# Pacific Journal of Mathematics

## IN THIS ISSUE—




**Vol. 9, No. 1** **May, 1959**

# PACIFIC JOURNAL OF MATHEMATICS




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

# ON THE STRUCTURE OF INFINITELY DIVISIBLE DISTRIBUTIONS

J. R. Blum and M. Rosenblatt

**1. Introduction and summary.** Let $F(x)$ be a distribution on the real line. Then we may write

$$F(x) = pF_1(x) + (1-p)F_2(x) \tag{1.1}$$

where $F_1(x)$ is a discrete distribution, $F_2(x)$ is a continuous distribution and $0 \leq p \leq 1$. We shall say that $F(x)$ is discrete if $p = 1$, $F(x)$ is continuous if $p = 0$ and $F(x)$ is a mixture if $0 < p < 1$.

Let $\varphi(s) = \int_{-\infty}^{\infty} e^{isx} dF(x)$ be the characteristic function corresponding to $F(x)$. It would be useful to give a convenient criterion on $\varphi(s)$ to determine when the corresponding distribution $F(x)$ is discrete, continuous, or a mixture. In § 2 we give such a criterion for the class of infinitely divisible (i.d.) distributions, utilizing the Khinchin representation of the characteristic function of such a distribution. In § 3 we apply the theorem of § 2 to characterize a certain class of stochastic processes.

**2. The structure theorem.** Let $\varphi(s)$ be the characteristic function of an i.d. distribution. The Khinchin representation of such a characteristic function takes the form

$$\varphi(s) = \exp\left\{i\gamma s + \int_{-\infty}^{\infty}\left[e^{ius} - 1 - \frac{ius}{1+u^2}\right]\frac{1+u^2}{u^2}dG(u)\right\} \tag{2.1}$$

where $\gamma$ is a real number and $G(u)$ is a real valued bounded nondecreasing function. $\gamma$ and $G(u)$ are uniquely determined by the conditions $G(-\infty) = 0$, $G(u+0) = G(u)$. We shall need the following two lemmas, the first of which is well known.

LEMMA 1. *Let $X$ and $Y$ be independent random variables. Then*
(i) *the distribution of $X + Y$ is discrete if and only if the distribution of each of the variables is discrete,*
(ii) *the distribution of $X + Y$ is a mixture if and only if one of the two distributions is a mixture and the other is either discrete or a mixture.*

Let $F(x)$ be a distribution. We shall define $F^{(k)}(x)$ as follows:

Received July 16, 1958. This research was sponsored by the Office of Ordnance Research, U. S. Army and the Office of Naval Research, U. S. Navy. Reproduction in whole or in part is permitted for any purpose of the U. S. Government.

$$F^{(0)}(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}, \quad F^{(1)}(x) = F(x) ,$$

and for $k \geq 2$, $F^{(k)}(x)$ denotes the $k$-fold convolution of $F(x)$ with itself.

LEMMA 2. *Let* $-\infty \leq a < b \leq \infty$, *and let* $F(x)$ *be a nondecreasing, bounded function defined for* $a \leq x \leq b$ *such that* $F(a) = 0$, $F(b) - F(a) = c > 0$. *Then*

$$\varphi(s) = \exp\left\{-c + \int_a^b e^{isx} dF(x)\right\} \tag{2.2}$$

*is a characteristic function corresponding to the* i.d. *distribution*

$$H(u) = e^{-c} \sum_{k=0}^{\infty} \frac{F^{(k)}(u)}{k\,!} . \tag{2.3}$$

*If* $F(x)$ *is a pure jump function then* $H(u)$ *is discrete. If* $F(x)$ *is continuous, then* $H(u)$ *is a mixture with a jump of magnitude* $e^{-c}$ *at the origin and continuous otherwise.*

*Proof.* For every positive integer $n$ let

$$H_n(u) = \sum_{k=0}^{n} \frac{F^{(k)}(u)}{k\,!} \Big/ \sum_{k=0}^{n} \frac{c^k}{k\,!}$$

and let

$$\varphi_n(s) = \sum_{k=0}^{n} \frac{1}{k\,!} \left[\int_a^b e^{isk}\, dF(x)\right]^k \Big/ \sum_{k=0}^{n} \frac{c^k}{k\,!} .$$

Then $H_n(u)$ is a distribution with characteristic function $\varphi_n(s)$. Since $H_n(u)$ converges to $H(u)$ and $\varphi(s)$ converges to the continuous function $\varphi(s)$ it follows that $H(u)$ is a distribution with characteristic function $\varphi(s)$. The fact that $H(u)$ is i.d. is immediate from the form of $\varphi(s)$. Now if $F(x)$ is a pure jump function then (2.2) becomes

$$\varphi(s) = e^{-c} \sum_{k=0}^{\infty} \frac{1}{k\,!} \left[\sum_{j=1}^{\infty} e^{isx_j} p_j\right]^k$$

where $F(x)$ has its jumps at the points $x_j$ with magnitudes $p_j$, and such a characteristic function clearly corresponds to a discrete distribution. Finally if $F(x)$ is continuous we may write

$$H(u) = e^{-c} F^{(0)}(u) + \frac{(1 - e^{-c})}{(e^c - 1)} \sum_{k=1}^{\infty} \frac{F^{(k)}(u)}{k\,!} ,$$

and since the infinite series converges uniformly it follows that $H(u)$ is the mixture of a continuous distribution and the distribution with a single jump at zero.

THEOREM 1. *Let $\varphi(s)$ be the characteristic function of an* i.d. *distribution $F(x)$. Let $G(u)$ be the function occurring in the representation* (2.1). *Then*

(i) *$F(x)$ is discrete if and only if $\int_{-\infty}^{\infty} \frac{1}{u^2}\, dG(u) < \infty$ and $G(u)$ is a pure jump function.*

(ii) *$F(x)$ is a mixture if and only if $\int_{-\infty}^{\infty} \frac{1}{u^2}\, dG(u) < \infty$ and $G(u)$ is not a pure jump function*

(iii) *$F(x)$ is continuous if and only if $\int_{-\infty}^{\infty} \frac{1}{u^2}\, dG(u) = \infty$ .*

*Proof.* Suppose first that $G(u)$ is a pure jump function with jumps at the points $u_j, j = 1, 2, \cdots$ and with corresponding magnitudes $\rho_j \geq 0$, such that $\sum_j \rho_j < \infty$. Then (2.1) (with $\gamma = 0$) takes the form

$$\varphi(s) = \exp\left\{\sum_j \left[e^{isu_j} - 1 - \frac{isu_j}{1 + u_j^2}\right]\frac{1 + u_j^2}{u_j^2}\rho_j\right\} . \tag{2.4}$$

Now if $\sum_j \rho_j / u_j^2 < \infty$ we may rewrite (2.4) in the form

$$\varphi(s) = \exp\left\{isb - c + \int_{-\infty}^{\infty} e^{isu} dM(u)\right\}$$

where

$$b = -\sum_j \frac{\rho_j}{u_j},\ c = \sum_j \frac{1 + u_j^2}{u_j^2}\rho_j ,$$

and where $M(u)$ is a bounded, nondecreasing, pure jump function with jumps at the points $u_j$ and corresponding magnitudes $((1 + u_j^2)/u_j^2)\rho_j$. Consequently it follows from Lemmas 1 and 2 that $F(x)$ is discrete.

Conversely we suppose that $F(x)$ is a discrete distribution. We shall show first that $G(u)$ is a pure jump function. To do this write $G(u) = G_1(u) + G_2(u)$ where $G_1(u)$ is a pure jump function and $G_2(u)$ is continuous. If $G(u)$ is not a pure jump function there will exist a closed interval $[a, b]$ not containing zero such that $G_2(a) < G_2(b)$. Then we may write $\varphi(s)$ in the form $\varphi(s) = M(s)N(s)$ where $M(s)$ is a characteristic function and

$$\begin{aligned} N(s) &= \exp\left\{\int_a^b\left[e^{isu} - 1 - \frac{isu}{1 + u^2}\right]\frac{1 + u^2}{u^2} dG_2(u)\right\} \\ &= \exp\left\{- is\int_a^b \frac{1}{u} dG_2(u) - \int_a^b \frac{1 + u^2}{u^2} dG_2(u) + \int_a^b e^{isu} dH(u)\right\} \end{aligned}$$

where $dH(u) = ((1 + u^2)/u^2)dG_2(u)$. From Lemma 2 it follows that $N(s)$ is the characteristic function of a mixture and from Lemma 1 it then follows that $F(x)$ is not discrete. Hence $G(u)$ is a pure jump function, and $\varphi(s)$ has the form (2.4).

We shall show that $\sum_j \rho_j/u_j^2 < \infty$. Since $\sum_j \rho_j < \infty$ it is sufficient to restrict attention to those $u_j$ for which $|u_j| \leq 1$. Since $F(x)$ is discrete it follows that $\varphi(s)$ is almost periodic and we have

$$\lim_{R\to\infty} \frac{1}{R}\int_0^R |\varphi(s)|^2 ds > 0 . \tag{2.5}$$

Now

$$|\varphi(s)|^2 = \exp\left\{\sum_j [\cos u_j s - \lambda 1]\frac{\lambda_j}{u_j^2}\right\}$$

where

$$\lambda_j = 2[1 + u_j^2]\rho_j .$$

Let

$$g(R) = \sum_{1/R \leq |u_j| \leq 1} \frac{\lambda_j}{u_j^2} .$$

We have

$$\begin{aligned} |\varphi(s)|^2 &\leq \exp\left\{\sum_{1/R \leq |u_j| \leq 1} [\cos u_j s - ]\frac{\lambda_j}{u_j^2} g(R)/g(R)\right\} \\ &\leq \frac{1}{g(R)} \sum_{1/R \leq |u_j| \leq 1} \frac{\lambda_j}{u_j^2} \exp\{[\cos u_j s - 1]g(R)\} . \end{aligned} \tag{2.6}$$

The first of these inequalities is immediate and the second is an application of Jensen's inequality.

From (2.6) we obtain

$$\begin{aligned} &\frac{1}{R}\int_0^R |\varphi(s)|^2 ds \\ &\leq \frac{1}{g(R)} \sum_{1/R \leq |u_j| \leq 1} \frac{\lambda_j}{u_j^2} \frac{1}{R}\int_0^R \exp\{[\cos u_j s - 1]g(R)\} ds . \end{aligned} \tag{2.7}$$

Suppose $R \geq 1$ and $|u_j| \geq 1/R$. Then for every $\varepsilon > 0$ there exists $\delta$ depending on $\varepsilon$ only with $0 < \delta < 1$ and with the following property: If $R_1(\varepsilon)$ is the subset of $[0, R]$ where $\cos u_j s < 1 - \delta$ and $R_2(\varepsilon)$ is the subset of $[0, R]$ where $\cos u_j s \geq 1 - \delta$, then the measure of $R_2(\varepsilon)$ does not exceed $\varepsilon R$. Using this and (2.7) we find

$$\frac{1}{R}\int_0^R |\varphi(s)|^2 ds \leq \varepsilon + e^{-\delta g(R)} . \tag{2.8}$$

Now if $\sum_j \rho_j/u_j^2 = \infty$, then clearly $\lim_{R\to\infty} g(R) = \infty$. This together with (2.8) contradicts (2.5), thus proving (i).

Now suppose $\int_{-\infty}^{\infty} 1/u^2 dG(u) < \infty$ and $G(u)$ is not a pure jump function. Then we may write $G(u) = G_1(u) + G_2(u)$ where $G_1(u)$ is a pure jump function and $G_2(u)$ is continuous. Of course we have

$$\int_{-\infty}^{\infty} \frac{1}{u^2} dG_i(u) < \infty \,, \qquad i = 1, 2 \,.$$

Then from (i)

$$\exp\left\{\int_{-\infty}^{\infty}\left[e^{isu} - 1 - \frac{isu}{1+u^2}\right]\frac{1+u^2}{u^2}\, dG_1(u)\right\}$$

is the characteristic function of a discrete distribution. Similarly from Lemma 2 it follows that

$$\exp\left\{\int_{-\infty}^{\infty}\left[e^{isu} - 1 - \frac{isu}{1+u^2}\right]\frac{1+u^2}{u^2}\, dG_2(u)\right\}$$

is the characteristic function of a mixture. Thus $F(x)$ is the convolution of a discrete distribution and a mixture and from Lemma 1 it follows that $F(x)$ is a mixture.

Conversely suppose $F(x)$ is a mixture. Then

$$\varphi(s) = p\varphi_1(s) + (1-p)\varphi_2(s) \tag{2.9}$$

where $0 < p < 1$, $\varphi_1(s)$ is the characteristic function of a discrete distribution and $\varphi_2(s)$ is the characteristic function of a continuous distribution. If we write $\varphi(s) = e^{\psi(s)}$ then $e^{\psi(s)/n}$ is a characteristic function for every positive integer $n$ because $F(x)$ is infinitely divisible. Clearly $e^{\psi(s)/n}$ must be the characteristic function of a mixture, i.e.

$$e^{\psi(s)/n} = p_n\varphi_{1,n}(s) + (1-p_n)\varphi_{2,n}(r) \tag{2.10}$$

where $0 < p_n < 1$, and $\varphi_{1,n}(s)$ and $\varphi_{2,n}(s)$ are of the same type as $\varphi_1(s)$ and $\varphi_2(s)$ respectively. From (2.9) and (2.10) we obtain

$$\varphi(s) = \left[e^{\frac{\psi(s)}{n}}\right]^n = p_n^n\varphi_{1,n}^n(s) + \sum_{k=1}^{n}\binom{n}{k}p_n^{n-k}(1-p_n)^k\varphi_{1,n}^{n-k}(s)\varphi_{2,n}^k(s)\,. \tag{2.11}$$

Now $\varphi_{1,n}^n(s)$ is the characteristic function of a discrete distribution and the sum occurring in (2.11) is the product of $(1-p_n^n)$ and a characteristic function of a continuous distribution. Thus $p_n = p^{1/n}$ and $[\varphi_{1,n}(s)]^n = \varphi_1(s)$ and we see that $\varphi_1(s)$ is the characteristic function of an i.d. distribution. Writing $\varphi_1(s) = e^{\psi_1(s)}$, $p = e^{-c}$ with $0 < c < \infty$ we have

$$e^{\frac{\psi(s)}{n}} = e^{-\frac{c}{n}} e^{\frac{\psi_1(s)}{n}} + (1 - e^{-\frac{c}{n}})\varphi_{2,n}(s) . \tag{2.12}$$

If we expand the exponentials in (2.12) we obtain

$$\lim_{n\to\infty} \varphi_{2,n}(s) = \varphi_{2,0}(s) = 1 + \frac{\psi(s) - \psi_1(s)}{c} . \tag{2.13}$$

Since $\psi(s)$ and $\psi_1(s)$ are continuous it follows that $\varphi_{2,0}(s)$ is a characteristic function, say $\varphi_{2,0}(s) = \int_{-\infty}^{\infty} e^{isx}dH(x)$, where $H(x)$ is a distribution.

Hence

$$\begin{aligned} \varphi(s) = e^{\psi(s)} &= e^{\psi_1(s)+\psi(s)-\psi_1(s)} \\ &= e^{\psi_1(s)+c[\varphi_{2,0}(s)-1]} = e^{\psi_1(s)+\int_{\infty}^{\infty}[e^{isu}-1]dcH(u)}. \end{aligned} \tag{2.14}$$

Now $e^{\psi_1(s)}$ is the characteristic function of a discrete distribution. If we equate formula (2.14) for $\varphi(s)$ with formula (2.1) for $\varphi(s)$ it follows from the first part of the theorem and the uniqueness of $G(u)$ that $\int_{-\infty}^{\infty} 1/u^2 dG(u) < \infty$. It is also a consequence of the first part of the theorem that $G(u)$ is not a pure jump function. Thus (ii) is proved and (iii) follows from (i) and (ii), proving the theorem.

From (2.14) we are able to deduce additional information in the mixed case.

COROLLARY. *Let $\varphi(s)$ be a characteristic function corresponding the* i.d. *distribution $F(x)$. If $F(x)$ is a mixture then $F(x)$ is the convolution of a discrete* i.d. *distribution and a* i.d. *distribution which has a jump at zero of magnitude less than one and is continuous otherwise.*

**3. A class of discrete processes.** Let $X_j(t), t \geq 0, j = 1, 2 \cdots$ be a sequence of independent stochastic processes such that for each $j$, $X_j(t)$ is a process with independent increments and such that for $0 \leq t_1 \leq t_2$ the random variable $X_j(t_2) - X_j(t_1)$ has characteristic function

$$\varphi_j(s, t_1, t_2) = \exp\left\{\left[e^{isu_j} - 1 - \frac{isu_j}{1 + u_j^2}\right] \frac{1 + u_j^2}{u_j^2}[\rho_j(t_2) - \rho_j(t_1)]\right\}$$

where $u_j$ is a real number and $\rho_j(t)$ is a nondecreasing function defined for $t \geq 0$ with $\rho_j(0) = 0$. Then each $X_j(t)$ is a generalized Poisson process, i.e. $X_j(t)$ assumes values of the form $y_k = ku_j - (\rho_j(t))/u_j$ with probability

$$P\{X_j(t) = y_k\} = \frac{e^{-\lambda_j(t)}\lambda_j^k(t)}{k!} ,$$

where $\lambda_j(t) = ((1 + u_j^2)/u_j^2)\rho_j(t)$. Now if $\sum_j \rho_j(t) < \infty$ for every $t \geq 0$, then we can define a process $X(t)$ as the sum of the processes $X_j(t)$, and the characteristic function of the process $X(t)$ will have the form

$$\varphi(s, t) = \exp\left\{\sum_j \left[e^{isu_j} - 1 - \frac{isu_j}{1 + u_j^2}\right]\frac{1 + u_j^2}{u_j^2}\rho_j(t)\right\} . \tag{3.1}$$

It is an immediate consequence of Theorem 1 that for any $t \geq 0, X(t)$ will be a discrete random variable if and only if $\sum_j (\rho_j(t))/u_j^2 < \infty$ .

Conversely suppose for $t \geq 0, X(t)$ is a stochastic process such that $X(0) = 0, X(t)$ is a discrete random variable for every $t \geq 0$, and the process has independent infinitely divisible increments. This will be true, e.g. if $X(t)$ is a discrete process with independent increments and such that $X(t)$ is continuous in probability. Then from Theorem 1 it follows that the characteristic function of the random variable $X(t)$ is essentially of the form (3.1) with $\rho_j(t)$ nondecreasing and $\sum_j (\rho_j(t))/u_j^2 < \infty$ for all $t$. Consequently $X(t)$ has the stochastic structure of a sum of independent generalized Poisson processes. We have

THEOREM 2. *Let $X(t)$ be a discrete stochastic process for $t \geq 0$, with $X(0) = 0$ and such that $X(t)$ has independent infinitely divisible increments. Then there exists a sequence of independent generalized Poisson processes $X_j(t), j = 1, 2 \cdots$ such that $X(t)$ has the same stochastic structure as $\sum_j X_j(t)$.*

In the case when $X(t)$ assumes only integer values Theorem 2 was already proved by Khinchin [1].

## REFERENCE

1. A Khinchin, *Nachwirkungsfreie Folgen von zufälligen Ereignissen,* Teor. Veroyatnost. i **1**, (1956).

INDIANA UNIVERSITY

# ASYMPTOTIC EXPRESSIONS FOR $\sum n^a f(n) \log^r n$

R. G. BUSCHMAN

In this paper some asymptotic expressions for sums of the type

$$\sum n^a f(n) \log^r n ,$$

where $f(n)$ is a number theoretic function, are presented. (The summations extend over $1 \leq n \leq x$ unless otherwise noted.) The method applied is to obtain the Laplace transformation,

$$\mathscr{L}\{F(t)\} = \int_0^\infty e^{-st}F(t)dt = f(s)$$

of the sum and then use a Tauberian theorem either from Doetsch [2] or its modification for a pole at points other than the origin, or from Delange [1] to obtain the asymptotic relation. If $f(n)$ is non-negative, then $F(t)$ is a non-negative, non-decreasing function and hence satisfies the conditions for the Tauberian theorems. In many cases the closed form of a Dirichlet series involving the functions are known, and in this case the relation

$$\mathscr{L}\left\{\sum_{1\leqq n \leqq e^t} n^a f(n) \log^r n\right\} = (-1)^r s^{-1}(d/ds)^r \sum_1^\infty n^{a-s} f(n)$$

can be used. The functions chosen for discussion and the Dirichlet series involving them can be found in Hardy and Wright [3], Landau [4], [5], or Titchmarsh [7]. We present first a few illustrations of the method and then a more extensive collection of results is presented at the end in a table.

First we choose $\sigma_k(n)$ as an example of a simpler type. Since

$$\sum_1^\infty n^{-s}\sigma_k(n) = \zeta(s)\zeta(s-k) ,$$

we have

$$\mathscr{L}\left\{\sum_{1\leqq n\leqq e^t} n^{b-1-k}\sigma_k(n)\log^r n\right\} = f(s) = (-1)^r s^{-1}(d/ds)^r\{\zeta(s+1-b)\zeta(s+1-b+k)\}.$$

For $k > 0$ the pole where $\Re s$ is greatest is at $s = b$ if $b \geq 0$. At that pole, since

$$\zeta^{(m)}(s+1-b) \sim (-1)^m m\,!\,(s-b)^{-m-1} ,$$

the Laplace transformation of the sum has the form

Received October 10, 1958.

$$f(s) \sim b^{-1}\zeta(1+k)r\,!\,(s-b)^{-r-1}\,.$$

Now if $b>0$, then by modifying Doetsch [2, p. 517] for poles not at the origin or from Delange [1, p. 235] we obtain

$$\sum_{1\leq n\leq e^t} n^{b-1-k}\sigma_k(n)\log^r n \sim b^{-1}\zeta(1+k)e^{bt}t^r\,,$$

or, if $x=e^t$

$$\sum n^{b-1-k}\sigma_k(n)\log^r n \sim b^{-1}\zeta(1+k)x^b\log^r x\,.$$

If $b=0$, then

$$f(s) \sim \zeta(1+k)r\,!\,s^{-r-2}\,,$$

so that form Doetsch [2, p. 517] after substituting $x=e^t$ we obtain

$$\sum n^{-1-k}\sigma_k(n)\log^r n \sim (r+1)^{-1}\zeta(1+k)\log^{r+1}x\,.$$

The expressions for $\sigma(n)$ can be obtained by setting $k=1$.

For $k=0$, $\sigma_k(n)$ becomes $d(n)$ which will be covered as a special case of $d_k(n)$.

For $k<0$ the pole where $\Re s$ is greatest is at $s=b-k$ so that for $b>k$

$$f(s) \sim (b-k)^{-1}\zeta(1-k)r\,!\,(s-b+k)^{-r-1}\,.$$

Hence

$$\sum n^{b-1-k}\sigma_k(n)\log^r n \sim (b-k)^{-1}\zeta(1-k)x^{b-k}\log^r x\,, \qquad \text{for } b>k\,;$$
$$\sum n^{-1}\sigma_k(n)\log^r n \sim (r+1)^{-1}\zeta(1-k)\log^{r+1}x\,, \qquad \text{for } b=k\,.$$

By analogy, since

$$\sum_1^\infty n^{-s}\phi(n) = \zeta(s-1)/\zeta(s)\,,$$

then

$$\sum n^{b-2}\phi(n)\log^r n \sim \{b\zeta(2)\}^{-1}x^b\log^r x\,, \qquad \text{for } b>0\,;$$
$$\sum n^{-2}\phi(n)\log^r n \sim \{(r+1)\zeta(2)\}^{-1}\log^{r+1}x\,, \qquad \text{for } b=0\,.$$

If $\chi_k(n)$ represents a character, mod $k$, then the Dirichlet series can be represented by

$$\sum_1^\infty n^{-s}\chi_k(n) = L_k(s)$$

so that if $\chi_k$ is a principal character then $L_k(s)$ has a pole at $s=1$ and

$$\sum n^{b-1}\chi_k(n)\log^r n \sim \phi(k)(kb)^{-1}x^b\log^r x\,, \qquad \text{for } b>0\,;$$
$$\sum n^{-1}\chi_k(n)\log^r n \sim \phi(k)\{(r+1)b\}^{-1}\log^{r+1}x\,, \qquad \text{for } b=0\,.$$

The Dirichlet series involving $d_k(n)$ yields a power of the $\zeta$-function, i.e.

$$\sum_1^\infty n^{-s} d_k(n) = \zeta^k(s) ,$$

so that for $k > 0$

$$\mathscr{L}\Big\{ \sum_{1 \leq n \leq e^t} n^{b-1} d_k(n) \log^r n \Big\} = (-1)^r s^{-1} (d/ds)^r \zeta^k(s+1-b) .$$

Now the Laplace transform can be written to show the behavior at the pole at $s = b$,

$$f(s) \sim (r+k-1)! \, \{b(k-1)!\}^{-1} (s-b)^{-r-k} .$$

Thus

$$\sum n^{b-1} d_k(n) \log^r n \sim \{b(k-1)!\}^{-1} x^b \log^{r+k-1} x , \qquad \text{for } b > 0 ;$$
$$\sum n^{-1} d_k(n) \log^r n \sim \{(r+k)(k-1)!\}^{-1} \log^{r+k} x , \qquad \text{for } b = 0 .$$

Special cases can be obtained for $k = 1, 2$, since $d_1(n) = 1$ and $d_2(n) = \sigma_0(n) = d(n)$.

In an analogous manner we can obtain from

$$\sum_1^\infty n^{-s} d(n^2) = \zeta^3(s)/\zeta(2s)$$

the expressions

$$\sum n^{b-1} d(n^2) \log^r n \sim \{2b\zeta(2)\}^{-1} x^b \log^{r+2} x , \qquad \text{for } b > 0 ;$$
$$\sum n^{-1} d(n^2) \log^r n \sim \{2(r+1)\zeta(2)\}^{-1} \log^{r+3} x , \qquad \text{for } b = 0 .$$

Certain of the common number-theoretic functions have not been considered and do not appear in the table (in particular $\mu(n)$, $\lambda(n)$, and $\chi_k(n)$ for non-principal characters) because the sum $F(t)$ fails to satisfy the non-decreasing hypothesis for the Tauberian theorems. $\lambda(n)$ has the additional bad characteristic as shown by the poles of the closed from of the Dirichlet series

$$\sum_1^\infty n^{-s} \lambda(n) = \zeta(2s)/\zeta(s)$$

in that the pole of the numerator is on the line $\Re s = 1/2$ which is critical for the determinator, and thus this is not the pole where $\Re s$ is greatest as required by the theorem from Delange.

Results which he has obtained for the case $r = 0$ and the functions $\sigma(n)$, $\sigma_k(n)$, $d(n)$, and $\phi(n)$, treated by a different method, have been communicated to me in advance of their publication by Mr. Swetharanyam [6].

Table

Asymptotic expressions for $\sum n^a f(n) \log^r n$

| General term of the sum | Asymptotic Expressions | |
|---|---|---|
| | $b>0$ | $b=0$ |
| $n^{b-1-k}\sigma_k(n)\log^r n$ $(k>0)$ | $b^{-1}\zeta(1+k)x^b\log^r x$ | $(r+1)^{-1}\zeta(1+k)\log^{r+1}x$ |
| $n^{b-1}\sigma_k(n)\log^r n$ $(k<0)$ | $(b-k)^{-1}\zeta(1-k)x^{b-k}\log^r x$ $(b>k)$ | $(r+1)^{-1}\zeta(1-k)\log^{r+1}x$ $(b=k)$ |
| $n^{b-2}\sigma(n)\log^r n$ | $b^{-1}\zeta(2)x^b\log^r x$ | $(r+1)^{-1}\zeta(2)\log^{r+1}x$ |
| $n^{b-1}d_k(n)\log^r n$ | $\{b(k-1)!\}^{-1}x^b\log^{r+k-1}x$ | $\{(r+k)(k-1)!\}^{-1}\log^{r+k}x$ |
| $n^{b-1}d(n)\log^r n$ | $b^{-1}x^b\log^{r+1}x$ | $(r+2)^{-1}\log^{r+2}x$ |
| $n^{b-1}\log^r n$ | $b^{-1}x^b\log^r x$ | $(r+1)^{-1}\log^{r+1}x$ |
| $n^{b-1}\wedge(n)\log^r n$ | $b^{-1}x^b\log^r x$ | $(r+1)^{-1}\log^{r+1}x$ |
| $n^{b-2}\phi(n)\log^r n$ | $\{b\zeta(2)\}^{-1}x^b\log^r x$ | $\{(r+1)\zeta(2)\}^{-1}\log^{r+1}x$ |
| $n^{b-1}q_k(n)\log^r n$ | $\{b\zeta(k)\}^{-1}x^b\log^r x$ | $\{(r+1)\zeta(k)\}^{-1}\log^{r+1}x$ |
| $n^{b-1}\mid\mu(n)\mid\log^r n$ | $\{b\zeta(2)\}^{-1}x^b\log^r x$ | $\{(r+1)\zeta(2)\}^{-1}\log^{r+1}x$ |
| $n^{b-1}2^{\omega(n)}\log^r n$ | $\{b\zeta(2)\}^{-1}x^b\log^{r+1}x$ | $\{(r+2)\zeta(2)\}^{-1}\log^{r+2}x$ |
| $n^{b-1}d(n^2)\log^r n$ | $\{2b\zeta(2)\}^{-1}x^b\log^{r+2}x$ | $\{2(r+3)\zeta(2)\}^{-1}\log^{r+3}x$ |
| $n^{b-1}d^2(n)\log^r n$ | $\{6b\zeta(2)\}^{-1}x^b\log^{r+3}x$ | $\{6(r+4)\zeta(2)\}^{-1}\log^{r+4}x$ |
| $\dfrac{\sigma_a(n)\sigma_d(n)\log^r n}{n^{1+a+d-b}}$ $(a>0)$ $(d>0)$ | $\dfrac{\zeta(1+a+d)\zeta(1+a)\zeta(1+d)}{b\zeta(2+a+d)}x^b\log^r x$ | $\dfrac{\zeta(1+a+d)\zeta(1+a)\zeta(1+d)}{(r+1)\zeta(2+a+d)}\log^{r+1}x$ |
| $\dfrac{\sigma_a(n)d(n)\log^r n}{n^{1+a-b}}$ $(a>0)$ | $\dfrac{\zeta^2(1+a)}{b\zeta(2+a)}x^b\log^{r+1}x$ | $\dfrac{\zeta^2(1+a)}{(r+2)\zeta(2+a)}\log^{r+2}x$ |
| $n^{b-2}a(n)\log^r n$ | $2(3b)^{-1}x^b\log^r x$ | $2\{3(r+1)\}^{-1}\log^{r+1}x$ |
| $n^{b-1}\chi_k(n)\log^r n$ | $\phi(k)(kb)^{-1}x^b\log^r x$ | $\phi(k)\{k(r+1)\}^{-1}\log^{r+1}x$ |
| $n^{b-1}r(n)\log^r n$ | $4b^{-1}L_4(1)x^b\log^r x$ | $4(r+1)^{-1}L_4(1)\log^{r+1}x$ |
| $n^{b-1}\wedge(n)\chi_k(n)\log^r n$ | $b^{-1}x^b\log^r x$ | $(r+1)^{-1}\log^{r+1}x$ |
| $n^{b-2}\phi(n)\chi_k(n)\log^r n$ | $\phi(k)\{kbL_k(2)\}^{-1}x^b\log^r x$ | $\phi(k)\{(r+1)kL_k(2)\}^{-1}\log^{r+1}x$ |
| $n^{b-1}2^{\omega(n)}\chi_k(n)\log^r n$ | $4\phi(k)\{3kb\zeta(2)\}^{-1}x^b\log^{r+1}x$ | $4\phi(k)\{3k(r+2)\zeta(2)\}^{-1}\log^{r+2}x$ |
| $n^{b-1}\{\pi(n)-\pi(n-1)\}\log^r n$ $(r>0)$ | $b^{-1}x^b\log^{r-1}x$ | $r^{-1}\log^r x$ |

## References


1. Hubert Delange, *Généralisation du théorème de Ikehara*, Ann. Sci. l'Ecole Norm. Sup. (3) **71** (1954), 213-242.
2. Gustav Doetsch, *Handbuch der Laplace-Transformation*, Basel, 1950.
3. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, Oxford, 1954.
4. Ddmund Landau, *Handbuch der Lehre von der Verteilung der Primzahlen*, 2nd ed., New York, 1953.
5. ———, *Vorlesungen über Zahlentheorie*, reprint, New York, 1947.
6. S. Swetharanyam, *Asymptotic expressions for certain type of sums involving the arithmetic functions of number theory*, J. Ind. Math. Soc. (to be published), Abstract: Math. Student, **25** (1957), p. 81.
7. E. C. Titchmarsh, *The theory of the Riemann Zeta-function*, Oxford, 1951.


University of Wichita and Oregon State College

# A CLASS OF RESIDUE SYSTEMS (mod $r$) AND RELATED ARITHMETICAL FUNCTIONS, I. A GENERALIZATION OF MÖBIUS INVERSION

ECKFORD COHEN

**1. Introduction.** Let $Z$ denote the set of positive integers and let $P$ and $Q$ be nonvacuous subsets of $Z$ such that if $n_1 \in Z, n_2 \in Z$, $(n_1, n_2) = 1$, then

$$n = n_1 n_2 \in P \rightleftarrows n_1 \in P, n_2 \in P; \tag{1.1}$$

suppose also that the elements $n$ in $Q$ satisfy the condition (1.1) with $P$ replaced by $Q$. If, in addition, every integer $n \in Z$ possesses a *unique* factorization of the form

$$n = ab, \quad a \in P, b \in Q, \tag{1.2}$$

then each of the sets $P$ and $Q$ will be called a *direct factor set* of $Z$, while $P$ and $Q$ together will be said to form a *conjugate pair*. In the rest of this paper $P$ will denote such a direct factor set with conjugate set $Q$. It is clear that 1 is the only integer common to both $P$ and $Q$. A simple example of a conjugate pair $P$, $Q$ is the set $P$ consisting of 1 alone and the set $Q = Z$.

Let $r$ be a positive integer. In this paper we shall generalize the notion of a reduced residue system (mod $r$). If $P$ is a given direct factor set, then the elements $a$ of a complete residue system (mod $r$) such that $(a, r) \in P$ will be called a *P-reduced residue system* (mod $r$) or simply a *P-system* (mod $r$). Any two $P$-system (mod $r$) are equivalent in the sense that they are determined by the residue classes of the integers (mod $r$). A $P$-system chosen from the numbers $1 \leq a \leq r$ will be called a *minimal* $P$-system (mod $r$). The number of elements in a $P$-system (mod $r$) will be denoted by $\phi_P(r)$ and called the *P-totient* of $r$. Clearly, if $P = 1, \phi_P(r)$ reduces to the ordinary Eulerian totient $\phi_1(r) = \phi(r)$, while $\phi_Z(r) = r$.

We summarize here the central points of the paper. Analogous to the generalization $\phi_P(r)$ of $\phi(r)$, we define in § 2 a function $\mu_P(r)$ extending the Möbius function $\mu(r)$ to arbitrary direct factor sets $P$. On the basis of this definition we prove in Theorem 3 an analogue of the Möbius inversion formula. This result is then applied in § 3 to yield an evaluation of $\phi_P(r)$. In § 4 a generalization $c_P(n, r)$ of Ramanujan's trigonometric sum $c(n, r)$ is defined and evaluated for arbitrary direct factor sets.

Received August 26, 1958.

In § 5 applications to two relative partition problems (mod $r$) are considered. In particular, in Theorem 12 we obtain a formula for the number of solutions (mod $r$) of the congruence

$$n \equiv x_1 + \cdots + x_s \pmod{r}, \tag{1.3}$$

such that $(x_i, r) \in P$, $(i = 1, \cdots, s)$. In Theorem 13 a formula is deduced for the number $\theta_P(n, r)$ of integers a (mod $r$) such that $(a, r) = 1$ and $(n - a, r) \in P$. These two theorems are wide generalizations of results proved by the author in [1], [2], and [3]. We remark that the method in § 5 and the latter part of § 4 is based on the theory of even functions (mod $r$) developed in the three papers cited above.

In § 6 the results of the preceding sections are specialized to the conjugate pair $P$, $Q$, where $P$ consists of the $k$-free integers and $Q$ is the set of $k$th powers. Precise criteria for the vanishing of $\theta_P(n, r)$ and $\theta_Q(n, r)$ in these cases will be found in Theorem 14.

Regarding the theoretical foundations of arithmetical inversion, we mention an investigation of Hölder [6]. Additional references to the literature appear in Hölder's paper.

Remark. It is noted that several of the results proved in this paper are valid for arbitrary sets $P$, as distinguished from direct factor sets (for example, Theorems 6, 8, 9, and 13). In the general case, however, the unifying method of arithmetical inversion is no longer applicable. The broader topic of arthmetical functions in relation to arbitrary sets $P$ will be treated in another paper.

**2. The inversion function $\mu_P(r)$.** We recall the following fundamental property of $\mu(r)$.

$$\sum_{d \mid r} \mu(d) = \rho(r) \equiv \begin{cases} 1 & (r = 1) \\ 0 & (r > 1) \end{cases}. \tag{2.1}$$

The $\mu$-function may be generalized to arbitrary direct factor sets by writing

$$\mu_P(r) = \sum_{\substack{d \mid r \\ d \in P}} \mu\left(\frac{r}{d}\right), \tag{2.2}$$

where the summation is over the divisors $d$ or $r$ contained in $P$. It will be observed that $\mu_1(r) = \mu(r)$ and $\mu_Z(r) = \rho(r)$.

By (2.2), (1.1), and the factorability of $\mu(r)$, it follows that $\mu_P(r)$ is a factorable function of $r$:

Theorem 1. *If* $r_1 \in J$, $r_2 \in J$, $(r_1, r_2) = 1$, *then*

$$\mu_P(r) = \mu_P(r_1)\mu_P(r_2) \, , \qquad (r = r_1 r_2) \, . \tag{2.3}$$

We next prove that the property (2.1) of $\mu(r)$ can be extended to the function $\mu_P(r)$.

THEOREM 2.

$$\sum_{\substack{d \mid r \\ d \in Q}} \mu_P\left(\frac{r}{d}\right) = \rho(r) \, . \tag{2.4}$$

*Proof.* On the basis of (2.1), (2.2) and the uniqueness of the factorization (1.2) one obtains

$$\begin{aligned} \sum_{\substack{d \mid r \\ d \in Q}} \mu_P\left(\frac{r}{d}\right) &= \sum_{\substack{de = r \\ d \in Q}} \sum_{\substack{e = \delta D \\ \delta \in P}} \mu(D) \\ &= \sum_{D \mid r} \mu(D) \sum_{\substack{\delta d = r/D \\ \delta \in P, d \in Q}} 1 = \sum_{D \mid r} \mu(D) = \rho(r) \, . \end{aligned}$$

This completes the proof.

By means of Theorem 2 we generalize the Möbius inversion formula to arbitrary direct factor sets.

THEOREM 3. *If $f(r)$ and $g(r)$ are arithmetical functions, then*

$$f(r) = \sum_{\substack{d \mid r \\ d \in Q}} g\left(\frac{r}{d}\right) \rightleftarrows g(r) = \sum_{d \mid r} f(d) \mu_P\left(\frac{r}{d}\right) \, . \tag{2.5}$$

*Proof.* Let $f(r)$ be defined as on the left of (2.5). By (2.4) one obtains

$$\begin{aligned} \sum_{d \mid r} f(d)\mu_P\left(\frac{r}{d}\right) &= \sum_{d \mid r} \left( \sum_{\substack{\delta e = d \\ \delta \in Q}} g(e) \right) \mu_P\left(\frac{r}{d}\right) \\ &= \sum_{e \mid r} g(e) \sum_{\substack{d\delta' = r \\ \delta e = d \\ \delta \in Q}} \mu_P(\delta') = \sum_{e \mid r} g(e) \sum_{\substack{\delta\delta' = r/e \\ \delta \in Q}} \mu_P(\delta') \\ &= \sum_{e \mid r} g(e) \rho\left(\frac{r}{e}\right) = g(r) \, . \end{aligned}$$

Conversely, let $g(r)$ be defined as on the right of (2.5). Then again by (2.4)

$$\begin{aligned} \sum_{\substack{d \mid r \\ d \in Q}} g\left(\frac{r}{d}\right) &= \sum_{\substack{de = r \\ d \in Q}} \left( \sum_{\delta \mid e} f(\delta) \mu_P\left(\frac{e}{\delta}\right) \right) \\ &= \sum_{\delta \mid r} f(\delta) \sum_{\substack{de = r \\ \delta\delta' = e \\ d \in Q}} \mu_P(\delta') = \sum_{\delta \mid r} f(\delta) \sum_{\substack{d\delta' = r/\delta \\ d \in Q}} \mu_P(\delta') \end{aligned}$$

$$= \sum_{\delta | q} f(\delta) \rho\left(\frac{r}{\delta}\right) = f(r) .$$

The proof is complete.

It is evident that if $P = 1$, $Q = Z$, Theorem 3 becomes the inversion formula of elementary number theory.

3. **The totient function $\phi_P(r)$.** The following principle is basic in considering $P$-totients.

THEOREM 4. *Let $d$ range over the divisiors of $r$ contained in $Q$, and for each such $d$ let $X$ range over the elements of a $P$-*system (mod *$r/d$). Then the set $d\ X$ forms a complete residue system* (mod r).

*Proof.* In the proof we suppose $n$ to range over the positive integers $\leq r$. For a fixed divisor $d$ of $r$, $d \in Q$, let $C_d$ represent the set of those $n$ for which $(n, r)$ is of the form $(n, r) = de$, $e \in P$. By the uniqueness of the factorization (1.2), a given $n$ lies in exactly one class $C_d$; hence the set of elements in the classes $C_d$ consists precisely of the integers $1, \cdots, r$. Moreover, for a fixed divisor $d$ of $r$ such that $d \in Q$, the elements $n = dx$ comprise $C_d$ if and only if $(x, r/d) \in P$, $1 \leq x \leq r/d$, that is, if and only if the elements $x$ form a minimal $P$-system (mod $r/d$). Replacing the particular $P$-system $x$ (mod $r/d$), by an arbitrary $P$-system $X$ (mod $r/d$) the theorem results.

Theorem 4 leads immediately to

THEOREM 5.

$$\sum_{\substack{d | r \\ d \in Q}} \phi_P\left(\frac{r}{d}\right) = r . \tag{3.1}$$

The evaluation of $\phi_P(r)$ follows from (3.1) on applying the inversion formula of Theorem 3:

THEOREM 6.

$$\phi_P(r) = \sum_{d | r} d \mu_P\left(\frac{r}{d}\right) . \tag{3.2}$$

In case $P = 1$, Theorem 6 becomes the well-known evaluation formula for $\phi(r)$.

Since $\mu_P(r)$ is factorable (Theorem 1) the same is true of $\phi_P(r)$, by (3.2):

THEOREM 7. *If* $(r_1, r_2) = 1$, *then*

$$\phi_P(r) = \phi_P(r_1)\phi_P(r_2)\,, \qquad (r = r_1 r_2)\,. \tag{3.3}$$

Next we show how $\phi_P(r)$ may be expressed in terms of the ordinary $\phi$-function.

THEOREM 8.

$$\phi_P(r) = \sum_{\substack{d \mid r \\ d \in P}} \phi\left(\frac{r}{d}\right). \tag{3.4}$$

*Proof.* By (2.2) and (3.2) it follows that

$$\phi_P(r) = \sum_{\delta \mid r} \delta \sum_{\substack{d \mid r/\delta \\ d \in P}} \mu\left(\frac{r/\delta}{d}\right) = \sum_{\substack{d \mid r \\ d \in P}} \sum_{\delta \mid r/d} \delta\mu\left(\frac{r/d}{\delta}\right).$$

and (3.4) results by (3.2) with $P = 1$.

**4. The exponential sum $c_P(n, r)$.** We define

$$c_P(n, r) = \sum_{(x, r) \in P} e(xn, r)\,, \qquad e(a, r) = e^{2\pi i a/r}, \tag{4.1}$$

where the summation is over a $P$-system (mod $r$). In case $P = 1$, $c_P(n, r)$ reduces to the Ramanujan sum, $c(n\ r)$. The next theorem generalizes the familiar evaluation of $c(n, r)$.

THEOREM 9.

$$c_P(n, r) = \sum_{d \mid (n, r)} d\mu_P\left(\frac{r}{d}\right). \tag{4.2}$$

*Proof.* Placing $\eta(n, r) = c_Z(n, r)$, we have

$$\eta(n, r) = \sum_{x (\mathrm{mod}\ r)} e(xn, r) = \begin{cases} r & (r \mid n) \\ 0 & (r \nmid n)\,. \end{cases} \tag{4.3}$$

Furthermore, by Theorem 4,

$$\eta(n, r) = \sum_{\substack{d \mid r \\ d \in Q}} \sum_{(x, r/d) \in P} e(dxn, r) = \sum_{\substack{d \mid r \\ d \in Q}} c_P\left(n, \frac{r}{d}\right). \tag{4.4}$$

Therefore, by the inversion theorem (§ 2),

$$c_P(n, r) = \sum_{d \mid r} \eta(n, d)\mu_P\left(\frac{r}{d}\right),$$

and the theorem follows on the basis of (4.3).

The function $c_P(n, r)$ is a generalization of both $\phi_P(r)$ and $\mu_P(r)$:

COROLLARY 9.1. *If* $n \equiv 0 \pmod{r}$, *then*

$$c_P(n, r) = \phi_P(r) . \tag{4.5}$$

COROLLARY 9.2 *If* $(n, r) = 1$, *then*

$$c_P(n, r) = \mu_P(r) . \tag{4.6}$$

By (4.2) and (2.3) we have, in addition,

THEOREM 10. *The function* $c_P(n, r)$ *is a factorable function of* $r$; *that is, if* $(r_1, r_2) = 1$, *then*

$$c_P(n, r) = c_P(n, r_1)c_P(n, r_2) , \qquad (r = r_1 r_2) . \tag{4.7}$$

In the proof of the next theorem we assume the results on even functions (mod $r$) proved in [1]. We first state a lemma which results on applying the Möbius-inversion formula to (2.2).

LEMMA 1.

$$\sum_{d|r} \mu_P(d) = \rho_P(r) \equiv \begin{cases} 1 & (r \in P) \\ 0 & (r \notin P) , \end{cases} \tag{4.8}$$

It is noted that $\rho_1(r) = \rho(r)$ .

THEOREM 11.

$$c_P(n, r) = \sum_{d|r} \rho_P\left(\frac{r}{d}\right)c(n, d) = \sum_{\substack{d|r \\ d \in P}} c\left(n, \frac{r}{d}\right) . \tag{4.9}$$

*Proof.* By (4.2), $c_P(n, r) = c_P((n, r), r)$, so that $c_P(n, r)$ is an even function of $n \pmod{r}$. Hence by Theorem 9 and [1, Theorem 4], $c_P(n, r)$ has a Fourier expansion,

$$c_P(n, r) = \sum_{d|r} \alpha(d, r)c(n, d) ,$$

where

$$\alpha(d, r) = \sum_{e|r/d} \mu_P(e) ,$$

and the theorem follows by (4.8).

We note that (4.9) reduces to (3.4) in case $n = 0$, thereby providing a new proof of Theorem 8, while in case $n = 1$, (4.9) becomes (2.2).

**5. Relative partitions** (mod $r$). In this section we assume the results of [2] and [3]. Let $A_p^{(s)}(n, r)$ denote the number of solutions (mod $r$) of (1.3), such that for each $x_i$, $(1 \leq i \leq s)$, $(x_i, r)$ is contained in a $P$-system (mod $r$). We deduce the following expansion for $A_P^{(s)}(n, r)$.

THEOREM 12. *For arbitrary positive integral s,*

$$A_P^{(s)}(n, r) = \frac{1}{r} \sum_{d|r} \left(c_P\left(\frac{r}{d}, r\right)\right)^s c(n, d) . \tag{5.1}$$

*Proof.* We prove (5.1) inductively on $s$. Obviously $A_P^{(1)}(n, r) = \rho_P((n, r))$. Hence applying [2, Theorem 3] to (4.9), one obtains

$$A_P^{(1)}(n, r) = \frac{1}{r} \sum_{d|r} c_P\left(\frac{r}{d}, r\right) c(n, d) . \tag{5.2}$$

This proves the theorem in case $s = 1$. We assume the theorem for $s = t \geq 1$. Then by [3, Theorem 1]

$$\begin{aligned} A_P^{(t+1)}(n, r) &= \sum_{n \equiv a+b \pmod r} A_P^{(t)}(a, r) A_P^{(1)}(b, r) \\ &= \frac{1}{r} \sum_{d|r} \left(c_P\left(\frac{r}{d}, r\right)\right)^{t+1} c(n, d) . \end{aligned}$$

This completes the induction.

Next we derive an arithmetical formula for the function $\theta_P(n, r)$ defined in the Introduction. Equivalently $\theta_P(n, r)$ may be defined as the number of solutions, $x, y$ (mod $r$) of

$$n \equiv x + y \pmod r , \quad (x, r) = 1 , \quad (y, r) \in P . \tag{5.3}$$

The proof will depend on the following lemma.

LEMMA 2. *Let e be a positive integer. Then*

$$\sum_{d|r} c\left(\frac{r}{d}, e\right)\mu(d) = \begin{cases} \mu\left(\dfrac{e}{r}\right) r & \textit{if } r \mid e , \\ 0 & \textit{otherwise.} \end{cases} \tag{5.4}$$

*Proof.* By the evaluation formula for $c(n, r)$,

$$\begin{aligned} \sum_{d|r} c\left(\frac{r}{d}, e\right)\mu(d) &= \sum_{d|r} \mu(d) \sum_{D|(r/d, e)} D\mu\left(\frac{e}{D}\right) \\ &= \sum_{D|(e,r)} \mu\left(\frac{e}{D}\right) D \sum_{d|r/D} \mu(d) , \end{aligned}$$

and (5.4) follows on applying (2.1) to the inner sum of the last expression.

THEOREM 13.

$$\theta_P(n, r) = \phi(r) \sum_{\substack{d \mid r \\ (d,n)=1}} \frac{\mu_P(d)}{\phi(d)}, \tag{5.5}$$

*where the summation is over the divisors of r prime to n.*

*Proof.* Using (5.2) we apply [2 Theorem 6] to $\theta_P(n, r)$ with $f(n, r) = A_P^{(1)}(n, r)$, obtaining on the basis of Theorem 11 and Lemma 2,

$$\begin{aligned}\phi^{-1}(r)\theta_P(n, r) &= \frac{1}{r} \sum_{\substack{d \mid r \\ (d,n)=1}} \frac{d}{\phi(d)} \Big( \sum_{\delta\delta' = r/d} c_P(\delta', r)\mu(\delta) \Big) \\ &= \frac{1}{r} \sum_{\substack{d \mid r \\ (d,n)=1}} \frac{d}{\phi(d)} \sum_{\substack{ee' = r \\ e' \in P}} \Big( \sum_{\delta\delta' = r/d} c(\delta', e)\mu(\delta) \Big) \\ &= \sum_{\substack{d \mid r \\ (d,n)=1}} \frac{1}{\phi(d)} \sum_{\substack{ee' = r \\ (r/d)E = e \\ e' \in P}} \mu(E) = \sum_{\substack{d \mid r \\ (d,n)=1}} \frac{1}{\phi(d)} \sum_{\substack{Ee' = d \\ e' \in P}} \mu(E),\end{aligned}$$

and the theorem follows by definition $\mu_P(r)$.

6. **Special cases.** For a fixed non-negative integer $k$, let $P$ be the set of all $k$-free numbers and let $Q$ be the set of all $k$th powers. Clearly $P$ and $Q$ form a conjugate pair of direct factor sets. We introduce the following notation for the functions corresponding to these sets: $\Phi_k(r) = \phi_P(r)$, $\mu_k(r) = \mu_P(r)$, $g_k(n, r) = c_P(n, r)$, and $\Psi_k(r) = \phi_Q(r)$, $\lambda_k(r) = \mu_Q(r)$, $h_k(n, r) = c_Q(n, r)$. If $(a, b)_k$ is defined to be the greatest $k$th power divisor of $a$ and $b$, then $\Phi_k(r)$ denotes the number of integers $a$ (mod $r$) such that $(a, r)_k = 1$, while $\Psi_k(r)$ denotes the number of $a$ (mod $r$) such that $(a, r)$ is a $k$th power, that is, $(a, r)_k = (a, r)$.

It is observed that, in case $k = 1$, $\Phi_k(r)$, $\mu_k(r)$, and $g_k(n, r)$ reduce to $\phi(r)$, $\mu(r)$, and $c(n, r)$, respectively. We also note that $\lambda_2(r) = \lambda(r)$, where $\lambda(r)$ represents the Liouville function. The conjugate totient functions $\Phi_k(r)$, and $\Psi_k(r)$ were introduced by Rogel [9]. Regarding the special case $k = 2$ of these two functions, $\Phi_2(r)$ was evaluated by Haviland [5] using a definition equivalent to that given here, while $\Psi_2(r)$ was evaluated by the author in [2, Corollary 4.2]. For a further discussion of the function $\Phi_k(r)$ we refer to McCarthy [7].

The following evaluation arise as corollaries of the results proved in §§ 3 and 4.

$$\Phi_k(r) = \sum_{d \mid r} d\mu_k\Big(\frac{r}{d}\Big) = \sum_{\substack{d \mid r \\ (d,r)_k=1}} \phi\Big(\frac{r}{d}\Big), \tag{6.1}$$

$$\Psi_k(r) = \sum_{d \mid r} d\lambda_k\Big(\frac{r}{d}\Big) = \sum_{d^k \mid r} \phi\Big(\frac{r}{d^k}\Big), \tag{6.2}$$

(6.3) $$g_k(n, r) = \sum_{d\mid(n,r)} d\mu_k\left(\frac{r}{d}\right) = \sum_{\substack{d\mid r \\ (d,r)_k=1}} c\left(n, \frac{r}{d}\right),$$

(6.4) $$h_k(n, r) = \sum_{d\mid(n,r)} d\lambda_k\left(\frac{r}{d}\right) = \sum_{d^k\mid r} c\left(n, \frac{r}{d}k\right).$$

By (2.2) the functions $\mu_k(r)$ and $\lambda_k(r)$ may be written

(6.5) $$\mu_k(r) = \sum_{\substack{d\mid r \\ (d,r)_k=1}} \mu\left(\frac{r}{d}\right), \quad \lambda_k(r) = \sum_{d^k\mid r} \mu\left(\frac{r}{d^k}\right),$$

In view of the factorability of $\mu_k(r)$ and $\lambda_k(r)$ it is sufficient to evaluate these functions for prime-power values of $r$, $r = p^m$ ($p$ prime, $m > 0$). In particular, it is easily deduced from (6.5) that

(6.6) $$\mu_k(p^m) = \begin{cases} -1 & (m = k) \\ 0 & (m \neq k), \end{cases}$$

while for $k \geq 2$,

(6.7) $$\lambda_k(p^m) = \begin{cases} 1 & (m \equiv 0 \pmod{k}) \\ -1 & (m \equiv 1 \pmod{k}) \\ 0 & (\text{otherwise}). \end{cases}$$

The functions $\mu_k(n)$ and $\lambda_k(n)$ were introduced by Gegenbauer [**4**]; for a further discussion we mention Hölder [**6**, §§ 6–7]. Note that $\lambda_1(r) = \mu_0(r) = \rho(r)$, $\lambda_0(r) = \mu(r)$.

The corresponding inversion formulas are contained in the following relations (Theorem 3):

(6.8) $$f(r) = \sum_{d^k\mid r} g\left(\frac{r}{d}\right) \rightleftarrows g(r) = \sum_{d\mid r} f(d)\mu_k\left(\frac{r}{d}\right);$$

(6.9) $$f(r) = \sum_{\substack{d\mid r \\ (d\ k-\text{free})}} g\left(\frac{r}{d}\right) \rightleftarrows g(r) = \sum_{d\mid r} f(d)\lambda_k\left(\frac{r}{d}\right).$$

The case $k = 1$ in (6.8) is the ordinary inversion theorem, while the case $k = 2$ in (6.9) yields the formula,

(6.9a) $$f(r) = \sum_{\substack{d\mid r \\ (\mu(d)\neq 0)}} g\left(\frac{r}{d}\right) \rightleftarrows g(r) = \sum_{d\mid r} f(d)\lambda\left(\frac{r}{d}\right),$$

the summation on the left ranging over the primitive (square-free) divisors of $r$.

We now specialize the additive results of § 5 to the particular sets $P$, $Q$ of this section. Placing $R_{k,s}(n, r) = A_P^{(s)}(n, r)$, $S_{k,s}(n, r) = A_Q^{(s)}(n, r)$, we observe that $R_{k,s}(n, r)$ represents the number of solutions of (1.3) such that $(x_i, r)_k = 1$, while $S_{k,s}(n, r)$ represents the number of solutions of (1.3) such that $(x_i, r)$ is a $k$th power ($i = 1, \cdots, s$). In particular, one obtains from Theorem 12,

$$(6.10) \qquad R_{k,s}(n, r) = \frac{1}{r} \sum_{d|r} \left(g_k\left(\frac{r}{d}, r\right)\right)^s c(n, d) ,$$

$$(6.11) \qquad S_{k,s}(n, r) = \frac{1}{r} \sum_{d|r} \left(h_k\left(\frac{r}{d}, r\right)\right)^s c(n, d) .$$

The case $k = 1$ in (6.10) is Theorem 6 of [1], (also cf. [2, § 2]), while the case $k = 2$ in (6.11) is Theorem 3 of [3] in an equivalent form.

If one places $\theta_P(n, r) = \theta_k(n, r)$ and $\theta_Q(n, r) = \varepsilon_k(n, r)$, then $\theta_k(n, r)$ denote the number of integers $a \pmod r$ such that $(a, r) = 1$ and $(n - a, r)_k = 1$, while $\varepsilon_k(n, r)$ denotes the number of $a \pmod r$ such that $(a, r) = 1$ and $(n - a, r)$ is a $k$th power. We deduce then from Theorem 13,

$$(6.12) \qquad \theta_k(n, r) = \phi(r) \sum_{\substack{d|r \\ (d,n)=1}} \frac{\mu_k(d)}{\phi(d)} ,$$

$$(6.13) \qquad \varepsilon_k(n, r) = \phi(r) \sum_{\substack{p|r \\ (d,n)=1}} \frac{\lambda_k(d)}{\phi(d)} .$$

The case $k = 1$ in (6.12) is [2, Corollary 21] while the case $k = 2$ in (6.13) is [3, Corollary 38].

Finally, we investigate the conditions under which $\theta_k(n, r)$ and $\varepsilon_k(n, r)$ vanish. It is sufficient to consider these functions when $r$ and $n$ are powers of the same prime $p$, $r = p^t$, $n = p^b$, $t > 0$, $t \geq b \geq 0$. A simple computation yields the following results. If $k \geq 1$, then

$$\theta_k(p^b, p^t) = \begin{cases} p^{t-k}(p^k - p^{k-1} - 1) & \text{if } b = 0, t \geq k , \\ p^{t-1}(p - 1) & \text{otherwise} . \end{cases}$$

Suppose $ak < t \leq (a + 1)k$ where $a$ is a (uniquely defined) non-negative integer. Then, if $k \leq 2$,

$$(p^k - 1)\varepsilon_k(p^b, p^t) = \begin{cases} p^{t-1}(p - 1)(p^k - 1) , \\ p^{t-k(a+1)}(p^{k-1} - 1) + p^{k+t-1}(p - 2) + p^{t-1} \\ p^{t+k-1}(p - 2) + p^{t-ak-1}(p^{ak} - p + 1) , \end{cases}$$

according as (i) $b > 0$, (ii) $b = 0, t = (a + 1)k$, or (iii) $b = 0, t < (a + 1)k$.

From these results it is easy to deduce that $\theta_k(p^b, p^t) = 0$ if and only if $p = 2, k = 1, b = 0$ and that $\varepsilon_k(p^b, p^t) = 0$ if and only if $p = 2, t < k, b = 0$. We are therefore led, on the basis of factorability considerations, to the following criterion in the general case.

THEOREM 14. *If* $k \geq 1$, *then* $\theta_k(n, r) = 0$ *if and only if* $k = 1$, $r$ *is even, and* $n$ *is odd.*

*If* $k \geq 2$, *then* $\varepsilon_k(n, r) = 0$ *if and only if* $r$ *is of the form* $2^t R$ *where* $R$ *is odd,* $0 < t < k$, *and* $n$ *is odd.*

The above result for $\theta_k(n, r)$ in case $k = 1$ is due to Ramanathan [8, p. 68]. The result for $\varepsilon_k(n, r)$ in case $k = 2$ was proved in [3, Corollary 38.1].

## REFERENCES

1. Eckford Cohen, *A class of arithmetical functions*, Proc. Nat. Acad. Sci., **41** (1955), 939-944.

2. ———, *Representation of even functions* (mod $r$), I. *Arithmetical identities*, Duke Math. J. **25** (1958), 401-421.

3. ———, *Representations of even functions* (mod $r$), II. *Cauchy products*, Duke Math. J., to appear.

4. Leopold Gegenbauer, *Zahlentheoretische Notiz*, S.-B. Akad. Wiss. (Vienna), **97** (1889), IIa, 420-426.

5. E. K. Haviland, *An analogue of Euler's $\phi$-function*, Duke Math. J., **11** (1944), 869-872.

6. O. Hölder, *Über gewisse der Möbiusschen Funktion $\mu(n)$ verwandte zahlentheoretische Funktionen, die Dirichletsche Multiplikation, und eine Verallgemeinerung der Umkehrungsformeln*, Verh. Sachs. Akad. Wiss (Leipzig), **85** (1933), 3-28.

7. Paul. J. McCarthy, *On a certain family of arithmetic functions*, Amer. Math. Monthly, **65** (1958), 586-590.

8. K. G. Ramanathan, *Some applications of Ramanujan's trigonometrical sum $C_m(n)$*, Proc. Indian Acad. Sci. (A), **20** (1944), 62-69.

9. F. Rogel, *Entwicklung einiger zahlentheoretischer Funktionen in unendliche, Reihen* S.-B. Kgl. Böhmischen Ges. Wiss. **30** (1900).

UNIVERSITY OF TENNESSEE

# NON-ABELIAN ORDERED GROUPS

PAUL CONRAD

**1. Introduction.** In this paper we prove some theorems about non-abelian o-groups, and give some methods of constructing such groups. Most of the literature on o-groups is concerned with abelian o-groups, and the examples in print of non-abelian o-groups are few. Iwasawa [8] proves that any free group can be ordered, and he also gives some additional examples of o-groups. Vinogradov [15] shows that the free product of two o-groups $A$ and $B$ can be ordered so as to preserve the given orders. Chehata [1] gives an example of an o-group that is simple. [3] and [11] contain examples of o-groups. Most of the theorems in this paper give methods for constructing o-groups. For example, in § 3 we study the o-automorphisms of an o-group $G$. For every group $A$ of o-automorphisms of $G$ that can be ordered we can construct a new o-group $H$ that contains $A$ and $G$. $H$ is the natural splitting extension of $G$ by $A$. In § 5 the relationship between central extensions and bilinear mappings is exploited. It is shown that any skew-symmetric real matrix can be used to construct o-groups. In § 6 some o-groups of rank 2 are constructed. In § 4 a study is made of the ordered extensions of a subgroup of the reals. One of the main results is a necessary and sufficient condition for such an extension to split. The principal tool used throughout is the extension theory of Schreier [14].

**2. Notation and Terminology.** The notation of [3] is used throughout. In particular, the notation and results from § 2 [3, pp. 517–518] are used repeatedly. Unless otherwise stated the group operation will always be addition and 0 will denote a group identity. $N$ and $N'$ are o-groups with elements $a, b, c, \cdots$ and $a', b', c', \cdots$ respectively. $G$ is a normal o-extension of $N$ by $N'$. We identify $G$ with its representation $G' = N' \times N$, where

$$(a', a) + (b', b) = (a' + b', f(a', b') + ar(b') + b)$$

and $(a', a)$ is positive if $a' > 0$ or $a' = 0$ and $a > 0$. See [3] for the properties of the factor mapping $f$ and the representative function $r$.

$\theta$ will always denote a trivial homomorphism of a group onto the identity element of some other group. For an o-group $H$, let $A(H)$ be the group of all o-automorphisms of $H$. For an abelian o-group $K$, let $D(K)$ be the $d$-closure or completion of $K$. In particular, $D(K)$ is a vector space over the rationals and there is a natural extension of the order

Received August 25, 1958. This work was supported by a grant from the National Science Foundation.

of $K$ to an order of $D(K)$. Finally let $\mathbf{R}$ be the additive group of all real numbers, $\mathbf{P}$ be the multiplicative group of all positive real numbers, $R$ be the additive group of all rational numbers, $\mathbf{P}$ be the multiplicative group of all positive rational numbers, and $I$ be the additive group of integers—*all with their natural order*.

3. **Order preserving automorphisms of G.** If $H$ is an o-group and $A$ is a group of o-automorphisms of $H$ that can be ordered, then the group $H' = A \times H$, where $(\alpha, a) + (\beta, b) = (\alpha\beta, a\beta + b)$ for $\alpha, \beta$ in $A$ and $a, b$ in $H$, can be ordered. Simply define $(\alpha, a)$ positive if $\alpha$ is positive in $A$ or $\alpha$ is the identity and $a$ is positive in $H$. Then clearly $H'$ is a splitting o-extension of $H$ by $A$. Thus if $A$ contains more than one element, then $H'$ is a non-abelian o-group. If $A$ is the group of all o-automorphisms of $H$, then $H'$ is called the o-*holomorph* of $H$. In [5] it has been shown that a certain class of o-groups with well ordered rank have ordered o-holomorphs. In this section we investigate the o-automorphisms of $G$.

Let $\pi$ be an o-automorphism of $G$ for which $(0 \times N)\pi = 0 \times N$. and let $\mathscr{A}$ be the group of all these o-automorphisms. If $G$ has well ordered rank or if $N'$ or $N$ has finite rank, then $\mathscr{A} = A(G)$. For $(a', a)$ and $(b', b)$ in $G$ we have

$$(a', a)\pi = [(a', 0) + (0, a)]\pi = (a', 0)\pi + (0, a)\pi$$
$$= (a'\alpha, a'\beta) + (0, a\gamma) = (a'\alpha, a'\beta + a\gamma) ,$$

where

$$(1) \qquad 0\beta = 0 .$$

$$[(a', a) + (b', b)]\pi = (a' + b', f(a', b') + ar(b') + b)\pi$$
$$= ((a' + b')\alpha, (a' + b')\beta + (f(a', b') + ar(b') + b)\gamma) .$$
$$(a', a)\pi + (b', b)\pi = (a'\alpha, a'\beta + a\gamma) + (b'\alpha, b'\beta + b\gamma)$$
$$= (a'\alpha + b'\alpha, f(a'\alpha, b'\alpha) + (a'\beta + a\gamma)r(b'\alpha) + b'\beta + b\gamma) .$$

Thus $\alpha \in A(N')$ and

$$(a' + b')\beta + (f(a', b') + ar(b') + b)\gamma$$
$$= f(a'\alpha, b'\alpha) + (a'\beta + a\gamma)r(b'\alpha) + b'\beta + b\gamma .$$

When $a' = b' = 0$ this reduces to $(a + b)\gamma = a\gamma + b\gamma$. Thus $\gamma \in A(N)$. The following two equations are the result of letting $a' = b = 0(a = b = 0)$.

$$(2) \qquad b'\beta + ar(b')\gamma = a\gamma r(b'\alpha) + b'\beta$$

$$(3) \qquad (a' + b')\beta + f(a', b')\gamma = f(a'\alpha, b'\alpha) + a'\beta r(b'\alpha) + b'\beta .$$

Conversely suppose that $\alpha \in A(N'), \gamma \in A(N), \beta : N' \to N,$ and (1), (2), (3) are satisfied. For $(a'\, a)$ in $G$ define $(a', a)\pi = (a'\alpha, a'\beta + a\gamma)$. Then by straightforward computation it follows that $\pi \in \mathscr{A}$.

For mappings $u$ and $v$ of $N'$ into $N$ and $a' \in N'$ we define $a'(u+v) = a'u + a'v$. Then each $\pi \in \mathscr{A}$ has a matrix representation

$$\begin{bmatrix} \alpha \beta \\ \theta \gamma \end{bmatrix}$$

where $\theta$ is the trivial homomorphism of $N$, into $N'$, and the mapping of $\pi$ onto its matrix representation is an isomorphism of $\mathscr{A}$ onto

$$\left\{\begin{bmatrix} \alpha \beta \\ \theta \gamma \end{bmatrix} : \alpha \in A(N), \gamma \in A(N'), \beta : N' \to N, \text{ and (1), (2), (3) are satisfied}\right\}.$$

For, let $\pi = (\alpha, \beta, \gamma)$ and $\bar{\pi} = (\bar{\alpha}, \bar{\beta}, \bar{\gamma})$, then

$$\begin{aligned}(a', a)\bar{\pi}\pi &= (a'\bar{\alpha}, a'\bar{\beta} + a\bar{\gamma})\pi = (a'\bar{\alpha}\alpha, a'\bar{\alpha}\beta + (a'\bar{\beta} + a\bar{\gamma})\gamma) \\ &= (a'\bar{\alpha}\alpha, a'(\bar{\alpha}\beta + \bar{\beta}\gamma) + a\bar{\gamma}\gamma)\end{aligned}$$

and

$$(4) \qquad \begin{bmatrix} \bar{\alpha}\bar{\beta} \\ \theta\bar{\gamma} \end{bmatrix}\begin{bmatrix} \alpha\beta \\ \theta\gamma \end{bmatrix} = \begin{bmatrix} \bar{\alpha}\alpha & \bar{\alpha}\beta + \bar{\beta}\gamma \\ \theta & \bar{\gamma}\gamma \end{bmatrix}$$

We shall frequently identity the elements of $\mathscr{A}$ with their matrix representation. Let $\mathscr{B}$ be the set of all $\beta : N' \to N$ that satisfy (1), (2), (3) when $\alpha$ and $\gamma$ are the identity automorphisms of $N'$ and $N$ respectively.

LEMMA 3.1. *$\mathscr{B}$ is an additive group that can be ordered.*

*Proof.* From the matrix representation of $\mathscr{A}$ it follows that $\mathscr{B}$ is a group. Well order the elements of $N'$ and define $\beta \in \mathscr{B}$ positive if $\beta \neq \theta$ and $a'\beta > 0$, where $a'$ is the first element in the well ordering for which $a'\beta \neq 0$. It is easy to check that this definition orders $\mathscr{B}$.

COROLLARY I. *The group of all mappings of a set onto an o-group can be ordered.*

COROLLARY II. *The group of all o-automorphisms of $G$ that induce the identity automorphism on $G/(0 \times N)$ and on $0 \times N$ can be ordered.*

Now suppose that $\mathscr{B}$, $A(N')$ and $A(N)$ are o-groups and let

$$\pi = \begin{bmatrix} \alpha\beta \\ \theta\gamma \end{bmatrix} \quad \bar{\pi} = \begin{bmatrix} \bar{\alpha}\bar{\beta} \\ \theta\bar{\gamma} \end{bmatrix}$$

be elements of $\mathscr{A}$. Then

$$(5) \quad \pi^{-1} = \begin{bmatrix} \alpha^{-1} & -\alpha^{-1}\beta\gamma^{-1} \\ \theta & \gamma^{-1} \end{bmatrix} \quad \pi^{-1}\bar{\pi}\pi = \begin{bmatrix} \alpha^{-1}\bar{\alpha}\alpha & \alpha^{-1}(\bar{\alpha}\beta + \bar{\beta}\gamma) - \alpha^{-1}\beta\gamma^{-1}\bar{\gamma}\gamma \\ \theta & \gamma^{-1}\bar{\gamma}\gamma \end{bmatrix}$$

DEFINITION 3.1. $\pi$ is *positive* if $\alpha$ is positive in $A(N')$ or $\alpha$ is the identity and $\gamma$ is positive in $A(N)$ or $\alpha$ ond $\gamma$ are identity automorphisms and $\beta$ is positive in $\mathscr{B}$.

Let $\mathscr{P}$ be the set of all positive elements in $\mathscr{A}$. It follows from (4) that $\mathscr{P}$ is closed with respect to multiplication. It follows from the first part of (5) that for each $\pi \in \mathscr{A}$, either $\pi$ is the identity or $\pi \in \mathscr{P}$ or $\pi^{-1} \in \mathscr{P}$. Unfortunately $\mathscr{P}$ is not in general normal. For suppose that $\bar{\pi} \in \mathscr{P}$, then if $\bar{\alpha}$ is positive or $\bar{\gamma}$ is positive, then $\pi^{-1}\bar{\pi}\pi$ is positive. Finally assume that $\bar{\alpha}$ and $\bar{\gamma}$ are identity automorphisms, then

$$\pi^{-1}\bar{\pi}\pi = \begin{bmatrix} \phi' & \alpha^{-1}(\beta + \bar{\beta}\gamma - \beta) \\ \theta & \phi \end{bmatrix},$$

where $\phi'(\phi)$ is the identity of $A(N')(A(N))$. Thus our definition orders $\mathscr{A}$ if and only if $\alpha^{-1}(\beta + \bar{\beta}\gamma - \beta) = \alpha^{-1}\beta + \alpha^{-1}\bar{\beta}\gamma - \alpha^{-1}\beta$ is positive. If we use the ordering of $\mathscr{B}$ defined in the proof of Lemma 3.1, then it suffices to show that $a'\alpha^{-1}\bar{\beta} > 0$, where $a'$ is the first element in the well ordering of $N'$ such that $a'\alpha^{-1}\bar{\beta} \neq 0$.

THEOREM 3.1. *If $A(N)$ can be ordered, then the group of all o-automorphisms $\pi$ of $G$ such that $(0 \times N)\pi = 0 \times N$ and $\pi$ induces the identity automorphism on $G/(0 \times N)$ can be ordered.*

We next consider the special cases where $G$ is a central extension of $N$ or where $G$ is a splitting extension of $N$. First assume that $N$ (actually $0 \times N$) is in the center of $G$. Then $r = \theta$ and $N$ is abelian. In particular, (1), (2), (3) reduce to

$$(a' + b')\beta + f(a', b')\gamma = f(a'\alpha, b'\alpha) + a'\beta + b'\beta$$

and $0\beta = 0$. Thus $\mathscr{B}$ is the torsion free abelian group $H(N', N)$ of all homomorphisms of $N'$ into $N$. Let $\Gamma$ be the set of all ordered pairs of convex subgroups $N'^{\gamma}$, $N'_{\gamma}$ of $N'$ such that $N'^{\gamma}$ covers $N'_{\gamma}$.

THEOREM 3.2. *Suppose that $G$ is a central extension of $N$, $A(N)$ can be ordered, $\Gamma$ is well ordered, and for each pair $\alpha \in A(N')$ and $\gamma \in \Gamma$ there exists a pair of positive integers $m$ and $n$ such that $ng\alpha \equiv mg$ modulo $N'_{\gamma}$ for all $g \in N'^{\gamma}$. Then $A(N')$ and $\mathscr{A}$ can be ordered.*

*Proof.* By the theorem in [5], $A(N')$ can be ordered. As in the proof of Theorem 3 [4 p. 388] we well order the elements of $N'$ so that

$$\underbrace{0 \to g_{11} \to g_{12} \to \cdots}_{N'^{1}} \quad \underbrace{g_{21} \to g_{22} \to \cdots}_{N'^{2}\backslash N'_{2}} \cdots \underbrace{g_{\omega 1} \to g_{\omega 2} \to \cdots}_{N'^{\omega}\backslash N'_{\omega}} \cdots$$

For each $\theta \neq \beta \in \mathscr{B}$ there exists a least element $L(\beta)$ in this well ordering such that $L(\beta)\beta \neq 0$. Define $\beta$ positive if $L(\beta)\beta > 0$. As before this orders $\mathscr{B}$. Thus to complete the proof it suffices to show that if $\beta$ is positive, then $\alpha\beta$ is positive for all $\alpha \in A(N')$. Let $g \in N'^{\gamma}/N'_{\gamma}$. Then there exist positive integers $m$ and $n$ such that $n(g\alpha) = mg + d$, where $d \in N'_{\gamma}$, hence $d \to g$. If $g \to L(\beta)$, then

$$n(g\alpha\beta) = (mg + d)\beta = m(g\beta) + d\beta = 0.$$

Thus $g\alpha\beta = 0$. If $g = L(\beta)$, then

$$n(L(\beta)\alpha\beta) = (mL(\beta) + d)\beta = m(L(\beta)\beta) + d\beta = m(L(\beta)\beta) > 0 .$$

Thus $L(\beta)\alpha\beta > 0$.

COROLLARY. *If $N$ is in the center of $G$, $A(N)$ can be ordered and $N' = R$, then $A(G)$ can be ordered.*

One should be careful not to place too many restrictions on $G$. For $A(G)$ may become trivial (consist of the identity only). de Groot [**6**] has shown that exist $2^c$ non-isomorphic archimedean o-groups that admit only the identity automorphism. Suppose that $G$ admits no proper o-automorphism and that $N'$ and $N$ are non-trivial. Then, since an inner automorphism is an o-automorphism, $G$ is abelian. Hence $N$ is in the center of $G$. Thus in order to construct a non-archimedean o-group that admits only the trivial o-automorphism, it suffices to find non-trivial subgroups $N'$ and $N$ of $\mathbf{R}$ such that neither admit proper o-automorphisms and the only homomorphism of $N'$ into $N$ is $\theta$. Then $G = N' \oplus N$ will do. One such pair is

$$N = I \text{ and } N' = \{m/2^n : m, n \in I\}e + \{p/3^q : p, q \in I\} ,$$

where $e$ is trancendental.

*For the remainder of this section assume that $G$ is a splitting extension of $N$ by $N'$ and that $N \subseteq$* $\mathbf{R}$. Without loss of generality $f(a', b') = 0$ for all $a', b'$ in $N'$ and $A(N) \subseteq \mathbf{P}$. Thus $r(b'), \gamma \in \mathbf{P}$, and $ar(b'), a\gamma$ represent ordinary multiplication, where $a \in N, b' \in N'$ and $\gamma \in A(N)$. In particular, (2) and (3) reduce to

$$r(b') = r(b'\alpha), \text{ and} \tag{2'}$$

$$(a' + b')\beta = a'\beta r(b') + b'\beta . \tag{3'}$$

Pick an element $k \in N$ and define $x'\beta = k(r(x') - 1)$ for all $x' \in N'$. $a'\beta r(b') + b'\beta = k(r(a') - 1)r(b') + k(r(b') - 1) = k(r(a')r(b') - 1) = k(r(a' + b') - 1) = (a' + b')\beta$. Thus $\beta \in \mathscr{B}$. Suppose that there exists an element $a'$ in the center of $N'$ such that $r(a') \neq 1$. Let $x'$ be any other element of $N'$, and let $\beta \in \mathscr{B}$. Then $x'\beta r(a') + a'\beta = (x' + a')\beta = (a' + x')\beta = a'\beta r(x') + x'\beta$. Thus $x'\beta(r(a') - 1) = a'\beta(r(x') - 1)$ or

$$(6) \qquad x'\beta = \left[\frac{a'\beta}{r(a') - 1}\right][r(x') - 1] .$$

Therefore $\beta$ is determined by $a'\beta$.

LEMMA 3.2. *If there exists an element $a'$ in the center of $N'$ such that $r(a') \neq 1$, then $\mathscr{B}$ is isomorphic to a subgroup of* **R** *that contains $N$.*

*Proof.* For $\beta \in \mathscr{B}$ we define $\beta\sigma = (a'\beta)/(r(a') - 1)$. Then

$$\begin{aligned}(\beta_1 + \beta_2)\sigma &= a'(\beta_1 + \beta_2)/(r(a') - 1) = (a'\beta_1)/(r(a') - 1) \\ &\quad + (a'\beta_2)/(r(a') - 1) = \beta_1\sigma + \beta_2\sigma .\end{aligned}$$

If $0 = \beta\sigma = (a'\beta)/(r(a') - 1)$, then $a'\beta = 0$. Thus by (6), $\beta = \theta$. Therefore $\sigma$ is an isomorphism of $\mathscr{B}$ into **R**, and by the preceding discussion $\mathscr{B}\sigma \supseteq N$.

If $r(a') < 1$, then $1 < r(a')^{-1} = r(-a')$. Thus we may assume that $r(a') - 1 > 0$. Define $\beta \in \mathscr{B}$ positive (notation) $\beta > \theta$) if $\beta\sigma > 0$. Then $\mathscr{B}$ is ordered and $A(N) \subseteq \mathbf{P}$ has a natural order. $\beta\sigma = (a'\beta)/(r(a') - 1) > 0$ if and only if $a'\beta > 0$. Thus $\beta > \theta$ if and only if $a'\beta > 0$. Suppose that $A(N')$ is also ordered. Then Definition 3.1 orders $A(G)$ if we can show that $\overline{\beta} > \theta$ implies that $\alpha^{-1}(\beta + \overline{\beta}\gamma - \beta) > \theta$ for all $\overline{\beta} \in \mathscr{B}$, and all $\pi = (\alpha, \beta, \gamma) \in A(G)$. But

$$\begin{aligned}a'\alpha^{-1}(\beta + \overline{\beta}\gamma - \beta) &= a'\alpha^{-1}\overline{\beta}\gamma = ((a'\alpha^{-1})\overline{\beta})\gamma \\ &= [(a'\overline{\beta})(r(a'\alpha^{-1}) - 1)/(r(a') - 1)]\,\gamma = a'\overline{\beta}\gamma .\end{aligned}$$

But since $a'\overline{\beta} > 0$ we have $a'\overline{\beta}\gamma > 0$.

THEOREM 3.3. *If $G$ splits over $N$, $N \subseteq$* **R**, *$A(N')$ can be ordered and there exists an element $a'$ in the center of $N'$ such that $r(a') \neq 1$, then $A(G)$ can be ordered.*

COROLLARY. *If $H$ is a non-abelian splitting o-extension of a subgroup of* **R** *by a subgroup of* **R**, *then $A(H)$ can be ordered.*

This is an immediate consequence of the theorem. If $N' = \mathbf{R}$, then (2′) is equivalent to $1 = r(b'(\alpha - 1))$. Hence either $r = \theta$ or $\alpha = 1$. Thus if $N' = \mathbf{R}$, then this corollary is an immediate consequence of Theorem 3.1.

**4. Ordered extension of subgroups of R.** Throughout this section assume that $N$ is a subgroup of $\mathbf{R}$ and that $N'$ is abelian. In particular, $r$ is a homomorphism of $N'$ into the group $A(N)$, and without loss of generality $A(N) \subseteq \mathbf{P}$ and $ar(b')$ is ordinary multiplication, where $a \in N$ and $b' \in N$.

$$(a', a) + (0, b) = (a', a + b) \text{ and } (0, b) + (a', a) = (a', br(a') + a) .$$

*These are equal if and only if $br(a') = b$. Thus $G$ is a central extension* of $N$ by $N'$ if and only if $r = \theta$.

LEMMA 4.1. *Suppose that $N'$ is d-closed. Then there exists a non-central o-extension of $N$ by $N'$ if and only if there exists $1 \neq p \in \mathbf{P}$ such that $p^s N = N$ for all $s \in R$.*

*Proof.* First suppose that $G$ is a non-central o-extension of $N$ by $N'$. Then $r \neq \theta$. Pick $a' \in N'$ so that $1 \neq r(a') = p \in \mathbf{P}$. For each positive integer $n$ there exists $b' \in N'$ such that $nb' = a'$. Hence $p = r(a') = r(nb') = r(b')^n$. Thus $r(b') = p^{1/n}$. For $m \in I$, we have $r(mb') = r(b')^m = p^{m/n}$. Thus $p^{m/n}N = N$ for all rational numbers $m/n$.

Conversely suppose that there exists $1 \neq p \in \mathbf{P}$ such that $p^s N = N$ for all $s \in R$. Pick $0 \neq b' \in N'$. Then $N' = Rb' \oplus D$, where $Ra'$ is the one dimensional subspace of $N'$ that contains $a'$ and $D$ is a subspace of $N'$. Each $a' \in N'$ has a unique representation $a' = sb' + d$, where $s \in R$ and $d \in D$. Define $q(a') = p^s$. Then $H = N' \times N$, where $(a', a) + (b', b) = (a' + b', aq(b') + b)$ is a splitting extension of $N$ by $N'$ that is not a central extension.

COROLLARY. *If $N'$ is d-closed and $N \subseteq R$, then $G$ is a central extension of $N$ by $N'$.*

THEOREM 4.1. *Suppose that $r \neq \theta$. Then $G$ splits over $N$ if and only if there exist $a' \in N'$ and $a \in N$ such that*

(a) $r(a') \neq 1$

(b) $[1/(r(a') - 1)][a(r(b') - 1) + f(a', b') - f(b', a')] \in N$ for all $b' \in N'$.

*Proof.* First suppose that $G$ splits. Choose a group $H$ of representatives of $G/N$, and pick one element $(a', a)$ of $H$ such that $r(a') \neq 1$. Let $(b', b)$ be any other element of $H$. Then since $H$ is abelian,

$$\begin{aligned}(b' + a', f(b', a') + br(a') + a) &= (b', b) + (a', a) = (a', a) + (b', b)\\ &= (a' + b', f(a', b') + ar(b') + b) .\end{aligned}$$

Thus

$$b(r(a') - 1) = a(r(b') - 1) + f(a', b') - f(b', a') .$$

$(b)$ is satisfied because

$$[1/(r(a') - 1)][a(r(b') - 1) + f(a', b') - f(b', a')] = b .$$

Note that

$$H = \{(b', [1/(r(a') - 1)][a(r(b') - 1) + f(a', b') - f(b', a')]) : b' \in N'\} .$$

Thus $H$ is uniquely determined by $(a', a)$.

Conversely suppose that $a' \in N'$ and $a \in N$ satisfy (a) and (b). Let

$$S = \{(b', b) \in G : (b', b) + (a', a) = (a', a) + (b', b)\} .$$

Clearly $S$ is a group. By the above computation it follows that $(b', b) \in S$ if and only if

$$b = [1/(r(a') - 1)][a(r(b') - 1) + f(a', b') - f(b', a')] .$$

Thus for each $b' \in N'$ there is one and only one $(b', b)$ in $S$. Therefore $S$ is a group of representatives for $G/N$.

The factor mapping $f$ is *symmetric* (*skew-symmetric*) if $f(a', b') = f(b', a')$ $(f(a', b') = -f(b', a'))$ for all $a', b'$ in $N'$.

COROLLARY I. *If $r \neq \theta$ and $f$ is symmetric, then $G$ splits. Moreover $f(a', b') = 0$ for all $a', b'$ in $N'$.*

*Proof.* Pick $a' \in N'$ such that $r(a') \neq 1$ and let $a = 0$. Then (a) and (b) are satisfied, hence $G$ splits. Also by the proof of the converse of the theorem, $S = \{(b', 0) : b' \in N'\}$ is a group of representatives. Thus $(a', 0) + (b', 0) = (a' + b', f(a', b')) \in S$. Therefore $f(a', b') = 0$

Let $f(N', N')$ denote the range of $f$.

COROLLARY II. *If there exists an $a' \in N'$ such that $r(a') \neq 1$ and $[1/(r(a') - 1)]f(N', N') \subseteq N$, then $G$ splits.*

*Proof.* Let $a = 0$. Then (a) and (b) are satisfied. Moreover, $\{(b', [1/(r(a') - 1)][f(a', b') - f(b', a')])\}$ is a group of representatives.

COROLLARY III. *If $N$ is a field and $r \neq \theta$, then $G$ splits.*

*Proof.* Pick $a' \in N'$ such that $r(a') \neq 1$. Since $1 \in N$ and $r(a')N = N$, $r(a') \in N$. Thus $1/(r(a') - 1) \in N$ and

$$[1/(r(a') - 1)]f(N', N') \subseteq [1/(r(a') - 1)]N = N .$$

REMARK. Rich [13] proved that if $N \subseteq \mathbf{R}$, $N' = \mathbf{R}$ and $r \neq \theta$, then

$G$ splits. This is a special case of Corollary III. Corollary III can be stated independently of the representation of $G$ as follows: If $H$ is an o-group, $C$ is a convex subgroup of $H$ that is o-isomorphic to the additive group of a subfield of $\mathbf{R}$, and $H/C$ is abelian, then either $H$ is a splitting extension of $C$ or $H$ is a central extension of $C$.

COROLLARY IV. *If there exists an* $a' \in N'$ *such that* $r(a') = (n+1)/n$ *for some positive integer* $n$, *then* $G$ *splits.*

*Proof.* $1/(r(a')-1)=n$. Thus $[1/(r(a')-1)]f(N', N')=nf(N', N')\subseteq N$.

COROLLARY V. *If* $N$ *is d-closed and there exists an* $a' \in N'$ *such that* $1 \neq r(a')$ *is rational, then* $G$ *splits.*

*Proof.* $1/(r(a')-1)$ is rational, hence $[1/(r(a')-1)]N \subseteq N$.

By Theorem 3.3 [3, p. 522] there exists an $a$-extension $H$ of $G$ such that the convex subgroup $K$ of $H$ that covers 0 is o-isomorphic to $\mathbf{R}$ and $H/K$ is o-isomorphic to $N'$. Thus by Theorem 4.1 either $H$ is a splitting extension of $K$ or $H$ is a central extension of $K$.

REMARK. If $H$ is a splitting o-extension of $K$, then without loss of generality $H = N' \times \mathbf{R}$, where $(a', a) + (b', b) = (a' + b', \quad as(b') + b)$. $s$ is a homomorphism of $N'$ into $\mathbf{P}$. For each $x$ in $D(N)$ there exists a positive integer $n$ such that $nx \in N'$. Define $t(x) = [s(nx)]^{1/n}$. Then $t$ is the unique extension of $s$ to a homomorphism of $D(N')$ into $\mathbf{P}$. $D(N')$, $\mathbf{R}$ and $t$ determine a splitting o-extension $M$ of $\mathbf{R}$ by $D(N)$. $M$ is an $a$-extension of $H$ and $M$ is $d$-closed. Thus by Theorem 3.2 [3 p. 519] there exists an $a$-closed a-extension $Q$ of $M$ with each component $o$-isomorphic to $\mathbf{R}$. $Q$ is an $a$-extension of $G$.

A mapping $g$ of $N' \times N'$ into $N$ is called *bilinear* if for all $x, y, z$ in $N'$

$$g(x+y, z) = g(x, z) + g(y, z)\ ,$$

and

$$g(x, y+z) = g(x, y) + g(x, z)\ .$$

Yamabe [16] and the Neumanns [12] have shown that if $N = I$, and the cardinality of $N'$ is at most $\aleph_1$, and $g$ is bilinear and satisfies $g(x, x) = 0$ only if $x = 0$, then $N'$ is a free abelian group. Hughes [7] has classified the groups of class 2 in terms of some special bilinear mappings. Iwasawa gives an example ([8] Example 2, p. 7) of an $o$-group that is determined by a bilinear mapping. For let $N'=I\times I$ and $N=I$. Define $g((a,b), (x,y))= ay$. Then $G=I\times I\times I$, where $(a,b,c)+(x,y,z)=(a+x, b+y, ay+c+z)$, and $(a,b,c)$ is positive if $a>0$ or $a=0$ and $b>0$ or $a=b=0$ and $c>0$, is an $o$-group of rank 3 that is isomorphic with Iwasawa's example. In fact, $G$ is generated by $a=(0,0,1)$, $b=(0,1,0)$ and $c=(1,0,0)$ and has generating relations $a+b=b+a$, $a+c=c+a$ and $c+b-c=a+b$.

The last example can be generalized because the bilinear form is a product of homomorphisms. For example, let $N$ be the additive group of an ordered ring, and let $\sigma$ and $\tau$ be homomorphisms of $N'$ into $N$. For $a', b'$ in $N'$ define $g(a', b') = \sigma(a')\tau(b')$. Then $H = N' \times N$, where $(a', a) + (b', b) = (a' + b', g(a', b') + a + b)$ is a central extension of $N$ by $N'$.

LEMMA 4.2. *If $f$ is bilinear, then $G$ is a splitting extension of $N$ or $G$ is a central extension of $N$.*

*Proof.* For $x, y, z$ in $N'$ we have

$$f(x,y)+f(x,z)+f(y,z)=f(x,y+z)+f(y,z)=f(x+y,z)+f(x,y)r(z)$$
$$=f(x,z)+f(y,z)+f(x,y)r(z)\,.$$

Therefore $f(x,y) \equiv f(x,y)r(z)$. Thus either $r(z) \equiv 1$ or $f(x,y) \equiv 0$.

COROLLARY. *If $N$ is abelian* (not necessarily a subgroup of $\mathbf{R}$), *$f$ is bilinear and $f(N', N')$ generates $N$, then $G$ is a central extension of $N$.*

**5. Central extensions and bilinear mappings.** Throughout this section assume that $N$ is in the center of $G$. Thus $G$ is determined by the $o$-group $N'$, the abelian $o$-group $N$, and the factor mapping $f: N' \times N' \rightarrow N$ that satisfies

(1) $$f(0, b') = f(a', 0) = 0, \text{ and}$$

(2) $$f(a' + b', c') + f(a', b') = f(a', b' + c') + f(b', c')\,.$$

In particular, any central extension of $N$ by $N'$ can be ordered. A central extension $H$ of $N$ by $N'$ with factor mapping $h$ is *equivalent* to $G$ (notation $H \sim G$) if there exists an isomorphism $\alpha$ of $H$ onto $G$ such that $(0, a)\alpha = (0, a)$ and $(a', a)\alpha \equiv (a', a)$ modulo $0 \times N$ for all $a$ in $N$ and all $a'$ in $N'$. If $H$ is ordered in the usual way, then $\alpha$ is an $o$-isomorphism. It is well known that $H \sim G$ if and only if there exists $t: N' \rightarrow N$ such that $t(0) = 0$ and

$$f(a', b') = h(a', b') - t(a' + b') + t(a') + t(b')$$

for all $a', b'$ in $N'$. In particular, $G \sim N' \oplus N$ if and only if there exists $t: N' \rightarrow N$ such that $t(0) = 0$ and $f(a', b') = -t(a' + b') + t(a') + t(b')$ for all $a', b'$ in $N'$.

It is easy to verify that if $g$ is a bilinear mapping of $N' \times N'$ onto $N$, then $g$ satisfies (1) and (2). Moreover, such a $g$ exists if and only if we can choose a representative function $r: N' \to G$ such that

$$r(a' + b' + c') = r(a' + b') + r(a' + c') + r(b' + c') - r(a') - r(b') - r(c')$$

for all $a', b', c'$ in $N'$. From (2) we have

$$f(a' + b', c') - f(a', c') - f(b', c') = f(a', b' + c') - f(a', b') - f(a', c') .$$

Thus $f$ is bilinear if it is linear in one variable.

LEMMA 5.1. *Suppose that $f$ is bilinear, then for $a, b$ in $N$ and $a', b', c'$ in $N'$ we have:*

(i) $-f(a', b') = f(-a', b') = f(a', -b')$.

(ii) $f(a', b') = f(-a', -b')$.

(iii) $(a', a) + (b', b) - (a', a) - (b', b) = (a' + b' - a' - b', f(a', b') - f(b', a'))$.

For $0 = f(a' - a', b') = f(a', b') + f(-a', b')$. Thus $-f(a', b') = f(-a', b')$ and similarly $-f(a', b') = f(a', -b')$. (ii) is an immediate consequence of (i), and (iii) follows by computing the left hand side.

Let $D(N)$ be the $d$-closure of $N$, and let $H = N' \times D(N)$. For $(a', a)$ and $(b', b)$ in $H$ define $(a', a) + (b', b) = (a' + b', f(a', b') + a + b)$. Then $H$ is a central extension of $D(N)$ by $N'$, and $G$ is a subgroup of $H$. There is a natural extension of the ordering of $G$ to an ordering of $H$. If $G \sim N' \oplus N$, then $H \sim N' \oplus D(N)$, but the converse is false. For in [2] there is an example where $N' = D(N) = R$, $H \sim N' \oplus N$ and $G x N' \oplus N$ [2, p. 862].

THEOREM 5.1. *Suppose that $N'$ is abelian and let $H = D(N') \times D(N)$. Also suppose that for all $a', b'$ in $N'$ and for all positive integers $n, f$ satisfies*

$$(3) \qquad nf(a', b') = f(na', b') = f(a', nb') .$$

*Then there exists a unique $g: D(N') \times D(N') \to D(N)$ that satisfies (3) and such that $g(a', b') = f(a', b')$ for all $a', b'$ in $N'$. For $(x, y)$ and $(u, v)$ in $H$ define $(x, y) + (u, v) = (x + u, g(x, u) + y + v)$.*

*(a) $H$ is a central extension of $D(N)$ by $D(N')$, and $G$ is a subgroup of $H$.*

*(b) $H$ is $d$-closed.*

*(c) For each $h$ in $H$ there exists a positive integer $n = n(h)$ such that $nh \in G$.*

*(d) There exists a unique extension of the ordering of $G$ to an ordering of $H$. $H$ will be called the $d$-closure of $G$.*

*Proof.* For each pair $x, y$ in $D(N')$ there exists a positive integer

$n = n_{x,y}$ such that $nx, ny \in N'$, define $g(x, y) = (1/n^2)f(nx, ny)$. This definition is independent of the particular choice of $n$. For if $mx, my \in N'$, then $m^2f(nx, ny) = f(mnx, mny) = n^2f(mx, my)$. Thus $(1/n^2)f(nx, ny) = (1/m^2)f(mx, my)$. Let $x, y, z \in D(N')$ and choose a positive integer $n$ such that $nx, ny, nz, n(x + y)$, and $n(y + z)$ belong to $N'$. Then

$$\begin{aligned} g(x + y, z) + g(x, y) &= (1/n^2)[f(nx + ny, nz) + f(nx, nz)] \\ &= (1/n^2)[f(nx, ny + nz) + f(ny, nz)] = g(x, y + z) + g(y, z) . \end{aligned}$$

By a similar argument $g$ satisfies (1) and (3). Also if $g'$ is any other extension of $f$ to $D(N') \times D(N')$ that satisfies (3), then $n^2g'(x, y) = g'(nx, ny) = f(nx, ny)$. Therefore $g'(x, y) = (1/n^2)f(nx, ny) = g(x, y)$ for all $x, y$ in $D(N')$.

Clearly (a) is satisfied. To prove (b) it suffices to show that $n(x, y) = (a, b)$ has a solution in $H$, where $n$ is a positive integer and $(a, b) \in H$. By induction

$$n(x, y) = (nx, [(n - 1)n/2]g(x, x) + ny) .$$

Thus $x = (1/n)a$ and

$$y = (1/n)(b - [(n - 1)n/2]g((1/n)a, (1/n)a))$$

is a solution. Consider $(x, y)$ in $H$, and let $m$ be a positive integer such that $mx \in N'$ and $my \in N$. Then

$$\begin{aligned} 2m^2(x, y) &= (2m(mx), (2m^2 - 1)m^2g(x, x) + 2m(my)) \\ &= (2m(mx), (2m^2 - 1)f(mx, my) + 2m(my)) \in G . \end{aligned}$$

Thus (c) is satisfied. The orderings of $N$ and $N'$ can be uniquely extended to orderings of $D(N)$ and $D(N')$. Define $(x, y) \in H$ positive if $x > 0$ or $x = 0$ and $y > 0$. This extends the ordering of $G$ to an ordering of $H$. But for any extension of the order of $G$, $h \in H$ is positive if and only if $nh$ is positive in $G$, where $n$ is a positive integer such that $nh \in G$. Thus this extension is unique.

REMARK. If $f$ is bilinear or symmetric or skew-symmetric, then so is $g$. By Theorem 3.2 [3, p. 519] there exists an $a$-closed $a$-extension of $H$ with each component $o$-isomorphic to $\mathbf{R}$.

Suppose that $f$ is bilinear. Let $x, y, z \in N'$ and let $w = x + y - x - y$. Then

$$f(w, z) + f(y, z) + f(x, z) = f(w + y + x, z) = f(x + y, z) = f(x, z) + f(y, z) .$$

Thus $f(w, z) = 0$. Similarly $f(z, w) = 0$. Therefore $f(c, z) = f(z, c) = 0$ for all $z$ in $N'$ and all $c$ in the commutator subgroup of $N'$.

LEMMA 5.2. *If $f$ is bilinear and $N'$ coincides with its commutator* *group, then* $G = N' \oplus N$.

Newmann [11] exhibits an o-group that coincides with its commutator group.

Suppose that $2N = N$ and $f$ is bilinear. Let $p(x, y) = (1/2)[f(x, y) + f(y, x)]$ and let $q(x, y) = (1/2)[f(x, y) - f(y, x)]$ for all $x, y$ in $N'$. Then $p(q)$ is a symmetric (skew-symmetric) bilinear mapping of $N' \times N'$ into $N$, and $f(x, y) = p(x, y) + q(x, y)$. Moreover, as in matrix theory, this representation is unique.

THEOREM 5.2. *If* $2N = N$ *and* $f$ *is bilinear, then* $G \sim H$, *where* $H$ *is the central extension of* $N$ *by* $N'$ *that is determined by the skew-symmetric part* $q$ *of. If* $f$ *is symmetric, then* $G \sim N' \oplus N$. *Thus if* $G$ *is abelian, then* $G \sim N' \oplus N$.

*Proof.* For each $x$ in $N'$ define $t(x) = (-1/2)f(x, x)$. Then

$$\begin{aligned} &-t(x + y) + t(x) + t(y) + q(x, y) \\ &\quad = (1/2)[f(x + y, x + y) - f(x, x) - f(y, y) + f(x, y) - f(y, x)] = f(x, y) . \end{aligned}$$

Thus $G \sim H$. If $f$ is symmetric, then $H = N' \oplus N$, and if $G$ is abelian, then $f$ is symmetric.

Suppose that $N$ and $N'$ are abelian and that $f$ is bilinear. Then by Theorem 5.1, we can embed $G$ into its $d$-closure $H = D(N') \times D(N)$. The factor mapping $g$ associated with $H$ is bilinear, and by Theorem 5.2 we may choose $g$ so that it is skew-symmetric and bilinear. Moreover, $sg(x, y) = g(sx, y) = g(x, sy)$ for all $s \in R$ and for all $x, y$ in $D(N)$. For

$$ng((m/n)x, y) = g(n(m/n)x, y) = g(mx, y) = mg(x, y) .$$

Thus $(m/n)g(x, y) = g((m/n)x, y)$. Let $\alpha_1, \alpha_2, \cdots$ be a basis for the rational vector space $D(N')$ and consider $X = x_1\alpha_{s_1} + \cdots + x_m\alpha_{s_m}$ and $Y = y_1\alpha_{t_1} + \cdots + y_n\alpha_{t_n}$ in $D(N')$. Then

$$g(X, Y) = \sum x_i g(\alpha_{s_i}, \alpha_{t_j}) y_j$$

Thus $g$ is determined by the skew symmetric matric $A = [g(\alpha_i, \alpha_j)]$ with components in $D(N)$. Conversely any such matric determines a bilinear skew-symmetric factor mapping of $D(N') \times D(N')$ into $D(N)$.

THEOREM 5.3. *If* $N'$ *is abelian and* $f$ *is bilinear, then* $G$ *is a subgroup of its d-closure* $H$ *and* $H$ *is completely determined by* $N, N'$ *and a skew symmetric matrix with entries from* $D(N)$. *The dimension of this matrix is equal to the rank of the vector space* $D(N')$.

If the rank of $D(N')$ is finite, say $n$, and $D(N) = R$, then by a suitable choice of coordinates for $D(N')$ we can get a canonical form for $A$.

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdot & \cdot & \cdot \\ -1 & 0 & 1 & \cdot & \cdot & \cdot \\ 0 & -1 & 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Thus $H$ is determined by $n$ and the rank of $A$. For example if $N' = R \times R \times R$ and $N = R$, then we have two non-trivial choices for $f$. One of which is

$$f((x_1, x_2, x_3), (y_1, y_2, y_3)) = [x_1 x_2 x_3] \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = -x_2 y_1 + (x_1 - x_3) y_2 + x_2 y_3 ,$$

and the other is obtained by using the cannonical matrix of rank 2. Thus for any ordering of $N'$ we have at least two non-trivial central $o$-extensions of $N$ by $N'$.

LEMMA 5.3. *If $A$ and $B$ are elements of an ordered semigroup $S$ and $A + B < B + A$, then $nA + nB < n(A + B) < n(B + A) < nB + nA$ for all integers $n$ greater than* 2.

*Proof.* If

$$A + (n-1)A + (n-1)B + B = nA + nB \geq n(A + B) = A + (n-1)(B + A) + B ,$$

then $(n-1)A + (n-1)B \geq (n-1)(B + A)$. If

$$B + (n-1)(A + B) + A = n(B + A) \geq nB + nA = B + (n-1)B + (n-1)A + A ,$$

then $(n-1)(A + B) \geq (n-1)B + (n-1)A$. Thus the lemma follows immediately by induction on $n$.

THEOREM 5.4. *If $1 \in N' \subseteq R$, then $G$ is abelian.*

*Proof.* By a simple induction argument (see [9] p. 265), $f(x, y) = f(y, x)$ for all integers $x$ and $y$. Let $A = (a', a)$ and $B = (b', b)$ be elements of $G$. Then since $a'$ and $b'$ are rational numbers, there exists a positive integer $n$ such that $nA = (x', x)$ and $nB = (y', y)$, where $x'$ and $y'$ are integers.

$$nA + nB = (x' + y', f(x', y') + x + y) = (y' + x', f(y', x') + y + x) = nB + nA .$$

Thus by Lemma 5.3, we have $A + B = B + A$.

**6. o-groups of rank 2.** Throughout this section we assume that $N$ and $N'$ are subgroups of $\mathbf{R}$. By Theorem 3.5 [3 p. 523] there exists an $a$-closed $a$-extension $H$ of $G$ such that both components are $o$-isomorphic to $\mathbf{R}$. By Theorem 4.1, either $H$ is a central extension of $\mathbf{R}$ or $H$ is a splitting extension of $\mathbf{R}$. A splitting $o$-extension of $\mathbf{R}$ by $\mathbf{R}$ is determined by a homomorphism of $\mathbf{R}$ into $\mathbf{P}$. If $H$ is a central extension of $\mathbf{R}$ by $\mathbf{R}$ with a bilinear factor mapping, then $H$ is determined by a skew-symmetric real matrix.

If $N'$ is cyclic, then $G$ is a splitting extension of $N$. Thus if $N'$ is cyclic and $N$ admits no proper $o$-automorphisms, then $G = N' \oplus N$. In particular, if $N' = N = I$, then $G = N' \oplus N$. In fact, as Loonstra [9] shows, there are only two normal extensions of $I$ by $I$ (not necessarily ordered) For if $H$ is a normal extension of $I$ by $I$, then $H$ splits over $I$. Thus $H = I \times I$ and $(a', a) + (b', b) = (a' + b', as(b') + b)$, where $s$ is a homomorphism of $I$ into the multiplicative group $\{1, -1\}$. Either $s(1) = 1$ or $s(1) = -1$. If $s(1) = 1$, then $s = \theta$ and $H = I \oplus I$. If $s(1) = -1$, then $s(2n) = 1$ and $s(2n+1) = -1$ for all $n \in I$. Thus the addition rule for $H$ is

$$(x, y) + (2m, n) = (x + 2m, y + n)$$
$$(x, y) + (2m + 1, n) = (x + 2m + 1, n - y)\ .$$

In this case $H$ can't be ordered because $-(1,0) + (0,1) + (1,0) = -(0,1)$. Thus $(0, 1)$ can't be positive or negative.

If $N = N' = R$, then $G$ is $o$-isomorphic to $R \oplus R$. For by Lemma 4.1, $G$ is a central extension of $N$ and by Theorem 5.4, $G$ is abelian. Thus $G$ is an abelian $o$-group of rank 2 with both components $o$-isomorphic to $R$. By Hahn's embedding theorem (see [2]) $G$ is $o$-isomorphic to $R \oplus R$.

*Example of a non-abelian o-group of rank* 2 *that is isomorphic to its group of o-automorphims.* Let $N = N' = \mathbf{R}$. For $a', b' \in N'$ define $f(a', b') = 0$ and $r(a') = e^{a'}$, where $e$ is transcendental. Then $(a', a) + (b', b) = (a' + b', ae^{b'} + b)$. By the remark at the end of § 3, an $o$-automorphism $\pi$ of $G$ has a representation $\pi = \begin{bmatrix} 1 & \beta \\ 0 & C \end{bmatrix}$, where $C \in \mathbf{P}$ and $x'\beta = 1\beta(e^{x'} - 1)/(e - 1) = \beta\sigma(e^{x'} - 1)$ for all $x' \in N'$. The mapping of $\pi$ onto $\begin{bmatrix} 1 & \beta\sigma \\ 0 & C \end{bmatrix}$ is an isomorphism of $A(G)$ onto the multiplicative group $A = \left\{\begin{bmatrix} 1 & B \\ 0 & C \end{bmatrix} : B \in \mathbf{R} \text{ and } C \in \mathbf{P}\right\}$. The mapping of $(a', a) \in G$ onto $\begin{bmatrix} e^{a'} & 0 \\ a & 1 \end{bmatrix}$ is an isomorphism of $G$ onto the multiplicative group $B = \left\{\begin{bmatrix} x & 0 \\ y & 1 \end{bmatrix} : x \in \mathbf{P} \text{ and } y \in \mathbf{R}\right\}$. The mapping of $\begin{bmatrix} x & 0 \\ y & 1 \end{bmatrix}$ onto $\begin{bmatrix} x & 0 \\ y & 1 \end{bmatrix}^{-1}$ is an isomorphism of $A$ onto $B$. Therefore $G$ is isomorphic to $A(G)$. In particular, there exists a non-trivial splitting $o$-extension of $G$ by $G$.

We conclude by giving an *example of an o-group of rank* 2 *that is not a central extension nor a splitting extension of its convex subgroup.* Let $G$ be the o-group of the last example, and let $H$ be the subgroup of $G$ that is generated by $\{(a, a): a \in R\}$. We have $(-1, -1) + (1, 1) = (0, 1 - e)$. Thus $H$ has rank 2.

$$(1, 1) + (0, 1 - e) = (1, 2 - e) \neq (1, e - e^2 + 1) = (0, 1 - e) + (1, 1) .$$

Thus $H$ is not a central extension.

LEMMA. *If* $(b', b) \in H$, *then* $b = \sum_1^m b_i e^{c_i}$, *where* $b_i, c_i \in R$ *and* $\sum_1^m b_i = b'$. *For* $(b', b) = P_1 + P_2 + \cdots + P_n$, *where* $P_i$ *or* $-P_i$ *is a generator. A simple induction on* $n$ *proves the lemma. In particular,* $(b', 0) \in H$ *only if* $b' = 0$. *It can be shown that* $H = \{(a, \sum a_i e^{b_i}): a, a_i, b_i \in R$ *and* $\sum a_i = a\}$, *but we will not need this.*

Now suppose (by way of contradiction) that $H$ is a splitting extension of its convex subgroup $C$. Pick a group $K$ of representatives of $H/C$, and let $(1, a)$ be the element in $K$ with first component 1. $a = \sum_1^j a_i e^{b_i}$, where $\sum_1^j a_i = 1$. In particular, $a \neq 0$. By the proof of Theorem 4.1

$$K = \{(b', a(e^{b'} - 1)/(e - 1)): b' \in R\} .$$

Let $d$ be the least common multiple of the denominators of the $a_i$ and let $b' = 1/p$, where $p$ is a prime and $p > d$. Then $d(\sum a_i e^{b_i}) = \sum c_i e^{b_i}$ has integral coefficients. By the above lemma

$$\frac{\left(\sum_1^j c_i e^{b_i}\right)(e^{b'} - 1)}{e - 1} = \sum_1^k e_i e^{d_i} \tag{1}$$

where $e_i, d_i \in R$. Let $q$ be a positive common multiple of $p$ and the denominators of the $b_i$ and the $d_i$. Then

$$\frac{\left[\sum_1^j c_i (e^{1/q})^{u_i}\right][(e^{1/q})^v - 1]}{(e^{1/q})^q - 1} = \sum_1^k e_i (e^{1/q})^{wq_i} \tag{2}$$

where $u_i, w_i, v \in I$. Without loss of generality we may assume that the $u_i$ and the $w_i$ are positive integers (multiply both sides of (2) by a suitable power of $e^{1/q}$). $e^{1/q}$ is trancendental. Thus (2) is essentially an equality of elements in the simple transcendental field extension $R(X)$ of $R$.

$$\frac{\left[\sum_1^j c_i X^{u_i}\right][X^v - 1]}{X^q - 1} = \sum_1^k e_i X^{w_i} \tag{3}$$

$b' = 1/p = v/q = v/pv$. Thus there exists a positive integer $n$ such that $p^n$ divides $q$, but $p^n$ does not divide $v$. The cyclotomic polynomial

$$f(X) = 1 + X^{p^{n-1}} + X^{2p^{n-1}} + \cdots + X^{(p-1)p^{n-1}}$$

is an irreducible factor of $X^q - 1$, but it does not divide $X^v - 1$. Therefore $f(X)$ divides $\sum c_i X^{u_i}$. Thus $\sum c_i X^{u_i} = f(X)g(X)$, where $g(X)$ is a polynomial with integral coefficients. Now let $X = 1$. Then $d = \sum_1^j c_i = f(1)g(1) = pg(1)$. Thus since $p$ and $d$ are positive and $g(1)$ is an integer, $d \geq p$. But this contradicts our choice of $p$.

Note that the example on page 526 of [3] is a splitting extension of $N$ by $N'$; and that $\{(a', -1): 0 \neq a' \in N'\} \cup \{0, 0)\}$ is a group of representatives.

## References


1. C. G. Chehata, *An algebraically simple ordered group*, Proc. London Math. Soc. **2** (1952), 183-197.
2. A. H. Clifford, *Note on Hahn's theorem on ordered abelian groups*, Proc. Amer. Math. Soc. **5** (1954), 860-863.
3. P. Conrad, *Extensions of ordered groups*, Proc. Amer. Math. Soc. **6** (1955), 516-528.
4. P. Conrad, *The group of order preserving automorphisms of an ordered abelian group*, Proc. Amer. Math. Soc. **9** (1958), 382-389.
5. ———, *A, correction and an improvement of a theorem on ordered groups.* To appear in Proc. Amer. Math. Soc.
6. J. de Groot, *Indecomposable abelian groups*, Indag. Math. **19** (1957), 137-145.
7. N. Hughes, *The use of bilinear mappings in the classification of groups of class* 2, Proc. Amer. Math. Soc. **2** (1951), 742-747.
8. K. Iwasawa, *On linearly ordered groups*, J. Math. Soc. Japan **1** (1948), 1-9.
9. F. Loonstra, *Sur les extensions du groupe additif des entiers rationnels per le mêeme groupe*, Indag. Math. **16** (1954), 263-272.
10. ———, *L'extension du groupe ordonné des entiers rationels par le mêeme groupe*, Indag. Math. **17** (1955), 41-49.
11. B. H. Neumann, *On ordered groups*, Amer. J. Math. **71** (1948), 1-18.
12. ———, and Hanna Neumann, *On a class of abelian groups*, Arch. Math. **4** (1953), 79-85.
13. R. P. Rich, Thesis (unpublished) The John Hopkins University (1950).
14. O. Schreier, *Über die Erweiterung von Gruppen*, Teil I, Monatsh. Math. Phys. **34** (1926), 165-180.
15. A. A. Vinogradov, *On the free product of ordered groups*, Math. Sb. N. S. **25** (1949), 163-168.
16. H. Yamabe, *A condition for an abelian group to be a free abelian group with a finite basis*, Proc. Japan Akad. **27** (1951), 205-207.


THE TULANE UNIVERSITY OF LOUISIANA

# ON THE VAN KAMPEN THEOREM

RICHARD H. CROWELL

**1. Introduction.** The van Kampen theorem provides a defining set of generators and relations for the fundamental group of the union of two topological spaces $X$ and Y where the fundamental groups of $X$, Y, and their intersection are given by defining sets of generators and relations. An intrinsic, purely group-theoretic formulation has been given by Fox using his direct limits of systems of groups [4]; however, the corresponding abstract proof had not been worked out. The present paper supplies such a proof (distilled from an earlier proof by Fox of the van Kampen theorem) to a natural generalization of the van Kampen theorem, which includes for example, in addition to the original theorem, the determination of the fundamental group of the union of an increasing nest of open sets each of whose groups is known [2].

In proving the principal result, Theorem (3.1), we depart from the usual development of the fundamental group in that paths and loops are not required to have the fixed unit interval as domains. In particular, a *path* $a$ is a continuous mapping of the interval $[0, \|a\|]$ into the space in question for some $\|a\| \geq 0$. For paths $a : [0, \|a\|] \to X$ and $b : [0, \|b\|] \to X$ which satisfy $a(\|a\|) = b(0)$, we define the *product path* $a \cdot b$ by

$$a \cdot b(t) = \begin{cases} a(t) \text{ for } 0 \leq t \leq \|a\| \\ b(t - \|a\|) \text{ for } \|a\| \leq t \leq \|a\| + \|b\| . \end{cases}$$

Thus, path multiplication is associative. Paths $a$ and $b$, having the same initial and terminal points, are *equivalent*, denoted by $a \simeq b$, iff there exists a collection of paths $h_s$: $[0, \|h_s\|] \to X$, $0 \leq s \leq 1$, such that

$h_0 = a$ and $h_1 = b$,
$h_s(0) = a(0) = b(0)$,
$h_s(\|h_s\|) = a(\|a\|) = b(\|b\|)$,
$\|h_s\|$ is a continuous function of $s$,
$h_s(t)$ is simultaneously continuous in $s$ and $t$.

We note that, for any path $a$ and positive number $t$, there is a path $b$ equivalent to $a$ with $\|b\| = t$. Furthermore, $\|h_s\|$ can always be taken as a linear function of $s$ and thus, in view of the preceding sentence, may be arranged to be constant. The induced multiplication of equivalence classes of paths and the definitions of the fundamental groupoid and group of $X$ are made in the usual way.

Received July 8, 1958.

**2. Systems of groups and direct limits** (cf. [4]). A *system* is any collection $\mathfrak{S}$ of groups and homomorphisms such that if $\theta ; G_\alpha \to G_\beta$ is in $\mathfrak{S}$, then $G_\alpha$ and $G_\beta$ are in $\mathfrak{S}$. A *homomorphism* $\Phi ; \mathfrak{S} \to G$ *of a system* $\mathfrak{S}$ *into a group* $G$ is a function which assigns to each group $G_\alpha$ in $\mathfrak{S}$ a homomorphism $\varphi_\alpha : G_\alpha \to G$ such that, for every $\theta : G_\alpha \to G_\beta$ in $\mathfrak{S}$, we have $\varphi_\alpha = \varphi_\beta \theta$. The *image* of $\Phi$ is the smallest subgroup of $G$ which contains the image of every homomorphism $\varphi_\alpha$ in $\Phi$, and $\Phi$ is *onto* iff its image is $G$ itself.

A homomorphism $\Phi : \mathfrak{S} \to G$ is a *direct limit* iff (i) $\Phi$ is onto and (ii) for any group $H$ and homomorphism $\Psi : \mathfrak{S} \to H$, there exists a homomorphism $\lambda : G \to H$ such that $\Psi = \lambda \Phi$, that is, for every group $G_\alpha$ in $\mathfrak{S}$, $\psi_\alpha = \lambda \varphi_\alpha$.

(2.1) THEOREM *Any system* $\mathfrak{S}$ *has a direct limit unique to within isomorphism.*

The proof is straightforward and is given in [4]. As a result of (2.1), one may relax the above terminology and speak simply of the group $G$ as the direct limit of the system $\mathfrak{S}$.

A given system $\mathfrak{S}$ may always be enlarged to a system $\mathfrak{S}'$ by adjoining all, or any number of, identity homomorphisms and finite compositions of homomorphisms of $\mathfrak{S}$. It is obvious that any homomorphism of $\mathfrak{S}$ is also a homomorphism of $\mathfrak{S}'$, and conversely. Thus,

(2.2) *Any direct limit* $\Phi : \mathfrak{S} \to G$ *is a direct limit* $\Phi : \mathfrak{S}' \to G$, *and conversely.*

**3. The generalized van Kampen theorem.** Consider a collection of pathwise-connected, open subsets $X_\alpha$ of a topological space $X$ closed under finite intersections and such that

$$X = \bigcup X_\alpha$$
$$p \in \bigcap X_\alpha, \text{ for some point } p$$

The set $\mathfrak{S}$ of fundamental groups $G_\alpha = \pi(X_\alpha, p)$ and all homomorphisms $\theta : G_\alpha \to G_\beta$ induced by inclusion is a system, and the homomorphisms $\varphi_\alpha : G_\alpha \to G = \pi(X, p)$ induced by inclusion constitute a homomorphism $\Phi : \mathfrak{S} \to G$.

(3.1) VAN KAMPEN THEOREM. $\Phi : \mathfrak{S} \to G$ *is a direct limit.*

*Proof.* There are two propositions to verify :

I. $\Phi$ *is onto.* Consider an arbitrary non-trivial element $A \in G$ and a loop $a$ representing $A$. Since $A \neq 1$, we know that $\| a \| > 0$. We construct a subdivision.

$$0 = t_0 < t_1 < \cdots < t_n = \|a\|$$

such that each $t_i - t_{i-1}$ is less than the Lebesgue number of the open covering of $[0, \|a\|]$ consisting of all inverse images $a^{-1}X_\alpha$. We then choose $X_{\alpha_i}$, $i = 1, \cdots, n$, such that

$$a[t_{i-1}, t_i] \subset X_{\alpha_i} \qquad i = 1, \cdots, n.$$

For each point $t_i$, $i = 0, \cdots, n$, of the subdivision, we select a path $b_i$ in $X$ subject to the conditions:

(i) $b_i(0) = p$ and $b_i(\|b_i\|) = a(t_i)$

(ii) If $a(t_i) = p$, then $b_i \equiv p$

(iii) $b_i(t) \in X_{\alpha_i} \cap X_{\alpha_{i+1}}$, $\quad 0 \le t \le \|b_i\|$ and $i = 1, \cdots, n-1$.

Note that (iii) uses the fact that the collection of subsets $X_\alpha$ is closed under finite intersections. Next, consider paths $a_i : [0, t_i - t_{i-1}] \to X$, $i = 1, \cdots, n$, defined by $a_i(t) = a(t + t_{i-1})$.
Clearly,

$$a = \prod_{i=1}^{n} a_i$$

and

$$a \simeq \prod_{i=1}^{n} b_{i-1} \cdot a_i \cdot b_i^{-1} .$$

Each path $b_{i-1} \cdot a_i \cdot b_i^{-1}$ is a $p$-based loop whose image lies entirely in $X_{\alpha_i}$ and which, therefore, is a representative loop of $\varphi_{\alpha_i} A_i$ for some $A_i \in G_{\alpha_i}$. Thus,

$$A = \prod_{i=1}^{n} \varphi_{\alpha_i} A_i$$

and the proof of I is complete.

II. *For any group $H$ and homomorphism $\Psi : \mathfrak{S} \to H$, there exists a homomorphism $\lambda : G \to H$ such that $\Psi = \lambda\Phi$.*
Proving II obviously amounts to proving that, for any $A_i \in G_{\alpha_i}$, $i = 1, \cdots, r$,

$$\prod_{i=1}^{r} \varphi_{\alpha_i} A_i = 1 \text{ implies } \prod_{i=1}^{r} \psi_{\alpha_i} A_i = 1 .$$

We select representative loops $a_i \in A_i$, $i = 1, \cdots, r$. Then the product

$$a = \prod_{i=1}^{r} \varphi_{\alpha_i} a_i$$

is contractible (We denote an inclusion mapping and its induced homomorphism of the fundamental groups by the same symbol), and there exists a homotopy $h : R \to X$, where $R = [0, \|a\|] \times [0, 1]$, which satisfies

$$h(t, 0) = a(t)$$
$$h(0, s) = h(t, 1) = h(\|a\|, s) = p$$

The vertical lines $t = \sum_{k=1}^{i} \|a_k\|$, $i = 1, \cdots, r$, provide a decomposition of $R$, and we consider a refinement

$$0 = t_0 < t_1 < \cdots < t_n = \|a\|$$
$$0 = s_0 < s_1 < \cdots < s_m = 1$$

into rectangles

$$R_{ij} = \{(t, s) \mid t_{i-1} \leq t \leq t_i \text{ and } s_{j-1} \leq s \leq s_j\}$$

the maximum of whose diameters is less than the Lebesgue number of the open covering of $R$ consisting of all inverse images $h^{-1}X_\alpha$. Consequently, there exists a function $\alpha(i, j)$ such that

$$h(R_{ij}) \subset X_{\alpha(i,j)} \qquad i = 1, \cdots, n \text{ and } j = 1, \cdots, m.$$

For each lattice point $(t_i, s_j)$, we select a path $e_{ij}$ in $X$ subject to the following conditions.

(iv) The initial and terminal points of $e_{ij}$ are $p$ and $h(t_i, s_j)$, respectively.

( v ) If $h(t_i, s_j) = p$, then $e_{ij} \equiv p$.

(vi) The image of $e_{ij}$ is contained in $X_{\alpha(i,j)} \cap X_{\alpha(i+1,j)} \cap X_{\alpha(i,j+1)} \cap X_{\alpha(i+1,j+1)}$ . (Assume $X_{\alpha(i,j)} = X$ if $i = 0, n+1$ or if $j = 0, m+1$).

(vii) If $\sum_{k=1}^{j-1} \|a_k\| \leq t_{i-1} \leq t_i \leq \sum_{k=1}^{j} \|a_k\|$, then the image of $e_{i1}$ contained in $X_{\alpha_j}$.

Next, cf. Fig. 1, consider paths

$$c_{ij}(t) = h(t + t_{i-1}, s_j) \qquad 0 \leq t \leq t_i - t_{i-1}$$
$$d_{ij}(s) = h(t_i, s + s_{j-1}) \qquad 0 \leq s \leq s_j - s_{j-1}$$

and set

$$a_{ij} = e_{i-1,j} \cdot c_{ij} \cdot e_{ij}^{-1} \qquad i = 1, \cdots, n \text{ and } j = 0, \cdots, m$$
$$b_{ij} = e_{i,j-1} \cdot d_{ij} \cdot e_{ij}^{-1} \qquad i = 0, \cdots, n \text{ and } j = 1, \cdots, m$$

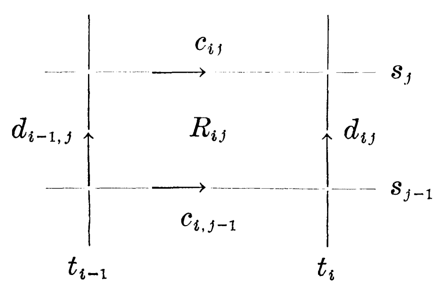



Fig. 1

The image points of the loops $a_{ij}, b_{ij}, a_{i,j-1}$, and $b_{i-1,j}$ all lie in $X_{\alpha(i,j)}$, Consequently, they define group elements $A_{ij}$, $B_{ij}$, $A'_{ij}$, and $\beta'_{ij}$ respectively. $G_{\alpha(i,j)}$. The product $a_{i,j-1} \cdot b_{ij} \cdot a_{ij}^{-1} \cdot b_{i-1,j}^{-1}$ is obviously contractible in $X$; moreover, since the image of $R_{ij}$ as well as the images of the four paths lies in $X_{\alpha(i,j)}$, the product is also contractible in $X_{\alpha(i,j)}$. We may conclude that

$$A'_{ij}B_{ij}A_{ij}^{-1}(B'_{ij})^{-1} = 1. \tag{1}$$

The central idea in the proof of II is the fact that *if group elements* $A \in G_\alpha$ *and* $B \in G_\beta$ *possess a common representative loop, then* $\psi_\alpha A = \psi_\beta B$.

The proof is easy: By assumption the system $\mathfrak{S}$ contains the fundamental group $G_\gamma$ of the intersection $X_\gamma = X_\alpha \cap X_\beta$ and the homomorphisms

$$G_\alpha \xleftarrow{\theta_1} G_\gamma \xrightarrow{\theta_2} G_\beta$$

induced by inclusion. The assertion that $A$ and $B$ possess a common representative loop states that there exists a $p$-based loop $c$ in $X_\gamma$ such that $\theta_1 c \in A$ and $\theta_2 c \in B$. Thus, if $c$ defines $C \in G_\gamma$, we have

$$\theta_1 C = A \text{ and } \theta_2 C = B$$

Since $\Psi$ is consistent with the mappings $\theta$,

$$\psi_\alpha A = \psi_\alpha \theta_1 C = \psi_\gamma C = \psi_\beta \theta_2 C = \psi_\beta B\ .$$

Applying the central assertion, we obtain

$$\begin{aligned} \psi_{\alpha(i,j)}A_{ij} &= \psi_{\alpha(i,j+1)}A'_{i,j+1} \\ \psi_{\alpha(i,j)}B_{ij} &= \psi_{\alpha(i+1,j)}B'_{i+1,j} \end{aligned} \tag{2}$$

Equation (1) says that the result of reading around each $R_{ij}$ under the homorphism $\psi_{\alpha(i,j)}$ is the identity. Equations (2) show that edges of adjacent rectangles will cancel. It follows (by induction) that the result of reading around the circumference of the large rectangle $R$ is the identity. Furthermore, only the elements along the bottom edge, $s = 0$, are non-trivial. We conclude, therefore, that

$$\prod_{i=1}^{n} \psi_{\alpha(i,0)}A_{ij} = 1\ .$$

Since each of the numbers $\sum_{k=1}^{j} \|a_k\|$, $j = 1, \cdots, r$, is a member of $\{t_1, \cdots, t_n\}$, there exists an index function $i(j)$ such that $i(0) = 0$, and

$$t_{i(j)} = \sum_{k=1}^{j} \|a_k\| \qquad j = 1, \cdots, r.$$

Then,

$$\prod_{i=i(j-1)+1}^{i(j)} a_{i0} \simeq \varphi_{\alpha_j} a_j \qquad j = 1, \cdots, r.$$

However, by virtue of (vii), we may assume that the equivalence is in $X_{\alpha_j}$. Thus, each loop $a_{i0}$, $i = i(j-1)+1, \cdots, i(j)$, determines a group element $A'_i \in G_{\alpha_j}$ and

$$\prod_{i=i(j-1)+1}^{i(j)} A'_i = A_j \ .$$

Since $A_{i0}$ and $A'_i$ possess a common representative loop $a_{i0}$, it follows from our central assertion that

$$\psi_{\alpha(i,0)} A_{ij} = \psi_{\alpha_j} A'_i \qquad i = i(j-1)+1, \cdots, i(j) \ ,$$

Finally, therefore,

$$1 = \prod_{j=1}^{r} \prod_{i=i(j-1)+1}^{i(j)} \psi_{\alpha(i,0)} A_{i0} = \prod_{j=1}^{r} \prod_{i=i(j-1)+1}^{i(j)} \psi_{\alpha_j} A'_i$$

$$= \prod_{j=1}^{r} \psi_{\alpha_j} A_j$$

and the proof of the generalized van Kampen theorem is complete.

**4. Generators and relations.** Since generators and relations describe a group only to within isomorphism, we shall speak of the image group of any direct limit of a system as the direct limit of the system. To obtain a presentation of the direct limit of a system of groups which are given by generators and relations is a simple matter of setting up the proper homomorphisms and chasing around a batch of consistent diagrams. Consider a system $\mathfrak{S}$, each group $G_\alpha$ of which has a presentation (cf [3])

$$G_\alpha = (x_\alpha^1, x_\alpha^2, \cdots : r_\alpha^1, r_\alpha^2, \cdots) \ .$$

Each mapping $\theta : G_\alpha \to G_\beta$ in $\mathfrak{S}$ is described by giving the assignment $\theta x_\alpha^i \in G_\beta$, $i = 1, 2, \cdots$. Then, the direct limit of $\mathfrak{S}$ has the presentation

$$\text{(1)} \qquad \mathrm{G} = (\{x_\alpha^i\} : \{r_\alpha^i\}, \{x_\alpha^i(\theta\, x_\alpha^i)^{-1}\})$$

i.e., all generators $x_\alpha^i$, all relators $r_\alpha^i$, and all elements $x_\alpha^i(\theta x_\alpha^i)^{-1}$ (a proof is given in [4]). The presentation (1) can be simplified in that, for each homomorphism $\theta : G_\alpha \to G_\beta$, the relators $r_\alpha^i$, $i = 1, 2, \cdots$, may be dropped. The reason is that, in the free group $F$ generated by all the generators in (1) and of which $G$ is the homomorphic image, the relators $r_\alpha^i$ are a consequence of the relators $r_\beta^i$ and the elements $x_\alpha^i(\theta x_\alpha^i)^{-1}$. To prove this assertion consider the diagram

$$\begin{array}{ccccc} & G_\alpha & \xrightarrow{\theta} & G_\beta & \\ & \zeta_\alpha \uparrow & & \uparrow \zeta_\beta & \\ & F_\alpha & \xrightarrow{\bar\theta} & F_\beta & \\ \text{inclusion} & & \searrow \ \swarrow & & \text{inclusion} \\ & & F & & \end{array}$$

$F_\alpha$ is the free group generated by $x_\alpha^i$, $i = 1, 2, \cdots$, and $\zeta_\alpha$ is the con-onical homomorphism whose kernel is the consequence of $r_\alpha^i$, $i = 1, 2, \cdots$. The mapping $\bar\theta$, which strictly speaking should be used in (1), is simply $\theta$ lifted to the free groups. Consider an arbitrary homomorphism $\eta$ of $F$ which maps $r_\beta^i$, $x_\alpha^i(\bar\theta x_\alpha^i)^{-1}$, $i = 1, 2, \cdots$, onto 1. Then, for any $u \in F_\alpha$,

$$\eta u = \eta \bar\theta u .$$

Since

$$\zeta_\beta \bar\theta r_\alpha^i = \theta \zeta_\alpha r_\alpha^i = 1 ,$$

each $\bar\theta r_\alpha^i$ is a consequence of the elements $r_\beta^i$. Hence,

$$\eta \bar\theta r_\alpha^i = \eta r_\alpha^i = 1$$

and the assertion is proved.

Consider a topogical space $X$ which is the union of two pathwise-connected open subsets $X_1$ and $X_2$ whose intersection $X_0 = X_1 \cap X_2$ is also pathwise-connected and contains a point $p$. Suppose we are given presentations of the fundamental groups $G_i = \pi(X_i, p)$, $i = 0, 1, 2$,

$$G_1 = (x_1, x_2, \cdots : r_1, r_2, \cdots)$$
$$G_2 = (y_1, y_2, \cdots : s_1, s_2, \cdots)$$
$$G_0 = (z_1, z_2, \cdots : t_1, t_2, \cdots)$$

and the mapping $\theta_i : G_0 \to G_i$, $i = 1, 2$, induced by inclusion are described by assignments $\theta_i z_j \in G_i$, $i = 1, 2$, $j = 1, 2, \cdots$. By our principal Theorem (3.1) and the results of the preceding paragraph, the fundamental group $G = \pi(X, p)$ has the presentation

$$G = (\{x_j\}, \{y_j\}, \{z_j\} : \{r_j\}, \{s_j\}, \{z_j(\theta_i z_j)^{-1}\})$$

This presentation is equivalent to (cf. [3])

$$G = (\{x_j\}, \{y_j\} : \{r_j\}, \{s_j\}, \{\theta_1 z_j (\theta_2 z_j)^{-1}\})$$

which is the assertion of the usnal formulation of the van Kampen theorem.

Consider a system $\mathfrak{S}$ of groups and mappings

$$G_1 \xrightarrow{\theta_1} G_2 \xrightarrow{\theta_2} G_3 \xrightarrow{\theta_3} \cdots \text{ with presentations}$$

$$(2) \qquad G_i = (x_i^1, x_i^2, \cdots : r_i^1, r_i^2, \cdots)$$

such that

$$(3) \qquad \theta_i x_i^j = x_{i+1}^j \text{ and } \theta_i r_i^j = r_{i+1}^j$$

($G_{i+1}$ may have more generators and relators than $G_i$). We may define a group

$$(4) \qquad G = (y_1, y_2, \cdots : s_1, s_2, \cdots)$$

and a homomorphism $\Phi : \mathfrak{S} \to G$ such that

$$\varphi_i x_i^j = y_j \text{ and } \varphi_i r_i^j = s_j .$$

It is easy to check that $G$ (more precisely, $\Phi : \mathfrak{S} \to G$) is the direct limit of $\mathfrak{S}$.

Finally, we consider an ascending chain of non-empty, open subsets $X_1 \subset X_2 \subset \cdots$ of some topological space. We have by (3.1) and (2.2) that the fundamental group $G$ of the union is the direct limit of the system

$$G_1 \xrightarrow{\theta_1} G_2 \xrightarrow{\theta_2} \cdots,$$

where $G_i = \pi(X_i, p)$ is the fundamental group and $\theta_1$ is induced by inclusion. Using the results of the preceding paragraph, we obtain a presentation (4) for $G$, if presentations (2) satisfying conditions (3), are given. This procedure is used in [1] to obtain (among other examples) a presentation of the group of the exterior of the Alexander Horned Sphere.

## REFERENCES

1. W. A. Blankinship and R. H. Fox, *Remarks on certain pathological open subsets of 3-space and their fundamental groups*, Proc. Amer. Math. Soc., **1**, (1950), 618–624.

2. R. H. Fox and E. Artin, *Some wild cells and spheres in three-dimensional space*, Ann. of Math., **49**, (1948), 979–990.

3. R. H. Fox, *Free differential calculus* II, Ann. of Math. **59**, (1954), 196–210.

4. R. H. Fox, *Discrete groups and their presentations*, Notes on lectures given spring term 1955 at Princeton University.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

# CONVOLUTION SEMIGROUPS OF MEASURES

IRVING GLICKSBERG

Let $S$ be a compact topological semigroup, $C(S)$ the Banach space of all continuous complex valued functions on $S$, and $\tilde{S}$ the normalized non-negative regular Borel measures on $S$. Under convolution and the $\omega^*$ topology of $C(S)^*$, $\tilde{S}$ and the unit ball $\tilde{\tilde{S}}$ of $C(S)^*$ each form a compact semigroup. The main purpose of this paper is the determination of all subgroups of $\tilde{S}$ and $\tilde{\tilde{S}}$ when $S$ is *abelian*.

In the case in which $S$ is a group, J. G. Wendel [10] has determined the idempotents in $\tilde{S}$: they are just the Haar measures of subgroups of $S$. This fails to hold for the general compact semigroup $S$, but does remain valid for compact *abelian* semigroups, due primarily to the fact that the least ideal in a compact *abelian* semigroup is a group. Indeed it is just this feature of the abelian case which allows one to complete the one point in Wendel's argument where essential use is made of a group structure, rather than a semigroup structure, for $S$, and further allows one to determine the subgroups of $\tilde{S}$.

The structure of the subgroups of $\tilde{S}$ (when $S$ is abelian or a group) is quite simple: each subgroup $\Gamma$ of $\tilde{S}$ consists of the $G-$ translates of Haar measure on $g$, where $G$ is a subgroup of $S$, and $g$ a normal subgroup of $G$. Thus $\Gamma$ is just the set of point masses on $G/g$ imbedded in $\tilde{S}$ in the natural fashion, and we arrive essentially at the fact that the only subgroups of $\tilde{S}$ are the obvious ones. But a consequence of this knowledge is an extension of the Weyl equidistribution theorem: for $\mu$ in $\tilde{S}$, $N^{-1}\sum_{n=1}^{N}\mu^n \xrightarrow[\omega^*]{}$ Haar measure of the least ideal of the sub-semigroup of $S$ generated by the carrier of $\mu$ (in the group situation this is convergence to Haar measure of the subgroup generated by the carrier).

Finally, in the abelian case, the determination of the subgroups of $\tilde{\tilde{S}}$ is obtained as a consequence; by virtue of a theorem of Eberlein [3] we can apply our results to obtain the subgroups of the convolution semigroup formed by the unit ball of $C_0(\mathscr{G})^*$ where $\mathscr{G}$ is a locally compact abelian group.

It is a pleasure to record the author's indebtedness to K. de Leeuw for his stimulating comments and suggestions, which were directly

Received July 14, 1958. Presented to the American Mathematical Society, June 21, 1958.

responsible for many of the results; our indebtedness to Wendel's paper will be self-evident.

**1. Preliminaries.** We begin with a résumé of the facts and notation we shall use in connection with semigroups, ideals, measures and convolution; for standard results on measure theory and topological groups the reader is referred to [5, 6, 9]. *Let $S$ be henceforth a compact semigroup,* i.e. a compact (Hausdorff) space with a jointly continuous (binary) operation (multiplication) under which it forms a semigroup. By a subsemigroup of $S$ we shall implicitly mean a *closed* subsemigroup; a not necessarily closed one will be called an *algebraic* subsemigroup. By a subgroup $G$ of $S$ we shall mean a (closed) subsemigroup which algebraically forms a group under our operation; since $G$ is compact, inversion (as is easily seen) is automatically continuous and *$G$ is a compact topological group.*

(1.1) Suppose now that $S$ is abelian. An ideal $I$ of $S$ is a nonvoid subset closed under multiplication from outside ($SI \subset I$), and a consequence of compactness is the fact that $S$ contains a least ideal $I = \bigcap_{x \in S} xS$; for $xyS \subset xS \cap yS$ implies $\{xS : x \in S\}$ has the finite intersection property while $xS$ is trivially closed so that $I \neq \phi$. And clearly $I$ is a (closed) ideal contained in any other ideal. Moreover

(1.11) *if $E$ is dense in $S$, then $I = \bigcap_{x \in E} xS$.*
For given an open set $V$ containing $I$ we have an $x$ in $S$ with $xS \subset V$ (otherwise the filter generated by $\{xS : x \text{ in } S\}$ has each of its elements meeting the compact complement $V'$ of $V$, whence $I \cap V' \neq \phi$). Thus by compactness and the continuity of multiplication we have a $y$ in $E$ near $x$ for which $yS \subset V$, $I \subset \bigcap_{y \in E} yS \subset V$, and (1.11) follows. Further *$I$ is a subgroup of $S$* as well [8]: for $x \in S \Rightarrow xI$ is an ideal contained in $I$, so $xI = I$. Thus if $x \in I$ we have an $e$ in $I$ for which $xe = x$, whence $yxe = yx$; since $Ix = xI = I$, $e$ is clearly an identity for $I$. On the other hand $yI = I$ implies that there is a $z$ in $I$ with $yz = e$, and $I$ is a group.

For a non-abelian $S$ we have the usual variety of ideals and the above facts are of course invalid; however it will be convenient to note that if $S$ is a group any sort of ideal must coincide with the full group $S$, and all of our remarks retain their full force.

(1.2) With $S$ abelian or not the fact that $S$ is compact allows us to identify $C(S)^*$ with the space of (integrals with respect to) complex regular Borel measures of $S$. We shall use the same letter to denote the functional and the measure, writing $\mu(f) = \int f(x)\mu(dx)$. The norm

$\|\mu\|$ of $\mu$ in $C(S)^*$ is of course its total variation, and the unit ball of $C(S)^*$, $\tilde{\tilde{S}} = \{\mu: \|\mu\| \leqq 1\}$ is compact in the $\omega^*$ topology, as is its subspace $\tilde{S} = \{\mu: \mu \geqq 0, \|\mu\| = 1\}$.

For $f \in C(S)$ let $f_x(y) = f(xy)$, $f^x(y) = f(yx)$, so that $f_x$ and $f^x$ are in $C(S)$. The compactness of $S$ and the continuity of multiplication combine to yield the maps $x \to f_x$, $x \to f^x$ of $S$ into $C(S)$ continuous, and thus for $\mu$ in $C(S)^*$, $\int f(xy)\mu(dx)$ is continuous in $y$. Consequently we can form the iterated integral $\iint f(xy)\mu(dx)\nu(dy)$ which, as a function of $f$, lies in $C(S)^*$. The corresponding measure $\mu\nu$, the convolution of $\mu$ and $\nu$, thus satisfies

$$\int f(x)\mu\nu(dx) = \iint f(xy)\mu(dx)\nu(dy) \tag{1.21}$$

for $f$ in $C(S)$. Moreover using the monotoneity arguments of [6][1] we have (1.21) *holding for bounded Baire functions* $f$. Since the associative law is easily verified, and $\mu, \nu \geqq 0$ implies $\mu\nu \geqq 0$ while $\|\mu\| = \mu(1) = \int \mu(dx)$ for $\mu \geqq 0$, $\tilde{S}$ clearly forms a semigroup under convolution, abelian if $S$ is (by Fubini's theorem); similarly $\tilde{\tilde{S}}$ forms a semigroup since clearly $\|\mu\nu\| \leqq \|\mu\| \cdot \|\nu\|$. If we now add the $\omega^*$ topology we obtain compact semigroups: for since $y \to f^y$ is continuous for an $f$ in $C(S)$, $F = \{f^y : y \in S\}$ is a compact subset of $C(S)$, and thus pointwise convergence of an equicontinuous bounded net of functions on $F$ implies uniform convergence by Ascoli's theorem. But $\tilde{S}$ and $\tilde{\tilde{S}}$ are equicontinuous sets of functions on $F$ and $\omega^*$ convergence amounts to pointwise convergence, so $\mu_\delta \to \mu$, $\nu_\delta \to \nu$ imply $\int f(xy)\mu_\delta(dx) \to \int f(xy)\mu(dx)$ uniformly in $y$ and therefore

$$\iint f(xy)\mu_\delta(dx)\nu_\delta(dy) \to \iint f(xy)\mu(dx)\nu(dy), \text{ or } \mu_\delta\nu_\delta \to \mu\nu .$$

Finally we note the existence, for each non-negative regular Borel measure $\mu$, of a unique closed set $A =$ carrier $\mu \subset S$ with the property that $\mu A = \|\mu\|$ and $\mu U > 0$ for each open $U$ with $A \cap U \neq \phi$ [10];

[1] For $\mu, \nu \geqq 0$ one argues as follows: the set of non-negative Baire $f$ for which

$$f'(y) = \int f(xy) \wedge 1 \, \mu(dx)$$

defines a Baire function $f'$ and for which

$$\iint f(xy) \wedge 1 \, \mu(dx)\nu(dy) = \int f(x) \wedge 1 \, \mu\nu(dx)$$

is clearly a monotone class containing the non-negative elements of $C^R(S)$, and thus includes all non-negative Baire $f$. For general $\mu, \nu$ the decomposition $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$, and the obvious distributivity of convolution suffice. Also monotoneity shows $f$ Baire on $S$ implies $\tilde{f}: (x, y) \to f(xy)$ is a Baire function on $S \times S$, and thus Fubini's theorem may be applied to $\tilde{f}$.

it is simply the complement of the union of all open sets of $\mu$ measure zero.

## 2. Idempotents and subgroups.

The fundamental tool in our analysis is the following extension of Wendel's Lemma 4.

LEMMA 2.1 *For $\mu$ and $\nu$ in $\tilde{S}$,*

$$\text{carrier } \mu\nu = (\text{carrier } \mu) \cdot (\text{carrier } \nu) .$$

*Proof.* Let $A$ and $B$ be the respective carriers of $\mu$ and $\nu$. Since each is compact so is $A \cdot B$, which in particular is then a Borel set. Thus by the regularity of $\mu\nu$, for $\varepsilon > 0$ we have an open $U$ containing $A \cdot B$ for which $\mu\nu(U) \leqq \mu\nu(A \cdot B) + \varepsilon$. Since $S$ is normal Urysohn's lemma applies to yield an $F$ in $C(S)$ with $\varphi_{A \cdot B} \leqq F \leqq \varphi_U$ (where $\mu_E$ is the characteristic function of $E$), i.e., $0 \leqq F \leqq 1$ and $F = 0$ on $U'$, $= 1$ on $A \cdot B$. But it is clear that $\varphi_A(x)\varphi_B(y) \leqq F(xy)$ for all $x$, $y$ in $S$, and thus

$$1 = \mu(A) \cdot \nu(B) = \iint \varphi_A(x)\varphi_B(y)\mu(dx)\nu(dy) \leqq \iint F(xy)\mu(dx)\nu(dy)$$
$$= \int F(x)\mu\nu(dx) \leqq \mu\nu(U) \leqq \mu\nu(A \cdot B) + \varepsilon \leqq 1 + \varepsilon .$$

Since $\varepsilon > 0$ is arbitrary, $\mu\nu(A \cdot B) = 1$. Moreover if $U$ is now anopen set with $(A \cdot B) \cap U \neq \phi$ then we can find open sets $V$ and $W$ for which $V \cap A \neq \phi$, $W \cap B \neq \phi$, and $V^- \cdot W^- \subset U$; choosing an $F$ in $C(S)$ with $\varphi_{V^- \cdot W^-} \leqq F \leqq \varphi_U$ again yields $\mu(V^-) \cdot \nu(W^-) \leqq \mu\nu(U)$, and this combines with $\mu(V^-) \geqq \mu(V) > 0$, $\nu(W^-) \geqq \nu(W) > 0$ to show $\mu\nu(U) > 0$. Hence $A \cdot B$ is indeed the carrier of $\mu\nu$.

If $\mu$ is now an idempotent in $\tilde{S}$,

$$(\text{carrier } \mu)^2 = \text{carrier } \mu^2 = \text{carrier } \mu .$$

In the group situation this guarantees the carrier is a group [4, 7], but in the case of an abelian semigroup $S$ we must go further.

THEOREM 2.2 *Let $S$ be abelian or a group, and $\mu^2 = \mu \in \tilde{S}$. Then* carrier $\mu$ *is a subgroup of $S$ and $\mu$ is its Haar measure*[2].

*Proof* (Following Wendel). For completeness we shall include both cases in our proof, although in the group situation we have Wendel's Theorem 1. Let $H =$ carrier $\mu$, so $H^2 = H$. For $f$ in $C^R(S)$ (the space

[2] For non-abelian $S$ this and our subsequent results fail in general. For take $S = [0, 1]$ under $\circ$, where $x \circ y = y$. Then $\mu\nu = \nu$, for $\mu$, $\nu \in \tilde{S}$ and each element of $\tilde{S}$ is an idempotent.

of real valued elements of $C(S)$) let $f'(x) = \int f(yx)\mu(dy)$, $x \in S$, so that $f' \in C^R(S)$. Since $H$ is compact $f'$ assumes its supremum over $H$ at some $x_0$ in $H$, and

$$f'(x_0) = \int f^{x_0}(y)\mu(dy) = \mu(f^{x_0}) = \mu^2(f^{x_0})$$
$$= \iint f(yzx_0)\mu(dy)\mu(dz) = \int f'(zx_0)\mu(dz) \leqq f'(x_0)$$

since $f' \leqq f'(x_0)$ on (carrier $\mu$) $\cdot\, x_0 = Hx_0 \subset H^2 = H$. Consequently $f'(x_0) = \int f'(zx_0)\mu(dz)$ and, since $f'$ is continuous and $H =$ carrier $\mu$, $f'$ assumes its supremum over $H$ on all of $Hx_0$; in particular then on the least ideal $I$ of the subsemigroup $H$ of $S$.

(In case $S$ is a group our proof is complete: for $H$ is a group, $I = H$, and $f'$ constant on $I = H \Rightarrow \mu$ is right invariant).

Now suppose $H \setminus I \neq \phi$. Then we can find an $x_1$ in $H$ and non-negative $f$ in $C(S)$ vanishing on $I$ for which $f^{x_1}$ does not vanish on all of $H$; otherwise for each $f \geqq 0$ in $C(S)$ with $f(I) = 0$ we have $f^x(H) = f(Hx) = 0$ for all $x$ in $H$, and thus $f(H) = f(H^2) = 0$. Hence for this $f$ and $x_1$ we have $\mu(f^{x_1}) > 0$ while, for $y$ in $I$, $f'(y) = \int f(xy)\mu(dx) = \int 0\mu(dx) = 0$ since $Hy \subset I$ and $f$ vanishes on $I$. But since $f'$ assumes its supremum over $H$ on $I$, $0 \geqq f'(x_1) = \mu(f^{x_1}) > 0$, the desired contradiction, and $H = I$, a subgroup (of $H$ and thus) of $S$. Moreover since $f'$ is constant on $I = H$, $\mu$ is invariant and our proof complete.

For a subset $E$ of $\tilde{S}$ we shall refer to $(\bigcup_{\mu \in E}\text{carrier}\,\mu)^-$ as the *carrier* of $E$, which is obviously consistent with our former use of the term. It should be noted that if $E$ is a subsemigroup of $\tilde{S}$ then carrier $E$ is a subsemigroup of $S$. For by Lemma 2.1 $\bigcup_{\mu \in E}\text{carrier}\,\mu$ is closed under multiplication, and therefore its closure is also. Moreover carrier $E^- =$ carrier $E$; for if carrier $\mu \not\subset$ carrier $E$ then there is an $f$ in $C(S)$ vanishing on carrier $E$ with $\mu(f) \neq 0$. But then $\nu(f) = 0$ for $\nu$ in $E$ and therefore for $\nu$ in $E^-$ as well, and $\mu \notin E^-$.

THEOREM 2.3 *Let $S$ be either abelian or a group, and let $\Gamma$ be a subgroup of $\tilde{S}$. Then the carrier $G$ of $\Gamma$ is a subgroup of $S$ while the carrier $g$ of the identity $\eta$ of $\Gamma$ is a normal subgroup of $G$. If $T_\eta$ denotes the map of $(G/g)^\sim \to \tilde{S}$ defined by*

$$T_\eta\nu(f) = \int_{G/g}\int_g f(xy)\eta(dx)\nu(dyg)\,, \qquad f \in C(S),$$

*then $T_\eta$ takes $(G/g)^\sim$ onto the $\omega^*$ closed convex hull $\mathscr{C}(\Gamma)$ of $\Gamma$, the point masses $(G/g)^p$ of $G/g$ onto $\Gamma$, and in each case is a (topological) isomorphism* *between these semigroups.*

COROLLARY 2.31. *$\Gamma$ is the set of $G$—translates of Haar measure on $g$.* For $\eta$ is Haar measure on $g$ by Theorem 2.2 and thus for $\nu =$ mass 1 at $gy \in G/g$ we have $T_\eta\nu(f) = \int_g f(xy)\eta(dx)$, which of course corresponds to the Haar measure of $g$ translated to the coset $gy$.

*Proof of Theorem* 2.3. Consider first the case in which $S$ is abelian. Let $S_0 = \bigcup_{\mu\in\Gamma}\text{carrier}\,\mu$, an algebraic subsemigroup of $S$ with $S_0^- =$ carrier $\Gamma = G$. Since $\mu = \eta\mu$ for $\mu$ in $\Gamma$, carrier $\mu =$ carrier $\eta$ · carrier $\mu$ $= g$ carrier $\mu$ by Lemma 2.1, and thus $gS_0 = S_0$ and therefore $gG = G$.

But for $x$ in $S_0$ we have $x \in$ carrier $\mu$, $\mu \in \Gamma$, so $x \cdot$ carrier $\mu^{-1} \subset$ carrier $\mu\mu^{-1} = g$ and $xG \cap g \neq \phi$. Further, since $xyG \subset xG \cap yG$ for $x, y \in S_0 \subset G$ we conclude from the compactness of $g$ that $g$ meets $\bigcap_{x\in S_0} xG$, the least ideal $I$ of the compact semigroup $G$ (cf. (1.11)). Consequently $g \subset I$; for $i \in g \cap I$ implies $g = ig \subset I$ since $g$ is a group and $I$ an ideal. Since $gG = G$ we obtain $G \subset I \subset G$, and since $I$ is algebraically a group, $G$ is a subgroup of $S$.

Now evidently $T_\eta$ maps $(G/g)^\sim$ into $\tilde{S}$. Let $f$ be in $C(S)$ and vanish on $G$. Then clearly $T_\eta\nu(f) = 0$, $\nu \in (G/g)^\sim$, so that $G$ contains the carrier of any measure in the range of $T_\eta$. The subset $M$ of $\tilde{S}$ of elements with carriers contained in $G$ may be considered as a subset of either $C(S)^*$ or $C(G)^*$; in each case we obtain the same $\omega^*$ topology since by Urysohn's lemma $C(G)$ is exactly the set of restrictions to $G$ of the elements of $C(S)$, and $\mu(f) = \mu(f \mid G)$ (where $\mu$ on the left is in $C(S)^*$, and on the right in $C(G)^*$). For the same reason we may evidently form the convolution of two elements of $M$ in either place, i.e. $M$ may be considered as a subsemigroup of either $\tilde{S}$ or $\tilde{G}$. Thus it will clearly suffice to consider $T_\eta$ as a map of $(G/g)^\sim$ into $\tilde{G}$.

But now we recognize $T_\eta$ as (a restriction of) the adjoint of the map $f \to f'$ of $C(G) \to C(G/g)$ defined by setting $f'(yg) = \int_g f(xy)\eta(dx)$. Thus $T_\eta$ is $(\omega^* \to \omega^*)$ continuous, and since $f \to f'$ is onto [6, 9], $T_\eta$ is one-to-one, hence a homeomorphism on $(G/g)^\sim$. Further $T_\eta$ is an isomorphism since for $f \in C(G)$

$$T_\eta\nu_1 \cdot T_\eta\nu_2(f) = \int_{G/g}\int_g\int_{G/g}\int_g f(xyzw)\eta(dx)\nu_1(dyg)\eta(dz)\nu_2(dwg)\ ,$$

$$T_\eta(\nu_1\nu_2)(f) = \int_{G/g}\int_g f(xy)\eta(dx)\nu_1\nu_2(dyg)$$

$$= \int_{G/g}\int_{G/g}\int_g f(xyw)\eta(dx)\nu_1(dyg)\nu_2(dwg)$$

$$= \int_{G/g}\int_{G/g}\int_g\int_g f(xzyw)\eta(dx)\eta(dz)\nu_1(dyg)\nu_2(dwg)$$

since $\eta^2 = \eta$, and thus multiplicativity follows from Fubini's theorem and commutativity.

Now let $\rho$ be the canonical homomorphism of $G \to G/g$ and, for $\mu$ in $\Gamma$ define $\bar{\mu} \in (G/g)^{\sim}$ by $\bar{\mu}(F) = \mu(F \circ \rho)$, $F \in C(G/g)$. Then for $f \in C(G)$,

$$\mu(f) = \eta\mu(f) = \int_G \int_g f(xy)\eta(dx)\mu(dy) = \int_{G/g} \int_g f(xy)\eta(dx)\bar{\mu}(dyg)$$
$$= T_\eta \bar{\mu}(f) \,, \quad \text{so} \quad \Gamma \subset T_\eta(G/g)^{\sim}.$$

Thus the (compact) preimage of $\Gamma$ is a subgroup of $(G/g)^{\sim}$ whose identity is the mass 1 at the identity $g$ of $G/g$ (for clearly this measure maps onto $\eta$ and $T_\eta$ is one-to-one). Since $G/g$ is a group, Lemma 2.1 implies each element of the preimage is a point mass; indeed the preimage consists of just those obtained from a closed subgroup of $G/g$ since as is well known the map from points to point masses (in the $\omega^*$ topology) is a homeomorphism [2] and trivially a group isomorphism. Hence we may identify the preimage as $(G_0/g)^p$, the point masses on $G_0/g$ where $G_0$ is a subgroup of $G$ containing $g$ ($G_0$ is closed since $G_0/g$ and $g$ compact imply $G_0$ is compact). But obviously the carrier of each element of $T_\eta(G_0/g)^p$ is contained in $G_0$ so that carrier $\Gamma = G \subset G_0$, and $G_0 = G$, $T_\eta(G/g)^p = \Gamma$.

To complete the proof in the abelian case we need only note the well known fact [2], that $(G/g)^{\sim}$ is the $\omega^*$ closed convex hull of $(G/g)^p$, so that $T_\eta[(G/g)^{\sim}] = \mathscr{C}(\Gamma)$ follows from linearity and continuity.

Now suppose $S$ is a (non-abelian) compact group with identity $e$. Since we clearly have $G =$ carrier $\Gamma = G^2$, $G$ is a subgroup of $S$ [4, 7]. Moreover $g$ is a normal subgroup of $G$. For $x \in$ carrier $\mu$, $\mu \in \Gamma$, implies $x$ carrier $\mu^{-1} \subset g$ by Lemma 2.1 so that if $y \in$ carrier $\mu^{-1}$,

$$xy = z \in g, \; x^{-1} = yz^{-1} \in (\text{carrier } \mu^{-1}) \cdot g = \text{carrier } \mu^{-1} \,.$$

Thus

$$x^{-1}gx \subset \text{carrier } \mu^{-1} \cdot g \cdot \text{carrier } \mu = g \,,$$

and $x^{-1}gx \subset g$ for a dense set of $x$ in $G$; if $y \in g$ then $x^{-1}yx \in g$ for all $x$ in $G$, by continuity, and $g$ is normal in $G$.

Now if we omit the first two paragraphs of the proof for the abelian case, each step will apply here with one exception: the proof that $T_\eta$ is multiplicative. But (applying Fubini's theorem) this follows from the fact that

$$\int_g f(xyzw)\eta(dz) = \int_g f(xzyw)\eta(dz)$$

or equivalently

$$\int_g f_x^w(yz)\eta(dz) = \int_g f_x^w(zy)\eta(dz)$$

and thus ultimately from $yg = gy$.

2.4 REMARK. If $\Gamma$ is an *algebraic* subgroup of $\tilde{S}$ then $\Gamma^-$ *is a subgroup of* $\tilde{S}$ so that $\Gamma$ consists of the $G$—translates of a Haar measure, where $G$ is an *algebraic* subgroup of $S$. For if the net $\{\mu_\delta\} \subset \Gamma$ converges to $\mu \in \Gamma^-$, then any cluster point $\nu$ of $\{\mu_\delta^{-1}\}$ must satisfy $\mu\nu = \eta$ as a cluster point of $\{\mu_\delta \mu_\delta^{-1}\}$ ; and clearly $\mu = \mu\eta$.

**3. Least ideals and carriers.** Our next result gives the relationship when $S$ is abelian, between the least ideal of a subsemigroup of $\tilde{S}$ and the least ideal of its carrier: *the carrier of the least ideal is the least ideal of the carrier.*

THEOREM 3.1. *Let $S$ be abelian and let $\Sigma$ be a subsemigroup of $\tilde{S}$ with least ideal $\mathscr{J}$ ; let $S_1 =$* carrier *$\Sigma$ with least ideal $I$. Then $I =$* carrier $\mathscr{J}$.

*Proof.* We know that $\mathscr{J}$ is a subgroup of $\Sigma$ and thus of $\tilde{S}$. Hence by Theorem 2.3 its carrier is a subgroup $G$ of $S$. Let $S_0 = \bigcup_{\mu \in \Sigma}$ carrier $\mu$, a dense algebraic subsemigroup of $S_1$. Let $x \in S_0$ so that $x \in$ carrier $\mu$ for some $\mu$ in $\Sigma$. For $\nu$ in $\mathscr{J}$, $\mu\nu \in \mathscr{J}$ so $x$ carrier $\nu \subset$ carrier $\mu\nu \subset G$ by Lemma 2.1, and thus $xS_0 \cap G \neq \phi$ and $xS_1 \cap G \neq \phi$. Since $I = \bigcap_{x \in S_0} xS_1$ by (1.11) we conclude as in the proof of 2.3 that $G \cap I \neq \phi$ and therefore $G \subset I$.

But the fact that $x$ carrier $\nu \subset G$ for $x \in S_0$, $\nu \in \mathscr{J}$ clearly implies $xG \subset G$ for $x \in S_0$. Consequently for $y$ in $G$, $xy \in G$ for all $x$ in $S_1$, by continuity, and thus $S_1 G \subset G$, or $G$ is an ideal in $S_1$. Hence $G$ contains the least ideal $I$ and $I = G =$ carrier $\mathscr{J}$.

THEOREM 3.2. *Let $\mu \in \tilde{S}$, with $S$ abelian. Then $N^{-1} \sum_{n=1}^{N} \mu^n \to$ Haar measure on the least ideal of the subsemigroup of $S$ generated by* carrier $\mu$. *If $S$ is a* (not necessarily abelian) *group, $N^{-1} \sum_{n=1}^{N} \mu^n \to$ Haar measure of the subgroup of $S$ generated by* carrier $\mu$.

*Proof.* Let $\Sigma_\mu$ be the subsemigroup of $\tilde{S}$ generated by $\mu$, $\nu_N = N^{-1} \sum_{n=1}^{N} \mu^n$, and let $\nu$ be any cluster point of $\{\nu_N\}$ which of course must lie in $\mathscr{C}(\Sigma_\mu)$. Since $\| \mu\nu_N - \nu_N \| \to 0$ we have $\mu\nu = \nu$ and thus $\lambda\nu = \nu$ for each $\lambda \in \mathscr{C}(\Sigma_\mu)$. Since $\mathscr{C}(\Sigma_\mu)$ is abelian this clearly implies $\nu$ is the unique cluster point of $\{\nu_N\}$ so that $\nu_N \to \nu$ by compactness. Moreover $\lambda\nu = \nu$, $\lambda \in \mathscr{C}(\Sigma_\mu)$, says $\{\nu\}$ is the least ideal of the subsemigroup $\mathscr{C}(\Sigma_\mu)$ of $\tilde{S}$, and an idempotent, so that $\nu$ is Haar measure of its carrier by 2.2.

Now if $S$ is abelian the carrier of $\nu$ is the least ideal $I$ of carrier $\mathscr{C}(\Sigma_\mu)$ by 3.1. Evidently the carrier of the algebraic convex hull of $\Sigma_\mu$ coincides with the carrier of $\Sigma_\mu$, and since carrier $E^-$ = carrier $E$, we have carrier $\mathscr{C}(\Sigma_\mu)$ = carrier $\Sigma_\mu$ and our proof is complete in this case.

If $S$ is a group with identity $e$, let $G$ be the subgroup of $S$ generated by carrier $\mu$. Since a subsemigroup of a compact group is a group, carrier $\{\mu^n : n \geqq 1\}$ is a subgroup of $S$ and clearly must coincide with $G$. Thus if $f(G) = 0$, $f \in C(S)$ we have $f$ vanishing on carrier $\mu^n$, hence $\mu^n(f) = 0$, all $n$, and $\nu(f) = 0$; consequently carrier $\nu \subset G$. On the other hand $\mu^n\nu = \nu$ so that by 2.1 carrier $\mu^n \cdot$ carrier $\nu =$ carrier $\nu$ and thus

$$\bigcup \text{carrier } \mu^n = \bigcup (\text{carrier } \mu^n e) \subset \bigcup (\text{carrier } \mu^n \cdot \text{carrier } \nu) = \text{carrier } \nu$$

so $G \subset$ carrier $\nu$ and our proof is complete.

3.3 REMARK. More generally we can follow Alaoglu and Birkhoff [1] to obtain a stronger assertion. Let $E$ be a commuting subset of $\tilde{S}$, and let $\Sigma$ be the abelian subsemigroup of $\tilde{S}$ generated by $E$. We can regard $\mathscr{C}(\Sigma)$ as partially ordered by $\mu \leqq \nu \Leftrightarrow \nu \in \mu\Sigma$, and then $\mathscr{C}(\Sigma)$ forms a directed set ($\mu\nu \geqq \mu, \nu$). If we regard $\mathscr{C}(\Sigma)$ as indexed by itself then *$\mathscr{C}(\Sigma)$ is a net and the net converges to Haar measure on the least ideal of* carrier $\Sigma$. For given $\mu \in \mathscr{C}(\Sigma)$ and $\varepsilon > 0$ there is a $\nu_0$ in $\mathscr{C}(\Sigma)$ for which $\|\mu\nu - \nu\| < \varepsilon$, $\nu \geqq \nu_0$: simply choose $\nu_0 = N^{-1}\sum_{n=1}^{N}\mu^n$ for $N$ large enough to yield $\|\mu\nu_0 - \nu_0\| < \varepsilon$; then $\nu \geqq \nu_0 \Rightarrow \nu = \nu_0\lambda$, $\|\mu\nu - \nu\| = \|\mu\nu_0\lambda - \nu_0\lambda\| \leqq \|\mu\nu_0 - \nu_0\| \cdot \|\lambda\| < \varepsilon$. Consequently we obtain a unique cluster point $\nu$ of our net to which the net must converge, with $\mu\nu = \nu$, $\mu \in \mathscr{C}(\Sigma)$ and the remainder of our proof applies.

3.4. **Our next result gives more explicit** information about the least ideal of a subsemigroup of $\tilde{S}$ when $S$ is abelian.

THEOREM 3.5. *Let $S$ be abelian, and $\Sigma$ be the subsemigroup of $\tilde{S}$ generated by a subset $E$ of $\tilde{S}$, with carrier $S_1$. Let $\mathscr{J}$ and $I$ be the respective least ideals of $\Sigma$ and $S_1$ with identities $\eta$ and $e$ respectively. Then $\mathscr{J}$ is the set of $I$-translates of Haar measure $\eta$ of the subgroup $h$ of $I$ generated by* $\{(e \text{ carrier } \mu)(e \text{ carrier } \mu)^{-1} : \mu \in E\}$.

*Proof.* We already know from Theorem 3.1 and Corollary 2.31 that $\mathscr{J}$ is the set of $I$—translates of Haar measure $\eta$ of some subgroup $g$ of $I$; we have only to show $g = h$. But each subgroup $g_0$ of $I$ is determined by its orthogonal subgroup $g_0^{\perp} = \{\alpha \in \hat{I} : \alpha(g_0) = 1\}$ in the character group $\hat{I}$ of $I$, so we need only show $g^{\perp} = h^{\perp}$. Moreover the elements of $g^{\perp}$ are just those $\alpha$ in $\hat{I}$ for which $\eta(\alpha) = \int \alpha(x)\eta(dx) = 1$ (for all others $\eta(\alpha) = 0$), hence $g^{\perp} = \{\alpha \in \hat{I} : |\mu(\alpha)| = 1, \mu \in \mathscr{J}\}$.

Now for $\mu$ in $\Sigma$, $e$ carrier $\mu \subset$ carrier $\eta$ carrier $\mu =$ carrier $\eta\mu \subset I$ and since $\eta\mu \in \mathscr{J}$, carrier $\eta\mu$ is a coset $yg \subset I$. Thus

$$(e \text{ carrier } \mu)(e \text{ carrier } \mu)^{-1} \subset yg(yg)^{-1} = g\ ,$$

and $\alpha \in g^{\perp}$ implies $\alpha \in h^{\perp}$. To see that $h^{\perp} \subset g^{\perp}$, note that each $\alpha$

in $\hat{I}$ has a continuous multiplicative extension $\alpha^*$ to $S_1$: simply set $\alpha^*(x) = \alpha(xe)$, $x \in S_1$. Further $\alpha^*$ has a continuous extension $\alpha'$ to all of $S$ by Urysohn's lemma, and, for $\mu$ in $\Sigma$, $\mu(\alpha') = \mu(\alpha^*)$; thus if $\mu_\delta \to \mu$ in $\Sigma$, $\mu_\delta(\alpha^*) \to \mu(\alpha^*)$. But since $\alpha^*$ is multiplicative if we define the (Fourier) transform $\hat{\mu} \in C(\hat{I})$ of $\mu$ in $\Sigma$ by setting $\hat{\mu}(\alpha) = \mu(\alpha^*) = \int \alpha^*(x)\mu(dx)$, we have $(\mu\nu)^\wedge = \hat{\mu} \cdot \hat{\nu}$ (ordinary product in $C(\hat{I})$), with $\hat{\mu}(\alpha) = \mu(\alpha)$ for $\mu$ in $\mathscr{F}$ (since carrier $\mu$ is then $\subset I$).

Let $\alpha \in h^\perp$. Then

$$\alpha((e \text{ carrier } \mu)(e \text{ carrier } \mu)^{-1}) = \alpha(e \text{ carrier } \mu) \cdot \overline{\alpha(e \text{ carrier } \mu)} = 1$$

for $\mu$ in $E$ which implies $\alpha$ is constant on the sets $e$ carrier $\mu$, $\mu$ in $E$. Thus

$$\hat{\mu}(\alpha) = \int \alpha^*(x)\mu(dx) = \int \alpha(xe)\mu(dx)$$

is a unimodular complex number. But then $(\mu\nu)^\wedge = \hat{\mu} \cdot \hat{\nu}$ implies that $|\hat{\mu}(\alpha)| = 1$ for $\mu$ in the algebraic subsemigroup of $S$ generated by $E$; since $\mu_\delta \to \mu$ in $\Sigma$ implies $\hat{\mu}_\delta(\alpha) = \mu_\delta(\alpha^*) \to \mu(\alpha^*) = \hat{\mu}(\alpha)$ the same must be true for all $\mu$ in $\Sigma$. In particular for $\mu$ in $\mathscr{F}$, $|\mu(\alpha)| = |\hat{\mu}(\alpha)| = 1$ whence $\alpha \in g^\perp$ and $g^\perp = h^\perp$.

**4. The semigroup $\tilde{\tilde{S}}$.** When $S$ is abelian the subgroups of $\tilde{\tilde{S}}$, the convolution semigroup formed by the unit ball of $C(S)^*$, can be determined from those of $\tilde{S}$.

Let $\Gamma$ be a non-trivial (i.e. $\neq \{0\}$) subgroup of $\tilde{\tilde{S}}$, with identity $\eta$. Clearly $0 \notin \Gamma$ and consequently $\|\mu\| = 1$ for $\mu$ in $\Gamma$; for $\|\mu\| < 1$ implies $\mu^n \to 0$ and thus $0 \in \Gamma$. Now by the Radon—Nikodym theorem we can associate with each complex measure $\mu$ a non-negative measure $|\mu|$ and a *unimodular* Baire function $\rho_\mu$ for which $\mu(dx) = \rho_\mu(x)|\mu|(dx)$ (we shall express this by writing $\mu = \rho_\mu \cdot |\mu|$) and $\||\mu|\| = \|\mu\|$. For write $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$ with $\mu_j \geqq 0$, and let $\nu = \mu_1 + \mu_2 + \mu_3 + \mu_4$. Each $\mu_j$ is absolutely continuous with respect to $\nu$ so there are functions $f_j$ in $L_1(\nu)$ (which we can take to be Baire functions since each $\nu$ integrable function is equivalent to a Baire function) for which $\mu_j = f_j \cdot \nu$. Set

$$f = f_1 - f_2 + i(f_3 - f_4)\,, \quad |\mu|(dx) = |f(x)|\nu(dx)$$

and $\rho_\mu(x) = f(x)/|f(x)|$ unless $f(x) = 0$ when we set $\rho_\mu(x) = 1$. Clearly $\rho_\mu$ and $|\mu|$ have the required properties.

Thus for $\mu$ and $\nu$ in $\Gamma$ we have

$$(4.1) \quad 1 = \||\mu\nu|\| = \int 1\,|\mu\nu|(dx) = \int \frac{1}{\rho_{\mu\nu}(x)}\mu\nu(dx) = \iint \frac{1}{\rho_{\mu\nu}(xy)}\mu(dx)\nu(dy)$$

$$= \iint \frac{\rho_\mu(x)\rho_\nu(y)}{\rho_{\mu\nu}(xy)}|\mu|(dx)|\nu|(dy) \leqq \iint 1\,|\mu|(dx)|\nu|(dy) = 1\,.$$

Consequently we have a $|\nu|$ null set $E$ for which $y \notin E$ implies there is a $|\mu|$ null set $E_y$ for which $x \notin E_y$ implies $\rho_\mu(x)\rho_\nu(y) = \rho_{\mu\nu}(xy)$. Hence for $f$ in $C(S)$

$$|\mu\nu|(f) = \int f(x)\frac{1}{\rho_{\mu\nu}(x)}\mu\nu(dx) = \iint f(xy)\frac{1}{\rho_{\mu\nu}(xy)}\mu(dx)\nu(dy)$$
$$= \iint f(xy)\frac{\rho_\mu(x)\rho_\nu(y)}{\rho_{\mu\nu}(xy)}|\mu|(dx)|\nu|(dy)$$
$$= \iint f(xy)|\mu|(dx)|\nu|(dy) = |\mu||\nu|(f),$$

so that $\mu \to |\mu|$ is an (algebraic) homomorphism of $\Gamma$ onto an algebraic subgroup $\Gamma_0$ of $\tilde{S}$, whose identity is obviously $|\eta|$. Let $G =$ carrier $\Gamma_0$ $=$ carrier $\Gamma_0^-$, so that $\Gamma_0^-$ consists of all $G$-translates of Haar measure $|\eta|$ of a subgroup $g$ of $G$. We shall see later that $\mu \to |\mu|$ is also continuous, so that $\Gamma_0$ is compact and coincides with $\Gamma_0^-$.

Now each Baire function $f$ on $S$ has its restriction to $g$ a Baire function of $g$ (for the set of real valued $f$'s for which this holds is a monotone class containing $C^R(S)$). Thus $\rho_\eta|g$ is a Baire function on $g$. Applying (4.1) to the special case $\mu = \nu = \eta$ we conclude that there is an $|\eta|$ null set $E$ of $g$ for which $y \notin E$ implies there is an $|\eta|$ null set $E_y$ of $g$ for which $x \notin E_y$ implies $\rho_\eta(x)\rho_\eta(y) = \rho_\eta(xy)$. For simplicity let us now write $\rho_\eta = \rho$, and, restricting our attention entirely to $g$, write $dx$ for $|\eta|(dx)$, the element of Haar measure on $g$, and speak of $|\eta|$ null sets as null.

For $f \in L_1(g)$ (which we take as a Baire function of $g$) let $M(f) = \int f(x)\rho(x)dx$. Since $y \cdot E_y$ is null by translation invariance, and $x \notin yE_y$ implies $y^{-1}x \notin E_y$ and thus $\rho(y^{-1}x)\rho(y) = \rho(y^{-1}xy) = \rho(x)$ for $y \notin E$, we can write (with $f * h$ the usual convolution in $L_1(g)$)

$$M(f*h) = \int f*h(x)\rho(x)dx = \iint f(y)h(y^{-1}x)\rho(x)dydx$$
$$= \iint f(y)h(y^{-1}x)\rho(x)dxdy = \iint f(y)h(y^{-1}x)\rho(y^{-1}x)\rho(y)dxdy$$
$$= \int f(y)M(h)\rho(y)dy = M(f)\cdot M(h)$$

so that $M$ is a multiplicative functional on $L_1(g)$. Thus we have a character $\beta$ of $g$ for which $\rho = \beta \bmod |\eta|$ on the carrier $g$ of $|\eta|$, and clearly then $\eta(dx) = \beta(x)|\eta|(dx)$. Moreover since $\hat{g} = \hat{G}/g^\perp$ we have an $\alpha$ in $\hat{G}$ for which $\alpha|g = \beta$, so that $\eta(dx) = \alpha(x)|\eta|(dx)$ as elements of $\tilde{S}$, or $\eta = \alpha\cdot|\eta|$. (Note that essential use is made here of the abelian nature of $S$).

Now $\alpha^{-1}$ can be extended continuously to all of $S$; since each element of $\Gamma$ vanishes on subsets of $G'$ the map $\mu \to \alpha^{-1} \cdot \mu$ is clearly a one-to-one continuous map on $\Gamma$, and therefore a homeomorphism. Further it is clearly multiplicative, so $\mu \to \alpha^{-1} \cdot \mu$ is an isomorphism of $\Gamma$ with a subgroup $\Gamma_1$ of $\tilde{\tilde{S}}$. The identity of $\Gamma_1$ is $|\eta|$, and thus for $\nu$ in $\Gamma_1$,

$$\iint \nu(dx)\nu^{-1}(dy) = \int |\eta|(dx) = 1$$

whence

$$1 \leqq \int \left| \int \nu(dx) \right| |\nu^{-1}|(dy) = |\nu(1)| \cdot ||\nu^{-1}|| \leqq 1 .$$

Consequently $\nu(1)$ is a unimodular complex number $\beta(\nu)$, and since $\nu_1\nu_2(1) = \nu_1(1)\nu_2(1)$, $\beta$ is multiplicative; evidently $\beta$ is continuous and thus a character of $\Gamma_1$. Moreover $\int (1/\beta(\nu))\nu(dx) = 1$ implies $(1/\beta(\nu))\nu \geqq 0$, so the map $\tau : \nu \to \beta(\nu)^{-1}\nu$ is a continuous homomorphism of $\Gamma_1$ into a subgroup of $\tilde{S}$. Further the composition of $\tau$ with $\mu \to \alpha^{-1} \cdot \mu$, taking

$$\mu \to \alpha^{-1} \cdot \mu \to \beta(\alpha^{-1} \cdot \mu)^{-1}\alpha^{-1} \cdot \mu = \lambda$$

clearly must map $\mu \to |\mu|$ since $\mu = \beta(\alpha^{-1}\mu)\alpha \cdot \lambda$ with $\lambda \geqq 0$ and $\beta(\alpha^{-1} \cdot \mu)\alpha$ unimodular on $G$. Thus our original map $\mu \to |\mu|$ was continuous and $\tau : \nu \to \beta(\nu)^{-1}\nu$ maps $\Gamma_1$ onto $\Gamma_0$, which now appears as the full set of $G$-translates of Haar measure $|\eta|$ on $g$.

For $\nu$ in $\Gamma_0$ let $\tau^{-1}\nu = \Theta_\nu \cdot \nu$ where $\Theta_\nu$ is a closed subset of the circle group $T$, and in particular $\Theta = \Theta_{|\eta|}$ is a subgroup of $T$. Since $\Gamma_1 = \bigcup_{\nu \in \Gamma_0} \Theta_\nu \cdot \nu$ it remains to find the $\Theta_\nu$. But $\nu \in \Gamma_0$, $t \in T$ and $t\nu \in \Gamma_1$, imply $\beta(t\nu) = t$ since $\nu \geqq 0$ and $\beta(t\nu)^{-1}t\nu \geqq 0$; thus $\beta(\Theta_\nu \cdot \nu) = \Theta_\nu$. Moreover since $\nu \to \Theta_\nu \cdot \nu$ maps $\Gamma_0$ (topologically and) isomorphically onto $\Gamma_1/\Theta|\eta|$ and $\beta$ (taking $\Theta|\eta|$ onto $\Theta$) maps the quotient group $\Gamma_1/\Theta|\eta|$ of cosets into $T/\Theta$ in a homomorphic fashion, and continuously (as is easily seen), the composition $\nu \to \beta(\Theta_\nu \cdot \nu) = \Theta_\nu \in T/\Theta$ is continuous. We now distinguish two cases: $\Theta = T$ so that, as we could have seen earlier, $\Gamma_1 = T\Gamma_0$ (clearly this occurs iff $\Gamma$ is circular in the sense that $T\Gamma \subset \Gamma$), or $\Theta$ is the group of $n$th roots of unity, $n \geqq 1$. In the latter case we may apply the natural isomorphism $\sigma(: \xi \to \xi^n)$ of $T/\Theta$ onto $T$ to map $\Theta_\nu$ into $T$. Writing $|\eta|_x(\varepsilon\Gamma_0)$ for the translate to $xg$ of $|\eta|$ we thus have $\varphi(x) = \sigma(\Theta_{|\eta|_x})$ defining a character of $G$ lying in $g^\perp$: for the map $x \to |\eta|_x$ of $G$ into $\Gamma_0$ is a continuous homomorphism as the composition of $G \to G/g$, the map from $G/g$ into $(G/g)^p$ $(\subset \tilde{G})$ (continuous by [2]) followed by $T_{|\eta|}$ (cf. 2.3). Consequently $\Theta_{|\eta|_x}$ consists of just the $n$th roots of $\varphi(x)$ and we may express our general element of $\Gamma_1$ as $\varphi(x)^{1/n}|\eta|_x$, where $x \in G$ and $\varphi(x)^{1/n}$ denotes any root. In summary then, we have

THEOREM 4.2. *Let $S$ be abelian and $\Gamma$ any non-trivial subgroup of*

$\tilde{\tilde{S}}$. *If $\Gamma$ is circular in the sense that $T\Gamma \subset \Gamma$ then there are subgroups $g$ and $G$ of $S$, with $g \subset G$, and a fixed $\alpha$ in $\hat{G}$ for which $\Gamma = \{t\alpha \cdot \mu_x : t \in T, x \in G\}$ where $\mu$ is Haar measure on $g$, $\mu_x$ its translate to $xg$. If $\Gamma$ is not circular then in addition to $g$, $G$ and $\alpha$ we have an integer $n \geqq 1$ and a $\varphi$ in $g^{\perp} \subset \hat{G}$ for which $\Gamma = \{\varphi(x)^{1/n}\alpha \cdot \mu_x : x \in G\}$ where $\varphi(x)^{1/n}$ runs over all* $n$th *roots of $\varphi(x)$. Conversely any such set of measures forms a subgroup of $\tilde{\tilde{S}}$.*

There remains only the last point which is fairly obvious in the circular case. In the non-circular case any subset $\Gamma$ of the type described is algebraically a group, and one need only verify its closure. But if $\varphi(x_\delta)^{1/n}\alpha \cdot \mu_{x_\delta} \to \nu$ then by virtue of the compactness of $G$ we can find a confinal subnet for which $x_{\delta'} \to x$, an element of $G$, whence $\alpha \cdot \mu_{x_{\delta'}} \to \alpha \cdot \mu_x$; since $\varphi(x_{\delta'}) \to \varphi(x)$ some $n$th root $\varphi(x)^{1/n}$ of $\varphi(x)$ is a cluster point of $\varphi(x_{\delta'})^{1/n}$ and $\varphi(x)^{1/n}\alpha \cdot \mu_x$ is thus a cluster point of our convergent net, hence $= \nu$ and $\nu \in \Gamma$.

REMARK 4.3. The first portion of our proof identifies the idempotents in $\tilde{\tilde{S}}$ when $S$ is a compact non-abelian group. For the argument shows $\mu^2 = \mu \neq 0 \Rightarrow |\mu|^2 = |\mu|$, so that $|\mu|$ is Haar measure of a subgroup $g$, while $\rho_\mu$ again appears as a multiplicative character of $g$.

**4.4. It may be worthwhile to note** the analogue for $\tilde{\tilde{S}}$ of Theorem 3.2: *for $S$ abelian and $\mu_0 \in \tilde{S}$, $N^{-1}\sum_{n=1}^N \mu_0^n \to 0$ unless there is an $\alpha$ in $C(S)$ which is unimodular and multiplicative on $S_{|\mu_0|}$* (the subsemigroup of $S$ generated by carrier $|\mu_0|$) *satisfying $\mu_0 = \alpha \cdot |\mu_0|$, in which case* $N^{-1}\sum_{n=1}^N \mu_0^n \to \alpha \cdot$ (Haar measure on the least ideal of $S_{|\mu_0|}$). For if $\Sigma$ is the subsemigroup of $\tilde{\tilde{S}}$ generated by $\mu_0$ then as before $N^{-1}\sum_{n=1}^N \mu_0^n \to \nu$, the unique element of the least ideal of the semigroup $\mathscr{C}(\Sigma)$. Clearly $\nu = 0$ if $0 \in \mathscr{C}(\Sigma)$; if $0 \notin \mathscr{C}(\Sigma)$ then

$$\mu \in \mathscr{C}(\Sigma) \Rightarrow ||\mu|| = 1 , \tag{4.41}$$

and

(4.42) the least ideal $\mathscr{I}$ of $\Sigma$ is a non-circular subgroup of $\tilde{\tilde{S}}$ as in 4.2 with $n = 1$.

For otherwise, in each case, we may conclude that $0 \in \mathscr{C}(\Sigma)$. Consequently $\mathscr{I} = \{\beta \cdot \eta_x : x \in G\}$ where $\beta \in \hat{G}$, $G$ is a subgroup of $S$ and $\eta$ is Haar measure of a further subgroup.

But just as in the proof of 4.2, $\mu \to |\mu|$ is an algebraic homomorphism of $\mathscr{C}(\Sigma)$ into $\tilde{S}$. Moreover the image of $\mathscr{I} = \{\eta_x : x \in G\}$ is closed and is easily seen to be the least ideal of the closure $\Sigma_1$ of the image of $\Sigma$. Thus by Theorem 3.1, $G$ is the least ideal of $S_1 = $ carrier $\Sigma_1$, and as in theorem 3.5 we can extend $\beta$ to a continuous unimodular multiplicative function $\beta_1$ on $S_1$ by setting $\beta_1(x) = \beta(xe)$, $x \in S_1$, where $e$ is the identity of the least ideal $G$ of $S_1$. Since $\beta_1^{-1}$ is multiplicative on $S_1$ and all $\mu$ in $\Sigma$ vanish on all Borel subsets of $S_1'$, $\mu \to \beta_1^{-1} \cdot \mu$ is a homomorphism on $\Sigma$ which in particular maps $\mathscr{I}$ into $\tilde{S}$. As a consequence it must map all of $\Sigma$ into $\tilde{S}$: for if $\nu = \beta_1^{-1} \cdot \mu$ is the image of $\mu \in \Sigma$ then, since $(\beta \cdot \eta)\mu \in \mathscr{I}$ we have $\eta\nu \in \tilde{S}$ whence $1 = \eta\nu(1) = \eta(1)\nu(1) = \nu(1)$, so $\|\nu\| = 1$, $\nu \geqq 0$. Evidently then $\beta_1^{-1} \cdot \mu = |\mu|$. In particular $\mu_0 = \beta_1 |\mu_0|$ and we may take $\alpha$ as any continuous extension to all of $S$ of $\beta_1$. Finally if such an $\alpha$ is available then

$$\alpha \cdot |\mu_0|^n = (\alpha \cdot |\mu_0|)^n \text{ so } N^{-1} \textstyle\sum_{n=1}^N \mu_0^n = \alpha \cdot (N^{-1} \sum_{n=1}^N |\mu_0|^n)$$

and the final assertion follows from theorem 3.2.

**5. Application to $C_0(\mathscr{G})^*$.** Let $\mathscr{G}$ be a locally compact abelian group and $C_0(\mathscr{G})$ the Banach space of continuous functions vanishing at $\infty$, so that $C_0(\mathscr{G})^*$ consists of the finite regular Borel measures on $\mathscr{G}$. Uniform continuity of each element of $C_0(\mathscr{G})$ allows one to define convolution just as in § 1, and $C_0(\mathscr{G})^*$ is easily seen to form an abelian semigroup. However, the natural choice of the $\omega^*$ topology of $C_0(\mathscr{G})^*$ will not yield the unit ball a topological semigroup[3]; rather it is the topology of pointwise convergence of Fourier—Stieltjes transforms (in which $\mu_\delta \to \mu \Longleftrightarrow \hat{\mu}_\delta(\alpha) \to \hat{\mu}(\alpha)$ for each $\alpha \in \hat{\mathscr{G}}$) which does, and it is this topology we shall adopt.

The possibility of applying our previous results to the (topological) semigroup we thus obtain from the unit ball of $C_0(\mathscr{G})^*$ arises from two facts, both due to Eberlein [3]. Let $\mathscr{G}^*$ be the almost periodic compactification[4] of $\mathscr{G}$. Then as Eberlein has noted there is an isometric imbedding of $C_0(\mathscr{G})^*$ into $C(\mathscr{G}^*)^*$: for $\mu \in C_0(\mathscr{G})^*$ let $\mu'(f) = \int f(x)\mu(dx)$ for $f$ almost periodic on $\mathscr{G}$. Since the almost periodic functions on $\mathscr{G}$ are isometrically isomorphic to $C(\mathscr{G}^*)$ we obtain $\mu' \in C(\mathscr{G}^*)^*$. The clearly linear map $\mu \to \mu'$ then preserves norms by the following argument: select a compact $K \subset \mathscr{G}$ for which $|\mu|(K') < \varepsilon$ and an element $f$ of the unit ball of $C_0(\mathscr{G})$ for which $\left|\int f(x)\mu(dx)\right| \geqq \|\mu\| - \varepsilon$, so that $\left|\int_K f(x)\mu(dx)\right| \geqq \|\mu\| - 2\varepsilon$. Since $\mathscr{G}$ has sufficiently many characters the map of $\mathscr{G}$ into $\mathscr{G}^*$ is one-to-one and thus a homeomorphism on $K$. Consequently[5] we can find an $F$ in the unit ball of $C(\mathscr{G}^*)$ which extends

---

[3] For example take $\mu_n$ = mass 1 at the integer $n \in R$; then $\mu_n \to 0$, and $\mu_{-n} \to 0$ in the $\omega^*$ topology of $C_0(R)^*$ as $n \to +\infty$ while $\mu_n\mu_{-n} = \mu_0$.

[4] It will be convenient to view $\mathscr{G}$ as a dense algebraic subgroup of $\mathscr{G}^*$, and the almost periodic functions on $\mathscr{G}$ as the restrictions, to $\mathscr{G}$, of elements of $C(\mathscr{G}^*)$, cf. **[6, 9]**.

[5] We can simply extend the real and imaginary parts of $f \mid K$ separately by Urysohn's lemma to obtain an extension $F'$, and set $F(x) = F'(x) \cdot (1 \wedge |F'(x)|^{-1})$ (= 0 of course if $F'(x) = 0$).

$f\,|\,K$, a continuous function on the compact subset $K$ of $\mathscr{G}^*$. Then we have

$$|\mu'(F)| = \left|\int_{\mathscr{G}} F(x)\mu(dx)\right| \geqq \left|\int_K f(x)\mu(dx)\right| - \varepsilon \geqq ||\mu|| - 3\varepsilon ,$$

and $||\mu'|| \geqq ||\mu|| - 3\varepsilon$. Evidently $||\mu'|| \leqq ||\mu||$, so $\mu \to \mu'$ is an isometry. Moreover it is clear that since both $\mathscr{G}$ and $\mathscr{G}^*$ have the same algebraic group of characters, the underlying group of $\hat{\mathscr{G}}$, we may write $\hat{\mu} = \hat{\mu}'$ since both of these Fourier—Stieltjes transforms coincide as functions on the set $\hat{\mathscr{G}}$, and thus (since for measures $\mu, \nu$ on either group $(\mu\nu)^\wedge = \hat{\mu} \cdot \hat{\nu}$) $(\mu\nu)' = \mu'\nu'$. Consequently the map $\mu \to \mu'$ is an algebraic isomorphism of the semigroup formed by the ball of $C_0(\mathscr{G})^*$ into that formed by the ball of $C(\mathscr{G}^*)^*$, $\tilde{\tilde{\mathscr{G}}}^*$. Further, our choice of topology is just that which makes the map topological as well.

The second and crucial fact for our application which we obtain from Eberlein is the following corollary of the main result[6] of [3]: *Consider $C_0(\mathscr{G})^*$ as imbeded in $C(\mathscr{G}^*)^*$. Then its elements are just those measures $\mu$ on $\mathscr{G}^*$ with $\hat{\mu}$ continuous on $\hat{\mathscr{G}}$.* Thus we can easily identify the range of $\mu \to \mu'$.

Suppose then we are given a non-trivial closed subgroup $\Gamma$ of the unit ball of $C_0(\mathscr{G})^*$, and let $\Gamma_0$ be its isomorphic image in $\tilde{\tilde{\mathscr{G}}}^*$. Then $\Gamma_0^-$ is a subgroup of $\tilde{\tilde{\mathscr{G}}}^*$ and thus by Theorem 4.2 each of its elements is of the form $t\alpha \cdot \eta_x$, where $t \in T$, $\alpha$ is a character of a subgroup $G_0$ of $\mathscr{G}^*$, $\eta$ is Haar measure of a further subgroup $g_0$ of $G_0$ and $\eta_x$ the translate of $\eta$ to the coset $xg_0 \subset G_0$; indeed since each character of $G_0$ extends to one of $\mathscr{G}^*$, we shall take $\alpha \in \mathscr{G}^{*\wedge}$, i.e. as a character of $\mathscr{G}$. But the identity $\alpha \cdot \eta$ of $\Gamma_0^-$ was already present in $\Gamma_0$ and thus has a continuous Fourier—Stieltjes transform on $\hat{\mathscr{G}}$, whence $\hat{\eta}$ is continuous on $\hat{\mathscr{G}}$. Since $\hat{\eta} = \varphi_{g_0^\perp}$, the characteristic function of the subgroup $g_0^\perp$ of $\mathscr{G}^{*\wedge}(= \hat{\mathscr{G}}$ in the discrete topology) orthogonal to $g_0$, we obtain the fact that $g_0^\perp$ is an open and closed subgroup of $\hat{\mathscr{G}}$, and $\hat{\mathscr{G}}/g_0^\perp$ is discrete. But $\hat{\mathscr{G}}/g_0^\perp$ is the character group of that subgroup $g$ of $\mathscr{G}$ orthogonal to $g_0^\perp$; consequently $g$ is a compact subgroup of $\mathscr{G}$. If $\mu$ denotes its Haar measure then $\hat{\mu} = \varphi_{g^\perp} = \varphi_{g_0^\perp} = \hat{\eta}$ so $\eta = \mu'$ by the one-to-one-ness of the Fourier-Stieltjes transformation.

Now consider a general element $t\alpha \cdot \eta_x$ of $\Gamma_0$. The fact that its

---

[6] Specifically Eberlein's result may be stated as follows: for $\mu \in C(\mathscr{G}^*)$, $\hat{\mu} \in L_\infty(\hat{\mathscr{G}})$ (in the usual sense) implies there is a $\nu$ in $C_0(\mathscr{G})^*$ for which $\hat{\nu}$ coincides with $\hat{\mu}$ in $L_\infty(\mathscr{G})$. Since here our $\hat{\mu}$ is continuous, as $\hat{\nu}$ must be, we obtain $\hat{\mu} = \hat{\nu}$ as functions and thus $\mu = \nu'$ by the (one-to-one)-ness of the Fourier-Stieltjes transformation.

transform is continuous implies $(\eta_x)^\wedge$ is continuous while

$$(\eta_x)^\wedge(\beta) = \beta(x)\hat{\eta}(\beta) = \beta(x)\varphi_{g_0^+}(\beta) \ ;$$

thus as a function of $\beta$, $\beta(x)$ is continuous on the open subset $g_0^+$ of $\hat{\mathscr{G}}$, hence on all of $\hat{\mathscr{G}}$. By duality we then have a $y$ in $\mathscr{G}$ for which $\beta(x) = \beta(y)$, all $\beta$, and we may identify $x$ as an element of $\mathscr{G} \cap G_0$. Conversely each $x$ in $\mathscr{G} \cap G_0$ gives rise to elements of $\Gamma_0^-$ which already lie in $\Gamma_0$ (for such measures lie in the image of $C_0(\mathscr{G})^*$ in which $\Gamma_0$ is relatively closed by hypothesis); thus $\Gamma_0$ consists of just those elements $t\alpha \cdot \eta_x$ of $\Gamma_0^-$ arising from $x$'s in $\mathscr{G} \cap G_0 = G$, algebraically a subgroup of $\mathscr{G}$. But clearly $G$ is closed in $\mathscr{G}$, and is thus a subgroup of $\mathscr{G}$, since the map $\mathscr{G} \to \mathscr{G}^*$ is continuous.

Finally it is clear that if $\Gamma$ (and thus $\Gamma_0$) is circular so is $\Gamma_0^-$; conversely if $\Gamma_0^-$ is circular then $T\alpha \cdot \eta \subset \Gamma_0^-$ and thus $T\alpha \cdot \eta \subset \Gamma_0$, whence $\Gamma_0$ and $\Gamma$ are circular. We have proved

THEOREM 5.1 *Let $\mathscr{G}$ be a locally compact abelian group and let $C_0(\mathscr{G})^*$ be topologized by pointwise convergence of Fourier—Stieltjes transforms. Then any closed convolution subgroup $\Gamma$ of the unit ball of $C_0(\mathscr{G})^*$ is determined as in Theorem 4.2 where $g$ is a compact subgroup of $\mathscr{G}$, $G$ is a closed subgroup, and $\alpha$ and $\varphi$ may be taken as elements of $\hat{\mathscr{G}}$.*

**5.2. It should be noted that** the convolution semigroup formed by the ball of $C_0(\mathscr{G})^*$, although not compact, shares some properties of compact semigroups: the closure of an algebraic subgroup is again a group, indeed a topological group in the relative topology (thus the last applies to an algebraic subgroup). For if $\Gamma$ is an algebraic subgroup its image $\Gamma_0$ in $C(\mathscr{G}^*)^*$ is an algebraic group, so that $\Gamma_0^-$ is a compact topological group. But of course $\Gamma^-$ is just the preimage of the intersection of $\Gamma_0^-$ with the image of $C_0(\mathscr{G})^*$.

Finally suppose $\Gamma$ is a non-trivial algebraic subgroup of the ball of $C_0(\mathscr{G})^*$ which in addition is $\omega^*$ closed (compact). Then $\Gamma$ is a closed subgroup as described in Theorem 5.1 with $G$ a compact subgroup of $\mathscr{G}$ (and conversely). For, changing our notation, let $G$ denote the subgroup of $\mathscr{G}$ produced via Theorem 5.1 for $\Gamma^-$. Then the set $H$ of $x$ in $G$ corresponding to elements $t\alpha \cdot \mu_x$ in $\Gamma$ forms a dense algebraic subgroup of $G$, as is easily seen. If $G$ is not compact then we have a net $\{x_\delta\} \subset H$ which tends to $\infty$, so that the corresponding net of measures $\{t_\delta\alpha \cdot \mu_{x_\delta}\}$ tends to 0 in the $\omega^*$ topology ($g$ being compact). But this implies $0 \in \Gamma$, which is clearly nonsense.

Consequently $G$ is compact and, since the elements of $\Gamma$ all vanish off $G$, the $\omega^*$ topology on $\Gamma$ reduces to the topology of pointwise convergence of Fourier—Stieltjes transforms (by virtue of the Stone—Weierstrass theorem and the existence of sufficiently many characters of $\mathscr{G}$). Therefore the image of $\Gamma$ in $C(\mathscr{C}^*)^*$ is compact and closed, whence $\Gamma = \Gamma^-$.

## REFERENCES


1. L. Alaoglu and G. Birkhoff, *General ergodic theorems*, Ann. of Math. (2) **41** (1940), 293-309.
2. R. F. Arens and J. L. Kelley, *Characterizations of the space of continuous functions over a compact Hausdorff space*, Trans. Amer. Math. Soc., **62** (1947), 499-508.
3. W. F. Eberlein, *Characterizations of Fourier—Stieltjes transforms*, Duke Math. J., **22** (1955), 465-468.
4. B. Gelbaum, G. K. Kalisch, and J. M. H. Olmsted, *On the imbedding of topological semigroups and integral domains*, Proc. Amer. Math. Soc., **2** (1951), 807-821.
5. P. R. Halmos, *Measure theory*, New York, 1950.
6. L. H. Loomis, *An introduction to abstract harmonic analysis*, New York, 1953.
7. K. Numakura, *On bicompact semigroups*, Math. J., Okayama University, **1** (1952), 99-108.
8. A. D. Wallace, *The structure of topological semigroups*, Bull. Amer. Math. Soc., **61** (1955), 95-112.
9. A. Weil, *L'Integration dans les groupes topologiques et ses applications*, Actualites Seientifiques et Industrielles, no. 869, Paris, 1940.
10. J. G. Wendel, *Haar measure and the semigroup of measures on a compact group*, Proc. Amer. Math. Soc., **5** (1954), 923-929.


UNIVERSITY OF NOTRE DAME

# LINEAR OPERATORS AND THEIR CONJUGATES

SEYMOUR GOLDBERG

**Introduction.** In a paper of Taylor and Halberg ([3]), a complete systematic account of the theorems about the range and inverse of a bounded linear operator $T$ and its conjugate $T'$ was presented. For example, questions concerning $T$ and the corresponding questions concerning $T'$ such as the following were answered:

Does $Tx = y$ have a solution $x$ for each given $y$? If not, for which $y$'s does a solution exist?

Does the operator $T$ have an inverse $T^{-1}$, and if so, is $T^{-1}$ bounded?

These matters were considered for a bounded linear operator $T$ defined on all of a normed linear space $X$ with values in a second normed linear space $Y$.

The purpose of this paper is to investigate the same questions for $T$ and $T'$, where now $T$ is defined on a linear manifold $\mathscr{D}$ dense in $X$, and moreover, $T$ need not be bounded. It is shown that most of the theorems are still valid under these weakened hypotheses. Examples are constructed to show which theorems no longer hold.

Next, by imposing the condition that $T$ be a closed linear operator on $\mathscr{D}$, we show that we obtain the same results as for the case that $T$ be bounded on all of $X$.

**1. The conjugate transformation.** Throughout this paper we shall use $X$ and $Y$ to denote normed linear spaces over the real or complex scalar field. The space of all continuous linear functionals on $X$ will be written as $X'$.

The following theorem is well known.

THEOREM 1.1. *Let $Y$ be complete. If $T$ is a bounded linear transformation on $\mathscr{D} \subset X$ to $Y$ with norm $\|T\|$, then $T$ has a unique extention* $\hat{T}$ on $\overline{\mathscr{D}}$ and $\|\hat{T}\| = \|T\|$.

DEFINITION 1. Let $T$ be a linear operator (not necessarily bounded) with domain $\mathscr{D}$ dense in $X$ and range $\mathscr{R} \subset Y$. The *conjugate transformation* $T'$ is defined as follows: Its domain $\mathscr{D}(T')$ consists of the sets of all $y \varepsilon Y'$ for which $y'T$ is continuous on $\mathscr{D}$; for such a $y'$ we define $T'y' = x'$ where $x'$ is the bounded linear extension of $y'T$ to $X$.

THEOREM 1.1 assures the existense of such an $x'$ which is unique. Thus $T'$ is well defined. It is easy to see that $\mathscr{D}(T')$ is a linear manifold and that $T'$ is a closed linear operator. We refer to $T'$ as the

Received June 17, 1958. The author gratefully acknowledges the advice and counsel of Professor Angus E. Taylor in the preparation of this paper.

conjugate of $T$.

Unless otherwise indicated. $T$ and $T'$ will be as in Definition 1.

DEFINITION 2. A set $F$ contained in the space of all linear functionals on $X$ is called *total* if $x'x = 0$ for all $x' \in F$ implies $x = 0$.

The following theorem is due to Phillips [2, Theorem 2.11.9, p. 43].

THEOREM 1.2. *If $T$ is closed, then $\mathscr{D}(T')$ is total.*

REMARKS. The converse of this theorem need not hold. For let $\mathscr{D}$ be such that $\overline{\mathscr{D}} = X$ but $\mathscr{D} \neq X = Y$, and let $T$ be tne identity operator on $\mathscr{D}$. However, we easily prove the following.

THEOREM 1.3. *If $\mathscr{D}(T) = X$ and $\mathscr{D}(T')$ is total, then $T$ is closed.*

*Proof*. Let $\lim x_n = x$ and $\lim Tx_n = y$. All we need show is that $y = Tx$. If this were not the case, there would exist a $y' \in \mathscr{D}(T')$ such that $y'(y - Tx) \neq 0$. Since $y'T$ is continuous on $X$, we have that

$$y'y = \lim y'Tx_n = y'Tx$$

which is a contradiction.

**2. The state of a linear operator and its conjugate.** To discuss the range of linear operator $T$, we consider the following three possibilities, where $\mathscr{R}(T)$ will denote the range of $T$.

I. $\mathscr{R}(T) = Y$,

II. $\mathscr{R}(T) \neq Y$, but $\mathscr{R}(T)$ is dense in $Y$,

III. $\mathscr{R}(T)$ is not dense in $Y$, that is $\overline{\mathscr{R}(T)} \neq Y$.

If $T$ has an inverse, then the inverse mapping $T^{-1}$ is a linear operator from the normed linear space $\mathscr{R}(T)$ into the normed linear space $X$. As regards the inverse of $T$, we consider the following three possibilities:

1. $T$ has a bounded inverse,
2. $T$ has an unbounded inverse,
3. $T$ has no inverse.

By the various pairings of I, II, or III with 1, 2, or 3, nine conditions can thus be described relating to $\mathscr{R}(T)$ and $T^{-1}$. For instance, it may be that $\mathscr{R}(T) = Y$, and that $T$ has a bounded inverse. This we will describe by saying that $T$ is in state $\mathrm{I}_1$, (written $T \in \mathrm{I}_1$).

Since $T'$ is a linear operator from $\mathscr{D}(T')$ into $X'$, we can use the above classifications for $\mathscr{R}(T')$ and the inverse of $T'$. To the ordered pair of operators $(T, T')$ we now make correspond an ordered pair of conditions which we call the "state" of $(T, T')$. Thus if $T \in \mathrm{I}_3$ and $T' \in \mathrm{III}_1$, we say that $(T, T')$ is in state $(\mathrm{I}_3, \mathrm{III}_1)$ (written $(T, T') \in (\mathrm{I}_3, \mathrm{III}_1)$).

At times we shall use a notation such as $(T, T') \in (\mathrm{I}_2, 3)$ to mean that $T \in \mathrm{I}_2$ and $T'$ has no inverse.

The question arises as to whether $(T, T')$ can be in each of the 81 states. It will be shown that only 16 states can occur if no additional assumptions are made about $X$, $Y$ or $T$. However, if we require that $X$ be reflexive, $Y$ complete and $T$ closed, the number of actually possible cases drops to 7.

We shall now exhibit several theorems which will enable us to determine which states can or cannot occur for the pair $(T, T')$.

THEOREM 2.1. *If $T'$ has a continuous inverse, then $\mathscr{R}(T')$ is closed.* ($T'$ cannot be in $\mathrm{II}_1$).

*Proof.* Suppose there exists a sequence $\{y'_n\}$ from $\mathscr{D}(T')$ with $T'y'_n \to x'$. The sequence $\{y'_n\}$ is a Cauchy sequence since $\|y'_n - y'_m\| \leqq M\|T'y'_n - T'y'_m\|$ where $M$ is the norm of $(T')^{-1}$ as an operator on $\mathscr{R}(T')$. But $Y'$ is complete, therefore there exists a $y' \in Y'$ such that $\lim y'_n = y'$. Hence $y' \in \mathscr{D}(T')$ and $T'y' = x'$ since $T'$ is closed.

Theorems 2.2 through 2.5 are due to Phillips [2 pp. 44–45].

THEOREM 2.2. *A necessary and sufficient condition that $\overline{\mathscr{R}(T)} = Y$ is that $T'$ have an inverse.*

THEOREM 2.3. *If $\mathscr{R}(T')$ is $w^*$ dense in $X'$, then $T$ has an inverse.*

THEOREM 2.4. *If $\overline{\mathscr{R}(T)} = Y$ and $T^{-1}$ exists, then $(T^{-1})' = (T')^{-1}$; furthermore, $T$ has a bounded inverse if and only if $T'$ has a bounded inverse defined on all of $X'$.*

THEOREM 2.5. *$\mathscr{R}(T') = X'$ if and only if $T$ has a bounded inverse.*

The following theorem will show that three more states for $(T, T')$ cannot exist if we require that $Y$ be complete.

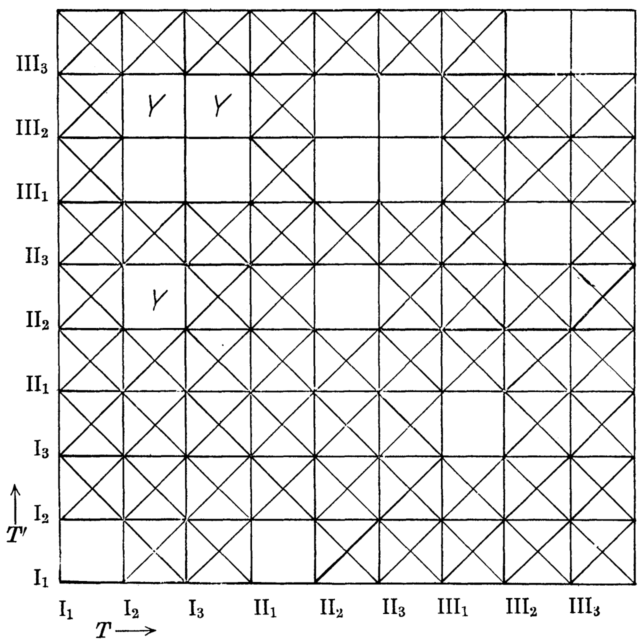
THEOREM 2.6. *If $Y$ is complete and $\mathscr{R}(T) = Y$, then $T'$ has a continuous inverse.* (States (I, 2) and (I, 3) cannot exist if $Y$ is complete).

*Proof.* If $T'$ did not have a continuous inverse, there would exist a sequence $y'_n$ in $Y'$ such that $\|y'_n\| \to \infty$ and $\|T'y'_n\| \to 0$. Since $\mathscr{R}(T) = Y$, it follows that $\|y'_n y\| \to 0$ for each $y \in Y$. Hence we can conclude that the sequence $\{\|y'_n\|\}$ is bounded, by a theorem due to Banach [1 p. 80, Theorem 5]. We have thus reached a contradiction.

3. **The state diagram of pairs $(T, T')$.** In order to present systematically which states can or cannot occur for pairs $(T, T')$, it will be

convenient to construct a "state diagram" conceived by Taylor [3 p. 100]. This diagram is a large square divided into 81 congruent smaller squares arranged in rows and columns. We label each column at the bottom denoting a given state for $T$, and each row by a "state" symbol placed at the left, denoting a certain state for $T'$. The square which is the intersection of a certain column and row will thus denote the state of the pair $(T, T')$. A square is crossed out by its diagonals if the corresponding state is impossible without requiring $X$ or $Y$ to be complete. As regards the remaining squares, we place the letter $Y$ in a square to indicate that the state cannot occur if $Y$ is complete.

First State Diagram




$Y$: Cannot occur if $Y$ is complete

**4. Example of states which can occur.** Excluding $(I_2, III_1)$, all of the examples in this section will be taken in the space $\ell^2$ with $X = Y = \ell^2$. The sequence space $\ell$ is defined to consist of all sequences $\{\xi_n\} = x$ such that $\sum_1^\infty |\xi_n|^2 < \infty$. The norm in $\ell^2$ is defined by

$$||x|| = (\sum_1^\infty |\xi_n|^2)^{1/2}.$$

It is well known that the conjugate space $(\ell^2)'$ of $\ell^2$ is congruent to $\ell^2$, whence $\ell^2$ is reflexive. In fact, every element in $(\ell^2)'$ is representable in one and only one way in the form

(1) $$x'x = \sum_1^\infty \alpha_n \xi_n$$

where the sequence $a = \{\alpha_n\}$ is an element of $\ell^2$. The correspondence between $x'$ and $a$ is a congruence between $(\ell^2)'$ and $\ell^2$. We shall write $x' = a$.

The set of vectors $u_k$, where $u_1 = (1, 0, \cdots)$, $u_2 = (0, 1, \cdots)$, etc. will frequently be used.

As the domain $\mathscr{D}$ of each linear operator $T$ in the examples to follow, excluding $(\mathrm{I}_2, \mathrm{III}_1)$, we take the linear combinations of the $u_k$. Clearly $\mathscr{D}$ is a subspace dense in $X = \ell^2$.

Taylor and Halberg [3 pp. 102–104] have shown that the seven states $(\mathrm{I}_1, \mathrm{I}_1)$, $(\mathrm{I}_3, \mathrm{III}_1)$, $(\mathrm{II}_2, \mathrm{II}_2)$, $(\mathrm{II}_3, \mathrm{III}_2)$, $(\mathrm{III}_1, \mathrm{I}_3)$, $(\mathrm{III}_2, \mathrm{II}_3)$, $(\mathrm{III}_3, \mathrm{III}_3)$ are all possible even when $Y = X = \mathscr{D} = \ell^2$ and $T$ is continuous.

We shall now demonstrate that the conditions corresponding to the 6 blank squares still unaccounted for in the state diagram can also occur.

$(\mathrm{II}_1, \mathrm{I}_1)$: It is clear that if we let $Y = X$ and $T$ be the identity operator on $\mathscr{D}$, then $(T, T')$ has the state $(\mathrm{II}_1, \mathrm{I}_1)$.

$(\mathrm{II}_2, \mathrm{III}_1)$: Let $Y = X$. If $x = (\xi_1, \xi_2, \cdots, \xi_n, 0, \cdots)$,

define

$$Tx = \left(\sum_1^n j\xi_j, \xi_2, \cdots, \xi_n, 0, \cdots\right).$$

Suppose

$$y' = (\alpha_1, \alpha_2, \cdots) \in \mathscr{D}(T') \subset \ell^2 .$$

From formula (1),

$$|y'Tu_k| = |\alpha_1 k + \alpha_k| \geqq |\alpha_1| k - |\alpha_k| \geqq |\alpha_1| k - \|y'\|$$

for $k > 1$. But $\|u_k\| = 1$ and $y'T$ is continuous on $\mathscr{D}$, therefore $\alpha_1$ must be zero. We now wish to determine the operator $T'$. If $T'y' = (\beta_1, \beta_2, \cdots) \in \ell^2$, then from formula (1),

$$\beta_k = T'y'u_k = y'Tu_k = \alpha_k$$

whence we see that $T'y' = y'$. Since $\alpha_1 = 0$, it is clear that $\overline{\mathscr{D}(T')} \neq X' = \ell^2$. Thus $T' \in \mathrm{III}_1$. Now $Tx = 0$ implies that $0 = \xi_2 = \xi_3 = \cdots = \xi_n = \xi_1 + 2\xi_2 + \cdots + n\xi_n$ or that $x = 0$, that is $T^{-1}$ exists; furthermore $\mathscr{R}(T) \neq Y$. An inspection of the state diagram shows that $T$ must be in $\mathrm{II}_2$. For the state $(\mathrm{I}_2, \mathrm{III}_1)$, we present two examples for the cases where $X$ is reflexive and $Y$ is not complete or where $X$ is complete and

$Y$ is reflexive. We do not have an example for $(I_2, III_1)$ where $X$ is reflexive and $Y$ is complete.

$(I_2, III_1)$: Take $Y = \mathscr{D}$ and let $T$ be the operator in the above example. If $y = (\eta_1, \eta_2, \cdots, \eta_p, 0, \cdots)$, then $Tx = y$ where $x = (\eta_1 - \sum_2^p k\eta_k, \eta_2, \cdots, \eta_p, 0, \cdots)$. Thus $\mathscr{R}(T) = \mathscr{D} = Y$. The above discussion now shows the existence of state $(I_2, III_1)$.

In [3 p. 108] it is shown that $(T, T')$ is in state $(I_2, III_1)$ where $T$ is a bounded linear operator from a normed linear space $Z$, which is not complete, into a reflexive normed linear sapce $Y$; for example, $Y = \ell^2$. Let $X$ be the completion of $Z$. Thus $\overline{Z} = X$ and $(T, T')$ is in state $(I_2, III_1)$ with respect to $X$ and $Y$.

$(II_3, III_1)$: Let $Y = X$. if $x = (\xi_1, \xi_2, \cdots, \xi_n, 0, \cdots)$, let

$$Tx = (2\xi_2, \cdots, n\xi_n, 0, \cdots) .$$

$T$ is clearly in $II_3$. Suppose $y' = (\alpha_1, \alpha_2, \cdots) \in \mathscr{D}(T')$ and that $T'y' = (\beta_1, \beta_2, \cdots) \in \ell^2$. Now

$$\beta_k = T'y'u_k = y'Tu_k = y'(ku_{k-1}) = k\alpha_{k-1} \qquad \text{if } k > 1 ,$$
$$\beta_1 = 0 .$$

Hence $\overline{\mathscr{R}(T')} \neq X' = \ell^2$. Moreover

$$\|T'y'\|^2 = \sum_2^\infty |k\alpha_{k-1}|^2 \geqq \sum_1^\infty |\alpha_j|^2 = \|y'\|^2 .$$

Thus $T' \in III_1$.

$(II_2, III_2)$: Let $Y = X$. If $x = (\xi_1, \cdots, \xi_n, 0, \cdots)$, let

$$Tx = \left(\xi_1 + \frac{1}{2}\xi_2 + 3\xi_3 + \cdots + b_n\xi_n, \xi_2/2, \cdots, \xi_n/n, 0, \cdots\right)$$

where

$$b_n = \begin{cases} 1/n \text{ if } n \text{ is even} \\ n \text{ if } n \text{ in odd.} \end{cases}$$

Clearly $T^{-1}$ exists; however $T^{-1}$ is not bounded since

$$\|Tu_{2k}\| = \|(u_1 + u_{2k})/2k\| = 1/k$$

and $\|u_{2k}\| = 1$. Furthermore, if $y = (\gamma_1, \gamma_2, \cdots, \gamma_N, 0, \cdots)$, then $Tx = y$ where

$$x = (\gamma_1 - \sum_2^N \gamma_j, 2\gamma_2 \cdots, N\gamma_N, 0, \cdots)$$

which shows that

$\mathscr{R}(T) = \mathscr{D}$, hence $T \in \text{II}_2$. Let $x_k = (1, 1/2, \cdots, 1/k, 0, \cdots)$. Then

$$Tx_k = (1 + 1/2^2 + 1 + \cdots + b_k/k, 1/2^2, \cdots, 1/k^2, 0, \cdots) .$$

Set $B_k = 1 + 1/2^2 + 1 + \cdots + b_k/k$. Obviously $B_k \to \infty$. If $y' = (\alpha_1, \alpha_2, \cdots) \in \mathscr{D}(T')$, then

$$|y'Tx_k| = |\alpha_1 B_k + \sum_2^k \alpha_j/j^2| \geqq |\alpha_1|B_k - \sum_2^k |\alpha_j/j^2| .$$

But

$$||x_k|| < \sum_1^\infty \frac{1}{n^2}, \sum_2^\infty | \frac{\alpha_j}{j^2} | \leqq ||y'|| \sum_1^\infty \frac{1}{n^2}$$

and $y'T$ is continuous on $\mathscr{D}$. Since $B_k \to \infty$, it follows that $\alpha_1 = 0$. If $T'y' = (\beta_1, \beta_2, \cdots) \in l^2$, then $\beta_1 = T'y'u_1 = y'Tu_1 = 0$ whence we see that $T' \in \text{III}$.

We shall now show that $T'$ does not have a bounded inverse.

Let $y'_k = \mu_k$ for $k > 1$. If $x = (\xi_1, \xi_2, \cdots, \xi_n, 0, \cdots)$ and $||x|| \leqq 1$, then $|y'_k Tx| = |\xi_k/k| \leqq 1$. Hence $y'_k \in \mathscr{D}(T')$. Now

$$T'y'_k u_j = y'_k Tu_j = \begin{cases} 0 & \text{if } k \neq j \\ 1/k & \text{if } k = j \end{cases}$$

or $||T'y'_k|| = ||u_k/k|| = 1/k$, which shows that $T'$ is not in state 1. This together with the fact that $T' \in \text{III}$ and $T \in \text{II}_2$ enable us to infer from the state diagram that $(T, T')$ is in state $(\text{II}_2, \text{III}_2)$.

$(\text{III}_2, \text{III}_3)$: Let $Y = l^2$. Similar to the preceding example, we define

$$Tx = (0, \xi_1 + 1/2\xi_2 + \cdots + b_n/n, \xi_2/2, \cdots, \xi_n/n, 0, \cdots) .$$

By the same procedure as above, we see that $T \in \text{III}_2$; also if $y' = (\alpha_1, \alpha_2, \cdots,) \in \mathscr{D}(T')$, it follows that $\alpha_2, = 0$ and therefore $\beta_1 = T'y'u_1 = y'u_2 = 0$ Hence $T' \in \text{III}$. From an inspection of the state diagram, it is clear that $T' \in \text{III}_3$.

We have now shown that twelve of the thirteen states are possible with $X$ and $Y$ reflexive. State $(\text{I}_2, \text{III}_1)$ is also possible with $X$ complete and $Y$ reflexive or with $X$ reflexive and $Y$ not complete. The state diagram assures us that no other states are possible as long as $Y$ is complete. If $X$ is complete and $Y$ is not required to be complete, then it is shown in [3] p. 106 that states $(\text{I}_2, \text{II}_2)$, $(\text{I}_2, \text{III}_2)$ and $(\text{I}_3, \text{III}_2)$ can occur; i.e. the squares which have the letter $Y$ become blank. Thus we have the justification of the entries is state diagram.

The question now arises as to whether in considering the same type of hypotheses on $X$ and $Y$, that is reflexivity and completeness, we can show that certain additional states are impossible if we put futher "reasonable" hypotheses on $T$, for example $T$ closed. The answer to this query is in the affirmative as we show in the next section. An assumption that $X$ be reflexive played no part in Theorems 2.1 through 2.6.

## 5. The State of a closed operator and its conjugate.

LEMMA. *If $T'$ has a continuous inverse, then for each $\alpha > 0$, 0 is an interior point of $\overline{TS_\alpha}$ where $S_\alpha = \{x \mid x \in \mathscr{D}, \|x\| \leqq \alpha\}$.*

*Proof.* An inspection of the first part of the proof of Theorem 6 [3 p. 97] will exhibit the proof of the lemma. It is to be noted that the argument does not depend on the hypothesis that $T$ be bounded.

THEOREM 5.1. *Suppose that $X$ is complete. If $T$ is closed and $T'$ has a continuous inverse, then $\mathscr{R}(T) = Y$. Moreover, if $T^{-1}$ exists, it is continuous.*

*Proof.* Define $S_n = \{x \mid x \in \mathscr{D}, \|x\| \leqq 1/2^n,\ n = 1, 2, \cdots\}$. By the lemma, we can choose a sequence of positive numbers $\{\varepsilon_n\}$ such that $\sum_1^\infty \varepsilon_n < \infty$, and $V_n = \{y \mid y \in Y, \|y\| < \varepsilon_n\} \subset \overline{TS_n}$. The existence of these $V_n$ and the arguments used in proving Theorem 2.12.1, p. 46.2 [2] will also prove this theorem. If, in the above theorem, $T$ were continuous on $X$, that is $\mathscr{D} = X$, one could conclude that $Y$ is complete. (cf. [3 Theorem 6 p. 97]) However, we cannot conclude that $Y$ is complete in Theorem 5.1 even if $\mathscr{D} = X$. The following example illustrates this assertion.

EXAMPLE. Let $X$ be any complete normed linear space of infinite dimension and let $H$ be a Hamel basis of $X$ with all elements $h \in H$ such that $\|h\| \leqq 1$. To each $x \in X$ there corresponds a unique finite set $h_1, h_2, \cdots, h_n$ in $H$ and unique scalars $\alpha_1, \alpha_2, \cdots, h_n$ such that $x = \sum_1^n \alpha_i h_i$. We now define another norm $\|x\|_1$ on $X$ by letting $\|x\|_1 = \sum_1^n |\alpha_i|$. Taylor and Halberg [3, p. 109] show that $X$ with this norm, which we designate by $X_1$, is not complete. Define $T$ as the identity mapping from $X$ onto $X_1$. $T$ has a bounded inverse, since

$$\|Tx\|_1 = \|x\|_1 = \sum_1^n |\alpha_i| \geqq \sum_1^n \|\alpha_i h_i\| \geqq \|x\|.$$

In addition, $T$ is also closed, for suppose $x_n \to x$ and $Tx_n \to y$. Since $\|x_n - y\| \leqq \|x_n - y\|_1 = \|Tx_n - y\|_1$ and $\|Tx_n - y\|_1 \to 0$, it follows that $Tx = x = y$. An inspection of the state diagram shows that $T'$ has a bounded inverse. Thus the hypotheses of Theorem 5.1 are satisfied, but $\mathscr{R}(T) = X_1$ is not complete.

This example also serves to illustrate that in the hypotheses of the "closed graph theorem" it is essential that the closed operator map into a complete normed linear space.

DEFINITION 3. If $E$ is a subset of $X$, define

$$E^\circ = \{x' \mid x' \in X' ;\ x'x = 0 \text{ for all } x \in E\} .$$

DEFINITION 4. If $S$ is a subset of $X'$, define

$$^\circ S = \{x \mid x'x = 0 \text{ for all } x' \in S\} .$$

The following known lemma is easy to prove.

LEMMA. *Let $X$ be reflexive. If $M$ is a closed linear subspace in $X'$, then $M = (^\circ M)^\circ$.*

THEOREM 5.2. *Let $X$ be reflexive. If $T^{-1}$ exists and $\mathscr{D}(T')$ is total, then $\overline{\mathscr{R}(T')} = X'$.*

*Proof.* We first show that $^\circ\overline{\mathscr{R}(T')} = (0)$. *If* $x \in {}^\circ\overline{\mathscr{R}(T')}$, then $y'Tx = T'y'x = 0$ for all $y' \in \mathscr{D}(T')$; but then $Tx = 0$ since $\mathscr{D}(T')$ is total. The fact that $T^{-1}$ exists implies that $x = 0$. Clearly $0 \varepsilon\ {}^\circ\overline{\mathscr{R}(T')}$, hence $^\circ\overline{\mathscr{R}(T')} = (0)$. Applying the preceding lemma, we see that

$$\overline{\mathscr{R}(T')} = (^\circ\overline{\mathscr{R}(T')})^\circ = (0)^\circ = X' .$$

COROLLARY. *Let $X$ be reflexive. If $T$ is closed and $T^{-1}$ exists, then $\overline{\mathscr{R}(T')} = X'$.*

*Proof.* Theorems 1.2 and 5.2.

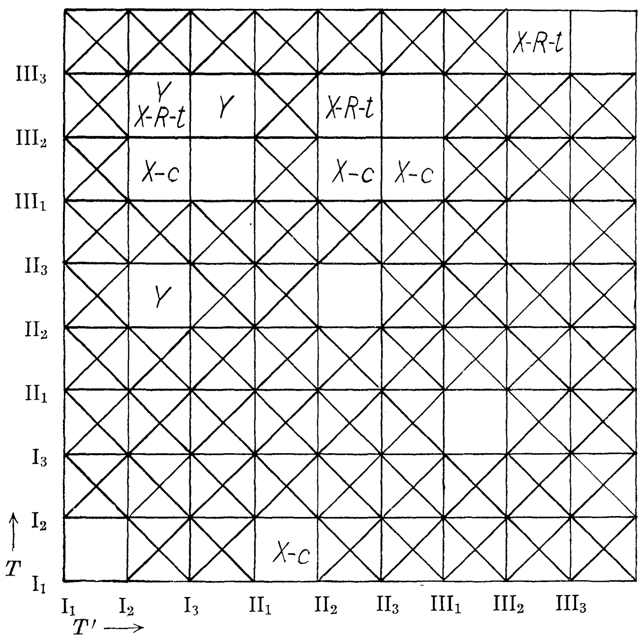
**6. The second state diagram.** The two theorems just proved as well as the state diagram in §3 enable us to determine the state diagram for a closed operator. We place *X-R-t* in a square to indicate that the state cannot occur if $X$ is reflexive and $\mathscr{D}(T')$ total. An *X-c* in a square will indicate that the state cannot occur if $X$ is complete and $T$ is closed.

This diagram is a generalization of the Taylor-Halberg state diagram for $T$ bounded on all of $X$.

**7. The spectrum of an operator and its conjugate.** In the present section we consider a linear transformation $T$, not necessarily bounded, with $\overline{\mathscr{D}} = X$ and $\mathscr{R}(T) \subset X$, where $X$ is a normed linear space. In this case, $T_\lambda = \lambda - T$ is well defined on $\mathscr{D}$, where $\lambda$ is a scalar.

DEFINITION 7.1. The values of $\lambda$ for which $T_\lambda$ has a bounded inverse with domain dense in $X$ form the *resolvent set* $\rho(T)$ of $T$; that is $T_\lambda \in \mathrm{I}_1 \cup \mathrm{II}_1$. The values of $\lambda$ for which $T_\lambda$ has an unbounded inverse

*Second State Diagram*




$Y$: Can't occur if $Y$ is complete.
$X$-$c$: Can't occur if $X$ is complete and $T$ is closed.
$X$-$R$-$t$: Can't occur if $X$ is reflexive and $\mathscr{D}(T')$ is total (in particular if $T$ is closed).

with domain dense in $X$ form the *continuous spectrum* $C\sigma(T)$, that is $T_\lambda \in \mathrm{I}_2 \cup \mathrm{II}_2$. The values of $\lambda$ for which $T_\lambda$ has an inverse whose domain is not dense in $X$ form the *residual spectrum* $R\sigma(T)$, that is $T_\lambda \in \mathrm{III}_1 \cup \mathrm{III}_2$. The values of $\lambda$ for which no inverse exists form the *point spectrum* $P\sigma(T)$, that is $\mathrm{I}_3 \cup \mathrm{II}_3 \cup \mathrm{III}_3$. The *spectrum* $\sigma(T)$ is defined to be the set of scalars not in $\rho(T)$.

These definitions can also be applied to $T'$. We would like now to draw inferences about the relationships between the above defined point sets for $T$ and $T'$. Since $\lambda - T' = (\lambda - T)'$, an appeal to the state diagram in §3 easily verify the following.

THEOREM 7.1. (a) *$\rho(T) = \rho(T')$ or equivalently, $\sigma(T) = \sigma(T')$.*

(b) $P\sigma(T) \subset R\sigma(T') \cup P\sigma(T')$ .

(c) $P(\sigma T') \subset R\sigma(T) \cup P\sigma(T)$ .

(d) $C\sigma(T) \subset R\sigma(T') \cup C\sigma(T')$ .

(e) $C\sigma(T') \subset C\sigma(T)$ .

(f) $R\sigma(T) \subset P\sigma(T')$ .

(g) $R\sigma(T') \subset C\sigma(T) \cup P\sigma(T)$ .

Suppose we now require that $T$ be closed, in addition to the other hypotheses mentioned at the beginning of the section. It is easy to see that $\lambda - T$ is also closed for $\lambda$ any scalar. Hence we can obtain the following theorem together with Theorem 7.1 by referring to the second state diagram in §6.

THEOREM 7.2. *If $T$ is a closed operator and $X$ is reflexive, then $C\sigma(T) = C\sigma(T')$ and $R\sigma(T') \subset P\sigma(T)$.*

REMARK. Let $X$ and $Y$ be Hilbert spaces. If $T^*$ is the adjoint of $T$, then $T^*$ may be put in place of $T'$ in using the first and second state diagrams. This is easy show by considering the fact that a Hilbert space is isometric to its conjugate space.

## REFERENCES

1. S. Banach, *Theorie des opérations linéarès*, Warsaw, 1932.

2. E. Hille and R. S. Phillips, *Functional analysis and semigroups*, Amer. Math. Soc. Colloquium Publ., **31**, (1957).

3. A. E. Taylor and C. J. A. Halberg, *General theorems about a linear operator and its conjugate*, J. Reine Angew. Math. **198** (1957), 93-111.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

# MEAN PLAY OF SUMS OF POSITIONAL GAMES

OLOF HANNER

**1. Introduction.** In 1953 Milnor studied certain positional 2-person games and defined what he called sums of such games [1]. He investigated the optimal strategies for these games and gave some information about them in terms of properties of the individual games.

In this paper we shall consider some other strategies for these sum games. They are in general not optimal. However, the difference between what a player gets when playing one of them instead of playing an optimal strategy can be estimated. For the sum of $n$ copies of the same game this difference is bounded for all $n$. Hence, in mean this difference is small for large $n$.

**2. Description of the games.** Essentially following Milnor [1] we describe the games as follows.

Each game contains a finite set of positions $P$. There are two players, $A_1$ and $A_2$. For each $p \in P$ and each player $A_i$, $i = 1, 2$, there is a set of possible moves $M_i(p) \subset P$. For each $p$ either both $M_1(p)$ and $M_2(p)$ contain at least one move or they are both vacuous. In the latter case $p$ is called an end position. For any chain $p_0, p_1, \cdots, p_l$ with $p_{j+1} \in M_1(p_j) \cup M_2(p_j)$, we shall have $p_j \neq p_k$ for $j \neq k$. The maximal number $l$ of steps in all such chains starting with $p_0 = p$ will be denoted by $l(p)$. Then

$$p_1 \in M_1(p) \cup M_2(p) \text{ implies } l(p_1) < l(p). \tag{2.1}$$

Note that a pass, $p \in M(p)$, is never possible. The positions with $l(p) = 0$ are just the end positions.

For each end position the payoff functions $k_1(p) = -k_2(p)$ are defined. They shall satisfy a condition given below. The players start with some position and move alternatively until an end position is reached. Then each player collects his payoff.

For each player $A_i$ and position $p$, let $v_i(p)$ be the value of the game for $A_i$ when it is his turn to move at position $p$. It is given by

$$\begin{aligned} v_i(p) &= k_i(p) && \text{for } l(p) = 0, \\ v_i(p) &= \max\{-v_{3-i}(p_1) \mid p_1 \in M_i(p)\} && \text{for } l(p) > 0. \end{aligned} \tag{2.2}$$

Because of (2.1) these formulas define $v_i(p)$ by induction on $l(p)$.

The numbers $k_i(p)$ are defined when $p$ is an end position. We require that they shall be given in such a way that

Received August 12, 1958.

(2.3) $$v_1(p) + v_2(p) \geqq 0 \qquad \text{for every } p \in P.$$

Since the value at $p$ for $A_i$ is $v_i(p)$ if he has the move and $-v_{3-i}(p)$ if the other player has the move, the amount $v_1(p) + v_2(p)$ is the gain for a player of having the move. Inequality (2.3) therefore says that it is at least as good to move as to pass (if this would be allowed).

**3. Sums of games.** We now define the sum of two games $G$ and $G'$. A position in the sum game $G + G'$ is a pair $(p, p') \in P \times P'$. A move in $G + G'$ is a move in one of the games $G$ and $G'$ and a pass in the other. Thus the moves in position $p + p' = (p, p')$ are

$$M_i(p + p') = M_i(p) \times p' \cup p \times M_i(p') .$$

We notice that

(3.1) $$l(p + p') = l(p) + l(p') .$$

In particular the condition is still satisfied that in a chain of successive positions all positions are different. The position $p + p'$ is an end position if and only if $p$ and $p'$ both are end positions. For the end positions we define $k_i(p + p')$ by

(3.2) $$k_i(p + p') = k_i(p) + k_i(p') .$$

It is not obvious that the sum of two games satisfying condition (2.3) also satisfies this condition. That this is the fact was proved by Milnor [1]. It will also be proved in §8 below as a consequence of Theorem 1.

It is clear that game addition is an associative and commutative operation and that the formulas corresponding to (3.1) and (3.2) hold for the sum of any finite number of games. A move in the sum of several games is a move in one of them and a pass in all the others.

**4. The main problem.** The problem for us will be to give good strategies for sums of games in terms of properties of the individual games. Then we must decide what kind of strategies we shall consider to be good.

One way to attack this problem is as follows. Consider $n$ copies of a game $G$ and take their sum $nG$. Let them all be started in the same position $p$. Then the value of the sum game is $v_i(np)$, where we have written $np$ instead of $p + \cdots + p$. Now, what happens to the mean value $v_i(np)/n$ when $n$ tends to infinity? In fact this number tends to a limit $m_i(p)$ which will be called the mean value of the game $G$ at $p$. In later sections we shall prove that $m_i(p)$ satisfies

(4.1) $$m_1(p) + m_2(p) = 0 ,$$

$$m_i(p) \leqq v_i(p) \ . \tag{4.2}$$

If we change $i$ to $3 - i$ in (4.2) and apply (4.1) we get

$$-v_{3-i}(p) \leqq m_i(p) \ . \tag{4.3}$$

Thus $m_i(p)$ lies between $v_i(p)$ and $-v_{3-i}(p)$ which represent the values for $A_i$ when the game is started at $p$ by him or by $A_{3-i}$ respectively.

Of a good strategy we now require that it guarantees at least $m_i(p)$. We see from (4.3) that though such a strategy may not guarantee $v_i(p)$, $A_i$ will nevertheless get more by playing it than by passing (if this would be allowed).

That the limit of $v_i(np)/n$ exists can be proved directly by an inequality given by Milnor [1, p. 294]:

$$v_i(p) - v_{3-i}(p') \leqq v_i(p + p') \leqq v_i(p) + v_i(p') \ .$$

We get

$$v_i((m + n)p) \leqq v_i(mp) + v_i(np) \ ,$$

and the existence of the limit of $v_i(np)/n$ follows (cf. [2], Erster Abschnitt, Aufgabe 98).

Another way of attacking our problem also leading to the number $m_i(p)$ will be used below. When a player shall move in a sum of games he chooses one game, say $G$, and there makes a move. Thereby he loses the possibility to make the move in one of the other games. If the value of this possibility is put equal to $t$ it is natural to compare the situation with the case when the player has to move in $G$ and pay the amount $t$ to the other player when moving. This will lead to the games $G_t$ and $G_t^*$ given in the next section. In this approach the value $m_i(p)$ is defined by induction on $l(p)$, thus by a finite procedure and not by a limit process.

*Conventions for the figures.* When giving examples of games by figures we use the following conventions. The positions are given by points and the moves indicated by segments joining them. A move by $A_1$ is a segment going down and to the left, a move by $A_2$ a segment going down and to the right. At an end position we put the value $k_1(p)$ and at any other position we put the two numbers $(m, \sigma)$, where $m = m_1(p)$ and $\sigma = \sigma(p)$ defined in the next section. Unless anything else is said, the game shall be played with the highest point as starting position.

EXAMPLE 1. Let $G$ be the game in Figure 1, and consider the sum of $n$ copies of $G$. First let us start at $p_2$ in all games. Then of course in about half of the games $A_1$ will get 7 and in the rest of them 3. Hence the mean value $m_1(p_2)$ is $(7 + 3)/2 = 5$. Analogously we get

$m_1(p_3) = -1$. If all games are started from $p_1$, it can be proved that an optimal play by both players is to choose the moves from $p_1$, $p_2$, and $p_3$ in this order of preference. Thus when both play optimally one move will first be made in all games. After these $n$ moves the players start attacking the positions $p_2$ in the games where $A_1$ made the move from $p_1$. At last the remaining games with positions $p_3$ are played. About 1/4 of the games will end in each of the four end positions. Hence the mean value $m_1(p_1)$ is $(7 + 3 + 0 - 2)/4 = 2$. The order of preference between $p_1$, $p_2$, $p_3$ is to be compared with the numbers $\sigma(p_1)$, $\sigma(p_2)$, $\sigma(p_3)$ which are defined in the next section. As given in the figure, $\sigma(p_1) = 3$, $\sigma(p_2) = 2$, $\sigma(p_3) = 1$. The number $\sigma(p)$ is in a sense the value of the move from position $p$.

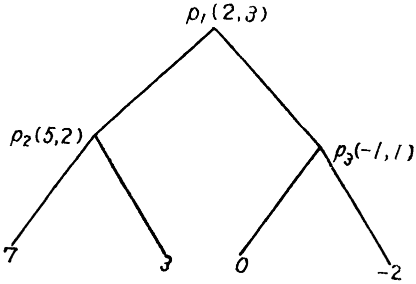


Figure 1

EXAMPLE 2. We change one of the payoff numbers in Figure 1 and get the game in Figure 2. Let us again consider the play of the sum of $n$ copies of the game. If all the games are started from $p_1$, the optimal play is now to choose the moves from $p_1$, $p_2$, $p_3$ in the order of preference: $p_2$, $p_1$, $p_3$, in accordance with the fact that $\sigma(p_2) = 5$, $\sigma(p_1) = 4$, and $\sigma(p_3) = 1$. Thus if $A_1$ moves from $p_1$ to $p_2$ in a game, $A_2$ will immediately move in the same game. Thus all games with only one possible exception will end in the position with payoff $k_1(p) = 3$. Thus $m_1(p_1) = 3$. Only if $A_2$ has the first move one game will end in another end position, the one with $k_1(p) = 0$.

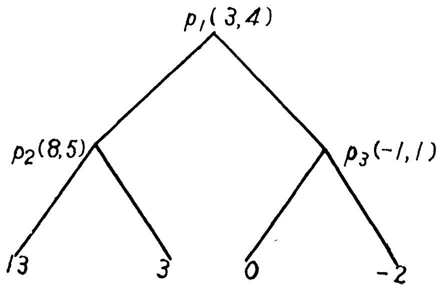


Figure 2

**5. The games $G_t$ and $G_t^*$.** Let $G$ be a game satisfying as usual the condition (2.3). Let $t$ be a real number $\geqq 0$. When $l(p) = 0$ put for $i = 1, 2$,

$$v_i(p\,;t) = k_i(p)\ ,$$

$$m_i(p) = k_i(p)\ ,$$

$$\sigma(p) = 0\ .$$

For each $p$ with $l(p) > 0$, we define four functions in $t : u_1(p\,;t)$, $u_2(p\,;t)$,

$v_1(p;t)$, $v_2(p;t)$ and three numbers $m_1(p)$, $m_2(p)$, and $\sigma(p)$. They shall satisfy (5.1)—(5.7).

(5.1) Each function $u_i(p;t)$ and $v_i(p;t)$ is a continuous function for $t \geqq 0$ with a derivative for all but a finite number of $t$-values. In each interval between these exception values the function is linear with derivative 0 or $-1$. For $t$ greater than the exception values the function $u_i(p;t)$ has derivative $-1$ and $v_i(p;t)$ has derivative 0,

(5.2) $$v_i(p;0) = v_i(p) ,$$

(5.3) $$u_i(p;t) = \max \{-v_{3-i}(p_1;t) \mid p_1 \in M_i(p)\} - t ,$$

(5.4) $$u_i(p;0) = v_i(p) ,$$

(5.5) $$\sigma(p) = \min \{t \mid t \geqq 0, u_1(p;t) + u_2(p;t) = 0\} ,$$

(5.6) $$m_i(p) = u_i(p;\sigma(p)) ,$$

(5.7) $$\begin{aligned} v_i(p;t) &= u_i(p;t) && \text{for } 0 \leqq t \leqq \sigma(p), \\ &= m_i(p) && \text{for } t > \sigma(p). \end{aligned}$$

We shall see below that these conditions are related to two games $G_t$ and $G_t^*$. Let us first show, however, that they define our functions and numbers by induction on $l(p)$.

For $l(p) = 0$ the function $v_i(p;t)$ is constant and equal to $v_i(p)$, hence it satisfies (5.1) and (5.2). Let $l(p) > 0$ and suppose that for each $p_1$ with $l(p_1) < l(p)$ and in particular for each $p_1 \in M_i(p)$ we have $v_i(p_1;t)$ defined satisfying (5.1) and (5.2). Then $u_i(p;t)$ can be defined by (5.3). By (5.1) for each $v_{3-i}(p_1;t)$ we get immediately (5.1) for $u_i(p;t)$ and by (5.2) for each $v_{3-i}(p_1;t)$ and by (2.2) we get (5.4). By (5.4) and (2.3) we have $u_1(p;0) + u_2(p;0) \geqq 0$ and by (5.1) for $u_i(p;t)$ we have $u_1(p;t) + u_2(p;t) \to -\infty$ when $t \to \infty$. Hence, since $u_i(p;t)$ is continuous, the set in (5.5) is not vacuous and $\sigma(p)$ is defined and $\geqq 0$. Then (5.6) and (5.7) will define $m_i(p)$ and $v_i(p;t)$. That $v_i(p;t)$ satisfies (5.1) and (5.2) follows from the corresponding facts for $u_i(p;t)$. Hence the induction will work.

EXAMPLE 3. We give in the diagram in Figure 4 the functions $u_1(p;t)$, $v_1(p;t)$, $-u_2(p;t)$, $-v_2(p;t)$ for the game in Figure 3 and also the values $m_1(p)$ and $\sigma(p)$ for the same game.

Properties (5.1)—(5.7) give some further formulas. Since (5.1)—(5.7) are only known to be true for $l(p) > 0$, we have to verify separately the case $l(p) = 0$ each time we get a formula which has a meaning even in this case. Note that $u_i(p;t)$ is not defined when $l(p) = 0$.

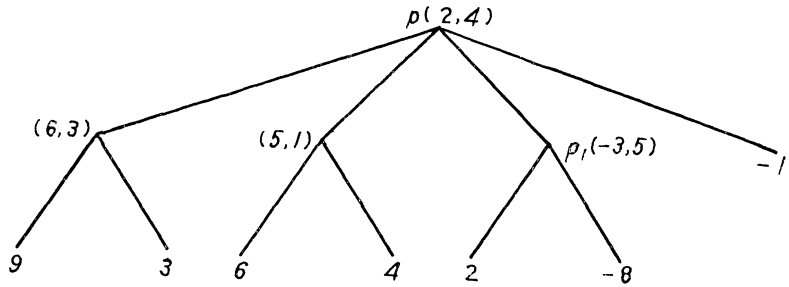
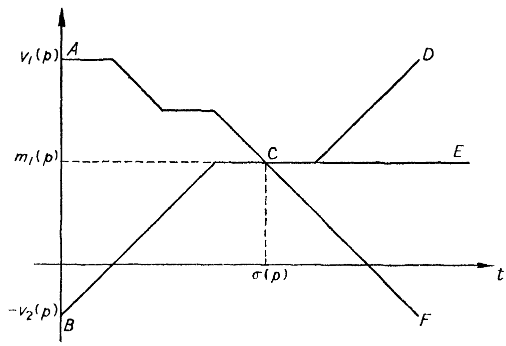



Figure 3




$u_1(p\,;t) : ACF, \quad -u_2(p\,;t) : BCD,$
$v_1(p\,;t) : ACE, \quad -v_2(p\,;t) : BCE.$

Figure 4

Since by (5.1) $u_i(p\,;t)$ is a decreasing function, (5.6) and (5.7) give

(5.8) $$v_i(p\,;t) = \max\,\{u_i(p\,;t),\, m_i(p)\}\ .$$

Hence, in particular

(5.9) $$v_i(p\,;t) \geqq m_i(p)\ .$$

By (5.5) and (5.6)

(5.10) $$m_1(p) + m_2(p) = 0\ .$$

Both (5.9) and (5.10) are true also when $l(p) = 0$ as is easily verified. For any $p$ they imply

(5.11) $$v_1(p\,;t) + v_2(p\,;t) \geqq 0\ .$$

By (5.2) and (5.9) we obtain

(5.12) $$v_i(p) \geqq m_i(p)\ .$$

Since $v_i(p\,;t)$ has derivative 0 or $-1$, we have for $t_1 < t_2$

$$0 \leqq v_i(p\,;t_1) - v_i(p\,;t_2) \leqq t_2 - t_1\ .$$

Apply this for $t_1 = 0$ and $t_2 = \sigma(p)$. Then by (5.2), (5.6), and (5.7)

$$v_i(p) \leqq m_i(p) + \sigma(p) . \tag{5.13}$$

Both (5.12) and (5.13) are also true when $l(p) = 0$. For any $p$ they give a lower and an upper bound for $v_i(p)$.

We are now ready to define the two games $G_t$ and $G_t^*$ mentioned above. Both are defined for each $t \geqq 0$. They are played with the positions in $G$. The players play alternatively. But each time a player makes a move into a new position in $G$ he has to pay $t$ to the other player. Thus for large $t$ it will be expensive to make a move. Therefore we introduce a new possibility. When $A_i$ has the move in $G_t$ he is allowed to stop the game instead of moving. In $G_t^*$ the same possibility is open except in the starting position, where the player who begins really must move (and pay $t$). When $A_i$ stops at $p$ he collects $m_i(p)$. Then $A_{3-i}$ gets $m_{3-i}(p)$ by (5.10). The value of $G_t$ at $p$ is $v_i(p\,;t)$ and the value of $G_t^*$ started at $p$ is $u_i(p\,;t)$. This is seen by induction from (5.3) and (5.8).

For large $t$ it is a disadvantage to have to start in $G_t^*$. The starting player will make a move and pay $t$ and the other player will then immediately stop the game. Thus if $t$ is great enough the starting player will always lose. Thus $G_t^*$ does not satisfy (2.3). The game $G_t$, however, satisfies (2.3) as is seen from (5.11). In fact we have introduced the number $m_i(p)$ and the possibility to stop just in order to save this property. The number $m_i(p)$ is defined by (5.5) and (5.6) as the value of $G_t^*$ with starting position $p$, when $t$ has become so large that it is no more an advantage to have the first move in $G_t^*$. The lowest $t$-value of this kind is $\sigma(p)$.

6. **The $t$-optimal moves.** We will call a move in $G$ a $t$-optimal move if it is optimal in $G_t$. Thus $p_1 \in M_i(p)$ is $t$-optimal if

$$v_i(p\,;t) = -v_{3-i}(p_1\,;t) - t . \tag{6.1}$$

There is a $t$-optimal move at $p$ for $A_i$ if $v_i(p\,;t) = u_i(p\,;t)$. Thus we get from (5.7) the following important fact: If $\sigma(p) \geqq t$ and if $p$ is not an end position there always exist $t$-optimal moves for both players.

If $\sigma(p) \leqq t$ we have $v_i(p\,;t) = m_i(p)$, and an optimal play of $G_t$ is to stop the game at $p$ and collect $m_i(p)$.

Now study a sequence $p_0, p_1, p_2, \cdots, p_l$ of positions that develop when the players play alternatively and make $t$-optimal moves. If $\sigma(p_l) > t$, there are $t$-optimal moves at $p_l$. Therefore the sequence can be continued and we can go on in this way until we reach a position $p$ with $\sigma(p) \leqq t$. We suppose this already done, so that $\sigma(p_l) \leqq t$.

We want to get some formulas for $m_i(p_k)$, $0 \leqq k \leqq l$. Since all moves in the sequence are $t$-optimal we know that a player cannot get more when playing $G_t$ by stopping at a position $p_k$, $k < l$, than by moving into $p_{k+1}$. Thus if $A_i$ makes the first move and if we put $v_i(p_0\,; t) = v$, we get

$$(6.2) \qquad m_i(p_{2k}) \leqq v \qquad \text{if } 0 \leqq 2k < l,$$

$$(6.3) \qquad m_i(p_{2k+1}) \geqq v + t \qquad \text{if } 1 \leqq 2k+1 < l,$$

where the term $+t$ in (6.3) is the amount $A_i$ shall have when the game is stopped after an odd number of moves as a compensation for the fact that he has made one more move than $A_{3-i}$, each player paying $t$ when moving in $G_t$. Since $\sigma(p_l) \leqq t$, an optimal play at $p_l$ in $G_t$ is to stop the game. Hence

$$(6.4) \qquad m_i(p_l) = v \qquad \text{if } l \text{ is even},$$

$$(6.5) \qquad m_i(p_l) = v + t \qquad \text{if } l \text{ is odd}.$$

Formulas (6.2)—(6.5) could also have been deduced from (6.1). Since all moves are $t$-optimal we get $v_i(p_0\,; t) = -v_{3-i}(p_1\,; t) - t = v_i(p_2\,; t) = -v_{3-i}(p_3\,; t) - t = \cdots$ and (6.2)—(6.5) follow if we apply (5.9) and (5.10) and the fact that since $\sigma(p_l) \leqq t$, we have by (5.7), $m_j(p_l) = v_j(p_l\,; t)$ for $j = 1, 2$.

EXAMPLE 4. The game in Figure 5 shows that strong inequality may hold in (6.2) and (6.3). All the moves which lead from $p_0$ to $p_5$ are 1-optimal and $v = v(p_0\,; 1) = 1$.

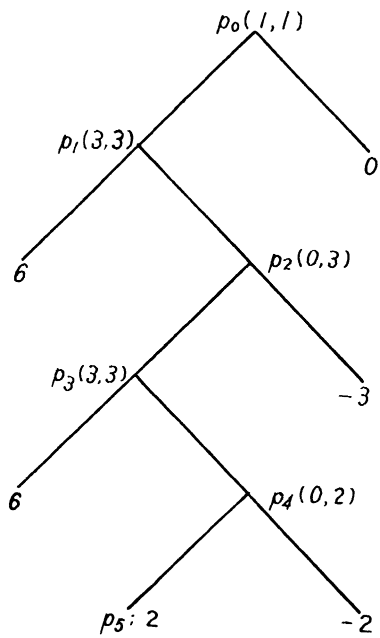


Figure 5

Let now only one player make $t$-optimal moves when playing $G_t$. He will get at least as much as when also the other player makes $t$-optimal moves. Thus we can get some formulas corresponding to (6.2)—(6.5). We put them together into two lemmas.

LEMMA 1. *Let $p_0, p_1, \cdots, p_l$ be a sequence of positions in $G$ such that $p_{2k+1} \in M_i(p_{2k})$, where $p_{2k+1}$ is a $t$-optimal move at $p_{2k}$, and such that $p_{2k+2} \in M_{3-i}(p_{2k+1})$. Then if $v_i(p_0\,; t) = v$, we have*

$$(6.6) \qquad m_i(p_{2k+1}) \geqq v + t\,,$$

$$(6.7) \qquad m_i(p_l) \geqq v$$

*if $\sigma(p_l) \leqq t$ and $l$ is even.*

LEMMA 2. *Let $p_0, p_1, \cdots, p_l$ be a sequence of positions in $G$ such that $p_{2k+1} \in M_i(p_{2k})$ and $p_{2k+2} \in M_{3-i}(p_{2k+1})$, where $p_{2k+2}$ is a $t$-optimal move at $p_{2k+1}$. Then if $v_i(p_0\,; t) = v$ we have*

$$m_i(p_{2k}) \leqq v\,, \tag{6.8}$$

$$m_i(p_l) \leqq v + t \quad \textit{if } \sigma(p_l) \leqq t \textit{ and } l \textit{ is odd.} \tag{6.9}$$

**7. The mean strategies for sum games.** We now go to our main subject, sums of games.

THEOREM 1. *Let us start the games $G_1, \cdots, G_n$ in positions $q_1, \cdots, q_n$. Put*

$$m_i = m_i(q_1) + \cdots + m_i(q_n)\,,$$

$$\sigma = \max\{\sigma(q_r) \mid 1 \leqq r \leqq n\}\,.$$

*Then the value $v_i(q_1 + \cdots + q_n)$ for $A_i$ when he starts at $q_1 + \cdots + q_n$ in $G_1 + \cdots + G_n$ satisfies*

$$m_i \leqq v_i(q_1 + \cdots + q_n) \leqq m_i + \sigma\,.$$

*Proof.* We proceed by induction on $l(q_1 + \cdots + q_n)$. When $l(q_1 + \cdots + q_n) = 0$, all $q_r$ are end positions and our theorem follows directly from $m_i(q_r) = k_i(q_r)$ and $\sigma(q_r) = 0$. By (2.1) we know that if one or several moves are made from $q_1 + \cdots + q_n$ we come to a position, say $p_1 + \cdots + p_n$, with

$$l(p_1 + \cdots + p_n) < l(q_1 + \cdots + q_n)\,.$$

Hence when proving our theorem we may assume that it is true for all positions obtainable from $q_1 + \cdots + q_n$ by one or several moves.

By symmetry we may specialize in the proof so that $i = 1$, i.e. $A_1$ makes the first move. We then want a strategy for him that secures the amount $m_1$ and a strategy for $A_2$ such that $A_1$ cannot get more than $m_1 + \sigma$. These strategies can be formulated together.

($\alpha$) Always make a $\sigma$-optimal move in one of the games $G_1, \cdots, G_n$.

($\beta$) Except for the first move, play in the game, in which the other player has just played.

In general it will not be possible to follow this strategy through the whole play of the game, since there are not $\sigma$-optimal moves in all positions. The strategy shall therefore be used during a period in the beginning of the play. In the position at the end of this period the induction hypothesis will be used. The length of the period depends upon the moves made. We give two possibilities to end the period.

($\gamma_1$) The other player plays in a game $G_r$ and there leaves a position $p_r$ with $\sigma(p_r) \leqq \sigma$.

($\gamma_2$) Positions $p_r$, $1 \leqq r \leqq n$, are reached for which $\sigma(p_r) \leqq \sigma$.

We have to show that a player can follow ($\alpha$) and ($\beta$) until ($\gamma_1$) or ($\gamma_2$) occurs. We first see that $A_1$ always can make his first move. In fact, by the definition of $\sigma$ there is a $q_r$ with $\sigma(q_r) = \sigma$. Thus there is a $\sigma$-optimal move in $G_r$. For all later moves the player following the strategy shall play in the position $p_r$ which the other player has just left. Then if ($\gamma_1$) does not occur, $\sigma(p_r) > \sigma$ and there is a $\sigma$-optimal move at $p_r$. Hence the game can be continued until ($\gamma_1$) occurs or until the player following the strategy ends the whole sum game by playing into an end position. Then $\sigma(p_r) = 0$ for all games $G_r$ and ($\gamma_2$) is satisfied. Hence it is possible to follow ($\alpha$) and ($\beta$) until ($\gamma_1$) or ($\gamma_2$) occurs.

In order to be able to use the induction hypothesis we have to compare $m_1$ with

$$m_1(p_1) + \cdots + m_1(p_n) ,$$

where $p_r$ is the position in $G_r$ at the end of the period. Therefore we first compare $m_1(q_r)$ with $m_1(p_r)$ for each $r$. Hence we are interested in those moves in the period that are made in $G_r$. Note that when at least one player follows the strategy, ($\beta$) implies that these moves are played alternatively by the players. Thus for each $G_r$ we are able to apply Lemmas 1 and 2 of the preceding section with $t = \sigma$. Since $\sigma \geqq \sigma(q_r)$, the number $v = v_i(q_r\,; \sigma)$ in these lemmas is $= m_i(q_r)$.

Let first $A_1$ follow the strategy. Denote by $p_r$ the position in $G_r$ at the end of the period. Then if the move into $p_r$ is made by $A_2$, we know, since $A_1$ follows ($\beta$), that this move is the last move in the period, and whether the period ends with ($\gamma_1$) or ($\gamma_2$) we get $\sigma(p_r) \leqq \sigma$ in this game $G_r$. Using the fact (5.10) for $q_r$ and $p_r$, $1 \leqq r \leqq n$, we apply Lemma 1 with $i = 1$ and Lemma 2 with $i = 2$. Then (6.6), (6.7), (6.8), and (6.9) imply respectively the following four formulas, depending upon who makes the first move and the last move in $G_r$.

(7.1) $$m_1(p_r) \geqq m_1(q_r) + \sigma \qquad A_1 \text{ first and last move,}$$

(7.2) $$m_1(p_r) \geqq m_1(q_r) \qquad A_1 \text{ first move, } A_2 \text{ last move,}$$

(7.3) $$m_1(p_r) \geqq m_1(q_r) \qquad A_2 \text{ first move, } A_1 \text{ last move,}$$

(7.4) $$m_1(p_r) \geqq m_1(q_r) - \sigma \qquad A_2 \text{ first and last move.}$$

We add the trivial fact

(7.5) $$m_1(p_r) = m_1(q_r) \qquad \text{if no move is made in } G_r.$$

Formulas (7.1)—(7.5) can be taken together in one formula

$$(7.6) \qquad m_1(p_r) \geqq m_1(q_r) + l_{1r}\sigma - l_{2r}\sigma \ ,$$

where $l_{ir}$ is the number of moves made by $A_i$ in $G_r$ during the period. Let us take the sum of the inequalities (7.6) for all $r$. Then

$$(7.7) \qquad m_1(p_1) + \cdots + m_1(p_n) \geqq m_1 + l_1\sigma - l_2\sigma \ ,$$

where $l_i$ is the number of moves made by $A_i$ during the period.

If the number of moves in the period is even we have $l_1 = l_2$. $A_1$ who makes the first move in the period shall also make the first move after the period (if there is any move to be made). $A_1$ can play so after the period that he secures $v_1(p_1 + \cdots + p_n)$. By the induction hypothesis this is $\geqq m_1(p_1) + \cdots + m_1(p_n)$ which by (7.7) is $\geqq m_1$. Hence we have shown that $A_1$ has been able to play from $q_1 + \cdots + q_n$ so as to secure $m_1$, and the left-hand inequality of our theorem is proved in this case.

We also have to consider the case that the period contains an odd number of moves. Then since $A_1$ makes the first move he also makes the last move and the period is not ended by $(\gamma_1)$, hence by $(\gamma_2)$. Thus $\sigma(p_r) \leqq \sigma$ for each $G_r$. We have now $l_1 = l_2 + 1$. $A_1$ can play so after the period that he secures $-v_2(p_1 + \cdots +$ By$p_n)$. the induction hypothesis, by $\sigma(p_r) \leqq \sigma$, and by (7.7) we get

$$\begin{aligned} -v_2(p_1 + \cdots + p_n) &\geqq -m_2(p_1) - \cdots - m_2(p_n) - \max\{\sigma(p_r)\} \\ &\geqq -m_2(p_1) - \cdots - m_2(p_n) - \sigma \\ &= m_1(p_1) + \cdots + m_1(p_n) - \sigma \\ &\geqq m_1 \ . \end{aligned}$$

Hence the left-hand inequality of the theorem is proved even in this case.

In order to prove the right-hand inequality of the theorem we let $A_2$ follow the strategy. Then by $(\beta)$ $A_1$ makes the first move in each $G_r$ (if there is any move in $G_r$ during the period). Lemma 2 with $i = 1$ gives now depending upon who makes the last move in $G_r$

$$(7.8) \qquad m_1(p_r) \leqq m_1(q_r) \qquad\qquad A_2 \text{ last move,}$$

$$(7.9) \qquad m_1(p_r) \leqq m_1(q_r) + \sigma \qquad\qquad A_1 \text{ last move.}$$

Proceeding as above we get a formula like (7.7), namely

$$(7.10) \qquad m_1(p_1) + \cdots + m_1(p_n) \leqq m_1 + l_1\sigma - l_2\sigma \ .$$

If the period contains an odd number of moves, $l_1 = l_2 + 1$. $A_2$ makes then the first move after the period (if there is any move to be made). He can therefore play so that $A_1$ gets at most $-v_2(p_1 + \cdots + p_n)$. By the induction hypothesis and by (7.10)

$$\begin{aligned}-v_2(p_1 + \cdots + p_n) &\leqq -m_2(p_1) - \cdots - m_2(p_n) \\ &= m_1(p_1) + \cdots + m_1(p_n) \\ &\leqq m_1 + \sigma ,\end{aligned}$$

so that the right-hand inequality is proved in this case.

Finally if the period contains an even number of moves, $l_1 = l_2$, and the period ends by $(\gamma_2)$, so that $\sigma(p_r) \leqq \sigma$. Then $A_1$ gets at most $v_1(p_1 + \cdots + p_n)$ and by the induction hypothesis and by (7.10)

$$\begin{aligned}v_1(p_1 + \cdots + p_n) &\leqq m_1(p_1) + \cdots + m_1(p_n) + \max\{\sigma(p_r)\} \\ &\leqq m_1(p_1) + \cdots + m_1(p_n) + \sigma \\ &\leqq m_1 + \sigma ,\end{aligned}$$

and the right-hand inequality is proved even in this case.

This completes the proof of Theorem 1.

In the proof just completed the strategy given by $(\alpha)$ and $(\beta)$ is used only in a period in the beginning of the play. When this period is ended we have used the induction hypothesis in the proof of the theorem. This means, however, that we shall start counting a new period and then again apply $(\alpha)$ and $(\beta)$. Continuing in this way we get the following consequence of the proof of Theorem 1.

THEOREM 2. *Make the same assumptions as in Theorem* 1. *Suppose one player, $A_k$, follows a strategy satisfying* (a)—(d) *below. Then $A_i$, the player making the first move, will get at least $m_i$ when $k = i$ and at most $m_i + \sigma$ when $k = 3-i$.*

(a) *Divide the moves made by the two players into periods.*

(b) *For each period let $\tau$ be the maximum of $\sigma(p_r)$ for the positions $p_r$ at the beginning of the period. With this $\tau$ defined for a period, always make $\tau$-optimal moves in the period.*

(c) *Except for the first move in a period play in the game in which the other player has just played.*

(d) *Start counting a new period when one of the following two situations occurs,*

($d_1$) *the other player plays in $G_r$ into a position $p_r$ with $\sigma(p_r) \leqq \tau$,*

($d_2$) *positions $p_r$ with $\sigma(p_r) \leqq \tau$ are reached in all $G_r$, $1 \leqq r \leqq n$.*

We call the strategies that satisfies (a)—(d) of this theorem mean strategies.

**8. Properties of $m_i(p)$ and $\sigma(p)$.** By Theorem 1 we easily prove the fact that the sum of games satisfying (2.3) also satisfies (2.3) (proved by Milnor [1, p. 294]). In fact by Theorem 1

$$v_i(q_1 + \cdots + q_n) \geqq m_i .$$

Since $m_1(q_r) + m_2(q_r) = 0$ for each $r$, we have $m_1 + m_2 = 0$. Hence

$$v_1(q_1 + \cdots + q_n) + v_2(q_1 + \cdots + q_n) \geqq 0 ,$$

which is (2.3) for $G_1 + \cdots + G_n$.

Thus $G_1 + \cdots + G_n$ is a game of the kind described in §2. We can therefore apply §5 and define e.g. $u_i(q_1 + \cdots + q_n ; t)$, $m_1(q_1 + \cdots + q_n)$, and $\sigma(q_1 + \cdots + q_n)$.

THEOREM 3. *Let us start the games $G_1, \cdots, G_n$ in positions $q_1, \cdots, q_n$. Then*

$$m_i(q_1 + \cdots + q_n) = m_i(q_1) + \cdots + m_i(q_n) , \tag{8.1}$$

$$\sigma(q_1 + \cdots + q_n) \leqq \max \{\sigma(q_r) \mid 1 \leqq r \leqq n\} . \tag{8.2}$$

The right-hand side of these formulas is just $m_i$ and $\sigma$ respectively defined in Theorem 1.

*Proof.* We need the following lemma.

LEMMA 3.

$$u_i(q_1 + \cdots + q_n ; \sigma) = m_i \quad \textit{when} \quad l(q_1 + \cdots + q_n) > 0.$$

Before proving the lemma let us see that Theorem 3 follows from it. If $l(q_1 + \cdots + q_n) = 0$, (8.1) and (8.2) are certainly true. If $l(q_1 + \cdots + q_n) > 0$ we get from Lemma 3, since $m_1 + m_2 = 0$,

$$u_1(q_1 + \cdots + q_n ; \sigma) + u_2(q_1 + \cdots + q_n ; \sigma) = 0 .$$

Then (8.2) follows from (5.5). We also see from (5.5) and the fact that $u_i(q_1 + \cdots + q_n ; t)$, $i = 1, 2$, are decreasing functions in $t$, that they are constant in the interval $(\sigma(q_1 + \cdots + q_n), \sigma)$. Then (8.1) follows from (5.6) and Lemma 3.

*Proof of Lemma* 3. The proof will be somewhat similar to that of Theorem 1. Without losing generality we put $i = 1$. We make the induction hypothesis that Theorem 3 is true for all $p_1 + \cdots + p_n$ obtainable from $q_1 + \cdots + q_n$ by one or several moves. We will prove

$$u_1(q_1 + \cdots + q_n ; \sigma) \geqq m_1 , \tag{8.3}$$

$$u_1(q_1 + \cdots + q_n ; \sigma) \leqq m_1 . \tag{8.4}$$

Of course they together will give Lemma 3. The number $u_1(q_1 + \cdots + q_n ; \sigma)$ is the value for $A_1$ in the game $(G_1 + \cdots + G_n)^*_\sigma$. To prove (8.3) and (8.4) we define strategies for $A_1$ and $A_2$ in this game: Follow ($\alpha$) and ($\beta$) of the proof of Theorem 1. Unless the other player stops the game in some position, continue until $(\gamma_1)$ occurs and then stop the game. When the game is stopped at $p_1 + \cdots + p_n$, $A_1$ collects $m_1(p_1 + \cdots + p_n)$. If then $A_1$ has made $l_1$ and $A_2$ $l_2$ moves ($l_1 = l_2$ or $l_1 = l_2 + 1$), $A_1$ has paid $l_1\sigma$ to $A_2$ and got $l_2\sigma$ from him. Hence the result will be that $A_1$ gets

$$m_1(p_1 + \cdots + p_n) - l_1\sigma + l_2\sigma .$$

Since by the induction we may apply Theorem 3, this is equal to

$$m_1(p_1) + \cdots + m_n(p_n) - l_1\sigma + l_2\sigma .$$

Thus in order to prove (8.3) and (8.4) we only need to verify that (7.7) and (7.10) are true when $A_1$ and $A_2$ respectively use the strategy described above.

Let $A_1$ follow the strategy, and let $p_r$ be the position in $G_r$ when the game is stopped. Then if the move into $p_r$ is made by $A_2$, we know since $A_1$ follows $(\beta)$, that this is the last move made before the game is stopped by $A_1$. Hence $(\gamma_1)$ is true, and we have $\sigma(p_r) \leqq \sigma$ for this game $G_r$. The proof of the formulas (7.1)—(7.4) now follows as in the proof of Theorem 1, and (7.7) will again be a consequence of these formulas. Hence we have given a strategy for $A_1$ in $(G_1 + \cdots + G_n)^*_\sigma$ which secures $m_1$. Thus (8.3) is proved.

Similarly if $A_2$ follows the strategy, we verify (7.8) and (7.9) thereby proving (7.10). Thus we have given a strategy for $A_2$ in $(G_1 + \cdots + G_n)^*_\sigma$ such that $A_1$ gets $\leqq m_1$. This proves (8.4). Thus Lemma 3 is proved and also Theorem 3.

Theorem 3 can be looked upon as a sharper form of Theorem 1. In fact we get Theorem 1 from Theorem 3 simply by applying (5.12) and (5.13) for $p = q_1 + \cdots + q_n$.

Let now the games $G_1, \cdots, G_n$ be $n$ copies of one and the same game $G$ and let $p_1, \cdots, p_n$ correspond to $p$ in $G$. We write $np$ for $p_1 + \cdots + p_n$. By Theorem 1

$$nm_i(p) \leqq v_i(np) \leqq nm_i(p) + \sigma(p) .$$

Divide by $n$ and let $n \to \infty$. Then, because of (5.10), we get the following result.

THEOREM 4. *The two expressions*

$$\frac{1}{n} v_i(np) \qquad \text{and} \qquad \frac{1}{n}(-v_{3-i}(np))$$

*which represent the mean value for $A_i$ in the sum of $n$ equal games when he or the other player has the first move, both tend to the same limit $m_i(p)$ when $n \to \infty$.*

This theorem justifies the name mean value for the number $m_i(p)$.

The name mean strategies for the strategies described in Theorem 2 is chosen, since it secures the mean value for the player who makes the first move.

We know by Theorem 3 that

$$m_i(p_1 + \cdots + p_n) = m_i(p_1) + \cdots + m_i(p_n) \tag{8.5}$$

and get from (5.10) and (5.12)

$$-v_{3-i}(p) \leqq m_i(p) \leqq v_i(p) . \tag{8.6}$$

Let us show that the two properties (8.5) and (8.6) determine $m_i(p)$ uniquely. Let $m(p)$ be given for all $p$ satisfying (8.5) and (8.6). We get

$$-v_{3-i}(np) \leqq nm(p) \leqq v_i(np) .$$

Divide by $n$ and let $n \to \infty$. Then, by Theorem 4 we get $m(p) = m_i(p)$, showing the uniqueness of $m_i(p)$.

## 9. Both players use mean strategies.

THEOREM 5. *Let in a sum $G_1 + \cdots + G_n$ both players follow a mean strategy, such as described by* (a)—(d) *in Theorem* 2. *Then*

(1) *the players will count the same periods,*

(2) *in each period both players will make all their moves in only one of the games $G_r$,*

(3) *the number $\tau$ defined by* (b) *of Theorem* 2 *is a decreasing function of the period,*

(4) *if to $m_i(q_1) + \cdots + m_i(q_n)$, where $q_r$ is the starting position of the game $G_r$, $1 \leqq r \leqq n$, we add $\tau$ for each move $A_i$ makes and $-\tau$ for each move $A_{3-i}$ makes, where $\tau$ is defined by* (b) *for the period containing the move, then the result will be $A_i$'s payoff.*

*Proof.* Here (1) will follow by induction if we show that the first period ends at the same moment for both players. When both players play in their first periods (c) implies that they both move in the same game, say in $G_s$. Then for $r \neq s$, $p_r = q_r$ for all positions $p_1 + \cdots + p_n$ that are reached in the period and therefore since $\sigma(q_r) \leqq \sigma$ by the definition of $\sigma$ (see Theorem 1), we get $\sigma(p_r) \leqq \sigma$, $r \neq s$. Thus when ($d_1$) occurs for one player ($d_2$) also occurs and since ($d_2$) is symmetric with respect to the two players the first period will be the same for both players. This proves (1).

When we know that the players count the same periods, (2) is a simple consequence of (c). (3) follows from the fact that each period ends with ($d_2$).

To prove (4) it will be sufficient to show that if in $G_r$, $q_r$ is the position at the beginning of a period and $p_r$ is the position at the end of the same period then whether $A_i$ or $A_{3-i}$ starts the period,

$$(9.1)\qquad m_i(p_1) + \cdots + m_i(p_n) = m_i(q_1) + \cdots + m_i(q_n) + l_i\tau - l_{3-i}\tau ,$$

where $l_i$ is the number of moves by $A_i$ in the period. Since $p_r = q_r$ for $r \neq s$, where $G_s$ is the game in which all moves are made during the period, (9.1) reduces to

$$(9.2)\qquad m_i(p_s) = m_i(q_s) + l_i\tau - l_{3-i}\tau .$$

If $A_i$ makes the first move in the period, (9.2) follows from (6.4) and (6.5). In fact these two formulas are proved for the case when both players make $t$-optimal moves until a position $p_l$ is reached with $\sigma(p_l) \leqq t$. But putting $t = \tau$ we get in our case by ($d_2$) that for the final position $p_s$ of the period, $\sigma(p_s) \leqq \tau$.

If $A_{3-i}$ makes the first move in the period, (9.2) is just proved with $3 - i$ substituted for $i$. However, the formula thus obtained reduces to (9.2) by the use of $m_i(p) + m_{3-i}(p) = 0$.

Thus Theorem 5 is proved.

Since $\tau$ is decreasing we see by (4) of Theorem 5 that $A_i$'s payoff is the sum of $m_i = m_i(q_1) + \cdots + m_i(q_n)$ and a sequence of terms with alternating signs and decreasing modules. If $A_i$ starts playing, the first term is positive and equal to $\sigma = \max\{\sigma(q_r)\}$ and the sum of the terms in the sequence is therefore $\geqq 0$ and $\leqq \sigma$, and $A_i$ will get at least $m_i$ and at most $m_i + \sigma$. This last result is of course contained in Theorem 2. Theorem 2 says even more, since it says that a mean strategy always guarantees a certain amount even if used against a player which plays any strategy, e.g. an optimal strategy.

**10. Some examples.** Conditions (a)—(d) of Theorem 2 do not in general determine a unique strategy. There are still some choices which the player may use to get as good result as possible. Thus there may be different $\tau$-optimal moves in the same game and, when the first move of a period shall be made, there may be several games in which there are $\tau$-optimal moves. In this connection it may be worth while to notice that there may be a $\tau$-optimal move even in a position $p$ with $\sigma(p) < \tau$. The number $\tau$ is determined as the maximum of $\sigma(p_r)$, $1 \leqq r \leqq n$ when the period starts, but it is not necessary to start the period in one of the games for which $\sigma(p_r)$ reaches this maximum. There may be $\tau$-optimal moves even in other games.

EXAMPLE 5. Let us study the game given in Figure 3. The move $p_1 \in M_2(p)$ is $t$-optimal for $A_2$ even when $4 < t \leqq 5$. In fact for these $t$-values $u_2(p\,; t) = v_2(p\,; t) = m_2(p)$ so that there must be a $t$-optimal move for $A_2$.

If a position has to be played in optimal way it is unimportant if this position is the starting position of the game or if it is a position which has developed during the play. This is not the case when mean strategies are used.

EXAMPLE 6. Compare the game in Figure 6 started by $A_2$ and the game in Figure 7 started by $A_1$. When $A_2$ has moved into $p_1$ in Figure 6 the situation for $A_1$ will be the same as when he starts in $p_1$ in Figure

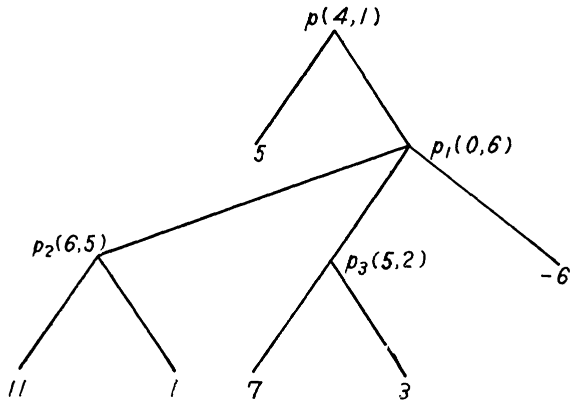



Figure 6

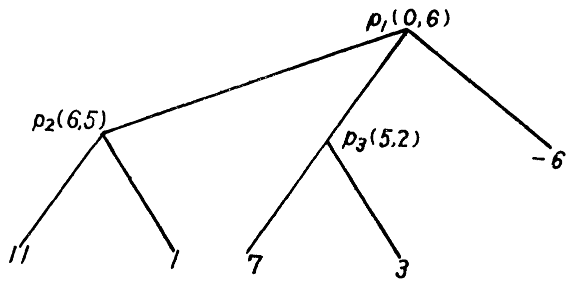



Figure 7

7. However, playing a mean strategy he will handle the two cases in different way. In Figure 6 $A_1$ plays in a period with $\tau = 1$. He will therefore make a 1-optimal move, the one into $p_3$. In Figure 7 he just starts a period with $\tau = 6$ and moves into $p_2$.

This difference may be explained thus. The move recommended by a mean strategy shall be a good move when played in the sum of $n$ copies of the game. We see readily that in $n$ copies of the game in Figure 6 the move into $p_3$ is the correct answer to $A_2$'s move into $p_1$. If $A_1$ always moves into $p_2$ he gets only about $5\frac{n}{2} + 1\frac{n}{2} = 3n$ though $m_1(p) = 4$. In $n$ copies of the game in Figure 7 the move into $p_2$ is correct. If $A_1$ always moves into $p_3$ he gets about $(-6)\frac{n}{2} + 7\frac{n}{4} + 3\frac{n}{4} = -\frac{n}{2}$ though $m_1(p_1) = 0$.

In a sense (4) of Theorem 5 means that the value of making a move is equal to the number $\tau$ for the period containing the move, where $\tau$ is max $\{\sigma(p_r)\}$ at the beginning of the period. One may try to change the rules for a mean strategy by requiring each move to be played at the position $p$ where $\sigma(p)$ is highest. The following example shows, however, that such a play does not guarantee the mean value.

EXAMPLE 7. Consider the sum game given in Figure 8. Suppose that $A_1$ starts and plays in the left game and that $A_2$ answers in the right game. Then $\sigma(p) = 7$ in the left game and $\sigma(p) = 6$ in the right

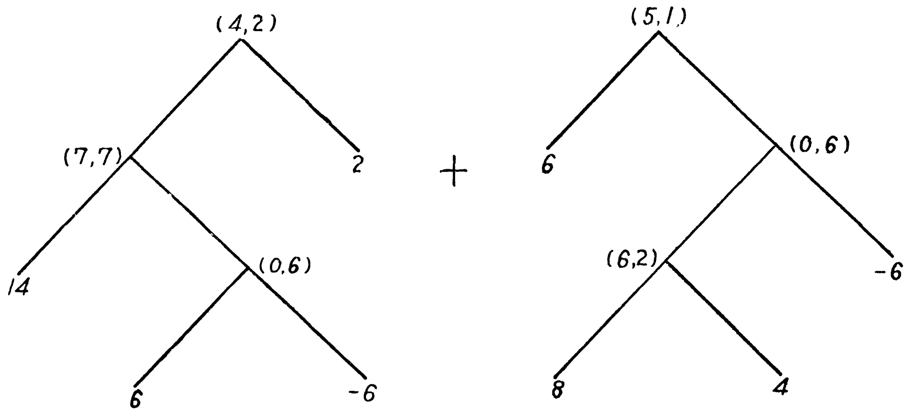



Figure 8

game. But if $A_1$ plays in the left game, where $\sigma(p)$ is highest he will get only $14 + (-6) = 8$ which is less than the mean value $4 + 5 = 9$. In fact $A_1$'s second move is made in a period with $\tau = 2$. Hence if he follows (a)—(d) he shall play a 2-optimal move in the game where the other player has just played, i.e. he shall play in the right game. Then he will get at least $6 + 4 = 10$ which is more than the mean value 9.

Let us in a final example show that an optimal move in a sum game need not be $t$-optimal for any $t$ in the summand $G$ in which it is made. Hence this move can never be recommended by a mean strategy.

EXAMPLE 8. The optimal move for $A_1$ in the sum game in Figure 9 is the move into $p_1$ in the left game, the mean strategy move is the move in the right game. The move into $p_1$ is never $t$-optimal in the left game for any $t$.

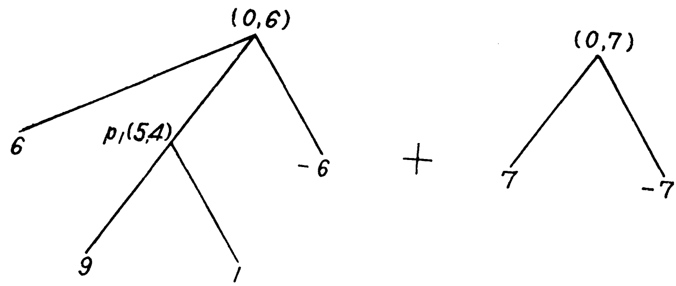



Figure 9

## REFERENCES

1. John Milnor, *Sums of positional games,* Contributions to the theory of games, Vol. **II**, p. 291-301, Princeton, 1953.
2. Pólya-Szegö, *Aufgaben und Lehrsätze aus der Analysis,* Bd. **1**, Berlin, 1925.

UNIVERSITY OF STOCKHOLM
SWEDEN

# ON ONE-TO-ONE HARMONIC MAPPINGS

ERHARD HEINZ

In this paper we shall prove the following:

THEOREM. *Let* $z = z(w)$ $(z = x + iy,\ w = u + iv)$ *be a one-to-one harmonic mapping of the disc* $|w| < 1$ *onto the disc* $|z| < 1$ *such that* $z(0) = 0$. *Then we have for* $|w| < 1$ *the estimate*

$$(1)\qquad |z_u|^2 + |z_v|^2 \geqq \frac{2}{\pi^2}\ .$$

As an improvement of an earlier result established in [1] J. C. C. Nitsche [4] showed that under the above conditions the inequality

$$(2)\qquad (|z_u|^2 + |z_v|^2)_{w=0} \geqq \frac{1}{2}$$

is satisfied[1]. In contrast to (2) the estimate (1) holds throughout the unit disc $|w| < 1$, but the constant involved is smaller than that of Nitsche.

In order to establish (1) we shall make use of a known result on harmonic functions (the analogue of the Schwarz Lemma)[2]. For the sake of completeness the proof of it will be given here.

LEMMA. *Let* $z = z(w) = x(w) + iy(w)$ *be a complex-valued harmonic function in the disc* $|w| < 1$. *Furthermore, let* $z(0) = 0$ *and* $|z(w)| < 1$ *for* $|w| < 1$. *Then we have the inequality*

$$(3)\qquad |z(w)| \leqq \frac{4}{\pi} \operatorname{arc\,tan} |w| \qquad |w| < 1.$$

*Proof*. Let $\theta$ be an arbitrary real number, and $f(w)$ be the function, which is regular-analytic in the disc $|w| < 1$ and satisfies the relations $f(0) = 0$ and

$$(4)\qquad \Re f(w) = x(w)\cos\theta + y(w)\sin\theta\ .$$

On account of our hypotheses we have

$$(5)\qquad |\Re f(w)| < 1 \qquad |w| < 1,$$

hence,

[1] For further references see [2].

[2] See Polya-Szegö [5], p. 140.

Received October 13, 1958.

$$(6)\qquad \Re\left(\exp\left[\frac{i\pi}{2}f(w)\right]\right) > 0 \qquad |w| < 1.$$

Consequently the function

$$(7)\qquad g(w) = \frac{\exp\left[\frac{i\pi}{2}f(w)\right] - 1}{\exp\left[\frac{i\pi}{2}f(w)\right] + 1}$$

satisfies the inequality

$$(8)\qquad |g(w)| < 1 \qquad |w| < 1,$$

and we have $g(0) = 0$. Applying now the Schwarz Lemma and the elementary inequality

$$(9)\qquad \left|\frac{e^{i\zeta} - 1}{e^{i\zeta} + 1}\right| \geqq \tan\frac{1}{2}|\Re\zeta| \qquad |\Re\zeta| \leqq \frac{\pi}{2}$$

we obtain the estimate

$$(10)\qquad \tan\frac{\pi}{4}|\Re f(w)| \leqq |g(w)| \leqq |w| ,$$

hence, by (4)

$$(11)\qquad |x(w)\cos\theta + y(w)\sin\theta| \leqq \frac{4}{\pi}\arctan|w|$$

for $|w| < 1$.

Since this holds for every real value of $\theta$ the inequality (3) follows, which proves the lemma.

*Proof of the theorem.* (I) We first prove (1) under the additional hypothesis that the function $z(w)$ and its first derivatives are continuous in the closed disc $|w| \leqq 1$. Since the mapping $w \to z(w)$ is one-to-one and harmonic, its Jacobian $|z_w|^2 - |z_{\bar{w}}|^2$ [3] cannot vanish, in virtue of a theorem of H. Lewy [3]. Furthermore, since hypothesis and conclusion of our theorem remain unchanged, if $z(w)$ is replaced by $\overline{z(w)}$, we may assume without loss of generality that

$$(12)\qquad |z_w|^2 - |z_{\bar{w}}|^2 > 0 \qquad |w| < 1.$$

Consequently, the function $z_w$ does not vanish in the disc $|w| < 1$. Furthermore, because of $z_{w\bar{w}} = 0$, it is regular-analytic. From these facts it follows that for $|w| \leqq 1$ the inequality

[3] Here and in the following considerations $\frac{\partial}{\partial w} = \frac{1}{2}\left(\frac{\partial}{\partial u} - i\frac{\partial}{\partial v}\right)$ and $\frac{\partial}{\partial \bar{w}} = \frac{1}{2}\left(\frac{\partial}{\partial u} + i\frac{\partial}{\partial v}\right)$ are the complex derivatives.

(13) $$|z_w| \geqq \min_{|w|=1} |z_w|$$

holds.

We shall now estimate the right-hand side of (13) from below by using our lemma. Let $\varphi$ and $r$ be two real numbers and $0<r<1$. Since by hypothesis the equation $|z(w)|=1$ holds for $|w|=1$ we have

(14) $$\left|\frac{z(e^{i\varphi})-z(re^{i\varphi})}{1-r}\right| \geqq \frac{1-|z(re^{i\varphi})}{1-r} \geqq \frac{1-4/\pi \operatorname{arc}\tan r}{1-r}$$

If we let $r$ tend to 1, we obtain

(15) $$\left(\left|\frac{\partial z(re^{i\varphi})}{\partial r}\right|\right)_{r=1} \geqq \frac{2}{\pi} \qquad 0 \leqq \varphi < 2\pi .$$

Furthermore, on account of (12) we have

(16) $$\left|\frac{\partial z(re^{i\varphi})}{\partial r}\right| = |z_w(re^{i\varphi})e^{i\varphi} + z_{\bar{w}}(re^{i\varphi})e^{-i\varphi}| \leqq |z_w| + |z_{\bar{w}}| \leqq 2|z_w|$$

for $0<r\leqq 1$. Combining this with (15) we infer that for $|w|=1$ the estimate

(17) $$|z_w| \geqq \frac{1}{\pi}$$

holds.

Hence, by (13) we obtain for $|w|\leqq 1$ the inequality

(18) $$\frac{1}{\pi} \leqq |z_w| = \frac{1}{2}|z_u - iz_v| \leqq 2^{-1/2}(|z_u|^2 + |z_v|^2)^{1/2} ,$$

which yields (1).

(II) Now let the mapping $z=z(w)$ merely satisfy the hypotheses of our theorem. Obviously there exists a sequence of numbers $\{R_n\}$ ($n\geqq 2$) such that the following conditions are satisfied:

(i) We have $0<R_n<1$ for all $n\geqq 2$, and

(19) $$\lim_{n\to\infty} R_n = 1 .$$

(ii) The disc $|z|<R_n$ is mapped by the inverse transformation $z\to w$ onto a simply-connected domain $D_n$ such that

(20) $$\left\{|w| \leqq 1-\frac{1}{n}\right\} \subset D_n \subset \{|w|<1\} .$$

Since the mapping $z\to w$ is analytic in $x$ and $y$, it follows that $D_n$ is bounded by an analytic Jordan curve. By the Riemann mapping theorem there exists a uniquely determined function $w=\Phi_n(\zeta)$, which maps the disc $|\zeta|<1$ ($\zeta=\xi+i\eta$) conformally onto $D_n$ such that $\Phi_n(0)=0$ and $\Phi_n'(0)>0$. Furthermore, $\Phi_n(\zeta)$ is analytic for $|\zeta|\leqq 1$. Consequently, the function

$$Z(\zeta) = \frac{z(\Phi_n(\zeta))}{R_n} \tag{21}$$

is harmonic for $|\zeta| < 1 + \delta$, where $\delta$ is a positive number, and satisfies all the hypotheses of the above theorem. From the facts established in (I) we conclude

$$\frac{|\Phi_n'(\zeta)|^2}{R_n^2}(|z_u|^2 + |z_v|^2) = |Z_\xi|^2 + |Z_\eta|^2 \geqq \frac{2}{\pi^2} . \tag{22}$$

Hence we have for $w = \Phi_n(\zeta)$ $(|\zeta| < 1)$ the inequality

$$|z_u|^2 + |z_v|^2 \geqq \frac{R_n^2}{|\Phi_n'(\zeta)|^2} \cdot \frac{2}{\pi^2} . \tag{23}$$

Furthermore, on account of (20) the estimates

$$\left(1 - \frac{1}{n}\right)|\zeta| \leqq |\Phi_n(\zeta)| \leqq |\zeta| \tag{24}$$

hold for $n \geqq 2$ and $|\zeta| < 1$. Applying the Schwarz Lemma it follows from (24) that there exists a sequence of integers $\{n_k\}$ such that the relations

$$\Phi_{n_k}'(\zeta) \to 1 \qquad (k \to \infty) \tag{25}$$

hold uniformly in every closed disc $|\zeta| \leqq \rho < 1$.

Now let $w^*$ be a fixed complex number with $|w^*| < 1$ and let us determine two positive numbers $k_0$ and $\rho$ such that the inequalities

$$\frac{|w^*|}{1 - \frac{1}{n_k}} \leqq \rho < 1 \tag{26}$$

are satisfied for $k \geqq k_0$. On account of (20) the point $w^*$ belongs to $D_{n_k}$ for $k \geqq k_0$. Hence there exists a sequence of complex numbers $\{\zeta_k\}$ with $|\zeta_k| < 1$ such that the equations

$$w^* = \Phi_{n_k}(\zeta_k) \tag{27}$$

hold for $k \geqq k_0$. By (24) we have

$$|\zeta_k| \leqq \frac{|w^*|}{1 - \frac{1}{n_k}} \leqq \rho < 1 \tag{28}$$

for $k \geqq k_0$. Applying now (23) and (25) we conclude

$$(|z_u|^2 + |z_v|^2)_{w=w^*} \geqq \frac{R_{n_k}^2}{|\Phi_{n_k}'(\zeta_k)|^2} \cdot \frac{2}{\pi^2} \to \frac{2}{\pi^2} \tag{29}$$

for $k \to \infty$. Since $w^*$ is an arbitrary point in the disc $|w| < 1$, our theorem is established.

## REFERENCES

1. E. Heinz, *Uber die Lösungen der Minimalflächengleichung*, Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl. (1952), 51-56.
2. E. Heinz, *On certain nonlinear elliptic differential equations and univalent mappings*, J. d'Analyse Math. Jerusalem **5** (1956/57), 197-272.
3. H. Lewy, *On the non-vanishing of the Jacobian in certain one-to-one mappings*, Bull. Amer. Math. Soc. **42** (1936), 689-692.
4. J. C. C. Nitsche, *On harmonic mappings*, Proc. Amer. Math. Soc., **9** (1958), 268-271.
5. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, I, Berlin, 1925.

STANFORD UNIVERSITY

# ON FINITE-DIMENSIONAL UNIFORM SPACES

J. R. ISBELL

**Introduction.** This paper has two nearly independent parts, concerned respectively with extension of mappings and with dimension in uniform spaces. It is already known that the basic extension theorems of point set topology are valid in part, and only in part, for uniformly continuous functions. The principal contribution added here is an affirmative result to the effect that every finite-dimensional simplicial complex is a uniform ANR, or ANRU. The complex is supposed to carry the uniformity in which a mapping into it is uniformly continuous if and only if its barycentric coordinates are equiuniformly continuous. (This is a metric uniformity.) The conclusion (ANRU) means that whenever this space $\mu A$ is embedded in another uniform space $\mu X$ there exist a uniform neighborhood $U$ of $A$ (an $\varepsilon$-neighborhood with respect to some uniformly continuous pseudometric) and a uniformly continuous retraction $r : \mu U \rightarrow \mu A$.

It is known that the real line is not an ARU. (Definition obvious.) Our principal negative contribution here is the proof that no uniform space homeomorphic with the line is an ARU. This is also an indication of the power of the methods, another indication being provided by the failure to settle the corresponding question for the plane. An ARU has to be uniformly contractible, but it does not have to be uniformly locally an ANRU. (The counter-example is compact metric and is due to Borsuk [2]). It does have to be uniformly locally connected, which is enough to give us a grip on the real line.

Smirnov has defined the *$\delta$-dimension* $\delta d$ of a uniform space as the least dimension of a cofinal family of finite uniform coverings and has shown that $\delta d$ has many of the properties of topological dimension functions and some novel ones [9, 10]. The *large dimension* $\Delta d$ is defined in the same manner, using arbitrary uniform coverings, in [6], where it is shown that $\Delta d$ is $\geqq \delta d$ and is (like $\delta d$) invariant under completion. The central result of the second part of this paper is that when $\delta d(\mu X) = n$ there are precisely two possible values for $\Delta d(\mu X)$, namely $n$ and $\infty$.

Two applications are made, the principal one being a considerable simplification of the proof of the main theorem of [10] (characterization of the $n$-dimensional uniform subspaces of $E^n$). Also there are two side conditions either of which implies $\Delta d(\mu X) = \delta d(\mu X)$: every uniform covering has a finite-dimensional uniform refinement ($\mu X$) is, so to speak, weakly finite-dimensional), or $\mu X$ is *locally fine* in the sense of [5]. The

Received May 26, 1958.

first of these includes the case of a weak uniformity induced by a family of real-valued functions, and the second includes the case of the finest uniformity compatible with the topology.

**1. Extension.** Most of the simplicial complexes considered in this paper will be finite-dimensional, largely because we do not have a convenient uniform structure for infinite-dimensional complexes. In any simplicial complex $X$, the points $x$ are determined by their barycentric coordinates $(x_\alpha)$. The function $d(x, y) = \max |x_\alpha - y_\alpha|$ is a distance function inducing a uniformity and topology on $X$; with this uniformity, $X$ is called a *uniform complex*. We recall from [7] that a finite-dimensional uniform complex is a complete space, the stars of vertices from a unifom covering $\{St_\alpha\}$, a mapping into $X$ is uniformly continuous if and only if its coordinate projections are equiuniformly continuous; and every finite-dimensional uniform covering of a uniform space has an equiuniformly continuous partition of unity subordinated to it, which then induces a canonical mapping into the nerve.

In fact, we can show the following.

1.1. *Every uniform covering has an equiuniformly continuous partition of unity subordinated to it.*

Because of the difficulty with infinite-dimensional complexes, we shall not get any use out of 1.1 excepting a very special application in the second section of the paper.

*Proof of* 1.1. Let $\{U_\alpha\}$ be a uniform covering of a uniform space $\mu X$. Let $d$ be a uniformly continuous pseudometric on $\mu X$ such that every set of $d$-diameter 2 or less is contained in some $U_\alpha$ [7]. Well order the indices $\alpha$. For each $\alpha$, we define a real-valued function $g_\alpha : g_\alpha(x)$ is the smaller of the numbers 1 and $\sup [d(x, X - U_\beta) | \beta < \alpha]$. The functions $g_\alpha$ increase monotonically to the pointwise limit 1 (continuously at limit ordinals), and each of them is uniformly continuous with respect to $d$, with modulus of continuity $\delta(\varepsilon) = \varepsilon$; that is, $d(x, y) < \varepsilon$ implies $|g_\alpha(x) - g_\alpha(y)| < \varepsilon$. Therefore the functions $f_\alpha = g_{\alpha+1} - g_\alpha$ form an equiuniformly continuous partition of unity; and if $x \in X - U_\alpha$, then $g_{\alpha+1}(x) = g_\alpha(x)$ and $f_\alpha(x) = 0$.

A partition of unity $\{f_\alpha\}$ will not yield a function with values in the nerve unless at each point all but finitely many $f_\alpha$ vanish. An obvious sufficient condition for this is that $\{U_\alpha\}$ is point-finite. We can rearrange the statement of this condition by using the following (essentially known) construction. A covering $\{V_\alpha\}$ with the same indexing set as $\{U_\alpha\}$ is called a *shrinking* of $\{U_\alpha\}$ if $V_\alpha \subset U_\alpha$ for all $\alpha$; let us call $\{V_\alpha\}$ a *strict shrinking* if there exists a uniform covering $w$ such that $St(V_\alpha, w) \subset U_\alpha$ for all $\alpha$.

1.2. *Every uniform covering has a strict shrinking which is uniform. If a uniform covering $w$ is a star-refinement (or merely a refinement) of $u$, then $u$ can be strictly shrunk (or merely shrunk) to a uniform covering $v$ whose elements are unions of disjoint families of elements of $w$.*

*Proof.* Well order the elements $U_\alpha$ of $u$ and define $V_\alpha$ as the union of all elements of $w$ whose stars (or merely whose selves) are contained in $U_\alpha$ but not contained in any preceding $U_\beta$.

At the moment, we want the following applications of 1.2. Every point-finite uniform covering has a uniformly locally finite uniform refinement (any strict shrinking); and if a covering has a uniformly locally finite refinement then it has a uniformly locally finite shrinking.

Note specifically that the nerve of the covering $v$ constructed in 1.2 is a subcomplex of the nerve of $u$ and is the image of the nerve of $w$ under a simplicial mapping. The dimension of $v$ is no greater than the dimension of $w$.

In normal topological spaces 1.2 is valid for coverings having locally finite refinements, and leads to the conclusion that such a covering admits a canonical mapping into the nerve for any reasonable topology on the nerve [3]; for continuity is easily deduced from the fact that a neighborhood of each point is mapped into a finite complex. A uniform space is called *locally fine* if every uniformly locally uniform covering is uniform [5]; it follows that every uniformly locally uniformly continuous function is uniformly continuous, and also that every uniform covering has a uniformly locally finite uniform refinement. Accordingly we have

1.3. *Relative to any uniformity for simplicial complexes which makes every finite subcomplex a uniform complex, the following is true: Corresponding to every uniform covering of a locally fine uniform space there is a canonical mapping into the nerve.*

We shall want to apply this with some structure on the nerve of $\{U_\alpha\}$ making $\{St_\alpha\}$ a uniform covering. It will suffice to use *uniform Whitehead complexes* (UW-complexes) defined as CW-complexes bearing the finest uniformity compatible with the topology.

We obtain also, from 1.1 and 1.2, certain mappings of any uniform space into the nerve (regarded as a uniform complex) of any covering having a point-finite uniform refinement. The mappings take $U_\alpha$ into $St_\alpha$, but when $\{St_\alpha\}$ is not uniform, this is of little value.

1.4. LEMMA. *Every bounded uniformly continuous pseudometric on a subspace of a uniform space many be extended to a bounded uniformly continuous pseudometric on the whole space.*

*Proof.* Suppose $e$ is a bounded pseudometric on $\mu A \subset \mu X$. We show first that there is a pseudometric $d$ on $\mu X$ satisfying $d(x, y) > e(x, y)$ for all $x$ and $y$ in $A$. For each integer $n$ (positive or negative) there is a uniform covering $u^n$ of $\mu X$ such that if $x$ and $y$ are in $A$ and in a common element of $u^n$ then $e(x, y) < 2^n$; and there is a pseudometric $d_n$ on $\mu X$ such that $d_n(x, y) < 2^{n+1}$ implies $x$ and $y$ are in a common element of $u^n$, but $d_n(x, y) \leq 2^{n+1}$ for all $x$ and $y$. If $2^k$ is a bound for values of $e$, then $\sum(d_n | k \geq n > -\infty]$ is a pseudometric $d$ uniformly continuous on $\mu X$, and $d(x, y) < 2^{n+1}$ implies $e(x, y) < 2^n$, so that $d > e$ in $A$. Finally define $m$ on $X$ by $m(x, y) = \min(d(x, y), \inf[d(x, a) + e(a, b) + d(b, y) | a$ and $b$ in $A])$. Examination of cases shows that $m$ is a pseudometric. Since $m \leq d$, $m$ is uniformly continuous. Then $m$ is the required extension of $e$.

1.5. COROLLARY. *For every uniformly continuous mapping $f$ of a subspace $\mu A$ of a uniform space $\mu X$ into a metric space $\nu B$, there exist a metric space $\nu Y$ containing $\nu B$ and a mapping $g$: $\mu X \to \nu Y$ extending $f$.*

For every metric is equivalent to a bounded one.

1.6. COROLLARY. *For every uniformly continuous mapping $f$ of a subspace $\mu A$ of a uniform space $\mu X$ into a uniform space $\nu B$, there exist a uniform space $\nu Y$ containing $\nu B$ and a mapping $g$: $\mu X \to \nu Y$ extending $f$.*

For every uniform space is a subspace of a product of metric spaces.

The definitions of absolute retract and absolute neighborhood retract write themselves, except that one must notice that uniform neighborhoods should be specified. If $A \subset U \subset \mu X$, $U$ is a *uniform neighborhood* of $A$ provided $U$ contains the star of $A$ with respect to some uniform covering. $\mu X$ is an ARU provided $\mu X \subset \mu Y$ implies the existence of a uniformly continuous retraction $\mu Y \to \mu X$; $\mu X$ is an ANRU provided $\mu X \subset \mu Y$ implies the existence of a uniformly continuous retraction onto $\mu X$ of some uniform neighborhood of $X$ in $\mu Y$. One point to be noticed is

1.7. *Every* ANRU *is complete; moreover, an incomplete space can be embedded as a closed subspace of a space in which there is no retraction of a uniform neighborhood.*

*Proof.* Given any incomplete space $\mu A$, let $\pi\mu A$ be its completion and $I$ a well-ordered space, in the order topology, having a last element $\omega$ such that every sequence converging to $\omega$ has a greater cardinal number than that of $A$. Embed $\mu A$ in $\pi\mu A \times I$, $\alpha \in A$ going to $(\alpha, \omega)$, and remove from the product the points $(x, \omega)$, $x$ not in $A$; evidently $\mu A$ becomes a closed subspace on which there is no continuous retraction of a neighborhood.

Among topological spaces one distinguishes between the property of being an absolute retract (or an ANR) and the stronger property of being an absolute (neighborhood) *extensor*: $Y$ is an absolute extensor for a class of spaces if when $A$ is a closed subspace of a space $X$ in this class, every continuous mapping of $A$ into $Y$ can be extended over $X$. From 1.4 and its corollaries we have the following: *Among uniform spaces, every absolute retract is an absolute extensor, i.e. every uniformly continuous mapping of a subspace of any space $\mu X$ into an* ARU *may be extended.* Similarly for ANRU's. Further, *for a metric space to be an* ARU *or* ANRU, *it suffices that it is a retract or neighborhood retract whenever it is embedded in a metric space.* Moreover, if we choose any convenient bounded distance function, we may assume the embedding is an isometry.

The reduction to the isometric case simplifies matters considerably, but there still remains some computation. We shall have to consider moduli of continuity explicitly. Recall that a *modulus*, in this context, is any function on the positive reals to the positive reals.

1.8. LEMMA. *For every modulus of continuity $\delta$ and every natural number $n$, there exists a modulus $\lambda$ such that every mapping of a subspace of a metric space into a uniform $n$-simplex, having the modulus $\delta$, can be extended to a mapping of the whole space into the simplex having modulus $\lambda$.*

*Proof.* The $n$-simplex $T$ is an ARU because it is uniformly equivalent to a product of intervals, each of which is an ARU by Katetov's extension theorem [8]. Now suppose the lemma is false, i.e. there exist $\delta$ and $n$ such that for each modulus $\lambda$ there exist metric spaces $A_\lambda \subset X_\lambda$ (with distance $d_\lambda$) and mappings $f_\lambda : A_\lambda \to T$, each having modulus of continuity $\delta$ but having no extension over $X_\lambda$ with modulus $\lambda$. Let $X$ be the union of disjoint copies of all $X_\lambda$, with the following distance function $d$ : $d(x, y) = 1$ unless $x$ and $y$ are in the same $X_\lambda$ and $d_\lambda(x, y) < 1$, in which case $d(x, y) = d_\lambda(x, y)$. Then with $A = \cup A_\lambda$, $f : A \to T$ defined by the values of the $f_\lambda$, $f$ is uniformly continuous with modulus of continuity $\min(\delta, 1)$. Therefore $f$ has an extension over $X$ which has a modulus of continuity $\lambda$. But $\min(\lambda, 1)$ is a modulus of continuity $\mu$, and the restriction of $f$ to $X_\mu$ has modulus $\mu$, a contradiction.

1.9. THEOREM. *Every finite-dimensional uniform complex is an ANRU.*

*Proof.* Suppose the $n$-dimensional complex $N$ is isometrically embedded in $X$ with distance $d$; on $N$, $d(x, y) = \max |x_\alpha - y_\alpha|$. We shall need the product of all the odd numbers up to $2n + 1$; for typographical convenience we take $(2n + 1)!$ Then let $N_k$ denote the $k$-skeleton of $N$, and $C_k$ the set of all $x$ in $X$ which satisfy

$$d(x, N_k) \leq \frac{(2n+1)!}{(2k+1)!} d(x, N) \quad \text{and} \quad d(x, N_k) \leq \frac{1}{(2k+3)!} .$$

Any $x$ in $C_0$ is within distance 1/6 of a unique vertex, which we define to be $f_0(x)$; thus $f_0 : C_0 \to N_0$ is a retraction with modulus of continuity $\delta \equiv 2/3$. Suppose the retraction $f_k : C_k \to N_k$ has been defined and has a definite modulus of continuity $\delta_k$. Now if $x$ is within distance $\theta$ of points $p$, $q$, in different $(k + 1)$-simplexes, $\sigma$, $\tau$, of $N$, then $p$ and $q$ are within $2\theta$ of each other. If $(2k + 4)\theta < 1$, define barycentric coordinates of a point $r$ by deleting those non-zero coordinates $p_\alpha$ of $p$ for which $q_\alpha = 0$ (whose sum is at most $(k + 1)2\theta$) and increasing accordingly one of the non-zero coordinates of $p$ which must be left. Then $r$ is common to $\sigma$ and $\tau$ and hence is in $N_k$; also, $d(x, r) \leq (2k + 3)\theta$. Thus if $x$ is in $C_{k+1}$ and not in $C_k$, then there is a unique $(k + 1)$-simplex $\sigma$ such that

$$d(x, \sigma) \leq \frac{(2n+1)!}{(2k+3)!} d(x, N) \quad \text{and} \quad d(x, \sigma) \leq \frac{1}{(2k+5)!} .$$

Let $C(\sigma)$ denote the union of the set of all such $x$ and $f_k^{-1}(\sigma)$; let $A(\sigma)$ denote $(C_k \cap C(\sigma)) \cup \sigma$; and define a retraction $\varphi : A(\sigma) \to \sigma$ to agree with $f_k$ on $C_k \cap C(\sigma)$. For pairs of points in $C_k$, $\varphi$ has modulus of continuity $\delta_k$; for pairs in $\sigma$, the identity function $\delta(\varepsilon) = \varepsilon$; and for $p$ in $C_k$, $q$ in $\sigma - C_k = \sigma^0$, we have

$$d(p, N_k) \leq \frac{(2n+1)!}{(2k+1)!} d(p, q) ,$$

which yields a point of $N_k$ near both $p$ and $q$ and establishes a modulus of continuity here also. By 1.8, $f_k$ can be extended over each $C(\sigma)$ separately, as a retraction with a definite modulus of continuity $\lambda$. Then $\delta_{k+1} = \lambda/(6k + 12)$ is a modulus of continuity for the whole mapping $f_{k+1}$, since two points at distance $\theta$ from each other in different $C(\sigma)$, $C(\tau)$, are within $(3k + 6)\theta$ of a point $r$ of $\sigma \cap \tau$, as above. Therefore the induction runs, and $f_n$ becomes defined on the entire $1/(2n + 3)!$ neighborhood of $N$.

It should be noted that the theorem as stated is trivially false for arbitrary uniform complexes, since some of them are incomplete. It is false for many complete ones also. It seems likely that strong results might be gotten by using some suitable uniformity for a complex, different from the one defined by $\max|x_\alpha - y_\alpha|$, though not necessarily different for finite-dimensional complexes. UW-complexes are different in the finite-dimensional case, and I do not know whether they satisfy 1.9.

One gets the homotopy extension lemma and the theorem ARU $\equiv$ uniformly contractible ANRU just as in the topological case. Precisely, if $I$ is the interval $[0, 1]$, the *cylinder* over $\mu X$ is the product $\mu X \times I$, and the *cone* over $\mu X$ is the quotient space of the cylinder obtained by collapsing $\{(x, 1)\}$ to a point. Homotopy and related concepts being defined as usual, we have

1.10. *If* $\mu A \subset \mu X$, $\nu Y$ *is an* ANRU, $h$: $\mu A \times I \to \nu Y$ *is a homotopy between* $h_0$ *and* $h_1$, *and* $g_0$ *is an extension of* $h_0$ *over* $\mu X$, *then* $h$ *can be extended to a homotopy of* $g_0$.

*Proof.* Define $f$ on $(\mu A \times I) \cup \{(x, 0)\}$ by $f(a, t) = h(a, t)$, $f(x, 0) = g_0(x)$. Let $f'$ be an extension of $f$ over a uniform neighborhood $U$, and let $p$ be a uniformly continuous real-valued function on $\mu X \times I$ vanishing outside $U$ and equal to 1 on the domain of $f$. Then $g(x, t) = f'(x, tp(x, t))$ defines the required extension.

1.11. *A uniform spaces is an* ARU *if and only if it is a uniformly contractible* ANRU.

*Proof.* An ARU is an ANRU by definition, and retraction of the cone over it defines a uniform contraction. Conversely, every mapping into a uniformly contractible ANRU is homotopic to a constant and therefore extensible.

1.12. *The cone over an* ANRU *is an* ARU.

*Proof.* Let $\nu Y$ be an ANRU, $C$ the cone over $\nu Y$, $\mu A$ a subspace of $\mu X$, and $f$: $\mu A \to C$ a uniformly continuous function. The construction of $C$ as a quotient space of $\nu Y \times I$ gives each point of $C$ a second coordinate in $I$, and each point except the vertex $v$ has a first coordinate in $Y$. Let $f_2$: $\mu A \to I$ be the second coordinate of $f$, which is uniformly continuous. Let $f_1$: $A - f_2^{-1}(1) \to Y$ be the first coordinate of $f$, and note that $f_1$ is uniformly continuous on each of the sets $A_n = f_2^{-1}[0, 1 - 2^{-n}]$. Let $g_2$: $\mu X \to I$ be a uniformly continuous extension of $f_2$. Let $g_1$ be a function with values in $Y$, defined on a subset $U$ of $X$, such that for each $n$, $U$ contains a uniform neighborhood of $A_n$ on which $g_1$ is a uniformly continuous extension of $f_1$. (The construction of $g_1$ requires a little care. Define $B_1 = A_2$, $j_1 = g_1 | B_1$. Inductively let $h_n$ be an extension of $j_n$ over a uniform neighborhood of $B_n$, and $i_n$ the restriction of $h_n$ to a uniform neighborhood $U_n$ of $B_n$ which is contained in $g_2^{-1}[0, 1 - 2^{-n-1}]$; let $B_{n+1} = A_{n+2} \cup U_n$, define $j_{n+1}: B_{n+1} \to Y$ by the values of $i_n$ and $g_1$, and continue.) If $U_n$ is such a neighborhood of $A_n$, there exist uniformly continuous pseudometrics $d_n$ on $\mu X$ relative to which $U_n$ is an $\varepsilon_n$-neighborhood of $A_n$; we may assume $d_n$ is bounded by 1 and form $d = \sum 2^{-n} d_n$, so that relative to $d$, $U_n$ is a $\delta_n$-neighborhood of $A_n$, for a sequence of positive numbers $\delta_n$. Let $g_0$ be a monotone decreasing continuous function on $I$ to $I$, vanishing only at 1, but satisfying $g_0(1 - 2^{-n}) < \delta_{n+1}$. Now define $g_2'$ on $\mu X$ to $I$ as follows. For $p$ in $A$, we have $g_2'(p) = f_2(p)$. If $d(p, A) > g_0(g_2(p))$ then $g_2'(p) = 1$. For all other $p$ we have

$$g_2'(p) = g_2(p) + \frac{d(p, A)}{g_0(g_2(p))}(1 - g_2(p)) .$$

One readily verifies that $g_2'$ is uniformly continuous. Since $g_2'$ takes the constant value 1 on the complement of $U$, we may define $g : \mu X \to C$ by $g(x) = (g_1(x), g_2'(x))$ where $g_2'(x) \neq 1$, $g(x) = v$ where $g_2'(x) = 1$. Then $g$ is a uniformly continuous extension of $f$.

1.12 shows that many ARU's exist. Also, a product of arbitrarily many ARU's is clearly an ARU. On the other hand, the product of a sequence of copies of the real line is not an ANRU; it is not a retract of any uniform neighborhood in the product of cones over the lines.

In the other direction, we have the following.

1.13. *There is no* ARU *homeomorphic with the real line.*

*Proof.* An ARU, and even an ANRU, must be uniformly locally connected; for it can be embedded in a product of metric spaces, and thus in a product of Banach spaces, where retraction of a uniform neighborhood establishes the assertion. Now since the only connected subsets of the line are intervals, a uniformly locally connected structure on the line is either incomplete or uniformly locally compact. In view of 1.7 and 1.11, the proof will be completed when we establish

1.14. *Every uniformly locally compact uniformly contractible space is compact.*

In turn 1.14 will be deduced from

1.15. *Every uniformly locally compact space has a basis of star-finite uniform coverings.*

*Proof of* 1.15. Observe that a uniformly locally compact space has a uniform covering $u$ such that the closures of the stars of elements of $u$ are compact; and the same is true for any refinement of $u$. There is a uniform refinement $v$ which has the property that no proper subfamily of $v$ is a covering of the space. To see this, take a pseudometric $d$ so that every set of $d$-diameter 2 or less is a subset of an element of $u$; choose a maximal family of points $p_\alpha$ with mutual distances $\geqq 1/2$; and define $V_\alpha$ as the set of all points within distance 1 of $p_\alpha$ *except* the other $p_\beta$. Now each spherical neighborhood of radius 1/4 is contained in one of the $V_\alpha$. If $x$ is a point within distance 1/4 of some $p_\alpha$, $V_\alpha$ contains the 1/4- neighborhood of $x$. For any other $x$, there is some $p_\alpha$ within distance 1/2 of $x$, and since the 1/4- sphere about $x$ contains no other $p_\beta$ by hopothesis, it is a subset of $V_\alpha$. Finally, the covering $\{V_\alpha\}$ must be star-finite since the closures of the stars are compact.

*Proof of* 1.14. Every uniformly contractible space is *finitely chainable* in the sense of [1]; that is, for any uniform covering $\{U_\alpha\}$ there exist finitely many indices $\alpha_1, \cdots, \alpha_n$ and a natural number $m$ such that every $U_\alpha$ can be joined to one of the $U_{\alpha_i}$ by a chain of $m$ or fewer intersecting sets $U_\beta$. (In fact, we may take $n = 1$.) If the covering is also star-finite, it is finite. Since a uniformly locally compact space is always complete, we have 1.14, and with it, 1.13.

It is not true that every ANRU is uniformly locally an ARU; at least, not in the sense that there are arbitrarily fine uniform coverings consisting of ARU's. The trouble is that a subspace which is an ARU must be closed, by 1.7. But Borsuk has exhibited [2] a compact 2-dimensional AR in $E^2$ in which no closed 2-dimensional proper subset is an ANR. (For compact spaces, AR $\equiv$ ARU and ANR $\equiv$ ANRU, since these spaces can be embedded in cubes.)

The converse is not true either. If $S^n$ denotes the finite complex which is the boundary of an $n$-simplex, then the uniform complex which is the separated sum of all $S^n$ is uniformly locally an ANRU but not itself an ANRU. With a little more care the same effect can be demonstrated with a metric space which is a discrete sum of ARU's (e.g. arcs $I_n$ in $S^n$ coming within distance $1/n$ of every point of the sphere).

An ARU considered as a topological space is an absolute extensor for paracompact spaces; and similarly for ANRU's. 1.13 seems to rule out any reasonable converse to the first part of this remark.

**2. Dimension.** The *δ-dimension* or *uniform dimension* $\delta d(\mu X)$ of a uniform space $\mu X$ is defined as the least $n$ such that every finite uniform covering has a (finite) uniform refinement whose nerve is $n$-dimensional; if there is no such $n$, we write $\delta d(\mu X) = \infty$. (In view of 1.2, it does not matter whether the parenthesis "(finite)" is included or not. If it is included, we have the original definition of Smirnov except for irrelevant changes in the concept of equivalence of two spaces—Smirnov speaks of proximity spaces—and we may quote his results [9, 10] freely.) Similarly the *large dimension* $\Delta d(\mu X)$ is the least $n$ such that every uniform covering has an $n$-dimensional uniform refinement [6]. The inequality $\delta d(\mu X) \leqq \Delta d(\mu X)$ is a trivial consequence of 1.2.

2.1. *A finite-dimensional uniform complex is uniformly equivalent to its first barycentric subdivision, by the identity mapping. The stars of vertices in successive barycentric subdivisions form a basis of uniform coverings.*

*Proof.* The first statement is a consequence of the second. For that, it is well known that the meshes of these coverings approach zero, and it remains to show that each is uniform. It is certainly uniform on a uniform neighborhood of the $O$-skeleton ; and the proof may be finished by an induction using the fact that the $(k-1)$-skeleton separates all the $k$-simplexes from each other.

2.2 LEMMA. *A uniform covering has a finite-dimensional uniform refinement if and only if it has a uniform refinement which is a union of finitely many uniformly discrete subcollections.*

*Proof.* Evidently if a covering $u$ is a union of $n$ subcollections which are uniformly discrete (or even merely collections of disjoint sets) then $u$ has dimension at most $n-1$. For the converse, consider the nerve $N(u)$ as a uniform complex, and let $f$ be a canonical mapping of the space $\mu X$ into $N(u)$. The stars of vertices in the first barycentric subdivision of $N(u)$ form a uniform covering $w$ which is a union of $n+1$ collections of disjoint sets, namely the collections of stars of vertices which are centroids of $i$-dimensional simplexes of $N(u)$, for $i = 0, \cdots, n$. If $w'$ is a uniform strict shrinking of $w$, then $w'$ is a union of $n+1$ uniformly discrete subcollections, and the same is true of $f^{-1}(w')$, which is a uniform refinement of $u$.

2.3. THEOREM. *If $\delta d(\mu X) = n$ then either $\Delta d(\mu X) = n$ or $\mu X$ has a uniform covering which has no finite-dimensional uniform refinement.*

*Proof.* We have observed already that $\Delta d(\mu X) \geqq n$. It remains to show that every finite-dimensional uniform covering $u$ has an $n$-dimensional uniform refinement. By 2.2 we may suppose $u$ is the union of uniformly discrete collections $u^0, \cdots, u^p$.

Let $U_i$ be the union of $u^i$. Then $u$ is a refinement of $\}U_i\}$, which is thus a finite uniform covering. By hypothesis $\{U_i\}$ has an $n$-dimensional uniform refinement, and therefore (by 1.2) it has an $n$-dimensional uniform shrinking $\{V_i\}$. Let $v^i$ be the restriction of $u^i$ to the subspace $V_i$. Each $v^i$ is a disjoint collection covering $V_i$; hence their union $v$ is an $n$-dimensional covering of the space which is finer than $u$. To show that $v$ is uniform it suffices to show that each $v^i$ is uniform on $V_i$ (since $\{V_i\}$ is uniform and finite). But with respect to some uniformly continuous pseudometric, the different elements of $u^i$ are at mutual distances $> \varepsilon$, and then in the subspace $V_i$ each element of $v^i$ is an $\varepsilon$-neigborhood of itself.

2.4. EXAMPLE. There are uniform spaces $\mu X$ for which $\Delta d(\mu X) = \infty$ and $\delta d(\mu X)$ has any desired value. Here is an example homeomorphic with a countable discrete space, and having a basis of star-finite uniform coverings. For the description of the structure of $X$, consider the metric space $K$ which is a union of cells $I^n$, each isometric to the unit ball in $E^n$, with the distance between two points is different cells defined to be 1. Identify the countable set $X$ with a countable dense subset of $K$, for the purpose of stating: a covering of $X$ is to be uniform on $\mu X$ provided it has a refinement of the form $\{U_{i\alpha}\}$, where the sets $X_i = U_\alpha U_{i\alpha}$ are finite in number, and on each $X_i$, considered as a subspace of $K$, $\{U_{i\alpha}\}$ is a uniform covering. One easily sees that this family of coverings satisfies Tukey's axioms and thus defines a uniformity; the associated topology is discrete, since some of the sets $X_i$ may be single points (lying in no other $X_j$).

Every finite covering of $X$ is uniform and may be refined by a finite partition, so that $\delta d(\mu X) = 0$. On the other hand, if $\{V_\alpha\}$ is a uniform covering of $K$ which has no finite-dimensional uniform refinement (e. g. the covering consisting of all sets of diameter $< 1$), then $\{V_\alpha \cap X\}$ is a uniform covering of $\mu X$. If it had a finite-dimensional uniform refinement, we should have $X$ partitioned into sets $X_1, \cdots, X_n$, each $X_i$ covered by a finite-dimensional covering $\{U_{i\beta}\}$ which is uniform on $X_i$ considered as a subspace of $K$. Using 2.2, we may as well assume each $\{U_{i\beta}\}$ is uniformly discrete. Moreover, we may assume $\{U_{i\beta}\}$ is a strict shrinking of $\{V_\alpha \cap X\}$, so that for some $\varepsilon > 0$ the $\varepsilon$-neighborhood of each $U_{i\beta}$ is contained in some $V_\alpha$. If $\varepsilon$ is small enough, any two sets $U_{i\beta}, U_{i\gamma}$, are $3\varepsilon$ apart. Then the $\varepsilon$-neighborhoods of the sets $U_{i\beta}$ form a finite-dimensional refinement of $\{V_\alpha\}$ and a uniform covering of the $\varepsilon/2$-neighborhood of a dense subset of $K$, that is, a uniform covering of $K$: a contradiction.

Let us record the last construction for later use.

2.5. *If $u$ is a uniform covering of $\mu X$ and $\{V_{i\beta}\}$ a uniform covering of $\mu A \subset \mu X$ which is a union of $n$ uniformly discrete families, and a strict* *shrinking of* $u$, *then there is a uniform covering* $\{W_{i\alpha}\}$ *of a uniform neighborhood of* $A$ *having these two properties and satisfying* $W_{i\alpha} \cap A = V_{i\alpha}$.

The proof is just as above except that one must introduce a suitable pseudometric.

Dowker's proof [3] that the same covering dimension is obtained for a normal topological space from its finite, star-finite, or locally finite open coverings goes by way of mappings into spheres and depends on (a) constructing canonical mappings and (b) modifying them to be essential onto each simplex. In uniform spaces, of course, the construction is impossible, since the conclusion is false. This need not mean that useful canonical mappings cannot be constructed. The other part of the construction, the removal of inessentiality, is definitely impossible, even in finite-dimensional spaces. This is easily verified for the subspace of the plane consisting of the vertical line segments $x = n$, $-1 \leqq y \leqq 0$, and $x=n, 1/n \leqq y \leqq 1$. In a sense, the construction of 1.15 yielding arbitrarily fine uniform coverings which have no proper subcoverings, is the best one can do in general.

In the case of locally fine uniform spaces, Dowker's argument goes through step by step. One has canonical mappings by 1.3 ; every uniform covering has a uniformly locally finite uniform shrinking; and modifications preserving uniform continuity uniformly locally preserve it in the large. Thus we have

2.6. *For locally fine uniform spaces, $\delta$-dimension coincides with large dimension.*

Note that the result applies to the topological dimension of non-normal completely regular spaces, provided the definitions are framed in terms of normal coverings ; these are the uniform coverings in the finest uniformity compatible with the topology, which is always a locally fine uniformity. [5] Smirnov has established some of the properties of the dimension defined in this way by finite normal coverings ; and also, for general uniform spaces, the dimension defined by extension of mappings into spheres is $\delta d$ [9].

2.7. *A uniform space which is a finite union of subspaces of large dimension* $\leqq n$ *has large dimension* $\leqq n$.

*Proof.* The $\delta$-dimensions of the subspaces coincide with the large dimensions ; then from the sum theorem for $\delta$-dimension [9] we know $\delta d(\mu X) = n$. From 2.3, $\Delta d(\mu X)$ is $n$ or $\infty$. However, every uniform covering of $\mu X$ may be refined by a union of finitely many finite-dimensional uniform coverings of the given subspaces, and hence, as in 2.5, by a finite-dimensional uniform covering of $\mu X$.

In this manner one can choose finite or infinite coverings according to convenience whenever the large dimension is known to be finite. This occurs for example in questions concerning subspaces of finite-dimensional spaces, such as

2.8. (Smirnov) *A subset $S$ of Euclidean space $E^n$ has $\delta$-dimension $n$ if and only if there exists $r>0$ such that for every $\varepsilon > 0$ there is a solid sphere of radius $r$ in which $S$ forms an $\varepsilon$-net.*

*Proof.* The conditions are sufficient, in view of 2.5, for they imply that any uniform neighborhood of $S$ contains a sphere of radius $r$. On the other hand, suppose the conditions not satisfied; thus for each $r>0$ there is $\varepsilon > 0$ such that every $r$-sphere contains a point distant by $\varepsilon$ from $S$. Consider the cell complex $K$ the walls of which are formed by the lattice hyperplanes $x_i = p$, $p$ integral. The first barycentric subdivision $K^1$ of $K$, and all successive barycentric subdivisions $K^m$, are simplicial complexes, with meshes approaching 0. Moreover, each is a uniform complex. In particular, on the $(n-1)$ skeleton of $K^m$, the stars of vertices form a uniform covering $u$. To see this, observe that $u$ is an open covering on any compact portion of space (say, all $|x_i| \leqq 2$), hence has a Lebesgue number there, and every point has a spherical neighborhood of radius 1 on which the restriction of $u$ is congruent to a part of $u$ contained in the specified portion.

To construct a uniform $(n-1)$-dimensional covering of $S$ of arbitrarily small mesh $2\alpha$, choose $m$ so that the mesh of $K^m$ is $\alpha$ or less. Let $\{St_i\}$ be the covering of the $(n-1)$-skeleton of $K^m$ by stars of vertices, and $\theta$ a Lebesgue number for this covering (relative to the $(n-1)$-skeleton). Now there exist, first, $r > 0$ such that every $n$-simplex $\sigma$ of $K^m$ contains a sphere of radius $r$; therefore, by hypothesis, $\varepsilon > 0$ such that each $\sigma$ contains a point distant by $\varepsilon$; finally, if $2\delta = \min(r, \varepsilon)$, each $\sigma$ contains a point $P_\sigma$ distant by $\delta$ from both $S$ and the frontier of $\sigma$. For each vertex $i$ of $K^m$, let $U_i$ consist of $St_i$ together with all open line segments $(p, p_\sigma)$ such that $p$ is a point of $St_i$ and a boundary point of $\sigma$ (thus $i$ is a vertex of $\sigma$). Relatively on the complement of the union of the spheres of radius $\delta$ about all $p_\sigma$ (a set which contain $S$), $\{U_i\}$ has a Lebesgue number, specifically $\theta\delta^2/\alpha^2$. To see this, observe that if $x$ is in the frontier of $\sigma$, and $y$ is an interior point of $\sigma$ within $\theta\delta/\alpha$ of $x$, we may construct two similar right triangles in the plane determined by $x$, $y$, and $p_\sigma$, as follows. Drop a perpendicular from $p_\sigma$ to the hyperplane of a face of $\sigma$ containing $x$, extend the ray from $p_\sigma$ to $y$ until it meets some face of $\sigma$ in a point $q$, and drop a perpendicular from $x$ to the line $qy$.

A sketch shows that $q$ must be within $\theta$ of $x$ and thus in this case some $U_i$ contains both $x$ and $y$. In the case of two interior points $y, z$, of one $n$-simplex $\sigma$, with $d(y, z) < \theta\delta^2/\alpha^2$ and both $y$ and $z$ distant by $\delta$ from $p_\sigma$, draw lines from $p_\sigma$ through $y$ and $z$ until they meet faces of $\sigma$, and observe that the distance between these lines measured parallel to $yz$ cannot increase beyond $\theta\delta/\alpha$ before one of the lines hits a face. The remaining case, that $y$ and $z$ are in different $n$-simplexes, $\sigma$, $\tau$, is similar; use the facts that $y$ and $z$ are nearer to the $(n-1)$-skeleton than to each other and that $\delta \leqq \alpha/2$ (since $r \leqq \alpha$). Finally a point common to $n+1$ or more sets $U_i$ would have to be interior to some $n$-simplex $\sigma$; but projection from $p_\sigma$ would give a point common to the corresponding sets $St_i$, which is absurd.

Let us conclude with a few further remarks. As the statement of 2.3 exhibits, we do not need to know that the $\Delta$-dimension is actually finite to know that $\Delta d$ and $\delta d$ are the same. In particular, they are the same in any space whose uniformity is the weak uniformity induced by a family of real-valued functions. I do not know whether $\Delta d$ and $\delta d$ coincide for all metric uniform spaces.

Dowker and Hurewicz have shown [4] that the covering dimension *dim* for a metrizable space coincides with the *sequential dimension ds* defined as the least $n$ such that there exists a sequence of locally finite open coverings, each of dimension $\leqq n$, of mesh converging to 0, each a closure-refinement of the preceding one. (In particular, the theorem shows that *ds* is a topological invariant, though the concept of mesh converging to 0 is not invariant.) Examination of their proof shows that one can replace the closure-refinements by star-refinements, and conclude: *For metrizable spaces, the covering dimension is the same as the least uniform dimension in any metric uniformity compatible with the topology.* I do not know whether the word "metric" can be omitted.

## References


1. M. Atsuji, *Uniform continuity of continuous functions of metric spaces*, Pacific J. Math. **8** (1958), 11-16.
2. K. Borsuk, *On an irreducible 2-dimensional absolute retract*, Fund. Math. **37** (1950), 137-160.
3. C. Dowker, *Mapping theorems for non-compact spaces*, Amer. J. Math. **69** (1947), 200-242.
4. C. Dowker and W. Hurewicz, *Dimension of metric spaces*, Fund. Math. **43** (1956), 83-88.
5. S. Ginsburg and J. Isbell *Some operators on uniform spaces*, Trans. Amer. Math. Soc., to appear.
6. J. Isbell, *Zero-dimensional spaces*, Tohoku Math. J. **7** (1955), 1-8.
7. Isbell, *Euclidean and weak uniformities*, Pacific J. Math. **8** (1958), 67-86.

8. Katetov, *On real-valued functions in topological spaces*, Fund. Math. **38** (1951), 85-91.
9. Yu. Smirnov, *On she dimension of proximity spaces*, Mat. Sb. **38** (1956), 283-302.
10. Yu. Smirnov, *Geometry of infinite uniform complexes and $\delta$-dimension of proximity spaces*, Mat. Sb. **40** (1956), 137-156.

UNIVERSITY OF WASHINGTON

# ON THE RADIUS OF UNIVALENCE OF THE FUNCTION

$$\exp z^2 \int_0^z \exp(-t^2) dt$$

Erwin Kreyszig and John Todd

**1. Introduction.** We shall determine the radius of univalence $\rho_u$ of the function

$$E(z) = e^{z^2} \int_0^z e^{-t^2} dt \ . \tag{1.1}$$

We shall write $E(z) = w = u(x, y) + iv(x, y)$. On the imaginary axis we have $u = 0$ and $v$, regarded as a function of $y$, has a single maximum at the solution $y = \rho$ of

$$2yv(0, y) = 1 \ .$$

The value of $\rho$ to eight decimal places has been determined by Lash Miller and Gordon [1] and is

$$\rho = 0.92413887 \ . \tag{1.2}$$

It is evident that $\rho_u \leqq \rho$. We shall prove the following theorem.

THEOREM. *The number $\rho$ is the radius of univalence of $E(z)$.*

Recently, the radius of univalence of the error function

$$erf(z) = \int_0^z e^{-t^2} dt$$

was determined [2]. It is interesting to note that when proceeding from $erf(z)$ to $E(z)$ we meet an entirely different situation. In the case of $erf(z)$, points $z_1, z_2$ closest to the origin and such that $erf(z_1) = erf(z_2)$ are conjugate complex and lie far apart from each other. In the case of $E(z)$ points of that nature can be found in an arbitrarily small neigborhood of the point $z = i\rho$.

The actual situation is made clear by the diagram and tables given below. In Fig. 1 we show the curves $R = |E| =$ constant and $\gamma = arg\, E =$ constant in the square $0 \leqq x \leqq 1.5$, $0 \leqq y \leqq 1.5$ of the $z$-plane. The table shows the values of $E$ for $z$ on the curve $C$ (defined below). The values given were obtained by summing an adequate number of terms of the power series on the Datatron 205 at the California Institute of Technology ; some were checked by comparison with the tables of Karpov [4, 5] from which values of $E(z)$ can be obtained.

Received September 3, 1958.

**2. Idea of proof.** Since

$$E(z) = \sum_{n=0}^{\infty} \frac{2^n}{1.3.5\cdots(2n+1)} z^{2n+1}, \qquad |z| < \infty, \tag{2.1}$$

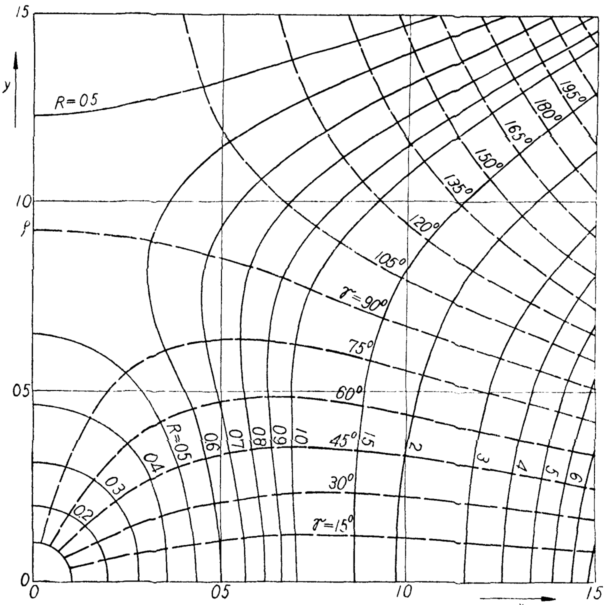



Fig. 1. Curves $R = |E| =$ const. and $\gamma = arg\, E =$ const. in the $z$-plane.

| $x$ | $E(x)$ |
|---|---|
| 0 | 0 |
| 0.1 | 0.1007 |
| 0.2 | 0.2054 |
| 0.3 | 0.3187 |
| 0.4 | 0.4455 |
| 0.5 | 0.5923 |
| 0.6 | 0.7671 |
| 0.7 | 0.9805 |
| 0.8 | 1.2473 |
| 0.9 | 1.5876 |

| $\phi$ | $E(\rho e^{i\phi})$ |
|---|---|
| 0° | 1.6837 |
| 10° | 1.4957+0.6121i |
| 20° | 1.0573+0.9759i |
| 30° | 0.6079+1.0473i |
| 40° | 0.2919+0.9463i |
| 50° | 0.1189+0.8024i |
| 60° | 0.0401+0.6817i |
| 70° | 0.0099+0.6003i |
| 80° | 0.0011+0.5553i |
| 90° | 0.5410i |

| $y$ | $E(iy)$ |
|---|---|
| 0 | 0 |
| 0.1 | 0.0993i |
| 0.2 | 0.1948i |
| 0.3 | 0.2826i |
| 0.4 | 0.3599i |
| 0.5 | 0.4244i |
| 0.6 | 0.4748i |
| 0.7 | 0.5105i |
| 0.8 | 0.5321i |
| 0.9 | 0.5407i |

we have $E(\bar{z}) = \overline{E(z)}$ and $E(-z) = -E(z)$ and may restrict our consideration to the first quadrant $x \geqq 0, y \geqq 0$ in the $z$-plane.
In the subsequent section we shall prove the following lemma.

LEMMA.

$$E(z_1) \neq E(z_2) \tag{2.2}$$

*for any two points on the boundary C of the open sector S of the circular disk* $|z| < \rho$ *in the first quadrant.*

From this it follows, since $E(z)$ is entire and thus regular in $S \cup C$ that $E(z)$ maps $S$ conformally and one-to-one onto the interior of the simple closed curve $C^*$ corresponding to $C$ in the $w$-plane [3, p. 121]. This establishes our theorem.

3. **Proof of the lemma.** Let $z = re^{i\phi}$. The curve $C$ consists of

the segment $S_1$: $\quad y = 0, \quad 0 < x < \rho ,$

the circular arc $K$: $\quad |z| = \rho , \quad 0 < \phi < \pi/2 ,$

the segment $S_2$: $\quad x = 0, \quad 0 < y < \rho .$

and the three common end points of these three arcs.

(A) On $S_1$, $E(z)$ is real and increases steadily with $x$.

(B) On $S_2$, $E(z)$ is imaginary, and $v$ increases steadily with $y$.

(C) $v \neq 0$ on $K$.

(D) On $K$, $|E(z)|$ decreases steadily with increasing $\phi$.

(A) is obvious from (2.1), and (B) follows from the definition of $\rho$.

*Proof of* (C). Integrating along segments parallel to the coordinate axes we have

$$v(x, y) = e^{-y^2}\Big[\cos 2xy \int_0^y e^{\tau^2} \cos 2x\tau d\tau + \sin 2xy \Big\{ e^{x^2}\int_0^x e^{-t^2} dt + \int_0^y e^{\tau^2} \sin 2x\tau d\tau \Big\}\Big] .$$

In $\{x > 0, y > 0\} \cap \{|z| \leqq \rho\}$ we have $\cos 2xy > 0$, $\sin 2xy > 0$. Therefore $v > 0$ on $K$.

*Proof of* (D). Integrating along a radius $\phi =$ constant from 0 to $\rho$ we have

$$E(z) = e^{i\phi}\int_0^\rho e^{h(r,\phi)} dr$$

where

$$h(r, \phi) = a(r, \phi) + ib(r, \phi) ,$$
$$a(r, \phi) = (\rho^2 - r^2) \cos 2\phi, \; b(r, \phi) = (\rho^2 - r^2) \sin 2\phi .$$

Therefore

$$|E|^2 = \int_0^\rho e^h dr \int_0^\rho e^{\bar{h}} dr .$$

Differentiating with respect to $\phi$ and setting

$$h^* = a^* + ib^*, \; a^* = a(r^*, \phi), \; b^* = b(r^*, \phi),$$
$$f = \cos(b^* - b) - i \sin(b^* - b)$$

we obtain

$$(|E|^2)_\phi = \int_0^\rho e^h h_\phi dr \int_0^\rho e^{\overline{h^*}} dr^* + \int_0^\rho e^{h^*} dr^* \int_0^\rho e^{\bar{h}} \bar{h}_\phi dr$$
$$= \int_0^\rho \int_0^\rho e^{a+a^*} \{ f h_\phi + \overline{f h_\phi} \} dr dr^* .$$

Now

$$a_\phi = -2(\rho^2 - r^2) \sin 2\phi , \; b_\phi = 2(\rho^2 - r^2) \cos 2\phi$$

and therefore

$$f h_\phi + \overline{f h_\phi} = 2\Re f h_\phi = 2[\cos(b^* - b) a_\phi + \sin(b^* - b) b_\phi]$$
$$= -4(\rho^2 - r^2) \sin(\alpha(\phi))$$

where

$$\alpha(\phi) = 2\phi + b - b^* = (r^{*2} - r^2) \sin 2\phi + 2\phi .$$

This yields

$$(3.1) \qquad (|E|^2)_\phi = -4 \int_0^\rho \int_0^\rho e^{a+a^*} (\rho^2 - r^2) \sin(\alpha(\phi)) dr dr^* .$$

Since from (1.2) we have $|r^{*2} - r^2| < 1$, we obtain

$$\alpha'(\phi) = 2 + 2(r^{*2} - r^2) \cos 2\phi > 0 .$$

Hence $\alpha(\phi), 0 \leqq \phi \leqq \pi/2$, has its maximum at $\phi = \pi/2$. Therefore $0 \leqq \alpha(\phi) < \pi$ when $0 \leqq \phi < \pi/2$ and $\sin(\alpha(\phi)) > 0$ when $0 < \phi < \pi/2$. This means that the integrand in (3.1) is positive in the region $0 \leqq r \leqq \rho$, $0 \leqq r^* \leqq \rho$ for all $\phi$ in the interval $0 < \phi < \pi/2$. Thus $(|E|^2)_\phi < 0$ when $0 < \phi < \pi/2$. This proves (D).

We note that (D) remains true if $K$ is replaced by quadrants of circles of radii somewhat larger than $\rho$; this, however, is of no interest here.

For $z_1 \in K, z_2 \in S_2$, or $z_1 \in K, z_2 \in K$, equation (2.2) holds, as follows from (D). For $z_1 \in K, z_2 \in S_1$ the same is true because of (C). In the other cases, $z_1 \in S_1, z_2 \in S_1$, etc., the validity of (2.2) is obvious. This proves the lemma.

## References

1. W. Lash Miller and A. R. Gordon, *Numerical evaluation of infinite series and integrals which arise in certain problems of linear heat flow, electrochemical diffusion etc.*, J. Phys. Chem. **35** (1931), 2785-2884. We are indebted to Dr. Walter Gautschi for this reference; Dr. Gautschi has recomputed the value of $\rho$ and communicated the following 10 D value

$$\rho = 0.92413\ 88730.$$

2. Erwin Kreyszig and John Todd, *The radius of univalence of erf z*, Bull. Amer. Math. Soc. **64** (1958), 363-364, and Numerische Mathematik **1** (1959) to appear.

3. J. E. Littlewood, *Lectures on the Theory of Functions*, Oxford, 1944.

4. K. A. Karpov, *Taklicy funkcii* $w(z) = e^{-z^2}\int_0^z e^{x^2}\,dx$ *v kompleksnoi oklasti*, Moscow, 1956.

5. K. A. Karpov *Taklicy funkcii* $F(z) = \int_0^z e^{x^2}dx$ *v kompleksnoi oklasti*, Moscow, 1958.

OHIO STATE UNIVERSITY
COLUMBUS, OHIO
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

# AN INTERPOLATION THEOREM IN THE PREDICATE CALCULUS

ROGER C. LYNDON

**1. Introduction.** In studying the formal structure of sentences whose validity is preserved under passage from an algebraic system to a homomorphic image of the system, we have had occasion to use a lemma from formal logic. A proof of this lemma, our Interpolation Theorem, can be given within the theory of deductive inference, as formalized by Gentzen. Gentzen's theory is rather complicated and perhaps not generally well known. Moreover, the use of any formalized system of deductive logic seems to an extent alien to the primarily algebraic nature of our intended application. Therefore we give here a proof of the Interpolation Theorem that lies entirely within the theory of models: our arguments are as far as possible in the spirit of abstract algebra, and, in particular, borrow nothing from formal logic beyond an understanding of the intended meaning, herein precisely defined, of the conventional symbolism.

The Interpolation Theorem deals with sentences of the Predicate Calculus. Roughly, these are sentences that can be build up using the usual logical connectives, symbols denoting operations (or functions), symbols denoting relations (or predicates), and variables whose range is individual elements of the systems under consideration, but no variables ranging over operations, relations, or sets. The theorem takes the same form whether or not we admit a predicate denoting identity, with suitable axioms, to the predicate calculus. For technical reasons we admit as sentential connectives only the signs for negation, conjunction and disjunction (regarding "if ⋯ then" as a defined concept), together with signs 0 and 1 for truth and falsehood. For each occurrence of a relation symbol in a sentence $S$, there is a unique maximal chain of well formed formulas, all containing the given occurrence and each occurring as a proper part of the next. The given occurrence of the relation symbol will be called *positive* if the number of formulas in this chain that begin with the negation sign is even, and *negative* if this number is odd. If $S$ is in prenex disjunctive form, this criterion takes the simpler form that an occurrence is negative if and only if it is preceded by the negation sign.

INTERPOLATION THEOREM, *Let $S$ and $T$ be sentences such that $S$ implies $T$. Then there exists a sentence $M$ such that $S$ implies $M$ and $M$*

Received October 22, 1958. Work supported in part under grants from the National Science Foundation,

1. See [5] and [9], Chapter XV.

*implies T, and that a relation symbol has positive occurrences in M only if it has positive occurrences in both S and T, and has negative occurrences in M only if it has negative occurrences in both S and T.*

This theorem is a generalization of a result of W. Craig [3, 4]; Craig's lemma is obtained from it by suppressing the distinction between positive and negative sentences. As indicated, our first proof of the Interpolation Theorem used the Gentzen calculus; it did not differ essentially from Craig's proof, at that time unpublished, of his lemma.

The leading idea of the present proof is to interpret $S$ implies $T$ to mean that $T$ holds in every model for which $S$ holds; we express this relation by writing $S \Rightarrow T$. By Gödel's Completeness Theorem [6], this semantic interpretation is equivalent to the interpretation $S \vdash T$, that $T$ is a formal consequence of $S$ in a deductive axiomatization of the predicate calculus. The crucial point in our argument is the Main Theorem, which serves as a substitute, under this interpretation, for results in the theory of proof due to Herbrand [8] and to Gentzen [5].

A theorem of the theory of proof may be taken, in general, as saying that if there exists any derivation of one set $\Delta$ of formulas from a set $\Gamma$, then there exists a derivation with certain special properties. A semantic counterpart of such a theorem will take the form of an 'interpolation theorem': if $\Gamma \Rightarrow \Delta$, then there exists a chain $\Gamma = \Gamma^1, \Gamma^2, \cdots, \Gamma^n = \Delta$ of sets of formulas, with certain special properties, such that $\Gamma^1 \Rightarrow \Gamma^2, \cdots, \Gamma^{n-1} \Rightarrow \Gamma^n$. Theorems of this sort will ordinarily require the occurrence in the $\Gamma^k$ of additional symbols (for the 'Skolem functions') that do not appear in $\Gamma$ or $\Delta$, although this is not true of the Interpolation Theorem. Our arguments abjure any formal use of the concept of deductive derivablilty, hence of the Completeness Theorem. In various special cases, where $\Gamma \vdash \Delta$ would be immediate, that $\Gamma \Rightarrow \Delta$ follows directly from our definitions. The more difficult half of the Completeness Theorem, that if $\Gamma \Rightarrow \Delta$ then $\Gamma \vdash \Delta$, is implicit in the Main Theorem, which guarantees the existence of a chain $\Gamma = \Gamma^1, \cdots, \Gamma^n = \Delta$ such that at each step the relation $\Gamma^k \vdash \Gamma^{k+1}$ is immediately evident.

I have profited much from discussions related to the present topic with A. Tarski and L. Henkin[2]; in particular, Tarski has emphasized the desirability of establishing the Interpolation Theorem by methods independent of the theory of proof. The idea of providing semantic proofs of results from the theory of proof is not new: a proof by E. Beth [1, 2], in a quite different formalism, of Craig's Lemma would certainly serve as well to prove the Interpolation Theorem; and A. Robinson has likewise provided semantic proofs of closely related results [10]. Unpublished results similar to those presented here have recently been

[2] In particular, while the author was visiting at the University of California, Berkeley.

obtained by A. Grzegorczyk, A. Mostowski and C. Ryll-Nardzewski, and by R. Vaught.

2. **Basic concepts.**[3] A *language* $L$ is determined by an ordered quadruple, $V, W, R, \rho$, where $V, W, R$ are disjoint sets, $V$ infinite, and $\rho$ is a function from $W \cup R$ to the natural numbers. The elements of $V$ will be called *variables*, those of $W$ *operation symbols*, and those of $R$ *relation symbols*; for $w$ in $W$, $r$ in $R$, $\rho(w)$ is the *rank* of $w$ and $\rho(r)$ the *rank* of $r$. The *logical symbols* are $0, 1, \sim, \wedge \vee, \forall, \exists$. The expressions of $L$ will be made up of these symbols together with parentheses and commas. A *term* is, recursively, any variable, and any expression $w(t_1, \cdots, t_{\rho(w)})$ where $w$ is an operation symbol and $t_1, \cdots, t_{\rho(w)}$ are terms. An *atomic formula* is any expression $r(t_1, \cdots, t_{\rho(r)})$ where $r$ is a relation symbol and $t_1, \cdots, t_{\rho(r)}$ are terms. A *formula* is, recursively, any atomic formula, and any expression $0, 1, \sim F, (F \wedge G), (F \vee G), \forall xF, \exists xF$ where $F$ and $G$ are formulas and $x$ is a variable. Formally, we define $L$ to be the set of its symbols, terms and formulas.

We introduce the abbreviations $F \supset G$ for $(\sim F \vee G)$, $\bigwedge_1^n F_i$ for $F_1 \wedge \cdots \wedge F_n$ with the convention $\bigwedge_1^0 F_i = 1$, and $\bigvee_1^n F_i$ for $F_1 \vee \cdots \vee F_n$ with $\bigvee_1^0 F_i = 0$, and write $\forall x_1 \cdots x_n$ for $\forall x_1 \cdots \forall x_n$. A *matrix* is a formula that does not contain $\forall$ or $\exists$. A *normal matrix* is a matrix of the form $\bigvee_{i=1}^{m} \bigwedge_{j=1}^{n_m} F_{ij}$ where each $F_{ij}$ is either $A_{ij}$ or $\sim A_{ij}$, for $A_{ij}$ an atomic formula. A *prenex* formula is one of the form $Q_1x_1 \cdots Q_qx_qM$ where each $Q_i$ is $\forall$ or $\exists$, each $x_i$ is a variable, and $M$ is a matrix; the formula is *normal* if the matrix $M$ is normal. An occurrence of a variable $x$ in a formula $F$ is *free* in the formula $F$ if it is not part of a subformula of the forms $\forall xG$ or $\exists xG$. A *sentence* is a formula without free occurrences of variables.

It is easily shown by induction that if $G$ is any part of a formula $F$, then there is a smallest part of $F$ that is a formula and contains $G$. It follows that there is a unique maximal chains of formulas $H_1, \cdots, H_n = F$, each a proper part of the next, and all containing $G$. The part $G$ is *positive* in $F$ if the number of $H_{i+1} = \sim H_i$ is even, and *negative* if it is odd. In what follows, $G$ will always be an occurrence of a relation symbol in $F$.

An *interpretation* of a language $L$ is determined by a set $A$ and a function $\mu$, defined on $V \cup W \cup R$, such that $\mu x \in A$ for $x \in V$, $\mu w \in A^{A^{\rho(w)}}$ for $w \in W$, and $\mu r \in 2^{A^{\rho(r)}}$ for $r \in R$. We regard 2 as the two element Boolean algebra with elements 0, 1 and operations $\sim, \wedge, \vee$, so that $\mu r$ is a function with values $(\mu r)(a_1, \cdots, a_{\rho(r)})$ equal to 0 or 1; but in practice we indulge in the harmless ambiguity of treating $\mu w$ as a subset of $A^{\rho(w)+1}$

and $\mu r$ of $A^{\rho(r)}$, and accordingly using such notation as $\mu w \subseteqq \mu w'$, $\mu r \subseteqq \mu r'$. Putting aside the trivial case that $L$ contains no relation symbols of positive rank, $\mu$ unambiguously determines its *domain* $A$.

The function $\mu$ determines a unique extension mapping all terms of $L$ into $A$, by the recursive definition

$$\mu[w(t, \cdots, t_{\rho(w)})] = (\mu w)(\mu t, \cdots, \mu t_{\rho(w)}) .$$

A further extension mapping all formulas of $L$ into 2 is determined by the conditions.

$$(1) \qquad \mu 0 = 0, \mu 1 = 1, \ \mu(\sim F) = \sim \mu F, \ \mu(F \wedge G) = \mu F \wedge \mu G, \\ \mu(F \vee G) = \mu F \vee \mu G ,$$

and

$$(2) \qquad \left.\begin{array}{l} \mu(\forall x F) = 1 \ \text{ if and only if } \ \lambda F = 1 \ \text{ for all } \lambda \\ \mu(\exists x F) = 1 \ \text{ if and only if } \ \lambda F = 1 \ \text{ for some } \lambda \end{array}\right\} \text{ such that}$$

$\lambda z = \mu z$ for all $z$ in $V \cup W \cup R - \{x\}$. Formally, we define an interpretation to be a function $\mu$ thus extended; in practice we shall say that $\mu$ and $\lambda$ agree except on $x$ when we mean that $\mu$ and $\lambda$ agree for all $z$ in $V \cup W \cup R - \{x\}$.

A *model* of $L$ is the restriction $\mathfrak{A}$ of an interpretation $\mu$ to the operation and relation symbols of $L$. The model $\mathfrak{A}$ may be regarded as a 'relational system'[4] consisting of a set $A$, its domain, together with a set of *operations* $\mathfrak{A}w$ indexed by the operation symbols $w$ of $L$, and a set of *relations* $\mathfrak{A}r$ indexed by the relation symbols $r$ of $L$. If $\mathfrak{A}$ is the restriction of $\mu$, we call $\mu$ an interpretation in the model $\mathfrak{A}$. If $\mu F = 1$, we say that $F$ *holds* for the interpretation $\mu$. Evidently $\mu F$ depends only on the domain $A$ of $\mu$, the values of $\mu$ on the operation and relation symbols that occur in $F$, and the values of $\mu$ on the variables that occur free in $F$. In particular, if $S$ is a sentence, $\mu S$ depends only on the model $\mathfrak{A}$ to which $\mu$ belongs, and if $\mu S = 1$ we say that $S$ holds in the model $\mathfrak{A}$.

If $\Gamma$ and $\Delta$ are sets of formulas of $L$, we say that $\Gamma$ *implies* $\Delta$ *in* $L$ if $\mu\Delta = \{1\}$ for all interpretations of $L$ such that $\mu\Gamma = \{1\}$. This interpretation is evidently independent of $L$, provided only that $\Gamma$ and $\Delta$ belong to $L$; we say simply that $\Gamma$ *implies* $\Delta$, and write $\Gamma \Rightarrow \Delta$. We write $\mu\Gamma = 1$ for $\mu\Gamma = \{1\}$, and employ such notation as $\Gamma_1, \Gamma_2 \Rightarrow F$ with the obvious meaning. If $\Gamma \Rightarrow \Delta$ and $\Delta \Rightarrow \Gamma$, then $\Gamma$ and $\Delta$ are *equivalent* and we write $\Gamma \Leftrightarrow \Delta$. That $1 \Rightarrow F$ expresses that $F$ is a *theorem*. A set $\Gamma$ is called *consistent* if there exists an interpretation $\mu$ such that $\mu\Gamma = 1$; thus $\Gamma \Rightarrow 0$ expresses that the set $\Gamma$ is *inconsistent*.

[4] See **[11]**, **[12]**.

**3. Preliminary propositions.** The set $\Phi = \Phi(L)$ of all formulas of $L$ constitutes, in an obvious sense, an algebraic system with operations $0, 1, \sim, \wedge, \vee$; in fact it is a 'word algebra', a free algebra without axioms. The relation $F \Leftrightarrow G$ is a congruence on $\Phi$, and the quotient system $\overline{\Phi}$ is a Boolean algebra, the Lindenbaum algebra of $L$. If $\kappa$ is the canonical map of $\Phi$ onto $\overline{\Phi}$, then every interpretation $\mu$ of $L$, when restricted to $\Phi$, can be factored uniquely in the form $\mu = \bar{\mu}\kappa$ where $\bar{\mu}$ is a homomorphism of $\overline{\Phi}$ onto 2.

The set $\Phi_0$ of all matrices of $L$ constitutes a subalgebra of $\Phi$, and its image $\overline{\Phi_0} = \kappa\Phi_0$ is a subalgebra of $\overline{\Phi}$. Every homomorphism $\theta$ of $\overline{\Phi_0}$ onto 2 can be extended to a homomorphism $\theta'$ of $\overline{\Phi}$ onto 2 such that $\theta'\kappa$ is an interpretation. To prove this we construct the special interpretation $\mu$ *induced* by $\theta$. For the domain $A$ of $\mu$ we take the set of all terms of $L$. For a variable $x$, define $\mu x = x$. For an operation symbol $w$ and terms $t_1, \cdots, t_{\rho(w)}$, we define $\mu w$ by assigning to $(\mu w)(t_1, \cdots, t_{\rho(w)})$ as value the term $w(t_1, \cdots, t_{\rho(w)})$. For a relation symbol $r$ and terms $t_1, \cdots, t_{\rho(r)}$ we define $\mu r$ by assigning to $(\mu r)(t_1, \cdots, t_{\rho(r)})$ the value $\theta\kappa[r(t_1, \cdots, t_{\rho(r)})]$ in 2. By virtue of the last definition, $\mu F = \theta\kappa F$ for all atomic formulas $F$. Since the images $\kappa F$ of the atomic formulas $F$ generate $\overline{\Phi_0}$, and $\bar{\mu}\kappa F = \theta\kappa F$ for atomic $F$, it follows that $\bar{\mu} = \theta$ on $\overline{\Phi_0}$ and $\bar{\mu}$ is an extension of $\theta$.

PROPOSITION 1. *If $\Gamma$ is a set of matrices, and $J$ the dual ideal in the Boolean algebra $\overline{\Phi_0}$ generated by $\kappa\Gamma$, then $\Gamma \Rightarrow 0$ if and only if $0 \in J$.*

*Proof.* Assume $0 \in J$. Then $0 = \kappa F_1 \wedge \cdots \wedge \kappa F_n$ for some $F_1, \cdots, F_n$ in $\Gamma$. If $\mu$ is an interpretation such that $\mu\Gamma = 1$, then each $\bar{\mu}\kappa F_i = \mu F_i = 1$, whence $1 = \bar{\mu}\kappa \bigwedge F_i = \bar{\mu} \bigwedge \kappa F_i = \bar{\mu}0 = 0$, a contradiction. Assume $0 \notin J$. Then $J \neq \Phi_0$ and $J \subseteq K$ for some maximal dual ideal $K$ in $\overline{\Phi_0}$ If $\theta$ is the canonical map of $\Phi_0$ onto 2 with kernel the maximal ideal $\overline{\Phi_0} - K$ complementary to the dual ideal $K$, then $\theta\kappa\Gamma \subseteq \theta J \subseteq \theta\kappa = 1$. If $\mu$ is the special interpretation of $L$ induced by the homomorphism $\theta$, then $\mu\Gamma = \bar{\mu}k\Gamma = \theta\kappa\Gamma = 1$, whence $\Gamma$ is consistent.

COROLLARY 1.1. *If $\Gamma$ is a set of matrices, then $\Gamma \Rightarrow 0$ if and only if $\Gamma_0 \Rightarrow 0$ for some finite subset $\Gamma_0$ of $\Gamma$.*

Every map $\sigma$ of the atomic formulas of $L$, as free generators of $\Phi_0$, into $\Phi_0$, extends to an endomorphism of $\Phi_0$, which in turn induces an endomorphism $\bar{\rho}$ of $\overline{\Phi_0}$. It follows that if $\Gamma \Rightarrow 0$ then $\sigma\Gamma \Rightarrow 0$. Every map $\sigma$ of the variables of $L$ into terms of $L$ extends in an obvious way to a map of the terms of $L$ into terms of $L$, hence of formulas of $L$ into formulas of $L$; a transformation induced in this fashion will be called a *substitution.*

PROPOSITION 2. *Let $\Gamma$ be set of sentences $S$ of the form $\forall x_1 \cdots x_n M$ where the $M$ are matrices, and $\Gamma'$ the set of all formulas $\sigma M$ where $\sigma$ is a substitution and $M$ is the matrix of some sentence $S$ in $\Gamma$. Then $\Gamma \Rightarrow 0$ if and only if $\Gamma' \Rightarrow 0$.*

*Proof.* Suppose that $\Gamma'$ is consistent. Then $\lambda\Gamma' = \bar{\lambda}\kappa\Gamma' = 1$ for some interpretation $\lambda$. Let $\mu$ be the special interpretation induced by the homomorphism $\lambda$ of $\Phi_0$ onto 2. Let $F = \forall x_1 \cdots x_n M$ be in $\Gamma$, and $\nu$ be an interpretation that agrees with $\mu$ except on $x_1, \cdots, x_n$. Since the values $\nu x$ for variables $x$ are terms, we may define a substitution by setting $\sigma x = \nu x$. Since $\mu x = x$ for all variables $x$, $\nu M = \mu\sigma M = \bar{\lambda}\sigma M = 1$. This establishes that $\mu F = 1$. Suppose $\Gamma'$ is inconsistent. Then for all interpretations $\mu$ there is some $F = \forall x_1 \cdots x_n M$ in $\Gamma$ and some substitution $\sigma$ such that $\mu\sigma M = 0$. Then setting $\lambda x_i = \mu\sigma x_i$, $i = 1, \cdots, n$ defines an interpretation $\lambda$ that agrees with $\mu$ except on $x_1, \cdots, x_n$, and such that $\lambda M = 0$. It follows that $\mu F = 0$.

COROLLARY 2.1. *If $\Gamma$ is a set of universal sentences, of the form $F = \forall x_1 \cdots x_n M$, where $M$ is a matrix, then $\Gamma \Rightarrow 0$ if and only if $\Gamma_0 \Rightarrow 0$ for some finite subset $\Gamma_0$ of $\Gamma$.*

A prenex sentence $S$ of the language $L$ may be written in the form

$$S = \forall x_{11} \cdots x_{1m_1} \exists y_1 \cdots \forall x_{n1} \ldots x_{nm_n} \exists y_n M$$

where $n, m_1, \cdots, m_n$ are natural numbers, the $x_{pq}$ and $y_r$ are variables, and $M$ is a matrix. The *Skolem matrix* of $S$ is the result $\sigma M$ of substituting $\sigma y_r = s_r(x_{11}, \cdots, x_{rm_r})$ and $\sigma z = z$ for all other variables $z$; here the $s_1, \cdots, s_n$ are new and distinct operation symbols which we may suppose uniquely associated with the pair consisting of $S$ and $L$. The *Skolem form* of $S$ is the sentence $\forall x_{11} \cdots x_{nm_n} \sigma M$. The Skolem form belongs to the language $L'$ obtained by adjoining the symbols $s_1, \cdots, s_n$ to $L$.

LEMMA 3. *Let $S$ be a sentence of the form*

$$S = \forall x_{11} \cdots x_{1m_1} \exists y_1 \cdots x_{n1} \forall \cdots x_{nm_n} \exists y F\ ,$$

*where the $x_{pq}$ and $y_r$ are distinct variables and $F$ is a formula in which all occurrences of these variables are free. Let $F'$ result from $F$ by substituting for each $y_r$ a term $\sigma y_r$ that contains no variables other than $x_{11}, \cdots, x_{rm_r}$. Let $S'$ be the sentence*

$$S' = \forall x_{11} \cdots x_{1m_1} x_{21} \cdots x_{nm_n} F'.$$

*Then* $S' \Rightarrow S$.

*Proof*[5]. We proceed by induction. For $n = 0$ the assertion is trivial. For $n = 1$ it suffices to observe that if $\mu$ is an interpretation such that $\mu F' = 1$, then defining an interpretation $\lambda$ to agree with $\mu$ except on $y_1$, and setting $\lambda y_1 = \mu\sigma y_1$, gives $\lambda F = \mu F'$, hence $\lambda F = 1$. For $n > 1$, form $F''$ from $F$ by substituting $\sigma y_r$ for $y_r$, all $y_r$ except $y_n$, and let $S'' = \forall x_{11} \cdots x_{nm_n} \exists y F''$. Then the case $n = 1$ applies to give $S' \Rightarrow S''$, and the case $n - 1$ to give $S'' \Rightarrow S$.

PROPOSITION 4. *Let* $\Gamma$ *be a set of prenex sentences of a language* $L$, *and* $\Gamma'$, *in an extended language* $L'$, *the set of all Skolem forms of the sentences in* $\Gamma$. *Then* $\Gamma$ *holds in a model* $\mathfrak{A}$ *of* $L$ *if and only if* $\Gamma'$ *holds in some extension* $\mathfrak{A}'$ *of* $\mathfrak{A}$ *to a model of* $L'$.

*Proof*. By an induction it evidently suffices to establish the conclusion under the assumption that $\Gamma'$ results from $\Gamma$ by replacing a single sentence $S$ by its Skolem form $S'$. If $\Gamma'$ holds in an extension $\mathfrak{A}'$ of $\mathfrak{A}$ to $L'$, it follows by Lemma 3 that $\Gamma$ holds in $\mathfrak{A}'$, and, since $\Gamma$ belongs to $L$, that $\Gamma$ holds in $\mathfrak{A}$. For the rest, by a second induction it suffices to establish the conclusion for $S = \forall x_1, \cdots, x_m \exists y F$, $S' = \forall x_1 \cdots x_m \sigma F$, $F$ a formula, $\sigma x_i = x_i$, $i = 1, \cdots, m$, and $\sigma y = s(x_1, \cdots, x_m)$, where $s$ does not belong to $L$ and $L'$ is obtained by adjoining $s$ to $L$.

Assume now that $\Gamma$ holds in $\mathfrak{A}$. For any $a_1, \cdots, a_m$ in the domain $A$ of $\mathfrak{A}$, there exists an interpretation $\mu$ in $\mathfrak{A}$ such that $\mu x_i = a_i$, $i = 1, \cdots, m$. Since $\mu S = 1$, it follows that $\mu(\exists y F) = 1$, and there exists an interpretation $\lambda$ that agrees with $\mu$ except on $y$ such that $\lambda F = 1$. By the axiom of choice we may define a function $f$ from $A^m$ into $A$ by choosing for all $a$ , $\cdots, a_m$ interpretations $\mu$ and $\lambda$ as above and setting $f(a_1, \cdots, a_m) = \lambda y$. Extend $\mathfrak{A}$ to $\mathfrak{A}'$ by defining $\mathfrak{A}'s = f$. If $\mu'$ is an interpretation in $\mathfrak{A}'$, then $\mu'$ agrees with some $\mu, \lambda$ as above on the variables $x_1, \cdots, x_m$. Moreover, $\mu'\sigma y = f(\mu' x_1, \cdots, \mu' x_m) = f(a_1, \cdots, a_m) = \lambda y$, whence $\mu'\sigma F = \lambda F = 1$. It follows that $\mu' S'$ 1 for all interpretations $\mu'$ in $\mathfrak{A}'$, whence $\Gamma'$ holds in $\mathfrak{A}'$.

COROLLARY 4.1. *If* $\Gamma$ *is any set of prenex sentences, then* $\Gamma \Rightarrow 0$ *if and only if* $\Gamma_0 \Rightarrow 0$ *for some finite subset* $\Gamma_0$ *of* $\Gamma$.

Every sentence is equivalent to a prenex sentence, and, indeed, a normal sentence. This follows by induction from various immediate consequences of the definitions, of which $\sim(F \wedge G) \Leftrightarrow (\sim F \vee \sim G)$ and $\forall x(F \wedge G) \Leftrightarrow (\forall x F \wedge \forall x G)$ are typical. In fact, it is easily seen that

[5] C. C. Chang pointed out to me a gap in an earlier version of this proof.

every sentence $S$ is equivalent to a normal sentence $S'$ such that a relation symbol occurs positively (negatively) in $S'$ only if it occurs positively (negatively) in $S$.

In view of this, Corollary 4.1 yields the Compactness Theorem.

PROPOSITION 5. *If $\Gamma$ is any set of sentences, then $\Gamma \Rightarrow 0$ if and only if $\Gamma_0 \Rightarrow 0$ for some finite subset $\Gamma_0$ of $\Gamma$.*

**4. The main theorem.** Let $S$ be a prenex sentence, of the form

$$S = \forall x_{11} \ldots \forall_{1m_1} \exists y_1 \cdots \forall x_{n1} \ldots x_{nm_n} \exists y_n M .$$

A second sentence $S_0$ will be said to arise from $S$ by *duplication* if

(i) $\pi_1, \cdots, \pi$ are substitutions such that all $\pi_i x_{pq} = x^i_{pq}$, $\pi_i y_r = y^i_r$, where the $x^i_{pq}$ and $y^i_r$ are distinct variables; and

(ii) $S_0$ results from $\pi_1 M \wedge \cdots \wedge \pi_a M$ by prefixing quantifiers $\forall x^i_{pq}$ and $\exists y^i_r$ in some order such that, for $p \leqq r$, $\forall x^i_{pq}$ precedes $\exists y^i_r$.

PROPOSITION 6. *If $S_0$ arises from $S$ by duplication, then $S \Rightarrow S_0$.*

*Proof.* Let $S$ have Skolem matrix $\sigma M$, in the language $L'$, where $\sigma x_{pq} = x_{pq}$ and $\sigma y_r = s_r(x_{11}, \cdots x_{rm_r})$. By Proposition 4, if $S$ holds in any model $\mathfrak{A}$, then its Skolem form $S'$ holds in some extension $\mathfrak{A}'$ of $\mathfrak{A}$ to $L'$. If $\mu$ is an interpretation of $L'$ in $\mathfrak{A}'$, then every substitution instance of $\sigma M$ holds in $\mu$; in particular, all $\pi_i \sigma M$ hold in $\mu$, whence $\bigwedge \pi_i \sigma M$ holds in $\mu$. But $\bigwedge \pi_i \sigma M$ results from $\bigwedge \pi_i M$ by substituting $s_r(x^i_{11}, \cdots, x^i_{rm_r})$ for each $y^i_r$, whence, by Lemma 3, $S_0$ holds in $\mathfrak{A}'$, and therefore in $\mathfrak{A}$.

For $S$ as before, a second sentence $S_0$ will be said to arise from $S$ by *specialization* if

(iii) $\theta$ is a substitution such that $\theta y_r = y_r$, while each $\theta x_{pq}$ is a term in certain new variables $u_1 \cdots, u_a$ together with the $y_r$ for $r < p$; and

(iv) $S_0$ results from $\theta M$ by prefixing quantifiers $\forall u_h$ and $\exists y_r$ in some order such that $\forall u_h$ precedes $\exists y_r$ if $u_h$ occurs in any $\theta x_{pq}$ for $p \leqq r$, and $\exists y_s$ precedes $\exists y_r$ if $y_s$ occurs in any $\theta x_{pq}$ for $p \leqq r$.

PROPOSITION 7. *If $S_0$ arises from $S$ by specialization, then $S \Rightarrow S_0$.*

*Proof.* Let $S$ have Skolem matrix $\sigma M$ in $L'$ as before. Define a substitution $\rho$ by setting $\rho z = z$ for all variables $z$ other than the $y_r$, and, by recursion on the order of quantification of the $y_r$ in $S_0$, defining $\rho y_r = \rho\theta\sigma y_r = s_r(\rho\theta x_{11}, \cdots, \rho\theta x_{rm_r})$. Since all $y_s$ that occur in $\theta\sigma y_r$ occur in some $\theta x_{pq}$ for $p \leqq r$, all such $y_s$ precede $y_r$ in $S_0$, and the recursion if legitimate. Since $\theta y_r = y_r$, $\rho\theta y_r = \rho y_r = \rho\theta\sigma y_r$ by the above definition, while for all other variables $z$, $\sigma z = z$ and again $\rho\theta z = \rho\theta\sigma z$. Suppose now that $S$ holds in a model $\mathfrak{A}$ of $L$, and hence, by Proposition 4, that the Skolem form $S'$ of $S$ holds in an extension $\mathfrak{A}'$ of $\mathfrak{A}$ to $L'$. Then, for every interpretation $\mu$ in $\mathfrak{A}'$, all instances of $\sigma M$ hold, and, in particular, $\rho\theta\sigma M$ holds. Since $\rho\theta\sigma = \rho\theta$, $\rho\theta\sigma M = \rho\theta M$. Now $\rho\theta M$ results from $\theta M$ by the substitution $\rho$, and $\rho u_h = u_h$, while $\rho y_r$ contains only those $u_h$ that occur in the $\rho\theta x_{pq}$ for $p \leqq r$; by induction, using (iii), these are among the $u_h$ that occur in $\theta x_{pq}$ for $p \leqq r$, and hence among the $u_h$ that precede $y_r$ in $S_0$. Therefore Lemma 3 applies to establish that $S_0$ holds in $\mathfrak{A}'$ and thus in $\mathfrak{A}$.

Let $S^1, S^2$ be prenex sentences of the form, for $\delta = 1, 2$,

$$S^\delta = \forall x^\delta_{11} \cdots x^\delta_{1m_1} \exists y^\delta_1 \cdots \forall x^\delta_{n1} \cdots x^\delta_{nm_n} \exists y^\delta_n M^\delta$$

with Skolem matrices $\sigma M^\delta$ in a language $L'$, where $\sigma x^\delta_{pq} = x^\delta_{pq}$, $\sigma y^\delta_r = s^\delta_r(x^\delta_{11}, \cdots x^\delta_{rm_r})$. Then $S^1$ and $S^2$ will be called *propositionally inconsistent* if there exists a substitution $\eta$ in $L'$ that is one-to-one on all atomic formulas of each $\sigma M^\delta$ such that $\eta\sigma M^1, \eta\sigma M^2 \Rightarrow 0$.

PROPOSITION 8. *If $S^1$, $S^2$ are propositionally inconsistent, then $S^1 S^2 \Rightarrow 0$.*

*Proof.* Suppose $S^1, S^2$ were consistent, hence both held in some model $\mathfrak{A}$ of $L$. Using Proposition 4, all instances of $\sigma M^1$ and $\sigma M^2$ would hold for all interpretations in a certain extension $\mathfrak{A}'$ of $\mathfrak{A}$ to a model of $L'$. Then $\eta\sigma M^1$ and $\eta\sigma M^2$ would hold for all such $\mu$, and $\mu 0 = 1$, a contradiction.

In propositions 6, 7 and 8 we have attempted to isolate the chief ideas that underly the Main theorem; the proof of this theorem can now be accomplished by easier and more natural stages, although at the cost of a small amount of repetition.

MAIN THEOREM. *Let $S^1$ and $S^2$ be prenex sentences such that $S^1, S^2 \Rightarrow 0$. Then there exist prenex sentences $T^1$, $T^2$, $U^1$ and $U^2$ such that* (1) *$T^1$ arises from $S^1$, and $T^2$ from $S^2$, by duplication*; (2) *$U^1$ arises from $T^1$, and $U^2$ from $T^2$, by specialization*; *and* (3) *$U^1$ and $U^2$ are propositionally inconsistent.*

*Proof.* Let $S^1, S^2, M^1, M^2, \sigma$ and $L, L'$ be as above. (There is clearly no loss of generality in taking common values of $n$ and the $m_r$, and a common substitution $\sigma$, for $S^1$ and $S^2$.) By Proposition 4, $S^1, S^2 \Rightarrow 0$ implies that their Skolem forms are inconsistent. By Proposition 2, the set of all instances of $\sigma M^1$ and $\sigma M^2$ is consistent. By Corollary 1.1 some finite set of these instances is inconsistent. Therefore there exist substitutions $\eta_1, \cdots, \eta_a$ in the language $L'$ such that

$$\eta_1\sigma M^1, \cdots, \eta_a\sigma M^1, \eta_1\sigma M^2, \cdots, \eta_a\sigma M^2 \Rightarrow 0 .$$

Define substitutions $\pi_1, \cdots, \pi_a$ such that all $\pi_i x^\delta_{pq} = x^{\delta i}_{pq}$ and $\pi_i y^\delta_r = y^{\delta i}_r$, where the $x^{\delta i}_{pq}$ and $y^{\delta i}_r$ are new and distinct variables. Define $\sigma'$ such that $\sigma' x^{\delta i}_{pq} = x^{\delta i}_{pq}$ and $\sigma' y^{\delta i}_r = s^\delta_r(x^{\delta i}_{11}, \cdots, x^{\delta i}_{rm_r})$; thus $\sigma'\pi_i M^\delta = \pi_i \sigma M^\delta$ for all $\pi_i$. Define $\eta$ such that $\eta x^{\delta i}_{pq} = \eta_i x^\delta_{pq}$; then $\eta\sigma'\pi_i M^\delta = \eta\pi_i\sigma M^\delta = \eta_i\sigma M^\delta$. Define $M^\delta_0 = \bigwedge \pi_i M$; then $\eta\sigma' M^\delta_0 = \bigwedge \pi_i \sigma M^\delta$, and $\eta\sigma' M^1_0, \eta\sigma' M^2_0 \Rightarrow 0$.

Let $S^\sigma_0$ be the sentence obtained from $M^\delta_0$ by prefixing quantifiers $\forall x^{\delta i}_{pq}$ and $\exists y^{\delta i}_r$ in an order such that, if $z$ and $z'$ are two of these variables and the term $\eta\sigma' z$ is shorter than the term $\eta\sigma' z'$, then the quantification of $z$ precedes that of $z'$. If $p \leqq r$, the term $\eta\sigma' x^{\delta i}_{pq} = \eta x^{\delta i}_{pq}$ is a proper part of the term $\eta\sigma' y^{\delta i}_r = s^\delta_r(\eta x^{\delta i}_{11}, \cdots, \eta x^{\delta i}_{rm_r})$, whence $\forall x^{\delta i}_{pq}$ precedes $\exists y^{\delta i}_t$ in $S^\delta_0$. Thus $S^\delta_0$ arises from $S^\delta$ by duplication.

Let $S^\delta_0$ have Skolem matrix $\sigma_0 M_0$ where $\sigma_0 x^{\delta i}_{pq} = x^{\delta i}_{pq}$ and $\sigma_0 y_r = s^{\delta i}_r(\cdots, x^{\delta j}_{pq}, \cdots)$, the arguments ranging, in order of occurrence in $S^\delta_0$, over all $x^{\delta j}_{pq}$ that precede $y_r$ in $S^\delta_0$. One has $\eta\sigma' x^{\delta i}_{pq} = \eta\sigma_0 x^{\delta i}_{pq}$, but $\eta\sigma' y_r = s^\delta_r(\eta x^{\delta i}_{11}, \cdots, \eta x^{\delta i}_{r\,m_r})$ while the term $\eta\sigma_0 y^{\delta i}_r = s^{\delta i}_r(\cdots, \eta x^{\delta j}_{pq}, \cdots)$ begins with a different operation symbol and contains additional arguments. To bring these into agreement, define a transformation $\chi$ on terms as follows:

(1) $\chi z = z$ for a variable $z$;

(2) $\chi\eta\sigma' y^{\delta i}_r = \chi\eta\sigma_0 y^{\delta i}_r$;

(3) for any term $t = w(t_1, \cdots, t_{\rho(w)})$ not of the form $\eta\sigma' y^{\delta i}_r$,
$$\chi t = w(\chi t_1, \cdots, \chi t_{\rho(w)}) .$$

The clause (2) if legitimate, by an induction on length of $\eta\sigma' y_r$. For $\chi\eta\sigma_0 y^{\delta i}_r = s^{\delta i}_r(\cdots, \chi\eta x^{\delta j}_{pq}, \cdots)$ contains $\chi\eta\sigma' y^{\delta k}_s$ only for those $\chi\eta\sigma' y^{\delta k}_r$ that occur in some $\chi\eta x^{\delta j}_{pq}$ for $p \leqq r$, and it follows by an induction that for all of these $s < p$. Let $L_0$ be the language obtained from $L$ by adjoining the symbols $s^{\delta i}_r$. Although neither $\chi$ nor $\chi\eta$ is in general a substitution, when applied to terms of $L_0$, which do not contain symbols $s^\delta_r$, the clause (2) is never invoked; consequently the restriction $\eta_0$ of $\chi\eta$ to $L_0$ is a substitution.

Since $\eta\sigma' M^1_0, \eta\sigma' M^2_0 \Rightarrow 0$, and $\chi$ induces a transformation on terms, it follows that $\chi\eta\sigma' M^1_0, \chi\eta\sigma' M^2_0 \Rightarrow 0$. Now $\chi\eta\sigma' y^{\delta i}_r = \chi\eta\sigma_0 y^{\delta i}_r$ by definition, while $\sigma' x^{\delta i}_{pq} = x^{\delta i}_{pq} = \sigma_0 x^{\delta i}_{pq}$ implies that $\chi\eta\sigma' x^{\delta i}_{pq} = \chi\eta\sigma_0 x^{\delta i}_{pq}$; it follows that $\chi\eta\sigma' M^\delta_0 = \chi\eta\sigma_0 M^\delta_0 = \eta_0\sigma_0 M^\delta_0$, the last since $\sigma_0 M^\delta_0$ belongs to $L_0$. Hence, $\eta_0\sigma^0 M^1_0, \eta_0\sigma_0 M^2_0 \Rightarrow 0$.

Dropping the subscripts on $S^\delta_0$, we now have the situation at the beginning of the proof, but with $a = 1$, that is with a single substitution $\eta$ such that $\eta\sigma M^1, \eta\sigma M^2 \Rightarrow 0$. From the set of all terms that occur in $\eta\sigma M^\delta$ obtain a set $B^\delta$ by deleting successively any term that is expressible, by means of the operation symbols of $L$, in terms of the rest. Since each $\eta\sigma y^\delta_r = s^\delta_r(\eta x^\delta_{11}, \cdots, \eta x^\delta_{rm_r})$ where $s^\delta_r$ does not belong to $L$, we can suppose that all the $\eta\sigma y_r^\delta$ belong to $B^\delta$. Let $b_1^\delta, \cdots, b_a^\delta$ be the remaining elements of $B^\delta$. Then for each $x_{pq}^\delta$ (that occurs in $M^\delta$) $\eta x_{pq}^\delta$ is expressible in terms of the $\eta\sigma y_r^\delta$ and $b_h^\delta$. More precisely, if $u_1^\delta, \cdots, u_a^\delta$ are new and distinct variables, and $\tau$ a substitution such that $\tau y_r^\delta = \eta\sigma y_r^\delta$, $\tau u_h^\delta = b_h^\delta$. then there exists in $L$ a term $\theta x_{pq}^\delta$ in the variables $y_r^\delta$ and $b_h^\delta$ such that $\tau\theta x_{pq}^\delta = \eta x_{pq}^\delta$. We extend $\theta$ to a substitution by setting $\theta z = z$ for all $z$ other than the $x_{pq}^1, x_{pq}^2$.

Let $S_0^\delta$ be the sentence obtained from $\theta M^\delta$ by prefixing the quantifiers $\forall u_h^\delta$ and $\exists y_r^\delta$ in an order such that if $z$ and $z'$ are two of these variables, and $\tau z$ is shorter than $\tau z'$, then $z$ precedes $z'$ in $S_0^\delta$. To verify that $S_0^\delta$ arises from $S^\delta$ by specialization, we observe that, for (iii), if $y_r^\delta$ occurs in $\theta x_{pq}$ then $\tau y_r^\delta = \eta\sigma y_r^\delta$ is a proper part of $\tau\theta x_{pq}^\delta = \eta x_{pq}^\delta$ whence $r < p$; and, for (iv), if $z$ is any $y_s^\delta$ or $u_h^\delta$ and $z$ occurs in $\theta x_{pq}^\delta$ for $p \leqq r$, then $\tau z$ is a part of $\eta x_{pq}^\delta$ which is in turn a proper part of $\eta\sigma y_r^\delta = \tau y_r^\delta$, whence $z$ precedes $y_r^\delta$ in $S_0^\delta$.

Let $S_0^\delta$ have Skolem matrix $\sigma_0\theta M^\delta$, where $\sigma_0 z_{pq} = z$ for all variables $z$ other than the $y_r^\delta$ and $\sigma_0 y_r^\delta = s_{0r}^\delta(\cdots, u_h^\delta, \cdots)$, the arguments ranging in order over all $u_h^\delta$ that precede $y_r^\delta$ in $S_0^\delta$. From $\eta\sigma M^1, \eta\sigma M^2 \Rightarrow 0$ it remains to construct $\eta_0$, one-to-one on the atomic formulas of $\sigma_0\theta M^1, \sigma_0\theta M^2$, such that $\eta_0\sigma_0\theta M^1, \eta_0\sigma_0\theta M^2 \Rightarrow 0$. For this define a transformation $\chi$ on terms as follows:

(1) $\chi z = z$ for a variable $z$;
(2) $\chi\theta\tau\sigma y_r^\delta = \chi\tau\sigma_0 y_r^\delta$;
(3) for any term $t = w(t_1, \cdots, t_{\rho(w)})$ not of the form $\tau\theta\sigma y_r^\delta$, $\chi t = w(\chi t_1, \cdots, \chi t_{\rho(w)})$.

As in an earlier situation, this definition is legitimate, and the restriction $\eta_0$ of $\chi\tau$ to the language $L_0$ obtained from $L$ by adjoining the symbols $s_{0r}^\delta$ is a substitution. As before we conclude from $\eta\sigma M^1, \eta\sigma M^2 \Rightarrow 0$ that $\chi\tau\theta\sigma M^1, \chi\tau\theta\sigma M^2 \Rightarrow 0$,
Now

$$\chi\tau\theta\sigma y_r^\delta = \chi\tau\sigma_0 y_r^\delta = \chi\tau\sigma_0\theta y_r^\delta = \eta_0\sigma_0\theta y_r^\delta ,$$

and

$$\chi\tau\theta\sigma x_{pq}^\delta = \chi\tau\theta x_{pq}^\delta = \chi\tau\theta\sigma_0 x_{pq}^\delta = \chi\tau\sigma_0\theta x_{pq}^\delta = \eta_0\sigma_0\theta x_{pq}^\delta .$$

It follows that $\chi\tau\theta\sigma M^\delta = \eta_0\sigma_0\theta M^\delta$, whence

$$\eta_0\sigma_0\theta M^1, \eta_0\sigma_0\theta M^2 \Rightarrow 0 .$$

It remains to show that $\eta_0\sigma_0 = \chi\tau\sigma_0$ is one-to-one on the terms of each $\theta M^\delta$. We show first that $\tau\theta\sigma$ is one-to-one on such terms. These terms are terms in the variables $u_h^\delta$ and $y_r^\delta$, containing only the operation symbols of $L$. Note that $\tau\theta\sigma u_h^\delta = \tau\theta u_h^\delta = \tau u_h^\delta = b_h^\delta$ and $\tau\theta\sigma y_r^\delta = \eta\sigma y_r^\delta$.

From the construction of $B$, it follows that, for two such terms $t$ and $t'$, $\tau\theta\sigma t = \tau\theta\sigma t'$ cannot hold for one of $t, t'$ a variable unless $t = t'$. Suppose now that $t = w(t_1, \cdots, t_{\rho(w)})$ and $t' = w'(t'_1, \cdots, t'_{\rho(w')})$. Comparing the first symbols we conclude from $\tau\theta\sigma t = \tau\theta\sigma t'$ that $w = w'$, and the arguments agree :

$$\tau\theta\sigma t_1 = \tau\theta\sigma t'_i \qquad i=1, \cdots, \rho(w)=\rho(w') .$$

By induction on the length of the shorter of $t, t'$ we conclude that each $t_i = t'_i$, whence $t = t'$.

Finally, $\chi\tau\sigma_0 y^\delta_r = \chi\tau\theta\sigma y^\delta_r$ by definition, and $\chi\tau\sigma_0 u^\delta_h = \chi\tau u^\delta_h = \chi\tau\theta\sigma u^\delta_h$. Hence $\chi\tau\sigma_0 = \chi\tau\theta\sigma$ on terms of $\theta M^\delta$. But $\chi$ is evidently one-to-one on terms that do not contain the symbols $s^\delta_{0r}$. Hence, for terms $t$ and $t'$ of $\theta M^\delta$, $\sigma\tau\sigma_0 t = \chi\tau\sigma_0 t^1$ implies $\chi\tau\theta\sigma t = \chi\tau\theta\sigma t'$, hence $\tau\theta\sigma t = \tau\theta\sigma t'$, and, by the property of $\tau\theta\sigma$ established above, $t = t'$. This completes the proof of the Main Theorem.

5. **The Interpolation theorem.** *Let $S$ and $T$ be sentences of a language $L$ such that $S \Rightarrow T$. Then there exists a sentence $S^0$ of the language $L$ such that $S \Rightarrow S^0$, $S^0 \Rightarrow T$, and that a relation symbol occurs positively in $S^0$ only if it occurs positively in both $S$ and $T$, and occurs negatively in $S^0$ only if it occurs negatively in both $S$ and $T$.*

*Proof.* $S$ is equivalent to a prenex sentence $S^1$ such that a relation symbol occurs positively (negatively) in $S^1$ only if it occurs positively (negatively) in $S$. And $\sim T$ is equivalent to a prenex sentence $S^2$ such that a relation symbol occurs positively (negatively) in $S^2$ only if it occurs negatively (positively) in $T$. Since $S^1, S^2 \Rightarrow 0$, by the Main Theorem there exist prenex sentencs $U^1$ and $U^2$ such that $S^1 \Rightarrow U^1$, $S^2 \Rightarrow U^2$, that $U^1$ contains the same kinds of occurrences of relation symbols as $S^1$ and $U^2$ as $S^2$, and that $\eta\sigma M^1, \eta\sigma M^2 \Rightarrow 0$ where $\sigma M^1, \sigma M^2$ are the Skolem matrices of $U^1, U^2$, and $\eta$ is a substitution that is one-to-one on the atomic formulas of each of $\sigma M^1, \sigma M^2$. All this is not altered if we modify $U^1$, $U^2$ by reducing $M^1, M^2$ to normal form.

It will suffice to find $S^0$ such that $U^1 \Rightarrow S^0$, and $S^0, U^2 \Rightarrow 0$, and a relation symbol occurs positively (negatively) in $S^0$ only if it occurs positively (negatively) in $U^1$ and negatively (positively) in $U^2$. Write $M^\delta = \bigvee M^\delta_i$, each $M^\delta_i = \bigwedge M^\delta_{ij}$, and each $M^\delta_{ij}$ either $A^\delta_{ij}$ or $\sim A^\delta_{ij}$ where $A^\delta_{ij}$ is an atomic formula. Define $M^0 = \bigvee M^0_i$ where $M^0_i = 0$ if $M^1_i \Rightarrow 0$, and otherwise $M^0_i$ results from $M^1_i$ by deleting all $M^1_{ij}$ such that $\sim\eta\sigma M^1_{ij}$ is not equivalent to some $\eta\sigma M^2_{hk}$. Let $S^0$ be the sentence obtained from $U^1$ by replacing its matrix $M^1$ by the matrix $M^0$. It is immediate that the occurences of relation symbols in $S^0$ are related to those in $U^1$ and $U^2$ in the required manner. Moreover. since $M^1 \Rightarrow M^0$ is immediate, it follows easily that $U^1 \Rightarrow S^0$.

It remains to show that $S^0, U^2 \Rightarrow 0$, and for this it will suffice to to show that $\eta\sigma M^0, \eta\sigma M^2 \Rightarrow 0$. Since $\eta\sigma M^1 \wedge \eta\sigma M^2 \Rightarrow 0$, then for all $i, h$, $\eta\sigma M^1_i \wedge \eta\sigma M^2_h \Rightarrow 0$. We want to conclude that for all $i, h$, $\eta\sigma M^0_i \wedge \eta\sigma M^2_h \Rightarrow 0$. Since $\sigma$ is clearly one-to-one on the terms of $M^1$, so is $\eta\sigma$, and $\eta\sigma M^1_i \Rightarrow 0$ implies $M^1_i \Rightarrow 0$, whence by definition $M^0_i = 0$, hence $\eta\sigma M^0_i = 0$ and the conclusion follows. If $\eta\sigma M^2_h \Rightarrow 0$ the conclusion is immediate. In the remaining case there exist $j$ and $k$ such that $\sim\eta\sigma M^1_{ij} \Leftrightarrow \eta\sigma M^2_{hk}$. But then, by definition, $M^0_i$ still contains the conjunct $M^1_{ij}$, and again $\eta\sigma M^0_i \wedge \eta\sigma M^2_h \Rightarrow 0$. Since $\eta\sigma M^0_i \wedge \eta\sigma M^2_h \Rightarrow 0$ for all $i, h$, it follows that $\eta\sigma M^0 \wedge \eta\sigma M^2 \Rightarrow 0$, completing the proof.

It was stated in the introduction that the Interpolation Theorem remains true for the predicate calculus with identity. Precisely, we restrict the definition of a language to apply only to those that contain a fixed relation symbol $e$ of rank two, and the definition of interpretation to admit only those $\mu$ for which $\mu e$ is the identity relation on the domain of $\mu$. The relation $S \Rightarrow T$ then acquires a stronger meaning. Nonetheless, the Interpolation Theorem as stated remains true in this new sense. (It may be well to note that $e$ is included among the relation symbols mentioned in the conclusion of the theorem.) In fact, all statements in this paper remain true in the new sense, apart from two modifications. First, Proposition 1 must be modified by enlarging $J$ to contain (the coset of) each formula $e(t, t)$, $t$ a term, and to contain any formula $F'$ obtainable from a formula $F$ in $J$ by replacing an occurrence of a term $t$ by a new term $t'$, provided that $e(t, t')$ is in $J$. Second, in the proof of the Interpolation Theorem, the $M^\delta_i$ as described above must be similarly enlarged by adjoining to each the finite set of all $M^\delta_{ij}$ of the form $A$ or $\sim A$, $A$ atomic, such that $M^\delta_i \Rightarrow M^\delta_{ij}$ in the present sense.

The Interpolation Theorem can be refined in other ways. Conditions can be imposed on the internal structure of the atomic formulas $r(t_1, \cdots, t_{\rho(r)})$ containing the relation symbol $r$. For example, define an $I$-occurrence of $r$ in $S$ to be one in which each $t_i$, for $i \in I \subseteq \{1, \cdots \rho(r)\}$, is a variable universally quantified in $S$. Then it can be required that $r$ have $I$-occurrences is $S^0$ only if it has $I'$-occurrences in $S$ and $I''$-occurrences in $T$, where $I'' \subseteq I \subseteq I'$. Alternatively, stronger conditions can be imposed on the external context in which a relation symbol occurs. For example, suppose all positive occurrences in $S$ of a relation symbol $r$ are in formulas $A' \supset A$ where $A$ and $A'$ are atomic formulas, and that none of the relation symbols appearing in the parts $A'$ of these formulas have positive occurrences in $S$, except possibly in parts $A$; then $S^0$ can be required to contain no positive occurrences of $r$. Such refinements of the Interpolation Theorem have proved useful in the study of homomorphisms and subdirect products of models, but because of their special nature it does not seem worthwhile to give separately formal statements and proofs of these results.

## REFERENCES

1. E. W. Beth, *On Padoa's method in the theory of definition*, Indag, Math. **15** (1953), 66.
2. ———, *The Foundations of Mathematics*, North Holland, Amsterdam, 1958.
3. W. Craig, *Linear reasoning. A new form of the Herbrand-Gentzen theorem*, J. Symbolic Logic **22** (1957), 250.
4. ———, *Three uses of the Herbrand-Gentzen theorem in relating model theory and proof theory*, J. Symbolic Logic **22** (1957), 269.
5. G. Gentzen, *Untersuchungen über das logische Schliessen*, Math. Z. **39** (1934-35), 176, 405.
6. K. Gödel, *Die Vollständigkeit der Axiome des logischen Funktionenkalküls*, Monatsh. Math. Phys. **37** (1930), 349.
7. L. Henkin, *The completeness of the first order functional calculus*, J. Symbolic Logic **14** (1949), 159.
8. J. Herbrand, *Recherches sur la théorie de la démonstration*, Travaux Soc. Sci. et Let. Varsovie, Cl. III, No. 33, 1930.
9. S. C. Kleene, *Introduction to Metamathematics*, Van Nostrand, New York, 1952.
10. A. Robinson, *A result on consistency and its application to the theory of definition*, Indag. Math. **18** (1956), 47.
11. A Tarski, *Some notions and methods on the borderline of algebra and metamathematics*, Proc. International Congress of Math. **1**, Cambridge, 1950, 705.
12. A. Tarski, *Contributions to the theory of models, I, II, III*, Indag. Math. **16** (1954) 572, 582; **17** (1955), 55.

UNIVERSITY OF MICHIGAN

# PROPERTIES PRESERVED UNDER HOMOMORPHISM

ROGER C. LYNDON

**1. Introduction.** The main result of this paper is a characterization of those sentences of the predicate calculus whose validity is preserved under passage from an abstract algebraic system to any homomorphic image of the system. An algebraic system is here construed to be a set together with certain operations and relations, including identity, defined for elements of the set. The sentences under consideration will contain symbols for these operations and relations, and variables whose range is the set of elements of the system, together with the usual logical symbols, but will contain no variables whose range consists of sets, relations, or functions. Such a sentence will be called *positive* if it contains the logical symbols for conjunction, disjunction and quantification only, but not the symbol for negation. It will be shown that:

(*) *A sentence of the predicate calculus is preserved under homomorphism if and only if it is equivalent to a positive sentence.*

An example is provided by the usual statement of the commutative law for multiplicative systems:

$$\forall xy \cdot xy = yx .$$

This is a positive sentence, and indeed every homomorphic image of a commutative system in commutative. As a second example, upon eliminating the symbol for "if ··· then", the left cancellation law takes the form

$$\forall xyz \cdot \sim(xy = xz) \vee y = z .$$

This sentence is not positive, and, indeed, from the fact that the left cancellation property is not preserved under homomorphism we conclude that it is not expressible by any positive sentence.

It is not difficult to show that every sentence equivalent to a positive sentence is preserved under homomorphism; although the converse seems nearly as obvious intuitively, to prove the converse appears to be a matter of considerable difficulty. That positive sentences are preserved was noted by the author [6], and also by E. Marczewski [9], who raised the question of the converse. A proof, by methods quite different from those used here, was announced by J. Łoś [5], but such a proof has not been published. The result has also been stated by A. I. Malcev [8], who appears to indicate a method of proof.

Received October 22, 1958. Work supported in part under grants from the National Science Foundation.

The central result of this paper, Theorem 5, is in fact a stronger form of the assertion (*) above. Some consequences and variants of this theorem are given, and examples to show that it can not be strengthened further in certain obvious ways.

The content of this paper lies within the theory of elementary classes as formulated by A. Tarski [12, 13][1]. We define and use here numerous concepts due to him, and, in particular, that of elementary extension, due to Tarski and R. Vaught [15]. We have tried to make this paper self contained to the extent that the main line of reasoning should be intelligible and convincing under any reasonable interpretation of the concepts employed; for the technical definitions necessary for rigor in the details of the proofs, we refer to an earlier paper [7]. Further, we borrow from that paper the relevant definitions and a proof for the following theorem, which is the cornerstone of the present paper:

INTERPOLATION THEOREM. *If $S$ and $T$ are sentences of the predicate calculus, and $S$ implies $T$, then there exists a sentence $M$ such that $S$ implies $M$ and $M$ implies $T$, and that a relation symbol occurs positively (negatively) in $M$ only if it occurs positively (negatively) in both $S$ and $T$.*

The author has profited from many discussions with L. Henkin and A. Tarski.[2] The relativization embodied in Theorem 5′ was suggested by A. Robinson.[3]

**2. Sentences increasing in a relation symbol.** Roughly, a property of a relation may be called increasing if, whenever it holds for a given relation it holds for any larger relation. Passing from properties to the sentences that express them, we make a precise definition. Let $Q$ be a subset of the set $R$ of all relation symbols in a language $L$, and let $Q'$ be a set of new and distinct relations symbols $q'$ in one-to-one correspondence with the symbols $q$ of $Q$ in such a way that $q'$ has the same rank as $q$. Let $I$ be the set of all sentences

$$I(q, q') = \forall x_1 \cdots x_{\rho(q)} \cdot q(x_1, \cdots, x_{\rho(q)}) \supset q'(x_1, \cdots, x_{\rho(q)}) \ ,$$

for all $q$ in $Q$. Let $\Gamma$ be a set of formulas of $L$, and $\Gamma'$ the result of replacing the symbols $q$ in $\Gamma$ by the corresponding $q'$. We call $\Gamma$ *increasing in $Q$* if $\Gamma, I \Rightarrow \Gamma'$.

[1] We use the word 'elementary' in preference to 'arithmetical', and, by an 'elementary class', mean always what is commonly called an 'arithmetical class in the wider sense ($AC_\Delta$)'.

[2] In particular, while the author was visiting at the University of California, Berkeley.

[3] At the American Mathematical Society Summer Institute, Ithaca, 1957.

PROPOSITION 1. *If a set $\Gamma$ of formulas is positive in all the relation symbols in a set $Q$, then $\Gamma$ is increasing in $Q$.*

*Proof.* It suffices to treat the case that $\Gamma$ consists of a single formula $F$. If $F$ is an atomic formula or, vacuously, the negation of an atomic formula, the conclusion is immediate. The general case follows by an obvious induction.

The converse is contained in the following.

PROPOSITION 2. *Let $L$, $Q$, $Q'$ and $I$ be as before. Let $\Sigma$, $\Gamma$, $\Delta$ be sets of sentences $L$, and let $\Sigma'$ result from $\Sigma$, and $\Delta'$ from $\Delta$, by replacing each $q$ by the corresponding $q'$. If $\Sigma$, $\Sigma'$, $\Gamma$, $I \Rightarrow \Delta'$, then there exists a set $\Pi$ of sentences $P$, positive in all the symbols of $Q$ and not containing the symbols of $Q'$, such that $\Sigma$, $\Gamma \Rightarrow \Pi$ and $\Sigma$, $\Pi \Rightarrow \Delta$.*

*Proof.* It suffices to treat the case that $\Delta$ consists of a single sentence $D$. By the Compactness Theorem (Corollary 4.1 of [1]), the hypothesis will hold with $\Sigma$, $\Gamma$, $I$ replaced by finite subsets, and hence, taking conjunctions, by single sentences: $S, S', C, J \Rightarrow D'$, where $C$ is positive in all the $q$ in $Q$, and $J$ is a conjunction of sentences $I(q, q')$. It follows directly that $S, C \Rightarrow J \wedge S' \supset D'$. The symbols $q'$ do not occur at all in $S$ or $C$. The symbols $q$ occur only in the part $J$ of $J \wedge S' \supset D'$, and since each occurrence of a symbol $q$ is negative in $J$, it is positive in $J \wedge S' \supset D'$. By the Interpolation Theorem there exists a sentence $P$, not containing the $q'$ and positive in the $q$, such that $S, C \Rightarrow P$ and $P \Rightarrow J \wedge S' \supset D'$. From $S, C \Rightarrow P$ we have $\Sigma, \Gamma \Rightarrow P$. From $P \Rightarrow J \wedge S' \supset D'$, replacing each $q'$ by $q$, it follows that $P \Rightarrow J^* \wedge S \supset D$ where $J^*$ is the result of replacing each $q'$ by $q$ in $J$. In fact, $J^*$ is a theorem, whence $P \Rightarrow S \supset D$, hence $P, S \Rightarrow D$, and $\Sigma, P \Rightarrow D$.

COROLLARY 2.1. *A set $\Gamma$ of sentences is increasing in the symbols of $Q$ if and only if it is equivalent to a set $\Pi$ of sentences positive in the symbols of $Q$.*

3. **Q-maps.** If $\Gamma$ is a set of sentences of the language $L$, let $\Gamma^*$ be the set of all models of $L$ in which all sentences of $\Gamma$ hold. If $K$ is a set of models of $L$, let $K^*$ be the set of all sentences of $L$ that hold in all models in $K$. It follows that $\Gamma^{**}$ is the 'logical closure' of $\Gamma$, the set of all sentences $S$ such that $\Gamma \Rightarrow S$. The *elementary closure* of $K$ is $K^{**}$, and $K$ is an *elementary class* if $K = K^{**}$, that is, if $K = \Gamma^*$ for any $\Gamma$. Two models $\mathfrak{A}$ and $\mathfrak{B}$ are *elementarily equivalent* if $\mathfrak{A}^* = \mathfrak{B}^*$, that is, if exactly the same sentences hold in $\mathfrak{A}$ as in $\mathfrak{B}$.

A model $\mathfrak{A}$ is *submodel* of a model $\mathfrak{B}$ if the domain $A$ of $\mathfrak{A}$ is a subset of the domain $B$ of $\mathfrak{B}$ and if each $\mathfrak{A}w$, $\mathfrak{A}r$ is the restriction of the corresponding $\mathfrak{B}w$, $\mathfrak{B}r$ to the subset $A$ of $B$. If $\mu$ is any interpretation in $\mathfrak{A}$, there is a unique interpretation $\lambda$ in $\mathfrak{B}$ such that $\mu$ and $\lambda$ agree on all variables of $L$. $\mathfrak{B}$ is an *elementary extension* of $\mathfrak{A}$ if for all $\mu, \lambda$ as above, and $F$ a formula of $L$, if $F$ holds in $\mu$ then $F$ holds in $\lambda$. In particular, $\mathfrak{A}$ and $\mathfrak{B}$ are elementarily equivalent.

If $\mathfrak{A}$ and $\mathfrak{B}$ are models with domains $A$ and $B$, a map $\theta$ of $A$ onto $B$ will be called a *Q-map*, for $Q \subseteq R$ a set of relation symbols, if, first of all, $\theta$ 'preserves' all operations and relations:

$$\theta[(\mathfrak{A}w)(a_1, \cdots, a_{\rho(w)})] = (\mathfrak{B}w)(\theta a_1, \cdots, \theta a_{\rho(w)}) , \quad \text{all } w \text{ in } W ,$$
$$(\mathfrak{A}r)(a_1, \cdots, a_{\rho(r)}) \Rightarrow (\mathfrak{B}r)(\theta a_1, \cdots, \theta a_{\rho(r)}) , \quad \text{all } r \text{ in } R ,$$

and, moreover, the implication in the last line is an equivalence for all $r$ not in $Q$. More concisely, $\theta(\mathfrak{A}w) = \mathfrak{B}w$, $\theta(\mathfrak{A}r) \subseteq \mathfrak{B}r$, with $\theta(\mathfrak{A}r) = \mathfrak{B}r$ for $r$ not in $Q$. If $\theta$ is one-to-one, we speak of a *Q-isomorphism*. An O-isomorphism, for O the empty set, is an *isomorphism* in the usual sense.[4]

If $\theta$ is any map of $\mathfrak{A}$ onto $\mathfrak{B}$, its *kernel* $k$, defined by $k(a, a')$ if and only if $\theta a = \theta a'$, is an equivalence relation on $A$. If $\theta$ is a O-map of $\mathfrak{A}$ onto $\mathfrak{B}$, then $k$ is a congruence relation on $\mathfrak{A}$, that is, it is substitutive with respect to all the $\mathfrak{A}w$ and $\mathfrak{A}r$. For any congruence $k$ on a model $\mathfrak{A}$, the operations $\mathfrak{A}w$ and relations $\mathfrak{A}r$ of $\mathfrak{A}$ induce operations $\mathfrak{A}w/k$ and relations $\mathfrak{A}r/k$ on the set $A/k$ of equivalence classes in $A$ under $k$; the *quotient model* $\mathfrak{A}/k$ is defined to have domain $A/k$, operations $(\mathfrak{A}/k)w = \mathfrak{A}w/k$, and relations $(\mathfrak{A}/k)r = \mathfrak{A}r/k$. It is immediate that the natural projection of $A$ onto $A/k$ is a O-map, and that if $\theta$ is any O-map of $\mathfrak{A}$ onto $\mathfrak{B}$, with kernel $k$, then $\theta$ induces a naturally an isomorphism of $\mathfrak{A}/k$ onto $\mathfrak{B}$.

We proceed to the statement of a proposition that contains all that we require about elementary extensions. For $\mathfrak{A}$ a model of the language $L$, with domain $A$, define a language $L_A$ by adjoining to $L$ new and distinct constants (operations of rank 0) $w_a$ for all $a$ in $A$, and a new relation $e_A$ of rank two. Extend $\mathfrak{A}$ to a model $\mathfrak{A}_A$ of $L_A$ by defining $\mathfrak{A}_A w_a = a$, that is, $\mathfrak{A}_A w_a$ is the constant operation with value $a$, and $\mathfrak{A}_A e_A$ to be the identity relation on $A$. Then $\mathfrak{A}_A^*$ is the set of all sentences of $L_A$ that hold in $\mathfrak{A}_A$.

PROPOSITION 3. *If $\mathfrak{A}$ and $\mathfrak{B}$ are models of $L$, and $\mathfrak{B}$ has an extension $\mathfrak{B}'$ to $L_A$ in which all sentences of $\mathfrak{A}_A^*$ hold, then $\mathfrak{B}$ has an O-image that is an elementary extension of $\mathfrak{A}$.*

[4] The concept of $Q$-map and that of elementary extension, as well as various results mentioned here, are special cases of more far-reaching ideas developed recently by H. J. Keisler [**4**]. The use of constants $w_a$, in the paragraph after next, derives from the 'diagrams' of A. Robinson [**11**]. Proposition 3 in contained in Th. 1.11 of [**15**].

*Proof.* Let $k = \mathfrak{B}'e_A$. Since the sentences expressing that $\mathfrak{A}_A e_A$, the identity on $A$, is a congruence on $\mathfrak{A}_A$ are in $\mathfrak{A}_A^*$, they hold in $\mathfrak{B}'$, whence $k$ is a congruence on $\mathfrak{B}'$. The quotient system $\mathfrak{C}' = \mathfrak{B}'/k$ then also satisfies $\mathfrak{A}_A^*$, and $\mathfrak{C}'e_A$ is the identity on the domain $C$ of $\mathfrak{C}'$. The restriction $\mathfrak{C}$ of $\mathfrak{C}'$ to $L$ is $\mathfrak{B}/k$, an O-image of $B$.

Define a map $\theta$ from $A$ into $C$ by setting $\theta a = \mathfrak{C}'w_a$. Now, $\mathfrak{A}_A w(a_1, \cdots, a_{\rho(w)}) = a'$ if and only if $e_A(w(w_{a_1}, \cdots, w_{a_{\rho(w)}}), w_{a'})$ is in $\mathfrak{A}_{A^*}$, hence if and only if this sentence holds in $\mathfrak{C}'$, that is, if

$$\mathfrak{C}'w(\theta a_1, \cdots, \theta a_{\rho(w)}) = \theta a' .$$

The same reasoning shows that $\mathfrak{A}_A r(a_1, \cdots, a_{\rho(r)})$ holds in $\mathfrak{A}$ if and only if $\mathfrak{C}'r(\theta a_1, \cdots, \theta a_{\rho(r)})$ in $\mathfrak{C}'$. This establishes that is an O-map of $\mathfrak{A}_A$ onto a subsystem $\theta\mathfrak{A}_A$ of $\mathfrak{C}'$, and, in fact, taking $r$ above to be $e_A$, $a = a'$ if and only if $\theta a = \theta a'$, whence $\theta$ is an isomorphism.

Since $\theta\mathfrak{A}_A$ is a submodel of $\mathfrak{C}'$, taking restrictions to $L$, $\theta\mathfrak{A}$ is a submodel of $\mathfrak{C}$. Let $\mu$ be an interpretation in $\theta\mathfrak{A}$, and $\lambda$ the interpretation in $\mathfrak{C}$ that agrees with $\mu$ on all variables. Let $F$ be a formula of $L$ with free variables $x_1, \cdots, x_n$, and $D$ the sentence that results from $F$ by replacing each $x_i$ by $w_{a_i}$ where $\mu x_i = \lambda x_i = \theta a_i$. If $\mu$ and $\lambda$ are extended to $L'$ in such a way that each $\mu w_{a_i} = \lambda w_{a_i} = \theta a_i$, then $\mu F = \mu D$ and $\lambda F = \lambda D$. Now, if $\mu F = 1$, $\mu D = 1$, and, since $D$ is a sentence, $D$ holds in $\theta\mathfrak{A}_A$, hence in $\mathfrak{A}_A$. Then $D$ is in $\mathfrak{A}_{A^*}$ and hence holds in $\mathfrak{C}'$, whence $\lambda D = 1$ and $\lambda F = 1$. This establishes that $\mathfrak{C}$ is an elementary extension of $\theta_A$.

It is now a trivial matter to construct $\mathfrak{D}$ from $\mathfrak{C}$ by replacing each element $\theta a$ in $\mathfrak{C}$ by $a$. Then $\mathfrak{D}$ is an elementary extension of $\mathfrak{A}$ itself, and the O-map of $\mathfrak{C}$ onto $\mathfrak{B}$ induces an O-map of $\mathfrak{D}$ onto $\mathfrak{B}$.

We come now to the main result concerning $Q$-maps.

THEOREM 4. *Let $\mathfrak{A}$ be a model of the language L, and K an elementary class of models of L. Then the following are equivalent:*

(1) *all Q-positive sentences of L that hold in K also hold in $\mathfrak{A}$;*

(2) *some elementary extension of $\mathfrak{A}$ is a Q-image of a model in K.*

*Proof.* Assume (1). Let $\Gamma = K^*$ and $\Delta = \mathfrak{A}_A^*$. Let $Q$, $Q'$, and $I$ be as before. Let $\Delta'$ result from $\Delta$ by replacing each relation symbol $q$ in $Q$ by the corresponding $q'$ in $Q'$. Suppose $\Gamma$, $I$, $\Delta'$ inconsistent. By the Compactness Theorem, $\Gamma$, $I$, $D' \Rightarrow 0$ where $D$ is a finite conjunction of sentences from $\Delta$, hence itself belongs to $\Delta$. Then $\Gamma, I \Rightarrow \sim D'$, and, by Proposition 2, and Compactness, there exists a $Q$-positive sentence $P$, not containing the symbols $q'$, $w_a$, $e_A$, that is, in $L$, such that $L \Rightarrow P$ and $P \Rightarrow \sim D$. But $\Gamma \Rightarrow P$ implies that $P$ holds in $K$, and, since $P$ is a $Q$-positive sentence of $L$, that $P$ holds in $\mathfrak{A}$. Therefore $P$ holds in $\mathfrak{A}_A$, and $P \Rightarrow \sim D$ gives a contradiction.

It has been shown that $\Gamma$, $I$, $\Delta'$ is consistent, hence holds for some model $\mathfrak{C}$ of the language $L'$ obtained from $L$ by adjoining the symbols $q'$, $w_a$, $e_A$. Let $\mathfrak{D}$ be the restriction of $\mathfrak{C}$ to the language $L'_A$ excluding the symbols $q$; since $\mathfrak{C}$ satisfies $\Delta'$, so does $\mathfrak{D}$. Define a model $\mathfrak{B}'$ of $L_A$ to agree with $\mathfrak{D}$ except that $\mathfrak{B}'q = \mathfrak{D}q'$; then $\mathfrak{B}'$ satisfies $\Delta$. By Proposition 3, some 0-image $\mathfrak{B}^*$ of the restriction $\mathfrak{B}$ of $\mathfrak{B}'$ to $L$ is an elementary extension of $\mathfrak{A}$.

Let $\mathfrak{E}$ be the restriction of $\mathfrak{C}$ to $L$; since $\mathfrak{C}$ satisfies $\Gamma$, so does $\mathfrak{E}$, and $\mathfrak{E}$ is in $K$. Now $\mathfrak{E}w = \mathfrak{C}w = \mathfrak{B}w$ for all $w$ in $W$, and $\mathfrak{E}r = \mathfrak{C}r = \mathfrak{B}r$ for all $r$ not in $Q$, while, for $q$ in $Q$, $\mathfrak{E}q = \mathfrak{C}q$ while $\mathfrak{B}q = \mathfrak{D}q' = \mathfrak{C}q'$, and, since $\mathfrak{C}$ satisfies the seetences $I$, $\mathfrak{E}q \subseteqq \mathfrak{B}q$. It follows that the identity map $\theta$ on the common domain $C$ of $\mathfrak{E}$ and $\mathfrak{B}$ is a $Q$-map of $\mathfrak{E}$ onto $\mathfrak{B}$. It follows that the 0-image $\mathfrak{B}^*$ of the $Q$-image $\mathfrak{B}$ of $\mathfrak{E}$ is a $Q$-image of $\mathfrak{E}$: the elementary extension $\mathfrak{B}^*$ of $\mathfrak{A}$ is the $Q$-image of $\mathfrak{E}$ in $K$.

To show that (2) implies (1), it suffices to show that if $\mathfrak{A}$ is a $Q$-image of some $\mathfrak{B}$ in $K$, and $P$ in $\Gamma$ is $Q$-positive, then $P$ holds in $\mathfrak{A}$. Define a model $\mathfrak{C}$ of the language $L'$, obtained from $L$ by adjoining the symbols $q'$, by taking as domain the common domain $A$ of $\mathfrak{A}$ and $\mathfrak{B}$; setting $\mathfrak{C}w$ equal to the common value $\mathfrak{A}w = \mathfrak{B}w$; for $r$ not in $Q$, setting $\mathfrak{C}r = \mathfrak{A}r = \mathfrak{B}r$; and defining $\mathfrak{C}q = \mathfrak{B}q$, $\mathfrak{C}q' = \mathfrak{A}q$. Since $\mathfrak{B}$ is in $K$, $\mathfrak{B}$ satisfies $P$ and so does $\mathfrak{C}$. Since $\mathfrak{A}$ is a $Q$-image of $\mathfrak{B}$, each $\mathfrak{B}q \subseteqq \mathfrak{A}q$, that is, each $\mathfrak{C}q \subseteqq \mathfrak{C}q'$, whence $\mathfrak{C}$ satisfies the sentences $I$. Since $P$ is $Q$-positive, it follows by Proposition 1 that $P, I \Rightarrow P'$, whence $P'$ holds in $\mathfrak{C}$, and, since $\mathfrak{C}P' = \mathfrak{A}P$, $P$ holds in $\mathfrak{A}$.

COROLLARY 4.1. *An elementary class $K$ is closed under $Q$-maps if and only if it is the set of all models for some set of $Q$-positive sentences.*

*Proof.* Assume $K$ closed under $Q$-maps. Let $K = \Gamma^*$, and let $\Pi$ be the set of all $Q$-positive consequences of $\Gamma$. If $\mathfrak{A}$ is in $\Pi^*$, some elementary extension $\mathfrak{B}$ of $\mathfrak{A}$ is a $Q$-image of a model $\mathfrak{C}$ in $K$. But then $\mathfrak{B}$ and therefore $\mathfrak{A}$ are in $K$. Thus $\Pi^* \subseteqq \Gamma^*$; since $\Pi \subseteqq \Gamma$ implies $\Gamma^* \subseteqq \Pi^*$, $K = \Gamma^* = \Pi^*$. The converse is immediate.

COROLLARY 4.2. *A set of sentences is preserved under $Q$-maps if and only if it is equivalent to a set of $Q$-positive sentences.*

**4. The Main Theorem.** We now choose once and for all a relation symbol $e$ of rank two, and consider henceforth only languages $L$ that contain this symbol. A model $\mathfrak{A}$ of $L$ will be called a *relational system* provided $\mathfrak{A}e$ is the identity relation on the domain $A$ of $\mathfrak{A}$. We shall speak of the set of all relational systems in an elementary class as an *elementary class of relational systems.*

The term *homomorphism* will be taken in the broad sense, for a map that preserves all functions and relations, that is, an $R$-map. The term *projection* will be used for the narrower concept of 0-map: $\mathfrak{B}$ is (the image under) a projection of $\mathfrak{A}$ if and only if $\mathfrak{B}$ is isomorphic to a quotient system of $\mathfrak{A}$. The other component of the concept of homomorphism in contained in that of *enlargement*, or $R$-isomophism: $\mathfrak{B}$ is (the image under) an enlargement of $\mathfrak{A}$ if and only if $\mathfrak{B}$ is isomorphic to a system obtained from $\mathfrak{A}$ by replacing its relations by more extensive relations. It is easily seen that if $\theta$ is any homomorphism of $\mathfrak{A}$ onto $\mathfrak{B}$, then $\mathfrak{A}$ has an enlargement $\mathfrak{A}'$, defined by taking $\mathfrak{A}'r = \theta^{-1}\mathfrak{B}r$ for all $r$ except, $e$, such that $\theta$ induces a projection of $\mathfrak{A}'$ outo $\mathfrak{B}$.

THEOREM 5. *Let $\mathfrak{A}$ be a relational system of the language L, and K an elementary class of systems of L. Then the following are equivalent:*

(1) $\mathfrak{A}$ *satisfies all sentences of L that hold in K and are*
$$\left\{\begin{array}{l}\textit{positive in all relation symbols}\\ \textit{positive in the symbol } e\\ \textit{positive in all relation symbols except } e\end{array}\right\};$$
(2) $\mathfrak{A}$ *has an elementary extension that is*
$$\left\{\begin{array}{l}\textit{a homomorphic image}\\ \textit{a projection}\\ \textit{an enlargement}\end{array}\right\} \textit{of a system in } K.$$

*Proof.* Let $Q_1 = R$, $Q_2 = \{e\}$, $Q_3 = R - \{e\}$.

If $\mathfrak{A}$ satisfies (2) it is a $Q_i$-image of a system in $K \subseteqq K^{**}$, and hence, by Theorem 4, $\mathfrak{A}$ satisfies all $Q_i$-positive sentences in $K^{***} = K^*$.

For the converse, suppose that $\mathfrak{A}$ is a relational system that satisfies all the $Q_i$-positive sentences in $K^*$. By Theorem 4, there exists a model $\mathfrak{C}$ (not necessarily a relational system) in $K^{**}$ and a $Q_i$-map $\theta$ of $\mathfrak{C}$ onto a model $\mathfrak{B}$ that is an elementary extension of $\mathfrak{A}$. Since $K$ is a class of relational systems, $K^*$ contains sentences requiring that $e$ be interpreted as a congruence, whence $\mathfrak{C}e$ is a congruence on $\mathfrak{C}$. Since $\mathfrak{A}$ is a relational system, $\mathfrak{A}e$ is a congruence, and, indeed, the identity on the domain $A$ of $\mathfrak{A}$. Since $\mathfrak{B}$ is an elementary extension of $\mathfrak{A}$, hence elementarily equivalent to $\mathfrak{A}$, $\mathfrak{B}e$ is a congruence on $\mathfrak{B}$, and its restriction to $A$ is the identity on $A$. It follows that $\mathfrak{B}/\mathfrak{B}e$ is an elementary extension of $\mathfrak{A}$.

The map $\theta$ induces a $Q_i$-isomorphism $\bar{\theta}$ of $\mathfrak{C}/\mathfrak{C}e$ onto $\mathfrak{B}/\mathfrak{C}e$. Since $\mathfrak{C}e \subseteqq \mathfrak{B}e$, there is a canonical projection $\kappa$ of $\mathfrak{B}/\mathfrak{C}e$ onto $\mathfrak{B}/\mathfrak{B}e$. Hence $\kappa\bar{\theta}$ is a $Q_i$-map of the relational system $\mathfrak{C}/\mathfrak{C}e$ onto the relational system $\mathfrak{B}/\mathfrak{B}e$. This completes the case of $Q_1 = R$. For $Q_2 = \{e\}$, $\mathfrak{C}$ and $\mathfrak{B}$ differ only in their values $\mathfrak{C}e$ and $\mathfrak{B}e$, whence $\mathfrak{C}/\mathfrak{C}e = \mathfrak{B}/\mathfrak{C}e$, and $\kappa$ is a projection of $\mathfrak{C}/\mathfrak{C}e$ onto $\mathfrak{B}/\mathfrak{B}e$. For $Q_3 = R - \{e\}$, $\mathfrak{C}e = \mathfrak{B}e$, whence $\mathfrak{B}/\mathfrak{C}e = \mathfrak{B}/\mathfrak{B}e$ and $\theta$ is a $Q_3$-isomorphism, that is, an enlargement, from $\mathfrak{C}/\mathfrak{C}e$ onto $\mathfrak{B}/\mathfrak{B}e$.

It would be possible, by the same arguments, to generalize Theorem 5 to $Q$-maps, where $Q \subseteq R$ may contain $e$ or not, and indeed to maps increasing in one set $Q$ of relation symbols and decreasing in a second set $Q'$. But, for simplicity, we shall rather restrict our attention to the entirely typical case of homomorphisms.

COROLLARY 5.1. *Let $K$ be an elementary class of relational systems. A sentence $S$ is true for all homomorphic images of systems in $K$ if and only if $S$ is a consequence of some positive sentence that holds for all systems in $K$.*

*Proof.* If $S$ is a consequence of a positive sentence $P$ that holds for all systems in $K$, it follows by the theorem that $P$, and therefore also $S$, hold for all homomorphic images of systems in $K$. Conversely, if $S$ holds for all homomorphic images of systems in $K$, and hence for all systems having such images as elementary extension, it follows by the theorem that $S$ holds for all systems that satisfy the set $\Pi$ of all positive sentences that hold for every system in $K$. Thus $\Pi \Rightarrow S$, and by the Compactness Theorem $P_1, \cdots, P_n \Rightarrow S$ for some finite set of $P_1, \cdots, P_n$ in $\Pi$, whence $P \Rightarrow S$ for $P = P_1 \wedge \cdots \wedge P_n$ in $\Pi$.

COROLLARY 5.2. *Let $K$ be an elementary class of relational systems. Every homomorphic image of a system in $K$ itself belongs to $K$ if and only if $K$ is the class of all systems satisfying a certain set of positive sentences.*

*Proof.* Let $K = \Gamma^*$, and suppose that $H(K) \subseteq K$, where $H(K)$ is the class of all homomorphic images of systems in $K$. Let $\Pi$ be the set of all positive sentences in $\Gamma$. Since $\Pi \subseteq \Gamma$, it is immediate that $\Gamma^* \subseteq \Pi^*$. By Corollary 5.1, every sentence $S$ in $\Gamma$ is a consequence of some sentence $P$ in $\Pi$, whence $\Gamma^* \subseteq \Pi^*$. It follows that $\Pi^* = \Gamma^* = K$.

COROLLARY 5.3. *A sentence has the property that whenever it holds for a system $\mathfrak{A}$ it holds for every homomorphic image of $\mathfrak{A}$ if and only if it is equivalent to a positive sentence.*

*Proof.* In Corollary 5.2, take $K$ to be the class characterized by a single sentence.

If a relational system $\mathfrak{A}$ satisfies the set of all positive sentences true for a system $\mathfrak{B}$, it follows from the theorem, with $K = \mathfrak{B}^{**}$, that some elementary extension $\mathfrak{A}'$ of $\mathfrak{A}$ is a homomorphic image of a system

$\mathfrak{B}'$ that is elementarily equivalent to $\mathfrak{B}$. In fact, by passing from the originally given language $L$ to the language $L_B$, there is no difficulty in establishing a stronger proposition, as follows: *$\mathfrak{A}$ satisfies all positive sentences true for $\mathfrak{B}$ if and only if $\mathfrak{A}$ has an elementary extension $\mathfrak{A}'$ and $\mathfrak{B}$ an elementary extension $\mathfrak{B}'$ such that $\mathfrak{A}'$ is a homomorphic image of $\mathfrak{B}'$.*

Let $H$ an elementary class of relational systems, and call a system in $H$ an *H-system*. The following relativized version of Theorem 5 is contained directly in that theorem.

THEOREM 5′. *If $\mathfrak{A}$ is an H-system and $K$ an elementary class of H-systems, then the following are equivalent:*

(1) *$\mathfrak{A}$ satisfies all positive sentences that hold in $K$;*

(2) *$\mathfrak{A}$ has an elementary extension that is a homomorphic image of a system in $K$.*

The relativized forms of the corollaries now follow as before, provided the relation $P \Rightarrow T$ is replaced by that of *H-implication*: $H, P \Rightarrow T$, and equivalence by *H-equivalence*. As an example, the relativized version of Corollary 5.3 asserts the equivalence of the following properties of a first order sentence $S$ of group theory:

(1) *if $\mathfrak{A}$ and $\mathfrak{B}$ are torsionfree groups, if $\mathfrak{B}$ is a homomorphic image of $\mathfrak{A}$, and $S$ holds for $\mathfrak{A}$, then $S$ holds for $\mathfrak{B}$;*

(2) *there exists a positive sentence $P$ such that, for each torsionfree group $\mathfrak{A}$, $S$ holds if and only if $P$ holds;*

and hence, further,

(2) *there exists a positive sentence $P$ such that the equivalence of $S$ and $P$ follows from the axioms for torsionfree groups.*

**5. Complementary examples.** We first note that the conclusion of Theorem 5 does not follow without the requirement that the class $K$ be elementary. For this, let $L$ be the language of elementary identity theory, without operation symbols and without relation symbols other than $e$. The relational systems for this language are simply sets. Let $K$ be the class of all finite systems; clearly $H(K) \subseteq K$. It is well known and easily seen that $K^*$ consists only of those sentences that are true in all models. A fortiori, every system, infinite or finite, satisfies all positive sentences that hold for $K$. But an elementary extension of an infinite system is itself infinite, hence cannot belong to $H(K)$.

Next we show that, even if $K$ is elementary, the class $H(K)$ of all homomorphic images of systems in $K$ need not be elementary; hence the reference to elementary extensions in Theorem 5 can not be deleted. For this, let $L$ contain no operation symbols, and only a single binary relation symbol $r$ in addition to $e$. Let $S$ be the following sentence:

$$S = \exists x \exists y \forall z \exists t \colon r(x, y) \wedge \cdot\, r(x, z) \supset r(x, t) \wedge r(z, t) \,.$$

For $n = 1, 2, \cdots$, let $S_n$ be the following sentence:

$$S_n = \exists x_0 x_1 \cdots x_n \cdot r(x_0, x_1) \wedge r(x_0, x_2) \wedge \cdots \wedge r(x_0, x_n) \\ \wedge r(x_1, x_2) \wedge r(x_2, x_3) \wedge \cdots \wedge r(x_{n-1}, x_n) \,.$$

We shall establish the following:

*If* $S \Rightarrow P$, *and* $P$ *is positive, then* $S_n \Rightarrow P$ *for some* $n = 1, 2, \cdots$; *hence, defining* $K = S^*$, $H(K)^* = \{S_1, S_2, \cdots\}^{**}$.

We use the Main Theorem of [7], with $S^1 = S$ and $S^2 = \sim P$. The Skolem matrix $M^1$ of $S$ has the form

$$r(s_0, s_1) \wedge \cdot\, r(s_0, z) \supset r(s_0, s(z)) \wedge r(z, s(z)) \,,$$

where $s_0, s_1$ are Skolem functions of rank 0, and $s$ of rank 1, in an extension $L'$ of $L$. The Skolem matrix $M^2$ of $\sim P$ is negative. If $N^1, N^2$ are the Skolem matrices of $U^1, U^2$, as in the Main Theorem of [7], evidently $N^2$ is negative, whence $N^1$ and hence $U^1$ are positive. We have that $U^1 \Rightarrow P$, and, from the relation of $U^1$ to $S^1$, that $U^1$ follows from a universal sentence with positive matrix $M$, where $M$ follows by propositional calculus alone from a set $\Sigma$ of instances of $M^1$. Define a sequence of terms $t_0, t_1 \cdots$ in $L'$ by setting $t_0 = s_0$, $t_1 = s_1$, and, inductively, $t_{n+1} = s(t_n)$ for all $n \geqq 1$. Define a substitution $\chi$ on the atomic formulas $F$ of $L'$ by setting $\chi F = F$ if $F$ is $r(t_n, t_{n+1})$ or $r(t_0, t_{n+1})$, for some $n = 0, 1, 2, \cdots$, and setting $\chi F = 0$ otherwise. Since $M$ is positive and each $\chi F \Rightarrow F$, $\chi M \Rightarrow M$. Since $\Sigma \Rightarrow M$ by propositional calculus, $\chi\Sigma \Rightarrow \chi M$. Thus $\chi\Sigma \Rightarrow M$. Now $\chi\Sigma$ is evidently equivalent to the set of all formulas $r(t_0, t_{n+1})$ and $r(t_n, t_{n+1})$, whence, by the Compactness Theorem, $M$ is a consequence of a finite set of them, and hence, for some $n$, of

$$r(t_0, t_1) \wedge \cdots \wedge r(t_0, t_n) \wedge r(t_1, t_2) \wedge \cdots \wedge r(t_{n-1}, t_n) \,.$$

But now $U^1$, which follows from $M$, follows equally from the Skolem matrix of $S^n$, hence from $S^n$ itself, and $S^n \Rightarrow P$.

Let $\mathfrak{A}$ be a relational system for $L$. The sentence $S$ evidently requires that the domain $A$ of $\mathfrak{A}$ contain an infinite chain of elements, not necessarily distinct, $a_0, a_1, \cdots$, such that $\mathfrak{A} r(a_0, a_n)$ and $\mathfrak{A} r(a_{n-1}, a_n)$ for all $n \geqq 1$. Since the image of such a chain in any homomorphic image of $\mathfrak{A}$ is again such a chain, every $\mathfrak{A}$ in $H(K)$ contains such a chain. On the other hand, each condition $S_n$ requires of a system $\mathfrak{A}$ that it contain a finite chain $a_0, a_1 \cdots, a_n$ related in this fashion, whence a system $\mathfrak{A}$ is in $H(K)^*$ if it contains such chains of unbounded lengths. If $\mathfrak{A}$ is a system with domain $A = \{a_0, a_1, \cdots\}$ and $\mathfrak{A} r(a_i, a_j)$ is true if and only if $j < i$, evidently $\mathfrak{A}$ contains chains $a_n, a_{n-1}, \cdots, a_0$ for all $n$, but no infinite chain of the kind required by $S$, whence $\mathfrak{A}$ is in $H(K)^*$ but not in $H(K)$.

We conclude by showing that it is not in general decidable whether a sentence of a first order language is equivalent to some positive sentence.[5] A first order *theory* $T$ may be taken as consisting of a language $L$ together with a consistent logically closed set $\Gamma = \Gamma^{**}$ of sentences of $L$, the *theorems* of $T$. The theory $T$ is *undecidable* if there is no effective method of deciding, for all sentences $S$ of $L$, whether $S$ is a theorem of $T$, that is, if the set $\Gamma$ is not recursive; this concept is of interest primarily in the case that there exists a finite, or at least recursive, set $\Gamma_0$ of *axioms*, such that $\Gamma_0^{**} = \Gamma$. We shall confine our attention to finitely axiomatizable undecidable theories that have the following additional property:

(*) every model in which $\Gamma$ holds has as homomorphic image some one-element model in which $\Gamma$ holds.

Two important examples of such theories are the following:

(1) $L$ contains at least one relation symbol (other than the identity symbol) of rank greater than one, $\Gamma$ empty;[6]

(2) $L$ contains the identity symbol and an operation symbol $w$ of rank 2, and $\Gamma$ is a set of axioms for group theory with $e$ interpreted as equality and $w$ as the group composition.[7]

Let $\Sigma$ be the set of all sentences $S$ of $L$ such that $S$ holds in every one-element model in which $\Gamma$ holds. Clearly $\Gamma \subseteq \Sigma$. Moreover, it is easily decidable, for $C$ the conjunction of all axioms in $\Gamma_0$, whether $C \supset S$ holds in all one-element models, and hence whether $S$ is in $\Sigma$. Consequently, it is not decidable whether a sentence in $\Sigma$ is a theorem.

Let $S$ be in $\Sigma$. Suppose first that $S$ is a theorem. Then $\Gamma \Rightarrow S$, whence $\sim S$ is $T$-equivalent to the false sentence 0, which is positive; that is, $\Gamma \Rightarrow \sim S \supset 0 \cdot \wedge \cdot 0 \supset \sim S$. Suppose now that $S$ is not a theorem. Then there exists a model $\mathfrak{A}$ in which $\Gamma$ holds while $S$ fails, and hence $\sim S$ holds. In view of the assumption (*) $\mathfrak{A}$ has as homomorphic image some one-element system $\mathfrak{B}$ in which $\Gamma$ holds. Since $S$ is in $\Sigma$, $S$ holds in $\mathfrak{B}$, that is, $\sim S$ fails in $\mathfrak{B}$. Since $\Gamma$ and $\sim S$ both hold in $\mathfrak{A}$, while $\Gamma$ holds and $\sim S$ fails in the homomorphic image $\mathfrak{B}$ of $\mathfrak{A}$, it follows from Theorem 5′ that $\sim S$ is not $T$-equivalent to any positive sentence. We have shown that, for $S$ in $\Sigma$, $\sim S$ is equivalent to a positive sentence (and, indeed, to the positive sentence 0) if and only if $S$ is a theorem. It follow that there exists no effective method of deciding, for sentences $S$ such that $\sim S$ is in $\Sigma$, nor, therefore, for all sentences of $L$, whether $S$ is $T$-equivalent to a positive sentence.

---

[5] For the main concepts of this paragraph, see [**14**].

[6] See Church [**1**].

[7] For the undecidability of the elementary theory of groups, see [**14**, p. 84] and the reference to Tarski given there; see also [**10**].

## References

1. A. Church, *A note on the Entscheidungsproblem*, J. Symbolic Logic **1** (1936), 40, 101.
2. L. Henkin, *The completeness of the first order functional calculus*, J. Symbolic Logic **14** (1949), 159.
3. A. Horn, *On sentences which are true of direct unions of algebras*, J. Symbolic Logic **16** (1951), 14.
4. H. J. Keisler, J. Symbolic logic, to appear.
5. J. Łoś, *Sur les classes définissables*, p. 107, in *Mathematical Interpretation of Formal Systems*, North Holland, Amsterdam, 1955.
6. R. C. Lyndon, review of [3], J. Symbolic Logic **16** (1951), 216.
7. R. C. Lyndon, *An interpolation theorem in the predicate calculus*, Pacific J. Math., **9** (1959),
8. A. I. Malcev, *Subdirect products of models*, Dokl. Akad. Nauk SSSR (N.S.) **109** (1956), 264.
9. E. Marczewski, *Sur les congruences et les proprietés positives d'algèbres abstraites*, Colloq. Math. **2** (1951), 220.
10. P. S. Novikov, *On the algorithmic unsolvability of the word problem in group theory*, Trudy Math. Inst. Steklov, No. 44, Moscow, 1955.
11. A. Robinson, *On the Metamathematics of Algebra*, North Holland, Amsterdam, 1951.
12. A. Tarski, *Some notions and methods on the borderline of algebra and metamathematics*, Proc. International Congress of Math. **1**, Cambridge, 1950, 705.
13. A. Tarski, *Contributions to the theory of models*, I, II, III, Indag. Math. **16** (1954), 572, 582; **17** (1955), 55.
14. A. Tarski, with A. Mostowski and R. M. Robinson, *Undecidable Theories*, North Holland, Amsterdam, 1953.
15. A. Tarski and R. L. Vaught, *Arithmetical extensions of relational systems*, Comp. Math. **13** (1957), 81.

UNIVERSITY OF MICHIGAN

# PROPERTIES PRESERVED IN SUBDIRECT PRODUCTS

ROGER C. LYNDON

**1. Introduction.** A characterization is obtained of those sentences $S$ of the predicate calculus such that $S$ holds for a subdirect product of general algebraic or relational systems[1] whenever it holds for each component system. We consider formulas in a first order language equipped with symbols for the operations and relations of the systems under consideration, and, in particular, with a symbol for the identity relation. An *atomic formula* is one obtained by inserting terms in the argument places of a relation symbol. A *positive formula* is one that can be built up from atomic formulas by means of conjunction, disjunction, and of universal and existential quantification (but without using negation). A *special Horn formula* is one of the from $P \supset F$ where $P$ is a positive formula and $F$ is an atomic formula, or any formula obtained from such formulas by conjunction and universal quantification. A *sentence* is a formula without free variables. As a corollary to our main theorem we obtain the following:

*A sentence has the property that it holds for a subdirect product of systems whenever it holds for each component system if and only if it is equivalent to a special Horn sentence.*

An example of a special Horn sentence is provided by the condition for an associative ring to be semisimple in the sense of Jacobson [7, Proposition 1, p. 9], which is expressed by the following sentence:

$$\forall z \cdot [\forall x \forall y \exists u \cdot xzy + u = xzyu \wedge uxzy = xzyu] \supset z = 0 .$$

We admit among subdirect products the subdirect product of an empty set of systems, which, from the definition, proves to be a trivial system with a single element and all relations universal. The sole effect of excluding this trivial case would be to admit in special Horn sentences clauses $\sim P$ along with the clauses $P \supset F$.

A. Horn [6] considered the more general class of all sentences obtained by universal and existential quantification from conjunctions of formulas of the type $P \supset F$ (or $\sim P$), where $P$ is a conjunction of atomic formulas and $F$ an atomic formula. Horn showed that all such sentences are preserved under (full) direct products, while C. C. Chang and Anne C. Morel [4] showed that there are sentences preserved under direct product that are not equivalent to any such Horn sentence. The problem of characterizing syntactically those sentences preserved under direct

Received October 22, 1958. Work supported in part under grants from the National Science Foundation

[1] This concept is due to Tarski; see [**13**], [**14**].

product, as well as that of determining under what algebraic processes Horn sentences are preserved, remains open. That the general Horn sentence is not preserved under subdirect product is shown by a simple example: the family of all finite subsets of an infinite set constitutes, in the usual sense, a ring without unity, that is, in which the Horn sentence $\exists x \forall y \cdot xy = y$ fails, although it is a subdirect product of two-element fields, in which this sentence holds.

The earliest result of the kind under consideration is that of G. Birkhoff [3] who showed that those classes of algebras that are closed under formation of direct products, subsystems, and homomorphic images are precisely those classes definable by universally quantified equations. In addition to the work of Horn, Chang, and Morel, properties preserved under direct products have been studied by K. Bing [2], K. Appel [1], and A. I. Taimanov [12], while subdirect products have been studied by A. Malcev [11].

We first proved the result stated above by means of the theory of Natural Inference of G. Gentzen [5]. The proof offered here seems preferable in that it is simpler, despite the fact that it contains a double induction (which could, with some artificiality, be removed), and in that it presupposes less. We have tried to make the present exposition readable as it stands to one familiar with the general ideas; but for various details, in particular, for precise definitions, and for an Interpolation Theorem which plays a central role in the argument, we refer to our earlier papers [9], [10].

**2. Preliminaries.** Let $L$ be a first order language, with operation symbols $w$ of prescribed ranks $\rho(w)$, and relation symbols $r$ of ranks $\rho(r)$, among which is the symbol $e$ for the identity relation, of rank $\rho(e) = 2$. A *model* $\mathfrak{A}$ for $L$ consists of a set of operations $\mathfrak{A}w$ on a certain non-empty domain $A$, and of relations $\mathfrak{A}r$ on $A$, indexed by the operation symbols $w$ and relation symbols $r$ of $L$, and of corresponding ranks. A *relational system* is a model such that $\mathfrak{A}e$ is the identity relation on the domain $A$ of $\mathfrak{A}$.

Let $\mathfrak{A}_i$, for all $i$ in an index set $I$, be relational systems for a language $L$. The *direct product* $\mathfrak{A}$ of the $\mathfrak{A}_i$ is defined as follows. The domain $A$ of $\mathfrak{A}$ is the Cartesian product of the domains $A_i$ of the $\mathfrak{A}_i$. For each $i$ in $I$ we denote by $\pi_i$ the *projection* carrying each $a$ in $\mathfrak{A}$ onto its component $\pi_i a$ in $\mathfrak{A}_i$. The operations $\mathfrak{A}w$ of $\mathfrak{A}$ are defined by specifying their components: for each $i$, and $a_1, \cdots, a_{\rho(w)}$ in $\mathfrak{A}$,

$$\pi_i[\mathfrak{A}w(a_1, \cdots, a_{\rho(w)})] = \mathfrak{A}_i w(\pi_i a_1, \cdots, \pi_i a_{\rho(w)});$$

the relations $\mathfrak{A}r$ of $\mathfrak{A}$ are defined by taking $\mathfrak{A}r(a_1, \cdots, a_{\rho(r)})$ to hold, for $a_1, \cdots, a_{\rho(r)}$ in $\mathfrak{A}$, if and only if $\mathfrak{A}_i r(\pi_i a_1, \cdots, \pi_i a_{\rho(r)})$ holds in $\mathfrak{A}_i$ for each $i$ in $I$. It must be noted that this last criterion is satisfied by the identity relation. A system $\mathfrak{A}'$ is a *subdirect product* of the systems $\mathfrak{A}_i$ if it is a subsystem of the direct product $\mathfrak{A}$ such that, for each $i$ in $I$, the projection $\pi_i$ maps the domain $A'$ of $\mathfrak{A}'$ *onto* the domain $A_i$ or $\mathfrak{A}_i$.

The usual criterion for an algegraic system to be isomorphic to a subdirect product of systems from a given collection carries over directly to relational systems, and takes the following form.

CRITERION. *A relational system $\mathfrak{A}$ is isomorphic to a subdirect product of systems belonging to a given collection $K$ if and only if there exists a family $\Theta$ of homomorphisms $\theta$ of $\mathfrak{A}$ onto systems $\theta\mathfrak{A}$ in $K$ such that for all $r$ and $a_1, \cdots, a_{\rho(r)}$ in $\mathfrak{A}$, $[(\theta\mathfrak{A})r](\theta a_1, \cdots, \theta a_{\rho(r)})$ for all $\theta$ in $\Theta$ implies that $\mathfrak{A}r\ (a_1, \cdots, a_{\rho(r)})$ in $\mathfrak{A}$.*

Before turning to the main theorem we establish a series of lemmas.

LEMMA 1. *Let $F$ be a formula with distinct free variables $x_1, \cdots, x_n$, and $F'$ the result of replacing in $F$ the $x_i$ by new and distinct constants (operation symbols of rank zero) $w_i$. If $C$ is any formula that does not contain the $w_i$, and $C \Rightarrow F'$, then $C \Rightarrow \forall x_1 \cdots x_n F$.*

*Proof.*[2] Let $F$ being to a language $L$ that does not contain the $w_i$; then $F'$ belongs to the language $L'$ obtained from $L$ by adjoining the symbols $w_i$. Let $\mu$ be an interpretation of $L$ such that $\mu C = 1$, and $\lambda$ an interpretation of $L$ that agrees with $\mu$ except on the variables $x_1, \cdots, x_u$. We must show that $\lambda F = 1$. Extend $\mu$ and $\lambda$ to interpretations $\mu'$ and $\lambda'$ of $L'$ by defining $\mu' w_i = \lambda' w_i = \lambda x_i$. Since $C$ belongs to $L$, $\mu' C = \mu C = 1$. Since $C \Rightarrow F'$, and $\mu' C = 1, \mu' F' = 1$. Since $F'$ does not contain the $x_i$, $\lambda' F' = \mu' F'$, and $\lambda' F' = 1$. By the construction of $F'$ and of $\lambda'$, $\lambda F = \lambda' F'$, whence $\lambda F = 1$.

Let $\mathfrak{A}$ be a model for the language $L$, and $L(A)$ the language obtained from $L$ by adjoining new and distinct constants $w_a$ for each element $a$ of the domain $A$ of $\overline{\mathfrak{A}}$. Let $\overline{\mathfrak{A}}$ be the extension of $\mathfrak{A}$ to $L(A)$ defined by setting $\overline{\mathfrak{A}} w_a = a$ for all $a$ in $\mathfrak{A}$. Let $\mu$ be an ordinal number, and $L_\mu$ the language obtained from $L(A)$ by adjoining new and distinct relations symbols $r_\nu$ of rank $\rho(r_\nu) = \rho(r)$, for all $r$ in $L$ and $\nu < \mu$. If $\mathfrak{A}_\mu$ is any model for $L_\mu$, and $\nu < \mu$, let $\mathfrak{A}_{\mu,\nu}$ be the model for $L$ defined by taking $\mathfrak{A}_{\mu,\nu} w = \mathfrak{A}_\mu w$ for all $w$, and $\mathfrak{A}_{\mu,\nu} r = \mathfrak{A}_\mu r_\nu$ for all $r$.

Let $K$ be an elementary class[3] of relational systems. We shall say that a model $\mathfrak{A}_\mu$ of $L_\mu$ has the property (*) if

(1) *the restriction of $\mathfrak{A}_\mu$ to the language $L(A)$ is an elementary extension of $\overline{\mathfrak{A}}$;*

[2] For concepts appearing in this paper without definition, see **[9]**, **[10]**.

[3] As in **[10]**, we use "elementary class" in the sense of Tarski's "arithmetical class in the wider sense ($AC_\Delta$)".

(2) *the restriction to* $L$ *of* $\mathfrak{A}_{\mu,\nu}$ *is in* $K$, *for all* $\nu < \mu$;
(3) $\mathfrak{A}_\mu r \subseteqq \mathfrak{A}_{\mu,\nu} r$ *for all* $r$ *in* $L$ *and all* $\nu < \mu$.

LEMMA 2. *For* $\mu = 0$, *the model* $\mathfrak{A}_0 = \overline{\mathfrak{A}}$ *of the language* $L_0 = L(A)$ *has the property* (*).

*Proof.* Condition (1) is trivial, and (2) and (3) are vacuous.

Let $\Sigma$ be the class of all special Horn sentences that hold for $K$, and $\Sigma^*$ the class of those models that satisfy all sentences in $\Sigma$.

LEMMA 3. *Let* $\mathfrak{A}$ *be in* $\Sigma^*$, *and* $F$ *an atomic sentence, that is, an atomic formula without free variables, of* $L(A)$ *that fails in* $\overline{\mathfrak{A}}$. *Let* $\mathfrak{A}_\mu$ *be a model for* $L_\mu$ *with property* (*). *Then there exists a model* $\mathfrak{A}_{\mu+1}$ *for* $L_{\mu+1}$ *with property* (*) *such that*

(1′) *the restriction of* $\mathfrak{A}_{\mu+1}$ *to* $L_\mu$ *is an elementary extension of* $\mathfrak{A}_\mu$; *and*

(4) $F$ *fails in* $\mathfrak{A}_{\mu+1}$.

*Proof.* Let $\Gamma$ be the set of all sentences of $L$ that hold in $K$. Let $\Delta$ be the set of all sentences of the language $L_\mu(A_\mu)$ that hold in $\overline{\mathfrak{A}}_\mu$. Let $\Gamma'$ result from $\Gamma$ and $F'$ from $F$ by replacing each $r$ by the corresponding $r_\mu$. Let $I$ be the set of all sentences

$$I(r, r_\mu) = \forall x_1 \cdots x_{\rho(r)} \cdot r(x_1, \cdots, x_{\rho(r)}) \supset r_\mu(x_1, \cdots, x_{\rho(r)}) ,$$

for all $r$ in $L$.

Suppose the set $\Delta, I, \Gamma', \sim F'$ is inconsistent. By the Compactness Theorem, there exists a conjunction of sentences from $\Gamma$, and hence a single sentence $C$ from $\Gamma$, such that $\Delta, I, C', \sim F'$ is inconsistent. Thus $\Delta, I \Rightarrow C' \supset F'$, where $C' \supset F'$ contains only the relation symbols $r_\mu$, while $\Delta$ does not contain the $r_\mu$, and $I$ contains the $r_\mu$ only positively. By the Interpolation Theorem of [9], there exists a positive sentence $P'$ containing only the $r_\mu$ such that $\Delta$, $I \Rightarrow P'$ and $P' \Rightarrow C' \supset F'$. If $P$ is the result of replacing each $r_\mu$ in $P'$ by the corresponding $r$, it follows that $\Delta \Rightarrow P$ and $P \Rightarrow C \supset F$. Thus $C \Rightarrow P \supset F$. Let $P_0$ and $F_0$ result from $P$ and $F$ by replacing all $w_{b_i}$ that occur in them by distinct variables $x_1, \cdots x_n$. Since $C$ is in $\Gamma$, and belongs to the language $L$ that does not contain the $w_b$, it follows by Lemma 1 that $C \Rightarrow H$, where $H = \forall x_1 \cdots x_n \cdot P_0 \supset F_0$. Since $H$ contains only the relation symbols $r$, and does not contain the $w_b$, it belongs to the language $L$. Since $H$ is a special Horn sentence, and a consequent of $C$ in $\Gamma$, $H$ is in $\Sigma$. Since $\mathfrak{A}$ is in $\Sigma^*$, $H$ holds in $\overline{\mathfrak{A}}$, and hence in $\mathfrak{A}$. It follows that $P \supset F$ holds in $\overline{\mathfrak{A}}$. On the other hand, from $\Delta \Rightarrow P$ we have that $P$ holds in $\overline{\mathfrak{A}}_\mu$, hence in $\overline{\mathfrak{A}}$. From the fact that $P$ and $P \supset F$ both hold in $\overline{\mathfrak{A}}$ it follows that $F$ holds in $\overline{\mathfrak{A}}$ which contradicts the hypothesis of the lemma.

It has been established that the set $\Delta$, $I, \Gamma'$, $\sim F'$ is consistent, and therefore holds in some model $\mathfrak{B}$. Let $\mathfrak{C}$ be the restriction of $\mathfrak{B}$ to the language $L$. From the fact that $\mathfrak{B}$ satisfies $\Delta$, it follows by Proposition 3 of [10] that the quotient model $\mathfrak{A}_{\mu+1} = \mathfrak{C}/\mathfrak{C}e$ is a relational system and an elementary extension of $\mathfrak{A}_\mu$. This establishes (1′), and, by virtue of the hypothesis that $\mathfrak{A}_\mu$ has the property (*), it follows that $\mathfrak{A}_{\mu+1}$ satisfies (1) and also (2) and (3) for all $\nu < \mu$. From the fact that $\mathfrak{B}$, and therefore $\mathfrak{A}_{\mu+1}$, satisfies $\Gamma$, it follows that the restriction to $L$ of $\mathfrak{A}_{\mu+1,\,\mu}$ is in $K$, which completes the proof of (2). From the fact that $\mathfrak{B}$, and therefore $\mathfrak{A}_{\mu+1}$, satisfies $I$, it follows that, for all $r$, $\mathfrak{A}_{\mu+1}r \subseteq \mathfrak{A}_{\mu+1,\mu}r$, which completes the proof of (3). Finally, from the fact that $\mathfrak{B}$ satisfies $\sim F'$ it follows that $F'$ fails in $\mathfrak{A}_{\mu+1}$, as required by (4).

LEMMA 4. *Let $\mu$ be a limit ordinal, and a family of systems $\mathfrak{A}_\nu$ for $L_\nu$, all $\nu < \mu$, be given, with the property (*) and such that*

*$(1:\mu)$ for all $\rho < \nu < \mu$, the restriction of $\mathfrak{A}_\nu$ to $L_\rho$ is an elementary extension of $\mathfrak{A}_\rho$.*

*Let $F_\nu$, all $\nu < \mu$, be a set of atomic sentences of $L(A)$ such that*

*$(4:\mu)$ for all $\nu + 1 < \mu$, $F_\nu$ fails in $\mathfrak{A}_{\mu,\nu+1}$.*

*Then there exists a model $\mathfrak{A}_\mu$ for $L_\mu$ with property (*) and such that*

*$(1:\mu+1)$ and $(4:\mu+1)$ hold.*

*Proof.* By virtue of $(1:\mu)$, the $\mathfrak{A}_\nu$, $\nu < \mu$ constitute an ascending chain of systems and their union is a well defined system $\mathfrak{A}_\mu$. Let $\rho < \mu$, and for all $\nu$, $\rho < \nu \leqq \mu$, let $\mathfrak{B}_\nu$ be the restriction of $\mathfrak{A}_\nu$ to $L_\rho$. Then $\mathfrak{B}_\mu$ is the union of the ascending chain of systems $\mathfrak{B}_\nu$, $\rho < \nu < \mu$. Since each $\mathfrak{B}_\nu$, $\rho < \nu < \mu$, is by $(1:\mu)$ an elementary extension of $\mathfrak{A}_\rho$, it follows directly from the definition of elementary extension that $\mathfrak{B}_\mu$ is an elementary extension of $\mathfrak{A}_\rho$.[4] This suffices to extend $(1:\mu)$ to $(1:\mu+1)$. That $\mathfrak{A}_\mu$ has property (*) follows from this directly. It remains only to note that, since $\mu$ is a limit ordinal, $(4:\mu+1)$ is in fact equivalent to $(4:\mu)$.

LEMMA 5. *Let $\mathfrak{A}$ be in $\Sigma^*$. Then there exists an ordinal $\mu$ and a model $\mathfrak{A}_\mu$ for $L_\mu$, with property (*) and such that*

*(4*) if any atomic sentence $F$ of $L(A)$ fails in $\overline{\mathfrak{A}}$, then it fails in some $\mathfrak{A}_{\mu,\nu}$, $\nu < \mu$.*

*Proof.* Let the $F_\nu$, all $\nu < \mu$, for some $\mu$, be the set of all atomic sentences of $L(A)$ that fail in $\overline{\mathfrak{A}}$. Let $\mathfrak{A}_0 = \overline{\mathfrak{A}}$ as in Lemma 2. For some $\nu$, $\nu < \mu$, suppose that systems $\mathfrak{A}_\rho$ have been constructed for all $\rho < \nu$ with property (*) and satisfying $(1:\nu)$, $(4:\nu)$. If $\nu$ is not a

[4] See Theorem 1.9 of [15].

limit ordinal, Lemma 3 with $F = F_\nu$ and $\nu - 1$ for $\mu$ assures us of $\mathfrak{A}_\nu$ with the required properties. If $\nu$ is a limit ordinal, Lemma 4 yields the same result. Thus transfinite induction yields a chain of $\mathfrak{A}_\nu$, all $\nu \geqq \mu$. The condition $(4 : \mu)$ now gives (4*).

LEMMA 6. *Let $\mathfrak{A}$ be in $\Sigma^*$. Then there exists an oridinal $\sigma$ and a system $\mathfrak{A}_\sigma$ for $L_\sigma$ with the property (*) and such that*

(4**) *if any atomic sentence $F$ of the language $L(A_\sigma)$ fails in $\overline{\mathfrak{A}}_\sigma$, then it fails in $\mathfrak{A}_{\sigma,\nu}$ for some $\nu < \sigma$.*

*Proof.* Iteration of Lemma 5 yields a sequence of ordinals $\mu_0 = 0 \leqq \mu_1 \leqq \mu_2 \leqq \cdots$ such that $\mathfrak{A}_0$ is $\overline{\mathfrak{A}}$; each $\mathfrak{A}_{\mu_n}$ has the property (*); for each $n$, the restriction of $\mathfrak{A}_{\mu_{n+1}}$ to $L_{\mu_n}$ is an elementary extension of $\mathfrak{A}_{\mu_n}$; and, finally, that if an atomic sentence $F$ of the language $L(A_{\mu_n})$ fails in $\overline{\mathfrak{A}}_{\mu_n}$, then it fails in some $\mathfrak{A}_{\mu_{n+1}}, \nu$, $\nu < \mu_{n+1}$. It follows directly that, for $\sigma = \lim_{n<\omega} \mu_n$, the union $\mathfrak{A}_\sigma$ of the ascending chain of $\mathfrak{A}_{\mu_n}$, $n < \omega$, has the required properties.

## 3. The Main Theorem.

THEOREM. *Let $\mathfrak{A}$ be a relational system of the language $L$, and $K$ an elementary class of systems of $L$. Then the following are equivalent.*

(1) *$\mathfrak{A}$ satisfies all special Horn sentences that hold in $K$;*

(2) *$\mathfrak{A}$ has an elementary extension that is a subdirect product of systems in $K$.*

*Proof.* To show that (2) implies (1), it clearly suffices to show that if $S$ is a sentence of the form $S = \forall x_1 \cdots x_n \cdot P \supset F$ where $P$ is positive and $F$ is atomic, and $\mathfrak{A}$ is a subdirect product of systems $\mathfrak{A}_\iota$ in which $S$ holds, then $S$ holds in $\mathfrak{A}$. Suppose then that $S$ holds in all the $\mathfrak{A}_i$, and yet $S$ fails in $\mathfrak{A}$. Then there exists an interpretation $\mu$ of $L$ in $\mathfrak{A}$ such that $\mu P = 1$ and $\mu F \neq 1$. Since each projection $\pi_i$ is a homomorphism of $\mathfrak{A}$ onto $\mathfrak{A}_i$, we have that, for all $r$ in $L$ and terms $t_1, \cdots, t_{\rho(r)}$, $\mathfrak{A}r(\mu t_1, \cdots, \mu t_{\rho(r)})$ implies $\mathfrak{A}_i r(\pi_i \mu t_1, \cdots, \pi_i \mu t_{\rho(r)})$. For each $\mathfrak{A}_i$, define an interpretation $\mu_i$ in $\mathfrak{A}_i$ by setting $\mu_i x = \pi_i \mu x$. Then $\mu G = 1$ implies $\mu_i G = 1$ for all $i$ if $G$ is an atomic formula, whence $\mu P = 1$ implies $\mu_i P = 1$ for all $i$. Since $S$ holds in each $\mathfrak{A}_i$, that $\mu_i P = 1$ implies $\mu_i F = 1$, all $i$. But $F$ is an atomic formula, and $\mu_i F = 1$ implies that $\mu F = 1$, a contradiction.

To show that (1) implies (2), assume that $\mathfrak{A}$ is in $\Sigma^*$, where $\Sigma$ is the set of all special Horn sentences true for $K$. By Lemma 6, for some ordinal $\sigma$ there exists a system $\mathfrak{A}_\sigma$ of $L_\sigma$ with properties (*) and (4*). Let $\mathfrak{A}'$ be the restriction of $\mathfrak{A}_\sigma$ to the language $L$; by virtue of (1), $\mathfrak{A}'$ is an elementary extension of $\mathfrak{A}$. For each $\nu < \sigma$, let $\mathfrak{B}_\nu$ be the restriction of $\mathfrak{A}_{\sigma,\nu}$ to the language $L$; by virtue of (2), each $\mathfrak{B}_\nu$ is in $K^{**}$, and consequently the quotient model $\mathfrak{C}_\nu = \mathfrak{B}_\nu/\mathfrak{B}_\nu e$ is a relational system in $K$. For each $\nu < \sigma$, $\mathfrak{A}'$ and $\mathfrak{B}_\nu$ have the same domain $A$, whence the canonical map $\theta_\nu$ of $\mathfrak{A}$ onto $\mathfrak{A}/\mathfrak{B}_\nu e$ maps $A$ onto the domain $C_\nu$ of $\mathfrak{C}_\nu$. By virtue of (3), for each $\nu < \sigma$ and each $r$ in $L$, $\mathfrak{A}'r = \mathfrak{A}_\sigma r \subseteqq \mathfrak{A}_{\sigma,\nu}r = \mathfrak{A}_\nu r$, whence $\theta_\nu$ defines a homomorphism of $\mathfrak{A}'$ onto $\mathfrak{C}_\nu$. To complete the proof that the family of $\theta_\nu$, $\nu < \sigma$, satisfies the Criterion for $\mathfrak{A}'$ to be a subdirect product of the $\mathfrak{C}_\nu$, $\nu < \sigma$, suppose that, for some $r$ in $L$ and $a_1, \cdots, a_{\rho(r)}$ in $\mathfrak{A}$, $[\mathfrak{C}_\nu r](\theta_\nu a_1, \cdots, \theta_\nu a_{\rho(r)})$ holds for all $\nu < \sigma$. If $\mathfrak{A}'r(a_1, \cdots, a_{\rho(r)})$ failed in $\mathfrak{A}'$, then the atomic sentence $F = r(w_{a_1}, \cdots, w_{a_{\rho(r)}})$ would fail in $\overline{\mathfrak{A}}$. By virtue of (4*), $F$ would fail in $\mathfrak{A}_{\sigma,\nu}$, for some $\nu < \sigma$, hence in $\mathfrak{B}_\nu$. Since $\mathfrak{B}_\nu w_{a_i} = \mathfrak{A}_\sigma w_{a_i} = \mathfrak{A} w_{a_i} = a_i$, $[\mathfrak{B}_\nu r](a_1, \cdots, a_{\rho(r)})$ would fail in $\mathfrak{B}_\nu$, whence $[\mathfrak{C}_\nu r](\theta_\nu a_1, \cdots, \theta_\nu a_{\rho(r)})$ would fail in $\mathfrak{C}_\nu$. This contradicts our hypothesis, and extablishes the desired conclusion, that $\mathfrak{A}'r(a_1, \cdots, a_{\rho(r)})$ holds in $\mathfrak{A}'$.

**4. A complementary example.** It will be shown that there exists an elementary class such that the set of all subdirect products of systems from this class is not an elementary class. In consequence, the reference to elementary extensions in the preceding theorem can not be deleted.

If $K$ is any class of systems, let $P(K)$ be the class of all systems isomorphic to some subdirect product of systems from $K$, and let $P_0(K)$ be the class of all systems isomorphic to some subdirect product of a non-empty family of systems from $K$. As was noted earlier, $P(K)$ will differ from $P_0(K)$ at most in containing all *trivial systems*, with domain a single element and all relations universal, which will not belong to $P_0(K)$ unless $K$ itself contains some trivial system. We suppose now that the language $L$ contains only a finite number of relation symbols, whence there is a single sentence $T$ characterizing the class of all trivial systems. Then

(1) *$P(K)$ is an elementary class if and only if $P_0(K)$ is an elementary class.*

If $K$ contains a trivial system, then $P(K) = P_0(K)$ and there is nothing to prove. Otherwise $P_0(K) = P(K) - T$. If $P(K)$ is elementary, say $P(K) = \Gamma^*$, then evidently $P_0(K) = \{\Gamma, \sim T\}^*$ and $P_0(K)$ is elementary. On the other hand, if $P_0(K) = \Gamma^*$, then $P_0(K) = \{C \vee T\colon \text{all } C \text{ in } \Gamma\}^*$.

If $K$ is any class of systems, let $H'(K)$ be the class of all those systems of which some homomorphic image lies in $K$. The following assertion can be obtained by dualizing the proof of the Main Theorem of [10], or may be deduced as a corollary to that theorem.

(2) *if $K$ is an elementary class, then $H'(K)^*$ is the set of all consequences of all negative sentences that hold in $K$.*

It is clear from the definitions that $P_0(K) \subseteqq H'(K)$. To obtain a partial converse, define an occurrence of a relation symbol in a sentence $S$ to be *universal* if no variable that occurs in the atomic formula containing the given occurrence is existentially quantified in $S$. Then

(3) *if* $K = \Gamma^*$, *where no sentence of* $\Gamma$ *contains a positive universal occurrence of any relation symbol, then* $H'(K)^* = P_0(K)^*$.

To establish (3), we first show that the argument used to establish the Interpolation Theorem in [9] in fact enables us to impose the following additonal conditions in the conclusion :

(4a) *a relation symbol has a positive universal occurrence in* $S^0$ *only if it has a positive universal occurrence in* $S$;

(4b) *a relation symbol has a positive non-universal occurrence in* $S^0$ *only if it has a positive non-universal occurrence in* $T$.

We refer to the proof of the Interpolation Theorem. To prove (4a), suppose that a relation symbol $r$ has no positive universal occurrence in $S = S^1$. Then each atomic formula of $S^1$ that contains $r$ positively also contains some variable that is existentially quantified in $S^1$, whence the corresponding atomic formula in the Skolem matrix for $S^1$ contains one of the Skolem functions $s_i^1$. It follows that each atomic formula in the Skolem matrix $M^1$ of $U^1$ that contains $r$ positively also contains one of the functions $s_i^1$, and the same is then true of $M^0$, whence it follows that the corresponding atomic formula in $S^0$ contains an existentially quantified variable. Since positive occurrences of $r$ in $S^0$ can arise only in this fashion, it follows that all such occurrences are non-universal.

To prove (4b), note first that an atomic formula containing a positive occurrence of $r$ in $S^0$ will correspond to an atomic formula $A$ in $M^0$ and hence in $M^1$, and that, if the occurrence is non-universal, then $A$ will contain one of the functions $s_{0j}^1$. Suppose now that every positive occurrence of $r$ in $T$ is universal; then in $S^2$, equivalent to $\sim T$, we may suppose that every variable that occurs in an atomic formula containing a negative occurrence of $r$ is existentially quantified. Passing to the Skolem matrix of $S^2$ and thence to $M^2$, it follows that if $B$ is any atomic formula of $M^2$ that contains a negative occurrence of $r$, then each occurrence of a variable of $r$ is subordinate to some one of the $s_{0j}^2$ in the sense of occurring in a term beginning with this symbol. From the construction of $M^0$ from $M^1$ and $M^2$ it results that an atomic formula $A$ of $M^1$, as above, will appear also in $M^0$ only in case $\eta A = \eta B$, for $B$ an atomic formula of $M^2$, as described. But this is impossible, since every occurrence of a symbol $s_{0k}^1$ in $\eta B$ is subordinate to some $s_{0i}^2$ while $A$ contains an ocurrence of some $s_{0j}^1$ that is not subordinate to any $s_{0i}^2$ in $A$, which does not contain the $s_{0i}^2$, and hence this occurrence of $s_{0j}^1$ is not subordinate to any $s_{0i}^2$ in $\eta A$.

Turn now to the proof of (3). From the theorem of § 3 it is easy to see that $P_0(K)^*$ consists of all consequences of 'generalized' special Horn sentences that hold in $K$, that is, of those sentences that hold in $K$ and are obtained by universal quantification and conjunction from formulas of the types $P \supset F$ and $\sim P$, for $P$ positive and $F$ atomic. From the hypothesis of (3), if $\Gamma \Rightarrow T$, where $T$ is a generalized special Horn sentence, then $\Gamma \Rightarrow S$ and $S \Rightarrow T$, where $S$, a conjunction of sentences from $\Gamma$, contains no positive universal occurrence of any relation symbol. Since $T$ contains no positive non-universal occurrences of any relation symbol, application of the Interpolation Theorem with the conditions (4a) and (4b) provides the existence of $S^0$ such that $S \Rightarrow S^0$ and $S^0 \Rightarrow T$, where $S^0$ contains no positive occurrences of any relation symbol, either universal or non-universal: in short, where $S^0$ is negative. Since $\Gamma \Rightarrow S^0$ and $S^0 \Rightarrow T$, it follows by (2) that $T \in H'(K)^*$. This establishes that $P_0(K)^* \subseteq H'(K)^*$, while the opposite inclusion follows from the fact that $P_0(K) \subseteq H'(K)$.

In § 5 of [10] an elementary class $K$ of systems, without operations and with a single binary relation (other than identity), was constructed, with the property that $H(K)$ is not elementary. Replacing, in each system in $K$, the relation in question by its complementary relation, yields an elementary class $K'$ of systems such that $H'(K')$ is not elementary. More explicitly, $K'$ is characterized by the single sentence

$$S' : \quad \exists xy \forall z \exists t : \sim r(x, y) \wedge : \sim r(x, z) \supset \sim r(x, t) \wedge \sim r(z, t)$$

It follows as in [10, § 5], or may be derived from the result there, that $H'(K)^* = \{S_1', S_2', \cdots\}^{**}$, where the $S_n'$ result from the $S_n$ by prefixing a negation sign to each occurrence of the symbol $r$. If $\mathfrak{A}'$ is the natural numbers with the relation $x \leqq y$, it contains descending chains of arbitrary length, hence satisfies the $S_n'$ and belongs to $H'(K')^{**}$. If $\mathfrak{A}'$ had a homomorphic image $\mathfrak{B}$ in $K'$, from $S'$ it would follow that $\sim \mathfrak{B}r(b_0, b_1), \cdots, \sim \mathfrak{B}r(b_n, b_{n+1}), \cdots$ for some $b_0, b_1, \cdots$ in $\mathfrak{B}$, and any set of inverse images $a_0, a_1, \cdots$ would constitute an infinite descending chain in $\mathfrak{A}'$, which is clearly a contradiction. Thus $\mathfrak{A}'$ is not in $H'(K')$, and $H'(K') \neq H'(K)^{**}$, that is, $H'(K')$ is not elementary.

Finally, the set $\Gamma = \{S'\}$ satisfies the hypothesis of (3); indeed, each atomic formula of $S'$ contains one of the existentially quantified variables $x, y$ or $t$. Thus, by (3), $P_0(K')^* = H'(K')^*$. It now follows that $P(K')$ is not elementary. For, by (1), this would imply that $P_0(K')$ were elementary, hence $P_0(K')^{**} = P_0(K')$. But $P^0(K')^* = H'(K')^*$ implies $H'(K')^{**} = P_0(K')^{**}$, and $P_0(K') \subseteq H'(K')$, which, together with $P_0(K')^{**} = P_0(K')$, would imply $H'(K')^{**} \subseteq H'(K')$ and hence that $H'(K')$ were elementary, a contradiction.

## REFERENCES

1. K. I. Appel, *Horn sentences in identity theory,* Notices of Amer. Math. Soc. **6** (1958), 343.
2. K. Bing, *On arithmetical classes not closed under direct union*, Proc. Amer. Math. Soc. **6** (1955), 836.
3. G. Birkhoff, *On the structure of abstract algebras,* Proc. Camb. Philos. Soc. **31** (1935), 433.
4. C. C. Chang and Anne C. Morel, *On closure under direct product*, J. Symbolic Logic **23** (1958), to appear.
5. G. Gentzen, *Untersuchungen über das logische Schliessen,* Math. Z. **39** (1934-35), 176, 405.
6. A. Horn, *On sentences which are true of direct unions of algebras,* J. Symbolic Logic **16** (1951), 14.
7. N. Jacobson, *The structure of rings,* Amer. Math. Soc., Providence, 1956.
8. S. R. Kogalovski, *On universal classes of algebras, closed under direct product,* Uspehi Mat. Nauk **13** (1958), 241.
9. R. C. Lyndon, *An interpolation theorem in the predicate calculus,* to Pacific J. Math. **9** (1959).
10. R. C. Lyndon, *Properties preserved under homomorphism,* to Pacific J. Math. **9** (1959).
11. A. I. Malcev, *Subdirect products of models,* Doklady Akad. Nauk SSSR (N.S.) **109** (1959), 264.
12. A. D. Taimanov, *On classes of models, closed under direct product,* Uspehi Mat. Nauk **13** (1958), 231.
13. A. Tarski, *Some notions and methods on the borderline of algebra and metamathematics,* Proc. Internationa Congress of Math. 1, Cambridge, 1950, 705.
14. A. Tarski, *Contributions to the theory of models,* I, II, III, Indag. Math. **16** (1954), 572, 582; **17** (1955), 55.
15. A. Tarski and R. L. Vaught, *Artithmetical extensions of relational systems*, Composition Math. **13** (1957), 81.

UNIVERSITY OF MICHIGAN

# A LEMMA ON ANALYTIC CURVES

ROBERT OSSERMAN

The following lemma turns out to be useful in many places in Riemann surface theory. It is generally sufficient to have piecewise analyticity, rather than analyticity, but the availability of analytic curves will invariably make proofs simpler. This is especially true of the second half of the lemma, since an analytic Jordan curve on a Riemann surface is defined by a mapping of the unit circle into the surface and this mapping can be extended to a one-to-one map of an annulus into the surface. The extended mapping can be used, for example, to define explicitly differentials on the surface having prescribed properties.

The proof which we give is perhaps not the most straightforward one but has certain advantages over the usual type of reasoning involving subdivision of the parameter interval.

LEMMA. *Every closed curve on a Riemann surface is homotopic to an analytic closed curve, and homologous to a finite sum of analytic Jordan curves.*

REMARK. By homology we mean singular homology. We are not concerned here with the choice of definition since we only use the following two properties of singular homology.

(1) Homotopic curves are homologous.

(2) If a closed curve $C$ is defined by a mapping $f(t)$ on the interval $I$ and if there is a subdivision of $I$ into intervals $I_k, k = 1, \cdots, n$, and if the restriction of $f(t)$ to $I_k$ defines a closed curve $C_k$, then $C$ is homologous to $\sum_{k=1}^{n} C_k$.

We may also word the lemma as follows:

(*a*) *Every homotopy class on a Riemann surface contains an analytic curve.*

(*b*) *Every singular homology class on a Riemann surface contains a cycle of the form $\sum_{k=1}^{n} C_k$ where the $C_k$ are analytic Jordan curves.*

*Proof.* Let $R$ be the Riemann surface and $C$ an arbitrary closed curve. Let $R^*$ be the universal covering surface of $R$ and let $p^*$ be a point on $R^*$ which projects onto a point $p$ of the curve $C$. If we continue along the curve $C$ starting at $p^*$ we arrive at a point $q^*$ which also projects onto $p$. Any arc joining $p^*$ to $q^*$ will, by definition of the universal covering surface, project onto a curve homotopic to $C$. In

Received July 15, 1958. This paper was written while the author was a Temporary member of the Institute of Mathematical Sciences at New York University.

particular, an analytic arc will project onto a curve which is analytic except possibly at $p$. To prove part (a) of the lemma we make a simple choice of arc which will give a curve analytic also at $p$.

We note first that if $p^*$ and $q^*$ coincide, then $C$ is homotopic to zero, and any small circle through $p$ will be homotopic (and homo logous) to zero, hence to $C$. In particular if $R^*$ is the sphere, then $R$ is itself the sphere and $C$ must have been homotopic to zero.

The other two possibilities are that $R^*$ is the plane or the interior of the unit circle. In the first case the cover transformation taking $p^*$ into $q^*$ is a linear transformation which takes the line through $p^*$ and $q^*$ into itself. The projection of the straight line segment joining $p^*$ to $q^*$ will therefore be analytic even at $p$.

If $R^*$ is the interior of the unit circle, then the cover transformation taking $p^*$ into $q^*$ will be a linear fractional transformation with either one or two fixed points on the unit circle. The circle through $p^*$ and $q^*$ and the fixed point (or points) will be mapped onto itself, so that again the arc of this circle joining $p^*$ to $q^*$ will project onto an analytic curve homotopic to $C$.

This proves part (a) of the lemma.

We have obtained in particular a simple proof that every closed curve is homologous to an analytic closed curve, which is in itself a useful fact.

The proof of part (b) requires a bit more effort. The case where $R^*$ is the plane is almost trivial since (except where $R$ is itself the plane, hence simply connected) the group of cover transformations is either of the form $z+nb$ where $b$ is a fixed complex number and $n$ runs through the integers, or else of the form $z+ma+nb$ where $a$ and $b$ are fixed and $m$ and $n$ run through the integers. In the first case $p^*$ and $q^*$ will correspond to a pair of points of the form $z_0$ and $z_0+nb$, so that the projection of the segment from $z_0$ to $z_0+nb$ will consist of the projection $C'$ of the segment from $z_0$ to $z_0+b$ covered $n$ times. Hence $C$ is homologous to $nC'$. But $C'$ is a Jordan curve since no two points on the segment between $z_0$ and $z_0+b$ are equivalent.

Similarly, in the second case if $p^*$ and $q^*$ correspond to $z_0$ and $z_0+ma+nb$, we simply take the straight line segment from $z_0$ to $z_0+ma$ followed by the segment from $z_0+ma$ to $z_0+ma+nb$.

Finally, in the case where $R^*$ is the interior of the unit circle, we consider the metric fundamental polygon $P$ consisting of all points in the unit circle which are nearer (in the non-euclidean metric) to the point $p^*$ than to any point equivalent to $p^*$ under a cover transformation[1]. If $T$ is the cover transformation which takes $p^*$ into $q^*$ we may represent

[1] A simplification in the proof at this point is due to the referee.

All the basic information about the fundamental polygon can be found, for example, in the book of Nevanlinna, **Uniformisierung,** Chapter VII, and in particular, §7.15.

it as $T = \prod_{k=1}^{n} T_k$, where each $T_k$ is a cover transformation taking one side $a_k$ of $P$ onto another side $b_k$. The sides $a_k$ and $b_k$ are circular arcs, and we can form the Riemann surface $R_k$ which consists of the interior points of $P$ together with inner points of the arcs $a_k$ and $b_k$ with equivalent points under $T_k$ identified. Then $R_k$ is doubly-connected and may be mapped conformally onto an annulus $r_1 < |w| < r_2$, where $r_1 = 0$ if $a_k$ and $b_k$ have a vertex in common. This annulus is one-to-one conformally equivalent to $R_k$ and hence to the surface $R$ cut along certain analytic arcs. Let $C_k$ be the curve in $R$ corresponding to that circle $|w| = r'$ which passes through the image of the point $p$. Then $C_k$ is a Jordan curve passing through $p$, and if we continue $R^*$ along $C_k$ starting at a point $p_k{}^*$ of $a_k$, we arrive at the point $q_k{}^*$ of $b_k$ which is the image of $p_k{}^*$ under $T_k$. Hence the curve $C_k$ generates the transformation $T_k$, and the curve $C' = \prod_{k=1}^{n} C_k$ is a closed curve at $p$ which generates the transformation $T$. Thus $C'$ is homotopic to $C$, and using the properties of singular homology remarked above we have that $C$ is homologous to $C'$ which is homologous to $\sum_{k=1}^{n} C_k$, proving the lemma.

STANFORD UNIVERSITY.

# ON A THEOREM DUE TO SZ.-NAGY

R. S. PHILLIPS

B. Sz.-Nagy [4] has proved the following theorem:

THEOREM A. *Let $[T_t;\ t \geqq 0]$ be a strongly continuous semi-group of contraction operators on a Hilbert space $H$. Then there exists a group of unitary operators $[\mathbf{U}_t, -\infty < t < \infty]$ on a larger Hilbert space $\mathbf{H}$ such that*

$$T_t y = \mathbf{P}\mathbf{U}_t y, \qquad y \in H,\ t \geqq 0\ ; \tag{1}$$

*here $\mathbf{P}$ is the projection operator with range $H$. Then space $\mathbf{H}$ can be chosen in a minimal fashion so that $[\mathbf{U}_t H;\ -\infty < t < \infty]$ spans $\mathbf{H}$. In this case $[\mathbf{U}_t]$ is strongly continuous and the structure $\{\mathbf{H}, \mathbf{U}_t, H\}$ is determined to within an isomorphism.*[1]

The infinitesimal generator $L$ of the semi-group $[T_t]$ is defined by

$$\lim_{\delta \to 0+} \delta^{-1}[T_\delta y - y] = Ly \tag{2}$$

for all $y \in H$ for which this limit exists. The operator $L$ is linear and closed with dense domain, $\mathfrak{D}(L)$ (see [1]). It is shown in [2] that $L$ is maximal dissipative in the sense that

$$(y, Ly) + (Ly, y) \leqq 0, \qquad y \in \mathfrak{D}(L)\ , \tag{3}$$

and $L$ being maximal with respect to this property. Since $[\mathbf{U}_t]$ is a semi-group as well as a group of operators, the infinitesimal generator $\mathbf{L}$ of $[\mathbf{U}_t]$ also shares these properties; however in the case of a group of unitary operators $i\mathbf{L}$ is in addition self-adjoint.

The purpose of this note is to study the relation between $L$ and $\mathbf{L}$. It turns out that $L$ is a restriction of $\mathbf{L}$ only when $L$ is maximal symmetric. In general $L$ is neither a restriction nor a projection of $\mathbf{L}$; in fact $\mathfrak{D}(\mathbf{L}) \cap H$ may contain only the zero element. Nevertheless we shall obtain $\mathbf{H}$, $\mathbf{L}$, and $[\mathbf{U}_t]$ directly from $L$, our principal tool being the discrete analogue of the above theorem, which is also due to Sz.-Nagy [4], namely

THEOREM B. *Let $J$ be a contraction operator on a Hilbert space $H$. Then there exists a unitary operator $\mathbf{J}$ on a larger Hilbert space $\mathbf{H}$ such that*

$$J^n y = \mathbf{P}\mathbf{J}^n y, \qquad y \in H,\ n \geqq 0; \tag{4}$$

*here $\mathbf{P}$ is the projection operator with range $H$. The space $\mathbf{H}$ can be*

Received September 24, 1958. This paper was written under the sponsorship of the National Science Foundation, contract NSF G-4231.

[1] Two structures $\{\mathbf{H}, \mathbf{U}_t, H\}$ and $\{\mathbf{H}', \mathbf{U}'_t, H\}$ are isomorphic if there is a unitary map $\mathbf{V}$ of $\mathbf{H}$ onto $\mathbf{H}'$ which is the identity on $H$ and is such that $\mathbf{V}\mathbf{U}_t y = \mathbf{U}'_t \mathbf{V} y$ for all $y \in \mathbf{H}$.

*chosen in a minimal fashion in the sense that* $[\mathbf{J}^n H; -\infty < n < \infty]$ *spans* $\mathbf{H}$. *In this case the structure* $\{\mathbf{H}, J, H\}$ *is determined to within an isomorphism.*

For a maximal dissipative operator $L$ with dense domain, it is shown in [2, §1.1] that $(I - L)$ is one-to-one with range $\Re(I - L) = H$ and that

$$(5) \qquad J = (I + L)(I - L)^{-1}$$

is a contraction operator with $\mathfrak{D}(J) = H$ and such that $(I + J)$ is one-to-one. Applying Theorem B we obtain the unitary operator $\mathbf{J}$ on the enlarged space $\mathbf{H}$ spanned by $[\mathbf{J}^n H; -\infty < n < \infty]$ with $\mathbf{J}$ satisfying the property (4).

LEMMA 1. *The operator* $(\mathbf{I} + \mathbf{J})$ *is one-to-one.*

*Proof.* Let $S$ be a contraction operator, set $\mathfrak{Z}(S) = [y; Sy + y = \theta]$, and denote the projection operator with range $\mathfrak{Z}(S)$ by $P_S$. Then the ergodic theorem (see [3, pp. 400-406]) asserts that

$$\operatorname{st.}\lim_{n\to\infty} (n + 1)^{-1} \sum_{n=0}^{n} (-S)^k = P_S$$

and that $SP_S = P_S S = -P_S$. We apply this result first to $J$ and then to $\mathbf{J}$. Making use of (4) we see that

$$\mathbf{P}\mathbf{P}_{\mathbf{J}} y = P_J y, \qquad y \in H .$$

As noted above $P_J = \Theta$, so that $\mathbf{P}\mathbf{P}_{\mathbf{J}}\mathbf{P} = \Theta$. Actually $\mathbf{P}_{\mathbf{J}}\mathbf{P} = \Theta$; for otherwise there would exist a $y \in H$ with $\mathbf{P}_{\mathbf{J}} y \neq \theta$ so that

$$(\mathbf{P}\mathbf{P}_{\mathbf{J}}\mathbf{P} y, y) = (\mathbf{P}_{\mathbf{J}} y, y) = \|\mathbf{P}_{\mathbf{J}} y\|^2 > 0 ,$$

which is impossible. Thus $\mathbf{P}_{\mathbf{J}}\mathbf{P} = \Theta$ and hence $\mathfrak{Z}(\mathbf{J})$ is orthogonal to $H$. But this means that

$$\mathbf{P}_{\mathbf{J}} \mathbf{J}^n H = \mathbf{J}^n \mathbf{P}_{\mathbf{J}} H = \theta ,$$

and we infer that $\mathbf{J}^n H$ is orthogonal to $\mathfrak{Z}(\mathbf{J})$ for all $n$. The minimal property of $\mathbf{H}$ therefore requires that $\mathfrak{Z}(\mathbf{J}) = \theta$.

REMARK. Associated with $\mathbf{J}$ is the resolution of the identity $[\mathbf{E}(\sigma); -\pi < \sigma \leqq \pi]$ and the integral representation

$$\mathbf{J}^n = \int_{-\pi}^{\pi} \exp(in\sigma) d\mathbf{E}(\sigma) .$$

Setting the restriction of $\mathbf{P}\mathbf{E}(\sigma)$ to $H$ equal to $F(\sigma)$ we see by (4) that

$$J^n = \int_{-\pi}^{\pi} \exp(in\sigma) dF(\sigma) .$$

The argument used in Lemma 1 applied to $S = \exp(i\mu)J$ shows that if

$J$ has no eigenvalues of absolute value one, then neither does $\mathbf{J}$ and hence that both $\mathbf{E}(\sigma)$ and $F(\sigma)$ are strongly continuous in $\sigma$. Conversely, $F(\sigma)$ is strongly continuous then as is readily verified

$$(n+1)^{-1}\sum_{k=0}^{n}[\exp(i\mu)J]^k y$$
$$= \int_{-\pi}^{\pi} K_n(\sigma+\mu)dF(\sigma)y \to \theta\,, \qquad y \in H;$$

here

$$K_n(\sigma) = (n+1)^{-1}\exp(in\sigma/2)\sin\left[\frac{n+1}{2}\sigma\right]\left[\sin\frac{\sigma}{2}\right]^{-1}.$$

It then follows from the ergodic theorem that $\mathfrak{Z}\{-\exp(i\mu)J\} = \theta$ and hence that $J$ has no eigenvalues of absolute value one.

THEOREM. *Set*

(6) $$\mathbf{L} = (\mathbf{J}-\mathbf{I})(\mathbf{J}+\mathbf{I})^{-1}\,.$$

*Then* $\mathbf{L}$ *generates a strongly continuous group of unitary operators* $[\mathbf{U}_t;\ -\infty < t < \infty]$ *such that*

(7) $$T_t y = \mathbf{P}\mathbf{U}_t y, \qquad y \in H,\ t \geqq 0$$

*and* $[\mathbf{U}_t H;\ -\infty < t < \infty]$ *spans* $\mathbf{H}$.

*Proof.* It follows from the above lemma that $(\mathbf{I}+\mathbf{J})$ is one-to-one and hence that $\mathbf{L}$ is well-defined. Morever $\mathfrak{D}(\mathbf{L}) = \mathfrak{R}(\mathbf{I}+\mathbf{J})$ is necessarily dense in $\mathbf{H}$ since otherwise $(\mathbf{I}+\mathbf{J}^*)$ would nullify some non-zero vector and since $\mathbf{J}^{-1} = \mathbf{J}^*$ the same would be true of $(\mathbf{I}+\mathbf{J})$. Further it is clear that $i\mathbf{L}$ is the Cayley tranform of $i\mathbf{J}$ and hence $\mathbf{L}$ generates a strongly continuous group of unitary operators which we shall denote by $[\mathbf{U}_t]$. In order to verify (7) we proceed to represent the resolvent $R(\lambda, L) = (\lambda I - L)^{-1}$ in terms of $J$ for $\lambda > 0$. We see from (5) that

(8) $$y = 2^{-1}(Ju+u) \text{ and } Ly = 2^{-1}(Ju-u), \qquad u \in H\,.$$

Suppose next that $\lambda y - Ly = f$. Replacing $y$ by $u$ as in (8) we obtain

$$2^{-1}\lambda(Ju+u) - 2^{-1}(Ju-u) = f$$

so that

$$u = 2(1+\lambda)^{-1}\sum_{n=0}^{\infty}[(1-\lambda)(1+\lambda)^{-1}]^n J^n f, \qquad \lambda > 0\,.$$

Again making use of (8) we get

$$y = 2^{-1}(Ju+u) = \sum_{n=0}^{n} a_n(\lambda)J^n f$$

where

$$a_0(\lambda) = (1+\lambda)^{-1} \text{ and } a_n(\lambda) = 2(1-\lambda)^{n-1}(1+\lambda)^{-n-1} \text{ for } n > 0 .$$

Thus $R(\lambda, L)$ can be represented by an absolutely convergent series in powers of $J$ for $\lambda > 0$. Taking powers of $R(\lambda, L)$ we see that

$$[R(y, L)]^k = \sum_{n=0}^{\infty} a_n^{(k)}(\lambda) J^n ,$$

where again the series is absolutely convergent. Similarly

$$\mathbf{R}(\lambda, \mathbf{L})^k = \sum_{n=0}^{\infty} a_n^{(k)}(\lambda) \mathbf{J}^n ,$$

and it follows from (4) that

$$[R(\lambda, L)]^k y = \mathbf{P}[\mathbf{R}(\lambda, \mathbf{L})]^k y, \qquad y \in H,\ k \geqq 0,\ \lambda > 0 . \tag{9}$$

According to Yosdia's proof of the Hille-Yosida theorem (see [1]),

$$T_t = \operatorname*{st.lim}_{\lambda\to\infty} \exp(tB_\lambda) \text{ and } \mathbf{U}_t = \operatorname*{st.lim}_{\lambda\to\infty} \exp(t\mathbf{B}_\lambda), \qquad t \geqq 0 , \tag{10}$$

where

$$B_\lambda = \lambda^2 R(\lambda, L) - \lambda I \text{ and } \mathbf{B}_\lambda = \lambda^2 R(\lambda, \mathbf{L}) - \lambda \mathbf{I} .$$

Thus for $y \in H$ the relation (9) implies

$$\exp(tB_\lambda) y = \mathbf{P} \exp(t\mathbf{B}_\lambda) y, \qquad y \in H,\ \lambda > 0 ,$$

and this together with (10) gives (7).

It remains to prove that $\mathbf{H}$ is the same as

$$\mathbf{H}_0 = \text{closed linear extension of } [\mathbf{U}_t H;\ -\infty < t < \infty] .$$

Let $\mathbf{P}_0$ be the projection of $\mathbf{H}$ onto $\mathbf{H}_0$. Then clearly $\mathbf{U}_t\mathbf{H}_0 \subset \mathbf{H}_0$ for all real $t$, and since $\mathbf{U}_t^* = \mathbf{U}_{-t}$ the same is true of the orthogonal complement to $\mathbf{H}_0$. As a consequence $\mathbf{P}_0\mathbf{U}_t = \mathbf{U}_t\mathbf{P}_0$ for all real $t$. Hence for $y \in \mathfrak{D}(\mathbf{L})$

$$\mathbf{P}_0\mathbf{L}y = \lim_{\delta\to 0+} \delta^{-1}(\mathbf{P}_0\mathbf{U}_\delta y - \mathbf{P}_0 y) = \lim_{\delta\to 0+} \delta^{-1}(\mathbf{U}_\delta\mathbf{P}_0 y - \mathbf{P}_0 y) = \mathbf{L}\mathbf{P}_0 y .$$

Thus $\mathbf{P}_0$ commutes with $\mathbf{L}$ and hence with $\mathbf{J}$. But since $H$ is obviously contained in $\mathbf{H}_0$ we have

$$\mathbf{J}^n H = \mathbf{J}^n \mathbf{P}_0 H = \mathbf{P}_0 \mathbf{J}^n H \subset \mathbf{H}_0 .$$

The minimal property of $\mathbf{H}$ asserted in Theorem B therefore implies that $\mathbf{H} = \mathbf{H}_0$. This concludes the proof of the theorem.

It should be noted that since $i\mathbf{L}$ is self-adjoint, the largest restriction to $H$ of $i\mathbf{L}$ will be symmetric. On the other hand if $iL$ is symmetric then it is easily verified that $J$ is an isometry and hence that $\mathbf{J}$ is an extension of $J$; in this case then $\mathbf{L}$ will be an extension of $L$. However in general if $u \in H$ and $y = \mathbf{J}u + u$, then $z = \mathbf{P}y = Ju + u \in \mathfrak{D}(L)$

and $LPy = \mathbf{P}\mathbf{L}y$; each $z \in \mathfrak{D}(L)$ can be so represented. A simple example shows that $\mathfrak{D}(L) \cap H$ may contain only the zero element.[2]

## REFERENCES

1. E. Hille and R. S. Phillips, *Functional analysis and semi-groups*, Amer. Math. Soc. Coll. Publ., XXXI (1957).

2. R. S. Phillips, *Dissipative operators and hyperbolic systems of partial differential equations*, Trans. Amer. Math. Soc., (1959).

3. F. Riesz and B. Sz.-Nagy, *Leçons d'analyse fonctionnelle*, 3rd edition Budapest (1955).

4. B. Sz.-Nagy, *Transformations de l'espace de Hilbert. fonctions de type positif sur un groupe*, Acta Sci. Math. Szeged, **15** (1954) 104–114.

THE UNIVERSITY OF CALIFORNIA,
LOS ANGELES

[2] Suppose $H$ is one-dimensional and $T_t = \exp(-t)$. The Sz.-Nagy construction for $\mathbf{H}$ in Theorem B then results in $\mathbf{H} = l_2$, the space of complex-valued sequences $y = \{\eta_n; -\infty < n < \infty\}$ with

$$(y, z) = \sum_{n=-\infty}^{\infty} \bar{\eta}_n \bar{\zeta}_n ,$$

$\mathbf{J}\{\eta_n\} = \{\eta_{n-1}\}$, and $\mathbf{P}\{\eta_n\} = \{\eta'_n\}$ ($\eta'_0 = \eta_0$; $\eta'_n = 0$ for $n \neq 0$). Then relation (8) as applied to $\mathbf{J}$ and $\mathbf{L}$ asserts that for each $\{\eta_n\} \in \mathfrak{D}(\mathbf{L})$ there is a $\{\mu_n\} \in \mathbf{H}$ such that

$$2\eta_n = \mu_{n-1} + \mu_n, \quad 2[\mathbf{L}\{\eta_n\}]_n = \mu_{n-1} - \mu_n .$$

If we also require that $\{\eta_n\} \in H$, then $\mu_{n-1} + \mu_n = 0$ for all $n \neq 0$ and this together with the condition $\sum |\mu_n|^2 < \infty$ implies that $\mu_n = 0$ for all $n$. It follows that $\mathfrak{D}(L) \cap H = \theta$.

# A GENERALIZATION OF ATOMIC BOOLEAN ALGEBRAS

R. S. PIERCE

**1. Introduction.** A Boolean algebra $B$ is called atomic if every non-zero element of $B$ contains an atom. A variant of this definition is the equivalence: $B$ is atomic if and only if $B$ contains a dense (i.e., coinitial in $B - \{0\}$) subset which is totally unordered. In this paper, we will investigate the properties of Boolean algebras which contain dense subsets of somewhat more general order type than the totally unordered sets.

DEFINITION 1.1. Let $\alpha$ be an infinite cardinal number. A partially ordered set $P$ will be called *$\alpha$-compact* if $P$ is closed under finite meets, contains a zero element and satisfies the condition that if $M \subseteqq P$ has cardinality $\leqq \alpha$ and no finite subset of $M$ has zero meet, then $M$ has a non-zero lower bound in $P$.

The use of the term "compact" is of course motivated by the topological analogy.

DEFINITION 1.2. A Boolean algebra $B$ will be called *$\alpha$-atomic* if $B$ contains a dense subset which is $\alpha$-compact.

Since a totally unordered set becomes $\alpha$-compact (for all $\alpha$) if a zero element is adjoined to it, an atomic Boolean algebra is $\alpha$-atomic for all cardinals $\alpha$.

The organization of the paper is as follows. Section two is devoted to the construction of examples of $\alpha$-atomic Boolean algebras. In section three, some properties of $\alpha$-atomic Boolean algebras are proved. Section four presents a representation theorem for $\alpha$-atomic algebras.

Throughout the paper, $\alpha$ will denote a fixed infinite cardinal number. The abbreviation $\alpha$-B.A. will be used for $\alpha$-complete Boolean algebra. The terms $\alpha$-subalgebra, $\alpha$-ideal, $\alpha$-homorphism, $\alpha$-field, etc. have their usual meanings. Thus, an $\alpha$-homorphism of an $\alpha$-B.A. is a homomorphism preserving $\alpha$-joins; an $\alpha$-subalgebra of an $\alpha$-B.A. is a subalgebra closed under formation of $\alpha$-joins in the enveloping algebra. It is sometimes convenient to use the symbol $\infty$ in place of $\alpha$ with the meaning that the corresponding property is to hold for all cardinals.

The lattice operations of join, meet and complement are designated by $\vee$, $\wedge$, and $(')$ respectively. The symbols 0 and $u$ denote the zero and unit in a Boolean algebra. Set operations are indicated by rounded symbols: $\cap$, $\cup$ and $\subseteqq$ stand for intersection, union and inclusion

Received July 24, 1957.

respectively. The empty set is denoted by $\emptyset$. The symbol $|A|$ represents the cardinality of the set $A$. For any cardinal number $\alpha$, the smallest cardinal greater than $\alpha$ is denoted $\alpha^+$.

**2. Examples.** We have already observed that any atomic Boolean algebra is $\infty$-atomic. The converse is also true.

THEOREM 2.1. *A Boolean algebra is $\infty$-atomic if and only if it is atomic.*

*Proof.* Let $B$ be $\infty$-atomic. If $|B| = \alpha$, it is possible to find a dense subset $D$ in $B$ which is $\alpha$-complete. Let $M$ be a maximal dual ideal in $D$. Then $M$ has the finite meet property and $|M| \leqq \alpha$. Hence, $M$ has a non-zero lower bound $a$ in $D$. By the maximality of $M$, it is clear that $a$ is an atom of $B$. Since, by Zorn's lemma, every non-zero element of $D$ is contained in a maximal dual ideal of $D$, it follows that every element of $D$ contains an atom. But $D$ is dense in $B$ so every element of $B$ contains an atom. Thus $B$ is atomic.

In order to construct an $\alpha$-atomic B.A., it is enough to exhibit an $\alpha$-compact partially ordered set $P$ which is disjunctive, that is, satisfies the condition that for any $a \nleqq b$, there exists $c \in P$ such that $0 \neq c \leqq a$ and $c \wedge b = 0$. Indeed, any disjunctive partially ordered set can be imbedded as a dense subset in a complete Boolean algebra (see [1]), and if the partially ordered set is $\alpha$-compact, then the B.A. will necessarily be $\alpha$-atomic. This complete B.A. is determined up to isomorphism by the disjunctive partially ordered set. In fact, a more precise statement is true.

LEMMA 2.2. *Let $B_1$ and $B_2$ be complete Boolean algebras. Let $P_1$ and $P_2$ be dense subsets of $B_1$ and $B_2$ which are closed under meets. Suppose $\varphi$ is an isomorphism of $P_1$ on $P_2$. Then $\varphi$ has a unique extension to an isomorphism of $B_1$ on $B_2$.*

This result is proved in [1] for example.

We describe a fairly general method of constructing partially ordered sets which are disjunctive and $\alpha$-compact.

Let $I$ be a non-empty index set. Let $\{X_i \mid i \in I\}$ be a collection of sets, each containing at least two elements. Put $X = \prod_{i\in I} X_i$. Suppose $\mathfrak{M}$ is a given non-empty collection of subsets of $I$ with the properties

(a) $\mathfrak{M}$ is closed under finite unions,

(b) $\mathfrak{M}$ is $\alpha$-directed: any subcollection of $\mathfrak{M}$ with cardinality $\leqq \alpha$ has an upper bound in $\mathfrak{M}$.

Let $M \in \mathfrak{M}$ and $\varphi \in \prod_{i\in M} X_i$, that is, $\varphi$ is a function on $M$ such that $\varphi(i) \in X_i$ for all $i \in M$. Denote

$$A_{M,\varphi} = \{\chi \in X \mid \chi|_M = \varphi\} .$$

Finally, let $\Omega = \Omega(I, X_i, \mathfrak{M})$ be the collection of all $A_{M,\varphi}$, together with the empty set.

LEMMA 2.3. *The set $\Omega$, ordered by inclusion, is an $\alpha$-compact, disjunctive partially ordered set.*

*Proof.* Observe that $A_{M,\varphi} \subseteqq A_{N,\psi}$ if and only if $M \supseteqq N$ and $\varphi|_N = \psi$. On the other hand, $A_{M,\varphi} \cap A_{N,\psi} = \emptyset$ if and only if there exists $i \in M \cap N$ such that $\varphi(i) \neq \psi(i)$. The fact that $\Omega$ is meet closed and disjunctive is a routine consequence of these observations and property (a) of $\mathfrak{M}$. The fact that $\Omega$ is $\alpha$-directed follows from property (b) of $\mathfrak{M}$.

A special case of this contruction is particularly interesting. If $I$ is arbitrary, each $X_i$ is a two element set and $\mathfrak{M}$ consists of all subsets $A$ of $I$ with $|A| \leqq \alpha$, then the conditions for the application of Lemma 2.3 are fulfilled. The $\alpha$-compact partially ordered set $\Omega$ in this case is completely determined by the cardinal numbers $\alpha$ and $\beta = |I|$. Thus we can designate this $\Omega$ simply as $\Phi_{\alpha\beta}$. Let $B$ be a complete Boolean algebra containing $\Phi_{\alpha\beta}$ as a dense subset and define $F_{\alpha\beta}$ to be the smallest $\alpha$-subalgebra of $B$ containing $\Phi_{\alpha\beta}$. It is clear from 2.2 that $F_{\alpha\beta}$ is determined up to isomorphism by $\Phi_{\alpha\beta}$.

THEOREM 2.4. *The Boolean algebra $F_{\alpha\beta}$ is isomorphic to the $\alpha$-field of subsets of $X$ generated by $\Phi_{\alpha\beta}$ and is a free $\alpha$-representable algebra with $\beta$ generators.*

REMARK. A Boolean algebra is called $\alpha$-representable if it is the $\alpha$-homomorphic image of an $\alpha$-field (see [2]). The fact that the class of all such algebras is equationally definable and therefore admits free algebras has been investigated in [5]. Indeed, Sikorski proves in [7] that the $\alpha$-field generated by $\Phi_{\alpha\beta}$ is a free $\alpha$-representable algebra with $\beta$ generators. Thus it is only necessary to prove the first assertion of 2.4.

*Proof.* Let $B_1$ be the $\alpha$-field (in $X$) generated by $\Phi_{\alpha\beta}$. Note that $\Phi_{\alpha\beta}$ is closed under $\alpha$-intersections and that the complement of any set of $\Phi_{\alpha\beta}$ is a union of sets of $\Phi_{\alpha\beta}$. Hence, by Lemma 5.2 of [4] (quoted in (4.3) below), $\Phi_{\alpha\beta}$ is dense in $B_1$. It follows from 2.2 that $B_1$ is isomorphic to $F_{\alpha\beta}$.

COROLLARY 2.5. *Every $\alpha$-representable Boolean algebra is an $\alpha$-homomorph of an $\alpha$-atomic, $\alpha$-field.*

It is easy to see that if $\beta \leqq \alpha$, then $\Phi_{\alpha\beta}$ is atomic and hence so is any Boolean algebra containing $\Phi_{\alpha\beta}$ as a dense subset. However, if $\beta > \alpha$, then a Boolean algebra $B$ containing $\Phi_{\alpha\beta}$ as a dense subset is not even $\alpha^+$-atomic. This is a consequence (by (3.3) below) of the stronger result that $B$ is not $\alpha^+$-distributive (see [9]).

THEOREM 2.6. *Let $B$ be a Boolean algebra containing $\Phi_{\alpha\beta}$ as a dense subset. Suppose also that $\beta > \alpha$. Then $B$ is not $\alpha^+$-distributive.*

*Proof.* Let $J \subseteqq I$ have cardinality $\alpha^+$. For $i \in I$, denote $a_{ik} = \{\chi \in X \mid \chi(i) = x_{ik}\}$ $(k = 1, 2)$, where $X_i = \{x_{i1}, x_{i2}\}$. Then $a_{i1} \vee a_{i2} = u$ (since no element of $\Phi_{\alpha\beta}$ is disjoint from both $a_{i1}$ and $a_{i2}$). But if $M \in \mathfrak{M}$, then $J \nsubseteqq M$, so there exists $j \in J - M$. For this index, $\bigcap_{i \in M} a_{i,\varphi(i)} \nsubseteqq a_{jk}$ $(k = 1, 2)$ for all $\varphi \in 2^M$. Thus

$$\bigwedge_{i \in J}(a_{i1} \vee a_{i2}) = u > 0 = \bigvee_{\varphi}\left(\bigwedge_{i \in J} a_{i,\varphi(i)}\right),$$

so $B$ is not $\alpha^+$-distributive.

REMARK. The referee has pointed out that 2.6 is related to the results in Scott's paper [6]. Scott constructs a Boolean algebra $B_\alpha$ (for each regular cardinal $\alpha$) which, when $\alpha = \beta^+$, is equivalent to the completion of $F_{\beta\beta^+}$.

## 3. Properties of $\alpha$-atomic algebras.

The term "covering" will be used to designate a subset of a Boolean algebra whose least upper bound is the unit element.

LEMMA 3.1. *Let $B$ be an $\alpha$-complete, $\alpha$-atomic Boolean algebra. Then $B$ has the following property:*

*(*) if $\{A_\sigma \mid \sigma \in S\}$ is a family of coverings of $B$ such that $|S| \leqq \alpha^+$ and if $b \neq 0$ in $B$, then there is a choice function $\varphi$ on $S$ such that $\varphi(\sigma) \in A_\sigma$ with the property that if $T \subseteqq S$ and $|T| \leqq \alpha$, then*

$$b \wedge \bigwedge_{\sigma \in T} \varphi(\sigma) \neq 0 .$$

*Proof.* Let $T \subseteqq B$ be dense and $\alpha$-compact. Denote by $\lambda$ the least ordinal of cardinality $\alpha^+$. We can assume that $S$ consists of the ordinals $\sigma < \lambda$. By transfinite induction, define functions $f: S \to T$ and $\varphi$ on $S$ with $\varphi(\sigma) \in A_\sigma$ having properties

(i) $\sigma < \tau$ implies $0 < f(\tau) \leqq f(\sigma) \leqq b$,
(ii) $f(\sigma) \leqq \varphi(\sigma)$.

These are constructed in the following way. Assume $f(\sigma)$ has been defined for all $\sigma < \tau$, where $\tau < \lambda$. By $\alpha$-compactness, $c = \bigwedge_{\sigma<\tau} f(\sigma) \neq 0$. If $\tau = 1$, then $c = u$. Let $\varphi(1) \in A_1$ satisfy $\varphi(1) \wedge b \neq 0$. Such an element exists, since $b = b \wedge u = b \wedge \bigvee A_1 = \bigvee\{b \wedge a | a \in A_1\}$. Let $f(1) \in T$ be chosen arbitrarily, satisfying $0 \neq f(1) \leqq \varphi(1) \wedge b$. If $\tau > 1$, then $c \leqq b$. Choose $\varphi(\tau) \in A_\tau$ so that $\varphi(\tau) \wedge c \neq 0$. As before, some element of $A_\tau$ will satisfy this requirement. Using the fact that $T$ is dense, it is possible to find $f(\tau) \in T$ such that $0 < f(\tau) \leqq \varphi(\tau) \wedge c$. With this construction, it is clear that (*i*) and (*ii*) are fulfilled.

Now if $T \subseteqq S$ and $|T| \leqq \alpha$, then since $\lambda$ is regular, there exists $\eta < \lambda$ such that $\tau < \eta$ for all $\tau \in T$. Hence,

$$b \wedge \bigwedge_{\sigma \in T}\varphi(\sigma) \geqq b \wedge \bigwedge_{\tau<\eta}\varphi(\tau) \geqq b \wedge \bigwedge_{\tau<\eta}f(\tau) \geqq f(\eta) > 0 .$$

This is the required conclusion.

COROLLARY 3.2. *Any $\alpha^+$-complete, $\alpha$-atomic* B.A. *is $\alpha^+$-representable.*

*Proof*. It is easy to see that the condition (*) of (3.2) implies Smith's property ($P_{\alpha+}$) (see [8]). Hence, (3.2) follows from Theorem 4.1 of [8].

COROLLARY 3.3. *Any $\alpha$-complete, $\alpha$-atomic* B.A. *is $(\alpha, \infty)$-distributive.*

For the definition of $(\alpha, \infty)$-distributivity, the reader is referred to [9] or [4]. The property (*), together with (2.3) of [4] implies (3.3).

If $B$ is a complete B.A. containing $\Phi_{\alpha\beta}$ as a dense subset, then $B$ is $\alpha^+$-representable by (3.2). If $\beta > \alpha$, then $B$ is $(\alpha, \infty)$-distributive, but not $\alpha^+$-distributive by (2.6). Hence, $B$ is not $2^{\alpha+}$-representable. If we admit the generalized continuum hypothesis, this means that $B$ is not $\gamma$-representable for any $\gamma > \alpha^+$, which partially answers a question raised by Chang [2; p. 213].

In [3], the author conjectures that any $\alpha$-distributive $2^\alpha$-complete B.A. is $2^\alpha$-representable. This conjecture now appears rather unlikely in view of Theorem 3.4 of [8], since its validity, together with the generalized continuum hypothesis, would imply a positive answer to Souslin's problem. However, *for an $\alpha$-subalgebra $B'$ of an $\alpha$-atomic Boolean algebra $B$, it is true that $2^\alpha$-completeness implies $\alpha^+$-representability.* For $B'$ is also an $\alpha$-subalgebra of $\overline{B}$, the normal completion of $B$. But $\overline{B}$ is $\alpha^+$-representable and $\alpha$-distributive by (3.2) and (3.3). Therefore, to see that $B'$ is $\alpha+$-representable, we have only to notice that it must actually be an $\alpha+$-subalgebra of $\overline{B}$.

THEOREM 3.4. *Let $\overline{B}$ be an $\alpha$-distributive, $2^\alpha$-complete Boolean algebra. Suppose $B$ is an $\alpha$-subalgebra of $\overline{B}$ which is $2^\alpha$-complete. Then $B$ is a $2^\alpha$-subalgebra of $\overline{B}$.*

REMARK. This theorem is well known for fields of sets. (See [9], Theorem 3.10 and the references given there.) We will give a proof, since the result seems to have been overlooked in [3] and [9].

*Proof.* Let $Q \subseteqq B$ and $|Q| \leqq 2^{\alpha}$. Let $b = \text{l.u.b. } Q$ in $B$ and $\bar{b} = \text{l.u.b. } Q$ in $\bar{B}$. Then $b \geqq \bar{b}$. Assume $b > \bar{b}$ and set $\bar{c} = b \wedge \bar{b}'$. Let $B_1$ be the $\alpha$-distributive Boolean algebra: $\{\bar{d} \in \bar{B} | \bar{d} \leqq \bar{c}\}$. Denote by $\bar{h}$ the mapping $\bar{B} \to B_1$ given by $\bar{h}(\bar{a}) = \bar{a} \wedge \bar{c}$. Evidently $\bar{h}$ is a complete homorphism of $\bar{B}$ and hence the restriction $h$ of $\bar{h}$ to $B$ is an $\alpha$-homomorphism. The image $h(B)$ is an $\alpha$-subalgebra of $B_1$ and therefore is $\alpha$-distributive. By Theorem 3.6 of [9] or (6.5) of [3], $h$ is a $2^{\alpha}$-homomorphism. But obviously $Q$ is contained in the kernel of $h$. Thus, $h(b) = \text{l.u.b.}_B\{h(a) | a \in Q\} = 0$. But $\bar{c} \leqq b$, so $\bar{c} = u \wedge \bar{c} = \bar{h}(u) = \bar{h}(b \vee \bar{c}) = \bar{h}(b) \vee \bar{h}(\bar{c}') = h(b) \vee (\bar{c}' \wedge \bar{c}) = 0$, a contradiction. Thus, $b = \bar{b}$, which is the desired conclusion.

COROLLARY 3.5. *Let $\beta$ be weakly attainable from the infinite cardinal $\alpha$. Suppose $\bar{B}$ is a $\beta$-distributive, $2^{\beta}$-complete* B.A. *and $B$ is an $\alpha$-subalgebra of $\bar{B}$ which is $2^{\beta}$-complete. Then $B$ is a $2^{\beta}$-subalgebra of $\bar{B}$.*

*Proof.* Clearly, if $\xi$ is a singular cardinal and $B$ is an $\eta$-subalgebra of $\bar{B}$ for all $\eta < \xi$, then $B$ is a $\xi$-subalgebra. Using this fact, 3.5 follows from 3.4 by transfinite induction.

**4. The representation theorems.** Not every $\alpha$-atomic B.A. is an $\alpha$-field, since the normal completion of an atomless $\alpha$-field will not, in general, be an $\alpha$-field ($\alpha$ being infinite). However, we will prove that every $\alpha$-atomic algebra is a dense subalgebra of the normal completion of an $\alpha$-field. Of course, since any B.A. is a dense subalgebra of its normal completion, it suffices to prove that any complete, $\alpha$-atomic B.A. contains a dense subalgebra $B_0$ which is isomorphic to an $\alpha$-field.

LEMMA 4.1. *Let be an $\alpha$-complete, $\alpha$-atomic* B.A. *Then $B$ contains a dense, $\alpha$-compact subset which is closed under $\alpha$-meets.*

*Proof.* By Definition 1.2, $B$ contains a dense $\alpha$-compact subset $T$. The set of all $\alpha$-meets of elements of $T$ will clearly be a dense, $\alpha$-compact subset of $B$ which is closed under $\alpha$-meets.

LEMMA 4.2. *Let $F$ be a disjunctive, $\alpha$-compact partially ordered set which is closed under $\alpha$-meets. Let $X$ be the set of all proper maximal dual ideals of $F$. For $a \in F$, let $\varphi(a) = \{P \in X | a \in P\}$. Then $\varphi$ is an $\alpha$-meet preserving, order isomorphism of $F$ into the Boolean algebra of* *all subsets of $X$. The image $\mathfrak{f}$ of $\varphi$ has the property that the complement of any set of $\mathfrak{f}$ is a union of sets of $\mathfrak{f}$.*

*Proof.* Since $F$ is $\alpha$-compact and closed under $\alpha$-meets, every maximal dual ideal of $F$ is also closed under $\alpha$-meets. Hence, $\varphi$ preserves $\alpha$-meets by the usual argument.

If $c \neq 0$ in $F$, then $\varphi(c) \neq \emptyset$, since every non-zero element is contained in a proper maximal dual ideal. Since $F$ is disjunctive, $a \nleqq b$ implies the existence of $c \in F$ with $0 \neq c \leqq a$ and $b \wedge c = 0$. Hence, $\emptyset \neq \varphi(c) \subseteqq \varphi(a)$ and $\varphi(b) \cap \varphi(c) = \emptyset$. Therefore, $\varphi(a) \nsubseteqq \varphi(b)$.

If $P \in X - \varphi(a)$, then $a \notin P$. By the maximality of $P$, there exists $b \in P$ such that $a \wedge b = 0$. Then $P \in \varphi(b) \subseteqq (\varphi(a))^c$. This shows that $(\varphi(a))^c$ is a union of sets of $\mathfrak{f}$.

For the proof of the main theorem of this section, we need a known result.

LEMMA 4.3. *Let $\mathfrak{f}$ be a family of subsets of a set $X$ with the properties that $\mathfrak{f}$ is closed under $\alpha$-intersections and the complement of any set of $\mathfrak{f}$ is a union of sets of $\mathfrak{f}$. Let $\mathfrak{L}$ be the $\alpha$-field generated by $\mathfrak{f}$. Then $\mathfrak{f}$ is dense in $\mathfrak{L}$.*

The proof of this fact can be found in [4].

THEOREM 4.5. *Let $B$ be an $\alpha$-atomic Boolean algebra. Then $B$ is isomorphic to a dense subalgebra of the normal completion of an $\alpha$-atomic $\alpha$-field of sets.*

*Proof.* Let $\overline{B}$ be the normal completion of $B$. Then $\overline{B}$ is $\alpha$-atomic. By (4.1), $\overline{B}$ contains a dense, $\alpha$-compact subset $F$ which is closed under $\alpha$-meets. Since $F$ is dense in the Boolean algebra $\overline{B}$, $F$ is disjunctive. By (4.2), there is an $\alpha$-isomorphism $\varphi$ of $F$ onto a family $\mathfrak{f}$ of subsets of a set $X$ with the two properties of (4.3). Let $\mathfrak{L}$ be the $\alpha$-field generated by $\mathfrak{f}$ and let $\mathfrak{L}_0$ be the normal completion of $\mathfrak{L}$. By (4.3), $\mathfrak{f}$ is dense in $\mathfrak{L}$ and hence in $\mathfrak{L}_0$. Consequently, by (2.2), $\varphi$ extends uniquely to an isomorphism of $\overline{B}$ on $\mathfrak{L}_0$. The restriction of this extension is an isomorphism of $B$ onto a dense sub-algebra of the normal completion of the $\alpha$-atomic $\alpha$-field $\mathfrak{L}$.

COROLLARY 4.6. *Any complete, $\alpha$-atomic Boolean algebra is isomorphic to the normal completion of an $\alpha$-atomic $\alpha$-field.*

## REFERENCES


1. J. R. Büchi, *Die Boolesche Partialordnung*, Portugaliae Mathematica **7** (1948), 118-177.
2. C. C. Chang, *On the representation of α-complete Boolean algebras*, Trans. Amer. Math. Soc. **85** (1957), 208-218.

3. R. S. Pierce, *Distributivity in Boolean algebras*, Pacific J. Math. **7** (1957), 983-992.
4. R. S. Pierce, *Distributivity and the normal completion of Boolean algebras*, Pacific J. Math. **8** (1958), 133-140.
5. D. Scott and A. Tarski, *Metamathematical proofs of some Boolean representation theorems*, in preparation.
6. D. Scott, *The independence of certain distributive laws in Boolean algebras*, Trans. Amer. Math. Soc. **84** (1957), 258-261.
7. R. Sikorski, *A note to Rieger's paper "On Free $\alpha\xi$-complete Boolean Algebras"*, Fund. Math. **38** (1951), 53-54.
8. E. C. Smith, Jr., *A distributivity condition for Boolean algebras*, Ann. of Math. **64** (1956), 551-561.
9. E. C. Smith, Jr. and A. Tarski, *Higher degrees of distributivity and completeness in Boolean algebras*, Trans. Amer. Math. Soc. **84** (1957), 230-257.

UNIVERSITY OF WASHINGTON

# ANALYTIC CONTINUATION OF MEROMORPHIC FUNCTIONS IN VALUED FIELDS

J. B. ROBERTS

In this paper* we consider analytic continuation of power series by matrix methods in arbitrary fields complete with respect to a valuation. In the complex field continuation can generally be achieved by a formal expansion of the given power series about a point in its circle of convergence. The new series (with power series coefficients) generally exists and converges over a circle extending beyond the circle of convergence of the original series.

When the field is non-Archimedean however the new circle of convergence is always contained in the old. Hence in this case we need have recourse to a summability method. In this paper we consider a certain class of matrix methods which can be applied to the power series coefficients appearing in the formal expansion of the original series about points outside the original circle of convergence. The methods will be applicable in Archimedean or non-Archimedean fields.

The work here is based upon Chapter 3 of the author's PhD dissertation written under the direction of Prof. G. K. Kalisch at the University of Minnesota in 1955.

**1. Notations and definitions.** Throughout the paper $k$ shall be a field which is complete with respect to a valuation, denoted by $|\ |$. Unless stated explicitely to the contrary the valuation may be either Archimedean or non-Archimedean. It is useful to note that, by a theorem of Ostrowski, if the valuation is Archimedean then $k$ is topologically isomorphic with the real or complex numbers.

We shall designate the collection of all infinite series with terms in $k$ by S. Further we introduce an operation, the Cauchy product, into $S$. If $C = \sum_{i=0}^{\infty} c_i$ and $C' = \sum_{i=0}^{\infty} c'_i$ are in $S$ then the Cauchy product $CC'$ is defined by

$$CC' = \sum_{i=0}^{\infty} \sum_{j=0}^{i} c_j c'_{i-j} .$$

This product is clearly in $S$; so $S$ is closed relative to this multiplication.

The subset of $S$ consisting of all unconditionally convergent series will be denoted by $T$. When $k$ is non-Archimedean $T$ coincides with the

* This paper was originally accepted by the Trans. Amer. Math. Soc., received by the editors of the Trans. Amer. Math. Soc. January 16, 1957, in revised form April 18, 1958. The author wishes to express his thanks to the referee who through his extensive comments on the first version has changed the character of the whole paper and has increased its generality in certain respects.

set of all convergent series since in this case a series converges if and only if its $n$th term goes to 0. When $k$ is Archimedean $T$ coincides with the collection of all absolutely convergent series. A theorem of Mertens in the Archimedean case (which remains true in the non-Archimedean case) assures us that $T$ is closed relative to the multiplication defined in $S$. Further by the same theorem if $C, C'$ converge respectively to $c, c'$ then $CC'$ converges to $cc'$.

The set of series in $T$ which converge to non-zero limits will be denoted by $T^*$. From the last sentence of the preceding paragraph we see that $T^*$ is closed under multiplication.

The set of infinite matrices $(a_{ij})$, $i = 0, 1, 2, \cdots; j = 0, 1, 2, \cdots$ where $a_{ij}$ is in $k$ for all $i, j$ will be denoted by $M$. We introduce into $M$ two operations—addition and multiplication. Addition is unrestictedly defined by the following:

$$m_1 = (a_{ij}), m_2 = (b_{ij}) \text{ then } m_1 + m_2 = (a_{ij} + b_{ij}) .$$

Clearly $m_1 + m_2$ is in $M$.

Multiplication is not unrestrictedly defined. We have the following definition ($m_1$ and $m_2$ as above):

$$m_1 m_2 = (c_{ij}) \text{ providing } c_{ij} = \sum_{q=0}^{\infty} a_{iq} b_{qj} \text{ converges for all } i, j .$$

We shall be interested in mappings from subsets of $S$ into $M$. A multiplicative homomorphism from a subset $V$ of $S$ into $M$ is a mapping $f$ of $V$ into $M$ such that when $C_1$, $C_2$ and $C_1C_2$ are in $V$ then $f(C_1)f(C_2)$ is defined and $f(C_1)f(C_2) = f(C_1C_2)$.

## 2. The matrices $A_C$ and $B_C$.

DEFINITION 1. Let $C = \sum_{i=0}^{\infty} c_i$ be in $S$.

(a) $B_C = (b_{ij})$ where $b_{ij} = c_{j-i}$ and $c_{j-i}$ is taken to be 0 when $j < i$;

(b) If $C$ converges to $c \neq 0$ then $A_C = B_C(c)^{-1}$ where $(c)$ is the diagonal matrix with all diagonal elements $c$.

LEMMA 1. *The map $C \to B_C$ is a multiplicative homomorphism of $S$ into $M$.*

*Proof.* Let $C = \sum_{i=0}^{\infty} c_i$, $C' = \sum_{i=0}^{\infty} c'_i$ be in $S$. Then $CC' = \sum_{i=0}^{\infty} \bar{c}_i$ where $\bar{c}_i = \sum_{j=0}^{i} c_j c'_{i-j}$. Thus $B_{CC'} = (\bar{c}_{j-i})$. Since $B_C$ and $B_{C'}$ each have only finitely many non-zero terms in each column $B_C B_{C'}$ is defined. Further $B_C B_{C'} = (d_{ij})$ where

$$d_{ij} = \sum_{q=0}^{\infty} c_{q-i} c'_{j-q} = \sum_{q=i}^{j} c_{q-i} c'_{j-q} = \sum_{s=0}^{j-i} c_s c'_{(j-i)-s} = \bar{c}_{j-i} .$$

Hence $B_C B_{C'} = B_{CC'}$ and the lemma is proved.

LEMMA 2. *Let $C$, $C'$ be in $S$ and suppose they converge respectively to the non-zero sums $c$, $c'$. Then $A_C A_{C'}$ exists and $A_C A_{C'} = A_{CC'}$ providing $CC'$ converges to $cc'$.*

*Proof.* $A_C = B_C(c)^{-1}$, $A_{C'} = B_{C'}(c')^{-1}$ and therefore

$$A_C A_{C'} = B_C(c)^{-1} B_{C'}(c')^{-1} = B_{CC'}(cc')^{-1} = A_{CC'} .$$

COROLLARY. *The map $C \to A_C$ is a multiplicative homomorphism of $T^*$ into $M$.*

We now introduce a norm into $T$ and two topologies into $M$.

DEFINITION 2. The *norm* of $C$, denoted by $|C|_T$, for $C = \sum_{i=0}^{\infty} c_i$ in $T$ is defined by:

$$|C|_T = \begin{cases} \max |c_i| & \text{for } k \text{ non-Archimedean;} \\ \sum_{i=0}^{\infty} |c_i| & \text{for } k \text{ Archimedean .} \end{cases}$$

By restricting our $C$ to be in $T$ we insure that this norm is defined. The following properties are valid for arbitrary $k$.

$$|C + C'|_T \leq |C|_T + |C'|_T ;$$
$$|CC'|_T \leq |C|_T |C'|_T ;$$
$$|aC|_T = |a| \; |C|_T \text{ for a in } k .$$

If $k$ is non-Archimedean the first two properties can be strengthened to read

$$|C + C'|_T \leq \max(|C|_T, |C'|_T) ;$$
$$|CC'|_T = |C|_T |C'|_T .$$

Defining addition in $T$ to be componentwise addition we see that $T$ is a normed ring.

DEFINITION 3. (a) The *weak topology* in $M$ is the topology induced on $M$ by making the sequence $m_n = (a_{ij}^{(n)})$ of matrices converge to the matrix $(a_{ij})$ if and only if for all $i$, $j$ we have $a_{ij}^{(n)} \to a_{ij}$. When this is true we say that the sequence $m_n$ *converges weakly* to $(a_{ij})$.

(b) If, for an arbitrary positive real number $r$, we denote the set of all matrices $(a_{ij})$ with $|a_{ij}| < r$, for all $i$, $j$, by $M_r$ then the set of $M_r$ gives a basis system for the open sets about the additive identity 0 in $M$. This induces on $M$ the topology of the additive group of $M$ and is called the *uniform topology*.

We note that addition and multiplication (when the latter is defined) are continuous in both topologies in $M$. Also if a sequence of matrices converges in the uniform topology it converges in the weak topology.

We shall denote by $\bar{M}$ the collection of matrices $m = (a_{ij})$ for which $\max |a_{ij}|$ exists. For $m$ in $\bar{M}$ we define $|m| = \max |a_{ij}|$. This induces the same topology on $\bar{M}$ which the uniform topology of $M$ induces on $\bar{M}$.

LEMMA 3. *The map $C \to B_C$ is a continuous map of $T$ into $M$ under either the uniform or weak topologies of $M$.*

*Proof.* Since $C$ is in $T$, $\max |c_i|$ exists and is $\leq |C|_T$. As $B_C = (c_{j-i})$ the norm of $B_C$, in $\bar{M}$, is given by

$$|B_C| = \max |c_{j-1}| = \max |c_i| \leq |C|_T .$$

Therefore the map of $T$ into $\bar{M}$ is continuous with respect to the norm topology of $\bar{M}$. Since this topology is induced by the uniform topology of $M$ this map is continuous relative to the uniform topology. This then implies continuity relative to the weak topology and the lemma is proved.

LEMMA 4. *The map $C \to A_C$ is a continuous map of $T^*$ into $M$ under either the uniform or weak topologies of $M$.*

*Proof.* $A_C = B_C(c)^{-1}$ where $c \neq 0$ is the sum of $C$. Since multiplication in $M$ is continuous in either topology as is the map $C \to B_C$ (by previous lemma) we need only show that the map $C \to (c)^{-1}$ is continuous. This is the product of three maps $C \to c \to c^{-1} \to (c)^{-1}$.

The first is continuous since it is an additive homomorphism and $|c| \leq |C|_T$. The second is a continuous map on $k^*$ (the non-zero elements of $k$). The third map is a ring isomorphism into $\bar{M}$ preserving norms. I.e. $|c^{-1}| = \max |c^{-1}| = |(c^{-1})|$. Hence this map is continuous into $\bar{M}$ relative to the norm in $\bar{M}$. As in the proof of Lemma 3 this concludes the proof.

We define the convergence of an infinite product $\prod_{n=1}^{\infty} C_n$, $C_n$ in $T$, in the usual way. That is, $\prod_{n=1}^{\infty} C_n$ converges providing $\lim_{q\to\infty} \prod_{n=1}^{q} C_n$ exists and is not the additive identity of $T$. Making use of the theorem:

$\prod_{n=1}^{\infty} C_n$, $C_n$ in $T$, converges if and only if $|1 - C_n|_T \to 0$ as $n \to \infty$ (where 1 is the multiplicative identity in $T$).

We deduce from Lemma 4 the following immediate consequence.

THEOREM 1. *Let $\sum_{n=1}^{\infty} C_n$ converge and suppose $C_n$ is in $T^*$ for all $n$. Then $\prod_{n=1}^{\infty} A_{C_n}$ converges relative to both weak and uniform topologies of $M$ and its limit is $A_{\prod_{n=1}^{\infty} C_n}$.*

**3. $T_2$ matrices and $C(x)$-continuation.** Each infinite matrix $m$ can be thought of as a mapping defined over a subset of $S$ and mapping this subset into $S$. In fact, let $m = (a_{ij})$ and suppose $C = \sum_{i=0}^{\infty} c_i$ is in $S$. Then if, for all $j$, the series $\sum_{i=0}^{\infty} c_i a_{ij}$ exists and equals $c'_j$ we shall say that the matrix $m$ maps $C$ onto $C' = \sum_{j=0}^{\infty} c'_j$. We shall write $mC = C'$.

(If we let $C^*$ be the "vector" $(c_0, c_1, c_2, \cdots)$ derived in the obvious way from $C$ then $C'^* = (mC)^* = C^* \cdot m$ where the right side is the ordinary matrix product of $C^*$ and $m$.)

When $C$ has sum $c$ then $C'$ has sum $c$ we call $m$ a $T_2$ matrix. Necessary and sufficient conditions in order that an infinite matrix be a $T_2$ matrix will be found in [2] for $k$ Archimedean and in [8ab] for $k$ non-Archimedean. (In the reference [2] the $T_2$ matrix is called an $\alpha$ matrix.)

Now suppose $C = \sum_{i=0}^{\infty} c_i$, $C' = \sum_{i=0}^{\infty} c'_i$ are in $T$ with sums $c, c'$ respectively. Then $CC'$ exists and

$$CC' = \sum_{j=0}^{\infty} \sum_{i=0}^{j} c'_i c_{j-i} = B_C C'$$

where $B_C$ is as defined in § 2.

Since in this case $CC'$ converges to $cc'$ we see that $B_C$ maps convergent series onto convergent series but alters the sum by a factor of $c$. Thus for $C$ in $T^*$, $A_C = B_C(c)^{-1}$ will map convergent series onto convergent series with the same sum. This proves the following.

LEMMA 5. *If $C$ is in $T^*$ then $A_C$ is a $T_2$ matrix.*

We wish now to consider series of functions. Let $C(x) = \sum_{i=0}^{\infty} c_i(x)$ and $U(x) = \sum_{i=0}^{\infty} u_i(x)$ where $x$ ranges over some subset $X$ of $k$. Suppose in addition that $C(x)$ is in $T^*$ for all $x$ in $D$ where $D$ is a subset of $X$. Further suppose there is a non-empty subset $\Delta$ of $D$ on which $U(x)$ converges. Then, by Lemma 5, $A_{C(x)}$ is a $T_2$ matrix for $x$ in $D$ and therefore transforms $U(x)$, for $x$ in $\Delta$, into a new series with the same sum. However it may be true that $A_{C(x)}U(x)$ is defined and converges for some $x$ in $D - \Delta$.

The sum function $u'(x)$, considered over the largest portion of $D$ on which $A_{C(x)}U(x)$ exists and converges, will be called the *$C(x)$-continuation* of $U(x)$ (or more accurately the $C(x)$-continuation of the sum function $u(x)$ of $U(x)$). The $C(x)$-continuation will be called *efficient* for $U(x)$ if there exists an $x$ in $D - \Delta$ for which $A_{C(x)}$ exists and converges.

In Archimedean fields it is possible for an infinite series to converge conditionally. If $C(x)$ converges conditionally for some $x$ then $A_{C(x)}$ is defined but is not necessarily a $T_2$ matrix (since the Cauchy product of conditionally convergent series may not converge). Considering $X$ now to be a topological space we can speak of the closure $\bar{D}$ of $D$ in $X$. Let $x$ be in $\bar{D}$ and suppose that when $y$ in $D$ converges to $x$ in certain prescribed ways then $C(x)$, $U(x)$, $A_{C(x)}U(x)$ converge respectively to

$$\lim_{y\to x} c(y),\ \lim_{y\to x} u(y),\ \lim_{y\to x} u'(y)$$

when these limits exist. When $x$ is in $D$ we know $u(x) = u'(x)$ so $A_{C(x)}U(x)$ converges when $U(x)$ does. Thus if $c(x)$, $u(x)$, $u'(x)$ are continuous over $D$ then for $x$ in $D$, whatever the prescribed ways $y$ in $D$ tends to $x$, we have $C(x)$, $U(x)$, $A_{C(x)}$, $U(x)$ converging respectively to

$$\lim_{y\to x} c(y),\ \lim_{y\to x} u(y),\ \lim_{y\to x} u'(y).$$

Let $D^*$ be the set of all $x$ in $\bar{D}$ for which $C(x)$, $U(x)$, $A_{C(x)}U(x)$ have the respective limits specified above as $y \to x$ in one of the prescribed ways. Then $D \subset D^* \subset \bar{D}$ when $c(x)$, $u(x)$, $u'(x)$ are continuous over $D$. The function $u'(x)$, considered over $D^*$, will be called a *generalized $C(x)$-continuation* of $U(x)$ (relative to the allowed modes of convergence of $y$ in $D$ to $x$ in $D^*$).

**4. Power series and the Weierstrass decomposition theorem.** In this section we take the $X$ of § 3 to be all of $k$ and suppose $C(x)$ and $U(x)$ to be power series about $\alpha$ in $k$. Then we may take, without loss of generality, the set $D$ to be a circle with center $\alpha$ from which have been excised all zeros of $C(x)$. Then $\Delta$ is the intersection of $D$ with some circle of center $\alpha$. When $k$ is non-Archimedean $\bar{D} = D$ and when $k$ is Archimedean $\bar{D}$ is the closed circle about $\alpha$ of the same radius as $D$. Thus (by Abel's theorem in the Archimedean case) if we prescribe $y$ in $D$ to converge to $x$ in $\bar{D} - D$ only radially we can take $D^* = \bar{D}$.

THEOREM 2. *For $x_0 \neq 0, n = 0, 1, 2, \cdots$ let $C_n(x_0) = \sum_{i=0}^{\infty} a_{ni}x_0^i$ and $C(x_0) = \sum_{i=0}^{\infty} c_i(x_0)$ be in $T$. Then if $C_n(x_0) \to C(x_0)$ the following are true.*

(i) *for each $i$ there is an $a_i$ such that $\lim_{n\to\infty} a_{ni} = a_i$;*

(ii) *$a_i x_0^i = c_i(x_0)$ .*

*That is, $C(x_0)$ is the term by term limit of $C_n(x_0)$ when it is the limit in the $T$ norm.*

*Proof.* Since $|C_n(x_0) - C(x_0)|_T \to 0$ we have $|C_n(x_0) - C_m(x_0)|_T \to 0$ as $n, m \to \infty$ independently. But

$$|C_n(x_0) - C_m(x_0)|_T = \left|\sum_{i=0}^{\infty} (a_{ni} - a_{mi})x_0^i\right| = \begin{cases} \max_i |a_{ni} - a_{mi}|\ |x_0|^i \\ \sum_{i=0}^{\infty} |a_{ni} - a_{mi}|\ |x_0|^i \end{cases}$$

in the cases where $k$ is non-Archimedean or Archimedean respectively. In either event $|a_{ni} - a_{mi}|\ |x_0|^i \leq |C_n(x_0) - C_m(x_0)|_T \to 0$. Hence by completeness of $k$ there is an $a_i$ such that $a_{ni} \to a_i$. This proves (i). To prove (ii) we have

$$|a_i x_0^i - c_i(x_0)| \leq |a_i x_0^i - a_{ni} x_0^i| + |a_{ni} x_0^i - c_i(x_0)|$$
$$\leq |a_i x_0^i - a_{ni} x_0^i| + |C_n(x_0) - C(x_0)|_T .$$

Since both terms on the right tend to zero the proof of (ii) is completed.

We now suppose that $k$ is algebraically closed and is non-Archimedean. If $C(x) = \sum_{i=0}^{\infty} a_i x^i$ is an entire power series (i.e. $C(x)$ is in $T$ for all $x$ in $k$) which is not identically zero then by the analogue of the Weierstrass decomposition theorem in algebraically closed non-Archimedean fields (see Schöbe [10] and Schnirelman [11]) we can express $C(x)$ as the formal limit of

$$a_0 x^{i_0} \prod_{q=0}^{n \leq \infty} (1 - x/z_q)$$

where $i_0$ is the multiplicity of the zero $x = 0$ of the sum $c(x)$ of $C(x)$ and where $z_q$ ranges over the set of non-zero zeros of $c(x)$, each factor $1 - x/z_q$ occuring a number of times equal to the multiplicity of $z_q$ as a zero of $c(x)$.

Schöbe [10] has also proved that $|z_q| \to \infty$ as $q \to \infty$. Therefore, since the terms of the product are power series and $1 + (1 - x/z_q) = x/z_q$ has $|x/z_q|_T \to 0$, the product $\prod_{q=0}^{n \leq \infty} (1 - x/z_q)$, when infinite, converges for every $x$ in $k$, relative to the topology of $T$. Hence by Theorem 2 above this product converges to $C(x)$. These remarks combined with Theorem 1 prove the following theorem. The notation is as above.

THEOREM 3. *Let $C(x)$ be an entire power series. Then for $x$ a non-zero of the sum function $c(x)$ of $C(x)$ we have*

$$A_{C(x)} = A_{a_0} A_x^{i_0} \prod_{q=0}^{n \leq \infty} A_{1-x/z_q}$$

*where, if $n$ is infinite, the right side converges to the left in both the uniform and weak topologies of $M$.*

In the case of the complex field the original Weierstrass decomposition theorem gives an analogous result where the $A_{1-x/z_q}$ are replaced by more complicated matrices corresponding to the primary factors of $C(x)$.

**5. Meromorphic functions and $C(x)$-continuation.** If the function $f(x)$ has a Taylor series expansion $\sum_{i=0}^{\infty} a_i (x - \alpha)^i$ about $\alpha$ in $k$ which converges to $f(x)$ in its circle of convergence we shall denote this series by $[f(x)]_\alpha$. If $D$ is the circle of convergence and $y$ is an interior point of $D$ we can expand $[f(x)]_\alpha$ about $y$ to obtain (formally) $[f(x)]_y$. Thus

$$[f(x)]_y \sim \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{i+j}{i}a_{i+j}(y-\alpha)^j(x-y)^i .$$

We shall denote $\sum_{j=0}^{\infty}\binom{i+j}{i}a_{i+j}(y-\alpha)^j$ by $[f_i(y)]_\alpha$. When the characteristic of $k$ is 0 we know

$$[f_i(y)]_\alpha = (1/i!)\left[\frac{d^i f(y)}{dy^i}\right]_\alpha .$$

(Although much of what we shall say is true for fields of arbitrary characteristic we confine ourselves to fields of characteristic 0 in order to simplify the discussion.)

Letting $f_i(y)$ be the sum function of $[f_i(y)]_\alpha$ we have

$$[f(x)]_y = \sum_{i=0}^{\infty} f_i(y)(x-y)^i .$$

It is known that in both the Archimedean and non-Archimedean case that for all $i$, $[f_i(y)]_\alpha$ converges for all $y$ in $D$. However in the Archimedean case it is often true that there is a circle $D_1$, not contained wholly within $D$, and in which, for all $i$, $[f_i(y)]_x$ converges and $\sum_{i=0}^{\infty}[f_i(y)]_x(x-y)^i$ converges in $D_1$ to $f(x)$.

This allows one to step by step recover the function $f(x)$ from a power series element $[f(x)]_x$ of the function. When $k$ is non-Archimedean it can be shown that no such circle as $D_1$ can ever exist. Thus the usual method of analytic continuation necessarily fails in such fields. In this section we shall show how $C(x)$-continuation can be applied in the case of the continuation of power series elements of meromorphic functions with known denominators (see below).

Let $D$ be a circle in $k$ ($D$ open if $k$ is Archimedean) with center $\alpha$. A function $f(x)$ defined over some subset of $k$ will be side to be *meromorphic over* $D$ if there exist two series $[g(x)]$ , $[h(x)]_x$ convergent on $D$ such that

(i) $f(x)$ is defined for $x$ in $D$ if and only if $h(x) \neq 0$;

(ii) $f(x) = g(x)/h(x)$ everywhere on $D$ where defined.

The function $h(x)$ will be called a *denominator* of $f(x)$ over $D$. If $D$ is the greatest such circle we call it the *circle of meromorphy* of $f(x)$.

LEMMA 6. *If $f(x)$ is meromorphic on $D$ with denominator $h(x)$ and if $\alpha$ is in $D$ then $f(x)$ is the $[h(x)]_\alpha$-continuation of $[f(x)]_x$.*

*Proof.* Let $x$ be in $D, h(x) \neq 0$. Then there is a $g(x)$ such that $[g(x)]_x$ converges on $D$ and $f(x) = g(x)/h(x)$. Now $f(x)$ is the sum function of

$$[g(x)]_\alpha/h(x) = ([h(x)]_\alpha/h(x))[f(x)]_\alpha = A_{[h(x)]_\alpha}[f(x)]_\alpha$$

and the lemma is proved.

THEOREM 4. *Let $f(x)$ be meromorphic on $D$ with denominator $h(x)$. Further for $\alpha$ in $D$ let*

$$[f(x)]_\alpha \sum_{i=0}^{\infty} a_i(x-\alpha)^i$$

*have circle of convergence contained in $D$. Then*

$$[f(x)]_y = \sum_{i=0}^{\infty} f_i(y)(x-y)^i$$

*when $y$ is in $D, h(y) \neq 0$, and $f_i(y)$ is the $[h(y)]_\alpha^{i+1}$-continuation of $i!\sum_{j=0}^{\infty}\binom{i+j}{i}a_{i+j}(y-\alpha)^j$.*

*Proof.* As seen above the formal expansion of $[f(x)]_\alpha$ about $y$ is given by

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\binom{i+j}{i}a_{i+j}(y-\alpha)^j(x-y)^i$$

where

$$i!\sum_{j=0}^{\infty}\binom{i+j}{i}a_{i+j}(y-\alpha)^j = \left[\frac{d^i f(y)}{dy^i}\right]_\alpha .$$

But

$$\frac{d^i f(y)}{dy^i} = \frac{d^i(g(y)/h(y))}{dy^i} = t(y)/h^{i+1}(y)$$

where $t(y)$ is a polynomial in $g(y)$ and $h(y)$. Thus $[t(y)]_\alpha$ converges over $D$ and $d^i f(y)/dy^i$ is meromorphic on $D$ with denominator $(t(y))^{i+1}$. Thus by Lemma 1, $d^i f(y)/dy^i$ is the $[(h(y))^{i+1}]_\alpha$-continuation of

$$[d^i f(y)/dy^i]_\alpha = i!\sum_{j=0}^{\infty}\binom{i+j}{i}a_{i+j}(y-\alpha)^j .$$

From Theorem 4 and Theorem 1 we have the

COROLLARY. *$f_i(y)$ is the sum function of $A^{i+1}_{[h(y)]_\alpha}[d^i f(y)/dy^i]_\alpha$.*

THEOREM 5. *If a function $f(x)$ defined over $D$ is the $C(x)$-continuation of a power series $[f(x)]_\alpha$ then $f(x)$ is meromorphic on $D$ when $C(x)$ is of the form $[h(x)]_\alpha$.*

*Proof. Let $f(x)$ be the sum function of*

$$A_{[h(x)]_\alpha}[f(x)]_\alpha = ([h(x)]_\alpha/h(x))[f(x)]_\alpha$$

which converges on $D$. Then $[g(x)]_\alpha = [h(x)]_\alpha[f(x)]_\alpha$ converges on $D$. Letting $g(x)$, $h(x)$ be the sums on $D$ of $[g(x)]_\alpha$, $[h(x)]_\alpha$ respectively gives $[f(x)]_\alpha$ the expansion about $\alpha$ of $g(x)/h(x)$ which is meromorphic on $D$.

There are many further questions which can be asked concerning these methods of continuation. In view of Theorem 5 one would wish to concentrate on $C(x)$-continuations where $C(x)$ is not a power series.

Further we can generalize the method so that instead of restricting ourselves to the use of $C(x)$-continuations we allow the use of arbitrary $T_2$ matrices. Some work has been done in this direction in [8a].

Vermes, making use of series to sequence methods, has dealt with similar problems for $k$ the field of complex numbers [13abc]. Some of his results in [13a] overlap some of the work done here. For further considerations of these and similar problems see the references to Chabauty, Krasner, Kürshák, Rychlik, Schöbe and Strassman cited below.

## References


1. C. Chabauty, *Sur la théorie des fonctions dans un corp valué* I, II, C.R. Acad. Sci., Paris, **231**, (1950), 396-7, 432-4.
2. R. G. Cooke, *Infinite Matrices and Sequence Spaces*, Macmillan, (1950).
3. J. Hadamard, *Essai sur l'étude des fonctions données par leur développement de Taylor*, J. Math. **4**, (1892), 8.
4. H. Hasse, *Zahlentheorie*, Akademie-Verlag, Berlin, (1949).
5. M. Krasner, (a) *Essai d'une théorie des fonctions analytiques dans les valués complets*, C. R. Acad. Sci., Paris **222**, (1946), 37-40, 165-7, 363-5, 581-3.

(b). *Certaines propriétés des séries de Taylor d'un ensemble au plus dénombrable de variables dans les corps valués complets et une demonstration structurale des formules de M. Pollaczek*, Bull. des Sci. Mathématique **71**, (1947), 123-152, 180-200.

(c) *Prolongement analytique dans les corps valués complets* C. R. Acad. Sci. Paris **238**, (1954), 2385-7 **239**, (1954), 468-470, 745-7, **244**, (1957), 1304-6, 1599-1602, 1996-9, 2510-2573, **245**, (1957), 1285-8.

6. J. Kürschák, *Limesbildung und allgemeine Körpertheorie*, J. reine angew. Math. **142**, (1913), 211-253.
7. F. Loonstra, *Analytische Untersuchungen über bewertete Körper* N.V. Noord-Hollandisch Uitgevers Maatschappij, Amsterdam (1941).
8. J. B. Roberts, (a) *Summability methods in valuation fields*, Thesis, Univ. of Minnesota (1955).

(b) *Matrix summability in F-fields*, Proc. Amer. Math. Soc. **8**, (1957) 541-3.

9. K. Rychlik, *Zur Bewertungstheorie der algebraischen Körper*, J. reine angew. Math. **153**, (1924) 94-107.
10. W. Schöbe, *Beiträge zur Functionentheorie in nicht-archimidisch bewerteten Körpern*, Universitas-Archiv, Band 42 des Archiv, Band 2 der Abteilung (1950).
11. L. Schnirelman, *Functions in algebraically closed valuation fields* (in Russian), Bull. Acad. Sci. URSS, Serié Math., **5**, **6**, (1938), 487-498.
12. R. Strassman, *Uber den Wertevorrat von Potenzreihen im Gebiet der p-adischen Zahlen*, J. reine angew. Math. **159**, (1928) 13-28.

13. P. Vermes, (a) *On γ-matrices and their application to the binomial series,* Proc. Ed. Math. Soc. **3, 8**, (1947), 1-13.

(b) *The application of γ-matrices to Taylor series,* Proc. Ed. Math. Soc. **2, 8,** (1948), 43-9

(c) *Series to series transformations and analytic continuation by matrix methods* Amer. J. Math. **71** (1949), 541-562.

REED COLLEGE

# IDEMPOTENT MEASURES ON ABELIAN GROUPS

WALTER RUDIN

## 1. Introduction.

1.1. All groups mentioned in this paper will be abelian, even when this is not explicitly stated, and will be written additively. For any locally compact abelian group $G$, $\mathscr{M}(G)$ will denote the set of all complex-valued Borel measures (sometimes called Radon measures) an $G$. The convolution of two such measures $\mu$ and $\lambda$ is the measure $\mu * \lambda$ defined by

$$(\mu * \lambda)(E) = \int_G \mu(E - x) d\lambda(x) \tag{1.1.1}$$

for every Borel set $E \subset G$, where $E - x$ is the set of all $y - x$ with $y \in E$.

With addition and scalar multiplication defined in the obvious way, and with convolution as multiplication, $\mathscr{M}(G)$ is a commutative algebra. A measure $\mu \in \mathscr{M}(G)$ is said to be *idempotent* if

$$\mu * \mu = \mu . \tag{1.1.2}$$

The set of all idempotent elements of $\mathscr{M}(G)$ will be denoted by $\mathscr{J}(G)$.

It would be interesting to have an explicit description of the idempotent measures on any locally compact group. For the circle group, this was obtained by Helson [1], and was of considerable help in the determination of the endomorphisms of the group algebra of that group [6]. In the present paper, the problem is completely solved for the finite-dimensional torus groups (Section IV) and for the discrete groups (Theorem 2.2). In Section II it is proved that every idempotent measure is concentrated on a compact subgroup (Theorem 2.1). In Section III the general problem is reduced to the study of the so-called irreducible idempotent measures on compact groups.

1.2. Apart from its intrinsic interest, this problem has another aspect: If $\mathscr{L}(G)$ is the Banach space of all complex Haar-integrable functions on $G$, and if $A$ is a bounded linear mapping of $\mathscr{L}(G)$ into $\mathscr{L}(G)$ which commutes with all translations of $G$, then it is known that there is a unique $\mu \in \mathscr{M}(G)$ such that

$$(Af)(x) = \int_G f(x - y) d\mu(y) \qquad (f \in \mathscr{L}(G)) . \tag{1.2.1}$$

Furthermore, $A^2 = A$ if and only if $\mu * \mu = \mu$.

Received May 22, 1958. The author is a Research Fellow of the Alfred P. Sloan Foundation.

Thus the determination of all idempotents in $\mathscr{M}(G)$ is equivalent to the determination of all projections in $\mathscr{L}(G)$ which commute with the translations of $G$.

1.3. Let $\Gamma$ be the dual group of $G$, that is, the group of all continuous characters of $G$. With every $\mu \in \mathscr{M}(G)$ there is associated its Fourier-Stieltjes transform $\hat{\mu}$, defined by

$$\hat{\mu}(\gamma) = \int_G (-x, \gamma) d\mu(x) \qquad (\gamma \in \Gamma), \tag{1.3.1}$$

where $(x, \gamma)$ denotes the value of the character $\gamma$ at the point $x$.

The correspondence $\mu \to \hat{\mu}$ is one-to-one, and the transform of a convolution is the pointwise product of the transforms of the factors. It follows that $\mu \in \mathscr{I}(G)$ if and only if $\hat{\mu}$ has 1 and 0 as its only values.

We associate with each $\mu \in \mathscr{I}(G)$ the set

$$S(\mu) = \{\gamma \in \Gamma \mid \hat{\mu}(\gamma) = 1\}. \tag{1.3.2}$$

*The problem of finding all* $\mu \in \mathscr{I}(G)$ *is then equivalent to the problem of finding all subsets of* $\Gamma$ *whose characteristic function is a Fourier-Stieltjes transform.*

1.4. In order to lead up to our conjecture concerning the structure of the measure in $\mathscr{I}(G)$, we present some relevant facts concerning subgroups, quotient groups, and measures.

(y) If $H$ is a closed subgroup of $G$, let $N(H)$ be the annihilator of $H$, that is, the set of all $\gamma \in \Gamma$ such that $(x, \gamma) = 1$ for all $x \in H$. Then $N(H)$ is a closed subgroup of $\Gamma$ and is the dual group of $G/H$. Also $\Gamma/N(H)$ is the dual group of $H$.

(b) For any $\mu \in \mathscr{M}(G)$, let $|\mu|$ be the measure defined by

$$|\mu|(E) = \sup \sum |\mu(E_i)|, \tag{1.4.1}$$

the sup being taken over all finite collections $\{E_i\}$ of pairwise disjoint Borel sets whose union is $E$; detailed proofs of the properties of the $|\mu|$ so defined can be found in [4]. The norm of $\mu$ is defined as $||\mu|| = |\mu|(G)$; with this norm, $\mathscr{M}(G)$ is a Banach algebra, and

$$||\mu|| \geq \sup |\hat{\mu}(\gamma)| \qquad (\gamma \in \Gamma, \mu \in \mathscr{M}(G)). \tag{1.4.2}$$

We say that $\mu$ is *concentrated* on $E$ if $|\mu|(E) = ||\mu||$. The *restriction* of $\mu$ to a set $A$ is the measure $\lambda$ defined by $\lambda(B) = \mu(A \cap B)$. The *support* of $\mu$ is the smallest closed set $F$ on which $\mu$ is concentrated. If $\mu$ is concentrated on a countable set, then $\mu$ is *discrete*; if $\mu(E) = 0$ for every countable set $E$, $\mu$ is *continuous*; if $\mu(E) = 0$ whenever $m(E) = 0$, where $m$ is the Haar measure of $G$, $\mu$ is *absolutely continuous*; finally, if $\mu$ is concentrated on a set $E$ with $m(E) = 0$, then $\mu$ is *singular*.

(c) If a closed subgroup $H$ contains the support of some $\mu \in \mathscr{M}(G)$, then $\mu$ may also be regarded as an element of $\mathscr{M}(H)$. Conversely, any $\lambda \in \mathscr{M}(H)$ may be regarded as an element of $\mathscr{M}(G)$, such that $\lambda(E) = 0$, whenever $E \cap H$ is empty.

It is important to note that if a group $H$ contains the support of two measures $\mu, \lambda \in \mathscr{M}(G)$, then the same is true of $\mu * \lambda$. This follows immediately from (1.1.1): Since $\lambda$ is concentrated on $H$, the integration extends over $H$ only, and if $E \cap H$ is empty, so is $(E - x) \cap H$ for every $x \in H$; hence $(\mu * \lambda)(E) = 0$.

(d) *If $H$ is a closed subgroup of $G$, if $\mu \in \mathscr{M}(G)$, and if $\hat{\mu}$ is constant on each coset of $N(H)$, then $\mu$ is concentrated on $H$.*

*We sketch the proof.* The assumption on $\mu$ implies that

$$\int_G (-x, \gamma')(-x, \gamma) d\mu(x) = \int_G (-x, \gamma') d\mu(x) \quad (\gamma' \in \Gamma, \gamma \in N(H)),$$

and the uniqueness theorem for Fourier-Stieltjes transforms shows that $(-x, \gamma) d\mu(x) = d\mu(x)$ for all $\gamma \in N(H)$. Hence $(-x, \gamma) = 1$ almost everywhere on the support of $\mu$, which means that this support lies in $H$.

(e) Suppose now that $G$ is *compact* (so that $\Gamma$ is *discrete*) and that $m_H$ is the Haar measure of a compact subgroup $H$ of $G$. Then $m_H \in \mathscr{M}(G)$ (see (c)), and it is easy to see that $m_H \in \mathscr{J}(G)$ and that $S(m_H) = N(H)$ (see (1.3.2)).

If, for some $\gamma \in \Gamma$, $d\mu(x) = (x, \gamma) dm_H(x)$, then $\mu$ is again idempotent, and $S(\mu) = N(H) + \gamma$.

*It follows that every coset of every subgroup is $S(\mu)$ for some $\mu \in \mathscr{J}(G)$.*

(f) Consider the family $W$ of all sets $S(\mu)$, for $\mu \in \mathscr{J}(G)$. If $\mu$ and $\lambda$ are idempotent, so are the measures $\mu * \lambda$, $\mu + \lambda - \mu * \lambda$, and $u - \mu$, where $u$ is the unit element of $\mathscr{M}(G)$ (i.e., $u$ is the point measure which assigns mass 1 to the identity elements of $G$; $\hat{u}(\gamma) = 1$ for all $\gamma \in \Gamma$). Since

$$S(\mu) \cap S(\lambda) = S(\mu * \lambda),$$
$$S(\mu) \cup S(\lambda) = S(\mu + \lambda - \mu * \lambda),$$

and the complement of $S(\mu)$ is $S(u - \mu)$, we see that $W$ is closed under finite intersections, finite unions, and complementation. That is to say, *$W$ is a ring of sets.*

1.5. Suppose again that $G$ is compact. Define the *coset-ring* of $\Gamma$ to be the smallest ring of sets which contains all cosets of all subgroups of $\Gamma$. We conclude from 1.4 (e), (f), that *every member of the coset-ring of $\Gamma$ is $S(\mu)$ for some $\mu \in \mathscr{J}(G)$.*

The structure of such a $\mu$ can be described as follows: Every subgroup of $\Gamma$ is $N(H)$ for some compact subgroup $H$ of $G$, and any finite union of cosets of $N(H)$ is $S(\mu)$ for a measure $\mu$ defined by

(1.5.1) $$d\mu(x) = \sum_{i=1}^{N} (x, \gamma_i) dm_H(x) ,$$

where $\gamma_1, \cdots, \gamma_N$ are distinct characters of $H$. Roughly speaking, $\mu$ is a trigonometric polynomial on $H$. If $S(\lambda)$ belongs to the coset ring of $\Gamma$, $\lambda$ can accordingly be obtained from measures of the form (1.5.1) by a finite sequence of the operations described in 1.4 (f).

It seems quite likely that there are no other idempotent measures:

CONJECTURE. *If $G$ is a compact abelian group and $\mu \in \mathscr{I}(G)$, then $S(\mu)$ belongs to the coset ring of $\Gamma$.*

The main result of this paper is the proof of this conjecture for the finite-dimensional torus groups.

## 2. Reduction to Compact Groups and Some Consequences[1].

Our first theorem shows that we may restrict our attention to measures defined on compact groups:

2.1. THEOREM. *Suppose $G$ is a locally compact abelian group and $\mu \in \mathscr{I}(G)$. Then $\mu$ is concentrated on a compact subgroup $K$ of $G$ (and hence $\mu \in \mathscr{I}(K)$).*

*Proof*. Let $G_0$ be the smallest closed subgroup of $G$ which contains the support of $\mu$; we wish to show that $G_0$ is compact. By 1.4(c) we may assume, without loss of generality, that $G_0 = G$, and 1.4(d) implies then that $\hat{\mu}$ is not constant on the cosets of any non-trivial closed subgroup of $\Gamma$. In other words, if we define $\mu_\gamma$ by

(2.1.1) $$d\mu_\gamma(x) = (x, \gamma) d\mu(x) \qquad (\gamma \in \Gamma) ,$$

then $\mu_\gamma \neq \mu$ if $\gamma \neq 0$. Since $\hat{\mu}_\gamma = 0$ or 1, (1.4.2) implies

(2.1.2) $$||\mu_\gamma - \mu|| \geq 1 \qquad (\gamma \neq 0) .$$

There exists a compact set $C \subset G$ such that $|\mu|(C') < 1/4$, where $C'$ denotes the complement of $C$. If $V$ is the set of all $\gamma \in \Gamma$ such that

(2.1.3) $$|1 - (x, \gamma)| < \frac{1}{3||\mu||}$$

for every $x \in C$, then $V$ is open (this is precisely the way in which the topology of $\Gamma$ is defined), and for every $\gamma \in V$ we have

(2.1.4) $$||\mu - \mu_\gamma|| \leq \int_G |1 - (x, \gamma)| d|\mu|(x) = \int_C + \int_{C'} \leq \frac{1}{3} + \frac{1}{2} < 1 .$$

[1] The proofs in this section are simpler than they were in the original version of this paper, due to welcome suggestions by the referee and by P. J. Cohen.

By (2.1.2), the open set $V$ thus consists of the identity element alone. Consequently, $\Gamma$ is discrete, and $G$ is compact.

2.2. THEOREM. *Suppose $G$ is discrete (so that $\Gamma$ is compact).*

(a) *If $\mu \in \mathscr{J}(G)$, then $\mu$ is concentrated on a finite subgroup $H$ of $G$, the annihilator $N(H)$ is an open-closed subgroup of $\Gamma$, and $S(\mu)$ is a finite union of cosets of $N(H)$.*

(b) *Conversely, every open-closed subset $E$ of $\Gamma$ is $S(\mu)$ for some $\mu \in \mathscr{J}(G)$.*

*Proof.* The first part of (a) follows from Theorem 2.1; for the rest, we observe that $\Gamma/N(H)$ is a finite group and that $\hat{\mu}$ is constant on the cosets of $N(H)$.

Next, if $E$ is an open-closed subset of $\Gamma$, there is a neighborhood $V$ of 0 in $\Gamma$ such that $E + V = E$. If $f$ and $g$ are the characteristic functions of $V$ and $E$, respectively, then

$$\int_\Gamma f(\gamma')g(\gamma - \gamma')dm(\gamma') = m(V)g(\gamma) , \tag{2.2.1}$$

where $m$ is the Haar measure of $\Gamma$. Since $f$ and $g$ are in $L^2(\Gamma)$, the Plancherel theorem implies that $g$ is a Fourier transform, and the result follows from the remark at the end of 1.3.

2.3. For technical reasons, which will become apparent in the next section, it is convenient to enlarge the class $\mathscr{J}(G)$ somewhat. We let $\mathscr{F}(G)$ be the class of all $\mu \in \mathscr{M}(G)$ such that $\hat{\mu}$ is an integer-valued function, and we can immediately prove the following proposition:

*If $\mu \in \mathscr{F}(G)$, then $\mu = a_1\mu_1 + \cdots + a_n\mu_n$, where $a_1, \cdots, a_n$ are integers and $\mu_1, \cdots, \mu_n \in \mathscr{J}(G)$.*

Indeed, let $a_1, \cdots, a_n$ be those integers which are different from 0 and which lie in the range of $\hat{\mu}$ (since $\hat{\mu}$ is bounded, this is a finite set). Let $P_i$ be a polynomial such that

$$P_i(0) = 0, P_i(a_j) = 0 \text{ if } j \neq i, P_i(a_i) = 1 ,$$

and put $\mu_i = P_i(\mu)$.

(We define $\mu^n = \mu * \mu^{n-1}$, and $P(\mu) = \sum_1^k c_n\mu^n$ if $P(x) = \sum_1^k c_n x^n$.) Then

$$\hat{\mu}_i(\gamma) = P_i(\hat{\mu})(\gamma) = \begin{cases} 1 \text{ if } \hat{\mu}(\gamma) = a_i, \\ 0 \text{ otherwise} \end{cases}$$

so that $\mu_i \in \mathscr{J}(G)$ and $\mu = a_1\mu_1 + \cdots + a_n\mu_n$.

2.4. THEOREM. *Suppose $G$ is a compact abelian group, $\mu \in \mathscr{F}(G)$,* *and $H$ is a closed subgroup of $G$. Let $\{H_i\}$ be the (evidently at most countable) collection of those cosets of $H$ for which*

$$|\mu|(H_i) > 0. \tag{2.4.1}$$

*Let $\tilde{H}$ be the smallest subgroup of $G$ which contains all these $H_i$, and let $\sigma$ be the restriction of $\mu$ to $\tilde{H}$. Then*

(i) $\sigma \in \mathscr{F}(G)$;

(ii) $\tilde{H}/H$ *is a finite group.*

Assertion (ii) implies, in particular, that $\{H_i\}$ is a finite collection, and that $\mu$ vanishes on every coset of $H$ which has infinite order in $G/H$.

*Proof.* We first claim that the following two statements are true for every Borel set $A \subset G$:

(a) If $A \cap \tilde{H}$ is empty, then $\sigma^n(A) = 0$ for $n = 1, 2, 3, \cdots$.

(b) If $A \subset \tilde{H}$, then $\sigma^n(A) = \mu^n(A)$ for $n = 1, 2, 3, \cdots$.

Note that $\tilde{H}$ is an at most countable union of cosets of $H$, hence in particular is a Borel set.

It is clear that (a) holds if $n = 1$, and we proceed by induction:

$$\sigma^n(A) = \int_G \sigma^{n-1}(A - x)d\sigma(x) = \int_{\tilde{H}} \sigma^{n-1}(A - x)d\sigma(x) .$$

If $A \cap \tilde{H} = 0$ and $x \in \tilde{H}$, then $(A - x) \cap \tilde{H} = 0$. Thus if (a) holds for $n - 1$, it also holds for $n$.

To prove (b), put $\tau = \mu - \sigma$, and expand $\mu^n = (\sigma + \tau)^n$ by the binomial theorem. We have to show that

$$(\tau^k * \sigma^{n-k})(A) = 0 \qquad (k = 1, 2, \cdots, n) \tag{2.4.2}$$

if $A \subset \tilde{H}$.

Since $\tau$ vanishes on every coset of $H$ and since $\tilde{H}$ is an at most countable union of such cosets, we have $\tau(A - x) = 0$ for every $x \in G$. Thus, for any $\lambda \in \mathscr{M}(G)$,

$$(\tau * \lambda)(A) = \int_G \tau(A - x)d\lambda(x) = 0 ,$$

and (2.4.2) follows.

From (a) and (b) we conclude that

$$\sigma^n(E) = \mu^n(E \cap \tilde{H}) \qquad (n = 1, 2, 3, \cdots) \tag{2.4.3}$$

for every Borel set $E \subset G$.

Let $a_1, \cdots, a_n$ be the non-zero values of $\hat{\mu}$, and let $P$ be the polynomial

$$P(t) = t \prod_{i=1}^{n} (t - a_i) .$$

Then $P(\mu) = 0$, and (2.4.3) implies that $P(\sigma) = 0$. From this it follows that every value of $\hat{\sigma}$ is a root of $P$, so that the range of $\hat{\sigma}$ lies in the set $\{0, a_1, \cdots, a_n\}$. This proves the first part of the theorem.

We now put a new topology on $\tilde{H}$: the neighborhoods of 0 are to be the sets $H \cap V$, for any neighborhood $V$ of 0 in $G$. Then $\tilde{H}$ is a locally compact abelian group, and $\sigma \in \mathscr{F}(\tilde{H})$. By Theorem 2.1 (extended from $\mathscr{I}(G)$ to $\mathscr{F}(G)$, via Proposition 2.3), we see that $\sigma$ is concentrated on a compact subgroup $K$ of $\tilde{H}$. The minimality which was one of the defining properties of $\tilde{H}$ shows that $K = \tilde{H}$. Thus $\tilde{H}$ is compact, and since $H$ is an open subgroup of $\tilde{H}$, we conclude that $\tilde{H}/H$ is finite.

## 3. Decomposition into Irreducible Measures.

3.1. If $G$ is compact, $\mu \in \mathscr{I}(G)$, and $H$ is a compact subgroup of $G$, the second part of Theorem 2.4 shows that $H$ has finite index in a compact subgroup $\tilde{H}$ such that $\mu$ vanishes on every translate of $\tilde{H}$ which is different from $H$. The existence of such $\tilde{H}$ suggests the following definitions.

($\alpha$) Suppose $G$ is compact, $\mu \in \mathscr{M}(G)$, and $K$ is a compact subgroup of $G$. We say that $K$ *is associated with* $\mu$ if

(i) $|\mu|(K + x) = 0$ for every $x \notin K$;

(ii) $|\mu|(H) < |\mu|(K)$ for every compact subgroup $H$ of $K$ which is different from $K$.

Note that (ii) implies $|\mu|(K) > 0$. Thus the null-measure has no group associated with it. If $\mu \neq 0$, the smallest compact group on which $\mu$ is concentrated is clearly associated with $\mu$.

($\beta$) If $\mu = 0$, or if there is precisely one compact subgroup associated with $\mu$, we say that $\mu$ is *irreducible.*

It should be pointed out that $\mu$ need not be concentrated on a subgroup which is associated with $\mu$. For example, let $G$ be the circle group $T^1$ (the one-dimensional torus), and set $\mu = m + \lambda$, where $m$ is the Haar measure of $T^1$, and $\lambda$ is a positive measure concentrated at the point $e^{ix}$. If $x/\pi$ is rational, then the finite cyclic group generated by $e^{ix}$ is associated with $\mu$, and so is $T^1$. If $x/\pi$ is irrational, then $T^1$ is the only group associated with $\mu$, and $\mu$ is irreducible.

3.2. The following two propositions may elucidate these concepts (we use the notations of 3.1):

(a) *If $K$ and $H$ are associated with $\mu$ and if $H$ is a proper subgroup of $K$, then $K/H$ is infinite.*

Indeed, suppose $K/H$ is finite, so that a finite number of translates of $H$ covers $K$. Since $|\mu|(H+x)=0$ whenever $x \notin H$, it follows that $|\mu|(K)=|\mu|(H)$, a contradiction.

(b) *Suppose $\mu \in \mathscr{F}(G)$ and $K$ is the smallest compact subgroup of $G$ such that $|\mu|(K)=\|\mu\|$. Then $\mu$ is irreducible if and only if $|\mu|(H)=0$ for every compact subgroup $H$ of $K$ such that $m(H)=0$, where $m$ is the Haar measure of $K$. If $\mu$ is irreducible, we also have $|\mu|(H+x)=0$ for every such $H$ and all $x \in G$.*

Note that $m(H)=0$ if and only if $K/H$ is infinite. Thus if $|\mu|(H)=0$ whenever $m(H)=0$, proposition (a) implies that $\mu$ is irreducible. On the other hand, if $|\mu|(H)>0$ for some $H$ with $m(H)=0$, the $\tilde{H}$ of Theorem 2.4 (with $K$ in place of $G$) is associated with $\mu$ and $m(\tilde{H})=0$; hence $\mu$ is not irreducible. The last assertion of (b) also follows from Theorem 2.4.

3.3 THEOREM. *Suppose $G$ is a compact abelian group and $\mu \in \mathscr{F}(G)$. Then there exist irreducible measures $\mu_i \in \mathscr{F}(G)$ and integers $a_i$ such that*

$$\mu = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n .$$

*Proof.* It is convenient to weaken the hypothesis somewhat and to assume merely that $\mu \in \mathscr{F}(G)$; assume also $\mu \neq 0$.

We first show that *there is a compact subgroup $H_1$ of $G$ which is associated with $\mu$, such that the restriction $\lambda_1$ of $\mu$ to $H_1$ is irreducible, and such that $\lambda_1 \in \mathscr{F}(G)$.*

If $\mu$ is irreducible, let $H_1$ be the smallest compact group on which $\mu$ is concentrated. If $\mu$ is not irreducible, there is a compact group $K_1$ which is associated with $\mu$, such that $|\mu|(K_1)<\|\mu\|$; let $\sigma_1$ be the restriction of $\mu$ to $K_1$. By Theorem 2.4, $\sigma_1 \in \mathscr{F}(G)$, and the same is thus true of $\mu-\sigma_1$. Since $\mu-\sigma_1 \neq 0$, (1.4.2) shows that

$$\|\mu-\sigma_1\| \geq 1 . \tag{3.3.1}$$

On the other hand, $\sigma_1$ and $\mu-\sigma_1$ are concentrated on disjoint sets, so that

$$\|\mu\| = \|\sigma_1\| + \|\mu-\sigma_1\| . \tag{3.3.2}$$

It follows that

$$\|\sigma_1\| \leq \|\mu\| - 1 . \tag{3.3.3}$$

If $\sigma_1$ is not irreducible, we repeat this process, with $\sigma_1$ in place of $\mu$: there is a compact $K_2 \subset K_1$ which is associated with $\sigma_1$ (hence with $\mu$), such that the restriction $\sigma_2$ of $\sigma_1$ to $K_2$ belongs to $\mathscr{F}(G)$ and satisfies the inequality

$$\|\sigma_2\| \leq \|\sigma_1\| - 1 \leq \|\mu\| - 2. \tag{3.3.4}$$

Since the norms decrease by at least 1 each time, we evidently obtain an irreducible measure ($\neq 0$) after repeating this process a finite number of times. Call this measure $\lambda_1$.

Since $\lambda_1$ is the restriction of $\mu$ to a group $H_1$, we see that $\lambda_1$ and $\mu - \lambda_1$ are concentrated on disjoint sets, so that $\|\mu\| = \|\lambda_1\| + \|\mu - \lambda_1\|$. We conclude, as above, that

$$\|\mu - \lambda_1\| \leqq \|\mu\| - 1 . \tag{3.3.5}$$

If $\mu - \lambda_1$ is irreducible, put $\lambda_2 = \mu - \lambda_1$. If not, repeat the preceding construction, finding an irreducible $\lambda_2 \in \mathscr{F}(G)$ such that

$$\|\mu - \lambda_1 - \lambda_2\| \leq \|\mu\| - 2 . \tag{3.3.6}$$

Again, this must stop after a finite number of steps, and we obtain a representation

$$\mu = \lambda_1 + \lambda_2 + \cdots + \lambda_p , \tag{3.3.7}$$

where each $\lambda_i$ is irreducible and belongs to $\mathscr{F}(G)$.

In 2.3, we saw that every $\lambda \in \mathscr{F}(G)$ is a finite linear combination, with integer coefficients, of idempotent measures $\mu_j$. The theorem thus follows from (3.3.7) if we can show that the $\mu_j$ are irreducible if $\lambda$ is irreducible; by 2.3, we therefore have to show that $P(\lambda)$ is irreducible if $P$ is a polynomial without constant term.

Suppose then that $K$ is associated with $\lambda$ and that $\lambda$ is irreducible. Let $m$ denote the Haar measure of $K$. If $H$ is a compact subgroup of $K$ with $m(H) = 0$, then $|\lambda|(H) = 0$ by 3.2 (b), and it easily follows that $|\lambda^n|(H) = 0$ for $n = 1, 2, 3, \cdots$, and hence that $|P(\lambda)|(H) = 0$. Thus if $P(\lambda) \neq 0$, $P(\lambda)$ is concentrated on a subgroup $K'$ of $K$ which has finite index in $K$, and $K'$ is associated with $P(\lambda)$. Applying 3.2 (b) again, we conclude that $P(\lambda)$ is irreducible.

This completes the proof.

3.4 Let $R$ denote the coset ring of $\Gamma$, and call a function $f$ defined on $\Gamma$ an $R$-function if

$$f(\gamma) = \sum_{i=1}^{N} c_i g_i(\gamma) ,$$

where each $g_i$ is the characteristic function of some member of $R$, and each $c_i$ is a complex number.

Suppose we know, for some compact group $G$, that $S(\mu) \in R$ for every irreducible $\mu \in \mathscr{F}(G)$. Then $\hat{\mu}$ is an $R$-function, and since a finite sum of $R$-functions is again an $R$-function, Theorem 3.3 shows that $\hat{\sigma}$ is an $R$-function for every $\sigma \in \mathscr{F}(G)$ (irreducible or not); this means that $S(\sigma) \in R$.

*Thus the conjecture made in* 1.5 *will be proved for a given compact* $G$ *if it is proved for every irreducible* $\mu \in \mathscr{J}(G)$.

According to the remarks made in 1.5, we therefore have to prove the following: *if* $K$ *is the subgroup associated with the irreducible* $\mu \in \mathscr{J}(G)$, *then* $\mu$ *is a trigonometric polynomial on* $K$. This is equivalent to the assertion that $\mu$ is absolutely continuous with respect to the Haar measure of $K$; for if $\mu$ is not a trigonometric polynomial, then $\hat{\mu}(\phi) = 1$ for infinitely many $\phi$ in the dual group of $K$, so that $\hat{\mu}(\phi)$ does not vanish at infinity.

## 4. The Idempotent Measure of the Torus Groups.

4.1. Let $T^r$ denote the $r$-dimensional torus group; the points of $T^r$ are of the form

$$x = (\xi_1, \cdots, \xi_r) , \tag{4.1.1}$$

the $\xi_i$ being real numbers mod $2\pi$. The dual group of $T^r$ is $\Lambda^r$, the group of all lattice points in $r$-dimensional euclidean space $R^r$, i.e., the set of all

$$n = (\nu_1, \cdots, \nu_r) \tag{4.1.2}$$

where the $\nu_i$ are integers. If we put

$$n \cdot x = \sum_{i=1}^{r} \nu_i \xi_i , \tag{4.1.3}$$

then $(x, n) = e^{in \cdot x}$, and the transform of any $\mu \in \mathscr{M}(T^r)$ is

$$\hat{\mu}(n) = \int_{T^r} e^{-in \cdot x} d\mu(x) \qquad (n \in \Lambda^r) . \tag{4.1.4}$$

We shall prove that every $\mu \in \mathscr{J}(T^r)$ has the structure described in 1.5:

4.2. THEOREM. *If* $\mu \in \mathscr{J}(T^r)$, *then* $S(\mu)$ *belongs to the coset ring of* $\Lambda^r$.

The discussion in 3.4 shows that Theorem 4.2 is a consequence of the following:

4.3 THEOREM. *Suppose* $G$ *is a compact subgroup of* $T^r$, *and suppose* $G$ *is associated with an irreducible measure* $\mu \in \mathscr{J}(T^r)$. *Then* $\mu$ *is absolutely continuous with respect to the Haar measure of* $G$.

*Proof.* We shall use induction on $r$.

For $r = 1$, the theorem is due to Helson [1] (our terminology differs from his, being adapted to a more general situation), but we include the proof for the sake of completeness.

If $G$ is a finite group, there is nothing to prove. If $G = T^1$, we have to prove that $S(\mu)$ is finite. Since $\mu$ is continuous, a well-known theorem of Wiener implies that

$$\lim_{N\to\infty} \frac{1}{2N+1} \sum_{-N}^{N} |\hat{\mu}(n)|^2 = 0 . \tag{4.3.1}$$

If $S(\mu)$ were infinite, it would therefore contain an infinite set $\{n_k\}$ such that none of the integers $n_k + 1, \cdots, n_k + k$ belong to $S(\mu)$, and a subsequence of the measures $\mu_k$ defined by

$$d\mu_k(x) = e^{-in_k x} d\mu(x) \tag{4.3.2}$$

would converge weakly (as functionals on the space of all continuous functions on $T^1$) to a measure $\sigma \in \mathscr{S}(T^1)$. The choice of $\{n_k\}$ shows that $\hat{\sigma}(0) = 1$ and $\hat{\sigma}(n) = 0$ for all $n > 0$. This latter fact implies, by a well-known theorem of F. and M. Riesz, that $\sigma$ is absolutely continuous. But every weak limit of the sequence (4.3.2) must be singular [2; p. 236]. Since $\sigma \neq 0$, this is a contradiction, and we conclude that $S(\mu)$ is finite.

We now assume that the theorem has been proved for $r \leq p - 1$ ($p = 2, 3, 4, \cdots$). To prove it for $r = p$, we consider two possibilities:

*Case* 1. *$G$ is the direct sum of $T^q$ (for some $q < p$) and a finite group $F$.*

*Case* 2. $G = T^p$.

*Case* 1. Let $f_1, \cdots, f_s$ be the elements of $F$, so that each element of $G$ can be written uniquely in the form $x + f$ with $x \in T^q$ and $f \in F$. Let $\phi_1, \cdots, \phi_s$ be the characters of $F$.

Let $\mu_1, \cdots, \mu_s$ be measures on $T^q$ defined by

$$\mu_k(E) = \sum_{j=1}^{s} (-f_j, \phi_k)\mu(E_j) \qquad (k = 1, \cdots, s) , \tag{4.3.3}$$

where $E$ is a Borel set in $T^q$ and $E_j = E + f_j$. Then, for $n \in \Lambda^q$, we have

$$\begin{aligned} \hat{\mu}_k(n) &= \int_{T^q} e^{-in\cdot x} d\mu_k(x) \\ &= \sum_{j=1}^{s} \int_{T^q + f_j} e^{-in\cdot x}(-f_j, \phi_k) d\mu(x + f_j) \\ &= \int_G e^{-in\cdot x}(-f, \phi_k) d\mu(x + f) = \hat{\mu}(n + \phi_k) = 0 \text{ or } 1 , \end{aligned}$$

so that $\mu_k \in \mathscr{F}(T^q)$.

Since $\mu$ is irreducible and $G$ is associated with $\mu$, (4.3.3) shows that $|\mu_k|(H) = 0$ for every proper compact subgroup of $T^q$; hence $\mu_k$ is irreducible and $T^q$ is associated with $\mu_k$, or $\mu_k = 0$. Our induction hypothesis now implies that $\mu_k$ is absolutely continuous $(1 \leq k \leq s)$.

If we multiply the equations (4.3.3) by $(f_i, \phi_k)$, add over $k$, and observe the orthogonality relations satisfied by the characters $\phi_k$, we obtain

$$\sum_{k=1}^{s} (f_i, \phi_k)\mu_k(E) = s \cdot \mu(E_i) \qquad (i = 1, \cdots, s) , \tag{4.3.4}$$

and thus the absolute continuity of the measures $\mu_k$ implies that $\mu$ is absolutely continuous with respect to the Harr measure of $G$. This settles Case 1.

*Case* 2. We now assume that $\mu \in \mathscr{F}(T^p)$ and that $|\mu|(H) = 0$ on every proper compact subgroup of $T^p$ (compare 3.2 (b)), and we wish to prove that $S(\mu)$ is finite. Our proof will be similar to that of the case $r = 1$, but we have to replace our reference to the theorem of F. and M. Riesz by a result recently proved by Helson and Lowdenslager[1] [3; Section 4, Lemma 3]:

*Let $Q$ be a subset of $\Lambda^p$ such that* (a) $n_1 + n_2 \in Q$ if $n_1 \in Q$ *and* $n_2 \in Q$, (b) $0 \notin Q$, (c) *if $n \neq 0$, then $n \in Q$ if and only if $-n \notin Q$. Suppose $\sigma$ is a singular measure on $T^p$ such that $\hat{\sigma}(n) = 0$ for all $n \in Q$. Then* $\hat{\sigma}(0) = 0$.

We think of $\Lambda^p$ as a subset of euclidean space $R^p$ (see 4.1). The theorem of Wiener (namely, (4.3.1)) extends without difficulty to Fourier-Stieltjes series in several variables and shows that $S(\mu)$ has density 0 in $\Lambda^p$, since $\mu$ is continuous. More precisely, the number of points of $S(\mu)$ in the $p$-dimensional cube with center at the origin and vertices $(\pm N, \pm N, \cdots \pm N)$ is $o(N^p)$.

If $S(\mu)$ is infinite, it follows that there exist spheres $V_k$ in $R^p$ with the following properties:

(i) The radius of $V_k$ is greater than $k$.

(ii) $V_k$ contains no point of $S(\mu)$ in its interior.

(iii) The boundary of $V_k$ contains a point $n_k \in S(\mu)$, and $n_k \to \infty$ as $k \to \infty$.

(iv) If $c_k$ is the center of $V_k$, the unit vectors

$$\frac{c_k - n_k}{|c_k - n_k|}$$

converge to some $b \in R^p$. (The absolute value sign denotes the length of the vector.)

---

[1] I wish to thank these two authors for letting me read their paper prior to its publication.

(v) The measures $\mu_k \in \mathscr{I}(T^p)$ defined by

$$d\mu_k(x) = e^{-in_k \cdot x} d\mu(x) \tag{4.3.5}$$

converge weakly to a measure $\sigma \in \mathscr{I}(T^p)$.

It is evident that $\{V_k\}$ can be chosen so as to statisfy (i), (ii), (iii), and a suitable subsequence will satisfy (iv) and (v) as well.

The passage from $\mu$ to $\sigma$, via (4.3.5), is such that $|\sigma|(E) \leq |\mu|(E)$ for every Borel set $E \subset T^p$; also, $\sigma$ is not changed if we replace $\mu$ by its singular component, since the Fourier transform of the absolutely continuous component tends to 0 (this is the argument used by Helson in [2; p. 236]).

Hence $\sigma$ is singular, $\sigma \in \mathscr{I}(T^p)$, and since $|\sigma|(H) \leq |\mu|(H) = 0$ for every proper compact subgroup $H$ of $T^p$, $\sigma$ is irreducible; our choice of $\{V_k\}$ shows that $\hat{\sigma}(0) = 1$ and that $\hat{\sigma}(n) = 0$ for every $n$ in the open half-space determined by $n \cdot b > 0$.

This last property is a consequence of the relation

$$S(\mu_k) = S(\mu) - n_k \tag{4.3.6}$$

and the fact that $S(\mu_k)$ therefore has no point in the interior of the sphere through 0 whose center is at $c_k - n_k$.

Let $\Lambda^q$ be the subgroup of $\Lambda^p$ which lies in the hyperplane $n \cdot b = 0$. Then $0 \leq q < p$. If $q = 0$, the theorem of Helson and Lowdenslager gives an immediate contradiction: take for $Q$ the set of all $n$ such that $n \cdot b > 0$; then $\hat{\sigma}(n) = 0$ in $Q$, $\sigma$ is singular, but $\hat{\sigma}(0) = 1$.

If $0 < q < p$, then $T^p$ is a direct sum $T^q + T^{p-q}$, where $\Lambda^q$ is the dual group of $T^q$ and the annihilator of $T^{p-q}$. Let $h$ be the natural homomorphism of $T^p$ onto $T^q$, and define a measure $\lambda \in \mathscr{M}(T^q)$ by

$$\lambda(E) = \sigma(h^{-1}(E)) \tag{4.3.7}$$

for all Borel sets $E \subset T^q$. Any $x \in T^p$ has a unique representation $x = x_1 + x_2$ with $x_1 \in T^q$, $x_2 \in T^{p-q}$. If $n \in \Lambda^q$, then $e^{in \cdot x_2} = 1$, so that

$$\hat{\lambda}(n) = \int_{T^q} e^{-in \cdot x_2} d\lambda(x_1) = \int_{T^p} e^{-in \cdot x} d\sigma(x) = \hat{\sigma}(n) = 0 \text{ or } 1\ . \tag{4.3.8}$$

Thus $\lambda \in \mathscr{I}(T^q)$ and the irreducibility of $\sigma$ shows that $\lambda$ vanishes on every proper compact subgroup of $T^q$. Our induction hypothesis now implies that $S(\lambda)$ is a finite subset of $\Lambda^q$.

Since $S(\lambda) = S(\sigma) \cap \Lambda^q$, we see that $S(\sigma)$ has only a finite number of points in the hyperplane $n \cdot b = 0$, and a suitable translation of $S(\sigma)$ by a vector in this hyperplane results in a singular measure $\sigma_1 \in \mathscr{I}(T^p)$ which has $\hat{\sigma}_1(n) = 0$ in a set $Q$ which satisfies the hypotheses of the Helson-Lowdenslager theorem, but which has $\hat{\sigma}_1(0) = 1$.

This contradiction shows that $S(\mu)$ is finite, and the proof is complete.

**5. Remarks.**

5.1 In the preceding section we have determined all idempotent measures on $T^r$, and incidentally also on all groups of the form $T^r + F$, where $F$ is any finite abelian group. We note that *these groups are the only compact abelian groups which have no infinite totally disconnected subgroups.*

Indeed, if $G$ is not $T^r + F$, then its dual group $\Gamma$ is not finitely generated (since every finitely generated abelian group is a direct sum of cyclic groups), and the well-known fact that a compact group is totally disconnected if and only if its dual group is a torsion group shows that the above-mentioned proposition is equivalent to the following purely algebraic theorem (compare 1.4 (a)).

5.2 THEOREM. *If $G$ is an abelian group which is not finitely generated, then $G$ can be mapped homomorphically onto an infinite torsion group.*

*Proof.* If $G$ has finite rank $p$, let $\{x_1, \cdots, x_p\}$ be an independent set in $G$. Factoring out the group generated by $x_1, \cdots, x_p$ gives a torsion quotient group, and the latter is infinite, since $G$ would otherwise be finitely generated.

If $G$ has infinite rank, let $\{x_1, x_2, x_3, \cdots\}$ be an independent set in $G$, and let $H$ be the group generated by $\{x_n\}$. Every $x \in H$ has a unique representation

$$x = \sum_1^\infty a_n(x)x_n , \tag{5.2.1}$$

where the coefficients $a_n(x)$ are integers; for each $x$, only finitely many $a_n(x)$ are different from 0.

Let $\{t_n\}$ be a sequence of distinct rational numbers, $0 < t_n < 1$, and define

$$h(x) = \sum_1^\infty a_n(x)t_n \qquad \text{(mod 1)} . \tag{5.2.2}$$

It is clear that $h$ is a homomorphism of $H$ into the group $Y$ of the rationals modulo the integers, and since $h(x_n) = t_n$, $h(H)$ is infinite. Since $Y$ is divisible, $h$ can be extended to a homomorphism of $G$ into $Y$ [5; p. 11]. Since $Y$ is a torsion group, the homomorphism $h$ has the desired properties.

We conclude with a result which gives further support to the conjecture stated in 1.5.

5.3 THEOREM. *Suppose $G$ is compact and connected, and $G$ is associated with an irreducible $\mu \in \mathscr{J}(G)$. Then $S(\mu)$ is a finite set.*

We shall omit the details of the proof, and merely give an outline: If $S(\mu)$ is infinite, then $S(\mu) \cap \Gamma_0$ is infinite, for some countable subgroup $\Gamma_0$ of $\Gamma$. Since $G$ is connected, $\Gamma$ has no element of finite order, and $\Gamma_0$ is isomorphic to a subgroup of the real line. If $\Delta$ is any finitely generated subgroup of $\Gamma_0$, then $S(\mu) \cap \Delta$ is finite, by the results of Section IV, and $\Gamma_0$ is the union of a countable increasing sequence of such $\Delta$'s. A translation argument, combined with a generalized version of the Helson-Lowdenslager Theorem (Section 6 of [3]) now leads to a contradiction, as in Section IV.

POSTSCRIPT (added in proof). About six months after the completion of this paper, P. J. Cohen has succeeded in proving the conjecture made in paragraph 1.5 in its full generality.

## REFERENCES

1. Henry Helson, *Note on harmonic functions*, Proc. Amer. Math. Soc. **4** (1953), 686-691.
2. ———, *On a theorem of Szegö*, Proc. Amer Math. Soc. **6** (1955), 235-242.
3. ———, and David Lowdenslager, *Prediction theory and Fourier series in several variables*, Acta Math. , **99** (1958), 165-202.
4. Edwin Hewitt, *Remarks on the inversion of Fourier-Stieltjes transforms*, Annals of Math. **57** (1953), 458-474.
5. Irving Kaplansky, *Infinite Abelian groups*, Ann Arbor, 1954.
6. Walter Rudin, *The automorphisms and the endomorphisms of the group algebra of the unit circle*, Acta Math. **95** (1956), 39-55.

UNIVERSITY OF ROCHESTER

# FREDHOLM EIGEN VALUES OF MULTIPLY-CONNECTED DOMAINS

M. SCHIFFER

**Introduction.** The solution of the boundary value problems of potential theory can be reduced, according to Poincaré, to an inhomogeneous integral equation of the second kind. It was the study of this particular problem which led, at the beginning of this century, to the development of the modern integral equation theory at the hands of Fredholm and Hilbert. From the beginning, attention was drawn to the eigen value problem for the homogeneous integral equation with the potential theoretical kernel [10]. The eigen functions of this problem can be extended as harmonic functions into the domain considered as well as extended into the complementary domain and give rise to interesting series developments and to a theory relating solutions of the interior and exterior boundary value problems of a closed curve or surface.

In a preceding paper [17], these Fredholm eigen functions were applied to problems of conformal mapping of simply-connected plane domains. Their connection with the dielectric Green's function of such domains was discussed and we showed the possibility of obtaining univalent functions by means of the dielectric Green's function. A variational formula for the Fredholm eigen values was established and an extremum problem for the latter was solved which permitted one to estimate the convergence of the Neumann-Liouville series solving the Dirichlet and Neumann boundary value problems.

In the present paper, the Fredholm eigen value problem is studied in the case of multiply-connected plane domains. Various new difficulties arise in this case. The complementary region of a multiply-connected domain is a domain set and the number of trivial solutions of the problem with the eigen values $|\lambda| = 1$ increases. This fact necessitates a brief restatement of the basic definitions and concepts in § 1. A certain repetition and overlap of material with the preceding paper could not be avoided; but, on the other hand, the presentation of this section makes the paper self-contained and should facilitate the understanding of it.

In § 2, the dielectric Green's functions $g_\varepsilon(z, \zeta)$ of a multiply-connected domain are discussed and their Fourier development in terms of the Fredholm eigen functions is given. The functions $g_\varepsilon$ are of geometric-physical significance by themselves; moreover, they represent a one-parameter $(0 < \varepsilon < \infty)$ family of harmonic positive-definite kernels which

Received October 1, 1958. Prepared under contract Nonr-225 (11) for Office of Naval Research.

have also the Fredholm functions as eigen functions. For $\varepsilon = 1$, $g_\varepsilon(z, \zeta)$ reduces to the fundamental singularity $-\log|z-\zeta|$ and leads to the classical kernel of potential theory. A power series development of the dielectric Green's function in terms of $(\varepsilon - 1)/(\varepsilon + 1)$ is given; the coefficient kernels are elementary and can be calculated explicitly by integration of simple functions over the boundary curve system.

The role of the one-parameter family $g_\varepsilon(z, \zeta)$ becomes particularly interesting when one studies the limit cases $\varepsilon = 0$ and $\varepsilon = \infty$. This is done in § 3. It appears that this function family interpolates between two well-known harmonic functions which determine two important canonical mappings of the domain considered; namely the radial-slit mapping and the circular-slit mapping.

In § 4 it is proved that not only the limit cases $\varepsilon = 0$ and $\varepsilon = \infty$ of $g_\varepsilon(z, \zeta)$ give rise to univalent functions in the domain but that each dielectric Green's function does so. We obtain one-parameter families of univalent functions which connect the radial-slit mapping function continuously with the circular-slit mapping function via any prescribed univalent function in the domain. This result is applied to give a new proof for the extremum properties which characterize the above two canonical slit mappings. Another type of one-parameter sets of univalent functions is constructed which interpolates between the canonical parallel-slit mappings.

In § 5, we use the dielectric Green's functions in order to define various norms and scalar products. These are quadratic and bilinear functionals defined for harmonic functions in the multiply-connected domain $\tilde{D}$ as well as for functions harmonic in the complementary domain set $D$. If one pair of argument functions is defined in $\tilde{D}$, the other pair in $D$, and if relations between their boundary values on the separating curve system are assumed, equations between the various scalar products are obtained. It is shown that these identities yield estimates and Ritz procedures for solution of boundary value problems in $\tilde{D}$ if the corresponding boundary value problems for the complementary set $D$ are already solved. In the special case $\varepsilon = 1$ the procedure becomes, of course, particularly easy to apply since the dielectric Green's function becomes trivial. It has, indeed, already been used in this form in order to prove interesting isoperimetric inequalities for polarization and for virtual mass [18–20]. The extension of the method to the case of general $\varepsilon$ should increase its flexibility and clarify its significance. The various quadratic forms are used, finally, in order to characterize each Fredholm eigen value $|\lambda| > 1$ by the solution of a simple maximum problem without side conditions. This result lays the groundwork for proving the variational formula for the Fredholm eigen values in the next section. The extremum definition is also used in order to prove that all positive

Fredholm eigen values of a subsystem of curves are never less than the corresponding positive eigen values of the full curve system.

In § 6, we derive the variational formula for the dielectric Green's functions under a small deformation of the domain. Through the maximum definition of the Fredholm eigen values, we can derive from this result also the variational formula for the Fredholm eigen values under the same deformation, This formula could also have been obtained immediately from the general perturbation theory of operators. But it seems of methodological interest to utilize fully the maximum property of each eigen value in order to give an elementary proof for this formula.

In order to avoid a discussion of possible degeneration of eigen values it is convenient to deal with symmetric functions of all eigen values and their variation, instead of considering individual eigen values. For this purpose, we define in § 7 the Fredholm determinant of a domain; this concept is rather natural when one comes from the general theory of integral equations. The variational formula for the Fredholm determinant is easily expressed in terms of a complex kernal closely connected with the dielectric Green's function which possesses, moreover, as limit case a kernel well-known in the theory of conformal mapping. Indeed, the variation of the Fredholm determinant for the particular value 1 of the argument is described by this classical kernel itself.

In § 8, at last, we apply the results of the preceding section in order to solve an extremum problem for univalent functions in a multiply-connected domain and involving the Fredholm determinant. This solution gives a new proof for the possibility to map every domain conformally onto a domain bounded by circumferences and characterizes this canonical domain as an extremum domain of a simple variational problem. The treatment of the variational problem for the Fredholm determinant seems also of interest from the methodological point of view and for the general theory of variations of domain functions. In general, one knows from the theory of normal families that a solution of an extremum problem for the family of functions, univalent in a given domain and with specified normalization, does exist; the method of variations has only the task to characterize the extremum domain. In our present problem, we had to restrict ourselves to univalent functions which are analytic in the closed domain in order to be sure of the existence of the Fredholm determinant. In this case, the theory of normal families does not guarantee the existence of an extremum function of equal character. We do not characterize, therefore, the extremum function by our variations, but rather an extremum sequence within the function class, considered. We prove from the very extremum property of the sequence that its limit function does, indeed, belong to the same class and has, moreover, certain characterizing properties. This procedure is very general and may have numerous analogous applications.

**1. The Fredholm eigen value problem.** Let $\tilde{D}$ be a domain in the complex $z$-plane containing the point at infinity; let its boundary $C$ consist of $N$ closed curves $C_j$ each of which is three times continuously differentiable. We denote the interior of each $C_j$ by $D_j$ and the union of the $N$ domains $D_j$ by $D$.

We define the kernel

$$k(z, \zeta) = \frac{\partial}{\partial n_\zeta} \log \frac{1}{|z - \zeta|} \qquad \zeta \in C \tag{1}$$

where $\boldsymbol{n}_\zeta$ denotes the normal of $C$ at $\zeta$ pointing into $D$. It is well known that, under our assumptions about $C$, the kernel $k(z, \zeta)$ is continuous in both its arguments as long as they are restricted to $C$.

We want to discuss the eigen value problem

$$\varphi_\nu(z) = \frac{\lambda_\nu}{\pi}\int_C k(z, \zeta)\varphi_\nu(\zeta)ds_\zeta\,, \qquad z \in C \tag{2}$$

which plays an important role in many boundary value problems of potential theory with respect to the multiply-connected domain $\tilde{D}$. The $\varphi_\nu(z)$ and the $\lambda_\nu$ are called the Fredholm eigen functions and the Fredholm eigen values, respectively, of the curve system $C$. The study of the Fredholm eigen value problem is facilitated by the fact that the kernel $k(z, \zeta)$ is, for fixed $\zeta \in C$, defined and harmonic for all values $z \neq \zeta$ in the complex plane. The integral in (2) represents, therefore, a harmonic function in $\tilde{D}$ and a set of different harmonic functions in $D$. We shall use the notation

$$\frac{\lambda_\nu}{\pi}\int_C k(z, \zeta)\varphi_\nu(\zeta)ds_\zeta = \begin{cases} h_\nu(z) & \text{for} \quad z \in D \\ \tilde{h}_\nu(z) & \text{for} \quad z \in \tilde{D}\,. \end{cases} \tag{3}$$

The set of harmonic functions $\tilde{h}_\nu(z)$ and $h_\nu(z)$ can be interpreted as the potential due to a double layer of logarithmic charges, spread along $C$ with the density $(\lambda_\nu/\pi)\varphi_\nu(\zeta)$. Hence, the well known discontinuity character of such potentials leads to the boundary relations at each point

$$\lim_{z\to z_0} h_\nu(z) = (1 + \lambda_\nu)\varphi_\nu(z_0)\,, \quad \lim_{z\to z_0} \tilde{h}_\nu(z) = (1 - \lambda_\nu)\varphi_\nu(z_0)\,, \tag{4}$$

and

$$\frac{\partial}{\partial n} h_\nu(z_0) = -\frac{\partial}{\partial \tilde{n}} \tilde{h}_\nu(z_0)\,, \tag{4'}$$

where $\tilde{n}$ denotes the normal of $C$ pointing into $\tilde{D}$.

The Fredholm eigen value problem may thus be formulated as the following question of potential theory which is of interest by itself:

To determine a harmonic function $\tilde{h}$ in $\tilde{D}$ and a set of harmonic functions $h$ in $D$ which have equal normal derivatives and proportional boundary values on $C$!. It is easily seen that the two problems are completely equivalent and that the possible factors of proportionality in the second problem are simple functions of the Fredholm eigen values $\lambda_\nu$.

Instead of the harmonic functions $h_\nu$ and $\tilde{h}_\nu$, we may consider their complex derivatives, i.e., the analytic functions

$$v_\nu(z) = \frac{\partial}{\partial z} h_\nu(z) , \quad \tilde{v}_\nu(z) = \frac{\partial}{\partial z} \tilde{h}_\nu(z) . \tag{5}$$

In view of definition (3) and by our assumption on $C$ it can be asserted that $v_\nu$ and $\tilde{v}_\nu$ are continuous in $D + C$ and $\tilde{D} + C$, respectively. In order to translate the relations (4) and (4′) into terms involving $v_\nu$ and $\tilde{v}_\nu$, we use the parametric representation $z = z(s)$ of $C$ by means of the arc length $s$ and introduce

$$z' = \frac{dz}{ds} , \tag{6}$$

the unit vector at $z(s)$ in direction of the tangent of $C$. We can then write (4) and (4′) in the form

$$\Re\left\{\frac{\partial h_\nu}{\partial z} z'\right\} = \frac{1 + \lambda_\nu}{1 - \lambda_\nu} \Re\left\{\frac{\partial \tilde{h}_\nu}{\partial z} z'\right\} , \quad \Im\left\{\frac{\partial h_\nu}{\partial z} z'\right\} = \Im\left\{\frac{\partial \tilde{h}_\nu}{\partial z} z'\right\} , \tag{7}$$

and combine these two equation into the one complex equation

$$v_\nu(z) z' = \frac{1}{1 - \lambda_\nu} \tilde{v}_\nu(z) z' + \frac{\lambda_\nu}{1 - \lambda_\nu} \overline{\tilde{v}_\nu(z) z'} , \qquad z = z(s) . \tag{8}$$

Introducing (8) into the Cauchy identity. We obtain for $\zeta \in D$

$$v_\nu(\zeta) = \frac{1}{2\pi i} \oint_C \frac{v_\nu(z)}{z - \zeta} dz = \frac{\lambda_\nu}{1 - \lambda_\nu} \frac{1}{2\pi i} \oint_C \frac{\overline{(\tilde{v}_\nu(z) dz)}}{z - \zeta} \tag{9}$$

while the use of the equation conjugate to (8) leads to

$$\frac{1}{2\pi i} \oint_C \frac{\overline{(v_\nu(z) dz)}}{z - \zeta} = \frac{1}{1 - \lambda_\nu} \frac{1}{2\pi i} \oint_C \frac{\overline{(\tilde{v}_\nu(z) dz)}}{z - \zeta} , \qquad \zeta \in D . \tag{10}$$

Combining (9) and (10), we arrive thus at the following integral equation for $v_\nu$:

$$v_\nu(\zeta) = \frac{\lambda_\nu}{2\pi i} \oint_C \frac{\overline{(v_\nu(z) dz)}}{z - \zeta} . \tag{11}$$

In the same way we prove the analogous equations

$$(9') \qquad \tilde{v}_\nu(\zeta) = \frac{\lambda_\nu}{1-\lambda_\nu} \cdot \frac{1}{2\pi i} \oint_C \frac{\overline{(v_\nu(z)dz)}}{z-\zeta}$$

and

$$(11') \qquad \tilde{v}_\nu(\zeta) = \frac{\lambda_\nu}{2\pi i} \oint_C \frac{\overline{(\tilde{v}_\nu(z)dz)}}{z-\zeta} .$$

In all these formulas the integration over the curve system $C$ has to be performed in the positive sense with respect to $D$.

The line integrals in (9), (9′) and (11), (11′) can be transformed into area integrals and the integral equations take the forms

$$(12) \qquad \frac{\lambda_\nu}{\pi} \iint_D \frac{\overline{v_\nu(z)}}{(z-\zeta)^2} d\tau_z = \begin{cases} v_\nu(\zeta) & \text{for } \zeta \in D \\ (1+\lambda_\nu)\tilde{v}_\nu(\zeta) & \text{for } \zeta \in \tilde{D} \end{cases}$$

and

$$(13) \qquad -\frac{\lambda_\nu}{\pi} \iint_{\tilde{D}} \frac{\overline{\tilde{v}_\nu(z)}}{(z-\zeta)^2} d\tau_z = \begin{cases} (1-\lambda_\nu)v_\nu(\zeta) & \text{for } \zeta \in D \\ \tilde{v}_\nu(\zeta) & \text{for } \zeta \in \tilde{D} . \end{cases}$$

In both integrals $d\tau_z$ denotes the area element with respect to the variable $z$ and the integrals have to be interpreted in the Cauchy principal sense whenever they become improper.

The transformation

$$(14) \qquad F(\zeta) = \frac{1}{\pi} \iint_E \frac{\overline{f(z)}}{(z-\zeta)^2} d\tau_z$$

carries every $L^2$-integrable function $f(z)$ defined in the complex plane $E$ into a new function $F(z)$ of the same class and with the same norm:

$$(15) \qquad \iint_E |F(z)|^2 d\tau = \iint_E |f(z)|^2 d\tau .$$

This functional transformation plays a role in many problems of function theory [1, 3, 4] and is called the "Hilbert integral transformation". The integral equations (12) and (13) show the close connection between the theories of the Fredholm eigen functions and of the Hilbert transforms of analytic functions.

We introduce next the Green's functions of the domain $\tilde{D}$ and of the set of domains $D$. While the Green's function $\tilde{g}(z, \zeta)$ of $\tilde{D}$ is defined as usual, the Green's function $g(z, \zeta)$ of $D$ is given by the equation

$$(16) \qquad g(z, \zeta) = \begin{cases} g_j(z, \zeta) & \text{for } z, \zeta \in D_j \\ 0 & \text{for } z \in D_j \, \zeta \in D_l, \, l \neq j . \end{cases}$$

Here, $g_j(z, \zeta)$ is the usual Green's function of the domain $D_j$. By complex differentiation, we derive from $g(z, \zeta)$ the analytic function

$$L(z, \zeta) = -\frac{2}{\pi}\frac{\partial^2}{\partial z \partial \zeta} g(z, \zeta) = \frac{1}{\pi(z-\zeta)^2} - l(z, \zeta) . \tag{17}$$

The kernels $L(z, \zeta)$ and $l(z, \zeta)$ are well known in the case that $D$ is a domain [3, 16]. We observe that our generalized kernel $l(z, \zeta)$ still preserves the following important property: If $f(z)$ is regular analytic in $D$, then

$$\frac{1}{\pi}\iint_D \frac{\overline{f(z)}}{(z-\zeta)^2} d\tau = \iint_D l(z, \zeta)\overline{f(z)} d\tau . \tag{18}$$

In fact, if $\zeta \in D_j$ then $l(z, \zeta) = l_j(z, \zeta)$ for $z \in D_j$ and $l(z, \zeta) = [\pi(z-\zeta)^2]^{-1}$ for $z \in D_l$, $l \neq j$. The identity (18) follows, therefore, directly from the corresponding property of the kernel $l_j(z, \zeta)$.

In particular, we may formulate the integral equations (12) and (13) for $v_\nu(z)$ and $\tilde{v}_\nu(z)$ as follows:

$$\lambda_\nu \iint_D l(z, \zeta)\overline{v_\nu(z)} d\tau = v_\nu(\zeta) , \qquad \zeta \in D \tag{19}$$

and

$$-\lambda_\nu \iint_{\tilde{D}} \tilde{l}(z, \zeta)\overline{\tilde{v}_\nu(\zeta)} d\tau = \tilde{v}_\nu(\zeta) , \qquad \zeta \in \tilde{D} . \tag{20}$$

From the symmetry of the kernels $l(z, \zeta)$ and $\tilde{l}(z, \zeta)$ we can conclude

$$\iint_D v_\nu \bar{v}_\mu d\tau = 0 \quad \text{if} \quad \lambda_\nu \neq \lambda_\mu \tag{21}$$

$$\iint_{\tilde{D}} \tilde{v}_\nu \bar{\tilde{v}}_\mu d\tau = 0 \quad \text{if} \quad \lambda_\nu \neq \lambda_\mu . \tag{21'}$$

Thus, using a familiar argument from theory of integral equation we may assume that any pair of different eigen functions $v_\nu, v_\mu$ (or $\tilde{v}_\nu. \tilde{v}_\mu$) are orthogonal upon each other:

$$\iint_D v_\nu \bar{v}_\mu d\tau = 0 ; \iint_{\tilde{D}} \tilde{v}_\nu \bar{\tilde{v}}_\mu d\tau = 0 \quad \text{for} \quad \nu \neq \mu . \tag{21''}$$

There remains the question of normalizing the $v_\nu$ and the $\tilde{v}_\nu$. We have obviously the free choice of a real multiplicator in the definition of $v_\nu$; however, this choice will already determine the function $\tilde{v}_\nu$ in a unique way, for example through equation (12). The relation between the norms of $v_\nu$ and $\tilde{v}_\nu$ is best understood by returning to the harmonic functions $h_\nu(z)$ and $\tilde{h}_\nu(z)$ and to their boundary relations (4) and (4′). In fact, we have

$$\iint_D |v_\nu|^2 d\tau = \frac{1}{4}\iint_D |\nabla h_\nu|^2 d\tau = -\frac{1}{4}\oint_C h_\nu \frac{\partial h_\nu}{\partial n} ds \tag{22}$$
$$= \frac{1}{4}\frac{1+\lambda_\nu}{1-\lambda_\nu}\oint_C \tilde{h}_\nu \frac{\partial \tilde{h}_\nu}{\partial \tilde{n}} ds = \frac{\lambda_\nu + 1}{\lambda_\nu - 1}\iint_{\tilde{D}} |\tilde{v}_\nu|^2 d\tau .$$

We can conclude first from (22) that

$$|\lambda_\nu| \geq 1 . \tag{23}$$

Let us consider the limit cases $\lambda_\nu = \pm 1$. For $\lambda_\nu = 1$ we have necessarily $\tilde{v}_\nu(z) \equiv 0$; the second equation (7) yields

$$\Im\{v_\nu(z) z'\} = 0 \quad \text{for} \quad z \in C . \tag{24}$$

Thus, the eigen function $v_\nu(z)$ is a real differential for each component domain $D_j$. But a simply-connected domain $D_j$ cannot have such real differentials; hence also $v_\nu(z) \equiv 0$. Thus, as far as the integral equation for $v_\nu$ and $\tilde{v}_\nu$ are concerned, $\lambda_\nu = 1$ cannot occur as an eigen value. The situation is, however, different when we return to the original integral equation (2) and to the harmonic functions $h_\nu$ and $\tilde{h}_\nu$. To $\lambda_\nu = 1$ must correspond

$$h_\nu(z) \equiv 2c_j \text{ in } D_j , \ \tilde{h}_\nu(z) \equiv 0 \tag{25}$$

and

$$\varphi_\nu(z) = c_j \quad \text{on} \quad C_j . \tag{25′}$$

In fact, it is immediately verified that for arbitrary choice of the constants $c_j$ the function $\varphi(z) = c_j$ on $C_j$ is a solution of the Fredholm eigen value problem (2) to the eigen value $\lambda_\nu = 1$. There exist thus $N$ linearly independent solutions of (2) to the eigen value $\lambda = 1$. These solutions disappear when we replace the original integral equation (2) by the integral equations for $v_\nu$ and $\tilde{v}_\nu$, say, by (12) and (13). It is easy to show that the eigen value $\lambda = 1$ is the only one lost in this transition.

We consider next the case $\lambda_\nu = -1$. We conclude now from (22) that $v_\nu(z) \equiv 0$. We find therefore, in view of (8)

$$\Im\{\tilde{v}_\nu(z) z'\} = 0 \quad \text{for} \quad z \in C , \tag{26}$$

i.e., $\tilde{v}_\nu(z)$ is a real differential of $\tilde{D}$. There are $N-1$ linearly independent differentials of this type in $\tilde{D}$ and we can construct a basis for them as follows. Let $\omega_j(z)$ be harmonic in $\tilde{D}$ and satisfy on $C$ the boundary condition

(27) $$\omega_j(z) = \delta_{jl} \quad \text{for} \quad z \in C_l \, .$$

$\omega_j(z)$ is called the harmonic measure of $C_j$ with respect to $z$ of $\tilde{D}$. Clearly, each function

(28) $$\tilde{w}_j(z) = i\frac{\partial \omega_j}{\partial z}$$

is a real differential in $\tilde{D}$. Since $\sum_{j=1}^{N} \omega_j \equiv 1$, we have $\sum_{j=1}^{N} \tilde{w}_j(z) \equiv 0$. But it is easily seen that apart from this relation no other linear condition between the $w_j$ does exist. Thus, we can select any $N-1$ of the $w_j(z)$ as a basis for all real differentials in $\tilde{D}$.

It is clear that each real differential in $\tilde{D}$ satisfies indeed the integral equations (12) and (13). However, there exists no corresponding single valued harmonic function $\tilde{h}_\nu(z)$ connected with the original Fredholm equation (2) which has this real differential as its complex derivative. Indeed, in view of (26) such function would have to satisfy the boundary condition

(29) $$\frac{\partial \tilde{h}_\nu}{\partial n} \equiv 0 \quad \text{on} \quad C$$

which admits only the solution $\tilde{h}_\nu = \text{const.}$ and could not lead to a non-vanishing differential. Thus, while we lost in the transition to (12) and (13) the $N$ eigen functions to the eigen value $\lambda = +1$, we have obtained $N-1$ new eigen functions to the eigen value $\lambda = -1$ which have no counterpart in the original Fredholm equation.

After discussing the exceptional cases $\lambda_\nu = \pm 1$, we consider now the eigen functions $v_\nu(z)$ and $\tilde{v}_\nu(z)$ which belong to eigen values $|\lambda_\nu| > 1$. Each such pair is obtained by complex differentiation from a pair of harmonic functions $h_\nu(z), \tilde{h}_\nu(z)$ connected with the original Fredholm problem. Since $h_\nu(z)$ is harmonic in each of the simply-connected domains $D_j$, it can be completed to a set of single-valued analytic functions in the set of domains $D_j$:

(30) $$V_\nu(z) = h_\nu(z) + ik_\nu(z) \, .$$

Similarly, we may complete $\tilde{h}_\nu$ in $\tilde{D}$ and define

(31) $$\tilde{V}_\nu(z) = \tilde{h}_\nu(z) + i\tilde{k}_\nu(z) \, .$$

From the boundary conditions (4) and (4′) and from the Cauchy-Riemann equations we derive the boundary conditions for the $k_\nu$:

(32) $$\tilde{k}_\nu(z) = k_\nu(z), \quad \frac{\partial}{\partial n} k_\nu(z) = \frac{1+\lambda_\nu}{1-\lambda_\nu} \frac{\partial}{\partial n} \tilde{k}_\nu(z) \, , \qquad z \in C \, .$$

Equations (32) guarantee that $\tilde{k}_\nu(z)$ is single-valued in $\tilde{D}$ since $k_\nu(z)$ is single-valued in each $D_j$. We may characterize the single-valued analytic functions $V_\nu(z)$ and $\tilde{V}_\nu(z)$ as follows: Their real parts have equal normal derivatives on $C$ while their boundary values are proportional in the ratio $(1+\lambda_\nu)/(1-\lambda_\nu)$. Their imaginary parts are equal on $C$ but their normal derivatives are proportional with the same ratio.

Let us write $k_\nu^{(1)} = (1-\lambda_\nu)k_\nu$ and $\tilde{k}_\nu^{(1)} = (1+\lambda_\nu)\tilde{k}_\nu$; we have on $C$

$$k_\nu^{(1)}(z) = \frac{1-\lambda_\nu}{1+\lambda_\nu}\tilde{k}_\nu^{(1)}(z), \quad \frac{\partial k_\nu^{(1)}}{\partial n} = -\frac{\partial \tilde{k}_\nu^{(1)}}{\partial \tilde{n}} . \tag{32'}$$

Thus, $k_\nu^{(1)}$ and $\tilde{k}_\nu^{(1)}$ may be conceived as a pair of $k$-functions belonging to eigen functions of the Fredholm problem (2) with the eigen value $-\lambda_\nu$. With each eigen value $\lambda_\nu$ with $|\lambda_\nu| > 1$ there occurs also its negative $-\lambda_\nu$ as an eigen value. Their corresponding $h$-functions are, up to a factor, conjugate harmonic functions.

Finally, we introduce the analytic functions

$$u_\nu(z) = \sqrt{\lambda_\nu - 1}\, v_\nu(z), \quad \tilde{u}_\nu(z) = i\sqrt{\lambda_\nu + 1}\, \tilde{v}_\nu(z) . \tag{33}$$

By virtue of (21″) and (22), we may assume that these functions form orthonormalized sets in $D$ and $\tilde{D}$; that is

$$\iint_D u_\nu \bar{u}_\mu d\tau = \delta_{\nu\mu}, \quad \iint_{\tilde{D}} \tilde{u}_\nu \bar{\tilde{u}}_\mu d\tau = \delta_{\nu\mu} . \tag{34}$$

Since the $u$-functions will be frequently used in this paper, we note here some formulas which follow immediately from the corresponding results for the $v$-functions. From (8) we derive the boundary relation

$$u_\nu(z)z' = \frac{i}{\sqrt{\lambda_\nu^2 - 1}}\tilde{u}_\nu(z)z' - \frac{\lambda_\nu i}{\sqrt{\lambda_\nu^2 - 1}}\overline{(\tilde{u}_\nu(z)z')} . \tag{35}$$

Equations (9), (9′) and (11), (11′) take on the form

$$\frac{\lambda_\nu}{2\pi i}\oint_C \frac{\overline{(\tilde{u}_\nu dz)}}{z-\zeta} = \begin{cases} i\sqrt{\lambda_\nu^2 - 1}\, u_\nu(\zeta) & \text{for } \zeta \in D \\ -\tilde{u}_\nu(\zeta) & \text{for } \zeta \in \tilde{D} . \end{cases} \tag{36}$$

and

$$\frac{\lambda_\nu}{2\pi i}\oint_C \frac{\overline{(u_\nu dz)}}{z-\zeta} = \begin{cases} u_\nu(\zeta) & \text{for } \zeta \in D \\ -i\sqrt{\lambda_\nu^2 - 1}\, \tilde{u}_\nu(\zeta) & \text{for } \zeta \in \tilde{D} . \end{cases} \tag{37}$$

From their connection with the Fredholm integral equation it can be shown that the $u_\nu(z)$ form a complete system of analytic functions in $D$, in the sense that every function $f(z)$ which is analytic in $D$ and for which $\iint_D |f|^2 d\tau < \infty$ can be represented in the form

$$f(z)=\sum_{\nu=1}^{\infty}a_\nu u_\nu(z),\ a_\nu=\iint_D f\bar{u}_\nu d\tau\,. \tag{38}$$

The series converges uniformly in each closed subdomain of $D$. In the same sense, the functions $\tilde{u}_\nu(z)$ form a complete orthonormal system within the class of all functions which are analytic in $\tilde{D}$, have a finite norm in $\tilde{D}$ and possess a finite single-valued integral in this multiply-connected domain. If we add to the $\{\tilde{u}_\nu\}$-set any $N-1$ linearly independent real differentials of $\tilde{D}$ we obtain a complete system for all analytic functions in $\tilde{D}$ with finite norm and vanishing at infinity [3, 21].

2. **The dielectric Green's function.** The theory of the Green's function of the domain $\tilde{D}$ is connected with the electrostatic problem of a point charge at a source point $\zeta$ in the presence of the system of grounded conductors $C_j$. We may consider also the problem to determine the electrostatic potential induced by the same point charge at $\zeta$ in the presence of $N$ isotropic dielectric media which are spread over the domains $D_j$ and have the dielectric constant $\varepsilon$. The corresponding potential $g_\varepsilon(z,\zeta)$ will now be defined in $D$ as well as in $\tilde{D}$ and will be characterized by the following properties:

(a) $g_\varepsilon(z,\zeta)$ is a harmonic function of $z$ in $D$ and in $\tilde{D}$, except for $z=\zeta$ and for $z=\infty$.

(b) If $\zeta\in\tilde{D}$, the function $g_\varepsilon(z,\zeta)+\log|z-\zeta|$ is harmonic at $\zeta$.

(b′) It $\zeta\in D$, the function $g_\varepsilon(z,\zeta)+\varepsilon\log|z-\zeta|$ is harmonic at $\zeta$.

(c) $g_\varepsilon(z,\zeta)$ is continuous through $C$.

(d) $\frac{\partial}{\partial n_z}g_\varepsilon(z,\zeta)+\varepsilon\frac{\partial}{\partial\tilde{n}_z}g_\varepsilon(z,\zeta)=0$ for $z\in C,\zeta$ in $D$ or in $\tilde{D}$.

(e) $g_\varepsilon(z,\zeta)+\log|z|\to 0$ as $z\to\infty$ for $\zeta$ fixed.

If such a function $g_\varepsilon(z,\zeta)$ exists it must be unique and symmetric in its two arguments, as is shown by the standard argument of potential theory based on the second Green's identity. In order to construct the Green's function, we set it up in the form

$$g_\varepsilon(z,\zeta)=\log\frac{1}{|z-\zeta|}+\int_C\mu(\eta,\zeta)\log|\eta-z|\,ds_\eta,\ \zeta\in\tilde{D} \tag{1}$$

and try to determine $\mu(\eta,\zeta)$ in such a way that the above requirements are fulfilled. We proceed analogously, if $\zeta\in D$; only the singularity term on the right side of (1) will now be $-\varepsilon\log|z-\zeta|$. By this formal set up, we have already fulfilled conditions (a) to (c). Condition (e) is satisfied if we require

$$\int_C\mu(\eta,\zeta)ds_\eta=\begin{cases}\varepsilon-1 & \text{for}\quad \zeta\in D\\ 0 & \text{for}\quad \zeta\in\tilde{D}\,.\end{cases} \tag{2}$$

Finally, we can satisfy (d) by choosing the density function $\mu$ of the line potential as solution of the integral equation

$$(3)\qquad \mu(z,\zeta)+\frac{\varepsilon-1}{\varepsilon+1}\frac{1}{\pi}\int_C \mu(\eta,\zeta)k(\eta,z)ds_\eta=\begin{cases}\dfrac{\varepsilon-1}{\varepsilon+1}\cdot\dfrac{\varepsilon}{\pi}k(\zeta,z) & \text{for } \zeta\in D\\[2mm] \dfrac{\varepsilon-1}{\varepsilon+1}\,\dfrac{1}{\pi}k(\zeta,z) & \text{for } \zeta\in\tilde{D}\,.\end{cases}$$

Here $k(\zeta, z)$ is defined by equation (1.1). We observe that

$$(4)\qquad \int_C k(\eta,z)ds_z=\begin{cases}0 & \text{for } \eta\in\tilde{D}\\ \pi & \text{for } \eta\in C\\ 2\pi & \text{for } \eta\in D\,.\end{cases}$$

Hence, if $\mu(z, \zeta)$ is a solution of the integral equation (3) we may integrate this equation with respect to $z$ over $C$ and verify that condition (2) is fulfilled automatically. It is sufficient, therefore, to concentrate upon the inhomogeneous integral equation (3).

For physical reasons, we shall assume $\varepsilon > 0$. In this case, we always have

$$(3')\qquad \left|\frac{\varepsilon-1}{\varepsilon+1}\right|<1\,.$$

Since we showed in § 1 that all eigen values of the kernel $k(z, \zeta)$ have absolute values $\geq 1$, it follows that integral equation (3) can always be solved by the usual process of iteration and that the solution can be represented by a Liouville-Neumann series. The convergence of this series will be the better, the nearer $\varepsilon$ will be to 1. We observe that

$$(5)\qquad g_1(z,\zeta)=\log\frac{1}{|z-\zeta|}$$

is trivially known.

The function

$$(6)\qquad \gamma_\varepsilon(z,\zeta)=g_\varepsilon(z,\zeta)-\log\frac{1}{|z-\zeta|}$$

is (for $\zeta \in D$ or for $\zeta \in \tilde{D}$) a regular harmonic function of $z$ in $\tilde{D}$, vanishes if $z$ tends to infinity and possesses a single-valued conjugate harmonic function in $\tilde{D}$. This last fact follows from the boundary condition (d) on the dielectric Green's function and the fact that each complementary domain $D_j$ is simply-connected. Let $\tilde{\Sigma}$ be the class of all functions $\tilde{h}(s)$ which are harmonic in $\tilde{D}$, vanish at infinity and have a single-valued conjugate harmonic function. It is easy to show that the harmonic functions $\tilde{h}_\nu(s)$ which belong to eigen values $|\lambda_\nu| > 1$ of the Fredholm problem in § 1 form a basis in the linear space $\tilde{\sum}$. By virtue of (1.5) and (1.21′), we have

$$\iint_{\tilde{D}} \nabla\tilde{h}_\nu \cdot \nabla\tilde{h}_\mu d\tau = 4\Re\left\{\iint_D \tilde{v}_\nu \bar{\tilde{v}}_\mu d\tau\right\} = 0 \qquad \text{for } \nu \neq \mu . \tag{7}$$

By a trivial renormalization we can then achieve that

$$\iint_{\tilde{D}} \nabla\tilde{h}_\nu \cdot \nabla\tilde{h}_\mu d\tau = \delta_{\nu\mu} . \tag{8}$$

We wish now to develop $\gamma_\varepsilon(z, \zeta)$ in terms of the complete orthonormal set $\{\tilde{h}_\nu\}$. In order to determine the Fourier coefficients or $\gamma_\varepsilon(z, \zeta)$, we consider the Dirichlet integrals

$$j_\nu(\zeta) = \iint_D \nabla g_\varepsilon(z, \zeta) \cdot \nabla h_\nu(z) d\tau_z + \iint_{\tilde{D}} \nabla g_\varepsilon(z, \zeta) \cdot \nabla \tilde{h}_\nu(z) d\tau_z . \tag{9}$$

We integrate first by parts with respect to $g_\varepsilon(z, \zeta)$ and use the continuity of this function across $C$ as well as the relation (1.4′) for the normal derivatives of $h_\nu$ and $\tilde{h}_\nu$ on $C$. We find

$$j_\nu(\zeta) \equiv 0 . \tag{10}$$

Next, we integrate by parts with respect to $h_\nu(z)$ and $\tilde{h}_\nu(z)$; we use (1.4) and the condition (d) on $g_\varepsilon(z, \zeta)$. We obtain the equations

$$j_\nu(\zeta) = 2\pi\varepsilon h_\nu(\zeta) - (1 + \varepsilon\rho_\nu) \int_C \frac{\partial g_\varepsilon(z, \zeta)}{\partial \tilde{n}_z} \tilde{h}_\nu(z) ds_z \quad \text{for} \quad \zeta \in D \tag{11}$$

and

$$j_\nu(\zeta) = 2\pi\tilde{h}_\nu(\zeta) - \frac{1 + \varepsilon\rho_\nu}{\varepsilon\rho_\nu} \int_C \frac{\partial g_\varepsilon(z, \zeta)}{\partial n_z} h_\nu(z) ds_z \quad \text{for} \quad \zeta \in \tilde{D} . \tag{11′}$$

Here, we have introduced the abbreviation

$$\rho_\nu = \frac{\lambda_\nu + 1}{\lambda_\nu - 1} ; \tag{12}$$

this simple function of $\lambda_\nu$ will occur frequently in our developments.

From (10), (11) and (11′) we deduce immediately

$$\iint_{\tilde{D}} \nabla g_\varepsilon(z, \zeta) \cdot \nabla\tilde{h}_\nu(z) d\tau_z = -\frac{2\pi\varepsilon}{1 + \varepsilon\rho_\nu} h_\nu(\zeta) \quad \text{for} \quad \zeta \in D \tag{13}$$

and

$$(13') \qquad \iint_{\tilde{D}} \nabla g_\varepsilon(z,\zeta)\cdot\nabla\tilde{h}_\nu(z)d\tau_z = \frac{2\pi\varepsilon\rho_\nu}{1+\varepsilon\rho_\nu}\tilde{h}_\nu(\zeta) \quad \text{for} \quad \zeta\in\tilde{D}.$$

When we specialize $\varepsilon = 1$, we obtain because of (5) the values of the left-hand integrals with $g_\varepsilon$ replaced by $\log 1/|z-\zeta|$. Hence, we obtain finally by subtraction

$$(14) \qquad \iint_{\tilde{D}} \nabla\gamma_\varepsilon(z,\zeta)\cdot\Delta\tilde{h}_\nu(z)d\tau_z = \frac{2\pi(1-\varepsilon)}{(1+\rho_\nu)(1+\varepsilon\rho_\nu)}h_\nu(\zeta) \quad \text{for} \quad \zeta\in D\,.$$

and

$$(14') \qquad \iint_{\tilde{D}} \nabla\delta_\varepsilon(z,\zeta)\nabla\tilde{h}_\nu(z)d\tau_z = \frac{2\pi(\varepsilon-1)}{(1+\rho_\nu)(1+\varepsilon\rho_\nu)}\tilde{h}_\nu(\zeta) \quad \text{for} \quad \zeta\in\tilde{D}\,.$$

Having expressed by (14) and (14′) the Fourier coefficients of $\gamma_\varepsilon(z,\zeta)$ with respect to the complete orthonormal system in $\tilde{\sum}$, we obtain thus the two series development for $z\in\tilde{D}$;

$$(15) \quad g_\varepsilon(z,\zeta) = \log\frac{1}{|z-\zeta|} + 2\pi(1-\varepsilon)\sum_{\nu=1}^{\infty}\frac{\tilde{h}_\nu(z)h_\nu(\zeta)}{(1+\rho_\nu)(1+\varepsilon\rho_\nu)} \quad \text{for} \quad \zeta\in D$$

$$(16) \quad g_\varepsilon(z,\zeta) = \log\frac{1}{|z-\zeta|} + 2\pi(\varepsilon-1)\sum_{\nu=1}^{\infty}\frac{\rho_\nu\tilde{h}_\nu(z)\tilde{h}_\nu(\zeta)}{(1+\rho_\nu)(1+\varepsilon\rho_\nu)} \quad \text{for} \quad \zeta\in\tilde{D}\,.$$

Both series converge uniformly in each closed subdomain of $\tilde{D}$.

We wish next to expand analogously $g_\varepsilon(z,\zeta)$ for $z\in D$ in terms of the functions $h_\nu(z)$. By (1.4), (1.4′) and the normalization (8), we have

$$(17) \qquad \iint_D \nabla h_\nu\cdot\nabla h_\mu d\tau = \rho_\nu\delta_{\nu\mu}\,.$$

Let $\omega_j(z)$ and $\tilde{g}(z,\infty)$ denote again the $j$-th harmonic measure and the Green's function with pole at infinity of $\tilde{D}$. We clearly have

$$(18) \qquad \int_C h_\nu\frac{\partial\omega_j}{\partial n}ds = 0, \quad \int_C h_\nu\frac{\partial\tilde{g}(z,\infty)}{\partial u}ds = 0\,.$$

Indeed, because of (1.4) these linear conditions are equivalent to those with $\tilde{h}_\nu$ and these in turn follow from the fact that all $\tilde{h}_\nu$ have single-valued harmonic conjugates in $\tilde{D}$ and that they all vanish at infinity.

Let $\sum$ be the linear space of functions $h(z)$ which are regular hamonic in $D$ and which satisfy the $N$ linear conditions (18). Observe that $\sum$ does not contain any function $h_0(z)$ which has a constant value $c_j$ in each $D_j$, except for $h_0(z)\equiv 0$. Indeed, the conditions (18) would yield for such a function $h_0(z)$

$$(18') \qquad \sum_{j=1}^{n} c_j p_{ij} = 0, \quad \sum_{j=1}^{n} c_j\omega_j(\infty) = 0$$

where

$$p_{ij} = \frac{1}{2\pi}\int_{C_j} \frac{\partial \omega_i}{\partial n} ds \tag{18''}$$

denotes the period matrix connected with the harmonic measures. But the first system of linear equations (18′) implies clearly [5, 15] $c_1 = c_2 = \cdots = c_N = c$ and the last equation yields

$$c \sum_{j=1}^{N} \omega_j(\infty) = c = 0 \,. \tag{19}$$

Thus, only the trivial function $h_0(z) \equiv 0$ of this type lies in $\sum$.

From this fact and the considerations of § 1, it follows that the functions $\{\rho_\nu^{1/2} h_\nu(z)\}$ form a complete orthonormal set in $\sum$. The function $\gamma_\varepsilon(z, \zeta)$ lies in $\sum$ if $\zeta \in \tilde{D}$; this follows at once from the conditions (c), (d) and (e) on the dielectric Green's function. If $\zeta \in D$, it is seen that $\gamma_\varepsilon(z, \zeta) + (1 - \varepsilon)g(z, \zeta)$ lies in $\sum$ where $g(z, \zeta)$ is the Green's function of $D$ defined by (1.16). The Fourier coefficients of $\gamma_\varepsilon(z, \zeta)$ are easily determined from (9), (10), (13) and (13′). Observe that for $\zeta \in D_j$

$$\iint_D \nabla g(z, \zeta) \cdot \nabla h_\nu(z) d\tau_z = - \int_{C_j} \frac{\partial h_\nu(z)}{\partial n} g(z, \zeta) ds_z = 0 \tag{20}$$

such that the correction term $(1 - \varepsilon)g(z, \zeta)$ does not affect the Fourier coefficients at all. We find without difficulty

$$\begin{aligned} g_\varepsilon(z, \zeta) = {} & \log \frac{1}{|z - \zeta|} + (\varepsilon - 1)g(z, \zeta) \\ & + 2\pi(\varepsilon - 1) \sum_{\nu=1}^{\infty} \frac{h_\nu(z) h_\nu(\zeta)}{\rho_\nu(1 + \rho_\nu)(1 + \varepsilon\rho_\nu)} \qquad \text{for} \quad \zeta \in D \end{aligned} \tag{21}$$

$$g_\varepsilon(z, \zeta) = \log \frac{1}{|z - \zeta|} + 2\pi(1 - \varepsilon) \sum_{\nu=1}^{\infty} \frac{h_\nu(z) \tilde{h}_\nu(\zeta)}{(1 + \rho_\nu)(1 + \varepsilon\rho_\nu)} \qquad \text{for} \quad \zeta \in \tilde{D} \,. \tag{22}$$

These series also converge uniformly in each closed subdomain of $D$. Equation (22) could have been derived from (15) and the property of symmetry of the dielectric Green's function in dependence of its two arguments.

The various series developments for $g_\varepsilon(z, \zeta)$ given so far are of theoretical interest and allow the derivation of numerous identities. They help little in the actual determination of the dielectric Green's function of a given domain since we know all Fredholm eigen functions and eigen values only in very few cases. In order to utilize the preceding formulas for actual calculations, we have to add the following considerations.

From the definition of the dielectric Green's functions and from Green's identity, one can derive the identity

$$(23)\qquad \frac{1}{e}\iint_D \nabla g_\varepsilon(z,\zeta)\cdot\nabla g_e(z,\eta)d\tau_z + \iint_{\tilde{D}} \nabla g_\varepsilon(z,\zeta)\cdot\nabla g_e(z,\eta)d\tau_z = 2\pi g_\varepsilon(\zeta,\eta)\ .$$

Interchanging $\varepsilon$ and $e$ in (23) and subtracting the new identitity, we obtain

$$(24)\qquad 2\pi[g_\varepsilon(\zeta,\eta) - g_e(\zeta,\eta)] = \left(\frac{1}{e} - \frac{1}{\varepsilon}\right)\iint_D \nabla g_\varepsilon(z,\zeta)\cdot\nabla g_e(z,\eta)d\tau_z\ .$$

In particular, passing to the limit $e\to\varepsilon$, we find

$$(25)\qquad \frac{\partial}{\partial\varepsilon}g_\varepsilon(\zeta,\eta) = \frac{1}{\varepsilon^2}\cdot\frac{1}{2\pi}\iint_D \nabla g_\varepsilon(z,\zeta)\nabla g_\varepsilon(z,\eta)d\tau_z\ .$$

We introduce the expression

$$(26)\qquad \Gamma(\zeta,\eta) = \frac{1}{2\pi}\iint_D\left(\nabla_z \log\frac{1}{|z-\zeta|}\cdot\nabla_z\log\frac{1}{|z-\eta|}\right)d\tau_z$$

which is a "geometric" functional of $D$, i.e., can be calculated from elementary functions by a simple process of integration and not by solving any boundary value problem of potential theory. Passing in (25) to the limit $\varepsilon = 1$, we find in view of (5)

$$(27)\qquad \left.\frac{\partial}{\partial\varepsilon}g_\varepsilon(\zeta,\eta)\right|_{\varepsilon=1} = \Gamma(\zeta,\eta)\ .$$

On the other hand, we can calculate this same $\varepsilon$-derivative directly from formulas (15), (16) and (21). Comparing results, we obtain

$$(28)\qquad \Gamma(\zeta,\eta) = -2\pi\sum_{\nu=1}^{\infty}\frac{h_\nu(\xi)\tilde{h}_\nu(\xi)}{(1+\rho_\nu)^2}\quad \text{for}\quad \zeta,\eta\in\tilde{D}$$

$$(28')\qquad \Gamma(\zeta,\eta) = 2\pi\sum_{\nu=1}^{\infty}\frac{\rho_\nu\tilde{h}_\nu(\zeta)\tilde{h}_\nu(\eta)}{(1+\rho_\nu)^2}\quad \text{for}\quad \zeta\in D, \eta\in\tilde{D}$$

$$(28'')\qquad \Gamma(\zeta,\eta) = g(\zeta,\eta) + 2\pi\sum_{\nu=1}^{\infty}\frac{h_\nu(\zeta)h_\nu(\eta)}{\rho_\nu(1+\rho_\nu)^2}\quad \text{for}\quad \zeta,\eta\in D\ .$$

The fact that these particular series in the $h$-functions have relatively elementary sums is of considerable interest. It leads to series developments for the dielectric Green's functions in terms of geometric expressions.

Let us define recursively

$$(29)\qquad \Gamma^{(n)}(z,\zeta) = \frac{1}{2\pi}\iint_D\left(\nabla_\eta\Gamma^{(n-1)}(\eta,z)\cdot\Delta_\eta\log\frac{1}{|\eta-\zeta|}\right)d\tau_\eta,\ \Gamma^{(1)}\equiv\Gamma\ .$$

Using equations (9), (10) and the Fourier formulas (13), (13′), we derive the series developments

$$\Gamma^{(n)}(z, \zeta) = -2\pi \sum_{\nu=1}^{\infty} \frac{\tilde{h}_\nu(z) h_\nu(\zeta)}{(1+\rho_\nu)^{n+1}} \quad \text{for} \quad z \in \tilde{D}, \zeta \in D \tag{30}$$

$$\Gamma^{(n)}(z, \zeta) = 2\pi \sum_{\nu=1}^{\infty} \frac{\rho_\nu \tilde{h}_\nu(z) \tilde{h}_\nu(\zeta)}{(1+\rho_\nu)^{n+1}} \quad \text{for} \quad z \in \tilde{D}, \zeta \in \tilde{D} \tag{31}$$

$$\Gamma^{(n)}(z, \zeta) = g(z, \zeta) + 2\pi \sum_{\nu=1}^{\infty} \frac{h_\nu(z) h_\nu(\zeta)}{\rho_\nu(1+\rho_\nu)^{n+1}} \quad \text{for} \quad z \in D, \zeta \in D . \tag{32}$$

We return now to the formulas (15), (16) and (21) for $g_\varepsilon(z, \zeta)$. We use the series development

$$\frac{\varepsilon - 1}{1 + \varepsilon\rho_\nu} = 2 \sum_{k=0}^{\infty} \left(\frac{\varepsilon - 1}{\varepsilon + 1}\right)^{k+1} \frac{(1-\rho_\nu)^k}{(1+\rho_\nu)^{k+1}} = \frac{2}{1-\rho_\nu} \sum_{k=0}^{\infty} \left(\frac{1}{\lambda_\nu} \frac{1-\varepsilon}{1+\varepsilon}\right)^{k+1} \tag{33}$$

which converges absolutely since $\varepsilon > 0$ and $|\lambda_\nu| > 1$. We insert this series into the above formulas for $g_\varepsilon(z, \zeta)$; interchanging the order of summation, we obtain in each case the representation:

$$g_\varepsilon(z, \zeta) = \log \frac{1}{|z - \zeta|} + \sum_{k=0}^{\infty} \left(\frac{\varepsilon - 1}{\varepsilon + 1}\right)^{k+1} M_k(z, \zeta) . \tag{34}$$

The kernels $M_k(z, \zeta)$ are defined as follows:

$$M_k(z, \zeta) = -4\pi \sum_{\nu=1}^{\infty} \frac{(1-\rho_\nu)^k}{(1+\rho_\nu)^{k+2}} \tilde{h}_\nu(z) h_\nu(\zeta) \quad \text{for} \quad z \in \tilde{D}, \zeta \in \tilde{D} \tag{35}$$

$$M_k(z, \zeta) = -4\pi \sum_{\nu=1}^{\infty} \frac{(1-\rho_\nu)^k \rho_\nu}{(1+\rho_\nu)^{k+2}} \tilde{h}_\nu(z) \tilde{h}_\nu(\zeta) \quad \text{for} \quad z \in \tilde{D}, \zeta \in \tilde{D} \tag{36}$$

$$M_k(z, \zeta) = 2g(z, \zeta) + 4\pi \sum_{\nu=1}^{\infty} \frac{(1-\rho_\nu)^k}{\rho_\nu(1+\rho_\nu)^{k+2}} h_\nu(z) h_\nu(\zeta) \quad \text{for} \quad z \in D, \zeta \in D . \tag{37}$$

By use of the geometric terms (30), (31) and (32), we can express $M_k(z, \zeta)$ in a uniform way, independently of the location of their arguments. We find

$$M_k(z, \zeta) = \sum_{\sigma=1}^{\infty} (-1)^k \binom{k}{\sigma} 2^{k-\sigma+1} \Gamma^{(k-\sigma+1)}(z, \zeta) . \tag{38}$$

Formulas (34) and (38) allow a series development for all dielectric Green's functions in the entire plane in terms of the known iterated Dirichlet integrals $\Gamma^{(n)}(z, \zeta)$. They are closely related to similar developments for the classical Green's function of a multiply-connected domain in terms of geometric expressions [3, 21]. The formulas are convenient for $|\varepsilon - 1|$ small. Observe also that the geometrical terms $M_k(z, \zeta)$ are independent of $\varepsilon$ and may be defined as the coefficients of the Taylor's series for $g_\varepsilon(z, \zeta)$ in terms of $(\varepsilon - 1)/(\varepsilon + 1)$.

**3. Limit values of the dielectric Green's function.** From the series developments for the dielectric Green's function, given in the preceding section, we can determine the limit values of $g_\varepsilon(z, \zeta)$ as $\varepsilon$ converges to zero or to infinity. For this purpose, we have to introduce additional functions of the classes $\sum$ and $\tilde{\sum}$ and to develop them into series of the $h$-functions.

(a) We suppose $\zeta \in \tilde{D}$ and consider the analytic function $\tilde{\varphi}(z, \zeta)$ of $z$ in $\tilde{D}$ which has a simple pole at $z = \zeta$, vanishes at infinity such that

$$(1) \qquad \lim_{z \to \infty} z\tilde{\varphi}(z, \zeta) = 1$$

and which maps $\tilde{D}$ in a one-to-one manner upon the complex plane slit along concentric circular arcs around the origin. These requirements determine $\tilde{\varphi}(z, \zeta)$ in a unique way.

Let now

$$(2) \qquad \tilde{G}(z, \zeta) = \log |\tilde{\varphi}(z, \zeta)| .$$

The function $\tilde{G}(z, \zeta) + \log |z - \zeta|$ is harmonic in $\tilde{D}$, has a single-valued harmonic conjugate there and vanishes as $|z| \to \infty$. Hence, this function lies in the class $\tilde{\sum}$.

We can construct $\tilde{G}(z, \zeta)$ explicitly in terms of the Green's function $\tilde{g}(z, \zeta)$ of $\tilde{D}$. In fact, it is evident that

$$(3) \qquad \tilde{G}(z, \zeta) = \tilde{g}(z, \zeta) - \tilde{g}(z, \infty) - \tilde{g}(\zeta, \infty) + \tilde{\gamma} - \sum_{j,k=1}^{\infty} \alpha_{jk}(\omega_j(z) - \omega_j(\infty))(\omega_k(\zeta) - \omega_k(\infty)) ,$$

with

$$(3') \qquad \tilde{\gamma} = \lim_{z \to \infty} (\tilde{g}(z, \infty) - \log |z|) .$$

The coefficient matrix $\alpha_{jk}$ has to be chosen in such a way as to make the conjugate of $\tilde{G}$ single-valued along each boundary curve $C_l$. Hence. we obtain for it the linear equations

$$(4) \qquad \omega_l(\zeta) - \omega_l(\infty) = \sum_{j,k=1}^{N-1} \alpha_{jk} p_{jl}[\omega_k(\zeta) - \omega_k(\infty)]$$

where the $p_{jl}$ are the elements of the period matrix defined in (2.18″). Hence, we conclude

$$(5) \qquad \sum_{j=1}^{N-1} \alpha_{jk} p_{jl} = \delta_{kl} ,$$

i.e., the $\alpha$-matrix is the inverse of the period matrix of rank $N - 1$.

We can develop $\tilde{G}(z, \zeta) + \log |z - \zeta|$ in terms of the complete orthonormal system $\{\tilde{h}_\nu\}$ in $\tilde{\sum}$. Since $\tilde{G}(z, \zeta)$ takes on each curve $C_l$ a constant boundary value

$$(6) \qquad \tilde{G}(z, \zeta) = c_l(\zeta) \quad \text{for} \quad z \in C_l, \zeta \in \tilde{D} ,$$

we have

$$(7) \qquad \iint_{\tilde{D}} \nabla \tilde{G}(z, \zeta) \cdot \nabla \tilde{h}_\nu(z) d\tau_z = \sum_{l=1}^{N} c_l(\zeta) \int_{C_l} \frac{\partial \tilde{h}_\nu}{\partial n} ds = 0 .$$

Thus, combining (7) with (2.13′) for $\varepsilon = 1$, we obtain

$$(8) \qquad \iint_{\tilde{D}} \nabla [\tilde{G}(z \zeta) + \log |z - \zeta|] \cdot \nabla \tilde{h}_\nu(z) d\tau_z = -\frac{2\pi\rho_\nu}{1 + \rho_\nu} \tilde{h}_\nu(\zeta) .$$

Consequently, we arrive at the following series development for $\tilde{G}(z,\zeta)$:

$$(9) \qquad \tilde{G}(z, \zeta) = \log \frac{1}{|z - \zeta|} - 2\pi \sum_{\nu=1}^{\infty} \frac{\rho_\nu}{1 + \rho_\nu} \tilde{h}_\nu(z) \tilde{h}_\nu(\zeta) .$$

We may now cast (2.16) into the form

$$(10) \qquad g_\varepsilon(z, \zeta) - \tilde{G}(z, \zeta) = 2\pi \sum_{\nu=1}^{\infty} \frac{\varepsilon\rho_\nu}{1 + \varepsilon\rho_\nu} \tilde{h}_\nu(z) \tilde{h}_\nu(\zeta) .$$

We recognize, in particular, that

$$(11) \qquad \lim_{\varepsilon \to 0} g_\varepsilon(z, \zeta) = \tilde{G}(z, \zeta) .$$

Thus, the logarithm of the important canonical map function $\tilde{\varphi}(z, \zeta)$ is closely related to the limit of the dielectric Green's function as $\varepsilon \to 0$.

Let next $\tilde{\psi}(z, \zeta)$ be analytic for $z \in \tilde{D}$ except for a simple pole at $z = \zeta \in \tilde{D}$, vanish at infinity such that

$$(1') \qquad \lim_{z \to \infty} z\tilde{\psi}(z, \zeta) = 1$$

and map $\tilde{D}$ univalently onto the entire plane slit along rectilinear segments which are all directed towards the origin. $\tilde{\psi}(z, \zeta)$ is uniquely determined and might be constructed explicitly in terms of the Neumann's function of $\tilde{D}$.

Let

$$(12) \qquad \tilde{N}(z, \zeta) = \log |\tilde{\psi}(z, \zeta)| .$$

Obviously, the function $\tilde{N}(z, \zeta) + \log |z - \zeta|$ lies in the class $\tilde{\sum}$. Since $\tilde{N}(z, \zeta)$ has, by its definition, vanishing normal derivatives on $C$, we have

(13) $$\iint_{\tilde{D}} \nabla \tilde{N}(z, \zeta) \cdot \nabla \tilde{h}_\nu(z) d\tau_z = 2\pi \tilde{h}_\nu(\zeta) ;$$

therefore, in view of (2.13′) for $\varepsilon = 1$

(14) $$\iint_{\tilde{D}} \nabla[\tilde{N}(z, \zeta) + \log |z - \zeta|] \cdot \nabla \tilde{h}_\nu(z) d\tau_z = \frac{2\pi}{1 + \rho_\nu} \tilde{h}_\nu(\zeta) .$$

Thus, we arrive at the series development

(15) $$\tilde{N}(z, \zeta) = \log \frac{1}{|z - \zeta|} + 2\pi \sum_{\nu=1}^{\infty} \frac{1}{1 + \rho_\nu} \tilde{h}_\nu(z) \tilde{h}_\nu(\zeta) .$$

We can transform (2.16) into

(16) $$\tilde{N}(z, \zeta) - g_\varepsilon(z, \zeta) = \sum_{\nu=1}^{\infty} \frac{\tilde{h}_\nu(z) \tilde{h}_\nu(\zeta)}{1 + \varepsilon\rho_\nu}$$

and read off the limit relation

(17) $$\lim_{\varepsilon \to \infty} g_\varepsilon(z, \zeta) = \tilde{N}(z, \zeta) .$$

The dielectric Green's function $g_\varepsilon(z, \zeta)$ yields thus in $\tilde{D}$ a continuous interpolation between the logarithms of two canonical map functions. The result is the more significant since we shall prove in the next section that each $g_\varepsilon(z, \zeta)$ is analogously related to a univalent function in $\tilde{D}$.

(b) From the fact that the function $\tilde{G}(z, \zeta) + \log |z - \zeta|$ lies in $\tilde{\Sigma}$, i.e., that it has a single-valued conjugate and that it vanishes at infinity, it follows by virtue of (6) that

(18) $$\sum_{l=1}^{N} c_l(\zeta) \int_{C_l} \frac{\partial \omega_j(z)}{\partial n} ds = \int_C \log \frac{1}{|z - \zeta|} \frac{\partial \omega_j}{\partial n} ds$$

and

(18′) $$\sum_{l=1}^{N} c_l(\zeta) \int_{C_l} \frac{\partial \tilde{g}(z, \infty)}{\partial n} ds = \int_C \log \frac{1}{|z - \zeta|} \frac{\partial \tilde{g}(z, \infty)}{\partial n} ds .$$

We define now for fixed $\zeta \in \tilde{D}$ the harmonic function $c(z, \zeta)$ of $z$ in $D$ by putting

(19) $$c(z, \zeta) = c_l(\zeta) \quad \text{for} \quad z \in D_l .$$

By (18), (18′) and the definition (2.18) of the class $\Sigma$, the function $-\log |z - \zeta| - c(z, \zeta)$ lies in this linear space. We may develop it, therefore, into a series of the $h_\nu(z)$. By use of (2.10) and (2.13′), we obtain

$$\log \frac{1}{|z-\zeta|} = c(z, \zeta) - 2\pi \sum_{\nu=1}^{\infty} \frac{h_\nu(z)\tilde{h}_\nu(\zeta)}{1+\rho_\nu} \qquad z \in D, \zeta \in \tilde{D}. \tag{20}$$

We may combine (20) with (2.22) and find

$$g_\varepsilon(z, \zeta) = c(z, \zeta) - 2\pi\varepsilon \sum_{\nu=1}^{\infty} \frac{1}{1+\varepsilon\rho_\nu} h_\nu(z)\tilde{h}_\nu(\zeta). \tag{21}$$

This leads to the limit relation

$$\lim_{\varepsilon\to 0} g_\varepsilon(z, \zeta) = c(z, \zeta) \quad \text{for} \quad z \in D, \zeta \in \tilde{D}. \tag{22}$$

The limit of $g_\varepsilon(z, \zeta)$ as $\varepsilon \to \infty$ does not seem to admit a simple geometric interpretation.

(c) Consider next the case $\zeta \in D$, say $\zeta \in D_l$. We define now the regular analytic functions $\tilde{\varphi}_l(z)$ which map $\tilde{D}$ univalently into a full circle around the origin which is slit along concentric circular arcs, such that $z = \infty$ goes into the center and that

$$\lim_{z\to\infty} z\tilde{\varphi}_l(z) = 1. \tag{1''}$$

The function $\tilde{\varphi}_l(z)$ is uniquely determined by the additional requirement that the special boundary curve $C_l$ shall correspond to the outer circumference.

Since the function

$$\tilde{G}_l(z) = \log|\tilde{\varphi}_l(z)| \tag{23}$$

is harmonic in $\tilde{D}$ except for a simple logarithmic pole at infinity and since

$$\tilde{G}_l(z) = c_{lj} \quad \text{for} \quad z \in C_j, \tag{24}$$

it is evident that $\tilde{G}_l(z)$ may again be expressed explicitly in terms of the Green's function $\tilde{g}(z, \zeta)$ of $\tilde{D}$ [5].

Since we assumed $\zeta \in D_l$, the function $\tilde{G}_l(z) + \log|z-\zeta|$ lies in the class $\tilde{\Sigma}$. We can develop it into a Fourier series of the system $\{\tilde{h}_\nu\}$. The same calculations as before lead to

$$\tilde{G}_l(z) = \log\frac{1}{|z-\zeta|} + 2\pi \sum_{\nu=1}^{\infty} \frac{\tilde{h}_\nu(z)h_\nu(\zeta)}{1+\rho_\nu}, \ z \in \tilde{D}, \zeta \in D_l. \tag{25}$$

From (2.15) we obtain

$$g_\varepsilon(z, \zeta) - \tilde{G}_l(z) = -2\pi\varepsilon \sum_{\nu=1}^{\infty} \frac{1}{1+\varepsilon\rho_\nu} \tilde{h}_\nu(z)h_\nu(\zeta); \tag{26}$$

hence

(27) $$\lim_{\varepsilon \to 0} g_\varepsilon(z, \zeta) = \tilde{G}_l(z) \quad \text{for} \quad z \in \tilde{D}, \zeta \in D_l\ .$$

We obtain again interesting canonical mappings from the dielectric Green's function by passing to the limit $\varepsilon = 0$.

(d) The expression $\tilde{G}_l(z) + \log|z - \zeta|$ satisfies the linear relations (2.18) if $\zeta \in D_l$. It has on $C$ the same boundary values as the function $g(z, \zeta) + \log|z - \zeta| + c_l(z)$ which is harmonic in $D$, with

(28) $$c_l(z) = c_{lj} \quad \text{for} \quad z \in D_j\ .$$

Thus, the new combination will belong to the class $\Sigma$ and can, therefore, be developed into a Fourier series in the $\{h_\nu\}$-system. An easy calculation leads to

(29) $$g(z, \zeta) = \log \frac{1}{|z - \zeta|} - c_l(z) - 2\pi \sum_{\nu=1}^{\infty} \frac{h_\nu(z) h_\nu(\zeta)}{\rho_\nu(1 + \rho_\nu)},\ z \in D, \zeta \in D_l\ .$$

From (29) and (2.21) follows

(30) $$g_\varepsilon(z, \zeta) - \varepsilon g(z, \zeta) = c_l(z) + 2\pi\varepsilon \sum_{\nu=1}^{\infty} \frac{1}{\rho_\nu(1 + \varepsilon\rho_\nu)} h_\nu(z) h_\nu(\zeta)\ .$$

Thus, we find the limit formulas, valid for $z \in D, \zeta \in D_l$:

(31) $$\lim_{\varepsilon \to 0} g_\varepsilon(z, \zeta) = c_l(z)\ , \quad \lim_{\varepsilon \to \infty} \frac{1}{\varepsilon} g_\varepsilon(z, \zeta) = g(z, \zeta)\ .$$

**4. Dielectric Green's functions and conformal mapping.** In this section, we shall show that the dielectric Green's function $g_\varepsilon(z, \zeta)$ leads to a univalent analytic function in $\tilde{D}$ and to a set of univalent analytic functions in $D$. Let us suppose, for the sake of definiteness, that the source point $\zeta$ lies in $\tilde{D}$. Let $p_\varepsilon(z, \zeta)$ be the analytic completion of $g_\varepsilon(z, \zeta)$ for $z$ in $\tilde{D}$; that is, $p_\varepsilon(z, \zeta)$ is analytic for $z \in D$ and we have

(1) $$g_\varepsilon(z, \zeta) = \Re\{p_\varepsilon(z, \zeta)\}\ .$$

$p_\varepsilon(z, \zeta)$ is regular analytic except for the two logarithmic poles at $\zeta$ and at $\infty$. The function has no periods with respect to the boundary curves $C_j$. Hence

(2) $$\tilde{f}_\varepsilon(z, \zeta) = \exp[-p_\varepsilon(z, \zeta)]\ , \qquad z \in \tilde{D}, \zeta \in \tilde{D}$$

is a single-valued analytic function of $z \in \tilde{D}$ and regular in this domain except for the simple pole at infinity. Since the analytic completion of a harmonic function is only determined up to an additive imaginary constant, we may choose $\tilde{f}_\varepsilon$ in such a way that

(2′) $$\tilde{f}'_\varepsilon(\infty, \zeta) = 1\ , \quad \tilde{f}_\varepsilon(\zeta, \zeta) = 0\ .$$

We may similarly complete $g_\varepsilon(z, \zeta)$ to analytic functions of $z$ in $D$. In order to determine the additive constants for the disjoint domains $D_j$ we proceed as follows. By condition (c) of § 2 on $g_\varepsilon(z, \zeta)$ and because of the Cauchy-Riemann equations, we have, whatever the analytic completion $p_\varepsilon(z, \zeta)$ of $g_\varepsilon(z, \zeta)$ in $D$:

$$(3) \qquad \Im\{p_\varepsilon(z, \zeta)\} = \varepsilon\Im\{\tilde{p}_\varepsilon(z, \zeta)\} + k_j \quad \text{for} \quad z \in C_j .$$

Here $\tilde{p}_\varepsilon$ and $p_\varepsilon$ shall denote the limits of $p_\varepsilon$ from $\tilde{D}$ and $D$, respectively; we shall use this more specific notation whenever discussing boundary relations. We dispose now of the additive constants in the domains $D_j$ by requiring $k_j = 0$. This convention fixes $p_\varepsilon(z, \zeta)$ in $D$ in a unique way.

In analogy to (2), we define

$$(4) \qquad f_\varepsilon(z, \zeta) = \exp\left[-\frac{1}{\varepsilon}p_\varepsilon(z, \zeta)\right] \quad \text{for} \quad z \in D, \zeta \in \tilde{D} .$$

We shall prove the

THEOREM. *The function $\tilde{f}_\varepsilon(z, \zeta)$ is univalent in $\tilde{D}$ and the set of functions $f_\varepsilon(z, \zeta)$ is univalent in $D$ in the sense that*

$$(5) \qquad f_\varepsilon(z_1, \zeta) = f_\varepsilon(z_2, \zeta) \text{ and } z_1, z_2 \in D \text{ implies } z_1 = z_2 .$$

In order to prove this theorem, we start with the

LEMMA. *The dielectric Green's function has no critical points. That is, the equation $p'_\varepsilon(z, \zeta) = 0$ is only satisfied at $z = \infty$ and this point is a pole of the Green's function. The dash denotes differentiation of $p_\varepsilon(z, \zeta)$ with respect to its analytic argument $z$.*

*Proof.* We denote again, more precisely, the analytic completion of $g_\varepsilon(z, \zeta)$ by $\tilde{p}_\varepsilon$ or by $p_\varepsilon$ according to the location of $z$ in $\tilde{D}$ or $D$, respectively. We combine the boundary conditions (c) and (d) of § 2 on the dielectric Green's function $g_\varepsilon(z, \zeta)$ into the one complex equation

$$(6) \qquad p'_\varepsilon(z, \zeta)z' = \frac{1+\varepsilon}{2}\tilde{p}'_\varepsilon(z, \zeta)z' + \frac{1-\varepsilon}{2}\overline{\tilde{p}'_\varepsilon(z, \zeta)z'} .$$

Since we assume throughout this paper $\varepsilon > 0$, equation (6) yields

$$(7) \qquad \Re e\{p'_\varepsilon(z, \zeta)/\tilde{p}'_\varepsilon(z, \zeta)\} > 0 \quad \text{for } z \in C .$$

This inequality implies, in particular:

$$(8) \qquad \oint_C d \arg p'_\varepsilon(z, \zeta) = \oint_C d \arg \tilde{p}'_\varepsilon(z, \zeta) .$$

The statement is evident if $p'_\varepsilon$ and $\tilde{p}'_\varepsilon$ are non-zero on $C$ but it can be upheld in the usual way even in the case that these two functions have common zeros on $C$

Let $Z, P$ and $\tilde{Z}, \tilde{P}$ denote the number of zeros and poles of $p'_\varepsilon$ and $\tilde{p}'_\varepsilon$ respectively, in their domains of definition. By the argument principle, we have

$$(9)\qquad \oint_C d\arg p'_\varepsilon(z,\zeta) = Z - P\,, \quad \oint_C d\arg \tilde{p}'_\varepsilon(z,\zeta) = \tilde{P} - \tilde{Z}$$

if $z$ runs through $C$ in the positive sense with respect to $D$. Combining (8) and (9), we obtain

$$(10)\qquad Z + \tilde{Z} = P + \tilde{P}\,.$$

But all poles of $p'_\varepsilon$ and $\tilde{p}'_\varepsilon$ are known; clearly $P = 0, \tilde{P} = 1$ and $Z \geq 1$. Hence, we conclude from (10):

$$(11)\qquad Z = 0\,, \quad \tilde{Z} = 1\,.$$

This proves our lemma.

In order to prove the theorem, we consider the lines defined by

$$(12)\qquad \Im\{\tilde{p}_\varepsilon(z,\zeta)\} = \alpha \text{ for } z \in \tilde{D}\,, \quad \Im\left\{\frac{1}{\varepsilon}p_\varepsilon(z,\zeta)\right\} = \alpha \text{ for } z \in D\,.$$

Each such line starts from the logarithmic pole $\zeta$ and runs to $\infty$. By virtue of our convention on the analytic completion of $g_\varepsilon(z,\zeta)$ these lines are continuous in the entire plane and, except on $C$, they are even analytic. Because of our lemma, there is no intersection between different lines except at $\zeta$ and $\infty$. The lines have the physical interpretation as lines of force for the corresponding electrostatic problem and the lemma asserts that there are no points of equilibrium in the field. The lines form for $0 \leq \alpha < 2\pi$ a non-intersecting system which covers the entire complex plane. Along each line, $g_\varepsilon(z,\zeta)$ decreases monotonically when we pass from $\zeta$ to $\infty$. These facts guarantee obviously that the analytic functions $\tilde{f}_\varepsilon(z,\zeta)$ and $f_\varepsilon(z,\zeta)$ have the above stated univalency properties. Thus, the theorem is proved.

Let us assume without loss of generality that $\zeta = 0$. Using the limit theorems of § 3, we can assert:

$$(13)\qquad \tilde{f}_0(z,0) = \tilde{\varphi}(z,0)^{-1}\,, \quad \tilde{f}_1(z,0) = z\,, \quad \tilde{f}_\infty(z,0) = \tilde{\psi}(z,0)^{-1}\,.$$

We have thus found a one-parameter family of univalent functions which connects continuously the circular slit mapping through the identity mapping with the radial slit mapping.

In order to illustrate the significance of this result, we calculate from (2.16) that

$$\log|\tilde{f}'_\varepsilon(\zeta,\zeta)| = 2\pi(\varepsilon-1)\sum_{\nu=1}^{\infty}\frac{\rho_\nu}{(1+\rho_\nu)(1+\varepsilon\rho_\nu)}\tilde{h}_\nu(\zeta)^2 . \tag{14}$$

Since all $\rho_\nu > 0$, this is a monotonically increasing function of $\varepsilon$ in the interval $[0, \infty)$; it is negative for $0 \leq \varepsilon < 1$ and positive for $1 < \varepsilon$. In particular:

$$|\tilde{f}'_0(\zeta,\zeta)| < 1 \quad |\tilde{f}'_\infty(\zeta,\zeta)| > 1 . \tag{15}$$

We define the family $\mathscr{F}_\zeta$ of all functions $\tilde{f}(z)$ which are analytic and univalent in $\tilde{D}$ and normalized by the requirements

$$\tilde{f}'(\infty) = 1 \qquad \tilde{f}(\zeta) = 0 . \tag{16}$$

Through the mapping $w = \tilde{f}(z)$ we obtain the new domain $\tilde{D}_w$; applying the inequalities (15) in this domain and returning to the original domain $\tilde{D}$, we obtain the inequality

$$|f'_0(\zeta,\zeta)| \leq |\tilde{f}'(\zeta)| \leq |\tilde{f}'_\infty(\zeta,\zeta)| \tag{17}$$

valid for each $f \in \mathscr{F}_\zeta$.

Inequality (16) asserts an extremum property of the canonical slit functions $\tilde{f}_0$ and $\tilde{f}_\infty$ which is well-known [13, 15]. It is, however, not obvious that all real values between the extrema are also possible values for $|f'(\zeta)|$ in $\mathscr{F}_\zeta$. We have now explicity constructed a one-parameter family in $\mathscr{F}_\zeta$ which interpolates between the two extremum values.

There are various other possibilities to obtain from the dielectric Green's function one-parameter families of univalent functions. Consider, for example, the analytic functions

$$A_\varepsilon(z,\zeta) = \frac{\partial}{\partial\xi}p_\varepsilon(z,\zeta) , \quad B_\varepsilon(z,\zeta) = \frac{1}{i}\frac{\partial}{\partial\eta'}p_\varepsilon(z,\zeta) \tag{18}$$

with $\zeta = \xi + i\eta$. Both functions are single-valued in $\tilde{D}$ and in $D$; they have for $z = \zeta$ simple poles with residue 1 and are else regular in $D$ and in $D$. We obtain from the identity (6) by differentiation

$$A'_\varepsilon(z,\zeta)z' = \frac{1+\varepsilon}{2}\tilde{A}'_\varepsilon(z,\zeta)z' + \frac{1-\varepsilon}{2}\overline{(\tilde{A}'_\varepsilon(z,\zeta)z')} \tag{19}$$

$$B'_\varepsilon(z,\zeta)z' = \frac{1+\varepsilon}{2}\tilde{B}'_\varepsilon(z,\zeta)z' - \frac{1-\varepsilon}{2}\overline{(\tilde{B}'_\varepsilon(z,\zeta)z')} . \tag{19'}$$

Let $a$ be an arbitrary point on $C$; integrating (19) along $C$ from $a$ to $z \in C$, we find

$$A_\varepsilon(z,\zeta) - A_\varepsilon(a,\zeta) = \frac{1+\varepsilon}{2}[\tilde{A}_\varepsilon(z,\zeta) - \tilde{A}_\varepsilon(a,\zeta)] + \frac{1-\varepsilon}{2}\overline{[\tilde{A}_\varepsilon(z,\zeta) - \tilde{A}_\varepsilon(a,\zeta)]} . \tag{20}$$

Hence, we have

$$\Re\left\{\frac{A_\varepsilon(z,\zeta)-A_\varepsilon(a,\zeta)}{\tilde{A}_\varepsilon(z,\zeta)-\tilde{A}_\varepsilon(a,\zeta)}\right\}>0 \quad \text{for} \quad z\in C\,. \tag{21}$$

Reasoning as before we can conclude by means of the argument principle that $A_\varepsilon(z,\zeta)$ takes the value $A_\varepsilon(a,\zeta)$ precisely once in $D+C$ and that $\tilde{A}_\varepsilon(z,\zeta)$, likewise, takes every boundary value precisely once. Thus, $A_\varepsilon(z,\zeta)$ and $\tilde{A}_\varepsilon(z,\zeta)$ are univalent in their respective domains of definition. The same reasoning applies to $B_\varepsilon(z,\zeta)$ and $\tilde{B}_\varepsilon(z,\zeta)$.

It is known, and easily verified, that

$$\tilde{A}_0(z,\zeta)=\frac{\partial}{\partial\xi}\log\tilde{\varphi}(z,\zeta)\,,\quad \tilde{B}_0(z,\zeta)=\frac{1}{i}\frac{\partial}{\partial\eta}\log\tilde{\varphi}(z,\zeta) \tag{22}$$

are univalent functions in $\tilde{D}$ with a simple pole at $z=\zeta$ and that they map $\tilde{D}$ onto the entire complex plane, slit along rectilinear segments parallel to the imaginary and the real axis, respectively [16]. Similarly, the analytic functions

$$\tilde{A}_\infty(z,\zeta)=\frac{\partial}{\partial\xi}\log\tilde{\psi}(z,\zeta)\,,\quad \tilde{B}_\infty(z,\zeta)=\frac{1}{i}\frac{\partial}{\partial\eta}\log\tilde{\psi}(z,\zeta) \tag{23}$$

are univalent in $\tilde{D}$ with the same singularity and map the domain onto the entire complex plane, slit along segments parallel to the real and the imaginary axis, respectively. Hence, by the uniqueness theorems on the canonical mappings of a domain, we must have

$$\tilde{A}_\infty(z,\zeta)=\tilde{B}_0(z,\zeta)+\kappa(\zeta)\,;\ \tilde{B}_\infty(z,\zeta)=\tilde{A}_0(z,\zeta)+\lambda(\zeta)\,. \tag{24}$$

Finally, clearly

$$\tilde{A}_1(z,\zeta)=\tilde{B}_1(z,\zeta)=\frac{1}{z-\zeta}\,. \tag{25}$$

Hence, $\tilde{A}_\varepsilon(z,\zeta)$ and $\tilde{B}_\varepsilon(z,\zeta)$ interpolate between the two parallel slit mappings through the simple rational mapping (25).

Using the series development (2.16) for $g_\varepsilon(z,\zeta)$, $\zeta\in\tilde{D}$, we may prove the well-known extremum properties of the canonical slit mappings in the same way, as we did above for the circular and the radial slit mapping.

We do not enter into a more detailed discussion of these families of univalent functions. We want to remark, however, that the dielectric Green's function is not, like the ordinary Green's function, a conformal invariant. By auxiliary mappings of $\tilde{D}$ into a domain $\tilde{D}_w$, one may obtain very different one-parameter families of univalent functions which interpolate between the canonical slit mappings.

**5. Dielectric Green's functions and norms in function spaces.** With each dielectric Green's function $g_\varepsilon(z, \zeta)$ we can connect a positive-definite quadratic form which may be interpreted as a norm in the linear function spaces $\Sigma$ and $\tilde{\Sigma}$, defined in § 2. This norm has remarkable properties for function pairs $h \in \Sigma$ and $\tilde{h} \in \tilde{\Sigma}$ which have on $C$ equal boundary values or equal normal derivatives. Useful inequalities and identities can be established which facilitate the solution of the boundary value problem in potential theory by utilizing auxiliary solutions in complementary domains. One can characterize the Fredholm eigen values $\lambda_\nu$ as solutions of certain extremum problems involving these quadratic forms. This characterization, in turn, will lead later to elegant variational formulas for the $\lambda_\nu$ under infinitesimal deformation of the curve system $C$.

Let $h$ and $\tilde{h}$ be two arbitrary functions of the classes $\Sigma$ and $\tilde{\Sigma}$, respectively. We have the Fourier developments

$$(1) \qquad h(z) = \sum_{\nu=1}^{\infty} x_\nu h_\nu(z) \ , \quad \tilde{h}(z) = \sum_{\nu=1}^{\infty} \tilde{x}_\nu \tilde{h}_\nu(z)$$

in terms of the complete orthonomal sets $\{\rho_\nu^{-1/2} h_\nu(z)\}$ and $\{\tilde{h}_\nu(z)\}$ of these linear spaces. The Fourier coefficients are given by

$$(2) \qquad x_\nu = \frac{1}{\rho_\nu} D(h, h_\nu) \ , \quad \tilde{x}_\nu = \tilde{D}(\tilde{h}, \tilde{h}_\nu)$$

where $D$ and $\tilde{D}$ denote the Dirichlet integral in $\Sigma$ and $\tilde{\Sigma}$:

$$(3) \qquad D(h, H) = \iint_D \nabla h \cdot \nabla H d\tau \ , \quad \tilde{D}(\tilde{h}, \tilde{H}) = \iint_{\tilde{D}} \nabla \tilde{h} \cdot \nabla \tilde{H} d\tau \ .$$

Let us consider now the particular case that

$$(4) \qquad h(z) = \tilde{h}(z) \quad \text{on} \quad C \ .$$

By Green's identity and (1.4′), we have obviously

$$(5) \qquad D(h, h_\nu) = -\int_C h \frac{\partial h_\nu}{\partial n} ds = -\int_C \tilde{h} \frac{\partial \tilde{h}_\nu}{\partial n} ds = -\tilde{D}(\tilde{h}, \tilde{h}_\nu)$$

which gives

$$(6) \qquad x_\nu = -\frac{1}{\rho_\nu} \tilde{x}_\nu \ .$$

We proceed analogously for two function $h \in \Sigma$ and $\tilde{h} \in \tilde{\Sigma}$ which satisfy on $C$ the relation

$$(7) \qquad \frac{\partial h}{\partial n} = \frac{\partial \tilde{h}}{\partial n} \ .$$

Now, Green's identity and (1.4) yield

$$(8)\qquad D(h, h_\nu) = -\int_C h_\nu \frac{\partial h}{\partial n} ds = \rho_\nu \int_C \tilde{h}_\nu \frac{\partial \tilde{h}}{\partial n} ds = \rho_\nu \tilde{D}(\tilde{h}, \tilde{h}_\nu)$$

and, consequently

$$(9)\qquad x_\nu = \tilde{x}_\nu .$$

Thus, both boundary relations (4) and (7) reflect themselves in a very simple manner in the relations (6) and (9) between the Fourier coefficients.

We define next the bilinear form

$$(10)\qquad \pi_\varepsilon(h, H) = \frac{1}{2\pi}\int_C\int_C g_\varepsilon(z, \zeta)\frac{\partial h(z)}{\partial n}\frac{\partial H(\zeta)}{\partial n} ds_z ds_\zeta$$

for any two elements of $\Sigma$ and in precisely the same manner we define the bilinear from $\tilde{\pi}_\varepsilon(\tilde{h}, \tilde{H})$ for any two elements in $\tilde{\Sigma}$.

By use of the Fourier type formulas (2.13) and (2.13′) we may express the bilinear forms in terms of the Fourier coefficients of the functions involved. Let us denote the Fourier coefficients of $h, \tilde{h}$ by $x_\nu, \tilde{x}_\nu$ and of $H, \tilde{H}$ by $y_\nu, \tilde{y}_\nu$; then a straightforward calculation shows that

$$(11)\qquad \pi_\varepsilon(h, H) = \sum_{\nu=1}^{\infty} \frac{\varepsilon\rho_\nu}{1+\varepsilon\rho_\nu} x_\nu y_\nu, \; \tilde{\pi}_\varepsilon(\tilde{h}, \tilde{H}) = \sum_{\nu=1}^{\infty} \frac{\varepsilon\rho_\nu}{1+\varepsilon\rho_\nu} \tilde{x}_\nu \tilde{y}_\nu .$$

We verify, first, from (11) that the quadratic forms $\pi_\varepsilon(h, h)$ and $\tilde{\pi}_\varepsilon(\tilde{h}, \tilde{h})$ are positive-definite. This fact allows us to interpret them, indeed, as norms in their corresponding function spaces.

On the other hand, we have because of the normalizations (2.8) and (2.17)

$$(12)\qquad D(h, H) = \sum_{\nu=1}^{\infty} \rho_\nu x_\nu y_\nu, \quad \tilde{D}(\tilde{h}, \tilde{H}) = \sum_{\nu=1}^{\infty} x_\nu y_\nu .$$

We define further the bilinear forms

$$(13)\qquad \Gamma_\varepsilon(h, H) = D(h, H) - \frac{1}{\varepsilon}\pi_\varepsilon(h, H), \quad \tilde{\Gamma}_\varepsilon(\tilde{h}, \tilde{H}) = \tilde{D}(\tilde{h}, \tilde{H}) - \tilde{\pi}_\varepsilon(\tilde{h}, \tilde{H})$$

and obtain for them the explicit representations :

$$(14)\qquad \Gamma_\varepsilon(h, H) = \sum_{\nu=1}^{\infty} \frac{\varepsilon\rho_\nu^2}{1+\varepsilon\rho_\nu} x_\nu y_\nu , \quad \tilde{\Gamma}_\varepsilon(\tilde{h}, \tilde{H}) = \sum_{\nu=1}^{\infty} \frac{1}{1+\varepsilon\rho_\nu} \tilde{x}_\nu \tilde{y}_\nu .$$

These formulas show that $\Gamma_\varepsilon$ and $\tilde{\Gamma}_\varepsilon$, too, are positive-definite and lead to norms in $\Sigma$ and $\tilde{\Sigma}$. We have the estimates :

(15) $$0 \leq \frac{1}{\varepsilon}\pi_{\varepsilon}(h, H) \leq D(h, h)\,; \quad 0 \leq \tilde{\pi}_{\varepsilon}(\tilde{h}, \tilde{h}) \leq \tilde{D}(\tilde{h}, \tilde{h})\,.$$

By the very definition of $\pi_\varepsilon$ and $\tilde{\pi}_\varepsilon$, we have

THEOREM I. *If*

(16) $$\frac{\partial h}{\partial n} = \frac{\partial \tilde{h}}{\partial n} \quad \textit{and} \quad \frac{\partial H}{\partial n} = \frac{\partial \tilde{H}}{\partial n} \quad \text{on} \quad C$$

*we have*

(17) $$\pi_\varepsilon(h, H) = \tilde{\pi}_\varepsilon(\tilde{h}, \tilde{H})\,.$$

From (4), (6) and (14), we derive :

THEOREM II. *If*

(18) $$h = \tilde{h} \text{ and } H = \tilde{H} \text{ on } C\,,$$

*we have*

(19) $$\Gamma_\varepsilon(h, H) = \varepsilon\tilde{\Gamma}_\varepsilon(\tilde{h}, \tilde{H})\,.$$

Finally, we verify from the explicit representations for the bilinear forms

THEOREM III. *If*

(20) $$h = \tilde{h} \text{ and } \frac{\partial H}{\partial n} = \frac{\partial \tilde{H}}{\partial n} \text{ on } C\,,$$

*we have*

(21) $$D(h, H) = -\tilde{D}(\tilde{h}, \tilde{H})$$

*and*

(22) $$\pi_\varepsilon(h, H) = -\varepsilon\tilde{\Gamma}_\varepsilon(\tilde{h}, \tilde{H}), \quad \Gamma_\varepsilon(h, H) = -\tilde{\pi}_\varepsilon(\tilde{h}, \tilde{H})\,.$$

Theorems I–III show a very symmetric interrelation between the various bilinear forms for elements with matching boundary data on $C$.

The significance of the preceding theorems lies in the fact that one has often to solve a boundary value problem, say in $\tilde{D}$, which is much easier to solve in the complementary domain $D$. In this case, the above theorems provide valuable information. Let us illustrate the method by the following applications.

(a) Given a function $h \in \Sigma$, to determine the function $\tilde{h} \in \tilde{\Sigma}$ which has on $C$ the same boundary values as $h$. In particular, we ask for the Dirichlet integral $\tilde{D}(\tilde{h}, \tilde{h})$.

This problem arises, for example, in two-dimensional electrostatics in connection with the question of polarization of a set of conductors in a homogeneous field [19, 22].

We derive inequalities for the Dirichlet integral in question by applying Theorems I–III. We start from the fact that $\pi_\varepsilon$ and $\Gamma_\varepsilon$ have definite quadratic forms and that they satisfy, therefore, the Schwarz inequalities

$$(23)\quad \pi_\varepsilon(h, H)^2 \leq \pi_\varepsilon(h, h) \cdot \pi_\varepsilon(H, H)\,; \quad \Gamma_\varepsilon(h, H)^2 \leq \Gamma_\varepsilon(h, h) \cdot \Gamma_\varepsilon(H, H)\,.$$

We select a pair of test functions $H \in \Sigma$ and $\tilde{H} \in \tilde{\Sigma}$ which have equal normal derivatives on $C$ and obtain from Theorem III and from (23)

$$(24)\qquad \Gamma_\varepsilon(h, H)^2 \leq \tilde{\pi}_\varepsilon(\tilde{h}, \tilde{h}) \cdot \tilde{\pi}_\varepsilon(\tilde{H}, \tilde{H})\,.$$

Using the definition (13) of $\tilde{\Gamma}_\varepsilon$ and Theorems I, II, we can transform (24) into

$$(25)\qquad \Gamma_\varepsilon(h, H)^2 \leq [\tilde{D}(\tilde{h}, \tilde{h}) - \frac{1}{\varepsilon}\Gamma_\varepsilon(h, h)]\pi_\varepsilon(H, H)\,.$$

This inequality contains the sought Dirichlet integral $\tilde{D}(\tilde{h}, \tilde{h})$ and else only the known function of $h \in \Sigma$ and the arbitrary test function $H \in \Sigma$. Thus:

$$(26)\qquad \tilde{D}(\tilde{h}, \tilde{h}) \geq \frac{\Gamma_\varepsilon(h, H)^2}{\pi_\varepsilon(H, H)} + \frac{1}{\varepsilon}\Gamma_\varepsilon(h, h)\,.$$

It is easily seen from our derivation that the inequality (26) is sharp if $H$ is chosen as that function in $\Sigma$ which has on $C$ the same normal derivative as $\tilde{h}$; in fact, in this case, the Schwarz inequality leading to (24) becomes an equality. Thus, we can express (26) as follows:

$$(26')\qquad \tilde{D}(\tilde{h}, \tilde{h}) = \max\frac{\Gamma_\varepsilon(h, H)^2}{\pi_\varepsilon(H, H)} + \frac{1}{\varepsilon}\Gamma_\varepsilon(h, h) \text{ for all } H \in \Sigma\,.$$

This representation permits us to determine the desired Dirichlet integral by a Ritz procedure in $\Sigma$.

It is sometimes more convenient to renounce a precise equation in order to obtain a simple and applicable estimate. We may select, for this purpose, the test function $H(z)$ as equal to the given function $h(z)$; in this case, we have by (13) and (26)

$$(27)\qquad \tilde{D}(\tilde{h}, \tilde{h}) \geq \frac{\Gamma_\varepsilon(h, h)}{\pi_\varepsilon(h, h)} D(h, h)\,.$$

This inequality holds for all pairs of functions $h \in \Sigma, \tilde{h} \in \tilde{\Sigma}$ which have equal boundary values at the same points of $C$.

In order to understand better the important inequality (27), we express it in terms of the corresponding Fourier coefficients. If we denote again by $x_\nu$ the coefficients of $h(z)$, we have by (6) the values $-\rho_\nu x_\nu = \tilde{x}_\nu$ for the Fourier coefficients of $\tilde{h}(z)$. Hence, using the explicit representations (11), (12) and (14) for the quadratic forms, we may write (27) as follows:

$$(27') \qquad \sum_{\nu=1}^{\infty} \rho_\nu^2 x_\nu^2 \cdot \sum_{\nu=1}^{\infty} \frac{\varepsilon\rho_\nu}{1+\varepsilon\rho_\nu} x_\nu^2 \geq \sum_{\nu=1}^{\infty} \rho_\nu x_\nu^2 \cdot \sum_{\nu=1} \frac{\varepsilon\rho_\nu^2}{1+\varepsilon\rho_\nu} x_\nu^2 .$$

We rearrange (27′) into the from

$$(27'') \qquad \sum_{\nu,\mu=1}^{\infty} \frac{\varepsilon^2\rho_\nu\rho_\mu(\rho_\nu-\rho_\mu)^2}{(1+\varepsilon\rho_\nu)(1+\varepsilon\rho_\mu)} x_\nu^2 x_\mu^2 \geq 0 .$$

Now the inequality has become evident; but, what is more important, we recognize that equality in (27″) and, hence in (27), holds if and only if all $x_\nu$ vanish except for those which belong to a fixed eigen value $\lambda_\mu$. Thus, equality in (27) holds for

$$(28) \qquad h(z) = h_\nu(z) \text{ and } \tilde{h}(z) = -\rho_\nu\tilde{h}_\nu(z) , \qquad \nu = 1, 2, \cdots,$$

and only for these functions.

It is interesting that the inequality (27) becomes precise infinitely often, namely for all functions of the sets $\{h_\nu\}$, $\{\tilde{h}_\nu\}$, which are complete in $\Sigma$ and $\tilde{\Sigma}$. On the other hand, this fact leads to a new characterization of the Fredholm eigen functions

(b) We deal next with the analogous question: given a function $h \in \Sigma$, to determine the function $\tilde{h} \in \tilde{\Sigma}$ which has at corresponding points of $C$ the same normal derivative as $h$. In particular, to determine the Dirichlet integral of $\tilde{h}$.

This problem occurs in the theory of a steady incompressible and irrotational fluid flow in the plane around the set of obstacles $C$. The sought Dirichlet integral, in this case, is the virtual mass of the curve system $C$ [19, 22].

We select now a pair of test functions $H \in \Sigma$, $\tilde{H} \in \tilde{\Sigma}$ which have equal boundary values on $C$. Starting again with the Schwarz inequality (23) and Theorem III, we have

$$(29) \qquad \pi_\varepsilon(h, H)^2 \leq \varepsilon^2\tilde{\Gamma}_\varepsilon(\tilde{h}, \tilde{h}) \cdot \tilde{\Gamma}_\varepsilon(\tilde{H}, \tilde{H}) .$$

We apply equation (13), make use of Theorems I and II and find

$$(30) \qquad \pi_\varepsilon(h, H)^2 \leq \varepsilon\Gamma_\varepsilon(H, H)[\tilde{D}(\tilde{h}, \tilde{h}) - \pi_\varepsilon(h, h)] .$$

Thus finally

$$\tilde{D}(\tilde{h}, \tilde{h}) \geq \frac{\pi_\varepsilon(h, H)^2}{\varepsilon \Gamma_\varepsilon(H, H)} + \pi_\varepsilon(h, h) . \tag{31}$$

We obtained thus again a lower bound for the Dirichlet integral in terms of the given function $h$ and the arbitrary test function $H$. If $H$ has on $C$ the same boundary values as $\tilde{h}$, the inequality (31) becomes an equality. This fact allows us again to approximate arbitrarily the Dirichlet integral from below by a Ritz sequence of test functions.

When we choose, on the other hand, $H(z) = h(z)$, we obtain

$$\tilde{D}(\tilde{h}, \tilde{h}) \geq \frac{\pi_\varepsilon(h, h)}{\Gamma_\varepsilon(h, h)} D(h, h) . \tag{32}$$

This inequality holds for every pair of functions $h \in \Sigma$, $\tilde{h} \in \tilde{\Sigma}$ with equal normal derivatives on $C$.

This inequality can be verified by means of the explicit Fourier representations (11), (12) and (14) as we did in the case of the inequality (27). We can further show as before that equality in (32) can hold if and only if

$$h(z) = h_\nu(z) , \quad \tilde{h}(z) = \tilde{h}_\nu(z) , \qquad \nu = 1, 2, \cdots . \tag{33}$$

Thus, inequality (32) leads to another characterization of the Fredholm eigen functions.

We obtain corresponding inequalities when we interchange the role of $D$ and $\tilde{D}$; the Dirichlet integral of a function $h \in \Sigma$ can then be estimated in terms of a function $\tilde{h} \in \tilde{\Sigma}$ which has on $C$ either the same boundary values or the same normal derivative as $h$.

The most convenient form in which the preceding theory can be applied is obtained by using $\varepsilon = 1$. For, in this case, the dielectric Green's function reduces to the elementary function $-\log |z - \zeta|$ and the bilinear forms can be easily evaluated. Indeed, the general method was first applied to obtain isoperimetric inequalities for polarization and virtual mass with this particular choice of $\varepsilon$ [18, 19, 20]. However, the flexibility of the method is obviously increased by considering arbitrary positive $\varepsilon$-values and the significance of the procedure is clarified in this way.

We shall now utilize the quadratic forms in order to obtain estimates for the Fredholm eigen values $\lambda_\nu$. Let $\lambda_1$ be the lowest positive Fredholm eigen value $> 1$. We have shown in § 1 that with $\lambda_1$ also $-\lambda_1$ is an eigen value. We denote $\lambda_2 = -\lambda_1$. By definition (2.12) of the $\rho_\nu$, we have obviously

$$\frac{1}{\rho_1} = \rho_2 \leq \rho_\nu \leq \rho_1 , \qquad \nu = 1, 2, 3, \cdots . \tag{34}$$

Using now the developments (11), (12) and (14) of the various bilinear forms, we verify by inspection the following theorems:

THEOREM IV. *For every function $h \in \Sigma$ the inequalities*

$$\frac{\varepsilon}{1+\varepsilon\rho_1} \leq \frac{\pi_\varepsilon(h, h)}{D(h, h)} \leq \frac{\varepsilon\rho_1}{\rho_1+\varepsilon} \tag{35}$$

*hold. The first equality sign holds only for those function $h_\nu \in \Sigma$ which belong to the eigen value $\lambda_1$; the second equality sign holds only for functions $h_\nu \in \Sigma$ which belong to the eigen value $\lambda_2$.*

THEOREM V. *Every function $\tilde{h} \in \tilde{\Sigma}$ satisfies the inequalities*

$$\frac{\varepsilon}{\rho_1+\varepsilon} \leq \frac{\tilde{\pi}_\varepsilon(\tilde{h}, \tilde{h})}{\tilde{D}(\tilde{h}, \tilde{h})} \leq \frac{\varepsilon\rho_1}{1+\varepsilon\rho_1}. \tag{36}$$

*Equality holds only if $\tilde{h} = \tilde{h}_\nu$ where $\tilde{h}_\nu$ belongs to the eigen values $\lambda_2$ and $\lambda_1$, respectively.*

We have thus characterized the lowest positive and non-trivial Fredholm eigen value $\lambda_1$ by a minimum and a maximum problem in $\Sigma$ and in $\tilde{\Sigma}$ for the ratio of two positive-definite quadratic forms. This characterization makes it possible to estimate this eigen value by the use of test functions in $\Sigma$ and in $\tilde{\Sigma}$. The most convenient case for applications is, of course, the case $\varepsilon = 1$.

It is clearly desirable to find analogous extremum problems which characterize the higher eigen values $\lambda_\nu$. For this purpose, we introduce the bilinear form

$$\Gamma_{\varepsilon,e}(h, H) = \frac{\Gamma_\varepsilon(h, H) - \Gamma_e(h, H)}{\varepsilon - e}, \qquad \varepsilon > 0, e > 0 \tag{37}$$

in $\Sigma$ and the bilinear form

$$\tilde{\pi}_{\varepsilon,e}(\tilde{h}, \tilde{H}) = \frac{\tilde{\pi}_\varepsilon(\tilde{h}, \tilde{H}) - \tilde{\pi}_e(\tilde{h}, \tilde{H})}{\varepsilon - e}, \qquad \varepsilon > 0, e > 0 \tag{37'}$$

in $\tilde{\Sigma}$. From (11) and (14), we obtain the Fourier representations

$$\Gamma_{\varepsilon,e}(h, H) = \sum_{\nu=1}^{\infty} \frac{\rho_\nu^2 x_\nu y_\nu}{(1+\varepsilon\rho_\nu)(1+e\rho_\nu)}; \quad \tilde{\pi}_{\varepsilon,e}(\tilde{h}, \tilde{H}) = \sum_{\nu=1}^{\infty} \frac{\rho_\nu \tilde{x}_\nu \tilde{y}_\nu}{(1+\varepsilon\rho_\nu)(1+e\rho_\nu)} \tag{38}$$

The quadratic forms $\Gamma_{\varepsilon,e}(h, h)$ and $\tilde{\pi}_{\varepsilon,e}(\tilde{h}, \tilde{h})$ are evidently positive-definite.

We observe that the function

$$f(x) = \frac{x}{(1+\varepsilon x)(1+ex)} \tag{39}$$

takes in the interval $0 \leq x < \infty$ its maximum value at the point

$$X_m = \frac{1}{\sqrt{\varepsilon e}} . \tag{40}$$

Hence, (38) yields the following theorems:

THEOREM VI. *Every function $h \in \Sigma$ satisfies the inequality*

$$\frac{\Gamma_{\varepsilon,e}(h, h)}{D(h, h)} \leq f(\rho_m) \tag{41}$$

*where $\rho_m$ is a value in the sequence of the $\rho_\nu$ which gives the largest value of $f$. Equality holds only for such $h_\nu$ which belong to such a value $\rho_m$.*

THEOREM VII. *For every function $\tilde{h} \in \tilde{\Sigma}$, the inequality*

$$\frac{\tilde{\pi}_{\varepsilon,e}(\tilde{h}, \tilde{h})}{\tilde{D}(\tilde{h}, \tilde{h})} \leq f(\rho_m) \tag{42}$$

*holds where $\rho_m$ is a value in the sequence of the $\rho_\nu$ which gives the largest possible value of $f(\rho)$. Equality holds only for such $\tilde{h}_\nu$ which belong to such a $\rho_m$.*

Given any specific $\rho_\nu$, we can always choose $\sqrt{\varepsilon e} = \rho_\nu^{-1}$ and the corresponding maximum problem will pick out this particular eigen value. We can apply Theorems VI and VII in order to obtain estimates for the location of $\rho_\nu$-values near any given point $x_m$ by the use of test functions in $\Sigma$ and in $\tilde{\Sigma}$. It is easily seen that Theorems IV and V are contained in Theorems VI and VII as limit cases.

We specialize in Theorem IV $\varepsilon = 1$ and obtain the particular result

$$\frac{1}{2\pi} \int_C \int_C \log \frac{1}{|z - \zeta|} \frac{\partial h(z)}{\partial n} \frac{\partial h(\zeta)}{\partial n} ds_z ds_\zeta \leq \frac{\rho_1}{1 + \rho_1} D(h, h) \tag{35'}$$

for every $h \in \Sigma$; equality holds only if $h = h_\nu$ and $h_\nu$ belongs to $\lambda_2$.

This result permits the following application. Consider the system of curves $C^*$ which consists of the subset $C_1, C_2, \cdots, C_{N^*}$ of $C$ with $N^* < N$. This system of boundaries determines a connected exterior $\tilde{D}^* \supset \tilde{D}$ and the set $D^*$ of the domains $D_j$, $j = 1, \cdots, N^*$. Let $\Sigma^*$ be the function class in $D^*$ which is analogous to the class $\Sigma$ in $D$ and let $h_2^*(z)$ correspond to the largest non-trivial negative eigen value $\lambda_2^*$ of $C^*$. We determine a function $h(z) \in \Sigma$ such that

$$\frac{\partial h}{\partial n} = \frac{\partial h_2^*}{\partial n} \text{ on } C^* , \quad \frac{\partial h}{\partial n} = 0 \text{ on } C - C^* . \tag{43}$$

Since the boundary conditions (43) determine $h(z)$ in each $D_j$ only up to an additive constant, we may adjust these constants in such a way that $h(z)$ satisfies the $N$ conditions (2.18) and thus belongs indeed to $\Sigma$. Observe that the Dirichlet integral of $h$ coincides in each $D_j$, $j \leq N^*$ with the corresponding Dirichlet integral of $h_2^*$, since $h$ and $h_2^*$ differ only by a constant in these domains. In each $D_j$ with $j > N^*$, $h(z)$ is a constant and has the Dirichlet integral zero. Hence:

$$D^*(h_2^*, h_2^*) = D(h, h) . \tag{44}$$

By (35′) we have

$$\frac{\rho_1^*}{1+\rho_1^*} D^*(h_2^*, h_2^*) = \frac{1}{2\pi}\int_{C^*}\int_{C^*} \log\frac{1}{|z-\zeta|}\frac{\partial h_2^*(z)}{\partial n}\frac{\partial h_2^*(\zeta)}{\partial n} ds_z ds_\zeta \tag{45}$$
$$= \frac{1}{2\pi}\int_C\int_C \log\frac{1}{|z-\zeta|}\frac{\partial h(z)}{\partial n}\frac{\partial h(\zeta)}{\partial n} ds_z ds_\zeta \leq \frac{\rho_1}{1+\rho_1} D(h, h) .$$

By virtue of (44), we conclude finally

$$\frac{\rho_1^*}{1+\rho_1^*} \leq \frac{\rho_1}{1+\rho_1}, \quad \rho_1^* \leq \rho_1 . \tag{46}$$

Thus, we proved:

THEOREM VIII. *The lowest positive and non-trivial eigen value $\lambda_1$ of a curve system $C$ is never larger than the corresponding eigen value $\lambda_1^*$ of any subsystem $C^*$ of $C$.*

Suppose all positive eigen values of $C$ arranged in increasing order, say $\lambda_{\nu'}$, such that $\nu' < \nu''$ implies $\lambda_{\nu'} \leq \lambda_{\nu''}$. Let us do the same with the positive eigen values $\lambda_{\nu'}^*$ of the subsystem $C^*$. By the above reasoning and by use of the standard methods of eigen value theory [cf. **11**], it is easily shown that quite generally

$$\lambda_{\nu'} \leq \lambda_{\nu'}^* \tag{47}$$

will be fulfilled.

We consider finally the bilinear form

$$B(h, H) = \frac{1}{2\pi}\int_C\int_C \Gamma(\zeta, \eta)\frac{\partial h(\zeta)}{\partial n}\frac{\partial H(\eta)}{\partial n} ds_\zeta ds_\eta \tag{48}$$

where $\Gamma(\zeta, \eta)$ is the geometric kernel defined in (2.26). For $h \in \Sigma$, $H \in \Sigma$ we have, in view of (2.28) the following Fourier representation for $B$:

$$B(h, H) = \sum_{\nu=1}^{\infty} \frac{\rho_\nu}{(1+\rho_\nu)^2} x_\nu y_\nu \tag{49}$$

and the same expression is also valid for $\tilde{h} \in \tilde{\Sigma}$, $\tilde{H} \in \tilde{\Sigma}$, From (11), (38) and (49) follows

(50) $$\pi_1(h, H) - B(h, H) = \sum_{\nu=1}^{\infty} \frac{\rho_\nu^2}{(1+\rho_\nu)^2} x_\nu y_\nu = \Gamma_{1,1}(h, H) ,$$

and

(51) $$B(\tilde{h}, \tilde{H}) = \tilde{\pi}_{1,1}(\tilde{h}, \tilde{H}) .$$

The function

(39′) $$f(x) = \frac{x}{(1+x)^2}$$

takes its maximum 1/4 for positive argument at the point $x_m = 1$ and we derive from (41) and (42) the inequalities

(52) $$0 \leq \pi_1(h, h) - B(h, h) \leq \frac{1}{4} D(h, h) , \qquad h \in \Sigma$$

and

(53) $$0 \leq B(\tilde{h}, \tilde{h}) \leq \frac{1}{4} \tilde{D}(\tilde{h}, \tilde{h}) , \qquad \tilde{h} \in \tilde{\Sigma} .$$

These inequalities are interesting since they yield estimates for the Dirichlet integrals of $h$ and $\tilde{h}$ by means of elementary integrations over $C$ which involve only the normal derivatives on $C$ of these functions and geometric terms. On the other hand, given only these normal derivatives, we could calculate the precise Dirichlet integrals only after solving a Neumann boundary value problem for the domains. We gave by inequality (32) another lower bound for $\tilde{D}(\tilde{h}, \tilde{h})$; but in this estimate we have to assume as known the solution of the corresponding boundary value problem for the complimentary region $D$ of $\tilde{D}$. The present inequalities are, therefore, often easier to apply.

The dielectric Green's functions $g_\varepsilon(z, \zeta)$ and $g_e(z, \zeta)$ which are needed in the calculation of $\tilde{\pi}_{\varepsilon,e}$ and $\Gamma_{\varepsilon,e}$ are known only for very few domains if $\varepsilon$ and $e$ are different from 1. We may, however, use the series developments (2.34) for these functions and utilize the partial sums in the development together with a simple estimate for the remainder terms in order to obtain estimates for $\rho_m$. The calculations are clearly quite laborious, but in principle feasible.

**6. Variational formulas for the dielectric Green's functions and for the Fredholm eigen values.** The properties (a)–(e) ennumerated in § 2 and defining the dielectric Green's functions $g_\varepsilon(z, \zeta)$ are all invariant under a conformal mapping $z^* = F(z)$ which is normalized at infinity such that $|F'(\infty)| = 1$. Unfortunately, the only conformal mapping of this kind which is regular in the entire complex plane has the trivial form $F(z) = az + b$, $|a| = 1$. We may consider, however, functions $F(z)$ which are analytic with isolated singularities. In this way, we are led naturally to a variational theory for the dielectric Green's functions.

The simplest possible choice of $F(z)$ is evidently

$$z^* = F(z) = z + \frac{\alpha}{z - z_0} \tag{1}$$

which has the right normalization at infinity but has a simple pole at $z = z_0$. We will choose $z_0$ arbitrarily in $D$ or in $\tilde{D}$ but not on the curve system $C$. Let $E(z_0)$ denote the entire complex plane from which a circle of radius $\sqrt{|\alpha|}$ around the center $z_0$ has been removed. It is easily seen that $F(z)$ is univalent in $E(z_0)$. Given, therefore, a fixed point $z_0$ in $D$ or in $\tilde{D}$, we can always choose $|\alpha|$ so small that $C$ lies in $E(z_0)$ and is mapped in a one-to-one manner into a new curve system $C^*$. Since $F(z)$ is regular analytic in $E(z_0)$ all differentiability properties of $C$ are transferred to $C^*$. We denote the dielectric Green's functions of the new curve system $C^*$ by $g_\varepsilon^*(z, \zeta)$. Our aim is to connect these new functions with the functions $g_\varepsilon(z, \zeta)$ of the original system $C$.

We introduce the function

$$d(z, \zeta) = g_\varepsilon^*(F(z)\,,\ F(\zeta)) - g_\varepsilon(z, \zeta)\,. \tag{2}$$

By the definition of $g_\varepsilon^*$ and of the curve system $C^*$, the function $d(z, \zeta)$ is symmetric and harmonic for $z, \zeta \in E(z_0)$, except along the curve set $C$. The function is still continuous across $C$ but its normal derivatives satisfy the discontinuity relation

$$\frac{\partial}{\partial n_z} d(z, \zeta) + \varepsilon \frac{\partial}{\partial \tilde{n}_z} d(z, \zeta) = 0 \quad \text{for} \quad z \in C, \zeta \in E(z_0) - C\,. \tag{3}$$

Observe that $d(z, \zeta)$ is still regular harmonic for $z = \zeta$ and that

$$\lim_{z \to \infty} d(z, \zeta) = 0\,. \tag{4}$$

We consider now the integral

$$J(\zeta, \eta) = \frac{1}{2\pi} \int_C \left[ d(z, \zeta) \frac{\partial}{\partial n_z} g_\varepsilon(z, \eta) - g_\varepsilon(z, \eta) \frac{\partial}{\partial n_z} d(z, \zeta) \right] ds_z\,. \tag{5}$$

We introduce the characteristic function $\delta(z)$ of $D$, i.e., we define

$$\delta(z) = \begin{cases} 1 & \text{if} \quad z \in D \\ 0 & \text{if} \quad z \notin D\,. \end{cases} \tag{6}$$

By Green's identity applied to $D$, we find

(7) $$J(\zeta, \eta) = \varepsilon d(\zeta, \eta)\delta(\eta) + T(\zeta, \eta)\delta(z_0) .$$

Here

(8) $$T(\zeta, \eta) = \frac{1}{2\pi} \int_{c(z_0)} \left[ d(z, \zeta) \frac{\partial}{\partial n_z} g_\varepsilon(z, \eta) - g_\varepsilon(z, \eta) \frac{\partial}{\partial n_z} d(z, \zeta) \right] ds_z ,$$

where $c(z_0)$ is the circumference of radius $\sqrt{|\alpha|}$ around $z_0$ and where $\mathbf{n}$ is its interior normal.

On the other hand, we may apply Green's identity to $J(\zeta, \eta)$ with respect to the complementary domain $\tilde{D}$. Taking notice of (4) and of the known discontinuity behavior of the various terms in the integrand, we find

(9) $$J(\zeta, \eta) = -\varepsilon d(\zeta, \eta)[1 - \delta(\eta)] - \varepsilon T(\zeta, \eta)[1 - \delta(z_0)] .$$

Subtracting (9) from (7), we obtain finally

(10) $$\varepsilon d(\zeta, \eta) = -T(\zeta, \eta)[\varepsilon + (1 - \varepsilon)\delta(z_0)] .$$

The difference function (2) of $g_\varepsilon^*$ and $g_\varepsilon$ is thus expressed in terms of an integral over the small circle $c(z_0)$ around the singularity point $z_0$.

A straightfoward calculation of the type usual in such variational problems [15, 21] yields

(11) $$g_\varepsilon^*(\zeta^*, \eta^*) = g_\varepsilon(\zeta, \eta) + \left[1 + \left(\frac{1}{\varepsilon} - 1\right)\delta(z_0)\right] \Re\{\alpha p_\varepsilon'(z_0, \zeta) p_\varepsilon'(z_0, \eta)\} + O(|\alpha|^2) ,$$

where $p_\varepsilon(z, \zeta)$ is the analytic function defined in § 4 whose real part is $g_\varepsilon(z, \zeta)$. The error term $O(|\alpha|^2)$ can be estimated uniformly for $\zeta$ and $\eta$ in $E(z_0)$ and for $z_0$ in any fixed closed domain which does not contain points of $C$.

We derived in (11) an interior variational formula for the dielectric Green's function which is very similar to the well-known variational formula for the ordinary Green's function of a domain [14, 15]. Observe that in the special case $\varepsilon = 1$ formula (11) reads

(11') $$\log \frac{1}{|\zeta^* - \eta^*|} = \log \frac{1}{|\zeta - \eta|} + \Re\left\{ \frac{\alpha}{(z_0 - \zeta)(z_0 - \eta)} \right\} + O(|\alpha|^2) .$$

In view of the identity

(11'') $$\log |\zeta^* - \eta^*| = |\zeta - \eta| + \log \left| 1 - \frac{\alpha}{(z_0 - \zeta)(z_0 - \eta)} \right| ,$$

we can verify (11') directly by means of the logarithmic series.

We shall not enter into the variational theory of the dielectric

Green's functions since it is entirely analogous to that given in the case of simply-connected domains [17]. We wish to utilize (11) in order to derive analogous variational formulas for the eigen values $\lambda_\nu$. For this purpose, we shall make use of the extremum principles (5.41) and (5.42) and of the method of transplanting the extremum function [6, 11].

Let us suppose that the singular point $z_0$ of our variation (1) lies in $\tilde{D}$; in this case, the function $F(z)$ is regular and univalent in $D$. If $h(z)$ is any analytic function in $D$, we can define by

$$h^*(z^*) = h(z) \tag{12}$$

a regular analytic function $h^*$ in each component $D_j^*$ of the varied domain set $D^*$. We call the definition (12) the transplantation of the function $h(z)$ from $D$ into $D^*$.

We define now the ratios

$$R(h) = \frac{\Gamma_{\varepsilon,e}(h, h)}{D(h, h)}, \qquad R^*(h^*) = \frac{\Gamma^*_{\varepsilon,e}(h^*, h^*)}{D^*(h^*, h^*)} \tag{13}$$

which occur in the extremum problem (5.41). In view of the conformal character of the transplantation, we have clearly

$$D(h, h) = D^*(h^*, h^*) \tag{14}$$

and

$$\frac{\partial h^*(z^*)}{\partial n^*} ds^* = \frac{\partial h(z)}{\partial n} ds, \qquad z \in C, z^* \in C^*. \tag{15}$$

It is, therefore, easy to calculate the ratio $R^*(h^*)$ by referring back to the original region $D$. By the definitions (5.10), (5.13) and (5.37), we find

$$\Gamma^*_{\varepsilon,e}(h^*, h^*) = \frac{1}{\varepsilon - e} \cdot \frac{1}{2\pi} \int_{C^*}\int_{C^*} \left[\frac{1}{e} g_e^*(\zeta^*, \eta^*) - \frac{1}{\varepsilon} g_\varepsilon^*(\zeta^*, \eta^*)\right] \cdot \\ \cdot \frac{\partial h^*(\zeta^*)}{\partial n^*} \frac{\partial h^*(\eta^*)}{\partial n^*} ds_\zeta^* ds_\eta^* . \tag{16}$$

Now, we use (11) and (15) in order to return to the curve system $C$ as the path of integration. We remember that $z_0 \in \tilde{D}$ and obtain

$$\Gamma^*_{\varepsilon,e}(h^*, h^*) = \Gamma^*_{\varepsilon,e}(h, h) + 2\pi\Re\left\{\alpha \frac{e^{-1}q_e(z_0)^2 - \varepsilon^{-1}q_\varepsilon(z_0)^2}{\varepsilon - e}\right\} + O(|\alpha|^2) \tag{17}$$

with

$$q_\varepsilon(z) = \frac{1}{2\pi} \int_C p_\varepsilon'(z, \zeta) \frac{\partial h(\zeta)}{\partial n} ds_\zeta . \tag{17'}$$

Since $z_0 \in \tilde{D}$, we can express $q_\varepsilon(z)$ as a surface integral

$$(18) \qquad q_\varepsilon(z_0) = -\frac{1}{\partial z_0}\left[\frac{1}{\pi}\iint_D \nabla g_\zeta(z_0, \zeta)\cdot \nabla h(\zeta) d\tau_\zeta\right].$$

The error term $O(|\alpha|^2)$ can be estimated uniformly for all functions $h(z)$ with bounded Dirichlet integral and for $z_0$ in a closed subdomain of $\tilde{D}$. We have to use the known error term in the variational formula (11) for the dielectric Green's function.

As a first result we can conclude that the eigen values of the ratio $R^*(h^*)$ depend continuously on $\alpha$ and converge with $|\alpha| \to 0$ to the corresponding eigen values of $R(h)$. We can, moreover, derive a precise asymptotic formula for these eigen values.

Let indeed $\rho_0$ be a particular $\rho_\nu$-value of the original curve system $C$ and let the function $f(x)$, defined in (5.39), be chosen in such a way that it takes its maximum at a point $x_m$ which is nearer to $\rho_0$ than to any other $\rho_\nu$. If $h_0 \in \Sigma$ is an eigen function which belongs to $\rho_0$, we will have

$$(19) \qquad R(h_0) = f(\rho_0) .$$

We may assume as before (see (2.17)) that

$$(19') \qquad D(h_0, h_0) = \rho_0 .$$

If $h_0^*$ is the transplantation of $h_0$ into $D^*$, we can use (14) and (17) in order to determine its ratio $R^*(h_0^*)$. But now we can use formulas (2.9), (2.10) and (2.13′) in order to express the analytic function $q_\varepsilon(z)$ by means of the analytic completion of $\tilde{h}_0(z)$ defined in (1.31). We have

$$(20) \qquad q_\varepsilon(z_0) = \frac{\varepsilon\rho_0}{1+\varepsilon\rho_0}\tilde{V}_0'(z_0) , \qquad z_0 \in \tilde{D} .$$

We can now combine (14), (17) and (19) in order to express $R^*(h_0^*)$. We make also use of (19) and of the definition (5.39) of $f(x)$; thus, we arrive finally at

$$(21) \qquad R^*(h_0^*) = f(\rho_0) - 2\pi\rho_0 f'(\rho_0)\Re\{\alpha \tilde{V}_0'(z_0)^2\} + O(|\alpha|^2) .$$

The function $h_0^*(z^*)$ defined by the transplantation of $h_0(z)$ will not, in general, belong to the class $\Sigma^*$ defined with respect to $D^*$ by linear conditions analogous to (2.18). However, we can add to every function $h^*(z^*)$ which is analytic in $D^*$ a different constant in each component $D_j^*$ in order to bring it into the class $\Sigma^*$. This trivial readjustment does not affect the Dirichlet integral nor the quadratic from $\Gamma_{\varepsilon,e}^*$ which depends only upon the normal derivatives of $h^*$. Thus, in the theory of the ratio $R^*(h^*)$ the restriction to the class $\Sigma^*$ is unessential, since easily achieved.

In particular, we may use $h_0^*$ as a competing function for the extremum problem regarding $R^*(h^*)$ and use the identity (21) in order to estimate the extremum values. Let us suppose that the value $\rho_0$ belongs to $k$ different eigen functions $h_\beta(z)$ of the unperturbed curve system $C$; we denote their analytic completions by $V_\beta(z)$. We restrict, at first, $h^*(z^*)$ to the linear sub-space spanned by the $k$ transplanted eigen functions $h_\beta^*(z^*)$. In this case, the ratio $R^*(h^*)$ will have precisely the $k$ stationary values

$$\tau_\beta = f(\rho_0) + 2\pi\rho_0 f'(\rho_0)\sigma_\beta + O(|\alpha|)\,, \qquad \beta = 1, 2, \cdots, k \tag{22}$$

where the $\sigma_\beta$ are the eigen values of the secular equation

$$\det \| \Re\{\alpha \tilde{V}_i'(z_0)\tilde{V}_j'(z_0)\} + \sigma\delta_{ij} \|_{i,j=1,2\cdots k} = 0\,. \tag{23}$$

Let us arrange the $\tau_\beta$ in decreasing order; likewise, we shall arrange the values $f(\rho_\beta^*)$ in decreasing order. Since the $k$ first values $f(\rho_\beta^*)$ are the largest stationary values of $R^*(h^*)$ for unrestricted argument function $h^*$, it follows from standard results on quadratic forms that

$$f(\rho_\beta^*) \geq f(\rho_0) + 2\pi\rho_0 f'(\rho_0)\sigma_\beta + O(|\alpha|^2), \qquad \beta = 1, \cdots, k\,. \tag{24}$$

Because of the continuous dependence of the eigen values $\rho_\nu^*$ on $\alpha$ there exists a positive constant $\delta$ such that for small enough $\alpha$ all eigen values $\rho_\gamma^*$ have from $\rho_0$ a distance larger than $\delta$, except for $k$ eigen values $\rho_\beta^*$ which can be brought arbitrarily near to $\rho_0$.

Having now chosen $|\alpha|$ sufficiently small, we can select $x_m$ to the left of $\rho_0$ and the $k$ neighboring $\rho_\beta^*$ but so near that all other $f(\rho_\gamma^*)$ are less than any of the $f(\rho_\beta^*)$. Since $f'(\rho) < 0$ for $\rho_0$ and all $\rho_\beta^*$, we derive from (24)

$$\rho_\beta^* \leq \rho_0 + 2\pi\rho_0\sigma_\beta + O(|\alpha|^2)\,, \qquad \beta = 1, 2, \cdots, k\,. \tag{24'}$$

Choosing, on the other hand, $x_m$ to the right of $\rho_0$ and the $\rho_\beta^*$ but again so near that $f(\rho_\beta^*)$ is still larger than all $f(\rho_\gamma^*)$, we obtain

$$\rho_\beta^* \geq \rho_0 + 2\pi\rho_0\sigma_\beta + O(|\alpha|^2)\,, \qquad \beta = 1, 2, \cdots, k\,. \tag{24''}$$

Thus, we proved:

The variation of an eigen value $\rho_0$ with degree of degeneracy $k-1$ is characterized by the formula

$$\rho_\beta^* = \rho_0 + 2\pi\rho_0\sigma_\beta + 0(|\alpha|^2) \tag{25}$$

where the $\sigma_\beta$ are the eigen values of the secular equation (23).

In the case that only one eigen function $h_\nu \in \Sigma$ belongs to $\rho_\nu$, we obtain the simpler variational formula

$$\delta\rho_\nu = -\Re\{2\pi\alpha\rho_\nu \tilde{V}_\nu'(z_0)^2\}\,. \tag{26}$$

By the relation (2.12) between $\rho_\nu$ and the Fredholm eigen value $\lambda_\nu$, we obtain in this case finally

$$\delta\lambda_\nu = (\lambda_\nu^2 - 1)\pi\Re\{\alpha \tilde{V}_\nu'(z_0)^2\} \ . \tag{27}$$

We can proceed in analogous fashion in the case that $z_0 \in D$. We will start then with $\tilde{h}_0 \in \tilde{\Sigma}$ which belongs to $\rho_0$ and which satisfies by (5.42) the equation

$$\tilde{R}(\tilde{h}_0) = \frac{\tilde{\pi}_{\varepsilon,e}(\tilde{h}_0, \tilde{h}_0)}{\tilde{D}(\tilde{h}_0, \tilde{h}_0)} = f(\rho_0) \ . \tag{28}$$

We transplant $\tilde{h}_0$ by an equation (12) into a comparison function $\tilde{h}_0^*$ in $\tilde{D}^*$. We assume the usual normalization

$$\tilde{D}(\tilde{h}_0, \tilde{h}_0) = 1 \tag{29}$$

and have, therefore, also

$$\tilde{D}^*(\tilde{h}_0^*, \tilde{h}_0^*) = 1 \ . \tag{29'}$$

The same chain of calculations as before leads to the asymptotic formula

$$\tilde{R}^*(\tilde{h}_0^*) = \frac{\tilde{\pi}^*_{\varepsilon,e}(\tilde{h}_0^*, \tilde{h}_0^*)}{\tilde{D}^*(\tilde{h}_0^*, \tilde{h}_0^*)} = f(\rho_0) + 2\pi f'(\rho_0)\Re\{\alpha V_0'(z_0)^2\} + O(|\alpha|^2) \ . \tag{30}$$

Here, $V_0(z)$ is the analytic completion of $h_0(z)$ in $D$. This formula is very similar to (21); it differs only by the factor $-\rho_0$. We obtain, therefore, the following result:

If $\rho_\nu$ is an eigen value of degeneracy $k-1$ it will change according to the formula

$$\rho_\beta^* = \rho_\nu + 2\pi\sigma_\beta + O(|\alpha|^2) \qquad \beta = 1, 2, \cdots, k \tag{31}$$

under a variation (1) of the curve system $C$. The $\sigma_\beta$ are the $k$ eigen values of the secular equation

$$\det \| \Re\{\alpha V_i'(z_0) V_j'(z_0)\} - \sigma\delta_{ij} \|_{i,j=1,\cdots,k} = 0 \tag{32}$$

and the $V_i(z)$ are the $k$ analytic functions whose real parts are the eigen functions $h_i(z)$ which belong to $\rho_\nu$.

In the particular case $k = 1$, i.e., non-degeneracy, we have

$$\delta\rho_\nu = \Re\{2\pi\alpha V_\nu'(z_0)^2\} \tag{32'}$$

and hence

$$\delta\lambda_\nu = -(\lambda_\nu - 1)^2\pi\Re\{\alpha V_\nu'(z_0)^2\} \ . \tag{33}$$

There is a lack of symmetry between the variational formulas (23), (25), on the one hand, and (31), (32) on the other. This fact is due to the different normalizations

$$\iint_D \left| V_i'(z) \right|^2 d\tau = D(h_i, h_i) = \rho_i \tag{34}$$

and

$$\iint_{\tilde{D}} \left| \tilde{V}_i'(z) \right|^2 d\tau = D(\tilde{h}_i, \tilde{h}_i) = 1 \; . \tag{35}$$

We were led to these normalizations from the theory of the Fredholm eigen functions $\varphi_i(z)$ through the representation (1.3). These normalizations were also used in the series developments of §§ 2 and 3. However, the variational formulas become symmetric when we define

$$u_\nu(z) = \rho_\nu^{-1/2} V_\nu'(z) \; , \quad \tilde{u}_\nu(z) = i \tilde{V}_\nu'(z) \; . \tag{36}$$

From the definition of the $V_\nu(z)$ and $\tilde{V}_\nu(z)$, their normalizations (34) and (35) and from the definitions (1.33), (1.34) it follows at once that the functions (36) are identical with the functions $u_\nu(z)$ and $\tilde{u}_\nu(z)$ defined at the end of § 1 and normalized by (1.34).

By means of the functions $u_\nu(z)$ and $\tilde{u}_\nu(z)$ we can express the law of variations of the eigen values $\lambda_\nu$ as follows:

THEOREM. *Let $\lambda_\nu$ be a Fredholm eigen value of the curve system $C$ and of degeneracy $k-1$; let $u_\beta(z), \tilde{u}_\beta(z) (\beta = 1, 2, \cdots, k)$ be the set of analytic eigen functions to this eigen value. If we subject the system $C$ to a variation* (1), *we have*

$$\frac{\delta\lambda_\nu}{\lambda_\nu^2 - 1} = \pi\sigma_\beta \tag{37}$$

*where $\sigma_\beta$ is an eigen value of the secular equation*

$$\det \| \Re\{\alpha u_i(z_0) u_j(z_0)\} + \sigma\delta_{ij} \| = 0 \quad \text{if} \quad z_0 \in D \tag{38}$$

*or of*

$$\det \| \Re\{\alpha \tilde{u}_i(z_0) \tilde{u}_j(z_0)\} + \sigma\delta_{ij} \| = 0 \quad \text{if} \quad z_0 \in \tilde{D} \; . \tag{39}$$

*In particular, we have in the case of non-degeneracy*

$$\frac{\delta\lambda_\nu}{\lambda_\nu^2 - 1} = -\pi\Re\{\alpha u_\nu^2(z_0)\} \quad \text{for} \quad z_0 \in D \tag{40}$$

*and*

$$(40') \qquad \frac{\delta\lambda_\nu}{\lambda_\nu^2 - 1} = -\pi\Re\{\alpha\tilde{u}_\nu^2(z_0)\} \quad \text{for} \quad z_0 \in \tilde{D}\,.$$

The preceding variational formulas can also be derived easily from the original integral equation (1.2) by means of the general theory of perturbations [17]. The above derivation is of interest since it allows a more detailed study of the error terms by means of the dielectric Green's function. It is also possible to obtain more precise statements by using the higher variational terms of these Green's functions. It is particularly easy to develop the higher variations for the lowest positive and non-trivial eigen value $\lambda_1$. Consider, for example, a variation (1) of the curve system $C$ with $z_0 \in \tilde{D}$. Let $h(z) \in \Sigma$ and $h^*$ its transplantation into $D^*$. By definition (5.10) and the identity (11″), we have

$$(41) \quad \pi_1^*(h^*, h^*) = \pi_1(h, h) - \frac{1}{2\pi}\int_C\int_C \log\left|1 - \frac{\alpha}{(z - z_0)(\zeta - z_0)}\right| \frac{\partial h(z)}{\partial n}\frac{\partial h(\zeta)}{\partial n}\, ds_z ds_\zeta\,.$$

Thus, $\pi_1(h, h)$ has a very simple transformation law under transplantation. The Dirichlet integral is invariant under transplantation. Since $\rho_1$ leads to the extremum values of the ratio (5.35) it is possible to determine the variations of higher order of $\lambda_1$ with relatively little labor.

We wish, finally, to add a simple algebraic remark to the variational formulas (37), (38) and (39). If $\lambda_\nu$ is of degeneracy $k - 1$ a variation (1) will, in general, reduce this degeneracy. It is, however, remarkable that the secular equations (38) and (39) have only two different eigen values such that even after the variation a degenerate eigen value can only split into two different eigen values, at least, up to the order $O(|\alpha|^2)$. Indeed, $\sigma$ is an eigen value, say of (38) if there exist $k$ real numbers $t_j$ such that the linear equations

$$(42) \qquad \sigma t_i + \sum_{j=1}^{k} \Re\{\alpha u_i u_j\} t_j = 0\,, \qquad i = 1, \cdots, k$$

hold while

$$(42') \qquad \sum_{j=1}^{k} t_j^2 = 1\,.$$

We denote

$$(43) \qquad \sum_{j=1}^{k} u_j t_j = M$$

and reduce (42) to

$$(44) \qquad \sigma t_i + \Re\{\alpha u_i M\} = 0\,, \qquad i = 1, \cdots, k\,.$$

Multiplying the $i$th equation (44) with $u_i$ and summing over all $i$-values, we find :

$$\sigma M + \frac{1}{2}\alpha M \sum_{i=1}^{k} u_i^2 + \frac{1}{2}\bar{\alpha}\bar{M} \sum_{i=1}^{k} | u_i |^2 = 0. \tag{45}$$

On the other hand, multiplying (44) with $t_i$ and summing over $i$, we obtain from (42′)

$$\sigma + \Re\{\alpha M^2\} = 0 . \tag{46}$$

From (45) and (46) we derive

$$- \sigma = \Re\{\alpha M^2\} = \frac{1}{2}\alpha \sum_{i=1}^{k} u_i^2 + \frac{1}{2}\frac{|\alpha M|^2}{\alpha M^2} \sum_{i=1}^{k} | u_i |^2 . \tag{47}$$

Let us put

$$\alpha M^2 = pe^{i\gamma} . \tag{48}$$

The real part and imaginary part of (47) are:

$$\begin{aligned} p \cos\gamma &= \frac{1}{2}\Re\left\{\alpha \sum_{i=1}^{k} u_i^2\right\} + \frac{|\alpha|}{2}\cos\gamma \cdot \sum_{i=1}^{k} | u_i |^2 \\ 0 &= \frac{1}{2}\Im\left\{\alpha \sum_{i=1}^{k} u_i^2\right\} - \frac{|\alpha|}{2}\sin\gamma \cdot \sum_{i=1}^{k} | u_i |^2 \end{aligned} \tag{47′}$$

Eliminating $\cos\gamma$ form the first equation by means of the second, we find

$$\begin{aligned} \sigma = &- \frac{1}{2}\Re\left\{\alpha \sum_{i=1}^{k} u_i(z_0)^2\right\} \\ &\pm \frac{1}{2}\sqrt{|\alpha|^2\left(\sum_{i=1}^{k} | u_i(z_0) |^2\right)^2 - \left[\Im\left\{\alpha \sum_{i=1}^{k} u_i(z_0)^2\right\}\right]^2} . \end{aligned} \tag{49}$$

We see, in particular, that the first variation of each eigen value, whatever its degree of degeneracy, depends only on

$$U(z_0) = \sum_{i=1}^{k} u_i(z_0)^2 \quad \text{and} \quad \Omega(z_0) = \sum_{i=1}^{k} | u_i(z_0)|^2 . \tag{50}$$

Observe that the product of the two possible $\sigma$-values (49) is

$$\frac{1}{4}\left|\alpha \sum_{i=1}^{k} u_i^2\right|^2 - \frac{1}{4}|\alpha|^2\left(\sum_{i=1}^{k} | u_i(z_0) |^2\right)^2 \leq 0 \tag{51}$$

such that under a variation (1) at least one component of a split up multiple eigen value is non-increasing. This is the reason why many maximum problems for positive eigen values lead to degenerate eigen values in the extremum case.

**7. The $L_\varepsilon$-kernels and the variation of the Fredholm determinants.** In this section, we shall discuss certain kernels obtained by complex differentiation of the dielectric Green's functions which will appear in certain variational formulas for important combinations of Fredholm eigen values. The significance of these kernels is best understood by considering the kernel obtained in an analogous way from the ordinary Green's function, say $\tilde{g}(z, \zeta)$ of $\tilde{D}$.

We defined already in (1.17) a kernel $L(z, \zeta)$ with respect to the Green's function $g(z, \zeta)$ of the domain set $D$ and observed its remarkable property (1.18). Analogously, we introduce the kernel

$$(1) \qquad \tilde{L}(z, \zeta) = -\frac{2}{\pi}\frac{\partial^2}{\partial z \partial \zeta}\tilde{g}(z, \zeta) = \frac{1}{\pi(z-\zeta)^2} - \tilde{l}(z, \zeta) .$$

$\tilde{l}(z, \zeta)$ is a regular analytic function for $z$ and $\zeta$ in $\tilde{D}$. We shall need two important facts about $\tilde{l}(z, \zeta)$ for later applications.
(a) For $\zeta \in C$ and $z \in \tilde{D}$, we have

$$(2) \qquad \frac{\partial \tilde{g}(z. \zeta)}{\partial z} \equiv 0 \quad \text{identically in } z \in \tilde{D}, \zeta \in C .$$

This identity remains even valid when $z$ moves onto $C$ but to a point different from $\zeta$. Let now $s$ be the length parameter on $C$, $\zeta(s)$ its parametric representation and $\zeta' = d\zeta/ds$ the local tangent unit vector. We differentiate the identity (2) with respect to $s$ and find

$$(3) \qquad \frac{\partial^2 \tilde{g}(z, \zeta)}{\partial z \partial \zeta}\zeta' + \frac{\partial^2 \tilde{g}(z, \zeta)}{\partial z \partial \bar{\zeta}}\bar{\zeta}' = 0 , \qquad z \in C, \zeta \in C .$$

We multiply this identity by $z'$ and using the symmetry of the first term in $z$ and $\zeta$ as well as the hermitian symmetry of the second term, we conclude:

$$(4) \qquad \tilde{L}(z, \zeta)z'\zeta' = \text{real} \quad \text{for } z \in C, \zeta \in C .$$

By use of (1), we may express this result also in the form

$$(5) \qquad \Im\{\tilde{l}(z, \zeta)z'\zeta'\} = \frac{1}{\pi}\Im\left\{\frac{z'\zeta'}{(z-\zeta)^2}\right\} .$$

This identity is of great interest since the left side expression is a differential depending on the Green's function while the right hand term depends only on the geometry of the curve system $C$. Moreover, it can be shown that $\tilde{l}(z, \zeta)$ is continuous in both variables in the closed domain $\tilde{D} + C$ [3, 21]. We may pass to the limit $z = \zeta$ on both sides of (5); an easy calculation yields the boundary condition

$$(6) \qquad \Im\{\tilde{l}(z, z)z'^2\} = \frac{1}{6\pi}\Im\left\{\frac{z'''}{z'} - \frac{3}{2}\left(\frac{z''}{z'}\right)^2\right\} .$$

Let us denote by $\kappa = \kappa(s)$ the curvature of $C$ at $z(s)$; then (6) obtains the elegant form

$$(7) \qquad \Im\{\tilde{l}(z, z)z'^2\} = \frac{1}{6\pi}\frac{d\kappa}{ds} .$$

In particular, we note that (7) and our assumptions on $C$ yield the

THEOREM. *The function $\tilde{l}(z, z)$ is a quadratic differential of $D$, i.e., satisfies*

$$(7') \qquad \tilde{l}(z, z)z'^2 = \text{real} \quad \textit{on } C$$

*if and only if $\tilde{D}$ is a domain bounded by circumferences $C_\nu$.*

(b) Let $z^* = f(z)$ be a univalent analytic function in $\tilde{D}$ which maps this domain into $\tilde{D}^*$. The conformal invariance of the Green's function is expressed by the identity

$$(8) \qquad \tilde{g}^*(z^*, \zeta^*) = \tilde{g}(z, \zeta)$$

which leads by differentiation to

$$(9) \qquad \tilde{L}^*(z^*, \zeta^*)f'(z)f'(\zeta) = \tilde{L}(z, \zeta) .$$

The $\tilde{l}$-kernel has, therefore, the transformation law

$$(10) \qquad \tilde{l}^*(z^*, \zeta^*)f'(z)f'(\zeta) = \tilde{l}(z, \zeta) + \frac{1}{\pi}\left[\frac{f'(z)f'(\zeta)}{(f(z) - f(\zeta))^2} - \frac{1}{(z - \zeta)^2}\right]$$

and, as a simple calculation shows, in particular

$$(11) \qquad \tilde{l}^*(z^*, z^*)f'(z)^2 = \tilde{l}(z, z) + \frac{1}{6\pi}\{f, z\}$$

where

$$(12) \qquad \{f, z\} = \frac{f'''(z)}{f'(z)} - \frac{3}{2}\left(\frac{f''(z)}{f'(z)}\right)^2$$

is the Schwarzian derivative of $f(z)$.

After these remarks on the kernel $\tilde{L}(z, \zeta)$, we introduce now a new kernel by the following formula which is modeled after (1):

$$(13) \qquad L_\varepsilon(z, \zeta) = -\frac{2}{\pi}\frac{\partial^2 g_\varepsilon(z, \zeta)}{\partial z \partial \zeta} .$$

This kernel is regular analytic and symmetric in both its arguments in $D$ and in $\tilde{D}$, except for a double pole for $z = \zeta$. We define further two kernels which are regular analytic for $z, \zeta \in D$ and for $z, \zeta \in \tilde{D}$, respectively:

(14) $$l_\varepsilon(z,\zeta) = \frac{1}{\pi(z-\zeta)^2} - \frac{1}{\varepsilon} L_\varepsilon(z,\zeta) \quad \text{in } D$$

and

(15) $$\tilde{l}_\varepsilon(z,\zeta) = \frac{1}{\pi(z-\zeta)^2} - L_\varepsilon(z,\zeta) \quad \text{in } \tilde{D} .$$

These kernels have elegant developments in terms of the complex eigen functions of the Fredholm integral equation. We start with the Fourier developments (2.16) and (2.21) for $g_\varepsilon(z,\zeta)$ in terms of the harmonic eigen functions $h_\nu(z)$ and $\tilde{h}_\nu(z)$. Using definition (1.17) and (2.21), we obtain by differentiation

(16) $$l_\varepsilon(z,\zeta) = \left(1 - \frac{1}{\varepsilon}\right)\left[ l(z,\zeta) + \sum_{\nu=1}^{\infty} \frac{V'_\nu(z)V'_\nu(\zeta)}{\rho_\nu(1+\rho_\nu)(1+\varepsilon\rho_\nu)} \right]$$

where the $V_\nu(z)$ are the analytic functions whose real part is $h_\nu(z)$. As pointed out in the preceding section, all $V'_\nu(z)$ have a different normalization and it is more convenient to introduce the functions $u_\nu(z)$ defined by (6.36) which have all the norm 1. Then (16) transforms to

(17) $$l_\varepsilon(z,\zeta) = \left(1 - \frac{1}{\varepsilon}\right)\left[ l(z,\zeta) + \sum_{\nu=1}^{\infty} \frac{u_\nu(z)u_\nu(\zeta)}{(1+\rho_\nu)(1+\varepsilon\rho_\nu)} \right].$$

We observe next that with each eigen value $\lambda_\nu > 0$ which belongs to $u_\nu(z)$, there occurs also the eigen value $-\lambda_\nu$ and it belongs to the eigen function $iu_\nu(z)$. This assertion can be verified directly from the complex integral equations (1.36) and (1.37); it is also a consequence of the fact, noted in § 1, that if $\lambda_\nu$ belongs to an eigen function $h_\nu(z)$ then $-\lambda_\nu$ will be an eigen value with the conjugate harmonic eigen function $k_\nu(z)$. Thus, in formula (17), each product $u_\nu(z)\, u_\nu(\zeta)$ occurs, therefore, twice; once coupled with $\rho_\nu$ and the other time with opposite sign and coupled with $1/\rho_\nu$. We combine these pairs of terms and sum now only over those $\nu$ which correspond to the positive eigen values $\lambda_\nu$. Using (2.12), we obtain finally

(18) $$l_\varepsilon(z,\zeta) = \left(1 - \frac{1}{\varepsilon}\right)\left[ l(z,\zeta) - \sum_{\nu=1}^{\infty} \frac{u_\nu(z)u_\nu(\zeta)}{\lambda_\nu} \right] + E^2 \sum_{\nu=1}^{\infty} \frac{\lambda_\nu^2 - 1}{\lambda_\nu^2 - E^2}\, \frac{u_\nu(z)u_\nu(\zeta)}{\lambda_\nu} .$$

with the notation

(19) $$E = \frac{\varepsilon - 1}{\varepsilon + 1} .$$

Passing to the limit $\varepsilon = 0$ and using the limit relation (3.31), we derive first from (18)

$$l(z, \zeta) = \sum_{\nu=1}^{\infty} \frac{u_\nu(z)u_\nu(\zeta)}{\lambda_\nu} \tag{20}$$

and, hence, (18) simplifies to

$$l_\varepsilon(z, \zeta) = E^2 \sum_{\nu=1}^{\infty} \frac{\lambda_\nu^2 - 1}{\lambda_\nu^2 - E^2} \frac{u_\nu(z)u_\nu(\zeta)}{\lambda_\nu} . \tag{21}$$

Similarly, we transform (15) by differentiation of (2.16) into the identity

$$\tilde{l}_\varepsilon(z, \zeta) = (\varepsilon - 1) \sum_{\nu=1}^{\infty} \frac{\rho_\nu \tilde{V}_\nu'(z) \tilde{V}_\nu'(\zeta)}{(1 + \rho_\nu)(1 + \varepsilon\rho_\nu)} \tag{22}$$

and replacing $\tilde{V}_\nu'(z)$ by $\tilde{u}_\nu(z)$ by means of (6.36), we find

$$\tilde{l}_\varepsilon(z, \zeta) = -(\varepsilon - 1) \sum_{\nu=1}^{\infty} \frac{\rho_\nu \tilde{u}_\nu(\zeta)\tilde{u}_\nu(z)}{(1 + \rho_\nu)(1 + \varepsilon\rho_\nu)} . \tag{23}$$

We combine again terms with $\rho_\nu$ and with $1/\rho_\nu$ and sum only over the positive eigen values $\lambda_\nu$; an easy calculation leads to

$$\tilde{l}_\varepsilon(z, \zeta) = E^2 \sum_{\nu=1}^{\infty} \frac{\lambda_\nu^2 - 1}{\lambda_\nu^2 - E^2} \frac{\tilde{u}_\nu(z)\tilde{u}_\nu(\zeta)}{\lambda_\nu} . \tag{24}$$

The complete symmetry between (21) and (24) is evident.

We consider the limit cases $\varepsilon = 0$ and $\varepsilon = \infty$ of formula (24) which correspond both to $E^2 = 1$. From (3.11) and (3.17) follows

$$\tilde{\lambda}(z, \zeta) = \sum_{\nu=1}^{\infty} \frac{\tilde{u}_\nu(z)\tilde{u}_\nu(\zeta)}{\lambda_\nu} = \frac{1}{\pi(z - \zeta)^2} + \frac{2}{\pi} \frac{\partial^2 \tilde{G}(z, \zeta)}{\partial z \partial \zeta} \tag{25}$$

$$= \frac{1}{\pi(z - \zeta)^2} + \frac{2}{\pi} \frac{\partial^2 \tilde{N}(z, \zeta)}{\partial z \partial \zeta} .$$

We can, therefore, express $\tilde{\lambda}(z, \zeta)$ by means of (3.3) in the form

$$\tilde{\lambda}(z, \zeta) = \tilde{l}(z, \zeta) - \frac{1}{2\pi} \sum_{j,k=1}^{N-1} \alpha_{jk} w_j'(z) w_k'(\zeta) \tag{26}$$

where $w_j(z)$ denotes the analytic completion of the harmonic measure $\omega_j(z)$. Formula (26) is the counterpart for $\tilde{D}$ of the relation (20) in $D$. The kernel $\tilde{\lambda}(z, \zeta)$ is composed of functions with single-valued integral in $\tilde{D}$; the kernel $\tilde{l}(z, \zeta)$ differs from it by a kernel which is composed of a basis of $N - 1$ functions in $\tilde{D}$ which do not have a single-valued integral and which are orthogonal in the Dirichlet metric to all functions in $\tilde{D}$ with single-valued integral.

For the sake of completeness, we give also the Fourier developments of the kernels

(27) $$K_\varepsilon(z, \bar{\zeta}) = -\frac{2}{\pi\varepsilon}\frac{\partial^2 g_\varepsilon(z, \zeta)}{\partial z \partial \bar{\zeta}} \quad \text{in } D$$

and

(28) $$\tilde{K}_\varepsilon(z, \bar{\zeta}) = -\frac{2}{\pi}\frac{\partial^2 g_\varepsilon(z, \zeta)}{\partial z \partial \bar{\zeta}} \quad \text{in } \tilde{D}\,.$$

Both kernels are analytic and have hermitian symmetry in their arguments. Putting

(29) $$K(z, \bar{\zeta}) = -\frac{2}{\pi}\frac{\partial^2 g(z, \zeta)}{\partial z \partial \bar{\zeta}}$$

we obtain by differentiation of (2.21) after the above combination of terms

(30) $$K_\varepsilon(z, \bar{\zeta}) = \left(1 - \frac{1}{\varepsilon}\right)\left[K(z, \bar{\zeta}) - \sum_{\nu=1}^{\infty} u_\nu(z)\overline{u_\nu(\zeta)}\right] + E\sum_{\nu=1}^{\infty}\frac{\lambda_\nu^2 - 1}{\lambda_\nu^2 - E^2}u_\nu(z)\overline{u_\nu(\zeta)}\,.$$

Again, we obtain by passage to the limit $\varepsilon = 0$ and in view of (3.31)

(31) $$K(z, \bar{\zeta}) = \sum_{\nu=1}^{\infty} u_\nu(z)\overline{u_\nu(\zeta)}$$

which reduces formula (30) to

(32) $$K_\varepsilon(z, \bar{\zeta}) = E\sum_{\nu=1}^{\infty}\frac{\lambda_\nu^2 - 1}{\lambda_\nu^2 - E^2}u_\nu(z)\overline{u_\nu(\zeta)}\,.$$

Similarly, we find by differentiation of (2.16) the identity

(33) $$\tilde{K}_\varepsilon(z, \bar{\zeta}) = E\sum_{\nu=1}^{\infty}\frac{\lambda_\nu^2 - 1}{\lambda_\nu^2 - E^2}\tilde{u}_\nu(z)\overline{\tilde{u}_\nu(\zeta)}\,.$$

Formulas (21), (24), (32) and (33) for the various kernels depend on $\varepsilon$ only through $E$ and this simple rational function of $\varepsilon$ has the symmetry property $E(1/\varepsilon) = -E(\varepsilon)$. This leads to the interesting identities:

(34) $$\frac{\partial^2 g_\varepsilon(z, \zeta)}{\partial z \partial \zeta} = \frac{\partial^2 g_{1/\varepsilon}(z, \zeta)}{\partial z \partial \zeta}, \quad \frac{\partial^2 g_\varepsilon(z, \zeta)}{\partial z \partial \bar{\zeta}} = -\frac{\partial^2 g_{1/\varepsilon}(z, \zeta)}{\partial z \partial \bar{\zeta}}$$

if $z, \zeta \in \tilde{D}$ and to a similar identity in $z, \zeta \in D$. These relations are known in the limit case $\varepsilon = 0$ where they represent differential relations between the Green's and the Neumann's function [2, 5, 21].

We define next the Fredholm determinant of the basic integral equation (1.2), Observe again that with each positive eigen value $\lambda_\nu$

occurs also the eigen value $-\lambda_\nu$ in equal multiplicity. We may thus write

$$D(E) = \prod_{\nu=1}^{\infty}\left(1 - \frac{E^2}{\lambda_\nu^2}\right) \tag{35}$$

where the product is to be extended over all positive eigen values $\lambda_\nu > 1$.

By use of the variational formulas (6.38) and (6.39) and of the identities (21) and (24) one can establish readily the

THEOREM. *If the curve system $C$ is varied according to* (6.1) *the Fredholm determinant $D(E)$ changes according to the variational formulas*

$$\delta \log D(E) = -2\pi\Re\{\alpha l_\varepsilon(z_0, z_0)\} \quad \text{for } z_0 \in D \tag{36}$$

*and*

$$\delta \log D(E) = -2\pi\Re\{\alpha \tilde{l}_\varepsilon(z_0, z_0)\} \quad \text{for } z_0 \in \tilde{D}\,. \tag{37}$$

*$E(\varepsilon)$ is the rational function* (19) *of $\varepsilon$.*

The elegant and symmetric variational formulas (36) and (37) show the theoretical interest of the Fredholm determinant (35). We observe that, in particular, for $\varepsilon = \infty$ and $E = 1$ we have by (20) and (25);

$$\delta \log D(1) = -2\pi\Re\{\alpha l(z_0, z_0)\} \quad \text{for } z_0 \in D \tag{38}$$

and

$$\delta \log D(1) = -2\pi\Re\{\alpha \tilde{\lambda}(z_0, z_0)\} \quad \text{for } z_0 \in \tilde{D}\,. \tag{38'}$$

The functional (35) is defined only for curve systems $C$ which are sufficiently differentiable. This fact creates difficulties in applications of the above variational formulas to extremum problems for the Fredholm determinant since it is not sure, a priori, that the extremum system $C$ will have the required smoothness. In many problems, however, it can be shown that the very property of being an extremum set guarantees already that the curve system $C$ is analytic. Thus, we may restrict ourselves from the beginning to the class of analytic curve systems $C$ and formulate the extremum problems only within this class. A first result for a general theory of extremum problems for the Fredholm determinants is the fact that $D(E)$ is semi-continuous from above in the class of all analytic curve systems $C$. In fact, we will prove the

THEOREM. *Let $\tilde{D}_n$ be a sequence of domains, each being bounded by an analytic curve system $C_n$ and with the Fredholm determinant $D_n(E)$. If the domains $\tilde{D}_n$ converge in the Carathéodory sense to a domain $\tilde{D}$ with analytic boundary $C$ and with the Fredholm determinant $D(E)$, then we have for all $E \geq 0$*

(39) $$\overline{\lim}\, D_n(E) \leq D(E)\ .$$

*Proof.* We define the kernel

(40) $$\tilde{\lambda}^{(2)}(z, \bar{\zeta}) = \iint_{\tilde{D}} \tilde{\lambda}(z, \eta)\overline{\tilde{\lambda}(z, \eta)} d\tau_\eta = \sum_{\nu=1}^{\infty} \frac{\tilde{u}_\nu(z)\overline{\tilde{u}_\nu(\zeta)}}{\lambda_\nu^2}$$

and define then recursively

(41) $$\tilde{\lambda}^{(2j)}(z, \bar{\zeta}) = \iint_{\tilde{D}} \tilde{\lambda}^{(2j-2)}(z, \bar{\eta})\tilde{\lambda}^{(2)}(\eta, \bar{\zeta}) d\tau_\eta = \sum_{\nu=1}^{\infty} \frac{\tilde{u}_\nu(z)\overline{\tilde{u}_\nu(\zeta)}}{\lambda_\nu^{2j}}\ .$$

We remark that

(42) $$\iint_{\tilde{D}} \tilde{\lambda}^{(2j)}(z, \bar{z}) d\tau_z = \sum_{\nu=1}^{\infty} \frac{1}{\lambda_\nu^{2j}} = S^{(2j)}\ .$$

We denote the corresponding expressions referring to the domain $\tilde{D}_n$ by the subscripts $n$. We assert, at first:

(43) $$\underline{\lim}\, S_n^{(2j)} \geq S^{(2j)}$$

To prove this assertion, we select a number $\delta > 0$ arbitrarily small and determine a closed subdomain $\tilde{\Delta}$ of $\tilde{D}$ such that

(44) $$\iint_{\tilde{\Delta}} \tilde{\lambda}^{(2j)}(z, \bar{z}) d\tau_z > S^{(2j)} - \delta\ .$$

By the definitions (25), (40), (41) and in view of the continuous dependence of the Green's function $\tilde{G}(z, \zeta)$ on its domain $\tilde{D}$, the kernels $\tilde{\lambda}_n^{(2j)}(z, \bar{\zeta})$ converge to $\tilde{\lambda}^{(2j)}(z, \bar{\zeta})$ uniformly in each closed subdomain of $\tilde{D}$, in particular in $\tilde{\Delta}$. Given $\delta$, we can choose $n(\delta)$ such that for $n > n(\delta)$ the domains $\tilde{D}_n$ contain $\tilde{\Delta}$ and that

(45) $$S_n^{(2j)} = \iint_{\tilde{D}_n} \tilde{\lambda}_n^{(2j)}(z, \bar{z}) d\tau_z > \iint_{\tilde{\Delta}} \tilde{\lambda}_n^{(2j)}(z, \bar{z}) d\tau_z$$
$$\geq \iint_{\tilde{\Delta}} \tilde{\lambda}^{(2j)}(z, \bar{z}) d\tau_z - \delta > S^{(2j)} - 2\delta\ .$$

Since $\delta$ can be chosen arbitrarily small, these inequalities imply (43).

We observe next that by definition (35)

(46) $$-\log D(E) = \sum_{j=1}^{\infty} \frac{1}{j} E^{2j} S^{(2j)}$$

and a corresponding representation is valid for $-\log D_n(E)$. Hence, from (43) follows immediately the asserted inequality (39) and the theorem is proved.

The significance of this theorem is the following. Let $\mathfrak{A}$ be a family of analytic curve systems $C$ and let us ask for the maximum of $D(E)$ within the family $\mathfrak{A}$, for some fixed value $E$. We know that by its definition $D(E) \leq 1$ and is thus trivially bounded in $\mathfrak{A}$. Let $U \leq 1$ denote the least upper bound of $D(E)$ in $\mathfrak{A}$; we can select an extremum sequence of curve sets $C_n$ in $\mathfrak{A}$ such that $D_n(E)$ converges to $U$. If it is possible to select a subsequence $C_{n1}$ of the $C_n$ such that the corresponding domains $D_{n1}$ converge to a domain $D_0$ with analytic boundary $C_0 \in \mathfrak{A}$, then $C_0$ is a maximum curve system. For, by our theorem (38), we have $D_0(E) \geq U$ and, hence, $D_0(E) = U$ since no $D(E)$ in $\mathfrak{A}$ can be larger than $U$. This argument will be applied in the following section to an interesting problem of conformal mapping.

**8. An extremum problem for Fredholm determinants and an existence proof for circular mappings.** In this section, we shall utilize the variational formulas for the Fredholm determinants in order to solve a specific maximum problem. The extremum domains of this problem will be characterized by the property that their boundary $C$ consists of circumferences. In this way, we will then prove that every plane domain can be mapped conformally upon a canonical domain whose boundaries are circumferences. This canonical mapping will appear as the solution of a simple extremum problem for the family of all univalent mappings of the given domain.

We formulate the following extremum problem:

*Let $\tilde{D}$ be a domain in the complex $z$-plane which contains the point at infinity and which is bounded by $N$ closed analytic curves $C$. Let $\mathscr{F}$ be the family of all functions $t = f(z)$ which are analytic in $\tilde{D} + C$, normalized at infinity by $f'(\infty) = 1$ and are univalent in $\tilde{D}$. Each $f(z) \in \mathscr{F}$ will map $\tilde{D}$ upon a domain $\tilde{\Delta}$ with analytic boundary $\Gamma$ and with the Fredholm determinants $\Delta(E)$. We ask for the functions $f(z) \in \mathscr{F}$ which lead to the maximum value of $\Delta(1)$.*

The existence of such maximum functions is by no means obvious. We can assert only that all determinants $\Delta(1)$ obtained by mappings of the family $\mathscr{F}$ have a least upper bound $U \leq 1$. Hence, we may select a sequence of mappings $f_n(z) \in \mathscr{F}$ such that

$$\lim_{n\to\infty} \Delta_n(1) = U\,. \tag{1}$$

Since the $f_n(z)$ are univalent in $\tilde{D}$ we can use the well-known normality properties of these functions and assume without loss of generality that the $f_n(z)$ converge to a limit function $f(z)$, uniformly in each closed subdomain of $\tilde{D}$. The limit function $f(z)$ provides a univalent map of $\tilde{D}$ into a domain $\tilde{\Delta}$ and is normalized at infinity. The image domains $\tilde{\Delta}_n$ converge in the Carathéodory sense to $\tilde{\Delta}$. If we could prove that $\tilde{\Delta}$ has an analytic boundary $\Gamma$, we would know that $f(z) \in \mathscr{F}$ and the semi-continuity from above of $\Delta(1)$ would insure $\Delta(1) = U$, i.e., that $f(z)$ is a maximum function.

In order to prove the fact $f(z) \in \mathscr{F}$ we consider the maximum sequence $f_n(z)$ which converges to $f(z)$. We want to characterize this sequence by comparing it with near-by sequences obtained by infinitesimal variations of their image domains $\tilde{\Delta}_n$. However, if we subject a multiply-connected domain $\tilde{\Delta}_n$ to an interior variation (6.1), we will, in general, obtain a domain $\tilde{\Delta}_n^*$ which is not conformally equivalent to $\tilde{\Delta}_n$ and cannot be obtained from $\tilde{D}$ by a mapping of the family $\mathscr{F}$. Let, indeed, $\omega_l(t)$ be the harmonic measure of the boundary component $\Gamma_l$ of $\Gamma$ with respect to $\tilde{\Delta}$ and let $((p_{jk}))$ denote the period matrix (2.18″) of this set of harmonic measures. The period matrix $((p_{jk}))$ is a conformal invariant and if we preserve the point at infinity under the conformal mappings, the numbers $\omega_l(\infty)$ must likewise be unchanged. On the other hand, it is well-known [**5**, **15**, **21**] that under a variation of the $t$-plane of the type (6.1) and with the singular point $t_0 \in \tilde{\Delta}$, we have

$$p_{jk}^* = p_{jk} + \Re\{\alpha w_j'(t_0) w_k'(t_0)\} + O(|\alpha|^2) \tag{2}$$

and

$$\omega_l^*(\infty) = \omega_l(\infty) + \Re\{\alpha p'(t_0, \infty) w_l'(t_0)\} + O(|\alpha|^2) \tag{3}$$

where again $w_l(t)$ and $p(t, \tau)$ denote the analytic completions in $t$ of the harmonic functions $\omega_l(t)$ and $g(t, \tau)$ in $\tilde{\Delta}$. We see that, in general, the numbers $p_{jk}$ and $\omega_l(\infty)$ will change under interior variations and that the domain $\tilde{\Delta}^*$ will not be obtained from $\tilde{D}$ by a mapping of the family $\mathscr{F}$.

Consider now $m$ points $t_\mu$ in $\tilde{\Delta}$ and the variation

$$t^* = t + \sum_{\mu=1}^{m} \frac{\alpha_\mu}{t - t_\mu} + O(|\alpha|^2), \quad |\alpha| = \max_\mu (|\alpha_\mu|) \tag{4}$$

where the error term is estimated uniformly in $\tilde{\Delta} + \Gamma$. We may choose the $\alpha_\mu$ and the correction term $O(|\alpha|^2)$ such that

$$\Re\left\{\sum_{\mu=1}^{m} \alpha_\mu w_j'(t_\mu) w_k'(t_\mu)\right\} = 0 \tag{5}$$

$$\Re\left\{\sum_{\mu=1}^{m} \alpha_\mu p'(t_\mu, \infty) w_l'(t_\mu)\right\} = 0 \tag{6}$$

and

$$p_{jk}^* = p_{jk}, \quad \omega_l^*(\infty) = \omega_l(\infty). \tag{7}$$

It can be shown, indeed, that given such values $t_\mu$ and $\alpha_\mu$, the variation (4) can be selected in such a way that $\tilde{\Delta}^*$ is conformally equivalent to $\tilde{\Delta}$ and that the points at infinity correspond [21]. Even now, we cannot assert that $\tilde{D}$ goes into $\tilde{\Delta}^*$ by a mapping of the family $\mathscr{F}$ which is normalized at infinity. However, the Fredholm determinants do not change under a homothetic mapping of a domain and, hence, the insistence on the normalization at infinity is unnecessary in our problem. Thus, the above variations (4) will transform the domains $\tilde{\Delta}_n$ of the extremum sequence into conformally equivalent domains $\tilde{\Delta}_n^*$ whose Fredholm determinants $\Delta_n^*(1)$ may be compared with the maximum sequence $\Delta_n(1)$.

We observe that the functions $w'_j(t)\cdot w'_k(t)$ and $p'(t,\infty)\cdot w'_l(t)$ are quadratic differentials of $\tilde{\Delta}$, i.e., functions $Q_k(t)$ which are regular analytic in $\tilde{\Delta}+\Gamma$ and satisfy on $\Gamma$ the boundary condition

$$Q_k(t)t'^2 = \text{real} . \tag{8}$$

At infinity all these functions satisfy the asymptotic relation

$$Q_k(t) = O(|t|^{-3}) . \tag{9}$$

All functions with the properties (8) and (9) from a linear space with real coefficients and of the dimension $3N$-3. We suppose that we have chosen from the above $N(N+1)$ quadratic differentials a fixed basis of $3N$-3 elements $Q_k(t)$, $k=1,2,\cdots,3N$-3.

After these preparations, we return to our maximum sequence of domains $\tilde{\Delta}_n$; we denote by $Q_k^{(n)}(t)$ the corresponding basis of quadratic differentials of $\tilde{\Delta}_n$ and by $Q_k(t)$ the basis for their limit domain $\tilde{\Delta}$. Clearly, we can choose the basis in each $\tilde{\Delta}_n$ and in $\tilde{\Delta}$ such that

$$\lim_{n\to\infty} Q_k^{(n)}(t) = Q_k(t) , \tag{10}$$

uniformly in each closed subdomain of $\tilde{\Delta}$. The determinant

$$\det \| \Re\{Q_k(t_l)\} \| , \qquad l, k = 1, 2, \cdots, 3N\text{-}3 \tag{11}$$

does not vanish identically in $\tilde{\Delta}$ because of the supposed real independence of the $Q_k(t)$. Hence, we can determine $3N$-3 points $t_\mu \in \tilde{\Delta}$ such that

$$\det \| \Re\{Q_k^{(n)}(t_\mu)\} \| \neq 0 \qquad k, \mu = 1, 2, \cdots, 3N\text{-}3 \tag{12}$$

for large enough $n$; we may even assume, without loss of generality, that (12) holds for all integers $n$.

Let $t_0$ be an arbitrary point in $\tilde{\Delta}_n$ and $\alpha^{(n)}$ be an arbitrary complex number. We determine $3N$-3 real numbers $x_\mu^{(n)}$ by the linear equations

$$\Re\{\alpha^{(n)}Q_k^{(n)}(t_0)\} = \sum_{\mu=1}^{3N-3} x_\mu^{(n)}\Re\{Q_k^{(n)}(t_\mu)\} , \quad k = 1, 2, \cdots, 3N\text{-}3 \tag{13}$$

which is always posible because of (12). Observe that $x_\mu^{(n)} = O(|\alpha^{(n)}|)$ for small values of $\alpha^{(n)}$. Consider then the interior variation of $\tilde{\Delta}_n$

(14) $$t^* = t + \frac{\alpha^{(n)}}{t - t_0} - \sum_{\mu=1}^{3N-3} \frac{x_\mu^{(n)}}{t - t_\mu} + O(|\alpha^{(n)}|^2) .$$

This variation is of the type (4), but by the choice (13) of the $x_\mu^{(n)}$, we are sure that the equations (5) and (6) will be fulfilled. We can, therefore, adjust the error term $O(|\alpha^{(n)}|^2)$ in such a way that the varied domain $\tilde{\Delta}_n^*$ is conformally equivalent to $\tilde{\Delta}_n$ and such that the points at infinity correspond. Hence, $\tilde{\Delta}_n^*$ may be used as a competing domain sequence to the maximum sequence $\tilde{\Delta}_n$. We apply now the variational formula (7.38′) in order to characterize the limit domain $\tilde{\Delta}$.

We derive from (7.38′) that the variation (14) of $\tilde{\Delta}_n$ yields

(15) $$\log \Delta_n^*(1) = \log \Delta_n(1) - 2\pi \Re\{\alpha^{(n)} \tilde{\lambda}_n(t_0, t_0)\} + 2\pi \sum_{\mu=1}^{3N-3} x_\mu^{(n)} \Re\{\tilde{\lambda}_n(t_\mu, t_\mu)\} + O(|\alpha^{(n)}|^2) .$$

Here, the $\tilde{\lambda}_n(t, t)$ denote the $\tilde{\lambda}$-kernels of $\tilde{\Delta}_n$. We denote

(16) $$\delta_n = \log U - \log \Delta_n(1) .$$

By the definition of the maximum sequence, we have $0 < \delta_n \to 0$. Since $\log \Delta_n^*(1) \leq \log U$, we infer from (15) the inequality

(17) $$\frac{1}{2\pi}\delta_n \geq - \Re\{\alpha^{(n)} \tilde{\lambda}_n(t_0, t_0)\} + \sum_{\mu=1}^{3N-3} x_\mu^{(n)} \Re\{\tilde{\lambda}_n(t_\mu, t_\mu)\} + O(|\alpha^{(n)}|^2) .$$

We choose finally

(18) $$\alpha^{(n)} = \delta_n r e^{i\tau} , \qquad r > 0$$

and define the real numbers $\xi_\mu$ by the system of linear equations

(19) $$\sum_{\mu=1}^{3N-3} \xi_\mu \Re\{Q_\kappa(t_\mu)\} = \Re\{e^{i\tau} Q_\kappa(t_0)\} , \qquad k = 1, \cdots, 3N\text{-}3 .$$

We divide equations (13) and (17) by $\delta_n$ and pass to the limit $n \to \infty$; comparing (13) with (19), we find

(20) $$\lim_{n\to\infty} \frac{x_\mu^{(n)}}{r\delta_n} = \xi_\mu ;$$

and since at $t_0, t_1, \cdots, t_{3N-3}$ holds

(21) $$\lim_{n\to\infty} \tilde{\lambda}_n(t_\mu, t_\mu) = \tilde{\lambda}(t_\mu, t_\mu) ,$$

we obtain from (17)

$$\frac{1}{2\pi r} \geq - \Re\{e^{i\tau}\tilde{\lambda}(t_0, t_0)\} + \sum_{\mu=1}^{3N-3} \xi_\mu \Re\{\tilde{\lambda}(t_\mu, t_\mu)\} . \tag{22}$$

This inequality holds for arbitrary values $r > 0$; hence, sending $r \to \infty$, we find

$$0 \geq - \Re\{e^{i\tau}\tilde{\lambda}(t_0, t_0)\} + \sum_{\mu=1}^{3N-3} \xi_\mu \Re\{\tilde{\lambda}(t_\mu, t_\mu)\} . \tag{23}$$

If we replace in (19) the signum $e^{i\tau}$ by $-e^{i\tau}$, the solution vector $\xi_\mu$ changes into $-\xi_\mu$. Since $e^{i\tau}$ is entirely arbitrary, the inequality (23) must also hold for inverted sign of the right hand term. Thus, we arrive finally at the equation

$$\Re\{e^{i\tau}\tilde{\lambda}(t_0, t_0)\} = \sum_{\mu=1}^{3N-3} \xi_\mu \Re\{\tilde{\lambda}(t_\mu, t_\mu)\} . \tag{24}$$

valid for arbitrary choice of the signum $e^{i\tau}$ and the corresponding choice (19) of the $\xi_\mu$. The fact that, for given fixed $t_1, \cdots, t_{3N-3}$ in $\tilde{\Delta}$ and for arbitrary $t_0 \in \tilde{\Delta}$, the linear equations (19) always imply the equation (24) for arbitrary $e^{i\tau}$, guarantees the existence of $3N$-3 real numbers $\beta_\mu(\mu = 1, \cdots, 3N\text{-}3)$ such that

$$\tilde{\lambda}(t, t) = \sum_{\mu=1}^{3N-3} \beta_\mu Q_\mu(t) . \tag{25}$$

This identity is then the condition which characterizes the limit domain $\tilde{\Delta}$ of an extremum sequence $\tilde{\Delta}_n$.

Since, in view of (7.26), the function $\tilde{\lambda}(t, t)$ coincides with the more fundamental kernel $\tilde{l}(t, t)$ except for a quadratic differential, we may express the result (25) as follows:

THEOREM I. *If $\tilde{\Delta}$ is the limit domain of a maximum sequence $\tilde{\Delta}_n$, its $\tilde{l}$-kernel satisfies the condition*

$$\tilde{l}(t, t) = Q(t) \tag{26}$$

*where $Q(t)$ is a quadratic differential of $\tilde{\Delta}$.*

From Theorem I, we can deduce

THEOREM II. *All boundary curves $\Gamma_l$ of $\tilde{\Delta}$ are analytic.*

*Proof.* Let us express equation (26) in terms of functionals of the original domain $\tilde{D}$ which is conformally equivalent to $\tilde{\Delta}$. By (7.11) and because of the covariance character of the quadratic differentials under conformal mapping, we can express (26) in the form

$$\tilde{l}(z, z) + \frac{1}{6\pi}\{f, z\} = Q(z) \tag{27}$$

where $Q(z)$ is the quadratic differential in $\tilde{D}$ which corresponds to $Q(t)$ under the mapping $t = f(z)$ of $\tilde{D}$ into $\tilde{\Delta}$ and $\tilde{l}(z, z)$ denotes the $\tilde{l}$-kernel of $\tilde{D}$. We have assumed that $\tilde{D}$ has analytic boundaries $C_\nu$; hence, we can assert that $\tilde{l}(z, z)$ and $Q(z)$ are analytic in the closed region $\tilde{D} + C$. By (7.12), we may now interpret the equation (27) as a linear differential equation with analytic coefficient in $\tilde{D} + C$:

$$\mu''(z) + 3\pi[Q(z) - \tilde{l}(z, z)]\mu(z) = 0 \tag{28}$$

for the unknown function

$$\mu(z) = [f'(z)]^{-1/2} . \tag{29}$$

From the general theory of ordinary differential equations we obtain that $\mu(z)$ is regular analytic in $\tilde{D} + C$ and can have only finitely many zeros on $C$. Hence, $f'(z)$ is analytic on $C$ except for poles which are at least of order 2. At such singular points on $C$, $f(z)$ would have poles too. But $f(z)$ is univalent in $\tilde{D}$ and has already a pole at infinity. It cannot have additional poles on $C$; hence, $f(z)$ and $f'(z)$ are regular analytic on $C$ and the theorem is proved.

In particular, we have now shown that the limit function $f(z)$ of the maximum sequence $f_n(z)$ belongs also to the family $\mathscr{F}$ considered and is, therefore, a maximum function of our problem.

Since we know now that the boundary curves $\Gamma_\iota$ of $\tilde{\Delta}$ are analytic, we can combine (26) with (7.7) and find:

$$\Im\{\tilde{l}(t, t)t'^2\} = \Im\{Q(t)t'^2\} = \frac{1}{6\pi}\frac{d\kappa}{ds} . \tag{30}$$

But $Q(t)$ is a quadratic differential of $\tilde{\Delta}$; thus we arrive at

$$\frac{d\kappa}{ds} = 0 \text{ on each } \Gamma_\iota . \tag{31}$$

This leads to

THEOREM III. *Each boundary curve $\Gamma_\iota$ of the maximum domain $\tilde{\Delta}$ is a circumference.*

Since in each given domain $\tilde{D}$ there exists at least one maximum sequence $f_n(z) \in \mathscr{F}$, we have given a new proof for the classical theorem [5, 7, 8, 9, 23]:

THEOREM IV. *Every plane domain $\tilde{D}$ can be mapped onto a domain bounded by circumferences.*

Since the domain $\tilde{\Delta}$ is the limit of a maximum sequence of domains

$\tilde{\Delta}_n$ and since it is analytically bounded, the semi-continuity of the Fredholm determinants leads to

THEOREM V. *Among all conformally equivalent domains, the circular domains have the largest value of the Fredholm determinant* $D(1)$.

## BIBLIOGRAPHY


1. L. V. Ahlfors, *Conformality with respect to Riemann metrics*, Ann. Acad. Sci. Fenn., Series A **202** (1955).
2. S. Bergman, and M. Schiffer, *A representation of Green's and Neumann's functions in the theory of partial differential equations of second order*, Duke Math. J. **14**, (1947), 609-638.
3. ———, *Kernel functions and conformal mapping*, Comp. Math. **8** (1951), 205-249.
4. I. E. Block, *Kernel functions and class $L^2$*, Proc. Amer. Math. Soc. **4** (1953) 110-117.
5. R. Courant, *Dirichlet's principle, conformal mapping, and minimal surfaces*, Appendix by M. Schiffer, New York 1950.
6. P. R. Garabedian, and M. Schiffer, *Convexity of domain functionals*, J. Analyse Math. **2** (1953), 281-368.
7. A. Hurwitz, and R. Courant, *Funktionentheorie*, Berlin, 1929.
8. P. Koebe, *Über die konforme Abbildung mehrfach zusammenhängender Bereiche*, Jahresber. D. Math. Ver. **19** (1910), 339-348.
9. L. Lichtenstein, *Neuere Entwicklung der Potentialtheorie, Konforme Abbildung*, Enzycl. d. Math. Wissensch. II, C. 3, (1918)
10. H. Poincaré, *Théorie du Potentiel Newtonien*, Paris, 1899.
11. G. Pólya, and M. Schiffer, *Convexity of functionals by transplantation*, J. Analyse Math. **3** (1954), 245–346.
12. R. de Possel, *Sur quelques problèmes de représentation conforme*, C. R. Paris **194** (1932), 42-44.
13. E. Rengel, *Existenzbeweise für schlichte Abbildungen mehrfachzusammenhängender Bereiche auf gewisse Normalbereiche*, Jahresber. D. Math. Ver. **44** (1934), 51-55.
14. M. Schiffer, *Variation of the Green's function and theory of the p-valued functions*, Amer. J. Math. **65** (1943), 341–360.
15. ———, *Hadamard's formula and variation of domain functions*, Amer. J. Math. **68** (1946), 417–448.
16. ———, *The kernel function of an orthonormal system*, Duke Math. J. **13** (1946), 529-540.
17. ———, *The Fredholm eigen values of plane domain*, Pacific J. Math. **7** (1957), 1187-1225.
18. ———, *Sur les rapports entre les solutions des problèmes intérieurs et celles des problèmes extérieurs*. C. R. Paris **244** (1957), 2680-2683.
19. ———, *Sur la polarization et la masse virtuelle*, C. R. Paris **244** (1957), 3118-3121,
20. ———, *Problèmes aux limites et fonctions propres de l'équation intégrale de Poincaré et de Fredholm*, C. R. Paris **245** (1957), 18–21.
21. M. Schiffer, and D. C. Spencer, *Functionals of finite Riemann surfaces*, Princeton 1954.
22. M. Schiffer and G. Szegö, *Virtual mass and polarization*, Trans. Amer. Math. Soc. **67** (1949), 130-205.
23. E. Schottky, *Über die konforme Abbildung mehrfach zusammenhängender ebener Flächen*, J.f.r.u. angew. Math. **83** (1877) 300-351.


STANFORD UNIVERSITY

# A NOTE ON THE COMPUTATION OF ALDER'S POLYNOMIALS

V. N. SINGH

In two recent papers [2, 3] I deduced and used the general transformation

$$(1)\qquad 1+\sum_{s=1}^{\infty}(-1)^s k^{Ms}x^{\frac{1}{2}s\{(2M+1)s-1\}}(1-kx^{2s})\frac{(kx;s-1)}{(x;s)}$$

$$=\prod_{n=1}^{\infty}(1-kx^n)\sum_{t=0}^{\infty}\frac{k^tG_{M,t}(x)}{(x;t)}\,,\qquad (M=2,3,\cdots)$$

to prove certain generalized identities of the type

$$(2)\qquad \prod_{n=1}^{\infty}\frac{(1-x^{(2M+1)n-s})(1-x^{(2M+1)n-(2M+1-s)})(1-x^{(2M+1)n})}{(1-x^n)}$$

$$=\sum_{t=0}^{\infty}\frac{A_s(x,t)G_{M,t}(x)}{(x;t)}\,,$$

where $A_s(x,t)$ and $G_{M,t}(x)$ are polynomials. For $s=M$ and $s=1$ respectively in (2), we get Alder's generalizations of the well-known Rogers-Ramanujan identities

$$\prod_{n=1}^{\infty}\frac{(1-x^{5n-2})(1-x^{5n-3})(1-x^{5n})}{(1-x^n)}=\sum_{t=0}^{\infty}\frac{x^{t^2}}{(x;t)}$$

and

$$\prod_{n=1}^{\infty}\frac{(1-x^{5n-1})(1-x^{5n-4})(1-x^{5n})}{(1-x^n)}=\sum_{t=0}^{\infty}\frac{x^{t(t+1)}}{(x;t)}$$

in the form [1]

$$\prod_{n=1}^{\infty}\frac{(1-x^{(2M+1)n-M})(1-x^{(2M+1)n-M-1})(1-x^{(2M+1)n})}{(1-x^n)}=\sum_{t=0}^{\infty}\frac{G_{M,t}(x)}{(x;t)}$$

and

$$\prod_{n=1}^{\infty}\frac{(1-x^{(2M+1)n-1})(1-x^{(2M+1)n-2M})(1-x^{(2M+1)n})}{(1-x^n)}=\sum_{t=0}^{\infty}\frac{x^tG_{M,t}(x)}{(x;t)}\,.$$

For the Alder polynomials $G_{M,t}(x)$ in (1), I gave the general form

$$(3)\qquad G_{M,t}(x)=x^{t^2}\sum_{t_1=0}^{\left[\frac{M-2}{M-1}t\right]}\frac{(x^{t-2t_1+1};2t_1)x^{-2t_1(t-t_1)}}{(x;t_1)}\prod_{n=2}^{M-2}T_{n,M}$$

Received June 3, 1958.

where

$$T_{n,M} = \sum_{t_n=0}^{\left[\frac{M-n-1}{M-n}t_{n-1}\right]} \frac{(x^{t_{n-1}-2t_n+1};\, 2t_n)x^{-2t_n(t_{n-1}-t_n)}}{(x;\, t_n)(x^{t_{n-2}-2t_{n-1}+1};\, t_n)} \qquad M \geq 2 ,$$

$[a]$ denoting the integral part of $a$.

Alder in his paper [1] states that the polynomials $G_{M,t}(x)$ do not seem to possess any striking properties, even for small values of $M$ and $t$. In the present note, using a simple recurrence relation, I prove beside other results the interesting property that

$$G_{M,t}(x) = x^t , \qquad t \leq (M-1) .$$

The form (3) is not very suitable for the actual computation of the polynomials $G_{M,t}(x)$ for particular values of $M$ and $t$ since certain factor have to be cancelled each time. Therefore, moving into the following series the factor $(x^{t-2t_1+1};\, t_1)$ from the first series and the factor $(x^{t_{n-1}-2t_n+1};\, t_n)$ from each of the $T_{n,M}$ series in (3), we put $G_{M,t}(x)$ in the form

$$(4) \qquad G_{M,t}(x) = x^{t^2} \sum_{t_1=0}^{\left[\frac{M-2}{M-1}t\right]} \frac{(x^{t-t_1+1};\, t_1)x^{-2t_1(t-t_1)}}{(x;\, t_1)} \prod_{n=2}^{M-1} \overline{T}_{n,M}$$

where

$$(5) \qquad \overline{T}_{n,M} = \sum_{t_n=0}^{\left[\frac{M-n-1}{M-n}t_{n-1}\right]} \frac{(x^{t_{n-1}-t_n+1};\, t_n)x^{-2t_n(t_{n-1}-t_n)}}{(x;\, t_n)} \times (x^{t_{n-2}-2t_{n-1}+t_n+1};\, t_{n-1}-t_n) .$$

Now if we put

$$(6) \qquad g_{M,t}(N, x) = \prod_{n=1}^{M-1} \overline{T}_{n,M} \qquad \text{(where } t_{-1} \equiv N) ,$$

then, since

$$(7) \qquad g_{M,t}(N, x) = \sum_{t_1=0}^{\left[\frac{M-2}{M-1}t\right]} \frac{(x^{t-t_1+1};\, t_1)(x^{N-2t+t_1+1};\, t-t_1)}{(x;\, t_1)} \times x^{-2t_1(t-t_1)} g_{M-1,t_1}(t, x) ,$$

it is easily seen by induction that for $t \leq M-1$, we have

$$(8) \qquad g_{M+1,t}(N, x) - g_{M,t}(N, x) = 0$$

because

$$(9) \qquad \left[\frac{M-2}{M-1}t\right] + 1 > \left[\frac{M-1}{M}t\right] \qquad t \leq M-1 .$$

From (4) we have

$$G_{M+1,t}(x) - G_{M,t}(x)$$

(10) $$= x^{t^2} \sum_{t=0}^{\left[\frac{M-2}{M-1}t\right]} \frac{(x^{t-t_1+1}; t_1)x^{-2t_1(t-t_1)}}{(x; t_1)} \{g_{M,t_1}(t, x) - g_{M-1,t_1}(t, x)\}$$

$$+ \sum_{t_1=\left[\frac{M-2}{M-1}t\right]+1}^{\left[\frac{M-1}{M}t\right]} \frac{(x^{t-t_1+1}; t_1)x^{-2t_1(t-t_1)}}{(x; t_1)} g_{M,t_1}(t, x) .$$

Hence from (8) and (9) it follows that, for $t \leq M - 1$,

$$G_{M,t}(x) = G_{M+1,t}(x)$$

that is,

$$G_{M,t}(x) = G_{M+1,t}(x) = \cdots = G_{\infty,t}(x) , \qquad t \leq M - 1 .$$

Now, for $k = 1$ and $M \to \infty$, (1) gives

$$\frac{1}{\prod_{n=1}^{\infty}(1 - x^n)} = \sum_{t=0}^{\infty} \frac{G_{\infty,t}(x)}{(x; t)}$$

whence $G_{\infty,t}(x) = x^t$, so that we finally get

(11) $$G_{M,t}(x) = x^t \qquad t \leq M - 1 .$$

(10) can be further used for the computation of polynomials $G_{M,t}(x)$ as follows.

We first find the general form for $G_{M,M}(x)$.

From (10) we have

(12) $$G_{M+1,M}(x) - G_{M,M}(x) = x_M x^{-2(M-1)} g_{M,M-1}(M, x) ,$$

where $x_n \equiv (1 - x^n)/(1 - x)$ for all $n$.

From (7) we find

(13) $$g_{M,M-1}(M, x) = (x; M - 1)x^{-(M-1)(M-2)} .$$

Using (13) in (12) we get

(14) $$G_{M,M}(x) = x^M\{1 - (x^2; M - 1)\}$$

since $G_{M+1,M}(x) = x^M$. Thus, for example,

$$G_{5,5}(x) = x^7 + x^8 + x^9 - x^{11} - 2x^{12} - x^{13} + x^{15} + x^{16} + x^{17} - x^{19} .$$

More generally, taking $t = M + r$ in (7), since

$$\left[\frac{M^2+(r-2)M-2r}{M-1}\right]=M+r-2 \qquad r\le M-2\,,$$

and

$$\left[\frac{M^2+(r-1)M-r}{M}\right]=M+r-2 \qquad 0<r\le M\,,$$

we easily get

$$g_{M+1,M+r}(N,x)-g_{M,M+r}(N,x)$$

(15) $$=\prod_{n=1}^{r}\overline{T}_{n,M}\{g_{M-r+1,M-r}(t_{r-1},x)-g_{M-r,M-r}(t_{r-1},x)\} \quad 0>r\le M-2\,,$$

where, in $\overline{T}_{n,M}$, $t=M+r$ and $t_r=M-r$. Thus for $t\le 2M-2(t\ne M)$ the second sum on the right of (10) does not exist and we may successively establish the general form of the polynomials $G_{M,t}(x)$ for $M<t\le 2(M-1)$. We thus find that

$$G_{M+1,M+1}(x)-G_{M,M+1}(x)=x^{M+3}(x^3;M-1)x_2 \qquad M\ge 3\,,$$

so that, using (14), we get

$$G_{M,M+1}(x)=x^{M+1}\{1-(x^3;M-1)(1+x^3)\} \qquad M\ge 3\,.$$

Similarly

$$G_{M,M+2}(x)=x^{M+2}\{1-(x^4;M-1)(1+x^4\cdot x_2)\} \qquad M\ge 4\,,$$

$$G_{M,M+3}(x)=x^{M+3}\{1-(x^5;M-1)(1+x^5\cdot x_3)\} \qquad M\ge 5\,,$$

The above values of the polynomials $G_{M,t}(x)$ suggest that probably,

(16) $$G_{M,t}(x)=x^t\{1-(x^{t-M+2};M-1)(1+x^{t-M+2}\cdot x_{t-M})\}\,,$$

for $$t\le 2(M-1)\,.$$

But I have not been able to verify the truth of this conjecture directly.

However, I intend to investigate these interesting polynomials more thoroughly in a future communication.

I am grateful to Dr. R. P. Agarwal for his kind help in the preparation of this note.

## REFERENCES

1. H. L. Alder *Generalizations of the Rogers-Ramanujan identities*, Pacific J. Math. **4** (1954), 161-168.
2. V. N. Singh, *Certain generalized hypergeometric identities of the Rogers-Ramanujan type*, Pacific J. Math. **7** (1957), 1011-1014.
3. ———, *Certain generalized hypergeometric identities of the Rogers-Ramanujan type* (II), Pacific J. Math. **7** (1957), 1691-99.

LUCKNOW UNIVERSITY

# ON INTEGRATION OF 1-FORMS

MAURICE SION

**1. Introduction.** It has been noted by several people that in order to define the integral of some differential 1-form $\omega$ along a curve $C$, the latter need not be of bounded variation. For example, in the extreme (and trivial) case where $\omega$ is the differential of some function $f$, the integral can be defined as the difference of the values assumed by $f$ at the end-points of $C$. No condition on $C$ is necessary. H. Whithney [4], with J. H. Wolfe, by the introduction of certain norms, has found general abstract spaces of curves along which the integral of 1-forms satisfying certain conditions can be defined. In fact, H. Whitney considers integration of $p$-forms with $p \geq 1$. In a previous paper [2], we obtained rather awkward conditions for a decent integral to exist that depended on the number of higher derivatives of $\omega$ on $C$.

In this paper, we consider 1-forms $\omega$ possessing 'higher derivatives' on $C$ in a sense somewhat different from that due to H. Whitney [3] which we used previously. A Lipschitz type condition on the remainders of the Taylor expansion is imposed (see 4.1.). We define the $\alpha$-variation of a curve as the supremum of sums of $\alpha$th powers of chords (see 2.7) and show that the integral of $\omega$ along $C$ exists if the $\alpha$-variation of $C$ is bounded, where $\alpha$ is related to the number of 'higher derivatives' of $\omega$ on $C$. Under somewhat stronger hypotheses on $C$, we show that this integral is an anti-derivative of $\omega$ on $C$.

**2. Notation and basic definitions.** Throughout this paper, $N$ is a positive integer and we use the following notation.

2.1. $E$ denotes Euclidean $(N+1)$-space.

2.2. $\|x\| = \left(\sum_{i=0}^{N} x_i^2\right)^{1/2}$ for $x \in E$.

2.3. diam $U = \sup\{d : d = \|x-y\|$ for some $x \in U$ and $y \in U\}$

2.4. $\varphi$ is a continuous function on the closed unit enterval to $E$ and $C = \text{range } \varphi$.

2.5. $\mathscr{S}$ is the set of all subdivisions of the unit interval, i.e. functions $T$ on $\{0, 1, \cdots, k\}$ for some positive integer $k$ such that: $T(0) = 0,\quad T(k) = 1,\quad T(i-1) < T(i)$ for $i = 1, \cdots, k$

2.6. $[T/a, b] = \{i : a \leq T(i-1) < T(i) \leq b\}$

2.7. $V_\alpha(a, b) = \sup_{T \in \mathscr{S}} \sum_{i \in [T/a,b]} \| \varphi(T(i-1) - \varphi(T(i)) \|^\alpha$

Received August 25, 1958.

## 3. Properties of $V_\alpha$.

3.1. LEMMA. *If* $0 \leq a \leq b \leq c \leq 1$, *then*

$$V_\alpha(a, b) + V_\alpha(b, c) \leq \alpha(a, c) \leq V_\alpha(a, b) + V_\alpha(b, c) + (\text{diam } C)^\alpha$$

3.2. LEMMA. *If* $\alpha < \beta$ and $V_\alpha(a, b) < \infty$, then $V_\beta(a, b) > \infty$.

*Proof.* Since $V_\alpha(a, b) < \infty$, there is an integer $n$ such that there are at most $n$ elements $i \in [T/a, b]$ with $\|\varphi(T(i-1)) - \varphi(T(i))\| \geq 1$ for any $T \in \mathscr{S}$. For any other $i \in [T/a, b]$ we have

$$\|\varphi(T(i-1)) - \varphi(T(i))\|^\beta < \|\varphi(T(i-1)) - \varphi(T(i))\|^\alpha .$$

Hence,

$$V_\beta(a, b) < V_\alpha(a, b) + n(\text{diam } C)^\beta < \infty .$$

## 4. Integration of 1-forms.

In this section, we first define the kind of differential form we shall be dealing with. Our definition is a variant of Whitney's definition of a function $m$ times differentiable on a closed set [3]. Next, we choose a special sequence of subdivisions and proceed to define the integral of the form over the curve $C$ by taking sums of polynomials of degree $m$ and then passing to the limit. Under conditions involving the generalized variation $V_\alpha$, we show that the integral exists and possesses, in particular, the properties of linearity and 'anti-derivative'.

Throughout this section, $m$ is a positive integer, $\eta \geq 0$, $K > 0$.

4.1. *The Differential Form.* Let

$$\sigma k = \sum_{i=0}^{N} k_i \text{ for any } (N+1)\text{-tuple } k .$$

A differential 1-form $\omega$ on $C$ is a function on the set of all $(N+1)$-tuples $k$, for which $k_i$ is a non-negative integer for $i = 0, \cdots, N$ and $1 \leq \sigma k \leq m$, to the set of real-valued functions on $C$ such that

$$\omega_k(y) = \sum_{\sigma j=0}^{m-\sigma k} \omega_{k+j}(x) \frac{(y_0 - x_0)^{j_0} \cdots (y_N - x_N)^{j_N}}{j_0! \cdots j_N!} + R_k(x, y)$$

where

$$|R_k(x, y)| < K\|x - y\|^{m+\eta-\sigma k} \text{ for } x \in C \text{ and } y \in C .$$

It is important to note that, in case $m = 1$ and $\eta > 0$, $\omega$ is a differential form on $C$ satisfying a Hölder condition. If however $m > 1$, then $\omega$ is also a closed differential form on $C$, that is, $d\omega = 0$ on $C$.

By taking $m = 1$ and $\eta = 1$, we get the sharp forms considered by Whitney. The conditions we impose on $C$, however, are quite different and, we feel, in practice easier to check than those obtained in [4].

4.2. *The sequence of subdivisions.* We define first, for each $(n+1)$-tuple of non-negative integers $(s_0, \cdots, s_n)$, a point $t(s_0, \cdots, s_n)$ by recursion on $n$ and on $s_n$. These will be the end-points of the $n$th subdivision of the unit interval.

4.2.1. DEFINITION. $t(0) = 0, \qquad t(1) = 1\,,$

$$t(s_0, \cdots, s_n, 0) = t(s_0, \cdots, s_n)\,,$$

$$t(s_0, \cdots, s_n, j+1) = \sup\{u : t(s_0, \cdots, s_n, j) \leq u \leq t(s_0, \cdots, s_n + 1)$$

$$\text{and } \|\varphi(u') - \varphi(t(s_0, \cdots, s_n, j))\| \leq \frac{1}{2^{n+1}} \text{ for } t(s_0, \cdots, s_n, j) \leq u' \leq u\}$$

for any non-negative integers $n$ and $j$.

We shall denote by $T$ the sequence of subdivisions of the unit interval such that:

$$\text{range } T_n = \{u : u = t(s_0, \cdots, s_n) \text{ for some } n\text{-tuple } (s_0, \cdots, s_n)\}\,.$$

4.2.2. LEMMA. *For any non-negative integers $n$ and $j$, we have*

$$t(s_0, \cdots, s_n) \leq t(s_0, \cdots, s_n, j) \leq t(s_0, \cdots, s_n + 1)\,.$$

4.2.3. LEMMA. *For any positive integer $n$, $i \in [T_n/0, 1]$, $j \in [T_{n-1}/0, 1]$ we have: $T_{n+1}$ is a refinement of $T_n$,* i.e. $\text{range } T_n \subset \text{range } T_{n+1}$;

*if*

$$T_n(i-1) \leq u \leq T_n(i)\,,$$

then

$$\|\varphi(T_n(i-1)) - \varphi(u)\| \leq \frac{1}{2^n}\,;$$

*if*

$$T_{n-1}(j-1) \leq T_n(i-1) < T_n(i) < T_{n-1}(j)\,,$$

*then*

$$\|\varphi(T_n(i-1)) - \varphi(T_n(i))\| = \frac{1}{2^n}\,.$$

4.2.4 LEMMA. *If $F(x, y)$ is a real number whenever $0 \leq x \leq y \leq 1$, $a \in$ range $T_n$, $b \in$ range $T_n$, and $a \leq b$, then*

$$\sum_{i \in [T_{n+1}/a, b]} F(T_{n+1}(i-1), T_{n+1}(i)) = \sum_{j \in [T_n/a, b]} \; \sum_{i \in [T_{n+1}/T_n(j-1), T_n(j)]} F(T_{n+1}(i-1), T_{n+1}(i))\,.$$

4.3. *The integral of* $\omega$. First, we define $\int_b^a \omega d\varphi$ as the limit of certain sums of polynomials.

4.3.1. *Definitions.*

$$
\begin{aligned}
P'(x, y) &= \sum_{\sigma k=1}^{m} \omega_k(x) \frac{(y_0 - x_0)^{k_0} \cdots (y_N - x_N)^{k_N}}{k_0 ! \cdots k_N !} , \\
P(a, b) &= P'(\varphi(a), \varphi(b)), \\
S_n(a, b) &= \sum_{i \in [T_n/a, b]} P(T_n(i-1), T_n(i)) , \\
\int_a^b \omega d\varphi &= \lim_{n \to \infty} S_n(a, b) .
\end{aligned}
$$

Next, in order to prove the existence of $\int_a^b \omega d\varphi$ and some of its properties under conditions involving $V_\alpha(a, b)$ for some $\alpha < m + \eta$, we introduce the following.

4.3.2. *Definitions.*

$$R(x, y, z) = P'(x, y) + P'(y, z) - P'(x, z) .$$

$$M = K \sum_{\sigma k=1}^{m} \frac{1}{k_0 ! \cdots k_N !} .$$

$$\beta = m + \eta .$$

4.3.3. LEMMA. *If* $x, y, z \in C$, $\| x - y \| \leq \delta$ *and* $\| y - z \| \leq \delta$, *then*

$$| R(x, y, z) | < M\delta^\beta .$$

*Proof.* Let $h(v) = P'(x, v)$ for $v \in E$. Then, $h$ is a polynomial of degree $m$. Let $\boldsymbol{O}_r = \{k : k$ is an $(N+1)$-tuple of non-negative integers and $1 \leq \sigma k \leq r\}$.
For $k \in \boldsymbol{O}_r$ and $p \in \boldsymbol{O}_r$, let $p \geq k$ iff $p_i \geq k_i$ for $i = 0, \cdots, N$, and let

$$D_k h(v) = \frac{\partial^{\sigma k} h(v)}{\partial^{k_0} v_0 \cdots \partial^{k_N} v_N} ,$$

then

$$D_k h(v) = \sum_{\substack{p \in \boldsymbol{O}_m \\ p \geq k}} \omega_p(x) \frac{(v_0 - x_0)^{p_0 - k_0} \cdots (v_N - x_N)^{p_N - k_N}}{(p_0 - k_0) ! \cdots (p_N - k_N) !} .$$

Hence, by Taylor's formula

$$h(z) = h(y) + \sum_{k \in \boldsymbol{O}_m} D_k h(y) \frac{(z_0 - y_0)^{k_0} \cdots (z_N - y_N)^{k_N}}{k_0 ! \cdots k_N !} = h(y) +$$

$$+\sum_{k\in O_m}\left\{\left[\sum_{\substack{p\in O_m\\ p\geq k}}\omega_p(x)\frac{(y_0-x_0)^{p_0-k_0}\cdots(y_N-x_N)^{p_N-k_N}}{(p_0-k_0)!\cdots(p_N-k_N)!}\right]\right.$$
$$\left.\cdot\frac{(z_0-y_0)^{k_0}\cdots(z_N-y_N)^{k_N}}{k_0!\cdots k_N!}\right\}.$$

On the other hand from 4.3.1 and 4.1 we have

$$P'(y,z)=\sum_{k\in O_m}\left\{\left[\omega_k(x)+\sum_{j\in O_{m-\sigma k}}\omega_{k+j}(x)\frac{(y_0-x_0)^{j_0}\cdots(y_N-x_N)^{j_N}}{j_0!\cdots j_N!}+R_k(x,y)\right]\right.$$
$$\left.\cdot\frac{(z_0-y_0)^{k_0}\cdots(z_N-y_N)^{k_N}}{k_0!\cdots k_N!}\right\}$$
$$=\sum_{k\in O_m}\left\{\left[\sum_{\substack{k\in O_m\\ p\geq k}}\omega_p(x)\frac{(y_0-x_0)^{p_0}-{}^{k_0}\cdots(y_N-x_N)^{p_N-k_N}}{(p_0-k_0)!\cdots(p_N-k_N)!}+R_k(x,y)\right]\right.$$
$$\left.\cdot\frac{(z_0-x_0)^{k_0}\cdots(z_N-y_N)^{k_N}}{k_0!\cdots k_N!}\right]\bigg\}$$
$$=h(z)-h(y)+\sum_{k\in O_m}R_k(x,y)\frac{(z_0-y_0)^{k_0}\cdots(z_N-y_N)^{k_N}}{k_0!\cdots k_N!}.$$

Making use of the condition on $R_k(x, y)$ stated in 4.1, we get

$$|P'(x,y)+P'(y,z)-P'(x,z)|<\sum_{k\in O_m}\frac{K\|y-x\|^{\beta-\sigma k}\|z-y\|^{\sigma k}}{k_0!\cdots k_N!}\leq M\delta^{\beta}.$$

4.3.4 LEMMA. *Suppose* $\|x(0)-x(i)\|\leq A$ and $\|x(i-1)-x(i)\|\leq A$ *for* $i=1,\cdots,p$, *whereas* $\|x(i-1)-x(i)\|=A/r$ *for* $i=1,\cdots,p-1$, *where all* $x(i)\in C$. *Then*

$$\left|\sum_{i=1}^{p}P'(x(i-1),x(i))-P'(x(0),x(p))\right|<Mr^{\alpha}A^{\beta-\alpha}\sum_{i=1}^{p}\|x(i-1)-x(i)\|^{\alpha}.$$

*Proof.* $\left|\sum_{i=1}^{p}P'(x(i-1),x(i))-P'(x(0),x(p))\right|$

$$\leq\sum_{i=2}^{p}|P'(x(0),x(i-1))+P'(x(i-1),x(i))-P'(x(0),x(i))|$$
$$=\sum_{i=2}^{p-1}|R(x(0),x(i-1),x(i))|<(p-1)MA^{\beta}=(p-1)Mr^{\alpha}A^{\beta-\alpha}\left(\frac{A}{r}\right)^{\alpha}$$
$$=Mr^{\alpha}A^{\beta-\alpha}\sum_{i=1}^{p-1}\|x(i-1)-x(i)\|^{\alpha}\leq Mr^{\alpha}A^{\beta-\alpha}\sum_{i=1}^{p}\|x(i-1)-x(i)\|^{\alpha}.$$

4.3.5 LEMMA. *Let* $n>1$, $a\in\text{range}\,T_n$, $b\in\text{range}\,T_n$, $a\leq b$, $[T_{n-1}/a,b]=0$. *Then* $|S_n(a,b)-P(a,b)|<M5^{\beta}V_{\beta}(a,b)$.

*Proof.* Let

$$a'=\sup\{u:u\in\text{range}\,T_{n-1}\text{ and }u\leq a\}$$
$$b'=\sup\{u:u\in\text{range}\,T_{n-1}\text{ and }u\leq b\}.$$

First, suppose $a \leq b' \leq b$. Then $a' < a$ and, by 4.2.3

$$\| \varphi(u) - \varphi(a') \| \leq \frac{1}{2^{n-1}} \qquad \text{for } a' \leq u \leq b'$$

$$\| \varphi(u) - \varphi(b') \| \leq \frac{1}{2^{n-1}} \qquad \text{for } b' \leq u \leq b .$$

Hence

$$\| \varphi(T_n(i)) - \varphi(a) \| \leq \frac{2}{2^{n-1}} \qquad \text{for } i \in [T_n/a, b] ,$$

$$\| \varphi(T_n(i)) - \varphi(b') \| \leq \frac{1}{2^{n-1}} \qquad \text{for } i \in [T_n/b', b] ,$$

$$\| \varphi(T_n(i-1)) - \varphi(T_n(i)) \| = \frac{1}{2^n} \qquad \text{for } i \in [T_n/a, b], T_n(i) \neq b', T_n(i) \neq b .$$

Replacing $\alpha$ by $\beta$ in 4.3.4 and using 4.3.3 and 3.1, we see that

$$\begin{aligned} &| S_n(a, b) - P(a, b) | = | S_n(a, b') + S_n(b', b) - P(a, b) | \\ &\leq | S_n(a, b') - P(a, b') | + | S_n(b', b) - P(b', b) | + | P(a, b') + P(b', b) - P(a, b) | \\ &< M4^\beta V_\beta(a, b') + M2^\beta V_\beta(b', b) + M V_\beta(a, b) \leq M5^\beta V_\beta(a, b) . \end{aligned}$$

Next suppose $b' < a$. Then, for $i \in [T_n/a, b]$,

$$\| \varphi(T_n(i)) - \varphi(a) \| \leq \frac{2}{2^{n-1}} ,$$

$$\| \varphi(T_n(i - 1)) - \varphi(T_n(i)) \| = \frac{1}{2^n} .$$

Hence, by 4.3.4,

$$| S_n(a, b) - P(a, b) | < M4^\beta V_\beta(a, b) .$$

4.3.6 LEMMA. *Let $a \in \text{range } T_n$, $b \in \text{range } T_n$, $a < b$. Then,*

$$| S_{n+1}(a, b) - S_n(a, b) | < M2^\alpha V_\alpha(a, b)\left(\frac{1}{2^{\beta-\alpha}}\right)^n .$$

*Proof*. Using 4.2.4, 4.2.3 and 4.3.4, we see that

$$\begin{aligned} &| S_{n+1}(a, b) - S_n(a, b) | \\ &= \left| \sum_{j \in [T_n/a, b]} \left[ \sum_{i \in [T_{n+1}/T_n(j-1), T_n(j)]} P(T_{n+1}(i - 1), T_{n+1}(i)) - P(T_n(j - 1), T_n(j)) \right] \right| \\ &< \sum_{j \in [T_n/a, b]} \left[ M2^\alpha \left(\frac{1}{2^n}\right)^{\beta-\alpha} \sum_{i \in [T_{n+1}/T_n(j-1), T_n(j)]} \| \varphi(T_{n+1}(i - )) - \varphi(T_{n+1}(i)) \|^\alpha \right] \\ &= M2^\alpha \left(\frac{1}{2^n}\right)^{\beta-\alpha} \sum_{i \in [T_{n+1}/a, b]} \| \varphi(T_{n+1}(i-1)) - \varphi(T_{n+1}(i)) \|^\alpha \leq M2^\alpha V_\alpha(a, b)\left(\frac{1}{2^{\beta-\alpha}}\right)^n . \end{aligned}$$

4.3.7. THEOREM. *If $0 \leq a \leq b \leq 1$, $\alpha < \beta$, $V_\alpha(a, b) < \infty$, then*

$$\left|\int_n^b \omega d\varphi\right| < \infty .$$

*Proof.* Let

$$a'_n = \inf\{u : u \in \text{range } T_n \text{ and } a \leq u\} ,$$
$$b'_n = \sup\{u : u \in \text{range } T_n \text{ and } u \leq b\} .$$

If $a = b$, the theorem is trivial. If $a < b$, for $n$ sufficiently large, we have

$$a \leq a'_{n+1} \leq a'_n \leq b'_n \leq b'_{n+1} \leq b ,$$
$$[T_n/a, a'_n] = 0 \qquad \text{and } [T_n/b'_n, b] = 0 ,$$
$$||\varphi(a'_{n+1})\varphi - (a'_n)|| \leq \frac{2}{2^n} \qquad \text{and } ||\varphi(b'_n) - \varphi(b'_{n+1})|| \leq \frac{1}{2^n} .$$

Hence

$$|S_{n+1}(a, b) - S_n(a, b)| = |S_{n+1}(a'_{n+1}, b'_{n+1}) - S_n(a'_n, b'_n)|$$
$$= |S_{n+1}(a'_{n+1}, a'_n) + S_{n+1}(a'_n, b'_n) + S_{n+1}(b'_n, b'_{n+1}) - S_n(a'_n, b'_n)|$$
$$\leq |S_{n+1}(a'_{n+1}, a'_n) - P(a'_{n+1}, a'_n)| + |S_{n+1}(a'_n, b'_n) - S_n(a'_n, b'_n)|$$
$$+ |S_{n+1}(b'_n, b'_{n+1})| - P(b'_n, b'_{n+1})| + |P(a'_{n+1}, a'_n)| + 1P(b'_n, b'_{n+1}| < \text{(by 4.3.5, 4.3.6)}$$
$$< M5^\beta V_\beta(a'_{n+1}, a'_n) + M2^\alpha V_\alpha(a'_n, b'_n)\left(\frac{1}{2^{\beta-\alpha}}\right)^n + M5^\beta V_\beta(b'_n, b'_{n+1}) + M'\frac{2}{2^n} + M'\frac{1}{2^n} ,$$

where

$$M' = \sup_{\substack{x \in C \\ 1 \leq \sigma k < m}} |\omega_k(x)| \sum_{\sigma k=1}^{m} \frac{1}{k_0 ! \cdots k_N !} .$$

Therefore, for any positive integer $p$ we have

$$|S_{n+p}(a, b) - S_n(a, b)| \leq \sum_{q=0}^{p-1} |S_{n+q+1}(a, b) - S_{n+q}(a, b)|$$
$$< M5^\beta \sum_{q=0}^{\infty} [V_\beta(a'_{n+q+1}, a'_{n+q}) + V_\beta(b'_{n+q}, b'_{+n+q+1})] + M2^\alpha V_\alpha(a, b) \sum_{q=0}^{\infty} \left(\frac{1}{2^{\beta-\alpha}}\right)^{n+q}$$
$$+ 3M' \sum_{q=0}^{\infty} \frac{1}{2^{n+q}} < M5^\beta(V_\beta(a, a'_n) + V_\beta(b'_n, b)) + M\frac{2^\beta}{2^{\beta-\alpha}_{-1}} V_\alpha(a, b)\left(\frac{1}{2^{\beta-\alpha}}\right)^n + \frac{6M'}{2^n} .$$

Since, by 3.2, $V_\beta(a, b) < \infty$, with the help of 3.1 we see that $V_\beta(a, a'_n) \to 0$ and $V_\beta(b'_n, b) \to 0$ as $n \to \infty$. Thus, the $S_n(a, b)$ form a Cauchy sequence and $\left|\int_a^b \omega d\varphi\right| < \infty$.

4.3.8. THEOREM. *Suppose* $\delta > 0$, $\alpha < \beta$, $L < \infty$, $||\varphi(a) - \varphi(b)|| < 1$, and

$$V_\alpha(a, b) < L\,||\varphi(a) - \varphi(b)||^\alpha$$

*whenever* $0 \leq a \leq b \leq 1$ *and* $b - a < \delta$. *Then, for some* $M' < \infty$,

$$\left|\int_a^b \omega d\varphi - P(a, b)\right| < M' \| \varphi(a) - \varphi(b) \|^\alpha$$

*whenever* $0 \le a \le b \le 1$ and $b - a < \delta$.

*Proof.* Given $0 \le a \le b \le 1$ and $b - a < \delta$, let

$$a'_q = \inf\{u : u \in \text{range } T_q \text{ and } a \le u\} ,$$
$$b'_q = \sup\{u : u \in \text{range } T_q \text{ and } u \le b\} ;$$

and let $n$ be the integer such that $[T_{n-1}/a, b] = 0$, $[T_n/a, b] \neq 0$.

Given $\varepsilon > 0$, we can choose $p$ so that

$$\left|\int_a^b \omega d\varphi - S_{n+p}(a'_{n+p}, b'_{n+p})\right| < \varepsilon$$

and

$$| P(a, b) - P('_{n+p}, b'_{n+p}) | < \varepsilon$$

and

$$| \| \varphi(a) - \varphi(b) \| - \| \varphi(a'_{n+p}) - \varphi(b'_{n+p}) \| | < \varepsilon .$$

Hence we need only to show that

$$| S_{n+p}(a'_{n+p}, b'_{n+p}) - P(a'_{n+p}, b'_{n+p}) | < M' \| \varphi(a'_{n+p}) - \varphi(b'_{n+p}) \|^\alpha$$

for some $M' < \infty$ and all positive integers $p$.

We can check that

$$\begin{aligned}
&| S_{n+p}(a'_{n+p}, b'_{n+p}) - P(a'_{n+p}, b'_{n+p}) | \\
&\le | S_n(a'_n, b'_n) - P(a'_n, b'_n) | + |P(a'_{n+p}, a'_n) + P(a'_n, b'_n) - P(a'_{n+p}, b'_n) | \\
&+ | P(a'_{n+p}, b'_n) + P(b'_n, b'_{n+p}) - P(a'_{n+p}, b'_{n+p}) | \\
&+ \sum_{k=0}^{p-1} \{ | P(a'_{n+p}, a'_{n+k+1}) + P((a'_{n+k+1}, a'_{n+k}) - P(a'_{n+p}, a'_{n+k}) | \\
&+ | P(b'_{n+k}, b'_{n+k+1}) + P(b'_{n+k+1}, b'_{n+p}) - P(b'_{n+k}, b'_{n+p}) | \\
&+ | S_{n+k+1}(a'_{n+k+1}, a'_{n+k}) - P(a'_{n+k+1}, a'_{n+k}) | \\
&+ | S_{n+k+1}(b'_{n+k}, b'_{n+k+1}) - P(b'_{n+k+1}, b'_{n+k}) | \\
&+ | S'_{n+k+1}(a'_{n+k}, b'_{n+k}) - S_{n+k}(a'_{n+k}, b'_{n+k}) | \} .
\end{aligned}$$

Now, we observe that

$$\| \varphi(u) - \varphi(v) \| \le \frac{2}{2^{n+k}} \qquad \text{for } a'_{n+p} \le u \le v \le a'_{n+k} ,$$

$$\| \varphi(u) - \varphi(v) \| \le \frac{1}{2^{n+k}} \qquad \text{for } b'_{n+k} \le u \le v \le b'_{n+p} ,$$

$$[T_{n+k}/a'_{n+k+1}, a'_{n+k}] = 0 ,$$
$$[T_{n+k}/b'_{n+k}, b'_{n+k+1}] = 0 .$$

Hence by 4.3.5, 4.3.3, 4.3.6 we have

$$
\begin{aligned}
&| S'_{n+p}(a'_{n+p}, b'_{n+p}) - P(a'_{n+p}, b'_{n+p}) | \\
&< M5^{\beta} V_{\beta}(a'_n, b'_n) + MV_{\beta}(a'_{n+p}, b'_n) + MV_{\beta}(a'_{n+p}, b'_{n+p}) \\
&+ M \sum_{k=0}^{p-1} \left\{ V_{\alpha}(a'_{n+p}, a'_{n+k}) \left( \frac{2}{2^{n+k}} \right)^{\beta-\alpha} + V_{\alpha}(b'_{n+k}, b'_{n+p}) \left( \frac{1}{2^{n+k}} \right)^{\beta-\alpha} \right. \\
&\left. + 5^{\beta} V_{\beta}(a'_{n+k+1}, a'_{n+k}) + 5^{\beta} V_{\beta}(b'_{n+k}, b'_{n+k+1}) + 2^{\alpha} V_{\alpha}(a'_{n+k}, b'_{n+k}) \left( \frac{1}{2^{\beta-\alpha}} \right)^{n+k} \right\} \\
&< M5^{\beta} V_{\beta}(a'_{n+p}, b'_{n+p}) + 2MV_{\beta}(a'_{n+p}, b'_{n+p}) \\
&+ MV_{\alpha}(a'_{n+p}, b'_{n+p})(2^{\beta-\alpha} + 1 + 2^{\alpha}) \sum_{k=0}^{\infty} \left( \frac{1}{2^{\beta-\alpha}} \right)^{n+k} \\
&< MV_{\alpha}(a'_{n+p}, b'_{n+p}) \left[ 5^{\beta} + 2 + (2^{\beta-\alpha} + 1 + 2^{\alpha}) \sum_{k=0}^{\infty} \left( \frac{1}{2^{\beta-\alpha}} \right)^{n+k} \right] \\
&< M' \| \varphi(a'_{n+p}) - \varphi(b'_{n+p}) \|^{\alpha}
\end{aligned}
$$

where

$$M' = ML\left[ 5^{\beta} + 2 + (2^{\beta-\alpha} + 1 + 2^{\alpha}) \sum_{k=0}^{\infty} \left( \frac{1}{2^{\beta-\alpha}} \right)^{n+k} \right] < \infty .$$

4.3.9. THEOREM. *If* $0 \leq a \leq b \leq c \leq 1, \left| \int_a^b \omega d\varphi + \int_b^c \omega d\varphi \right| < \infty$, then

$$\int_a^c \omega d\varphi = \int_a^b \omega d\varphi + \int_b^c \omega d\varphi .$$

*Proof.* Let

$$
\begin{aligned}
a'_n &= \sup\{u : u \in \text{range}\, T_n \text{ and } u \leq b\} \\
b'_n &= \inf\, \{u : u \in \text{range}\, T_n \text{ and } b \leq u\} .
\end{aligned}
$$

We have $\lim_{n\to\infty} P(a'_n, b'_n) = 0$ and for sufficiently large $n$

$$S_n(a, c) = S_n(a, b) + P(a'_n, b'_n) + S_n(b, c) .$$

Taking the limit on both sides we get the desired result.

4.3.10. REMARK. If $\omega$ and $\omega'$ are both 1-forms in the sense of 4.1, then so is $(\omega + \omega')$ and

$$\int_a^b (\omega + \omega') d\varphi = \int_a^b \omega d\varphi + \int_a^b \omega' d\varphi$$

provided the right hand side is bounded. This is an immediate consequence of the definitions.

## REFERENCES

1. G. Glaeser, *Etudes de quelques algèbres tayloriennes*, J. d'Analyse Math., Jerusalem, **6**, (1958), 1–124.

2. M. Sion, *On the existence of functions having given partial derivatives on a curve,* Trans. Am. Math. Soc. **77** (1954), 179-201.
3. H. Whitney, *Analytic extensions of differentiable functions defined on closed sets,* Trans. Am. Math. Soc. **36** (1934), 63-89.
4. ———, *Geometric integration theory,* Princeton Univ. Press, 1957.

UNIVERSITY OF CALIFORNIA

# SUBDIRECT SUMS AND INFINITE ABELIAN GROUPS

ELBERT A. WALKER

**1. Definitions.** Let $G$ be a group, and suppose $G$ is a subgroup of the direct sum $\sum_{a \in I} \oplus H_a$ of the collection of groups $\{H_a\}_{a \in I}$. If the projection of $G$ into $H_a$ is onto $H_a$ for each $a \in I$, then $G$ is said to be a *subdirect sum* of the groups $\{H_a\}_{a \in I}$. (Only weak direct and subdirect sums are considered here.) If a group $G$ is isomorphic to a subdirect sum of the groups $\{H_a\}_{a \in I}$, then $G$ is said to be *represented* as a subdirect sum of the groups $\{H_a\}_{a \in I}$. A group is called a *rational group* if it is a subgroup of a $Z(p^\infty)$ group or a subgroup of the additive group of rational numbers.

2. THEOREM. *Every Abelian group can be represented as a subdirect sum of rational groups where the subdirect sum intersects each of the rational groups non-trivially.*

*Proof.* $G$ is isomorphic to a subgroup of some divisible group, and thus can be represented as a subdirect sum $G'$ of rational group $\{H_a\}_{a \in I}$. Let $(h_1, h_2, \cdots, h_a, \cdots)$ be an element of $G'$. Let $(h_1, h_2, \cdots, h_a, \cdots)\beta_1 = (k_1, h_2, \cdots, h_a, \cdots)$, where $k_1 = h_1$ if $G' \cap H_1 \neq 0$, and $k_1 = 0$ if $G' \cap H_1 = 0$. Assume $\beta_c$ has been defined for $c < b$. Define

$$(h_1, h_2, \cdots, h_a, \cdots)\beta_b = (k_1, k_2, \cdots, k_b, h_{b+1}, \cdots)$$

where $k_b = h_b$ if $H_b \cap (\bigcup_{c<b} G'\beta_c) \neq 0$, and $k_b = 0$ otherwise. Each $\beta_a$ preserves addition because each is a projection. Let $(h_1, h_2, \cdots, h_a, \cdots) \neq (0, 0, \cdots, 0, \cdots)$ and let

$$(h_1, h_2, \cdots, h_a, \cdots)\beta_a = (k_1, k_2, \cdots, k_a, h_{a+1}, h_{a+2}, \cdots) .$$

Only a finite number of the coordinates of $(h_1, h_2, \cdots, h_a, \cdots)$ are not 0. Let them be $h_{a_1}, h_{a_2}, \cdots, h_{a_n}$, where $a_1 < a_2 < \cdots < a_n$. If $a < a_n$, then

$$\begin{aligned}(h_1, h_2, \cdots, h_a, \cdots)&\beta_a \\ &= (k_1, k_2, \cdots, k_a, h_{a+1}, \cdots, h_{a_n}, h_{a_n+1}, \cdots) \neq (0, 0, \cdots, 0, \cdots)\end{aligned}$$

since $h_{a_n} \neq 0$. Assume $a \geqq a_n$. If $n=1$ and $a_1=1$, then $(h_1, h_2, \cdots, h_a, \cdots) = (h_{a_1}, 0, 0, \cdots, 0, \cdots) \in G'$ and $G' \cap H_1 \neq 0$ so that $(h_{a_1}, 0, 0, \cdots$ [illegible] $=$ $(h_{a_1}, 0, 0, \cdots, 0, \cdots)$. That is, $k_{a_1} = h_{a_1} \neq 0$, and hence $(h_1, h_2, \cdots$ [illegible] $\neq$ $(0, 0, \cdots, 0, \cdots)$. If $n = 1$ and $a_n \neq 1$, then $(0, 0, \cdots, h_{a_1}, 0, 0, \cdots) \in G'$ and also in $G'\beta_c$ for all $c < a_1$. Thus $H_{a_1} \cap (\bigcup_{c<a_1} G'\beta_c) \neq 0$, and

Received September 19, 1958, in revised form October 20, 1958.

$$(h_1, h_2, \cdots, h_a, \cdots)\beta_a = (0, 0, \cdots, 0, h_{a_1}, 0, 0, \cdots)\beta_a$$
$$= (0, 0, \cdots, 0, h_{a_1}, 0, 0, \cdots)\beta_{a_1} = (0, 0, \cdots 0, h_{a_1}, 0, 0, \cdots)$$
$$\neq (0, 0, \cdots, 0, \cdots) .$$

Assume $n > 1$. If $(h_1, h_2, \cdots, h_a, \cdots)\beta_a = (0, 0, \cdots, 0, \cdots)$, then $k_c = 0$ for $c \leqq a_n$, and

$$(h_1, h_2, \cdots, h_a, \cdots)\beta_{a_{n-1}} = (0, 0, \cdots, 0, h_{a_n}, 0, 0, \cdots) .$$

Therefore $H_{a_n} \cap (G'\beta_{a_{n-1}}) \neq 0$, and so $H_{a_n} \cap (\bigcup_{c<a} G'\beta_c) \neq 0$. Hence $k_{a_n} = h_{a_n} \neq 0$, and this contradicts $k_c = 0$ for $c \leqq a_n$. Therefore

$$(h_1, h_2, \cdots, h_a, \cdots)\beta_a \neq (0, 0, \cdots, 0, \cdots) ,$$

and the kernel of $\beta_a$ is 0. Hence each $\beta_a$ is an isomorphism. Now let $(h_1, h_2, \cdots, h_a, \cdots)\beta = (k_1, k_2, \cdots, k_a, \cdots)$. Clearly $\beta$ is a homomorphism of $G'$ into $\sum_{a \in I} \oplus H_a$. But the kernel of $\beta$ is 0 because every element in $G'$ has only a finite number of non-zero coordinates. Let $I'$ be the set of indices such that $a \notin I'$ implies that the image of the projection of $G'\beta$ into $H_a$ is 0. $G'\beta$ is isomorphic to a subdirect sum of the groups $\{H_a\}_{a \in I'}$. If $G'\beta \cap H_1 = 0$, then for $(h_1, h_2, \cdots, h_a, \cdots) \in G'$ we have $(h_1, h_2, \cdots, h_a, \cdots)\beta_1 = (0, h_2, \cdots, h_a, \cdots)$, so that

$$(h_1, h_2, \cdots, h_a, \cdots)\beta = (0, k_2, k_3, \cdots, k_a, \cdots) .$$

Hence the image of the projection of $G'\beta$ into $H_1$ is 0. Therefore $1 \notin I'$. Let $a > 1$. Suppose $G'\beta \cap H_a = 0$ and $H_a \cap (\bigcup_{c<a} G'\beta_c) \neq 0$. Then there exists $b<a$ such that $H_a \cap G'\beta_b \neq 0$. Let $(0, 0, \cdots, 0, k_a, 0, 0, \cdots) \in H_a \cap G'\beta_b$, where $k_a \neq 0$. Let $(h_1, h_2, \cdots, h_a, \cdots)\beta_b = (0, 0, \cdots, 0, k_a, 0, 0, \cdots)$. Then $(h_1, h_2, \cdots, h_a, \cdots)\beta = (0, 0, \cdots, 0, k_a, 0, 0, \cdots)$, and so $G'\beta \cap H_a \neq 0$. Therefore if $G'\beta \cap H_a = 0$, then $H_a \cap (\bigcup_{c<a} G'\beta_c) = 0$. This implies for every $(h_1, h_2, \cdots h_a, \cdots) \in G'$ that

$$(h_1, h_2, \cdots, h_a, \cdots)\beta_a = (k_1, k_2, \cdots, k_a, h_{a+1}, h_{a+2}, \cdots) ,$$

where $k_a = 0$, and hence that

$$(h_1, h_2, \cdots, h_a, \cdots)\beta = (k_1, k_2, \cdots, 0, k_{a+1}, k_{a+2}, \cdots) .$$

Thus the image of the projection of $G'\beta$ into $H_a$ is 0 so that $a \notin I'$. Hence for $a \in I'$, $G'\beta \cap H_a \neq 0$. Since $G$ is isomorphic to $G'\beta$, the theorem follows.

3. Remarks. Theorem 9 in [1] is an immediate corollary of the preceding theorem, as are some other known theorems in Abelian group theory. In [2], Scott proves that every uncountable Abelian group $G$ has, for every possible infinite index $\alpha$, $2^{o(G)}$ subgroups of order equal to $o(G)$ and of index $\alpha$, and that for each given infinite index, their intersection is 0. The following theorem shows that if $G$ is torsion free, one can say more.

4. THEOREM. *Every torsion free Abelian group $G$ of infinite rank has, for every possible infinite index $\alpha$, $2^{o(G)}$ pure subgroups of order equal to $o(G)$ and of index $\alpha$. Furthermore, the intersection of these pure subgroups of index $\alpha$ is 0.*

*Proof.* Represent $G$ as a subdirect sum $G'$ of rational groups $\{H_a\}_{a\in I}$ such that for each $a \in I$, $G' \cap H_a \neq 0$. Let $\alpha$ be an infinite cardinal such that $\alpha \leqq o(G)$. $o(I) = o(G)$ since $G$ has infinite rank. Let $I = S_1 \cup S_2$ where $o(S_1) = \alpha$, $o(S_2) = o(G)$, and $S_1 \cap S_2 = \phi$. Let $T$ be a subset of $S_2$ such that $o(S_2 - T) = o(G)$. There are $2^{o(G)}$ such $T$'s. Let $(h_1, h_2, \cdots, h_a, \cdots)$ be in $G'$, and let

$$(h_1, h_2, \cdots, h_a, \cdots)t = \left(\sum_{j\in T} h_j, k_1, k_2, \cdots, k_a, \cdots\right),$$

where $k_i = h_i$ if $i \in S_1$ and $k_i = 0$ otherwise. The mapping $t$ is a homomorphism and the order of its image is equal to $o(S_1)$. That is, the index of the kernel of $t$ is $\alpha$. The order of the kernel of $t$ is equal to $o(G)$ since $o(S_2 - T) = o(G)$, and $G' \cap H_a \neq 0$ for all $a \in I$. Let $T, T' \supseteqq S_2$, $T \neq T'$. Then there is a $j \in T$ such that $j \notin T'$, say. Let $h_j \in G'$, $h_j \neq 0$. Then

$$(0, 0, \cdots, h_j, 0, 0, \cdots)t = (h_j, 0, \cdots).$$

However, $(0, 0, \cdots, h_j, 0, 0, \cdots)t' = (0, 0, 0, \cdots)$. Hence the kernel of $t$ is not the same as the kernel of $t'$. These kernels are pure in $G'$ since the quotient groups are torsion free. Thus $G$ has $2^{o(G)}$ pure subgroups of index $\alpha$, and of order equal to $o(G)$. Suppose $(h_1, h_2, \cdots, h_a, \cdots)$ is in the intersection of all these pure subgroups of index $\alpha$. Then if $b \in S_1$, $h_b = 0$. Hence if $h_c \neq 0$, letting $T = \{c\}$, we have

$$(h_1, h_2, \cdots, h_c, \cdots, h_a, \cdots)t = (h_c, 0, 0, \cdots) \neq 0,$$

which is impossible. Therefore for each $a \in I$, $h_a = 0$, and this shows that the intersection of these subgroups is 0.

5. REMARKS. Every torsion free divisible group $D$ of rank $\alpha$ is a direct sum of $\alpha$ copies of the additive group of rational numbers, and $D$ contains an isomorphic copy of every torsion free Abelian group of rank $\alpha$. The following theorem says that if $\alpha$ is infinite, every torsion free Abelian group of rank $\alpha$ is represented in a special way in $D$.

6. THEOREM. *Every torsion free Abelian group $G$ of infinite rank can be represented as a subdirect sum $G'$ of copies of the additive group of rational numbers, and in such a way that $G'$ intersects each subdirect summand non-trivially.*

*Proof.* Represent $G$ as a subdirect sum $G'$ of the rational groups

$\{H_a\}_{a\in I}$ such that for each $a \in I$, $G' \cap H_a \neq 0$. Suppose first that $G$ has countably infinite rank. That is, suppose $I$ is the set of positive integers. Each $H_a$ is a subgroup of the additive group of rational numbers, since $G$ is torsion free. Let $k_1, k_2, k_3, \cdots$ be a sequence of non-zero rational numbers such that $k_i \in G' \cap H_i$. Let $r_1, r_2, r_3, \cdots$ be the non-zero rational numbers arranged in a sequence. Let $s_i = r_i/k_i$. Let $(h_1, h_2, \cdots, h_n, \cdots)$ be an element of $G'$. Let

$$(h_1, h_2, \cdots, h_n, \cdots)\beta = \left(\sum_{i=1}^{\infty} s_i h_i, \sum_{i=2}^{\infty} s_i h_i, \cdots, \sum_{i=n}^{\infty} s_i h_i, \cdots\right).$$

Since only a finite number of the $h_i$'s are non-zero, for each $k$, $\sum_{i=k}^{\infty} s_i h_i$ is a rational number, and for only a finite number of $k$'s is $\sum_{i=k}^{\infty} s_i h_i$ non-zero.

$$\begin{aligned}
&((h_1, h_2, \cdots, h_n, \cdots) + (g_1, g_2, \cdots, g_n, \cdots))\beta \\
&= (h_1 + g_1, h_2 + g_2, \cdots, h_n + g_n, \cdots)\beta \\
&= \left(\sum_{i=1}^{\infty} s_i(h_i + g_i), \cdots, \sum_{i=n}^{\infty} s_i(h_i + g_i), \cdots\right) \\
&= \left(\sum_{i=1}^{\infty} s_i h_i + \sum_{i=1}^{\infty} s_i g_i, \cdots, \sum_{i=n}^{\infty} s_i h_i + \sum_{i=n}^{\infty} s_i g_i, \cdots\right) \\
&= (h_1, h_2, \cdots, h_n, \cdots)\beta + (g_1, g_2, \cdots, g_n, \cdots)\beta .
\end{aligned}$$

Hence $\beta$ is a homomorphism of $G'$ into a direct sum of copies of the additive group $R$ of rationals. Let $R_n$ be the set of $n$th coordinates of elements of $G'\beta$. $R_n$ is a subgroup of $R$ since it is the image of the projection of $G'\beta$ onto its $n$th coordinates. Let $m \geqq n$.

$$(0, 0, \cdots, 0, k_m, 0, 0, \cdots) \in G'$$

and

$$(0, 0, \cdots, 0, k_m, 0, 0, \cdots)\beta = (r_m, r_m, \cdots, r_m, 0, 0, \cdots) ,$$

so that $r_m \in R_n$. Thus $R_n$ contains all but at most a finite number of elements of $R$, and being a subgroup of $R$, must then be $R$. Therefore $G'\beta$ is a subdirect sum of copies of $R$. Let $x \in G'$, $x \neq 0$, and let $h_r$ be the last non-zero coordinate of $x$. Then the $r$th coordinate of $x\beta$ is $s_r h_r \neq 0$. Hence the kernel of $\beta$ is 0 and $\beta$ is an isomorphism of $G$ onto a subdirect sum of copies of $R$. Now consider the case where $I$ is not countable. Let $I$ be the union of the set of mutually disjoint countably infinite sets $\{I_j\}_{j\in J}$. Denote by $S_j$ the image of the projection of $G'$ into $\sum_{a\in I_j} \oplus H_a$. Then $G'$ is a subdirect sum of the set of groups $\{S_j\}_{j\in J}$, and each $S_j$ is of countably infinite rank. Hence each $S_j$ may be represented as a subdirect sum of copies of the additive group of rational numbers, and it follows that $G$ may be so represented. In light of the proof of 2, this representation may be assumed to intersect each subdirect summand non-trivially.

## References

1. W. R. Scott, *Groups, and cardinal numbers*, Amer. J. Math. **74** (1952), 187–197.
2. W. R. Scott, *The number of subgroups of given index in nondenumerable Abelian groups*, Proc. Amer. Math. Soc. **5** (1954), 19–22.

University of Kansas and
New Mexico State University

# HOMOGENEOUS STOCHASTIC PROCESSES*

JOHN W. WOLL, JR.

**Summary.** The form of a stationary translation-invariant Markov process on the real line has been known for some time, and these processes have been variously characterized as infinitely divisible or infinitely decomposable. The purpose of this paper is to study a natural generalization of these processes on a homogeneous space $(X, G)$. Aside from the lack of structure inherent in the very generality of the spaces $(X, G)$, the basic obstacles to be surmounted stem from the presence of non trivial compact subgroups in $G$ and the non commutativity of $G$, which precludes the use of an extended Fourier analysis of characteristic functions, a tool which played a dominant role in the classical studies. Even in the general situation there is a striking similarity between homogeneous processes and their counterparts on the real line.

A homogeneous process is a process in the terminology of Feller [3] on a locally compact Hausdorff space $X$, whose transition probabilities $P(t, x, dy)$ are invariant under the action of elements $g \in G$ of a transitive group of homeomorphisms of $X$, in the sense that $P(t, g[x], g[dy]) = P(t, x, dy)$. It is shown that if every compact subset of $X$ is separable or $G$ is commutative the family of measures $t^{-1}P(t, x, \cdot)$ converges to a not necessarily bounded Borel measure $Q_x(\cdot)$ on $X$-$\{x\}$ as $t \to 0$, meaning that for every bounded continuous, complex valued function $f$ on $X$ which vanishes in a neighborhood of $x$ and is constant at infinity $t^{-1}P(t, x, f) \to Q_x(f)$.

In 3 we show that the paths of a separable homogeneous process are bounded on every bounded $t$-interval and have right and left limits at every $t$ with probability one. If the action of $G$ on $X$ is used to translate the origin of each jump to $x$, it is shown for suitably regular compact sets $D$ that the probability of a jump into $D$ while $t \in [0, T]$ is given by 1–exp $\{-TQ_x(D)\}$. The maps $f \to P(t, \cdot, f) = (T_t f)(\cdot)$ map the Banach space, $C(X)$, of continuous functions generated by the constants and functions with compact support into itself, and by a suitable normalization can be assumed strongly continuous for $t \geq 0$. Indeed, $T_t$ is a strongly continuous semi-group. The domain $D(A)$ of the infinitesimal generator $A$ of $T_t$ admits a smoothing operation whose precise

* This paper was originally accepted by the Trans. Amer. Math. Soc.; received by the editors of the Trans. Amer. Math. Soc. November 5, 1956, in revised form February 24, 1958. This paper is a thesis, submitted in partial fulfillment of the requirements for the Ph. D. degree, Princeton University, June 1956. During its preparation the author was a National Science Foundation Predoctoral Fellow.

nature is described in Corollary 1 of Theorem 2.2. Roughly speaking, if $g \in D(A), \varepsilon > 0$, and $f \in C(X)$ we can find an $h \in D(A)$ such that $\|h - f\|_\infty < \varepsilon$ and $f = h$ on any preassigned compact subset of interior $(\{x \mid f(x) = g(x)\})$.

A family of measures $P(x, A)$ which is a probability measure on the Borel subsets of $X$ for fixed $x \in X$, measurable in $x$ for fixed $A$, and invariant under $G$(i.e. $P(g[x], g[A]) = P(x, A)$), is called a homogeneous transition probability of norm one. Such a family generates a continuous endomorphism $f \to P(\cdot, f) = (Pf)(\cdot)$ of $C(X)$, and a homogeneous process $T_t = \exp\{tr(P-1)\}$. This latter process is called a compound Poisson process. In 4 we study strong convergence for compound Poisson processes (Definition 4.1) and prove among other facts that every homogeneous process is a strong limit of compound Poisson processes $\exp\{tr_i(P_i - 1)\}$, and if $r_i \leq M < +\infty$ the limit process is necessarily compound Poisson. If $X$ is given the discrete topology every homogeneous process on $X$ is compound Poisson. In case the $Q_x$ associated with $P(t, x, dy)$ vanishes identically, or equivalently $P(t, x, dy)$ has continuous paths, we show in 5 that $P(t, x, dy)$ is the strong limit of compound Poisson processes whose $P_i(x, dz)$ have support arbitrarily closed to $x$.

In 7 we study subordination of homogeneous processes as defined by Bochner [1]. By phrasing the definition in terms of a probabilistically run clock it is shown that many processes are maximal in the partial order induced by subordination. If we follow the notation in (7.2) where $\exp\{tS(b, A, F, z)\}$ denotes the family of characteristic functions of a homogeneous process $X(t, \omega)$ on Euclidean $n$-space, we obtain the following type of result. When support $(F)$ is compact, $X(t, \omega)$ is not subordinate to any process but itself unless support $(F)$ is also contained in the half line $R^+b$ and $A = 0$. In this latter case $X(t, \omega)$ is subordinate to the Bernoulli process $Z(t, \omega) = tb$. Actually somewhat stronger statements can be made but they are proven only for the real line.

In our notation we have not distinguished between the application of a measure $\mu$ to a function $f$ and the measure of a measurable set $E$, denoting these respectively by $\mu(f)$ and $\mu(E)$. The set $X$–$E$ is denoted by $E^c$ and the usual convention is adopted in letting $C, R, R^+, Z$, and $Z^+$ represent respectively the complex numbers, the reals, the non-negative reals, the integers, and the non-negative integers.

I would like to take this opportunity to express my gratitude to Professor Bochner who has patiently encouraged this work, and whose own ideas are at the base of § 7.

**1. Introduction.** Let $G$ be a Hausdorff, locally compact topological group and $H$ one of its compact subgroups. We consider $G$ as a group of homeomorphisms of the left cosets of $G$ modulo $H$, which we denote by $G/H$, the left coset $xH$ being mapped by $a \in G$ into the left coset $axH$. Moreover, if we are primarily interested in the space $G/H$ and the action of $G$ on this space, there is no reason why the subgroup $H$ should play a dominant role, for the homeomorphism $x \to xb$ of $G$ carries the left coset $xH$ of $H$ into the left coset $xbb^{-1}Hb$ of $b^{-1}Hb$. As far as left multiplication by elements of $G$ is concerned this is an operator homeomorphism and $G/H$ is equivalent to $G/b^{-1}Hb$. For this reason we use the neutral letter $X$ for the space $G/H$ and denote the operation of $a \in G$ on $x \in X$ by $a[x]$. We call the system $(X, G)$ a homogeneous space and note that $X$ is naturally homeomorphic to any coset space of the form $G/G_z$ where $G_z = \{a \in G \mid a[z] = z\}$.

For the sake of exposition let $N(X)$ be the Banach space of regular bounded complex Borel measures on $X$; $C_c(X)$ the linear space of continuous complex valued functions with compact support; $C_\infty(X)$ the closure of $C_c(X)$ in the uniform norm; $C(X)$ the Banach space of functions generated by $C_\infty(X)$ and the constant functions with the uniform norm. We use the current notation of W. Feller and denote linear transformations of $N(X)$ by postmultiplication. In this notation a linear transformation $T$ of $C_\infty(X)$ is denoted by the same letter as its adjoint transformation on $N(X)$, viz. $\mu T(f) = \mu(Tf)$. By the expression $\mu \geq 0$, $\mu \in N(X)$, we mean $\mu$ is a real valued non-negative measure, and by the transformation $L_a$ we refer to the isometries of $C_c(X)$, $C_\infty(X)$, $C(X)$ and $N(X)$ generated by translation of $X$ by $a \in G$, viz. $(L_a f)(x) = f(a^{-1}[x])$, $\mu L_a(E) = \mu(a[E])$. When $x \in X$ we denote by $\delta^x$ the measure placing a unit mass at $x$, so that $\delta^x(E) = 1$ if $x \in E$ and 0 otherwise. For example, we shall often use the relationship $\delta^x L_{a^{-1}} = \delta^{a[x]}$. Finally, when we say that a directed sequence $(\mu_q)_{q \in Q}$ of measures-commonly called a net in $N(X)$-converges weakly to $\mu \in N(X)$, in symbols $\mu_q \to \mu$, we mean for every $f \in C_c(X)$, $\mu_q(f) \to \mu(f)$ as complex numbers.

DEFINITION 1.1. A *homogeneous transition probability* is a continuous endomorphism $P: N(X) \to N(X)$ satisfying:

(i) $\mu \geq 0$ implies $\mu P \geq 0$;
(ii) $\mu_q \to \mu$ implies $\mu_q P \to \mu P$;
(iii) $PL_a = L_a P$.

An endomorphism, $P$, with properties (i) and (ii) is usually called a transition probability and (iii) makes the transition probability homogeneous.

When $f \in C_\infty(X)$ it follows from (ii) that $\delta^x P(f)$ is continuous in $x$. In addition $\delta^x P(f) \in C_\infty(X)$. To show this let $a_i[z]$ be any directed sequence in $X$ tending to infinity, then by virtue of the assumed compactness of $G_z$, $a_i[D] \to$ infinity for any compact set $D \subset X$ and

$L_{a_i^{-1}}f(y) = f(a_i[y]) \to 0$ boundedly and uniformly on every compact set. It follows immediately that

$$\delta^{a_i[z]}P(f) = \delta^z L_{a_i^{-1}}P(f) = \delta^z P(L_{a_i^{-1}}f) \to 0 ,$$

proving $\delta^x P(f) \in C_\infty(X)$. Now for any given $\mu \in N(X)$ let $\mu_q = \sum_i m_{qi}\delta^{x_{qi}}$ be a bounded, directed sequence of purely atomic measures in $N(X)$ approaching $\mu$ weakly, $\mu_q \to \mu$. Then separate calculations show that on the one hand $\mu_q P(f) \to \mu P(f)$, while on the other

$$\mu_q P(f) = \sum_i m_{qi}\delta^{x_{qi}}P(f) = \mu_q(\delta^x P[f]) \to \mu(\delta^x P[f]) .$$

Accordingly, for $f \in C_\infty(X)$ and $\mu \in N(X)$

$$\mu(\delta^x P[f]) = \mu P(f) ; \tag{1.1}$$

the adjoint of $P$ transforms $C_\infty(X)$ into itself ; and when $f \in C_\infty(X)$

$$(Pf)(x) = \delta^x P(f) . \tag{1.2}$$

By letting $f \uparrow 1$ in (1.1) we see that (1.1) and (1.2) hold for $f \in C(X)$ as well. If $P$ and $Q$ are homogeneous transition probabilities

$$\|P\| = \delta^x P(1), \tag{1.3}$$

$$\|PQ\| = \|P\| \; \|Q\| . \tag{1.4}$$

We obtain (1.3) by noting first from (iii) that $\delta^x P(1)$ does not depend on $x \in X$, and then it follows from (i) that $\mu \geq 0$ implies $\|\mu P\| = \mu P(1) = \|\mu\|\delta^x P(1)$ ; so $\|P\| \geq \delta^x P(1)$. The opposite inequality is obtained by using the preceeding remarks on a Jordan decomposition of $\mu \in N(X)$, while (1.4) follows easily from (1.3).

We remark at this point that if $z \in X$ and $m$ is the normalized Haar measure of $G_z$, then the map $P \to \tilde{P} \in N(G)$, given for $g \in C_\infty(G)$ by $\tilde{P}(g) = \int \delta^z P(dy) \int m(dw) g(yw)$, maps the Banach algebra generated by the homogeneous transition probabilities isometrically onto the subalgebra $m * N(G) * m$ of $N(G)$.

DEFINITION 1.2. A *homogeneous process* is a one-parameter semigroup, $(T_t)_{t>0}$, of homogeneous transition probabilities which is temporally continuous in the sense

(iv) $\mu \in N(X)$ and $f \in C(X)$ imply $\mu T_t(f)$ is continuous in $t$.

Using (1.4), (1.3) and (iv) we see that $\|T_t\| = e^{pt}$. Therefore, we replace $T_t$ by the equivalent process $e^{-pt}T_t$ and assume in the rest of this paper that (v) below is satisfied unless explicitely stated otherwise.

(v) $\|T_t\| = 1$,

The requirement of temporal continuity for a homogeneous process is equivalent to the weak continuity of the restricted adjoint process $T_t: C(X) \to C(X)$, and by well known results, [5], implies the strong continuity of this last process for $t > 0$. The theory of semi-groups shows that $T_t$ is strongly continuous at $t = 0$ on the closure of $\bigcup_{t>0} T_t C(X)$. This may be a proper subspace of $C(X)$. For example, let $(G, G)$ be the homogeneous space of a group acting on itself by left translations with $\mu T_t = \mu * m$, where $m$ is the normalized Haar measure of a non trivial compact subgroup. G. Hunt [7, pp. 291–293] has shown that by enlarging the subgroups $G_z$ if necessary we can always assume a homogeneous process possesses the following quality.

*Property* I. For any fixed $x \in X$ and any Borel measurable neighborhood $N$ of $x$, $\delta^x T_t(N) \to 1$ or equivalently $\delta^x T_t(N^c) \to 0$ as $t \to 0$.

It is a routine calculation to show that Property I is equivalent to the strong continuity of $T_t: C(X) \to C(X)$ at the origin, so we henceforth assume our processes satisfy Property I, and by Hunt's result we can do this without loss of generality.

In view of the above statements we can apply the Hille-Yosida theory of strongly continuous semi-groups to the semi-group $T_t: C(X) \to C(X)$. An elementary application of this theory shows that there exists a dense linear subspace of $C(X)$ which we denote by $D(A)$, and a closed linear operator $A: D(A) \to C(X)$ with the property that for $f \in D(A)$ $\limsup_{t\to 0} \| (Tf - f)t^{-1} - Af \|_\infty = 0$.

2. **Properties of** $D(A)$. In this section we investigate the domain of the infinitesimal generator for homogeneous processes which satisfy Property II below. Later we show that Property II is automatically satisfied if either $X$ is separable or $G$ is commutative.

*Property* II. There is a regular Borel measure $Q_z$ on $X - \{z\}$, such that $t^{-1}(T_t f)(z) \to Q_z(f)$ as $t \to 0$ for all $f \in C(X)$ which vanish on any neighborhood of $z$.

$Q_z$ is positive and $Q_z L_{a^{-1}} = Q_{a[z]}$. In general, of course, $Q_z$ will be unbounded, although its values on any set $E$ lying in the complement of a fixed neighborhood $N$ of $z$ must be bounded, or equivalently $\delta^z T_t(E) = O(|t|)$ as $t \to 0$. This is easily checked if one notes that $D(A)$ includes the constant functions and is invariant under $G$. For then we can choose an $f \in D(A)$ which is everywhere positive on $X$, vanishes at $z$, and is greater than 1 on $N^c$. Clearly

$$Q_z(E) \leq (Af)(z) = \lim_{t\to 0} t^{-1}(T_t f)(z)$$

is bounded independently of $E \subset N^c$.

The homogeneity of the process $T_t$ entails a uniformity in this convergence to $Q_z$ which may be stated as follows.

THEOREM 2.1. *If $J$ is an open subset of $X$, $f \in C(X)$, and $f(J) = 0$; then $t^{-1}T_t f$ converges to its limit as $t \to 0$ uniformly on every compact subset of $J$. If in addition $J^c$ is compact, this convergence is even uniform on every closed subset of $J$.*

*Proof.* To prove the first assertion it suffices to show that the approach is uniform on a neighborhood of $z \in J$ of the form $N[z]$ where $N^2[z] \subset J$ and $N = N^{-1}$ is compact. If $a, b \in N$, and $K$ is chosen so that $t^{-1}\delta^z T_t(N[z]^c) \leq K$, then

$$|t^{-1}\delta^{a[z]}T_t(f) - t^{-1}\delta^{b[z]}T_t(f)| \leq K\|L_{ba^{-1}}(f) - f\|_\infty .$$

The family $\{h_t(x) | t > 0, h_t(x) = t^{-1}\delta^x T_t(f)\}$ is accordingly, equicontinuous on $N[z]$, and $h_t(x) \to Q_x(f)$ as $t \to 0$. It follows that this approach is uniform on $N[z]$. The second assertion is a consequence of the fact that for each $\varepsilon > 0$ there is a compact set $D_\varepsilon$ and a $t_\varepsilon > 0$, such that $x \notin D_\varepsilon$ and $t \leq t_\varepsilon$ imply $|t^{-1}(T_t f)(x)| \leq \varepsilon$. Assume on the contrary that there is a sequence $a_i[z] \to$ infinity, together with a sequence $t_i \to 0$, such that $|t_i^{-1}(T_{t_i} f)(a_i[z])| > \varepsilon$. Now choose a bounded sequence of functions $g_j$ from $C(X)$ which converges to zero monotonely while their supports approach infinity, and which satisfy the crucial inequalities $g_j \geq \sup_{i \geqq j} |L_{a_i^{-1}}(f)|$. The inequalities

$$\varepsilon \leq \liminf_i |t_i^{-1}(T_{t_i} f)(a_i[z])| \leq \lim_j Q_z(g_j) = 0$$

yield the contradiction which proves the second assertion.

Suppose $h \in D(A)$ and on some open subset $J$ of $X$ we have $f = h$, where $f \in C(X)$. Suppose further that either $\bar{J}$ or $J^c$ is compact. Let $D$ be a closed subset of $J$, containing a neighborhood of infinity if $J^c$ is compact; and let $m \in C(X)$ be constructed so that $0 \leq m \leq 1$, $m(D) = 0$, and $m(J^c) = 1$. The map $S: X \times X \times (0, \infty) \to R$ defined by $S(x, y, t) = (T_{m(y)t} f)(x)$ is, because of the strong continuity of $T_t$, defined and continuous in $(x, y, t)$. Consequently its restriction to the diagonal in the first two components is continuous on $X \times (0, \infty)$. When $x \in J^c$, $(T_{m(x)t} f)(x) = (T_t f)(x)$, while if $J^c$ is compact $(T_{m(x)t} f)(x) = (T_0 f)(x) = f(x)$ close to infinity. The maps $W_t: C(X) \to C(X)$ defined by $f(x) \to T_{m(x)t} f(x)$ form a strongly continuous family of bounded linear transformations, so that we may form the Rieman integral $g(x) = r^{-1}\int_0^r (W_t f)(x)dt$. If $x \in D$, $m(x) = 0$ and $g(x) = f(x)$; while for any $x$,

$$\begin{aligned} |g(x) - f(x)| &= r^{-1}\left|\int_0^r \{W_t f(x) - f(x)\}dt\right| \\ &\leq \sup_{t<r} \| W_t f - f \|_\infty \to 0 \text{ as } r \to 0 . \end{aligned}$$

Thus for any given $\varepsilon > 0$ we may guarantee that $\|g - f\|_\infty < \varepsilon$ by a sufficiently small choice of $r$.

THEOREM 2.2. *Let $D$ be a closed subset of the open set $J$ in $X$, where either $\bar{J}$ or $J^c$ is compact. Let $h \in D(A)$, $f \in C(X)$, and $f = h$ on $J$. Then for each $\varepsilon > 0$ there exists a $g_\varepsilon \in D(A)$, such that* (i) $h = f = g_\varepsilon$ *on* $D$, *and* (ii) $\|g_\varepsilon - f\|_\infty < \varepsilon$.

*Proof.* We only need to prove that the $g$ defined above lies in $D(A)$. To do this we show that

$$s^{-1}\{T_s g - g\} = r^{-1}\int_0^r s^{-1}W_t\{T_s f - f\}dt$$

converges to its limit uniformly on $X$. The computations are divided into two cases.

*Case* I. $m(x) \leq 1/2$. On this closed subset of $J$

$$s^{-1}\{T_s f - f\} = s^{-1}\{T_s h - h\} + s^{-1}\{T_s(f - h) - (f - h)\} .$$

Now $s^{-1}\{T_s h - h\} \to Ah$ in the uniform norm as $s \to 0$, and by Theorem 2.1 $s^{-1}\{T_s(f - h) - (f - h)\} \to Q_x(f - h)$ uniformly on $\{x \mid m(x) \leq 1/2\}$. Accordingly,

$$s^{-1}\{T_s g - g\} \to r^{-1}\int_0^r W_t\{Ah + Q_x(f - h)\}dt$$

uniformly on $\{x \mid m(z) \leq 1/2)$.

*Case* II. $m(x) \geq 1/4$. On this set

$$\begin{aligned} s^{-1}\{T_s g - g\} &= (rs)^{-1}\int_0^r \{T_{s+m(x)t}f - T_{m(x)t}f\}dt \\ &= \{rsm(x)\}^{-1}\left[\int_{m(x)r}^{m(x)r+s}\{T_u f\}du - \int_0^s T_u f du\right] \\ &\to \{rm(x)\}^{-1}\{T_{m(x)r}f(x) - f(x)\} \end{aligned}$$

as $s \to 0$, uniformly on $\{x \mid m(x) \geq 1/4\}$. These observations show that $g \in D(A)$.

In the above proof we did not really need the fact that $h \in D(A)$. Indeed, had we replaced $h$ by $h + u$, where $u \in C(X)$ and $u(J) = 0$, there would have been no change in the proof. Furthermore, since $g(x) = r^{-1}\int_0^r (T_{m(x)t}f)(x)dt$, if $f$ is real valued we have the relation $\inf_x f(x) \leq \inf_y g(y) \leq \sup_z g(z) \leq \sup_x f(x)$. These remarks allow us to state a corollary.

COROLLARY 1. *Let $J_1, J_2, \cdots, J_n$ be disjoint open sets of $X$ and let $D_i$ be a closed subset of $J_i$. Suppose that for each $i$ either $\bar{J}_i$ is compact or for at most one $i$, $J_i^c$ is compact. Let $h_i \in D(A)$ and $f \in C(X)$ satisfy $f = h_i$ on $J_i$. Then for each $\varepsilon > 0$ there exists a $g_\varepsilon \in D(A)$, such that*

(i) $g_\varepsilon = h_i = f$ on $D_i$,

(ii) $\|g_\varepsilon - f\|_\infty < \varepsilon$, and

(iii) *if $f$ is real valued then* $\inf_x f(x) \leq \inf_y g_\varepsilon(y) \leq \sup_z g_\varepsilon(z) \leq \sup_x f(x)$.

Specializing the preceding we get the following.

COROLLARY 2. *If $B$ is compact in $X$, $H$ is closed, and $H \cap B = \phi$; there exists an $f \in D(A)$, such that $0 \leq f \leq 1, f(H) = 0$,* and *$f(B) = 1$.*

The above results have been derived under Property II and we now show that we can replace Property II by a condition on $D(A)$.

*Property* III. For each $\varepsilon > 0$ and $f \in C(X)$ which is zero in a neighborhood $J$ of $x$, there exists an $f_\varepsilon \in D(A)$, such that $f_\varepsilon = 0$ in a neighborhood $U$ of $x$ which is independent of $\varepsilon$, $\|f_\varepsilon - f\|_\infty < \varepsilon$, and if $f$ is real valued so is $f_\varepsilon$ with $\inf_x f(x) \leq \inf_y f_\varepsilon(y) \leq \sup_z f_\varepsilon(z) \leq \sup_x f(x)$.

THEOREM 2.3. *Property* III *is equivalent to Property* II *for homogeneous processes.*

*Proof.* We have already seen that Property II implies Property III and will now demonstrate the converse. First, let $J$ be a neighborhood of $x$, and choose a function $h \in C(X)$, such that $h = 0$ in a neighborhood $U \subset J$ of $x$, and $h(J^c) = 1$. We also require that $h \geq 0$ everywhere. Now for $0 < \varepsilon < 1$ choose an $h_\varepsilon$ in accordance with Property III.

$$t^{-1}\delta^x T_t(J^c) \leq (1 - \varepsilon)^{-1}\delta^x T_t(h_\varepsilon)t^{-1} \leq K(U) < + \infty$$

for some constant $K(U)$ depending on $U$. If $f \in C(X)$ and $f(J) = 0$, Property II is equivalent to the fact that $t^{-1}\delta^x T_t(f) \rightarrow a$ limit as $t \rightarrow 0$. The inequalities

$$\begin{aligned} &|\lim_{t\rightarrow 0} t^{-1}\delta^x T_t(f_{\varepsilon_1}) - \lim_{t\rightarrow 0} t^{-1}\delta^x T_t(f_{\varepsilon_2})| \leq \\ &\quad \leq \limsup_{t\rightarrow 0} t^{-1}\delta^x T_t(|f_{\varepsilon_1} - f_{\varepsilon_2}|) \\ &\quad \leq (\varepsilon_1 + \varepsilon_2) \limsup_{t\rightarrow 0} t^{-1}\delta^x T_t(U^c) \leq (\varepsilon_1 + \varepsilon_2)K(U) \end{aligned}$$

show that $\lim_{\varepsilon\rightarrow 0} \lim_{t\rightarrow 0} t^{-1}\delta^x T_t(f_\varepsilon)$ exists. If we call this limit $b$, the inequality

$$|t^{-1}\delta^x T_t(f) - b| \leq |t^{-1}\delta^x T_t(f_\varepsilon) - b| + \varepsilon K(U)$$

shows by first letting $t \to 0$ and then $\varepsilon \to 0$, that $t^{-1}(T_t f)(x) \to b$ as $t \to 0$, completing the proof.

By the use of well-known smoothing techniques which exist on a differentiable manifold we can conclude that if $G$ is a Lie group or if $X$ is a differentiable manifold on which $G$ acts differentiably, and $D(A)$ includes $C^\infty(X) \cap C(X)$, then Property III and Property II are automatically satisfied.

We close this section by remarking that the only place where we have used the homogeneity of $T_t$ was in the proof of Theorem 2.1, so that a strongly continuous semi-group on $C(X)$ satisfying the conclusions of Theorem 2.1 also satisfies the conclusions of Corollary 1 if its adjoint preserves positivity.

**3. The paths of a homogeneous process and Property II.** Before we discuss the nature of the paths of a homogeneous process it is necessary to take certain precautions which will assure us that the properties we want to discuss can be handled by the theory of probability. Given a consistent set of transition probabilities for a particle moving in a locally compact Hausdorff space it is possible, using a theorem of Kolmogoroff's, to construct an infinite product space in which these transition probabilities determine the finite dimensional distributions. From this we can construct, in the usual manner, a set of paths and a probability measure on this path space. Alternatively, we can consider a process as a family $X(t, \omega)$ of measurable transformations from some abstract sample space $\Omega$ to the range space $X$. Since there is some freedom in the definition of $X(t, \omega)$ given only the finite dimensional distributions, it is important to notice that the concept of separability as used on the real line is available in this case.

DEFINITION 3.1. A process $\{X(t, \omega), 0 \leq t\}$ will be called *separable relative to the class $A$* of closed subsets of the locally compact Hausdorff space $X$ if, and only if, (1) there exists a denumerable subset $\{t_j\} \subset [0, \infty)$, and (2) an event $K \subset \Omega$ with $P\{K\} = 0$, such that for every open interval $I \subset [0, \infty)$, and every set $F \in A$, the event

$$\{\omega | X(t_j, \omega) \in F, t_j \in I \cap \{t_j\}\} - \{\omega | X(t, \omega) \in F, t \in I\} \subset K .$$

The importance of the concept of separability rests on the validity of the following theorem.

THEOREM 3.1. *Let $X$ be a separable locally compact Hausdorff space, and let $\{X(t, \omega), t \leq 0\}$ be an $X$-valued process. Then there exists an $X^*$-valued stochastic process $\{X(t, \omega), t \geq 0\}$ such that*: (1) $X(t, \omega)$ *is* *defined on the same $\omega$-space, $\Omega$, as $X(t, \omega)$ and takes values in the one point compactification, $X^*$, of $X$*; (2) *$X(t, \omega)$ is separable relative to the class of closed sets of $X^*$*; (3) *for every $t \geq 0$ $\{P\omega \mid X(t, \omega) = X(t, \omega)\} = 1$.*

*Proof.* This theorem may be proved in a manner entirely analogous to the comparable theorem on the real line, and its proof is given, for example, in Doob [2 p. 57]. We remark that the separability of the space is necessary for this proof.

In the following we discuss the displacements of the path $X(t, \omega)$ during a closed interval of time. The success of our technique depends directly on the possibility of comparing two such displacements with different origins. If our homogeneous space $(X, G)$ is that of a group acting on itself, we can translate all displacements origins to the identity and their endpoints are uniquely determined. On a general homogeneous space, however, the endpoint of a translated displacement is not determined by its new origin. We introduce several concepts from the calculus of relations and a space of displacements to handle this ambiguity.

By a relation $(U)$ on $X$ we mean a subset of $X \times X$ which contains the diagonal. The notations $(U)^{-1} = \{(y, x) \mid (x, y) \in (U)\}$, $(W) \circ (U) = \{(x, y) \mid \text{for some } z, (x, z) \in (U) \text{ and } (z, y) \in (W)\}$, and $(U)[A] = \{x \mid (y, x) \in (U) \text{ for some } y \in A\}$ are standard. It is sometimes convenient to substitute $U_x$ for $(U)[\{x\}]$ and in this notation $(W) \circ (U)_x = \bigcup_{y \in U_x} W_y$. A relation $(U)$ is called homogeneous whenever $(x, y) \in (U)$ implies $(a[x], a[y]) \in (U)$ for all $a \in G$, and it is convenient to describe a relation by giving a property possessed by all the sets $U_x$. For example, we call a relation $(U)$ compact, open, closed, or a neighborhood relation if each $U_x$ is compact, open, closed or a neighborhood of $x$ respectively. Using a double coset representation it is easily shown that the class of homogeneous neighborhood relations forms a base for the natural uniformity of the homogeneous space $(X, G)$. We say the displacement from $x$ to $y$ belongs to the homogeneous relation $(U)$ if $(x, y) \in (U)$. The reason for this mention of relations is that they appear to be exactly what is needed to generalize the statement and proof of a theorem from Kinney [9], p. 292–293.

THEOREM 3.2. *Let $(X, G)$ be a homogeneous space satisfying the first axiom of countability, and let $X^*$ be the space $X$ compactified by adding a single point at infinity. Let $\{X(t, \omega), t \geq 0\}$ be an $X^*$-valued homogeneous stochastic process governed by the transition probabilities $\delta^x T_t$. and separable with respect to the closed subsets of $X^*$. Then if $T > 0$, there is an $\omega$-set $E_T$ with $P\{E_T\} = 0$, such that $\omega \notin E_T$ implies the statements below.*

(1) *$X(t, \omega)$ is bounded on $t \in [0, T)$. By which we mean*

$$Cl[\{X(t, \omega) | t \in [0, T)\}]$$

*is a compact subset of* $X$.

(2) $X(t, \omega)$ *has finite right and left hand limits at every* $t \in [0, T)$.

(3) *The number of jumps in* $[0, T)$ *whose displacements lie outside a homogeneous neighborhood relation* $(U)$ *is finite. Furthermore, for any homogeneous neighborhood relation,* $(U)$, *the maximum number of disjoint subintervals* $(t, s)$ *of* $[0, T)$ *for which* $(X(t^-, \omega), X(s^-, \omega)) \notin (U)$ *is finite, where* $X(t^-, \omega) = \lim_{h \uparrow 0} X(t + h, \omega)$.

In particular the use of the one point compactification of $X$ was only necessary to cover the processes constructed in Theorem 3.1 and may be eliminated as soon as (1) is proved.

The displacement or jump of a particle from $x$ to $y$ can be considered as a point in the space $X \times X$. It is natural to consider classes of similar displacements, and this involves the introduction of an equivalence relation on $X \times X$, two points, $(x, y)$ and $(x', y')$, being considered equivalent if there is an $a \in G$ such that $(a[x], a[y]) = (x', y')$. This is a closed equivalence relation and the quotient space is homeomorphic to the space of double cosets $\{G_x a G_x\} = Y$. Let $p: X \times X \to Y$ be the canonical projection. $Y$ is a locally compact Hausdorff space known as the space of displacements, and $p$ is a continuous open mapping. If we fix the first component at $x$, so we only consider jumps origination at $x$, we get a map $p': X \to Y$ given by $z \to p(x, z)$. Using this map the commutativity of $T_t$ with $L_a$ shows that we can very properly place the measures $\delta^x T_t$ and $Q_x$ on the space $Y$ without losing a thing. If $f \in C_c(X)$, and $m$ is the normalized Harr measure of $G_x$, the equation $\delta^x T_t(f) = \int f(a[z]) m(da) \delta^x T_t(dz)$ indicates a means of returning $\delta^x T_t$ and $Q_x$ to $X$ from $Y$. The following theorem is the key to the results of this section. It and Theorem 3.4 are generalizations of similar results for homogeneous processes on the real line which may be found, for example, in Doob [2, p. 422–424].

THEOREM 3.3. *Let* $(X, G)$ *be a homogeneous space satisfying the first axiom of countability, and* $X(t, \omega)$ *a separable homogeneous process on* $X$ *governed by the transition probabilities* $\delta^X T_t(\cdot)$. *Suppose that for some sequence of t-values,* $t_j \to 0$, *there is a not necessarily bounded regular Borel measure* $Q_X$ *on* $X - \{x\}$; *for which* $f \in C_c(X), x \notin$ *support* $(f)$, *imply* $t_j^{-1}(T_{t_j} f)(x) \to Q_x(f)$ *as* $t_j \to 0$. *Let* $Y$ *be the space of displacements of* $(X, G)$, *and denote the measures* $\delta^x T_t$ *and* $Q_x$ *after transference to* $Y$ *by the same symbols. Suppose* $X(0, \omega) = x$ a.s,, $C(D) = \{\omega | \lim_{n \to \infty} p(X(t - n^{-1}, \omega), X(t + n^{-1}, \omega)) \in D$ *for some* $t \in [0, T)\}$, *and let* $P_*$ *and* $P^*$ *be the inner and outer measures induced by* $P$ *on* $\Omega$.

(1) *Then for any compact subset* $D$ *of* $Y - \{x\}$:

$$1 - \exp\{T \limsup_j t_j^{-1}\delta^x T_t(D)\} \leq P_*\{C(D)\}\,;\, P^*\{C(D)\} \leq 1 - e^{-TQ_x(D)}\ .$$

(2) *If $D$ is a compact subset of $Y - \{x\}$ satisfying, either*

(i) *there is a $U$ open $\subset D$ for which $\delta^x T_t(U - D) = o(t)$ as $t \to 0$, or*

(ii) *there exists a sequence $C_k$ compact $\subset$ interior $(D)$ for which $Q_x(C_k) \to Q_x(D)$ as $k \to \infty$, then*

$$P_*\{C(D)\} = P^*\{C(D)\} = 1 - e^{-TQ_x(D)}$$

*Proof.* Using the monotone sequence $t_j \to 0$ we define a sequence of partitions of $[0, T)$. The $j$th partition being given by

$$\{[0, t_j), [t_j, 2t_j), \cdots, [(k_j - 1)t_j, k_j t_j), [k_j t_j, T)\}\ ,$$

where $k_j$ is the largest integer $< T/t_j$. Define $Y$-valued random variables

$$H(j, n) = p(X(\{n-1\}t_j^-, \omega), X(nt_j^-, \omega)) \qquad 1 \leq n \leq k_j$$
$$H(j, k_j + 1) = p(X(k_j t_j^-, \omega) X(T^-, \omega))\ ,$$

where as usual $X(t^-, \omega) = \lim_{s \uparrow t} X(s, \omega)$. For any measurable subset $F \subset Y$ we put $F(j, n) = \{\omega \mid H(j, n) \in F\}$ and $F(j) = \bigcup_{n=1}^{k_j+1} F(j, n)$. Since $\{H(j, n), 1 \leq n \leq k_j\}$ are independent and identically distributed random variables, it follows that

$$P\{F(j)\} = 1 - \{1 - \delta^x T_{T-k_j t_j}(F)\}\{1 - \delta^x T_{t_j}(F)\}^{k_j}\ ,$$

where $1 \geq \varepsilon_j \to 0$, $0 \leq T - k_j t_j < t_j \downarrow 0$, and $k_j = Tt_j^{-1} - \varepsilon_j$. Let $D$ be a compact subset of $Y$, and $U$ an open neighborhood of $D$ whose closure does not contain $x$. Our knowledge that the paths have right and left hand limits at every point shows that $C(D) \subset \liminf_j U(j)$. If $\omega \in \limsup_j D(j)$, choose a sequence of semiclosed intervals $[n_j t_j, (n_j + 1)t_j)$ which converge to a point, and across whose length the path $\omega$ has a displacement $H(j, n_j) \in D$. By passing to further subsequences if necessary we can assume that the sequences of endpoints are monotone. There are four cases;

(a) $n_j t_j \uparrow$, $(n_j + 1)t_j \downarrow$ leads to $\omega \in C(D)$;

(b) $n_j t_j \downarrow$, $(n_j + 1)t_j \uparrow$ is impossible; while

(c) $n_j t_j \uparrow$, $(n_j + 1)t_j \uparrow$ and

(d) $n_j t_j \downarrow$, $(n_j + 1)t_j \downarrow$ both lead to $\omega$'s having infinitely many displacements close to $D$ in $[0, T)$, and, accordingly, occuring with probability zero by condition (3) of Theorem 3.2. Therefore, $\limsup_j D(j) \subset C(D) \subset \liminf_j U(j)$ which in turn implies

$$1 - \exp\{-T \limsup_j t_j^{-1}\delta^x T_{t_j}(D)\} \leq P_*\{C(D)\}\ ,$$

and

$$P^*\{C(D)\} \leq 1 - \exp\{-T \liminf_j t_j^{-1}\delta^x T_{t_j}(U)\}\ .$$

It can be shown by a standard arguement that if $V$ open $\supset U$ and $x \notin \overline{V}$, then $Q_x(D) \leq \liminf_j t_j^{-1}\delta^x T_{t_j}(U) \leq Q_x(V)$, and by letting $V$ and $U$ shrink to $D$ it follows that

$$P^*\{C(D)\} \leq 1 - e^{-TQ_x(D)} .$$

If in addition $D$ satisfies either (i) or (ii) in (2), it is easy to see that

$$Q_x(D) \leq \limsup_j t_j^{-1}\delta^x T_{t_j}(D) ,$$

which implies the conclusion of (2).

We now show that the conditions (i) and (ii) placed on the compact set $D$ in (2) of Theorem 3.3 are sufficiently unrestrictive for us to prove Property II for homogeneous processes.

THEOREM 3.4. *Let $(X, G)$ be a separable locally compact homogeneous space, and let $T_t$ be a homogenous process on $(X, G)$. Then there exists a unique not necessarily bounded Borel measure $Q_x$ on $Y - \{x\}$, the space of displacements of $(X, G)$, such that $f \in C(X)$ and $x \notin$ support $(f)$ imply*

$$t^{-1}(T_t f)(x) \to Q_x(f) \qquad \text{as } t \to 0 .$$

*Proof.* The use of the Hille-Yosida theory of strongly continuous semi-groups shows that when restricted to the complement of any neighborhood $U$ of $x \in Y$, the family of measures $t^{-1}\delta^x T_t$ is bounded. We compactify $Y$ by adding a point at infinity and denote the compactified space by $Y^*$. Using the compactness of bounded sets in the weak star topology, and the first axiom of countability for $Y$, we can find a sequence $t_j \to 0$ and a not necessarily bounded Borel measure $Q_x$ on $Y^* - \{x\}$, such that for any $f \in C(Y) = C(Y^*)$, $x \notin$ support $(f)$ implies $t_j^{-1}\delta^x T_{t_j}(f)$ as $t_j \to 0$. There remain two problems.

(a) to prove $Q_x$ is unique, and

(b) to show that $Q_x(\{\infty\}) = 0$. Using the separability of $(X, G)$ construct a representation, $X(t, \omega)$, of the paths of $T_t$ satisfying all the conditions in Theorem 3.2.

Suppose that $s_j^{-1}\delta^x T_{s_j} \to Q_x'$ were another limit and let $C$ be any compact subset of $Y - \{x\}$. For an arbitrary $\varepsilon > 0$ let $U_\varepsilon$ be chosen so that $C \subset U_\varepsilon$ open $\subset \overline{U}_\varepsilon$ compact $\subset Y - \{x\}$, $Q_x(U_\varepsilon) < Q_x(C) + \varepsilon$, and $Q_x'(U_\varepsilon) < Q_x'(C) + \varepsilon$. Now construct a function $f \in C_c(Y)$, such that $f(U_\varepsilon{}^c) = 0, 0 \leq f \leq 1$, and $f(C) = 1$. Select a $b \in (0, 1)$ such that

$$Q_x(\{x \mid f(x) = b\}) = Q_x'(\{x \mid f(x) = b\}) = 0$$

and put $D = \{x \mid f(x) \leq b\}$. Note that

$$C \text{ compact} \subset \text{interior } (D) \subset D \text{ compact} \subset U_\varepsilon ,$$

$Q_x$ (interior $(D)$) $= Q_x(D)$, and $Q_x'$ (interior $(D)$) $= Q_x'(D)$. Accordingly, $D$ satisfies condition (ii) of Theorem 3.3 (2), and thus for any fixed $T > 0$,

$$P^*\{C(D)\} = 1 - e^{-T}Q_x(D) = 1 - e^{-TQ_x'(D)} ,$$

so that $Q_x(D) = Q_x'(D)$. This shows

$$|Q_x'(C) - Q_x(C)| \leq |Q_x(D) - Q_x(D - C) + Q_x'(D - C) - Q_x'(D)| \leq 2\varepsilon$$

for every $\varepsilon > 0$; so $Q_x = Q_x'$.

To show that $Q_x$ is really a Borel measure on $Y - \{x\}$, and not on $Y^* - \{x\}$, we must show that $Q_x(\{\infty\}) = 0$. In order to do this we make a final appeal to the paths. Except for an event of probability zero we know that $\mathrm{Cl}\,[\{X(t, \omega) | t \in [0, T)\}]$ is a compact subset of $X$, and, consequently, its projection on $Y$ will also be compact. Using the method above, choose a sequence of compact sets $D_n \downarrow \{\infty\}$ in $Y^*$ and satisfying condition (ii) of Theorem 3.3 (2). Then if $Q_x(\{\infty\}) \neq 0$,

$$P\{C(D)\} = 1 - e^{-T}Q_x(D_n) \geq r > 0 ,$$

and $P\{\bigcap_{n=1}^{\infty} C(D_n)\} > 0$. Any path with a jump in every $D_n$ during the time interval $[0, T)$ certainly contains $\infty$ as a limit point. Thus our process violates the condition that the paths are a.s. bounded. Hence $Q_x(\{\infty\}) = 0$.

The temporal continuity (weak continuity) of $T_t$ enables us to restrict consideration to a sigma-compact subset of $X$, namely $\bigcup_{t>0}$ support $(\delta^x T_t)$. As a consequence of the preceeding theorem this remark proves the following corollary.

COROLLARY. *Let $(X, G)$ be a homogeneous space where every compact subset of $X$ is separable. Then every homogeneous process on $(X, G)$ possesses Property* II.

The following theorem gives an accurate description of the important set support $(Q_x)$.

THEOREM 3.5. *Let $(X, G)$, $Y$, $X(t, \omega)$, $\delta^x T_t$, and $Q_x$ be as in Theorem* 3.4. *For those $\omega$'s which have one sided limits let*

$$F(\omega) = \{\lim_{n\to\infty} p(X(t - n^{-1}, \omega), X(t + n^{-1}, \omega)) | t \geq 0\} .$$

*Then* $\overline{F(\omega)} =$ *support* $(Q_x) \cup \{x\}$ a.s..

*Proof*. By definition $C(\bigcup_\alpha F_\alpha) = \bigcup_\alpha C(F_\alpha)$ for $F_\alpha$ measurable $\subset Y$. Now choose a sequence of compact sets $D_i$, such that $\bigcup_i D_i = \{\text{support } (Q_x) \cup \{x\}\}^c$. $P^*\{C(D)\} \leq 1 - e^{-T\cdot 0} = 0$, so

$$P^*\{C(\bigcup_i D_i)\} = P^*\{\bigcup_i C(D_i)\} \leq \sum_i P^*\{C(D_i)\} = 0 ,$$

which shows $F(\omega) \subset \text{support}\,(Q_x) \cup \{x\}$. Let $U$ open $\subset \overline{U}$ compact $\subset Y - \{x\}$, and $U \cap$ support $(Q_x) \neq \phi$. Choose $f \in C_c(Y)$, such that $0 \leq f \leq 1$, support $(f) \subset U$, and $Q_x(f) > 0$. Let

$$r(U) = \limsup_{t \to 0} t^{-1}\delta_x T_t(\overline{U} \geq Q_x(f) > 0 \,.$$

If $\{U_i, i = 1, 2, \cdots, n\}$ is a finite class of such $U$'s, let

$$P_*\{\omega \,|\, \omega \text{ has a jump in every } \overline{U}_i \text{ when } t < T\} = A(T) \,;$$
$$P_*\{\omega \,|\, \omega \text{ has a jump in } \overline{U}_k \text{ when } T(k-1)n^{-1} \leq t < Tkn^{-1}\} = B(k, T) \,.$$

We have the following inequalities relating the above numbers

$$A(T) \geq B(1, T)B(2, T) \cdots B(n, T)$$
$$\geq (1 - \exp\{-n^{-1}Tr(\overline{U}_1)\}) \cdots (1 - \exp\{-n^{-1}Tr(\overline{U}_n)\}) \,.$$

Letting $T$ approach infinity one sees that

$$P_*\{\omega \,|\, \omega \text{ has a jump in each } \overline{U}_i \text{ for some } t > 0\} \geq 1.$$

If we now choose a countable sequence of finite classes of the type $\{U_i\}$ above, and let their sets become arbitrarily fine while their unions swell out and eventually cover support $(Q_x)$,

$$P_*\{\omega \,|\, \omega \text{ has a jump in each } \overline{U}_i \text{ of the } k\text{th covering}\} = 1 \,.$$

Hence the inner measure of their intersection is one. Now the paths in their intersection have jumps in the closure of each of the finer and finer covering sets, and consequently for these paths $\mathrm{Cl}\,[\overline{F(\omega)}] \supset$ support $(Q_x)$. The existence of left hand limits for $X(t, \omega)$ implies $x \in [F(\omega)]$.

**4. Compound Poisson processes.** The poisson process with rate parameter $r \geq 0$ on the real line is a homogeneous process with transition probabilities $\delta^0 T_t(E) = \exp\{tr(\delta^1 - \delta^0)\}(E)$. It can be generalized to a compound Poisson process by replacing $\delta^1$ by any positive regular Borel measure $\mu$ of norm one. Probabilistically one thinks of a compound Poisson process in the following manner. A simple Poisson process is run at a rate $r$, and when a jump occurs in this simple process, the particle ruled by the compound Poisson process jumps from its position $x$ into the set $E + x$ with probability $\mu(E)$.

Suppose we observe two Poisson processes, $\exp\{tr_i(\delta^1 - \delta^0)\}$ $i = 1, 2$, running simultaneously. We can then define a new process as follows. The state of our process will be described by a finite sequence of $x_1$'s and $x_2$'s called a word, and we change states when a jump occurs in either of our two simple Poisson processes. If a jump occurs in the $i$th process we lengthen our state by placing the symbol $x_i$ to the right of the current word. One can calculate that the probability of starting with the empty state at time zero, and being at a fixed state with $n_1 x_1$'s and $n_2 x_2$'s at time $T$, is independent of their order, and given by $\exp\{-(r_1 + r_2)T\}(r_1 T)^{n_1}(r_2 T)^{n_2}/(n_1 + n_2)!$. To give an alternative description of this process, let $H$ be the free group generated by the symbols $\{x_1, x_2\}$. Topologize $H$ with the discrete topology, and consider the compound Poisson process on $H$ given by

$$\delta^e T_t = \exp\{t(r_1 + r_2)(p\delta^{x_1} + q\delta^{x_2} - \delta^e)\ ,$$

where $p = r_1(r_1 + r_2)^{-1}$ and $q = r_2(r_1 + r_2)^{-1}$. An elementary expansion shows this process is identical with the word generating process defined above. It would seem natural to define the superposition of the two Poisson processes $\exp\{tr_i(\delta^{x_i} - \delta^e)\}$ to be the process $\delta^e T_t$. This symbol generating process can also be interpreted by running a simple Poisson process at the rate $r_1 + r_2$, and each time a jump occurs, multiplying on the right by $x_1$ with probability $p$ and by $x_2$ with probability $q$.

The analogue of the compound Poisson processes for a homogeneous space $(X, G)$ is the class of processes of the form $T_t = \exp\{tr(P - 1\}$, where $P$ is a homogeneous transition probability of norm one. We shall, accordingly, call these processes compound Poisson processes. An easy computation shows that the infinitesimal generator of such a process is $A = r(P - 1)$ and $D(A) = C(X)$. Thus the superposition of the two processes in the preceeding paragraph corresponds to the addition of their infinitesimal generators. We use this last remark to define the superposition of an arbitrary homogeneous process and a compound Poisson process.

DEFINITION 4.1. The sequence $T_t^{(n)}$ of semi-groups on $C(X)$ is said to *converge in the sense of Bernoulli* $\{$*strongly*$\}$ *to the semi-group* $T_t$ if, and only if, whenever $f \in C(X)$, $(T_t^{(n)} f)(x) \to (T_t f)(x)$ for each fixed $x$ and $t$ as $n \to \infty$ $\{$if, and only if, the following condition is satisfied. For each $\delta > 0$ and each $f \in D(A)$ where $A$ is the infinitesimal generator of $T_t$, there exists an integer $N_{\delta,f}$, such that $n \geq N_{\delta,f}$ implies $\|(T_t^{(n)} f - T_t f\|_\infty \leq \delta t$ for all $t > 0\}$.

It is an elementary consequence of this definition that $T_t^{(n)} \to T_t$ strongly implies $T_t^{(n)} \to T_t$ in the sense of Bernoulli. We now recall a fact from the theory of semi-groups which we need in the proof of the next theorem. Put $A_\varepsilon = \varepsilon^{-1}(T_\varepsilon - 1)$, then $\|e^{tA_\varepsilon} f - T_t f\|_\infty \to 0$ as $\varepsilon \to 0$ for any $f \in C(X)$, and uniformly for $t \in (0, M)$, $M < \infty$. More precisely,

$$\text{(4.1)} \qquad \|(e^{tA_\varepsilon} - T_t)f\|_\infty \leq t \limsup_{s \to 0} \|(A_\varepsilon - A_s)f\|_\infty\ ,$$

which for $f \in D(A)$ becomes $\|(e^{tA_\varepsilon} - T_t)f\|_\infty \le t\|(A_\varepsilon - A)f\|_\infty$.

THEOREM 4.1. *Every homogeneous process is a strong limit of a sequence of compound Poisson processes.*

*Proof.* Let $T_t$ be a homogeneous process, $f \in D(A)$, and $A_{t_0}$ be as above. Then $\|(\exp\{tA_{t_0}\} - T_t)f\|_\infty \le t\delta_{t_0}$, where $\delta_{t_0} \to 0$ as $t_0 \to 0$.

We now study the concept of strong convergence in more detail for compound Poisson processes. Our main results are stated in Theorems 4.2 – 4.5, but before proceeding to these theorems we establish the following useful lemma.

LEMMA 4.1. *Let $\pi_{n,t}$ $n = 1, 2, \cdots, \infty$ be positive Borel measures on $(X, G)$ while $t$ ranges over the compact separable set $F$. Suppose for each $f \in C(X)$, $\pi_{n,t}(f)$ is continuous in $t$, and $\pi_{n,t}(f) \to \pi_{\infty,t}(f)$ as $n \to \infty$ uniformly for $t \in F$. Under these conditions*

$$\sup\{|\pi_{n,t}(L_a f) - \pi_{\infty,t}(L_a f)| : a \in G\} \to 0$$

*uniformly on $F$ as $n \to \infty$.*

*Proof.* We show first that given $\varepsilon > 0$ there exists a compact set $B$, for which $\pi_{n,t}(B^c) < \varepsilon$ whatever $n$ and $t$. Choose $\{t_i\}$ as a countable dense set in $F$, and consider the union, $E$, of the supports of all $\pi_{n,t_i}$. $\overline{E}$ is sigma-compact as the closure of a sigma-compact set in a uniformly locally compact space, and includes the supports of all the measures $\pi_{n,t}$; because by the continuity of $\pi_{n,t}(f)$ if a point does not lie in $\overline{E}$ it cannot lie in the support of any $\pi_{n,t}$. Select a sequence of functions $f_n \in C(X)$, such that $f_n \downarrow 0$ on $\overline{E}$, $f_n(0_n^c) = 1$, and $f_n(V_n) = 0$, where $V_n$ is a compact set contained in the open set $0_n \subset V_{n+1}$. For fixed $k$ the sequence of continuous $t \to \pi_{k,t}(f_n)$ converges monotonically to zero on the compact set $F$ as $n \to \infty$. This convergence is then uniform. Accordingly, we can find an $M_k$, such that $n \ge M_k$ implies $0 \le \pi_{k,t}(f_n) < \varepsilon/2$ all $t \in F$. Using the hypotheses of the lemma, we can find a $K_n$, such that $k \ge K_n$ implies $|\pi_{k,t}(f_n) - \pi_{\infty,t}(f_n)| < \varepsilon/2$ for all $t$. Put $n' = M_\infty$. Then for any $K_{n'} \le k \le \infty$, and every $t \in F$,

$$\begin{aligned}\pi_{k,t}(0_{n'}^c) &\le \pi_{k,t}(f_{n'}) \\ &\le \pi_{\infty,t}(f_{n'}) + |\pi_{k,t}(f_{n'}) - \pi_{\infty,t}(f_{n'})| \le \varepsilon .\end{aligned}$$

Whereas for $k \le K_{n'}$, and every $t \in F$,

$$\pi_{k,t}(0_{M_k}^c) \le \pi_{k,t}(f_{M_k}) < \varepsilon/2 .$$

If we put $W_\varepsilon$ equal to the union of $0_{n'}$ and $\bigcup_{j=0}^{K_{n'}} 0_{M_j}$, $B_\varepsilon = \overline{W}_\varepsilon$ satisfies the desired condition.

Now put $f = 1$ and observe that from $\pi_{n,t}(f) \rightarrow \pi_{\infty,t}(f)$ uniformly on $F$, we can conclude there is an $M$, such that $0 \leq ||\pi_{n,t}|| \leq M$, and $||\pi_{n,t}|| \rightarrow ||\pi_{\infty,t}||$ uniformly on $F$. This shows we need only to prove the lemma for $f \in C_\infty(X)$. Choose a $\delta > 0$, an $f \in C_\infty(X)$, and put $D_\delta = \{x : |f(x)| \geq \delta\}$. $D_\delta$ is compact. If $H_\delta = \{a \in G : a[D_\delta] \cap B_\varepsilon = \phi\}$, $H_\delta$ is open, and $a \in H_\delta$ implies.

$$|\pi_{n,t}(L_{a^{-1}}f) - \pi_{\infty,t}(L_{a^{-1}}f)| \leq 2\varepsilon ||f||_\infty + 2\delta M .$$

In this paragraph we show that

$$\limsup_n \sup \{|\pi_{n,t}(L_{a^{-1}}f) - \pi_{\infty,t}(L_{a^{-1}}f)| : t \in F, a \in H_\delta^c\} = 0 .$$

Regard $a \rightarrow |\pi_{n,t}(L_{a^{-1}}f) - \pi_{\infty,t}(L_{a^{-1}}f)| = h_{a,n}(t)$ as a map from the compact set $H_\delta^c$ to the space of continuous real valued functions on $F$. If $b$ is sufficiently close to $a$, $||L_{a^{-1}}f - L_{b^{-1}}f||_\infty < \varepsilon'$, from which $|h_{a,n}(t) - h_{b,n}(t)| \leq 2M\varepsilon'$, so that the maps $h_{a,n}(\cdot)$ are equicontinuous in $a$. Now $\lim_{n\rightarrow\infty} h_{a,n}(t) = 0$ uniformly in $t$ by hypotheses, so that we have a sequence of equicontinuous functions, $a \rightarrow h_{a,n}(\cdot)$, defined on a compact set, $H_\delta^c$, with values in a normed vector space and converging pointwise to zero. As a trivial consequence, they converge uniformly to zero and

$$\limsup_n \sup \{h_{a,n}(t) : t \in F, a \in H_\delta^c\} = 0 .$$

Collecting results we have shown

$$\limsup_n \sup \{h_{a,n}(t) : t \in F, a \in G\} \leq 2\varepsilon ||f||_\infty + 2\delta M .$$

Since $\varepsilon$ and $\delta$ are arbitrary, this gives the conclusion of Lemma 4.1.

THEOREM 4.2. *If $r_n \rightarrow r$ and $P^{(n)}$, $P$ are homogeneous transition probabilities, such that for each $f \in C(X)$, $P^{(n)}f \rightarrow Pf$ pointwise as $n \rightarrow \infty$. Then.*

$$\exp\{tr_n(P^{(n)} - 1)\} \rightarrow \exp\{tr(P - 1)\}$$

strongly as $n \rightarrow \infty$ .

*Proof*. Given $f \in C(X)$ define $W(n, t)$ by

$$tW(n, t) = ||(\exp\{tr_n(P^{(n)} - 1)\} - \exp\{tr(P - 1)\})f||_\infty .$$

We must show that $\limsup_n \sup \{W(n, t) : t > 0\} = 0$. An expansion gives

$$\begin{aligned} W(n, t) \leq\ & |t^{-1}(e^{-tr_n} - e^{-tr})|\ ||f||_\infty \\ & + \sup \{|r_n e^{-tr_n}\delta^x P^{(n)}(L_a f) - re^{-tr}\delta^x P(L_a f)| : a \in G\} \\ & + \sup \{|t^{-1}Q^x_{n,t}(L_a f)| : a \in G\} , \end{aligned}$$

where $Q^x_{n,t}$ is a difference of two positive measures and satisfies: $||Q^x_{n,t}|| \leq 2$; $||t^{-1}Q^x_{n,t}|| \to 0$ uniformly in $n$ as $t \to 0$; $t^{-1}Q^x_{n,t}(g) \to 0$ for each $g \in C(X)$ as $n \to \infty$ uniformly in every compact $t$ subinterval of $(0, \infty)$. We handle these three terms separately. Clearly

$$\limsup_n \sup \{|t^{-1}(e^{-tr_n} - e^{-tr})| : t > 0\} = 0 .$$

Lemma 4.1 shows that the second term is also zero in the limit. For the last terms we let $t$ range over a bounded interval $[t_0, T]$ where $0 < t_0$ and again use Lemma 4.1. If $t_0$ is small enough and $T$ large enough the extra pieces are arbitrarily small, so that for sufficiently large $n$ we can make this last term small too. This completes the proof of Theorem 4.2.

For later use we weaken the hypothesis of Theorem 4.2 by generalizing the concept of a homogeneous process to allow an escape of mass to infinity. Let $X^* = \{X, x^*\}$ be the cannonical one point compactification of $X$ with $x^*$ denoting the point at infinity. Extend the operations of $G$ to $X^*$ as proper maps, so that $G[x^*] = x^*$. In order to use our earlier notation we consider the spaces $C_\infty(X)$, $C_c(X)$ as imbedded in $C(X^*) = C(X)$. A homogeneous transition probability on $(X^*, G)$ is a continuous endomorphism $P: N(X^*) \to N(X^*)$ satisfying (i), (ii) and (iii) of Definition 1.1. It follows as before that $P$ satisfies equations (1.1), (1.2), (1.3) and (1.4) of § 1 with $f \in C(X)$ and $\mu \in N(X^*)$. On the subspace $N(X)$ of $N(X^*)$, $P$ can be expressed as

$$P = P' + kS , \tag{4.1}$$

where $P'$ is a homogeneous transition probability on $(X, G)$ and $\mu S = \mu(X)\delta^{x^*}$.

A homogeneous process on $(X^*, G)$ is a weakly continuous one-parameter semi-group of homogeneous transition probabilities of norm one acting on $C(X)$, or as adjoints on $N(X^*)$. If $X$ is not compact it is easy to show that whenever $x \neq x^*$ there is an $r \geq 0$, such that

$$\delta^x T_t(E) = \delta^x T_t(X \cap E) + \{1 - e^{-rt}\}\delta^{x^*}(E) , \tag{4.2}$$

and

$$\delta^{x^*} T_t(E) = \delta^{x^*}(E) . \tag{4.3}$$

THEOREM 4.3. *If $r_n \to r$ and $P^{(n)}$, $P$ are homogeneous transition probabilities on $(X^*, G)$, such that for each $f \in C(X)$, $P^{(n)}f \to Pf$ pointwise as $n \to \infty$. Then*

$$\exp\{tr_n(P^{(n)} - 1)\} \to \exp\{tr(P - 1)\}$$

*strongly as $n \to \infty$.*

*Proof.* An expansion shows immediately that one only need consider those $f \in C_\infty(X)$ in the criterion for strong convergence. Since for these $f$, $Pf = P'f$, the conclusion of Theorem 4.3 follows by the methods used in the proof of Theorem 4.2.

THEOREM 4.4. *Let* $\exp\{tr_n(P^{(n)} - 1)\}$ *be a sequence of compound Poisson processes on* $(X, G)$. *Suppose* $r_n \leq K < \infty$ *and* $\exp\{tr_n(P^{(n)} - 1)\}$ *converges in the sense of Bernoulli to the homogeneous process* $T_t$ *on* $(X, G)$, *where* $\|T_t\| = 1$. *Then* $T_t = \exp\{tr(P - 1)\}$ *is a compound Poisson process on* $(X, G)$, *and for some subsequence* $\{n_i\}$ *of* $\{n\}$, $r_{n_i} \to r$, *and* $P^{(n_i)}f \to Pf$ *pointwise whenever* $f \in C(X)$

*Proof.* Choose a subsequence $\{n_i\}$ so that $r_{n_i}$ approaches some $r$ and $\delta^x P^{(n_i)} \to \delta^x P' + k\delta^{x^*}$ in the weak topology generated by $C(X)$. By Theorem 4.2 it suffices to show $k = 0$. If $k \neq 0$ it follows from Theorem 4.3. that

$$T_t = \exp\{tr(P' - 1)\} + (1 - e^{-ktr})\delta^{x^*} ,$$

violating the condition that $\|T_t\| = 1$ on $(X, G)$.

The following theorem is known for homogeneous processes on a commutative group. A proof based on an analysis of characteristic functions is given in Bochner [1, p. 76].

THEOREM 4.5. *Let* $X$ *be a discrete space so that every one point subset of* $X$ *is an open set, and let* $T_t$ *be a homogeneous process on* $(X, G)$. *Then* $T_t$ *is a compound Poisson process.*

*Proof.* Select a sequence of compound Poisson processes

$$\exp\{tr_n(P^{(n)} - 1)\}$$

converging strongly to $T_t$ and normalized by $\delta^x P^{(n)}(\{x\}) = 0$. Using the discretness of $X$ the characteristic function of the set $X - \{x_0\}$, $I_{X-\{x_0\}}$, is in $C(X)$. For any $\varepsilon > 0$ we can choose an $f \in D(A)$, such that $\|I_{X-\{x_0\}} - f\|_\infty < \varepsilon$. By addition of the constant $-f(x_0)$ to $f$ we obtain an $h = f - f(x_0) \in D(A)$ which vanishes at $x_0$ and is greater than $1 - 2\varepsilon$ elsewhere. Using this $h$ in the definition of strong convergence we find that on the one hand

$$\|t^{-1}(\exp\{tr_n(P^{(n)} - 1)\} - T_t)(h)\|_\infty < \delta_n \to 0$$

as $n \to \infty$, while on the other this first expression approaches

$$\|r_n(P^{(n)} - 1)(h) - Ah\|_\infty$$

as $t \to 0$. From this it follows that

$$(1-2\varepsilon)r_n \le r_n\delta^{x_0}P^{(n)}(h) \le \delta_n + Ah(x_0) ,$$

which implies

$$r_n \le \{\delta_n + Ah(x_0)\}(1-2\varepsilon)^{-1} \le K < \infty .$$

Then $T_t$ is a compound Poisson process by Theorem 4.4.

5. **Processes with continuous paths.** We define in this section a class of processes having continuous paths, and we give conditions for a process to belong to this class.

DEFINITION 5.1. We say that a homogeneous process $T_t$ on $(X, G)$ has the property $G_w$ $\{G_s\}$, or belongs to the class of processes designated by the symbol $G_w$ $\{G_s\}$, if, and only if, for any fixed $x_0 \in X$ and any neighborhood $N_{x_0}$ of $x_0$, there is a sequence of compound Poisson processes $\exp\{tr_n(P^{(n)}-1)\}$ converging in the sense of Bernoulli {strongly} to $T_t$, with support $(\delta^{a[x_0]}P^{(n)}) \subset a[N_{x_0}]$ for all $a \in G$

THEOREM 5.1. *If $T_t \in G_s$ and satisfies Property* II *of* § 2, *it follows that $\delta^x T_t(N_x^c) = o(t)$ as $t \to 0$ for every measurable neighborhood $N_x$ of $x$.*

*Proof.* Choose a compact neighborhood, $V_x$, of $x$ contained in the interior of $N_x$, and in accordance with Corollary 2 of Theorem 2.2 let $f \in D(A)$, $0 \le f \le 1$, $f(V_x) = 0$, and $f(N_x^c) = 1$. Now select a sequence of compound Poisson processes $\exp\{tr_n(P^{(n)}-1)\}$ converging strongly to $T_t$ and satisfying support $(\delta^x P^{(n)}) \subset V_x$. It suffices to show $(T_t f)(x) = o(t)$ as $t \to 0$, or that for every $t_n \to 0$, $\limsup_n t_n^{-1}(T_{t_n} f)(x) = 0$. It is no restriction to assume $t_n r_n$ and $t_n r_n^2$ are both less than $n^{-1}$. The condition of strong convergence applied to $f$, then shows that

$$\begin{aligned}\limsup_n t_n^{-1}(T_{t_n} f)(x) &= \limsup_n t_n^{-1}[\exp\{t_n r_n(P^{(n)}-1)\}(f)](x)\\ &\le \limsup_n t_n^{-1}e^{-t_n r_n}(e^{t_n r_n} - 1 - r_n t_n)\\ &\le \limsup_n t_n^{-1}e^{-t_n r_n}(t_n r_n)^2\\ &\le \limsup_n n^{-1}e^{-t_n r_n} = 0\end{aligned}$$

as we desired to prove.

In the proof of a partial converse to Theorem 5.1 we will need the following lemma.

LEMMA 5.1. *If $t, \varepsilon, \delta > 0$* and *$k \ge (te^2+1)\varepsilon^{-1} + 1 - \log\delta$,* then

$$\sum_{n=k}^{+\infty} \exp\{-t\varepsilon^{-1}\}(t\varepsilon^{-1})^n/n! \le \varepsilon\delta .$$

*Proof.* Rather than prove this lemma in detail we will indicate a method of proof. If the sum is overestimated by an integral and the $n!$ in that integral is underestimated using $(2\pi)^{1/2}n^{n+(1/2)}e^{-n} \leq n!$, one obtains the statement that $k \geq \varepsilon^{-1}te^2$ implies

$$\sum_{n=k+1}^{+\infty} \exp\{-t\varepsilon^{-1}\}(t\varepsilon^{-1})/n! \leq (2\pi k)^{-1/2}\exp\{-(k+t\varepsilon^{-1})\} .$$

The desired conclusion follows easily from this estimate.

Lemma 5.1 gives more than we need to prove a weak converse to Theorem 5.1, but not enough to prove a converse. The best we are able to achieve in this direction using the above estimate is stated below.

THEOREM 5.2. (i) *If* $\delta^x T_t(N_x^c) = o(t)$ *as* $t \to 0$ *for every neighborhood* $N_x$ *of* $x \in X$, $T_t \in G_w$.

(ii) *If* $\delta^x T_t(N_x^c) = o(t^2)$ *as* $t \to 0$ *fore very neighborhood* $N_x$ *of* $x \in X$, $T_t \in G_s$.

*Proof.* These results are stated together because their proofs parallel one another. Let $T_t$ be (at first) any homogeneous process, and suppose $N_{x_0}$ is a compact neighborhood of $x_0$. Put $W_{x_0} = \bigcap_{a \in G_{x_0}} a[N_{x_0}]$ and $W_y = b[W_{x_0}]$ where $b[x_0] = y$. This choice of $W_{x_0}$ insures that $W_y$ is well-defined, $W_{a[x_0]} \subset a[N_{x_0}]$, and $W_x$ is a compact neighborhood of $x$. Now define a homogeneous transition probability, $P_\varepsilon$, by $\delta^y T_\varepsilon(F) = \delta^y T_\varepsilon(W_y \cap F)$. Let $s(\varepsilon) = \delta^x T_\varepsilon(W_x^c)$, and $q(\varepsilon) = 1 - s(\varepsilon) = \delta^x T_\varepsilon(W_x)$. We show that the compound Poisson processes $\exp\{t\varepsilon^{-1}([1-s(\varepsilon)]^{-1}P_\varepsilon - 1)\}$ approximate $T_t$ in the desired sense as $\varepsilon \to 0$. Since support $(\delta^x P_\varepsilon) \subset W_x \subset a[N_{x_0}]$ whenever $a[x_0] = x$, this will be sufficient to prove Theorem 5.2. Since it is known that the processes $\exp\{t\varepsilon^{-1}(T_\varepsilon - 1)\}$ approximate $T_t$ in the strong sense, it suffices to show

$$U(t, \varepsilon) = \|\exp\{t\varepsilon^{-1}[q(\varepsilon)^{-1}P_\varepsilon - 1]\} - \exp\{t\varepsilon^{-1}(T_\varepsilon - 1)\}\| \to 0$$

as $\varepsilon \to 0$ for fixed $t$ in conclusion (i), and that $t^{-1}U(t, \varepsilon) \to 0$ uniformly in $t$ as $\varepsilon \to 0$ in conclusion (ii). Calculation shows

$$U(t, \varepsilon) \leq \sum_{n=0}^{+\infty} \exp\{-t\varepsilon^{-1}\}(t\varepsilon^{-1})^n B(n, \varepsilon)/n! ,$$

where

$$B(n, \varepsilon) = \|\{q(\varepsilon)^{-1}P_\varepsilon\}^n - T_\varepsilon^n\| .$$

Since $\delta^x(T_\varepsilon^n - P_\varepsilon^n) \geq 0$, it follows that $\|T_\varepsilon^n - P_\varepsilon^n\| = 1 - q(\varepsilon)^n$; so

$$B(n, \varepsilon) \leq 2q(\varepsilon)^{-n}\{1 - q(\varepsilon)^n\} .$$

For large $n$ there is a better bound because $B(n, \varepsilon) \leq 2$ in any case. We will also use the fact that $1 \geq s \geq 0$ implies $\log(1-s) \geq -s(1-s)^{-1}$.

*Proof of* (i). Let $s(\varepsilon) = f(\varepsilon)\varepsilon$ where $f(\varepsilon) \to 0$ as $\varepsilon \to 0$, and suppose $t$ is fixed and $\varepsilon$ small enough so that $\varepsilon^{-1}K \geq (t\varepsilon^2 + 1)\varepsilon^{-1} + 1$. Putting $\delta = 1$ in Lemma 5.1, and using the substitution

$$E(n, r) = \exp\{r\varepsilon^{-1}\log[1 - \varepsilon^n f(\varepsilon)]\} ,$$

we find

$$\begin{aligned} U(t, \varepsilon) &\leq \sum_{n \geqq K\varepsilon^{-1}} \{\text{terms above}\} + \sum_{K\varepsilon^{-1} < n} \{\text{terms above}\} \\ &\leq E(1, -K)\{1 - E(1, K)\} + 2\varepsilon \to 0 \end{aligned}$$

as $\varepsilon \to 0$ .

*Proof of* (ii). For $\delta > 0$. We will show $\limsup_{\varepsilon \to 0} \sup_{t>0} t^{-1}U(t, \varepsilon) \leq \delta$.

Since $U(t, \varepsilon) \leq 2$, we only need to consider those $t \leq 2\delta^{-1}$. For these $t$ choose $K > 0$ and a range of $\varepsilon$'s sufficiently close to zero, so that the hypothesis of Lemma 5.1 is satisfied for all $\varepsilon$'s and $t$'s which come under consideration. Let $s(\varepsilon) = \varepsilon^2 f(\varepsilon)$ where $f(\varepsilon) \to 0$ as $\varepsilon \to 0$. Computing as before and noting that

$$\sum_{n>0} \exp\{-t\varepsilon^{-1}\}(t\varepsilon^{-1})^{n-1}\varepsilon^{-1}/n! \leq \varepsilon^{-1} ,$$

we find

$$\begin{aligned} t^{-1}U(t, \varepsilon) &\leq \varepsilon^{-1}\{1 - E(2, K)\}E(2, -K) + \varepsilon K\varepsilon^{-1} \sum_{n > K\varepsilon^{-1} - 1} \{\text{terms above}\} \\ &\leq \varepsilon^{-1}(1 - \exp\{-Kf(\varepsilon)\varepsilon[1 - \varepsilon^2 f(\varepsilon)]^{-1}\}E(2, -K) \\ &\quad + 2\varepsilon\delta K^{-1} \to 0 \end{aligned}$$

as $\varepsilon \to 0$, which completes the proof of Theorem 5.2.

In Euclidean spaces it is easy to see that $\delta^0 T_t(N_0^c) = o(t)$ implies $\delta^0 T_t(N_0^c) = o(t^m)$ for any $m \to 0$ as $t \to 0$. Accordingly, in these spaces the $o(t)$ condition and the $G_s$ condition are equivalent. This may be true in general but we will not explore the question further here.

**6. The commutative case.** This section contains an independent proof of Property II that is more general than the proof in § 3 in that $X$ need not be separable, and more restrictive because $G$ must be commutative. Let $G = \{x, y, z, \cdots\}$ be a commutative, Hausdorff, locally compact topological group, and let $\hat{G} = \{a, b, c, \cdots\}$ be its character group. Denote the Fourier-Stieltjes transform of the measure $\mu \in N(G)$ by $\check{\mu}(x) = \int (x, a)\mu(da)$, and write $\mu_n \xrightarrow{c} \mu$ if, and only if, for every $g \in C(G)$, $\mu_n(g) \to \mu(g)$. If $f_n$, $f$ are uniformly continuous bounded functions on $G$, we write $f_n \xrightarrow{p} f$ as a substitute for the fact that $f_n \to f$ pointwise and uniformly on compact sets. It is convenient to denote the Haar measure of a Borel set $E$ by $|E|$, and to let $N_+(\hat{G})$ be an abbreviation for the cone of positive measures in $N(\hat{G})$. We need the following results which we state without proof from harmonic analysis.

(a) If $\mu_n, \mu \in N_+(\hat{G})$, then $\mu_n \xrightarrow{c} \mu$ is equivalent with $\check{\mu}_n \xrightarrow{c} \check{\mu}$.

(b) If $\mu_n \in N_+(\hat{G})$ and $\check{\mu}_n \to f$ pointwise, $f$ being continuous at $e \in G$, then $f$ is continuous on $G$ and there is a $\mu \in N_+(G)$, such that $f = \check{\mu}$ and $\check{\mu}_n \xrightarrow{c} \check{\mu}$.

(c) If $\mu_n \in N_+(\hat{G})$, $\check{\mu}_n \to \hat{\mu}$ almost everywhere with respect to the Harr measure on $G$, and $\check{\mu}_n(e) \to \mu(e)$; then $\check{\mu}_n \xrightarrow{p} \check{\mu}$.

(d) If $G$ is connected, $\mu_n \in N_+(\hat{G}), \|\mu_n\| \leq M < +\infty$, and $\check{\mu}_n \to M$ on a set, $A$, of positive Harr measure in $G$; then $\check{\mu}_n \xrightarrow{c} M$ on $G$.

(e) If $\|\mu_n\| \leq M < +\infty$, then $\mu_n \to \mu$ if, and only if,

$$\int_E \check{\mu}_n(x)\,|dx| \to \int_E \mu(x)\,|dx|$$

for every compact $E \subset G$.

Let $P^+(\hat{G})$ be the cone of regular not necessarily bounded Borel measures, $Q$, on $\hat{G}-\{\hat{e}\}$ for which the integral $Q'(x) = \int \{1 - (x, a)\} Q(da)$ exists and is continuous on $G$. If $U$ is a compact symmetric neighborhood of $e \in G$, we define the function $h$ on $G$ by

$$h(a) = |U|^{-1} \int_U \{1 - (y, a)\}\,|dy|,$$

and observe that this $h$ has the following properties:

(i) $h$ is real valued and continuous;

(ii) $0 \leq h(a) < 2$;

(iii) it $G$ is connected, $h(a) = 0$ implies $a = \hat{e}$;

(iv) $h(a) \to 1$ as $a \to$ infinity, so that $h \in C(\hat{G})$.

By choosing if necessary a new Haar measure we can assume $|U| = 1$. We do this in the proofs below, and in addition denote the measure $\mu(F) = \int_F h(a)Q(da)$ by $hQ$.

LEMMA 6.1. *If* $Q \in P^+(\hat{G}), 0 \leq \int h(a)Q(da) = \|hQ\| < +\infty$.

*Proof.* $\int_U Q'(x)|dx| = \int_U |dx| \int_{\hat{G}-\{\hat{e}\}} \{1-(x,a)\}Q(da) = \int h(a)Q(da).$

The first term above is clearly bounded, so $\|hQ\|$ is finite.

Let $\{Q_q, q \in L\}$ be a directed set with $Q_q \in P^+(\hat{G})$, and $g$ a continuous complex valued function on $G$. We say $Q'_q \to g$ boundedly on every compact set if $Q'_q \to g$ pointwise, and for every compact set $D \subset G$ there is a positive $K_D$ and an $N_D \in L$, such that $q > N_D$ and $x \in D$ imply $|Q'_q(x)| \leq K_D$. We denote this by $Q'_q \xrightarrow{b} g$. The following theorem collects the Fourier analysis we need for $P^+(\hat{G})$.

THEOREM 6.1. *For $Q \in P^+(\hat{G})$ define $Q^h(x) = \int (x, a)h(a)Q(da)$, and let $Q_q$ be a directed sequence from $P^+(\hat{G})$. Then*

(1) $Q'_q \xrightarrow{b} Q' \Rightarrow Q^h_q \to Q^h$ pointwise $\Leftrightarrow Q^h_q \xrightarrow{p} Q^h \Leftrightarrow hQ_q \xrightarrow{c} hQ$;

(2) $Q'_q \xrightarrow{b} g$ continuous $\Rightarrow Q^h_q \to$ some continuous $f \Leftrightarrow Q^h_q \xrightarrow{p} f \Leftrightarrow hQ_q \xrightarrow{c}$ some $\mu \in N_+(\hat{G})$.

*Proof.* Only the first implications need proof in each case. Calculation shows

$$\begin{aligned} Q^h(x) &= \int (x, a)h(a)Q(da) = \int_U |dy| \int Q(da)\{(x, a)[1-(y, a)]\} \\ &= \int_U |dy| \int Q(da)\{[(x, a)-1]+[1-(yx, a)]\} \\ &= \int_U Q'(yx)|dy| - Q'(x)\,. \end{aligned}$$

The implications follow after an application of the Lebesgue bounded convergence criterion.

THEOREM 6.2. *Let $T_t$ be a homogeneous process on $G$. Put $\delta^{\hat{e}} T_t = w_t$. Then $\hat{w}_t(x) = e^{tf(x)}$ where $f$ is a uniquely determined continuous complex function on $G$.*

*Proof.* This is an immediate consequence of the semi-group property $w_t(x)\check{w}_s(x) = \check{w}_{t+s}(x)$ and Property I, $w_t \xrightarrow{p} 1$ as $t \to 0$. For small enough values of $t$ we can even define $f$ directly at a particular $x \in G$ by putting $f(x) = t^{-1}\log \check{w}_t(x)$ and using the principal branch of the logarithm.

We can now prove a slightly weaker statement than Property II for commutative groups.

THEOREM 6.3. *If $T_t$ is a homogeneous process on the locally compact abelian group $\hat{G}$, there is an $m \in N_+(\hat{G})$, such that $t^{-1}h\delta^{\hat{e}}T_t \xrightarrow{c} m$ as $t \to 0$.*

*Proof.* $\int(x, a)\delta^{\hat{e}}T_t(da) = e^{-tg(x)}$ from Theorem 6.2. If we put $Q_t = t^{-1}\delta^e T_t$,

$$t^{-1}(1 - \exp\{-tg(x)\}) = Q_t'(x) \xrightarrow{b} g(x)$$

as $t \to 0$, and Theorem 6.1 shows $hQ_t \to$ some $m \in N_+(\hat{G})$.

In general it is not possible to choose a single $h$ in the above manner which vanishes just at $\hat{e}$. For example, if $G$ is the real line with the discrete topology, denoted by $R_d$, then in order to satisfy the condition $|U| < +\infty$, $U$ must be a finite point set and generate a proper subgroup $K$ of $R_d$. In this case the associated $h$ will surely vanish on $K^* = R_d/K$ which is certainly not equal to $\{\hat{e}\}$. By $K^*$ we mean, as usual, the set of characters identically equal to one on $K$. By varying $U$ we can, however, prove Property II. In the following we denote by $J$ the class of $h$'s defined by $h(a) = |U|^{-1}\int_U \{1 - (y, a)\}|dy|$ for some compact symmetric neighborhood $U$ of $e$.

LEMMA 6.2. *If $h \in J$, then $H_h = \{a : h(a) = 0\}$ is a compact subgroup of $\hat{G}$.*

*Proof.* $H_h$ is closed and compact because $h \to 1$ as $a \to$ infinity. Since $a \in H_h$ is equivalent to $(x, a) = 1$ for all $x \in U$, $H_h$ is also a subgroup of $\hat{G}$

LEMMA 6.3. *For each $a \in \hat{G} - \{\hat{e}\}$ there is an $h \in J$ with $h(a) > 0$.*

*Proof.* Choose a $y \in G$ for which $(y, a) \neq 1$, and for any $U$ satisfying the above conditions construct a new $U' = yU \cup U \cup Uy^{-1}$. $U'$ satisfies the required conditions and since $(x, a) \neq 1$ for every $x \in U'$, the $h$ corresponding to $U'$ satisfies $h(a) > 0$.

With this preparation Property II is immediate.

THEOREM 6.4. *Given any open neighborhood $N_{\hat{e}}$ of $\hat{e} \in \hat{G}$ and any homogeneous process $T_t$ on $\hat{G}$, $t^{-1}\delta^e T_t \xrightarrow{c}$ some $\mu \in N_+(N_{\hat{e}}^c)$ when restricted to $N_{\hat{e}}^c$.*

*Proof.* Compactify $\hat{G}$ by adding a point at infinity. Then taking complements in the compactified space $\{H_h^c \cap N_{\hat{e}}^c : h \in J\}$ forms an open covering of the compact $N^c_{\hat{e}}$. Choose a finite subcovering $\{H^c_{h_i} \cap N^c_{\hat{e}}, 1 \leq i \leq n\}$, then $\sum_{i=1}^n h_i > 0$ on $N^c_{\hat{e}}$. By Theorem 6.3 $h_i t^{-1}\delta^{\hat{e}} T_t \xrightarrow{c} m_{h_i}$ on $\hat{G}$ as $t \to 0$ and, accordingly,

$$\sum_{i=1}^n h_i t^{-1}\delta^{\hat{e}} T_t \xrightarrow{c} \sum_{i=1}^n m_{h_i} ,$$

so that

$$t^{-1}\delta^{\hat{e}} T_t \xrightarrow{c} \left(\sum_{i=1}^n h_i\right)^{-1} \sum_{i=1}^n m_{h_i}$$

on $N^c_{\hat{e}}$.

**7. Subordination of stochastic processes.** In this section we give a new definition of the concept of subordination introduced by Bochner for homogeneous processes on Euclidean spaces. Using this definition we are able to show that there are a great many processes which are not subordinate to any but themselves. We introduce this topic by discussing subordination from the characteristic function viewpoint.

Bochner calls a map $v: (0, +\infty) \to (0, +\infty)$ a completely monotone mapping if for every completely monotone function $f: R^+ \to R$, the function $f \cdot v: R^+ \to R$ is completely monotone. A mapping, $v$, is a completely monotone mapping if, and only if, $dv/dx$ is a completely monotone function, or equivalently $e^{-tv}$ is a completely monotone function for every $t > 0$. If $v$ is a completely monotone mapping, it can be extended to a map $v: R^+ + iR \to R^+ + iR$ of the closed right half complex plane into itself, which is analytic on the interior of its domain, and is of the form $v(z) = c_0 + cz + Q_v(z)$ where $c_0, c \geq 0$;

(i) $\Re\mathrm{e}\{Q_v(z)\} \geq 0$;

(ii) $\Re\mathrm{e}\{Q_v(z)\} = 0, \Re\mathrm{e}(z) > 0$ implies $Q_v \equiv 0$;

(iii) $\Re\mathrm{e}\{Q_v(z)\} = 0, \Re\mathrm{e}\{z\} = 0$ implies $\Im\mathrm{m}\{Q_v(z)\} = 0$;

(iv) $\Re\mathrm{e}\{Q_v(z)\} = o(|z|)$ as $|z| \to +\infty$ with $\Re\mathrm{e}(z) \geq 0$. If $v$ is a completely monotone mapping $v(0+)$ exist, and we call $v$ a subordinator when $v(0+) = 0$. $v$ is a subordinator if, and only if, $e^{-tv(x)} = \int_0^\infty e^{-xs}\pi_t(ds)$, where $\pi_t \geq 0$, $\|\pi_t\| = 1$, $\pi_0(\{0\}) = 1$, $\pi_t{}^*\pi_s = \pi_{s+t}$, and $s \to t$ implies $\pi_s \to \pi_t$. If $v$ is a subordinator and $e^{-th(x)} = \int (x, a)\delta^e T_t(da)$ is the family or Fourier-Stieltjes transforms of a homogeneous process on $\hat{G}$, then $\exp\{-tv[h(x)]\}$ is also the family of Fourier-Stieltjes transforms of a homogeneous process on $\hat{G}$.

DEFINITION 7.1. (Bochner) A process $e^{-th(x)} = \int (x, a)\delta^{\hat{e}} T_t(da)$ is called *subordinate* to a process $e^{-th(x)} = \int (x, a)\delta^{\hat{e}} T_t(da)$ if, and only if, one of the three following equivalent conditions is satisfied.

(1) There is a subordinator $v$, with $e^{-tv(x)} = \int_0^\infty e^{-sx}\pi_t(ds,)$ such that

$$v\{h(x)\} = h(x).$$

(2) $e^{-th(x)} = \int_0^\infty e^{-sh(x)}\pi_t(ds)$ .

(3) $\hat{\delta^e} T_t = \int_0^\infty \hat{\delta^e} T_s \pi_t(ds)$ .

If $T_t$ is a homogeneous process with infinitesimal generator $A$ and $A_\varepsilon = \varepsilon^{-1}(T_\varepsilon - 1)$, then the equation $\exp\{-tv(-A_\varepsilon)\} = \int \exp\{sA_\varepsilon\}\pi_t(ds)$ shows that subordination leads from a process with infinitesimal generator $A$ to a process with infinitesimal generator $-v(-A)$ when suitable interpreted. Bochner has proven that an infinitely decomposable process with a Fourier-Stieltjes transform of the form $\exp\{-t(bx^2 + ib'x)\}$ on the real line is not subordinate to any processes but those with ch.f.$(X_t) = \exp\{-tc(bx^2 + ib'x)\}$ and $c > 0$. In this case subordination is simply a linear change of time scale. In general we denote the relationship $T_t'$ is subordinate to $T_t'$ by $T_t' < T_t$. If processes differing only by a linear change of time scale are identified, it is easy to show using Bochner's result that the relation $<$ is a proper partial ordering of homogeneous processes.

An alternative definition of subordination rests on a probabilistic choice of time scale. If $X(t, \omega)$ is a measurable stochastic process and $Y(t, \omega)$ is a non negative stochastic process which is independent of $\{X(s, \omega)\}$, and whose paths are almost surely non decreasing in $R^+$, we can form the composite process $Z(t, \omega) = X(Y(t, \omega), \omega)$. Under special circumstances this composition corresponds to subordination of the $X(t, \omega)$ process by the $Y(t, \omega)$ process. In general the transformation $X(t, \omega) \to Z(t, \omega)$ will preserve any of those properties of the $X(t, \omega)$ process which depend only on the order relations of the time scale. For example, if $X(t, \omega)$ is Markov, a semi-martingale, a martingale, or spatially homogeneous, so is $Z(t, \omega)$. The stationarity of $X(t, \omega)$, a property depending not only on the order of the time scale but on the magnitude of certain time differences, will in general not be preserved unless $Y(t, \omega)$ is a homogeneous process (i.e. spatially homogeneous and stationary). In this last case if $Y(0, \omega) = 0$ with probability one, we say $Z(t, \omega)$ is subordinate to $X(t, \omega)$ with subordinator $Y(t, \omega)$.

After making the observation that $P\{Y(t, \omega) \in E \mid Y(0, \omega) = 0\}$ should correspond to $\pi_t(E)$ in Bochner's definition, we can show the coincidence of transition probabilities in the two concepts of subordination. If $X(t, \omega)$ is a homogeneous process on the commutative group $G$ with $X(t, \omega) = e$ and

$$P\{X(t+s,\omega)\in E\,|\,X(s,\omega)\} = \delta^{X(s,\omega)}T_t(E)\ ,$$

then subordinating by $Y(t,\omega)$ according to Definition 7.1, we get $\delta^e T'_t(E) = \int \delta^e T_s(E)\pi_t(ds)$. If we subordinate the same process using the definition in terms of paths, we find

$$\delta^e T'_t(E) = P\{Z(t,\omega)\in E\,|\,Z(0,\omega)=e\}$$

$$\delta^e T'_t(E) = \int P\{X(u,\omega)\in E\,|\,Y(t,\omega)=u\}P\{Y(t,\omega)\in du\}$$

$$= \int P\{X(u,\omega)\in E\}\pi_t(du) = \int \delta^e T_u(E)\pi_t(du)\ .$$

Consider a homogeneous process on a not necessarily commutative locally compact topological group, $G$, described by the transition probabilities $\delta^x T_t$. For such a process, define the set $F(T_t) = \bigcap_{t>0}$ support $(\delta^e T_t)$. $F(T_t)$ is a closed possibly empty semi-group of $G$.

THEOREM 7.1. *Let $T'_t$ be a homogeneous process subordinate to $T_t$ on $G$, and supporse $t^{-1}\delta^e T'_t \to Q_e$ on $G - \{e\}$ in the sense of Theorem* 3.4 *as $t\to 0$. The either $T'_t = T_{ct}$ for some $c>0$, or support $(Q_e) \supset F(T_t)$.*

*Proof.* Since the paths of $Y(t,\omega)$ are non decreasing and accordingly, of finite length in any time interval,

$$E\{\exp\{iuY(t,\omega)\}\} = \exp\{t(icu + \int_0^\infty [e^{ixu}-1]J(dx))\}\ , \tag{7.1}$$

where $c \geq 0$ and $\int_0^\varepsilon xJ(dx) < +\infty$ for any $\varepsilon > 0$. If $J=0$, $Y(t,\omega) = ct$ a.s. and $T'_t = T_{ct}$. If $J \neq 0$, the $Y(t,\omega)$ process will have jumps with probability one, and during any of these jumps there is a positive probability that the $T_t$ process will move from its position to a neighborhood of any specific point in $F(T_t)$. This means the subordinated process may have jumps anywhere in $F(T_t)$, then from Theorem 3.3 $F(T_t) \subset$ support $(Q_e)$.

That support $(Q_e) \neq F(T'_t)$ in general can be seen by subordinating a Bernoulli process on the real line (see Bochner [1] for definitions). It is easy to refine Theorem 7.1.

THEOREM 7.2. *Let $T'_t$ be a homogeneous process subordinate to $T_t$ on $G$. Let $t^{-1}\delta^e T'_t \to Q'_e$ and $t^{-1}\delta^e T_t \to Q_e$ as $t\to 0$ in the sense of Theorem* 3.4. *Let $Y(t,\omega)$ be the subordinating process as in* (7.1). *Then if $c<0$,*

$$\text{support}\ (Q'_e) = \text{Cl}\,[\,\cup\,\{\text{support}\ (\delta^e T_s): s\in \text{support}\ (J)\}\cup \text{support}\ (Q_e)]\ ,$$

*and if $c = 0$ ,*

$$\text{support}\ (Q'_e) = \text{Cl}\,[\,\cup\ \{\text{support}\ (\delta^e T_s): s \in \text{support}\ (J)\}]\ .$$

*In particular,* support $(Q'_e) = \phi$ *if, and only if,* $J = 0$ *and either* $Q_e$ *or* $c$ *equals zero.*

*Proof.* Let $z \in$ support $(\delta^e T_s)$ and $s \in$ support $(J)$. Then given any neighborhood $U_z$ of $z$ there is a neighborhood $V_s$ of $s$, such that $t \in V_s$ implies $\delta^e T_t(U_z) > 2^{-1}\delta^e T_s(U_z) > 0$. We can show this by choosing a compact set $D \subset U_z$ and a function $f \in C_c(G)$, such that $f(D) = 1, f(U_z^c) = 0$ and $0 \leq f \leq 1$. If $D$ is chosen so that $\delta^e T_s(f)$ is close to $\delta^e T_s(U_z)$ the assertion follows from the continuity of $\delta^e T_t(f)$. There is a positive probability that $Y(t, \omega)$ will have a jump while $t \in V_s$ and the $T'_t$ process a corresponding jump in $U_z$. Thus $z \in$ support $(Q'_e)$. If $c > 0$ it is clear from the definition of subordination, and the fact that the paths of a homogeneous process can be chosen to have limits from both the left and the right at all time points, that support $(Q_e) \subset$ support $(Q'_e)$. If, conversely, $y \in$ support $(Q'_e)$, the subordinated paths will have a jump in every neighborhood of $y$ and the only manner in which this can happen is for $y$ to lie in the right hand side of the above expressions.

COROLLARY 1. *If* $F(T_t) =$ support $(\delta^e T_s)$ *for every* $s > 0$, *and* $J \neq 0$, *then* support $(Q'_e) = F(T_t)$. *If* $J = 0$ support $(Q'_e) =$ support $(Q_e)$.

COROLLARY 2. *If all motion in the* $T_t$ *process occurs by jumps, as is the case if* $T_t$ *is compound Poisson, and* $J \neq 0$, *then*

$$\text{support}\ (Q'_e) = F(T_t) = \text{Cl}\,[\,\textstyle\bigcup_{k=1}^{\infty} \{\text{support}\ (Q_e)\}^k]\ .$$

*Proof.* In this case $F(T_t) =$ support $(\delta^e T_s)$ for every $s > 0$ and the first corollary applies.

If $G$ is commutative we can use a different method of description.

THEOREM 7.3. *Let* $T'_t$ *be a homogeneous process subordinate to* $T_t$ *on the commutative group* $G$. *Let* $t^{-1}\delta^e T'_t \to Q'_e$ *and* $t^{-1}\delta^e T_t \to Q_e$ *as* $t \to 0$. *Then*

$$\text{support}\ (Q'_e) \supset \text{support}\ (Q'_e) \cdot \text{support}\ (Q_e)\ .$$

*Proof.* In this case it is clear that

$$\text{support}\ (\delta^e T_t) \cdot \text{support}\ (Q_e) \subset \text{support}\ (\delta^e T\ )$$

and the conclusion follows from Theorem 7.2.

Let us now restrict our attention to Euclidean $n$-space. In $R^n$ we denote the inner product by $\langle x, y\rangle$ and the norm by $|x|$. The characteristic functions of a homogeneous process $X(t, \omega)$ have the form

$$(7.2) \qquad E[\exp\{i\langle z, X(t, \omega)\rangle\}] = \exp\{tS(b, A, F, z)\}:$$

where $b \in R^n$, $A$ is a positive semi-definite linear transformation of $R^n$, $F$ is a positive, bounded, regular Borel measure on $R^n - \{0\}$,

$$(7.3) \qquad S(b, A, F, z) = i\langle b, z\rangle - \langle Az, z\rangle + Q(F, z),$$

and

$$(7.4) \qquad Q(F, z) = \int_{|u|>0}\left[\exp\{i\langle u, z\rangle\} - 1 - \frac{i\langle u, z\rangle}{1+|u|^2}\right]\frac{1+|u|^2}{|u|^2}F(du).$$

Corollary 1 shows that if $X'(t, \omega)$ is a process subordinate to $X(t, \omega)$ in a non trivial manner and

$$(7.5) \qquad E[\exp\{i\langle z, X'(t, \omega)\rangle\}] = \exp\{tS(b', A', F', z)\},$$

then support $(F)$ contains the subspace of $R^n$ orthogonal to $\{x | Ax = 0\}$. If, in particular, $A$ is positive definite support $(F') = R^n$. Theorem 7.3 states

$$\text{support}(F') + \text{support}(F) \subset \text{support}(F').$$

This shows if support $(F')$ is compact that $X(t, \omega)$ is not subordinate to any process but itself and possibly a Bernoulli process of the form $X(t, \omega) = tb$. In the latter case $X'(t, \omega) = Y(t, \omega)b$ and all displacement of the $X'(t, \omega)$ process takes place along the ray $\{sb : s \geq 0\} = R^+b$. These observations are summarized below.

THEOREM 7.4. *Let $X'(t, \omega)$ be a homogeneous process on $R^n$ for which support $(F')$ is compact, then $X'(t, \omega)$ is not subordinate to any process but itself unless support $(F') \subset R^+b$, and $A' = 0$. In the latter case $X'(t, \omega)$ is subordinate to the Bernoulli process $Z(t, \omega) = tb$.*

Theorem 7.4 does not exhaust the results which can be obtained by the above technique. In particular we will improve Theorem 7.4 for the real line.

Let $X'(t, \omega)$ be a homogeneous process on the real line subordinate to $X(t, \omega)$ as above. For convenience put

$$L(f) = \int_{x\neq 0}[f(x)]^{-1}F(dx),$$

and

$$B(s) = [b + L(x)]s + \mathrm{Cl}\left[\bigcup_{n=1}^{+\infty}\{\text{support}(F)\}^n\right]$$

when $L(|x|) < +\infty$. Then if $\delta^0 T_s$ denotes the transition probabilities of $X(t, \omega)$, and $X(t, \omega)$ has no Gaussian component, it is clear that support $(\delta^0 T_s) = B(s)$ when $L(|x|) < +\infty$, and support $(\delta^0 T_s) = R$ otherwise. Using Theorem 7.2 this leads to the following.

THEOREM 7.5. *Let the homogeneous process $X'(t, \omega)$ on the real line be subordinated by $Y(t, \omega)$ to the process $X(t, \omega)$, where their Fourier-Stieltjes transforms are given by* (7.1) *and* (7.2).

(1) *If $A \neq 0$ or $L(|x|) = +\infty$*, support $(F') = R$.

(2) *If $A = 0$ and $L(|x|) < +\infty$*, then

$$\text{support}\,(F') = [\,\cup\,\{B(s) : s \in \text{support}\,(J)\}]\cup \text{support}\,(J)$$

*if $c > 0$*, and

$$\text{support}\,(F') = \cup\,\{B(s) : s \in \text{support}\,(J)\}$$

*if $c = 0$.*

It should be noted that we are using the notation of a multiplicative group, so that $\bigcup_{n=1}^{+\infty} \{\text{support}\,(F)\}^n$ in $B(s)$ refers to the additive semi-group generated by support $(F)$.

If we use the easily proved fact that a closed additive semi-group of the real line which contains both a positive and a negative number is necessarily a subgroup of $R$, the following corollary of Theorem 7.5 is immediate.

COROLLARY. *If $X'(t, \omega)$ is non trivially subordinate to a homogeneous process, then support $(F')$ has one of the forms $(H + S) \cup W$ or $H + S$, where $W$ is a closed set which generates the closed additive semi-group $S$, and $H$ is not empty and is contained in either* $[0, +\infty)$ *or* $(-\infty, 0]$. *If $S$ is not a closed subgroup of $R$ it is contained in either* $[0, +\infty)$ *or* $(-\infty, 0]$.

This rules out, among others, sets like $\{\ldots, -2, -1, 0\} \cup (0, +\infty)$ as the support of the $F$ of a non trivially subordinated process.

## REFERENCES

1. S. Bochner, *Harmonic analysis and the theory of probability*, University of California Press Berkeley and Los Angeles, 1955.

2. J. L. Doob, *Stochastic processes*, John Wiley and Sons, New York, 1953.

3. W. Feller, *The general diffusion operator and positivity preserving semi-groups in one dimension*, Ann. of Math., **60** (1954), 417-436.

4. W. Feller, *Semi-groups of transformations in general weak topologies*, Ann. of Math., **57** (1953), 287-308.

5. E. Hille, *Functional analysis and semi-groups*. Amer. Math. Soc. Colloquium Publications, Vol. 31. New York, 1948.

6. G. A. Hunt, *Positive transformations on group spaces*, Dept. of Math. Cornell University, Ithaca, New York.
7. G. A. Hunt, *Semi-groups of measures on Lie groups*. Trans. Amer. Math. Soc., **81** (1956), 264-293.
8. L. H. Loomis, *An introduction to abstract harmonic analysis*. D. Van Nostrand Co., New York (1953).
9. J. R. Kinney, *Continuity properties of sample functions of Markov processes*, Trans. Amer. Math. Soc. **74** (1939), 280-302

PRINCETON UNIVERSITY

# Pacific Journal of Mathematics

Vol. 9, No. 1 May, 1959