Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 9, No. 4 **August, 1959**

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be typewritten (double spaced), and the author should keep a complete copy. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, L. J. Paige at the University of California, Los Angeles 24, California.

50 reprints per author of each article are furnished free of charge; additional copies may be obtained at cost in multiples of 50.

The *Pacific Journal of Mathematics* is published quarterly, in March, June, September, and December. The price per volume (4 numbers) is $12.00; single issues, $3.50. Back numbers are available. Special price to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: $4.00 per volume; single issues, $1.25.

Subscriptions, orders for back numbers, and changes of address should be sent to Pacific Journal of Mathematics, 2120 Oxford Street, Berkeley 4, California.

Printed at Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.), No. 6, 2-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

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

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

A NOTE ON KATO'S UNIQUENESS CRITERION FOR SCHRÖDINGER OPERATOR SELF-ADJOINT EXTENSIONS

F. H. BROWNELL

1. Introduction. Kato [2] has shown local square integrability with boundedness at ∞ of the potential coefficient function to be a sufficient condition for the Schrödinger operator in $L_2(R_n)$ to have a unique self-adjoint extension in case dimension $n = 3$. His statement is for $n = 3p$, thus with p factors R_3, but with the condition on V stated separately for each R_3 factor as is natural for application to quantum mechanics; this in essence amounts to $n = 3$ from our standpoint. Using the Young-Titchmarsh theorem on Fourier transforms, we generalize Kato's argument to general dimension $n \geq 1$. We show the connection of the resulting criterion with our earlier construction [1] of a self-adjoint extension as the inverse of a modified Green function integral operator. We also give a variational characterization of the spectrum here.

2. Uniqueness condition. Let $V(\boldsymbol{x})$ be a given, real-valued, measurable function over $\boldsymbol{x} \in R_n$, euclidean n-space. We consider the following additional conditions upon V, using the notation $(\boldsymbol{x} \cdot \boldsymbol{y}) = \sum_{j=1}^{n} x_j y_j$ and $|\boldsymbol{x}| = \sqrt{(\boldsymbol{x} \cdot \boldsymbol{x})}$ for $\boldsymbol{x}$ and $\boldsymbol{y} \in R_n$, and also denoting n dimensional Lebesque measure on R_n by μ_n.

CONDITION I. *For some* $b < +\infty$ *let* $V(\boldsymbol{x})$ *be essentially bounded* ($A = [\text{ess sup} \mid V(\boldsymbol{x}) \mid] < +\infty$) *over* $\{\boldsymbol{x} \in R_n \mid |\boldsymbol{x}| \geq b\}$, *and let*

$$(1) \qquad \int_{\{x \mid |\boldsymbol{x}| \leq b\}} |V(\boldsymbol{x})|^{(1/2)(n+\rho)} d\mu_n(\boldsymbol{x}) = M_\rho < +\infty$$

for some $\rho > 0$ *satisfying also* $n + \rho \geq 2$.

CONDITION II. *Let* $V(\boldsymbol{x})$ *satisfy Condition* I *with in addition* $n + \rho = 4$ *in* (1) *if dimension* $n < 4$.

Condition II is our generalization of Kato's uniqueness criterion, our following Theorem T. 1 in the special case $n = 3$ thus being due to Kato [2]. Following Kato, we define $\mathscr{D}_1 \subseteq L_2(R_n)$ as the linear manifold of Hermite functions, polynomials in the coordinates x_j multiplied by $\exp(-1/2 \mid \boldsymbol{x} \mid^2)$. Assuming Condition II), clearly the pointwise product $Vu \in L_2(R_n)$ for all $u \in \mathscr{D}_1$. Hence

Received December 30, 1958, amalgamation with addendum May 29, 1959. This work was supported by an Office of Naval Research Contract, and reproduction in whole or in part by U.S. Federal agencies is permitted.

$$(2) \qquad [H_1 u](\boldsymbol{x}) = -\nabla^2 u(\boldsymbol{x}) + V(\boldsymbol{x})u(\boldsymbol{x})$$

with $\nabla^2 = \sum_{j=1}^{n}(\partial^2/\partial x_j^2)$ the Laplacian, defines H_1 as a linear operator in $L_2(R_n)$ with dense domain $\mathscr{D}_1$. Also the easily established Green's identity for u and $w \in \mathscr{D}_1$ shows that H_1 is symmetric (see [3], p. 28–41, p. 48–50 for terminology and theorems used hereafter).

Next for $u \in L_2(R_n)$ we have existent (see [4]) the Fourier-Plancherel transform $\hat{u} \in L_2(R_n)$ defined by

$$(3) \qquad \hat{u}(\boldsymbol{y}) = \lim_{N\to\infty}\left(\frac{1}{2\pi}\right)^{n/2}\int_{\{\boldsymbol{x}\,|\,|\boldsymbol{x}|\le N\}} e^{-i(\boldsymbol{x}\cdot\boldsymbol{y})}u(\boldsymbol{x})d\mu_n(\boldsymbol{x}) ,$$

with the limit in the $L_2(R_n)$ norm sense over $\boldsymbol{y} \in R_n$; similarly

$$(4) \qquad u(\boldsymbol{x}) = \lim_{N\to\infty}\left(\frac{1}{2\pi}\right)^{n/2}\int_{\{\boldsymbol{y}\,|\,|\boldsymbol{y}|\le N\}} e^{i(\boldsymbol{x}\cdot\boldsymbol{y})}\hat{u}(\boldsymbol{y})d\mu_n(\boldsymbol{y}) ,$$

with the limit also in the $L_2(R_n)$ norm sense. In terms of (3) and (4), define $\mathscr{D}$ as the set of $u \in L_2(R_n)$ such that $|\boldsymbol{y}|^2\hat{u}(\boldsymbol{y})$ is also in $L_2(R_n)$ over $\boldsymbol{y}$. Define T as a linear operation in $L_2(R_n)$ with domain $\mathscr{D}$ by $Tu = w$, $\hat{w}(\boldsymbol{y}) = |\boldsymbol{y}|^2\hat{u}(\boldsymbol{y})$ for $u \in \mathscr{D}$, $w \in L_2(R_n)$ existing uniquely for such u since (3) and (4) define a unitary operator and its inverse on $L_2(R_n)$.

We may now state the main theorem of this section as follows. Actually, since Condition II will be seen at the end of the next section to imply Condition S stated there, this theorem is a consequence of Stummel's theorem ([5], Th 4.2), p. 171), except for an awkward but essentially trivial change of basic domain. Also our proof is rather different, being much closer to Kato's original argument. See also [6].

THEOREM T.1. *Let V satisfy condition* II. *Then the pointwise product $[Vu](\boldsymbol{x}) = V(\boldsymbol{x})u(\boldsymbol{x})$ has $Vu \in L_2(R_n)$ for $u \in \mathscr{D}$, and $Hu = Tu + Vu$ for $u \in \mathscr{D}$ has H to be a self-adjoint operator in $L_2(R_n)$ with dense domain $\mathscr{D}$. Furthermore, $\mathscr{D}_1 \subseteq \mathscr{D}$, $H_1 \subseteq H$, and H is the unique self-adjoint extension of H_1.*

Here $\mathscr{D}_1 \subseteq \mathscr{D}$, and hence $\mathscr{D}$ is dense, follows clearly from the fact ([4], p. 81, Theorem 57) that $S^*(\mathscr{D}_1) \subseteq \mathscr{D}_1$, where S denotes the unitary operator from $L_2(R_n)$ onto itself given by (4), $S\hat{u} = u$, and where $S^*u = \hat{u}$ in (3) represents the adjoint and inverse S^*. Thus $T_1 u = -V^2 u$ for $u \in \mathscr{D}_1$ has $T_1 u = Tu$ for $u \in \mathscr{D}_1$ from $[S^*(T_1 u)](\boldsymbol{y}) = |\boldsymbol{y}|^2\hat{u}(\bar{\boldsymbol{y}})$ by integration by parts; hence $T_1 \subseteq T$. Thus $H_1 \subseteq H$ follows from the following lemma (Lemma 4 of Kato [2]), which represents the heart of our argument.

LEMMA T. 2. *Let V satisfy Condition* II. *Then for $u \in \mathscr{D}$ follows both $Vu \in L_2(R_n)$ and the $L_2(R_n)$ norm inequality*

$$(5) \qquad \| Vu \| \leq \alpha \| Tu \| + \beta \| u \|$$

for some α and β positive and finite, for which α may be chosen as small as desired with β depending on α.

To prove this lemma, we will first establish that $\mathscr{D} \subseteq L_{r'}(R_n)$ with $r' = 2(n+\rho)/(n-4+\rho)$ and $\rho > 0$ given in Condition II if dimension $n \geq 4$, and that $\mathscr{D} \subseteq L_\infty(R_n)$ if $n = 1, 2$, or 3. For this purpose we start, for $u \in \mathscr{D}$ and arbitrary $\omega > 0$ and with $\rho > 0$ as in Condition II, with the Schwarz-Hölder estimate

$$(6) \int_{R_n} |\hat{u}(\boldsymbol{y})|^{2(n+\rho)/(n+\rho+4)} d\mu_n(\boldsymbol{y})$$
$$\leq \left[\int_{R_n} |\hat{u}(\boldsymbol{y})|^2 (1+\omega^4 |\boldsymbol{y}|^4) d\mu_n(\boldsymbol{y})\right]^{1/p} \left[\int_{R_n} (1+\omega^4|\boldsymbol{y}|^4)^{-(p'/p)} d\mu_n(\boldsymbol{y})\right]^{1/p'}$$
$$= [\| u \|^2 + \omega^4 \| Tu \|^2]^{1/p} \left[\sigma_n \int_0^\infty \frac{r^{n-1}}{(1+\omega^4 r^4)^{(n+\rho)/4}} dr\right]^{1/p'}$$
$$= [\| u \|^2 + \omega^4 \| Tu \|^2]^{1/p} C_{n,\rho} \omega^{-(n/p')}$$
$$= C_{n,\rho} [\omega^{-(4n/n+\rho)} \| u \|^2 + \omega^{4\rho/(n+\rho)} \| Tu \|^2]^{1/p}$$

where

$$C_{n,\rho} = \left[\sigma_n \int_0^\infty \frac{t^{n-1}}{(1+t^4)^{(n+\rho)/4}} dt\right]^{1/p'} < +\infty ,$$

where $\sigma_n = 2\pi^{n/2}[\Gamma(n/2)]^{-1}$ is $n-1$ dimensional "area" measure of the unit spherical shell in R_n, where $1/p + 1/p' = 1$ with $2 = p[2(n+\rho)/(n+\rho+4)]$ and thus $1 < p = 1 + 4/(n+\rho) \leq 2$, $p'/p = 1/(p-1) = (n+\rho)/4$, $-np/p' = -4n/(n+\rho)$, and $4 - np/p' = 4\rho/(n+\rho)$.

Now if dimension $n = 1, 2$, or 3, then $n + \rho = 4$ in Condition II and (6) yields for $u \in \mathscr{D}$

$$(7) \qquad \left(\operatorname*{ess\,sup}_{\boldsymbol{x} \in R_n} |u(\boldsymbol{x})|\right) \leq \left(\frac{1}{2\pi}\right)^{n/2} \int_{R_n} |\hat{u}(\boldsymbol{y})| \, d\mu_n(\boldsymbol{y})$$
$$\leq \left(\frac{1}{2\pi}\right)^{n/2} C_{n,\rho} \left[\omega^{-n} \| u \|^2 + \omega^{4-n} \| Tu \|^2\right]^{1/2}$$

using also (4) with convergence almost (μ_n) everywhere for a subsequence from $L_2(R_n)$ norm convergence.

Now if dimension $n \geq 4$, then in (6) define $r = 2(n+\rho)/(n+\rho+4) = 2/p$, and hence $1 < r < 2$ from $1 < p = 1 + 4/(n+\rho) < 2$. Now $1/r + 1/r' = 1$ has

$$r' = \frac{1}{1-\frac{1}{r}} = \frac{1}{1-\left(\frac{1}{2}+\frac{2}{n+\rho}\right)} = \frac{2(n+\rho)}{n-4+\rho}\,.$$

Hence the Young-Hausdorff-Titchmarsh theorem ([4], Theorem 74), p. 96), generalizing with negligible changes in proof from R_1 to R_n, using subsequences convergent almost everywhere to show that the known existent $L_2(R_n)$ and $L_{r'}(R_n)$ norm limits in (4) must agree, yields in (6) for $u \in \mathscr{D}$ if $n \geq 4$

$$(8) \qquad \left[\int_{R_n} |\, u(\boldsymbol{x})\,|^{r'} d\mu_n(\boldsymbol{x})\right]^{1/r'}$$
$$\leq \left(\frac{1}{2\pi}\right)^{n(1/2-1/r)}\left[\int_{R_n} R_n\,|\, \hat{u}(\boldsymbol{y})\,|^{r}\, d\mu_n(\boldsymbol{y})\right]^{1/r}$$
$$\leq \left(\frac{1}{2\pi}\right)^{n(1/2-1/r')}(c_{n,\rho})^{1/r}[\omega^{-4n/(n+\rho)}\,||\, u\,||^2 + \omega^{4\rho/(n+\rho)}\,||\, Tu\,||^2]^{1/2}$$

Thus we see if dimension $n = 1, 2,$ or 3 that (7) with Condition II, $n + \rho = 4$, yields for $u \in \mathscr{D}$

$$(9) \qquad ||\, Vu\,||^2 \leq \left(\frac{1}{2\pi}\right)^{n}(c_{n,\rho})^2 M_\rho[\omega^{4-n}\,||\, Tu\,||^2 + \omega^{-n}\,||\, u\,||^2 + A^2\,||\, u\,||^2$$

over all $\omega > 0$. Thus, since $\sqrt{|\, a\,|^2 + |\, b\,|^2} \leq |\, a\,| + |\, b\,|$, (5) follows with α arbitrarily small as desired for Lemma T. 2, since $4 - n \geq 1$ here.

If dimension $n \geq 4$, then we use (8), Condition II, and over the $|\,\boldsymbol{x}\,| \leq b$ portion of the integral a Schwarz-Hölder estimate with $2\tilde{r} = r' = 2(n+\rho)/(n-4+\rho) > 2$ from $1 < r < 2$, $1/\tilde{r} + 1/\tilde{r}' = 1$, and thus

$$2\tilde{r}' = \frac{2}{1-\left(\frac{n-4+\rho}{n+\rho}\right)} = \frac{2(n+\rho)}{4} = \frac{1}{2}(n+\rho)\,.$$

Hence, if $n \geq 4$, for $u \in \mathscr{D}$

$$(10) \qquad ||\, Vu\,||^2 \leq (M_\rho)^{4/(n+\rho)}\left(\frac{1}{2\pi}\right)^{n(1-(2/r'))}(c_{n,\rho})^{2/r}$$
$$\times\ [\omega^{4\rho/(n+\rho)}\,||\, Tu\,||^2 + \omega^{-4n/(n+\rho)}\,||\, u\,||^2] + A^2\,||\, u\,||^2$$

for all $\omega > 0$. Thus again (5) follows in this case $n \geq 4$ with α arbitrarily small, since $\omega^{4\rho/(n+\rho)} \to 0$ as $\omega \to 0^+$. Thus the proof of Lemma T. 2 is complete.

Returning to the proof of our Theorem T. 1, from the remarks preceding Lemma T. 2 we see this lemma permits H to be defined on $\mathscr{D}$ dense, and $H_1 \subseteq H$ from $T_1 \subseteq T$. Also T is self-adjoint with domain $\mathscr{D}$.

For by definition S^*TS is a purely multiplicative operator, $[S^*TS\hat{u}](\boldsymbol{y}) = |\boldsymbol{y}|^2\hat{u}(\boldsymbol{y})$, with the natural domain of all $\hat{u} \in L_2(R_n)$ such that $|\boldsymbol{y}|^2 \hat{u}(\boldsymbol{y})$ is in $L_2(R_n)$. It is well known and easy to see that this makes S^*TS self-adjoint, and hence so is T since S is unitary.

Next $(Tu, u) = \int_{R_n} |\boldsymbol{y}|^2 |\hat{u}(\boldsymbol{y})|^2 d\mu_n(\boldsymbol{y}) > 0$ for $\|u\| > 0$ shows that the spectrum of T is confined to $[0, +\infty]$. Hence $(T + \lambda^2 I)^{-1}$ is for real $\lambda > 0$ a bounded Hermitian operator on $L_2(R_n)$ with range $\mathscr{D}$, $(T+\lambda^2 I)\mathscr{D} = L_2(R_n)$ following from the spectral theorem for self-adjoint T. Thus (much as in Kato [2], Lemma 5)), from (5), we have for all $u \in L_2(R_n)$

$$\begin{aligned} \| V(T + \lambda^2 I)^{-1}u \| &\leq \alpha \| T(T + \lambda^2 I)^{-1}u \| + \beta \|(T + \lambda^2 I)^{-1}u \| \\ &\leq \left(\alpha + \frac{\beta}{\lambda^2}\right)\| u \| , \end{aligned} \tag{11}$$

since $\| T(T + \lambda^2 I)^{-1} \| < 1$ and $\| (T + \lambda^2 I)^{-1} \| \leq 1/\lambda^2$ are clear from the spectral representation of T. Thus choosing $\alpha < 1/2$ in (5), and then λ sufficiently positive so that $\frac{\beta}{\lambda^2} < \frac{1}{2}$, we see from (11) that the operator $\tilde{V}$ defined on $\mathscr{D}$ by $[\tilde{V}u](\boldsymbol{x}) = V(\boldsymbol{x})u(\boldsymbol{x})$ satisfies

$$\| \tilde{V}(T + \lambda^2 I)^{-1} \| \leq \left(\alpha + \frac{\beta}{\lambda^2}\right) < 1 . \tag{12}$$

Hence $I + \tilde{V}(T + \lambda^2 I)^{-1}$ is a bounded linear operator on $L_2(R_n)$ with range $L_2(R_n)$, since

$$[I + \tilde{V}(T + \lambda^2 I)^{-1}]^{-1} = I + \sum_{p=1}^{\infty} (-1)^p [\tilde{V}(T + \lambda^2 I)^{-1}]^p$$

also exists bounded. Thus, for λ large so (12) holds,

$$H + \lambda^2 I = T + \lambda^2 I + \tilde{V} = [I + \tilde{V}(T + \lambda^2 I)^{-1}](T + \lambda^2 I) \tag{13}$$

takes $\mathscr{D}$ onto $L_2(R_n)$, since $T + \lambda^2 I$ has already been seen to do so. Since $T = T^*$ has been shown and since $\tilde{V}$ is obviously symmetric, it follows that $H = T + \tilde{V}$ and $H + \lambda^2 I$ are symmetric, $H + \lambda^2 I \subseteq (H + \lambda^2 I)^*$. But $(H + \lambda^2 I)\mathscr{D} = L_2(R_n)$ in (13) thus makes $H+\lambda^2 I = (H+\lambda^2 I)^* = H^*+\lambda^2 I$, $H = H^*$, and hence H is self-adjoint (see [3], p. 35).

In order to complete the proof of Theorem T. 1, it remains only to show that the self-adjoint extension H of H_1 is the unique self-adjoint extension. Since here $H_1 \subseteq H_1^{**} \subseteq H = H^* \subseteq H_1^*$ is well-known [3], and likewise $H_1^{**} \subseteq \tilde{H} \subseteq H_1^*$ for any other self-adjoint extension $\tilde{H}$, since $H = H_1^{**}$ will make $H_1^* = (H_1^*)^{**} = (H_1^{**})^* = H^* = H = H_1^{**}$, and since $H_1^{**} = \bar{H}_1$ the closure of H_1, it suffices for this uniqueness to show $H \subseteq \bar{H}_1$.

In order to do so, we first (Lemma 1), Kato [2]) notice that orthogonality of nonzero $u_0 \in L_2(R_n)$ to $(I + T_1)u = (I + T)u$ for all $u \in \mathscr{D}_1$ would require $\hat{u}_0$ to be orthogonal to all $(1 + |\boldsymbol{y}|^2)\hat{u}(\boldsymbol{y})$; equivalently, since $S^*(\mathscr{D}_1) \subseteq \mathscr{D}_1$ and $S(\mathscr{D}_1) \subseteq \mathscr{D}_1$ makes $S^*(\mathscr{D}_1) = \mathscr{D}_1 = S(\mathscr{D}_1)$, this would require $\hat{u}_0(\boldsymbol{y})(1 + |\boldsymbol{y}|^2)\exp(-(1/4)|\boldsymbol{y}|^2)$, an element of $L_2(R_n)$, to be orthogonal to all polynomials in y_j multiplied by $\exp(-(1/4)|y|^2)$. But the density of $\mathscr{D}_1$ in $L_2(R_n)$ and a change of scale by the factor $\sqrt{2}$ shows this to be impossible. Hence $(I + T)\mathscr{D}_1$ is dense in $L_2(R_n)$.

Thus given $u \in \mathscr{D}$ and $\delta > 0$ there exists $u_1 \in \mathscr{D}_1$ such that

$$\begin{aligned}\delta > \|(I + T)u - (I + T)u_1\| &= \|(I + S^*TS)(\hat{u} - \hat{u}_1)\| \\ &= \left[\int_{R_n} (1 + |\boldsymbol{y}|^2)^2 |\hat{u}(\boldsymbol{y}) - \hat{u}_1(\boldsymbol{y})|^2 d\mu_n(\boldsymbol{y})\right]^{1/2} \\ &\geq (\max \|u - u_1\|, \|T(u - u_1)\|)\ .\end{aligned}$$

Thus by (5),

$$\begin{aligned}\|Hu - H_1u_1\| = \|H(u - u_1)\| &\leq \|T(u - u_1)\| + \|V(u - u_1)\| \\ &\leq (1 + \alpha)\|T(u - u_1)\| + \beta\|u - u_1\| < (1 + \alpha + \beta)\delta\ .\end{aligned}$$

Hence the graph of H is contained in the closure of the graph of H_1, $H \subseteq \overline{H}_1$, and H is the unique self-adjoint extension of H_1 as desired. Thus Theorem T. 1 is completely proved.

3. Connection with other conditions. We will show in this section that Condition I, which is always implied by (and for $n \geq 4$ coincides with) Condition II, implies our earlier one (Condition III, see eq. 19) for the construction of a self-adjoint extension as the inverse of a modified Green function integral operator. In fact, it is easy to verify for $V(\boldsymbol{x}) = |\boldsymbol{x}|^{-\eta}$ that Condition I and Condition III are each equivalent to $0 \leq \eta < (\min 2, n)$, so that in this sense they have the same strength. We remark that Condition I is the natural one, used in a forthcoming joint paper, for an asymptotic formula for the distribution of eigenvalues of the bottom part of the Schrödinger operator spectrum. Finally we will show, as noted before T. 1, that

$$\text{Condition II} \Rightarrow \text{Condition S} \Rightarrow \text{Condition III}\ .$$

In order to give this connection with the modified Green function, we need to introduce the fundamental singularity ${}_nK_\omega(r)$ for $-\nabla^2 + \omega^2 I$ with constant $\omega > 0$. This may be defined (see [1], p. 555) uniquely by the requirements that ${}_nK_\omega(r)$ be continuous over $r > 0$, that ${}_nK_\omega(|\boldsymbol{x}|) \in L_1(R_n)$ over $\boldsymbol{x}$, and that $[\omega^2 + |\boldsymbol{y}|^2]^{-1} = \int_{R_n} {}_nK_\omega(|\boldsymbol{x}|)e^{i(\boldsymbol{x}\cdot\boldsymbol{y})}d\mu_n(\boldsymbol{x})$ over $y \in R_n$. Such ${}_nK_\omega(r) > 0$ over $r > 0$ and $\omega > 0$. We define

$$ {}_n\tilde{K}_\omega(r) = M_n r^{-(n-2)} \exp\left(-\frac{\omega r}{4}\right) \quad \text{for} \quad n \geq 3 \,, $$

$$ {}_2\tilde{K}_\omega(r) = M_2[1 + \sqrt{\omega r}\,]^{-1}[1 + \ln(1 + (\omega r)^{-1})]e^{-\omega r}, $$

and

$$ {}_1\tilde{K}_\omega(r) = (2\omega)^{-1}e^{-\omega r} = {}_1K_\omega(r) \,, $$

with M_n the least possible real constant having ${}_nK_\omega(r) \leq {}_n\tilde{K}_\omega(r)$ over all $r > 0$ and $\omega > 0$, such positive finite M_n always existing. Finally define for $\omega > 0$,

$$ \overline{|V|}_\omega = \operatorname*{ess\,sup}_{x \in R_n} \int_{R_n} {}_n\tilde{K}_\omega(|\,\boldsymbol{x} - \boldsymbol{y}\,|)\,|\,V(\boldsymbol{y})\,|\,d\mu_n(\boldsymbol{y}) \,. \tag{14} $$

THEOREM T. 3. *Let V satisfy Condition* I. *Then* $\overline{|V|}_\omega < +\infty$ *for all* $\omega > 0$ *and*

$$ \lim_{\omega \to +\infty} \overline{|V|}_\omega = 0 \,. \tag{15} $$

Moreover for all $\omega > 0$

$$ \lim_{p \to +\infty} \overline{|V - V_p|}_\omega = 0 \,, \tag{16} $$

where $V_p(\boldsymbol{x}) = V(\boldsymbol{x})$ *if* $|V(\boldsymbol{x})| \leq p$, $V_p(\boldsymbol{x}) = p$ *if* $V(\boldsymbol{x}) > p$, *and* $V_p(\boldsymbol{x}) = -p$ *if* $V(\boldsymbol{x}) < -p$.

The proof is rather elementary, using for $n \geq 2$ the Schwarz-Hölder inequality with $r = (1/2)(n + \rho) > 1$ and $1/r + 1/r' = 1$, and hence

$$ r' = \frac{1}{1 - \dfrac{2}{n+\rho}} = \frac{n+\rho}{n-2+\rho} \,. $$

Thus Condition I yields in (14) for $n \geq 2$, the Schwarz-Hölder inequality being used on the $|\,\boldsymbol{y}\,| \leq b$ portion, and also ${}_n\tilde{K}_\omega(t/\omega) = \omega^{n-2}\,{}_n\tilde{K}_1(t)$ and $(n-2)(n+\rho)/(n-2+\rho) - n = -2\rho/(n-2+\rho) < 0$,

$$ \begin{aligned} \overline{|V|}_\omega \leq{} & (M_\rho)^{1/r}\omega^{-2\rho/(n+\rho)}\left[\sigma_n\int_0^\infty \{{}_n\tilde{K}_1(t)\}^{(n+\rho)/(n-2+\rho)}\,t^{n-1}dt\right]^{1/r'} \\ & + A\omega^{-2}\sigma_n\int_0^\infty {}_n\tilde{K}_1(t)t^{n-1}dt \,. \end{aligned} \tag{17} $$

In (17) the second integral is obviously finite, and so is the first for $n = 2$. For $n > 2$ we see in the first integral that only the portion $0<t<1$ is in doubt, and here we have to consider the integrand factor t raised to the exponent

$$ -(n-2)\frac{n+\rho}{n-2+\rho} + n - 1 = \frac{2\rho}{n-2+\rho} - 1 > -1 \,. $$

Thus the first integral in (17) is also finite for $n > 2$ as well as for $n=2$, and (17) shows $\overline{|V|}_\omega < +\infty$ for all $\omega > 0$ and also that (15) follows for $n \geq 2$.

Finally for (16), taking $p > A$ so that $V(\boldsymbol{x}) - V_p(\boldsymbol{x}) = 0$ almost (μ_n) everywhere over $|\boldsymbol{x}| \geq b$ by Condition I, we see that in place of (17) we have, with $c_n < +\infty$ by the finiteness of the first integral in (17), for $n \geq 2$

$$(18) \qquad \overline{|V - V_p|}_\omega \leq c_n \omega^{-2\rho/(n+\rho)} \left[\int_{\{x \mid |x| \leq b\}} |V(\boldsymbol{x}) - V_p(\boldsymbol{x})|^{(1/2)(n+\rho)} d\mu_n(\boldsymbol{x}) \right]^{1/x} .$$

Since $\lim_{p\to\infty} |V(\boldsymbol{x}) - V_p(\boldsymbol{x})| = 0$ for all $\boldsymbol{x} \in R_n$, and since $|V(\boldsymbol{x}) - V_p(\boldsymbol{x})| \leq |V(\boldsymbol{x})|$, we see Condition I and dominated convergence in (18) yields (16) as desired for $n \leq 2$.

Finally consider $n = 1$, ${}_1\tilde{K}_\omega(r) = {}_1K_\omega(r) = (2\omega)^{-1}e^{-\omega r}$. Notice that Condition I with $1 + \rho \geq 2$ clearly implies itself with ρ replaced by $\rho' = 1$. Thus in place of (17) and (18) we have for $n = 1$

$$(17)' \qquad \overline{|V|}_\omega \leq M_1(2\omega)^{-1} + A\omega^{-2} ,$$

$$(18)' \qquad \overline{|V - V_p|}_\omega \leq (2\omega)^{-1} \int_{\{x \mid |x| \leq b\}} |V(\boldsymbol{x}) - V_p(\boldsymbol{x})| \, d\mu_n(\boldsymbol{x}) ,$$

which clearly yield (15) and (16) in the same way as above. Thus the proof of Theorem T.3 is complete.

Now consider the following condition on V. As stated in Corollary T.4 immediately thereafter, this condition is implied by Condition I, as we see from (15) above.

CONDITION III. *There exists some $\omega, 0 < \omega < +\infty$, such that*

$$(19) \qquad \overline{|V|}_\omega < 1 .$$

COROLLARY T.4. *If Condition* I *is satisfied, then so is Condition* III.

Condition III is our earier condition in [1] mentioned above. For our modified Green function, consider the formulae

$$(20) \qquad G_\omega(\boldsymbol{x}, \boldsymbol{y}) = {}_nK_\omega(|\boldsymbol{x} - \boldsymbol{y}|) + \sum_{p=1}^{\infty} (-1)^p \int_{R_n} \int_{R_n} \cdots \int_{R_n} {}_nK_\omega(|\boldsymbol{x} - {}_1\boldsymbol{z}|) V({}_1\boldsymbol{z}) {}_nK_\omega(|{}_1\boldsymbol{z} - {}_2\boldsymbol{z}|) V({}_2\boldsymbol{z}) \cdots V({}_p\boldsymbol{z}) {}_nK_\omega(|{}_p\boldsymbol{z} - \boldsymbol{y}|) d\mu_n({}_1\boldsymbol{z}) \cdots d\mu_n({}_p\boldsymbol{z}) ,$$

$$(21) \qquad [G_\omega u](\boldsymbol{x}) = \int_{R_n} G_\omega(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{y}) d\mu_n(\boldsymbol{y}) .$$

By virtue of our earlier work ([1], p. 560, 567, Lemma 3.4, Theorem 3.5, Theorem 4.5), we have the following theorem, using $|\overline{V}|_\omega \leq |\overline{V}|_{\omega'}$ from ${}_n\tilde{K}_\omega(r) \leq {}_n\tilde{K}_{\omega'}(r)$ in (14) for $\omega \geq \omega'$.

THEOREM T.5. *Let the conditions of Theorem* T.3 *hold and let* ω_1, $0 < \omega_1 < +\infty$, *be chosen so that* (19) *holds. Then for* $\omega \geq \omega_1$ *the right side of* (20) *converges almost* $(\mu_n \times \mu_n)$ *everywhere as a definition of* $G_\omega(\boldsymbol{x}, \boldsymbol{y})$, $G_\omega(\boldsymbol{x}, \boldsymbol{y}) = G_\omega(\boldsymbol{y}, \boldsymbol{x})$ *almost* $(\mu_n \times \mu_n)$ *everywhere, in* (21) *the right side exists finite almost* (μ_n) *everywhere and is in* $L_2(R_n)$ *for* $u \in L_2(R_n)$, *and the operator* G_ω *on* $L_2(R_n)$ *so defined is bounded Hermitian* $\| G_\omega \| \leq \omega^{-2}(1 - |\overline{V}|_\omega)^{-1}$. *Moreover the operator* H_2 *defined by*

$$H_2 = G_\omega^{-1} - \omega^2 I \tag{22}$$

exists as a self-adjoint operator in $L_2(R_n)$ *independent of* $\omega \geq \omega_1$.

Now under Condition I here, which is less than Condition II if $n \leq 3$, the linear manifold $\mathscr{N} = \{u \in L_2(R_n) \mid Vu \in L_2(R_n)\}$ need no longer contain $\mathscr{D}_1$, and hence H_1 may not exist as an operator in $L_2(R_n)$. Thus define $\tilde{\mathscr{D}}_1 = \mathscr{N} \cap \mathscr{D}_1$, and as in (2)

$$[\tilde{H}_1 u](\boldsymbol{x}) = -\nabla^2 u(\boldsymbol{x}) + V(\boldsymbol{x})u(\boldsymbol{x}) \tag{23}$$

for $u \in \tilde{\mathscr{D}}_1$; thus $\tilde{H}_1$ satisfies $(\tilde{H}_1 u, w) = (u, \tilde{H}_1 w)$ for $u, w \in \tilde{\mathscr{D}}_1$. Note $\tilde{\mathscr{D}}_1 = \mathscr{D}_1$ and $\tilde{H}_1 = H_1$ if $n \geq 4$, Condition I and II coinciding. Hence, after proving the following theorem, $H_2 = H$ follows for $n \geq 4$.

THEOREM T.6. *Let* V *satisfy Condition* I. *Then the self-adjoint operator* H_2 *defined by* (22), *known existent by Corollary* T.4 *and Theorem* T.5, *is an extension of* $\tilde{H}_1$, $\tilde{H}_1 \subseteq H_2$

We note here that $\tilde{\mathscr{D}}_1$ need not be dense in $L_2(R_n)$ if $n \leq 3$, although $\tilde{H}_1$ will not be a very respectable operator from the Hilbert space viewpoint if $\tilde{\mathscr{D}}_1$ is not dense, in particular not being symmetric. This theorem is the same as our earlier one ([1]), Theorem 5.3, p. 572) except for change in the initial domain from $\mathscr{G}_0 = \mathscr{N} \cap \mathscr{G}$ there to $\tilde{\mathscr{D}}_1 = \mathscr{N} \cap \mathscr{D}_1$ here. Merely sketching the proof, we first see

$$(u, \varphi) = \int_{R_n} [G_\omega u](\boldsymbol{x})\{(\omega^2 + V(\boldsymbol{x}))\overline{\varphi(\boldsymbol{x})} - \nabla^2\overline{\varphi(\boldsymbol{x})}\}\, d\mu_n(\boldsymbol{x}) \tag{24}$$

follows for $\varphi \in \mathscr{D}_1$ and $u \in L_1(R_n) \cap L_2(R_n)$, the proof being unchanged from the earlier one ([1], Theorem 5.1, p. 568) for φ having continuous second partials and vanishing outside a bounded set. Taking $\varphi \in \tilde{\mathscr{D}}_1 = \mathscr{N} \cap \mathscr{D}_1$ in (24) and using the facts that G_ω is bounded Hermitian and that $L_1(R_n) \cap L_2(R_n)$ is dense in $L_2(R_n)$, we obtain from (23)

$$G_\omega(\omega^2 I + \tilde{H}_1)\varphi = \varphi \tag{25}$$

for $\varphi \in \tilde{\mathscr{D}}_1$. Thus $\tilde{\mathscr{D}}_1 \subseteq$ (range of G_ω) = (domain of G_ω^{-1}), and $\omega^2 I + \tilde{H}_1 \subseteq G_\omega^{-1}$, $\tilde{H}_1 \subseteq G_\omega^{-1} - \omega^2 I = H_2$ as desired, proving T.6.

THEOREM T.7. *Let* V *satisfy Condition* I *and define* $h_1 = \lim_{r\to\infty}(\operatorname{ess\,inf}_{|\boldsymbol{x}|\geq r} V(\boldsymbol{x}))$. *Then this limit exists satisfying* $-A \leq h_1 \leq A$ *and the spectrum* Σ *of the self-adjoint operator* H_2 *defined by* (22), *known existent by Corollary* T.4 *and Theorem* T.5, *has* $(-\infty, h_1) \cap \Sigma$ *to consist of pure point spectra with* $(-\infty, h) \cap \Sigma$ *finite and having a finite dimensional eigenspace for all* $h < h_1$. *If also* $h_0 = [\operatorname{ess\,inf}_{\boldsymbol{x}\in R_n} V(\boldsymbol{x})] > -\infty$, *then* $(-\infty, h_0) \cap \Sigma$ *is empty.*

Since (19) and (16) follow from Condition I for large ω by Theorem T.3, this theorem follows from our earlier one ([1], Theorem 6.4, p. 579).

Finally we finish this section by proving in the following Theorems T.8 and T.9 the implications asserted before, namely II $\Rightarrow$ S $\Rightarrow$ III. Since Condition S, as noted before Theorem T.1, implies the conclusion of that theorem, from II $\Rightarrow$ S we have an alternate proof of Theorem T.1. For knowledge of this work of Stummel [5] we are indebted to the referee. Although Theorems T.8 and T.9 seem of sufficient interest to record, their proofs are simple exercises in the use of the Schwarz-Hölder inequality.

We start by stating Stummel's Condition S.

CONDITION S.

$$\left\{\sup_{\boldsymbol{x}\in R_n}\int_{\{\boldsymbol{y}\,|\,|\boldsymbol{x}-\boldsymbol{y}|\leq 1\}} |V(\boldsymbol{y})|^2 \,|\boldsymbol{x}-\boldsymbol{y}|^{-\gamma} d\mu_n(\boldsymbol{y})\right\} < +\infty \tag{26}$$

for some real γ *satisfying* $\gamma > n - 4$ *and* $\gamma \geq 0$.

THEOREM T.8. *If* V *satisfies Condition* II, *then it also satisfies Condition* S.

THEOREM T.9. *If* V *satisfies Condition* S, *then equation* (15) *and hence Condition* III *are satisfied by* V.

To prove T.8 first, Condition II clearly yields (26) with $\gamma = 0$, which thus takes care of the trivial case $1 \leq n < 4$.

Now consider dimension $n \geq 4$. Then for the $\rho > 0$ in (1) of the given Condition II, we may choose real γ to satisfy

$$n - 4\left(\frac{n}{n+\rho}\right) > \gamma > n - 4 \tag{27}$$

and must then verify (26). Take $p = (1/4)(n+\rho) > 1$ and then $1/p + 1/p' = 1$, for which

$$n[(n+\rho)-4] > (n+\rho)\gamma\ , \qquad n+\rho > 4+(n+\rho)\frac{\gamma}{n}\ ,$$

$$\left(1-\frac{\gamma}{n}\right)(n+\rho) > 4\ , \quad p=(1/4)(n+\rho) > \left(1-\frac{\gamma}{n}\right)^{-1},$$

$$1-\frac{1}{p'}=\frac{1}{p}<1-\frac{\gamma}{n}\ ,$$

and hence $\gamma p' < n$. Thus for (26) we have the Schwarz-Hölder estimate

$$\begin{aligned}(28)\qquad &\int_{\{\boldsymbol{y}\,||\boldsymbol{x}-\boldsymbol{y}|\leqq 1\}} |V(\boldsymbol{y})|^2\,|\boldsymbol{x}-\boldsymbol{y}|^{-\gamma} d\mu_n(\boldsymbol{y}) \\ &\quad\leq \left[\int_{\{\boldsymbol{y}\,||\boldsymbol{x}-\boldsymbol{y}|\leqq 1\}} |V(\boldsymbol{y})|^{2p} d\mu_n(\boldsymbol{y})\right]^{1/p} \left[\sigma_n \int_0^1 r^{-\ p'} r^{n-1} dr\right]^{1/p'}\end{aligned}$$

with $2p=(1/2)(n+\rho)$, $\gamma p' < n$, and σ_n as in (6). Thus the second factor on the right of (28) is a finite constant, Condition II assures that the first factor is bounded over $\boldsymbol{x} \in R_n$, and (27) and (28) yield (26) for Condition S. This completes the proof of T.8.

Now for Theorem T.9 it suffices to prove that Condition S implies $\lim_{\omega\to+\infty}|\overline{V}|_\omega = 0$, since equation (15) yields the conclusion of Corollary T.4 as noted there. Considering first the general case $n \geq 4$, and taking $\beta = n-2-\gamma/2 < n-2-(n-4)(1/2) = n/2$ so that $2\beta < n$ from $\gamma > n-4 \geq 0$, the Schwarz inequality yields

$$\begin{aligned}(29)\quad &\int_{\{\boldsymbol{y}\,||\boldsymbol{x}-\boldsymbol{y}|\leqq 1\}} |V(\boldsymbol{y})| \frac{e^{-\omega|\boldsymbol{x}-\boldsymbol{y}|}}{|\boldsymbol{x}-\boldsymbol{y}|^{n-2}} d\mu_n(\boldsymbol{y}) \\ &\quad\leq \left[\int_{\{\boldsymbol{y}\,||\boldsymbol{x}-\boldsymbol{y}|\leq 1\}} |V(\boldsymbol{y})|^2\,|\boldsymbol{x}-\boldsymbol{y}|^{-\gamma} d\mu_n(\boldsymbol{y})\right]^{1/2} \left[\sigma_n \int_0^1 e^{-\omega r} r^{n-1-2\beta} dr\right]^{1/2}.\end{aligned}$$

On the right here the second factor is $\leq [\omega^{-(n-2\beta)}\sigma_n\Gamma(n-2\beta)]^{1/2} \to 0$ as $\omega \to +\infty$ since $n-2\beta > 0$; the first factor is independent of ω and bounded over $\boldsymbol{x} \in R_n$ according to Condition S. Hence we see that the left side of (29) converges to zero uniformly over $\boldsymbol{x} \in R_n$ as $\omega \to +\infty$ for $n \geq 4$.

In order to estimate $|\overline{V}|_\omega$, we must also consider the left side of (29) with the range of integration replaced by its complement in R_n. For this we define

$B(\boldsymbol{j}) = \{\boldsymbol{x} \in R_n \,||\, x_i - 2j_i(n)^{-1/2}| \leq (n)^{-1/2}$ for $1 \leq i \leq n\}$, $\boldsymbol{j} = (j_1, j_2, \cdots, j_n)$ for integer j_i, and also $r(\boldsymbol{j}) = \inf_{\boldsymbol{x}\in B(\boldsymbol{j})} |\boldsymbol{x}|$. Noting that $B(0) \subseteq \{\boldsymbol{x} \,||\, \boldsymbol{x}| \leq 1\}$ makes $\{\boldsymbol{x}\,||\,\boldsymbol{x}| > 1\} \subseteq \bigcup_{\boldsymbol{j}\neq 0} B(\boldsymbol{j})$, we see with n ≥ 4

$$(30)\qquad \int_{\{y\,|\,|x-y|>1\}} |V(y)| \frac{e^{-\omega|x-y|}}{|x-y|^{n-2}} d\mu_n(y)$$
$$\leq \int_{\{z\,|\,|z|>1\}} e^{-\omega|z|} |V(x-z)|\, d\mu_n(z)$$
$$\leq \sum_{j\neq 0} e^{-\omega r(j)} \int_{B(j)} |V(x-z)|\, d\mu_n(z)$$
$$\leq \left(\frac{2}{\sqrt{n}}\right)^{n/2}\left[\sup_{x\in R_n} \int_{\{y\,|\,|x-y|\leq 1\}} |V(y)|^2\, d\mu_n(y)\right]^{1/2} \left\{\sum_{j\neq 0} e^{-\omega r(j)}\right\}.$$

Since $|x-y|^{-\gamma} \geq 1$ in (26), we see that Condition S assures that the first factor on the far right side of (30) is a finite constant. Moreover, we see that the second factor

$$\left\{\sum_{j\neq 0} e^{-\omega r(j)}\right\} \to 0 \quad \text{as} \quad \omega \to +\infty,$$

using

$$r(j) \geq \left(\sup_{x\in B(j)} |x|\right) - 2$$

to estimate the portion of this sum where

$$r(j) \geq 3 \quad \text{by} \quad \left(\frac{2}{\sqrt{n}}\right)^{-n} \sigma_n \int_3^\infty e^{-\omega(r-2)} r^{n-1} dr$$

which $\to 0$ by dominated convergence, and using $r(j) \geq 1/\sqrt{n} > 0$ for $j \neq 0$ to estimate the remaining finite sum portion. Thus the left side of (30) converges to zero uniformly over $x \in R_n$ as $\omega \to +\infty$, which when combined with the same conclusion about (29) proved above yields $\overline{|V|}_\omega \to 0$ and completes the proof of T.9 for dimension $n \geq 4$.

For dimension $n < 4$, we see Condition S becomes just (26) with $\gamma = 0$. Hence $\int_{R_n} [{}_n\tilde{K}_\omega(|x|)]^2 d\mu_n(x) = c_n \omega^{-(4-n)}$, easily seen with $c_n < +\infty$ for $n < 4$ from the definition preceding (14), gives in place of (29)

$$(31)\qquad \left[\sup_{x\in R_n} \int_{\{y\,|\,|x-y|\leq 1\}} |V(y)|\, {}_n\tilde{K}_\omega(|x-y|) d\mu_n(y)\right] = 0\left(\omega^{-(4-n)1/2}\right)$$

as $\omega \to +\infty$. Also (30) still shows the integral over the complimentary region to converge to zero uniformly over $x \in R_n$ as $\omega \to +\infty$ if $n = 3$, and a very similar computation gives the same result if $n = 1$ or 2. Hence $\lim_{\omega\to+\infty} \overline{|V|}_\omega = 0$ follows from Condition S when dimension $n < 4$ as well as when $n \geq 4$, and the proof of T.9 is complete.

4. Variational characterization of the spectrum. In this section we will show (see T.13 following) that a variational characterization of the spectrum, well-known at least for continuous V and bounded domains, also holds for H_2 with V subject only to Condition I. This is rather easy to obtain ([2], p. 209, eq. (23)) under Condition II, and the major effort in our argument amounts to showing that Condition I, which is weaker for $1 \leq n \leq 3$, actually suffices.

We start with the following theorem, where by the L_1 sense of the Fourier transform $\hat{u}$ for $u \in L_1(R_n)$ we mean (3) with no limit and $\int_{\{x \mid |x| \leq N\}}$ replaced by the oridinary Lebesque integral $\int_{R_n}$. Notice that if $u \in L_1(R_n) \cap L_2(R_n)$, then by taking subsequences we may be sure that the two definitions of $\hat{u}(\boldsymbol{y})$ are equal almost (μ_n) everywere. Hereafter $\| u \|_r$ denotes the $L_r(R_n)$ norm of u, and $\| u \|$ or $\| u \|_2$ the $L_2(R_n)$ norm.

THEOREM T.10. *Let V satisfy Condition* I *and let λ_0 be in the point spectrum of H_2, defined by* (22), *with eigenvector $u_0 \in \mathscr{D}_2 = G_\omega(L_2(R_n))$, $H_2 u_0 = \lambda_0 u_0$ and $\| u_0 \| = 1$. Then $Vu_0 \in L_1(\{\boldsymbol{x} \mid |\boldsymbol{x}| \leq b\})$ and over $\boldsymbol{y} \in R_n$*

$$|\boldsymbol{y}|^2 \hat{u}_0(\boldsymbol{y}) + \psi_0(\boldsymbol{y}) = \lambda_0 \hat{u}_0(\boldsymbol{y}) \tag{32}$$

where $\psi_0 = \hat{f}_0 + \hat{g}_0$, $\hat{f}_0$ is the L_1 sense transform of $f_0(\boldsymbol{x}) = V(\boldsymbol{x}) u_0(\boldsymbol{x}) \chi_b(\boldsymbol{x})$ with $\chi_b(\boldsymbol{x})$ the characteristic function of $\{\boldsymbol{x} \in R_n \mid |\boldsymbol{x}| \leq b\}$, and $\hat{g}_0$ is the usual L_2 transform of $g_0 = Vu_0 - f_0$.

If $(n + \rho) \geq 4$, then Condition II follows from Condition I, $H_2 = H$ and $u_0 \in \mathscr{D}_2 = \mathscr{D}$ by Theorems T.1 and T.6, $Vu_0 \in L_2(R_n)$ by Lemma T.2 and hence $\varepsilon L_1(\{\boldsymbol{x} \mid |\boldsymbol{x}| \leq b\})$, ψ_0 exists as defined and $= \widehat{Vu_0}$ defined in the usual L_2 sense, and (32) follows from $Hu_0 = \lambda_0 u_0$ and the definition of H.

The proof of T. 10 thus being complete for $(n + \rho) \geq 4$ and hence for $n \geq 4$, we now consider the remaining case $2 \leq n + \rho < 4$, for which $1 \leq n \leq 3$. Since $G_\omega u_0 = (\lambda_0 + \omega^2)^{-1} u_0$ with $\lambda_0 + \omega^2 > 0$ for $\omega \geq \omega_1$ follows from (22) and $H_2 u_0 = \lambda_0 u_0$, we see ([1], (3.5), (3.6), and (3.21), p. 558 and 562) by using the Schwarz inequality that u_0 is essentially bounded, $u_0 \in L_\infty(R_n)$ and $\|u_0\|_\infty = \operatorname{ess\,sup}_{x \in R_n} |u_0(x)| < +\infty$. Thus by Condition I, $Vu_0 \in L_r(\{\boldsymbol{x} \mid |\boldsymbol{x}| \leq b\}) \subseteq L_1(\{\boldsymbol{x} \mid |\boldsymbol{x}| \leq b\})$ with $r = \frac{1}{2}(n + \rho)$ satisfying $1 \leq r < 2$, and ψ_0 exists as defined.

Now, $L_1 \cap L_2$ being dense in L_2, there exists a sequence $u'_k \in L_1(R_n) \cap L_2(R_n)$ such that the L_2 norm $\|u_0 - u'_k\|_2 \to 0$. Hence as above, $u_k = (\lambda_0 + \omega^2) G_\omega u'_k$ has $u_k \in L_\infty \cap L_2$ and both $\|u_0 - u_k\|_2 \to 0$ and also $\|u_0 - u_k\|_\infty \to 0$. Actually ([1], Lemma 4.1, p. 565), u_k and $Vu_k \in L_1(R_n)$ also, and

$$(|\boldsymbol{y}|^2 - \lambda_0) \hat{u}_k(\boldsymbol{y}) + \psi_k(\boldsymbol{y}) = (\lambda_0 + \omega^2) \{\hat{u}'_k(\boldsymbol{y}) - \hat{u}_k(\boldsymbol{y})\} \tag{33}$$

with $\psi_k = \widehat{Vu_k}$ in the L_1 sense. Defining f_k and g_k from u_k analogously to f_0 and g_0 from u_0, $\psi_k = \hat{f}_k + \hat{g}_k$ defined in the L_1 sense. Moreover,

$$\|\hat{f}_0 - \hat{f}_k\|_\infty \leq (2\pi)^{-n/2} \|f_0 - f_k\|_1 \leq (2\pi)^{-n/2} \|V\|_{1,b} \|u_0 - u_k\|_\infty \to 0$$

with

$$||V||_{1,b} = \int_{\{x \mid |x| \le b\}} |V(x)|\, d\mu_n(x) ,$$

and $||g_0 - g_k||_2 \le A||u_0 - u_k||_2 \to 0$ by using Condition I. Thus, after taking subsequences, we may assume almost (μ_n) everywhere that

$$\psi_k(y) = \hat{f}_k(y) + \hat{g}_k(y) \to \hat{f}_0(y) + \hat{g}_0(y) = \psi_0(y) , \quad \hat{u}_k(y) \to \hat{u}_0(y) ,$$

and $\hat{u}'_k(y) \to \hat{u}_0(y)$, since $||\hat{u}_k - \hat{u}_0||_2 = ||u_k - u_0||_2 \to 0$ and $||\hat{u}'_k - \hat{u}_0||_2 = ||u'_k - u_0||_2 \to 0$. Thus (33) yields (32), and the proof of theorem T. 10 is complete.

We next give some approximation lemmas.

LEMMA T. 11. *Let V satisfy Condition* I *with $n + \rho \ge 4$, and hence Condition* II *also; let $u_0 \in \mathscr{D}$. Then there exists a sequence of $u_k \in \mathscr{D}$, satisfying simultaneously $||u_0 - u_k|| \to 0$, $||T(u_0 - u_k)|| \to 0$, $||V(u_0 - u_k)|| \to 0$ for these $L_2(R_n)$ norms.*

This was proved in the last two paragraphs of § 2. In the following we denote $(z \cdot \xi) = \sum_{j=1}^{n} z_j \bar{\xi}_j$, $|z| = \sqrt{(z \cdot z)}$ for z and $\xi \in C_n$, unitary n space. $\mathscr{D}_2 = G_\omega(L_2(R_n))$ for $\omega \ge \omega_1$ is the domain of H_2 as usual.

LEMMA T. 12. *Let V satisfy Condition* I *with $2 \le n + \rho < 4$ and let $u_0 \in \mathscr{D}_2$ satisfy $H_2 u_0 = \lambda_0 u_0$ and $||u_0|| = 1$. Then $|y|\hat{u}_0(y) \in L_2(R_n)$ and $u_0 \in L_\infty(R_n)$ and $\hat{u}_0 \in L_1(R_n)$, and there exists a sequence of $u_k \in \mathscr{D}_1$ such that simultaneously $||u_0 - u_k||_2 \to 0$, $||u_0 - u_k||_\infty \to 0$,*

$$\int_{R_n} |V(x)|\, |u_0(x) - u_k(x)|^2 d\mu_n(x) \to 0 ,$$

and

$$\int_{R_n} |\nabla_{\text{gen}} u_0(x) - \nabla u_k(x)|^2 d\mu_n(x) \to 0 ,$$

where ∇ denotes the ordinary gradient differential operator and $\nabla_{\text{gen}} u$ the C_n vector valued function whose components are in $L_2(R_n)$ and have the components of $iy\hat{u}(y)$ as their L_2 sense Fourier transforms.

To prove T. 12, first notice $2 \le n + \rho < 4$ makes $1 \le n \le 3$, and hence, as shown in proving T. 10, $u_0 \in L_\infty(R_n)$ and $f_0 \in L_r(R_n)$ with $r = \frac{1}{2}(n + \rho)$, $1 \le r < 2$. Thus, using the Young-Hausdorff-Titchmarsh theorem as in (8), the L_1 sense $\hat{f}_0 \in L_{r'}(R_n)$ with

$$r' = \frac{1}{1 - 1/r} = \frac{n + \rho}{n + \rho - 2} > 2 , \text{ and } r' = \infty$$

if $n + \rho = 2$.

Next notice that for $0 < \nu < 2$ we have from (32)

$$(34)\qquad \left(\frac{|\boldsymbol{y}|}{1+|\boldsymbol{y}|}\right)^{2-\nu}|\boldsymbol{y}|^{\nu}\hat{u}_0(\boldsymbol{y}) = (1+|\boldsymbol{y}|)^{-2+\nu}[\lambda_0\hat{u}_0(\boldsymbol{y}) - \hat{g}_0(\boldsymbol{y})] - (1+|\boldsymbol{y}|)^{-2+\nu}\hat{f}_0(\boldsymbol{y})\,.$$

Thus we may conclude $|\boldsymbol{y}|^{\nu}\hat{u}_0(\boldsymbol{y}) \in L_2(R_n)$, as desired, whenever this holds for both terms on the right of (34). The first term is obviously in $L_2(R_n)$. For the second term we use $\hat{f}_0 \in L_{r'}(R_n)$ and the Schwarz-Hölder inequality with

$$2\alpha = r' = \frac{n+\rho}{n+\rho-2} > 2\,, \quad \alpha' = \frac{1}{1-1/\alpha} = \frac{n+\rho}{4-(n+\rho)}$$

holding even for $n+\rho=2$, for which $\alpha = \infty$ and $\alpha' = 1$. Thus, with σ_n as in (6),

$$(35)\quad \int_{R_n}\frac{|\hat{f}_0(\boldsymbol{y})|^2}{(1+|\boldsymbol{y}|)^{2(2-\nu)}}d\mu_u(\boldsymbol{y}) \le \|\hat{f}_0\|_{r'}^2\left[\sigma_n\int_0^\infty\frac{t^{n-1}}{(1+t)^{2(2-\nu)\alpha'}}dt\right]^{1/\alpha'} < +\infty$$

provided that

$$n < (2-\nu)2\alpha' = \frac{2(n+\rho)(2-\nu)}{4-(n+\rho)}\,.$$

This last inequality is equivalent to

$$2-\nu > \frac{[4-(n+\rho)]n}{2(n+\rho)}\,,$$

and this to

$$\nu < \frac{4\rho+n(n+\rho)}{2(n+\rho)} = \frac{n}{2} + \frac{2\rho}{n+\rho}\,.$$

We see for our $n = 1, 2,$ or 3, $\rho > 0$, $2 \le n+\rho < 4$, that this last inequality is always satisfied for $\nu = 1$ and for $\nu = \nu_1 = n/2 + \rho/(n+\rho)$. Note $n/2 < \nu_1 < 2$. Thus we have shown $|\boldsymbol{y}|\hat{u}_0(\boldsymbol{y})$ and $|\boldsymbol{y}|^{\nu_1}\hat{u}(\boldsymbol{y})$ to be $\in L_2(R_n)$.

Next for any finite set of $\nu_p > 0$, define

$$[L\hat{u}](\boldsymbol{y}) = \left(1 + \sum_p |\boldsymbol{y}|^{\nu_p}\right)\hat{u}(\boldsymbol{y})\,.$$

As in the last two paragraphs of § 2, $L\mathscr{D}_1$ is dense in $L_2(R_n)$, since any $\hat{u} \in L_2(R_n)$ has

$$\left(1 + \sum_p |\boldsymbol{y}|^{\nu_p}\right)\exp\left(-\tfrac{1}{4}|\boldsymbol{y}|^2\right)\hat{u}(\boldsymbol{y}) \in L_2(R_n)$$

and therefore is not orthogonal to all $Q(\boldsymbol{y})\exp(-\frac{1}{2}|\boldsymbol{y}|^2)$ with polynomial Q, and thus $\hat{u}$ cannot be orthogonal to $L\mathscr{D}_1$. Hence, for any $u \in L_2(R_n)$ such that $L\hat{u} \in L_2(R_n)$ there exists (since $\mathscr{D}_1$ transforms onto $\mathscr{D}_1$) a sequence $u_k \in \mathscr{D}_1$ such that $\|L(\hat{u} - \hat{u}_k)\|_2 \to 0$, and thus simultaneously

$$\int_{R_n} |\boldsymbol{y}|^{2\nu_p} |\hat{u}(\boldsymbol{y}) - \hat{u}_k(\boldsymbol{y})|^2 d\mu_n(\boldsymbol{y}) \to 0$$

for the finite set of ν_p as well as $\|u - u_k\|_2 = \|\hat{u} - \hat{u}_k\|_2 \to 0$. Applying this result to $u_0 \in L_2(R_n)$ with the finite set $\{1, \nu_1\}$ of ν's, since $|\boldsymbol{y}|\hat{u}_0(\boldsymbol{y})$ and $|\boldsymbol{y}|^{\nu_1}\hat{u}_0(\boldsymbol{y})$ were shown to be in $L_2(R_n)$, there thus exists a sequence of $u_k \in \mathscr{D}_1$ such that simultaneously

$$\|u_0 - u_k\|_2 = \|\hat{u} - \hat{u}_k\|_2 \to 0 \, , \quad \int_{R_n} |\boldsymbol{y}|^2 |\hat{u}_0(\boldsymbol{y}) - \hat{u}_k(\boldsymbol{y})|^2 d\mu_n(\boldsymbol{y}) \to 0 \, ,$$

and

$$\int_{R_n} |\boldsymbol{y}|^{2\nu_1} |\hat{u}_0(\boldsymbol{y}) - \hat{u}_k(\boldsymbol{y})|^2 d\mu_n(\boldsymbol{y}) \to 0$$

with

$$\nu_1 = \frac{n}{2} + \frac{\rho}{n+\rho} > \frac{n}{2} \, .$$

From the second limit statement just proved, and from $|\boldsymbol{y}|\hat{u}_0(\boldsymbol{y}) \in L_2(R_n)$, we see that $\boldsymbol{\nabla}_{\text{gen}} u_0$ exists as defined and that, since $i\boldsymbol{y}\hat{u}_k(\boldsymbol{y})$ clearly has its components the L_2 transforms of the C_n vector valued function $\boldsymbol{\nabla} u_k(\boldsymbol{x})$,

$$\int_{R_n} |\boldsymbol{\nabla}_{\text{gen}} u_0(\boldsymbol{x}) - \boldsymbol{\nabla} u_k(\boldsymbol{x})|^2 d\mu_n(\boldsymbol{x}) = \int_{R_n} |\boldsymbol{y}|^2 |\hat{u}_0(\boldsymbol{y}) - \hat{u}_k(\boldsymbol{y})|^2 d\mu_n(\boldsymbol{y}) \to 0 \, .$$

Next for $u \in L_2(R_n)$ having $|\boldsymbol{y}|^{\nu_1}\hat{u}(\boldsymbol{y}) \in L_2(R_n)$,

$$\begin{aligned} (36) \qquad \|\hat{u}\|_1 &\le M\Big[\int_{R_n} (1 + |\boldsymbol{y}|^{2\nu_1}) |\hat{u}(\boldsymbol{y})|^2 d\mu_n(\boldsymbol{y})\Big]^{1/2} \\ &\le M\|\hat{u}\|_2 + M\Big[\int_{R_n} |\boldsymbol{y}|^{2\nu_1} |\hat{u}(\boldsymbol{y})|^2 d\mu_n(\boldsymbol{y})\Big]^{1/2} , \\ M &= \Big[\sigma_n \int_0^\infty \frac{t^{n-1}}{1 + t^{2\nu_1}} dt\Big]^{1/2} < +\infty \, , \end{aligned}$$

using the Schwarz inequality and $2\nu_1 > n$. Thus from $|\boldsymbol{y}|^{\nu_1}\hat{u}_0(\boldsymbol{y}) \in L_2(R_n)$ we conclude $\|\hat{u}_0\|_1 < +\infty$ and $\hat{u}_0 \in L_1(R_n)$, and likewise $\|\hat{u}_0 - \hat{u}_k\|_1 \to 0$ follows from

$$\int_{R_n} |\boldsymbol{y}|^{2\nu_1} |\hat{u}_0(\boldsymbol{y}) - \hat{u}_k(\boldsymbol{y})|^2 d\mu_n(\boldsymbol{y}) \to 0$$

shown above. Thus we have

$$||u_0 - u_k||_\infty \leq (2\pi)^{-n/2}||\hat{u}_0 - \hat{u}_k||_1 \to 0$$

from the L_1 sense of (4) agreeing here with the L_2 sense as usual. Hence finally

$$\text{(37)} \qquad \int_{R_n} |V(\boldsymbol{x})|\,|u_0(\boldsymbol{x}) - u_k(\boldsymbol{x})|^2 d\mu_n(\boldsymbol{x}) \\ \leq ||u_0 - u_k||_\infty^2 \int_{\{\boldsymbol{x}|\ |\boldsymbol{x}|\leq b\}} |V(\boldsymbol{x})| d\mu_n(\boldsymbol{x}) + A||u_0 - u_k||_2^2$$

by Condition I with the right side $\to 0$ as $k \to +\infty$. Thus the proof of Lemma T. 12 is complete.

We now are ready to give our variational characterization of the spectrum Σ of H_2, assuming only Condition I. Define

$$h_1 = \lim_{r\to\infty}\left(\operatorname*{ess\,inf}_{|\boldsymbol{x}|\geq r} V(\boldsymbol{x})\right),$$

and by Theorem T. 7 we know that $\Sigma \cap (-\infty, h)$ for $h < h_1$ consists of a finite set of λ which are each in the point spectrum of H_2 with finite multiplicity. Thus there is uniquely defined a finite or countable set $\{\lambda_p\} = \Sigma \cap (-\infty, h_1)$, $\lambda_p \leq \lambda_{p+1}$, and the $\lambda_p = \lambda$ repeat according to the multiplicity of each λ in the point spectrum of H_2. In the statement following, $u \perp S$ means $(u, w) = 0$ for all $w \in S$.

THEOREM T. 13. *Let V satisfy Condition* I *and let $\{\lambda_p\}$, possibly empty, be defined as above. Then each such λ_p satisfies*

$$\text{(38)} \qquad \lambda_p = \sup_{\substack{S\subseteq L_2(R_n),\\ \text{card}\,S<p}}\left\{\inf_{\substack{u\in\mathscr{D}_1\\ ||u||=1,u\perp S}} \int_{R_n}(|\nabla u(\boldsymbol{x})|^2 + V(\boldsymbol{x})|u(\boldsymbol{x})|^2)d\mu_n(\boldsymbol{x})\right\},$$

and such λ_p exists for any integer $p \geq 1$ for which the right side of (38) *is $<h_1$. Moreover, in this statement $\mathscr{D}_1$ may be replaced by $\mathscr{D}_0$, the set of all $u \in L_2(R_n)$ which possess continuous second partials everywhere and such that $u(\boldsymbol{x})$ together with all its partial derivatives of order ≤ 2 is $O([1 + |\boldsymbol{x}|^m]\exp(-\frac{1}{2}|\boldsymbol{x}|^2))$ over $\boldsymbol{x} \in R_n$ for some integer $m > 0$ depending on u.*

For integer $p \geq 1$ define $\tau_p(\mathscr{D}_1)$ as the right side of (38), and similarly $\tau_p(\mathscr{D}_0)$ with $\mathscr{D}_1$ replaced by $\mathscr{D}_0$. $\mathscr{D}_0 \supseteq \mathscr{D}_1$ clearly makes $\tau_p(\mathscr{D}_0) \leq \tau_p(\mathscr{D}_1)$. Thus to prove theorem T. 13 we need only show first that any existing λ_p has $\lambda_p \geq \tau_p(\mathscr{D}_1)$, and secondly that $\tau_p(\mathscr{D}_0) < h_1$ has λ_p existing with $\tau_p(\mathscr{D}_0) \geq \lambda_p$.

Now for each λ_p we may choose $\varphi_p \in \mathscr{D}_2$, the domain of H_2, such that $H_2\varphi_p = \lambda_p\varphi_p$ and $(\varphi_p, \varphi_{p'}) = \delta_{p,p'}$, since H_2 is self-adjoint. Thus using T. 10 and multiplying (32) by $\varphi_{p'}(\boldsymbol{y})$ and integrating over R_n we have, since $(\hat{\varphi}_p, \varphi_{p'}) = (\varphi_p, \varphi_{p'}) = \delta_{p,p'}$,

(39) $$\lambda_p \delta_{p,p'} = \int_{R_n} \{|\boldsymbol{y}|^2 \hat{\varphi}_p(\boldsymbol{y}) \overline{\hat{\varphi}_{p'}(\boldsymbol{y})} + \hat{\psi}_p(\boldsymbol{y}) \overline{\hat{\varphi}_{p'}(\boldsymbol{y})}\} d\mu_n(\boldsymbol{y}) ,$$

the integral of each term in (39) existing finite in the Lebesque sense. This finiteness is clear if $n + \rho \geq 4$, since then Condition II holds and $\varphi_p \in \mathscr{D}_2 = \mathscr{D}$, $|\boldsymbol{y}|\hat{\varphi}_p(\boldsymbol{y}) \in L_2(R_n)$ and $\psi_p \in L_2(R_n)$ by T. 2. Otherwise $2 \leq n + \rho < 4$, and T. 12 yields $|\boldsymbol{y}|\hat{\varphi}_p(\boldsymbol{y}) \in L_2(R_n)$ and $\hat{\varphi}_p \in L_1(R_n) \cap L_2(R_n)$; hence $\hat{\psi}_p = \hat{f}_p + \hat{g}_p$ with $\hat{g}_p \in L_2(R_n)$ and $\hat{f}_p \in L_\infty(R_n)$ from $f_p \in L_1(R_n)$ also makes the second term integral be finite as well as the first. Also Parseval's equality applied to the terms on the right side of (39) yields

(40) $$\lambda_p \delta_{p,p'} = \int_{R_n} \{(\boldsymbol{\nabla}_{\text{gen}} \varphi_p(\boldsymbol{x}) \cdot \overline{\boldsymbol{\nabla}_{\text{gen}} \varphi_{p'}(\boldsymbol{x})}) + V(\boldsymbol{x})\varphi_p(\boldsymbol{x})\overline{\varphi_{p'}(\boldsymbol{x})}\} d\mu_n(\boldsymbol{x}) ,$$

provided that in addition we show

$$\int_{R_n} \hat{f}_p(\boldsymbol{y}) \overline{\hat{\varphi}_{p'}(\boldsymbol{y})} d\mu_n(\boldsymbol{y}) = \int_{R_n} f_p(\boldsymbol{x}) \overline{\varphi_p(\boldsymbol{x})} d\mu_n(\boldsymbol{x})$$

in the case $2 \leq n + \rho < 4$, where as usual $f_p(\boldsymbol{x}) = V(\boldsymbol{x})\varphi_p(\boldsymbol{x})\chi_b(\boldsymbol{x})$ as in T. 10. Replacing V by the truncate V_q defined for (16) and defining ${}_qf_p = V_q \varphi_p \chi_b$, then ${}_qf_p \in L_2(R_n)$ and $({}_q\hat{f}_p, \hat{\varphi}_{p'}) = ({}_qf_p, \varphi_{p'})$ follows by Parseval's equality. Clearly Condition I, $\varphi_p \in L_\infty(R_n)$ by T. 12, and dominated convergence over $\{\boldsymbol{x} \mid |\boldsymbol{x}| \leq b\}$ yields $||f_p - {}_qf_p||_1 \to 0$ as $q \to +\infty$, and hence also $||\hat{f}_p - {}_q\hat{f}_p||_\infty \to 0$. Thus $\varphi_{p'} \in L_\infty(R_n)$ and $\hat{\varphi}_{p'} \in L_1(R_n)$ by T. 12 in our case $2 \leq n + \rho < 4$ gives the desired result

(41) $$\int_{R_n} \hat{f}_p(\boldsymbol{y}) \overline{\hat{\varphi}_{p'}(\boldsymbol{y})} d\mu_n(\boldsymbol{y}) = \lim_{q \to \infty} ({}_q\hat{f}_p, \hat{\varphi}_{p'}) = \lim_{q \to \infty} ({}_qf_p, \varphi_{p'}) = \int_{R_n} f_p(\boldsymbol{x}) \overline{\varphi_{p'}(\boldsymbol{x})} d\mu_n(\boldsymbol{x}) ,$$

and (40) is completely proved.

Now from (40), for $u = \sum_{j=1}^{p} c_j \varphi_j$ we have

(42) $$\int_{R_n} \{|\boldsymbol{\nabla}_{\text{gen}} u(\boldsymbol{x})|^2 + V(\boldsymbol{x})|u(\boldsymbol{x})|^2\} d\mu_n(\boldsymbol{x}) = \sum_{j=1}^{p} \lambda_j |c_j|^2 \leq \lambda_p \left[\sum_{j=1}^{p} |c_j|^2 \right] = \lambda_p ||u||^2 .$$

Next by T. 11 and T. 12, since $\mathscr{D}_2 = \mathscr{D}$ if $n + \rho \geq 4$, for each $\varphi_j \in \mathscr{D}_2$, $1 \leq j \leq p$, we can choose a sequence ${}_k\varphi_j \in \mathscr{D}_1$ having $||\varphi_j - {}_k\varphi_j||_2 \to 0$,

$$\int_{R_n} |V(\boldsymbol{x})| \, |\varphi_j(\boldsymbol{x}) - {}_k\varphi_j(\boldsymbol{x})|^2 \, d\mu_n(\boldsymbol{x}) \to 0 ,$$

and

$$\int_{R_n} |\nabla_{\text{gen}}\varphi_j(x) - \nabla_k\varphi_j(x)|^2 d\mu_n(x) \to 0$$

as $k \to \infty$, and also satisfying $\|\varphi_j - {}_k\varphi_j\|_2 < 1/(3p)$ for all k. This last requirement assures that $|({}_k\varphi_j, {}_k\varphi_{j'}) - \delta_{j,j'}| \leq \theta/p$ for some fixed $\theta < 1$ (actually $\theta = \frac{8}{9}$ here), and hence the set $\{{}_k\varphi_j\}$ over $1 \leq j \leq p$ is linearly independent and thus spans a p dimensional manifold $\mathscr{M}_k$ of $\mathscr{D}_1$. Thus given $S \subseteq L_2(R_n)$ with card $S < p$, the orthogonal projection of S into the subspace $\mathscr{M}_k$ spans at most a $p-1$ dimensional manifold, and hence there exists $u_k \in \mathscr{M}_k$, $\|u_k\| = 1$, $u_k \perp S$. Also

$$u_k = \sum_{j=1}^{p} {}_kc_j\, {}_k\varphi_j$$

has

$$1 = \|u_k\|^2 = \sum_{j,j'=1}^{p} {}_kc_j\, {}_k\bar{c}_{j'}({}_k\varphi_j, {}_k\varphi_{j'}) \geq \sum_{j=1}^{p} |{}_kc_j|^2 - \frac{\theta}{p}\left(\sum_{j=1}^{p} |{}_kc_j|\right)^2 \geq (1-\theta)\sum_{j=1}^{p} |{}_kc_j|^2$$

by the Schwarz inequality,

$$\sum_{j=1}^{p} |{}_kc_j|^2 \leq (1-\theta)^{-1} < +\infty ,$$

and hence by taking subsequences we can assume ${}_kc_j \to {}_0c_j$ for some complex ${}_0c_j$ as $k \to +\infty$ for each j, $1 \leq j \leq p$. Thus $u_0 = \sum_{j=1}^{p} {}_0c_j\, \varphi_j$ has $u_k \to u_0$ in each of the three quadratic form norms for which ${}_k\varphi_j \to \varphi_j$ above, using the Minkowski inequality. Hence (42) for u_0 has the left side to be equal the limit as $k \to +\infty$ of the same expression with u_k replacing u_0. Since $u_k \in \mathscr{M}_k \subseteq \mathscr{D}_1$, $\|u_k\| = 1$, and $u_k \perp S$, we thus see that $\tau_p(\mathscr{D}_1) \leq \lambda_p$ holds for existing $\lambda_p < h_1$, which completes the first part of our proof.

In order to complete the proof Theorem T. 13, we must show $\tau_p(\mathscr{D}_0) < h_1$ has $\tau_p(\mathscr{D}_0) \geq \lambda_p$ with λ_p existing. Consider fixed $u_0 \in \mathscr{D}_0$. The truncate V_q, defined as for (16), with $q > A$ satisfies Condition II clearly, and thus defines the self-adjoint ${}_qH$ with domain $\mathscr{D} \supseteq \mathscr{D}_0$ as in T. 1, and ${}_qH \supseteq {}_qH_0$ defined on $\mathscr{D}_0$ by (2) with V_q. Hence by integrating by parts, and using the exponential bounds in the definition of $\mathscr{D}_0$, ${}_qE$ being the spectral measure for ${}_qH$,

$$\text{(43)} \quad \int_{R_n} \{|\nabla u_0(x)|^2 + V_q(x)\,|u_0(x)|^2\} d\mu_n(x) = ({}_qHu_0, u_0)$$

$$= \int_{-\infty}^{\infty} \lambda d({}_qE(\lambda)u_0, u_0)$$

$$= \left\{ \tilde{\sum_{{}_q\lambda_j < h}}' {}_q\lambda_j({}_qE(\{{}_q\lambda_j\})u_0, u_0)\right\} + \int_{\lambda \geq h} \lambda d({}_qE(\lambda)u_0, u_0)$$

$$\geq \left\{ \tilde{\sum_{{}_q\lambda_j < h}} {}_q\lambda_j({}_qE(\{{}_q\lambda_j\})u_0, u_0)\right\} + h\left\{\|u_0\|^2 - \tilde{\sum_{{}_q\lambda_j < h}} ({}_qE(\{{}_q\lambda_j\})u_0, u_0)\right\}$$

for any $h < h_1$, the sum $\sum^{\sim}_{{}_q\lambda_j<h}$ being finite then by T. 7 and here being defined to give one term for each distinct $\lambda \in \Sigma_q$.

Now taking $q \to +\infty$ in (43), by Condition I and dominated convergence the limit of the left side is obtained by replacing V_q by V. On the right side $\lceil V - V_q \rceil_\omega \to 0$ by (16) under Condition I, and hence $\| G_\omega - {}_qG_\omega \| \to 0$ ([1], 3.20, p. 561). Defining F_ω as the spectral measure of G_ω and $f_\omega(\lambda) = 1/(\lambda + \omega^2)$, we have ([1], Theorem 4.5, p. 567) $E(B) = F_\omega(f_\omega(B))$ for Borel subsets B of the spectrum Σ of H_2; also the usual loop integral formula

$$F_\omega([a, c]) = \frac{1}{2\pi i}\int_C (zI - G_\omega)^{-1} dz$$

holds in the weak sense, where C is a rectangular curve in the complex plane with sides parallel to the axes whose interior region intersects the real axis in (a, c), provided both "a" and "c" are at a positive distance from $f_\omega(\Sigma)$. Thus $\| G_\omega - {}_qG_\omega \| \to 0$ implies $\| E(B) - {}_qE(B) \| \to 0$ for any closed interval $B \subseteq (-\infty, h_1)$ whose endpoints are not in $\{\lambda_p\}$. Hence ${}_q\lambda_j \to \lambda_j$ for λ_j existing, and (43) becomes

$$(44) \qquad \begin{aligned} &\int_{R_n} \{|\nabla u(x)|^2 + V(x)\,|u(x)|^2\} d\mu_n(x) \\ &\geq \Big\{ \sum^{\sim}_{\lambda_j<h} \lambda_j (E(\{\lambda_j\})u, u) \Big\} + h \Big\{ \| u \|^2 - \sum^{\sim}_{\lambda_j<h} (E(\{\lambda_j\})u, u) \Big\} \\ &= \Big\{ \sum_{\lambda_j<h} \lambda_j \,|(u, \varphi_j)|^2 \Big\} + h \Big\{ \| u \|^2 - \sum_{\lambda_j<h} |u, \varphi_j)|^2 \Big\} \end{aligned}$$

for $u \in \mathscr{D}_0$ and $h < h_1$, the sum $\sum_{\lambda_j<h}$ meaning as usual one term for each index j satisfying $\lambda_j < h$.

Now assume $\tau_p(\mathscr{D}_0) < h_1$ for some integer $p \geq 1$, set $h' = \frac{1}{2}[h_1 + \tau_p(\mathscr{D}_0)]$, and thus $\tau_p(\mathscr{D}_0) < h' < h_1$. Now consider the particular $S = \{\varphi_j \,|\, \lambda_j < h'$ exists and $j < p\} \subseteq L_2(R_n)$, for which (card S) $< p$ clearly. Thus (44) with $\| u \| = 1$, $(u, \varphi_j) = 0$ for $\varphi_j \in S$, and $h = h'$ would give $\tau_p(\mathscr{D}_0) \geq h'$ if either λ_p did not exist or else $\lambda_p \geq h'$, yielding the contradiction $h' \leq \tau_p(\mathscr{D}_0) < h'$. Thus $\lambda_p < h' < h_1$ must exist, and (44) with $\| u \| = 1$, $(u, \varphi_j) = 0$ for $j < p$, and $h = \lambda_p$ gives $\tau_p(\mathscr{D}_0) \geq \lambda_p$ as desired. Thus the proof of Theorem T. 13 is complete.

REFERENCES

1. F. H. Brownell, *Spectrum of the static potential Schrödinger equation over E_n*, Annals of Math., **54** (1951), 554-594.

2. T. Kato, *Fundamental properties of Hamiltonian operators of Schrödinger type*, Trans. Amer. Math. Soc., **70** (1951), 195-211.

3. B. v. S. Nagy, *Spectraldarstellung linearer Transformationen des Hilbertshen Raumes*, Ergebnisse Math., **5** (1942).

4. E. C. Titchmarsh, *Introduction to the theory of Fourier integrals*, Oxford U. Press, London, 1937.
5. F. Stummel, *Singulare elliptische Differentialoperatoren in Hilbertschen Raumen*, Math. Annalen, **132** (1956), 150-176.
6. *(added in proof)* N. Nilsson, *Essential self-adjointness and the spectral resolution of Hamiltonian operators*, Kungl. Fysiogr. Sällsk. i Lund Förh., **29** (1959), no. 1, 1-19.

UNIVERSITY OF WASHINGTON

THE RING OF NUMBER-THEORETIC FUNCTIONS

E. D. CASHWELL AND C. J. EVERETT

Introduction. The set Ω of all functions $\alpha(n)$ on $N = \{1, 2, 3, \cdots\}$ to the complex field F forms a domain of integrity under ordinary addition, and *arithmetic product* defined by: $(\alpha \cdot \beta)(n) = \sum \alpha(d)\beta(n/d)$, summed over all $d \mid n$, $d \in N$. The group of units of this domain contains as a subgroup the set of all multiplicative functions. Against this background, the "inversion theorems" of number theory appear as obvious consequences of ring operations, and generalizations of the standard functions arise in a natural way. The domain Ω is isomorphic to the domain P of formal power series over F in a countable set of indeterminates. The latter part of the paper is devoted to proving that the theorem on unique factorization into primes, up to order and units, holds in P and hence in Ω.

1. Definition. The class Ω of all number-theoretic functions α, [4; Ch. IV], i.e., functions $\alpha(n)$ on the set N of natural numbers $n = 1, 2, 3 \cdots$ to the complex field F, forms a domain of integrity (commutative, associative ring with identity and no proper divisors of zero) under ordinary addition: $(\alpha + \beta)(n) \equiv \alpha(n) + \beta(n)$, and an operation, frequently occurring in number theory in various disguises, which we call the arithmetic product:

$$(\alpha \cdot \beta)(n) \equiv \sum \alpha(d)\beta(d')$$

the summation extending over all ordered pairs (d, d') of natural numbers such that $dd' = n$.

The commutativity $\alpha \cdot \beta = \beta \cdot \alpha$ follows from the fact that the correspondence $(d, d') \to (d', d)$ is one-to-one on such a set of ordered pairs to (all of) itself, while the associative law $\alpha \cdot (\beta \cdot \gamma) = (\alpha \cdot \beta) \cdot \gamma$ can be verified by observing that, in either association, $(\alpha \cdot \beta \cdot \gamma)(n) = \sum \alpha(d)\beta(d')\gamma(d'')$, summed over all ordered triples (d, d', d'') with $dd'd'' = n$.

The zero 0 and additive inverse $-\alpha$ of α are of course the functions defined by $0(n) \equiv 0$, and $(-\alpha)(n) \equiv -\alpha(n)$, and one sees at once that the function ε with $\varepsilon(1) = 1$, $\varepsilon(n) = 0$ for $n > 1$, is the identity: $\varepsilon \cdot \alpha = \alpha$ for all α of Ω.

That the ring Ω has no proper divisors of zero may be seen in various ways, three of which occur incidentally in the following sections (2, 4, 5).

2. A norm for number-theoretic functions. A function $N(\alpha)$ on

Received March 8, 1959.

Ω to the set of non-negative integers $0, 1, 2, \cdots$ which is zero if and only if $\alpha = 0$, and has the property $N(\alpha \cdot \beta) = N(\alpha)N(\beta)$ for *all* α, β of Ω, may be defined by setting $N(0) = 0$, and, for all $\alpha \neq 0$, taking $N(\alpha)$ to be the least natural number n for which $\alpha(n) \neq 0$.

Indeed, we find that, if α and β are non-zero functions of Ω with $N(\alpha) = a$ and $N(\beta) = b$, then $(\alpha \cdot \beta)(n) \equiv 0$ for all (if any) n of N with $n < ab$, and $(\alpha \cdot \beta)(ab) = \alpha(a)\beta(b) \neq 0$. It follows that Ω is domain of integrity, and that the norm $N(\alpha)$ has the multiplicative property.

3. Group of units. If for α, β in the domain of integrity Ω, there exists a γ in Ω such that $\alpha = \beta \cdot \gamma$, we say β divides α and write $\beta | \alpha$. The set Υ of all units ν, i.e., elements of Ω which divide the identity ε, forms a commutative group under $(\cdot)$ with identity ε. Two functions α, β of Ω are called associates (notation $\alpha \sim \beta$) in case there is a unit ν such that $\beta = \alpha \cdot \nu$. One sees that $\alpha \sim \beta$ if and only if $\alpha | \beta$ and $\beta | \alpha$, and that $(\sim)$ is an equivalence relation which splits Ω into disjoint classes [ ] of associates. For example, the class [0] contains only 0, while $[\varepsilon] = \Upsilon$. These trivial properties are shared by all domains of integrity.

In our ring Ω, an element α is a unit if and only if $\alpha(1) \neq 0$, equivalently $N(\alpha) = 1$. For, if $\alpha\alpha' = \varepsilon$, $1 = \varepsilon(1) = \alpha(1)\alpha'(1)$ implies $\alpha(1) \neq 0$. To see that this is also sufficient, we first introduce the (number-theoretic) function $\lambda(n)$ defined by $\lambda(1) = 0$, $\lambda(p_1 \cdots p_l) = l$ for any product of l (not necessarily distinct) primes. We have $\lambda(a) = 0$ if and only if $a = 1$, and $\lambda(ab) = \lambda(a) + \lambda(b)$ always. This function has the property of classifying all natural numbers according to their *length*. We have now to construct a function α' in Ω with $(\alpha \cdot \alpha')(n) \equiv \varepsilon(n)$ from a given α for which $\alpha(1) = A \neq 0$. Manifestly, for $n > 1$, this relation itself defines the value of $\alpha'(n)$ unambiguously for each n of length $\lambda(n) = l$ in terms of values $\alpha'(d')$ with $\lambda(d') < l$. Thus, if we define $\alpha'(1) = 1/A$ for the single n of length 0, and proceed inductively on $\lambda(n)$, we automatically obtain the desired α'.

We note in passing that if α, β are any two number-theoretic functions and $\nu \cdot \nu' = \varepsilon$, then $\beta = \alpha \cdot \nu$ if and only if $\alpha = \beta \cdot \nu'$. This trivial relation between associates is the basis for the so-called inversion theorems of number theory. (Cf. § 7).

4. The degree of a number-theoretic function. Just as a natural order $1 < 2 < 3 < \cdots$ of the set N permitted the definition of a norm, so does the order implicit in the λ function enable us to introduce what we may call the *degree* $D(\alpha)$ of a non-zero function α of Ω.

Specifically we take $D(\alpha) = d$ to mean that $\alpha(n) = 0$ for all (if any) n of N with $\lambda(n) < d$, and that there exists an n with $\lambda(n) = d$ for which $\alpha(n) \neq 0$. Thus $D(\alpha)$ is a function on all non-zero α of Ω to the non-negative integers, with $D(\alpha) = 0$ if and only if α is a unit, and $D(\alpha \cdot \beta) = D(\alpha) + D(\beta)$ for all non-zero α, β.

We may indeed show somewhat more. Let $D(\alpha) = d$, $D(\beta) = e$, and suppose a and b are respectively the least integers with $\lambda(a) = d$, $\lambda(b) = e$, for which $\alpha(a) \neq 0$, $\beta(b) \neq 0$. Then $(\alpha \cdot \beta)(n) = 0$ for all (if any) n with $\lambda(n) < d + e$; $(\alpha \cdot \beta)(ab) = \alpha(a)\beta(b) \neq 0$, where, of course, $\lambda(ab) = d + e$; and finally, indeed, $(\alpha \cdot \beta)(n) = 0$ for all $n < ab$ with $\lambda(n) = d + e$, that is to say, ab is itself the least integer of its length at which $\alpha \cdot \beta$ does not vanish.

5. A second norm. The final remarks of the preceding section make it clear that another norm $M(\alpha)$ is available. Specifically, set $M(0) = 0$, and for $\alpha \neq 0$ with $D(\alpha) = d$, set $M(\alpha) = a$, where a is the least integer of length $\lambda(a) = d$ for which $\alpha(a) \neq 0$. It follows that $M(\alpha)$ is a function on all α of Ω to the non-negative integers such that $M(\alpha) = 0$ if and only if $\alpha = 0$, $M(\alpha) = 1$ if and only if α is a unit, and $M(\alpha \cdot \beta) = M(\alpha)M(\beta)$ always.

Thus $M(\alpha)$ has all the properties proved for $N(\alpha)$ and moreover determines $D(\alpha) = \lambda(M(\alpha))$ for $\alpha \neq 0$.

6. The multiplicative functions. This and the following few sections (7–10) are to some extent expository, our object being to observe how familiar results appear when considered from the point of view of the ring Ω or to propose some natural generalizations suggested by the new notation. After this we return to the "arithmetic" of the domain Ω itself.

A number-theoretic function α is said to be multiplicative in case $(a, b) = 1$ implies $\alpha(ab) = \alpha(a)\alpha(b)$ and (to exclude the trivial $\alpha = 0$) there is an integer n for which $\alpha(n) \neq 0$. In the presence of the former property, the latter is equivalent to the condition $\alpha(1) = 1$, which signifies for us that the set M of all multiplicative functions is a subset of the group Υ of units of Ω.

Clearly (1) a function α for which $\alpha(1) = 1$ and $\alpha(\Pi p^a) = \Pi \alpha(p^a)$ is multiplicative, $\alpha(p^a)$ being quite arbitrary for each power $a = 1, 2, \cdots$ of each prime p; and (2) two *multiplicative* functions identical on all such p^a are equal.

That $M \cdot M \subset M$ follows readily from the definition of M, and the identity ε is in M, seen perhaps most trivially from (1) above. To see that M is a *subgroup* of Υ requires only the further fact that the inverse α' of a multiplicative function α, which we know exists uniquely, is itself multiplicative. This we prove in a way which provides a second construction of the inverse in the case of a multiplicative function. [5; p. 89]

Given α in M, define a function β in Ω as follows. Set $\beta(1) = 1$. For each p, define $\beta(p^a)$ for $a = 1, 2, \cdots$ successively by the relation $\sum \alpha(d)\beta(d') = 0$, summed over all pairs (d, d') with $dd' = p^a$. Finally, define $\beta(\Pi p^a) = \Pi \beta(p^a)$. The β thus defined is in M by (1) above. Since α is also in M, we know $\alpha \cdot \beta \in M \cdot M \subset M$. To verify that the functions $\alpha \cdot \beta$ and ε of M are equal, it suffices, by (2) above, to observe that $(\alpha \cdot \beta)(p^a) = \varepsilon(p^a) = 0$, which is the defining equation for $\beta(p^a)$. Since the inverse of any unit is unique, the β so constructed must coincide with that obtainable by the λ construction of § 3.

7. The special multiplicative functions n^k. Define the (multiplicative) function ν_k for arbitrary real k by $\nu_k(n) = n^k$. Its inverse ν_k' is seen by the preceding construction to be: $\nu_k'(1) = 1$, $\nu_k'(n) = (-1)^l n^k$ when n is a product of l *distinct* primes, and zero otherwise.

Now (a) $\nu_k' \cdot \nu_k = \varepsilon$, and (b) if α, β are any two number-theoretic functions, we have $\beta = \alpha \cdot \nu_k$ if and only if $\alpha = \beta \cdot \nu_k'$. For the special case $k = 0$, (a) yields the familiar equation $\sum_{d|n}\mu(d) = \varepsilon(n)$, and (b) becomes the "Möbius inversion theorem" [Cf. 4; Th. 35, 38], since ν_0' is the Möbius function μ. Indeed, we may write $\nu_k'(n) \equiv \mu(n)n^k$ for all k, n.

We may note one further generalization in this direction. If α and β are any two number-theoretic functions, we see that

$$(1) \qquad \sum_{m=1}^{n}(\alpha \cdot \beta)(m) = \sum_{m=1}^{n}\sum_{d|m}\alpha(d)\beta(m/d) = \sum_{d=1}^{n}\alpha(d)\sum_{l=1}^{[n/d]}\beta(l)\,.$$

In particular, if β is a unit, and $\alpha = \beta'$, we obtain

$$1 = \sum_{d=1}^{n}\beta'(d)\sum_{l=1}^{[n/d]}\beta(l)\,.$$

Further specializing to $\beta = \nu_k$,

$$1 = \sum_{d=1}^{n}\mu(d)d^k\sum_{l=1}^{[n/d]}l^k\,.$$

Finally, $k = 0$ gives the familiar [4; Th. 36]

$$1 = \sum_{d=1}^{n}\mu(d)[n/d]\,.$$

8. The sum of the k-th powers of the divisors. It is clear that the transform $\beta(n) = \sum_{d|n}\alpha(d)$ of number theory [5, Th. 6–8] appears in our notation as $\beta = \alpha \cdot \nu_0$. Thus in particular the number theoretic function $\sigma_k(n) = \sum_{d|n}d^k$ is seen to be the (multiplicative) function $\sigma_k = \nu_k \cdot \nu_0 \in M \cdot M \subset M$. The most familiar are $\tau = \sigma_0 = \nu_0 \cdot \nu_0$, the number of divisors, and $\sigma = \sigma_1 = \nu_1 \cdot \nu_0$, the sum of the divisors.

As an illustration, note that equation (1) of the preceding section yields

$$\sum_{m=1}^{n}(\alpha \cdot \nu_0)(m) = \sum_{d=1}^{n}\alpha(d)[n/d] ;$$

in particular, for $\alpha = \nu_0$,

$$\sum_{m=1}^{n}\tau(m) = \sum_{d=1}^{n}[n/d] ,$$

and for $\alpha = \nu_1$,

$$\sum_{m=1}^{n}\sigma(m) = \sum_{d=1}^{n}d[n/d] .$$

The inverse $\sigma'_k(n)$ is 1 for $n=1$, $(-1)^{\lambda}\Pi_i(p_i^k+2-a_i)$ for $n=p_1^{a_1}\cdots p_l^{a_l}$, where $1 \leqq a_i \leqq 2$ and $\lambda = \lambda(n)$, and zero otherwise. This may be seen from $\sigma'_k = \nu'_0 \cdot \nu'_k$ and the value of $(\nu'_0 \cdot \nu'_k)(p^a)$ obtained from § 7. For the special case $k = 0$, we may write $\tau'(n)$, for n of the second type, as $(-1)^{\lambda}2^l/a_1 \cdots a_l$.

We note that the relation $\sigma'_k = \nu'_k \cdot \nu'_0$, besides determining the function σ'_k explicitly as indicated above, yields also the equation $\sigma'_k(n) \equiv \sum_{d|n}d^k\mu(d)\mu(n/d)$, in particular $\tau'(n) \equiv \sum_{d|n}\mu(d)\mu(n/d)$.

9. A generalized φ-function. The well-known relations $\varphi \cdot \nu_0 = \nu_1$ and $\varphi = \nu'_0 \cdot \nu_1$ satisfied by the Euler φ-function [4; Th. 39, 40] suggest definition of a general function $\varphi_{k,l} = \nu'_k \cdot \nu_l$, specifically

$$\varphi_{k,l}(n) = n^l \sum_{d|n}\mu(d)d^{k-l}$$

which has the value $n^l\Pi_i(1 - p_i^{k-l})$ for $n = p_1^{a_1}\cdots p_t^{a_t}$. We should then have the relation $\nu_k \cdot \varphi_{k,l} = \nu_l$ or $\sum_{d|n}\varphi_{k,l}(d)d^{-k} = n^{l-k}$.

It is clear that the derivation of relations between arithmetic functions becomes simplified by employing the algebra of the ring Ω, or of the groups Γ or M. Consider for instance how easily $\sigma = \nu_0 \cdot \nu_1$, $\nu_1 = \nu_0 \cdot \varphi$, and $\nu_0 \cdot \nu_0 = \tau$ implies $\sigma = \tau \cdot \varphi$.

Not quite so elegant is the generalization:

$$n^k\sigma_{l-k}(n) = (\nu_k \cdot \nu_l)(n) , \tag{1}$$

$$\nu_l = \nu_k \cdot \varphi_{k,l} , \tag{2}$$

$$\nu_k \cdot \nu_k(n) = n^k\tau(n) \qquad \text{(special case of (1))}, \tag{3}$$

imply $n^k\sigma_{l-k}(n) = \sum_{d|n}d^k\tau(d)\varphi_{k,l}(n/d)$.

10. The Φ-function. Define the number-theoretic function $\Phi(n)$ to be the *sum* of the integers in N which are prime to n and do not exceed n. Obviously $\Phi(n) = n\varphi(n)/2$ unless $n = 1$ and $\Phi(1) = 1$. Although

Φ is thus a unit in Γ, $\Phi(ab) = 2\Phi(a)\Phi(b)$ for $(a, b) = 1$, $a > 1$, $b > 1$, and therefore Φ is not in M.

If we classify the integers $1, 2, \cdots, n$ according to their greatest common divisor d with n, we find in the d-class the integers a with $(a, n) = d$, $1 \leqq a \leqq n$. There are exactly as many such a as there are b with $(b, n/d) = 1$, $1 \leqq b \leqq n/d$. This yields for Landau [4; Th. 39] the relation $\sum_{d|n}\varphi(n/d) = n$ and the formula for φ by Möbius inversion. We may note that the same partition suggests the additional relation:

$$\kappa(n) = \frac{n(n+1)}{2} = \sum_{a=1}^{n} a = \sum_{d|n} d\Phi(n/d) = (\Phi \cdot \nu_1)(n) .$$

As a final example, we note that, since $\nu_1 \cdot \nu_0 = \sigma$,

$$\kappa \cdot \nu_0 = \Phi \cdot \sigma .$$

11. Primes. A number-theoretic function α is said to be a prime in case $\alpha \neq 0$, α is not a unit, and $\alpha = \beta \cdot \gamma$ implies β or γ is a unit. The associates of a prime are also prime. The remaining functions, neither 0, units, nor primes, are called composite. The associates of a composite function are composite.

Any function with $N(\alpha)$ a prime natural number is prime; more generally any function with $M(\alpha)$ a prime, or equivalently, any function with $D(\alpha) = 1$. As an example, note that from § 9 $\delta \equiv \sigma - \nu_1 = \tau \cdot \varphi - \nu_0 \cdot \varphi = (\tau - \nu_0) \cdot \varphi$. Since $\delta(1) = 0$ and $\delta(2) = 1$, we see that $M(\delta) = 2$ and so $\sigma - \nu_1$ and $\tau - \nu_0$ are associated primes. If two non-unit functions α, β are associates, we see that $\beta(p) = (\nu \cdot \alpha)(p) = \nu(1)\alpha(p)$ for all prime p, where $\nu(1) \neq 0$. Hence there is a continuum of non-associated primes even of this simple type.

Naturally there are many other kinds of primes, a fact which will become glaringly obvious in § 16.

12. The chain condition. If $\alpha_0 \neq 0$, $\alpha_1 | \alpha_0$, and in the corresponding equation $\alpha_0 = \alpha_1 \cdot \beta_1$ the (uniquely determined) β_1 is not a unit, we say α_1 properly divides α_0 and write $\alpha_1 || \alpha_0$. For example, every composite element α has a factorization $\alpha = \beta \cdot \gamma$ in which $\beta || \alpha$ and $\gamma || \alpha$. If in a domain of integrity, every chain of proper divisors $\cdots \alpha_2 || \alpha_1 || \alpha_0 \neq 0$ is finite, we say the domain satisfies the chain condition. In any such domain it is easy to see [2; p. 117] first that every α not zero and not a unit has a prime divisor, and from this that every such α is expressible as a finite product of primes.

That our ring satisfies the chain condition is an obvious consequence of the properties of either the norm or the degree functions. For example, $\alpha_1 || \alpha_0 \neq 0$, $\alpha_0 = \alpha_1 \cdot \beta_1$, β_1 not a unit, implies $D(\beta_1) > 0$ and $D(\alpha_0) = D(\alpha_1) + D(\beta_1) > D(\alpha_1)$, where D has non-negative integral values.

Having come this far, it is natural to ask whether the expression of a non-zero, non-unit number-theoretic function as a product of primes is unique (up to order and units). We have been unable to find a reference for such a theorem, and offer a proof in the remaining sections.

In the presence of the chain condition, the existence of a greatest common divisor for every two elements is necessary and sufficient for the uniqueness property. [2; p. 120]. Although we have an abundance of norms, we cannot hope to obtain a Euclidean algorithm, since we certainly could not have linear expressibility of the g.c.d. For suppose α, β are non-associated primes. Then (α, β) certainly exists and is ε. whereas a linear relation $\varepsilon = \gamma \cdot \alpha + \delta \cdot \beta$ is impossible (consider $n = 1$),

13. A reduction theorem. It simplifies matters to show first that if the uniqueness of factorization fails, it must fail in a particularly simple way. Suppose indeed that uniqueness in false in Ω. Following an argument of Lindemann and Davenport [1; § 2.11] let us divide the set of all non-zero non-unit elements of Ω into normal elements, whose factorization into primes *is* unique, and *abnormal* elements, which can be factored into primes in two essentially different ways. Clearly a prime α is normal by definition.

We prove that if α is an abnormal element of minimal norm $N(\alpha)$, and $\alpha = \sigma_1 \cdots \sigma_m = \tau_1 \cdots \tau_n$ are two essentially different factorizations of α into primes, σ_i, τ_j, then necessarily $m = n = 2$ and $\sigma_1, \sigma_2, \tau_1, \tau_2$ all have the same norm N.

Note first that neither m nor n is unity, since a prime is normal. Moreover, no σ_j is the associate of any τ_j, for if so, cancellation would produce an abnormal element of norm $N < N(\alpha)$. Without loss of generality, we may assume $N(\sigma_1) \leqq N(\sigma_2) \leqq \cdots \leqq N(\sigma_m)$, $N((\tau_1) \leqq N(\tau_2) \leqq \cdots \leqq N(\tau_n)$, and $N(\sigma_1) \leqq N(\tau_1)$. Then $N(\sigma_1 \cdot \tau_1) = N(\sigma_1) \cdot N(\tau_1) \leqq N(\tau_1)N(\tau_1) \leqq N(\tau_1)N(\tau_2) \leqq N(\alpha)$. If any one of these ($\leqq$) relations is actually ($<$), we have $N(\sigma_1 \cdot \tau_1) < N(\alpha)$, which we will see leads to a contradiction.

Suppose indeed that $N(\sigma_1 \cdot \tau_1) < N(\alpha)$, and consider $\beta = \alpha - \sigma_1 \cdot \tau_1$. Certainly $\beta \neq 0$, for $\alpha = \sigma_1 \cdot \tau_1$ implies $\sigma_2 \cdots \sigma_m = \tau_1$, and since τ_1 is prime, we have $m = 2$ and $\tau_1 \sim \sigma_2$, contradiction. Also β is not a unit, since $\sigma_1 | \beta$. From the definition of norm N and the assumption $N(\sigma_1 \cdot \tau_1) < N(\alpha)$ it follows that $N(\beta) = N(\sigma_1 \cdot \tau_1) < N(\alpha)$. Hence β is *normal*. However, the non-associates σ_1, τ_1 both divide β, and, β being normal, $\sigma_1 \cdot \tau_1 | \beta$. Hence $\sigma_1 \cdot \tau_1 | \alpha = \sigma_1 \cdots \sigma_m = \sigma_1 \cdot \tau_1 \cdot \gamma$. Thus $\sigma_2 \cdots \sigma_m = \tau_1 \cdot \gamma$. But $N(\sigma_2 \cdots \sigma_m) < N(\alpha)$, and $\sigma_2 \cdots \sigma_m$ is not zero and not a unit ($m \geqq 2$). It follows that $\sigma_2 \cdots \sigma_m = \tau_1 \cdot \gamma$ is normal and τ_1 is associated with some σ_j, a contradiction.

We are forced to conclude that $N(\sigma_1)N(\tau_1) = N(\tau_1)N(\tau_1) = N(\tau_1)N(\tau_2) = N(\alpha)$ and so $N(\sigma_1) = N(\tau_1) = N(\tau_2) \equiv N$ and $n = 2$. Hence $N^2 =$

$N(\tau_1)N(\tau_2) = N(\alpha) = N(\sigma_1) \cdots N(\sigma_m) \geqq N^m$ implies $m \leqq 2$. But $m > 1$ so $m = 2$, $N(\sigma_2) = N$, and all is proved.

Thus if unique prime factorization fails in Ω, we should have an element of form $\alpha \cdot \beta = \gamma \cdot \delta$, α, β, γ, δ primes (of identical norm N) and α not associated with either γ or δ.

14. The ring of formal power series. Let the primes p of N be listed in *any* definite order $p_1, p_2, p_3, \cdots$. Then every integer n may be written uniquely in the form $n = p_1^{a_1}p_2^{a_2} \cdots$ and uniquely described by a vector $(a_1, a_2, \cdots)$ with non-negative integral components, finitely many of which are non-zero, all such vectors being realized as n ranges over N. Hence a number-theoretic function $\alpha = \alpha(n)$ may be associated with a definite "formal power series" in a countably infinite number of indeterminates $x_1, x_2, \cdots$, having coefficients in the complex field F, by means of the correspondence

$$\alpha \to P(\alpha) = \Sigma\alpha(n)x_1^{a_1}x_2^{a_2} \cdots .$$

Here, the summation extends over all $n = p_1^{a_1}p_2^{a_2} \cdots$ of N.

This correspondence is clearly one to one on Ω to the set $F_\omega = F\{x_1, x_2, \cdots\}$ of *all* such power series. Moreover, addition is preserved, and $P(\alpha \cdot \beta) = P(\alpha)P(\beta)$, the latter operation being the usual formal operation on power series involving multiplication and collection of (finite numbers of) "like terms."

Thus the ring of all number-theoretic functions is isomorphic to the ring of all formal power series $F_\omega = F\{x_1, x_2, \cdots\}$. We emphasize that the only restriction on these series is that only a *finite* number of x_i actually appear (i.e., have $a_i > 0$) in any *term*. However, infinitely many x_i may well occur (in terms with non-zero coefficients) in the *same* series, so that we have here a more general ring than that discussed by Krull [3; § 4]. Indeed, each series of Krull's ring of power series (over F) corresponds to a number theoretic function zero except on a set of integers generated by *some* finite set of primes.

15. Some preliminaries. We deal in the remainder of the paper only with the power series representation $A = A\{x_1, x_2, \cdots\} = \Sigma\alpha(n)x_1^{a_1}x_2^{a_2} \cdots$ of number-theoretic functions. The domain $F_\omega = F\{x_1, x_2, \cdots\}$ contains (in the sense of isomorphism) for every $l = 1, 2, \cdots$ the domain $F_l = F\{x_1, \cdots, x_l\}$ of power series in l "variables." For the latter domains, the theorem on unique factorization into primes is known. [3; § 4 and 6; § 2]. The units of F_l are again the series with non-zero constant term.

If l is any integer $1, 2, \cdots$ and if $A = A\{x_1, x_2, \cdots\}$ is in F_ω or some F_m with $m \geqq l$, we mean by $(A)_l$ the series $A\{x_1, \cdots, x_l, 0, 0, \cdots\}$

obtained from A by deleting all terms of A actually involving any x_i with $i > l$. Indeed, the mapping $A \to (A)_l$ is a ring homomorphism of F_ω or F_m onto F_l. One can write $A = (A)_l + A_l^*$, where the latter series involves only terms containing at least one x_i with $i > l$, and in this way one sees that $(AB)_l = (A)_l(B)_l$.

In reality all series we consider are actually in F_ω, but we do not hesitate to say $A\{x_1, \cdots, x_l, 0, 0, \cdots\}$ is "in F_l." Our objective is to throw the proof of unique factorization in F_ω back onto the rings F_l, $l = 1, 2, \cdots$, in which the theorem is known to be true. But first we have to show that the primes of F_ω are all of a special kind.

16. The nature of a prime. If a series A of F_ω is neither zero nor a unit, then there is some minimal $L = L(A)$ for which $(A)_l$ is neither zero nor a unit of F_l, $l \geqq L$. For $A\{0, 0, \cdots\} = 0$, and since $A \neq 0$, A must contain with non-zero coefficient some product $x_1^{a_1}x_2^{a_2}\cdots$ with $(a_1, a_2, \cdots) \neq (0, 0, \cdots)$. If in this term x_k is the last variable with $a_k > 0$, then $(A)_k \neq 0$. Hence there is a *minimal* L with $(A)_L \neq 0$, $L \geqq 1$. But then $(A)_l$ is not zero or a unit for any $l \geqq L$.

Now if A is not zero or a unit in F_ω, and any $(A)_l$ is prime in F_l, where of course $l \geqq L = L(A)$, then $(A)_m$ is prime in F_m for all $m \geqq l$, and also A is prime in F_ω. For example, if $(A)_m = R_mS_m$, where R_m, S_m are non-units in F_m, then $(A)_l = (A_m)_l = (R_m)_l(S_m)_l$, where neither of the latter factors in F_l are units. For such A, there is a minimal integer $P = P(A) \geqq L(A)$ such that $(A)_l$ is prime in F_l for all $l \geqq P(A)$. We say such primes are *finitely prime.*

The remaining logical possibility is that for some A, not zero or a unit, we have $(A)_l$ composite in F_l for all $l \geqq L(A)$. We shall show that such an A is composite in F_ω, and hence the

Principal Lemma: all primes of F_ω are finitely prime.

17. Proof of the principal lemma. Let A be a fixed non-zero non-unit series in F_ω with $L = L(A)$, and suppose that, for every $l \geqq L$, $(A)_l = R_lS_l$ where R_l and S_l are non-units of F_l. We say R_l and S_l are true factors of $(A)_l$ and R_lS_l is a true factorization of $(A)_l$. A true factor of $(A)_l$ is thus a non-unit proper divisor of $(A)_l$ in F_l, and so has a companion of the same kind.

We shall call any chain $[R_L, R_{L+1}, \cdots, R_M]$ of true factors of the corresponding $(A)_l$, $l = L, \cdots, M$ *telescopic* if each $R_{l-1} = R_l(x_1, \cdots, x_{l-1}, 0) = (R_l)_{l-1}$. Now observe that any true factorization $(A)_m = R_mS_m$, $m > L$ induces a true factorization of $(A)_{m-1} = ((A)_m)_{m-1} = (R_m)_{m-1}(S_m)_{m-1} \equiv R_{m-1}S_{m-1}$ and so down to $(A)_L = R_LS_L$, where the chain of true factors $[R_L, \cdots, R_m]$ *is* telescopic. Thus we have from the original assumption on A, the existence of a sequence

$$\kappa_0 = [R_{00}]$$
$$\kappa_1 = [R_{10}, R_{11}]$$
$$\kappa_2 = [R_{20}, R_{21}, R_{22}]$$
$$\vdots$$

of *telescopic* chains κ_i of *true* factors R_{ij}, $j = 0, 1, \cdots, i$ of $(A)_{L+j}$.

We want to prove the existence of an *infinite* chain of *true* factors $\kappa^* = [R_0^*, R_1^*, R_2^*, \cdots]$ which is telescopic throughout. If we could do so, we should have $(A)_{L+j} = R_j^* S_j^*$ for all $j \geqq 0$. Clearly the chain $[S_0^*, S_1^*, \cdots]$ is also telescopic, since $(R_{j-1}^* S_{j-1}^*) = (R_j^* S_j^*)_{L+j-1} = (R_j^*)_{L+j-1} \cdot (S_j^*)_{L+j-1} = R_{j-1}^*(S_j^*)_{L+j-1}$. But any infinite telescopic chain *defines* unambiguously a series of F_ω. If R^* and S^* are the (non-unit) series defined by the R_j^* and S_j^* chains, we must have $A = R^* S^*$, since we can prove identity of the left and right coefficients of any term by regarding $(A)_{L+j} = R_j^* S_j^*$ for suitable j. Thus the principal lemma would be proved.

Since unique factorization holds in F_l, there are only a finite number of classes of associates into which the true factors of any $(A)_l$ can fall. Hence (pigeon-hole principal!) an *infinite* set of the chains κ_i have their *first* entry equivalent to some *one* true factor T_0 of $(A)_L$. Choose one of these and call it κ_0'. Of *this* infinite set, there is an infinite *subset* of κ_i whose *second* entry is equivalent to some one true factor T_1 of $(A)_{L+1}$. Choose one and call it κ_1'. Continuing in this way we are led to a *subsequence* of (telescopic) chains

$$\kappa_0' = [R_{00}', \cdots]$$
$$\kappa_1' = [R_{10}', R_{11}', \cdots]$$
$$\kappa_2' = [R_{20}', R_{21}', R_{22}', \cdots]$$

each of which extends at least to the main diagonal, such that the entries on *this diagonal and below* have the property that, for each $j = 0, 1, 2, \cdots$ $R_{ij}' \sim T_j$ for all $i \geqq j$.

We can now construct the telescopic infinite chain κ^* working only with the main diagonal and the diagonal next below it, as follows. Define $R_0^* = R_{00}'$. Since $R_{10}' \sim T_0 \sim R_0^*$ in F_L, there is a unit U_L of F_L such that $R_0^* = R_{10}' U_L = (R_{11}' U_L)_L$. Define $R_1^* = R_{11}' U_L$ in F_{L+1}, and note that R_1^* is a true factor of $(A)_{L+1}$, $(R_1^*)_L = R_0^*$, and $R_1^* \sim T_1$ in F_{L+1}.

To make the process perfectly clear and to avoid a formal induction, we carry the construction through one more step. Since $R_{21}' \sim T_1 \sim R_1^*$ in F_{L+1}, there is a unit U_{L+1} of F_{L+1} such that $R_1^* = R_{21}' U_{L+1} = (R_{22}' U_{L+1})_{L+1}$. Define $R_2^* = R_{22}' U_{L+1}$ in F_{L+2} and note that R_2^* is a true factor of $(A)_{L+2}$, $(R_2^*)_{L+1} = R_1^*$, and $R_2^* \sim T_2$ in F_{L+2}. The proof of the lemma is now clear.

18. Proof of unique factorization. Suppose unique factorization into primes fails in $\Omega \cong F_\omega$. By § 13, we must have a series of the form $AB = CD$ where A, B, C, D are primes in F_ω and A is not associated with C or D. Since all primes are of finite type, there exists an integer P such that, in the equation $(AB)_l = (A)_l(B)_l = (C)_l(D)_l = (CD)_l$, $(A)_l$, $(B)_l$, $(C)_l$, $(D)_l$ are primes in F_l for all $l \geqq P$. Since factorization in each F_l is unique, $(A)_l$ must be associated with either $(C)_l$ or $(D)_l$ in F_l for each $l \geqq P$. Hence there must be an infinite increasing subsequence $\sigma = \{m\}$ of integers $m \geqq P$ such that either $(A)_m \sim (C)_m$ in F_m or $(A)_m \sim (D)_m$ in F_m for all $m \in \sigma$. Without loss of generality we may suppose the former case. Then $(A)_m = U_m(C)_m$, where U_m is a unit of F_m, for each m of σ. If $m < n$ are any two integers of the sequence σ, $U_m(C)_m = (A)_m = (A_n)_m = (U_n)_m(C_n)_m = (U_n)_m(C)_m$, and U_n is an extension of U_m by terms each of which involves a variable x_i with $i > m$ and so does not occur in U_m. Thus the sequence U_m, $m \in \sigma$ defines a unit U of F_ω, and $A = UC$, by the same type of argument used in the preceding section in showing $A = R^*S^*$. But then $A \sim C$ in F_ω, which is a contradiction. Hence *factorization into primes exists and is unique in the rings* Ω *and* F_ω, *up to order and units.*

References

1. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 3rd Ed. (1954) Oxford Univ. Press, London.

2. N. Jacobson, *Lectures in abstract algebra*, **1** (1951) D. Van Nostrand Co., N. Y.

3. W. Krull, *Beiträge zur Arithmetik kommutativer Integritätsbereiche, III. zum Dimensionsbegriff der Idealtheorie*, Math. Zeit. **42** (1937), 745-766.

4. E. Landau, *Elementary number theory*, (1927) Chelsea Pub. Co., N. Y.

5. W. J. LeVeque, *Topics in number theory*, **1** (1956), Addison-Wesley Pub. Co., Reading, Mass.

6. W. Rückert, *Zum Eliminationsproblem der Potenzreihenideale*, Math. Ann. **107** (1933), 259-281.

UNIVERSITY OF CALIFORNIA
LOS ALAMOS SCIENTIFIC LABORATORY
LOS ALAMOS, NEW MEXICO

ON CONTINUATION OF BOUNDARY VALUES FOR PARTIAL DIFFERENTIAL OPERATORS

H. O. CORDES

Let

$$L = \sum_{i=1}^{n} a_i(x)\partial/\partial x_i + b(x) \tag{1}$$

be a first order partial differential operator acting on m-component vector functions and defined in a bounded domain D with smooth boundary Γ. Suppose the $m \times m$-matrices $a_i(x)$ are hermitian symmetric and continuously differentiable in $D + \Gamma$. Further let the $m \times m$-matrix $b(x)$ be bounded and measurable over $D + \Gamma$.

Recently K. O. Friedrichs [3] has developed a theory of boundary value problems of the type

$$\begin{aligned} (L - \alpha)u &= f, \qquad x \in D \\ Tu &= 0, \qquad x \in \Gamma \end{aligned} \tag{2}$$

where α denotes a nonvanishing real constant and T a certain $m \times m$-matrix defined all over the boundary T and satisfying certain further conditions. Concurrently the author worked on the same type of boundary value problem from a different approach extending Friedrich's results to the case of nonlocal boundary conditions [1].

Study of these extensions showed that investigation of the following problem is of basic importance for the author's method:

The question is asked whether a given m-component vector function φ defined on the boundary Γ can be continued into the domain D to become a classical solution u of the equation

$$L(u) = f$$

where f is any arbitrary measurable function defined and squared integrable over D, which is not given in advance but may be defined after φ has been fixed.

Obviously this question is trivially answered "yes" if the boundary and the boundary function are sufficiently smooth. On the other hand if this is not the case, counter examples can be given. It is trivial to find counter examples for special nonelliptic systems but one also can find some for elliptic systems. For instance if the boundary functions u_0, v_0 on the periphery of the unit circle $x^2 + y^2 = 1$ are defined by

$$u_0 = \alpha(\vartheta)\sin\vartheta/2, \qquad v_0 = -\alpha(\vartheta)\cos\vartheta/2, \qquad 0 \le \vartheta \le 2\pi \tag{3}$$

Received January 12, 1959. This paper has been prepared under the sponsorship of the Office of Naval Research, Contract No. Nonr 2-228(09).

and if $\alpha(\vartheta)$ is piecewise continuous and has a jump for any $\vartheta_0 \neq 0, 2\pi$, then it will be shown in §4 that there does not exist any couple u, v of real or complex valued functions both being defined and continuously differentiable in the open unit disk $x^2 + y^2 < 1$ and such that

(a) $$-u_x + v_y = f, \qquad u_y + v_x = g$$

both are squared integrable over $x^2 + y^2 < 1$;

(b) u, v are uniformly bounded on $x^2 + y^2 < 1$ and

(c) $$\lim_{r \to 1} u(r \cos \vartheta, r \sin \vartheta) = u_0(\vartheta)$$
$$\lim_{r \to 1} v(r \cos \vartheta, r \sin \vartheta) = v_0(\vartheta)$$

almost everywhere on $0 \leq \vartheta \leq 2\pi$.

Considering this problem more carefully it shows that the essential reason for this continuation to be impossible is the following:

The above problem can be connected with the differential operator

(5) $$L = a_1 \partial/\partial x + a_2 \partial/\partial y$$

with a_1, a_2 being the matrices

(6) $$a_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad a_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Using this operator notation we can say that the equation

(7) $$L\varphi = \psi$$

with φ, ψ being two component vector functions has no classical solution, defined in the unit disk and achieving the boundary values defined by

(8) $$\varphi(x, y) = (u_0(\vartheta), v_0(\vartheta)) \qquad x = \cos \vartheta, y = \sin \vartheta$$

in the sense of the conditions (a), (b), and (c) mentioned above.

If we define

(9) $$A(\vartheta) = a_1 \cos \vartheta + a_2 \sin \vartheta$$

(10) $$\tilde{A}(\vartheta) = a_2 \cos \vartheta - a_1 \sin \vartheta$$

then

(11) $$L = A(\vartheta)\partial/\partial r + r^{-1}\tilde{A}(\vartheta)\partial/\partial\vartheta .$$

Hence $A(\vartheta)$ is the coefficient of the derivative in the direction normal to the boundary.

We note that $A(\vartheta)$ is a non-singular (even orthogonal) matrix for every ϑ. It will follow from our development that this is the reason why a continuation of discontinuous boundary values becomes impossible. If for some more general operator L the matrix which corresponds to $A(\vartheta)$ is singular on a point or on a set of points then this set can be allowed to contain discontinuities of certain types. And conversely it will be our main result that if φ_0 is bounded measurable only at the boundary and if in addition $A\varphi_0$ is Lipschitz continuous then a continuation in the above sense is possible.

The main result is stated in Theorem 3.1. Essentially we will obtain the continuation by use of the elementary solution of the parabolic equation

$$\nabla^2 u = \partial u / \partial t \ . \tag{12}$$

We shall use this for a kind of mollifier. In §§1 and 2 we prove some auxiliary results most of which will be known. In order to keep the paper as self contained as possible most of the facts required have been proved explicitely.

1. Auxiliary results. In this section we will establish some known results which have to be used essentially in the following. Let

$$s^2 = s_1^2 + s_2^2 + \cdots + s_p^2 \tag{1.1}$$

and let the function

$$\Phi(s; t) = \Phi(s_1, \cdots, s_p; t) \tag{1.2}$$

be defined by

$$\Phi(s; t) = (4\pi t)^{-p/2} \exp(-|s|^2/4t) \ . \tag{1.3}$$

It is known that this function $\Phi(s; t)$ is the elementary solution of the parabolic equation

$$\nabla^2 u = \sum_{i=1}^{p} \partial^2 u / \partial s_i^2 = \partial u / \partial t \ . \tag{1.4}$$

First we note

LEMMA 1.1.

$$\int |s|^{2k} e - |s|^{2k} e^{-|s|^2} ds = 2^{-k} \pi^{-p/2} p(p+2)(p+4) \cdots (p+2k-2) \ . \tag{1.5}$$

Here the integral extends over the whole $(s_1, \cdots, s_p)$-space.

The proof of Lemma 1.1 can be obtained by repeated application of Green's formula.

LEMMA 1.2. *Let*

(1.6) $$f(s) = f(s_1, \cdots, s_p)$$

be a (scalar) complex valued bounded measurable function defined and nonnegative for

(1.7) $$-\infty < s_j < \infty, \qquad j = 1, \cdots, p .$$

Let s_0 be any point and let Δ denote the cube

(1.8) $$|s_j - s_j^0| \leq \delta, \qquad j = 1, \cdots, p .$$

Statement. If

(1.9) $$\lim_{\delta \to 0} \delta^{-p} \int_\Delta f(s) ds = 0$$

then

(1.10) $$\lim_{t \to 0} \int_\Delta \Phi(s_0 - s'; t) f(s') ds' = 0$$

the integral in (1.10) being taken over the whole s-space.

Proof. It is obvious that we can restrict ourself to the case $s_0 = 0$. Now, (1.9) being satisfied, let

(1.11) $$\beta(\delta_0) = \sup_{0 < \delta \leqq \delta_0} \left\{ \delta^{-p} \int_\Delta f(s) ds \right\}^{(p+1)^{-1}}$$

and let

(1.12) $$\gamma(\delta) = \delta(\delta + \beta(\delta))$$

$\gamma(\delta)$ is a strictly monotonically increasing function of δ, and $\gamma(0) = 0$. Hence the inverse function $\delta = \delta(\gamma)$ exists in some right neighborhood of $\gamma = 0$ and $\delta(0) = 0$. Also

(1.13) $$\begin{aligned} \gamma^{-p} \int_\Delta f(s) ds &\leqq (\delta + \beta(\delta))^{-p} \beta(\delta)^{p+1} \\ &\leq \beta(\delta) \longrightarrow 0, \delta \longrightarrow 0 . \end{aligned}$$

Hence

(1.14) $$\lim_{\gamma \to 0} \gamma^{-p} \int_\Delta f(s) ds = 0 .$$

Let

(1.15) $$\pi = \delta / \gamma ,$$

then

(1.16) $$\lim_{\gamma \to 0} \tau(\gamma) = \infty .$$

Let $\varDelta'$ be the cube $|s_j| \leq \gamma, j = 1, \cdots, p$. Then by (1.15) $\varDelta$ can be written in the form

$$\varDelta = \tau \varDelta' \tag{1.17}$$

and (1.14) reads

$$\lim_{\gamma \to 0} \gamma^{-p} \int_{\tau \varDelta'} f(s) ds = 0 \ . \tag{1.18}$$

Now for any given $t > 0$ set $\gamma = t^{1/2}$, then

$$\int \Phi(s_0 - s'; t) f(s')\, ds' = \int \Phi(s; t) f(s) ds = \int_{\tau \varDelta'} + \int_{C(\tau \varDelta')} \tag{1.19}$$

where $C(\tau\varDelta')$ denotes the complement of the cube $\tau\varDelta'$ with respect to the whole s-space. But remembering the definition of $\Phi(s; t)$ we obtain for the first integral

$$\leq (4\pi)^{-p/2} \gamma^{-p} \int_{\tau \varDelta'} f(s) ds \tag{1.20}$$

and hence for $t \to 0$, i.e., $\gamma \to 0$ the first integral tends to zero by (1.18). On the other hand $f(s)$ is assumed to be uniformly bounded, hence the second integral can be estimated by

$$\begin{aligned} & c_0 \int_{C(\tau \varDelta')} \Phi(s; t) ds \\ & \leqq c_0 (4\pi)^{-p/2} \gamma^{-p} \left\{ \int_{|\sigma| \geqq \tau\gamma} e^{-\sigma^2/\gamma^2} d\sigma \right\}^p \\ & = c_0 (4\pi)^{-p/2} \left\{ \int_{|\sigma| \geqq \tau} e^{-\sigma^2} d\sigma \right\}^p \ . \end{aligned} \tag{1.21}$$

But by (1.16)

$$\pi = \tau(\gamma) = \tau(t^{1/2}) \tag{1.22}$$

tends to ∞ at $t \to 0$. Therefore the second integral also tends to zero. This proves the lemma.

LEMMA 1.3. *Let $\Phi(s; t)$ be as defined in* (1.3) *and let*

$$\Psi_i(s; t) = \partial / \partial s_i \Phi(s; t) \ . \tag{1.23}$$

Then

$$\int ds \Phi(s - s'; t) \Phi(s - s''; t) = (8\pi t)^{-p/2} \exp(-|s' - s''|^2 / 8t) \tag{1.24}$$

$$\begin{aligned} & \int ds \Psi_i(s - s'; t) \Psi_i(s - s''; t) \\ & = (4t)^{-1} (8\pi t)^{-p/2} (1 - (s_i' - s_i'')^2/4t) \exp(-|s' - s''|^2/8t) \end{aligned} \tag{1.25}$$

both integrals being taken over the whole $(s_1, \cdots, s_p)$-space.

Proof. We only remark that

$$\begin{aligned}(1.26)\qquad &\exp(-|s-s'|^2/4t)\exp(-|s-s''|^2/4t)\\ &=\exp(-|s'-s''|^2 8t)\exp(-|\hat{s}|^2/2t)\end{aligned}$$

where we denote

$$(1.27)\qquad \hat{s} = s - 1/2(s' + s'') .$$

Therefore the integral (1.24) equals to

$$(1.28)\qquad (4\pi t)^{-p}\exp(-|s'-s''|^2/8t)\int\exp(-|\hat{s}|^2/2t)d\hat{s}$$

and clearly

$$(1.29)\qquad \int\exp(-|\hat{s}|^2/2t)d\hat{s} = (2\pi t)^{p/2} .$$

This proves the first formula. For the second formula we note that

$$(1.30)\qquad \Psi_i(s;t) = -(2t)^{-1}(4\pi t)^{-p/2}s_i\exp(-|s|^2/4t) .$$

Now

$$(1.31)\qquad (s_i - s_i')(s_i - s_i'') = \hat{s}_i^2 - 1/4(s_i' - s_i'')^2 .$$

Hence the integral (1.25) gets the form

$$\begin{aligned}(1.32)\qquad &(2t)^{-2}(4\pi t)^{-p}\exp(-|s'-s''|^2/8t)\\ &\times\left\{\int\hat{s}\exp(-|\hat{s}|^2/2t)d\hat{s} - 1/4(s_i'-s_i'')^2\int\exp(-|\hat{s}|^2/2t)d\hat{s}\right\} .\end{aligned}$$

But

$$(1.33)\qquad \int\hat{s}_i^2\exp(-|\hat{s}|^2/2t)d\hat{s} = t(2\pi t)^{p/2} .$$

If we substitute (1.29) and (1.33) into (1.32) then we get

$$(1.34)\qquad = (4t)^{-1}(8\pi t)^{-p/2}(1-(s_i'-s_i'')^2/4t)\exp(-|s'-s''|^2/8t)$$

which completes the proof.

LEMMA 1.4. *Let*

$$(1.35)\qquad \Omega_1(s;t) = (2t)^{-1}(4\pi t)^{-p/2}\exp(-|s|^2/4t)$$

$$(1.36)\qquad \Omega_2(s;t) = |s|^2(2t)^{-2}(4\pi t)^{-p/2}\exp(-|s|^2/4t) .$$

Statement.

$$(1.37)\qquad \int ds\Omega_1(s-s';t)\Omega_1(s-s'';t)\sum_{i=1}^{p}(s_i-s_i')(s_i-s_i'') = -1/2d/dt((8\pi t)^{-p/2}\exp(-|s'-s''|^2/8t))$$

$$\int ds\Omega_2(s-s';t)\Omega_2(s-s'';t)\sum_{i=1}^{p}(s_i-s_i')(s_i-s_i'')$$

$$(1.38)\qquad = -1/2d/dt[(8\pi t)^{-p/2}\{(8t)^{-2}|s'-s''|^4 + p(8t)^{-1}|s'-s''|^2 + 1/4(p+2)(p+4)\}\exp(-|s'-s''|^2/8t)]\ .$$

Proof. We introduce the notation

$$(1.39)\qquad \hat{\sigma} = (2t)^{-1/2}(s-1/2(s'+s'')),\qquad \sigma^* = (8t)^{-1/2}(s'-s'')$$

and we observe that

$$(1.40)\qquad \sum_{i=1}^{p}(s_i-s_i')(s_i-s_i'') = 2t(|\hat{\sigma}|^2 - |\sigma^*|^2)\ .$$

Now if we substitute (1.36) and (1.40) into the integral (1.37) this integral equals

$$(1.41)\qquad (2t)^{-1}(8\pi^2 t)^{-p/2}\exp(-|\sigma^*|^2)\int(|\hat{\sigma}|^2-|\sigma^*|^2)\exp(-|\hat{\sigma}|^2 d\hat{\sigma}) = (8\pi)^{-p/2}(p/4t^{-p/2-1} - 1/16|s'-s''|^2 t^{-p/2-2})\exp(-|s'-s''|^2/8t)$$

Here for the evaluation of

$$(1.42)\qquad \int|\hat{\sigma}|^2\exp(-|\hat{\sigma}|^2)d\hat{\sigma}$$

Lemma 1.1 has been applied. Now (1.41) is equal to the derivative in (1.37) as can be proved by differentiation. Therefore (1.37) is proved. For the second integral we get in a similar way the expression

$$(1.43)\qquad (2t)^{-1}(8\pi^2 t)^{-p/2}\exp(-|\sigma^*|^2) \times\int|\hat{\sigma}-\sigma^*|^2|\hat{\sigma}+\sigma^*|^2(|\hat{\sigma}|^2-|\sigma^*|^2)\exp(-|\hat{\sigma}|^2)d\hat{\sigma}\ .$$

Here we were using that

$$(1.44)\qquad s-s' = (2t)^{1/2}(\hat{\sigma}-\sigma^*),\qquad s-s'' = (2t)^{1/2}(\hat{\sigma}+\sigma^*)\ .$$

We observe that

$$(1.45)\qquad |\hat{\sigma}-\sigma^*|^2|\hat{\sigma}+\sigma^*|^2 = (|\hat{\sigma}|^2+|\sigma^*|^2)^2 - 4(\hat{\sigma}\sigma^*)^2$$

and further that

$$(1.46)\qquad \int(\hat{\sigma}\sigma^*)^2(|\hat{\sigma}|^2-|\sigma^*|^2)\exp(-|\hat{\sigma}|^2)d\hat{\sigma}$$

$$= \sum_{i=1}^{p} \left\{ (\sigma_i^*)^2 \int (\hat\sigma_i)^2 (|\hat\sigma|^2 - |\sigma^*|^2) \exp(-|\hat\sigma|^2) d\hat\sigma \right\}$$
$$= 1/p |\sigma^*|^2 \int |\hat\sigma|^2 (|\hat\sigma|^2 - |\sigma^*|^2) \exp(-|\hat\sigma|^2) d\hat\sigma \ .$$

Here we used that

$$\int \hat\sigma_i \hat\sigma_k (|\hat\sigma|^2 - |\sigma^*|^2) \exp(-|\hat\sigma|_0^2) d\hat\sigma = 0, \ i \neq k \ . \tag{1.47}$$

Substituting (1.45) and (1.46) into (1.43) we get the expression

$$(2t)^{-1}(8\pi^2 t)^{-p/2} e^{-|\sigma^*|^2} \int (|\hat\sigma|^4 + |\sigma^*|^4 + 2(p-2)/p |\hat\sigma|^2 |\sigma^*|^2)(|\hat\sigma|^2 - |\sigma^*|^2) e^{-|\hat\sigma|^2} d\hat\sigma \ . \tag{1.48}$$

Further

$$\begin{aligned} &(|\hat\sigma|^4 + |\sigma^*|^4 + 2(p-2)/p |\hat\sigma|^2 |\sigma^*|^2)(|\hat\sigma|^2 - |\sigma^*|^2) \\ &= |\hat\sigma|^6 + (p-4)/p |\hat\sigma|^4 |\sigma^*|^2 - (p-4)/p |\hat\sigma|^2 |\sigma^*|^4 - |\sigma^*|^6 \ . \end{aligned} \tag{1.49}$$

We substitute this into (1.38) and then use Lemma 1.2 to evaluate the integral, then this integral equals

$$\begin{aligned} &\pi^{-p/2}\{1/8p(p+2)(p+4) \\ &+ 1/4(p+2)(p-4)|\sigma^*|^2 - 1/2(p-4)|\sigma^*|^4 - |\sigma^*|^6\} \ . \end{aligned} \tag{1.50}$$

On the other hand by calculating the derivative (1.38) we get the expression

$$\begin{aligned} &-1/2(8\pi)^{-p/2}\{-1/2(p+4)t^{-p/2-1}|\sigma^*|^4 - 1/2p(p+2)t^{-p/2-1}|\sigma^*|^2 \\ &\qquad - 1/8p(p+2)(p+4)t^{-p/2-1}\} \exp(-|\sigma^*|^2) \\ &-1/(2t)(8\pi t)^{-p/2}\{|\sigma^*|^6 + p|\sigma^*|^4 + 1/4(p+2)(p+4)|\sigma^*|^2\} \exp(-|\sigma^*|^2) \\ &= -(2t)^{-1}(8\pi t)^{-p/2} \exp(-|\sigma^*|^2)\{|\sigma^*|^6 + 1/2(p-4)|\sigma^*|^4 \\ &\qquad -1/4(p+2)(p-4)|\sigma^*|^2 - 1/8p(p+2)(p+4)\} \ . \end{aligned} \tag{1.51}$$

If we substitute (1.50) into (1.49) and then compare the obtained expression with (1.51) we find that both are equal. Therefore formula (1.38) is proved.

2. Lemmata about special integral operators. The following lemma was used earlier by K. O. Friedrichs [2]. It can be considered to be a translation of a theorem about infinite matrices going back to I. Schur [6].

LEMMA 2.1. *Let*

$$X(s; s') = X(s_1, \cdots, s_p; s_1', \cdots, s_1', \cdots, s_p') \tag{2.1}$$

be defined and continuous for $s, s' \in D_0$, D_0 being any region of $(s_1, \cdots, s_p)$-space, and let

$$\gamma = \sup_{s \in D_0} \int_{D_0} |X(s;s')| ds' \tag{2.2}$$

$$\delta = \sup_{t' \in D_0} \int_{D_0} |X(s;s')| ds . \tag{2.3}$$

Statement.

$$\int_{D_0} ds \left| \int_{D_0} X(s, s') u(s') ds' \right|^2 \leq \gamma^\delta \int_{D_0} |u(s)|^2 ds \tag{2.4}$$

holds for every complex valued measurable function $u(s)$ which is squared integrable over D_0.

Proof. By Schwarz' inequality

$$\begin{aligned} \int_{D_0} ds \left| \int_{D_0} X(s;s') u(s') ds' \right|^2 &\leq \int_{D_0} ds \left(\int_{D_0} |X(s;s')| \, |u(s')| \, ds' \right)^2 \\ &\leq \int_{D_0} ds \left\{ \int_{D_0} |X(s;s')| ds' \int_{D_0} |X(s;s')| \, |u(s')|^2 ds' \right\} \\ &\leq \gamma \int_{D_0} |u(s')|^2 \left(\int_{D_0} |X(s;s')| ds \right) ds' \leq \gamma^\delta \int_{D_0} |u(s')|^2 ds' . \end{aligned}$$

Now let $\Phi(s;t)$, $\Psi_i(s;t)$; $\Omega_1(s;t)$, $\Omega_2(s;t)$ be defined as in (1.1), (1.23), (1.35), and (1.36).

LEMMA 2.2.

$$\sum_{i=1}^{p} \iint dsdt \left| \int \Psi_i(s-s';t) u(s') ds' \right|^2 \leq \int |u(s)|^2 ds \tag{2.5}$$

for every $u(s)$ squared integrable over the whole s-space and having a compact carrier. Here the integral $\int dt$ is taken over the interval $0 \leq t \leq 1$, the integrals $\int ds$ and $\int ds'$ are considered to be taken over the whole s-space.

Proof. First of all by Lemma 1.3:

$$\begin{aligned} &\sum_{i=1}^{p} \iint dsdt \left| \int \Psi_i(s-s';t) u(s') ds' \right|^2 \qquad (2.6) \\ &= \lim_{\varepsilon \to 0} \iint ds' ds'' \overline{u(s')} u(s'') \int_\varepsilon^1 dt \sum_{i=1}^{p} \int ds \Psi_i(s-s';t) \Psi_i(s-s'';t) \\ &= \lim_{\varepsilon \to 0} \iint ds' ds'' \overline{u(s')} u(s'') \int_\varepsilon^1 dt (4t)^{-1} (8\pi t)^{-p/2} \\ &\quad \times (p - (4t)^{-1} |s'-s''|^2) \exp(-|s'-s''|^2/8t) . \end{aligned}$$

But as we saw in the proof of Lemma 1.4 (formula (1.51)) this integrand is equal to

$$(2.7) \qquad -1/2d/dt\{(8\pi t)^{-p/2}\exp(-|s'-s''|^2/8t)\}$$

and hence the right hand side equals to

$$\begin{aligned}
&= -1/2\lim_{\varepsilon\to 0}\iint ds'ds''\overline{u(s')}u(s'') \\
&\quad\times\{(8\pi)^{-p/2}\exp(-|s'-s''|^2/8)-(8\pi\varepsilon)^{-p/2}\exp(-|s'-s''|^2/8)\} \\
&\le 1/2\lim_{\varepsilon\to 0}\int ds'\overline{u(s')}\int ds''(8\pi\varepsilon)^{-p/2}\exp(-|s'-s''|^2/8\varepsilon)u(s'') \\
&\le 1/2\lim_{\varepsilon\to 0}\left\{\int|u(s)|^2ds(8\pi\varepsilon)^{-p}\right. \\
&\quad\left.\times\int ds'\left|\int\exp(-|s'-s''|^2/8\varepsilon)u(s'')ds''\right|^2\right\}^{1/2}.
\end{aligned}$$

Here we were using that the kernel $\exp(-|s'-s''|^2/8)$ is positive definite as can be easily seen by Lemma 1.3. Since

$$(2.8) \qquad \int\exp(-|s'-s''|^2/8\varepsilon)ds'=\int\exp(-|s'-s''|^2/8\varepsilon)ds=(8\pi\varepsilon)^{p/2}$$

Lemma 2.1 yields

$$(2.9) \qquad (8\pi\varepsilon)^{-p}\int ds'\left|\int\exp(-|s'-s''|^2/8\varepsilon)u(s''\right|^2\le\int|u(s)|^2ds\,.$$

This completes the proof of Lemma 2.2.

LEMMA 2.3. *Let*

$$(2.10) \qquad \Omega(s;t)=d/dt\Phi(s;t)$$

and let $v(s)$ be Lipschitz continuous over the whole $(s_1,\cdots,s_p)$-space and with compact carrier.

Statement.

$$(2.11) \qquad \iint dsdt\left|\int ds'\Omega(s-s';t)v(s)\right|^2\le p\int\sum_{i=1}^{p}|\partial v/\partial s_i|^2ds\,.$$

Proof. Since $\Phi(s;t)$ is a solution of the parabolic equation (1.29) we get

$$(2.12) \qquad \Omega(s;t)=\sum_{i=1}^{p}\partial/\partial s_i\Psi_i(s;t)$$

and hence by Green's formula

$$\int ds'\Omega(s-s';t)v(s') = \sum_{i=1}^{p}\int \Psi_i(s-s';t)v_i(s')ds' \tag{2.13}$$

where we denote

$$v_i(s) = \partial/\partial s_i(v(s)) . \tag{2.14}$$

Consequently

$$\begin{aligned}\iint dsdt\left|\int ds'\Omega(s-s';t)v(s')\right|^2 &\le p\sum_{i=1}^{p}\iint dsdt\left|\int ds'\Psi_i(s-s';t)v_i(s')\right|^2 \\ &\le p\sum_{i=1}^{p}\left(\sum_{k=1}^{p}\iint dsdt\left|\int ds'\Psi_k(s-s';t)v_i(s')\right|^2\right) \\ &\le p\sum_{i=1}^{p}\int |\partial v/\partial s_i|^2 ds\end{aligned} \tag{2.15}$$

which prove the lemma.

In the following c always denotes a constant not depending on $u(s)$.

LEMMA 2.4.

$$\iint dsdt\left|\int ds'\Omega(s-s';t)(s_i-s_i')u(s')\right|^2 \le c\int |u(s)|^2 ds \tag{2.16}$$

for any arbitrary $u(s)$ with compact carrier and squared integrable over the s-space.

Proof. Clearly

$$\begin{aligned}\Omega(s;t) &= d/dt\Phi(s;t) \\ &= (4\pi t)^{-p/2}(|s|^2/(4t)^2 - p/(2t))\exp(-|s|^2/4t) \\ &= \Omega_2(s;t) - p\Omega_1(s;t) .\end{aligned} \tag{2.17}$$

Hence the integral in (2.16) can be estimated by

$$\begin{aligned}&2\sum_{i=1}^{p}\iint dsdt\left|\int ds'\Omega_2(s-s';t)(s_i-s_i')u(s')\right|^2 \\ &\quad + 2p^2\sum_{i=1}^{p}\iint dsdt\left|\int ds'\Omega_1(s-s';t)(s_i-s_i')u(s')\right|^2 .\end{aligned} \tag{2.18}$$

Now this can be written in the form

$$\begin{aligned}&2\lim_{\varepsilon\to 0}\iint ds'ds''\overline{u(s')}u(s'') \\ &\quad\times\int_\varepsilon^1 dt\sum_{i=1}^{p}\int ds\Omega_2(s-s';t)\Omega_2(s-s'';t)(s_i-s_i')(s_i-s_i'') \\ &+ 2p^2\lim_{\varepsilon\to 0}\iint ds'ds''\overline{u(s')}u(s'')\end{aligned} \tag{2.19}$$

$$\times \int_\varepsilon^1 dt \sum_{i=1}^{p} \int ds\Omega_1(s-s';t)\Omega_1(s-s'';t)(s_i-s_i')(s_i-s_i'') .$$

We apply Lemma 1.4 and this equals

$$\begin{aligned}(2.20) \qquad & -\lim_{\varepsilon\to 0}\int ds'ds''\overline{u(s')}u(s'') \\ & \times \{(8\pi)^{-p/2}\Xi_1(|s-s'|^2/8)-(8\pi\varepsilon)^{-p/2}\Xi_1(|s'-s''|^2/8\varepsilon)\} \\ & \times \exp(-|s'-s''|^2/8\varepsilon)\end{aligned}$$

where $\Xi_1(\alpha)$ means a certain polynomial in α with constant coefficients and of degree two, the coefficients only depending on p. By a treatment similar to the last expression of Lemma 2.3 we get the final statement.

LEMMA 2.5.

$$\begin{aligned}(2.21) \qquad & \iint dsdt\left|\int ds'|\Omega(s-s';t)||s-s'|^{1+\varepsilon}u(s')\right|^2 \\ & \leq c(\varepsilon)\int|u(s)|^2ds\end{aligned}$$

for any positive ε and for any arbitrary $u(s)$ with compact carrier and squared integrable over the whole space, $c(\varepsilon)$ being a constant independent of $u(s)$.

Proof. Clearly it suffices to prove the corresponding inequality with $\Omega(s-s';t)$ replaced by $\Omega_j(s-s';t)$, $j=1,2$. In order to achieve these estimates we again use the notation (1.49) and estimate as follows:

$$\begin{aligned}(2.22) \qquad & \int ds|\Omega_1(s-s';t)||\Omega_1(s-s'';t)|[|s-s'|^2|s-s''|^2]^{(1+\varepsilon)/2} \\ & =(2t)^{-1+\varepsilon}(8\pi^2t)^{-p/2}\exp(-|s'-s''|^2/8t)\int d\hat{\sigma}e^{-|\hat{\sigma}|^2} \\ & \quad\times\{(|\hat{\sigma}|^2+|\sigma^*|^2)^2-4(\hat{\sigma}\sigma^*)^2\}^{(1+\varepsilon)/2} \\ & =(2t)^{-1+\varepsilon}(8\pi^2t)^{-p/2}\exp(-|s'-s''|^2/8t)J(s'-s''|^2/8t)\end{aligned}$$

where

$$\begin{aligned}(2.23) \qquad J(|\sigma^*|^2) & =\int d\hat{\sigma}e^{-|\hat{\sigma}|^2}\{(|\hat{\sigma}|^2+|\sigma^*|^2)^2-4(\hat{\sigma}\sigma^*)^2\}^{(1+\varepsilon)/2} \\ & \leq\int d\hat{\sigma}\exp(-|\hat{\sigma}|^2)\{|\hat{\sigma}|^2+|\sigma^*|^2\}^{(1+\varepsilon)} \\ & \leq 2^\varepsilon\int d\hat{\sigma}\exp(-|\hat{\sigma}|^2)|\hat{\sigma}|^{2+2\varepsilon}+2^\varepsilon|\sigma^*|^{2+2\varepsilon}\int d\hat{\sigma}\exp(-|\hat{\sigma}|^2) \\ & \leq\gamma_1(\varepsilon)t^{-1+\varepsilon}(8\pi t)^{-p/2}\exp(-|s'-s''|^2 8/t) \\ & \quad+\gamma_2(\varepsilon)t^{-1+\varepsilon}(8\pi t)^{-p/2}[|s'+s''|^2/8t]^{1+\varepsilon}\exp(-|s'-s''|^2/8t) .\end{aligned}$$

Here Hoelders inequality has been employed. Hence (2.22) can be estimated as follows:

$$\iint dsdt\left|\int ds'\left|\Omega_1(s-s';t)\right||s-s'|^{1+\varepsilon}u(s')\right|^2$$
$$+\gamma_1(\varepsilon)\int_0^1 dtt^{\varepsilon-1}\iint ds'ds''\overline{u(s')}u(s'')(8\pi t)^{-p/2}\exp(-|s'-s''|^2/8t)$$
$$+\gamma_2(\varepsilon)\int_0^1 dtt^{\varepsilon-1}\iint ds'ds''\overline{u(s')}u(s'')$$
$$\times(8\pi t)^{-p/2}(|s'-s''|^2/8t)^{1+\varepsilon}\exp(-|s'-s''|^2/8t)$$
$$\leq\gamma(\varepsilon)\int_0^1 dtt^{\varepsilon-1}\int|u|^2ds=\gamma(\varepsilon)\varepsilon^{-1}\int|u(s)|^2ds\ .$$

Here again Lemma 2.1 and Lemma 1.1 were employed. A quite analogous argument is possible for $\Omega_2(s-s';t)$; therefore Lemma 2.5 is proved.

LEMMA 2.6.

$$\iint t^2dsdt\left|\int\Omega(s-s';t)u(s')ds'\right|^2\leq c\int|u(s)|^2ds \tag{2.25}$$

for arbitrary $u(s)$ with compact carrier squared integrable over the whole s-space.

Proof. Again it suffices to prove this inequality for Ω replaced by Ω_2 and Ω_2. Now

$$\begin{aligned}&\int ds\Omega_1(s-s';t)\Omega_1(s-s'';t)\\&\quad=(2t)^{-2}(8\pi^2t)^{-p/2}\exp(-|s'-s''|^2/8t)\int d\hat{\sigma}\exp(-|\hat{\sigma}|^2)\\&\quad=(2t)^{-2}(8\pi t)^{p/2}\exp(-|s'-s''|^2/8t)\ .\end{aligned}\tag{2.26}$$

Hence by Lemma 2.1:

$$\begin{aligned}&\iint ds'ds''\overline{u(s')}u(s'')\int ds\Omega_1(s-s';t)\Omega_1(s-s'';t)\\&\quad\leq(2t)^{-2}\int|u(s)|^2ds\ .\end{aligned}\tag{2.27}$$

Consequently

$$\begin{aligned}&\iint t^2dsdt\left|\int\Omega_1(s-s';t)u(s')ds'\right|^2\\&\quad\leq 1/4\int|u(s)|^2ds\ .\end{aligned}\tag{2.28}$$

Again a similar argument proves the corresponding inequality for Ω_2; therefore Lemma 2.6 is proved.

We finally use the preceding lemmata to establish

LEMMA 2.7. *Let*

$$A(s;t)=((a_{ik}(s;t))) \tag{2.29}$$

be an $m\times m$*-matrix with coefficients* $a_{ik}(s;t)$ *having uniformly Hoelder continuous and uniformly bounded first partial derivatives in the domain*

$$D_0=\{s_1,\cdots,s_p;t\ni-\infty<s_k<+\infty,k=1,\cdots,p;0<t<1\}\ . \tag{2.30}$$

Let

$$u(s)=(u_1(s),\cdots,u_m(s)) \tag{2.31}$$

be an m-component vector function having a compact carrier and being squared integrable over the whole $(s_1,\cdots,s_p)$-space. Let the vector function

$$A(s;0)u(s)=v(s) \tag{2.32}$$

be Lipschitz continuous over the whole $(s_1,\cdots,s_p)$-space.

Statement. There exist two constants c_1, c_2 which are independent of $u(s)$ such that

$$\iint dsdt\left|A(s;t)\int ds'\Omega(s-s';t)u(s')\right|^2 \le c_1\int|u(s)|^2ds+c_2\sum_{i=1}^p\int|\partial v/\partial s_i|^2ds\ . \tag{2.33}$$

Proof. We decompose as follows:

$$\begin{aligned}A(s;t)\int ds'\Omega(s-s';t)u(s') &= \int\Omega(s-s';t)v(s')ds'\\ &+(A(s;t)-A(s;0))\int\Omega(s-s';t)u(s')ds'\\ &+\sum_{i=1}^p\int\Omega(s-s';t)(s_i-s_i')u_i(s')ds'\\ &+\int\Omega(s-s';t)[A(s;0)-A(s';0)\\ &-\sum_{i=1}^p(s_i-s_i')\partial/\partial s_i'A(s';0)]u(s')ds'\end{aligned} \tag{2.34}$$

where

$$v(s) = A(s;0)u(s),\ u_i(s) = [\partial/\partial s_i(A(s;0))]u(s)\ . \tag{2.35}$$

By our assumption for $A(s; t)$ we get

$$|(A(s;t) - A(s;0))w| \leq ct|w| \tag{2.36}$$

and

$$\begin{aligned} &|[A(s;0) - A(s';0) - \sum_{i=1}^{p}(s_i - s_i')\partial/\partial s_i(A(s';0))]u(s')| \\ &\leq c|s-s'|^{1+\varepsilon}|u(s')|\ . \end{aligned} \tag{2.37}$$

Therefore we can use the Lemmata 2.3, 2.4, 2.5, and 2.6 respectively to estimate the integrals in (2.33) for the succeeding terms in (2.34) by either $c\int|u(s)|^2\,ds$ or $\int|\partial v/\partial s_i|^2 ds$. Hence Lemma 2.7 is proved.

LEMMA 2.8. *Let $u(s)$ be a bounded measurable m-component vector function defined in the whole s-space and let it have a compact carrier. Further, with the notations of Lemma 2.7, let*

$$v(s) = A(s;0)u(s) \tag{2.38}$$

be Lipschitz continuous over the whole s-space.

Let

$$u(s;t) = \int \Phi(s-s';t)u(s')ds'\ . \tag{2.39}$$

Then

$$\lim_{t\to 0} u(s;t) = u(s) \text{ almost everywhere} \tag{2.40}$$

and

$$v(s;t) = A(s;t)u(s;t) \tag{2.41}$$

is continuous all over in the domain D_0 defined in (2.30) and its boundary.

Proof. Let $\varepsilon > 0$ be given. Since $u(s)$ is bounded and measurable, by Lusin's theorem a measurable set E_ε of p-dimensional measure $m(E_\varepsilon)$ less than ε exists such that $u(s)$ is continuous on the complement $C(E_\varepsilon)$ of E_ε with respect to the s-space. If $\chi(s)$ denotes the characteristic function of E_ε and if Δ denotes the cube with sides 2δ defined in (1.8), then by well known facts

$$\lim_{\delta\to 0}\delta^{-p}\int_\Delta \chi(s)ds = 0 \tag{2.42}$$

for every $s_0 \in C(E_\varepsilon + N_\varepsilon)$ where N_ε denotes a certain nullset. We will show that for every $s_0 \in C(E_\varepsilon + E_\varepsilon)$ relation (2.40) holds. This will prove the first statement of the lemma, because then obviously it is possible to construct a monotonically decreasing sequence of sets which converges toward a nullset and such that after exempting any set of the sequence the statement (2.40) holds.

Now, $s_0 \in C(N_\varepsilon + E_\varepsilon)$ being given, decompose as follows:

$$\int \Phi(s_0 - s'; t)u(s')ds' = \int_{C(E_\varepsilon)\cap\Delta_0} + \int_{E_\varepsilon\cap\Delta_0} + \int_{C(\Delta_0)} \tag{2.43}$$

where Δ_0 denotes the cube (1.8) with side $\delta = \delta_0$. Then

$$\int_{C(E_\varepsilon)\cap\Delta_0} = \mu_{\Delta_0}\int_{C(E_\varepsilon)\cap\Delta_0} \Phi(s_0 - s'; t)ds' \tag{2.44}$$

where μ_{Δ_0} denotes a mean value of $u(s)$ in the cube Δ_0.

But since u is continuous in $C(E_\varepsilon) \cap \Delta$ it follows that

$$|\mu_{\Delta_0} - u(s_0)| < \varepsilon' \tag{2.45}$$

if $\delta_0 > 0$ is sufficiently small. Also

$$\int_{C(E_\varepsilon)\cap\Delta_0} \Phi(s_0 - s'; t)ds' \le \int \Phi(s_0 - s'; t)ds' = 1 . \tag{2.46}$$

Consequently, using (2.44) and (2.46) we get

$$\begin{aligned} &\left|\int_{C(E_\varepsilon)\cap\Delta_0} \Phi(s_0 - s'; t)u(s')ds' - u(s_0)\right| \\ &\le |\mu_{\Delta_0} - u(s_0)| + c\int_{\Delta_0} \Phi(s_0 - s'; t)\chi(s')ds' \\ &+ c\int_{C(\Delta_0)} \Phi(s - s'; t)ds' \end{aligned} \tag{2.47}$$

with $c = \sup|u(s)|$. Finally for the second and third integral in (2.43) we obtain estimates

$$\left|\int_{E_\varepsilon\cap\Delta_0}\right| \le c\int_{\Delta_0} \Phi(s - s'; t)\chi(s')ds' \tag{2.48}$$

and

$$\left|\int_{C(\Delta_0)}\right| \le c\int_{C(\Delta_0)} \Phi(s - s'; t)ds' . \tag{2.49}$$

Hence by (2.43), (2.47), (2.48), and (2.49)

$$\begin{aligned} &\left|\int \Phi(s - s';\ t)u(s')ds' - u(s_0)\right| \\ &\le |\mu_{\Delta_0} - u(s_0)|\, 2c\int_{\Delta_0} \Phi(s - s';\ t)\chi(s')ds' \\ &+ 2c\int_{C(\Delta_0)} \Phi(s - s';\ t)\, ds' . \end{aligned} \tag{2.50}$$

Choosing first δ_0 sufficiently small the first term can be made arbitrarily small; then keeping δ_0 fixed by Lemma 1.2 and (2.42) the second term also can be made arbitraily small by choosing t small. Also the last term for fixed δ_0 becomes arbitrarily small if t tends to zero. Hence formula (2.40) is proved.

In order to prove the continuity of (2.41) we decompose

$$v(s;\ t) = \int \Phi(s - s';\ t) v(s')\, ds' + \int \Phi(s - s';\ t)\,(A(s;\ t) - A(s';\ 0)) u(s')\, ds' \,. \tag{2.51}$$

Since $v(s)$ is assumed to be Lipschitz continuous, the first term obuiously is a continuous function in D_0. The second term is also continuous for every $t > 0$. But since $u(s)$ is assumed to be bounded we get

$$\begin{aligned} &\int \Phi(s - s';\ t)\,(A(s;\ t) - A(s';\ 0)) u(s')\, ds' \\ &\leq ct \int \Phi(s - s';\ t)\, ds' + c' \int \Phi(s - s';\ t)\, |\, s - s' \,|\, ds' \\ &= c''t + c'/\, t^{1/2} \longrightarrow 0,\ t \longrightarrow 0\ . \end{aligned} \tag{2.52}$$

Therefore the continuity is also proved for $t = 0$. This proves the lemma.

3. A continuation theorem. Let D be a bounded domain of the $(x_1, \cdots, x_n)$-space with a twice continuously differentiable boundary Γ which consits of a finite number of simple nonintersecting hyper surfaces. More specifically we assume that the boundary Γ has second derivatives satisfying a uniform Hoelder condition. Let

$$a_i(x) = ((a^i_{jl}(x))),\ i = 1, \cdots, n,\ b(x) = ((b_{ik}(x))) \tag{3.1}$$

be $m \times m$-matrices with complex coefficients defined in $D + \Gamma$. Let $a_i(x)$ be hermitian symmetric and its coefficients be continuously differentiable in $D + \Gamma$ and let the derivatives satisfy a uniform Hoelder condition in $D + \Gamma$. Let $b(x)$ have continuous coefficients in $D + \Gamma$. Let $A(x)$, $x \in D + \Gamma$ be any hermitian symmetric $m \times m$-matrix having continuously differentiable coefficients in $D + \Gamma$ and such that

$$A(x) = \sum_{i=1}^{n} a_i(x)\, \nu_i(x),\ x \in \Gamma \tag{3.2}$$

where $\nu(x) = (\nu_1(x), \cdots, \nu_n(x))$ denotes the exterior normal on Γ. We define the differential operator L_1 in $\mathfrak{D}_{L_1}$ by

$$L_1 u = \sum_{i=1}^{n} a_i(x) \partial u / \partial x_i + b(x) u(x) \tag{3.3}$$

for complex valued m-component vector functions

$$u(x) = (u_1(x), \cdots, u_n(x)) \tag{3.4}$$

where $\mathfrak{D}_{L_1}$ is the space of all $u(x)$ satisfying the following conditions:

(a) u, $\partial u/\partial x_i$, $i = 1, \cdots, n$, continuous in D.

(b) $u(x)$ uniformly bounded in D.

(c) $\lim_{\varepsilon\to 0} u(x - \varepsilon\nu) = u(x)$ for every $x \in \Gamma$, except possibily on an n-1-dimensional null set.

(d) $v(x) = A(x)\, u(x)$ is continuous on $D + \Gamma$

(e) $\int_D |L_1 u|^2 dx < \infty$.

We prove the following

THEOREM 3.1. *Let $u_0(x)$ be an m-component vector function which is defined measurable and bounded on Γ and for which*

$$v_0(x) = A(x)u_0(x) \tag{3.5}$$

is Lipschitz continuous on Γ.

Then there exists a function $u(x) \in \mathfrak{D}_{L_1}$ such that

$$u(x) = u_0(x) \text{ on } \Gamma . \tag{3.6}$$

Proof. We consider any arbitray point $x_0 \in \Gamma$. There is a certain neighborhood

$$U_{x_0} = \{x \ni |x - x_0| \leq \varepsilon\} \tag{3.7}$$

which can be mapped by a twice Hoelder continuously differentiable one to one mapping

$$y = y(x) \tag{3.8}$$

onto a bounded region in the $(y_1, \cdots, y_n)$-space in such a way that the point x_0 goes into the origin $y = (0, \cdots, 0)$, the intersection

$$\Gamma_x = \Gamma_0 \cap U_{x_0} \tag{3.9}$$

into a certain neighborhood of $(0, \cdots, 0)$ on the hyperplane $y_1 = 0$, and the intersection

$$D_{x_0} = (D + \Gamma) \cap U_{x_0} \tag{3.10}$$

into a certain half neighborhood of $(0, \cdots, 0)$ satisfying $y_1 \geq 0$. We also can assume that the Jacobian does not vanish.

$$\det ((\partial y_i/\partial x_k)) \neq 0, \ y \in D_{x_0} + \Gamma_{x_0} . \tag{3.11}$$

The image $y(D_{x_0})$ of D_{x_0} under this transformation contains a cube of the type

$$\mathfrak{Q}_{x_0} = \{y \in 0 \leq y_1 \leq \eta(x_0),\ |y_\nu| \leq 1/2\eta(x_0),\ \nu = 2, \cdots, n\} . \tag{3.12}$$

We denote the intersection of $\mathfrak{Q}_{x_0}$ with the hyperplane $y_1 = 0$ by $\mathfrak{q}_{x_0}$ and we set

$$x(\mathfrak{Q}_{x_0}) = \mathfrak{Q}'_{x_0},\ x(\mathfrak{q}_{x_0}) = \mathfrak{q}'_{x_0} \tag{3.13}$$

where $x = x(y)$ denotes the inverse transformation of (3.8). There is a hypersphere

$$U'_{x_0} = \{x \ni |x - x_0| \leq \eta'(x_0)\} \tag{3.14}$$

such that

$$D'_{x_0} = D_{x_0} \cap U'_{x_0} \subset \mathfrak{Q}'_{x_0} \tag{3.15}$$

and such that the same inclusion still holds for $\eta'(x_0)$ being replaced by a somewhat larger number.

This construction can be employed for every $x_0 \in \Gamma$. Since Γ is a bounded closed set, the whole Γ can be covered by the interior points of a finite number of spheres

$$U'_{x_\nu},\ \nu = 1, \cdots, N. \tag{3.16}$$

There is a decomposition of the identity, i.e., a set of N functions

$$\varphi_\nu(x),\ \nu = 1, \cdots, N \tag{3.17}$$

being defined and infinitely differentiable in the whole $(x_1, \cdots, x_n)$-space and such that

$$\varphi_\nu(x) = 0 \text{ outside of } U'_{x_\nu} \tag{3.18}$$

and

$$\sum_{\nu=1}^{N} \varphi_\nu(x) = 1 \text{ on } \Gamma . \tag{3.19}$$

Now any vector function $u_0(x)$ being given which satisfies the conditions of the Theorem 3.1, define

$$u_{\nu,0}(x) = u_0(x)\varphi_\nu(x),\ x \in \Gamma,\ \nu = 1, \cdots, N . \tag{3.20}$$

Clearly $u_{\nu,0}(x)$ also satisfies the assumptions of Theorem 3.1, especially because

$$A(x)u_{\nu,0}(x) = (A(x)u_0(x))\varphi_\nu(x) . \tag{3.21}$$

We will prove that every $u_{\nu,0}(x)$ can be continued to a function $u_\nu(x) \in \mathfrak{D}_{L_1}$

in the sense of the assertion. This obviously will prove Theorem 3.1, because the sum of all $u_\nu(x)$ will be the desired continuation of $u_0(x)$.

Now, if we apply the mapping just defined in each particular neighborhood D_{x_ν} then the vector function $u_{\nu,0}(x)$ will be transformed into a certain function

$$w_{\nu,0}(y) = u_{\nu,0}(x(y)) \tag{3.22}$$

defined and measurable on $y(\Gamma_{x_\nu})$ which contains the cube $\mathfrak{q}_{x_\nu}$. Since by definition $u_{\nu,0}(x) = 0$ outside of D'_{x_ν} and since

$$y(D'_{x_\nu}) \subset \mathfrak{Q}_{x_\nu} \tag{3.23}$$

holds, the function $w_{\nu,0}(y)$ is defined for $y \in \mathfrak{q}_{x_\nu}$ and has its carrier in the interior of this n-1-dimensional cube. We can consider $w_{\nu,0}(y)$ as being defined on the whole hyperplane $y_1 = 0$ by setting it equal to zero outside of $\mathfrak{q}_{x_\nu}$. We would like to apply the various lemmata of § 2. In order to do this we first transform the operator L_1 to the new variables y.

$$L_1 = \sum_{i=1}^{n} \tilde{a}_i(y)\partial/\partial y_i + \tilde{b}(y) \ , y \in y(D_{x_\nu}) \tag{3.24}$$

where

$$\tilde{a}_i(y) = \sum_{k=1}^{n} \partial y_i/\partial x_k a_k(x(y)); \ \tilde{b}(y) = b(x(y)) \ . \tag{3.25}$$

Further we define

$$\tilde{A}(y) = A(x(y)), \ y \in y(D_{x_\nu}) \ , \tag{3.26}$$

Clearly it is possible to continue the matrix $\tilde{A}(y)$ to a matrix function being defined, bounded and continuously differentiable on the whole semispace

$$y_1 \geq 0, \ -\infty < y_\nu < +\infty, \ \nu = 2, \cdots, n \ ; \tag{3.27}$$

its first derivatives satisfying a uniform Hoelder condition in every compact subregion. Now we remark that for

$$y_1 = t, \ y_2 = s_1, \ y_3 = s_2, \cdots, y_n = s_p; \ p = n-1 \tag{3.28}$$

the functions $w_{\nu,0}(y)$ and $\tilde{A}(y)$ satisfy every assumption necessary for application of Lemma 2.2, Lemma 2.7, and Lemma 2.8. Hence the function

$$w_\nu(y) = \int \Phi(s-s'; \ y_1) w_{\nu,0}(s')\, ds' \tag{3.29}$$

satisfies the following conditions:

(α) w_ν, $\partial w_\nu/\partial y_i$ continuous for $y_1 > 0$.
(β) w_ν uniformly bounded for $y_1 \geq 0$.
(γ) $\lim_{\varepsilon\to 0} w_\nu(y - \varepsilon z)$ exists for every y with $y_1 = 0$ and every vector $z_1 = 1$, $z_j = 0$, $j = 2, \cdots, n$ with the possible exemption of a set of n-1-dimensional measure zero which is contained in $\mathfrak{q}_{x_\nu}$.
(δ) $v_\nu(y) = \tilde{A}(y)w_\nu(y)$ is continuous for $y_1 \geq 0$.
(ε)

$$\int_{y_1\geq 0}\left\{| w_\nu(y) |^2 + | \tilde{A}(y)\partial w_\nu/\partial y_1 |^2 + \sum^n | \partial w_\nu/\partial y_i |^2\right\}dy < \infty . \tag{3.30}$$

Finally take any infinitely differentiable function $\tilde{\varphi}_\nu(y)$ being $= 1$ on $y(D'_{x_\nu})$ and having its carrier in $y(D_{x_\nu})$ and take

$$\tilde{w}_\nu(y) = \tilde{\varphi}_\nu(y)w(y) . \tag{3.31}$$

Clearly $\tilde{w}_\nu(y)$ also has the properties (α), $\cdots$, (ε). Transform this function back to the old variables and continue it zero outside of $D_{x_\nu}(x)$. Call the new function $u_\nu(x)$. Then it is clear that

$$u_\nu(x) = u_{\nu,0}(x) \text{ on } \Gamma . \tag{3.32}$$

Also $u_\nu(x)$ satisfies the conditions (a), (b), (c), and (d). Since

$$| L_1 u_\nu |^2 \leq c\left[| \tilde{A}(y)\partial u_\nu/\partial y_1 |^2 + \sum_{i=2}^n | \partial u_\nu/\partial y_i |^2 + | u_\nu |^2\right] \tag{3.33}$$

(3.30) yields the condition (e) too. Hence $u_\nu(x)$ is the desired continuation and Theorem 3.1 is proved.

4. A counterexample. Let D be the unit circle $x_1^2 + x_2^2 < 1$ and accordingly Γ be the periphery of the unit circle $x_1^2 + x_2^2 = 1$. In D we consider the operator defined in formula (5) of the introduction

$$L_1 = a_1\partial/\partial x_1 + a_2\partial/\partial x_2 \tag{4.1}$$

with

$$a_1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \tag{4.2}$$

Then the equation

$$L_1 u = f \tag{4.3}$$

for the 2-component vector functions

$$u = \{u_1, u_2\}, \ f = \{f_1, f_2\} \tag{4.4}$$

defined in $D + \Gamma$ is equivalent to the system

$$(4.5)\qquad -\partial u_1/\partial x_1 + \partial u_2/\partial x_2 = f_1$$
$$\partial u_1/\partial x_2 + \partial u_2/\partial x_1 = f_2\ .$$

Hence for real valued u_1, u_2 we get

$$(4.6)\quad \int_D (f_1^2 + f_2^2)\,dx = \int_D (\partial u_1/\partial x_1 - \partial u_2/\partial x_2)^2 + (\partial u_1/\partial x_2 + \partial u_2/\partial x_1)^2\,dx$$
$$= \int_D [(\partial u_1/\partial x_1)^2 + (\partial u_1/\partial x_2)^2 + (\partial u_2/\partial x_1)^2 + (\partial u_2/\partial x_2)^2]\,dx$$
$$+ 2\int_D (\partial u_1/\partial x_2 \partial u_2/\partial x_1 - \partial u_1/\partial x_1 \partial u_2/\partial x_2)\,dx\ .$$

Now, assuming u twice continuously differentiable in $D + \Gamma$ we can apply Green's formula to the last integral:

$$(4.7)\qquad \int_D (\partial u_1/\partial x_2 \partial u_2/\partial x_1 - \partial u_1/\partial x_1 \partial u_2/\partial x_2)\,dx$$
$$= \int_\Gamma u_1(x_2 \partial u_2/\partial x_1 - x_1 \partial u_2/\partial u_2/\partial x_2)\,d\sigma\ .$$

Hence the last integral in (4.6) is equal to

$$(4.8)\qquad 2\int_\Gamma u_1(x_2\partial u_2/\partial x_1 - x_1 \partial u_2/\partial x_2)\,d\sigma = -\,2\int_0^{2\pi} u_1 \partial u_2/\partial\vartheta\, d\vartheta$$

where

$$(4.9)\qquad \vartheta = \text{arc } tg\, x_2/x_1\ .$$

Now we impose on u the condition

$$(4.10)\qquad u_1 \sin\vartheta/2 + u_2\cos\vartheta/2 = 0\ .$$

Then

$$(4.11)\qquad -2\int_0^{2\pi} u_1\partial u_2/\partial\vartheta d\vartheta = \int_0^{2\pi} [\partial/\partial\vartheta(u_2^2)]\cot\vartheta/2$$
$$= -\int_0^{2\pi} u_2^2 \partial/\partial\vartheta(\cot\vartheta/2)\,d\vartheta = 1/2\int_0^{2\pi} u_2^2 \sin^{-2}\vartheta/2\,d\vartheta\ .$$

This integration by parts is legitimate because the condition (4.10) implies $u_2 = 0$ at $\vartheta = 0,\ 2\pi$. Since u is supposed to have continuous first derivatives it follows that $u_2^2\sin^{-2}\vartheta/2$ remains bounded also for $\vartheta = 0,\ 2\pi$. Consequently

$$(4.12)\qquad \int_D |\,L_1 u\,|^2\,dx = \int_D |\,f\,|^2\,dx$$
$$= \int_D [(\partial u_1/\partial x_1)^2 + (\partial u_1/\partial x_2)^2 + (\partial u_2/\partial x_1)^2 + (\partial u_2/\partial x_2)^2]\,dx$$
$$+ 1/2\int_0^{2\pi} u_2^2 \sin^{-2}\vartheta/2\,d\vartheta\ .$$

Since the last integral is nonnegative we obtain

$$\int_D |L_1 u|^2 dx \geq \int_D [(\partial u_1/\partial x_1)^2 + (\partial u_1/\partial x_2)^2 + (\partial u_2/\partial x_1)^2 + (\partial u_2/\partial x_2)^2]\, dx \;. \tag{4.13}$$

Next assume $\varphi = \{\varphi_1, \varphi_2\}$ to be some function satisfying the conditions (a), (b), (c), and (e), of Theorem 3.1 applied to the special operator L_1 defined in (4.1). Also assume that on the boundary Γ:

$$\varphi_1 = \alpha(\vartheta)\cos\vartheta/2,\ \ \varphi_2 = -\ \alpha(\vartheta)\sin\vartheta/2,\ \ 0 \leq \vartheta \leq 2\pi \;. \tag{4.14}$$

Let $\alpha(\vartheta)$ be real valued and piecewise continuous but not continuous. Then we will show that this leads to a contradiction.

First of all the vector function φ can be assumed to be real valued in $D + \Gamma$ because any complex valued such φ being given, $1/2(\varphi + \bar{\varphi})$ would satisfy the same conditions as φ and would be real valued.

Now, if L_- in $\mathfrak{D}_{L_-}$ denotes the restriction of the operator L_1 in $\mathfrak{D}_{L_1}$ to the space $\mathfrak{D}_{L_-}$ of all functions twice continuously differentiable in $D + \Gamma$ and satisfying the boundary conditions (4.10) then we obtain a dissipative operator in the sense of R. S. Phillips [4], which is characterized by local boundary conditions. For the matrix

$$A = \sum_{i=1}^{2} a_i \nu_i = a_1 \cos\vartheta + a_2 \sin\vartheta \tag{4.15}$$

we get the representation

$$A(\vartheta) = \begin{pmatrix} -\cos\vartheta & \sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix} = \begin{pmatrix} \sin^2\vartheta/2, & \sin\vartheta/2\cos\vartheta/2 \\ \sin\vartheta/2\cos\vartheta/2, & \cos^2\vartheta/2 \end{pmatrix} - \begin{pmatrix} \cos^2\vartheta/2, & -\sin\vartheta/2\cos\vartheta/2 \\ -\sin\vartheta/2\cos\vartheta/2, & \sin^2\vartheta/2 \end{pmatrix} \tag{4.16}$$

and it is easy to see that the two matrices of this last decomposition are identical with the matrices P_0 and N_0 respectively which project orthogonally onto the spaces of all eigenvectors corresponding to the eigenvalues $+1$ and -1 respectively. The boundary condition

$$P_0 u = 0 \ \text{ on } \Gamma \tag{4.17}$$

obviously is equivalent to the condition (4.10). Hence the inner product $\bar{u}Au$ is ≤ 0 for all u satisfying the condition (4.17) (or (4.10)). Hence

$$Q(u,\ u) = 2Re\int_D \bar{u} L_1 u\, dx = \int_\Gamma \bar{u} A u\, d\sigma \leq 0 \;, \tag{4.18}$$

which proves that L_- in $\mathfrak{D}_{L_-}$ is dissipative. On the other hand in the sense of K. O. Kriedrichs [3] this boundary condition is "admissible", because

$$A = P_0 - N_0,\ P_0 \geq 0,\ N_0 \geq 0\ . \tag{4.19}$$

Also the rank of A is constantly equal to two.

Hence if L_-^* in $\mathfrak{D}_{L_-^*}$ denotes the adjoint of L_- in $\mathfrak{D}_{L_-}$ with respect to the inner product

$$\langle u,\ v\rangle = \int_D \bar{u}v\, dx \tag{4.20}$$

and if L_+ in $\mathfrak{D}_{L_+}$ denotes the operator analogous to L_- in $\mathfrak{D}_{L_-}$ with the boundary condition (4.17) replaced by $N_0 u = 0,\ x \in \Gamma$, then

$$L_-^{**} = L_+^*\ . \tag{4.21}$$

But φ is a function of L_+^* because from the conditions (a), (b), (c) and (e) it follows immediately that

$$\langle \varphi,\ Lu\rangle + \langle L\varphi,\ u\rangle = \int_\Gamma \bar{\varphi} Au\, d\sigma = 0 \tag{4.22}$$

for all $u \in \mathfrak{D}_{L_+}$. Hence (4.21) implies

$$\varphi \in \mathfrak{D}_{L_-^{**}}\ . \tag{4.23}$$

Therefore a sequence $\varphi^n \in \mathfrak{D}_{L_-}$ exists such that

$$\langle \varphi^n - \varphi,\ \varphi^n - \varphi\rangle \longrightarrow 0,\ n \longrightarrow \infty \tag{4.24}$$

$$\langle L_1(\varphi^n - \varphi),\ L_1(\varphi^n - \varphi)\rangle \longrightarrow 0,\ n \longrightarrow \infty\ . \tag{4.25}$$

Now (4.25) implies

$$\langle L_1(\varphi^n - \varphi^m),\ L_1(\varphi^n - \varphi^m)\rangle \longrightarrow 0,\ n, m \longrightarrow \infty\ . \tag{4.26}$$

Let

$$\varphi^{n,m} = \varphi^n - \varphi^m \tag{4.27}$$

then (4.13) yields

$$\langle \partial\varphi^{nm}/\partial x_1,\ \partial\varphi^{nm}/\partial x_1\rangle + \langle \partial\varphi^{nm}/\partial x_2,\ \partial\varphi^{nm}/\partial x_2\rangle \longrightarrow 0, n, \mathrm{m} \to \infty\ . \tag{4.28}$$

Hence $\partial\varphi^n/\partial x_1$, $\partial\varphi^n/\partial x_2$ converges in the square mean. Let

$$\partial\varphi^n/\partial x_1 \longrightarrow \psi,\ n \longrightarrow \infty\ , \tag{4.29}$$

and let u be any vector function continuously differentiable in $D + \Gamma$ and vanishing outside of some circle $|\,x\,| \leq r < 1$. Then

$$\langle \partial\varphi^n/\partial x_1,\ u\rangle = -\langle \varphi^n,\ \partial u/\partial x_1\rangle\ . \tag{4.30}$$

For $n \to \infty$ we get

$$\langle \varphi, u \rangle = - \langle \varphi, \partial u / \partial x_1 \rangle . \tag{4.31}$$

But φ is continuously differentiable for $|x| < 1$. Hence, using the special properties of u, we get

$$\langle \psi, u \rangle = - \langle \varphi, \partial u / \partial x_1 \rangle = \langle \partial \varphi / \partial x_1, u \rangle . \tag{4.32}$$

Or

$$\langle \psi - \partial \varphi / \partial x_1, u \rangle = 0 \tag{4.33}$$

for all u with the above properties. But the set of all such u is dense in the space L_2; hence

$$\psi = \partial \varphi / \partial x_1 . \tag{4.34}$$

In the same manner we obtain the relation

$$\partial \varphi^n / \partial x_2 \longrightarrow \partial \varphi / \partial x_2 . \tag{4.35}$$

Hence the derivatives $\partial \varphi / \partial x_1$, $\partial \varphi / \partial x_2$ are squared integrable and the Dirichlet-integral of φ exists.

But it is a well known fact that a function φ with the properties (a), (b), (c) which is piecewise continuous on the periphery of the unit circle and has a jump, cannot have the Dirichlet integral existing.

References

1. H. O. Cordes, *On maximal first order partial differential operators*; to appear.

2. K. O. Friedrichs, *The identity of weak and strong solutions of differential operators*, Trans. Amer. Math. Soc., **55** (1944), 132–151.

3. K. O. Friedrichs, *Symmetric positive linear differential equations*, *Dissipative* Domm. Pure Appl. Nath., 11, (1958), 333-418.

4. R. R. Phillips, *Dissipative operators and hyperbolic systems of partial differential operators*; to appear in Pacific. J. Math,

5. I. Schur, *Bemerkungen zur Theorie der beschrankten Billinearformen mit unendlich vielen Veranderlichen*; J. f. d. reine u. angew. Math., **140**, (1911), 1–28; especially page 6.

University of California
Berkeley, California

n-PARAMETER FAMILIES AND BEST APPROXIMATION

PHILIP C. CURTIS, JR.

1. Introduction. Let $f(x)$ be a real valued continuous function defined on a closed finite interval and let F be a class of approximating functions for f. Suppose there exists a function $g_0 \in F$ such that $||f - g_0|| = \inf_{g \in F} ||f - g||$ where $||f|| \equiv \sup_{x \in [a,b]} |f(x)|$. The problem of characterizing g_0 and giving conditions that it be unique is classical and has received attention from many authors. The well-known results for polynomials were generalized by Bernstein [2] to "Chebyshev" systems. Later Motzkin [10] and Tornheim [15] further extended these theorems to not necessarily linear families of continuous functions. The only essential requirement was that to any n-points in the plane with distinct abscissae lying in a finite interval $[a, b]$, there should be a unique function in the class F passing through the given points. Such a system F is called an n-parameter family. Constructive methods for determining the function from F of best approximation to f, due to Remes [14] in the polynomial case, were extended to the above situation by Novodvorskii and Pinsker [13]. In this paper and in the paper of Motzkin two apparently additional requirements were placed on the system F. One, a continuity condition, was shown by Tornheim to follow from the axioms of F. The other, a condition on the multiplicity of the roots of $f - g, f, g \in F$, also follows from the definitions as will be shown in § 2. In § 3 the characterization of g_0 is discussed. Methods for constructing g_0 are given in § 4. These are based on the maximization of a certain function of $n + 1$ variables. In § 5 it is shown that an n-parameter familiy has a unique function of best approximation to an arbitrary continuous function in the $L_{p,N}$ norm if and only if F is the translate of a linear n-parameter family. The problem of the existence of n-parameter families on general compact spaces S is discussed in § 6. Under additional hypotheses on F it is shown that S must be homeomorphic to a subset of the circumference of the unit circle. If n is even this subset must be proper.

2. n-parameter families functions. Following Tornheim we define, for a fixed integer $n \geq 1$, an n-parameter family of functions F to be a class of real valued continuous functions on the finite interval $[a, b]$ such that for any real numbers

$$x_1, \cdots, x_n, y_1, \cdots, y_n, a \leq x_1 < x_2 < \cdots < x_n \leq b$$

Received February 17, 1959. This research was supported in part by the Space Technology Laboratories Inc..

there exists a unique $f \in F$ such that $f(x_i) = y_i \; i = 1, \cdots, n$. For convenience we will usually take $[a, b]$ to be the interval $[0, 1]$. We will include the possibility that 0 and 1 are identified. Then of course $x_1 \neq x_n$, and the functions of F are periodic of period 1. We call such a family a periodic n-parameter family. If we wish to consider specifically the case when 0 and 1 are not identified, we will refer to F as an ordinary n-parameter family. If F is a linear vector space of functions then we will call F a linear n-parameter family (e.g., polynomials of degree $\leq n-1$). The following continuity theorem of Tornheim [15] is a generalization of a result of Beckenbach [1] for $n = 2$.

THEOREM 1. *Let F be an n-parameter family on* $[0, 1]$. *For*

$$k = 1, 2, \cdots, \textit{let } x_1^{(k)}, \cdots, x_n^{(k)}, y_1^{(k)}, \cdots, y_n^{(k)}, 0 \leq x_1^{(k)} < \cdots < x_n^{(k)} \leq 1$$

be given sequences of real numbers and let f_k be the unique function from F such that

$$f_k(x_i^{(k)}) = y_i^{(k)} \qquad i = 1, \cdots, n .$$

Suppose for each

$$i, \lim_{k\to\infty} x_i^{(k)} = x_i, \lim_{k\to\infty} y_i^{(k)} = y_i \textit{ and } 0 \leq x_1 < \cdots < x_n \leq 1 .^1$$

Let f be the unique function from F such that $f(x_i) = y_i \; i = 1, \cdots, n$. Then $\lim_{k\to\infty} f_k = f$ uniformly on $[0, 1]$.

Proof. If 0 and 1 are not identified the proof is given in [15]. Therefore, let 0 and 1 be identified and the functions of F be periodic. Suppose f_k does not tend uniformly to f. For some $\varepsilon > 0$, there exists a sequence $\{u_k\} \subset [0, 1]$ such that for each k, $|f(u_k) - f_k(u_k)| \geq \varepsilon$. Since a subsequence of $\{u_k\}$ converges, we may assume $\{u_k\}$ does and let $u = \lim_{k\to\infty} u_k$. By a suitable rotation of $[0, 1]$ we may assume $u, x_1, \cdots, x_n$ all lie in the interior of an interval $[a, b]$, $0 < a < b < 1$. But F forms an ordinary n-parameter family on $[a, b]$ and hence $f_k \to f$ uniformly on $[a, b]$ which is a contradiction. This completes the proof.

We now verify that n-parameter families are unisolvent in the sense of Motzkin [10]. Let $f, g \in F$ and let x be an interior point of $[0, 1]$. If x is a zero of $f - g$ and if $f - g$ does not change sign in a suitably small neighborhood about x then we will say the zero x has multiplicity 2, otherwise we say x has multiplicity 1. If 0 and 1 are not identified and either is a zero of $f - g$, then the multiplicity is taken to be 1. We shall denote the sum of the multiplicities of the zeros of $f - g$ within an interval $[a, b]$ by $m_{a,b}(f, g)$. The following generalized con-

[1] If 0, 1 are identified we assume $x_n^{(k)} < 1$ and $x_n < 1$,

vexity notion is also useful. A continuous function h will be said to be convex to F if h intersects no function of F at more than n points. The following result extends Theorems 2 and 3 of [15].

THEOREM 2. *Let F be an n-parameter family on* $[0, 1]$ *and let h be convex to F. Then for any $f, g \in F, m_{0,1}(f, h) \leq n$ and $m_{0,1}(f, g) \leq n-1$.*

Proof. We assume first that 0 and 1 are not identified and that F is an ordinary n-parameter family. We verify the first statement by induction on n. For $n = 1$ the result follows by [15] Theorem 2. Hence, let h be a continuous function convex to a $k + 1$ parameter family F and assume the conclusion holds for all k-parameter families. For $f \in F$ let $x_i, i = 1, \cdots, m$, be the zeros of $f - h$ ordered from left to right and assume $m_{0,1}(f, h) > k + 1$. Choose a point u such that $x_1 < u < x_2$. If $F_1 = \{g \in F \mid g(x_1) = h(x_1)\}$, then F_1 is a k-parameter family on $[u, 1]$. $f \in F_1$ and h is convex to F_1. By our inductive assumption $m_{u,1}(f, h) \leq k$. Therefore x_1 must be a zero of $f - h$, and $m_{0,1}\,(f, h) = k + 2$. By the same reasoning we may assume x_m is a double zero of $f - h$.

We now construct a set E of k points from $[0, 1]$ in the following manner. First choose an $\varepsilon > 0$ such that $x_i + 2\varepsilon < x_{i+1} - 2\varepsilon, i = 1, \cdots, m - 1$. If x is a single zero of $f - h$ then let x belong to E. If x is a double zero of $f - h, x \neq x_1, x_m$ let $x + \varepsilon$, and $x - \varepsilon$ belong to E. We add the points $x_1 + \varepsilon, x_m - \varepsilon$. Since $m_{x_1+\varepsilon, x_m-\varepsilon}(f, h) = k - 2$ it is clear that E contains exactly k points. Choose a point $x', x_1 + \varepsilon < x' < x_2 - \varepsilon$. Let f_n be the unique function in F such that

$$f_n(x) = f(x),\ x \in E$$

$$f_n(x') = f(x') + \frac{1}{n}\operatorname{sgn}\,[f(x') - h(x')]$$

Now $f_n - f$ has k zeros which must all be simple by [15] Theorem 3. Within the interval $[x_1, x_m]\, f_n - h$ has exactly k simple zeros since f_n was chosen so that at the points $x_i \pm 2\varepsilon, i = 2, \cdots, m - 1, x_1 + 2\varepsilon, x_m - 2\varepsilon, f$ lies between f_n and h. Hence for $0 \leq x < x_1$ and $x_m < x \leq 1, f_n$ and h are on the same side of f (i.e., $\operatorname{sgn}[f_n(x) - f(x)] = \operatorname{sgn}[h(x) - f(x)]$. But by Theorem 1, f_n tends uniformly to f as $n \to \infty$. Hence for n sufficiently large $f_n - h$ must have at least $k + 2$ zeros which is a contradiction.

The case when 0 and 1 are identified and F is periodic causes no difficulty. For if $x_1, \cdots, x_m$ are the zeros of $f - h$, using a suitable rotation we may assume that there is an interval $[a, b]$, such that $0 < a < x_1 < \cdots < x_m < b < 1$. F is an ordinary n-parameter family on $[a, b]$ and $m_{0,1}(f, h) = m_{a,b}(f, h) \leq n$.

The verification of the second assertion is very similar to the above, and we leave the details to the reader.

COROLLARY. *There are no periodic n-parameter families when n is an even integer.*

Proof. Suppose false. Let F be a periodic n-parameter family and n an even integer. Let $f \in F$ and choose x_i $i = 1, \cdots, n$ such that $0 < x_1 < x_2 < \cdots < x_n < 1$. Choose $g \in T$ such that $g(x_i) = f(x_i)$ $i = 1, \cdots, n-1$, $g(x_n) = f(x_n) + 1$. By Theorem 2, $f - g$ changes sign at each of the points x_i, $i = 1, \cdots n-1$; and since $f - g$ can have no other zeros within $[0, 1]$, $g(1) > f(1)$. On the other hand $g(0) < f(0)$ which is a contradiction, since f, g are periodic of period 1.

3. Best approximation in the L_∞ norm. If g is continuous on $[0, 1]$, $g \notin F$, then $\{g - f\}$ forms a new n-parameter family. Hence without loss of generality we may consider the characterization and construction of the function $\hat{f} \in F$ such that

$$\|\hat{f}\| = \inf_{f \in F} \|f\| \equiv \delta$$

We first adopt the following notation. If $S \subset [0, 1]$

$$\delta_S = \inf_{f \in F} \sup_{t \in S} |f(t)| .$$

Let T denote the class of vectors $\boldsymbol{u} = (u_1, \cdots, u_{n+1})$ satisfying the condition that $0 \leq u_1 < u_2 < \cdots u_{n+1} \leq 1$. The statements and proofs of the results of this section are valid when F consists of continuous periodic functions on $[0, 1]$. We shall assume, however, that F is an ordinary n-parameter family and leave the details in the periodic case to the reader.

The following two lemmas are appropriate generalizations of results of de la Vallee Poussin [6] for polynomials. Where possible we refer the reader to [13] for proofs.

LEMMA 1. *For any $\boldsymbol{u} = (u_1, \cdots, u_{n+1}) \in T$ there exists a unique $f \in F$ and unique real number λ such that $f(u_i) = (-1)^i \lambda \cdot i = 1, \cdots, n+1$. Moreover $|\lambda| = \delta_{\boldsymbol{u}}$ and f is the only function in F with the property that $\max_{i=1,\dots,n+1} |f(u_i)| = \delta_{\boldsymbol{u}}$. In addition suppose for $k = 1, 2, \cdots$ that*

$$\boldsymbol{u}^{(k)} = (u_1^{(k)}, \cdots, u_{n+1}^{(k)}) \in T \textit{ and } f_k\ (u_i^{(k)}) = (-1)^i \lambda^{(k)} .$$

Then if $\boldsymbol{u}^{(k)} \to \boldsymbol{u}$ and $\boldsymbol{u} \in T$, it follows that $f_k \to f$ uniformly on $[0, 1]$ and $\lambda^{(k)} \to \lambda$.

LEMMA 2. *Let $\boldsymbol{u} \in T$ and a sequence of non-negative numbers λ_i $i = 1, \cdots, n+1$ be given. If there exists an $f \in F$ such that*

$$f(u_i) = (-1)^i \lambda_i \; i = 1, \cdots, n+1 \text{ or } f(u_i) = (-1)^{i+1} \lambda_i \; i = 1, \cdots, n+1$$

then either $\min \lambda_i < \delta_{\boldsymbol{u}} < \max \lambda_i$ or $\lambda_i = \delta_{\boldsymbol{u}}$ $i = 1, \cdots, n+1$.

Proof. Lemma 2 is a restatement of Lemma 1 of [13]. Everything in Lemma 1 except the facts that $|\lambda| = \delta_{\boldsymbol{u}}$ and the function f satisfying $\max_{i=1,\ldots,n+1} |f(u_i)| = \delta_{\boldsymbol{u}}$ is unique is proved explicitly in [13]. To prove the latter statements observe that if there is a $g \in F$ satisfying $|g(u_i)| < |\lambda|$ then $f(u_i) - g(u_i) = (-1)^i \lambda_i$ $i = 1, \cdots, n+1$ where either $\lambda_i \geq 0$, $i = 1, 2, \cdots, n+1$ or $\lambda_i \leq 0$ $i = 1, 2, \cdots, n+1$. In either case by [12], Lemma 1, $f - g$ must have at least n zeros between u_1 and u_{n+1} counting multiplicity which is a contradiction.

For $\boldsymbol{u} \in T$ we will usually denote the function f of Lemma 1 by $f_{\boldsymbol{u}}$. Next we define a function $\delta(u_1, \cdots, u_{n+1})$ of $n+1$ variables.

$$\begin{aligned} \delta(\boldsymbol{u}) \equiv \delta(u_1, \cdots, u_{n+1}) &= \delta_{\boldsymbol{u}} \text{ if } \boldsymbol{u} = (u_1, \cdots, u_{n+1}) \in T \\ &= 0 \text{ otherwise .} \end{aligned}$$

If we restrict the points u_i to lie in some subset $S \subset [0, 1]$, then $\delta(u_1, \cdots, u_{n+1})$ will be denoted $\delta_S(u_1, \cdots, u_{n+1})$.

LEMMA 3. *$\delta(u_1, \cdots, u_{n+1})$ is continuous on R^{n+1}*

Proof. Assume that $\delta(u_1, \cdots, u_{n+1})$ is not continuous at some point $\boldsymbol{u} = (u_1, \cdots, u_{n+1})$. We may assume $0 \leq u_1 \leq u_2 \leq \cdots \leq u_{n+1} \leq 1$, and by Lemma 1 we may assume that $m (\leq n)$ of the points u_i are distinct. Consequently $\delta(u_1, \cdots, u_{n+1}) = 0$. Suppose there exists an $\varepsilon > 0$ and a sequence $\{\boldsymbol{u}_k\} \subset T$ such that $\boldsymbol{u}_k \to \boldsymbol{u}$ and $\delta_{\boldsymbol{u}_k} \geq \varepsilon$. Let $u_i^{(k)}$ be the ith coordinate of $\boldsymbol{u}_k$. Choose n points u_i', $0 \leq u_i' < \cdots < u_n' \leq 1$ such that m of the points u_i' coincide with the m distinct points u_i. Let f_0 be the unique function in F such that $f_0(u_i') = 0$. Choose η such that for any i $|u_i' - u_i| < \eta$ implies $|f_0(u_0)| < \varepsilon/2$. Choose k so large that all coordinates of $\boldsymbol{u}_k$ are within η neighborhoods of some coordinate of $\boldsymbol{u}'$. Then $f_{\boldsymbol{u}_k}(u_i^{(k)}) - f_0(u_i^{(k)}) = (-1)^i \lambda_i$ where $\operatorname{sgn} \lambda_i^{(k)} = \operatorname{sgn} \lambda_{i+i}^{(k)}$ $i = 1, \cdots, n$. As in the proof of Lemma 1 it follows that $f_{\boldsymbol{u}_k} - f_0$ must have at least n zeros within $[0, 1]$ which is a contradiction.

Using the function $\delta(u_1, \cdots, u_{n+1})$ one can give a simple proof of the Theorem of Motzkin and Tornheim characterizing the function $\hat{f}$ which has minimum deviation from zero.

THEOREM 3. *There exists a unique $\hat{f} \in F$ such that $\|\hat{f}\| = \inf_{f \in F} \|f\|$. $\hat{f}$ is uniquely characterized by the fact that for some $\boldsymbol{u} = (u_1, \cdots, u_{n+1}) \in T$*

$\|\hat{f}\| = \delta_u$. *u will have this property if and only if $\delta(u_1, \cdots, u_{n+1})$ is an absolute maximum, and then $\hat{f} = f_u$.*

Proof. Since $\delta(u_1, \cdots, u_{n+1})$ is a continuous function on a compact set, its maximum is attained for some $u = (u_1, \cdots, u_{n+1}) \in T$. Assert $\|f_u\| = \delta_u$. If $\|f_u\| > \delta_u$, then there is a point x' in $[0, 1]$ for which $|f_u(x')| = \|f_u\|$. We form a new vector $u' \in T$ by replacing one coordinate u_i of u by x' in the following way. If $u_i < x' < u_{i+1}$ $i = 1, \cdots, n$ and $\operatorname{sgn} f_u(u_i) = \operatorname{sgn} f_u(x')$ then let $u'_j = u_j$, $j \neq i$, and $u'_i = x'$. If $\operatorname{sgn} f_u(u_i) = (-1) \operatorname{sgn} f_u(x')$ let $u'_j = u_j$ $j \neq i+1$ and $u'_{i+1} = x'$. If $x' < u_1 (x' > u_{n+1})$ and $\operatorname{sgn} f_u(u_1) = \operatorname{sgn} f_u(x')$ $(\operatorname{sgn} f_u(u_{n+1}) = \operatorname{sgn} f_u(x'))$ let $u'_j = u_j$ $j \neq 1$ $(j \neq n+1)$ and $u'_1 = x'$ $(u'_{n+1} = x')$. If $\operatorname{sgn} f_u(u_1) = (-1) \operatorname{sgn} f_u(x')$ $(\operatorname{sgn} f_u(u_{n+1}) = (-1) \operatorname{sgn} f_u(x'))$ then let $u'_1 = x'$, $u'_j = u_{j-1}$ $j = 2, \cdots, n+1$ $(u'_j = u_{j+1}, j = 1, \cdots, n, u'_{n+1} = x')$. Now either $f_u(u'_i) = (-1)^i \lambda_i$ $i = 1, \cdots, n+1$ or $f_u(u'_1) = (-1)^{i+1} \lambda_i$ $i = 1, \cdots, n+1$ where $\lambda_i = \delta_u$ or $\lambda_i = \|f_u\|$. Therefore by Lemma 2, $\delta_u < \delta_{u'} < \|f_u\|$ which contradicts the maximality of δ_u.

It now follows immediately that $\|f_u\| = \inf_{f \in F} \|f\|$ and that f_u is the only such function with this property. For if $f_0 \in F$ and $\|f_0\| \leq \|f_u\|$ then $\|f_0\| \leq \delta_u$ which contradicts Lemma 1. Moreover the same argument shows that if there exists an $f_0 \in F$ and a $v \in T$ such that $\|f_0\| = \delta_v$ then $\|f_0\| = \inf_{f \in F} \|f\|$. It is clear that $\delta(v_1, \cdots, v_{n+1})$ must be an absolute maximum.

In the above theorem if $\|f\|$ is replaced by $\|f\|_S = \sup_{t \in S} |f(t)|$ where S is any closed set of $[0, 1]$ containing at least $n+1$ points, then the same conclusions hold. Here of course, the function $\delta(u_1, \cdots, u_{n+1})$ is replaced by $\delta_S(u_1, \cdots, u_{n+1})$ and the points u_k are assumed to be in S. The following generalization of [11] Theorem 7.1 is therefore relevant.

THEOREM 4. *Let S_k, S be closed sets of $[0, 1]$ such that for each k, S_k, contains at least $n+1$ points; S contains infinitely many points, and $S_k \subset S$. Let $\hat{f}_k, \hat{f}_0$ be functions from F which minimize $\|f\|_{S_k}, \|f\|_S$ respectively. If for each $\varepsilon > 0$ there exists an integer k_0 such that for $k > k_0$ each point $u \in S$ is at a distance less than ε from some point of S_k, than $\hat{f}_k \to \hat{f}_0$ uniformly on $[0, 1]$.*

Proof. We assume $\delta_S > 0$. $S_k \subset S$ implies $\delta_{S_k} \leq \delta_S$. Choose $u = (u_1, \cdots, u_{n+1}) \in T$, $u_i \in S$ such that $\delta_S(u_1, \cdots, u_{n+1})$ is an absolute maximum. Let $u_k = (u_1^{(k)}, \cdots, u_{n+1}^{(k)}) \in T$, $u_j^{(k)} \in S_k$ be chosen such that $u_k \to u$. By Lemma 1, $\delta_{u_k} \to \delta_u$ and since $\delta_{u_k} \leq \delta_{S_k}$, $\delta_{S_k} \to \delta_u = \delta_S$. Let $v_k = (v_1^{(k)}, \cdots, v_{n+1}^{(k)}) \in T$, $v_i^{(k)} \in S_k$ be chosen so that for each k, $\delta_{S_k}(v_1^{(k)}, \cdots, v_{n+1}^{(k)})$ is an absolute maximum. Extract any convergent subsequence v_{k_j} with limit v.

If $\boldsymbol{v} = (v_1, \cdots, v_{n+1})$, then $v_i \in S$ and $\delta_v = \delta_S$. Also $\hat{f}_{k_j} = f_{v_{k_j}}$ tends uniformly to f_v, the function from F with minimum deviation on $\boldsymbol{v}$. But by the uniqueness of $f_v, f_v = \hat{f}_0$. The above argument shows that any subsequence of $\{\hat{f}_k\}$ contains a refinement which converges to $\hat{f}_0$. Hence $\lim_{k\to\infty} \hat{f}_k = \hat{f}_0$ uniformly on $[0, 1]$.

4. The estimation of f. In [13] Novodovorskii and Pinsker consider a direct method, due to Remes [14] in the polynomial case, for the estimation of $\hat{f}$. However the following Lemma shows that $\hat{f}$ is continuously dependent on estimates of the best approximation. Hence if $\boldsymbol{u}$ is a vector in T for which $\delta(\boldsymbol{u})$ is an estimate of $\inf_{h\in F} \|f\|$, then the solution of the equation $f(u_i) = (-1)^i\lambda\ i = 1, \cdots, n+1$ is the appropriate estimate of $\hat{f}$.

LEMMA 4. *Let $\{\delta_n\}$ be a sequence of non-negative numbers converging to $\delta = \inf_{f\in F} \|f\|$ from below. If $\boldsymbol{u}_n$ are vectors in T for which $\delta(\boldsymbol{u}_n) = \delta_n$, then $\lim_{n\to\infty} f_{\boldsymbol{u}_n} = \hat{f}$ uniformly on $[0, 1]$.*

Proof. If the conclusion is false there exists a subsequence $\{\boldsymbol{u}_{k_j}\}$ and a number $\varepsilon > 0$ such that $\|\hat{f} - f^{\boldsymbol{u}_{k_j}}\| \geq \varepsilon$. But $\{\boldsymbol{u}_{k_j}\}$ may be further refined to obtain a convergent subsequence of vectors. Calling this $\{\boldsymbol{u}_{k_j}\}$ and letting $\boldsymbol{u}_0 = \lim_{j\to\infty} \boldsymbol{u}_{k_j}$ we have by Lemma 1 $\delta(\boldsymbol{u}_0) = \lim_{j\to\infty} \delta(\boldsymbol{u}_{k_j})$. By Theorem 3 $f_{\boldsymbol{u}_0} = \hat{f}$ which is a contradiction.

We shall consider two algorithms for estimating δ and prove convergence of both.

Each of these algorithms can be used efficiently for actual numerical calculations. A detailed description of method 2 for polynomials on a finite point set can be found in [5]. Also for polynomials on an interval a maximization procedure has been announced by Bratton [3].

For both methods the following notation is convenient. For $\boldsymbol{u} = (u_1, \cdots, u_{n+1}) \in T$ define for $j = 1, \cdots, n+1$.

$$\begin{aligned} \delta_{\boldsymbol{u}}^{(j)}(x) &= \delta(u_1, \cdots, u_{j-1}, x, u_{j+1}, \cdots, u_{n+1}) \text{ if } u_{j-1} \leq x \leq u_{j+1} \\ &= 0 \text{ otherwise} \end{aligned}$$

where we take $u_0 = 0, u_{n+2} = 1$. We now form $\eta_{\boldsymbol{u}}(x) \equiv \max_{j=1,\ldots,n+1} \delta_{\boldsymbol{u}}^j(x)$. From the continuity of $\delta(u_1, \cdots, u_{n+1})$ it follows that for each j, $\delta_u^{(j)}(x)$ is continuous, and hence $\eta_{\boldsymbol{u}}(x)$ is continuous. Therefore there exists a point x', $0 \leq x' \leq 1$ and integer $1 \leq m \leq n+1$ such that

$$\delta_{\boldsymbol{u}}^m(x') = \max_{j=1,\cdots,n+1} \|\delta_{\boldsymbol{u}}^j\| = \|\eta_{\boldsymbol{u}}\| .$$

For a given vector $\boldsymbol{u}$ we define $\boldsymbol{u}' = (u_1', \cdots, u_{n+1}')$ by setting $u_j' = u_j$, $j \neq m$, $u_m' = x'$.

THEOREM 5. *If vectors $\boldsymbol{u}_k$ are defined inductively in the above fashion with $\boldsymbol{u}_1 \in T$ chosen arbitrarily, then $\lim_{k\to\infty} \delta(\boldsymbol{u}_k)$ exists and there exists $\boldsymbol{u}_0 \in T$ such that $\delta(\boldsymbol{u}_0) = \lim_{k\to\infty} \delta(\boldsymbol{u}_k)$. Furthermore $\delta(\boldsymbol{u}_0)$ is an absolute maximum of the function $\delta(\boldsymbol{u})$.*

Proof. $\{\delta(\boldsymbol{u}_k)\}$ is a monotonically increasing, bounded sequence hence convergent. If $\delta = \lim_{k\to\infty} \delta(\boldsymbol{u}_k)$, then a suitable subsequence $\{\boldsymbol{u}_{k_j}\}$, converges to $\boldsymbol{u}_0$ and $\delta(\boldsymbol{u}_0) = \delta$. We now assert $\eta_{\boldsymbol{u}_{k_j}}(x)$ converges uniformly to $\eta_{\boldsymbol{u}_0}(x)$. It suffices to assume $u_i \le x \le u_{i+1}$. Then

$$\begin{aligned} |\eta_{\boldsymbol{u}_0}(x) - \eta_{\boldsymbol{u}_{k_j}}(x)| &= |\max(\delta^i_{\boldsymbol{u}_0}(x), \delta^{i+1}_{\boldsymbol{u}_0}(x),)) - \max(\delta^i_{\boldsymbol{u}_{k_j}}(x), \delta^{i+1}_{\boldsymbol{u}_{k_j}}(x))| \\ &\le |\delta^i_{\boldsymbol{u}_0}(x) - \delta^i_{\boldsymbol{u}_{k_j}}(x)| + |\delta^{i+1}_{\boldsymbol{u}_0}(x) - \delta^{i+1}_{\boldsymbol{u}_{k_j}}(x)| . \end{aligned}$$

Since $\delta(\boldsymbol{u})$ is a uniformly continuous function the latter expression tends to zero uniformly in x.
Hence

$$\|\eta_{\boldsymbol{u}_0}\| = \lim_{j\to\infty} \|\eta_{\boldsymbol{u}_{k_j}}\| .$$

But

$$\|\eta_{\boldsymbol{u}_{k_j}}\| = \delta(\boldsymbol{u}_{k_j+1}) \le \delta(\boldsymbol{u}_{k_{j+1}}) \le \|\eta_{\boldsymbol{u}_{k_{j+1}}}\|$$

Therefore $\|\eta_{\boldsymbol{u}_0}\| = \lim_{j\to\infty} \delta(\boldsymbol{u}_{k_j}) = \delta(\boldsymbol{u}_0)$. It now follows by the same argument as in the proof of Theorem 3 that $\|f_{\boldsymbol{u}_0}\| = \delta(\boldsymbol{u}_0)$ and by Theorem 3, $\delta(\boldsymbol{u}_0)$ is a maximum.

For the second method of estimation of f we alter slightly our definition of $\delta^1_{\boldsymbol{u}}(x)$ and $\delta^{n+1}_{\boldsymbol{u}}(x)$. We now define

$$\begin{aligned} \delta^1_{\boldsymbol{u}}(x) &= \delta(x, u_2, \cdots, u_{n+1}) \text{ if } 0 \le x \le u_2 . \\ &= \delta(u_2, u_3, \cdots, u_{n+1}, x) \text{ if } u_{n+1} \le x \le 1 \end{aligned}$$

$$\begin{aligned} \delta^{n+1}_{\boldsymbol{u}}(x) &= \delta(u_1, \cdots, u_n, x) \text{ if } u_n \le x \le 1 \\ &= \delta(x, u_1, \cdots, u_n) \text{ if } 0 \le x \le u_1 . \end{aligned}$$

The algorithm proceeds as follows. First let $\varepsilon > 0$ be chosen. Select an arbitrary vector $\boldsymbol{u} \in T$. Maximize $\delta^2_{\boldsymbol{u}}(x)$ over its domain of definition. Let x' be a point for which $\delta^2_{\boldsymbol{u}}(x)$ is a maximum. If $\delta^2_{\boldsymbol{u}}(x') \ge (1+\varepsilon)\delta(\boldsymbol{u})$, replace u_2 by x' forming a new vector u'. If not, let $\boldsymbol{u}' = \boldsymbol{u}$. We now maximize $\delta^3_{\boldsymbol{u}'}(x)$ and continue inductively. Special attention is necessary for $\delta^{n+1}_{\boldsymbol{u}}(x)$ and $\delta^1_{\boldsymbol{u}}(x)$. If x' is a point for which $\delta^{n+1}_{\boldsymbol{u}}(x)$ is a maximum and $\delta^{n+1}_{\boldsymbol{u}}(x) \ge (1+\varepsilon)\delta(\boldsymbol{u})$, then $\boldsymbol{u}'$ is formed in the following way. If $x' \ge u_n$ then $u'_i = u_i$, $i = 1, \cdots, n$, $u'_{n+1} = x'$; if $x' \le u_1$ then $u'_1 = x'$ $u'_i = u_{i-1}$ $i = 2, \cdots, n+1$. In the latter case, the next function maximized is $\delta^2_{\boldsymbol{u}'}(x)$. If the first case occurs then $\delta^1_{\boldsymbol{u}'}(x)$ is maximized. Let x'' be a point for which $\delta^1_{\boldsymbol{u}'}(x)$.

is a maximum and $\delta^1_{u'}(x'') \geq (1+\varepsilon)\delta(\boldsymbol{u}')$. If $x'' \leq u'_2$ then $u''_1 = x''$ and $u''_i = u'_i$ $i = 2, 3, \cdots, n+1$. If $x'' \geq u'_{n+1}$ then $u''_i = u_{i+1}$ $i = 1, \cdots, n$ and $u''_{n+1} = x''$. For the first case the next function maximized is $\delta^2_{u''}(x)$; the second case, $\delta^{(1)}_{u''}(x)$. If

$$\delta^{n+1}_{\boldsymbol{u}}(x') < (1+\varepsilon)\delta(\boldsymbol{u})\,(\delta^1_{u'}(x'') < (1+\varepsilon)\delta(\boldsymbol{u}'))$$

then we take $\boldsymbol{u}' = \boldsymbol{u}\,(\boldsymbol{u}'' = u')$. When there have been $n+1$ consecutive maximizations with no change in the vector $\boldsymbol{u}$, ε is now replaced by $\varepsilon/2$ and the process is repeated. We now continue inductively and pass to the limit as $\varepsilon/2^k \to 0$.

THEOREM 6. *The conclusions of Theorem* 5 *hold if the sequence* $\{\boldsymbol{u}_k\}$ *is chosen inductively in accordance with the above algorithm.*

Proof. As before, $\lim_{k\to\infty}\delta(\boldsymbol{u}_k) = \delta$ exists. We choose a particular convergent subsequence $\{\boldsymbol{u}_{k_j}\}$ of $\{\boldsymbol{u}_k\}$. For each j let $\boldsymbol{u}_{k_j}$ be a vector of $\{\boldsymbol{u}_k\}$ such that for each i, $i = 1, \cdots, n+1$ and all appropriate x, $\delta^i_{\boldsymbol{u}_{k_j}}(x) < (1+\varepsilon/2^j)\delta(\boldsymbol{u}_{k_j})$. The algorithm guarantees that for each integer j such a vector $\boldsymbol{u}_{k_j}$ exists in the sequence $\{\boldsymbol{u}_k\}$. Since a refinement of this sequence is convergent, we assume $\{\boldsymbol{u}_{k_j}\}$ converges. Then if $\boldsymbol{u}_{k_j} \to \boldsymbol{u}_0$, $\delta(\boldsymbol{u}_0) = \delta$. Suppose $\delta(\boldsymbol{u}_0)$ is not a maximum of $\delta(\boldsymbol{u})$, then $\|f_{\boldsymbol{u}_0}\| > \delta(\boldsymbol{u}_0)$. Choose x' so that $|f_{\boldsymbol{u}}(x')| = \|f\|$, and form $\boldsymbol{u}'$ by replacing one point, the ith say, of $\boldsymbol{u}_0$ by x' in the manner of the proof of Theorem 3. Form $\boldsymbol{u}'_{k_j}$ by replacing the ith coordinate of $\boldsymbol{u}_{k_j}$ by x' Then $\boldsymbol{u}'_{k_j} \to \boldsymbol{u}'$ and $\delta(\boldsymbol{u}'_{k_j}) \to \delta(\boldsymbol{u}')$. Therefore for j sufficiently large, since $\delta(\boldsymbol{u}') > \delta$,

$$\delta(\boldsymbol{u}'_{k_j}) > \frac{\delta(\boldsymbol{u}') + \delta}{2}$$

On the other hand for each j there is a point x and an integer m such that

$$\delta(\boldsymbol{u}'_{k_j}) = \delta^m_{\boldsymbol{u}_{k_j}}(x) \leq \left(1 + \frac{\varepsilon}{2^j}\right)\delta(\boldsymbol{u}_{k_j}) \leq \left(1 + \frac{\varepsilon}{2^j}\right)\delta\,.$$

For j sufficiently large this is a contradiction, therefore $\|f_{\boldsymbol{u}_0}\| = \delta(\boldsymbol{u}_0)$ and $\delta(\boldsymbol{u}_0)$ is an absolute maximum.

5. Approximation in $L_{p,N}$ norm. For $N \geq n$ let $x_1, \cdots, x_N$ be N distinct points of $[0,1]$. In place of the sup norm let $\|f\| = \{\sum_{i=1}^N |f(x_i)|^p\}^{1/p}$ and assume $p > 1$. The fundamental problem to be considered here is to give necessary and sufficient conditions that the function $\hat{f} \in F$ for which $\|\hat{f}\| = \inf_{f\in F}\|\hat{f}\|$ is unique. Now the image of F under the mapping $f \to (f(x_1), \cdots, f(x_N))$ is a closed set in N dimensional Euclidean space. By a theorem of Motzkin [9] as generalized by Busemann [4], to each point $x \in E_N$ there will exist a unique nearest point in a given set $S \subset E_N$ with respect to a strictly convex metric if and only if S is closed and convex. Hence $\hat{f}$ will be unique if and only if F is convex, but for n-parameter families we can say more.[2]

THEOREM 7. *An n-parameter family F is convex if and only if F is the translate of a linear n-parameter family.*

Proof. If F is the translate of a linear n-parameter family, i.e., there exists a continuous g on $[0, 1]$ and a linear n-parameter family F_0 such that each $f \in F$ can be written uniquely as $f = g + f', f \in F_0$, then F is obviously convex. Conversely suppose F is convex. Choose n distinct points $x_1, \cdots, x_n$ in $[0, 1]$. Let $f_0, f_1, \cdots, f_n$ be the unique functions of F such that $f_0(x_j) = 0, j = 1, \cdots, n; f_k(x_j) = \delta_{kj}$ for $k, \ j = 1, \cdots, n$ where δ_{kj} is the Kronecker delta. We assert that each $f \in F$ has a representation as

$$f = f_0 + \sum_{k=1}^{n} \lambda_k(f_k - f_0) \text{ where } \lambda_k = f(x_k) \ .$$

If such a representation exists it is obviously unique. Also the vector space spanned by $f_1 - f_0, \cdots, f_n - f_0$, is obviously an n-parameter family and the theorem is proved. To prove the assertion let

$$F_k = \{f \in F \mid f(x_{k+1}) = f(x_{k+2}) = \cdots = f(x_n) = 0\}$$
$$F'_k = \{f \in F \mid f(x_j) = 0 \ \ j \neq k\} \ .$$

From the convexity of F, F'_k is a convex one parameter family on a suitably small interval containing x_k. We assert $f \in F'_k$ implies $f = f_0 + \lambda_k(f_k - f_0)$ where $\lambda_k = f(x_k)$. By convexity this is obviously true for $0 \leq \lambda_k \leq 1$. For $\lambda_k > 1$ if $f \in F'_k, f(x_k) = \lambda_k$ then by convexity

$$f_k = \frac{1}{\lambda_k} f + \left(1 - \frac{1}{\lambda_k}\right) f_0$$

or $f = f_0 + \lambda_k(f_k - f_0)$. If $\lambda_k < 0$,

$$f_0 = \frac{1}{1 - \lambda_k} f + \frac{(-\lambda_k)}{1 - \lambda_k} f_k$$

or $f = f_0 + \lambda_k(f_k - f_0)$. To finish the proof we apply an induction. Assume $f \in F_k$ implies that $f = f_0 + \sum_{j=1}^{k} \lambda_j(x_j - x_0)$ where $f(x_j) = \lambda_j$ and

[2] For a discussion of related results see the article by Motzkin in the Symposium on Numerical Approximation, University of Wisconsin Press, 1959.

suppose $g \in F_{k+1}$ and $g(x_j) = \mu_j, j = 1, \cdots, k+1$. Then if $g_1 = f_0 + \sum_{j=1}^{k} 2\mu_j(f_j - f_0)$, $g_2 = f_0 + 2\mu_{k+1}(f_{k+1} - f_0)$ it follows that

$$g' = \frac{g_1 + g_2}{2} \in F_{k+1}$$

and $g'(x_j) = \mu_j, j = 1, \cdots, k+1$. Therefore

$$g = g' = f_0 + \sum_{j=1}^{k+1} \mu_j(f_j - f_0) .$$

6. The existence of n-parameter families on compact space. Let $f_1, \cdots, f_n$, be n linearly independent real valued continuous functions defined on a compact set S in finite dimensional Euclidean space. Let V be the span of the functions $f_1, \cdots, f_n$. In 1918 Haar [7] showed that to each continuous real valued function g defined on S, there is a unique $\hat{f} \in V$ satisfying $\|\hat{f} - g\| = \inf_{f \in V} \|f - g\|$ where $\|f\| = \sup_{s \in S} |f(s)|$ if and only if no non-zero function in V vanished at more than $n - 1$ points of S. Haar noted that the existence of such a set of functions V placed a severe restriction on the set S. In 1956 Mairhuber [8] proved that if V satisfied the above condition of Haar then S is a homeomorphic image of a subset of the circumference of the unit circle. If n is even this subset must be proper. It is clear that V satisfies the condition of Haar if and only if V is a linear n-parameter family. The characterization of those compact Hausdorff spaces on which there exist n-parameter families F for $n > 1$ seems to be quite difficult. One can give a characterization if one imposes a rather strong local condition on F. The result presented here includes the one of Mairhuber, and is proved by somewhat different means. The following fundamental lemma is perhaps of independent interest.

LEMMA 5. *Let S be a compact connected Hausdorff space with the property that for each point $x \in S$ there exists a neighborhood U_x and continuous real valued functions f_1, f_2 defined on U_x such that for $y, z \in U_x$, $y \neq z$*

$$\begin{vmatrix} f_1(y) & f_1(z) \\ f_2(y) & f_2(z) \end{vmatrix} \neq 0 . \tag{1}$$

Then S may be embedded homeomorphically into the circumference C of the unit circle.

Proof. Without loss of generality we assume U_x is a closed, therefore compact neighborhood of x. f_1, f_2 never vanish simultaneously on U_x and therefore f_1/f_2 defines a continuous mapping of U_x into the compactified real line. (1) guarantees that the mapping is one to one and $\phi_x(u) = \mathrm{Arctan}\,(f_1/f_2)(u)$ gives a homeomorphism of U_x into C.

We next verify that S is locally connected. To do this it suffices to show that for each $x \in S$ there exists a connected neighborhood which can be mapped homeomorphically into C. In fact if ϕ_x is the homeomorphism for a point $x \in S$ constructed above, and if $C_x = \phi_x(U_x)$, it is enough to show that there exists a connected neighborhood V_x in C_x of $\lambda_x \equiv \phi_x(x)$. For then $\phi_x^{-1}(V_x)$ is a connected neighborhood of x contained in U_x. But C_x is a compact subset of C. Therefore let I_x be the component of λ_x in C_x. I_x is a compact connected subset of C. I_x is then either an interval or all of C. If I_x is the latter we are through. Also if I_x is an interval and λ_x an interior point (relative to C) then $\phi_x^{-1}(I_x)$ is the required neighborhood. Hence assume that λ_x is an end point of I_x. This will include that degenerate case when I_x is just one point. We may also assume that there does not exist a suitably small connected neighborhood N of λ_x in C such that $N \cap C_x \subset I_x$. For then $\phi_x^{-1}(N \cap N_x)$ is an appropriate neighborhood of x. Therefore it now must follow that for any connected neighborhood N of λ_x in C there exists λ_1, λ_2 in the interior of N such that $\lambda_1, \lambda_2 \notin C_x$ and $(\lambda_1, \lambda_2) \cap C_x \neq \phi$. If we let $F = \phi_x^{-1}[(\lambda_1, \lambda_2) \cap C_x]$ and $G = \phi_x^{-1}[C_x \sim (\lambda_1, \lambda_2)]$ then $F \cup (S \sim U_x)$ and G separate S which is a contradiction.

We note that S is certainly a separable metric since a finite number of homeomorphic images of subsets of C cover S. Hence by [16] Theorem 5.1, S is arc wise connected.

We now assert S is homeomorphic to a subset of C. Let $U_1, \cdots, U_n$ be a finite collection of connected neighborhoods covering S each of which is homeomorphic to a subset of C. By a suitable rearrangement we may assume that $U_2 \cap U_1 \neq \phi$ and $U_2 \not\subset U_1$. Let $x_1 \in U_1 \sim U_2$, $x_2 \in U_2 \sim U_1$ $x \in U_1 \cap U_2$. Let A be the maximal subset of $U_1 \cup U_2$ connecting x_1, x, x_2. This must be all of $U_1 \cup U_2$, for if $y \in U_1 \cup U_2$ and $y \notin A$, then y may be connected to any point in A by an arc in $U_1 \cup U_2$. If y is connected to A at an end point of A, this is an enlargement of A which contradicts maximality. If y is connected to A at a point other than an end point, then no neighborhood of this point is homeomorphic to a subset of C. This also is a contradiction. If $U_1 \cup U_2$ is not all of S then $U_1 \cup U_2$ is homeomorphic to an arc, and by induction the homeomorphism may be extended to all of S.

THEOREM 8. *For $n > 1$ let F be an n-parameter family of functions defined on a compact Hausdorff space S. Suppose in addition that to each point $x \in S$ there exists a neighborhood N_x and functions $f_1, f_2 \in F$ such that*

$$\begin{vmatrix} f_1(y) & f_1(z) \\ f_2(y) & f_2(z) \end{vmatrix} \neq 0$$

for $y, z \in N_x$, $y \neq z$. Then there exists a homeomorphism of S into the circumference of the unit circle. If n is even the image of S must be a proper subset of C.

Proof. First we note that S cannot have a proper subset W homeomorphic to C. If n is even this follows directly from the Corollary to Theorem 2. If n is odd, choose $x \in S \sim W$ and let $F' = \{f \in F \mid f(x) = 0\}$; then F' is an $n-1$ parameter family defined on W. Since $n-1$ is even this is a contradiction. We may therefore assume that if n is even S is not homeomorphic to C.

If I is a component of S then by Lemma 5 there exists a homeomorphism ϕ of I onto the closed interval $[0, 1]$ considered as a subset of C. We assert that if I is not all of S, then ϕ can be extended to an open and closed set $U \supset I$. U and its complement then separate S. If I is itself open in S then we take $U = I$. If not, let $x = \phi^{-1}(0)$, $y = \phi^{-1}(1)$. Let N_x, N_y be compact neighborhoods of x and y respectively and let ϕ_x, ϕ_y be homeomorphisms of N_x and N_y respectively into C. We may assume $\phi_x(x) = 0$, $\phi_y(y) = 1$ and

$$\phi_x(N_x \cap I) \subset [0, 1] \text{ and } \phi_y[N_y \cap I] \subset [0, 1] .$$

If we define ϕ' by

$$\begin{aligned} \phi'(z) &= \phi(z) \quad \text{if } z \in I \\ &= \phi_x(z) \text{ if } z \in N_x \sim I \\ &= \phi_y(z) \text{ if } z \in N_y \sim I \end{aligned}$$

then ϕ' is a homeomorphism of $N_x \cup N_y \cup I \equiv N$ into C. Also int. $N \supset I$. Now $[0, 1] = \phi'(I)$ is the maximal connected subset of $\phi'(N)$ containing $\phi'(I)$. Therefore there exist sequences $\{\lambda_n\}$, $\{\mu_n\}$ of real numbers tending monotonically to 0 from below, and monotonically to 1 from above, respectively such that $\{\lambda_n\} \cap \phi'(N) = \phi$ and $\{\mu_n\} \cap \phi'(N) = \phi$. Choose n large enough that $\phi'^{-1}[\lambda_n, 0] \subset$ interior of N_x and $\phi'^{-1}[1, \mu_n] \subset$ interior of N_y. Clearly $J_n = \phi'^{-1}[\lambda_n, \mu_n]$ is a closed set containing I. J_n is open in the interior of N. Hence J_n is open in S.

Let T be the class of open sets O of S which can be mapped homeomorphically into C. We partially order T in the following way. If $O_1, O_2 \in T$ then $O_1 \leq O_2$ if $O_1 \subset O_2$ and if there exist homeomorphisms ϕ_1, ϕ_2 of O_1, O_2 respectively into C such that ϕ_2 agrees with ϕ_1 on O_1. By Zorn's lemma there exists a maximal element O of T. We assert $O = S$. If not, let $x \in S \sim O$. Then there exists an open and closed set $U \ni x$ and mapping ϕ such that ϕ maps U homeomorphically into C.

$O \cap U$ and $O \sim U$ are separated open sets of S. Hence if ϕ' is any homeomorphism of O into C such $\phi'(O) \cap \phi(U) = \phi$. ϕ'' defined by $\phi''(x) \equiv \phi(x)$, $x \in O \cap U$, $\phi''(x) \equiv \phi'(x)$, $x \in O \sim U$ is also a homeomorphism of O into C. ϕ'' has an obvious extension to $U \cup O$ which contradicts the maximality of O.

COROLLARY. *If F is a linear n-parameter family $(n > 1)$ defined on the compact Hausdorff space S, then S is homeomorphic to a subset of C. If n is even the subset must be proper.*

Proof. We assume S contains more than n points. For a given $x \in S$ choose $n-2$ distinct points $x_1, \cdots, x_{n-2}$ of S outside a suitably small compact neighborhood N_x of x. If $F_x = \{f \in F \mid f(x_i) = 0, i = 1, \cdots, n-2\}$ then F_x is a linear 2-parameter family defined on N_x. Therefore, for any two linearly independent functions f_1, f_2 in F_x,

$$\begin{vmatrix} f_1(y) & f_1(z) \\ f_2(y) & f_2(z) \end{vmatrix} \neq 0 \text{ for } y, z \in N_x, y \neq z .$$

We now apply the theorem.

BIBLIOGRAPHY

1. E. F. Beckenbach, *Generalized convex functions*, Bull, Amer. Math. Soc. **43** (1937), 363-371.
2. S. Bernstein, *Leçons sur les properties extremals et la meilleur approximation des functions analytiques d'une variable réelle*, Paris, Gauthier-Villars, 1926.
3. D. Bratton, *New results in the theory and techniques of Chebyshev fitting*, Abstract No. 546-34, Notices of the Amer. Math. Soc. **5** (1958), 210.
4. H. Busemann, *Note on a theorem of convex sets*, Mathematisk Tidsskrift B (1947), 32-34.
5. P. C. Curtis, Jr. and W. L. Frank, *An algorithm for the determination of the polynomial of best minimum approximation to a function defined on a finite point set*, Jour. Assoc. Comp. Mach. **6** (1959), 395-404.
6. C. J. de la Valle Poussin, *Leçons sur l'approximations des functions d'une variable réelle*, Paris, Gauthier-Villars, 1919.
7. A. Haar, *Die Minkowskische Geometie and die Annäherung an stetige Funcktionen*, Math. Ann. **18** (1918), 294-311.
8. J. C. Mairhuber, *On Harr's theorem concerning Chebyshev approximation problems having unique solutions*, Proc. Am. Math. Soc. **7** (1956), 609-615.
9. T. S. Motzkin, *Sur quelques properiétes caracteristiqutes des ensembles convexes*, Atti. Acad. Naz. Naz. Lincei Rend 6, **21** (1935), 562-567.
10. ———, *Approximation by curves of a unisolvent family*, Bull, Amer. Math. Soc. **55** (1949), 789-793.
11. ——— and J. L. Walsh, *The least pth power polynomials on a finite point set*, Trans. Amer. Math. Soc. **83** (1956), 371-396.
12. ———, *Polynomials of best approximation on a real finite point set I*, Trans. Amer. Math. Soc. **91** (1959), 231-245.

13. E. N. Novodvorskii and I. S. Pinsker, *On a process of equalization of maxima*, Usp. Mat. Nauk **6** (1951), 174-181 (Russian).
14. Ya. L. Remes, *On a method of Chebyshev type approximation of functions*, Ukr. An. 1935.
15. L. Tornheim, *On n-parameter families of functions and associated convex functions*, Trans. Amer. Math. Soc. **69** (1950), 457-467.
16. G. T. Whyburn *Analytic Topology*, Amer Math. Soc. Colloquium Publications, **38**, (1942)

UNIVERSITY OF CALIFORNIA LOS ANGELES AND
YALE UNIVERSITY

PROBLEMS IN SPECTRAL OPERATORS

URI FIXMAN

Introduction. An important problem in the theory of spectral operators in Banach spaces initiated by N. Dunford [5; 6] is that of deciding whether the linear operators of the types encountered in analysis are spectral. Various conditions for spectral operators have been given in [5], but further research is needed in order to apply them to specific cases. J. Schwarz [11] has shown that a class of operators arising from, not necessarily self adjoint, integro-differential boundary-value problems consists of spectral operators. The present investigation originated in a problem on stationary sequences in Banach spaces which led to the study of unitary operators, namely linear isometries of the space onto itself, from this point of view. Accordingly, attention was focused on the class of unitary operators, and the limitations imposed on the operators under study were designed to include it.

Section 1 contains a summary of definitions and results from [5; 6]. A distinction, significant only in non-reflexive spaces, is made between spectral and merely prespectral operators according to the topology in which σ-additivity of the resolutions of the identity is required. As shown in § 2, a resolution of the identity of a prespectral operator uniquely determines the resolutions of the identity of its spectral restrictions. A simple example shows how this can be used to prove that certain operators are not spectral.

Known results are combined in § 3 to yield a necessary condition for spectral operators of scalar type, which involves only the norms of rational functions of the operators. If the space is reflexive and the spectrum an R-set [1, p. 397], the condition is also sufficient. Using the results of § 2 this condition is localised to "cyclic" subspaces generated by single elements. A much more general approach to localization, via the notion of vector measures associated with the operator, is expounded in [3]. It is felt though that the present considerations retain their interest owing to the explicit conditions given. The method of [3] also implies the results of § 2 on restrictions for the case of a reflexive space. Section 3 ends with some characterizations of finite dimensional cyclic subspaces.

The above results are specialized in § 4 to unitary operators which, if the space is reflexive, satisfy all the subsidiary conditions. As a corollary it follows that in a reflexive space a unitary operator is spectral

Received January 26, 1959. This paper is based on the author's doctoral thesis written at the Hebrew University. The author wishes to express his gratitude to Professor A. Dvoretzky for his guidance and encouragement.

if and only if every stationary sequence it generates is spectral.

The final section contains examples of non-spectral unitary operators. It is shown that a unitary operator U in the space of continuous functions defined on a compact Hausdorff space is not spectral provided the homeomorphism determined by U is non-periodic. Using the boundness of the norms of the values of a resolution of the identity, examples are given of non spectral unitary operators in the spaces l_p, $1 \leq p \leq \infty$, $p \neq 2$. The two methods used are combined to show that if in particular the permutation of the basis, determined by a unitary operator in the last mentioned spaces has an infinite "cycle", the operator is not spectral. Examples of non spectral unitary operators in the spaces L_p, $p \neq 2$, (c) and (c_0) follow as corollaries.

1. **Preliminaries.** Let $\mathfrak{X}$ denote a complex B-space and $\mathfrak{B}$ the Boolean algebra of Borel subsets of the complex plane p. A *spectral measure* in $\mathfrak{X}$ is a homomorphism E of $\mathfrak{B}$ onto a Boolean algebra of projections of $\mathfrak{X}$ such that: $E(p) = I =$ identity operator, $E(\phi) = 0$, and $||E(\sigma)|| \leq M < \infty$, M independent of $\sigma \in \mathfrak{B}$. The Boolean operations on commuting projections A, B are defined, as usual, by

$$A \cap B = AB \, , \qquad A \cup B = A + B - AB \, .$$

A spectral measure E in $\mathfrak{X}$ is said to be *of class* Γ in case Γ is a total linear manifold in $\mathfrak{X}^*$ and $x^*E(\ )x$ is σ-additive on $\mathfrak{B}$ for $x \in \mathfrak{X}$, $x^* \in \Gamma$.

Let $B(\mathfrak{X})$ be the B-algebra of bounded linear operators of $\mathfrak{X}$ into itself. If $T \in B(\mathfrak{X})$ and $\mathfrak{Y}$ is a (closed) subspace of $\mathfrak{X}$, we denote by $T \mid Y$ the restriction of T to $\mathfrak{Y}$, and by $\sigma(T)$ and $\rho(T)$ respectively the spectrum and resolvent set of T. Thus, if $\mathfrak{Y}$ is, invariant under T, $\sigma(T \mid Y)$ denotes the spectrum of T considered as an operator in $\mathfrak{Y}$. For $\zeta \in \rho(T)$, $(\zeta - T)^{-1}$ is abbreviated to $T(\zeta)$.

An operator $T \in B(\mathfrak{X})$ is called a *prespectral operator* (of class Γ) in case there exists a spectral measure E of some class Γ such that

$$TE(\sigma) = E(\sigma)T \, , \qquad \sigma(T \mid E(\sigma)\mathfrak{X}) \subseteq \bar{\sigma} \, , \quad \sigma \in \mathfrak{B} \, .$$

E is then called a *resolution of the identity* for T.

An operator in $\mathfrak{B}(\mathfrak{X})$ is called a *spectral operator* if it is prespectral of class $\mathfrak{X}^*$. In this case, E is σ-additive on $\mathfrak{B}$ in the strong operator topology, and the boundness of its range is a consequence of the other requirements [6, p. 325]. A spectral operator T has a unique resolution of the identity E[6, Th. 6]. If $A \in B(\mathfrak{X})$ commutes with T, then it commutes with E[6, Th. 5].

It may also easily be shown that if the bounded subsets of $\mathfrak{X}$ are weakly sequentially conditionally compact, in particular if $\mathfrak{X}$ is reflexive, then every prespectral operator in $\mathfrak{X}$ is spectral.

Let $T \in B(\mathfrak{X})$, $x \in \mathfrak{X}$. By an abuse of language, an $\mathfrak{X}$-valued function f defined and analytic on an open set $D(f) \subseteq p$ is called an *analytic extension* of $T(\zeta)x$ if

$$(\zeta - T)f(\zeta) = x \,, \qquad \zeta \in D(f) \,.$$

$f(\zeta) = T(\zeta)x$ on $D(f) \cap \rho(T)$ for otherwise $(\zeta - T)(f(\zeta) - T(\zeta)x) = x - x = 0$ would imply $\zeta \in \sigma(T)$. Further we have

1.1. THEOREM. *If T is a prespectral operator, and f, g are analytic extensions of $T(\zeta)x$, then $f(\zeta) = g(\zeta)$ for $\zeta \in D(f) \cap D(g)$.* ([6, Th. 2], The further assumption $D(f) \supseteq \rho(T)$, which is made there, is not used in the proof).

Hence there exists a maximal open set which may serve as a domain of definition of an analytic extension of $T(\zeta)x$. This set is called the *resolvent set* of x, and is denoted by $\rho(x)$ (or $\rho_T(x)$, when more then one operator is involved in the discussion). Its complement $\sigma(x)$ (or $\sigma_T(x)$) is called the *spectrum* of x. The maximal analytic extension itself is denoted by $x(\zeta)$ (or $x_T(\zeta)$).

The main use of the concepts above is through the following characterization of spectral subspaces [6, Th. 4]:

1.2. THEOREM. *Let T be a prespectral operator in $\mathfrak{X}$ with a resolution of the identity E, and let $\sigma \subseteq p$ be closed. Then*

$$E(\sigma)\mathfrak{X} = \{x \mid \sigma(x) \subseteq \sigma\} \,.$$

Let E be a spectral measure which vanishes on the complement of a compact set σ, and let f be a complex valued function continuous on σ. Then the Riemann integral $\int_\sigma f(\zeta)E(d\zeta)$ exists in the uniform operator topology [6, Th. 7]. An operator S is said to be of *scalar type* if it is spectral and satisfies

1.3. $$S = \int \zeta E(d\zeta) \left(= \int_{\sigma(S)} \zeta E(d\zeta) \right) ,$$

where E is the resolution of the identity of S [6, Def. 1].

The reader is referred to [4] for the definition and properties of $f(T)$, where $T \in B(\mathfrak{X})$ and f belongs to a certain class of locally holomorphic functions. In the sequel, f will in general be a rational function with poles in $\rho(T)$. If S is of scalar type with the resolution of the identity E, then we have the functional calculus

1.4. $$f(S) = \int f(\zeta)E(d\zeta) \,.$$

We refer to [6] for the general case of a spectral operator.

Finally we shall need the concept of the *cyclic subspace* $[x]$ generated by an element $x \in \mathfrak{X}$. By this is meant the subspace spanned by $\{T(\zeta)x \mid \zeta \in \rho(T)\}$ [5, Def. 1.4]. It has the following properties [5, Lemma 1.5]:

1.5. LEMMA.
1.5.1. $x \in [x]$.
1.5.2. $f(T)[x] \subseteq [x]$.
1.5.3. *If* $y \in [x]$, *then* $[y] \subseteq [x]$.

2. Restrictions of prespectral operators. The following is a generalization of the uniqueness theorem for spectral operators mentioned in § 1.

2.1. THEOREM. *Let T be a prespectral operator in the B-space $\mathfrak{X}$, and let E be a resolution of the identity for T. Let $\mathfrak{Y}$ be a subspace of $\mathfrak{X}$ invariant under T. Then if $T \mid \mathfrak{Y}$ is spectral, its resolution of the identity equals the restriction $E \mid \mathfrak{Y}$ of E to $\mathfrak{Y}$.*

Proof. Let $y \in \mathfrak{Y}$. The function $y_{T\mid\mathfrak{Y}}(\zeta)$ is an analytic extension of $T(\zeta)y$ with domain $\rho_{T\mid\mathfrak{Y}}(y)$. Thus $\rho_{T\mid\mathfrak{Y}}(y) \subseteq \rho_T(y)$, or

$$\sigma_T(y) \subseteq \sigma_{T\mid\mathfrak{Y}}(y) . \tag{2.1.1}$$

Let F denote a resolution of the identity for $T \mid \mathfrak{Y}$. If σ is a closed subset of the complex plane, we have by 1.2

$$\sigma_{T\mid\mathfrak{Y}}(F(\sigma)y) \subseteq \sigma .$$

Therefore, by (2.1.1),

$$\sigma_T(F(\sigma)y) \subseteq \sigma,$$

and again by 1.2

$$E(\sigma)F(\sigma)y = F(\sigma)y . \tag{2.1.2}$$

If τ is a closed set disjoint from σ, we get, operating with $E(\sigma)$ on $E(\tau)F(\tau)y = F(\tau)y$,

$$E(\sigma)F(\tau)y = 0 . \tag{2.1.3}$$

(2.1.3) and the σ-additivity of F in the strong operator topology show that $E(\sigma)F(\sigma')y = 0$ (σ' denotes the complement of σ with respect to p). This together with (2.1.2) gives

$$E(\sigma)y = F(\sigma)y , \qquad \sigma \text{ closed.}$$

The properties of E and F now yield the same equality for every Borel set.

The theorem above shows that invariance of $\mathfrak{Y}$ under E (i.e., under every value of E) is a necessary condition in order that $T \mid \mathfrak{Y}$ be spectral. This condition is by no means automatically fulfilled, and this fact can be used to show that an operator is not spectral:

2.2. EXAMPLE. Let Ω be a compact topological space. We consider the B-space $C(\Omega)$ of all complex valued functions f continuous on Ω with $\|f\| = \max_{\omega \in \Omega} |f(\omega)|$. Let $\mu \in C(\Omega)$, and let S be the operator of multiplication by μ. Heuristically, S cannot in general be spectral because projections which "ought" to belong to the resolution of the identity are not members of $B(C(\Omega))$. This is made precise as follows. Let T be the extension of the multiplication to the space $\mathfrak{X} = M(\Omega)$ of complex valued functions f bounded on Ω with $\|f\| = \sup |f(\omega)|$. T is prespectral with a resolution of the identity: $E(\sigma)$ is the multiplication by $\chi_\sigma(\mu(\ ))$, where χ_σ is the characteristic function of σ. The σ-additivity may be verified with respect to the total linear manifold generated by the functionals x_ω^*, $\omega \in \Omega$, defined by $x_\omega^* x = x(\omega)$, $x \in M(\Omega)$. To see that $\sigma(T \mid E(\sigma)\mathfrak{X}) \subseteq \bar{\sigma}$, observe that if $\zeta \in \bar{\sigma}'$, $(T \mid E(\sigma)\mathfrak{X})(\zeta)$ is the multiplication by $\chi_\sigma(\mu(\ ))(\zeta - \mu)^{-1}$ (here $0/0 = 0$). We omit the details. Now suppose, for instance, that μ is not constant on a connected component of Ω, and that ω_1, ω_2 are two points in the component such that $\mu(\omega_1) \neq \mu(\omega_2)$. Then taking $\sigma = \{\mu(\omega_1)\}$ we see that $E(\sigma)$ does not leave $C(\Omega)$ invariant. Hence $S = T \mid C(\Omega)$ is not spectral.

The next theorem is a partial converse of Theorem 2.1. We need two lemmas.

2.3. LEMMA. *Let T be a prespectral operator in the B-space $\mathfrak{X}$, and let $A \in B(\mathfrak{X})$ commute with T. If $x \in \mathfrak{X}$, then $\sigma(Ax) \subseteq \sigma(x)$ and $(Ax)(\zeta) = Ax(\zeta)$, $\zeta \in \rho(x)$.*

Proof. For $\zeta \in \rho(x)$, $(\zeta - T)Ax(\zeta) = A(\zeta - T)x(\zeta) = Ax$. The conclusion follows by the definition of $\sigma(Ax)$ and 1.1.

2.4. LEMMA. *If T is prespectral in $\mathfrak{X}$, $x \in \mathfrak{X}$ and τ is a connected component of $\rho(x)$ such that $\tau \cap \rho(T) \neq \phi$, then $x(\zeta) \in [x]$, $\zeta \in \tau$.*

Proof. Since $\rho(x)$ is open in the complex plane, τ has the same property and is therefore a region. Let $x^* \in \mathfrak{X}$ vanish on $[x]$. For $\zeta \in \rho(T)$, $x(\zeta) = T(\zeta)x \in [x]$; thus $f(\zeta) = x^*x(\zeta)$ vanishes on the open subset $\tau \cap \rho(T)$ of τ. Being regular, f vanishes identically on τ. A well known corollary of the Hahn-Banach extension theorem yields the conclusion.

It may also be shown that $\{\zeta \in \rho(x) | x(\zeta) \in [x]\}$ is open and closed in $\rho(x)$. If $\rho(T)$ is dense in the plane, then $x(\zeta) \in [x]$ for every $\zeta \in \rho(x)$ [5, Lemma 1.5.3]. Cf. however Example 2.6 below.

2.5. THEOREM. *Let T be a prespectral operator in $\mathfrak{X}$ with a resolution of the identity E. Let $\mathfrak{Y}$ be a subspace of $\mathfrak{X}$ invariant under $T(\zeta)$, $\zeta \in \rho(T)$, and under E. Then $T | \mathfrak{Y}$ is prespectral with a resolution of the identity $E | \mathfrak{Y}$. If T is spectral or spectral of type m (v. [6, p. 336]), $T | Y$ has the same property.*

Proof. Since $T = \frac{1}{2\pi i}\int_C T(\zeta)d\zeta$, where C is a circle containing $\sigma(T)$ in its interior and the integral is in Riemann's sense and in the uniform operator topology, $\mathfrak{Y}$ is invariant under T, and $T | \mathfrak{Y}$ is well defined. If T is spectral, we may assume invariance under T instead of under $T(\zeta)$, $\zeta \in \rho(T)$, using [6, Lemma 3].

All the assertions of the theorem are easily verified, except: For every $\sigma \in \mathfrak{B}$, $\sigma((T|\mathfrak{Y})|(E|Y)(\sigma)\mathfrak{Y}) = \sigma(T|E(\sigma)\mathfrak{Y}) \subseteq \bar{\sigma}$. We have to show that if $\zeta \in \bar{\sigma}'$, then $\zeta - T$ induces a one-to-one mapping of $E(\sigma)\mathfrak{Y}$ onto itself. Since $\sigma(T | E(\sigma)\mathfrak{X}) \subseteq \bar{\sigma}$, there is no $z \neq 0$ in $E(\sigma)\mathfrak{X}$ and hence in $E(\sigma)\mathfrak{Y}$ such that $(\zeta - T)z = 0$. It remains to show that the range of $(\zeta - T)|E(\sigma)\mathfrak{Y}$ is $E(\sigma)\mathfrak{Y}$. Let $z \in E(\sigma)\mathfrak{Y}$. Then $E(\sigma)z = z$, hence $E(\bar{\sigma})z = z$, and therefore by 1.2 $\sigma(z) \subseteq \bar{\sigma}$. Therefore $\zeta \in \rho(z)$, and since $(\zeta - T)z(\zeta) = z$ it suffices to show that $z(\zeta) \in E(\sigma)\mathfrak{Y}$. Let π be an open half plane with ζ on its boundary. From 1.2 it follows that $\sigma(E(\pi')z) \subseteq \pi' \cup \sigma(z)$, and therefore $\{\zeta\} \cup \pi \subseteq \rho(E(\pi')z)$. Since $\rho(E(\pi')z)$ is open, it follows that ζ belongs to a component of $\rho(E(\pi')z)$ which contains arbitrarily distant points of the complex plane and thus points of $\rho(T)$. 2.4 now implies $(E(\pi')z)(\zeta) \in [E(\pi')z]$. The assumptions of the invariance of $\mathfrak{Y}$ show that $[E(\pi')z] \subseteq \mathfrak{Y}$. Therefore $(E(\pi')z)(\zeta) \in \mathfrak{Y}$. But by 2.3, we have $(E(\pi')z)(\zeta) = E(\pi')z(\zeta)$; therefore

$$E(\pi')z(\zeta) \in \mathfrak{Y} .$$

Similarly one shows $E(\pi)z(\zeta) \in \mathfrak{Y}$. Therefore $z(\zeta) = E(\pi')z(\zeta) + E(\pi)z(\zeta) \in \mathfrak{Y}$. On the other hand, $E(\sigma)z = z$ implies by 2.3 $E(\sigma)z(\zeta) = z(\zeta)$. Therefore $z(\zeta) \in E(\sigma)\mathfrak{Y}$ as required.

It follows from the proof above that, under the conditions of the theorem, $z \in Y$ implies $z(\zeta) \in \mathfrak{Y}$, $\zeta \in \rho(z)$. The following example shows that without invariance of $\mathfrak{Y}$ under E, this need not hold even if T is a normal operator in Hilbert space. This, in turn, amplifies Example 2.2 by showing that even if T is spectral, and not merely prespectral, $\mathfrak{Y}$ is not necessarily invariant under E.

2.6. EXAMPLE. Let $\mathfrak{X}$ be the Hilbert space $L^2(\Omega)$, where Ω is the disc $\{\omega \mid |\omega| \leq 1\}$ in the complex plane. Let T be the operator of multiplication by ω. Then T is a bounded normal operator and spectral. We define $x \in \mathfrak{X}$ by

$$x[\omega] = \begin{cases} 1 & \text{if } \frac{1}{2} \leq |\omega| \leq 1 \\ 0 & \text{if } \quad |\omega| < \frac{1}{2} . \end{cases}$$

The maximal analytic extension of $T(\zeta)x$, $x(\zeta)$, exists for ζ not in the ring $\sigma(x) = \{\zeta \mid \frac{1}{2} \leq |\zeta| \leq 1\}$, and then

$$x(\zeta)[\omega] = \begin{cases} \dfrac{1}{\zeta - \omega} & \text{if } \frac{1}{2} \leq |\omega| \leq 1 \\ 0 & \text{if } \quad |\omega| < \frac{1}{2} . \end{cases}$$

We consider the subspace $\mathfrak{Y} = [x]$, which is invariant under $T(\zeta)$, $\zeta \in \rho(T)$ by 1.5.2, and contains x by 1.5.1. $[x]$ is the closure of the finite linear combinations of the functions $T(\zeta)x = x(\zeta)$ for $\zeta \in \rho(T) = \Omega'$.

Now, suppose that for a fixed ζ, $|\zeta| < \frac{1}{2}$, $x(\zeta)$ were approximable by these linear combinations in the Hilbertian norm. Since all these functions are holomorphic in $\sigma(x)$, $x(\zeta)$ would be uniformly approximable by the linear combinations on a closed ring τ concentric with and inner to $\sigma(x)$ [13, p. 96]. But this is impossible, since the approximants are rational functions with poles in the unbounded component of τ', while the only analytic continuation of $x(\zeta) \mid \tau$ to the other component is $1/(\zeta - \omega)$, which is not regular at ζ (v. [13, p. 25, Th. 16]).

It may also be directly shown that there exist x and σ such that $E(\sigma)x \notin [x]$.

We now give an example to show that the assumption that $T \mid \mathfrak{Y}$ is spectral cannot be dropped in Theorem 2.1 even if $\mathfrak{Y} = \mathfrak{X}$; i.e., a pre-spectral operator may have more than one resolution of the identity.

2.7. Example. We specialize Example 2.2, retaining its notation. We take for Ω the set of positive integers. Thus $\mathfrak{X} = M(\Omega)$ is the space usually denoted by (m). For μ we chose a function belonging to $\mathfrak{X}$ which satisfies

(2.7.1) $$\mu(1) = 1 ;$$

(2.7.2) $$\mu(j) \neq 1 , \qquad j > 1 ;$$

(2.7.3) $$\lim_j \mu(j) = 1 .$$

As is well known [2, p. 34], there exists a real bounded linear functional $\lim_R$, defined on the space $(m)_R$ of all real bounded sequences, which has the following properties:

$(2.7.4)_R$ If $x, y \in (m)_R$ and $y(j) = x(j+1)$, $\qquad j = 1, 2, \cdots,$

then $\lim_R y = \lim_R x$;

$$(2.7.5)_R \qquad \underline{\lim}\, x(j) \leq \lim_R x \leq \overline{\lim}\, x(j)\ .$$

We define a functional lim on $\mathfrak{X}$ by $\lim x = \lim_R x' + i \lim_R x''$, where $x = x' + ix''$, $x', x'' \in (m)_R$. Evidently, lim is a bounded linear functional which enjoys the property (2.7.4) analogous to $(2.7.4)_R$. Further we have (T defined as in 2.2)

$$(2.7.6) \qquad \lim Tx = \lim x\ .$$

To see this, we write $(Tx)(j) = (\mu(j) - 1)x(j) + x(j)$. By the linearity of lim, it suffices to show that $\alpha(j) \to 0 (\alpha(j) = \mu(j) - 1)$ implies $\lim \alpha x = 0$. This follows from $(2.7.5)_R$ on separating α and x into their real and imaginary parts. We define an operator $A \in B(\mathfrak{X})$ by $Ax = \lim x \cdot x_0$, where $x_0(j) = \delta_{1j}$ (Kronecker's symbol). Using (2.7,1), (2.7.6) we get $TA = AT$. On the other hand, A does not commute with E (defined in 2.2). Taking $\sigma = \{1\}$ we have, using (2.7.2), (2.7.4), $AE(\sigma)x = 0$ while $E(\sigma)Ax = \lim x \cdot x_0$. Hence the function F, defined by

$$F(\sigma) = E(\sigma) + AE(\sigma) - E(\sigma)A\ , \qquad \sigma \in \mathfrak{B}\ ,$$

differs from E. We show that F is a resolution of the identity for T. A straightforward calculation, based on the fact E is a spectral measure, shows that F is a spectral measure (In verifying that $F(\sigma)F(\delta) = F(\sigma \cap \delta)$, one uses the fact that $AE(\tau)A = 0$, $\tau \in \mathfrak{B}$). F is σ-additive with respect to the total linear manifold generated by the functionals x_j^* $(x_j^* x = x(j))$, $j \geq 2$ and $x^* = x_1^* - \lim$; since $x_j^* F(\sigma)x = x_j^* E(\sigma)x$ for $j \geq 2$, while $x^* F(\sigma)x = \chi_\sigma(1)(x(1) - \lim x)$. Since T commutes with E and A, T commutes with F. Finally, to see that $\sigma(T \mid F(\sigma)\mathfrak{X}) \subseteq \bar{\sigma}$, we assert that the restriction of $(T \mid E(\bar{\sigma})\mathfrak{X})(\zeta)$ to $F(\sigma)\mathfrak{X}$, $\zeta \in \bar{\sigma}'$, is an inverse of $(\zeta - T) \mid F(\sigma)\mathfrak{X}$. As shown in the proof of [6, Th. 5], the prespectrality of T implies $E(\bar{\sigma})AE(\bar{\sigma}) = AE(\bar{\sigma})$. Hence $E(\bar{\sigma})AE(\sigma) = AE(\sigma)$, whence it follows that $E(\bar{\sigma})F(\sigma) = F(\sigma)$. Therefore $F(\sigma)\mathfrak{X} \subseteq E(\bar{\sigma})\mathfrak{X}$, and the mentioned restriction is well defined. Let $x \in F(\sigma)\mathfrak{X}$. Then $\sigma(x) \subseteq \bar{\sigma}$, by 1.2, since $x \in E(\bar{\sigma})\mathfrak{X}$. Further, 1.1 and 2.3 imply

$$(T \mid E(\bar{\sigma})\mathfrak{X})(\zeta)x = x(\zeta) = (F(\sigma)x)(\zeta) = F(\sigma)x(\zeta)\ .$$

Thus the range of the restriction is included in $F(\sigma)\mathfrak{X}$. The truth of our assertion is now evident.

3. Conditions for operators of scalar type. If $T \in B(\mathfrak{X})$, the *full algebra* generated by T, denoted by $\mathfrak{A}(T)$, is the smallest subalgebra of $B(\mathfrak{X})$ which is closed in the norm topology of $B(\mathfrak{X})$, which is inverse-closed and which contains T and I [6, Def. 5]. Let σ be a compact subset of the complex plane. We denote by $R(\sigma)$ the set of rational functions regular on σ. $CR(\sigma)$ denotes the closure of $R(\sigma)$ in $C(\sigma)$. Following [1, p. 397], a compact nowhere dense set σ in the complex plane is called an *R-set* if and only if $CR(\sigma) = C(\sigma)$. For properties of R-sets used in the sequel see [1, p. 398] and the references there given.

3.1. THEOREM. *Let $S \in B(\mathfrak{X})$, then the following equivalent conditions are necessary in order that S be of scalar type*:

3.1.1. *There exists a constant $H < \infty$ such that for every $f \in R(\sigma(S))$*

$$\|f(S)\| \leq H \max_{\zeta \in \sigma(S)} |f(\zeta)| = H \|f(S)\|_{sp} = H \lim_n \|f(S)^n\|^{1/n} .$$

3.1.2. *There exists a constant $K < \infty$ such that for every $f \in R(\sigma(S))$*

$$\|f(S)\|^2 \leq K \|f(S)^2\| .$$

If $\mathfrak{X}$ is reflexive and $\sigma(S)$ is an R-set, each of the mentioned conditions is sufficient. Each of the following conditions implies 3.1.1:

3.1.3. *For every $x \in \mathfrak{X}$ there exists a constant $H(x)$ (independent of f) such that for every $f \in R(\sigma(S))$*

$$\|f(S)x\| \leq H \max_{\zeta \in \sigma(S)} |f(\zeta)| \cdot \|x\| = H(x) \|f(S)\|_{sp} \|x\| .$$

3.1.4. *The same; with $h(x)$, $f \in R(\sigma(S \mid [x]))$ and*

$$\|f(S \mid [x]) \leq h(x) \max_{\zeta \in \sigma(S \mid [x])} |f(\zeta)| = h(x) \|f(S \mid [x])\|_{sp}$$

3.1.5. *The same; with $k(x)$, $f \in R(\sigma(S \mid [x]))$ and*

$$\|f(S \mid [x])\|^2 \leq k(x) \|f(S \mid [x])^2\| .$$

3.1.3 *is implied by* 3.1.1. 3.1.4 *and* 3.1.5 *are necessary if S is of scalar type and satisfies the following condition*:

3.1.6. *If E is the resolution of the identity of S, $x \in \mathfrak{X}$ and $\sigma \in \mathfrak{B}$, then*

$$E(\sigma)x \in [x] .$$

Proof. For the equivalence of 3.1.1 and 3.1.2 see [9, p. 78] and for the necessity see the beginning of the proof of [6, Th. 13]. If one of them holds, then $\mathfrak{A}(S)$ is equivalent to $CR(\sigma(S))$, hence if $\sigma(S)$ is an R-set, to $C(\sigma(S))$. Therefore if $\mathfrak{X}$ is reflexive, S is of scalar type by [6, Th. 18 (IV)]. Since, from 1.5.2, $\sigma(S \mid [x]) \subseteq \sigma(S)$ and $\|f(S)x\| \leq \|f(S \mid [x])\| \cdot \|x\|$, the equivalent conditions 3.1.4, 3.1.5 imply 3.1.3. The proof that 3.1.3 implies 3.1.1 is much like the proof of the uniform boundness theorem. 3.1.3 and Baire's category theorem imply that at least one of the sets

$$G_j = \{x \in \mathfrak{X} \mid \|f(S)x\| \leq j \|f(S)\|_{sp} \|x\|\ ,\ f \in R(\sigma(S))\} \quad j = 1, 2, \cdots,$$

let it be the nth, contains a sphere $\{x \in \mathfrak{X} \mid \|x - x_0\| < r\}$, $r > 0$. 3.1.1 then easily follows with $H = n(2\|x_0\| + r)/r$. If S is of scalar type and satisfies 3.1.6, then every $[x]$ is invariant under E (because if $y \in [x]$, then $E(\sigma)x \in [y] \subseteq [x]$ by 1.5.3) and $S(\zeta)$ (by 1.5.2). Therefore, by 2.5, $S \mid [x]$ is of scalar type, and the necessity of 3.1.4, 3.1.5, which are 3.1.1, 3.1.2 for $S \mid [x]$, follows.

REMARKS. In case the conclusion of 1.2 holds, it may be convenient to replace $\sigma(S \mid [x])$ by $\sigma(x)$ in 3.1.4, 3.1.5. One always has $\sigma(x) \subseteq \sigma(S \mid [x])$. By slight modifications in the proof of [5, Lemma 1.10], one shows that, provided S is spectral, $\sigma(x) = \sigma(S \mid [x])$ (for every x) if and only if for every x and $\zeta \in \rho(x)$, $x(\zeta) \in [x]$. As remarked after 2.5, this is the case if 3.1.6 holds.

Taking S as in 2.2, 3.1.1 is obviously fulfilled. By an appropriate choice of Ω and μ, we may achieve that S is not spectral although $\sigma(S) =$ range of μ is an R-set. Thus these conditions fail to assure scalarity if $\mathfrak{X}$ is not reflexive.

We conclude the present section with some characterizations of finite dimensional cyclic subspaces.

3.2. THEOREM. *If S is of scalar type, satisfies* 3.1.6 *and* $x \in \mathfrak{X}$, *then the following conditions are equivalent*:

3.2.1. $[x]$ *is of finite dimension.*

3.2.2. $\mathfrak{A}(S)x$ *is of the second category in* $[x]$ (*or* $x = 0$).

3.2.3. *For each* $y \in [x]$ *there exists a* $U(y) \in B(\mathfrak{X})$, *commuting with S, such that* $U(y)x = y$.

3.2.4. *For each* $y \in [x]$ *there exists a* $V(y) \in B([x])$, *commuting with* $S \mid [x]$, *such that* $V(y)x = y$.

3.2.5. $\sigma(x)$ *is finite* (*equivalent to* 3.2.1 *by mere scalarity*).

Proof. Evidently we may assume $x \neq 0$. 3.2.1 $\Rightarrow$ 3.2.2 and 3.2.3: Since $\{f(S) \mid f \in R(\sigma(S))\}$ is dense in $\mathfrak{A}(S)$, $\mathfrak{A}(S)x$ is a dense linear submanifold of $[x]$. By 3.2.1, $\mathfrak{A}(S)x$ is of finite dimension; hence closed. Therefore $\mathfrak{A}(S)x = [x]$, whence 3.2.2 and 3.2.3 follow.

3.2.2 or 3.2.3 $\Rightarrow$ 3.2.4: Under either hypothesis the set

$$Z = \{z = U(z)x \mid U(z) \in B([x]), U(z)S = SU(z)\}$$

is of the second category in $[x]$. Suppose $f_n \in R(\sigma(S))$, $\|f_n(S)x\| = 1$

and $z \in Z$. Then $\{f_n(S)z\}$ is bounded since

$$\|f_n(S)z\| = \|f_n(S)U(z)x\| = \|U(z)f_n(S)x\| \\ \leq \|U(z)\| \|f_n(S)x\| = \|U(z)\| .$$

Therefore, by the uniform boundness theorem, $\{\|f_n(S) | [x]\|\}$ is bounded. Hence, if $f_n \in R(\sigma(S))$, $\|f_n(S)x\| = 1$ and $y \in [x]$, then $\{\|f_n(S)y\|\}$ is bounded. This shows that there exists a constant $c(y) < \infty$ such that $\|f(S)y\| \leq c(y)\|f(S)x\|$, $f \in R(\sigma(S))$. We define the transformation $V(y)$ on $\{f(S)x \,|\, f \in R(\sigma(S))\}$ by

$$V(y)f(S)x = f(S)y .$$

$V(y)$ is bounded by $c(y)$ on a dense linear submanifold of $[x]$. Therefore it is uniquely defined, and can be extended by continuity to a bounded operator on $[x]$. Evidently, this operator satisfies our requirements.

3.2.4 $\Rightarrow$ 3.2.5: We first show that for each $y \in [x]$ there exists a constant $c(y)$ such that

$$\|E(\sigma)y\| \leq c(y)\|E(\sigma)x\| , \qquad \sigma \in \mathfrak{B} .$$

As in the proof of 3.1, $S | [x]$ is of scalar type with the resolution of the identity $E | [x]$. By the commutativity theorem, mentioned in § 1, $E | [x]$ commutes with $V(y)$. Therefore for every Borel set σ

$$\|E(\sigma)y\| = \|E(\sigma)V(y)x\| = \|(E(\sigma) | [x])V(y)x\| \\ = \|V(y)(E(\sigma) | [x])x\| \leq \|V(y)\| \|E(\sigma)x\| .$$

This proves our statement. Hence, if we define

$$G_j = \{y \in [x] \,|\, \|E(\sigma)y\| \leq j\|E(\sigma)x\|, \sigma \in \mathfrak{B}\} , \qquad j = 1, 2, \cdots,$$

we have $\bigcup_j G_j = [x]$. Since the G_j's are closed, it follows by the usual category argument that there exists a constant $c < \infty$ such that

$$\|E(\sigma) | [x]\| \leq c\|E(\sigma)x\| , \qquad \sigma \in \mathfrak{B}.$$

Since the norm of a non null projection is at least 1, it follows that

$E(\sigma_n)x \to 0$, $\sigma_n \in \mathfrak{B} \Rightarrow$ There exists an n_0 such that $E(\sigma_n) | [x] = 0$ for $n \geq n_0$.

Now, suppose $\sigma(x)$ were infinite. Then we could represent it in the form $\sigma(x) = \bigcup_{n=0}^{\infty}\sigma_n$, where the σ_n are pairwise disjoint, $\sigma_0 \in \mathfrak{B}$ and σ_n, $n \leq 1$ are non void sets open relative to $\sigma(x)$ (we omit the easy proof). From the σ-additivity of E in the strong operator topology it follows that $E(\sigma_n)x \to 0$. Hence, by what was proved above, there exists an $m \geq 1$ such that $E(\sigma_m)x = 0$. $\sigma_m = \sigma(x) \cap \tau$, where τ is open in the complex plane. We have

$$E(\tau)x = E(\tau)E(\sigma(x))x = E(\tau \cap \sigma(x))x = E(\sigma_m)x = 0 .$$

Therefore $E(\tau')x = x$. Since τ' is closed, 1.2 implies $\sigma(x) \subseteq \tau'$. Thus we get $\sigma_m = \sigma(x) \cap \tau = \phi$, contradicting the choice of σ_m.

3.2.5 $\Rightarrow$ 3.2.1: Since we assumed $x \neq 0$, we have $\sigma(x) \neq \phi$. Let $\sigma(x) = \{\zeta_1, \cdots, \zeta_r\}$. If $y \in [x]$ there exist $f_n \in R(\sigma(S))$ such that $f_n(S)x \to y$. By 1.4, $f_n(S) = \int f_n(\zeta)E(d\zeta)$. Using Riemann's sums approximating the integral, we get

$$f_n(S)E(\sigma(x)) = \sum_{j=1}^{r} f_n(\zeta_j)E(\{\zeta_j\}) .$$

But $f_n(S)E(\sigma(x))x = f_n(S)x$; therefore

$$(*) \qquad \sum_{j=1}^{r} f_n(\zeta_j)E(\{\zeta_j\})x \to y .$$

Now

$$(**) \qquad E(\{\zeta_j\})x,\ j = 1, \cdots, r \text{ are linearly independent:}$$

If $\sum_{j=1}^{r}\alpha_j E(\{\zeta_j\})x = 0$, then operating with $E(\{\zeta_k\})$, we get $\alpha_k E(\{\zeta_k\})x = 0$. But $E(\{\zeta_k\})x \neq 0$ for otherwise

$$x = E(\sigma(x))x = E(\sigma(x) - \{\zeta_k\})x + E(\{\zeta_k\})x = E(\sigma(x) - \{\zeta_k\}x$$

would imply by 1.2 the contradiction $\sigma(x) \subseteq \sigma(x) - \{\zeta_k\}$. Therefore $\alpha_k = 0$. From (**) and (*) it follows by a well known argument that the sequences $\{f_n(\zeta_j)\}_{n=1}^{\infty}$ are bounded; hence compact. Therefore there exists a subsequence $\{n_k\}$ of the indices such that $f_{n_k}(\zeta_j) \to \alpha_j$, $j = 1, \cdots, r$. So

$$y = \sum_{j=1}^{r} \alpha_j E(\{\zeta_j\})x .$$

The vectors $E(\{\zeta_j\}x$, $j = 1, \cdots, r$, are independent of y, and thus span $[x]$.

4. Applications to unitary operators. To render the results of § 3 conveniently applicable, one should know beforehand of an operator that if it is spectral, it is of scaler type and satisfies Condition 3.1.6. We shall show that this is the case for a class of operators which includes the unitary operators in reflexive spaces. We lean heavily on [5]; and although some familiarity with this paper is assumed in the present section, it will be convenient to cite the pertinent definitions.

4.1. DEFINITION. Let the spectrum $\sigma(T)$ of an operator $T \in B(\mathfrak{X})$ lie in a closed rectifiable Jordan curve Γ_0. Suppose that Γ_0 is embedable in a family Γ_δ, $-\delta_0 \leq \delta \leq \delta_0$ $(0 < \delta_0 \leq \frac{1}{2})$, of closed rectifiable Jordan curves which satisfies the following conditions: Γ_{δ_1} is interior to Γ_{δ_2} for $-\delta_0 \leq \delta_1 < \delta_2 \leq \delta_0$. The curve Γ_δ is defined by a function $\zeta(\lambda, \delta)$, $-1 \leq \lambda \leq 1$, with $\zeta(-1, \delta) = \zeta(1, \delta)$. As λ increases from -1 to 1, the point $\zeta(\lambda, \delta)$ traces Γ_δ in a counterclockwise direction. For different values of λ, the arcs $\zeta(\lambda, \delta)$, $-\delta_0 \leq \delta \leq \delta_0$ do not intersect. They are rectifiable, and $|\delta|$ is the length of the subarc with endpoints $\zeta(\lambda, 0)$ and $\zeta(\lambda, \delta)$. Under these assumptions a nonnegative integer-valued function $\nu(\lambda)$ satisfying the condition

$$\| \delta^{\nu(\lambda)} T(\zeta(\lambda, \delta)) \| \leq 1 , \quad 0 < |\delta| < \delta_0, \quad -1 \leq \lambda \leq 1 ,$$

is called an *index function* for T.

4.2. THEOREM. *If U is a unitary spectral operator, it is of scalar type.*

Proof. This is essentially proved in [5]: It is easy to show that the spectrum of U lies in the unit circle and that if we embed the unit circle in the family of circles Γ_δ, $-\frac{1}{2} \leq \delta \leq \frac{1}{2}$, defined by $\zeta(\lambda, \delta) = (1 + \delta)e^{\pi i \lambda}$, $-\lambda \leq \lambda \leq 1$, then $\nu(\lambda) \equiv 1$ is an index function for U. Since $\zeta(\lambda, \delta)$ has continuous second partial derivatives, and the assumptions of [5, Lemma 3.16] hold, it follows from [5, Lemma 3.18] that $\int_{\sigma(U)} (U - \zeta)E(d\zeta) = 0$ or $U = \int \zeta E(d\zeta)$.

4.3. LEMMA. *Let $S \in B(\mathfrak{X})$ be spectral with index function $\nu(\lambda) \equiv 1$ with respect to $\zeta(\lambda, \delta)$ which has continuous second partial derivatives. Let $\mathfrak{X}$ be reflexive. Then $E(\{\zeta\})x \in [x]$, $x \in \mathfrak{X}$, $\zeta \in \Gamma_0$.*

Proof. Let $\zeta_0 \in \Gamma_0$. Then ζ_0 is of the form $\zeta_0 = \zeta(\lambda_0, 0)$. It is shown in the proof of [5, Th. 3.12 (III)] that there is a $y \in \mathfrak{X}$ and a sequence $\delta_n \to 0$ such that for $\zeta_n = \zeta(\lambda_0, \delta_n)$ we have

$$(4.3.1) \qquad (\zeta_n - \zeta_0)S(\zeta_n)x(\rightarrow)y .$$

Further, (4.3.2) $(\zeta_0 - S)y = 0$,

$$(4.3.3) \qquad x - y \in \overline{(\zeta_0 - S)\mathfrak{X}} .$$

$y \in [x]$ since, by (4.3.1), it is a weak limit of vectors in $[x]$, hence a strong limit of their linear combinations [2, p. 134. Th. 2]. (4.3.2) implies, by [6, Lemma 1], $E(\{\zeta_0\})y = y$. By (4.3.3), there exist z_n such that $(\zeta_0 - S)z_n \to y - x$, and by [5, Lemma 3.17] $E(\{\zeta_0\})(\zeta_0 - S) = 0$; therefore $E(\{\zeta_0\})(y - x) = 0$. It follows that $E(\{\zeta_0\})x = y \in [x]$.

4.4. LEMMA. *Under the hypotheses of* 4.3, *if $\zeta, \xi \in \Gamma_0$, $\zeta \neq \xi$ and*

$x \in \mathfrak{X}$, then there exist $z_k \in [x]$ such that $E(\{\zeta'\})E(\{\xi'\})x = \lim_k (S-\zeta)^2 \cdot (S-\xi)^2 z_k$.

Proof. Let $\zeta_0, \xi_0 \in \Gamma_0$, $\zeta_0 \neq \xi_0$. Then, by 1.5.1, 1.5.3 and 4.3, $u = E(\{\zeta_0\}')E(\{\xi_0\}')x \in [x]$. Therefore by 1.5.3 it is sufficient to establish the representation for u with $z_k \in [u]$. The argument follows closely part of the proof of [5, Lemmas 2.6, 2.10]. As in the proof of 4.3, there exist $\zeta_n \to \zeta_0$ such that

$$(\zeta_n - \zeta_0)S(\zeta_n)u(\rightarrow)E(\{\zeta_0\})u = 0 .$$

Thus

$$(\zeta_0 - S)S(\zeta_n)u = (\zeta_0 - \zeta_n)S(\zeta_n)u + u(\rightarrow)u .$$

Since $S(\zeta_n)u \in [u]$ and since weak convergence to u implies strong convergence of linear combinations, it follows that there exist $u_k \in [u]$ such that

$$(\zeta_0 - S)u_k \to u . \tag{4.4.1}$$

Operating on u_k with the identity

$$(\zeta_0 - S)^2 S(\zeta_n) = (\zeta_0 - \zeta_n)^2 S(\zeta_n) + (\zeta_0 - \zeta_n) + (\zeta_0 - S)$$

and letting n tend to infinity, we get

$$(\zeta_0 - S)u_k = \lim_n (\zeta_0 - S)^2 S(\zeta_n)u_k . \tag{4.4.2}$$

But $S(\zeta_n)u_k \in [u_k] \subseteq [u]$, hence (4.4.1), (4.4.2) show that there are $v_k \in [u]$ such that

$$(\zeta_0 - S)^2 v_k \to u . \tag{4.4.3}$$

Operating on (4.4.3) with $E(\{\xi_0\}')$, we get

$$(\zeta_0 - S)^2 E(\{\xi_0\}')v_k \to u . \tag{4.4.4}$$

But by 4.3 $E(\{\xi_0\}')v_k \in [v_k] \subseteq [u]$; therefore, by what has been proved thus far, $E(\{\xi_0\}')v_k$ is of the form

$$E(\{\xi_0\}')v_k = \lim (\xi_0 - S)^2 v_{kn} , \quad v_{kn} \in [u] . \tag{4.4.5}$$

From (4.4.4), (4.4.5) our lemma follows.

4.5. THEOREM. *If S is a spectral operator which satisfies the assumptions of* 4.3, *in particular if S is a spectral unitary operator in a reflexive space, then it satisfies Condition* 3.1.6.

Proof (After [5, Th. 2.11]). Since $E(\sigma) = E(\sigma \cap \Gamma_0)$, $\sigma \in \mathfrak{B}$, and since E is σ-additive in the strong operator topology, it suffices to show that $E(\sigma)x \in [x]$ for σ the closed proper subarcs of Γ_0. Let $\zeta = \zeta(\lambda, 0)$, $\xi = \zeta(\mu, 0)$, $\lambda \neq \mu$, be the ends of the arc

$$[\zeta, \xi] = \{\zeta(\alpha, 0) \mid \lambda \leq \alpha \leq \mu \text{ if } \lambda < \mu;\ \alpha \notin (\mu, \lambda) \text{ if } \mu < \lambda\} .$$

We show that $E([\zeta, \xi])x \in [x]$ (the case $\lambda = \mu$ cared for by 4.3). Since $I = E(\{\zeta\}) + E(\{\xi\}) + E(\{\zeta\}')E(\{\xi\}')$, we have

$$E([\zeta, \xi])x = E(\{\zeta\})x + E(\{\xi\})x + E([\zeta, \xi])E(\{\zeta\}')E(\{\xi\}')x .$$

By 4.3 we have to show that $E(\zeta, \xi])u \in [x]$, where $u = E(\{\zeta\}')E(\{\xi\}')x$. But by 4.4 there exists a sequence $z_k \in [x]$ such that

$$E([\zeta, \xi])u = \lim_k E([\zeta, \xi])(S - \zeta)^2(S - \xi)^2 z_k .$$

Thus we have only to show that $z \in [x]$ implies

$$E([\zeta, \xi])(S - \zeta)^2(S - \xi)^2 z \in [x] .$$

Let $\zeta_n = \zeta(\lambda_n, 0)$, $\xi_n = (\mu_n, 0)$, where the sequences $\lambda_n \to \lambda$, $\mu_n \to \mu$ are so chosen that if $\lambda < \mu$ then $\lambda_n < \lambda < \mu < \mu_n$, while if $\mu < \lambda$ then $\mu < \mu_n < \lambda_n < \lambda$. It is shown during the proof of [5, Th. 2.4] that, since S has 1 as an index function, $(S - \zeta)^2(S - \xi)^2$ is of the form

$$(S - \zeta)^2(S - \xi)^2 = \lim_n (I(\lambda, \mu) + I(\mu_n, \lambda_n)) , \tag{4.5.1}$$

where $I(\alpha, \beta)$, $-1 \leq \alpha, \beta \leq 1$, $\alpha \neq \beta$, are certain operators, the manner of definition of which is explained in [5, Lemma 2.4], which enjoy the properties:

(4.5.2) $I(\alpha, \beta)[x] \subseteq [x]$ ($I(\alpha, \beta)$ being a line integral of $S(\zeta)$).

(4.5.3) $\sigma(I(\alpha, \beta)y) \subseteq [\zeta(\alpha, 0), \zeta(\beta, 0)]$, $y \in \mathfrak{X}$ [5, Lemma 2.4].

Let $z \in [x]$. Then, by (4.5.1),

$$\begin{aligned} E([\zeta, \xi])&(S - \zeta)^2(S - \xi)^2 z \\ &= \lim_n (E([\zeta, \xi]I(\lambda, \mu)z + E([\zeta, \xi])I(\mu_n, \lambda_n)z) . \end{aligned} \tag{4.5.4}$$

But by (4.5.3) $\sigma(I(\lambda, \mu)z) \subseteq [\zeta, \xi]$, $\sigma(I(\mu_n, \lambda_n)z) \subseteq [\xi_n, \zeta_n]$ and hence by 1.2 $E([\zeta, \xi])I(\lambda, \mu)z = I(\lambda, \mu)z$ and

$$\begin{aligned} E([\zeta, \xi])I(\mu_n, \lambda_n)z &= E([\zeta, \xi])E([\xi_n, \zeta_n])I(\mu_n, \lambda_n)z \\ &= E(\phi)I(\mu_n, \lambda_n)z = 0 . \end{aligned}$$

Thus (4.5.4) takes the form

$$E([\zeta, \xi[)(S - \zeta)^2(S - \xi)^2 z = I(\lambda, \mu)z ,$$

and we may conclude $E([\zeta, \xi])(S - \zeta)^2(S - \xi)^2 z \in [x]$ from (4.5.2).

Generalizing the Hilbert space terminology, a two sided sequence of vectors $\{x_n\}_{n=-\infty}^{\infty}$ is called stationary if and only if the norm of any finite linear combination $\sum_{j=1}^{k}\alpha_j x_{j+h}$ is independent of h. If U is a unitary operator in $\mathfrak{X}$ and $x \in \mathfrak{X}$, then the sequence $\{U^n x\}_{n=-\infty}^{\infty}$ is stationary. Conversely, if $\{x_n\}_{n=-\infty}^{\infty}$ is stationary and $\mathfrak{Y}$ is the subspace spanned by this sequence, there exists a unique operator $U \in B(\mathfrak{Y})$ which satisfies $Ux_n = x_{n+1}$, an integer. U is unitary in $\mathfrak{Y}$ and is termed the shift operator of $\{x_n\}$. We call a stationary sequence spectral in case its shift operator is spectral.

The final statement of the following theorem replaces the problem of characterization of reflexive spaces every unitary operator of which is spectral by that of characterizing spectral stationary sequences. This "local" form of the problem seems more appropriate since the spectrality of every unitary operator in a space $\mathfrak{X}$ may depend not on "regular" properties of $\mathfrak{X}$ but on an irregularity which renders the class of unitary operators very sparse.

4.6. THEOREM. *Let U be a unitary operator in $\mathfrak{X}$. Then conditions* 3.1.1, 3.1.2 *and* 3.1.3 *are necessary in order that U be spectral. If $\mathfrak{X}$ is reflexive, then each of the conditions* 3.1.1 *to* 3.1.5 *is necessary and sufficient; and it is sufficient to let f in these conditions range over polynomials. For a reflexive $\mathfrak{X}$, U is spectral if and only if every stationary sequence it generates is spectral.*

Proof. The first statement follows from 4.2. and 3.1. It follows from 4.5 and from the fact that $\sigma(U)$, being a subset of the unit circle, s an R-set that if $\mathfrak{X}$ is reflexive, all the parts of Theorem 3.1 are applicable. Let $g \in R(\sigma(U))$. Using Cauchy's integral formula, it may be proved that there exists an admissible domain τ (in the sense of [4, Def. 2.2]) which contains $\sigma(U)$, such that g is uniformly approximable on τ by functions of the form

$$h(\zeta) = \sum_{j=1}^{k} \frac{1}{\zeta - \lambda_j}, \quad \lambda_j \in \rho(U)$$

(τ may depend on g, but not on the approximants. Cf. [1, p. 398]). Since $\sigma(U)$ is contained in the unit circle, we may assume, diminishing τ if necessary, that the complement of $\bar{\tau}$ is either connected or consists of two components at most, one of which contains the point $\zeta = 0$. In either case it follows from [13, p. 47, Th. 15] that the functions h, and hence g, are uniformly approximable on τ by polynomials f in ζ and ζ^{-1}. Thus these polynomials form a dense subalgebra of $R(\sigma(U))$, and by the continuity of the functional calculus, the corresponding $f(U)$'s are dense in $\{g(U) \mid g \in R(\sigma(U))\}$ in the uniform operator topology. From the proof of 3.1 it is seen that we may replace $R(\sigma(U)$ and $R(\sigma(U \mid [x]))$

by any subalgebra of $R(\sigma(U))$ with these properties. Since U is unitary, the conditions of 3.1 remain invariant if the involved functions are multiplied by ζ^k, k an integer. Therefore polynomials in ζ will do. Finally it follows from what has been shown above that the subspace spanned by a stationary sequence $\{U^n x\}_{n=-\infty}^{\infty}$ is $[x]$. Thus the final statement follows from the fact that 3.1.4 is the same as 3.1.1 for the shift operator.

5. Examples of non spectral unitary operators. Let Ω be a compact Hausdorff space. The unitary operators in $C(\Omega)$ are the operators of the form $(Ux)(\omega) = \mu(\omega)x(h(\omega))$, $\omega \in \Omega$, where h is a homeomorphism of Ω on itself, $\mu \in C(\Omega)$ and $|\mu(\omega)| \equiv 1$. This is proved in [12, pp. 469–472] for the real case, but the proof can be modified to apply to the complex case too by the use of an argument of Arens in a similar situation (v. [9, p. 88]). The following theorem treats only the case that h is non-periodic; for the case that h is the identity mapping Cf. Example 2.2 above.

5.1. THEOREM. *Let Ω be a compact Hausdorff space, and let U of the form $(Ux)(\omega) = \mu(\omega)x(h(\omega))$ (h, μ as above) be a unitary operator in $C(\Omega)$. If h is non-periodic, then U is not spectral.*

Proof. By 4.2, 3.1 and the fact that $\sigma(U)$ is contained in the unit circle (actually, coincides with it), it is sufficient to show that there exists no finite constants H such that

(5.1.1) $\|f(U))\| \leq H \max_{|\zeta|=1} |f(\zeta)|$, f a polynomial in ζ.

Let us calculate $\|f(U)\|$. If $f(\zeta) = \sum_{k=0}^{n} \alpha_k \zeta^k$, then

$$(f(U)x)(\omega) = \sum_{k=0}^{n} \alpha_k \mu(\omega)^k x(h^{0k}(\omega)) , \tag{5.1.2}$$

where h^{0k} denotes the kth iterate by substitution of $h(h^{00}(\omega) \equiv \omega)$. By hypothesis there exists an $\omega_0 \in \Omega$ such that the points $h^{0k}(\omega_0)$, $k = 0, 1, \cdots, n$ are distinct. Since Ω is Hausdorff, there exist pairwise disjoint open sets π_k, $k = 0, 1, \cdots, n$ such that $h^{0k}(\omega_0) \in \pi_k$. Since a compact Hausdorff space is normal, it follows by Urysohn's lemma that there exist functions $y_k \in C(\Omega)$ such that $y_k(h^{0k}(\omega_0)) = 1$, $y_k(\omega) = 0$ for $\omega \in \pi'_k$ and $0 \leq y_k(\omega) \leq 1$ on Ω. We define $x_0 \in C(\Omega)$ by

$$x_0(\omega) = \sum_{k=0}^{n} \overline{\operatorname{sgn}(\alpha_k \mu(\omega_0)^k)} y_k(\omega) .$$

Substitution in (5.1.2) gives

$$(f(U)x_0)(\omega_0) = \sum_{k=0}^{n} |\alpha_k \mu(\omega_0)^k| = \sum_{k=0}^{n} |\alpha_k| .$$

Since $\|x_0\| = 1$, $\|f(U)\| \geq \sum_{k=0}^{n} |\alpha_k|$ (actually $\|f(U)\| = \sum_{k=0}^{n} |\alpha_k|$). The necessary condition (5.1.1) now takes the form: there exists an $H < \infty$ such that for every polynomial $f(\zeta) = \sum_{k=0}^{n} \alpha_k \zeta^k$,

$$\sum_{k=0}^{n} |\alpha_k| \leq H \max_{|\zeta|=1} |f(\zeta)| .$$

To contradict this statement we use the following example of Hardy [8, § 14]. The series

$$\sum_{k=2}^{\infty} \frac{(-1)^k \binom{-i}{k}}{\log k} \zeta^k$$

converges uniformly for $|\zeta| = 1$, while the sum of the absolute values of its coefficients diverges. Therefore the polynomials which form its partial sums furnish us with the required counter example.

5.2. THEOREM. *In each of the sequence spaces l_p, $1 \leq p \leq \infty$, $p \neq 2$, there exists a non spectral unitary operator.*

Proof. If U is a unitary spectral operator in $\mathfrak{X}$, then necessarily [5, Assumption 1.14]:

(5.2.1) $M(U) = \sup \{\|x\| \mid x, y \in \mathfrak{X}, \|x + y\| = 1, \sigma(x) \cap \sigma(y) = \phi\} < \infty$.

This follows from the boundness of E by 1.2. Even if U is not spectral, the conclusion of 1.1 holds because $\sigma(U)$ is nowhere dense; and thus $\sigma(x)$ and $M(U)$ are definable. We show that in each of the considered spaces there exists a unitary operator U with $M(U) = \infty$.

Let p, $1 \leq p \leq \infty$, $p \neq 2$, be given. We denote by $\mathfrak{X}_j$ a space of the type $l_{p,n}$ or l_p (the last possibility is needed only for the remarks made after the theorem). If $\{\mathfrak{X}_j\}_{j=1}^{\infty}$ is a sequence of such spaces, we denote by $\sum_{j=1}^{\infty} \oplus \mathfrak{X}_j$ the Banach space of all sequences $\{x_j\}$ with $x_j \in \mathfrak{X}_j$ and

$$\|\{x_j\}\| = \left(\sum_{j=1}^{\infty} \|x_j\|^p\right)^{1/p} < \infty \quad (\text{if } p = \infty,\ \|\{x_j\}\| = \sup \|x_j\| < \infty).$$

If for each j, $T_j \in B(\mathfrak{X}_j)$, we denote by $\sum_{j=1}^{\infty} \oplus T_j$ the transformation T defined on (part of) $\sum_{j=1}^{\infty} \oplus \mathfrak{X}_j$ by $T\{x_j\} = \{T_j x_j\}$

5.3. LEMMA. *If $\mathfrak{X} = \sum_{j=1}^{\infty} \oplus \mathfrak{X}_j$ and for each j U_j is a unitary operator in $\mathfrak{X}_j$, then $U = \sum_{j=1}^{\infty} \oplus U_j$ is a unitary operator in $\mathfrak{X}$ and $M(U) \geq \sup_j M(U_j)$.*

Proof. That U is unitary is obvious. If $x_j \in \mathfrak{X}_j$ for a definite j, we denote by x_j^* the vector $\{y_k\} \in \mathfrak{X}$ defined by $y_j = x_j$, $y_k = 0$ for $k \neq j$.

Since the operation * is linear and norm preserving, $(\zeta - U_j)x_j(\zeta) = x_j$ for $\zeta \in \rho U_j(x_j)$ implies $(\zeta - U)x_j(\zeta)^* = x_j^*$ where $x_j(\zeta)^*$ is analytic on $\rho_{U_j}(x_j)$. Therefore $\sigma_U(x_j^*) \subseteq \sigma_{U_j}(x_j)$. It is obvious how to complete the proof.

Since l_p is linearly isometric to $\sum_{j=1}^{\infty} \oplus l_{p,n_j}$ where n_j are arbitrary natural numbers, 5.3 shows that we have only to find indices n_j and unitary operators U_j in l_{p,n_j} such that $\sup_j M(U_j) = \infty$. Let φ_j, $j = 1, \cdots, n$, be the natural basis of $l_{p,n}$. Henceforth U_n will denote the unitary operator in $l_{p,n}$ determined by the requirements $U_n\varphi_j = \varphi_{j+1(\text{mod}\, n)}$. The following lemmas will show that $\sup_n M(U_n) = \infty$, which will finish the proof.

We now use tensorial products as in [10]. If $x = (x_1, \cdots, x_n) \in l_{p,n}$, $y = (y_1, \cdots, y_m) \in l_{p,m}$, we define $x \otimes y$ to be the vector $(x_1y_1, x_1y_2, \cdots, x_1y_m, x_2y_1, x_2y_2, \cdots, x_2y_m, \cdots, x_ny_1, x_ny_2, \cdots, x_ny_m)$ of $l_{p,nm}$. This is a Kronecker product [7, p. 208], and the norm is a cross norm with respect to it, that is $\|x \otimes y\| = \|x\| \|y\|$. The tensorial product of linear operators, T in $l_{p,n}$ and S in $l_{p,m}$, is uniquely defined by the requirements $(T \otimes S)(x \otimes y) = Tx \otimes Sy$.

5.4. LEMMA. *If T, S are linear operators in $l_{p,n}$, $l_{p,m}$ respectively, then $\sigma_{T\otimes S}(x \otimes y) = \{\eta\theta \mid \eta \in \sigma_T(x),\ \theta \in \sigma_S(y)\}$.*

Proof. If T is an operator in a finite dimensional space and f is the minimum polynomial of x with respect to T, then $\sigma_T(x)$ is the set of zeros of f (cf. [5, p. 589]). We may assume that neither $\sigma_T(x)$ nor $\sigma_S(y)$ is empty since this case is trivial. In case $\sigma_T(x) = \{\eta\}$, $\sigma_S(y) = \{\theta\}$ the minimum polynomials are of the respective forms $(\zeta - \eta)^t$, $(\zeta - \theta)^s$ $(t, s \geq 1)$. By induction on t and s and use of the identity

$$(T \otimes S - \eta\theta)(x \otimes y) = (T - \eta) \otimes Sy + \eta x \otimes (S - \theta)y ,$$

one shows that the minimum polynomial of $x \otimes y$ with respect to $T \otimes S$, is of the form $(\zeta - \eta\theta)^r$, $r \geq 1$, and therefore $\sigma_{T\otimes S}(x \otimes y) = \{\eta\theta\}$ (actually we need only the case $t = s = 1$). In the general case $\sigma_T(x) = \{\eta_1, \cdots, \eta_a\}$, $\sigma_S(y) = \{\theta_1, \cdots, \theta_b\}$ we have by the finite dimensional case of the spectral theorem ([4, § 1] or [7, p. 132]) the resolutions $x = \sum_{i=1}^{a} x_i$, $y = \sum_{j=1}^{b} y_j$, where $\sigma_T(x_i) = \{\eta_i\}$, $\sigma_S(y_j) = \{\theta_j\}$. Let $\{\eta\theta \mid \eta \in \sigma_T(x), \theta \in \sigma_S(y)\} = \{\kappa_1, \cdots, \kappa_c\}$ and let z_k be the sum of the vectors $x_i \otimes y_j$ such that $\eta_i\theta_j = \kappa_k$. Since the x_i's are linearly independent and the y_j's are different from zero (by our assumption $x \neq 0$, $y \neq 0$), it follows that $z_k \neq 0$. Therefore, by the case of one point spectra, $\sigma_{T\otimes S}(z_k) = \{\kappa_k\}$. Since $x \otimes y = \sum_{k=1}^{c} z_k$ and since the minimum polynomial of a sum of vectors with minimum polynomials relatively prime in pairs is their product [7, p. 68], the statement of the lemma follows.

5.5. LEMMA. *If* $(m, n) = 1$, *then* $M(U_{nm}) \geq M(U_n)M(U_m)$.

Proof. $U_n \otimes U_m$ is determined by requirements of the form $(U_n \otimes U_m)\varphi_j = \varphi_{j\pi}$, where φ_j, $1 \leq j \leq nm$, is the natural basis of $l_{p,nm}$ and π is a permutation of the indices. Since $(m, n) = 1$, π is cyclic, and it is easily verified that there exists a unitary operator V in $l_{p,nm}$ such that $U_{nm} = V(U_n \otimes U_m)V^{-1}$, which implies that $M(U_{nm}) = M(U_n \otimes U_m)$. Since $l_{p,n}$ is of finite dimension, there exist vectors $x^{(1)}$, $y^{(1)}$ satisfying: $\sigma_{U_n}(x^{(1)}) \cap \sigma_{U_n}(y^{(1)}) = \phi$, $\| x^{(1)} + y^{(1)} \| = 1$ and $\| x^{(1)} \| = M(U_n)$. Let $x^{(2)}$, $y^{(2)}$ play a similar role with respect to U_m. Consider the vectors $x = x^{(1)} \otimes x^{(2)}$, $y = x^{(1)} \otimes y^{(2)} + y^{(1)} \otimes x^{(2)} + y^{(1)} \otimes y^{(2)}$. Since $\sigma(U_n)$ is the set of roots of unity of order n, $\sigma_{U_n}(x^{(1)})$ and $\sigma_{U_n}(y^{(1)})$ are sets of roots of unity of order n. Similarly for $\sigma_{U_m}(x^{(2)})$ and $\sigma_{U_m}(y^{(2)})$. Since $(m, n) = 1$, the representation of a root of unity order mn as a product of a root of unity of order n by one of order m is unique. Therefore it follows from 5.4 that $\sigma_{U_n \otimes U_m}(x) \cap \sigma_{U_n \otimes U_m}(y) = \phi$. By the cross property of the norm $\| x + y \| = \| (x^{(1)} + y^{(1)}) \otimes (x^{(2)} + y^{(2)}) \| = 1$ and $\| x \| = \| x^{(1)} \otimes x^{(2)} \| = M(U_n)M(U_m)$. Thus $M(U_{nm}) = M(U_n \otimes U_m) \geq M(U_n)M(U_m)$.

5.6. LEMMA. *For every given* p, $1 \leq p \leq \infty$, $p \neq 2$, *there exist an* $\eta > 1$ *and positive integers* k *and* m_0 *such that* $M(U_{km+1}) > \eta$ *for* $m \geq m_0$.

Proof. By calculating the eigenvectors of U_n, one shows that the vectors $x = (x_1, \cdots, x_n)$ with $\sigma_{U_n}(x)$ disjoint from $\sigma_{U_n}(y)$, where $y = (1, 1, \cdots, 1)$, are those which satisfy $\sum x_j = 0$. Thus

$$M(U_n) \geq \sup \left\{ \frac{\| x \|}{\| x + \alpha y \|} \,\middle|\, \sum x_j = 0,\ \alpha \text{ arbitrary} \right\} .$$

For $2 < p < \infty$ we chose $x = (1, \cdots, 1, -m/(n - m), \cdots, -m/(n - m)$, where 1 is repeated m times, and

$$\alpha = \frac{\left(\dfrac{m}{n - m}\right)^{(p-2)/(p-1)} - 1}{1 + \left(\dfrac{n - m}{m}\right)^{1/(p-1)}} .$$

Then if $n = km + 1$, $k \geq 2$ and

$$m \to \infty, \frac{\| x \|^p}{\| x + \alpha y \|^p}$$

tends to

$$\frac{(1 + t^{1/(p-1)})^p(1 + t^{1-p})}{(t^{1/(p-1)} + t^{-(p-2)/(p-1)})^p + t\left(\dfrac{t + 1}{t}\right)^p}$$

where $t = k - 1$. Although the last expression tend to 1 as $t \to \infty$, it is not difficult to verify that it is greater than 1 for all sufficiently large values of t; hence a suitable integer $k = t + 1$ can be found. The case $1 < p < 2$ follows by duality: If $1/p + 1/q = 1$ then $M_q(U_n)$, where the subscript indicates that U_n is to be considered as an operator in $l_{q,n}$, is the maximum of the norms of the values of the resolution of the identity E of U_n. The resolution of the identity of $U_n^* = U_n^{-1}$ is E. Therefore $M_p(U_n^{-1}) = M_q(U_n)$. But U_n is unitarily equivalent in $l_{p,n}$ to U_n^{-1}. Therefore $M_p(U_n) = M_q(U_n)$; and since $2 < q < \infty$ the lemma is true in this case too. If $p = 1$, we may take $x = (1, \cdots, 1, -n + 1)$, $\alpha = -1$; while if $p = \infty$, we take the same x and $\alpha = n/2$.

Finally to see that 5.5 and 5.6 imply $\sup_n M(U_n) = \infty$, we have only to use the fact that each sequence $a_m = km + 1$, $m = 1, 2, \cdots$, contains an infinite subsequence of pairwise prime integers. As pointed out by Dr. Dov Jarden such a subsequence is obtained by defining inductively $m_1 = 1$, $m_{j+1} = a_{m_1} a_{m_2} \cdots a_{m_j}$.

REMARKS. For $p = 1, \infty$ the proof of 5.2 yields unitary operators which are not even prespectral. It applies also to subspaces which contain all finite sequences. It also follows from 5.2 that if Ω is a measure space which is not a finite union of atoms, then there exist non spectral unitary operators in the space $L_p(\Omega)$, $1 \leq p < \infty$, $p \neq 2$. An operator U in l_p, $1 \leq p < \infty$, $p \neq 2$, is unitary only if determined by $U\varphi_j = \lambda_j \varphi_{j\pi}$, $j = 1, 2, \cdots$, where $\{\varphi_j\}$ is the natural basis, π a permutation and $|\lambda_j| = 1$ ([2, p. 178]. The proof goes easily over to the complex case). We decompose π into disjoint cycles (including the possibility of infinite "cycles") and consider the unitary operators induced by U in the subspaces spanned by the φ_j's with j belonging to a definite cycle. One shows that $M(U) = \sup M(V)$, where V runs over the induced operators. Moreover, if we change the λ_j's into 1 and the cycle of V into a standard one, we obtain an operator W with $M(W) = M(V)$. Hence Condition (5.2.1) depends only on the length of the cycles determined by π. From Theorem 5.7 it will follow that if in particular at least one of these cycles is infinite, (5.2.1) does not hold. On the other hand, it follows from [5, Th. 3.11, Th. 3.12 (III)] that this condition is sufficient for spectrality of U if $1 < p < \infty$.

5.7. THEOREM. *Let* $\bar{l}_p$, $1 \leq p \leq \infty$, $p \neq 2$, *be the space of two-sided sequences* $\{\alpha_j\}_{j=-\infty}^{\infty}$ *with the obvious norm. Let* φ_j, $-\infty < j < \infty$, *be the natural basis of* $\bar{l}_p$ *and* U *the unitary operator defined by* $U\varphi_j = \varphi_{j+1}$. *Then* U *is not spectral.*

Proof. To facilitate the writing we assume $p < \infty$. From the proof of 4.6 through [6, Th. 18 (IV)] (cf. 3.1), it follows that if $H(U_n)$ is the infimum of possible constants in Condition 3.1.1 for polynomials, then $M(U_n) \leq H(U_n)$. Let K be a positive number. Then by the proof of 5.2 there exists an n such that $M(U_n) > 2K$, and therefore there exists a polynomial g such that $\| g(U_n \| > 2K \max_{\zeta^n=1} | g(\zeta) |$. If f is a polynomial, $f(U_n)$ depends only on the values f assumes at the nth roots of unity, and in a continuous manner. Therefore, by the approximation theorem of Weierstrass, there exists a polynomial $f(\zeta) = \sum_{k=0}^{s} \beta_k \zeta^k$ such that $\| f(U_n) \| > 2K \max_{\zeta^n=1} | f(\zeta) |$ and $2 \max_{\zeta^n=1} | f(\zeta) | > \max_{|\zeta|=1} | f(\zeta) |$; hence $\| f(U_n) \| > K \max_{|\zeta|=1} | f(\zeta) |$. Identifying $l_{p,n}$ with the subspace of $\bar{l}_p$ spanned by $\varphi_1, \cdots, \varphi_n$, we see that there exists an $x = \sum_{j=1}^{n} \alpha_j \varphi_j$ such that

$$\frac{\| f(U_n)x \|}{\| x \|} > K \max_{|x|=1} | f(\zeta) | . \tag{5.7.1}$$

It will simplify the notation if we assume, as we may, that the formal degree s of f is of the form $s = rn$, $r > 1$. Let t be a positive integer and consider the vector $x' = \sum_{m=1}^{r+t} \sum_{j=1}^{n} \alpha_j \varphi_{(m-1)n+j}$. Then

$$\begin{aligned} f(U)x' &= \sum_{k=0}^{s} \sum_{m=1}^{r+t} \sum_{j=1}^{n} \beta_k \alpha_j \varphi_{(m+1)n+j+k} \\ &= \sum_{u=1}^{2r+t} \sum_{v=1}^{n} (\sum \beta_k \alpha_j) \varphi_{(u-1)n+v} \end{aligned} \tag{5.7.2}$$

where the inner summation in the r.h.s. extends over the pairs j, k satisfying $(m-1)n + j + k = (u-1)n + v$, where $1 \leq j \leq n$, $0 \leq k \leq s = rn$ and $1 \leq m \leq r + t$. On the other hand

$$f(U_n)x = \sum_{k=0}^{s} \sum_{j=1}^{n} \beta_k \alpha_j \varphi_{j+k(\mathrm{mod}\, n)} = \sum_{v=1}^{n} (\sum \beta_k \alpha_j) \varphi_v , \tag{5.7.3}$$

where here the inner summation is over the pairs j, k satisfying $j + k = v(\mathrm{mod}\, n)$ with the same inequalities. For $r + 1 \leq u \leq r + t$, the coefficient of $\varphi_{(u-1)n+v}$ in (5.7.2) equals the coefficient of φ_v in (5.7.3). Therefore

$$\| f(U)x' \| \geq t^{1/p} \| f(U_n)x \| , \tag{5.7.4}$$

and on the other hand

$$\| x' \| = (r + t)^{1/p} \| x \| . \tag{5.7.5}$$

(5.7.1), (5.7.4) and (5.7.5) imply

$$\| f(U) \| > \left(\frac{t}{r+t}\right)^{1/p} K \max_{|\zeta|=1} | f(\zeta) | .$$

Letting t tend to infinity, we get $\| f(U) \| \geq K \max_{|\zeta|=1} | f(\zeta) |$. Since K is arbitrary and $\sigma(U)$ is contained in the unit circle, this shows that U does not satisfy Condition 3.1.1; hence it is not spectral.

REFERENCES

1. W. G. Bade, *Weak and strong limits of spectral operators*, Pacific J. Math. **4** (1954), 393-413.

2. S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932.

3. E. Bishop, *Spectral theory for operators on a Banach space*, Trans. Amer. Math. Soc. **86** (1957), 414-445.

4. N. Dunford, *Spectral Theory I. Convergence to projections*, Trans. Amer. Math. Soc. **54** (1943), 185-217.

5. ———, *Spectral Theory II. Resolutions of the identity*, Pacific J. Math. **2** (1952), 559-614.

6. ———, *Spectral operators*, Pacific J. Math. **4** (1954), 321-354.

7. N. Jacobson, *Lectures in abstract algebra.* Vol. II—Linear algebra, New York, 1953.

8. E. Landau, *Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie*, 2te Auf., Berlin, 1929.

9. L. H. Loomis, *Abstract harmonic analysis*, New York, 1953.

10. F. J. Murray, *On complementary manifolds and projections in spaces L_p and l_p*, Trans. Amer. Math. Soc. **41** (1937), 138-152.

11. J. Schwarz, *Perturbations of spectral operators, and applications*, Pacific J. Math. **4** (1954), 415-458.

12. M. H. Stone, *Applications of the theory of Boolean rings to general topology*, Trans. Amer. Math. Soc. **41** (1937), 375-481.

13. J. L. Walsh, *Interpolation and appoximation by rational functions in the complex domain*, New York, 1935.

THE HEBREW UNIVERSITY

UNIFORMIZABLE SPACES WITH A UNIQUE STRUCTURE

I. S. GÁL

Here we shall study only uniformizable Hausdorff spaces. In general if a topological space X is uniformizable then there are many uniform structures $\mathscr{U}$ compatible with the topology of X. If X is compact then there is only one uniform structure for X and there are also non-compact spaces whose structures are uniquely determined by their topology. (See [1] and [2].) The purpose of this note is to give a necessary and sufficient condition that $\mathscr{U}$ be uniquely determined by X. Let $C(X)$ be the algebra of bounded real valued continuous functions on X and let $C(X)$ be topologized by the topology of uniform convergence on the whole space X. By $A(X)$ we denote the subalgebra of those real valued continuous functions which are constant on the complement of some compact set in X. We shall prove the following

THEOREM. *The uniformizable Hausdorff space X admits only one uniform structure if and only if $A(X)$ is dense in $C(X)$.*

Another necessary and sufficient condition for uniqueness was found earlier by R. Doss [3]: The closed sets C_1 and C_2 are called *normally separable* if there exists a continuous real valued function f on X which takes the value 1 on C_1 and the value 2 on C_2. Doss proved the following:

Uniqueness takes place if and only if of any two normally separable sets at least one is compact.

The following proof of the Theorem makes no use of this criterion given by Doss. However at the end it will be proved that the criterion stated in the Theorem and the criterion due to Doss are equivalent. This gives a new, simpler proof of Doss's theorem. Approximately at the same time when [3] was published P. Samuel in [5] and T. Shirota in [6] proved that

Among the uniform structures compatible with the topology of X there is a weakest if and only if X is locally compact.

The two halfs of this theorem are stated as of Lemma 3 and Lemma 6 below. Their proofs are independent of the rest of the paper and so they furnish a simple proof for the Samuel-Shirota theorem.

A space X is said to be *normally imbedded* in the space Y if every real valued continuous function on X admits a continuous extension to Y. If this property is supposed to hold only for bounded functions one speaks about a *bounded normal imbedding*. E. Hewitt in [4] proved that

Received December 24, 1958.

The Hausdorff space X is normally imbedded in every uniformizable space containing X as a dense subspace if and only if of any two disjoint sets at least one is compact.

Among all uniform structures compatible with the topology of a uniformizable space X there is a strongest called the Weil structure or the universal structure of X. Its existence follows from the fact that the union of all uniform structures compatible with X is a subbase for a uniform structure which is compatible with X. The Weil structure $\mathscr{U}_W$ is uniquely determined by the following property: If $\mathscr{V}$ is a uniform structure for Y and $f: X \to Y$ is continuous with respect to the topology of X and the uniform topology associated with $\mathscr{V}$ then f is uniformly continuous with respect to $\mathscr{U}_W$ and $\mathscr{V}$. In general $\mathscr{U}_W$ is not a precompact structure.

Let X satisfy the criterion given by Doss and let $\mathscr{U}$ be the unique structure compatible with its topology. The uniqueness implies that $\mathscr{U}$ is identical with the Weil structure of X. Let X be a dense subspace of the uniformizable space Y and let $\mathscr{V}$ be a uniform structure for Y. The restriction of $\mathscr{V}$ to X is the Weil structure of X and so every real valued continuous function f on X is uniformly continuous with respect to $\mathscr{V}$. Consequently f can be extended to a uniformly continuous function on Y and so X is normally imbedded in Y. Thus by Hewitt's theorem one of any two disjoint closed sets of X must be compact. Combining the present Theorem with the theorms of Doss and Hewitt we obtain:

Any two of the following statements are equivalent:

(i) *X has a unique uniform structure.*

(ii) *If C_1 and C_2 are normally separable closed sets in X then at least one of them is compact.*

(iii) *If C_1 and C_2 are disjoint closed sets in X then at least one of them is compact.*

(iv) *$A(X)$ is dense in $C(X)$.*

(v) *X is normally imbedded in every uniformizable space containing X as a dense subspace.*

We omitted the analogue of (v) concerning bounded normal imbeddings. For we have:

If X has a unique uniform structure then every real valued continuous function is bounded on X.

This follows from Lemma 1 below. Using (iii) one can also prove that any two disjoint closed sets are normally separable.

The following notations will be used: Open sets will be denoted by O, closed sets by C, neighborhoods by N_x and open neighborhoods by O_x. For the closure of a set A we write $\bar{A}$ and cA stands for the complement of A with respect to a given set containing A. Uniform structures will be denoted by $\mathscr{U}$, $\mathscr{V}$, $\cdots$; the completion of a uniform space

X with respect to a structure $\mathscr{U}$ will be denoted by $\bar{X}$ and the complete structure will be denoted by $\bar{\mathscr{U}}$. As usual $U \circ V$ is the composition of the vicinities $U, V \in \mathscr{U}$ and $U[x] = [y : (x, y) \in U]$. If $\mathscr{U}_i$ $(i \in I)$ are uniform structures for X then lub $\mathscr{U}_i$ denotes the uniform structure generated by the subbase $\cup \mathscr{U}_i$. It is the weakest structure which is stronger than any $\mathscr{U}_i$ $(i \in I)$. If f is uniformly continuous on X then $\bar{f}$ denotes its extension to $\bar{X}$. The structure $\mathscr{U}_C$ used in the proof of Lemma 5 is the so-called Čech structure which was introduced by Samuel in [5]. The fact that the definition given in [5] is equivalent to the present simpler definition follows from Lemma 4. $\mathscr{U}_C$ is the strongest precompact structure compatible with the topology of X and its completion is the Stone-Čech compactification βX.

LEMMA 1. *If $A(X)$ is dense in $C(X)$ then every uniform structure $\mathscr{U}$ compatible with the topology of X is precompact.*

Proof. This follows by a simple argument which is used also in [3]: Suppose that X is a topological space and $\mathscr{U}$ is a non-precompact structure compatible with the topology of X. Then there is a symmetric vicinity $U \in \mathscr{U}$ and a sequence of points $x_1, x_2, \cdots$ in X such that $(x_m, x_n) \in U$ only if $m = n$. We choose a symmetric $V \in \mathscr{V}$ satisfying $V \circ V \subseteq U$ and a symmetric $W \in \mathscr{U}$ satisfying $W \circ W \subseteq V$. Since X is completely regular there is a real valued continuous function f_n on X with the property that $|f_n(x)| \leq 1$ for every $x \in X$, the closure of $W[x_n]$ is a support of f_n and $f_n(x) = \pm 1$ according as n is even or odd. By $W \circ W \subseteq V$ the closure of $W[x_n]$ is contained in $V[x_n]$ and by $V \circ V \subseteq U$ the sets $V[x_m]$ and $V[x_n]$ intersect only if $m = n$. Therefore the series $\Sigma f_n(x)$ contains for each $x \in X$ at most one non-vanishing term and it defines a bounded continuous function f on X. Neither $\{x_1, x_3, \cdots\}$ nor $\{x_2, x_4, \cdots\}$ is compact and so f can not be approximated uniformly on X by elements of $A(X)$. Hence the existence of non-precompact structure implies that $A(X)$ is not dense in $C(X)$.

LEMMA 2. *If $A(X)$ is dense in $C(X)$ then X is locally compact.*

Proof. Let O_x be an open neighborhood of the point $x \in X$ and let f be a real valued continuous function on X such that $0 \leq f(\xi) \leq 1$ for every $\xi \in X$, $f(x) = 1$ and $f(\xi) = 0$ if $\xi \notin O_x$. Since $A(X)$ in dense in $C(X)$ there is a continuous function g which is constant on the complement O of a compact set C and is such that $|f(\xi) - g(\xi)| \leq \varepsilon$ for every $\xi \in X$. If $N_x = [\xi\colon f(\xi) \geq 1 - \varepsilon]$ is a subset of C then N_x is a compact neighborhood of x. If this is not the case then O and N_x have a common point ξ. Then for every $\eta \in O$ we have

$$f(\eta) \geq g(\eta) - \varepsilon = g(\xi) - \varepsilon \geq f(\xi) - 2\varepsilon \geq 1 - 3\varepsilon > 0$$

and so $\eta \in O_x$. Since $O \subseteq O_x$ where $C = cO$ is compact we see that the complement of O_x is compact. If this is the situation for every open neighborhood O_x of x then X is compact. Hence either N_x is a compact neighborhood of x for every $x \in X$ or X is a compact space.

Let f map X into Y and let $\mathscr{V}$ be a uniform structure for Y. The sets $f^{-1}(V) = [(x_1, x_2) : (f(x_1), f(x_2)) \in V]$ $(V \in \mathscr{V})$ form a base for a uniform structure $\mathscr{U}$ for X, called the inverse image of $\mathscr{V}$ under f. If f is a real valued function on X and $\mathscr{V}$ is the usual structure of the reals the inverse structure will be denoted by $\mathscr{U}_f$. It is a pseudo-metric structure which is generated by the pseudo-metric $d_f(x_1, x_2) = |f(x_1) - f(x_2)|$. If f is bounded then $\mathscr{U}_f$ is precompact. If $\{f\}$ is a family of real valued functions on X we call lub $\mathscr{U}_f$ the *uniform structure generated by the family* $\{f\}$. Every f in $\{f\}$ is uniformly continuous with respect to lub $\mathscr{U}_f$. Moreover if $\mathscr{U}$ is a uniform structure for X and if every f in $\{f\}$ is uniformly continuous with respect to $\mathscr{U}$ then lub $\mathscr{U}_f \leq \mathscr{U}$. If every $f \in \{f\}$ is bounded then lub $\mathscr{U}_f$ is a precompact structure for X. These simple consequences are presented in greater detail in Chapter IX of [1].

Some interesting uniform structures are structures generated by families of real valued functions $\{f\}$. For example let X be locally compact and let $\{f\}$ be the family $A(X)$. Given $x \in X$ and a compact neighborhood C_x of x there is a real valued continuous function f on X such that $f(x) = 1$ and C_x is a support of f. Hence C_x is a neighborhood of x in the uniform topology associated with $\mathscr{U}_f$. It follows that $\mathscr{U} =$ lub $\mathscr{U}_f$ is compatible with the topology of X. Every $f \in A(X)$ is constant on the complement of a compact set and so it is uniformly continuous with respect to any uniform structure $\mathscr{V}$ which is compatible with X. Therefore $\mathscr{U} \leq \mathscr{V}$ and so $\mathscr{U}$ is the weakest structure compatible with the topology given on X. Hence we proved the following lemma, which incidentally is an exercise in [1]. (See Chap. IX. p. 16 Exercise 11.)

LEMMA 3. *If X is a locally compact Hausdorff space then there is a weakest uniform structure which is compatible with the topology of X. It is the uniform structure generated by the family $A(X)$.*

The weakest structure if it exists is necessarily precompact. Now we show that every precompact separated structure can be generated by families of real valued functions. For let $\mathscr{U}$ be a precompact separated structure for X and let $\bar{X}$ be the completion of X with respect to $\mathscr{U}$. The completed structure will be denoted by $\bar{\mathscr{U}}$. Let $\mathscr{U}_{\bar{f}}$ denote the uniform structure generated on $\bar{X}$ by the real valued function $\bar{f}$ given on $\bar{X}$. It is clear that the restriction of $\mathscr{U}_{\bar{f}}$ to X is the same as the structure $\mathscr{U}_f$ generated on X by the restriction f of $\bar{f}$ to X. More generally if $\{\bar{f}\}$ is a family of real valued functions on $\bar{X}$ then the restriction of lub $\mathscr{U}_{\bar{f}}$ to X is the structure lub $\mathscr{U}_f$. If $\{\bar{f}\}$ is the family of all real valued continuous functions on $\bar{X}$ then lub $\mathscr{U}_{\bar{f}}$ is compatible with the topology of $\bar{X}$ and so by the compactness of $\bar{X}$ we have $\bar{\mathscr{U}} =$ lub $\mathscr{U}_{\bar{f}}$. Therefore $\mathscr{U} =$ lub $\mathscr{U}_f$ where $\{f\}$ is the family of the restrictions of continuous functions $\bar{f}$ to X. Since f is the restriction of some $\bar{f}$ if and only if f is uniformly continuous with respect to $\mathscr{U}$ we have

LEMMA 4. *Every precompact separated structure $\mathscr{U}$ is generated by the family of those real valued functions which are uniformly continuous with respect to $\mathscr{U}$.*

The topology of uniform convergence on X is meaningful on the linear space L of all real valued functions on the set X: The ε-neighborhood of 0 consists of those functions f on X for which supr $|f(x)| < \varepsilon$. Let $A, C \subseteq L$ and let A be dense relative to C. By $\mathscr{U}_A$ and $\mathscr{U}_C$ we denote the uniform strutures generated by the families A and C, respectively. Then for every $c \in C$ and $\varepsilon > 0$ there is an $a \in A$ such that $| a(x) - c(x) | < \varepsilon/4$ for every $x \in X$ and so

$$[(x, y) : | c(x) - c(y) | < \varepsilon] \supseteq \left[(x, y): | a(x) - a(y) | < \frac{\varepsilon}{2}\right]$$

This implies that every vicinity of $\mathscr{U}_C$ contains a vicinity of $\mathscr{U}_C$ so that $\mathscr{U}_C \leq \mathscr{U}_A$. If in addition $A \subseteq C$ then $\mathscr{U}_A \leq \mathscr{U}_C$ and so we have

LEMMA 5. *If A is dense in C then they generate the same uniform structure.*

Now it is easy to show that if $A(X)$ is dense in $C(X)$ then there is only one uniform structure which is compatible with the topology of X: By Lemma 2 the space X is locally compact and so by Lemma 3 it has a weakest uniform structure $\mathscr{U}_w$ which is compatible with its topology. By the same lemma $\mathscr{U}_w$ is generated by $A(X)$. It will be sufficient to show that $\mathscr{U}_w$ is identical with the Weil structure $\mathscr{U}_W$ of X. By Lemma 1 $\mathscr{U}_W$ is precompact and so by Lemma 4 it is generated by the family of those real valued functions on X which are uniformly continuous with respect to $\mathscr{U}_W$. By the precompactness and by the definition of $\mathscr{U}_W$ this family is $C(X)$. Since $A(X)$ is dense in $C(X)$ by Lemma 5 they generate the same structure, that is $\mathscr{U}_w = \mathscr{U}_W$. This proves the sufficiency of the condition given in the Theorem.

Now we shall prove that the condition stated in the Theorem is also necessary. First we suppose that X is a locally compact Hausdorff space. Let $\bar{X} = X \cup \{\infty\}$ be the Alexandroff compactification of X and let $\mathscr{U}_A$ be the uniform structure obtained for X by restricting the unique structure of $\bar{X}$ to X. We prove that a real valued function f is uniformly continuous with respect to $\mathscr{U}_A$ if and only if f belongs to the uniform closure of $A(X)$. For compact X this is obvious so we may assume that X is not a compact space. Since the elements of $A(X)$ are uniformly continuous with respect to any structure compatible with the topology of X the same holds for the elements of its closure $\overline{A(X)}$ and so it will be sufficient to show that if f is uniformly continuous with respect to $\mathscr{U}_A$ then $f \in \overline{A(X)}$. However if f is uniformly continuous with respect to $\mathscr{U}_A$ then it has a continuous extension $\bar{f}$ to $\bar{X}$. By the continuity of $\bar{f}$ at ∞ for every $\varepsilon > 0$ there is a compact set $C \subset X$ such that $|\bar{f}(x) - \bar{f}(\infty)| < \varepsilon$ for every $x \notin C$. Let O be an open neighborhood of C which does not contain ∞. Since $\bar{X}$ is normal there is a real valued continuous function $\bar{g}$ on $\bar{X}$ which takes the value 1 on C, vanishes outside of O and satisfies $0 \leq g(x) \leq 1$ on $\bar{X}$. Then $h = (f - \bar{f}(\infty))g + \bar{f}(\infty)$ belongs to $A(X)$ and is such that $|h(x) - f(x)| < \varepsilon$ for every $x \in X$.

Let us now suppose that $A(X)$ is not dense in $C(X)$. Then there is an $f \in C(X)$ which is not in $\overline{A(X)}$ and so it is not uniformly continuous with respect to $\mathscr{U}_A$. Since every element of $C(X)$ is uniformly continuous with respect to the uniform structure $\mathscr{U}_C$ generated by $C(X)$ we see that $\mathscr{U}_A$ and $\mathscr{U}_C$ are distinct structures compatible with the topology of X. This proves the necessity of the condition in the case of locally compact spaces.

The proof of the Theorem will be completed by showing

LEMMA 6. *If the uniformizable Hausdorff space X is not locally compact then there is no weakest among the uniform structures which are compatible with the topology of X.*

Proof. First we notice that if X is a Hausdorff space and if the Hausdorff space $\bar{X}$ is a compactification of X which contains only finitely many more elements than X then X is locally compact. Now let $\mathscr{U}$ be a uniform structure which is compatible with the topology of the uniformizable Hausdorff space X. We assume that X is not locally compact. Let B denote the family of those bounded real valued functions on X which are uniformly continuous with respect to $\mathscr{U}$ and let $\mathscr{V}$ be the uniform structure generated on X by B. Then $\mathscr{V}$ is precompact and is not stronger than $\mathscr{U}$. Let $\bar{X}$ be the compact completion of X with respect to $\mathscr{V}$. By the foregoing remark $\bar{X} - X$ is an infinite set. We consider the space $\bar{Y}$ obtained from $\bar{X}$ by identifying a finite number of distinct points $x_1, \cdots, x_n$ $(n > 1)$ of $\bar{X} - X$. The identification space $\bar{Y}$ will be compact and separated, so it has a unique uniform structure whose restriction to X will be denoted by $\mathscr{W}$. Then $\bar{Y}$ is the completion of X with respect to $\mathscr{W}$ and $\bar{X}$ is the completion of X with respect to $\mathscr{V}$. By Lemma 4 both $\mathscr{V}$ and $\mathscr{W}$ are generated by their families of real valued uniformly continuous functions. A real valued function is uniformly continuous with respect to $\mathscr{W}$ if and only if it is uniformly continuous with respect to $\mathscr{V}$ and its extension to $\bar{X}$ assumes the same value at $x_1, \cdots, x_n$. Hence $\bar{X}$ being separated there are real valued functions on X which are uniformly continuous with respect to $\mathscr{V}$ but not with respect to $\mathscr{W}$. Therefore $\mathscr{W} < \mathscr{V} \leq \mathscr{U}$ and so X has no weakest structure compatible with its topology. Lemma 6 and the Theorem are proved.

We finish by proving that the condition given in the Theorem is equivalent to the condition of Doss. First suppose that $A(X)$ is dense in $C(X)$. Let C_0 and C_1 be normally separated by f. We may assume that $0 \leq f(x) \leq 1$ for every $x \in X$, f is 0 on C_0 and 1 on C_1. We choose a $g \in A(X)$ satisfying $|f(x) - g(x)| < \varepsilon < \frac{1}{2}$ everywhere on X. Let g be constant on the complement of the compact set C. If this constant value is neither 0 nor 1 then both C_0 and C_1 are compact. Otherwise we may restrict ourselves to the case when C is a compact support of g. If $x \notin C$ then $g(x) = 0$ so $f(x) < \varepsilon$ and $x \in C_{0\varepsilon} = [x : f(x) \leq \varepsilon]$. Therefore $cC \subseteq C_{0\varepsilon} \subseteq cC_{1\varepsilon} = [x : f(x) < 1 - \varepsilon]$. This shows that $C_1 = [x : f(x) = 1] \subseteq C$ and so C_1 is compact.

Next we suppose that X satisfies Doss's condition. Let $f \in C(X)$ and $\varepsilon > 0$ be given. We consider the closed sets $C_k = [x : |f(x) - k\varepsilon| \leq \frac{1}{2}]$ where $k = 0, \pm 1, \pm 2, \cdots$. Their union is X. Any two of the sets C_{2k} $(k = 0, \pm 1, \pm 2, \cdots)$ are normally separable so at most one of them is not compact. Similarly at most one of the sets C_{2k+1} $(k = 0, \pm 1, \pm 2, \cdots)$ can be non-compact. Moreover if C_{2k} and C_{2l+1} are not compact they must have common points and so

$$C_{2k} \cup C_{2l+1} = \left[x : \varepsilon\left(m - \frac{1}{2}\right) \leq f(x) \leq \varepsilon\left(m + \frac{3}{2}\right)\right]$$

for some m. We define

$$g(x) = \begin{cases} f(x) & \textit{if } x \in C_k \textit{ and } k < m \\ \varepsilon(m - \frac{1}{2}) & \textit{if } x \in C_{2k} \cup C_{2l+1} \\ f(x) - \varepsilon & \textit{if } x \in C_k \textit{ and } k > m + 1 \,. \end{cases}$$

Then $|f(x) - g(x)| \leq \varepsilon$ for every $x \in X$ and $g \in A(X)$ because f being bounded there are only finitely many non-void sets among the sets C_k. If only C_{2k} or only C_{2l+1} is non-compact the construction is similar.

References

1. N. Bourbaki, *Topologie générale*, Chap. I, II et IX; 2nd edition Hermann and Cie, Editeurs Paris (1951) and (1958).
2. J. Dieudonné, *Sur les espaces uniformes complets*, Ann. Ec. Normal Sup. 3e série **56** (1939), 277-291.
3. R. Doss, *On uniform spaces with a unique structure*, Amer. J. Math. **71** (1949), 19-23.
4. E. Hewitt, *A note on extensions of continuous real functions*, Assais Acad. Brasil Ci. **21** (1949), 175-179.
5. P. Samuel, *Ultrafilters and compactification of uniform spaces*, Trans. Amer. Math. Soc. **64** (1948), 100-132.
6. T. Shirota, *On systems of structures of a completely regular space*, Osaka Math. Journal **2** (1950), 131-143.

YALE UNIVERSITY AND CORNELL UNIVERSITY

HIGHER DIMENSIONAL CYCLIC ELEMENTS

JOHN GARY

Introduction. Whyburn, in 1934, introduced the higher dimensional cyclic elements [5]. He gave an analysis of the structure of the homology groups of a space in terms of its cyclic elements. His results were for finite dimensional spaces, and he used the integers modulo two as the coefficient group. Puckett generalized some of Whyburn's results to compact metric spaces [3]. Simon has shown that if E is a closed subset of a compact space M, which contains all the $(r-1)$-dimensional cyclic elements of M, then $H^r(E) \approx H^r(M)$[4]. He also obtained a direct sum decomposition of $H^r(M)$ using the cyclic elements of M. We will extend some of these results.

The properties of zero-dimensional cyclic elements in locally connected spaces, and the relation of these cyclic elements to monotone mappings, is basic in the applications of zero-dimensional cyclic element theory. We shall give some counter-examples concerning the generalization of these properties to higher dimensional cyclic elements.

1. Preliminaries. Throughout this paper M will always denote a compact Hausdorff space. We shall use the augmented Cech homology and cohomology with a field as coefficient group. Results stated in terms of cohomology may be given a dual expression in terms of homology by means of the dot product duality for the Cech theory.

DEFINITION 1.1. A T_r set in M is a closed subset T of M such that $H^r(K) = 0$, for all closed subsets K of T.

DEFINITION 1.2. An E_r set in M is a non-degenerate subset of M which is maximal with respect to the property that it can not be disconnected by a T_r set of M.

The proofs of Lemmas 1.3 through 1.9 can be found in the papers by Whyburn [5] and Simon [4]. The proofs given by Whyburn are for subsets of Euclidean space, but they can be carried over to our case without difficulty.

LEMMA 1.3. *Let K be a subset of M which can not be disconnected by a T_r set. If $M = M_1 \cup M_2$, T_r-separated (by this we mean M_1 and M_2 are proper closed subsets and $M_1 \cap M_2$ is a T_r set), then $K \subset M_1$ (or, $K \subset M_2$).*

Received December 24, 1958. The results of this paper are contained in the authors doctoral dissertation, University of Michigan, 1956. The author wishes to thank Professor R. L. Wilder for his advice and encouragement.

LEMMA 1.4. *If K is an E_r set, then K is closed and connected.*

LEMMA 1.5. *If K_1 and K_2 are both E_r sets and $K_1 \neq K_2$, then $K_1 \cap K_2$ is a T_r set. Any T_r set is also a T_{r+1} set.*

LEMMA 1.6. *If K is a non-degenerate subset of M, which can not be disconnected by a T_r set, then K is contained in a unique E_r set in M.*

DEFINITION 1.7. If $\gamma^r \in H^r(M)$ and D is a minimal, closed subset of M such that $i^*(\gamma^r) \neq 0$ (where $i^*: H^r(M) \to H^r(D)$ is the inclusion map), then D is called a *floor* for γ^r.

LEMMA 1.8. *If $\gamma^r \in H^r(M)$ and $\gamma^r \neq 0$, then there exists a floor for γ^r.*

LEMMA 1.9. *If D is a floor for γ^r, then D can not be disconnected by a T_{r-1} set.*

LEMMA 1.10. *If $\{E^1, \cdots, E^n\}$ is a finite collection of E_{r-1} sets in M, with $M \neq \bigcup_{i=1}^n E^i$, then there exist proper, closed subsets, M_1 and M_2, of M such that* (1) $M = M_1 \cup M_2$, (2) *$M_1 \cap M_2$ is the union of a finite number of T_{r-1} sets (therefore, $M_1 \cap M_2$ is a T_r set),* (3) $M_1 \supset \bigcup_{i=1}^n E^i$.

Proof. The proof will be by induction on n. The case $n = 1$ follows from Lemma 1.3.

Assume the lemma is true up to $n - 1$. Since M is not an E_{r-1} set, we have $M = M_1 \cup M_2$, T_{r-1}-separated. Let $E = \bigcup_{i=1}^n E^i$. If $(M - E) \cap (M - (M_1 \cap M_2)) = \phi$, then the desired T_{r-1}-separation of M could be obtained by using the boundary of an open set in $M_1 \cap M_2$. Therefore, we can assume $(M - E) \cap (M - M_1) \neq \phi$. By Lemma 1.3, we can assume $\bigcup_{i=1}^s E^i \subset M_2$ and $\bigcup_{i=s+1}^n E^i \subset M_1$, where $1 \leq s < n$. We must have $E^i \subset (\overline{M - M_1})$, for $1 \leqq i \leqq s$. Otherwise, we could separate E^i by the T_{r-1} set $(\overline{M - M_1}) \cap (M_1 \cap M_2)$. Since $(M - E) \cap (M - M_1) \neq \phi$, $(\overline{M - M_1}) \neq \bigcup_{i=1}^s E^i$. Thus, by the induction assumption, $(\overline{M - M_1}) = M_4 \cup M_5$, where $\bigcup_{i=1}^s E^i$ is contained in M_4 and $M_4 \cap M_5$ is the union of a finite number of T_{r-1} sets. If we let $\tilde{M}_1 = M_1 \cup M_4$ and $\tilde{M}_2 = M_5$, then

(1) $M = \tilde{M}_1 \cup \tilde{M}_2$,

(2) $\tilde{M}_1 \cap \tilde{M}_2$ is the union of a finite number of T_{r-1} sets,

(3) $\bigcup_{i=1}^n E^i \subset \tilde{M}_1$,

(4) $\tilde{M}_1$ and $\tilde{M}_2$ are proper closed subsets of M.

2. Cyclic elements and the structure of M.

DEFINITION 2.1. A closed subset A of M is called a L_r set if every E_{r-1} set, whose intersection with A is not a T_r set, is contained in A.

The proofs of the following theorems are given below.

THEOREM 2.2. *If A is a L_r set, then $i^*\colon H^r(M) \to H^r(A)$ is onto. Thus, by duality, $i_*\colon H_r(A) \to H_r(M)$ is one-to-one.*

THEOREM 2.3. *Let A be a closed subet with the following property: if E is an E_{r-1} set and $H^r(E) \neq 0$, then E is contained in A. Then the map $i^*\colon H^r(M) \to H^r(A)$ is one-to-one and, by duality, $i_*\colon H_r(A) \to H_r(M)$ is onto.*

THEOREM 2.4. *Suppose there are only a finite number, say $\{E^1, \cdots, E^n\}$, of E_{r-1} sets such that $H^r(E^i) \neq 0$. Let $A = \bigcup_{i=1}^n E^i$. Then the mappings $i^*\colon H^r(M) \to H^r(A)$ and $i_*\colon H_r(A) \to H_r(A)$ are isomorphisms.*

REMARK. Theorem 2.4 can not be generalized to an infinite number of E_{r-1} sets, as the following example shows. In Euclidean space let $M = D \cup [\bigcup_{i=1}^\infty C_i]$, where $D = \{(x, y, z) | z = 0, x^2 + y^2 \leqq 1\}$ and $C_i = \{(x, y, z) | z = 1/i, x^2 + y^2 = 1\}$. We do not have $H_1(\bigcup_{i=1}^\infty C_i) \approx H_1(M)$, under the inclusion mapping.

THEOREM 2.5. *Let $\gamma^r \in H^r(M)$ and suppose U is an open set, such that if D is a floor for γ^r, then D is contained in U (see Definition 1.7). Then there exists a $\gamma_u^r \in H^r(M, M - U)$ such that $\gamma^r = j^*(\gamma_u^r)$, where $j^*\colon H^r(M, M - U) \to H^r(M)$.*

THEOREM 2.6. *Assume E is an E_{r-1} set in M and N is a closed subset of M, where $N \cap E = \phi$. Then the composite mapping $j_* i_*\colon H_r(E) \to H_r(M, N)$ is one-to-one. Here, $i_*\colon H_r(E) \to H_r(M)$ and $j_*\colon H_r(M) \to H_r(M, N)$ are the natural mappings.*

LEMMA 2.7. *Let (M, A) be a compact pair with $\gamma^r \in H^r(A)$. If $\delta^*(\gamma^r) \neq 0$, where $\delta^*\colon H^r(A) \to H^{r+1}(M, A)$, then there is a minimal closed set B such that $B \subset A$, and $\delta_B^*(\gamma_B^r) \neq 0$. Here, $\delta_B^*\colon H^r(B) \to H^{r+1}(M, B)$ and $\gamma_B^r = i^*(\gamma^r)$, where $i^*\colon H^r(A) \to H^r(B)$.*

LEMMA 2.8. *Let B be a minimal set defined in Lemma 2.7. There exists a minimal closed set N such that $\delta^*(\gamma_B^r) \neq 0$, where $\delta^*\colon H^r(B) \to H^{r+1}(N, B)$.*

Proof. The proof of these lemmas is obtained from the continuity of the Cech theory and Zorn's lemma.

LEMMA 2.9. *The set N, in Lemma 2.8, can not be disconnected by a T_{r-1} set.*

Proof. Suppose $N = N_1 \cup N_2$, where $N_1 \cap N_2$ is a T_{r-1} set. We will show this to be impossible, unless $N = N_1$. Let B be as defined in Lemma 2.8, and define $B_i = N_i \cap B$ $(i = 1, 2)$. We will show that the mapping induced by inclusion

$$K^*\colon H^{r+1}(N, B) \to H^{r+1}(N_1, B_1) \oplus H^{r+1}(N_2, B_2)$$

is an isomorphism. We use the relative Mayer-Vietoris sequence given below; note that $T = N_1 \cap N_2$ is a T_{r-1} set [2].

$$\begin{array}{ccccc}
& H^{r+1}(N_2, B_2) & \oplus \; H^{r+1}(N_1, B_1) & \xrightarrow{K^*} & H^{r+1}(N, B) \\
& \uparrow i_1^* & \uparrow i_2^* & & \uparrow i^* \\
H^r(N, N) \to & H^{r+1}(N, N_1 \cup B) & + \; H^{r+1}(N, N_2 \cup B) & \xrightarrow{\bar{K}^*} & H^{r+1}(N, B \cup T) \\
& & & & \to H^{r+1}(N, N) .
\end{array}$$

The mappings i_1^* and i_2^* are isomorphisms by excision, the map $\bar{K}^*$ by exactness. Using the three exact sequences given below we see that i^* is an isomorphism.

$$\begin{gathered}
H^{s-1}(B \cap T) \to H^s(B \cup T) \to H^s(B) \oplus H^s(T) \to H^s(B \cap T) \\
H^s(B \cup T) \to H^s(B) \to H^{s+1}(B \cup T, B) \to H^{s+1}(B \cup T) \to H^{s+1}(B) \\
H^r(B \cup T, B) \to H^{r+1}(N, B \cup T) \to H^{r+1}(N, B) \to H^{r+1}(B \cup T, B) .
\end{gathered}$$

The first is a Mayer-Vietoris sequence, the second is a sequence for a pair, the third is a sequence for a triple. Thus K^* is an isomorphism.

In the diagram below, since $\delta_N^*(\gamma_B^r) \neq 0$, we may assume $\delta_{N_1}^*\phi_1^*(\gamma_B^r) \neq 0$.

$$\begin{array}{ccccc}
H^{r+1}(N, B) & \xrightarrow{\qquad K^* \qquad} & & H^{r+1}(N_1, B_1) & \oplus \; H^{r+1}(N_2, B_2) \\
\uparrow & & \nearrow i_1^* & \uparrow & \uparrow \\
\delta_N^* & & H^{r+1}(M, B_1) & \delta_{N_1}^* & \delta_{N_2}^* . \\
\Big| & & \nwarrow \delta_1^* & \Big| & \Big| \\
H^r(B) & \xrightarrow{\qquad \phi_1^* \oplus \phi_2^* \qquad} & & H^r(B_1) & \oplus \; H^r(B_2)
\end{array}$$

We now have $\delta_1^*\phi_1^*(\gamma_B^r) \neq 0$, since $i_1^*\delta_1^*\phi_1^*(\gamma_B^r) = \delta_{N_1}^*\phi_1^*(\gamma_B^r) \neq 0$. This implies $B_1 = B$, by the definition of B. Therefore, $\phi_1^*(\gamma_B^r) = \gamma_B^r$ and $\delta_{N_1}^*\phi_1^*(\gamma_B^r) = \delta_{N_1}^*(\gamma_B^r) \neq 0$. Since N is minimal, we must have $N_1 = N$. Thus, N can not be disconnected by a T_{r-1} set.

Proof of Theorem 2.2. We will show $\delta^*(\gamma^r) = 0$, for all $\gamma^r \in H^r(A)$, where $\delta^*\colon H^r(A) \to H^r(M, A)$. Suppose not; then choose N and B according to Lemma 2.8. Then there exists an E_{r-1} set containing N, by Lemma 1.6. Let E denote this E_{r-1} set. Since E contains N, we have $E \cap A \supset B$. Since $H^r(B) \neq 0$, B is not a T_r set. Therefore $E \subset A$, because A is an L_r set. This implies that N is contained in A. But this is impossible, as the diagram below shows. By the definition of the pair (N, B), $\delta^* i^*(\gamma^r) \neq 0$.

$$\begin{array}{ccc} H^r(A) & \xrightarrow{\delta_1^*} & H^{r+1}(A, B) \\ \downarrow{\scriptstyle i^*} & & \downarrow \\ H^r(B) & \xrightarrow{\delta^*} & H^{r+1}(N, B)\ . \end{array}$$

Proof of Theorem 2.3. Consider the exact sequence:

$$H^r(M, A) \xrightarrow{j^*} H^r(M) \xrightarrow{i^*} H^r(A)\ .$$

Suppose $j^*(\gamma^r) \neq 0$, where $\gamma^r \in H^r(M, A)$. By Lemmas 1.9 and 1.6 there is an E_{r-1} set which contains a floor for $j^*(\gamma^r)$. Let E be this E_{r-1} set. Since E contains a floor for $j^*(\gamma^r)$, $H^r(E) \neq 0$. Therefore, $E \subset A$; which implies $i^*j^*(\gamma^*) \neq 0$, since E contains a floor for $j^*(\gamma^r)$. Therefore j^* is a trivial map and i^* is one-to-one.

Proof of Theorem 2.4. By Theorem 2.3, $i^*\colon H_r(A) \to H_r(M)$ is onto. If $i_*(Z_r) = 0$, for some $Z_r \in H_r(A)$; then there is a minimal set K such that

(1) $K \supset A$, and

(2) $i_*^K(Z_r) = 0$, where $i_*^K : H_r(A) \to H_r(K)$ [2]. If $K \neq A$; then, by Lemma 1.10, we have $K = K_1 \cup K_2$, T_r-separated. The Mayer-Vietoris sequence below implies $i_*^{K_1}(Z_r) = 0$, where $i_*^{K_1} : H_r(A) \to H_r(K_1)$.

$$H_r(K_1 \cap K_2) \to H_r(K_1) \oplus H_r(K_2) \to H_r(K)\ .$$

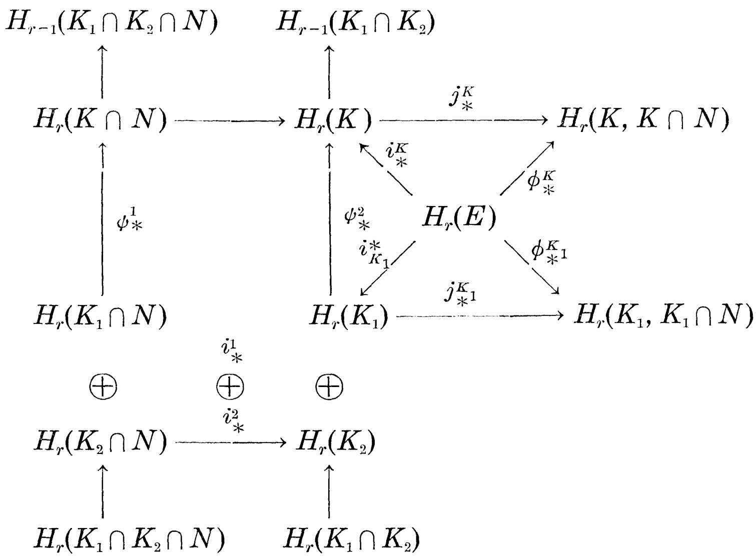
Therefore, $K = A$ and $Z_r = 0$, or i_* is one-to-one.

Proof of Theorem 2.5. Consider the exact sequence,

$$H^r(M, M - U) \xrightarrow{j^*} H^r(M) \xrightarrow{i^*} H^r(M - U)\ .$$

We will show $i^*(\gamma^r) = 0$, where γ^r is the element of $H^r(M)$ given in the theorem. Suppose $i^*(\gamma^r) \neq 0$; then, by Lemma 1.8, there exists a floor for $i^*(\gamma^r)$ contained in $M - U$. If D is this floor, then D is a floor for γ^r, since $i_D^* = i_{DU}^* i^*$. Here, $i_D^*\colon H^r(M) \to H^r(D)$ and $i_{DU}^*\colon H^r(M - U) \to H^r(D)$ are inclusion mappings. Therefore, by the definition of U, D is contained in U. This is impossible, hence $i^*(\gamma^r) = 0$.

Proof of Theorem 2.6. Let $\phi_* = j_* i_*$, and suppose $\phi_*(Z_r) = 0$, for some $Z_r \in H_r(E)$. Then there exists a minimal closed set K in M such that $K \supset E$ and $\phi^K_*(Z_r) = 0$, where $\phi^K_*\colon H_r(E) \to H_r(K, K \cap N)$ is analogous to ϕ_* defined above. This follows from Zorn's lemma and continuity. We will assume $K \neq E$. Since E is an E_{r-1} set, we can write $K = K_1 \cup K_2$, T_{r-1}-separated. Also, we can assume $E \subset K_1$. Consider the following commutative diagram:



The two vertical sequences are Mayer-Vietoris sequences. Also, the two horizontal sequences are exact. We have

$$H_{r-1}(K_1 \cap K_2) = H_{r-1}(K_1 \cap K_2 \cap N) = H_r(K_1 \cap K_2) = H_r(K_1 \cap K_2 \cap N) = 0\ ,$$

since $K_1 \cap K_2$ is a T_{r-1} set. Since $\phi^K_*(Z_r) = j^K_* i^K_*(Z_r) = 0$, there exists a $Z^3_r \in H_r(K \cap N)$ such that $i^3_*(Z^3_r) = i^K_*(Z_r)$. There exists

$$(Z^1_r, Z^2_r) \in H_r(K_1 \cap N) \oplus H_r(K_2 \cap N)$$

such that $\psi^1_*(Z^1_r, Z^2_r) = Z^3_r$. By commutativity,

$$\psi^2_*(i^1_*(Z^1_r), i^2_*(Z_r)) i^2_* \psi^1_*(Z^1_r, Z^2_r) = i^3_*(Z^3_r) = i^K_*(Z_r)\ ,$$

and

$$\psi^*_2(i^{K_1}_*(Z_r), 0) = i^K_*(Z_r)\ .$$

By exactness, ψ^2_* is an isomorphism, hence $i^1_*(Z^1_r) = i^{K_1}_*(Z_r)$. Therefore, $j^{K_1}_* i^{K_1}(Z_r) = j^{K_1}_* i^1_*(Z^1_r) = 0$. But this is impossible, since K is minimal. Thus, $K = E$ and ϕ^* is one-to-one.

3. Cyclic elements in locally connected spaces. The zero-dimensional cyclic elements in a locally connected continuum have several useful properties. For example, if the continuum M is locally connected, then the zero-dimensional cyclic elements of M are also locally connected and these cyclic elements form a null sequence. Also, the simple 0-links (definition below) are identical with the E_0 sets in an lc^0 space [6]. The examples below show that these properties do not generalize.

DEFINITION 3.1. A non-degenerate subset K of M is called a simple r-link of M, if K is maximal with respect to the following property: if $M = M_1 \cup M_2$, T_r-separated, then $K \subset M_1$ (or $K \subset M_2$). In other words, K is a maximal subset which can not be separated by a T_r set that also separates M.

LEMMA 3.2. *All simple r-links in M are closed. If K_1 and K_2 are two distinct simple r-links in M, then $K_1 \cap K_2$ is a T_r set. If L is a non-degenerate subset of M that is not disconnected by any T_r set which also disconnects M, then L is contained in a simple r-link of M.*

Proof. The proof is similar to those for the corresponding lemmas for cyclic elements.

EXAMPLE. We will construct an lc^r space M in which the collection of E_r sets does not form a null sequence. This example will also show that, in an lc^r space, the simple r-links need not be the same as the E_r sets.

For each positive integer n, let R_n be a solid, three dimensional rod of height one and diameter $1/2^n$. In Euclidean three-space, define I by $I = \{(x, y, z) | x = 0, y = 0, 0 \leqq z \leqq 1\}$. Imbed R_n in three-space so that R_n is tangent to R_{n+1} and the sequence of sets R_n converges to I (i.e. $R_n = \{(x, x, z) | x^2 + (y - 3/2^{n+1})^2 \leqq 1/2^{2n+2}, 0 \leqq z \leqq 1\}$). Let M be the set $[\bigcup_{n=1}^{\infty} R_n] \cup I$. Then M is a compact $1c^1$ space, each R_n is an E_1 set in M, but the collection $\{R_n\}$ is not a null sequence. Also, I is a simple 1-link, but is not an E_1 set.

THEOREM 3.3. *If M is $s - lc$ and E is an E_r set of M, where $s \geqq r$, then E is $s - lc$.*

Proof. Given any $x \in E$, and an open set U^0 of E containing x, then there exists an open set U of M such that $U \cap E = U^0$. Since M is $s - lc$, there exists an open set V, containing x, such that $\bar{V} \subset U$ and any compact s-cycle in V bounds on a compact subset of U. Let Z_s be a compact cycle on $V \cap E = V^0$. Then there exists a minimal closed set K in M such that $\bar{V}^0 \subset K \subset U$, and Z_s bounds on K. By using the Mayer-Vietoris sequence, as it was used in the proof of Theorem 2.4, we can show $K \subset U^0$. Therefore Z_s bounds in U^0 and E is $s - lc$.

EXAMPLE. We will construct a compact lc^r space which contains an E_r set which is not lc^r. Consider the following curve in three-space:

$$x = 0, y = t, z = \sin(\pi/t), \text{ for } 0 < t \leqq 1 .$$

Expand this curve slightly so that it becomes a solid, three dimensional figure, which oscillates as it approaches the origin. Let N be this space, along with its limiting line segment on the z-axis. Let $P = \{(x, y, z) \mid x = 0, 0 \leqq y \leqq 1, -1 \leqq z \leqq 1\}$; then define $M = P \cup N$. Thus N is an E_1 set in M and M is lc^1 but N is not $0 - lc$.

4. Cyclic elements and monotone mappings. A very basic property of the zero-dimensional cyclic element theory is the following: if $f: M \rightarrow N$ is a monotone mapping (*i.e.* the inverse image of any point is connected), M and N are lc^0, and E_N is an E_0 set in N; then there is an E_0 set in M whose image under f contains E_N. This result does not hold in higher dimensions, as the example below demonstrates. The best result we have obtained in this direction is Theorem 4.2.

DEFINITION 4.1. A mapping f: $M \rightarrow N$ is r-monotone, if $H^s(f^{-1}(y)) = 0$, for all $y \in N$ and $0 \leqq s \leqq r$.

THEOREM 4.2. *Let f be an $(r-1)$-monotone mapping of M onto N, where M and N are compact Hausdorff spaces. If D_N is a floor for $\gamma_N^r \in H^r(N)$, then there exists a floor D_M for $f^*(\gamma_N^r)$ such that $f(D_M) = D_N$.*

Proof. Since f is $(r-1)$-monotone, f^*: $H^r(N) \rightarrow H^r(M)$ is a one-to-one mapping [1]. Therefore, $f^*(\gamma_N^r) \neq 0$. Consider the commutative diagram below. The vertical mappings are inclusion mappings; and D_M is defined below.

$$\begin{array}{ccc} M & \xrightarrow{\quad f \quad} & N \\ \uparrow{\scriptstyle i_M} & & \uparrow{\scriptstyle i_N} \\ f^{-1}(D_N) & \xrightarrow{\quad f_1 \quad} & D_N \\ \uparrow{\scriptstyle j_M} & & \uparrow{\scriptstyle j_N} \\ D_M & \xrightarrow{\quad f_2 \quad} & f(D_M) \end{array} .$$

The mapping f_1 is the restriction of f to $f^{-1}(D_N)$. Therefore, f_1 is $(r-1)$-monotone. Since D_N is a floor for γ_N^r, $i_N^*(\gamma_N^r) \neq 0$. Since

f_1^*: $H^r(D_N) \to H^r(f^{-1}(D))$ is one-to-one, $i_M^* f^*(\gamma_N^r) = f_1^* i_N^*(\gamma_N^r) \neq 0$. Therefore, $f^{-1}(D_N)$ contains a floor for $f^*(\gamma_N^r)$. Denote this floor by D_M and let f_2 be the restriction of f to D_M. By the definition of a floor, $j_M^* i_M^* f^*(\gamma_M^r) \neq 0$. Since $j_M^* i_M^* f^*(\gamma_N^r) = f_2^* j_N^* i_N^*(\gamma_N^r)$, we have $j_N^* i_N^*(\gamma_N^r) \neq 0$. This implies $f(D_M) = D_N$, since D_N is a floor for γ_N^r.

We shall omit the proofs of Lemmas 4.3 and 4.5.

LEMMA 4.3. *Let N_1 and N_2 be subsets of M which can not be disconnected by a T_r set. Suppose that $\bar{N}_1 \cup \bar{N}_2$ is not a T_r set. Then $\bar{N}_1 \cup \bar{N}_2$ can not be disconnected by a T_r set.*

LEMMA 4.4. *Let $f: M \to N$, and suppose $T \subset N$ is a T_r set such that $f^{-1}(T)$ is also a T_s set. Also, assume f is a homeomorphism of $M - f^{-1}(T)$ onto $N - T$. Then, if T^N is a T_r set in N, $f^{-1}(T^N)$ is a T_r set in M.*

Proof. Let K be a closed subset of $f^{-1}(T^N)$. Denote $f^{-1}(T)$ by T^{-1}. In the commutative diagram below f_1^* is an isomorphism, by excision. Therefore, by exactness, $H^r(K) = 0$.

$$\begin{array}{ccccc} H^r(K, K \cap T^{-1}) & \to & H^r(K) & \to & H^r(K \cap T^{-1}) \\ \uparrow f_1^* & & \uparrow f^* & & \\ H^r(f(K), f(K \cap T^{-1})) & \to & H^r(f(K)) & \to & \end{array}$$

LEMMA 4.5. *Assume f is a mapping of M onto N such that the inverse image of any T_r set in N is a T_r set in M. If $K \subset M$ can not be disconnected by a T_r set in M, then $f(K)$ can not be disconnected a T_r set in N.*

EXAMPLE. If f is an r-monotone mapping of M onto N, where M and N are lc^∞ spaces and E^N is an E_r set in N; there may not be an E_r set, E^M, in M such that $f(E^M) \supset E^N$.

We will construct the example in three space. Consider the following solid cylinders:

$$M_1 = \{(x, y, z) \mid x^2 + y^2 \leqq 1,\ 0 \leqq z \leqq 1\}$$
$$M_2 = \{(x, y, z) \mid x^2 + (y - 2)^2 \leqq 1,\ 0 \leqq z \leqq 1\}\ .$$

The cylinders M_1 and M_2 are tangent along $I = \{(x, y, z) \mid x = 0, y = 1, 0 \leqq z \leqq 1\}$. Let M_3 be an arc joining the endpoints of I, which does not meet $M_1 \cup M_2$ except at these endpoints. Let $M = \bigcup_{i=1}^3 M_i$. We will define a decomposition of M, and will let $f: M \to N$ be the decomposition mapping.

To form N, identify all the points in M_3 into a single point. Then the mapping $f: M \to N$ is r-monotone for all r and the restriction of f to $M - M_3$ is a homeomorphism.

We will show that N is an E_1 set. First, neither M_1 nor M_2 can be disconnected by a T_1 set. Lemmas 4.4 and 4.5 imply that neither $f(M_1)$ nor $f(M_2)$ can be disconnected by a T_1 set. By Lemma 4.3, $N = f(M_1) \cup f(M_2)$ can not be disconnected by a T_1 set, since $f(M_1) \cup f(M_2)$ contains an essential 1-cycle. If K is a closed subset of M such that $f(K) \supset N$, then $K \supset M_1 \cup M_2$. Then K can be disconnected by a T_1 set, namely $M_1 \cap M_2$. Therefore, there is no E_1 set in M whose image is N.

Note that M is obviously lc^r for all r. Therefore N is also lc^r, for all r, since f is r-monotone, for all r.

References

1. E. G. Begle, *The Vietoris mapping theorem for bicompact spaces*, Annals of Math., (2) **51** (1950), 534-542.

2. Samuel Eilenberg and Norman Steenrod, *Foundations of Algebraic Topology*, Princeton, 1952.

3. W. J. Puckett, Jr., *A theorem concerning homologies in a compact space*, Fundamentae Mathematicae, **31-32** (1938), 32.

4. Arthur Simon, *N-Cyclic elements*, Duke Math. Journal, **24** (1957), 1-7.

5. G. T. Whyburn, *Cyclic elements of higher order*, Amer. J. Math., **56** (1934), 133-146.

6. ———, *Analytic Topology*, Amer. Math. Soc. Coll. Pub., 1942.

7. R. L. Wilder, *Topology of Manifolds*, Amer. Math. Soc. Coll. Pub. 1949.

California Institute of Technology

ON DIOPHANTINE APPROXIMATION AND TRIGONOMETRIC POLYNOMIALS

RICHARD P. GOSSELIN

The usefulness of Diophantine approximation in achieving both positive and negative results in the subject of trigonometric interpolating polynomials is well established (cf. e.g. [1], [4]). The trigonometric polynomials, hereafter called simply polynomials, which we shall consider mainly and designate by $I_{n,u}(x;f)$ are those of order n taking on the values of a given function f at the points $u + 2\pi k/(2n+1)$, $k = 0, 1, \cdots, 2n$. Thus

$$I_{n,u}(x;f) = \frac{2}{2n+1}\sum_{k=0}^{2n} f(u + x_k^{(n)})D_n(x - u - x_k^{(n)}),$$

$$D_n(x) = \frac{\sin(2n+1)x/2}{2\sin(x/2)},\; x_k^{(n)} = \frac{2\pi k}{2n+1}\,.$$

It is assumed that f is periodic and defined almost everywhere so that for almost every u, $I_{n,u}(x;f)$ is defined for all n. Marcinkiewicz and Zygmund [4] have shown that each p, $1 \leqq p < 2$, there is a function f of class L^p such that for almost every point of the square $0 \leqq x \leqq 2\pi$, $0 \leqq u \leqq 2\pi$, $I_{n,u}(x;f)$ diverges. They made strong use of the following classical result of Diophantine approximation: for each x there are infinitely many rationals p/q such that $|x - p/q| \leqq 1/q^2$.

Our aim in this paper is to generalize the result of Marcinkiewicz and Zygmund. The chief tool of proof is a result proved in the next section, concerning the approximation of reals by rationals in which the range of the denominators is restricted. In the third section we give our main theorem to the effect that for any increasing function ψ defined on $(0, \infty)$ there is an f such that $\psi(|f|)$ is integrable over $0 \leqq x \leqq 2\pi$ and such that $I_{n,u}(x;f)$ diverges for almost every (x, u). In the last section we show this result holds for Jackson polynomials.

2. We begin with a preliminary lemma. If F is a measurable set, $|F|$ will denote its measure. We shall let C, C_1, and C_2 denote constants, independent of the values of the integers N, M, and m.

LEMMA 1. *Let N, M, and m be three integers such that $0 \leqq N < M \leqq m/2$. Let F be the subset of $(0, 1)$ such that for each x in F there is an irreducible rational p/q, $0 < p < q$, $N < q \leqq M$ satisfying $|x - p/q| \leqq$*

Received December 10, 1958. This work was supported by the National Science Foundation through Research Grant NSF G-2789.

$1/qm$. *Then*

$$\frac{12(M-N)}{\pi^2 m} - \frac{C}{m}\log^2(M+1) \leqq |F| \leqq \frac{12(M-N)}{\pi^2 m} + \frac{C}{m}\log^2(M+1)\,.$$

If only $0 \leqq N < M \leqq m$, *then the second inequality above holds.*

F is the union of intervals of the form $(p/q - 1/qm, p/q + 1/qm)$. The number of irreducible rationals with denominator q of the above form is $\phi(q)$ where ϕ is the Euler function. The contribution to the measure of F from a given q is no more that $2\phi(q)/qm$ so that the measure of F does not exceed

$$\frac{2}{m}\sum_{q=N+1}^{M}\frac{\phi(q)}{q}$$

Let $\psi(0) = 0, \psi(n) = \sum_{q=1}^{n}\phi(q)$. Applying Abel's transformation to the above sum, we obtain

$$(1) \qquad |F| \leqq \frac{2}{m}\sum_{q=N+1}^{M}\frac{\psi(q)}{q(q+1)} + \frac{2}{m}\left\{\frac{\psi(M)}{M+1} - \frac{\psi(N)}{(N+1)}\right\}.$$

By a known theorem (cf. e.g. [3, p. 120])

$$(2) \qquad \frac{3q^2}{\pi^2} - C_1 q\log(q+1) \leqq \psi(q) \leqq \frac{3q^2}{\pi^2} + C_1 q\log(q+1)\,.$$

Substitution of (2) into (1) gives

$$|F| \leqq \frac{6}{\pi^2 m}\sum_{q=N+1}^{M}\frac{q}{q+1} + \frac{6M}{\pi^2 m} - \frac{6N^2}{\pi^2 m(N+1)} + \frac{C_2}{m}\log^2(M+1)\,.$$

This implies the second statement of the lemma. In case $M \leqq m/2$, there is no overlapping of the (open) intervals $(p/q - 1/qm, p/q + 1/qm)$. For otherwise, there are distinct rational $r/s, p/q$ (let us say $r/s > p/q$) of the required form such that

$$0 < \frac{r}{s} - \frac{p}{q} < \frac{1}{sm} + \frac{1}{qm} \text{ and } 0 < rq - ps < \frac{q+s}{m} \leqq 1\,.$$

This contradicts the fact that $rq - ps$ is an integer. Thus

$$|F| = \frac{2}{m}\sum_{q=N+1}^{M}\frac{\phi(q)}{q}\,.$$

Now the inequality (2) implies the lemma.

THEOREM 1. (i) *Let* m *be a sufficiently large positive integer, and let* γ *be a real number such that* $0 < \gamma < \pi^2/12$. *Let* E *be the subset of*

$(0, 1)$ *such that for each* x *in* E *there exists an irreducible rational* $p/q, 0 < p < q, \gamma m < q \leqq m$ *for which* $|x - p/q| \leqq 1/\gamma m^2$. *Then there is an absolute constant* C *such that*

$$|E| \geqq 1 - \frac{12\gamma}{\pi^2} - Cm^{-1}\log^2 m .$$

(ii) *Let* γ *be a real number such that* $0 < \gamma < \pi^2/24$. *Let* E_1 *be the subset of* $(0, 1)$ *such that for each* x *in* E_1 *there exists an irreducible rational* $p/q, 0 < p < q, \gamma m < q \leqq m$, *with* q *odd for which* $|x - p/q| \leqq 2/\gamma^2 m^2$. *Then there is an absolute constant* C *such that*

$$|E_1| \geqq 1 - \frac{24\gamma}{\pi^2} - Cm^{-1}\log^2 m .$$

As in the proof of the theorem mentioned in the introduction (cf. [6, p. 43]) we may find for each x in $(0, 1)$ an irreducible rational p/q such that

$$(3) \qquad |x - p/q| \leqq \frac{1}{qm}, \; 0 < q \leqq m .$$

If x is restricted to the (open) interval $I = (1/m, 1 - 1/m)$, then $0 < p < q$. We shall say q and x are associated if (3) holds with x in I and with p/q irreducible, $0 < p < q, 0 < q \leqq m$. Let F_1 be the subset of I for which all q associated with x do not exceed γm. Since each x is associated with some q, the set F_1 is a subset of the set F of Lemma 1 for which $N = 0$ and $M = [\gamma m]$, the greatest integer not exceeding γm. We may assume without loss of generality that $\gamma m > 1$. Let E be the complement of F_1 with respect to I. Since the measure of F does not exceed $12\gamma/\pi^2 + Cm^{-1}\log^2 m$, part (i) follows from (3) and the inequality $q > \gamma m$.

Let F_2 be the subset of I for which all q associated with an x in F_2 are such that $(1 - \gamma)m < q \leqq m$. F_2 is a subset of the set F of Lemma 1 for which $M = m, N = [(1 - \gamma)m]$. Let E_1 be the complement of $F_1 \cup F_2$ with respect to I. Then $|E_1| \geqq 1 - 24\gamma/\pi^2 - Cm^{-1}\log^2 m$. If x belongs to E_1, there is a q associated with x such that $\gamma m \leqq q \leqq m(1 - \gamma)$. If q is even, we may find integers η and ξ such that

$$(4) \qquad \eta p - \xi q = 1$$

where η must be odd, and automatically ξ/η is irreducible. Let η_0 be the least positive solution of (4) (cf. [1] for a similar argument). If $\eta_0 \geqq \gamma m$, it follows that

$$\left|\frac{p}{q} - \frac{\xi_0}{\eta_0}\right| = \frac{1}{q\eta_0} \leqq \frac{1}{\gamma^2 m^2}$$

and

$$(5)\qquad \left|x-\frac{\xi_0}{\eta_0}\right| \leqq \left|x-\frac{p}{q}\right| + \left|\frac{p}{q}-\frac{\xi_0}{\eta_0}\right| \leqq \frac{1}{qm}+\frac{1}{\gamma^2 m^2} \leqq \frac{2}{\gamma^2 m^2}$$

If $\eta_0 < \gamma m$, let $\eta_1 = \eta_0 + q$. Then $\gamma m \leqq q \leqq \eta_1 \leqq \gamma m + q \leqq m$, and (5) holds with ξ_0/η_0 replaced by ξ_1/η_1. We may assume that $\gamma^2 > 1/m$ so that $0 < \xi < \eta \leqq m$ as required.

3. We begin this section with a lemma which is related to the results of the preceding section, but it contains only as much information as will be used in the proof of the next theorem.

LEMMA 2. *Let m be a sufficiently large integer, A_m a real satisfying $1 \leqq A_m \leqq \log m$, and $d \log m$ an integer with $8 < d < 10$. Let $\mathcal{N}$ be the set of odd positive integers $2n+1$ not exceeding m and such that*

$$(6)\qquad \left|\frac{\mu m}{\nu} - (2n+1)\right| \leqq \frac{4A_m^{1/2}}{\nu} \text{ for some } (\mu, \nu) \text{ such that}$$

$$0 < \mu \leqq \nu \leqq d \log m.$$

Let G be the subset of $(0, 1)$ such that for x in G, there is a $2n+1$ in $\mathcal{N}$ and a k, $0 < k < 2n+1$ for which $|x - k/(2n+1)| \leqq 2A_m^{1/2}/m^2$. Then

$$|G| \leqq \frac{36 d^2 \log^3 m}{m}.$$

For a given μ and ν, no more than $1 + 8A_m^{1/2}/\nu$ integers $2n+1$ satisfy (6). For a given ν, no more than $\nu + 8A_m^{1/2}$ integers may satisfy (6) for some $\mu \leqq \nu$. Hence N, the number of distinct integers in $\mathcal{N}$, does not exceed $d^2 \log^2 m + 8dA_m^{1/2} \log m$. If x belongs to G, x is contained in an interval of length $4A_m^{1/2}/m^2$ centered about some point $k/(2n+1)$. For each $2n+1$, the total length of the intervals is no more than $4A_m^{1/2}/m$. Thus,

$$|G| \leqq \frac{4NA_m^{1/2}}{m} \leqq \frac{36 d^2 \log^3 m}{m}.$$

THEOREM 2. *Let ψ be a monotone increasing function defined on $(0, \infty)$. There exists a function f such that $\psi(|f|)$ is integrable on $(0, 2\pi)$ and such that the sequence $I_{n,u}(x; f)$ diverges for almost all points of the square $0 \leqq x \leqq 2\pi$, $0 \leqq u \leqq 2\pi$.*

Let A_m be a positive number satisfying the inequality $16 \leqq A_m \leqq (\log m)^{1/2}$. A more exact specification of A_m will be given at a later point. The function f will be a sum of periodic, step functions f_m of the following form. When x belongs to one of the intervals

$$|x - 2\pi j/m| \leqq 4\pi A_m^{1/2}/m^2, \qquad j = 0, 1, \cdots, m-1$$

let $f_m(x) = A_m$; when x belongs to one of the complementary intervals of $(0, 2\pi)$ let $f_m(x) = 0$. Let E_1 be the set of Theorem 1, part (ii), corresponding to m and $\gamma = A_m^{-1/4}$, and expanded to the interval $(0, 2\pi)$ on the u axis. For m sufficiently large, $|E_1| \geqq 2\pi(1 - 25/\pi^2 A_m^{1/4})$. Let G be the set of Lemma 2 expanded to the interval $(0, 2\pi)$ on the u axis. Let E_m be the difference set $E_1 - (G \cup G_1)$ where G_1 is the set of u such that $|u| \leqq 2\pi/(\log m)^{1/2}$. By our above estimates

$$(7) \qquad |E_m| \geqq 2\pi\left[1 - \frac{25}{\pi^2 A_m^{1/4}} - \frac{36d^2 \log^3 m}{m} - \frac{2}{(\log m)^{1/2}}\right].$$

Let $E_{m,j}$ be the set E_m translated by $-2\pi j/m$, $j = 0, 1, \cdots, m-1$: *i.e.* u belongs to $E_{m,j}$ if and only if $u + 2\pi j/m$ (modulo 2π) belongs to E_m. Let $-u$ belong to $E_{m,j}$ We may assume that $-u + 2\pi j/m$ belongs to the interval $(0, 2\pi)$. Since E_m is a subset of E_1, there exists, according to Theorem 1, part (ii), an odd integer, $2n+1$, $m/A_m^{1/4} \leqq 2n + 1 \leqq m$, and an integer k, $0 < k < 2n+1$, such that

$$(8) \qquad \left|u - \frac{2\pi j}{m} + \frac{2\pi k}{2n+1}\right| \leqq \frac{4\pi A_m^{1/2}}{m^2}.$$

This inequality implies that $f_m(u + 2\pi k/(2n+1)) = A_m$. Since $-u + 2\pi j/m$ does not belong to the set G, the integer $2n+1$ cannot belong to the set $\mathscr{N}$ defined by (6). If $f_m(u + 2\pi(k+\mu)/(2n+1)) = A_m$ for some nonzero integer μ, then there must be a nonzero integer ν such that

$$(9) \qquad \left|u - \frac{2\pi(j+\nu)}{m} + \frac{2\pi(k+\mu)}{2n+1}\right| \leqq \frac{4\pi A_m^{1/2}}{m^2}.$$

We may assume that $\mu > 0, \nu > 0$. The inequalities (8) and (9) imply that

$$(10) \qquad \left|\frac{\mu}{2n+1} - \frac{\nu}{m}\right| \leqq \frac{4A_m^{1/2}}{m^2}$$

and (10) implies that $\mu \leqq \nu$. For if $\mu > \nu$, then

$$\frac{\mu}{2n+1} - \frac{\nu}{m} \geqq \frac{1}{2n+1} > \frac{4A_m^{1/2}}{m^2}.$$

It also follows from (10) that

$$\left|\frac{\mu m}{\nu} - (2n+1)\right| \leqq \frac{4A_m^{1/2}(2n+1)}{\nu m} \leqq \frac{4A_m^{1/2}}{\nu}\,.$$

Comparison of this inequality with (6) shows that $|\nu| > d(\log m)$. Our analysis shows, in fact, that if $f_m(u + x_i^{(n)}) = A_m$, then $f_m(u + x_{i+\nu}^{(n)}) = 0$ when $|\nu| \leqq d(\log m)$ and $2n+1$ does not belong to $\mathcal{N}$. For each $j = 0, 1, \cdots, m-1$, let I_j be the set of the x axis defined by

$$\frac{\pi}{mA_m^{1/4}} \leqq \left|x - \frac{2\pi j}{m}\right| \leqq \frac{\pi}{m}\,.$$

If x belongs to I_j, and if $-u$ belongs to $E_{m,j}$, then we find from (8) that

$$\left|x - u - \frac{2\pi k}{2n+1}\right| \leqq \left|x - \frac{2\pi j}{m}\right| + \left|u + \frac{2\pi k}{2n+1} - \frac{2\pi j}{m}\right|$$
$$\leqq \frac{\pi}{m} + \frac{4\pi A_m^{1/2}}{m^2} < \frac{3\pi}{2m}$$

for some k and for some n for which $m/A_m^{1/4} \leqq 2n+1 \leqq m$. Furthermore

$$\left|x - u - \frac{2\pi k}{2n+1}\right| \geqq \left|x - \frac{2\pi j}{m}\right| - \left|u + \frac{2\pi k}{2n+1} - \frac{2\pi j}{m}\right|$$
$$\geqq \frac{\pi}{mA_m^{1/4}} - \frac{4\pi A_m^{1/2}}{m^2} > \frac{\pi}{2mA_m^{1/4}}\,.$$

These inequalities imply that

(11)
$$\frac{1}{\left|\sin\frac{1}{2}\left(x - u - \frac{2\pi k}{2n+1}\right)\right|} > \frac{m}{3};\ \left|\sin\left(n + \frac{1}{2}\right)(x-u)\right|$$
$$\geqq \sin\frac{\pi}{4A_m^{1/2}} \geqq \frac{1}{2A_m^{1/2}}\,.$$

Now we are ready to estimate $I_{n,u}(x; f_m)$ with x in I_j and $-u$ in $E_{m,j}$.

(12)
$$I_{n,u}(x; f_m) = \frac{f_m(u + x_k^{(n)})}{2n+1}\,\frac{(-1)^k \sin(n+1/2)(x-u)}{\sin\frac{1}{2}(x - u - x_k^{(n)})}$$
$$+ \frac{\sin(n+1/2)(x-u)}{2n+1}\sum_{i\neq k}\frac{f_m(u + x_i^{(n)})(-1)^i}{\sin\frac{1}{2}(x - u - x_i^{(n)})}\,.$$

Denote the first and second terms on the right by D_1 and D_2 respectively.

By (8) and (11)

$$(13) \qquad |D_1| \geqq \frac{2|\sin(n+1/2)(x-u)|mA_m}{3(2n+1)} \geqq |\sin(n+1/2)(x-u)|\frac{A_m}{3}.$$

We may assume that for the terms of D_2, $|x_i^{(n)} - x_k^{(n)}| \leqq \pi$ so that except possibly for one term of the sum which can be ignored, $|x - u - x_i^{(n)}| \leqq \pi$. Hence for the terms of D_2, $|\sin 2^{-1}(x-u-x_i^{(n)}| \geqq |x-u-x_i^{(n)}|/\pi$, and

$$|D_2| \leqq \frac{\pi|\sin(n+1/2)(x-u)|}{2n+1} \sum_{i \neq k} \frac{f_m(u + x_i^{(n)})}{|x - u - x_i^{(n)}|}.$$

The denominator of the terms in the sum increases with $|i-k|$. Furthermore if i and i' are distinct values of the index for which the numerator is nonzero, then $|i-k| > d\log m$, $|i'-k| > d\log m$, and $|i - i'| > d\log m$. Thus we find that

$$\begin{aligned} |D_2| &\leqq \frac{2\pi|\sin(n+1/2)(x-u)|A_m}{2n+1} \sum_{r=1}^{M} \frac{2n+1}{2\pi r d\log m} \\ &\leqq \frac{|\sin(n+1/2)(x-u)|}{d\log m} A_m \log(M+1), M = \left\langle \frac{2n+1}{2d\log m} \right\rangle. \end{aligned}$$

We denote by $\langle y \rangle$ the least integer $\geqq y$. From this inequality and from (11), (12), and (13), we deduce that if x belongs to I_j, and if $-u$ belongs to $E_{m,j}$, there exists an integer, $2n+1$, and a positive constant C such that

$$(14) \qquad |I_{n,u}(x; f_m)| \geqq C|\sin(n+1/2)(x-u)|A_m \geqq \frac{CA_m^{1/2}}{2}.$$

The product set $I_j \times E_{m,j}$ of the xu-plane has two dimensional measure equal to $2\pi|E_m|(1 - A_m^{-1/4})/m$. There are m such mutually disjoint sets, and the total measure of their union, H_m, is $2\pi|E_m|(1 - A_m^{-1/4})$. Thus if $(x, -u)$ belongs to H_m, then (14) holds for the proper n. We note here that $\lim|H_m| = 4\pi^2$ if $\lim A_m = \infty$.

Let

$$(15) \qquad f(x) = \sum_{i=1}^{\infty} f_{m(i)}(x).$$

We shall impose various conditions on the sequence of positive integers $m(i)$, all related to the rapidity of its growth. Let $\{B_i\}$ be a sequence of reals going to ∞ so that $\sum_{j<i} B_j \leqq B_i^{1/4}$. Let $m(i)$ increase so rapidly that $\log m(i) \geqq B_i^2$ and that

$$(16) \qquad \psi(2B_i)\frac{(\log m(i))^{1/4}}{m(i)} \leqq 2^{-i}.$$

Let $A_{m(i)} = B_i$ so that $A_{m(i)} \leqq (\log m(i))^{1/2}$ as required. Now $f_m(x)$ is 0 except on a set of measure not exceeding $4\pi A_m^{1/2}/m \leqq 4\pi(\log m)^{1/4}/m$. Let

$$\rho_j = \sum_{i=j}^{\infty} \frac{(\log m(i))^{1/4}}{m(i)}, \sum_{j=1}^{\infty} \rho_j < \infty .$$

It follows that the series in (15) converges almost everywhere and that $\sum_{j=i}^{\infty} f_{m(j)}(x)$ is 0 outside a set of measure $4\pi\rho_i$. Let K_i be the set of x values for which $f_{m(i)}(x) \neq 0$, and $f_{m(j)}(x) = 0$ when $j > i$. The K_i's are mutually disjoint, and their union is, except for a set of measure 0, the set where $f(x) \neq 0$. Moreover, $|K_i| \leqq 4\pi(\log m(i))^{1/4}/m(i)$. When x is in K_i,

$$\psi(|f(x)|) \leqq \psi\Big(\sum_{j=1}^{i} f_{m(j)}(x)\Big) \leqq \psi(2B_i) .$$

Thus by (16)

$$\int_0^{2\pi} \psi(|f(x)|)dx \leqq \sum_{i=1}^{\infty} \psi(2B_i)|K_i| < \infty .$$

In the estimation of the interpolating polynomials, we shall require certain other conditions. Thus we assume that f belongs to L^p for some $p > 1$ and that

$$|K_j|(2B_j)^p \leqq \frac{2^{-j}}{m(j-1)}, \qquad j > 1 .$$

From this it follows that

$$\int_0^{2\pi} |\sum_{j>i} f_{m(j)}(x)|^p dx \leqq \sum_{j>i} |K_j|(2B_j)^p \leqq \frac{2^{-i}}{m(i)} . \tag{17}$$

Furthermore we note that $\sum_{j=1}^{i-1} f_{m(j)}(x)$ is a function of bounded variation so that, for each u, the interpolating polynomials converge to the function at every point of continuity, *i.e.* outside a finite set [6; p. 36]. Thus, given $m(1), m(2), \cdots, m(i-1)$, we choose $m(i)$ so large that for $2n + 1 \geqq m(i)/B_i^{1/4}$,

$$\Big|I_{n,u}(x; \sum_{j=1}^{i-1} f_{m(j)})\Big| \leqq 2 \max_x \Big(\sum_{j=1}^{i-1} f_{m(j)}(x)\Big) \leqq 2A_{m(i)}^{1/4} \tag{18}$$

for (x, u) outside a set of two dimensional measure not exceeding 2^{-i}. Finally since $\lim|H_m| = 4\pi^2$, the $m(i)$ can be spread out so sparsely that

$$\sum_{i=1}^{\infty} |H'_{m(i)}| < \infty \tag{19}$$

where H'_m is the complement of H_m.

To estimate $I_{n,u}(x;f)$, we let $m(i)/B_i^{1/4} \leqq 2n+1 \leqq m(i)$. Then

$$(20)\qquad I_{n,u}(x;f) = I_{n,u}(x;\sum_{j<i} f_{m(j)}) + I_{n,u}(x;f_{m(i)}) + I_{n,u}(x;\sum_{j>i} f_{m(j)}) .$$

Let $g(x, u)$ be the maximum of the absolute value of the last term on the right for $2n+1 \leqq m(i)$. A result of Marcinkiewicz and Zygmund [4] implies that

$$\int_0^{2\pi}\int_0^{2\pi} |g(x, u)|^p dxdu \leqq \sum_{2n+1 \leqq m(i)} \int_0^{2\pi}\int_0^{2\pi} |I_{n,u}(x;\sum_{j>i} f_{m(j)})|^p dxdu$$
$$\leqq C_p m(i)\int_0^{2\pi} |\sum_{j>i} f_{m(j)}(x)|^p dx$$

and the last term on the right does not exceed $C_p 2^{-i}$ by (17). C_p is a constant depending only on p. Thus

$$\max_{2n+1 \leqq m(i)} |I_{n,u}(x;\sum_{j>i} f_{m(j)})| \leqq C_p^{1/p}$$

outside a set of measure 2^{-i}. This, together with (18) and (20), implies

$$|I_{n,u}(x;f)| \geqq |I_{u,u}(x;f_{m(i)})| - 2A_{m(i)}^{1/4} - C_p^{1/p}$$

outside a set of measure 2^{-i+1}. Combining the above with (14) implies that for each $(x, -u)$ outside a set of measure $|H'_{m(i)}| + 2^{-i+1}$, there exists an n and a positive constant C such that

$$|I_{n,u}(x;f)| \geqq CA_{m(i)}^{1/2} .$$

From (19) this inequality is true for almost every $(x, -u)$ with sufficiently large i and appropriate n, and the theorem follows.

4. That Theorem 2 holds for Jackson polynomials is relatively easy to prove. We have

$$J_{n,u}(x;f) =$$
$$\frac{1}{(n+1)^2}\sum_{i=0}^{n} f(u+t_i^{(n)})\left\{\frac{\sin 2^{-1}(n+1)(x-u-t_i^{(n)})}{\sin 2^{-1}(x-u-t_i^{(n)})}\right\}^2,\ t_i^{(n)} = \frac{2\pi i}{n+1} .$$

If $f(x) \geqq f_m(x) \geqq 0$,

$$(21)\qquad J_{n,u}(x;f) \geqq \frac{f_m(u+t_k^{(n)})}{(n+1)^2}\left\{\frac{\sin 2^{-1}(n+1)(x-u)}{\sin 2^{-1}(x-u-t_k^{(n)})}\right\}^2 .$$

Thus all of the previous proof devoted to showing that there was not undue interference with one dominant term is now unnecessary. The rest of the proof is very much like the previous one. With some adjustments in the function, we gain additional information.

THEOREM 3. *Given ψ as before, there exists f such that $\psi(|f|)$ is integrable over $(0, 2\pi)$ and such that the sequence $J_{n,u}(x;f)$ diverges for almost every point of the square $0 \leqq x \leqq 2\pi, 0 \leqq u \leqq 2\pi$. Furthermore for any $p \geqq 1$ and $\varepsilon, 0 < \varepsilon < 1$, there is a function f of class L^p such that for almost every point (x, u)*

$$\overline{\lim_{n}} \frac{|J_{n,u}(x;f)|}{n^{1-\epsilon}} > 0 .$$

Let α, β, and A_m be positive reals to be specified at a later point. Let f_m be a periodic step function of the following form. When x belongs to one of the intervals

$$\left|x - \frac{2\pi j}{m}\right| \leqq \frac{2\pi A_m^\alpha}{m^2}, \qquad j = 0, 1, \cdots, m-1$$

let $f_m(x) = A_m$; when x belongs to one of the complementary intervals of $(0, 2\pi)$, let $f_m(x) = 0$. Let E_m be the set E of Theorem 1, part (i), corresponding to m and $\gamma_m = A_m^{-\alpha}$, expanded to $(0, 2\pi)$ of the u axis. Let $E_{m,j}$ be the translation of E_m by $-2\pi j/m$; and let I_j be the set of the x axis such that for some j satisfying $0 \leqq j \leqq m-1$,

$$\frac{2\pi}{mA_m^\beta} \leqq \left|x - \frac{2\pi j}{m}\right| \leqq \frac{\pi}{m} .$$

Given $-u$ in $E_{m,j}$ and x in I_j, there exists an n, $mA_m^{-\alpha} \leqq n+1 \leqq m$, and a k such that

$$\left|u - \frac{2\pi j}{m} + t_k^{(n)}\right| \leqq \frac{2\pi A_m^\alpha}{m^2} .$$

For proper choice of A_m, we have as before

$$\frac{\pi}{mA_m^\beta} \leqq |x - u - t_k^{(n)}| \leqq \frac{3\pi}{2m}$$

so that from (21) $J_{n,u}(x;f)$ exceeds $A_m^{1-2\beta}/10$. Since $||f_m||_p^p = 4\pi A_m^{p+\alpha}/m$, we need only have $A_m^{p+\alpha} = o(m)$ to write $f(x) = \sum_{i=1}^\infty f_{m(i)}(x)$ with the $m(i)$ spread out sufficiently. If α and β are small, the result follows.

Since the sequence of Jackson polynomials corresponding to a continuous function converges uniformly to that function [6; p. 47], it is essentially only for the class of bounded functions that the question of the behaviour of the Jackson polynomials on the square $0 \leqq x \leqq 2\pi$, $0 \leqq u \leqq 2\pi$ is unresolved. However this is no longer true for the ordinary polynomials $I_{n,u}(x;f)$ which may act in a quite irregular way (cf. [2], [5]); and the behaviour of $I_{n,u}(x;f)$ for f continuous still presents a problem of considerable interest.

REFERENCES

1. P. Erdös, *On divergence properties of Lagrange interpolating parabolas,* Ann. of Math. **42** (1941), 309-315.
2. R. P. Gosselin, *On the convergence behaviour of trigonometric interpolating polynomials,* Pacific. J. Math. **5**, Supp. 2 (1955), 915-922.
3. W. J. Leveque, *Topics in number theory,* vol. 1 Reading (1956).
4. J. Marcinkiewicz and A. Zygmund, *Mean values of trigonometric polynomials,* Fund. Math. **28** (1937), 131-166.
5. ———, *Sur la divergence des polynomes d'interpolation,* Acta Univ. Szeged **8** (1936-37), 131-135.
6. A. Zygmund, *Trigonometric interpolation,* (mimeographed notes) University of Chicago, 1950.

UNIVERSITY OF CONNECTICUTT

GENERATING SETS OF ELEMENTS IN COMPACT GROUPS

GILBERT HELMBERG

1. Preliminaries. It is well known that compact topological groups have many properties similar to those of finite groups, which are of course special cases of compact topological groups under the discrete topology. The program of this paper is to characterize sets of elements in a compact topological group which generate a given subgroup and, conversely, to determine properties of the subgroup generated by a given set of elements by an investigation of the properties of this set. Tools for our investigation are the convolution algebra of continuous complex-valued functions on the group and the system of irreducible representations of the group. We shall also formulate the results using those concepts. Our results are straightforward generalizations of known theorems on generating sets of elements in finite groups[1].

From now on G will denote a compact topological group which, as a topological space, is T_1. It follows that G is Hausdorff and, therefore, also normal. Let e denote the identity of G. A subset H of G will be called a subgroup of G if it is an abstract subgroup of G and closed, unless the contrary is specifically stated. Let μ denote the normalized Haar measure on G: $\mu(G) = 1$.

A subgroup H with positive measure $\mu(H) > 0$ is necessarily both open and closed, as are all (left) cosets of H. Thus a compact group G with such a subgroup is disconnected and the quotient-spaces G/H (with respect to left cosets of H) is finite and discrete in the quotient topology. Then $1/\mu(H)$ is the index of H in G. The quotient space of G with respect to left cosets of a subgroup of measure 0 contains infinitely many elements and is again compact, Hausdorff and normal.

Let C denote the field of complex numbers and $C(G)$ the set of all complex-valued continuous functions on G. Defining scalar multiplication and addition in $C(G)$ pointwise as usual, $C(G)$ becomes a Banach-space under the uniform norm: $\|f\| = \sup_{x\in G}\{|f(x)|\}$ $(f \in C(G))$. Defining multiplication in $C(G)$ by convolution,

$$(f * g)(x) = \int_G f(xy^{-1})g(y)dy \ ,$$

$C(G)$ becomes a Banach algebra. Left and right translations of $f \in C(G)$ by $s \in G$ are defined by ${}_sf(x) = f(sx)$ and $f_s(x) = f(xs)$ respectively. Both ${}_sf$ and f_s are functions in $C(G)$ and every $f \in C(G)$ is both left

[1] See [2].

Received April 28, 1958. Presented at the 65th Annual Meeting of the American Mathematical Society in Cincinnati, Ohio. January 28-30, 1958.

and right uniformly continuous.

DEFINITION 1. The subgroup H of G is said to be generated by a set $M \subset G$ if it is the smallest subgroup of G containing M.

The subgroup generated by M will be denoted by $H(M)$. It is evidently the closure of the set of all finite products of positive and negative powers of elements in M. From a theorem of Numakura[2] about compact semigroups it follows that $H(M)$ is already the closure of the set of all finite products of positive powers of elements of M.

2. Subsets of G and corresponding ideals in $C(G)$. With every non-void subset M of G we shall associate the set $F(M)$ of all functions $f \in C(G)$ invariant under right translation by every element $s \in M$.

$$F(M) = \{f : f \in C(G), f_s = f \text{ for all } s \in M\} .$$

Obviously $F(M)$ is non-void, since it contains the constant functions. It is clearly a linear subspace of $C(G)$, and it contains with every $f \in F(M)$ the function $a * f$ if $a \in C(G)$ since

$$\begin{aligned}(a * f)_s(x) = (a * f)(xs) &= \int_G a(xsy^{-1})f(y)dy \\ &= \int_G a(xy^{-1})f(ys)dy = (a * f)(x) .\end{aligned}$$

$F(M)$ is therefore a left ideal in $C(G)$.

It is clear that $M_1 \subset M_2$ implies $F(M_1) \supset F(M_2)$. If $\overline{M}$ is the closure of M in G we have therefore $F(M) \supset F(\overline{M})$.

LEMMA 1. $F(M) = F(\overline{M})$.

Proof. We have to show $F(M) \subset F(\overline{M})$. Assume that there is $f \in F(M)$ such that $f \notin F(\overline{M})$. Then there is $\overline{m} \in \overline{M}$ such that $f_{\overline{m}} \neq f$ and

$$(1) \qquad ||f_{\overline{m}} - f|| > a \text{ for some } a > 0 .$$

Because of the uniform continuity of f, we can choose a neighborhood V of e such that

$$|f(x) - f(y)| < \frac{a}{2} \text{ if } x^{-1}y \in V .$$

The set $\overline{m}V$ is a neighborhood of $\overline{m}$ and contains a point $m \in M$. Then

[2] See **[6]** p. 102.

$$|f(x\overline{m}) - f(xm)| < \frac{a}{2} \text{ for all } x \in G$$

since $(x\overline{m})^{-1}xm = \overline{m}^{-1}m \in V$. Since $f(x\overline{m}) = f_{\overline{m}}(x)$ and $f(xm) = f(x)$ it follows that $||f_{\overline{m}} - f|| < a/2$ which contradicts our assumption (1). Hence $f_{\overline{m}} = f$ and $f \in F(\overline{M})$ for all $f \in f(M)$ and the Lemma follows.

Now let $f \in F(M)$ and $a \in M, b \in M$. Clearly $f_e = f$. Since $f(xa) = f(x)$ for all $x \in G$, we also have $f(xa^{-1}a) = f(xa^{-1})$ for all $x \in G$ or $f_{a^{-1}} = f$. Moreover $f_{ab}(x) = f_b(xa) = f(xa) = f(x)$ for all $x \in G$. If we denote by $H'(M)$ the abstract (not necessarily closed) subgroup of G generated by M then evidently $F(M) \subset F(H'(M))$. On the other hand, $M \subset H'(M)$ implies $F(M) \supset F(H'(M))$ and therefore $F(M) = F(H'(M))$. Now $H(M)$ is the closure of $H'(M)$ in G, and by Lemma 1 we obtain

LEMMA 2. $F(M) = F(H(M))$.

This result allows us to infer some further properties of the functions of $F(M)$. To simplify the notation, we shall in the rest of this paragraph write H instead of $H(M)$. Let $\{g_rH: r \in R\}$ be the decomposition of G into distinct left cosets of H and G/H be the corresponding quotient space. For $f \in F(H)$ and arbitrary $h \in H$, we have $f(g_rh) = f(g_r)$, so that f is constant on every coset g_rH. Conversely every continuous function on G constant on every left coset of H has clearly the property $f_h = f$ for all $h \in H$ and belongs to $F(H)$. Hence $F(M)$ is the set of all continuous functions on G that are constant on left cosets of the subgroup generated by M.

Let us denote by $C(G/H)$ the set of all continuous complex-valued functions on G/H. If we associate with every $f \in F(H)$ the function f' on G/H defined by $f'(g_rH) = f(g_r)$ then $f' \in C(G/H)$ and the mapping $f \to f'$ is a linear one-to-one mapping of $F(H)$ as a linear space onto the linear space $C(G/H)$.[3]

To identify the dimension of $C(G/H)$ as a linear space we have to distinguish two cases.

(a) $\mu(H) > 0$. G/H is finite and discrete. The $i = 1/\mu(H)$ characteristic functions of the points of G/H form a basis in $C(G/H)$. Therefore $F(H)$ is finite-dimensional and closed in the uniform norm in $C(G)$.

(b) $\mu(H) = 0$. G/H is a normal Hausdorff space with infinitely many points. Therefore $C(G/H)$ and $F(H)$ are infinite-dimensional. Let $\overline{F}(H)$ be the closure of $F(H)$ in $C(G)$ and $\bar{f} \in \overline{F}(H)$. Assume $\bar{f}_h \neq \bar{f}$ for some $h \in H$, or

$$(2) \qquad ||\bar{f}_h - \bar{f}|| > a \text{ for some } a > 0 .$$

[3] See [5] p. 110, 111.

There is $f \in F(H)$ such that $||\bar{f} - f|| < a/2$ or

$$|\bar{f}(xh) - f(xh)| < \frac{a}{2} \quad \text{for all } x \in G$$

$$|\bar{f}_h(x) - f(x)| < \frac{a}{2} \quad \text{for all } x \in G$$

$$||\bar{f}_h - f|| < \frac{a}{2}\,.$$

But then

$$||\bar{f}_h - \bar{f}|| \leqq ||\bar{f}_h - f|| + ||\bar{f} - f|| < a$$

which contradicts (2). Therefore $\bar{f}_h = \bar{f}$ for all $h \in H$ and $\overline{F}(H) \subset F(H)$ which shows that $F(H)$ is again closed in $C(G)$.

The results of our discussion are summed up in

THEOREM 1. *$F(M)$ is a closed left ideal in $C(G)$ consisting exactly of all continuous functions on G which are constant on each left coset of the subgroup $H(M)$. As linear subspace of $C(G)$, $F(M)$ is $1/\mu(H(M))$-dimensional if $\mu(H(M)) > 0$ and infinite-dimensional if $\mu(H(M)) = 0$.*

Analogous statements hold for the set of all continuous functions on G that are invariant under left-translation by every element $m \in M$.

3. Subgroups of G and corresponding ideals in $C(G)$. Let the subset M of G be a subgroup H. We can reverse the correspondence between H and $F(H)$ by observing that H is completely characterized by $F(H)$ as the set of all elements of G which right translate every $f \in F(H)$ into itself. In order to see this we have only to show that for every $m \notin H$ there is $f \in F(H)$ such that $f_m \neq f$. Since $m^{-1} \notin H$ we have $H \neq m^{-1}H$. By the complete regularity of G/H, there is $f' \in C(G/H)$ such that $f'(H) = 1$ and $f'(m^{-1}H) = 0$. Defining $f \in F(H)$ by the relation $f(x) = f'(xH)$ for all $x \in G$, we have $f(m^{-1}) = 0$ and $f_m(m^{-1}) = f(e) = 1$. Hence $f_m \neq f$.

It follows that for two arbitrary subgroups H_1 and H_2 of G $F(H_1) \supset F(H_2)$ implies $H_1 \subset H_2$. The converse is obviously true. We conclude:

LEMMA 3. *If H_1 and H_2 are subgroups of G_1 then $H_1 \subset H_2$ if and only if $F(H_1) \supset F(H_2)$.*

Taking $\{e\}$ and G as subgroups of G we have in particular $F(e) = C(G)$ and $F(G) = \{\alpha 1\}$ i.e., the (left) ideal consisting of all constant functions.

Let now N be a normal subgroup of G, $n \in N$ and $f \in F(N)$. For every $x \in G$ we have ${}_nf(x) = f(nx) = f(xn_1) = f_{n_1}(x) = f(x)$ where $n_1 \in N$.

Therefore every element of $F(N)$ is both left and right invariant under translation by elements of N. For an arbitrary $a \in C(G)$ we then have:

$$\begin{aligned}(f * a)_n(x) = (f * a)(xn) &= \int_G f(xny^{-1})a(y)dy = \int_G f(n_1xy^{-1})a(y)dy \\ &= (f * a)(x) \qquad \text{for all } x \in G\end{aligned}$$

$F(N)$ is then a right ideal and therefore a two sided ideal in $C(G)$.

Suppose now that H is non-normal. Then $gH \neq Hg$ for some $g \in G$. We can assume that there is $h \in H$ such that $hg \notin gH$. (Otherwise there would be $h_1 \in H$ such that $gh_1 \notin Hg$ or $h_1g^{-1} \notin g^{-1}H$, and we could take h_1 and g^{-1} in place of h and g.) Then $hgH \cap gH = 0$. We shall exhibit functions $f \in F(H)$ and $a \in C(G)$ such that $f * a \notin F(H)$. It will follow that $F(H)$ is not a two-sided ideal in $C(G)$. Again we distinguish two cases.

(a) $\mu(H) > 0$. The sets gH and Hg^{-1} are both open and closed. Let f be the characteristic function of gH and a be the characteristic function of Hg^{-1}. Then $f \in F(H)$ and $a \in C(G)$.

Let us now consider $f_1(y) = f(hy^{-1})a(y)$ as a function of y. Plainly f_1 is continuous. If $y \in Hg^{-1}$ then $hy^{-1} \in hgH$ and $f(hy^{-1}) = 0$, since $hgH \cap gH = 0$. Therefore $f_1(y) = 0$ for $y \in Hg^{-1}$. However, for $y \notin Hg^{-1}$, $a(y) = 0$ and again $f_1(y) = 0$. We see that

$$(f * a)(h) = \int_G f(hy^{-1})a(y)dy = 0 \,. \tag{3}$$

On the other hand, using the function $f_2(y) = f(y^{-1})a(y)$, we see that $f_2 \in C(G), f_2 \geqq 0$ and $f_2(g^{-1}) = f(g)a(g^{-1}) = 1$. Since the Haar integral is strictly positive on $C(G)$ we conclude that

$$(f * a)(e) = \int_G f(y^{-1})a(y)dy > 0 \,. \tag{4}$$

Comparison of (3) and (4) shows that $f * a$ is not constant on H. Therefore it cannot belong to $F(H)$.

(b) $\mu(H) = 0$. Since G/H is Hausdorff and normal, there are disjoint open neighborhoods U_1 and U_2 of gH and hgH respectively. In view of the complete regularity of G/H, we can find $f' \in C(G/H)$ such that $f' \geqq 0, f'(gH) = 1$, and f' vanishes on the (closed) complement of U_1 in G/H, which contains in particular the open neighborhood U_2 of hgH.

Defining $f(x) = f'(xH)$, we obtain a non-negative function $f \in F(H)$ assuming the value 1 on gH and vanishing on an open set U (the pre-image of U_2 under the mapping $x \to xH$) containing hg. We now choose a symmetric open neighborhood V of e such that $hgV \subset U$ and a non-negative function $a \in C(G)$ assuming the value 1 at g^{-1} and vanishing outside the open set Vg^{-1}. This choice again is possible by the complete regularity of G.

We again consider the continuous function $f_1(y) = f(hy^{-1})a(y)$. For $y \in Vg^{-1}$ we have $hy^{-1} \in hgV \subset U$ so that $f(hy^{-1}) = 0$ and $f_1(y) = 0$. On the other hand, $y \notin Vg^{-1}$ implies $a(y) = 0$ and $f_1(y) = 0$. So

$$(f * a)(h) = \int_G f(hy^{-1})a(y)dy = 0 \ . \tag{3'}$$

Considering $f_2(y) = f(y^{-1})a(y)$, we see that $f_2 \geqq 0, f_2 \in C(G)$ and $f_2(g^{-1}) = f(g)a(g^{-1}) = 1 > 0$. Therefore

$$(f * a)(e) = \int_G f(y^{-1})a(y)dy > 0 \ . \tag{4'}$$

Comparing (3') and (4'), we see again that $f * a$ is not constant on H and does not belong to $F(H)$.

As a result we obtain

LEMMA 4. *A subgroup H of G is normal if and only if $F(H)$ is a two sided ideal in $C(G)$.*

The correspondence between $F(M)$ and $H(M)$ for arbitrary subsets $M \subset G$ leads yet to another useful result.

LEMMA 5. *Let M_1 and M_2 be any subsets of G. Then $M_2 \subset H(M_1)$ if and only if $F(M_1 \cup M_2) = F(M_1)$.*

Proof. Assume first $M_2 \subset H(M_1)$. Then $H(M_1 \cup M_2) = H(M_1)$ and by Lemma 2, we have

$$F(M_1 \cup M_2) = F(H(M_1 \cup M_2)) = F(H(M_1)) = F(M_1) \ .$$

Let us now assume that $F(M_1 \cup M_2) = F(M_1)$. It is clear that $F(M_2) \supset F(M_1 \cup M_2)$. Using Lemma 2, we get $F(H(M_2)) \supset F(H(M_1))$ and by Lemma 3 $M_2 \subset H(M_2) \subset H(M_1)$.

Lemma 5 states in particular that an element $m \in G$ can be approximated by finite products of positive powers of elements in M if and only if the set of all function of $C(G)$ which are invariant under right translation by all elements of M is not reduced by joining m to M.

Taking $M_2 = G$, we obtain as a necessary and sufficient condition for the set M_1 to generate G that $F(M_1)$ be the set of all constant functions on G.

Taking for M_1 a subset of a given subgroup $H = M_2$, Lemma 5 states that M_1 generates H if and only if $F(M_1) = F(H)$.

4. Irreducible representations of G. We now list some definitions and facts concerning representations which we shall have to use in the following.[4]

Let $\{R^{(\lambda)} : \lambda \in \Lambda\}$ be a complete system of inequivalent irreducible unitary continuous representations of G of degrees r_λ respectively. Let $R^{(\lambda)}(s)$ be the matrix associated with the element s in $R^{(\lambda)}$ for a given basis in the corresponding vector space and $R^{(0)}$ the identity representation. Denoting by $u_{ik}^{(\lambda)} \in C(G)$ the coefficient in the ith row and kth column in $R^{(\lambda)}$, we have $u_{ik}^{(\lambda)}(s^{-1}) = \overline{u_{ki}^{(\lambda)}(s)}$ and

$$\begin{aligned} &\sum_{k=1}^{r_\lambda} u_{ik}^{(\lambda)}(s)\overline{u_{ik}^{(\lambda)}(s)} = \delta_{ij} \\ (5)\qquad &\int_G u_{ij}^{(\lambda)}(x)\overline{u_{pq}^{(\lambda')}(x)}dx = \delta_{\lambda\lambda'}\delta_{ip}\delta_{jq} \cdot \frac{1}{r_\lambda} \\ &u_{ij}^{(\lambda)} * u_{pq}^{(\lambda')} = \delta_{\lambda\lambda'}\delta_{jp} \cdot \frac{1}{r_\lambda} u_{iq}^{(\lambda)} , \end{aligned}$$

since the $R^{(\lambda)}$ are unitary.

The functions $u_{ij}^{(\lambda)}$ are linearly independent and form a basis for the linear space $R(G)$ of all complex linear combinations

$$(6)\qquad l = \sum_{\lambda=\lambda_1}^{\lambda_n} \sum_{i,k=1}^{r_\lambda} \alpha_{ik}^{(\lambda)} u_{ik}^{(\lambda)} , \qquad \alpha_{ik}^{(\lambda)} \in C .$$

(5) shows that $R(G)$ is a subalgebra of $C(G)$. The Peter-Weyl theorem says that $R(G)$ is dense in $C(G)$ under the uniform norm. More specifically[5], every $f \in C(G)$ can be uniformly approximated by functions of the form

$$(7)\qquad l = \sum_{\lambda=\lambda_1}^{\lambda_n} \alpha_\lambda \sum_{i=1}^{r_\lambda} (u_{ii}^{(\lambda)} * f)$$

which belong to $R(G)$ as shown below.

Using the notation $(a, b) = \int_G a(x)\overline{b(x)}dx$ for $a \in C(G), b \in C(G)$ we have, as can be verified easily,

$$(8)\qquad u_{ik}^{(\lambda)} * f = \sum_{j=1}^{r_\lambda} (f, u_{kj}^{(\lambda)}) u_{ij}^{(\lambda)} \in R(G)$$

$$(9)\qquad f * u_{ik}^{(\lambda)} = \sum_{j=1}^{r_\lambda} (f, u_{ji}^{(\lambda)}) u_{jk}^{(\lambda)} \in R(G) .$$

From (5) and (8) we can conclude that for fixed λ and i the functions $u_{ik}^{(\lambda)}$ $(k = 1, 2, \cdots, r_\lambda)$ form a basis for a minimal right ideal $R_i^{(\lambda)}$ of $R(G)$

[4] See **[5]** §§ 39, 40.

[5] See **[5]** Theorem 39D. As pointed out be Prof. Edwin Hewitt in a lecture, one can choose the approximate identity in the center of $C(G)$ by taking $u(x) = \int_G v(y^{-1}xy)dy$ and having $v \in C(G)$ $(v \geqq 0)$ vanish outside a sufficiently small neighborhood of e.

and $C(G)$. Analogously it follows from (5) and (9) that for fixed λ and k, the functions $u_{ik}^{(\lambda)}$ $(i = 1, 2, \cdots, r_\lambda)$ form a basis for a minimal left ideal $L_k^{(\lambda)}$ of $R(G)$ and $C(G)$. Finally it follows from (5), (8) and (9) that for fixed λ the functions $u_{ik}^{(\lambda)}$ $(i, k = 1, 2, \cdots, r_\lambda)$ form a basis for a minimal two sided ideal $T^{(\lambda)}$ in $R(G)$ and $C(G)$. Each of these ideals is closed because of its finite dimensionality.

Taking $l \in R(G)$ as in (6) we have

$$u_{ii}^{(\lambda)} * l = \frac{1}{r_\lambda} \sum_{k=1}^{r_\lambda} \alpha_{ik}^{(\lambda)} u_{ik}^{(\lambda)} \in R_i^{(\lambda)}$$

(10)
$$l * u_{kk}^{(\lambda)} = \frac{1}{r_\lambda} \sum_{i=1}^{r_\lambda} \alpha_{ik}^{(\lambda)} u_{ik}^{(\lambda)} \in L_k^{(\lambda)}$$

$$\left[\sum_{i=1}^{r_\lambda} u_{ii}^{(\lambda)}\right] * l = \frac{1}{r_\lambda} \sum_{i,k=1}^{r_\lambda} \alpha_{ik}^{(\lambda)} u_{ik}^{(\lambda)} \in T^{(\lambda)}$$

and

(11)
$$l = \sum_{\lambda=\lambda_1}^{\lambda_r} r_\lambda \sum_{i=1}^{r_\lambda} (u_{ii}^{(\lambda)} * l) = \sum_{\lambda=\lambda_1}^{\lambda_r} r_\lambda \sum_{k=1}^{r_\lambda} (l * u_{kk}^{(\lambda)})$$
$$= \sum_{\lambda=\lambda_1}^{\lambda_r} r_\lambda \left[\left(\sum_{i=1}^{r_\lambda} u_{ii}^{(\lambda)}\right) * l\right].$$

We see that $R(G)$ is the direct sum of the minimal two sided ideals $T^{(\lambda)}$ which in turn are direct sums of minimal right ideals $R_i^{(\lambda)}$ and, in the same way, of minimal left ideals $L_k^{(\lambda)}$.

(12)
$$R(G) = \sum_{\lambda \in \Lambda} \oplus T^{(\lambda)}$$
$$T^{(\lambda)} = \sum_{i=1}^{r_\lambda} \oplus R_i^{(\lambda)} = \sum_{k=1}^{r_\lambda} \oplus L_k^{(\lambda)}.$$

$R(G)$ is itself a two sided ideal in $C(G)$ but is not closed unless it coincides with $C(G)$. (This occurs if and only if G is finite).

The numbers $(f, u_{ik}^{(\lambda)})$ appearing in (8) and (9) can be regarded as the Fourier coefficients of the function $f \in C(G)$. For non-zero f there exist only a countable number of non-zero Fourier coefficients (and at least one).

Every element $a = \sum_{k=1}^{r_\lambda} \alpha_k u_{ik}^{(\lambda)} \in R_i^{(\lambda)}$ can be written in vector notation as a scalar product $\mathfrak{u}_i^{(\lambda)}\mathfrak{a}$ where $\mathfrak{u}_i^{(\lambda)}$ stands for the basis vector $(u_{i1}^{(\lambda)}, u_{i2}^{(\lambda)}, \cdots, u_{ir_\lambda}^{(\lambda)})$ and $\mathfrak{a}$ for the coefficient vector $(\alpha_1, \alpha_2, \cdots, \alpha_{r_\lambda})$, written as column vector. By the definition of $u_{ik}^{(\lambda)}$ we obtain under right translation by any $s \in G$

(13)
$$[u_{ik}^{(\lambda)}]_s(x) = u_{ik}^{(\lambda)}(xs) = \sum_{j=1}^{r_\lambda} u_{ij}^{(\lambda)}(x) u_{jk}^{(\lambda)}(s) \quad \text{or}$$
$$\mathfrak{u}_{i_s}^{(\lambda)} = \mathfrak{u}_i^{(\lambda)} \cdot R^{(\lambda)}(s).$$

Right translation by s evidently induces a linear transformation in $R_i^{(\lambda)}$ whose matrix with respect to $\mathfrak{u}_i^{(\lambda)}$ as a basis is just $R^{(\lambda)}(s)$, and $R_i^{(\lambda)}$ is invariant under right translation. For any function $a \in R_i^{(\lambda)}$, the effect of the translation is given by the formulas

$$a_s = \mathfrak{u}_{i_s}^{(\lambda)}\mathfrak{a} = \mathfrak{u}_i^{(\lambda)}R^{(\lambda)}(s)\mathfrak{a} = \mathfrak{u}_i^{(\lambda)}\mathfrak{a}_s \tag{14}$$
$$\mathfrak{a}_s = R^{(\lambda)}(s)\mathfrak{a}$$

where $\mathfrak{a}_s$ is the coefficient vector of a_s.

5. Generating sets in G and irreducible representations of G. We investigate for a given subgroup H of G the intersection of $F(H)$ with the ideals of $R(G)$, introduced above. If $f \in F(H)$ and $f \neq 0$, then $(f, u_{ik}^{(\lambda)}) \neq 0$ for some λ, i, k. The function

$$u_{ii}^{(\lambda)} * f = \sum_{j=1}^{r_\lambda} (f, u_{ij}^{(\lambda)})u_{ij}^{(\lambda)}$$

is different from zero, lies in $F(H)$, and by (8) also in $R(G)$ (in fact in $R_i^{(\lambda)}$), therefore in $F'(H) = F(H) \cap R(G)$ (also in $F(H) \cap R_i^{(\lambda)}$). $F'(H)$ is again a left ideal in $C(G)$ since $R(G)$ is a two sided ideal in $C(G)$ and contains all functions of the form $u_{ii}^{(\lambda)} * f$ for a given $f \in F(H)$. From (7), we obtain as an immediate consequence

LEMMA 6. *$F'(H) = F(H) \cap R(G)$ is dense in $F(H)$.*

Let now $f' \in F'(H)$. By (11), f' can be written as a linear combination of functions of the form $u_{ii}^{(\lambda)} * f$ which are by (10) contained in $F(H) \cap R_i^{(\lambda)}$. On the other hand, every linear combination of functions in $F(H) \cap R_i^{(\lambda)}$ is again a function of $F'(H)$. On account of the direct decomposition of $R(G)$ with respect to the minimal right ideals $R_i^{(\lambda)}$, we see that $F'(H)$ is, as a linear space, the direct sum of the linear spaces $F(H) \cap R_i^{(\lambda)}$,

$$F'(H) = \sum_{\lambda \in \Lambda} \oplus \sum_{i=1}^{r_\lambda} \oplus [F(H) \cap R_i^{(\lambda)}] \tag{15}$$

some of which may consist only of zero.

Let now $F(H) \cap R_i^{(\lambda)}$ be non-zero (we have already seen that there must be at least one non-zero $F(H) \cap R_i^{(\lambda)}$) and let $f_i^{(\lambda)} \in F(H) \cap R_i^{(\lambda)}$. We can write $f_i^{(\lambda)}$ as a scalar product of the basis vector $\mathfrak{u}_i^{(\lambda)}$ of $R_i^{(\lambda)}$ and the coefficient vector $\mathfrak{f}^{(\lambda)}$

$$f_i^{(\lambda)} = \mathfrak{u}_i^{(\lambda)}\mathfrak{f}^{(\lambda)} . \tag{16}$$

The function $f_i^{(\lambda)}$ is invariant under right translation by all elements $h \in H$. In view of (14) this means that

$$\mathfrak{f}^{(\lambda)} = R^{(\lambda)}(h)\mathfrak{f}^{(\lambda)} \text{ for all } h \in H \tag{17}$$

i.e., $\mathfrak{f}^{(\lambda)}$ is an eigenvector of $R^{(\lambda)}(h)$ with eigenvalue 1 for all $h \in H$. Conversely, for fixed λ, every eigenvector with eigenvalue 1 common to all $R^{(\lambda)}(h)$ $(h \in H)$ determines by (16) a function $f_i^{(\lambda)} \in F(H) \cap R_i^{(\lambda)}$.

Since for a given i, λ linear independence of functions $f_i^{(\lambda)}, g_i^{(\lambda)}$ is equivalent to linear independence of the corresponding coefficient vectors $\mathfrak{f}^{(\lambda)}, \mathfrak{g}^{(\lambda)}$ we see that the dimension of $F(H) \cap R_i^{(\lambda)}$ as a linear space is precisely the number of linearly independent eigenvectors $\mathfrak{f}^{(\lambda)}$ common to all $R^{(\lambda)}(h)$ $(h \in H)$ with eigenvalue 1.

DEFINITION 2. For any non-void subset M of G and for any fixed λ, let $d^{(\lambda)}(M)$ denote the maximal number of linearly independent eigenvectors common with eigenvalue 1 to $R^{(\lambda)}(m)$ for all $m \in M$.

The inequalities $0 \leqq d^{(\lambda)}(M) \leqq r_\lambda$ necessarily hold. In the present case, we see that $d^{(\lambda)}(H)$ is the dimension of $F(H) \cap R_i^{(\lambda)}$ for all $i = 1, 2, \cdots, r_\lambda$ since it obviously does not depend on i. Taking $d^{(\lambda)}(H)$ linearly independent functions of $F(H) \cap R_i^{(\lambda)}$ and $r - d^{(\lambda)}(H)$ properly chosen $u_{ik}^{(\lambda)}$ $(i, \lambda$ fixed) as a basis for $R_i^{(\lambda)}$ amounts to transforming the representation $R^{(\lambda)}$ to an equivalent one, $R'^{(\lambda)} = S^{-1}R^{(\lambda)}S$, in which $R'^{(\lambda)}$ restricted to the elements of H, becomes reducible as representation of H and is found to contain the identity-representation of H exactly $d^{(\lambda)}(H)$ times. Thus $d^{(\lambda)}(H)$ can also be defined an the multiplicity with which the identity representation of H is contained in $R^{(\lambda)}$, restricted to the elements of H and considered as a representation of H.

$F(H) \cap R_i^{(\lambda)}$ has the dimension $d^{(\lambda)}(H)$ for given λ, as we have seen. The subspace $F(H) \cap T^{(\lambda)}$ is the direct sum of all $F(H) \cap R_i^{(\lambda)}$ $(i=1, 2, \cdots, r_\lambda)$ and has therefore dimension $r_\lambda d^{(\lambda)}(H)$. If there is only a finite number of non-zero $d^{(\lambda)}(H)$, then there are only a finite number of non-zero $F(H) \cap R_i^{(\lambda)}$ and $F(H) \cap T^{(\lambda)}$. By (15), we see that $F'(H)$ is a linear space of dimension $\sum_{\lambda \in \Lambda} r_\lambda d^{(\lambda)}(H)$ which is finite-dimensional, and therefore $F'(H)$ is closed. But then $F'(H) = F(H)$ by Lemma 6, and $F(H)$ is of finite dimension $\sum_{\lambda \in \Lambda} r_\lambda d^{(\lambda)}(H)$. If infinitely many $d^{(\lambda)}(H)$ are non-zero then $F'(H)$ is an infinite dimensional linear space and the same must be true of $F(H)$. Combining this result with the results of Theorem 1, we obtain:

THEOREM 2. *If $d^{(\lambda)}(H)$ is the multiplicity with which the identity representation of a subgroup H of G is contained in $R^{(\lambda)}$, restricted to the elements of H and considered as a representation of H, then*

$$\sum_{\lambda \in \Lambda} r_\lambda d^{(\lambda)}(H) = \frac{1}{\mu(H)} \qquad \textit{if } \mu(H) > 0 \,.$$

If $\mu(H) = 0$ then the series $\sum_{\lambda \in \Lambda} r_\lambda d^{(\lambda)}(H)$ diverges.

The sum $\sum_{\lambda \in \Lambda} r_\lambda d^{(\lambda)}(H)$ can therefore be considered as giving the "index" of H in G. A subgroup H has measure 0 if and only if $d^{(\lambda)}(H) > 0$ for infinitely many $\lambda \in \Lambda$.

Let N be a normal subgroup of G and $d^{(\lambda)}(N) > 0$ for a certain λ. Then $F(N) \cap R_i^{(\lambda)}$ contains a non-zero function $f = \sum_{k=1}^{r_\lambda} \alpha_k u_{ik}$. Assume that $\alpha_l \neq 0$. The set $F(N)$ is a two sided ideal by Lemma 4, and so is $F'(N) = F(N) \cap R(G)$. Therefore $F'(N)$ contains together with f the function

$$f * u_{lj}^{(\lambda)} = \frac{\alpha_l}{r_\lambda} u_{ij}^{(\lambda)} \text{ for arbitrary } j\,, \qquad 1 \leqq j \leqq r_\lambda\,.$$

This means that $R_i^{(\lambda)} \subset F'(N)$ and $d^{(\lambda)}(N) = r_\lambda$. On the other hand, supposing that for a given subgroup H $d^{(\lambda)}(H)$ assumes only the values 0 or r_λ for all $\lambda \in \Lambda$, we see that $F(H) \cap R_i^{(\lambda)}$ is either zero or $R_i^{(\lambda)}$. Then $F(H) \cap T^{(\lambda)}$ is either zero or $T^{(\lambda)}$ and $F'(H)$ is the direct sum of two sided ideals and itself a two sided ideal in $C(G)$. Its closure $F(H)$ must also be two sided and by Lemma 4, H is normal.

THEOREM 3. *A subgroup H of G is normal if and only if $d^{(\lambda)}(H)$ assumes only the values 0 or r_λ for all $\lambda \in \Lambda$.*[6]

Trivial illustrations of this fact are given by the entire group G ($d^{(0)}(G) = 1$ and $d^{(\lambda)}(G) = 0$ for $\lambda \neq 0$) and by the group consisting of $\{e\}$ only ($d^{(\lambda)}(e) = r_\lambda$ for all $\lambda \in \Lambda$).

We proceed now to characterize the generating properties of an arbitrary subset M of G by means of the representations $R^{(\lambda)}$. Since $M \subset H(M)$, there are by the definition of $d^{(\lambda)}(H(M))$ at least $d^{(\lambda)}(H(M))$ linearly independent functions in $R_i^{(\lambda)}$ that are invariant under right translation by all elements of M and $d^{(\lambda)}(M) \geqq d^{(\lambda)}(H(M))$. Conversely, as seen in the proof of Lemma 2, any such function of $R_i^{(\lambda)}$ is also invariant under right translation by all elements of $H(M)$ and $d^{(\lambda)}(M) \leqq d^{(\lambda)}(H(M))$. Together with the previous result, we now have

LEMMA 7. *If M is an arbitrary subset of G, then $d^{(\lambda)}(M) = d^{(\lambda)}(H(M))$ for all $\lambda \in \Lambda$.*

The main result which we can now prove is

THEOREM 4. *If M_1 and M_2 are arbitrary subsets of G, then $M_2 \subset H(M_1)$ if and only if $d^{(\lambda)}(M_1 \cup M_2) = d^{(\lambda)}(M_1)$ for all $\lambda \in \Lambda$.*

Proof. Let $M_2 \subset H(M_1)$. Then $H(M_1) = H(M_1 \cup M_2)$ and $d^{(\lambda)}(M_1) = d^{(\lambda)}(M_1 \cup M_2)$ for all $\lambda \in \Lambda$ by Lemma 7. On the other hand, the

[6] See also [**4**] and [**1**] Theorem 1.

equality $d^{(\lambda)}(M_1) = d^{(\lambda)}(M_1 \cup M_2)$ for all $\lambda \in \Lambda$ implies by Lemma 7 that

$$F(H(M_1)) \cap R_i^{(\lambda)} = F(H(M_1 \cup M_2)) \cap R_i^{(\lambda)} \text{ for all } \lambda \in \Lambda ,$$

$$F(H(M_1)) \cap T^{(\lambda)} = F(H(M_1 \cup M_2)) \cap T^{(\lambda)} \text{ for all } \lambda \in \Lambda ,$$

$$F'(H(M_1)) = F(H(M_1)) \cap R(G) = F(H(M_1 \cup M_2)) \cap R(G) \\ = F'(H(M_1 \cup M_2)) \text{ (by (15))} ,$$

$$F(H(M_1)) = F(H(M_1 \cup M_2)) \text{ (by Lemma 6) and} \\ M_2 \subset H(M_1) \text{ (by Lemmas 2 and 5)} .$$

A number of corollaries are easily obtained. Putting $M_2 = G$ in Theorem 4 and noting that $d^{(\lambda)}(G)$ is positive only for $\lambda = 0$ we obtain

COROLLARY 4.1. *The subset M of G generates G if and only if $d^{(\lambda)}(M) = 0$ for all $\lambda \neq 0$.*

Taking as M_2 a subgroup H and as M_1 a subset M of H, we get

COROLLARY 4.2. *The subset M of the subgroup H of G generates H if and only if $d^{(\lambda)}(M) = d^{(\lambda)}(H)$ for all $\lambda \in \Lambda$.*

Finally, combining the results of Theorem 2, 3 and Lemma 7, we obtain

COROLLARY 4.3. *The subset M of G generates a normal subgroup of G if and only if $d^{(\lambda)}(M)$ assumes only the values 0 and r_λ for all $\lambda \in \Lambda$. If $d^{(\lambda)}(M) > 0$ for only a finite number of $\lambda \in \Lambda$, then M generates a subgroup of measure $1/\sum_{\lambda \in \Lambda} r_\lambda d^{(\lambda)}(M)$; otherwise M generates a subgroup of measure 0.*

6. Finite generating sets in G. The preceding results are in particular valid for finite groups. In that case we are only concerned with the investigation of generating properties of finite sets of elements. Schreier and Ulam[7] have shown that a connected compact metric group G is generated by almost every pair of elements. Since the component of the identity in any compact group G is a connected normal subgroup of finite index in G, it is clear that there are always a finite number of generators for a compact metric group.

For the case of a finite set M, there is a simple way to determine $d^{(\lambda)}(M)$ and to state the conditions of the last theorems and corollaries, based on the following lemma.

LEMMA 8. *Let $B^{(\lambda)}(m_1, \cdots, m_s)$ be the rectangular matrix with r_λ rows and sr_λ columns obtained by joining horizontally the s matrices $R^{(\lambda)}(m_k) - R^{(\lambda)}(e)$ $(k = 1, 2, \cdots, s)$. Let $b^{(\lambda)}(m_1, \cdots, m_s)$ be the rank of*

[7] See [7] and [8].

$B^{(\lambda)}(m_1, \cdots, m_s)$. *Then* $d^{(\lambda)}(\{m_k : k = 1, \cdots, s\}) = r_\lambda - b^{(\lambda)}(m_1, \cdots, m_s)$.

Since this Lemma has been stated by the author in [1] without proof it may be suitable to set down a proof here.

Proof. Let $B^{*(\lambda)}(m_1, \cdots, m_s)$ be the conjugate transpose of $B^{(\lambda)}(m_1, \cdots, m_s)$. Its rank is the same as that of $B^{(\lambda)}(m_1, \cdots, m_s)$. Since $R^{(\lambda)}$ is unitary, $B^{*(\lambda)}(m_1, \cdots, m_s)$ could have been obtained by placing the s matrices $R^{(\lambda)}(m_k^{-1}) - R^{(\lambda)}(e)$ $(k = 1, \cdots, s)$ below each other. Since $d^{(\lambda)}(\{m_k : k = 1, \cdots, s\}) = d^{(\lambda)}(\{m_k^{-1} : k = 1, \cdots, s\})$ we have to show that the rank of $B^{*(\lambda)}(m_1, \cdots, m_s)$ is equal to $r_\lambda - d^{(\lambda)}(\{m_k^{-1} : k = 1, \cdots, s\})$. In order to simplify the notation, we shall from now on omit the index λ and the indication of the group elements when possible.

If we denote by A_s the $rs \times rs$ matrix obtained by placing the non-singular $r \times r$ matrix, A, s times along the principal diagonal in a $rs \times rs$ zero-matrix, then A_s is non-singular and $A_s^{-1}B^*A$ has again rank b. If $\mathfrak{u} = (u_1, \cdots, u_r)$ is the basis of the r-dimensional linear space corresponding to the matrix-representation R, then the transition to a new basis $\mathfrak{u}'$ in which the d first basis vectors are invariant under the transformations corresponding to $m_1^{-1}, \cdots, m_s^{-1}$ is given by the formula $\mathfrak{u}P = \mathfrak{u}'$ where P is a non-singular $r \times r$ matrix. In the new basis these transformations are given by the matrices $P^{-1}R(m_k^{-1})P$. The d first columns in each of these have as their only non-zero elements 1's in the main diagonal. In each of the matrices $P^{-1}(R(m_k^{-1}) - R(e))P$ those columns are therefore zero columns. Placing those s matrices one below the other we obtain, as one can readily see, exactly the matrix $P_s^{-1}B*P$. The rank of this matrix can therefore not exceed $r - d$ and we have $b \leqq r - d$.

Assume that $b < r - d$. Then one of the columns $C'_{d+1}, \cdots, C'_r$ in $P_s^{-1}B^*P$, say C'_c, would be a linear combination of the other ones. By a permutation of the vectors u_{d+1} and u_c in $\mathfrak{u}'$ given by $\mathfrak{u}'Q = \mathfrak{u}''$, where Q is the matrix of the corresponding permutation, we obtain as above a matrix $Q_s^{-1}P_s^{-1}B^*PQ$ with rank b in which the d first columns vanish and the $(d+1)$-th column appears as a linear combination of the remaining ones $C''_{d+1} = \sum_{j=d+2}^{r} \alpha_j C''_j$.

Define R as the matrix obtained from $R(e)$ by replacing in the $(d+1)$-th column the zeros below the principal diagonal by $-\alpha_{d+2}, \cdots, -\alpha_r$ in that order. Passing to a new basis by the formula $\mathfrak{u}''R = \mathfrak{u}'''$, we obtain as above the matrix $R_s^{-1}Q_s^{-1}P_s^{-1}B^*PQR$ in which, as one can see easily, the first $d+1$ columns vanish. But then the first $d+1$ columns in $(PQR)^{-1}R(m_k^{-1})PQR$ have as their only non-zero elements 1's in the main diagonal. This in turn means that the first $d+1$ basis vectors in $\mathfrak{u}'''$ are invariant under the transformations corresponding to all elements $m_k^{-1}(k = 1, \cdots, s)$. But this contradicts our assumption that there are not more than d linearly independent vectors of that property. So $b = r - d$, and the lemma is proved.

Lemma 8 allows us to determine $d^{(\lambda)}(\{m_1, \cdots, m_s\})$ if the matrices $R^{(\lambda)}(m_k)(k = 1, \cdots, s)$ are given. Applying Lemma 8 to a single element m, we see that $d^{(\lambda)}(\{m\})$ is exactly the multiplicity of the eigenvalue 1 in $R^{(\lambda)}(m)$. If $R^{(\lambda)}(m)$ does not have 1 as an eigenvalue, then $b^{(\lambda)}(m) = r_\lambda$.

Using Lemma 8, we can also reformulate the preceding results. *e.g.* Corollary 4.1 takes the following form: the elements $m_1, \cdots, m_s$ generate G if and only if $b^{(\lambda)}(m_1, \cdots, m_s) = r_\lambda$ for $\lambda \neq 0$. This condition is in particular satisfied if for every $\lambda \neq 0$ there is at least one $m^{(\lambda)}$ among the $m_1 \cdots m_s$ for which $R^{(\lambda)}(m^{(\lambda)})$ does not have 1 as an eigenvalue. In this case, however, we can even say that the products of the form $m_1^{a_1} \cdots m_s^{a_s}$ $(0 \leqq a_k : k = 1, \cdots, s)$ are dense in G and, arranged in a certain order, form a sequence which is equidistributed in G.[8] Similarly we can see that the hypothesis of Corollary 4.2 is satisfied if for every $\lambda \in \Lambda$ there is at least one $m^{(\lambda)}$ such that the multiplicity of the eigenvalue 1 in $R^{(\lambda)}(m^{(\lambda)})$ is exactly $d^{(\lambda)}(H)$, i.e., the multiplicity with which $R^{(\lambda)}$ restricted to H contains the identity-representation of H. Again in this case we can make the stronger statement that the products of the form $m_1^{a_1} \cdots m_s^{a_s}$ $(0 \leqq a_k : k = 1, \cdots, s)$ are dense in H and, arranged in a certain order, form a sequence which is equidistributed in H.

REFERENCES

1. A. H. Clifford, *Representations induced in an invariant subgronp*, Annals of Math., **38** No. 3 (1937), 533-550.

2. G. Helmberg, *Strukturbeziehungen zwischen endlicher Gruppe, Gruppenring und irreduziblen Darstellungen*, Monatshefte fuer Math. Wien, **58** (1954), 241-257.

3. ———, *A theorem on equidistribution in compact groups*, Pacific J. Math., **8** (1958), 227-241.

4. A. Kulakoff, Rec. Math. Soc. Moscow **36** (1929), 129-134.

5. L. Loomis, *An Introduction to Abstract Harmonic Analysis*, van Nostrand, New York, 1953.

6. Katsumi Numakura, *On bicompact semigroups*, Math. Journal of Okayama Univ. **1** (1952), 99-108.

7. T. Schreier and L. Ulam, *Sur le nombre des generateurs d'un groupe topologique compact et connexe*, Fund. Math. **24** (1935), 302-304.

8. ———, *Eine Bemerkung ueber Erzeugende in kompakten Gruppen*, Fund. Math. **25** (1925), 198-199.

TULANE UNIVERSITY

[8] See [3].

THE H_p-PROBLEM AND THE STRUCTURE OF H_p-GROUPS

D. R. HUGHES AND J. G. THOMPSON

1. Introduction. Let G be a group, p a prime, and $H_p(G)$ the subgroup of G generated by the elements of G which do not have order p. In a research problem in the Bulletin of the American Mathematical Society, one of the authors posed the following problem: is it always true that $H_p(G) = 1$, $H_p(G) = G$, or $[G : H_p(G)] = p$? This problem is easily settled in the affirmative for $p = 2$, and a similar answer was recently given for $p = 3$ ([5]). In this paper (Section 2) we give an affirmative answer for the case that G is finite and not a p-group. Furthermore (Section 3) we are able to give a rather precise description of the structure of G in the most interesting case, when $[G : H_p(G)] = p$. This structure theorem depends heavily on the deep results of Hall and Higman ([4]) and Thompson ([6]) on finite groups. If $H\ (\neq 1)$ is a finite group and there exists a group G such that $H_p(G)$ is isomorphic to H, where $H_p(G) \neq G$, then we call H an H_p-group; it is seen that H_p-groups are natural generalizations of "Frobenius groups." By a Frobenius group we mean a finite group G possessing an automorphism σ of prime order p such that $x^\sigma = x$ if and only if $x = 1$. It is easy to show that this implies

$$x^{1+\sigma+\cdots+\sigma^{p-1}} = x(x^\sigma)\cdots(x^{\sigma^{p-1}}) = 1\ ,$$

for all x in G. This last equation characterizes H_p-groups,[1] and as a generalization of Thompson's result ([6]) that Frobenius groups are nilpotent, we show that H_p-groups are solvable, among other things.

Throughout the paper, if B is a group, A a subgroup of B, then $N_B(A)$ and $C_B(A)$ mean, respectively, the normalizer and centralizer of A in B. By $Z(A)$ we mean the center of A.

2. The H_p-problem. Let G be a group, and let $H = H_p(G)$. Suppose

(1) G is finite,

(2) G is not a p-group,

(3) the index of H in G is greater than p,

(4) G is a group of minimal order satisfying (1), (2), (3). Note that every element of G which is not in H has order p.

Let q be a prime dividing $[G:1]$, $q \neq p$, and let Q be a Sylow q-group of G; then Q is also a Sylow q-group of H. Let $N = N_G(Q)$; then

Received January 16, 1959. The first author was supported in part by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under contract No. AF 18 (600)-1383.

[1] Unless the group is a p-group; see Theorem 2.

by the Frattini argument (see [1], p. 117, for instance), $G = NH$. Thus $[G:1] = [NH:1] = [N:1][H:1]/[N \cap H:1]$.

First, let us suppose $N \neq G$. Then clearly $H_p(N) \subseteq H_p(G)$, so $H_p(N) \subseteq H \cap N$. Since $Q \subseteq H_p(N)$, it follows that $H_p(N) \neq 1$, so $[N:H_p(N)] \leq p$, and hence $[N:N \cap H] \leq p$. So $p^2 = [G:H] = [G:1]/[H:1] = [N:1]/[N \cap H:1] = [N:N \cap H] \leq p$. This is impossible, so we must have $N = G$, and thus Q is normal in G.

Now let Q_1 ($\neq 1$) be any subgroup of Q, normal in G, and consider G/Q_1. Clearly $H_p(G/Q_1) = 1$ or $H_p(G/Q_1)$ has index p in G/Q_1, unless G/Q_1 is a p-group. Indeed, it is obvious that $H_p(G/Q_1) \subseteq H/Q_1$. But $[G/Q_1 : H/Q_1] = [G:H] = p^2$, so $[G/Q_1 : H_p(G/Q_1)] \geq [G/Q : H/Q_1] = p_2$ implies $H_p(G/Q_1) = 1$. So G/Q_1 is a p-group.

LEMMA 1. *If $[G:H] = p^2$, then Q is an elementary abelian q-group, none of whose proper subgroups ($\neq 1$) is normal in G, Q is normal in G, and $G = PQ$, where P is a Sylow p-group of G.*

Proof. We have shown that Q is normal. If Q_1 above is taken to be the Frattini subgroup of Q, then Q_1 is normal in G, since it is characteristic in Q. Since $Q_1 \neq Q$, G/Q_1 cannot be a p-group, so we must have $Q_1 = 1$. Thus Q is elementary abelian. Since G/Q is a p-group, it is clear that $G = PQ$, and the rest of the lemma follows similarly.

In what follows, P is a Sylow p-group of G and $P_0 \subseteq P$ is a Sylow p-group of H; clearly $[P:P_0] = p^2$ and P_0 is normal in P, since $P_0 = P \cap H$.

If x ($\neq 1$) is in Q, while a is in G, not in H, and if $ax = xa$, then ax has order pq. But ax is not in H, since a is not in H, and thus ax has order p; hence $ax \neq xa$. If $P_0 = 1$, then P, of order p^2, is an automorphism group of $H = Q$ such that no non-identity element of P fixes any non-identity element of Q. But by ([2], pp. 334–335) this means that P is cyclic, whereas P is clearly elementary abelian in this case (for all its elements have order p). So $P_0 \neq 1$.

Since P_0 is normal in P, $P_0 \cap Z(P) \neq 1$ (see [3], p. 35, for instance). Let z be an element of $P_0 \cap Z(P)$, chosen to have order p, and let Z_0 be the subgroup (of order p) generated by z; note that z and Z_0 are contained in H. Let $K = Z_0Q$, and observe that $[K:1] = p[Q:1]$. Let a be an element of G, not in H, and $G_1 = \{a, K\}$ = the group generated by a and K. Then $Q \subseteq H_p(G_1) \subseteq H \cap G_1 \neq G_1$, so $[G_1 : H_p(G_1)] = p$, by induction. Hence $Z_0 \subseteq K \subseteq H_p(G_1)$, so there must be an element y in K of order pq. Then y^p is in Q and y^q is in $x^{-1}Z_0x$, for some x in K, since Z_0 is a Sylow p-group of K. By adjusting our choice of P, we can assume that y^q is in Z_0; let $u = y^p$, $v = y^q$. Then $u \neq 1$, $v \neq 1$, u is in Q, v is in Z_0, and $uv = vu$. So if $Q_1 = \{u\}$, we have $Z_0 \subseteq C_G(Q_1)$. But then $x^{-1}Z_0x \subseteq C_G(x^{-1}Q_1x)$, and if x is in P, this implies $Z_0 \subseteq C_G(x^{-1}Q_1x)$, for all x in P. But, from Lemma 1, the subgroup generated by all

$x^{-1}Q_1x$, as x ranges over P, must be Q, and so $Z_0 \subseteq C_G(Q)$. Since Z_0 is in the center of P, it follows that Z_0 is normal in G, so we consider G/Z_0. One easily sees that $H_p(G/Z_0) \subseteq H/Z_0$, and $H_p(G/Z_0)$ equals neither 1 nor G/Z_0. Hence $p^2 = [G:H] = [G/Z_0 : H/Z_0] \leq [G/Z_0 : H_p(G/Z_0)] = p$, which is a contradiction. So:

THEOREM 1. *If $H_p(G) \neq 1$ or G, and if G is finite and not a p-group, then $[G : H_p(G)] = p$.*

If G is a p-group, or is infinite, the situation seems more inaccessible; as remarked earlier, Theorem 1 still holds if $p = 2$ or 3, no matter what G is. But the proof for $p = 3$ (see [5]) utilizes the Burnside theorem (for $p = 3$) and this strongly suggests that the infinite case at least is considerably harder.

3. Structure of H_p-groups. Let us suppose that G is a finite group, and that $H = H_p(G)$ has index p in G. Then we say that H is an H_p-group.

THEOREM 2. *If H is not a p-group, then H is an H_p-group if and only if H has an automorphism σ of order p such that*

$$x^{1+\sigma+\cdots+\sigma^{p-1}} = 1 ,$$

for all x in H.

Proof. If $H = H_p(G)$, let a be in G, a not in H, and define $x^\sigma = a^{-1}xa$, for x in H. Since $(ax)^p = 1$, while $(ax)^p = a^p(x)(x^\sigma) \cdots (x^{\sigma^{p-1}})$, the equation of the theorem follows immediately.

Conversely, if σ exists satisfying the hypotheses of the theorem, then let G be the holomorph of H by the automorphism group $\{\sigma\}$. It is easy to see that $H_p(G) \subseteq H$. Since $H_p(G) \neq 1$ (for H is not a p-group), it follows that $[G : H_p(G)] = p$, from Theorem 1, so $H_p(G) = H$.

Note that if $x^\sigma = x$, then the equation of Theorem 2 implies $x^p = 1$. So if p does not divide the order of the H_p-group H, then H is even a Frobenius group, and so is nilpotent ([6]).

THEOREM 3. *If H is an H_p-group, then $H = PK$, where P is a Sylow p-group of H, K is normal in H and is nilpotent, and $P \cap K = 1$. In particular, H is solvable.*

Proof. We can assume that $P \neq 1$, and that H is not a p-group. Inductively, suppose the theorem is true for all H_p-groups whose order is less than the order of H, and (using Theorem 2) let γ be an automorphism of H, of order p, such that

$$x^{1+\gamma+\cdots+\gamma^{p-1}} = 1 , \quad \text{all } x \text{ in } H.$$

If A is a γ-invariant subgroup of H, then A is an H_p-group or is a p-group, while if B is a γ-invariant normal subgroup of H, then H/B is an H_p-group or is a p-group.

Now let B be any γ-invariant subgroup of P, B normal in P, $B \neq 1$; let $N = N_H(B)$. If $N = H$, then H/B is an H_p-group, so $H/B = (P/B)(K_1/B)$, where K_1/B is normal in H/B and is nilpotent. So K_1 is normal in H and since K_1/B is γ-invariant in H/B, so is K_1 γ-invariant in H. So K_1 is an H_p-group. If $K_1 \neq H$, then $K_1 = BK$, where K is normal in K_1 and is nilpotent, and $K \cap B = 1$. But then K is characteristic in K_1, hence is normal in H; every Sylow q-group of H, $q \neq p$, is in K. So K is characteristic in H and clearly $H = PK$, $P \cap K = 1$.

If $K_1 = H$ for every such B, then $B = P$ is the only γ-invariant normal subgroup of P, other than 1. Hence in particular P is elementary abelian. Then H/P is an H_p-group, and even a Frobenius group, so is nilpotent. Furthermore (since H is then solvable), $H = PK$, where K is isomorphic to H/P. Let $K = Q_1Q_2 \cdots Q_t$, where Q_i is a Sylow q_i-group of K (and of H) for distinct primes $q_1, q_2, \cdots, q_t$.

Now let G be the holomorph of H with the group $\{\gamma\}$. Then, by the Frattini argument, $N_G(Q_i) \cap H \neq N_G(Q_i)$, so by an appropriate choice of γ_i in G, γ_i not in H, we can assume that Q_i is γ_i-invariant. Thus PQ_i is γ_i-invariant and so it is an H_p-group (it is straightforward to check that any element of G, not in H, can play the role of γ).[2]

If $t > 1$, then PQ_i has order smaller than H, so Q_i is normal in PQ_i. Thus both P and K are contained in $N_H(Q_i)$, so Q_i is normal in H, hence K, which is the direct product of the Q_i, is normal in H, so we are done.

If $t = 1$, let $Q = Q_1$, and as above, choose γ in G, not in H, so that Q is γ-invariant. If $Q_0 \neq 1$ is a γ-invariant normal subgroup of Q, then PQ_0 is an H_p-group, smaller than $H = PQ$ if $Q_0 \neq Q$; thus P normalizes Q_0, so Q_0 is normal in H. Then by considering H/Q_0, we find that Q/Q_0 is normal, so Q is normal in H, and again we are done. Thus we can assume that Q is elementary abelian with only trivial γ-invariant normal subgroups.

Now we consider the holomorph G again. The maximal normal p-group of G is P, since $\{\gamma\}$ (as part of G) is not normalized modulo P by Q. Then G/P is a solvable (and in particular, p-solvable) group of automorphisms of the elementary abelian group P, and G/P has no normal p-group ($\neq 1$). Furthermore, this representation of G/P as a linear transformation group on P is faithful, since $C_H(P) \cap Q = 1$ (otherwise $C_H(P) \cap Q$ would be a non-trivial γ-invariant normal subgroup of Q). Thus we can utilize Theorem B of Hall and Higman ([4]); since Q is abelian, Theorem B asserts that γ, as a linear transformation of P, has the minimal

[2] In these references to the holomorph G, we are not making a distinction between an element as an automorphism of H and as an element of G; the automorphism is actually identified with an element of G which induces the prescribed automorphism in H.

polynomial $(x-1)^p$. But in fact, γ has a minimal polynomial which divides $1 + x + \cdots + x^{p-1}$, since

$$b^{1+\gamma+\cdots+\gamma^{p-1}} = 1 ,$$

for all b in P. Thus we have a contradiction, and so Q is normal in H, and we are done.

Now we must consider the case that if B $(\neq 1)$ is any γ-invariant subgroup of P, normal in P, then $N = N_H(B)$ is never equal to H. Hence N, being γ-invariant, is an H_p-group or is a p-group, so $N = P_1K_1$, where P_1 is a Sylow p-group of N, K_1 is normal in N and is nilpotent, and $K_1 \cap P_1 = 1$. Since B is normal in N, K_1 is contained in $C_N(B)$, and thus contained in $C_H(B)$, so $N_H(B)/C_H(B)$ is a p-group (i.e., is isomorphic to P_1/P_0, for some subgroup P_0 of P_1). But then, since this holds for all such B, Thompson's theorem ([**6**]) asserts that P has a normal complement K in H; i.e., $H = PK$, where $P \cap K = 1$ and K is normal in H. Since K consists exactly of the elements of H whose order is prime to p, K is characteristic. Thus K is an H_p-group (even a Frobenius group) and is nilpotent, so we are done.

BIBLIOGRAPHY

1. Reinhold Baer, *Classes of finite groups and their properties*, Illinois J. Math., **1** (1957), 115-187.

2. W. Burnside, *Theory of Groups of Finite Order*, New York, 1955.

3. P. Hall, *A contribution to the theory of groups of prime-power order*, Proc. London Math. Soc., **36** (1933), 29–95.

4. P. Hall and G. Higman, *On the p-length of p-soluble groups and reduction theorem for Burnside's problem*, Proc. London Math. Soc., **6** (1956), 1-42.

5. E. G. Straus and G. Szekeres, *On a problem of D. R. Hughes*, Proc. Amer. Math. Soc., **9** (1958), 157–158.

6. J. G. Thompson, *Finite groups with fixed-point-free automorphisms of prime order*, Math. Z. (to appear).

UNIVERSITY OF CHICAGO
DE PAUL UNIVERSITY

PROJECTIVE INJECTIVE MODULES

J. P. JANS

1. Introduction. In this paper we prove several theorems about rings having a generous supply of projective injective modules. This is a curious class of rings. For instance, every module over a semisimple ring with minimum condition is both projective and injective, while over the integers only the zero module has this property. On the other hand, for some non-semisimple rings, Quasi Frobenius rings [5], every projective module is injective. For others no non-trivial projective module is injective (for example, a primary algebra over a field with radical square zero and having vector space dimension greater than two).

We begin our study in § 2 by considering primitive rings. We give (Theorem 2.1) a necessary and sufficient condition for a primitive ring to have a faithful projective injective irreducible module. By means of this condition we prove a structure theorem (Corollary 2.3) for rings having both a left and a right injective projective irreducible module with the same anihilator.

In § 3 we generalize both halves of a theorem originally proved by Thrall for finite dimensional algebras [10, Theorem 5]. This theorem states that a necessary and sufficient condition for the minimal injective [3] of the ring to be projective is that the ring have a faithful injective module which is a direct summand of every faithful module. We prove this theorem in one direction for semi-primary rings and, in the other direction, for rings with the ascending chain condition. It should be noted that we have rephrased the theorem to eliminate the duality given by the field. We find that this can be replaced by the dual concepts, projective and injective.

Throughout the paper we shall only consider rings with identity 1 and modules over such rings on which 1 acts like identity. " Minimum condition " means minimum condition on left ideas [1].

The author wishes to express his appreciation to John Walter for many stimulating conversations which contributed to the formulation of this paper. We also wish to thank Alex Rosenberg for suggesting clear concise proofs of Theorems 2.1 and 3.2.

2. Projective injective irreducibles. We shall begin by considering primitive rings. Recall that a (right) primitive ring R has a faithful irreducible right module M [7, p. 4]. The module M is always the homomorphic image of R, and if M is projective then M is induced by

Received Jaunary 28, 1959. The author was supported by the National Science Foundation.

a minimal right ideal of R. That is, R is a primitive ring with minimal right ideals. Conversely, if R is a primitive ring with minimal right ideals then the faithful irreducible module is induced by an idempotent generated (= direct summand) right ideal of R. Thus, the faithful irreducible is projective.

In the following we shall study primitive rings with minimal right ideals and we shall establish a necessary and sufficient condition for the faithful irreducible module of such a ring R to be injective. We are greatly aided in this study by the rich structure theory for these rings; see for example Jacobson's book [7, Chapter IV].

Using the notation and the structure theorem from [7, p. 75], we have $S = F(M, N) \subset R \subset L(M, N)$ where M, N are dual spaces over a division ring D and M (N) is a right (left) irreducible faithful projective R-module. S is the socle of R.

THEOREM 2.1. *The module M is R injective if and only if* $M = N^* = \mathrm{Hom}_D(N, D)$.

Proof. If $M = \mathrm{Hom}_D(N, D)$ then by Prop. 1.4 p. 107 of [2], M is R injective.

For the converse, assume that M is R injective. In this case, it is enough to show that for every maximal right ideal J of S there is a nonzero element a of S such that $aJ = 0$. Then the left ideal Sa contains an idempotent $e \neq 0$ such that $eJ = 0$ and J is a modular [7] (called regular in [9]) right ideal. But Rosenberg has shown [9, p. 131] that if every maximal right ideal of S is modular then $M = N^* = \mathrm{Hom}_D(N, D)$.

Identify M with a minimal right ideal of S. Since J is maximal in S we can consider the R exact sequence of modules

$$O \longrightarrow J \longrightarrow S \xrightarrow{\theta} M \longrightarrow O\ .$$

Since M is R injective by [2, Th 3.1, p. 8] the homomorphism θ has the form $\theta(s) =$ as for some $a \neq 0$ in the right ideal M of S. But since $\mathrm{Ker}\ \theta = J, aJ = 0$. Theorem 2.1 then follows from the remarks above.

One should note that the corresponding theorem with right and left interchanged is proved analogously, hence we have the following

COROLLARY 2.2. *If R is a primitive ring then R is a simple ring with minimum condition if and only if R has both a left and a right faithful irreducible projective injective module.*

Proof. If R is a simple ring with minimum condition then it has faithful irreducible left and right modules [7, p. 39] and every module over such a ring is both projective and injective [2, p. 11].

To show the converse, we appeal to the theorem. Using the notation of the theorem, $M = N^*$ and $M^* = N$. But we know [7, p. 68], that this can only happen when both have finite dimension over D. In this case R is isomorphic to *all* transformations on M and is a simple ring with minimum condition [7, p. 39].

The theorem and its corollary also have applications to any ring having left and right projective injective irreducibles. It is clear that if a ring R can be written as a ring direct sum $S + K$ where S is a simple ring with minimum condition, then R has both a left and a right projective injective irreducible module, each having anihilator K. It is interesting to note that the converse is also true.

COROLLARY 2.3. *If R has both a left and a right projective injective irreducible, each having anihilator K, then $R = S + K$* (ring direct sum) *where S is a simple ring with minimum condition.*

Proof. Under the above assumptions R/K is both a left and a right primitive ring and the faithful irreducible left and right modules considered as R/K modules are still projective and injective. Thus, by Corollary 2.2, R/K is a simple ring with minimum condition and both as an R module and as an R/K module is the direct sum of a finite number of copies of the left irreducible projective injective module. Thus the sequence of left R modules $0 \to K \to R \to R/K \to 0$ splits and $R = S \oplus K$, left R direct. The proof will be established if we can show that S is really an ideal of R.

Certainly, $KS = (0)$ because S is the direct sum of modules anihilated by K. Let k belong to K and consider the left ideal Sk contained in K. It is clear that $(Sk)^2 = SkSk = (0)$ because k anihilates S on the left. Suppose that Sk is not zero. In this case, Sk is the homomorphic image of the completely reducible module S and is the direct sum of a finite number of injective irreducible modules. But that makes Sk injective and a direct summand of R. However, this contradicts the fact that Sk is square zero, since direct summands of R are idempotent generated. Thus we have established that $Sk = (0)$ and that the decomposition given above is a ring direct sum.

REMARK. There is a one-sided version of Corollary 2.3, in which one assumes only the existence of a projective injective irreducible left module plus the ascending chain condition on left ideals in R modulo its Jacobson radical. The conclusion is the same. However, the conclusion is two sided, so the existence of a projective injective left irreducible and the above mentioned chain condition (or semi-primary, etc.) implies the existence of a projective injective right irreducible.

3. Minimal faithfuls and minimal injectives. Following Thrall's paper [10], we shall say that the ring R has a *minimal faithful left module* M if M is a faithful injective module and if M appears as a direct summand of every faithful module. It is clear that M must be projective, for the ring itself is a faithful projective module. M will always be isomorphic to some left ideal direct summand of R.

If T is any R module, the minimal injective $Q(T)$ of T is the unique "smallest" injective module containing T as a submodule, [3]. Using these two concepts, we can prove a generalization of one half of a theorem of Thrall [10, Theorem 5]. Thrall proved it for finite dimensional algebras over a field.

THEOREM 3.1. *If R is right Noetherian and if R has a minimal faithful left module M then $Q(R)$, the left minimal injective of R, is projective.*

Proof. As noted above M must be isomorphic to a projective injective left ideal which we also denote by M. In R consider the collection of right ideals generated by finite sets of elements of M. Since we have assumed R to be right Noetherian, there is in this collection a maximal right ideal H generate by $x_1, \cdots, x_n$ belonging to M. Since H is maximal with respect to this property, we know that $M \subset H$. For if not, H could be enlarged by adjoining another generator from M.

If x is in R and $xx_i = 0$ for $i = 1, \cdots, n$, then $xH = (0)$ and consequently $xM = (0)$. But M is faithful, so $x = 0$. Now let Q be the direct sum of n copies of M and for x in R define $\theta: R \to Q$ by letting the ith component of $\theta(x)$ be xx_i. This is a left module homomorphism of R into Q and, by the remark above, is a monomorphism. Q is projective and injective since it is the direct sum of a finite number of projective injective modules. The minimal injective of R is a direct summand of Q and is therefore projective.

We should note that if R is both left and right Noetherian and has a minimal faithful left module then the minimal injective of any projective module is projective. This follows from the fact that every free module can be embedded in a projective injective module, a direct sum of copies of M. We need the assumption that R is left Noetherian to insure that the direct sum of left injectives is injective. Compare this to the definition of Quasi Frobenius ring [5]: "Every projective *is* injective".

To prove the other half of Thrall's theorem we consider the class of semi-primary rings. The ring R is said to be semi-primary if it has a nilpotent Jacobson radical N and R/N has minimum condition on left ideals. An important property of semi-primary rings is the fact that every module over such a ring has minimal submodules. For, if M is a module over the semi-primary ring R with radical N then in the sequence $M \supset NM \supset \cdots \supset N^rM = (0)$ of submodules of M there is a point where $N^kM \neq (0)$ but $N^{k+1}M = (0)$. N^kM, a module over R/N, is the direct sum of irreducibles each of which is minimal. Note also that R has only a finite number of nonisomorphic irreducible modules.

THEOREM 3.2. *If R is a semi-primary ring and if the left minimal injective $Q(R)$ of R is projective then R has a minimal faithful module.*

Proof. By the remark above, we know that R itself has minimal left ideals. Let $M_1, \cdots, M_n$ be one each of the non-isomorphic minimal left ideals of R. From [8], we know that the minimal injective $Q(M_i)$ of M_i is indecomposable. In addition each $Q(M_i)$ is projective since it appears as a direct summand of $Q(R)$. But the projective indecomposable modules over a semi-primary ring actually appear as left ideal direct summands of the ring [4, p. 331]. Thus each $Q(M_i)$ is isomorphic to a projective injective indecomposable left ideal L_i of R. Note that for $i \neq j$, L_i is not isomorphic to L_j since each has a unique minimal submodule [8] and these are not isomorphic.

Let M be the direct sum of the modules L_i, we wish to show that M is the minimal faithful module for R. From its definition it is projective and injective. If M_α is a minimal ideal of R, M_α is isomorphic to a minimal submodule of M. Since M is injective that isomorphism has the form $x \to xm$ for some m in M [2, p. 8]. Hence M_α does not anihilate M. If no minimal left ideal of R anihilates M, then no nonzero left ideal anihilates M and M is faithful.

Now let T be an R module such that $M_iT \neq 0$. Then there exists t in T such that $M_it \neq 0$. Consider the homomorphism $\sum(x) = xt$ of L_i into T. This homomorphism restricted to M_i is not zero and since M_i is the unique minimal submodule of L_i, Σ is actually a monomorphism of L_i into T. L_i is injective so $T = L_i \oplus T_1$.

From the preceding argument we conclude that for $i \neq j$ $M_iL_j = 0$ since L_i and L_j are indecomposable and not isomorphic. Now let F be a faithful R module. Since $M_1F \neq 0$, the argument above shows that $F = L_1 \oplus F_1$ where $M_iF_1 \neq 0$ for $i > 1$. Continuing inductively, $F_{i-1} = L_i \oplus F_i$ where $M_jF_i \neq 0$ for all $j > i$. Thus we see $F = M \oplus F_n$ and M appears as a direct summand of every faithful R module. This completes the proof of Theorem 3.2.

REMARK. Since a ring with minimum condition is both semi-primary and Noetherian, both halves of Thrall's theorem hold for these rings.

REFERENCES

1. Artin, Nesbitt and Thrall, *Rings with Minimum Condition*, Ann Arbor (1944).
2. Cartan and Eilenberg, *Homological Algebra*, Princetion (1956).
3. Eckmann and Schopf, *Uber Injective Moduln*, Arch. Math. **4** (1953).
4. S. Eilenberg, *Homological dimension and syzygies*, Ann. of Math. **64** (1956), 328.
5. Eilenberg, Nagao and Nakayama, *On the dimensions of modules and algebras II* Nagoya Math. Journal **9** (1955), 1-16.
6. N. Jacobson, *Lectures in Abstract Algebra I*, Van Nostrand: 1951.
7. ———, *Structure of rings*, Amer. Math. Soc. Coll. Publication XXXVII, 1956.
8. E. Matlis, *On injective modules*, Pacific J. Math. **8** (1958), 511-528.
9. A. Rosenberg, *Subrings of simple rings with minimal ideals*, Trans. Amer. Math. Soc. **73** (1952), 115-138.
10. R. M. Thrall, *Some generalizations of quasi Frobenius algebras*, Trans. Amer. Math. Soc. **64** (1948) 173-183.

UNIVERSITY OF WASHINGTON

COINCIDENCE PROPERTIES OF BIRTH AND DEATH PROCESSES

SAMUEL KARLIN AND JAMES MCGREGOR

A birth and death process (for brevity referred to henceforth as process B) is a stationary Markov process whose state space is the non-negative integers and whose transition probability matrix

$$P_{ij}(t) = \Pr\{x(t) = j \mid x(0) = i\} \tag{1}$$

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

$$P_{ij}(t) = \begin{cases} \lambda_i t + o(t) & \text{if} \quad j = i + 1 \\ \mu_i t + o(t) & \text{if} \quad j = i - 1 \\ 1 - (\lambda_i + \mu_i)t + o(t) & \text{if} \quad j = i \end{cases} \tag{2}$$

where $\lambda_i > 0$ for $i \geq 0$, $\mu_i > 0$ for $i \geq 1$ and $\mu_0 \geq 0$. We further assume that $P_{ij}(t)$ satisfies the foward and backward equation in the usual form. In this paper we restrict attention to the case $\mu_0 = 0$ so that when the particle enters the state zero it remains there a random length of time according to an exponential distribution with parameter λ_0 and then moves into state one etc.

In order to avoid inessential difficulties we assume henceforth that the infinitesimal birth and death rates λ_i and μ_i uniquely determine the process. This is equivalent to the condition $\sum_{n=0}^{\infty} (\pi_n + 1/\lambda_n \pi_n) = \infty$ where

$$\pi_0 = 1 \text{ and } \pi_n = \frac{\lambda_0 \lambda_1 \lambda_2 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \mu_3 \cdots \mu_n} \qquad [2].$$

In the companion paper we show that for all $t > 0$

$$\det(P_{i_\mu, j_\nu}(t)) = P\begin{pmatrix} t\,; & i_1, i_2, \cdots, i_n \\ & j_1, j_2, \cdots, j_n \end{pmatrix} \qquad \begin{matrix} i_1 < i_2 < i_3 < \cdots < i_n \\ j_1 < j_2 < j_3 < \cdots < j_n \end{matrix} \tag{3}$$

has the following interpretation: Start n labled particles at time zero in states $i_1, i_2, \cdots, i_n$ respectively, each governed by the transition law (1) and acting independently. The determinant (3) is equal to the probability that at time t particle 1 is located in state j_1, particle 2 is located in state j_2 etc., without any two of these particles having occupied simultaneously a common state at some earlier time $\tau < t$. We refer to this event as a transition in time t of n particles from initial states

Received December 18, 1958. This work was supported in part by an Office of Naval Research Contract at Stanford University.

occur. This problem is completely solved in § 4. By means of trivial arguments it is shown that coincidence is certain if the original birth and death process is recurrent, while coincidence is not certain if the original process is strongly transient. If the original process is weakly transient coincidence may or may not be certain, and this case presents a much more difficult problem. A criterion is given which expresses the necessary and sufficient condition that coincidence be certain, in terms of the constants of the original birth and death process. Finally in § 3 some interesting examples are considered. A technique for computing the distribution of the time until coincidence is developed, and applied to the telephone trunking model and some linear growth models.

1. Positivity properties of $Q\binom{i_1, i_2, \cdots, i_n}{x_1, x_2, \cdots, x_n}$.

Let M, K and L be functions of two variables satisfying

$$M(\xi, \eta) = \int_a^b K(\xi, \zeta)L(\zeta, \eta)d\sigma(\zeta) \tag{8}$$

where ξ traverses X, ζ ranges through Y and η varies over Z all of which are linearly ordered sets and where $\sigma(\zeta)$ denotes a measure defined in Y. X can denote an interval of the real line or a set of discrete points on the line. In the latter case, the set will usually consist of the integers. The same applies to Y and Z. When Y consists of a discrete space then, of course, the integral sign of (8) is interpreted as a sum. We define the Fredholm determinant

$$M\binom{x_1, x_2, \cdots, x_n}{z_1, z_2, \cdots, z_n} = \begin{vmatrix} M(x_1, z_1), & M(x_1, z_2), & \cdots, & M(x_1, z_n) \\ M(x_2, z_1), & M(x_2, z_2), & \cdots, & M(x_2, z_n) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ M(x_n, z_1), & M(x_n, z_2), & \cdots, & M(x_n, z_n) \end{vmatrix} \tag{9}$$

with $x_1 < x_2 < \cdots < x_n$ and $z_1 < z_2 < \cdots < z_n$ and analogously for K and L.

If the formula (8) is viewed as a continuous version of a matrix product, then the extension of the multiplication rule which evaluates subdeterminants of M in terms those of K and L becomes

$$M\binom{x_1, x_2, \cdots, x_n}{z_1, z_2, \cdots, z_n} = \int \cdots \int_{a \le y_1 < y_2 < \cdots < y_n \le b} K\binom{x_1, x_2, \cdots, x_n}{y_1, y_2, \cdots, y_n} L\binom{y_1, y_2, \cdots, y_n}{z_1, z_2, \cdots, z_n} \cdot d\sigma(y_1)d\sigma(y_2) \cdots d\sigma(y_n) . \tag{10}$$

For the proof of (10) we refer to Pólya and Szëgo I [8 p. 48]

$$(6)\qquad Q\begin{pmatrix} i_1, i_2, \cdots, i_n \\ x_1, x_2, \cdots, x_n \end{pmatrix} = \begin{vmatrix} Q_{i_1}(x_1), Q_{i_1}(x_2), \cdots, Q_{i_1}(x_n) \\ Q_{i_2}(x_1), Q_{i_2}(x_2), \cdots, Q_{i_2}(x_n) \\ \cdots \\ Q_{i_n}(x_1), Q_{i_n}(x_2), \cdots, Q_{i_n}(x_n) \end{vmatrix}$$

where $i_1 < i_2 < \cdots < i_n$ and $x_1 < x_2 < \cdots < x_n$ we obtain by virtue of (5) that

$$(7)\qquad P\left(t\,;\begin{matrix} i_1, i_2, \cdots, i_n \\ j_1, j_2, \cdots, j_n \end{matrix}\right)$$

$$= \pi_{j_1}\pi_{j_2}\cdots\pi_{j_n} \int_{x_1<x_2<\cdots<x_n}\cdots\int e^{-(x_1+x_2+\cdots+x_n)t} Q\begin{pmatrix} i_1, i_2, \cdots, i_n \\ x_1, x_2, \cdots, x_n \end{pmatrix} Q\begin{pmatrix} j_1, j_2, \cdots, j_n \\ x_1, x_2, \cdots, x_n \end{pmatrix} \cdot$$

$$\cdot\, d\psi(x_1)d\psi(x_2)\cdots d\psi(x_n)\,.$$

(See Paragraph A of Section 1.)

The above formula displays in the simplest possible way the dependence of $P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right)$ on the time t, the initial state $(i_1, \cdots, i_n)$ and final state $(j_1, \cdots, j_n)$. For the birth and death process itself formula (5) has proven to be a very powerful tool in analyzing the statistical properties of the process. It may be expected that formula (7) will be of comparable utility in the study of the compound process. However certain technical details stand in the way of such a study. While the general properties of the orthogonal polynomials $\{Q_n(x)\}$ have been intensively investigated by numerous mathematicians, the somewhat more complicated polynomials $\left\{Q\begin{pmatrix} i_1, \cdots, i_n \\ x_{1i} \cdots, x_n \end{pmatrix}\right\}$ appear to be new objects of study. At the present time we possess numerous interesting theorems about these polynomials but our results are still incomplete. In a separate publication we will elaborate on the structure of this determinantal polynomial system. In the present paper we develop only those properties directly relevant to our analysis.

We investigate two types of problems associated with the compound process. The first problem is concerned with the behavior of the ratio

$$\frac{P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right)}{P\left(t\,;\begin{matrix} k_1, \cdots, k_n \\ l_1, \cdots, l_n \end{matrix}\right)}$$

as $t \to \infty$. This requires some knowledge of positivity properties of the polynomials $Q\begin{pmatrix} i_1, \cdots, i_n \\ x_1, \cdots, x_n \end{pmatrix}$. In § 1 these required positivity properties are developed, and in § 2 it is shown that the above ratio converges to a finite positive limit as $t \to \infty$. The second problem is that of determining for which processes coincidence in a finite state is certain to

$i_1, i_2, \cdots, i_n$ to the states $j_1, j_2, \cdots, j_n$ respectively, without coincidence. In particular, for $t > 0$ the expression (3) is always positive. For continuous time discrete state space processes, the converse proposition is also true. Specifically, if (3) is always positive, then $P_{ij}(t)$ is the transition matrix of a birth and death process [6].

In this paper, we investigate certain aspects of the structure of the Markov process describing the transitions of n particles conditioned that no coincidence takes place.

We refer to this process as the compound birth and death process of order n. Frequently, when no ambiguities arise the terms "birth and death" and "order n" will be suppressed. The basis of the subsequent analysis is principally an integral representation for

$$P\left(t\,;\ \begin{matrix} i_1, i_2, \cdots, i_n \\ j_1, j_2, \cdots, j_n \end{matrix}\right)$$

which is derived from a corresponding representation formula for $P_{ij}(t)$.

Let $Q_n(x)$ denote a sequence of polynomials of degree n defined by the recursive relations

$$\begin{aligned} -xQ_n(x) &= -(\lambda_n + \mu_n)Q_n(x) + \lambda_n Q_{n+1}(x) + \mu_n Q_{n-1}(x) \qquad n \geq 0 \\ Q_0(x) &\equiv 1 \qquad\qquad Q_{-1}(x) \equiv 0\,. \end{aligned} \tag{4}$$

These equations may be written in compact form as

$$-xQ = AQ$$

where Q represents the vector $(Q_0(x), Q_1(x), Q_2(x), \cdots,)$ and A is the infinitesimal matrix of the process

$$A = \begin{Vmatrix} -\lambda_0 & \lambda_0 & & & & \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & & & \\ & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & & \\ & & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 & \\ & & & \cdot & \cdot & \cdot \\ & & & & \cdot & \cdot & \cdot \\ & & & & & \cdot & \cdot & \cdot \end{Vmatrix}$$

Let $\psi(x)$ denote the unique measure on $[0, \infty)$ with respect to which $Q_n(x)$ are orthogonal. (The measure ψ is unique because of the assumption $\sum(\pi_k + 1/\lambda_k\pi_k) = \infty$.) Then

$$P_{nm}(t) = \pi_m \int_0^\infty e^{-xt} Q_n(x) Q_m(x) d\psi(x)\,. \tag{5}$$

Introducing the notation

The relevance and utility of this identity will be abundantly clear. We record several relations which are applications of it.

(A) The derivation of (7) from (5) is a special case of (10).

(B) The identity

$$\sum_{j=0}^{n} \pi_j Q_j(x) = \frac{\lambda_n \pi_n [Q_{n+1}(x) - Q_n(x)]}{-x} = H_n(x) \tag{11}$$

can be expressed in the form (8) with $\xi = n, \zeta = j$, and $\eta = x$

$$M(\xi, \eta) = H_n(x)$$
$$K(\xi, \zeta) = \begin{cases} 1 & \zeta \le \xi \\ 0 & \zeta > \xi \end{cases}$$
$$L(\zeta, \eta) = Q_j(x)$$

($Q_j(x) \equiv 0$ for j a negative integer and $d\sigma(\zeta) = \pi_j$ when $\zeta = j$.)

Since

$$K\begin{pmatrix} i_1, i_2, \cdots, i_n \\ l_1, l_2, \cdots, l_n \end{pmatrix} = 0$$

unless $0 \le l_1 \le i_1, i_1 < l_2 \le i_2, \cdots, i_{n-1} < l_n \le i_n$, in which case its value is one, we obtain by applying (10) to (11)

$$\begin{aligned} &H\begin{pmatrix} i_1, i_2, \cdots, i_n \\ x_1, x_2, \cdots, x_n \end{pmatrix} \\ &= \sum_{l_1=0}^{i_1} \sum_{l_2=i_1+1}^{i_2} \cdots \sum_{l_n=i_{n-1}+1}^{i_n} \pi_{l_1}\pi_{l_2} \cdots \pi_{l_n} Q\begin{pmatrix} l_1, l_2, \cdots, l_n \\ x_1, x_2, \cdots, x_n \end{pmatrix} \end{aligned} \tag{12}$$

(C) We shall need to evaluate determinants of the form

$$\begin{vmatrix} Q_{n_0}(0), & Q'_{n_0}(0), & \cdots, & Q^{(k)}_{n_0}(0) \\ Q_{n_1}(0), & Q'_{n_1}(0), & \cdots, & Q^{(k)}_{n_1}(0) \\ \vdots & \vdots & & \vdots \\ Q_{n_k}(0), & Q'_{n_k}(0), & \cdots, & Q^{(k)}_{n_k}(0) \end{vmatrix}, \quad n_0 < n_1 < \cdots < n_k \tag{13}$$

which for convenience of writing we give the name $\alpha(n_0, n_1, \cdots, n_k)$.

We assume tentatively in what follows that Q_n are normalized such that $Q_n(0) = 1$. This can be accomplished with no loss of generality since $Q_n(0)$ are different from zero. The value of the determinant $\alpha(n_0, n_1, \cdots, n_k)$ in the general situation would be altered by the multiplicative factor $1/Q_{n_0}(0)Q_{n_1}(0) \cdots Q_{n_k}(0)$.

A more convenient expression for (13) is obtained as follows: By subtracting the first row from each of the succeeding rows and using the fact that $Q_n(0) = 1$ for all n we have

$$\alpha(n_0, n_1, n_2, \cdots, n_k) = \begin{vmatrix} Q'_{n_1}(0) - Q'_{n_0}(0) \cdots Q^{(k)}_{n_1}(0) - Q^{(k)}_{n_0}(0) \\ Q'_{n_2}(0) - Q'_{n_0}(0) \cdots Q^{(k)}_{n_2}(0) - Q^{(k)}_{n_0}(0) \\ \vdots \qquad\qquad\qquad \vdots \\ Q'_{n_k}(0) - Q'_{n_0}(0) \cdots Q^{(k)}_{n_k}(0) - Q^{(k)}_{n_0}(0) \end{vmatrix} .$$

We next observe that relation (11) provided with successive differentiation yields

$$(14) \quad Q^{(r+1)}_{n+1}(0) - Q^{(r+1)}_{n_0}(0) = -(r+1) \sum_{\nu=n_0+1}^{n} \frac{1}{\lambda_\nu \pi_\nu} \sum_{\mu=0}^{\nu} \pi_\mu Q^{(r)}_\mu(0)$$

$$(n > n_0, r = 0, 1, 2, \cdots,) .$$

In order to apply (10) to (14), we may identify

$$M(\xi, \eta) = Q^{(r+1)}_{n+1}(0) - Q^{(r+1)}_{n_0}(0)$$

$$K(\xi, \zeta) = \begin{cases} 1 & \zeta \leq \xi \\ 0 & \zeta > \xi \end{cases}$$

$$L(\zeta, \eta) = \sum_{\mu=0}^{\nu} \pi_\mu Q^{(r)}_\mu(0)$$

and $d\sigma(\zeta) = 1/\lambda_\nu \pi_\nu$ where ξ, ζ, η each traverse the set of non-negative integers. By virtue of (10) utilizing the representation (14) we obtain

$$(15) \quad \alpha(n_0, n_1, n_2, \cdots, n_k)$$

$$= (-1)^k (k\,!) \sum_{l_1=n_0+1}^{n_1} \sum_{l_2=n_1+1}^{n_2} \cdots \sum_{l_k=n_{k-1}+1}^{n_k} \frac{1}{\lambda_{l_1}\pi_{l_1}} \frac{1}{\lambda_{l_2}\pi_{l_2}} \cdots \frac{1}{\lambda_{l_k}\pi_{l_k}} L\binom{l_1, l_2, \cdots, l_k}{0, 1, \cdots, k-1}$$

where we have employed the specific evaluations of the Fredholm subdeterminants based on $K(\xi, \eta)$.

Another application of (10) shows that

$$(16) \quad L\binom{l_1, l_2, \cdots, l_k}{0, 1, \cdots, k-1}$$

$$= \sum_{\mu_0=0}^{l_1} \sum_{\mu_1=l_1+1}^{l_2} \cdots \sum_{\mu_{k-1}=l_{k-1}+1}^{l_k} \pi_{\mu_0}\pi_{\mu_1} \cdots \pi_{\mu_{k-1}} \alpha(\mu_0, \mu_1, \cdots, \mu_{k-1}) .$$

Putting (15) and (16) together establishes the recursive relation

$$(17) \quad \alpha(n_0, n_1, \cdots, n_k) = (-1)^k k\,! \sum_{l_1=n_0+1}^{n_1} \cdot \sum_{l_2=n_1+1}^{n_2} \cdots \sum_{l_k=n_{k-1}+1}^{n_k} \frac{1}{\lambda_{l_1}\pi_{l_1}} \frac{1}{\lambda_{l_2}\pi_{l_2}}$$

$$\cdots \frac{1}{\lambda_{l_k}\pi_{l_k}} \sum_{\mu_0=0}^{l_1} \cdot \sum_{\mu_1=l_1+1}^{l_2} \cdots \sum_{\mu_{k-1}=l_{k-1}+1}^{l_k} \pi_{\mu_0}\pi_{\mu_1} \cdots \pi_{\mu_{k-1}} \alpha(\mu_0, \mu_1, \mu_2, \cdots, \mu_{k-1}) .$$

Note that the range of summations guarantee that $\mu_0 < \mu_1 < \mu_2 < \cdots < \mu_{k-1}$.

Furthermore, (17) exhibits determinants $\alpha(n_0, n_1, \cdots, n_k)$ of order $k+1$ in terms of corresponding determinants of order k. Consequently, the procedure may be iterated out of which follows whenever $Q_n(0) > 0$ that

$$(18) \qquad (-1)^{k(k+1)/2} \cdot \alpha(n_0, n_1, n_2, \cdots, n_k) > 0$$

for all choices of n_i provided $n_0 < n_1 < n_2 < \cdots < n_k$. It is also routine to calculate the explicit value of $\alpha(n_0, n_1, n_2 < \cdots < n_k)$ by iteration of (17). In particular

$$\alpha(n_0, n_1) = -\sum_{k=n_0+1}^{n_1} \frac{1}{\lambda_k \pi_k} \sum_{l=0}^{k} \pi_l$$

$$\alpha(n_0, n_1, n_2) = \sum_{k=n_0+1}^{n_1} \frac{1}{\lambda_k \pi_k} \sum_{l=n_1+1}^{n2} \frac{1}{\lambda_l \pi_l} \sum_{\mu_0=0}^{k} \pi_{\mu_0} \sum_{\mu_1=k+1}^{l} \pi_{\mu_1} \alpha(\mu_0, \mu_1) .$$

The derivation of the identity (17) was predicated upon the fact that $Q_n(0) = 1$ for all n. If all the $Q_n(0)$ are of negative sign then the sign of (13) is altered by the factor $(-1)^{k+1}$ where $k+1$ is the order of the matrix. Indeed, all we need do is replace $Q_n(x)$ by $Q_n(x)/Q_n(0) = P_n(x)$ and apply the argument to $P_n(x)$. The value of $\alpha(n_0, n_1, \cdots, n_k)$ based on $P_n(x)$ differs only by an obvious multiplying factor from that based on $Q_n(x)$.

(D) Following the same lines of argument as above we shall show

$$(19) \qquad (-1)^{(k-1)k/2} Q\binom{n_1, n_2, \cdots, n_k}{x_1, x_2, \cdots, x_k} > 0$$

provided $x_1 < x_2 < x_3 < \cdots < x_k \leq a$ where a denote the smallest value in the spectrum of ψ, and where $Q_n(0) > 0$ by our normalization condition. The result expressed in (19) may be regarded as a generalization to the compounded polynomial system of the property that $Q_n(x)$ for $x < a$ is of one sign.

Suppose for definiteness that the polynomials $Q_n(x)$ are orthogonal functions with respect to a measure ψ on $[0, \infty)$. The proof is by induction on k. Since $Q_n(x)$ are normalized to be positive at 0, it follows that $Q_n(x) > 0$ for all $x \leq a$ which is the assertion of (19) when $k = 1$. We shall assume that the validity of (19) for kth order determinants has been demonstrated for any system of orthogonal polynomials whose weight function concentrates on the interval $[0, \infty)$, and proceed to show the result is valid for the $k+1$st order determinants. Let $x_1, x_2, \cdots, x_{k+1}$ denote a set of values arranged in increasing order with $x_{k+1} \leq a$. Replacing $Q_n(x)$ by $Q_n(x + x_{k+1})/Q_n(x_{k+1})$ we may, without loss of generality assume $x_{k+1} = 0$ and that $Q_n(x_{k+1}) = 1$ for all n. This alters the original determinants by a positive multiplicative factor, provided we evaluate the changed matrix polynomial system at the points $y_i = x_i - x_{k+1}$. Hence

$$Q\binom{n_1, n_2, \cdots, n_{k+1}}{x_1, x_2, \cdots, x_{k+1}} = \begin{vmatrix} Q_{n_1}(x_1), & Q_{n_1}(x_2) & \cdots & Q_{n_1}(x_k) & 1 \\ Q_{n_2}(x_1), & Q_{n_2}(x_2) & \cdots & Q_{n_2}(x_k) & 1 \\ \vdots & \vdots & & \vdots & \vdots \\ Q_{n_{k+1}}(x_1), & Q_{n_{k+1}}(x_2) & \cdots & Q_{n_{k+1}}(x_k) & 1 \end{vmatrix}$$

Subtracting the kth row from the $k+1$ row, the k-1th row from the kth row, etc., and finally the first from the second row we have

$$(20) \qquad Q\binom{n_1, n_2, \cdots, n_{k+1}}{x_1, x_2, \cdots, x_{k+1}}$$

$$= (-1)^k \begin{vmatrix} Q_{n_2}(x_1) - Q_{n_1}(x_1), & Q_{n_2}(x_2) - Q_{n_1}(x_2), & \cdots, & Q_{n_2}(x_k) - Q_{n_1}(x_k) \\ \vdots & & & \\ Q_{n_{k+1}}(x_1) - Q_{n_k}(x_1), & \cdots & & Q_{n_{k+1}}(x_k) - Q_{n_k}(x_k) \end{vmatrix}.$$

Observe that

$$(21) \qquad Q_{n_l}(x) - Q_{n_{l-1}}(x) = -x \sum_{i=n_{l-1}}^{n_l - 1} \frac{1}{\lambda_i \pi_i} H_i(x)$$

where

$$H_i(x) = \frac{\lambda_i \pi_i [Q_{i+1}(x) - Q_i(x)]}{-x}$$

comprise an orthogonal system of polynomials with respect to the measure $x d\psi / \lambda_0$ which concentrates its measure on $(0, \infty)$ since ψ does [2, p 504]. Therefore,

$$(-1)^{k(k-1)/2} H\binom{m_1, m_2, \cdots, m_k}{x_1, x_2, \cdots, x_k} > 0$$

whenever $x_1 < x_2 < x_3 < \cdots < x_k \le 0$ and $m_1 < m_2 < \cdots < m_k$ since $H_i(0) = \sum_{j=0}^{i} \pi_j Q_j(0) > 0$. Inserting (21) into (20) shows that $(-1)^k Q\binom{n_1, n_2, \cdots, n_{k+1}}{x_1, x_2, \cdots, x_{k+1}}$ can be written as

$$(-1)^k x_1 x_2 \cdots x_k \sum_{\mu' s} \gamma_{\mu_1, \mu_2, \cdots, \mu_k} H\binom{\mu_1, \mu_2, \cdots, \mu_k}{x_1, x_2, \cdots, x_k}$$

where the μ's traverse the sets $n_j \le \mu_j \le n_{j+1} - 1$, $j = 1, 2, \cdots, k$ respectively, and $\gamma_{\mu_1, \cdots, \mu_k} > 0$. Taking account of the inequality $x_j < 0$, $j = 1, 2, \cdots, k$ and the induction hypothesis which insures the inequality $(-1)^{k(k-1)/2} H\binom{\mu_1, \cdots, \mu_k}{x_1, \cdots, x_k} > 0$ we obtain

$$(-1)^k (-1)^{k(k-1)/2} Q\binom{n_1, n_2, \cdots, n_{k+1}}{x_1, x_2, \cdots, x_{k+1}} > 0$$

as asserted. This completes the proof.

A little manipulation of (19) will show that

$$(22)\qquad (-1)^{k(k-1)/2}\begin{vmatrix} Q_{n_1}(\gamma) & Q'_{n_1}(\gamma) & \cdots & Q^{(k-1)}_{n_1}(\gamma) \\ Q_{n_2}(\gamma) & & \cdots & \\ \vdots & & & \\ Q_{n_k}(\gamma) & Q'_{n_k}(\gamma) & \cdots & Q^{(k-1)}_{n_k}(\gamma) \end{vmatrix} \geq 0$$

true for every $\gamma \leq a$. This is verified by subtracting the last column of (19) from the next to last and using the mean value theorem. Repeating this k times and afterwards letting all the x_i converge to γ produces (22). Subject to the correct normalization the argument employed in paragraph (C) above shows that these determinants are actually strictly positive.

A further sharpening of the relation (19) and (22) is possible. In order to describe this extension we must assign a special meaning to the determinant

$$Q^*\binom{n_1, n_2, \cdots, n_k}{x_1, x_2, \cdots, x_k}$$

where $n_1 < n_2 < \cdots < n_k$ and $x_1 \leq x_2 \leq \cdots \leq x_k$ and distinguished in that several of the x's can be equal. (The asterisk sign on the Q shall always occur when one or more of the x's are equal and indicates that a special interpretation is to be made.) Let us illustrate by means of an example.

If $x_1 < x_2 = x_3 < x_4 = x_5 = x_6$ then

$$Q^*\binom{n_1, n_2, \cdots, n_6}{x_1, x_2, \cdots, x_6} = \begin{vmatrix} Q_{n_1}(x_1) & Q_{n_1}(x_2) & Q'_{n_1}(x_2) & Q_{n_1}(x_4) & Q'_{n_1}(x_4) & Q''_{n_1}(x_4) \\ Q_{n_2}(x_1) & Q_{n_2}(x_2) & Q'_{n_2}(x_2) & Q_{n_2}(x_4) & Q'_{n_2}(x_4) & Q''_{n_2}(x_4) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Q_{n_6}(x_1) & Q_{n_6}(x_2) & Q'_{n_6}(x_2) & Q_{n_6}(x_4) & Q'_{n_6}(x_4) & Q''_{n_6}(x_4) \end{vmatrix}$$

In general, when there is a block of equal x values present, the successive columns, corresponding to these x values in forming Q^* are determined by the successive derivatives, *i.e.* (Q_n), (Q'_n), (Q''_n), $\cdots$, $(Q_n^{(r-1)})$ where r is the number of equal x values.

One can show by a more tedious elaboration of the methods in (C) and (D) that generally

$$(23)\qquad (-1)^{k(k-1)/2}Q^*\binom{n_1, n_2, \cdots, n_k}{x_1, x_2, \cdots, x_k} > 0$$

when $x_1 \leq x_2 \leq \cdots \leq x_k \leq a$ with the emphasis on strict inequality in (23).

We do not indicate the details since an analogous argument will be used in the proof of Theorem 1.

(E) With the aid of the results of (C) and (D) we shall deduce determinantal inequalities valid for special choices of positive x's. Let $Q_n(x)$ be a system of orthogonal polynomials normalized as usual so that $Q_n(0)>0$ and ψ its measure on $[0, \infty)$. Let us suppose the measure ψ begins with isolated jumps located at $a_1 < a_2 < \cdots < a_r$ followed by a non-isolated point in the spectrum starting at a_{r+1} where r may be $0, 1, 2, \cdots$. In particular, when $r = 0$ then the first point in the spectrum of ψ is not an isolated jump. On the other extreme if $r = \infty$ then the first portion of the spectrum ψ consists of an infinite number of isolated jumps which could include the full spectrum. It is not necessary, in what follows, to describe more precisely the spectrum beyond a_{r+1}.

THEOREM 1. *Let $0 \le n_1 < n_2 < n_3 < \cdots < n_k$ and Q_n be normalized as usual such that $Q_n(0) > 0$; then for $k \le r$,*

$$(-1)^{k(k-1)/2} Q\binom{n_1, n_2, \cdots, n_k}{a_1, a_2, \cdots, a_k} > 0 \tag{24}$$

and for $k > r$,

$$(-1)^{k(k-1)/2} Q^*\binom{n_1, n_2, \cdots, n_r, n_{r+1}, n_{r+2}, \cdots, n_k}{a_1, a_2, \cdots, a_r, a_{r+1}, a_{r+1}, \cdots, a_{r+1}} > 0 \tag{25}$$

where Q^* is defined as above.

Proof. The proof is by induction on the order of the determinant k. The case where $r = 0$ has already been completely examined in paragraph (C). Hence, we assume $r \ge 1$. We suppose furthermore that the theorem has been established with regard to any orthogonal polynomial system whose spectral measure concentrates on the non-negative axis with the number of initial isolated jumps totalling less than r. Let r be fixed and ≥ 1 and suppose we have established the theorem for determinants of size $< k$. Denote by $P_n(x) = Q_n(x + a_1)/Q_n(a_1)$. These polynomials constitute an orthogonal system with respect to the measure $\psi(x + a_1)$ whose first mass points occur at $0, a_2 - a_1, a_3 - a_1, \cdots, a_r - a_1$. Observe that

$$P^*\binom{n_1, n_2, \cdots, n_k}{b_1, b_2, \cdots, b_k} = Q^*\binom{n_1, n_2, \cdots, n_k}{a_1, a_2, \cdots, a_k} C(n_1, n_2, \cdots, n_k)$$

where $C(n_1, n_2, \cdots, n_k) > 0$ and

$$b_1 = 0, b_2 = a_2 - a_1, \cdots, b_r = a_r - a_1, b_j = a_{r+1} - a_1 \quad \text{for} \quad j \ge r + 1 .$$

Subtracting the k-1th row from the kth row, the k-2th row from the k-1th row etc., we obtain

$$P^*\binom{n_1, n_2, \cdots, n_k}{b_1, b_2, \cdots, b_k}$$
$$= \begin{vmatrix} P_{n_2}(b_2) - P_{n_1}(b_2), & \cdots, P_{n_2}(b_{r+1}) - P_{n_1}(b_{r+1}), & P'_{n_2}(b_{r+1}) - P'_{n_1}(b_{r+1}) & \cdots \\ \vdots & \vdots & \vdots & \\ P_{n_k}(b_2) - P_{n_{k-1}}(b_2), & \cdots, P_{n_k}(b_{r+1}) - P_{n_{k-1}}(b_{r+1}), & P'_{n_k}(b_{r+1}) - P'_{n_{k-1}}(b_{r+1}) & \cdots \end{vmatrix}$$

The right-hand side is a determinant of size $k - 1$. Dividing the respective columns by $-1/b_\nu$, $\nu = 2, 3, \cdots$, (remember $b_2 > 0$ since $r \geq 1$), we have

$$(26) \qquad (-1)^{k-1} \operatorname{sign} P^*\binom{n_1, n_2, \cdots, n_k}{b_1, b_2, \cdots, b_k}$$
$$= \operatorname{sign} \begin{vmatrix} \frac{P_{n_2}(b_2) - P_{n_1}(b_2)}{-b_2}, \cdots, \frac{P_{n_2}(b_{r+1}) - P_{n_1}(b_{r+1})}{-b_{r+1}}, & \frac{P'_{n_2}(b_{r+1}) - P'_{n_1}(b_{r+1})}{-b_{r+1}} \cdots \\ \vdots & \\ \frac{P_{n_k}(b_2) - P_{n_{k-1}}(b_2)}{-b_2}, \cdots, \frac{P_{n_k}(b_{r+1}) - P_{n_{k-1}}(b_{r+1})}{-b_{r+1}}, & \frac{P'_{n_k}(b_{r+1}) - P'_{n_{k-1}}(b_{r+1})}{-b_{r+1}} \cdots \end{vmatrix}.$$

Let

$$H_r(x) = \lambda_r^* \pi_r^* \left[\frac{P_{r+1}(x) - P_r(x)}{-x} \right]$$

and set

$$M_r^{(l)}(x) = \frac{\lambda_r^* \pi_r^* [P_{r+1}^{(l)}(x) - P_r^{(l)}(x)]}{-x}$$

$l = 0, 1, 2, \cdots$, so that $M_r^{(0)} = H_r$ and λ_r^* and μ_r^* are the parameters corresponding to the polynomial system $P_n(x)$. Finally, for $0 \leq \mu_1 < \mu_2 < \cdots < \mu_{k-1}$ define

$$L\binom{\mu_1, \cdots, \mu_{k-1}}{b_2, \cdots, b_k} = \begin{vmatrix} H_{\mu_1}(b_2) & \cdots H_{\mu_1}(b_{r+1}), & M'_{\mu_1}(b_{r+1}) & \cdots & M_{\mu_1}^{(k-1-r)}(b_{r+1}) \\ \vdots & & & & \\ H_{\mu_{k-1}}(b_2) & \cdots H_{\mu_{k-1}}(b_{r+1}), & M'_{\mu_{k-1}}(b_{r+1}) & \cdots & M_{\mu_{k-1}}^{(k-1-r)}(b_{r+1}) \end{vmatrix}.$$

Expanding the right-hand side of (26), using (21) and an analogous formula for the successive derivatives of $Q_n(x)$, we obtain

$$\text{(27)}\quad \operatorname{sign}(-1)^{k-1}P^*\begin{pmatrix} n_1, n_2, \cdots, n_k \\ b_1, b_2, \cdots, b_k \end{pmatrix} = \operatorname{sign}\sum_{\mu}\gamma_{\mu_1,\mu_2,\cdots,\mu_{k-1}}L\begin{pmatrix} \mu_1, \mu_2, \cdots, \mu_{k-1} \\ b_2, b_3, \cdots, b_k \end{pmatrix},$$

where the γ's are positive and $n_i \le \mu_i < n_{i+1}$, $(i = 1, \cdots, k-1)$. But

$$H_r^{(l)}(x) = M_r^{(l)}(x) + \sum_{t=0}^{l-1} c_t(x) M_r^{(t)}(x)$$

Hence by suitable operations on the columns of L we obtain

$$L\begin{pmatrix} \mu_1, \cdots, \mu_{k-1} \\ b_2, \cdots, b_k \end{pmatrix} = H^*\begin{pmatrix} \mu_1, \mu_2, \cdots, \mu_{k-1} \\ b_2, b_3, \cdots, b_k \end{pmatrix}$$

where the H^* determinant is formed from the polynomial system H_n in the same way that Q^* is constructed in terms of Q_n.

The H system represent orthogonal polynomials with respect to a measure $d\alpha(x) = C \cdot x d\psi(x + a_1)$. The jump at the origin of $d\psi(x + a_1)$ is obliterated due to the factor x. Otherwise, α possesses $r - 1$ initial jumps located at $a_2 - a_1, \cdots, a_r - a_1$ and the non-isolated portion of the spectrum begins at the point $a_{r+1} - a_1$. By the induction hypothesis $(-1)^{(k-1)(k-2)/2}H^*\begin{pmatrix} \mu_1, \mu_2, \cdots, \mu_{k-1} \\ b_2, b_3, \cdots, b_k \end{pmatrix} > 0$. This fact in conjunction with (27) shows that

$$(-1)^{k(k-1)/2}P^*\begin{pmatrix} n_1, n_2, \cdots, n_k \\ b_1, b_2, \cdots, b_k \end{pmatrix} > 0$$

as desired. The proof of the theorem is complete.

What is essential for the validity of (25) is that the first r choices of $y_i > 0$ used in evaluating (25) should coincide with the first spectral points a_i of ψ (here r has the same meaning as in the theorem). Otherwise the values of y_j $(j \ge r + 1)$ can be arbitrarily chosen from the interval $a_r < y \le a_{r+1}$ with the restriction that they are arranged in ascending order even allowing equalities. Actually, more is true. A careful examination of the above arguments shows that

$$\text{(28)}\qquad (-1)^{k(k-1)/2}Q^*\begin{pmatrix} n_1, n_2, \cdots, n_s, n_{s+1}, \cdots, n_k \\ a_1, a_2, \cdots, a_s, y_{s+1}, \cdots, y_k \end{pmatrix} > 0 \quad (s \le r)$$

where n_i strictly increase and y_j for $j \ge s + 1$ satisfy $a_s < y_{s+1} \le y_{s+2} \le \cdots \le y_k \le a_{s+1}$.

To complete the story we note without proof that it is possible to construct examples which show that $Q\begin{pmatrix} n, n+1 \\ x, y \end{pmatrix}$ does not possess a fixed sign for all n when x and y satisfy $x < a_1$ and $a_1 < y < a_2$.

2. The compound process. The infinite matrix $P(t)$ satisfies the differential equations

$$\frac{dP(t)}{dt} = AP(t) ,$$

$$\frac{dP(t)}{dt} = P(t)A ,$$

called respectively the backward and the forward equations of the birth and death process. Either equation may be derived from the other when it is known that both $P(t)$ and A satisfy the symmetry relations

$$P_{ij}(t)\pi_i = P_{ji}(t)\pi_j, \ \ a_{ij}\pi_i = a_{ji}\pi_j \ .$$

As a consequence of these equations we deduce the backward and forward equations of the compound process:

$$\begin{aligned}
\frac{d}{dt} P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right) &= \sum_{r=1}^{n} \left\{ \mu_{i_r} P\left(t\,;\begin{matrix} i_i, \cdots, i_{r-1}, i_r - 1, i_{r+1}, \cdots, i_n \\ j_1, \cdots, j_{r-1}, j_r, j_{r+1}, \cdots, j_n \end{matrix}\right) \right. \\
&\quad \left. - (\lambda_{ir} + \mu_{ir}) P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right) + \lambda_{i_r} P\left(t\,;\begin{matrix} i_1, \cdots, i_{r-1}, i_r + 1, i_{r+1}, \cdots, i_n \\ j_1, \cdots, j_{r-1}, j_r, j_{r+1}, \cdots, j_n \end{matrix}\right)\right\}
\end{aligned}$$

(29)

$$\begin{aligned}
\frac{d}{dt} P\left(t\,;\begin{matrix} i_1, \cdots, j_n \\ j_1, \cdots, j_n \end{matrix}\right) &= \sum_{r=1}^{n} \left\{ \lambda_{j_r - 1} P\left(t\,;\begin{matrix} i_1, \cdots, i_{r-1}, i_r, i_{r+1}, \cdots, i_n \\ j_1, \cdots, j_{r-1}, j_r - 1, j_{r+1}, \cdots, j_n \end{matrix}\right) \right. \\
&\quad \left. - (\lambda_{j_r} + \mu_{j_r}) P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right) + \mu_{j_r + 1} P\left(t\,;\begin{matrix} i_1, \cdots, i_{r-1}, i_r, i_{r+1}, \cdots, i_n \\ j_1, \cdots, j_{r-1}, j_r + 1, j_{r+1}, \cdots, j_n \end{matrix}\right)\right\} .
\end{aligned}$$

Here we employ the natural convention that $P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right)$ for $i_1 \leq \cdots \leq i_n$ and $j_1 \leq \cdots \leq j_n$ is zero if any two i_ν or any two j_ν are equal or if $i_1 = -1$ or $j_1 = -1$. The first of the above equations (backward equation) follows at once from

$$\begin{aligned}
&\frac{d}{dt} P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right) \\
&= \textstyle\sum_\sigma (\operatorname{sign} \sigma) P_{i_1 j_{\sigma_1}}(t) \cdots P_{i_{r-1} j_{\sigma_{r-1}}}(t) \left\{ \dfrac{d}{dt} P_{i_r j_{\sigma_r}}(t) \right\} P_{i_{r+1} j_{\sigma_{r+1}}}(t) \cdots P_{i_n j_{\sigma_n}}(t)
\end{aligned}$$

on applying the backward equation, $P'(t) = AP(t)$. Here $\sum_\sigma$ denotes summation over all permutations $\sigma = \begin{pmatrix} 1, \cdots, n \\ \sigma_1, \cdots, \sigma_n \end{pmatrix}$ of $1, 2, \cdots, n$. The forward equation may be obtained in a similar way from the forward equation of the original process. Alternatively either of the two equations is a consequence of the other one together with the symmetry relations

$$P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right)\pi_{i_1}\cdots\pi_{i_n} = P\left(t\,;\begin{matrix} j_1, \cdots, j_n \\ i_1, \cdots, i_n \end{matrix}\right)\pi_{j_1}\cdots\pi_{j_n}\,,$$

and $\lambda_r\pi_r = \mu_{r+1}\pi_{r+1}$.

The backward and forward equations of the compound process may also be derived from the representation (7) and the fact the determinantal polynomials $Q\begin{pmatrix} i_1, \cdots, i_n \\ x_1, \cdots, x_n \end{pmatrix}$ satisfy the recurrence formula

$$\begin{aligned} -(x_1+\cdots+x_n)Q\begin{pmatrix} i_1, \cdots, i_n \\ x_1, \cdots, x_n \end{pmatrix} &= \sum_{r=1}^{n}\Big[\mu_r Q\begin{pmatrix} i_1, \cdots, i_{r-1}, i_r-1, i_r, \cdots, i_n \\ x_1, \cdots\cdots\cdots\cdots\cdots, x_n \end{pmatrix} \\ -(\lambda_r+\mu_r)Q\begin{pmatrix} i_1, \cdots, i_n \\ x_1, \cdots, x_n \end{pmatrix} &+ \lambda_r Q\begin{pmatrix} i_1, \cdots, i_{r-1}, i_r+1, i_{r+1}, \cdots, i_n \\ x_1, \cdots\cdots\cdots\cdots\cdots, x_n \end{pmatrix}\Big] \end{aligned} \tag{30}$$

where $Q\begin{pmatrix} j_1, \cdots, j_n \\ x_1, \cdots, x_n \end{pmatrix}$ for $j_1 \le \cdots \le j_n$ is taken to be zero if any two j_ν are the same or if $j_1 = -1$. This recurrence formula follows at once by applying the basic recurrence formula $-xQ(x) = AQ(x)$ to the right member of the identity

$$\begin{aligned} &-(x_1+\cdots+x_n)Q\begin{pmatrix} i_1, \cdots, i_n \\ x_1, \cdots, x_n \end{pmatrix} \\ &= \sum_{r=1}^{n}\sum\nolimits_\sigma (\operatorname{sign}\sigma)Q_{i_1}(x_{\sigma_1})\cdots Q_{i_{r-1}}(x_{\sigma_{r-1}})[-x_{\sigma_r}Q_{i_r}(x_{\sigma_r})]\cdots Q_{i_n}(x_{\sigma_n})\,. \end{aligned}$$

It is not difficult to see that $P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right)$ converges to zero as $t \to \infty$. In fact if the original birth and death process is either transient or recurrent null then $P_{ij}(t) \to 0$ for each i and j so the determinant $\to 0$. On the other hand if the original birth and death process is recurrent (either ergodic or recurrent null) and $F_{i0}(t)$ is the probability that first passage from state i to state 0 occurs in time $\le t$ then $F_{i,0}(t) \to 1$ and from probabilistic considerations

$$P\left(t\,;\begin{matrix} i_1. \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right) \le 1 - F_{i_n,0}(t) \to 0 \text{ as } t \to \infty\,.$$

Thus we have two reasons why the determinants may $\to 0$ and at least one of them is always in force.

According to the Doeblin-Chung ratio theorem [1]

$$\lim_{t\to\infty}\frac{\int_0^t P\left(\tau\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right)d\tau}{\int_0^t P\left(\tau\,;\begin{matrix} k_1, \cdots, k_n \\ l_1, \cdots, l_n \end{matrix}\right)d\tau}$$

exists and is finite and positive. For the compound process of the birth and death process we are able to make the following considerably sharper statement.

THEOREM 2.

$$\lim_{t\to\infty} \frac{P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right)}{P\left(t\,;\begin{matrix} k_1, \cdots, k_n \\ l_1, \cdots, l_n \end{matrix}\right)}$$

exists and is finite and positive.

Proof. It is evidently sufficient to consider the case when $(k_1, \cdots, k_n) = (l_1, \cdots, l_n) = (0, 1, \cdots, n-1)$. Let $f(x_1, \cdots, x_n)$ be the polynomial such that

$$Q\begin{pmatrix} i_1, & \cdots, & i_n \\ x_1, & \cdots, & x_n \end{pmatrix} Q\begin{pmatrix} j_1, & \cdots, & j_n \\ x_1, & \cdots, & x_n \end{pmatrix} = f(x_1, \cdots, x_n)\left[Q\begin{pmatrix} 0, & \cdots, & n-1 \\ x_1, & \cdots, & x_n \end{pmatrix}\right]^2 .$$

We wish to show that

$$(31)\quad \frac{\displaystyle\int_{0\le x_1<\cdots<x_n}\!\!\!\cdots\int e^{-(x_1+\cdots+x_n)t} f(x_1, \cdots, x_n)\left[Q\begin{pmatrix} 0, & \cdots, & n-1 \\ x_1, & \cdots, & x_n \end{pmatrix}\right]^2 d\psi(x_1)\cdots d\psi(x_n)}{\displaystyle\int_{0\le x_1<\cdots<x_n}\!\!\!\cdots\int e^{-(x_1+\cdots+x_n)t} Q\left[\begin{pmatrix} 0, & \cdots, & n-1 \\ x_1, & \cdots, & x_n \end{pmatrix}\right]^2 d\psi(x_1)\cdots d\psi(x_n)}$$

converges to a finite positive limit as $t \to \infty$. Suppose there are x values $0 \le a_1 < \cdots < a_{r+1}$ such that the function $\psi(x)$ has positive jumps at $a_1, \cdots, a_r$ but no other spectrum in $0 \le x < a_{r+1}$ while ψ has infinitely many points of increase in every interval $a_{r+1} < x < a_{r+1} + \varepsilon$. We consider separately the cases $r \ge n$, $1 \le r < n$ and $r = 0$. The case $1 \le r < n$, which exhibits all the necessary arguments, will be discussed in detail and the other two cases are left as an exercise for the interested reader. When $1 \le r < n$ integrals of the form

$$\int_{0\le x_1<\cdots<x_n}\!\!\!\cdots\int F(x_1, \cdots, x_n) d\psi(x_1)\cdots d\psi(x_n)$$

may be written in the form

$$\int_{\substack{a_1\le x_1\le\cdots\le x_n \\ x_2 \ge a_2 \\ \vdots \\ x_{r+1} \ge a_{r+1}}}\!\!\!\cdots\int F(x_1, \cdots, x_n) d\psi(x_1)\cdots d\psi(x_n) .$$

For large t the main contributions to the integrals in (31) therefore come from the neighborhood of the point $(x_1, \cdots, x_n) = (a_r, \cdots, a_r, a_{r+1}, \cdots, a_{r+1})$. To make this precise we first observe that Theorem 1 shows that $f(a_1, \cdots, a_r, a_{r+1}, \cdots, a_{r+1}) = c$ is positive and that the measure

$$\left[Q\binom{0, \cdots, n-1}{x_1, \cdots, x_n}\right]^2 d\psi(x_1) \cdots d\psi(x_n) = d\theta(x_1, \cdots, x_n)$$

has positive mass in every "right-hand" neighborhood of the point $(a_1, \cdots, a_r, a_{r+1}, \cdots, a_{r+1})$. The expression (31) can be written in the form $c + (I_1/I_2)$ where

$$I_1 = \int \cdots \int e^{-(x_1+\cdots+x_n)t+(a_1+\cdots+a_r+a_{r+1}+\cdots+a_{r+1})t} \cdot [f(x_1, \cdots, x_n) - c] d\theta(x_1, \cdots, x_n) ,$$

$$I_2 = \int \cdots \int e^{-(x_1+\cdots+x_n)t+(a_1+\cdots a_r+a_{r+1}+\cdots+a_{r+1})t} d\theta(x_1, \cdots, x_n) .$$

Given $\varepsilon > 0$ we choose $\delta > 0$ so $|f(x_1, \cdots, x_n) - c| < \varepsilon$ for $|x_1 - a_1| + \cdots + |x_r - a_r| + |x_{r+1} - a_{r+1}| + \cdots + |x_n - a_{r+1}| \leq \delta$. Let R_δ and R'_δ denote the parts of the region $0 \leq x_1 < \cdots < x_n$ where

$x_1 + \cdots + x_n \leq a_1 + \cdots + a_r + (n-r)a_{r+1} + \delta$ and where
$x_1 + \cdots + x_n > a_1 + \cdots + a_r + (n-r)a_{r+1} + \delta$ respectively.

Then

$$\begin{aligned} |I_1| &= \left|\int_{R_\delta} \cdots \int + \int_{R'_\delta} \cdots \int\right| \\ &\leq \varepsilon \int \cdots \int e^{-(x_1+\cdots+x_n)t+(a_1+\cdots)t} d\theta(x_1, \cdots, x_n) \\ &\quad + e^{-\delta t} \int_{R'_\delta} \cdots \int |f(x_1, \cdots, x_n) - c| \, d\theta(x_1, \cdots, x_n) \end{aligned}$$

while

$$\begin{aligned} |I_2| &\geq \int_{R_\delta} \cdots \int e^{-(x_1+\cdots)t+(a_1+\cdots)t} d\theta(x_1, \cdots, x_n) \\ &\geq \int_{R_{\frac{1}{2}\delta}} \cdots \int e^{-(x_1+\cdots)t+(a_1+\cdots)t} d\theta(x_1, \cdots, x_n) \geq Be^{-\frac{1}{2}\delta t} \end{aligned}$$

where $B > 0$. Consequently $\limsup_{t\to\infty} |(I_1)/(I_2)| \leq \varepsilon$ and the theorem follows.

3. Some examples of the probability distribution of the time until coincidence. A random variable of natural interest to the study of the compound process of order n is the time t^* until coincidence. To expedite the discussion we restrict attention to the case of the compound process involving two particles. The obvious extensions are left to the reader. In general, coincidence need not occur with certainty. We define t^* to be the time of first coincidence if this is finite and to be $+\infty$ otherwise. In the next section the condition that coincidence be a certain event is expressed in terms of the parameters of the birth and death process. In this section the explicit distribution of t^* is determined for some important examples.

We begin with a few remarks concerning the general character of this problem. We may consider a two-dimensional birth and death process whose states are all pairs (i, j) with $i \geq 0$, $j \geq 0$ and transition probability law

$$P_{ij;kl} = P_{ik}(t)P_{jl}(t) \ .$$

In this formulation the problem is to determine the distribution of the time of first hitting the diagonal ray $i = j$.

Alternatively, we may consider the compound process with state space (i, j), $0 \leq i < j$ and transition probability law $P\left(t\,;\begin{smallmatrix} i, & j \\ k, & l \end{smallmatrix}\right)$. In this formulation, coincidence occurs if the particle is in some state $(k, k+1)$ and is then absorbed—the process terminates at $(k, k+1)$. The problem is then to determine the distribution of the time until the process terminates in this manner.

Let $S^{ij}(t)$, $(0 \leq i < j)$ denote the probability distribution of the time until coincidence when the initial states of the particles are respectively i and j; *i.e.*

$$S^{ij}(t) = \Pr\{t^* \leq t \mid x(0) = i,\ y(0) = j,\ i < j\}$$

Because the path functions are continuous (a particle moving from state i to state j in time t must occupy all the intermediate states in the intervening time), coincidence can only occur following a transition from a state $(k, k+1)$ for some k. More exactly, the probability that coincidence happens during the time interval $[t, t+h]$ with h sufficiently small requires that the two particles occupy adjacent states before coincidence at time t and at the next transition the particles meet. The probability of this event is clearly

$$\sum_{k=0}^{\infty} P\left(t\,;\begin{matrix} i, & j \\ k, & k+1 \end{matrix}\right)(\lambda_k + \mu_{k+1})h + o(h)$$

and the density function of the time until coincidence is

$$R^{ij}(t) = \frac{dS^{ij}}{dt}(t) = \sum_{k=0}^{\infty} (\lambda_k + \mu_{k+1})P\left(t\,;\begin{matrix} i, & j \\ k, & k+1 \end{matrix}\right) .$$

The method we use to compute $S^{ij}(t)$ consists of determining explicitly the generating function

$$G(z, w) = \sum_{0\le i<j} R^{ij}(t)(z^i w^j - z^j w^i)$$

or sometimes more conveniently

$$H(z, w) = \sum_{0\le i<j} \pi_i \pi_j R^{ij}(t)(z^i w^j - z^j w^i)$$

and then reading off the coefficient of $z^i w^j$, $(i < j)$.

If we have available

$$f_k(z, t) = \sum_{l=0}^{\infty} P_{kl}(t) z^l \tag{32}$$

and hence

$$\pi_k f_k(z, t) = \sum_{l=0}^{\infty} \pi_l P_{lk}(t) z^k$$

we obtain employing (10) the determinantal identity

$$\begin{aligned} M(k, z, w, t) &= \pi_k \pi_{k+1} \begin{vmatrix} f_k(z, t) & f_{k+1}(z, t) \\ f_k(w, t) & f_{k+1}(w, t) \end{vmatrix} \\ &= \sum_{0\le l_1<l_2} \pi_{l_1} \pi_{l_2} P\begin{pmatrix} t\,; & l_1, & l_2 \\ & k, & k+1 \end{pmatrix} (w^{l_2} z^{l_1} - w^{l_1} z^{l_2}) \end{aligned}$$

where $z < w$.

Direct summation gives

$$\sum_{k=0}^{\infty} (\lambda_k + \mu_{k+1}) M(k, z, w, t) = \sum_{0\le l_1<l_2} \pi_{l_1} \pi_{l_2} R^{l_1 l_2}(t)[w^{l_2} z^{l_1} - w^{l_1} z^{l_2}]\,. \tag{33}$$

In many cases it is possible to recognize the left-hand side of (33) in terms of classical functions and then obtain $R^{l_1 l_2}(t)$ by picking out the proper coefficient in the series expansion. We record several important examples.

EXAMPLE 1. Consider the telephone trunking model ($\lambda_n = \lambda$, $\mu_n = n\mu$, $n \ge 0$) [4]. The orthogonal polynomials are the Poisson Charlier polynomials. The generating function of the transition probabilities is known to be

$$f_k(z, t) = e^{-a(1-z)(1-e^{-\mu t})}[1 - (1 - z)e^{-\mu t}]^k = \alpha_t(z)[\beta_t(z)]^k$$

where $a = \lambda/\mu$ and $\alpha_t(z)$ and $\beta_t(z)$ are defined in the obvious fashion. The preceding calculations in this case yield

$$\sum_{0 \le l_1 < l_2} \pi_{l_1}\pi_{l_2} R^{l_1 \cdot l_2}(t)[w^{l_2}z^{l_1} - w^{l_1}z^{l_2}] \tag{34}$$
$$= a(w - z)e^{-\mu t}\alpha_t(z)\alpha_t(w) \sum_{k=0}^{\infty} [(k+1)\mu + \lambda]\frac{[a^2\gamma_t(w, z)]^k}{k!\,(k+1)!}$$

where $\gamma_t(w, z) = \beta_t(z)\beta_t(w)$. This is a combination of Bessel functions viz. $\mu I_0(2\sqrt{a^2\gamma}) + \lambda/\sqrt{a^2\gamma}\ I_1(2\sqrt{a^2\gamma})$ where I_ν denotes the usual Bessel function with imaginary argument. If we specialize to the coefficient of z^0w^1 we get

$$R^{01}(t) = e^{-\mu t}e^{-2a\sigma}\left[I_0(2a\sigma) + \frac{\lambda}{a\sigma}I_1(2a\sigma)\right] \text{ where } \sigma = 1 - e^{-\mu t}\ .$$

EXAMPLE 2. Consider the linear growth birth and death process where

$$\lambda_n = (n + 1 + \alpha)\kappa \text{ and } \mu_n = n\kappa \qquad n \ge 0$$

and α is real, $\alpha > -1$. The associated orthogonal polynomials are the Laguerre system normalized at the origin equal to 1. Utilizing the generating function of [5 eq. (25)] we obtain

$$\sum_{0 \le l_1 < l_2} \pi_{l_1}\pi_{l_2} R^{l_1 l_2}(t)[w^{l_2}z^{l_1} - w^{l_1}z^{l_2}] \tag{35}$$
$$= \kappa(\alpha + 1)\delta_t(z)\delta_t(w)[\gamma_t(w) - \gamma_t(z)]\{2F(\alpha + 1, \alpha + 2, 1, u_t(z, w))$$
$$+ \alpha F(\alpha + 1, \alpha + 2, 2, u_t(z, w)\}$$

where F denotes the standard hypergeometric function

$$F(a, b, c, t) = \sum_{n=0}^{\infty} \frac{(a)_n(b)_n}{(c)_n n!} t^n$$

and $(a)_n = \Gamma(a + n)/\Gamma(a)$. Here,

$$\delta_t(z) = \left[\frac{1}{(1 + \kappa t)\left(1 - \frac{\kappa t}{1 + \kappa t}z\right)}\right]^{\alpha+1},\quad \gamma_t(z) = \frac{\kappa t}{1 + \kappa t}\frac{\left(1 + \frac{1 - \kappa t}{\kappa t}z\right)}{\left(1 - \frac{\kappa t}{1 + \kappa t}z\right)}$$

and $u_t(z, w) = \gamma_t(z)\gamma_t(w)$. The coefficient of z^0w^1 in (35) reduces to

$$\frac{\kappa(\alpha + 1)}{(1 + \kappa t)^{2\alpha+4}}\left\{2F\left(\alpha + 1,\ \alpha + 2,\ 1, \left(\frac{\kappa t}{1 + \kappa t}\right)^2\right)\right. \tag{36}$$
$$\left. + \alpha F\left(\alpha + 1, \alpha + 2, 2, \left(\frac{\kappa t}{1 + \kappa t}\right)^2\right)\right\}\ .$$

The coincidence time density function $R^{01}(t)$ is the expression (36) apart from the constant factor $1/(\alpha + 1)$. When α is a non-negative integer the coincidence time density function reduces to a rational function. In the particular case $\alpha = 0$, we obtain

$$R^{01}(t) = \frac{2\kappa}{(1 + 2\kappa t)^2}$$

which shows that coincidence is certain with the expected time until coincidence infinite. This is true of all the linear growth processes introduced in this example.

If we examine a linear growth process where there exists a permanent absorbing state at -1 then obviously coincidence is never certain. It is of some interest to compute the probability of coincidence before absorption. Let us illustrate by considering the model where $\lambda_n = (n+1)\kappa$ and $\mu_n = (n + 1)\kappa$ for $n \geq 0$. A calculation similar to that above gives

$$R^{01}(t) = \frac{2\kappa}{(1 + 2\kappa t)^2(1 + \kappa t)^2} + \frac{\kappa}{(1 + \kappa t)^5} .$$

It is easy to evaluate $\int_0^\infty R^{01}(t)dt = 25/4 - 8 \log 2$. The reader may verify that this lies between 0 and 1 (approximately. 71).

4. The probability of coincidence. In this section we shall determine the exact conditions which imply that coincidence in a finite state is certain to occur. Our results apply to the case of n independent particles moving simultaneously subject to the transition law of the same birth and death process (B). Our methods may be extended in the obvious way to treat the case in which the particles are subject to different independent birth and death laws. Such a generalization is left to the reader.

If the process (B) is recurrent then coincidence is clearly certain. In fact, if two particles originate in states i and $j > i$, respectively then the second particle reaches the state 0 in finite time with probability one and coincidence must precede this event because of "continuity" of paths. Thus it remains to decide the probability of coincidence when the process (B) is transient.

In [3] we classified two kinds of transient processes. A transient birth and death process is said to be "weakly transient" if $\sum_{j=0}^\infty P_{ij}(t) \equiv 1$ for all t and some i. In terms of the birth and death rates this is equivalent to the divergence to infinity of the sequence

$$\sum_{n=0}^{m-1} \frac{1}{\lambda_n \pi_n} \left(\sum_{k=0}^{n} \pi_k \right) = - Q_m^{(\prime)}(0)$$

where Q_m are the associated polynomials of the process (B).

A birth and death process is said to be strongly transient if for some t and i, $\sum_{j=0}^\infty P_{ij}(t) < 1$. A necessary and sufficient condition for the process to be strongly transient is that, for any starting position and for any positive time value t, with positive probability the diffusing particle reaches infinity in time t.

It becomes evident that for strongly transient processes coincidence is not a certain event, since with positive probability one particle may stay in a given state (say i) in any specified length of time while the other particle moves to infinity without touching state i during this same period of time. An analogous argument will prove that the probability of coincidence for the case of n independently moving particles is not a certain event when the process is strongly transient. We shall determine in Theorem 3 the exact condition for coincidence to be certain. It will be clear that the criteria is the same for two, three or n particles.

We concentrate in what follows on the case of two particles. It is tempting to proceed as follows. Let w_{ij} denote the probability of no coincidence in finite time when two particles start respectively in states i and j $(i < j)$. We set $w_{kk} = 0$. Writing out a recursion relation in terms of the first transition, we obtain

$$
\begin{aligned}
w_{ij} = {} & \frac{\lambda_i}{\lambda_i + \lambda_j + \mu_i + \mu_j} w_{i+1,j} + \frac{\mu_i}{\lambda_i + \lambda_j + \mu_i + \mu_j} w_{i-1,j} \\
& \frac{\lambda_j}{\lambda_i + \lambda_j + \mu_i + \mu_j} w_{i,j+1} + \frac{\mu_j}{\lambda_i + \lambda_j + \mu_i + \mu_j} w_{i,j-1}
\end{aligned}
\tag{37}
$$

valid for all $0 \leq i \leq j$. A sufficient condition guaranteeing that coincidence is certain is that the only bounded positive solution of the system (37) is the identically zero solution. In the situation of non-certain coincidence it would also be of interest to calculate the probability of no coincidence w_{ij}. The investigation of this problem is complicated by the abundance of positive solutions that (37) possesses.

The study of (37) is interesting in itself and indicative of the difficulties associated with solving two-dimensional difference equation systems even in comparatively simple cases having probabilistic significance.

To illustrate this we exhibit several solutions of (37). Suppose the spectral measure ψ of (B) is located in the interval $[a, \infty)$ where $a \geq 0$. Then

$$
w_{ij}(\alpha) = -\frac{1}{2\alpha} \begin{vmatrix} Q_i(-\alpha), & Q_i(\alpha) \\ Q_j(-\alpha), & Q_j(\alpha) \end{vmatrix} = -\frac{1}{2\alpha} Q\begin{pmatrix} i, & j \\ -\alpha, & \alpha \end{pmatrix}
\tag{38}
$$

for each α satisfying $0 \leq \alpha \leq a$ is positive by virtue of Theorem 1. when $\alpha = 0$, $w_{ij}(0)$ is interpreted as $-Q'_j(0) + Q'_i(0) = \sum_{k=i}^{j-1} 1/\lambda_k \pi_k \sum_{r=0}^{k} \pi_r$. The verification that for all α, $w_{ij}(\alpha)$ is a solution of (37) is accomplished by choosing $x_1 = -\alpha$ and $x_2 = \alpha$ in the recursion law (30).

Unfortunately, there is no natural ordering among the solutions $w_{ij}(\alpha)$. We show first that $w_{0j}(\alpha)$ is increasing in α $(0 \leq \alpha \leq a)$ for each j. To this end, observe that

$$w_{0j}(\alpha) = Q_j(-\alpha) - Q_j(\alpha) = \sum_{k=1}^{j} a_{jk}\alpha^k$$

where a_{jk} is positive for k odd and zero for k even. Hence, $w_{0j}(\alpha)$ increases as asserted. On the other hand, we show that $w_{j,j+1}(\alpha)$ is decreasing in the same range of α. In fact, by virtue of a known representation [9 p. 42] we have

$$w_{j,j+1}(\alpha) = \frac{1}{\lambda_j\pi_j}\sum_{r=0}^{j}\pi_r Q_r(\alpha)Q_r(-\alpha) .$$

Hence, $w'_{j,j+1}(\alpha) = 1/\lambda_j\pi_j \sum_{r=0}^{j}\pi_r[-Q_r(\alpha)Q'_r(-\alpha) + Q'_r(\alpha)Q_r(-\alpha)]$. It is enough to show since $Q_r(\alpha)Q_r(-\alpha)$ is positive that

$$(39) \qquad \frac{-Q_r(\alpha)Q'_r(-\alpha) + Q'_r(\alpha)Q_r(-\alpha)}{Q_r(\alpha)Q_r(-\alpha)} = -\frac{Q'_r(-\alpha)}{Q_r(-\alpha)} + \frac{Q'_r(\alpha)}{Q_r(\alpha)} < 0 .$$

But the roots of $Q'_r(x)$ are separated by the roots of $Q_r(x)$ and since $Q_r(x)$ has no roots in $[-\infty, a)$ [9, p. 43] we conclude that $-Q'_r(x)/Q_r(x)$ is increasing.

The lack of order and the multiplicity of natural positive solutions seem to be the main sources of difficulty in proving the non-existence of any bounded positive solutions of (37). The solution $w_{ij}(0)$ should be singled out because it is always present (as $a \geq 0$) and also $\lim_{j\to\infty} w_{0j}(0) = \infty$ is precisely the condition that the process be weakly transient.

It should be added that the one parameter family of solutions, displayed in (38), when α is a positive number, does not exhaust in terms of linear span the totality of solutions. It appears that one can always construct at least a three parameter family of determinantal extremal solutions. The problem of characterizing all solutions of (37) in general remains open and relates to the problem of determining all determinantal polynomial systems satisfying the recursion law of (30).

We now turn to a discussion of the main theorem of this section.

THEOREM 3. *If the process (B) is recurrent or weakly transient then coincidence is certain if and only if*

$$(40) \qquad v_n = \sum_{k=1}^{n}\frac{1}{\lambda_k\pi_k}\sum_{r=0}^{k}\pi_r(w_{k+1} - w_r) \to \infty$$

where $w_m = \sum_0^{m-1} 1/\lambda_i\pi_i \sum_{j=0}^{i}\pi_j$ and $w_0 = 0$.

Before embarking on a proof of the theorem, it is necessary to interpret condition (40). To this end, denote by t_i the random variable which represents the length of time for a particle subject to the transition law of the process (B) to move from state i to state $i+1$. In other words t_i denotes the first passage time from state i to state $i+1$. In the same way, since the path functions are continuous, $z_n = t_0 + t_1 + \cdots + t_{n-1}$ represents the first passage time from state 0 to state n. The t_i are evidently independent but not identically distributed random variables.

The Laplace transform $\varphi_n(s)$ of the distribution of z_n is given by

$$\varphi_n(s) = \frac{1}{Q_n(-s)}$$

when Q_n is the nth orthogonal polynomial. More generally, the Laplace transform of the distribution of $t_m + t_{m+1} + \cdots + t_{n-1}$ is

$$\frac{Q_m(-s)}{Q_n(-s)} .$$

These formulae are proved as follows: The well-known Laplace transform formula, which expresses the first passage time distribution from state i to state j in terms of the transition probability function is

$$\hat{F}_{ij}(s) = \frac{\hat{P}_{ij}(s)}{\hat{P}_{jj}(s)} \qquad i \neq j. \tag{41}$$

Inserting the formula of [2 p. 522] in (41) gives the desired result.

From knowledge of the Laplace transform it is routine (successive differentiation of $\varphi_n(s)$ at zero) to determine the moments of z_n. In particular,

$$E(z_n) = \sum_0^{n-1} \frac{1}{\lambda_k \pi_k} \sum_{r=0}^{k} \pi_r = w_n = -Q_n'(0)$$

and

$$\text{variance } (z_n) = -Q_n''(0) + [Q_n'(0)]^2 . \tag{42}$$

From identity (11), we get

$$+\frac{Q_n''(0)}{2!} = \sum_{t=0}^{n-1} \frac{1}{\lambda_t \pi_t} \sum_{r=0}^{t} \pi_r[-Q_r'(0)] = \sum_{t=0}^{n-1} \frac{1}{\lambda_t \pi_t} \sum_{r=0}^{t} \pi_r w_r .$$

Inserting this in (42) leads to

$$\frac{\operatorname{Var}(z_n)}{2} = \frac{w_n^2}{2} - \sum_{t=0}^{n-1} \frac{1}{\lambda_t \pi_t} \sum_{r=0}^{t} \pi_r w_r .$$

But

$$\frac{1}{2}w_n^2 = \sum_{r=0}^{n-1}(w_{r+1}-w_r)w_{r+1} - \frac{1}{2}\sum_{r=0}^{n-1}(w_{r+1}-w_r)^2$$
$$= \sum_{r=0}^{n-1}\left(\frac{1}{\lambda_r\pi_r}\sum_{k=0}^{r}\pi_k\right)w_{r+1} - \frac{1}{2}\sum_{r=0}^{n-1}(w_{r+1}-w_r)^2$$

$$\frac{1}{2}\operatorname{Var}(z_n) = \sum_{r=0}^{n-1}\frac{1}{\lambda_r\pi_r}\sum_{k=0}^{r}\pi_k(w_{r+1}-w_k) - \frac{1}{2}\sum_{r=0}^{n-1}(w_{r+1}-w_r)^2 \tag{43}$$
$$= v_{n-1} - \frac{1}{2}\sum_{r=0}^{n-1}(w_{r+1}-w_r)^2$$

Since w_i is increasing in i

$$v_{n-1} \geq \sum_{r=0}^{n-1}\frac{1}{\lambda_r\pi_r}\sum_{k=0}^{r}\pi_k(w_{r+1}-w_r) = \sum_{r=0}^{n-1}(w_{r+1}-w_r)^2$$

and hence

$$\frac{1}{2}\operatorname{Var}(z_n) \geq \frac{1}{2}\sum_{r=0}^{n-1}(w_{r+1}-w_r)^2 .$$

If the series $\sum_0^\infty (w_{r+1}-w_r)^2$ is divergent then $v_{n-1}\to\infty$ and $\frac{1}{2}\operatorname{Var}(z_n)\to\infty$, but if the series is convergent then $v_{n-1}-\frac{1}{2}\operatorname{Var}(z_n)$ is bounded. In any case $\{v_n\}$ and $\{\operatorname{Var}(z_n)\}$ either both converge or both diverge.

It is possible for w_n to increase to infinity while at the same time v_n stays uniformly bounded. For example, let

$$\pi_r = \frac{e^{e^r}}{r} \text{ and } \frac{1}{\lambda_r\pi_r} = \frac{1}{e^{e^r}} \qquad \text{for } r \geq 1.$$

A straightforward calculation shows that

$$w_{n+1} \sim \sum_{r=0}^{n}\frac{1}{e^{e^r}}\sum_{k=1}^{r}\cdot\frac{e^{e^k}}{k} = \sum_{r=0}^{n}\frac{1}{r} + \text{a convergent series}$$
$$\sim \log n + c .$$

Also

$$v_n \sim \sum_{k=1}^{n-1}\frac{1}{e^{e^k}}\sum_{l=1}^{k}\frac{e^{e^l}}{l}(w_{k+1}-w_l) .$$

The inner sum grows like its largest term and we have

$$v_n = \sum_{k=1}^{n-1}\frac{1}{k}[\log(k+1)-\log k] + a \text{ convergent sequence}$$

which clearly exhibits v_n as uniformly bounded.

A class of examples in which $v_n\to\infty$ can be constructed as follows. Suppose, π_n and $\dfrac{1}{\lambda_n\pi_n}$ obey the asymptotic relations

$$\pi_n \sim n^{\alpha-1}L(n)\ (\alpha \neq 0) \text{ and } \frac{1}{\lambda_n\pi_n} \sim n^{\beta}L^*(n)$$

where $L(n)$ and $L^*(n)$ are slowly oscillating sequences ($L(n)$ is said to be slowly oscillating if for every $c > 1$, $\frac{L([cn])}{L(n)} \to 1$), $\Sigma \frac{1}{\lambda_n\pi_n} < \infty$ and w_n tends to infinity. Under these conditions we show that v_n tends to infinity. In fact

$$w_n \sim \sum_{r=1}^{n} r^{\beta+\alpha}\tilde{L}(r) \sim n^{\alpha+\beta+1}\tilde{L}(n)$$

where $\tilde{L}(n) = c \cdot L(n)L^*(n)$, ($c$ is a constant) and provided $\alpha+\beta+1>0$. (Similar conclusions hold even in the cases $\alpha = 0$ and $\alpha + \beta = -1$ involving iterates of $L(n)$.)
We next observe that

$$(44)\qquad \begin{aligned} v_n &\geq A\sum_{r=0}^{n} r^{\beta}L^*(r)\sum_{k=0}^{r} k^{\alpha-1}L(k)[w_{r+1} - w_k] \\ &\geq A'\sum_{r=0}^{n} r^{\beta}L^*(r)\sum_{k=1}^{r/2} k^{\alpha-1}L(k) \end{aligned}$$

where A and A' stand for fixed constants. The estimate in (44) is valid since w_r grows like $r^{\alpha+\beta+1}\tilde{L}(r)$. Finally,

$$\begin{aligned} v_n &\geq A''\sum_{r=0}^{n} r^{\beta}L^*(r)r^{\alpha}L\left(\frac{1}{2}r\right) \\ &\geq A'''n^{\alpha+\beta+1}L(n) \end{aligned}$$

and the proof is finished.

Some other useful conditions that assure the validity of (40) are as follows: If the spectral measure ψ of the birth and death procsse (B) has either

(a) positive measure in every neighborhood of the origin, or if

(b) ψ has an infinite number of points of increase, contained in a bounded interval I, then v_n tends to infinity.

The proof of these statements depend on an alternative representation of the quantity Var z_n. To this effect, we observe that the Laplace transform of z_n can be factored in the form

$$(45)\qquad \varphi_n(s) = \frac{1}{Q_n(-s)} = \frac{1}{\prod_{i=1}^{n}\left(1 + \frac{s}{\alpha_{ni}}\right)}$$

where α_{ni} are the roots of Q_n (recall that the α_{ni} are real and positive). A direct calculation shows that

$$\operatorname{Var} z_n = \sum_{i=1}^{n} \frac{1}{\alpha_{ni}^2} \text{ and } w_n = \sum_{i=1}^{n} \frac{1}{\alpha_{ni}}\,.$$

In case (a), the first root α_{n1} tends to zero and hence Var z_n becomes unbounded. In case (b) as n increases the interval I must contain an unbounded number of roots α_{ni}, and therefore Var z_n is unbounded. Several notable applications may be recorded.

Queueing models, defined by the parameters $\lambda_n = \lambda$, $n \geq 0$, $\mu_n = \mu$, $n \geq k_0$, $\mu_0 = 0$, and k_0 a prescribed positive integer, have the property that coincidence is a certain event. In fact, for these examples case (b) applies (see [4]).

The situation of linear growth, birth and death processes, (*i.e.* $\lambda_n = \lambda n + a$, $n \geq 0$, $\mu_n = \mu n + b$, $n > 0$, $\mu_0 = 0$) with regard to the probability of coincidence is as follows. If $\mu = \lambda$, then coincidence is always certain (case (b) above). If $\mu > \lambda$ then the process is recurrent and coincidence is trivially certain. If $\mu < \lambda$ then the process is weakly transient and coincidence is not certain. This last assertion is proved as follows. The spectral measure is discrete with mass points located essentially at an arithmetic series. The roots of $Q_n(-s)$ for any n are separated by the mass points of ψ and hence always $\Sigma\left(\frac{1}{\alpha_{ni}}\right)^2 \leq k\Sigma\frac{1}{n^2} < C$.

We turn now to the proof of the theorem. The arguments are divided into a series of lemmas.

DEFINITION (Levy [7]). A series of independent random variables $x_1 + \cdots + x_n = s_n$ is *essentially divergent* if there exists no sequence of constants a_n such that $s_n - a_n$ converges almost surely to a finite random variable.

LEMMA 1. *If v_n is divergent then the series of independent random variables $t_0 + t_1 + \cdots + t_{k-1} = z_k$ is essentially divergent. (The meaning of t_r is as before.)*

Proof. Suppose we can find a sequence of constants a_n such that $z_n - a_n$ converges. In particular, its characteristic function

$$\frac{e^{ia_n\lambda}}{Q_n(-i\lambda)} \text{ converges for each real } \lambda$$

to a characteristic function $\varphi(\lambda)$. It follows that the corresponding symmetrized random variable with characteristic function

$$\frac{1}{|Q_n(-i\lambda)|^2} \text{ converges to } |\varphi(\lambda)|^2$$

for each real λ and uniformly in any finite interval. But, by virtue of (45) for $\lambda > 0$

$$|Q_n(-i\sqrt{\lambda})|^2 = \prod_{i=1}^{n}\left(1 + \frac{\lambda}{\alpha_{ni}^2}\right) \geq 1 + \lambda\sum_{i=1}^{n}\frac{1}{\alpha_{ni}^2}$$
$$= 1 + \lambda \operatorname{Var} z_n .$$

Hence, for $\lambda \neq 0$, $|Q_n(-i\lambda)|^2$ tends to infinity and $|\varphi(\lambda)|^2 \equiv 0$. Thus $\varphi(\lambda)$ is not a characteristic function as required. The contradiction implies that $z_n - a_n$ cannot converge for any sequence of constants and consequently z_n is essentially divergent as was to be shown.

COROLLARY 1. Suppose v_n is divergent and let t_i and t'_i represent independent observations of the first passage time from state i to state $i+1$. Then

$$t_0 + t_1 + \cdots + t_{k-1} - t'_0 - t'_1 - \cdots - t'_{k-1} = z_k - z'_k$$

is essentially divergent.

Proof. This is clear since the characteristic function of $z_k - z'_k - a_k$ ($a_k =$ a sequence of constants) is

$$\frac{e^{i\lambda a_k}}{|Q_k(-i\lambda)|^2}$$

which for real $\lambda \neq 0$ tends to zero as shown in the proof of Lemma 1.

LEMMA 2. *With the same notation as in Corollary* 1, *if* v_n *diverges then for every fixed* r

(46) $$\Pr \{[t_0 + t_1 + \cdots + t_k] - [t'_r + t'_{r+1} + \cdots + t'_k] < 0 \text{ i.o.}\} = 1$$

(i.o. is an abbreviation of infinitely often).

Proof. With r held fixed it will be sufficient to prove that

(47) $$\Pr \{t'_r + t'_{r+1} + \cdots + t'_k - t_r - t_{r+1} - \cdots - t_k > C \text{ i.o.}\} = 1$$

for every positive constant C. Indeed, the validity of (47) implies that for almost every value of $t_0 + t_1 + \cdots + t_{r-1}$

$$1 = \Pr \{t_0 + \cdots + t_k - (t'_r + \cdots + t'_k) < 0 \text{ i.o.} \mid t_0 + t_1 + \cdots + t_{r-1}\} .$$

Invoking the law of total probabilities leads immediately to the conclusion (46).

We devote ourselves now to the proof of relation (47). Since the series $(t'_r - t_r) + (t'_{r+1} - t_{r+1}) + \cdots + (t'_k - t_k) = T_k$ (the dependence of T_k on r is suppressed since we are keeping r fixed) is essentially divergent we may appeal to a theorem of P. Levy [7 p. 147] and deduce that if A_k is any sequence of constants

(48) $$\Pr \{T_k \geq A_k \text{ i.o.}\}$$

is either 0 or 1. We select for our purpose all $A_k = 0$. Since T_k constitute a series of symmetric random variables the value of the expression (48) is clearly 1. By virtue of a second theorem of P. Levy [7 p. 147],

$$\Pr\{T_k \geq C \text{ i.o.}\} = 1$$

for any constant C and the proof of the lemma is finished.

Proof of the Sufficiency of Theorem 3. Suppose for definiteness that particle labeled (i) starts in state 0 and particle labeled (ii) starts in state r, each independently subject to the same transition law. Let t_i and t'_i, for particles (i) and (ii) respectively, represent as previously the first passage time from state i to state $i+1$. Lemma 2 assures that with probability 1 there is a state k such that the particle labeled (i), having started at zero, reaches k for the first time earlier than the particle labeled (ii) whose initial state was r. Since the path functions are continuous, the two particles necessarily cross and coincidence is certain.

Necessity. The proof of necessity will likewise be written in the form of a series of lemmas.

LEMMA 3. *If v_n is bounded then*

$$\Pr\{t_0 + t_1 + \cdots + t_{k-1} - t'_r - t'_{r+1} - \cdots - t'_k > 0 \quad \textit{for all } k \geq r\} > 0 . \tag{49}$$

Proof. Consider $T_k = (t'_r - t_r) + \cdots + (t'_k - t_k)$, $k = r, r+1, \cdots$, which is a partial sum composed of independent symmetrically distributed random variables. The hypothesis (see (43)) means that the variance of T_k is uniformly bounded. Therefore, invoking the three series theorem (because $t'_i - t_i$ are symmetric only the convergence of the series formed by the variances of the successive terms has to be verified), we may conclude that T_k converges almost surely to a finite valued random

Let t^* denote the limit of T_k. Take any value C such that

$$\Pr\{|t^*| < C\} > 0 .$$

Since T_k converges almost surely to t^* there is a k_0 such that

$$\Pr\{|T_k| < C \text{ for all } k \geq k_0\} > 0 .$$

Making C even larger (say C') if necessary we can assure

$$\Pr\{|T_k| < C' \text{ for all } k = r, r+1, \cdots, \} > 0 . \tag{50}$$

Consider now the random variable $t_0 + t_1 + \cdots + t_{r-1}$ which is independent of all T_k, $k \geq r$. Since t_0 is exponentially distributed it follows that

$$\Pr\{t_0 + t_1 + \cdots + t_{r-1} > C'\} > 0 \tag{51}$$

for any C' sufficiently large. Combining (50) and (51) yields the estimate

$$\begin{aligned} &\Pr\{t_0 + t_1 + \cdots + t_{r-1} - T_k > 0 \text{ for all } k \geq r\} \\ &\geq \Pr\{t_0 + \cdots + t_{r-1} > C'\} \Pr\{T_k < C' \text{ for all } k \geq r\} > 0 \end{aligned}$$

for an appropiate positive constant C.

This means that with positive probability a particle starting at zero never reaches a state $k \geq r + 1$ for the first time at an earlier time then a particle beginning in state r. The proof of the lemma is finished.

LEMMA 4. *If coincidence is a certain event when the particles have a prescribed pair of initial states r, s ($r < s$) then coincidence is a certain event for any pair of initial states.*

Proof. This is a direct consequence of the fact that with positive probability any pair of state i, j ($i < j$) can be attained starting from the initial states r and s without the occurrence of coincidence.

Consequently, if there exists positive probability of no coincidence starting from i and j, respectively then the same is true for r and s contrary to the hypothesis.

LEMMA 5. *Let coincidence be a certain event. Suppose the initial states of the two particles* (i) *and* (ii), *respectively are i_0 and $j_0 > i_0$. Then the event that particle* (ii) *reaches every state k ($k \geq k_0$) for the first time ahead of particle* (i) *has probability zero.*

Proof. We shall prove the lemma by producing an infinite sequence of states $k_1 < k_2 < \cdots$ with the following properties (called A). If the initial states of the particles (i) and (ii) are any pair r and s where $r < s$ and $s \leq k_i$ then the probability exceeds 1/4 that particle (i) will reach state k_{i+1} ahead of particle (ii).

Let us suppose statement (A) is established and now show how to finish the proof of the lemma. To this end, we have

$$\begin{aligned} &\Pr\{\text{(ii) reaches state } k \text{ prior to (i) for all } k \geq k_0\} \\ &\leq \Pr\{\text{(ii) reaches state } k_i \text{ prior to (i) for all } k_i \geq k_0\} \\ &\leq \prod_{k_i > k_0} (1 - \Pr\{\text{(i) reaches state } k_{i+1} \text{ prior to (ii)} \mid \text{(ii) reaches state} \\ &\qquad k_i \text{ prior to (i))}\} . \end{aligned}$$

The infinite product is zero since on account of statement (A) infinitely many factors are $\leq 3/4$.

It remains to prove statement (A).

Suppose we have already constructed $k_1, k_2, \cdots, k_i$. Since coincidence is a certain event regardless of the pair of initial states r and s, ($r < s$

and $s \leq k_i$) there exists a time value t_0 so that with probability $\geq 1 - \varepsilon$ coincidence occurs sometime earlier then t_0. The value of t_0 may be determined for each pair of initial states r and s. However, since there are only a finite number of possibilities r, s $(r < s)$ where $s \leq k_i$ we can choose t_0 large enough so that the same value of t_0 applies for any of these pairs of starting states. By further reducing to a subset of paths of probability $\geq 1 - 2\varepsilon$ (ε can be specified in advance as small as desired) we can determine a state $k_{i+1} > k_i$ which is not entered by either particle in the time duration $(0, t_0)$. The existence of k_{i+1} is guaranteed since the hypothesis of the lemma postulates that coincidence is certain and hence the process cannot be strongly transient. Restricting consideration to this set of paths we note that at the first instance of coincidence the two particles are indistinguishable and hence, with probability 1/2, particle (i) will enter state k_{i+1} ahead of particle (ii). Let E_i denote the event that (i) reaches state k_{i+1} ahead of (ii) when the initial states respectively are any pair r and s, $s \leq k_i$.

The above argument establishes that $\Pr\{E_i\} \geq (1 - 2\varepsilon)/2 > 1/4$ and the proof is hereby complete.

Proof of Necessity. This is immediate by comparing Lemmas 3 and 5.

The problem of computing the probability of coincidence for the case when v_n is bounded remains open.

We close with some observations regarding the problem of determining criteria which guarantee finite expected time for coincidence. First it is evident that for an ergodic birth and death process the expected time until coincidence is finite. To decide when the event of coincidence has a finite expected time is in general an open question.

The following two examples are of some interest. In the case of the linear growth processes associated with the Laguerre polynomials, we were able to determine a double generating function for the explicit distribution of the coincidence time (33). Here, it is easy to show by direct calculation that the expected coincidence time is infinite.

We shall now prove that for the recurrent null or transient queueing model (labeled B) the expected coincidence time is infinite. For definiteness $\lambda_n = \lambda$, $n \geq 0$ and $\mu_n = \mu$, $n \geq 1$, $\mu_0 = 0$.

We consider for the situation of two particles starting in states i and j, $0 < i < j$, the following induced random walk W whose state space is composed of the non-negative integers. We say that W is in state r if $j - i = r$. Transitions in W are engendered whenever one of the particles of process B changes its state. Explicitly a transition of W occurs from state r to $r - 1$ if and only if after the first change the state labels of the two particles, undergoing the process B, are either $(i + 1, j)$ or $(i, j - 1)$. A movement from r to $r + 1$ occurs in the contrary case. The motion on W, thus induced by the birth and death process will be understood to apply only when $i > 0$. The homogeneity of the queueing model implies that the changes engendered in W are independent of the specific states occupied by the two particles of the process B and only depend on their distance $(j - i)$ apart provided $i > 0$. Hence

$$\Pr_W\{r \to r - 1\} = \Pr_W\{r \to r + 1\} = 1/2 \text{ for } r > 0 .$$

It is well known that for this random walk the time until first passage into the state 0 from any non-zero initial state has an infinite expected value [3]. Moreover, first passage into 0 obviously corresponds to the event of coincidence for the original birth and death process. There is one slight complication in the above argument arising from the fact that when one of the particles of process B starting at i reaches zero, the transition probabilities of the induced random walk do not agree with the probabilities of the changes in distance between the particles. This is due to the reflecting character of state zero, *i.e.* when one of the particle of its process is in state 0 then this particle can only move to state 1. We will show that this complication is of no consequence in deciding whether coincidence in B occurs with finite expected time.

Let the particles begin in states i and $j, (i < j)$. Since coincidence is certain let us consider all those paths E where coincidence occurs without either particle ever reaching zero. Conditioned in this way the induced random walk describes the changes of the "distance" (number of states separating the two particles) until coincidence. But, for the random walk W the expected number of transitions for the first passage into zero is infinite. Since the expected time between transitions for the birth and death process is $1/(\lambda + \mu)$, the expected time until coincidence averaged over the paths of E is infinite. Next, let F denote the set of paths in the process B where the particle, starting in state $i < j$, reaches state zero before coincidence. Since the process B is null recurrent or transient, again the expected time length of the paths of F is infinite. Hence, under either circumstance the expected time until coincidence is infinite.

The above argument may be extended to prove that if a birth and death process is null recurrent of transient with certain coincidence, then the expected time of coincidence is infinite provided $\sum_{i=1}^{\infty} 1/(\lambda_i + \mu_i) = \infty$, and

$$1 - p_n = q_n = \max_{i>0} \frac{\lambda_i + \mu_{i+n}}{\lambda_i + \mu + \lambda_{i+n} + \mu_{i+n}}, \qquad p_0 = 1$$

are the transition parameters of a recurrent null or transient random walk W on the integers (*i.e.* $\Pr_W \{n \to n + 1\} = p_n$).

On the other hand, the expected coincidence time is finite whenever

$$1 - p_n = q_n = \min_{i>0} \frac{\lambda_i + \mu_{i+n}}{\lambda_i + \mu_i + \lambda_{i+n} + \mu_{i+n}}$$

describes an ergodic random walk W and $\max_j 1/(\lambda_j + \mu_j)$ is bounded.

REFERENCES

1. K. L. Chung, *Foundations of the theory of continuous parameter Markoff chains*, Proc. Third. Berkeley Symposium, 2, (1956), 29-40.
2. S. Karlin and J. L. McGregor, *The differential equations of birth and death processes and the Stieltjes moment problem*, Trans. Amer. Math. Soc., **86** (1957), 489-546.
3. Samuel Karlin and J. L. McGregor, *The classification of birth and death processes*, Trans. Amer. Math. Soc., **86** (1957), 366–401.
4. ———, *Many server queueing processes with Poisson input and exponential service times*, Pacific J. Math., **8** (1958), 87–118.
5. ———, *Linear growth, birth and death processes*, Journal of Math. and Mechanics, **7** (1958), 643-662.
6. ———, *A characterization of the birth and death process*, Proc. Nat. Acad. of Sciences **45** (1959), 375-379.
7. P. Levy, *Théorie de l'addition des variables aléatorires*, Paris, Gauthier-Villars, 1937.
8. G. Pólya and G. Szegö, *Aufgaben und lehrsätze aus der analysis*, I, Berlin, 1925.
9. G. Szegö, *Orthogonal polynomials*, Amer. Math. Soc. Colloquium Publications, 23 (1939).

STANFORD UNIVERSITY

COINCIDENCE PROBABILITIES

SAMUEL KARLIN AND JAMES MCGREGOR

1. Introduction. It was shown in [14] that if $P(t) = (P_{ij}(t))$ is the transition probability matrix of a birth and death process, then the determinants

$$(1) \qquad P\left(t\,;\begin{matrix} i_1 \cdots i_n \\ j_1 \cdots j_n \end{matrix}\right) = \begin{vmatrix} P_{i_1 j_1}(t) \cdots P_{i_1 j_n}(t) \\ \vdots \qquad\qquad \vdots \\ P_{i_n j_1}(t) \cdots P_{i_n j_n}(t) \end{vmatrix}$$

where $i_1 < i_2 < \cdots < i_n$ and $j_1 < j_2 < \cdots < j_n$ are strictly positive when $t > 0$. In this paper it is shown that these determinants have an interesting probabilistic significance.

(A) *Suppose that n labelled particles start out in states $i_1, \cdots, i_n$ and execute the process simultaneously and independently. Then the determinant* (1) *is equal to the probability that at time t the particles will be found in states $j_1, \cdots, j_n$ respectively without any two of them ever having been coincident* (simultaneously in the same state) *in the intervening time.*

From this statement it follows that the determinant is non-negative, and as will be seen strict positivity can be deduced from natural hypotheses, for example if $P_{i_\alpha j_\alpha}(t) > 0$ for $\alpha = 1, \cdots, n$ and every $t > 0$.

The truth of the above statement rests chiefly on the facts that the process is *one-dimensional*—its state space is linearly ordered, and that the path functions of the process are *everywhere "continuous"*. Of course the path functions are discontinuous in the ordinary sense but the discontinuities are only of magnitude one. Thus when a transition occurs the diffusing particle moves from a given state only into one of the two neighboring states, and even if the particle goes off to infinity in a finite time it either remains there or else it returns in a continuous way and does not suddenly reappear in one of the finite states. These two properties of one-dimensionality and "continuity" have the effect that when several particles execute the process simultaneously and independently, a change in the order of the particles cannot occur unless a coincidence first takes place. (The states are all stable so that with probability one a transition involves only one of the particles.)

It is also important for our results that the processes involved have the strong Markoff property of Hunt [10], [11], (see also [19]). However it is a consequence of theorems of Chung [3] that any continuous time

Received December 18, 1958. This work was supported in part by an Offiice of Naval Research Contract at Stanford University.

parameter Markoff chain whose states are all stable has the strong Markoff property.

There exist processes of birth-and-death type whose path functions may have discontinuities at infinity. Such processes have been described in some detail by Feller. Although the above result (A) does not apply to these processes they fall within a more general class of processes which we discuss next.

We consider a stationary Markoff process whose state space is a set of integers and whose states are all stable. Let $(P_{ij}(t))$ be the transition probability matrix. Then

(B) *Suppose that n labelled particles start in states $i_1, \cdots, i_n$ and execute the process simultaneouly and independently. For each permutation σ of $1, \cdots, n$ let A_σ denote the event that at time t the particles are in states $j_{\sigma(1)}, \cdots, j_{\sigma(n)}$ respectively, without any two of them ever having been coincident in the intervening time. Then*

$$P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right) = \sum_\sigma (\operatorname{sign} \sigma) \Pr\{A_\sigma\}$$

where the sum runs over all permutations of $1, \cdots, n$ and sign $\sigma = 1$ or -1 *according as σ is an even or an odd permutation.*

The first stated result is seen to be a special case of this one. For if the path functions are "continuous" and $i_1 < \cdots < i_n, j_1 < \cdots < j_n$ then $\Pr\{A_\sigma\}$ is zero except when σ is the identity permutation. There is one other case in which the general formula permits an interesting simplification, namely when the process is a local cyclic process. By this we mean that the states may be viewed as $N+1$ points $0, 1, \cdots, N$ on a circle and transitions occur only between neighboring states, 1 and N being neighbors of zero and $N-1$ and 0 neighbors of N. We take $0 \le i_1 < \cdots < i_n \le N$ and $0 \le j_1 < \cdots < j_n \le N$ and then $\Pr\{A_\sigma\}$ is zero unless σ is a cyclic permutation. Since the cyclic permutations of an odd number of objects are all even permutations we have in this situation

$$P\left(t\,;\begin{matrix} i_1, \cdots, i_n \\ j_1, \cdots, j_n \end{matrix}\right) = \sum_{\text{cyclic } \sigma} \Pr\{A_\sigma\}, \qquad n \text{ odd}\,. \tag{3}$$

This determinant is therefore non-negative.

Analogous results hold for one dimensional diffusion processes. Let $P(t, x, E)$ be the transition probability function of a stationary process whose state space is an interval on the extended real line. It will be assumed that the process has the strong Markoff property and that its path functions are *continuous* everywhere. Given two Borel sets E, F the inequality $E < F$ will denote that $x < y$ for every $x \in E, y \in F$. We take n states $x_1 < x_2 < \cdots < x_n$ and n Borel sets $E_1 < E_2 < \cdots < E_n$

and form the determinant

$$(4)\qquad P\left(t;\ \begin{matrix} x_1, \cdots, x_n \\ E_1, \cdots, E_n \end{matrix}\right) = \begin{vmatrix} P(t, x_1, E_1) \cdots P(t, x_1, E_n) \\ \vdots \\ P(t, x_n, E_1) \cdots P(t, x_n, E_n) \end{vmatrix}$$

(C) *Suppose that n labelled particles start in states $x_1, \cdots, x_n$ and execute the process simultaneously and independently. Then the determinant* (4) *is equal to the probability that at time t the particles will be found in the sets $E_1, \cdots, E_n$ respectively without any two of them ever having been coincident in the intervening time.*

Next consider a stationary strong Markoff process whose state space is a metric space and whose path functions are continuous on the right. We take n states $x_1, \cdots, x_n$ and n Borel sets $E_1, \cdots, E_n$ and again form the determinant (4).

(D) *Suppose that n labelled particles start in the states $x_1, \cdots, x_n$ and execute the process simultaneously and independently. For each permutation σ of $1, 2, \cdots, n$ let A_σ denote the event that at time t the particles are in the states $E_{\sigma_1}, \cdots, E_{\sigma_n}$ respectively without any two of them ever having been coincident in the intervening time. Then*

$$(5)\qquad P\left(t;\ \begin{matrix} x_1, \cdots, x_n \\ E_1, \cdots, E_n \end{matrix}\right) = \sum_\sigma (\operatorname{sign} \sigma) \Pr\{A_\sigma\}$$

where the sum runs over all permutations σ.

The last result contains all of the preceding ones as special cases. It has another interesting special case, namely when the state space is a circle and the path functions are continuous.

There is a mapping $\theta \to e^{i\theta} = x$ of the closed interval $0 \leq \theta \leq 2\pi$ onto the circle. Given n Boral sets $E_1, \cdots, E_n$ on the circle we say $E_1 < \cdots < E_n$ if there are n Borel sets $E'_1 < \cdots < E'_n$ in the interval $(0, 2\pi]$ or $[0, 2\pi)$ which are mapped onto $E_1, \cdots, E_n$ respectively by the above mapping. Specializing the sets to be one point sets gives the meaning for $x_1 < \cdots < x_n$ when $x_1, \cdots, x_n$ are n points on the circle.

Now let $P(t, x, E)$ be the transition probability function of a strong Markoff process on the circle with *continuous* path functions. Because of the continuity of paths a change in the cyclic order of several diffusing particles on the circle cannot occur unless a coincidence first takes place. Thus the terms in (5) corresponding to non-cyclic permutations σ will all be zero. Finally we take advantage of the fact that the cyclic permutations of an *odd* number of objects are all even permutations, and obtain the following.

(E) *Suppose $x_1 < \cdots < x_n$, $E_1 < \cdots < E_n$ and n labelled parti-*

cles start at $x_1, \cdots, x_n$ respectively and execute the process simultaneously and independently. If n is odd and A_σ is defined as before then

$$P\left(t;\ \begin{matrix} x_1, \cdots, x_n \\ E_1, \cdots, E_n \end{matrix}\right) = \sum_{\text{cyclic } \sigma} \Pr\{A_\sigma\} \tag{6}$$

where the sum runs over all cyclic permutations.

Similar but more complicated results are valid in still more general situations. For example we restrict our discussion to stationary processes although both the methods and the results can be extended to non-stationary processes. A generalization of another type which has interesting applications is obtained when the n particles execute *different processes.*

Let $P_\alpha(t, x, E), \alpha = 1, \cdots, n$ be transition probability functions of n strong Markoff process on the real line with continuous path functions. Choose n states $x_1 < \cdots < x_n$ and n Borel sets $E_1 < \cdots < E_n$ and form the determinant

$$\det P_\alpha(t, x_\alpha, E_\beta)\ . \tag{7}$$

If n labelled particles start in states $x_1, \cdots, x_n$ respectively, and execute the processes simultaneously and independently, the ith particle executing the ith process, then the determinant (7) is the probability that at time t the particles will be found in the sets $E_1, \cdots, E_n$ respectively, without any two of them ever having been coincident in the intervening time.

The formal proofs of formulas (5) and (6) and of the interpretation of $P\left(t,\ \begin{matrix} x_1, x_2 \cdots, x_n \\ E_1, E_2 \cdots, E_n \end{matrix}\right)$ are elaborated in § 5. For this purpose the relevant preliminaries and definitions concerning Markoff processes are summarized in § 4.

In § 6 we offer some observations on the problem of determining when the strong Markoff property applies to direct products of processes. In this connection we direct attention to those aspects of this problem relevant to our analysis of the main theorem of § 5.

Section 2 contains a brief heuristic proof of (C) in the situation of two particles. This is inserted in order to motivate the formal proof of § 5. Section 3 discusses the connections of the concept of total positivity, to statements (A) - (E).

Total positivity is significant in relation to the theory of vibrations of mechanical systems [8], the method of inversion of convolution transforms [9], and the techniques of mathematical economics [13]. In this paper total positivity is shown to be also important in describing the structure of one dimensional strong Markoff processes whose path functions are continuous. In a vague sense the most general totally positive kernel can be built from convolutions of stochastic processes whose path functions are continuous. In principle, the representation desired is similar to the representation formula which applies to Pólya frequency functions discovered by Schoenberg [20]. A detailed discussion of this idea will be published separately. In this connection we mention that Loewner has completely analyzed the generation of totally positive matrices from infinitesimal elements [18].

In § 7 we investigate conditions which insure that the determinant (4) is strictly positive. We find that this is the case if $P(t, x, E) > 0$ whenever $t > 0$, E is any open set and $P(t, x, E)$ represents the transition probability function of a strong Markoff process on the real line with continuous path functions.

The following converse proposition is of interest. Suppose the transition function $P(t, x, E)$ of a Markoff process has the property that all determinants of the form (4) are non-negative. Does there exist a realization of the process such that almost all path functions are continuous? This is true with some mild further restrictions. In § 8 with the aid of a theorem of Ray [19] we are able to establish a partial converse based on a restriction about the local character of $P(t, x, E)$. It will be recognized that most cases of Markoff processes obey this requirement.

In § 9 we characterize the most general one dimensional spatially homogeneous process whose transition kernel is totally positive.

The final section presents a series of examples of totally positive kernels derived from Markoff processes with continuous path functions.

2. A heuristic argument. In this section we give a non-rigorous outline of the method of proof for the case of two particles. Let $P(t, x, E)$ be the transition probability function of a stationary Markoff process on the real line. Suppose that two distinguishable particles start at x_1 and $x_2 > x_1$ and let $E_1 < E_2$ be two Borel sets. The determinant

$$P\left(t\,;\ \begin{matrix} x_1 & x_2 \\ E_1 & E_2 \end{matrix}\right) = P(t, x_1, E_1)P(t, x_2, E_2) - P(t, x_1, E_2)P(t, x_2, E_1)$$

is equal to $\Pr\{A'_1\} - \Pr\{A'_2\}$ where A'_1 is the event that at time t the first particle is in E_1, the second in E_2 and A'_2 is the event that at time t the first particle is in E_2, the second in E_1. Each event A'_i, regarded as a collection of paths, may be split up into two disjoint sets $A_i + A''_i$ where A_i consists of all the paths in A'_i for which no coincidence occurs before time t and A''_i consists of the paths in A'_i with at least one coincidence before time t. We assume the paths are sufficiently smooth so that for each path in A''_1 and A''_2 there is a first coincidence time. This will certainly be the case if all paths are continuous on the right. Choose a path in A''_1 and at the time of first coincidence interchange the labels of the two particles. This converts the given path into a path in A_2'' and the resulting map of A_1'' into A_2'' is clearly one-to-one and onto. Because of the Markoff property and because the particles act independently it is plausible that this map is measure preserving so that

$$\Pr\{A_1''\} = \Pr\{A_2''\}$$

and granting this it follows that

$$P\left(t\,;\ \begin{matrix} x_1, & x_2 \\ E_1 & E_2 \end{matrix}\right) = \Pr\{A_1'\} - \Pr\{A_2'\}$$
$$= \Pr\{A_1\} - \Pr\{A_2\}\ ,$$

which is the general form of the result. If the path functions are all continuous then $\Pr\{A_2\} = 0$ and the formula becomes

$$P\left(t\,;\ \begin{matrix} x_1, & x_2 \\ E_1 & E_2 \end{matrix}\right) = \Pr\{A_1\}\ .$$

3. **Total positivity.** A matrix is called (strictly) *totally positive* if all of its minors of all orders are (strictly positive) non-negative. Such matrices and their continuous analogues the totally positive kernels occur in a variety of applications and have been studied by numerous authors. A lucid outline of the theory together with an extensive bibliography has been given by Schoenberg [21], Krein and Gantmacher [8]. Our results indicate the existence of large natural classes of semi-groups of totally positive matrices and totally positive kernels. One simply takes the transition probability function of a one dimensional diffusion process with continuous path functions. A number of interesting examples are given in § 10.

Conversely the total positivity of the transition function may be used to draw conclusions regarding continuity of the path functions. A program along these lines has already been carried out by the authors for the case of birth and death processes [12]. (see also § 8.)

Our attention was first drawn to total positivity in connection with diffusion processes by unpublished results of C. Loewner who showed that the fundamental solution of

$$\frac{\partial u}{\partial t} = a\frac{\partial^2 u}{\partial x^2} + b\frac{\partial u}{\partial x}$$

on a finite interval with smooth a and b and classical boundary conditions, is totally positive.

4. **Definitions.** As indicated in the introduction we are chiefly concerned with processes on the integers, the real line, or the circle. In

order to deal with all cases at once it is convenient to discuss certain results for a more general process whose state space is a metric space X.

Let X be a metric space, $\mathfrak{B}$ the Borel field generated by the open sets of X, and $\mathfrak{B}'$ the Borel ring generated by the finite intervals on $0 \leq t < \infty$. Suppose there is given a set Ω called the sample space and an X-valued function $x(t, \omega), 0 \leq t < \infty, \omega \in \Omega$. Let $\mathfrak{M}$ be the Borel field of subsets of Ω generated by the sets of the form $\{\omega; x(t, \omega) \in E\}$ where $t \geq 0$ and $E \in \mathfrak{B}$. Suppose that for each $x \in X$ there is given a probability measure P_x on X such that $P_x\{\omega; x(0, \omega) = x\} = 1$. Then the function $x(t, \omega)$ is called a stochastic process on X with sample space Ω and distributions $\{P_x\}$.

The stochastic process is said to have right continuous path functions if for every fixed ω the function $x(\cdot, \omega)$ is right continuous on $0 \leq t < \infty$.

Let $\mathscr{M}_t$ denote the Borel field generated by all sets $\{\omega; x(s, \omega) \in E\}$ where $E \in \mathfrak{B}$ and $0 \leq s \leq t$. Conditional probabilities relative to $\mathfrak{M}_t$ will be denoted by $P_x\{\cdots | x(s), s \leq t\}$. The stochastic process is called a stationary Markoff process if for every fixed t

$$P_x\{x(t_i + t, \omega) \in E_i, i = 1, \cdots, n \mid x(s), s \leq t\}$$
$$= P_{x(t,\omega)}\{x(t_i, \omega) \in E_i, i = 1, \cdots, n\}$$

with probability one when $0 < t_1 < \cdots < t_n$ and $E_1, \cdots, E_n \in \mathfrak{B}$.

We will be concerned only with stationary Markoff processes in X with right continuous path functions. It will always be assumed that the function

$$P(t, x, E) = P_x\{x(t, \omega) \in E\}$$

is measurable relative to $\mathfrak{B}' \otimes \mathfrak{B}$. This function satisfies the Chapman-Kolmogoroff equation:

$$P(t + s, x, E) = \int_x P(t, x, dy) P(s, y, E) .$$

Let F be a closed set in X. The time of first hitting F is defined as

$$\tau_F(\omega) = \inf \{t; x(t, \omega) \in F\}$$

where the inf of the void set is taken to be $+\infty$. The place of first hitting F is defined, if $\tau_F(\omega) < \infty$, as

$$\xi_F(\omega) = x(\tau_F(\omega), \omega) .$$

The Markoff process will be called a *strong Markoff process* if for any closed set F we have the first passage relation

$$P_x\{x(t, \omega) \in E\} = P_x\{x(t, \omega) \in E, \tau_F(\omega) > t\} + \int_0^t \int_F P_y\{x(t-s, \omega) \in E\} P_x \begin{Bmatrix} \tau_F(\omega) \in ds \\ \xi_F(\omega) \in dy \end{Bmatrix} .$$

In this relation it is implicitly assumed that the sets $\{\omega; \tau(\omega) < t\}$ and $\{\omega; \tau(\omega) < t, \xi(\omega) \in H\}$ where H is a closed subset of F, are $\mathfrak{W}_t$ measurable for each t. A discussion of the validity of these assumptions made in § 6. It is there shown that under very slight conditions on the transition function the assumption holds.

It seems reasonable to believe that the direct product of a finite number of strong Markoff processes is again a strong Markoff process. At the present time we are not able to prove that this is generally true, although in the proof of the main theorem we assume this result. On the other hand proofs can be given which cover the vast majority of the special cases of interest. As noted above it follows from theorems of Chung that the strong Markoff property is preserved under direct products for processes with countably many states all of which are stable. This includes the birth and death case. In § 6 we give a proof for direct products of a one dimensional diffusion process whose transition probability function $P(t, x, E)$ is jointly continuous in t and x. This covers the case when $P(t, x, E)$ comes from a diffusion equation

$$\frac{\partial u}{\partial t} = a(x)\frac{\partial^2 u}{\partial x^2} + b(x)\frac{\partial u}{\partial x}$$

with $a(x)$, $b(x)$ continuous and $a(x) > 0$. References to other theorems of this kind are given in § 6.

Let X_i, $i = 1, \cdots, n$ be metric spaces and for each i let $x_i(t, \omega_i)$ be a stationary Markoff process in X_i with sample space Ω_i and distributions $\{P^{(i)}_{x_i}\}$. We form the product space $\overline{X} = X_1 \otimes \cdots \otimes X_n$ in which the generic point is an n-tuple $\bar{x} = (x_1, \cdots, x_n)$ with $x_i \in X_i$. The space $\overline{X}$ with the distance $\rho(\bar{x}, \bar{y}) = \Sigma\rho(x_i, y_i)$ is a metric space. The vector valued function $\bar{x}(t, \bar{\omega}) = (x_1(t, \omega_1), \cdots, x_n(t, \omega_n))$ is a stationary Markoff process in $\overline{X}$ whose sample space is the direct product $\overline{\Omega}$ of the Ω_i and whose distributions are the direct product measures

$$\overline{P}_{\bar{x}} = \prod_i P^{(i)}_{x_i}, \qquad \bar{x} = (x_1, \cdots, x_n) .$$

$\bar{x}(t, \bar{\omega})$ is called the *direct product* of the given processes.

5. The main theorem. Let X be a metric space, and $x(t, \omega)$ a stationary strong Markoff process in X with right continuous sample functions, sample space Ω and distributions $\{P_x\}$. We form the direct products $\overline{X}$, $\overline{\Omega}$ of n copies of X and Ω respectively and the direct product

$\bar{x}(t, \bar{\omega})$ of n copies of the given process. We say this direct product process represents "n labelled particles executing the $x(t, \omega)$ process simultaneously and independently", and this is the sense in which that phrase is to be interpreted in statements (A)-(E) of the introduction. *We assume $\bar{x}(t, \bar{\omega})$ is a strong Markoff process* (see § 6).

The associated distributions are

$$P_{\bar{x}} = \prod_{i=1}^{n} P_{x_i}, \qquad \bar{x} = (x_1, \cdots, x_n) .$$

The set F of coincident states consists of the points $\bar{x} = (x_1, \cdots, x_n)$ with at least two of the x_i equal to one another. A permutation λ of the n letters $1, 2, \cdots, n$ is called a transposition if there are two letters $i < j$ such that $\lambda(i) = j$, $\lambda(j) = i$, and $\lambda(r) = r$ if $i \neq r \neq j$. In this case we use the notation $\lambda = (i, j)$. A coincident state $\bar{x} = (x_1, \cdots, x_n)$ is said to belong to the transposition $\lambda = (i, j)$, $i < j$ if $x_1, \cdots, x_{j-1}$ are all different but $x_i = x_j$. Thus every coincident state belongs to a unique transposition, and for a given λ the set of all coincident states belonging to λ will be denoted by $F(\lambda)$. The group of all $n!$ permutations of $1, 2, \cdots, n$ will be denoted by S and the set of all transpositions by Λ.

Given n Borel sets $E_1, \cdots, E_n$ in X and a permutation $\sigma \in S$, the direct product set

$$E_\sigma = E_{\sigma(1)} \otimes \cdots \otimes E_{\sigma(n)}$$

is a Borel set in $\bar{X}$. Let $A'_\sigma = \{\bar{\omega}; \bar{x}(t, \bar{\omega}) \in E_\sigma\}$ where $t > 0$ is fixed. Then if $\bar{x} = (x_1, \cdots, x_n)$

$$P\left(t;\ \begin{matrix} x_1, \cdots, x_n \\ E_1, \cdots, E_n \end{matrix}\right) = \sum_{\sigma \in S} (\operatorname{sign} \sigma) P_{\bar{x}}\{A'_\sigma\}$$

by definition of the determinant and of $P_{\bar{x}}$.

The time $\tau(\bar{\omega})$ of first coincidence is defined as the time of first hitting F:

$$\tau(\bar{\omega}) = \tau_F(\bar{\omega}) = \inf\{t;\ x(t, \bar{\omega}) \in F\} .$$

The place of first coincidence is $\xi(\bar{\omega}) = \bar{x}(\tau(\bar{\omega}), \bar{\omega})$. Our main result can now be stated very simply as follows.

THEOREM 1. *The sets*

$$A_\sigma = \{\omega;\ \omega \in A'_\sigma, \tau(\bar{\omega}) > t\}$$

are all measurable and

$$P\left(t;\ \begin{matrix} x_1, \cdots, x_n \\ E_1, \cdots, E_n \end{matrix}\right) = \sum_{\sigma \in S} (\operatorname{sign} \sigma) P_{\bar{x}}\{A_\sigma\} .$$

Proof. Since τ is measurable the sets A_σ are also measurable. For each σ we apply the strong Markoff property to obtain

$$P_{\bar{x}}\{A'_\sigma\} = P_{\bar{x}}\{A_\sigma\} + \int_0^t d\Phi(s) \int_F P_{\bar{y}}\{\bar{x}(t-s, \bar{\omega}) \in \bar{E}_\sigma\}\mu(dy)$$

where

$$\Phi(s) = P_{\bar{x}}\{\tau(\bar{\omega}) \leq s\} ,$$
$$\mu(M) = P_{\bar{x}}\{\xi(\bar{\omega}) \in M \mid \tau(\bar{\omega}) = s\} .$$

Now F is the union of the disjoint Borel sets $F(\lambda)$, $\lambda \in \Lambda$, and if $\bar{y} \in F(\lambda)$ then $P_{\bar{y}}\{\bar{x}(t-s, \omega) \in \bar{E}_\sigma\} = P_{\bar{y}}\{\bar{x}(t-s, \bar{\omega}) \in \bar{E}_{\lambda\sigma}\}$. Hence

$$\begin{aligned}
&\sum_{\sigma \in S} (\text{sign}\,\sigma) [P_{\bar{x}}\{A'_\sigma\} - P_{\bar{x}}\{A_\sigma\}] \\
&\quad = \sum_{\sigma \in S} \sum_{\lambda \in \Lambda} (\text{sign}\,\sigma) \int_0^t d\Phi(s) \int_{F(\lambda)} P_{\bar{y}}\{\bar{x}(t-s, \bar{\omega}) \in \bar{E}_\sigma\}\mu(dy) \\
&\quad = \sum_{\sigma \in S} \sum_{\lambda \in \Lambda} - (\text{sign}\,\lambda\sigma) \int_0^t d\Phi(s) \int_{F(\lambda)} P_{\bar{y}}\{x(t-s, \bar{\omega}) \in \bar{E}_{\lambda\sigma}\}\mu(dy) \\
&\quad = - \sum_{\sigma \in S} (\text{sign}\,\sigma) [P_{\bar{x}}\{A'_\sigma\} - P_{\bar{x}}\{A_\sigma\}] .
\end{aligned}$$

This quantity is therefore zero and

$$\begin{aligned}
P\left(t;\ \begin{matrix} x_1, \cdots, x_n \\ E_1, \cdots, E_n \end{matrix}\right) &= \sum_{\sigma \in S} (\text{sign}\,\sigma) P_{\bar{x}}\{A'_\sigma\} \\
&= \sum_{\sigma \in S} (\text{sign}\,\sigma) P_{\bar{x}}\{A_\sigma\} .
\end{aligned}$$

The various assertions (A) - (D) of the introduction can be obtained by specializing the above theorem in the appropriate way.

6. Strong Markoff property for direct products. For the vast majority of one-dimensional diffusion processes which are met in applications one finds that the transition probability function $P(t, x, E)$ is jointly continuous in t and x. It will be shown that the direct product of n-copies of such a process has the strong Markoff property. The proof imitates the proof of a theorem of Dynkin and Jushkevich [7].

THEOREM 2. *Let $x(t, \omega)$ be a stationary Markoff process on the real line with continuous path functions and transition probability function $P(t, x, E)$ which is jointly continuous in t, x. Then the direct product $\bar{x}(t, \bar{\omega})$ of n copies of this process is a strong Markoff process.*

Proof. Let F be a closed set in the n-dimensional space, $\tau(\bar{\omega})$ the time of first hitting F for the direct product process, and $\xi(\bar{\omega})$ the place of first hitting F. The fact that $\tau(\omega)$ and $\xi(\bar{\omega})$ are measurable functions is a trivial consequence of the continuity of the path functions. With a given integer $m \geq 1$ let $\tau_m(\bar{\omega}) = k/m$, where k is the integer such that

$$\frac{k-1}{m} < \tau(\bar{\omega}) \leq \frac{k}{m},$$

and let $\xi_m(\bar{\omega}) = \bar{x}(\tau_m(\bar{\omega}), \bar{\omega})$. Then for any Borel set $\bar{E}$

$$P_{\bar{x}}(\bar{x}(t, \omega) \in \bar{E}\} = P_{\bar{x}}\{\bar{x}(t, \bar{\omega}) \in \bar{E}, \tau_m(\bar{\omega}) > t\} + \sum_{1 \leq k \leq mt} P_{\bar{x}}\left\{\bar{x}(t, \bar{\omega}) \in \bar{E}, \tau_m(\bar{\omega}) = \frac{k}{m}\right\}.$$

Let

$$A_k(\bar{\omega}) = \begin{cases} 1 & \text{if} \quad \tau_m(\bar{\omega}) = \dfrac{k}{m}, \\ 0 & \text{if} \quad \tau_m(\bar{\omega}) \neq \dfrac{k}{m}, \end{cases}$$

and

$$f(\bar{y}) = \begin{cases} 1 & \text{if} \quad \bar{y} \in \bar{E}, \\ 0 & \text{if} \quad \bar{y} \notin \bar{E}. \end{cases}$$

Then

$$\begin{aligned} P_{\bar{x}}\left\{\bar{x}(t, \bar{\omega}) \in \bar{E}, \ \tau_m(\bar{\omega}) = \frac{k}{m}\right\} &= E_{\bar{x}}\{A_k(\bar{\omega}) f(x(t, \omega))\} \\ &= E_{\bar{x}}\left\{E\left\{A_k(\bar{\omega}) f(\bar{x}(t, \bar{\omega})) \mid x(s), \, s \leq \frac{k}{m}\right\}\right\} \\ &= E_{\bar{x}}\left\{A_k(\bar{\omega}) E\left\{f(\bar{x}(t, \bar{\omega})) \mid x(s), \, s \leq \frac{k}{m}\right\}\right\} \\ &= E_{\bar{x}}\left\{A_k(\bar{\omega}) P_{\bar{x}(k/m, \bar{\omega})}\left\{\bar{x}\left(t - \frac{k}{m}, \bar{\omega}\right) \in \bar{E}\right\}\right\} \\ &= \int_{\bar{X}} P_{\bar{y}}\left\{\bar{x}\left(t - \frac{k}{m}, \bar{\omega}\right) \in \bar{E}\right\} P_{\bar{x}}\left\{\xi_m(\bar{\omega}) \in dy, \ \tau_m(\bar{\omega}) = \frac{k}{m}\right\} \end{aligned}$$

and hence we have the first passage relation for τ_m:

$$P_{\bar{x}}\{\bar{x}(t, \bar{\omega}) \in \bar{E}\} = P_{\bar{x}}\{\bar{x}(t, \bar{\omega}) \in \bar{E}, \tau_m(\bar{\omega}) > t\} + \int_0^t \int_{\bar{X}} P_{\bar{y}}\{\bar{x}(t - s, \bar{\omega}) \in \bar{E}\} P_{\bar{x}}\left\{\begin{matrix} \tau_m(\bar{\omega}) \in ds \\ \xi_m(\bar{\omega}) \in dy \end{matrix}\right\}.$$

For every $\bar{\omega}$ we have $\tau_m(\bar{\omega}) \geq \tau_{m+1}(\bar{\omega}) \downarrow \tau(\bar{\omega})$ and by continuity of path functions $\xi_m(\bar{\omega}) \to \xi(\bar{\omega})$ as $m \to \infty$. Hence $\tau_m(\bar{\omega})$, $\xi_m(\bar{\omega})$ converge in measure to $\tau(\bar{\omega})$, $\xi(\bar{\omega})$. Since $P_{\bar{y}}\{\bar{x}(t-s, \bar{\omega}) \in \bar{E}\}$ is jointly continuous in $\bar{y}$ and s and is bounded we may let $m \to \infty$ in the above formula and obtain the first passage relation for $\tau(\bar{\omega})$. This completes the proof.

The referee has brought to our attention the following stronger theorem of Blumenthal, [1, Theorem 1.1], which is slightly reworded here.

THEOREM. *If the process has right continuous path functions and if for every bounded continuous function f the function* $\int f(y)P(t, x, dy)$ *is continuous in x for each* $t > 0$, *then the process has the strong Markoff property.*

In this theorem the state space X is any metric space. Naturally this theorem requires more involved arguments than the above Theorem 2. Finally we mention that a very thorough discussion of the Markoff chain case has been given by Chung [4].

7. Strict total positivity. Let X be the non-negative integers and $x(t, \omega)$ a stationary strong Markoff process on X with all states stable and "continuous" path functions. If $P(t) = (P_{ij}(t))$ is the transition probability matrix of the process then it follows from assertion (A) that this matrix is totally positive. Let us call the process *a* strict process if $P_{ij}(t) > 0$ for every i, j and all $t > 0$. We will prove

THEOREM 3. *If the process is strict then its transition probability matrix is strictly totally positive for every* $t > 0$.

Proof. The proof is similar to the proof of a related theorem in [14], namely Theorem 20 on page 543. It is seen from the proof of that theorem that it is sufficient for our purposes to prove that if $i_1 < i_2 < \cdots < i_n$ then

$$P\left(t;\ \begin{matrix} i_1, \cdots, i_n \\ i_1, \cdots, i_n \end{matrix}\right) > 0$$

for every $t > 0$, that is the principal subdeterminants are strictly positive. However since

$$P\left(2t;\ \begin{matrix} i_1, \cdots, i_n \\ i_1, \cdots, i_n \end{matrix}\right) \geq \left[P\left(t;\ \begin{matrix} i_1, \cdots, i_n \\ i_1, \cdots, i_n \end{matrix}\right)\right]^2$$

it is enough to show that these determinants are strictly positive for sufficiently small $t > 0$. Because the path functions are right continuous, if $\{r_k\}$ is an ordering of the positive rationals, the set

$$\bigcup_{m=1}^{\infty} \bigcap_{r_k \leq 1/m} \{\omega ; x(r_k, \omega) = i \mid x(0, \omega) = i\}$$

has probability one. Hence for some $m = m(i) > 0$ there is a positive probability R_i that a path starting at i remains at i for at least up to time $1/m(i)$. Now if $0 < t < \max_{1 \leq k \leq n} 1/m(i_k)$ then we have

$$P\left(t ; \begin{matrix} i_1, \cdots, i_n \\ i_1, \cdots, i_n \end{matrix}\right) \geq \prod_{k=1}^{n} R_{i_k} > 0$$

and this proves the theorem.

Now let $x(t, \omega)$ be a stationary strong Markoff process on the real line with continuous path functions satisfying the hypothesis of Theorem 1. Let $P(t, x, E)$ be the transition probability function of the process. The process will be called *strict* if $P(t, x, E) > 0$ whenever $t > 0$ and E is any non-void open set. We will prove

THEOREM 4. *If the process is strict then its transition probability function is strictly totally positive in the sense that if $t>0$, $x_1 < \cdots < x_n$ and $E_1 < \cdots < E_n$ are non-void open sets then*

$$P\left(t ; \begin{matrix} x_1, \cdots, x_n \\ E_1, \cdots, E_n \end{matrix}\right) > 0 .$$

We begin with two lemmas in which the hypotheses of the theorem are assumed.

DEFINITION. If a, b are two points on the real line then

$$\tau_a(\omega) = \inf\{t ; x(t, \omega) = a\} ,$$
$$M(t, x, a) = P_x\{\tau_a(\omega) \leq t\} ,$$
$$M(t, x, a, b) = P_x\{\tau_a(\omega) \leq t, \tau_b(\omega) > t\} .$$

LEMMA. *If $a < x < b$ then $M(t, x, a, b) > 0$ and $M(t, x, b, a) > 0$ for every $t > 0$.*

Proof. Assume that $M(t, x, b, a) = 0$ for some $t = t_0 > 0$ and hence for every $t < t_0$. Then if $J = [b, \infty)$ we have for every $t < t_0$

$$P(t, x, J) = \int_0^t P(t - s, a, J) d_s M(s, x, a)$$

and in virtue of the continuity of paths

$$P(t, a, J) = \int_0^t P(t - s, x, J) d_s M(s, a, x) .$$

Now because of the continuity of paths we can choose t_1 so $0 < t_1 < t_0$ and

$$M(t, a, x) M(t, x, a) \leq 1/2 \qquad \text{for } 0 \leq t \leq t_1 .$$

Since $P(s, a, J) \leq 1$ for all $s \leq t_1$ it follows from the integral equations that

$$P(s, a, J) \leq 1/2 \qquad \text{for } s \leq t_1 ,$$

and by an iteration argument we obtain $P(t_1, a, J) = 0$ which contradicts the hypothesis. Hence $M(t, x, b, a) > 0$ for $t > 0$. Similarly $M(t, x, a, b) > 0$ for $t > 0$.

DEFINITION. Given an open interval $V = (a, b)$ let

$$R(t, x, V) = P_x\{\tau_a(\omega) > t, \tau_b(\omega) > t\} .$$

LEMMA 2. *If* $x \in V = (a, b)$ *then* $R(t, x, V) > 0$ *for all* $t > 0$.

Proof. Assume that for some $x \in V$ and $t' > 0$ we have $R(t', x, V) = 0$. Then $R(t, x, V) = 0$ for all $t > t'$. Because of continuity of paths $t_0 = \inf\{t ; R(t, x, V) = 0\}$ is positive. Now choose any $y \in V, y \neq x$. To fix the ideas we assume $x < y < b$. If $\varepsilon > 0$ is so small that $M(t', x, y, a) - M(\varepsilon, x, y, a) > 0$ then the inequality

$$0 = R(t', x, V) \geq \int_\epsilon^{t'} R(t' - \tau, y, V) d_\tau M(\tau, x, y, a)$$

shows that $R(t' - \epsilon, y, V) = 0$. Consequently if $t_1 = \inf\{t ; R(t, y, V) = 0\}$ then

$$0 < t_1 \leq t_0 - \varepsilon < t_0 .$$

But we can now repeat the argument and show that $t_0 < t_1$. This contradiction proves the lemma.

Proof of the Theorem. Let $x_1 < \cdots < x_n$ and $E_1 < \cdots < E_n$ be non-void open sets. The *index* of the determinant

$$P\left(t ; \begin{matrix} x_1, \cdots, x_n \\ E_1, \cdots, E_n \end{matrix}\right)$$

is defined to be the number k of values of i for which x_i is not in E_i. Thus the index of an nth order determinant of this kind is an integer between 0 and n inclusive.

In each set E_i choose a non-void open interval U_i such that $x_i \in U_i$ if $x_i \in E_i$ but $\bar{U}_i$ contains no x_j if $x_i \notin E_i$. Because of the probabilistic interpretation

$$P\begin{pmatrix} t\,; & x_1, \cdots, x_n \\ & E_1, \cdots, E_n \end{pmatrix} \geq P\begin{pmatrix} t\,; & x_1, \cdots, x_n \\ & U_1, \cdots, U_n \end{pmatrix} .$$

These two determinants have the same index k. If $k = 0$, then from the probabilistic interpretation and the second lemma above

$$P\begin{pmatrix} t\,; & x_1, \cdots, x_n \\ & U_1, \cdots, U_n \end{pmatrix} \geq \prod_{i=1}^{n} R(t, x_i, U_i) > 0 .$$

Thus the subdeterminants with index zero are positive. Now suppose the index is $k > 0$. We can find n open intervals $U_1', \cdots, U_n'$ whose closures are mutually disjoint such that $x_i \in U_i'$ for every i and $U_i' = U_i$ if $x_i \in U_i$. We can choose n points $x_1', \cdots, x_n'$ such that $x_i' \in U_i$ for every i and $x_i' = x_i$ if $x_i \in U_i$. Now in the collection $U_1, \cdots, U_n, U_1', \cdots, U_n'$ there are exactly $m = n + k$ distinct intervals and they are disjoint. Denote them by $V_1 < \cdots < V_m$. Similary in $x_1, \cdots, x_n, x_1', \cdots, x_n'$ there are exactly $n + k$ distinct points. Denote them by $y_1 < \cdots < y_m$ and then $y_i \in V_i$ for each i. Let $B(t)$ be the m-square matrix with elements

$$b_{ij}(t) = P(t, y_i, V_j) .$$

The determinant $P\begin{pmatrix} t\,; & x_1, \cdots, x_n \\ & U_1, \cdots, U_n \end{pmatrix}$ is a minor of $B(t)$. Moreover $B(t)$ is totally positive, all of its elements are strictly positive, and its principal minors have index zero and are therefore strictly positive. Hence by Lemma 14 of [**14**] all minors of $B(t)$ of index one are strictly positive. This proves that $P\begin{pmatrix} t\,; & x_1, \cdots, x_n \\ & E_1, \cdots, E_n \end{pmatrix} > 0$ if the index of this determinant is ≤ 1. We now assume that for some integer r, $1 \leq r < n$, all the determinants of the type $P\begin{pmatrix} t\,; & x_1, \cdots, x_n \\ & E_1, \cdots, E_n \end{pmatrix}$ with index $\leq r$ are strictly positive.

Let $1 \leq i_1 < \cdots < i_n \leq m$, $1 \leq j_1 < \cdots < j_n \leq m$ and

$$\sum_{\nu=1}^{n} | i_\nu - j_\nu | = r + 1 .$$

Then

$$P\left(t+s;\ \begin{matrix} y_{i_1}, \cdots, y_{i_n} \\ V_{j_1}, \cdots, V_{j_n} \end{matrix}\right)$$

$$\geq \sum_{1 \leq \alpha_1 < \cdots < \alpha_n \leq m} \int_{v_1 \in V_{\alpha_1}} \cdots \int_{v_n \in V_{\alpha_n}} P\left(t;\ \begin{matrix} y_{i_1}, \cdots, y_{i_n} \\ dv_1, \cdots, dv_n \end{matrix}\right) P\left(s;\ \begin{matrix} v_1, \cdots, v_n \\ V_{j_1}, \cdots, V_{j_n} \end{matrix}\right)$$

and in this sum there is at least one term with

$$\sum_{\nu=1}^{n} | i_\nu - \alpha_\nu | \leq r\ , \qquad \sum_{\nu=1}^{n} | \alpha_\nu - j_\nu | \leq r\ .$$

For this term the integrand $P\left(s;\ \begin{matrix} v_1, \cdots, v_n \\ V_{j_1}, \cdots, V_{j_n} \end{matrix}\right)$ is positive for every $v_1, \cdots, v_n$ in the range of integration because $v_\nu \in V_{j_\nu}$ for at least $n-r$ values of ν. Also for this term the integrator $P\left(t;\ \begin{matrix} y_{i_1}, \cdots, y_{i_n} \\ dv_1, \cdots, dv_n \end{matrix}\right)$ has positive measure on the range of integration because $y_{i_\nu} \in V_{\alpha_\nu}$ for at least $n - r$ values of ν. Hence the special term and also the entire sum is strictly positive. This proves that $P\left(t;\ \begin{matrix} x_1, \cdots, x_n \\ E_1, \cdots, E_n \end{matrix}\right) > 0$ if the index of this determinant is $\leq r + 1$, and the theorem follows by induction on the index.

8. Local character of $P(t, x, E)$ and continuity of path functions. Let $P(t, x, E)$ be the transition probability function of a stationary Markoff process on the real line. Given $\delta > 0$ we define

$$\begin{aligned} V(x, \delta) &= [x + \delta, \infty)\ , \\ U(x, \delta) &= (-\infty, x - \delta]\ , \\ I'(x, \delta) &= U(x, \delta) \cup V(x, \delta)\ . \end{aligned}$$

The transition probabilities are called of *local character* if $P(t, x, I'(x, \delta)) = o(t)$ for each x and $\delta > 0$. They are called *uniformly of local character* if for each $\delta > 0$ and each compact set F on the real line the relation $P(t, x, I'(x, \delta)) = o(t)$ holds uniformly for $x \in F$. We will prove that if the transition probabilities are positive of order two (see Theorem 5) and if for some $\alpha > 0$ we have $P(t, x, I'(x, \delta)) = o(t^\alpha)$ for each x and each $\delta > 0$ then the transition probabilities are uniformly of local character, and in fact for every $\beta > 0$ the relation $P(t, x, I'(x, \delta)) = o(t^\beta)$ holds uniformly on compact sets. This is of interest in connection with a theorem of Ray [19] to the effect that if the transition probabilities are uniformly of local character and if $P(t, x, X) = 1$ where X is the real line (not the extended real line) then the process has path functions continuous except possibly at $+\infty$ and $-\infty$.

THEOREM 5. *Let $P(t, x, E)$ be stationary transition probabilities on the real line such that $P(t, x, E) \to 1$ as $t \to 0+$ if x is an interior point of E. If $P(t, x, E)$ is positive of order two (i. e. the second order determinants of* (4) *are non-negative) and if there is an $\alpha > 0$ such that for every x and every $\delta > 0$ we have $P(t, x, I'(x, \delta)) = o(t^\alpha)$ then for every compact set F on the real line and every $\beta > 0$, $\delta > 0$ there is a constant $M = M(F, \delta, \beta)$ such that*

$$P(t, x, I'(x, \delta)) \leq Mt^\beta$$

for every $x \in F$.

Proof. Given a point x on the real line and $\delta > 0$ let $y = x + \delta/2$ and $N = (y - \delta/4, y + \delta/4)$. Then because of the second order positivity

$$P(t, x, V(x, \delta))P(t, y, N) \leq P(t, x, N)P(t, y, V(x, \delta)) .$$

Both factors of the right member of this inequality are $O(t^\alpha)$ while $P(t, y, N) \to 1$ as $t \to 0$. Hence $P(t, x, V(x, \delta)) = O(t^{2\alpha})$.

This is valid for arbitrary x and δ, so the argument can be iterated, and for any integer $n \geq 1$ we have

$$P(t, x, V(x, \delta)) = O(t^{2^n\alpha}) .$$

The O symbol so far may depend on x and certainly depends on δ. A similar argument applies to $P(t, x, U(x, \delta))$ and combining them we have

$$P(t, x, I'(x, \delta)) = O(t^\beta)$$

for any $\beta > 0$.

Now suppose $x < y < z$, let $E = (z, \infty)$ and let W be an open interval containing y but whose closure does not contain z. Then

$$\begin{aligned} P(t, x, E)P(t, y, W) &\leq P(t, x, W)P(t, y, E) \\ &\leq P(t, y, E) \end{aligned}$$

There is a positive $t_0 = t_0(y, E)$ such that $P(t, y, W) \geq 1/2$ for $t \leq t_0$ and therefore

$$P(t, x, E) \leq 2P(t, y, E) \qquad \text{if } t \leq t_0 .$$

Similarly if $z < y < x$ and $E = (-\infty, z)$ then there is a positive $t_1 = t_1(y, E)$ such that

$$P(t, x, E) \leq 2P(t, y, E) \qquad \text{if } t \leq t_1 .$$

Now let $F = [a, b]$ be a finite interval and $\delta > 0$. Choose a finite number of points $y_1, \cdots, y_m$ such that every open subinterval of $(a - \delta,$

$b + \delta)$ of length $(1/2)\delta$ contains at least one of the points y_i. Given $x \in F$ there are indices α, β such that

$$x - \frac{1}{2}\delta < y_\alpha < x < y_\beta < x + \frac{1}{2}\delta .$$

Since $U(x, \delta) \subseteq U(y_\alpha, \delta/4)$ and $V(x, \delta) \subseteq V(y_\beta, \delta/4)$ we have

$$P(t, x, U(x, \delta)) \leq 2P\Big(t, y_\alpha, U\Big(y_\alpha, \frac{\delta}{4}\Big)\Big) ,$$

$$P(t, x, V(x, \delta)) \leq 2P\Big(t, y_\beta, V\Big(y_\beta, \frac{\delta}{4}\Big)\Big)$$

for sufficiently small t. In fact these inequalities are valid if t is less than the least of the numbers $t_0(y_i, V(y_i, \delta/4)), t_1(y_i, U(y_i, \delta/4)), i=1, 2, \cdots, m$. Since each of the finite collection of functions $P(t, y_i, V(y_i, \delta/4))$, $P(t, y_i, U(y_i, \delta/4)), i = 1, 2, \cdots, m$ is $o(t^\beta)$ for any $\beta > 0$, it follows at once that for fixed $\delta > 0, \beta > 0$ $P(t, x, I'(x, \delta)) = O(t^\beta)$ uniformly for $x \in F$.

9. Homogeneous processes. A process on the real line will be called a homogeneous process if it is a stationary strong Markoff process with right continuous path functions and its transition probability function satisfies the homogeneity relation

$$P(t, x + h, E) = P(t, x, E - h)$$

where $E - h = \{y\,; y + h \in E\}$. This class of processes includes all the processes with stationary independent increments and is slightly more general. If X denotes the real line then for any homogeneous process the function

$$P(t, x, X) = P(t, 0, X) = \alpha(t)$$

is independent of x. From the Chapman-Kolmogoroff equation $\alpha(t+s)= \alpha(t)\alpha(s)$ and then because of monotonicity $\alpha(t) = e^{-\beta t}$ where $0 \leq \beta \leq +\infty$. The case $\beta = 0$ gives the processes with stationary independent increments. The general homogeneous process is obtained by taking a process with stationary independent increments and stopping it after a random time T with $\Pr\{T > t\} = e^{-\beta t}$. The trivial case $\beta = +\infty$ is excluded in the remainder of this section.

There are two special kinds of homogeneous processes of particular interest from our point of view. First the *essentially determined* ones for which, if E is any open set

$$P(t, x, E) = \begin{cases} e^{-\beta t} & \text{if} \quad x + vt \in E \\ 0 & \text{otherwise} \end{cases}$$

where v is a real constant and $0 \leq \beta < \infty$. And second, those *derived from the Wiener process,* for which

$$P(t, x, E) = \frac{e^{-\beta t}}{\sqrt{2\pi\sigma t}} \int_E \exp\left[-\frac{(x + vt - y)^2}{2\sigma t}\right] dy$$

where v is a real and σ a positive constant and $0 \leq \beta < \infty$. These two types are interesting because they have continuous path functions and the transition probability functions are therefore totally positive. For those derived from the Wiener process it is strictly totally positive, while for the essentially determined ones it is not. The main result in this section is the following.

THEOREM 6. *If the transition probability function of a homogeneous process is totally positive then the process is either an essentially determined one or else one derived from the Wiener process.*

Together with the results of § 5 this theorem shows that for homogeneous processes total positivity is equivalent to continuity of the path functions. At the close of this section we show by a different method that for homogeneous processes positivity of order two is already equivalent to continuity of the path functions. This assertion is probably true not only for homogeneous processes but for arbitrary one dimensinal strong Markoff processes with right continuous path functions. Although we are not yet able to prove the result in this generality, we do have a proof for the case of birth and death processes, which is published separately [12].

Proof. Let $P(t, x, E)$ be the transition probability function of a totally positive homogeneous process and let $P(t, x, (-\infty, \infty)) = e^{-\beta t}$. We form the function

$$P_\varepsilon(t, x, E) = \int_{-\infty}^{\infty} e^{\beta t} P(t, y, E) q_\varepsilon(t, y - x) dy$$

where $\varepsilon > 0$ and $q_\varepsilon(t, x) = (2\pi\varepsilon t)^{-1/2} \exp[-(x^2/2\varepsilon t)]$. Then P_ε is a homogeneous strictly totally positive kernel for $t > 0$, it satisfies the Chapman Kolmogoroff equation, and is analytic in its dependence on x. There is therefore a density function $p_\varepsilon(t, x)$ such that

$$P_\varepsilon(t, x, E) = \int_E p_\varepsilon(t, y - x) dy \;.$$

For fixed ε, p_ε is measurable in t, x and is analytic in x for fixed ε, t. From the formula

$$P_\varepsilon(t, y - x) = \lim_{h\to 0+} \frac{1}{h} P_\varepsilon(t, x, (y, y + h))$$

we deduce that if $x_1 < x_2 < \cdots < x_n$ and $y_1 < y_2 < \cdots < y_n$ then det $p_\varepsilon(t, x_i - y_j) \geq 0$ for $t > 0$. Thus for fixed t and ε the function $p_\varepsilon(t, x)$ is a Pólya frequency function $\left(\text{we have } \int_{-\infty}^{\infty} p_\varepsilon(t, x)dx = 1\right)$ in the sense of Schoenberg [20] and the Laplace transform

$$\frac{1}{\psi(s, t)} = \int_{-\infty}^{\infty} e^{-xs} p_\varepsilon(t, x)dx$$

converges in a strip $-a < \mathrm{Re}[s] < a$ with $a > 0$, and has there a representation

$$\psi(s, t) = e^{-\gamma s^2 + \delta s} \prod_{\nu=1}^{\infty} (1 + \delta_\nu s)e^{-\delta_\nu s}$$

where $\gamma \geq 0, \delta, \delta_\nu$ are real, $0 < \gamma + \sum \delta_\nu^2 < \infty$. The constants $\gamma, \delta, \delta_\nu$ will of course depend on t. From the Chapman-Kolmogoroff equation we have $\psi(s, t) = [\psi(s, t/n)]^n$ where n is any positive integer. Consequently any zero of $\psi(s, t)$ must be of order at least n and n being arbitrary there can be no zeros. Hence

$$\psi(s, t) = e^{\delta s - \gamma s^2}, \qquad \gamma > 0 .$$

Again using the Chapman Kolmogoroff equation in the form $\psi(s, t+\tau) = \psi(s, t)\psi(s, \tau)$ we deduce that $\delta = at, \gamma = b^2 t$ where a, b are real and independent of t. Now if $t > 0$ is fixed $F(x) = e^{\beta t} P(t, x, (0, \infty))$ is non-decreasing, $F(-\infty) = 0, F(+\infty) = 1$, that is F is a distribution function, and the above result shows that the convolution of F with the normal density $q_\epsilon(t, x)$ is a distribution of normal type. By a well known theorem [17], F is also of normal type and we have

$$\int_{-\infty}^{\infty} e^{-sx} dF(x) = e^{-ats} + (b^2 - \varepsilon)ts^2$$

with $b^2 - \varepsilon \geq 0$. If $b^2 - \varepsilon = 0$ the given homogeneous process is an essentially determined one while if $b^2 - \varepsilon > 0$ it is one derived from the Wiener process.

Another approach to the problem of determining when homogeneous processes or equivalentely infinitely divisible processes are totally positive is based on the Levy Khintchine representation. We consider an infinitely divisible process $x(t)$ properly centered with no fixed points of discontinuities whose characteristic function $\varphi(t, s)$ has an expression

$$(1) \qquad \log \varphi(t, s) = t\left[i\gamma s + \int_{-\infty}^{\infty}\left(e^{isx} - 1 - \frac{isx}{1 + x^2}\right)dG(x)\right]$$

$$= t\psi(s)$$

with the aid of (1) we are able to establish

$$\lim_{t\to 0}\frac{1}{t}\Pr\{|x(t)-x(0)|\geq\lambda\}=\int_{|x|\geq\lambda>0}dG(x) \tag{2}$$

when λ and $-\lambda$ are continuity points of G. This limit relation is essentially known but for lack of any available specific reference we sketch a proof.

The proof consists of defining

$$H(t,\lambda)=\begin{cases}\dfrac{\Pr\{x(t)-x(0)\leq\lambda\}-1}{t}; & \text{for}\quad \lambda>0\\[2ex] \dfrac{\Pr\{x(t)-x(0)\leq\lambda\}}{t} & \text{for}\quad \lambda<0\end{cases}$$

and forming the Fourier Stieljes transform of H which reduces to $(\varphi(t,s)-1)/t$. This clearly converges pointwise as $t\to 0$ to $\psi(s)$. Invoking the Levy convergence criteria following comparison with (1) establishes (2).

An alternative proof of (2) can be based on verifying the validity of (2) first for the case of a finite composition of independent Poisson processes and afterwards passing to a limit to obtain the general infinitely divisible process.

The truth of (2) also follows by exploiting the properties of the infinitely divisible process $U_{\lambda,t}$ which counts the number of jumps of magnitude exceeding λ that the process $x(t)$ executes in time t. (See [5] page 424).

Because of (2) and Theorem 5, we see that $x(t)$ is totally positive of order 2 if and only if $\int_{|x|\geq\lambda}dG(x)=0$ for all $\lambda>0$. Hence the only totally positive infinitely divisible process is the Wiener process except for a drift factor.

10. Examples. In this section we present some examples of totally positive semigroups of matrices and kernels. These matrices and kernels are fundamental solutions of parabolic differential equations (or differential difference equations).

In generating examples of totally positive kernels it is useful to note that if $P(t,x,E)$ represents a totally positive kernel and $P(t,x,E)$ possesses a continuous density $p(t,x,y)$ with respect to a σ-finite measure μ then $p(t,x,y)$ is totally positive in the sense that

$$\det p(t,x_i,y_j)\geq 0$$

where $x_1<x_2<\cdots<x_n$ and $y_1<y_2<\cdots<y_n$. The proof consists of

selecting $E_1 < E_2 < \cdots < E_n$ where E_i is a sufficiently small open set enclosing y_i and computing

$$\lim\left[\frac{1}{\mu(E_1)\mu(E_2)\cdots\mu(E_n)}\det P(t, x_i, E_j)\right]$$

the limit taken as $\mu(E_i)$ tends to 0 for all i.

Ex. (i) The analytic properties of birth and death matrices have already been investigated in detail by the authors [14]. In Theorem 20 of that paper it is shown that with every solvable Stieltjes moment problem there is associated one or more strictly totally positive semigroups of matrices. A few examples of interest are recorded:

(a) Let $L_n^\alpha(x)$ be the usual Laguerre polynomials $\left(\text{normalized so that } L_n^\alpha(0) = \binom{n+\alpha}{n}\right)$, and let $P(t)$ be the infinite matrix with elements

$$P_{nm}(t) = \int_0^\infty e^{-xt} L_n^\alpha(x) L_m^\alpha(x) x^\alpha e^{-x} dx .$$

Then $P(t)$ is strictly totally positive for $t > 0, \alpha > -1$.

(b) Let $c_n(x, a)$ be the Poisson-Charlier polynomials [15] and $P(t)$ the matrix with elements

$$P_{nm}(t) = \sum_{k=0}^\infty e^{-kt} c_n(k, a) c_m(k, a) \frac{a^k}{k!} .$$

Then $P(t)$ is strictly totally positive for $t > 0, a > 0$.

Ex (ii) The Wiener process on the real line is a strong Markoff process with continuous path functions. The direct product of n copies of this process is the n-dimensional Wiener process which is known to be a strong Markoff process. Therefore the kernel

$$P(t, x, E) = \frac{1}{\sqrt{4\pi t}} \int_E \exp\left[\frac{-(x-y)^2}{4t}\right] dy$$

is totally positive for $t > 0$ (strictly, since $P(t, x, E) > 0$ when E is an open set).

Ex. (iii) If $Y(t) = (Y_1(t), \cdots, Y_k(t))$ is the k-dimensional Wiener process and $X(t)$ is its radial part, *i.e.*,

$$X(t) = \left[\sum_1^k Y_i^2(t)\right]^{1/2}$$

then $X(t)$ is a process on $0 \le x < \infty$ with continuous path functions. These processes have been studied by Levy [16], Spitzer [22] and others. The corresponding diffusion equation and transition function are

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{2\gamma}{x}\frac{\partial u}{\partial x},$$

$$P(t, x, E) = \int_E p(t, x, y)d\mu(y),$$

where

$$\gamma = \frac{k-1}{2},$$

$$p(t, x, y) = \int_0^\infty e^{-\alpha^2 t}\, T(\alpha y)T(\alpha y)d\mu(\alpha),$$

$$T(x) = \Gamma\left(\gamma + \frac{1}{2}\right)\left(\frac{x}{2}\right)^{1/2-\gamma} J_{\gamma-1/2}(x),$$

$$\mu(x) = \frac{x^{2\gamma+1}}{2^{\gamma+1/2}\Gamma(\gamma+3/2)},$$

where J stands for the usual Bessel function.

These formulas make sense for arbitrary $\gamma \geq 0$ and have been studied by Bochner [2]. The density may be written in the form

$$p(t, x, y) = (2t)^{-(\gamma+1/2)} \exp\left(\frac{-x^2}{4t}\right)\exp\left(\frac{-y^2}{4t}\right)T\left(\frac{ixy}{2t}\right).$$

Now $T(ix/2t)$ is a power series with positive coefficients, in fact

$$T\left(\frac{ix}{2t}\right) = \sum_{k=0}^{\infty} a_k x^{2k} = \int_{0-}^{\infty} x^s d\sigma(s)$$

where

$$a_k = \frac{\Gamma(\gamma+1/2)}{k!\,\Gamma(\gamma+k+1/2)(4t)^{2k}}$$

and $\sigma(s)$ is an increasing step function whose jumps occur at the even integers. Let $0 \leq x_1 < x_2 < \cdots < x_n$ and $0 \leq y_1 < y_2 < \cdots < y_n$. If $0 \leq s_1 < s_2 < \cdots < s_n$ then the Vandermonde determinant

$$\Delta\begin{pmatrix} x_1, \cdots, x_n \\ s_1, \cdots, s_n \end{pmatrix} = \det\{(x_\alpha)^{s_\beta}\}$$

is known to be non-negative, positive if $x_1 > 0$. From the formula

$$\det T\left(\frac{ix_\alpha y_\beta}{2t}\right) = \iint_{0 \leq s_1 < s_2 \cdots < s_n < \infty} \Delta\begin{pmatrix} x_1 \cdots x_n \\ s_1 \cdots s_n \end{pmatrix} \Delta\begin{pmatrix} y_1 \cdots y_n \\ s_1 \cdots s_n \end{pmatrix} d\sigma(s_1)d\sigma(s_2)\cdots d\sigma(s_n)$$

it readily follows that $T(ixy/2t)$ and hence also $p(t, x, y)$ is strictly totally positive.

Ex. (iv) If we consider Brownian motion on the circle the transition density function has the form

$$p(t, \theta, \psi) = 1 + 2 \sum_{n=1}^{\infty} e^{-4\pi^2 n t} \cos 2\pi n(\theta - \psi)$$

where θ and ψ traverse the unit interval. This formula may be derived as the fundamental solution of the heat equation on the circle. In this case the hypothesis of Theorem 1 are fulfilled and we deduce that all odd order determinants of $p(t, \theta, \psi)$ are non-negative (actually strictly positive); viz

If $0 \leqq \theta_1 < \theta_2 < \cdots < \theta_{2n+1} \leq 1$ and $0 \leq \psi_1 < \psi_2 < \cdots < \psi_{2n+1} \leq 1$ then $\det p(t, \theta_i, \psi_j) > 0$.

References

1. R. M. Blumenthal, *An extended Markov Property*, Trans, A. M. S., **85** (1957), 52-72.
2. S. Bochner, *Sturm Liouville and heat equations etc.*, Proc. Conf. Diff. Eqns. Maryland, (1955), 23-48.
3. K. L. Chung, *Foundations of the theory of continuous parameter Markoff chains*, Proc. Third Berkeley Symposium, **2** (1956), 29-40.
4. K. L. Chung, *On a basic property of Markov chains*, Annals of Math., **68** (1958), 126-149.
5. J. L. Doob, *Stochastic processes*, New York 1953.
6. E. B. Dynkin, *Infinitesimal operators of Markoff processes*, Theory of probability and its applications, **1** (1956), 38-60, (in Russian).
7. E. B. Dynkin, A. Juskevitch, *Strong Markov processes*, Theory of probability and its applications, **1** (1956), 149-155, (in Russian).
8. F. Gantmacher and M. Krein, *Oscillatory matrices and kernels and small vibrations of mechanical systems*, (in Russian), 2nd ed., Moscow 1950.
9. I. I. Hirschman and D. V. Widder, *The convolution transform*, Princeton 1955.
10. G. A. Hunt, *Some theorems concerning Brownian motion*, Trans. Amer. Math. Soc., **81** (1956), 294-319.
11. G. A. Hunt, *Markoff processes and potentials*, Illinois Jour. Math., **1** (1957), 44-93.
12. S. Karlin and J., McGregor, *A characterization of birth and death processes*, Proc. Nat. Acad. Sci., March (1959), 375-379.
13. S. Karlin, *Mathematical methods and theory in games, programming and economics*, Addison Wesley, to appear.
14. 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.
15. S. Karlin and J. McGregor, *Many server queuing processes with Poisson input and exponential service times*, Pacific J. Math., **8** (1958), 87-118.
16. P. Lèvy, *Processus stochastiques et mouvement brownien*, Paris 1948.
17. M. Loève, *Probability theory*, van Nostrand, 1955 (p. 271, Theorem A).
18. C. Loewner, *On totally positive matrices*, Math. Zeit., **63** (1955), 338-340.
19. D. Ray, *Stationary Markov processes with continuous path functions*, Trans. Amer. Math. Soc., **82** (1956) 452-493.
20. I. J. Schoenberg, *On Pólya frequency functions*, Jour. d'Anal. Math., **1** (1951), 331-374.
21. I. J. Schoenberg, *On smoothing operations and their generating functions*, Bull. Am. Math. Soc., **59** (1953), 199-230.
22. F. Spitzer, *Some theorems concerning 2-dimensional Brownian motion*, **87** (1958), 187-197.

STANFORD UNIVERSITY

MEASURES ON BOOLEAN ALGEBRAS

J. L. KELLEY

This paper is concerned with the general problem of the existence of measures on Boolean algebras. A *measure* on a Boolean algebra $\mathscr{A}$ is a finitely additive, non-negative function on $\mathscr{A}$ which assumes the value one at the unit element of the algebra $\mathscr{A}$. It is known that measures on Boolean algebras always exist, and in some profusion (see, for example, [2]). We are concerned primarily with the existence of measures which are *strictly positive*; that is measures which vanish only at the zero element of the algebra. Not all Boolean algebras possess strictly positive measures, and workable necessary and sufficient conditions for the existence of a strictly positive measure have not been given. We shall give such conditions. Our results seem to represent definite progress on the general problem, although the relationship between our conditions and various conjectures is not clear. In particular, I do not know whether there is necessarily a strictly positive measure on an algebra $\mathscr{A}$ which satisfies the condition: $\mathscr{A} - \{0\}$ is the union of a countable family $\{\mathscr{A}_n\}$, such that each disjoint subclass of the class $\mathscr{A}_n$ contains at most n members. Tarski has conjectured that this is the case.

In the first section we define, combinatorially, for each subset $\mathscr{B}$ of a Boolean algebra $\mathscr{A}$ a number, $I(\mathscr{B})$, called the intersection number of $\mathscr{B}$. It is then showed that there is a strictly positive measure on $\mathscr{A}$ if and only if $\mathscr{A} - \{0\}$ is the union of a countable number of sets, each of which has positive intersection number. The intersection number is also evaluated precisely in terms of measures on $\mathscr{A}$; $I(\mathscr{B})$ is the maximum, for all measures m on $\mathscr{A}$, of inf $\{m(B) : B \in \mathscr{B}\}$. A dualized formulation of these results in terms of coverings is obtained.

The second section is concerned with the existence of countably additive measures. Necessary and sufficient conditions for the existence of such measures have been given by Maharam [3], but these conditions are not entirely satisfactory. The contribution to the problem made here is simply this: an algebra supports a countably additive strictly positive measure if and only if it has a strictly positive measure and is weakly countably distributive. (See the second section for definitions). The condition of weak countable distributivity appears thus as a very natural requirement which enables one to derive countably additive measures from finitely additive ones; the fundamental difficulties lie in the finitely additive case.

It has been shown that some of the natural conjectures on the

Received February 2, 1959. This investigation was made during the tenure of a Fulbright Research Professorship at Cambridge University.

existence of measures, or of countably additive measures, imply the Souslin hypothesis (cf. [2] and [3]). In the third section of this paper some results of this sort are obtained. Assuming the falsehood of the Souslin hypothesis, a linear continuum is constructed that has appallingly pathological properties. A condition in terms of Boolean algebras which is equivalent to the Souslin hypothesis is given. The last section deals with measures which dominate, or are dominated by, a non-negative function having certain convexity properties.

We shall assume, unless specifically stated otherwise, that the Boolean algebra $\mathscr{A}$ is the algebra of all open and closed subsets of a compact totally disconnected Hausdorff space X; the Boolean operations of (finite) join and meet are then simply union and intersection, and X is the unit of the algebra. This assumption is justified since, via the classical Stone theorem, every Boolean algebra can be isomorphically represented in this fashion.

Intersection numbers and covering numbers. Let $\mathscr{B}$ be an arbitrary non-void subclass of a Boolean algebra $\mathscr{A}$. For each finite sequence $S = \langle S_1, \cdots, S_n \rangle$ of (not necessarily distinct) members of $\mathscr{B}$ denote by $n(S)$ the number n of terms in the sequence and let $i(S)$ be the maximum number of members with non-void intersection. If K_i is the characteristic function of S_i then $i(S) = \sup \{\sum \{K_i(x); i = 1, \cdots n\}: x \in X\}$. The *intersection number* $I(\mathscr{B})$ is defined to be $\inf \{i(S)/n(S);$ S a finite sequence in $\mathscr{B}\}$.

1. PROPOSITION. If m is a measure on $\mathscr{A}$ and $\mathscr{B}$ is a non-void subclass of $\mathscr{A}$ then $\inf \{m(B) : \in \mathscr{B}\} \leq I(\mathscr{B})$, where $I(\mathscr{B})$ is the intersection number of $\mathscr{B}$.

Proof. Let $r = \inf\{m(B) : B \in \mathscr{B}\}$, and let K_i be the charcteristic function of S_i where $S = \langle S_1, \cdots, S_n \rangle$ is a finite sequence in $\mathscr{B}$. Then $\int \sum \{K_i : i = 1, \cdots, n\}\, dm = \sum \{m(S_i) : i = 1, \cdots, n\} \geq rn$, and therefore $\sum \{K_i(x) : i = 1, \cdots, n\} \geq rn$ for some x in X. Hence the maximum number $i(S)$ of elements of S which intersect is at least rn, and $i(S)/n(S) \geq r$. Taking the infimum for all S, we have $I(\mathscr{B}) \geq r$.

The preceding proposition can be restated: $I(\mathscr{B}) \geq \sup \inf \{m(B):$ $B \in \mathscr{B}\}$, where the supremum is taken over all measures m on $\mathscr{A}$. The principal result of this section implies that equality holds here and the supremum is assumed.

2. THEOREM. *If $\mathscr{B}$ is a non-void subclass of a Boolean algebra $\mathscr{A}$ then there is a measure m on $\mathscr{A}$ such that $\inf\{m(B) : B \in \mathscr{B}\} = I(\mathscr{B})$, where $I(\mathscr{B})$ is the intersection number of $\mathscr{B}$.*

Proof. Let C be the class of all continuous real valued functions on X, with the usual supremum norm, let F be the class of characteristic functions of members of $\mathscr{B}$, and let G be the convex hull of F. We assert that if $f \in G$ then $\|f\| \geq I(\mathscr{B})$. For suppose that $\|\sum \{t_i K_i : i=1, \cdots, q\}\|=r$, where K_i is the characteristic function of a member S_i of $\mathscr{B}$, $0 < t_i \leq 1$ for each i, and $\sum \{t_i : i = 1, \cdots, q\} = 1$. For arbitrary $e > 0$ we then have $\|(1/n) \sum \{p_i K : i = 1, \cdots, q\}\| < r + e$ for suitably chosen positive integers $p_1, \cdots, p_q$ with $\sum \{p_i : i = 1, \cdots, q\} = n$. Upon considering the sequence in $\mathscr{B}$ obtained by counting each S_i the integer p_i times, we see that at least $n\, I(\mathscr{B})$ of the members of that sequence intersect, whence $I(\mathscr{B}) \leq \|(1/n) \sum \{p_i K_i : i = 1, \cdots, q\}\| < r + e$. Consequently $I(\mathscr{B}) \leq r$ and the assertion is proved.

Let H be the open sphere in C about 0 of radius $I(\mathscr{B})$; i.e. $H = \{f : \|f\| < I(\mathscr{B})\}$. Since the norm of each member of G is at least $I(\mathscr{B})$, each member of the combinatorial sum $G + H$ is somewhere positive. If P is the class of non-negative members of C and Q is the cone $\{s(g+h)+tp : s \geq 0,\ t \geq 0,\ g \in G,\ h \in H,\ p \in P\}$, then no member of Q is everywhere negative. In view of the Hahn-Banach theorem, there is therefore a hyperplane separating Q and the function which is identically -1; that is, there is a linear functional ϕ such that $\phi(-1) < \phi(f)$ for all f in Q. We may suppose that $\phi(-1) = -1$, and because Q is closed under multiplication by positive scalars, we must have $\phi(f) \geq 0$ for f in Q. For $e > 0$, the function constantly equal to $-I(\mathscr{B}) + e$ belongs to H, hence $f - I(\mathscr{B}) + e \in Q$ for all f in G, and therefore $\phi(\mathrm{f} - I(\mathscr{B}) + e) \leq 0$. Thus $\phi(f) \geq I(\mathscr{B}) - e$ for all positive e, and hence $\phi(f) \geq I(\mathscr{B})$ for all f in G. Finally, for each A in $\mathscr{A}$ let $m(A)$ be ϕ of the characteristic function K_A of A. Then m is non-negative, finitely additive, $m(X) = 1$, and $m(B) \geq I(\mathscr{B})$ if $B \in \mathscr{B}$, because in this case the characteristic function K_B belongs to G. The theorem is then proved.

3. COROLLARY. *For each non-void subset $\mathscr{B}$ of $\mathscr{A}$ the intersection number $I(\mathscr{B})$ is the maximum of the numbers* $\inf \{m(B) : B \in \mathscr{B}\}$ *for all measures m on $\mathscr{A}$.*

4. THEOREM. *There is a strictly positive measure on a Boolean algebra $\mathscr{A}$ if and only if $\mathscr{A} - \{0\}$ is the union of a countable number of classes, each of which has a positive intersection number.*

Proof. If $\mathscr{A} - \{0\}$ is the union of classes $\mathscr{B}_n$ with $I(\mathscr{B}_n) > 0$ then, choosing measures m_n on $\mathscr{A}$ with $\inf \{m_n(B) : B \in \mathscr{B}_n\} > 0$, the sum $\sum \{2^{-n} m_n : n \text{ an integer}\}$ is a strictly positive measure. On the other hand, if m is strictly positive on $\mathscr{A}$ then $\mathscr{A} - \{0\}$ is the union of the classes $\{B : m(B) \geq 1/n\}$, and each of these classes has a positive intersection number by virtue of Proposition 1.

We may derive directly from the preceding results a necessary and sufficient condition for the existence of a measure which is "small" on a subclass $\mathscr{C}$ of an algebra $\mathscr{A}$, since $m(A) \leq r$ for each A in $\mathscr{C}$ if and only if $m(B) \geq 1 - r$ for each member B of the class $\mathscr{B}$ of complements of members of $\mathscr{C}$. This dualization leads to an interesting description of the intersection number of $\mathscr{B}$ in terms of the class $\mathscr{C}$, which we now give.

For each finite sequence $S = \langle S_1, \cdots, S_n \rangle$ in $\mathscr{A}$ let $m(S)$, the *multiplicity of covering,* be the minimum number of times that each point of X is covered by elements of S. If K_i is the characteristic function of S_i then $m(S) = \inf \{\sum \{K_i(x) : i = 1, \cdots, n\} : x \in X\}$. Let $n(S)$, as before, be the number n of elements of S, and for each non-void class $\mathscr{C}$ of $\mathscr{A}$ let the *covering number* $C(\mathscr{C})$ be the supremum of $m(S)/n(S)$ for all finite sequences S in $\mathscr{C}$. Intuitively, this may be interpreted: $1/C(\mathscr{C})$ is the outer measure of X obtained by using coverings by elements of $\mathscr{C}$, assigning members of $\mathscr{C}$ measure one, and permitting "multiple" coverings.

The connection between intersection and covering numbers is given by the

5. PROPOSITION. If $\mathscr{B}$ is a non-void subclass of the algebra $\mathscr{A}$ and $\mathscr{C}$ is the class of complements of members of $\mathscr{B}$, then $I(\mathscr{B}) + C(\mathscr{C}) = 1$.

Proof. Let S be a finite sequence in $\mathscr{A}$. The number $n(S)-i(S)$ can be described: it is the smallest number such that some set of (S) elements of S intersect, and, in terms of the sequence $S' = \langle S_1',$ $\cdots, S_n' \rangle$ of complements S of $S = \langle S_1, \cdots, S_n \rangle$, the number $n(S) - i(S)$ is the smallest such that the remainder, after omission of some set of $n(S)-i(S)$ elements from S', does not cover X. In brief, $n(S)-i(S)$ is the multiplicity of covering of $\langle S_1', \cdots, S_n' \rangle$. We then have: $1 - I(\mathscr{B}) = 1 - \inf \{i(S)/n(S) : S$ a finite sequence in $\mathscr{B}\} = \sup \{[n(S)-i(S)]/n(S) : S$ a finite sequence in $\mathscr{B}\} = \sup \{m(S')/n(S') : S'$ a finite sequence in $\mathscr{C}\} = C(\mathscr{C})$.

In view of Corollary 3, the preceding proposition implies the

6. COROLLARY. *For each non-void subclass $\mathscr{C}$ of $\mathscr{A}$ the covering number $C(\mathscr{C})$ is the minimum of the numbers* $\sup\{m(A) : A \in \mathscr{C}\}$, *where the minimum is taken over all measures m on $\mathscr{A}$.*

In view of Theorem 4, we have immediately the

7. COROLLARY. *There is a strictly positive measure on $\mathscr{A}$ if and only if $\mathscr{A} - \{X\}$ is the union of a countable family of subclasses, each of which has covering number less than* 1.

Countably additive measures. A Boolean algebra $\mathscr{A}$ is *complete* if and only if each subclass of $\mathscr{A}$ has a supremum in $\mathscr{A}$ (or equivalently, each subclass has an infimum). We shall denote $\inf \{B : B \in \mathscr{B}\}$ by

$\bigwedge \{B : B \in \mathscr{B}\}$, and sup $\{B : B \in \mathscr{B}\}$ by $\bigwedge \{B : B \in \mathscr{B}\}$. In general these differ from $\bigcap \{B : B \in \mathscr{B}\}$ and $\bigcup \{B : B \in \mathscr{B}\}$; if the infimum and supremum exist, then $\bigwedge \{B : B \in \mathscr{B}\}$ is the interior of $\bigcap \{B : B \in \mathscr{B}\}$, and $\bigvee \{B : B \in \mathscr{B}\}$ is the closure of $\bigcup \{B : B \in \mathscr{B}\}$. In case $\mathscr{A}$ is complete, the interior of each closed subset of X is closed, and the closure of each open subset is open.

In this section it will always be assumed that $\mathscr{A}$ is complete and satisfies the *countable chain condition* (that is, each disjoint subclass of $\mathscr{A}$ is countable). The set of positive integers will be denoted by ω, and the class of all sequences of positive integers by ω^ω. For $n \in \omega^\omega$, and $i \in \omega$, n_i is the ith member of the sequence n.

The algebra $\mathscr{A}$ is *weakly countably distributive* iff for every double sequence $A_{i,j}$ of members of $\mathscr{A}$ such that $A_{i,j+1} \subset A_{i,j}$ for all i and j, it is true that $\bigvee \{\bigwedge \{A_{i,j} : j \in \omega\} : i \in \omega\} = \bigwedge \{\bigvee \{A_{i,n_i} : i \in \omega\}$ $n \in \omega^\omega\}$. The topological condition on X which is equivalent to weak countable distributivity is simple and striking. Dixmier has shown [1] that each first category subset of the Stone space of a hyperstonian* algebra is nowhere dense, and the results of Horn and Tarski [2] imply that a hyperstonian algebra is weakly countably distributive. We now show that these two properties are equivalent for complete algebras satisfying the countable chain condition. (I do not believe the following theorem has been published previously, although it has been discovered independently by John Oxtoby.)

8. THEOREM. *A complete Boolean algebra $\mathscr{A}$ which satisfies the countable chain condition is weakly countably distributive if and only if each subset of the Stone X which is of category I is nowhere dense.*

Proof. We first note that a subset A of X is nowhere dense if and only if $\bigwedge \{B : B \in \mathscr{A}$ and $B \supset A\} = 0$. Moreover, because $\mathscr{A}$ satisfies the countable chain condition, the infimum of any subclass $\mathscr{B}$ of $\mathscr{A}$ is identical with the infimum of some countable subclass of $\mathscr{B}$, and hence A is nowhere dense if and only if there is a sequence $\{B_n\}$ of members of $\mathscr{A}$ such that $A \subset B_n$ and $\bigwedge_n B_n = 0$. Of course, $\{B_n\}$ may be assumed to be monotonically decreasing.

Assuming $\mathscr{A}$ is weakly countably distributive, suppose C is a subset of X which is of the first category. Then $C = \bigcup_i A_i$, where each A_i is nowhere dense, and there are members $B_{i,j}$ of $\mathscr{A}$, monotonically decreasing in j, such that $A_i \subset \bigcap_j B_{i,j}$ and $0 = \bigwedge_j B_{i,j}$ for each i. Hence $0 = \bigvee_i \bigwedge_j B_{i,j} = \bigwedge_n \bigvee_i B_{i,n_i}$. But $\bigvee_i B_{i,n_i} \supset \bigcup_i A_i = C$ for each n, and it follows that C is nowhere dense.

* That is, the Stone space of a measure algebra. The term "hyperstonian" seems unfortunate. In spite of my affection and admiration for Marshall Stone, I find the notion of a Hyper-Stone downright appalling.

To prove the converse, suppose that each subset of X which is of the first category is nowhere dense, and that $A_{i,j} \in \mathscr{A}$ and $A_{i,j+j} \subset A_{i,j}$ It is easy to see that $\bigvee_i \bigwedge_j A_{i,i} \subset \bigwedge_n \bigvee_i A_{i,n_i}$. Suppose that B is a non-void member of $\mathscr{A}$ such that $B \subset \bigwedge_n \bigvee_i A_{i,n_i}$ and $B \cap \bigvee_i \bigwedge_j A_{i,j}$ is void. Since $B \cap \bigwedge_j A_{i,j}$ is void, B does not intersect the interior of $\bigcap_j A_{i,j}$, and hence $B \cap \bigcap_j A_{i,j}$ is nowhere dense. Therefore $B \cap \bigcup_i \bigcap_j A_{i,j}$ is of category one and hence nowhere dense. We may then choose a non-zero member C of $\mathscr{A}$ such that $C \subset B$ and $C \cap \bigcup_i \bigcap_j A_{i,j}$ is empty. Using compactness, choose n_i such that $C \cap A_{i,n_i}$ is empty. Then $C \cap \bigwedge_n \bigvee_i A_{i,n_i}$ is void, which contradicts the fact $C \subset B$.

Making use of this proposition, we have no difficulty in establishing the

9. THEOREM. *Let $\mathscr{A}$ be a complete Boolean algebra with a strictly positive measure m. Then there is a strictly positive countably additive measure on $\mathscr{A}$ if and only if $\mathscr{A}$ is weakly countably distributive.*

Proof. Horn and Tarski [2] have established the fact that each algebra (not necessarily complete) which has a strictly positive countably additive measure is weakly countably distributive. On the other hand, let m be a strictly positive measure on $\mathscr{A}$ and let $\mathscr{B}$ be the σ-algebra of subsets of X which is generated by $\mathscr{A}$. Then m has a unique extension n which is a countably additive measure on $\mathscr{B}$. (This may be established by using m and coverings by members of $\mathscr{A}$ to define an outer measure p on $\mathscr{B}$, and showing that the σ-algebra of p-measurable sets contains $\mathscr{A}$; or alternatively, one may use m to define a positive linear functional on the class of continuous linear functions on X, and use the Riesz-Kakutani representation theorem for such functionals.) Let $b = \sup \{n(B) : B \in \mathscr{B}$ and B nowhere dense$\}$. We assert that this supremum is attained, for if $\{B_i\}$ is a sequence such that $b = \sup \{n(B_i) : i \in \omega\}$ then $D = \bigcup \{B_i : i \in \omega\}$ is also nowhere dense, in view of our hypothesis, and clearly $b = n(D)$. Now define n_c by $n_c(A) = n(A - D)$. Then, in view of its definition, n_c vanishes at every nowhere dense set. Moreover, n_c is strictly positive on $\mathscr{A}$, for if A is a non-void member of $\mathscr{A}$ there is a non-void member B of $\mathscr{A}$ such that $B \subset A - D$, and $n_c(A) \geq n_c(B) = m(B) > 0$. Finally, n_c is countably additive on $\mathscr{A}$, for if $\{A_i\}$ is a disjoint sequence in $\mathscr{A}$ then the set $E = \bigvee \{A_i : i \in \omega\} - \bigcup \{A_i : i \in \omega\}$ is nowhere dense, hence $n_c(E) = 0$, and $n_c(\bigvee \{A_i : i \in \omega\}) = n_c(\bigcup \{A_i : i \in \omega\}) = \sum \{n_c(A_i) : i \in \omega\}$. The theorem is then proved.

Souslin lines. We first review a few definitions. A *linear continum* is a non-void set X with a linear ordering $>$ such that X has a first and a last element and such that there are no jumps or gaps. In terms of the order topology, these requirements can be stated: X is compact and connected. A linearly ordered set X is of *real type* iff there is an order isomorphism of X onto a subset of the class of real numbers. The *interval algebra* of a linearly ordered set is the Boolean algebra generated by the half-open intervals $(a : b] = \{x : a < x \leq b\}$, where a and b are arbitrary points of X. It is well known (and easy to prove) that a linear continuum X is of real type iff the order topology is separable (that is, there is a countable dense subset of X), and this is the case iff there is a strictly positive measure on the interval algebra of X. If the interval algebra satisfies the countable chain condition, then X is said to satisfy the *countable interval* condition.

We shall call a linear continuum X a *Souslin line* iff X satisfies the countable interval condition, and no non-void interval in X is of real type. We now show that, if the Souslin hypothesis fails, then there exists a Souslin line.

10. PROPOSITION. Let X be a linear continuum which satisfies the countable interval condition but is not of real type. If R is the relation $\{(y, y) : x = y$ or the interval $[y, x]$ is of real type or the interval $[x : y]$ is of real type$\}$, then the quotient X/R is, with the induced order, a Souslin line.

Proof. As a preliminary we show that if $\mathscr{T}$ is a family of intervals in X which is linearly ordered by inclusion, and if each member of $\mathscr{T}$ is of real type, then $\bigcup\{T : T \in \mathscr{T}\}$ is also of real type. If $\mathscr{T}$ is countable this result follows from the fact that a countable union of separable subsets of X is itself separable. But $\mathscr{T}$ may always be assumed to be countable for: we may by transfinite induction choose a subfamily $\{T_a\}$ of $\mathscr{T}$ which is well ordered by $\subset$ and covers $\bigcup\{T : T \in \mathscr{T}\}$, and such a well ordered family must be countable since otherwise the class of sets of the form $T_{a+1} - T_a$ yields an uncountable disjoint family of intervals.

It follows easily from the above that an equivalence class modulo the relation R is either a closed interval or consists of a single point, and that the family $\mathscr{T}$ of such intervals must be countable and disjoint. The quotient map Q of X onto X/R is then continuous relative to the order topologies for X and X/R, and X/R is therefore compact and connected. Finally, suppose that I is a separable interval in X/R, that A is a countable dense subset of I, and that B is a countable dense subset of $\bigcup\{F : F \in \mathscr{T})$. Let C be a countable subset of X such that $B \subset C$ and C intersects $Q^{-1}[x]$ for each x in A. We assert that the closure C^- of C contains $Q^{-1}[I]$, for: if J is an open interval disjoint from C^- then J is disjoint from $\bigcup\{F : F \in \mathscr{T}\}$, hence Q is a homeomorphism on J, whence $Q[J]$ is disjoint from A, and therefore $Q[J]$

is disjoint from I. Thus $Q^{-1}[I]$ is separable, and this, in view of the definition of R, contradicts the fact that I is a proper interval. Consequently X/R can contain no separable interval, and is therefore a Souslin line.

It is now easy to show that a Souslin line has very curious properties. Recall that a *regular open* set is an open set which is the interior of its closure. For a compact Hausdorff space (or for any space which is of the second category at each of its points) the Boolean algebra of regular open sets is naturally isomorphic to the algebra of Borel sets modulo the ideal of Borel sets of the first category, for: the class of all Borel sets A such that the Boolean sum of A and some first category set B is regular and open is easily seen to contain the Borel algebra, and no non-void regular open set is of category I.

11. LEMMA. *Let X be a Souslin line. Then a subset A of X is separable if and only if it is nowhere dense, and this is the case if and only if the set A is of the first category.*

Proof. A separable subset of X is nowhere dense in view of the definition of a Souslin line; conversely, if A is nowhere dense the set E of endpoints of intervals complementary to A^- is countable, and choosing a member of A between each pair of points of E whenever possible yields a countable dense subset of A. Finally, each nowhere dense set is of the first category, and the countable union of sets which are nowhere dense, hence separable, is separable, and hence nowhere dense.

12. THEOREM. *The algebra $\mathscr{A}$ of regular open subsets of a Souslin line X has the properties:*
(i) each disjoint subfamily of $\mathscr{A}$ is countable,
(ii) the algebra $\mathscr{A}$ is complete,
(iii) the algebra $\mathscr{A}$ is weakly countably distributive,
(iv) there is no strictly positive measure on $\mathscr{A}$, and in fact
(v) if m is any measure on $\mathscr{A}$ then there is a countable subfamily $\mathscr{B}$ of $\mathscr{A}$ such that $m(B) = 0$ for B in $\mathscr{B}$ and $\bigvee \{B : B \in \mathscr{B}\} = X$.

Proof. The assertion (i) is clear, and (ii) follows from (i) together with the fact that $\mathscr{A}$ is countably complete. To show that $\mathscr{A}$ is weakly countably distributive let us consider $\mathscr{A}$ as the algebra of Borel subsets of X modulo sets of the first category, and suppose that $A_{i,j}$ is a double sequence of Borel sets such that $A_{i,j} \supset A_{i,j+1}$ for all i and j. We may suppose that each $A_{i,j}$ is closed. It is always the case that $B = \bigwedge \{\bigvee \{A_{i,n_i} : i \in \omega\} : n \in \omega^\omega\} \supset \bigvee \{\bigwedge \{A_{i,j} : j \in \omega\} : i \in \omega\} = C$. Let us suppose that I is a non-void open interval whose intersection with C is of the first category, (and hence nowhere dense). Then I contains a non-void open interval J whose closure is disjoint from C. Via compactness, choose for each i an integer n_i such that A_{i,n_i} fails to intersect J. Then B, in view of its definition, is disjoint from J, and we have showed that $B \cap (X - C)$ is nowhere dense. Thus B is congruent to C modulo the class of sets of the first category, and $\mathscr{A}$ is weakly countable distributive.

If there were a strictly positive measure on $\mathscr{A}$ then the map carrying each point x of X into the measure of the interval $[a, x]$, where a is the first point of X, would be an order isomorphism of X into a set of real numbers. Thus (iv) is proved, and assertion (v) follows from a simple argument based on the fact that every closed interval in X is itself a Souslin line.

REMARK. Part (iv) above may easily be strengthened; for example: there is clearly no strictly monotonic real valued function on $\mathscr{A}$.

It is not difficult to give a precise equivalence to the Souslin hypothesis in terms of properties of Boolean algebras. Let us call a maximal chain in a Boolean algebra (that is, a maximal linearly ordered subclass) a *segment*. A Boolean algebra is *atomless* if and only if each non-zero element is the sum of two disjoint non-zero elements.

13. THEOREM. *The following statements are equivalent:*
(i) (the Souslin hypothesis) *each linear continuum which satisfies the countable interval condition is of real type,*
(ii) *each segment in an atomless Boolean algebra which satisfies the countable chain condition is of real type, and*
(iii) *if $\mathscr{A}$ is the algebra of regular open subsets of a linear continuum, and if $\mathscr{A}$ is complete, atomless, weakly countably distributive, and satisfies the countable chain condition, then each segment in $\mathscr{A}$ is of real type.*

Proof. We first show that (i) implies (ii). If $\mathscr{S}$ is a segment in atomless Boolean algebra $\mathscr{A}$ which satisfies the countable chain condition, then it is easily seen that there is no uncountable disjoint family of intervals in $\mathscr{S}$. Moreover, since $\mathscr{A}$ is atomless, there are no gaps in $\mathscr{S}$ (that is, between any two distinct members of $\mathscr{S}$ there is a third member which is distinct from both). It follows that the (order) completion of $\mathscr{S}$ is a linear continuum satisfying the countable interval condition, and, assuming the Souslin hypothesis, is of real type. Thus $\mathscr{S}$ is of real type.

Clearly (ii) implies (iii). That (iii) implies (i) is an immediate consequence of the properties of a Souslin line, and the fact (Proposition 10) that a Souslin line exists if the Souslin hypothesis fails.

Dominated measures. Maharam has showed [3] that if a Boolean algebra $\mathscr{A}$ satisfies certain conditions then there is a continuous outer measure on $\mathscr{A}$; that is, there is a non-negative real valued function p such that if $A = \bigvee \{A_n : n \in \omega\}$ then $p(A) \leq \sum \{p(A_n : n \in \omega\}$, and, if $\{A_n)$ is a monotonically decreasing sequence with $0 = \bigwedge \{A_n : n \in \omega\}$ then $p(A_n)$ converges to zero. It seems possible that this result might be strengthened by showing that each outer measure dominates a measure. This leads to the general problem of Hahn-Banach type: if $\mathscr{B}$ is a subalgebra of $\mathscr{A}$ and if m is a measure on $\mathscr{B}$ which is dominated there by a non-negative function p such that $p(A) + p(B) \geq p(A \cup B)$, is it then possible to find a measure on $\mathscr{A}$ which is an extension of m and is everywhere dominated by p? The extension theorem just proposed is false, even for finite algebras. However, a similar result can be established if the premises concerning the function p are strengthened, and, although the result fails to apply to Maharam's theorem, it appears to be of some interest in itself.

14. THEOREM. *Let $\mathscr{A}$ be a Boolean algebra, let p be a non-negative monotonic real valued function on $\mathscr{A}$ such that $p(A)+p(B) \geq p(A \cup B) + p(A \cap B)$ for all members A and B of $\mathscr{A}$, and let m be a measure on a subalgebra $\mathscr{B}$ of $\mathscr{A}$ such that $m(B) \leq p(B)$ for B in $\mathscr{B}$. Then there is a measure n on $\mathscr{A}$, which is an extension of m, such that $n(A) \leq p(A)$ for all A in $\mathscr{A}$.*

Proof. The proof is first reduced, by means of a compactness argument, to the case of a finite Boolean algebra. For each pair $\mathscr{C}$ and $\mathscr{D}$ of finite subalgebras of $\mathscr{A}$ such that $X \in \mathscr{C} \subset \mathscr{B}$ and $\mathscr{C} \subset \mathscr{D}$ let $Q(\mathscr{C}, \mathscr{D})$ be the set of all non-negative functions q on $\mathscr{A}$ which satisfy the requirements: q is finitely additive on $\mathscr{D}$, $q(C) = m(C)$ for C in $\mathscr{C}$, and $q(A) \leq p(A)$ for A in $\mathscr{A}$. The class $Q(\mathscr{C}, \mathscr{D})$ is, by virtue of the Tychonoff product theorem, compact relative to the topology of pointwise convergence on $\mathscr{A}$, and $Q(\mathscr{C}', \mathscr{D}') \supset Q(\mathscr{C}, \mathscr{D})$ if $\mathscr{C}' \subset \mathscr{C}$ and $\mathscr{D}' \subset \mathscr{D}$. Proof of the theorem is equivalent to showing that the intersection of the classes $Q(\mathscr{C}, \mathscr{D})$, for all $\mathscr{C}$ and $\mathscr{D}$, is non-void, and in view of compactness it is sufficient to show that each class $Q(\mathscr{C}, \mathscr{D})$ is non-void.

The problem is then reduced to that of extending m from a subalgebra $\mathscr{C}$ to a finite containing subalgebra $\mathscr{D}$, and we may assume that $\mathscr{D}$ is a minimal algebra properly containing $\mathscr{C}$. In this case, using the known structure of finite Boolean algebras, $\mathscr{C}$ is generated by a finite class $C, C_1, \cdots, C_n$ of disjoint non-void sets, and $\mathscr{D}$ is generated by $D, D', C_1, \cdots, C_n$, where $D \cap D'$ is void and $D \cup D' = C$. The extension of m requires the choice of a number $m(D)$ such that the following inequalities are satisfied:

$m(D) + m(A) \leq p(D \cup A)$ for all A in $\mathscr{C}$ with $A \cap C$ void, and $m(D \cup D') - m(D) + m(B) \leq p(D' \cap B)$ for all B in $\mathscr{C}$ with $B \cup C$ void. Thus extension is possible if and only if $m(D)$ can be chosen so that $m(C) + m(B) - p(D' \cup B) \leq m(D) \leq p(D \cup A) - m(A)$ for all members A and B of $\mathscr{C}$ which are disjoint from C, and this inequality can be attained if the left hand member never exceeds the right hand for all such choices of A and B. Rewriting, the proof reduces to establishing that $m(C) + m(A) + m(B) \leq p(D \cup A) + p(D' \cup B)$ for all members A and B of $\mathscr{C}$ which are disjoint from C. But

$$m(C) + m(A) + m(B) = m(C \cup A \cup B) + m(A \cap B)$$
$$\leq p(C \cup A \cup B) + p(A \cap B) \leq p(D \cup A) + p(D' \cap B),$$

the last inequality being derived from the assumption on p as applied to the sets $D \cup A$ and $D' \cup B$. Thus the extension of m is always possible, and the theorem is proved.

There is a dual to the preceding theorem which may be obtained as follows: Suppose m is a measure on the subalgebra $\mathscr{B}$ and that m *dominates* a non-negative function p such that $p(A) + p(B) \leq p(A \cup B) + p(A \cap B)$ for all A and B in $\mathscr{A}$. Then, setting $q(A) = m(X) - p(X - A)$, it is easily verified that m and q satisfy the conditions of the preceding theorem. There is therefore an extension n of m which is everywhere dominated by q, and it follows that n dominates p. Hence:

15. COROLLARY. *Let $\mathscr{A}$ be a Boolean algebra, let p be a non-negative monotonic real valued function on $\mathscr{A}$ such that $p(A) + p(B) \leq p(A \cup B) + p(A \cap B)$ for all A and B in $\mathscr{A}$, and let m be a measure on a subalgebra $\mathscr{B}$ of $\mathscr{A}$ such that $m(B) \geq p(B)$ for B in $\mathscr{B}$. Then there is a measure n on $\mathscr{A}$, which is an extension of m, such that $n(A) \geq p(A)$ for all A in $\mathscr{A}$.*

ADDENDUM

Since writing this paper I have received the following communications. From Professor A. Horn (June 18, 1959):

A. "···Lemma 11-this property is actually characteristic of Souslin lines: A linear countinuum in which the separable subsets coincide with the nowhere dense subsets is a Souslin line, and conversely. This is true because if we have an uncountable disjoint family of intervals, then a set formed by choosing one point from each interval is nowhere dense and non separable. Thus we have a new and intersting formulation of Souslin's problem····"

B. [···· Incidentally, it is interesting that Theorem 14 is not valid (even for monotonic p), if $\mathscr{B}$ is not a subalgebra and m is merely a partial measure (in the sense of [2]) on $\mathscr{B}$ ···]

From Professor Roman Sikorski:
March 25, 1959.

A. (1) The proof of Theorem 2 can be simplified. Theorem 2 is a particular case of the following general theorem:

(*) Let X be a partially ordered Banach space such that $0 \leq x \leq y$ implies $|x| \leq |y|$. For every convex set S of non-negative elements there exists a functional $f \geq 0$[1] such that $|f| = 1$ and $\inf_{x \in S} f(x) = \inf_{x \in S} |x|$.

Theorem (*) follows immediately from a general theorem on the existence of a functional satisfying a given set of inequalities. This general theorem is due to Mazur and Orlicz (Studia Math. 13, (1953), 137-179). A simple proof of Mazur-Orlicz's theorem was given by me (ibidem, p. 180) and by Ptak (also in Studia Math.).

(2) Your Theorem 9 can be proved simply without using Stone spaces. In fact, suppose that m is a finite measure on a Boolean algbra $\mathscr{B}$. The formula

$m'(A) = \inf (m(A_1) + m(A_2) + \cdots)$ for $A \in \mathscr{B}$

(where inf is extended over all disjoint decompostions $A = A_1 + A_2 + \cdots$) defines a σ -measure on $\mathscr{B}$ (viz. m' is the greatest σ -measure $\leq m$). It is easy to verify that if m is strictly positive and B is weakly σ -distributive, then m' is strictly positive.

The remark (2) is due to Professor Ryll-Nardzewski.
April 3, 1959

B. I would like to inform you that Professor Ryll-Nardzewski has found the following analogue of your Theorem 4 for σ-measures:

(*) There exists a strictly positive σ-measure μ on a Boolean algebra $\mathscr{B}$ if and only if $\mathscr{B} - (0)$ is the union of a sequence $\{\mathscr{B}_n\}$ such that, for every n,

(1) the intersection number $I(\mathscr{B}_n)$ is positive;

(2) if $A_m \subset A_{m+1}$ $(m = 1, 2, \cdots)$ and $A_1 + A_2 + \cdots \in \mathscr{B}_n$, then there exists an m such that $A_m \in \mathscr{B}_n$.

Necessity. Take as $\mathscr{B}_n$ the class of all $A \in \mathscr{B}$ such that $\mu(A) > 1/n$.

Sufficiency. There exists a measure μ'_n such that $\mu'_n(A) \geq I(\mathscr{B}_n)$ for every $A \in \mathscr{B}_n$. Let

$$\mu_n(A) = \inf \lim_m \mu_n'(A_m)$$

where inf is extended over all sequences $A_m \in \mathscr{B}$ such that $A = A_1 + A_2 + \cdots$ and $A_m \subset A_{m+1}$. By definition, μ_n is a σ-measure and $\mu_n(A) \geq I(\mathscr{B}_n)$ for $A \in \mathscr{B}_n$ on account of (2). The σ-measure

[1] i.e. such that $f(x) \geq 0$ for $x \geq 0$.

$$\mu(A) = \frac{1}{2}\mu_1(A) + \frac{1}{2^2}\mu_2(A) + \cdots$$

is strictly positive on B.

REFERENCES

1. J. Dixmier, *Sur certains espaces considérés par M. H. Stone,* Summa Brasil. Math. **2** (1951), 151-182.
2. A. Horn and A. Tarski, *Measures in Boolean algebras,* Trans. Amer. Math. Soc. **64** (1948), 467-497.
3. D. Maharam, *An algebraic characterization of measure algebras,* Annals of Math., **48** (1947), 154-167.

UNIVERSITY OF CALIFORNIA
CAMBRIDGE UNIVERSITY

GENERALIZED RANDOM VARIABLES

JOHN G. KEMENY

We will consider random variables on a denumerably infinite sample space. However, the range $\boldsymbol{R}$ of the random variables will not necessarily be a set of real numbers. In Part I the range will be a subset of a given metric space, and in Part II it will be an arbitrary set. Since each distribution on the sample space determines a distribution on $\boldsymbol{R}$ (for a given random variable), the sample space may be ignored entirely, and we may restrict our attention to distributions on $\boldsymbol{R}$. Thus, instead of discussing means and variances of random variables on the sample space, we will discuss means and variances of distributions on the set $\boldsymbol{R}$.

In classical probability theory $\boldsymbol{R}$ would be a set of real numbers, and the mean and variance of a distribution on $\boldsymbol{R}$ would also be real numbers. Of these restrictions only one will be kept, namely that the variance will always be a non-negative real number. As indicated above, $\boldsymbol{R}$ may be a more general space, and the means will also be selected from more general spaces. The defining property of a mean will be the property of minimizing the variance of the given distribution. It will be shown that these means still have many of the classical properties, though in general means are not unique, and in certain circumstances there may be no mean.

While the mean is classically taken to be a real number, it need not be an element of $\boldsymbol{R}$. For example, the mean of a set of integers may be a fraction. This approach is extended in Part I, where the means may be arbitrary points of a certain metric space $\boldsymbol{T}$, and $\boldsymbol{R}$ is any subset of $\boldsymbol{T}$. Even the form chosen for the variance is the same as in classical probability theory.

In Part II the concept of a random variable and of means is further generalized. Here $\boldsymbol{R}$ is an arbitrary set, and the topological space $\boldsymbol{T}$ from which means are chosen need not be metric and need bear no relation to $\boldsymbol{R}$. The variance is still a numerical function on $\boldsymbol{T}$, but of a much more general form than in Part I. In both frameworks an analogue of the strong law of large numbers is proved, to show that classical results can be generalized to these new kinds of random variables.

In Part III we consider certain generalizations. The positive result in this part is that the restriction to independent random variables in Parts I and II is unnecessary; the results hold for any metrically

Received February 11, 1959. The author wishes to acknowledge the aid of the National Science Foundation through a grant given to the Dartmouth Mathematics Project. He also wishes to thank H. Mirkil for a number of valuable suggestions.

transitive stationary process. There are also certain negative results, showing that some "obvious generalizations" fail.

PART I

We consider a metric space $\boldsymbol{T}$, from which our means will be selected. Let φ be the metric on $\boldsymbol{T}$. The range $\boldsymbol{R} = \{r_i\}$ of our random variables may be any denumerable subset of $\boldsymbol{T}$. We impose one restriction on the space $\boldsymbol{T}$:

(1) The closed spheres in $\boldsymbol{T}$ are compact.

As indicated above, instead of considering the random variables themselves, we will consider only distributions $P = \{p_i\}$ on $\boldsymbol{R}$. For each such distribution we define two numerical functions on $\boldsymbol{T}$:

$$W_P(t) = \sum_i \varphi(r_i, t) \cdot p_i \text{ and } V_P(t) = \sum_i \varphi^2(r_i, t) \cdot p_i .$$

The former may be thought of as a mean distance to t with respect to P, and the latter as a variance computed with respect to t. Both functions are non-negative real valued, with $+\infty$ as a possible value. We will, however, consider only distributions that satisfy the condition:

(2) There is a $t_0 \in \boldsymbol{T}$ such that $V_P(t_0)$ is finite.

DEFINITION. The *variance* of the distribution P is defined as

$v_P = \inf_{t \in \boldsymbol{T}} V_P(t)$. An element t of $\boldsymbol{T}$ is a *mean* of P if $V_P(t) = v_P$. We denote the set of means of P by M_P.

In view of this definition we note that (2) is equivalent to the assumption that P has finite variance. Should P have infinite variance, then all points of $\boldsymbol{T}$ would be means of P according to the definition. While the theorems to be proven below would all be true, they would become trivial. It is possible to give a more sophisticated definition of the mean for the case of an infinite variance (see [3]), but this leads into problems beyond the scope of the present paper.

LEMMA 1. *If $W_P(t_1)$ and $V_P(t_1)$ are finite, then*

$$| W_P(t_1) - W_P(t_2)| \leq \varphi(t_1, t_2) \text{ and}$$
$$| V_P(t_1) - V_P(t_2)| \leq \varphi(t_1, t_2) \cdot [2V_P(t_1) + 2 + \varphi(t_1, t_2)] .$$

Proof. $$| W_P(t_1) - W_P(t_2)| \leq \sum_i |\varphi(r_i, t_1) - \varphi(r_i, t_2)| \cdot p_i \leq \sum_i \varphi(t_1, t_2) \cdot p_i = \varphi(t_1, t_2)$$

using the triangle inequality on φ.

$$
\begin{aligned}
|V_P(t_1) - V_P(t_2)| &\le \sum_i |\varphi^2(r_i, t_1) - \varphi^2(r_i, t_2)| \cdot p_i \\
&\le \varphi(t_1, t_2) \cdot \sum_i [\varphi(r_i, t_1) + \varphi(r_i, t_2)] \cdot p_i \\
&= \varphi(t_1, t_2) \cdot [W_P(t_1) + W_P(t_2)] \\
&\le \varphi(t_1, t_2) \cdot [2W_P(t_1) + \varphi(t_1, t_2)]
\end{aligned}
$$

where in the second step we factored and applied the triangle inequality, while in the last step we used the result proved above. The second part of the lemma then follows if we observe that $W_P(t) \le V_P(t) + 1$ for any t.

LEMMA 2. *$W_P(t)$ and $V_P(t)$ are finite for all t.*

Proof. This is a consequence of the restriction (2). Choose t_0 as in (2). Then $W_P(t_0)$ is also finite. Lemma 1 yields that $|V_P(t_0) - V_P(t)|$ is finite, hence the lemma follows.

LEMMA 3. *$V_P(t)$ is a continuous function of t.*

Proof. Suppose that $t_1, t_2, \cdots$ is a sequence converging to t; then by Lemma 1, $|V_P(t) - V_P(t_k)| \le \varphi(t, t_k) \cdot [2V_P(t) + 2 + \varphi(t, t_k)]$. But $\varphi(t, t_k) \to \varphi(t, t) = 0$, hence $V_P(t_k) \to V_P(t)$.

There is one closed sphere that will occur frequently below. Let $S_1 = \mathrm{Sph}\,(\sqrt{6V_P(r_1)/p_1}, r_1)$, that is the set of all points in $\boldsymbol{T}$ whose φ-distance from r_1 is at most the specified amount. Then S_1 is compact by (1).

THEOREM I. *M_P is a non-empty compact set.*

Proof. If $t \notin S_1$, then $V_P(t) \ge \varphi^2(r_1, t) \cdot p_1 > 6V_P(r_1)$. Hence $V_P(t)$ is bounded away from v_P. Thus the inf of V_P on all of T is the same as on S_1. But V_P is continuous, by Lemma 3, and hence $V_P(S_1)$ is compact. This means that V_P actually takes on its inf on S_1, hence P has at least one mean. Furthermore $M_P = V_P^{-1}(\{v_p\})$, hence it is a closed subset of S_1, and thus compact.

We will suppose that a sequence of point $x_1, x_2, \cdots$ is selected from $\boldsymbol{R}$. The points are selected independently, at random, according to a distribution Q satisfying (2). [We may consider $\boldsymbol{R}$ to be our new sample space, and the x_j to be identity functions on $\boldsymbol{R}$. Then they are independent, identically distributed generalized random variables.] For each n, we associate a distribution $H^n = \{h_i^h\}$ with the first n points in this sequence; namely h_i^n is the fraction of the first n points that are equal to r_i, or h_i^n is the frequency of occurrence of r_i among the first n random variables. It will be convenient to write V_n in place of

V_{H^n}, to write v_n for the variance of H^n, and M_n for the set of means of H^n. Clearly H^n has the property (2).

LEMMA 4. *For any $t \in T$, $V_n(t) \to V_Q(t)$ with probability* 1.

Proof. We may consider $\varphi^2(x_j, t)$, for a fixed t, to be a sequence of ordinary random variables on $\boldsymbol{R}$. They are independent and identically distributed. Their mean is $V_Q(t)$. Since this is finite by Lemma 2, the ordinary strong law of large numbers applies to them. But this states precisely that $V_n(t) \to V_Q(t)$ with probability 1. (See [2], p. 208).

LEMMA 5. *For any compact set C there is probability* 1 *that $V_n(t) \to V_Q(t)$ uniformly on C.*

Proof. Since C is compact and V_Q is continuous, $V_Q \leq A$ on C. For any integer k we can find a finite set of points C_k such that the spheres of radius $1/k$ about points in C_k cover C. Since the union of the sets C_k is denumerable, it follows from Lemma 4 that there is probability 1 that $V_n(t) \to V_Q(t)$ for all points in all the C_k. [The set of sequences on which convergence fails at one point has measure 0, hence the union of all these denumerably many sequences has measure 0, and hence the complement of the union has measure 1.] We restrict ourselves to such sequences of x_j. Let t be any point in C. Select a point $t_k \in C_k$ so that $\varphi(t, t_k) < 1/k$. Then

$$\begin{aligned}
|V_n(t) - V_Q(t)| &\leq |V_n(t) - V_n(t_k)| + |V_n(t_k) - V_Q(t_k)| \\
&\qquad + |V_Q(t_k) - V_Q(t)| \\
&\leq \frac{1}{k}\left[2V_n(t_k) + 2 + \frac{1}{k}\right] + |V_n(t_k) - V_Q(t_k)| \\
&\qquad + \frac{1}{k}\left[2V_Q(t_k) + 2 + \frac{1}{k}\right] \\
&\leq \frac{2}{k}[2A + 3] + 3|V_n(t_k) - V_Q(t_k)|
\end{aligned}$$

where Lemma 1 was used in step 2, and the uniform bound A applied and terms combined in step 3.

Given $\varepsilon > 0$, we choose k large enough to make the first term less than $\varepsilon/2$. Since C_k has only a finite number of elements, for sufficiently large n, $|V_n(t_k) - V_Q(t_k)| < \varepsilon/2$ for all $t_k \in C_k$. Hence for sufficiency large n, $|V_n(t) - V_Q(t)| < \varepsilon$ for all $t \in C$.

LEMMA 6.[1] *With probability* 1, *for sufficiently large* n $M_n \subseteq S_1$.

Proof. By the ordinary strong law of large numbers, $h_1^n \to q_1$ with probability 1; and by Lemma 4, $V_n(r_1) \to V_Q(r_1)$ with probability 1. Hence we can select a sequence with probability 1 on which both events take place. On such a sequence, for sufficiently large n,

$$v_n \leq V_n(r_1) < 2V_Q(r_1) \quad \text{and} \quad h_1^n > q_1/2 .$$

Hence, if $t \notin S_1$, then for sufficiently large n,

$$V_n(t) \geq \varphi^2(r_1, t) \cdot h_1^n > (6V_Q(r_1)/q_1) \cdot (q_1/2) = 3V_Q(r_1) ,$$

which is bounded away from v_n. Hence if $t \notin S_1$, then $t \notin M_n$. And hence $M_n \subseteq S_1$.

We are now in a position to prove a version of the strong law of large numbers. This states that the sequence of sample means converges to the mean of the distribution with probability 1. In our more general framework we do not have unique means, though we do have assurance from Theorem I that the set of means is non-empty both for the samples and for the distribution. We thus want to prove that the sequence of sets M_n converges to the fixed set M_Q with probability 1. As the criterion for convergence we require that every open set containing M_Q should contain almost all M_n. If the means happen to be unique, this is equivalent to ordinary convergence.

THEOREM II. $M_n \to M_Q$ *with probability* 1.

Proof. By Lemma 5, there is probability 1 that $V_n(t) \to V_Q(t)$ uniformly on S_1. By Lemma 6, almost all M_n are subsets of S_1 with probability 1. Hence with probability 1 we may restrict ourselves to x_j-sequences on which both events occur. Let O be an open set containing M_Q. Then $S_1 \cap \tilde{O}$ is compact, and hence V_Q takes on a minimum value v on it. But no mean is in this set, hence $v > v_Q$. Let $m \in M_Q$ and $t \in S_1 \cap \tilde{O}$.

$$V_n(t) - V_n(m) = [V_n(t) - V_Q(t)] + [V_Q(t) - V_Q(m)] + [V_Q(m) - V_n(m)] .$$

From the uniform convergence of V_n we know that for sufficiently large n the first and third terms will both be less than $(v - v_Q)/3$ in absolute value. The middle term is at least $v - v_Q$. Hence for sufficiently large n the difference is positive, and hence $t \notin M_n$. Hence no element of $S_1 \cap \tilde{O}$ is in M_n, and we also know that $M_n \subseteq S_1$ for almost all n. Hence $M_n \subseteq O$ for almost all n.

[1] S_1 is here defined with respect to the Q-distribution, that is, V_Q and q_1 take the place of V_P and p_1 in the definition. This will be the sphere used from here on in Part I.

THEOREM III. *$v_n \to v_Q$ with probability* 1.

Proof. $|v_n - v_Q| \leq |V_n(m_n) - V_Q(m_n)| + |V_Q(m_n) - V_Q(m)|$ if $m_n \in M_n$, $m \in M$.

As in the previous theorem, we may combine Lemmas 5 and 6 to assure that the first term tends to 0 with probability 1. The sequence of m_n's will, with probability 1, have a limit point, by Lemma 6 and the compactness of S_1. And by Theorem II this limit point will, with probability 1, be in M_Q. It then follows from the continuity of V_Q that if we choose as m this limit point, the second term goes to 0 with probability 1.

One interesting set of applications of these theorems may be obtained by choosing for $\boldsymbol{T}$ a metric space with compact spheres, and choosing for φ a suitable function of the metric. If d is the metric, and f is a numerical function such that $f(0) = 0, f' > 0$ and $f'' \leq 0$, then $\varphi(t_1, t_2) = f(d(t_1, t_2))$ is also a metric on $\boldsymbol{T}$. In particular, we may choose $\varphi = d^k$, for $k \leq 1$. The choice of $k = 1$ yields the generalization of the ordinary arithmetic mean, and $k = 1/2$ yields a generalization of the median.

If for $\boldsymbol{T}$ we choose Euclidean n-space, and let $\varphi = d$, then Theorem II yields the classical strong law of large numbers for the case of discrete random variables with a finite variance.

Condition (1) is a natural condition to impose when generalizing results from Euclidean n-space. But it is reasonable to ask whether the condition is really necessary. For example, could one replace it by the assumption that $\boldsymbol{T}$ is locally compact? The following example shows that local compactness does not suffice: Let $\boldsymbol{T} = \boldsymbol{R} \cup \boldsymbol{S}$, with $\boldsymbol{S} = \{s_i\}$ for $i = 1, 2, \cdots$. We introduce the metric φ as follows.

$$\varphi(r_i, r_j) = \varphi(s_i, s_j) = 2(1-\delta_{ij}) \quad \text{and} \quad \varphi(r_i, s_j) = \begin{cases} 1 \text{ if } j \geq i \\ 2 \text{ if } j < i \end{cases}.$$

Let $p_i = 1/2^i$. Then $v_P = 1$, and r_1 is the unique mean. Suppose that H^n is a close approximation of P, with $h_1^n \leq 1/2$. This has positive probability. If i_0 is the last i for which $h_1^n > 0$, then s_j is a mean of H^n for all $j \geq i_0$. Hence M_n does not converge to $M_P = \{r_1\}$. This metric topology, which happens to be discrete, violates condition (1), but $\boldsymbol{T}$ is locally compact.

PART II

We will now consider a more general framework in which $\boldsymbol{R}$ is an arbitrary set, and $\boldsymbol{T}$ any topological space. We will consider the space $\boldsymbol{P}$ of all possible measures $P = \{p_i\}$ on $\boldsymbol{R}$. But since $\boldsymbol{R}$ is an arbitrary denumerable infinite set, we may—without loss of generality—take P to be a measure on the integers. The basic tool in Part I was a numerical function $V_P(t)$ on $\boldsymbol{T}$, for each measure $\boldsymbol{P}$, satisfying certain conditions.

We will again assume that there is a function V_P corresponding to each P in $\boldsymbol{P}$.

We will introduce metrics on two basic spaces. On the space $\boldsymbol{P}$ of all measures on the integers we define $d(P, Q) = \sum_i |p_i - q_i|$. On the space $\boldsymbol{F}$ of all non-negative bounded real-valued functions on $\boldsymbol{T}$ we define $d(f, g) = \sup_{t \in \boldsymbol{T}} |f(t) - g(t)|$.

Our basic assumptions concern the mapping $P \to V_P$ from $\boldsymbol{P}$ into $\boldsymbol{F}$. We require that:

(1) Each image V_P is a function that takes on a minimum on every closed subset of $\boldsymbol{T}$.

(2) The mapping is continuous.

We may then introduce means and variances as in the definition in Part I. We may prove near analogues of the previous theorems.

THEOREM I′. *M_P is non-empty for each $P \in \boldsymbol{P}$.*

Proof. V_P takes on a minimum on every closed subset of $\boldsymbol{T}$, by (1), hence it takes on its minimum on $\boldsymbol{T}$.

We will again consider sequences x_j, selected independently at random according to a distribution Q. We define the sample distributions, means, and variances as in Part I.

LEMMA. *$H^n \to Q$ with probability 1.*

Proof. The lemma asserts that $d(H^n, Q) \to 0$ with probability 1. From the definition of the metric on P we see that this asserts that $\sum_i |h_i^n - q_i| \to 0$ with probability 1. This was proved by Parzen in a paper that has not yet appeared (see [4]).

THEOREM II′. *$M_n \to M_Q$ with probability 1.*

Proof. Let O be an open set containing M_Q. Then $\tilde{O}$ is closed, and hence V_Q takes on a minimum value on it, by (1), say v. Since no mean of Q is in $\tilde{O}$, $v > v_Q$.

Suppose that $H^n \to Q$ in $\boldsymbol{P}$, which occurs with probability 1 by the lemma. Then by (2), $V_n \to V_Q$ in $\boldsymbol{F}$. But this means that $V_n(t)$ converges uniformly to $V_Q(t)$. Let $t \in \tilde{O}$, $m \in M_Q$, then

$$V_n(t) - V_n(m) = [V_n(t) - V_Q(t)] + [V_Q(t) - V_Q(m)] + [V_Q(m) - V_n(m)] .$$

By the uniform convergence of V_n we can make the first and third terms less in absolute value than $(v - v_Q)/3$, for all $t \in \tilde{O}$, for sufficiently large n. The middle term is at least $v - v_Q$, hence for all sufficiently large n the difference is positive, and hence for these n, $M_n \subseteq O$.

THEOREM III′. *$v_n \to v_Q$ with probability* 1.

Proof. $v_n \leq V_n(m) \leq v_Q + |V_n(m) - V_Q(m)|$, for $m \in M_Q$.

And $v_Q \leq V_Q(m_n) \leq v_n + |V_Q(m_n) - V_n(m_n)|$, for $m_n \in M_n$.
Hence, $|v_n - v_Q| \leq \sup_{t \in T} |V_n(t) - V_Q(t)| = d(V_n, V_Q)$.

But this tends to 0 with probability 1, by the lemma and (2).

Let us consider some applications of these theorems. First we will suppose that $V_P(t) = \Sigma \varphi^2(r_i, t) \cdot p_i$, where φ is a numerical function on $\boldsymbol{R} \times \boldsymbol{T}$. This is the nearest analogue we have to Part I. But even in this case the assumptions made in Part II are not comparable to those in Part I. The easiest way to assure that (2) is satisfied is to require that $|\varphi| \leq B$ on $\boldsymbol{R} \times \boldsymbol{T}$. Then V_p is always bounded, and

$$|V_P(t) - V_Q(t)| \leq \sum_i \varphi^2(r_i, t) \cdot |p_i - q_i| \leq B^2 \sum_i |p_i - q_i| .$$

Hence $d(V_P, V_Q) \leq B^2 \cdot d(P, Q)$. Hence the mapping $P \to V_P$ is continuous.

There are various ways of fulfilling (1). One very interesting case is where $\boldsymbol{T}$ is compact and $\varphi(r_i, t)$ is lower semi-continuous on $\boldsymbol{T}$ for each $r_i \in \boldsymbol{R}$. Then every closed subset is compact, and hence a lower semi-continuous function will take on its minimum on it. And V_P is the uniform limit of a sequence of monotone increasing lower semi-continuous functions, hence it itself has this property.

Thus if $\boldsymbol{T}$ is compact, we may choose as φ any function bounded on $\boldsymbol{R} \times \boldsymbol{T}$, such that each $\varphi(r_i, t)$ is lower semi-continuous on $\boldsymbol{T}$. Obvious examples of this may be found by choosing $\boldsymbol{R} \subseteq \boldsymbol{T}$, where $\boldsymbol{T}$ is a compact metric space and φ a continuous function of the distance. Thus we see that if we are willing to assume that $\boldsymbol{T}$ is compact, we are allowed to choose φ in much greater generality than in Part I.

If, in particular, $\boldsymbol{T}$ is a finite metric space, then Theorem II′ has an interesting corollary. Since the topology is discrete in this case, $M_n \to M_Q$ implies that $M_n \subseteq M_Q$ for sufficiently large n. Hence there is probability 1 that for sufficiently long sample sequences all sample means are means of the distribution. If the distribution has a unique mean, then there is probability 1 that all sufficiently large samples have this mean as their unique mean.[2]

Let us now consider an example of a compact space with a bounded φ, where φ is only lower semi-continuous. Let $\boldsymbol{T}$ be the set of all vectors $\{a_i\}$, $i = 1, 2, \cdots$, where $a_i \geq 0$ and $\sum_i a_i \leq 1$. We define the distance $\varphi(A, B)$ between two vectors as $\sum_i |a_i - b_i|$. However, $\boldsymbol{T}$ is

[2] An interesting application of this result is worked out by the author and J. L. Snell in a forthcoming book: It can be shown that there is a "natural" metric for the space of all rankings of k individuals. Thus our result allows certain statistical procedures for rankings.

not compact with respect to the metric topology. So we choose for $\boldsymbol{T}$ a weaker topology, namely the topology of componentwise convergence of vectors. $\boldsymbol{T}$ is compact with respect to this weaker topology, and φ is lower semi-continuous on the resulting topological space. Clearly, $\varphi \leq 2$, hence all our conditions are satisfied, and hence the theorems are applicable.

Let us next consider an example where φ is bounded, but $\boldsymbol{T}$ is not compact. Let $\boldsymbol{R} \subseteq \boldsymbol{T}$, and $\boldsymbol{T}$ an arbitrary topological space. Define $\varphi(r, t) = 1 - \delta_{rt}$. Then $V_P(t) = 1$ if $t \notin \boldsymbol{R}$, and $1 - p_i$ if $t = r_i$. Hence on a set not intersecting $\boldsymbol{R}$ the minimum of 1 is taken on at all points, while otherwise the minimum is taken on where p_i is largest. Hence V_P satisfies condition (1), and we see that M_P is always a non-empty, finite set. It is the set of *modes* of the distribution P.

Finally, let us consider one example where V_P is not of the general form of Part I. Let $\boldsymbol{T}$ be a compact metric space, and $\boldsymbol{R} \subseteq \boldsymbol{T}$. If d is the metric, we define $V_P(t)$ as the inf of $\sum_{i \in J} d(r_i, t) \cdot p_i$ over all sets J such that $\sum_{i \in J} p_i > .9$. Here V_P is lower semi-continuous on the compact set, hence (1) holds. Since d is bounded on $\boldsymbol{T}$, say by B, a change of ε in P will produce a change of at most $B\varepsilon$ in V_P; hence (2) holds.

This result has the following "practical" application. Suppose that a state legislature decides to establish a state university. They may insist that the University service at least 90 percent of the state's population, and that it be in the "most convenient location" for the population. This may be interpreted by introducing as a metric distance between r_i and t the distance a person at r_i has to travel to reach a university located at t. Then we find the mean of the V_P described above, with p_i taken proportional to the population at location r_i. Theorem II′ then states that if the college population is a cross-section of the entire population, and if the university is large enough, then there is an excellent chance that the location "most convenient for the entire population" will be "most convenient for the freshman class" in any given year. While the practicality of this procedure is debatable, it is more reasonable than the location of a state university in the geometric center of the state. It also shows that the theorems of Part II lend themselves to many unorthodox applications.

It is worth remarking that if V_P is lower semi-continuous, then M_P will be closed. So in all the examples discussed above where V_P was a lower semi-continuous function on a compact space, we obtain the full equivalent of Theorem I, since M_P is compact. And in the example of the modes M_P was finite.

But we can't always expect this to happen in the very general framework of Part II. As a matter of fact, M_P may be any subset of

$\boldsymbol{T}$. Let S be a subset of $\boldsymbol{T}$, and define $f(t)$ to be 0 on S and 1 on $\tilde{S}$. If we assign this same function to all distributions, that is $V_P = f$ for all P, then both conditions (1) and (2) are fulfilled. But $M_P = S$.

It may be worth pointing out that the mapping $P \rightarrow V_P$ need not be defined on all of $\boldsymbol{P}$. It suffices if it is defined and continuous on a subspace, as long as this subspace includes all measures having only a finite number of positive p_i. The theorems then apply to measures in the subspace. This extra freedom is convenient in a situation where the desired definition of V_P leads to unbounded functions for certain distributions.

PART III

In conclusion we will show that certain other classical ideas fail to generalize. If X_1, X_2 are two random variables, we can introduce a *mean random variable* X which corresponds to $1/2(X_1 + X_2)$ in the classical case. We define the value of X to be the mean of the values of X_1, X_2, if there is a unique mean. If there is more than one mean, we assume that X is equally likely to take on each of these values. We would at least expect that if X_1 and X_2 have the same unique mean, then X also has this mean. However, Figure 1 shows a distribution on a metric space with eight points (each line represents a unit distance), which provides a counter-example. If X_1 and X_2 each have the distribution of Figure 1, then X has the distribution of Figure 2. While X_1 and X_2 have the unique mean A, X has the unique mean B.

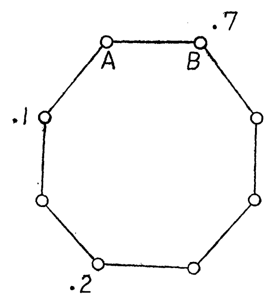


Fig. 1.

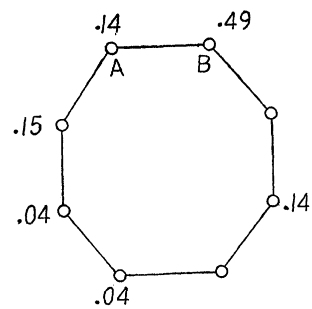


Fig. 2.

Next we will consider classical proofs using Chebyshev's inequality. We may state a version of this inequality, in the terminology of Part I as

$$\Pr[\varphi(x, M_Q) > k] \leq v_Q/k^2 .$$

This inequality may be proved by an exact analogue of the classical argument. However, the usual method for obtaining the weak law of large numbers from it fails. We would need to show that if we define a mean random variable X, as above, its variance tends to 0. However, this is rarely the case. If, for example, R consists of two points, and we have probability 1/2 for each point, then the variance of X tends to 1/2.

On the other hand, it is easy to extend our results to stochastic processes more general than those considered so far. In Parts I and II only identically distributed independent generalized random variables were considered. However, the only property of the process used in Part I was that the strong law of large numbers held. In Part II only the Parzen result was used. Both of these hold for metrically transitive stationary processes (see [1], Ch. X Sects. 1–2, and [4]). Hence all our results hold for these stochastic processes.

Bibliography

1. J. L. Doob, *Stochastic processes*, John Wiley and Sons, 1953.
2. W. Feller, *Probability theory and it applications*, John Wiley and Sons 1950.
3. M. Fréchet, *Les éléments aléatoires de nature quelconque dans un espace distancié*, Institut Henri Poincaré, Vol. X, No. IV, 1948
4. E. Parzen, *On the convergence of distribution functions at points of discontinuity* forthcoming.

Dartmouth College

CONCERNING THE COHOMOLOGY RING OF A SPHERE BUNDLE

D. G MALM

1. Introduction. This paper is concerned with the problem of determining the cohomology ring of an orientable fibre space whose fibere is a sphere, in terms of the cohomology ring of the base space and invariants of the fibre space.

When the fibering sphere is of even dimension $k-1$, an invariant P in the $(2k-2)$-dimensional cohomology group of the base space is defined, which is closely related to one of the Pontrjagin characteristic classes if the fibre space is a fibre bundle. If the $(2k-2)$-dimensional cohomology group of the base space B has no elements of order two, then two $(k-1)$-sphere spaces over B with the same Stiefel-Whitney classes W_k and W_{k-1} and the same invariant P have isomorphic integral cohomology rings.

In the other case, when k is even, if $H^{2k-2}(B, Z)$ has no two-torsion, then two $(k-1)$-sphere spaces over B with the same Stiefel-Whitney classes W_k, W_{k-1}, and W_{k-2} have isomorphic integral cohomology rings.

If $H^{2k-2}(B, Z)$ has elements of order two the situation seems to be more complicated and no results are obtained. Also, the problem of determining the cohomology ring with mod 2 coefficients is not touched upon here.

The method is based upon the algebraic mapping cylinder of the map $x \rightarrow x\mathscr{U}$, where $\mathscr{U}$ is Thom's class, and thus parallels Thom's construction of the Gysin sequence using the mapping cylinder.

In conclusion, I wish to thank Professor W. S. Massey for his generous advice and encouragement in the preparation of this paper, which contains the essential parts of a dissertation submitted to Brown University.

2. Notation and terminology. We define a fibre space as an ordered quadruple (E, p, B, F) such that E, B, and F are topological spaces, $p: E \rightarrow B$ is a continuous map, and such that the following condition holds: For each $x \in B$, there is a neighborhood U of x and a homeomorphism ϕ mapping $U \times F$ onto $p^{-1}(U)$ such that $(p\phi)(y, z) = y$ for each $y \in U$ and $z \in F$. We call E the total space, F the fibre, and B the base space.

By a fibre bundle is meant a fibre space with a structural group, as defined in Steenrod's book [8]. A fibre bundle whose fibre is an

Received January 19, 1959. The author received financial and other support from the National Science Foundation under Grant G1348 during the preparation of part of this paper.

n-sphere and whose group is the group of all $(n+1) \times (n+1)$ real orthogonal matrices of determinant $+1$ (denoted by $SO(n+1)$) will be called an n-sphere bundle. An n-sphere space is a fibre space whose fibre is an n-sphere.

We assume that all n-sphere spaces with which we are concerned satisfy the following orientability condition: If S^n_x denotes the fibre over the point $x \in B$, then the local system of groups defined by $H^n(S^n_x)$, for $x \in B$, is a simple system.

We also assume that the base space of any fibre space or fibre bundle we consider is compact, and we will use Čech-Alexander-Spanier cohomology with compact supports. Unless otherwise indicated, all cohomology groups are with integer coefficients.

In [11], R. Thom showed that the Gysin sequence of a $(k-1)$-sphere space (E, p, B, S^{k-1}) may be obtained in the following manner: There is associated to the given $(k-1)$-sphere space another fibre space (A, p_0, B, F) whose fibre F is a k-cell, for which we may suppose $E \subset A$. (A is the mapping cylinder [10] of $p : E \to B$). Thom showed that there is an element $\mathscr{U} \in H^k(A-E) = H^k(A, E)$ such that the homomorphism $\theta : H^{q-k}(A) \to H^q(A-E)$ defined by

$$\theta(x) = x\mathscr{U} \tag{2.1}$$

(the cup product) is an isomorphism onto. In addition $p_0^* : H^q(B) \to H^q(A)$ is an isomorphism onto. In fact, there is a cross section $s : B \to A$ where $s(x)$ is the center point of the fibre over x, and s^* and p_0^* are inverse to each other. We thus obtain the following commutative diagram of exact sequences, where all the vertical arrows are isomorphisms onto:

$$\begin{array}{ccccccccccc}
\cdots \longrightarrow & H^q(A-E) & \xrightarrow{j^*} & H^q(A) & \xrightarrow{m^*} & H^q(E) & \xrightarrow{\delta^*} & H^{q+1}(A-E) & \longrightarrow \cdots \\
& \uparrow{\scriptstyle\theta} & & \uparrow{\scriptstyle id.} & & \uparrow{\scriptstyle id.} & & \uparrow{\scriptstyle\theta} & \\
\cdots \longrightarrow & H^{q-k}(A) & \xrightarrow{\lambda} & H^q(A) & \xrightarrow{m^*} & H^q(E) & \xrightarrow{\nu} & H^{q-k+1}(A) & \longrightarrow \cdots \\
& \uparrow{\scriptstyle p_0^*} & & \uparrow{\scriptstyle p_0^*} & & \uparrow{\scriptstyle id.} & & \uparrow{\scriptstyle p_0^*} & \\
\cdots \longrightarrow & H^{q-k}(B) & \xrightarrow{\mu} & H^q(B) & \xrightarrow{p^*} & H^q(E) & \xrightarrow{\psi} & H^{q-k+1}(B) & \longrightarrow \cdots
\end{array}$$

Figure 1

Here the homomorphisms λ, μ, ν, and ψ are defined by $\lambda = j^*\theta$; $\mu = p_0^{*-1}\lambda p_0^*$; $\nu = \theta^{-1}\delta^*$; and $\psi = p_0^{*-1}\nu$; the top horizontal sequence is the cohomology sequence of the pair (A, E), and the bottom sequence is the Gysin sequence. Thus according to the results of Thom, the Gysin sequence of (E, p, B, S^{k-1}) is isomorphic to the cohomology sequence of the pair (A, E).

In addition, if we let

$$w_k = j^*(\mathscr{U})$$

and
$$W_k = p_0^{*-1}(w_k)\,,$$

then W_k is the kth Stiefel-Whitney class (the characteristic class) and

$$\mu(x) = xW_k \quad \text{for} \quad x \in H^*(B)\,,$$
$$\lambda(y) = yw_k \quad \text{for} \quad y \in H^*(A)\,.$$

Define w_i by

$$\theta(w_i) = S_q^i(\mathscr{U}) \tag{2.2}$$

where $S_q^i : H^k(A - E) \to H^{k+i}(A - E)$ denotes the Steenrod squaring operation (see [9], or [3], *exposé* 14).* Then also the Stiefel-Whitney classes W_i are given by

$$W_i = p_0^{*-1}(w_i)\,.$$

Thus $W_i \in H^i(B, Z)$ for i odd and $W_i \in H^i(B, Z_2)$ for i even and less than k. W_k is always an integral cohomology class. In addition, $2W_i = 0$ for i odd. For more details, the reader is again referred to the paper of Thom [11].

We will regard $C^*(A, E)$ as a subgroup of $C^*(A)$. It is actually a two-sided ideal in $C^*(A)$, with respect to both the cup product and Steenrod's cup-i products [9]. Since $C^*(E) \approx C^*(A)/C^*(A, E)$ (we are using Alexander-Čech cohomology), we identify these two cochain rings. Note that the map j^* of Figure 1 is then induced by the inclusion $C^*(A, E) \subset C^*(A)$.

The notation introduced here will remain constant, for example, A will always be the mapping cylinder of $p : E \to B$, $\mathscr{U}$ will always be the cohomology class introduced by Thom, etc.

Another important property of the Stiefel-Whitney classes is the following: The Bockstein homomorphism maps the even dimensional ones onto the odd dimensional ones (see [8], p. 195).

Finally, the map ν of Figure 1 satisfies the following equations: If $x \in H^q(A)$ and $y \in H^p(E)$, then

$$\nu[m^*(x) \cdot y] = (-1)^q x \cdot (\nu y)$$

and
$$\nu[y \cdot m^*(x)] = (-1)^{kq}(\nu y) \cdot x\,.$$

This is Lemma 1 of [7].

3. The algebraic mapping cylinder of θ. If (E, p, B, S^{k-1}) is a sphere space, then using $p : H^*(B) \to H^*(E)$, $H^*(E)$ is a module over $H^*(B)$ with the definition $x \cdot y = (p^*x)y$ for $x \in H^*(B)$ and $y \in H^*(E)$. The

* Here, for i even, we let θ operate on $H^*(A, Z_2)$ in the obvious way.

following is an unpublished result of A. Shapiro: If (E, p, B, S^{k-1}) and (E', p', B, S^{k-1}) are $(k-1)$-sphere spaces over the same base space B with the same characteristic class W_k, then $H^*(E)$ and $H^*(E')$ are isomorphic as $H^*(B)$-modules. According to W. S. Massey, this may be proved in the following manner.

Let A be the mapping cylinder of $p: E \to B$ and let $U \in C^*(B)$ be such that $p_0^\# U \in \mathscr{U} \in H^k(A - E)$. Let M be the algebraic mapping cylinder (see [5], page 159, Exercise D) of the map $x \to xU$ for $x \in C^*(B)$, that is

$$M^p = C^p(B) \times C^{p-k+1}(B)\,,$$
$$M = \Sigma_p M^p\,,$$

and $$\delta(x, y) = (\delta x + yU, -\delta y) \text{ for } (x, y) \in M.$$

It is easily seen that (M, δ) is a differential group. In the diagram of exact sequences

$$\begin{array}{ccccccccc} 0 & \longrightarrow & C^p(B) & \xrightarrow{i} & M^p & \xrightarrow{j} & C^{p-k+1}(B) & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle\delta} & & \downarrow{\scriptstyle\delta} & & \downarrow{\scriptstyle\delta} & & \\ 0 & \longrightarrow & C^{p+1}(B) & \xrightarrow{i} & M^{p+1} & \xrightarrow{j} & C^{p-k+2}(B) & \longrightarrow & 0 \end{array}$$

where $i(x) = (x, 0)$ and $j(x, y) = y$, the left square commutes and the right square anti-commutes. We obtain the exact sequence

$$\cdots \longrightarrow H^{p-k}(B) \xrightarrow{\mu} H^p(B) \xrightarrow{i^*} H^p(M) \xrightarrow{j^*} H^{p-k+1}(B) \xrightarrow{\mu} \cdots,$$

where μ is the map induced by $x \to xU$, in other words, μ is the map μ of Figure 1.

Now define $\eta: M^p \to C^p(A)/C^p(A, E) = C^p(E)$ by letting $\eta(x, y)$ be the equivalence class of $p_0^\#(x)$ in $C^p(A)/C^p(A, E)$. η commutes with δ and thus induces $\eta^*: H^p(M) \to H^p(E)$. We then have the diagram

$$\begin{array}{ccccccccccc} \cdots & \longrightarrow & H^{q-k}(B) & \xrightarrow{\mu} & H^q(B) & \xrightarrow{i^*} & H^q(M) & \xrightarrow{j^*} & H^{q-k+1}(B) & \longrightarrow & \cdots \\ & & \downarrow{\scriptstyle id.} & & \downarrow{\scriptstyle id.} & & \downarrow{\scriptstyle\eta^*} & \otimes & \downarrow{\scriptstyle id.} & & \\ \cdots & \longrightarrow & H^{q-k}(B) & \xrightarrow{\mu} & H^q(B) & \xrightarrow{p^*} & H^q(E) & \xrightarrow{\psi} & H^{q-k+1}(B) & \longrightarrow & \cdots \end{array}$$

It is easily verified that all the squares commute except the square marked "$\otimes$", which anti-commutes. By the five-lemma, η^* is an isomorphism onto.

We now make M into a module over $C^*(B)$ by the definition

$$x(v, w) = (xv, (-1)^p xw)$$

for $x \in C^p(B)$ and $(v, w) \in M$. It is easily verified that $\delta\{x(v, w)\} = (\delta x)(v, w) + (-1)^p x\delta(v, w)$ and thus $H^*(M)$ is an $H^*(B)$-module. For

$x \in H^*(B)$ and $y \in H^*(M)$, we have $\eta^*(xy) = (p^*x)(\eta^*y)$, that is, η^* preserves the module product. Consequently $H^*(M)$ and $H^*(E)$ are isomorphic as $H^*(B)$-modules.

Now suppose U' is any other representative for W_k. Then the maps $x \to xU$ and $x \to xU'$ are chain homotopic, and it is easily seen that both algebraic mapping cylinders have isomorphic cohomology as $H^*(B)$-modules. Consequently, both $H^*(E)$ and $H^*(E')$ are isomorphic as $H^*(B)$-modules to $H^*(M)$, where M is the algebraic mapping cylinder obtained from any representative for W_k.

We remark that it is misleading to say that "$H^*(E)$ depends only upon $H^*(B)$ and the characteristic class W_k." It is possible to give examples of polyhedra B_1 and B_2 such that their integral cohomology rings are isomorphic, and then construct 1-sphere bundles $(E_1, p_1, B_1 S^1)$ and $(E_2, p_2, B_2 S^1)$ such that the characteristic classes of the bundles correspond under the isomorphism, yet $H^*(E_1, Z)$ and $H^*(E_2, Z)$ are not isomorphic. The reason is that $H^*(B_1, Z_2)$ and $H^*(B_2, Z_2)$ are non-isomorphic, hence the mod two Gysin sequences are non-isomorphic, and $H^*(E_1, Z_2)$ and $H^*(E_1, Z_2)$ and $H^*(E_2, Z_2)$ do not have the some additive structure.

The theorems that follow concerning the cohomology ring of the total space E will be obtained by introducing a multiplication in the algebraic mapping cylinder and proving that under certain circumstances the map η^* (or rather, a similar map) is a ring homomorphism. For simlicity, we will work with the cochains of A instead of B.

4. Adjusted triples and the multiplication in the mapping cylinder. We wish to define a bilinear function (product) from $M^p \times M^q$ into M^{p+q} which obeys the familiar coboundary formula

$$\delta(\alpha \cdot \beta) = (\delta\alpha) \cdot \beta + (-1)^p \alpha \cdot (\delta\beta)$$

for $\alpha \in M^p$ and $\beta \in M^q$. The problem may be simplified by observing that if (x, y) and (v, w) are elements of M, we require that

$$\begin{aligned}(x, y) \cdot (v, w) &= [(x, 0) + (0, y)][(v, 0) + (0, w)] \\ &= (x, 0)(v, w) + (0, y)(v, 0) + (0, y)(0, w).\end{aligned}$$

Thus we divide the problem into three simpler ones; for each of these products we require that the coboundary formula holds. Furthermore, we know what the first product should be, for we want to preserve the module structure. Thus we want $(x, 0)(v, w) = (xv, (-1)^p xw)$ for $(x, o) \in M^p$. By a careful study of the last two products, we arrive at the following definitions.

DEFINITION 4.1. *An adjusted triple (U, W, N) for the sphere space (E, p, B, S^{k-1}) is a triple of coohains of A for which:*

(i) $U \in \mathscr{U} \in H_k(A - E)$,
(ii) W *is a cochain representing* w_{k-1},
(iii) $\delta N = U \smile_1 U - WU,\ N \in C^{2k-2}(A, E)$,
(iv) $\delta w = \begin{cases} 0 \text{ if } k \text{ is even} \\ -2U \text{ if } k \text{ is odd.} \end{cases}$

DEFINITION 4.2. *Let* (U, W, N) *be adjusted and let* M *be the algebraic mapping cylinder of the map* $x \to xU$. *For* $(x, y) \in M^p$ *and* $(v, w) \in M^q$, *we define*

$$(x, y) \cdot (v, w) =$$
$$(xv + (-1)^{p+q}y(U \smile_1 v) + (-1)^{p+kq+k+1}y[(wU) \smile_2 U] + (-1)^{p+k(q+1)}ywN,$$
$$(-1)^p xw + (-1)^{kq}yv + (-1)^{k+(k+1)q}y(w \smile_1 U) + (-1)^{k+(k+1)q}ywW).$$

We will now prove two propositions which will justify the above definitions.

LEMMA 4.3. *For any sphere space* (E, p, B, S^{k-1}), *there exists an adjusted triple* (U, W, N).

THEOREM 4.4. *The product of Definition* 4.2 *is a bilinear function from* $M^p \times M^q$ *to* M^{p+q} *for which* $\delta[(x, y)(v, w)] = [\delta(x, y)](v, w) + [(-1)^p(x, y)]\delta(v, w)$. *Consequently a product* $H^p(M) \times H^q(M) \to H^{p+q}(M)$ *is induced. In addition, the additive isomorphism* $\eta^* : H^p(M) \to H^p(E)$ *induced by* $\eta : M^p \to C^p(A)/C^p(A, E)$ *where* $\eta(x, y)$ *is the equivalence class of* x *in* $C^p(A)/C^p(A, E)$, *preserves products.*

We prove Lemma 4.3 first. Suppose k is even. Let W be any cocycle representing w_{k-1} and let U be any cocycle representing $\mathscr{U} \in H^k(A - E)$. By equation (2.2), $U \smile_1 U$ and WU represent the same element of $H^{2k-1}(A - E)$. Thus there is a cochain N in $C^*(A, E)$ for which $\delta N = U \smile_1 U - WU$.

Now suppose that k is odd. It is known that in this case, if $\Delta : H^{k-1}(A, Z_2) \to H^k(A, Z)$ is the Bockstein homomorphism, then $\Delta(w_{k-1}) = w_k = -w_k$. Let $U \in \mathscr{U} \in H^k(A - E)$. If W^1 is any integral cochain representing w_{k-1}, then there is a cochain $R \in C^*(A)$ for which $\delta W^1 = -2U + \delta R$, whence $(\delta W^1)U = 2U^2 + (\delta R)U$. From (2.2) we see that $U \smile_1 U = W^1U + \delta N^1 + 2Q$ for some $N^1, Q \in C^*(A, E)$. Taking the coboundary, we get $-2U^2 = (\delta W^1)U + \delta(2Q)$ and so $\delta(2Q) = \delta(-RU)$, that is $2Q + RU$ is a cocycle of $C^*(A, E)$. Since the map θ of (2.1) is an isomorphism, there is a cocycle X of $C^*(A)$ and a cochain $\Xi \in C^*(A, E)$ for which $2Q + RU = XU + \delta\Xi$. Consequently $U \smile_1 U = (W^1 + X - R)U + \delta(N^1 + \Xi)$. By taking cohomology classes, we see that $W^1 + X - R$ represents w_{k-1}, and $(U, W^1 + X - R, N^1 + \Xi)$ is adjusted. This completes the proof of Lemma 4.3.

To prove Theorem 4.4, we note first that the product is clearly bilinear, since the cup product and the cup-i products are. To prove the coboundary formula, we compute

$$\begin{aligned}\delta[(x, y)(v, w)] &= (\delta(xv) + (-1)^{p+q}\delta[y(U\smile_1 v)] + (-1)^{p+kq+k+1}\delta\{y[(wU)\smile_2 U]\} \\ &+ (-1)^{p+k(q+1)}\delta(ywN) + (-1)^p xwU + (-1)^{kq} yvU + (-1)^{k+(k+1)q} y(w\smile_1 U)U \\ &+ (-1)^{k+(k+1)q} ywWU,\ -\{(-1)^p\delta(xw) + (-1)^{kq}\delta(yv) \\ &+ (-1)^{k+(k+1)q}\delta[y(w\smile_1 U)] + (-1)^{k+(k+1)q}\delta(ywW)\})\end{aligned}$$

and

$$\begin{aligned}&[\delta(x, y)](v, w) + [(-1)^p(x, y)]\delta(v, w) = (\delta x + yU,\ -\delta y)(v, w) \\ &\quad + (-1)^p(x, y)(\delta v + wU,\ -\delta w) \\ &= ((\delta x + yU)v + (-1)^{p+q+1}(-\delta y)(U\smile_1 v) + (-1)^{p+kq+k}(-\delta y)[(wU)\smile_2 U] \\ &\quad + (-1)^{p+1+k(q+1)}(-\delta y)wN,\ (-1)^{p+1}(\delta x + yU)w + (-1)^{kq}(-\delta y)v \\ &\quad + (-1)^{k+(k+1)q}(-\delta y)(w\smile_1 U) + (-1)^{k+(k+1)q}(-\delta y)wW) \\ &\quad + (-1)^p(x(\delta v + wU) + (-1)^{p+q+1}y(U\smile_1[\delta v + wU]) \\ &\quad + (-1)^{p+k(q+1)+k+1}y[(-(\delta w)U)\smile_2 U] + (-1)^{p+k(q)}y(-\delta w)N, \\ &(-1)^p x(-\delta w) + (-1)^{k(q+1)}y(\delta v + wU) + (-1)^{k+(k+1)(q+1)}y((-\delta w)\smile_1 U) \\ &\quad + (-1)^{k+(k+1)(q+1)}y(-\delta w)W)\ .\end{aligned}$$

Thus the difference of the first components is

$$\begin{aligned}&(-1)^{p+q}\delta[y(U\smile_1 v)] + (-1)^{p+kq+k+1}\delta\{y[(wU)\smile_2 U]\} + (-1)^{p+k(q+1)}\delta(ywN) \\ &\quad + (-1)^{kq}yvU + (-1)^{k+(k+1)q}y(w\smile_1 U)U + (-1)^{k+(k+1)q}ywWU - yUv \\ &\quad + (-1)^{p+q+1}(\delta y)(U\smile_1 v) + (-1)^{p+kq+k}(\delta y)[(wU)\smile_2 U] + (-1)^{p+1+k(q+1)}(\delta y)wN \\ &\quad + (-1)^q y(U\smile_1(\delta v + wU)) + (-1)^{k(q+1)+k+1}y[(\delta w \cdot U)\smile_2 U] \\ &\quad + (-1)^{kq}y(\delta w)N\ .\end{aligned}$$

We now use the formula

$$\begin{aligned}(4.5)\qquad \delta(u\smile_i v) &= (-1)^{p+q-i}u\smile_{i-1}v + (-1)^{pq+p+q}v\smile_{i-1}u + (\delta u)\smile_i v \\ &\quad + (-1)^p u\smile_i \delta v\end{aligned}$$

for u a p-cochain and v a q-cochain (see [9]), and a formula due to G. Hirsch [6],

$$(4.6)\qquad (uv)\smile_1 w = u(v\smile_1 w) + (-1)^{q(r+1)}(u\smile_1 w)v$$

where v and w are q- and r-cochains respectively. Thus

$$\begin{aligned}(-1)^{p+q}\delta[y(U\smile_1 v)] &= (-1)^{p+q}(\delta y)(U\smile_1 v) \\ &+ (-1)^{q+k+1}y\{(-1)^k U\smile_1(\delta v) + (-1)^{q+k+1}Uv + (-1)^{k+q+kq}vU\}\ .\end{aligned}$$

Also,

$$(-1)^{p+k(q+1)}\delta(ywN) = (-1)^{p+k(q+1)}(\delta y)(wN) + (-1)^{kq+1}y\{(\delta w)N + (-1)^{q+k+1}w(U\smile_1 U - WU)\} .$$

Consequently the difference of the first components is seen to reduce to

$$\begin{aligned}&(-1)^{p+kq+k+1}\delta\{y[(wU)\smile_2 U]\} + (-1)^{k+(k+1)q}y(w\smile_1 U)U \\ &+ (-1)^{p+kq+k}(\delta y)[(wU)\smile_2 U] + (-1)^q y(U\smile_1(wU)) \\ &+ (-1)^{k(q+1)+k+1}y[(\delta w\cdot U)\smile_2 U] + (-1)^{kq+k+q}yw(U\smile_1 U) .\end{aligned}$$

Since

$$\begin{aligned}&(-1)^{p+kq+k+1}\delta\{y[(wU)\smile_2 U]\} = (-1)^{p+kq+k+1}(\delta y)[(wU)\smile_2 U] \\ &+ (-1)^{kq}y\{(-1)^{q+k+1}(wU)\smile_1 U + (-1)^{q+k+1+k(q+1)}U\smile_1(wU) \\ &+ ((\delta w\cdot U)\smile_2 U\} ,\end{aligned}$$

this difference becomes

$$\begin{aligned}&(-1)^{k+(k+1)q}y(w\smile_1 U)U + (-1)^{kq+k+q}yw(U\smile_1 U) \\ &+ (-1)^{kq+q+k+1}y[(wU)\smile_1 U] ,\end{aligned}$$

which is zero by the formula of G. Hirsch.

On the other hand, the difference of the second components of the two expressions is

$$\begin{aligned}&(-1)^{k+(k+1)q+1}\delta[y(w\smile_1 U)] + (-1)^{k+(k+1)q+1}\delta(ywW) + (-1)^p yUw \\ &+(-1)^{k+(k+1)q}(\delta y)(w\smile_1 U) + (-1)^{k+(k+1)q}(\delta y)wW \\ &+ (-1)^{p+k(q+1)+1}ywU + (-1)^{p+k+(k+1)(q+1)}y[(\delta w)\smile_1 U] \\ &+ (-1)^{p+k+(k+1)(q+1)}y(\delta w)W .\end{aligned}$$

But

$$\begin{aligned}\delta[y(w\smile_1 U)] = (\delta y)(w\smile_1 U) + (-1)^{p+k+1}y[(\delta w)\smile_1 U + (-1)^q wU \\ + (-1)^{q+1+(q+k+1)k}Uw] ,\end{aligned}$$

and

$$\delta(ywW) = (\delta y)(wW) + (-1)^{p+k+1}y[(\delta w)\cdot W + (-1)^{q+k+1}w(\delta W)] .$$

Thus this difference reduces to

$$\begin{aligned}&(-1)^{p+k(q+1)+1}ywU + (-1)^{p+kq}ywU + (-1)^{p+k(q+1)+1}yw(\delta W) \\ &= (-1)^{p+kq}[(-1)^{k+1}ywU + ywU + (-1)^{k+1}yw\delta W] = 0 ,\end{aligned}$$

since $\delta W = 0$ for k even and $\delta W = -2U$ for k odd.

Thus the cochain formula holds and a product is induced on the cohomology level. Since $C^*(A, E)$ is an ideal in $C^*(A)$, and since U and N are in $C^*(A, E)$, we see immediately from the definitions of η and the product that η^* preserves products. This completes the proof of Theorem 4.4.

We remark that the product of Definition 4.2 is *not* associative, though, of course, the induced product on the cohomology level is.

5. The invariant P. We now define, when k is odd, an invariant P' of the $(k-1)$-sphere space (E, p, B, S^{k-1}), which is an element of $H^{2k-2}(A, Z)$, and its image $P = p_0^{*-1}(P') \in H^{2k-2}(B, Z)$. P' and P will be called the P'-invariant and the P-invariant, respectively, of the sphere space.

Let (U, W, N) be an adjusted triple. A straightforward computation, using equation (4.5), shows that $W^2 + W\smile_1(\delta W) - 4N - U\smile_2 U$ is a $(2k-2)$-cocycle. We define P' to be its cohomology class in $H^{2k-2}(A, Z)$, and $P = p_0^{*-1}(P')$.

THEOREM 5.1. *Let (E, p, B, S^{k-1}) be a sphere space for which k is odd, and let (U, W, N) and (U', W', N') be adjusted triples for this sphere space. Then $W^2 + W\smile_1(\delta W) - 4N - U\smile_2 U$ and $W'^2 + W'\smile_1(\delta W') - 4N' - U'\smile_2 U'$ represent the same element of $H^{2k-2}(A)$, and consequently P' is independent of the choice of nice triple made in its definition.*

This theorem, which states that P is an invariant of the sphere space, is proved with the help of the following lemma.

LEMMA 5.2. *Let k be odd, and let (U, W, N) be an adjusted triple. Then (U', W', N') is an adjusted triple if, and only if, there exist $\beta \in C^{k-1}(A, E)$, $\gamma \in C^{k-2}(A)$, $\varphi \in C^{k-2}(A)$ a cocycle* mod 2, *and $\rho \in C^{2k-2}(A, E)$ a cocycle, for which*

$$U' = U + \delta\beta ,$$
$$W' = W - 2\beta + \delta(\varphi + \gamma) ,$$

and

$$N' = N + \beta\smile_1 U' - U\smile_1\beta - (\varphi + \gamma)U' + \beta^2 - W\beta + \rho .$$

We first prove that if (U, W, N) and (U', W', N') are adjusted, then there exist β, γ, φ, and ρ with the stated properties. Since U and U' both represent $\mathscr{U}$, there exists $\beta \in C^{k-1}(A, E)$ for which $U' = U + \delta\beta$. Now $\delta W' = -2U' = -2(U + \delta\beta) = \delta W - \delta(2\beta)$, or $W' - W + 2\beta$ is a cocycle. Let $\alpha' = W' - W + 2\beta$. Taking cohomology mod 2 in A, (denoted by brackets) we see that $0 = [W' - W] = [2\beta - \alpha'] = [\alpha']$. Thus there exist $\gamma \in C^{k-2}(A)$ and $\alpha \in C^{k-1}(A)$ for which $\alpha' = \delta(\gamma) + 2\alpha$, and $\delta\alpha = 0$. Then $W' = W + 2(\alpha - \beta) + \delta\gamma$. Now

$$\begin{aligned}\delta(N' - N) &= U'\smile_1 U' - U\smile_1 U + WU - W'U' \\ &= (\delta\beta)\smile_1 U + U\smile_1(\delta\beta) + (\delta\beta)\smile_1(\delta\beta) - 2(\alpha - \beta)U \\ &\quad -\delta(\gamma U) - W(\delta\beta) - 2(\alpha - \beta)\delta\beta - \delta(\gamma\delta\beta) .\end{aligned}$$

Using equation (4.5), we have

$$\delta(\beta\smile_1 U - U\smile_1\beta) = (\delta\beta)\smile_1 U + U\smile_1(\delta\beta) + 2\beta U - 2U\beta\ .$$

Consequently

$$\begin{aligned}\delta(N' - N) = \delta(\beta\smile_1 U - U\smile_1\beta) - \delta(W\beta) + (\delta\beta)\smile_1(\delta\beta) - 2\alpha U\\ -\delta(\gamma U) - 2\alpha(\delta\beta) - \delta(\gamma\delta\beta) + 2\beta\delta\beta\ .\end{aligned}$$

But since $(\delta\beta)\smile_1(\delta\beta) = \delta(\beta\smile_1(\delta\beta)) - \beta\delta\beta + (\delta\beta)\beta$, we have

$$\begin{aligned}\delta(N' - N) &= \delta\{\beta\smile_1 U - U\smile_1\beta - W\beta - \gamma U + \beta^2 - \gamma\delta\beta + \beta\smile_1(\delta\beta)\}\\ &\quad - 2\alpha U'\\ &= \delta\{\beta\smile_1 U' - U\smile_1\beta - \gamma U' - W\beta + \beta^2\} - 2\alpha U'\ ,\end{aligned}$$

which states that $2\alpha U'$ is a coboundary of $C^*(A, E)$, since N', β, U', U, and N are in $C^*(A, E)$. Since θ of (2.1) is an isomorphism, this means that there is $\varphi \in C^{k-2}(A)$ such that $2\alpha = \delta\varphi$, and $-2\alpha U' = -\delta(\varphi U')$. We then have

$$\delta(N' - N) = \delta(\beta\smile_1 U' - U\smile_1\beta - \gamma U' - W\beta + \beta^2)\ .$$

This gives the stated result immediately.

Now suppose (U, W, N) is adjusted and β, γ, φ, and ρ have the stated properties. Then clearly U' represents $\mathscr{U}$ and W' represents w_{k-1}. Also $\delta W' = \delta W - 2\delta\beta = -2U - 2\delta\beta = -2U'$. Finally,

$$\begin{aligned}\delta N' &= U\smile_1 U - WU + \delta(\beta\smile_1 U') - \delta(U\smile_1\beta) - \delta(\varphi + \gamma)U' + \beta\delta\beta\\ &\quad + (\delta\beta)\beta - W\delta\beta + 2U\beta\\ &= U\smile_1 U - WU' - \delta(\varphi + \gamma)U' + (\delta\beta)\smile_1 U' + \beta U' - U'\beta\\ &\quad + U\smile_1\delta\beta - U\beta + \beta U + \beta\delta\beta + (\delta\beta)\beta + 2U\beta\\ &= U'\smile_1 U' - WU' - \delta(\varphi + \gamma)U' + 2\beta U'\\ &= U'\smile_1 U' - W'U'\ .\end{aligned}$$

Consequently (U', W', N') is adjusted. This completes the proof of Lemma 5.2.

We now prove Theorem 5.1. Let (U, W, N) and (U', W', N') be adjusted triples related by β, γ, φ, and ρ as in Lemma 5.2. Let

$$a = W^2 + W\smile_1(\delta W) - 4N - U\smile_2 U$$

and

$$a' = W'^2 + W'\smile_1(\delta W') - 4N' - U'\smile_2 U'\ .$$

Then

$$\begin{aligned}a' - a = W'^2 - W^2 + W'\smile_1(\delta W') - W\smile_1(\delta W) + 4(N - N')\\ + U\smile_2 U - U'\smile_2 U'\ .\end{aligned}$$

Since $U\smile_2 U$ and $U'\smile_2 U'$ both represent $w_{k-2}\mathscr{U}$, we may let $\delta z = U\smile_2 U - U'\smile_2 U'$, for some $z \in C^*(A, E)$. Then it is easily verified, using Lemma 5.2 and equation (4.5), that

$$a' - a = -4\rho + \delta\{z - (\varphi+\gamma)\smile_1(2U') + W'(\varphi+\gamma) - (\varphi+\gamma)\delta(\varphi+\gamma) + (\varphi+\gamma)W' - W\smile_1(2\beta)\} .$$

Thus, taking cohomology in $H^*(A)$, we have $[a'-a] = [-4\rho]$. But ρ is a cocycle of $A - E$ and thus for some $X \in H^*(A)$ we have $[\rho] = X\mathscr{U}$ where the cohomology class of ρ is here taken in $H^*(A-E)$. Consequently, now taking cohomology classes in $H^*(A)$, we have $[-4\rho] = -4(Xw_k)$ which is zero since $2w_k = 0$. This completes the proof of Theorem 5.1.

We now turn to some properties of P. We shall prove the following theorem:

THEOREM 5.3. *Let k be odd, and let $(E, p, B, S^{k-1}, SO(k))$ be a $(k-1)$-sphere bundle, with B a finite polyhedron. If $H^{2k-2}(B, Z)$ has no elements of order two, then $P = P_{2k-2}$, the Pontrjagin class in dimension $2k-2$.*

The hypothesis that the fibre space admit $SO(k)$ as structural group is needed in order that the Pontrjagin class be defined.

The proof of this theorem requires several lemmas and the use of the universal Gysin sequence.

We recall [8] that given any topological group G, there exists a universal principal G-bundle (E_G, p, B_G, G, G) which has the following property:

Given a polyhedron B, any principal G-bundle over B is isomorphic to the bundle induced by some map $f: B \to B_G$. B_G is called the classifying space for G.

Suppose now that G_0 is a closed subgroup of G. The following lemma is proved by H. Cartan in [3], *exposè* 7.

LEMMA 5.4. *If (E, p, B, G, G) is a principal G-bundle, and $\pi: E/G_0 \to E/G = B$ is the natural projection, then $(E/G_0, \pi, B, G/G_0, G)$ is a fibre bundle which is associated with (E, p, B, G, G), where G operates on G/G_0 in the natural way.*

It is known that if (E_G, p, B_G, G, G) is a universal principal G-bundle, and G_0 is a closed subgroup of G, then in the associated fibre bundle $(E_G/G_0, \pi, B, G/G_0, G)$ given by Lemma 5.4, the total space E_G/G_0 is of the same homotopy type as the classifying space B_{G_0}. For a proof, see [7], Lemma 6. Taking $G = SO(k)$, $G_0 = SO(k-1)$, we have $G/G_0 = S^{k-1}$.

We will call $(B_{SO(k-1)}, \pi, B_{SO(k)}, S^{k-1}, SO(k))$ the universal $(k-1)$-sphere bundle. It has the following pleasant property: Any bundle

$(E, p, B, S^{k-1}, SO(k))$ is isomorphic to the induced bundle $f^{-1}(B_{SO(k-1)}, \pi, B_{SO(k)}, S^{k-1}, SO(k))$ for some map $f: B \to B_{SO(k)}$. This follows from the fact that the operation of taking induced bundles and of taking associated bundles commute. This is easily proved if one uses the definition of "induced bundle" and "associated bundle" in terms of the coordinate transformations ([8]).

LEMMA 5.5. *Let $(E', p', B', S^{k-1}, SO(k))$ be a $(k-1)$-sphere bundle, with P-invariant β', and let $(E, p, B, S^{k-1}, SO(k))$ be the bundle induced by $f: B \to B'$, with P-invariant β. Then $f^*(\beta') = \beta$.*

Proof. Let $F: E \to E'$ be the map of the total spaces corresponding to f, so that the following diagram is commutative:

$$\begin{array}{ccc} E & \xrightarrow{F} & E' \\ \downarrow{\scriptstyle p} & & \downarrow{\scriptstyle p'} \\ B & \xrightarrow{f} & B' \end{array}$$

This diagram may be imbedded in a commutative diagram

$$\begin{array}{ccc} A & \xrightarrow{\mathscr{F}} & A' \\ \uparrow{\scriptstyle i} & & \uparrow{\scriptstyle i'} \\ E & \xrightarrow{F} & E' \\ \downarrow{\scriptstyle p} & & \downarrow{\scriptstyle p'} \\ B & \xrightarrow{f} & B' \end{array}$$

where A and A' are the mapping cylinders of p and p' respectively, and i and i' are inclusion maps. A is a quotient space of $(E \times I) \cup B$, where I is the closed unit interval, and similarly for A'. Letting square brackets denote equivalence classes in the quotient spaces, $\mathscr{F}$ is defined by

$$\mathscr{F}[(x, t)] = [(Fx, t)] \qquad \text{for } x \in E,\ t \in I,$$

and

$$\mathscr{F}[b] = [fb] \qquad \text{for } b \in B.$$

Also $i(e) = [(e, 0)]$ for $e \in E$. It is easily verified that $\mathscr{F}$ is a continuous function and the diagram commutes. Let $\mathscr{F}^{\#}: C^*(A') \to C^*(A)$ be the cochain homomorphism induced by $\mathscr{F}$.

Passing to the cochain level we have the commutative diagram

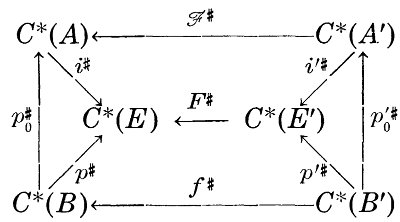


Figure 2

where $p_0 : A \to B$ and $p_0' : A' \to B'$ are the projections.

Now let (U', W', N') be an adjusted triple for (E', p', B', S^{k-1}). Let $U = \mathscr{F}^\#(U')$, $W = \mathscr{F}^\#(W')$, and $N = \mathscr{F}^\#(N')$. Clearly $\mathscr{F}^\#(C^*(A', E')) \subset C^*(A, E)$. Since the Stiefel-Whitney classes and Thom's class $\mathscr{U}$ are preserved by f (or $\mathscr{F}$), we see that (U, W, N) is an adjusted triple for (E, p, B). A representative cocycle for the P'-invariant of (E', p', B') is $W'^2 + W' \smile_1 \delta W' - 4N' - U' \smile_2 U'$, under $\mathscr{F}^\#$ this goes into $W^2 + W \smile_1 \delta W - 4N - U \smile_2 U$, a representative cocycle for the P'-invariant of (E, p, B). Consequently $f^*(\beta') = \beta$, in view of the commutativity of Figure 2.

The following two lemmas together imply Theorem 5.3.

LEMMA 5.6. *Let k be odd, and $(E, p, B, S^{k-1}, SO(k))$ a $(k-1)$-sphere bundle, with B a finite polyhedron. Using the rationals or the integers mod n, n odd, as coefficients for cohomology, $P = P_{2k-2}$, the Pontrjagin class in dimension $2k-2$ with rational or mod n coefficients.*

LEMMA 5.7. *Let G be a finitely generated abelian group with no elements of order two. Let $a \in G$ be such that for each odd integer n there is an $\alpha \in G$ for which $a = n\alpha$. Then $a = 0$.*

We omit the proof of Lemma 5.7, which is quite simple.

Proof of Lemma 5.6. In view of Lemma 5.5, it suffices to prove Lemma 5.6 for the universal $(k-1)$-sphere bundle $(B_{SO(k-1)}, \pi, B_{SO(k)}, S^{k-1}, SO(k))$.

Since the base space B of our bundle is a finite polyhedron, we need only use an n-universal bundle for sufficiently large n. For this bundle, the base space may be chosen to be compact (see [8], Section 19), and we may use Alexander-Spanier cohomology with compact supports.

Let W_k be the characteristic class of this bundle, thus $W_k \in H^k(B_{SO(k)})$, and let $W_{k-1}(k-1) \in H^{k-1}(B_{SO(k-1)})$ be the universal Euler-Poincaré class (for the cohomology of the classifying spaces see the article by A. Borel [1]; for a review of the results we need, see the article by W. S. Massey [7]). Since k is odd, $2W_k = 0$ and $W_k = 0$.

Choose (U, W, N) adjusted for this sphere bundle and let M be the algebraic mapping cylinder associated with this adjusted triple, with a multiplication defined as in §4. We then have the following commutative diagram of exact sequences, where all the vertical arrows are isomorphisms onto (the square marked "$\otimes$" anticommutes). The notation is that used previously.

$$
\begin{array}{ccccccccc}
0 & \xrightarrow{\lambda} & H^q(A) & \xrightarrow{i^*} & H^q(M) & \xrightarrow{j^*} & H^{q-k+1}(A) & \xrightarrow{\lambda} & 0 \\
 & & \Big\downarrow id. & & \Big\downarrow \eta^* & \otimes & \Big\downarrow id. & & \\
0 & \xrightarrow{\lambda} & H^q(A) & \xrightarrow{m^*} & H^q(B_{SO(k-1)}) & \xrightarrow{\nu} & H^{q-k+1}(A) & \xrightarrow{\lambda} & 0 \\
 & & \Big\uparrow \pi_0^* & & \Big\uparrow id. & & \Big\uparrow \pi_0^* & & \\
0 & \xrightarrow{\mu} & H^q(B_{SO(k)}) & \xrightarrow{\pi^*} & H^q(B_{SO(k-1)}) & \xrightarrow{\psi} & H^{q-k+1}(B_{SO(k)}) & \xrightarrow{\mu} & 0
\end{array}
$$

Fig. 3.

In what follows, an integer n is to be taken as $n\omega$ or $\bar{n}\omega'$ if the coefficients are the rationals or the integers mod an odd integer respectively. Here, ω is the unit of $C^*(A, \text{rationals})$, ω' is the unit of $C^*(A, Z_{2m+1})$, and $\bar{n}$ is n reduced mod $2m+1$.

We note that $(W, 2)$ is a cocycle of M^{k-1} and compute that $(W, 2)^2 = (W^2 - (\delta W)\smile_1 W + 4U\smile_2 U - 4N, 0)$. Since $(\delta W)\smile_1 W = \delta(W\smile_1 W) - W\smile_1(\delta W)$, we have

$$(W, 2)^2 = (Z + 5U\smile_2 U - \delta(W\smile_1 W), 0)$$

where $Z = W^2 + W\smile_1(\delta W) - 4N - U\smile_2 U$ is a representative cocycle for the invariant $P' \in H^{2k-2}(A)$. Since U is a coboundary, $5U\smile_2 U$ is and

$$i^*(P') = [(W, 2)^2] \in H^{2k-2}(M)\,,$$

where the square brackets denote cohomology classes. Thus $m^*(P') = (\eta^* i^*)(P') = \eta^*[(W, 2)]\eta^*[(W, 2)] = [W]^2 \in H^{k-1}(B_{SO(k-1)})$, and $(\pi^* \circ \pi_0^{*-1})(P') = [W]^2$, or $\pi^*(P) = [W]^2$. We now need the following lemmas.

LEMMA 5.8. *With integral coefficients, $\psi(W_{k-1}(k-1))$ is twice a generator of $H^0(B_{SO(k)})$.*

LEMMA 5.9. *With integral coefficients, $\pi^*(P_{2k-2}) = (W_{k-1}(k-1))^2$.*

For proofs, see [7], Lemmas 7 and 8.

Thus, using the rationals or the integers mod an odd integer for coefficients, we have $\pi^*(P_{2k-2}) = (W_{k-1}(k-1))^2$. Now π^* is an isomorphism and we complete the proof that $P = P_{2k-2}$ by showing that $(W_{k-1}(k-1))^2 = [W]^2$.

By Lemma 5.8, we may choose $\varepsilon = \pm 1$ so that $\nu(\varepsilon W_{k-1}(k-1)) = -2 \in H^0(A)$. But $\nu([W]) = -j^*[(W, 2)] = -2$. By exactness there is a $y \in H^*(A)$ such that $\varepsilon W_{k-1}(k-1) = [W] + m^*(y)$. Multiplying by $[W]$

and $\varepsilon W_{k-1}(k-1)$ respectively, we get

$$\varepsilon W_{k-1}(k-1)[W] = [W]^2 + m^*(y)[W]$$

and

$$(W_{k-1}(k-1))^2 = \varepsilon W_{k-1}(k-1)[W] + m^*(y)\varepsilon W_{k-1}(k-1) .$$

Together these give us

$$[W]^2 - (W_{k-1}(k-1))^2 = -m^*(y)([W] + \varepsilon W_{k-1}(k-1)) .$$

We now apply ν to this equation, remembering that $[W]^2$ and $W_{k-1}(k-1))^2$ are in image $m^* = \text{kernel}\, \nu$. Then

$$0 = -y\nu([W] + \varepsilon W_{k-1}(k-1)) = 4y .$$

Thus $y = 0$ and $(W_{k-1}(k-1))^2 = [W]^2$.

It is possible to prove the following theorem, which immediately implies Lemma 5.6.

If k is odd, and $(E, p, B, S^{k-1}, SO(k))$ is a $(k-1)$-sphere bundle with $W_k = 0$ and B a polyhedron, then $P = P_{2k-2}$, the Pontrjagin class in dimension $2k-2$.

This is a direct consequence of Theorem IV of [7]. It is only necessary to prove that P is Massey's invariant $4\alpha + \beta^2$, which can be done by a computation in the mapping cylinder.

According to W. T. Wu [12], for a $(k-1)$-sphere bundle, if $\mathscr{P}^2$ denotes the Pontrjagin squaring operation, then

$$P_{2k-2} = \mathscr{P}^2(W_{k-1}) - W_{k-2}W_k ,$$

reduced mod 4. If (U, W, N) is adjusted for the sphere bundle, $W^2 + W \smile_1 (\delta W)$ represents $\mathscr{P}^2(w_{k-1})$ and $U \smile_2 U$ represents $w_{k-2}w_k$. Consequently,

$$P = P_{2k-2} \text{ reduced mod } 4 .$$

Let G_{k-1} denote the group of all homeomorphisms of S^{k-1}, and $B_{G(k-1)}$ the classifying space for G_{k-1}. It would be of interest to know whether the invariant P comes from a cohomology class in $H^{2k-2}(B_{G_{k-1}})$.

6. The main theorem for *k* odd. In this section we assume that k is odd and (E, p, B, S^{k-1}) is a $(k-1)$-sphere space. We consider the effect of dropping the conditions that N and U be in $C^*(A, E)$, where (U, W, N) is an adjusted triple for (E, p, B, S^{k-1}). A check of the proof of Theorem 4.4 shows that the product of Definition 4.2 still induces a product in the mapping cylinder. However, in general η^* no longer preserves products. To retrieve (in part) this property of η^* we add a requirement that U, W, and N be connected with the invariant P'. Before stating the main theorem we require several lemmas.

LEMMA 6.1. *Let $U \in w_k \in H^k(A)$, $Z \in P' \in H^{2k-2}(A)$, and let W be any integral cochain representing w_{k-1} for which $\delta W = -2U$. Then there exist $N \in C^{2k-2}(A)$ and $Q \in C^{2k-3}(A)$ for which*

$$4N + \delta Q = W^2 + W \smile_1 (\delta W) - Z - U \smile_2 U .$$

Proof. Let (U', W', N') be adjusted, and let $Z' = W'^2 + W' \smile_2 (\delta W') - 4N' - U' \smile_2 U$, a representative cocycle for P'. Then there are cochains α, β, and γ for which $U = U' + \delta\beta$, $W = W' + \delta\gamma - 2\beta$, and $Z = Z' + \delta\alpha$. Let

$$N = N' - W'\beta + \beta \smile_1 U' + \beta^2 - U' \smile_1 \beta - (\delta\gamma)\beta + \beta \smile_1 (\delta\beta) - \gamma U'$$

and

$$Q = -\gamma \smile_1 (2U') - 2W' \smile_1 \beta - 2(\delta\gamma) \smile_1 \beta - \alpha + U' \smile_2 \beta - \beta \smile_2 U' + (\delta\beta) \smile_2 \beta + \beta \smile_1 \beta + \gamma W' + W'\gamma + (\delta\gamma)\gamma .$$

A straightforward computation of $4N + \delta Q$ completes the proof.

We now prove a similar lemma for the cochains of B instead of A. The fibre space (A, p_0, B, k-cell) has a cross section $s : B \to A$. On the cochain level we have

$$C^*(B) \underset{p_0^\#}{\overset{s^\#}{\rightleftarrows}} C^*(A)$$

with $s^\# \circ p_0^\#$ the identity.

LEMMA 6.2. *Let $U \in W_k \in H^k(B)$, $Z \in P \in H^{2k-2}(B)$, and let W be any integral cochain representing W_{k-1} for which $\delta W = -2U$. Then there exist $N \in C^{2k-2}(B)$ and $Q \in C^{2k-3}(B)$ such that*

$$4N + \delta Q = W^2 + W \smile_1 (\delta W) - Z - U \smile_2 U .$$

Proof. $p_0^\# U, p_0^\# W$, and $p_0^\# Z$ satisfy the conditions of Lemma 6.1. Let N' and Q' be the cochains of A given by Lemma 6.1, and $N = s^\# N'$, $Q = s^\# Q'$. Then

$$\begin{aligned} 4N + \delta Q &= s^\#(4N' + \delta Q') = (s^\# p_0^\# W)^2 \\ &+ s^\#((p_0^\# W) \smile_1 (\delta p_0^\# W)) - s^\# p_0^\# Z \\ &- s^\#\{(p_0^\# U) \smile_2 (p_0^\# U)\} = W^2 + W \smile_1 (\delta W) - Z - U \smile_2 U . \end{aligned}$$

We remark that since $4\delta N = 4(U \smile_1 U - WU)$ we have $\delta N = U \smile_1 U - WU$. Also, if N, Q and N', Q' satisfy Lemma 6.2 or Lemma

6.1, then $N - N'$ is a cocycle and $4(N - N')$ is a coboundary, for $4(N - N') = \delta(Q' - Q)$.

LEMMA 6.3. *If (U', W', N') is adjusted, and if (U, W, N) is as in Lemma 6.1, there exist cochains β, γ, and a cocycle $T \in C^{2k-2}(A)$ such that $4T$ is a coboundary, $U = U' + \delta\beta$, $W = W' + \delta\gamma - 2\beta$,*

and
$$N = N' - W'\beta + \beta\smile_1 U' + \beta^2 - U'\smile_1\beta - (\delta\gamma)\beta + \beta\smile_1(\delta\beta) - \gamma U' + T\,.$$

This follows directly from the proof of Lemma 6.1 and the above remark.

Now let U, W, and N be any cochains of B which satisfy Lemma 6.2, and let M be the algebraic mapping cylinder of the map $x \to xU$, with a product given by Definition 4.2. We then have a product in $H^*(M)$. For the remainder of this section, we will use square brackets to denote the natural map $C^*(A) \to C^*(A)/C^*(A, E) = C^*(E)$. The main theorem follows.

THEOREM 6.4. *There exist $\eta: M^p \to C^p(E)$ an allowable homomorphism and a cocycle $T \in C^{2k-2}(B)$ such that $4T$ is a coboundary which have the property that if (x, y) and (v, w) are p and q-cocycles, respectively, of M, then*

$$\eta\{(x, y)(v, w)\} - \eta(x, y)\eta(v, w) = [(-1)^{p+q+1}p_0^\sharp(ywT)] + \delta X \tag{6.5}$$

for some cochain X of E.

Proof. The homomorphism η is defined as follows: Choose (U', W', N') adjusted. We apply Lemma 6.3 to $p_0^\sharp U, p_0^\sharp W, p_0^\sharp N$ to obtain $\beta \in C^{k-1}(A)$ for which $p_0^\sharp U = U' + \delta\beta$. Define $\eta(x, y) = [p_0^\sharp x + p_0^\sharp(\bar{y}) \cdot \beta]$ for $(x, y) \in M$. Then

$$\begin{aligned}\delta\eta(x, y) &= [\delta p_0^\sharp x + \{\delta p_0^\sharp(\bar{y})\}\beta + p_0^\sharp(y)(\delta\beta)]\\ &= [p_0^\sharp\delta x + p_0^\sharp(\delta\bar{y})\beta + p_0^\sharp(y)(p_0^\sharp U - U')]\,,\end{aligned}$$

while

$$\begin{aligned}\eta\delta(x, y) &= \eta(\delta x + yU, -\delta y)\\ &= [p_0^\sharp\delta x + (p_0^\sharp y)(p_0^\sharp U) + p_0^\sharp(\overline{-\delta y})\beta]\,.\end{aligned}$$

Since $\overline{-\delta y} = \delta\bar{y}$ and $U' \in C^*(A, E)$, $\eta\circ\delta = \delta\circ\eta$ and η is allowable.

Let $\Gamma = \eta\{(x, y)(v, w)\} - \eta(x, y)\eta(v, w)$, where (x, y) and (v, w) are p and q-cocycles respectively of M. Then

$$\begin{aligned}\Gamma = [&(-1)^{p+q}p_0^\sharp(y)(p_0^\sharp U\smile_1 p_0^\sharp v) + (-1)^{p+q}p_0^\sharp y\{(p_0^\sharp(wU))\smile_2 p_0^\sharp U\}\\ &+ (-1)^{p+q+1}p_0^\sharp(y\, w\, N) + (-1)^p p_0^\sharp(y\, v)\beta\\ &+ (-1)^{p+q+1}(p_0^\sharp y)\{(p_0^\sharp w)\smile_1(p_0^\sharp U)\}\beta\\ &+ (-1)^{p+q+1}p_0^\sharp(y\, w\, W)\beta - (p_0^\sharp\bar{y})\beta(p_0^\sharp v) - (p_0^\sharp\bar{y})\beta(p_0^\sharp\bar{w})\beta]\,.\end{aligned}$$

We now replace $p_0^\# U$ by $U' + \delta\beta$. Note that all terms involving U' drop out, for $C^*(A, E)$ is an ideal. For simplicity, we write $x' = p_0^\#(x)$, etc. Then

$$\begin{aligned}\Gamma = [&(-1)^{p+q}y'((\delta\beta)\smile_1 v') + (-1)^{p+q}y'\{(w'\delta\beta)\smile_2(\delta\beta)\}\\ &+ (-1)^{p+q+1}y'w'p_0^\#(N) + (-1)^p y'v'\beta + (-1)^{p+q+1}y'(w'\smile_1\delta\beta)\beta\\ &+ (-1)^{p+q+1}y'w'p_0^\#(W)\beta + (-1)^{p+1}y'\beta v' + (-1)^{p+q+1}y'\beta w'\beta] .\end{aligned}$$

Now

$$\delta(\beta\smile_1 v') = (\delta\beta)\smile_1 v' + \beta\smile_1(\delta v') + (-1)^{q+1}\beta v' + (-1)^q v'\beta .$$

Since $\delta v = -wU$, $\delta(p_0^\# v) = -p_0^\# w p_0^\# U = -w'U' - w'\delta\beta$. Thus

$$\begin{aligned}(-1)^{p+q}y'\{(\delta\beta)\smile_1 v'\} &= (-1)^{p+q}y'\delta(\beta\smile_1 v') + (-1)^{p+q}y'(\beta\smile_1 w'U')\\ &+ (-1)^{p+q}y'\{\beta\smile_1(w'\delta\beta)\} + (-1)^p y'\beta v' + (-1)^{p+1}y'v'\beta\end{aligned}$$

and

$$\begin{aligned}\Gamma = [&(-1)^{p+q}y'\{\beta\smile_1(w'\delta\beta)\} + (-1)^{p+q}y'\{(w'\delta\beta)\smile_2(\delta\beta)\}\\ &+ (-1)^{p+q+1}y'w'p_0^\#(N) + (-1)^{p+q+1}y'(w'\smile_1\delta\beta)\beta\\ &+ (-1)^{p+q+1}y'w'p_0^\#(W)\beta + (-1)^{p+q+1}y'\beta w'\beta] + \text{coboundaries} ,\end{aligned}$$

for y' is a cocycle and $(-1)^{p+q}y'\delta(\beta\smile_1 v')$ a coboundary. It is easily checked that

$$\begin{aligned}(-1)^{p+q+1}y'\beta w'\beta &= (-1)^p y'\delta\{\beta\smile_1 w')\beta\\ &+ (-1)^{p+1}y'\{(\delta\beta)\smile_1 w'\}\beta + (-1)^{p+q+1}y'w'\beta^2 .\end{aligned}$$

From Lemma 6.3 we have $p_0^\#(W) = W' + \delta\gamma - 2\beta$. From these we get

$$\begin{aligned}\Gamma = [&(-1)^{p+q}y'\{\beta\smile_1(w'\delta\beta) + (-1)^{p+q}y'\{(w'\delta\beta)\smile_2(\delta\beta)\}\\ &+ (-1)^{p+q+1}y'w'p_0^\#(N) + (-1)^{p+q+1}y'(w'\smile_1\delta\beta)\beta\\ &+ (-1)^{p+q+1}y'w'(W' + \delta\gamma - 2\beta)\beta + (-1)^p y'\delta\{\beta\smile_1 w'\}\beta\\ &+ (-1)^{p+1}y'\{(\delta\beta)\smile_1 w'\}\beta + (-1)^{p+q+1}y'w'\beta^2] + \text{coboundaries} .\end{aligned}$$

Since

$$\begin{aligned}(-1)^{p+1}y'\{(\delta\beta)\smile_1 w'\}\beta &+ (-1)^{p+q+1}y'\{w'\smile_1\delta\beta\}\beta\\ &+ (-1)^p y'\delta\{\beta\smile_1 w'\}\beta = (-1)^{p+1}y'\delta\{w'\smile_1\beta\}\beta ,\end{aligned}$$

we have

$$\begin{aligned}\Gamma = [&(-1)^{p+q}y'\{\beta\smile_1(w'\delta\beta)\} + (-1)^{p+q}y'\{(w'\delta\beta)\smile_2\delta\beta\}\\ &+ (-1)^{p+q+1}y'w'p_0^\#(N) + (-1)^{p+q+1}y'w'(W' + \delta\gamma)\beta\\ &+ (-1)^{p+q}y'w'\beta^2 + (-1)^{p+1}y'\delta\{w'\smile_1\beta\}\beta] + \text{coboundaries} .\end{aligned}$$

Now

$$\begin{aligned}(-1)^{p+q}y'\{(w'\delta\beta)\smile_2\delta\beta\} &= (-1)^{p+1}y'\delta\{(w'\delta\beta)\smile_2\beta\}\\ &+ (-1)^{p+q+1}y'\{(w'\delta\beta)\smile_1\beta\} + (-1)^{p+q+1}y'\{\beta\smile_1(w'\delta\beta) ,\end{aligned}$$

and thus

$$\begin{aligned}\Gamma = (-1)^{p+q+1}y'\{(w'\delta\beta)\smile_1\beta\} + (-1)^{p+q+1}y'w'(W' + \delta\gamma)\beta \\ + (-1)^{p+q}y'w'\beta^2 + (-1)^{p+1}y'\delta\{w'\smile_1\beta\}\beta \\ + (-1)^{p+q+1}y'w'\{-W'\beta + \beta^2 - (\delta\gamma)\beta + \beta\smile_1(\delta\beta) + T\}] \\ + \text{coboundaries, where we have used Lemma 6.3 on } p_0^{\#}(N)\ .\end{aligned}$$

Thus

$$\begin{aligned}\Gamma = [(-1)^{p+q+1}y'\{(w'\delta\beta)\smile_1\beta\} + (-1)^{p+1}y'\delta\{w'\smile_1\beta\}\beta \\ + (-1)^{p+q+1}y'w'(\beta\smile_1\delta\beta) + (-1)^{p+q+1}y'w'T] + \text{coboundaries}\ .\end{aligned}$$

But

$$(-1)^{p+1}y'\delta\{w'\smile_1\beta\}\beta = (-1)^{p+1}y'\delta\{(w'\smile_1\beta)\beta\} + (-1)^{p+q+1}y'(w'\smile_1\beta)\delta\beta\ ,$$

and so

$$\begin{aligned}\Gamma &= [(-1)^{p+q+1}y'\{(w'\delta\beta)\smile_1\beta\} + (-1)^{p+q+1}y'(w'\smile_1\beta)\delta\beta \\ &\quad + (-1)^{p+q+1}y'w'(\beta\smile_1\delta\beta) + (-1)^{p+q+1}y'w'T] + \text{coboundaries} \\ &= [(-1)^{p+q+1}y'\{(w'\delta\beta)\smile_1\beta + (w'\smile_1\beta)\delta\beta + w'(\beta\smile_1\delta\beta) \\ &\qquad + (-1)^{p+q+1}y'w'T] + \text{coboundaries}\ .\end{aligned}$$

By Hirsch's formula 4.6,

$$\begin{aligned}\Gamma = [(-1)^{p+q+1}y'w'\{\beta\smile_1\delta\beta + (\delta\beta)\smile_1\beta\} + (-1)^{p+q+1}y'w'T] \\ + \text{coboundaries}\ .\end{aligned}$$

Since $\delta(\beta\smile_1\beta) = \beta\smile_1\delta\beta + (\delta\beta)\smile_1\beta$,

$$\Gamma = [(-1)^{p+q+1}y'w'T] + \text{coboundaries}\ .$$

In view of the fact that $p_0^{\#}\circ s^{\#}$ is homotopic to the identity, we have

$$\Gamma = [(-1)^{p+q+1}p_0^{\#}(yw\, s^{\#}T)] + \text{coboundaries}$$

as asserted. This completes the proof of Theorem 6.4.

REMARK 6.6. *The following diagram commutes except for the square marked "$\otimes$" which anti-commutes.*

$$\begin{array}{ccccccccc}\cdots \longrightarrow & H^{q-k}(B) & \xrightarrow{\mu} & H^q(B) & \xrightarrow{p^*} & H^q(E) & \xrightarrow{\phi} & H^{q-k+1}(B) & \longrightarrow \cdots \\ & \uparrow id. & & \uparrow id. & & \uparrow \eta^* & \otimes & \uparrow id. & \\ \cdots \longrightarrow & H^{q-k}(B) & \xrightarrow{\mu} & H^q(B) & \xrightarrow{i^*} & H^q(M) & \xrightarrow{j^*} & H^{q-k+1}(B) & \longrightarrow \cdots\end{array}$$

Thus by the five-lemma, η^* is one-to-one and onto.

Proof. Let $x' \in x \in H^*(B)$. Then $(x', 0)$ represents $i^*(x)$ and $[p_0^\# x']$ represents $\eta^* i^*(x)$. It also represents $p^*(x)$. For the other square, let $(x, y) \in z \in H^*(M)$. Then y represents $j^*(z)$, and $[p_0^\# x + (p_0^\# \bar{y})\beta]$ represents $\eta^*(z)$. Referring to Figure 1, $\psi = p_0^{\#-1}\theta^{-1}\delta^*$. Now $\delta\{p_0^\#(x) + (p_0^\# \bar{y})\beta\} = p_0^\#(-yU) + p_0^\#(y)\delta\beta = -(p_0^\# y) \cdot U'$. Thus $(\psi\eta^*)(z)$ is represented by $-y$.

From equation 6.5 we see that if $H^{2k-2}(B)$ has no elements of order two, $\eta^* : H^q(M) \to H^q(E)$ is a ring isomorphism.

THEOREM 6.7. *Let (E, p, B, S^{k-1}) and (E', p', B, S^{k-1}) be two (orientable) $(k-1)$-sphere spaces over the same (compact) base space with k odd. Suppose $H^{2k-2}(B, Z)$ has no elements of order two. Then if the sphere spaces have the same P-invariant and the same Stiefel-Whitney classes W_k aud W_{k-1}, their integral cohomology rings are isomorphic.*

To prove this, we observe that both cohomology rings are isomorphic to the cohomology of the mapping cylinder M of Theorem 6.4.

If the rationals or the integers mod n, n odd, are used as coefficients for cohomology, then η^* is always a ring isomorphism since $H^{2k-2}(B)$ will have no elements of order two. Consequently the cohomology ring with these coefficients of a sphere space is always given by Theorem 6.4.

7. The case k even. In this section we suppose (E, p, B, S^{k-1}) is a $(k-1)$-sphere space, with k even. Suppose $V \in C^{k-2}(A)$ is any integral cochain representing w_{k-2}. Then for some W, $\delta V = -2W$ and W represents w_{k-1}. Let $U \in w_k$. Then VU and $U \smile_2 U$ both represent $w_{k-2}w_k$ and so $VU + U \smile_2 U$ is a coboundary mod 2, *i.e.*, there exist N and Q cochains of A for which

$$2N + \delta Q = VU + U \smile_2 U .$$

From this it follows that $\delta N = U \smile_1 U - WU$. If also $2N' + \delta Q' = VU + U \smile_2 U$, then $N - N'$ is a cocycle and $2(N - N')$ a coboundary.

LEMMA 7.1. *Let (U, W, N) be adjusted for the sphere space (E, p, B, S^{k-1}) and let V be an integral cochain representing w_{k-2} for which $\delta V = -2W$. Then there exist a cocycle $Y \in C^{k-2}(A)$ and a cochain $X \in C^{2k-2}(A, E)$ for which*

$$VU + U \smile_2 U - 2N = 2YU + \delta X .$$

Proof. We first remark that it is possible to find such (U, W, N) and V. One chooses V to be any integral cochain representing w_{k-2} and defines W by $\delta V = -2W$. Then W represents w_{k-1}. Choose $U \in \mathscr{U} \in H^k(A - E)$, and $N \in C^{2k-2}(A, E)$ such that $\delta N = U \smile_1 U - WU$. Now let

$b = VU + U \smile_2 U - 2N$. Then $b \in C^{2k-2}(A, E)$ and it is easly seen that b is a cocycle. For some $x, y \in C^*(A, E)$, $U \smile_2 U = VU + 2x + \delta y$ since $U \smile_2 U$ and VU both represent $w_{k-2}\mathscr{U}$ mod 2. Thus $b = 2(VU + x - N) + \delta y$. Since b is a cocycle, $VU + x - N$ is a cocycle of $C^*(A, E)$. The map θ of (2.1) is an isomorphism, consequently there is a cocycle $Y \in C^{k-2}(A)$ and a cochain $Z \in C^{2k-2}(A, E)$ such that $VU + x - N = YU + \delta Z$. Then $b = 2YU + \delta(2Z + y)$.

The following crucial lemma may be interpreted as giving a standard form for the cochains N described in the opening paragraphs of this section.

LEMMA 7.2. *Suppose U is any representative cocycle for w_k, V is any cochain representing w_{k-2}, $\delta V = -2W$, so W represents w_{k-1}, $N \in C^{2k-2}(A)$, $Q \in C^{2k-3}(A)$, and $2N + \delta Q = VU + U \smile_2 U$. Suppose also that (U', W', N') is adjusted, V' represents w_{k-2}, and $\delta V' = -2W'$. Let X and Y be chosen by Lemma 7.1 so that $V'U' + U' \smile_2 U' - 2N' = 2YU' + \delta X$. Then there exist β, α, γ, and T, cochains of A of degrees $k-1$, $k-2$, $k-3$, and $2k-2$ respectively so that T is a cocycle, $2T$ is a coboundary, $U = U' + \delta\beta$, $W = W' + \delta\alpha$, $V = V' + \delta\gamma - 2\alpha$, and*

$$N = N' - \alpha\delta\beta - \alpha U' + (\delta\beta) \smile_2 U' + \beta \smile_1 (\delta\beta) + \beta^2 + W'\beta + yU' + T\,.$$

Proof. The existence of α, β, and γ so that the first three equations are satisfied is trivial. To prove the lemma it is only necessary to verify that

$$\begin{aligned} &2(N' - \alpha\delta\beta - \alpha U' + (\delta\beta) \smile_2 U' + \beta \smile_1 (\delta\beta) + \beta^2 + W'\beta + YU') + \delta Q' \\ &\quad = VU + U \smile_2 U \end{aligned}$$

for some cochain Q'. We choose

$$Q' = \gamma U' + V'\beta + \gamma\delta\beta + (\delta\beta) \smile_3 U' + \beta \smile_2 (\delta\beta) + \beta \smile_1 \beta + X\,.$$

The computation is omitted since it is straightforward.

For the next theorem, we return to the cochains of the base space B. We suppose $U \in W_k \in H^k(B)$, V is a cochain representing $W_{k-2} \in H^{k-2}(B, Z_2)$, and $\delta V = -2W$. Then W represents $W_{k-1} \in H^{k-1}(B)$. We obtain N and Q in $C^*(B)$ for which $2N + \delta Q = VU + U \smile_2 U$. Let M be the algebraic mapping cylinder of the map $x \to xU$ for $x \in C^*(B)$, with a product given by Definition 4.2. This product satisfies the coboundary formula and induces a product in $H^*(M)$. We will use square brackets to denote the natural map $C^*(A) \to C^*(A)/C^*(A, E)$.

THEOREM 7.3. *There exists an allowable homomorphism $\eta : M^p \to C^p(E)$ and a cocycle $T \in C^{2k-2}(B)$ for which $2T$ is a coboundary with the* *following property: If* (x, y) *and* (v, w) *are* p *and* q*-cocycles, respectively, of* M*, then*

$$(7.4) \qquad \eta\{(x, y)(v, w)\} - \eta(x, y)\eta(v, w) = [(-1)^p p_0^\#(y\, w\, T)] + \delta Z$$

for some cochain Z *of* E.

REMARK 7.5. The following diagram commutes except for the square marked "$\otimes$" which anti-commutes:

$$\begin{array}{ccccccccccc}
\cdots \longrightarrow & H^{q-k}(B) & \xrightarrow{\mu} & H^q(B) & \xrightarrow{i^*} & H^q(M) & \xrightarrow{j^*} & H^{q-k+1}(B) & \longrightarrow \cdots \\
 & \downarrow{\scriptstyle id.} & & \downarrow{\scriptstyle id.} & & \downarrow{\scriptstyle \eta^*} & \otimes & \downarrow{\scriptstyle id.} & \\
\cdots \longrightarrow & H^{q-k}(B) & \xrightarrow{\mu} & H^q(B) & \xrightarrow{p^*} & H^q(E) & \xrightarrow{\phi} & H^{q-k+1}(B) & \longrightarrow \cdots
\end{array}$$

Consequently, by the five-lemma, η^* is one-to-one and onto.

To prove Theorem 7.3, we first choose (U', W', N') adjusted and obtain X and Y from Lemma 7.1. Then apply Lemma 7.2 to obtain α, β, γ, and T for which $p_0^\# U = U' + \delta\beta$, $p_0^\# W = W' + \delta\alpha$, $p_0^\# V = V' + \delta\gamma - 2\alpha$, and

$$p_0^\# N = N' - \alpha\delta\beta - \alpha U' + (\delta\beta)\smile_2 U' + \beta\smile_1(\delta\beta) + \beta^2 + W'\beta + YU' + T\,.$$

Define, for (x, y) in M,

$$\eta(x, y) = [p_0^\# x + (p_0^\# y)\beta]\,.$$

Then η is allowable, *i.e.*, $\delta\eta = \eta\delta$.

The remainder of the proof is omitted, as it is a tedious computation similar to the proof of Theorem 6.4. The proof that the diagram of Remark 7.5 commutes has been given in the proof of Remark 6.6.

From equation (7.4) it follows that if $H^{2k-2}(B)$ has no elements of order two, η^* is a ring isomorphism.

THEOREM 7.6. *Suppose* (E, p, B, S^{k-1}) *and* (E', p', B, S^{k-1}) *are two (orientable)* $(k-1)$*-sphere spaces over the same compact base space with* k *even. Suppose* $H^{2k-2}(B, Z)$ *has no elements of order two. Then if the sphere spaces have the same Stiefel-Whitney classes* W_k, W_{k-1}, *and* W_{k-2}, *their integral cohomology rings are isomorphic.*

This follows because both cohomology rings must be isomorphic to the cohomology ring $H^*(M)$.

The following theorem generalizes a result of R. Thom ([4], *exposé* 17, *Théorème* 3).

THEOREM 7.7. *Suppose* (E, p, B, S^{k-1}) *is a* $(k-1)$*-sphere space, for* k *even. Using the rational numbers or the integers mod* n, n *odd, as* *coefficients for cohomology, the cohomology of the base space and the characteristic class W_k determine the cohomology ring of the total space E.*

Proof. $W_{k-1} = 0$ since $2W_{k-1} = 0$. Let U be any representative cocycle for $W_k \in H^k(B)$, and let M be the algebraic mapping cylinder of $x \to xU$, for $x \in C^*(B)$. We introduce a multiplication in M by choosing $W = 0$ and $N = \frac{1}{2}(U \smile_2 U)$ or, specifically, the multiplication is defined by

$$(x, y)(v, w) = (xv + (-1)^{p+q}y(U \smile_1 v) + (-1)^{p+1}y\{(wU)\smile_2 U\} + (-1)^p yw\tfrac{1}{2}(U\smile_2 U), (-1)^p xw + yv + (-1)^q y(w \smile_1 U))$$

for $(x, y) \in M^p$ and $(v, w) \in M^q$. Since $\delta\frac{1}{2}(U\smile_2 U) = U \smile_1 U$, this multiplication induces a multiplication in $H^*(M)$.

Let $U' \in \mathscr{U} \in H^k(A - E)$. Then for some $\beta \in C^{k-1}(A)$, $p_0^\# U = U' + \delta\beta$. Define $\eta : M^p \to C^p(E)$ by $\eta(x, y) = [p_0^\# x + (p_0^\# \bar{y})\beta]$. Then η is allowable and induces $\eta^* : H^p(M) \to H^p(E)$. Let (x, y) and (v, w) be p and q-cocycles, respectively, of M and let $\Gamma = \eta\{(x, y)(v, w)\} - \eta(x, y)\eta(v, w)$. Then, letting $x' = p_0^\# x$, etc., as before, we have

$$\begin{aligned}\Gamma = [(-1)^{p+q}y'\{(p_0^\# U)\smile_1 v'\} + (-1)^{p+1}y'\{(w'p_0^\# U)\smile_2 p_0^\# U\} \\ + (-1)^p y'w'\tfrac{1}{2}\{(p_0^\# U)\smile_2 p_0^\# U\} + (-1)^{p+q+1}y'v'\beta + (-1)^{p+1}y'(w'\smile_1 p_0^\# U)\beta \\ + (-1)^p y'\beta v' + (-1)^{p+q+1}y'\beta w'\beta]\ .\end{aligned}$$

Exactly as in the proof of Theorem 6.4, reduce this to

$$\Gamma = [(-1)^{p+1}y'w'(\beta\smile_1\delta\beta) + (-1)^p y'w'\tfrac{1}{2}\{(\delta\beta)\smile_2\delta\beta\} + (-1)^{p+1}y'w'\beta^2] + \text{coboundaries}\ .$$

Since $(\delta\beta)\smile_2(\delta\beta) = \delta(\beta\smile_2\delta\beta) + \beta\smile_1(\delta\beta) + (\delta\beta)\smile_1\beta$,

$$\Gamma = [(-1)^{p+1}y'w'\tfrac{1}{2}(\beta\smile_1\delta\beta) + (-1)^p y'w'\tfrac{1}{2}(\delta\beta\smile_1\beta) + (-1)^{p+1}y'w'\beta^2] + \text{coboundaries}\ ,$$

and so Γ is a coboundary since

$$2\beta^2 = \delta(\beta\smile_1\beta) + (\delta\beta)\smile_1\beta - \beta\smile_1(\delta\beta)\ .$$

Thus η^* preserves products. η^* is shown to be one-to-one and onto exactly as in the proof of Remark 6.6.

In conclusion, we would like to point out that the remarks at the end of Chapter 3 apply also to Theorems 6.7 and 7.6. The question of what one needs to know about $H^*(B)$ in addition to the product structure (and various characteristic classes) to determine $H^*(E)$ seems to be rather complicated (see [7], Part 1, and [4], *exposé* 17). Certainly various higher order operations are needed.

References

1. A. Borel, *Topology of Lie groups and characteristic classes*, Bull. Amer. Math. Soc. **61** (1955), 397-432.

2. ———, *Selected topics in the homology theory of fibre bundles*, Department of Mathematics, University of Chicago, 1954, (mimeographed lecture notes).

3. H. Cartan, *Espaces fibrés et homotopie*, Ecole Normale Supérieure, Paris, 1956 (mimeographed seminar notes).

4. ———, *Algebres d'Eilenberg-MacLane et homotopie*, Ecole Normale Supérieure, Paris, 1955 (mimeographed seminar notes).

5. S. Eilenberg and N. E. Steenord, *Foundations of algebraic topology*, Princeton University Press, 1953.

6. G. Hirsch, *Quelques proprietés des produits de Steenrod*, C. R. Acad. Sci. Paris **241** (1955), 923-925.

7. W. S. Massey, *On the cohomology ring of a sphere bundle*, Jour. of Math. and Mech. **7** (1958), 265-290.

8. N. E. Steenrod, *The topology of fibre bundles*, Princeton University Press, 1951.

9. ———, *Products of cocycles and extensions of mappings*, Ann. of Math. **48** (1947), 290-320.

10. ———, *Cohomology invariants of mappings*, Ann. of Math. **50** (1949), 954-988.

11. R. Thom, *Espaces fibrés en spheres et Carrés de Steenrod*, Ann. Ecole Norm. Sup. **69** (1952), 109-182.

12. W. T. Wu, *On Pontrjagin classes, III*, Acta. Math. Sinica **4** (1954), 323-346 (in Chinese, English Summary).

Brown University
Rutgers, The State University

TRANSFORMATIONS ON TENSOR PRODUCT SPACES

MARVIN MARCUS AND B. N. MOYLS

1. Introduction. Let U and V be m- and n-dimensional vector spaces over an algebraically closed field F of characteristic 0. Then $U \otimes V$, the tensor product of U and V, is the dual space of the space of all bilinear functionals mapping the cartesian product of U and V into F. If $x \in U$, $y \in V$ and w is a bilinear functional, then $x \otimes y$ is defined by: $x \otimes y(w) = w(x, y)$. If $e_1, \cdots, e_m$ and $f_1, \cdots, f_n$ are bases for U and V, respectively, then the $e_i \otimes f_j$, $i = 1, \cdots, m$, $j = 1, \cdots, n$, form a basis for $U \otimes V$.

Let $M_{m,n}$ denote the vector space of $m \times n$ matrices over F. Then $U \otimes V$ is isomorphic to $M_{m,n}$ under the mapping ψ where $\psi(e_i \otimes f_j) = E_{ij}$, and E_{ij} is the matrix with 1 in the (i, j) position and 0 elsewhere. An element $z \in U \otimes V$ is said to be of rank k if $z = \sum_{i=1}^{k} x_i \otimes y_i$, where $x_1, \cdots, x_k$ are linearly independent and so are $y_1, \cdots, y_k$. If $R_k = \{z \in U \otimes V \mid \operatorname{rank}(z) = k\}$, then $\psi(R_k)$ is the set of matrices of rank k, in $M_{m,n}$. In view of the isomorphism any linear map T of $U \otimes V$ into itself can be considered as a linear map of $M_{m,n}$ into itself.

In [2] and [3], Hua and Jacob obtained the structure of any mapping T that preserves the rank of every matrix in $M_{m,n}$ and whose inverse exists and has this property (coherence invariance). (In [3] F is replaced by a division ring, and T is shown to be semi-linear by appealing to the fundamental theorem of projective geometry.) In [4] we obtained the structure of T when $m = n$, T is linear and T preserves rank 1, 2 and n. Specifically, there exist non-singular matrices M and N such that $T(A) = MAN$ for all $A \in M_{nn}$, or $T(A) = MA'N$ for all A, where A' designates the transpose of A. Frobenius (cf. [1], p. 249) obtained this result when T is a a linear map which preserves the determinant of every A. In [5] it was shown that this result can be obtained by requiring only that T be linear and preserve rank n. In the present paper we show that rank 1 suffices (Theorem 1), or rank 2 with the side condition that T maps no matrix of rank 4 or less into 0 (Theorem 2). Thus our hypothesis will be that T is linear and $T(R_1) \subseteq R_1$. We remark that T may be singular and still its kernel may have a zero intersection with R_1; e.g., take $U = V$ and $T(x \otimes y) = x \otimes y + y \otimes x$.

2. Rank one preservers. Throughout this section T will be a linear transformation (l.t.) of $U \otimes V$ into $U \otimes V$ such that $T(R_1) \subseteq R_1$. Here

Received March 2, 1959. The work of the first author was sponsored by U.S. National Science Foundation Grant G. 5416.

U and V are m- and n-dimensional vector spaces over F. Let $e_1, \cdots, e_m$ aud $f_1 \cdots, f_n$ be fixed bases for U and V, and set

$$(1) \qquad T(e_i \otimes f_j) = u_{ij} \otimes v_{ij}, \qquad i = 1, \cdots, m;\ j = 1, \cdots, n.$$

Note that no u_{ij} or v_{ij} can be zero. We shall show, in case $m \neq n$ that there exist vectors u_i and v_j such that $T(e_i \otimes f_j) = u_i \otimes v_j$, and hence that the l.t. T is a tensor product of transformations on U and V separately. In case $m = n$ it will be shown that a slight modification of T is a tensor product.

Denote by $L(x_1, \cdots, x_t)$ the subspace spanned by the vectors $x_1, \cdots, x_t$, and let $\rho(x_1, \cdots, x_t)$ be the dimension of $L(x_1, \cdots, x_t)$.

LEMMA 1. *Let $x_1, \cdots, x_r, w_1, \cdots, w_s$ be vectors in U, and let $y_1, \cdots, y_r, z_1, \cdots, z_s$ be vectors in V. Let*

$$(2) \qquad \sum_{i=1}^{r} (x_i \otimes y_i) = \sum_{j=1}^{s} (w_j \otimes z_j) .$$

If $\rho(x_1, \cdots, x_r) = r$, then $y_i \in L(z_1, \cdots, z_s)$, $i = 1, \cdots, r$; and similarly if $\rho(y_1, \cdots, y_r) = r$, then $x_i \in L(w_1, \cdots, w_s)$, $i = 1, \cdots, r$.

Proof. Suppose that $\rho(x_1, \cdots, x_r) = r$. Let θ be a linear functional on U such that $\theta(x_1) = 1$, $\theta(x_i) = 0$, $i \neq 1$, and let α be an arbitrary linear functional on V. For $x \in U$, $y \in V$, define

$$(3) \qquad g(x, y) = \theta(x)\alpha(y) .$$

Applying (2) to g, we get

$$\alpha(y_1) = \sum_{i=1}^{s} \theta(w_j)\alpha(z_j) = \alpha\Big(\sum_{j=1}^{s} \theta(w_j)z_j\Big)$$

where each $\theta(w_j)$ is a scalar. Since α is arbitrary, y_1, and similarly $y_2, \cdots, y_r$, are contained in $L(z_1, \cdots, z_s)$. The second part of the lemma is proved in the same way.

LEMMA 2. *If $T(R_1) \subseteq R_1$, and T satisfies* (1), *then for $i = 1, \cdots, m$, either*

$$(4) \qquad \rho(u_{i1}, \cdots, u_{in}) = n \quad and \quad \rho(v_{i1}, \cdots, v_{in}) = 1 ,$$

or

$$(5) \qquad \rho(u_{i1}, \cdots, u_{in}) = 1 \quad and \quad \rho(v_{i1}, \cdots, v_{in}) = n .$$

Similarly, for $j = 1, \cdots, n$, either

$$(6) \qquad \rho(u_{1j}, \cdots, u_{mj}) = m \quad and \quad \rho(v_{1j}, \cdots, v_{mj}) = 1 ,$$

or

$$(7) \qquad (u_{1j}, \cdots, u_{mj}) = 1 \quad \textit{and} \quad (v_{1j}, \cdots, v_{mj}) = m .$$

Proof. Suppose that $u_{i\alpha}$ and $u_{i\beta}$ are independent. Then

$$T(e_i \otimes (f_\alpha + f_\beta)) = (u_{i\alpha} \otimes v_{i\alpha}) + (u_{i\beta} \otimes v_{i\beta})$$

must be a tensor product $u \otimes v$. By Lemma 1, $v_{i\alpha}, v_{i\beta} \in L(v)$. Since all $v_{ij} \neq 0$, $L(v_{i\alpha}) = L(v_{i\beta})$. For $\gamma \neq \alpha, \beta$, $L(v_{i\gamma}) = L(v_{i\alpha})$, since $u_{i\gamma}$ must be independent of at least one of $u_{i\alpha}, u_{i\beta}$. We have shown that if $\rho(u_{i1}, \cdots, u_{in}) \geq 2$, then $\rho(v_{i1}, \cdots, v_{in}) = 1$.

Suppose next that $\rho(u_{i1}, \cdots, u_{in}) = 1$, viz., $u_{i\alpha} = c_\alpha u_{i1}$, $c_\alpha \neq 0$, $\alpha = 1, \cdots, n$. If

$$\rho(v_{i1}, \cdots, v_{in}) < n , \quad \text{let} \quad \sum_{\alpha=1}^{n} a_\alpha v_{i\alpha} = 0$$

be a non-trivial dependence relation. Then

$$T\left(e_i \otimes \left(\sum_{\alpha=1}^{n} \frac{a_\alpha}{c_\alpha} f_\alpha\right)\right) = \sum_{\alpha=1}^{n} \left(c_\alpha u_{i1} \otimes \frac{a_\alpha}{c_\alpha} v_{i\alpha}\right) = u_{i1} \otimes \left(\sum_{\alpha=1}^{n} a_\alpha v_{i\alpha}\right) = 0 ,$$

which is impossible by the nature of T. Hence $\rho(u_{i1}, \cdots, u_{in}) = 1$ implies $\rho(v_{i1}, \cdots, v_{in}) = n$.

It follows by a similar argument that if $\rho(v_{i1}, \cdots, v_{in}) = 1$, then $\rho(u_{i1}, \cdots, u_{in}) = n$. Hence either (4) or (5) must hold. The second part of the lemma is proved similarly.

We remark that if $m < n$ (or $n < m$), then (4) (or (7)) cannot hold.

LEMMA 3. *Either* (4) *and* (7) *hold for all* i, j; *or* (5) *and* (6) *hold for all* i, j.

Proof. We show first that either (4) or (5) holds uniformly in i. Suppose that for some i and k, $1 \leq i \leq k \leq m$, $\rho(u_{i1}, \cdots, u_{in}) = n$ while $\rho(u_{k1}, \cdots, u_{kn}) = 1$. Then for some α, $1 \leq \alpha \leq n$, $\rho(u_{i\alpha}, u_{k\alpha}) = 2$. For $\beta \neq \alpha$ consider

$$\begin{aligned} \eta &= T[(e_i + e_k) \otimes (cf_\alpha + f_\beta)] \\ &= c(u_{i\alpha} \otimes v_{i\alpha}) + (u_{i\beta} \otimes v_{i\beta}) + c(u_{k\alpha} \otimes v_{k\alpha}) + (u_{k\beta} \otimes v_{k\beta}) , \end{aligned}$$

where c is an arbitrary scalar.

By hypothesis and Lemma 2, $v_{i\alpha} = av_{k\alpha}$ and $v_{i\beta} = b_1 v_{i\alpha} = bv_{k\alpha}$ for suitable non-zero scalars a and b, while $\rho(v_{k\alpha}, v_{k\beta}) = 2$. Thus $\eta = (acu_{i\alpha} + bu_{i\beta} + cu_{k\alpha}) \otimes v_{k\alpha} + (u_{k\beta} \otimes v_{k\beta})$, and by Lemma 1, $\rho(acu_{i\alpha} + bu_{i\beta} + cu_{k\alpha}, u_{k\beta}) = 1$ for all scalars c. Since $\rho(u_{k\alpha}, u_{k\beta}) = 1$, this implies that $\rho(cu_{i\alpha} + u_{i\beta}, u_{k\beta}) = 1$ for all c. This is impossible, since $\rho(u_{i\alpha}, u_{i\beta}) = 2$. Thus either (4) is true for all i, or (5) is true for all i. A similar argument applies to (6) and (7).

If (4) and (6) hold for all i and j, then there exist non-zero scalars c_{ij} such that $v_{ij} = c_{ij}v_{11}$, $i = 1, \cdots, m$, $j = 1. \cdots, n$. For a_j, b scalars, consider

$$T\left[\left(\sum_{i=1}^{m} a_i e_i\right) \otimes (f_1 - bf_2)\right] = \left(\sum_{i=1}^{m} a_i c_{i1} u_{i1} - b \sum_{i=1}^{m} a_i c_{i2} u_{i2}\right) \otimes v_{11} .$$

Let $z_1, \cdots, z_m$ and $w_1, \cdots, w_m$ be the m-column vectors which are respectively the representations of $u_{11}, \cdots, u_{m1}$ and $u_{12}, \cdots, u_{m2}$ with respect to the basis $e_1, \cdots, e_m$. Let C be the m-square matrix whose columns are $c_{11}z_1, \cdots, c_{m1}z_m$ and let W be the m-square matrix whose columns are $c_{12}w_1, \cdots, c_{m2}w_m$. Then with respect to the basis $e_1, \cdots, e_m$ the vector $\sum_{i=1}^{m} a_i c_{i1} u_{i1} - b \sum_{i=1}^{m} a_i c_{i2} u_{i2}$ has the representation $(C - bW)a$ where a is the column m-tuple $(a_1, \cdots, a_m)$. Now C and W are non-singular since $\rho(u_{11}, \cdots, u_{m1}) = \rho(u_{12}, \cdots, u_{m2}) = m$, so choose b to be an eigenvalue of $W^{-1}C$ and choose a to be the corresponding eigenvector. Then $(C - bW)a = 0$ and hence there exist scalars $a_1, \cdots, a_m$ not all 0 and b such that

$$T\left(\sum_{i=1}^{m} a_i e_i \otimes (f_1 - bf_2)\right) = 0 ,$$

a contradiction since $T(R_1) \subseteq R_1$.

Hence (4) and (6) cannot hold for all i and j. Similarly both (5) and (7) cannot hold for all i and j. This completes the proof of the lemma.

In view of the remark preceding this lemma, (5) and (6) must hold when $m \neq n$.

THEOREM 1. *Let U and V be m- and n-dimensional vector spaces respectively. Let T be a linear transformation on $U \otimes V$ which maps elements of rank one into elements of rank one. Let T_1 be the* l.t. *of $V \otimes U$ into $U \otimes V$ which maps $y \otimes x$ onto $x \otimes y$. If $m = n$, let φ be any non-singular* l.t. *of U onto V. Then if $m \neq n$, there exist non-singular* l.t.'s *A and B on U and V, respectively, such that $T = A \otimes B$. If $m = n$, there exist non-singular A and B such that either $T = A \otimes B$ or $T = T_1(\varphi A \otimes \varphi^{-1}B)$.*

Proof. By (1) and Lemma 3, $T(e_i \otimes f_j) = u_{ij} \otimes v_{ij}$, $i = 1, \cdots, m$, $j = 1, \cdots, n$, where either (5) and (6) hold or (4) and (7) hold. Suppose first that the former is the case; in particular, $\rho(u_{i1}, \cdots, u_{in}) = 1$ for $i = 1, \cdots, m$ and $\rho(v_{1j}, \cdots, v_{mj}) = 1$ for $j = 1, \cdots, n$. Then there exist non-zero scalars s_{ij}, t_{ij} such that $u_{ij} = s_{ij}u_{i1}$ and $v_{ij} = t_{ij}v_{1j}$. Thus

$$T(e_i \otimes f_j) = c_{ij} u_i \otimes v_j , \tag{8}$$

where $u_i = u_{i1}$, $v_j = v_{1j}$, and $c_{ij} = s_{ij}t_{ij}$. For $i = 2, \cdots, n$,

$$T\left[(e_1 + e_i) \otimes \left(\sum_{j=1}^{n} f_j\right)\right] = u_1 \otimes \sum_{j=1}^{n} c_{1j} v_j + u_i \otimes \sum_{j=1}^{n} c_{ij} v_j$$

must be a direct product $x \otimes w$. By (6) and Lemma 1, $\sum_{j=1}^{n} c_{ij} v_j = d_i \sum_{j=1}^{n} c_{1j} v_j$ for some constant d_i. By (5), $c_{ij} = d_i c_{1j}$. Hence

$$T(e_i \otimes f_j) = x_i \otimes y_j \,, \tag{9}$$

where $x_i = d_i u_i$ and $y_j = c_{1j} v_j$. Since the $\{x_i\}$ and $\{y_j\}$ are each linearly independent sets, there non-singular linear transformations A and B such that $x_i = Ae_i$ and $y_j = Bf_j$. Then $T = A \otimes B$.

When $m = n$, (4) and (7) may hold; in particular,

$$\rho(v_{i1}, \cdots, v_{in}) = 1 \text{ and } \rho(u_{1j}, \cdots, u_{nj}) = 1 \quad \text{for} \quad i, j = 1, \cdots, n.$$

As in the preceding case, there exist linearly independent sets $x_1, \cdots, x_n$ and $y_1, \cdots, y_n$ such that

$$T(e_i \otimes f_j) = x_j \otimes y_i \,. \tag{10}$$

There exist non-singular transformations A and B of U and V, respectively, such that $Ae_i = \varphi^{-1} y_i$ and $Bf_j = \varphi x_j$, $i, j = 1, \cdots, n$. Thus $T_1^{-1} T(e_i \otimes f_j) = \varphi Ae_i \otimes \varphi^{-1} Bf_j$. Q.E.D.

In matrix language we have the following.

COROLLARY. *Let T be a* l.t. *on the space M_{nn} of n-square matrices. If the set of rank one matrices is invariant under T, then there exist non-singular matrices A and B such that either $T(X) = AXB$ for all $X \in M_{nn}$ or $T(X) = AX'B$ for all $X \in M_{nn}$.*

3. **Rank two preservers.** In this section T will be a l.t. of $U \otimes V$ such that $T(R_2) \subseteq R_2$. We shall show that under certain conditions $T(R_1) \subseteq R_1$.

LEMMA 4. *If W is a subspace of $U \otimes V$ such that, for some integer r, $1 \leq r \leq \min(m, n)$,*

$$\dim W \geq mn - r \max(m, n) + 1 \,, \tag{11}$$

then $W \cap \bigcup_{j=1}^{r} R_j \neq \phi$.

Proof. Suppose that $m = \max(m, n)$. The products $e_i \otimes f_j$, $i = 1, \cdots, m$, $j = 1, \cdots, r$, are linearly independent and span a space W_1 of dimension mr. Furthermore, $W_1 \subseteq \bigcup_{j=1}^{r} R_j$. Then $\dim(W_1 \cap W) = \dim W_1 + \dim W - \dim(W_1 \cup W) \geq mr + (mn - rm + 1) - mn = 1$. The result follows, since $W_1 \cap W \subseteq \bigcup_{j=1}^{r} R_j \cap W$.

LEMMA 5. *If $T(R_2) \subseteq T(R_2) \subseteq R_2$, then $T(R_1) \subseteq R_1 \cup R_2$.*

Proof. Suppose $x_1 \otimes y_1 \in R_1$, and choose $x_2 \otimes y_2 \in R_1$ such that $\rho(x_1, x_2) = \rho(y_1, y_2) = 2$. Then $\alpha = sT(x_1 \otimes y_1) + tT(x_2 \otimes y_2) \in R_2$ for all non-zero scalars s, t. Now suppose that $T(x_1 \otimes y_1) = \sum_{j=1}^{p} u_j \otimes v_j$, where $\rho(u_1, \cdots, u_p) = \rho(v_1, \cdots, v_p) = p$, and that $T(x_2 \otimes y_2) = \sum_{j=1}^{q} z_j \otimes w_j$, where $\rho(z_1, \cdots, z_q) = \rho(w_1, \cdots, w_q) = q$. Let $u_{p+1}, \cdots, u_m$ be a completion of $u_1, \cdots, u_p$ to a basis for U. It follows that

$$\sum_{j=1}^{q} z_j \otimes w_j = \sum_{j=1}^{m} u_j \otimes h_j$$

for some vectors $h_j \in V$, $j = 1, \cdots, m$. Then

$$\begin{aligned}\alpha &= \sum_{j=1}^{p} u_j \otimes sv_j + \sum_{j=1}^{p} u_j \otimes th_j + \sum_{j=p+1}^{m} u_j \otimes th_j \\ &= \sum_{j=1}^{p} u_j \otimes (sv_j + th_j) + \sum_{j=p+1}^{m} u_j \otimes th_j \,.\end{aligned}$$

Since $\alpha \in R_2$, it follows by Lemma 1 that

$$\rho(sv_1 + th_1, \cdots, sv_p + th_p) \leq 2 \quad \text{for} \quad st \neq 0 \,.$$

The vectors $sv_1 + th_1, \cdots, sv_p + th_p$ are linearly independent when $s = 1$ and $t = 0$. By continuity, they remain independent for small values of t. Hence $p \leq 2$ and $T(x_1 \otimes y_1) \in R_1 \cup R_2$.

THEOREM 2. *If $T(R_2) \subseteq R_2$ and $0 \notin T(\bigcup_{j=1}^{4} R_j)$, then $T(R_1) \subseteq R_1$.*

Proof. Suppose $x_1 \otimes y_1 \in R_1$ and $T(x_1 \otimes y_1) \notin R_1$. By Lemma 5, $T(x_1 \otimes y_1) \in R_2$, since $0 \notin T(R_1)$. Thus $T(x_1 \otimes y_1) = (u_1 \otimes v_1) + (u_2 \otimes v_2)$, where $\rho(u_1, u_2) = \rho(v_1, v_2) = 2$. Let $x_1, \cdots, x_m$ and $y_1, \cdots, y_n$ be bases for U and V respectively. Then for $st \neq 0$

$$sT(x_1 \otimes y_1) + tT(x_i \otimes y_j) \in R_1 \cup R_2 \qquad \text{for } i = 1, \cdots, m,\ j = 1, \cdots, n. \tag{12}$$

At this point it seems simpler to regard the images $T(x_i \otimes y_j)$ as elements of M_{mn}. It is clear that there is no loss in generality in taking $T(x_1 \otimes y_1) = E_{11} + E_{22}$.

Let i and j be fixed for this discussion, and let $A = T(x_i \otimes y_j)$. Let $a_1, \cdots, a_n$ be the m-dimensional vectors which are the columns of A, and let ε_k be the unit vector with 1 in the kth position. It follows from (12) that

$$\rho(s\varepsilon_1 + ta_1, s\varepsilon_2 + ta_2, ta_3, \cdots, ta_n) = 2 \tag{13}$$

for $st \neq 0$. The Grassmann products

$$(s\varepsilon_1 + ta_1) \wedge (s\varepsilon_2 + ta_2) \wedge ta_k \,, \quad 3 \leq k \leq n \tag{14}$$

must be zero for $st \neq 0$. In the expansion of (14) the coefficient of s^2t is 0; that is, $\varepsilon_1 \wedge \varepsilon_2 \wedge a_k = 0$.

Thus the matrix A has non-zero entries only in the first two rows and columns. It follows immediately that the dimension of the range of $T \leq 2(m+n) - 4$. Hence the dimension of the kernel of $T \geq mn - 2(m+n) + 4 > mn - 4\max(m, n) + 1$.

By Lemma 4, there exists an element of $\bigcup_{j=1}^{4}$ whose image is zero. This contradicts the hypothesis; hence $T(R_1) \subseteq R_1$.

We see then that the form of T satisfying Theorem 2 is given in the conclusions of Theorem 1.

REMARK. We feel that the hypothesis $0 \notin T(\bigcup_{j=1}^{4} R_j)$ of Theorem 2 should not be necessary, but we have not been able to prove the theorem without it. More generally, we conjecture that $T(R_k) \subseteq R_k$ for some fixed k, $1 \leq k \leq n$, should suffice to prove that T is essentially a tensor product.

REFERENCES

1. E. B. Dynkin, *Maximal subgroups of the classical groups*, A.M.S. Translations, Series 2, **6** (1957), 245-378. (Original: Trudy Moskov. Mat. Obsc. **1** (1952), 39-166.)

2. L. K. Hua, *A theorem on matrices and its application to Grassmann space*, Science Reports of the National Tsing Hua University, Ser. A, **5** (1948), 150-181.

3. H. G. Jacob, *Coherence invariant mappings on Kronecker products*, Amer. J. Math. **77** (1955), 177-189.

4. Marvin Marcus and B. N. Moyls, *Linear transformations on algebras of matrices*, Can. J. Math. **11** (1959) 61-66.

5. Marvin Marcus and R. Purves, *Linear transformations on algebras of matrices: The invariance of the elementary symmetric functions*, Can. J. Math. **11** (1959), 383-396.

UNIVERSITY OF BRITISH COLUMBIA

THE NILPOTENT PART OF A SPECTRAL OPERATOR

CHARLES A. McCARTHY

1. Introduction. Throughout this paper, $\mathfrak{X}$ is a Banach space, T a bounded spectral operator on $\mathfrak{X}$ with scalar part S, nilpotent part N, and resolution of the identity $E(\sigma)$ for σ a Borel set in the complex plane. M is the bound for the norms of the $E(\sigma)$; $|E(\sigma)| \leq M$ for all Borel sets σ. The resolvent function for $T, (\lambda - T)^{-1}$, is denoted by $R(\lambda, T)$. The operator $R(\lambda, T)E(\sigma)$ has an unique analytic extension from the resolvent set of T to the complement of $\bar{\sigma}$, and on the subspace $E(\sigma)\mathfrak{X}$ it is equal to the operator $R(\lambda, T_\sigma)$ where T_σ is the restriction of T to $E(\sigma)\mathfrak{X}$. For material on spectral operators, we refer to the papers on N. Dunford [1], [2]. $\chi_\sigma(\xi)$ is the characteristic function of the Borel set σ: $\chi_\sigma(\xi) = 1$ if $\xi \in \sigma$, $\chi_\sigma(\xi) = 0$ if $\xi \notin \sigma$. For p a non-negative real number, μ_p is Hausdorff p-dimensional measure [3, pp. 102 *ff.*]; μ_2 is Lebesgue planar measure multiplied by $\pi/4$, and μ_1 restricted to an arc is majorized by arc length.

We assume throughout that there is an integer m for which the resolvent function for T satisfies the mth order rate of growth condition

$$|R(\lambda, T)E(\sigma)| \leq K \cdot d(\lambda, \sigma)^{-m}, \lambda \notin \bar{\sigma}, |\lambda| \leq |T| + 1 ,$$

where $d(\lambda, \sigma)$ is the distance from λ to σ and K is a constant independent of σ. If $\mathfrak{X}$ is Hilbert space, it is known that this growth condition implies $N^m = 0$ [1, p. 337]. In an arbitrary Banach space, this is no longer true; the best that can be done is $N^{m+2} = 0$. If $\mathfrak{X}$ is weakly complete, $N^{m+1} = 0$; or if σ is a set of μ_2 measure zero, $N^{m+1}E(\sigma) = 0$. If σ lies in an arc and either $\mathfrak{X}$ is weakly complete or σ has μ_1 measure zero, then $N^m E(\sigma) = 0$. Examples show that we cannot obtain lower indices of nilpotency in general.

2. The fundamental lemma and some easy consequences. If $f(\xi)$ is a bounded, scalar valued Borel function, the operator $\int f(\xi)E(d\xi)$ exists as a bounded operator with norm at most $4M \cdot \sup_\xi |f(\xi)|$ [1, p. 341], so that uniform convergence of a sequence of bounded Borel functions $f_n(\xi)$ implies convergence in the uniform operator topology of the operators $\int f_n(\xi)E(d\xi)$. Thus for a given bounded Borel function $f(\xi)$ and a given positive number η, there exist a finite number of disjoint Borel sets σ_i and points $\xi_i \in \sigma_i$ such that

Received February 5, 1959. This paper is a portion of a doctoral dissertation presented to Yale University, written under the direction of Professor E. Hille, while the author was an NSF fellow. Particular thanks are due to W. G. Bade who read the manuscript and discovered an error in the author's original proof of Theorem 3.1.

$$\left|\int f(\xi)E(d\xi) - \sum_i f(\xi_i)E(\sigma_i)\right| < \eta .$$

Similiarly if A_n are a finite number of bounded operators and $f_n(\xi)$ are bounded Borel functions, for any positive number η, there exist a finite number of disjoint Borel sets σ_i and points $\xi_i \in \sigma_i$ such that

$$\left|\sum_n \int A_n f_n(\xi)E(d\xi) - \sum_i \sum_n A_n f_n(\xi_i)E(\sigma_i)\right| < \eta;$$

in particular, for an integer k and a positive number η, there exist a finite number of disjoint Borel sets σ_i and points $\xi_i \in \sigma_i$ such that

$$\left|\int (T-\xi)^k E(d\xi) - \sum_i (T-\xi_i)^k E(\sigma_i)\right| < \eta .$$

LEMMA 2.1. *There exist constants M_k such that $|N^k E(\sigma)| \leq M_k \varepsilon^{k+1-m}$ for any choice of $\varepsilon, 0 < \varepsilon \leq 1$, and Borel set σ of diameter no greater than ε.*

Proof. Pick $\varepsilon, 0 < \varepsilon \leq 1$, and let σ be any Borel set of diameter no greater than ε. We have [1, p, 338]

$$N^k E(\sigma) = \int_\sigma (T-\xi)^k E(d\xi) .$$

For any positive number η, there is a decomposition of σ into a finite number of disjoint Borel sets $\sigma_i \subset \sigma$, and points $\xi_i \in \sigma_i$ such that

$$\left|\int (T-\xi)^k E(d\xi) - \sum_i (T-\xi_i)^k E(\sigma_i)\right| < \eta .$$

Since σ is of diameter at most ε, there is a circle Γ of diameter 3ε which encloses σ and for which $|\gamma - \xi| \geq \varepsilon$ for all $\gamma \in \Gamma$ and $\xi \in \sigma$. Then

$$(T-\xi_i)^k E(\sigma_i) = \frac{1}{2\pi i}\int_\Gamma (\gamma - \xi)^k R(\gamma, T)E(\sigma_i)d\gamma ,$$

so that

$$\sum_i (T-\xi_i)^k E(\sigma_i) = \frac{1}{2\pi i}\int_\Gamma R(\gamma, T)\sum_i (\gamma-\xi_i)^k E(\sigma_i)d\gamma ,$$

which in norm in no greater than

$$(*) \qquad \frac{1}{2\pi}\cdot \sup_{\gamma\in\Gamma}|R(\gamma, T)E(\sigma)| \cdot \sup_{\gamma\in\Gamma}\left|\sum_i (\gamma-\xi_i)^k E(\sigma_i)\right| \cdot \text{length of } \Gamma .$$

The mth order rate of growth condition gives

$$\sup_{\gamma\in\Gamma}|R(\gamma, T)E(\sigma)| \leq K\varepsilon^{-m} .$$

For any $\gamma \in \Gamma$,

$$|\textstyle\sum_i (\gamma - \xi_i)^k E(\sigma_i)| \leq 4M \cdot \max_i |\gamma - \xi_i|^k \leq 4M(2\varepsilon)^k ,$$

so that (*) is no greater than

$$\frac{1}{2\pi} K\varepsilon^{-m} \cdot 4M(2\varepsilon)^k \cdot 6\pi\varepsilon = M_k \varepsilon^{k+1-m} ,$$

where $M_k = 3 \cdot 2^{k+2}KM$, and is independent of $\eta, \varepsilon, \sigma$, and the manner in which σ is decomposed. Thus

$$|N^k E(\sigma)| \leq M_k \varepsilon^{k+1-m} + \eta$$

for every positive η, which proves the lemma.

THEOREM 2.2. *Let σ be a Borel set whose Hausdorff p-measure is zero for a given p. Then $N^k E(\sigma) = 0$ where k is an integer and $k \geq p + m - 1$.*

Proof. Since σ has p-measure zero, for every $\varepsilon > 0$, there is a covering of σ by disjoint sets σ_i of diameter ε_i such that $\sum_i \varepsilon_i^p < \varepsilon$. By Lemma 2.1 we have

$$\begin{aligned} |N^k E(\sigma)| &\leq \textstyle\sum_i |N^k E(\sigma_i)| \leq M_k \sum_i \varepsilon_i^{k+1-m} \\ &\leq M_k \textstyle\sum_i \varepsilon_i^{(p+m-1)+1-m} \leq M_k \sum_i \varepsilon_i^p \leq M_k \varepsilon . \end{aligned}$$

Since ε may be chosen arbitrarily small, $N^k E(\sigma) = 0$.

COROLLARY 2.3. $N^{m+2} = 0$.

Proof. Taking σ to be the spectrum of T and $p = 3$, $N^{m+2}E(\sigma(T)) = 0$; but $E(\sigma(T))$ is the identity mapping on $\mathfrak{X}$.

COROLLARY 2.4. *If σ has planar measure zero, then $N^{m+1}E(\sigma) = 0$.*

COROLLARY 2.5. *If σ has μ_1-measure zero, then $N^m E(\sigma) = 0$.*

3. **The case of weakly complete $\mathfrak{X}$.** Let σ be a Borel set in the plane. For any $\varepsilon > 0$, we can cover σ with disjoint Borel sets. σ_i of diameter ε_i, $\varepsilon_i \leq 1$, such that

$$\textstyle\sum_i \varepsilon_i^2 \leq \mu_2(\sigma) + \varepsilon .$$

Thus by Lemma 2.1,

$$|N^{m+1}E(\sigma)| \leq \sum_i |N^{m+1}E(\sigma_i)| \leq M_{m+1} \sum_i \varepsilon_i^2$$
$$\leq M_{m+1}(\mu_2(\sigma) + \varepsilon) .$$

Since ε and σ are arbitrary, we have for all Borel sets σ,

$$|N^{m+1}E(\sigma)| \leq M_{m+1}\mu_2(\sigma) .$$

As a consequence, all the scalar measures $x^*N^{m+1}E(\cdot)x = [(N^*)^{m+1}E^*(\cdot)x^*]x$, $x \in \mathfrak{X}$, $x^* \in \mathfrak{X}^*$, are absolutely continuous with respect to μ_2, and have derivative bounded by $M_{m+1}|x^*||x|$.

Suppose that $f(\xi) = \sum_{p=1}^{P} \alpha_p \chi_{\sigma_p}(\xi)$ is a simple Borel function; α_p are scalar constants and σ_p are disjoint Borel sets. We have

$$\left|\int f(\xi)(N^*)^{m+1}E^*(d\xi)\right| \leq \sum_{p=1}^{P} |\alpha_p(N^*)^{m+1}E^*(\sigma_p)|$$
$$\leq \sum_{p=1}^{P} |\alpha_p| M_{m+1}\mu_2(\sigma_p)$$
$$= M_{m+1}|f|_{L_1(\mu_2)} .$$

Thus if $f_n(\xi)$ are simple Borel functions converging in $L_1(\mu_2)$ to $f(\xi)$, the operators $\int f_n(\xi)(N^*)^{m+1}E^*(d\xi)$ converge in the uniform operator topology to an operator which we denote by $\int f(\xi)(N^*)^{m+1}E^*(d\xi)$; this limit operator has norm bounded by $M_{m+1}|f|L_1(\mu_2)$.

THEOREM 3.1. *If $\mathfrak{X}$ is weakly complete, then $N^{m+1} = 0$.*

Proof. Assume that $N^{m+1} \neq 0$, so that also $(N^*)^{m+1} \neq 0$. We will first obtain a bicontinuous map of an infinite dimensional L_1 space into $\mathfrak{X}^*$. An analogous map into $\mathfrak{X}$ would show then that $\mathfrak{X}$ cannot be reflexive, since the image in $\mathfrak{X}$ of this L_1 space would be a closed nonreflexive subspace of $\mathfrak{X}$; however, the map into $\mathfrak{X}^*$ is needed for the slightly more general case of $\mathfrak{X}$ weakly complete.

Let the Borel set $\sigma, x_0 \in \mathfrak{X}$, and $x_0^* \in \mathfrak{X}^*$ be chosen so that $[(N^*)^{m+1}E^*(\sigma)x_0^*]x_0 \neq 0$, and let the derivative of the measure $[(N^*)^{m+1}E^*(\cdot)x_0^*]x_0$ be denoted by $g(\xi)$. We can then find a subset τ of σ and a constant $a > 0$ such that $\mu_2(\tau) > 0$ and $|g(\xi)| \geq a$ on τ.

Define the map Φ of $L_1(\tau, \mu_2)$ into $\mathfrak{X}^*$ by

$$\Phi(f) = \int_\tau f(\xi)(N^*)^{m+1}E^*(d\xi)x_0^* .$$

Φ is a linear map with bound $M_{m+1}|x_0^*|$. Now take

$$x = \int_\tau [g(\xi)]^{-1} \operatorname{sgn} \overline{f(\xi)} E(d\xi)x_0 ;$$

The norm of x is no greater than $4M \cdot a^{-1} \cdot |x_0|$. But we have

$$\begin{aligned}[\Phi(f)](x) &= \int_\tau f(\xi)[g(\xi)]^{-1} \operatorname{sgn} \overline{f(\xi)}[(N^*)^{m+1}E^*(d\xi)x_0^*]x_0 \\ &= \int_\tau |f(\xi)|[g(\xi)]^{-1}g(\xi)\mu_2(d\xi) \\ &= |f|_{L_1} ,\end{aligned}$$

which shows that

$$|\Phi(f)| \geq |f|_{L_1} \cdot a \cdot (4M|x_0|)^{-1} ,$$

so that Φ is one-to-one and has a continuous inverse.

Now let Ψ be the map of $L_\infty(\tau, \mu_2)$ into $\mathfrak{X}$:

$$\Psi(h) = \int_\tau [g(\xi)]^{-1}h(\xi)E(d\xi)x_0 ,$$

Ψ is a continuous map with bound no greater than $4M \cdot a^{-1}|x_0|$; we will show that Ψ is one-to-one and bicontinuous. We have

$$\begin{aligned}\Phi(f)\Psi(h) &= \int_\tau f(\xi)[g(\xi)]^{-1}h(\xi)[(N^*)^{m+1}E^*(d\xi)x_0^*]x_0 \\ &= \int_\tau f(\xi)h(\xi)\mu_2(d\xi) ,\end{aligned}$$

so that

$$\begin{aligned}\sup_{|f|_{L_1}\leq 1} |\Phi(f)\Psi(h)| &= \sup_{|f|_{L_1}\leq 1} \left|\int_\tau f(\xi)h(\xi)\mu_2(d\xi)\right| \\ &= |h|_{L_\infty} .\end{aligned}$$

But since Φ is bounded,

$$\begin{aligned}\sup_{|f|_{L_1}\leq 1} |\Phi(f)\Psi(h)| &\leq \sup_{\substack{x^*\in X^* \\ |x^*|\leq|\Phi|}} |x^*\Psi(h)| \\ &= |\Phi|\,|\Psi(h)| ,\end{aligned}$$

so that

$$|h|_{L_\infty} \leq |\Phi|\,|\Psi(h)| ;$$

thus Ψ is one-to-one and bicontinuous, The range $\mathfrak{Y}$ of Ψ in $\mathfrak{X}$ is then a closed non weakly complete subspace of $\mathfrak{X}$. But this is impossible, because every closed subspace of a weakly complete Banach space is again weakly complete; the proof of this last remark is as follows.

Let $\mathfrak{X}$ be a weakly complete Banach space, $\mathfrak{Y}$ a closed subspace. Let y_n be a weakly Cauchy sequence in $\mathfrak{Y}$, so that y^*y_n is a Cauchy sequence of numbers for every y^* in Y^*. Since any x^* in X^*, when restricted to $\mathfrak{Y}$, is an element of $\mathfrak{Y}^*$, x^*y_n is a Cauchy sequence of numbers for every x^* in $\mathfrak{X}^*$. Since $\mathfrak{X}$ is weakly complete, there is an x_0 in $\mathfrak{X}$ such that $\lim_{n\to\infty} x^*y_n = x^*x_0$ for every x^* in $\mathfrak{X}^*$; and since $\mathfrak{Y}$ is strongly closed in $\mathfrak{X}$, it is weakly closed, so that x_0 must lie in $\mathfrak{Y}$. Finally since every y^* in $\mathfrak{Y}^*$ is, by the Hahn-Banach theorem, the restriction of an x^* in $\mathfrak{X}^*$, $\lim y^*y_n = y^*x_0$ for every y^* in $\mathfrak{Y}^*$, so that $\mathfrak{Y}$ is weakly complete.

THEOREM 3.2. *If $\mathfrak{X}$ is weakly complete, then $N^mE(\sigma) = 0$ for every set σ of finite μ_1-measure.*

Proof. Follow exactly the same discussion above, replacing the number $m+1$ by m and the measure μ_2 by μ_1.

Note that Theorems 3.1 and 3.2 also hold if $\mathfrak{X}$ is assume to be separable instead of weakly complete, for the image of the L_∞ space in $\mathfrak{X}$ would be a nonseparable closed subspace of $\mathfrak{X}$; but every closed subspace of a separable space is again separable.

4. Examples. In the following examples we will need two computational lemmas.

LEMMA 4.1. *For each real number $p \geq 1$ and Borel set σ,*

$$\int_\tau |\lambda - \xi|^{-(p+2)}\mu_2(d\xi) \leq 8d(\lambda, \sigma)^{-p}, \text{ for all } \lambda \notin \bar{\sigma} .$$

Proof.

$$\begin{aligned} &\int_\sigma |\lambda - \xi|^{-(p+2)}\mu_2(d\xi) \\ &\leq \int_{|\lambda-\xi|\geq d(\lambda,\sigma)} |\lambda - \xi|^{-(p+2)}\mu_2(d\xi) \\ &= \frac{4}{\pi}\int_0^{2\pi} d\theta \int_{d(\lambda,\sigma)}^\infty r^{-(p+2)} r\, dr \qquad (\lambda - \xi = re^{i\theta}) \\ &\leq 8d(\lambda, \sigma)^{-p} . \end{aligned}$$

LEMMA 4.2. *For each real number $p \geq 1$ and Borel subset σ of the real line,*

$$\int_\sigma |\lambda - \xi|^{-(p+1)}\mu_1(d\xi) \leq 2^{p+1}\pi d(\lambda, \sigma)^{-p} ,$$

where μ_1 is Lebesgue measure along the line, and λ is any complex number, $\lambda \notin \bar{\sigma}$.

Proof. Let $\lambda = \alpha + i\beta$, α, β real. Then either, (i), $d(\alpha, \sigma) \geq d(\lambda, \sigma)/2$ or, (ii) $|\beta| \geq d(\lambda, \sigma)/2$. In case (i) we have

$$\int_\sigma |\lambda - \xi|^{-(p+1)}\mu_1(d\xi) \leq \int_{d(\lambda,\sigma)/2}^\infty \eta^{-(p+1)}d\eta \qquad (\lambda - \xi = \eta)$$
$$\leq 2^{p+1}p^{-1}d(\lambda, \sigma)^{-p} .$$

In case (ii) we have

$$\int_\sigma |\lambda - \xi|^{-(p+1)}\mu_1(d\xi) \leq \int_{-\infty}^\infty |\xi - i\beta|^{-(p+1)}d\xi$$
$$\leq \int_{-\infty}^\infty (\xi^2 + \beta^2)^{-\frac{1}{2}(p+1)}d\xi$$
$$\leq 2^{p+1}\pi d(\lambda, \sigma)^{-p} .$$

EXAMPLE 4.3. Let Σ be a disc in the plane with μ_2-measure 1. Let

$$x = L_\infty(\Sigma) \oplus L_2(\Sigma) \oplus \cdots \oplus L_2(\Sigma) \oplus L_1(\Sigma) ,$$

where m copies of $L_2(\Sigma)$ are taken. Let T be the operator $S + N$ where S and N are defined as

$$S[f(\xi) \oplus g_1(\xi) \oplus \cdots \oplus g_m(\xi) \oplus h(\xi)]$$
$$= [\xi f(\xi) \oplus \xi g_1(\xi) \oplus \cdots \oplus \xi g_m(\xi) \oplus \xi h(\xi)] ,$$
$$N[f(\xi) \oplus g_1(\xi) \oplus \cdots \oplus g_m(\xi) \oplus h(\xi)]$$
$$= [0 \oplus f(\xi) \oplus g_1(\xi) \oplus \cdots \oplus g_m(\xi)] .$$

Since Σ has measure 1, any function in L_r is in L_s for all $s \leq r$, and the L_s norm is no greater than the L_r norm; thus N is a bounded operator with norm 1. Also N is a nilpotent for which $N^{m+1} \neq 0$. The operator T is a spectral operator with resolution of the identity

$$E(\sigma)[f(\xi) \oplus g_1(\xi) \oplus \cdots \oplus g_m(\xi) \oplus h(\xi)]$$
$$= [f(\xi)\chi_\sigma(\xi) \oplus g_1(\xi)\chi_\sigma(\xi) \oplus \cdots \oplus g_m(\xi)\chi_\sigma(\xi) \oplus h(\xi)\chi_\sigma(\xi)] .$$

The resolvent function is

$$R(\lambda, T)E(\sigma)[f(\xi) \oplus g_1(\xi) \oplus \cdots \oplus g_m(\xi) \oplus h(\xi)]$$
$$= \Big[\frac{f(\xi)\chi_\sigma(\xi)}{\lambda - \xi} \oplus \Big(\frac{g_1(\xi)\chi_\sigma(\xi)}{\lambda - \xi} + \frac{f(\xi)\chi_\sigma(\xi)}{(\lambda - \xi)^2}\Big) \oplus \cdots \oplus$$
$$\Big(\frac{g_m(\xi)\chi_\sigma(\xi)}{\lambda - \xi} + \cdots + \frac{g_1(\xi)\chi_\sigma(\xi)}{(\lambda - \xi)^m} + \frac{f(\xi)\chi_\sigma(\xi)}{(\lambda - \xi)^{m+1}}\Big)$$
$$\oplus \Big(\frac{h(\xi)\chi_\sigma(\xi)}{\lambda - \xi} + \frac{g_m(\xi)\chi_\sigma(\xi)}{(\lambda - \xi)} + \cdots + \frac{g_1(\xi)\chi_\sigma(\xi)}{(\lambda - \xi)^{m+1}} + \frac{f(\xi)\chi_\sigma(\xi)}{(\lambda - \xi)^{m+2}}\Big)\Big] .$$

All the terms are clearly of mth order rate of growth except possibly for

$$\text{(a)}\left|\frac{f(\xi)\chi_\sigma(\xi)}{(\lambda-\xi)^{m+1}}\right|_{L_2},\quad \text{(b)}\left|\frac{f(\xi)\chi_\sigma(\xi)}{(\lambda-\xi)^{m+2}}\right|_{L_1},\ \text{and}\ \text{(c)}\left|\frac{g_1(\xi)\chi_\sigma(\xi)}{(\lambda-\xi)^{m+1}}\right|_{L_1}.$$

For (a) we have

$$\left\{\int_\sigma |f(\xi)(\lambda-\xi)^{-(m+1)}|^2\mu_2(d\xi)\right\}^{1/2} \le |f|_{L_\infty}\left\{\int_\sigma |\lambda-\xi|^{-2m-2}\mu_2(d\xi)\right\}^{1/2}$$
$$\le |f|_{L_\infty}\sqrt{8}\,d(\lambda,\sigma)^{-m},$$

for (b) we have

$$\int_\sigma |f(\xi)(\lambda-\xi)^{-(m+2)}|\,\mu_2(d\xi) \le |f|_{L_\infty}\int_\sigma |\lambda-\xi|^{-(m+2)}\mu_2(d\xi)$$
$$\le |f|_{L_\infty}\cdot 8d(\lambda,\sigma)^{-m},$$

and for (c) we have

$$\int_\sigma |g_1(\xi)(\lambda-\xi)^{-(m+1)}|\,\mu_2(d\xi) \le \left\{\int_\sigma |g_1(\xi)|^2\mu_2(d\xi)\right\}^{1/2}\left\{\int_\sigma |\lambda-\xi|^{-2m-2}\mu_2(d\xi)\right\}^{1/2}$$
$$\le |g_1|_{L_2}\cdot\sqrt{8}\cdot d(\lambda,\sigma)^{-m}.$$

Thus each term of the resolvent, and hence the resolvent itself satisfies the mth order rate of growth condition; this shows that Corollary 2.3 cannot be improved.

EXAMPLE 4.4. Let Σ be as in the previous example and let

$$\mathfrak{X} = L_r(\Sigma)\oplus\cdots\oplus L_r(\Sigma)\oplus L_s(\Sigma)$$

where m copies of L_r are taken. r and s are to satisfy $1 < s < r < \infty$ and $rs \le 2(r-s)$. Let $T = S + N$, where S and N are defined in essentially the same way as in the previous example. The resolvent function is given by

$$R(\lambda,T)E(\sigma)[f_1(\xi)\oplus\cdots\oplus f_m(\xi)\oplus g(\xi)]$$
$$=\left[\frac{f_1(\xi)\chi_\sigma(\xi)}{\lambda-\xi}\oplus\cdots\oplus\left(\frac{f_m(\xi)\chi_\sigma(\xi)}{\lambda-\xi}+\cdots+\frac{f_1(\xi)\chi_\sigma(\xi)}{(\lambda-\xi)^m}\right)\right.$$
$$\left.\oplus\left(\frac{g(\xi)\chi_\sigma(\xi)}{\lambda-\xi}+\frac{f_m(\xi)\chi_\sigma(\xi)}{(\lambda-\xi)^2}+\cdots+\frac{f_1(\xi)\chi_\sigma(\xi)}{(\lambda-\xi)^{m+1}}\right)\right].$$

Each of the terms is clearly of mth order rate of growth except possibly for the L_s norm of $f_1(\xi)(\lambda-\xi)^{-(m+1)}\chi_\sigma(\xi)$, and for this we have

$$\left\{\int_\sigma |f_1(\xi)(\lambda-\xi)^{m+1}|^s\mu_2(d\xi)\right\}^{1/s}$$
$$\le\left\{\int_\sigma |f_1(\xi)|^r\mu_2(d\xi)\right\}^{1/r}\left\{\int|\lambda-\xi|^{-\frac{(m+1)rs}{r-s}}\mu_2(d\xi)\right\}^{\frac{r-s}{rs}}$$
$$\le |f_1|_{L_r}\cdot 8^{\frac{s-r}{rs}}\cdot d(\lambda,\sigma)^{-m-\left(1-\frac{2(r-s)}{rs}\right)}$$
$$\le |f_1|_{L_r}\cdot 8^{\frac{r-s}{rs}}d(\lambda,\sigma)^{-m}$$

Thus the resolvent satisfies the mth order rate of growth condition, and $N^m = 0$. Since $\mathfrak{X}$ is reflexive, this shows that Theorem 3.1 cannot be improved. Note that $\mathfrak{X}$ is also separable.

EXAMPLE 4.5. Let Σ be the interval $[0, 1]$ endowed with μ_1-measure, and let

$$\mathfrak{X} = L_\infty(\Sigma) \oplus \cdots \oplus L_\infty(\Sigma) \oplus L_1(\Sigma)$$

where m copies of L_∞ are taken. Let $T = S + N$ where S and N are defined in essentially the same way as in the previous examples. The resolvent function is given by

$$\begin{aligned} &R(\lambda, T)E(\sigma)[f_1(\xi) \oplus \cdots \oplus f_m(\xi) \oplus g(\xi)] \\ &\quad = \Big[\frac{f_1(\xi)\chi_\sigma(\xi)}{\lambda - \xi} \oplus \cdots \oplus \Big(\frac{f_m(\xi)\chi_\sigma(\xi)}{\lambda - \xi} + \cdots + \frac{f_1(\xi)\chi_\sigma(\xi)}{(\lambda - \xi)^m}\Big) \\ &\qquad \oplus \Big(\frac{g(\xi)\chi_\sigma(\xi)}{\lambda - \xi} + \frac{f_m(\xi)\chi_\sigma(\xi)}{(\lambda - \xi)^2} + \cdots + \frac{f_1(\xi)\chi_\sigma(\xi)}{(\lambda - \xi)^{m+1}}\Big)\Big] . \end{aligned}$$

Each of the terms is clearly of mth order rate of growth except for the L_1 norm of $f_1(\xi)(\lambda - \xi)^{-(m+1)}\chi_\sigma(\xi)$, and for this we have

$$\begin{aligned} \int_\sigma |f(\xi)(\lambda - \xi)^{-(m+1)}|\mu_1(d\xi) &\leq |f|_{L_\infty} \int_\sigma |\lambda - \xi|^{-(m+1)}\mu_1(d\xi) \\ &\leq |f|_{L_\infty} 2^{m+1}\pi d(\lambda, \sigma)^{-m} . \end{aligned}$$

Thus we have an example of an operator with spectrum in a rectifiable arc which satisfies the mth order rate of growth condition, but for which $N^m \neq 0$.

BIBLIOGRAPHY

1. N. Dunford, *Spectral operators*, Pacific J. Math. **4** (1954), 321-354.
2. ———, *A survey of the theory of spectral operators*, Bul. AMS **64** (1958), 217-274.
3. Hurewicz and Wallman, *Dimension theory*, Princeton University press, 1948.

YALE UNIVERSITY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

ON A CRITERION FOR THE WEAKNESS OF AN IDEAL BOUNDARY COMPONENT

KÔTARO OIKAWA

1. Exhaustion. Let F be an open Riemann surface. An *exhaustion* $\{F_n\}$ of F is an increasing (i.e., $\overline{F}_n \subset F_{n+1}$) sequence of subregions with compact closures such that $\bigcup_{n=1}^{\infty} F_n = F$. We assume that ∂F_n consists of a finite number of closed analytic curves and that each component of $F - F_n$ is noncompact. This is the most common definition used in the theory of open Riemann surfaces. Sometimes, however, we shall add the restriction that each component of ∂F_n is a dividing cycle; if this is the case we shall call the exhaustion *canonical*.

2. Weak boundary component. Let γ be an ideal boundary component of F, and let $\{F_n\}$ be a canonical exhaustion of F. Then there exists a component γ_n of ∂F_n which separates γ from F_n. Let n_0 be a fixed number and consider the component G_n of $\overline{F}_n - F_{n_0}$ $(n > n_0)$ such that $\gamma_n \subset \partial G_n$. There exists a harmonic function $s_n(p)$ on $\overline{G}_n$ which satisfies the following conditions:

(i) $s_n = 0$ on γ_{n_0} and $\int_{\gamma_{n_0}} *ds_n = 2\pi$, $(\gamma_{n_0} = \partial F_{n_0} \cap \partial G_n)$

(ii) $s_n = \log r_n = \text{const.}$ on γ_n,

(iii) $s_n = \text{const.}$ on each component $\beta_{n\nu}$ of $\partial G_n - \gamma_n - \gamma_{n_0}$ and $\int_{\beta_{n\nu}} *ds_n = 0$.

The condition $\lim_{n\to\infty} r_n = \infty$ depends neither on n_0 nor on the exhaustion. If it is satisfied, γ is said to be *weak*.

Weak boundary components were introduced for plane regions by Grötzch [1] in connection with the so-called Kreisnormierungsproblem. He called them vollkommen punktförmig. They were generalized for open Riemann surfaces by Sario [6] and discussed also by Savage [7] and Jurchescu [2]. The above definition was given by Jurchescu [2].

A noncompact subregion N whose relative boundary ∂N consists of a finite number of closed analytic curves is called a *neighborhood of* γ if γ is an ideal boundary component of N as well. Let $\{c\}$ be the family of all cycles c (i.e., unions of finite numbers of closed curves) which are in N and separate γ from ∂N. Jurchescu [2] showed that $\lambda\{c\} = 0$ if and only if γ is weak, where $\lambda\{c\}$ is the extremal length of the family $\{c\}$.

Received October 22, 1958. This paper was prepared under Contract No. DA-04-495-ORD-722, OOR Project No. 1517 between the University of California, Los Angeles and the Office of Ordnance Research, U. S. Army.

3. Savage's criterion. Let $\{F_n\}$ be an arbitrary exhaustion. Let E_n be the smallest union of components of $F_n - \overline{F}_{n-1}$ such that $\gamma_{n-1} = \partial E_n \cap \partial F_{n-1}$ is a cycle which separates γ from F_{n-1} $(n = 2, 3, \cdots)$. Evidently $\gamma_n \subset \partial E_n$. If $\{F_n\}$ is canonical, E_n is connected and γ_n is a closed analytic curve.

There exists a harmonic function $u_n(p)$ on $\overline{E}_n$ such that

(i) $u_n = 0$ on γ_{n-1} and $\int_{\gamma_{n-1}} *du_n = 2\pi$,

(ii) $u_n = \log \mu_n = \text{const.}$ on $\partial E_n - \gamma_{n-1} = \partial E_n \cap \partial F_n$.

The quantity $\log \mu_n$ is called the *modulus of* E_n (cf. Sario [4,5], who called μ_n the modulus). It is expressed in terms of extremal length as follows:

$$\log \mu_n = \frac{2\pi}{\lambda\{c\}_n},$$

where $\{c\}_n$ is the family of cycles in E_n homologous to γ_{n-1}.

Since $\sum^\infty 1/\lambda\{c\}_n \leqq 1/\lambda\{c\}$, we get the following criterion:

THEOREM 1 (Savage [7]). *If there exists an exhaustion such that* $\prod_{n=2}^\infty \mu_n = \infty$, *then* γ *is weak.*

The purpose of the present note is to discuss the converse of this theorem.

4. Jurchescu's criterion. Suppose the exhaustion $\{F_n\}$ is canonical. There exists a harmonic function $U_n(p)$ on $\overline{E}_n$ such that

(i) $U_n = 0$ on γ_{n-1} and $\int_{\gamma_{n-1}} *dU_n = 2\pi$,

(ii) $U_n = \log M_n = \text{const.}$ on γ_n,

(iii) $U_n = \text{const.}$ on each component $\beta_{n\nu}$ of $\partial E_n - \gamma_n - \gamma_{n-1}$ and $\int_{\beta_{n\nu}} *dU_n = 0$.

Jurchesch's paper [2] contains implicitly the following result:

THEOREM 2 (Jurchescu). *A boundary component* γ *is weak if and only if there exists a canonical exhaustion such that* $\prod_{n=2}^\infty M_n = \infty$.

Proof. Sufficiency: Let $\{c'\}_n$ be the family of cycles in E_n separating γ_n from γ_{n-1}. It is not difficult to see that $\log M_n = 2\pi/\lambda\{c'\}_n$. Since $\sum^\infty 1/\lambda\{c'\}_n \leqq 1/\lambda\{c\}$, we conclude that $\sum_{n=2}^\infty \log M_n = \infty$ implies $\lambda\{c\} = 0$.

Necessity: Consider a canonical exhaustion $\{F_n^0\}$. The desired exhaustion $\{F_n\}$ is obtained by taking its subsequence as follows:

$F_1 = F_1^0$. To define F_2, consider the quantity r_n introduced in No. 2 with respect to $F_n^0 - \overline{F}_1^0$ $(n = 2, 3, \cdots)$. Take n_2 so large that $r_{n_2} \geqq 2$, and put $F_2 = F^0_{n_2}$. Evidently $M_2 = r_{n_2}$. Similarly, $F_3 = F^0_{n_3}$ is defined by considering $F^0_n - \overline{F}^0_{n_2}$ $(n = n_2 + 1, n_2 + 2, \cdots)$ and by taking $n_3 > n_2$ so large that $r_{n_3} \geqq 2$ where r_{n_2} is the quantity r_n introduced in No. 2 with respect to $F^0_n - \overline{F^0_{n_2}}$. We have $M_3 = r_{n_3}$. On continuing this process, we obtain a canonical exhaustion such that $\sum_{n=2}^{\infty} \log M_n \geqq \sum_{n=2}^{\infty} \log 2 = \infty$. The idea of this proof was first used by Noshiro [3].

5. The converse of Savage's criterion. We shall now show that Savage's criterion in Theorem 1 is also necessary.

THEOREM 3. *If γ is weak, then there exists an exhaustion such that $\prod_{n=2}^{\infty} \mu_n = \infty$. It is not necessarily canonical.*

Proof. By Theorem 2 there exists a canonical exhaustion $\{F^0_n\}$ such that $\prod_{n=2}^{\infty} M^0_n = \infty$. From this we construct a canonical exhaustion $\{F^*_n\}$ as follows:

$F^*_1 = F^0_1$. To construct F^*_2, let $\partial E^0_2 - \gamma^0_1 - \gamma^0_2 = \beta_{21} \cup \beta_{22} \cup \cdots \cup \beta_{2k_2}$ be the decomposition into components, and let H^ν_3 be the component of $F^0_3 - F^0_2$ such that $\partial H^\nu_3 \cap \overline{F}^0_2 = \beta_{2\nu}$ $(\nu = 1, 2, \cdots, k_2)$. F^*_2 is the union of $F^*_1, E^0_2 \cup \gamma^0_1$, all the other components of $F^0_2 - F^0_1$, and $\bigcup_{\nu=1}^{k_2} H^\nu_3$. In this way, F^*_n is defined as the union of F^*_{n-1}, $E^0_n \cup \gamma^0_{n-1}$, every component of $F^0_{m+1} - F^0_m$ $(m \geqq n)$ which is adjacent to F^*_{n-1}, and $\bigcup_{\nu=1}^{k_n} H^\nu_{n+1}$. By construction, $E^*_n = E^0_n \cup \bigcup_{\nu=1}^{k_n} H^\nu_{n+1}$.

The desired exhaustion $\{F_n\}$ is obtained by taking a refinement of $\{F^*_n\}$ as follows: Consider E^0_n and the function U^0_n for the exhaustion $\{F^0_n\}$. Let $\partial E^0_n - \gamma^0_n - \gamma^0_{n-1} = \beta_{n1} \cup \beta_{n2} \cup \cdots \cup \beta_{nk_n}$ be the decomposition into components and let $U^0_n \equiv a_\nu$ on $\beta_{n\nu}$ $(\nu = 1, 2, \cdots, k_n)$. We may assume, without loss of generality, that the a_ν's are different by pairs. We suppose that

$$0 \equiv a_0 < a_1 < \cdots < a_{k_n} < a_{k_n+1} \equiv \log M^0_n.$$

Take a'_ν $(a_{\nu-1} < a'_\nu < a_\nu\,;\ \nu = 1, 2, \cdots, k_n, a'_{k_n+1} \equiv \log M^0_n)$ and a''_ν $(a_\nu < a''_\nu < a_{\nu+1}\,;\ \nu = 1, \cdots, k_n, a''_0 \equiv 0)$ so close to a_ν that

$$(1) \qquad \sum_{\nu=1}^{k_n+1} (a'_\nu - a''_{\nu-1}) \geqq \log M^0_n - 2^{-n} .$$

Consider the sets

$$D^\nu_n = \{p\,; a''_{\nu-1} < U^0_n(p) < a''_\nu\},\ \nu = 1, 2, \cdots, k_n + 1,\ (a''_{k_n+1} \equiv \log M^0_n)$$
$$D'^\nu_n = \{p\,; a''_{\nu-1} < U^0_n(p) < a'_\nu\},\ \nu = 1, 2, \cdots, k_n + 1 .$$

The modulus $\log \mu'^{(\nu)}$ of D'^ν_n with respect to $\beta^\nu = \{p\,; U^0_n(p) = a''_{\nu-1}\}$ and $\partial D'^\nu_n - \beta^\nu$ is equal to $a'_\nu - a''_{\nu-1}$, since the function $U^0_n(p) - a''_{\nu-1}$ plays the role of $u_n(p)$ introduced in No. 3. Let $\log \mu^{(\nu)}$ be the modulus of D^ν_n

with respect to β^{ν} and $\partial D_n^{\nu} - \beta^{\nu}$. Since $\mu^{(\nu)} \geqq \mu'^{(\nu)}$, we obtain, by (1),

$$\sum_{\nu=1}^{k_n+1} \log \mu^{(\nu)} \geqq \log M_n^0 - 2^{-n} . \tag{2}$$

We have decomposed E_n^0 into $k_n + 1$ subsets D_n^{ν}. $E_n^* - E_n^0$ consists of components H_{n+1}^{ν} such that $\beta_{n\nu} = \partial H_{n+1}^{\nu} \cap \partial E_n^0 \, (\nu = 1, 2, \cdots, k_n)$. By decomposing H_{n+1}^{ν} into $k_n - \nu + 1$ slices, we obtain a decomposition of E_n^* into $k_n + 1$ parts. It is possible to divide each of the other components of $F_n^* - \overline{F}_{n-1}^*$ into $k_n + 1$ pieces so that we get an exhaustion $\{F_n\}$ which is a refinement of $\{F_n^*\}$. D_n^{ν} plays the role of E_n with respect to this exhaustion. Therefore, by (2), we get

$$\sum_{n=2}^{\infty} \log \mu_n \geqq \sum_{n=2}^{\infty} \log M_n^0 - 1 = \infty .$$

6. **Remark.** On a "schlichtartig" surface, every exhaustion is canonical. If F is an arbitrary Riemann surface, the question arises whether or not Savage's criterion is still necessary under the restriction that $\{F_n\}$ is canonical. The answer is given by

THEOREM 4. *There exist a γ of an F which is weak and such that $\prod_{n=2}^{\infty} \mu_n < \infty$ for every canonical exhaustion.*

Construction of F: In the plane $|z| < \infty$, consider the closed intervals

$$I_k : [2^{k^2}, 2^{k^2} + 1] \qquad (k = 2, 3, \cdots)$$

on the positive real axis, and the circular arcs

$$\alpha_{\nu} : |z| = \nu, |\arg z| \leqq \frac{\pi}{2}$$

$$(\nu = 2^{k2} + 2, 2^{k^2} + 3, \cdots, 2^{(k+1)^2} - 1 ; k = 2, 3, \cdots) .$$

Take two replicas of the slit plane $(|z| < \infty) - \bigcup_{k=2}^{\infty} I_k$ and connect them crosswise across I_k $(k = 2, 3, \cdots)$. From the resulting surface, delete all the α_{ν}'s on both sheets. This is a Riemann surface F of infinite genus.

F has an ideal boundary component γ over $z = \infty$, which is evidently weak.

Let $\{F_n\}$ be an arbitrary canonical exhaustion. Consider E_n corresponding to γ (No. 3). The interval I_k determines a closed analytic curve C_k on F. Since $\gamma_{n-1} = \partial E_n \cap \overline{F}_{n-1}$ is a dividing cycle, the intersection number $\gamma_{n-1} \times C_k$ vanishes and, therefore, $\gamma_{n-1} \cap C_k$ consists of an even number of points whenever it is not void.* Take two consecutive points

* *Added in proof.* We should have mentioned the case where γ_{n-1} tangents C_k. The following discussion covers this case if the number of the points of $\gamma_{n-1} \cap C_k$ is counted with the multiplicity of tangency and case $p=q$ is not excluded.

p and q in $\gamma_{n-1} \cap C_k$. There are two possibilities according as the arc $\widehat{pq} \subset \gamma_{n-1}$ is homotopic to $\widehat{pq} \subset C_k$ or not. If the latter case happens for at least one pair of p and q, we shall say that γ_{n-1} intersects C_k *properly*.

Since γ_{n-1} is a closed curve separating γ from F_{n-1}, there exists a number k such that γ_{n-1} intersects C_k properly. If there is more than one k, we take the greatest one and denote it by $k(n)$.

To estimate μ_n, let $\{c\}_n$ be the family of all cycles in E_n separating γ_{n-1} from $\partial E_n - \gamma_{n-1}$. We have mentioned that $\log \mu_n = 2\pi/\lambda\{c\}_n$. Let C_k be a curve for which there are numbers n with $k(n) = k$. Evidently these n are finite in number and consecutive. Let n_k be the greatest.

I. If $k(n) = k$ and $n < n_k$ then γ_{n-1} and γ_n intersect C_k properly. Since every $c \in \{c\}_n$ separates γ_{n-1} from γ_n, it has a component which intersects C_k and is not completely contained in the doubly connected region $\varDelta_k$ consisting of all points that lie over $\{z\,;\ 2^{k^2} - 1 < |z| < 2^{k^2} + 2, |\arg z| < \pi/2\}$. Therefore, every c contains a curve in $\{c'\}^{(k)}$ which is the family of all curves in the right half-plane connecting I_k with the imaginary axis. Consequently

$$(3) \qquad \sum_{\substack{k(n)=k \\ n \neq n_k}} \frac{1}{\lambda\{c\}_n} \leqq \frac{1}{\lambda\{c'\}^{(k)}} .$$

II. $k(n) = k$ and $n = n_k$. Consider all the α_ν ($\nu \geqq 2^{k^2} + 2$) on the upper sheet. Let G_{n-1} be the component of $F - \overline{F}_{n-1}$ such that $\partial G_{n-1} = \gamma_{n-1}$. For a sufficiently large ν, α_ν is an ideal boundary component of G_{n-1}. Let $\nu(k)$ be the least ν with this property. If $\nu(k) = 2^{k^2} + 2$, then every $c \in \{c\}_n$ separates γ_{n-1} from $\alpha_{\nu(k)}$ and, therefore, it has a component intersects either C_k or one of four line segments over $[2^{k^2} - 1, 2^{k^2}]$ or $[2^{k^2} + 1, 2^{k^2} + 2]$. When $\nu(k) = 2^{l^2} + 2$ for some $l > k$, then γ_{n-1} separates $\alpha_{\nu(k)-3}$ from $\alpha_{\nu(k)}$ and every $c \in \{c\}_n$ separates γ_{n-1} from $\alpha_{\nu(k)}$, so that c has a component with the above property. If $\nu(k)$ is not of the form $2^{l^2} + 2$, then, for the same reason, every $c \in \{c\}_n$ has a component which intersects the line segment on the upper sheet lying over $[\nu(k) - 1, \nu(k)]$, and is not contained in the simply connected region on the upper sheet consisting of all points over $\{z\,;\ \nu(k) - 1 < |z| < \nu(k), |\arg z| < \pi/2\}$. In any case, every $c \in \{c\}_n$ contains a curve in $\{c''\}^{(k)}$ which is the family of all curves in the right half-plane connecting $[\nu(k) - 3, \nu(k)]$ with the imaginary axis. Therefore,

$$(4) \qquad \frac{1}{\lambda\{c\}_n} \leqq \frac{1}{\lambda\{c''\}^{(k)}} .$$

By (3) and (4), we obtain

$$(5) \qquad \sum_{n=2}^{\infty} \log \mu_n = 2\pi \sum_{n=2}^{\infty} \frac{1}{\lambda\{c\}_n} \leqq 2\pi \sum_{k=2}^{\infty} \left(\frac{1}{\lambda\{c'\}^{(k)}} + \frac{1}{\lambda\{c''\}^{(k)}} \right) .$$

To show the convergence of $\sum_{k=2}^{\infty} 1/\lambda\{c'\}^{(k)}$, we make use of the transformation $z \rightarrow z^2$. It is immediately seen that $\lambda\{c'\}^{(k)}$ is equal to the extremal distance between $[-\infty, 0]$ and $I'_k = [2^{2k^2}, (2^{k^2}+1)^2]$ with respect to the region $A = \{[-\infty, 0] \cup I'_k\}^c$. Since A is conformally equivalent to Teichmüller's extremal region $\{[-1, 0] \cup [P, \infty]\}^c$ where

$$P = \frac{2^{2k^2}}{(2^{k^2}+1)^2 - 2^{2k^2}} ,$$

we have (Teichmüller [8])

$$\begin{aligned} \lambda\{c'\}^{(k)} &\sim \frac{\log P}{2\pi} \qquad (P \rightarrow \infty) \\ &\sim \frac{k^2 \log 2}{2\pi} \qquad (k \rightarrow \infty) , \end{aligned}$$

and, therefore, $\sum_{k=2}^{\infty} 1/\lambda\{c'\}^{(k)} < \infty$. Similarly $\sum_{k=2}^{\infty} 1/\lambda\{c''\}^{(k)} < \infty$ because $\nu(k) \geqq 2^{k^2} + 2$. We conclude that

$$\sum_{n=2}^{\infty} \log \mu_n < \infty .$$

References

1. H. Grötzsch, *Eine Bemerkung zum Koebeschen Kreisnormierungsprinzip*, Ber. Verh. Sächs. Akad. Wiss. Leipzig. Math.-Nat. Kl., **87** (1935), 319-324.
2. M. Jurchescu, *Modulus of a boundary component*, Pacific J. Math., **8** (1958), 791-809.
3. K. Noshiro, *Open Riemann surface with null-boundary*, Nagoya Math. J., **3** (1951), 73-79.
4. L. Sario, *Über Riemannsche Flächen mit hebbarem Rand*, Ann. Acad. Sci. Fenn. Ser. A. I, **50** (1948), pp. 79.
5. ———, *Modular criteria on Rieman surfaces*, Duke Math. J., **20** (1953), 279-286.
6. ———, *Capacity of the boundary and of a boundary component*, Ann. Math., **59** (1954), 135-144.
7. N. Savage, *Weak boundary components of an open Riemann surface*, Duke Math. J., **24** (1957), 79-96.
8. O. Teichmüller, *Untersuchungen über konforme und quasikonforme Abbibldung*, Deutsche Math., **3** (1938), 621-678.

UNIVERSITY OF CALIFORNIA, LOS ANGELES,
AND
TOKYO INSTITUTE OF TECHNOLOGY

AN ALGEBRAIC CRITERION FOR IMMERSION

BARRETT O'NEILL

Let R be the curvature tensor of a simply connected d-dimensional ($d \geq 4$) Riemannian manifold M. T. Y. Thomas [2] has proved that if the rank of R is not too small, there exist conditions expressed in terms of polynomials in the coordinates of R which are satisfied if and only if M can be immersed in the Euclidean space R^{d+1}. The proof is existential; the polynomials are not all given explicitly. Using the notion of Grassmann algebra we shall find a single, rather simple condition on R necessary and sufficient for the existence of an immersion $i: M \to \bar{M}(K)$ with second fundamental form of rank at least four, where $\bar{M}(K)$ is a complete $(d+1)$-dimensional Riemannian manifold of constant curvature K. If coordinates are introduced this condition can be expressed algebraically in terms of polynomial equations and inequalities in the coordinates of R. The case $K = 0$ yields an explicit variant of Thomas' result.

1. A differential criterion for immersion. Following [1] we fix the following notation for the structural elements associated with a d-dimensional C^∞ Riemannian manifold M: $F(M)$, the bundle of frames on M: R_a, right-multiplication of $F(M)$ by $a \in O(d)$, the group of $d \times d$ orthogonal matrices; φ, the 1-form of the Riemannian connection. Thus $\varphi = (\varphi_{ij})$ is a vertical equivariant 1-form on $F(M)$ with values in the Lie algebra of $d \times d$ skew-symmetric matrices. (We assume throughout that $1 \leq i, j, k \leq d$.) Let $\omega = (\omega_i)$ be the usual horizontal equivariant R^d-valued 1-form on $F(M)$ defined by $\omega_i(x) = \langle d\pi(x), f_i \rangle$, where x is in the tangent space $F(M)_f$ to $F(M)$ at $f = (f_1, \cdots, f_d)$ and π is the natural projection. The curvature form $\Phi = (\Phi_{ij})$ is by definition $D\varphi$, the horizontal part of $d\varphi$. In the case of 1-forms or 1-vectors we write xy, rather than $x \wedge y$, for the Grassmann product.

THEOREM 1. *Let M be a simply connected d-dimensional Riemannian manifold, $\bar{M}$ a complete $(d+1)$-dimensional Riemannian manifold of constant curvature K. Then M can be immersed in $\bar{M}$ if and only if there exists a horizontal equivariant R^d-valued 1-form $\sigma = (\sigma_i)$ on $F(M)$ such that*

$$(1)\qquad \begin{cases} \Sigma_k \sigma_k \omega_k = 0 & \\ \Phi_{ij} = \sigma_i \sigma_j + K\omega_i \omega_j & \text{(Gauss equation)} \\ D\sigma_i = 0 & \text{(Codazzi equation)}. \end{cases}$$

Received February 27, 1959.

Proof. Suppose there exists an immersion $i\colon M \to \overline{M}$. Since M is simply connected, there is a unit normal vector field on the immersed manifold, N being a differentiable ($= C^\infty$) map from M to the tangent of $\overline{M}$. Then the formula $\psi(m, f_1, \cdots, f_d) = (i(m), di(f_1), \cdots, di(f_d), N(m))$ defines a differentiable map $\psi\colon F(M) \to F(\overline{M})$. (Denote by $\overline{R}_a, \overline{\varphi}, \cdots$ the structural elements of $\overline{M}$.) Note that $\psi \circ R_a = \overline{R}_a \circ \psi$ if $a \in O(d) \subset O(d+1)$. This fact plus the uniqueness of the Riemannian connection of M are used in the proof that

$$(2)\qquad \begin{cases} \omega_i = \overline{\omega}_i \circ d\psi \\ 0 = \overline{\omega}_{d+1} \circ d\psi \\ \varphi_{ij} = \overline{\varphi}_{ij} \circ d\psi \, . \end{cases}$$

Furthermore, the R^d-valued 1-form defined by (3) $\sigma_i = \varphi_{i,d+1} \circ \psi$ satisfies the conditions stated in the theorem. This form is, of course, one expression for the second fundamental form of the immersed manifold.

Conversely, given a form σ on $F(M)$ with the stated properties we must produce an immersion $i\colon M \to \overline{M}$. To do this we first find a differentiable map $\varphi\colon F(M) \to F(\overline{M})$ satisfying the differential equations (2) and (3). Consider the 1-forms $\overline{\omega}_i - \omega_i, \overline{\omega}_{d+1}, \overline{\varphi}_{ij} - \varphi_{ij}, \overline{\varphi}_{i,d+1} - \sigma_i$ on $F(M) \times F(\overline{M})$, where we use the same notation for a form on one factor and that form pulled back to the product manifold by a projection. We want to apply the Frobenius theorem to these forms. Taking account of the structural equations one sees that its hypothesis holds provided $\Sigma_k \sigma_k \omega_k = 0$; $d\varphi_{ij} = -\Sigma_k \varphi_{ik}\varphi_{kj} + \sigma_i\sigma_j + K\omega_i\omega_j$; and $d\sigma_i = -\Sigma_k \varphi_{ik}\sigma_k$. But these conditions follow from the corresponding equations in (1)—in the case of the last one because for σ (or any other R^d-valued horizontal equivariant 1-form on $F(M)$ we have $d\sigma_i = -\Sigma\varphi_{ik}\sigma_k + D\sigma_i$. Then if $(g, \overline{g}) \in F(M) \times F(\overline{M})$, an integral manifold through $(g, \overline{g})$ given by the Frobenius theorem is the graph of a differentiable function ψ' defined on a neighborhood U of $g \in F(M)$, carrying g to $\overline{g}$, and satisfying (2) and (3). Subject to these conditions φ' is unique, except for the size of its domain. Further, one can show that φ' commutes with right-multiplication in the sense that, where meaningful, $\varphi' \circ R_a$ and $\overline{R}_a \circ \varphi'$ agree. This fact permits us to extend the local solution ψ' by right-multiplication (in an obvious way) to a solution $\varphi\colon \pi^{-1}(V) \to F(\overline{M})$, where $V = \pi(U) \subset M$. Thus there exists a unique differentiable map $j\colon V \to \overline{M}$ such that $j \circ \pi = \overline{\pi} \circ \psi$ on $\pi^{-1}(V)$. We claim that j is an immersion: In fact, suppose $f \in F(M)$ projects to $m \in V$, and let $\psi(f) = \overline{f} \in F(\overline{M})$. Now if $y \in F(M)_f$ projects to $x \in M_m$ we have

$$\langle x, f_i \rangle = \omega_i(y) = \overline{\omega}_i(d\psi(y)) = \langle d\pi(d\psi(y)), \overline{f}_i \rangle$$
$$= \langle dj(x), f_i \rangle, \text{ and } \langle dj(x), f_{d+1} \rangle = \overline{\omega}_{d+1}(d\psi(x)) = 0 \, .$$

This proves $j\colon V \to \bar{M}$ is an immersion; similarly one checks that its second fundamental form is $\sigma|\pi^{-1}(V)$. But an immersion is controlled by its second fundamental form; explicitly in the case at hand, if j' is another such immersion of V in $\bar{M}$ with $j(m) = j'(m)$ and $dj_m = dj'_m$ for some one $m \in V$, then $j = j'$. This uniqueness property, the simple connectedness of M, and the special character of $\bar{M}$ are the essential points in a proof (which we omit) that out of local immersions as above a global immersion $i\colon M \to \bar{M}$ can be constructed of which σ is the second fundamental form.

2. The Gauss equation. Of the conditions (1) imposed on σ, the crucial one is the Gauss equation. Under the usual translation [1] of horizontal equivariant objects on $F(M)$ into objects on M, the curvature form becomes a function which to each $x, y \in M_m$ assigns a linear transformation $R_{xy}\colon M_m \to M_m$. Then then equation $\langle R_{xy}(u), v\rangle = \langle R_m(xy), uv\rangle$ defines *the curvature transformation* R_m as a linear operator on the Grassmann space $\wedge^2 M_m$. The function $m \to R_m$ is for our purposes the most convenient form of the curvature tensor R of M. The form σ translates to a function S on M with S_m a linear operator on M_m, and the Gauss equation becomes $R = S \wedge S + K$, where K denotes scalar multiplication by the constant curvature K of $\bar{M}$.

Reversing the process, suppose that S is a differentiable field of linear operators on the tangent spaces of M such that $R = S \wedge S + K$. Let σ be the horizontal, equivariant R^d-valued 1-form on $F(M)$ corresponding to S. Then $\Phi_{ij} = \sigma_i\sigma_j + K\omega_i\omega_j$. The other two conditions on σ follow automatically if the rank of $R - K$, that is, the minimum rank of $R_m - K$ for $m \in M$, is not too small. Explicitly:

LEMMA 1. (notation as above) *Let* $R = S \wedge S + K$. *If rank* $(R - K) \geq 3$, *then* $\Sigma_k \sigma_k \omega_k = 0$. *If rank* $(R - K) \geq 4$, *then* $D\sigma_i = 0$.

Proof. By a symmetry of R, shared by K, we have $\mathfrak{S}\langle S(x)S(u), yv\rangle = 0$, where $\mathfrak{S}$ denotes the sum over the cyclic permutations of x, u, y. Eliminating v we get $\mathfrak{S}\{(\langle S(y), x\rangle - \langle y, S(x)\rangle)S(u)\} = 0$. But since rank $S \wedge S \geq 3$, the same is true for S, and it follows that $\langle S(y), x\rangle = \langle y, S(x)\rangle$. But the symmetry of S is equivalent to $\Sigma_k \sigma_k \omega_k = 0$.

To prove the second assertion (due essentially to T. Y. Thomas), we apply D to the equation $\Phi_{ij} = \sigma_i\sigma_j + K\omega_i\omega_j$. Since $D\omega = 0$ and $D\Phi = 0$ (Bianchi identity) we get $D\sigma_i \wedge \sigma_j = \sigma_i \wedge D\sigma_j$. The rank condition implies rank $S \geq 4$, hence rank $\sigma \geq 4$. Thus the result is a consequence of the following.

LEMMA 2. *Let* $x_1, \cdots, x_d \in V$, *a finite-dimensional real vector space, and let* $w_1, \cdots, w_d \in \wedge^2 V$. *If* $x_i \wedge w_j = w_i \wedge x_j$ *for all* $1 \leq i, j \leq d$, *and the vectors* $x_1, \cdots, x_a$ *span a subspace of dimension* ≥ 4, *then* $w_1 = \cdots = w_a = 0$.

Proof. We may suppose that x_1, x_2, x_3, x_4 are the first four elements of a basis $e_1, e_2, \cdots$ for V. Let $P = \{1, 2, 3, 4\}$, and fix an index $p \in P$. By a standard Grassmann argument one can show that there is a $y_p \in V$ such that $w_p = y_p e_p$. Then $e_p \wedge w_q = w_p \wedge e_q$ implies $(y_p + y_q)e_p e_q = 0$ for all $q \in P$. Thus $2y_p = (y_p + y_q) + (y_p + y_r) - (y_q + y_r)$ is in the subspace spanned by e_p, e_q, e_r, where q and r are any elements of P such that p, q, r are all different. It follows that y_p is a multiple of e_p, and thus $w_p = 0$. But if $i > 4$, then $e_p \wedge w_i = w_p \wedge e_i = 0$ for all $p \in P$, so that $w_i = 0$ also.

Summarizing, if M *and* $\bar{M}$ *are as in Theorem* 1 *and rank* $(R - K) \geq 4$, *then* M *can be immersed in* $\bar{M}$ *if and only if* $R - K$ *is decomposable,* i.e. *expressible as* $S \wedge S$ *with* S *a differentiable field of linear operators on the tangent spaces of* M.

In the following section we consider the purely Grassmannian question of the decomposability of $R_m - K$ at a single point of M.

3. Decomposability. Let V and W be finite-dimensional real vector spaces, and let $T\colon \wedge^2 V \to \wedge^2 W$ be a linear transformation. To determine whether T is decomposable we use the following definition: Three bivectors are *crossed* if any two, but not all three, are collinear, (a set of bivectors being called collinear if all have a common non-zero divisor, i.e. all are decomposable and the planes of the non-zero ones have a line in common.) One easily proves:

LEMMA 3. *Bivectors* w_1, w_2, w_3 *are crossed if and only if there exist linearly independent vectors* x, y, z *and non-zero numbers* K, L, M *such that*

$$\begin{cases} w_1 = K\, xy \\ w_2 = L\, xz \\ w_3 = M\, yz\,. \end{cases} \tag{4}$$

If w_1, w_2, w_3 are crossed, then in any expression (4) the sign of the product KLM is always the same. (In fact, the vectors x, y, z are unique up to non-zero scalar multiplication, so we need only check that changing the signs of any subset of $\{x, y, z\}$ does not change the sign of KLM.) In case $KLM > 0$ we say that w_1, w_2, w_3 are *coherently crossed.* Note that if T is decomposable then T carries coherently crossed bivectors to bivectors which are either coherently crossed or coplanar. Our aim is to prove the converse when rank $T \geq 4$. (We do not need the easy cases of lower rank.)

LEMMA 4. *The following conditions on T are equivalent*:

(a) *T carries decomposable bivectors to decomposable bivectors.*

(b) *T carries two collinear bivectors to two collinear bivectors.*

(c) *$T(xy) \wedge T(uv) \in \wedge^2 W$ is skew-symmetric in its arguments.*

LEMMA 5. *If rank $T \geq 4$ and T carries crossed to crossed or coplanar bivectors, then R carries collinear to collinear bivectors.*

Proof. It is sufficient to prove collinearity is preserved in the case of three bivectors. Thus we must show that $T(e_1e_2)$, $T(e_1e_3)$, $T(e_1e_4)$ are collinear. Now any two of these bivectors are collinear, hence all three are either crossed or collinear. We assume the former and get a contradiction. If they are crossed there is a unique subspace U of W, with dimension 3, such that the bivectors are in $\wedge^2 U \subset \wedge^2 W$. We may also assume that e_1, e_2, e_3, e_4 are linearly independent for otherwise we can reduce to the case of two collinear bivectors. Thus these vectors are part of a basis for V.

Case I. There is an index i such that $T(e_1e_i) \notin \wedge^2 U$.
Consider $T(e_1e_2)$, $T(e_1e_3)$, $T(e_1(e_4 + \delta e_i))$, where δ is an arbitrarily small non-zero number. Now the last of these three bivectors is not in $\Lambda^2 U$, while the union of the planes of the first two spans U. Hence all three are not in the second Grassmann product of any 3-dimensional subspace of W. Thus they are not crossed. On the other hand, any two are collinear, so all three are collinear. But this is a contradiction, for an arbitrarily samll change in the crossed bivectors $T(e_1e_2)$, $T(e_1e_3)$, $T(e_1e_4)$ cannot produce collinear bivectors.

Case II. For all i, $T(e_1e_i) \in \Lambda^2 U$.
We prove the contradiction rank $T \leq 3$ by showing that $T(e_pe_q) \in \Lambda^2 U$ for all p, q. If $T(e_1e_p)$ and $T(e_1e_q)$ are independent, then by hypothesis, $T(e_pe_q)$ is crossed with these two bivectors, hence is in $\Lambda^2 U$. If they are dependent and $T(e_1e_p) \neq 0$, then by hypothesis $T(e_1e_p)$ and $T(e_pe_q)$ are coplanar and $T(e_pe_q) \in \Lambda^2 U$. Finally, if $T(e_1e_p) = 0$, then by Lemma 5 $0 = T(e_1e_p) \wedge T(e_re_q) = T(e_1e_r) \wedge T(e_pe_q)$ for $r = 2, 3, 4$. But since $T(e_1e_2)$, $T(e_1e_3)$, $T(e_1e_4)$ are crossed one easily deduces from these equations that $T(e_pe_q) \in \Lambda^2 U$.

THEOREM 2. *Let $T\colon \wedge^2 V \to \wedge^2 W$ be a linear transformation of rank ≥ 4. Then there exists a linear transformation $S\colon V \to W$ such that $T = S \wedge S$ if and only if T carries coherently crossed to coherently* *crossed or coplanar bivectors.*

Proof. We may choose a basis $e_1, \cdots, e_d$ for V such that T is never zero on the corresponding canonical basis for $\Lambda^2 V$. Fix an index $1 \leq i \leq d$. By the preceding lemma there is a non-zero vector $u_i \in W$ such that u_i divides each $T(e_i e_j)$, $j = 1, \cdots, d$. Furthermore this vector is unique up to scalar multiplication. To see this we need only show that these bivectors $T(e_i e_j)$ are not all coplanar. But if they were, then $T(e_i e_j)$, $T(e_i e_k)$, $T(e_j e_k)$, since not crossed, would have to be coplanar for all j, k, implying rank $T \leq 1$.

Now let i, j be different indices. We claim that $T(e_i e_j) = K_{ij} u_i u_j$, In fact, since there is an index k such that the bivectors $T(e_i e_j)$ and $T(e_i e_k)$ are not coplanar, they are crossed with $T(e_j e_k)$. By Lemma 3 and the divisibility properties of u_i, u_j, u_k, it follows that these crossed bivectors may be written as $K u_i u_j$, $L u_i u_k$, $M u_j u_k$ respectively.

By changing the signs of $u_2, \cdots, u_d$ where necessary, we shall now arrange to have the number K_{ij} $(i < j)$ all positive. We can certainly get all $K_{ij} > 0$ in this way. Consider $T(e_1 e_i)$, $T(e_1 e_j)$, $T(e_i e_j)$. If the first two bivectors are not coplanar, then all three are coherently crossed, hence the product $K_{1i} K_{1j} K_{ij}$, and consequently K_{ij}, are positive. If $T(e_1 e_i)$ and $T(e_1 e_j)$ are coplanar, we argue as follows: Since rank $T > 1$ there is an index k (say $k > j$) such that u_k is not in the plane spanned by u_1, u_i, u_j. Thus $T(e_1 e_i)$ and $T(e_1 e_k)$ are not coplanar, so $K_{ik} > 0$. Similarly $K_{jk} > 0$. And since u_i, u_j, u_k are independent, it follows that $K_{ij} > 0$.

To complete the proof it will suffice to find numbers $\lambda_1, \cdots, \lambda_d$ such that for any $i < j$ we have $K_{ij} = \lambda_i \lambda_j$. For then the equation $T(e_i e_j) = K_{ij} u_i u_j$ becomes $T(e_i e_j) = (\lambda_i u_i)(\lambda_j u_j)$, and by definding $S \colon V \to W$ to be the linear transformation such that $S(e_i) = \lambda_i u_i$ we get $T = S \wedge S$.

Call a set i, j, k of indices a *triple* if $i < j < k$ and u_i, u_j, u_k are independent. For each triple consider the equations $K_{ij} = \lambda_i \lambda_j$, $K_{ij} = \lambda_i \lambda_k$, $K_{jk} = \lambda_j \lambda_k$. Since the K's are positive there is a unique positive solution $\lambda_i, \lambda_j, \lambda_k$. Since each index i is in at least one triple we get at least one such value for λ_i. We must show that the values obtained from two different triples containing i are the same. We need only consider triples of the form i, j, p and i, j, q, for it will be clear from the proof in this case that the position of i in a triple is immaterial and that the case where five indices are involved may be reduced to the present one using rank $T \geq 4$. We know that

$$\begin{aligned} T(e_i e_j) &= \lambda_i \lambda_j u_i u_j & T(e_i e_j) &= \mu_i \mu_j u_i u_j \\ T(e_i e_p) &= \lambda_i \lambda_p u_i u_p & T(e_i e_q) &= \mu_i \mu_q u_i u_q \\ T(e_j e_p) &= \lambda_j \lambda_p u_j u_p & T(e_j e_q) &= \mu_j \mu_q u_j u_q \,. \end{aligned}$$

First consider the case in which the vectors u_i, u_j, u_p, u_q are linearly independent. By Lemma 4, $T(e_ie_p) \wedge T(e_je_q) = -T(e_je_p) \wedge T(e_ie_q)$, but since $u_iu_pu_ju_q \neq 0$ this implies $\lambda_i\mu_j = \mu_i\lambda_j$. But also $\lambda_i\lambda_j = \mu_i\mu_j$, and since the numbers in the last two equations are all positive we get $\lambda_i = \mu_i$. Now suppose $u_i, u_j u_p, u_q$ are dependent, hence span a 3-dimensional subspace. Since rank $T \geq 4$ there must exist an index r (say $r > p, q$) such that u_i, u_j, u_p, u_r and $u_i\, u_j\, u_q, u_r$ are each linearly independent. Thus the values of λ_i determined by i, j, p and i, j, p are the same as that determined by i, j, r.

This shows the existence of S such that $T = S \wedge S$; uniqueness up to sign is implicit in the proof, for the only ultimate element of choice is in the orientation of u_1, i.e. the use of u_1 rather than $-u_1$.

4. Coordinate criteria for decomposability. With notation as in the preceding section, fix bases $e_1, \cdots, e_d$ for V and $f_1, \cdots, f_{\bar{d}}$ for W. Let $T_{ij} = T(e_ie_j) = \Sigma_{\alpha<\beta} T_{ij\alpha\beta} f_\alpha f_\beta$. What conditions on T_{ij} are necessary and sufficient for T to be decomposable, or alternatively (if rank $T \geq 4$) for T to carry coherently crossed to coherently crossed or coplanar bivectors? Necessary is that T carry decomposable to decomposable bivectors, and this is easily proved equivalent to

$$(5) \qquad T_{ij} \wedge T_{kl} = T_{kj} \wedge T_{il} \text{ for all } 1 \leq i, j, k, l \leq d$$

This condition as well as the condition rank $T \geq 4$ are standardly expressible in terms of polynomials in $T_{ij\alpha\beta}$.

LEMMA 6. *Suppose that any two of the bivectors $a, b, c \in \wedge^2 W$ are collinear, and let $a = \Sigma_{\alpha<\beta} A_{\alpha\beta} f_\alpha f_\beta$, similar expressions for b, c. Then a, b, c are coherently crossed if and only if there exist indices $1 \leq \alpha < \beta < \gamma \leq \bar{d}$ such that*

$$\Delta(\alpha\beta\gamma) = \begin{vmatrix} A_{\alpha\beta} & A_{\alpha\gamma} & A_{\beta\gamma} \\ B_{\alpha\beta} & B_{\alpha\gamma} & B_{\beta\gamma} \\ C_{\alpha\beta} & C_{\alpha\gamma} & C_{\beta\gamma} \end{vmatrix} > 0$$

Proof. The bivectors a, b, c are either crossed or collinear. We show:

(1) if crossed, then for some α, β, γ we have $\Delta(\alpha\beta\gamma) \neq 0$,

(2) if coherently crossed, then each such non-zero determinant is positive,

(3) if collinear, then each such determinant is zero.

For α, β, γ let $x \to \bar{x}$ be the natural projection of W onto the subspace U spanned by $f_\alpha, f_\beta, f_\gamma$; same notation for the induced projection of $\wedge^2 W$ onto $\wedge^2 U$. In the first two cases above we can write a, b, c in the form of (4), hence $\bar{a} = K\bar{x}\bar{y}$, $\bar{b} = L\bar{x}\bar{z}$, $\bar{c} = M\bar{y}\bar{z}$. For (1), since x, y, z independent there are indices α, β, γ such that $\bar{x}, \bar{y}, \bar{z}$ are independent, hence $\bar{a}, \bar{b}, \bar{c}$ are independent, and the result follows. For (2), suppose $\varDelta(\alpha\beta\gamma) \neq 0$. Using the above notation we have $KLM > 0$. Notice that any two canonical bases (lexicographic order) for $\wedge^2 U$ have the same orientation. Thus $\varDelta(\alpha\beta\gamma) > 0$. The proof of (3) is similar.

A further necessary condition for decomposability of T is that T_{ij}, T_{ik}, T_{jk} be coherently crossed or coplanar. Assuming (1), this is equivalent to

(6) If $1 \leq i < j < k \leq d$, then either T_{ij}, T_{ik}, T_{jk} are coplanar or there exist indices $1 \leq \alpha < \beta < \gamma \leq \bar{d}$ such that

$$\begin{vmatrix} T_{ij\alpha\beta} & T_{ij\alpha\gamma} & T_{ij\beta\gamma} \\ T_{ik\alpha\beta} & T_{ik\alpha\gamma} & T_{ik\beta\gamma} \\ T_{ik\alpha\beta} & T_{jk\alpha\gamma} & T_{jk\beta\gamma} \end{vmatrix} > 0.$$

If the basis $e_1, \cdots, e_d$ is such that all $T_{ij} \neq 0$, then (5) and (6) are necessary and sufficient for the decomposability of T, for Lemma 5 and Theorem 2 use no more than this. For an arbitrary basis, however, they are not enough, as one can see from simple examples. We must add, say

(7) If $T_{ij} = T_{ik} = 0$, then either $T_{jk} = 0$, or, for all r, $T_{ir} = 0$.

Now one can prove the following lemma by reducing to the case in which all $T_{ij} \neq 0$.

LEMMA 7. *Let $T\colon \wedge^2 V \to \wedge^2 W$ be a linear transformation with rank $T \geq 4$. Then T is decomposable if and only if, relative to arbitrary canonical bases for $\wedge^2 V$ and $\wedge^2 W$, conditions* (5), (6), (7) *hold.*

5. Summary. Again let R be the curvature transformation of the simply connected manifold M. For simplicity we discuss the case $\bar{M} = R^{d+1}$. Assume that (at each point) rank $R \geq 4$ and R carries coherently crossed to coherently crossed or coplanar bivectors. It is clear that the proof of Theorem 2 applies simultaneously to all R_n with n any point of a convex neighborhood C of $m \in M$. One need only use the simple connectedness of C to choose the orientations of the various choices of u_1 consistently. We thus obtain a *differentiable* field of linear operators S such that $R = S \wedge S$, first locally, then as usual, globally. When rank $R \geq 3$ we can still prove R decomposable, but the Codazzi equation may fail; thus our criterion for immersion, while always sufficient, is necessary only in the case of immersions for which the second fundamental form S has rank at least four. Call such an immersion 4-regular.

This same argument, with $R - K$ in place of K, proves

THEOREM 3. *Let M be a simply connected d-dimensional manifold* $(d \geq 4)$ *with curvature transformation R. Let $\bar{M}(K)$ be a complete* $(d+1)$-*dimensional manifold of constant curvature K. Then M has a* 4-*regular immersion in $\bar{M}(K)$ if and only if rank* $(R - K) \geq 4$ *and* $R - K$ *carries coherently crossed to coherently crossed or coplanar bivectors*, i.e. *conditions* (5), (6), (7) *hold at eaeh point of M.*

For a given M one may ask for the set $\mathscr{K}$ of numbers K such that M has a regular immersion in an $\bar{M}(K)$. Consider two cases:

(i) If R does not preserve decomposability, say $R(xy)^2 \neq 0$, then M is not immersible in R^{d+1} and $\mathscr{K}$ contains at most the number K determined by the necessary condition $R(xy) = S(x)S(y) + Kxy$. We check as above whether $K \in \mathscr{K}$.

(ii) If R preserves decomposability, so that (5) holds, $\mathscr{K}$ may well be infinite. By studying conditions (6), (7) one can characterize $\mathscr{K}$ in terms of polynomials in an unknown K and the coordinates of R.

REFERENCES

1. W. Ambrose and I.M. Singer, *On homogeneous Riemannian manifolds*, Duke Math. J., **25** (1958), 647-669.

2. T. Y. Thomas, *Riemannian spaces of class one and their charactetization*, Acta Math., **67** (1936), 169-211.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY,
UNIVERSITY OF CALIFORNIA, LOS ANGELES

VIBRATION OF A NONHOMOGENEOUS MEMBRANE

M. H. PROTTER

1. Introduction. We consider a simply connected two dimensional domain D with a nonhomogeneous membrane M stretched across D and fixed at the boundary Γ. Let $p(x, y) \geq 0$ be the density function of the membrane. We shall be concerned with the first eigenvalue λ_0 of the equation

$$(1) \qquad u_{xx} + u_{yy} + \lambda p(x, y)u = 0$$

subject to the condition $u = 0$ on Γ. Let K be the circle with boundary C on which a homogeneous membrane M_1 of the same mass as M is stretched. Let λ_1 be the first eigenvalue of

$$(2) \qquad v_{xx} + v_{yy} + \lambda v = 0$$

with $v = 0$ on C. In a recent paper Nehari [1] established the following interesting result.

THEOREM. (Nehari) *If* $\log p(x,y)$ *is subharmonic then*

$$(3) \qquad \lambda_0 \geq \lambda_1 .$$

Nehari further showed that relaxation to the condition that $p(x, y)$ be subharmonic is not possible. In fact for the case that D is a circle and $p(x, y)$ is superharmonic the inequality in (3) is shown to be reversed.

It is the purpose of this paper to establish comparison theorems for the first eigenvalue of homogeneous and nonhomogeneous membranes *of the same shape*. That is, we shall consider the first eigenvalue of equations (1) and (2) in the same domain D subject to the boundary condition $u = 0$ and $v = 0$ on Γ respectively. We denote the first eigenvalue of the latter problem by μ and consider comparisons between λ_0 and μ. We of course have the completely trivial comparison

$$\lambda_0 \geq \mu$$

if $0 \leq p(x, y) \leq 1$ throughout D. Nehari's result pertained to the case where $p(x, y)$ had average value 1 and thus we wish to obtain relations between λ_0 and μ for density functions which may become large.

A general technique for obtaining lower bounds for the first eigenvalue for a homogeneous membrane in a domain D follows from the

Received March 12, 1959. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command under Contract No. AF 49 (638)-398.

inclusion principle. If D is contained in D_0 then the first eigenvalue for D is larger than that for D_0. If D is bounded then we can enclose D in a rectangle or circle for which the first eigenvalue is known. This technique is also possible for nonhomogeneous membranes as will be readily seen from the basic inequalities established in § 2. In § 3 comparison theorems are established when the density function is assumed to satisfy various conditions involving the behavior of the second derivative of $p(x, y)$. Section 4 discusses comparison theorems between two nonhomogeneous membranes.

2. Basic inequalities. Let u be any function which vanishes on Γ, and let $a(x, y)$ be an arbitrary C^2 function in D. We apply Green's theorem to the expression

$$\iint_D au(u_{xx} + u_{yy})dxdy$$

and obtain

$$\iint_D au(u_{xx} + u_{yy})dxdy = -\iint_D a(u_x^2 + u_y^2)dxdy + \frac{1}{2}\iint_D u^2(a_{xx} + a_{yy})dxdy \tag{4}$$

The boundary integrals vanishing in virtue of $u = 0$ on Γ. Further we let $P(x, y)$, $Q(x, y)$ be arbitrary C' functions in D and note that

$$\iint [Pu^2)_x + (Qu^2)_y]dxdy = 0 . \tag{5}$$

Performing the differentiations in (5) and adding the result to (4) we get

$$\begin{aligned} -\iint_D au(u_{xx} + u_{yy})dxdy \\ = + \iint_D \Big\{ a(u_x^2 + u_y^2) + 2Puu_x + 2Quu_y \\ + \left[P_x + Q_y - \frac{1}{2}(a_{xx} + a_{yy})\right]u^2\Big\}dxdy . \end{aligned}$$

If u were the first eigenfunction and λ the first eigenvalue of the nonhomogeneous membrane, then (1) would hold and the above expression would be

$$\begin{aligned} \iint_D \Big\{a(u_x^2 + u_y^2) + 2Puu_x + 2Quu_y \\ + \left[P_x + Q_y - \frac{1}{2}(a_{xx} + a_{yy}) - a\lambda p\right]u^2\Big\}dxdy = 0 \end{aligned} \tag{6}$$

On the other hand this integrand is a quadratic form in u_x, u_y, u. It will be a positive definite form if $a > 0$ and

$$P_x + Q_y \geq \frac{1}{a}(P^2 + Q^2) + \frac{1}{2}(a_{xx} + a_{yy}) + ap\lambda \;. \tag{7}$$

If a, P, Q, λ happen to satisfy (7) then clearly it is impossible that (6) holds. However if (7) holds for any value $\bar{\lambda}$, it obviously holds for $0 \leq \lambda \leq \bar{\lambda}$ and thus (6) cannnot hold for any function $u(x, y)$ with $0 \leq \lambda \leq \bar{\lambda}$. This implies that $\bar{\lambda}$ is a lower bound for the first eigenvalue of (1).

We shall therefore be concerned with the possibility of selection of functions P, Q, a such that inequality (7) holds for some value $\bar{\lambda}$. For convenience we assume the bounded domain D is in the first quadrant. We select the function $a(x, y)$ to be

$$a(x, y) = \sin \alpha x \sin \beta x$$

where α and β are constants selected so that $a(x, y)$ is positive throughout $\bar{D}$. We define the quantities

$$m_0 = \min_{\bar{D}} a$$

and $M_0 = m_0^{-1}$. Inequality (7) is implied by the inequality

$$P_x + Q_y \geq M_0(P^2 + Q^2) + \frac{1}{2}(a_{xx} + a_{yy}) + ap\lambda \tag{8}$$

and if we define

$$P_1 = M_0 P,\ Q_1 = M_0 Q$$

(8) is equivalent to

$$P_{1x} + Q_{1y} \geq P_1^2 + Q_1^2 + \frac{1}{2}M_0(a_{xx} + a_{yy}) + M_0 ap\lambda \;. \tag{9}$$

Let $\phi(x, y)$ be the first eigenfunction for equation (2) in the domain D subject to the condition $v = 0$ on Γ. That is,

$$\phi_{xx} + \phi_{yy} + \mu\phi = 0 \;.$$

We make the following selection:

$$P_1 = -\frac{\phi_x}{\phi}, \quad Q_1 = -\frac{\phi_y}{\phi}$$

and obtain from (9)

$$\mu \geq M_0 \sin \alpha x \sin \beta y \left[-\frac{1}{2}(\alpha^2 + \beta^2) + \lambda p(x, y) \right] \tag{10}$$

Define the quantity

$$N_0 = \max_{\bar{D}} p(x, y) \sin \alpha x \sin \beta y$$

and we obtain the following result.

THEOREM 1. *Let λ_0 be the first eigenvalue for the nonhomogeneous membrane with density function $p(x, y)$ spanning a domain D and μ the first eigenvalue for the homogeneous membrane spanning the same domain. Then*

$$\lambda_0 \geq \frac{\mu + \frac{1}{2}(\alpha^2 + \beta^2)}{M_0 N_0} . \tag{11}$$

The theorem is an immediate consequence of inequality (10) which exhibits the positive definiteness of the integrand (6). Inequality (11) is a statement that (10) must be violated.

We note that (11) is a useful relation if N_0 is particularly small; hence this states that $p(x, y)$ should be small near the center of the membrane, but may be large near the outer edge and still (11) will be a significant lower bound for λ_0. The basic distinction between (11) and other results lies in the fact that $p(x, y)$ has no restriction except positivity.

A word should be said about the selection of the function $a(x, y)$. We chose for this function the first eigenfunction for the equation (2) applied to a rectangle which contains D in its interior. We could have selected for $a(x, y)$ the first eigenfunction for any including domain, e.g., a circle, equilateral triangle, etc. with a resulting inequality similar to (11). Finally the selection $a \equiv 1$ yields the standard result

$$\lambda_0 \geq \frac{\mu}{\max_{\bar{D}} p(x, y)} .$$

3. Bounds with condition on the density function. We return to inequality (7) and the selection of a, P, and Q. We recall that these functions may be arbitrary except that $a(x, y)$ must be positive. We make the choice

$$a(x, y) = \frac{1}{p(x, y)} \tag{12}$$

Then (7) becomes

$$P_x + Q_y \geq p(x, y)(P^2 + Q^2) + \frac{1}{2}\Delta\left(\frac{1}{p}\right) + \lambda\ .$$

We define

$$p_0 = \max_{\overline{D}} p(x, y) \tag{13}$$

and select

$$P = -\frac{\phi_x}{p_0\phi}, \quad Q = -\frac{\phi_y}{p_0\phi}$$

where, as before, ϕ is the first eigenfunction of (2) for the domain D. We obtain

$$\frac{\mu}{p_0} \geq \frac{1}{2}\Delta\left(\frac{1}{p}\right) + \lambda\ .$$

If we assume the function $1/p$ is superharmonic and set

$$N_1 = -\max_{\overline{D}} \frac{1}{2}\Delta\left(\frac{1}{p}\right) \tag{14}$$

we obtain the following result.

THEOREM 2. *Let* λ_0 *be the first eigenvalue for the nonhomogeneous membrane with density function* $p(x, y)$ *and* μ *the corresponding first eigenvalue for the homogeneous membrane spanning the same domain* D. *If* $1/p$ *is superharmonic in* D *we have the inequality*

$$\lambda_0 \geq \frac{\mu}{p_0} + N_1 \tag{15}$$

where p_0 *and* N_1 *are given by* (13) *and* (14) *respectively.*

It is possible to obtain a comparison theorem for the case where $\log p$ is subharmonic. To see this we make the choice

$$a(x, y) = \log\frac{1}{p}$$

and we assume $0 < p(x, y) < 1$ in $\overline{D}$. With this selection we take

$$P = -\frac{\phi_x}{p_0\phi}, \quad Q = -\frac{\phi_y}{p_0\phi}$$

as before and obtain

$$\frac{\mu}{p_0} \geq \frac{1}{2}\Delta\left(\log \frac{1}{p}\right) + \lambda p \log \frac{1}{p} .$$

We assume $\log p$ is subharmonic and define

$$N_2 = \frac{1}{2} \min_{\bar{D}} \ \Delta(\log p) \tag{16}$$

$$N_3 = \max_{\bar{D}} p \log \frac{1}{p} . \tag{17}$$

THEOREM 3. *Let* λ_0 *and* μ *be as in Theorem* 2. *If* $\log p$ *is subharmonic in* D *then*

$$\lambda_0 \geq \frac{\mu}{p_0 N_3} + \frac{N_2}{N_3}$$

where N_2 *and* N_3 *are given by* (16) *and* (17).

A final application of this type which we exhibit results from the selection

$$a = e^{\alpha p(x,y)}$$

where α is a constant which remains to be chosen. If we suppose that p is strictly superharmonic and select α so that

$$\frac{1}{2}\Delta p + \alpha(p_x^2 + p_y^2) \leq 0$$

we obtain the relation

$$\lambda_0 \geq \mu \max_{\bar{D}} \left(\frac{e^{-\alpha p}}{p}\right) .$$

4. Comparison of two nonhomogeneous membranes. Let $q(x, y)$ be a second density function corresponding to a membrane spanning D and let ν be the first eigenvalue for

$$w_{xx} + w_{yy} + \nu q(x, y) w = 0 \tag{18}$$

with boundary condition $w = 0$ on Γ. We denote the corresponding first eigenfunction by $\psi(x, y)$. It is possible to compare λ_0 and ν when the functions p and q satisfy various relations. Let

$$q_0 = \max_{\bar{D}} q(x, y) \tag{19}$$

$$r_0 = \max_{\bar{D}} \frac{p(x, y)}{q(x, y)} \tag{20}$$

and

$$N_4 = - \max_{\bar{D}} \Delta\left(\frac{q}{p}\right). \tag{21}$$

We make the selections

$$a = \frac{q}{p}, \quad P = -\frac{\psi_x}{r_0\psi}, \quad Q = -\frac{\psi_y}{r_0\psi}$$

and find

$$\frac{\nu q}{r_0} \geq \frac{1}{2}\Delta\left(\frac{q}{p}\right) + q\lambda .$$

THEOREM 4. *Let* λ_0 *and* ν *be the first eigenvalue corresponding to density functions* p *and* q *respectively. If* q/p *is superharmonic then we have the inequality*

$$\lambda_0 \geq \frac{\nu}{r_0} + \frac{1}{2}\frac{N_4}{q_0}$$

where q_0, r_0 *and* N_4 *are given by* (19), (20) *and* (21).

Additional inequalities, analogous to those obtained in §§ 2 and 3 may be obtained by other selections for a, P and Q.

BIBLIOGRAPHY

1. Z. Nehari, *On the principal frequency of a membrane*, Pacific J. Math. **8** (1958), 285-293.

UNIVERSITY OF CALIFORNIA
BERKELEY

INTRINSIC OPERATORS IN THREE-SPACE

VICTOR L. SHAPIRO

1. Introduction. In Euclidean three-space there are three important classical intrinsic operators, namely the intrinsic curl, the intrinsic divergence, and the intrinsic (or generalized) Laplacian. Usually they are given in terms of differential operators, but the occasion arises sometimes when they cannot be so defined. In particular if u is the Newtonian potential due to a continuous distribution, then in general u is only a function in class C^1, and consequently the usual Laplacian of u, the usual curl of grad u, and the usual divergence of grad u cannot be defined. Nevertheless, as it is easy to show, the intrinsic curl of grad u is equal to zero, the intrinsic (or generalized) Laplacian of u equals the intrinsic divergence of grad u, and furthermore Poisson's equation holds. The question arises whether the converse is true. The answer to questions of this nature is the subject matter of this paper. In particular we shall establish the following result (with the precise definitions given in the next section):

THEOREM 1. *Let D be a domain in Euclidean three-space and let v be a continuous vector field defined in D. Then a necessary and sufficient condition that v be locally in D the gradient of a potential of a distribution with continuous density is that the intrinsic curl of v be zero in D and the intrinsic divergence of v be continuous in D.*

2. Definitions and notation. We shall use the following vectorial notation: $x = (x_1, x_2, x_3)$, $\alpha x + \beta y = (\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2, \alpha x_3 + \beta y_3)$, $(x, y) =$ the usual scalar product, $x \times y =$ the usual cross product, and $|x| = (x, x)^{1/2}$.

Let $v(x) = [v_1(x), v_2(x), v_3(x)]$ be a continuous vector field defined in the neighborhood of the point x_0. Then we define the upper intrinsic curl of v at x_0 to be the vector, curl $^*v(x_0) = [w_1^*(x_0), w_2^*(x_0), w_3^*(x_0)]$ where $w_j^*(x_0) = \limsup_{r\to 0}(\pi r^2)^{-1}\int_{C_j(x_0,r)}(v, dx)$, $j = 1, 2, 3$, with $C_j(x_0, r)$ the circumference of the circle of radius r and center x_0 in the plane through x_0 normal to the x_j-axis where $C_j(x_0, r)$ is oriented in the counterclockwise direction when seen from the side in which the x_j-axis points. In a similar manner using lim inf, we define the lower intrinsic curl of v at x_0, $\operatorname{curl}_* v(x_0)$. If $\operatorname{curl}^* v(x_0) = \operatorname{curl}_* v(x_0)$ is finite, we call this

Received February 9, 1959. 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. AF49 (638)-253. Reproduction in whole or in part is permitted for any purpose of the United States Government.

common value the intrinsic curl of v at x_0 and designate it by curl $v(x_0)$. This definition is essentially the intrinsic definition of the curl as given in [4, p. 71].

Next, we define the intrinsic divergence. Let $v(x)$ be a continuous vector field defined in a neighborhood of the point x_0. Then with $S(x_0, r)$ the spherical surface with center x_0 and radius r, we define the upper intrinsic divergence of v at x_0 as follows

$$\operatorname{div}^* v(x_0) = \limsup_{r\to 0} 3(4\pi r^3)^{-1}\int_{S(x_0,r)}(v, n)dS$$

where n is the outward pointing unit normal on $S(x_0, r)$ and dS is the natural surface area element on $S(x_0, r)$. Similarly we define the lower intrinsic divergence, $\operatorname{div}_* v(x_0)$, using lim inf. If $\operatorname{div}_* v(x_0) = \operatorname{div}^* v(x_0)$ is finite, we call this common value the intrinsic divergence of v at x_0 and designate it by $\operatorname{div} v(x_0)$ (see [9]).

If $u(x)$ is a continuous function defined in a neighborhood of the point x_0, then the upper intrinsic (or generalized) Laplacian of u at the point x_0, $\operatorname{Lap} u(x_0)$, is usually defined as

$$\operatorname{Lap}^* u(x_0) = \limsup_{r\to 0}\left[(4\pi r^2)^{-1}\int_{S(x_0,r)} u dS - u(x_0)\right]6r^{-2} .$$

Similarly we define $\operatorname{Lap}_* u(x_0)$ using lin inf. If $\operatorname{Lap}^* u(x_0) = \operatorname{Lap}_* u(x_0)$ is finite, we call this common value the intrinsic (or generalized) Laplacian of u at x_0 and designate it by $\operatorname{Lap} u(x_0)$.

It is clear that if $v(x)$ is in class C^1 and $u(x)$ is in class C^2, then $\operatorname{curl} v(x)$, $\operatorname{div} v(x)$, and $\operatorname{Lap} u(x)$ exist and equal the usual curl, divergence, and Laplacian respectively, defined in terms of the partial derivatives.

If $f(x)$ is a function defined in a neighborhood of the point x_0 and if $f(x)$ is in L^1 in $S_1(x_0, r)$ for some $r > 0$ where $S_1(x_0, r)$ is the open solid sphere with center x_0 and radius r, we shall designate by $A^*f(x_0)$ the following upper limit:

$$A^*f(x_0) = \limsup_{r\to 0}(4\pi r^3)^{-1}3\int_{S_1(x_0,r)} f(x)dx .$$

Similarly, we shall designate by $A_*f(x_0)$ the corresponding value obtained by using lim inf. As is well-known, for almost all x in $S_1(x_0, r)$, $A_*f(x) = A^*f(x)$.

Given $v(x)$ a continuous vector field defined in a domain D, we shall say that $v(x)$ is locally in D the gradient of a potential of a distribution with bounded density if for each point x_0 in D there exists an $S_1(x_0, r)$ contained in D and two functions $f(x)$ and $h(x)$ defined in $S_1(x_0, r)$ with $f(x)$ bounded in $S_1(x_0, r)$ and $h(x)$ harmonic in $S_1(x_0, r)$ such that

$$(1) \quad u(x) = -(4\pi)^{-1}\int_{S_1(x_0,r)} f(y)|x-y|^{-1}dy + h(x) \text{ for } x \text{ in } S_1(x_0, r) ,$$

and $v(x) = \operatorname{grad} u(x)$ for x in $S_1(x_0, r)$. It is understood that $f(x)$ is bounded in $S_1(x_0, r)$ but need not be bounded in D.

It is well-known that if $u(x)$ is defined by (1), then $u(x)$ is in class C^1 in $S_1(x_0, t)$, and furthermore $\operatorname{Lap} u(x) = f(x)$ (see [7]) at every point where $A^*f(x) = A_*f(x)$. We shall show that $\operatorname{curl} \operatorname{grad} u(x) = 0, \operatorname{div}^* \operatorname{grad} u(x) = A^*f(x)$, and $\operatorname{div}_* \operatorname{grad} u(x) = A_*f(x)$.

$\bar{E}$ will designate the closure of the set E.

3. Statement of main results.

We shall prove the theorems stated below.

THEOREM 2. *Let D be a bounded domain in Euclidean three-space, and let $v(x)$ be a continuous vector field defined in D. Then a necessary and sufficient condition that $v(x)$ be locally in D the gradient of a potential of a distribution with bounded density is that*

(i) $\operatorname{curl}_* v(x)$ *and* $\operatorname{curl}^* v(x)$ *be finite-valued in D.*

(ii) $\operatorname{curl}_* v(x) = \operatorname{curl}^* v(x) = 0$ *almost everywhere in D.*

(iii) $\operatorname{div}_* v(x)$ *and* $\operatorname{div}^* v(x)$ *be locally bounded in D.*

In the next theorem, the definitions of regular curves and regular surfaces are those given in [4, Chapter 4].

THEOREM 3. *Let D be a bounded domain in Euclidean three-space, and let $v(x)$ be a continuous vector field defined in D. Suppose that*

(i) $\operatorname{curl}^* v(x)$ *and* $\operatorname{curl}_* v(x)$ *are finite valued in D.*

(ii) *there exists a continuous vector-field $w(x)$ such that* $w(x) = \operatorname{curl}_* v(x) = \operatorname{curl}^* v(x)$ *almost everywhere in D.*

Then $\operatorname{curl} v(x)$ *exists everywhere in D and is equal to $w(x)$. Furthermore Stokes' theorem with respect to v and* $\operatorname{curl} v$ *holds for every open two-sided regular surface contained in the interior of D, that is*

$$\int_C (v, dx) = \int_S (\operatorname{curl} v, n) dS \tag{2}$$

where C is the regular curve which is the boundary of S oriented in the counter-clockwise sense when seen from the side of S towards which n points.

The sufficiency conditions of Theorems 1 and 2 follow as corollaries of Theorem 5 to be stated in § 5. As a further corollary of Theorem 5, we obtain the following extension of a theorem of Beckenbach's [1, Theorem 1] (i.e. we remove the uniformity conditions stated in his theorem).

THEOREM 4. *Let $v(x)$ be a continuous vector field defined in a bounded domain D of Euclidean three-space. Then a sufficient condition that $v(x)$ be a Newtonian vector field in D is that*

(i) $\operatorname{curl} v(x) = 0$ *in* D

(ii) $\operatorname{div} v(x) = 0$ *in* D.

The curl of a vector field which is only assumed continuous in a domain can be defined in a different manner than that given above, namely by using spherical surfaces and the cross product. We shall consider this definition and the analogues of Theorem 1, 2, 3, and 4 in the concluding section of this paper.

4. Proof of Theorem 3. Since we need the result of Theorem 3 in order to establish Theorems 1, 2, and 4, we shall prove the former theorem first. In order to do this, we need the following lemma:

LEMMA 1. *Let* $v(x) = [v_1(x), v_2(x), v_3(x)]$ *be a continuous vector field defined and continuous in a neighborhood of the point* x_0, *and let* $\lambda(x)$ *be a non-negative function in class* C^1 *in a neighborhood of the point* x_0. *Let* $v'(x) = \lambda(x)v(x)$, *that is* $v'_j(x) = \lambda(x)v_j(x)$, $j = 1, 2, 3$. *Then*

(a) $\operatorname{curl}^* v'(x_0) = \lambda(x_0) \operatorname{curl}^* v(x_0) + \operatorname{grad} \lambda(x_0) \times v(x_0)$

(b) $\operatorname{curl}_* v'(x_0) = \lambda(x_0) \operatorname{curl}_* v(x_0) + \operatorname{grad} \lambda(x_0) \times v(x_0)$

where $\lambda(x_0) \operatorname{curl}^* v(x_0) = \lambda(x_0) \operatorname{curl}_* v(x_0) = 0$ *in case* $\lambda(x_0) = 0$.

To prove the lemma, it is sufficient to prove (a) for (b) will follow on considering $-v(x)$. To prove (a), we have to show with $w^*(x_0) = \operatorname{curl}^* v(x_0)$ that

$$\lambda(x_0)w_k^*(x_0) + v_j(x_0)\lambda_{x_i}(x_0) - v_i(x_0)\lambda_{x_j}(x_0)$$
$$= \limsup_{r\to 0}(\pi r^2)^{-1}\int_{\sigma_k(x_0, r)} \lambda(x)v_i(x)dx_i + \lambda(x)v_j(x)dx_j$$

where (i, j, k) is a cyclic permutation of (1, 2, 3) and $\lambda(x_0)w_k^*(x_0) = 0$ in case $\lambda(x_0) = 0$. But this follows immediately from [9, Lemma 8].

To prove Theorem 3, it is sufficient to establish

$$\int_\sigma (v, dx) = \int_S (w, n)dS \tag{3}$$

for every open two-sided regular surface S contained in the interior of D. For once (3) is established, it holds in the particular case when S is a disc. Consequently the assumed continuity of w in D and (3) implies that

$$\lim_{r\to 0}(\pi r^2)^{-1}\int_{\sigma_j(x_0, r)} (v, dx) = w_j(x_0) \qquad j = 1, 2, 3 .$$

Therefore $\operatorname{curl} v$ exists everywhere in D and is equal to w, and consequently (3) is equivalent (2).

We shall now proceed to establish (3). In order to do this, we first notice that with no loss of generality (since we are going to use Fourier series to prove (3)) we can assume that the closure of our domain D is contained in the interior of the three-dimensional torus $T_3 = \{x, -\pi < x_j \leqq \pi, j = 1, 2, 3\}$. Now let S be a given open two-sided regular surface contained in the interior of D. Since S itself is a closed point set, between S and D we can put two domains D' and D'' with the following property:

$$S \subset D' \subset \overline{D}' \subset D'' \subset \overline{D}'' \subset D \subset \overline{D} \subset T_3 .$$

Letting $\lambda(x)$ be a localizing function which is non-negative and in class C^∞ on T_3 and which takes the value one on D' and the value zero on $T_3 - \overline{D}''$, we set $v'(x) = \lambda(x)v(x)$ and $w'(x) = \lambda(x)w(x) + \operatorname{grad}\lambda(x) \times v(x)$ for x in D and $v'(x) = w'(x) = 0$ for x in $T_3 - D$. Since $v'(x) = v(x)$ and $w'(x) = w(x)$ for x on S, (3) will be established once we can show that

$$\int_C (v', dx) = \int_S (w', n)dS . \tag{4}$$

In order to establish (4), we first observe from Lemma 1 and (i) and (ii) of Theorem 3 that

(5) $\operatorname{curl}^* v'(x)$ and $\operatorname{curl}_* v'(x)$ are finite-valued in T_3

(6) $\operatorname{curl}^* v'(x) = \operatorname{curl}_* v'(x) = w'(x)$ almost everywhere in T_3.

Next we designate the multiple Fourier series of v'_j and w'_j respectively by

$$\begin{aligned} v'_j(x) &\sim \sum_m a^j_m e^{i(m,x)} \\ w'_j(x) &\sim \sum_m b^j_m e^{i(m,x)} \end{aligned} \qquad j = 1, 2, 3 \tag{7}$$

where m represents an integral lattice point in three-space.

The essential step in proving (4) is to show that

$$b^\alpha_m = i(m_\beta a^\gamma_m - m_\gamma a^\beta_m) \tag{8}$$

where (α, β, γ) is a cyclic permutation of $(1, 2, 3)$.

In order to do this we fix x_α and observe that

$$v'_j(x) \sim \sum_{m_\beta} \sum_{m_\gamma} a^j_{m_\beta m_\gamma}(x_\alpha) e^{i(m_\beta x_\beta + m_\gamma x_\gamma)} \text{ for } j = \beta, \gamma \tag{9}$$

where

$$a^j_{m_\beta m_\gamma}(x_\alpha) = (4\pi^2)^{-1} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{-i(m_\beta x_\beta + m_\gamma x_\gamma)} v'_j(x) dx_\beta dx_\gamma . \tag{10}$$

Now by (5),

(11) $$\limsup_{r\to 0}(\pi r^2)^{-1}\int_{C_\alpha(x,r)} v'_\beta(x_\alpha, y_\beta, y_\gamma)dy_\beta + v'_\gamma(x_\alpha, y_\beta, y_\gamma)dy_\gamma$$

is finite-valued in T_3 with a similar statement holding for lim inf, and by (6),

(12) $$\lim_{r\to 0}(\pi r^2)^{-1}\int_{C_\alpha(x,r)} v'_\beta(x_\alpha, y_\beta, y)dy_\beta + v'_\gamma(x_\alpha, y_\beta, y_\gamma)dy_\gamma$$
$$= w'_\alpha(x_\alpha, x_\beta, x_\gamma) \text{ for almost every } (x_\beta, x_\gamma) \text{ if } x_\alpha \text{ lies in } (-\pi, \pi] - E_\alpha \text{ where } E_\alpha \text{ is a set of linear measure zero.}$$

Consequently it follows from (10), (11), (12), a modified version of [9, Lemma 8], and [9, Theorem 2] that for m_β and m_γ any two integers and x_α in $(-\pi, \pi] - E_\alpha$ that

(13) $$i[m_\beta a^\gamma_{m_\beta m_\gamma}(x_\alpha) - m_\gamma a^\beta_{m_\beta m_\gamma}(x_\alpha)] = (4\pi^2)^{-1}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{-i(m_\beta x_\beta + m_\gamma x_\gamma)} w'_\alpha(x_\alpha, x_\beta, x_\gamma)dx_\beta dx_\gamma \ .$$

Letting m_α be any integer, multiplying both sides of (13) by $(2\pi)^{-1}e^{-im_\alpha x_\alpha}$, and then integrating over $(-\pi, \pi]$, we conclude from (10), the fact that E_α is of linear measure zero, and (7) that

$$i(m_\beta a^\gamma_m - m_\gamma a^\beta_m) = b^\beta_m \ ,$$

which is (8).

(4) follows now fairly easily. We introduce for $t > 0$, the vector fields $v'(x, t)$ and $w'(x, t)$ where

(14) $$v'_j(x, t) = \Sigma a^j_m e^{i(m,x)-|m|t}$$
$$w'_j(x, t) = \Sigma b^j_m e^{i(m,x)-|m|t} \qquad j = 1, 2, 3 \ .$$

Then, since $v'(x, t)$ and $w'(x, t)$ are vector fields in class C^∞ on T_3 and since we can differentiate under the summation signs in (14), we conclude from (8) that $\operatorname{curl} v'(x, t) = w'(x, t)$. Consequently,

(15) $$\int_C (v'(x, t), dx) = \int_S (w'(x, t), n)dS \qquad \text{for } t > 0.$$

But as is well-known [2], $v'(x, t) \to v'(x)$ and $w'(x, t) \to w'(x)$ as $t \to 0$ uniformly for x in T_3. Therefore from the definition of a regular curve, it follows that $\int_C (v'(x, t), dx) \to \int_C (v', dx)$, and from the definition of a regular surface, it follows that $\int_S (w'(x, t), n)dS \to \int_S (w', n)dS$. We conclude from (15) that

$$\int_C (v', dx) = \int_S (w', n)dS$$

which is precisely (4), and the proof of Theorem 3 is complete.

5. **Proof of Theorem 1, 2, and 4.** The necessary conditions of Theorems 1 and 2 follow immediately from the following lemma (for an analogous two-dimensional result, see [3]), which we shall prove:

LEMMA 2. *Let* $u(x) = -(4\pi)^{-1}\int_{S_1(x_0,r_0)} f(y)\,|x-y|^{-1}dy$ *where* $f(x)$ *is bounded in* $S_1(x_0, r_0)$ *with* $r_0 > 0$. *Then for* x *in* $S_1(x_0, r_0)$

(a) $\operatorname{curl}\operatorname{grad} u(x) = 0$

(b) $A_* f(x) = \operatorname{div}_* \operatorname{grad} u(x)$ *and* $A^* f(x) = \operatorname{div}^* \operatorname{grad} u(x)$

(c) $\operatorname{div}_* \operatorname{grad} u(x) \leqq \operatorname{Lap}_* u(x) \leqq \operatorname{Lap}^* u(x) \leqq \operatorname{div}^* \operatorname{grad} u(x)$

To prove the lemma, it is clearly sufficient to prove it in the case $x = x_0$, and furthermore with no loss of generality, we can assume x_0 is the origin.

Setting $v(x) = \operatorname{grad} u(x)$, we observe that

$$v_j(x) = (4\pi)^{-1}\int_{S_1(0,r_0)} f(y)(x_j - y_j)\,|x-y|^{-3}dy \qquad j = 1, 2, 3\,, \tag{16}$$

and $v_j(x)$ is a continuous function. Observing that

$$\int_{C_j(0,r)} (\operatorname{grad}|x-y|^{-1}, dx) = 0$$

for y not on $C_j(0, r)$ $j = 1, 2, 3$, we conclude from (16) and Fubini's theorem that $\int_{C_j(0,r)} (v, dx) = 0$ for $j = 1, 2, 3$. Consequently (a) of the lemma is established.

Observing the $-\int_{S(0,r)} (\operatorname{grad}|x-y|^{-1}, n)dS = 4\pi$ if y is in $S_1(0, r)$ and $= 0$ if y is not in $\overline{S}_1(0, r)$, we obtain from (16) and Fubini's theorem that for $0 < r < r_0$.

$$\int_{S(0,r)} (v, n)dS = \int_{S_1(0,r)} f(y)dy\,. \tag{17}$$

Dividing both sides of (17) by $4\pi r^3/3$ and then taking $\liminf_{r\to 0}$ of both sides and next $\limsup_{r\to 0}$, gives us precisely part (b) of the lemma.

(c) follows from (b), the boundedness of f, and [5].

Theorem 4 and the sufficiency conditions of Theorems 1 and 2 follow from the following more general theorem:

THEOREM 5. *Let* D *be a bounded domain in Euclidean three-space, and let* $v(x)$ *be a continuous vector field defined in* D. *Suppose that*

(i) $\operatorname{curl}_* v(x)$ *and* $\operatorname{curl}^* v(x)$ *are finite-valued in* D

(ii) $\operatorname{curl}_* v(x) = \operatorname{curl}^* v(x) = 0$ *almost everywhere in* D

(iii) $\operatorname{div}_* v(x)$ *and* $\operatorname{div}^* v(x)$ *are finite-valued in* D

(iv) *there exists a function* $f(x)$ *such that* $f(x)$ *is in* L^1 *on every closed subdomain of* D *and such that* $\operatorname{div}_* v(x) \geqq f(x)$ *for* x *in* D.

Then (a) $\operatorname{div} v(x)$ *exists almost everywhere in* D

(b) $\operatorname{div} v(x)$ *is in* L^1 *on every closed subdomain of* D

(c) *for every closed sphere* $\overline{S}_1(x_0, r_0)$ *contained in* D, *there exists a function* $u(x)$ *in class* C^1 *in* $S_1(x_0, r_0)$ *such that for* x *in* $S_1(x_0, r_0)$, $v(x) = \operatorname{grad} u(x)$ *and furthermore*
$u(x) = -(4\pi)^{-1}\int_{S_1(x_0,r_0)} \operatorname{div} v(y)\,|x-y|^{-1}dy + h(x)$ a.e. in $S_1(x_0, r_0)$
where $h(x)$ *is harmonic in* $S_1(x_0, r_0)$.

In order to prove Theorem 5, we first need the following lemma (see [8, p. 381]):

LEMMA 3. *Let* $u(x)$ *be in class* C^1 *in* $S_1(x_0, r_0)$. *Then* $\operatorname{div}_* \operatorname{grad} u(x_0) \leqq \operatorname{Lap}_* u(x_0) \leqq \operatorname{Lap}^* u(x_0) \leqq \operatorname{div}^* \operatorname{grad} u(x_0)$

With no loss in generality, we assume that x_0 is the origin. Then by the mean value theorem

$$\left[(4\pi)^{-1}\int_0^{\pi}\int_0^{2\pi} u(t\sin\theta\cos\varphi,\, t\sin\theta\sin\varphi,\, t\operatorname{con}\theta)\sin\theta\, d\theta d\varphi - u(0)\right]/t^2 6^{-1}$$
$$= (4\pi)^{-1}\int_0^{\pi}\int_0^{2\pi} u_t(r\sin\theta\cos\varphi,\, r\sin\theta\sin\varphi,\, r\cos\theta)\,\sin\theta\, d\theta d\varphi / r 3^{-1}$$

where $0 < r < t$. We conclude that

$$\sup_{0<r<t}\left[(4\pi r^2)^{-1}\int_{S(0,r)} u\,dS - u(0)\right]/r^2 6^{-1}$$
$$\leqq \sup_{0<r<t}(4\pi r^3)^{-1} 3\int_{S(0,r)}[\operatorname{grad} u, n]\,dS\,.$$

Consequently from their very definitions, $\operatorname{Lap}^* u(0) \leqq \operatorname{div}^* \operatorname{grad} u(0)$. Similarly we show that $\operatorname{div}_* \operatorname{grad} u(0) \leqq \operatorname{Lap}_* u(0)$, and the proof to the lemma is complete.

It follows immediately from the three-dimensional analogue of [9, Theorem 2] that (a) and (b) of Theorem 5 hold. To obtain (c) of Theorem 5, we observe that there exists a positive ε such $S_1(x_0, r_0+\varepsilon) \subset D$. Let x be in $S_1(x_0, r_0+\varepsilon)$, and let $P(x)$ be the line segment connecting x_0 with x and directed to x. Then we define $u(x) = \int_{P(x)}(v, dy)$, and observe, since by Theorem 3 $\operatorname{curl} v = 0$ everywhere in $S_1(x_0, r_0+\varepsilon)$ and Stokes' theorem with respect to v and $\operatorname{curl} v$ holds in this domain, that $u(x)$ is in class C^1 in $S_1(x_0, r_0+\varepsilon)$ and furthermore that $v(x) = \operatorname{grad} u(x)$. Consequently by Lemma 3 and (iii) of the theorem

(18) $\operatorname{Lap}_* u(x)$ and $\operatorname{Lap}^* u(x)$ are finite-valued in $S_1(x_0, r_0+\varepsilon)$,

and by (a) and (b) of the theorem and Lemma 3

(19) $\operatorname{Lap} u(x) = \operatorname{div} v(x)$ almost everywhere in $S_1(x_0, r_0 + \varepsilon)$.

Therefore by (b) of the theorem, (18), (19), and the three-dimensional analogue of [6, Theorem 1], it follows that for almost all x in $S_1(x_0, r_0)$

$$u(x) = -(4\pi)^{-1}\int_{S_1(x_0,r_0)} \operatorname{div} v(y)\,|x - y|^{-1}dy + h(x)$$

where $h(x)$ is harmonic in $S_1(x_0, r_0)$. But this is precisely (c) of Theorem 5, and the proof to the theorem is complete.

6. The spherical intrinsic curl. Let $v(x)$ be a continuous vector field defined in a neighborhood of the point x_0. Then as mentioned earlier, the upper and lower intrinsic curl of v at x_0 can be defined by means of the cross product and spherical surfaces. In short, we define the upper spherical intrinsic curl to be the component-wise upper limit, $\operatorname{curl}_S^* v(x_0) = \limsup_{r\to 0}(4\pi r^3)^{-1}3\int_{S(x_0,r)} (n \times v)dS$. Similarly we define the lower spherical intrinsic curl, $\operatorname{curl}_{*S} v(x_0)$, using $\liminf_{r\to 0}$. In case $\operatorname{curl}_S^* v(x_0) = \operatorname{curl}_{*S} v(x_0)$ is finite, we say the spherical intrinsic curl of v exists at the point x_0, and we designate this common value by $\operatorname{curl}_S v(x_0)$.

We shall prove the following theorems:

THEOREM 6. *Theorems* 1, 2, 3, 4, *and* 5 *continue to hold if in each of these theorems* $\operatorname{curl}^* v$, $\operatorname{curl}_* v$, *and* $\operatorname{curl} v$ *are replaced by* $\operatorname{curl}_S^* v$, $\operatorname{curl}_{*S} v$, *and* $\operatorname{curl}_S v$ *respectively.*

THEOREM 7. *Let D be a bounded domain in Euclidean three-space, and let $v(x)$ be a continuous vector field defined in D. Then*

(a) *if* $\operatorname{curl}_S v(x)$ *exists and is continuous in D, then* $\operatorname{curl} v(x)$ *exists everywhere in D and equals* $\operatorname{curl}_S v(x)$.

(b) *if* $\operatorname{curl} v(x)$ *exists and is continuous in D, then* $\operatorname{curl}_S v(x)$ *exists everywhere in D and equals* $\operatorname{curl} v(x)$.

To prove Theorem 6, it follows from the proofs of Theorems 1, 2, 4, and 5 that it is sufficient just to prove Theorem 3 and Lemma 2(a) when $\operatorname{curl}^* v$, $\operatorname{curl}_* v$, and $\operatorname{curl} v$ are replaced respectively by $\operatorname{curl}_S^* v$, $\operatorname{curl}_{*S} v$, and $\operatorname{curl}_S v$.

The analogue of Lemma 2(a) follows immediately from Fubini's theorem and the fact that $\int_{S(x_0,r)} n \times \operatorname{grad}|x - y|^{-1}dS = 0$ if y is not on $S(x_0, r)$.

To prove the new version of Theorem 3, we designate by p^j the unit vector in the direction of the x_j-axis and set $v^j = v \times p^j$ for $j = 1, 2, 3$. Then it follows from the definition of spherical intrinsic curl and intrinsic divergence that the jth component of $\operatorname{curl}_S^* v$ is $\operatorname{div}^* v^j$ with a similar remark holding for $\operatorname{curl}_{*S} v$. Consequently by (i) and (ii) of the new version of Theorem 3 and by the three-dimensional analogue of [9, Theorem 2], we obtain that for $\overline{S}_1(x_0, r)$ contained in D,

$$\int_{S(x_0,r)} (v^j, n)dS = \int_{S_1(x_0,r)} w_j(x)dx \qquad j = 1, 2, 3 . \tag{20}$$

But (20) implies that $\operatorname{curl}_S v(x)$ exists everywhere in D and equals $w(x)$, giving the first part of the theorem.

The last part follows in a manner similar to the original version of Theorem 3, and it suffices to give a sketch of the proof. We first establish the analogue of Lemma 1 for the spherical intrinsic curl. Next with $\overline{D}$ contained in the interior of T_3 and S contained in D, we introduce the periodic vector fields $v'(x) = \lambda(x)v(x)$ and $w'(x) = \lambda(x)w(x) + \operatorname{grad}\lambda(x) \times v(x)$ where $\lambda(x)$ is a non-negative localizing function in class C^∞ which takes the value one in a neighborhood of S and the value zero outside another neighborhood of S for points in T_3. Then with $v'(x, t)$ and $w'(x, t)$ as in Theorem 3, it follows using the three dimensional analogues of the results in [9] that $\operatorname{curl} v'(x, t) = w'(x, t)$. But, as before, this implies that $\int_C (v, dx) = \int_S (w, n)dS$, which fact completes the proof of the theorem.

Theorem 7(a) follows immediately from Theorem 6.

To prove Theorem 7(b), we assume that $\overline{D}$ is contained in the interior of T_3, and we set $w(x) = \operatorname{curl} v(x)$. Then with $\overline{S}_1(x_0, 3r_0)$ contained in D and $\lambda(x)$ a non-negative localizing function of class C^∞ which takes the value one in $S_1(x_0, r_0)$ and the value zero in $T_3 - S_1(x_0, 2r_0)$, we introduce, as before, the periodic vector fields $v'(x) = \lambda(x)v(x)$, $w'(x) = \lambda(x)w(x) + \operatorname{grad}\lambda(x) \times v(x)$, $v'(x, t)$, and $w'(x, t)$. Exactly as in Theorem 3, we obtain that $\operatorname{curl} v'(x, t) = w'(x, t)$. But then on setting $v'^j(x) = v'(x) \times p^j$ and $v'^j(x, t) = v'(x, t) \times p^j$ for $j = 1, 2, 3$, we obtain that

$$\int_{S(x_0,r)} (v'^j(x, t), n)dS = \int_{S_1(x_0,r)} w'_j(x, t)dx \text{ for } r > 0 ,$$

and consequently that

$$\int_{S(x_0,r)} (v'^j(x), n)dS = \int_{S_1(x_0,r)} w'_j(x)dx .$$

This last fact, however, implies that $\operatorname{curl}_S v(x_0) = w(x_0)$, and therefore completes the proof to Theorem 7(b).

BIBLIOGRAPHY

1. E. F. Beckenbach, *Vector analogues of Morera's theorem*, Bull. Amer. Math. Soc., **48** (1942), 937-941.

2. S. Bochner, *Summation of multiple Fourier series by spherical means*, Trans. Amer.

Math. Soc., **40** (1936), 175-207.
3. G. C. Evans, *An elliptic system corresponding to Poisson's equation.* Acta Litt. Sci Szeged, **6** (1932-34), 27-33.
4. O. D. Kellogg, *Foundations of potential theory*, Berlin, 1929.
5. I. Privaloff, *On a theorem of S. Saks*, Rec. Math. (Mat. Sbornik), **51** (1941), 457-460.
6. W. Rudin, *Integral representations of continuous functions*, Trans. Amer. Math. Soc., **68** (1950), 278-286.
7. S. Saks, *On the operators of Blashke and Privaloff for subharmonic functions*, Rec. Math. (Math. Sbornik), **51** (1941), 451-456.
8. ———, *Note on defining properties of harmonic functions*, Bull. Amer. Math. Soc., **38** (1932), 380-382.
9. V. L. Shapiro, *The divergence theorem for discontinuous vector fields*, Ann. of Math., **68** (1958), 604-624.

THE INSTITUTE FOR ADVANCED STUDY
RUTGERS UNIVERSITY

TESTS FOR PRIMALITY BASED ON SYLVESTERS CYCLOTOMIC NUMBERS

MORGAN WARD

Introduction. Lucas, Carmichael [1] and others have given tests for primality of the Fermat and Mersenne numbers which utilize divisibility properties of the Lucas sequences (U) and (V); in this paper we are concerned only with the first sequence;

$$(U)\colon\ U_0,\ U_1,\ U_2,\ \cdots,\ U_n = \frac{\alpha^n - \beta^n}{\alpha - \beta},\ \cdots .$$

Here α and β are the roots of a suitably chosen quadratic polynomial $x^2 - Px + Q$, with P and Q coprime integers. (For an account of these tests, generalizations and references to the early literature, see Lehmer's Thesis [2]).

I develop here a test for primality of a less restrictive nature which utilizes a divisibility property of the Sylvester cyclotomic sequence [3]:

$$(Q) : Q_0 = 0,\ Q_1 = 1,\ Q_2,\ \cdots,\ Q_n = \prod_{\substack{1 \le r \le n \\ (r,n)=1}} (\alpha - e^{\frac{2\pi i r}{n}}\beta),\ \cdots$$

Here α and β have the same meaning as before. (U) and (Q) are closely connected [4]; in fact

$$U_n = \prod_{d \mid n} Q_d . \tag{1.1}$$

The divisibility property is expressed by the following theorem proved in § 3 of this paper.

THEOREM. *If m is an odd number dividing some cyclotomic number Q_n whose index n is prime to m, then every divisor of m greater than one has the same rank of apparition n in the Lucas sequence (U) connected with (Q).*

Here the rank of apparition or rank, of any number d in (U) means as usual the least positive index x such that $U_x \equiv 0 \pmod{d}$.

The following primality test is an immediate corollary.

Primality test. If m is odd, greater than two, and divides some cyclotomic number Q_n whose index n is both prime to m and greater than the square root of m, then m is a prime number except in two trivial cases: $m = (n-1)^2$, $n - 1$ a prime greater than 3, or $m = n^2 - 1$ with $n - 1$ and $n + 1$ both primes.

Received January 14, 1959.

The primality tests of Lucas and Carmichael are the special case when $n = m \pm 1$ is a power of two (which allows Q_n to be expressed in terms of V_n) with $X^2 - Px + Q$ suitably specialized.

2. **Notations.** We denote the rational field by R, and the ring of rational integers by I. The polynomial

$$f(x) = x^2 - Px + Q\ , \qquad P,\ Q, \text{ in } I \text{ and co-prime} \tag{2.1}$$

is assumed to have distinct roots α and β.

We denote the root field of $f(x)$ by $\mathscr{A}$ and the ring of its integers by $\mathscr{J}$. Thus $\mathscr{A}$ is either R itself, or a simple quadratic extension of R.

Let p be an odd prime of I, and $\mathfrak{p}$ a prime ideal factor of p in $\mathscr{J}$. Every element ρ of $\mathscr{A}$ may be put in the form $\rho = \alpha/a$ with α in $\mathscr{J}$ and a in I. The totality of such ρ with $(a, p) = 1$ forms a subring $\mathscr{J}_p$ of $\mathscr{A}$. Evidently $\mathscr{A} \supset \mathscr{J}_p \supset \mathscr{J} \supseteq I$. If we extend $\mathfrak{p}$ into $\mathscr{J}_p$ in the obvious way, we obtain a prime ideal $\mathfrak{P}$. The homomorphic image of $\mathscr{J}_p$ modulo $\mathfrak{P}$ is a field, $\mathscr{F}_p$. We denote the mapping of $\mathscr{J}_p$ onto $\mathscr{F}_p$ by $(\mathfrak{P})$.

Let $F_n(z)$ denote the cyclotomic polynomial of degree $\phi(n)$. $F_n(z)$ has coefficients in I, and if n is greater than one, then (Lehmer [2], Carmichael [1])

$$Q_n = \beta^{\phi(n)} F_n\left(\frac{\alpha}{\beta}\right)\ , \tag{2.2}$$

Furthermore

$$z^n - 1 = \prod_{d|n} F_n(z)\ . \tag{2.3}$$

3. **Proof of theorem.** Let m be an odd number greater than one which divides some term of (Q) whose index n is prime to m, so that

$$Q_n \equiv 0 \pmod{m}\ , \qquad (n,\ m) = 1. \tag{3.1}$$

Throughout the next three lemmas, p stands for a fixed prime factor of m.

LEMMA 1. *If $\mathfrak{p}$ is any ideal factor of p in $\mathscr{J}$, then*

$$(Q,\ p) = (\alpha,\ \mathfrak{p}) = (\beta,\ \mathfrak{p}) = (1)\ . \tag{3.2}$$

Proof. It suffices to prove that $(Q, p) = (1)$. Assume the contrary. Then $(p, P) = 1$. Since $U_1 = 1$ and $U_{x+2} = PU_{x+1} - QU_x \equiv PU_{x+1} \pmod{p}$, it follows by induction that $U_n \not\equiv 0 \pmod{p}$. Then by (1.1), $Q_n \not\equiv 0$

(mod p). But p divides m so that by (3.1) $Q_n \equiv 0 \pmod{p}$ a contradiction.

LEMMA 2. *The rank of apparition of p in (U) is n.*

Proof. Since $U_n \equiv 0 \pmod{p}$, p has a positive rank of apparition in (U), r say. Then r divides n. But by (1.1), $U_r = \prod_{d|n} Q_d$. Hence $Q_d \equiv 0 \pmod{p}$ for some d dividing both r and n. Clearly, if $d = n$, then $r = n$ and we are finished. Assume that d is less than n.

The number $\alpha/\beta = \alpha^2/Q$ is in $\mathscr{I}_p$ by Lemma 1. Let τ be its image in $\mathscr{F}_p$ under the mapping $(\mathfrak{P})$. Then by (2.2) and Lemma 1 $F_n(\tau) = F_d(\tau) = 0$ in $\mathscr{F}_p$. Consequently the resultant of the polynomials $F_n(z)$ and $F_d(z)$ is zero in $\mathscr{F}_p$. Therefore its inverse image under the mapping is in $\mathfrak{P}$. But this resultant is evidently in I. Therefore it must be divisible by p. But by formula (2.3), since $d < n$ the resultant of $F_n(z)$ and $F_d(z)$ must divide the discriminant $\pm n^{n-1}$ of $z^n - 1$. Thus $n \equiv 0 \pmod{p}$ so that $(n, m) \equiv 0 \bmod p$ which contradicts (3.1) and completes the proof.

LEMMA 3. *The rank of apparition in (U) of any positive power of p which divides m is n.*

Proof. Let p^k divide m, $k \geq 1$ and let the rank of p^k in (U) be r. Now $U_n = \prod_{d|n} Q_d \equiv 0 \pmod{p^k}$. But by Lemma 2, each Q_d with $d < n$ is prime to p. Hence r must equal n.

The theorem proper now follows easily. For let m' be any divisor of m other than one. By Lemma 3, every prime power dividng m' has rank of apparition n in (U). But the rank of apparition of m' in (U) is the least common multiple of the ranks of the prime powers of maximal order diving m'. (Carmichael [1]). Hence m' also has rank of apparition n in (U).

4. Proof of primality test. Assume that (3.1) holds for some n greater than $\sqrt{m}$. If m is not a prime, it has a prime factor $\leq \sqrt{m}$. Let p be the smallest such factor, and let

$$m = pq\,, \qquad q \geq 3\,. \tag{4.1}$$

Then p has rank n in (U) by Lemma 3. But by a classical result of Lucas, $U_{p\pm1} \equiv 0 \pmod{p}$. Hence n divides $p \pm 1$. If n is less than $p+1$, $\sqrt{m} < p \leq \sqrt{m}$, a contradiction. Hence $n = p + 1$. If $p = \sqrt{m}$, then $m = (n-1)^2$ and $n - 1$ is a prime. Since m is odd, $n \geq 4$. This is the first trivial case.

If $p < \sqrt{m}$, then $q \geq p + 2$ and $m \geq p(p+2)$. But if $m > p(p+2)$, then $n^2 > m \geq (p+1)^2 = n^2$, a contradiction. Hence $m = p(p+2)$ where $p+2$ has no prime factor smaller than p. Hence $p+2$ is a prime and $m = n^2 - 1$ with both $n-1$ and $n+1$ primes. This is the second trivial case. In every other case then, m must be a prime.

5. Conclusion. The two trivial cases can actually occur. For if $P = 22$ and $Q = 3$, then $Q_6 = \alpha^2 - \alpha\beta + \beta^2 = P^2 - 3Q = 475$. Hence $Q_6 \equiv 0 \pmod{25}$ and $25 = (6-1)^2$. Again, if $P = 17$ and $Q = 3$, then $Q_6 = 280$. Hence $Q_6 \equiv 0 \pmod{35}$ and $35 = 6^2 - 1 = 5 \times 7$. It is worth noting that these trivial cases cannot occur if α and β are rational integers. (See [1], Theorem XII and remark.)

To illustrate the theorem, note that if $P = 2$ and $Q = 1$, $Q_9 = 73$. Since $\sqrt{73} < 9$ and $(9, 73) = 1$, 73 is a prime. But for $P = 3$ and $Q = 1$, $Q_9 = 91$. But $9 < \sqrt{91}$ so the test is inapplicable. As a matter of fact, 91 is the product of two primes. Evidently the test may be extended to cover such a case. That is, if $Q_n \equiv 0 \pmod{m}$, $(n, m) = 1$ and $n > \sqrt[3]{m}$, m will usually be either a prime, or the product of two primes.

References

1. R. D. Carmichael, *On the numerical factors of arithmetic forms*, Ann. of Math., **15** (1913-14), 30-70.

2. D. H. Lehmer, *An extended theory of Lucas functions*, Ann. of Math. **31** (1930), 419-448.

3. J. J. Sylvester, *On certain ternary cubic form equations*, Amer. J. Math. **2** (1879), 357-83.

4. Morgan Ward, *The mappings of the positive integers into themselves which preserve division*, Pacific J. Math. **5** (1955), 1013-1023.

CALIFORNIA INSTITUTE OF TECHNOLOGY

A FIXED POINT THEOREM FOR CHAINED SPACES

L. E. WARD, JR.

1. **Introduction.** There are a number of theorems in the literature of the following type: if a topological space is acyclic in the sense of containing no simple closed curve, and if other appropriate conditions are satisfied then the space has the fixed point property, that is, each continuous function f of the space into itself admits a solution of the equation $x = f(x)$. For example, if the space is compact metric and locally connected (i.e., a dendrite) then it has the fixed point property. There are many generalizations of this theorem. Appropriate to this discussion are of those of Borsuk [1], Plunkett [2], Wallace [3], the author [5] and [6], and Young [8]. A common characteristic of these generalizations is their requirement, explicit or implicit, of rather strong unicoherence conditions. But it is clear that many relatively simple acyclic spaces possessing the fixed point property are not unicoherent. As an example consider the following sets in the Cartesian plane:

$$A = \{(x, y) : 0 < x \leq 1, y = \sin(\pi/x)\} ,$$
$$B = \{(0, y) : -2 \leq y \leq 1\} ,$$
$$C = \{(x, -2) : 0 \leq x \leq 1\} ,$$
$$D = \{(1, y) : -2 \leq y \leq 0\} .$$

The continuum $M = A \cup B \cup C \cup D$ is not unicoherent but it is arcwise connected, acyclic, and has the fixed point property. It is the purpose of this note to formulate and prove a fairly general result which includes this and related examples. In so doing we shall generalize the theorems of Borsuk and Young cited above. As in our earlier papers the methods used here are order-theoretic in character. Section 2 is devoted to the partial order stucture of the spaces to be considered, and may be regarded as an addendum to [4], [6] and [7].

2. **Chained spaces.** Throughout all spaces to be considered are Hausdorff. By a *topological chain* or, more simply, a *chain,* we mean a continuum (=compact connected set) which has exactly two non-cutpoints. These two points are, of course, endpoints and a chain is simply the natural analogue of an arc in spaces which are not assumed to be metric. A space is *topologically chained* or *chained* provided each two distinct points lie in some chain. Obviously each two distinct points of a chained space are the endpoints of some chain. If a space has the property that each two distinct points are the endpoints of at most one chain, then it is said to be *acyclic.* In this case the unique chain whose

Received February 4, 1959.

endpoints are x and y is denoted $[x, y]$. It is convenient to define $[x, x]$ to be the set whose only element is x.

Acyclic chained spaces have an inherent partial order structure which facilitates their study. By a *partial order* on a set we mean a binary, reflexive, transitive relation $\leqq$ between elements of the set which, in addition, satisfies the rule

$$x \leqq y \textit{ and } y \leqq x \textit{ implies } x = y.$$

If $x \leqq y$ but $x \neq y$ we write $x < y$, and if P is a partially ordered set we define

$$L(x) = \{y \in P : y \leqq x\} , \quad M(x) = \{y \in P : x \leqq y\} .$$

In order to characterize acyclic chained spaces we recall a related theorem from [7]. A *dendritic* space is a connected and locally connected space in which each two distinct points can be separated by the omission of some third point.

THEOREM 1. *A necessary and sufficient condition that a locally connected space be dendritic is that it admit a partial order satisfying*

(i) *$L(x)$ and $M(x)$ are closed sets for each point x,*

(ii) *if $x < y$ then there exists z such that $\times < z$ and $z < y$,*

(iii) *for each x and y the set $L(x) \cap L(y)$ is nonempty, compact and simply ordered,*

(iv) *for each x the set $M(x) - x$ is open.*

Although many chained spaces are not locally connected (e.g., the space M of § 1) they can be made locally connected by properly altering the topology. This change of topology preserves the original chain structure of the space, and functions which are continuous in the orginal topology remain continuous in the new one. This technique appears to have originated with Young [8]. If X is a Hausdorff space let us say that a *chain component* of X is any subset of X which is maximal with respect to being chained. The *chain topology* is that topology which results from taking the chain components of open sets of the given topology as a basis for the chain topology. It is easily seen (and was proved in [8]) that any space is locally connected in its chain topology.

LEMMA 1. *An acyclic chained space is dendritic with respect to its chain topology.*

Proof. Let x and y be distinct points of the acyclic chained space X and let $z \in [x, y] - x \cup y$. Since X is acyclic no chain in $X - z$ contains both x and y, and therefore z separates x and y in the chain topology. Since X is connected and locally connected in the chain topology it is dendritic.

From Theorem 1 and Lemma 1 we infer that each acyclic chained space is endowed with an intrinsic partial order structure which can aptly be called the *chain cutpoint ordering*. It can be described in the following way (compare with [7]). Select an element e and define $x \leqq y$ if and only if $x \in [e,y]$. We now prove that the chain cutpoint ordering characterizes the acyclic spaces.

THEOREM 2. *A necessary and sufficient condition that the Hausdorff space X be acyclic and chained is that it be dendritic in its chain topology.*

Proof. The necessity was established in Lemma 1. To prove the sufficiency of the condition let X be a space which is dendritic in its chain topology. By Theorem 1 X admits a partial order which satisfies (i) – (iv) relative to the chain topology. If x and y are distinct points of X then by (ii) and (iii) they are contained in a continuum $L(x) \cup L(y)$ and by Theorem 3 of [7] that continuum is a tree. Since a tree is chained, so is X. If two distinct chains C_1 and C_2 have common endpoints, let A_1 be a component of $C_1 - C_2$, x and y the endpoints of $\bar{A}_1$, and $\bar{A}_2$ the minimal subchain of C_2 which joins x and y. Obviously no point can separate x and y in the chain topology, for it would have to lie in $A_1 \cap \bar{A}_2 = 0$. Since this is a contradiction we conclude that X is acyclic.

3. **A condition on rays.** Let X be a space and $e \in x$. A *ray of X with endpoint e* is the union of a maximal nest of chains which have e as a common endpoint. Thus, in a Euclidean space a half line emanating from the origin is a ray in this sense. In the example of § 1 the set A is a ray of M with endpoint (1,0).

If R is a ray with endpoint e in the space X and $x \in R$, let $A(R,x)$ be the closure of $(R - [e, x]) \cup x$, We then define

$$K_R = \bigcap \{A(R, x) : x \in R\} .$$

In a Euclidean space a ray R consisting of a half line emanating from the origin has $K_R = 0$. However, in the example of § 1 the set A has K_A equal to a closed line segment.

The crux of our fixed point argument is the following. If $f : X \rightarrow X$ is continuous where X is acyclic and chained, we examine the points x such that $x \leqq f(x)$. Either there is a "last" such point in a restricted order-theoretic sense, in which case that point is fixed by a continuity argument, or else such points form a ray R. Then we can show that $f(K_R) \subset K_R$, so that the fixed point property follows provided each K_R has the fixed point property.

We begin by formalizing this condition on rays.

(F_A) If R *is a ray with endpoint* a *then* K_R *has the fixed point property.*

In the example of § 1 let $a = (1, -2)$. Then there are two rays with endpoint a, $B \cup C$ and $A \cup D$. Since $K_{B \cup C}$ is a point and $K_{A \cup D}$ is a line segment the space M satisfies (F_a).

THEOREM 3. *If* X *is an arcwise connected space in which the union of any nest of arcs is contained in an arc then* X *is acyclic and* X *satisfies* (F_a) *for each* $a \in X$.

Proof. Since the union of any nest of arcs is contained in an arc, X is acyclic; and if R is a ray then $\overline{R}$ is evidently an arc so that K_R is a point.

The substance of Young's fixed point theorem [8] is that the spaces of Theorem 3 have the fixed point property; hence, Theorem 5 below is truly a generalization.

THEOREM 4. *If* X *is an arcwise connected, hereditarily unicoherent continuum then* X *satisfies* (F_a) *for each* $a \in X$.

Proof. We note that each subcontinuum of X is arcwise connected, for if x and y are elements of the subcontinuum Y and $[x, y] - Y$ is not empty then $[x, y] \cup Y$ would not be unicoherent. Now if R is a ray of X then K_R, being the intersection of a nest of continua, is a continuum and hence is itself arcwise connected and hereditarily unicoherent. Borsuk's theorem [1] asserts that such sets have the fixed point property.

This result demonstrates that all continua statisfying the hypothesis of Borsuk's fixed point theorem are included in Theorem 5.

If A and B are subsets of a partially ordered set with $A \subset B$ then A is *cofinal* in B provided for each $b \in B$ there exists $a(b) \in A$ such that $b \leqq a(b)$.

THEOREM 5. *Let* X *be a topologically chained acyclic space and suppose there exists* $e \in X$ *such that* (F_e) *is satisfied. Then* X *has the fixed point property.*

Proof. We give X the chain cutpoint ordering with minimal element e and let $f : X \to X$ be a continuous function. Consider the family $\mathscr{S}$ of all pairs (S, S') satisfying the following six conditions:

(i) S *is a nonempty simply ordered subset of* X,

(ii) *S and S' are connected,*
(iii) *S' is cofinal in S,*
(iv) $e \in S$,
(v) *$x \leqq f(x)$ for each $x \in S$,*
(vi) *$S \cup f(S')$ is simply ordered.*

Obviously the pair (e, e) is a member of $\mathscr{S}$. We can partially order $\mathscr{S}$ by defining $(S_\gamma, S'_\gamma) \prec (S_\delta, S'_\delta)$ if and only if $S_\gamma \subset S_\delta$ and $S_\delta \cup f(S'_\gamma)$ is simply ordered. If $\mathscr{N} = \{(S_\gamma, S'_\gamma)\}$ is a $\prec$ – simple subfamily of $\mathscr{S}$ and $S = \bigcup \{S_\gamma\}$, $S' = \bigcup \{S'_\gamma\}$ then it is clear that $(S, S') \in \mathscr{S}$ and that (S, S') is a $\prec$ – upper bound of $\mathscr{N}$. Thus Zorn's lemma can be applied; let (S_0, S'_0) be a $\prec$ – maximal member of $\mathscr{S}$.

If $x_0 = \sup S_0$ exists we assert that $x_0 \leqq f(x_0)$. For suppose there is $t \in S_0$ such that $f(x_0)$ is not a successor of t. We may assume $t < f(t)$; if $T = [t, x_0]$ then $f(T)$ is a tree and t separates $f(t)$ and $f(x_0)$ in $f(T)$. If W is the component of $f(T) - t$ which contains $f(x_0)$ then W is a neighborhood of $f(x_0)$ in the relative topology of $f(T)$ and hence there is $q \in S_0$, $t < q < x_0$ such that $f(q) \in W$. But this implies that $f(q)$ is not a successor of q, a contradiction. Therefore, $t \leqq f(x_0)$ for each $t \in S_0$ and hence $x_0 \leqq f(x_0)$. If $x_0 < f(x_0)$ let U be a connected neighborhood of $f(x_0)$ *relative to the chain topology* such that $\overline{U} \subset X - x_0$. Then there exists $x_1 \in X - \overline{U}$ such that $x_0 < x_1 < f(x_0)$ and $f([x_0, x_1]) \subset U$. But then each point $p \in [x_0, x_1]$ satisfies $p \leqq f(p)$; letting $S_1 = S_0 \cup [x_0, x_1]$, $S'_1 = x_1$, it is apparent that $(S_0, S'_0) \prec (S_1, S'_1)$ in contradiction of the maximality of S_0. Hence $x_0 = f(x_0)$.

On the other hand if S_0 has no supremum it is a ray R with endpoint e and it remains only to show that $f(K_R) \subset K_R$. By (vi) and the fact that $S_0 = R$ is a ray we have $f(S'_0) \subset R$. Moreover, $A(R, x) \subset \overline{S'_0}$ for each $x \in S'_0$ and hence $K_R \subset \overline{S'_0}$. Therefore, $f(K_R) \subset \overline{R}$. Now suppose $f(y) \in \overline{R} - K_R$ for some $y \in K_R$. Let V be a neighborhood of $f(y)$ such that $\overline{V}$ and K_R are disjoint; then $\overline{V}$ and $A(R, x)$ are disjoint for some $x \in R$ and there exists $a \in R - [e, x]$ such that $f(a) \in V$. Moreover, it is clear by (ii) and (iii) that a may be so chosen that $a \in S'_0$ and hence $f(a) \in A(R, x)$, a contradiction. Therefore, $f(K_R) \subset K_R$ and the proof is complete.

References

1. K. Borsuk, *A theorem on fixed points*, Bull. Acad. Polon. Sci. CI III, **2** (1954), 17-20.
2. R. L. Plunkett, *A fixed point theorem for continuous multivalued transformations*, Proc. Amer. Math. Soc., **7** (1956), 160-3.
3. A. D. Wallace, *A fixed point theorem for trees*, Bull. Amer. Math. Soc., **47** (1941), 757-60.
4. L. E. Ward, Jr., *A note on dendrites and trees*, Proc. Amer. Math. Soc., **5** (1954), 992-4.

5. ———, *Mobs, trees and fixed points*, Proc. Amer. Math. Soc., **8** (1957), 798-804.

6. ———, *A fixed point theorem for multi-valued functions*, Pacific J. Math., **8** (1958), 921-927.

7. ———, *On dendritic sets*, Duke Math. J., **25** (1958), 505-14.

8. G. S. Young, *The introduction of local connectivity by change of topology*, Amer. J. Math., **68** (1946), 479-94.

U. S. NAVAL ORDNANCE TEST STATION
CHINA LAKE, CALIFORNIA
UNIVERSITY OF OREGON

SILOV TYPE C ALGEBRAS OVER A CONNECTED LOCALLY COMPACT ABELIAN GROUP

ALFRED B. WILLCOX

A certain class of commutative Banach algebras of functions on a compact abelian group has been studied by G. E. Silov [6]. His algebras, which he calls homogeneous rings, are partially characterized by the property of containing arbitrary translates of elements. The most interesting examples are various algebras of complex functions on the circle or torus of any dimension with various differentiability properties and algebras of continuous functions on a compact abelian group which have absolutely convergent Fourier series. Silov's results have been extended by Mirkil [5] to algebras over non-abelian compact groups. We present here some results which generalize parts of the theory to translation closed algebras over connected locally compact abelian groups. The major problem in an extension in this direction centers about a replacement for the type of classical Fourier analysis for continuous functions on compact groups which has no satisfactory analog even in the abelian non-compact case. Our approach to this problem is to recapture *locally* some of the compact case when it becomes necessary. This approach makes it necessary to add to Silov's conditions various additional assumptions. Nevertheless, a considerable portion of the theory survives; enough, in fact, to include analogs of all the interesting examples from the compact case. In § 1 we present the basic construction on which the structure theorems of § 2 are based. In § 3 various examples are discussed. It will be assumed that the reader is familiar with the general theory of commutative regular Banach algebras. An account assuming an identity can be found in [6]. The results extend easily to algebras without identity. Such extensions can be found in [2], [3], [4], or, for certain non-commutative algebras, in [8].

1. In this section we describe a method of constructing a Banach algebra from the following ingredients:

(i) a connected locally compact abelian group G,

(ii) a primary commutative Banach algebra K with identity, maximal ideal Q, and norm $|\cdot|$, and

(iii) a homomorphism ω of the character group $\hat{G}$ of G into the coset of the identity in K modulo Q.

By well-known structure theorems [7, section 29] $G = E_p \times G_c$ where E_p is the p-dimensional vector group and G_c is compact abelian. From

Received August 8, 1958, and in revised form February 27, 1959.

this it follows easily that G is σ-compact, i.e., G contains a sequence $\{C_n\}$ of compact neighborhoods of the identity 0 such that

(1) C_n is contained in C_{n+1} for all n and

(2) $G = \bigcup_{n=1}^{\infty} C_n$.

Such a sequence $\{C_n\}$ will be called a σ-covering of G. If f is a complex function defined on G and $\{C_n\}$ is a fixed σ-covering we denote by $[f]^{(n)}$ the function defined by

$$[f]^{(n)}(t) = f(t)\ , \qquad t \in C_n$$
$$[f]^{(n)}(t) = 0\ , \qquad t \notin C_n\ .$$

Now suppose that for each $n = 1, 2, \cdots$ we have a linear combination of characters $\sum_{i=1}^{k_n} c_{in}\chi_{in}$, c_{in} complex, $\chi_{in} \in \hat{G}$. Form the sequence $\{f^{(n)}\}$ with $f^{(n)} = [\sum_{i=1}^{k_n} c_{in}\chi_{in}]^{(n)}$. Such a sequence will be called *ω-Cauchy* if it is Cauchy in the metric

$$N(f^{(n)} - f^{(m)}) = \sup_{t \in G} | \sum c_{in}[\chi_{in}]^{(n)}(t)\omega(\chi_{in}) - \sum c_{jm}[\chi_{jm}]^{(m)}(t)\omega(\chi_{jm}) |\ .$$

$N(f^{(n)})$ is defined in the obvious way, and it is clear that

$$| N(f^{(n)}) - N(f^{(m)}) | \leqq N(f^{(n)} - f^{(m)})\ .$$

Thus the complex sequence $\{N(f^{(n)}\}$ is Cauchy if $\{f^{(n)}\}$ is ω-Cauchy. We define $|| \{f^{(n)}\} ||$ to be $\lim N(f^{(n)})$, $n \to \infty$. If $\{f^{(n)}\}$ and $\{g^{(n)}\}$ are ω-Cauchy then $\{(f - g)^{(n)}\}$ is also ω-Cauchy. $\{f^{(n)}\}$ and $\{g^{(n)}\}$ will be called equivalent if $|| \{(f - g)^{(n)}\} || = 0$. *The resulting set of equivalence classes of ω-Cauchy sequences $\{f^{(n)}\}$ will be denoted by $K_\omega(G)$.* In $K_\omega(G)$ we introduce the obvious operations $\alpha\{f^{(n)}\}$, $\{f^{(n)}\} + \{g^{(n)}\}$ and $\{f^{(n)}\} \cdot \{g^{(n)}\}$. With the above norm $K_\omega(G)$ is clearly a normed complex algebra.

THEOREM 1.1. *$K_\omega(G)$ is a Banach algebra independent of the choice of the σ-covering $\{C_n\}$.*

We omit the details of the proof of this theorem. The second statement follows readily from remark (A) below, and a more or less standard diagonalization process shows that $K_\omega(G)$ is complete.

Two remarks on the structure of $K_\omega(G)$ are immediate.

(A) $K_\omega(G)$ is isomorphic and isometric to an algebra of continuous K-valued functions defined on G and vanishing at ∞, the norm being the usual sup norm. This can be seen as follows. Each element $\{f^{(n)}\}$ of $K_\omega(G)$ is a Cauchy sequence in the Banach algebra of all bounded K-valued functions on G with the sup norm. Assign to $\{f^{(n)}\}$ its limit $\tilde{f}$ in this algebra. $\tilde{f}(t)$ is necessarily continuous since any $t_0 \in G$ has a neighborhood within which $f^{(n)}(t)$ is continuous for all sufficiently large

n. $\tilde{f}(t) \to 0$ as $t \to \infty$ since each $f^{(n)}(t)$ has compact support. The mapping $\{f^{(n)}\} \to \tilde{f}$ is clearly a homomorphism. Moreover,

$$\begin{aligned} \| \{f^{(n)}\} \| = \lim_n N(f^{(n)}) &= \lim_n \sup_t | \textstyle\sum c_{in}[\chi_{in}]^{(n)}(t)\omega(\chi_{in}) | \\ &= \sup_t \lim_n | \textstyle\sum c_{in}[\chi_{in}]^{(n)}(t)\omega(\chi_{in}) | \\ &= \sup_t |\tilde{f}(t)| . \end{aligned}$$

so the correspondence is an isometry.

(B) Since $\omega(\chi)(Q) = 1$ for each $\chi \in \hat{G}$ we have $| \sum c_{in}[\chi_{in}]^{(n)}(t) | \leqq | \sum c_{in}[\chi_{in}]^{(n)}(t)\omega(\chi_{in}) |$. Thus each element of $K_\omega(G)$ determines uniquely a complex function $f(t)$ such that $\sup |f(t)| \leqq \| \{f^{(n)}\} \|$. The mapping $\{f^{(n)}\} \to f$ is a continuous homomorphism of $K_\omega(G)$ onto a subalgebra of $C_0(G)$, the Banach algebra of all continuous complex functions vanishing at ∞ on G. $K_\omega(G)$ will be said to be *radical* or to *separate points of* G accordingly as the corresponding subalgebra of $C_0(G)$ is zero or separates points of G.

In the sequel we shall denote a general element of $K_\omega(G)$ by $\tilde{f}$ as suggested by (A) and the image of this element in the corresponding subalgebra of $C_0(G)$ by f.

EXAMPLES. (1) Remark (B) and the Stone-Weierstrass theorem show that if ω is the trivial homomorphism sending each χ into the identity in K then $K_\omega(G) = C_0(G)$.

(2) Let $G = E_1$ and K be the Banach algebra with two generators 1, x with $x^2 = 0$. K is the set of all polynomials $\alpha_0 + \alpha_1 x$, α_i complex, with norm defined by $|\alpha_0 + \alpha_1 x| = |\alpha_0| + |\alpha_1|$. K is primary with Q the subalgebra generated by x. $\hat{G} = E_1$ and a general character is $\chi(t) = e^{i\lambda t}$, $\lambda \in E_1$. Define ω by $\omega(\chi) = \omega(\lambda) = 1 + i\lambda x$. ω is clearly a continuous homomorphism. A general element $\{f^{(n)}\}$ of $K_\omega(G)$, with $f^{(n)} = [\sum c_{pn}\chi_{pn}]^{(n)}$, is a function $\tilde{f}(t) = f(t) + g(t)x$ where

$$f(t) = \lim_n \sum_p c_{pn}\chi_{pn}(t) ,$$

$$g(t) = \lim_n \sum_p c_{pn}\chi'_{pn}(t)$$

and both limits are uniform in a neighborhood of each $t_0 \in E_1$. Thus $g(t) = f'(t)$ and both $f(t)$ and $f'(t)$ tend to 0 at ∞. $K_\omega(G)$ is the algebra $D_1(E_1)$ of Example 1, § 3. Various properties of $K_\omega(G)$ are immediate from standard theorems on Fourier series. We point out several which play roles in subsequent theorems of this section. The homomorphism $\tilde{f} \to f$ of remark (B) is clearly an isomorphism in this case. Moreover, if f is any complex continuously differentiable function on E_1 with compact support then $\tilde{f} \in K_\omega(G)$. This is obvious if we take for a σ-covering the collection of intervals $[-n, n]$ and look at the Fourier series for such a function on an arbitrary interval $[-n, n]$ containing the support of f. To obtain a sequence $\{f^{(n)}\}$ defining $\tilde{f}$ we need only take, for each sufficiently large n, a suitable partial sum of the Fourier series for f on $[-n, n]$. Thus $K_\omega(G)$ contains elements $\tilde{f}$ such that $f(t) = 1$ on an arbitrary compact subset of G and $f(t) = 0$ on a disjoint closed set. By Theorem 1.5 below G is the space of maximal regular ideals of $K_\omega(G)$ so $K_\omega(G)$ is a regular Banach algebra. In fact, by the definition of the norm $K_\omega(G)$ contains a *bounded* sequence $\{\tilde{f}_n\}$ for which $\tilde{f}_n(t) = 1$ on $[-n, n]$ and $\tilde{f}_n(t)$ has compact support. Such a sequence is an "approximate identity" in $K_\omega(G)$, i.e., $\lim \tilde{f}\tilde{f}_n = \tilde{f}$ for any $\tilde{f} \in K_\omega(G)$. Thus the elements with compact support are dense in $K_\omega(G)$. Finally, any element $\tilde{f}$ whose support is contained in $[-n, n]$ can be approximated uniformly on $[-n, n]$ by K-valued functions of the form $\Sigma c_p \omega(\chi_p)\chi_p(t)$ where each χ_p is constant on the subgroup $\{0, \pm n, \pm 2n, \cdots\}$, or, equivalently, each χ_p is an integral multiple of $2\pi/n$ (cf. condition (A) below). This, too, follows from a glance at the Fourier series for the image f on the interval $[-n, n]$.

LEMMA 1.2. *For any $K_\omega(G)$ we have the following:*

(a) *$f(t) = \tilde{f}(t)(Q)$ for any $\tilde{f} \in K_\omega(G)$,*

(b) *$K_\omega(G)$ is closed under multiplication by $\hat{G}$ in the sense that for $\tilde{f} \in K_\omega(g)$ and $\chi \in \hat{G}$ there exists an element $\chi\tilde{f} \in K_\omega(G)$ such that $[\chi\tilde{f}](t) = \chi(t)\omega(\chi)\tilde{f}(t)$ for all $t \in G$.*

(c) *$K_\omega(G)$ is closed under translation in the sense that for $\tilde{f} \in K_\omega(G)$ and $s \in G$ there exists an element $\tilde{f}_s \in K_\omega(G)$ such that $\tilde{f}_s(t) = \tilde{f}(t-s)$ for all $t \in G$.*

Proof. For each $t \in G$,

$$\begin{aligned}\tilde{f}(t)(Q) &= [\lim \Sigma_i c_{in}[\chi_{in}]^{(n)}(t)\omega(\chi_{in})](Q) \\ &= \lim \Sigma_i c_{in}[\chi_{in}]^{(n)}(t)\{\omega(\chi_{in}(Q)\} = f(t)\,,\end{aligned}$$

since $\omega(\chi_{in})(Q) = 1$. This proves (a). (b) is clear: if $\tilde{f} \leftrightarrow \{f^{(n)}\}$ then $\chi\tilde{f} \leftrightarrow \{[\chi f]^{(n)}\}$. (c) would be equally trivial if it were true that $[\chi]^{(n)}(t-s) = \chi(-s)[\chi]^{(n)}(t)$ for all $t \in G$. Since this is not the case a slight extra argument is necessary. Let $\tilde{f} \in K_\omega(G)$ with

$$\sup_{t \in G} | \Sigma_i c_{in}[\chi_{in}]^{(n)}(t)\omega(\chi_{in}) - \tilde{f}(t) | \to 0 \quad \text{as} \quad n \to \infty .$$

For each n pick an integer n' in such a way that $n' \to \infty$ as $n \to \infty$ and $C_{n'} \supset C_n - s$ for all n. Then for any $t \in C_n$

$$[\chi_{in'}]^{(n')}(t-s) = [\chi_{in'}]^{(n)}(t) \cdot \chi_{in'}(-s) \ .$$

We may assume that $|\tilde{f}(t)| < \varepsilon$ for $t \notin C_n$, n sufficiently large, so it follows that

$$\sup_{t \in G} | \Sigma_i c_{in'} \chi_{in'}(-s)[\chi_{in'}]^{(n)}(t)\omega(\chi_{in'}) - \tilde{f}(t-s) | < \varepsilon$$

for sufficiently large n. This means that $\tilde{f}_s \in K_\omega(G)$.

LEMMA 1.3. *$K_\omega(G)$ is either radical or separates points of G.*

This follows immediately from Lemma 1.2, parts (b) and (c) together with the fact that $\hat{G}$ separates points of G. This lemma together with remark (B) yield the following lemma. Again we omit the details of the easy proof. We denote the structure space of maximal regular ideals of $K_\omega(G)$ by $\mathfrak{M}_K$.

LEMMA 1.4. *For $t \in G$ the set $M_t = \{\tilde{f} \in K_\omega(G) \,|\, f(t) = 0\}$ is a maximal regular ideal of $K_\omega(G)$. Given an arbitrary $\tilde{f} \in K_\omega(G)$ the image $\tilde{f}(M_t)$ of $\tilde{f}$ modulo the maximal regular ideal M_t is $f(t)$. If $K_\omega(G)$ is not radical then the mapping $t \rightarrow M_t$ is one-to-one of G into $\mathfrak{M}_K$.*

Denote by $TK_\omega(G)$ the "Tauberian part" of $K_\omega(G)$, that is, the closed subalgebra of $K_\omega(G)$ generated by the elements $\tilde{f}(t)$ which have compact support. Lemmas 1.2, 1.3, and 1.4 hold for the algebra $TK_\omega(G)$, and we denote its structure space by $\mathfrak{M}_{TK}$. Given the conditions of Lemma 1.4 we will identify G with its image in $\mathfrak{M}_{TK}$ or $\mathfrak{M}_K$. We will be interested in algebras $K_\omega(G)$ and $TK_\omega(G)$ primarily when they are regular. Whether there actually exists a non-regular $K_\omega(G)$ is an interesting open question to which we will refer again in some remarks at the end of this section.

THEOREM 1.5. *Let ω be continuous. If $TK_\omega(G)$ is not radical then $G = \mathfrak{M}_{TK}$. If $TK_\omega(G)$ is regular then the group topology in G is the same as the $\mathfrak{M}_{TK}$-topology.*

Proof. The proof of the first statement is very similar to Silov's proof of the analogous theorem for the compact case so we omit most of the details. If $M_0 \in M_{TK}$ consider $\tilde{e} \in TK_\omega(G)$ such that $\tilde{e}(t)$ has compact support and $\tilde{e}(M_0) = 1$. Let $m(\chi) = [\chi\tilde{e}](M_0)$. One shows that $m(\chi)$ is a homomorphism of $\hat{G}$ into the complexes of modulus 1. Since ω is continuous it follows that m is continuous. Thus by the duality theorem $m(\chi) = \chi(t_0)$ for some $t_0 \in G$. This says that $\tilde{f}(M_0) = f(t_0)$ for any element which is a linear combination of elements $\chi\tilde{e}$, hence, by definition of $TK_\omega(G)$, for any element $\tilde{g}\tilde{e}$ with $\tilde{g} \in TK_\omega(G)$. The desired result follows since $\tilde{e}(M_0) = 1$. The second statement in the theorem follows from standard theorems in topology. By definition of the Gelfond topology, the $\mathfrak{M}_{TK}$-topology is weaker than the group topology on G. Both are Hausdorff and locally compact, and if $TK_\omega(G)$ is regular then an $\mathfrak{M}_{TK}$-compact set K is G-compact (since $TK_\omega(G)$ has a unit modulo the kernel of K and all elements tend to zero at ∞ on G). Thus the topologies are the same.

The last part of the above proof also yields the following.

COROLLARY 1.6. *If $K_\omega(G)$ is regular then G is closed in $\mathfrak{M}_K$ and its topology is inherited from $\mathfrak{M}_K$.*

We can now formulate a necessary and sufficient condition for any regular $TK_\omega(G)$ to be semi-simple. Recall that $G = E_n \times G_c$ so that G clearly contains a discrete subgroup D for which G/D is compact (D is essentially the group I_n, where I is the group of integers) and a compact neighborhood C of the identity such that the natural map of C into G/D is one-to-one. $TK_\omega(G)$, or, more, generally, any algebra R of continuous K-valued functions on G, will be said to satisfy *Condition* (A) if:

(1) $TK_\omega(G)$ (or R) contains elements $\tilde{f}(t)$ with $f(t)$ not identically zero such that $\tilde{f}(t)$ has support contained in C, and

(2) every $\tilde{f} \in TK_\omega(G)$ (or R) with support in C is a uniform limit on C of functions of the form $\Sigma c_i\chi_i(t)\omega(\chi_i)$ where the χ_i are elements of $\hat{G}$ which are constant on D, i.e., each χ_i is a character of G/D.

Condition (A) implies that any $\tilde{f} \in TK_\omega(G)$ supported by C determines uniquely a function $\tilde{f}(\bar{t})$ on G/D such that $\tilde{f}(\bar{t})$ is an element of $K_{\bar{\omega}}(G/D)$ where $\bar{\omega}$ is the homomorphism of the character group of G/D into K which is induced by ω. Thus $TK_\omega(G)$ is locally rather firmly tied to the compact case.

The following lemma is stated in a form in which it will be applicable both in the present discussion and later in § 2.

LEMMA 1.7. *Let R be a semi-simple regular Banach algebra of continuous functions $\tilde{f}$ from G to K vanishing at ∞ with $\|\tilde{f}\| = \sup|\tilde{f}(x)|$; $x \in G$. Suppose $\mathfrak{M}(R) = G$ and that R is closed under translation and multiplication by $\hat{G}$ in the sense of Lemma 1.2. Then*

(a) *for any $\tilde{f} \in R$, $\tilde{f}(t)$ vanishes on any open set in G on which $f(t) = \tilde{f}(M_t)$ vanishes, and*

(b) *R satisfies Condition* (A).

Proof. The proof of (a) is exactly the proof of the corresponding lemma (4.7.1) in [6] so we omit the details.

Denote a general element of G by (s, t) where $s = (\alpha_1, \alpha_2, \cdots, \alpha_n) \in E_n$, $t \in G_c$. For real $\alpha > 0$ define $S(\alpha) = \{(s, t) \mid |\alpha_i| \leqq \alpha, t \in G_c\}$. For the discrete subgroup D we can take the direct product of the usual discrete subgroup I_n of E_n and the identity subgroup of G_c. G/D is then the product of an n-torus and G_c. We may further assume that the compact neighborhood $S(\alpha)$ of 0 with the usual identifications, operations and topology is isomorphic and homeomorphic to G/D. If C is a compact subset of G containing $S(\alpha)$ then $\tilde{f} \in R$ is said to be D-periodic on C if for any $x \in C$, $d \in D$ for which $x + d \in C$ we have $\tilde{f}(x + d) = \tilde{f}(x)$. Clearly any D-periodic element on C determines uniquely both a continuous K-valued function on G/D and a similar complex valued function. R contains D-periodic functions on any compact set in G since regularity and part (a) of the theorem provide elements whose support is in $S(\alpha)$ and these can be extended to all of C by a finite number of translations by elements of D. (The possibility of multiplying a unit modulo the kernel of C by characters also yields D-periodic functions, but for reasons of later applicability we prefer not to make use of this hypothesis until later in the proof.) Suppose $\tilde{f}$ is D-periodic on $S(3\alpha)$ and that $h \in S(\alpha)$. Then the element $\tilde{f}_h$ is D-periodic on $S(2\alpha)$. Let I be the kernel of the subset $S(\alpha)$ of $\mathfrak{M} = G$ and let $\bar{R} = R/I$. Denote the image in $\bar{R}$ of a general $\tilde{f} \in R$ by $\bar{f}$. The norm of $\bar{f}$ in $\bar{R}$ is $\|\bar{f}\| = \|f\|_{S(\alpha)} = \inf \|g\|$; $g(x) = f(x)$ all $x \in S(\alpha)$. Let $\bar{R}_p$ be the closed subalgebra of $\bar{R}$ generated by all $\bar{f}_h$ with $\tilde{f}$ and h as above. Clearly $\bar{R}_p$ can be represented as an algebra of continuous complex functions on the compact abelian group G/D. Consider one of the generators $\bar{g} = \bar{f}_{h_1}$ and an element h in the interior of $S(\alpha)$. By adjusting h_1 by an element of D without changing the image $\bar{f}_{h_1}$ we can arrange to have $h_1 + h \in S(\alpha)$. Then $\tilde{g}_h = [\tilde{f}_{h_1}]_h$ is D-periodic on $S(2\alpha)$ and its image $\bar{g}_h$ is in $\bar{R}_p$. It is an easy exercise to show that if $\bar{t}$ denotes the image in G/D of $t \in G$ then $\bar{g}_h(\bar{t}) = \bar{g}(\bar{t} - \bar{h})$ so $\bar{g}_h$ is a translate of $\bar{g}$ in $\bar{R}_p$. The translation operator $T_{\bar{h}}$ is then defined on a dense subset of $\bar{R}_p$. We show that $T_{\bar{h}}$ is bounded. Let $\bar{f}$ be a general element of this dense set, i.e., $\tilde{f} = \Sigma_i [\tilde{f}_i]_{h_i}$ with $\tilde{f}_i$ D-periodic on $S(3\alpha)$, $h_i \in S(\alpha)$. Consider $\bar{f}_{\bar{h}}$, the image of $\tilde{f}_h$ as above. We must show that $\|\tilde{f}_h\|_{S(\alpha)} \leqq k \|\tilde{f}\|_{S(\alpha)}$ where k is independent of $\tilde{f}$. Let $S = S(\alpha) + h$. Clearly S is in the interior of $S(2\alpha)$. Choose a closed set T such that $S(\alpha) \cup S \subset T \subset$ interior $S(2\alpha)$. It is obvious that $\|\tilde{f}\|_{S(\alpha)} = \|\tilde{f}_h\|_S$. We show that $\|\tilde{f}_h\|_{S(\alpha)} \leqq \|\tilde{f}_h\|_T \leqq k \|\tilde{f}_h\|_S$. The first inequality is clear since $T \supset S(\alpha)$. Pick $\tilde{e} \in R$ such that $e(x) = 1$ on T, $e(x) = 0$ outside $S(2\alpha)$. Then $f_h e(x) = f_h(x)$ on T and

$$\|\tilde{f}_h \tilde{e}\| = \sup |\tilde{f}_h \tilde{e}(x)| \, (x \in S(2\alpha)) \leqq \|\tilde{e}\| \sup |\tilde{f}_h(x)| \, (x \in S)$$

by D-periodicity of $\tilde{f}_h$ on $S(2\alpha)$. By part (a) together with continuity of elements of R we see that $\|\tilde{f}_h\|_S \geqq \sup|\tilde{f}_h(x)|(x \in S)$ so we have $\|\tilde{f}_h\tilde{e}\| \leqq \|\tilde{f}_h\|_S \cdot \|\tilde{e}\|$. But $\|\tilde{f}_h\|_T \leqq \|\tilde{f}_h\tilde{e}\|$ so $\|\tilde{f}_h\|_T \leqq \|\tilde{e}\| \cdot \|\tilde{f}_h\|_S$. Hence T_h is bounded, hence extendible to $\bar{R}_p$ where it clearly defines the ordinary translate $\bar{f}_{\bar{h}}$ of an arbitrary $\bar{f} \in \bar{R}$. If h is on the boundary of $S(\alpha)$ we write $h = h_1 + h_2$, $h_i \in$ interior of $S(\alpha)$ and proceed as above. Since all $f \in R$ are *unjformly* continuous K-valued functions it follows that all elements of $\bar{R}_p$ are *continuous under translation*, that is, for any $\bar{f}$ and $\varepsilon > 0$, $\|\bar{f} - \bar{f}_{\bar{h}}\| < \varepsilon$ for all $\bar{h}$ in some neighborhood of $\bar{0}$. Thus $\bar{R}_p$ is a homogeneous space of functions in the sense of Silov satisfying the conditions of [6, 2.7]. We can therefore conclude that linear combinations of character of G/D are dense in $\bar{R}_p$.

If $\tilde{e} \in R$ is chosen so that $e(t) = 1$ on $S(3\alpha)$ and if χ_i are characters of G constant on D, then if $\tilde{g} = \Sigma c_i[\chi_i\tilde{e}]$ $\tilde{g}$ is in $\bar{R}_p$ and is the corresponding linear combination of characters in that algebra. $\tilde{g}(x) = \Sigma c_i \chi_i(x)\omega(\chi_i)$ for each $x \in S(\alpha)$ so Condition (A) follows from the fact, noted above, that $\|\bar{f}\| \geqq \sup|\tilde{f}(x)|(x \in S(\alpha))$.

THEOREM 1.8. *Let ω be continuous. If $TK_\omega(G)$ is regular then it is semi-simple if and only if it satisfies Condition* (A). *If $K_\omega(G)$ is regular then it satisfies Condition* (A) *if and only if it is semi-simple and $\mathfrak{M}_K = G$.*

Proof. Suppose $TK_\omega(G)$ is regular. Necessity of the condition is contained in Lemma 1.7 in view of the results of Theorem 1.5 and Lemma 1.2. Sufficiency follows readily from the fact that any $K_\omega(G)$ with G compact abelian is semi-simple [6, Theorem 4.6]. Suppose $\tilde{f} \in TK_\omega(G)$ and $f(t) = 0$ for all $t \in G$. Pick $\tilde{e} \in TK_\omega(G)$ with support contained in C and with $e(t_0) \neq 0$ (by Condition (A)). Then $\tilde{e}\tilde{f}$ is supported by C so $\tilde{e}\tilde{f}(\bar{t}) \in K_\omega(G/D)$ and $ef(\bar{t}) = 0$ for all $\bar{t}$. Thus $\tilde{e}\tilde{f}(\bar{t}) = 0$ for all $\bar{t}$ so that $\tilde{e}\tilde{f}(t_0) = 0$. $\tilde{e}(t_0)$ has an inverse in K since it is contained in no maximal ideal of K so we must have $\tilde{f}(t_0) = 0$. Thus, for each $s \in G$, $f_s(t_0) = 0$ which implies that $\tilde{f} = 0$. The statement for $K_\omega(G)$ follows by the same argument if we observe that we have actually proved that Condition (A) is equivalent to the vanishing of the kernel of G. For $TK_\omega(G)$ this is semi-simplicity since $G = \mathfrak{M}_{TK}$, For $K_\omega(G)$ the vanishing of this kernel is equivalent to semi-simplicity plus the condition that $G = \mathfrak{M}_K$, since we know by Corollary 1.6 that G is closed in $\mathfrak{M}_K$.

THEOREM 1.9. *Let ω be continuous. If $TK_\omega(G)$ $(K_\omega(G))$ is regular* *and semi-simple then it is an algebra of type C.*[1]

Proof. Using part (a) of Lemma 1.7 one easily proves that the set $\{\tilde{f} \mid \tilde{f}(t_0) = 0\}$ is a closed primary ideal. It is immediate, then, that the norm in $K_\omega(G)$ is smaller than the type C norm. But the opposite inequality always holds.

Before turning to some structure theorems based on the above construction we mention several questions concerning the algebras $TK_\omega(G)$ and $K_\omega(G)$. The first one concerns the connectivity assumption on G. The results in this section hold in slightly more generality. The definitions and most of the early results require only that G be σ-compact. Condition (A), Lemma 1.7, and Theorem 1.8 require only that G be generated by a compact neighborhood of the identity (so that $G = E_n \times G_c \times G_d$, G_d discrete [7, section 29]). Full use of connectivity is used only in the next section in the proof of Theorem 2.3. Whether connectivity could be dropped in favor of, say, σ-compactness is an open question. Further open questions concern some of the separation conditions we have employed. Does there exist a radical $K_\omega(G)$? Does there exist a non-regular $K_\omega(G)$? Does a $K_\omega(G)$ exist for which $TK_\omega(G) \neq K_\omega(G)$? These questions are closely related to the question of regularity of $K_\omega(G)$ in the compact case, and a complete answer to this question is not known. Silov has sufficient conditions for regularity of $K_\omega(G)$ for compact G [6, section 5.8], but no necessary conditions. In case $G = E_n$ and K is finite dimensional there is some evidence which suggests that $TK_\omega(G)$ is regular and equal to $K_\omega(G)$. This is true, for instance, for dimension $\leqq 3$, but the proof requires a classification of primary algebras of these dimensions. This approach is not promising in the general finite dimensional case, however, since a classification of all finite dimensional primary algebras is not known. (Such a classification would involve a classification of finite dimensional nilpotent algebras, a more familiar unsolved problem.) In case $G = E_1$ it is not hard to exhibit sufficient conditions for regularity of $TK_\omega(G)$ or $K_\omega(G)$ by reducing to the compact case where Silov's conditions can be applied. We state one such result without proof. If $G = E_1$ we may identity $\hat{G}$ with E_1, the circle group C with $E_1/I(p)$ where $I(p)$ is the subgroup of integral multiples of p, p a positive integer, and $\hat{C}$ with the group of integers. The homomorphism ω of E_1 into K induces, for each p, a homomorphism ω_p of C into K: $\omega_p(n) = \omega(n/p)$. *If $K_{\omega_p}(C)$ is regular for each $p = 1, 2, 3, \cdots$ then $K_\omega(E_1)$ and $TK_\omega(E_1)$ are regular.*

[1] A commutative regular B-algebra R is of type C if its norm is equivalent to the norm $|||f||| = \sup \|f\|_M$, where M ranges over the structure space of maximal regular ideals and $\|f\|_M$ is the norm of the image of f in the difference algebra $R/J(M)$ (see section 2).

2. In § 1 we have seen that under certain conditions algebras $TK_\omega(G)$ or $K_\omega(G)$ are semi-simple commutative Banach algebras of type C closed under multiplication by $\hat{G}$ and under translation. In this section we consider the converse problem.

We follow Silov in calling a Banach algebra R *homogeneous over* G if R satisfies the following conditions: R is a semi-simple regular commutative Banach algebra whose space of maximal regular ideals is a locally compact abelian group G, R is closed under translation, the norm in R is translation invariant, and the elements of R are continuous under translation in the norm of R (it is sufficient to assume that R contains a set of generators continuous under translation). Further, in case G is not compact, we assume that R is Tauberian in the sense that the elements with compact support are dense in R.

For $t_0 \in G$ let the corresponding maximal regular ideal be M_{t_0}. M_{t_0} contains a unique minimal closed primary ideal $J(t_0)$ which can be characterized as the closure of the set of all $f \in R$ such that $f(t) = 0$ in a neighborhood of t_0. (If R were not Tauberian the above f would have to be assumed in addition to have compact support.) Also, since R is Tauberian, it is easy to see that an element e with compact support for which $e(t) = 1$ for all t in a neighborhood of t_0 is a unit modulo $J(t_0)$.

Later in this section we will make use of the extensions to algebras without unit element of the theorems on regular commutative Banach algebras contained in [6, section 3]. As far as we know, some of these generalizations are not available in the literature (in particular, the results of sections 3.5-3.9 on algebras of type C). However, they are all routine, and under the Tauberian condition the facts mentioned above make Silov's proofs applicable almost without change.

If 0 is the identity element of G let $K = R/J(0)$. K is a commutative primary Banach algebra with identity and maximal ideal $Q = M_0/J(0)$. As before, denote the norm in K by $|\cdot|$.

LEMMA 2.1. *If R is a homogeneous algebra over the locally compct abelian group G then for all $s \in G$, $R/J(s)$ is isomorphic and isometric to $R/J(0) = K$, and R can be represented as an algebra of continuous K-valued functions on G vanishing at ∞.*

Proof. The isomorphism is $f + J(0) \rightarrow f_s + J(s)$. Clearly it is a homomorphism of K onto $R/J(s)$. It is an isomorphism since by definition $f \in J(0)$ implies $f_s \in J(s)$. By invariance of the norm in R under translation it is immediate that $\|f\|_0 = \|f_s\|_s$ where $\|g\|_t$ denotes the norm of the image of g in $R/J(t)$. For $f \in R$, $t \in G$ let $\tilde{f}(t)$ be the image of f under the mapping $R \to R/J(t) \to K$. The collection of functions $\tilde{f}(t)$ is the algebra isomorphic to R. Since $\|f\| \equiv \|\tilde{f}\| \geqq \sup |\tilde{f}(t)|$ $(t \in G)$, continuity of the functions $\tilde{f}$ follows from continuity of the elements of R under translation. Since R is Tauberian it is an easy exercise to show that each $\tilde{f}(t)$ vanishes at ∞.

LEMMA 2.2. *Let R be homogeneous over G and let R' be the set of all elements of R with compact support. Suppose R' is closed under multiplication by $\hat{G}$,* i.e., *for each $f \in R'$ and $\chi \in \hat{G}$ there exists an element $\chi f \in R'$ such that $\chi f(t) = \chi(t)f(t)$ for all t. Then*

(a) *R determines a homomorphism ω of $\hat{G}$ into the coset of* 1 *in $K = R/J(0)$ modulo $Q = M_0/J(0)$,*

(b) *for any $f \in R$ and any $\chi \in \hat{G}$ for which $g = \chi f$ exsists in R, $\tilde{g}(t) = \chi(t)\omega(\chi)\tilde{f}(t)$ for all $t \in G$, and*

(c) *if the mapping $\chi \to \chi f$ is continuous then ω is continuous.*

Proof. Pick $e \in R$ with compact support and with $e(t) = 1$ on a compact neighborhood C of 0. As we have remarked above, e is a unit modulo $J(0)$. For $\chi \in \hat{G}$ consider the element χe. If $\omega(\chi)$ denotes the image of χe in $R/J(0) = K$ the homomorphism is $\chi \to \omega(\chi)$. Clearly $\omega(\chi)(Q)=1$. Since, for $\chi_1, \chi_2 \in \hat{G}$, $[\chi_1\chi_2 e - \chi_1 e \cdot \chi_2 e](t) = 0$ in a neighborhood of 0 and outside a compact set we have $\omega(\chi_1\chi_2) = \omega(\chi_1)\omega(\chi_2)$. A similar argument shows that ω is independent of the choice of C and the choice of e. Let $h = \chi e$, then $h(t) = \chi(t)e(e) = \chi(s)\chi(t-s)e(t-s)$ provided both t and $t-s$ are in C. If s is in the interior of C then let U be a neighborhood of 0 such that $U \subset C$, $U + s \subset C$ then the above holds for all $t \in U + s$. Thus $h - \chi(s)h_s \in J(s)$ so via the mapping $R \to R/J(s) \to K$ we have $h \to h + J(s) = \chi(s)h_s + J(s) \to \chi(s)h + J(0) = \chi(s)\omega(\chi)$, this is, $\tilde{h}(s) = \chi(s)\omega(\chi)$ for s in the interior of C. The equality extends to all of C by continuity. Now let $g = \chi f$ for any $f \in R$ for which the product is defined. Fix $t_0 \in G$, let C be a compact neighborhood of 0 containing t_0 in its interior, and consider the corresponding e and $h = \chi e$. It follows easily that $\tilde{g}(t_0) = \tilde{g}\tilde{e}(t_0) = \tilde{h}\tilde{f}(t_0) = \chi(t_0)\omega(\chi)\tilde{f}(t_0)$. Part (c) is obvious.

Two Banach algebras R_1 and R_2 with the same structure space $\mathfrak{M}$ will be called *locally isomorphic* in case for each $t \in \mathfrak{M}$ there exist homeomorphic neighborhoods U_1 and U_2 of t such that every element of R_1 restricted to U_1 is carried by the homeomorphism into an element of R_2 restricted to U_2, and conversely. Two algebras of K-valued functions on G will be called *locally K-isomorphic* in case the analogous condition holds for the K-valued functions.

THEOREM 2.3. *Let R be a homogeneous Banach algebra over a* *connected locally compact abelian group. If R is of type C then R is closed under multiplication by $\hat{G}$. R can be represented as a closed subalgebra of $TK_\omega(G)$ where $K = R/J(0)$ and ω is the homomorphism given in Lemma 2.2. $TK_\omega(G)$ is semi-simple and R and $TK_\omega(G)$ are locally K-isomorphic. If ω is continuous then R and $TK_\omega(G)$ are locally isomorphic.*

Proof. Several remarks on Lemma 1.7 and its proof will produce a large part of the proof of the present theorem. In the first place, we know by Lemma 2.1 that R satisfies all the conditions of Lemma 1.7 except closure under multiplication by $\hat{G}$. This hypothesis is expendable, however. It was needed in the lemma only because we lacked the machinery for an intrinsic construction of the homomorphism ω. The proof of 1.7 shows (without using the hypothesis in question) the existence in $\bar{R}_p$ of a generating set X of characters which distinguish between points of G/D. Since the set $S(\alpha)$ is the structure space of $\bar{R}$ and $\chi(t) \neq 0$ for all t, it follows from standard Banach algebra theorems that with each $\chi \in X$ $\bar{R}$ contains its complex conjugate χ^{-1}. But the only subgroup of $(G/D)^\wedge$ which separates points of G/D is $(G/D)^\wedge$ itself (by Stone-Weierstrass and orthonormality of $(G/D)^\wedge$) so $\bar{R}$ contains all characters of G/D. Thus for any character χ which is identically 1 on D, R contains an element which is $\chi(t)$ on $S(\alpha)$. Furthermore, in the proof of 1.7 more general "rectangles"

$$S(\xi_1, \xi_2, \cdots, \xi_n) = \{(s, t) \in G \mid |\alpha_i| \leqq \xi_i, t \in G_c\} ,$$

with the obvious corresponding discrete subgroups D, could have been used in place of the sets $S(\alpha)$. Since $\hat{G} = E_n \times \hat{G}_c$ [2, 35A] it is clear that *any* $\chi \in \hat{G}$ is identically 1 on some such D. It follows that for any $\chi \in \hat{G}$ there exists a set $S(\xi_1, \xi_2, \cdots, \xi_n)$ such that R contains a sequence $\tilde{f}_k$, $k = 1, 2, \cdots$, with $f_k(t) = \chi(t)$ on $S(k\xi_1, k\xi_2, \cdots, k\xi_n)$. Since this latter collection of compact sets is a σ-covering of G we conclude that for any $\chi \in \hat{G}$ and compact set $C \subset G$ R contains an element which is $\chi(t)$ on C. Any element of R with compact support can therefore be multiplied by a character, so Lemma 2.2 applies and the homomorphism ω is defined. The second part of 2.2, together with the fact that R is of type C, implies that if $f_k \to f$, f_k with compact support, then $\{\chi f_k\}$ is Cauchy and $\chi f_k \to \chi f$. Thus R is closed under multiplication by $\hat{G}$. Conclusion (b) of 1.7 implies that R is a subalgebra of $TK_\omega(G)$. For it is clear that if $\{C_n\}$ is any σ-covering there exist discrete subgroups D_n such that the mapping $C_n \to G/D_n$ (compact) is one-to-one and Condition (A) holds for each pair C_n, D_n. If $f \in R$, $f = \lim f_n$, with the support of f_n contained in C_n, and if $\tilde{f}_n(t)$ is approximated to within

$1/n$ uniformly on C_n by a function $\tilde{f}^{(n)}$ of the form $\Sigma c_i \chi_i(t)\omega(\chi)$, then clearly f corresponds to the element $\{f^{(n)}\}$ of $K_\omega(G)$. Since R is Tauberian it is in $TK_\omega(G)$, and R is closed since its norm is the $K_\omega(G)$ norm. The local K-isomorphism and resulting semi-simplicity of $TK_\omega(G)$ follow from Lemma 2.2 and regularity of R, and the final conclusion follows from Theorem 1.5.

THEOREM 2.4. *Let R be a homogeneous Banach algebra of type C over the connected locally compact abelian group G with R closed under multiplication by $\hat{G}$. Suppose that for some σ-covering $\{C_n\}$ of G there exists a bounded sequence $\{e_n\}$ of elements of R with compact support such that $e_n(t) = 1$ on C_n. Then $R = TK_\omega(G) = K_\omega(G)$.*

Proof. By Theorem 2.3 we need only show that $R \supset K_\omega(G)$. Let $k = \sup \| e_n \|$ and suppose that $e_n(t)$ vanishes outside $C_{n'}$. Let $\{f^{(n)}\}$ be any ω-Cauchy sequence of linear combinations of characters defining an element of $K_\omega(G)$. Consider the sequence $\{f^{(n')}e_n\}$ in R. Choose $\varepsilon > 0$, then since $\{f^{(n)}\}$ is ω-Cauchy it follows from Lemma 2.2 that there exists a compact set C_ε such that for sufficiently large n $| f^{(n')}\tilde{e}_n(t) | < \varepsilon k$ for $t \notin C_\varepsilon$. It is also clear that if m and n are sufficiently large $(m > n)$ then $| f^{(n')}\tilde{e}_n(t) - f^{(m')}\tilde{e}_m(t) | < \varepsilon$ for $t \in C_n$. Thus, for sufficiently large m and n

$$\| f^{(n')}e_n - f^{(m')}e_m \| < \max(\varepsilon, 2k\varepsilon) ,$$

so $\{f^{(n')}e_n\}$ is Cauchy. Its limit is the element we seek.

3. In this section we exhibit three examples of algebras of the type discussed above.

(1) Let $G = E_1$ and $R = D_m(E_1)$ be the algebra of all complex functions f on E_1 which have m continuous derivatives all of which tend to zero together with f at ∞. $\| f \| = \sup \sum_{i=0}^{m} 1/i! \, | f^{(i)}(t) | \; (-\infty < t < \infty)$. It is easy to verify that $\mathfrak{M}(D_m) = E_1$ and that $J(t) = \{f \in D_m \,|\, f^{(i)}(t) = 0, i = 1, 2, \cdots, m\}$. D_m is locally isomorphic to $D_m[a, b]$, which is thoroughly discussed by Silov and to $D_m(C)$, C the circle group [6]. $D_m/J(t_0)$ is easily seen to be an $(m+1)$-dimensional "truncated" polynomial algebra generated by images of functions which are $(t - t_0)^k$, $k = 0, 1, \cdots, m$ in a neighborhood of t_0. D_m is of type C; indeed, the norm of f modulo $J(t_0)$ is exactly $\sum 1/i! \, | f^{(i)}(t_0) |$. It is also clearly closed under multiplication by $\hat{G}$. Since each $f^{(i)}(t) \to 0$ at ∞ it is uniformly continuous on E_1. Consequently, for each $f \in D_m$ $\| f - f_s \| \to 0$ as $s \to 0$. D_m is regular and Tauberian by easy proofs. Finally, it is clear that there exist $e_n \in D_m$ with $\| e_m \|$ *constant* and $e_n(t) = 1$ on $[-n, n]$, $e_n(t) = 0$ outside $[-n-1, n+1]$. This is true for any σ-covering of E_1 provided

that the distance between C_n and the complement of the support of e_n is bounded away from zero. Thus $D_m = K_\omega(G)$. Here $K = \{\alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_m x^m \mid \alpha_i \text{ complex}, x^{m+1} = 0\}$ and for $f \in D_m$ $\tilde{f}(t) = f(t) + f'(t)x + \cdots + (1/m!)f^{(m)}(t)x^m$. ω is given, then, by $e^{i\lambda t} \to 1 + i\lambda x + \cdots + (1/m!)(i\lambda)^m x^m$ and is clearly continuous.

(2) Let G be any direct product of copies of E_1 and the circle group C. One can define a wide variety of algebras on G analogous to $D_m(E_1)$. For the circle and torus examples have been discussed by Silov [6]. We illustrate by considering the algebra D_m^θ $(-\pi/2 \leqq \theta \leqq \pi/2)$ of all continuous functions on the cylinder $E_1 \times C$ which have m continuous directional derivatives in the direction making an angle θ from the generating circle C, all vanishing at ∞. D_m^θ can easily be seen to be homogeneous of type C over $E_1 \times C$ and to have a bounded set of units modulo a σ-covering of $E_1 \times C$. Thus $D_m^\theta = K_\omega(E_1 \times C)$. It is easily seen that K is the same $(m+1)$-dimensional algebra which occurred in (1) and that ω is given by

$$\omega[e^{i\lambda t}e^{int}] = 1 + \sum_{k=1}^{m} (1/k!)[(i\lambda)^k \cos\theta + (in)^k \sin\theta]x^k .$$

All D_m^θ, m fixed, are locally isomorphic. If we call a curve in $E_1 \times C$ which intersects each generating circle in a constant angle α an α-curve then it is clear that given non-zero $\alpha \neq \beta$ there is a homeomorphism of G onto itself sending each α-curve into a β-curve and each β-curve into an α-curve, but that no homeomorphism can send a $\pi/2$-curve into a 0-curve. From this it is easy to see that all D_m^θ, $\theta \neq 0$ are isomorphic to each other, but that D_m^0 is *not* isomorphic to D_m^θ, $\theta \neq 0$.

In the next example we introduce the C-completion R^c of a non-type C Banach algebra R, that is, the completion of R relative to the type C norm. The general situation is somewhat as follows: Silov has shown that if R^c is semi-simple then it is an algebra of type C, and he has examined the connections between R and R^c for regular commutative Banach algebras ([6] contains an account assuming an identity, and the results generalize easily to algebras without identity.). If R is a homogeneous algebra over a compact abelian group R^c is automatically a $K_\omega(G)$ and is therefore semi-simple. No such clear cut answers appear to be available in the non-compact case, but given various additional bits of information about R it is possible to obtain information about R^c from the results in § 2. The algebra of the next example is one for which such additional bits are available.

(3) Let G be a σ-compact abelian group, R the Banach algebra of Fourier transforms f of elements $\hat{f}$ of $L_1(\hat{G})$. If $f \in R$ with $f(t) = \int \hat{f}(\chi)\overline{\chi(t)}d\chi$ then for $\|f\|$ we use the L_1-norm of $\hat{f}$. Multiplication in R

is pointwise and R is isomorphic and isometric to $L_1(\hat{G})$ with convolution as multiplication. Several properties of R are immediate or well-known.

(a) G is the structure space of R and R is semi-simple, regular and Tauberian [2]. If $\hat{f} \in L_1(\hat{G})$ and $h \in G$ then the function $\chi(h)\hat{f}(\chi)$ is also in $L_1(\hat{G})$. But this function corresponds to the function $f_h(t) = f(t-h)$ in R so

(b) R is closed under translation. Clearly $\|f\| = \|f_h\|$. It is easy to verify that $\|f - f_h\|$ tends to 0 at $h = 0$ so

(c) the elements of R are continuous under translation. If $f \in R$ and $\chi_0 \in \hat{G}$ then $\chi_0(t)f(t)$ is the Fourier transform of the translate $\hat{f}_{\chi_0} \in L_1(\hat{G})$, so

(d) R is closed under multiplication by $\hat{G}$. Moreover, by a well-known theorem on the Haar integral, if $\hat{f}$ and $\tilde{e}$ are in $L_1(\hat{G})$ then $\hat{f} * \hat{e}$ can be L_1-approximated by linear combinations of translates of $\hat{e}$. In R this means that

(e) Re is generated by $\hat{G}e$. Finally

(f) $\|\chi f\| = \|f\|$ for all $f \in R$, $\chi \in \hat{G}$ by an easy proof. From properties (a)-(d) it can easily be seen that R^c satisfies all the conditions of Lemmas 2.1 and 2.2 with the possible exception of semi-simplicity. The fact that for any unit e modulo a compact set of G Re is generated by $\hat{G}e$ enables one to show directly that $R^c \subset TK_\omega(G)$; the type C condition on R and the connectivity condition on G were used in Theorem 2.3 essentially to establish property (e). Property (f) (or, more generally, $\|\chi^{\pm n}f\| = o(n)$ for all χ, f) implies that $TK_\omega(G) = C_0(G)$. For if e is such that $\chi e \to \omega(\chi)$ in K then $|\omega(\chi)| \leqq \|\chi e\|$. Thus $|\omega(\chi^{\pm n})| = |\omega(\chi)^{\pm n}|$ is $o(n)$ and a theorem of Gelfond-Hille [1, p. 715] shows that this implies in a primary algebra that $\omega(\chi) = 1$. Example 1 of § 1 completes the proof. Thus R^c is semi-simple, hence homogeneous of type C so by Theorem 2.3 R^c is locally isomorphic to $C_0(G)$. By theorems of Silov [6; 3.5, 3.9] extended to algebras without identity $R/J(0)$ is isomorphic to the corresponding difference algebra in $C_0(G)$, but this is the complex field. Thus $J(0)$ and hence each $J(t)$ is maximal. This provides a proof of the well known theorem (first proved by Beurling and Segal for the real line and then by Kaplansky in general) which says that in the group algebra of a locally compact abelian group closed primary ideals are maximal. Finally, if G is connected then $R^c = C_0(R)$. For R contains elements with compact support for which $f(t) = 1$ on a compact set and $0 \leqq f(t) \leqq 1$ for all t. Since $J(t) = M_t$ this says that the type C norm of f is 1, so Theorem 2.4 applies to R^c.

REFERENCES

1. E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc. Colloquium publication, **31** (1957).

2. L. H. Loomis, *An introduction to abstract harmonic analysis*, New York (1953).

3. G. W. Mackey, *Commutative Banach algebras*, (edited by A. Blair), Harvard University (1952), (mimeographed lecture notes).

4. H. Mirkil, *The work of Silov on commutative semi-simple Banach algebras*, University of Chicago (1952), (mimeographed).

5. ———, *Translation-invariant function algebras on compact groups*, University of Chicago thesis (1957).

6. G. E. Silov, *Homogeneous rings of functions*, Uspehi Math. Nauk, (N.S.) 6, No. 1 (41), 91-137, (1951) in Russian; Amer. Math. Soc. Transl. No. 92.

7. A. Weil, *L'Integration dans les groupes topologiques*, Paris (1940).

8. A. B. Willcox, *Some structure theorems for a class of Banach algebras*, Pacific J. Math. **6** (1956), 177-192.

AMHERST COLLEGE

CORRECTON TO "EQUIVALENCE AND PERPENDICULARITY OF GAUSSIAN PROCESSES"

J. FELDMAN

It has been kindly pointed out to me by D. Lowdenslager that, as it stands, the argument in [1] only works when $\boldsymbol{L}_2(\mu)$ and $\boldsymbol{L}_2(\nu)$ are separable. In particular, the theorem of von Neumann from [2], which is used there, only holds in separable Hilbert spaces. Our theorem nevertheless holds in the non-separable case; an argument will be supplied here enabling one to go from the separable to the general case. We retain notation and terminology of [1].

For any *countable* subset $\boldsymbol{C}$ of $\boldsymbol{L}$, let $\mathscr{S}_C$ be the σ-subalgebra of $\mathscr{S}$ generated by $\boldsymbol{C}$, $\boldsymbol{L}_C$ the linear subspace of $\boldsymbol{L}$ spanned by $\boldsymbol{C}$, and μ_C, ν_C the restrictions of μ, ν to $\mathscr{S}_C$. $\bigcup_C \mathscr{S}_C$ is a σ-algebra contained in $\mathscr{S}$, and, since each $x \in \boldsymbol{L}$ is in some $\boldsymbol{L}_C$, each x in $\boldsymbol{L}$ is measurable with respect to $\bigcup_C \mathscr{S}_C$. Therefore $\mathscr{S} = \bigcup_C \mathscr{S}_C$. Now, suppose, under the assumptions of the theorem of [1], that μ and ν are not equivalent. Then there is some set in $\mathscr{S}$ with μ-measure 0 and ν-measure > 0 (or vice versa). This set is in some $\mathscr{S}_C$. So μ_C and ν_C are not equivalent. By the separable case of the theorem, they are mutually perpendicular, i.e., there is some set in $\mathscr{S}_C$ with μ-measure 0 and ν-measure 1. Thus μ and ν are mutually perpendicular.

Next we show that $\mu \sim \nu$ implies that the correspondence $x^\nu \xrightarrow{T} x^\mu$ between equivalence classes of functions has the property that T extends to an equivalence operator between the linear subspaces $\bar{\boldsymbol{L}}_\mu$ and $\bar{\boldsymbol{L}}_\nu$ of $\boldsymbol{L}_2(\mu)$, $\boldsymbol{L}_2(\nu)$ generated by $\boldsymbol{L}$. Assume, then, that $\mu \sim \nu$. By using the separable case, we easily see that T and T^{-1} are bounded. An argument on p. 704 of [1] still works to show that the extension of T to an operator from $\bar{\boldsymbol{L}}_\mu$ onto $\bar{\boldsymbol{L}}_\nu$ still has the property that, given ξ in $\bar{\boldsymbol{L}}_\mu$, there is an $\mathscr{S}$-measurable x such that $x^\mu = \xi$ and $x^\nu = T\xi$. Write $T^* T$ as $\int \lambda \, d\,\mathrm{F}(\lambda)$. Let $E_n = F\left(1 + \frac{1}{n}\right) - F\left(1 - \frac{1}{n}\right)$, $n = 2, 3, 4, \cdots$ Let $E = \bigcap_n E_n$. I now assert $(I - E)\,\bar{\boldsymbol{L}}_\mu$ is separable. For otherwise $(I - E_n)\,\bar{\boldsymbol{L}}_\mu$ would be inseparable for some n, and one could therefore find a countable orthonormal infinite set $\xi_1, \xi_2, \cdots$ of elements of $\bar{\boldsymbol{L}}_\mu$ for which $\|(T^* T - I)\,\xi_i\| \geqq \frac{1}{n}\|\xi_i\|$, all i. Let $\boldsymbol{H}$ be the Hilbert space spanned by the ξ_i. Let $\tilde{\boldsymbol{L}}$ be the set of μ-measurable functions x on S such that $x^\mu \in \boldsymbol{H}$. Let $\tilde{\mathscr{S}}$ be the σ-algebra spanned by them. Let $\tilde{\mu}$, $\tilde{\nu}$ be the completions of μ and ν, restricted to $\tilde{\mathscr{S}}$. Then the Hilbert spaces $\bar{\tilde{\boldsymbol{L}}}_{\tilde{\mu}}$, $\bar{\tilde{\boldsymbol{L}}}_{\tilde{\nu}}$ are isometric to $\boldsymbol{H}$ and $T(\boldsymbol{H})$, respectively, in a natural way. Therefore they are separable, and, since $\tilde{\mu} \sim \tilde{\nu}$, the operator $\tilde{T}$ induced by the correspondence $\tilde{x}^{\mu} \longrightarrow \tilde{x}^{\nu}$ is an equivalence operator. But T is unitarily equivalent to $T|\boldsymbol{H}$, and $T|\boldsymbol{H}$ was constructed so as *not* to be an equivalence operator, giving a contradiction.

To show T is an equivalence operator, it suffices to show this for $T|(I - E)\ \bar{\boldsymbol{L}}_{\mu}$. Since $(I - E)\ \bar{\boldsymbol{L}}_{\mu}$ is separable, we can reduce to the separable case exactly as in the last five sentences of the previous paragraph, with $(I - E)\ \bar{\boldsymbol{L}}_{\mu}$ playing the role played there by $\boldsymbol{H}$ to show that T is an equivalence operator.

Finally, suppose that, for $x \in \boldsymbol{L}$, $x^{\mu} = 0 \Longleftrightarrow x^{\nu} = 0$, and that the one-to-one operator T from $\boldsymbol{L}_{\mu}$ to $\boldsymbol{L}_{\nu}$ induced thereby extends to an equivalence operator from $\bar{\boldsymbol{L}}_{\mu}$ to $\bar{\boldsymbol{L}}_{\nu}$. It must be shown that $\mu \sim \nu$. If μ is not equivalent to ν, then as shown in the first paragraph (and using the notation established there) there is some countable subset $\boldsymbol{C}$ of $\boldsymbol{L}$ such that μ_C and ν_C are not equivalent. But the operator T_C induced by sending x^{μ} to x^{ν} for $x \in \boldsymbol{L}_C$ is precisely the restriction of T to those elements in $\boldsymbol{L}_{\mu}$ which come from $\boldsymbol{L}_C$. Now, the restriction of T to a subspace is again an equivalence operator, so T_C extends to an equivalence operator from $(\overline{\boldsymbol{L}_C})_{\mu}$ to $(\overline{\boldsymbol{L}_C})_{\nu}$, which contradicts the separable case of the theorem.

Also, in reviewing [1], E. Nelson noticed that Lemma 1 is misstated. It should read "positive" instead of "self-adjoint," and, in (b), "$A^2 - I$" rather than "$(A - I)^2$."

Bibliography

1. J. Feldman, *Equivalence and perpendicularity of Gaussian processes,* Pacific J. Math., Vol. 8 No. 4, 1958.

2. J. von Neumann, *Charakterisierung des spektrums eines integral-operatoren,* Actualites Sci. Ind. 229, Paris, 1935.

Pacific Journal of Mathematics
Vol. 9, No. 4 August, 1959