S admits a uniformly continuous unbounded function if and only if S is not finitely chainable.*

Proof. Verification of "only if" part is analogous to that stated in the above example, hence is passed over. Let S be not finitely chainable, then there is a positive number e such that, for any finitely many points and a positive integer n, there is a point which cannot be bound with any one of points selected above by an e-chain with length n. We denote by A_0^n the set of all points which can be bound with a fixed x_0 by an e-chain with length n.

(1) When $A_0^n\neq A_0^{n+1}$ for every n, we put

$$f(x)=(n-1)e+d(x,\ A_0^{n-1})$$

for x belonging to A_0^n and not to A_0^{n-1}, and $f(x)=0$ for $x\notin A_0=\bigcup_n A_0^n$ ($f(x)=d(x_0,\ x)$ for $x\in A_0^1$). Since $S(A_0,\ e)=A_0$, $f(x)$ is uniformly continuous on S if it is so on A_0. Let $A_0^n\ni x\notin A_0^{n-1}$ and $d(x,\ y)<e'<e$, then $A_0^{n+1}\ni y\notin A_0^{n-2}$. (i) When y is in A_0^{n-1}, then

$$f(y)=(n-2)e+d(y,\ A_0^{n-2})$$

and $d(x,\ A_0^{n-1})<e'$, $d(y,\ A_0^{n-2})<e$, hence $f(y)\leqq f(x)$. If $d(y,\ A_0^{n-2})<e-e'$, then $d(y,\ y')<e-e'$ for some y' of A_0^{n-2} and $d(x,\ y')\leqq d(x,\ y)+d(y,\ y')<e$, so that x is in A_0^{n-1}, which is excluded. Therefore $d(y,\ A_0^{n-2})\geqq e-e'$ and

$$\begin{aligned}|f(x)-f(y)| &=f(x)-f(y)=e+d(x,\ A_0^{n-1})-d(y,\ A_0^{n-2})\\ &<e+e'-(e-e')=2e'\ .\end{aligned}$$

(ii) When y is in A_0^n and not in A_0^{n-1}, then

$$f(y)=(n-1)e+d(y,\ A_0^{n-1}),$$

and we have

$$|f(x)-f(y)|=|d(x,\ A_0^{n-1})-d(y,\ A_0^{n-1})|\leqq d(x,\ y)<e'$$

(cf. the proof of Prop. 3 of §2, [1]). (iii) The remaining case for y is similar to (i). Consequently $f(x)$ is uniformly continuous on A_0.

(2) When $A_0^n=A_0^{n+1}$ for some n, then $A_0^m=A_0^n$ for every $m\geqq n$, and, in the similar way to (1), $A_1=\cup A_1^n$ is obtained from a point of $S-A_0$. If we can make an unbounded function which is uniformly continuous on A_1, our proof will be complete.

(3) When we cannot, for every $m(0\leqq m\leqq n)$, construct a desired function on A_m obtained in the same way as (2), $A_0, \cdots, A_n$ cannot cover the space, because the space is not finitely chainable; namely we have a sequence of infinitely many subsets $A_0,\ A_1, \cdots$ when our proof is not complete in the similar way to (2). Then we put $f(x)=n$ for x of A_n and $f(x)=0$ for x which is not in any A_n. Then, since $S(A_m,\ e)\cap A_n=\phi$ for any $m\neq n$ and $S(\cup A_n,\ e)=\cup A_n$, $f(x)$ is uniformly continuous.

THEOREM 3. *If S is a connected metric space which is not finitely chainable, then the set of all uniformly continuous functions of S does not form a ring.*

Proof. The following verification is essentially due to Isiwata [3]. There is, by Theorem 2, a uniformly continuous unbounded function $f(x)$ of the space, and we have a sequence $A=\{x_n\ ;\ n=1,\ 2, \cdots\}$ such that $f(x_n)=a_n$, $a_{n+1}-a_n\geqq 1$, $a_1\geqq 1$; A has no accumulation point. For some positive number α, $d(x,\ y)<\alpha$ imlies $|f(x)-f(y)|<1/3$, and so $S(x_m,\ \alpha)\cap S(x_n,\ \alpha)=\phi$ for $m\neq n$. We put

$$h(x)=1-d(A,\ x)/\alpha \qquad \text{and} \qquad G=\bigcup_n S(x_n,\ \alpha)$$

and

$$f'(x) = \begin{cases} h(x) & \text{for } x \in G\,, \\ 0 & \text{for } x \notin G\,. \end{cases}$$

$h(x)$ is uniformly continuous on the space, because $d(A, x)$ is so (cf. Prop. 3 of §2, [1]). $h(x) > 0$ and $h(y) \leqq 0$ for x of G and y of $S-G$ respectively, so we have

$$|h(x)-h(y)| = h(x)-h(y) \geqq h(x) = |f'(x)-f'(y)|\,.$$

Hence $f'(x)$ is uniformly continuous on the space. $g(x)=f(x)f'(x)$ is not uniformly continuous. In fact, if it is uniformly continuous, $d(x, y)<\beta$ implies

$$(*) \qquad |g(x)-g(y)|<1 \quad \text{and} \quad |f(x)-f(y)|<1$$

for some β ($\leqq \alpha$). We select a positive integer n such that a_n is greater than $1+4\alpha/\beta$, and take a point y such that $\beta/2 \leqq d(x_n, y)<\beta$ (it is possible to take such a point because of the connectedness of the space). Then, by $(*)$, we have $|a_n - f(y)|<1$, $f(y)>a_n-1\geqq 0$, and

$$\begin{aligned} |g(x_n)-g(y)| &= |a_n-(1-d(A, y)/\alpha)f(y)| = |a_n - f(y) + d(x_n, y)f(y)/\alpha| \\ &\geqq |d(x_n, y)f(y)/\alpha| - |a_n - f(y)| > d(x_n, y)f(y)/\alpha - 1 \\ &> \beta(a_n-1)/2\alpha - 1 > \beta(1+4\alpha/\beta-1)/2\alpha - 1 = 1\,, \end{aligned}$$

which contradicts $(*)$.

REFERENCES

1. N. Bourbaki, *Topologie générale*, Chap. IX, Paris Hermann, 1948.
2. T. Isiwata, *On uniform continuity of* $C(X)$ (Japanese), Sugaku Kenkyu Roku of Tokyo Kyoiku Daigaku **2** (1955), 36-45.
3. ———, *Structures of a uniform space* X *and* $C(X)$, Sci. Rep. Tokyo Kyoiku Daigaku **5** (1956), 174-184.
4. J. Nagata, *On the uniform topology of bicompactifications*, J. Inst. Polytech. Osaka City Univ. **1** (1959), 28-38.

KANAZAWA JAPAN

A NUMERICAL CONDITION FOR MODULARITY OF A LATTICE

S. P. AVANN

1. Introduction. In this note a simple numerical condition (θ) is presented which is necessary for modularity of a finite lattice L. Though not sufficient (θ) appears to be a condition imposing a strong tendency toward modularity.

NOTATION. Covering, proper inclusion, and inclusion will be denoted by $>$, $\supset$, $\supseteqq$ respectively. $N[S]$ will denote the order of the set S. The unit and zero elements will be denoted by u and z respectively.

DEFINITION 1. A finite lattice L is upper semi-modular [1: p. 100] if and only if

(ξ') $\qquad$ a and $b > a \cap b$ imply $a \cup b > a$ and b.

L is lower semi-modular if and only if

(ξ'') $\qquad$ $a \cup b > a$ and b imply a and $b > a \cap b$.

DEFINITION 2. In a finite lattice let $C(a) = \{x \in L | x < x \cup a > a\}$ and $D(a) = \{x \in L | x > x \cap a < a\}$.

2. Tests for modularity An immediate consequence of Definitions 1 and 2 is the following theorem.

THEOREM 1. *In a finite lattice L condition* (ξ') *is equivalent to* $D(a) \subseteqq C(a)$ *for all* $a \in L$ *and both imply* $N[D(a)] \leqq N[C(a)]$. *Dually,* (ξ'') *is equivalent to* $D(a) \supseteqq C(a)$ *for all* $a \in L$ *and both imply* $N[D(a)] \geqq N[C(a)]$. *Moreover, modularity,* (ξ') *and* (ξ''), *is equivalent to* $D(a) = C(a)$ *for all* $a \in L$ *and both imply the condition* (θ):

(θ) $\qquad$ $N[D(a)] = N[C(a)]$ for all $a \in L$.

The contrapositive of the last statement of Theorem 1 serves as a useful test for non-modularity :

THEOREM 2. *If there exists* $a \in L$ *for which* $N[D(a)] \neq N[U(a)]$, *then* L *is non-modular.*

When either (ξ') or (ξ'') is known to hold in L, the verification of the condition (θ) is a test often easiest to apply. It merely requires counting coverings.

Received December 5, 1957.

THEOREM 3. *In a finite lattice* L (ξ') *and* (θ) *together imply modularity and dually* (ξ'') *and* (θ) *together likewise imply modularity.*

Proof. From Theorem 1 condition (ξ') implies $D(a) \subseteqq C(a)$, and along with (θ) we obtain $D(a)=C(a)$ for all $a \in L$. Hence L is modular.

Condition (θ) appears to be a very strong condition toward modularity. It would be useful to know a much weaker but easily applicable condition than (ξ') or (ξ'') to serve along with (θ) as a set of necessary and suffiicient conditions for modularity.

3. Near-modular lattices.

DEFINITION 3. A finite lattice L is *near-modular*, henceforth abbreviated NM, if and only if (θ) is valid and the Jordan-Dedekind chain condition is satisfied.

REMARK. It is conceivable that the JD chain condition is implied by (θ), though no proof was readily found. The imposition of the JD condition seems desirable, since it is satisfied in all finite semi-modular and modular lattices. Hence each element of a NM lattice L will possess a uniquely determined rank.

THEOREM 4. *In a* NM *lattice* L *we have* $D(x)=C(x)$ *whenever* x *is a point* (*atom*) *or dual point. Condition* (ξ') *is satisfied by all pairs of points and* (ξ'') *by all pairs of dual points.*

Proof. Let p be an arbitrary point of L and q any element in $C(p)$. By consideration of rank, q is also a point which is distinct from p. Hence $q > q \cap p = z$, the zero element of L, and $q \in D(p)$. Thus $C(p) \subseteqq D(p)$. Equality of orders yields $C(p)=D(p)$. Any pair of points p and q cover their meet z so that $q \in D(p)$. Hence $q \in C(p)$ so that (ξ') is valid for p and q. The remainder of Theorem 4 follows by duality.

COROLLARY. *All* NM *lattices of rank less than* 4 *are modular.*

THEOREM 5. *There exist* NM *lattices of rank* 4 *that are nonmodular.*

The smaller example L_4 of the two examples found was constructed from the finite projective geometry PG(2, 2) as follows. If the points of PG(2, 2) are designated by 1, 2, 3, 4, 5, 6, 7, the lines, considered as sets of points, can be taken as 356, 467, 571, 612, 723, 134, 245, and $u=$ 1234567. For L_4 take $u=1234567$, and the 7 dual points as 1247, 2351, 3462, 4573, 5614, 6725, 7136, namely the complementary sets to the dual points of PG(2, 2). The remaining elements of L_4 are generated by taking all point set meets of its dual points. The lines of L_4 are the $21=\binom{7}{2}$ pairs of points 12, 13, $\cdots$, 67; the points are 1, 2, 3, 4, 5, 6, 7; and z is the null set.

The automorphism group of L_4 is easily seen from the manner of its construction to be the same as that for PG(2, 2), of order 168. L_4 possesses dual automorphisms, one of which carries the planes in the order indicated above into the points 7, 6, 5, 4, 3, 2, 1 respectively. Moreover, L_4 is a complemented point (atomic) lattice. It possesses no non-trivial homomorphic images, since all prime quotients are projective.

When the above procedure of construction of L_4 from PG(2, 2) was applied to PG(2, 3), PG(2, 2^2), PG(3, 2) and other PG(k, p^n), the lattices obtained were all found to violate (θ). Some of them violated also the JD chain condition.

The structure of L_4 suggested a method of obtaining additional examples of non-modular NM lattices as follows. Let L_t consist of z; n points $p_1, p_2, \cdots, p_n$ where $n=1+\binom{t}{2}$; $\binom{n}{2}$ lines consisting of all pairs of points: $p_1 p_2, \cdots, p_{n-1} p_n$; n planes of which the first is the set of t points $p_{i_1} p_{i_2} \cdots p_{i_r} \cdots p_{i_t}$ where $i_r=1+\binom{r}{2}$, $(r=1, 2, \cdots, t)$ and the remaining planes are obtained from the first by repeated applications of the cyclic permutation $(123 \cdots n)$ to the subscripts; and $u=p_1 p_2 \cdots p_n$. This procedure yields for $t=1, 2, 3$ the Boolean algebras B^1, B^2, B^4 respectively. For $t=4$, the lattice L_4 described above is obtained. For $t=5$ a second example, L_5, of a non-modular NM lattice of length 4 is obtained. For $t \geqq 6$ one fails to obtain a lattice. It can readily be shown by consideration of certain congruences that for $t \geqq 6$ there always exist at least two pairs of planes, as described in the construction, which intersect in three or more points and other pairs of planes that intersect in less than two points. When two planes have three points p, q, r in common, the lines pq, pr, and qr have each of the planes as upper bound, but fail to have a least upper bound.

4. Extensions, In this section, methods of construction of other NM lattices from given ones are presented.

THEOREM 6. *The direct product of* NM *lattices is also an* NM *lattice.*

Proof. Let $L=L_1 \times L_2 \times \cdots \times L_n$ where the components are NM lattices. Represent each $a \in L$ in the usual way as the n-tuple $(a_1, \cdots, a_n)$ with $a \in L_i$ $(i=1, \cdots, n)$, so that $a \cup b$ and $a \cap b$ are obtained by taking joins and meets respectively component-wise. Let $C(a)$ and $D(a)$ be the functions of Definition 2. Define $H(a)$ as the set of elements covering $a \in L$ and $K(a)$ as the set of elements covered by a. Let

$C(a_i)$, $D(a_i)$, $H(a_i)$, $K(a_i)$ be the corresponding sets with respect to $a_i \in L_i$. Now $a>b$ in L if and only if $a_j>b_j$ for some j and $a_i=b_i$ for $i\neq j$. It follows readily that

$$(1) \qquad N[C(a)]=\sum_{i=1}^{n} N[C(a_i)]+\sum_{i\neq j} N[H(a_i)]\cdot N[K(a_j)]$$

$$(2) \qquad N[D(a)]=\sum_{i=1}^{n} N[D(a_i)]+\sum_{i\neq j} N[K(a_j)]\cdot N[H(a_i)]$$

The last summations of the two equations are equal. By hypothesis (θ_i): $N[C(a_i)]=N[D(a_i)]$ for $i=1, \cdots, n$. Hence (θ) is valid in L.

NOTATION. $L=a/b$ indicates a lattice with unit element a and zero element b. The set sum and product of lattices L_1 and L_2, considered only as sets of elements, will be denoted by L_1+L_2 and $L_1\cdot L_2$ respectively.

LEMMA 1. *If $L_1=a_1/z$ possesses a dual ideal a_1/b_1 isomorphic to an ideal a_2/b_2 of a second lattice $L_2=u/b_2$, then by identifying as x each pair of elements $x_1 \in a_1/b_1$ and $x_2 \in a_2/b_2$ that correspond under the isomorphism, a lattice $L=u/z$ can be constructed having L_1 as an ideal and L_2 as a dual ideal such that $L=L_1+L_2$ and $a/b\cong L_0=L_1\cdot L_2$.*

The elements of L are taken as the identified elements $x \in a/b$ and the remaining elements of L_1 and L_2. Join $\cup$ and meet $\cap$ in L are defined in terms of $\cup_1$, $\cap_1$ in L_1 and $\cup_2$, $\cap_2$ in L_2 according to the cases:

$$\left.\begin{array}{l} r\cup s=r\cup_i s \\ r\cap s=r\cap_i s \end{array}\right\} \quad r, s\in L_i, \quad (i=1, 2)$$

$$\left.\begin{array}{l} r\cup s=s\cup r=(r\cup_1 b)\cup_2 s \\ r\cap s=s\cap r=r\cap_1(a\cap_2 s) \end{array}\right\} \quad r\in L_1, \; s\in L_2 \; .$$

The verification of the lattice postulate is routine and is omitted. This method of extension was first employed systematically by M. Hall and R. P. Dilworth [2; Lemma 4.1].

In Lemmas 2, 3, 4 and Theorem 7 following, let $L=u/z$, ideal $L_1=a/z$, dual ideal $L_2=u/b$, and quotient sublattice $L_0=a/b$ be related as in Lemma 1: $L=L_1+L_2$ and $L_0=L_1\cdot L_2$. We note that $L-L_2$, L_0, $L-L_1$ is a partitioning of L into disjoint subsets.

LEMMA 2. *If $s>r$ in L, then s and r are both in L_1 or both in L_2.*

Proof. Obviously impossible is the case $s\notin L_2$, $r\notin L_1$. Assume that $s\notin L_1$, $r\notin L_2$; that is, $b\subseteq s\not\subseteq a$, $a\supseteq r\not\supseteq b$. Then $s=s\cup r=s\cup(b\cup r)\supset b\cup r$, otherwise $s\subseteq b\cup r\subseteq a$, a contradiction. Furthermore $b\cup r\supset r$, otherwise $b\subseteq r$, a contradiction. Thus the covering $s>r$ is violated and the only possible cases are as stated.

LEMMA 3. *If $D_i(x)$, $C_i(x)$ are the functions of Definition 2 relative to L_i $(i=0, 1, 2)$, then*

$$\begin{aligned}&(3)\ D_0(x)=D_1(x)\cdot D_2(x)\\&(3')\ C_0(x)=C_1(x)\cdot C_2(x)\\&(4)\ D(x)=D_1(x)+D_2(x)\\&(4')\ C(x)=C_1(x)+C_2(x)\end{aligned}$$

Proof. (3) holds since $r\in D_0(x)$: $r>r\cap x<x$ with r, $r\cap x$, x all in $L_0=L_1\cdot L_2$ if and only if $r>r\cap x<x$ with $r, r\cap x, x$ all in both L_1 and L_2; that is, $r\in D_1(x)$, $D_2(x)$, $D_1(x)\cdot D_2(x)$. Next, $r\in D_1(x)$: $r>r\cap x<x$ with $r, r\cap x, x$ all in L_1 implies $r\in D(x)$: $r>r\cap x<x$ with $r, r\cap x, x$ all in L. Thus $D_1(x)\subseteqq D(x)$. Similarly $D_2(x)\subseteqq D(x)$ so that $D_1(x)+D_2(x)\subseteqq D(x)$. For demonstration of the less trivial reverse inclusion let $r\in D(x)$: $r>r\cap x<x$ in L. If $r\cap x\notin L_2$, both r and x are in L_1, along with $r\cap x_1$, by Lemma 2. For this case $r\in D_1(x)$. If $r\cap x\in L_2$, then certainly also are both r and x. Whence $r\in D_2(x)$. We thus obtain $D(x)\subseteqq D_1(x)+D_2(x)$ and therefore (4). Dually (3′) and (4′) are valid.

LEMMA 4. *The following statements of orders are valid*:

$$(5)\qquad N[D(x)]+N[D_0(x)]=N[D_1(x)]+N[D_2(x)]$$

$$(5')\qquad N[C(x)]+N[C_0(x)]=N[C_1(x)]+N[C_2(x)]$$

This follows immediately from Lemma 3.

THEOREM 7. *If any three of L, L_1, L_2, L_0, related as in* Lemmas 1–4, *are near-modular, then all are near-modular.*

Proof. Equality of three pairs of corresponding members of (5) and (5′) implies equality of the remaining pair.

REMARKS. It is no doubt possible to construct non-modular NM lattices in other ways; for example, by piecing together several NM lattices to become the ideals of L and several others to become the dual ideals of L. Such a construction would require perhaps a more precise knowledge of the basic structure of a NM lattice.

A sublattice, and even a quotient sublattice, of a NM lattice is not necessarily near-modular. It is an open question whether or not the homomorphic image of a NM lattice is near-modular.

REFERENCES

1. G. Birkhoff, *Lattice theory*, rev. ed., Amer. Math. Soc. Colloquium Publications, New York, 1948.
2. M. Hall and R. P. Dilworth, *The imbedding theorem for modular lattices*, Ann. of Math., **45** (1944), 450–456.

UNIVERSITY OF WASHINGTON

A MEAN VALUE THEOREM FOR QUADRATIC FIELDS

RAYMOND G. AYOUB

1. Introduction. Let K be an algebraic extension of the rationals of degree k, $F(n)$ denote the number of ideals whose norm is the rational integer n, $H(x)=\sum_{n\leqq x}F(n)$. Let $\zeta(s, K)$ denote the Dedekind zeta function for the field K, that is,

$$\zeta(s, K)=\sum_{\mathfrak{A}}\frac{1}{N(\mathfrak{A})s}=\sum_{n=1}^{\infty}\frac{F(n)}{n^s}$$

and α the residue of $\zeta(s, K)$ at its simple pole at $s=1$.

It has long been known [8] that

$$H(x)=\alpha x+\Delta_k(x)$$

where

$$\Delta_k(x) = 0(x^{1-1/k})$$

and Landau [3] proved that

$$\Delta_k(x) = 0(x^{1-2/(k+1)})$$

The precise nature of the error term $\Delta_k(x)$ seems rather intractable and seems to be intimately related to the behavior of the function $\zeta(s, K)$ in the critical strip. Of considerable interest is the particular case when K is the Gaussian field $R(i)$, for in that case $\Delta_k(x)$ is the error term in the classical problem of the number of lattice points in a circle.

Using some results of class field theory, Suetuna [4] has obtained an improvement of Landau's result in the case when the field is normal and has abelian Galois group and $k\geqq 4$. For, when the field is abelian, the theorems of Weber-Takagi tell us that $\zeta(s, K)$ is the product of k Dirichlet L-functions belonging to primitive characters. Applying his approximate functional equation for the Dirichlet L-functions, and using refined estimates for these in the critical strip, Suetuna then obtains the desired result.

In the light of more recent techniques for dealing with the Riemann zeta function, further improvements are possible. The devices for handing the zeta function are used for the L-functions and the class

Received February 12, 1957.

field theorems again applied. We omit details.

It is our object here to study the problem of a mean value for $\Delta_k(x)$. We are able to obtain a precise result but only for quadratic fields. Some known results follow as corollaries when the quadratic field is specified: for example when the field is $R(i)$.

We use as our tools a result of Suetuna [4] and a technique devised by Titchmarsh [5], [6] for the corresponding results for the closely allied problem of $\sum_{n\leq x} d_k(n)$ where $d_k(n)$ is the number of solutions of $n=n_1n_2\cdots n_k$. We follow closely Titchmarsh's method.

2. Notations and statement of Main theorem. Let $k=2$, $\Delta_2(x)=\Delta(x)$, $s=\sigma+it$. Following Hardy [6], we define the mean value of $\Delta(x)$ as the least number β such that

$$\frac{1}{x}\int_0^x \Delta^2(t)dt=0(x^{2\beta+\varepsilon})$$

It is our object to prove the following.

MAIN THEOREM. $\beta=\dfrac{1}{4}$.

We first relate β to the convergence of an integral.

THEOREM 2.1 *Let γ be the lower bound of positive numbers σ for which*

$$I=\int_{-\infty}^{\infty}\frac{|\zeta(\sigma+it,K)|^2}{|\sigma+it|^2}dt \tag{1}$$

converges. Then $\beta=\gamma$ and if $\sigma>\beta$, then

$$2\pi\int_0^\infty \Delta^2(x)x^{-2\sigma-1}dx=I \tag{2}$$

Proof. Using the classical formula for the sum of the coefficients of a Dirichlet series, we have,

$$H(x)=\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty}\frac{\zeta(s,K)}{s}x^s ps \qquad (c>1)$$

$$=\frac{1}{2\pi i}\lim_{T\to\infty}\int_{c-iT}^{c+iT}\frac{\zeta(s,K)}{s}x^s ds$$

We move the line of integration to $\sigma=\delta$, where $0<\delta<1$. Using Cauchy's theorem and taking account of the residue at $s=1$, we get, if δ is chosen appropriately close to 1,

$$\varDelta(x)=\frac{1}{2\pi i}\lim_{T\to\infty}\int_{\delta-iT}^{\delta+iT}\frac{\zeta(s,\ K)}{s}x^s ds$$

The bound on δ follows from the implied calculation but we do not need it since we now prove the validity of (3) for the range $\gamma<\delta<1$. To do this, we note that by some general theorems of analysis [2], and taking account of (1), $\frac{\zeta(s,\ K)}{s}$ tends uniformly to 0 as $t\to\pm\infty$. With this established, we integrate around the rectangle defined by $\delta'-iT$, $\delta-iT$, $\delta+iT$, $\delta'+iT$ with $\gamma<\delta'<\delta<1$, let $T\to\infty$, and deduce the desired result.

With Titchmarsh, we now use the theory of Mellin transforms. The Parseval theorem for the Mellin integral gives [7],

$$(4)\qquad \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|\zeta(\delta+it),\ k|^2}{|\delta+it|^2}=\int_0^{\infty}\varDelta^2\left(\frac{1}{x}\right)x^{2\delta-1}dx$$

$$=\int_0^{\infty}\varDelta^2(x)x^{-2\delta-1}dx$$

as long as $\gamma<\delta<1$. This implies that $\beta\leqq\gamma$: by (4)

$$\int_{\frac{x}{2}}^{x}\varDelta^2(x)x^{-2\delta-1}dx<C(\delta)=C_1$$

that is,

$$\int_{\frac{x}{2}}^{x}\varDelta^2(x)dx<C_2x^{2\delta+1}$$

Replacing x by $x/2$, $x/4$, $x/8,\cdots$ and adding, we deduce

$$\int_1^{x}\varDelta^2(x)dx<C_3x^{2\delta+1}$$

whence $\beta\leqq\delta$, that is, $\beta\leqq\delta$.

To prove the reverse inequality, we have by Plancherel's form of the inverse Mellin transform [7],

$$(5)\qquad \frac{\zeta(s,\ K)}{s}=\int_0^{\infty}\varDelta(x)x^{-s-1}dx$$

where the right hand integral exists in the mean square sense for $\gamma<\delta<1$. Actually the right hand side is uniformly convergent for the range $\beta'<\sigma<\beta''$ where $\beta<\beta'<\beta''<1$. For, using the Schwartz inequality,

$$\int_{\frac{x}{2}}^{x}|\Delta(x)|x^{-\sigma-1}dx \leqq \left(\int_{\frac{x}{2}}^{x}\Delta^2(x)dx\right)^{1/2}\left(\int_{\frac{x}{2}}^{x}x^{-2\sigma-2}\right)^{1/2} = O(x^{2\beta+1+\varepsilon})^{1/2}(x^{-2\sigma-1})^{1/2} = O(x^{\beta-\sigma+\varepsilon})$$

putting $x=2, 4, 8, \cdots$ and adding, we get

$$\int_{1}^{\infty}|\Delta(x)|x^{-\sigma-1}dx<\infty$$

By a similar argument,

$$\int_{0}^{\infty}\Delta^2(x)x^{-2\delta-1}$$

converges for $\beta<\delta<1$. Now

$$\int_{0}^{\infty}\Delta(x)x^{-s-1}dx$$

is an analytic function for $\beta<\sigma<1$, and hence

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{|\zeta(\delta+it, K)|^2}{|\delta+it|^2}dt<\infty$$

for $\beta<\delta<1$, whence $\gamma\leqq\delta$ that is, $\gamma\leqq\beta$ and the theorem is proved.

So far we have not made use of the condition $k=2$. Indeed Titchmarsh's method applies in quite a general setting. We require the condition however in the proof of the main theorem.

LEMMA (Suetuna). *If* $\sigma>\frac{1}{2}$, *then*

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}|\zeta(s, K)|^2dt=\sum_{n=1}^{\infty}\frac{F^2(n)}{n^{2\sigma}}$$

Proof of the Main Theorem. We first prove that $\beta\geqq\frac{1}{4}$. By the above lemma, we have for $\frac{1}{2}<\sigma<1$,

$$\int_{\frac{T}{2}}^{T}|\zeta(s, K)|^2<CT$$

Therefore for $0<\sigma<\frac{1}{2}$, $T>1$ and using Hecke's functional equation for $\zeta(s, K)$, (see for example. Landau [3]), we get

$$\int_{-\infty}^{\infty}\frac{|\zeta(\sigma+it, K)|^2}{|\sigma+it|^2}dt>\int_{\frac{T}{2}}^{T}\frac{|\zeta(\sigma+it, K)|^2}{|\sigma+it|^2}dt>\frac{C_1}{T^2}\int_{\frac{T}{2}}^{T}|\zeta(\sigma+it, K)|^2dt$$

$$>C_2T^{-4\sigma}\int_{\frac{T}{2}}^{T}|\zeta(1-\sigma-it,\ K)|^2dt>C_3T^{1-4\sigma}$$

The right hand side tends to infinity if $\sigma<\frac{1}{4}$, whence $\beta\geqq\frac{1}{4}$.

Again by the above lemma, for $\frac{1}{2}<\sigma<1$

$$\int_1^T|\zeta(\sigma+it,\ K)|^2dt=O(T)$$

Using the functional equation, we get for $0<\sigma<\frac{1}{2}$

$$\int_1^T|\zeta(\sigma+it,\ K)|^2dt=O(T^{2-4\sigma})\int_1^T|\zeta(1-\sigma-it,\ K)|^2dt$$
$$=O(T^{3-4\sigma})$$

Hence

$$\int_{\frac{T}{2}}^{T}\frac{|\zeta(\sigma+it,\ K)|^2}{|\sigma+it|^2}dt=O(T^{-\eta}) \qquad \eta>0$$

provided that $\sigma>\frac{1}{4}+\varepsilon$. It then follows by a simple argument that

$$\int_T^\infty\frac{|\zeta(\sigma+it,\ K)|^2}{|\sigma+it|^2}dt\longrightarrow 0$$

for $\sigma>\frac{1}{4}+\varepsilon$, and therefore that $\gamma\leqq\frac{1}{4}$ that is, that $\beta\leqq\frac{1}{4}$.

References

1. G. H. Hardy, *The average order of the arithmetical functions* $P(x)$ *and* $\Delta(x)$, Proc. London Math. Soc. Series 2, **15** (1916), 192–213.
2. ———, A. Ingham, and G. Polya, *Theorem concerning mean values of analytic functions*, Proc. Royal Soc A, **113** (1926) 542–569.
3. Landau, *Einfuhrung in die elementare und analyische Theorie der algebrischen Zahlen und der Ideale*, (Göttingen 1917).
4. Z. Suetuna, *On the product of L-functiom*, Japanese J. Math. **2**, (1925), 1937
5. E. C. Titchmarsh, *On divisor problems*, Quarterly J. Math. (Oxford) **9.** (1938) 216–220.
6. ———, *The theory of the Riemann zeta function*, (Oxford, 1951).
7. ———, *Introduction to the thoery of Fourier integrals*, (Oxford, 1931).
8. H. Weber, *Lehrbuch der Algebra* **2**, (Braunschweig, 1908).

PENNSYLVANIA STATE UNIVERSITY

SUBALGEBRAS OF FUNCTIONS ON A RIEMANN SURFACE

ERRETT BISHOP

1. Introduction and preliminaries. A set of problems, which has attracted much attention in recent years, treats the question of what functions can be approximated in some given topology by a given function algebra on a given set of points. The classical Weierstrass approximation theorem, and its generalization, the Stone-Weierstrass approximation theorem, are well-known results of this type which have proved very useful in analysis. Very important work has more recently been done by Lavrentiev, Keldys, and Mergelyan, and their results generalize the classical theorem of Runge (see Saks and Zygmund [4] for Runge's theorem).

The theorem of Mergelyan states that every continuous function on a compact set C of the complex plane, which is analytic at interior points, can be uniformly approximated on C by polynomials, if C does not separate the plane, i.e., if the complement of C is connected. We prove a theorem which generalizes this result in two respects: the plane is replaced by an arbitrary separable Riemann surface (without boundary, but not necessarily connected), and the algebra of all polynomials is replaced by what we call a total subalgebra of the algebra R of all functions which are everywhere analytic on the Riemann surface. The subalgebra R' is called total if it contains the constant functions and if the set $\{p \mid p \in C$ and there exists $q \neq p$ in C, with $f(p) = f(q)$ for all f in $R'\} \cup \{p \mid p \in C$ and no function in R' is one-to-one in any neighborhood of $p\}$, called the singular set of C relative to R', is finite for all compact sets C. (It can be shown that when R' is not total, but contains the constant functions, one can identify points on the surface to obtain a new surface on which R' is total.)

Our methods are highly measure-theoretic, and we make constant use of the fact that any bounded linear functional Λ on the space $\Omega(C)$ of all continuous complex-valued functions on a compact set C of our surface can be represented as a Borel measure μ on C. This means that $\int f\, d\mu = \Lambda(f)$ for all f in $\Omega(C)$. We shall somewhat loosely identify Λ and μ, so that by the value of μ on f we shall mean $\int f\, d\mu$, and by saying that μ is orthogonal to f we shall mean $\int f\, d\mu = 0$. For a compact

Received October 25, 1957. The author wishes to express his thanks to the Office of Naval Research for the partial support of the work through ONR 041 157.

set C, $\Phi(C)$ will denote the set of all continuous functions on C which are analytic at interior points. We are actually interested in bounded linear functionals Λ on $\Phi(C)$, but by means of the Hahn-Banach theorem every such Λ can be extended to $\Omega(C)$, and therefore can be represented by a measure μ on C. If R' is a subalgebra of R, then $R'(C)$ will denote the set of all continuous functions on C which are uniform limits on C of functions in R'. Obviously $R'(C)\subset\Phi(C)$, and the problem, roughly speaking, is to determine by how much $R'(C)$ differs from $\Phi(C)$. We do this via an investigation of those measures μ on C which are orthogonal to $R'(C)$, that is, we see how much these measures miss being orthogonal to $\Phi(C)$.

We proceed to some definitions, which are necessary to the statement of the theorem to be proved. If C is a compact set, and if R' is a subalgebra of R, then $\mathscr{S}(C, R')$ will denote the set $\{p|$ for each f in R', there exists q in C with $|f(q)|\geqq|f(p)|\}$. The condition $\mathscr{S}(C, R')=C$ is the natural extension of Mergelyan's condition-that C not separate the plane-to the more general situation considered here. The bounded linear functional Λ on $\Phi(C)$ will be called an R'-local differential operator on $\Phi(C)$, of order not exceeding N, if (1) Λ is orthogonal to $R'(C)$, and (2) there exists a finite subset S of the singular set of C with respect to R', such that $f(p)=f(q)$ for all f in R' and all p and q in S, and such that $\Lambda(g)=0$ whenever g is a function in $\Phi(C)$ which vanishes at all points of S and vanishes to order at least N at all points of S which are interior to C. The bounded linear functional Λ on $\Phi(C)$ will be called a R'-homogeneous differential operator on $\Phi(C)$, of order not exceeding N, if it is a finite sum of R'-local differential operators on $\Phi(C)$, of orders not exceeding N. The result to be proved reads: If R' is a total subalgebra of R, if C is a compact set with $\mathscr{S}(C, R')=C$, and if Λ is a bounded linear functional on $\Phi(C)$ which is orthogonal to $R'(C)$, then Λ is a R-homogeneous differential operator on $\Phi(C)$, of order not exceeding N, where N depends only on R' and C. Since it will be easy to show that the only R-homogeneous differential operator on $\Phi(C)$ is 0, this will have the corollary that $R(C)=\Phi(C)$ whenever $\mathscr{S}(C, R)=C$. In general, we shall only be able to conclude that the vector space $R'(C)$ (over the complex field) is of finite codimension in the vector space $\Phi(C)$. It will be possible to describe $R'(C)$ exactly in case C has no interior points.

Of the six preparatory lemmas to be proved, Lemmas 4 and 6 are of some interest in themselves. Lemma 6, in particular, seems to be a very useful tool in the theory of approximation by polynomials, and the author will give other applications of this lemma elsewhere.

We develop more notation for later use. If C is compact, and if the function f in R generates the subalgebra R', then $\mathscr{S}(C, f)$ will mean $\mathscr{S}(C, R')$, so that p will be in the complement $\mathscr{S}'(C, f)$ of $\mathscr{S}(C, f)$ if and only if $f(p)$ is in the unbounded component of the complement of $f(C)$. If C_1 has compact closure and if C_2 is compact, we say that f in R is schlicht on C_1 relative to C_2 if there exists a neighborhood U of the closure of C_1 such that no point in U is identified with any other point of $U \cup C_2$ by f. If C_2 is void, we simply say that f is schlicht on C_1, and if also C_1 is a point $\{q\}$, we say that f is schlicht at q (or that f is one-to-one in some neighborhood of q). Since a separable Riemann surface is metrizable, we assume the existence of a metric ρ on the surface. If S_1 and S_2 are compact and $S_1 \supset S_2$, we define

$$\rho(S_1, S_2) = \sup\{\inf\{\rho(p, q) | q \in S_2\} | p \in S_1\} .$$

An arc is a homeomorphic image of $[0, 1]$, and an open arc is an arc minus its endpoints. A closed disc is a homeomorphic image of $\{z | |z| \leqq 1\}$, and a disc is a closed disc minus its boundary.

2. Preparatory lemmas.

LEMMA 1. *Let F be a compact set of the complex plane with connected complement, and let 0 be in the boundary of F. Let N be a positive integer. Then the function z can be uniformly approximated on F by polynomials which vanish at 0 to order at least N.*

Proof. If there is a sequence $\{h_n\}$ of polynomials whose derivatives vanish at 0 and which converge uniformly to z on F, then the sequence $\{h_n - h_n(0)\}$ of polynomials vanishes at 0 to order at least 2 and converges uniformly to z on F. Now assume that z cannot be uniformly approximated on F by polynomials which vanish at 0 to order at least 2. Then z cannot be uniformly approximated on F by polynomials whose derivatives vanish at 0. If we let $\Omega(F)$ be the Banach space of all continuous complex-valued functions on F, this means that z is not in the subspace of $\Omega(F)$ generated by the polynomials whose derivatives vanish at 0. Thus there will exist a bounded linear functional Λ on $\Omega(F)$ which will vanish on all polynomials whose derivatives vanish at 0, but with $\Lambda(z) = a \neq 0$. It follows that $\Lambda(h) = ah'(0)$ for all polynomials h. We may assume that the bound of Λ is 1 and that $a > 0$. Let U be a simply connected open set containing F, the distance η of whose boundary to 0 is less than $a/16$. Let ϕ be the conformal map of $|z| < 1$ onto U, with $\phi(0) = 0$ and $\phi'(0) > 0$. Since the boundary of U contains points at a distance η from 0, it is known (see [1], page 75) that $\phi'(0) \leqq 4\eta$. If we let Ψ be the map of U onto $|z| < 1$ which is inverse to ϕ, then $\Psi'(0) = [\phi'(0)]^{-1} \geqq (4\eta)^{-1}$. If we define f on U to be the analytic function $f = (2 - \Psi)^{-1}$, we have $|f(z)| \leqq 1$ for z in F, so that $|\Lambda(f)| \leqq 1$. Also

$$f'(0)=\frac{\Psi'(0)}{|2-\Psi(0)|^2}=\frac{1}{4}\;|\Psi'(0)|\geqq\frac{1}{16\eta}>\frac{1}{a}\;.$$

Since f is analytic on U, there will exist a sequence $\{g_n\}$ of polynomials converging uniformly to f on some neighborhood of F. Therefore $g_n'(0)$ will converge to $f'(0)$. Thus,

$$1<a|f'(0)|=a\;\lim_n|g_n'(0)|=\lim_n|\Lambda(g_n)|=|\Lambda(f)|\leqq 1\;.$$

This contradiction shows that z is the subspace T_2 of $\Omega(F)$ generated by polynomials h which vanish at 0 to order at least 2. Thus $z^2=z\cdot z$ is in the subspace T_3 of $\Omega(F)$ generated by polynomials $z\cdot h$ which vanish at 0 to order at least 3. Thus all polynomials which vanish to order at least 2 at 0 are in T_3, so that $T_2=T_3$. Thus $z\in T_3$. By a continuation of this process, it can be shown that z is in the subspace T_N consisting of the closure in $\Omega(F)$ of all polynomials which vanish at 0 to order at least N. This completes the proof.

LEMMA 2. *Let R' be a total subalgebra of R, let C be a compact set with $\mathscr{S}(C, R')=C$, and let S be the singular set of C relative to R' (so that S is finite). Then there exists a closed C-neighborhood C' of S and a positive integer N, such that any function in $\Phi(C')$ which vanishes at all points of S which are interior to C, to order at least N, and which vanishes at all points of S, is in $R'(C')$, and such that C' is the union of disjoint closed sets $\{C_p\}$, each containing exactly one point p of S.*

Proof. Let p and q be any two distinct points of S. Let f be any non-zero function in R' which vanishes on S but which does not vanish identically in a neighborhood of any point of S. Such a function can be found because R' is total. Let n be the exact order to which f vanishes at p. Then it is easy to find a closed disc U containing p in its interior, and an analytic function ϕ which is defined and one-to-one on some neighborhood of U, which maps U onto $\{z\,|\,|z|\leqq c\}$ for some $c>0$, which vanishes at p, and for which $[\phi(r)]^n=f(r)$ for all r in U. Since f vanishes on S, we can also find a closed neighborhood H of S containing U such that $f(H)=f(U)$. Since R' is total, we can in addition take U and H to be so small that S will be the singular set of H relative to R'. Let q_0 be any point in the component of the interior of H which contains q, except q itself, with $f(q_0)\neq 0$, and let p_0 be any point of U with $f(p_0)=f(q_0)$. Let ζ be a primitive nth root of unity, and let π be the map of U onto itself defined by $\phi(\pi r)=\zeta\phi(r)$. Obviously $f(r)=f(\pi r)$ for all r in U. Since S alone is the singular set of H relative to R', there exists g in R' taking distinct values at the points q_0, p_0, and the first n-1 images, $p_1=\pi(p_0)$, $p_2=\pi(p_1)$, $\cdots$, $p_{n-1}=\pi(p_{n-2})$ of p_0 under

π. Note that $f(p_j)=f(p_0)$ for $1\leqq j\leqq n-1$.

Now g^k, for $0\leqq k\leqq n$, can be expanded on U as a uniformly convergent power series in powers of ϕ, which implies that g^k can be written on U in the form $g^k=\sum_{i=0}^{n-1} f_{ki}\phi^i$, where f_{ki} is the sum on U of a power series in powers of $f=\phi^n$ which converges uniformly on U. The series defining f_{ki} will actually converge uniformly on H, because $f(H)=f(U)$. Thus we may extend the definition of f_{ki} to H, where it will be a function in $R'(H)$ which identifies all pairs of points in H that are identified by f. Therefore $f_{ki}(p_j)=f_{ki}(p_0)$ for $0\leqq j\leqq n-1$, $0\leqq k\leqq n$, and $0\leqq i\leqq n-1$, and consequently

$$[g(p_j)]^k=\sum_{i=0}^{n-1} f_{ki}(p_0)[\phi(p_j)]^i .$$

This implies that the product of the matrices

$$(f_{ki}(p_0)),\ 0\leqq k\leqq n-1,\ 0\leqq i\leqq n-1 ,$$

and $([\phi(p_j)]^i)$, $0\leqq i\leqq n-1$, $0\leqq j\leqq n-1$, is the non-singular Vandermonde matrix $([g(p_j)]^k)$, $0\leqq k\leqq n-1$, $0\leqq j\leqq n-1$. Therefore, the function M in $R'(H)$ defined by $M=\det(f_{ki})$, $0\leqq k\leqq n-1$, $0\leqq i\leqq n-1$, does not vanish at p_0. Now for each r in U the linear system

$$-[g(r)]^k x_0+\sum_{i=0}^{n-1} f_{ki}(r)x_{i+1}=0 , \qquad 0\leqq k\leqq n,$$

has the non-trivial solution $x_0=1$, $x_1=1$, $x_2=\phi(r)$, $\cdots$, $x_n=[\phi(r)]^{n-1}$. Thus the function h in $R'(H)$ defined by

$$h=\begin{vmatrix} -1 & f_{00} & \cdots & f_{0\ n-1} \\ -g & f_{10} & & f_{1\ n-1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ -g^n & f_{n0} & \cdots & f_{n\ n-1} \end{vmatrix}$$

vanishes identically on U. On the other hand, we have just seen that the coefficient $(-1)^{n+1} M$ of g^n in this determinant does not vanish at p_0. We may therefore write h in the form $\sum_{k=0}^{n} h_k g^k$, where h_k is a function in $R'(H)$ which identifies any pair of points which is identified by f, and where $h_n(p_0)\neq 0$. By substitution of $p_0, \cdots, p_{n-1}$ into this expression for h, we obtain

$$0=h(p_j)=\sum_{k=0}^{n} h_k(p_j)[g(p_j)]^k=\sum_{k=0}^{n} h_k(p_0)[g(p_j)]^k ,$$

since f identifies p_j and p_0. Thus $g(p_0), \cdots, g(p_{n-1})$ are n distinct roots of the nth degree equation $\sum_{k=0}^{n} h_k(p^0)x^k=0$, so that $g(q_0)$, which is distinct from these roots, does not satisfy the equation. Therefore

$$h(q_0)=\sum_{k=0}^{n} h_k(q_0)[g(q_0)]^k=\sum_{k=0}^{n} h_k(p_0)[g(q_0)]^k\neq 0 .$$

Thus h does not vanish identically in any neighborhood of q, or it would vanish in the component of the interior of H containing q, and therefore it would vanish at q_0.

Thus we see that for distinct points p and q in S there exists a closed neighborhood H of S and a function h in $R'(H)$ such that h vanishes identically in a neighborhood of p but does not vanish identically in any neighborhood of q. By multiplying together such functions, we see that for all points p in S there exists a closed neighborhood K of S, and a function f in $R'(K)$ which does not vanish in any neighborhood of p, but vanishes in some neighborhood of every other point of S. With this new function f, whose multiplicity at p we call n, choose U, ϕ, π, and H in the same way as they were chosen for the old function f. In addition, we may assume that H is so small that f vanishes on $H-U$. We now extend the definition of ϕ to all of H by defining ϕ to vanish on $H-U$. Let p_0 be any point in U distinct from p, and define $p_1, p_2, \cdots, p_{n-1}$ as above. Choose any function g in R' which takes distinct values at $p_0, p_1, \cdots, p_{n-1}$. Let the functions $f_{.k}$ be defined as before, so that $g^k=\sum_{i=0}^{n-1} f_{ki}\,\phi^i$ on U, for $0\leqq k\leqq n-1$. (We shall not need the equation for g^n.) We have seen that the determinant M defined above is in $R'(H)$ and does not vanish identically on U. Applying Cramer's rule to the set of equations for the g^k, we can solve them for ϕ, obtaining M in the denominator and some function of $R'(U)$ in the numerator. It follows that the restriction of the function $\phi\cdot M$ to U is in $R'(U)$. Now M, being a polynomial in the f_{ki}, is equal on U to the sum of a power series in powers of f which converges uniformly on U. Let the first non-zero term of this power series be $a_0 f^t$. Then f^t/M will be a uniformly convergent power series on some neighborhood $U'=\{q|q\in U, |\phi(q)|\leqq c'<c\}$ of p in powers of f. Since f vanishes on $H-U$, the series will converge uniformly on $H'=U'\cup(H-U)$ to a function f_0 in $R'(H')$ which equals f^t/M on U' and vanishes on $H'-U'$. Since $\phi\cdot M$ is in $R'(U')$, it follows that the function $(\phi\cdot M)\cdot f_0 =\phi\cdot f^t=\phi^{nt+1}$ is in $R'(H')$. Since $f=\phi^n$ is also in $R'(H')$, and since the exponents n and $nt+1$ are relatively prime, the function ϕ^i will be in $R'(H')$ if i is sufficiently large, say if $i\geqq N$. Therefore, any function in $\Phi(H')$ which vanishes on $H'-U'$ and which vanishes to order at least N at p will be in $R'(H')$.

Now let p be a boundary point of C, and we shall show that the last statement continues to hold with $N=1$ if H' is replaced by $H' \cap C$. From $\mathscr{S}(C, R')=C$, it follows that none of the components of $U'-C$ lies interior to U', since $\mathscr{S}(C, R')$ would contain such a component. Therefore every component of $U'-C$ contains boundary points of U'. Since ϕ is a homeomorphism on U', it follows that the complement of $\phi(U' \cap C)=\phi(H' \cap C)=F$ is connected. Since $\phi(p)=0$, the number 0 is in the boundary of F, By Lemma 1, there exists a sequence $\{h_n\}$ of polynomials which vanish at 0 to order at least N and which converge uniformly to z on F. The function $h_n \circ \phi$, for each n, is therefore in $R(H')$, by the last statement of the preceding paragraph, and $h_n \circ \phi \to \phi$ uniformly on $H' \cap C$ as $n \to \infty$. Therefore $\phi \in R'(H' \cap C)$. By Mergelyan's theorem, any function which is continuous on $\phi(H' \cap C)$ and analytic at interior points can be uniformly approximated by polynomials h. Therefore, any function in $\Phi(H' \cap C)$ which vanishes at p and vanishes on $(H' \cap C)-U'$ can be uniformly approximated by functions of the form $h \circ \phi$, and so belongs to $R'(H' \cap C)$.

It follows from what we have just proved that there exist disjoint closed C-neighborhoods $\{C_p\}$, one for each point p in S, whose union we denote by C', and a positive integer N, such that any function f in $\Phi(C')$ which vanishes on S, which vanishes on $C'-C_p$ for some p, and which vanishes to order at least N at p if p is interior to C, will be in $R'(C')$. Since any function in $\Phi(C')$ which satisfies the conditions of the lemma can be written as a sum of such functions f, the conclusion of the lemma follows.

LEMMA 3. *Let C be compact, and let R' be a total subalgebra of R with $\mathscr{S}(C, R')=C$. Let Λ be a bounded linear functional on $\Phi(C)$, which is orthogonal to $R'(C)$ and which can be represented as a measure on an arbitrary C-neighborhood of the singular set S of C relative to R'. Then Λ is a R'-homogeneous differential operator on $\Phi(C)$, whose order does not exceed an integer N depending on R' and C but not on Λ.*

Proof. Partition S into equivalence classes $S_1, S_2, \cdots, S_n$, by defining $p \equiv q$ to mean $g(p)=g(q)$ for all g in R'. Then there exist functions $f_1, f_2, \cdots, f_n$ in R' such that $f_i(p)=0$ for p in $S-S_i$ and $f_i(p)=1$ for p in S. Thus, by Runge's theorem, there exist disjoint closed C-neighborhoods $U_1, U_2, \cdots, U_n$, of $S_1, S_2, \cdots, S_n$ respectively, such that, for $1 \leqq i \leqq n$, there exists a sequence of functions in R' which converges uniformly on $U=U_1 \cup U_2 \cup \cdots \cup U_n$ to a function g_i which has the value 1 on U_i and the value 0 on $U-U_i$. Since Λ can be realized as a measure on U, it can be extended to be a bounded linear functional Λ' on $\Phi(U)$. Obviously Λ' will vanish on $R'(U)$. Therefore, if we define the functionals $\Lambda_1, \cdots, \Lambda_n$ by $\Lambda_i(f)=\Lambda'(fg_i)$, for all f in $\Phi(C)$, we obtain bounded linear functionals on $\Phi(C)$ which vanish on $R'(C)$ and have sum Λ. For each i, $1 \leqq i \leqq n$, let V_i be any closed C-neighborhood of S_i which is a subset of U_i. By hypothesis, there will exist a measure μ on $V=V_1 \cup \cdots \cup V_n$ which represents Λ. For each i, $1 \leqq i \leqq n$, let $\{g_{ik}\}_{k=1}^{\infty}$ be a sequence of functions in R' converging uniformly on U to g_i. Then for each f in $\Phi(C)$ we have

$$\Lambda_i(f)=\Lambda'(fg_i)=\lim_{k\to\infty} \Lambda'(fg_{ik})=\lim_{k\to\infty} \Lambda(fg_{ik})$$
$$=\lim_{k\to\infty}\int fg_{ik}\, d\mu=\int fg_i\, d\mu=\int_{U_i} f\, d\mu=\int_{V_i} f\, d\mu .$$

Therefore Λ_i is represented by the restriction of μ to V_i, from which it follows that Λ_i can be represented as a measure on an arbitrary C-neighborhood of S_i. To finish the proof, it is only necessary to show that Λ_i is a R'-local differential operator on $\Phi(C)$ of order not exceeding some positive integer N depending only on R' and C. Let the closed C-neighborhood C' of S and the positive integer N have the properties stated in Lemma 2. If we write $C_i=\cup\{C_p|p \in S_i\}$, then C_i is a closed C-neighborhood of S_i such that any function in $\Phi(C_i)$ which vanishes on S_i, and which vanishes at all points of S_i which are interior to C, to order at least N, is in $R'(C_i)$. Since Λ_i can be represented as a measure on C_i, and since Λ_i is orthogonal to $R'(C)$, we see that Λ_i will be orthogonal to any function in $\Phi(C)$ which agrees on C_i with a function in $R'(C_i)$. Thus $\Lambda_i(f)=0$ whenever f is a function in $\Phi(C)$ which vanishes on S_i and which vanishes to order at least N at all points of S_i which are interior to C. Since $g(p)=g(q)$ for all p and q in S_i and all g in R', it follows from the definition that Λ_i is a R'-local differential operator on $\Phi(C)$ of order not exceeding N, as was to be proved.

LEMMA 4. *Let C be a compact set whose intersection with a disc U is an open analytic arc A which divides $U-C$ into components U_1 and U_2. Let R' be a total subalgebra of R, and let μ be a Borel measure on C which is orthogonal to R'. Let there exist functions f and g in R' which are schlicht relative to C on U. Let $f(A)$ be in the outside boundary of $f(C \cup \overline{U_2})$, where $\overline{U_2}$ is the closure of U_2, and let $g(A)$ be in the outside boundary of $g(C \cup \overline{U_1})$. Then μ vanishes on all subsets of A.*

Proof. Consider any open sub-arc B of A, which has endpoints a and b in A with $\mu(\{a\})=\mu(\{b\})=0$. Let B_1 be any closed sub-arc of A which contains the closure of B in its interior. Since the analytic arc $f(A)$ forms part of the outside boundary of $f(C \cup \overline{U_2})$, we can find a function ϕ on $f(C \cup U_2)$ which is a uniform limit of polynomials, which maps $f(C \cup U_2 - B_1)$ into $\{z|\Im(z)>0\}$, which maps $f(B_1)$ in one-to-one fashion onto a subset of the real axis, and which maps the endpoints of $f(B)$ onto 0 and 1. To find ϕ, let J be a simple closed curve about the set $f(C\cup U_2)$ which has $f(B_1)$ as part of its boundary and which has no other points of $f(C\cup U_2)$ in its boundary. Let ϕ_1 be the Riemann map of the interior of J into the unit disc. Then by Bieberbach [1], it follows that ϕ_1 can be extended to be continuous on J and to map J homeomorphically onto $\{z \mid |z|=1\}$. By Mergelyan [3], ϕ_1 is the uniform limit of polynomials. Then we can find a function ϕ_2 which is analytic on the unit disc and continuous on the closed unit disc, which maps the closed unit disc in a one-to-one fashion into $\{z \mid \Im(z)\geqq 0\}$, which maps the arc $\phi_1(f(B_1))$ in one-to-one fashion onto a subset of the real axis, and which maps $\phi_1(f(a))$ and $\phi_1(f(b))$ (but not necessarily in that order) onto 0 and 1. The composite function $\phi=\phi_2\circ\phi_1$ will have the desired properties. Thus the function $f'=\phi\circ f$ is the uniform limit on $C\cup U_2$ of functions in R', maps $(C\cup U_2)-B_1$ into $\{z \mid \Im(z)>0\}$, maps B_1 in onet to-one fashion onto a subset of the real line, and maps B onto the uni- interval (0, 1). The function f' can be extended to be analytic and schlicht in some neighborhood of the closure of B because it maps U_2 into $\{z \mid \Im(z)>0\}$ and maps B_1 in one-to-one fashion into the real line.

In the same way we can find a function g' on $C\cup U_1$ which is the uniform limit of functions in R', which maps $C\cup U_1-B_1$ into $\{z \mid \Im(z)<0\}$, which maps B_1 in one-to-one fashion into the real axis, and which maps B onto (0, 1). As above, g' can be extended to be schlicht on some neighborhood of the closure of B, and the values of the extended function at points of U_2 sufficiently near to B will lie in the set $\{z \mid \Im(z)>0\}$. Thus both f' and g' have positive imaginary part at points of U_2 near B. Therefore f' and g' increase in the same direction along B. We may therefore label the endpoints a and b of B in such a way that $f'(a)=g'(a)=0$ and $f'(b)=g'(b)=1$. It is clear that the algebra T generated on $C-\{a, b\}$ by f' and g' is orthogonal to the measure μ, because $\mu(\{a\})=\mu(\{b\})=0$. The function

$$h_1=\frac{f'}{g'}\,\frac{g'-1}{f'-1},$$

defined on $C-\{a, b\}$, can be extended to a continuous function h_1 on C, because both numerator and denominator vanish only at a and b, about which points they can be extended to be analytic with simple zeros. For $\alpha>0$ consider the function

$$\frac{f'}{g'-\alpha i}\,\frac{g'-1}{f'-1+\alpha i},$$

defined on C. Its absolute value will be less than the absolute value of h_1. Therefore, as $\alpha\to 0$, it converges boundedly to h_1 on $C-\{a, b\}$.

Now $1/(g'-\alpha i)$ and $1/(f'-1+\alpha i)$ are uniform limits on C of polynomial functions of g' and f' respectively, so that h_1 is a bounded limit on $C-\{a, b\}$ of functions in the algebra T. Therefore all powers of h_1 are orthogonal to the measure μ. Now f' has positive imaginary part on $C-B_1$, so that $f'/(f'-1)$ has negative imaginary part on $C-B_1$. Similarly, $(g'-1)/g'$ has negative imginary part on $C-B_1$. Thus it is possible to define the arguments of $f'/(f'-1)$ and $(g'-1)/g'$ to be continuous on the set $C-B_1$ and to have values in the interval $(-\pi, 0)$. Since these functions are real on $B_1-\{a, b\}$, we may therefore define the arguments on $C-\{a, b\}$ to be continuous and to have values in the interval $(-\pi, 0]$. Thus the argument of h_1, since h_1 is the product of the functions just considered, can be defined continuously on C to have values in the interval $(-2\pi, 0]$. Since C is compact, the values will actually lie in the interval $(\epsilon-2\pi, 0]$ for some $\epsilon>0$. We may therefore obtain the function $\log h_1$ on C as a uniform limit of polynomial functions of h_1, so that the real part of $\log h_1$ will be $\log\left|\dfrac{f'(g'-1)}{g'(f'-1)}\right|$ and the imaginary part will have values in $(-2\pi, 0]$ and will vanish on B_1. It follows that $\int \log h_1 \, d\mu=0$.

For each $\alpha>0$, by an argument similar to the one just given, the function

$$h_\alpha=\frac{f'+\alpha i}{f'-1+\alpha i}\quad\frac{g'-1-\alpha i}{g'-\alpha i}$$

will be a uniform limit on C of polynomial functions of f' and g', and will have an argument function with values in the interval $(-2\pi, 0)$. Thus $\log h_\alpha$ can be defined to be a function on C which is a uniform limit of polynomial functions of f' and g', and whose imaginary part has values in the interval $(-2\pi, 0)$. Therefore, $\int \log h_\alpha \, d\mu=0$. The real part of $\log h_\alpha$ converges uniformly on $C-B_1$ to $\log|h_1|$, as $\alpha\to 0$, because g' and $f'-1$ are bounded away from 0 on $C-B_1$. Also the real part of $\log h_\alpha$ converges boundedly on $B_1-\{a, b\}$ to the same function, since the reality of f' and g' on B_1 implies that the absolute values of the functions $\dfrac{f'+\alpha i}{g'-\alpha i}$ and $\dfrac{g'-1-\alpha i}{f'-1+\alpha i}$ are nearer to 1 on B_1 than are $\left|\dfrac{f'}{g'}\right|$ and $\left|\dfrac{g'-1}{f'-1}\right|$ respectively. It follows that the real part of $\log h_\alpha$ converges boundedly on $C-\{a, b\}$ to $\log|h_1|=\Re(\log h_1)$. The imaginary part of $\log h_\alpha$, on the other hand, must converge boundedly on $C-B_1$ to $\Im(\log h_1)$, because h_α converges to h_1 on $C-B_1$ and both $\Im(\log h_1)$ and $\Im(\log h_\alpha)$ have values in the interval $(-2\pi, 0)$ on $C-B_1$.

On the sub-arc B of B_1, f' and g' are positive whereas $f'-1$ and $g'-1$ are negative, so that the argument of h_α will be a small positive number, on B, modulo 2π, if α is small, which means the argument of h_α will be near -2π on B. Thus, as $\alpha \to 0$, we see that $\Im(\log h_\alpha)$ converges to -2π on B. Similarly, we see that $\Im(\log h_\alpha)$ converges to 0 on $B_1 - B - \{a, b\}$. Thus $\log h_\alpha$ converges boundedly as $\alpha \to 0$ to a function h_2 on $C - \{a, b\}$, for which $\log h_1 - h_2$ has the value $2\pi i$ on B and the value 0 on $C - B - \{a, b\}$. Since $\int \log h_\alpha \, d\mu = 0$, we must have $\int h_2 \, d\mu = 0$. Therefore $0 = \int (\log h_1 - h_2) d\mu = 2\pi i \mu(B)$. Since this is true whenever μ vanishes at the endpoints of B, it follows that μ vanishes on all subsets of A, as was to be proved.

LEMMA 5. *Let R' be a total subalgebra of R. Let S be a compact set and C a compact subset of S. Let q_0 be a non-isolated point of $S-C$. Let g_0 be a function in R' which assumes its maximum modulus for S at the point q_0, and at no points of C. Let g_0 be non-constant on every component of the Riemann surface which contains points of S. Then there exists a function g in R' which assumes its maximum modulus for S at a unique point q, lying in $S-C$, and there exists a neighborhood W of q on which g is schlicht relative to S.*

Proof. Let

$$\Gamma_1 = \{p | p \in S, \ g_0 \text{ is not schlicht at } p\} .$$

Since, by the hypothesis, the points of S at which g_0 is not schlicht must be isolated, it follows that Γ_1 is finite. Therefore the set Γ, defined to be the union of Γ_1 and the singular set of S relative to R', is finite. Thus $g_0(S)$ is a compact subset of the complex plane, $g_0(C)$ is a compact subset of $g_0(S)$, and $g_0(q_0)$ is a point of maximum modulus of $g_0(S)$ which is a non-isolated point of $g_0(S) - g_0(C)$. Thus $g_0(q_0)$ is in the outside boundary of $g_0(S)$, and since $g_0(q_0)$ is a non-isolated point of $g_0(S)$, there must exist points z_0 distinct from $g_0(q_0)$ but arbitrarily near to $g_0(q_0)$ which lie in the outside boundary of $g_0(S)$. By taking z_0 sufficiently close to $g_0(q_0)$, we may assume that z_0 is not in $g_0(C)$, nor in the finite set $g_0(\Gamma)$. We may therefore find a point w in the unbounded component of the complement of $g_0(S)$ whose distance to z_0 is less than its distance to $g_0(C) \cup g_0(\Gamma)$. The minimum distance of w to $g_0(S)$ is therefore attained at no point of $g_0(C) \cup g_0(\Gamma)$. The function $(z-w)^{-1}$ of z therefore attains its maximum modulus for $g_0(S)$ at no point of $g_0(C) \cup g_0(\Gamma)$. Since w is in the unbounded component of the complement of $g_0(S)$, it follows that $(z-w)^{-1}$ can be uniformly approximated on some neighborhood N of $g_0(S)$ by polynomials h. If the approximation is sufficiently good, h will be schlicht on $g_0(S)$ because $(z-w)^{-1}$ is schlicht on N, and h will attain its maximum modulus for $g_0(S)$ at a point z_1 in $g_0(S)-g_0(C)-g_0(\Gamma)$. Therefore the function $g_1=h\circ g_0$ is in R' and attains its maximum modulus for S at a point q_1 (any point of S with $g_0(q_1)=z_1$) of $S-C-\Gamma$. Since q_1 is not in Γ, g_0 is schlicht at q_1. Since h is schlicht on $g_0(S)$, the function g_1 will therefore be schlicht at q_1.

Let the finite set S' consist of all those points p in S, except q_1, for which $g_1(p)=g_1(q_1)$. By replacing g_1 by $g_1+g_1(q_1)$, if necessary, we may assume that g_1 attains its maximum modulus for S only at q_1 and at points of S'. Since q_1 is not in Γ, we can find a function g_2 in R' with $g_2(q_1)=0$, $g_2(p)=-g_1(q_1)$ for all p in S'. Let ϵ be a positive number, and consider the function $g=g_1+\epsilon g_2$ of R'. Since g_1 is schlicht at q_1, there will exist a neighborhood U of q_1 such that g will be schlicht on U for all ϵ sufficiently small. Also there will exist a neighborhood V of the set S' such that $|g_2(p)+g_1(p)|<|g_1(q_1)|$ for all p in V, because we have $g_2(p)+g_1(p)=g_2(p)+g_1(q_1)=0$ for all p in S'. Thus for all p in $V\cap S$ we have

$$\begin{aligned}|g(p)|=|g_1(p)+\epsilon g_2(p)|=|(1-\epsilon)g_1(p)+\epsilon(g_2(p)+g_1(p))|<(1-\epsilon)|g_1(p)|\\ +\epsilon|g_1(q_1)|\leqq|g_1(q_1)|\leqq \sup\ \{|g(r)|\,|\,r\in S\}\ .\end{aligned}$$

Thus g does not attain its maximum modulus for S on the set V. If ϵ is sufficiently small, on the other hand, g can attain its maximum modulus for S only near S' or near q_1, since g_1 attains its maximum modulus only at S' and at q_1. Therefore g can attain its maximum modulus for S only at points of U, if ϵ is sufficiently small. The point q of U where this happens may not be unique, but if we take such a point q and replace g by $g+g(q)$, then q will be the unique point where g attains its maximum modulus for S, because g is schlicht on U. Since g assumes its maximum modulus at the unique point q in S and is schlicht on U, there will exist a disc W in U containing q on which g is schlicht relative to S. This completes the proof of the lemma.

LEMMA 6. *If F is a compact subset of the complex plane, and ν is a measure on F which is orthogonal to all polynomials, then for almost all real numbers x_0 there exists a measure β on the set $L=\{z|\Re(z)=x_0$ and z is not in the unbounded component of the complement of $F\}$, such that*

$$\int_{F_1} h\ d\nu=-\int_{F_2} h\ d\nu=\int h\ d\beta$$

for all polynomials h, where

$$F_1=F\cap\{z|\Re(z)\geqq x_0\} \text{ and } F_2=F\cap\{z|\Re(z)\leqq x_0\} .$$

Proof. There will exist a measure μ on F which assumes non-negative values and which dominates the complex-valued measure ν in the sense that $|\nu(s)|\leqq\mu(s)$ for all Borel sets S. Let ϕ be the non-negative, non-decreasing function of the real variable x_0 defined by $\phi(x_0)=\mu(\{x+iy|x\leqq x_0\})$. Then $\phi'(x_0)$ will exist for almost all x_0. Assume x_0 is such that $\phi'(x_0)$ exists. Then the equation $\int_{F_1} h\, d\nu=-\int_{F_2} h\, d\nu$ is a consequence of the equation $\int_F h\, d\nu=0$ and the fact that ν, because $\phi'(x_0)$ exists, vanishes on all subsets of $F_1\cap F_2$. By Runge's theorem, we will then have $\int_{F_1} g\, d\nu=-\int_{F_2} g\, d\nu$, whenever g is any function analytic on some neighborhood of the set consisting of the union of F and the bounded components of the complement of F. Choose ϵ with $0<\epsilon<1$. Write $T=\{z=x_0+iy|$ the distance from z to L does not exceed $\epsilon\}$, and $V=\{y|x_0+iy\in T\}$. Let h be any polynomial, and write $\|h\|=\sup\{|h(z)||z\in T\}$. For $\Re(z)>x_0$, define

$$h_1(z)=\frac{1}{2\pi i}\int_T h(\zeta)(\zeta-z)^{-1}d\zeta ,$$

where the direction of integration along T is upward. For $\Re(z)<x_0$, let

$$h_2(z)=\frac{1}{2\pi i}\int_T h(\zeta)(\zeta-z)^{-1}d\zeta .$$

Then it is well known and easy to see that both h_1 and h_2 have continuous boundary values at points z_0 of T which are interior points of T, relative to the line $\{z|\Re(z)=x_0\}$, and that the difference of those boundary values, $h_1(z_0)-h_2(z_0)$, is $h(z_0)$. Therefore, if we define $h_1(z)=h(z)+h_2(z)$ for $\Re(z)<x_0$, and $h_2(z)=h_1(z)-h(z)$ for $\Re(z)>x_0$, then by extending to the interior of T by continuity, we obtain analytic functions h_1 and h_2 on some neighborhood of the set consisting of the union of F and the bounded components of the complement of F, such that $h=h_1-h_2$.

Thus we have

$$\begin{aligned}\int_{F_1} h(z)d\nu(z)&=\int_{F_1} h_1(z)d\nu(z)-\int_{F_1} h_2(z)d\nu(z)\\ &=\int_{F_1} h_1(z)d\nu(z)+\int_{F_2} h_2(z)d\nu(z) .\end{aligned}$$

We consider the first term of this sum, and obtain

$$\left|\int_{F_1} h_1(z)d\nu(z)\right| = \left|\int_{F_1} \frac{1}{2\pi i}\int_T h(\zeta)(\zeta - z)^{-1}d\zeta\, d\nu(z)\right|$$

$$\leqq \|h\| \int_{F_1}\int_{v\in V}[(x_0-x)^2+(v-y)^2]^{-\frac{1}{2}}\, dv\, d\mu(z)$$

$$= \|h\| \int_{F_1}\int_{t\in V-y}[(x_0-x)^2+t^2]^{-\frac{1}{2}}dt\, d\mu(z)$$

$$\leqq \|h\| \int_{F_1}\int_{-M}^{M}[(x_0-x)^2+t^2]^{-\frac{1}{2}}dt\, d\mu(z)$$

$$= \|h\| \int_{x_0}^{K}\int_{-M}^{M}[(x_0-x)^2+t^2]^{-\frac{1}{2}}dt\, d\phi(x)\ ,$$

where M is some constant not depending on ϵ and where

$$K = \sup\{x | \Re(z) = x,\ z \in F_1\}\ .$$

Since $\phi'(x_0)$ exists, the difference quotient $[\phi(x)-\phi(x_0)](x-x_0)^{-1}$ will be bounded, so that there will exist a constant η such that $\phi(x)-\phi(x_0) < \eta(x-x_0)$ for all $x > x_0$. Thus the function ψ defined for all $x > x_0$ by $\psi(x) = \eta(x-x_0) - [\phi(x)-\phi(x_0)]$ is positive. Also

$$f(x) = \int_{-M}^{M}[(x-x_0)^2+t^2]^{-\frac{1}{2}}dt$$

is a positive decreasing function of x for $x > x_0$, and

$$\psi(x)f(x) \leqq \eta(x-x_0)f(x) \longrightarrow 0 \text{ as } x \longrightarrow x_0\ .$$

It follows by integration by parts that

$$\int_{x_0}^{K} f(x)d\psi(x) \geqq 0, \text{ or}$$

$$\int_{x_0}^{K} f(x)\, d\phi(x) \leqq \eta\int_{x_0}^{K} f(x)dx\ .$$

Therefore

$$\left|\int_{F_1} h_1(z)\, d\nu(z)\right| \leqq \eta\|h\| \int_{x_0}^{K}\int_{-M}^{M}[(x-x_0)^2+t^2]^{-\frac{1}{2}}dt\, dx\ .$$

Now the last integral is finite, as may be seen by transforming to polar coordinates. Now since a similar estimate can be obtained for

$$\left|\int_{F_2} h_2(z)\, d\nu(z)\right|\ ,$$

we see that there exists a constant Q, not depending on ϵ, such that $\left|\int_{F_1} h(z)d\nu(z)\right| \leqq Q\|h\|$, for all polynomials h. Since Q does not depend on ϵ, we see that

$$\left|\int_{F_1} h(z)\,d\nu(z)\right| \leqq Q \ \sup\ \{|h(z)|\,|\,z \in L\}\ ,$$

for all polynomials h. Since the linear functional $h \to \int_{F_1} h(z)\,d\nu(z)$ can be extended, by the Hahn-Banach theorem, to a linear functional of bound Q on $\Phi(L)$, we see that the measure β exists, as was required to prove.

LEMMA 7. *Let C be compact, and μ a measure on C orthogonal to the total subalgebra R'. Let $\mathscr{S}(C, R')=C$. Let f be a function in R'. Let a and c be real numbers, $a<c$, and let D be a closed disc containing the sets $C \cap \{q|\Re(f(q)) \geqq a\}$ and $\mathscr{S}(C, f) \cap \{q|\Re(f(q)) \geqq a\} \cap D$ in its interior, such that f is schlicht on D relative to C, and such that $D \cap \{q|\Re(f(q))=b\}$ is non-void whenever $a<b<c$. Then, for every b with $a<b<c$, there exists a measure μ' on $C \cap \{q|\Re(f(q)) \leqq b\}$ such that $\int g d\mu = \int g d\mu'$ for all g in $\Phi(C)$.*

Proof. Define a measure ν on $F=f(C)$ by $\nu(S)=\mu(f^{-1}(S))$. Then if h is any polynomial, we have $\int h d\nu = \int h \circ f d\mu = 0$, since $h \circ f \in R'$. Now let x_0 be chosen as in Lemma 6, where we may impose the additional requirement that $a<x_0<b$. It follows that the sets

$$E=\mathscr{S}(C, f) \cap \{q|\Re(f(q))=x_0\} \cap D$$

and $C_1=C \cap \{q|\Re(f(q)) \geqq x_0\}$ are contained in the interior of D. Write $C_2=C \cap \{q|\Re(f(q)) \leqq x_0\}$, so that $f(C_1)=F_1$ and $f(C_2)=F_2$, in the notation of Lemma 6. By the definition of ν, we see that $\int_{C_1} h \circ f d\mu = \int_{F_1} h d\nu$ for all polynomials h. Consider the complex number z_0 not in $f(E)$ with $\Re(z_0)=x_0$. There are two cases to consider, depending on whether z_0 is in $f(D)$ or not. In case $z_0 \in f(D)$, then $z_0=f(q_0)$ for q_0 in

$$(D-E) \cap \{q|\Re(f(q))=x_0\} \subset \mathscr{S}'(C, f)\ ,$$

by definition of E. Therefore, z_0 is in the unbounded component of the complement of $F=f(C)$. In case z_0 is not in $f(D)$, then z_0 can be joined to a point z_1 in the boundary of $f(D)$ by a closed interval I whose interior lies in $\{z|\Re(z)=x_0\}-f(D)$, because $\{z|\Re(z)=x_0\} \cap f(D)$ is non-void by the hypotheses of the theorem. Now $F \cap \{z|\Re(z)=x_0\}$ is contained in the interior of $f(D)$, because $C \cap \{q|\Re(f(q))=x_0\}$ is contained in the interior of D. It follows that the interval I lies in the complement of F. Since we have already seen that a point z_1 with $\Re(z_1)=x_0$ and $z_1 \in f(D)-f(E)$ must lie in the unbounded component of the complement of F, it follows that z_0 lies in the unbounded component of the complement of F. Thus, from a consideration of the two possible cases, we see that the set $\{z|\Re(z)=x_0\}-f(E)$ is a subset of the unbounded component of the complement of F. It follows that $L\subset f(E)$, where the set L is defined in Lemma 6. Thus, since f is schlicht on D, we may define the measure α on E by $\alpha(S)=\beta(f(S))$, where β is the measure on L defined in Lemma 6, and obtain $\int hd\beta=\int h\circ f d\alpha$ for all polynomials h. Thus

$$\int_{C_1} h\circ f d\mu=\int_{F_1} hd\nu=\int hd\beta=\int h\circ f d\alpha$$

for all polynomials h. Since both E and C_1 are subsets of D, and since any analytic function on D can be uniformly approximated on D by polynomial functions of f (because f is schlicht on D), we therefore see that $\int_{C_1} gd\mu=\int gd\alpha$ for all g in R. Since ν vanishes on all subsets of $F_1\cap F_2$, then μ will vanish on all subsets of $C_1\cap C_2$, so that

$$\int_{C_1} gd\mu=-\int_{C_2} gd\mu$$

for all g in R'. We therefore see that $\int_{C_2} gd\mu=-\int gd\alpha$ for all g in R'. Thus if H is the carrier of the measure α and if g is in R', we see that $\int_{C_1\cup H} gd(\mu-\alpha)=0$, and $\int_{C_2\cup H} gd(\mu+\alpha)=0$.

We now show that α, which we know is a measure on E, is actually a measure on $E\cap C$, that is, that the carrier H of α is a subset of C. Assume first that $H-C$ contains an isolated point r. Then r is isolated point of $H\cup C_1$, and since f is schlicht on the subset $H\cup C_1$ of D, the point $f(r)$ is an isolated point of $f(H\cup C_1)$. Also $\Re(f(r))=x_0\leqq\Re(z)$ for all z in $f(H\cup C_1)$. It follows that the function θ on $f(H\cup C_1)$ which has value 1 at $f(r)$ and vanishes elsewhere is a uniform limit of polynomials. Thus $\theta\circ f$ is in $R'(H\cup C_1)$. By the equation derived at the end of the last paragraph, it follows that $\alpha(\{r\})=-\int_{C_1\cup H}\theta\circ f d(\mu-\alpha)=0$. This contradicts the fact that r is an isolated point of the carrier H of α, and hence $H-C$ has no isolated points. There exists a function g_0 in R' which assumes its maximum modulus for $H\cup C$ at no point of C, if $H-C$ is non-void, because $\mathscr{S}(C, R')=C$. Since $H\cup C$ is compact, there are only a finite number of components of the Riemann surface which intersect $H\cup C$.

Since R' is total, we can find g_1 in R' which is non-constant on each component of the surface which intersects $H\cup C$. Therefore, if ϵ is sufficiently small, the function $g_2=g_0+\epsilon g_1$ in R' will be non-constant on each component of the surface which intersects $H \cup C$, and will assume its maximum modulus for $H \cup C$ at no point of C. Therefore, by Lemma 5, there exists g in R' assuming its maximum modulus for $H \cup C$ at a unique point q of $H-C$ which has a neighborhood on which g is schlicht relative to $H \cup C$. Since $q \in E-C$, we can find an arc B of $\{r|\Re(f(r))=x_0\} \cap D$ which contains q in its interior, which is disjoint from C, and which lies in some disc $N \subset D-C$ on which g is schlicht relative to $H \cup C$. We may choose N and B so that $N \cap \{r|\Re(f(r))>x_0\}$ and $N \cap \{r|\Re(f(r))<x_0\}$ are connected. Write $S=H \cup C \cup B$. Then we can find a point q_0 in N such that

$$|g(q_0)| > \max\{|g(r)||r \in B\} = \max\{|g(r)||r \in S\} .$$

By moving q_0 slightly, we may actually assume that

$$q_0 \in N - \{r|\Re(r))=x_0\} .$$

Let U be a disc contained in N and containing q_0 and q such that $U \cap S$ is an open sub-arc A of B dividing $U-A$ into components

$$U_1 = U \cap \{r|\Re(f(r))>x_0\} \text{ and } U_2 = U \cap \{r|\Re(f(r))<x_0\} ,$$

with $\overline{U} \cap S = \overline{A}$, where $\overline{U}$ is the closure of U. Since f is schlicht on D relative to C, and since $U \subset D$ and $S \subset D \cup C$, then f is schlicht on U relative to S.

Let q_1 be any point of $S \cup \overline{U}$, at which g assumes its maximum modulus. Since $|g(q_1)| \geqq |g(q_0)| > \max\{|g(r)|r \in S\}$, we have $q_1 \in \overline{U}-S$. Thus either $q_1 \in \overline{U}_1$ or $q_1 \in \overline{U}_2$, but q_1 is not in $\overline{U}_1 \cap \overline{U}_2 \subset \overline{A} \subset S$. Assume $q_1 \in \overline{U}_1$. Then $g(q_1)$ is in the boundary of the unbounded component of the complement of $g(S \cup \overline{U})$, since it is a point of maximum modulus of $g(S \cup \overline{U})$. Since $g(q_1)$ is not in $g(S \cup \overline{U}_2)$, it is therefore in the unbounded component of the complement of $g(S \cup \overline{U}_2)$. The set $g(\overline{U}_1 - B)$ is connected and disjoint from $g(S \cup \overline{U}_2)$, because $\overline{U}_1 - B$ is disjoint from $S \cup \overline{U}_2$ and g is schlicht on $\overline{U}$ relative to S. Since $g(q_1) \in g(\overline{U}_1 - B)$, it follows that $g(\overline{U}_1 - B)$ is in the unbounded component of the complement of $g(S \cup \overline{U}_2)$. Since $g(A) = g(B \cap U)$ is in the boundary of $g(\overline{U}_1 - B)$, it follows that $g(A)$ is in the outside boundary of $g(S \cup \overline{U}_2)$, in this case. In case $q_1 \in \overline{U}_2$, it similarly follows that $g(A)$ is in the outside boundary of $g(S \cup \overline{U}_1)$.

First consider the case in which $g(A)$ is in the outside boundary of $g(S \cup \overline{U}_1)$. Then $g(A)$ is in the outside boundary of $g(H \cup C_2 \cup B \cup \overline{U}_1)$. Since the real part of f equals x_0 on A and is less than or equal to x_0 on $H \cup C_2 \cup \overline{U}_2$, the open arc $f(A)$ is in the outside boundary of

$$f(H \cup C_2 \cup B \cup \overline{U}_2) .$$

Since $\int_{H\cup C_2 \cup B} h\, d(\mu+\alpha)=0$ for all h in R', we can apply Lemma 4, to the compact set $H\cup C_2 \cup B$, to the measure $\mu+\alpha$, to the disc U, and to the functions f and g in R', to conclude that the measure $\mu+\alpha$, and therefore α itself, vanishes on all subsets of U. Next consider the case in which $g(A)$ is in the outside boundary of $g(S\cup \overline{U}_2)$. Then $g(A)$ is in the outside boundary of $g(H\cup C_1 \cup B \cup \overline{U}_2)$. Since the real part of f equals x_0 on A and is greater than or equal to x_0 on $H\cup C_1\cup B\cup \overline{U}_1$, in this case $f(A)$ is in the outside boundary of $f(H\cup C_1 \cup B \cup \overline{U}_1)$. Since $\int_{H\cup C_1\cup B} h\, d(\mu-\alpha)=0$ for all h in R', we see by Lemma 4 again that the measure $\mu-\alpha$, and therefore α, vanishes on all subsets of U. Thus, in either case, we see that α vanishes on all subsets of U. This contradicts the fact that the point q in U is in the carrier H of α. This contradiction shows that $H-C$ is void, so that α is a measure on $E\cap C$.

Now $\mathscr{S}(C_1, R')\subset \mathscr{S}(C, R')=C$. Moreover, if $q\in C-C_1$ then $q\in \mathscr{S}'(C_1, R')$ because $\Re(f(q))<x_0\leqq \Re(f(q'))$ for all q' in C_1. Hence $\mathscr{S}(C_1, R')=C_1$. If $D-C_1$ were not connected, there would exist a component of $D-C_1$ containing only interior points of D (because C_1 is a subset of the interior of D), so that $\mathscr{S}(C_1, R')$ would contain all points of this component, contradicting the fact that $\mathscr{S}(C_1,R')=C_1$. Thus $D-C_1$ is connected. Since f is schlicht on D, it follows that $F_1=f(C_1)$ has a connected complement. By the theorem of Mergelyan, every continuous function on F_1 which is analytic at interior points can therefore be uniformly approximated by polynomials. From this it follows that every continuous function on C_1 which is analytic at interior points can be uniformly approximated by polynomial functions of f, so that $\Phi(C_1)=R'(C_1)$. Since $H\subset E\cap C\subset C_1$, and since we have already seen that $\int_{C_1} g d\mu=\int_H g d\alpha$ for all g in R', it follows that

$$\int_{C_1} g d\mu=\int_H g d\alpha$$

for all g in $\Phi(C_1)$. If we define the measure μ' on

$$(C-C_1)\cup H\subset C_2\subset C\cap \{q|\Re(f(q))\leqq b\}$$

by $\mu'(S)=\mu(S-C_1)+\alpha(S)$, we obtain

$$\int g d\mu'=\int_{C-C_1} g d\mu+\int_H g d\alpha=\int_{C-C_1} g d\mu+\int_{C_1} g d\mu=\int g d\mu$$

for all g in $\Phi(C)$, as was to be proved.

3. The main theorem and its consequences.

THEOREM 1. *Let R' be a total subalgebra of R. Let C be a compact set with $\mathscr{S}(C, R')=C$. Let Λ be a bounded linear functional on $\Phi(C)$ which is orthogonal to $R'(C)$. Then Λ is a R'-homogeneous differential operator on $\Phi(C)$, whose order does not exceed some positive integer N depending only on R' and C.*

Proof. We know that Λ can be represented as a measure on C. Therefore the class Γ, consisting of all compact subsets S of C for which Λ can be represented as a measure on S and for which $\mathscr{S}(S, R')=S$, is non-void, because $C\in\Gamma$. We construct a sequence $\{S_n\}$ of sets from Γ by taking $S_1=C$, and choosing S_{n+1} such that $S_{n+1}\subset S_n$ and

$$\rho(S_n,\ S_{n+1})\geqq\frac{1}{2}\sup\{\rho(S_n,\ S)|S\subset S_n,\ S\in\Gamma\}\ .$$

Then $\rho(S_n,\ S_{n+1})\to 0$ as $n\to\infty$, because otherwise the compact set C would contain an infinite set of points whose mutual distances were larger than some fixed positive number. Write $S=\cap S_n$, and assume that there exists a point q_0 in S not in the singular set T of C relative to R'. Then there exists a function g_0 in R' which vanishes on T but does not vanish on q_0. Since S is compact, there exist only a finite number of components of the surface which intersect S. Since R' is total, there exists a function g_1 in R' which is non-constant on every component of the surface which intersects S. Thus, if ϵ is sufficiently small, $g_2=g_0+\epsilon g_1$ will be non-constant on every component of the surface which intersects S, and the set K consisting of those points of S where g_2 attains its maximum modulus will not intersect T. If there exists a point in K which is a non-isolated point of S, then by Lemma 5 there exists a function f in R' which attains its maximum modulus for S at a unique point p, and which is schlicht relative to S on some closed disc D containing p in its interior. On the other hand, if all points of K are isolated, then K is finite, and since K does not intersect T, there exists a function g_3 in R' which has the value $g_2(p)$ at some point p of K, which has the value $-g_2(r)$ at all other points r of K, and which is schlicht at p. For a sufficiently small positive number ϵ, it follows that the function $f=g_2+\epsilon g_3$ will attain its maximum modulus for S at the unique point p and will be schlicht relative to S on some closed disc D containing p in its interior. Thus, if we assume that S is not a subset of the singular set of C relative to R', we may find f, p, and D which have the properties described. We may assume also that $f(p)>0$.

Let a_0 be some real number less than $f(p)$ such that the set

$$D \cap \{q | \Re(f(q)) = a\}$$

is non-void whenever $a_0 < a < f(p)$. For each real number a with $a_0 < a < f(p)$, consider the compact sets $V_a = \mathscr{S}(S, f) \cap \{q | \Re(f(q)) \geqq a\} \cap D$ and $W_a = S \cap \{q | \Re(f(q)) \geqq a\}$. The intersection of the V_a is $V_{f(p)} = \{p\}$, and the intersection of the W_a is $W_{f(p)} = \{p\}$. Thus, if a is sufficiently near to $f(p)$, the sets V_a and W_a will be contained in the interior of D. Having chosen such a value of a, define the compact sets

$$V_n = \mathscr{S}(S_n, f) \cap \{q | \Re(f(q)) \geqq a\} \cap D$$

and $W_n = S_n \cap \{q | \Re(f(q)) \geqq a\}$, for each positive integer n. Since $\cap S_n = S$, we have $\cap V_n = V_a$ and $\cap W_n = W_a$. Thus, if n is sufficiently large, the sets V_n and W_n will be contained in the interior of D. Let b be any number with $a < b < f(p)$, and choose a value of n for which V_n and W_n are contained in the interior of D, for which f is schlicht on D relative to S_n, and for which $2\rho(S_n, S_{n+1})$ is less than the distance d of p to $\{q | \Re(f(q)) \leqq b\}$. Then by Lemma 7, we see that there exists a measure ν on $S_n \cap \{q | \Re(f(q)) \leqq b\} = S'_n$ which represents Λ, because there exists such a measure on S_n. Now $\mathscr{S}(S'_n, R') \subset \mathscr{S}(S_n, R') = S_n$. Also, if $q \in S_n - S'_n$, then $\Re(f(q)) > b \geqq \sup \{\Re(f(q')) | q' \in S'_n\}$, so that $q \in \mathscr{S}'(S'_n, R')$. Thus $\mathscr{S}(S'_n, R') = S'_n$, and so $S'_n \in \Gamma$. Also $\rho(S_n, S'_n) \geqq d > 2\rho(S_n, S_{n+1})$. This contradicts the choice of S_{n+1}. Therefore S is a subset of the singular set of C relative to R'. Since $\cap S_n = S$ and since Λ can be represented as a measure on S_n, then Λ can be represented as a measure on an arbitrary C-neighborhood of S. It follows from Lemma 3 that Λ is a R'-homogeneous differential operator on $\Phi(C)$, of order not exceeding N, as was to be proved.

COROLLARY 1. *If C is compact, and if R' is a total subalgebra of R, with $\mathscr{S}(C, R') = C$, then there exists a positive integer N such that $R'(C)$ contains the ideal $I(C, R', N)$ of $\Phi(C)$ consisting of those functions in $\Phi(C)$ which vanish on the singular set S of C relative to R' and which vanish to order at least N at those points of S which are interior to C. The ideal $I(C, R', N)$, and therefore $R'(C)$ itself, has finite codimension when considered as a vector subspace of $\Phi(C)$.*

Proof. Choose N as in Theorem 1. Then, by Theorem 1, it follows that every bounded linear functional on $\Phi(C)$ which vanishes on $R'(C)$ will vanish on $I(C, R', N)$. It follows from the Hahn-Banach theorom that $I(C, R', N) \subset R'(C)$. The last statement of the corollary is obvious.

COROLLARY 2. *If C is compact, if R' is a total subalgebra of R with $\mathscr{S}(C, R') = C$, and if the singular set of C relative to R' is void, then $R'(C) = \Phi(C)$.*

Proof. This corollary is in immediate consequence of Corollary 1. This corollary applies to R itself, if no component of the surface is compact, since then it is known that R is total, and that the singular set of R relative to C is void, for all C.

COROLLARY 3. *Let C be compact and without interior points. Let R' be a total subalgebra of R with $\mathscr{S}(C, R')=C$. Let f be a continuous function on C for which $f(p)=f(q)$ whenever p and q are points in C for which $h(p)=h(q)$ for all h in R'. Then $f\in R'(C)$.*

Proof. Let Λ be a bounded linear functional on $\Phi(C)$ which is orthogonal to $R'(C)$. We must show that $\Lambda(f)=0$, and the Hahn-Banach theorem will do the rest. Since Λ, by Theorem 1, is a R'-homogeneous differential operator on $\Phi(C)$, and since C has no interior points, we see that Λ is a finite sum $\Lambda=\Sigma\Lambda_i$, where Λ_i is orthogonal to $R'(C)$ and has the form $\Lambda_i(g)=\sum_{j=1}^{n_i} a_{ij}\,g(p_{ij})$, with p_{ij} in C and with $h(p_{ij})=h(p_{i1})$ for $1\leqq j\leqq n_i$ and all h in R'. Thus $f(p_{ij})=f(p_{i1})$ for $1\leqq j\leqq n_i$. Since the function 1 is in R', this implies $\sum_{j=1}^{n_i} a_{ij}=0$. Thus we have

$$\Lambda_i(f)=\sum_{j=1}^{n_i} a_{ij}f(p_{ij})=\sum_{j=1}^{n_i} a_{ij}f(p_{i1})=f(p_{i1})\sum_{j=1}^{n_i} a_{ij}=0 .$$

This completes the proof.

The hypothesis that R' contain the constant functions, which is made in Theorem 1 (because R' is required to be total), is undesirable, since, for instance, it rules out the case of an ideal R'. We now show that this hypothesis is not necessary to the validity of Theorem 1. To this end, let R' and C satisfy the hypotheses of Theorem 1, except that we weaken the word "total" by dropping the requirement that R' contain the constant functions. Let Λ be any bounded linear functional on $\Phi(C)$ which is orthogonal to $R'(C)$. Let the original Riemann surface be enlarged by the addition of the extra disc $\{z||z|<1\}$ as a new component, and let the algebra T' on the new surface consist of all functions of the form $c+f$, where c is a constant, and where f is any analytic function on the new surface which vanishes at the center $z=0$ of the extra disc and which agrees on the original surface with some function in R'. Let H be the union of C and the subset $\left\{z||z|\leqq\frac{1}{2}\right\}$ of the extra disc. Then Λ can be considered as a bounded linear functional on $\Phi(H)$, and obviously the functional Λ' on $\Phi(H)$ defined by $\Lambda'(g)=\Lambda(g-g(0))$ will vanish on $T'(H)$. By Theorem 1, we see that Λ' is a T'-homogeneous differential operator on $\Phi(H)$ of order not exceeding some constant N depending on T' and H (and, therefore, depending on R' and C). It follows that Λ is a R'-homogeneous differential operator on $\Phi(C)$ of order not exceeding N, as was to be proved.

REFERENCES

1. L. Bierberbach, *Lehrbuch der Funktionentheorie* Vol. 2, Leipzing, 1930.
2. J. Kelley, *General topology*, New York, 1955.
3. S. Mergelyan, *On the representation of functions by series of polynomials on closed sets*, Amer. Math. Soc. Translation, No. 85, 1953.
4. S. Saks and A. Zygmund, *Analytic functions*, Warsaw, 1952.

UNIVERSITY OF CALIFORNIA, BERKELEY

THE RELATIONS BETWEEN A SPECTRAL OPERATOR AND ITS SCALAR PART

S. R. FOGUEL

1. Introduction. It is shown in Dunford's theory of spectral operators, that every spectral operator T can be decomposed into the sum of a scalar operator S, and a generalized nilpotent N [1]. We study here properties which are inherited by S from T. The main results are:

1. If the spectral operator T is compact, weakly compact, or has a closed range, then respectively S is compact, weakly compact, or has a closed range.

2. The relations between the point spectra, continuous spectra, and residual spectra of S and T are investigated.

3. If the sum of two commuting spectral operators is spectral, then the sum of their scalar parts is scalar.

2. Notation. Most of the notation is taken from [1]. Let X be a complex Banach space. A spectral measure is a set function $E(\cdot)$, defined on Borel sets in the complex plane, whose values are projections on X, which satisfy:

(α) For any two Borel sets σ and δ $E(\sigma)E(\delta)=E(\sigma\cap\delta)$.

(β) Let Φ be the void set and p the complex plane.

Then

$$E(\Phi)=0 \text{ and } E(p)=I .$$

(γ) There exists a constant M such that $|E(\sigma)|\leq M$, for every Borel set σ.

(δ) The vector valued set function $E(\cdot)x$ is countable additive for each $x\in X$.

The operator T is a spectral operator, whose resolution of the identity is the spectral measure $E(\cdot)$ if

(a) for every Borel set σ $E(\sigma)T=TE(\sigma)$.

(b) Let T_α denote the restriction of T to the subspace $E(\alpha)X$, (T_α $=T|E(\alpha)X$) then

Received March 13, 1957, in revised form August 19, 1957. This paper is a part of a dissertation presented for the degree of Doctor of Philosophy in Yale University. The author wishes to express his gratitude to Professor N. Dunford for his guidance and kind encouragement.

$$\sigma(T_\alpha)\subset\overline{\alpha}$$

where $\sigma(A)$ is the spectrum of A.

Throughout the paper T denotes a spectral operator, $E(\cdot)$ its resolution of the identity, S its scalar part given by $S=\int_p \lambda E(d\lambda)$, N its radical given by $N=T-S$. The operator N is a generalized nilpotent, and the operators N, S, T, $E(\alpha)$ commute [1]. A spectral operator is of finite type, if for some integer n, $N^{n+1}=0$. We shall denote $N\cdot E(\langle 0\rangle)$ by N_0, hence $N_0=TE(\langle 0\rangle)=E(\langle 0\rangle)T$.

3. **Topological properties.** In this section, several topological properties will be shown to be valid for S whenever they are valid for T. The following lemma will be used.

LEMMA 1. *S is in the uniformly closed operator algebra generated by the projections $E(\alpha)$ with $0\notin\overline{\alpha}$.*

Proof. $S=\int_{\sigma(T)}\lambda E(d\lambda)$ and $\sigma(T)$ is bounded, see [1] Theorem 1. Given $\varepsilon>0$ let $\sigma(T)$ be divided into the disjoint sets $\alpha_0, \alpha_1, \cdots, \alpha_n$ with

$$\begin{aligned} &0\in\alpha_0\,, \qquad 0\notin\overline{\alpha}_i\,, \qquad i=1,2,\cdots,n \text{ and}\\ &\operatorname{diam}(\alpha_i)<\varepsilon \qquad\qquad i=0,1,2,\cdots,n\,. \end{aligned}$$

Let $\lambda_0=0$ and $\lambda_i\in\alpha_i$. Then

$$\left|S-\sum_{i=1}^n \lambda_i E(\alpha_i)\right|=\left|\int_{\sigma(T)}\Bigl(\lambda-\sum_{i=0}^n \lambda_i\chi_{\alpha_i}(\lambda)\Bigr)E(d\lambda)\right|\,.$$

If $\lambda\in\sigma(T)$ then

$$\left|\lambda-\sum_{i=0}^n \lambda_i\chi_{\alpha_i}(\lambda)\right|\leqq\varepsilon\,.$$

Now by [1], p. 330, for every bounded measurable function defined on $\sigma(T)$

$$\left|\int_{\sigma(T)}f(\lambda)E(d\lambda)\right|\leqq\sup\{|f(\lambda)|\,, \qquad \lambda\in\sigma(T)\}\cdot 4M\,.$$

Hence

$$\left|S-\sum_{i=1}^n \lambda_i E(\alpha_i)\right|\leqq 4M\varepsilon\,.$$

THEOREM 1. *Let $\mathfrak{A}$ be a uniformly closed right (left) ideal in the algebra of operators on X. If T belongs to $\mathfrak{A}$ so do S, N, and $E(\alpha)$ with $0\notin\overline{\alpha}$.*

Proof. By condition b of §2 T_α with $0\notin\overline{\alpha}$ possesses a bounded everywhere defined inverse T_α^{-1}. Let us define P_α by $P_\alpha x = T_\alpha^{-1}E(\alpha)x$, $x \in X$, $0 \notin \bar{\alpha}$. P_α is a bounded everywhere defined operator. Now

$$TP_\alpha x = T(T_\alpha^{-1}E(\alpha)x) = (TT_\alpha^{-1})(E(\alpha)x) = E(\alpha)x \ .$$

Also

$$P_\alpha Tx = T_\alpha^{-1}E(\alpha)Tx = T_\alpha^{-1}TE(\alpha)x = (T_\alpha^{-1}T)E(\alpha)x = E(\alpha)x \ .$$

Hence if $0 \notin \bar{\alpha}$ then $E(\alpha) \in \mathfrak{A}$. Note that this fact remains true even if $\mathfrak{A}$ is not uniformly closed. Now by Lemma 1 $S \in \mathfrak{A}$ and therefore $N \in \mathfrak{A}$ too.

COROLLARY 1. *If T is compact then so are S, N and $E(\alpha)$ $(0 \notin \bar{\alpha})$.*

COROLLARY 2. *If T is weakly compact then so are S, N and $E(\alpha)$ with $0 \notin \bar{\alpha}$.*

COROLLARY 3. *If $TX \subset Y$ where Y is a closed subspace of X, then $SX \subset Y$ and $NX \subset Y$ and $E(\alpha)X \subset Y$, $0 \notin \bar{\alpha}$. Hence*

$$SX \cup NX \cup \cup (E(\alpha)X \mid 0 \notin \bar{\alpha}) \subset \overline{TX}$$

and if the range of T is separable so are the ranges of S, N and $E(\alpha)$, $0 \notin \bar{\alpha}$.

COROLLARY 4. *If $A_0T = 0$ $(TA_0 = 0)$ then $A_0S = A_0N = 0$ and $A_0E(\alpha) = 0$, $0 \notin \bar{\alpha}$ $(SA_0 = NA_0 = E(\alpha)A_0 = 0$ if $0 \notin \bar{\alpha})$. In particular T is a spectral operator of finite type if and only if some power of N annihilates T.*

COROLLARY 5. *If $Tx = 0$ then $Nx = Sx = E(\alpha)x = 0$ where $\bar{\alpha}$ does not contain 0.*

COROLLARY 6. *If (x_n) is a bounded sequence of vectors, and the sequence (Tx_n) has a limit then the sequences (Sx_n), (Nx_n) and $(E(\alpha)x_n)$ with $0 \notin \bar{\alpha}$ have limits.*

To prove these corollaries one has to note that:

(a) The classes of compact and weakly compact operators are uniformly closed two-sided ideals. (See [3] Chapter 6).

(b) The classes of operators A satisfying $AX \subset Y$ or $A_0A = 0$ are uniformly closed right ideals.

(c) The classes of operators A satisfying $Ax = 0$ or $AA_0 = 0$ or the limit of Ax_n exists are uniformly closed left ideals.

REMARK TO COROLLARY 6. By the proof of Theorem 1 the sequence $(E(\alpha)x_n)$, $0 \notin \bar{\alpha}$, has a limit whenever the sequence (Tx_n) has, even if the sequence (x_n) is not bounded.

THEOREM 2. *$AT=0$ if and only if $AE(p-\langle 0\rangle)=0$ ($A=AE(\langle 0\rangle)$) and $AN_0=0$. Similarly $TA=0$ if and only if $E(p-\langle 0\rangle)A=N_0A=0$.*

Proof. If $AN_0=AE(p-\langle 0\rangle)=0$ then $AE(\alpha)=AE(p-\langle 0\rangle)E(\alpha)=0$ if $0\notin\bar{\alpha}$, thus by Lemma 1 $AS=0$. Now

$$AN=ANE(\langle 0\rangle)+ANE(p-\langle 0\rangle)=AN_0+(AE(p-\langle 0\rangle))N=0 .$$

Thus $AT=AS+AN=0$. Conversely if $AT=0$ then $AN_0=ATE(\langle 0\rangle)=0$, and $AE(\alpha)=0$ if $0\notin\bar{\alpha}$. Now for each $x\in X$

$$AE(p-\langle 0\rangle)x=\lim AE\left\{z\,\middle|\,\frac{1}{n}\leqq|z|\right\}x=0$$

by countable additivity.

The second half of the theorem is proved in the same way.

Using Corollary 5 one can prove in the same way that $Tx=0$ if and only if $N_0x=E(p-\langle 0\rangle)x=0$.

COROLLARY 1. *If $E(\langle 0\rangle)=0$, then $AT=0$ or $TA=0$ if and only if $A=0$.*

Proof. By Theorem 2 if $AT=0$ or $TA=0$ then $A=AE(\langle 0\rangle)$ or $A=E(\langle 0\rangle)A$.

COROLLARY 2. *If $E(\langle 0\rangle)=0$ then $\overline{TX}=X$.*

Proof. If $\overline{TX}\neq X$ then there exists a bounded functional $x^*\neq 0$ such that $x^*(TX)=0$. Let $Ax=x^*(x)x_1$ where x_1 is any vector different from 0. $AT=0$ and $A\neq 0$ which contradicts Corollary 1.

THEOREM 3. *If T has a closed range so does S.*

1. *Proof.* Let $E(\langle 0\rangle)=0$ then Corollary 2 of Theorem 2 shows that $\overline{TX}=X$. But by assumption $\overline{TX}=TX$, thus $TX=X$. Also, the operator T is one-to-one by [1] p. 327 and thus T possesses a bounded everywhere defined inverse. Thus $0\notin\sigma(S)=\sigma(T)$ and $SX=X$.

2. Let $E(\langle 0\rangle)\neq 0$. The operator $T_{p-\langle 0\rangle}$ is a spectral operator whose resolution of the identity $F(\cdot)$ is given by $F(\alpha)=E(\alpha)E(p-\langle 0\rangle)=E(\alpha-\langle 0\rangle)$, hence $F(\langle 0\rangle)=0$. Now if $T_{p-\langle 0\rangle}x_n\to y$ ($y\in E(p-\langle 0\rangle)X$), then, there exists a vector x in X such that $Tx=y$, because T has a closed range. Therefore

$$T_{p-\langle 0\rangle}(E(p-\langle 0\rangle)x)=TE(p-\langle 0\rangle)x=E(p-\langle 0\rangle)Tx=E(p-\langle 0\rangle)y=y .$$

Hence $T_{p-\langle 0\rangle}$ satisfies the same conditions assumed for T in the first part and therefore $0 \notin \sigma(T_{p-\langle 0\rangle})$ and

$$S_{p-\langle 0\rangle}X = E(p-\langle 0\rangle)X\ , \quad \text{but} \quad S_{p-\langle 0\rangle}X = SX\ ,$$

so S has a closed range.

By the proof of the last theorem it follows that if T has a closed range then $0 \notin \sigma(T_{p-\langle 0\rangle})$, hence 0 is an isolated point of the spectrum of T.

THEOREM 4. *The operator T has a closed range if and only if*
1. *0 is an isolated point of $\sigma(T)$.*
2. *The operator N_0 has a closed range.*

Proof. We proved that Condition 1 is necessary. Now if $N_0x_n \to y$ then $E(\langle 0\rangle)N_0x_n \to E(\langle 0\rangle)y$ but $E(\langle 0\rangle)N_0 = N_0$ thus $E(\langle 0\rangle)y = y$. Also $N_0 = TE(\langle 0\rangle)$ and T has a closed range, thus if $T(E(\langle 0\rangle)x_n) \to y$ then for some x, $Tx = y$. Hence $TE(\langle 0\rangle)x = N_0x = E(\langle 0\rangle)y = y$. Conversely if 1. and 2. are satisfied let $Tx_n \to y$. Then

$$\begin{aligned} TE(p-\langle 0\rangle)x_n + TE(\langle 0\rangle)x_n &= TE(p-\langle 0\rangle)x_n + N_0x_n \\ &\to y = E(p-\langle 0\rangle)y + E(\langle 0\rangle)y\ . \end{aligned}$$

Multiplying this equation by $E(p-\langle 0\rangle)$ and $E(\langle 0\rangle)$ one gets the following two equations

$$\begin{aligned} TE(p-\langle 0\rangle)x_n &\to E(p-\langle 0\rangle)y \\ N_0x_n &\to E(\langle 0\rangle)y \end{aligned}$$

By 1. $T_{p-\langle 0\rangle}$ possesses a bounded everywhere defined inverse. Hence, for some x_1 in $E(p-\langle 0\rangle)X$, $Tx_1 = E(p-\langle 0\rangle)y$.
By 2. for some vector x_2, $N_0x_2 = E(\langle 0\rangle)y$. Thus

$$T(x_1 + E(\langle 0\rangle)x_2) = Tx_1 + N_0x_2 = y.$$

4. **Properties of spectral points.** Let A be a bounded linear operator on X, define

$\sigma_p(A) = \{\lambda \mid \lambda I - A$ is not one-to-one$\}$

$\sigma_c(A) = \{\lambda \mid \lambda I - A$ is one-to-one and $(\lambda I - A)X$ is dense in X, but not equal to $X\}$.

$\sigma_r(A) = \{\lambda \mid \lambda I - A$ is one-to-one and $\overline{(\lambda I - A)X} \neq X\}$.

(See [6] p. 292.)

The sets $\sigma_p(A)$, $\sigma_c(A)$ and $\sigma_r(A)$ are disjoint and

$$\sigma(A) = \sigma_p(A) \cup \sigma_c(A) \cup \sigma_r(A)\ .$$

THEOREM 1. *If T is a spectral operator of finite type, then $\lambda \in \sigma_p(T)$ if and only if $E(\langle\lambda\rangle)\neq 0$, and $\lambda \in \sigma_c(T)$ if and only if $E(\langle\lambda\rangle)=0$, and $\lambda \in \sigma(T)$. Thus $\sigma(T)=\sigma_p(T) \cup \sigma_c(T)$.*

Proof. If $E(\langle\lambda\rangle)\neq 0$ let $x \in E(\langle\lambda\rangle)X$, $x \notin 0$, then

$$Sx=\int_{\sigma(T)}\mu E(d\mu)x=\int_{\sigma(T)}\mu E(d\mu)E(\langle\lambda\rangle)x=\lambda x .$$

Let ν be the first integer such that $N^{\nu}x=0$, then

$$TN^{\nu-1}x=SN^{\nu-1}x+N^{\nu}x=N^{\nu-1}Sx=\lambda N^{\nu-1}x ,$$

therefore $\lambda \in \sigma_p(T)$. If $E(\langle\lambda\rangle)=0$ then Corollary 2 of Theorem 2, §3, applied to $\lambda I-T$, shows that $\overline{(\lambda I-T)X}=X$. Also, by [1] Lemma 1, $\lambda I-T$ is one-to-one and thus $\lambda \in \sigma_c(T)$.

THEOREM 2. *$\sigma_c(S)\subset\sigma_c(T)$ and $\sigma_p(T) \cup \sigma_r(T)\subset\sigma_p(S)$.*

Proof. If $\lambda \in \sigma_c(S)$ then $E(\langle\lambda\rangle)=0$, and by the last part of the proof of Theorem 1, $\lambda \in \sigma_c(T)$. Thus $\sigma_c(S)\subset\sigma_c(T)$ and

$$\sigma_p(T)\cup\sigma_r(T)=\sigma(T)-\sigma_c(T)\subset\sigma(T)-\sigma_c(S)=\sigma(S)-\sigma_c(S)=\sigma_p(S) .$$

If $E(\langle\lambda\rangle)=0$ then $\lambda \in \sigma_c(T)$. Let us examine therefore the case where $E(\langle\lambda\rangle)\neq 0$. To simplify notation assume that $\lambda=0$.

THEOREM 3. *Let $E(\langle 0\rangle)\neq 0$ then*

1. *$0 \in \sigma_p(T)$ if N_0 is not one-to-one on $E(\langle 0\rangle)X$.*
2. *$0 \in \sigma_c(T)$ if N_0 is one-to-one on $E(\langle 0\rangle)X$ and $\overline{N_0(E(\langle 0\rangle)X)}=E(\langle 0\rangle)X$.*
3. *$0 \in \sigma_r(T)$ if N_0 is one-to-one on $E(\langle 0\rangle)X$ and $\overline{N_0(E(\langle 0\rangle)X}\neq E(\langle 0\rangle)X$.*

Proof.

1. If there exists a vector x such that $x\neq 0$, $x=E(\langle 0\rangle)x$ and $N_0x=0$ then

$$Tx=TE(\langle 0\rangle)x=N_0x=0$$

2. The operator $T_{p-\langle 0\rangle}$ is one-to-one on $E(p-\langle 0\rangle)X$ by [1] Lemma 1. Now if N_0 is one-to-one on $E(\langle 0\rangle)X$ then T is one-to-one on X: If $Tx=0$ then $E(\langle 0\rangle)Tx=N_0x=N_0E(\langle 0\rangle)x=0$ and $TE(p-\langle 0\rangle)x=T_{p-\langle 0\rangle}E(p-\langle 0\rangle)x=0$. Thus $E(\langle 0\rangle)x=0$ and $E(p-\langle 0\rangle)x=0$, but then $x=E(\langle 0\rangle)x+E(p-\langle 0\rangle)x=0$. Now by Corollary 2 of Theorem 2, §3

$$\overline{T_{p-\langle 0\rangle}E(p-\langle 0\rangle)X}=E(p-\langle 0\rangle)X$$

and by assumption

$$\overline{N_0X}=E(\langle 0\rangle)X$$

but

$$\overline{TX}\supset\overline{T_{p-\langle 0\rangle}E(p-\langle 0\rangle)X}$$

and

$$\overline{TX}\supset\overline{N_0X}$$

therefore

$$\overline{TX}\supset X\,.$$

3. By Part 2, T is one-to-one. Let x be a vector in $E(\langle 0\rangle)X$ whose distance from N_0X is greater than some positive number r. Let y be any vector in X. Then

$$|x-Ty|=|x-N_0y-TE(p-\langle 0\rangle)y|\,.$$

Hence

$$|x-Ty|\geqq\frac{1}{M}|E(\langle 0\rangle)[x-N_0y-TE(p-\langle 0\rangle)y]|$$
$$=\frac{1}{M}|x-N_0E(\langle 0\rangle)y|\geqq\frac{r}{M}\,.$$

Hence

$$x\notin\overline{TX}\,.$$

The next theorem is valid for separable spaces only.

THEOREM 4. *If X is separable, then $\sigma_p(T)\cup\sigma_r(T)$ is countable.*

Proof. Theorems 1 and 2 show that $\sigma_p(T)\cup\sigma_r(T)\subset\sigma_p(S)=\{\lambda\,|\,E(\langle\lambda\rangle)\neq 0\}$. For any λ in $\sigma_p(S)$ let x_λ be a vector satisfying $|x_\lambda|=1$ and $E(\langle\lambda\rangle)x_\lambda=x_\lambda$. Now if $\lambda_1\neq\lambda_2$ then

$$|x_{\lambda_1}-x_{\lambda_2}|\geqq\frac{1}{M}|E(\langle\lambda_1\rangle)(x_{\lambda_1}-x_{\lambda_2})|=\frac{|x_{\lambda_1}|}{M}=\frac{1}{M}\,.$$

The set $\{x_\lambda\,|\,\lambda\in\sigma_p(S)\}$ is separable because X is, hence the set is countable.

We conclude this discussion by studying another subset of the spectrum.

DEFINITION. Let A be a bounded linear operator on X, then $\sigma_0(A)$

$=\{\lambda \mid \text{there exists a sequence } (x_n) \text{ such that } |x_n|=1 \text{ and } (\lambda I-A)x_n \to 0\}$. See [5] p. 51.

LEMMA 1. $\sigma_p(S) \subset \sigma_0(T)$.

Proof. Let $x \neq 0$ satisfy $Sx=\lambda x$. If for some n, $N^n x=0$, let us take the first such integer. Then

$$TN^{n-1}x=(S+N)N^{n-1}x=N^{n-1}Sx=\lambda N^{n-1}x ,$$

and thus $\lambda \in \sigma_p(T) \subset \sigma_0(T)$. If for every n, $N^n x \neq 0$ then

$$T\frac{(N^n x)}{|N^n x|}=(S+N)\frac{N^n x}{|N^n x|}=\lambda\frac{N^n x}{|N^n x|}+\frac{N^{n+1}x}{|N^n x|} .$$

It is enough to show that for some subsequence n_i

$$\frac{|N^{n_i+1}x|}{|N^{n_i}x|} \to 0 .$$

Let us assume, to the contrary, that for some $\varepsilon>0$ $|N^{n+1}x| \geqq \varepsilon |N^n x|$ for all n, then

$$|x| \leqq \frac{|Nx|}{\varepsilon} \leqq \frac{|N^2 x|}{\varepsilon^2} \leqq \cdots \leqq \frac{|N^n x|}{\varepsilon^n} ,$$

but this would imply that

$$\lim_{n\to\infty} \sqrt[n]{|N^n|}=\lim_{n\to\infty} \sqrt[n]{|N^n|}\sqrt[n]{|x|} \geqq \limsup_{n\to\infty} \sqrt[n]{|N^n x|}$$

$$\geqq \limsup_{n\to\infty} \varepsilon\sqrt[n]{|x|}=\varepsilon .$$

But N is a generalized nilpotent and thus $\lim_{n\to\infty}\sqrt[n]{|N^n|}=0$.

THEOREM 5. $\sigma(T)=\sigma_0(T)$.

Proof. By Theorem 2 and Lemma 1 $\sigma_p(T) \cup \sigma_r(T) \subset \sigma_p(S) \subset \sigma_0(T)$. Thus it is enough to show that $\sigma_c(T) \subset \sigma_0(T)$. Let $\lambda \in \sigma_c(T)$ we may assume that $\lambda=0$. If $0 \notin \sigma_0(T)$ then $|Tx| \geqq \varepsilon |x|$, $x \in X$, for some positive ε. This implies that TX has a closed range, but $\overline{TX}=X$ hence $TX=X$, which contradicts the assumption that $0 \in \sigma_c(T)$.

Let us conclude this section with a few examples.

1. Define in l_1 the generalized nilpotent operator N by

$$N(x_1, x_2, x_3, \cdots)=(x_2, 0, x_4, 0, \cdots)$$

and let $S=0$. S is compact while T is not weakly compact.

2. Let X be the space of continuous functions on $[0, 1]$ vanishing

at the point 0. Define N by $Nf=g$, $g(x)=\int_0^x f(s)\,ds$, and let $S=0$. S has a closed range while T does not. $0 \in \sigma_p(S)$ but $0 \in \sigma_c(T)$.

3. Let N be defined as in 2, and $S=I$. T and S have closed ranges but the range of N is not closed.

5. **Decompositions of spectral operators.** Let $T_1, \cdots, T_n$ be n commuting operators. There exists a minimal algebra of operators $\mathfrak{A}$, with the properties:

1. $T_i \in \mathfrak{A}$, $i=1, 2, \cdots, n$.
2. If $U \in \mathfrak{A}$ and U^{-1} is a bounded everywhere defined operator then $U^{-1} \in \mathfrak{A}$.
3. The algebra $\mathfrak{A}$ is uniformly closed.

This algebra will be called the full algebra generated by $T_1, \cdots, T_n$, and it is a commutativealgebra. Let $\Delta_{\mathfrak{A}}$ denote the space of homomorphisms from $\mathfrak{A}$ to the algebra of complex numbers. By Condition 2, and the Gelfand theory [4], if $U \in \mathfrak{A}$ then $\sigma(U)=\{\mu(U) \mid \mu \in \Delta_{\mathfrak{A}}\}$; thus if $\mu(U)=0$ for each $\mu \in \Delta_{\mathfrak{A}}$ then U is a generalized nilpotent.

LEMMA 1. *Every scalar operator S is the sum S_1+iS_2 where S_1 and S_2 are scalar operators and*

1. $S_1S_2=S_2S_1$.
2. *$\sigma(S_1)$ and $\sigma(S_2)$ are sets of real numbers.*
3. *The Boolean algebra of projections generated by the resolutions of the identity of S_1 and S_2 is bounded.*

Proof. Let $E(\cdot)$ be the resolution of the identity of S; then

$$S=\int zE(dz)=\int(x+iy)E(dz)=\int xE(dz)+i\int yE(dz)$$
$$=\int \lambda E_1(d\lambda)+i\int \lambda E_2(d\lambda)$$

where

$$E_1(\alpha)=E\{z \mid z=x+iy \text{ and } x \in \alpha\}$$

$$E_2(\alpha)=E\{z \mid z=x+iy \text{ and } y \in \alpha\}$$

Conditions 1, 2, and 3 are readily verified.

THEOREM 1. *Let T be a spectral operator. Then there exist two operators R and J such that*

1. *$T=R+iJ$ and $RJ=JR$*
2. *The sets $\sigma(R)$ and $\sigma(J)$ are real sets.*
3. *R is a scalar operator and J is a spectral operator.*

4. *The Boolean algebra of projections generated by the resolutions of the identity of R and J is bounded.*

If R_1 and J_1 satisfy Conditions 1 *and* 2, *then they are spectral operators and there exists a generalized nilpotent M such that*

$$R_1=R+M\ , \qquad J_1=J+iM\ .$$

REMARK. By the last assertion and Theorem 8 of [1] Conditions 1, 2, and 3 insure uniqueness. We shall call R the real part of T and J the imaginary part of T.

Proof. Let $T=S+N$. Using the notation of Lemma 1, put $R=S_1$, $J=S_2-iN$, and Conditions 1., 2., 3., and 4. follow by Lemma 1. Now, if R_1 and J_1 satisfy 1., and 2., then by Theorem 5 of [1], the operators R, J, R_1, J_1 commute. Let $\mathfrak{A}$ be the full algebra generated by these operators, if $\mu\in\Delta_{\mathfrak{A}}$ then

$$0=\mu(T-T)=\mu(R-R_1)+i\mu(J-J_1)$$

but $\mu(R-R_1)$ and $\mu(J-J_1)$ are real numbers by Condition 2. Hence

$$\mu(R-R_1)=\mu(J-J_1)=0\ .$$

Thus if $M=R-R_1$ then M is a generalized nilpotent and $J-J_1=iM$.

LEMMA 2. *Every scalar operator S can be written as the product of two scalar operators T_1 and T_2 which satisfy*

1. $T_1T_2=T_2T_1=S$.

2. *$\sigma(T_1)$ is a set of non-negative numbers and $\sigma(T_2)$ is a subset of the unit circle.*

3. *The Boolean algebra of projections generated by the resolutions of the identity of T_1 and T_2 is bounded.*

Proof. It follows from the multiplicative property of the spectral measure $E(\cdot)$ of S that

$$S=\int\lambda E(d\lambda)=\int|\lambda|E(d\lambda)\int\operatorname{sgn}\lambda E(d\lambda)\ .$$

Thus $S=T_1T_2$ where

$$T_1=\int|\lambda|E(d\lambda)=\int\mu E_1(d\mu)\ \text{if}\ E_1(\cdot)$$

is defined by

$$E_1(\alpha)=E\{\lambda\,|\,|\lambda|\in\alpha\}$$

and

$$T_2=\int \operatorname{sgn} \lambda E(d\lambda)=\int \mu E_2(d\mu)$$

where

$$E_2(\alpha)=E\{\lambda \mid \operatorname{sgn} \lambda \in \alpha\} .$$

It is easy to verify Conditions 1, 2, and 3.

THEOREM 2. *Let T be a spectral operator. Then there exist two operators P and U such that*

1. *$T=PU=UP$.*
2. *$\sigma(P)$ is a set of non-negative numbers and $\sigma(U)$ is a subset of the unit circle.*
3. *U is a scalar operator and P is spectral.*
4. *The Boolean algebra of projections generated by the resolutions of the identity of P and U is bounded.*

If P_1 and U_1 satisfy 1. and 2., then they are spectral operators and $U_1=U+N_1$ $P_1=P+N_2$ where N_1 ond N_2 are generalized nilpotents and

$$N_2=\sum_{n=0}^{\infty}(-N_1U^{-1})^{n+1}P .$$

REMARK. By the last assertion Conditions 1, 2, and 3 insure uniqueness. The operator P will be called the absolute value of T and U the argument of T.

Proof. Let $T=S+N$. Using the notation of Lemma 2 put $P=(T_1+T_2^{-1}N)$ and $U=T_2$, then $PU=T$ because $T_2N=NT_2$ (Theorem 8 of [1]). Now, Conditions 1, 2, 3, and 4 follow by Lemma 2. Let P_1 and U_1 satisfy 1 and 2; then by Theorem 8 of [1], P_1, U_1, P, U commute. Let $\mathfrak{A}$ be the full algebra generated by these operators. If $\mu \in \Delta_{\mathfrak{A}}$ then $\mu(T)=\mu(P)\mu(U)=\mu(P_1)\mu(U_1)$ and by Condition 2 $\mu(P)=\mu(P_1)$ and $\mu(U)=\mu(U_1)$. Thus $N_1=U_1-U$ and $N_2=P_1-P$ are generalized nilpotents. Now

$$T=UP=(U+N_1)(P+N_2)=UP+N_1P+UN_2+N_1N_2$$

or

$$-PN_1=(U+N_1)N_2$$

hence

$$\begin{aligned} N_2&=-(U+N_1)^{-1}N_1P \\ &=-\Big(\sum_{n=0}^{\infty}(-1)^n(U^{-1})^{n+1}N_1^n\Big)N_1P=\sum_{n=0}^{\infty}(-U^{-1}N_1)^{n+1}P . \end{aligned}$$

In order to apply these theorems we need the following result.

THEOREM 3. *A spectral operator T is a scalar operator whose spectrum lies on the unit circle if and only if: T^{-1} is a bounded everywhere defined operator, and there exists a constant M such that*

$$|T^n| \leqq M \qquad n = \pm 1, \pm 2, \cdots .$$

Proof. If $T = \int_{|\lambda|=1} \lambda E(d\lambda)$ then

$$|T^n| = \left| \int_{|\lambda|=1} \lambda^n E(d\lambda) \right| \leqq 4 \sup \{ |E(\alpha)| \mid \alpha \text{ a Borel set} \} ,$$

by [1], p. 341. Conversely assume that $|T^n| \leqq M$ $n = \pm 1, \pm 2, \cdots$ then

$$R(\lambda ; T) = \begin{cases} \sum_{n=0}^{\infty} \dfrac{T^n}{\lambda^{n+1}} , & |\lambda| > 1 \\ -\sum_{n=0}^{\infty} \lambda^n (T^{-1})^{n+1} , & |\lambda| < 1 \end{cases}$$

because that two series converge. Thus $\sigma(T) \subset \{\lambda \mid |\lambda| = 1\}$ and $|R(\lambda ; T)| \leqq M/|1 - |\lambda||$ if $|\lambda| \neq 1$. By Lemma 3.16 of [2] if $T = S + N$, where S is scalar and N is a generalized nilpotent, then $N^2 = 0$. Hence

$$T^n = S^n + nNS^{n-1} .$$

Therefore $nN = (T^n - S^n) S^{-(n-1)}$.

Thus nN is a bounded sequence of operators and therefore $N = 0$.

LEMMA 3. *Let S_1 and S_2 be two commuting scalar operators with real spectra, if $S_1 + S_2$ is spectral then it is scalar.*

Proof. Let $S_1 + S_2 = S + N$ where S is scalar and N is a generalized nilpotent. By Theorem 3 the operator $e^{i(S+N)} = e^{iS_1} \cdot e^{iS_2}$ is a scalar operator, but

$$e^{i(S+N)} = e^{iS} e^{iN} = e^{iS} + iNe^{iS} \sum_{n=1}^{\infty} \frac{(iN)^{n-1}}{n!} ,$$

hence

$$iNe^{iS} \sum_{n=1}^{\infty} \frac{(iN)^{n-1}}{n!} = 0$$

but the operator $ie^{iS} \sum_{n=1}^{\infty} \dfrac{(iN)^{n-1}}{n!}$ possesses an inverse and thus $N = 0$.

THEOREM 4. *Let S_1 and S_2 be two commuting scalar operators, if $S_1 + S_2$ is spectral then*

1. *$S_1 + S_2$ is a scalar operator.*

2. *The real (imaginary) part of* S_1+S_2 *is the sum of the real (imaginary) parts of* S_1 *and* S_2.

Proof. Let S_1, S_2 and S_1+S_2 be decomposed into real and imaginary parts as in Theorem 1. Then

$$S_1=R_1+iJ_1\ , \qquad S_2=R_2+iJ_2\ , \qquad S_1+S_2=R+iJ$$

where R_1, J_1, R_2, J_2 and R are scalar operators, while J is spectral, and would be scalar if and only if S_1+S_2 is a scalar operator. The operators R_1, J_1, R_2, J_2 commute and thus by the Gelfand theory [4] R_1+R_2 and J_1+J_2 have real spectra. By Theorem 1 $R_1+R_2=R+M$ and $J_1+J_1=J+iM$, where M is a generalized nilpotent. By Lemma 3 the operator R_1+R_2 is a scalar operator, but R is scalar too, thus by Theorem 8 of [1] $M=0$. Now $J_1+J_2=J$ which is a spectral operator and, again, by Lemma 3, J is scalar. Thus S_1+S_2 is scalar and $R_1+R_2=R$, $J_1+J_2=J$.

THEOREM 5. *Let* S_1 *and* S_2 *be two commuting scalar operators. If* S_1S_2 *is spectral then*

1. S_1S_2 *is a scalar operator.*
2. *The absolute value (argument) of* S_1S_2 *is the product of the absolute values (arguments) of* S_1 *and* S_2.

Proof. Let S_1, S_2 and S_1S_2 be decomposed as in Theorem 2.

$$S_1=P_1U_1\ , \qquad S_2=P_2U_2 \qquad S_1S_2=PU\ .$$

The operators U_1, U_2, U, P_1 and P_2 are scalar, and P is a spectral operator, which is scalar if and only if S_1S_2 is scalar. Using commutativity of the operators in question and Theorem 2 we derive that

$$P_1P_2=P+N_2\ , \qquad U_1U_2=U+N_1\ ,$$

where N_1 and N_2 are generalized nilpotents and $N_2=\sum_{n=0}^{\infty}(-N_1U^{-1})^{n+1}P$. By Theorem 3, $N_1=0$ and hence $N_2=0$ too, which proves the second assertion. In order to complete the proof it remains to show that P_1P_2 is scalar. Now P is spectral, let $P=P_1P_2=S+M$ where S is scalar and M a generalized nilpotent. Let $E(\cdot)$ and $F(\cdot)$ be the resolutions of the identity of P_1 and P_2 respectively. Denote $E\{\lambda\,|\,\lambda>\varepsilon_1\}=E_{\varepsilon_1}$ and $F\{\lambda\,|\,\lambda>\varepsilon_2\}=F_{\varepsilon_2}$, then the spectrum of $E_{\varepsilon_1}P_1F_{\varepsilon_2}P_2=SE_{\varepsilon_1}F_{\varepsilon_2}+ME_{\varepsilon_1}F_{\varepsilon_2}$ on $E_{\varepsilon_1}F_{\varepsilon_2}X$ is contained in the set $\{\lambda\,|\,\lambda\geqq\varepsilon_1\varepsilon_2\}$ by the Gelfand theory. The operator $\log(E_{\varepsilon_1}P_1E_{\varepsilon_2}P_2)$ is thus well defined and it is not difficult to show that it is equal to $\log(E_{\varepsilon_1}P_1)+\log(E_{\varepsilon_2}P_2)$. This sum is spectral by [1], p. 340, and by Theorem 4 it is scalar. Thus $E_{\varepsilon_1}P_1F_{\varepsilon_2}P_2$ is scalar and therefore $ME_{\varepsilon_1}F_{\varepsilon_2}=0$. By countable additivity $ME_0F_0=0$ but $P_1E_0=P_1$ and $P_2F_0=P_2$. Thus

$$P_1P_2=P_1E_0P_2F_0=SE_0F_0+ME_0F_0=SE_0F_0\ ,$$

but $P_1P_2=S+M$, hence $S+M=SE_0F_0$, therefore $S=SE_0F_0$ and $M=0$ by Theorem 8 of [1]. Hence $P_1P_2=S$ is a scalar operator.

REMARK. From Theorems 4 and 5 it follows that the sum or product of two commuting spectral operators is spectral, if and only if, the sum or product of their scalar parts is scalar.

A decomposition of a non-spectral operator A into real and imaginary parts is possible in some cases.

THEOREM 6. *Let A be an operator and $\sigma(A)\subset K$ where K satisfies*

1. *There exists a function f which is analytic and one-to-one in a neighborhood of K.*
2. *The image of K is a subset of the unit circle.*
3. *The inverse function of f exists and is analytic in a neighborhood of the unit circle, let us denote this function by g.*
4. $g(\bar{z})=\overline{g(z)}$ *if* $|z|=1$.

Then $A=A_1+iA_2$ where $\sigma(A_1)$ and $\sigma(A_2)$ are sets of real numbers and $A_1A_2=A_2A_1$. If $A=B_1+iB_2$ where B_1 and B_2 satisfy the same conditions then $B_1=A_1+N$ and $B_2=A_2+iN$ and N is a generalized nilpotent.

Proof. Let $\varphi(z)=g(1/f(z))$ then φ is analytic in a neighborhood of K and for $z\in K$, $\varphi(z)=\bar{z}$. Define

$$A_1=\frac{A+\varphi(A)}{2} \quad \text{and} \quad A_2=\frac{A-\varphi(A)}{2i}\ .$$

If $\mathfrak{A}$ is the full algebra generated by A and $\mu\in\Delta_{\mathfrak{A}}$,

$$\mu(A_1)=\frac{\mu(A)+\varphi(\mu(A))}{2}$$

is the real part of $\mu(A)$, and $\mu(A_2)$ is the imaginary part of $\mu(A)$. Thus the first part of the theorem is proved. The second part is proved as in Theorem 1.

We conclude this section by a study of roots of operators. The operator B is said to be an nth root of A if $B^n=A$. The operators A and B commute $AB=BA=B^{n+1}$. Let $\mathfrak{A}$ be the full algebra generated by B. If $\mu\in\Delta_{\mathfrak{A}}$ then $\mu(B)^n=\mu(A)$ thus

$$\sigma(B)\subset(\sigma(A))^{1/n}$$

Thus if $B^n=I$ then $\sigma(B)\subset\{\lambda\,|\,\lambda^n=1\}$ and hence is a finite set. By Theorem VII. 3.20 of [3], B is spectral and by Theorem 3, B is a scalar operator. Thus

$$B=\sum_{k=0}^{n-1} e^{2k\pi i/n}E_k \qquad \text{where } E_k^2=E_k,\ E_kE_j=0$$

if $k\neq j$, and $\sum_{k=0}^{n-1} E_k=I$.

THEOREM 7. *Let S be a scalar operator with real spectrum whose resolution of the identity is $E(\cdot)$. Let $S_1=\int \lambda^{1/n}E(d\lambda)$ where* $\arg \lambda^{1/n} =(\arg \lambda)/n$. *If S_2 satisfies $S_2^n=S$, then $\sigma(S_2)\subset(\sigma(S))^{1/n}$, and if $\sigma(S_2)\subset\{\lambda^{1/n}\mid \lambda \in \sigma(S)$ and $\arg \lambda^{1/n}=(\arg \lambda)/n\}$ then*

$$S_2=S_1+N \quad \textit{and} \quad N=NE(\langle 0\rangle) \quad \textit{and} \quad N^n=0\ .$$

Proof. The operators S_1 and S_2 commute by [1] p. 329. Let $\mathfrak{A}$ be the full algebra generated by them. If $\mu \in \Delta_{\mathfrak{A}}$ then $\mu(S_1)=\mu(S_2)$ and thus $S_2-S_1=N$ is a generalized nilpotent. Now

$$S=S_2^n=S_1^n+nNS_1^{n-1}+\frac{n(n-1)}{2}N^2S_1^{n-2}+\cdots+N^n=S_1^n \tag{1}$$

therefore

$$N\left(nS_1^{n-1}+\frac{n(n-1)}{2}NS_1^{n-2}+\cdots+N^{n-1}\right)=0$$

but by Corollary 4 of Theorem 1, Section 3, $NS_1^{n-1}=0$. Thus by Theorem 2 of §3, $N=NE(\langle 0\rangle)$, but then $NS_1^q=0$ for every integer q. Instead of (1) we have, therefore,

$$S=S_1^n+N^n \quad \text{or} \quad N^n=0$$

which completes the proof.

REFERENCES

1. N. Dunford, *Spectral operators*. Pacific J. Math. **4** (1954), 321–354.
2. ———, *Spectral theory*. II *Resolutions of the identity*. Pacific J. Math. **2** (1952), 559–614.
3. ———, and J. Schwartz, *Linear operators*. Interscience Pub., New York.
4. I. M. Gelfand, *Normierte Ringe*. Mat Sb. N. S. **9** (51), (1941), 3–24.
5. P. R. Halmos, *Introduction to Hilbert space and the theory of spectral multiplicity*. Chelsea, New York, 1951.
6. A. C. Zaanen, *Linear analysis*. P. Noordhoff, Groningen, and Interscience Pub., New York, 1953.

YALE UNIVERSITY

EUCLIDEAN AND WEAK UNIFORMITIES

J. R. ISBELL

Introduction. This paper is a study in the structure of some special classes of uniform spaces. In outline, machinery is developed in successive stages, roughly two stages. The first stage is illustrated by an unsuccessful attack on the characterization of subspaces of Euclidean spaces, in the usual uniform structure. The second stage leads to a characterization of those uniform spaces which are subspaces of Euclidean spaces in the finest structure consistent with the topology.

The main tool in the second stage is a covariant functor on uniform spaces to uniform spaces which is closely analogous to the derivative, the main tool employed by Ginsburg and the author in [4]. It yields also a number of results which complement, and a couple which improve, results of [4] and of [5].

That tool is inapplicable to the study of the usual Euclidean uniform structure. The approach attempted is to get a subspace of E^n as the inverse limit of the nerves of its uniform covering, or of any basis of uniform coverings. Indeed there is a basis of coverings whose nerves are uniformly equivalent to subspaces of E^n—*Euclidean coverings*, let us say, and the nerves, *Euclidean complexes*—and in some sense one can set up an inverse system of mappings on these nerves "uniformly" within E^n. The contribution of this paper is to formalize this approach and clear away imaginary difficulties, leaving the very real difficulties of characterizing Euclidean complexes and formulating reasonable criteria for a whole sequence of complexes connected by mappings to fit smoothly in E^n. Beyond this, it is shown that for a simplicial complex to be Euclidean, it is sufficient that its 1-skeleton should be Euclidean.

The author has profited from discussions of this material with Ernest Michael, G. D. Mostow, and Edward Nelson.

1. Coverings. We follow the usual practice of designating a topological space (X, T) by the abbreviation X. For a uniform space (X, μ) we write μX. As is fairly well known, the uniformity is determined by a knowledge of

(a) which relations in X are entourages, or

(b) which coverings of X are uniform, or

(c) which pseudometrics on X are uniformly continuous. In this

Received August 20, 1957. Supported by a National Science Foundation fellowship. Presented in part to the American Mathematical Society, December 27, 1956, under the title "Weak uniformities induced by families of real-valued functions".

paper we are concerned mostly with coverings, and therefore we adopt the convention that μ is the family of all uniform coverings. It is convenient to choose the convention according to which a uniform covering need not consist of open sets. Let us recall the defining conditions: X is a completely regular topological space and μ is a family of coverings of the set X such that

(i) If $u \in \mu$ and u is a refinement of v, then $v \in \mu$;

(ii) The intersection $u \wedge v$ of two coverings in μ is in μ;

(iii) Every covering in μ has a star-refinement in μ;

(iv) If $\{U_\alpha\} \in \mu$ then the interiors U_α^0 form a covering and this covering is in μ;

(v) For any point x the stars of x with respect to coverings in μ form a neighborhood basis at x. (The reader who is unfamiliar with the terminology should consult Tukey. [10])

Recall the notation $u <^* v$ for "u is a star-refinement of v", and $St(A, u)$ for the star of a set A with respect to a covering u. A *normal sequence* of coverings is a sequence (u^n) satisfying $u^{n+1*} < u^n$ for all n. Recall that a function is uniformly continuous if and only if the inverse image of every uniform covering is uniform.

We need the fundamental result

1.0. *For every uniform covering u of a uniform space μX there is a uniformly continuous pseudometric d on μX such that for each x in X, the set of all y such that $d(x, y) < 1$ is a subset of some element of u.* Exactly this result does not seem to be in print, though Bourbaki has a proof [3] of the corresponding statement connecting entourages with pseudometrics. It will suffice to sketch the similar proof of 1.0. Take a normal sequence (u^n) of uniform coverings, with $u^0 = u$. For each x, y, in X, let $g(x, y)$ be 0 if $St(x, u^n)$ contains y for all n, 2 if $St(x, u^n)$ never contains y, and otherwise 2^{1-n}, where n is the largest index for which $y \in St(x, u^n)$. Let $d(x, y)$ be the infimum of 1 and all the various finite sums $\sum g(p_i, p_{i+1})$, where $p_1 = x$ and $p_n = y$. By the form of the definition, d is a pseudometric. To see that d is uniformly continuous on $\mu X \times \mu X$, it suffices to observe that $u^n \times u^n$ is a uniform covering on each element of which d varies no more than 2^{1-n}. Finally, suppose $\sum g(p_i, p_{i+1}) \leqq 1$, $p_1 = x$, $p_n = y$. If we pick p and q respectively so that

(1) p is the last p_i such that $g(p_1, p_2) + \cdots + g(p_{i-1}, p_i) \leqq 1/2$, and

(2) q is the last p_j such that $g(p_i, p_{i+1}) + \cdots + g(p_{j-1}, p_j) \leqq 1/2$, then computation shows that also $g(p_j, p_{j+1}) + \cdots + g(p_{n-1}, p_n) \leqq 1/2$. If x and y are not both in some element of u^1, then one of the pairs (x, p), (p, q), (q, y), fails to be contained in any element of u^2. Then induction leads to a contradiction which completes the proof.

A family of functions f_α all defined on one uniform space μX into one uniform space νY is *equiuniformly continuous* if for each uniform covering v of νY there is a uniform covering u of μX which is at once a refinement of all $f_\alpha^{-1}(v)$. We wish to regard the nerve of a covering, or any simplicial complex, as a uniform space in the structure in which a mapping f into the complex is uniformly continuous if and only if the functions f_α, into the real line, which are the barycentric coordinates of f, form an equiuniformly continuous family. This dictates the following definition. A *uniform complex* μX is a simplicial complex X consisting of points x with barycentric coordinates x_α, provided with the distance function $d(x, y) = \max|x_\alpha - y_\alpha|$, and the uniformity μ induced by d.

In this paper the nerve $N(u)$ of a covering u is always regarded as a uniform complex. The general vertex of the nerve of $\{U_\alpha\}$ is called α. The *star* of a vertex α_0 is the union of the incident simplexes, that is, the set of all points q in $N(\{U_\alpha\})$ with nonzero α_0th coordinate. Note that the stars of vertices always form an open covering $\{St(\alpha)\}$, but this covering is uniform if and only if the complex is finite-dimensional. For any function h with values in a uniform complex, the coordinate functions h_α constitute a partition of unity. If $\{h_\alpha\}$ is a partition of unity subordinated to the covering $\{U_\alpha\}$, this means precisely that for all α, $h^{-1}(St(\alpha)) \subset U_\alpha$. In the finite-dimensional case we may summarise as follows. A covering w is *realized* by a mapping into a uniform space if it is refined by the inverse image of some uniform covering. We have

1.1. *An equiuniformly continuous partition of unity subordinated to a finite-dimensional uniform covering of a uniform space determines a realization of the covering by a uniformly continuous mapping into its nerve.*

It should be noted that for infinite-dimensional complexes it might well be desirable to employ a different uniformity, and perhaps even a different topology. In this paper we shall be concerned only with finite-dimensional complexes, and the choice of definitions is partially justified by

1.2. THEOREM. *To every finite-dimensional uniform covering of a uniform space there is subordinated an equiuniformly continuous partition of unity.*

Proof. For every uniform covering u of μX there is a uniformly continuous pseudometric d, as given by 1.0, such that each point x is in at least one $U_\alpha \in u$ which contains the sphere of d-radius 1 about x. If u is finite-dimensional, so that each x is in at most n sets U_α, consider the functions $d_\alpha(x) = d(x, Y - U_\alpha)$. For each x, $\sum d_\alpha(x)$ is a finite sum and hence a definite real number $e(x)\geqq 1$. Let $f_\alpha(x)=d_x(x)/e(x)$. The functions f_α form a partition of unity subordinated to u. For any $\varepsilon>0$, the covering of X consisting of all apheres of d-radius ε is uniform; and on such a sphere no f_α varies more than $4n\varepsilon$ (by a computation). Thus $\{f_\alpha\}$ is equiuniformly continuous.

It follows, of course, that a uniform covering can be realized by a mapping into a Euclidean space if its nerve is uniformly equivalent to a subspace of a Euclidean space. Let us call such a uniform complex a *Euclidean* complex, and such a covering a *Euclidean* covering.

Smirnov has defined [9] a "uniform complex" as a geometric complex K in a Euclidean space E^n such that the diameters of the simplexes of K are bounded above and the distances between pairs of disjoint simplexes of K are bounded away from zero. Because of the overlapping terminology, it should be observed that *an abstract complex K is Euclidean, as defined above, if and only if it can be embedded in some E^n as a uniform complex in the sense of Smirnov.* The proof of "if" is trivial; the converse is an exercise which we may omit, since it will follow from 1.8.

A covering u is *star-bounded*, of *density* at most n, if each element of u meets at most n other elements of u. (The term "star-bounded" is due to Mostow [8], "density" to Boltyanski [1].) Obviously a star-bounded covering is star-finite and finite-dimensional, but not conversely. A collection v of sets is said to be *discrete* relative to a covering u if no element of u meets two different elements of v. (Note that a subspace of μX is discrete in the induced uniformity if and only if it is a discrete collection of points relative to some covering in μ.) A covering u may be a finite union of collections, $u^1, u^2, \cdots$, each of which is discrete relative to u. Clearly such a covering is starbounded; conversely.

1.3. *Every star-bounded covering u is the union of finitely many subcollections each of which is discrete relative to u.*

Proof. In $u=\{U_\alpha\}$ let $\{U^1_\beta\}$ be a maximal subset such that no set U_α meets more than one U^1_β. Evidently $\{G^1_\beta\}$ is discrete relative to u. Now in $\{U_\alpha\}$, for each U_0, there are at most m sets U_α meeting U_0, and each of these meets at most $m-1$ more sets U_γ; let this family of $1+m+(m^2-m)$ or fewer sets be called F_0. Each F_α meets $\{U^1_\beta\}$, since otherwise U_α could be added to the supposedly maximal family. Having $u^1=\{U^1_\beta\}, u^2, \cdots, u^k$, let u^{k+1} be a maximal subset of $\{U_\alpha\}$ disjoint from $u^1, \cdots, u^k$, and such that no element of $\{U_\alpha\}$ meets more than one element of n^{k+1}. For each U_α which is not in $u^1, \cdots, u^k$, necessarily u^{k+1} meets F_α (as above). Therefore if U_α is not in $u^1, \cdots, u^{m^2}$, then F_α is exhausted and U_α is in u^{m^2+1}.

REMARK. The properties just shown to be equivalent are graph-theoretic, that is, they depend only on the 1-skeleton of the nerve of the covering.

Tukey has defined a *star-finite collection* of coverings as a collection, the union of any two of whose members is star-finite [9]. He proved (though he states less) that a uniform star-finite covering has a uniform star-refinement such that the union of the two coverings is star-finite, and hence by induction one has a normal sequence which is a star-finite collection [10, pp. 49–50]. Similarly we define a *star-bounded* collection of coverings as a collection, the union of any two of whose members is star-bounded; the corresponding result is given below (1.6).

A Euclidean covering is star-bounded, and more. Let us say that the covering u is *of polynomial growth* if there is a real polynomial P such that, for each $U \in u$, for all natural numbers k, the number of elements V of u such that there is a chain $U=U_0, U_1, \cdots, U_k=V$, all U_i in u, $U_i \cap U_{i+1}$ nonempty for all i, is bounded by $P(k)$.

1.4. *Every Euclidean covering is of polynomial growth.*

Proof. Suppose the nerve $N(u)$ is embedded in E^n by a uniform equivalence. Let d be the distance function in $N(u)$ and $e(x, y)$ the Euclidean distance between the images of x and y. There is $\varepsilon>0$ such that $d(x, y) \geqq 1$ implies $e(x, y) \geqq \varepsilon$; and there is $\delta>0$ such that $d(x, y) \leqq \delta$ implies $e(x, y) \leqq 1$. If x and y are vertices of $N(u)$ corresponding to members of u which are joined by a chain of length k, then $e(x, y) \leqq k/\delta$. Then for each vertex x, the set of all such y is a set of points whose mutual e-distances are all at least ε, packed in a Euclidean sphere of radius k/δ; hence their number is bounded by a polynomial in k.

Call a covering *linear* if its nerve is uniformly equivalent to a subspace of the real line R.

1.5. *A covering u is linear if and only if it can be indexed with integers, $u=\{U_i\}$, so that U_m meets U_n only if $|n-m| \leqq 1$. This is equivalent to the conditions that u is,*

(a) *countable,*

(b) *one-dimensional,*

(c) *acyclic,*

(d) *atriodic, that is, of density* 2 *or less, and*;

(e): (i) *the nerve of u does not contain three disjoint half-lines*;

(ii) *if the nerve contains a whole line then it is connected*;

(iii) *if the nerve contains two disjoint half-lines then it has only finitely many components.*

The proof is omitted. Note that connectedness implies (e).

By a standard argument (cf. 1.1 of [4]) we obtain

1.6. *Let the covering v be a star-refinement of u, that is, $v<^*u$. If u is*

(a) *star-bounded, or.*

(b) *of polynomial growth, or.*

(c) *linear, then there exists a covering w which also satisfies* (a), (b), *or* (c) *such that $v<w^*<u$. Further, $u\cup w$ is star-bounded; thus if u is a uniform star-bounded covering of a uniform space then there is a star-bounded normal sequence of uniform coverings u^n such that $u^1=u$.*

Proof. Let C be the set of all subsets γ of u such that there is at least one point common to all the members of γ. For each ordered pair (γ, δ) of elements of C, let $W_{\gamma\delta}$ be the union of all $V\in v$ such that the set of all elements of u which contain V is precisely γ, and the set of all elements of u which contain $St(V, v)$ is precisely δ. Let $w=\{W_{\gamma\delta}\}$.

Clearly $v<w$. For any nonempty $W_{\gamma\delta}$, δ is nonempty, and any V which meets $W_{\gamma\delta}$ is contained in every member of δ. Thus $St(W_{\gamma\delta}, w)\subset U$ for any member U of δ, and $w<^*u$.

If u is star-bounded of density m, then for each $W_{\gamma\delta}$ choose $U\in\delta$. No $W_{\alpha\beta}$ can meet $W_{\gamma\delta}$ unless every element of α and of β meets U; therefore there are at most 2^{2m} such $W_{\alpha\beta}$, and w is star-bounded. Clearly $u\cup w$ is star-bounded, and the last statement of the theorem follows by induction.

If the growth of u is bounded by a polynomial $P(n)$, then u is star-bounded of density $m\leqq P(1)$, and the growth of w is bounded by $2^{2m}p$. It may be of interest to note that this is a polynomial of the same degree as P.

Now suppose u is linear. We must modify the above covering $\{W_{\gamma\delta}\}$. Observe that if $W_{\gamma\delta}$ is not empty then each of γ and δ consists of one or two elements. If u is indexed as in 1.5, $u=\{U_n\}$, then there are four possibilities:

(a) $\gamma=\delta=\{n\}$, for some n;

(b) $\gamma=\{n, n+1\}$, $\delta=n$;

(c) $\gamma=\delta=\{n, n+1\}$;

(d) $\gamma=\{n, n+1\}$, $\delta=\{n+1\}$. For each n, replace the two sets described under (b) and (c) with their union. One readily verifies that the modified w is a linear covering satisfying $v<w<^*$u.

From 1.6 we may deduce that, for any uniformity μ, the set of all star-bounded coverings in μ forms a basis for a uniformity, say $b\mu$. The axioms on coarsening (i), intersection (ii), and interiors (iv) are obvious; star-refinement (iii) follows from 1.6, and the neighborhood basis axiom (v) from the fact that every finite covering is star-bounded.

Since the inverse image of a star-bounded covering, under any function, is star-bounded, therefore when f: $\mu X \to \gamma Y$ is uniformly continuous, f: $b\mu X \to b\gamma Y$ is also uniformly continuous. We summarize this (as in [4]) in the slightly elliptical statement that b is a *functor*. All this is true also for coverings of polynomial growth. However, linear coverings do not in general suffice, for the set of all linear coverings in μ is not closed under finite intersection. The finite intersections of linear coverings in μ do form a basis for a uniformity, which is the familiar uniformity $c\mu$ induced by real-valued uniformly continuous functions. To see this it suffices to observe that, by 1.2 and 1.1, to every linear uniform covering u one may asssciate a mapping into $N(u) \subset R$ which realizes u.

1.7. F*or any uniform space μX, the star-bounded coverings in μ, as well as those of polynomial growth, form a basis for a uniformity consistent with the topology. Both of these transformations are functors. The weak uniformity $c\mu$ induced by the real-valued uniformly continuous functions on μX has a basis consisting of all the Euclidean coverings in μ, and a sub-basis consisting of all the linear coverings in μ.*

The proof that Euclidean coverings from a basis for $c\mu$ is again by 1.2 and 1.1. Whether any purely combinatorial result such as 1.6 is valid for Euclidean coverings is not known. (Of course 1.6 applies if it is true that every countable covering of polynomial growth is Euclidean.)

Let mE^n denote Euclidean n-space, mR the line, in the usual uniformity. Note that mE^n is the product of n copies of mR. Beyond this we may omit the "m" for the present, since no other uniformities on these spaces are being considered.

1.8. THEOREM. *A necessary and sufficient condition that a uniform complex X be Euclidean is that the vertices of X may be identified with a set of points in some E^n, any two of which are at distance greater than* 1, *so that the distances between pairs of vertices which are joined by an edge* ($1-$*simplex*) *of X are bounded. In fact, this is the necessary and sufficient condition that there exist a uniform equivalence φ of X into the product of E^n and a cell of some dimension*; *and φ may be taken to be semilinear.*

Proof. The necessity (both statements) is evident. Suppose conversely that f maps the vertices α of X into E^n, with the distance from $f(\alpha)$ to $f(\beta)$ greater than 1 for all $\alpha \neq \beta$, and less than M when α and β are joined by an edge. For any $x=(x_\alpha)$ in X, define $(\varphi_1(x), \cdots, \varphi_n(x)) = g(x) = \sum x_\alpha f(\alpha) \in E^n$. Evidently g is uniformly continuous.

Let C_α be the sphere of radius $2M+1$ about $g(\alpha)$; let K_α be the least subcomplex of X which contains $g^{-1}(C_\alpha)$. The vertices of K_α are mapped by f into points of distance 1 or more from each other in a sphere of radius $3M+1$, and hence their number has a bound $q+1$. Then each K_α may be embedded by an isometry k_α in the abstract q-dimensional simplex; embedding the simplex in a cell in E^q, we obtain mappings $h_\alpha: K_\alpha \to E^q$ which are semilinear uniform equivalences, having a common modulus of continuity, and such that the mappings h_α^{-1} have a common modulus of continuity. Define an extension i_α of h_α over X as follows: every x in X can be expressed uniquely as a convex combination $ty+(1-t)z$, where y is in the subcomplex K_α and z has coordinate $z_\beta=0$ for all β in K_α; let $i_\alpha(x)=th_\alpha(y)$. Then $\{i_\alpha\}$ is an equiuniformly continuous family of semilinear mappings. Further, there is a cell in E^q which contains all their ranges.

Finally, $\{g(K_\alpha)\}$ is a star-bounded covering of $g(X)$, and thus, by 1.3, it is a union of subcollections $v^1, \cdots, v^s$, each discrete relative to the whole covering. For $j=1, \cdots, s$, let $d_j = \sum [i_\alpha \mid g(K_\alpha) \in v^j]$. Observe that on each star $St(\beta)$ in X, d_j coincides with one i_α (that is, at most one fails to vanish; for $St(\beta) \subset K_\beta$, and $g(K_\beta)$ meets at most one $g(K_\alpha)$ in v^j). Therefore $d_j: X \to E^q$ is uniformly continuous. The definition of $\varphi: X \to E^{n+qs}$ is completed by putting $(\varphi_k(x))$, $n+(j-1)q+1 \leqq k \leqq n+jq$, equal to the vector $d_j(x)$, for each $j=1, \cdots, s$.

We have a uniformly continuous semilinear mapping φ of X into the product of E^n and a qs-dimensional cell. Uniform continuity of φ^{-1} means that for each $\varepsilon>0$ there is $\delta>0$ such that two points at distance $>\varepsilon$ in X are mapped by φ into points at distance $>\delta$. For any two points, x, y, in X, either g maps them into points at distance >1 (and so does φ), or they lie in a common K_α. But then some d_j coincides on K_α with the embedding h_α. Thus φ is a uniform equivalence.

1.9. COROLLARY. *If the* 1-*skelton of a uniform complex X is Euclidean then X is Euclidean.*

1.10. THEOREM. *The following conditions on a uniform complex X are equivalent.*

(a) *X is a countable, star-finite, finite-dimensional complex.*

(b) *X is a locally compact, σ-compact, finite-dimensional space.*

(c) *X is homeomorphic with a closed subset of a Euclidean space.*

(d) *There is a distance-increasing homeomorphism of X into a Euclidean space.*

(e) *There is a uniformly continuous homeomorphism of X upon a closed subset of a Euclidean space.*

Proof. The implications (e)$\Rightarrow$(c), (d)$\Rightarrow$(c), and (c)$\Rightarrow$(b)$\Rightarrow$(a), are evident. From the hypothesis (a) that X is a countable star-finite uniform complex of dimension n, we shall construct the mappings of X into E^{2n+3} required for (d) and (e). Clearly it suffices to construct such a mapping into E^{2n+2} for each component of X. Let Y be a component of X, and let φ be a one-to-one semilinear mapping of Y into E^{2n+1}. (To construct φ it suffices to map the vertices of Y upon a set of points in general position in E^{2n+1}.) Since Y is star-finite, φ is continuous.

Choose a vertex Y_0 of the complex Y and let f_0 be a one-to-one semilinear mapping of Y into E^{2n+2} which sends Y_0 to the origin and all of Y into the hyperplane $x_1=0$. Let Y_1 be the subcomplex which is the closure of the star of Y_0; inductively let Y_{k+1} be the subcomplex spanned by Y_k and the vertices which are joined to Y_k by 1-cells of Y. Since Y is star-finite, each Y_k is a finite complex; and since Y is connected, the union of all Y_k is Y. Let Z_k be the span of the vertices not in Y_k. For each k, each point p of Y can be written uniquely (in barycentric coordinates) as $\lambda p_1+(1-\lambda)p_2$, with λ and $1-\lambda$ nonnegative, p_1 in Y_k, p_2 in Z_k. Inductively, let f_k be a piecewise linear one-to-one mapping of Y into E^{2n+2}, sending Y into the halfspace $x_1\leqq c_k$ and Z_{k-1} into the hyperplane $x_1=c_k$ and increasing distances in Y_k. Write $f_k(p)=g(p)+h(p)$, where $g(p)$ is the projection of $f_k(p)$ on the x_1-axis, $h(p)$ the projection on $x_1=0$. For p in Z_k, $f_{k+1}(p)$ is to be $\alpha g(p)+\beta h(p)$, where α and β are large constants to be determined. For p in Y_k, $f_{k+1}(p)=f_k(p)$; and for general $p=\lambda p_1+(1-\lambda)p_2$ (as above), $f_{k+1}(p)$ must be $\lambda f_{k+1}(p_1)+(1-\lambda)f_{k+1}(p_2)$. On Z_k, g is constant, and h is one-to-one, piecewise linear, and continuous. The common part of Z_k and Y_{k+1} is a finite complex, and hence there exists β so large that βh increases distances on this complex. Similarly, if α and β are large enough, f_{k+1} will increase distances on Y_{k+1}, and the induction runs. Finally we have a sequence (f_k) of continuous mappings of Y into E^{2n+2}, converging locally uniformly to a limit ψ. Then ψ is continuous; and ψ increases distances, which implies that ψ^{-1} is continuous. Thus (a) implies (d).

Since each Y_k is compact, one can go back and modify the constants α and β at each step so as to end with a uniformly continuous homeomorphism g upon an image which is not necessarily a closed set. Define a real-valued function h on Y as follows. For the distinguished vertex Y_0, $h(Y_0)=0$. For any other point y there is just one k such that y is in Y_{k+1} but not in y_k; and there is a unique relation $y=\lambda p_1+(1-\lambda)p_2$, $p_1\in Y_k$, $p_2\in Z_k$. Let $h(y)=k+1-\lambda$. Evidently h is uniformly continuous. Let $h'(y)$ be the point in E^{2n+2} whose first coordinate is $h(y)$, with all other coordinates zero; then $g+h'$ is a uniformly continuous homeomorphism upon a closed set. This completes the proof.

The complexes satisfying (*d*) (in slightly different words) are called

Lebesgue complexes by Smirnov. [9] Evidently in any fixed E^n, (d) and (e) are not equivalent (if $n>1$).

2. **Bases.** This section is primarily a discussion of the subspaces of the line mR, including a characterization; it concludes with a formulation of the same approach to subspaces of mE^n.

Let us first suppose given the topological space R, and characterize m among its uniformities. Evidently m is

(a) metric, that is, it has a countable basis of coverings. It has

(b) a star-bounded basis, and it is

(c) uniformly locally connected, that is, there is a basis of coverings whose elements are connected sets. We shall see that these properties are shared by m only with the uniformities induced by metrizing R as $(0, 1)$ or as a half-infinite interval; thus m can be characterized by adding the condition (d): the space is complete.

These are evidently not the conditions to apply to subspaces of mR, (c) being invalid. We shall have to replace (c) with some sort of conditions on the nerves of the coverings. It is not enough to say (c′) there is a basis of linear coverings, even on the topological space R. This is shown by the following subspace of mE^2. Take the half-line consisting of all points $(x, 0)$, $x\leqq 3$, and for $n=3, 4, \cdots$, take the four line segments running successively from $(n, 0)$ to $(n+1-3/n, 1)$ to $(n+1-2/n, 0)$ to $(n+1-1/n, 1)$ to $(n+1, 0)$. A sketch shows that this metric space satisfies conditions (a), (b), (c′), and (d), but not (c); it is homeomorphic but not uniformly equivalent to mR.

We have indicated some uniformities on R satisfying (a), (b), and (c), but not (d). For (a), (c), and (d), consider the following distance function f. For notational convenience let e indicate $e(x, y)=|x-y|$; let $\min(x, y)=m$. If $e\geqq 1$, or if $m\leqq 1$, then $f(x, y)=e$; otherwise $f(x, y)=e^{1/m}$. Finally, to construct a nonmetric uniformity on R satisfying (b), (c), and (d), let (a_n) designate a (variable) sequence of positive numbers coverging to zero. For each natural number m, define the covering $u(m, (a_n))$ to consist of the following intervals.

(1) For every integer t such that neither $t-1, t$, nor $t+1$ is a positive integral multiple of m, the interval $(t-1/m, t+1/m)$.

(2) For each positive integer n, the intervals $(n+a_n, n+2/m)$ and $(n-2/m, n-a_n)$.

(3) For $-m\leqq t\leqq m$, and for all n, the intervals $(n+(t-1/m)a_n, n+(t+1/m)a_n)$. Consider the collection of all $u(m, (a_n))$ such that $m\geqq 4$ and $a_n<1/m+1$ for all n. One readily verifies that this collection is a basis of a uniformity having the required properties. One may note also that all the above examples have bases consisting of linear coverings.

2.1. *Every uniformly locally connected metric space which is homeomorphic to the real line R and has a star-bounded basis of uniform coverings is uniformly equivalent either to mR or to an open interval of mR.*

Proof. We may call the space μR; it is required to construct a uniform equivalence of μR into mR. We are given a countable basis $\{u^n\}$ for μ, a star-bounded basis $\{v^\alpha\}$, and also a basis consisting of coverings with connected elements. The interiors of these connected sets are open intervals, still forming uniform coverings, which still constitute a basis $\{w^\beta\}$. If for each n we choose w^n refining u^n, we have a countable basis of coverings with open intervals. Evidently we may suppose each w^{n+1} is a star-refinement of w^n (since some w^{n+k} is), and we may suppose $w^1 < v^\alpha$ for some α. Next we interpolate linear coverings z^n consisting of open intervals, $w^{n+1} < z^n < w^n$, as follows. Choose a point p and let Z_0^n be $St(p, w^{n+1})$. Evidently Z_0^n is an open interval (p_1, q_1). Having points p_k and q_k, define Z_k^n as $St(q_k, w^{n+1})$ and Z_{-k}^n as $St(p_k, w^{n+1})$. At some stage an improper interval may be obtained, so that p_k or q_k does not exist; in that case omit so much of the construction as involves the missing points. Evidently the union of all Z_k^n, n fixed, $k = 0, \pm 1, \cdots$, is a nonempty open and closed subset of R, hence all of R. Since every intervel in $w^{n+}{}_{\text{I}}$ contains at most one of the points $p, p_k, q_k, w^{n+1} < z^n = \{Z_k^n\}$; and since $w^{n+1} <^* w^n$, hence $z^n < w^n$. Clearly z^n is linear.

To see that $\{z^n\}$ is a star-bounded family, consider any $m < n$. Each element of z^n meets at most two other elements of z^n and at most three elements of z^m. Thus if $z^m \cup z^n$ is not star-bounded then there exist sets Z in z^m meeting arbitrarily many elements of z^n. A fortiori there exist sets V in v^α meeting arbitrarily many elements of z^n. Since z^n is linear, one can find for each positive integer r a set $V \in v^\alpha$ countaining r points, no two of which lie in a common element of z^n. But there is a covering $v^\beta < z^n$ such that $v^\alpha \cup v^\beta$ is star-bounded, since the v's form a star-bounded basis. The contradiction establishes that $z^m \cup z^n$ is star-bounded; and the famıly $\{z^n\}$ is a countable star-bounded basis consisting of linear coverings each of which consists of open intervals.

Now index the elements of z^n with rational numbers s, $z^n = \{Z_s^n\}$, as follows. For $n=1$, the values of s are the integers k assigned above; thus Z_s^n does not meet Z_t^n if $|s-t| > 1$. Having indexed z^n, consider each Z_s^n. There is a next rational number $t > s$ such that some element of z^n is called Z_t^n, except possibly for one (greatest) value of s; if there is such en exceptional s, assign to it the value $t = s + 2^{-n}$. There are finitely many elements Z of z^{n+1} such that s is the least index such that $Z \subset Z_s^n$; and the number of them, $h(s)$ is a bounded function of s (n fixed). Furthermore, exactly one of them meets an element of z^{n+1}

which is contained in the next Z_q^n before Z_s^n (with a possible exception if there is no such q) and exactly one meets an element of z^{n+1} which is contained in Z_t^n but not in Z_s^n. Index these elements of z^{n+1} in order from Z_q^n toward Z_t^n as Z_i^{n+1}, for $i=s, s+(t-s/h(s)), \cdots, t-(t-s)/h(s)$ (equal steps). This completes the indexing. Then routine computation shows that for each point x in μR, the numbers $g_n(x)=\max [s|x \in Z_s^n]$ converge to a limit $g(x)$, and that g is a uniformly continuous function realizing all of the coverings z^n. Since $\{z^n\}$ is a basis, g is one-to-one and g is a uniform equivalence.

If one tries to carry out the construction of 2.1 on the example given previously of a complete metric space homeomorphic to R having a star-bounded basis of linear uniform coverings, it breaks down because ultimately z^{n+1} must be "crooked" in z^n. It is not crooked in the strong sense familiar from the construction of the pseudo-arc; indeed, with a suitable choice, one can arrange that near any point in the space almost all z^{n+1} are "straight" in z^n. Up to some critical value N the chains z^n follow an approximating smooth path; then z^{N+1} and all subsequent z^n follow the kinds in the curve. This means that we must impose a very strong straightness condition in order to characterize the subspaces of mR. Let us use the term *chain in u* for a subset of a covering u whose elements correspond to the vertices of a chain of edges in $N(u)$.

2.2. *The following conditions on a uniform space μX are necessary and sufficient in order that μX should be uniformly equivalent to a subspace of mR.*

(a) *μ has a basis which is a countable sequence of linear coverings u^n, with $u^{n+1}<u^n$, such that (1) if $(U_1, \cdots, U_p)$ is a chain in u^{n+1}, with U_1 and U_p both contained in one element U of u^n, then all U_i are contained in U; and (2) if $(U_1, \cdots, U_p)$ and $(V_1, \cdots, V_q)$ are two chains in u^{n+1} having no common elements, some element U of u^n contains both U_1 and V_1, and some element V of u^n contains both U_p and V_q, then U meets V.*

(b) *μ has a star-bounded basis.*

The necessity of the conditions is obvious, and the proof of sufficiency is an easy modification of 2.1. However, the proof as given above does not look ready to be generalized to E^n. We conclude this section with some easily proved remarks outlining another version which might have brighter prospects.

First, it suffices to work with the completion. Second, if a complete uniform space has a countable basis consisting of finite-dimensional coverings, (a) there is a natural inverse system of semilinear mappings on the nerves of these coverings, and (b) the space is the inverse limit of this system. I have in mind the mappings defined, for a sequence $\{u^n\}, u^{n+1}<^*u^n$, as follows. Since u^n is finite-dimensional, each element

U_α of u^{n+1} is contained in only finitely many elements of u^n, and the corresponding vertices in $N(u^n)$ span a simplex; the vertex α of $N(u^{n+1})$ can be taken to the center of gravity of that simplex, in a uniformly continuous semilinear mapping. Third, if all the nerves can be embedded in one complete space in such a way that the mappings $N(u^{n+1}) \rightarrow N(u^n)$ move no point more than ε_n, where $\varepsilon_n \rightarrow 0$, then of course the inverse limit space is embedded in the same containing space. This is clearly possible under the hypotheses of 2.2.

3. **The weak derivative.** In this section we describe an operation on uniformities which generalizes the passage from the usual uniformity m on a Euclidean space to the finest uniformity a. It is not known whether this operation is applicable to general uniformities[1]; the main results of this section apply only to *weak uniformities* induced by families of real-valued functions.

For any weak uniformity μ on a space X, we define the *weak derivative* $w\mu$ of μ as the family of all coverings of X which have a refinement of the form $\{U^\alpha \cap V_i^\alpha\}$, where $\{U^\alpha\}$ is a covering in μ and the families $V^\alpha = \{V_i^\alpha\}$, for each α, are finite coverings in μ of bounded dimension. (This is a modification of an operation called the *derivative* in [4]. We might as well have required v^α only to cover the subspace U_α; the equivalence follows from the simple proposition 3.6 below.) If we recall that since μ is a weak uniformity, the covering $\{U_\alpha\}$ may be supposed Euclidean, we see that the typical covering $\{U_\alpha \cap V_i^\alpha\}$ is (1) uniformly locally uniform (on μX), (2) uniformly locally finite, and (3) finite-dimensional.

The proof that $w\mu$ is a uniformity will be a demonstration that $w\mu$ is the weak uniformity induced by a certain family of functions. Let $C(\mu X)$ denote the family of all real-valued uniformly continuous functions on μX (uniformly continuous into mR). The term *composition* will be used with the specific meaning of a functional composition $g(f_1 \cdots, f_n)$, where $f_1, \cdots, f_n$, are in $C(\mu X)$ and g is any continuous real-valued function on E^n. In particular, the family of all such functions on X to R is the *closure*, under composition, of $C(\mu X)$. (Cf. [5].)

3.1. *For each Euclidean space E^n, the weak derivative of the usual metric uniformity, m, is the finest uniformity consistent with the topology*;

---

1. Specifically, applying the definition of $w\mu$ in the next paragraph to a general uniformity μ, it is not known wheter $w\mu$ is always a uniformity in the present sense. The referee points out that it is certainly a regular uniformity in the sense of Morita and [7]; and there is a non-trivial theory of such structures. In that theory, the referee observes, 3.5 is valid without restriction on μ.

that is, the uniformity a defined by all open coverings.

Proof. Evidently any covering in wm has an open refinement. Conversely, for any open covering $\{W_\gamma\}$ of E^n, consider a uniform covering $\{U_\alpha\}$ consisting of closed metric spheres. Since each U_α is a compact space, there is a finite uniform covering $\{G_i^\alpha\}$ of U_α refining the open covering $\{U_\alpha \cap W_\gamma | \text{all } \gamma\}$. We may take $\{G_i^\alpha\}$ n-dimensional. Let p_α be the center of the sphere U_α, and for each G_i^α meeting the boundary B of U_α, let V_i^α consist of G_i^α together with all points q outside U_α such that the intersection of the segment pq with B is a point of G_i^α; otherwise let $V_i^\alpha = G_i^\alpha$. Evidently $\{V_i^\alpha\}$ is a uniform finite n-dimensional covering of mE^n, and $\{U_\alpha \cap V_i^\alpha\}$ is a refinement of $\{W_\gamma\}$. Thus $wmE^n = aE^n$.

3.2. *For any open covering u of a Euclidean space E^n, there is a homeomorphism of E^n onto itself which takes u onto a uniform covering of mE^n.*

This is obvious.

3.3. THEOREM. *The weak derivative of a weak uniformity is a weak uniformity. Specifically, if μ is weak, then $w\mu$ is the weak uniformity induced by the closure under composition of $C(\mu X)$.*

Proof. If $f_1, \cdots, f_n$, are in $C(\mu X)$, $g: E^n \to R$ is continuous, and u is any uniform (even any open) covering of R, then the inverse image of u under g is uniform in $aE^n = wmE^n$ and hence the inverse image of u under $g(f_1, \cdots, f_n)$ is in the family of coverings $w\mu$. Evidently $w\mu$ is closed under intersection; therefore $w\mu$ contains the weak uniformity induced by the closure under composition of $C(\mu X)$.

Conversely, since μ is weak, each covering $\{U_\alpha \cap V_i^\alpha\}$ in $w\mu$ may be refined by a covering $\{U_j \cap V_i^j\}$ so that the following is true. There is a uniformly continuous function f: $\mu X \to mE^n$ realizing $\{U_j\}$. Each $v^j = \{V_i^j\}$ is finite and at most k-dimensional and is realized by a bounded uniformly continuous function g_j: $\mu X \to mE^q$ (here $q = 2k+1$). Also, $\{U_j\}$ is star-bounded and can be written as the union of p relatively discrete subcollections u^r; and finally, $\{U_j\}$ is countable. We shall construct a mapping h of μX into $(n+pq)$–space.

Choose positive numbers c_j such that $c_j |g_j(x)| < 2^{-j}$ for all x. For each j, let d_j be a uniformly continuous real-valued function on μX with values in $[0, c_j]$, vanishing outside the star of U_j and having the constant value c_j on U_j. For each x, define the first n coordinates of $h(x)$ to be the coordinates of $f(x)$. Let the q coordinates of $h(x)$ from the $(n+q(r-1)+1)$th through the $(n+qr)$th be $\sum [d_j(x)g_j(x) | U_j \in n^r]$.

Since the series $\sum c_j \|g_j\|$ converges (absolutely), $h: \mu X \to mE^s (s=n+pq)$ is uniformly continuous. It is clear from the construction that $h: X \to aE^s$ realizes $\{U_j \cap V_i^j\} = z$, that is, z is refined by the inverse image of some open covering. By 3.2, there is a continuous function $T: E^s \to E^s$ such that $T(h): X \to mE^s$ realizes z. But each coordinate of $T(h)$ is a composition of a continuous coordinate projection of T with the uniformly continuous coordinates of h, and the proof is complete.

From an approximation theorem proved in [5, Theorem 1.7] we have

3.4. COROLLARY. *$C(w\mu X)$ contains all the compositions $g(f_1, \cdots, f_n)$, g continuous and f_i in $C(\mu X)$, and consists of all uniform limits of such compositions.*

In [5] there is an example of a family of functions A such that the uniform closure of the closure under composition of A is not itself closed under composition. That example A is not $C(\mu X)$ for any μ, but this is inessential. We describe an example of a uniform space μX such that μ is weak and $w(w\mu) \neq w\mu$, omitting the details of the verification.

Example. Let X be the set of all ordered triples (i, j, k) of positive integers, with the discrete topology. Let μ be the set of all coverings u of X such that (1) for some n', for each $n > n'$, there is an element U_n of u which contains all (n, j, k); and (2) for each n $(\leqq n')$, for some m', for each $m > m'$ there is an element U_{nm} of u which contains all $(n, m, k,)$. Observe that μ has a basis consisting of discrete coverings; thus μ is weak, and $w\mu$ and $ww\mu$ can be computed without worrying about dimension. One may verify that a covering u is in $w\mu$ if and only if (a) for each n there is $m' = m'(n)$ such that for each $m > m'$ there are finitely many elements of u whose union contains all (n, m, k), and (b) for some n', for each $n > n'$, (i) there are finitely many elements of u whose union contains all (n, j, k), and (ii) for each $m > m'(n)$ all points (n, m, k) are in one element of u. Then $ww\mu$ is determined by the conditions (a) and (b), (i); in particular, $ww\mu \neq w\mu$.

Powers of w are defined by $w^{\alpha+1} = ww^\alpha$; for limit ordinals α, $w^\alpha\mu$ is the union of the increasing sequence of families of coverings $w^\beta\mu$, $\beta < \alpha$. Since the uniformities $w^\alpha\mu$ are successively finer, there must be an α such that $w^{\alpha+1}\mu = w^\alpha\mu$. (By 3.4, the first uncountable ordinal is such an α.)

3.5. *Applied to uniform spaces with weak uniformities, the weak derivative and all its powers are functors commuting with completion.*

Proof. If $f: \mu X \to \nu Y$ is uniformly continuous then, since f^{-1} preserves finiteness and dimension of coverings, $f: w\mu X \to w\nu Y$ is uniformly continuous; thus w is a functor. If F is a Cauchy filter in μX and $\{U_\alpha \cap V_i^\alpha\}$ a typical covering in $w\mu$, then F contains some U_α and, for that α, some V_i^α; being a filter, F contains $U_\alpha \cap V_i^\alpha$. Thus the same filters are Cauchy in μ and in $w\mu$, and the completions $\pi\mu X$ and $\pi w\mu X$ have the same points. Obviously every covering in $w\pi\mu$ is in $\pi w\mu$; the converse is a routine application of Morita's demonstration [7; Lemma 7, Th. 3, Th. 9] that every uniform covering $\{V_\beta\}$ of μX can be extended to a uniform covering $\{V_\beta^*\}$ of $\pi\mu X$ such that $V_\beta = V_\beta^* \cap X$ and the correspondence $V_\beta \longleftrightarrow V_\beta^*$ preserves the nerve. Thus $\pi w = w\pi$. Therefore if w^α is a functor commuting with π, so is $w^{\alpha+1}$. The proof is completed by the observation that every covering in w^α, for α a limit ordinal, is already in some w^β for $\beta<\alpha$.

The next four propositions amount to a closer analysis of the theorem of [5] that if $C(\mu X)$ is closed under composition then for any subspace Y of X, in the induced uniformity μ^*, $C(\mu^* Y)$ contains only the restrictions of function in $C(\mu X)$.

3.6. *Let $\mu^* Y$ be a subspace of μX, and $\{U_i\}$ a finite uniform covering of $\mu^* Y$ of dimension k. There is a finite uniform covering $\{V_j\}$ of μX, of dimension $2k+1$ or less, such that $\{V_j \cap Y\}$ is a refinement of $\{U_i\}$.*

3.7. *The weak derivative preserves subspaces; that is, if $\mu^* Y$ is a subspace of μX (and μ is weak) then the uniformity induced on Y by $w\mu$ is $w\mu^*$.*

3.8. *If f is a real-valued function on X, μ a weak uniformity on X, and $\{U_\alpha\}$ a uniform covering of μX such that on each U_α, f is bounded and uniformly continuous, then f is uniformly continuous on $w\mu X$.*

3.9. *If $\mu^* Y$ is a subspace of μX (μ a weak uniformity) and f a uniformly continuous real-valued function on $\mu^* Y$, then f has an extension in $C(w\mu X)$.*

Proof of 3.6. This is a corollary of a theorem of Katětov [6]: every bounded real-valued uniformly continuous function on a subspace of any uniform space has a bounded uniformly continuous extension over the whole space. If $\{U_i\}$ is a finite k-dimensional covering of $\mu^* Y \subset \mu X$, then $\{U_i\}$ cen be realized by a mapping into a compact subset of E^{2k+1}; each coordinate can be extended, by Katětov's theorem, and the conclusion follows.

Proof of 3.7. If $\mu^* Y$ is a subspace of μX and $\{U_\alpha \cap V_i^\alpha\}$ a typical covering in $w\mu$, then $\{U_\alpha \cap Y\}$ is in μ^*, the coverings $y^\alpha = \{V_i^\alpha \cap Y\}$ are finite coverings in μ^* of bounded dimension, and hence $\{U_\alpha \cap V_i^\alpha \cap Y\}$ is in $w\mu^*$. The converse is clear in the light of 3.6.

Proof of 3.8. If f is bounded and uniformly continuous on each element of the uniform covering $\{U_\alpha\}$, then the inverse image of any uniform covering of mR is refined by a covering $\{U_\alpha \cap V_i^\alpha\}$, where for each α, $\{V_i^\alpha\}$ is a uniform finite 1-dimensional covering of the subspace U_α. By 3.6, each $\{V_i^\alpha\}$ may be extended to a uniform finite 3-dimensional covering of μX, and hence f is uniformly continuous on $w\mu X$. (Actually, by the method of 3.6, these coverings $\{V_i^\alpha\}$ may be extended so as to remain 1-dimensional.)

We may note that the hypothesis that μ is weak was not needed for these proofs; thus if w can be satisfactorily interpreted for more general spaces, 3.7 and 3.8 will carry over. (Cf. the footnote.[1]) The hypothesis will be used for 3.9, though one could avoid it by a use of results of [4]. It should be noted that the proof of 3.9 is almost the same as the proof of a similar extension theorem in [4].

Proof of 3.9. Note first (*) that a function h which is defined on a uniform space ρA into a uniform space σB, and uniformly continuous on each of a finite family of subspaces of ρA which make up a uniform covering, is uniformly continuous on ρA. Now consider the given hypothesis, f: $\mu^* Y \rightarrow mR$ uniformly continuous, $\mu^* Y$ a subspace of μX. Let $V_n = f^{-1}((n-1, n+1))$ in Y, and let $U_n = V_n \cup (X - Y)$. Since $\{V_n\}$ is in μ^*, therefore $\{U_n\}$ is in μ. Since μ is weak, $\{U_n\}$ has a countable uniform star-refinement $\{W_i\} = w$.

The function f is defined, in particular, on the subspace $Y \cap St(W_1, w)$ of the space $St(W_1, w) \cap (Y \cup W_1)$. On that subspace f is uniformly continuous and, since $St(W_1, w) \subset U_n$ for some n, f is bounded there. By Katětov's theorem [6] there is a bounded uniformly continuous function g_1 on $St(W_1, w) \cap (Y \cup W_1)$ to mR, such that g_1 and f agree on their common domain $Y \cap St(W_1, w)$. Therefore, by (*), the function f_1 on $Y \cup W_1$ whose values are those of f and of g_1 is uniformity induced by μ).

Having extended f to f_n, defined on the union of Y and $W_1, \cdots, W_n$, uniformly continuous there, and bounded on each W_{ni}, one constructs by the same argument an extension f_{n+1} which is defined on W_{n+1} also. By induction one has a well-defined function $\bar{f}$ extending f over all of X. On each W_i, $\bar{f}$ agrees with f_i and thus is bounded and uniformly continuous. By 3.8, $\bar{f}$ is uniformly continuous on $w\mu X$.

The next result is also based on a similar theorem in [4]. Let us quote a lemma [4, proposition 2.3]: every uniform space which is not precompact has an infinite uniformly discrete subspace.

3.10. *For a metric space μX, μ can be finer than the weak derivative of some weak uniformity ν on X only if (1) the set C of all non-isolated points of X forms a precompact subspace of μX, and (2) for any complete subset S of $X-C$, the distances of different points of S are bounded away from zero. Unless X has uncountably many isolated points, these conditions imply that μ is a weak uniformity and $\mu = w\mu$.*

Proof. First suppose that μX satisfies (1) and (2) and has only countably many isolated points. Then every uniform covering has a uniform refinement which consists of a finite covering of an ε-neighborhood of C and a countable discrete covering of the rest of X; thus μ is a weak uniformity. Consider the completion $\pi\mu X$ of μX. If $\pi\mu$ is not the finest uniformity consistent with the topology, then there is a non-uniform open covering $\{U_\alpha\}$. This means that there is a sequence of points z_n such that for each n, no U_α contains the sphere of radius 2^{-n} about z_n. Since X is a dense subspace, we may choose x_n in X within distance 2^{-n} of z_α, so that no U_α contains the 2^{1-n}-sphere about x_n. Since $\{U_\alpha\}$ is an open covering, the sequence (x_n) can have no accumulation point in $\pi\mu X$. Since C is precompact, it is not possible that infinitely many x_n are in C. Then we may choose a subsequence—to simplify notation, suppose it is the whole sequence—so that $\{x_n\}$ is an infinite subset of $X-C$, which is closed in $\pi\mu X$ and thus complete, but such that no U_α contains the 2^{1-n}-sphere about x_n. This means that we can choose y_n in X within distance 2^{1-n} of x_n, so that no U_α contains the 2^{2-n}-sphere about y_n. It is therefore impossible (as before) that infinitely many y_n are in C. But now we have a complete subset of $X-C$, consisting of all the x_n and all but finitely many y_n, in which distances are not bounded from zero. The contradiction proves the untenability of the hypothesis that $\pi\mu$ is not the finest uniformity consistent with the topology of $\pi\mu X$. It follows that $w\pi\mu = \pi\mu$, and since w preserves subspaces, $w\mu = \mu$.

Suppose next that μ is finer than $w\nu$ for some ν, but C is not precompact in the uniformity induced by μ. Since w preserves subspaces, it is clear that C is not precompact in νX either. Therefore C has an infinite uniformly discrete (in νX) subspace, by the proposition 2.3 of [4] which was pointed out above. This means there are an infinite subset $\{x_i\}$ of C and a covering u in ν such that the sets $St(x_i, u)$ are disjoint. Choose $v <^* u$ in ν, so that the sets $S_i = St(x_i, v)$ form a uniformly discrete collection. Choose points z_i in S_i, $z_i \neq x_i$, such that for some metric d inducing the uniformity μ, $d(z_i, x_i)$ converges to zero. For each i, there is a bounded uniformly continuous real-valued function f_i on νX such that $f_i(x_i)=0, f_i(z_i)=1$, and on $X-S_i, f_i$ is identically 1. Define the real-valued function g on X as follows: if x is in some S_i, $g(x) = f_i(x)$; if x is in no S_i, $g(x)=1$. Then on every element of v, g coincides with some f_i; therefore by 3.8, g is uniformly continuous on $w\nu X$. Supposedly $w\nu$ is coarser than the metric uniformity μ; but since (dz_i, x_i) converges to zero while $g(z_i)-g(x_i)=1$, this is absurd. There remains the possibility that $\mu \supset w\nu$ and X contains a complete set of isolated points S which contains a sequence of pairs (x_i, y_i) such that $d(x_i, y_i)$ converges to zero; but evidently the above argument can be repeated in the set consisting of all x_i and y_i, to lead to the same contradiction.

A corollary of 3.10 is that in the sequence of uniformities $\mu, w\mu, w^2\mu, \cdots$, at most two can be metric, the first and the last.

3.11. *If f: $\mu X \rightarrow mE^n$ is a uniformly continuous homeomorphism upon a closed set, and μ is weak, then f: $w\mu X \rightarrow aE^n$ is a uniform equivalence.*

Proof. We have that f: $w\mu X \rightarrow aE^n$ is uniformly continuous. Let νY be the image of X, regarded as a subspace of aE^n. Since f is a homeomorphism, every open covering of X is the inverse image of an open covering of Y; since Y is closed in E^n, every open covering of Y is in ν. Therefore every open covering of X (a fortiori, every uniform covering of $w\mu X$) is realized by f: $w\mu X \rightarrow \nu Y$. Thus f is a uniform equivalence.

3.12. THEOREM. *A uniform space μX is uniformly equivalent to a [closed] subspace of aE^n, for some n, if and only if (1) μ is a weak uniformity, (2) $\mu = w\mu$, and (3) μX has a uniform covering whose elements are precompact [compact] metric speces which have finite-dimensional completions of bounded dimension.*

Proof. Since w preserves subspaces, the necessity of the conditions is evident. Since w commutes with completion, it suffices to prove the sufficiency in case μX is complete; and by 3.11, it suffices to construct a uniformly continuous homeomorphism upon a closed set. We may replace the covering given by (3) with a Euclidean refinement $u=\{U_i\}$, realized by a uniformly continuous mapping f: $\mu X \rightarrow mE^n$, and partitioned into p discrete subcollections u^r. Each U_i is a compact k-dimensional metric space and hence is homeomorphic to a bounded subset of E^q, $q=2k+1$. It remains to build a mapping into $(n+pq)-$space, as in 3.3.

Let e_i be a homeomorphism of U_i into E^q, g_i a bounded uniformly continuous extension of e_i over μX. Let d_i be a uniformly continuous function on μX with values in $[0, 1]$, vanishing outside the star of U_i and identically 1 on U_i. For each x, let the first n coordinates of $h(x)$ be those of $f(x)$; let the $(n+q(r-1)+1)$th through $(n+qr)$th coordinates form the vector $\sum[d_i(x)g_i(x) \mid U_i \in u^r]$. On each U_i, h is a finite sum and thus is uniformly continuous; therefore by 3.8, h is uniformly continuous on μX. On each U_i, hence on X, h is a homeomorphism. Finally, since f realizes u, the sets $h(U_i)$ form a locally finite collection of compact sets, and therefore their union is closed. By 3.11, the proof is complete.

REFERENCES

1. V. Boltyanski, *On a property of two-dimensional compacta,* Dokl. Akab. Nauk **75** (1950), 605–608.
2. N. Bourbaki, *Topologie Générale,* Act. Sci. Ind. 858, Paris, 1940.
3. ———, *Topologie Générale,* Act. Sci. Ind. 1045, Paris, 1948.
4. S. Ginsburg and J. Isbell, *Some operators on uniform spaces,* to appear.
5. J. Isbell, *Algebras of uniformly continuous functions,* to appear.
6. M. Katětov, *On real-valued functions in topological spaces,* Fund. Math. **38** (1951), 85–91.
7. K. Morita, *On the simple extension of a space with respect to a uniformity,* I, Proc. Japan Acad. **27** (1951), 65–72.
8. G. Mostow, *Equivariant embeddings in euclidean space,* Ann. Math. **65** (1957), 432–446.
9. Yu. Smirnov, *Geometry of infinite uniform complexes and δ-dimension of proximity spaces,* Mat. Sb. **40** (82), (1956), 137–156.
10. J. Tukey, *Convergence and uniformity in topology.* Ann. Math. Study 2, Princeton, 1940.

UNIVERSITY OF WASHINGTON

MANY SERVER QUEUEING PROCESSES WITH POISSON INPUT AND EXPONENTIAL SERVICE TIMES

SAMUEL KARLIN and JAMES MCGREGOR

1. Introduction. A birth and death process is a stationary Markoff process whose state space is the non-negative integers and whose transition probability matrix

$$P_{ij}(t)=\Pr\{x(t)=j\,|\,x(0)=i\} \tag{1.1}$$

satisfies the conditions (as $t\to 0$)

$$p_{ij}(t)=\begin{cases}\lambda_i t+o(t) & \text{if } j=i+1\,,\\ \mu_i t+o(t) & \text{if } j=i-1\,,\\ 1-(\lambda_i+\mu_i)t+o(t) & \text{if } j=i\,,\end{cases} \tag{1.2}$$

where $\lambda_i>0$ for $i\geqq 0$, $\mu_i>0$ for $i\geqq 1$, and $\mu_0\geqq 0$. The process is called a queueing process if $\mu_0=0$ and $\lambda_i=\lambda$ for all i. The state of the system is then interpreted as the length of a queue for which the inter-arrival times have a negative exponential distribution with parameter λ, and for which the service times have a negative exponential distribution whose parameter μ_n depends on the length of the line. The classical case of a single server queue corresponds to $\mu_n=\mu$, $n\geqq 1$, and has been discussed by Reuter and Lederman [9] and Bailey [1].

The so-called telephone trunking problem (Feller [3]) arises from a queueing process with infinitely many servers, each of whose service time distribution has the same parameter μ, so that $\mu_n=n\mu$, $n\geqq 1$. Besides these two special cases, we discuss a queue with n servers, each of whose service time has a negative exponential distribution with the same parameter μ, so that $\mu_k=k\mu$ for $1\leqq k\leqq n$, $\mu_k=n\mu$ for $k\geqq n$. Our methods can also be used to study queueing processes with several servers whose service times have negative exponential distributions not all with the same parameter.

A sample of the type of problems treated is as follows:

(1) to obtain a usable formula for the transition probability $P_{ij}(t)$;

(2) to compute the distribution of the length of a busy period;

(3) to compute the distribution of the number of customers served during a busy period;

(4) to compute the distribution of the maximum length of the queue during a busy period; and similar questions.

Received May 29, 1957.

At this point it would be of some interest to tie the investigations of this paper together with the other work in this field. It is important to emphasize that we are concerned primarily with the analysis of non-stationary problems associated with the n server queuing process. The equilibrium distribution of length of line for the case of exponential service time and Poisson input is trivial to determine. The equilibrium situation for the general input process with exponential service time and n servers was completely resolved by Kendall [7] who, in addition, evaluated explicitly the distribution of waiting time for a randomly arriving customer. A non-constructive existence theorem for the stationary distribution of a general input process and a general service time distribution was given in [8]. In contrast, a considerable amount of insight regarding transient behavior has been attained in the case of the one server queue. For an elegant treatment of this case the reader is referred to the work of Takács [10].

Part of the significance in resolving the problems related to the n server queue even subject to the special assumptions of exponential service time and Poisson input, in addition to its independent interest, rests on the following two observations:

(1) the general queueing process with the corresponding appropriate parameters behaves on an average like the exponential case, and

(2) the solution for the exponential case may be suggestive as to the nature of the answers in the general case.

Our detailed analysis regarding queueing processes with exponential service time, Poisson input, and many servers derives from our knowledge of the refined structure of birth and death processes developed in [4] and [5]. We rely primarily on the theory of recurrence and absorption for a birth and death process as spelled out in [5].

In this connection, although the parameter μ_0 is zero for a queueing process it is convenient to consider, along with a queueing process, related birth and death processes for which μ_0 is positive. Such a process has an ignored absorbing state at -1, a state in which the system remains forever once it arrives there. When the system is in the zero state and a transition occurs, the system moves to state 1 with probability $\lambda_0/(\lambda_0+\mu_0)$ and is absorbed with probability $\mu_0/\lambda_0+\mu_0$.

The infinitesimal matrix of the general birth and death process is of the form

$$A=\begin{bmatrix} -(\lambda_0+\mu_0) & \lambda_0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \\ \cdots & & & \end{bmatrix} \tag{1.3}$$

This matrix determines a system of polynomials by means of the recursion relations

$$(1.4)\qquad \begin{cases} Q_0(x)=1\,, \\ -xQ_0(x)=-(\lambda_0+\mu_0)Q_0(x)+\lambda_0 Q_1(x)\,, \\ -xQ_n(x)=\mu_n Q_{n-1}(x)-(\lambda_n+\mu_n)Q_n(x)+\lambda_n Q_{n+1}(x)\,. \end{cases}$$

It is shown in [4] that there is a positive regular measure ψ on $0 \leqq x < \infty$ for which the orthogonality relations

$$(1.5)\qquad \int_0^\infty Q_i(x)Q_j(x)d\psi(x)=\frac{\delta_{ij}}{\pi_j} \qquad i, j=0, 1, \cdots$$

where $\pi_0=1$, $\pi_n=\dfrac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}$, are valid. In the case of a queueing process, the measure ψ is unique [4], and moreover the transition probability matrix $P(t)=(P_{ij}(t))$ of the process is uniquely determined by A. It has the representation

$$(1.6)\qquad P_{ij}(t)=\pi_j\int_0^\infty e^{-xt}Q_i(x)Q_j(x)\,d\psi(x)\,.$$

This is an extremely useful form of expression for the transition probability function, and our first task will be to compute the polynomials $\{Q_n(x)\}$ and the spectral measure ψ belonging to the various queueing processes.

This is accomplished in the following section based on a formula which connects the Stieltjes transform of the spectral measure of the process and the Stieltjes transform of the spectral measure of the associated process. Once the Stieltjes transform of the spectral measure is known, then recourse to the classical inversion formulas of Stieltjes transforms enables us to determine the spectral measure itself. This is done in § 4. Previous to that in § 3 a discussion of the infinite server queueing process is made. Here we recognize the corresponding polynomials as the classical Poisson-Charlier polynomials which are known to be orthogonal with respect to an appropriate Poisson distribution. Some remarks are appended describing the nature of first passage distributions of the states of the system in this case.

In the following section the spectral measure and the polynomials of the n server queueing process are explicitly determined. The polynomials are found to be expressible as combinations of the familiar Chebycheff polynomials of the first and second kind and Poisson-Charlier polynomials.

In § 5 the previous theory is specialized to the one and two server process. Further detailed information regarding these processes is collected.

§ 6 is devoted to a complete study of various probability distributions associated with queueing problems of one and two servers. The transition probabilities of the Markov process describing the waiting line are explicitly determined. The distribution to the length of the busy period, the distribution of the number of customers served during a busy period, and other such distributions are exhibited. In the following section the corresponding results for the n server queue are written out. The proofs of these assertions for the general case, exceedingly more complicated in detail, are carried out in the discussion of Appendix A. In § 8 we derive the distribution of the maximum length of line during a busy period. The second appendix summarizes the properties of a new system of polynomials related to the Poisson-Charlier polynomials.

2. The related processes. From a given birth and death process with infinitesimal matrix (1.3) a new process is obtained by stopping the given process whenever the state 0 is reached. For this new process the state 0 is an absorbing state, and if we ignore this state the process is a birth and death process for which the parameter μ_0 is positive. The waiting time in any state $i \geqq 1$ has the same distribution for both the original and the new process, and moreover both processes have the same post exit distributions for each state $i \geqq 1$. Consequently the infinitesimal matrix of the new process (with the state 0 ignored) is

$$\begin{bmatrix} -(\lambda_1+\mu_1) & \lambda_1 & 0 & 0 & \cdots \\ \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & 0 & \cdots \\ \cdots & & & & \end{bmatrix}, \tag{2.1}$$

which is obtained from (1.3) by removing the zero row and zero column. The polynomials defined by

$$\begin{cases} Q_0^{(0)}(x)=0,\ Q_1^{(0)}(x)=-\dfrac{1}{\lambda_0}, \\ -xQ_n^{(0)}(x)=\mu_n Q_{n-1}^{(0)}(x)-(\lambda_n+\mu_n)Q_n^{(0)}(x)+\lambda_n Q_{n+1}^{(0)}(x),\ n\geqq 1 \end{cases} \tag{2.2}$$

are called the *associated polynomials* of the system $\{Q_n(x)\}$. It is seen that, except for the constant factor $-\dfrac{1}{\lambda_0}$, they are the polynomials belonging to the new birth and death process. Consequently the transition probability matrix $(\tilde{P}_{ij}(t))$, $i, j \geqq 1$, of the new process is given by

$$\tilde{P}_{ij}(t)=\frac{\mu_1}{\lambda_0}\pi_j\int_0^\infty e^{-xt}[-\lambda_0 Q_i^{(0)}(x)][-\lambda_0 Q_j^{(0)}(x)]d\alpha(x) \tag{2.3}$$

where α is the spectral measure of the new process. In [5, § 8] it is shown that the Stieltjes transforms of the spectral measures ψ and α of the two processes,

$$(2.4)\qquad B(s)=\int_0^\infty \frac{d\psi}{x-s}\,,\qquad C(s)=\int_0^\infty \frac{d\alpha}{x-s}\,,$$

are related by the identity

$$(2.5)\qquad B(s)=\frac{1}{\lambda_0+\mu_0-s-\lambda_0\mu_1 C(s)}\,.$$

This identity is the basic tool used in calculating the spectral measure of the n server queueing process. Once the function $B(s)$ has been found the measure can be computed by means of known formulas for inverting the Stieltjes transform. See [5] for a discussion of this inversion relative to the identity (2.5), and [12], [11] for the general inversion problem.

By iterating (2.5) a relation will be obtained between the spectral measure of the original process and the spectral measure of the process obtained from the original one by stopping it whenever the state n is reached. Denote the spectral measure of the original process by ψ_0, and the spectral measure of the process obtained from the original one by stopping it whenever the nth state is reached, by ψ_{n+1}. Then if

$$(2.6)\qquad B_k(s)=\int_0^\infty \frac{d\psi_k(x)}{x-s}\,,$$

(2.5) gives

$$(2.7)\qquad B_n(s)=\frac{\dfrac{1}{\lambda_n}}{\dfrac{\lambda_n+\mu_n-s}{\lambda_n}-\mu_{n+1}B_{n+1}(s)}\,.$$

It is clear that

$$(2.8)\qquad B_0(s)=\frac{\alpha_n B_n(s)+\beta_n}{\gamma_n B_n(s)+\delta_n}$$

where $\alpha_n, \beta_n, \gamma_n, \delta_n$ are (not uniquely determined) functions of s. Permissible choices for $n=0, 1$ are $\alpha_0=1, \beta_0=\gamma_0=0, \delta_0=1$ and $\alpha_1=0, \beta_1=1/\lambda_0$, $\gamma_1=-\mu_1$, $\delta_1=(\lambda_0+\mu_0-s)/\lambda_0=Q_1(s)$. Substituting (2.7) into (2.8) it is found that the coefficient functions can be determined by the relations

$$\alpha_{n+1}=-\mu_{n+1}\beta_n\,,$$
$$\lambda_n\beta_{n+1}=\alpha_n+(\lambda_n+\mu_n-s)\beta_n\,,$$

$$\gamma_{n+1} = -\mu_{n+1}\delta_n ,$$
$$\lambda_n\delta_{n+1} = \gamma_n + (\lambda_n + \mu_n - s)\delta_n ,$$

and hence

$$B_0(s) = -\frac{\mu_n Q^{(0)}_{n-1}(s) B_n(s) - Q^{(0)}_n(s)}{\mu_n Q_{n-1}(s) B_n(s) - Q_n(s)} . \tag{2.9}$$

3. **The queue with infinitely many servers.** The polynomials belonging to this process will be denoted by $p_n(x) = p_n(x, \lambda, \mu)$. They satify

$$\begin{cases} p_0(x) = 1, \\ -xp_0(x) = -\lambda p_0(x) + \lambda p_1(x), \\ -xp_n(x) = n\mu p_{n-1}(x) - (\lambda + n\mu)p_n(x) + \lambda p_{n+1}(x), \qquad n \geqq 1 . \end{cases} \tag{3.1}$$

They can be identified in terms of the Poisson-Charlier polynomials $c_n(x, a)$, [2, Vol. 2, p. 226], which satisfy

$$\begin{cases} c_0(x, a) = 1, \\ -xc_0(x, a) = -ac_0(x, a) + ac_1(x, a) , \\ -xc_n(x, a) = nc_{n-1}(x, a) - (n + a)c_n(x, a) + ac_{n+1}(x, a), \qquad n \geqq 1. \end{cases} \tag{3.2}$$

Thus

$$p_n(x, \lambda, \mu) = c_n\left(\frac{x}{\mu}, \frac{\lambda}{\mu}\right) , \tag{3.3}$$

The measure with respect to which the Poisson-Charlier polynomials are orthogonal consists of masses

$$j(x) = e^{-a}\frac{a^x}{x\,!} \text{ at } x = 0, 1, 2, \cdots .$$

Hence the spectral measure ψ of the infinite server queue consists of masses

$$d\psi(x) = \frac{e^{-a}a^n}{n\,!} \text{ at } x_n = n\mu,\ n = 0, 1, \cdots \tag{3.4}$$

where $a = \frac{\lambda}{\mu}$. From well-known properties of the Poisson-Charlier polynomials [2] it is found that

$$p_n(k\mu) = p_k(n\mu) , \tag{3.5}$$

$$(3.6) \qquad \sum_{n=0}^{\infty} p_n(\mu x)\frac{z^n}{n\,!}=e^z\Big(1-\frac{z}{a}\Big)^x, \quad a=\frac{\lambda}{\mu}.$$

The representation of the transition probability matrix is

$$(3.7) \qquad \begin{aligned} P_{nk}(t)&=\pi_k\int_0^\infty e^{-xt}p_n(x)p_k(x)d\psi(x) \\ &=\frac{a^k}{k\,!}\sum_{r=0}^{\infty} e^{-r\mu t}p_n(r\mu)p_k(r\mu)e^{-a}\frac{a^r}{r\,!}. \end{aligned}$$

In particular

$$(3.8) \qquad \begin{aligned} P_{n0}(t)&=e^{-a}\sum_{r=0}^{\infty} p_r(n\mu)\frac{(ae^{-\mu t})^r}{r\,!} \\ &=e^{-a(1-e^{-\mu t})}(1-e^{-\mu t})^n, \end{aligned}$$

and

$$(3.9) \qquad \begin{aligned} \sum_{k=0}^{\infty} P_{nk}(t)z^k&=e^{-a}\sum_{r=0}^{\infty} p_r(n\mu)\frac{(ae^{-\mu t})^r}{r\,!}\sum_{k=0}^{\infty} p_k(r\mu)\frac{(az)^k}{k\,!} \\ &=e^{-a(1-z)(1-e^{-\mu t})}[1-(1-z)e^{-\mu t}]^n. \end{aligned}$$

The last two formulas are well-known and can be found by generating function techniques [3, p. 396].

Now consider the spectral measure α of the process obtained by stopping the infinitely many server process when the zero state is reached. Writing (2.5) in the form

$$(3.10) \qquad C(s)=\int_0^\infty \frac{d\alpha(x)}{x-s}=\frac{1}{\lambda\mu}\left[\lambda-s-\frac{1}{\displaystyle\int_0^\infty \frac{d\psi(x)}{x-s}}\right]$$

and noting that

$$(3.11) \qquad B(s)=\int_0^\infty \frac{d\psi(x)}{x-s}=e^{-a}\sum_{n=0}^{\infty}\frac{a^n}{n\,!(n\mu-s)},$$

we see that $C(s)$ is a metomorphic function whose poles are simple poles at the zeros of $B(s)$. Thus α is a discrete distribution whose masses are located at the zeros of $B(s)$. The zeros $s_0<s_1<s_2<\cdots$ of $B(s)$ are all simple, $n\mu<s_n<(n+1)\mu$, and the mass α_n of the distribution α which is located at s_n is

$$(3.12) \qquad \alpha_n=\frac{+1}{\lambda\mu B'(s_n)}.$$

(See [5] for a more complete discussion. The function here denoted by $B(s)$ is there denoted by $B(-s)$). For many purposes it is sufficient to know s_n and α_n for only the first few values of n. For example the first passage time distribution

$$F_{10}(t)=Pr\{X(\tau)=0 \text{ for some } \tau,\ 0<\tau\leqq t\,|\,X(0)=1\}$$

of the original process is

$$F_{10}(t)=\mu\int_0^\infty \frac{1-e^{-xt}}{x}\,d\alpha(x) \tag{3.13}$$

$$=1-\mu\sum_{n=0}^\infty \frac{\alpha_n}{s_n}e^{-s_nt}$$

and for large t only the first few terms are important.

For purposes of numerical computation the following facts, stated for the case $\mu=1$, are useful:

(i) for all $a>0$, $\frac{ds_n}{da}<0$, and

(ii) $s_n<s_{n+1}-1$.

To prove (i) it is noted that

$$\sum_{k=0}^\infty \frac{a^k}{k\,![k-s_n(a)]}\equiv 0\,, \qquad a>0\,,$$

and hence

$$\sum_{k=0}^\infty \frac{ka^{k-1}}{k\,![k-s_n(a)]}+\sum_{k=0}^\infty \frac{a^k}{k\,![k-s_n(a)]^2}\left(\frac{ds_n}{da}\right)=0\,. \tag{3.14}$$

Consequently it is sufficient to show that

$$\sum_{k=0}^\infty \frac{ka^k}{k\,!(k-s_n)}>0\,.$$

Now $n<s_n<n+1$ so

$$\sum_{k=n}^\infty \frac{a^k}{k\,!(k-s_n)}=-\sum_{k=0}^{n-1}\frac{a^k}{k\,!(k-s_n)}>0$$

and

$$\sum_{k=0}^\infty \frac{ka^k}{k\,!(k-s_n)}=\sum_{k=1}^{n}\sum_{r=k}^\infty \frac{a^r}{r\,!(r-s_n)}+\sum_{k=n+1}^\infty \frac{(k-n)a^n}{k\,!(k-s_n)}>0\,.$$

To prove (ii) it is observed that

$$B(s, a) = e^{-a} \sum_{k=0}^{\infty} \frac{a^k}{k!(k-s)}$$

satisfies the functional equation

$$\frac{\partial B(s, a)}{\partial a} = -[B(s, a) - B(s-1, a)] . \tag{3.15}$$

Because of (i) and (3.14) $\left.\frac{\partial B(s, a)}{\partial a}\right|_{s=s_n} > 0$ which, together with (3.15), gives $B(s_n - 1, a) > B(s_n, a) = 0$. Now $B(s, a) = B(s)$ is monotone increasing in each interval $n < s < n+1$, $n = 0, 1, 2, \cdots$. Consequently

$$s_{n-1} < s_n - 1 .$$

The following table gives s_n and $B'(s_n)$ for $n = 0, 1, 2$ and several values of a.

a	$\hat{s}_1$ First root	$B'(\hat{s}_1)$ At first root	$\hat{s}_2$ Second root	$B'(\hat{s}_2)$	$\hat{s}_3$	$B'(\hat{s}_3)$
.5	.65116	6.54006	1.88388	10.21023	2.97092	25.00957
1.0	.45027	8.47902	1.72376	8.90911	2.88131	12.91379
1.5	.31745	13.63762	1.58297	11.60410	2.77136	13.41379
2.0	.22517	23.92535	1.46574	17.38949	2.66252	17.55924

4. The spectral measure and the polynomials of the n server queue. For the n server process

$$\begin{cases} \lambda_k = \lambda , \\ \mu_k = \begin{cases} k\mu, & k \leqq n , \\ n\mu, & k \geqq n . \end{cases} \end{cases} \tag{4.1}$$

Hence

$$Q_k(x) = p_k(x, \lambda, \mu) , \qquad k \leqq n , \tag{4.2}$$

where p_n is given by (3.3). The polynomials for $k \geqq n$ will be determined presently. As in § 2 we denote the spectral measure of the process by ψ_0 and the spectral measure of the process obtained if the given process is stopped whenever the state k $(k \geqq 0)$ is reached by ψ_{k+1}. If

$$B_k(s) = \int_0^{\infty} \frac{d\psi_k(x)}{x - s} \tag{4.3}$$

then from (2.9)

$$B_0(s) = -\frac{n\mu Q_{n-1}^{(0)}(s) B_n(s) - Q_n^{(0)}(s)}{n\mu Q_{n-1}(s) B_n(s) - Q_n(s)} . \tag{4.4}$$

Because of (4.1), $B_n(s)=B_{n+1}(s)$ and hence (2.7) gives

$$B_n(s)=\frac{\lambda+n\mu-s-\sqrt{(\lambda+n\mu-s)^2-4n\lambda\mu}}{2n\lambda\mu}, \tag{4.5}$$

where, in accordance with (4.3), the square root is taken positive for $s<0$. Substituting (4.5) into (4.4) and rationalizing we obtain, with the use of the identity $\lambda_{n-1}\pi_{n-1}[Q_nQ_{n-1}^{(0)}-Q_n^{(0)}Q_{n-1}]=1$, where $\lambda_n\pi_n=\lambda\left(\frac{\lambda}{\mu}\right)^n\frac{1}{n!}$,

$$B_0(s)=-\frac{L_n(s)}{K_n(s)} \tag{4.6}$$

where

$$\begin{aligned} L_n(s)=&\,4\lambda^2Q_n(s)Q_n^{(0)}(s)+4n\lambda\mu Q_{n-1}(s)Q_{n-1}^{(0)}(s)\\ &-2\lambda(\lambda+n\mu-s)[Q_n(s)Q_{n-1}^{(0)}(s)+Q_n^{(0)}(s)Q_{n-1}(s)]\\ &+2(n-1)!\left(\frac{\mu}{\lambda}\right)^{n-1}\sqrt{(\lambda+n\mu-s)^2-4n\lambda\mu}, \end{aligned} \tag{4.7}$$

$$\begin{aligned} K_n(s)=&\,4\lambda^2[Q_n^2(s)-Q_{n-1}(s)Q_{n+1}(s)]\\ =&\,4n\lambda\mu[Q_{n-1}^2(s)-Q_n(s)Q_{n-2}(s)]-4\lambda\mu Q_n(s)[Q_{n-1}(s)-Q_{n-2}(s)]. \end{aligned} \tag{4.8}$$

$$\begin{aligned} L_n(s)=&\,4\lambda^2\left[Q_n(s)Q_n^0(s)-\frac{1}{2}\{Q_{n-1}^{(0)}(s)Q_{n+1}(s)+Q_{n-1}(s)Q_{n+1}^{(0)}(s)\}\right]\\ &+2(n-1)!\left(\frac{\mu}{\lambda}\right)^{n-1}\sqrt{(\lambda+n\mu-s)^2-4n\lambda\mu}. \end{aligned}$$

It is seen that $K_n(s)$ is a polynomial in s of exact degree $2n-1$, with a root at $x=0$, and that the polynomial part of $L_n(s)$ is of degree 2n−2. The Stieltjes inversion formula

$$\psi_0(x)=\frac{1}{\pi}\lim_{\eta\to0+}\int_{-\varepsilon}^{x}\Im B_0(\xi+i\eta)d\xi \tag{4.9}$$

gives ψ_0 at all of its points of continuity. The above formulas show that $\Im B(x+i\eta)$ converges uniformly to zero as $\eta\to0^+$ if x is in any closed interval containing no zeros of $K_n(x)$ and disjoint from the interval $|\lambda+n\mu-x|\leqq\sqrt{4n\lambda\mu}$. Consequently over the interval

$$|\lambda+n\mu-x|<\sqrt{4n\lambda\mu} \tag{4.10}$$

the measure ψ_0 has a continuous density $\psi'_0(x)$ given by

$$\psi'_0(x)=\frac{(n-1)!}{2\pi\lambda^2}\left(\frac{\mu}{\lambda}\right)^{n-1}\frac{\sqrt{4n\lambda\mu-(\lambda+n\mu-x)^2}}{Q_n^2(x)-Q_{n-1}(x)Q_{n+1}(x)}. \tag{4.11}$$

In addition ψ_0 may have discrete masses at some or all of the zeros of $K_n(x)$. This possibility is discussed in § 7, and treated in detail in the appendix.

To determine the polynomials $Q_k(x)$ for $k \geqq n$, let $R_k(x)=Q_{n+k}(x)$, $k \geqq -1$. Then

$$R_0(x)=Q_n(x)\,, \qquad R_{-1}(x)=Q_{n-1}(x) \tag{4.12}$$

and

$$-xR_k(x)=n\mu R_{k-1}(x)-(\lambda+n\mu)R_k(x)+\lambda R_{k+1}(x)\,, \quad k \geqq 0\,, \tag{4.13}$$

which is a recurrence formula in which the coefficients are independent of k. The $\{R_k(x)\}$ can be expressed in terms of the Chebycheff polynomials $\{T_k(x)\}$, $\{U_k(x)\}$ which satisfy

$$xP_k(x)=\frac{1}{2}P_{k-1}(x)+\frac{1}{2}P_{k+1}(x)\,, \qquad k \geqq 1\,, \tag{4.14}$$

and

$$T_0(x)=1\,, \quad T_{-1}(x)=x\,, \quad U_0(x)=1\,, \quad U_{-1}(x)=0\,. \tag{4.15}$$

In fact, since

$$V_k(x)=\left(\frac{n\mu}{\lambda}\right)^{k/2} T_k\left(\frac{\lambda+n\mu-x}{\sqrt{4n\lambda\mu}}\right)$$

and

$$W_k(x)=\left(\frac{n\mu}{\lambda}\right)^{k/2} U_k\left(\frac{\lambda+n\mu-x}{\sqrt{4n\lambda\mu}}\right)$$

are solutions of (4.13) for which

$$V_0(x)=1\,, \quad V_{-1}(x)=\frac{\lambda+n\mu-x}{2n\mu}\,, \quad W_0(x)=1\,, \quad W_{-1}(x)=0\,,$$

we have

$$\begin{aligned} R_k(x)&=\frac{2n\mu}{\lambda+n\mu-x}Q_{n-1}(x)V_k(x) \\ &\quad+\left[Q_n(x)-\frac{2n\mu}{\lambda+n\mu-x}Q_{n-1}(x)\right]W_k(x) \\ &=Q_n(x)W_k(x)-\frac{n\mu}{\lambda}Q_{n-1}(x)W_{k-1}(x)\,, \quad k \geqq 0\,. \end{aligned} \tag{4.16}$$

Hence for $k \geqq 0$

$$Q_{n+k}(x)=\left(\frac{n\mu}{\lambda}\right)^{k/2}\left[Q_n(x)U_k\left(\frac{\lambda+n\mu-x}{\sqrt{4n\lambda\mu}}\right)\right. \tag{4.17}$$

$$-\sqrt{\frac{n\mu}{\lambda}}Q_{n-1}(x)U_{k-1}\left(\frac{\lambda+n\mu-x}{\sqrt{4n\lambda\mu}}\right)\Bigg]$$

where $U_{-1}(\xi)=0$ and

$$U_k(\xi)=\frac{\sin(k+1)\theta}{\sin\theta}\ ,\quad \xi=\cos\theta\ ,\quad k\geqq 0\ .$$

The system of polynomials $\{Q_k(x)\}$ is completely determined by (4.2) and (4.17).

A similar argument shows that the associated polynomials $\{Q_k^{(0)}(x)\}$ satisfy

$$Q_{n+k}^{(0)}(x)=\left(\frac{n\mu}{\lambda}\right)^{k/2}\Bigg[Q_n^{(0)}(x)U_k\left(\frac{\lambda+n\mu-x}{\sqrt{4\lambda\mu}}\right) \tag{4.18}$$

$$-\sqrt{\frac{n\mu}{\lambda}}Q_{n-1}^{(0)}(x)U_{k-1}\left(\frac{\lambda+n\mu-x}{\sqrt{4\lambda\mu}}\right)\Bigg]$$

for $k\geqq 0$.

5. The spectral measure of the one server and two server processes. For the case of one server

$$Q_0(x)=1\ ,\quad Q_1(x)=\frac{\lambda-x}{\lambda}\ ,\quad Q_2(x)=\frac{\lambda+\mu-x}{\lambda}\cdot\frac{\lambda-x}{\lambda}-\frac{\mu}{\lambda}\ , \tag{5.1}$$

$$Q_0^{(0)}(x)=0\ ,\quad Q_1^{(0)}(x)=-\frac{1}{\lambda}\ ,\quad Q_2^{(0)}(x)=-\frac{\lambda+\mu-x}{\lambda^2}\ . \tag{5.2}$$

Using these values in (4.7), (4.8) gives

$$B_0(s)=-\frac{2(\mu-\lambda+s)+2\sqrt{(\lambda+\mu-s)^2-4\lambda\mu}}{4\mu s}\ . \tag{5.3}$$

The only possible pole of $B_0(s)$ is at $s=0$, and

$$\lim_{s\to 0}-sB_0(s)=\frac{1}{2\mu}[(\mu-\lambda)+|\mu-\lambda|]$$

$$=\begin{cases}0 & \text{if } \mu\leqq\lambda\\ \dfrac{\mu-\lambda}{\mu} & \text{if } \mu>\lambda\ .\end{cases}$$

Thus the spectral measure ψ has the continuous density

$$\psi'(x)=\frac{1}{2\pi\mu}\,\frac{\sqrt{4\lambda\mu-(\lambda+\mu-x)^2}}{x} \tag{5.4}$$

on the interval $|\lambda+\mu-x|<\sqrt{4\lambda\mu}$ and has in addition a mass of amount

$\frac{\mu-\lambda}{\mu}$ located at $x=0$ if $\mu>\lambda$, but has no extra mass if $\mu\leqq\lambda$. The polynomials are given by

$$Q_{k+1}(x)=\left(\frac{\mu}{\lambda}\right)^{k/2}\left[\frac{\lambda-x}{\lambda}U_k\left(\frac{\lambda+\mu-x}{\sqrt{4\lambda\mu}}\right)\right.$$
$$\left.-\sqrt{\frac{\mu}{\lambda}}U_{k-1}\left(\frac{\lambda+\mu-x}{\sqrt{4\lambda\mu}}\right)\right], \quad k\geqq 0\,. \tag{5.5}$$

The associated polynomials are given by

$$Q_{k+1}^{(0)}(x)=-\frac{1}{\lambda}\left(\frac{\mu}{\lambda}\right)^{k/2}U_k\left(\frac{\lambda+\mu-x}{\sqrt{4\lambda\mu}}\right), \quad k\geqq 0\,, \tag{5.6}$$

and using (4.5) the function $B_1(s)$ is

$$B_1(s)=\frac{\lambda+\mu-s-\sqrt{(\lambda+\mu-s)-4\lambda\mu}}{2\lambda\mu}\,. \tag{5.7}$$

Hence the spectral measure ψ_1 of the associated process consists simply of the continuous density

$$\psi_1'(x)=\frac{1}{2\pi\lambda\mu}\sqrt{4\lambda\mu-(\lambda+\mu-x)^2}\,, \quad |\lambda+\mu-x|\leqq\sqrt{4\lambda\mu}\,. \tag{5.8}$$

We now turn to the two server case. The polynomials Q_k, $Q_k^{(0)}$ are again given by (5.1) and (5.2) for $k=0, 1, 2$ and a straightforward computation gives

$$K_2(s)=4\mu s\left[\left(\frac{\lambda-s}{\lambda}\right)^2+\frac{\mu}{\lambda}\frac{\lambda-s}{\lambda}+\frac{\mu}{\lambda}\right]$$
$$=\frac{4\mu s}{\lambda^2}[s^2-(2\lambda+\mu)s+\lambda(\lambda+2\mu)]\,,$$

$$L_2(s)=\frac{2\mu}{\lambda^2}[(\lambda-s)(2\mu-\lambda-2s)+\lambda\sqrt{(\lambda+2\mu-s)^2-8\lambda\mu}]\,,$$

and hence

$$B_0(s)=-\frac{(\lambda-s)(2\mu-\lambda-2s)+\lambda\sqrt{(\lambda+2\mu-s)^2-8\lambda\mu}}{2s[s^2-(2\lambda+\mu)s+\lambda(\lambda+2\mu)]}\,. \tag{5.9}$$

Consequently the spectral measure ψ of the two server queue has the density

$$\psi'(x)=\frac{\lambda}{2\pi}\frac{\sqrt{8\lambda\mu-(\lambda+2\mu-x)^2}}{x[x^2-(2\lambda+\mu)x+\lambda(\lambda+2\mu)]} \tag{5.10}$$

on the interval $|\lambda+2\mu-x|\leqq\sqrt{8\lambda\mu}$, and in addition may have jumps at one or more of the zeros of the denominator of (5.9). Considering first the zero at $s=0$, we find that

$$\lim_{s\to 0} -sB_0(s)=\frac{(2\mu-\lambda)+\sqrt{(2\mu-\lambda)^2}}{2(2\mu+\lambda)}$$

$$=\begin{cases}\dfrac{2\mu-\lambda}{2\mu+\lambda} & \text{if} \quad 2\mu>\lambda\ , \\ 0 & \text{if} \quad 2\mu\leqq\lambda\ .\end{cases}$$

Hence ψ has a jump of magnitude $\dfrac{2\mu-\lambda}{2\mu+\lambda}$ at $x=0$ if $2\mu>\lambda$, but no jump at $x=0$ if $2\mu\leqq\lambda$. The other zeros of the denominator are

$$s_1=\frac{2\lambda+\mu}{2}-\frac{1}{2}\sqrt{\mu(\mu-4\lambda)}\ ,$$

$$s_2=\frac{2\lambda+\mu}{2}+\frac{1}{2}\sqrt{\mu(\mu-4\lambda)}\ .$$

These two roots are non-real if $\mu<4\lambda$. Since $B_0(s)$ has no non-real poles we assume $\mu\geqq 4\lambda$. If $\mu=4\lambda$ the denominator has a double zero at $s=3\lambda$. A simple computation shows that in this case the numerator also has a double zero at $s=3\lambda$, and hence no jump of ψ is involved. If $\mu>4\lambda$ the residue at s_2 is easily computed. In fact

$$(\lambda+2\mu-s_2)^2-8\lambda\mu=\frac{1}{4}[10\mu^2-6\mu\sqrt{\mu(\mu-4\lambda)}-36\lambda\mu]$$

$$=\left[\frac{\mu-3\sqrt{\mu(\mu-4\lambda)}}{2}\right]^2,$$

and hence

$$(\lambda-s_2)(2\mu-\lambda-2s_2)+\lambda\sqrt{(\lambda+2\mu-s_2)^2-8\lambda\mu}$$

$$=-\frac{\lambda}{2}\{\mu-3\sqrt{\mu(\mu-4\lambda)}-|\mu-3\sqrt{\mu(\mu-4\lambda)}|\}$$

$$=\begin{cases}0 & \text{if} \quad \mu\geqq 3\sqrt{\mu(\mu-4\lambda)} \\ \lambda[3\sqrt{\mu(\mu-4\lambda)}-\mu] & \text{if} \quad 3\sqrt{\mu(\mu-4\lambda)}>\mu\ .\end{cases}$$

The condition $3\sqrt{\mu(\mu-4\lambda)}>\mu$ is equivalent to $\dfrac{\mu}{\lambda}>\dfrac{9}{2}$. Consequently ψ has a jump at $x=s_2$ of magnitude

$$\frac{\lambda[3\sqrt{\mu(\mu-4\lambda)}-\mu]}{\sqrt{\mu(\mu-4\lambda)}\,[2\lambda+\mu+\sqrt{\mu(\mu-4\lambda)}\,]}$$

if $\frac{\mu}{\lambda}>\frac{9}{2}$, but no jump there if $\frac{\mu}{\lambda}\leqq\frac{9}{2}$. A similar calculation shows that ψ never has a jump at $x=s_1$. If $\mu>4\lambda$ then

$$s_2<\lambda+2\mu-\sqrt{8\lambda\mu}$$

except that equality holds when $\frac{\mu}{\lambda}=\frac{9}{2}$. The polynomials are given by

$$\text{(5.11)}\qquad Q_{k+2}(x)=\left(\frac{2\mu}{\lambda}\right)^{k/2}\left[\frac{(\lambda+\mu-x)(\lambda-x)}{\lambda^2}U_k\left(\frac{\lambda+2\mu-x}{\sqrt{8\lambda\mu}}\right)\right.$$
$$\left.-\sqrt{\frac{2\mu}{\lambda}}\cdot\frac{\lambda-x}{\lambda}U_{k-1}\left(\frac{\lambda+2\mu-x}{\sqrt{8\lambda\mu}}\right)\right]$$

for $k\geqq 0$.

6. Probability distributions of various random quantities Associated with the one and two server processes.

In this section we compute the distributions of some interesting random variables connected with the one and two server processes. The transition probability function of the one server process is

$$\text{(6.1)}\qquad P_{ij}(t)=Y\left(\frac{\mu}{\lambda}\right)\left(\frac{\lambda}{\mu}\right)^j\left(\frac{\mu-\lambda}{\mu}\right)$$
$$+\left(\frac{\lambda}{\mu}\right)^j\int_{\lambda+\mu-\sqrt{4\lambda\mu}}^{\lambda+\mu+\sqrt{4\lambda\mu}}e^{-xt}Q_i(x)Q_j(x)\frac{\sqrt{4\lambda\mu-(\lambda+\mu-x)^2}}{2\pi\mu x}dx$$

where $Y(z)$ is 0 if $z\leqq 1$, 1 if $z>1$, and the polynomials are given by (5.5). The explicit expression for the distribution of waiting time, $W(t,\xi)$, of a customer arriving at time t in the case of the one server queue may be readily derived from the integral representation (6.1). This is accomplished as follows: If at time t the length of line (state of the process) consists of n people with $n\geqq 1$ then the density of the waiting time of a person arriving at the moment t is the gamma density of order n whose scale parameter is μ. The probability that at time t the length of line is n where initially the state of the process was i is given by $P_{in}(t)$. Consequently, for $\xi>0$

$$\text{(6.2)}\qquad d_\xi W_i(t,\xi)=\sum_{n=1}^{\infty}P_{i,n}(t)\frac{\mu^n\xi^{n-1}e^{-\mu\xi}}{(n-1)!}d\xi\,.$$

Inserting the detailed formula (6.1) into the summation of (6.2) and performing the calculation, we obtain the formula

$$\text{(6.3)}\quad dW_i(t,\xi)=\frac{\lambda}{\mu}(\mu-\lambda)e^{-(\mu-\lambda)\xi}d\xi+$$

$$+e^{-\mu\xi}\sqrt{\frac{\lambda}{\mu}}\frac{1}{\pi}\int_{\lambda+\mu-\sqrt{4\lambda\mu}}^{\lambda+\mu+\sqrt{4\lambda\mu}}\frac{e^{-xt}}{x}Q_i(x)\Im\left[\{(\lambda-x)e^{i\theta}-\sqrt{\lambda\mu}\}\,e^{\xi\sqrt{\lambda\mu}e^{i\theta}}\right]dx\,d\xi$$

where $\cos\theta=\frac{\lambda+\mu-x}{\sqrt{4\lambda\mu}}$ and $\Im$ stands for the imaginary part. We have tacitly assumed that $\frac{\mu}{\lambda}>1$ which of course is the interesting and practical case.

The evaluation of the sum is direct once it is realized that $Q_n(x)$ can be expressed according to (5.5) in terms of Chebycheff polynomials.

$$Q_n(x)=\left(\frac{\mu}{\lambda}\right)^{\frac{n-1}{2}}\left[\left(\frac{\lambda-x}{x}\right)U_{n-1}\left(\frac{\lambda+\mu-x}{\sqrt{4\lambda\mu}}\right)-\sqrt{\frac{\mu}{\lambda}}\,U_{n-2}\left(\frac{\lambda+\mu-x}{\sqrt{4\lambda\mu}}\right)\right], \quad n\geqq 1$$

and

$$U_k(z)=\frac{\sin(k+1)\theta}{\sin\theta}=\frac{\mathscr{I}(e^{i(k+1)\theta})}{\sin\theta} \quad \text{where } \cos\theta=z\ .$$

Of course $P_{00}(t)$ evaluates the probability that a person arriving at time t doesn't have to wait for his service to begin.

For computational purposes it might be remarked that the integrals of (6.1) and (6.3) may be expressed in terms of combinations of Bessel functions with imaginary arguments. This follows from the familiar fact that the Laplace transform of $\sqrt{1-t^2}$ for $-1\leqq t\leqq 1$ involves Bessel functions [9]. This indeed is true of the majority of formulas connected with queueing. However, from the point of view of an understanding of the theory, and also for many practical purposes, we prefer the answer in the form of the integral representation.

The integral representation also enables us to determine directly the rate of approach to equilibrium in the ergodic case. The conclusion is immediate from relation (6.1) which implies the inequality

$$\left|P_{ij}(t)-\left(\frac{\mu-\lambda}{\mu}\right)\left(\frac{\lambda}{\mu}\right)^j\right|\leqq Me^{-t(\lambda+\mu-\sqrt{4\lambda\mu})}\ .$$

The asymptotic behavior of $P_{ij}(t)$ for large t is easily obtained from formula (6.1). For example, for the case when $\lambda=\mu$

$$P_{00}(t)=\frac{1}{2\pi\mu}\int_0^{4\lambda}e^{-xt}\sqrt{\frac{4\lambda-x}{x}}\,dx\ ,$$

and when t is large the main contribution to this integral is from the immediate neighborhood of $x=0$. In fact

$$P_{00}(t)\sim\frac{1}{\pi\sqrt{\mu}}\int_0^1\frac{e^{-xt}}{\sqrt{x}}\,dx=\frac{1}{\pi\sqrt{t\mu}}\int_0^t\frac{e^{-\xi}\,d\xi}{\sqrt{\xi}}$$

and hence

$$P_{00}(t) \sim \frac{1}{\sqrt{\pi \mu t}} \quad \text{as } t \to \infty \ .$$

The cases when $\lambda > \mu$ or $\lambda < \mu$ can be dealt with in a similar way.

Now consider the distribution of the length of a busy period, or what is the same thing, the distribution of the first passage time from state 1 to state 0, or what is the same again, the distribution of the time of absorption into the zero state for the related process (§2), given that the related process starts in state 1. If $\tilde{P}_{ij}(t)$ is the transition probability function of the related process, and ψ_1 is its spectral measure, then the probability $F_{10}(t)$ that absorption occurs before time t is

$$F_{10}(t) = \mu \int_0^t \tilde{P}_{11}(\tau)\, d\tau = \mu \int_0^\infty \frac{1-e^{-xt}}{x}\, d\psi_1(x) \ .$$

The Stieltjes transform of ψ_1 is given by (4.5) with $n=1$, and hence ψ_1 consists of the density

$$\psi_1'(x) = \frac{1}{2\pi\lambda\mu} \sqrt{4\lambda\mu - (\lambda+\mu-x)^2}$$

on the interval $|\lambda+\mu-x| \leqq \sqrt{4\lambda\mu}$. Consequently

$$F_{10}(t) = \frac{1}{2\pi\lambda} \int_{\lambda+\mu-\sqrt{4\lambda\mu}}^{\lambda+\mu+\sqrt{4\lambda\mu}} \frac{1-e^{-xt}}{x} \sqrt{4\lambda\mu-(\lambda+\mu-x)^2}\, dx \tag{6.4}$$

is the probability that the length of a busy period for the one server queue is $\leqq t$. In a similar way the probability $F_{k0}(t)$ that the queue will become idle before time t when there are k customers at time zero can be computed. Using the fact that the associated polynomials are given by

$$Q_{n+1}^{(0)}(x) = -\frac{1}{\lambda}\left(\frac{\mu}{\lambda}\right)^{n/2} U_n\left(\frac{\lambda+\mu-x}{\sqrt{4\lambda\mu}}\right) , \tag{6.5}$$

one obtains

$$\begin{aligned} F_{k0}(t) = \frac{1}{2\pi\lambda} \int_{\lambda+\mu-\sqrt{4\lambda\mu}}^{\lambda+\mu+\sqrt{4\lambda\mu}} & \frac{1-e^{-xt}}{x} \\ & \cdot \left(\frac{\mu}{\lambda}\right)^{\frac{k-1}{2}} U_{k-1}\left(\frac{\lambda+\mu-x}{\sqrt{4\lambda\mu}}\right) \sqrt{4\lambda\mu-(\lambda+\mu-x)^2}\, dx \ . \end{aligned} \tag{6.6}$$

It is also possible to compute the distribution of the number N of customers arriving during a busy period, or more generally the number N_k of customers arriving before the queue becomes idle given that initially there were k customers in the queue. For this purpose we consider the random walk whose possible states are the integers $1, 2, 3, \cdots$ and an ignored absorbing state at 0. The one step transition probabilities of the random walk are

$$P_{ij}=\begin{cases} p_i & \text{if } j=i+1\ , \\ q_i & \text{if } j=i-1\ , \\ 0 & \text{if } j=i \text{ or } |j-i|>1\ , \end{cases}$$

where

$$p_i=\frac{\lambda_i}{\lambda_i+\mu_i}=\frac{\lambda}{\lambda+\mu}\ , \qquad q_i=\frac{\mu_i}{\lambda_i+\mu_i}=\frac{\mu}{\lambda+\mu}\ .$$

These quantities are independent of i and we denote then by p, q. When the particle executing the random walk is in state 1 and a transition occurs, the particle goes to state 2 with probability p and is absorbed into the zero state with probability q.

Each sample function of the associated queueing process generates in an obvious way a sample function of the random walk process, and it is clear that the random variable N_k, which is the number of customers arriving before the queue becomes idle, is the same as the total number of steps to the right made by the random walk before absorption at zero occurs. The total number of transitions of the random walk process which occur before absorption is a random variable M_k related to N_k and the initial state k in such a way that

$$M_k=k+2N_k\ . \tag{6.7}$$

If P_{ij}^n denotes the n step transition probability of the random walk, then

$$\Pr\{M_k=m\}=qP_{k,1}^{m-1} \tag{6.8}$$

and hence

$$\begin{aligned}\Pr\{N_k=n\}&=Pr\{M_k=k+2n\}\\ &=qP_{k,1}^{2n+k-1}\ .\end{aligned} \tag{6.9}$$

Thus the distribution of N_k is known if P_{ij}^n is known.

An integral representation for P_{ij}^n is obtained as follows. The random walk determines a system of polynomials by means of the recursion relations

$$\begin{cases} R_1(x)=1\ , \qquad R_0(x)=0 \\ xR_n(x)=qR_{n-1}(x)+pR_{n+1}(x)\ , \quad n\geq 1\ . \end{cases} \tag{6.10}$$

It is seen that $R_n(x)$ is a polynomial in x of exact degree n. It can be shown that the polynomials $R_n(x)$ are orthogonal on $-1\leq x\leq 1$ with respect to a uniquely determined measure α of total mass 1. A proof of this fact which covers not merely the queueing case, but also the random walk arising in a similar way from a general birth and death process, is outlined in [5]. It is rather obvious that when $x^n R_i(x)$ is written as a linear combination of the polynomials $\{R_k(x)\}$ the coefficient of $R_j(x)$ is P_{ij}^n. Since it can be shown from the recurrence formulas that

$$\int_{-1}^{1} R_n^2(x)\, d\alpha(x) = \frac{1}{\pi_n^*}\ , \qquad n \geqq 1\ ,$$

where $\pi_n^* = \left(\frac{p}{q}\right)^{n-1}$, it follows that

$$P_{ij}^n = \pi_j^* \int_{-1}^{1} x^n R_i(x) R_j(x)\, d\alpha(x)\ , \tag{6.11}$$

which is the desired representation of P_{ij}^n. Combining (6.9) and (6.11) we get an expression for the probability distribution of N_k in terms of the measure α. In particular the distribution of $N=N_1$, the number of customers arriving during a busy period, is

$$\Pr\{N=n\} = q\int_{-1}^{1} x^{2n}\, d\alpha(x)\ . \tag{6.12}$$

The polynomials which satisfy the recurrence relation (6.10) are easily found to be

$$R_n(x) = \left(\frac{\mu}{\lambda}\right)^{\frac{n-1}{2}} U_{n-1}\left(\frac{(\lambda+\mu)x}{\sqrt{4\lambda\mu}}\right)\ , \tag{6.13}$$

from which it follows that the measure α consists of the density

$$\alpha'(x) = \frac{2}{\pi}\,\frac{\lambda+\mu}{\sqrt{4\lambda\mu}}\sqrt{1-\left(\frac{\lambda+\mu}{\sqrt{4\lambda\mu}}x\right)^2} \tag{6.14}$$

on the interval $|x| \leqq \frac{\sqrt{4\lambda\mu}}{\lambda+\mu}$. Consequently

$$\begin{aligned} Pr\{N=n\} &= \frac{\mu}{\lambda+\mu}\cdot\frac{2}{\pi}\cdot\frac{\lambda+\mu}{\sqrt{4\lambda\mu}}\int_{-\frac{\sqrt{4\lambda\mu}}{\lambda+\mu}}^{\frac{\sqrt{4\lambda\mu}}{\lambda+\mu}} x^{2n}\sqrt{1-\left(\frac{\lambda+\mu}{\sqrt{4\lambda\mu}}x\right)^2}\, dx \\ &= \frac{\mu}{\lambda+\mu}\cdot\left(\frac{\sqrt{4\lambda\mu}}{\lambda+\mu}\right)^{2n}\cdot\frac{2}{\pi}\int_{-1}^{1}\xi^{2n}\sqrt{1-\xi^2}\, d\xi\ . \end{aligned} \tag{6.15}$$

We now turn to the two server queue. The transition probability function is

$$\begin{aligned} P_{ij}(t) = {} & Y\left(\frac{2\mu}{\lambda}\right)\cdot\frac{2\mu-\lambda}{2\mu+\lambda}\cdot 2\cdot\left(\frac{\lambda}{2\mu}\right)^j \\ & + Y\left(\frac{2\mu}{9\lambda}\right)\frac{\lambda[3\sqrt{\mu(\mu-4\lambda)}-\mu]}{\sqrt{\mu(\mu-4\lambda)}\,[2\lambda+\mu+\sqrt{\mu(\mu-4\lambda)}\,]}\cdot 2\cdot\left(\frac{\lambda}{2\mu}\right)^j e^{-s_2 t} Q_i(s_2) Q_j(s_2) + \end{aligned} \tag{6.16}$$

$$+2\left(\frac{\lambda}{2\mu}\right)^j \frac{\lambda}{2\pi}\int_{\lambda+2\mu-\sqrt{8\lambda\mu}}^{\lambda+2\mu+\sqrt{8\lambda\mu}} e^{-xt} Q_i(x)Q_j(x)\frac{\sqrt{8\lambda\mu-(\lambda+2\mu-x)^2}}{x[x^2-(2\lambda+\mu)x+\lambda(\lambda+2\mu)]}\,dx$$

(unless $j=0$ in which case $2\left(\frac{\lambda}{2\mu}\right)^j$ is to be replaced by 1), where $Y(z)$ is 1 if $z>1$, 0 if $z\leqq 1$, and

$$s_2=\frac{2\lambda+\mu}{2}+\frac{1}{2}\sqrt{\mu(\mu-4\lambda)}\,, \tag{6.17}$$

and where the polynomials $Q_i(x)$ are given by (5.11). Once again the asymptotic behavior of $P_{ij}(t)$ for large t is clearly exhibited by (6.16). In fact the first term on the right is either zero or else the largest term, and the second term, if not zero, is the second largest term. Finally the asymptotic behavior of the third term is a simple matter to investigate. For example if $\lambda=2\mu$ it is found that

$$P_{00}(t)\sim\frac{\lambda}{4}\frac{1}{\sqrt{\pi t}}\,.$$

By arguments entirely analogous to those used in the derivation of (6.3), we may obtain the form of the distribution of waiting time for a customer arriving at time t in the two server queue. In fact, if $W_i(t,\xi)$ represents the cumulative distribution of waiting time for a person arriving at time t where at time zero the state of the process was i, then

$$dW_i(t,\xi)=\sum_{n=2}^{\infty}P_{i,n}(t)\frac{(2\mu)^{n-1}\xi^{n-2}}{(n-2)!}e^{-2\xi}\,d\xi\,,\quad \xi>0\,,$$

with $P_{i,n}(t)$ given by (6.16). We restrict attention only to the ergodic case when $2\mu>\lambda$. Use of (6.16) in conjunction with (5.11) establishes the ultimate formula

$$\begin{aligned}
dW_i(t,\xi)\,d\xi&=\frac{2\mu-\lambda}{2\mu+\lambda}\cdot\frac{\lambda^2}{\mu}e^{-(2\mu-\lambda)\xi}\\
&+Y\left(\frac{2\mu}{9\lambda}\right)\frac{2\lambda^3}{\mu}\frac{Q_i(s_2)e^{-s_2 t}\sqrt{8\lambda\mu}\,e^{-2\mu\xi}}{\sqrt{\mu(\mu-4\lambda}\,[2\lambda+\mu+\sqrt{\mu(\mu-4\lambda)}\,]}\\
&\cdot\Im\left\{\left[\frac{(\lambda+\mu-s_2)(\lambda-s_2)e^{i\theta*}}{\lambda^2}-\sqrt{\frac{2\mu}{\lambda}}\left(\frac{\lambda-s_2}{\lambda}\right)\right]e^{\xi\sqrt{2\mu\lambda}\,e^{i\theta*}}\right\}\\
&+e^{-2\mu\xi}\frac{\lambda^3}{2\mu\pi}\sqrt{8\lambda\mu}\int_{\lambda+2\mu-\sqrt{8\lambda\mu}}^{\lambda+2\mu+\sqrt{8\lambda\mu}}\frac{e^{-xt}Q_i(x)}{x[x^2-(2\lambda+\mu)x+\lambda(\lambda+2\mu)]}\\
&\cdot\Im\left\{\left[\frac{(\lambda+\mu-x)(\lambda-x)}{\lambda^2}\,e^{i\theta}-\sqrt{\frac{2\mu}{\lambda}}\left(\frac{\lambda-x}{\lambda}\right)\right]e^{\xi\sqrt{2\mu\lambda}\,e^{i\theta}}\right\}dx
\end{aligned} \tag{6.18}$$

where $\cos\theta^* = \dfrac{\lambda+2\mu-s_2}{\sqrt{8\lambda\mu}}$ and $\cos\theta = \dfrac{\lambda+2\mu-x}{\sqrt{8\lambda\mu}}$.

A busy period of the two server queue can now mean either a time interval during which both servers are busy, or else a time interval during which at least one server is busy.

Considering first the busy period for both servers, suppose the process is initially in state 2, and let T be the time at which the process first reaches state 1. Then $\Pr\{T<t\}$ is the probability of absorption before time t for the second associated process. Now the second associated process is similar to the first associated process of a one server queue in which the parameter μ has been replaced by 2μ. Hence using (6.4),

$$\Pr\{T<t\} = \frac{1}{2\pi\lambda}\int_{\lambda+2\mu-\sqrt{8\lambda\mu}}^{\lambda+2\mu+\sqrt{8\lambda\mu}} \frac{1-e^{-xt}}{x}\sqrt{8\lambda\mu-(\lambda+2\mu-x)^2}\,dx\,. \tag{6.19}$$

The distribution of the time before state 1 is reached when the initial state is k ($k\geqq 2$) can be obtained from (6.6) in a similar way. By another argument of this kind it is seen that the distribution of the number N of customers arriving during the busy period T is obtained by replacing μ with 2μ in (6.15). Thus

$$\Pr\{N=n\} = \frac{2\mu}{\lambda+2\mu}\left(\frac{\sqrt{8\lambda\mu}}{\lambda+2\mu}\right)^{2n}\int_{-1}^{1}\xi^{2n}\sqrt{1-\xi^2}\,d\xi\,. \tag{6.20}$$

Next let us study the time during which at least one server is busy. Thus we suppose the initial state of the process is 1 and we denote by T^* the time when the zero state is first reached. If ψ_1 is the spectral measure of the associated process, then by our previous argument

$$\Pr\{T^*<t\} = \mu\int_0^\infty \frac{1-e^{-xt}}{x}\,d\psi_1(x)\,. \tag{6.21}$$

Now the Stieltjes transforms $B_1(s)$ and $B_2(s)$ of the spectral measures of the first and second associated processes are related by

$$B_1(s) = \frac{1}{\lambda+\mu-s-2\lambda\mu B_2(s)} \tag{6.22}$$

and from (4.5)

$$B_2(s) = \frac{\lambda+2\mu-s-\sqrt{(\lambda+2\mu-s)^2-8\lambda\mu}}{4\lambda\mu}\,. \tag{6.23}$$

Hence

$$B_1(s)=\frac{1}{2\mu}\cdot\frac{\lambda-s-\sqrt{(\lambda+2\mu-s)^2-8\lambda\mu}}{s-(\mu-\lambda)}. \tag{6.24}$$

It follows that $\psi_1(x)$ has the continuous density

$$\psi_1'(x)=\frac{1}{2\pi\mu}\cdot\frac{\sqrt{8\lambda\mu-(\lambda+2\mu-x)^2}}{x-(\mu-\lambda)}$$

on the interval $|\lambda+2\mu-x|\leqq\sqrt{8\lambda\mu}$, and in addition has a jump of magnitude $\frac{\mu-2\lambda}{\mu}$ at $x=\mu-\lambda$ if $\mu>2\lambda$, but has no jump if $\mu\leqq 2\lambda$. Thus (6.21) becomes

$$\Pr\{T^*<t\}=Y\left(\frac{\mu}{2\lambda}\right)(\mu-2\lambda)\frac{1-e^{-t(\mu-\lambda)}}{\mu-\lambda} \tag{6.25}$$
$$+\frac{1}{2\pi}\int_{\lambda+2\mu-\sqrt{8\lambda\mu}}^{\lambda+2\mu+\sqrt{8\lambda\mu}}\frac{1-e^{-xt}}{x}\cdot\frac{\sqrt{8\lambda\mu-(\lambda+2\mu-x)^2}}{x-(\mu-\lambda)}\,dx,$$

where, as usual, $Y(z)=1$ if $z>1$, 0 if $z\leqq 1$.

It is natural to next ask for the distribution of the number N^* of customers arriving during the busy period T^*. This again leads to the study of a random walk on the integers $1, 2, \cdots$ with an ignored absorbing state at 0. The polynomials of the random walk satisfy the recursion relations

$$\begin{cases} R_1(x)=1, \quad R_0(x)=0 \\ xR_1(x)=p_1R_2(x) \\ xR_n(x)=qR_{n-1}(x)+pR_{n+1}(x), \quad n\geqq 2, \end{cases} \tag{6.26}$$

where

$$p=\frac{\lambda}{\lambda+2\mu}, \quad q=\frac{2\mu}{\lambda+2\mu}, \quad p_1=\frac{\lambda}{\lambda+\mu}.$$

To compute the spectral measure α of this random walk we consider also the associated random walk obtained if the given one is stopped whenever state 1 is reached. Denoting the spectral measure of the associated random walk by β we look for a relation between the Stieltjes transforms

$$\int_{-1}^{1}\frac{d\alpha(x)}{x-z}, \quad \int_{-1}^{1}\frac{d\beta(x)}{x-z}$$

analogous to the relation (2.5) for the spectral measures of a birth and death process and its associated process. Such a relation, applicable to a general random walk and its associated process, is proved in [6] and may be stated as follows. If the state space of the random walk is $0, 1, 2, \cdots$ and the one step transition probabilities are

$$(6.27)\qquad P'_{ij}=\begin{cases} q_i & \text{if } j=i-1\ , \\ r_i & \text{if } j=i\ , \\ p_i & \text{if } j=i+1\ , \\ 0 & \text{if } |j-i|>1\ , \end{cases}$$

where $p_i>0$, $q_i>0$, $r_i\geqq 0$, then the spectral measure α of the process and the spectral measure β of the associated process are connected by the identity

$$(6.28)\qquad \int_{-1}^{1}\frac{d\alpha(x)}{x-s}=\frac{1/p_0}{\dfrac{r_0-s}{p_0}-q_1\displaystyle\int_{-1}^{1}\frac{d\beta(x)}{x-s}}\ .$$

Now let α_0 denote the spectral measure of the random walk determined by the polynomials (6.26) ; let α_1 denote the spectral measure of the associated process and α_2 the spectral measure of the second associated process. First applying (6.28) with $\alpha=\alpha_0$, $\beta=\alpha_1$, we get

$$(6.29)\qquad \int_{-1}^{1}\frac{d\alpha_0(x)}{x-s}=\frac{-1}{s+p_1q\displaystyle\int_{-1}^{1}\frac{d\alpha_1(x)}{x-s}}\ .$$

Then applying (6.28) with $\alpha=\alpha_1$, $\beta=\alpha_2$, we get

$$(6.30)\qquad \int_{-1}^{1}\frac{d\alpha_1(x)}{x-s}=\frac{-1}{s+pq\displaystyle\int_{-1}^{1}\frac{d\alpha_2(x)}{x-s}}$$

But clearly $\alpha_1=\alpha_2$ so from (6.30)

$$(6.31)\qquad \int_{-1}^{1}\frac{d\alpha_1(x)}{x-s}=\frac{-s+\sqrt{s^2-4pq}}{2pq}$$

where the radical must be determined by analytic continuation from positive values for $s>1$. Now (6.31) and (6.29) give

$$(6.32)\qquad \int_{-1}^{1}\frac{d\alpha_0(x)}{x-s}=\gamma\frac{(\gamma-1)s-\sqrt{s^2-4pq}}{\{1-(\gamma-1)^2\}s^2-4pq}\ ,$$

where $\gamma=2p/p_1=2(\lambda+\mu)/(\lambda+2\mu)$ and $\gamma-1=\lambda/(\lambda+2\mu)$ is positive and less than one. The Stieltjes inversion formula giving α_0 at all of its points of continuity is

$$(6.33)\qquad \alpha_0(x)=\frac{1}{\pi}\lim_{\eta\to 0+}\int_{-1-\varepsilon}^{x}\Im\left[\int_{-1}^{1}\frac{d\alpha_0(y)}{y-\xi-i\eta}\right]d\xi$$

and of course α has a jump at a point x_0 if and only if its Stieltjes transform has a pole there. A simple computation shows that the right side of (6.32) has no poles if $1 \leqq \gamma < 2$, which is the case in our problem. Thus α_0 consists of the continuous density

$$\alpha_0'(x) = \frac{\gamma}{\pi} \frac{\sqrt{4pq - x^2}}{4pq - \{1 - (\gamma - 1)^2\} x^2} \tag{6.34}$$

on the interval $|x| \leqq \sqrt{4pq}$, with $\gamma = 2p/p_1$. It is easy to express the probability distribution of N^* in terms of α_0, the result being

$$\begin{aligned} \Pr(N^* = n\} &= \frac{\mu}{\lambda + \mu} \int_{-\sqrt{4pq}}^{\sqrt{4pq}} x^{2n} \, d\alpha_0(x) \\ &= \frac{2}{\pi} \frac{\mu}{\lambda + 2\mu} \left(\frac{\sqrt{8\lambda\mu}}{\lambda + 2\mu} \right)^{2n} \int_{-1}^{1} \xi^{2n} \frac{\sqrt{1 - \xi^2}}{1 - \frac{4\mu(\lambda + \mu)}{(\lambda + 2\mu)^2} \xi^2} \, d\xi \, . \end{aligned} \tag{6.35}$$

7. Results concerning the n server queue. The method used in § 5 to compute the spectral measures for the one and two server queues can be used in the same way to compute the spectral measure of a queue with three or more servers. Although the description of the spectral measure ψ in terms of the parameters λ, μ, and the number n of servers becomes more and more complicated as n increases, it is nevertheless possible to deduce certain general features of ψ. These general features are stated without proof in the next paragraph, and the proofs are supplied in the appendix.

The spectral measure ψ of the n server queue consists of a continuous density $\psi'(x)$ on the interval

$$\lambda + n\mu - \sqrt{4n\lambda\mu} < x < \lambda + n\mu + \sqrt{4n\lambda\mu} \, ,$$

and in addition there may be a finite number of isolated jumps. The number of such isolated jumps is one of the integers $0, 1, 2, \cdots, n$ and these jumps all lie in the half-open interval

$$0 \leqq x < \lambda + n\mu - \sqrt{4n\lambda\mu} \, .$$

If $n\mu > \lambda$ there is a jump at $x = 0$ of magnitude ρ given by

$$\frac{1}{\rho} = \sum_{r=0}^{n-1} \frac{1}{r!} \left(\frac{\lambda}{\mu} \right)^r + \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n \frac{n\mu}{n\mu - \lambda} \, , \tag{7.1}$$

but if $n\mu \leqq \lambda$ there is no jump at $x = 0$. We form the polynomial

$$F(\sqrt{b}) = Q_n(\lambda(\sqrt{nb} - 1)^2) - \sqrt{nb} \, Q_{n-1}(\lambda(\sqrt{nb} - 1)^2) \tag{7.2}$$

which is of degree $2n$ in $\sqrt{b}$. It has a zero of order n at $\sqrt{b}=0$ and n simple zeros $\sqrt{b_1}, \sqrt{b_2}, \cdots, \sqrt{b_n}$ with $1=b_1<b_2<\cdots<b_n$. The spectral measure of the n server queueing process has exactly k jumps to the left of $x=\lambda+n\mu-\sqrt{4n\lambda\mu}$ if and only if

$$b_k<\frac{\mu}{2}\leqq b_{k+1}$$

where we take $b_0=0$ and $b_{n+1}=+\infty$.

For the case $n=3$ the three critical values b_k occur at $b=1/3$ and the two roots of $12b^2-112b+147=0$.

In discussing the busy period distributions for an n server queue, one has to distinguish n different cases. In the simplest case, one observes the time interval T during which all n servers are busy—that is, at time zero the process is in state n and T is the first time at which the process is in state n-1. The distribution of T is of course obtained from (6.4) by replacing μ with $n\mu$, so that

$$\text{(7.3)} \qquad \Pr\{T\leqq t\}=\frac{1}{2\pi\lambda}\int_{\lambda+n\mu-\sqrt{4n\lambda\mu}}^{\lambda+n\mu+\sqrt{4n\lambda\mu}}\frac{1-e^{-xt}}{x}\sqrt{4n\lambda\mu-(\lambda+n\mu-x)^2}\,dx$$

is the distribution of a busy period for all n servers. Similarly, the distribution of the number N of customers arriving during a busy period for all n servers is obtained from (6.15):

$$\text{(7.4)} \qquad \Pr\{N=k\}=\frac{n\mu}{\lambda+n\mu}\left(\frac{\sqrt{4n\lambda\mu}}{\lambda+n\mu}\right)^{2k}\frac{2}{\pi}\int_{-1}^{1}\xi^{2k}\sqrt{1-\xi^2}\,d\xi\,.$$

In the next simplest case one observes the time interval T^* during which at least $n-1$ servers are busy—that is, at time zero the process is in state $n-1$ and T^* is the first time at which the process is in state $n-2$. After a computation similar to that in (6.21)—(6.25) we find

$$\text{(7.5)} \qquad \Pr\{T^*<t\}=Y\left(\frac{\mu}{n\lambda}\right)(\mu-n\lambda)\frac{1-e^{-t(n-1)(\mu-\lambda)}}{(n-1)(\mu-\lambda)}$$
$$+\frac{1}{2\pi}\int_{\lambda+n\mu-\sqrt{4n\lambda\mu}}^{\lambda+n\mu+\sqrt{4n\lambda\mu}}\frac{1-e^{-xt}}{x}\cdot\frac{\sqrt{4n\lambda\mu-(\lambda+n\mu-x)^2}}{x-(n-1)(\mu-\lambda)}\,dx$$

where $Y(z)$ has its usual significance. If now N^* is the number of customers arriving during a busy period for $(n-1)$ of the servers, then from (6.34) with

$$p=\frac{\lambda}{\lambda+n\mu},\quad q=\frac{n\mu}{\lambda+n\mu},\quad \gamma=2\frac{\lambda+(n-1)\mu}{\lambda+n\mu}$$

we obtain

$$(7.6)\quad \Pr\{N^*=k\} = \frac{2}{\pi}\frac{(n-1)\mu}{\lambda+n\mu}\left(\frac{\sqrt{4n\lambda\mu}}{\lambda+n\mu}\right)^{2k}\int_{-1}^{1}\xi^{2k}\frac{\sqrt{1-\xi^2}}{1-\dfrac{4[\lambda+(n-1)\mu]}{(\lambda+n\mu)^2}\xi^2}\,d\xi\,.$$

Using the same kind of techniques it is possible to find the distribution of the length of a busy period for m of the n servers, and the distribution of the number of customers arriving during such a busy period.

8. Maximum length of the queue during a busy period.

Consider for the moment a general birth and death process with parameter λ_n, μ_n and with $\mu_0>0$. Suppose the initial state is i and let $j>i$. It was shown in [5] that the probability that absorption at zero occurs without the state j ever being visited is

$$(8.1)\qquad {}_jA_i=\mu_0\int_0^\infty Q_i(x)\frac{d\psi^{(j)}(x)}{x}$$

$$=\frac{\mu_0\sum_{k=i}^{j-1}\dfrac{1}{\lambda_k\pi_k}}{1+\mu_0\sum_{k=0}^{j-1}\dfrac{1}{\lambda_k\pi_k}}$$

where $\psi^{(j)}$ is the spectral measure of the process on the states $0, 1, 2, \cdots, j-1$ which is obtained from the original process by stopping it whenever the j state is reached.

We first use this result to compute the probability $\zeta_{n,j}$ that during a busy period for all servers in an n-server queue the maximum length of the queue is always *less* than $n+j$. This of course is just the probability that when the nth associated process starts out in its zero state, absorption occurs before it ever visits the jth state, and hence (8.1) gives

$$\zeta_{n,j}=\frac{n\mu}{\lambda}\cdot\frac{\left(\dfrac{n\mu}{\lambda}\right)^{j}-1}{\left(\dfrac{n\mu}{\lambda}\right)^{j+1}-1}\,.$$

A similar application of (8.1) to the $(n-1)$ st associated process gives the probability $\zeta^*_{n,j}$ that for the n server queue during a busy period for at least $n-1$ of the servers the length of the queue is always less than $n-1+j$. The result is

$$\zeta^*_{n,j}=\frac{\dfrac{(n-1)\mu}{\lambda}\cdot\dfrac{\left(\dfrac{n\mu}{\lambda}\right)^j-1}{\dfrac{n\mu}{\lambda}-1}}{1+\dfrac{(n-1)\mu}{\lambda}\cdot\dfrac{\left(\dfrac{n\mu}{\lambda}\right)^j-1}{\left(\dfrac{n\mu}{\lambda}\right)-1}}\,.$$

Appendix A. The nature of the spectral measure for the n server queue.

In this section we present the proofs of the statements made concerning the structure of the spectral measure for the n server queue.

Let ψ_0 be the spectral measure of the n server queue, ψ_k be the spectral measure of the kth associated process, and let $B_k(s)$ be the Stieltjes transform defined by (4.3). The relation between B_k and B_{k+1} is

$$B_k(s)=\frac{1}{\lambda+k\mu-s-(k+1)\lambda\mu B_{k+1}(s)} \tag{A.1}$$

and $B_n(s)$ is given by (4.5). In the interval $s_0=\lambda+n\mu-\sqrt{4n\lambda\mu}<s<\lambda+n\mu+\sqrt{4n\lambda\mu}=s_1$ the imaginary part of $B_n(s+i\tau)$ converges to a positive limit as $\tau\to 0^+$. Consequently $\psi_n(x)$, and by induction each $\psi_k(x)$, $0\leqq k\leqq n$, has a continuous spectrum in this interval. From (A.1) it is seen that ψ_k has a jump at each point $x=s$ where the denominator $\lambda+k\mu-s-(k+1)\lambda\mu B_{k+1}(s)$ has a simple zero. These jumps cannot occur in the interior of the interval of the continuous spectrum because there the imaginary part of the denominator is negative, and they cannot occur at the ends of the interval because there $B_n(s)$, and by induction each $B_k(s)$, has singularities which are not poles. From (4.5) it is found that ψ_n has no jumps; in fact $B_n(s)$ increases steadily from zero at $s=-\infty$ to the value $(n\lambda\mu)^{-1/2}$ at $s=s_0$ and increases steadily from the value $-(n\lambda\mu)^{-1/2}$ at $s=s_1$ to zero at $s=+\infty$.

To locate the jumps, if any, of ψ_{n-1}, consider the places where the graph of the straight line $y=\lambda+(n-1)\mu-x$ intersects the graph of $y=n\lambda\mu B_n(x)$. No intersection occurs for $x>s_1$ because

$$\lambda+(n-1)\mu-s_1-n\lambda\mu B_n(s_1)=-\mu-\sqrt{n\lambda\mu}<0\,.$$

Moreover, since

$$\lambda+(n-1)\mu-s_0-n\lambda\mu B_n(s_0)=-\mu+\sqrt{n\lambda\mu}\,,$$

and in view of the monotonicity of the two graphs, there is one intersection to the left of $x=s_0$ if $-\mu+\sqrt{n\lambda\mu}<0$, or equivalently if $\mu>n\lambda$, and no intersection if $\mu \leqq n\lambda$. Thus ψ_{n-1} never has a jump to the right of the continuous spectrum and has one jump to the left of the continuous spectrum if $\mu > n\lambda$, no jump if $\mu \leqq n\lambda$. The jump, if $\mu > n\lambda$, is easily found to be at $x=(n-1)(\mu-\lambda)$.

It will be shown that none of the measures ψ_k have any jumps to the right of the continuous spectrum. This has already been verified for ψ_n and ψ_{n-1}, and we proceed by induction. Suppose it has been established for $k+1 \leqq r \leqq n-1$ that ψ_r has no jumps to the right of s_1 and that $B_r(s_1)$ is finite, negative, and greater than $-(n\lambda\mu)^{-1/2}$. Since $B_{n-1}(s_1) = -(\mu+\sqrt{n\lambda\mu})^{-1}$, the inequality is certainly valid for $r=n-1$. From (A.1) we get

$$B_k(s_1) = \frac{-1}{(n-k)\mu + 2\sqrt{n\lambda\mu} + (k+1)\lambda\mu B_{k+1}(s_1)},$$

and by virtue of the assumed inequality for $r=k+1$ it follows that $B_k(s_1)$ is finite, negative, and greater than $-(n\lambda\mu)^{-1/2}$. Since $B_{k+1}(s)$ increases steadily from its finite negative value at $s=s_1$ to zero at $s=\infty$, it follows that the denominator of (A.1) is not zero for $s>s_1$ and ψ_k has no jump to the right of s_1. This completes the induction.

Now suppose it has been established that for some k, $1 \leqq k+1 \leqq n-1$, and some choice of λ and μ, the measure ψ_{k+1} has exactly r jumps. Let these jumps be at $x_1 < x_2 < \cdots < x_r$. Then $x_r < s_0$ and in each of the r intervals $-\infty < s < x_1$, $x_1 < s < x_2, \cdots, x_{r-1} < s < x_r$, the function $(k+1)\lambda\mu B_{k+1}$ increases steadily to $+\infty$, and thus in each interval its graph intersects the graph of $\lambda + k\mu - s$ exactly once. Consequently ψ_k has exactly one jump in each of these intervals. In the interval $x_r < s < s_0$ the function $(k+1)\lambda\mu B_{k+1}$ increases steadily from $-\infty$ to its possibly finite value at s_0, and in this interval ψ_k has either one or no jumps. Thus ψ_k has at least r and at most $r+1$ jumps. It follows that for any λ, μ the number of jumps of ψ_k is at most $n-k$.

Setting $s=s_0$ in (A.1)

$$B_k(s_0) = \frac{1}{2\sqrt{n\lambda\mu} - (n-k)\mu - (k+1)\lambda\mu B_{k+1}(s_0)}. \tag{A.2}$$

The necessary and sufficient condition that ψ_k have one jump more than ψ_{k+1} is that this expression be negative. Now it follows by induction starting from

$$B_{n-1}(s_0) = \frac{1}{\sqrt{n\lambda\mu} - \mu}$$

that for $k \leqq n-1$,

$$B_k(s_0) \sim -\frac{1}{(n-k)\mu}$$

as $\mu\to\infty$ with λ fixed. Consequently, for any fixed λ, ψ_k has exactly $n-k$ jumps for all sufficiently large μ. On the other hand it follows by induction that for each $k\leqq n-1$,

$$\lim_{\mu\to 0} \sqrt{\mu}B_k(s_0),\ \lambda \text{ fixed },$$

exists and is positive but less than $(n\lambda)^{-1/2}$. Consequently, for any fixed λ, ψ_k has no jumps for all sufficiently small μ.

In order to make a more careful study of the number of jumps of ψ_k we introduce the associated families of polynomials $\{Q_m^{(k)}(x)\}$ defined for $k=-1$ by

$$Q_m^{(-1)}(x)=-\frac{1}{\lambda}Q_m(x)$$

and for $k\geqq 0$ by the recursion formulas

$$\begin{aligned}
&Q_r^{(k)}(x)=0 \quad \text{for} \quad r\leqq k,\\
&Q_{k+1}^{(k)}(x)=-\frac{1}{\lambda}\,,\\
-xQ_r^{(k)}(x)&=\mu_r Q_{r-1}^{(k)}(x)-(\lambda+\mu_r)Q_r^{(k)}(x)+\lambda Q_{r+1}^{(k)}(x),\ r\geqq k+1\,.
\end{aligned}$$

It is seen that except for the constant factor $-(1/\lambda)$, the polynomials $Q_m^{(k)}(x)$ with k fixed are the polynomials belonging to the $(k+1)$th associated process and are orthogonal with respect to ψ_{k+1}. Applying (2.9) to the kth associated process we obtain

$$B_k(s)=\frac{1}{\lambda}\cdot\frac{n\mu Q_{n-1}^{(k)}(s)B_n(s)-Q_n^{(k)}(s)}{n\mu Q_{n-1}^{(k-1)}(s)B_n(s)-Q_n^{(k-1)}(s)}\,.$$

In terms of the variable $b=\mu/\lambda$ we have $s_0=\lambda(1-\sqrt{nb})^2$, $n\mu B_n(s_0)=\sqrt{nb}$. If we let

$$P_r^{(k)}(\sqrt{b})=\frac{-1}{\lambda}b^{-(r-k-1)/2}Q_r^{(k)}(\lambda(1-\sqrt{nb})^2)$$

then

$$B_k(s_0)=\frac{1}{\lambda}\cdot\frac{P_n^{(k)}(\sqrt{b})-\sqrt{n}\,P_{n-1}^{(k)}(\sqrt{b})}{P_n^{(k-1)}(\sqrt{b})-\sqrt{n}\,P_{n-1}^{(k-1)}(\sqrt{b})}\,. \tag{A.3}$$

The quantities $P_r^{(k)}(\xi)$ satisfy

$$\begin{cases} P_k^{(k)}(\xi)=0,\ P_{k+1}^{(k)}(\xi)=1\,,\\ -(n-r)\xi P_r^{(k)}(\xi)=rP_{r-1}^{(k)}(\xi)-2\sqrt{n}\,P_r^{(k)}(\xi)+P_{r+1}^{(k)}(\xi)\,, \end{cases} \tag{A.4}$$

for $k+1\leqq r\leqq n$. By virtue of the form of this recurrence formula it follows that for each fixed k the polynomials $P_r^{(k)}(\xi)$, $k+1\leqq r\leqq n$, form a finite system of orthogonal polynomials and that the polynomials $P_r^{(k+1)}(\xi)$, $k+2\leqq r\leqq n$, are the corresponding associated polynomials. Writing (A.4) in the form

$$\begin{aligned} P_{r+1}^{(k)}(\xi)-\sqrt{n}\,P_r^{(k)}(\xi)=\sqrt{n}\,[P_r^{(k)}(\xi)-\sqrt{n}\,P_{r-1}^{(k)}(\xi)] \\ +(n-r)[P_{r-1}^{(k)}(\xi)-\xi P_r^{(k)}(\xi)] \end{aligned}$$

it is easily shown by induction that all of the polynomials $P_r^k(\xi)$, $P_r^{(k)}(\xi)-\sqrt{n}\,P_{r-1}^{(k)}(\xi)$, $k+1\leqq r\leqq n$, are strictly positive for $\xi\leqq 0$.

Now the polynomial

$$G_n^{(k-1)}(\xi)=P_n^{(k-1)}(\xi)-\sqrt{n}\,P_{n-1}^{(k-1)}(\xi)$$

is a quasi-orthogonal polynomial, of exact degree $n-k$, belonging to the system of polynomials $P_r^{(k-1)}$, and the corresponding associated polynomial is $G_n^{(k)}(\xi)$. Consequently $G_n^{(k-1)}(\xi)$ has exactly $(n-k)$ distinct positive roots, say $\xi_1^{k-1}<\xi_2^{k-1}<\cdots<\xi_{n-k}^{k-1}$, and the $n-k-1$ roots of $G_n^{(k)}(\xi)$ lie one in each of the open intervals $\xi_r^{k-1}<\xi<\xi_{r+1}^{k-1}$.

The quantities $B_k(s_0)$ can be computed by using the recurrence formulas for the polynomials $P_n^{(k)}(\xi)$. In particular

$$\begin{aligned} B_{n-1}(s_0)&=\frac{-1}{\lambda\sqrt{b}}\cdot\frac{1}{(2\sqrt{n}-\sqrt{b}\,)-\sqrt{n}} \\ &=\frac{-1}{\lambda\sqrt{b}\,(\sqrt{n}-\sqrt{b}\,)} \end{aligned}$$

which checks with an earlier computation. Thus we see that the root ξ_1^{n-2} of $G_n^{n-2}(\xi)$ is $\sqrt{n}$, and we already know that ψ_{n-1} has no jump if $\sqrt{b}\leqq\xi_1^{n-2}$ and has one jump if $\sqrt{b}>\xi_1^{n-2}$. Suppose it has been established that ψ_{k+1} has no jump if $\sqrt{b}\leqq\xi_1^k$, has r jumps if $\xi_r^k<\sqrt{b}\leqq\xi_{r+1}^k$ for $r=1, 2, \cdots, n-k-2$, and has $n-k-1$ jumps if $\sqrt{b}>\xi_{n-k-1}^k$. This property is easily extended to ψ_k by induction. In fact ψ_k always has either the same number of jumps or else one more jump than ψ_{k+1}, and it has one more jump than ψ_{k+1} if and only if the expression

$$B_k(s_0)=\frac{1}{\lambda b^{1/2}}\,\frac{G_n^{(k)}(\sqrt{b}\,)}{G_n^{(k-1)}(\sqrt{b}\,)}$$

is finite and negative. The result follows because of the interlacing of the roots of $G_n^{(k-1)}(\xi)$ and $G_n^{(k)}(\xi)$.

Summarizing, the number of jumps of ψ_k to the left of the continuous spectrum is equal to the number of roots of $G_n^{(k-1)}(\xi)$ which are less than $\sqrt{b}$. In particular the number of jumps of ψ_0, the spectral measure of the n server process, is equal to the number of roots of the polynomial

$$G_n^{(-1)}(\xi)=P_n^{(-1)}(\xi)-\sqrt{n}\,P_{n-1}^{(-1)}(\xi)$$
$$=\xi^{-n}[Q_n(\lambda(1-\xi\sqrt{n})^2)-\xi\sqrt{n}\,Q_{n-1}(\lambda(1-\xi\sqrt{n})^2)]$$

which are less than $\sqrt{b}$.

Appendix B. The random walk polynomials derived from the infinitely many server queue.

In § 6 we had occasion to consider, along with a birth and death process, the imbedded random walk. A system of polynomials, which are useful in a variety of problems, arises in this way from the infinitely many server process. These polynomials depend on a parameter $a=\lambda/\mu>0$ and will be denoted by $r_n(x, a)$ or sometimes by $r_n(x)$. The purpose of this appendix is to list their useful properties.

The polynomials are defined by the recursion formulas

$$r_0(x, a)=1\ ,$$
$$xr_0(x, a)=r_1(x, a)\ ,$$
$$(n+a)xr_n(x, a)=nr_{n-1}(x, a)+ar_{n+1}(x, a),\quad n\geqq 1\ .$$

They are orthogonal on the interval $-1\leqq x\leqq 1$ with respect to a measure ψ which consists entirely of jumps. If we let

$$x_k=\sqrt{\frac{a}{k+a}},\quad k=0, 1, 2, \cdots$$

then ψ has equal jumps at x_k and at $-x_k$ of magnitude

$$\sigma_k=\frac{1}{2}\,\frac{a}{k+a}\,\frac{(k+a)^k}{k!e^{k+a}},\quad k=0, 1, \cdots$$

and these have been normalized so the sum of all the jumps is one. The orthogonality relation is

$$\sum_{k=0}^{\infty} r_n(x_k)r_m(x_k)\sigma_k+\sum_{k=0}^{\infty} r_n(-x_k)r_m(-x_k)\sigma_k=\frac{\delta_{n,m}}{\pi_n}$$

where

$$\pi_n=\frac{n+a}{a}\cdot\frac{a^n}{n!}\ .$$

A generating function is

$$\sum_{n=0}^{\infty} r_n(x, a)\frac{z^n}{n!}=e^{z/x}\left(1-\frac{xz}{a}\right)^{a(1-x^2)/x^2}\ ,$$

and from this explicit representations of the polynomials can be obtained. For example

$$r_n(x, a)=\frac{1}{x^n}c_n\left(a\,\frac{1-x^2}{x^2},\ \frac{a}{x^2}\right)$$

where $c_n(x, a)$ are the Poisson-Charlier polynomials.

REFERENCES

1. N. T. J. Bailey, *A continuous time treatment of a simple queue using generating functions*, J. Roy. Stat. Soc., Ser. B **16** (1954), 288-291.
2. A. Erdélyi, et al, *Higher transcendental functions*, McGraw-Hill vol. II (1953).
3. W. Feller, *An introduction to probability theory and its applications*, Wiley (1950).
4. S. Karlin and J.McGregor, *The differential equations of birth-and-death processes, and the Stieltjes moment problem*, Trans. Amer. Math. Soc., **85** (1957), 489-546.
5. ———, *The classification of birth and death processes*, Trans. Amer. Math. Soc., **86** (1957), 366-400.
6. ———, *Representation theory for random walks*, to be published.
7. D. G. Kendall, *Stochastic processes in the theory of queues*. Ann. Math. Stat., **24** (1953) 333-354.
8. J. Kiefer and J. Wolfowitz, *On the theory of queues with many servers*, Trans. Amer. Math. Soc., **78** (1955), 1-18.
9. W. Lederman and G. E. H. Reuter, *Spectral theory for the differential equations of simple birth-and-death processes*, Phil. Trans. Roy. Soc., London-Ser. A vol. 246, No. 914 (1954), 321-369.
10. L. Takács, *Investigation of waiting time problems by reduction to Markov processes*, *Acta Math*. Acad. Sci. Hung., **6** (1955).
11. E. C. Titchmarsh, *Theory of Fourier integrals*, Oxford (1948).
12. D. V. Widder, *The Laplace transform*, Princeton (1941).

STANFORD UNIVERSITY

CURVATURE IN HILBERT GEOMETRIES

PAUL KELLY AND ERNST STRAUS

For every pair of points, p and q, interior to a simple, closed, convex curve C in the Euclidean plane, the line $\xi = p \times q$ cuts C in a pair of points u and v. If C has at most one segment then the Hilbert distance from p to q, defined by

$$h(p, q) = \left|\log\left(\frac{up}{uq} \cdot \frac{vq}{vp}\right)\right| ,$$

is a proper metric (where up denotes the Euclidean distance from u to p), and is invariant under projective transformations. The geometry induced on the interior of C is a Hilbert geometry, and the Hilbert lines are carried by Euclidean lines [2].

We shall be concerned here with curvature at a point defined in a qualitative rather than a quantitative sense (cf. [1, p 237]).

DEFINITION 1. The *curvature at p* is *positive* or *negative* if there exists a neighborhood U of p such that for every x, y in U we have

$$2\ h(\bar{x}, \bar{y}) \geqq h(x, y) ,$$

respectively

$$2\ h(\bar{x}, \bar{y}) \leqq h(x, y) ,$$

where $\bar{x}$, $\bar{y}$ are the Hilbert midpoints respectively of the segments from p to x and p to y. If there is neither positive nor negative curvature at a point then the curvature is *indeterminate* at that point. This qualitative curvature is clearly a projective invariant.

In order to state our result we need one more concept.

DEFINITION 2. A point p is a *projective center* of C if there exists a projective transformation, π, of the plane so that πp is the affine center of πC.

A projective center is characterized by the following. Let ξ be a line through p, and let $\xi \cap C = \{u, v\}$, and let p'_ξ be the harmonic conjugate of p with respect to u and v. Finally, let L_p be the locus of all p'_ξ. Then p is a projective center if and only if L_p is a straight line.

Conic sections are characterized by the fact that every point in their interior is a projective center [3]. We can now state our main result, which solves a problem of H. Busemann [1, Problem 34, p. 406].

THEOREM. *If p is a point of determinate curvature then it is* *a projective center of C. In particular, if the curvature is determinate everywhere then C is an ellipse and the Hilbert geometry is hyperbolic.*

We first establish some lemmas.

LEMMA 1. *For any point p, interior to C, there exists* a *line η* (possibly the line at infinity) *which intersects L_p in at least two points and does not intersect C.*

Proof. There is at least one chord of C which is bisected by p. If ξ_1 is the line of such a chord then ξ_1 intersects L_p at q_1 on the line at infinity. If L_p has a second point at infinity then the line at infinity satisfies the lemma. If L_p has only one point at infinity then L_p is a connected curve. It cannot lie within the strip formed by the two supporting lines of C which are parallel to ξ_1 for then it would intersect C. There is therefore a point q_2 of L_p outside this strip and the line $\eta = q_1 \times q_2$ satisfies the lemma.

COROLLARY. *For every p in the interior of C there exists a projective transformation, π, so that πC is a closed, convex curve, and so that πp is the midpoint of two mutually perpendicular chords of πC whose endpoints are points of differentiability of πC.*

Proof. Since all but a denumerable set of points of C are points of differentiability, we may choose the line η of Lemma 1 so that $\eta \cap L_p$ contains p'_{ξ_1} and p'_{ξ_2} and so that C is differentiable at its points of intersection with ξ_1 and ξ_2. Now let π_1 be a projective transformation which maps η into the line at infinity, and let π_2 be an affine transformation which maps $\pi_1 \xi_1$ and $\pi_1 \xi_2$ into perpendicular lines. Then $\pi = \pi_2 \pi_1$ has the required properties.

LEMMA 2. *If a chord of C, of* (Euclidean) *length $2k$, has p for its midpoint and if q is a neighboring point on the chord at* (Euclidean) *distance ds from p, then $dS = (2/k)\, ds + O(ds^3)$, where $dS = h(p, q)$.*

Proof. If the endpoints of the chord are u and v, and the order of the points on the chords is u, p, q, v, then, by definition,

$$\begin{aligned} dS &= \log\left(\frac{up+pq}{up}\right)\left(\frac{vp}{vp-pq}\right) = \log\left(\frac{k+ds}{k}\right)\left(\frac{k}{k-ds}\right) \\ &= \log\left(1+\frac{ds}{k}\right) - \log\left(1-\frac{ds}{k}\right) \\ &= \left[\frac{ds}{k} - \frac{1}{2}\left(\frac{ds}{k}\right)^2 + \frac{1}{3}\left(\frac{ds}{k}\right)^3 - \cdots\right] \\ &\qquad - \left[-\frac{ds}{k} - \frac{1}{2}\left(\frac{ds}{k}\right)^2 - \frac{1}{3}\left(\frac{ds}{k}\right)^3 - \cdots\right] \end{aligned}$$

$$= \frac{2}{k} ds + O(ds^3) .$$

LEMMA 3. *Let (r, θ) be polar coordinates whose pole p is an interior point of C at which the curvature is determinate. If C is differentiable at the ends of two perpendicular chords which bisect each other at p, then C satisfies the "one-sided" differential relations*

$$\begin{aligned} \frac{d}{d\theta}\left(\frac{\csc 2\theta}{r^2}\right)\bigg|_{\theta_0^+} &= \frac{d}{d\theta}\left(\frac{\csc 2\theta}{r^2}\right)\bigg|_{(\theta_0+\pi)^+} \\ \frac{d}{d\theta}\left(\frac{\csc 2\theta}{r^2}\right)\bigg|_{\theta_0^-} &= \frac{d}{d\theta}\left(\frac{\csc 2\theta}{r^2}\right)\bigg|_{(\theta_0+\pi)^-} \end{aligned} \tag{1}$$

for all θ_0.

Proof. We first introduce Cartesian coordinates, with origin p, so that the y-axis intersects C at points of second order differentiability, and so that the axes do not coincide with the two given chords bisected by p. The curve C is then given by an "upper" arc $y=y_1(x)$ and a "lower" arc $y=-y_2(x)$. Let the bisected chords lie on the lines $\xi_1: y=ax$ and $\xi_2: y=(1/a)x$ respectively. Let $b_1=(dx, a\,dx)$ and $c_1=(2\,dx, 2a\,dx)$ on ξ_1, and $b_2=(dx, -(1/a)\,dx)$ and $c_2=(2\,dx, -(2/a)\,dx)$ on ξ_2, where dx is positive and chosen so that b_1, b_2, c_1, and c_2 lie inside C. Assume that p is a point of negative curvature. Then.

$$2\,h(m_1, m_2) \leqq h(c_1, c_2). \tag{2}$$

where m_i is the Hilbert midpoint of the segment from p to c_i.

To show that $h(m_i, b_i)=O(dx^3)$, we define $dS_1=h(p, c_1)$ and $ds_1=pc_1$. With $2k$ representing the Euclidean length of the chord on ξ_1, it follows from Lemma 2 that $dS_1=(2/k)\,ds_1+O(ds_1^3)$, and hence that

$$h(p, m_1)=\frac{1}{2}dS_1=\frac{1}{k}ds_1+O(ds_1^3) . \tag{3}$$

Also, from Lemma 2 and the relation $ds_1=2\,pb_1$, it follows that

$$h(p, b_1)=\frac{2}{k}pb_1+O[(pb_1)^3]=\frac{1}{k}(ds_1)+O(ds_1^3) . \tag{4}$$

Since $h(m_1, b_1)=|h(p, m_1)-h(p, b_1)|$, equations (3) and (4) imply that $h(m_1, b_1) =O(ds_1^3)$. But $ds_1=dx(1+a^2)^{1/2}=O(dx)$, hence $h(m_1, b_1)=O(dx^3)$. Similarly, $h(m_2, b_2)=O(dx^3)$, and therefore

$$h(m_1, b_1)+h(m_2, b_2)=O(dx^3) . \tag{5}$$

From the triangle inequality,

$$h(m_1, m_2) \geqq h(b_1\, b_2) - h(m_1, b_1) - h(m_2, b_2)\ . \tag{6}$$

This, together with (5), yields

$$h(m_1, m_2) \geqq h(b_1, b_2) - O(dx^3)\ , \tag{7}$$

and from (1) and (7) we obtain

$$2\ h(b_1, b_2) < h(c_1, c_2) + O(dx^3)\ . \tag{8}$$

We now wish to calculate the distances in (8). First, we have

$$\begin{aligned} h(b_1, b_2) &= h[(dx, a\,dx),\ (dx, -\frac{1}{a}dx)] \\ &= \log\left[\frac{y_1(dx)+\frac{1}{a}dx}{y_1(dx)-a\,dx}\cdot\frac{y_2(dx)+a\,dx}{y_2(dx)-\frac{1}{a}dx}\right] \\ &= \log\left[1+\frac{dx}{a\,y_1(dx)}\right]+\log\left[1+\frac{a\,dx}{y_2(dx)}\right] \\ &\qquad -\log\left[1-\frac{a\,dx}{y_1(dx)}\right]-\log\left[1-\frac{dx}{a\,y_2(dx)}\right]\ . \end{aligned} \tag{9}$$

Using the Maclaurin expansion of the logarithms, and collecting first and second degree terms, we obtain

$$\begin{aligned} h(b_1, b_2) &= dx\left(a+\frac{1}{a}\right)\left[\frac{1}{y_1(dx)}+\frac{1}{y_2(dx)}\right] \\ &\qquad + \frac{dx^2}{2}\left(a^2-\frac{1}{a^2}\right)\left[\frac{1}{y_1^2(dx)}-\frac{1}{y_2^2(dx)}\right]+O(dx^3)\ . \end{aligned} \tag{10}$$

Because both of the functions $y_1(x)$ and $y_2(x)$ are convex and have second derivatives at $x=0$, they can be represented in the form

$$y_i(dx) = y_i(0) + y_i'(0)dx + O(dx^2)\ , \qquad i=1, 2, \tag{11}$$

and hence

$$\begin{aligned} \frac{1}{y_i(dx)} &= \frac{1}{y_i(0)} - \frac{y_i'(0)}{y_i^2(0)}dx + O(dx^2) \\ \frac{1}{y_i^2(dx)} &= \frac{1}{y_i^2(O)} + O(dx)\ . \end{aligned} \tag{12}$$

The substitution of (12) in (10) gives

$$\begin{aligned} h(b_1, b_2) &= dx\left(a+\frac{1}{a}\right)\left[\frac{1}{y_1}-\frac{y_1'dx}{y_1^2}+\frac{1}{y_2^2}-\frac{y_1'dx}{y_2^2}\right] \\ &\qquad + \frac{dx^2}{2}\left(a^2-\frac{1}{a^2}\right)\left(\frac{1}{y_1^2}-\frac{1}{y_2^2}\right)+O(dx^3)\ , \end{aligned} \tag{13}$$

where $y_i = y_i(0)$. Hence

$$2\,h(b_1,b_2) = 2\,dx\left(a+\frac{1}{a}\right)\left[\frac{1}{y_1}+\frac{1}{y_2}-\frac{y_1'dx}{y_1^2}-\frac{y_2'dx}{y_2^2} + \frac{dx}{2}\left(a-\frac{1}{a}\right)\left(\frac{1}{y_1^2}-\frac{1}{y_2^2}\right)\right]+O(dx^3)\,. \tag{14}$$

By the substitution of $2\,dx$ for dx we obtain

$$h(c_1,c_2) = 2\,dx\left(a+\frac{1}{a}\right)\left[\frac{1}{y_1}+\frac{1}{y_2}-\frac{2\,y_1'\,dx}{y_1^2}-\frac{2\,y_2'\,dx}{y_2^2} + dx\left(a-\frac{1}{a}\right)\left(\frac{1}{y_1^2}-\frac{1}{y_2^2}\right)\right]+O(dx^3)\,. \tag{15}$$

Substituting this and (14) in (8) we have

$$\begin{aligned} &2dx\left(a+\frac{1}{a}\right)\left[\frac{1}{y_1}+\frac{1}{y_2}-\frac{y_1'dx}{y_1^2}-\frac{y_2'dx}{y_2^2} + \frac{dx}{2}\left(a-\frac{1}{a}\right)\left(\frac{1}{y_1^2}-\frac{1}{y_2^2}\right)\right] \\ &< 2\,dx\left(a+\frac{1}{a}\right)\left[\frac{1}{y_1}+\frac{1}{y_2}-\frac{2\,y_1'\,dx}{y_1^2}-\frac{2\,y_2'\,dx}{y_2^2} + dx\left(a-\frac{1}{a}\right)\left(\frac{1}{y_1^2}-\frac{1}{y_2^2}\right)\right]+O(dx^3)\,. \end{aligned} \tag{16}$$

By dividing both sides of this inequality by $2dx(a+1/a)$, and then rearranging the terms, we obtain

$$dx\left(\frac{y_1'}{y_1}+\frac{y_2'}{y_2}\right)-\frac{dx}{2}\left(a-\frac{1}{a}\right)\left(\frac{1}{y_1^2}-\frac{1}{y_2^2}\right)<O(dx^2)\,. \tag{17}$$

Division of both sides of (17) by dx yields a new inequality whose right side is $O(dx)$ but whose left side is independent of dx. From this it follows that

$$\frac{y_1'}{y_1}+\frac{y_2'}{y_2}-\frac{1}{2}\left(a-\frac{1}{a}\right)\left(\frac{1}{y_1^2}-\frac{1}{y_2^2}\right)\leqq 0\,. \tag{18}$$

Consider now a reflection in the y-axis taking C into a curve $\overline{C}$ which is divided by the x-axis into an "upper" arc $z=z_1(x)$ and a "lower" arc $z=-z_2(x)$. With the lines $z=(1/a)$ and $z=-ax$ playing the roles of ξ_1 and ξ_2, and with $\bar{b}_1$, $\bar{c}_1$, $\bar{b}_2$, $\bar{c}_2$ defined respectively by $(dx, (1/a)\,dx)$, $(2\,dx, (2/a)\,dx)$, $(dx, -a\,dx)$, and $(2\,dx, -2a\,dx)$, a repetition of the former argument leads to

$$\frac{z_1'}{z_1^2}+\frac{z_2'}{z_2^2}-\frac{dx}{2}\left(\frac{1}{a}-a\right)\left(\frac{1}{z_1^2}-\frac{1}{z_2^2}\right)\leqq 0\,. \tag{19}$$

Since $z_i=y_i$ and $z_i'=-y_i'$, (19) is also

(20) $$-\frac{y_1'}{y_1^2}-\frac{y_2'}{y_2^2}+\frac{dx}{2}\left(a-\frac{1}{a}\right)\left(\frac{1}{y_1^2}-\frac{1}{y_2^2}\right)\leqq 0 .$$

Combining the opposite inequalities (18) and (20), we obtain

(21) $$\frac{y_1'}{y_1^2}+\frac{y_2'}{y_2^2}-\frac{1}{2}\left(a-\frac{1}{a}\right)\left(\frac{1}{y_1^2}-\frac{1}{y_2^2}\right)=0 .$$

Since (21) is an equality, it is clear that the same result would have been obtained if all preceding inequalities has been reversed. In other words (21) holds if p is a point of determinate curvature.

To express (21) in polar coordinates, let the polar axis be ξ_1 and let θ_0 designate the angle between the polar axis and the upper half-line of the y-axis. The angles of inclination to the x-axis of the tangent lines to C at $(0, y_1)$ and $(0, y_2)$ are α_1 and α_2 respectively and the clockwise angles from the radius vectors to the tangent lines at these points are ω_1 and ω_2. From the standard relationships between polar and Cartesian coordinates, it follows that

(22) $$y_1'(0)=\tan\alpha_1=-\cot\omega_1=\left[-\frac{1}{r}\frac{dr}{d\theta}\right]\theta_0$$
$$y_2'(0)=-\tan\alpha_2=\cot\omega_2=\left[\frac{1}{r}\frac{dr}{d\theta}\right]\theta_0+\pi$$

Also, by definition, $a=\cot\theta_0$ so $\frac{1}{2}\left(a-\frac{1}{a}\right)=\cot 2\theta_0$. Substituting this and (22) in (21) we obtain

(23) $$\left[-\frac{1}{r^3}\frac{dr}{d\theta}\right]_{\theta_0}+\left[\frac{1}{r^3}\frac{dr}{d\theta}\right]_{\theta_0+\pi}-(\cot 2\theta_0)\left[\frac{1}{r^2(\theta_0)}-\frac{1}{r^2(\theta_0+\pi)}\right]=0 ,$$

and hence

(24) $$\left[\frac{1}{r^3}\frac{dr}{d\theta}+\frac{1}{r^2}\cot 2\theta\right]_{\theta_0}=\left[\frac{1}{r^3}\frac{dr}{d\theta}+\frac{1}{r^2}2\theta\right]_{\theta_0+\pi} .$$

Multiplying both sides of (24) by $2\,csc 2\theta_0=2\,csc 2(\theta_0+\pi)$ we have

(25) $$\frac{d}{d\theta}\frac{(csc\,2\theta)}{r^2}\bigg|_{\theta_0}=\frac{d}{d\theta}\frac{(csc\,2\theta)}{r^2}\bigg|_{\theta_0+\pi} .$$

Since (25) involves only first derivatives, it holds for all θ_0 for which r is differentiable at both θ_0 and $\theta_0+\pi$. Since the one-sided derivative exists everywhere, we get the desired relations in (1), for all θ_0, from the semi-continuity of the one sided derivative.

Proof of the Theorem. According to the corollary of Lemma 1 there is always a projective transformation such that, after the transformation,

p satisfies the conditions of Lemma 3. From (1) we obtain

$$\int_{\theta_0}^{\theta} d\left(\frac{csc\ 2\theta}{r^2}\right)=\int_{\theta_0+\pi}^{\theta+\pi} d\left(\frac{csc\ 2\theta}{r^2}\right), \tag{26}$$

where the integrals are Stieltjes intergrals and the interval (θ_0, θ) does not contain a multiple of $\pi/2$. Hence

$$\frac{1}{r^2(\theta)}=\frac{1}{r^2(\theta+\pi)}+k_j\sin 2\theta,\ k_j=\text{constant} \tag{27}$$

where $$(j-1)\frac{\pi}{2}\leqq\theta\leqq j\frac{\pi}{2},(j=1, 2, 3, 4)\ .$$

Since r is differentiable at the points for which $\theta=0,\pi/2, \pi,3\pi/2$, we obtain from (27), upon differentiation at these points, the relations $k_1=k_2=k_3=k_4$. On the other hand, if we replace θ by $\theta+\pi$ in (27) we obtain the relations $k_1=-k_3$, and $k_2=-k_4$. In other words, $k_j=0$ and $r(\theta)=r(\theta+\pi)$. Since this shows p to be a metric center, it was initially a projective center.

The last statement in the theorem is well known (see [3] and e.g. [2, p.164]).

If a Hilbert metric is defined in the interior of an n-dimensional, convex surface S, the definitions for curvature and projective centers are unchanged. The metric for the space induces, on any plane through an interior point p, a two-dimensional Hilbert geometry. If p is a point of determinate curvature, it is a two-dimensional projective center for every plane through it. Since the L_p locus for every plane section is a line, it is easily seen that the total L_p locus must be a plane and hence that p is a projective center of S. If curvature is determinate everywhere then S is an ellipsoid and the geometry is hyperbolic.

It seems probable that a Hilbert geometry can contain no points of positive curvature.

References

1. H. Busemann, *The geometry of geodesics*, Academic Press, N.Y. 1957.

2. ———, and P. Kelly, *Projective geometry and projective metrics*, Academic Press, N.Y. 1953.

3. T. Kajima, *On a characteristic property of the conic and the quadric,*, Sci. Rep., Tohoku Univ. 8, 1919

University Of California, Santa Barbara And Los Angeles.

STATIONARY MEASURES FOR CERTAIN STOCHASTIC PROCESSES

JOHN LAMPERTI

Introduction. In a recent paper [1], T.E. Harris has studied stationary processes $\{Z_n\}$ with a finite number of states, taken to be the integers $0, 1, \cdots, D-1$. His technique is to map the half-infinite sample sequences $Z_n, Z_{n-1}, \cdots$ onto the unit interval by means of the correspondence

$$X_{n+1} = Z_n/D + Z_{n-1}/D^2 + \cdots . \tag{1}$$

The X_n then form a stationary Markov process. In § 5 of [1] Harris shows (Theorem 7) that if the process $\{Z_n\}$ is of mixing type, then either the stationary distribution $G(x) = \Pr(X_n \leqq x)$ has a unit step, or is the uniform distribution, or $G(x)$ is continuous and totally singular.

The purpose of this paper is to investigate correspondences such as (1) in general, using two simple lemmas in ergodic theory which are given in the next section. If $g(\{i_0, i_1, \cdots\})$ is any essentially one-to-one and measurable mapping of the space of sequences $\{i_0, i_1, \cdots\}$ onto another measurable space X, then a correspondence similar to (1) may be defined between stochastic processes with states i and processes on X:

$$X_{n+1} = g(\{Z_n, Z_{n-1}, \cdots\}) . \tag{2}$$

Theorem 1 describes the resulting distributions on X; Theorem 2 is a specialization to the case of (1). Finally an additional application (Theorem 3) is made to certain of the processes studied by Karlin in [3]. Theorem 2 contains Theorem 7 of [1], and Theorem 3 overlaps with § 7 of [3]. In addition to a unified approach, some extension of the previous results is obtained in both cases.

2. Ergodic theory lemmas.

LEMMA 1. *Let (Ω, W) be a measurable space and T a measurable transformation of Ω onto itself. Let μ_1 and μ_2 be two sigma-finite measures on (Ω, W) such that for each, T is a measure preserving, metrically-transitive transformation. Then if μ_1 and μ_2 are not proportional, they are orthogonal (i.e., have their positive mass on disjoint sets).*

Proof. Suppose μ_1 and μ_2 are both finite measures, and assume they have been normalized. Let A be a set such that $\mu_1(A) \neq \mu_2(A)$. Define

$$B_i = \left\{ \omega \in \Omega \,|\, \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \phi_A(T^j\omega) = \mu_i(A) \right\} , \qquad i = 1, 2,$$

Received July 8, 1957. This paper is based on a technical report prepared under contract Nonr 220 (16) with the Office of Naval Research.

where $\phi_A(\cdot)$ is the characteristic function of the set A. By the individual ergodic theorem, $\mu_i(B_i)=1$. Hence B_1 and $B_1^c \supset B_2$ are a decomposition of Ω demonstrating the orthogonality of μ_1 and μ_2. In the case where one or both of the measures are infinite, the same idea may be carried out using Hopf's ratio ergodic theorem (see, for instance, [2]).

Now let Ω be the $\prod_{i=-\infty}^{\infty} Y_i$ of sequences $\{\omega_i\}$ where $\omega_i \in Y_i$ with each $Y_i=Y$, Y a measurable space. Let W be the Borel field generated by the "cylinder sets" of Ω[2]. Denote by S the "shift" transformation

$$S\{\omega_i\}=\{\nu_i\} \text{ where } \nu_i=\omega_{i+1}\,. \tag{3}$$

LEMMA 2. *Let μ be a measure on (Ω, W) such that S is measure preserving and metrically-transitive and $\mu\ (\Omega)=1$. Then one of the following is the case*:

(a) *there is a finite sequence* $a_1, a_2, \cdots, a_m$ *of points of Y such that μ has mass* $\frac{1}{\text{m}}$ *on each of the m points of Ω given by*

$$\omega_i=a_j \text{ for } i\equiv j+k \bmod m\,;\ k=0, 1, \cdots, m-1.$$

(b) $\mu\{\omega \in \Omega | \omega_0=a_0, \omega_{-1}=a_1, \cdots\}=0$ *for any sequence* $\{a_n\}$ *of points of Y.*

Proof. Suppose that case (a) does not hold. Then we shall show that $\mu(A)=0$, where

$$A=\bigcup_{n=-\infty}^{\infty} A_n,\ A_n=\{\omega \in \Omega | \omega_n=a_0\omega_{n-1}=a_1, \cdots\}\,.$$

Now A is invariant under S, and so $\mu(A)=0$ or 1. Assume $\mu(A)=1$. It is not hard to see that if $\{a_i\}$ were a periodic sequence, case (a) would hold. But if $\{a_i\}$ is not periodic, a value of n such that $\omega_n=a_0$, $\omega_{n-1}=a_1, \cdots$ must be unique, and so the A_n are disjoint. Since S is measure preserving, $\mu(A_n)$ are all equal. This contradicts the assumption that $\mu(A)=1$.

Finally we remark that, speaking somewhat less precisely, Lemma 2, may be re-expressed as : A stationary ergodic stochastic process either executes deterministically a certain periodic motion, or else each path function has probability 0. We shall refer to these alternatives as case *a* and *b*.

3. Induced Markov processes. In this section we continue to use the notation $\Omega=\prod_{i=-\infty}^{\infty} Y_i$, but Y is restricted to be a fixed (not necessarily

[2] In the applications we shall make of this lemma, Y will be restricted (in fact, will be denumerable), so that the Kolmogorov extension theorem will hold.

finite) set of integers. Let Ω_0 denote the space of half-infinite sequences $\{i_0, i_1, \cdots\}$ where $i_n \in Y$; the measurable sets of both Ω and Ω_0 are again those belonging to the Borel fields generated by the cylinder sets. If $\omega^{(0)}=\{i_0, i_1, \cdots\}$ is an element of Ω_0, the notation $\{i, \omega^{(0)}\}$ will mean the sequence $\{i, i_0, i_1, \cdots\}$.

Let (X, F) be a measurable space, and suppose that there exists a mapping $g(\omega^{(0)})$ of Ω_0 onto X which is one to one if sets at most denumerable are deleted from Ω_0 and from X; suppose also that both g and g^{-1} are measurable. Let μ be any probability measure on the space Ω, and let $\{Z_n\}$ mean the stochastic process consisting of the random variables

$$Z_n(\omega)=i_n \ . \tag{4}$$

A new process $\{X_n\}$, with state-space X, may be defined by (2) and (4).

LEMMA 3. *Assume that for each particular sequence $\omega^{(0)}$ in Ω_0, and for each n,*

$$\Pr(Z_n=i_0, Z_{n-1}=i_1, \cdots)=0 \ . \tag{5}$$

Then $\{X_n\}$ is a Markov process whose (not necessarily stationary) transition probabilities are given with probability one by

$$\begin{aligned} &X_{n+1}=g[\{i, g^{-1}(X_n)\}] \text{ with probability} \\ &f_i(X_n)=\Pr[Z_n=i|\{Z_{n-1}, Z_{n-2}, \cdots\}=g^{-1}(X_n)] \ . \end{aligned} \tag{6}$$

Proof. Let $E\subset\Omega$ denote the set of all "path functions" for the $\{Z_n\}$ process such that some segment $\{Z_n, Z_{n-1}, \cdots\}$ belongs to the (denumerable) set which must be deleted from Ω_0 in order to secure a one-to-one map onto X; it follows from (5) that $\mu(E)=0$. Therefore with probability one, knowledge of X_n determines the sequence $\{Z_{n-1}, Z_{n-2}, \cdots\}=g^{-1}(X_n)$ uniquely. Then X_n also determines $X_{n-1}, X_{n-2}, \cdots$ and so the process is Markovian. That (6) gives the transition law is clear. There are, of course, many cases where the Markov property and (6) hold even though (5) does not.

Consider now measures μ such that the shift operation (3) s measure preserving and metrically-transitive ; in other words, measures such that $\{Z_n\}$ is a stationary, ergodic stochastic process. In this case, $\{X_n\}$ will also be stationary ; let Q_μ denote the stationary probability measure of (each) X_n.

THEOREM 1. *If $\{Z_n\}$ executes deterministically a cycle of period m (case a), then Q_μ concentrates its mass upon at most m points of X. Otherwise (case b) $\{X_n\}$ is a Markov process, the measure Q_μ is non-atomic,* *and any two measures of this type resulting from different μ's are orthogonal.*

Proof. In case a, the measure μ concentrates on m points, and so Q_μ will concentrate on the images of these points, which may or may not all be distinct. (If they are distinct, then $\{X_n\}$ must be a Markov process.) Otherwise, (case b), it follows from Lemma 2 that (5) holds, and hence that Q_μ is non-atomic. Lemma 3 then implies that $\{X_n\}$ is a Markov process. Under the mapping g the relation of orthogonality of non-atomic measures is preserved, and so Lemma 1 yields the remaining assertion of the theorem.

A remark about infinite measures will conclude this section. Suppose the shift operation is measure-preserving and metrically-transitive for sigma-finite measures μ on Ω which have the property (b) of Lemma 2. Let Q_μ denote the perhaps infinite measures which are then induced by g on (X, F). Lemma 1 is still available, and so the conclusion of orthogonality of distinct Q_μ's remains valid.

4. Applications.

EXAMPLE 1. We now consider the particular case studied in [**1**]. The set Y will consist of the intergers $0, 1, \cdots, D-1$, and (X, F) will be the unit interval and the field of Borel sets. Let

$$g(\{i_0, i_1, \cdots\}) = i_0/D + i_1/D^2 + \cdots . \tag{7}$$

Then the correspondence between a process $\{Z_n\}$ with states Y and a process $\{X_n\}$ is given by (1). In this situation we have

THEOREM 2. *Let $\{Z_n\}$ be a metrically-transitive, stationary process with state-space Y; let $G(x) = \Pr(X_n \leqq x)$ be the stationary distribution function of X_n. Then $\{X_n\}$ is always a Markov process, and one of the following holds:*

(i) *$\{Z_n\}$ executes deterministically a cycle of period m; in this case, $G(x)$ has m discontinuities each of leap $1/m$.*

(ii) *Each Z_n is independently uniformly distributed $\{0, 1, \cdots, D-1\}$. In this case $G(x) = x$, $0 \leqq x \leqq 1$.*

(iii) *$G(x)$ is continuous and singular with respect to Lebesgue measure. Finally, any two continuous distributions $G(x)$ are orthogonal.*

Proof. The fact that $\{X_n\}$ is always a Markov process, and the statement (i), follow since the mapping (7) cannot map two sequences of Ω_0 having positive measure into the same point. Statement (ii) is easily verified, and then the remainder of the theorem follows from Theorem 1.

EXAMPLE 2. In this application, the $\{X_n\}$ process is the primary object of interest; it is a type of learning model [3]. Let Y consist of the integers 0 and 1, and again take (X, F) as the unit interval and Borel field. Let σ and α be two numbers between 0 and 1 such that $\sigma+\alpha\leqq 1$. (The present approach does not seem to apply when $\sigma+\alpha>1$.) Define inductively a family of subintervals of the unit interval as follows:

$$A(\{0\})=[0, \sigma],\ A(\{1\})=[1-\alpha, 1],$$

and if $\omega_m^0=\{i_0, i_1, \cdots, i_{m-1}\}$, then

$$A(\{0, \omega_m^0\})=\sigma A(\omega_m^0) \text{ and } A(\{1, \omega_m^0\})=1-a+aA(\omega_m^0).$$

Now since both $\sigma<1$ and $\alpha<1$, for any sequence $\omega^{(0)}$ the $A(\omega_m^0)$ are a sequence of nested intervals of length approaching zero. Therefore the following definition is meaningful:

$$g(\{i_0, i_1, \cdots\})=\bigcap_{m=1}^{\infty} A(\omega_m^0)\ . \tag{8}$$

Let $\{Z_n\}$ be a stationary stochastic process with states Y, and define $\{X_n\}$ by (2) and (8). Let $G(x)$ again denote the stationary distribution of the X_n.

THEOREM 3. *$\{X_n\}$ is a Markov process with transition law*

$$X_{n+1}=\begin{cases}\sigma X_n \\ 1-\alpha+\alpha X_n\end{cases} \text{with probability } \begin{matrix} f_0(X_n) \\ f_1(X_n)=1-f_0(X_n)\ . \end{matrix} \tag{9}$$

Any stationary Markov process of this form is induced by some process $\{Z_n\}$: If $\{Z_n\}$ is in addition metrically transitive, one of the following cases must hold:

(i) *$\{Z_n\}$ executes deterministically a cycle of period m; $G(x)$ has m discontinuities each of leap $1/m$.*

(ii) *$G(x)=x$; this occurs if and only if $\sigma+\alpha=1$ and $f_0(x)\equiv\sigma$, $f_1(x)\equiv 1-\sigma$.*

(iii) *$G(x)$ is continuous and singular with respect to Lebesque measure.*

Any two continuous distributions $G(x)$ arising from different metrically-transitive processes $\{Z_n\}$ but the same mapping g (that is, the same σ and α) are orthogonal.

Proof. If $\sigma+a<1$, then the mapping g is not onto the whole unit interval, but onto a cantor-like subset of measure zero; it is precisely one-to-one onto this set. Therefore X_n must be a Markov process, and the transition law (9) is obtained from (6) and (8). The continuity of $G(x)$ if (i) does not hold follows from Lemma 2. Since $G(x)$ is a distribution on a set of measure zero, it must automatically be a singular distribution for any process $\{Z_n\}$.

Now suppose $\sigma+\alpha=1$. In this case there are some points x corresponding under g^{-1} to two points of Ω_0; however, just as in Theorem 2 the ambiguities of the mapping do not affect either the Markov property or the m distinct discontinuities of statement (i). Statement (ii) is readily verified, and then (iii) follows from Theorem 1. Whether $\sigma+\alpha=1$, or <1, the last statement of the theorem also follows from Theorem 1.

If $\{X_n\}$ is a stationary Markov process of the form (9), the distribution $G(x)=\Pr(X_n \leqq x)$ concentrates positive mass only on the domain of g^{-1}; hence a stationary measure is induced by g^{-1} on Ω_0, which extends to a measure μ on Ω. The process $\{Z_n\}$ inducing $\{X_n\}$ is then defined by (4). This completes the proof.

References

1. T. E. Harris, *On chains of infinite order*, *Pacific J. of Math*, **5** (1955), 707–724.
2. E. Hopf, *Ergodentheorie*, Chelsea, 1948.
3. S. Karlin, *Some random walks arising in learning models I*, *Pacific J. Math.* **3** (1953), 725–756.

California Institute of Technology

DISTRIBUTIVITY AND THE NORMAL COMPLETION OF BOOLEAN ALGEBRAS

R. S. PIERCE

1. Introduction. In a recent paper, [4], Smith and Tarski studied the interrelations between completeness and distributivity properties of a Boolean algebra. Independently, the author also obtained some of the results of Smith and Tarski. This work was reported in [2]. The present paper continues the study of distributivity in Boolean algebras. Specifically, it deals with the problem of imbedding a Boolean algebra B in an α-distributive, β-complete algebra, α and β being infinite cardinal numbers. If it is required that the imbedding be regular, that is, preserve existing joins and meets, then (see [3]) the problem is equivalent to the question of when the normal completion of B (or a subalgebra of the completion) is α-distributive. Our two main results can be briefly stated as follows:

THEOREM 3.1. *Every α-distributive Boolean algebra can be regularly imbedded in an α-complete, α-distributive Boolean algebra.*

THEOREM 5.1. *There exists an α-field of sets whose normal completion is not α-distributive.*

Between these principal results, we obtain two simple conditions, one of which is necessary, the other sufficient for the normal completion of a Boolean algebra to be α-distributive. These appear naturally as particular cases of more general facts relating properties which are similar to, but not identical with α-distributivity and β-completeness.

2. Preliminary results. The notation of this paper will be the same as that of [2]. The Greek letters α, β and γ always denote cardinal numbers, while ρ, σ and τ are used as indices belonging to sets R, S and T respectively. The symbol ∞ will be used as though it were a largest cardinal. This is a notational convenience, and in no case involves questionable logic. As in [2], a subset A of an arbitrary Boolean algebra B is called a *covering* (of B) if the least upper bound of A in B is the unit u of B. If the elements of the covering A are disjoint, then A is termed a *partition*. Finally, if the covering (partition) A is of cardinality less than, or equal to α, symbolically $\overline{\overline{A}} \leqq \alpha$, then A is called an α-covering (respectively, α-partition). If A and $\tilde{A}$ are subsets of B, then $\tilde{A}$ is said to refine A when every $\tilde{a} \in \tilde{A}$ is $\leqq$ some $a \in A$.

Received December 28, 1956. In revised form July 11, 1957.

DEFINITION 2.1 (Smith-Tarski). A Boolean algebra B is called (α, β)-distributive if

$$\wedge_{\sigma\in S}\vee_{\tau\in T}b_{\sigma\tau}=\vee_{\varphi\in F}\wedge_{\sigma\in S}b_{\sigma\varphi(\sigma)}\ ,\ F=T^{S}$$

holds identically when $\overline{\overline{S}}\leqq\alpha$, $\overline{\overline{T}}\leqq\beta$ and the bounds are assumed to exist in B.

Some elementary consequences of this definition are worth noting:

(2.2) If B is (α, β)-distributive and $\alpha'\leqq\alpha$, $\beta'\leqq\beta$, then B is (α', β')-distributive. Any regular[1] subalgebra of an (α, β)-distributive Boolean algebra is (α, β)-distributive. Every Boolean algebra is (n, β)-distributive, where n is finite and β is arbitrary.

The last assertion of (2.2) is a variant of the Tarski-von Neumann theorem (see [1], p. 165). This infinite distributivity is a property of Boolean algebras which we use repeatedly and without mention.

A useful characterization of (α, β)-distributive Boolean algebras is given by the following theorem, which, in somewhat different terms, appears in [4]. Since this characterization is used often in the sequal, we sketch a proof.

THEOREM 2.3. *Let α and β be arbitrary cardinal numbers. A Boolean algebra B is (α, β)-distributive if and only if, for any family $\{A_\sigma|\sigma\in S\}$ of β-coverings of B with $\overline{\overline{S}}\leqq\alpha$, there is a covering of B which refines every A_σ.*

Proof. Suppose B is (α, β)-distributive. Let $\{A_\sigma|\sigma\in S\}$ be a given family of β-coverings with $\overline{\overline{S}}\leqq\alpha$. It can be assumed that every A_σ is indexed by the same set T: $A_\sigma=\{a_{\sigma\tau}|\tau\in T\}$. Let $A=\{a\in B|\{a\}$ refines every $A_\sigma\}$. Clearly A refines every A_σ. If A is not a covering of B, there exists $b\neq 0$ (the zero of B) which is disjoint from every $a\in A$. Setting $b_{\sigma\tau}=a_{\sigma\tau}\wedge b$, it is easy to see that $\wedge_\sigma\vee_\tau b_{\sigma\tau}=b>0=\vee_\psi\wedge_\sigma b_{\sigma\psi(\sigma)}$. This contradicts (α, β)-distributivity. Thus A is a covering.

Conversely, let B satisfy the condition of the theorem. Suppose $\vee_{\tau\in T}b_{\sigma\tau}$, $\wedge_{\sigma\in S}\vee_{\tau\in T}b_{\sigma\tau}=b$ and $\wedge_{\sigma\in S}b_{\sigma\varphi(\sigma)}$ exist for all $\sigma\in S$ and all $\varphi\in F=T^S$. Let ω be a symbol not in T. Put $T'=T\cup\{\omega\}$, $b_{\sigma\omega}=b'$, $\overline{A}_\sigma=\{b_{\sigma\tau}|\tau\in T'\}$, $b_\varphi=\wedge_{\sigma\in S}b_{\sigma\varphi(\sigma)}$ for all $\varphi\in F$. Then each $\overline{A}_\sigma$ is a β-covering, so by assumption there is a covering A which refines every $\overline{A}_\sigma$. If $a\in A$, then either $a\leqq b_\varphi$ for some $\varphi\in F$, or else $a\leqq b'$. Thus, if $c\geqq b_\varphi$ for all φ, $c\vee b'\geqq$ l.u.b. $A=u$ (the unit of B). Hence, $c\geqq b$. Since b is obviously an upper bound of all b_φ, it follows that $b=\wedge_{\varphi\in F}b_\varphi$.

For simplicity, an (α, α)-distributive B. A. is just called α-distributive.

[1] A subalgebra $\overline{B}$ of a Boolean algebra B is called regular (see [3]) if, whenever a = l.u.b. A in $\overline{B}$ ($a\in\overline{B}$, $A\leqq\overline{B}$), then a= l.u.b. A in B also. Of course, in a Boolean algebra, this property implies its dual and conversely.

COROLLARY 2.4. *A Boolean algebra B is α-distributive if and only if every family $\{A_\sigma | \sigma \in S\}$ of binary partitions with $\overline{\overline{S}} \leq \alpha$ has a common refining covering.*

Indeed, if $\{\overline{A}_\sigma | \sigma \in S\}$ $(\overline{\overline{S}} \leq \alpha)$ is a family of α-coverings, say $\overline{A}_\sigma = \{a_{\sigma\tau} | \tau \in T\}$, then, setting $A_{\sigma\tau} = [a_{\sigma\tau}, (a_{\sigma\tau})']$, the set $\{A_{\sigma\tau} | \sigma \in S, \tau \in T\}$ is a family of no more than α binary partitions of B and any covering which refines all $A_{\sigma\tau}$ is a common refinement of all $\overline{A}_\sigma$ (because $\wedge_\tau (a_{\sigma\tau})' = 0$.

For future reference, we list some of the well known properties of the normal completion (or completion by "cuts") of a Boolean algebra. The Stone-Glivenko theorem ((2.5) below) is proved in the standard reference [1]. The proofs of (2.6) to (2.8) are conveniently collected in [3].

(2.5) (Stone-Glivenko) The normal completion of a Boolean algebra is a Boolean algebra.

(2.6) Let $\overline{B}$ be the normal completion of the Boolean algebra B. Then B is a regular subalgebra of $\overline{B}$.

(2.7) Any Boolean algebra B is dense in its normal completion $\overline{B}$. That is, if $0 \neq \overline{b} \in \overline{B}$, then there exists $b \in B$ with $0 \neq b \leq \overline{b}$.

(2.8) If the Boolean algebra B is a dense subset of the complete Boolean algebra $\overline{B}$, then $\overline{B}$ is isomorphic to the normal completion of B. Moreover, if $B \subset \tilde{B} \subset \overline{B}$ and $\tilde{B}$ is complete, then $\tilde{B} = \overline{B}$.

DEFINITION 2.9. Let B be a Boolean algebra. Let $\overline{B}$ be the normal completion of B. Let α be an infinite cardinal number. The *normal α-completion* of B is the intersection of all α-complete subalgebras of $\overline{B}$ which contain B. Denote this algebra B^α. It will also be convenient to write B^∞ for $\overline{B}$.

Clearly, B^α is the smallest α-complete subalgebra of B^∞ containing B. Moreover, B is dense in B^α and is regularly imbedded in B^α.

3. The imbedding theorem.

The primary purpose of this section is to prove Theorem 3.1 (stated in the introduction). However, the method of the proof is used several times in the following sections, so it behooves us to present it in a form which is sufficiently general to cover all future needs.

LEMMA 3.2. *Let $\overline{B}$ be a complete Boolean algebra. Let $\mathfrak{A}$ be a non-empty family of partition of $\overline{B}$ such that if $\{A_\sigma | \sigma \in S\} \subseteqq \mathfrak{A}$ and $\overline{\overline{S}} \leq \alpha$, then some $A \in \mathfrak{A}$ refines every A_σ. Let $\tilde{B}$ be the set of all joins of subsets of the partitions A in $\mathfrak{A}$. Then $\tilde{B}$ is an α-complete Boolean algebra such* *that* $A \subseteqq \tilde{B}$ *for every* $A \in \mathfrak{A}$ *and every* α*-covering of* $\tilde{B}$ *is refined by some* $A \in \mathfrak{A}$. *Hence,* $\tilde{B}$ *is* α*-distributive.*

Proof. If $C \subseteqq A \in \mathfrak{A}$, then (l.u.b. $C)' =$ l.u.b. $(A - C)$, since A is a partition. Hence $\tilde{B}$ is closed under complementation. Suppose $\{c_\sigma | \sigma \in S\}$ is a subset of $\tilde{B}$ with $\overline{\overline{S}} \leqq \alpha$. By definition of $\tilde{B}$, for each $\sigma \in S$, there exists a partition $A_\sigma \in \mathfrak{A}$ and a subset $C_\sigma \subseteqq A_\sigma$ such that $c_\sigma =$ l.u.b. C_σ. Then A_σ refines the binary partition $\{c_\sigma, (c_\sigma)'\}$. Let $A \in \mathfrak{A}$ be a common refinement of all A_σ. Then A is a common refinement of all $\{c_\sigma, (c_\sigma)'\}$ and g.l.b. $\{c_\sigma | \sigma \in S\} =$ l.u.b. $\{a \in A | a \leqq c_\sigma$ all $\sigma \in S\} \in \tilde{B}$. Indeed, $c =$ l.u.b. $\{a \in A | a \leqq c_\sigma$, all $\sigma \in S\} \leqq \vee_{\sigma \in S} c_\sigma$ is clear. But also, $c' =$ l.u.b. $\{a \in A | a \leqq (c_\sigma)'$, seme $\sigma \in S\} \leqq \wedge_{\sigma \in S} (c_\sigma)' = (\wedge_{\sigma \in S} c_\sigma)'$. Hence, $\tilde{B}$ is an α-complete B.A. Obviously, $A \subseteqq \tilde{B}$ for all $A \in \mathfrak{A}$. If $\tilde{A}$ is an α-covering of $\tilde{B}$, then, as proved above, every binary partition $\{c, c'\}$ with $c \in \tilde{A}$ is refined by some $A_c \in \mathfrak{A}$. Choosing $A \in \mathfrak{A}$ to be a refinement of all these A_c gives a refinement of $\tilde{A}$. In fact, any $a \in A$ satisfies either $a \leqq c$, or $a \leqq c'$ for all $c \in \tilde{A}$. If $a \leqq c'$ for every c, than $a \leqq \wedge_{c \in \tilde{A}} c' = ($l.u.b. $\tilde{A})' = 0$, since $\tilde{A}$ is a covering. Thus every $a \in A$ satisfies $a \leqq c$ for some $c \in \tilde{A}$.

Proof of (3.1). Let $\overline{B}$ be the normal completion of B. Let $\mathfrak{A}$ be the set of all partitions of $\overline{B}$, which are of the form $\Pi_{\sigma \in S} A_\sigma = \{b_\varphi | \varphi \in 2^S\}$, where the $A_\sigma = \{a_{\sigma 0}, a_{\sigma 1}\}$ are binary partitions of B and $b_\varphi = \wedge_{\sigma \in S} a_{\sigma \varphi(\sigma)} \in \overline{B}$. The fact that $\Pi_{\sigma \in S} A_\sigma$ is a partition follows directly from the assumed distributivity of B. If $A_\tau = \Pi_{\sigma \in S(\tau)} A_{\sigma\tau} \in \mathfrak{A}$ for all $\tau \in T$ with $\overline{\overline{T}} \leqq \alpha$, then $A = \Pi_{\tau \in T} \Pi_{\sigma \in S(\tau)} A_{\sigma\tau} \in \mathfrak{A}$ is a common refinement of all A_τ. Thus, the hypotheses of (3.2) are satisfied. Consequently, there is an α-complete, α-distributive Boolean algebra $\tilde{B}$ with $B \subseteqq \tilde{B} \subseteqq \overline{B}$. Since B is a regular subalgebra of $\overline{B}$, it is also a regular subalgebra of $\tilde{B}$.

4. Conditions for distributivity. In this section, we will examine the following five properties of a Boolean algebra B:

(I_α) B is α-complete;
(II_α) every subset of an α-partition of B has a l.u.b. in B;
(III_β) every β-covering of B can be refined by a β-partition;
($IV_{\alpha\beta}$) B is (α, β)-distributive;
($V_{\alpha\beta}$) If $\{A_\sigma | \sigma \in S\}$ is a set of β-partitions of B with $\overline{\overline{S}} \leqq \alpha$, then there is a covering of B which is a common refinement of every A_σ.

Certain relations between these properties are more or less evident.

(4.1) (a) I_α and II_α are hereditary in α, that is, I_α implies I_γ and II_α implies II_γ for all $\gamma \leqq \alpha$;

(b) $IV_{\alpha\beta}$ and $V_{\alpha\beta}$ are hereditary in both α and β;

(c) I_α implies II_α;

(d) $IV_{\alpha\beta}$ implies $V_{\alpha\beta}$;

(e) $V_{\alpha\beta}$ and III_β together imply $IV_{\alpha\beta}$;
(f) if I_α holds for all $\alpha<\beta$, then III_β, is satisfied;
(g) $IV_{\alpha\alpha}$ is equivalent to $V_{\alpha 2}$ and hence to $V_{\alpha\alpha}$;
(h) III_∞ is always satisfied, so $IV_{\alpha\infty}$ is equivalent to $V_{\alpha\infty}$.

Proofs. The statements (a)-(e) are obvious. If B is α-complete for all $\alpha<\beta$, and $A=\{a_\xi\}$ is a β-covering of B indexed by the set of all ordinals ξ of cardinality less than β, then $\{c_\xi|\bar{\xi}<\beta\}$ will be a β-partition refining A if $c_\xi=a_\xi\wedge(\vee_{\eta<\xi}a_\eta)'$. The assertion of (g) is a restatement of (2.4). Finally, with the help of Zorn's lemma, it is always possible to construct a partition to refine any covering. This construction, the details of which we omit, proves (h).

It appears from (4.1) (e)-(h) that the condition $V_{\alpha\beta}$ is only slightly weaker than $IV_{\alpha\beta}$. On the other hand, the condition II_α is substantially weaker than I_α, as the following example indicates. Let X be a set of cardinality β; let B be the Boolean algebra of finite subsets of X and their complements. If α is any cardinal number less than β, then any α-partition of B is finite. Consequently, B satisfies II_α. In one case however, the properties I_α and II_α are equivalent, namely:

(4.2) II_∞ is equivalent to I_∞.

Proof. Let C be an arbitrary subset of B. Let $C'=\{d\in B|d\wedge c=0$, all $c\in C\}$. Then clearly, u is the only upper bound of the set $C\cup C'$, that is, $C\cup C'$ is a cover. By (4.1) (h), there is a partition A refining $C\cup C'$. If $D=\{a\in A|\{a\}$ refines $C\}$, then $A-D=\{a\in A|a\wedge c=0$, all $c\in C\}$. Hence l.u.b. $C=$l.u.b. D exists by II_∞.

It is appropriate now to explain the object of studying the various properties listed above. Our main interest, of course, is the relation between I_∞ and $IV_{\alpha\alpha}$, and specifically we would like to find simple necessary and sufficient conditions for the normal completion of a Boolean algebra to satisfy $IV_{\alpha\alpha}$. It is rather easy to prove that $IV_{\alpha\infty}$ is sufficient and $IV_{\alpha\,\exp(\alpha)}$ is necessary for α-distributivity in B^∞. The effort to fit these two facts into a broader pattern leads to consideration of conditions II_β and $V_{\alpha\beta}$. It turns out that properties II_β and $V_{\alpha\beta}$ are tied together rather closely. Unfortunately I_β and $IV_{\alpha\beta}$ do not enjoy such an intimate relationship and the two conditions mentioned above are the more or less accidental offspring of II_β and $V_{\alpha\beta}$ rather than the progeny of I_β and $IV_{\alpha\beta}$.

THEOREM 4.3. *If the Boolean algebra B satisfies $V_{\alpha\beta}$ and II_γ, where $\gamma=\beta^\alpha$, then B satisfies $V_{\alpha\gamma}$.*

Proof. The theorem is trivial if α is finite, so it will be assumed that α is an infinite cardinal number. Let A be a γ-partition of B. Then A can be indexed by a subset of T^S, where $\overline{\overline{T}}=\beta$ and $\overline{\overline{S}}=\alpha$, say $A=\{a_\varphi\}$. Since B satisfies II_γ, it is meaningful to define $b_{\sigma\tau}=$l.u.b. $\{a_\varphi|\varphi(\sigma)=\tau\}$ for each $\sigma\in S$, $\tau\in T$. Then $A_\sigma=\{b_{\sigma\tau}|\tau\in T\}$ is a β-partition of B and it is easy to see that any common refinement of all A_σ is also a refinement of A. Now suppose $\{A_\rho|\rho\in R\}$ is a set of γ-partitions of B and $\overline{\overline{R}}\leqq\alpha$. For each ρ in R, define (as above) a set of β-partitions $\{A_{\rho\sigma}|\sigma\in S_\rho\}$ with the property that a common refinement of every $A_{\rho\sigma}$ with $\sigma\in S_\rho$ is also a refinement of A_ρ. Consider the set of all β-partitions $\{A_{\rho\sigma}|\sigma\in S_\rho,\ \rho\in R\}$. There are at most $\alpha^2=\alpha$ of these, so by property $V_{\alpha\beta}$, there is a covering A which refines every $A_{\rho\sigma}$. But then A refines every A_ρ. Thus, B satisfies $V_{\alpha\gamma}$.

COROLLARY 4.4 (Smith-Tarski [4]). *If B is α-distributive and 2^α-complete, then B is $(\alpha,\ 2^\alpha)$-distributive.*

COROLLARY 4.5. *A necessary condition that B^β be α-distributive, where $\beta\geqq 2^\alpha$, is that B be $(\alpha,\ 2^\alpha)$-distributive.*

Indeed, if B^β is α-distributive, then by (4.4) it is $(\alpha, 2^\alpha)$-distributive. But B is a regular subalgebra of B^β and hence (by (2.2)) B is also $(\alpha, 2^\alpha)$-distributive.

We do not know whether the converse of 4.5 holds. That is, if B is $(\alpha, 2^\alpha)$-distributive, does it follow that B^{2^α} is α-distributive? This seems doubtful, but if the goal of 2^α-completeness (that is, property I_{2^α}) is replaced by the property II_{2^α}, then a positive result is obtained (in Corollary 4.8 below).

THEOREM 4.6. *Let B be an arbitrary Boolean algebra. Define $\overline{B}$ to be the intersection of all algebras $\tilde{B}$ with the property II_β such that $B\subseteqq\tilde{B}\subseteqq B^\infty$. Then $\overline{B}$ satisfies II_β. Moreover, $\overline{B}$ has property $V_{\alpha\beta}$ if and only if B has property $V_{\alpha\beta}$. Also, if B is a-complete and satisfies $V_{\alpha\beta}$, where $\beta^\alpha=\beta$, then $\overline{B}$ is α-complete.*

Proof. Clearly $\overline{B}$ satisfies II_β. Since B is a regular subalgebra of $\overline{B}$, the property $V_{\alpha\beta}$ for $\overline{B}$ implies the same property for B. To establish the converse, it is sufficient to show that every β-partition of $\overline{B}$ can be refined by a β-partition of B.

Let $\mathfrak{A}$ be the set of all β-partitions of B. By (2.5), every $A\in\mathfrak{A}$ can be considered as a partition of B^∞. By (2.2), every finite subset of $\mathfrak{A}$ has a common refinement in $\mathfrak{A}$. Let $\tilde{B}$ be the set of all joins in B^∞ of a subset of some $A\in\mathfrak{A}$. By (3.2), $\tilde{B}$ is a Boolean algebra containing B. Clearly $\tilde{B}\subseteqq\overline{B}$. Suppose $\tilde{A}$ is a β-partition of $\tilde{B}$, say $\tilde{A}=\{a_\tau|\tau\in T\}$. Then

$a_\tau = \bigvee \{b_{\sigma\tau} | \sigma \in S_\tau\}$ with $b_{\sigma\tau} \in B$, $b_{\sigma\tau} \wedge b_{\sigma'\tau} = 0$ for $\sigma \neq \sigma'$, and $\overline{\overline{S}}_\tau \leqq \beta$. Consequently, $A = \{b_{\sigma\tau} | \sigma \in S_\tau, \tau \in T\}$ is a β-partition of B which refines $\tilde{A}$. The join of any subset of $\tilde{A}$ is also the join of a subset of A and therefore in $\tilde{B}$. Since $\tilde{A}$ was an arbitrary β-partition, $\tilde{B}$ has property II_β. Consequently, $\overline{B} \leqq \tilde{B}$. Thus every β-partition of $\overline{B} = \tilde{B}$ can be refined by a β-partition of B.

Finally, suppose B is α-complete and satisfies $V_{\alpha\beta}$, with $\beta^\alpha = \beta$. If $\{A_\sigma | \sigma \in S\}$, $\overline{\overline{S}} \leqq \alpha$ is a set of β-partitions of B, then $\prod_{\sigma \in S} A_\sigma = \{\bigwedge_{\sigma \in S} b_\sigma | b_\sigma \in A_\sigma\}$ is a $\beta^\alpha = \beta$-partition. Hence, by (3.2), $\tilde{B} = \overline{B}$ is α-complete.

COROLLARY 4.7. *The normal completion of a Boolean algebra B is (α, ∞)-distributive if and only if B is (α, ∞)-distributive.*

Proof. By (4.6), (4.1) and (4.2).

COROLLARY 4.8. *If the continuum hypothesis is true for the infinite cardinal α (that is, 2^α covers α), then an α-complete Boolean algebra B can be regularly imbedded in an α-complete, α-distributive algebra satisfying II_{2^α} if and only if B is $(\alpha, 2^\alpha)$-distributive.*

Proof. The sufficiency of $(\alpha, 2^\alpha)$-distributivity is a consequence of (4.6) and (4.1). The necessity follows from (4.3), (4.1) and (2.2).

5. **An example.** Because of (4.5), the Theorem (5.1) of the introduction can be proved by constructing an α-field which is not $(\alpha, 2^\alpha)$-distributive.

Let X be a set of cardinality 2^α. Denote by Y the set of all ordinal numbers of cardinality less than α. Let Z be the set of all *bounded* functions in Y^X, that is, functions f for which there is an $\eta \in Y$ such that $f(x) < \eta$ for all x in X. Let $\mathscr{L}$ be the collection of all sets of the form

$$L = L_{W,\varphi} = \{f \in Z | f|_W = \varphi\} ,$$

where $W \subseteqq X$, $\overline{\overline{W}} \leqq \alpha$ and $\varphi \in Y^W$. It is obvious that $\mathscr{L}$ contains the empty set and is closed under α-intersections.

Let $\mathscr{F}$ be the α-field generated by $\mathscr{L}$. It is to be shown that $\mathscr{F}$ is not $(\alpha, 2^\alpha)$-distributive. The proof hinges on a lemma, which is useful in its own right.

LEMMA 5.2. *Let Z be a set. Suppose $\mathscr{L}$ is a nonempty family of subsets of Z with the following properties:*
(i) *every α-intersection of sets in $\mathscr{L}$ is in $\mathscr{L}$;*
(ii) *the complement of any set of $\mathscr{L}$ is a union of sets of $\mathscr{L}$.*
Let $\mathscr{F}$ be the α-field generated by $\mathscr{L}$. Then $\mathscr{L}$ is dense in $\mathscr{F}$.

Proof. Let $\overline{\mathfrak{B}}$ be the complete B. A. of all subsets of Z. Let $\mathfrak{A}$ be the collection of all partitions A of $\overline{\mathfrak{B}}$ with $A\subseteqq\mathscr{L}$. If $\{A_\sigma|\sigma\in S\}\subseteqq\mathfrak{A}$, $\overline{\overline{S}}\leqq\alpha$ and say $A_\sigma=\{L_{\sigma\tau}|\tau\in T_\sigma\}$, then by (i), $\prod_{\sigma\in S}A_\sigma=\{\cap_{\sigma\in S}L_{\sigma\varphi(\sigma)}|\varphi\in\prod_{\sigma\in S}T_\sigma\}$ is in $\mathfrak{A}$ and refines every A_σ. Let $\tilde{\mathfrak{B}}$ consist of all sets $V\subseteqq Z$ such that both V and V^c are disjoint unions of set of $\mathscr{L}$. By (3.2) and (ii), $\mathscr{L}\subseteqq\tilde{\mathfrak{B}}+\overline{\mathfrak{B}}$ and $\tilde{\mathfrak{B}}$ is an α-field. Thus, $\mathscr{L}\subseteqq\mathscr{F}\subseteqq\tilde{\mathfrak{B}}$. Since every set of $\tilde{\mathfrak{B}}$ is a union of sets of $\mathscr{L}$, the same is true of $\mathscr{F}$ and in particular, $\mathscr{L}$ is dense in $\mathscr{F}$.

We now proceed to prove that $\mathscr{F}$ is not $(\alpha, 2^\alpha)$-distributive. For each pair (x,η) with $x\in X$ and $\eta\in Y$, define $T_{(x,\eta)}=\{f\in Z|f(x)=\eta\}$. Clearly $T_{(x,\eta)}\in\mathscr{L}$. For each $\eta\in Y$, let $A_\eta=\{T_{(x,\eta)}|x\in X\}$. The argument is completed by showing

(1) A_η is a 2^α-covering of $\mathscr{F}$;

(2) no covering of $\mathscr{F}$ refines every A_η.

Proof of (1). Evidently, $\overline{\overline{A}}_\eta=2^\alpha$, so the only thing to prove is that the l.u.b. of A_η in $\mathscr{F}$ is Z. The first step is to show that the conditions (i) and (ii) of (5.2) are fulfilled, so that $\mathscr{L}$ is dense in $\mathscr{F}$. Condition (i) is clear. For condition (ii), let $L=L_{W,\varphi}\in\mathscr{L}$. Then $L^c=\cup_{x\in W}\{f\in Z|f(x)\neq\varphi(x)\}=\cup_{x\in W}(\cup\{T_{(x,\eta)}|\eta\neq\varphi(x)\})$ is a union of sets of $\mathscr{L}$.

Since $\mathscr{L}$ is dense in $\mathscr{F}$, it is enough, in proving (1), to show that if $L\in\mathscr{L}$ satisfies $L\cap T_{(x,\eta)}=\phi$ for all x, then $L=\phi$. Suppose $L\neq\phi$ and say $L=L_{W,\varphi}$. Pick $f\in L$ and let $x\in X-W$. Define $g\in Z$ by $g(x)=\eta$, $g(y)=f(y)$ if $y\neq x$. Then $g\in T_{(x,\eta)}$ and $g\in L$. Hence, $L\cap T_{(x,\eta)}\neq\phi$, which is the required conclusion.

Proof of (2). First note that $\cap_{\eta\in Y}(\cup A_\eta)=\phi$. For otherwise there would be an $f\in Z$ whose range included every $\eta\in Y$, contray to the boundedness of the functions of Z. But if A is a subset of $\mathscr{F}$ which refines every A_η, then $\cup A\subseteqq\cup A_\eta$ for all η. Hence, $\cup A\subseteqq\cap_{\eta\in Y}(\cup A_\eta)=\phi$, so A cannot be a covering.

References

1. G. Birkhoff, *Lattice theory*, revised edition, 1948.
2. R. S. Pierce, *Distributivity in Boolean algebras*, Pacific J. Math. **5** (1957).
3. R. Sikorski, *Products in Boolean algebras*, Fund. Math. **37** (1950), 25–54.
4. E. C. Smith and A. Tarski, *Higher degrees of completeness and distributivity in Boolean algebras*, Trans. Amer. Math. Soc. **84** (1957), 230–257.

UNIVERSITY OF WASHINGTON

TRANSCENDENTAL ADDITION THEOREMS FOR THE HYPERGEOMETRIC FUNCTION OF GAUSS

F. M. RAGAB

1. Introduction. In this paper, integrals involving products of two Gauss functions, regarded as functions of their parameters, are evaluated it terms of other functions of the same kind. In all these integrals it is assumed that $|x|<1$. Also the integrals are taken up the entire length of the imaginary axis with loops, if necessary, to separate the increasing and decreasing sequences of poles. These formulae are:

$$(1)\quad \frac{1}{2\pi i}\int\frac{\Gamma(\alpha+s)\Gamma(\alpha'-s)}{\Gamma(\gamma+s)\Gamma(\gamma'-s)}F(\alpha+s,\beta;\gamma+s;x)F(\alpha'-s,\beta';\gamma'-s;x)\,ds$$
$$=\frac{\Gamma(\alpha+\alpha')\Gamma(\gamma+\gamma'-\alpha-\alpha'-1)}{\Gamma(\gamma+\gamma'-1)\Gamma(\gamma-\alpha)\Gamma(\gamma'-\alpha')}F(\alpha+\alpha',\beta+\beta';\gamma+\gamma'-1;x)\,,$$

where $\gamma+\gamma'-1\neq 0,-1,-2,\cdots;$ $\gamma-\alpha\neq 0,-1,-2,\cdots$ and $\gamma'-\alpha'\neq 0,-1,-2,\cdots;$

$$(2)\quad \frac{1}{2\pi i}\int\frac{\Gamma(\alpha+s)\Gamma(\alpha'-s)}{\Gamma(\beta+s)\Gamma(\beta'-s)}F(\beta-\alpha,\gamma;\beta+s;x)F(\beta'-\alpha',\gamma';\beta'-s;x)\,ds$$
$$=\frac{\Gamma(\alpha+\alpha')\Gamma(\beta+\beta'-\alpha-\alpha'-1)}{\Gamma(\beta-\alpha)\Gamma(\beta'-\alpha')\Gamma(\beta+\beta'-1)}F(\beta+\beta'-\alpha-\alpha'-1,\gamma+\gamma';\beta+\beta'-1;x)\,,$$

where $\beta-\alpha\neq 0,-1,-2,\cdots;$ $\beta'-\alpha'\neq 0,-1,-2,\cdots$ and $\beta+\beta'-1\neq 0,-1,-2,\cdots;$

$$(3)\quad \frac{1}{2\pi i}\int\Gamma(\alpha+s)\Gamma(\alpha'+s)\Gamma(\beta-s)\Gamma(\beta'-s)F\Big(\alpha+s,\beta-s;\frac{1}{2}\alpha+\frac{1}{2}\beta+\frac{1}{2};x\Big)$$
$$\times F\Big(\alpha'+s,\beta'-s;\frac{1}{2}\alpha'+\frac{1}{2}\beta'+\frac{1}{2};x\Big)ds$$
$$=\Gamma(\alpha+\beta')\Gamma(\alpha'+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha'+\beta')\{\Gamma(\alpha+\alpha'+\beta+\beta'\}^{-1}$$
$$\times F\Big(\alpha+\beta',\alpha'+\beta;\frac{1}{2}\alpha+\frac{1}{2}\alpha'+\frac{1}{2}\beta+\frac{1}{2}\beta'+\frac{1}{2};x\Big)\,,$$

where $\alpha+\alpha'+\beta+\beta'\neq 0,-1,-2,\cdots;$

$$(4)\quad \frac{1}{2\pi i}\int\Gamma(\gamma+s)\Gamma(\gamma'-s)\Gamma(\alpha+s)\Gamma(\alpha'-s)$$
$$\times F(\alpha+s,\beta;\alpha+\gamma';x)F(\alpha'-s,\beta',\alpha'+\gamma;x)\,ds$$
$$=\Gamma(\gamma+\gamma')\Gamma(\alpha+\alpha')\Gamma(\alpha+\gamma')\Gamma(\alpha'+\gamma)\{\Gamma(\alpha+\alpha'+\gamma+\gamma')\}^{-1}$$
$$\times F(\alpha+\alpha',\beta+\beta';\alpha+\alpha'+\gamma+\gamma';x)\,,$$

Received March 22, 1957, and in revised form July 19, 1957.

where $\alpha+\alpha'+\gamma+\gamma'\neq 0, -1, -2, \cdots$; and

(5) $$\frac{1}{2\pi i}\int \frac{\Gamma(\beta+s)\Gamma(\beta'-s)}{\Gamma(\gamma+s)\Gamma(\gamma'-s)} F(\alpha+s, \beta+s ; \gamma+s ; x)F(\alpha'-s, \beta'-s ; \gamma'-s ; x)\, ds$$
$$=\frac{\Gamma(\gamma+\gamma'-\beta-\beta'-1)\Gamma(\beta+\beta')}{\Gamma(\gamma-\beta)\Gamma(\gamma'-\beta')\Gamma(\gamma+\gamma'-1)} F(\beta+\beta', \alpha+\alpha'-1 ; \gamma+\gamma'-1 ; x)\,,$$

where $\gamma-\beta\neq 0, -1, -2, \cdots$; $\gamma'-\beta'\neq 0, -1, -2, \cdots$ and $\gamma+\gamma'-1\neq 0, -1, -2, \cdots$.

All we need is the following two formulae [E. C. Titchmarch, Fourier integrals, p. 194]:

(6) $$\frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty} \Gamma(\alpha+s)\Gamma(\beta+s)\Gamma(\gamma-s)\Gamma(\delta-s)\, ds$$
$$=\Gamma(\alpha+\gamma)\Gamma(\alpha+\delta)\Gamma(\beta+\gamma)\Gamma(\beta+\delta)\{\Gamma(\alpha+\beta+\gamma+\delta)\}^{-1}$$

where $(-\alpha<k, -\beta<k, \gamma>k, \delta>k)$; and

(7) $$\frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty} \frac{\Gamma(\alpha-s)\Gamma(\gamma-s)}{\Gamma(\beta+s)\Gamma(\delta-s)}\, ds = \frac{\Gamma(\alpha+\gamma)\Gamma(\beta+\delta-\alpha-\gamma-1)}{\Gamma(\beta-\alpha)\Gamma(\delta-\gamma)\Gamma(\beta+\delta-1)}\,,$$

where $(-\alpha<k, -\beta<k, \gamma>k, \delta>k)$.

It may be noted that the restrictions, needed for (6) and (7), on the parameters in formulae (1)—(5), can be removed later on by the theory of analytical continuation. The proofs and two other formulae will be given in §2., while some confluent forms of addition theorems will be deduced, as a limiting case, in §3.

2. **Proof.** On expanding each hypergeometric function on the left hand side of (1) and changing the order of integration and summation it becomes

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{(\beta ; m)(\beta' ; n)}{m!\, n!} x^{m+n} \cdot \frac{1}{2\pi i}\int \frac{\Gamma(\alpha+m+s)\Gamma(\alpha'+n-s)}{\Gamma(\gamma+m+s)\Gamma(\gamma'+n-s)}\, ds\,.$$

From (7), it follows that the last integral is equal to

$$\frac{\Gamma(\alpha+\alpha'+m+n)\Gamma(\gamma+\gamma'-\alpha-\alpha'-1)}{\Gamma(\gamma-\alpha)\Gamma(\gamma'-\alpha')\Gamma(\gamma+\gamma'+m+n-1)}\,.$$

Thus the left hand side of (1) becomes

$$\frac{\Gamma(\alpha+\alpha')\Gamma(\gamma+\gamma'-\alpha-\alpha'-1)}{\Gamma(\gamma'-\alpha')\Gamma(\gamma-\alpha)\Gamma(\gamma+\gamma'-1)}$$
$$\times\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} \frac{(\beta ; m)(\beta' ; n)(\alpha+\alpha' ; m+n)}{m!\, n!(\gamma+\gamma'-1 ; m+n)} x^{m+n}$$

$$=\frac{\Gamma(\alpha+\alpha')\Gamma(\gamma+\gamma'-\alpha-\alpha'-1)}{\Gamma(\gamma'-\alpha')\Gamma(\gamma-\alpha)\Gamma(\gamma+\gamma'-1)}$$
$$\times\sum_{p=0}^{\infty}\frac{(\alpha+\alpha'\,;p)(\beta\,;p)}{p!(\gamma+\gamma'-1\,;p)}x^pF(\beta',\ -p\,;1-\beta-p\,;1)\ ,$$

and from this formula (1) follows by applying Gauss's theorem. The proof of (2) is the same as the proof of (1).

To prove (3), expand each hypergeometric function on the left hand side of (3) and change the order of integration and summation ; then it becomes

$$\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{x^{m+n}}{m!\ n!\left(\frac{1}{2}\alpha+\frac{1}{2}\beta+\frac{1}{2}\,;m\right)\left(\frac{1}{2}\alpha'+\frac{1}{2}\beta'+\frac{1}{2}\,;n\right)}$$
$$\times\frac{1}{2\pi i}\int\Gamma(\alpha+m+s)\Gamma(\alpha'+n+s)\Gamma(\beta+m-s)\Gamma(\beta'+n-s)\,ds\ .$$

From (6), it follows that the last integral is equal to

$$\frac{\Gamma(\alpha+\beta+2m)\Gamma(\alpha+\beta'+m+n)\Gamma(\alpha'+\beta+m+n)\Gamma(\alpha'+\beta'+2n)}{\Gamma(\alpha+\alpha'+\beta+\beta'+2m+2n)}\ .$$

Thus the left hand side of (3) becomes

$$(2\sqrt{\pi})^{-1}\Gamma\left(\frac{1}{2}\alpha+\frac{1}{2}\beta+\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\alpha'+\frac{1}{2}\beta'+\frac{1}{2}\right)$$
$$\times\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\frac{\Gamma\left(\frac{1}{2}\alpha+\frac{1}{2}\beta+m\right)\Gamma\left(\frac{1}{2}\alpha'+\frac{1}{2}\beta'+n\right)}{m!\ n!\ \Gamma\left(\frac{1}{2}\alpha+\frac{1}{2}\alpha'+\frac{1}{2}\beta+\frac{1}{2}\beta'+\frac{1}{2}+m+n\right)}$$
$$\times\frac{\Gamma(\alpha+\beta'+m+n)\Gamma(\alpha'+\beta+m+n)}{\Gamma\left(\frac{1}{2}\alpha+\frac{1}{2}\alpha'+\frac{1}{2}\beta+\frac{1}{2}\beta'+m+n\right)}x^{m+n}$$
$$=\Gamma(\alpha+\beta')\Gamma(\alpha'+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha'+\beta')\{\Gamma(\alpha+\alpha'+\beta+\beta')\}^{-1}$$
$$\times\sum_{p=0}^{\infty}\frac{(\alpha+\beta'\,;p)(\beta+\alpha'\,;p)\left(\frac{1}{2}\alpha'+\frac{1}{2}\beta'\,;p\right)}{p!\left(\frac{1}{2}\alpha+\frac{1}{2}\alpha'+\frac{1}{2}\beta+\frac{1}{2}\beta'\,;p\right)\left(\frac{1}{2}\alpha+\frac{1}{2}\alpha'+\frac{1}{2}\beta+\frac{1}{2}\beta'+\frac{1}{2}\,;p\right)}x^p$$
$$\times F\left(\frac{1}{2}\alpha+\frac{1}{2}\beta,\ -p\,;1-\frac{1}{2}\alpha'-\frac{1}{2}\beta'-p\,;1\right)\,;$$

and from this, formula (3) follows by applying Gauss's theorem. From the proof of (3), the following formula can be deduced :

$$(8)\quad \frac{1}{2\pi i}\int \Gamma(\alpha+s)\Gamma(\beta-s)\Gamma(\alpha'+s)\Gamma(\beta'-s)$$
$$\times F\left(\alpha+s,\ \beta-s\ ;\ \frac{1}{2}\alpha+\frac{1}{2}\beta\ ;\ x\right)F\left(\alpha'+s,\ \beta'-s\ ;\ \frac{1}{2}\alpha'+\frac{1}{2}\beta'\ ;\ x\right)ds$$
$$=\Gamma(\alpha+\beta)\Gamma(\alpha'+\beta')\Gamma(\alpha+\beta')\Gamma(\alpha'+\beta)\{\Gamma(\alpha+\alpha'+\beta+\beta')\}^{-1}$$
$$\times {}_3F_2\left[\begin{matrix}\alpha+\beta',\ \alpha'+\beta,\ \frac{1}{2}(\alpha+\alpha'+\beta+\beta'+2)\ ;\ x\\ \frac{1}{2}(\alpha+\alpha'+\beta+\beta'),\ \frac{1}{2}(\alpha+\alpha'+\beta+\beta'+1)\end{matrix}\right],$$

where $\alpha+\alpha'+\beta+\beta'\neq 0,\ -1,\ -2,\ \cdots$.

The proof of (4) is the same as the proof of (3), while formula (5) can be deduced by substituting for each hypergeometric function on the left hand side an integral of Barnes's type and changing the order of integration.

Finally, I may mention the following formula which involves a generalized hypergeometric function,

$$(9)\quad \frac{1}{2\pi i}\int \Gamma(\alpha+s)\Gamma(\beta+s)\Gamma(\alpha'-s)\Gamma(\beta'-s)$$
$$\times F(\alpha+s,\ \beta+s\ ;\ \gamma\ ;\ x)F(\alpha'-s,\ \beta'-s\ ;\ \gamma'\ ;\ x)\,ds$$
$$=\Gamma(\alpha+\alpha')\Gamma(\alpha+\beta')\Gamma(\beta+\alpha')\Gamma(\beta+\beta')\{\Gamma(\alpha+\alpha'+\beta+\beta')\}^{-1}$$
$$\times {}_6F_5\left[\begin{matrix}\alpha+\alpha',\ \alpha+\beta',\ \beta+\alpha',\ \beta+\beta',\ \frac{1}{2}(\gamma+\gamma'-1),\ \frac{1}{2}(\gamma+\gamma')\ ;\ x\\ \gamma,\ \gamma',\ \gamma+\gamma'-1,\ \frac{1}{2}(\alpha+\alpha'+\beta+\beta'),\ \frac{1}{2}(\alpha+\alpha'+\beta+\beta'+1)\end{matrix}\right],$$

where $\gamma+\gamma'-1\neq 0,\ -1,\ -2,\ \cdots,\ \alpha+\alpha'+\beta+\beta'\neq 0,\ -1,\ -2,\ \cdots$ and either γ or γ' is not zero or a negative integer.

3. Confluent forms of addition theorems. In (1) take $\beta'=\beta$, write x/β for x and let $\beta\to\infty$ to get

$$(10)\quad \frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty}\frac{\Gamma(\alpha+s)\Gamma(\alpha'-s)}{\Gamma(\gamma+s)\Gamma(\gamma'-s)}F(\alpha+s\ ;\ \gamma+s\ ;\ x)F(\alpha'-s\ ;\ \gamma'-s\ ;\ x)\,ds$$
$$=\frac{\Gamma(\alpha+\alpha')\Gamma(\gamma+\gamma'-\alpha-\alpha'-1)}{\Gamma(\gamma+\gamma'-1)\Gamma(\gamma-\alpha)\Gamma(\gamma'-\alpha')}F(\alpha+\alpha'\ ;\ \gamma+\gamma'-1\ ;\ 2x),$$

where $\mathfrak{R}(k+\alpha)>0$, $\mathfrak{R}(k+\gamma)>0$, $\mathfrak{R}(\alpha'-k)>0$, $\mathfrak{R}(\gamma'-k)>0$, $\gamma+\gamma'-1\neq 0,\ -1,\ -2,\ \cdots$; $\gamma-\alpha\neq 0,\ -1,\ -2,\ \cdots$ and $\gamma'-\alpha'\neq 0,\ -1,\ -2,\ \cdots$.

In (2), take $\gamma'=\gamma$, write x/γ for x and let $\gamma\to\infty$, to get

$$(11)\quad \frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty}\frac{\Gamma(\alpha+s)\Gamma(\alpha'-s)}{\Gamma(\beta+s)\Gamma(\beta'-s)}F(\beta-\alpha\ ;\ \beta+s\ ;\ x)F(\beta'-\alpha'\ ;\ \beta'-s\ ;\ x)\,ds$$

$$= \frac{\Gamma(\alpha+\alpha')\Gamma(\beta+\beta'-\alpha-\alpha'-1)}{\Gamma(\beta-\alpha)\Gamma(\beta'-\alpha')\Gamma(\beta+\beta'-1)} F(\beta+\beta'-\alpha-\alpha'-1\,;\beta+\beta'-1\,;2x)\,,$$

where $\Re(k+\alpha)>0$, $\Re(k+\beta)>0$, $\Re(\alpha'-k)>0$, $\Re(\beta'-k)>0$, $\beta-\alpha\neq 0, -1, -2, \cdots$; $\beta'-\alpha'\neq 0, -1, -2, \cdots$ and $\beta+\beta-1\neq 0, -1, -2, \cdots$.

Finally in (4), take $\beta'=\beta$; write x/β for x and let $\beta\to\infty$, to get

$$\begin{aligned} (12)\qquad & \frac{1}{2\pi i}\int_{k-i\infty}^{k+i\infty}\Gamma(\gamma+s)\Gamma(\gamma'-s)\Gamma(\alpha+s)\Gamma(\alpha'-s) \\ & \qquad \times F(\alpha+s\,;\alpha+\gamma'\,;x)F(\alpha'-s\,;\alpha'+\gamma\,;x)\,ds \\ & \quad =\Gamma(\gamma+\gamma')\Gamma(\alpha+\alpha')\Gamma(\alpha+\gamma')\Gamma(\alpha'+\gamma)\{\Gamma(\alpha+\alpha'+\gamma+\gamma')\}^{-1} \\ & \qquad \times F(\alpha+\alpha'\,;\alpha+\alpha'+\gamma+\gamma'\,;2x)\,, \end{aligned}$$

where $\Re(k+\gamma)>0$, $\Re(\alpha+k)>0$, $\Re(\gamma'-k)>0$, $\Re(\alpha'-k)>0$, and $\alpha+\alpha'+\gamma+\gamma'\neq 0, -1, -2, \cdots$.

INSTITUTE FOR ADVANCED STUDY, PRINCETON, N. J., U.S.A. AND
FACULTY OF SCIENCE, EIN SHAMS UNIVERSITY ABBASSIA CAIRO EGYPT

PRINCIPAL SOLUTIONS OF NON-OSCILLATORY SELF-ADJOINT LINEAR DIFFERENTIAL SYSTEMS

William T. Reid

1. Introduction. In their study of real quadratic functionals

$$\int_a^b [r(x)y'^2+2q(x)yy'+p(x)y^2]dx$$

admitting a singularity at the end-point $x=a$ Morse and Leighton [**11**] showed that if $x=a$ is not its own first conjugate point then the corresponding Euler differential equation

$$(r(x)y'+q(x)y)'-(q(x)y'+p(x)y)=0, \quad a<x\leqq b , \tag{1.1}$$

possesses a non-trivial solution $u(x)$ such that $u(x)/y(x)\to 0$ as $x\to a^+$ for each solution $y(x)$ of (1.1) that is independent of $u(x)$. Such a solution $u(x)$ was termed a focal solution belonging to $x=a$ by Morse and Leighton [**11**], but in a subsequent continuation of the study by Leighton [**8**] the terminology was changed to principal solution.

If $f(t)$ is a real-valued continuous function on $t_0\leqq t<\infty$ and

$$x''+f(t)x=0 , \qquad t_0\leqq t<\infty , \tag{1.2}$$

is non-oscillatory, Hartman and Wintner [**4**] have termed a non-trivial solution $x(t)$ a principal solution if

$$\int_{t_0}^\infty |x(t)|^{-2}dt=\infty , \tag{1.3}$$

for t_0 greater than the largest zero of $x(t)$, and proved that a non-oscillatory equation (1.2) has a principal solution that is unique to an arbitrary non-zero constant factor; moreover, if $x(t)\not\equiv 0$ is a solution of (1.2) which is not principal then every solution $y(t)$ of (1.2) is of the form $y(t)=Cx(t)+o(|x(t)|)$ as $t\to\infty$, where the constant C is or is not zero according as $y(t)$ is or is not principal. In view of this latter result, for a non-oscillatory equation (1.2) a solution $x(t)$ is principal in the sense of Hartman and Wintner if and only if it is principal in the sense of Morse and Leighton.

Recently Hartman [**5**] has considered a self-adjoint vector differential equation

---

Received August 12, 1957. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 18(603)-86. Reproduction in whole or in part is permitted for any purpose of the United States Government.

$$(1.4) \qquad (R(t)x')' + F(t)x = 0\ , \qquad 0 \leqq t < \infty\ ,$$

where $R(t)$, $F(t)$ are $n\times n$ matrices which are continuous and hermitian, while $R(t)$ is positive definite on the interval of consideration. An $n\times n$ matrix solution of the corresponding matrix differential equation

$$(1.4') \qquad (R(t)X')' + F(t)X = 0$$

is termed "prepared" by Hartman if $X^*(t)R(t)X'(t)$ is hermitian. Under the assumption that the class Γ of solutions $X = X(t)$ of (1.4′) which are prepared and non-singular on a corresponding interval $a_X < t < \infty$ is non-empty, Hartman showed that in Γ there exists a solution which is principal in the sense that the least proper value $\lambda_X(t)$ of the positive definite hermitian matrix

$$(1.5) \qquad \int_{t_0}^{t} (X^*X)^{-1} ds\ , \qquad (t_0 \textit{ sufficiently large}\ ;\ t > t_0),$$

satisfies $\lambda_X(t) \to \infty$ as $t \to \infty$, and this principal prepared solution is unique up to multiplication on the right by an arbitrary non-singular constant matrix, while there also exist in Γ solutions that are non-principal in the sense that the greatest proper value $\mu_X(t)$ of (1.5) remains finite as $t\to\infty$; moreover, if $Y(t)$ and $X(t)$ are matrices of Γ which are principal and non-principal, respectively, then $X^{-1}(t)Y(t) \to 0$ as $t\to\infty$.

Hartman's assumption that the above defined class Γ is non-empty is indeed an hypothesis of non-oscillation, since in view of the results of a recent paper of Reid [13] the class Γ is non-empty if and only if (1.4) is non-oscillatory for large t in the sense that there exists a t_0 such that if $x(t)$ is a solution of (1.4) satisfying $x(t_1) = 0 = x(t_2)$ with $t_0 < t_1 < t_2$ then $x(t) \equiv 0$.

It is to be noted that Hartman's definition of principal solution for an equation (1.4) which is non-oscillatory for large t has the undesirable feature of limiting the considered matrix solutions of (1.4′) to the class Γ; indeed, Hartman [5 ; §11] gives an example of a non-prepared solution $X(t)$ of (1.4′) that is non-singular for large t, and such that the least proper value $\lambda_X(t)$ of the corresponding hermitian matrix (1.5) satisfies $\lambda_X(t) \to \infty$ as $t\to\infty$. Moreover, as Hartman points out, his classification of principal and non-principal solutions does not present a disjunctive alternative in the class Γ.

For a self-adjoint vector differential equation of somewhat more general type than that considered by Hartman, and which is non-oscillatory for large values of the independent variable, the present paper presents a generalized definition of principal solution that distinguishes such solutions in the class Γ_0 of all matrix solutions which are non-singular for large values of the independent variable. In addition, it is shown that principal solutions possess on Γ_0 certain properties that are extensions of properties established by Hartman for the class Γ. It is to be commented also that the presented determination of a principal solution is by variational methods and is direct in nature, in contrast to the indirect character of the proofs of the existence of a principal solution in the above-cited papers of Hartman, Hartman and Wintner, and Morse and Leighton ; in this connection it is to be remarked that although the existence of a principal solution for (1.1) was established indirectly by Morse and Leighton [**11**], the properties of principal solutions derived in their Theorem 2.2 permit a ready direct determination of such a solution.

Sections 2 - 8 of the present paper deal with a self-adjoint n-dimensional vector equation with complex coefficients that is a direct generalization of the scalar equation (1.1) ; Section 9 is devoted to a more general differential system with complex coefficients that is of the general form of the accessory differential equations for a variational problem of Bolza type.

Matrix notation is used throughout ; in particular, matrices of one column are termed vectors, and for a vector $y=(y_\alpha)$, $(\alpha=1, \cdots, n)$, the norm $|y|$ is given by $(|y_1|^2+\cdots+|y_n|^2)^{1/2}$. The symbol E is used for the $n\times n$ identity matrix, while 0 is used indiscriminately for the zero matrix of any dimensions ; the conjugate transpose of a matrix M is denoted by M^*. Moreover, the notation $M\geqq N$, $(M>N)$, is used to signify that M and N are hermitian matrices of the same dimensions and $M-N$ is a nonnegative (positive) hermitian matrix.

2. Formulation of the problem.

For x on a given interval X: $a<x<\infty$ let $\omega(x, y, \pi)$ denote the hermitian form

$$\omega(x, y, \pi)=\pi^*R(x)\pi+\pi^*Q(x)y+y^*Q^*(x)\pi+y^*P(x)y , \tag{2.1}$$

in the $2n$ variables $y, \pi=(y_1, \cdots, y_n, \pi_1, \cdots, \pi_n)$. *It will be assumed throughout Sections 2 - 8 that $R(x)$, $Q(x)$, $P(x)$ are $n\times n$ matrices having complex-valued continuous elements on X, with $R(x)$, $P(x)$ hermitian, and $R(x)$ non-singular on this interval.*

If c, d are points of X the symbol $I[y\,;\,c, d]$ will denote the hermitian functional

$$I[y\,;\,c,\ d]=\int_c^d \omega(x,\ y,\ y')\,dx . \tag{2.2}$$

For the functional (2.2) the vector Euler equation is

$$L[u]\equiv(R(x)u'+Q(x)u)'-(Q^*(x)u'+P(x)u)=0 , \tag{2.3}$$

which may be written in terms of the canonical variables

$$u(x),\ v(x)=R(x)u'(x)+Q(x)u(x)$$

as the first order system

$$(2.4) \qquad u' = A(x)u + B(x)v \ , \quad v' = C(x)u - A^*(x)v \ ,$$

where the $n \times n$ coefficient matrices of (2.4) are continuous on X and given by $A = -R^{-1}Q$, $B = R^{-1}$, $C = P - Q^*R^{-1}Q$; in particular, the matrices $B(x)$, $C(x)$ are hermitian on X and $B(x)$ is non-singular on this interval.

Corresponding to (2.3) and (2.4) are the respective matrix equations

$$(2.3') \qquad L[U] \equiv (R(x)U' + Q(x)U)' - (Q^*(x)U' + P(x)U) = 0 \ ,$$

$$(2.4') \qquad U' = A(x)U + B(x)V, \qquad V' = C(x)U - A^*(x)V \ .$$

In [13] the author has discussed various criteria of oscillation and non-oscillation for an equation (2.3) in which the coefficient matrices satisfy weaker conditions than those imposed above; although the results of the present paper hold for equations of the generality discussed in [13], for simplicity specific attention is restricted to the case described above.

Throughout the subsequent discussion of Sections 2 - 8 we shall deal consistently with the cononical system (2.4) and associated matrix system (2.4′), instead of the equivalent respective equations (2.3) and (2.3′), since in Section 9 there is considered a vector differential system more general than (2.3), but with associated canonical system still of the form (2.4).

If $U(x) \equiv \| U_{\alpha j}(x) \|$, $V(x) = \| V_{\alpha j}(x) \|$, $(\alpha = 1, \cdots, n\ ;\ j = 1, \cdots, r)$ are $n \times r$ matrices, for typographical simplicity the symbol $(U(x)\ ;\ V(x))$ will be used to denote the $2n \times r$ matrix whose j-th column has elements $U_{1j}(x), \cdots, U_{nj}(x), V_{1j}(x), \cdots, V_{nj}(x)$. In the major portion of the following discussion we shall be concerned with matrices $(U(x)\ ;\ V(x))$ which are solutions of the matrix differential system (2.4′).

If $(U_1(x)\ ;\ V_1(x))$ and $(U_2(x)\ ;\ V_2(x))$ are individually solutions of (2.4′) then, as noted in Lemma 2.1 of [13], the matrix $U_1^*(x)V_2(x) - V_1^*(x)U_2(x)$ is a constant. This matrix will be denoted by $\{U_1, U_2\}$; it is to be remarked that for the problem formulated above there is no ambiguity in this notation, since the $V(x)$ belonging to a solution $(U(x)\ ;\ V(x))$ of (2.4′) is uniquely determined as $V(x) = R(x)U'(x) + Q(x)U(x)$. As in [13], two solutions $(u_1(x)\ ;\ v_1(x))$ and $(u_2(x)\ ;\ v_2(x))$ of (2.4) are said to be (mutually) *conjoined* if $\{u_1, u_2\} = 0$. If $(U(x)\ ;\ V(x))$ is a solution of (2.4′) whose column vectors are conjoined solutions of (2.4), then $(U(x);\ V(x))$ will be termed a matrix of conjoined solutions. In particular, if $U(x)$, $V(x)$ are $n \times n$ matrices such that $(U(x)\ ;\ V(x))$ is a matrix of conjoined solutions of (2.4), then $U(x)$ is a prepared solution of (2.3′) in the sense of Hartman [5]. If the coefficients of (2.1) are real-valued, then two real-valued solutions of (2.4) are conjoined if and only if they are conjugate in the sense introduced originally by von Escherich. The reader is referred to [13; pp. 737, 743] for comments on the use of the synonym "conjoined" for the case of (2.1) with complex-valued coefficients.

Two points s, t of X are said to be (mutually) *conjugate*, (with respect to (2.3) or (2.4)), if there exists a solution $(u(x); v(x))$ with $u(x)\not\equiv 0$ on $[s, t]$ and satisfying $u(s)=0=u(t)$. The system (2.4) will be termed *non-oscillatory on a given interval* provided no two distinct points of this interval are conjugate; moreover, (2.4) will be called *non-oscillatory for large* x if there exists a subinterval $a_0 < x < \infty$ of X on which this system is non-oscillatory.

3. **Related matrix solutions of** (2.4′). Suppose that $(U(x); V(x))$ is a solution of (2.4′) with $U(x)$ non-singular on a given subinterval X_0 of X, and denote by K the $n\times n$ constant matrix such that $\{U, U\}\equiv K$. If $(U_0(x); V_0(x))$ is a $2n\times r$ matrix solution of (2.4′) on X_0, and K_0 is the $n\times r$ constant matrix such that $\{U, U_0\}\equiv K_0$, then from this latter relation it follows that the $n\times r$ matrix $H(x)=U^{-1}(x)U_0(x)$ is such that

$$(3.1)\qquad U_0(x)=U(x)H(x),\ \ V_0(x)=V(x)H(x)+U^{*-1}(x)[K_0-KH(x)]\ ,$$

and in view of the relation $K=-K^*$ it may be verified readily that

$$(3.2)\qquad \{U_0,\ U_0\}\equiv -H^*(x)KH(x)+H^*(x)K_0-K_0{}^*H(x)\equiv K_1\ ,$$

where K_1 is a constant $r\times r$ matrix. Moreover, from the differential equations $U_0'=AU_0+BV_0$, $U'=AU+BV$ it follows that

$$(3.3)\qquad H'(x)=U^{-1}(x)B(x)U^{*-1}(x)[K_0-KH(x)],\ \ x\in X_0\ .$$

Conversely, if K_0 is an arbitrary $n\times r$ constant matrix, and $H(x)$ is an $n\times r$ matrix satisfying the corresponding matrix differential equation (3.3), then it follows readily that the $2n\times r$ matrix $(U_0(x); V_0(x))$ defined by (3.1) is a solution of (2.4′) with $\{U, U_0\}\equiv K_0$, and $\{U_0, U_0\}$ given by (3.2).

Now if $x=s$ is a point of X and $T(x)=T(x, s; U)$ is the solution of the matrix differential system

$$(3.4)\qquad T'=-U^{-1}(x)B(x)U^{*-1}(x)KT,\ \ T(s)=E\ ,$$

then by the method of variation of parameters it follows immediately that $H(x)$ is a solution of (3.3) for a given $n\times r$ matrix K_0 if and only if there is an $n\times r$ constant matrix $H_0=H(s)$ such that

$$(3.5)\qquad H(x)=T(x,\ s\ ;\ U)[H_0+S(x,\ s\ ;\ U)K_0]\ ,$$

where

$$(3.6) \qquad S(x, s; U) = \int_s^x T^{-1}(t, s; U) U^{-1}(t) B(t) U^{*-1}(t)\, dt, \quad x, s \in X_0 .$$

The corresponding solution $(U_0(x); V_0(x))$ of (2.4′) determined by (3.1) is such that

$$(3.7) \qquad U_0(x) = U(x) T(x, s; U)[U^{-1}(s) U_0(s) + S(x, s; U)\{U, U_0\}] .$$

In general, if $F(x)$ is a continuous $n \times n$ matrix and $Y(x)$ is the fundamental matrix of $Y' = F(x)Y$ satisfying $Y(s) = E$, then $Z = Y^{*-1}(x)$ is the fundamental matrix solution of $Z' = -F^*(x)Z$ satisfying $Z(s) = E$. As $K = \{U, U\}$ satisfies $K = -K^*$ it follows that $T^{*-1}(x) = T^{*-1}(x, s; U)$ is the solution of $(T^{*-1})' = -KU^{-1}(x)B(x)U^{*-1}(x)T^{*-1}$ satisfying $T^{*-1}(s) = E$. Now if $H(x)$ is a solution of (3.3) then

$$[K_0 - KH(x)]' = -KU^{-1}(x)B(x)U^{*-1}(x)[K_0 - KH(x)] ,$$

and hence $K_0 - KH(x) = T^{*-1}(x, s; U)[K_0 - KH_0]$. Since $K = \{U, U\}$ and $K_0 = \{U, U_0\}$, this latter relation may be written as the following identity for solutions $(U_0(x); V_0(x))$ and $(U(x); V(x))$ of (2.4′), with $U(x)$ nonsingular on the interval of consideration X_0 and x, s arbitrary values on this interval,

$$(3.8) \qquad \begin{aligned} \{U, U_0\} - \{U, U\} U^{-1}(x) U_0(x) \equiv T^{*-1}(x, s; U)[\{U, U_0\} \\ - \{U, U\} U^{-1}(s) U_0(s)] . \end{aligned}$$

In particular, if $\{U, U\} = 0$ then

$$(3.9) \qquad K = 0, \ T(x, s; U) \equiv E, \ H(x) = H_0 + \int_s^x U^{-1}(t) B(t) U^{*-1}(t)\, dt ,$$

and the $U_0(x)$, $V_0(x)$ given by (3.1) satisfy $\{U_0, U_0\} = 0$ if and only if the $r \times r$ constant matrix $H_0^* K_0$ is hermitian. In case $\{U, U\} = 0$ the formula (3.7) reduces to a relation that may be found in various recent papers, (see Sternberg and Kaufman [14]; Barrett [1 and 2]; Hartman [5]). For future reference the above results are collected in the following theorem.

THEOREM 3.1. *If* $(U(x); V(x))$ *is a solution of* (2.4′) *with* $U(x)$ *nonsingular on a subinterval* X_0 *of* X, *and* K *is the constant* $n \times n$ *matrix such that* $\{U, U\} \equiv K$, *then an* $n \times r$ *matrix* $U_0(x)$ *belongs to a solution* $(U_0(x); V_0(x))$ *of* (2.4′) *on* X_0 *if and only if* $U_0(x) = U(x)H(x)$, *where* $H(x)$ *is of the form* (3.5) *with* $T(x, s; U)$ *and* $S(x, s; U)$ *determined by* (3.4) *and* (3.6), *respectively, and* H_0, K_0 *are* $n \times r$ *constant matrices. Moreover, for such a* $U_0(x)$ *the corresponding* $V_0(x)$ *is given by* (3.1), $\{U, U_0\} = K_0$, $\{U_0, U_0\}$ *has the value* (3.2), *and the identities* (3.7), (3.8) *hold for* $x, s \in X_0$; *in particular, if* $K = 0$ *then* $T(x, s; U) \equiv E$ *and* $\{U_0, U_0\} \equiv 0$ *if and only if the constant* $r \times r$ *matrix* $H_0^* K_0$ *is hermitian.*

It is to be emphasized that the above theorem is quite independent of any non-oscillatory character of (2.4). For example, the scalar equation $u''+u=0$ has solution $u(x)=\exp(ix)$ which satisfies $u(x)\neq 0$ on $(-\infty, \infty)$, and with $\{u, u\}\equiv 2i$, $T(x, s; u)=\exp(-2i(x-s))$, $S(x, s; u)=\sin(x-s)\exp(i(x-s))$; moreover, $u_0(x)=\sin x$ is a second solution of this equation for which $\{u, u_0\}\equiv 1$, and one may verify readily the identities (3.7) and (3.8).

THEOREM 3.2. *Suppose that* $(U(x); V(x))$ *is a solution of* (2.4′) *with* $U(x)$ *non-singular on a subinterval* X_0 *of* X. *If* $s\in X_0$ *then for* $t\in X_0$, $t\neq s$, *the matrix* $S(t, s; U)$ *is singular if and only if* t *is conjugate to* s. *In particular, if* (2.4) *is non-oscillatory on a subinterval* $X_0: a_0<x<\infty$, *and* $(U(x); V(x))$ *is a solution of* (2.4′) *with* $U(x)$ *non-singular on* X_0, *then for* $s\in X_0$ *the matrix* $S(t, s; U)$ *is non-singular for* $t\in X_0$, $t\neq s$; *moreover, if there exists an* $s\in X_0$ *such that* $S^{-1}(x, s; U)\to 0$ *as* $x\to\infty$ *then* $S^{-1}(x, r; U)\to 0$ *as* $x\to\infty$ *for arbitrary* $r\in X_0$.

As $B(x)$ is non-singular, if $u(x)\equiv 0$, $v(x)$ is a solution of (2.4) on a given subinterval of X then $v(x)\equiv 0$ on this subinterval. In view of this condition, which is a property of "normality" of (2.4), it follows that if $(U_0(x); V_0(x))$ is a solution of (2.4′) with $U_0(s)=0$ and $V_0(s)$ non-singular then t is conjugate to s if and only if $U_0(t)$ is singular. Now if $(U(x); V(x))$ is a solution of (2.4′) with $U(x)$ non-singular on X_0, then for $s\in X_0$ the above-defined $(U_0(x); V_0(x))$ is such that $\{U, U_0\}$ is the non-singular matrix $U^*(s)V_0(s)$, and from (3.7) it follows that $U_0(x)=U(x)T(x, s; U)S(x, s; U)U^*(s)V_0(s)$ for $x\in X_0$, and thus $S(t, s; U)$ is singular for a value $t\in X_0$, $t\neq s$, if and only if t is conjugate to s. Consequently, if (2.4) is non-oscillatory on a subinterval X_0, and $(U(x); V(x))$ is a solution of (2.4′) with $U(x)$ non-singular on X_0, then $S(t, s; U)$ is non-singular for $t\in X_0$, $t\neq s$. Now the fundamental matrix $T(x, s; U)$ of (3.4) satisfies the well-known relation $T(x, s; U)=T(x, r; U)T(r, s; U)$ for $r, s\in X_0$, and by direct computation it follows that

$$S(x, s; U)=T(s, r; U)[S(x, r; U)-S(s, r; U)] \tag{3.10}$$

for $r, s, x\in X_0$. If for a general non-singular matrix M the supremum and infimum of $|My|$ on the sphere $|y|=1$ are denoted by $\mu(M)$ and $\lambda(M)$, respectively, then the relation

$$\mu(M^{-1})|My|\geqq|M^{-1}(My)|=|y|=|M(M^{-1}y)|\geqq\lambda(M)|M^{-1}y|\ ,$$

implies that $1=\lambda(M)\mu(M^{-1})$. As the condition that $S^{-1}(x, s; U)\to 0$ as $x\to\infty$ is equivalent to $\mu(S^{-1}(x, s; U))\to 0$ as $x\to\infty$, this condition holds if and only if $\lambda(S(x, s; U))\to\infty$ as $x\to\infty$. Now in view of the non-singularity of $T(s, r; U)$ it follows from (3.10) that for $r, s\in X_0$ we have $\lambda(S(x, s; U))\to\infty$ as $x\to\infty$ if and only if $\lambda(S(x, r; U))\to\infty$ as $x\to\infty$.

In view of the result of Theorem 3.2, for an equation (2.4) that is non-oscillatory for large x a solution $(U(x)\,;\ V(x))$ of (2.4′) will be termed a *principal solution* if $U(x)$ is non-singular for x on some interval X_U: $a_U < x < \infty$ and $S^{-1}(x,\ s\ ;\ U) \to 0$ as $x \to \infty$ for at least one (and consequently all) $s \in X_U$. If $(U(x)\,;\ V(x))$ is a matrix of conjoined solutions of (2.4) with $U(x)$ non-singular for large x this definiton clearly reduces to that of Hartman [5]. In the following sections it will be shown that if $R(x)$ is positive definite on X, and (2.4) is non-oscillatory for large x, then there does exist a principal solution of (2.4′), and this principal solution is unique up to multiplication on the right by a non-singular constant matrix. In general, however, one has the following theorem, which shows that if (2.4) is non-oscillatory for large x then a solution of (2.4′) which is principal in the sense defined above possesses a property corresponding to that used as a definitive property by Morse and Leighton [11] for the scalar eqution (1.1).

THEOREM 3.3. *If* (2.4) *is non-oscillatory for large x, then a solution* $(U(x)\,;\ V(x))$ *of* (2.4′) *is a principal solution if $U(x)$ is non-singular for large x and there exists a solution* $(U_0(x)\,;\ V_0(x))$ *of* (2.4′) *with $U_0(x)$ non-singular for large x and such that for some value $s \in X$,*

$$U_0^{-1}(x)U(x)T(x,\ s\ ;\ U) \to 0 \ \ as\ \ x \to \infty\ ; \tag{3.11}$$

moreover, $\{U,\ U_0\}$ is non-singular for any such $(U_0(x)\,;\ V_0(x))$. *Conversely, if* (2.4) *is non-oscillatory for large x, and* $(U(x)\,;\ V(x))$ *is a principal solution of* (2.4′), *then any solution* $(U_0(x)\,;\ V_0(x))$ *of* (2.4′) *with* $\{U,\ U_0\}$ *non-singular is such that $U_0(x)$ is non-singular for large x and* (3.11) *holds for arbitrary $s \in X$.*

Suppose that (2.4) is non-oscillatory for large x, and that there is a solution $(U(x)\,;\ V(x))$ of (2.4′) with $U(x)$ non-singular on an interval X_0: $a_0 < x < \infty$. If $(U_0(x)\,;\ V_0(x))$ is also a solution of (2.4′) then by (3.7),

$$[U(x)T(x,\ s\ ;\ U)]^{-1}U_0(x) = U^{-1}(s)U_0(s) + S(x,\ s\ ;\ U)\{U,\ U_0\}\ ; \tag{3.12}$$

moreover, if $U_0(x)$ is non-singular and satisfies (3.11) for some $s \in X_0$, then $\lambda([U(x)T(x,\ s\ ;\ U)]^{-1}U_0(x)) \to \infty$ as $x \to \infty$ and from (3.12) it follows that $\{U,\ U_0\}$ is non-singular and $\lambda(S(x,\ s\ ;\ U)) \to \infty$ as $x \to \infty$, so that $(U(x)\,;\ V(x))$ is a principal solution of (2.4′).

On the other hand, if (2.4) is non-oscillatory for large x, and $(U(x);\ V(x))$ is a principal solution of (2.4′), then for s sufficiently large we have that $\lambda(S(x,\ s\ ;\ U)) \to \infty$ as $x \to \infty$. For such a value s, and $(U_0(s)\,;\ V_0(x))$ a solution of (2.4′) with $\{U,\ U_0\}$ non-singular, we have $\lambda(U^{-1}(s)U_0(s) + S(x,\ s\ ;\ U)\{U,\ U_0\}) \to \infty$ as $x \to \infty$, and hence from (3.12) it follows that $\lambda([U(x)T(x,\ s\ ;\ U)]^{-1}U_0(x)) \to \infty$ as $x \to \infty$, which is equivalent to the condition that $U_0(x)$ is non-singular for large x and satisfies (3.11). As

$T(x, s; U)=T(x, r; U)T(r, s; U)$, if (3.11) holds for one value s then this condition holds for arbitrary $s \in X$.

4. Certain basic results of the calculus of variations. For the functional (2.2) an n-dimensional vector function $y(x)$ will be termed *differentially admissible on a subinterval* of X if on this subinterval $y(x)$ is continuous and has piecewise continuous derivatives. For brevity, if $[c, d]$ is a compact subinterval of X the symbol $H_+[c, d]$ will signify the condition that $I[y; c, d]>0$ for arbitrary $y(x)$ differentially admissible on $[c, d]$, and such that $y(x)\not\equiv 0$ on $[c, d]$, $y(c)=0=y(d)$. We shall also denote by H_R the condition that $R(x)>0$ on X; in view of the basic assumption that $R(x)$ is non-singular on X the condition H_R holds whenever there is a single s of X such that $R(s)>0$.

For the subsequent discussion the following known variational results are basic.

THEOREM 4.1. *If $[c, d]$ is a compact subinterval of X then a necessary and sufficient condition for $H_+[c, d]$ is that H_R hold, together with one of the following conditions*:

(i) (2.4) *is non-oscillatory on* $[c, d]$;

(ii) *there exists a matrix* $(U(x); V(x))$ *of conjoined solutions of* (2.4) *with $U(x)$ non-singular on* $[c, d]$.

THEOREM 4.2. *If $[c, d]$ is a compact subinterval of X such that $H_+[c, d]$ holds, then for arbitrary vectors y_c, y_d there is a unique solution $(u(x); v(x))$ of* (2.4) *satisfying $u(c)=y_c$, $u(d)=y_d$, and $I[y; c, d]>I[u; c, d]$ for arbitrary differentially admissible $y(x)$ with $y \not\equiv u$ on $[c, d]$, $y(c)=u(c)$, $y(d)=u(d)$.*

THEOREM 4.3. *Suppose that $[c, d]$ is a compact subinterval of X such that $H_+[c, d]$ holds. If $(U_c(x); V_c(x))$, $[(U_d(x); V_d(x))]$, is the solution of* (2.4′) *determined by $U_c(c)=E$, $U_c(d)=0$, $[U_d(d)=E$, $U_d(c)=0]$, and $(U(x); V(x))$ is a solution of* (2.4′) *satisfying $U(c)=E$, $V(c)>V_c(c)$, $[U(d)=E$, $V(d)<V_d(d)]$, then $(U(x); V(x))$ is a matrix of conjoined solutions of* (2.4) *with $U(x)$ non-singular on $[c, d]$.*

For the case in which the coefficient matrices of (2.1) are real-valued the results of Theorems 4.1 and 4.2 are classical results in the calculus of variations, (see, for example, Morse [10; Chapter I], or Bliss [3; Chapter IV]; for the general case of complex coefficients these results are contained in Theorems 2.1 and 2.2 of Reid [13]. In connection with Theorem 4.2 it is to be commented that if

$$I[\eta, u; c, d]=\int_c^d [\eta^{*\prime}(Ru'+Qu)+\eta^*(Q^*u'+Pu)]dx$$

for differentially admissible $\eta(x)$, $u(x)$, then in case $(u(x)\,;\,v(x))$ is a solution of (2.3) on $[c, d]$ we have

$$I[\eta,\, u\,;\, c,\, d]=\eta^*(x)v(x)\Big|_c^d. \tag{4.1}$$

A ready consequence of (4.1) is that if $(u(x)\,;\,v(x))$ and $y(x)$ satisfy the conditions of Theorem 4.2 then

$$I[y\,;\, c,\, d]=I[u\,;\, c,\, d]+I[y-u\,;\, c,\, d]\,, \tag{4.2}$$

which is the well-known "integral formula of Weierstrass" for the functional (2.2).

Theorem 4.3 is a comparison theorem of Sturmian type that is a special case of results of Morse [9; §10, or 10; Chapter IV, §8] in case the coefficients of (2.1) are real-valued, and Morse's method may be extended readily to prove the stated result. The method introduced by Hestenes [6], (see also Bliss [3; §§86–87]), to establish the corresponding result for variational problems of Bolza type yields the following brief and elegant proof of the statement of the theorem involving $(U_c(x);\, V_c(x))$; the statement involving $(U_d(x)\,;\, V_d(x))$ follows by a similar argument. By Theorem 4.2 the condition $H_+[c, d]$ implies the existence of the solution $(U_c(x)\,;\, V_c(x))$ of (2.4′) satisfying $U_c(c)=E$, $U_c(d)=0$; the end condition $U_c(d)=0$ clearly implies that $(U_c(x)\,;\, V_c(x))$ is a matrix of conjoined solutions and consequently $V_c(c)=U_c^*(c)\, V_c(c)$ is hermitian. For $(U(x)\,;\, V(x))$ a solution of (2.4′) satisfying $U(c)=E$, $V(c)>V_c(c)$ the matrix $U(d)$ is non-singular, since if $U(d)\xi=0$ then $u(x)=(U(x)-U_c(x))\xi$, $v(x)=(V(x)-V_c(x))\xi$ is a solution of (2.4) satisfying $u(c)=0=u(d)$ so that $u(x)\equiv0$ by Theorem 4.1, and hence $(V(c)-V_c(c))\xi=0$ and $\xi=0$. Moreover, $U(x)$ is non-singular on $c<x<d$, since if $c<b<d$ and $U(b)\xi=0$ then $y(x)$ defined as $y(x)=(U(x)-U_c(x))\xi$, $c\leqq x\leqq b$, and $y(x)=-U_c(x)\xi$, $b\leqq x\leqq d$, satisfies $y(c)=0=y(d)$ and is differentially admissible on $[c, d]$, while in view of the hermitian character of $U_c^*(b)V_c(b)$ we have

$$\begin{aligned} I[y\,;\, c,\, d]&=\xi^*[U^*(b)-U_c^*(b)][V(b)-V_c(b)]\xi-\xi^*V_c^*(b)U_c(b)\xi\\ &=-\xi^*U_c^*(b)[V(b)-V_c(b)]\xi+\xi^*V_c^*(b)[U(b)-U_c(b)]\xi\\ &=-\xi^*\{U_c,\, U-U_c\}\xi\\ &=-\xi^*[V(c)-V_c(c)]\xi\,, \end{aligned}$$

and consequently $I[y\,;\, c,\, d]<0$ unless $\xi=0$, so that $\xi=0$ in view of $H_+[c, d]$.

5. Systems (2. 4) that are non-oscillatory for large x. For a system satisfying H_R and non-oscillatory for large x, the following theorem determines a particular matrix of conjoined solutions which subsequently will be shown to be a principal solution, as defined in Section 3.

THEOREM 5.1. *Suppose that* (2.4) *satisfies* H_R *and is non-oscillatory on a subinterval* X_0: $a_0<x<\infty$ *of* X. *If* $s\in X_0$ *and for* $t\in X_0$, $t\neq s$, *the matrix* $(U_{st}(x);\ V_{st}(x))$ *is the solution of* (2.4′) *determined by* $U_{st}(s)=E$, $U_{st}(t)=0$, *then* $U_{s,\infty}(x)=\lim_{t\to\infty} U_{st}(x)$, $V_{s,\infty}(x)=\lim_{t\to\infty} V_s(x)$ *exist and* $(U_{s,\infty}(x);\ V_{s,\infty}(x))$ *is a matrix of conjoined solutions of* (2.4) *with* $U_{s,\infty}(x)$ *non-singular on* X_0; *moreover,* $U_{r,\infty}(x)=U_{s,\infty}(x)U_{r,\infty}(s)$ *and* $V_{r,\infty}(x)=V_{s,\infty}(x)U_{r,\infty}(s)$ *for* $r,\ s,\ x\in X_0$.

As the initial condition $U_{st}(t)=0$ implies $\{U_{st},\ U_{st}\}=0$, it follows that if $s,\ t\in X_0$, $s\neq t$, then $(U_{st}(x);\ V_{st}(x))$ is a matrix of conjoined solutions, so that the matrix $U^*_{st}(x)V_{st}(x)$ is hermitian for $x\in X$; in particular, $V_{st}(s)$ is hermitian. For a given $s\in X_0$ let $r,\ t$ be points of X_0 satisfying $r<s<t$, and for an arbitrary non-zero constant vector ξ let $y(x)$ denote the vector function defined on $[r,\ t]$ as

$$y(x)=U_{sr}(x)\xi \text{ on } [r,\ s]\,;\ y(x)=U_{st}(x)\xi \text{ on } [s,\ t]\,. \tag{5.1}$$

Now this vector function $y(x)$ is differentially admissible and $y(r)=0=y(t)$, so that under the hypothesis that (2.4) satisfies H_R and is non-oscillatory on X_0 it follows from Theorem 4.1 that

$$0<I[y\,;\ r,\ t]=\xi^*U^*_{sr}(s)V_{sr}(s)\xi-\xi^*U^*_{st}(s)V_{st}(s)\xi=\xi^*[V_{sr}(s)-V_{st}(s)]\xi\,.$$

As this relation holds for arbitrary non-zero vectors ξ we have

$$V_{st}(s)<V_{sr}(s) \text{ for } r,\ s,\ t\in X_0,\ r<s<t\,. \tag{5.2}$$

For $s<t<d$, and ξ an arbitrary non-zero constant vector, let $u(x)=U_{sd}(x)\xi$, $v(x)=V_{sd}(x)\xi$ and $y(x)=U_{st}(x)\xi$ on $[s,\ t]$, $y(x)\equiv0$ on $[t,\ d]$. Then $(u(x);\ v(x))$ is a solution of (2.4), while $y(x)$ is differentially admissible and satisfies $y(s)=u(s)$, $y(d)=u(d)$, $y(x)\not\equiv u(x)$ on $[s,\ d]$, so that

$$-\xi^*V_{sd}(s)\xi=I[u\,;\ s,\ d]<I[y\,;\ s,\ d]=I[y\,;\ s,\ t]=-\xi^*V_{st}(s)\xi \tag{5.3}$$

in view of Theorem 4.2; that is,

$$V_{st}(s)<V_{sd}(s) \text{ for } s,\ t,\ d\in X_0,\ s<t<d\,. \tag{5.4}$$

By a similar argument it follows that

$$V_{sc}(s)<V_{sr}(s) \text{ for } c,\ r,\ s\in X_0,\ c<r<s\,. \tag{5.5}$$

From (5.2), (5.4) it follows that for fixed $s\in X_0$ the one-parameter family of hermitian matrices $V_{st}(s)$, $s<t<\infty$, is monotone increasing and bounded, so that there is an hermitian matrix $V_{s,\infty}$ such that $V_{sd}(s)\to V_{s,\infty}$ as $d\to\infty$. Moreover, in view of (5.2), (5.4), (5.5) it follows that

$$V_{st}(s)<V_{s,\infty}<V_{sr}(s) \text{ for } r,\ s,\ t\in X_0,\ r<s<t\,, \tag{5.6}$$

If $(U_{s,\infty}(x)\,;\ V_{s,\infty}(x))$ is the solution of (2.4′) determined by the initial values $U_{s,\infty}(s)=E$, $V_{s,\infty}(s)=V_{s,\infty}$ then clearly $(U_{st}(x)\,;\ V_{st}(x))\rightarrow(U_{s,\infty}(x)\,;\ V_{s,\infty}(x))$, while the hermitian character of $V_{s,\infty}=U^*_{s,\infty}(s)V_{s,\infty}(s)$ implies that $\{U_{s,\infty},\ U_{s,\infty}\}=0$, and $(U_{s,\infty}(x)\,;\ V_{s,\infty}(x))$ is a matrix of conjoined solutions. Moreover, in view of Theorem 4.3, inequality (5.6) implies that $U_{s,\infty}(x)$ is non-singular on each subinterval $[r,\ t]$ of X_0 with $r<s<t$, and hence $U_{s,\infty}(x)$ is non-singular on X_0.

The final statement of the theorem is an immediate consequence of the fact that $U_{st}(x)=U_{rt}(x)U^{-1}_{rt}(s)$, $V_{st}(x)=V_{rt}(x)U^{-1}_{rt}(s)$ for $r,\ s,\ t\in X_0$, $r\neq t,\ s\neq t$.

If (2.4) is oscillatory on X then there exists a t such that there are points s of X which precede t and are conjugate to t, and consequently there is a largest such conjugate point $s=c(t)$ preceding t. For a system (2.4) satisfying H_R it follows from Theorem 4.1 that if $c(t)$ exists for a value $t=t_1$ then $c(t)$ exists for $t_1<t<\infty$ and increases with t. In accordance with the terminology introduced by Morse and Leighton [11] for a scalar second order linear differential equation, the first conjugate point $c(\infty)$ of $x=\infty$ on X is defined as the limit of $c(t)$ as $t\rightarrow\infty$. Clearly such a system (2.4) is non-oscillatory for large x if and only if either (2.4) is non-oscillatory on X or $c(\infty)$ exists and is finite. If $c(\infty)$ exists and is finite then (2.4) is non-oscillatory on $(c(\infty),\ \infty)$, so that the interval X_0 of Theorem 5.1 may be chosen as this interval, and consequently for $c(\infty)<s<\infty$ the matrix of conjoined solutions $(U_{s,\infty}(x)\,;\ V_{s,\infty}(x))$ has $U_{s,\infty}(x)$ non-singular on $(c(\infty),\ \infty)$. On the other hand, the definition of $c(\infty)$ implies that (2.4) is oscillatory on an arbitrary subinterval $(a_0,\ \infty)$ of X with $a_0<c(\infty)$, and Theorem 4.1 implies that $U_{s,\infty}(x)$ is singular at some point of such a subinterval $(a_0,\ \infty)$, so that by continuity $U_{s,\infty}(x)$ is singular for $x=c(\infty)$. That is, if H_R holds and (2.4) is non-oscillatory for large x then the matrix of conjoined solutions $(U_{s,\infty}(x)\,;\ V_{s,\infty}(x))$ of Theorem 5.1 is such that $c(\infty)$ exists on X if and only if $U_{s,\infty}(x)$ is singular at some point of X, in which case $c(\infty)$ is the largest value of x for which $U_{s,\infty}(x)$ is singular.

6. Principal solutions. From Theorem 5.1 it follows that if (2.4) satisfies H_R and is non-oscillatory on X_0: $a_0<x<\infty$ then there exist matrix solutions $(U(x)\,;\ V(x))$ of (2.4′) with $U(x)$ non-singular on X_0. The basic result on principal solutions for such a system (2.4) is contained in the following theorem.

THEOREM 6.1. *Suppose that the equation* (2.4) *satisfies* H_R *and is non-oscillatory on a subinterval* X_0 : $a_0<x<\infty$ *of* X. *If* $(U(x)\,;\ V(x))$ *is a solution of* (2.4′) *with* $U(x)$ *non-singular on an interval* X_U : $a_U<x<\infty$ *then for* s *a point common to* X_0 *and* X_U *the matrix*

$$(6.1) \qquad M(s\,;\,U)=\lim_{t\to\infty} S^{-1}(t,\,s\,;\,U)$$

exists and is finite. Moreover, $M(s\,;\,U)=0$ *and* $(U(x)\,;\,V(x))$ *is a principal solution of* (2.4′) *if and only if* $U(x)=U_{r,\infty}(x)C$, $V(x)=V_{r,\infty}(x)C$, *where* r *is any fixed value on* X_0, $(U_{r,\infty}(x)\,;\,V_{r,\infty}(x))$ *is the matrix of conjoined solutions as determined by Theorem* 5.1, *and* C *is a non-singular constant matrix.*

In view of Theorems 3.2 and 5.1 it clearly suffices to establish the result of the above theorem for $s=r$ a point common to X_0 and X_U. For such a value s it follows from Theorem 3.1 that

$$U_{s,\infty}(x)=U(x)T(x,\,s\,;\,U)[U^{-1}(s)+S(x,\,s\,;\,U)\{U,\,U_{s,\infty}\}]\,,$$

$$U_{st}(x)=U(x)T(x,\,s\,;\,U)[E-S(x,\,s\,;\,U)S^{-1}(t,\,s\,;\,U)]U^{-1}(s)\,,$$

and since $U_{st}(x)\to U_{s,\infty}(x)$, $V_{st}(x)\to V_{s,\infty}(x)$ as $x\to\infty$ it follows that $M(s\,;\,U)$ defined by (6.1) exists and has the finite value

$$(6.2) \qquad M(s\,;\,U)=-\{U,\,U_{s,\infty}\}U(s)\,.$$

In particular, (6.2) implies that $M(s\,;\,U)=0$ if and only if $\{U,\,U_{s,\infty}\}=0$. As $0=\{U_{s,\infty},\,U_{s,\infty}\}=V_{s,\infty}(s)-V^*_{s,\infty}(s)$ it follows that $0=\{U,\,U_{s,\infty}\}=U^*(s)V_{s,\infty}(s)-V^*(s)U_{s,\infty}(s)=U^*(s)V^*_{s,\infty}(s)-V^*(s)$ if and only if $(U(s)\,;\,V(s))$ satisfies with the non-singular matrix $C=U(s)$ the initial conditions $U(s)=U_{s,\infty}(s)C$, $V(s)=V_{s,\infty}(x)C$, and therefore $U(x)\equiv U_{s,\infty}(x)C$, $V(x)\equiv V_{s,\infty}(x)C$.

In particular, under the hypotheses of Theorem 6.1 it follows that if $(U(x)\,;\,V(x))$ is a principal solution of (2.4′) then $(U(x)\,;\,V(x))$ is a matrix of conjoined solutions of (2.4), and therefore $T(x,\,s\,;\,U)\equiv E$. As the first conclusion of Theorem 3.3 with $U_0(x)=U(x)$ implies that if (2.4) has a solution $(U(x)\,;\,V(x))$ with $U(x)$ non-singular for large x, and $T(x,\,s\,;\,U)\to 0$ as $x\to\infty$, then $(U(x)\,;\,V(x))$ is a principal solution, the following corollary is direct consequence of the results of Theorems 3.3, 6.1, and formula (6.2).

COROLLARY. *In case* (2.4) *satisfies* H_R, *and is non-oscillatory for large* x, *then*:

(i) *if* $(U(x)\,;\,V(x))$ *is a solution of* (2.4′) *with* $U(x)$ *non-singular on* X_0: $a_0<x<\infty$, *and* $s\in X_0$, *then it is not true that* $T(x,\,s\,;\,U)\to 0$ *as* $x\to\infty$;

(ii) *if* $(U(x)\,;\,V(x))$ *is a principal solution of* (2.4′), *then for a solution* $(U_0(x)\,;\,V_0(x))$ *of* (2.4′) *the matrix* $\{U,\,U_0\}$ *is non-singular if and only if* $U_0(x)$ *is non-singular for large* x *and* $U_0^{-1}(x)U(x)\to 0$ *as* $x\to\infty$, *moreover, if* $\{U,\,U_0\}$ *is non-singular then, for* s *sufficiently large,* $\lim_{t\to\infty} S(t,\,s,\,U_0)$ *exists and is non-singular.*

Finally, we shall establish the following result; in particular, conclusion (v) generalizes a result of Hartman [5].

THEOREM 6.2. *Suppose that* (2.4) *satisfies* H_R *and is non-oscillatory on a subinterval* $X_0: a_0<x<\infty$ *of* X, *while* $(U_{s,\infty}(x); V_{s,\infty}(x))$, $s\in X_0$, *is the matrix of conjoined solutions as determined by Theorem* 5.1. *If* $(U(x); V(x))$ *is a solution of* (2.4′) *with* $U(x)$ *non-singular on* X_0, *and* $S(\infty, r; U)=\lim_{v\to\infty} S(x, r; U)$ *exists and is finite for some* $r\in X_0$, *then for arbitrary* $s\in X_0$:

(i) $S(\infty, s; U)$ *exists, and*

$$(6.3)\quad S(\infty, s; U)=T(s, r; U)[S(\infty, r; U)-S(s, r; U)] \text{ for } s, x\in X_0;$$

(ii) $\{U, U_{s,\infty}\}$ *is non-singular*;

(iii) $U^{-1}(x)U_{s,\infty}(x)\to 0$ *as* $x\to\infty$;

(iv) $\{U, U_{s,\infty}\}-\{U, U\}U^{-1}(s)$ *is non-singular, and* $T(\infty, s; U)=\lim_{x\to\infty} T(x, s; U)$ *exists and is equal to the non-singular matrix* $\{U_{s,\infty}, U\}^{-1}[\{U_{s,\infty}, U\}-U^{*-1}(s)\{U, U\}]$;

(v) $U_{s,\infty}(x)=-U(x)S(\infty, x; U)\{U, U_{s,\infty}\}$.

Conclusion (i) is an immediate consequence of relation (3.10). Now, as established in the proof of Theorem 6.1, the matrix $M(s; U)=\lim_{t\to\infty} S^{-1}(t, s; U)$ exists and has the finite value $-\{U, U_{s,\infty}\}\ U(s)$, so if $S(\infty, s; U)$ exists and is finite we have

$$(6.4)\qquad E=-S(\infty, s; U)\{U, U_{s,\infty}\}U(s),$$

and hence $\{U, U_{s,\infty}\}$ is non-singular; in turn it follows from the Corollary to Theorem 6.1 that (ii) implies (iii).

In order to establish conclusion (iv), it is noted that the non-singularity of $U(x)$ on X_0 implies the validity of (3.8) with $U_0=U_{s,\infty}(x)$, so that

$$(6.5)\qquad \begin{aligned}\{U, U_{s,\infty}\}-\{U, U\}U^{-1}(x)U_{s,\infty}(x)\\ =T^{*-1}(x, s; U)[\{U, U_{s,\infty}\}-\{U, U\}U^{-1}(s)]\end{aligned}$$

for $s, x\in X_0$. From conclusions (ii), (iii) and relation (6.5) it follows that if ξ is a constant vector satisfying $[\{U, U_{s,\infty}\}-\{U, U\}U^{-1}(s)]\xi=0$ then $\xi=0$, so that $\{U, U_{s,\infty}\}-\{U, U\}U^{-1}(s)$ is non-singular for $s\in X_0$. This result, together with conclusions (ii), (iii) and relation (6.5), imply that for $s\in X_0$ the matrix $T^{*-1}(x, s; U)$ approaches the non-singular matrix $\{U, U_{s,\infty}\}[\{U, U_{s,\infty}\}-\{U, U\}U^{-1}(s)]^{-1}$, which is equivalent to the final statement of conclusion (iv).

Finally, it is to be noted that (6.4) is equivalent to

$$E=-U(x)S(\infty, x; U)\{U, U_{x,\infty}\}, \text{ for } x\in X_0,$$

and as $U_{s,\infty}(t)=U_{x,\infty}(t)U_{s,\infty}(x)$, $V_{s,\infty}(t)=V_{x,\infty}(t)U_{s,\infty}(x)$ for $s, t, x\in X_0$ it follows that $\{U, U_{x,\infty}\}U_{s,\infty}(x)=\{U, U_{s,\infty}\}$ and $U_{s,\infty}(x)=-U(x)S(\infty, x; U)$ $\{U, U_{s,\infty}\}$ for $x, s\in X_0$, thus establishing conclusion (v).

7. **An example.** In the notation of the preceding sections, the example of Section 11 of Hartman [5] shows that for an equation (2.4) which satisfies H_R, and is non-oscillatory for large x, there may exist solutions $(U(x); V(x))$ of (2.4′) with $U(x)$ non-singular for large x and such that

$$\left[\int_s^x U^{-1}(t)B(t)U^{*-1}(t)dt\right]^{-1}\to 0 \text{ as } x\to\infty , \tag{7.1}$$

while $(U(x); V(x))$ is not a principal solution. As shown by Theorem 6.1, for general solutions $(U(x); V(x))$ of (2.4′) with $U(x)$ non-singular for large x the discriminating property for principal solutions is not (7.1), but rather $S^{-1}(x, s; U)\to 0$ as $x\to\infty$. We shall proceed to illustrate the results of the preceding sections by the example of Hartman.

For typographical simplification a 2×2 matrix $||M_{\alpha\beta}||$, $(\alpha, \beta=1, 2)$, will be displayed as $M=(M_{11}; M_{12}; M_{21}; M_{22})$. In this notation the two-dimensional vector equation of Hartman's example is

$$u''+P(x)u=0,\ 0<x<\infty,\ \textit{with}\ P(x)=(0;\ 0;\ 0;\ (4x^2)^{-1}) . \tag{7.2}$$

For (7.2) the matrix solutions $(U_{st}(x); V_{st}(x)=U'_{st}(x))$ of Theorem 5.1 have

$$U_{st}(x)=((x-t)/(s-t);\ 0;\ 0;\ (x/s)^{1/2}\ (\ln t-\ln x)/(\ln t-\ln s)) .$$

and consequently $(U_{s,\infty}(x); V_{s,\infty}(x))$ has $U_{s,\infty}(x)=(1; 0; 0; (x/s)^{1/2})$. Hartman's example involves the principal solution $(U_{1,\infty}(x); V_{1,\infty}(x))$ for which $U_{1,\infty}(x)=(1; 0; 0; x^{1/2})$, and the matrix solution $(U(x); V(x))$ having $U(x)=(1; x; 0; x^{1/2})$. For these matrix solutions one may compute readily the following quantities;

$$\begin{aligned}
&S(x, s; U_{s,\infty})=(x-s;\ 0;\ 0;\ s(\ln x-\ln s)) ,\\
&\{U, U\}=(0;\ 1;\ -1;\ 0),\ \{U_{1,\infty}, U\}=(0;\ 1;\ 0;\ 0) ,\\
&T(x, 1; U)=(1-x\ln x;\ 1-x-x\ln x;\ \ln x;\ 1+\ln x) ,\\
&S(x, 1; U)=(x-1+x\ln x;\ -\ln x;\ -x\ln x;\ \ln x) ,\\
&M(1; U)=(0;\ 0;\ 1;\ 1);\ U^{-1}(x)U_{1,\infty}(x)=(1;\ -x;\ 0;\ 1) .
\end{aligned}$$

It is to be noted that $\{U_{1,\infty}, U\}$ is singular, so that the corollary to Theorem 6.1 implies that the matrix $U^{-1}(x)U_{1,\infty}(x)$ does not tend to 0 as $x\to\infty$, a fact that is obvious from the specific value of this matrix.

To illustrate further the results of the preceding section, consider the solution $(U_1(x); V_1(x))$ of (2.4′) with $U_1(x)=(x; 1; 0; x^{1/2}\ln x)$. For this solution $U_1(x)$ is non-singular for $x>1$, and one has

$$U_1^{-1}(x)U_{s,\infty}(x)=(1/x\,;\ -1/(s^{1/2}\,x\,\ln\,x)\,;\ 0\,;\ 1/(s^{1/2}\,\ln\,x))\,,$$
$$\{U_1,\ U_1\}=(0\,;\ -1\,;\ 1\,;\ 0),\ \{U_{s,\infty},\ U_1\}=(1\,;\ 0\,;\ 0\,;\ s^{-1/2})\,,$$
$$x^3(\ln\,x)^2U_1^{-1}(x)B(x)U_1^{*-1}(x)=(1+x(\ln\,x)^2\,;\ -x\,;\ -x\,;\ x^2)\,.$$

Moreover, if $\theta=\theta(x,\ s)=(1/\ln\,x)-(1/\ln\,s)$, it may be verified that

$$T(x,\ s\,;\ U_1)=(1-\theta/x\,;\ (x-s-\theta)/(sx)\,;\ \theta\,;\ 1+\theta/s)\,,$$
$$S(x,\ s\,;\ U_1)=((x-s-\theta)/(sx)\,;\ \theta/s\,;\ \theta/x\,;\ -\theta)\,,$$
$$(x-s)S^{-1}(x,\ s\,;\ U_1)=(xs\,;\ x\,;\ s\,;\ 1-(x-s)/\theta)\,,$$

from which one may verify readily that for $1<s<\infty$,

$$T(\infty,\ s\,;\ U_1)=(1\,;\ 1/s\,;\ -1/\ln\,s\,;\ 1-1/(s\,\ln\,s))\,,$$
$$S(\infty,\ s\,;\ U_1)=(1/s\,;\ -1/(s\,\ln\,s)\,;\ 0\,;\ 1/\ln\,s)\,,$$
$$M(s\,;\ U_1)=(s\,;\ 1\,;\ 0\,;\ \ln\,s)\,.$$

8. **Further properties of principal solutions.** Suppose that (2.4) satisfies H_R, and is non-oscillatory on a subinterval $X_0:\ a_0<x<\infty$ of X; for $s,\ t \in X_0,\ s<t$, let $Y_{st}(x)=U_{st}(x)$ on $x\leqq t$, and $Y_{st}(x)\equiv 0$ on $x\geqq t$, where, as in Theorem 5.1, $(U_{st}(x)\,;\ V_{st}(x))$ is the solution of (2.4′) satisfying $U_{st}(s)=E,\ U_{st}(t)=0$.

For brevity, if $y(x)$, $u(x)$ are differentially admissible vector functions on $[s.\ \infty)$ such that

$$\lim_{t\to\infty}I[y,\ u\,;\ s,\ t] \tag{8.1}$$

exists and is finite, the value of (8.1) will be denoted by $I[y,\ u\,;\ s]$, moreover, for brevity we shall write $I[y\,;\ s]$ in place of $I[y,\ y\,;\ s]$. In particular, for arbitrary constant vectors ξ we have $I[Y_{st}\xi\,;\ s]=I[U_{st}\xi\,;\ s,\ t]$. Now from relations (5.3) and (4.2) it follows that

$$0<\xi^*[V_{sd}(s)-V_{st}(s)]\xi=I[Y_{st}\xi\,;\ s]-I[Y_{sd}\xi\,;\ s]=I[Y_{st}\xi-Y_{sd}\xi\,;\ s]$$

for $s<t<d,\ s\in X_0$, and since $V_{st}(s)\to V_{s,\infty}$ as $t\to\infty$ it follows that for $s\in X_0$, and ξ an arbitrary constant vector,

$$I[Y_{st}\xi-Y_{sd}\xi\,;\ s]\to 0 \ \textit{as}\ t,\ d\to\infty\,. \tag{8.2}$$

It is to be emphasized that in general it is not true that

$$-\xi^*V_{s,\infty}(s)\xi=I[U_{s,\infty}\xi\,;\ s],\ \textit{for}\ s\in X_0\,, \tag{8.3}$$

although $-\xi^*V_{st}(s)\xi=I[Y_{st}\xi\,;\ s]$ for $t>s$, and $Y_{st}(x)\xi\to U_{s,\infty}(x)\xi$ as $t\to\infty$; moreover, in general it is not true that the vector function $U_{s,\infty}(x)\xi$ is bounded on $[s,\ \infty)$, although $Y_{st}(x)\xi\equiv 0$ for $x\geqq t$. The statements are illustrated by the well-known scalar second order equation $u''+u/(4x^2)=0$, which is non-oscillatory on $(0,\ \infty)$; for this equation $u_{1,\infty}(x)=x^{1/2}$

and $v_{1,\infty}(1)=1/2$, while $\omega(x, u_{1,\infty}, u'_{1,\infty})\equiv 0$. However, much more can be said about the principal solutions $(U_{s,\infty}(x); V_{s,\infty}(x))$ in case the hermitian integrand function ω is such that

(8.4) $\omega(x, y, \pi)\geqq 0$ *for arbitrary* x, y, π *with* $x \in X_0$.

In view of the continued understanding that $R(x)$ is non-singular on X, it is clear that (8.4) implies H_R, as well as the result that $H_+[s, t]$ holds for arbitrary compact subintervals $[s, t]$ of X_0, so that (2.4) is non-oscillatory on X_0.

THEOREM 8.1. *If condition* (8.4) *holds on a subinterval* $X_0 : a_0<x<\infty$ *of* X *then* (8.3) *is valid; moreover,* $U^*_{s,\infty}(x)V_{s,\infty}(x)\leqq 0$ *on* $s\leqq x<\infty$ *and* $U^*_{s,\infty}V_{s,\infty}\to 0$ as $x\to\infty$.

Since $V_{st}(s)\to V_{s,\infty}(s)$, and the vector function $Y_{st}(x)\xi$ tends to $U_{s,\infty}(x)\xi$ uniformly on each compact subinterval of $[s, \infty)$ as $t\to\infty$, whenever condition (8.4) holds on X_0 it follows readily from the relation $-\xi^* V_{st}(s)\xi =I[Y_{st}\xi ; s]$ that $I[U_{s,\infty}\xi ; s]$ exists and

$$-\xi^* V_{s,\infty}(s)\xi\geqq I[U_{s,\infty}\xi ; s] .$$

Now $V_{s,\infty}(s)$ is hermitian and by (4.1) we have

$$-\xi^* V_{s,\infty}(s)\xi=I[Y_{st}\xi, U_{s,\infty}\xi ; s, t]=I[Y_{st}\xi, U_{s,\infty}\xi ; s] .$$

Moreover, whenever (8.4) holds we have the Schwarz inequality

$$|I[Y_{st}\xi, U_{s,\infty}\xi ; r]|^2\leqq I[Y_{st}\xi ; r]I[U_{s,\infty}\xi ; r] \textit{ for } s<r<\infty,$$

and as $I[Y_{st}\xi ; r]\leqq I[Y_{st}\xi ; s]\leqq I[Y_{sp}\xi ; s]$ for $t\geqq p>s$ it follows that for given $p>s$, $\varepsilon>0$ there exists a value $r=r_\varepsilon>s$ such that

$$-\xi^* V_{s,\infty}(s)\xi\leqq \Re(I[Y_{st}\xi, U_{s,\infty}\xi ; s, r])+\varepsilon \textit{ for } t\geqq p .$$

As $\Re(I[Y_{st}\xi, U_{s,\infty}\xi ; s, r])\to I[U_{s,\infty}\xi ; s, r]$ as $t\to\infty$, and $I[U_{s,\infty}\xi ; s, r] \leqq I[U_{s,\infty}\xi ; s]$ by (8.4), it follows that $-\xi^* V_{s,\infty}(s)\xi\leqq I[U_{s,\infty}\xi ; s]$, thus completing the proof of (8.3). Finally, condition (8.4) implies that for ξ a non-zero constant vector the integral $I[U_{s,\infty}\xi ; s, r]=\xi^*[U^*_{s,\infty}(r)V_{s,\infty}(r) -V_{s,\infty}(s)]\xi$ is a monotone increasing function of r on $s<r<\infty$ which tends to $I[U_{s,\infty}\xi ; s]=-\xi^* V_{s,\infty}(s)\xi$ as $r\to\infty$, and consequently $U^*_{s,\infty}(r) V_{s,\infty}(r)\leqq 0$ on (s, ∞) and $U^*_{s,\infty}(r)V_{s,\infty}(r)\to 0$ as $r\to\infty$.

In particular, if $R(x)\equiv E$, $Q(x)\equiv 0$ and $P(x)\geqq 0$ on X, then the above theorem implies that $(|U_{s,\infty}(x)\xi|^2)'=2\xi^*U^*_{s,\infty}(x)V_{s,\infty}(x)\xi\leqq 0$, so that for such an equation (2.4) the norm of the vector function $U_{s,\infty}(x)\xi$ tends to a limit as $x\to\infty$. This particular result has been established by Wintner [16].

It is to be emphasized that condition (8.4) does not imply that

$U_{s,\infty}\to 0$ as $x\to\infty$. For example, (8.4) holds for the scalar equation

$$(u'/(e^x+2))' - 2u/(e^x+2)^2 = 0$$

with general solution $u = c_1(1+e^{-x}) + c_2 e^x$, and principal solution $u_{0,\infty}(x) = (1+e^{-x})/2$.

THEOREM 8.2. *If H_R holds and* (2.4) *is non-oscillatory on a subinterval X_0: $a_0 < x < \infty$ of X then $U_{s,\infty}(x)\to 0$ as $x\to\infty$ if there exists a constant $k>0$ and a continuous positive function $h(x)$ such that if s, $d \in X_0$, $s<d$, then*

$$I[y\,;\,s,\,d] \geqq k\int_s^d [h(x)|y'|^2 + |y|^2/h(x)]dx \tag{8.5}$$

for arbitrary $y(x)$ which are differentially admissible on $[s, d]$ and satisfy $y(s)=0=y(d)$.

If the vector function $y(x)$ is differentially admissible on $[s, d]$, and $y(s)=0=y(d)$, then

$$\begin{aligned} 2|y(x)|^2 &= \int_s^x (y^*y' + y^{*\prime}y)dx - \int_x^d (y^*y' + y^{*\prime}y)dx\,, \\ &\leqq 2\int_s^d |y|\,|y'|dx \leqq \int_s^d [h(x)|y'|^2 + |y|^2/h(x)]dx\,, \end{aligned}$$

the last inequality holding for arbitrary continuous positive functions $h(x)$. Consequently the hypothesis of Theorem 8.2 implies that there is a positive constant k such that

$$2k|y(x)|^2 \leqq I[y\,;\,s,\,d] \text{ for } s \leqq x \leqq d \tag{8.6}$$

holds if s, $d \in X_0$, $s<d$, and $y(x)$ is a differentially admissible vector function on $[s, d]$ with $y(s)=0=y(d)$. In particular, if $s<t<d$ and ξ is a constant vector, then $y(x) = Y_{st}(x)\xi - Y_{sd}(x)\xi$ is such a vector function with $y(x)\equiv 0$ for $x \geqq d$ and $I[y\,;\,s,\,d] = I[y\,;\,s]$, so that

$$2k|Y_{st}(x)\xi - Y_{sd}(x)\xi|2 \leqq I[Y_{st}\xi - Y_{sd}\xi\,;\,s], \quad s \leqq x < \infty. \tag{8.7}$$

Inequalities (8.2), (8.7) then imply that as $t\to\infty$ the convergence of $Y_s(x)\xi$ to $U_{s,\infty}(x)\xi$ is uniform on $s \leqq x < \infty$. As $Y_{st}(x)\xi \equiv 0$ for $x \geqq t$ it then follows that $U_{s,\infty}(x)\xi \to 0$ as $x\to\infty$ for arbitrary constant vectors ξ, so that $U_{s,\infty}(x)\to 0$ as $x\to\infty$.

THEOREM 8.3. *If on a subinterval X_0: $a_0 < x < \infty$ of X we have $Q(x) \equiv 0$, $R(x)$ of class C' with $R(x)>0$, $R'(x)\leqq 0$, and there is a non-negative continuous function $k(x)$ such that $\int^\infty k(x)dx$ is divergent and $y^*P(x)y \geqq k(x)y^*R(x)y$ for arbitrary vectors y, then $U^*_{s,\infty}(x)R(x)U_{s,\infty}(x)\to 0$ as $x\to\infty$.*

The hypotheses of the theorem clearly imply condition (8.4) on X_0. Now if $Q(x)\equiv 0$ and $R(x)$ is of class C' we have $V_{s,\infty}(x)=R(x)U'_{s,\infty}(x)$, and as $\{U_{s,\infty}, U_{s,\infty}\}=0$ it follows that $(U^*_{s,\infty}RU_{s,\infty})'=2U^*_{s,\infty}V_{s,\infty}+U^*_{s,\infty}R'U_{s,\infty}$, so that in view of the condition $R'(x)\leqq 0$ and the last conclusion of Theorem 8.1 we have $(U^*_{s,\infty}RU_{s,\infty})'\leqq 0$ on X_0. Consequently, for an arbitrary constant vector ξ the non-negative function $\xi^*U^*_{s,\infty}(x)R(x)U_{s,\infty}(x)\xi$ is non-increasing on X_0, and thus tends to a non-negative limit as $x\to\infty$. Moreover, by Theorem 8.1 the integral $I[U_{s,\infty}\xi\,;\,s]$ exists and is finite, so that in view of the relation

$$I[U_{s,\infty}\xi\,;\,s]\geqq\int_s^\infty \xi^*U^*_{s,\infty}PU_{s,\infty}\xi\;dx\geqq\int_s^\infty k(x)[\xi^*U^*_{s,\infty}RU_{s,\infty}\xi]dx\;,$$

and the divergent character of $\int^\infty k(x)dx$, it follows that $\xi^*U^*_{s,\infty}(x)R(x)U_{s,\infty}(x)\xi\to 0$ as $x\to\infty$, for ξ an arbitrary constant vector.

As a particular instance of the above theorem we have the following result.

COROLLARY. *If on a subinterval* X_0: $a_0<x<\infty$ *of* X *we have* $Q(x)\equiv 0$, $R(x)$ *a constant matrix* $R>0$, *and there is a non-negative continuous function* $k_1(x)$ *such that* $\int^\infty k_1(x)dx$ *is divergent and* $y^*P(x)y\geqq k_1(x)|y|^2$ *for arbitrary vectors* y, *then for* $s\in X_0$ *we have* $U_{s,\infty}(x)\to 0$ *as* $x\to\infty$.

For the case of a scalar equation the result of the above corollary in essence dates from Kneser [7], as has been pointed out by Wintner [15].

Added November 20, 1957. P. Hartman has pointed out to the author that the following argument establishes the conclusion of Theorem 8.3 with the hypothesis that $\int^\infty k(x)dx$ is divergent replaced by the weaker condition that $\int^\infty xk(x)dx$ is divergent. Since Theorem 8.1 implies that $U^*_{s,\infty}V_{s,\infty}\leqq 0$, from the condition $U^*_{s,\infty}R'U_{s,\infty}\leqq 0$ and the expression given for $(U^*_{s,\infty}RU_{s,\infty})'$ in the proof of Theorem 8.3 it follows that the integral $\int^\infty U^*_{s,\infty}V_{s,\infty}\,dx$ exists. From Theorem 8.1 it follows that $U^*_{s,\infty}V_{s,\infty}\to 0$ and

$$-\xi^*U^*_{s,\infty}(u)V_{s,\infty}(u)\xi=I[U_{s,\infty}\xi\,;\,u]\geqq\int_u^\infty \xi^*U^*_{s,\infty}PU_{s,\infty}\xi dx$$

for $a_0<u<\infty$ and arbitrary constant vectors ξ, and as $U^*_{s,\infty}PU_{s,\infty}\geqq 0$ the integrals $\int_u^\infty U^*_{s,\infty}PU_{s,\infty}dx$ and $\int_u^\infty\left[\int_x^\infty U^*_{s,\infty}PU_{s,\infty}dt\right]dx$ exist for $a_0<u<\infty$; an integration by parts then yields the existence of the integral $\int_u^\infty xU^*_{s,\infty}(x)P(x)U_{s,\infty}(x)dx$. Consequently the condition that $y^*P(x)y\geqq k(x)y^*R(x)y$ for arbitrary vectors y implies that the integral

$\int_u^\infty xk(x)U^*_{s,\infty}(x)R(x)U_{s,\infty}(x)dx$ exists, and in view of the relations $U^*_{s\infty}RU_{s,\infty} \geqq 0$, $(U^*_{s,\infty}RU_{s,\infty})' \leqq 0$ it follows that $U^*_{s,\infty}RU_{s,\infty} \to 0$ whenever $\int^\infty xk(x)dx$ is divergent.

9. **A more general differential system.** In this section we shall consider a differential system with complex coefficients that is of the general form of the accessory differential equations for a variational problem of Bolza type, (see, for example, Bliss [3 ; § 81] and Reid [12]). As in § 2, $\omega(x, y, \pi)$ will denote an hermitian form (2.1) with $R(x)$, $Q(x)$, $P(x)$ $n \times n$ matrices having complex-valued continuous elements on X: $a < x < \infty$, and $R(x)$, $P(x)$ hermitian on this interval. In addition, consider a vector linear form

$$\Phi(x, y, \pi) \equiv \varphi(x)\pi + \theta(x)y \, , \tag{9.1}$$

where $\varphi(x)$ and $\theta(x)$ are $m \times n$, $(m < n)$, matrices with complex-valued continuous elements on X. Instead of the hypothesis of Section 2 that $R(x)$ is non-singular, it is now assumed that the $(n+m) \times (n+m)$ hermitian matrix

$$\begin{Vmatrix} R(x) & \varphi^*(x) \\ \varphi(x) & 0 \end{Vmatrix} \tag{9.2}$$

is non-singular on X; in particular, the non-singularity of (9.2) on X implies that $\varphi(x)$ is of rank m on this interval.

For the variational problem involving the functional (2.2) subject to the auxiliary m-dimensional vector differential equation

$$\Phi(x, y, y') = 0 \tag{9.3}$$

the Euler-Lagrange differential equations are in vector form

$$\begin{gathered} (R(x)u' + Q(x)u + \varphi^*(x)\mu)' - (Q^*(x)u' + P(x)u + \theta^*(x)\mu) = 0 \, , \\ \Phi(x, u, u') = 0 \, , \end{gathered} \tag{9.4}$$

where $u(x)$ is an n-dimensional vector function and $\mu(x)$ is an m-dimensional "multiplier" vector function.

The inverse of the non-singular matrix (9.2) is of the form

$$\begin{Vmatrix} T(x) & \tau^*(x) \\ \tau(x) & t(x) \end{Vmatrix} ,$$

where $T(x)$ and $t(x)$ are hermitian matrices of orders n and m, respectively, and $\tau(x)$ is an $m \times n$ matrix. In terms of the canonical variables

$$u(x),\ v(x)=R(x)u'(x)+Q(x)u(x)+\varphi^*(x)\mu(x)$$

the Euler-Lagrange equations (9.4) become a vector differential system (2.4), with now

$$(9.5)\qquad A=-(TQ+\tau^*\theta),\ B=T,\ C=P-Q^*TQ-Q^*\tau^*\theta-\theta^*\tau Q-\theta^*t\theta\ ;$$

the matrices B and C of (9.5) are hermitian on X, while B is a non-negative definite matrix of rank $n-m$ with $B\varphi^*=0$ throughout this interval. Throughout this section we shall continue to refer to the vector equation (2.4) and the corresponding matrix equation (2.4′), *with the understanding that the coefficient matrices are given by* (9.5).

As in Section 2, if $(U_1(x)\,;\ V_1(x))$ and $(U_2(x)\,;\ V_2(x))$ are solutions of (2.4′) then the matrix $U_1^*(x)V_2(x)-V_1^*(x)U_2(x)$ is a constant; to denote this matrix by $\{U_1,\ U_2\}$ now in general involves an ambiguity, however, since if $(U(x)\,;\ V(x))$ is a solution of (2.4′) there may exist other matrices $V_0(x)\neq V(x)$ such that $(U(x)\,;\ V_0(x))$ is also a solution of (2.4′). This ambiguity does not exist, however, if (2.4) is such that whenever $u(x)\equiv 0$, $v(x)$ is a solution of this equation on a non-degenerate subinterval of X then $v(x)\equiv 0$ on this subinterval; if this property holds the equation (2.4) is said to be *identically normal*, or to be *normal on every subinterval*, on X. It is to be commented that this condition of normality was used in Section 3 to show that if (2.4) is non-oscillatory on X_0, and $(U(x)\,;\ V(x))$ is a solution of (2.4′) with $U(x)$ non-singular on this interval, then $S(t,\ s\,;\ U)$ is non-singular for $s,\ t\in X_0,\ s\neq t$.

For the equation (2.4) now under consideration one may define the concepts of conjugate point, non-oscillation on a subinterval, and non-oscillation for large x, in precisely the language of Section 2. For the problem involving the functional (2.2) subject to the differential equation (9.3) an n-dimensional vector function $y(x)$ will now be said to be differentially admissible on a subinterval of X if on this subinterval $y(x)$ is continuous, has piecewise continuous derivatives, and satisfies (9.3); for a compact subinterval $[c,\ d]$ of X the symbol $H_+[c,\ d]$ will again denote the condition that $I[y\,;\ c,\ d]>0$ for arbitrary differentially admissible $y(x)$ which are not identically zero on $[c,\ d]$ and satisfy $y(c)=0=y(d)$. For the problem now considered the symbol H_R signifies the condition that for all $x\in X$ we have $\pi^*R(x)\pi>0$ for arbitrary non-zero vectors π satisfying the restraint $\varphi(x)\pi=0$; in view of the basic assumption that (9.2) is non-singular throughout X it follows that H_R holds whenever there is a single $s\in X$ such that $\pi^*R(s)\pi>0$ for arbitrary non-zero vectors π satisfying $\varphi(s)\pi=0$.

With the above definitions, the result of Theorem 4.1 is valid for the equation (2.4) now under consideration. In this connection, it is to be commented that if we write $y=(y^1_\alpha+iy^2_\alpha)$, $(\alpha=1,\cdots,n)$, and denote by z the real $2n$-dimensional vector function with components $(y^1_1,\cdots,y^1_n, y^2_1,\cdots,y^2_n)$, then $\omega(x, y, y')$ is a quadratic form $\omega_0(x, z, z')$ in (z, z') with real coefficients, and (9.3) is equivalent to a real $2m$-dimensional vector differential equation $\Phi_0(x, z, z')=0$. Moreover, $H_+[c, d]$ and H_R are individually equivalent to the corresponding conditions $H^0_+[c, d]$ and H^0_R for the associated real problem in z, and for this latter problem the conclusion that $H^0_+[c, d]$ implies H^0_R is a well-known result of the calculus of variations, (see, for example, Bliss [3; Theorem 78.2 and Lemma 81.2]). For a problem of the sort formulated above which satisfies H_R, the method of proof of Lemma 89.1 of Bliss [3] yields the result that $H_+[c, d]$ holds if and only if there is a matrix $(U(x); V(x))$ of conjoined solutions of (2.4) with $U(x)$ non-singular on $[c, d]$, and the method of proof of Lemma 89.2 of Bliss [3] establishes that $H_+[c, d]$ holds if and only if (2.4) is non-oscillatory on $[c, d]$.

For a differential system (2.4) of the type now under consideration, the result of Theorem 4.2 is valid only if this system is normal on the interval $[c, d]$, since if $y(x)$ is differentially admissible then $y(c)$, $y(d)$ must satisfy $v^*(d)y(d)-v^*(c)y(c)=0$ with all vector functions $v(x)$ belonging to abnormal solutions $u\equiv 0$, $v(x)$ of (2.4) on $[c, d]$. On the other hand, if (2.4) is normal on every subinterval of X then Theorems 4.2 and 4.3 hold, as well as relations (4.1) and (4.2) for vector functions that are differentially admissible for the problem of this section.

From the above remarks it follows that for systems (2.4) with coefficient matrices given by (9.5), and *which are normal on every subinterval of* X, the various theorems of Sections 3–6 remain valid, with no changes in proofs required. An important illustration of this class of systems (2.4) is afforded by certain systems (2.4) that are equivalent to self-adjoint scalar differential equations of even order. Indeed, suppose that $p_j(x)$, $(j=0, 1, \cdots, 2n)$, are real-valued functions with $p_{2n}(x)\neq 0$ on X and $p_j(x)$ of class $C^{(j/2)}$ or $C^{((j+1)/2)}$ according as j is even or odd, and let $R(x)$, $Q(x)$, $P(x)$ be diagonal matrices with $P_{\alpha\alpha}(x)=(-1)^{\alpha-1}p_{2\alpha-2}(x)$, $Q_{\alpha\alpha}(x)=i(-1)^\alpha p_{2\alpha-1}(x)$, $(\alpha=1,\cdots,n)$, $R_{\alpha\alpha}(x)\equiv 0$ for $\alpha<n$ and $R_{nn}(x)=(-1)^n p_{2n}(x)$, while $\Phi(x, y, \pi)=(\pi_\beta-y_{\beta+1})$, $(\beta=1,\cdots, n-1)$. The corresponding vector differential system (2.4) is readily seen to be normal on every subinterval, and $(u(x); v(x))$ is a solution of this system if and only if $u_\alpha(x)=y^{(\alpha-1)}(x)$, $(\alpha=1,\cdots,n)$, where $y(x)$ is a solution of the self-adjoint differential equation

$$\sum_{\alpha=0}^{n}[p_{2\alpha}(x)y^{(\alpha)}]^{(\alpha)}+i\sum_{\alpha=1}^{n}([p_{2\alpha-1}(x)y^{(\alpha-1)}]^{(\alpha)}+[p_{2\alpha-1}(x)y^{(\alpha)}]^{(\alpha-1)})=0\ .$$

It is to be noted also that for a system (2.4) normal on every subinterval the results of Theorems 8.1 and 8.2 are valid, with (8.4) replaced by the condition that $\omega(x, y, \pi) \geqq 0$ for arbitrary $(x, y; \pi)$ with $x \in X_0$, and satisfying $\Phi(x, y, \pi)=0$.

Finally, it is to be remarked that for an equation (2.4) with coefficients given by (9.5), and which is not normal on every subinterval of X, there do exist suitable modifications of Theorems 4.2 and 4.3 which with an altered definition of principal solution enable one to establish certain results corresponding to those of Sections 5,6 ; however, the details of these results will not be presented here.

BIBLIOGRAPHY

1. J. H. Barrett, *Matrix systems of second order differential equations*, Portugaliae Mathematica, **14** (1955), 79–89.
2. ———, *A Prüfer transformation for matrix differential equations*, Proc. Amer. Math. Soc., **8** (1957), 510-518.
3. G. A. Bliss, *Lectures on the calculus of variations*, University of Chicago Press, 1946.
4. P. Hartman and A. Wintner, *On the assignment of asymptotic values for the solutions of linear differential equations of the second order*, Amer. J. Math., **77** (1955), 475–483.
5. P. Hartman, *Self-adjoint, non-oscillatory systems of ordinary, second order, linear differential equations*, Duke Math. J., **24** (1957), 25–35.
6. M. R. Hestenes, *Sufficient conditions for the problem of Bolza in the calculus of variations*, Trans. Amer. Math. Soc., **36** (1934), 793–818.
7. A. Kneser, *Untersuchung und asymptotische Darstellung der Integrale gewisser Differentialgleichungen bei grossen reellen Werthen des Arguments, I.*, J. reine angew. Math., **116** (1896), 178–212.
8. W. Leighton, *Principal quadratic functionals*, Trans. Amer. Math. Soc. **67** (1949), 253-274.
9. M. Morse, *A generalization of the Sturm separation and comparison theorems in n-space*, Math. Ann., **103** (1930), 72–91.
10. M. Morse, *The calculus of variations in the large*, Amer. Math. Coll. Publ. XVIII, (1934).
11. M. Morse and W. Leighton, *Singular quadratic functionals*, Trans. Amer. Math. Soc., **40** (1936), 252–286.
12. W. T. Reid, *A matrix differential equation of Riccati type*, Amer. J. Math., **68** (1946), 237–246; *Addendum*, ibid, **70** (1948), 460.
13. W. T. Reid, *Oscillation criteria for linear differential systems with complex coefficients*, Pacific J. Math., **6** (1956), 733–751.
14. R. L. Sternberg and H. Kaufman, *Applications of the theory of systems of differential equations to multiple non-uniform transmission lines*, J. Math. Phys. **31** (1952–53), 244–252.
15. A. Wintner, *On the Laplace-Fourier transcendents occurring in mathematical physics*, Amer. J. Math., **69** (1947), 87-98.
16. A. Wintner, *On linear repulsive forces*, Amer. J. Math., **71** (1949), 362–366.

NORTHWESTERN UNIVERSITY.

ON GENERAL MINIMAX THEOREMS

MAURICE SION

1. Introduction. von Neumann's minimax theorem [10] can be stated as follows: if M and N are finite dimensional simplices and f is a bilinear function on $M \times N$, then f has a saddle point, i. e.:

$$\max_{\mu \in M} \min_{\nu \in N} f(\mu, \nu) = \min_{\nu \in N} \max_{\mu \in M} f(\mu, \nu) .$$

There have been several generalizations of this theorem. J. Ville [9], A. Wald [11], and others [1] variously extended von Neumann's result to cases where M and N were allowed to be subsets of certain infinite dimensional linear spaces. The functions f they considered, however, were still linear. M. Shiffman [8] seems to have been the first to have considered concave-convex functions in a minimax theorem. H. Kneser [6], K. Fan [3], and C. Berge [2] (using induction and the method of separating two disjoint convex sets in Euclidean space by a hyperplane) got minimax theorems for concave-convex functions that are appropriately semi-continuous in one of the two variables. Although these theorems include the previous results as special cases, they can also be shown to be rather direct consequences of von Neumann's theorem. H. Nikaidô [7], on the other hand, using Brouwer's fixed point theorem, proved the existence of a saddle point for functions satisfying the weaker algebraic condition of being quasi-concave-convex, but the stronger topological condition of being continuous in each variable.

Thus, there seem to be essentially two types of argument: one uses some form of separation of disjoint convex sets by a hyperplane and yields the theorem of Kneser-Fan (see 4.2), and the other uses a fixed point theorem and yields Nikaidô's result.

In this paper, we unify the two streams of thought by proving a minimax theorem for a function that is quasi-concave-convex and appropriately semi-continuous in each variable. The method of proof differs radically from any used previously. The difficulty lies in the fact that we cannot use a fixed point theorem (due to lack of continuity) nor the separation of disjoint convex sets by a hyperplane (due to lack of convexity). The key tool used is a theorem due to Knaster, Kuratowski, Mazurkiewicz based on Sperner's lemma.

It may be of some interest to point out that, in all the minimax theorems, the crucial argument is carried out on spaces M and N that

Received June 26, 1957. This research was supported by the United States Air Force, Office of Scientific Research and Development Command, under contract No. AF18(600)-1109 [Supplemental Agreement No. 4 (56-339)].

are finite dimensional simplices. When concave-convexlike functions are considered, the topological conditions of compactness and semi-continuity are used only in reducing the problem to the finite dimensional case. For quasi-concave-convex functions, however, semi-continuity is needed in a more crucial way, as can be seen from the example in 3.6.

2. **Fundamental notions and definitions.** The following definitions of concavelike and convexlike functions were first considered by K. Fan [3]. They generalize the concepts of concavity and convexity and are valid for spaces without linear structure.

2.1. A function f on $M \times N$ is *concavelike* in M if for every $\mu_1, \mu_2 \in M$ and $0 \leqq t \leqq 1$, there is a $\mu \in M$ such that

$$tf(\mu_1, \nu) + (1-t)f(\mu_2, \nu) \leqq f(\mu, \nu) \quad \text{for all } \nu \in N .$$

2.2. A function f on $M \times N$ is *convexlike* in N if for every $\nu_1, \nu_2 \in N$ and $0 \leqq t \leqq 1$, there is a $\nu \in N$ such

$$tf(\mu, \nu_1) + (1-t)f(\mu, \nu_2) \geqq f(\mu, \nu) \quad \text{for all } \mu \in M .$$

2.3. A function f on $M \times N$ is *concave-convexlike* if it is concavelike in M and convexlike in N.

2.4. A function f on $M \times N$ is *quasi-concave* in M if $\{\mu : f(\mu, \nu) \geqq c\}$ is a convex set for any $\nu \in N$ and real c.

2.5. A function f on $M \times N$ is *quasi-convex* in N if $\{\nu : f(\mu, \nu) \leqq c\}$ is a convex set for any $\mu \in M$ and real c.

2.6. A function f on $M \times N$ is *quasi-concave-convex* if it is quasi-concave in M and quasi-convex in N.

2.7. A function f on $M \times N$ is u. s. c.-l. s. c. if $f(\mu, \nu)$ is upper semi-continuous in μ for each $\nu \in N$ and lower semi-continuous in ν for each $\mu \in M$.

2.8. For a function f on $M \times N$, we set

$$\sup \inf f = \sup_{\mu \in M} \inf_{\nu \in N} f(\mu, \nu) ,$$

$$\inf \sup f = \inf_{\nu \in N} \sup_{\mu \in M} f(\mu, \nu) .$$

2.9. The convex hull of X will be denoted by $\ulcorner X \urcorner$.

2.10. The closure of X will be denoted by $\overline{X}$.

3. **Minimax theorems for quasi-concave-convex functions.** The aim of this section is Theorem 3.4. The method of proof, making use of 3.1, 3.2, and 3.3, is very different from any argument used previously in obtaining minimax theorems.

3.1. THEOREM. *Let S be an n-dimensional simplex with vertices $a_0, \cdots, a_n$. If $A_0, \cdots, A_n$ are open sets such that $S \subset \bigcup_{i=0}^{n} A_i$, $S - A_i$ is convex, and $a_i \notin A_j$ for $i \neq j$ $(i, j, = 0 \cdots, n)$, then $\bigcap_{i=0}^{n} A_i \neq 0$.*

Proof. We can set $A_i = \bigcup_{k=0}^{\infty} B_{i,k}$ where the $B_{i,k}$ are open and $\bar{B}_{i,k} \subset B_{i,k+1}$. Since S is compact, there is an integer N such that $S \subset \bigcup_{i=0}^{n} \bar{B}_{i,N}$. By a theorem of Knaster, Kuratowski, Mazurkiewicz [5], we have $\bigcap_{i=0}^{n} A_i \supset \bigcap_{i=0}^{n} \bar{B}_{i,N} \neq 0$.

3.2. THEOREM. *Let $\mathfrak{A} = \{a_0, \cdots, a_n\}$ consist of $n+1$ points in a linear space of dimension $k < n$. Then $\bigcap_{i=0}^{n} \ulcorner(\mathfrak{A} - \{a_i\})\urcorner \neq 0$.*

Proof. $\bigcap_{\substack{i=0\\ i\neq j}}^{n} \ulcorner(\mathfrak{A} - \{a_i\})\urcorner \supset \{a_j\} \neq 0$ for $j = 0, \cdots, n$.

Hence by Helly's Theorem [14], we have the desired result.

3.3. LEMMA. *Let M be a convex set, Y a finite set, and f a function on $M \times Y$, quasi-concave and upper semi-continuous in M. Suppose, in addition, that Y is minimal with respect to the property: for each $\mu \in M$ there is a $y \in Y$ with $f(\mu, y) < c$. Then there exists $\mu_0 \in M$ such that $f(\mu_0, y) < c$ for all $y \in Y$.*

Proof. Let $Y = \{y_0, \cdots, y_n\}$ and set $A_i = \{\mu :\ f(\mu, y_i) < c\}$ for $i = 0, \cdots, n$. Then the A_i are open and $M - A_i$ convex. By hypothesis, for each i, there exists $a_i \in M$ such that $a_i \in M - A_j$ for $j \neq i$. Let $\mathfrak{A} = \{a_0, \cdots, a_n\}$. Then $\ulcorner(\mathfrak{A} - \{a_i\})\urcorner \subset M - A_i$ and, since $M \subset \bigcup_{i=0}^{n} A_i$, we must have $\bigcap_{i=0}^{n} \ulcorner(\mathfrak{A} - \{a_i\})\urcorner = 0$. Hence, by 3.2., $\mathfrak{A}$ spans an n-dimensional simplex in M and, by 3.1, there exists a $\mu_0 \in \bigcap_{i=0}^{n} A_i$.

3.3′. LEMMA. *Let N be a convex set, X a finite set, and f a function on $X \times N$, quasi-convex and lower semi-continuous in N. Suppose, in addition, that X is minimal with respect to the property: for each $\nu \in N$ there is an $x \in X$ with $f(x, \nu) > c$. Then there exists $\nu_0 \in N$ such that $f(x, \nu_0) > c$ for all $x \in X$.*

3.4. THEOREM. *Let M and N be convex, compact spaces, and f a function on $M \times N$, quasi-concave-convex and* u. s. c.-l. s. c.. Then sup inf $f = \inf$ sup f.

Proof. Suppose sup inf $f < c < \inf$ sup f. Let $A_\mu = \{\nu : f(\mu, \nu) > c\}$ and $B_\nu = \{\mu : f(\mu, \nu) < c\}$. The A_μ are open and cover N. Since N is compact, a finite number of the A_μ cover N. Similarly, a finite number of the B_ν cover M. We can therefore choose finite subsets $X_1 \subset M$ and $Y_1 \subset N$ such that for each $\nu \in N$, and hence for each $\nu \in \ulcorner Y_1 \urcorner$, there is an $x \in X_1$ with $f(x, \nu) > c$; and for each $\mu \in M$, and hence for each $\mu \in \ulcorner X_1 \urcorner$, there is a $y \in Y_1$ with $f(\mu, y) < c$.

Let X_2 be a minimal subset of X_1 such that for each $\nu \in \ulcorner Y_1 \urcorner$ there is an $x \in X_2$ with $f(x, \nu) > c$. Next, let Y_2 be a minimal subset of Y_1 such that for each $\mu \in \ulcorner X_2 \urcorner$ there is a $y \in Y_2$ with $f(\mu, y) < c$.

Thus, by repeating this process of alternately reducing the X_i and Y_i, after a finite number of steps, we can choose finite subsets $X \subset M$ and $Y \subset N$ such that X is minimal with respect to the property : for each $\nu \in \ulcorner Y \urcorner$ there is an $x \in X$ with $f(x, \nu) > c$; and Y is minimal with respect to the property : for each $\mu \in \ulcorner X \urcorner$ there is a $y \in Y$ with $f(\mu, y) < c$. By 3.3, there exists $\mu_0 \in \ulcorner X \urcorner$ such that $f(\mu_0, y) < c$ for all $y \in Y$ and hence (by quasi-convexity) $f(\mu_0, \nu) < c$ for all $\nu \in \ulcorner Y \urcorner$. By 3.3′, there exists $\nu_0 \in \ulcorner Y \urcorner$ such that $f(x, \nu_0) > c$ for all $x \in X$ and hence (by quasi-concavity) $f(\mu, \nu_0) > c$ for all $\mu \in \ulcorner X \urcorner$. Then $c < f(\mu_0, \nu_0) < c$, which is impossible.

3.3. COROLLARY. *Let M and N be convex spaces one of which is compact, and f a function on $M \times N$, quasi-concave-convex and* u. s. c. -l. s. c.. *Then* sup inf $f = \inf$ sup f.

Proof. Suppose M is compact and sup inf $f < c < \inf$ sup f. Then there exists a finite set $Y \subset N$ such that for any $\mu \in M$ there is a $y \in Y$ with $f(\mu, y) < c$. Taking $f' = f|(M \times \ulcorner Y \urcorner)$, we get sup inf $f' < c < \inf$ sup f' in contradiction to 3.4 with N replaced by $\ulcorner Y \urcorner$ and f by f'.

3.6. REMARK. In Theorem 3.4, the condition that f be u. s. c.-l s. c. cannot be removed nor appreciably weakened even if the spaces M, N are finite dimensional. To see this, we consider the following example. Let $M = N = [0, 1]$ and $f(\mu, \nu) = 0$ for $0 \leqq \mu < 1/2$ and $\nu = 0$ or $1/2 \leqq \mu \leqq 1$ and $\nu = 1$; $f(\mu, \nu) = 1$ otherwise. We easily check that f is quasi-concave-convex ; for each μ, $f(\mu, \nu)$ is lower semi-continuous in ν ; however $f(\mu, 1)$ is not upper semi-continuous in μ. We also have : sup inf $f = 0$ and inf sup $f = 1$.

4. Minimax theorems for concave-convexlike functions. For concave-convexlike functions, the topology for the spaces on which they are defined plays only a secondary role. Theorem 4.2 (4.2′) below, which is the generalization of Kneser's theorem to concave-convexlike functions due to K. Fan [3], is not a special case of 3.4 since the concepts of concave-convexlike and quasi-concave-convex are independent of each other (see [7]). It is however a special case of 4.1′ (4.1), which is itself an immediate consequence of 3.4 (actually, von Neumann's theorem).

4.1. THEOREM. *Let M and N be any spaces, f a function on $M\times N$ that is concave-convexlike. If for any $c<\inf\sup f$ there exists a finite subset $X\subset M$ such that for any $\nu\in N$ there is an $x\in X$ with $f(x,\nu)>c$, then* $\sup\inf f=\inf\sup f$.

4.1′. THEOREM. *Let M, N be any spaces, f a function on $M\times N$ that is concave-convexlike. If for any $c>\sup\inf f$ there exists a finite set $Y\subset N$ such that for any $\mu\in M$ there is a $y\in Y$ with $f(\mu,y)<c$, then* $\sup\inf f=\inf\sup f$.

4.2. THEOREM. (Kneser, Fan). *Let M be compact, N any space, f a function on $M\times N$ that is concave-convexlike. If $f(\mu,\nu)$ is upper semi-continuous in μ for each ν, then* $\sup\inf f=\inf\sup f$.

Proof. If $c>\sup\inf f$, let $A_\nu=\{\mu: f(\mu,\nu)<c\}$ for each $\nu\in N$. The A_ν are open and cover M, hence a finite number of them cover M. We may therefore apply 4.1′.

4.2′ THEOREM. *Let M be any space, N compact, f a function on $M\times N$ that is concave-convexlike. If $f(\mu,\nu)$ is lower semi-continuous in ν for each μ, then* $\sup\inf f=\inf\sup f$.

REFERENCES

1. *Contributions to the theory of games*, vol. **I**, Ann. of Math. Studies, No. 24, 1950, Princeton Univ. Press, Princeton.
2. C. Berge, *Sur une convexité régulière et ses applications à la théorie des jeux*, Bull. Soc. Math. France, **82** (1954), 301–319.
3. K. Fan, *Minimax Theorems*, Proc. Nat. Acad. Sci., **39** (1953), 42–47.
4. E. Helly, *Über Mengen könvexer Korper mit gemeinschaftlichen Punkten*, J. Deutsch. Math. Vereinig., **32** (1923), 175–176.
5. B. Knaster, C., Kuratowski, S. Mazurkiewicz, *Ein Beweis des Fixpunktsatzes für n-dimensionale Simplexe*, Fund. Math., **14** (1929), 132–138.
6. H. Kneser, *Sur un théorème fondamental de la thérie des jeux*, C. R. Acad. Sci. Paris **234** (1952), 2418–2420.
7. H. Nikaidô, *On von Neumann's minimax theorem*, Pacific J. Math., **4** (1954), 65–72.
8. M. Shiffman, *On the equality min max=max min, and the theory of games*, RAND Report RM-243, 1949.

9. J. Ville, *Sur la théorie générale des jeux ou intervient l'habilité des joueurs. Traité du calcul des probabilites et de ses applications*, IV, **2** (1938), 105-113; by E. Borel and collaborators.
10. J. von Neumann, *Zur Theorie der Gesellshaftsphiele*, Math. Ann. **100** (1928) 295-320.
11. A. Wald, *Generalization of a theorem by von Neumann concerning zero-sum two-person games*. Ann. of Math., **46** (1945), 281-286.

THE INSTITUTE FOR ADVANCED STUDY

ON SEMI-NORMED *-ALGEBRAS

CHIEN WENJEN

1. Introduction. The notion of semi-normed algebras was introduced by Arens as a generalization of Banach algebras [**2, 5**]. They are called locally multiplically-convex algebras by Michael [**16**]. Various properties of Banach algebras have been generalized to semi-normed algebras [**5, 16, 21, 22, 23**].

We repeat here a few definitions. Let A be a linear algebra over the field K of complex or real numbers. A nonnegative real-valued function V defined on A is called a semi-norm if it satisfies the following conditions:

$V(x+y)\leqq V(x)+V(y)$, $V(xy)\leqq V(x)V(y)$, $V(\lambda x)=|\lambda|V(x)$. Suppose there is a family $\mathscr{V}$ of semi-norms such that $V(x)=0$ for all $V\in\mathscr{V}$ only if $x=0$. A is a semi-normed algebra if all the translations of the sets on which $V(x)<e$, where e is real and $V\in\mathscr{V}$, are taken as a subbase of topology, and is complete if it is complete with respect to the uniform structure defined by the various relations $V(x-y)<e$. A is called an *-algebra if there is a semi-linear operation * such that $(\lambda x-yz)^*=\bar{\lambda}x^*-z^*y^*$, $x^{**}=x$. A subset U of A is called idempotent if $UU\subset U$; it is called multiplicatively convex (m-convex) if it is convex and idempotent. A is locally m-convex if there exists a basis for the neighbourhoods of the origin consisting of sets which are m-convex and symmetric.

The present paper is devoted to generalizing the representation theorems for commutative and noncommutative Banach algebras to semi-normed algebras. An application of the Gelfand-Neumark-Arens representation theorem for commutative Banach algebras yields a simple proof of the spectral theorem for bounded self-adjoint operators in Hilbert space [**14**, p. 95]. Our generalized representation theorem for commutative semi-normed algebras gives rise to a similar proof of the spectral theorem for unbounded self-adjoint operators.

The characterization of the algebra $C(T, K)$ of all complex-valued continuous functions on a locally compact, paracompact Hausdorff space T has been treated by Arens [**5**, p. 469]. We have a characterization theorem for $C(T, K)$ where T is a locally compact completely regular space and also a uniqueness theorem for the space T [cf. the Banach-Stone theorem, **6**, p. 170, 20, p. 469]: If $C(T_1, K)$, $C(T_2, K)$ are topo-

Received July 19, 1954, and in revised form September 1, 1957. This paper is part of a thesis submitted by the writer to the graduate division of the University of California, Los Angeles, (Summer, 1953) in partial satisfaction for the Ph. D. degree. The writer is indebted to Professor Richard Arens for his encouragement and valuable advice.

logically isomorphic, then T_1 and T_2 are homeomorphic. If T_1, T_2 are Hewitt's Q-spaces [11, p. 85], the topological equivalence between the spaces follows from the algebraic isomorphism between $C(T_1, K)$ and $C(T_2, K)$, but not in general.

2. Functional representation.

2.1. THEOREM. *Let A be a complete commutative semi-normed *-algebra* (with or without a unit) *over the complex numbers K such that*

2.2. $V(xx^*) \geqq k_V V(x^*)$, *for all $V \in \mathscr{V}(k_V > 0)$. Then A is topologically isomorphic to a complete self-adjoint subalgebra S of the algebra $C(T, K)$ of all continuous complex-valued functions (vanishing at infinity if A has no unit) on T with k-topology, where T is the union of the members of a family of pairwise disconnected and closed-open sets. (compact if A has a unit, otherwise locally compact).*

Proof. The elements x in A satisfying $V(x)=0$ form an ideal Z_V, a kernel ideal of A. The quotient algebra A/Z_V is a normed algebra when V is used to define a norm, and the completion B_V of A/Z_V is a commutative Banach *-algebra. By Gelfand-Neumark-Arens representation theorem [3, Theorem 1, p. 278], there exists a Hausdorff space (compact if A has a unit, otherwise locally compact) $Q_V = V$-neighbourhood homomorphism, for which B_V is the class of all complex-valued continuous functions (vanishing at infinity if A has no unit) on Q_V such that

$$x_V{}^*(q) = \overline{x_V(q)} \qquad (q \in Q_V,\ x \in B_V).$$

and

$$k_V V(x_V) \leqq \sup_{q \in Q_V} |x_V(q)| \leqq V(x_V) \,. \tag{2.3}$$

Let

$$T = \bigcup_{V \in \mathscr{V}} Q_V \,.$$

Retaining the original weak* topology for Q_V and regarding all Q_V as pairwise disconnected and closed-open subsets, we have a locally compact completely regular space T. The complex-valued continuous functions on T are of the form $f(t) = \{f_V\}$, where $f_V(t) \in C(Q_V, K)$ and $f(t) = f_V(t)$ if $t \in Q_V$.

The mapping

$$P: \qquad x \in A \rightarrow x(t) = \{x_V(t)\} \in C(T, K)$$

maps A onto a subalgebra S of $C(T, K)$. P is isomorphic; for, if x maps to zero functional, then $V(x)=0$ for all $V \in \mathscr{V}$ and x is the zero element of A.

In fact, P is a homeomorphism. Denote the open set in A consisting of all x such that $V(x)<e$ by $0(V, e)$ and the open set in $C(T, K)$ defined by $\sup_{q\in Q_V} |f(q)|<k_V e$ by $0'(Q_V, e)$. It follows from the inequalities 2.3 that P maps $0(V, e)$ onto a subset of $C(T, K)$ containing $0'(Q_V, e)$. This proves the continuity of the inverse mapping of P from S onto A.

Let W be a compact subset in T contained in the union of $Q_{V_1},\cdots, Q_{V_n}$. It is clear that P maps the intersection of $0(V_1, e),\cdots, 0(V_n, e)$ onto a subset in $C(T, K)$ contained in the intersection of $0'(V_1, e/k_V),\cdots, 0(V_n, e/k_V)$, and S, that is, in the intersection of $0'(W, e/k_V)$ and S. P is therefore continuous.

The completeness of S is an immediate consequence of the completeness of A and inequalities 2.3.

2.4. COROLLARY. *Let M_V be a maximal ideal in B_V (the completion of the quotient ring $A_V=A/Z_V$) and let $f(t)$ be a complex-valued continuous function on the space T. Then $f(t)$ belongs to S if $f_V(M_U)=f_U(M_U)$ whenever $U\leqq V$.*

Proof. M_V is actually a point in Q_V and $f_V(M_U)$ belongs to $C(Q_V, K)$. Let $\overline{\Pi}_{UV}$ be the natural mapping of B_V into B_U when $U\leqq V$. Then $\overline{\Pi}_{UV}(f_V)=f_U$ whenever $U\leqq V$ if $f_V(M_U)=f_U(M_U)$. Hence the corollary [**16**, Theorem 5.1].

This immediately yields the following result [cf. 5, p. 471].

2.5. THEOREM. *Let A be a commutative complete semi-normed $*-$algebra with a unit* (without unit) *satisfying* 2.2. *Then an element x in A has an inverse* (reverse) *if $x(M)\neq 0$ ($x(M)\neq -1$) for each closed maximal ideal M in A.*

3. **Spectrum.** An element h in a complete semi-normed *-algebra A satisfying 2.2 is called Hermitian, if $h^*=h$; and an Hermitian element h is called positive, if its spectrum consists of nonnegative numbers.

3.1. THEOREM. *The spectrum of every Hermitian element h is real.*

Proof. Suppose A has a unit. Let A_1 be the minimal complete *-subalgebra of A containing h. Then A_1 is commutative. By Theorem 2.1 A_1 is equivalent to a closed subalgebra S of $C(T, K)$. The corresponding function $h(M)$ of the element h in A is real-valued. For any nonreal number λ, the function $h(M)-\lambda$ is not equal to zero anywhere. The theorem follows from Theorem 2.5.

3.2. THEOREM. *Every closed self-adjoint subalgebra A_0 of a complete semi-normed *-algebra A with a unit* (without unit) *satisfying* 2.2 *contains inverses* (reverses).

Proof. Rickart has proved that $x_V \in A_{0V}$ (the completion of $A_{0v}=A_0/Z_V$) has an inverse (reverse) iff both $x_V^* x_V$ and $x_V x_V^*$ have inverses (reverses) and that the inverse (reverse) of x_V is contained in A_{0V} iff the inverses (reverses) of $x_V^* x_V$ and $x_V x_V^*$ are contained in A'_{0V} [18, pp. 531–532]. Since every closed maximal ideal in A contains a kernel ideal [5, p. 466], it follows from Theorem 2.5 that A_0 contains inverses (reverses) of its Hermitian elements, and hence of all its elements which have inverses (reverses) in A.

3.3. COROLLARY. *Let A_0 be any closed self-adjoint subalgebra of A. Then the spectrum of $x \in A_0$ relative to A_0 is identical with the spectrum relative to A.*

3.4. THEOREM. *Let x be a normal element, that is, $xx^*=x^*x$, of A* (with or without a unit) *and let $f(\lambda)$ be a complex-valued continuous function* (vanishing at infinity, if A has no unit) *defined on the spectrum σ of x. Then $f(x)$ defines an element contained in every commutative closed self-adjoint subalgebra of A which contains x.*

Moreover if $s(\lambda)=f(\lambda)+g(\lambda)$, $p(\lambda)=f(\lambda)g(\lambda)$, $q(\lambda)=\bar{f}(\lambda)$, $r(\lambda)=\lambda$, *then* $s(x)=f(x)+g(x)$, $p(x)=f(x)g(x)$, $q(x)=f(x)^*$, $r(x)=x$.

Proof. Let A_0 be a commutative closed self-adjoint subalgebra of A containing x and let M_V be a maximal ideal in A_{0V}. Then A_0 is equivalent to a closed self-adjoint subalgebra S of the algebra $C(T, K)$ of all complex-valued continuous functions on a locally compact completely regular space T and $f(x_V(M_U))=f(x_U(M_U))$ whenever $U \leqq V$. By Corollary 2.4, $f(x(M))$ determines a unique element, denoted by $f(x)$, contained in A_0. The first part of the theorem is proved.

The second part of the theorem is obvious.

3.5. THEOREM. *The sum of two positive elements is positive.*

Proof. Suppose A has a unit. Let h and k be two positive elements in A and let A_0 be the minimal closed self-adjoint subalgebra of A containing $h+k$. Since the inverse of $h_V+k_V+\lambda e$ for any nonnegative number λ and each $V \in \mathscr{V}$. [**13**, p. 52] the function $h(M)+k(M)+\lambda$ does not vanish at any M. The theorem follows from Theorem 2.5.

3.6. THEOREM. *The Hermitian elements of a complete seminormed *-algebra satisfying the condition* 2.2 *constitute a lattice.*

Proof. To any Hermitian h, there is a positive element $|h|$ corresponding to the function $|\lambda|$ by Theorem 3.4. Let h and k be arbitrary Hermitian elements and define.

$h \vee k = \frac{1}{2}(h+k+|h-k|)$, $h \wedge k = \frac{1}{2}(h+k-|h-k|)$. Then the Hermitian elements constitutes a lattice.

4. Closed self-adjoint subalgebras.

4.1. THEOREM. *A commutative complete semi-normed *-algebra A satisfying the condition* 2.2 *is equivalent to a closed, separating self-adjoint subalgebra S of the algebra $C(T_0, K)$ of all complex-valued continuous functions* (vanishing at infinity, if A has no unit) *on a completely regular space T_0 with a topology which has at most the open sets of the k-topology, that is, with a topology $\rho \leqq k$.*

Proof. By Theorem 2.1, A is equivalent to a closed self-adjoint subalgebra S of $C(T, K)$, where T is a union of pairwise disconnected and closed-open sets (compact if A has a unit, otherwise locally compact). Let $x(t)$ be the corresponding function in S of the element x in A. Denote by T_0 the class of all subsets of T:

$$L_a = \{t \,;\, x(t) = x(a) \text{ for each } x \in A\} .$$

Following Cěch's notation, Let ρ denote the mapping:

$$a \in T \rightarrow L_a$$

and let $[f, I]$ denote those elements $\rho(t)$ of T_0 such that $f(t) \in I$, where $f(t)$ is a continuous real function belonging to S end I is an open interval. The topology generated by considering all these $[f, I]$ as a subbase is called ρ-topology.

It is easy to see that ρ is a continuous mapping and that for any $a \in T$, there is an $[f, I]$ containing $\rho(a)$. Let $[f_1, I_1]$ and $[f_2, I_2]$ be any two open sets in T_0 containing $\rho(a)$. If both $f_1(a)$ and $f_2(a)$ are different from zero, we can assume without loss of generality that $f_1(a) = f_2(a)$ and that I_1 and I_2 are identical. We define $g_i(t) = f_i(t)$ if $f_i(t) \leqq f_i(a)$, and $g_i(t) = 2f_i(a) - f_i(t)$ if $f_i(t) > f_i(a)$, $i = 1, 2$. Then $g_1(t)$ and $g_2(t)$ are continuous functions. Let $g(t) = g_1(t) \wedge g_2(t)$. It is clear that $[g, I] \subset [f_1, I] \cap [f_2, I]$. In case $f_1(a) = 0$ and $f_2(a) \neq 0$, we can assume that $f_1(t)$ and $f_2(t)$ are nonnegative. Let $g(t) = f_2(t) - f_1(t)$. An interval I can be so chosen that $[g, I] \subset [f_1, I] \cap [f_2, I]$. Hence T_0 is a topological space. Čech has proved that T_0 is Hausdorff and completely regular [8, p. 827].

Now the closed subalgebra S of $C(T, K)$ is a closed, separating subalgebra of $C(T_0, K)$.

4.2. REMARK. It is clear that the elements in the space T_0 are the closed maximal ideals in the algebra A and the ρ-topology is the weak* topology. Professor Arens has constructed examples to show that T_0 is not necessarily locally compact. He has also constructed a completely regular space T such that $C(T, K)$ with k-topology is not complete. [4, p. 234]. We have, however, the following.

4.3. THEOREM. *The necessary and sufficient condition that a commutative complete semi-normed *-algebra A satisfying the condition* 2.2 *be equivalent to $C(T, K)$, with k-topology, of all complex-valued continuous functions on a locally compact completely regular space T is:*

To any closed maximal ideal M_0 in A, there are an $x \in A$ and an $\varepsilon > 0$ such that the intersection of the maximal ideals M satisfying the relation $|x(M_0) - x(M)| \leqq \varepsilon$ contains a kernel ideal.

Proof. The necessity is obvious. The sufficiency follows from Theorem 4.1 and Corollary 2.4.

4.4. REMARK. Theorem 4.3 generalizes the theorem of Arens characterizing the algebra $C(T, K)$, where T is a locally compact, paracompact Hausdorff space. [5, p. 469]. Let A be an algebra with a locally finite partition of unity. (For definition and notation, see 5, p. 463) To any maximal closed ideal M_0, there exists an u_V such that $u_V(M_0) = \delta \neq 0$, since M_0 contains a kernel ideal. There are only a finite number of W such that $W(u_V) \neq 0$, say, $W_1, \cdots, W_n$. Let $W_0 = \max.(W_1, \cdots, W_n)$. The intersection of the closed maximal ideals M satisfying $|u_V(M_0) - u_V(M)| \leqq \delta/2$ evidently contains Z_{W_0}.

4.5. THEOREM. *For the algebra $C(T, K)$ of all complex-valued continuous functions* (vanishing at infinity) *on a locally compact completely regular space T with k-topology, there is one-to-one correspondence between closed ideals in $C(T, K)$ and the closed subsets of T.*

This is a generalization of a theorem due to Stone [20, Theorem 85] and the proof is straightforward.

4.6. COROLLARY. *For the glgebra of all complex-valued continuous functions* (vanishing at infinity) *on a locally compact completely regular space with k-topology, there is one-to-one correspondence between the closed maximal* (regular) *ideals of the algebra and the points of the space* (the point at infinity is not included).

4.7. THEOREM. *The necessary and sufficient condition two locally compact completely regular spaces T and T' be homeomorphic is that the* *algebras* $C(T, K)$ *and* $C(T', K)$ *of all complex-valued continuous functions* (vanishing at infinity) *on the spaces with k-topology be topologically isomorphic.*

Proof. Following Stone's idea, we define the closure of a family of closed maximal (regular) ideals in $C(T, K)$ as the hull of the kernel of the family [**14**, p. 56]. It is clear that a subset of the space T is closed iff it is equal to the hull of its kernel when it is considered as a set of the maximal (regular) ideals in $C(T, K)$.

4.8. REMARK. The homeomorphism between the spaces T and T' does not follow from the algebraic isomorphism between $C(T, K)$ and $C(T', K)$. For example, the space $T_{\Omega+1} \otimes T_{\omega+1} - (\Omega, \omega)$ [**11**, p. 69] is pseudo-compact, completely regular, locally compact, and $C(T, K)$ and $C(\beta T, K)$ are algebraically isormorphic, while T and βT are not homeomorphic.

5. Spectral theorem for unbounded self-adjoint operators in Hilbert space.

5.1. Let L be the algebra of all real-valued continuous functions defined on a locally compact Hausdorff space T and vanishing off compact sets. It is well-known that every nonnegative linear functional on L is an integral [**14**, p. 44].

A family of real-valued functions on a space is called monotone if it is closed under the operations of taking monotone increasing and decreasing limits. The functions belonging to the smallest monotone family including L are called Baire functions.

A topological space T is called hemi-compact by Arens [**1**, p. 486] if there exists a sequence T_i of compact subsets of T such that $\bigcup_{i=1}^{\infty} T_i = T$ and every compact subset of T is contained in some T_i. Every topological space which is both σ-compact and locally compact is hemi-compact.

5.2. LEMMA. *Let G be a *-representation of the algebra $C_0(T, K)$ of all complex-valued continuous functions vanishing outside compact sets on a hemi-compact Hausdorff space T, which is a union of pairwise disconnected, closed-open compact sets $T_1, T_2, \cdots$, by a family $\mathfrak{B}$ of operators in a Hilbert space H. Let H be spanned by a sequence of closed linear manifolds $H_1, H_2, \cdots$, orthogonal in pairs, such that each operator of $\mathfrak{B}$ is bunded on H_i and G is a bounded *-representation of the algebra $C(T_i, K)$ of all complex-valued continuous functions on T_i by a family of* *operators on* H_i. *Then* G *can be extented to a* **-representation of the algebra* $B(T, K)$ *of all Baire functions bounded on compact subsets of* T, *and the extension is unique, subject to the condition that* $J_{x,y}(f)=(G_f x, y)$ *is a complex-valued integral for every* $x \in H, y \in H^*$.

Proof. The function $F(f_i, x, y)=(G_{f_i}x, y)$, defined for $f_i \in C(T_i)$, $x \in H_i$, $y \in H_i^*$, is a bounded integral on $C(T_i)$ and thus is uniquely extensible to $B(T_i)$. [**14**, p. 93]. Hence the lemma [**17**, p. 312].

5.3. THEOREM. *To any self-adjoint operator* R *in a Hilbert space* H, *there exists a unique family of projections* $\{E_\lambda\}$ *depending on the parameter* λ, *satisfying*

(a) $$E_\lambda < E_\mu \quad \text{or} \quad E_\lambda = E_\lambda E_\mu \quad \text{for} \quad \lambda < \mu ,$$

(b) $$E_{\lambda+0} = E_\lambda ,$$

(c) $$\lim_{\lambda \to -\infty} E_\lambda = 0 \quad \textit{and} \quad \lim_{\lambda \to \infty} E_\lambda = I ,$$

such that

$$R = \int_{-\infty}^{\infty} \lambda dE_\lambda .$$

Proof. Let b_i be a set of real numbers, $i=0, \pm 1, \pm 2, \cdots$, such that

(1) $$\text{for all } i, b_i > b_{i-1} ;$$

(2) $$\lim_{i \to \infty} b_i = \infty ;$$

(3) $$\lim_{i \to -\infty} b_i = -\infty .$$

Then there exists a set of closed linear manifolds $\{H_i\}, i=1, 2, \cdots$, orthogonal in pairs, spanning H, and such that R is defined on H_i and satisfies the relation [**15**, 17]

$$b_i I \geqq R \geqq b_{i-1} I .$$

Let P_i be a projection on H such that $P_i x = x$ if $x \in H_i$, and $P_i x = 0$ otherwise. Now $P_1, P_2, \cdots$, and R generate a commutative semi-normed *-algebra A, the semi-norms of its elements being the norms of the operators in H_i. By Theorem 2.1, A is equivalent to a closed self-adjoint subalgebra S of the algebra $C(T, K)$ of all complex-valued continuous functions on a hemi-compact Hausdorff space T, which is a union of a sequence of pairwise disconnected, closed-open compact subsets $T_1, T_2, \cdots$. S is, in fact, the algebra $C(T, K)$ itself.

Any real continuous function $f(t)$ on the space T is a Baire function. Define a continuous function f_n such that $f_n(t)=f(t)$ if $t\in T_1\cup\cdots\cup T_n$ and $f_n(t)=0$ otherwise. Let $g_n^m\in L$ so that $g_n^m\uparrow mf_n$, where g_n^m vanish outside the sets $T_1,\cdots,T_n$, and let $g_n=g_1^n\vee\cdots\vee g_n^n$. Then $g_n\uparrow f$ and f is a Baire function. Also the characteristic functions of closed subsets in T are Baire functions.

Let $\hat{R}$ be the image of the operator R. Given $\varepsilon>0$, we can choose λ_i, $i=0, \pm1, \pm2,\cdots$ such that $\lambda_i\to\infty$, $\lambda_{-i}\to-\infty$ as $i\to\infty$ and, for all i, $\lambda_i>\lambda_{i-1}$, $\lambda_i-\lambda_{i-1}<\varepsilon$. Let $\hat{E}_\lambda$ be the characteristic function of the closed set where $\hat{R}\leqq\lambda$, and choose λ_i' from the interval $[\lambda_{i-1}, \lambda_i]$.
Then

$$\left\|\hat{R}-\sum_{-\infty}^{\infty}\lambda_i'(\hat{E}_{\lambda_i}-\hat{E}_{\lambda_{i-1}})\right\|_\infty<\varepsilon$$

and hence

$$\left\|R-\sum_{-\infty}^{\infty}\lambda_i'(E_{\lambda_i}-E_{\lambda_{i-1}})\right\|_V<\varepsilon \text{ for each } V\in\mathscr{V}.$$

The theorem is proved.

6. Imbedding algebras into rings of operators in Hilbert space.

6.1. THEOREM. *Every complete semi-normed *-algebra A with or without a unit, satisfying the condition $V(xx^*)=V(x)V(x^*)$ for each $V\in\mathscr{V}$, can be isomorphically mapped onto a closed self-adjoint subalgebra A_1 of the algebra of all linear operators in a Hilbert space $H=\sum_{V\in\mathscr{V}}H_V$ such that if $x\in A$ maps to $X\in A_1$, then X is bounded in each H_V and $V(x)=\|x\|_V$ for each $V\in\mathscr{V}$, where $\|x\|_V$ denotes the norm of X in H_V.*

Proof. By Gelfand-Neumark representation theorem [10, Theorem 1; 12, p. 409], the completed quotient algebra A_V can be isometrically mapped onto a closed self-adjoint subalgebra of the algebra of all bounded operators in Hilbert space H_V.

Let

$$H=\sum_{V\in\mathscr{V}}H_V$$

be the set of all complexes $h=\{h_V\}$, $h_V\in H_V$, with

$$\sum_{V\in\mathscr{V}}\|h\|_V^2<\infty .$$

The algebraic operations and inner products are defined as follows:

$$\lambda h=\{\lambda h_V\},\ h_1+h_2=\{h_{1V}+h_{2V}\}\ ,\quad (h_1, h_2)=\sum_{V\in\mathscr{V}}(h_{1V}-h_{2V})\ .$$

Let $h_i=\{h_{iV}\}$. Then $\|h_i-h_j\|^2=\sum_{V\in\mathcal{V}}\|h_{iV}-h_{jV}\|^2$. $\|h_i-h_j\|\to 0$ implies $\|h_{iV}-h_{jV}\|\to 0$ for each V. For any fixed V, h_{iV} approaches to an element h_{0V} in H_V as a limit when i approaches infinity. Then $h_i\to h_0=\{h_{0V}\}$ which belongs to H, and H is complete.

The corresponding operator X in H of an element $x\in A$ is defined as $X=\{X_V\}$, where X_V is the operator in H_V corresponding to $x_V\in \bar{A}_V$. Now $Xh=\{X_Vh_V\}$ with

$$\sum_{V\in\mathcal{V}}\|X_Vh_V\|^2<\infty\ .$$

The domain of X is dense in H, for it contains all those elements $\{h_V\}$ where h_V are 0 except for a finite number of them. It is clear that $X(H)\subset H$ and $X(H_V)\subset H_V$.

Bibliography

1. R. Arens, *A topology for spaces of transformations*, Ann. of Math. **47** (1946), 480-495.
2. ———, *The space L^ω and convex topological rings*, Bull. Amer, Math. Soc. **52** (1946), 931-935.
3. ———, *Representations of *-algebras*, Duke Math. J. **14** (1947), 269-282.
4. ———, *Note on convergence in topology*, Math. Mag. **23** (1950), 229-234.
5. ———, *A generalization of normed rings*, Pacific J. Math. **2** (1952), 455-471.
6. S. Banach, *Theories des operations lineares*, Monog. Mat., Warsaw, 1932.
7. G. Birkhoff, *Lattice theory*, Amer. Math. Soc. Coll. Publ. 25, New York, 1948.
8. E. Čech, *On bicompact spaces*, Ann. of Math. **38** (1937), 823-844.
9. I. Gelfand, *Normierte Ringe*, Rec. Math. (Mat. Sb.) N. S. **9**, (1941), 3-24.
10. I. Gelfand and M. Neumark, *On the imbedding of normed rings of operators in Hilbert space*, Rec. Math. (Mat. Sb.) N. S. **12** (1943), 197-213.
11. E. Hewitt, *Rings of real-valued continuous functions*, I, Trans. Amer. Math. Soc. **64** (1948), 45-99.
12. I. Kaplansky, *Normed algebras*, Duke Math. J. **16** (1949), 399-418.
13. J. L. Kelley and R. L. Vaught, *The positive cone in Banach algebras*, Trans. Amer. Math. Soc. **74** (1953), 44-55.
14. L. H. Loomis, *An introduction to abstract harmonic analysis*, New York, Van Nostrand, 1953.
15. E. R. Lorch, *Return to the self-adjoint transformations*, Acta Sci. Math. Szeged, **12**B (1950), 137-144.
16 E. Michael, *Locally multiplicativelly-convex topological algebras*, Mem. Amer. Math. Soc., no. 11, 1951.
17. F. Riesz et B. Sz.-Nagy, *Lecons d'analyse fonctionnelle*, 1953.
18. C. E. Rickart, *Banach algebras with an adjoint operation*, Ann. of Math. **47** (1946), 528-550.
19. I. E. Segal, *Representations of certain commutative Banach algebras*, Bull. Amer. Math. Soc. (abstract no. 130) **52** (1946), 421.
20. M. H. Stone, *Applications of the theory of Boolean rings to general topology*, Trans. Amer. Math. Soc. **41** (1937), 375-481.
21. S. Warner, *Weak locally multiplically-convex algebras*, Pacific J. Math. **5** (1955), 1025-1032.
22. ———, *Polynomial completeness in locally multiplicativelly-convex algebras*, Duke Math. J. **23** (1956), 1-12.
23. ———, *Inductive limits of algebras*, Trans. Amer. Math. Soc. **82** (1956), 190-216.

UNIVERSITY OF CALIFORNIA, LOS ANGELES AND TEXAS TECHNOLOGICAL COLLEGE, LUBBOCK.

Pacific Journal of Mathematics

Vol. 8, No. 1 March, 1958