Pacific Journal of Mathematics

IN THIS ISSUE—

Vol. 3, No. 2 **April, 1953**

PACIFIC JOURNAL OF MATHEMATICS

EDITORS

R. M. ROBINSON
University of California
Berkeley 4, California

R. P. DILWORTH
California Institute of Technology
Pasadena 4, California

E. HEWITT
University of Washington
Seattle 5, Washington

E. F. BECKENBACH
University of California
Los Angeles 24, California

ASSOCIATE EDITORS

H. BUSEMANN
HERBERT FEDERER
MARSHALL HALL
P. R. HALMOS
HEINZ HOPF
R. D. JAMES
BØRGE JESSEN
PAUL LÉVY
GEORGE PÓLYA
J. J. STOKER
E. G. STRAUS
KÔSAKU YOSIDA

SPONSORS

UNIVERSITY OF BRITISH COLUMBIA
CALIFORNIA INSTITUTE OF TECHNOLOGY.
UNIVERSITY OF CALIFORNIA, BERKELEY
UNIVERSITY OF CALIFORNIA, DAVIS
UNIVERSITY OF CALIFORNIA, LOS ANGELES
UNIVERSITY OF CALIFORNIA, SANTA BARBARA
UNIVERSITY OF NEVADA
OREGON STATE COLLEGE
UNIVERSITY OF OREGON
UNIVERSITY OF SOUTHERN CALIFORNIA
STANFORD RESEARCH INSTITUTE
STANFORD UNIVERSITY
WASHINGTON STATE COLLEGE
UNIVERSITY OF WASHINGTON

* * *

AMERICAN MATHEMATICAL SOCIETY
NATIONAL BUREAU OF STANDARDS, INSTITUTE FOR NUMERICAL ANALYSIS

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be typewritten (double spaced), and the author should keep a complete copy. Manuscripts may be sent to any of the editors except Robinson, whose term expires with the completion of the present volume; they might also be sent to M. M. Schiffer, Stanford University, Stanford, California, who is succeeding Robinson. All other communications to the editors should be addressed to the managing editor, E. F. Beckenbach, at the address given above.

Authors are entitled to receive 100 free reprints of their published papers and may obtain additional copies at cost.

The *Pacific Journal of Mathematics* is published quarterly, in March, June, September, and December. The price per volume (4 numbers) is $8.00; single issues, $2.50. Special price to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: $4.00 per volume; single issues, $1.25.

Subscriptions, orders for back numbers, and changes of address should be sent to the publishers, University of California Press, Berkeley 4, California.

Printed at Ann Arbor, Michigan. Entered as second class matter at the Post Office, Berkeley, California.

UNIVERSITY OF CALIFORNIA PRESS • BERKELEY AND LOS ANGELES

COPYRIGHT 1953 BY PACIFIC JOURNAL OF MATHEMATICS

AN OPERATIONAL CALCULUS FOR OPERATORS WITH SPECTRUM IN A STRIP

WILLIAM G. BADE

1. Introduction. Let X be a complex Banach space, and T be a closed distributive operator whose domain and range are in X. We suppose the spectrum $\sigma(T)$ of T does not cover the whole plane, and write

$$(\lambda I - T)^{-1} = R_\lambda(T)$$

for $\lambda \not\in \sigma(T)$. In the case that T is bounded, N. Dunford [2] and A. E. Taylor [13] have defined an operational calculus for T by the formula

$$f(T) = \frac{1}{2\pi i} \int_C f(\lambda) R_\lambda(T) d\lambda, \tag{1.1}$$

where f is analytic on $\sigma(T)$, and C is a suitable bounded contour enclosing $\sigma(T)$. Such functions f form an algebra, and the mapping $f \longrightarrow f(T)$ is a homomorphism of this algebra into the algebra of bounded operators on X.

When T is assumed to be closed but not bounded, the problem of developing an operational calculus for T meets with the difficulties that the domain $D(T)$ is a proper subspace, and $\sigma(T)$ is in general unbounded. A modification of (1.1),

$$f(T) = f(\infty)I + \frac{1}{2\pi i} \int_C f(\lambda) R_\lambda(T) d\lambda, \tag{1.2}$$

has been used by Taylor [14] when f is analytic on $\sigma(T)$ and at infinity. Here C is a bounded contour enclosing the singularities of f. Although most of the theory for the bounded case may be carried over, the class of functions f is restricted; and polynomials in T, being unbounded operators, need a separate

Received June 13, 1952. This paper is a revised version of the author's doctoral dissertation submitted to the University of California, Los Angeles. He wishes to thank Professor Angus E. Taylor for constant help and encouragement.

Pacific J. Math. 3 (1953), 257-290

treatment.

In this paper we consider the case that $\sigma(T)$ lies in a strip S of finite width, and $\|R_\lambda(T)\|$ is bounded outside any strip containing S in its interior. This is a common situation for differential operators (for example, $T = d/dt$ in $L_p(-\infty, \infty)$, $p \geq 1$). These assumptions enable us to define an operator corresponding to any function analytic and of finite order in a strip containing S. The operator $f(T)$ is bounded or unbounded depending on the growth behavior of f. In this it resembles the operational calculus for unbounded self-adjoint operators in Hilbert space ([8], [12]), and in fact reduces to it in this case.

In §§ 2-4 the calculus is constructed from a postulated set of conditions on T and $R_\lambda(T)$. If f is absolutely integrable in a strip containing S, the operator $f(T)$ is defined by a variant of formula (1.1),

$$f(T) = \frac{1}{2\pi i}\int_\Gamma f(\lambda)R_\lambda(T)\,d\lambda,$$

where Γ is an infinite contour running up one side of S and down the other. If f is of order $n-2$, roughly speaking, then $f(T)x$ is defined for x in the subspace

$$D_n(T) = (x \mid x, Tx, \ldots, T^{n-1}x \in D(T))$$

by the formula

$$f(T)x = \frac{(\alpha I - T)^n}{2\pi i}\int_\Gamma \frac{f(\lambda)R_\lambda(T)x}{(\alpha - \lambda)^n}\,d\lambda,$$

where α is any point exterior to Γ. Equivalently,

$$f(T)x = \sum_{i=0}^{n-1}\frac{f^{(i)}(0)}{i!}T^i x + \frac{1}{2\pi i}\int_\Gamma \frac{f(\lambda)R_\lambda(T)T^n_x}{\lambda^n}\,d\lambda.$$

The usual homomorphism rules hold, and the results are consistent with those of Taylor. A closed extension of $f(T)$ is obtained which coincides with the Stone-von Neumann operator in the case of self-adjoint T in Hilbert space.

In § 5 we assume a further growth condition on $\|R_\lambda(T)\|$ near $\sigma(T)$, and investigate operators corresponding to bilateral transforms. This section is largely a reformulation for our situation of results of Hille [3, Chap. 15] for an operational calculus for the case that $\sigma(T)$ is confined to a half plane and f is a one-sided transform. In § 7 we prove a theorem on the construction of inverses of such operators by limits of polynomials in T.

As an application we take T the operation of differentiation in the spaces C and L_p $(1 \leq p < \infty)$ on the real line and the unit circle. For the case of the line where $\sigma(T)$ is the imaginary axis, if

$$f(\lambda) = \int_{-\infty}^{\infty} e^{\lambda\xi} G(\xi)d\xi$$

converges absolutely in a strip containing $\sigma(T)$, then

$$f(T)x(t) = \int_{-\infty}^{\infty} G(\xi)x(t-\xi)d\xi.$$

Consider the Stieltjes transform

$$\phi(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \operatorname{sech}\frac{\xi}{2}\, x(t-\xi)d\xi,$$

for which $f(\lambda) = [\cos \pi\lambda]^{-1}$. Writing

$$\cos \pi\lambda = \prod_{k=1}^{\infty}\left(1 - \frac{\lambda^2}{(k-1/2)^2}\right),$$

we see from the inversion theorem of § 7 that

$$\lim_{n\to\infty}\prod_{k=1}^{n}\left(1 - \frac{1}{(k-1/2)^2}\frac{d^2}{dt^2}\right)\phi(t) = x(t), \tag{1.3}$$

where convergence is in the norm topology of any of the spaces mentioned. This example is typical of a class of inversion theorems for which the theorem of §7 gives a uniform treatment. Inversion formulas of this sort have been proved by different methods for $L_2(-\infty, \infty)$ by Pollard [9] and for $C[-\infty, \infty)$ by Widder [16] (see Hirschman and Widder [4, 5, 6 and 7] for extensive results on the corresponding local problem). The case $p \neq 2$ does not seem to have been considered before. Our method also applies to inversion of transforms

$$\int_{-\pi}^{\pi} H(\xi)x(t-\xi)d\xi$$

in the corresponding spaces on the circle.

2. Construction of the calculus. Let T be a closed operator whose domain $D(T)$ is a prescribed subspace. We suppose:

A. The spectrum $\sigma(T)$ lies in the vertical strip

$$-\gamma \leq \sigma \leq \gamma \qquad (\lambda = \sigma + i\tau,\ 0 \leq \gamma < \infty).$$

B. The resolvent $R_\lambda(T) = (\lambda I - T)^{-1}$ satisfies

$$\|R_\lambda(T)\| \leq M(t),\ |\sigma| \geq t,\ t > \gamma.$$[1]

The strip containing $\sigma(T)$ is taken to be vertical merely for convenience. The symbol $\rho(T)$ will denote the resolvent set, and $[X]$ the set of bounded linear operators mapping X into itself. We shall need the following known results:

(a) As T is closed, $R_\lambda(T)$ is in $[X]$ for $\lambda \in \rho(T)$ [14, p. 110].

(b) If $D_n(T)$ are the subspaces defined by

$$D_0(T) = X,\ D_n(T) = (x \mid x, Tx, \cdots, T^{n-1}x \in D(T)) \qquad (n \geq 1),$$

then for any polynomial

$$P(\lambda) = a_0\lambda^n + a_1\lambda^{n-1} + \cdots + a_n \qquad (a_0 \neq 0),$$

the corresponding operator $P(T)$ with domain $D_n(T)$ is closed, and [14, p. 202].

$$R_\lambda(T)P(T)x = P(T)R_\lambda(T)x$$

(c) If $x \in D_n(T)$, and α is any complex number, then

$$R_\lambda(T) = \sum_{i=0}^{n} \frac{(T-\alpha I)^i x}{(\lambda-\alpha)^{i+1}} + \frac{(T-\alpha I)^{n+1} R_\lambda(T)x}{(\lambda-\alpha)^{n+1}}, \tag{2.1}$$

and for any m and n, $R_\lambda^m(T)$ maps $D_n(T)$ one-to-one onto $D_{m+n}(T)$ [14, p. 204-205].

We shall also need the following elementary consequence of the definition

[1] One may show through the Phragmen-Lindelöf theorem (the proof in [15, p. 177] holds for operator-valued functions) that B is implied by the apparently much weaker condition $\|R_\lambda(T)\| = O(e^{\beta|\sigma|})$, $\beta < 1$.

of a closed operator.

LEMMA 2.1. *Let K be closed with domain D, and let*

$$H_n \in [X] \qquad (n = 1, 2, \cdots),$$

with

$$Hx = \lim_{n \to \infty} H_n x$$

defined for each $x \in X$. If $x \in D$, and

$$H_n Kx = KH_n x$$

for each n, then $Hx \subset D$ and $KHx = HKx$.

Our procedure for assigning operators $f(T)$ to functions $f(\lambda)$ will be a variant of the contour integral approach of Dunford and Taylor. It will be convenient to set up the correspondence first for a particular class of functions and use this class to treat less restrictive cases.

DEFINITION 2.1. We denote by $\mathfrak{L}(0, \gamma)$ the class of functions f satisfying:

(a) f is analytic in a strip $-r < \sigma < r$, $r > \gamma$ (r may vary with f).

(b) As $\tau \longrightarrow \pm\infty$, $f(\sigma + i\tau) \longrightarrow 0$ uniformly with respect to σ, $-r_1 \leq \sigma \leq r_1$ for any $r_1 < r$.

(c) $$\int_{-\infty}^{\infty} |f(\sigma + i\tau)| d\tau < \infty \qquad (-r < \sigma < r).$$

The class $\mathfrak{L}(0, \gamma)$ is an algebra, although strictly speaking not an algebra of functions, since the functions f do not have a common domain. To get condition (c) for products, we note that if $f \in \mathfrak{L}(0, \gamma)$, $f(\sigma + i\tau)$ is bounded in τ for fixed σ $(-r < \sigma < r)$. Thus if $|f_1(\sigma + i\tau)| \leq M$ (σ fixed),

$$\int_{-\infty}^{\infty} |f_1(\sigma + i\tau)\, f_2(\sigma + i\tau)| \leq M \int_{-\infty}^{\infty} |f_2(\sigma + i\tau)| d\tau < \infty.$$

For convenience we adopt a convention with regard to contour integrals. If f is analytic in $-r < \sigma < r$, where $r > \gamma$, the symbol $\Gamma_c(\omega)$ will denote a contour composed of the two parallel line segments $\sigma = \pm c$, $-\omega \leq \tau \leq \omega$, where $\gamma < c < r$; the positive sense along $\sigma = c$ will be upward, and that along $\sigma = -c$

downward. The symbol Γ_c will denote the contour obtained by letting $\omega \longrightarrow \infty$.

We now define operators corresponding to functions in $\mathcal{L}(0, \gamma)$.

DEFINTION 2.2. *For* $f \in \mathcal{L}(0, \gamma)$, *we set*

$$f(T)x = \frac{1}{2\pi i} \int_{\Gamma_c} f(\lambda) R_\lambda(T) x d\lambda. \tag{2.2}$$

This formula defines an operator in $[X]$, the integral converging absolutely and uniformly in x. It is easily seen to be independent of c except for the restriction $\gamma < c < r$.

THEOREM 2.1. *If* $f, g \in \mathcal{L}(0, \gamma)$, *then*

(a) $(f + g)(T) = f(T) + g(T)$,

(b) $(fg)(T) = f(T)g(T)$.

Proof. Statement (a) is obvious. To prove (b) let f and g both satisfy the the conditions of Definition 2.1 in the strip $[-r. r]$, and let c and c' be chosen so that $\gamma < c < c' < r$. Using the functional equation

$$R_\lambda(T) - R_\mu(T) = (\mu - \lambda) R_\lambda(T) R_\mu(T), \tag{2.3}$$

we see readily that $f(T)g(T)$ is given by the expression

$$\frac{1}{2\pi i} \int_{\Gamma_c} f(\lambda) R_\lambda(T) d\lambda \, \frac{1}{2\pi i} \int_{\Gamma_{c'}} \frac{g(\mu)}{\mu - \lambda} d\mu$$
$$+ \frac{1}{2\pi i} \int_{\Gamma_{c'}} g(\mu) R_\mu(T) d\mu \, \frac{1}{2\pi i} \int_{\Gamma_c} \frac{f(\lambda)}{\lambda - \mu} d\lambda .$$

Since $c < c'$,

$$\frac{1}{2\pi i} \int_{\Gamma_c} \frac{f(\lambda)}{\mu - \lambda} d\lambda = 0, \quad \frac{1}{2\pi i} \int_{\Gamma_{c'}} \frac{g(\mu)}{\mu - \lambda} d\mu = g(\lambda).$$

Formula (b) now follows.

THEOREM 2.2. *Let* $f \in \mathcal{L}(0, \gamma)$, $x \in D_n(T)$, *and* $P(T)$ *be a polynomial in* T *of degree* n. *Then* $f(T)x \in D_n(T)$, *and*

$$P(T)f(T)x = f(T)P(T)x.$$

Proof. For fixed m let $\pi(i, m)$ $(i = 1, 2, \cdots)$ be a sequence of partitions of the contour $\Gamma_c(m)$ whose meshes go to zero as $i \longrightarrow \infty$. Setting

$$H_{im}x = \sum\nolimits_{\pi(i,\, m)} f(\lambda_k)(\lambda_k - \lambda_{k-1})R_{\lambda_k}(T)x,$$

we have

$$P(T)H_{im}x = H_{im}P(T)x$$

for each i. Now letting $i \longrightarrow \infty$, we see from Lemma 2.1 that

$$H_m x = \frac{1}{2\pi i}\int_{\Gamma_c(m)} f(\lambda)R_\lambda(T)x\,d\lambda \in D_n(T),$$

and

$$P(T)H_m x = H_m P(T)x.$$

We now apply the lemma again as $m \longrightarrow \infty$.

We note also the following useful consequence of (2.3).

LEMMA 2.2. *Let $f \in \mathfrak{L}(0, \gamma)$ and $|\Re(\alpha)| > c$. Then*

$$R_\alpha(T)f(T)x = \frac{1}{2\pi i}\int_{\Gamma_c} \frac{f(\lambda)R_\lambda(T)x}{\alpha - \lambda}\,d\lambda.$$

To assign operators to functions with less restrictive growth properties than those of $\mathfrak{L}(0, \gamma)$, we must overcome the problem of the convergence of the integral in (2.2). As motivation, suppose that $f \in \mathfrak{L}(0, \gamma)$ and $x \in D_n(T)$. Then, by the Lemma 2.2,

$$R_\alpha^n(T)f(T)x = \frac{1}{2\pi i}\int_{\Gamma_c} \frac{f(\lambda)R_\lambda(T)x}{(\alpha - \lambda)^n}\,d\lambda,$$

or, by use of Theorem 2.2,

$$f(T)x = \frac{1}{2\pi i}\int_{\Gamma_c} \frac{f(\lambda)R_\lambda(T)(\alpha I - T)^n x}{(\alpha - \lambda)^n}\,d\lambda.$$

The convergence-producing factor in the denominator suggests the following development.

DEFINITION 2.3. For each $n \geq 0$, let

$$\mathfrak{L}(n, \gamma) = (f \mid f(\lambda)(\alpha - \lambda)^{-n} \in \mathfrak{L}(0, \gamma) \text{ for } |\Re(\alpha)| > \gamma),$$

and let $\mathfrak{L}(\infty, \gamma) = \bigcup_{n=0}^{\infty} \mathfrak{L}(n, \gamma)$.

The definition of $\mathfrak{L}(n, \gamma)$ does not depend on α, since if $f \in \mathfrak{L}(n, \gamma)$ for one α it is in for any other with $|\Re(\alpha)| > \gamma$. We note that

$$\mathfrak{L}(0, \gamma) \subset \mathfrak{L}(1, \gamma) \subset \cdots;$$

$\mathfrak{L}(\infty, \gamma)$ is an algebra.

THEOREM 2.3. *If $f, g \in \mathfrak{L}(n, \gamma)$, then $\alpha f + \beta g \in \mathfrak{L}(n, \gamma)$. If $f \in \mathfrak{L}(m, \gamma)$ and $g \in \mathfrak{L}(n, \gamma)$, then $fg \in \mathfrak{L}(m + n, \gamma)$.*

We omit the proof, which follows from the fact $\mathfrak{L}(0, \gamma)$ is an algebra.

DEFINITION 2.4. For $f \in \mathfrak{L}(n, \gamma)$, and x in $D_n(T)$, we choose α such that $|\Re(\alpha)| > \gamma$, and define

$$f(T)x = \frac{1}{2\pi i} \int_{\Gamma_c} \frac{f(\lambda) R_\lambda(T)(\alpha I - T)^n x}{(\alpha - \lambda)^n} \, d\lambda \tag{2.4}$$

where $c < |\Re(\alpha)|$.

To show that this definition is independent of α, let $f \in \mathfrak{L}(n, \gamma)$, $n \geq 1$, $x \in D_n(T)$, and $|\Re(\alpha)|, |\Re(\beta)| > c$. By Lemma 2.2,

$$\begin{aligned}
R_\alpha^n(T) \frac{1}{2\pi i} \int_{\Gamma_c} & \frac{f(\lambda) R_\lambda(T)(\beta I - T)^n x}{(\beta - \lambda)^n} \, d\lambda \\
&= \frac{1}{2\pi i} \int_{\Gamma_c} \frac{f(\lambda) R_\lambda(T)(\beta I - T)^n x}{(\beta - \lambda)^n (\alpha - \lambda)^n} \, d\lambda \\
&= \frac{1}{2\pi i} \int_{\Gamma_c} \frac{f(\lambda) R_\lambda(T) R_\beta^n(T)(\beta I - T)^n x}{(\alpha - \lambda)^n} \, d\lambda \\
&= R_\alpha^n(T) \frac{1}{2\pi i} \int_{\Gamma_c} \frac{f(\lambda) R_\lambda(T)(\alpha I - T)^n x}{(\alpha - \lambda)^n} \, d\lambda.
\end{aligned}$$

Since $R_\alpha^n(T)$ has an inverse,

$$\frac{1}{2\pi i}\int_{\Gamma_c} \frac{f(\lambda)R_\lambda(T)(\alpha I - T)^n x}{(\alpha - \lambda)^n}\, d\lambda = \frac{1}{2\pi i}\int_{\Gamma_c} \frac{f(\lambda)R_\lambda(T)(\beta I - T)^n x}{(\beta - \lambda)^n}\, d\lambda.$$

We remark also that when $x \in \mathfrak{L}(n, \gamma)$, it is also in $\mathfrak{L}(n+1, \gamma)$; and if $x \in D_{n+1}(T)$, Lemma 2.2 shows that formula 2.1 for both n and $n+1$ yields the same operator.

THEOREM 2.4. *Let* $f \in \mathfrak{L}(m, \gamma)$, $g \in \mathfrak{L}(n, \gamma)$, *and* $x \in D_{m+n}(T)$. *Then* $g(T)x \in D_n(T)$, *and* $f(T)g(T)x = (fg)(T)x$.

Proof. We note that

$$fg \in \mathfrak{L}(m+n, \gamma) \text{ and } g(T)x \in D_m(T)$$

by Theorems 2.2 and 2.3. Now if we write

$$h(\lambda) = f(\lambda)(\alpha - \lambda)^{-m},\ k(\lambda) = g(\lambda)(\alpha - \lambda)^{-n}, \text{ and } |\Re(\alpha)| > \gamma,$$

then, by Theorems 2.1 and 2.2,

$$\begin{aligned} f(T)g(T)x &= h(T)(\alpha I - T)^m k(T)(\alpha I - T)^n x \\ &= (\alpha I - T)^{m+n} h(T)k(T)x \\ &= (\alpha I - T)^{m+n}(hk)(T)x = (fg)(T)x. \end{aligned}$$

We are led to another formula for $f(T)$ in the following way. Suppose first that $f \in \mathfrak{L}(0, \gamma)$ and $x \in D_n(T)$. Then in the integral (2.2) we may substitute

$$R_\lambda(T)x = \sum_{i=0}^{n-1} \frac{T^i x}{\lambda^{i+1}} + \frac{R_\lambda(T)T^n x}{\lambda^n}$$

to obtain [2]

---

[2] A formula of this type for $n = 2$ is used by Hille [3, p. 239].

$$f(T)x = \sum_{i=0}^{n-1} \frac{f^{(i)}(0)T^i x}{i!} + \frac{1}{2\pi i}\int_{\Gamma_c} \frac{f(\lambda)R_\lambda(T)T^n x}{\lambda^n}\,d\lambda.$$

This formula has meaning when $f \in \mathfrak{L}(n, \gamma)$ instead of $\mathfrak{L}(0, \gamma)$, and we shall establish it. In fact we prove more generally:

THEOREM 2.5. *Let $f \in \mathfrak{L}(n, \gamma)$ and $x \in D_n(T)$. If $|\mathfrak{R}(\alpha)| < c$, then*

$$(2.5)\quad f(T)x = \sum_{i=0}^{n-1} \frac{f^{(i)}(\alpha)(T-\alpha I)^i x}{i!} + \frac{1}{2\pi i}\int_{\Gamma_c} \frac{f(\lambda)R_\lambda(T)(T-\alpha I)^n x}{(\lambda-\alpha)^n}\,d\lambda.$$

Proof. We suppose first that $|\mathfrak{R}(\alpha)| > \gamma$, and choose c' such that $\gamma < c' < |\mathfrak{R}(\alpha)| < c$. Then

$$f(T) = \frac{1}{2\pi i}\int_{\Gamma_{c'}} \frac{f(\lambda)R_\lambda(T)(T-\alpha I)^n x}{(\lambda-\alpha)^n}\,d\lambda$$

$$-\frac{1}{2\pi i}\int_C \frac{f(\lambda)R_\lambda(T)(T-\alpha I)^n x}{(\lambda-\alpha)^n}\,d\lambda,$$

where C is a small circle described counterclockwise enclosing α and not intersecting Γ_c or $\Gamma_c{'}$. Substituting from (2.1) in the second integral, we have

$$-\frac{1}{2\pi i}\int_C \frac{f(\lambda)R_\lambda(T)(T-\alpha I)^n x\,d\lambda}{(\lambda-\alpha)^n}$$

$$= \frac{1}{2\pi i}\int_C f(\lambda)R_\lambda(T)x\,d\lambda + \sum_{k=0}^{n-1}\frac{1}{2\pi i}\int_C \frac{f(\lambda)}{(\lambda-\alpha)^k}\,d\lambda\,(T-\alpha I)^k x$$

$$= \sum_{k=0}^{n-1}\frac{f^{(k)}(\alpha)}{k!}(T-\alpha I)^k x.$$

This establishes (2.5) when $|\mathfrak{R}(\alpha)| > \gamma$. However, the right side of (2.5) is independent of α and analytic in α. Thus (2.5) holds in the larger region.

The calculus developed above has the disadvantage that if $P(\lambda)$ is a polynomial of degree n, and hence is at best in $\mathfrak{L}(n+2, \gamma)$, the operator corresponding to $P(\lambda)$ is defined only on $D_{n+2}(T)$. We check that this operator is really the formal polynomial in T in the typical case $f(\lambda) = \lambda$. For x in $D_3(T)$, using the series formula, we have

$$f(T)x = Tx + \frac{1}{2\pi i}\int_{\Gamma_c} \frac{R_\lambda(T)T^3x}{\lambda^2}\, d\lambda,$$

since $f(0) = f''(0) = 0$. But

$$\left\| \int_{\Gamma_c} \frac{R_\lambda(T)T^3x}{\lambda^2}\, d\lambda \right\| \leq 2M(c)\,\|T^3x\| \int_{-\infty}^{\infty} \frac{d\tau}{c^2+\tau^2} \tag{2.6}$$

$$= \frac{2\pi\,\|T^3x\|\,M(c)}{c}.$$

The left side is independent of c, and must vanish since $M(c)$ is nonincreasing.

In passing we note without proof some facts which we shall not need later. If for $x \in D_n(T)$ we write

$$\|x\|_n = \sum_{i=0}^{n} \|T^i x\|,$$

then $D_n(T)$ with this norm is a Banach space $X^{(n)}$; and if $f \in \mathfrak{L}(n, \gamma)$, the operator $f(T)$ defined on $X^{(n)}$ to X is bounded. If $D(T)$ is dense in X (a condition of § 5), then $D_m(T)$ is dense in $X^{(n)}$ for $m > n$.

3. Consistency with Taylor's operators. For an arbitrary closed operator with nonempty resolvent set, Taylor [14] has defined an operator corresponding to any function f analytic on $\sigma(T)$ and at infinity. In our situation, let us denote by $\mathfrak{G}(\gamma)$ the class of functions whose singularities lie in bounded sets in the two half-planes $\sigma < -\gamma$ and $\sigma > \gamma$. If $f \in \mathfrak{G}(\gamma)$, and Ω_1 and Ω_2 are clockwise contours in $\rho(T)$ enclosing these sets, then Taylor's formula defines

$$f[T] = f(\infty)I + \frac{1}{2\pi i}\int_{\Omega_1+\Omega_2} f(\lambda)R_\lambda(T)\,d\lambda. \tag{3.1}$$

This operator is bounded, and the correspondence $f \longrightarrow f[T]$ preserves sums and products. In our theory, $\mathfrak{G}(\gamma) \subset \mathfrak{L}(2, \gamma)$; and the corresponding operator

$f(T)$ are defined only on $D_2(T)$. We shall show that $f[T]$ is an extension of $f(T)$. First we note the relationship between $\mathcal{G}(\gamma)$ and the classes $\mathcal{L}(n, \gamma)$.

LEMMA 3.1. *Let* $f \in \mathcal{G}(\gamma)$. *Then* $f \in \mathcal{L}(k, \gamma)$ $(k = 0, 1)$ *if and only if it has a zero of order at least* $2 - k$ *at infinity.*

Proof. First let $f \in \mathcal{L}(0, \gamma) \cap \mathcal{G}(\gamma)$, and let R be the radius of a circle about the origin containing the singularities of f in its interior. Since clearly $f(\infty) = 0$, we may write

$$f(\lambda) = \frac{a}{\lambda} + \frac{g(\lambda)}{\lambda^2},$$

where g is analytic, and

$$|g(\lambda)| < M \text{ for } |\lambda| \geq R.$$

Then if $a \neq 0$,

$$\int_R^n |f(i\tau)|\, d\tau \geq |a| \log \frac{n}{R} - M \int_R^n \frac{d\tau}{\tau^2},$$

or

$$\lim_{n \to \infty} \int_R^n |f(i\tau)|\, d\tau = \infty,$$

which contradicts Condition (C) of Definition 2.1. Thus f has a zero of order two at infinity. The converse is clear. For the case $k = 1$, we use the foregoing argument on $f(\lambda)(\alpha - \lambda)^{-1}$.

THEOREM 3.1. *Let* $f \in \mathcal{L}(0, \gamma) \cap \mathcal{G}(\gamma)$. *Then* $f(T) = f[T]$.

Proof. Let K be a large circle of radius r centered at zero, and let c be chosen $(c > \gamma)$ so that the singularities of f lie interior to K but exterior to Γ_c. Let

$$\Gamma_c(\omega) \qquad \left(\omega = \sqrt{r^2\ c^2}\right)$$

denote that part of Γ_c cut off by K, and let S be the two arcs of K (described counterclockwise) which lie exterior to Γ_c. In formula (4.1), for $f[T]$ we may take

$$\Omega_1 + \Omega_2 = S + \Gamma_c(\omega).$$

Now, by Lemma 3.1,

$$|f(\lambda)| = O(\lambda^{-2})$$

for large values of λ. Also by Condition B of § 2, $\|R_\lambda(T)\|$ is bounded if $|\sigma| \geq c$. Thus

$$\left\| \int_S f(\lambda) R_\lambda(T)\, d\lambda \right\| = O\left(\frac{1}{r}\right).$$

As

$$\lim_{r \to \infty} \Gamma_c(\omega) = \Gamma_c,$$

the theorem is proved.

COROLLARY 3.1. *If* $f \in \mathcal{G}(\gamma)$, *then*

$$f[T]x = f(T)x$$

for x *in* $D_2(T)$.

Proof. Let

$$g(\lambda) = f(\lambda)(\alpha - \lambda)^{-2}.$$

Then

$$g \in \mathcal{L}(0, \gamma) \cap \mathcal{G}(\gamma) \text{ and } g[T] = g(T).$$

But (see [14, p. 203])

$$f(T)x = (\alpha I - T)^2 g(T) = (\alpha I - T)^2 g[T] = f[T].$$

Similarly one shows if $f \in \mathcal{L}(1, \gamma) \cap \mathcal{G}(\gamma)$ then

$$f(T)x = f[T]x$$

for x in $D(T)$.

4. A closed extension of $f(T)$. Let f be in $\mathcal{L}(n, \gamma)$, but not in $\mathcal{L}(n-1, \gamma)$.

The operator $f(T)$, defined on $D_n(T)$, need not in general be closed on this domain. However, we can describe an extended domain on which it will be closed. As before, we set

$$h(\lambda) = f(\lambda)(\alpha - \lambda)^{-n}, \quad |\Re(\alpha)| > \gamma,$$

and write

$$f(T)x = (\alpha I - T)^n h(T)x.$$

DEFINITION 4.1. We define

$$D(\tilde{f}(T)) = (x \mid h(T)x \in D_n(T)),$$

and define

$$\tilde{f}(T)x = (\alpha I - T)^n h(T)x$$

for x in $D(\tilde{f}(T))$.

Since $h \in \mathfrak{L}(0, \gamma)$, $D_n(T) \subset D(\tilde{f}(T))$ by Theorem 2.2. The inclusion may be proper. We shall need a lemma.

LEMMA 4.1. *Let f in $\mathfrak{G}(\gamma)$ have a zero of order n (but not $n+1$) at infinity. Then $x \in D_m(T)$ $(m \geq 0)$ if and only if $f[T]x \in D_{m+n}(T)$.*

Proof. Necessity is proved by Taylor [14, p. 203]. To prove sufficiency first suppose that $n = 0$, i.e., $f(\infty) \neq 0$, and that

$$f[T]x \in D_m(T).$$

If $x \in D_k(T)$, where $0 \leq k < m$, then

$$f[T]x = f(\infty)x + \frac{1}{2\pi i}\int_{\Omega_1+\Omega_2} f(\lambda)R_\lambda(T)x\,d\lambda$$

$$= f(\infty)x + h[T]x,$$

where $h(\lambda) = f(\lambda) - f(\infty)$. As h has a zero at infinity, $h[T]x \in D_{k+1}(T)$ by the first half of the lemma. But as $k < m$,

$$f[T]x \in D_{k+1}(T) \subset D_m(T),$$

and hence $x \in D_{k+1}(T)$. By repeating the argument, we see that $x \in D_m(T)$. If $n \geq 1$ and

$$f[T]x \in D_{m+n}(T),$$

we set

$$g(\lambda) = (\alpha - \lambda)^n f(\lambda)$$

and note that

$$f[T]x = R_\alpha^n(T)g[T]x.$$

As $R_\alpha^n(T)$ maps $D_{m+n}(T)$ one-to-one onto $D_m(T)$,

$$g[T]x \in D_m(T).$$

But since $g(\infty) \neq 0$, $x \in D_m(T)$ by the case $n = 0$ just proved.

THEOREM 4.1. *The subspace $D(\tilde{f}(T))$ is independent of α, and $\tilde{f}(T)$ with domain $D(\tilde{f}(T))$ is a closed extension of $f(T)$.*

Proof. If $|\Re(\alpha)|$, $|\Re(\beta)| > \gamma$, $\alpha \neq \beta$ we denote by $D_\alpha(\tilde{f}(T))$, $D_\beta(\tilde{f}(T))$, h_α, and h_β the respective domains and functions $f(\lambda)(\alpha-\lambda)^{-n}$ and $f(\lambda)(\beta-\lambda)^{-n}$. Then

$$h_\alpha(\lambda) - h_\beta(\lambda) = \frac{P(\lambda)h_\beta(\lambda)}{(\alpha-\lambda)^n} = g(\lambda)h_\beta(\lambda),$$

where P is a polynomial of degree $n-1$. The function g is in $\mathcal{G}(\gamma)$, and $g[T]$ carries $D_n(T)$ into itself (in fact into $D_{n+1}(T)$) by the lemma. Since

$$(gh_\beta)(T) = g[T]h_\beta(T)$$

(apply $R_\beta(T)$ to both sides),

$$h_\alpha(T) = h_\beta(T) + g[T]h_\beta(T)$$

and

$$D_\beta(\tilde{f}(T)) \subset D_\alpha(\tilde{f}(T)).$$

Symmetry gives the reverse. Now fixing α and associated $h(\lambda)$, let

$$x_n \in D(\widetilde{f}(T)) \qquad (n = 1, 2, \cdots),$$

and suppose

$$\lim x_m = x_0 \text{ and } \lim \widetilde{f}(T)x_m = y_0.$$

Then

$$\lim h(T)x_m = h(T)x_0,$$

as $h(T)$ is continuous. As $h(T)x_m \in D_n(T)$, and $(\alpha I - T)^n$ is closed on $D_n(T)$, $h(T)x_0 \in D_n(T)$; i.e. $x_0 \in D(\widetilde{f}(T))$, and

$$\widetilde{f}(T)x_0 = (\alpha I - T)^n h(T)x_0 = y_0.$$

When

$$f(\lambda) = \sum_{i=0}^{n} a_i \lambda^i \qquad (a_n \neq 0),$$

the operator $\widetilde{f}(T)$ is the formal polynomial in T. For if

$$h(\lambda) = f(\lambda)(\alpha - \lambda)^{-(n+2)}$$

then $h \in \mathcal{G}(\gamma)$, with a zero of order exactly two at infinity. By Lemma 4.1,

$$h[T]x \;(= h(T)x) \in D_{n+2}(T)$$

if and only if $x \in D_n(T)$. Hence

$$D_n(T) = D(\widetilde{f}(T)).$$

If $x \in D_n(T)$,

$$R_\alpha^{n+2}(T)\widetilde{f}(T)x = h[T]x = R_\alpha^{n+2}(T) \sum_{i=0}^{n} a_i T^i x,$$

so

$$\widetilde{f}(T) = \sum_{i=0}^{n} a_i T^i.$$

By similar reasoning one shows that $\widetilde{f}(T) = f[T]$ for $f \in \mathcal{G}(\gamma)$.

We now identify $\widetilde{f}(T)$ in the case that X is a Hilbert space and $A = -iT$ is self-adjoint [8, p. 44]. If $\{E_\tau\}$ $(-\infty < \tau < \infty)$ is the resolution of I associated with A, $D(A) = D(T)$ is the set of x for which

$$\lim_{k \to \infty} \int_{-\infty}^{\infty} \tau e_k(\tau) \, dE_\tau \, x$$

exists, where $e_k(\tau)$ is one if $-k \leq \tau \leq k$ and zero otherwise. If $F(\tau)$ is continuous, we define D_F as the set of x for which

$$F(A)x = \lim_{k \to \infty} \int_{-\infty}^{\infty} F(\tau) e_k(\tau) \, dE_\tau x$$

exists. The operator $F(A)$ is normal, and hence closed on D_F. We write

$$D_F = D(A^n) \text{ if } F(\tau) = \tau^n.$$

Taking $F(\tau) = f(i\tau)$, we easily see that $f(T) = F(A)$ when $f \in \mathcal{L}(0, \gamma)$ (here $\gamma = 0$) and that $D(A^n) = D_n(T)$. Now let $f \in \mathcal{L}(n, \gamma)$,

$$h(\lambda) = f(\lambda)(\alpha - \lambda)^{-n}, \ |\Re(\alpha)| > 0, \text{ and } H(\tau) = h(i\tau).$$

If

$$h(T)x \in D_n(T),$$

that is

$$H(A)x \in D(A^n) = D((\alpha I - iA)^n),$$

then

$$\lim_{k \to \infty} \int_{-\infty}^{\infty} F(\tau) e_k(\tau) \, dE_\tau x$$

$$= \lim_{k \to \infty} \int_{-\infty}^{\infty} (\alpha - i\tau)^n e_k(\tau) \, dE_\tau \int_{-\infty}^{\infty} H(\mu) \, dE_\mu x$$

exists and $x \in D_F$. The argument may be reversed. Thus

$$D_F = D(\tilde{f}(T))$$

and

$$F(A) = (\alpha I - iA)H(A) = (\alpha I - T)^n h(T) = \tilde{f}(T).$$

5. Operators corresponding to bilateral transforms. A class of operators of particular interest in applications is the set corresponding to the subclass $\mathfrak{V}(\gamma) \subset \mathfrak{L}(2, \gamma)$ of absolutely convergent bilateral Laplace-Stieltjes transforms $\int_{-\infty}^{\infty} e^{-\lambda\xi}\, d\beta(\xi)$. By a well-known theorem of Hamburger [17, p. 265], any function in $\mathfrak{L}(0, \gamma)$ is in $\mathfrak{V}(\gamma)$. While functions in $\mathfrak{L}(0, \gamma)$ yield bounded operators, functions from $\mathfrak{V}(\gamma)$ may yield unbounded operators in the absence of additional assumptions on T and $R_\lambda(T)$. In this section and throughout the rest of the paper we shall assume

$$(5.1) \qquad \begin{aligned} &\mathrm{B}': \quad \|R_\lambda(T)\| \le \frac{1}{|\sigma| - \gamma}, \\ &\mathrm{C}: \quad D(T) \text{ is dense in } X,^3 \end{aligned}$$

which will ensure boundedness. The results here are essentially due to Hille; in his book [3, Chap. 15] he constructs a calculus for operators with spectrum in the left half-plane, the class of functions being one-sided transforms converging absolutely in a half-plane containing the spectrum. The details of the construction in our case will differ sufficiently to justify giving an outline of the development. We take the following result from Hille.

THEOREM 5.1. [3, p. 307, 322]. *If T is a closed operator satisfying* A, B′, *and* C, *the formulas*

$$(5.2) \qquad \mathfrak{J}(\xi)x = \begin{cases} \dfrac{1}{2\pi i}\,(C, 1) - \displaystyle\int_{r-i\infty}^{r+i\infty} e^{\lambda\xi} R_\lambda(T)x\, d\lambda & (\xi > 0), \\ x & (\xi = 0), \\ \dfrac{1}{2\pi i}\,(C, 1) - \displaystyle\int_{-r+i\infty}^{-r-i\infty} e^{\lambda\xi} R_\lambda(T)x\, d\lambda & (\xi < 0), \end{cases}$$

[3] Condition C implies $D_n(T)$ is dense for $n > 1$. For if $\lambda \in \rho(T)$, $D_n(T) = R_\lambda^{n-1}(T)D(T)$ is dense in $R_\lambda^{n-1}(T)X = D_n(T)$, and thus by a repetition of the argument is dense in $D(T)$, and hence in X. It will follow from Theorem 5.1 [3, p. 228] that $D_\infty(T) = \bigcap_{n=1}^{\infty} D_n(T)$ is also dense.

where $x \in X$, $r > \gamma$, *define a group of bounded operators* $\mathfrak{J}(\xi)$, $-\infty < \xi < \infty$, *where*

$$\mathfrak{J}(\xi + \eta) = \mathfrak{J}(\xi)\mathfrak{J}(\eta), \quad \|\mathfrak{J}(\xi)\| \leq e^{|\xi|\gamma};$$

$\mathfrak{J}(\xi)$ *is strongly continuous in* ξ, *and, for each* $x \in D(T)$,

$$\lim_{\xi \to 0} \frac{\mathfrak{J}(\xi)x - x}{\xi} = Tx.$$

$R_\lambda(T)$ *has the representation*

$$R_\lambda(T)x = \begin{cases} \int_0^\infty e^{-\lambda\xi}\,\mathfrak{J}(\xi)x\,d\xi & (\sigma > \gamma), \\ -\int_{-\infty}^0 e^{-\lambda\xi}\,\mathfrak{J}(\xi)x\,d\xi & (\sigma < -\gamma), \end{cases} \tag{5.3}$$

$x \in X$, *the integrals converging absolutely.*

The operator $\mathfrak{J}(\xi)$ will prove to be $f(T)$ for $f(\lambda) = e^{\lambda\xi}$.

Let Ψ denote the vector space of complex-valued functions β satisfying:

(a) β is normalized and of bounded variation on $(-\infty, \infty)$.

(b) $\beta(-\infty) = \lim_{\xi \to -\infty} \beta(\xi) = 0$.

(c) There is an $r > \gamma$ (r depending on β), such that $\int_{-\infty}^{\infty} e^{-\lambda\xi}\,d\beta(\xi)$ converges absolutely for $-r \leq \sigma \leq r$.

We denote by Ψ_c the subclass of Ψ of β continuous at zero and write

$$\Psi_+ = (\beta \mid \beta \in \Psi_c,\ \beta(\xi) = 0, \quad \xi \leq 0),$$

$$\Psi_- = (\beta \mid \beta \in \Psi_c,\ \beta(\xi) = \beta(0),\ 0 < \xi).$$

If u denotes the function

$$u(\xi) = \begin{cases} 0 & (\xi < 0), \\ \frac{1}{2} & (\xi = 0), \\ 1 & (0 < \xi), \end{cases}$$

then each $\beta \in \Psi$ has a unique representation

$$\beta = \beta_+ + \beta_- + [\beta(0+) - \beta(0-)]u,$$

where $\beta_+ \in \Psi_+$, $\beta_- \in \Psi_-$. The latter functions are given by

$$\beta_+(\xi) = \beta(\xi) - \beta(0+) \qquad (0 < \xi)$$

and

$$\beta_-(\xi) = \beta(\xi) \qquad (\xi < 0),$$

and $\beta(0-)$ otherwise.

DEFINITION 5.1. A function f is in $\mathfrak{V}(\gamma)$ if and only if

$$f(\lambda) = \int_{-\infty}^{\infty} e^{-\lambda\xi}\, d\beta(\xi),$$

where $\beta \in \Psi$. It is in $\mathfrak{A}(\gamma)$ if and only if $f \in \mathfrak{V}(\gamma)$ and β is absolutely continuous.

We write $f = f_+ + f_- + [\beta(0+) - \beta(0-)]$ corresponding to the decomposition above, and note that

$$(5.4) \qquad f_+(\lambda) = \int_{-\infty}^{0} e^{-\lambda\xi}\, d\beta_+(\xi), \quad f_-(\lambda) = \int_{0}^{\infty} e^{-\lambda\xi}\, d\beta_-(\xi).$$

Thus f_+ and f_- are analytic for $\sigma < r$ and $\sigma > -r$ respectively.

THEOREM 5.2. *If $f \in \mathfrak{V}(\gamma)$, then the bounded operator $f\{T\}$ defined by*

$$(5.5) \qquad f\{T\}x = \int_{-\infty}^{\infty} \mathfrak{T}(-\xi)x\, d\beta(\xi) \qquad (x \in X)$$

is also given by

$$(5.6) \qquad f\{T\}x = \frac{1}{2\pi i}\,(C, 1) - \int_{r-i\infty}^{r+i\infty} f(\lambda) R_\lambda(T)x\, d\lambda$$

$$+ \frac{1}{2\pi i}\,(C, 1) - \int_{-r+i\infty}^{-r-i\infty} f(\lambda) R_\lambda(T)x\, d\lambda.$$

Moreover,

$$f_+\{T\}x = \int_{-\infty}^{0} \mathfrak{I}(-\xi)x\,d\beta_+(\xi) = \frac{1}{2\pi i}\,(C,1) - \int_{r-i\infty}^{r+i\infty} f_+(\lambda)R_\lambda(T)x\,d\lambda,$$

$$f_-\{T\}x = \int_{0}^{\infty} \mathfrak{I}(-\xi)x\,d\beta_-(\xi) = \frac{1}{2\pi i}\,(C,1) - \int_{-r+i\infty}^{-r-i\infty} f_-(\lambda)R_\lambda(T)x\,d\lambda;$$

and if $\beta = u$, $f\{T\}x = x$, *the two integrals of* (5.6) *each yield* $x/2$. *If* $x \in D_2(T)$, $f\{T\}x = f(T)x$ *in the sense of* §2.

We sketch the proof. Consider the first integral of (5.6). After substituting from (5.3) and (5.5), and interchanging the order of integration, we may express the partial integral

$$\frac{1}{2\pi}\int_{-\omega}^{\omega}\left[1+\frac{|\tau|}{\omega}\right] f(r+i\tau)R_{r+i\tau}(T)x\,d\tau$$

$$(5.7) \qquad = \int_0^\infty e^{-r\xi}\,v(\xi,\omega)\,d\beta_+(\xi) + \int_{-\infty}^{0} e^{-r\xi}\,v(\xi,\omega)\,d_{\beta_-}(\xi)$$

$$+\,[(\beta(0+) - \beta(0-)]\,v(0,\omega),$$

where

$$v(\xi,\omega) = \int_{-\infty}^{\infty} v(\alpha)\cdot\frac{2\sin^2\omega(\xi-\alpha)/2\,d\alpha}{\omega(\xi-\alpha)^2}$$

and

$$v(\alpha) = \begin{cases} e^{r\alpha}\,\mathfrak{I}(-\alpha)x & (\alpha < 0), \\ x/2 & (\alpha = 0), \\ 0 & (0 < \alpha). \end{cases}$$

The classical theorem on the Fejer integral holds for vector-valued functions (see [3, p. 49]). Thus

$$\lim_{n\to\infty} v(\xi,\omega) = v(\xi)$$

for each ξ and, in fact, uniformly in any bounded interval of continuity of $v(\xi)$.

Since β_+ and β_- are continuous at zero, one shows easily that the first integral on the right of (5.7) vanishes in the limit, and

$$\frac{1}{2\pi i}\,(C, 1) - \int_{r-i\infty}^{r+i\infty} f(\lambda)R_\lambda(T)x\,d\lambda = \int_{-\infty}^{0} \mathfrak{T}(-\xi)x\,d\xi + \frac{[\beta(0+)-\beta(0-)]}{2}x$$

$$= f_-\{T\}x + \frac{[\beta(0+)-\beta(0-)]}{2}x.$$

The second integral of (5.6) yields $f_+\{T\}x + [\beta(0+)-\beta(0-)]x$, and the sum of the two is $f\{T\}x$.

Finally, when $x \in D_2(T)$ we may substitute

$$R_\lambda(T)x = \frac{x}{\lambda} + \frac{Tx}{\lambda^2} + \frac{R_\lambda(T)T^2x}{\lambda^2}$$

in (5.6). Calculation of residues yields

$$f\{T\}x = f(0)x + f'(0)Tx + \frac{1}{2\pi i}\int_{\Gamma_r} \frac{f(\lambda)R_\lambda(T)T^2x}{\lambda^2}\,d\lambda = f(T)x.$$

It follows from the foregoing theorem that $\mathfrak{T}(\xi)$ is $f\{T\}$ for $f(\lambda) = e^{\lambda\xi}$. If $f_1, f_2 \in \mathfrak{V}(\gamma)$, one shows easily [3, p. 309] that

$$(f_1 f_2)\{T\}x = \int_{-\infty}^{\infty} \mathfrak{T}(-\xi)x\,d\theta(\xi) = f_1\{T\}f_2\{T\}x,$$

where

$$\theta(\xi) = \int_{-\infty}^{\infty} \beta_1(\xi-\eta)\,d\beta_2(\eta).$$

We also note that

$$\|f\{T\}\| \le \int_{-\infty}^{\infty} e^{|\xi|\gamma}\,|d\beta(\xi)|,$$

and if $x \in D_n(T)$, $f \in \mathfrak{V}(\gamma)$ then $f\{T\}x \in D_n(T)$. The proof of the latter follows that of Theorem 2.2.

6. A class of kernels. We shall denote by $\mathfrak{G}_0(\gamma)$ those functions in $\mathfrak{G}(\gamma)$

(see §3) which vanish of infinity. Any f in $\mathcal{G}_0(\gamma)$ is in $\mathcal{Q}(\gamma)$, i.e., is an absolutely convergent bilateral Laplace transform. Our purpose is to characterize the kernels $G(\xi)$ for which

$$f(\lambda) = \int_{-\infty}^{\infty} G(\xi) e^{-\lambda\xi}\, d\xi, \quad f \in G_0(\gamma).$$

For this we shall need certain well-known results.

An entire function F is said to be of exponential type δ if

$$\max_{|z|=r} |F(z)| = O\left(e^{(\delta+\epsilon)r}\right)$$

for every positive ϵ and no negative ϵ. Polya has shown [11, p. 585] that there is a one-to-one correspondence between entire functions of exponential type and functions analytic at infinity as follows: If

(6.1) $$f(\lambda) = \sum_{n=0}^{\infty} \frac{a_n}{\lambda^{n+1}} \qquad (|\lambda| > C)$$

where C is the natural radius of convergence, then

$$f(\lambda) = \int_0^{\infty} e^{-\lambda\xi} F(\xi)\, d\xi,$$

where F is the entire function

(6.2) $$F(\xi) = \sum_{n=0}^{\infty} \frac{a_n \xi^n}{n!} \quad ;$$

F is of exponential type C. Conversely, F determines f. Further let K be the set of singularities of f, and

$$k(\phi) = \max_{\lambda \in K} \mathcal{R}(\lambda e^{-i\phi})$$

be its support function. The function

$$h(\phi) = \overline{\lim_{r\to\infty}} \frac{\log |F(r e^{i\phi})|}{r}$$

is called the indicator of F as it measures the growth of F in the direction ϕ. Polya shows that $k(\phi) = h(-\phi)$.

THEOREM 6.1. *A function f is in $\mathfrak{G}_0(\gamma)$ if and only if it is of the form*

$$f(\lambda) = \int_{-\infty}^{\infty} e^{-\lambda\xi}\, G(\xi)\, d\xi,$$

where

$$G(\xi) = \begin{cases} F_+(\xi) & (0 \le \xi < \infty), \\ F_-(\xi) & (-\infty < \xi < 0), \end{cases}$$

and F_+ and F_- are entire functions of exponential type satisfying

$$|F_+(\xi)| = O(e^{\xi\delta_+}), \quad \delta_+ < -\gamma, \quad as \quad \xi \longrightarrow +\infty,$$

$$|F_-(\xi)| = O(e^{\xi\delta_-}), \quad \delta_- > \gamma, \quad as \quad \xi \longrightarrow -\infty.$$

Proof. Let $f \in \mathfrak{G}_0(\gamma)$. Then we may write

$$f(\lambda) = \frac{1}{2\pi i}\int_{\Omega_1} \frac{f(\lambda)}{\xi - \lambda}\, d\xi + \frac{1}{2\pi i}\int_{\Omega_2} \frac{f(\xi)}{\xi - \lambda}\, d\xi$$

$$= f_-(\lambda) + f_+(\lambda),$$

where Ω_1 and Ω_2 are bounded clockwise contours lying in $\Re(\lambda) > \gamma$ and $\Re(\lambda) < -\gamma$, respectively, which enclose the sets of singularities of f which are in these two half-planes. The functions f_- and f_+ are in $\mathfrak{G}_0(\gamma)$, and the decomposition is unique. Then

$$f_+(\lambda) = \int_0^{\infty} e^{-\lambda\xi}\, F_+(\xi)\, d\xi, \quad \Re(\lambda) > c_+, \tag{6.3}$$

where $F_+(\xi)$ is entire of type c_+. Since the singularities of f_+ lie in $\Re(\lambda) < -\gamma$,

$$k(\pi) = h(0) < -\gamma,$$

and thus

$$|F_+(\xi)| = O(e^{\xi\delta_+}),$$

where $\delta_+ < -\gamma$ as $\xi \longrightarrow +\infty$, and the integral (6.3) converges absolutely for $\Re(\lambda) > -\gamma$.

One shows that

$$f_-(\lambda) = \int_{-\infty}^{0} e^{-\lambda\xi} F_-(\xi)d\xi,$$

where $-F_-(\xi)$ is the entire function associated with f_- by (6.1) and (6.2) by setting

$$\lambda = -\mu, \ g(\mu) = f_-(\lambda),$$

as the singularities of g lie in $\Re(\mu) < -\gamma$.

Conversely, if

$$f_+(\lambda) = \int_{0}^{\infty} e^{-\lambda\xi} F_+(\xi)d\xi,$$

where F_+ is an entire function of exponential type with the indicated order property, f_+ is analytic at infinity and its singularities lie in $\Re(\lambda) < -\gamma$. The case of f_- is similar.

7. An inversion theorem. We now prove a result which, when T is the operator of differentiation in spaces of functions on the real line, will yield the inversion of many common convolution transforms by differential operators of infinite order (see (1.2) and (8.3)).

Let f be in $\mathcal{U}(\gamma)$ and $[f\{T\}]^{-1}$ exist. If $1/f$ is in $\mathcal{L}(m, \gamma)$ for some m, the calculus shows that

$$[f\{T\}]^{-1} = (f^{-1})(T).$$

When this is not the case, however, we can often construct the inverse as a pointwise limit of polynomials or other operators. The idea is to find a sequence $\{h_n\}$ of functions in $\mathcal{L}(\infty, \gamma)$ such that

$$h_n(\lambda)f(\lambda) \longrightarrow 1$$

suitably near $\sigma(T)$. The functions $h_n(\lambda)f(\lambda)$ may be treated by the calculus, and under proper conditions the sequence

$$(h_n f)(T) = h_n(T)f(T)$$

should converge strongly to I. We shall call a sequence $\{h_n\}$ an inverting sequence for $f \in \mathfrak{V}(\gamma)$ if

(1) $h_n(\lambda)f(\lambda) \in \mathfrak{V}(\gamma)$ $n = 1, 2, \cdots$ with a common strip $[-r, r]$ $(r > \gamma)$ of absolute convergence;

(2) $\lim_{n\to\infty} h_n(\lambda)f(\lambda) = 1$ $(-r \leq \sigma \leq r)$;

(3) for some integer $k \geq 0$,

$$\lim_{n \to \infty} \int_{\Gamma_r} |\lambda^{-k}[1 - h_n(\lambda)f(\lambda)]| \, |d\lambda| = 0.$$

Note that uniform boundedness of $h_n(\lambda)f(\lambda)$ implies (3) if $k = 2$.

LEMMA 7.1. *If $h \in \mathfrak{L}(m, \gamma)$ and f and hf are in $\mathfrak{V}(\gamma)$, then*

$$(hf)\{T\} = \widetilde{h}(T)f\{T\}.$$

Proof. Given x in X we pick x_i $(i = 1, 2, \cdots,)$ in $D_{m+2}(T)$ converging to x. Then $f\{T\}\, x_i$ and $(hf)\{T\}\, x_i$ converge to $f\{T\}\, x_i$ and $(hf)\{T\}\, x_i$ respectively. Since for each i,

$$(hf)\{T\}x_i = h(T)f\{T\}x_i = \widetilde{h}(T)f\{T\}x_i,$$

and $\widetilde{h}(T)$ is closed, the result follows.

THEOREM 7.1. *If $\{h_n\}$ is an inverting sequence for $f \in \mathfrak{V}(\gamma)$, then*

$$\lim_{n \to \infty} \widetilde{h}_n(T)f\{T\}x = x$$

for each x in $D_k(T)$. The limit holds for all x in X if and only if the transformations $(h_nf)\{T\}$ are uniformly bounded.

Proof. For each n,

$$\widetilde{h}_n(T)f\{T\} = (h_nf)\{T\}$$

by the lemma. Now if $x \in D_k(T)$ we may write

$$(h_nf)(T)x = \sum_{i=0}^{k-1} \frac{(h_nf)^{(i)}(0)\,T^ix}{i!} + \frac{1}{2\pi i}\int_{\Gamma_r} \frac{h_n(\lambda)f(\lambda)R_\lambda(T)T^kx}{\lambda^k}\, d\lambda.$$

By condition (2), all terms in the summation go to zero except the first which converges to x. The last term by (3) converges to

$$\frac{1}{2\pi i}\int_{\Gamma_r}\frac{R_\lambda(T)T^k x}{\lambda^k}\,d\lambda.$$

But this integral vanishes by the argument to establish (2.6). Since $D_k(T)$ is dense in X we conclude the last statement from the Banach-Steinhaus theorem and the principle of uniform boundedness [3, p. 25].

In applications we shall take $h_n(\lambda)$ of the form $p_n(\lambda)$ or $e^{c_n\lambda}\,p_n(\lambda)$, where p_n is a polynomial and c_n is real. In each case,

$$h_n(T) = \widetilde{h}_n(T).$$

For the former this was proved at the end of § 4. The latter case is left to the reader.

8. Examples. An important application of the theory of the preceding sections is found by taking for T the operator of differentiation in certain spaces of complex-valued functions defined on the boundary of the unit circle or on the real line. Where functions in the spaces we consider are defined on the line, the spectrum of T is the imaginary axis, whereas $\sigma(T)$ consists of the integral points of the imaginary axis when the functions in the space are defined on the circumference of the circle. For this reason we shall call these two groups of spaces the continuous and discrete cases respectively.

Continuous:

1. $C[-\infty, \infty]$,

   $\|x\| = \sup_t|x(t)|$,

   $D(T) = (x\,|\,x'(t) \in C[-\infty, \infty])$.

2. $L_p(-\infty, \infty)$ $(1 \leq p < \infty)$,

   $\|x\| = \left(\int_{-\infty}^{\infty}|x(t)|^p\,dt\right)^{1/p}$,

   $D(T) = (x\,|\,x(t)$ is absolutely continuous on each finite interval and $x'(t) \in L_p(-\infty, \infty))$.

Discrete:

1. $C[-\pi, \pi]$,

$||x|| = \sup_t |x(t)|$,

$D(T) = (x \,|\, x'(t) \in C[-\pi, \pi])$.

2. $L_p(-\pi, \pi)$ $(1 \leq p < \infty)$,

$||x|| = \left(\int_{-\pi}^{\pi} |(t)|^p \, dt \right)^{1/p}$,

$D(T) = (x \,|\, x(t)$ is absolutely continuous and $x'(t) \in L_p(-\pi, \pi))$.

In the discrete case the functions are, of course, periodic, and an x in $C[-\infty, \infty]$ has limits at $+\infty$ and $-\infty$.

Well-known theorems show that $D(T)$ and, in fact, $D_\infty(T)$ are dense in all these spaces. One shows easily that T is closed. It follows that $(\lambda I - T)$ is closed for any λ, and $(\lambda I - T)^{-1}$ is closed when it exists. Since a closed transformation defined on the whole space is bounded [3, p. 30], the resolvent $R_\lambda(T)$ will exist if and only if λ is such that the differential equation

$$\lambda y(t) - y'(t) = x(t)$$

has a unique solution y in X for each x in X. Then

$$y = R_\lambda(T)x.$$

One shows easily in the continuous case (compare with 5.3) that

$$R_\lambda(T)x(t) = \begin{cases} \int_0^\infty e^{-\lambda\xi}\, x(t+\xi)\, d\xi & (\Re(\lambda) > 0), \\ -\int_{-\infty}^0 e^{-\lambda\xi}\, x(t+\xi)\, d\xi & (\Re(\lambda) < 0), \end{cases}$$

and

$$\mathfrak{T}(\xi)x(t) = x(t+\xi).$$

Also

$$||R_\lambda(T)|| \leq \frac{1}{|\sigma|}.$$

When $f \in \mathfrak{A}(\gamma)$ (as $\gamma = 0$ we shall write just $\mathfrak{A}$ hereafter), that is,

$$f(\lambda) = \int_{-\infty}^{\infty} e^{-\lambda\xi} G(\xi) d\xi,$$

then

$$f(T)x(t) = \int_{-\infty}^{\infty} G(\xi)x(t-\xi)d\xi.$$

Also

$$\|f(T)\| \leq \int_{-\infty}^{\infty} |G(\xi)| d\xi$$

in all the spaces.

Now in the discrete case the formulas above for the continuous case are all valid if one interprets $x(t) = x(t + 2n\pi)$. However, with this identity, they may be given the more convenient form

$$R_\lambda(T)x(t) = \begin{cases} \dfrac{1}{1-e^{-2\pi\lambda}} \displaystyle\int_0^{2\pi} e^{-\lambda\xi} y(t+\xi)d\xi & (\Re(\lambda) > 0), \\[2ex] \dfrac{-1}{1-e^{2\pi\lambda}} \displaystyle\int_{-2\pi}^{0} e^{-\lambda\xi} y(t+\xi)d\xi & (\Re(\lambda) < 0). \end{cases}$$

Another representation is

$$R_\lambda(T)x(t) = \frac{2e^{\lambda t}}{\sinh \pi\lambda} \left[e^{-\lambda\pi} \int_{-\pi}^{t} e^{-\lambda\xi} x(\xi)d\xi + e^{\lambda\pi} \int_{t}^{\pi} e^{-\lambda\xi} x(\xi)d\xi \right]$$

$$(\lambda \neq in,\ n = 0, \pm 1, \pm 2, \cdots).$$

For $f \in \mathfrak{a}$,

$$f(T)x(t) = \int_{-\pi}^{\pi} H(\xi)x(t-\xi)d\xi,$$

where

$$H(\xi) = \sum_{n=-\infty}^{\infty} G(\xi + 2n\pi).$$

If we use the Fourier representation

$$x(t) \sim \sum_{n=-\infty}^{\infty} x_n e^{int},$$

then

$$R_\lambda(T)x(t) \sim \sum_{n=-\infty}^{\infty} \frac{x_n e^{int}}{\lambda - in},$$

and, more generally,

$$f(T)x(t) \sim \sum_{n=-\infty}^{\infty} f(in)x_n e^{int},$$

where the numbers $f(in)$ are the Fourier coefficients of $H(\xi)$. Again,

$$\|R_\lambda(T)\| \le \frac{1}{|\sigma|}$$

and

$$\|f(T)\| \le \int_{-\pi}^{\pi} |H(\xi)|\, d\xi = \int_{-\infty}^{\infty} |G(\xi)|\, d\xi.$$

When $p = 2$, one may show further that

$$\|f(T)\| = \sup_{\lambda \in \sigma(T)} |f(\lambda)|$$

in both the discrete and continuous cases. This and the above facts are well known. The transformations $f(T)$ are special cases of factor transforms for which one may refer to [3, p. 344, 361]; see also [1, p. 99].

In view of these remarks we may state a corollary of Theorem 7.1 in the following convenient form.

THEOREM 8.1. *Let*

$$f(\lambda) = \int_{-\infty}^{\infty} e^{-\lambda\xi}\, G(\xi)\, d\xi$$

and

$$h_n(\lambda) = e^{c_n \lambda} \; p_n(\lambda),$$

where the numbers c_n *are real and the* p_n *are polynomials. If* $f(\lambda)$ *and*

$$h_n(\lambda) f(\lambda) = \int_{-\infty}^{\infty} e^{-\lambda \xi} \; G_n(\xi) d\xi$$

have a common strip $[-r, r]$, $r > 0$ *of absolute convergence, and in this strip*

$$|h_n(\lambda) f(\lambda)| \leq M$$

and

$$\lim_{n \to \infty} h_n(\lambda) f(\lambda) = 1,$$

then:

(a) *if* $x \in L_p(-\infty, \infty)$ $(p \geq 1)$, *or* $C[-\infty, \infty]$, *and* $x \in D_2(T)$,

$$\lim_{n \to \infty} h_n(T) \int_{-\infty}^{\infty} G(\xi) x(t - \xi) d\xi = x(t) \tag{8.1}$$

in norm. If for some M',

$$\int_{-\infty}^{\infty} |G_n(\xi)| d\xi \leq M' \qquad (n = 1, 2, \cdots), \tag{8.2}$$

then (8.1) *holds for all* x.

(b) *If* $x \in L_2(-\infty, \infty)$, *the limit* (8.1) *holds for all* x *without* (8.2).

Actually the theorem is too restrictive in $L_2(-\infty, \infty)$. Since $-iT$ is self-adjoint in this space [12, p. 441], we may use the calculus of Stone and von Neumann (see § 4). With this calculus it is sufficient for the conclusion that $h_n(\lambda) f(\lambda)$ be defined and converge boundedly to unity on the imaginary axis.

As an application of the theorem above we obtain for function spaces an inversion theorem due to Hirschman and Widder [7]. Let

$$f(\lambda) = [E(\lambda)]^{-1}, \; E(\lambda) = \prod_{k=1}^{\infty} \left(1 - \frac{\lambda}{a_k}\right) e^{\lambda / b_k},$$

where $a_k = b_k + i c_k$ $(k = 1, 2, \cdots)$ is a sequence of complex numbers such that

$$\sum_{k=1}^{\infty}\left(\frac{1}{b_k}\right)^2 < \infty, \quad \sum_{k=1}^{\infty}\left(\frac{c_k}{b_k}\right)^2 < \infty.$$

If

$$h_n(\lambda) = e^{d_n\lambda}\prod_{k=1}^{n}\left(1 - \frac{\lambda}{a_k}\right)e^{\lambda/b_k} \qquad (n = 1, 2, \cdots),$$

where the d_n are real numbers approaching zero, then the conditions of the theorem on the functions $f(\lambda)$, $h_n(\lambda)f(\lambda)$ $(n = 1, 2, \cdots)$ are satisfied in any closed vertical strip free of zeros of $E(\lambda)$. Letting

$$G(\xi) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} f(\lambda)\, e^{\lambda\xi}\, d\lambda,$$

$$G_n(\xi) = \frac{1}{2\pi i}\int_{-i\infty}^{i\infty} h_n(\lambda) f(\lambda)\, e^{\lambda\xi}\, d\lambda$$

these authors show that

$$\int_{-\infty}^{\infty} |G_n(\xi)|\, d\xi \le M < \infty.$$

Thus

$$(8.3) \quad \lim_{n\to\infty} \mathfrak{Z}(d_n)\prod_{k=1}^{n}\left(1 - \frac{T}{a_k}\right)\mathfrak{Z}\left(\frac{1}{b_k}\right)\int_{-\infty}^{\infty} G(\xi)x(t-\xi)\,d\xi = x(t) \quad \left(T = \frac{d}{dt}\right),$$

in norm for x in any of the spaces. We list below the kernels of some common transforms with their bilateral transforms which fall under this discussion:

Laplace	$e^{\xi}\, e^{-e^{\xi}}$	$\Gamma(1-\lambda)$
Stieltjes	$\dfrac{1}{2\pi}\operatorname{sech}\dfrac{\xi}{2}$	$\dfrac{1}{\cos \pi\lambda}$
Meijer	$\dfrac{2}{\pi}\cos\dfrac{\pi\nu}{2}\, e^{\xi}\, K_{\nu}(\xi)$	$\dfrac{\cos(\pi\nu/2)\Gamma((1-\nu-\lambda)/2)\Gamma((1+\nu-\lambda)/2)}{\pi 2^{\lambda}}$

$$\left(-\frac{1}{2} \le \nu \le \frac{1}{2}\right).$$

Hirschman and Widder have studied inversion formulas of this sort in great detail (see [4, 5, 6, 7, 16]; see also Pollard [10]). Their results involve the formal differential operator $D = d/dt$, and are concerned with inversion at particular points rather than in the norm topology of function spaces. Their proofs are quite different, involving a convergence argument with the kernels $G_n(\xi)$ in contrast to the present method of first proving inversion for a dense set of functions. On the other hand, the present method seems unsuitable for obtaining local results.

Similar inversion formulas have been proved for $L_2(-\infty, \infty)$ by H. Pollard [9] by use of Fourier transform methods. He needs only to prove that the products $h_n(\lambda)\, f(\lambda)$ converge boundedly to unity on the axis. However, in each case he considers one can show this is true in a strip of positive width. In several cases we are unable to show (8.2), for example in the case of the Weierstrass transform where

$$G(\xi) = \frac{e^{-\xi^2}}{\sqrt{\pi}}, \; f(\lambda) = e^{\lambda^2/4}, \text{ and } h_n(\lambda) = \left(1 - \frac{\lambda^2}{4n}\right)^n,$$

or the partial sums in the series for $f(\lambda)$, and in the case of the Stieltjes and Laplace transforms when h_n are the partial sums.

As a final application, we give an example for the case of the circle. Here the Weierstrass transform takes the form either of a transformation of series

$$x(t) \sim \sum_{n=-\infty}^{\infty} x_n e^{int}, \quad f(T)x(t) = \sum_{n=-\infty}^{\infty} x_n e^{int - n^2/4},$$

or a finite convolution

$$f(T)x(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta_3(\xi, 1/4)\, x(t-\xi)\, d\xi,$$

where

$$\theta_3(\xi, \alpha) = \sqrt{\pi/\alpha} \sum_{n=-\infty}^{\infty} e^{-(\xi + 2n\pi)^2/4\alpha}$$

$$= 1 + 2 \sum_{n=1}^{\infty} e^{-n^2\alpha} \cos n\xi$$

is the theta function occurring in the theory of heat conduction [3, p. 402]. For any x in $L_2(-\pi, \pi)$ or x in $D_2(T)$ of the other spaces,

$$\lim_{n \to \infty} \left(1 - \frac{T^2}{4n}\right)^n f(T)x(t) = x(t)$$

in norm.

REFERENCES

1. S. Bochner and K. Chandrasekharan, *Fourier transforms*, Ann. of Math. Studies No. 19, Princeton, 1949.

2. N. Dunford, *Spectral theory I. Convergence to projections*, Trans. Amer. Math. Soc. 54 (1943), 185-217.

3. E. Hille, *Functional analysis and semi-groups*. Amer. Math. Soc. Coll. Publications, vol. 31, New York, 1948.

4. I. I. Hirschman, Jr. and D. V. Widder, *An inversion and representation theory for convolution transforms with totally positive kernels*, Proc. Nat. Acad. Sci. U.S.A. 34 (1948), 152-156.

5. ———, *Generalized inversion formulas for convolution transforms*, Duke Math. J. 15 (1948), 659-696.

6. ———, *The inversion of a general class of convolution transforms*, Trans. Amer. Math. Soc. 66 (1949), 135-201.

7. ———, *Convolution transforms with complex kernels*, Pacific J. Math. 1(1951), 211-225.

8. B. Nagy, *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes*, Berlin, 1942.

9. H. Pollard, *Integral transforms*, Duke Math. J. 13 (1946), 307-330.

10. ———, *Integral transforms II*. Ann. of Math. 49 (1948), 956-965.

11. G. Polya, *Untersuchungen uber Lücken und Singularitäten von Potenzreihen* Math. Z. 29 (1929), 549-640.

12. M. H. Stone, *Linear transformations in Hilbert space and their applications to analysis*, Amer. Math. Soc. Coll. Publications, vol. 15, New York, 1932.

13. A. E. Taylor, *Analysis in complex Banach spaces*, Bull. Amer. Math. Soc. 49 (1943), 652-669.

14. ———, *Spectral theory of closed distributive operators*, Acta Math. 84 (1950), 189-224.

15. E. C. Titchmarsh, *The theory of functions*, 2nd ed., Oxford, 1939.

16. D. V. Widder, *Inversion formulas for convolution transforms*, Duke Math J. 14 (1947), 217-249.

17. ———, *The Laplace transform*, Princeton, 1946.

UNIVERSITY OF CALIFORNIA, LOS ANGELES AND BERKELEY

SUBFUNCTIONS OF SEVERAL VARIABLES

E. F. BECKENBACH AND L. K. JACKSON

1. Introduction. Convex functions have been generalized in the following two directions: to subharmonic functions [5] of two (or more) independent variables, by replacing the dominating family $\{F(x)\}$ of linear functions, or solutions of the differential equation

$$\frac{d^2 F}{dx^2} = 0,$$

with a family of harmonic functions $\{F(x, y)\}$, or solutions of the partial differential equation

$$\Delta F \equiv \frac{\partial^2 F}{\partial x^2} + \frac{\partial^2 F}{\partial y^2} = 0; \tag{1}$$

and to subfunctions [1] of one variable, by replacing the dominating family of linear functions with a more general family of functions of one variable having certain geometric features in common with the family of linear functions.

We shall here combine the foregoing considerations, generalizing subharmonic functions by replacing the dominating family of harmonic functions with a more general family of functions of two (or more) independent variables.

Bonsall [2] has recently considered some properties of subfunctions of two independent variables relative to the family of solutions of the second-order elliptic partial differential equation

$$\Delta F + a(x, y)\frac{\partial F}{\partial x} + b(x, y)\frac{\partial F}{\partial y} + c(x, y)F = 0.$$

Received February 17, 1952. This research was supported in part by an Office of Naval Research contract.

Pacific J. Math. 3 (1953), 291-313

Tautz [6] has considered a more general situation; but he restricted the dominating family of functions to being linear, and its members to having no positive maxima or negative minima at interior points of the domain of definition.

After developing some properties of subfunctions from a few postulates for our dominating family of functions (§2), we shall introduce further postulates as we need them in studying a Dirichlet problem relative to the dominating family of functions (§§3-5). Applications to elliptic partial differential equations will be made in a subsequent publication.

2. **Generalized subharmonic functions.** Let D be a plane domain (nonnull connected open set), and let $\{\gamma\}$ be a family of closed contours γ bounding subdomains Γ of D such that

a) $\overline{\Gamma} \equiv \Gamma + \gamma \subset D$,

b) $\overline{\Gamma}$ is closed,

c) $\{\gamma\}$ includes all circles κ which lie together with their interiors K in D,

$$\overline{K} = K + \kappa \subset D,$$

and have radii less than a fixed $\rho > 0$.

Let $\{F(x, y)\}$ be a family of functions whose domains of definition lie in D and which satisfy the following postulates:

POSTULATE 1. For any member γ of $\{\gamma\}$ and any continuous boundary-value function $f(x, y)$ on γ, there is a unique $F(x, y; f; \gamma) \in \{F(x, y)\}$ such that

a) $F(x, y; f; \gamma) = f(x, y)$ on γ,

b) $F(x, y; f; \gamma)$ is continuous in $\overline{\Gamma}$.

POSTULATE 2. For each constant $M \geq 0$, if

$$F_1(x, y),\ F_2(x, y) \in \{F(x, y)\},$$

and

$$F_1(x, y) \leq F_2(x, y) + M \qquad \text{on } \gamma, \tag{2}$$

then

$$F_1(x, y) \leq F_2(x, y) + M \qquad \text{in } \overline{\Gamma};$$

further, if the strict inequality holds at a point of γ then the strict inequality holds throughout Γ.

REMARK. We note that the second part of Postulate 2 might have been restricted to the case $M = 0$ without actual loss of generality. For if the strict inequality in (2) holds at a point of γ then also

$$F_1(x, y) \leq F(x, y; F_2 + M; \gamma) \qquad \text{on } \gamma,$$

with the strict inequality holding at a point of γ. It follows from the second part of Postulate 2, restricted to the case $M = 0$, that

$$F_1(x, y) < F(x, y; F_2 + M; \gamma) \leq F_2(x, y) + M \qquad \text{in } \Gamma;$$

or

$$F_1(x, y) < F_2(x, y) + M \qquad \text{in } \Gamma,$$

for $M \geq 0$.

DEFINITION 1. A function $g(x, y)$ is a *continuous sub-*$\{F(x, y)\}$ *function* in D, or briefly a *subfunction*, provided

a) $g(x, y)$ is continuous in D,

b) the inequality

$$g(x, y) \leq F(x, y) \qquad \text{on } \gamma$$

implies the inequality

$$g(x, y) \leq F(x, y) \qquad \text{in } \Gamma.$$

NOTATION. In the sequel we shall restrict use of the symbols D, γ, Γ, $\overline{\Gamma}$, κ, K, $\overline{K}$, $f(x, y)$, $F(x, y)$, $F(x, y; f; \gamma)$, and $g(x, y)$ to the foregoing designations.

THEOREM 1. *If $g(x, y)$ is a subfunction in D, then either*

$$g(x, y) \equiv F(x, y; g; \gamma) \qquad \textit{in } \overline{\Gamma}, \tag{3}$$

or

$$g(x, y) < F(x, y; g; \gamma) \qquad \textit{throughout } \Gamma.$$

Proof. Suppose that for a point (x_0, y_0) of Γ we have

$$g(x_0, y_0) = F(x_0, y_0; g; \gamma). \tag{4}$$

Let $\kappa \in \{\gamma\}$ be a circle with center at (x_0, y_0) and lying together with its interior K in Γ. Then we have

$$g(x, y) = F(x, y; g; \kappa) \leq F(x, y; g; \gamma) \qquad \text{on } \kappa,$$

and therefore

$$g(x, y) \leq F(x, y; g; \kappa) \leq F(x, y; g; \gamma) \qquad \text{in } \overline{K}. \tag{5}$$

In particular, by (4), we have

$$g(x_0, y_0) \leq F(x_0, y_0; g; \kappa) \leq F(x_0, y_0; g; \gamma) = g(x_0, y_0),$$

so that

$$F(x_0, y_0; g; \kappa) = F(x_0, y_0; g; \gamma),$$

and therefore, by (5) and Postulate 2,

$$g(x, y) = F(x, y; g; \kappa) = F(x, y; g; \gamma) \qquad \text{on } \kappa.$$

Since Γ is a connected open set, for any point (x, y) of Γ there is a finite chain of circles in $\{\gamma\}$, each lying in Γ, each with its center on the circumference of the preceding one, and such that the chain begins with κ and ends with a circle through (x, y). Continued repetition of the foregoing analysis shows that (4) implies (3).

COROLLARY. *If $g(x, y)$ is a subfunction in D, and, for a fixed γ, $\overline{\Gamma}$ is interior to the domain of definition of $F(x, y; g; \gamma)$, then either*

$$g(x, y) \equiv F(x, y; g; \gamma),$$

or every neighborhood of each point of γ contains points exterior to $\overline{\Gamma}$ for which

$$g(x, y) > F(x, y; g; \gamma).$$

REMARK. For a subfunction $g(x)$ of one variable, relative to a family $\{F(x)\}$ of functions defined on an interval $a < x < b$, the corollary can be strengthened as follows [1]: If $a < x_1 < x_2 < b$, and

$$g(x_1) = F(x_1), \qquad g(x_2) = F(x_2),$$

then either

$$g(x) \equiv F(x)$$

or

$$g(x) > F(x)$$

for $a < x < x_1$ and for $x_2 < x < b$. But, as independently observed in conversation by R. H. Bing and M. H. Heins, the stronger result does not generally hold for subfunctions of more than one variable. Thus the function

$$V(z) \equiv \log \left| z\left(z - \frac{1}{2}\right) \right|, \qquad (z = x + iy)$$

is subharmonic in $|z| < 1$. For large $M > 0$, the set of points where

$$V(z) < -M$$

has exactly two components, one containing the point $z = 0$, the other containing $z = 1/2$. Let $W(z)$ be defined by

$$W(z) \equiv \begin{cases} -M \text{ on the component containing } z = 0, \\ \max[V(z), -2M] \text{ elsewhere in } |z| < 1. \end{cases}$$

Now $W(z)$ is continuous and subharmonic in $|z| < 1$, coincides with the harmonic function $-M$ on small circles with center at the origin, but is strictly dominated by $-M$ in the neighborhood of $z = 1/2$.

THEOREM 2. *If $g_n(x, y)$ is a subfunction in D for $n = 1, 2, \cdots$, and*

$$g_n(x, y) \longrightarrow g_0(x, y)$$

uniformly on each closed and bounded subset of D, then $g_0(x, y)$ is a subfunction in D.

Proof. Clearly, $g_0(x, y)$ is continuous in D. For any $\gamma \in \{\gamma\}$ and any $\epsilon > 0$, there is an $N = N(\epsilon)$ such that for $n \geq N$ we have

$$|g_n(x, y) - g_0(x, y)| < \epsilon \qquad \text{in } \overline{\Gamma}.$$

Then for $n \geq N$ and $(x, y) \in \Gamma$, we have

$$g_0(x, y) \leq g_n(x, y) + \epsilon$$
$$\leq F(x, y; g_n; \gamma) + \epsilon \leq F(x, y; g_0; \gamma) + 2\epsilon.$$

Therefore, since $\epsilon > 0$ is arbitrary, we have

$$g_0(x, y) \leq F(x, y; g_0; \gamma) \qquad \text{in } \overline{\Gamma}.$$

The following result is a generalization of Littlewood's theorem [4, pp. 152-157] concerning subharmonic functions.

THEOREM 3. *A function $g(x, y)$, continuous in D, is a subfunction in D if and only if corresponding to each $(x_0, y_0) \in D$ there exists a sequence of circles $\kappa_n = \kappa_n(x_0, y_0) \in \{\gamma\}$ with centers at (x_0, y_0) and radii $\rho_n(x_0, y_0) \longrightarrow 0$, such that*

$$g(x_0, y_0) \leq F(x_0, y_0; g; \kappa_n).$$

Proof. We shall prove only the sufficiency of the condition, since the necessity is obvious by definition.

Suppose that the condition holds but that $g(x, y)$ is not a subfunction; then there is a $\gamma \in \{\gamma\}$ and an $F(x, y) \in \{F(x, y)\}$ such that

$$g(x, y) \leq F(x, y) \qquad \text{on } \gamma$$

but

$$g(x, y) > F(x, y)$$

at some point of Γ. Then the set of points of Γ on which

$$\max_{(x, y) \in \Gamma} g(x, y) - F(x, y) \equiv M > 0$$

is attained is a closed nonnull interior set E in Γ.

Let (x_0, y_0) be a point of E such that

$$\text{dist}\,[(x_0, y_0), \gamma] = \min_{(x, y) \in E} \text{dist}\,[(x, y), \gamma].$$

By hypothesis, we have

$$g(x_0, y_0) \leq F(x_0, y_0; g; \kappa_n);$$

but, by our selection of (x_0, y_0), for sufficiently large n there is an arc of κ_n on which

$$g(x, y) - F(x, y) < M.$$

Thus on κ_n we have

$$F(x, y; g; \kappa_n) = g(x, y) \leq F(x, y) + M,$$

with the strict inequality holding at some point, so that, by Postulate 2, at each point inside κ_n we have

$$F(x, y; g; \kappa_n) < F(x, y) + M;$$

in particular, we have

$$g(x_0, y_0) \leq F(x_0, y_0; g; \kappa_n) < F(x_0, y_0) + M = g(x_0, y_0),$$

a contradiction.

REMARK. A method similar to that used in proving Theorem 3 can be used to show that Postulate 2, restricted to the case $\gamma = \kappa$, implies the result stated in Postulate 2 for general $\gamma \in \{\gamma\}$. Thus Postulate 2 might have been restricted to the case $\gamma = \kappa$ without actual loss of generality.

THEOREM 4. *If $g_1(x, y)$, $g_2(x, y)$, $\cdots$, $g_n(x, y)$ are subfunctions in D, then $g_0(x, y)$, defined by*

$$g_0(x, y) \equiv \max [g_1(x, y), g_2(x, y), \cdots, g_n(x, y)],$$

is a subfunction in D.

Proof. Since the functions $g_j(x, y)$ $(j = 1, 2, \cdots, n)$ are continuous, it follows that $g_0(x, y)$ also is continuous. Let $\gamma \in \{\gamma\}$, and let $(x_0, y_0) \in \Gamma$. Then for some j, with $1 \leq j \leq n$, we have

$$g_0(x_0, y_0) = g_j(x_0, y_0) \leq F(x_0, y_0; g_j; \gamma) \leq F(x_0, y_0; g_0; \gamma).$$

THEOREM 5. *If $g(x, y)$ is a subfunction in D, then, for any fixed $\gamma \in \{\gamma\}$, the function $g(x, y; \gamma)$, defined by*

$$g(x, y; \gamma) \equiv \begin{cases} g(x, y) & \text{for } (x, y) \in D - \Gamma, \\ F(x, y; g; \gamma) & \text{for } (x, y) \in \Gamma, \end{cases}$$

is a subfunction in D.

Proof. It follows from Theorem 3 that we need to test the behavior of $g(x, y; \gamma)$ only relative to small circles $\kappa \in \{\gamma\}$ with centers at points (x_0, y_0) of the given γ. But then we immediately have the desired inequality

$$g(x_0, y_0; \gamma) = g(x_0, y_0) \le F(x_0, y_0; g; \kappa) \le F(x_0, y_0; g(x, y; \gamma); \kappa).$$

DEFINITION 2. *Superfunctions* are defined by reversing the inequalities in the definition (Definition 1) of subfunctions relative to the family $\{F(x, y)\}$.

It is easy to show that results analogous to Theorems 1-5, with suitable alterations, hold for superfunctions: in addition to writing "superfunction" for "subfunction," we reverse the inequality in the last line of Theorem 1 and in the last line of Theorem 3, and replace "max" by "min" in Theorem 4.

3. **A Dirichlet problem.** We now introduce some additional symbols.

NOTATION. Let Ω be a bounded connected open subset of D with boundary ω such that

$$\overline{\Omega} \equiv \Omega + \omega \subset D.$$

To distinguish points of Ω from points of ω, we shall often designate points of Ω by capital letters A, B, and points of ω by α, β; and we shall write $f(A)$ for $f(x, y)$, where (x, y) are the coordinates of A, and so on.

Let $h(\alpha)$ be a bounded, but not necessarily continuous, function defined on ω, and define $h_*(\alpha)$ and $h^*(\alpha)$ by

$$h_*(\alpha) \equiv \liminf_{\beta \to \alpha} h(\beta),$$

$$h^*(\alpha) \equiv \limsup_{\beta \to \alpha} h(\beta).$$

DEFINITION 3. By a *solution of the Dirichlet problem* for Ω relative to $\{F(x, y)\}$ and relative to the given bounded boundary-value function $h(\alpha)$ on ω, we shall mean a function $H(x, y)$ which is continuous in Ω, satisfies

$$h_*(\alpha) \leq \liminf_{A \to \alpha} H(A) \leq \limsup_{A \to \alpha} H(A) \leq h^*(\alpha), \tag{6}$$

and is such that for each $\gamma \in \{\gamma\}$ with $\overline{\Gamma} \subset \Omega$ we have

$$H(x, y) \equiv F(x, y; H; \gamma) \qquad \text{in } \Gamma. \tag{7}$$

DEFINITION 4. We shall say that a function $H(x, y)$ which is continuous in Ω, and which satisfies (7) for each $\gamma \in \{\gamma\}$ with $\overline{\Gamma} \subset \Omega$, is an $\{F(x, y)\}$-*function in* Ω, though of course in the given family $\{F(x, y)\}$ there might be no member whose domain of definition contains Ω; the given domains of definition might for instance be small circles. Clearly, the only functions which are both subfunctions and superfunctions are the $\{F(x, y)\}$-functions.

DEFINITION 5. The function $\phi(x, y)$ is an *under-function* provided

a) $\phi(x, y)$ is continuous in $\overline{\Omega}$,

b) $\phi(A)$ is a subfunction in Ω,

c) $\phi(\alpha)$ satisfies

$$\phi(\alpha) \leq h_*(\alpha) \qquad \text{on } \omega.$$

DEFINITION 6. The function $\psi(x, y)$ is an *over-function* provided

a) $\psi(x, y)$ is continuous in $\overline{\Omega}$,

b) $\psi(A)$ is a superfunction in Ω,

c) $\psi(\alpha)$ satisfies

$$\psi(\alpha) \geq h^*(\alpha) \qquad \text{on } \omega.$$

THEOREM 6. *If $\phi(x, y)$ is an under-function and $\psi(x, y)$ is an over-function, then*

$$\phi(x, y) \leq \psi(x, y) \qquad \textit{in } \overline{\Omega}.$$

Proof. The result can be established by a method similar to that used in proving Theorem 3.

THEOREM 7. *If* $\phi_1(x, y), \phi_2(x, y), \cdots, \phi_n(x, y)$ *are under-functions, then* $\phi(x, y)$*, defined by*

$$\phi(x, y) \equiv \max [\phi_1(x, y), \phi_2(x, y), \cdots, \phi_n(x, y)],$$

is an under-function.

THEOREM 8. *If* $\psi_1(x, y), \psi_2(x, y), \cdots, \psi_n(x, y)$ *are over-functions, then* $\psi(x, y)$*, defined by*

$$\psi(x, y) \equiv \min [\psi_1(x, y), \psi_2(x, y), \cdots, \psi_n(x, y)],$$

is an over-function.

Proof. Property b) of Definition 5 holds for $\phi(x, y)$ by Theorem 4; the other properties of Definition 5 hold for $\phi(x, y)$ since they hold for each $\phi_j(x, y)$. Thus Theorem 7 is valid; and Theorem 8 can be proved similarly.

THEOREM 9. *If* $\phi(x, y)$ *is an under-function, and* $\gamma \in \{\gamma\}$, $\overline{\Gamma} \subset \Omega$, *then* $\phi(x, y; \gamma)$*, defined by*

$$\phi(x, y; \gamma) \equiv \begin{cases} \phi(x, y) & \text{for } (x, y) \text{ in } \overline{\Omega} - \Gamma, \\ F(x, y; \phi, \gamma) & \text{for } (x, y) \text{ in } \Gamma, \end{cases}$$

is an under-function.

THEOREM 10. *If* $\psi(x, y)$ *is an over-function, then* $\psi(x, y; \gamma)$ *is an over-function.*

Proof. Theorem 9 follows immediately from Definition 5 and Theorem 5, and Theorem 10 holds similarly.

POSTULATE 3. For any $\kappa \in \{\gamma\}$, and for any collection $\{f_\nu(x, y)\}$ of functions $f_\nu(x, y)$ which are continuous and uniformly bounded on κ, the functions $F(x, y; f_\nu; \kappa)$ are equicontinuous in K.

We shall use the following well-known and easily established result in connection with Postulate 3.

LEMMA 1. *For any collection* $\{U_\nu(x, y)\}$ *of functions* $U_\nu(x, y)$ *which are uniformly bounded and equicontinuous on a set* E*, the functions* $S(x, y)$ *and* $I(x, y)$*, defined by*

$$S(x, y) \equiv \sup_{U_\nu \in \{U_\nu\}} [U_\nu(x, y)],$$

$$I(x, y) \equiv \inf_{U_\nu \in \{U_\nu\}} [U_\nu(x, y)],$$

are continuous on E.

POSTULATE 4. For any bounded connected open subset Ω of D with boundary ω such that $\overline{\Omega} \subset D$, and for any bounded function $h(\alpha)$ defined on ω, there exists an under-function $\phi(x, y)$, and there exists an over-function $\psi(x, y)$.

DEFINITION 7. By $H_*(x, y)$ and $H^*(x, y)$ we shall denote the functions defined by

$$H_*(x, y) \equiv \sup_{\phi \in \{\phi\}} [\phi(x, y)],$$

$$H^*(x, y) \equiv \inf_{\psi \in \{\psi\}} [\psi(x, y)],$$

where $\{\phi\}$ and $\{\psi\}$ denote the families of under- and over-functions, respectively.

The existence of the functions $H_*(x, y)$ and $H^*(x, y)$ follows from Postulate 4 and Theorem 6.

THEOREM 11. *The function* $H_*(x, y)$ *is a subfunction in* Ω.

Proof. We shall show first that for each $\kappa \in \{\gamma\}$, with $\overline{K} \subset \Omega$, the function $H_*(x, y)$ is continuous in K, so that $H_*(x, y)$ is continuous in Ω.

Let $\phi_0(x, y)$ and $\psi_0(x, y)$ be fixed members of $\{\phi(x, y)\}$ and $\{\psi(x, y)\}$, respectively; and for each $\phi(x, y) \in \{\phi(x, y)\}$ define

$$\Phi(x, y) \equiv \max [\phi(x, y), \phi_0(x, y)].$$

Using Theorem 4, we readily verify that $\Phi(x, y)$ satisfies the conditions of Definition 5, so that $\Phi(x, y)$ is an under-function; also, by Theorem 9, $\Phi(x, y; \kappa)$ is an under-function.

Since $\Phi(x, y; \kappa)$ is an under-function, and

$$\phi(x, y) \leq \Phi(x, y) \leq \Phi(x, y; \kappa),$$

we have

$$H_*(x, y) = \sup_{\Phi \in \{\Phi\}} \Phi(x, y; \kappa); \tag{8}$$

further, using Theorem 6, for (x, y) in K we obtain

$$\phi_0(x, y) \leq \Phi(x, y; \kappa) = F(x, y; \Phi; \kappa) \leq \psi_0(x, y). \tag{9}$$

It now follows from (8), (9), Postulate 3, and Lemma 1 that $H_*(x, y)$ is continuous in K, so that $H_*(x, y)$ is continuous in Ω.

Now, for any circle $\kappa \in \{\gamma\}$ with center (x_0, y_0) and $\overline{K} \subset \Omega$, and for any $\phi \in \{\phi\}$, by the definition of $H_*(x, y)$ we have

$$\phi(x, y) \leq H_*(x, y) \qquad \text{on } \kappa;$$

therefore, since $\phi(x, y)$ is a subfunction in Ω, we have

$$\phi(x_0, y_0) \leq F(x_0, y_0; \phi, \kappa) \leq F(x_0, y_0; H_*; \kappa),$$

whence, again by the definition of $H_*(x, y)$, it follows that

$$H_*(x_0, y_0) \leq F(x_0, y_0; H_*; \kappa).$$

Accordingly, by Theorem 3, $H_*(x, y)$ is a subfunction in Ω.

THEOREM 12. *The function $H_*(x, y)$ is a superfunction in Ω.*

Proof. Let there be given any $\epsilon > 0$ and any $\gamma \in \{\gamma\}$ with $\overline{\Gamma} \subset \Omega$. Then, by the definition of $H_*(x, y)$, for any $(x_0, y_0) \in \overline{\Gamma}$ there is a $\phi_0 \in \{\phi\}$ such that

$$\phi_0(x_0, y_0) > H_*(x_0, y_0) - \frac{\epsilon}{2};$$

therefore, by continuity, there is a circle $\kappa_0 \in \{\gamma\}$, with center (x_0, y_0) and $\overline{K}_0 \subset \Omega$, such that

$$\phi_0(x, y) > H_*(x, y) - \epsilon \qquad \text{in } \overline{K}_0. \tag{10}$$

Since the circular discs K_0 form an open covering of $\overline{\Gamma}$, by the Heine-Borel Theorem there exists a finite subcovering; let $\phi_1(x, y), \phi_2(x, y), \cdots, \phi_n(x,y)$ be the associated under-functions, and let $\phi(x, y)$ be defined by

$$\phi(x, y) \equiv \max\,[\phi_1(x, y), \phi_2(x, y), \cdots, \phi_n(x, y)].$$

Then, by Theorem 7, $\phi(x, y)$ is an under-function; and, by (10), we have

$$\phi(x, y) > H_*(x, y) - \epsilon \qquad \text{in } \bar{\Gamma}.$$

By Postulate 2, then, we obtain

$$F(x, y; \phi; y) > F(x, y; H_* - \epsilon; y) \geq F(x, y; H_*; y) - \epsilon \qquad \text{in } \bar{\Gamma}. \tag{11}$$

Since for $(x, y) \in \bar{\Gamma}$ and any $\phi \in \{\phi\}$ we also have

$$H_*(x, y) \geq \phi(x, y; y) = F(x, y; \phi; y), \tag{12}$$

it follows from (11) and (12) that

$$H_*(x, y) \geq F(x, y; H_*; y) - \epsilon \qquad \text{in } \bar{\Gamma}.$$

Thus since $\epsilon > 0$ is arbitrary, we have

$$H_*(x, y) \geq F(x, y; H_*; y) \qquad \text{in } \bar{\Gamma}.$$

so that $H_*(x, y)$ is a superfunction in Ω.

Since Theorems 11 and 12 hold also for the function $H^*(x, y)$, and since the only functions which are both subfunctions and superfunctions in Ω are $\{F(x, y)\}$-functions in Ω (see Definition 4), we have the following result:

THEOREM 13. *The functions $H_*(x, y)$ and $H^*(x, y)$ are $\{F(x, y)\}$-functions in Ω.*

We now turn our attention to the behavior of the functions $H_*(x, y)$ and $H^*(x, y)$ at the boundary ω of Ω.

4. **Regular boundary points; barrier functions.** We make the following definition.

DEFINITION 8. The point $\alpha_0 \in \omega$ is a *regular* boundary point of Ω relative to $\{F(x, y)\}$ provided that for every bounded function $h(\alpha)$ on ω the functions $H_*(x, y)$ and $H^*(x, y)$ satisfy (6) at α_0:

$$h_*(\alpha_0) \leq \liminf_{A \to \alpha_0} H_*(A) \leq \limsup_{A \to \alpha_0} H_*(A) \leq h^*(\alpha_0), \tag{13a}$$

(13b) $$h_*(\alpha_0) \leq \liminf_{A \to \alpha_0} H^*(A) \leq \limsup_{A \to \alpha_0} H^*(A) \leq h^*(\alpha_0).$$

THEOREM 14. *If all points of ω are regular boundary points of Ω, and $h(\alpha)$ is continuous on ω, then the Dirichlet problem for Ω, relative to $\{F(x, y)\}$ and $h(\alpha)$, has a unique solution.*

Proof. From (13) and the continuity of $h(\alpha)$ on ω, we see that $H_*(x, y)$ and $H^*(x, y)$ are continuous in $\overline{\Omega}$ and satisfy

$$H_*(\alpha) = H^*(\alpha) = h(\alpha) \qquad \text{on } \omega.$$

Accordingly, by Definitions 5 and 6 and Theorems 11 and 12, $H_*(x, y)$ is both an under-function and an over-function; similarly, $H^*(x, y)$ is both an under-function and an over-function. Therefore, by Theorem 6, we have

$$H_*(x, y) \equiv H^*(x, y) \qquad \text{in } \overline{\Omega}.$$

For the same reason, any other solution of the Dirichlet problem must coincide with $H_*(x, y)$ and $H^*(x, y)$ in $\overline{\Omega}$.

We shall now give local sufficient conditions in terms of barrier functions (see [3, pp. 326-328]) in order that a point $\alpha \in \omega$ be a regular point of Ω; in the next section we shall study conditions under which barrier functions exist.

DEFINITION 9. For a point $\alpha_0 = (x_0, y_0) \in \omega$, a circle κ with center at α_0 and with $\overline{K} \subset D$, and constants $\epsilon > 0$, M, and N, a function

$$s(x, y) \equiv s(x, y; \kappa; \epsilon, M, N)$$

is a *barrier subfunction* provided:

a) $s(x, y)$ is continuous in $\overline{\Omega} \cap \overline{K}$,

b) $s(x, y)$ is a subfunction in $\Omega \cap K$,

c) $s(\alpha_0) \geq N - \epsilon$,

d) $s(x, y) \leq N + 2\epsilon$ on $\omega \cap K$,

e) $s(x, y) \leq M$ on $\overline{\Omega} \cap \kappa$.

DEFINITION 10. With the notation of Definition 9, a function

$$S(x, y) \equiv S(x, y; \kappa; \epsilon, M, N)$$

is a *barrier superfunction* provided

a) $S(x, y)$ is continuous in $\overline{\Omega} \cap \overline{K}$,

b) $S(x, y)$ is a superfunction in $\Omega \cap K$,

c) $S(\alpha_0) \leq N + \epsilon$

d) $S(x, y) \geq N - 2\epsilon$ on $\omega \cap K$,

e) $S(x, y) \geq M$ on $\overline{\Omega} \cap \kappa$.

THEOREM 15. *If for the point $\alpha_0 \in \omega$, and for each set of constants $\epsilon > 0$, M, and N, there exists a sequence of circles $\kappa_n = \kappa_n(\alpha_0)$ with center at α_0 and radii $\rho_n(\alpha_0) \longrightarrow 0$ for which barrier subfunctions $s(x, y; \kappa_n; \epsilon, M, N)$ and barrier superfunctions $S(x, y; \kappa_n; \epsilon, M, N)$ exist, then α_0 is a regular boundary point of Ω relative to $\{F(x, y)\}$.*

Proof. For a given bounded function $h(\alpha)$ defined on ω, it follows from Theorem 6 and Definition 7 that

$$H_*(x, y) \leq H^*(x, y) \qquad \text{in } \overline{\Omega},$$

so that

$$\liminf_{A \to \alpha_0} H_*(A) \leq \liminf_{A \to \alpha_0} H^*(A)$$

and

$$\limsup_{A \to \alpha_0} H_*(A) \leq \limsup_{A \to \alpha_0} H^*(A).$$

Accordingly, in order to verify (12) and thus prove the theorem, we need only show that

$$h_*(\alpha_0) \leq \liminf_{A \to \alpha_0} H_*(A) \tag{14}$$

and

$$\limsup_{A \to \alpha_0} H^*(A) \leq h^*(\alpha_0). \tag{15}$$

For a given $\epsilon > 0$ there is a circle κ satisfying the hypotheses of the theorem and for which

$$h_*(\alpha_0) - \epsilon \le h_*(\alpha) \le h^*(\alpha) \le h^*(\alpha_0) + \epsilon \qquad \text{on } \omega \cap \overline{K}.$$

For a fixed $\delta > 0$ and for any under-function $\phi(x, y)$, let

$$M = \min_{(x,y) \in \overline{\Omega} \cap \kappa} [\phi(x, y) - \delta],$$

$$N = h_*(\alpha_0) - 3\epsilon,$$

and

$$s(x, y) = s(x, y; \kappa; \epsilon, M, N).$$

Consider the function $\Phi(x, y)$, defined by

$$\Phi(x, y) \equiv \begin{cases} \max[\phi(x, y), s(x, y)] & \text{in } \overline{\Omega} \cap \overline{K}, \\ \phi(x, y) & \text{in } \overline{\Omega} - \overline{K}; \end{cases}$$

we shall show that $\Phi(x, y)$ is an under-function. Since we have

$$s(x, y) \le M < \phi(x, y) \qquad \text{on } \overline{\Omega} \cap \kappa,$$

it follows readily that

$$\Phi(x, y) = \phi(x, y) \qquad \text{on } \overline{\Omega} \cap \kappa,$$

and accordingly that $\Phi(x, y)$ is continuous in $\overline{\Omega}$. Further, $\Phi(x, y)$ is a subfunction in $\Omega - \overline{K}$, since $\Phi(x, y) = \phi(x, y)$ there; $\Phi(x, y)$ is a subfunction in $\Omega \cap K$ by Theorem 4 and Definitions 5 and 9; and for a point $\alpha \in \Omega \cap \kappa$ we have $s(\alpha) < \phi(\alpha) - \delta$, so that there is a circle about α in which $\Phi(x, y) = \phi(x, y)$; thus, by Theorem 3, $\Phi(x, y)$ is a subfunction in Ω; also, by Definition 9 and the choice of κ and N, we have

$$s(\alpha) \le N + 2\epsilon = h_*(\alpha_0) - \epsilon \le h_*(\alpha) \qquad \text{on } \omega \cap \overline{K},$$

and therefore, since

$$\phi(\alpha) \le h_*(\alpha) \qquad \text{on } \omega,$$

we have

$$\Phi(\alpha) \leq h_*(\alpha) \qquad \text{on } \omega.$$

By Definition 5 we have thus shown that $\Phi(x, y)$ is an under-function.

By the choice of N and the definitions of $s(x, y)$ and $\Phi(x, y)$, we have

$$h_*(\alpha_0) - 4\epsilon = N - \epsilon \leq s(\alpha_0) \leq \Phi(\alpha_0),$$

so that by continuity there is a neighborhood of α_0 in whose intersection with $\overline{\Omega} \cap \overline{K}$ we have

$$h_*(\alpha_0) - 5\epsilon \leq \Phi(x, y)$$

and consequently

$$h_*(\alpha_0) - 5\epsilon \leq H_*(x, y).$$

Since $\epsilon > 0$ is arbitrary, (14) now follows; and (15) can be established similarly.

5. **The existence of barrier functions.** Relative to the Laplace partial differential equation (1), a criterion of Poincaré [3, p. 329] for α_0 to be a regular boundary point of Ω is that there should exist a circle κ with

$$\overline{\Omega} \cap \overline{K} = \alpha_0. \tag{16}$$

We shall now adjoin postulates concerning the family $\{F(x, y)\}$ under which (16) is a sufficient condition for the existence of barrier sub- and superfunctions at α_0, and therefore, by Theorem 15, for α_0 to be a regular boundary point of Ω relative to $\{F(x, y)\}$.

POSTULATE 5. For any circle $\kappa \in \{\gamma\}$, and any real number M, there exist continuous functions $f_1(x, y)$, $f_2(x, y)$, defined on κ, such that

$$F(x, y; f_1; \kappa) \leq M, \; F(x, y; f_2; \kappa) \geq M \qquad \text{in } \overline{K}.$$

POSTULATE 6. For any circle $\kappa \in \{\gamma\}$, and any real numbers $\epsilon > 0$ and N, there exists a continuous function $f(x, y)$ defined on κ such that

$$|F(x_0, y_0; f; \kappa) - N| \leq \epsilon,$$

where (x_0, y_0) is the center of κ.

POSTULATE 7. For any circle $\kappa \in \{\gamma\}$; if the functions $f_j(x, y)$ $(j = 0, 1, \cdots)$, defined on κ, are continuous and uniformly bounded on κ, and

$$\lim_{j \to \infty} f_j(x, y) = f_0(x, y)$$

for all but at most a finite number of points of κ, then

$$\lim_{j \to \infty} F(x, y; f_j; \kappa) = F(x, y; f_0; \kappa)$$

for all points of K.

POSTULATE 8. For any circle $\kappa \in \{\gamma\}$, if the functions $f_j(x, y)$ $(j = 1, 2, \cdots)$, defined on κ, are continuous on κ and equicontinuous at a point $(x_0, y_0) \in \kappa$, then the functions $F(x, y; f_j; \kappa)$ $(j = 1, 2, \cdots)$, defined in $\overline{K}$, are equicontinuous at (x_0, y_0).

THEOREM 16. *If for the point $\alpha_0 \in \omega$ there exists a circle κ, with $\overline{K} \subset D$, such that*

$$\overline{\Omega} \cap \overline{K} = \alpha_0,$$

then α_0 is a regular boundary point of Ω relative to $\{F(x, y)\}$.

Proof. Since the conclusions of Theorems 15 and 16 are identical, in order to prove Theorem 16 we need only to show that its hypothesis implies that of Theorem 15. Explicitly, we shall give the construction of a barrier subfunction for a suitable circle $\kappa_1(\alpha_0)$ with center at α_0 and inside an arbitrary circle κ_0 with center at α_0, as prescribed in Theorem 15; the existence of barrier superfunctions can be treated similarly.

Let the circle $\kappa_0 \subset D$ be drawn with center at $\alpha_0 = (x_0, y_0)$ and intersecting κ. By Postulate 6 there is a continuous function $f(x, y)$ defined on κ_0 such that

$$|F(x_0, y_0; f; \kappa_0) - N| < \epsilon. \tag{17}$$

By continuity, there is a circle $\kappa_1 \subset K_0$, with center at α_0, such that

$$F(x, y; f; \kappa_0) \leq N + 2\epsilon \qquad \text{in } \overline{K}_1. \tag{18}$$

Now we define

$$R = \min_{(x, y) \in \overline{K}_1} F(x, y; f; \kappa_0)$$

and

$$M_* = \min (M, N, R).$$

By Postulate 5, there exists a continuous function $f_1(x, y)$ defined on κ_1 such that

$$F(x, y; f_1; \kappa_1) \leq M_* \qquad \text{in } \overline{K}_1. \tag{19}$$

Let B be the intersection of the line of centers of κ and κ_1 with the arc of κ_1 lying outside $\overline{K}$, and let B_1', B_1'', B_2', B_2'' be points of κ_1 near B arranged in the order $B_2'\ B_1'\ B\ B_1''\ B_2''$ around κ_1.

We define the function $f_2(x, y)$ on κ_1 as follows:

$$f_2(x, y) = f_1(x, y) - 1 \qquad \text{on arc } B_1'\ B\ B_1'';$$

$$f_2(x, y) = F(x, y; f; \kappa_0) \qquad \text{on long arc } B_2'\ B_2'';$$

$$f_2(x, y) = l'(\theta) \qquad \text{on arc } B_1'\ B_2';$$

$$f_2(x, y) = l''(\theta) \qquad \text{on arc } B_1''\ B_2'';$$

the functions $l'(\theta)$ and $l''(\theta)$ are linear functions of the central angle of κ_1, such that $f_2(x, y)$ is continuous on κ_1.

For (x, y) on κ_1, we set

$$f_3(x, y) = \min [f_2(x, y), F(x, y; f; \kappa_0)].$$

By (17) and Postulate 7, we can take the arc $B_2'\ B\ B_2''$ small enough that

$$| F(x_0, y_0; f_3; \kappa_1) - N | \leq \epsilon. \tag{20}$$

Further, since

$$f_3(x, y) \leq F(x, y; f; \kappa_0) \qquad \text{on } \kappa_1,$$

by (18) and Postulate 2 we have

$$F(x, y; f_3; \kappa_1) \leq N + 2\epsilon \qquad \text{in } \overline{K}_1. \tag{21}$$

Let $Q \in K_1$ be a point on the open line-segment $\alpha_0 B$, and sufficiently close to B that

$$F(Q; f_3; \kappa_1) < F(Q; f_1; \kappa_1). \tag{22}$$

Let κ' and κ'' be the two circles through Q and α_0, and tangent to κ_1. Let ρ be the length of the common chord $\alpha_0 Q$ of κ' and κ'', or the length of the common chord of κ' and κ, whichever is less, and choose the constant C so that

$$C\rho > M^*, \tag{23}$$

where

$$M^* = \max_{(x,y)\in \overline{K}_1} |F(x, y; f_1; \kappa_1)| + \max_{(x,y)\in \overline{K}_1} |F(x, y; f_3; \kappa_1)|.$$

We now define continuous functions $h_n'(x, y)$ and $h_n''(x, y)$ on κ' and κ'', respectively, as follows:

$$\begin{aligned} h_1'(x, y) &= F(x, y; f_3; \kappa_1) - C[(x - x_0)^2 + (y - y_0)^2]^{1/2} \qquad \text{on } \kappa', \\ h_1''(x, y) &= F(x, y; f_3; \kappa_1) - C[(x - x_0)^2 + (y - y_0)^2]^{1/2} \qquad \text{on } \kappa'', \end{aligned} \tag{24}$$

and, for $n = 2, 3, \cdots$,

$$\begin{aligned} h_n'(x, y) &= \begin{cases} h_1'(x, y) & \text{on } \kappa' - \overline{K}'', \\ F(x, y; h_{n-1}''; \kappa'') & \text{on } \kappa' \cap \overline{K}'', \end{cases} \\ h_n''(x, y) &= \begin{cases} h_1''(x, y) & \text{on } \kappa'' - \overline{K}' \\ F(x, y; h_n'; \kappa') & \text{on } \kappa'' \cap \overline{K}'. \end{cases} \end{aligned} \tag{25}$$

Let

$$G = \max_{(x,y)\in \overline{K}'} |F(x, y; h_1'; \kappa')| + \max_{(x,y)\in \overline{K}''} |F(x, y; h_1''; \kappa'')|;$$

then by Postulate 5 there is a continuous function $f_4(x, y)$ defined on κ_1 such that

$$F(x, y; f_4; \kappa_1) \leq -G \qquad \text{in } \overline{K}_1.$$

It follows from Postulate 2 and the definitions of the $h_n'(x, y)$ and $h_n''(x, y)$ that for each positive integer n we have

$$F(x, y; f_4; \kappa_1) \leq F(x, y; h_n'; \kappa') \leq F(x, y; f_3; \kappa_1) \qquad \text{in } \overline{K}',$$

and

$$F(x, y; f_4; \kappa_1) \leq F(x, y; h_n''; \kappa'') \leq F(x, y; f_3; \kappa_1) \qquad \text{in } \overline{K}''.$$

Hence, by Postulate 3 and Lemma 1, the functions $u'(x, y)$ and $u''(x, y)$, defined by

$$u'(x, y) = \sup_{h_n' \in \{h_n'\}} F(x, y; h_n'; \kappa')$$

and

$$u''(x, y) = \sup_{h_n'' \in \{h_n''\}} F(x, y; h_n''; \kappa'')$$

are continuous in K' and K'', respectively; indeed, by Postulate 8 and by Lemma 1 applied to the sets $(\overline{K}' - \alpha_0 - Q)$ and $(\overline{K}'' - \alpha_0 - Q)$, the functions $u'(x, y)$ and $u''(x, y)$ are continuous in $\overline{K}'$ and $\overline{K}''$, respectively, except possibly at the points α_0 and Q. As for the behavior of these functions at α_0, since by our construction we have

$$F(x, y; h_1'; \kappa') \leq u'(x, y) \leq F(x, y; f_3; \kappa_1) \qquad \text{in } \overline{K}', \tag{26}$$

and

$$F(x, y; h_1''; \kappa'') \leq u''(x, y) \leq F(x, y; f_3; \kappa_1) \qquad \text{in } \overline{K}'', \tag{27}$$

and since the functions at the extremes of these inequalities have equal values at α_0 and are continuous at α_0, it follows that $u'(x, y)$ and $u''(x, y)$ are continuous also at α_0.

As in the last part of the proof of Theorem 11, $u'(x, y)$ and $u''(x, y)$ can easily be shown to be subfunctions in K' and K'', respectively.

Now we define the function $u(x, y)$ in $\overline{K}' \cup \overline{K}''$ as follows:

$$(28) \qquad u(x, y) = \begin{cases} u'(x, y) & \text{in } \overline{K}' - K'' \\ u''(x, y) & \text{in } \overline{K}'' - \overline{K}' \\ \max [u'(x, y), u''(x, y)] & \text{in } \overline{K}' \cap K''. \end{cases}$$

Since $u'(x, y)$ and $u''(x, y)$ coincide on $\kappa' \cap K''$, on $\kappa'' \cap K'$, and at α_0, and both are continuous at α_0, it follows that $u(x, y)$ is continuous in $\overline{K}' \cup \overline{K}''$ except possibly at Q.

Clearly $u(x, y)$ is a subfunction in $K' - \overline{K}''$ and in $K'' - \overline{K}'$. By Theorem 4, $u(x, y)$ is a subfunction in $K' \cap K''$. Since in addition the hypothesis of Theorem 3 holds for each point of $\kappa' \cap K''$ and for each point of $\kappa'' \cap K'$, it follows that $u(x, y)$ is a subfunction throughout $K' \cup K''$.

To conclude the proof, we shall show that the function

$$(29) \qquad s(x, y) \equiv \begin{cases} F(x, y; f_1; \kappa_1) \text{ for } (x, y) \in \overline{\Omega} \cap [K_1 - (\overline{K}' \cup \overline{K}'')], \\ \max [F(x, y; f_1; \kappa_1), u(x, y)] \text{ for } (x, y) \in \overline{\Omega} \cap (\overline{K}' \cup \overline{K}'')], \end{cases}$$

satisfies all the conditions of Definition 9 for being a barrier subfunction for $\kappa_1 = \kappa_1(\alpha_0)$ as prescribed in Theorem 15.

Since, by (23), (24), (25), and the definitions of $u'(x, y)$, $u''(x, y)$, and $u(x, y)$, we have

$$(30) \qquad u(x, y) < F(x, y; f_1; \kappa_1)$$

on the part of $\kappa' \cup \kappa''$ which lies in Ω; since $u(x, y)$ is continuous on $\kappa' \cup \kappa''$ except possibly at Q; and since, by (22), (26), and (27), there is a neighborhood of Q in which (30) holds, it follows that $s(x, y)$ is continuous in $\overline{\Omega} \cap \overline{K}_1$.

That $s(x, y)$ is a subfunction follows from exactly the same kind of argument as the one used in discussing $u(x, y)$.

By (20), (24), (25), (26), and (27), we have

$$u(x_0, y_0) = F(x_0, y_0; f_3; \kappa_1) \geq N - \epsilon,$$

whence, by (29), we have also

$$s(x_0, y_0) \geq N - \epsilon.$$

It follows from (19) that

$$F(x, y; f_1; \kappa_1) \leq N + 2\epsilon \qquad \text{in } \overline{K}_1$$

and from (21), (26), (27), and (28) that

$$u(x, y) \leq N + 2\epsilon \qquad \text{in } \overline{K}' \cup \overline{K}'',$$

whence, by (29),

$$s(x, y) \leq N + 2\epsilon \qquad \text{on } \Omega \cap K.$$

Finally, by (19), (23), (24), (25), and the definitions of $u'(x, y)$, $u''(x, y)$, $u(x, y)$, and $s(x, y)$, we have

$$s(x, y) \leq M \qquad \text{on } \overline{\Omega} \cap \kappa_1.$$

Thus $s(x, y)$ satisfies all the conditions of Definition 9, and is a barrier subfunction $s(x, y; \kappa_1; \epsilon, M, N)$ as desired.

REFERENCES

1. E. F. Beckenbach, *Generalized convex functions*, Bull. Amer. Math. Soc. 43 (1937), 363-371.

2. F. F. Bonsall, *On generalized subharmonic functions*, Proc. Cambridge Philos. Soc. 46 (1950), 387-395.

3. O. D. Kellogg, *Potential theory*, Springer, Berlin, 1929.

4. J. E. Littlewood, *Theory of functions*, Oxford University Press, Oxford, 1944.

5. F. Riesz, *Über subharmonische Funktionen und ihre Rolle in der Funktionentheorie und in der Potentialtheorie*, Acta Sci. Math. Szeged 2 (1925), 87-100.

6. Georg Tautz, *Zur Theorie der elliptischen Differentialgleichungen. I.*, Math. Ann. 117 (1939-41), 694-726.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

EXTENSION OF A RENEWAL THEOREM

DAVID BLACKWELL

1. Introduction. A chance variable x will be called a *d-lattice variable* if

$$(1) \qquad \sum_{n=-\infty}^{\infty} \Pr\{x = nd\} = 1$$

and

$$(2) \qquad d \text{ is the largest number for which (1) holds.}$$

If x is not a d-lattice variable for any d, x will be called a *nonlattice variable.* The main purpose of this paper is to give a proof of:

THEOREM 1. *Let $x_1, x_2, \cdots$ be independent identically distributed chance variables with $E(x_1) = m > 0$ (the case $m = +\infty$ is not excluded); let $S_n = x_1 + \cdots + x_n$; and, for any $h > 0$, let $U(a, h)$ be the expected number of integers $n \geq 0$ for which $a \leq S_n < a + h$. If the x_n are nonlattice variables, then*

$$U(a, h) \longrightarrow \frac{h}{m}, \ 0 \qquad \textit{as } a \longrightarrow +\infty, -\infty.$$

If the x_n are d-lattice variables, then

$$U(a, d) \longrightarrow \frac{d}{m}, \ 0 \qquad \textit{as } a \longrightarrow +\infty, -\infty.$$

(*If $m = +\infty$, h/m and d/m are interpreted as zero.*)

This theorem has been proved (A) for nonnegative d-lattice variables by Kolmogorov [5] and by Erdös, Feller, and Pollard [4]; (B) for nonnegative nonlattice variables by the writer [1], using the methods of [4]; (C) for d-lattice variables by Chung and Wolfowitz [3]; (D) for nonlattice variables such that the distribution of some S_n has an absolutely continuous part and $m < \infty$ by Chung

Received June 28, 1952. This paper was written under an Office of Naval Research contract.

Pacific J. Math. 3 (1953), 315-320

and Pollard [2], using a purely analytical method; and (E) in the form given here by Harris (unpublished). Harris' proof does not essentially use the results of the special cases (A), (B), (C), (D); the proof given here obtains Theorem 1 almost directly from the special cases (A) and (B) by way of an integral identity and an equation of Wald.

2. **An integral identity.** Let N_1 be the smallest n for which $S_n > 0$, and write $z_1 = S_{N_1}$; let N_2 be the smallest $n > 0$, for which $S_{N_1+n} - S_{N_1} > 0$, and write $z_2 = S_{N_1+N_2} - S_{N_1}$, and so on. Continuing in this way, we obtain sequences N_1, $N_2, \cdots; z_1, z_2, \cdots$ of independent, positive, identically distributed chance variables such that

$$S_{N_1+\cdots+N_K} = z_1 + \cdots + z_K .$$

Let $V(t)$, $R(t)$ denote the expected number of integers $n \geq 0$ for which

$$T_n = z_1 + \cdots + z_n \leq t \text{ and } -t \leq S_n \leq 0,$$

$n < N_1$, respectively. That $V(t) < \infty$ follows from a theorem of Stein [6], and that $R(t) < \infty$ follows from $E(N_1) < \infty$, which we show in the next section. The integral identity is:

THEOREM 2. $U(a, h) = \int_0^\infty [R(t-a) - R(t-a-h)]\, dV(t)$.

Proof. If n_K is the number of integers n with

$$N_1 + \cdots + N_K \leq n < N_1 + \cdots + N_{K+1} \text{ and } a \leq S_n < a + h,$$

we have

$$E(n_K \mid T_K = t) = R(t-a) - R(t-a-h),$$

so that

$$E(n_K) = \int_0^\infty [R(t-a) - R(t-a-h)]\, dF_K(t),$$

where $F_K(t) = \Pr\{T_K \leq t\}$. Summing over $K = 0, 1, 2, \cdots$, and using the fact that

$$V(t) = \sum_{K=0}^{\infty} F_K(t),$$

we obtain the theorem.

3. Wald's equation. The main purpose of this section is to note that $E(N_1)$ is finite, so that an equation of Wald [7, p. 142] holds.

THEOREM 3. *$E(N_1) < \infty$ and $mE(N_1) = E(z_1)$, so that m, $E(z_1)$ are both finite or both infinite.*

Proof. In showing $E(N_1)$ finite, we may suppose $\{x_n\}$ bounded above; for defining $x_n^* = \min\{s_n, M\}$ yields an $N_1^* \geq N_1$; choosing M sufficiently large makes $E(x_n^*) > 0$, and $E(N_1^*) < \infty$ implies $E(N_1) < \infty$. Since

$$\frac{T_K}{K} = \frac{S_{N_1 + \cdots + N_K}}{N_1 + \cdots + N_K} \cdot \frac{N_1 + \cdots + N_K}{K},$$

we obtain, letting $K \longrightarrow \infty$ and using the strong law of large numbers, first that $E(z_1) = mE(N_1)$ and next since if $\{x_n\}$ is bounded above and $\{z_n\}$ is bounded, that $E(N_1)$ is finite in this case and consequently in general.

4. The d-lattice case. For d-lattice variables, Theorem 2 yields

$$U(nd, d) = \sum_{s=0}^{\infty} r(s-n)\, v(s) = \sum_{s=0}^{\infty} r(s)\, v(s+n), \tag{3}$$

where $r(s) = R(sd) - R([s-1]d)$ and $v(s) = V(sd) - V([s-1]d)$. Now

$$\sum_{s=0}^{\infty} r(s) = \lim_{t \to \infty} R(t) = E(N_1) < \infty.$$

Theorem (A) asserts that

$$v(n) \longrightarrow \frac{d}{E(z_1)}, 0 \qquad \text{as } n \longrightarrow \infty, -\infty;$$

applying this to (1) yields

$$U(nd, d) \longrightarrow \frac{dE(N_1)}{E(z_1)}, 0 \qquad \text{as } n \longrightarrow \infty, -\infty,$$

and Wald's equation yields Theorem 1 for d-lattice variables.

5. The nonlattice case. For nonlattice variables we have, rewriting Theorem

2 with a change of variable,

$$U(a, h) = \int_M^\infty [R(t) - R(t-h)] \, dV(t+a).$$

For any $M > 0$, write

$$U(a, h) = I_1(M, a, h) + I_2(M, a, h),$$

where

$$I_1 = \int_0^M [R(t) - R(t-h)] \, dV(t+a)$$

and

$$I_2 = \int_0^\infty [R(t) - R(t-h)] \, dV(t+a).$$

Theorem B applied to $\{z_n\}$ yields

$$V(t+h) - V(t) \longrightarrow \frac{h}{E(z_1)}$$

for all $h > 0$ as $t \longrightarrow \infty$, so that, since $R(t)$ is monotone,

$$I_1 = \int_0^M R(t) \, dV(t+a) - \int_0^{M-h} R(t) \, dV(t+a+h)$$

$$\longrightarrow \frac{1}{E(z_1)} \cdot \int_{M-h}^M R(t) \, dt, \ 0 \qquad \text{as } a \longrightarrow \infty, -\infty$$

for fixed M, h. We now show that, for fixed h, $I_2(M, a, h) \longrightarrow 0$ as $M \longrightarrow \infty$ uniformly in a. We have

$$I_2 = \sum_{n=0}^\infty \int_{M+nh}^{M+(n+1)h} [R(t) - R(t-h)] \, dV(t+a)$$

$$\leq \sum_{n=0}^\infty R_1(M, n) \, [V(a+M+(n+1)h) - V(a+M+nh)],$$

where

$$R_1(M, n) = \sup [R(t) - R(t-h)]$$

as t varies over the interval $(M + nh, M + (n+1)h)$. Since, by Theorem (B),

$$V(b+h) - V(b) \longrightarrow \frac{h}{E(z_1)} \qquad \text{as } b \longrightarrow \infty,$$

there is a constant c (for the given h) such that

$$I_2(M, a, h) \leq c \sum_{n=0}^{\infty} R_1(M, n) \qquad \text{for all } M \text{ and } a.$$

Now

$$\sum_{n=0}^{\infty} R_1(M, 2n) \leq E(N_1) - R(M) \text{ and } \sum_{n=0}^{\infty} R_1(M, 2n+1) \leq E(N_1) - R(M),$$

and $R(M) \longrightarrow E(N_1)$ as $M \longrightarrow \infty$. Thus

$$|U(a, h) - I_1(M, a, h)| < \epsilon(M, h)$$

for all a, where $\epsilon(M, h) \longrightarrow 0$ *as* $M \longrightarrow \infty$ for fixed h. Then

$$\left|U(a, h - \frac{hE(N_1)}{E(z_1)}\right| \leq \epsilon(m, h) + \left|I_1(M, a, h) - \frac{1}{E(z_1)} \int_{M-h}^{M} R(t)\, dt\right| + \left|\frac{1}{E(z_1)} \int_{M-h}^{M} R(t)\, dt - hE(N_1)\right|,$$

so that

$$\limsup_{a \to \infty} \left|U(a, h) - \frac{hE(N_1)}{E(z_1)}\right| \leq \epsilon(M, h) + \frac{1}{E(z_1)} \left|\int_{M-h}^{M} R(t)\, dt - hE(N_1)\right|.$$

Letting $M \longrightarrow \infty$ yields

$$U(a, h) \longrightarrow \frac{hE(N_1)}{E(z_1)} \qquad \text{as } a \longrightarrow \infty,$$

and Wald's equation yields Theorem 1 for $a \longrightarrow \infty$. Similarly,

$$U(a, h) \leq \epsilon(M, h) + |I_1(M, a, h)|$$

for all a, so that

$$\limsup_{a \to -\infty} U(a, h) \leq \epsilon(M, h)$$

and $U(a, h) \longrightarrow 0$ as $a \longrightarrow -\infty$. This completes the proof.

References

1. D. Blackwell, *A renewal theorem*, Duke Math. J. 15 (1948), 145-151.

2. K. L. Chung and Harry Pollard, *An extension of renewal theory*, Proc. Amer. Math. Soc. 3 (1952), 303-309.

3. K. L. Chung and J. Wolfowitz, *On a limit theorem in renewal theory*, Ann. of Math. 55 (1952), 1-6.

4. P. Erdös, W. Feller, and H. Pollard, *A theorem on power series*, Bull. Amer. Math. Soc. 55 (1949), 201-204.

5. A. Kolmogorov, *Anfangsgründe der Theorie der Markoffschen Ketten mit unendlich vielen möglichen Zuständen*, Mat. Sbornik N.S. 1 (1936), 607-610.

6. C. Stein, *A note on cumulative sums*, Ann. Math. Statistics 17 (1946), 498-499.

7. A. Wald, *Sequential tests of statistical hypotheses*, Ann. Math. Statistics 16 (1945) 117-186.

STANFORD UNIVERSITY

SOME THEOREMS ON THE SCHUR DERIVATIVE

L. CARLITZ

1. Introduction. Given the sequence $\{a_m\}$ and $p \neq 0$, Schur [5] defined the derivative a_m' by

$$(1.1) \qquad a_m' = \Delta a_m = (a_{m+1} - a_m)/p^{m+1};$$

higher derivatives are defined by means of

$$a_m^{(r)} = \Delta^r a_m = \Delta(a_m^{(r-1)}), \quad a_m^{(0)} = a_m.$$

In particular if p is a prime, a an integer and $a_m = a^{p^m}$, then by Fermat's theorem

$$a_m' = (a^{p^{m+1}} - a^{p^m})/p^{m+1}$$

is integral. Schur proved that if $p \nmid a$, then also the derivatives

$$\Delta^2 a^{p^m},\ \Delta^3 a^{p^m},\ \cdots,\ \Delta^{p-1} a^{p^m}$$

are all integral. Moreover if $a_0' \equiv 0 \pmod p$ then all the derivatives $\Delta^r a^{p^m}$ are integral, while if $a_0' \not\equiv 0 \pmod p$ then every number of $\Delta^p a^{p^m}$ has the denominator p.

A. Brauer [1] gave another proof of Schur's results. About the same time Zorn [6] proved these results by p-adic methods and indeed proved the following stronger theorem. For $x \equiv 1 \pmod p$, define

$$X_m = (x^{p^m} - 1)/p^{m+1},$$

and as above let $\Delta^r X_m$ denote the r-th derivative of X_m; then

$$(1.2) \qquad \Delta^r X_m \equiv \frac{(p-1)(p^2-1)\cdots(p^r-1)}{(r+1)!} X_m^{r+1} \pmod{p^m}$$

provided $r < p$; for $r < p-2$, the congruence (1.2) holds $(\text{mod } p^{m+1})$. It is also shown that Schur's theorem is an easy consequence of Zorn's results.

Received November 16, 1951.

Pacific J. Math. 3 (1953), 321-332

In the present paper we shall give a simple elementary proof of Zorn's congruences. In addition we prove, for example, that for $r \leq p$,

$$\Delta^r a^{p^m} \equiv \frac{1}{r!} a^{p^m} q_m^r \frac{\prod_{i=1}^r (p^i - 1)}{(p-1)^r} \pmod{p^m}, \tag{1.3}$$

where

$$a^{(p-1)p^m} = 1 + p^{m+1} q_m;$$

for $r < p - 1$, (1.3) holds $\pmod{p^{m+1}}$.

We next (§4) extend Schur's and Zorn's theorems to algebraic numbers. In §5 we consider a generalization of another kind suggested by the arithmetic function (see for example [2, p. 84-86])

$$F(a, m) = \sum_{de=m} \mu(d) a^e. \tag{1.4}$$

Finally (§6), we give some applications of Schur's theorem to the Euler and Bernoulli polynomials and numbers; the results are analogous to Kummer's congruences [3, Ch. 12]. In particular $\Delta^r E_{k+p^m}$ is integral (mod p) for $p > 2, r < p$, $r \leq m$; also $\Delta^r (B_{k+p^m}/(k + p^m))$ is integral (mod p) for $p - 1 \nmid k + 1$, $r < p$, $r \leq m$. Here E_k and B_k denote the Euler and Bernoulli numbers in the notation of Nörlund [3].

2. Formulas for $\Delta^r a_m$. We shall require some preliminary results.

LEMMA 1. *The following identity holds*:

$$\prod_{i=0}^{r=1} (x - p^i) = \sum_{i=0}^{r} (-1)^i \left[{r \atop i}\right] p^{i(i-1)/2} x^{r-i}, \tag{2.1}$$

where

$$\left[{r \atop i}\right] = \frac{(p^r - 1)(p^{r-1} - 1) \cdots (p^{r-i+1} - 1)}{(p-1)(p^2-1) \cdots (p^i - 1)} = \left[{r \atop r-i}\right], \left[{r \atop 0}\right] = 1. \tag{2.2}$$

LEMMA 2. *Put*

$$W_{k,r} = \sum_{i=0}^{k} (-1)^i \left[{k \atop i}\right] \binom{p^{k-i}}{r} p^{i(i-1)/2},$$

where $\binom{m}{r}$ *denotes a binomial coefficient. Then*

$$W_{k,r} = \begin{cases} 0 & (r<k) \\ \dfrac{1}{r!} \prod_{i=0}^{r-1} (p^r - p^i) & (r=k) \\ \dfrac{1}{r!}\, p^{k(k-1)/2}\, U_{k,r} & (r>k), \end{cases} \tag{2.3}$$

where $U_{k,r}$ *is an integer.*

Lemma 1 is will known. To prove Lemma 2, we note first that the binomial coefficient $\binom{x}{r}$ is a polynomial in x of degree r. Since by (2.1)

$$\sum_{i=0}^{k} (-1)^i \left[{k \atop i}\right] p^{i(i-1)/2}\, p^{r(k-i)} = \prod_{i=0}^{k-1} (p^r - p^i),$$

the several parts of (2.3) follow without much difficulty.

LEMMA 3. *For an arbitrary sequence* $\{a_m\}$,

$$\Delta^r a_m = p^{-rm-r(r+1)/2} \sum_{i=0}^{r} (-1)^i \left[{r \atop i}\right] p^{i(i-1)/2}\, a_{m+r-i}\,. \tag{2.4}$$

This formula, which is given by Schur, is easily proved. In view of (2.1) it can be put in the following symbolic form:

$$\Delta^r a_m = p^{-rm-r(r+1)/2}\, a^m \prod_{i=0}^{r-1} (a - p^i), \tag{2.5}$$

where it is understood that after expansion of the right member a^k is to be replaced by a_k.

Suppose now that $p \nmid a$ and put

$$a^{(p-1)p^m} = 1 + p^{m+1} q_m, \tag{2.6}$$

so that q_m is integral. Then by the binomial theorem we have

$$a^{(p-1)p^{m+s}} = \sum_{i=0}^{p^r} \binom{p^s}{i} p^{(m+1)i} q_m^i \qquad (r \geq s),$$

and by (2.4) this implies

$$p^{rm+r(r+1)/2}\ \Delta^r a^{(p-1)p^m}$$

$$= \sum_{s=0}^{r} (-1)^{r-s} \left[{r \atop s}\right] p^{(r-s)(r-s-1)/2} \sum_{i=0}^{p^r} \binom{p^s}{i} p^{(m+1)i} q_m^i$$

$$= \sum_{i=0}^{p^r} p^{(m+1)i} q_m^i \sum_{s=0}^{r} (-1)^{r-s} \left[{r \atop s}\right] \binom{p^s}{i} p^{(r-s)(r-s-1)/2}$$

$$= \sum_{i=0}^{p^r} p^{(m+1)i} q_m^i W_{r,i}$$

$$= \frac{1}{r!} p^{rm+r(r+1)/2} q_m^r \prod_{i=1}^{r} (p^i - 1)$$

$$+ \sum_{i=r+1}^{p^r} \frac{1}{i!} p^{(m+1)i+r(r-1)/2} q_m^i U_{r,i},$$

by (2.3); $W_{r,i}$ and $U_{r,i}$ have the same meaning as in Lemma 2. We thus get

$$\Delta^r a^{(p-1)p^m} = \frac{1}{r!} q_m^r \prod_{i=1}^{r} (p^i - 1) + \sum_{i=r+1}^{p^r} \frac{1}{i!} p^{(m+1)(i-r)} q_m^i U_{r,i}. \tag{2.7}$$

We next set up a similar formula for $\Delta^r q_m$, where q_m is defined by (2.6). Indeed substitution in (2.4) gives

$$p^{rm+r(r+1)/2}\ \Delta^r q_m = \sum_{s=0}^{r} (-1)^{r-s} \left[{r \atop s}\right] p^{(r-s)(r-s-1)/2-(m+s+1)} \left(a^{(p-1)p^{m+s}} - 1\right)$$

$$= \sum_{s=0}^{r} (-1)^{r-s} \left[{r \atop s}\right] p^{(r-s)(r-s-1)/2-(m+s+1)} \sum_{i=1}^{p^r} \binom{p^s}{i} p^{(m+1)i} q_m^i$$

$$= \sum_{i=1}^{p^r} p^{(m+1)(i-1)} q_m^i \sum_{s=0}^{r} (-1)^{r-s} \left[{r \atop s}\right] \binom{p^s}{i} p^{(r-s)(r-s-1)/2-s}$$

$$= \frac{1}{(r+1)!} p^{rm+r(r+1)/2} q_m^{r+1} \prod_{i=1}^{r} (p^i - 1)$$

$$+ \sum_{i=r+2}^{p^r} \frac{1}{i!} p^{(m+1)(i-1)+r(r-1)/2} q_m^i U'_{r,i},$$

by a slight modification of Lemma 2; the coefficient $U'_{r,i}$ is integral and is defined by

$$\frac{1}{i!}\, p^{r(r-1)/2}\, U'_{r,i} = \sum_{s=0}^{r} (-1)^s \left[{r \atop s} \right] (p_i^{r-s})\, p^{s(s-1)/2-(r-s)}.$$

Hence

$$(2.8)\quad \Delta^r q_m = \frac{1}{(r+1)!}\, q_m^{r+1} \prod_{i=1}^{r} (p^i - 1) + \sum_{i=r+2}^{p^r} \frac{1}{i!}\, p^{(m+1)(i-r-1)}\, q_m^i\, U'_{r,i}.$$

Using the same method we can also evaluate $\Delta^r a^{p^m}$. It follows from (2.6) that

$$(2.9)\qquad a^{p^{m+s}} = a^{p^m} (1 + p^{m+1} q_m)^{e_s} \qquad \left(e_s = \frac{p^s - 1}{p - 1} \right),$$

and thus substitution in (2.4) yields

$$p^{rm+r(r+1)/2}\, \Delta^r a^{p^m} = a^{p^m} \sum_{s=0}^{r} (-1)^{r-s} \left[{r \atop s} \right] p^{(r-s)(r-s-1)/2} \sum_{i=0}^{e_r} \binom{e_s}{i} p^{(m+1)i} q_m^i$$

$$= a^{p^m} \sum_{i=0}^{e_r} p^{(m+1)i}\, q_m^i \sum_{s=0}^{r} (-1)^{r-s} \left[{r \atop s} \right] \binom{e_s}{i} p^{(r-s)(r-s-1)/2}.$$

Since $\binom{e_s}{i}$ is a polynomial in p^s of degree i, the same reasoning as before applies and we get after a little manipulation

$$(2.10)\qquad \Delta^r a^{p^m} = \frac{1}{r!}\, a^{p^m}\, q_m^r\, \frac{\prod_{i=1}^{r} (p^i - 1)}{(p-1)^r} + a^{p^m} \sum_{i=r+1}^{e_r} \frac{1}{i!}\, p^{(m+1)(i-r)}\, q_m^i\, U''_{r,i},$$

where $U''_{r,i}$ is integral.

Comparison of (2.7) and (2.10) shows that (2.7) is included in (2.10). Indeed it is easy to set up the following formula which includes both (2.7) and (2.10):

$$(2.11) \qquad \Delta^r a^{kp^m} = \frac{1}{r!} a^{kp^m} q_m^r k^r \frac{\prod_{i=1}^r (p^i - 1)}{(p-1)^r} + a^{kp^m} \sum_{i=r+1}^{e_r} \frac{1}{i!} p^{(m+1)(i-r)} q_m^i V_{r,i},$$

where $V_{r,i} = V_{r,i}^{(k)}$ is integral and $k \geq 1$. The proof of (2.11) is exactly like the proof of (2.10); the first step is to raise both members of (2.9) to the k-th power.

3. The main results. In order to make use of (2.7) and (2.10) it is evidently necessary to examine $p^{(m+1)(i-r)}/i!$. We suppose $i > r$, $r \leq p$. Then in the first place it is easily seen [6, p. 462] that $p^{i-r}/i!$ is integral (mod p), and a simple discussion shows that $p^{i-r}/i!$ is divisible by p unless (i) $i = p$, $r = p - 1$, or (ii) $i = p + 1$, $r = p$. We now state:

THEOREM 1. *The derivative $\Delta^r a^{(p-1)p^m}$ is integral for $1 \leq r \leq p - 1$, while $\Delta^p a^{(p-1)p^m}$ has the denominator p provided $a^{p-1} \not\equiv 1 \pmod{p^2}$; if $a^{p-1} \equiv 1 \pmod{p^2}$ then all $\Delta^r a^{(p-1)p^m}$ are integral.*

THEOREM 2. *For $1 \leq r \leq p$, $m \geq 0$,*

$$(3.1) \qquad \Delta^r a^{(p-1)p^m} \equiv \frac{1}{r!} q_m^r \prod_{i=1}^{r} (p^i - 1) \pmod{p^m};$$

if $r < p - 1$, the congruence is valid $\pmod{p^{m+1}}$.

THEOREM 3. *The derivative $\Delta^r a^{p^m}$ is integral for $1 \leq r \leq p - 1$, while $\Delta^p a^{p^m}$ has the denominator p provided $a^{p-1} \not\equiv 1 \pmod{p^2}$; if $a^{p-1} \equiv 1 \pmod{p^2}$ then all $\Delta^r a^{(p-1)p^m}$ are integral.*

THEOREM 4. *For $1 \leq r \leq p$, $m \geq 0$,*

$$(3.2) \qquad \Delta^r a^{p^m} \equiv \frac{1}{r!} a^{p^m} q_m^r \frac{\prod_{i=1}^r (p^i - 1)}{(p-1)^r} \pmod{p^m};$$

if $r < p - 1$, the congruence is valid $\pmod{p^{m+1}}$.

If we make use of (2.11) rather than (2.7) or (2.10) we get the following more general result.

THEOREM 4′. *For $1 \leq r \leq p$, $m \geq 0$*

$$\Delta^r a^{kp^m} \equiv \frac{1}{r!}\, a^{kp^m}\, q_m^r k^r\, \frac{\prod_{i=1}^r (p^i - 1)}{(p-1)^r} \pmod{p^m};$$

if $r < p - 1$, *the congruence is valid* $\pmod{p^{m+1}}$.

To apply (2.8) we first examine $p^{i-r-1}/i!$ for $i > r + 1$, $r + 1 \leq p$. We have:

THEOREM 5. *The derivative* $\Delta^r q_m$ *is integral for* $1 \leq r \leq p - 2$, *while* $\Delta^{p-1} q_m$ *has the denominator* p *provided* $a^{p-1} \not\equiv 1 \pmod{p^2}$; *if* $a^{p-1} \equiv 1 \pmod{p^2}$ *then all* $\Delta^r q_m$ *are integral.*

THEOREM 6. *For* $1 \leq r \leq p - 1$, $m \geq 0$,

$$\Delta^r q_m \equiv \frac{1}{(r+1)!}\, q_m^{r+1} \prod_{i=1}^{r} (p^i - 1) \pmod{p^m}; \tag{3.3}$$

if $r < p - 2$, *the congruence is valid* $\pmod{p^{m+1}}$.

Theorem 3 is of course Schur's theorem; Theorems 5 and 6 are due to Zorn. The remaining theorems are presumably new.

4. Generalization for algebraic numbers. Let k be an algebraic number field of degree n and let $\mathfrak{p}$ denote a prime ideal of k; also let

$$N\mathfrak{p} = p^f; \qquad \mathfrak{p}^e \mid p, \qquad \mathfrak{p}^{e+1} \nmid p; \tag{4.1}$$

for simplicity we assume $p > n$. If α k is integral $\pmod{\mathfrak{p}}$ and $\mathfrak{p} \nmid \alpha$, then by Fermat's Theorem

$$\alpha^{p^f - 1} = 1 + \beta, \qquad \beta \equiv 0 \pmod{\mathfrak{p}}. \tag{4.2}$$

It follows from (4.2) that

$$\alpha^{(p^f-1)p^m} = 1 + \beta_m, \qquad \beta_m \equiv 0 \pmod{\mathfrak{p}^{me+1}}, \tag{4.3}$$

while (4.3) implies

$$\alpha^{(p^f-1)p^{m+s}} = \sum_{i=0}^{p^r} \binom{p^s}{i} \beta_m^i \qquad (r \geq s). \tag{4.4}$$

Then, exactly as in § 2,

$$p^{rm+r(r+1)/2}\ \Delta^r\ \alpha^{(p^f-1)p^m} = \sum_{s=0}^{r} (-1)^{r-s} \left[{r \atop s}\right] p^{(r-s)(r-s-1)/2} \sum_{i=0}^{p^r} \binom{p^s}{i} \beta_m^i$$

$$= \sum_{i=0}^{p^r} \beta_m^i \sum_{s=0}^{r} (-1)^{r-s} \left[{r \atop s}\right] \binom{p^s}{i} p^{(r-s)(r-s-1)/2};$$

application of Lemma 2 now leads to

$$(4.5)\ \Delta^r\ \alpha^{(p^f-1)p^m} = \frac{1}{r!}\ p^{-r(m+1)}\ \beta_m^r \prod_{i=1}^{r}(p^i-1) + \sum_{i=r+1}^{p^r} \frac{1}{i!}\ p^{-r(m+1)}\ \beta_m^i\,\omega_{r,i},$$

where $\omega_{r,i}$ is integral. Note that for $e > 1$ the right member of (4.5) need not be integral. Accordingly we assume $e = 1$; the assumption $p > n$ is then no longer needed.

We now have:

THEOREM 7. *Let* $N\mathfrak{p} = p^f$, $\mathfrak{p}^2 \nmid p$, $\mathfrak{p} \nmid \alpha$; *then* $\Delta^r\ \alpha^{(p^f-1)p^m}$ *is integral for* $1 \le r \le p-1$, *while* $\Delta^p\ \alpha^{(p^f-1)p^m}$ *has the denominator* p *provided* $\alpha^{p^f-1} \not\equiv 1 \pmod{\mathfrak{p}^2}$; *if* $\alpha^{p^f-1} \equiv 1 \pmod{\mathfrak{p}^2}$ *then all* $\Delta^r\ \alpha^{(p^f-1)p^m}$ *are integral.*

THEOREM 8. *With the hypotheses of Theorem* 7,

$$(4.6) \qquad \Delta^r\ \alpha^{(p^f-1)p^m} \equiv \frac{1}{r!}\left(\frac{\beta_m}{p^{m+1}}\right)^r \prod_{i=1}^{r} (p^i-1) \pmod{\mathfrak{p}^m}$$

for $r \le p$; *if* $r < p-1$ *the congruence is valid* $\pmod{\mathfrak{p}^{m+1}}$.

In order to extend Theorems 3 and 4′ it is convenient to suppose that $\mathfrak{p}$ is a prime ideal of the first degree. The following two theorems may be proved.

THEOREM 9. *Let* $N\mathfrak{p} = p$, $\mathfrak{p}^2 \nmid p$, $\mathfrak{p} \nmid \alpha$; *then* $\Delta^r\ \alpha^{p^m}$ *is integral for* $1 \le r \le p-1$, *while* $\Delta^p\ \alpha^{p^m}$ *has the denominator* p *provided* $\alpha^{p-1} \not\equiv 1 \pmod{\mathfrak{p}^2}$; *if* $\alpha^{p-1} \equiv 1 \pmod{\mathfrak{p}^2}$ *then all* $\Delta^r\ \alpha^{p^m}$ *are integral.*

THEOREM 10. *With the hypotheses of Theorem* 9,

$$(4.7) \qquad \Delta^r\ \alpha^{kp^m} \equiv \frac{1}{r!}\left(\frac{k\,\beta_m}{p^{m+1}}\right)^r \frac{\prod_{i=1}^{r}(p^i-1)}{(p-1)^r} \pmod{\mathfrak{p}^m}$$

for $r \le p$; *if* $r < p-1$ *the congruence is valid* $\pmod{\mathfrak{p}^{m+1}}$.

For brevity we omit the extension of Theorems 5 and 6 for algebraic numbers.

5. Another generalization. Changing slightly the notation (1.1) we put

$$\Delta_p a_{mp^i} = (a_{mp^{i+1}} - a_{mp^i})/p^{i+1}, \tag{5.1}$$

and

$$\Delta_p^r a_{mp^i} = (\Delta_p^{r-1} a_{mp^{i+1}} - \Delta_p^{r-1} a_{mp^i})/p^{i+1}.$$

Then clearly $\Delta_p \Delta_q = \Delta_q \Delta_p$. If a and k are arbitrary integers then if follows from a well-known theorem concerning (1.4) that

$$\delta_k a^k = \Delta_{p_1} \cdots \Delta_{p_s} a^k \qquad (k = p_1^{e_1} \cdots p_s^{e_s}) \tag{5.2}$$

is integral. In view of Schur's theorem we can state the following generalization.

THEOREM 11. *Let $(a,k) = 1$ and let $r <$ the smallest prime dividing k; define*

$$\delta_k^r a^k = \delta_k \delta_k^{r-1} a^k. \tag{5.3}$$

Then $\delta_k^r a_k$ is integral for $k > 1$.

Indeed because of the commutativity of the operators Δ_{p_i} we need only observe that (5.2) and (5.3) imply

$$\delta_k^r a^k = \Delta_{p_1}^r \cdots \Delta_{p_s}^r a^k \tag{5.4}$$

and the theorem follows immediately.

The restriction $(a,k) = 1$ can be removed by taking k sufficiently large as we shall see below.

A slight extension of Theorem 11 is contained in:

THEOREM 12. *Let*

$$(a,k) = 1, \quad k = p_1^{e_1} \cdots p_s^{e_s},$$

and let $r_i < p_i$, $j = 1, \cdots, s$; then

$$\Delta_{p_1}^{r_1} \cdots \Delta_{p_s}^{r_s} a^k \tag{5.5}$$

is integral for all $k > 1$.

We remark that the function defined in (5.2) can also be expressed in the form

$$\delta_k a^k = \frac{(-1)^s}{k_1} \sum_{d \mid k} \mu(d)\, a^{dk},$$

where $\mu(d)$ is the Möbius function and

$$k_1 = p_1^{e_1+1} \cdots p_s^{e_s+1};$$

similarly (5.3) becomes

$$\delta_k^r a^k = \frac{(-1)^s}{k_1} \sum_{d \mid k} \mu(d)\, \delta_k^{r-1} a^{dk}.$$

Formulas of a different kind can be obtained by applying (2.4) to (5.4) and (5.5); for example, (2.5) suggests the following symbolic formula:

$$\delta_k^r a^k = k^{-r} \prod_{j=1}^{s} p_j^{r(r+1)/2} \cdot \prod_{j=1}^{s} a_j^{e_j} \prod_{i=0}^{r=1} (a_j - p_j^i),$$

where after expansion $a_1^{f_1} \cdots a_s^{f_s}$ is to be replaced by a^m,

$$m = p_1^{f_1} \cdots p_s^{f_s}.$$

A similar but slightly more complicated formula can be stated for (5.5). We shall omit the generalization of Theorems 11 and 12 to algebraic numbers.

6. Applications. In the theorems of § 2 it is assumed that $p \nmid a$. However Theorem 3, for example, is easily extended to the case $p \mid a$. We can state that $\Delta^r a^{p^m}$ is integral for $r \leq p-1$ and arbitrary a provided $m \geq r$. For let $p \mid a$; then, in view of (2.4), it is only necessary to verify that

$$p^{m+r-i} + \frac{1}{2}\, i(i-1) \geq rm + \frac{1}{2}\, r(r+1)$$

for $0 \leq i \leq r \leq p-1$, $r \geq m$. This can be proved by induction with respect to m. In the next place since Theorem 11 is a direct consequence of Theorem 3 we infer that it also holds for all a provided $r \leq \min(e_1, \cdots, e_s)$ in the notation of Theorem 11.

Now consider the number

$$C_k = \sum_{a=1}^{n} A_a a^k, \tag{6.1}$$

where A_a denote integers (mod p) and $n \geq 1$ is arbitrary. Then

(6.2) $$\Delta^r C_{k+p^m} = \sum_{a=1}^{n} A_a \Delta^r a^{k+p^m} \qquad (k \geq 0),$$

so that by the remark in the previous paragraph $\Delta^r C_{p^m}$ is certainly integral (mod p) provided $r \leq p-1$ and $r \leq m$. In the second place we may apply the operator δ_k^r defined in (5.2) and (5.3) and get

(6.3) $$\delta_k^r C_{h+k} = \sum_{a=1}^{n} A_a \delta_k^r a^{h+k};$$

we infer that $\delta_k^r C_k$ is integral provided $r <$ the smallest prime dividing k and $r \leq \min(i_1, \cdots, i_s)$, the notation being that of (5.2). Indeed a somewhat more general result can be obtained by applying Theorem 15, namely,

(6.4) $$\Delta_{p_1}^{r_1} \cdots \Delta_{p_s}^{r_s} C_{h+k} \qquad (h \geq 0)$$

is integral provided $r_t < p_t$, $r_t \leq e_t$, $t = 1, \cdots, s$.

As an instance of (6.1) we take the well-known formula for the Euler polynomial

(6.5) $$E_m(x) = \sum_{s=0}^{m} \frac{1}{2^s} \sum_{i=0}^{s} (-1)^i \binom{s}{i} (x+i)^m.$$

(We use the notation of Nörlund [4] for the Euler and Bernoulli polynomials.) If $p > 2$ and x is integral (mod p) the preceding discussion applies. In particular using (2.4) we have:

THEOREM 13. *Let $p > 2$ and x be integral* (mod p). *Then*

$$\Delta^r E_{k+p^m}(x) = p^{-rm-r(r+1)/2} \sum_{i=0}^{r} (-1)^i \left[{r \atop i}\right] p^{i(i-1)/2} E_{k+p^{m-i}}(x)$$

is integral (mod p) *provided* $r < p$, $r \leq m$.

For brevity we omit the generalizations corresponding to (6.3) and (6.4). The special case

(6.6) $$\sum_{de=m} \mu(d) E_{k+e}(x) \equiv 0 \qquad (\text{mod } m)$$

may be noted

As for the Bernoulli polynomials, it can be shown that if $p \nmid a$ and x is integral (mod p) then a formula of the type (6.1) holds for

$$\beta_k(x) = \frac{a^{k+1}-1}{k+1} B_{k+1}(x). \tag{6.7}$$

(See for example Nielsen [3, Ch. 14].) Thus it follows that

$$\Delta^r \beta_{k+p^m}(x) = p^{-rm-r(r+1)/2} \sum_{i=0}^{r} (-1)^i \left[{r \atop i}\right] p^{i(i-1)/2} \beta_{k+p^{m-i}}(x)$$

is integral for $r < p$, $r \leq m$. If now we assume $p-1 \nmid k$ and take a a primitive root (mod p) such that $a^{p-1} \equiv 1 \pmod{p^r}$ we get:

THEOREM 14. *Let $p > 2$ and x be integral* (mod p); *put* $H_k(x) = B_k(x)/k$. *Then if $p-1 \nmid k+1$,*

$$\Delta^r H_{k+p^m}(x) = p^{-rm-r(r+1)/2} \sum_{i=0}^{r} (-1)^i \left[{r \atop i}\right] p^{i(i-1)/2} H_{k+p^{m-i}}(x)$$

is integral for $r < p$, $r \leq m$.

Finally corresponding to (6.6) we state

$$\sum_{de=m} \mu(d) \beta_{k+e}(x) \equiv 0 \pmod{m},$$

for $\beta_k(x)$ as defined in (6.7).

REFERENCES

1. A. Brauer, *Über eine Erweiterung des kleinen Fermatschen Satzes*, Math. Z. 42 (1937), 255-262.

2. L. E. Dickson, *History of the theory of numbers*, vol. 1, Washington, 1919.

3. Niels Nielsen, *Traité élémentaire des nombres de Bernoulli*, Paris, 1923.

4. N. E. Nörlund, *Differenzenrechnung*, Berlin, 1924.

5. I. Schur, *Ein Beitrag zur elementaren Zahlentheorie*, Sitzungsberichte der Preussischen Akademie der Wissenschaften (1933), 145-151.

6. Max Zorn, *p-adic analysis and elementary number theory*, Ann. of Math. (2) 38(1937) 451-464.

DUKE UNIVERSITY

GENERALIZED CONVEXITY AND SURFACES OF NEGATIVE CURVATURE

PAUL A. CLEMENT

Introduction. In a study [4] of surfaces whose Gaussian, or total, curvature K satisfies the relation $K \leq 0$, a number of functions having geometrical significance have been shown to be convex. In the present paper, a study of surfaces whose Gaussian curvature satisfies $K \leq K_0$, where K_0 is a negative constant, leads to the determination of a class of functions which are subfunctions (defined in § 1.1) of a two-parameter family of functions determined by the bound K_0. This is a natural generalization because the convexity property is equivalent to the subfunction property with respect to the particular two-parameter family (nonvertical straight lines) determined by the bound $K_0 = 0$.

A main objective will be to exhibit functions which have a geometrical significance and also have the subfunction property for surfaces with $K \leq K_0$. This property then implies certain inequality relations for functions associated with certain geometrical configurations on such surfaces.

I. SUBFUNCTIONS

1.1. Definitions. A real-valued function $g(x)$ of a single real variable x defined on an open interval (a, b), with $-\infty \leq a < x < b \leq +\infty$, is said to be a *convex function* of x provided $g(x)$ satisfies the inequality

$$g[tx_1 + (1-t)x_2] \leq tg(x_1) + (1-t)g(x_2) \tag{1.11}$$

for all x_1, x_2 in (a, b) and for all t on the range $0 \leq t \leq 1$. If $g(x)$ is of class C^2, it is convex if and only if $g''(x) \geq 0$ throughout the interval.

Geometrically, (1.11) indicates that no part of the graph of the curve $y = g(x)$ lies above the chord joining two points upon it within the interval (a, b).

A generalization of the foregoing characteristic geometric property of convex

Received July 6, 1952. This paper is largely a condensation of a June, 1949, doctoral dissertation, University of California, Los Angeles; I gratefully acknowledge the guidance of Professor Beckenbach.

Pacific J. Math. 3 (1953), 333-368

functions leads to the theory of subfunctions [3]. Let $\{h_{\alpha\beta}(x)\}$ be a two-parameter family of continuous functions such that for all x_1, x_2 in (a, b) and every y_1, y_2 there exists a unique member $h_{\alpha_0\beta_0}(x)$ of the family such that $h_{\alpha_0\beta_0}(x_i) = y_i$ $(i = 1, 2)$. Then a function $g(x)$ is said to be a *subfunction* of the given family on (a, b) provided[1] we have

$$(1.12) \qquad g[tx_1 + (1-t)x_2] \leq h_{\alpha_1\beta_1}[tx_1 + (1-t)x_2]$$

for all x_1, x_2 in (a, b) and for all t on the range $0 \leq t \leq 1$, and where

$$h_{\alpha_1\beta_1}(x_i) = g(x_i) \qquad (i = 1, 2).$$

Geometrically, (1.12) indicates that in the subinterval (x_1, x_2) no part of the graph of the curve $y = g(x)$ lies above the member of the parameter family joining the points $[x_1, g(x_1)]$ and $[x_2, g(x_2)]$. We note that if $g(x)$ is convex, it is a subfunction of the two-parameter family of nonvertical straight lines.

1.2. A fundamental theorem. Necessary and sufficient conditions that a function $g(x)$ be a subfunction of a certain type of two-parameter family have been obtained by Shniad [10]. The following lemma and theorem are due to him; proofs are included because of the fundamental use made of the theorem in subsequent developments.

LEMMA 1.1. *If $\phi(x)$ is a positive continuous function of x, and $\psi(x)$ is a strictly increasing continuous function of x, on $a < x < b$, then the condition that $g(x)$ be a subfunction of the family $A\phi + B\phi\psi$, where A and B are parameters of the family, is equivalent to the condition that g/ϕ be a convex function of ψ.*

Proof. The hypotheses on ϕ and ψ ensure that g/ϕ is a continuous function of ψ. To prove the existence of a unique member of the family through any two points (x_i, y_i) $(i = 1, 2)$, with the x_i distinct and in the interval, it suffices to note that

$$\begin{vmatrix} \phi(x_1) & \phi(x_1)\psi(x_1) \\ \phi(x_2) & \phi(x_2)\psi(x_2) \end{vmatrix} = \phi(x_1)\phi(x_2)[\psi(x_2) - \psi(x_1)] \neq 0.$$

Let x_1 and x_2 satisfy $a < x_1 < x_2 < b$, and let

[1] A more general definition of the subfunction property is given in [5]; in [3] it is shown that a function satisfying (1.12) necessarily is continuous on (a, b).

$$h_{\alpha_1\beta_1}(x) = A_1\phi(x) + B_1\phi(x)\psi(x)$$

with

$$h_{\alpha_1\beta_1}(x_i) = g(x_i) \qquad (i = 1, 2).$$

Then the condition

$$h_{\alpha_1\beta_1}(x) \geq g(x) \qquad \text{for } x_1 < x < x_2$$

is equivalent to the condition

$$A_1 + B_1\psi(x) \geq \frac{g(x)}{\phi(x)} \qquad \text{for } x_1 < x < x_2,$$

or that $g(x)/\phi(x)$ be a convex function of ψ on the range $\psi(a+) < \psi < \psi(b-)$.

THEOREM 1.2.[2] *Let $\phi(x)$, $\psi(x)$, and $g(x)$ be functions having the following properties on an interval $a < x < b$:*

a) *the functions ϕ, ψ, and g have continuous second derivatives,*

b) *the inequalities $\phi(x) > 0$ and $\psi'(x) > 0$ hold, and*

c) *each of the functions $\phi(x)$ and $\phi(x)\psi(x)$ is a solution of the differential equation*

$$h'' + Ph' + Qh = 0,$$

where P and Q are continuous on the interval.

Then a necessary and sufficient condition that $g(x)$ be a subfunction of the family $A\phi + B\phi\psi$ on the given interval is that

$$g'' + Pg' + Qg \geq 0$$

on the interval.

Proof. From Lemma 1.1 it follows that g is a subfunction of the family if and only if g/ϕ is a convex function of ψ. Since g/ϕ has a continuous second derivative with respect to ψ, the latter condition is equivalent to

---

[2] The conclusion of this theorem is obtained in a more general setting in [9]. However, the proof is immediate for the theorem as stated here, and this form is sufficient for our purposes.

$$\frac{1}{(\psi')^2\phi}\left\{g'' + g'\left[-\frac{\psi''}{\psi'} - \frac{2\phi'}{\phi}\right] + g\left[\frac{\psi''\phi'}{\psi'\phi} - \frac{\phi''}{\phi} + 2\left(\frac{\phi'}{\phi}\right)^2\right]\right\} \geq 0.$$

From the Wronskian relation we easily verify that ϕ and $\phi\psi$ are linearly independent solutions of the differential equation. Then the theorem follows from uniqueness properties of linearly independent solutions of this type of differential equation.

1.3. Sub-K_0 functions. The differential equation we are to consider is

$$h'' + K_0 h = 0,$$

where K_0 is a negative constant, and the interval of definition is $0 \leq x < b \leq \infty$. The two-parameter family of solutions of the equation is given by

$$\{h_{\alpha\beta}(x)\} = \{\alpha \cosh(\sqrt{-K_0}x) + \beta \sinh(\sqrt{-K_0}x)\}, \tag{1.31}$$

where α and β are the parameters. A property of this family is given in the following lemma; we omit the proof.

LEMMA 1.3. *If* A: (x_1, y_1) *and* B: (x_2, y_2) *are two points with* $x_1 \neq x_2$, *then there is one and only one curve of the family* $\{h_{\alpha\beta}(x)\}$ *passing through* A *and* B. *Thus, if* $y_1 \geq 0$ *and* $y_2 \geq 0$, *the curve* $h_{\alpha_1\beta_1}(x)$ *passing through* A *and* B *satisfies* $h_{\alpha_1\beta_1}(x) \geq 0$ *for* $x_1 \leq x \leq x_2$.

DEFINITION. A function $g(x)$ will be said to be a *sub-K_0 function* of x if it is a subfunction of the family $\{h_{\alpha\beta}(x)\}$ of (1.31) on the interval $0 \leq x < b \leq \infty$. Moreover, $g(x)$ will be said to be a *K_0-function* if the sign of equality of its subfunction relation (1.12) holds throughout the interval; and it will be a *strictly sub-K_0 function* if the strict inequality holds throughout for $0 < t < 1$.

It is convenient to introduce a second-order differential operator $\mathfrak{S}$ defined by

$$\mathfrak{S} \equiv D^2 + K_0,$$

where K_0 is a negative constant; we may write $\mathfrak{S}_x$ to indicate the variable for differentiation.

REMARK. With the choices

$$\phi(x) \equiv \cosh(\sqrt{-K_0}x) \quad \text{and} \quad \psi(x) \equiv \tanh(\sqrt{-K_0}x),$$

the family $\{\alpha\phi+\beta\phi\psi\}$ coincides with the family (1.31), and these functions ϕ and ψ satisfy the hypotheses of Theorem 1.2. Hence a function $g(x)$ of class C^2 is a sub-K_0 function (K_0-function) if and only if $\mathfrak{S}g(x) \geq 0$ ($\mathfrak{S}g(x) = 0$) on the interval.

Certain elementary properties of sub-K_0 functions are given in the following theorems. The proofs are omitted as they merely involve applying the foregoing remark to appropriate members of the family $\{h_{\alpha\beta}(x)\}$.

THEOREM 1.4. *Any linear combination of sub-K_0 functions with nonnegative coefficients is a sub-K_0 function.*

THEOREM 1.5. *Let $f(x)$ be a nonnegative sub-K_0 function, and let k be a constant ≥ 1. Then $[f(x)]^k$ is a sub-K_0 function; in fact, $[f(x)]^k$ is a sub-kK_0 function.*

THEOREM 1.6. *Let $f_i(x)$ $(i = 1, 2, \cdots, n)$ be convex functions of x which are nonnegative and monotonic nondecreasing and at least one of which is a sub-K_0 function. Then the product function $f_1 f_2 \cdots f_n$ is a sub-K_0 function.*

II. SURFACES OF NEGATIVE CURVATURE

2.1. Geodesic parameters. Let an analytic surface S be represented by geodesic parameters [7, p. 174] (u, v), so that

$$ds^2 = du^2 + \mu^2(u, v)\,dv^2 \qquad (\mu \geq 0), \tag{2.11}$$

and

$$da = \mu(u, v)\,du\,dv, \tag{2.12}$$

where the curves $v = constant$ are the geodesics, and the curves $u = constant$ are the *geodesic parallels*. The surface S is said to be given in *geodesic representation.*

Singular points of the geodesic family are points where $\mu = 0$; other points, where $\mu > 0$, are *regular points*.

The Gaussian curvature K of S exists at all regular points. If S is given in geodesic representation, the Gaussian curvature is given [7, p. 181] by the formula

$$K = -\frac{1}{\mu}\frac{\partial^2 \mu}{\partial u^2}. \tag{2.13}$$

DEFINITION. An analytic surface S will be said to be *a sub-K_0 surface* if its Gaussian curvature is bounded from above by K_0, a negative constant, at all regular points of S. Moreover, S will be said to be *a K_0-surface* if its Gaussian curvature everywhere is K_0. If S is a sub-K_0 surface which is not a K_0-surface, it will be said to be *a strictly sub-K_0 surface*.

2.2. Geodesic parallels. We have the following lemma.

LEMMA 2.1. *If an analytic surface S is given in geodesic representation, then a necessary and sufficient condition that S be a sub-K_0 surface is that the function $\mu(u, v_0)$ be a sub-K_0 function of u for each line-segment $u_1 < u < u_2$, $v = v_0$ in the (u, v)—domain of definition.*

Proof. The result follows directly from (2.13) and Lemma 1.3 by an argument analogous to that in [4, p. 286]. The proof reveals that μ is a strictly sub-K_0 function of u if and only if S is a strictly sub-K_0 surface, and that μ is a K_0-function of u if S is a K_0-surface.

Let S be a sub-K_0 surface given in geodesic representation. Then we have the following results.

THEOREM 2.2. *Let the arcs $C(u)$ $(u_1 \leq u \leq u_2)$, of length $l(u)$, be arcs of geodesic parallels between geodesics $v = v_1$ and $v = v_2$ $(v_1 < v_2)$ on S. Then the length $l(u)$ is a sub-K_0 function of u (that is, of the geodesic length $u - u_1$); $l(u)$ is a strictly sub-K_0 function if S is a strictly sub-K_0 surface, and $l(u)$ is a K_0-function if S is a K_0-surface.*

Proof. A proof may be patterned on that of the related theorem in [4, p. 287], in which we substitute the appropriate member (which is of class C^2) of the family (1.31) in a subfunction inequality in place of the convexity inequality.

THEOREM 2.3. *Let the arcs $C(u)$ $(u_1 - W < u < u_1 + W)$, of length $l(u)$, be arcs of geodesic parallels between geodesics $v = v_1$ and $v = v_2$ $(v_1 < v_2)$ on S, and let $a(w)$ denote the area of the part of S enclosed by $v = v_1$, $C(u_1 + w)$, $v = v_2$, $C(u_1 - w)$ $(0 \leq w < W)$. Then $a(w)$ is a sub-K_0 function of w; $a(w)$ is a strictly sub-K_0 function if S is a strictly sub-K_0 surface, and is a K_0-function of w if S is a K_0-surface.*

Proof. The proof is similar to that in [4, p. 288] when we consider subfunction properties instead of convexity properties.

2.3. Geodesic polar coordinates. Let the analytic surface S be represented in geodesic polar coordinates [7, p. 181] (u, v), that is, coordinates for which (2.11), (2.12), and

$$\mu(0, v) = 0, \quad \left[\frac{\partial \mu}{\partial u}\right]_{u=0} = 1 \tag{2.31}$$

are satisfied. The curve $u = u_0$ is a *geodesic circle* with center at the *pole* P of the representation and *geodesic radius* u_0.

We shall write r, θ for u, v, respectively.

Hereafter we indicate functions determined by, or calculated for, a K_0-surface by a subscript zero. Some such functions can be determined explicitly.

LEMMA 2.4. *Let S_0 be a K_0-surface, and let $l_0(r)$ and $a_0(r)$ denote the circumference and area, respectively, of the geodesic circle on S_0 with fixed center P and geodesic radius r. Then*

$$l_0(r) \equiv \frac{2\pi}{\sqrt{-K_0}} \sinh(\sqrt{-K_0}r) \tag{2.32}$$

and

$$a_0(r) \equiv \frac{2\pi}{-K_0}[\cosh(\sqrt{-K_0}r) - 1]. \tag{2.33}$$

Moreover,

$$l_0(r) > 2\pi r \qquad (r > 0 \text{ on } S),$$

and

$$a_0(r) > \pi r^2 \qquad (r > 0 \text{ on } S).$$

Proof. Since $\mathfrak{S}_u \mu = 0$, we find that the function $\mu_0(r)$ of the family (1.31) satisfying (2.31) is

$$\mu_0(r) = \frac{1}{\sqrt{-K_0}} \sinh(\sqrt{-K_0}r).$$

When we evaluate (2.11) and (2.12) for a geodesic circle using this expression for μ, we obtain the formulas of the lemma. The inequalities are easily established; cf. [4, p. 291-292].

We remark that the functions $l_0(r)$ and $a_0(r)$ will occur in formulas which refer to a sub-K_0 surface S; in such cases, (2.32) and (2.33) provide definitions of these functions on S.

III. SUBFUNCTIONS FOR GEODESIC CIRCLES

3.1. Definition. Some functions of geometrical significance involving the geodesic radius have certain properties in common. We collect these properties in the following definition.

CONDITION C. For a given sub-K_0 surface S and for a given pole P of geodesic polar coordinates on S, a function $\phi(r)$ of the geodesic radius r satisfies *Condition* C provided: $\phi(0) = 0$; for $r \geq 0$ on S, $\phi(r)$ is a continuous, nondecreasing sub-K_0 function of r; $\phi(r) \equiv 0$ if S is a K_0-surface, but otherwise $\phi(r)$ is a strictly sub-K_0 function.

If we let $K_0 = 0$, the K_0-surface becomes a developable surface, and the "sub-K_0 function of r" property becomes the usual "convex function of r" property. Thus our Condition C specializes to Condition A of [4, p. 289] when $K_0 = 0$.

It follows from the theorems of § 1.3 that sums and products of functions which satisfy Condition C also satisfy Condition C.

3.2. The length function. Hereafter we assume that $\mu(r, \theta)$ is of class C^2, which ensures the existence of the derivatives we write. We now consider a geodesic circle C_r on S with fixed center P and geodesic radius r.

LEMMA 3.1. *Let S be an analytic sub-K_0 surface, and let $l(r)$ denote the length of the circumference of C_r. Then $l(r)$ satisfies the differential relation*

$$\mathfrak{S}l(r) \equiv l''(r) + K_0 l(r) \geq 0 \qquad (r \geq 0 \text{ on } S). \tag{3.21}$$

Proof. The result is immediate since $\mathfrak{S}_r \mu(r, \theta) \geq 0$ for $r \geq 0$ on S. We note that equality holds in (3.21) if and only if S is a K_0-surface, that is, in our notation, if and only if we have $\mathfrak{S}l_0(r) = 0$, where $l_0(r)$ is given by (2.32).

LEMMA 3.2. *Let S be an analytic sub-K_0 surface, and let $a(r)$ denote the area of C_r. Then $a(r)$ satisfies the differential relation*

$$a''(r) + K_0 a(r) - 2\pi \equiv l'(r) + K_0 a(r) - 2\pi \geq 0 \qquad (r \geq 0 \text{ on } S). \tag{3.22}$$

Proof. By differentiating the area function

$$a(r) = \int_0^r \int_0^{2\pi} \mu(\rho, \theta)\, d\rho\, d\theta,$$

we get

$$a''(r) = \int_0^{2\pi} \frac{\partial \mu}{\partial r}\, d\theta \equiv l'(r).$$

Since $a(0) = 0$, and $l'(0) = 2\pi$ by (2.31), we have equality in (3.22) for $r = 0$. The derivative of the function

$$l'(r) + K_0 a(r) - 2\pi$$

is $\mathfrak{S}l(r)$, which is nonnegative by Lemma 3.1; hence the left member of (3.22) is monotonic nondecreasing, and (3.22) holds. It is readily seen that equality holds in (3.22) if and only if S is a K_0-surface.

THEOREM 3.3. *Let S be an analytic sub-K_0 surface, and let $l(r)$ denote the length of the circumference of C_r. Then the function*

$$\phi_1(r) \equiv l(r) - l_0(r) \qquad (r \geq 0 \text{ on } S),$$

satisfies Condition C.

Proof. The functions $\mu(r, \theta)$ and $\mu_0(r, \theta)$ associated with the surfaces S and S_0, respectively, both satisfy (2.31), and are such that

$$\frac{\partial^2 \mu}{\partial r^2} + K\mu = 0 \qquad (r \geq 0 \text{ on } S),$$

$$\frac{\partial^2 \mu_0}{\partial r^2} + K_0 \mu_0 = 0 \qquad (r \geq 0 \text{ on } S),$$

where $K \leq K_0$. By Sturm's oscillation theorems [8, Chap. X], it follows that

$$\mu(r, \theta) - \mu_0(r, \theta) \geq 0 \qquad (r \geq 0 \text{ on } S),$$

and

$$\frac{\partial \mu}{\partial r} - \frac{\partial \mu_0}{\partial r} \geq 0 \qquad (r \geq 0 \text{ on } S).$$

Hence we find that $\phi_1(0) = 0$, $\phi_1(r) \equiv 0$ on S if and only if S is a K_0-surface, and

$$\phi_1'(r) \equiv l'(r) - l_0'(r) = \int_0^{2\pi} \left(\frac{\partial \mu}{\partial r} - \frac{\partial \mu_0}{\partial r} \right) d\theta \geq 0 \qquad (r \geq 0 \text{ on } S).$$

Then calculation shows that

$$\mathfrak{S}\phi_1(r) = \mathfrak{S}l(r),$$

whence

$$\mathfrak{S}\phi_1(r) \geq 0 \qquad (r \geq 0 \text{ on } S),$$

by Lemma 3.1. Thus, by Theorem 1.2, $\phi_1(r)$ satisfies Condition C.

COROLLARY 3.4. *If S is an analytic sub-K_0 surface, then $l(r)$ is a monotonic increasing sub-K_0 function of r and satisfies the inequality*

$$l(r) \geq l_0(r) \qquad (r \geq 0 \text{ on } S);$$

$l(r)$ is a strictly sub-K_0 function if and only if S is not a K_0-surface.

Proof. The inequality follows from Theorem 3.3 and the identity

$$l(r) = l_0(r) + \phi_1(r) \qquad (r \geq 0 \text{ on } S).$$

REMARK. The function $\phi_1(r)$ may be modified to form a new function in the following way: replace the function $l_0(r)$ (Lemma 2.4) in $\phi_1(r)$ by its Maclaurin series expansion from which has been deleted any finite or infinite number of terms. The new function so obtained is a *nonnegative, monotonic increasing, sub-K_0 function* of r. In like manner, similar functions may be formed from subsequent ϕ functions which involve subtractive functions $l_0(r)$ and $a_0(r)$. We omit proofs.

3.3. The area function. On a surface where $K \leq K_0$, the area function $a(r)$ for a geodesic circle C_r has properties similar to those given for $l(r)$.

THEOREM 3.5. *Let S be an analytic sub-K_0 surface, and let $a(r)$ denote the area of C_r. Then the function*

$$\phi_2(r) \equiv a(r) - a_0(r) \qquad (r \geq 0 \text{ on } S)$$

satisfies Condition C.

Proof. Verification is immediate by use of Lemma 3.2 and Theorem 1.2.

COROLLARY 3.6. *If S is an analytic sub-K_0 surface, then $a(r)$ is a monotonic increasing sub-K_0 function of r and satisfies the inequality*

$$a(r) \geq a_0(r) \qquad (r \geq 0 \text{ on } S);$$

$a(r)$ is a strictly sub-K_0 function if and only if S is not a K_0-surface.

Proof. The inequality follows from Theorem 3.5 and the identity

$$a(r) \equiv a_0(r) + \phi_2(r) \qquad (r \geq 0 \text{ on } S).$$

We shall find additional theorems for the functions $a(r)$ and $\phi_2(r)$ showing certain subfunction properties of these functions when an additional assumption is made for the surface S. In the sequel we use the following lemma, which shows that certain conditions which clearly imply the sub-K_0 function property for a function also imply this property for its square root.

LEMMA 3.7. *If $g(r)$ is a nonnegative function for which $g'''(r)$ exists in the interval $\alpha \leq r < \beta$, and $g(r)$ satisfies*

$$h(\alpha) \equiv 2g(\alpha)g''(\alpha) - [g'(\alpha)]^2 + 4K_0[g(\alpha)]^2 \geq 0 \tag{3.31}$$

and

$$g'''(r) + 4K_0 g'(r) \geq 0 \qquad (\alpha \leq r < \beta),$$

then $[g(r)]^{1/2}$ is a sub-K_0 function in $\alpha \leq r < \beta$ and is a strictly sub-K_0 function there provided

$$g'''(r) + 4K_0 g'(r) > 0 \qquad (\alpha < r < \beta). \tag{3.32}$$

Proof. If we let $f(r) \equiv [g(r)]^{1/2}$, then at points where $f(r) \neq 0$ we have

$$\mathfrak{S}f(r) = \frac{1}{4}[g(r)]^{-3/2} h(r).$$

Moreover,

$$h'(r) = 2g(g''' + 4K_0 g'), \tag{3.33}$$

so that from the hypotheses we get

$$h(\alpha) \geq 0, \quad h'(r) \geq 0 \qquad (\alpha \leq r < \beta),$$

whence $h(r) \geq 0$. Thus $\mathfrak{S}f(r) \geq 0$ at points where $f(r) \neq 0$. And, since the nonnegative function $f(r)$ satisfies the subfunction inequality (1.12) for points where $f(r) = 0$, it follows that the continuous function $f(r)$ is a sub-K_0 function for $\alpha \leq r < \beta$.

With (3.31) and (3.32) the nonnegative sub-K_0 (and hence convex) function $g(r)$ can vanish at no more than one point of $\alpha \leq r < \beta$, whence, by (3.33), we have $h(r) > 0$ ($\alpha < r \not< \beta$). It follows that we have $\mathfrak{S}f(r) > 0$ except for at most one point of $\alpha < r < \beta$, so that $f(r)$ is a strictly sub-K_0 function for $\alpha \leq r < \beta$. This completes the proof of the lemma.

An additional assumption on the surface S causes certain functions immediately to satisfy (3.33) for $r \geq 0$ on S. Thus, if S satisfies $K \leq 4K_0$ for its Gaussian curvature, then a modification of the proof of Theorem 3.3 indicates that we have

$$a'''(r) + 4K_0 a'(r) \equiv l''(r) + 4K_0 l(r) \geq 0 \qquad (r \geq 0 \text{ on } S),$$

with equality holding if and only if S is a $4K_0$-surface. We now determine some functions which have certain subfunction properties in common; these properties are collected in:

CONDITION D. For a given sub-$4K_0$ surface S and for a given pole P of the geodesic polar coordinates on S, a function $\psi(r)$ of the geodesic radius r satisfies *Condition* D provided: $\psi(0) = 0$; for $r \geq 0$ on S, $\psi(r)$ is a continuous monotonic nondecreasing sub-K_0 function of r; and $\psi(r)$ is a strictly sub-K_0 function except possibly when S is a $4K_0$-surface.

THEOREM 3.8. *Let S be an analytic sub-$4K_0$ surface, and let $a(r)$ denote the area of the geodesic circle C_r. Then*

$$\psi_1(r) \equiv [a(r)]^{1/2}$$

and

$$\psi_2(r) \equiv [\phi_2(r)]^{1/2} \equiv [a(r) - a_0(r)]^{1/2}$$

satisfy Condition D, *and $\psi_1(r)$ is a K_0-function if S is a $4K_0$-surface.*

Proof. We have

$$a'(r) = l(r), \quad a''(r) = l'(r), \quad a'''(r) = l''(r);$$

hence, beside $a(0) = 0$, we have

$$a'(0) = 0, \quad a'''(r) \geq 0 \qquad (r \geq 0 \text{ on } S),$$

with

$$a'''(r) + 4K_0 a'(r) > 0$$

for $r > 0$ on S unless $K \equiv 4K_0$. Then for $r \geq 0$ on S, $a(r)$ satisfies the hypotheses on $g(r)$ of Lemma 3.7, so that $\psi_1(r)$ satisfies Condition D for $r \geq 0$ on S, and is a strictly sub-K_0 function if S is not a $4K_0$-surface. If S is a $4K_0$-surface, then

$$[a(r)]^{1/2} \equiv \left[\frac{2\pi}{-4K_0} \{\cosh(\sqrt{-4K_0}r) - 1\}\right]^{1/2} \equiv \left(\frac{\pi}{-K_0}\right)^{1/2} \sinh(\sqrt{-K_0}r),$$

and thus it is a K_0-function.

The proof for $\psi_2(r)$ is similar in method and is omitted.

We can find other functions which satisfy Condition D. Let $l_1(r)$ and $a_1(r)$ denote the length of circumference and area, respectively, of the geodesic circle C_r on a $4K_0$-surface. Formulas for $l_1(r)$ and $a_1(r)$ can be written (see Lemma 2.4), and these expressions serve to define $l_1(r)$ and $a_1(r)$ for a surface S having arbitrary curvature. If S is a sub-$4K_0$ surface, then, by methods analogous to those of Lemma 3.1 and Lemma 3.2, we find the relations

$$l'(r) \geq 2\pi - 4K_0 a(r) \qquad (r \geq 0 \text{ on } S),$$

and

$$l''(r) + 4K_0 l(r) \geq 0 \qquad (r \geq 0 \text{ on } S),$$

with the equality sign holding for $r > 0$ on S if and only if S is a $4K_0$-surface; that is,

$$l_1'(r) = 2\pi - 4K_0 a_1(r), \tag{3.34}$$

and

$$l_1''(r) + 4K_0 l_1(r) = 0.$$

THEOREM 3.9. *Let* S *be an analytic sub-*$4K_0$ *surface, and let* $l(r)$ *and* $a(r)$ *denote the circumference and area function, respectively, of* C_r *on* S. *Then the functions*

$$\psi_3(r) \equiv l(r) - l_1(r),$$

$$\psi_4(r) \equiv a(r) - a_1(r),$$

and

$$\psi_5(r) \equiv [a(r) - a_0(r)]^{1/2}$$

satisfy Condition D.

Proof. The method is that used in earlier theorems wherein now we apply the four relations which immediately precede Theorem 3.9.

REMARK. It was indicated earlier that our Condition C reduces to Condition A of [4, p. 289] if $K_0 = 0$. Now if $K_0 = 0$, the assumption that S satisfies $K \leq 4K_0$ imposes no new requirement upon the surface. In fact, our Condition D becomes Condition A if $K_0 = 0$ and if the function $\psi(r)$ is identically zero when the surface is developable.

The role played by the condition $K \leq 4K_0$, when $K_0 \neq 0$ for "square root" functions is indicated in the following theorem.

THEOREM 3.10. *Let* S *be an analytic sub-*K_0 *surface, and let* $a(r)$ *denote the area of* C_r *on* S. *Then in order that the function*

$$\psi_1(r) \equiv [a(r)]^{1/2}$$

be a sub-K_0 *function of* r *for every possible pole* P, *it is necessary and sufficient that* S *be a sub-*$4K_0$ *surface.*

Proof. The sufficiency has been established in Theorem 3.8.

Now let P_1 be a point of S where $K_1 > 4K_0$, and let P_1 be the pole of a geodesic polar coordinate system. Since S is analytic, there exists a neighborhood of P_1 in which $K > 4K_0$, and hence a value $r_1 > 0$ such that the geodesic circle of radius r_1 lies entirely within this neighborhood. In this coordinate system we have

$$\frac{\partial^2 \mu}{\partial r^2} + 4K_0 \mu < 0 \qquad (0 < r \leq r_1 \text{ on } S),$$

and then it easily follows that

$$l''(r) + 4K_0 l(r) < 0 \qquad (0 < r \leq r_1 \text{ on } S).$$

By calculation we get that

$$\mathfrak{S}\psi_1(r) = \frac{1}{4} a^{-3/2} [2al' - l^2 + 4K_0 a^2] \equiv \frac{1}{4} a^{-3/2} h(r),$$

where $h(r)$ is the bracketed expression. Then we have that $h(0) = 0$, and

$$h'(r) = 2a(l'' + 4K_0 l) < 0 \qquad (0 < r \leq r_1 \text{ on } S);$$

hence $h(r) < 0$ for $0 < r \leq r_1$, and thus also

$$\mathfrak{S}\psi_1(r) < 0 \qquad (0 < r \leq r_1 \text{ on } S).$$

Then by Theorem 1.2, $\psi_1(r)$, when evaluated in a coordinate system with such a pole, cannot be a sub-K_0 function.

IV. The Isoperimetric Inequality and Related Functions

4.1. The isoperimetric inequality. Let L and A denote the perimeter and area, respectively, of a simply connected region bounded by an analytic curve on a surface of nonpositive curvature. The isoperimetric inequality

$$\theta \equiv \frac{L^2}{4\pi} - A \geq 0 \tag{4.11}$$

holds for such a region. In fact, the following theorem [6, p. 670-672] has been established:

For an analytic surface S, a necessary and sufficient condition that (4.11) *hold for all simply connected regions bounded by analytic curves on S is that $K \leq 0$ on S. Further, if $K \leq 0$ but $K \not\equiv 0$ on S, then the strict sign of inequality holds in* (4.11); *while if $K \equiv 0$ on S, then the sign of equality holds in* (4.11) *only for geodesic circles on S.*

We shall study the function θ of (4.11) and some modifications of it for sub-K_0 function properties when S is assumed to be a sub-K_0 surface and the region is that determined by a geodesic circle. A well-known generalization of the function θ for geodesic circles on surfaces of constant negative curvature K_0 is the function

$$\phi_3(r) \equiv \frac{l^2(r)}{4\pi} + \frac{K_0 a^2(r)}{4\pi} - a(r), \tag{4.12}$$

which we shall call the *isoperimetric function.*

THEOREM 4.1. *Let S be an analytic sub-K_0 surface, and let $l(r)$ denote the length of the circumference, and $a(r)$ the area, of the geodesic circle C_r on S. Then the isoperimetric function $\phi_3(r)$ satisfies Condition* C.

Proof. Squaring the inequality (3.22) and using (3.21), we obtain

$$4\pi \mathfrak{S}\phi_3(r) \geq [l'^2 + K_0 l^2 - 4\pi^2] \equiv h(r),$$

where $h(r)$ is the function in brackets. Then we see that $h(0) = 0$, and that $h'(r) \geq 0$ for $r \geq 0$ by (3.21); hence $\mathfrak{S}\phi_3(r) \geq 0$ for $r \geq 0$ on S, and thus $\phi_3(r)$ is a sub-K_0 function by Theorem 1.2. The other requirements of Condition C are easily found to be satisfied by $\phi_3(r)$.

COROLLARY 4.2. *If S is an analytic sub-K_0 surface, and $l(r)$ and $a(r)$ as in the theorem, then the function*

$$\theta_1(r) \equiv \frac{l^2(r)}{4\pi} - a(r)$$

is a continuous monotonic nondecreasing sub-K_0 function of r.

Proof. It follows from the proof of Lemma 3.2 that $a(r)$ is a continuous monotonic nondecreasing sub-K_0 function of r, and then that $a^2(r)$ also has these properties by Theorem 1.5. Then, using the positive coefficient $-K_0/4\pi$, we may apply Theorem 1.4 and get

$$\theta_1(r) \equiv \phi_3(r) - \frac{K_0 a^2(r)}{4\pi},$$

so that $\theta_1(r)$ has the properties stated in Corollary 4.2.

COROLLARY 4.3. *If S is an analytic sub-K_0 surface, then*

$$(4.13)\qquad \frac{l^2(r)}{4\pi} - a(r) > \frac{l^2(r)}{4\pi} + \frac{K_0 a^2(r)}{4\pi} - a(r) \geq 0 \qquad (r > 0 \text{ on } S),$$

where the sign of equality holds for $r > 0$ on S if and only if S is a K_0-surface.

Proof. This corollary is an immediate consequence of Theorem 4.1 and Corollary 4.2.

4.2. Modifications of the isoperimetric function. We shall consider modifications of the isoperimetric function, $\phi_3(r)$, which are produced by adding certain functions to it and/or by replacing $l(r)$ by $l_0(r)$ or $2\pi r$ and $a(r)$ by $a_0(r)$ or zero. For example, the function $\theta_1(r)$ may be considered a modification of $\phi_3(r)$ formed by replacing the $a^2(r)$ function in $\phi_3(r)$ by zero.

THEOREM 4.4. *Let S be an analytic sub-K_0 surface, and let $l(r)$, $l_0(r)$, and $a(r)$, $a_0(r)$ denote length and area functions associated with the geodesic circle C_r. Then the functions*

$$\phi_4(r) \equiv \frac{l_0(r)\,l(r)}{4\pi} + \frac{K_0\, a_0^2(r)}{4\pi} - a(r)$$

and

$$\phi_5(r) \equiv \frac{l_0(r)l(r)}{4\pi} + \frac{K_0 a_0(r) a(r)}{4\pi} - \frac{a_0(r)}{2} - \frac{a(r)}{2}$$

satisfy Condition C.

Proof. We establish the result that

$$(4.21)\qquad l_0\, l' - l\, l_0' \equiv l_0\, l' + K_0 a_0\, l - 2\pi l \geq 0 \qquad (r \geq 0 \text{ on } S).$$

The function on the left is zero when $r = 0$, and its derivative is the nonnegative (by Lemma 3.1) function $l_0(r)\,\mathfrak{S}\, l(r)$; hence (4.21) holds.

Now $\phi_4(0) = 0$, and $\phi_4'(r) \geq 0$ by (4.21) and Theorem 3.3; thus $\phi_4(r)$ is monotonic nondecreasing. The calculation for $\mathfrak{S}\,\phi_4(r)$ may be arranged so that

$$4\pi\, \mathfrak{S}\,\phi_4(r) \equiv [\,l_0 \mathfrak{S}\, l - K_0 l(l - l_0) - 2K_0 a_0(l' - l_0') - 4\pi K_0(a - a_0) + K_0(l_0^2 + K_0\, a_0^2 - 4\pi a_0)\,].$$

Then we have $\mathfrak{S}\phi_4(r) \geq 0$ for $r \geq 0$ on S, since each parenthesis above is nonnegative by previous results—the last one, in particular, being identically zero according to Corollary 4.3. Thus $\phi_4(r)$ is a sub-K_0 function by Theorem 1.2. Finally, $\phi_4(r)$ satisfies Condition C since the signs of equality hold in the relations above if and only if S is a K_0-surface, and obviously $\phi_4(r) \equiv 0$ if S is a K_0-surface.

For $\phi_5(r)$, we find that $\phi_5(0) = 0$ and that $\phi_5'(r) \geq 0$ by Lemma 3.2. We may arrange the calculation so that

$$4\pi\mathfrak{S}\phi_5(r) \equiv [l_0\mathfrak{S}l + (l_0'l' + K_0 l_0 l - 2\pi l_0' - 2\pi K_0 a_0)]$$

$$\geq l_0'l' + K_0 l_0 l - 4\pi^2 \equiv h(r),$$

where $h(r)$ is the function after the inequality sign. Clearly $h(0) = 0$, and we find that

$$h'(r) \equiv l_0'\mathfrak{S}l \geq 0.$$

Hence, $h(r) \geq 0$ for $r \geq 0$ on S, and then Theorem 1.2 ensures that $\phi_5(r)$ is a sub-K_0 function. The other conditions to complete the proof are easily verified.

Theorem 4.4 then admits a corollary which is analogous to the isoperimetric inequality for the functions $\phi_4(r)$ and $\phi_5(r)$; we omit its statement, but we remark that the inequality for $\phi_5(r)$ is sharper than the isoperimetric inequality (4.13) in that it presents a better estimate (greater lower bound) for $l(r)$.

The next theorem presents another function determined by the modification process.

THEOREM 4.5. *For a surface and functions as in Theorem* 4.4, *the function*

$$\phi_6(r) \equiv \frac{rl(r)}{2} - a(r) - \frac{rl_0(r)}{2} + a_0(r) \equiv \frac{r}{2}[l(r) - l_0(r)] - [a(r) - a_0(r)]$$

satisfies Condition C.

We omit the computations and also the corollary stating the inequality satisfied by $\phi_6(r)$.

It may be noted that with $\phi_6(r)$ satisfying Condition C it readily follows that $\phi_4(r)$ does. For if the function

$$\frac{1}{4\pi}\,(l_0 - 2\pi r)\,(l - l_0),$$

which satisfies Condition C (in part by Theorem 1.6), is added to $\phi_6(r)$, we obtain $\phi_4(r)$, which then satisfies Condition C (in part by Theorem 1.4).

Theorem 4.5 suggests a consideration of the substitution of $l(r) - l_0(r)$ and $a(r) - a_0(r)$ for the functions $l(r)$ and $a(r)$. When this substitution is made in the isoperimetric function, we find that the new function does not satisfy our conditions. Nevertheless, in the next theorem we have a result of this procedure.

THEOREM 4.6. *For a surface and functions as in Theorem* 4.4, *the function*

$$\phi_7(r) \equiv \frac{1}{4\pi}\,[l(r) - l_0(r)]^2 + \frac{K_0}{4\pi}\,[a(r) - a_0(r)]^2$$

satisfies Condition C.

Proof. We find that $\phi_7(0) = 0$ and that $\phi_7'(r) \geq 0$ by Lemma 3.2 and Theorem 3.3; thus $\phi_7(r)$ is monotonic nondecreasing. By computation we find that

$$4\pi\mathfrak{S}\phi_7(r) \geq [2(l' - l_0')(l' + K_0 a - 2\pi) + K_0(l - l_0)^2 + K_0^2(a - a_0)^2] \equiv h(r),$$

where $h(r)$ is the bracketed expression. We see that $h(0) = 0$, and that its derivative satisfies

$$h'(r) \geq 2[(l'' - l_0'')(l' + K_0 a - 2\pi) + K_0(l - l_0)(l' - l_0') + K_0^2(a - a_0)(l - l_0)]$$

$$= 2(l'' + K_0 l)(l' + K_0 a - 2\pi) \geq 0$$

by Lemmas 3.1 and 3.2. Hence $h(r) \geq 0$ for $r \geq 0$ on S, with equality holding if and only if S is a K_0-surface. It follows that $\phi_7(r)$ satisfies Condition C since obviously $\phi_7(r) \equiv 0$ if S is a K_0-surface.

Theorem 4.6 admits refinements of the inequalities which appear in Corollary 4.3 and Corollary 3.4.

COROLLARY 4.7. *Let an analytic sub-K_0 surface S be referred to a geodesic polar coordinate system with pole P. Then, for geodesic circles, the isoperimetric function $\phi_3(r)$ and the functions $\phi_5(r)$ and $\phi_7(r)$ satisfy the inequalities*

$$\phi_3(r) \geq 2\phi_5(r) \geq 0 \tag{4.22}$$

and

$$\phi_3(r) \geq \phi_7(r) \geq 0,$$

where the signs of equality hold for $r > 0$ on S if and only if S is a K_0-surface, in which case all functions are identically zero.

Proof. It is easily seen that

$$\phi_7(r) \equiv \phi_3(r) - 2\phi_5(r),$$

and the corollary then follows from Theorems 4.6 and 4.3.

COROLLARY 4.8. *Let an analytic sub-K_0 surface S be referred to a geodesic polar coordinate system with pole P. Then the length of the circumference of a geodesic circle of radius r satisfies the inequality*

$$l(r) \geq l_0(r) + \sqrt{-K_0}\,[a(r) - a_0(r)], \tag{4.23}$$

where the sign of equality holds for $r > 0$ on S if and only if S is a K_0-surface.

Proof. Since $K_0 < 0$, $4\pi\phi_7(r)$ has real factors. The factor $\phi_8(r)$, where $\phi_8(r) \equiv l - l_0 + \sqrt{-K_0}\,(a - a_0)$, satisfies $\phi_8(r) \geq 0$ by Corollaries 3.4 and 3.6; hence so also does the other factor by Corollary 4.7. This other factor yields (4.23).

Less precise relations may be obtained from the isoperimetric function by using the theorems of § 1.3.

THEOREM 4.9. *Let S be an analytic sub-K_0 surface with length and area functions relating to geodesic circles on S as previously defined. Then the functions*

$$\phi_9(r) \equiv \frac{1}{4\pi}\,[l^2(r) - l_0^2(r)] - [a(r) - a_0(r)],$$

$$\phi_{10}(r) \equiv \frac{1}{4\pi}[l^2(r) - l_0^2(r)] + \frac{K_0}{4\pi}\,[a^2(r) - a_0^2(r)],$$

and

$$\phi_{11}(r) \equiv l^2(r) - \frac{a(r)}{a_0(r)}\, l_0^2(r) \equiv l^2(r) - 4\pi a(r)\cosh\left(\frac{\sqrt{-K_0}\, r}{2}\right)$$

satisfy Condition C.

Proof. We refer to § 1.3 and merely indicate the verification of the desired subfunction property of these functions. Thus, $\phi_9(r)$ results from adding the function $-(K_0/4\pi)[a^2(r) - a_0^2(r)]$, which satisfies Condition C, to the isoperimetric function $\phi_3(r)$. The function $\phi_{10}(r)$ is obtained by adding the function $\phi_2(r) \equiv a(r) - a_0(r)$ to $\phi_3(r)$. And the function $\phi_{11}(r)$ is obtained by adding $-K_0 a(r)[a(r) - a_0(r)]$, which satisfies Condition C, to $4\pi\phi_3(r)$.

4.3. Another kind of modification. The properties of the isoperimetric function and its modifications which we have developed now enable us to introduce new functions which satisfy our conditions. These new functions are produced by replacing each term of an expression by its square root.

THEOREM 4.10. *Let S be an analytic sub-$4K_0$ surface with length and area functions relating to a geodesic circle on S as previously defined. Then the functions*

$$\psi_6(r) \equiv \sqrt{a(r)} - \sqrt{a_0(r)}$$

and

$$\psi_7(r) \equiv \sqrt{a(r)} - \sqrt{a_1(r)}$$

satisfy Condition D.

Proof. We have $\psi_7(0) = 0$, and

$$\psi_7'(r) = \frac{1}{2}\left[\frac{l}{\sqrt{a}} - \frac{l_1}{\sqrt{a_1}}\right] \geq 0$$

for $r \geq 0$ on S by the properties of $\phi_{11}(r)$ of Theorem 4.9, since now $l_1(r)$ and $a_1(r)$ behave analogously to $l_0(r)$ and $a_0(r)$ of that theorem. Hence $\psi_7(r)$ is a monotonic nondecreasing function of r. Then using (3.34) and the isoperimetric identity satisfied by $l_1(r)$, we get

$$\mathfrak{S}\psi_7(r) = \frac{a^{-3/2}}{4}\,[2al' - l^2 + 4K_0 a^2] \geq 0$$

for $r \geq 0$ on S, since the function in brackets is identical with that which would occur for the function $\psi_1(r)$ of Theorem 3.8. Thus $\psi_7(r)$ satisfies Condition D.

The proof for $\psi_6(r)$ is similar to this for $\psi_7(r)$.

The next theorem presents a modification of the function $\phi_{11}(r)$ of Theorem 4.9.

THEOREM 4.11. *Let S be an analytic sub-K_0 surface with length and area functions relating to geodesic circles on S as previously defined. Then the function*

$$\phi_{12}(r) \equiv l(r)\sqrt{a_0(r)} - l_0(r)\sqrt{a(r)}$$

satisfies Condition C.

Proof. We first establish the inequality

$$2a(r)l'(r) - l^2(r) + K_0 a^2(r) \geq 0 \qquad (r \geq 0 \text{ on } S), \tag{4.31}$$

where the sign of equality holds for $r > 0$ on S if and only if S is a K_0-surface. The result is immediate, since the function on the left in (4.31) is zero at $r = 0$, and its derivative is nonnegative for $r \geq 0$.

Clearly $\phi_{12}(0) = 0$, and $\phi_{12}(r) \geq 0$ for $r > 0$ on S since $\phi_{11}(r)$ satisfies Condition C. Then, by substituting for $l'(r)$ and $l_0'(r)$ from (4.31), we find that

$$\phi_{12}'(r) \geq \frac{1}{2}\left[\left(\frac{l}{a} + \frac{l_0}{a_0}\right)\left(l\sqrt{a_0} - l_0\sqrt{a}\right) - K_0\sqrt{a a_0}\;(\sqrt{a} - \sqrt{a_0})\right] \geq 0.$$

Thus $\phi_{12}(r)$ is monotonic nondecreasing. Then using (3.21) we find that

$$2\mathfrak{S}\phi_{12}(r) \geq \left(\frac{l' l_0}{\sqrt{a_0}} - \frac{l l_0'}{\sqrt{a}}\right) + \left(\frac{l_0' l}{\sqrt{a_0}} - \frac{l_0^2 l}{2a_0\sqrt{a_0}} - \frac{l_0' l}{\sqrt{a}}\right)$$
$$+\left(\frac{l' l_0}{\sqrt{a_0}} - \frac{l' l_0}{\sqrt{a}} + \frac{l^2 l_0}{2a\sqrt{a}}\right).$$

Now using (4.31) in the last two parentheses, we get

$$4\mathfrak{S}\phi_{12}(r) \geq 2\left(\frac{l'l_0}{\sqrt{a_0}} - \frac{ll_0'}{\sqrt{a}}\right) + \frac{ll_0}{aa_0}(l\sqrt{a_0} - l_0\sqrt{a}) - K_0 l_0\sqrt{\frac{a}{a_0}}(\sqrt{a} - \sqrt{a_0})$$

$$\geq \frac{2}{\sqrt{a}}\left(l'l_0\sqrt{\frac{a}{a_0}} - ll_0'\right) \geq \frac{2}{\sqrt{a}}(l'l_0 - ll_0').$$

Hence, by (4.21), it follows that $\mathfrak{S}\phi_{12}(r) \geq 0$ for $r \geq 0$ on S. Thus, on citing Theorem 1.2 and the obvious fact that $\phi_{12}(r) \equiv 0$ if S is a K_0-surface, we have shown that $\phi_{12}(r)$ satisfies Condition C.

V. Extensions and Generalizations

5.1. Geodesic circular sectors. The generalization from a basic configuration of geodesic circles to one of geodesic circular sectors is indicated in [4, p. 296], and its relations apply immediately to this study.

We state some representative results.

THEOREM 5.1. *Let S be an analytic sub-K_0 surface, and let $l(r;\theta_1,\theta_2)$ and $a(r;\theta_1,\theta_2)$ denote respectively the length of the bounding arc and the area of the geodesic circular sector on S with fixed pole P, fixed angle from θ_1 to θ_2, $\theta_1 < \theta_2$, and geodesic radius r. Then the functions*

$$\phi_{13}(r) \equiv l(r;\theta_1,\theta_2) - l_0(r;\theta_1,\theta_2),$$

$$\phi_{14}(r) \equiv a(r;\theta_1,\theta_2) - a_0(r;\theta_1,\theta_2),$$

and

$$\phi_{15}(r) \equiv \frac{l^2(r;\theta_1,\theta_2)}{2(\theta_2-\theta_1)} + \frac{K_0 a^2(r;\theta_1,\theta_2)}{2(\theta_2-\theta_1)} - a(r;\theta_1,\theta_2),$$

$$(r \geq 0,\ \theta_1 \leq \theta \leq \theta_2 \ \text{on}\ S),$$

satisfy Condition C.

The proof for each function is similar to the proof of the analogous result for the corresponding function for geodesic circles, and will not be given here.

Other functions which satisfy Condition C or Condition D for geodesic circular sectors (the analogues of those for geodesic circles) obviously could be written. It is clear that, as corollaries, we then obtain certain inequality relations between the length and area functions for a suitably restricted surface.

5.2. Regular super-K_0 surfaces. The preceding results concerning sub-K_0 surfaces hold in the large and are unaffected by singular points. We now describe somewhat analogous results for surfaces whose Gaussian curvature satisfies $K \geq K_0$; such surfaces will be called super-K_0 surfaces. We still assume $K_0 < 0$, although some of the results hold, in the small, for K_0 any constant. In general, our results will hold only on parts of S where there are no singular points of the surface, or of the family of geodesics, other than at the pole of geodesic polar coordinates; and some of the results hold only in the small even where there are no singular points.

A function $f(x)$ is said to be a *super-K_0 function* provided $-f(x)$ is a sub-K_0 function.

A surface S given in geodesic coordinates, or in geodesic polar coordinates, will be said to be *regular* provided there are no singular points on S except, in the case of geodesic polar coordinates, at the pole P.

Lemma 2.1 holds if we add the restriction that S is regular, and replace "sub-K_0" by "super-K_0." Theorems 2.2 and 2.3 hold with the same alterations, and the inequality relations given by (3.21), (3.22), and (4.21) hold with the inequality signs reversed.

THEOREM 5.2. *Let S be a regular analytic super-K_0 surface, and let $l(r)$ and $a(r)$ denote the length and area functions for a geodesic circle C_r. Then the functions $-\phi_j(r)$ $(j = 1, 2, 4, 5, 6, 8)$ satisfy Condition* C.

Proof. The theorem follows in routine fashion by an examination of earlier calculations for these functions in relation to (3.21), (3.22), and (4.21) with the inequality signs reversed.

Now consider the isoperimetric function $\phi_3(r)$. We compute $\phi_3'(r)$, and find that *$\phi_3(r)$ is monotonic nonincreasing on any regular super-K_0 surface S, and is monotonic decreasing if S is not a K_0-surface.* Actually, since $l'(0) = 2\pi$, it follows from a consideration of $\mathfrak{S}\phi_3(r)$ that there is an $r_0 = r_0(S, P)$ such that $\phi_3(r)$ is a super-K_0 function for $0 \leq r \leq r_0$.

From the properties of the functions $\phi_j(r)$ we obtain results for $l(r)$ and $a(r)$. We have

$$l(r) = l_0(r) + \phi_1(r),$$

$$l'(r) = l_0'(r) + \phi_1'(r),$$

and

$$\mathfrak{S}\, l(r) = \mathfrak{S}\, \phi_1(r).$$

Since the functions $-\phi_j(r)$ satisfy Condition C on regular super-K_0 surfaces, we have

$$\phi_1(r) \leq 0, \quad \mathfrak{S}\, \phi_1(r) \leq 0 \qquad (r \geq 0 \text{ on } S).$$

It follows that *on regular analytic super-K_0 surfaces the function $l(r)$ is a super-K_0 function and satisfies $l(r) \leq l_0(r)$; $l(r)$ is a strictly super-K_0 function and satisfies the strict inequality for $r > 0$ on S if S is not a K_0-surface.* Also, on these surfaces we have $\phi_1'(0) = 0$, so that, since $\phi_1(r)$ is a super-K_0 function, *for a given regular analytic super-K_0 surface and for a given pole P on S, either $l(r)$ is monotonic increasing on S or there is an $r_0 = r_0(S, P) > 0$ such that $l(r)$ is monotonic increasing for $0 \leq r \leq r_0$ and monotonic decreasing for $r \geq r_0$ on S.*

Again, we have

$$a(r) = a_0(r) + \phi_2(r),$$

$$a'(r) = l_0(r) + \phi_1(r) = l(r),$$

and

$$\mathfrak{S}\, a(r) = 2\pi + \mathfrak{S}\, \phi_2(r).$$

On regular analytic super-K_0 surfaces we have

$$\phi_2(r) \leq 0, \quad \phi_2''(0) = 0, \quad \mathfrak{S}\, \phi_2(r) \leq 0 \qquad (r \geq \text{ on } S).$$

Hence *on regular super-K_0 surfaces, $a(r)$ satisfies $a(r) \leq a_0(r)$; the strict inequality holds for $r > 0$ on S if S is not a K_0-surface. Further, for a given regular analytic super-K_0 surface, and for a given pole P on S, either $a(r)$ is a strictly sub-K_0 function, or there is an $r_0 = r_0(S, P) > 0$ such that $a(r)$ is a strictly sub-K_0 function for $0 \leq r \leq r_0$ and a strictly super-K_0 function for $r \geq r_0$ on S.* The interval $0 \leq r \leq r_0$ on which $a(r)$ is a sub-K_0 function coincides with the interval on which $l(r)$ is increasing.

From the properties of $\phi_4(r)$, $\phi_5(r)$ and $\phi_6(r)$ described in Theorem 5.2, we deduce some inequalities of interest. Thus, *on regular analytic super-K_0 surfaces we have the inequalities*

$$a(r) \geq \frac{1}{4\pi} [l_0(r)l(r) + K_0 a_0^2(r)],$$

$$a(r) \geq \frac{1}{l_0'(r)} [l_0(r)l(r) - 2\pi a_0(r)],$$

and

$$a(r) \geq a_0(r) - \frac{r}{2} [l_0(r) - l(r)],$$

associated with $\phi_4(r)$, $\phi_5(r)$, and $\phi_6(r)$ respectively, with the signs of equality holding for $r > 0$ on S if and only if S is a K_0-surface.

When the proof in Theorem 4.6 is examined in light of the new basic inequalities for regular analytic super-K_0 surfaces, we find that $\phi_7(r)$ *remains* a monotonic, nondecreasing sub-K_0 function. The function $4\pi\phi_7(r)$ is factorable in such a way that $\phi_8(r)$ is a factor; then, by Theorem 5.2, the other factor satisfies the inequality

$$l(r) - l_0(r) - \sqrt{-K_0}\, [a(r) - a_0(r)] \leq 0.$$

Hence, using this last relation, *on regular analytic super-K_0 surfaces we have the inequalities*

$$l(r) \leq l_0(r) - \sqrt{-K_0}\, [a_0(r) - a(r)] \leq l_0(r),$$

and

$$a_0(r) \geq a(r) \geq a_0(r) - \frac{1}{\sqrt{-K_0}} [l_0(r) - l(r)],$$

with the signs of equality holding for $r > 0$ on S if and only if S is a K_0-surface.

The ϕ functions related to geodesic circular sectors (see Theorem 5.1) have analogous properties on regular super-K_0 surfaces.

5.3. Surface characterization. Heretofore we have assumed S to be either a sub-K_0 surface or a super-K_0 surface. In certain instances we have obtained, in the two cases, conclusions which are distinct except for the dividing class of K_0-surfaces. Thus by logical exclusion we obtain several characterizations of the indicated classes of surfaces.

For example, *a regular analytic surface S is a super-K_0 surface, but not a K_0-surface, if and only if for each pole P on S we have*

$$l(r) < l_0(r) \tag{5.31}$$

for all $r > 0$ on S.

Proof. In §5.2 we have shown that the condition $K \geq K_0$, $K \not\equiv K_0$, on S implies (5.31). Conversely, if we should have $K_1 < K_0$ at some P_1 on S, then we would have $K < K_0$ throughout some neighborhood of P_1, and therefore, in the neighborhood, we would have $l(r) > l_0(r)$; also, if we should have $K \equiv K_0$ on S, then we would have $l(r) \equiv l_0(r)$; hence (5.31) implies $K \geq K_0$, $K \not\equiv K_0$ on S.

In the same way we could establish similar results for each function in the following theorem.

THEOREM 5.3. *The regular analytic surface S is* i) *a sub-K_0 surface, but not a K_0-surface,* ii) *a super-K_0 surface, but not a K_0-surface, or* iii) *a K_0-surface, if and only if we have*

$$\text{i) } \phi_j(r) > 0, \qquad \text{ii) } \phi_j(r) < 0, \qquad \textit{or} \quad \text{iii) } \phi_j(r) \equiv 0, \tag{5.32}$$

$$(j = 1, 2, \cdots, 6, 8, 9, \cdots, 15),$$

respectively, for all poles P and all $r > 0$ on S.

By Theorem 1.2, it is evident that we might replace (5.32) with the differential conditions

$$\text{i) } \mathfrak{S}\phi_j(r) > 0, \qquad \text{ii) } \mathfrak{S}\phi_j(r) < 0, \qquad \text{or} \quad \text{iii) } \mathfrak{S}\phi_j(r) \equiv 0$$

$$(j = 1, 2, 4, 5, 6, 8, 13, 14).$$

5.4. Geodesically similar curves. The preceding theory may be applied to more general configurations than geodesic circles and sectors. Thus we may study comparison functions which involve length and area functions relating to a class of curves upon an arbitrary surface S as compared to the corresponding curves upon a K_0-surface or in the plane.

It is evident that r has heretofore played a dual role: it has served as the parameter for the family of geodesic circles (sectors) on S with centers at the pole P, and it also has been a variable of the geodesic polar coordinate system. We now rephrase the previous conditions in terms of the parameter of the family of curves to be considered.

CONDITION A(k). For a given surface S of nonpositive Gaussian curvature, and for a given one-parameter family of curves $C(k)$, a function $\lambda(k)$ of the parameter k satisfies *Condition* A(k) provided: $\lambda(0) = 0$; for $k \geq 0$, $\lambda(k)$ is a continuous monotonic nondecreasing convex function of k; $\lambda(k) \equiv 0$ if S is a developable surface, but otherwise is monotonic increasing and strictly convex.

CONDITION C(k). For a given sub-K_0 surface S, and for a given one-parameter family of curves $C(k)$, a function $\tau(k)$ of the parameter k satisfies *Condition* C(k) provided: $\tau(0) = 0$; for $k \geq 0$, $\tau(k)$ is a continuous monotonic nondecreasing sub-K_0 function of k; $\tau(k) \equiv 0$ if S is a K_0-surface, but otherwise $\tau(k)$ is a strictly sub-K_0 function of k.

On a surface S referred to geodesic polar coordinates (r, θ) with a given pole P, we first consider the family of curves $C(k)$ of parameter k given by

$$r = kf(\theta), \quad k \geq 0, \tag{5.41}$$

where $f(\theta)$ admits a continuous derivative and $f(\theta) \geq 1$. We remark that the condition $f(\theta) \geq 1$ is merely a normalization; for, if $f(\theta_0) \leq 1$ and $f(\theta) \neq 0$ in a closed interval, $\alpha \leq \theta \leq \beta$, then $f(\theta)$ is bounded away from zero in (α, β), say $f(\theta) \geq m > 0$ in (α, β). Then a new parameter k_1 may be introduced by setting $k = mk_1$, so that

$$r = k_1 f_1(\theta) = k_1[mf(\theta)],$$

and this representation satisfies our requirements. It may be noted that $f(\theta) \equiv 1$ presents the case of geodesic circles (sectors). The curves $C(k)$ of the family given by (5.41) are said to be *similarly situated* or *homothetic*, and we shall call them geodesically similar.

THEOREM 5.4. *Let S be an analytic surface of nonpositive Gaussian curvature referred to geodesic polar coordinates with given pole P_0. Let $l_P(k; \alpha, \beta)$ and $l_S(k; \alpha, \beta)$ denote the lengths of the curve of the family $C(k)$ of (5.41) from $\theta = \alpha$ to $\theta = \beta$, $(\alpha < \beta)$, for the parameter value k in the plane and on the surface S, respectively, and let $a_P(k; \alpha, \beta)$ and $a_S(k; \alpha, \beta)$ denote the* *areas of the sectors formed by the curve of the family $C(k)$, $\theta = \alpha$, and $\theta = \beta$ ($\alpha < \beta$), for the parameter value k in the plane and on the surface S respectively. Then the functions*

$$\lambda_1(k;\ \alpha,\ \beta) \equiv l_S(k;\ \alpha,\ \beta) - l_P(k;\ \alpha,\ \beta)$$

and

$$\lambda_2(k;\ \alpha,\ \beta) \equiv a_S(k;\ \alpha,\ \beta) - a_P(k;\ \alpha,\ \beta)$$

satisfy Condition $A(k)$.

Proof (outlined). For θ_0 fixed and $\alpha < \theta_0 < \beta$, let

$$\lambda_0(k,\ \theta_0) \equiv \left[\mu^2 + \left(\frac{dr}{d\theta}\right)^2\right]^{1/2} - \left[r^2 + \left(\frac{dr}{d\theta}\right)^2\right]^{1/2},$$

where $r = r(\theta)$ is given by (5.41). We find that $\lambda_0(0,\ \theta_0) = 0$, and that $\partial\lambda_0/\partial k$ and $\partial^2\lambda_0/\partial k^2$ are nonnegative since $\mu(r,\ \theta_0)$ is a convex function of r. On verifying the other requirements, we have that, for each fixed value θ_0, $\lambda_0(k,\ \theta_0)$ satisfies Condition A(k).

Since

$$\lambda_1(k;\ \alpha,\ \beta) \equiv \int_\alpha^\beta \lambda_0(k,\ \theta)\, d\theta,$$

it follows (See [4, Theorem 1, p. 287].) that $\lambda_1(k;\ \alpha,\ \beta)$ satisfies Condition A(k). If the function $l_S(k;\ \alpha,\ \beta)$ alone is considered, then the relations used also indicate that $l_S(k;\ \alpha,\ \beta)$ is a convex function of k, that it is strictly convex if S is not a developable surface, and that it is linear (as a function of k) if S is developable.

Now, with (5.41),

$$\lambda_2(k;\ \alpha,\ \beta) \equiv \int_\alpha^\beta \int_0^r [\mu(\rho,\ \theta) - \rho]\, d\rho\, d\theta,$$

and its first and second derivatives are found to be nonnegative by use of the convexity of $\mu(r,\ \theta)$. The remainder of the argument is direct.

We find other results for the area functions:

THEOREM 5.5. *Let S be an analytic sub-K_0 surface referred to geodesic polar coordinates with given pole P_0. Let $a_S(k;\alpha,\beta)$, $a_{S_0}(k;\alpha,\beta)$, and $a_P(k;\alpha,\beta)$ denote the areas of the sectors formed by the curve of the family $C(k)$ of (5.41) for the parameter value k, $\theta=\alpha$, and $\theta=\beta$ $(\alpha<\beta)$ on the surface S, the K_0-surface S_0, and in the plane respectively. Then the function*

$$\lambda_2(k;\alpha,\beta)\equiv a_S(k;\alpha,\beta)-a_P(k;\alpha,\beta)$$

is a monotonic nondecreasing sub-K_0 function of k, and the function

$$\tau_1(k;\alpha,\beta)\equiv a_S(k;\alpha,\beta)-a_{S_0}(k;\alpha,\beta)$$

satisfies Condition C(k).

Proof. By Theorem 5.4, $\lambda_2(k;\alpha,\beta)$ is nonnegative and monotonic nondecreasing. By calculation,

$$\mathfrak{S}_k\lambda_2(k;\alpha,\beta)=\int_\alpha^\beta\int_0^r\left[\mathfrak{S}_\rho\,\mu+\frac{\partial^2\mu}{\partial\rho^2}\,(f^2-1)-K_0\rho\right]d\rho\,d\theta>0$$

for $r>0$ $(k>0)$ on S since $f(\theta)\geq 1$. Hence $\lambda_2(k;\alpha,\beta)$ is a sub-K_0 function of k.

For the other function, we find that

$$\frac{\partial\tau_1}{\partial k}=\int_\alpha^\beta\left[\mu(r,\theta)-\mu_0(r,\theta)\right]f(\theta)\,d\theta$$

and

$$\mathfrak{S}_k\tau_1(k;\alpha,\beta)=\int_\alpha^\beta\int_0^r\left[f^2\,\mathfrak{S}_\rho(\mu-\mu_0)-K_0(\mu-\mu_0)(f^2-1)\right]d\rho\,d\theta.$$

These are nonnegative by the proof of Theorem 3.3, and the rest of the argument is immediate.

5.5. The Steiner configuration. Let C be an arbitrary closed convex curve in the plane, of length L and area F, and let $C(\rho)$ be a curve parallel to C at a distance ρ from it, ρ being measured along the outward normal to C, of length $L(\rho)$ and area $F(\rho)$. The family of curves $C(\rho)$ will be called a *Steiner configuration*; it is a classical result of Steiner [2, p. 128] that

$$L(\rho) = L + 2\pi\rho$$

and

$$F(\rho) = F + \rho L + \pi\rho^2.$$

Generalizations of these formulas for curves lying on a curved surface have been given in [1; 2], and explicit formulas found in the case of surfaces of constant curvature. We shall establish the sub-K_0 function property of some functions which involve the $L(\rho)$ and $F(\rho)$ functions for the Steiner configuration associated with a suitable curve C on an arbitrary sub-K_0 surface, $K_0 < 0$. It is evident that our preceding theory for geodesic circles of center P_0 on S is obtained from a Steiner configuration on S if the curve C is a geodesic circle of center P_0 on S.

Let the curve C be a simple, closed, bounding, and differentiable curve on the surface S. Introduce a geodesic representation with coordinates (u, v) in which $u = 0$ is the curve C, and $v = constant$ are the geodesics orthogonal to C; further, let v be the arc length of C measured positively for motion on the curve which keeps the bounded area to the left, and let u be the arc length of geodesics normal to C. Sufficient conditions for the validity of such a coordinate system in a region of S have been given [1; 2]. We shall assume that our coordinate system is valid and term *admissible* those curves which satisfy the above conditions.

Then, for an admissible curve C of length L and area F, and for ρ fixed, the length $L(\rho)$ of $C(\rho)$ is given by

$$L(\rho) = \int_C \mu(\rho, v)\,dv, \tag{5.51}$$

and the area $F(\rho)$ of $C(\rho)$ is given by

$$F(\rho) = F + \int_C \int_0^\rho \mu(u, v)\,du\,dv. \tag{5.52}$$

For a K_0-surface, $K_0 < 0$, Abascal [1, p. 843] has shown that these relations simplify to (in our notation)

$$L_0(\rho) = L + l_0(\rho) - \frac{K_0}{2\pi}[F\,l_0(\rho) + L\,a_0(\rho)], \tag{5.53}$$

and

$$F_0(\rho) = F + a_0(\rho) + \frac{L\,l_0(\rho)}{2\pi} - \frac{K_0\,F\,a_0(\rho)}{2\pi}, \tag{5.54}$$

where $l_0(\rho)$ and $a_0(\rho)$ are given in Lemma 2.4.

LEMMA 5.6. *The functions $L_0(\rho)$ and $F_0(\rho)$ satisfy the relations*

$$L_0'(\rho) + K_0 F_0(\rho) - 2\pi = 0,$$

$$L_0''(\rho) + K_0 L_0(\rho) = 0,$$

and

$$\frac{L_0^2(\rho)}{4\pi} + \frac{K_0 F_0^2(\rho)}{4\pi} - F_0(\rho) = M = \text{constant} \tag{5.55}$$

for $\rho \geq 0$ on the K_0-surface S_0.

Proof. The first two relations follow easily from (5.53), (5.54), and the properties of $l_0(\rho)$ and $a_0(\rho)$. The third relation is immediate since the derivative of its left member is zero.

THEOREM 5.7. *Let S be an analytic sub-K_0 surface, and let $C(\rho)$ denote the curves of the Steiner configuration for an admissible curve C on S. Then the length function $L(\rho)$ is a sub-K_0 function of ρ; $L(\rho)$ is a strictly sub-K_0 function if S is a strictly sub-K_0 surface, and it is a K_0-function of ρ if S is a K_0-surface. Further, the area function $F(\rho)$ is a strictly sub-K_0 function.*

Proof. These properties of $L(\rho)$ were established in Theorem 2.2.

By calculation from (5.52) we get

$$\mathfrak{S}_\rho F(\rho) = \int_C \int_0^\rho \mathfrak{S}_u\, \mu\, du\, dv + \int_C \left(\frac{\partial \mu}{\partial u}\right)_{\rho=0} dv + K_0 F. \tag{5.56}$$

For our geodesic representation, it is known [7, p. 188] that

$$K_g(v) = \left[\frac{\partial \mu(u, v)}{\partial u}\right]_{u=0},$$

where $K_g(v)$ is the geodesic curvature of C. By the Gauss-Bonnet theorem [7, p. 191], noting that C has no exterior angles, we get

$$\int_C K_g(v)\,dv = 2\pi - \iint_F K\mu\,du\,dv \geq 2\pi - K_0 F. \tag{5.57}$$

since $K \leq K_0$. With $\mathfrak{S}_u \mu \geq 0$ and (5.57), it follows from (5.56) that

$$\mathfrak{S}_\rho F(\rho) \equiv L'(\rho) + K_0 F(\rho) \geq 2\pi, \tag{5.58}$$

and then Theorem 1.2 ensures the result of the theorem.

We shall now make comparison between the length and area functions for Steiner configurations on a sub-K_0 surface S and on the K_0-surface. However, our expressions may be considered to be functions formed with respect to S alone because of (5.53), (5.54), and the known formulas for $l_0(\rho)$ and $a_0(\rho)$.

THEOREM 5.8. *Let S be an analytic sub-K_0 surface, and let $C(\rho)$ denote the curves of the Steiner configuration for an admissible curve C of length L and area F on S. Let $C_0(\rho)$ denote the curves of the Steiner configuration for any admissible curve C_0 of length $L_0 = L$ and area $F_0 = F$ on the K_0-surface S_0. Then the functions*

$$\tau_2(\rho) \equiv L(\rho) - L_0(\rho)$$

and

$$\tau_3(\rho) \equiv F(\rho) - F_0(\rho)$$

satisfy Condition $\mathrm{C}(\rho)$, *where ρ is the parameter of the family.*

Proof. There is equality in (5.57) if S is a K_0-surface, and the proof using (5.51) and (5.52) is similar to those of Theorem 3.3 and Theorem 3.5.

Theorem 5.8 admits the corollary that *the functions $L(\rho)$ and $F(\rho)$ satisfy the inequalities*

$$L(\rho) \geq L_0(\rho) \equiv L + l_0(\rho) + \frac{|K_0|}{2\pi}\,[F\,l_0(\rho) + L a_0(\rho)]$$

and

$$F(\rho) \geq F_0(\rho) \equiv F + a_0(\rho) + \frac{L\,l_0(\rho)}{2\pi} + \frac{|K_0|\,F\,a_0(\rho)}{2\pi}\ ;$$

both functions are strictly sub-K_0 functions and satisfy the strict inequalities *for* $\rho > 0$ *on* S *if* S *is not a* K_0*-surface, and they satisfy the equalities if* S *is a* K_0*-surface.* We remark that the conditions $L_0 = L$ and $F_0 = F$ were imposed to meet the requirements of Condition $\mathrm{C}(\rho)$. The sub-K_0 function properties and inequality relations above would hold equally well for any admissible C_0 such that $L_0 \le L$ and $F_0 \le F$.

We shall now establish some results for functions involving $L(\rho)$ and $F(\rho)$ which are analogous to the isoperimetric function and to its modifications.

THEOREM 5.9. *Let* $L(\rho)$ *and* $F(\rho)$ *be the length and area functions, respectively, of the curves of a Steiner configuration on an analytic sub-*K_0 *surface* S. *Then the function*

$$\theta(\rho) \equiv \frac{L^2(\rho)}{4\pi} - F(\rho)$$

is a positive monotonic strictly increasing sub-K_0 *function of* ρ; *further, if* C_0 *on the* K_0*-surface* S_0 *satisfies* $L_0 = L$ *and* $F_0 = F$, *then the function*

$$\tau_4(\rho) \equiv \frac{L^2(\rho) - L_0^2(\rho)}{4\pi} - [F(\rho) - F_0(\rho)]$$

satisfies Condition $\mathrm{C}(\rho)$.

Proof. It is known [6] that $\theta(\rho) > 0$ on sub-K_0 surfaces. With (5.58) and Theorem 5.7, routine computations show that $\theta'(\rho)$ and $\mathfrak{S}\,\theta(\rho)$ are positive, establishing the properties of $\theta(\rho)$. The properties of $\tau_4(\rho)$ are established in routine manner by the use of (5.58) and Theorem 5.8.

THEOREM 5.10. *Let* $L(\rho)$ *and* $F(\rho)$ *be the length and area functions, respectively, of the curves of a Steiner configuration on a sub-*K_0 *surface* S, *and let* $L_0(\rho)$ *and* $F_0(\rho)$ *be the length and area functions, respectively, of the curves of a Steiner configuration on a* K_0*-surface* S_0. *Let the admissible curve* C_0 *on* S_0 *satisfy* $L_0 = L$ *and* $F_0 = F$. *Then the function*

$$\tau_5(\rho) \equiv \left[\frac{L^2(\rho)}{4\pi} + \frac{K_0 F^2(\rho)}{4\pi} - F(\rho)\right] - \left[\frac{L_0^2(\rho)}{4\pi} + \frac{K_0 F_0^2(\rho)}{4\pi} - F_0(\rho)\right]$$

satisfies Condition $\mathrm{C}(\rho)$.

Proof. Obviously $\tau_5(0) = 0$, and using (5.55) we get $\tau_5'(\rho) \ge 0$ by (5.58).

By another calculation we find that

$$4\pi\mathfrak{S}\tau_5(\rho) \geq \{2L'(\rho)[L'(\rho)+K_0F(\rho)-2\pi]+K_0L^2(\rho)+K_0^2F^2(\rho)$$
$$-4\pi K_0F(\rho)-4\pi K_0M\},$$

where the constant M is given by (5.55). We then use (5.58) just as we used (3.22) in the proof of Theorem 4.1, and we get

$$(5.59)\qquad 4\pi\mathfrak{S}\tau_5(\rho) \geq [L'^2(\rho)+K_0L^2(\rho)-4\pi^2-4\pi K_0M] \equiv h(\rho)$$

where $h(\rho)$ is the function in brackets. By (5.53) and (5.54), we verify that

$$4\pi K_0M = L_0'^2(\rho) + K_0L_0^2(\rho) - 4\pi^2,$$

and when this is substituted in (5.59), it follows that $h(0) \geq 0$ since $L'(0) \geq L_0'(0)$ by (5.58). By computation and use of Theorem 5.7 we find that $h'(\rho) \geq 0$ for $\rho \geq 0$; hence $h(\rho) \geq 0$ for $\rho \geq 0$, and $\tau_5(\rho)$ is a sub-K_0 function. Since further considerations show that the signs of equality hold above if and only if S is a K_0-surface, we have, with the final remark that $\tau_5(\rho) \equiv 0$ if S is a K_0-surface, the result that $\tau_5(\rho)$ satisfies Condition C(ρ).

We remark that the last two theorems imply inequalities for the functions $L(\rho)$ and $F(\rho)$ somewhat similar to (4.13); we omit the formal statements.

Let the symbol $\Phi_k(\rho)$ denote the new functions produced from the functions $\phi_k(r)$ when the functions r, $a(r)$, $a_0(r)$, $l(r)$, and $l_0(r)$, associated with geodesic circles, are replaced by the functions ρ, $F(\rho)$, $F_0(\rho)$, $L(\rho)$, and $L_0(\rho)$, respectively, associated with the Steiner configuration of an admissible curve C. For example,

$$\tau_5(\rho) \equiv \Phi_3(\rho) - M,$$

where M is the constant of (5.55). It may then be verified (indeed almost solely by inspection of the proof that the corresponding $\phi(r)$ function satisfies Condition C) that *the functions*

$$\Phi_k(\rho) - M \qquad (k = 4, 5),$$

and

$$\Phi_k(\rho) \qquad (k = 6, 7, 8, 9, 10, 11),$$

satisfy Condition C(ρ).

Again, we might formulate a Condition D(ρ) which is analogous to Condition D in the way that Condition C(ρ) corresponds to Condition C. Then large parts of the theory in § 3.3 on "square-root" functions are found to apply to similar functions associated with a Steiner configuration. Finally, it may be shown that much of the theory in §§ 5.2, 5.3 can be generalized to hold for appropriate functions associated with Steiner configurations.

REFERENCES

1. E. V. Abascal, *A generalization of Steiner's formulae*, Bull. Amer. Math. Soc. 53 (1947), 841-844.

2. C. B. Allendoerfer, *Steiner's formulae on a general* S^{n+1}, Bull. Amer. Math. Soc. 54 (1948), 128-135.

3. E. F. Beckenbach, *Generalized convex functions*, Bull. Amer. Math. Soc. 43 (1937), 363-371.

4. ———, *Some convexity properties of surfaces of negative curvature*, Amer. Math. Monthly 55 (1948), 285-301.

5. E. F. Beckenbach and R. H. Bing, *On generalized convex functions*, Trans. Amer. Math. Soc. 58 (1945), 220-230).

6. E. F. Beckenbach and T. Radó, *Subharmonic functions and surfaces of negative curvature*, Trans. Amer. Math. Soc. 35 (1933), 662-674.

7. L. P. Eisenhart, *An introduction to differential geometry*, Princeton, 1940.

8. E. L. Ince, *Ordinary differential equations*, New York, 1944.

9. M. M. Peixoto, *Generalized convex functions and second order differential inequalities*, Bull. Amer. Math. Soc. 55 (1949) 563-572.

10. H. Shniad, *Convexity properties of means of moduli of analytic functions*, Thesis, University of California, Los Angeles, 1948.

UNIVERSITY OF CALIFORNIA, LOS ANGELES, AND
STATE COLLEGE OF WASHINGTON

ON THE HITCHCOCK DISTRIBUTION PROBLEM

MERRILL M. FLOOD

1. Introduction. Frank L. Hitchcock [1] has offered a mathematical formulation of the problem of determining the most economical manner of distribution of a product from several sources of supply to numerous localities of use, and has suggested a computational procedure for obtaining a solution of his system in any particular case. L. Kantorovitch [2], Tjalling C. Koopmans [3], George B. Dantzig [4b], C. B. Tompkins [5], Julia Robinson [7; 8], Alex Orden [6], and others [4] have also discussed the computational aspects of this problem; paper [5] illustrates the use of the "projection method," due to C. B. Tompkins, as a computational process applicable to either of the Fundamental Problems of the present paper.

We shall be concerned only with the mathematical justification of computational procedure, and shall limit our attention to one specific method of solution of general validity. No attempt will be made to compare the various methods already proposed, either as to their mathematical similarity or as to their relative efficiency in any particular case.

2. The problem. The problem is to find a set of values of the mn variables x_{ij}, subject to the following conditions:

$$\sum_{i=1}^{m} x_{ij} = c_j, \quad \sum_{j=1}^{n} x_{ij} = r_i, \tag{2.1}$$

$$x_{ij} \geq 0, \tag{2.2}$$

Received January 25, 1952. The author's interest in the problem was aroused by papers on transportation theory presented by Koopmans [4a] and Dantzig [4b] at a conference on linear programming in Chicago during June, 1949, under the auspices of the Cowles Commission for Research in Economics of the University of Chicago. Several other papers presented at this conference are of closely related interest. Professor Koopmans, in his Introduction to the Conference Proceedings [4], also discussed the background and interrelationship of the conference papers — including the bearing of some of these on the Hitchcock distribution problem. The results of the present paper have been presented in three seminar lectures: once in December, 1949, at The RAND Corporation in Santa Monica, once in July, 1950, at the Institute for Numerical Analysis of the National Bureau of Standards in Los Angeles, and once in June, 1951, at the National Bureau of Standards in Washington, D.C. The author is especially indebted to Dr. D.R. Fulkerson, who has given real assistance in simplifying notation and proofs of theorems, for a careful reading of the manuscript.

Pacific J. Math. 3 (1953), 369-386

$$\sum_{i,j} x_{ij}\, d_{ij} = \text{minimum}. \tag{2.3}$$

The numbers m, n, r_i, c_j, and d_{ij} are given positive integers with $\sum c_j = \sum r_i$. The indices i and j are understood always to range over these same integers m and n, respectively; it is also assumed, for convenience, that $m \geq n$. Any set of values x_{ij} that satisfies all these conditions is called a *solution* of the problem.

There is no loss of generality in assuming that the d_{ij} are positive integers, rather than rational numbers, since the problem is essentially unchanged if d_{ij} is replaced by $ad_{ij} + b$, where a and b are any positive rational numbers. We have not examined the case in which some of the quantities r_i, c_j, and d_{ij} are irrational. The only effect of irrationality on the results of the present paper is a possible lack of convergence of the iterative process of solution. These considerations are not of importance in the usual applications.

It will sometimes be more convenient to use an alternative statement of the problem, in matrix notation, as follows:

$$M'y \geq b, \tag{2.4}$$

$$y \geq 0, \tag{2.5}$$

$$a'y = \text{minimum}. \tag{2.6}$$

It is easily seen that the two formulations are equivalent if y, a, b, and M' are defined as follows:

$$y_{n(i-1)+j} = x_{ij},$$

$$a_{n(i-1)+j} = d_{ij},$$

$$M' = \begin{Vmatrix} I_n & I_n & \cdots & I_n \\ -I_n & -I_n & \cdots & -I_n \\ J_1 & J_2 & \cdots & J_m \\ -J_1 & -J_2 & \cdots & -J_m \end{Vmatrix}, \qquad b = \begin{Vmatrix} c \\ -c \\ r \\ -r \end{Vmatrix},$$

where I_n is the identity matrix of order n, and J_i is the $m \times n$ matrix with all elements zero except for the ith row in which each element is unity. Of course, y, a, c, and r are column matrices (or vectors) with components $y_{n(i-1)+j}$, $a_{n(i-1)+j}$, c_j, and r_i, respectively, and a prime denotes the transpose of a matrix (or vector).

3. Fundamental theorems. There are several fundamental theorems concerning systems of linear inequalities that are useful for this paper. I reproduce their statements here in a form given by A. W. Tucker in an unpublished note dated December, 1949. The interested reader can find proofs of these theorems, and of others of similar type, in a paper by Gale, Kuhn, and Tucker [4c].

FUNDAMENTAL PROBLEMS. (Here lower case roman letters denote one-column vectors, while capitals denote rectangular matrices; M, a, and b are given, but d is to be determined.)

PROBLEM I. To satisfy the constraints $Mx \leq a$, $x \geq 0$, and make $b'x = d$ for d *maximal* in the sense that no x satisfying the constraints makes $b'x > d$.

PROBLEM II. To satisfy the constraints $M'y \geq b$, $y \geq 0$, and make $a'y = d$ for d *minimal* in the sense that no y satisfying the constraints makes $a'y < d$.

Problems I and II are said to be *dual*.

FUNDAMENTAL FEASIBILITY THEOREM. *The constraints in a problem are feasible (that is, satisfied by some x or y) if and only if the dual problem in homogeneous form (that is, with $b = 0$ or $a = 0$) has a null solution.*

FUNDAMENTAL EXISTENCE THEOREM. i. *The vectors x and y are solutions of Problems* I *and* II *if and only if they satisfy their constraints in the two problems and make $a'y = b'x$. Such x and y exist if the constraints in both problems are feasible.*

ii. *A problem has a solution if and only if its constraints are feasible and its homogeneous form has a null solution.*

FUNDAMENTAL DUALITY THEOREM. *A problem has a solution (for a unique d) if and only if the dual problem has a solution (for the same d).*

4. The dual and combined problems. We note that the problem, as stated in relations (2.4)-(2.6), is a Fundamental Problem of form II. The dual problem is:

$$Mx \leq a, \tag{4.1}$$

$$x \geq 0, \tag{4.2}$$

$$b'x = \text{maximum}. \tag{4.3}$$

This can be rewritten in a more convenient form, for our present purposes, as follows:

$$v_j - u_i \leq d_{ij}, \tag{4.4}$$

$$\sum c_j v_j - \sum r_i u_i = \text{maximum}, \tag{4.5}$$

where $v_j = x_j - x_{n+j}$, and $u_i = - x_{2n+i} + x_{2n+m+i}$; we omit the condition (4.2), that $x \geq 0$, since this imposes no limitation on u_i and v_j.

THEOREM 1. *The problem has a solution.*

Proof. By the Fundamental Existence Theorem, there is a solution if and only if the constraints are feasible and $y = 0$ is a solution of the problem when $b = 0$. Now

$$\sum c_j = \sum r_i,$$

so

$$x_{ij} = \frac{r_i c_j}{\sum r_i}$$

satisfies the constraints. When $b = 0$, obviously the only values that satisfy the constraints are $x_{ij} = 0$, and so the theorem is proved.

By the Fundamental Duality Theorem, we see:

COROLLARY 1A. *The dual problem has a solution.*

THEOREM 2. *The numbers x_{ij}, and u_i, v_j, are solutions of the problem and the dual, respectively, if and only if they satisfy*:

$$\sum_j x_{ij} = r_i, \quad \sum_i x_{ij} = c_j, \quad x_{ij} \geq 0, \tag{4.6}$$

$$d_{ij} + u_i - v_j \geq 0, \tag{4.7}$$

$$x_{ij}(d_{ij} + u_i - v_j) = 0. \tag{4.8}$$

Proof. Since (4.6) and (4.7) are simply the constraints for the problem and the dual, respectively, it remains only to show that (4.8) is equivalent to the condition $a'y - b'x = 0$. Now

$$a'y - b'x = \sum_{i,j} x_{ij} d_{ij} - \sum_j c_j v_j + \sum_i r_i u_i$$

$$= \sum_{i,j} x_{ij} d_{ij} - \sum_{i,j} x_{ij} v_j + \sum_{i,j} x_{ij} u_i = \sum_{i,j} x_{ij}(d_{ij} + u_i - v_j).$$

Since each term in this sum is nonnegative,

$$a'y - b'x = 0$$

if and only if

$$x_{ij}(d_{ij} + u_i - v_j) = 0.$$

We refer to the problem of finding values for x_{ij}, u_i, and v_j that satisfy (4.6)-(4.8) as the "combined problem", and note that the combined problem always has a solution.

5. **Linear graphs.** It will be convenient, for some purposes, to associate linear graphs [9] with certain subsets of the elements of a matrix $S = ||s_{hk}||$. If I is a given subset of the elements of S, we define the *I-graph* L of S as follows: the *vertices* of L are all the points (h, k) in the Cartesian plane for which $s_{hk} \in I$; the *arcs* of L are all line-segments joining pairs of *neighboring* vertices with either equal abscissas or equal ordinates, where two vertices with equal abscissas (ordinates) are neighboring if they are not separated by another vertex of L with the same abscissa (ordinate). For the moment, denote the vertices of L by symbols $a, b, c, \cdots, f$, and the arcs by symbols such as $ab, bc, \cdots, cf$ (no distinction is made between the arcs ab and ba). Then a *chain* is a set of one or more distinct arcs that can be arranged as $ab, bc, \cdots, de$, ef, where vertices denoted by different symbols are distinct. A *cycle* is a set of distinct arcs (at least four are necessary) that can be ordered as ab, $bc, \cdots, ef$, fa, the vertices being distinct as in the case of a chain. A graph is *connected* if each pair of vertices is joined by a chain. A *forest* is a graph containing no cycles, and a *tree* is a connected forest.

If L contains v vertices, a arcs, and p connected pieces, the number $\mu = a - v + p$ is known as the *cyclomatic number* (or first Betti number) of L. It follows from a well-known theorem [9] concerning linear graphs in general that: (i) L is a forest if and only if $\mu = 0$, and (ii) L contains just one cycle if and only if $\mu = 1$.

Note that L contains a cycle if and only if there is a subset of I that can be arranged as a sequence

$$s_{h_1 k_1},\ s_{h_1 k_2},\ s_{h_2 k_2},\ s_{h_2 k_3},\ \cdots,\ s_{h_\sigma k_\sigma},\ s_{h_\sigma k_1},$$

where the h's and k's are distinct among themselves; and L contains a single cycle if and only if I contains just one subset that can be arranged in the displayed form. We call such a subset of I an *I-circuit* on S, and denote it by $[S_\sigma]$. For a particular arrangement of $[S_\sigma]$, we also refer to the terms $s_{h_\alpha k_\alpha}$ as *odd-terms*, the others as *even-terms*.

In case I consists of all $s_{hk} > 0$, as it frequently will, we speak of the *positive graph* of S, *positive circuits* on S, and abbreviate such statements as "the positive graph of S is a forest" to "S is a forest".

6. The method of solution. In the method of solution to be developed for the problem, we start with a special set of values $X^0 = ||x_{ij}^0||$ that satisfy the constraints (4.6). We then test to determine whether or not there exist u_i and v_j satisfying the relations (4.7) and (4.8) for the given X^0. If so, then X^0 is a solution, otherwise not. The method next yields a new trial matrix $X^1 = ||x_{ij}^1||$, if X^0 is not a solution, such that

$$\sum_{i,j} (x_{ij}^0 - x_{ij}^1) d_{ij} \geq 1.$$

After a finite number of steps this process necessarily must terminate, and it leads to an exact integral solution of the problem.

The first trial matrix X^0 is a forest of t trees, and has $m + n - t$ nonzero elements. According as $t = 1$ or $t > 1$, two essentially different cases may be met at each stage of the solution process.[1]

At each stage when $X = ||x_{ij}||$ is a tree, the equations (4.8) have a general solution for u_i and v_j with one free parameter, say u_1. However, the quantities $d_{ij} + u_i - v_j$ are uniquely determined in this case, so it is sufficient to calculate them and note whether or not they are all nonnegative in order to decide whether or not X is a solution. If some

[1] These are the nondegenerate and degenerate cases in the work of Dantzig [4b]. We shall use these terms also. The method of solution developed by Dantzig [4b] for the nondegenerate case is essentially the same as the one in the present paper, although the derivations of the results are quite different. Orden [6] has subsequently given an elegant method for reducing the degenerate case to the nondegenerate one, as an extension of the ϵ-method proposed by Dantzig [4b]. The author believes that the treatment of the degenerate case provides the only results in the present paper that are new, or at least fresh for the Hitchcock problem, and also of some mathematical interest. It also seems likely that the method given here will often be more efficient computationally, in the degenerate case, than the Dantzig-Orden ϵ-method.

$$d_{i_1 j_1} + u_{i_1} - v_{j_1} < 0,$$

then there is a unique I-circuit $[X_s]$ on X, where I consists of $x_{i_1 j_1}$ and all positive x_{ij}, that may be arranged with $x_{i_1 j_1}$ as the second term, say. Let g denote the smallest odd-term of $[X_s]$. Then the new trial matrix X^* is obtained from X by adding g to the even terms of $[X_s]$, subtracting g from the odd-terms, and leaving the other elements of X unchanged.

At each stage when X is a forest of $t > 1$ trees, the equations (4.8) have a general solution for u_i and v_j with t independent parameters, and the quantities $d_{ij} + u_i - v_j$ involve $t - 1$ independent parameters. The rows and columns of the matrix X are rearranged so that it can be represented as a square matrix of order t whose t^2 elements are submatrices X_{ab} such that $X_{ab} = 0$ if $a \neq b$, and X_{aa} is a tree with $m_a + n_a - 1$ nonzero elements and is of order $m_a \times n_a$. It may also be assumed that each X_{aa} is a solution of its subproblem. We can select

$$u_1, \; u_{m_1+1}, \; \cdots, \; u_{m_1+\cdots+m_{t-1}+1}$$

to be the t parameters. If we assign these the value zero and denote this particular solution of (4.8) by $\bar{u}_i$ and $\bar{v}_j$, then we may define numbers

$$\bar{p}_{ij} = d_{ij} + \bar{u}_i - \bar{v}_j .$$

We partition the matrix $\bar{P} = ||\bar{p}_{ij}||$ into submatrices corresponding to the X_{ab} and denote them $\bar{P}_{ab}$. Let p_{ab} be the smallest element in $\bar{P}_{ab}$ and define the square matrix P of order t by $P = ||p_{ab}||$. To designate the position of p_{ab} in the matrix $\bar{P} = ||\bar{p}_{ij}||$, we may write p_{ab} alternatively as

$$\bar{p}_{a\,b}^{\,i_a\, j_b},$$

the subscripts referring to the submatrix and the superscripts to the rows and columns in the submatrix. When it introduces no ambiguity, the subscripts on the superscripts will be omitted in order to simplify the notation.

The test as to whether or not X is a solution consists of forming all sums

$$p_{a_1 a_2 \cdots a_h} = p_{a_1 a_2} + p_{a_2 a_3} + \cdots + p_{a_h a_1}$$

for $h = 2, 3, \cdots, t$, where $(a_1\, a_2 \cdots a_h)$ is any permutation of h different positive integers, none greater than t; X is a solution if and only if all such sums are nonnegative.

If any

$$p_{a_1 a_2 \cdots a_h} < 0,$$

then there is a unique I-circuit $[X_s]$ on X, where I consists of all positive x_{ij} together with all x_{ij} that correspond to the terms

$$\bar{p}^{i}_{a_k}\,{}^{j}_{a_{k+1}} \quad \text{of } p_{a_1 a_2 \cdots a_h},$$

which can be arranged to involve all

$$x^{i}_{a_k}\,{}^{j}_{a_{k+1}}$$

as even-terms. If g is the smallest odd-term in $[X_s]$, then (as in the nondegenerate case) the new trial matrix X^* is obtained by adding g to the even-terms of $[X_s]$, subtracting g from the odd-terms, and leaving the other elements of X unchanged.

7. The initial trial solution. An X that satisfies (4.6) will be called a *trial solution.* It would be all right to take the positive values

$$\frac{r_i c_j}{\sum r_i}$$

for the initial trial solution $X^0 = ||x^0_{ij}||$. An alternative is to construct an initial trial solution that is a forest. It is always possible to do this in integral values. The following theorem certifies the existence of such an integral trial solution. The method of proof shows how to construct one.

THEOREM 3. *There is a matrix* $X^0 = ||x^0_{ij}||$ *with integral elements that satisfies* (4.6) *and is a forest.*

Proof. The theorem is trivial for $m = 1$. Assume the theorem is true for m and consider the case $m + 1$.

Let the notation be chosen so that

$$r_1 \geq r_2 \geq \cdots \geq r_{m+1} > 0, \text{ and } c_1 \geq c_2 \geq \cdots \geq c_n > 0.$$

If $n < m + 1$, then $c_1 > r_{m+1}$. If $n = m + 1$ then $c_1 > r_{m+1}$ unless $c_i = r_j = \lambda$

(for all i and j); in this latter case $X^0 = \lambda$ satisfies the conditions of the theorem. Hence, by the induction hypothesis, there is a set of nonnegative integers x^*_{ij} $(i = 1, \cdots, m)$ such that

$$\sum_i x^*_{ij} = c_j - \delta_{1j} r_{m+1}, \quad \sum_j x^*_{ij} = r_i, \quad \text{and} \quad X^* = ||x^*_{ij}||$$

is a forest. Then X^0, defined by

$$x^0_{ij} = x^*_{ij}, \quad x^0_{m+1\ j} = \delta_{1\,j}\, r_{m+1}\,,$$

satisfies (4.6). Now since the $(m+1)$st row, with only one positive element, clearly cannot contribute terms to a positive circuit, X^0 is also a forest; the theorem is proved.

To apply this method, in the construction of a trial solution, search for the smallest r_{i_1} and the largest c_{j_1}, and then set

$$x^0_{i_1\, j_1} = r_{i_1}\,.$$

In effect, this deletes the i_1st row, after c_{j_1} is replaced by $c_{j_1} - r_{i_1}$, and the process is repeated (with interchanged rows and columns as necessary) until all x^0_{ij} have been determined. For automatic machine calculation, the procedure is easily made unique, for any one starting order of rows and columns, by specifying that the search is first on row-totals when the number of rows is the same as the number of columns at any stage, and that the row-total or column-total with the smallest index is chosen whenever at any stage there are several equal values to choose from. This initial trial solution will be called "preferred" for identification.[2]

THEOREM 4. *A trial solution that is a forest of t trees has $m + n - t$ nonzero elements.*

Proof. Observe first that if the trial solution X is a forest of t trees, the rows and columns of X can be rearranged so that X has the form

[2]Sometimes, as in this instance, we indicate how to make a unique choice among possible alternatives at each computational step, but usually do not. It is necessary to do this in order completely to routinize the computing steps, of course, but the matter presents no difficulty and we omit it here.

$$\begin{Vmatrix} X_{11} & 0 & \cdots & 0 \\ 0 & X_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & X_{tt} \end{Vmatrix},$$

where each X_{aa} is a tree. Consequently, the theorem amounts to proving that an $m \times n$ matrix with no zero rows or columns, which is a tree, has $m + n - 1$ positive elements. If $m + n = 2$, this is obvious, so assume the statement to be true for all matrices for which $m + n = k$ and consider one for which $m + n = k + 1$. Since $m \geq n$, clearly some row has only one positive element, as otherwise there would be a positive circuit. Delete this row and apply the induction hypothesis.

In actual cases when m and n are relatively small, or when there is other reason to believe that an initial trial solution better than the preferred one can be found by trial and error, it may be better to construct the initial trial solution in some other way than the one given in the proof of Theorem 3, in order to reduce the number of steps required in the iterative process.

The methods developed in this paper apply directly for any trial solution that is a forest, and are readily extended for other cases. It is easy to see that there must be at least one solution which is a forest.

8. Nondegenerate case. We consider now the case of a trial solution X which is a tree. Let the positive elements of X be

$$x_{i_a j_a} \qquad (a = 1, \cdots, m + n - 1).$$

We shall need the following theorem.

THEOREM 5. *If X is a trial tree, the set of equations*

$$(8.1) \qquad d_{ij} + u_i - v_j = 0 \text{ for } (i, j) = (i_a, j_a),$$

has the general solution

$$u_i = u_i^* + z, \quad v_j = v_j^* + z,$$

where (u_i^*, v_j^*) is a particular solution and z is arbitrary.

Proof. The theorem is apparent for $m = 1$, and we proceed by induction.

Suppose the theorem is true for all trial trees of m rows, and let X be an $(m+1)\times n$ trial tree. Obviously, there must be at least one row of X that has exactly one nonzero element; we may suppose it to be $x_{m+1\,n}$ without loss of generality—also that

$$i_{m+n-1} = m+1 \text{ and } j_{m+n-1} = n.$$

Since X is a trial tree, the matrix obtained from X by deleting the last row (or, if $m+1=n$, its transpose) is also. The induction hypothesis implies that the general solution of (8.1), with the final equation omitted, is of the form

$$u_i = u_i^* + z, \; v_j = v_j^* + z.$$

We note next that this final equation becomes

$$u_{m+1} = (v_n^* - d_{m+1\,n}) + z = u_{m+1}^* + z.$$

The theorem follows easily.

It will be convenient to call the particular solution $\overline{u}_i$, $\overline{v}_j$ of (8.1) obtained by setting $u_1 = 0$ the *preferred* trial solution of the dual problem corresponding to the trial tree X. As an obvious consequence of Theorem 5, we state:

COROLLARY 5A. *If X is a trial tree, then it is a solution of the problem if and only if the corresponding preferred trial solution* $(\overline{u}_i, \overline{v}_j)$ *of the dual problem satisfies*

$$d_{ij} + \overline{u}_i - \overline{v}_j \geq 0$$

for all i and j.

All that is needed now in order to establish the method for the nondegenerate case is to show how to construct a new trial matrix X^*, if X is not a solution, such that

$$\sum_{i,j} (x_{ij} - x_{ij}^*) d_{ij} \geq 1.$$

In this case, it follows by Corollary 5A that

$$d_{kl} + \overline{u}_k - \overline{v}_l < 0$$

for at least one pair (k, l) and, of course, $x_{kl} = 0$.

THEOREM 6. *If the trial solution X is a tree, and $x_{kl} = 0$, then there is a unique I-circuit on X, where I consists of all positive x_{ij} together with x_{kl}.*

Proof. It suffices to show that the I-graph of X has cyclomatic number $\mu = 1$. By assumption, the positive graph of X has cyclomatic number zero; and since X must have positive elements x_{al} and x_{kb} for some a and b, the I-graph of X has two more arcs, one more vertex, and the same number (one) of connected pieces. Hence $\mu = 1$, and the proof is complete.

Now arrange this unique I-circuit $[X_s]$ with x_{kl} as the second term, and let g be the minimum of the odd-terms of $[X_s]$ in this arrangement. If we subtract g from the odd-terms, add g to the even terms, and leave the remaining elements of X unchanged, we get a matrix X^* that satisfies (4.6) and is a forest (since $[X_s]$ is unique).

THEOREM 7. *The following relation holds:*

$$\sum_{i,j} (x_{ij} - x_{ij}^*) d_{ij} \geq 1.$$

Proof. Let

$$[X_s] = [x_{i_1 j_1}, x_{i_1 j_2}, x_{i_2 j_2}, x_{i_2 j_3}, \cdots, x_{i_s j_s}, x_{i_s j_1}],$$

where

$$x_{i_1 j_2} = x_{kl}.$$

Then

$$\sum_{i,j} (x_{ij} - x_{ij}^*) d_{ij} = g(d_{i_1 j_1} - d_{i_1 j_2} + d_{i_2 j_2} - d_{i_2 j_3} + \cdots + d_{i_s j_s} - d_{i_s j_1})$$

$$= -g(d_{i_1 j_2} + u_{i_1} - v_{j_2}) \geq 1.$$

The theorem follows.

If X^* is a tree, then the whole process is repeated until at some stage a trial matrix is obtained that either (i) is a solution, or (ii) is not a solution and is a forest of $t > 1$ trees. We shall now discuss (ii).

9. The degenerate case. Let X be a trial matrix which is a forest of $t > 1$ trees. As we have seen, we may suppose that the rows and columns of X are ordered so that

$$X = \left\| \begin{matrix} X_{11} & 0 & \cdots & 0 \\ 0 & X_{22} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & X_{tt} \end{matrix} \right\|,$$

where each submatrix X_{aa} of order $m_a \times n_a$ is a tree. We can apply the methods of the nondegenerate case to the subproblems corresponding to the submatrices X_{aa}, and either obtain a solution to each subproblem or further decompose the matrix X; thus we may also assume that each X_{aa} is a solution to its subproblem.

By Corollary 5A, we know that

$$(9.1) \qquad d^{i\,j}_{a\,a} + \overline{u}^i_a - \overline{v}^j_a \geq 0 \qquad (a = 1, \cdots, t;\ i_a = 1, \cdots, m_a;\ j_a = 1, \cdots, n_a),$$

where $\overline{u}_a$, $\overline{v}_a$ is the preferred trial solution of the dual subproblem corresponding to the solution X_{aa}, and that

$$(9.2) \qquad d^{i\,j}_{aa} + \overline{u}^i_a - \overline{v}^j_a = 0 \quad \text{if} \quad x^{i\,j}_{aa} > 0.$$

We recall also that the most general values for u^i_a and v^j_a are given by

$$u^i_a = \overline{u}^i_a + z_a, \qquad v^j_a = \overline{v}^j_a + z_a,$$

where the z_a are arbitrary parameters.

It follows from Theorem 2 that X is a solution if and only if there are values of z_a that satisfy inequalities corresponding to (4.7), or in our present notation:

$$(9.3) \qquad d^{i\,j}_{ab} + u^i_a - v^j_b \geq 0 \qquad \text{for all } a,\ b,\ i_a, \text{ and } j_b.$$

But (9.3) has a solution for z_a if and only if the following inequalities have a solution for z_a:

$$(9.4) \qquad p_{ab} + z_a - z_b \geq 0,$$

where

$$\overline{p}_{ab}^{ij} = d_{ab}^{ij} + \overline{u}_a^i - \overline{v}_b^j, \quad p_{ab} = \min_{(i,j)} \{\overline{p}_{ab}^{ij}\}.$$

We have proved:

LEMMA A. *The matrix X is a solution if and only if there are real numbers z_a such that*

$$p_{ab} + z_a - z_b \geq 0 \qquad (a, b = 1, \cdots, t) .$$

In order to establish a criterion for the solvability of (9.4), we consider a special case of the original problem, defined as follows:

$$d_{ab} = p_{ab}, \; r_a = c_b = 1, \; a, b = 1, \cdots, t.$$

We call this the *special problem,* the corresponding dual the *special dual,* and now consider the *special combined problem*:

$$\sum_b y_{ab} = \sum_a y_{ab} = 1 \qquad (y_{ab} \geq 0),$$

$$p_{ab} + z_a - w_b \geq 0, \quad y_{ab}(p_{ab} + z_a - w_b) = 0.$$

If we set $y_{ab} = \delta_{ab}$, then for this trial solution the conditions reduce to:

$$p_{ab} + z_a - w_b \geq 0 \qquad \text{for } a \neq b,$$

$$p_{aa} + z_a - w_a = 0.$$

Since $p_{aa} = 0$, it follows that $z_a = w_a$, and so these conditions are equivalent to (9.4). Hence, by Theorem 2, (9.4) has a solution if and only if $\|\delta_{ab}\|$ is a solution of the special problem. Using Lemma A, we now have:

LEMMA B. *The matrix X is a solution of the original problem if and only if the identity matrix is a solution of the special problem.*

THEOREM 8. *The matrix X is a solution of the problem if and only if*

$$p_{a_1 a_2 \cdots a_h} \geq 0 \qquad (h = 2, 3, \cdots, t),$$

where $(a_1, a_2, \cdots, a_h)$ *is any permutation of h different positive integers, none greater than t, and*

$$p_{a_1 a_2 \cdots a_h} = p_{a_1 a_2} + p_{a_2 a_3} + \cdots + p_{a_h a_1}.$$

Proof. By Lemma B, it suffices to show that the condition of the theorem is equivalent to the statement that $||\delta_{ab}||$ is a solution of the special problem.

First of all, it is easy to see that at least one solution $Y = ||y_{ab}||$ of the special problem is a forest, and hence has less than $2t$ nonzero elements. That the elements of Y are all either zero or unity can be seen by induction as follows. The basis of the induction is obvious, and we consider the case $t+1$, assuming the statement for t. There must be at least one element of Y that is unity, as otherwise Y would have at least $2t$ nonzero elements. We may suppose that this element is $y_{t+1\, t+1}$. But then the induction hypothesis implies that each element y_{ab} $(a, b = 1, \cdots, t)$ is zero or one. It follows that there are exactly t elements of Y that are unity, whence we can write

$$\sum_{a,b} y_{ab}\, p_{ab} = p_{a_1 b_1} + p_{a_2 b_2} + \cdots \quad p_{a_t b_t},$$

where $(a_1\, a_2 \cdots a_t)$ and $(b_1\, b_2 \cdots b_t)$ are permutations of the first t integers. Then $||\delta_{ab}||$ is a solution of the special problem if and only if always

$$p_{a_1 b_1} + p_{a_2 b_2} + \cdots + p_{a_t b_t} \geq p_{11} + p_{22} + \cdots + p_{tt} = 0.$$

The proof is completed by noting that this sum can be written as

$$p_{a_1 a_2} + p_{a_2 a_3} + \cdots + p_{a_h a_1},$$

with $(a_1\, a_2 \cdots a_h)$ as described in the theorem.

We now need to show how to construct an improved trial solution X^* in the event that X is not a solution. In this case, we know from Theorem 8 that there is a sum

$$\bar{p}^{\,i^0 j^0}_{a_1 a_2} + \bar{p}^{\,i^0 j^0}_{a_2 a_3} + \cdots + \bar{p}^{\,i^0 j^0}_{a_h a_1} < 0.$$

Let I consist of all positive elements x^{ij}_{aa} together with all

$$x_{a_k a_{k+1}}^{i^0 j^0}$$

of X. Then we assert:

THEOREM 9. *There is a unique I-circuit on X that can be arranged to involve as even-terms all the*

$$x_{a_k a_{k+1}}^{i^0 j^0}.$$

Proof. The positive graph of X has $m+n-t$ vertices, $m+n-2t$ arcs, and t connected pieces. Also, for each

$$x_{a_k a_{k+1}}^{i^0 j^0}$$

there are nonzero elements

$$x_{a_k a_k}^{i^0 j^1}, \quad x_{a_{k+1} a_{k+1}}^{i^1 \quad j^0}.$$

Hence in passing from the positive graph to the I-graph, h vertices and $2h$ arcs are added, and the number of connected pieces is decreased from t to $t-h+1$. Thus the cyclomatic number of the I-graph is

$$\mu = (2h+m+n-2t) - (h+m+n-t) + (t-h+1) = 1,$$

so there is a unique I-circuit $[X_s]$ on X. Since the graph obtained by omitting from I any

$$x_{a_k a_{k+1}}^{i^0 j^0}$$

clearly has no cycle, $[X_s]$ contains all of these.

Evidently $[X_s]$ can be arranged, for example, as

$$\left[x_{a_1 a_1}^{i^0 j^1}, \; x_{a_1 a_2}^{i^0 j^0}, \; x_{a_2 a_2}^{i^1 \; j^0}, \cdots, x_{a_2 a_2}^{i^0 \; j^1}, \; x_{a_2 a_3}^{i^0 j^0}, \cdots\right]$$

so that all

$$x_{a_k a_{k+1}}^{i^0 j^0}$$

appear as even-terms.

As in the nondegenerate case, let g be the smallest odd-term in $[X_s]$ (hence $g > 0$), and define a new trial matrix X^* by replacing the elements of X that appear in $[X_s]$ by new ones increased by g for even-terms and decreased by g for odd-terms; the other elements of X are left unchanged. Again X^* satisfies the conditions for a trial matrix. To complete the discussion of the degenerate case, it remains only to prove:

THEOREM 10. *The following relation holds*:

$$\sum_{i,j} (x_{ij} - x_{ij}^*) d_{ij} \geq 1.$$

Proof. Since X and X^* differ only on

$$[X_s] = [x_{i_1 j_1},\ x_{i_1 j_2},\ x_{i_2 j_2},\ x_{i_2 j_3},\ \cdots,\ x_{i_s j_s},\ x_{i_s j_1}],$$

then

$$\sum_{i,j} (x_{ij} - x_{ij}^*) d_{ij} = -g(d_{i_1 j_1} - d_{i_1 j_2} + d_{i_2 j_2} - d_{i_2 j_3} + \cdots + d_{i_s j_s} - d_{i_s j_1}).$$

The proof is completed by noting that

$$d_{ij} = \overline{p}_{ij} + \overline{v}_j - \overline{u}_i \quad \text{and} \quad \overline{p}_{ij} = 0 \quad \text{if} \quad x_{ij} > 0,$$

so that

$$\sum_{i,j} (x_{ij} - x_{ij}^*) d_{ij} = -g(p_{a_1 a_2 \cdots a_h}) \geq 1.$$

REFERENCES

1. Frank L. Hitchcock, *The distribution of a product from several sources to numerous localities*, j. Math. Physics, 20 (1941), 224-230.

2. L. Kantorovitch, *On the translocation of masses*, Doklady Akad. Nauk SSSR, 37 (1942), 199-201.

3. Tjalling C. Koopmans, *Optimum utilization of the transportation system*, Econometrica, 17, Supplement (1949), 136-146.

4. *Activity analysis of production and allocation* (Proceedings of the Linear Programming Conference held at the University of Chicago, June, 1949, Tjalling C. Koopmans, Ed., John Wiley and Sons, Inc., New York (1951).

a. Tjalling C. Koopmans and Stanley Reiter, *A model of transportation* (pp. 222-259).

b. George B. Dantzig, *Application of the simplex method to a transportation problem* (pp. 359-373).

c. David Gale, Harold W. Kuhn, and Albert W. Tucker, *Linear programming and the theory of games* (pp. 317-329).

5. J. M. McLynn and C. B. Tompkins, *Application of a duality theorem in the calculation of some linear maximizing and minimizing problems*, Engineering Research Associates, Inc. (Unpublished).

6. Alex Orden, *A procedure for handling degeneracy in the transportation problem*, Headquarters U.S. Air Force, April, 1951 (Unpublished).

7. Julia Robinson, *A note on the Hitchcock-Koopmans problem*, The RAND Corporation, June 1950 (Unpublished).

8. ———, *On the Hamiltonian game—a traveling salesman problem*, The RAND Corporation, December 1949 (Unpublished).

9. Oswald Veblen, *Analysis Situs*, American Mathematical Society Colloquium Publications, New York (1931).

THE RAND CORPORATION

ON THE UNIQUE DETERMINATION OF SOLUTIONS OF THE HEAT EQUATION

W. FULKS

1. Introduction. Recently it has been shown independently by Hartman and Wintner [5] and by the present author [4] that if $u(x, t)$ has continuous derivatives u_{xx} and u_t, and is a nonnegative solution of the heat equation

$$(1) \qquad u_{xx}(x, t) - u_t(x, t) = 0$$

in a rectangle R: $\{0 < x < 1;\ 0 < t < k \leq \infty\}$, then $u(x, t)$ can be represented in the form

$$(2) \quad u(x, t) = \int_{0+}^{1-0} G(x, t; y, 0)\, dA(y) + \int_0^t G_y(x, t; 0, s)\, dB(s) - \int_0^t G_y(x, t; 1, s)\, dC(s),$$

where

$$(3) \qquad G(x, t; y, s) = \frac{1}{2}\left[\vartheta_3\left(\frac{x-y}{2}, t-s\right) - \vartheta_3\left(\frac{x+y}{2}, t-s\right)\right],$$

and where ϑ_3 is the Jacobi theta function. The integrals are Riemann-Stieltjes integrals with nondecreasing integrator functions, A, B, and C. The first integral may be improper but is absolutely convergent. It was further shown (see [5] and [3]) that

$$(4) \qquad u(x, 0+) = A'(x)$$

and

$$(5) \qquad u(0+, t) = B'(t-0);\ u(1-0, t) = C'(t-0)$$

at every point where the derivatives in question exist.

Received January 30, 1952. The preparation of this paper was sponsored, in part, by the Office of Naval Research, Contract Nonr-386(00).

Pacific J. Math. 3 (1953), 387-391

2. Theorem. As to the question of the extent to which (4) and (5) uniquely determine $u(x, t)$, it is clear that they do not do so completely, for the singular solution $G_y(x, t; 0, 0)$, called a heat explosion by Doetsch [2], has normal boundary values identically zero on the three boundaries $x = 0$, $x = 1$, and $t = 0$ of R. Yet A, B, C, through formula (2), do uniquely determine u; hence one might expect that by proper choice of the path of approach to the boundary, zero boundary values would assure the vanishing of u. In particular, because of the central role played by G and G_y in the representation (2), one might expect those paths to be the curves along which these functions become unbounded. This leads us to the following:

THEOREM. *Suppose*

(a) $u(x, t)$ *is a nonnegative solution of* (1) *in* R;

(b) u_{xx} *and* u_t *are continuous in* R;

(c) $u(x, 0+) = 0$ $\qquad (0 < x < 1)$;

(d) *for every* s $(0 \le s < k)$, $\lim u(x, t) = 0$ *as* (x, t) *tends to* $(0, s)$ *along some parabolic arc of the form* $t - s = ax^2$, $a > 0$, *and* $\lim u(x, t) = 0$ *as* (x, t) *tends to* $(1, s)$ *along some parabolic arc of the form* $t - s = a(x-1)^2$, $a > 0$.

Then $u(x, t) \equiv 0$ *in* R.

3. Proof. As we remarked in the first sentence, conditions (a) and (b) permit representation of u in the form (2). From the formula

$$\vartheta_3(x/2, t) = (\pi t)^{-1/2} \sum_{n=-\infty}^{\infty} \exp\left[\frac{-(x+2n)^2}{4t}\right], \tag{6}$$

which can be found in [2], it is easily seen that for $0 < x < 1$ the two latter integrals in formula (2) $\longrightarrow 0$ as $t \longrightarrow 0+$. Furthermore,

$$\int_{0+}^{1-0} G(x, t; y, 0)\, dA(y) = \int_{0+}^{\delta} G(x, t; y, 0)\, dA(y)$$
$$+ \int_{\delta}^{1-\delta} G(x, t; y, 0)\, dA(y) + \int_{1-\delta}^{1-0} G(x, t; y, 0)\, dA(y),$$

where $\delta < (1/2) \min[x, 1-x]$ and is taken so small that, given $\epsilon > 0$,

$$\left|\int_{0+}^{\delta} G(x, t; y, 0)\, dA(y)\right| < \epsilon \text{ and } \left|\int_{1-\delta}^{1-0} G(x, t; y, 0)\, dA(y)\right| < \epsilon$$

uniformly in t, for $0 < t \le t_0$ for some t_0. Possibility to do this is ensured by [5, Lemma 2, p. 385]. Now

$$\int_{\delta}^{1-\delta} G(x,\, t;\, y,\, 0)\; dA(y) = \int_{\delta}^{1-\delta} (4\pi t)^{-1/2} \exp\left[\frac{-(x-y)^2}{4t}\right] dA(y)$$

$$+ \int_{\delta}^{1-\delta} \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} (4\pi t)^{-1/2} \exp\left[\frac{-(x-y+2n)^2}{4t}\right] dA(y)$$

$$- \int_{\delta}^{1-\delta} \sum_{n=-\infty}^{\infty} (4\pi t)^{-1/2} \exp\left[\frac{-(x+y+2n)^2}{4t}\right] dA(y).$$

The two latter integrals are easily seen to vanish with t. Since also the left side of (2) $\longrightarrow 0$ as $t \longrightarrow 0$, it follows that, if $\delta' < \delta$,

$$\overline{\lim_{t\to 0+}} \int_{\delta'}^{1-\delta'} (4\pi t)^{-1/2} \exp\left[\frac{-(x-y)^2}{4t}\right] dA(y)$$

$$\leq \overline{\lim_{t\to 0+}} \int_{\delta}^{1-\delta} (4\pi t)^{-1/2} \exp\left[\frac{-(x-y)^2}{4t}\right] dA(y) \leq 2\epsilon.$$

Let $\epsilon \longrightarrow 0$ and obtain

$$\lim_{t\to 0+} \int_{\delta'}^{1-\delta'} (4\pi t)^{-1/2} \exp\left[\frac{-(y-x)^2}{4t}\right] dA(y) = 0.$$

By [6, Th. 7], we see that $A(y)$ is constant between δ', and $1-\delta'$. Let $\delta' \longrightarrow 0$. This ensures the vanishing of the first integral of (2).

Now let us turn to the boundary $x = 0$. Suppose that for some t_0 the boundary function $B(s)$ is not continuous. If σ is the jump (positive since $B(s)$ is increasing) in $B(s)$ at $s = t_0$, then for $t > t_0$, since $G_y(x,\, t;\, 0,\, s) \geq 0$ (see [5, p. 370]).

$$u(x,\, t) \geq \int_0^t G_y(x,\, t;\, 0,\, s)\; dB(s) \geq \sigma\, G_y(x,\, t;\, 0,\, t_0)$$

$$= \frac{1}{2}\, \sigma x \pi^{-1/2} (t-t_0)^{-3/2} \exp\left[\frac{-x^2}{4(t-t_0)}\right]$$

$$+ \frac{1}{2}\, \sigma \pi^{-1/2} (t-t_0)^{-3/2} \sum_{\substack{n=-\infty \\ n\neq 0}}^{\infty} (2n+x) \exp\left[\frac{-(2n+x)^2}{4(t-t_0)}\right].$$

Since $u(x, t) \longrightarrow 0$ as $(x, t) \longrightarrow (0, t_0)$ along $t - t_0 = ax^2$ for some $a > 0$, we have

$$u(x, t) \geq \frac{1}{2} \sigma \pi^{-1/2} x^{-2} a^{-3/2} \exp\left[\frac{-1}{4a}\right]$$

$$+ \frac{1}{2} \sigma \pi^{-1/2} a^{-3/2} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \frac{2n + x}{x^3} \exp\left[\frac{-(2n+x)^2}{4ax^2}\right],$$

As $x \longrightarrow 0+$, the sum clearly $\longrightarrow 0$; but

$$\lim_{(x,t)\to(0,t_0)} u(x, t) = 0 \geq \underline{\lim}_{x \to 0} \frac{1}{2} \sigma \pi^{-1/2} x^{-2} a^{-3/2} \exp\left[\frac{-1}{4a}\right] = \infty .$$

This is a contradiction. Hence $\sigma = 0$, and $B(s)$ is continuous for $0 \leq s < k$.

Now let $t = t_0 + ax^2$. Then

$$u(x, t) \geq \int_{t_0}^{t_0 + ax^2/2} G_y(x, t; 0, s)\, dB(s)$$

$$= \int_{t_0}^{t_0+ax^2/2} \frac{1}{2} x \pi^{-1/2} (t-s)^{-3/2} \exp\left[\frac{-x^2}{4(t-s)}\right] dB(s)$$

$$+ \int_{t_0}^{t_0+ax^2/2} \frac{1}{2} \pi^{-1/2} (t-s)^{-3/2} Q(x, t; s)\, dB(s),$$

where

$$Q(x, t; s) = \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} (2n + x) \exp\left[\frac{-(2n+x)^2}{4(t-s)}\right]$$

Clearly the latter integral vanishes with x, Since in the interval of integration we have

$$\exp\left[\frac{-x^2}{4(t-s)}\right] \geq \exp\left[\frac{-x^2}{4(ax^2/2)}\right] = \exp\left[\frac{-1}{2a}\right]$$

and

$$t - s \leq ax^2,$$

it follows that

$$u(x,t) \geq \frac{1}{2} \pi^{-1/2} a^{-3/2} x^{-2} \exp\left[\frac{-1}{2a}\right]\left[B\left(t_0 + \frac{ax^2}{2}\right) - B(t_0)\right] + o(1)$$

$$\geq K \frac{B(t_0 + ax^2/2) - B(t_0)}{ax^2/2} + o(1),$$

where K is a positive constant. Letting $x \longrightarrow 0$, we obtain

$$0 \geq \overline{\lim_{x \to 0}} \frac{B(t_0 + ax^2/2) - B(t_0)}{ax^2/2} = D^+[B(t_0)].$$

Hence, by [1, p.580], $B(s)$ is a monotone decreasing function. Since it is non-decreasing, it must be constant. Similarly it can be shown that $C(s)$ is constant. This completes the proof.

It seems probable that conditions (b), (c) and (d) would ensure the vanishing of $u(x,t)$ if it were represented by (2) with A, B, C of bounded variation, but the proof eludes the author.

References

1. C. Caratheodory, *Vorlesungen Über Reelle Funktionen*, Berlin, 1927.

2. G. Doetsch, *Theorie und Anwendung der Laplace-Transformation*, New York, 1943.

3. W. Fulks, *On the boundary values of solutions of the heat equation*, Pacific J. Math. 2 (1952), 141-145.

4. ———, *On integral representations and uniqueness of solutions of the heat equation*, University of Minnesota Thesis, June, 1949.

5. P. Hartman and A. Wintner, *On the solutions of the equation of heat conduction*, American J. Math. 72 (1950), 367-395.

6. D. V. Widder, *Positive temperatures on an infinite rod*, Trans. Amer. Math. Soc. 55 (1944), 85-95.

THE UNIVERSITY OF MINNESOTA

LENGTH AND AREA OF A CONVEX CURVE UNDER AFFINE TRANSFORMATION

JOHN W. GREEN

1. Introduction. We consider in the plane the class of all convex curves into which a given convex curve can be affinely transformed, and seek the minimum of L^2/A, where L denotes perimeter and A the area. This amounts to finding the minimum length for a fixed area, or, what is the same thing, to finding the minimum length under area-preserving affine transformations. In § 2 are found necessary conditions on the supporting function that a given curve yield the minimum of L^2/A, and in § 3 these are shown to be sufficient. In § 4 is derived a property of the minimizing curves; namely that if they are sufficiently smooth, they have at least six vertices. In § 5 is derived an integral representation of the supporting function of a convex curve, and another lemma to be used in § 6. In 6 we study the problem of finding the maximum, over all convex curves, of the minimum over affine transformations of L^2/A; in other words, we seek that curve of given area, which when affinely transformed so as to minimize its length, gives the greatest length. We show that the extreme curve is a polygon of not more than five sides, but fail to show what is extremely likely, that the solution is a triangle.

For general facts about convex figures and their supporting functions which are used, see [3].

2. Necessary conditions. Consider a convex curve K and its area-preserving affine transforms. Since rigid motions can be ignored, any transformation in which we are interested can be written in the form

$$(1) \qquad T: \begin{cases} x = e^{\lambda} x', \\ y = \mu x' + e^{-\lambda} y'. \end{cases}$$

The length $L(\lambda, \mu)$ of the transformed curve $K(\lambda, \mu)$ is a continuous function of λ and μ, and tends to ∞ as $(\lambda^2 + \mu^2)^{1/2}$ becomes large. Thus $L(\lambda, \mu)$ has a minimum value, which we take for the moment to be at $\lambda = \mu = 0$.

Received June 2, 1952.

Pacific J. Math. 3 (1953), 393-402

In order to find $L(\lambda, \mu)$ we need the supporting function $p(\lambda, \mu; \theta)$ of $K(\lambda, \mu)$. If $p(\theta) = p(0, 0, \theta)$ is the supporting function of K, then a supporting line to K is

$$x \cos \theta + y \sin \theta = p(\theta). \tag{2}$$

The transformation (1) carries (2) into

$$x'(e^{\lambda} \cos \theta + \mu \sin \theta) + y' e^{-\lambda} \sin \theta = p(\theta), \tag{3}$$

which is a supporting line to $K(\lambda, \mu)$.

To convert (3) into normal form we set

$$\begin{cases} e^{\lambda} \cos \theta + \mu \sin \theta = k \cos \phi, \\ \qquad\qquad e^{-\lambda} \sin \theta = k \sin \phi, \end{cases} \tag{4}$$

or

$$\begin{aligned} \cot \phi &= e^{2\lambda} \cot \theta + \mu e^{\lambda}, \\ k^2 &= (e^{\lambda} \cos \theta + \mu \sin \theta)^2 + e^{-2\lambda} \sin^2 \theta. \end{aligned} \tag{5}$$

The normal form of (3) is then

$$x' \cos \phi + y' \sin \phi = p(\theta)/k,$$

and so

$$p(\lambda, \mu, \phi) = p(\theta)/k.$$

From (5) and (4) we see that

$$\csc^2 \phi \, d\phi = e^{2\lambda} \csc^2 \theta \, d\theta, \; e^{2\lambda} k^2 \sin^2 \phi = \sin^2 \theta,$$

and so $d\phi = d\theta/k^2$. Thus[1]

$$L(\lambda, \mu) = \int p(\lambda, \mu, \phi) \, d\phi = \int p(\theta) \frac{d\theta}{k^3}. \tag{6}$$

Now let λ and μ be functions of a parameter t, with $\lambda(0) = \mu(0) = 0$. Then

$$L(\lambda(t), \mu(t)) = L(t),$$

and direct computation from (6) results in

[1] All integrals go from 0 to 2π unless otherwise noted.

$$(7) \qquad \frac{-L'(0)}{3} = \int p(\theta)\left\{\lambda_0' \cos 2\theta + \frac{1}{2}\mu_0' \sin 2\theta\right\} d\theta = 0.$$

Since λ_0' and μ_0' may be taken at pleasure, it is clear that in order for $t = 0$ to yield a minimum, we must have

$$(8) \qquad \int p(\theta) \cos 2\theta \, d\theta = \int p(\theta) \sin 2\theta \, d\theta = 0.$$

In other words, a necessary condition that K give a minimum length is that the second Fourier coefficients of p be zero.

3. Sufficiency. Suppose now that $\lambda = \mu = 0$ is a critical value of $L(\lambda, \mu)$, not necessarily the minimum. Then, as in § 2, we see that

$$\int p \cos 2\theta \, d\theta = \int p \sin 2\theta \, d\theta = 0.$$

Futher differentiation of (6), with the use of (8) and certain trigonometric identities, results in

$$(9)\ L''(0) = \frac{3}{2}\int p(\theta)\{x^2(1+5\cos 4\theta) + 10xy \sin 4\theta + y^2(1-5\cos 4\theta)\}d\theta,$$

where $x = \lambda_0'$, $2y = \mu_0'$. Setting

$$(10) \qquad K(\theta) = x^2\left(1 - \frac{1}{3}\cos 4\theta\right) - \frac{2}{3}xy \sin 4\theta + y^2\left(1 + \frac{1}{3}\cos 4\theta\right),$$

we may rewrite (9) as

$$(11) \qquad L''(0) = \frac{3}{2}\int p(\theta)\{K + K''\}\, d\theta.$$

Suppose now that p is twice differentiable, and integrate the K'' term in (11) by parts twice. We get

$$(12) \qquad L''(0) = \frac{3}{2}\int (p + p'')\, K \, d\theta.$$

The discriminant of the quadratic form (10) is equal to $-32/9$, and the form is positive definite. Let M be its minimum value for $x^2 + y^2 = 1$, and all θ. The quantity $p + p''$ is the radius of curvature, $ds/d\theta$, of K, and so

$$(13) \qquad L''(0) \geq \frac{3}{2}\int M\, ds = \frac{3}{2} ML.$$

If p is not twice differentiable, we approximate it uniformly by supporting functions which are. The right member of (9), for these approximating functions, is at least $3ML/2$, where L is computed for the approximating function; thus, passing to the limit, we see that (13) is satisfied in this case also.

Because of (13), we now see that if $\lambda = \mu = 0$ is a critical point for $L(\lambda, \mu)$, then it is a proper relative minimum. Consider now any transformation T_0, corresponding to parameters λ_0, μ_0, which yields a

$$K_0 = K(\lambda_0, \mu_0)$$

for which the second Fourier coefficients of the supporting function vanish. We may write T in the form $(TT_0^{-1})T_0$; that is, in studying the length of the transforms of K as function of T, we may study instead the length of the transforms of K_0 as function of TT_0^{-1}. We may write

$$TT_0^{-1}: \begin{cases} x = e^{(\lambda - \lambda_0)} x' & = e^{\xi} x', \\ y = (\mu e^{-\lambda_0} - \mu_0 e^{-\lambda}) x' + e^{-(\lambda - \lambda_0)} y' & = \eta x' + e^{-\xi} y', \end{cases}$$

where

$$(14) \qquad \begin{cases} \xi = \lambda - \lambda_0, \\ \eta = \mu e^{-\lambda_0} - \mu_0 e^{-\lambda}. \end{cases}$$

Now

$$L(\lambda, \mu) = \mathfrak{L}(\xi, \eta),$$

and, by the foregoing analysis, $\mathfrak{L}(\xi, \eta)$ has a proper relative minimum at $\xi = \eta = 0$. But the transformation (14) is nonsingular, and so $L(\lambda, \mu)$ has a proper relative minimum at λ_0, μ_0. Thus every critical point of $L(\lambda, \mu)$ is a proper relative minimum. But an (analytic) function in the plane which has only minima for critical points and which tends to ∞ at great distance can have only one critical point [6]. Thus $L(\lambda, \mu)$ has only one critical point, and this must be at the minimum.

THEOREM 1. *A necessary and sufficient condition that K have the least length of all curves into which it can be transformed by an area-preserving affine transformation is that*

$$\int p \cos 2\theta \, d\theta = \int p \sin 2\theta \, d\theta = 0.$$

Henceforth we shall refer to such K as *extreme curves*.

4. A six-vertex theorem. A vertex on a convex curve is a point where the radius of curvature has an extremum. It is a theorem of Kneser (see for example [1, p. 160]) that every convex curve, if sufficiently smooth, has at least four vertices.

THEOREM 2. *Each extreme curve with a continuous radius of curvature has at least six vertices.*[2]

The radius of curvature ρ is given in terms of the supporting function by $\rho = p + p''$. Now

$$\int \rho \cos\theta \, d\theta = \int \frac{ds}{d\theta} \cos\theta \, d\theta = \int \cos\theta \, ds = \oint dy = 0,$$

and similarly for $\int \rho \sin\theta \, d\theta$. Also

$$\int \rho \cos 2\theta \, d\theta = \int (p + p'') \cos 2\theta \, d\theta = 0,$$

by two integrations by parts. Thus we see that

$$\rho \sim \frac{L}{2\pi} + \sum_{3}^{\infty} (a_n \cos n\theta + b_n \sin n\theta). \tag{15}$$

It has been known since Liouville ([5, p. 264]) that (15) implies that $\rho - L/2\pi$ has at least six alternations in signs, and hence ρ six extrema.

In a very similar manner one can prove the following theorem, which however, will only be stated.

THEOREM 3. *Each extreme curve intersects a certain circle, of radius $L/2\pi$, at least six times.*

5. Some lemmas. If $H(\xi, \mu)$ is the Minkowski Stützfunktion of a convex curve, then

$$p(\theta) = H(\cos\theta, \sin\theta).$$

Now H is a convex function of ξ, η; $p(\theta)$ is not convex, but has the somewhat

[2] Blaschke [2] has already shown that a convex curve K may be affinely transformed until its radius of curvature is in the form (15), and thus that it has six vertices. However, the vanishing of the coefficients a_2 and b_2 was attained in an entirely different way. Namely, he found that ellipse K_1, of area equal to that of K, whose mixed volume with K is a minimum. Transforming affinely so that K_1 becomes a circle, we see that K becomes a curve satisfying (15). We have not been able to discover that Blaschke or others made any application of this result to the present problem.

analogous property of being sub-sine. A function $f(\theta)$ is sub-sine if, provided

$$f(\theta) = A \cos\theta + B \sin\theta \text{ at } \theta_1 \text{ and } \theta_2, \text{ where } \theta_1 < \theta_2 < \theta_1 + \pi,$$

then

$$f(\theta) \le A \cos\theta + B \sin\theta \text{ for } \theta_1 \le \theta \le \theta_2.$$

A necessary and sufficient condition [4] that a periodic function $p(\theta)$ be the supporting function of a convex curve is that it be sub-sine, or, if it is of class C'', that $p + p'' \geq 0$.

LEMMA 1. *A necessary and sufficient condition that a function $p(\theta)$ of period 2π be the supporting function of a convex curve is that it be expressible in the form*

$$(16) \qquad p(\theta) = \int_{\theta_0}^{\theta} \sin(\theta - t)\, d\alpha(t) + A \cos\theta + B \sin\theta,$$

where α is a nondecreasing function.

First let a supporting function $p \subset C''$; then

$$p + p'' = g(\theta) \geq 0.$$

The solution of the differential equation $p + p'' = g(\theta)$ is readily verified to be

$$(17) \qquad p(\theta) = \int_{\theta_0}^{\theta} \sin(\theta - t)\, g(t)\, dt + p(\theta_0) \cos(\theta - \theta_0) + p'(\theta_0) \sin(\theta - \theta_0),$$

which is of the form (16) with

$$\alpha(\theta) = \int_{\theta_0}^{\theta} g(t)\, dt.$$

Note that

$$\alpha(\theta_0) = 0 \text{ and } \alpha(\theta_0 + 2\pi) = \int (p + p'')\, d\theta = L.$$

Now if $p \not\subset C''$, it is the uniform limit of supporting functions p_n which are. We put each p_n in the representation (17), and apply the Helly selection theorem and the Bray-Helly theorem ([7, p. 29-31]) to obtain the result immediately. The factors $p_n'(\theta_0)$ offer no difficulty, since one easily shows that they are bounded for all n.

The converse is proved similarly. If a periodic p is given by (16), we can approximate α by a sequence of smooth monotone functions α_n which give periodic functions p_n; these p_n are sub-sine since they satisfy

$$p_n'' + p_n' = \alpha_n' \geq 0.$$

Again using the Bray-Helly theorem, we have that $p = \lim p_n$; that is, p is a limit of sub-sine functions, and so is sub-sine.

LEMMA 2. *If $p(\theta)$ is a supporting function, and if there exist at least six disjoint intervals in $0 \leq \theta \leq 2\pi$, interior to each of which p is not identically of the form $A \cos \theta + B \sin \theta$, then there exists a function $\eta(\theta)$ with the following properties:*

(a) *$p + \lambda\eta$ is a supporting function for small $|\lambda|$,*

(b) $\int \eta \, d\theta = \int \eta \cos 2\theta \, d\theta = \int \eta \sin 2\theta \, d\theta = 0$,

(c) $\eta \not\equiv A \cos \theta + B \sin \theta$.

Let I_j: $a_j < \theta < b_j$, $j = 1, 2, \cdots, 6$, be the disjoint intervals mentioned, and let p be given by (16). We may assume that $\alpha(\theta)$ is continuous at a_j and b_j. Define

$$\beta_j(\theta) = \begin{cases} \alpha(a_j) & \text{for } 0 \leq \theta < a_j, \\ \alpha(\theta) & \text{for } a_j \leq \theta < b_j, \\ \alpha(b_j) & \text{for } b_j \leq \theta \leq 2\pi. \end{cases} \tag{18}$$

while outside $(0, 2\pi)$ we make $d\beta_j$ periodic. Set

$$\beta = \sum \lambda_j \beta_j, \text{ where } |\lambda_j| \leq 1.$$

Then $\alpha(\theta) + \lambda\beta(\theta)$ is nondecreasing if $|\lambda| \leq 1$, as simple computation reveals. We set

$$\eta_j = \int_0^\theta \sin(\theta - t) \, d\beta_j(t) \text{ and } \eta = \sum \lambda_j \eta_j.$$

Then $p + \lambda\eta$ is of the form (16), with $\alpha + \lambda\beta$ in place of α. In order that η have period 2π, and thus that (a) be satisfied, we demand that

$$\sum \lambda_j \int \sin \theta \, d\beta_j(\theta) = \sum \lambda_j \int \cos \theta \, d\beta_j(\theta) = 0. \tag{19}$$

To satisfy conditions (b) of the lemma, we set

$$(20)\qquad \sum \lambda_j \int \eta_j \, d\theta = \sum \lambda_j \int \eta_j \cos 2\theta \, d\theta = \sum \lambda_i \int \eta_i \sin 2\theta \, d\theta = 0.$$

Equations (19) and (20) comprise five homogeneous equations in the six unknowns λ_j. They always have a nontrivial solution, which we employ for the construction of β. If $\lambda_k \neq 0$, then η is equal in I_k to a nonzero multiple of $p(\theta)$, plus sine and cosine terms, and this by hypothesis is not of the form $A \cos \theta + B \sin \theta$. Thus (c) is satisfied, and the lemma is proved.

6. The minimax problem. We now restrict our attention to extreme curves, and seek the maximum m of L^2/A. A crude estimate of m can be obtained as follows. If K is any convex curve of area 1, inscribe in K a triangle Δ of maximum area, $A(\Delta)$. Then at each vertex of Δ, K must have a supporting line parallel to the opposite side of Δ, and these three supporting lines form a triangle Δ_1. Transform the plane in an area-preserving affine way so that Δ and Δ_1 are carried into equilateral triangles Δ' and Δ_1', and K into K'. The perimeter $L(\Delta')$ of Δ' is given by

$$L(\Delta') = 6\sqrt{A(\Delta')/\sqrt{3}}.$$

Then

$$L(K') \leq L(\Delta_1') = 2L(\Delta') = 12\sqrt{A(\Delta')/\sqrt{3}} \leq 12/\sqrt[4]{3}.$$

Thus for the transform K' of K, we have

$$L^2/A \leq 48\sqrt{3}, \text{ and so } m \leq 48\sqrt{3}.$$

On the other hand, the equilateral triangle gives

$$L^2/A = 12\sqrt{3}, \text{ and so } m \geq 12\sqrt{3}.$$

We now normalize our problem by considering extreme curves of length 1, and try to minimize the area. By the usual compactness argument ([2, p. 62]), there does exist a minimizing curve K. Let p be the supporting function of K. Suppose there exists a function $\eta(\theta)$ satisfying conditions (a), (b) of Lemma 2. Consider the area $A(\lambda)$ of the extreme curve, of unit length, whose supporting function is $p + \lambda\eta$. We have

$$(21)\qquad 2A(\lambda) = \int \{(p+\lambda\eta)^2 - (p'+\lambda\eta')^2\}\, d\theta$$
$$= 2A(0) + 2\lambda \int (p\eta - p'\eta')\, d\theta + \lambda^2 \int (\eta^2 - \eta'^2)\, d\theta.$$

Because of the extreme nature of K, the term $\int(p\eta - p'\eta')\, d\theta = 0$. Because of conditions (b) of Lemma 2, the Fourier series of η will be as follows.

$$\eta = a_1 \cos\theta + b_1 \sin\theta + \sum_3^\infty (a_j \cos j\theta + b_j \sin j\theta),$$

and by Parseval's relation,

$$\frac{1}{\pi}\int \eta^2\, d\theta = (a_1^2 + b_1^2) + \sum_3^\infty (a_i^2 + b_i^2).$$

Similarly (η' being bounded),

$$\frac{1}{\pi}\int \eta'^2\, d\theta = (a_1^2 + b_1^2) + \sum_3^\infty j^2 (a_i^2 + b_i^2),$$

and so

$$(22)\qquad \int (\eta^2 - \eta'^2)\, d\theta = \pi \sum_3^\infty (1 - j^2)(a_i^2 + b_i^2).$$

Since $A(\lambda) \geq A(0)$, we see from (21) and (22) that $a_j = b_j = 0$ for $j \geq 2$, so that $\eta \equiv a_1 \cos\theta + b_1 \sin\theta$. Thus it is not possible to satisfy (a), (b), and (c) simultaneously.

Now if K is a polygon, p is piecewise of the form $A\cos\theta + B\sin\theta$, and conversely. If K is not a polygon it is clear that one can find as many intervals as desired in each of which p is not of that form, and Lemma 2 applies. Lemma 2 also applies if K is a polygon of six or more sides. Thus it is not possible for K to be other than a polygon of five or fewer sides.

It appears very likely that K is an equilateral triangle and that $m = 12\sqrt{3}$. To eliminate the cases of four and five sides is just a problem in the calculus, but possibly a very difficult one. In these cases there are not enough sides to construct the variations used above, which consist of sliding sides in and out parallel to themselves, so if a variational method is to be used, a different kind of variation, involving changing the angles, must be found.

REFERENCES

1. W. Blaschke, *Kreis und Kugel*, Leipzig, 1916.

2. ———, *Eine Minimumaufgabe über Eilinien*, Christiaan Huygens, Internationaal Math. Tijdschrift, 2 (1922-23) 150-154.

3. T. Bonnesen und W. Fenchel, *Theorie der Konvexen Köper*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 3, no. 1, Berlin, 1934.

4. J. W. Green, *Sets subtending a constant angle on a circle*, Duke Math. J. **17** (1950) 263-267.

5. J. Liouville, *Sur le developpement des fonctions, etc.*, J. Math. Pures Appl. **1**, (1836), 253-265.

6. M. Morse, *The calculus of variations in the large*, Colloquium Publications of the Amer. Math. Soc., New York, 1934.

7. D. V. Widder, *The Laplace transform*, Princeton University Press, 1941.

UNIVERSITY OF CALIFORNIA, LOS ANGELES, AND THE
INSTITUTE FOR ADVANCED STUDY.

AN ISOPERIMETRIC MINIMAX

WILLIAM GUSTIN

Introduction. In the preceding paper J. W. Green considers for a given convex body K in the euclidean plane the minimum of the isoperimetric ratio r (ratio of squared perimeter l^2 to area a) taken over all affine transforms k of K. He then investigates the maximum value taken over all K of this minimum ratio, shows by variational methods that such a maximum is attained by some polygon of five or fewer sides, and conjectures that it is, in fact, attained by a triangle with $12\sqrt{3}$, the isoperimetric ratio of an equilateral triangle, as the minimax ratio. I shall prove this conjecture directly by refining an estimation used by Green, the precise statement of results being as follows:

I. *Let K be an nontriangular plane convex body; there then exists an affine transform k of K with $r(k) < 12\sqrt{3}$.*

II. *Let T be a nonequilateral triangle; then $r(T) > 12\sqrt{3}$.*

Before taking up the proof of these results we dispose of a lemma.

III. *Let k be a possibly degenerate convex body with $s \subset k \subset t$, wherein t is an equilateral triangle, and s a side of t; there then exists a number x with $0 \leqq x \leqq 1$ such that*

$$l(k) \leqq (2/3 + x/3)\ l(t)$$

$$a(k) \geqq x\ a(t),$$

simultaneous equality occurring if and only if either $x = 0$, $k = s$ or $x = 1$, $k = t$.

Proof of III. Let p be that supporting strip of k parallel to the line-segment s; and let x be the ratio of the width of p to the width or altitude of t. Thus $0 \leqq x \leqq 1$, with $x = 0$ or $x = 1$ according as $k = s$ or $k = t$. Choose a point at which k touches the side of p opposite s, and define k_* to be the triangle with this point as apex and s as base. Define k^* to be the trapezoid formed by intersection of p and t. Clearly $s \subset k_* \subset k \subset k^* \subset t$; and $k_* = k = k^*$ if and only if $k = s$ or $k = t$.

Received October 16, 1952.
Pacific J. Math. 3 (1953), 403-405

Since $k \supset k_*$, it follows that $a(k) \geqq a(k_*)$, with equality if and only if $k = k_*$. And since $k \subset k^*$, it follows that $l(k) \leqq l(k^*)$ with equality if and only if $k = k^*$. These inequalities become, upon the easy computation of $a(k_*)$ and $l(k^*)$, the asserted inequalities of III.

Proof of I. Let K be the given nontriangular convex body. Since the area functional is continuous, it easily follows from a compactness argument that a triangle T of maximal area can be inscribed in K. Let the three sides of T be labelled S_i $(i = 1, 2, 3)$, and let V_i be that vertex of T opposite S_i. Because the area of T is maximal, the line L_i through V_i and parallel to S_i is a line of support of K. The triangle formed by the three lines L_i then circumscribes K and also T; it is composed of four nonoverlapping congruent triangles T and T_i, where T_i is labelled so as to have S_i as a side. That part K_i of K in T_i is a possibly degenerate convex body with $S_i \subset K_i \subset T_i$. Now any triangle can be affinely transformed into any other triangle. In particular, T can be affinely transformed into an equilateral triangle t, with T_i going into t_i, S_i into s_i, K_i into k_i, and K into k. Therefore $s_i \subset k_i \subset t_i$, and t_i is congruent to t. According to III, ratios x_i exist giving inequalities on $l(k_i)$ and $a(k_i)$. Furthermore, since K and hence k is nontriangular, not all $x_i = 0$ and not all $x_i = 1$. Therefore $0 < x < 1$, where $x = \sum x_i/3$. Evidently k is composed of the four nonoverlapping sets t and k_i in such a way that

$$l(k) = \sum l(k_i) - l(t) \leqq (1 + x)\, l(t),$$

$$a(k) = \sum a(k_i) + a(t) \geqq (1 + 3x)\, a(t),$$

whereupon

$$r(k) \leqq \frac{(1+x)^2}{1+3x}\, r(t) = \left[1 - \frac{x(1-x)}{1+3x}\right] 12\sqrt{3} < 12\sqrt{3},$$

as was to be shown.

Proof of II. Through II is merely a matter of trigonometry, and very likely can be verified by exhibiting a neat but perhaps unperspicuous trigonometric identity, I shall here prove it by the sort of methods used above.

Let T be a nonequilateral triangle. Define S_i, V_i, L_i as above. Since T is nonequilateral, some two of its sides, say S_1 and S_2, are unequal. Let v_3 be that point on the line L_3, regarded as a linear mirror, at which $v_1 = V_1$ is reflected when viewed from $v_2 = V_2$; and let t be the so symmetrized isosceles triangle with vertices v_i and sides s_i. Then the path $s_1 s_2$ is shorter than $S_1 S_2$, so $l(t) < l(T)$; and, since both triangles have the same base and altitude, $a(t) = a(T)$. Therefore $r(t) < r(T)$. Consequently if the minimum isoperimetric ratio among triangles is attained, it is attained by an equilateral triangle only; whereupon it would follow that $r(T) > 12\sqrt{3}$, as was to be shown. Now all possible triangle isoperimetric ratios are realized by triangles of fixed perimeter containing a fixed point. By a compactness argument, some such triangle achieves a maximum area and hence a minimum isoperimetric ratio. This completes the proof.

INDIANA UNIVERSITY

SOME HAUSDORFF MEANS WHICH EXHIBIT THE GIBBS' PHENOMENON

ARTHUR E. LIVINGSTON

1. Introduction. The regular Hausdorff mean of order n with kernel $g(x)$ for the sequence (s_k) is defined by

$$h_n = h_{n,\,g} = \sum_{k=0}^{n} \binom{n}{k} s_k \int_0^1 t^k (1-t)^{n-k}\, dg(t),$$

where $g(x)$ is of bounded variation on the interval $0 \le x \le 1$, $g(1) - g(0) = 1$, and $g(0+) = g(0)$. The integral in the definition being a Stieltjes integral, it is clear that $g(0)$ may be taken to be zero.

For the sequence

$$s_n(x) = \sum_{k=1}^{n} \frac{\sin kx}{k},$$

Otto Szász [3] has proved the following result: If, as $n \longrightarrow \infty$, $x_n \longrightarrow 0+$ and $nx_n \longrightarrow A \le \infty$, then

$$h_{n,g}(x_n) \longrightarrow \int_0^1 \mathrm{Si}(Ax)\, dg(x),$$

where

$$\mathrm{Si}(x) = \int_0^x \frac{\sin t}{t}\, dt.$$

He defines the Gibbs' ratio for the kernel $g(x)$ to be

$$F(g) = \max_{A>0} \frac{2}{\pi} \int_0^1 \mathrm{Si}(Ax)\, dg(x).$$

Received March 14, 1952. The work in this paper was done during the period when the author held Atomic Energy Commission Predoctoral Fellowship OR-21478.

Pacific J. Math. 3 (1953), 407-415

If $F(g) > 1$, then the sequence $\{h_{n,g}(x)\}$ exhibits the Gibbs' phenomenon on the right at $x = 0$.

It is here proved that (1) if $\alpha(x)$ is a regular Hausdorff step-function kernel whose points of jump are linearly independent over the rationals, then $F(\alpha) > 1$; (2) if $\alpha(x)$ is regular and has precisely two jumps, then $F(\alpha) > 1$. It seems reasonable that the first result is true without the hypothesis of linear independence, but the author has been unable to show this.

The Euler method of summability (ϵ, p), $0 < p \leq 1$, is a regular Hausdorff method having for its kernel the one-step function $\epsilon_p(x)$ which vanishes for $0 \leq x < p$, and has the value one for $p \leq x \leq 1$; the method (ϵ, p) is ordinarily denoted by $(E, (1-p)/p)$. Clearly,

$$F(\epsilon_p) = \frac{2}{\pi} \operatorname{Si}(\pi) > 1 \qquad (0 < p \leq 1),$$

so that the one-step case of (1) above follows trivially (this was shown by Szász [2, 3]).

2. Notation. It is convenient to collect here some notations which will be used throughout this paper.

(a) $\alpha(x)$ is a step-function defined as follows:

$$\begin{aligned} \alpha(x) &= a_1 = 0 && \text{for } 0 \leq x < \beta_1, \\ &= a_k && \text{for } \beta_{k-1} \leq x < \beta_k \text{ and } k = 2, \cdots, N, \\ &= a_{N+1} = 1 && \text{for } \beta_N \leq x \leq 1, \end{aligned}$$

where $a_k \neq a_{k+1}$ for $k = 1, \cdots, N$;

(b) $\operatorname{Si}(x) = \int_0^x \frac{\sin t}{t}\, dt;$

(c) $\operatorname{si}(x) = \operatorname{Si}(x) - \frac{1}{2}\pi = \int_\infty^x \frac{\sin t}{t}\, dt;$

(d) $f(x) = f_\alpha(x) = \frac{2}{\pi} \int_0^1 \operatorname{Si}(xy)\, d\alpha(y) = \frac{2}{\pi} \sum_{k=1}^{N} A_k \operatorname{Si}(x\beta_k),$

where $A_k = a_{k+1} - a_k$;

(e) $F(\alpha) = \max\limits_{x > 0} f_\alpha(x).$

It is clear that it is no restriction to assume that all regular step-function kernels are of the form (a).

3. The zeros of si(x). It is well known that $\mathrm{si}[(2n+1)\pi] > 0$ and $\mathrm{si}(2n\pi) < 0$ for $n = 0, 1, \cdots$, and that $\mathrm{si}(x)$ has precisely one zero, call it z_n, in each interval $n\pi < x < (n+1)\pi$ $(n = 0, 1, \cdots)$. It is intuitively clear and easy to prove rigorously that

$$z_n - \left(n + \frac{1}{2}\right)\pi > 0.$$

It will be shown in this section that even more is true, namely, that

$$z_n - \left(n + \frac{1}{2}\right)\pi \downarrow 0.$$

The tables [4] for the sine integral show that

$$1.9264 < z_0 < 1.9265 \text{ and } 4.893 < z_1 < 4.894.$$

It therefore follows that the following statement is true:

THEOREM 3.1. *The function* $\mathrm{si}(x)$ *is positive whenever*

$$-1.2150 < x - (2n+1)\pi < \frac{1}{2}\pi,$$

and is negative whenever

$$x \geq 0 \text{ and } -1.389 < x - 2n\pi < \frac{1}{2}\pi, \qquad (n = 0, 1, \cdots).$$

This result is needed in § 5.

It will now be shown that the zeros modulo π of $\mathrm{si}(x)$ form a strictly decreasing sequence with limit $\pi/2$. The formal statement is:

THEOREM 3.2. *Let* $(n + 1/2 + x_n)\pi$ *be the zero of* $\int_x^\infty u^{-1} \sin u \, du$ *in the interval*

$$n\pi < x < (n+1)\pi \qquad (n = 0, 1, \cdots).$$

Then the sequence (x_n) *is strictly decreasing with limit zero.*

(The first two paragraphs of the following proof are due to Harry Pollard, the fourth to the referee. Both Pollard and the referee point out that the relation

$$\frac{d}{dx}\,[F'(x)/F(x)] > 0$$

of the fourth paragraph can be deduced from general theorems on completely monotonic functions [5, pp. 144, 145, 167]. I. I. Hirschman, Jr., has observed that the zeros modulo π in the interval $0 < x < \infty$ of $\int_x^\infty g(u) \sin u\, du$ are monotone decreasing for any $g(u)$ which is completely monotonic on $0 < u < \infty$).

Proof. Let

$$F(x) = \int_0^\infty e^{-xu}(1+u^2)^{-1}\, du \quad \text{for } x > 0.$$

Then

$$(1) \qquad \int_x^a u^{-1} \sin u\, du = [F(u) \cos u - F'(u) \sin u]_x^a$$

for $a > 0$. To prove this, let $L(x)$ and $R(x)$ denote, respectively, the left and right sides of (1). Since $L(a) = R(a)$, it is sufficient to show that $L'(x) = R'(x)$ for $x > 0$. But this is immediate, for

$$L'(x) = -x^{-1} \sin x,$$

$$R'(x) = -\sin x\,[F(x) + F''(x)] = -\sin x \int_0^\infty e^{-xu}\, du.$$

Now taking the limit in (1) as $a \longrightarrow \infty$ gives

$$(2) \qquad -\int_x^\infty u^{-1} \sin u\, du = F(x) \cos x - F'(x) \sin x,$$

for $F(\infty) = F'(\infty) = 0$.

Since $F(x) > 0$ and $F'(x) < 0$, it follows from (2) that the finite zeros of $\int_x^\infty u^{-1} \sin u\, du$ occur at the points where

$$\frac{F'(x)}{F(x)} = \cot x.$$

Therefore, to complete the proof of the theorem, it is sufficient to show that

$F'(x)/F(x)$ is strictly increasing to zero as $x \longrightarrow \infty$.

Employing the usual derivative notation, one has

$$-(-x)^{n+1}F^{(n)}(x) = x^{n+1}\int_0^\infty \frac{u^n e^{-xu}}{1+u^2}\,du = \int_0^\infty \frac{u^n e^{-u}}{1+(u/x)^2}\,du,$$

so that

$$-(-x)^{n+1}F^{(n)}(x) \longrightarrow n! \quad \text{as} \quad x \longrightarrow \infty.$$

Therefore,

$$\frac{F'(x)}{F(x)} = x^{-1}\left[\frac{x^2F'(x)}{xF(x)}\right] \longrightarrow 0 \quad \text{as } x \longrightarrow \infty.$$

All that remains to be shown, then, is that $F'(x)/F(x)$ is strictly increasing, and this will follow if

$$\frac{d}{dx}\left[\frac{F'(x)}{F(x)}\right] > 0$$

or, equivalently, if

$$[F'(x)]^2 - F(x)F''(x) < 0.$$

Now

$$F(x) - 2F'(x)y + F''(x)y^2 = \int_0^\infty \frac{e^{-xu}}{1+u^2}(1+yu)^2\,du > 0,$$

so that the discriminant of the quadratic expression in y on the left must be negative. Since this discriminant is $[F'(x)]^2 - F(x)F''(x)$, the proof is complete.

4. The main theorem.

Two lemmas are needed.

LEMMA 4.1. *If* $0 < a_k < 1$ *for* $k = 1, \cdots, n$, *and* $a_1, \cdots, a_n, 1$ *are linearly independent over the rationals, then, given* $\epsilon > 0$, *there exist odd positive integers* $x, l_1, \cdots, l_m$, $m \leq n$, *and there exist even positive integers* $l_{m+1}, \cdots, l_n$, *such that* $0 < xa_k - l_k < \epsilon$ *for* $k = 1, \cdots, n$.

Proof. If Red u denotes the fractional part of u, then it is known that the vectors (Red $ja_1, \cdots$, Red ja_n), $j = 0, 1, \cdots$, are dense in the n-dimensional unit-cube [1, p. 83]. Hence there is a positive integer j such that

$$\frac{1}{2}(1 - a_k) < \text{Red } ja_k < \min\left(\frac{1 - a_k + \epsilon}{2}, 1\right) \qquad (k = 1, \cdots, m),$$

$$\frac{1}{2}(2 - a_k) < \text{Red } ja_k < \min\left(\frac{2 - a_k + \epsilon}{2}, 1\right) \qquad (k = m + 1, \cdots, n).$$

The conclusion of the lemma is satisfied by taking

$$x = 2j + 1,\ l_k = 2(ja_k - \text{Red } ja_k) + 1 \quad \text{for} \quad k = 1, \cdots, m,$$

and

$$l_k = 2(ja_k - \text{Red } ja_k + 1) \quad \text{for } k = m + 1, \cdots, n.$$

LEMMA 4.2. *Let* $\alpha(x)$ *be defined as in* $2(a)$. *If* $\beta_1, \cdots, \beta_N$, 1 *are linearly independent over the rationals, then* $F(\alpha) > 1$.

Proof. Let P, Q be the sets of positive integers $k \leq N$ for which $A_k > 0$, $A_k < 0$, respectively. Then

$$f(x) = \frac{2}{\pi}\left(\sum_{k \in P} + \sum_{k \in Q}\right) A_k \text{ Si}(x\beta_k).$$

By hypothesis, $0 < \beta_k < 1$ for $k = 1, \cdots, N$. Therefore, Lemma 4.1, with $\epsilon = 1/2$, asserts the existence of a positive x_0 and nonnegative integers n_k such that

$$0 < \pi x_0 \beta_k - (2n_k + 1)\pi < \frac{1}{2}\pi \quad \text{for } k \in P$$

and

$$0 < \pi x_0 \beta_k - 2(n_k + 1)\pi < \frac{1}{2}\pi \quad \text{for } k \in Q.$$

By Theorem 3.1, $\text{si}(\pi x_0 \beta_k) > 0$ for $k \in P$ and $\text{si}(\pi x_0 \beta_k) < 0$ for $k \in Q$. Recalling that $\sum A_k = 1$, one obtains that $f(\pi x_0) > 1$, which is sufficient.

Since

$$\lim_{A \to \infty} \mathrm{Si}(Ax) = \frac{1}{2}\pi \ \text{sign } x$$

boundedly, it follows that $F(g) \geq 1$ for every regular Hausdorff kernel.

Let now $\alpha(x)$ be a regular N-jump Hausdorff kernel. It will be shown that if $F(\alpha) = 1$, then $\beta_1, \cdots, \beta_N$ are linearly dependent over the rationals, and this will prove:

THEOREM 4.1. *If $\alpha(x)$ is defined as in 2(a) with $\beta_1, \cdots, \beta_N$ linearly independent over the rationals, then $F(\alpha) > 1$.*

Proof. Let $\beta = (\beta_1, \cdots, \beta_N)$ and $r = (r_1, \cdots, r_N)$, r_k rational. Set

$$|\beta| = \max_{1 \leq k \leq N} \beta_k,$$

and let x be a scalar such that $0 < x < |\beta|^{-1}$. Let Λ be the zero N-tuple. The inner product of N-tuples A and B is defined in the usual way and is denoted by $(A \mid B)$. Let

$$\alpha^x(t) = 1 \quad \text{for } x\beta_N \leq t \leq 1$$

and $\alpha^x(t) = \alpha(xt)$ otherwise. Then α^x is also a regular N-jump Hausdorff kernel, and $F(\alpha^x) = F(\alpha)$.

Suppose now that $F(\alpha) = 1$. According to Lemma 4.2, there corresponds to each x in the interval $0 < x < |\beta|^{-1}$ an $r_x \neq \Lambda$ and a rational number R_x such that

$$(x\beta \mid r_x) = R_x.$$

But the available r_x, R_x are countable while the permissible x are uncountable. Hence, there is an uncountable set X of x associated with an $r \neq \Lambda$ and a rational R. If $x, x' \in X$, then

$$(x - x')(\beta \mid r) = 0.$$

Taking $x \neq x'$ gives $(\beta \mid r) = 0$; that is, $\beta_1, \cdots, \beta_N$ are linearly dependent over the rationals.

5. The two-step case. The theorem to be proved is:

THEOREM 5.1. *If $\alpha(x)$ is a regular two-jump Hausdorff kernel, then $F(\alpha) > 1$.*

Proof. If β_1 and β_2 are linearly independent over the rationals, then Theorem 4.1 gives the result.

If $\alpha(x)$ is not an increasing function, then either $A_1 > 1$ and $A_2 < 0$ or $A_1 < 0$ and $A_2 > 1$. Suppose that it is the first. Recalling that $A_1 + A_2 = 1$, one obtains

$$f(x) = \frac{2}{\pi}\,\mathrm{Si}(x\beta_1) - \frac{2}{\pi} A_2[\mathrm{Si}(x\beta_1) - \mathrm{Si}(x\beta_2)].$$

Since $A_2 < 0$, and $\mathrm{Si}(\pi)$ is the absolute maximum of $\mathrm{Si}(x)$, it follows that

$$f(\pi/\beta_1) \geq \frac{2}{\pi}\,\mathrm{Si}(\pi) > 1.$$

The remaining two-jump kernels are those which are increasing and for which

$$\frac{\beta_2}{\beta_1} = \frac{p}{q},$$

with p and q integral and $(p, q) = 1$. If p and q are odd, there is no problem, for then $f(\pi q/\beta_1) > 1$. Otherwise, one of p, q is odd and the other even. To treat this situation, the following lemma, whose proof offers no difficulty, is useful:

LEMMA 5.1. *Let $0 < b_1 < b_2 \leq 1$. If l_1 and l_2 are odd positive integers such that*

$$|l_1 b_2 - l_2 b_1| < \frac{\epsilon}{\pi}(b_1 + b_2),$$

then there exists a positive number x such that

$$|xb_k - \pi l_k| < \epsilon \text{ for } k = 1, 2.$$

By Theorem 3.1, the proof of Theorem 5.1 will be complete if a positive x and odd positive integers l_1 and l_2 exist such that

$$|x\beta_k - \pi l_k| < 1.215 \text{ for } k = 1, 2.$$

By the above lemma, then, one wishes to find odd positive integers $l_1 = 2i + 1$ and $l_2 = 2j + 1$ such that

$$|pl_1 - ql_2| = |2pi - 2qj + p - q| < \frac{1.215}{\pi}(p + q).$$

Since p and q have unlike parity, $p + q \geq 3$. It will therefore be sufficient to find nonnegative integers i and j such that $2pi - 2qj + p - q = 1$.

If $p - q = 1$, simply take $i = q$ and $j = p$.

If $p - q \geq 3$, then the Diophantine equation

$$pi - qj = \frac{1}{2}(1 - p + q)$$

makes sense and, furthermore, has positive solutions i and j.

6. Remark. According to Theorem 3.2, the zeros modulo π of $\mathrm{si}(x)$ tend to $\pi/2$. Therefore, the method of proof used in this paper can not be expected to handle all step-function kernels omitted by Theorem 4.1.

References

1. J. F. Koksma, *Diophantische Approximationen*, Chelsea, New York.

2. Otto Szász, *On the Gibbs' phenomenon for Euler means*, Acta Sci. Math. Szeged 12 Part B (1950), 107-111.

3. ———, *Gibbs' phenomenon for Hausdorff means*, Trans. Amer. Math. Soc. 69 (1950), 440-456.

4. *Tables of Sine, Cosine, and Exponential Integrals*, vol. II, Federal Works Agency, Works Progress Administration for the City of New York, 1940.

5. D. V. Widder, *The Laplace Transform*, Princeton, 1941.

University of Oregon

ON GENERATION OF SOLUTIONS OF THE BIHARMONIC EQUATION IN THE PLANE BY CONFORMAL MAPPINGS

CHARLES LOEWNER

Introduction. The study of harmonic functions in the plane is essentially facilitated by the invariance of the Laplace equation

$$\nabla^2 u \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

under the group of conformal mappings. The transformations leaving the biharmonic equation

$$\nabla^4 u \equiv \frac{\partial^4 u}{\partial x^4} + 2\,\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = 0 \tag{1}$$

invariant are much more restricted; they only form the group of similarity transformations in the (x, y)-plane. On the other hand, more general transformations leaving the biharmonic equation invariant may be obtained if u is not treated as a scalar which does not change its value under the transformations, but transformations of the more general type

$$\begin{aligned} x' &= \phi(x, y) \\ y' &= \psi(x, y) \\ u' &= \chi(x, y)\, u \end{aligned} \tag{2}$$

are permitted. We assume the functions ϕ, ψ, and χ to be four times continuously differentiable, and $\chi \neq 0$. That such nontrivial transformations exist follows immediately from the well-known representation of a biharmonic function u in the form

---

Received June 23, 1952. This paper is a partly simplified version of work done under sponsorship of the Office of Naval Research. See the Technical Report with the same title, No. 13, N6ori-106, Task V (NR-043-992), with two notes by G. Szegö.

Pacific J. Math. 3 (1953), 417-436

$$u = h_1(x, y)r^2 + h_2(x, y), \tag{3}$$

with suitable harmonic functions h_1 and h_2, and $r^2 = x^2 + y^2$. If a transformation by reciprocal radii is applied, which in polar coordinates is given by

$$r' = \frac{1}{r}, \quad \theta' = \theta, \tag{4a}$$

and u is transformed according to the formula

$$u' = \frac{1}{r^2} u, \tag{4b}$$

then u' becomes

$$u' = h_1'(x', y') + h_2'(x', y')r'^2, \tag{3a}$$

with h_1' and h_2' being the harmonic functions of x', y' obtained from h_1 and h_2 by the transformation (4a). This shows that u' is biharmonic in x' and y'.

By combination of the transformation obtained with arbitrary similarities, more general transformation of type (2) may be obtained. In order to write them in a simple way we set

$$z = x + iy, \quad z' = x' + iy'.$$

One sees easily that the composed transformations can be written in one of the following forms:

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta}, \quad u' = k \left| \frac{dz'}{dz} \right| u; \tag{5'}$$

$$z' = \frac{\alpha \bar{z} + \beta}{\gamma \bar{z} + \delta}, \quad u' = k \left| \frac{dz'}{d\bar{z}} \right| u. \tag{5''}$$

The constants α, β, γ, δ, k are only subjected to the conditions

$$\begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} \neq 0, \quad k \neq 0, \quad k \text{ real}.$$

Each Moebius transformation in the (x, y)-plane may, therefore, be extended to a transformation in the (x, y, u)-space leaving the biharmonic equation invariant. The extended transformations are analogues of those introduced by

W. Thomson in his study of the Laplace equation in 3-space.

In the first section of this paper we shall show that formulas (5′) and (5″) represent the *only* transformations of type (2) leaving the biharmonic equation invariant. They form a group M in the (x, y, u)-space depending on seven real parameters.

The introduction of M has the advantage that if a problem concerning the biharmonic equation is solved for a domain B of the (x, y)-plane, it can also be solved for any domain B' obtained from B by a Moebius transformation. A further advantage consists in the possibility of introducing domains having $z = \infty$ in the interior or on the boundary. All definitions regarding the behavior of a biharmonic function u at $z = \infty$ are obtained by using one of the transformations (5) transforming $z = \infty$ into a finite point $z' = a'$, and considering the transformed biharmonic function u' at $z' = a'$. For example, u is called *regular* at $z = \infty$ if u' is regular at $z' = a'$. Also the concept of a biharmonic Green's function $\Gamma(x, y)$ with the boundary conditions $u = 0$ and $\partial u/\partial n = 0$ requiring that u and the normal derivative are zero on the boundary (Green's function of the clamped plate) may be extended to the case where the domain considered contains $z = \infty$ in its interior, and ∞ should be the pole of Γ. The singular part of Γ, belonging to a finite pole a', is given by $r'^2 \log r'$, r' denoting the distance of $z' = x' + iy'$ from a'. By using a transformation (5) transforming a' into infinity, one obtains a biharmonic function satisfying the same boundary conditions in the transformed domain whose singular part at $z = \infty$ is $-c \log r$, with a positive constant c, and r representing the distance of z from $z = 0$ or from any other fixed point of the z-plane. *In order to make the definition definite we set* $c = 1$.

This extension of the concept of Green's function will be utilized in §2, which is concerned with a question of Hadamard [3] regarding the sign of the Green's function. He asked whether it may oscillate in sign. R. J. Duffin [1] indicated that the answer is affirmative by constructing solutions of the biharmonic Poisson equation

$$\nabla^4 u = \rho(x, y)$$

in an infinite straight strip satisfying the boundary conditions

$$u = 0, \quad \frac{\partial u}{\partial n} = 0$$

which oscillate in sign although ρ is positive. In §2 simple examples of domains bounded by *analytic Jordan curves* are constructed in which Green's function for suitable choice of the pole may oscillate in sign. Other examples were found by G. Szegö [4] and P. R. Garabedian [2].

There are indications that in the exterior of a *convex* curve with the *pole at infinity* a change of the sign of the Green's function cannot occur. In the last part of § 2 we prove only that this conjecture is *equivalent to positivity* of the *harmonic* function $\nabla^2\Gamma$.

The fact that the biharmonic equation is absolutely invariant only under the group of similarities does not exclude the possibility that for an *individual* biharmonic function u other conformal mappings exist which transform u into a new biharmonic function. Indeed, we shall show in § 1 that in general there exists a one-parameter family of conformal mappings which are not similarities and which transform u into biharmonic functions. In particular one can construct in this way from one Green's function a one-parameter family of Green's functions of *nonsimilar* domains. (Only the case of a circle has to be excluded here.) This also will be discussed in § 1 and applied in § 2.

1. Transformations of biharmonic functions. We shall prove that the transformations of type (2) leaving the biharmonic equation invariant form the group M described by equations (5). All transformations are assumed to be one-to-one and four times continuously differentiable, and the Jacobian shall never vanish.

We make use of the well-known fact that the biharmonic equation is the Euler-Lagrange equation of the variational problem

$$\delta \iint (\nabla^2 u)^2 \, dx \, dy = 0. \tag{6}$$

If the integral of (6) is subjected to a transformation of type (2), an integral in the (x', y')-plane must be obtained whose Euler-Lagrange equation must again be the biharmonic equation

$$\nabla'^4 u' = 0.$$

The new integrand is a quadratic expression in the second derivatives of u' with respect to x' and y', and the second degree terms are evidently given by

$$\frac{1}{\chi^2} \left| J\begin{pmatrix} x & y \\ x' & y' \end{pmatrix} \right| \left\{ \frac{\partial^2 u'}{\partial x'^2} \left[\left(\frac{\partial x'}{\partial x} \right)^2 + \left(\frac{\partial x'}{\partial y} \right)^2 \right] \right. \tag{7}$$

$$\left. + 2 \frac{\partial^2 u'}{\partial x' \partial y'} \left[\frac{\partial x'}{\partial x} \frac{\partial y'}{\partial x} + \frac{\partial x'}{\partial y} \frac{\partial y'}{\partial y} \right] + \frac{\partial^2 u'}{\partial y'^2} \left[\left(\frac{\partial y'}{\partial x'} \right)^2 + \left(\frac{\partial y'}{\partial y} \right)^2 \right] \right\}^2,$$

J being the Jacobian of the transformation from the (x', y')-plane into the (x, y)-plane.

Already from the expression (7) one can derive the fourth order terms of the Euler-Lagrange equation, which by assumption is again the biharmonic equation. This leads by a simple computation to the equations

$$(8) \qquad \left(\frac{\partial x'}{\partial x}\right)^2 + \left(\frac{\partial x'}{\partial y}\right)^2 = \left(\frac{\partial y'}{\partial x}\right)^2 + \left(\frac{\partial y'}{\partial y}\right)^2,$$

$$\frac{\partial x'}{\partial x}\frac{\partial y'}{\partial x} + \frac{\partial x'}{\partial y}\frac{\partial y'}{\partial y} = 0,$$

and we see that the mapping must be conformal.

In order to obtain further conditions on the transformation, we specialize u to an arbitrary *harmonic* function of x and y. Since it is then also a harmonic function of x' and y', we have

$$(9') \qquad \nabla'^2 u' = u \nabla'^2 \chi + 2\left\{\frac{\partial \chi}{\partial x'}\frac{\partial u}{\partial x'} + \frac{\partial \chi}{\partial y'}\frac{\partial u}{\partial y'}\right\},$$

and further,

$$(9'') \qquad 0 = \nabla'^4 u' = u \nabla'^4 \chi + 4\left\{\frac{\partial \nabla'^2 \chi}{\partial x'}\frac{\partial u}{\partial x'} + \frac{\partial \nabla'^2 \chi}{\partial y'}\frac{\partial u}{\partial y'}\right\}$$

$$+ 4\left\{\frac{\partial^2 \chi}{\partial x'^2}\frac{\partial^2 u}{\partial x'^2} + 2\frac{\partial^2 \chi}{\partial x' \partial y'}\frac{\partial^2 u}{\partial x' \partial y'} + \frac{\partial^2 \chi}{\partial y'^2}\frac{\partial^2 u}{\partial y'^2}\right\}.$$

Since $\nabla'^2 u = 0$ represents the *only* relation between the derivatives of u' with respect to x' and y' up to the second order, we may conclude from (9″) that

$$(10) \qquad \frac{\partial^2 \chi}{\partial x'^2} = \frac{\partial^2 \chi}{\partial y'^2}, \quad \frac{\partial^2 \chi}{\partial x' \partial y'} = 0.$$

The only functions satisfying these conditions are those of the form

$$(11') \qquad \chi = c_0(x'^2 + y'^2) + 2c_1 x' + 2c_2 y' + c_3$$

(c_0, c_1, c_2, c_3 arbitrary constants).

Application of the same considerations to the inverse transformation leads to a similar formula for the reciprocal of χ:

$$\frac{1}{\chi} = c_0'(x^2 + y^2) + 2c_1'x + 2c_2'y + c_3', \tag{11''}$$

with suitable constants c_0', c_1', c_2', c_3'. Consider now first the case of a *nonconstant* χ. According to (11′) and (11″), the level lines of χ are in both planes systems of concentric circles each of which may degenerate into a system of parallel lines. But a conformal mapping transforming such systems into each other must be, as is well known, *a proper or improper Moebius transformation.*

In the case of a constant χ we may proceed as follows: We compose the transformation with one of the transformations (5) having a nonconstant χ, and apply to the composed transformation the previous result saying that now the (x, y)-plane is subject to a Moebius transformation. Using the group property of the Moebius transformation, we conclude that also the *original transformation of the* (x, y)*-plane is a Moebius transformation.*

We now have to investigate how the coefficients c_i in χ depend on the Moebius transformation to which the (x, y)-plane is subjected. Since we already know the transformations (5), it is sufficient to consider only the identity transformation. We know already that χ must have the form

$$\chi = c_0(x^2 + y^2) + 2c_1x + 2c_2y + c_3,$$

and multiplication of any biharmonic function u by χ must again lead to a biharmonic function. Setting

$$u = x^2 + y^2$$

gives immediately the result $c_0 = 0$. Setting further

$$u = x(x^2 + y^2) \text{ or } u = y(x^2 + y^2)$$

gives then $c_1 = c_2 = 0$, and we see that χ *must be a constant.* We have thus arrived at:

THEOREM 1. *The most general transformations of type* (2) *which leave the biharmonic equation invariant are represented by formulas* (5′) *and* (5″).

As was already mentioned in the introduction, there exist in general for an *individual* biharmonic function $u(x, y)$ conformal mappings which are not similarities, and which transform u again into a biharmonic function. In order to derive them we make use of the well-known Goursat representation of a biharmonic function,

$$u = \Re\{\bar{z}p(z) + q(z)\}, \tag{12}$$

where $p(z)$ and $q(z)$ are analytic functions of $z = x + iy$, and the symbol $\Re$ indicates the real part of the quantity in parentheses. The representation is unique modulo a change of p and q into

$$p_1 = p + a + icz, \quad q_1 = q - \bar{a}z + id \tag{13}$$

(a, c, d constants, c and d real).

We write further the Laplacian in the more convenient form

$$\nabla^2 u = 4\,\frac{\partial^2 u}{\partial z\,\partial\bar{z}}\,. \tag{14}$$

Without loss of generality we may restrict our attention to proper conformal mapping. Let

$$z' = f(z) \tag{15}$$

be such a mapping transforming u into a biharmonic function in the ($z' = x' + iy'$)-plane. We have

$$\frac{\partial^2 u}{\partial z'\partial\bar{z}'} = \frac{\partial^2 u}{\partial z\,\partial\bar{z}}\,\frac{dz}{dz'}\,\frac{d\bar{z}}{d\bar{z}'} = \Re\left\{\frac{dp}{dz}\,\frac{dz}{dz'}\,\frac{d\bar{z}}{d\bar{z}'}\right\} = \Re\left\{\frac{dp}{dz'}\,\frac{d\bar{z}}{d\bar{z}'}\right\} \tag{16'}$$

and

$$\frac{\partial^4 u}{\partial z'^2\partial\bar{z}'^2} = \Re\left\{\frac{d^2 p}{dz'^2}\,\frac{d^2\bar{z}}{d\bar{z}'^2}\right\}, \tag{16''}$$

and, therefore,

$$\Re\left\{\frac{d^2 p}{dz'^2}\,\frac{d^2\bar{z}}{d\bar{z}'^2}\right\} = 0. \tag{16}$$

Excluding now the trivial case of linear mappings characterized by

$$\frac{d^2z}{dz'^2} \equiv 0,$$

we conclude from (16) that

$$\frac{d^2p}{dz'^2} = ic\,\frac{d^2z}{dz'^2} \qquad (c \text{ a real constant}), \tag{17}$$

and hence

$$p = icz + \alpha z' + \beta \qquad (\alpha, \beta \text{ constants}). \tag{18}$$

We may exclude also the possibility $\alpha = 0$, since then $p = icz + \beta$ and according to (12) our function u is harmonic. But in this case (15) may be *any* conformal mapping. If $\alpha \neq 0$ we have

$$z' = a(p - icz) + b \qquad (a, b, c \text{ constants}, a \neq 0, c \text{ real}). \tag{19}$$

We have thus arrived at:

THEOREM 2. *A proper conformal mapping transforming a given biharmonic but not harmonic function u with the Goursat representation* (12) *again into a biharmonic function is either a similarity or one of the transformations* (19).

REMARK 1. For the functions u with *constant* Laplacian $\nabla^2 u$, the mappings (19) coincide with the *similarity mappings*, and these are the only biharmonic functions of this type.

REMARK 2. By combination of the transformations (19) with transformations of type (5), more general mappings may be obtained transforming u into a biharmonic function.

2. A question of Hadamard regarding the sign of Green's function of the clamped plate. As was already discussed in the introduction, Hadamard asked whether Green's function of the clamped plate may change its sign. We shall construct here very elementary examples showing that this is the case. In order not to interrupt further considerations, we shall first derive several simple lemmas which will be used in our constructions.

Consider first a finite domain B, and let $\Gamma(z_1, z_2)$ be its biharmonic

Green's function now considered as function of a pair of points z_1 and z_2 in B.[1] Because of the well-known symmetry

$$\Gamma(z_1, z_2) = \Gamma(z_2, z_1),$$

it is irrelevant which of the points is considered as pole. We have:

LEMMA 1. *For any choice of m points $z_1, z_2, \cdots, z_m$ in B, the determinant*

$$(20) \qquad D \equiv \begin{vmatrix} \Gamma(z_1, z_1), & \Gamma(z_1, z_2), \cdots, & \Gamma(z_1, z_m) \\ \Gamma(z_2, z_1), & \Gamma(z_2, z_2), \cdots, & \Gamma(z_2, z_m) \\ \cdots & \cdots & \cdots \\ \Gamma(z_m, z_1), & \Gamma(z_m, z_2), \cdots, & \Gamma(z_m, z_m) \end{vmatrix}$$

satisfies

$$D \geq 0.$$

This is an immediate consequence of the well-known fact that $\Gamma(z_1, z_2)$ represents a positive definite kernel.

In particular, $\Gamma(a, a) \geq 0$; but the equality sign cannot hold since then the inequality

$$\begin{vmatrix} \Gamma(a, a) & \Gamma(a, z) \\ \Gamma(z, a) & \Gamma(z, z) \end{vmatrix} = -\Gamma^2(a, z) \geq 0$$

would lead to $\Gamma(a, z) = 0$ for all z in B, which is evidently impossible. We have, therefore:

LEMMA 2.[2] *For all points z in B, we have*

$$(21) \qquad \Gamma(z, z) > 0.$$

We assume now B to contain ∞ in its interior, and state, for its Green's function with the pole at infinity, which we will call $\Gamma(z)$:

[1] It is, for example, sufficient to assume the boundary of B to be three times continuously differentiable to ensure the existence of Γ.

[2] Hadamard ascribes the first proof of inequality (21) to M. Boggio. His formula for the variation of $\Gamma(z, z)$ on p. 28 of the already quoted paper also implicitly contains a proof.

LEMMA 3. *$\Gamma(z)$ can be represented in the neighborhood of $z = \infty$ in the form*

$$\Gamma(z) = -\log r + a_0 r^2 + 2a_1 x + 2a_2 y + \Gamma_1(z) \tag{22}$$

$$(r = |z|, a_0, a_1, a_2 \text{ constants}),$$

with a remainder Γ_1 bounded in the neighborhood of $z = \infty$, and the constant a_0 is positive.

Proof. If we apply to $\Gamma(z)$ the transformation (5′) corresponding to $z' = 1/z$, $\Gamma(z)$ changes into Green's function $\Gamma'(z')$ of the transformed domain B' with the pole in $z' = 0$. But for Γ' we have

$$\Gamma'(z') = r'^2 \log r' + a_0 + 2a_1 x' + 2a_2 y' + \cdots,$$

the dots indicating quantities of at least second order. The constant a_0 is positive by Lemma 2. Transforming back, we obtain the contents of Lemma 3.

We introduce now the Goursat representation of Γ, writing it in the form

$$\Gamma(z) = \Re\{\bar{z}\, p(z)\} - h, \tag{23}$$

where $p(z)$ is analytic and h harmonic.[3] From Lemma 3 we can easily conclude that the free constants in the choice of p and h can be selected so that the following conditions are satisfied:

(a) The function $p(z)$ has at infinity a simple pole with a *positive* $p'(\infty)$.

(b) The function h differs from $\log r$ by a harmonic function in B, *regular also at infinity.*

By the conditions (a) and (b), the functions p and h are uniquely determined.

We shall now derive *properties of $p(z)$ characterizing it independently of h.* We use the analytic function

$$w(z) = 2\frac{\partial h}{\partial z}.$$

---

[3]We assume from now on that B is simply connected, in which case $p(z)$ is always single-valued.

Its expansion at infinity starts with the term $1/z$. From (23) and the boundary conditions satisfied by Γ we conclude that, on the boundary of B,

$$0 = \frac{\partial \Gamma}{\partial z} = \frac{1}{2}\left\{\overline{p(z)} + \frac{dp}{dz}\bar{z}\right\} - \frac{\partial h}{\partial z}. \tag{24}$$

This equation evidently completely expresses the boundary conditions on Γ modulo an additive constant. We have proved, therefore:

LEMMA 4. *The function $p(z)$ is characterized by the two properties:*

(a) *It has at infinity a simple pole with a positive derivative $p'(\infty)$.*

(b) *The function*

$$\overline{p(z)} + \frac{dp}{dz}\bar{z}$$

coincides on the boundary of B with a function which is analytic in B and whose expansion at ∞ starts with $1/z$.

We know from Theorem 2 that $p(z)$ maps our domain B onto a domain B_1 (not necessarily schlicht) with preservation of the Green's function with pole at infinity. In order to bring Lemma 4 into a form *in which B and B_1 play a symmetric role*, we introduce the function $z = g(\zeta)$ which maps the exterior of the unit circle $|\zeta| = 1$ onto B so that

$$g(\infty) = \infty, \quad g'(\infty) > 0. \tag{25}$$

In a similar way,

$$f(\zeta) = p(g(\zeta))$$

maps the exterior of $|\zeta| = 1$ onto the domain B_1, and we have again

$$f(\infty) = \infty, \quad f'(\infty) > 0. \tag{25}$$

Lemma 4 can now be expressed by saying that

$$\overline{f(\zeta)} + \frac{df}{d\zeta}\frac{d\zeta}{dz}\overline{g(\zeta)}$$

coincides on $|\zeta| = 1$ with the boundary values of a function analytic in $|\zeta| > 1$

whose Taylor expansion at infinity starts with the term $(g'(\infty)\zeta)^{-1}$. Multiplication with $dz/d\zeta$ leads, therefore, to:

LEMMA 5. *The function* $f(\zeta)$ *is characterized by the following properties* :

(a) *It is analytic in* $|\zeta| > 1$ *and has at infinity a simple pole with* $f'(\infty) > 0$.

(b) *The function*

$$\overline{f(\zeta)}\,\frac{dg(\zeta)}{d\zeta} + \overline{g(\zeta)}\,\frac{df(\zeta)}{d\zeta} \tag{26}$$

coincides on $|\zeta| = 1$ *with the boundary values of a function* $\omega(\zeta)$ *analytic in* $|\zeta| > 1$ *whose expansion at infinity starts with the term* $1/\zeta$.

As soon as $f(\zeta)$ is determined, it is easy to construct the Green's function by using equation (24). It gives, after transformation into the ζ-plane,

$$2\,\frac{\partial h}{\partial \zeta} = \omega(\zeta) \qquad (|\zeta| \geq 1). \tag{27}$$

An integration of $\omega(\zeta)$ determines h modulo an additive constant which is to be adjusted to the boundary condition $\Gamma = 0$.

Lemma 5 will now be utilized to find simple examples of domains whose Green's function oscillates in sign. The simplest choice of $g(\zeta)$ one might try would be

$$g(\zeta) = \zeta + \frac{\beta_1}{\zeta} \qquad (|\beta_1| \leq 1),$$

which maps $|\zeta| \geq 1$ onto the exterior of an ellipse or, in the limiting case $|\beta_1| = 1$, onto a slit domain. In this case, one verifies easily that

$$f(\zeta) = \frac{\zeta - \beta_1/\zeta}{2(\beta_1\bar{\beta}_1 + 1)} \;; \tag{28}$$

and a simple computation gives, for Γ written as function of ζ, the expression

$$\Gamma = \frac{1}{2(1+\beta_1\bar{\beta}_1)}\left(\rho^2 - \frac{\beta_1\bar{\beta}_1}{\rho^2} - 1 + \beta_1\bar{\beta}_1\right) - \log\rho, \; (|\zeta| = \rho,\ \rho \geq 1). \tag{29}$$

But one verifies easily that Γ is *here always positive*. We try, therefore, for $g(\zeta)$ an expression[4]

$$g(\zeta) = \zeta + \frac{\beta_1}{\zeta} + \frac{\beta_2}{\zeta^2} \tag{30}$$

and shall show that for suitable choice of the constants β_1 and β_2 we obtain a schlicht map of $|\zeta| \geq 1$ on a domain B with a Green's function *whose sign oscillates.*

First we shall show that the corresponding $f(\zeta)$ is of the form

$$f(\zeta) = c\left\{\zeta + \alpha_0 + \frac{\alpha_1}{\zeta} + \frac{\alpha_2}{\zeta^2}\right\}, \quad c > 0. \tag{31}$$

In order to verify this we introduce the analytic functions

$$\tilde{g}(\zeta) = \frac{1}{\zeta} + \overline{\beta}_1\,\zeta + \overline{\beta}_2\,\zeta^2, \tag{32′}$$

$$\tilde{f}(\zeta) = c\left\{\frac{1}{\zeta} + \overline{\alpha}_0 + \overline{\alpha}_1\,\zeta + \overline{\alpha}_2\,\zeta^2\right\}, \tag{32″}$$

which coincide on $|\zeta| = 1$ with $\overline{g(\zeta)}$ and $\overline{f(\zeta)}$, respectively. According to Lemma 5, we have to show that the constants c, α_0, α_1, α_2 can be chosen in such a way that

$$\omega(\zeta) = \tilde{f}(\zeta)\,\frac{dg}{d\zeta} + \tilde{g}(\zeta)\,\frac{df}{d\zeta} \tag{33}$$

has an expansion in $1/\zeta$ starting with the term $1/\zeta$. We have

$$\omega(\zeta) = c\left\{\frac{1}{\zeta} + \overline{\alpha}_0 + \overline{\alpha}_1\,\zeta + \overline{\alpha}_2\,\zeta^2\right\}\left\{1 - \frac{\beta_1}{\zeta^2} - \frac{2\beta_2}{\zeta^3}\right\} \tag{34}$$

$$+ c\left\{\frac{1}{\zeta} + \overline{\beta}_1\,\zeta + \overline{\beta}_2\,\zeta^2\right\}\left\{1 - \frac{\alpha_1}{\zeta^2} - \frac{2\alpha_2}{\zeta^3}\right\},$$

or

[4] N. Mouskhelichvili gave a general procedure telling how to construct Green's func tion of a domain whose mapping function $g(\zeta)$ is rational. We have to compute it in cas((30) in all details. See [4]. N. Mouskhelichvili, *Application des intégrales analogues* (*celles de Cauchy a quelques problème de la physique mathématique*, Tiflis, 1922.

$$(34') \qquad \omega(\zeta) = c\left\{(\overline{\alpha}_2 + \overline{\beta}_2)\,\zeta^2 + (\overline{\alpha}_1 + \overline{\beta}_1)\,\zeta - (\overline{\alpha}_2\,\beta_1 + \overline{\beta}_2\,\alpha_1 - \overline{\alpha}_0) - \frac{2\overline{\alpha}_2\,\beta_2 + 2\overline{\beta}_2\,\alpha_2 + \overline{\alpha}_1\,\beta_1 + \overline{\beta}_1\,\alpha_1 - 2}{\zeta} + \cdots\right\}.$$

The conditions on $f(\zeta)$ are, therefore, satisfied if

$$(35') \qquad \alpha_1 + \beta_1 = 0, \quad \alpha_2 + \beta_2 = 0, \quad \alpha_0 = \overline{\alpha}_1\,\beta_2 + \overline{\beta}_1\,\alpha_2,$$

and

$$c\,(2\overline{\alpha}_2\,\beta_2 + 2\overline{\beta}_2\,\alpha_2 + \overline{\alpha}_1\,\beta_1 + \overline{\beta}_1\,\alpha_1 - 2) = -1,$$

which gives

$$(35) \qquad \alpha_1 = -\beta_1, \quad \alpha_2 = -\beta_2, \quad \alpha_0 = -2\overline{\beta}_1\,\beta_2,$$

and

$$c^{-1} = 2(1 + \beta_1\,\overline{\beta}_1 + 2\beta_2\,\overline{\beta}_2).$$

Our function $f(\zeta)$ is, therefore, given by

$$(36) \qquad f(\zeta) = \left\{2(1 + \beta_1\,\overline{\beta}_1 + 2\beta_2\,\overline{\beta}_2)\right\}^{-1} \cdot \left\{\zeta - 2\overline{\beta}_1\,\beta_2 - \frac{\beta_1}{\zeta} - \frac{\beta_2}{\zeta^2}\right\}.$$

In order to obtain Γ, we have to compute, according to formula (23), the real part of $f(\zeta)\overline{g(\zeta)}$. Using the expression (36), we obtain

$$c^{-1}f(\zeta)\overline{g(\zeta)} = \left\{\zeta - 2\overline{\beta}_1\,\beta_2 - \frac{\beta_1}{\zeta} - \frac{\beta_2}{\zeta^2}\right\}\left\{\overline{\zeta} + \frac{\overline{\beta}_1}{\overline{\zeta}} + \frac{\overline{\beta}_2}{\overline{\zeta}^2}\right\}$$

$$= \zeta\overline{\zeta} - 2\overline{\beta}_1\,\beta_2\,\overline{\zeta} + \overline{\beta}_1\,\frac{\zeta}{\overline{\zeta}} - \beta_1\,\frac{\overline{\zeta}}{\zeta} + \overline{\beta}_2\,\frac{\zeta}{\overline{\zeta}^2} - 2\,\frac{\overline{\beta}_1^2\,\beta_2}{\overline{\zeta}} - \beta_2\,\frac{\overline{\zeta}}{\zeta^2}$$

$$- \frac{2\overline{\beta}_1\,\overline{\beta}_2\,\beta_2}{\overline{\zeta}^2} - \frac{\beta_1\,\overline{\beta}_1}{\zeta\overline{\zeta}} + \frac{\beta_1\,\overline{\beta}_2}{\zeta\overline{\zeta}^2} - \frac{\overline{\beta}_1\,\beta_2}{\overline{\zeta}\zeta^2} - \frac{\beta_2\,\overline{\beta}_2}{\zeta^2\overline{\zeta}^2},$$

and

$$(37)\qquad c^{-1}\,\Re\{f(\zeta)\overline{g(\zeta)}\} = \rho^2 - \frac{\beta_1\overline{\beta}_1}{\rho^2} - \frac{\beta_2\overline{\beta}_2}{\rho^4}$$

$$- 2\,\Re\left\{\beta_1\overline{\beta}_2\,\zeta + \frac{\beta_1^2\overline{\beta}_2}{\zeta} + \frac{\beta_1\beta_2\overline{\beta}_2}{\zeta^2} + \frac{\overline{\beta}_1\beta_2}{\overline{\zeta}\,\zeta^2}\right\}.$$

We substitute further into the formula (34) for $\omega(\zeta)$ the expressions (35) for α_0, α_1, α_2, and obtain

$$(38')\qquad \omega(\zeta) = c\left\{\frac{1}{\zeta} - 2\beta_1\overline{\beta}_2 - \overline{\beta}_1\zeta - \overline{\beta}_2\zeta^2\right\}\left\{1 - \frac{\beta_1}{\zeta^2} - \frac{2\beta_2}{\zeta^3}\right\}$$

$$+ c\left\{\frac{1}{\zeta} + \overline{\beta}_1\zeta + \overline{\beta}_2\zeta^2\right\}\left\{1 + \frac{\beta_1}{\zeta^2} + \frac{2\beta_2}{\zeta^3}\right\},$$

or, after a simple computation,

$$(38)\qquad \omega(\zeta) = \frac{1}{\zeta} + c\left\{\frac{2(\beta_1^2\overline{\beta}_2 + 2\overline{\beta}_1\beta_2)}{\zeta^2} + \frac{4\beta_1\beta_2\overline{\beta}_2}{\zeta^3}\right\}.$$

In order to obtain h we may, according to (27), integrate $\omega(\zeta)$:

$$(39)\qquad \int\omega(\zeta)\,d\zeta = \log\zeta - 2c\left\{\frac{\beta_1^2\overline{\beta}_2 + 2\overline{\beta}_1\beta_2}{\zeta} + \frac{\beta_1\beta_2\overline{\beta}_2}{\zeta^2}\right\} + C \quad (C \text{ a constant}).$$

Since the real part of (39) must coincide on $|\zeta| = 1$ with $\Re\{f(\zeta)\overline{g(\zeta)}\}$, we obtain, by comparison of (37) and (39),

$$(40')\qquad c\{1 - \beta_1\overline{\beta}_1 - \beta_2\overline{\beta}_2\} = \Re(C),$$

or

$$(40)\qquad \Re(C) = \frac{1 - \beta_1\overline{\beta}_1 - \beta_2\overline{\beta}_2}{2(1 + \beta_1\overline{\beta}_1 + 2\beta_2\overline{\beta}_2)}.$$

From the foregoing formulas we finally obtain, by substitution into (23), the following expression for $\Gamma(z)$ in terms of ζ:

$$c^{-1}\Gamma = \rho^2 - \frac{\beta_1\bar\beta_1}{\rho^2} - \frac{\beta_2\bar\beta_2}{\rho^4} - 2\,\Re\left\{\beta_1\bar\beta_2\zeta - \frac{2\bar\beta_1\beta_2}{\zeta} + \frac{\bar\beta_1\beta_2}{\rho^2\zeta}\right\} \tag{41}$$

$$- \frac{1}{c}\log\rho - (1-\beta_1\bar\beta_1 - \beta_2\bar\beta_2),$$

with

$$c^{-1} = 2(1+\beta_1\bar\beta_1 + 2\beta_2\bar\beta_2). \tag{42}$$

We can now show that the constants β_1 and β_2 can be chosen in such a way that $g(\zeta)$ represents a schlicht mapping of $|\zeta| \geq 1$, and still $\Gamma(\zeta)$ oscillates in sign. Evidently we obtain an oscillating Γ if the normal derivative

$$\left(\frac{\partial^2\Gamma}{\partial\rho^2}\right)_{\rho=1}$$

on the unit circle $|\zeta| = 1$ becomes *negative* in some of its points. But we have, from (41),

$$c^{-1}\left(\frac{\partial^2\Gamma}{\partial\rho^2}\right)_{\rho=1} = 2 - 6\beta_1\bar\beta_1 - 20\beta_2\bar\beta_2$$

$$-2\Re\left\{-\frac{4\bar\beta_1\beta_2}{\zeta} + \frac{6\bar\beta_1\beta_2}{\zeta} + \frac{2\bar\beta_1\beta_2}{\zeta} + \frac{4\bar\beta_1\beta_2}{\zeta}\right\} + 2(1+\beta_1\bar\beta_1 + 2\beta_2\bar\beta_2),$$

or

$$c^{-1}\left(\frac{\partial^2\Gamma}{\partial\rho^2}\right)_{\rho=1} = 4 - 4\beta_1\bar\beta_1 - 16\beta_2\bar\beta_2 - 16\,\Re\,\frac{\beta_1\bar\beta_2}{\zeta} \qquad (|\zeta|=1). \tag{43}$$

This expression becomes negative on the unit circle if

$$16|\beta_1\beta_2| > 4 - 4\beta_1\bar\beta_1 - 16\beta_2\bar\beta_2,$$

or

$$|\beta_1| + 2|\beta_2| > 1. \tag{44}$$

The function $g(\zeta)$ represents a schlicht mapping of the exterior of the unit circle if the difference quotient

$$\frac{g(\zeta_1) - g(\zeta_2)}{\zeta_1 - \zeta_2}$$

has positive real part there. This is the case if

$$|\beta_1| + 2|\beta_2| \leq 1.$$

But even the exterior of slightly smaller circle is mapped schlicht under this condition if only the cases $|\beta_1| + 2|\beta_2| = 1$, and β_1^3 and β_2^2 of equal argument, are excluded. If, therefore, $|\beta_1| + 2|\beta_2| = 1$, and β_1^3 and β_2^2 have different arguments, all sufficiently close values β_1, β_2 give schlicht mappings of $|\zeta| \geq 1$, and can be chosen so that (44) is satisfied. *We thus obtain examples of domains where* Γ *oscillates in sign.*

We conjecture that, for the exterior of a *convex* curve and pole at infinity, Γ is *positive*. We shall support this conjecture by proving:

THEOREM 3. *For the exterior of a convex curve and pole at infinity the positivity of* Γ *is equivalent to the positivity of* $\nabla^2\Gamma$.

This is important because $\nabla^2\Gamma$ is a *harmonic* function.

First we shall prove that the positivity of Γ implies the positivity of $\nabla^2\Gamma$. Assume first that the boundary curve is analytic. Then Γ can be analytically continued beyond the boundary curve, and we can speak of derivatives of higher order on the curve itself. From the positivity of Γ and the boundary condition it follows immediately that the normal derivative of second order on the boundary satisfies

$$\frac{\partial^2\Gamma}{\partial n^2} \geq 0.$$

But the second derivative in the tangential direction is zero again on account of the boundary conditions. We have, therefore, $\nabla^2\Gamma \geq 0$ on the boundary. Since $\nabla^2\Gamma$ is *harmonic* and, on account of Lemma 3,

$$[\nabla^2\Gamma]_{z=\infty} > 0,$$

the Laplacian $\nabla^2\Gamma$ is *positive* in the whole domain.

The condition of analyticity of the boundary curve now can easily be dropped by a limiting process.

We shall now prove the converse: $\nabla^2\Gamma > 0$ implies $\Gamma > 0$. For the proof we need some preliminary considerations regarding the following question: Which differential operators of second order,

$$(45)\qquad v(x, y) = a_{11}(x, y)\frac{\partial^2 u}{\partial x^2} + 2a_{12}(x, y)\frac{\partial^2 u}{\partial x\,\partial y} + a_{22}(x, y)\frac{\partial^2 u}{\partial y^2}$$
$$+ 2a_1(x, y)\frac{\partial u}{\partial x} + 2a_2(x, y)\frac{\partial u}{\partial y} + a_3(x, y)u,$$

transform an arbitrary *biharmonic* function $u(x, y)$ into a *harmonic* function $v(x, y)$? The answer is given in:

LEMMA 6. *The most general operator of the required type is of the form*

$$(46)\quad v(x, y) = m(x, y)\nabla^2 u - 2\left\{\frac{\partial m(x, y)}{\partial x}\frac{\partial u}{\partial x} + \frac{\partial m(x, y)}{\partial y}\frac{\partial u}{\partial y}\right\} + \nabla^2 m \cdot u,$$

where the function $m(x, y)$ *is of the form*

$$(47)\quad m(x, y) = c_0(x^2 + y^2) + 2c_1 x + 2c_2 y + c_3 \quad (c_0, c_1, c_2, c_3 \text{ constants}).$$

Proof. We form

$$(48')\qquad \nabla^2 v = a_{11}\frac{\partial^2\nabla^2 u}{\partial x^2} + 2a_{12}\frac{\partial^2\nabla^2 u}{\partial x\,\partial y} + a_{22}\frac{\partial^2\nabla^2 u}{\partial y^2} + \text{terms of lower order.}$$

Since the only relation between the derivatives of u up to the fourth order is given by $\nabla^4 u = 0$, we see already from (48′) that

$$(49')\qquad a_{12} = 0,\ a_{11} = a_{22}.$$

Calling $a_{11} = a_{22} = m$, we can write (45) in the form

$$(45')\qquad v = m\nabla^2 u + 2a_1\frac{\partial u}{\partial x} + 2a_2\frac{\partial u}{\partial y} + a_3 u.$$

From (45′) we obtain

$$(48'') \qquad \nabla^2 v = 2\left\{\frac{\partial m}{\partial x}\frac{\partial \nabla^2 u}{\partial x} + \frac{\partial m}{\partial y}\frac{\partial \nabla^2 u}{\partial y}\right\} + 2\left\{a_1 \frac{\partial \nabla^2 u}{\partial x} + a_2 \frac{\partial \nabla^2 u}{\partial y}\right\}$$

+ terms of lower order.

From (48″) we conclude that

$$(49'') \qquad a_1 = -\frac{\partial m}{\partial x}, \quad a_2 = -\frac{\partial m}{\partial y},$$

and (45′) can be rewritten as

$$(45'') \qquad v = m\nabla^2 u - 2\left\{\frac{\partial m}{\partial x}\frac{\partial u}{\partial x} + \frac{\partial m}{\partial y}\frac{\partial u}{\partial y}\right\} + a_3 u.$$

From this equation we derive

$$(48''') \qquad \nabla^2 v = \nabla^2 m \nabla^2 u - 4\left\{\frac{\partial^2 m}{\partial x^2}\frac{\partial^2 u}{\partial x^2} + 2\frac{\partial^2 m}{\partial x \partial y}\frac{\partial^2 u}{\partial x \partial y} + \frac{\partial^2 m}{\partial y^2}\frac{\partial^2 u}{\partial y^2}\right\}$$

$+ a_3 \nabla^2 u$ + term of lower order ,

which gives

$$(49'') \qquad \frac{\partial^2 m}{\partial x^2} = \frac{\partial^2 m}{\partial y^2}, \quad \frac{\partial^2 m}{\partial x \partial y} = 0, \quad a_3 = \nabla^2 m,$$

and we obtain (45) in the form (46) with m given by an expression (47). We verify easily that for any such choice of m the Laplacian satisfies

$$\nabla^2 v = 0.$$

Let B now be the exterior of a closed *convex* curve, and Γ its Green's function with the singular point at infinity, and assume that $\nabla^2\Gamma > 0$ in B. Take a straight line which does not penetrate into the interior of the boundary curve. By change of the coordinate system we can make this line the y-axis, so that B lies to its left. We now apply Theorem 5 with the special choice $m = x$, and obtain that

$$(50) \qquad v = x\nabla^2\Gamma - 2\frac{\partial \Gamma}{\partial x}$$

is *harmonic* in B. On the boundary of B,

$$\frac{\partial \Gamma}{\partial x} = 0, \quad \nabla^2 \Gamma \geq 0 \tag{51}$$

the latter inequality being a consequence of the assumption that $\nabla^2 \Gamma > 0$ in B. Since our domain lies completely in the half-plane $x \leq 0$, we conclude from (51) that v has nonpositive boundary values. From the behavior of Γ at $z = \infty$ expressed by Lemma 2, we see that v is regular at infinity. From the extremum properties of harmonic functions we now conclude that $v \leq 0$ *in the whole* B. In particular, on the y-axis where $x = 0$ we obtain the result

$$\frac{\partial \Gamma}{\partial x} \geq 0. \tag{52}$$

Let l now be a half line originating in a point C of the boundary curve of B orthogonal to the tangent line at C. Laying the y-axis perpendicular to l through any of its points, we see that (52) holds in all points of l; and, since $\Gamma = 0$ at C, we arrive at the inequality $\Gamma \geq 0$ along the whole l. But the equality cannot hold, for otherwise Γ would be zero along a whole segment of l and, since it is analytic, along the whole l. This contradicts Lemma 2, which implies that $\Gamma \longrightarrow \infty$ as $z \longrightarrow \infty$. The whole domain B can be covered with half lines having the properties of l. The inequality $\Gamma > 0$ holds, therefore, in the whole B.

References

1. R. J. Duffin, *On a question of Hadamard concerning superbiharmonic functions*, J. Math. Physics, 27 (1949), 253-258.

2. P. R. Garabedian, *A partial differential equation arising in conformal mapping*, O.N.R. Technical Report No. 8, N6ori-106, Task V (NR-043-992).

3. J. Hadamard, *Mémoire sur le problème d'analyse rélatif à l'équilibre des plaques élastiques encastrées*, Mémoire présentée par divers savants à l'Académie des Sciences, ser. 2, 33 (1908), 128 pages.

4. N. Mouskhelichvili, *Application des intégrales analogues a celles de Cauchy a quelques problème de la physique mathématique*, Tiflis, 1922.

5. G. Szegö, See the paper following the present paper.

STANFORD UNIVERSITY

REMARK ON THE PRECEDING PAPER OF CHARLES LOEWNER

G. Szegö

1. Introduction. In the preceding paper, Charles Loewner constructed certain Jordan curves with the property that the clamped plates bounded by such Jordan curves have an oscillating Green's function. The question concerning the sign of the Green's function has been raised by J. Hadamard, and this problem has been pursued recently by R. J. Duffin and P. R. Garabedian. The construction of Loewner is based on a method due to N. Muskhelichvili using appropriate conformal mappings. [1]

The purpose of the present note is to construct such Jordan curves in an elementary manner. For the sake of completeness we repeat a few definitions to be found in the preceding paper.

A function of $u(x, y)$ defined in a domain g and having therein continuous partial derivatives of the fourth order is called a biharmonic function in g if it satisfies the biharmonic equation

$$\nabla^4 u = \nabla^2 \nabla^2 u = \frac{\partial^4 u}{\partial x^4} + 2 \frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = 0. \tag{1}$$

Let g be a connected domain bounded by a finite number of analytic arcs. Let q be a fixed point in g. The Green's function $\Gamma(p) = \Gamma(p; q)$ of g with respect to q is a function of the variable point $p = p(x, y)$ satisfying the following conditions:

(a) Γ is a biharmonic function of p except at the singular point q. Denoting by r the distance of p from q, we have

$$\Gamma = r^2 \log r + k, \tag{2}$$

where $k(p) = k(x, y)$ is biharmonic in g without exception.

(b) On the boundary of g we have the conditions:

[1] See the References given in the paper of Loewner.

Received June 23, 1952.
Pacific J. Math. 3 (1953), 437-446

$$\Gamma = \frac{\partial \Gamma}{\partial n} = 0. \tag{3}$$

A function $u(x, y)$ biharmonic in the neighborhood of $x = 0$, $y = 0$ can be written in the form:

$$u(x, y) = (x^2 + y^2)\, u_1(x, y) + u_2(x, y), \tag{4}$$

where u_1 and u_2 are harmonic functions in the neighborhood of $x = y = 0$.

The previous concepts can be extended to infinite domains by using an analogue of Thomson's transformation. As is obvious from the representation (4), we have:

Let $u(x, y) = u(r, \phi)$ be harmonic in the neighborhood of the origin $r = 0$ (r, ϕ are polar coordinates). We apply the inversion

$$\begin{aligned} x &= x'(x'^2 + y'^2)^{-1}, \qquad & y &= y'(x'^2 + y'^2)^{-1}, \\ r &= (r')^{-1}, & \phi &= \phi'. \end{aligned} \tag{5}$$

The function

$$U(x', y') = (r')^2\, u(x, y) \tag{6}$$

will be then a biharmonic function in the neighborhood of $x' = \infty$, $y' = \infty$.

A function biharmonic in the neighborhood of the origin can be presented as a linear combination of the basic biharmonic functions

$$\begin{aligned} &r^n \cos n\phi, \qquad r^n \sin n\phi, \\ &r^{n+2} \cos n\phi, \qquad r^{n+2} \sin n\phi, \qquad\qquad n = 0, 1, 2, \cdots, \end{aligned} \tag{7}$$

to which $r^2 \log r$ has to be added if the function is singular at the origin (as for instance is the case for the Green's function with respect to the origin).

A function biharmonic in the neighborhood of $x = \infty$, $y = \infty$, can be represented as a linear combination of the basic biharmonic functions

$$\begin{aligned} &r^{-n} \cos n\phi, \qquad r^{-n} \sin n\phi, \\ &r^{2-n} \cos n\phi, \qquad r^{2-n} \sin n\phi, \qquad\qquad n = 0, 1, 2, \cdots, \end{aligned} \tag{8}$$

to which $\log r$ has to be added if the function is singular at infinity.

By use of the inversion (5), (6) there is no difficulty in defining the Green's function of an infinite domain with singular point at infinity provided that this point is an interior point of the domain.

2. **Results.** In order to prepare the construction announced above, we consider the infinite plane, which we interpret as the complex z-plane, $z = x + iy = re^{i\phi}$, cut along the following circular arc of the unity circle:

$$(1) \qquad r = 1, \quad \pi - \alpha \leqq \phi \leqq \pi + \alpha.$$

Here α is given, $0 < \alpha < \pi$. We map this infinite slit domain onto the exterior of the unit circle of the ζ-plane, $\zeta = \rho e^{i\psi}$, in such a way that $z = \infty$ and $\zeta = \infty$ correspond to each other. Furthermore, we assume that $dz/d\zeta$ is positive at $z = \zeta = \infty$. This mapping has the following form:

$$(2) \qquad z = \frac{\zeta(\lambda\zeta - 1)}{\zeta - \lambda}, \qquad 0 < \lambda < 1.$$

Here λ is an appropriate function of α.

First we note that the real point $z = -1$ of the slit corresponds to $\zeta = 1$ and $\zeta = -1$. Now let $\lambda = \cos\psi_0$, $0 < \psi_0 < \pi/2$; since

$$(3) \qquad \frac{dz}{d\zeta} = \lambda \frac{1 - 2\lambda\zeta + \zeta^2}{(\zeta - \lambda)^2},$$

we see that the points $\zeta = e^{\pm i\psi_0}$ correspond to the end-points $z = -e^{\pm i\alpha}$ of the slit. More precisely, $e^{i\psi_0}$ corresponds to $e^{i(\pi-\alpha)} = -e^{-i\alpha}$. As $\zeta = e^{i\psi}$ describes the unit circle in the positive sense, z surrounds the circular slit; the arc $-\psi_0 < \psi < \psi_0$ corresponds to the inner (concave) side of the slit, and the remaining arc to the outer (convex) side of it. In particular, $\zeta = 1$ and $\zeta = -1$ are transformed into the point $z = -1$ on the concave and convex border of the slit, respectively.

Inserting $\zeta = e^{i\psi_0}$ in (2), we find that

$$\frac{e^{i\psi_0}(\cos\psi_0 \cdot e^{i\psi_0} - 1)}{e^{i\psi_0} - \cos\psi_0} = \frac{e^{2i\psi_0}(\cos\psi_0 - e^{-i\psi_0})}{e^{i\psi_0} - \cos\psi_0} = e^{2i\psi_0} = -e^{-i\alpha},$$

so that $2\psi_0 = \pi - \alpha$; hence

$$(4) \qquad \lambda = \sin(\alpha/2).$$

We denote the image of the circle $|\zeta| = R$, $R > 1$, by C_R. This is an analytic Jordan curve.

The principal results of this note are the following:

I. Let $\Gamma(p)$ be the Green's function of the infinite slit domain of the z-plane bounded by the circular arc (1), having at $z=\infty$ its singular point. *This function changes its sign in the slit domain just defined.*

II. Let $\Gamma(p)$ be the Green's function of the infinite domain outside of the curve C_R, $R>1$, having at $z=\infty$ its singular point. *This function will change its sign in the infinite domain outside of C_R provided R is sufficiently near to* 1.

From the last example it is easy to derive an example of the kind announced in the introduction: we have to apply an inversion to the curve C_R with respect to any fixed interior point. Here we must use the results of Chapter 1.

3. Circular slit. We seek the Green's function $\Gamma(p)$ of the circular slit domain in the following form:

$$\Gamma(p) = \log\frac{1}{\rho} + A\left(\rho - \frac{1}{\rho}\right)\cos\psi + f(\rho,\psi) + (r^2-1)\, g(\rho,\psi). \tag{1}$$

Here A is a constant, and $f(\rho,\psi)$ and $g(\rho,\psi)$ are harmonic functions regular for $\rho>1$, including $\rho=\infty$. The point p is represented by the complex number $z=re^{i\phi}$ defined in 2. The relation between $z=re^{i\phi}$ and $\zeta=\rho e^{i\psi}$ is given by 2 (2).

The boundary conditions of the clamped plate amount to the fact that the function (1) and its derivative with respect to ρ vanish as $\rho=1$. But $\rho=1$ implies that $r=1$, so that we have:

(I) $f(1,\psi)=0$; i.e., $f(\rho,\psi)\equiv 0$.

(II) $-1+2A\cos\psi+\left(\dfrac{\partial(r^2)}{\partial\rho}\right)_{\rho=1} g(1,\psi)=0.$

Now we note the following formulas which will be useful in our later work:

$$\begin{aligned} r^2 &= \rho^2\,\frac{\lambda^2\rho^2-2\lambda\rho\cos\psi+1}{\rho^2-2\lambda\rho\cos\psi+\lambda^2},\\ r^2-1 &= (\rho^2-1)\,\frac{-2\lambda\rho\cos\psi+\lambda^2(\rho^2+1)}{\rho^2-2\lambda\rho\cos\psi+\lambda^2}. \end{aligned}\tag{2}$$

From the second we conclude that

$$\left(\frac{\partial(r^2)}{\partial\rho}\right)_{\rho=1} = 2\left(\frac{-2\lambda\rho\cos\psi+\lambda^2(\rho^2+1)}{\rho^2-2\lambda\rho\cos\psi+\lambda^2}\right)_{\rho=1} = \frac{4\lambda(-\cos\psi+\lambda)}{1-2\lambda\cos\psi+\lambda^2}. \tag{3}$$

Hence condition (II) can be written as follows:

$$(4) \qquad -1 + 2A \cos\psi + \frac{4\lambda(-\cos\psi + \lambda)}{1 - 2\lambda\cos\psi + \lambda^2}\, g(1, \psi) = 0.$$

We determine A according to the condition

$$-1 + 2A \cos\psi + \frac{4\lambda(-\cos\psi + \lambda)}{4\lambda^2} = 0,$$

$$A = (2\lambda)^{-1},$$

and (4) yields

$$(5) \qquad \begin{aligned} g(1, \psi) &= \frac{1 - 2\lambda\cos\psi + \lambda^2}{4\lambda^2}; \\ g(\rho, \psi) &= \frac{1 + \lambda^2}{4\lambda^2} - \frac{1}{2\lambda}\,\frac{\cos\psi}{\rho}. \end{aligned}$$

Recapitulating, we find the following expression for the Green's function $\Gamma(p)$:

$$(6) \quad \Gamma(p) = \log\frac{1}{\rho} + \frac{1}{2\lambda}\left(\rho - \frac{1}{\rho}\right)\cos\psi + (r^2 - 1)\left(\frac{1 + \lambda^2}{4\lambda^2} - \frac{\cos\psi}{2\lambda\rho}\right).$$

In the limiting case $\alpha \longrightarrow \pi$, $\lambda \longrightarrow 1$, we obtain of course the Green's function of the exterior of the unit circle, namely,

$$(7) \qquad \log\frac{1}{\rho} + \frac{1}{2}(\rho^2 - 1).$$

4. Conclusion. (a) The dominant term in 3 (6) is

$$r^2\,\frac{1 + \lambda^2}{4\lambda^2},$$

so that Γ is positive as $z \longrightarrow \infty$. Now we write

$$z = r = 0, \; \zeta = \rho = \frac{1}{\lambda},$$

and have

$$(1)\qquad \begin{aligned}\Gamma &= \log\lambda + \frac{1}{2\lambda}\left(\frac{1}{\lambda} - \lambda\right) - \left(\frac{1+\lambda^2}{4\lambda^2} - \frac{1}{2}\right)\\ &= \log\lambda + \frac{1-\lambda^2}{4\lambda^2}\,.\end{aligned}$$

This quantity is certainly negative if λ is sufficiently near 1, more precisely if $\lambda > \lambda_0$ where λ_0 is the only root of the equation

$$(2)\qquad \log\lambda + \frac{1-\lambda^2}{4\lambda^2} = 0$$

on the range $0 < \lambda < 1$.

(b) We can show however that Γ must change its sign *for all* λ, $0 < \lambda < 1$. For this purpose we compute the following second derivative at the point $z = -1$ on the *concave* side of the slit:

$$\left(\frac{d^2\Gamma}{d\rho^2}\right)_{\rho=1,\,\psi=0} = 1 - \frac{1}{\lambda} + \left(\frac{d^2(r^2)}{d\rho^2}\right)_{\rho=1,\,\psi=0} \frac{(1-\lambda)^2}{4\lambda^2} + \left(\frac{d(r^2)}{d\rho}\right)_{\rho=1,\,\psi=0} \cdot \frac{1}{\lambda}\,.$$

From the second formula in 3 (2) we find

$$\left(\frac{d^2(r^2)}{d\rho^2}\right)_{\rho=1,\,\psi=0} = 2\,\frac{-2\lambda+2\lambda^2}{(1-\lambda)^2} + 4\,\frac{(-2\lambda+2\lambda^2)(1-\lambda)^2 - (-2\lambda+2\lambda^2)(2-2\lambda)}{(1-\lambda)^4} = \frac{4\lambda(1+3\lambda)}{(1-\lambda)^2}\,,$$

so that, in view of 3 (3),

$$(3)\qquad \left(\frac{d^2\Gamma}{d\rho^2}\right)_{\rho=1,\,\psi=0} = 1 - \frac{1}{\lambda} + \frac{1+3\lambda}{\lambda} - \frac{4}{1-\lambda} = -\frac{4\lambda}{1-\lambda}\,,$$

which is indeed negative.

(c) It is interesting to compute this second derivative for all values of λ. We obtain from 3 (6):

$$
\begin{aligned}
\left(\frac{\partial^2 \Gamma}{\partial p^2}\right)_{\rho=1} &= 1 - \frac{\cos\psi}{\lambda} + 2\,\frac{-2\lambda\cos\psi + 2\lambda^2}{1 - 2\lambda\cos\psi + \lambda^2}\left(\frac{1+\lambda^2}{4\lambda^2} - \frac{\cos\psi}{2\lambda}\right) \\
&\quad + 4\,\frac{d}{d\rho}\left(\frac{-2\lambda\rho\cos\psi + \lambda^2(\rho^2+1)}{\rho^2 - 2\lambda\rho\cos\psi + \lambda^2}\right)_{\rho=1}\left(\frac{1+\lambda^2}{4\lambda^2} - \frac{\cos\psi}{2\lambda}\right) \\
&\quad + \frac{2\cos\psi}{\lambda}\,\frac{-2\lambda\cos\psi + 2\lambda^2}{1 - 2\lambda\cos\psi + \lambda^2} \\
&= \frac{4\lambda(\lambda - \cos\psi)}{1 - 2\lambda\cos\psi + \lambda^2}\,.
\end{aligned}
\tag{4}
$$

Hence this second derivative is positive on the convex side of the circular arc, and negative on the concave side of this arc. On the convex side Γ is positive, and on the concave side Γ is negative, provided p is sufficiently near to the arc in question.

5. On the Green's function of the infinite domain which is the exterior of C_R. We denote now by $\Gamma(p)$ the Green's function of the infinite domain which is the exterior of C_R, having its singular point at infinity. We seek this function in the form:

$$
\begin{aligned}
\Gamma(p) &= \log\frac{R}{\rho} + A\,\frac{\rho}{R}\cos\psi + B \\
&\quad + C\,\frac{\rho^2 - \lambda^2}{\rho^2 - 2\lambda\rho\cos\psi + \lambda^2} + |z|^2\left(D + E\,\frac{R}{\rho}\cos\psi\right),
\end{aligned}
\tag{1}
$$

where A, B, C, D, E are appropriate constants depending on R and λ; here again the point p is represented by the complex number $z = re^{i\phi}$, and the relation between $z = re^{i\phi}$ and $\zeta = \rho e^{i\psi}$ is the same as above. We show that the constants $A, \cdots, E$ can be determined in a unique way so that Γ satisfies the boundary conditions of the clamped plate provided that R is sufficiently near to 1; more precisely, there must be $0 < R - 1 < \epsilon = \epsilon(\lambda)$.

The conditions

$$
\Gamma = \frac{\partial \Gamma}{\partial \rho} = 0 \qquad \text{for } \rho = R \tag{2}
$$

are equivalent to the following:

$$(3)\qquad \Gamma_1 = \frac{\partial \Gamma_1}{\partial \rho} = 0 \qquad \text{for } \rho = R,$$

where the function Γ_1 is defined by

$$\begin{aligned}\Gamma_1 &= (\rho^2 - 2\lambda\rho \cos\psi + \lambda^2)\,\Gamma \\ (4)\qquad &= \left(\log \frac{R}{\rho} + A\frac{\rho}{R}\cos\psi + B\right)(\rho^2 - 2\lambda\rho\cos\psi + \lambda^2) + C(\rho^2 - \lambda^2) \\ &\quad + \rho^2(1 - 2\lambda\rho\cos\psi + \lambda^2\rho^2)\left(D + E\frac{R}{\rho}\cos\psi\right).\end{aligned}$$

Here we used the first formula in 3 (2). Now (4) is a quadratic expression in $\cos\psi$. Conditions (3) can be replaced by the corresponding set of equations for the coefficients of $\cos\psi$ in (4). These equations are somewhat simplified if in (4) we replace $\cos\psi$ by $\rho^{-1}\cos\psi$. The resulting coefficients are:

$$\begin{aligned}M_1(\rho) &= \left(\log\frac{R}{\rho} + B\right)(\rho^2 + \lambda^2) + C(\rho^2 - \lambda^2) + D\rho^2(1 + \lambda^2\rho^2), \\ (5)\ M_2(\rho) &= \frac{A}{R}(\rho^2 + \lambda^2) - 2\lambda\left(\log\frac{R}{\rho} + B\right) - 2\lambda D\rho^2 + ER(1 + \lambda^2\rho^2), \\ M_3(\rho) &= -\frac{2\lambda A}{R} - 2\lambda ER.\end{aligned}$$

The boundary conditions are equivalent to the following set of conditions:

$$(6)\qquad M_1(R) = M_1'(R) = M_2(R) = M_2'(R) = M_3(R) = M_3'(R) = 0.$$

(b) The last of these six equations can be disregarded since $M_3(\rho)$ is independent of ρ. The resulting five equations are linear in the five unknown quantities $A, \cdots, E$. They are as follows:

$$\begin{aligned}(7)\qquad & B(R^2 + \lambda^2) + C(R^2 - \lambda^2) + DR^2(1 + \lambda^2R^2) = 0, \\ & 2RB - \frac{1}{R}(R^2 + \lambda^2) + 2CR + D(2R + 4\lambda^2R^3) = 0, \\ & \frac{A}{R}(R^2 + \lambda^2) - 2\lambda B - 2\lambda DR^2 + ER(1 + \lambda^2R^2) = 0, \\ & 2A + \frac{2\lambda}{R} - 4\lambda DR + 2E\lambda^2R^2 = 0, \quad \frac{A}{R} + ER = 0.\end{aligned}$$

In order to show that the unknowns are uniquely determined, we have to discuss the determinant of this system. As $R \longrightarrow 1$ the elements of this determinant approach those of the following determinant (the second and fourth equations are divided by 2):

$$(8) \qquad \begin{vmatrix} 0 & 1+\lambda^2 & 1-\lambda^2 & 1+\lambda^2 & 0 \\ 0 & 1 & 1 & 1+2\lambda^2 & 0 \\ 1+\lambda^2 & -2\lambda & 0 & -2\lambda & 1+\lambda^2 \\ 1 & 0 & 0 & -2\lambda & \lambda^2 \\ 1 & 0 & 0 & 0 & 1 \end{vmatrix} .$$

Subtracting here the first column from the last we obtain:

$$(1-\lambda^2) \begin{vmatrix} 0 & 1+\lambda^2 & 1-\lambda^2 & 1+\lambda^2 \\ 0 & 1 & 1 & 1+2\lambda^2 \\ 1+\lambda^2 & -2\lambda & 0 & -2\lambda \\ 1 & 0 & 0 & 0 \end{vmatrix}$$

$$= -(1-\lambda^2) \begin{vmatrix} 1+\lambda^2 & 1-\lambda^2 & 1+\lambda^2 \\ 1 & 1 & 1+2\lambda^2 \\ -2\lambda & 0 & -2\lambda \end{vmatrix}$$

$$= 2\lambda^2(1-\lambda^2) \begin{vmatrix} 1+\lambda^2 & 1-\lambda^2 \\ -2\lambda & 0 \end{vmatrix} = 4\lambda^3(1-\lambda^2) \neq 0 .$$

6. Conclusion. From 5 (7) it is obvious that the parameters $A, \cdots, E$ are rational functions of R and λ. Let λ be fixed, $0 < \lambda < 1$. Then these parameters are rational functions of R, and the evaluation of the determinant 5 (8) shows that they are regular in a certain neighborhood of $R = 1$. Incidentally, we find from 3 (6) that

$$(1) \qquad A = -E = \frac{1}{2\lambda}, \quad D = -B = \frac{1+\lambda^2}{4\lambda^2}, \quad C = 0 \qquad \text{as } R = 1.$$

Inserting $z = 0$, $\rho = 1/\lambda$, $\psi = 0$ in 5 (1), we obtain an elementary function of R which is regular at $R = 1$. It is a combination of $\log R$ and the rational functions A, B, C of R. Now this function is negative for $R = 1$ (provided λ is sufficiently small). From this the same property follows for the function 5 (1) provided R is sufficiently near to 1. This yields the desired property of the domain outside of the level curve C_R of the conformal mapping of the circular slit domain onto the exterior of the unit circle.

In order to prove the same property for all λ (and for sufficiently small values of $R-1$), we compute

$$\left(\frac{d^2\Gamma}{d\rho^2}\right)_{\rho=R,\ \psi=0}. \tag{2}$$

We note that the curve C_R intersects the real axis in two points; the curve is convex at the left point and concave at the right point of the intersection, provided $R-1$ is small enough. The second derivative (2) we consider is associated with the concave point of intersection. Now (2) has the same sign as

$$\left(\frac{d^2\Gamma_1}{d\rho^2}\right)_{\rho=R,\ \psi=0}, \tag{3}$$

where Γ_1 is defined as in 5 (4). From 5 (4) we see again that (3) is a function of R which is regular at $R=1$. Since it is negative for $R=1$, it must be negative for all $R>1$ sufficiently near to 1.

This establishes the assertions.

STANFORD UNIVERSITY

LIPSCHITZ FUNCTIONS OF CONTINUOUS FUNCTIONS

IMANUEL MARX AND GEORGE PIRANIAN

1. Introduction. The present paper was suggested by a note of W. S. Loud [2] in which the following theorem on functions of a real variable is proved.

THEOREM 1. *If* α *is a constant* $(0 < \alpha < 1)$, *there exist a continuous function* $f(t)$ *and a pair of positive constants* K_1 *and* K_2 *such that*

$$|f(t+h) - f(t)| < K_1 \, |h|^{\alpha}$$

for all t *and all* h, *and such that*

$$\limsup_{h \to 0} \frac{|f(t+h) - f(t)|}{|h|^{\alpha}} > K_2$$

for all t.

It is natural to examine the possibility of a variable exponent $\alpha(t)$ and to consider various definitions that associate with every continuous function $f(t)$ a "Lipschitz function" $\alpha(t; f)$. For a reasonable choice of the definition, Loud's result implies that every constant $\alpha\,(0 < \alpha < 1)$ is the Lipschitz function of some continuous function. The following sections offer two different definitions of Lipschitz functions, and deal with the problem of characterizing the functions that are Lipschitz functions of continuous functions.

2. The point Lipschitz function of a function. Let $f(t)$ be a continuous, real-valued function of t. Consider the quantity

$$Q(\alpha, t_0; f) = \limsup_{h \to 0} \frac{|f(t_0 + h) - f(t_0)|}{|h|^{\alpha}}.$$

If $Q(\alpha, t_0; f)$ is finite for $\alpha = \alpha'$, it is zero for all α less than α'; if Q is greater than zero for $\alpha = \alpha'$, it has the value $+\infty$ for all α greater than α'. Let $\alpha(t_0; f)$ denote the least upper bound of all α for which $Q(\alpha, t_0; f)$ is

Received November 26, 1951.
Pacific J. Math. 3 (1953), 447-459

finite. Then $\alpha(t; f)$ shall be called the *point Lipschitz function* of $f(t)$. A simple computation shows that

$$(1) \qquad \alpha(t; f) = \liminf_{h \to 0} \frac{\log |f(t+h) - f(t)|}{\log |h|},$$

where the fraction on the right is to be interpreted as having the value $+\infty$ whenever $f(t+h) = f(t)$.

THEOREM 2. *If $f(t)$ is continuous $(-\infty < t < \infty)$, then $\alpha(t; f)$ is the inferior limit of a sequence of continuous functions of t; if $\alpha(t; f) > 1$ throughout some open interval of the t-axis, then $\alpha(t; f) = \infty$ throughout the interval.*

Let $L(t, h; f)$ denote the fraction in the right member of (1), let ϵ denote a constant $(0 < \epsilon < 1/2)$, and let

$$(2) \qquad \alpha_\epsilon(t; f) = \min \left[1/\epsilon, \min_{\epsilon \le |h| \le 2\epsilon} L(t, h; f) \right].$$

Then $\alpha_\epsilon(t; f)$ is a continuous function of t, because of the restriction on h and the truncation of $L(t, h; f)$ imposed in the right member of (2). If ϵ is assigned the successive values $1/4, 1/8, 1/16, \cdots$, the first part of Theorem 2 follows from the formula (2).

For the second part of the theorem, consider an open interval I on the t-axis throughout which $\alpha(t; f) > 1$. Because $f'(t) = 0$ throughout I, $f(t)$ is constant in I, and the proof of the theorem is complete.

THEOREM 3. *Let $\{\alpha_n(t)\}$ be a sequence of continuous functions $(0 \le \alpha_n(t) \le 1)$, and let*

$$\alpha(t) = \liminf_{n \to \infty} \alpha_n(t).$$

Then there exists a continuous function $f(t)$ such that $\alpha(t; f) = \alpha(t)$.

The theorem will be proved by a construction analogous to that used by Loud in his proof of Theorem 1. Let $g(t, s)$ be the continuous function which takes the value 0 when t is an even multiple of s, takes the value 1 when t is an odd multiple of s, and is linear between consecutive multiples of s. Let α be a constant between 0 and 1, and let $A > [2(1-\alpha)]^{-1}$ be an integer. Loud proved that the series

$$\sum_{n=1}^{\infty} 2^{-2Aan} g(t, 2^{-2An}), \tag{3}$$

converges to a function $f(t)$ which has the properties promised in Theorem 1. Roughly, the principal intuitive idea behind Loud's proof is that for every pair of values t and h at most one or two terms of the series (3) make a significant contribution to the difference $f(t+h) - f(t)$; the contribution is always small enough so that the first inequality in Theorem 1 is satisfied; and for every t there exist arbitrarily small values of h for which the contribution is so large that the second inequality in Theorem 1 is satisfied. The following three sections will be devoted to elaborations of Loud's method that lead to a proof of Theorem 3. We first summarize our construction of a continuous function $f(t)$ whose Lipschitz function $\alpha(t)$ is the inferior limit of a given sequence of continuous functions $\alpha_n(t)$. The construction is then described in full detail in the following two sections.

The function $f(t)$ will be written as an infinite series

$$f(t) = \sum_{m=1}^{\infty} G_m(t),$$

with G_m depending on the function $\alpha_m(t)$ alone. For every m, we divide the t-axis into intervals I over each of which the function $\alpha_m(t)$ lies within fixed bounds to be specified. The function $G_m(t)$ is defined separately in each interval. Over a fixed interval I, the graph Γ_m of the function $G_m(t)$ consists of rows of saw-teeth completely filling I. There is a row of relatively high and wide teeth in the central portion of the interval, flanked by two rows of somewhat lower and much narrower teeth, which are in turn flanked by two rows of still lower and narrower teeth, and so on. All the teeth of the central row have equal height and equal width, all the teeth of the two flanking rows have equal height and equal width, and so on. Toward the end-points of the interval I, the heights and the widths of the teeth of Γ_m approach zero. The function $G_m(t)$ is continuous for all t, is differentiable except at a countable number of points, and is not constant in any interval. The heights and the widths of the saw-teeth are so chosen that

$$\alpha(t; f) = \alpha(t)$$

for the function

$$f(t) = \sum_{m=1}^{\infty} G_m(t).$$

Details of the proof follow.

3. Classes of intervals. For each m, we denote by I_m a class of intervals to be constructed, with I_m depending on $I_1, I_2, \cdots, I_{m-1}$ as well as on the functional values of $\alpha_m(t)$. The class I_1 consists of finitely many or infinitely many disjoint, open intervals that meet the following three requirements: each point of the t-axis lies in the closure of one of the intervals; no point of the t-axis is a limit point of end-points of intervals of the class I_1; and throughout each of the intervals one of the conditions

$$0 \leq \alpha_1(t) \leq 3/4,$$

$$1/4 \leq \alpha_1(t) \leq 1$$

is satisfied.

When the classes $I_1, I_2, \cdots, I_{m-1}$ of intervals have been defined, we choose the intervals of the class I_m subject to the following four requirements: each point of the t-axis lies in the closure of one of the intervals of the class; no point of the t-axis is a limit point of end-points of intervals of the class; throughout each interval of the class, one of the conditions

$$\begin{aligned} 0 &\leq \alpha_m(t) \leq 3/2^{m+1}, \\ 1/2^{m+1} &\leq \alpha_m(t) \leq 5/2^{m+1}, \\ 3/2^{m+1} &\leq \alpha_m(t) \leq 7/2^{m+1}, \\ &\cdots\cdots\cdots\cdots, \\ 1-3/2^{m+1} &\leq \alpha_m(t) \leq 1 \end{aligned} \tag{4}$$

is satisfied; and no end-point of an interval of the class I_m is an end-point of an interval of a previously defined class.

4. The saw-tooth functions. The function $f(t)$ to be constructed will be of the form

$$f(t) = \sum_{m=1}^{\infty} G_m(t) = \sum_{n=1}^{\infty} g_n(t), \tag{5}$$

where

$$G_1(t) = g_1 + g_3 + g_6 + g_{10} + \cdots,$$

$$G_2(t) = g_2 + g_5 + g_9 + \cdots,$$

$$G_3(t) = g_4 + g_8 + \cdots,$$

$$G_4(t) = g_7 + \cdots,$$

$$\cdots.$$

The indices $n = 1, 3, 6, 10, \cdots$, are said to belong to $m = 1$, the indices $n = 2, 5, 9, \cdots$, are said to belong to $m = 2$, and so on. The graphs of the functions $G_m(t)$ and $g_n(t)$ will be denoted by Γ_m and γ_n, respectively.

We summarize the construction of the functions $g_n(t)$. Corresponding to the function G_m we have selected a class I_m of intervals, and throughout each interval I of the class I_m the function $\alpha_m(t)$ satisfies one of 2^m inequalities specified in § 3. We choose a class of exponents c in accordance with these inequalities, so that c is fixed throughout each interval I, and a decreasing sequence $\{s_n\}$ of positive numbers, where s_n depends only on $s_1, s_2, \cdots, s_{n-1}$ and on the value of m to which the index n belongs. The function $g_n(t)$ is continuous, with a graph γ_n consisting alternately of rows of equally high and equally wide saw-teeth and of segments of the t-axis. There are at most two subintervals of each interval I of the class I_m where $g_n(t)$ differs from zero, and in these subintervals it has the form

$$g_n(t) = s_n^c\, g(t, s_n), \tag{6}$$

with the exponent c corresponding to the interval I. (The function $g(t, s)$ was defined in § 2, immediately after the statement of Theorem 3.) The terms $g_n(t)$ of the series for $G_m(t)$ are so chosen that every point interior to an interval of the class I_m lies on the base of a tooth of the graph Γ_m of $G_m(t)$. It remains to describe the choice of the exponents c in (6), to select the sequence $\{s_n\}$, and to determine the position of the teeth in γ_n.

Let $g_n(t)$ be a term of $G_m(t)$, and let I be an interval of the class I_m. The graph γ_n of $g_n(t)$ shall contain rows of equal saw-teeth over at most two subintervals of I, and the exponent c in (6) determining the height of the teeth shall have the value $1/2^{m+1}$, $3/2^{m+1}, \cdots$, or $1 - 1/2^{m+1}$, according as the function $\alpha_m(t)$ satisfies in I the first, second, $\cdots$, or last of the conditions (4). It follows that the height s_n^c of a tooth in γ_n has one of finitely many values and depends only on the range of values taken by $\alpha_m(t)$ in the interval I where the tooth appears, while the width $2s_n$ is the same for all teeth appearing anywhere in γ_n. It also follows that the derivative $g_n'(t)$ exists (except at denumerably many points), and takes only finitely many values.

The number s_1 can be chosen arbitrarily, subject to the condition $0 < s_1 < 1$. Once the numbers $s_1, s_2, \cdots, s_{n-1}$ are determined, s_n is chosen subject to the following three conditions:

1) We require that s_{n-1} be an even multiple of s_n.

2) We require that the inequality

$$s_n^{2^{-m-1}} < \frac{s_{n-1}}{10}$$

be satisfied, where m refers to the function $G_m(t)$ of which $g_n(t)$ is a term. If this requirement is met, the height of each tooth in γ_n is no greater than 1/20 the base width of any tooth in γ_r $(r = 1, 2, \cdots, n-1)$, and every tooth in γ_{n-1} is more than 10 times as high as every tooth in γ_n. It follows that the series $\sum_{n=1}^{\infty} |g_n(t)|$ converges uniformly on the interval $(-\infty, \infty)$.

3) We require finally that the slopes of the sides of the lowest teeth in γ_n be in absolute value greater than 10 times the sum of the greatest slopes that can possibly occur in γ_r $(r = 1, 2, \cdots, n-1)$. Because the side of a tooth of height s^c and width $2s$ has a slope numerically equal to $1/s^{1-c}$, this requirement is met provided s_n is chosen small enough.

We turn now to the disposition of the teeth in γ_n. Let I be any interval of the class I_1. Then γ_1 shall have as many teeth in I as possible, subject to the restriction that the distance from either end of I to any tooth of the graph shall be greater than twice the height of the tooth.

Again, if I is an interval of the class I_1, and if γ_1 has no teeth in I, then γ_3 shall have as many teeth in I as possible, subject again to the restriction mentioned above. If γ_1 has beeth in I, then γ_3 shall have, in I, two rows of teeth flanking the row of teeth of γ_1; again, the distance from either end of I to any tooth in γ_3 shall be greater than twice the height of the tooth.

Next, if I is an interval of the class I_1, then γ_6 shall have teeth in the middle portion of I provided that γ_3 has no teeth in I and I is sufficiently long. If γ_3 has teeth in I, then γ_6 shall have two rows of teeth: each of these rows shall be adjacent to a previously constructed row, and shall extend as near as possible to the nearer end of I, subject to the condition that the remaining distance be greater than twice the height of the teeth.

The construction of γ_n $(n = 10, 15, 21, \cdots)$ proceeds according to the pattern that has been described. The construction of γ_n $(n = 2, 5, 9, \cdots)$ is similar to the previous construction, with these modifications: the construction is carried out with reference to intervals of the class I_2, and with reference to the values of the function $\alpha_2(t)$; and the distance from each tooth to the end of the interval of the class I_2 in which it stands is required to be 2^2 times the height of the tooth (2^m times the height of the tooth in the construction of teeth belonging to the graph Γ_m). The construction of Γ_m is entirely independent of the construction of Γ_k $(k \neq m)$. Further details are superfluous, and we must only prove that the function defined by (6) has the required properties.

5. Arithmetical estimates. First we show that if t_0 is a fixed point, $\alpha_0 = \alpha(t_0)$, and $\epsilon > 0$, then

$$\limsup_{h \to 0} \frac{|f(t_0 + h) - f(t_0)|}{|h|^{\alpha_0 + \epsilon}} = \infty. \tag{7}$$

In other words, the point Lipschitz function of $f(t)$ has at t_0 a value not greater than α_0.

Let m be any integer such that t_0 is interior to an interval of the class I_m (at most one positive integer m fails to satisfy this requirement); then there exists an integer n such that $g_n(t)$ is a term in the series defining $G_m(t)$ and such that t_0 lies on the base of a tooth of the graph γ_n of $g_n(t)$. Therefore $g_n(t)$ is linear in a sufficiently small interval with t_0 as end-point; that is, there exists a number h, with

$$\frac{1}{2} s_n \leq |h| < s_n, \tag{8}$$

such that the function $g_n(t)$ is linear in the interval joining the points $t = t_0$ and $t = t_0 + h$. For this number h we have further

$$|G_m(t_0 + h) - G_m(t_0)| = |g_n(t_0 + h) - g_n(t_0)| = |h| \; s_n^{c-1}$$

where

$$c < \alpha_0 + 2^{-m} + 2^{-m-1}. \tag{9}$$

For all teeth that cover the segment joining t_0 and $t_0 + h$ and belong to graphs γ_r with $r < n$, the requirement 3) on $\{s_n\}$ implies that

$$s_n^{c-1} > 10 \sum_{r<n} \max |g_r'(t)|,$$

and therefore that

$$\sum_{r<n} |g_r(t_0+h) - g_r(t_0)| < |h|\, s_n^{c-1}/10.$$

For the functions g_r with $r > n$, the requirement 2) on $\{s_n\}$ gives

$$\sum_{r>n} |g_r(t_0+h) - g_r(t_0)| < s_n\,(1/10 + 1/100 + \cdots) < s_n/5.$$

From (8) and from the estimates obtained thus far it follows that

$$\frac{|f(t_0+h) - f(t_0)|}{|h|^{\alpha_0+\epsilon}} > \frac{|h|\, s_n^{c-1} - |h|\, s_n^{c-1}/10 - s_n/5}{|h|^{\alpha_0+\epsilon}}$$

$$> s_n^c\,(9/20 - 1/5)\, s_n^{\alpha_0+\epsilon} = \frac{1}{4}\, s_n^{c-\alpha_0-\epsilon}.$$

By (9), the exponent in the last member is less than $-\epsilon/2$ if m is large enough, and therefore the relation (7) is established.

Secondly we must prove that, for every $\epsilon > 0$,

$$\limsup_{h\to 0} \frac{|f(t_0+h) - f(t_0)|}{|h|^{\alpha_0-\epsilon}} = 0. \tag{10}$$

We choose an integer m_0 such that $2^{-m} + 2^{-m-1} < \epsilon$ for all $m > m_0$, and a positive quantity h_0 such that the interval

$$t_0 - h_0 \le t \le t_0 + h_0$$

contains no end-point of an interval of any of the classes $I_1, I_2, \cdots$, or I_{m_0}, except possibly the point t_0 itself. Without restricting the generality of the proof, we suppose that $|h| < h_0$.

To establish (10), we make separate estimates of the variation of $G_m(t)$

for the following three cases: the point t_0 is an end-point of an interval of the class I_m; t_0 is not an end-point, and $m \leq m_0$; or t_0 is not an end-point, and $m > m_0$.

If t_0 is an end-point of an interval of the class I_m, then the disposition of the saw-teeth ensures that

$$|G_m(t_0 + h) - G_m(t_0)| < |h|/2^m. \tag{11}$$

For each $m \leq m_0$ for which t_0 is not an end-point of an interval of the class I_m, the quantity $G_m(t_0 + h) - G_m(t_0)$ can be written as the sum of finitely many terms of the form $g_n(t_0 + h) - g_n(t_0)$; the set of indices n that occur in this sum depends only on h_0, not on h. Since the corresponding derivatives $g_n'(t)$ are bounded, it follows that, for every $\alpha < 1$,

$$\lim_{h \to 0} \sum_{m \leq m_0} \frac{|G_m(t_0 + h) - G_m(t_0)|}{|h|^{\alpha}} = 0.$$

For $m > m_0$ and $|h| < h_0$, two possibilities arise, for each index m: either t_0 and $t_0 + h$ both lie in the same interval of the class I_m, or they do not lie in the same interval. In the latter case, the inequality (11) holds, and it is therefore sufficient to discuss the former case.

Let k be the least integer such that $s_k \leq |h|$ and such that the term $g_k(t)$ occurs in the series defining a function $G_m(t)$ with $m > m_0$. The choice of the exponents in the definition (6) of $g_n(t)$, and the requirement 2) on the sequence $\{s_n\}$, imply the inequality

$$\sum_{r \geq k} |g_r(t_0 + h) - g_r(t_0)| < 2\, s_k^c \leq 2\, |h|^c,$$

and the quantity c is greater than $\alpha_0 - \epsilon$ because of our choice of m_0. Finally we estimate the contribution from those terms $g_r(t)$, occurring in the series that define functions $G_m(t)$ with $m > m_0$, for which $s_r > |h|$ while t_0 and $t_0 + h$ lie in the same interval of the class I_m. Let p be the greatest value of the index r for such terms. We find

$$|g_p(t_0 + h) - g_p(t_0)| < |h|\, s_p^{c-1} < |h|^c\, |s_p/h|^{c-1} < |h|^c,$$

and the sum of the remaining terms of the same category is less than half of the last member. The inequality (10) is established, and the proof of Theorem 3 is complete.

We observe that Theorem 3 no longer holds if the restriction $0 \leq \alpha_n(t) \leq 1$ is removed. For if $f(t)$ is a continuous function, the set of points where $\alpha(t; f) \leq 1$ cannot have isolated points. The complete characterization of the functions that are point Lipschitz functions of continuous functions appears to be difficult.

6. The local Lipschitz function of a function. Let $f(t)$ be a continuous, real-valued function of t, and let h be a variable taking positive values. Denote by $B(t_0, h; f)$ the least upper bound of all numbers β for which the quantity

$$\frac{|f(t'') - f(t')|}{(t'' - t')^{\beta}}$$

remains bounded as long as the variables t' and t'' satisfy the restriction $t_0 - h \leq t' < t'' \leq t_0 + h$. For each value t_0, $B(t_0, h; f)$ is a nonincreasing function of h. The quantity

$$\beta(t) = \beta(t; f) = \lim_{h \to 0+} B(t, h; f) \tag{12}$$

shall be called the *local Lipschitz function* of $f(t)$.

It follows at once from the definition that

$$\beta(t; f) \leq \alpha(t; f),$$

for every continuous function $f(t)$. That equality does not always occur is seen from the following example. Consider the function

$$f(t) = t \sin 1/t \ (t \neq 0), \ f(0) = 0.$$

In every closed interval that does not contain $t = 0$, $f(t)$ has a bounded derivative, so that $\alpha(t; f) \geq 1$ for $t \neq 0$. Since $\alpha(0; f) = 1$, it follows that the point Lipschitz function of $f(t)$ is everywhere equal to 1 [except at the zeros of $f'(t)$, where $\alpha(t; f) = 2$]. On the other hand, the local Lipschitz function $\beta(t; f)$ has the value 1 everywhere except at $t = 0$, and $\beta(0; f) = 1/2$ (for details, see [1]). It follows that equations (1) and (12) do not define equivalent Lipschitz functions.

7. Characterization theorems. The following two theorems provide a characterization of bounded local Lipschitz functions of continuous functions.

THEOREM 4. *If $f(t)$ is continuous $(-\infty < t < \infty)$, then*

i) *$\beta(t; f)$ is lower semi-continuous;*

ii) *for each point t, either $0 \le \beta(t; f) \le 1$, or $\beta(t; f) = \infty$;*

iii) *the set of points t where $\beta(t; f) \ne \infty$ is a perfect set.*

Suppose that $\beta(t) = \beta(t; f)$ is the local Lipschitz function of a continuous function $f(t)$. If $\beta(t_0) > 1$, the point t_0 is interior to an interval in which $f(t)$ satisfies a Lipschitz condition with an exponent greater than 1. Then $f(t)$ is constant in an interval about t_0, so that $\beta(t_0; f) = \infty$. Part ii) of the theorem is proved.

The set of points where $0 \le \beta(t; f) \le 1$ cannot have isolated points, and the set of points where $\beta(t; f) = \infty$ is open. Therefore part iii) of the theorem is proved, as well as part i) for those points t where β is infinite. Finally, if $\beta(t_0) \le 1$, for every $\epsilon > 0$ there exists an interval $|t - t_0| < \delta$ in which

$$\beta(t) > \beta(t_0) - \epsilon.$$

It follows that

$$\beta(t_0) \le \liminf_{t \to t_0} \beta(t)$$

for all t_0, and the proof of the theorem is complete.

THEOREM 5. *Let $\beta(t)$ $(0 \le \beta(t) \le 1)$ be a lower semi-continuous function. Then there exists a strictly increasing continuous function $F(t)$ such that*

$$\beta(t; F) = \beta(t).$$

Let $\{I_r\}$ denote a sequence of closed intervals on the t-axis with the following property: for each point t_0 and for every $\epsilon > 0$, there exists an interval I_r of length less than ϵ, covering t_0. In each interval I_r we select a point t_r at which $\beta(t)$ assumes its minimum value in the interval. The function $F(t)$ will be chosen as an infinite series

$$F(t) = \sum_{r=1}^{\infty} f_r(t).$$

If the point t_r coincides with one of the points $t_1, t_2, \cdots, t_{r-1}$, we choose

$f_r(t) \equiv 0$; otherwise we choose $f_r(t)$ as a strictly increasing function of t whose local Lipschitz function has the value $\beta(t_r)$ for $t = t_r$ and the value 1 everywhere else.

The term $f_r(t)$ is constructed from the function

$$f(t;\, \gamma) = \frac{|t|^{\gamma} \operatorname{sgn} t}{1 + |t|^{\gamma}}\,.$$

This function is strictly increasing; for

$$f'(t;\, \gamma) = \frac{|t|^{\gamma-1}}{(1 + |t|^{\gamma})^2}\,,$$

when $t \neq 0$, and the function is continuous at $t = 0$. Furthermore, the local Lipschitz function of $f(t;\, \gamma)$ has the value γ at $t = 0$, since

$$(13) \qquad |f(t'';\, \gamma) - f(t';\, \gamma)| \le 2\,|t'' - t'|^{\gamma}$$

for all distinct t' and t''; and it has the value 1 everywhere else, since $f'(t;\, \gamma)$ is continuous for $|t| > 0$. In order to adapt $f(t;\, \gamma)$ to our needs we require a sequence $\{\rho_r\}$ of positive numbers, chosen as follows. We set $\rho_1 = 1$. If t_r coincides with $t_1, t_2, \cdots$, or t_{r-1}, the function $f_r(t)$ is identically zero, and no number ρ_r is needed. If all the quantities $|t_r - t_1|, |t_r - t_2|, \cdots, |t_r - t_{r-1}|$ $(r > 1)$ exceed 1, we set $\rho_r = 1$; otherwise we set ρ_r equal to the least of these quantities.

The nonzero terms $f_r(t)$ of the function $F(t)$ are given by the formula

$$(14) \qquad f_r(t) = 2^{-r}\, \rho_r\, f\left(\frac{t - t_r}{\rho_r}\,;\ \beta(t_r)\right).$$

We prove first that

$$(15) \qquad \beta(t_0;\, F) \le \beta(t_0),$$

for each point t_0. For every $h > 0$, the interval

$$t_0 - h < t < t_0 + h$$

has a subinterval I_r which contains the point t_0. Since

$$\beta(t_r) \le \beta(t_0),$$

and since the function $F(t)$ does not satisfy in I_r a Lipschitz condition with an exponent greater than $\beta(t_r)$, the relation (15) is established.

Secondly, we show that

$$\beta(t_0; F) \geq \beta(t_0) - \epsilon \tag{16}$$

for each point t_0 and every $\epsilon > 0$. Because $\beta(t)$ is lower semi-continuous,

$$\beta(t) > \beta(t_0) - \epsilon$$

in some interval $|t - t_0| < \delta$. We choose a pair of integers r_1 and r_2 $(r_2 > r_1)$ such that

$$t_0 - \delta < t_{r_1} < t_0 < t_{r_2} < t_0 + \delta,$$

and denote by I the closed interval (t_{r_1}, t_{r_2}). We make separate estimates for those terms $f_r(t)$ for which t_r lies in I, for those terms with $r > r_2$ for which t_r is exterior to I, and for those terms with $r < r_2$ for which t_r is exterior to I.

By (13) and (14), the inequality

$$|f_r(t'') - f_r(t')| \leq 2^{1-r}\ |t'' - t'|^{\beta(t_r)}$$

holds for all distinct t' and t''. This inequality implies that the sum of all terms $f_r(t)$ for which t_r lies in I satisfies throughout I a Lipschitz condition with exponent $\beta(t_0) - \epsilon$. If $r > r_2$ and if t_r does not lie in I, then $f_r'(t) < 2^{r-1}$ in I, so that the sum of all terms $f_r(t)$ corresponding to such values of r has a bounded derivative in I, that is, satisfies throughout I a Lipschitz condition with exponent 1. Finally, let I' be a subinterval of I containing t_0 and sufficiently small to exclude all points t_r for which $r < r_2$, except those coinciding with t_0. The sum of the corresponding terms $f_r(t)$ also has a bounded derivative in I'. The inequality (16) is established, and the proof of Theorem 5 is complete.

REFERENCES

1. J. H. Curtis *et al.*, Amer. Math. Monthly 47 (1940), p. 53, and Amer. Math. Monthly 48 (1941), pp. 413-414.

2. W. S. Loud, *Functions with prescribed Lipschitz condition*, Proc. Amer. Math. Soc. 2 (1951), pp. 358-360.

UNIVERSITY OF MICHIGAN

THE SPHERICAL CURVATURE OF A HYPERSURFACE IN EUCLIDEAN SPACE

T. K. Pan

1. Introduction. Let V_n be a hypersurface immersed in a Euclidean space S_{n+1}. Let P be a point of V_n corresponding to the point P' of the hyperspherical representation G_n of V_n. Let V denote the extension of a region ϕ of V_n, and V' the extension of the corresponding hyperspherical region ϕ' of G_n. If the region around P tends to zero, the ratio V'/V tends to a limit Γ, which is called the *spherical curvature* of V_n at P [1, pp. 258-261]. It is found that $\Gamma = |\Omega/g|$, where $g = |g_{ij}|$ and $\Omega = |\Omega_{ij}|$ are respectively the determinants of the coefficients of the first and the second fundamental forms of V_n. In this note, some properties of the spherical curvature are studied, and new interpretations of the Gaussian curvature are derived.

The notation of Eisenhart [2] will be used for the most part.

2. Some properties. Let a real and analytic hypersurface V_n be defined by

$$y^\alpha = y^\alpha(x^1, \cdots, x^n) \qquad (\alpha = 1, \cdots, n+1),$$

referred to a Cartesian coordinate system y^α in a Euclidean space S_{n+1}. Let a vector-field v in V_n be defined by

$$v^\alpha = p^i \partial y^\alpha / \partial x^i \qquad (i = 1, \cdots, n),$$

where the v^α are real and analytic functions of the x^i. Let C be a curve of V_n. The normal curvature vector of v with respect to C at P is defined as the normal component of the derived vector of the vector-field v along C at P [3]. Let κ denote a nonzero extreme value of the magnitudes of the normal curvature vectors of v with respect to all curves of V_n at P. Then κ, which is called a principal curvature of v at P, is defined by

$$(2.1) \qquad |\Psi_{ij} - \kappa^2 g_{ij}| = 0,$$

Received May 13, 1952.
Pacific J. Math. 3 (1953), 461-466

where

$$\Psi_{ij} = \Omega_{ik}\,\Omega_{jl}p^k p^l / g_{kl}p^k p^l.$$

Since $||\Psi_{ij}||$ is of rank 1, there is one such extreme corresponding to a vector-field v. Its value is evidently equal to

$$\kappa = (\Psi_{ij}\, g^{ij})^{1/2} = (H_{ij}\, p^i\, p^j / g_{ij}\, p^i\, p^j)^{1/2}, \tag{2.2}$$

where H_{ij} is the fundamental tensor of the hyperspherical representation G_n.

The extreme of the principal curvature of a vector-field v at P, as the field varies, is defined by

$$|H_{ij} - \kappa^2 g_{ij}| = 0. \tag{2.3}$$

There are n such extremes $\bar{\kappa}_i$ corresponding to the principal directions for the tensor H_{ij}. Their product is found to be

$$\prod_{i=1}^{n} \bar{\kappa}_i = |H/g|^{1/2} = |\Omega/g|,$$

since $H = |H_{ij}| = \Omega^2/g$, [1, p. 260]. The principal directions for the tensor H_{ij} and those determined by the tensor Ω_{ij} are identical, since the principal curvature of a principal vector-field can easily be shown equal to the normal curvature of the corresponding line of curvature. Hence we have:

THEOREM 2.1. *The spherical curvature of a V_n at P is equal to the product of the extreme principal curvatures of vector-fields in V_n at P, which is the same as the product of principal curvatures of V_n at P.*

Since S_{n+1} is Euclidean, the equations of Gauss are

$$R_{ijkl} = \Omega_{ik}\,\Omega_{jl} - \Omega_{il}\,\Omega_{jk}\ . \tag{2.4}$$

Multiplying (2.4) by g^{ik} and summing with respect to i and k, we obtain

$$H_{jl} = M\Omega_{jl} + R_{jl}, \tag{2.5}$$

where M is the mean curvature of V_n, and where R_{jl} is the Ricci tensor. When V_n is a minimal hypersurface, we have $M = 0$, and the Ricci tensor is identical with the fundamental tensor of G_n. If $M \neq 0$, we have

$$H_{ij}\, p^i\, p^j = R_{ij}\, p^i\, p^j \tag{2.6}$$

if and only if v is an asymptotic vector-field. If v is a unit asymptotic vector-field, we notice, from (2.2), (2.6), and the equality

$$R_{ij}\, \lambda_{h|}{}^{i}\, \lambda_{h|}{}^{j} = - \sum_{k=1}^{n} \gamma_{hk},$$

that the square of the principal curvature of v at P is numerically equal to the sum of the Riemannian curvatures determined by v and $n-1$ other mutually orthogonal unit vectors orthogonal to v at P. Hence we have established the following result:

THEOREM 2.2. *The square of the principal curvature of an asymptotic vector-field at P in V_n is numerically equal to the mean curvature of V_n at P for the corresponding asymptotic direction.*

The extreme of the principal curvatures κ of asymptotic vector-fields at P in V_n is defined by

$$|R_{ij} - \kappa^2\, g_{ij}| = 0.$$

There are n such extreme values corresponding to the principal directions for the Ricci tensor R_{ij}. Their product is evidently equal to $|\Omega/g|$, if V_n is minimal. Hence we have:

THEOREM 2.3. *The principal curvatures of asymptotic vector-fields at P in V_n attain their extreme values in the principal directions for the Ricci tensor.*

THEOREM 2.4. *The spherical curvature of a minimal V_n at P is the product of the principal curvatures of the n vector-fields at P corresponding to the principal directions for the Ricci tensor.*

3. The Gaussian curvature. When $n = 2$, Γ is called the spherical curvature of a surface S in an ordinary space. It coincides in absolute value with the Gaussian curvature K of S. The principal curvature of a vector-field v in V_n for $n = 2$ coincides in absolute value with the principal curvature of v in S, [3]. The extreme principal curvatures of vector-fields in V_n for $n = 2$ coincide in absolute value with the principal curvatures of S. The mean curvature of V_n for

$n = 2$ is identical with the Gaussian curvature of S. Hence Theorems 2.1 and 2.2 lead directly to the following new interpretations of the Gaussian curvature:

THEOREM 3.1. *The Gaussian curvature of S at P is the product of the extreme principal curvatures of vector fields of S at P, and is the negative of the square of the magnitude of the Gaussian representation of a unit arc along an asymptotic line from P in S.*

Let p^α and q^α be two distinct conjugate vector fields in S. Then we have

$$q^\beta = e^{\beta\mu}\, d_{\alpha\mu}\, p^\alpha \qquad (\alpha,\beta,\mu = 1, 2),$$

where $d_{\alpha\mu}$ is the second fundamental tensor of S. The principal curvatures of the vector-fields p^α and q^α are respectively equal to

$$e\rho_p = (h_{\alpha\beta}\, p^\alpha p^\beta / g_{\alpha\beta}\, p^\alpha p^\beta)^{1/2},$$

$$e\rho_q = (h g_{\alpha\beta}\, p^\alpha p^\beta / g h_{\alpha\beta}\, p^\alpha p^\beta)^{1/2},$$

where $h_{\alpha\beta}$ is the third fundamental tensor of S. Hence their product is

$$(3.1) \qquad (e\rho_p)(e\rho_q) = (h/g)^{1/2}.$$

The expression on the right side of (3.1) is equal to eK, where e is $+1$ or -1 according as K is positive or negative at the point under consideration. At an elliptic point, the principal curvatures of all vector-fields are of the same sign. At a hyperbolic point, the principal curvatures of two vector-fields are different in sign if they lie in different sections separated by the asymptotic lines of S. Consequently, the principal curvatures of two conjugate vector-fields have opposite signs, since conjugate directions are separated by the asymptotic directions of the surface. Hence at an elliptic point of S, the product of the principal curvatures of two conjugate vector-fields is positive; while at a hyperbolic point of S, it is negative. At a parabolic point the normal curvature of any vector-field with respect to any curve is zero. We may consider that every direction in S at a parabolic point is both an asymptotic direction and a principal direction of a vector-field which is to be considered. Hence at a parabolic point the principal curvature of any vector-field is zero; consequently, the product of the principal curvatures of two conjugate vector-fields is zero. Thus the following theorem is proved:

THEOREM 3.2. *The Gaussian curvature of S at P is the product of the principal curvatures of any two distinct conjugate vector-fields in S at P.*

The sum of the squares of the principal curvatures of the two conjugate vector-fields is found to be

$$(e\rho_p)^2 + (e\rho_q)^2 = M(\kappa_p + \kappa_q) - 2K,$$

where κ_p and κ_q are the normal curvatures of the curves of the two fields, and where M is the mean curvature of S. By Theorem 3.2 the above equation can be written as

$$(e\rho_p + e\rho_q)^2 = M(\kappa_p + \kappa_q). \tag{3.2}$$

Since the product of the normal radii at a point in conjugate directions is a maximum for characteristic lines, and a minimum for lines of curvature, and since the sum of normal radii in conjugate directions is constant, we obtain from (3.2) the following result:

THEOREM 3.3. *The sum of the principal curvatures of two conjugate vector-fields at P is the mean proportional between the mean curvature at P of S and the sum of the normal curvatures in the two conjugate directions at P. The square of the sum of the principal curvatures of two conjugate vector-fields at P is a maximum for the principal vector-fields of S, and a minimum for the characteristic vector-fields of S.*

Let m $(m > 2)$ directions be such that the angle of two adjoining directions is $2\pi/m$. Let the principal curvatures of the vector-fields in such directions be denoted by $e\rho_1, e\rho_2, \cdots, e\rho_m$. Then

$$\frac{1}{m}\sum_{i=1}^{m>2} (e\rho_i)^2 = \frac{1}{2}M^2 - K,$$

since

$$\frac{1}{m}\left(\sum_{i=1}^{m>2} \kappa_{p_i}\right) = \frac{1}{2}M,$$

where κ_{p_i} are the normal curvatures of the curves of the corresponding vector-fields.

THEOREM 3.4. *One mth of the sum of the squares of the principal curvatures of m (>2) vector-fields at P, such that the angle of two adjoining vectors of these fields at P is $2\pi/m$, is constant and is the same for any m greater than two. The constant is half of the square of the mean curvature of S minus the Gaussian curvature of S at P.*

It is easy to prove that the principal direction of a vector-field in S is orthogonal to the curve of the field if and only if the vector-field is an asymptotic field. Let p^{α} be an asymptotic vector-field in S. Then its orthogonal trajectories are defined by

$$du^{\beta} = e^{\beta\mu} g_{\alpha\mu} p^{\alpha}.$$

The principal curvature of the asymptotic vector-field p^{α} is given by

$$(e\rho_p) = d_{\alpha\beta} p^{\alpha} e^{\beta\mu} g_{\gamma\mu} p^{\gamma} \Big/ [(g_{\alpha\beta} p^{\alpha} p^{\beta})\ (g_{\alpha\beta} e^{\alpha\mu} g_{\gamma\mu} p^{\gamma} e^{\beta\lambda} g_{\sigma\lambda} p^{\sigma})]^{1/2},$$

which after simplification becomes

$$(e\rho_p) = \epsilon^{\beta\mu} d_{\alpha\beta} g_{\gamma\mu} p^{\alpha} p^{\gamma} \Big/ g_{\alpha\beta} p^{\alpha} p^{\beta} = \tau_g ,$$

where τ_g is the geodesic torsion of the curve of the asymtotic vector-field.

THEOREM 3.5. *The principal curvature of an asymptotic vector-field at P in S is equal to the geodesic torsion at P of the curve of the field, or simply the torsion at P of the corresponding asymptotic line.*

From Theorem 3.1 and Theorem 3.5 we immediately obtain the first part of the theorem of Enneper, that the square of the torsion of a real asymptotic line at a point is equal to the absolute value of the total curvature of the surface at the point. By the second part of the same theorem we notice that *the principal curvatures of the asymptotic vector-fields in S are different in sign.*

REFERENCES

1. T. Levi-Civita, *The absolute differential calculus*, Blackie & Son, London, 1947.

2. L. P. Eisenhart, *Riemannian geometry*, Princeton University Press, Princeton, 1949.

3. T. K. Pan, *Normal curvature of a vector field*, Amer. J. Math. 74 (1952), 955-966.

UNIVERSITY OF CALIFORNIA, BERKELEY

ON SELF-ADJOINT DIFFERENTIAL EQUATIONS OF SECOND ORDER

RUTH LIND POTTER

Introduction. This paper is concerned with the behavior near $x = \infty$ of solutions of the self-adjoint differential equation

$$(1) \qquad [r(x)y']' + p(x)y = 0,$$

where $r(x) > 0$ and $r(x)$ and $p(x)$ are continuous for positive values of x. A solution is said to oscillate near $x = \infty$ if it has no largest zero. We study the oscillation and boundedness of solutions of equations of the form (1). Repeated use is made throughout the paper of the Sturm comparison and separation theorems and of two theorems due to Leighton [6; 5]. Leighton's theorems are the following.

THEOREM L_1. *If $r(x)$ and $p(x)$ are continuous and $r(x) > 0$ on the interval $0 < x < \infty$, and*

$$\lim_{x\to\infty} \int_1^\infty \frac{dx}{r(x)} = \infty \quad \text{and} \quad \lim_{x\to\infty} \int_1^x p(x)dx = \infty,$$

then every solution of (1) vanishes infinitely often on the interval $(1, \infty)$.

THEOREM L_2. *If $r(x)$ and $p(x)$ are continuous, and $r(x)p(x)$ is a positive monotone function of x for x large, a necessary condition that solutions of (1) be oscillatory near $x = \infty$ is that not both limits*

$$\lim_{x\to\infty} \int_1^x \frac{dx}{r(x)}, \qquad \lim_{x\to\infty} \int_1^x p(x)dx$$

exist and are finite.

We proceed to the study of conditions under which solutions of equation (1)

Received May 5, 1952. The author is indebted to Professor Walter Leighton for helpful suggestions in the preparation of this paper. Part of the work was done while the author was employed under contract N9onr-95100 with the Office of Naval Research.

Pacific J. Math. 3 (1953), 467-491

are oscillatory.

1. Oscillation theorems. In this first section we consider the so-called "normal" form of equation (1) in which $r(x) \equiv 1$. It will be useful to set

$$p(x) = h^{-2}(x),$$

where $h(x)$ is positive and of class C^2 when $x > a > 0$. Equation (1) then becomes

$$y'' + h^{-2}(x)y = 0. \tag{1.1}$$

To study the oscillation of solutions of equation (1.1), it is useful to consider also the equations

$$[h^2(x)z']' + z = 0, \tag{1.2}$$

$$[h(x)\eta']' + \left[\frac{1}{h(x)} - \frac{h'^2(x)}{4h(x)} + \frac{h''(x)}{2}\right]\eta = 0, \tag{1.3}$$

$$[h(x)\zeta']' + \left[\frac{1}{h(x)} - \frac{h'^2(x)}{4h(x)} - \frac{h''(x)}{2}\right]\zeta = 0. \tag{1.4}$$

Nonnull solutions of these four differential equations are oscillatory[1] or nonoscillatory simultaneously, for one may readily verify that the derivative of a solution of (1.1) is a solution of (1.2), equation (1.3) is obtained from (1.1) by the substitution $\eta = h^{-1/2}(x)y$, and (1.4) is obtained from (1.2) by the substitution $\zeta = h^{1/2}(x)z$.

We define

$$H_1(x) = \left[\frac{1}{h(x)} - \frac{h'^2(x)}{4h(x)} + \frac{h''(x)}{2}\right] \tag{1.5}$$

and

$$H_2(x) = \left[\frac{1}{h(x)} - \frac{h'^2(x)}{4h(x)} - \frac{h''(x)}{2}\right]. \tag{1.6}$$

[1]A solution is said to be oscillatory on an interval if it vanishes infinitely often on the interval.

It follows from Sturm's comparison theorem that if $H_1(x) \leq 0$ or $H_2(x) \leq 0$ for large values of x, the solutions of (1.1) are nonoscillatory. Similarly, it follows from Theorem L_1 that if[2]

$$\lim \int_a^x \frac{dx}{h(x)} = \infty,$$

solutions of (1.1) are oscillatory if

$$\lim \int_a^x H_1(x)dx = +\infty,$$

or if

$$\lim \int_a^x H_2(x)dx = +\infty.$$

We proceed with a proof of the following result.

THEOREM 1.1. *If*

$$\lim \int_a^x H_2(x)dx = +\infty,$$

the solutions of (1.1) *are oscillatory.*

Note first that $\lim \int_a^x h^{-1}(x)dx$ cannot be finite, for then $h'(x) \longrightarrow -\infty$; and $h(x)$ could not be positive, as assumed. An application of Theorem L_1 completes the proof of the theorem.

The following lemma will be useful in the sequel.

LEMMA 1.2. *If $h(x)$ is a positive monotone function, a necessary condition that solutions of* (1.1) *be oscillatory is that*

$$\lim \int_a^x \frac{dx}{h(x)} = \infty.$$

To prove the lemma let us suppose that its conclusion is false; that is, suppose

$$\lim \int_a^x \frac{dx}{h(x)} < \infty.$$

[2]All limits taken in this paper will be limits as $x \longrightarrow \infty$. Unless otherwise indicated, a is a suitably chosen positive number.

Then by a well-known theorem on infinite integrals, $\lim xh^{-1}(x) = 0$, so that for any fixed value of n, $h^{-1}(x) < (nx)^{-1}$, for x sufficiently large. Since solutions of the equation

$$y'' + (nx)^{-2} y = 0$$

are nonoscillatory whenever $n \geq 2$, an application of Sturm's comparison theorem yields the contradiction, and the truth of the lemma is established.

THEOREM 1.2. *If*

$$\lim \int_a^x H_1(x)\,dx = \infty,$$

a necessary and sufficient condition that the solutions of (1.1) *be oscillatory is that*

$$\lim \int_a^x \frac{dx}{h(x)} = \infty.$$

The sufficiency of the condition follows from Theorem L_1 applied to equation (1.3).

To prove the necessity, let us suppose that

$$\lim \int_a^x \frac{dx}{h(x)} < \infty.$$

Since

$$\lim \int_a^x H_1(x)\,dx = \infty,$$

it is readily seen that $\lim h'(x) = +\infty$, and hence $h(x)$ is monotone for large values of x. It follows from Lemma 1.2 that then the solutions of (1.1) are nonoscillatory, contrary to the hypothesis.

The proof of the theorem is complete.

THEOREM 1.3. *If* $\lim \int_a^x h^{-1}(x)\,dx = \infty$ *and, for large values of* x,

$$[h'(x)]^2 \leq k^2 < 4,$$

solutions of (1.1) *are oscillatory.*

Under the hypotheses of the theorem,

$$\lim \int_a^x H_1(x)\,dx \geq \lim \left[\frac{1}{2} h'(x) - \frac{1}{2} h'(a) + \int_a^x \left(1 - \frac{1}{4k^2}\right) \frac{dx}{h(x)} \right] = \infty,$$

so that Theorem L_1 implies that solutions of (1.3), and hence that solutions of (1.1), are oscillatory.

THEOREM 1.4. *If $h(x)H_2(x)$ is a positive monotone function, a necessary condition that solutions of* (1.1) *be oscillatory is that*

$$\lim \int_a^x \frac{dx}{h(x)} = +\infty.$$

To prove the theorem note first that it follows from Theorem L_2 that not both the limits

$$\lim \int_a^x \frac{dx}{h(x)}, \qquad \lim \int_a^x H_2(x)\,dx$$

can be finite. Suppose the conclusion of the theorem were false. Then the positiveness of $H_2(x)$ would imply that the second limit above would also be finite. From this contradiction we may infer the truth of the theorem.

The following result is useful in the application of the theory.

THEOREM 1.5. *If* $\lim h'(x) = L$ *exists, solutions of* (1.1) *are nonoscillatory if $L > 2$, and oscillatory if $L < 2$.*[3]

This theorem is proved by using Sturm's comparison theorem with the aid of the relation

$$h(x) = h(a) + \int_a^x h'(x)\,dx.$$

If $L = 2$, solutions may or may not be oscillatory depending on $h(x)$, as the following example shows.

EXAMPLE 1.1. For the equation

$$y'' + \frac{a^2 + 1/4 \log^2 x}{x^2 \log^2 x}\, y = 0,$$

[3] Part of this theorem is contained in a theorem of Hartman and Wintner [1].

we have

$$h(x) = \frac{x \log x}{(a^2 + 1/4 \log^2 x)^{1/2}}$$

and $\lim h'(x) = 2$, whereas the solutions of the equation are oscillatory or not according as $a^2 > 1/4$ or $a^2 \leq 1/4$.

COROLLARY 1.5. *If* $\lim \int_a^x h^{-1}(x)\,dx = \infty$, *and* $H_1(x)$ *and* $H_2(x)$ *are nonnegative and not identically zero for large values of* x, *a necessary condition that solutions of* (1.1) *be nonoscillatory is* $\lim h'(x) = 2$.

If

$$\lim \int_a^x \frac{dx}{h(x)} = \infty,$$

and either

$$\lim \int_a^x H_1(x)\,dx = \infty$$

or

$$\lim \int_a^x H_2(x)\,dx = \infty,$$

application of Theorem L_1 to (1.3) or (1.4), as the case may be, shows that solutions of (1.1) are oscillatory. Therefore, if solutions of (1.1) are assumed nonoscillatory,

$$\lim \int_a^x H_1(x)\,dx < \infty$$

and

$$\lim \int_a^x H_2(x)\,dx < \infty,$$

in which case $\lim h'(x)$ may be seen to exist. Since

$$\int_a^x H_1(x)\,dx = \frac{1}{2}h'(x) - \frac{1}{2}h'(a) + 1/4\int_a^x \frac{1}{h(x)}\,[4 - h'^2(x)]\,dx,$$

$$\int_a^x H_2(x)\,dx = -\frac{1}{2}h'(x) + \frac{1}{2}h'(a) + 1/4\int_a^x \frac{1}{h(x)}\,[4 - h'^2(x)]\,dx,$$

and the limit of the difference of the two integrals exists, lim $h'(x)$ exists. Moreover, since

$$0 < \int_a^x [H_1(x) + H_2(x)]\,dx,$$

lim $h'^2(x) \leq 4$. Therefore, by Theorem 1.5, lim $h'(x) = 2$, and the corollary is established.

An extension of Theorem 1.5 to the more general equation (1) can be made if either

$$\lim \int_a^x \frac{dx}{r(x)} = \infty$$

or

$$\lim \int_a^x p(x)\,dx = \infty.$$

We assume that $r(x) > 0$ and $p(x) \geq 0$, and that $r(x)$ and $p(x)$ are functions of class C' when $0 < a < x$.

THEOREM 1.6. *If*

$$\lim \int_a^x \frac{dx}{r(x)} = \infty,$$

and

$$\lim r(x)\frac{d}{dx}[r(x)p(x)]^{-1/2} = L,$$

the solutions of (1) *are oscillatory if* $L < 2$, *and nonoscillatory if* $L > 2$.

Transforming equation (1) by the substitution

$$t = \int_a^x \frac{dx}{r(x)}$$

leads to the equation

$$\frac{d^2y}{dt^2} + r(x)p(x)y = 0.$$

The theorem follows immediately upon application of Theorem 1.5 to this equation. (Note that $L < 0$ is incompatible with the assumption $r(x)p(x) \geq 0$.)

THEOREM 1.7. *If $p(x)$ is positive for large values of x, and*

$$\lim \int_a^x p(x)dx = \infty$$

and

$$\lim r(x) \frac{d}{dx} [r(x)p(x)]^{-1/2} = M,$$

the solutions of (1) *are oscillatory if $M > -2$ and nonoscillatory if $M < -2$.*

If y is a solution of equation (1), $z = r(x)y'$ is a solution of the differential equation

$$\left[\frac{1}{p(x)}z'\right]' + \frac{1}{r(x)}z = 0. \tag{1.7}$$

Thus the solutions of (1) and those of (1.7) are oscillatory or nonoscillatory together. Application to equation (1.7) of the procedure used on equation (1) in the proof of Theorem 1.6 establishes the stated result.

The examples which follow indicate the sensitivity of the results of this section.

EXAMPLE 1.2. For

$$y'' + a^2x^ny = 0, \tag{1.8}$$

we note that

$$h'(x) = -\frac{n}{2a}x^{-(n+2)/2}.$$

To study the equation we distinguish three cases.

Case 1: $n > -2$. Then $\lim h'(x) = 0$, so that the solutions of (1.8) are seen to be oscillatory by Theorem 1.5.

Case 2: $n < -2$. Then $\lim h'(x) = \infty$, and Theorem 1.5 can again be applied, showing the solutions of (1.8) to be nonoscillatory here.

Case 3: $n = -2$. Then $\lim h'(x) = 1/a$. The solutions are oscillatory if $a^2 > 1/4$ and nonoscillatory if $a^2 < 1/4$, by Theorem 1.5. Theorem 1.5 fails to give any information if $a^2 = 1/4$ ($\lim h'(x) = 2$). In this case, however, $H_1(x) \equiv 0$, and the solutions are nonoscillatory. The equation

$$y'' + 1/4x^{-2}y = 0$$

is thus in a sense a limiting equation.

EXAMPLE 1.3. For

$$y'' + (1/4x^{-2} + e^{-x})\,y = 0,$$

since $\lim h'(x) = 2$, Theorem 1.5 gives no information about the solutions. However, for large values of x, $H_1(x) < 0$ and the solutions are accordingly nonoscillatory.

EXAMPLE 1.4. Another equation for which $\lim h'(x) = 2$ is

$$y'' + 1/4\, x^{-2}\log^{-1}x\,(1 + \log x)y = 0.$$

The solutions of this equation are oscillatory by Theorem 1.1 since

$$\lim \int_a^x H_2(x)\,dx = \infty.$$

The limitations of the theory of this section are indicated by the fact that from the theorems which have been given here it is not possible to determine whether the solutions of the equation in Example 1.1 are oscillatory or not.

2. Counting the zeros of a solution. We consider first the differential equation (1.1), where $h(x) > 0$ and of class C' on the interval $0 < x < \infty$. Let $N(a, x)$ represent the number of zeros of a solution $y(x)$ of (1.1) on the interval[4] (a, x) where $a > 0$. This number differs by at most one for all solutions, and hence for the present purpose can be considered as depending only

[4]In designating intervals it will be convenient to use the following conventions: $[a, b]$ means the interval $a \leq x \leq b$, $(a, b]$ means the interval $a < x \leq b$, $[a, b)$ means the interval $a \leq x < b$, (a, b) means the interval $a < x < b$.

on the differential equation and not on a particular solution.

In the preceding section it was shown that the solutions of equation (1.1) are oscillatory whenever $\lim h'(x) < 2$, and are nonoscillatory whenever $\lim h'(x) > 2$. There are equations with oscillatory solutions and others with nonoscillatory solutions for which $\lim h'(x) = 2$. Wiman [8] has given an asymptotic formula for $N(a, x)$ when $\lim h'(x) = 0$:

$$N(a, x) \sim \frac{1}{\pi} \int_a^x \frac{dx}{h(x)}.$$

An asymptotic formula is readily found whenever $0 < \lim h'(x) < 2$, by considering the set of differential equations

$$y'' + (m^2 + 1/4)x^{-2}y = 0,$$

where m is any real number. For a particular value of m, $N(a, x) = \frac{m}{\pi} \log x + k$ (k is a constant), and $h'(x) = (m^2 + 1/4)^{-1/2}$.

THEOREM 2.1. *If, in equation* (1.1), $\lim h'(x) = m < 2$,

$$N(a, x) \sim \frac{1}{\pi}\left(\frac{1}{m^2} - \frac{1}{4}\right)^{1/2} \log x. \tag{2.1}$$

Any differential equation included in Theorem 2.1 is also included in the stronger Theorem 2.3 given below.

The Wiman formula can be extended to an equation of the form (1).

THEOREM 2.2. *If*

$$\lim \int_a^x \frac{dx}{r(x)} = \infty \ \textit{or} \ \lim \int_a^x p(x)dx = \infty,$$

then whenever

$$\lim r(x)\frac{d}{dx}[r(x)p(x)]^{-1/2} = 0$$

the relation

$$N(a, x) \sim \frac{1}{\pi} \int_a^x \sqrt{p(x)r^{-1}(x)}\,dx$$

holds.

If $\lim \int_a^x r^{-1}(x)dx = \infty$, we apply to (1) the transformation

$$t = \int_a^x r^{-1}(x)dx$$

and obtain

$$\frac{d^2y}{dt^2} + r(x)p(x)y = 0. \tag{2.2}$$

According to the Wiman theorem, the number of zeros $N(a_1, t)$ of a solution $y(t)$ of (2.2) is asymptotically equal to

$$\frac{1}{\pi}\int_{a_1}^{t} [r(x)p(x)]^{1/2}dt,$$

provided

$$\lim \frac{d}{dt}[r(x)p(x)]^{-1/2} = 0.$$

But this is equivalent, under the transformation, to the first half of the theorem.

If

$$\lim \int_a^x p(x)dx = \infty,$$

we apply the transformation

$$s = \int_a^x p(x)dx$$

to equation (1.7), noting that the zeros of a solution of (1) and those of a solution of (1.7) separate each other, and proceed as above.

An application of a variant of the foregoing method yields a generalization of the Wiman theorem for equation (1.1).

THEOREM 2.3. *If the function* $g(x) = [x^2h^{-2}(x) - 1/4]^{1/2}$ *is real and positive, and*

$$\lim x[g^{-1}(x)]' = 0, \tag{2.3}$$

then

$$N(a, x) \sim \frac{1}{\pi} \int_a^x \frac{g(x)}{x} dx.$$

To prove the theorem we transform (1.1) by the substitution $y = x^{1/2}z$ and obtain

$$[xz']' + g^2(x)x^{-1}z = 0. \tag{2.4}$$

The proof of the theorem may now be completed by applying Theorem 2.2 to equation (2.4).

Theorem 2.3 is more general than the Wiman theorem. Applying the law of the mean to $h^{-1}(x)$, we see that the Wiman condition, $\lim h'(x) = 0$, implies $\lim xh^{-1}(x) = \infty$, and it is readily verified that whenever the Wiman condition is satisfied equation (2.3) holds. On the other hand, Theorem 2.3 applies to differential equations for which the Wiman theorem is not available; e.g.,[5]

$$y'' + \frac{1 + \log x}{4x^2 \log x} y = 0. \tag{2.5}$$

It should be observed that Theorem 2.3 includes all equations covered by Theorem 2.1, whereas Theorem 2.1 is not applicable to equation (2.5) since $\lim h'(x) = 2$.

Still more refined results are obtainable if instead of using the transformation which led to equation (2.4), we use the substitution $y = q^{1/2}(x)z$, where $q(x)$ is so chosen that $\int_a^x q^{-1}(x)dx$ diverges more slowly than $\log x$. This suggests the use of the sequence

$$x \log x,\ x \log x \log_2 x,\ \cdots,\ x \log x \cdots \log_n x,\ \cdots$$

(cf. [6]).

To show that such a sequence can be used, the following theorem is included.

THEOREM 2.4. *In the differential equation*

$$[r_{n-1}(x)y']' + p(x)y = 0, \tag{2.6}$$

[5]See Example 1.4. This equation was shown to have oscillatory solutions.

where

$$r_0(x) = x, \quad r_n(x) = r_{n-1}(x)\log_n x,$$

if

$$\lim r_{n-1}(x) \frac{d}{dx}[r_{n-1}(x)p(x)]^{-1/2} = 0, \tag{2.7}$$

and

$$p^{-1}(x) = o[r_{n-1}(x)\log_n^2 x], \tag{2.8}$$

then

$$\lim r_n(x) \frac{d}{dx}[r_n(x)p_1(x)]^{-1/2} = 0, \tag{2.9}$$

where

$$p_1(x) = \log_n^{1/2}x\,\{[r_{n-1}(x)(\log_n^{1/2}x)']' + p(x)\log_n^{1/2}x\}.$$

Moreover, (2.9) *does not imply* (2.7).

The proof is clear once the limits in question are evaluated.

3. Boundedness of the solutions of a particular equation. In this section we study the question of boundedness near $x = +\infty$ of solutions of the self-adjoint differential equation

$$[r(x)y']' - p(x)y = 0. \tag{3.1}$$

We assume that $r(x)$ and $p(x)$ are positive continuous functions of x for x large, and that $r'(x)$ is continuous.

A CANONICAL FORM. It is useful to develop a canonical form for the solutions of (3.1). This form is suggested by the special case

$$r(x)p(x) = k^4.$$

In this instance the general solution of (3.1) may be written

$$c_1 e^{v(x)} + c_2 e^{-v(x)},$$

where $v(x) = k^{-2} \int_a^x p(x)dx$, and c_1 and c_2 are arbitrary constants.

Direct computation and an application of the fundamental existence theorem for systems of differential equations yield the following result.

THEOREM 3.1. *The general solution of* (3.1) *may be written*

$$c_1 u(x) e^{v(x)} + c_2 u(x) e^{-v(x)},$$

where $u(x)$ *and* $v(x)$ *are functions of class* C^2 *which satisfy the pair of equations*

$$ru^3[(ru')' - pu] = -1, \tag{3.2}$$

$$ru^2 v' = 1. \tag{3.3}$$

Since $u(x)$ is a function of class C^2 satisfying (3.2) and (3.3), $u(x)$ cannot vanish.

THEOREM 3.2. *The general solution of* (3.2) *is given by the relation*

$$u^2 = ay_1^2 + by_2^2 + 2cy_1 y_2 \tag{3.4}$$

where y_1 *and* y_2 *are linearly independent solutions of* (3.1) *and* a, b, *and* c *are any constants satisfying the relation*

$$ab = -k^{-2} + c^2,$$

if k *is the constant*

$$r(x)[y_1(x)y_2'(x) - y_1'(x)y_2(x)].$$

To prove the theorem, the solution given by (3.4) can be substituted directly in (3.2).

BOUNDEDNESS OF SOLUTIONS OF (3.2). We first prove a lemma.

LEMMA 3.3. *Let* $r(x)$, $r'(x)$, *and* $p(x)$ *be continuous and* $r(x)p(x)$ *be positive and monotone for large values of* x. *If* $u(x)$ *is a positive solution of equation* (3.2), *the relations* $\overline{\lim}\, u(x) = \infty$ *and* $\underline{\lim}\, u(x) < \infty$ *cannot hold simultaneously.*

Suppose that the hypotheses of the theorem are satisfied when $x > a$, and that $\overline{\lim}\, u(x) = \infty$ and $\underline{\lim}\, u(x) < \infty$. Since $u(x)$ is of class C^2, there are an infinite number of relative maximum points and of relative minimum points of $u(x)$ on (a, ∞). We rewrite the equation (3.2) in the form

$$(3.2)' \qquad (ru')' = r^{-1}u^{-3}(rpu^4 - 1).$$

From the hypotheses of the lemma, $\lim\,[r(x)p(x)]$ exists and is nonnegative. If $\lim\,[r(x)p(x)] > 0$, there exists a relative maximum point x_M of $u(x)$ for which

$$[r(x_M)u'(x_M)]' > 0 \text{ and } r(x_M)u'(x_M) = 0.$$

This implies that there is a positive number ϵ such that $u'(x) > 0$ when $x_M < x < x_M + \epsilon$, which is impossible. If $\lim\,[r(x)p(x)] = 0$, from equation (3.2)$'$ we see there is a relative minimum point x_m of $u(x)$ for which

$$[r(x_m)u'(x_m)]' < 0 \text{ and } r(x_m)u'(x_m) = 0.$$

This implies that there is a positive number ϵ' such that $u'(x) < 0$ when $x_m < x < x_m + \epsilon'$, which is impossible.

Thus, in any case, the assumption $\overline{\lim}\, u(x) = \infty$ and $\underline{\lim}\, u(x) < \infty$ leads to a contradiction. The truth of the lemma follows.

THEOREM 3.3. *Let $r(x)$, $r'(x)$, and $p(x)$ be continuous, and $r(x)p(x)$ be positive and monotone increasing for large values of x. Then every solution $u(x)$ of equation (3.2) is bounded near $x = \infty$.*

We recall that a solution $u(x)$ of equation (3.2) cannot vanish, and note that if $u(x)$ is a solution, so is $-u(x)$. Let a be a positive real number such that the hypotheses of the theorem are satisfied when $x > a$. Suppose then that $u(x) > 0$ and let $m(x)$ and $M(x)$ be respectively the minimum value and the maximum value of $u(x)$ on $[a, x]$. Let x_m and x_M be such that

$$u(x_m) = m(x), \quad u(x_M) = M(x).$$

Since

$$[r(x)u'(x)]^2 - r(x)p(x)u^2(x) + r(a)p(a)u^2(a)$$

$$+ \int_a^x u^2(x)[r(x)p(x)]'\,dx + u^{-2}(a) = u^{-2}(x) + [r(a)u'(a)]^2,$$

it follows that

$$[r(x_m)u'(x_m)]^2 + c_1^2 \leq m^{-2}(x) + r(a)p(a)m^2(x) + c_2^2, \tag{3.5}$$

$$[r(x_M)u'(x_M)]^2 + c_1'^2 \geq M^{-2}(x) + r(a)p(a)M^2(x) + c_2'^2, \tag{3.6}$$

where c_1, c_2, c_1', c_2' are real constants. We identify two cases according as $[r(x)u'(x)]^2$ is bounded or not.

Case 1. If $[r(x)u'(x)]^2$ is bounded, it follows from inequality (3.6) that $M(x)$ is bounded, and hence that $u(x)$ is bounded.

Case 2. If $\overline{\lim}\,[r(x)u'(x)]^2 = \infty$, we assume $u(x)$ is not bounded. Then

$$\lim M(x) = \infty \quad \text{and} \quad \lim u(x) = \infty$$

by Lemma 3.3. But equation (3.2)′ then implies that $(ru')'$ is eventually positive or $\lim\,[r(x)u'(x)]^2 = \infty$. It follows from inequality (3.5) that $\lim m^{-1}(x) = \infty$ and hence that $\lim m(x) = 0$. Then by Lemma 3.3, $u(x)$ is bounded.

We give a companion result when $r(x)p(x)$ is monotone decreasing.

THEOREM 3.4. *If $r(x)$, $r'(x)$, and $p(x)$ are continuous, and $r(x)p(x)$ is positive and monotone decreasing for large values of x, every solution $u(x)$ of* (3.2) *is bounded away from zero.*

The proof of this theorem is similar to that of Theorem 3.3.

The following theorem specializes Theorem 3.3 to the "normal form" of (1).

THEOREM 3.5. *If $p(x) > 0$ and monotone increasing for large values of x, then the differential equation*

$$y'' - p(x)y = 0 \tag{3.7}$$

is such that all solutions are monotone, and there is one solution $y_1(x)$ which approaches zero as $x \longrightarrow \infty$. Every solution of the differential equation (3.7) *which is linearly independent of $y_1(x)$ is unbounded on* $(0, \infty)$.

The monotone character of the solutions is apparent from the fact that if $y(x)$ is a solution of (3.7), $y''(x)$ is eventually of one sign and hence so is $y'(x)$. The general solution can be written

$$y(x) = c_1 u(x) e^{v(x)} + c_2 u(x) e^{-v(x)}, \tag{3.8}$$

where $u(x)$ is positive, and $u(x)$ and $v(x)$ are functions of class C^2 satisfying (3.2) and (3.3). By Theorem 3.3, $u(x)$ is bounded, and hence $[v'(x)]^{-1}$ is bounded. Thus $v'(x)$ is positive and bounded away from zero. This implies

$$\overline{\lim}\, u(x) e^{-v(x)} = \lim u(x) e^{-v(x)} = 0.$$

We set

$$y_1(x) = u(x) e^{-v(x)}$$

and let $y_2(x)$ be a positive solution linearly independent of $y_1(x)$. Since

$$\lim y_1'(x) = 0$$

and

$$y_1(x) y_2'(x) - y_1'(x) y_2(x) = c,$$

where c is a nonzero constant,

$$\lim y_2(x) = \infty,$$

and the theorem follows.

COROLLARY 3.5. *If* $\lim \int_a^x r^{-1}(x)dx = \infty$, *and* $r(x)p(x)$ *is a positive monotone increasing function of* x *for* x *large, there is one solution* $y_1(x)$ *of equation* (3.1) *which approaches zero as* $x \longrightarrow \infty$. *All solutions linearly independent of* $y_1(x)$ *approach* $+\infty$ *or* $-\infty$ *as* $x \longrightarrow \infty$. *The solutions are all monotone.*

We let $t = \int_a^x r^{-1}(x)dx$. Equation (3.1) becomes

$$\frac{d^2y}{dt^2} - r(x)p(x)y = 0, \tag{3.9}$$

where r and p are to be considered functions of t. The theorem follows from Theorem 3.5 applied to equation (3.9).

THEOREM 3.6. *If* $r(x)$, $r'(x)$, *and*

$$P(x) = \frac{p(x)}{r(x)} + \frac{1}{2}\frac{r''(x)}{r(x)} - \frac{1}{4}\left[\frac{r'(x)}{r(x)}\right]^2$$

are continuous, $r(x)$ and $P(x)$ are positive, and $P(x)$ is monotone increasing for x large, there is one solution $y_1(x)$ of equation (3.1) such that

$$\lim r^{1/2}(x)y_1(x) = 0.$$

For every solution $y(x)$ which is linearly independent of $y_1(x)$, $r^{1/2}(x)y(x)$ is not bounded.

To prove this theorem we transform equation (3.1) by means of the substitution $z = r^{1/2}(x)y$. The resulting differential equation is

$$z'' - P(x)z = 0.$$

Application of Theorem 3.5 to this equation yields the theorem.

EXAMPLE 3.1. For

$$(x^2y')' - x^2y = 0,$$

all solutions of the corresponding equation

$$x^2u^3[(x^2u')' - x^2u] = -1$$

are bounded since $[r(x)p(x)]' > 0$ (Theorem 3.3). The general solution of the given equation is

$$\frac{1}{x}(c_1e^x + c_2e^{-x}).$$

EXAMPLE 3.2. For the equation

$$\left(\frac{1}{x^2}y'\right)' - \frac{x^2-2}{x^4}y = 0,$$

$[r(x)p(x)]' < 0$. By Theorem 3.4, therefore, all solutions of

$$\frac{1}{x^2}u^3\left[\left(\frac{1}{x^2}u'\right)' - \frac{1}{x^4}(x^2 - 2)u\right] = -1$$

are bounded away from zero. Moreover, since

$$\frac{p(x)}{r(x)} + \frac{1}{2}\frac{r''(x)}{r(x)} - \frac{r'^2(x)}{4r^2(x)} = 1,$$

by Theorem 3.6 there is a solution $y_1(x)$ of the given equation such that

$$\lim x^{-1} y_1(x) = 0,$$

and for every linearly independent solution $y(x)$,

$$\lim x^{-1} y(x) = \infty .$$

The general solution of this equation is

$$c_1 x e^{x} + c_2 x e^{-x}.$$

EXAMPLE 3.3. Consider the differential equation

$$y'' - \left(x + \frac{3}{4x^2}\right) y = 0.$$

By Theorem 3.5, one solution of this equation approaches zero, and all solutions which are linearly independent of this solution become infinite. The general solution of the equation is

$$c_1 x^{-1/2} e^{x^2/2} + c_2 x^{-1/2} e^{-x^2/2}.$$

THE RICCATI EQUATION ASSOCIATED WITH EQUATION (3.1). Since the solutions of equation (3.1) are nonoscillatory, the transformation $w = r(x) y'/y$ applied to this equation leads to the relationship

$$w' = p(x) - r^{-1}(x) w^2, \tag{3.10}$$

which is valid for each solution y of (3.1) when x is sufficiently large. The differential equation (3.10) can be used to obtain additional information on the question of boundedness of solutions of (3.1).

THEOREM 3.7. *If* $\lim \int_a^x r^{-1}(x)dx < \infty$ *and* $\lim \int_a^x p(x)dx < \infty$, *all solutions of* (3.1) *are bounded, and there is a positive constant* M *such that*

$$|y'(x)| < M r^{-1}(x).$$

It is sufficient to consider only positive solutions of equation (3.1). Accordingly, we suppose $y(x)$ is any solution of (3.1) which is positive for x large, and let b be a positive number such that both $y(x)$ and $y'(x)$ are of one sign when $x \geq b$. Equation (3.10) is then valid for such values of x, and $w(x)$ is of one sign. If $w(x) < 0$ when $x \geq b$, $y(x)$ is bounded. If $w(x) > 0$, when $x \geq b$ it follows from equation (3.10) that

$$w'(x) < p(x)$$

and

$$w(x) < w(b) + \int_b^x p(x)dx < K,$$

where K is a constant. Hence

$$\frac{y'(x)}{y(x)} < Kr^{-1}(x)$$

and

$$\log\left|\frac{y(x)}{y(b)}\right| < K\int_b^x \frac{dx}{r(x)} < \infty.$$

Thus, $y(x)$ is bounded.

To prove the last statement of the theorem we apply the first part to the equation

$$[p^{-1}(x)z']' - r^{-1}(x)z = 0,$$

for which $z = r(x)y'$ is a solution if y is a solution of (3.1).

Examples 3.1 and 3.2 show that the hypotheses of Theorem 3.7 cannot be weakened to the convergence of only one of the integrals

$$\int_a^\infty \frac{dx}{r(x)}, \qquad \int_a^\infty p(x)dx.$$

4. Boundedness of nonoscillatory solutions of an equation of the form (1). In this section we study the boundedness of solutions of an equation of the form (1) when its solutions are nonoscillatory and both $r(x)$ and $p(x)$ are positive and continuous functions of x for large values of x. [6] It is known that a necessary condition for the solutions of (1) to be nonoscillatory is that not both

[6] Sections 3 and 4 together discuss boundedness of nonoscillatory solutions whenever $p(x)$ is eventually one sign, for $p(x)$ negative and positive respectively. The case where $p(x)$ is not of one sign is not studied in this paper.

$$\lim \int_a^x \frac{dx}{r(x)} = \infty$$

and

$$\lim \int_a^x p(x)dx = \infty.$$

If $r(x)p(x)$ is a monotone function, the convergence of both of the aforementioned integrals is a sufficient condition that (1) have nonoscillatory solutions [5].

We state the principal theorem:

THEOREM 4.1. *A necessary and sufficient condition that an equation of the form* (1) *with nonoscillatory solutions have all solutions bounded near* $x = \infty$ *is that*

$$\lim \int_a^x \frac{dx}{r(x)} < \infty.$$

Whenever the solutions of equation (1) are nonoscillatory, the transformation $w(x) = r(x)y'/y$ leads to the Riccati equation

$$w' = -p(x) - \frac{w^2}{r(x)}, \tag{4.1}$$

which is valid for each solution $y(x)$ of (1) when x is sufficiently large.

Let $y(x)$ be a nonoscillatory solution of (1) such that $y(x) > 0$ and $y'(x) \neq 0$ whenever $x \geq a > 0$, where a has been chosen sufficiently large that $p(x) > 0$ when $x \geq a$. It is sufficient to consider only solutions which are eventually positive since the negative of a solution of (1) is also a solution. Then if $x \geq a$, equation (4.1) is valid as noted above, and

$$\frac{w'(x)}{w^2(x)} < -\frac{1}{r(x)}.$$

Hence,

$$\frac{1}{w(x)} > \int_a^x \frac{dx}{r(x)} + \frac{1}{w(a)}. \tag{4.2}$$

We assume that

$$\lim \int_a^x \frac{dx}{r(x)} < \infty.$$

If $w(x) > 0$,

$$\frac{y'(x)}{y(x)} < \frac{1}{r(x)} \left[\int_a^x \frac{dx}{r(x)} + \frac{1}{w(a)} \right]^{-1},$$

by equation (4.2), so that

$$\log y(x) < \log \left| \int_a^x \frac{dx}{r(x)} + \frac{1}{w(a)} \right| + c,$$

where c is a constant. Accordingly, $y(x)$ is bounded. If $w(x) < 0$, then $y'(x) < 0$, and $y(x)$ is bounded. Thus, whenever

$$\lim \int_a^x \frac{dx}{r(x)} < \infty,$$

all solutions of (1) are bounded.

If $\lim \int_a^x r^{-1}(x)dx = \infty$, it follows from equations (4.1) and (4.2) that $w(x)$ is a positive, monotone decreasing function with $\lim w(x) = 0$. Therefore,

$$\lim \frac{r(x)y'(x)}{y(x)} = 0 \tag{4.3}$$

for all solutions $y(x)$ of (1). Let $y_1(x)$ and $y_2(x)$ be any two linearly independent solutions of (1) which are positive for x large. From equation (4.3),

$$\lim \frac{r(x)y_1'(x)}{y_1(x)} = 0, \qquad \lim \frac{r(x)y_2'(x)}{y_2(x)} = 0.$$

If c is the nonzero constant such that

$$r(x)\,[y_1'(x)y_2(x) - y_1(x)y_2'(x)] = c,$$

then

$$\frac{r(x)y_1'(x)}{y_1(x)} - \frac{r(x)y_2'(x)}{y_2(x)} = \frac{c}{y_1(x)y_2(x)}. \tag{4.4}$$

The limit of the left side of the equation (4.4) is zero. Therefore, since all positive solutions of (1) are monotone increasing, at least one of $y_1(x)$ and $y_2(x)$ becomes infinite, and if

$$\lim \int_a^x \frac{dx}{r(x)} = \infty,$$

not all of the solutions of (1) are bounded.

THEOREM 4.2. *Solutions of* (1) *are nonoscillatory and bounded near* $x = \infty$ *if* $r(x)p(x)$ *is monotone decreasing and* $\lim \int_a^x r^{-1}(x)dx < \infty$, *or if* $r(x)p(x)$ *is monotone increasing and* $\lim \int_a^x p(x)dx < \infty$.

A solution of (1) can be written in the canonical form $u(x)\sin v(x)$, where $u(x)$ and $v(x)$ are functions of class C^2 satisfying [5], the pair of equations

$$ru^3[(ru')' + pu] = 1, \quad ru^2v' = 1.$$

If $r(x)p(x)$ is monotone decreasing, there exists [5] a positive number m such that $u(x) > m$. Since $\lim \int_a^x r^{-1}(x)dx < \infty$,

$$\lim v(x) = \lim \left[v(a) + \int_a^x \frac{dx}{r(x)u^2(x)}\right] < v(a) + m^{-2}\lim \int_a^x \frac{dx}{r(x)} < \infty.$$

Thus, solutions of (1) are nonoscillatory. An application of Theorem 4.1 now yields the first part of the theorem.

The proof of the second half of the theorem is obtained by considering equation (1.7). If $y(x)$ is a solution of (1), $r(x)y'(x)$ is a solution of (1.7). But by the preceding paragraph, $r(x)y'(x)$ is nonoscillatory and bounded when $r^{-1}(x)p^{-1}(x)$ is monotone decreasing and $\lim \int_a^x p(x)dx < \infty$. Therefore, $y(x)$ is nonoscillatory. That $y(x)$ is bounded follows from a theorem of Leighton [3].

From equation (4.1) it is evident that whenever $\lim \int_a^x p(x)dx = \infty$, $w(x)$ is negative for every solution $y(x)$ of (1). This remark, together with the fact noted in the proof of Theorem 4.1 that, when $\lim \int_a^x r^{-1}(x)dx = \infty$, $w(x)$ is positive for every solution $y(x)$ of (1), proves the following theorem:

THEOREM 4.3. *If* $\lim \int_a^x p(x)dx = \infty$, *all nonoscillatory solutions* $y(x)$ *of* (1) *have the property that* $y^2(x)$ *is monotone decreasing. If* $\lim \int_a^x r^{-1}(x)dx = \infty$, *all nonoscillatory solutions have the property that* $y^2(x)$ *is monotone increasing.*

It should be observed that the restriction to nonoscillatory solutions in

Theorem 4.1 is not superfluous. This is illustrated by the following example.

EXAMPLE 4.1. The differential equation

$$(xy')' + \frac{1}{x}y = 0$$

has the general solution

$$y(x) = c_1 \sin \log |x| + c_2 \cos \log |x|.$$

All solutions of the equation are bounded near $x = \infty$, whereas

$$\lim \int_a^x \frac{dx}{r(x)} = \lim \log \left|\frac{x}{a}\right| = \infty.$$

From the theorems of this section it is evident that whenever the so-called "normal form" of equation (1) with $r(x) = 1$ has nonoscillatory solutions, these solutions cannot all be bounded.

5. Remarks on a theorem of Leighton. We recall that in Theorem L_1 Leighton gives, as a sufficient condition for solutions of (1) to be oscillatory, that

$$\lim \int_a^x \frac{dx}{r(x)} = \infty, \quad \lim \int_a^x p(x)dx = \infty.$$

In the paper [6] containing this theorem there was established the existence of a sequence of tests for oscillation, each more sensitive than the preceding. This sequence was obtained by successively transforming an equation of the form (1) into an equation

$$[r_n(x)y']' + p_n(x)y = 0,$$

where

$$r_0(x) = x, \quad r_n(x) = r_{n-1}(x)\log_n x.$$

It might be asked whether there is some positive function $R(x)$ with the property that whenever (1) is transformed into an equation

$$[R(x)y']' + P(x)y = 0,$$

the relations

$$\lim \int_a^x \frac{dx}{R(x)} = \infty, \quad \lim \int_a^x P(x)dx = \infty$$

would give a necessary as well as a sufficient condition for oscillation. That there is no such function is shown by the following theorem.

THEOREM 5.1. *If $r(x)$ is a positive continuous function such that*

$$\lim \int_a^x r^{-1}(x)dx = \infty,$$

there exists a positive continuous function $p(x)$ such that $\lim \int_a^x p(x)dx < \infty$, *and solutions of the differential equation*

$$[r(x)y']' + p(x)y = 0$$

are oscillatory.

We set

$$p(x) = \frac{1}{r(x)} \left[1 + \int_a^x \frac{dx}{r(x)}\right]^{-2}.$$

The truth of the theorem then follows from Theorem 1.6.

REFERENCES

1. P. Hartman, and A. Wintner, *On non-conservative linear oscillations of low frequency*, Amer. J. Math. 70 (1948), 529-539.

2. E. L. Ince, *Ordinary differential equations*, Longmans, London, 1927.

3. W. Leighton, *Bounds for the solutions of a second-order linear differential equation*, Proc. Nat. Acad. Sci., U.S.A. 35 (1949), 190-191.

4. ———, *Principal quadratic functionals and self-adjoint second-order differential equations*, Proc. Nat. Acad. Sci., U.S.A. 35 (1949), 192-193.

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

6. ———, *The detection of the oscillation of solutions of a second-order linear differential equation*, Duke Math. J. 17 (1950), 57-61.

7. C. Sturm, *Sur les équations différentielles linéaires du second ordre*, J. Math. Pures Appl. 1 (1836), 106-186.

8. A. Wiman, *Über die reellen Lösungen der linearen Differentialgleichungen zweiter Ordnung*, Ark. Mat. Astr. Fys. 12, no. 14 (1917).

WASHINGTON UNIVERSITY

A NOTE ON THE BANACH SPACES OF CALKIN AND MORREY

E. H. ROTHE

1. Introduction. Let G be a bounded domain in an n-dimensional real Euclidean space, and for $\alpha > 1$, let L_α be the space of real-valued functions f such that f^α is summable over G. The class $\mathfrak{P}_\alpha$ as defined by Calkin and Morrey [2; 6; 7] is then the class of functions which together with their first "generalized derivatives" [2, Def. 3.4; 7, p. 4] are in L_α. With a suitable norm, $\mathfrak{P}_\alpha$ becomes a Banach space [2, p. 185]. Morrey proved [7, p. 8] that in this Banach space the solid sphere V of radius K and with the origin as center is "weakly compact"[1]). Using this fact together with lower semicontinuity theorems, he obtained very general existence theorems for minima of multiple integrals[2]).

The object of the present note is to point out that some of the results in this direction may be obtained by the use of general Banach space theory: the starting point is the simple remark that the Banach space $\mathfrak{P}_\alpha$ is reflexive (§ 2). The weak compactness of the solid sphere V is, by Alaoglu's theorem [1], a corollary to this remark [§ 3]. It now follows almost immediately that a real-valued function $I(x)$, which is "weakly" lower semicontinuous, takes a minimum in V (Theorem 3.1). In § 4 some sufficient conditions for weak lower semicontinuity are given. Finally, as an example of the applicability of these considerations to calculus of variation problems, a theorem on the existence of minima of multiple integrals is given which is related to, but not identical with, the results of Morrey referred to at the end of the previous paragraph (§ 5).

2. The uniform convexity and reflexivity of the space $\mathfrak{P}_\alpha$. Let t denote the point with coordinates $t_1, t_2, \cdots, t_n$ of the domain G of § 1. Let $f(t) = f^{(0)}(t)$ be an element of $\mathfrak{P}_\alpha$, and $f^{(i)}(t)$ $(i = 1, 2, \cdots, n)$ its first generalized derivative with respect to t_i. Let $\|f\|$ be defined by the equation

$$\|f\|^\alpha = \int_G \sum_{i=0}^{n} |f^{(i)}(t)|^\alpha dt. \tag{2.1}$$

[1] See [7] for Morrey's definition of weak compactness. The weak topology used in the present paper is defined in § 3.

[2] See [7, Chap. III], where also the relation to the results of Tonelli is discussed.

Received April 28, 1952.

Pacific J. Math. 3 (1953), 493-499

We then have:

LEMMA 2.1. *Let $\alpha > 1$. With the norm defined by (2.1) the space of classes of functions of $\mathfrak{P}_\alpha$ equivalent under this norm is a Banach space*[3].

From now on $\mathfrak{P}_\alpha$ will always denote the Banach space of Lemma 2.1, and it will always be supposed that $\alpha > 1$.

THEOREM 2.1. *The space $\mathfrak{P}_\alpha$ is uniformly convex*[4].

Proof. Let L_α be the Banach space of classes of equivalent functions f which are defined in G and for which

$$\left\{ \int_G |f|^\alpha dt \right\}^{1/\alpha} \tag{2.2}$$

exists. L_α is uniformly convex [3, p.403, Corollary]. Since a finite "uniformly convex" direct product of uniformly convex spaces is uniformly convex [3, p. 397-398] it follows that the direct product of L_α taken $(n+1)$ times by itself, that is, the space $\tilde{\mathfrak{P}}_\alpha$ of $(n+1)$-tuples $f_0, f_1, \cdots, f_n$ $(f_\nu \in L_\alpha;\ \nu = 0, 1, \cdots n)$ with the norm

$$\left\{ \int_G \sum_{\nu=0}^{n} |f_\nu|^\alpha dt \right\}^{1/\alpha}$$

is likewise uniformly convex. This proves the theorem since $\mathfrak{P}_\alpha$ is obviously a linear subspace of $\tilde{\mathfrak{P}}_\alpha$.

Since a uniformly convex space is reflexive [5; 8], we have the following corollary to Theorem 2.1.

COROLLARY. *For $\alpha > 1$, $\mathfrak{P}_\alpha$ is reflexive.*

3. **The compactness of the sphere V.** We recall first a few well-known definitions and facts. Let E be an arbitrary Banach space in the strong topology, that is in the topology induced by the norm of the space. Let K be a positive number, and V be the solid sphere $||x|| \leqq K$ of E. By V_K we denote then the topological space whose elements are those of V and whose topology is induced by the following neighborhood definition: A neighborhood of the point x_0 of V_K is determined by a positive number ϵ and a finite number of linear continuous functionals $l_1(x), \cdots, l_n(x)$, and consists of all points x of V_K for which

[3] See [2, p. 185]. The definition of the norm given by Calkin is slightly different from the one used in the present paper. However, the proof of Lemma 2.1 is essentially unaltered.

[4] For the definition of the term "uniformly convex" see [3].

$$|l_i(x) - l_i(x_0)| < \epsilon \qquad (i = 1, 2, \cdots, n).$$

If E is the conjugate space of another Banach space F, $E = F^*$, we denote by V_K^* the topological space whose elements are again those of V, but whose topology is induced by the following neighborhood definition: A neighborhood of a point x_0 of V_K^* is determined by a positive number ϵ and a finite number of elements $f_1, \cdots, f_n$ of F, and consists of all points x of V_K for which

$$|x(f_i) - x_0(f_i)| < \epsilon \qquad (i = 1, 2, \cdots, n).$$

A well-known theorem of Alaoglu [1, Theorem 1.3] states that V_K^* is compact. Since for a reflexive space we have $V_K = V_K^*$, we obtain:

LEMMA 3.1. *If E is reflexive then V_K is compact.*

Since a strongly closed convex subset of V is also closed in the weak topology (that is, in the topology of V_K) we have as a consequence of Lemma 3.1 the following:

LEMMA 3.2. *Let C be a convex subset of V which is closed in the strong topology, and C_K the same set in the topology of V_K. Then C_K is compact.*

An easy consequence of Lemma 3.2 is:

LEMMA 3.3. *Let C and C_K have the same meaning as in Lemma 3.2, and let $I(x)$ be a real-valued function which is lower semicontinuous in C_K. Then $I(x)$ reaches a minimum in some point of C.* [5]

The preceding lemmas, together with the corollary to Theorem 2.1, now yield the main result of the present section:

THEOREM 3.1. *Let C be a bounded closed convex subset of $\mathfrak{P}_\alpha$. Let the norms of the elements of C be bounded by the positive constant K. Let V and V_K have the same meaning as in the first paragraph of this section, with E replaced by $\mathfrak{P}_\alpha$, and denote the set C in the topology of V_K by C_K. Then C_K is compact, and a real-valued function $I(x)$, which is lower semicontinuous in C_K, reaches a minimum in C.*

4. Sufficient conditions for lower semicontinuity. We prove now:

THEOREM 4.1. *Let C and C_K have the same meaning as in Theorem 3.1,*

[5] For a proof that Lemma 3.2 implies Lemma 3.3 see [9, p. 423-424].

and let $I(x)$ be a real-valued function defined on C. Then the following condition is sufficient for the lower semicontinuity of $I(x)$ on C_K (and therefore, by Theorem 3.1, for the existence of a minimum of $I(x)$ on C): to each $x_0 \in C$ there exists a bounded linear functional $l(x)$ such that

$$I(x) - I(x_0) \geqq l(x - x_0) \tag{4.1}$$

for all $x \in C$.

Proof. By definition of the lower semicontinuity we have to prove: to any given $\epsilon > 0$ there exists a neighborhood $N(x_0)$ of x_0 in V_K such that

$$I(x) - I(x_0) \geqq -\epsilon \tag{4.2}$$

for all x in the intersection, $N(x_0) \cap C$. But by (4.1) the inequality (4.2) will certainly be satisfied if we choose

$$N(x_0) = \{x \mid |l(x) - l(x_0)| < \epsilon,\ x \in V_K\}.$$

THEOREM 4.2. *With the same notations as in Theorem 4.1 let $I(x)$ have first and second order Fréchet differentials $D(x, h)$ and $D^2(x, h, k)$ at every point x of C. Moreover, let*

$$D^2(x, h, h) \geqq 0 \quad \textit{for} \quad x \in C. \tag{4.3}$$

Then $I(x)$ is lower semicontinuous in C_K.

Proof. From the Taylor expansion [4, Theorem 5],

$$I(x_0 + h) - I(x_0) = D(x_0, h) + \frac{1}{2} \int_0^1 D^2(x_0 + th, h, h)\, dt,$$

together with (4.3), we obtain

$$I(x_0 + h) - I(x_0) \geqq D(x_0, h).$$

This inequality shows that the assumptions of Theorem 4.1 are satisfied with

$$l(x - x_0) = D(x_0, x - x_0).$$

5. **An application to a multiple integral variational problem.** Let G be the domain of § 1 with points $t = (t_1, \cdots, t_n)$. For each $\mu = 1, \cdots, m$ let $z_\mu(t) \in \mathfrak{P}_\alpha$ and let Π_α be the space of classes of equivalent m-tuples $z = (z_1(t), \cdots, z_m(t))$ with the norm

$$\|z\| = \left[\int_G \sum_{\mu=1}^{m} \left\{|z|^\alpha + \sum_{\nu=1}^{n} \left|\frac{\partial z_\mu}{\partial t_\nu}\right|^\alpha\right\} dt\right]^{1/\alpha}. \tag{5.1}$$

LEMMA 5.1. *Theorems* 3.1 *and* 4.1 *still hold if* $\mathfrak{P}_\alpha$ *is replaced by* Π_α.

This lemma is obvious from the proofs of the theorems in question.

THEOREM 5.1. *Let*

$$f(t, z, p) = f(t_1, \cdots, t_n, z_1, \cdots, z_m, p_{11}, \cdots, p_{mn})$$

be a real-valued function of the indicated variables with the following properties:

(1) *f is defined for $t = (t_1, \cdots, t_n) \in G$ and for all values of the real variables z_μ, $p_{\mu\nu}$ ($\mu = 1, \cdots, m$; $\nu = 1, \cdots, n$), and for the same domain of the varibles $\partial f/\partial t_\nu$, $\partial f/\partial z_\mu$, and $\partial f/\partial p_{\mu\nu}$ are supposed to exist*;

(2) *if $z_\mu(t) \in \mathfrak{P}_\alpha$ then the functions of t obtained by replacing z by $z_\mu(t)$ and $p_{\mu\nu}$ by $\partial z_\mu/\partial t_\nu$ in f, $\partial f/\partial z_\mu$, and $\partial f/\partial p_{\mu\nu}$ are in L_β, where β is defined by*

$$1/\beta + 1/\alpha = 1;$$

(3) $e(t, z, z^0, p, p^0) \geqq 0$,

where by definition

$$e(t, z, z^0, p, p^0) = f(t, z, p) - f(t, z^0, p^0)$$
$$- \sum_{\mu=1}^{m} \left\{f_{z_\mu}(t, z^0, p^0)(z_\mu - z_\mu^0) + \sum_{\nu=1}^{n} f_{p_{\mu\nu}}(t, z^0, p^0)(p_{\mu\nu} - p_{\mu\nu}^0)\right\}.$$

Under these assumptions, if

$$I(z) = \int_G f\left[t_1, \cdots, t_n, z_1(t), \cdots, z_m(t), \frac{\partial z_1}{\partial t_1}, \cdots, \frac{\partial z_m}{\partial t_n}\right] dt,$$

then there exists a

$$z^{(1)} = z^{(1)}(t) = [z_1^{(1)}(t), \cdots, z_m^{(1)}(t)]$$

in the sphere

$$\|z\| \leqq K \tag{5.2}$$

such that

$$I(z) \geqq I(z^{(1)})$$

for all z in the sphere (5.2).

Proof. By Lemma 5.1 and Theorem 3.1, it will be sufficient to prove that $I(z)$ is lower semicontinuous at each point z^0 of the sphere (5.2). To such z^0 we define the linear functional $l(\zeta)$ of

$$\zeta = [\zeta_1(t), \cdots, \zeta_m(t)]$$

by setting

$$l(\zeta) = \int_G \sum_{\mu=1}^{m} \left\{ (f_{z_\mu})_0 \, \zeta_\mu + \sum_{\nu=1}^{n} (f_{p_{\mu\nu}})_0 \frac{\partial \zeta_\mu}{\partial t_\nu} \right\} dt,$$

where the symbol $(\;)_0$ indicates that the arguments are

$$t_1, \cdots, t_n, z_1^0(t), \cdots, z_m^0(t), \partial z_1^0/\partial t_1, \cdots, \partial z_m^0/\partial t_n,$$

and where

$$\zeta = [\zeta_1(t), \cdots, \zeta_m(t)] \in \Pi_\alpha.$$

The assumption (2) assures us that the linear functional $l(\zeta)$ is bounded. From the definition of $l(\zeta)$ and the assumption (3) we obtain

$$I(z) - I(z^0) = l(z - z^0) + e \geqq l(z - z^0).$$

Thus the assumption (4.1) of Theorem 4.1 is satisfied, and the theorem to be proved follows from Theorem 4.1 in conjunction with Lemma 5.1.

REFERENCES

1. L. Alaoglu, *Weak topologies of normed linear spaces*, Ann. of Math. 41 (1940), 252-267.

2. J. W. Calkin, *Functions of several variables and absolute continuity I*, Duke Math. J. 6 (1940), 170-186.

3. J. A. Clarkson, *Uniformly convex spaces*, Trans. Amer. Math. Soc. 40 (1936), 396-414.

4. L. M. Graves, *Riemann integration and Taylor's theorem in general analysis*, Trans. Amer. Math. Soc. 29 (1927), 163-177.

5. Sh. Kakutani, *Weak topology and regularity of Banach spaces*, Proc. Imp. Acad. Tokyo, 15 (1939), 169-173.

6. C. B. Morrey, Jr., *Functions of several variables and absolute continuity*, Duke Math. J. 6 (1940), 187-215.

7. ———, *Multiple integral problems in the calculus of variations and related topics*, University of California Publications in Mathematics, New Series, Vol. 1 No. 1 (1943), 1-130.

8. B. J. Pettis, *A proof that every uniformly convex space is reflexive*, Duke Math. J. 5 (1939), 249-253.

9. E. H. Rothe, *Gradient mappings and extrema in Banach spaces*, Duke Math. J. 15 (1948), 421-431.

UNIVERSITY OF MICHIGAN

A GENERALIZATION OF THE CENTRAL ELEMENTS OF A GROUP

EUGENE SCHENKMAN

1. Introduction. If a and g are elements of a group G, we shall denote by $a^{(1)}(g)$ or $a(g)$ the element $g^{-1}ag$, and then for $n = 2, 3, 4, \cdots$ define $a^{(n)}(g) = a(a^{(n-1)}(g))$.

If for some n and all $g \in G$, $a^{(n)}(g) = a$ then a will be called *weakly central of order n* or simply *weakly central.* Thus the center elements of G are weakly central of order 1.

As usual, let

$$[g, a] = a^{-1} g^{-1} ag = a^{-1} \cdot a(g);$$

then it can readily be verified by induction on n that

$$a^{-1} \cdot a^{(n)}(g) = a^{-1} \cdot [\overbrace{a \cdots [a}^{n-1 \text{ times}}, g] \cdots]^{-1} \cdot a \cdot [\overbrace{a \cdots [a}^{n-1 \text{ times}}, g] \cdots]$$

$$= [\overbrace{a \cdots [a}^{n \text{ times}}, g] \cdots].$$

Thus $a^{(n)}(g) = a$ is equivalent to

$$[\overbrace{a \cdots [a}^{n \text{ times}}, g] \cdots] = e,$$

where e is the identity of G. It follows that if a is an element of a normal nilpotent finite subgroup of G then a is weakly central. Another easy consequence of the definition is that if a is weakly central in G then a is its own normalizer in G if and only if $\{a\} = G$; here $\{a\}$ denotes the subgroup generated by a. It should also be noted that if a is weakly central in G, then $\overline{a}$ is weakly central in $\overline{G}$, where $\overline{a}$ is the image of a under a homomorphism which takes G onto $\overline{G}$.

Received July 2, 1952.

Pacific J. Math. 3 (1953), 501-504

2. Theorems. We shall establish the following results.

THEOREM 1. *If in a locally finite group G all elements whose orders are powers of a certain prime p are weakly central then they comprise a normal subgroup of G.*

An immediate consequence is the following analogue to Engel's Theorem for Lie algebras.

COROLLARY 1. *If all the elements of a locally finite group are weakly central then the group is the direct product of p-groups.*

THEOREM 2. *If G is a locally finite solvable group then the weakly central elements comprise a normal subgroup of G which is the direct product of p-groups.*

This result can also be stated as follows for finite groups.

THEOREM 2*a*. *If G is a finite solvable group then an element is weakly central if and only if it is in the nil radical of the group. Here the nil radical refers to the largest of all the nilpotent normal subgroups – nilpotent in the sense that $H^n = e$, where $H^n = [H^{n-1}, H]$* (cf. [1, pp. 98-102]).

It has not been determined whether solvability must necessarily be assumed for Theorems 2 and 2*a* to be true.

3. Proof of Theorem 1. We shall first consider the case where G is finite, and use induction on the order of G. Let p be the prime such that all elements of G whose orders are a power of p are weakly central. We must show that if S_0 is a p-Sylow subgroup of G then S_0 is the only p-Sylow subgroup, and hence is normal in G. We do this by obtaining a contradiction in case S_0 is not normal in G. Let $S_1, \cdots, S_k$ be the conjugate Sylow subgroups of S_0, and suppose first that $S_i \cap S_0 = \{e\}$ for $i = 1, \cdots, k$. If $N \neq G$ is the normalizer of S_0, then every element y of G not in N transform S_0 into one of the S_i. But then for $e \neq a \in S_0$ we have $a(y) \notin S_0$ and consequently $a(y) \notin N$, since

$$N \cap S_i = S_0 \cap S_i = \{e\},$$

and hence, for all positive integers n, $a^{(n)}(y) \notin S_0$. It follows that $a^{(n)}(y) \neq a$ for all n, and a is not weakly central, contrary to hypothesis.

Accordingly we need only consider the case where $S_i \cap S_0 = \{e\}$ for some i. Let D be a maximal intersection of two different Sylow subgroups. Then the normalizer N_D of D in G must have more than one p-Sylow subgroup [2, Chap. IV, Theorem 7]. It follows by our induction assumption that N_D must equal G; for if N_D were properly contained in G it would have but one p-Sylow subgroup, contrary to the above. But now if $N_D = G$, then D is normal in G, and the order of G/D is less than that of G; consequently, again by the induction assumption, G/D has but one p-Sylow subgroup. On the other hand $N_D = G$ has more than one p-Sylow subgroup containing D, and therefore so also has G/D; this again leads us to a contradiction. Thus in this case S_0 must be normal in G as the theorem asserts.

REMARK The above proof shows that a weakly central element of prime power order must lie in the intersection of at least 2 p-Sylow subgroups if the number of p-Sylow subgroups is greater than one.

We return to the proof of Theorem 1 and consider the case where G is locally finite. This means that any finite set of elements of G generates a finite subgroup of G. Now we are assuming that the elements belonging to a certain prime p are weakly central, and wish to show that they comprise a normal subgroup of G. It is obvious that they form an invariant set, and hence they generate a normal subgroup of G. Furthermore, the product of any two elements whose orders are powers of p has also order a power of p because of the local finiteness of G and because the theorem is true for finite groups. It follows that the elements whose orders are powers of p actually comprise the group they generate. This completes the proof of Theorem 1.

4. Proof of Theorem $2a$. From a previous remark we know that if an element is in the nil radical then it is weakly central. We must show conversely that if an element is weakly central then it is in the nil radical. The proof will be made by induction on the order of G. If the order is one then the theorem is obviously true. We now assume the theorem true for groups whose orders are less than k, and let G be a group of order k. Let N be the nil radical, and g a weakly central element of G. If $\{g, N\} \neq G$ then gN is weakly central in G/N, and hence by the induction assumption gN is contained in a proper normal subgroup M/N of G/N; (if the nil radical of G/N is not a proper subgroup of G/N then G/N is nilpotent and the statement is true since every proper subgroup of a finite nilpotent group is contained in a proper normal subgroup). It follows that g and N are contained in a proper normal subgroup M of G, and therefore by the induction assumption g is in the nil radical N_M of M; but N_M is contained in N since the nil radical is a characteristic subgroup (cf. [1, p.102]), and hence N_M is a normal nilpotent subgroup of G; therefore when $\{g, N\} \neq G$ then $g \in N$ as we wished to show.

We now consider the case where $\{g, N\} = G$. Let Z be the center of N; then Z is normal in G. Now if z is any element of Z such that $\{g, z\} \neq G$, then by the induction assumption $\{g, z\}$ is nilpotent and hence has a center Q. But $\{g, z\} \cap Z$ is normal in $\{g, z\}$ since Z is normal in G, and therefore

$$Q \cap \{g, z\} \cap Z \neq \{e\}$$

[2, Chap. IV, Theorem 14]. But then, if

$$H = Q \cap \{g, z\} \cap Z,$$

then H is in the center of G since $G = \{g, N\}$, and hence H is normal in G. It follows by the induction assumption that G/H is nilpotent, whence, for some k, $G^k \subseteq H$. But since H is in the center of G,

$$G^{k+1} = [G^k, G] \subseteq [H, G] = \{e\},$$

and therefore G is nilpotent; $G = N$, and $g \in N$ as was to be shown.

Accordingly we need now only consider the case where $\{g, z\} = G$ for every $z \in Z$. Since g is weakly central then $\{g\}$ cannot be its own normalizer in G; that is, $\{g\}$ is normal in R, where $R \neq \{g\}$. On the other hand, since $G = \{g, Z\}$, it follows that R or a subgroup of R is of the form $\{g, z\} = G$, so that $R = G$. Hence g is in a cyclic normal subgroup of G, and consequently is in the nil radical N as we wished to show. This completes the proof of Theorem $2a$.

5. **Proof of Theorem 2.** We first note that the product of two weakly central elements is weakly central since they generate a finite group in which Theorem $2a$ is applicable. Thus the weakly central elements comprise a subgroup which is obviously normal. It is the direct product of p-groups by Corollary 1.

References

1. H. Fitting, *Beiträge zur Theorie der Gruppen endlicher Ordnung*, Jber. Deutsch. Math. Verein. 48 (1938), 77-141.

2. H. Zassenhaus, *Theory of groups*, translated, Chelsea, New York, 1949.

Louisiana State University

A NOTE ON THE DIMENSION THEORY OF RINGS

A. SEIDENBERG

1. Introduction. Let O be an integral domain. If in O there is a proper chain

$$(0) \subset P_1 \subset P_2 \subset \cdots \subset P_n \subset (1)$$

of prime ideals, but no such chain

$$(0) \subset P_1' \subset \cdots \subset P_{n+1}' \subset (1),$$

then O will be said to be *n-dimensional*. Let O be of dimension n: the question is whether the polynomial ring $O[x]$ is necessarily $(n+1)$-dimensional. Here, as throughout, x is an indeterminate.

By an *F-ring* we shall mean a 1-dimensional ring O such that $O[x]$ is not 2-dimensional (i. e., the proposed assertion that $O[x]$ is necessarily 2-dimensional fails). Given an F-ring, we try by definite constructions to pass to a larger F-ring having the same quotient field: this restricts the class of rings in which to look for an F-ring—a priori we do not know they exist. In this way we also come (in Theorem 8 below) to a complete characterization of F-rings: if O is 1-dimensional, then $O[x]$ is 2-dimensional if and only if every quotient ring of $\bar{O}$, the integral closure of O, is a valuation ring. The rings $\bar{O}$ thus coincide (for dimension 1) with Krull's Multiplikationsringe [5; p. 554].

2. Preliminary results. The first five theorems are of a preparatory character, and the proofs offer no difficulties.

THEOREM 1. *Let O be an arbitrary commutative ring with 1, P_1, P_2, P_3 distinct ideals in $O[x]$. If $P_1 \subset P_2 \subset P_3$, and P_2 and P_3 are prime ideals, then P_1, P_2, P_3 cannot have the same contraction to O.*

Proof. Let

$$P_1 \cap O = P_2 \cap O = p,$$

Received May 15, 1952.
Pacific J. Math. 3 (1953), 505-512

and consider

$$O[x]/P_2 = \bar{O}[\bar{x}],$$

where $\bar{x}$ is the residue of x and $\bar{O} \simeq O/p$. Since

$$\bar{O}[x] \cdot p \subseteq P_1 \subset P_2,$$

$\bar{x}$ is algebraic over the integral domain $\bar{O}$. Let $\bar{P}_3$ be the image of P_3; then $\bar{P}_3 \neq (0)$; but also $\bar{P}_3 \cap \bar{O} \neq (0)$. In fact, let $\gamma \in \bar{P}_3$, $\gamma \neq 0$. Then

$$c_0 \gamma^n + c_1 \gamma^{n-1} + \cdots + c_n = 0$$

for some $c_i \in \bar{O}$, $c_n \neq 0$; and $c_n \in \bar{P}_3 \cap \bar{O}$. Hence also $P_3 \cap O \neq p$,

COROLLARY. *If O is* 1*-dimensional, and P_1, P_2, P_3 are distinct prime ideals in $O[x]$ different from* (0) *with $P_1 \subset P_2 \subset P_3$, then $P_1 \cap O = (0)$, P_2 is the extension of its contraction to O, and P_3 is maximal.*

Proof. If $P_1 \cap O \neq (0)$, then P_1, P_2, P_3 would all have to contract to the same maximal ideal in O. So

$$P_1 \cap O = (0) \text{ and } P_2 \cap O = p \neq (0).$$

Were $O[x] \cdot p \subset P_2$ properly, then, since $O[x] \cdot p$ is prime,

$$O[x] \cdot p \cap O = (0),$$

whereas

$$O[x] \cdot p \cap O = p.$$

So $O[x] \cdot p = P_2$. Were P_3 not maximal, we would have $P_2 \cap O = (0)$.

For the foregoing theorem, see also [4; Th. 10, p. 375].

THEOREM 2. *If O is n-dimensional, then $O[x]$ is at least $(n+1)$-dimensional and at most $(2n+1)$-dimensional.*

Proof. Let

$$(0) \subset P_1 \subset P_2 \subset \cdots \subset P_n \subset (1)$$

be a proper chain of prime ideals in O. Then

$$(0) \subset O[x] \cdot P_1 \subset O[x] \cdot P_2 \subset \cdots \subset O[x] \cdot P_n \subset (1)$$

is also a proper chain of prime ideals in $O[x]$; and $O[x] \cdot P_n$ is not maximal, since, for example,

$$O[x] \cdot P_n \subset (O[x] \cdot P_n, x) \subset (1).$$

(Here, as throughout, we use the symbol $\subset$ for proper inclusion.) Hence $O[x]$ is at least $(n+1)$-dimensional. Let now O be n-dimensional, and consider a chain

$$(0) \subset P_1' \subset \cdots \subset P_m' \subset (1)$$

of prime ideals in $O[x]$. Let there be s distinct ideals among the contractions

$$(0) \cap O, P_1' \cap O, \cdots, P_m' \cap O.$$

Then

$$m + 1 < 2s \leq 2(n+1), \text{ so } m \leq 2n + 1.$$

THEOREM 3. *If O is n-dimensional but $O[x]$ is not $(n+1)$-dimensional, then for at least one minimal prime ideal p of O either the quotient ring O_p is an F-ring or O/p is m-dimensional and $O/p[x]$ is not $(m+1)$-dimensional, and $m < n$.*

Proof. Suppose that for some minimal prime ideal p of O, $O[x] \cdot p$ is not minimal in $O[x]$; that is, there exists a prime ideal P such that

$$(0) \subset P \subset O[x] \cdot p.$$

Then

$$(0) \subset O_p[x] \cdot P \subset O_p[x] \cdot p$$

is also a chain of prime ideals in $O_p[x]$, as one easily verifies. Since $O_p[x] \cdot p$ is not maximal, this shows that O_p is an F-ring. We pass then to the case that $O[x] \cdot p$ is minimal for every minimal prime ideal p of O. Let

$$(0) \subset P_1' \subset \cdots \subset P_{n+2}' \subset (1)$$

be a chain of prime ideals in $O[x]$. If

$$P_1' \cap O = p \neq (0),$$

then O/p is at most $(n-1)$-dimensional, and $O[x]/O[x] \cdot p$ is a polynomial ring in one variable over O/p and is at least $(n+1)$-dimensional. So we must suppose

$$P_1' \cap O = (0);$$

but then

$$P_2' \cap O = p_2 \neq (0);$$

let p be a minimal prime ideal contained in p_2 — such exists since O is finite dimensional; then $O[x] \cdot p \subset P_2'$, properly, since $O[x] \cdot p$ is minimal but P_2' is not. Replacing P_1' by $O[x] \cdot p$, we come back to a previous case, and the proof is complete.

COROLLARY. *If O is an F-ring, then so is some quotient ring of O.*

The foregoing theorem shows that if for some n there exists a ring O which is n-dimensional, while $O[x]$ is not $(n+1)$-dimensional, then there exist F-rings. Thus we may provisionally confine our attention to 1-dimensional rings O.

THEOREM 4. *If O is 1-dimensional, and O is a valuation ring, then $O[x]$ is 2-dimensional.*

Proof. Let p be a proper prime ideal of O, and let

$$(0) \subset P \subseteq O[x] \cdot p,$$

where P is prime. Let

$$f(x) \in P, \quad f(x) \neq 0.$$

Then one can factor out from $f(x)$ a coefficient of least value, that is, write

$$f(x) = c \cdot g(x),$$

where $c \in p$, and $g(x)$ has at least one coefficient equal to 1; in particular, then $g(x) \notin O[x] \cdot p$; hence $c \in P$. So $P \cap O \neq (0)$, whence

$$P \cap O = p \quad \text{and} \quad P = O[x] \cdot p.$$

This proves that $O[x]$ is 2-dimensional (see Corollary to Theorem 1).

Theorem 4 restricts the size of an F-ring, since a maximal ring is a valuation ring. The following theorem reduces the considerations to integrally closed rings.

THEOREM 5. *Let $\overline{O}$ be the integral closure of the integral domain O. Then O is an F-ring if and only if $\overline{O}$ is an F-ring.*

Proof. Let R be an integral domain integrally dependent on O; a basic theoem of Krull (see, for example, [2; Th. 4, p. 254]) says that if $P_1 \subset P_2$ are prime ideals in R, then they contract to distinct prime ideals in O; hence $\dim R \leq \dim O$. Another theorem (loc. cit., p. 254) says that if $p_1 \subset p_2$ are prime ideals in O, and p_1 is a prime ideal in R contracting to p_1, then there exists a prime ideal P_2, $P_2 \supset P_1$, contracting to p_2. Hence $\dim R \geq \dim O$, and so $\dim R = \dim O$. Hence $\overline{O}$ is 1-dimensional if and only if O is 1-dimensional, and $\overline{O}[x]$ is 2-dimensional if and only if $O[x]$ is 2-dimensional.

Thus if there exist F-rings, then there exist integrally closed F-rings, and, taking an appropriate quotient ring, we see that there would exist an integrally closed F-ring O having just one proper prime ideal. In view of Theorem 4 (and the close association of integrally closed rings with valuation rings) one may ask whether an integrally closed ring with only one proper prime ideal is necessarily a valuation ring. Were it so, there would be no F-rings, but it is not so: Krull has an example [6; p. 670f]. For convenience, we may mention the example: let K be an algebraically closed field, x and y indeterminates; O consists of the rational functions $r(x, y)$ which, when written in lowest terms, have denominators not divisible by x, and which are such that $r(0, y) \in K$.

3. Principal results. We now establish:

THEOREM 6. *If O is integrally closed with only one maximal ideal p, α an element of the quotient field of O, and $1/\alpha \notin O$, then $O[\alpha] \cdot p$ is prime. If also $\alpha \notin O$, then $O[\alpha] \cdot p$ is not maximal.*

Proof. We first observe that

$$(O[\alpha] \cdot p, \alpha) \neq (1),$$

as an equation

$$1 = c_0 + c_1 \alpha + \cdots + c_s \alpha^s \qquad (c_0 \in p,\ c_i \in O),$$

leads to an equation of integral dependence for $1/\alpha$ over O. Let now $g(x) \in O[x]$ be a monic polynomial of positive degree. We may assume, trivially, that $\alpha \notin O$; then $g(\alpha) = c \in O$ is impossible, as $g(\alpha) - c = 0$ would be an equation of integral dependence for α over O; in particular, $g(\alpha) \neq 0$. Also $1/g(\alpha) \notin O$, for if it were in O, it would be a nonunit in O, and hence would be in p, so that

$$1 \in g(\alpha) \cdot p \subseteq O[\alpha] \cdot p,$$

and this is not so. By the result on α,

$$(O[g(\alpha)] \cdot p,\ g(\alpha)) \neq (1).$$

Since α satisfies $g(x) - g(\alpha) = 0$, $O[\alpha]$ is integral over $O[g(\alpha)]$; over any prime ideal in $O[g(\alpha)]$ containing $(O[g(\alpha)] \cdot p,\ g(\alpha))$, there lies a prime ideal in $O[\alpha]$, hence

$$(O[\alpha] \cdot p,\ g(\alpha)) \neq (1).$$

Since $1 + g(x)$ is monic of positive degree, also

$$(O[\alpha] \cdot p,\ 1 + g(\alpha)) \neq (1).$$

This shows that $g(\alpha) \notin O[\alpha] \cdot p$, a conclusion that also holds if $g(x)$ is of degree zero; that is, $g(x) = 1$.

We now prove that under the homomorphism $g(x) \longrightarrow g(\alpha)$ of $O[x]$ onto $O[\alpha]$, the inverse image of $O[\alpha] \cdot p$ is $O[x] \cdot p$; this will complete the proof, as $O[x] \cdot p$ is prime but not maximal. Let, then,

$$g(x) \in O[x],\ g(x) \notin O[x] \cdot p.$$

We write

$$g(x) = g_1(x) + g_2(x),$$

where $g_2(x) \in O[x] \cdot p$ and no coefficient of $g_1(x)$ is in p; in particular, this is so for the leading coefficient c. Then $g_1(\alpha)/c \notin O[\alpha] \cdot p$, since $g_1(x)/c$ is monic. A fortiori, $g_1(\alpha) \notin O[\alpha] \cdot p$, whence also $g(\alpha) \notin O[\alpha] \cdot p$.

COROLLARY. *In the case $\alpha \notin O$, if $g(x) \in O[x]$ and $g(\alpha) \in O[\alpha] \cdot p$, then $g(x) \in O[x] \cdot p$.*

THEOREM 7. *Let O be an integrally closed integral domain, p a proper ideal therein, a an element in the quotient-field of O, but $a \notin O_p$, $1/a \notin O_p$. Then $O[a] \cdot p$ is prime but not maximal; in fact,*

$$O[\alpha] \cdot p \cap O = p \quad \text{and} \quad O[\alpha]/O[\alpha] \cdot p \simeq O/p[x].$$

Proof. We know that $O_p[\alpha] \cdot p$ is prime, and

$$O_p[\alpha] \cdot p \cap O[\alpha] = O[\alpha] = O[\alpha] \cdot p$$

by the last corollary (and the fact that $O_p \cdot p \cap O = p$). Hence $O[\alpha] \cdot p$ is prime. Also here, as in the corollary, we have that if $g(x) \in O[x]$ and $g(\alpha) \in O[\alpha] \cdot p$, then $g(x) \in O[x] \cdot p$; the required isomorphism follows at once.

Theorem 7 is known in the case that O is a finite discrete principal order [3, §49, p.134-136]. The class of rings dealt with in the theorem includes this class properly; for example, the ring O of the example of Krull is not a finite discrete principal order, as $xy^{\rho} \in O$ for all ρ, but $y \notin O$.

THEOREM 8. *If O is 1-dimensional, then $O[x]$ is 2-dimensional if and only if every quotient ring of the integral closure of O is a valuation ring.*

Proof. By Theorem 5, we may assume O to be integrally closed. If O is an F-ring, then so is one of its quotient rings (Theorem 3, Corollary). This quotient ring is not a valuation ring (Theorem 4). Conversely, suppose some quotient ring $O_1 = O_p$ is not a valuation ring. Let α be an element of the quotient field of O_1 such that $\alpha \notin O_1$ and $\alpha^{-1} \notin O_1$. Then $O_1[\alpha]$ is at least 2-dimensional, by Theorem 6, and $O_1[x]$ is at least 3-dimensional, as one sees by considering the homomorphism of $O_1[x]$ onto $O_1[\alpha]$ determined by mapping x into α. So O_1 is an F-ring. Thus $O_p[x]\cdot p$ is not minimal in $O_p[x]$, and it follows at once that $O[x]\cdot p$ is not minimal in $O[x]$, whence O is an F-ring.

Let O be the ring of Krull's example above, and let X be an indeterminate. The single prime ideal p in O is constituted by the rational fractions $r(x, y)$ which, when written in lowest terms, have numerator divisible by x, i.e., are of the form $x\ g(x, y)$, where $g(x, y) \in K[x, y]$. The polynomials in $O[X]$ which vanish for $X = y$ form a prime ideal, different from (0) since $xX - xy$ is in it, properly contained in $O[X]\cdot p$.

The following theorem is well known [4, Th. 13, p.376].

THEOREM 9. *If O is a Noetherian ring of dimension n, then $O[x]$ is $(n+1)$-dimensional.*

Proof. Taking a quotient ring or residue class does not destroy the Noetherian character of O, so by Theorem 3 we may suppose O is 1-dimensional. Let then p be a proper prime ideal in O. Then $O[x]\cdot p$ is minimal for every principal ideal $O[x]\cdot(a)$, where $a \in p$, $a \neq 0$, so by the Principal Ideal Theorem [3, p.37], $O[x]\cdot p$ is minimal in $O[x]$, and $O[x]$ is 2-dimensional by Theorem 1, Corollary. — Instead of the Principal Ideal Theorem, one could use instead that the integral closure $\overline{O}$ is also Noetherian (see, for example, [1, Th. 3, p.29]; see also [3, §39, p.108]). Neither proof makes use of the full force of the quoted theorems, so it might be of some interest to find a direct proof using less technical means.

NOTE. In a forthcoming paper we will show that if O is a 1-dimensional ring such that $O[x]$ is 2-dimensional, then $O[x_1, \cdots, x_n]$ is $(n+1)$-dimensional. Theorem 2, above, will also be completed by examples showing that for any m, n with $n+1 \leq m \leq 2n+1$, there exist n-dimensional rings such that $O[x]$ is m-dimensional.

REFERENCES

1. I. S. Cohen, *Commutative rings with restricted minimum condition*, Duke Math. J. 17 (1950), 27-42.

2. I. S. Cohen and A. Seidenberg, *Prime ideals and integral dependence*, Bull. Amer. Math. Soc. 52 (1946), 252-261.

3. W. Krull, *Idealtheorie*, Ergebnisse der Mathematik und ihrer Grenzgebiete, 4, no. 3, Berlin, 1935.

4. ———, *Jacobsonsche Ringe, Hilbertscher Nullstellensatz, Dimensionstheorie*, Math. Z. 54 (1951), 354-387.

5. ———, *Beiträge zur Arithmetik kommutativer Integritätsbereiche, I. Multiplikationsringe, ausgezeichnete Idealsysteme, Kroneckersche Funktionalringe.* Math. Z. 41 (1936), 545-577.

6. ———, *Beiträge zur Arithmetik kommutativer Integritätsbereiche*, II. *v-Ideale und vollständig ganz abgeschlossene Integritätsbereiche.* Math. Z. 41 (1936), 665-679.

UNIVERSITY OF CALIFORNIA, BERKELEY.

Pacific Journal of Mathematics

Vol. 3, No. 2 April, 1953