UNIVERSAL
LIBRARY

OU_158853

UNIVERSAL
LIBRARY

OSMANIA UNIVERSITY LIBRARY

Call No. 517.38/ 151 T Accession No. 38768

Author

Title Symposium on partial differential equations. 1956.

This book should be returned on or before the date last marked below

TRANSACTIONS OF THE SYMPOSIUM

ON

PARTIAL DIFFERENTIAL EQUATIONS

held at the University of California, at Berkeley

June 20-July 1, 1955

Sponsored by

OFFICE OF NAVAL RESEARCH

UNIVERSITY OF CALIFORNIA, BERKELEY, CALIF

UNIVERSITY OF KANSAS, LAWRENCE, KAN

and

AMERICAN MATHEMATICAL SOCIETY

Editorial Committee

N ARONSZAJN - A. DOUGLIS - C B. MORREY, JR.

Reprinted from *Communications on Pure and Applied Mathematics*, a journal issued quarterly by the Institute of Mathematical Sciences of New York University; Volume IX, Number 3 (1956)

ALL RIGHTS RESERVED
Reproduction in whole or in part is permitted for any purpose of the United States Government

Library of Congress Catalog Card Number 56-13341

INTERSCIENCE PUBLISHERS, INC., 250 Fifth Avenue, New York 1, N.Y.

For Great Britain and Northern Ireland:
Interscience Publishers Ltd., 88/90 Chancery Lane, London W.C.2

PRINTED IN THE NETHERLANDS BY KONINKLIJKE VERENIGDE DRUKKERIJEN HOITSEMA N.V. GRONINGEN

NOTE

This volume contains all except four of the 45-minute lectures presented at the Berkeley Symposium on Partial Differential Equations which was sponsored by the Office of Naval Research, the University of California, the University of Kansas, and the American Mathematical Society. Of the four lectures not included, that of Professor P. D. Lax already has been published in Communications on Pure and Applied Mathematics, Vol. 8, p.p. 615—633 (1955), and that of Professor Paul Garabedian is to be published elsewhere. Garabedian's lecture also has appeared as Technical Report No. 42 of the Applied Mathematics and Statistics Laboratory at Stanford University. In addition, this volume contains a manuscript of an informal evening lecture by Professor Maynard Arsove and the manuscript of the lecture which was to have been given by Dr. J. Lehner of Los Alamos who was unable to attend at the last moment.

Besides the 45-minute lectures included in this volume, there were five series of hour lectures of a somewhat more expository character as follows:

M. M. SHIFFER, Linear elliptic differential equations I (three lectures)

PAUL GARABEDIAN, Linear elliptic differential equations II (three lectures)

E. HOPF, Non-linear partial differential equations (five lectures)

MARCEL RIESZ, Linear hyperbolic differential equations (five lectures)

P. D. LAX, Approximations to solutions of differential equations, etc. (four lectures)

These series are being published by the University of Kansas.

The editorial work for the conference was done by Professors N. Aronszajn of the University of Kansas, A. Douglis of New York University, and C. B. Morrey, Jr., of the University of California with the assistance of other members of the departments at the respective universities.

CONTENTS

PAGE

Note iii

Some criteria for normality of families of continuous functions.
By Maynard G. Arsove 1

Some hints on Huygens' principle and Hadamard's conjecture.
By Leifur Ásgeirsson 9

Multivalued harmonic functions in three variables.
By Stefan Bergman 29

Survey of local properties of solutions of elliptic partial differential equations.
By Lipman Bers 41

On the regularity properties of solutions of elliptic differential equations.
By Felix E. Browder 53

On the calculus of variations in two variables.
By Lamberto Cesari 65

Discontinuity of solutions of quasi-linear differential equations in two variables.
By Y. W. Chen 75

On singular and regular Cauchy problems.
By J. B. Diaz 85

A criterion for the validity of Huygens' principle.
By Avron Douglis 93

Singularities of generalized axially symmetric potentials.
By Arthur Erdélyi 105

Growth properties of solutions of non-linear elliptic equations.
By Robert Finn 117

Difference methods on a digital computer for Laplacian boundary value and eigenvalue problems.
By George E. Forsythe 127

On the local behavior of solutions of $\Delta u = g(x, u, \nabla u)$.
By Philip Hartman 137

The scope and limitations of the method of Wiener and Hopf.
By Albert E. Heins 149

On the existence problem for surfaces of constant mean curvature.
By Erhard Heinz 169

PAGE

On the reflection principle for polyharmonic functions.
By Alfred Huber 173

On linear differential equations in Banach spaces.
By Tosio Kato 181

An unsymmetric operator arising in the theory of neutron diffusion.
By Joseph Lehner 189

A variational method in the theory of harmonic integrals.
By Charles B. Morrey, Jr. 201

Estimates and existence of solutions of elliptic equations.
By Louis Nirenberg 211

New isoperimetric inequalities for eigenvalues and other physical quantities.
By L. E. Payne 233

Remarks on Courant's nodal line theorem.
By Åke Pleijel 245

Remarks on the application of gradient mappings to the calculus of variations and the connected boundary value problems in partial differential equations.
By E. H. Rothe 253

Mean values and continuity of Riesz potentials.
By Kennan T. Smith 271

On radiation conditions.
By J. J. Stoker 279

A direct approach to the problem of stability in the numerical solution of partial differential equations.
By John Todd 299

Upper and lower bounds for eigenvalues by finite difference methods.
By H. F. Weinberger 315

Some new methods in two-dimensional variational problems with special reference to minimal surfaces.
By L. C. Young 327

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 299–305 (1956)

Some Criteria for Normality of Families of Continuous Functions*

MAYNARD G. ARSOVE
University of Washington

1. Introduction

Restricting attention to the case of finite normality, we shall call a family $\mathcal{F}$ of real- or complex-valued functions on a region Ω a *normal family* provided every sequence of functions in $\mathcal{F}$ has a subsequence converging uniformly on compact subsets of Ω. For the sake of notational brevity Ω will be assumed to lie in the plane, but the arguments adapt readily to more general regions.

In practice, interest centers on families of functions which are at least continuous, and the usual approach to normality is based on the following corollary of Ascoli's theorem [10, pp. 54—55].

THEOREM 1. *A family $\mathcal{F}$ of continuous functions on Ω is normal if and only if for every compact subset K of Ω*

1) $\mathcal{F}$ is bounded on K and

2) $\mathcal{F}$ is equicontinuous on K.

Simple calculations reveal that for analytic and harmonic functions condition 2) can be omitted. What we propose to do, however, is to replace the equicontinuity condition 2) in Theorem 1 by a condition which is *trivially* satisfied by harmonic functions.[1] There results a criterion for normality (Theorem 2) which not only bears directly on the harmonic case but also shows that in every normal family of continuous functions the deviation from harmonicity must be uniformly small on compact sets (in a sense made precise by the theorem). Moreover, in this form it is easy to see that boundedness in condition 1) can be replaced by L^p boundedness for $p \geqq 1$.

A discussion of the normality criteria is given in §2, and in §3 we summarize the results of [4] applying these ideas to families of subharmonic functions.

---

*The research reported upon here was supported in part by the National Science Foundation.

[1]Analytic functions are included, as a special class of complex harmonic functions, and will henceforth not be referred to explicitly.

2. The Normality Criteria

For K a compact subset of Ω, r a positive number less than the distance from K to the boundary of Ω, and w a continuous function on Ω we write

$$\Lambda_w^K(r) = \sup_{z \in K} |w(z) - \alpha_r w(z)|,$$

where $\alpha_r w(z)$ denotes the areal mean of w over the neighborhood of radius r about z. Then if $\mathcal{F}$ is a family of such functions w, we define $\Lambda_{\mathcal{F}}^K(r)$ as the least upper bound of $\Lambda_w^K(r)$ as w varies over $\mathcal{F}$.

Since Λ_w^K vanishes for w harmonic, we can regard this function as expressing the extent to which w deviates from harmonicity on K. Similarly, $\Lambda_{\mathcal{F}}^K$ is a measure of the uniform deviation from harmonicity on K of the functions in $\mathcal{F}$.

As we proceed to show, a characterization of normal families of continuous functions is obtained when the equicontinuity condition of Theorem 1 is replaced by the condition that $\alpha_r w(z)$ converge to $w(z)$, as $r \to 0$, uniformly in z on compact sets and in w on $\mathcal{F}$.

THEOREM 2. *A family $\mathcal{F}$ of continuous functions on Ω is normal if and only if for all compact subsets K of Ω*

1) $\mathcal{F}$ is bounded on K and

2) $\lim_{r \to 0} \Lambda_{\mathcal{F}}^K(r) = 0.$

That *1)* and *2)* here are necessary for normality is obvious from Theorem 1. For the sufficiency we prove a result that is actually stronger than the one stated, by showing that continuity of the functions need not be assumed.[2] Indeed, continuity follows at once from *2)*, since for $z_1, z_2 \in K$ we have

$$|w(z_1) - w(z_2)| \leqq |\alpha_r w(z_1) - \alpha_r w(z_2)| + 2\Lambda_{\mathcal{F}}^K(r)$$

and the areal mean function $\alpha_r w$ is continuous.

It is well known that continuity of w gives rise to continuous partial differentiability of $\alpha_r w$; and, in fact, the partial derivatives of $\alpha_r w$ can be calculated according to the formulas [9, p. 11]

$$\frac{\partial \alpha_r w}{\partial x} = \frac{1}{\pi r} \int_0^{2\pi} w(z + re^{i\vartheta}) \cos \vartheta \, d\vartheta,$$

$$\frac{\partial \alpha_r w}{\partial y} = \frac{1}{\pi r} \int_0^{2\pi} w(z + re^{i\vartheta}) \sin \vartheta \, d\vartheta.$$

We next take M as an upper bound for all $|w|$ on a compact subset K' of Ω containing K in its interior and restrict r to be smaller than the distance from K to the boundary of K'. The mean-value theorem (applied separately

[2] We require only that the functions be summable on compact subsets of Ω.

to the real and imaginary parts of w) then yields

$$|\alpha_r w(z_1)-\alpha_r w(z_2)| \leqq \frac{8M}{r}|z_1-z_2|,$$

so that

$$|w(z_1)-w(z_2)| \leqq \frac{8M}{r}|z_1-z_2|+2\Lambda^K_{\mathscr{F}}(r).$$

Thus, $\mathscr{F}$ is equicontinuous on K, and the proof is complete.

It should be observed that condition *2)* of Theorem 2 does not by itself imply equicontinuity, a fact which is apparent from the consideration of non-equicontinuous families of harmonic functions (such as $w_n(x, y) = nx$, $n = 1, 2, \cdots$).

A further variant of Theorem 1 is obtained by replacing boundedness of the family $\mathscr{F}$ on K by L^p boundedness for $p \geqq 1$. We deal explicitly with the corresponding variant of Theorem 2.

THEOREM 3. *A family $\mathscr{F}$ of continuous functions on Ω is normal if and only if for every compact subset K of Ω*

1) $\mathscr{F}$ is a bounded subset of $L^p(K)$ for some (equivalently, all) $p \geqq 1$ and

2) $\lim_{r\to 0} \Lambda^K_{\mathscr{F}}(r) = 0.$

Conditions *1)* and *2)* are obviously necessary for normality, in view of Theorem 2. To show that they are also sufficient, we choose K' as a compact subset of Ω containing K in its interior and require r to be smaller than the distance from K to the boundary of K'. The Hölder inequality leads at once to the boundedness condition

$$|w(z)| \leqq (\pi r^2)^{-1/p}||w||_p+\Lambda^K_{\mathscr{F}}(r), \qquad z \in K, \tag{2.1}$$

where the indicated L^p norm is taken relative to K'. Normality then follows from Theorem 2.

The similar inequality

$$|w_j(z)-w_k(z)| \leqq (\pi r^2)^{-1/p}||w_j-w_k||_p+2\Lambda^K_{\mathscr{F}}(r), \qquad z \in K,$$

serves to establish

THEOREM 4. *Let $\mathscr{F}$ be a normal family of continuous functions on Ω, and let $\{w_n\}$ be a sequence of functions in $\mathscr{F}$. If $\{w_n\}$ converges in the L^p sense ($p \geqq 1$) on compact subsets of Ω, then $\{w_n\}$ converges uniformly on compact subsets of Ω.*

3. Applications to Subharmonic Functions

Normal families of subharmonic functions have been considered by Montel, who proved [7, p. 38] that the upper envelope of a normal family of

continuous subharmonic functions is also a continuous subharmonic function.

As one might expect from the basic role played by the mean-value operator α_r in the theory of subharmonic functions, the criteria for normality developed in §2 can be applied directly to subharmonic, and (more generally) δ-subharmonic,[3] functions. A number of these applications have been worked out in detail in [4], and we present here the main results.

Let us suppose that w is a δ-subharmonic function on Ω having the mass distribution m, and let us denote by $\Phi_w^z(t)$ the total variation of m on the neighborhood $N_t(z)$ of radius t about z. We then set

$$\Psi_w^K(r) = \sup_{z\in K} \int_0^r \Phi_w^z(t)\, d \log t.$$

Further, if $\mathscr{F}$ is any family of functions δ-subharmonic on Ω, we define $\Psi_{\mathscr{F}}^K(r)$ as the least upper bound of $\Psi_w^K(r)$ as w varies over $\mathscr{F}$. In this, as well as the ensuing discussion, it is assumed without further mention that r is always taken small enough so that all the neighborhoods involved have closures in Ω.

For any δ-subharmonic function w the areal mean function has the representation

$$\alpha_r w(z) = w(z) - \frac{2}{r^2}\int_0^r \varrho \int_{N_\varrho(z)} \log \frac{\varrho}{|z-\zeta|}\, dm(\zeta)\, d\varrho.$$

An integration by parts leads to

$$A_{\mathscr{F}}^K(r) \leqq \Psi_{\mathscr{F}}^K(r), \tag{3.1}$$

and from this it is plain that the uniform restriction on the behavior of the mass distributions expressed by $\lim_{r\to 0} \Psi_{\mathscr{F}}^K(r) = 0$ yields condition 2) of Theorems 2 and 3. There results

THEOREM 5. *Let $\mathscr{F}$ be a family of δ-subharmonic functions on Ω such that for all compact subsets K of Ω, $\mathscr{F}$ is bounded on K and $\lim_{r\to 0} \Psi_{\mathscr{F}}^K(r) = 0$. Then*

1) *$\mathscr{F}$ is a normal family,*
2) *all functions in $\mathscr{F}$ can be expressed as differences of continuous subharmonic functions, and*
3) *the mass distributions for functions in $\mathscr{F}$ have total variations uniformly bounded on compact subsets of Ω.*

If, further, $\{w_n\}$ is a sequence of functions in $\mathscr{F}$ converging to a function w on Ω, then

[3] For brevity, functions representable as differences of subharmonic functions are called *δ-subharmonic*; an introduction to the theory of such functions can be found in [3].

4) w can be expressed as the difference of two continuous subharmonic functions,

5) the mass distributions for $\{w_n\}$ converge vaguely[4] *to that for w, and*

6) $\Psi_w^K(r) \leqq \Psi_{\mathscr{F}}^K(r)$.

Assertions 4)—6) are important in establishing compactness[5] of certain families of δ-subharmonic functions.

THEOREM 6. *Let M and M' be constants, and let $\mathscr{F}$ be a family of δ-subharmonic functions w on Ω having mass distributions given by density functions ϱ_w. If the functions w satisfy*

1) $|w| \leqq M$ and 2) $|\varrho_w| \leqq M'$ almost everywhere,

then $\mathscr{F}$ is normal. If all functions satisfying 1) and 2) belong to $\mathscr{F}$, then it is compact.

Normality here follows readily from Theorem 5, since

$$\Psi_{\mathscr{F}}^K(r) \leqq \int_0^r M'\pi t^2\, d\log t = \frac{\pi r^2 M'}{2}.$$

Also, a result analogous to Theorem 6 can be derived for the case of mass distributions which all lie on a straight line and are given by uniformly bounded linear densities.

Turning next to the part played by the L^p norm in connection with normal families of δ-subharmonic functions, we make fundamental use of the following generalization of a theorem of Myškis [8].[6]

THEOREM 7. *Let $\mathscr{F}$ be a family of δ-subharmonic functions w on Ω having mass distributions given by density functions ϱ_w in L^p for $p > 1$. Then there is a positive constant $r(K, p)$, depending only on K and p, such that for $0 < r < r(K, p)$*

$$\Psi_w^K(r) \leqq ||\varrho_w||_p (2\pi r)^{1-1/p}$$

holds for all w in $\mathscr{F}$.[7]

Consequently, when the density functions form a bounded subset of $L^p(\Omega)$, we have $\lim_{r\to 0} \Lambda_{\mathscr{F}}^K(r) = 0$ for all compact sets K. A glance at Theorem 3 then suggests consideration of the space $\mathscr{L}^p$, $p > 1$, of all δ-subharmonic functions w on Ω with mass distributions given by density functions ϱ_w such that $||w||_p + ||\varrho_w||_p < +\infty$. Under the norm

[4] A sequence $\{m_n\}$ of mass distributions on a region Ω is said to converge *vaguely* to a mass distribution m on Ω provided $\lim_{n\to\infty} \int f\, dm_n = \int f\, dm$ for all continuous functions f supported by compact subsets of Ω (terminology of H. Cartan).

[5] If a normal family contains the limit of every convergent sequence of its elements, it is called *compact*

[6] A similar derivation has been employed by Frostman to prove an equicontinuity theorem for the α-potentials of M. Riesz (see [6] pp. 92–93).

[7] Here, and in what follows, $||\ ||_p$ will denote the L^p norm over Ω.

$$||w|| = [(||w||_p)^p + (||\varrho_w||_p)^p]^{1/p}$$

$\mathscr{L}^p$ is a separable, uniformly convex Banach space, in which the infinitely differentiable functions of finite norm are dense and all bounded subsets are normal.

Moreover, from (2.1), (3.1), and Theorem 7 we infer the inequality

$$|w(z)| \leqq M_z ||w||,$$

where M_z is a positive constant depending on z. Thus, the point functionals on $\mathscr{L}^p$ are continuous, and it follows [1, p. 138] that $\mathscr{L}^2$ admits a reproducing kernel.[8]

Although the membrane equation $\Delta w + \lambda w = 0$ has already been generalized in a variety of ways, we introduce a further generalization, closely related to $\mathscr{L}^p$ spaces. The generalized equation is

$$2\pi \varrho_w = \lambda w, \tag{3.2}$$

where λ is a bounded measurable function on Ω and we seek as solutions δ-subharmonic functions w having mass distributions with density ϱ_w. An important aspect of this equation is that the $\mathscr{L}^p$ norm defines the same topology as the L^p norm on the space of solutions w of (3.2). We therefore have at our disposal the normal family properties of $\mathscr{L}^p$ spaces when the usual L^p norm ($p > 1$) is used.

For an application of the above results we turn to the family of elliptic equations

$$\Delta w + \lambda w = 0$$

obtained by fixing constants $M > m > 0$ and taking λ as an arbitrary continuous function such that

$$m \leqq \lambda \leqq M.$$

In their study of this family of equations Bergman and Schiffer have shown [5, pp. 548—551], by means of kernel function theory, that the solutions which satisfy

$$\int_\Omega [(\text{grad } w)^2 + \lambda w^2] da \leqq 1 \qquad (a = \text{area}) \tag{3.3}$$

form a normal family. This is, however, also a consequence of the properties of $\mathscr{L}^p$, and we note that normality persists when (3.3) is replaced by the L^p boundedness condition $||w||_p \leqq 1$, $p > 1$.

[8] The subspace of $\mathscr{L}^2$ consisting of C'' functions satisfying certain special boundary conditions on regions Ω admitting a Green's function has been studied by Aronszajn [2] p. 407.

Bibliography

[1] Aronszajn, N., *La théorie des noyaux reproduisants et ses applications, I*, Proc. Cambridge Phil. Soc., Vol. 39, 1943, pp. 133-153.

[2] Aronszajn, N., *Green's functions and reproducing kernels*, Proceedings of the Symposium on Spectral Theory and Differential Problems, Oklahoma Agriculture and Mechanical College, Stillwater, Oklahoma, 1951, pp. 355–411.

[3] Arsove, M. G., *Functions representable as differences of subharmonic functions*, Trans. Amer. Math. Soc., Vol. 75, 1953, pp. 327–365.

[4] Arsove, M. G., *Normal families of subharmonic functions*, Proc. Amer. Math. Soc., Vol. 7, 1956, pp. 115–126.

[5] Bergman, S., and Schiffer, M, *Kernel functions in the theory of partial differential equations of elliptic type*, Duke Math. J., Vol. 15, 1948, pp. 535–566.

[6] Frostman, O., *Distributions de masses normées par la métrique de L^p*, Lund Univ. Math. Sem. Meddelanden, supplementary volume dedicated to Marcel Riesz, 1952, pp. 90–100.

[7] Montel, P., *Sur les fonctions convexes et les fonctions sousharmoniques*, J. Math. Pures Appl., Vol. 7, 1928, pp 29–60.

[8] Myškis, A. D., *A theorem on the convergence of a sequence of functions*, Uspekhi Matem. Nauk, Vol. 7, 1952, pp 186–190.

[9] Radó, T., *Subharmonic Functions*, Springer, Berlin, 1937.

[10] Saks, S., and Zygmund, A., *Analytic Functions*, Nakl. Polskiego Tow Matematycznego, Warsaw, 1952.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 307–326 (1956)

Some Hints on Huygens' Principle and Hadamard's Conjecture

LEIFUR ÁSGEIRSSON

Háskóli Íslands (University of Iceland), Reykjavik and New York University

This paper contains a summary of results on Huygens' Principle and Hadamard's Conjecture which I have obtained during the last twenty years in continuation of my work on the mean value theorem for ultrahyperbolic equations.[1] This is, however, the first public account of this work apart from some talks in a seminar on mathematical physics at New York University last fall. Some of my results overlap with the findings of other authors[2], but, it is hoped, may still be of interest because of the different procedures followed.

I would like to acknowledge my great indebtedness for information and inspiration to Hadamard's book "Le Problème de Cauchy" (Paris, 1932).

Huygens' Principle in the theory of second order linear partial differential equations of hyperbolic type has been formulated by Hadamard in Le Problème de Cauchy, pp. 75—77. We shall take it in the sense of his "minor premise", or rather, in the precise form which it can be given by an interpretation of solution formulae of Cauchy's Problem, as will be explained below.

I shall have to suppose my listeners and readers to be acquainted with the general theory of Cauchy's Problem, of which I shall recall only a few basic things.

Let $L[u]$ be the function in $x_1, \cdots, x_n$-space $E_n(X)$ which is produced by the operator

$$L \equiv \sum_{i,k=1}^{n} a_{ik} \frac{\partial^2}{\partial x_i \partial x_k} + 2\sum_{i=1}^{n} b_i \frac{\partial}{\partial x_i} + c$$

acting upon any regular function $u = u(x_1, \cdots, x_n) = u(X)$ in $E_n(X)$.

[1] Cf. Leifur Ásgeirsson, Über eine Mittelwertseigenschaft von Lösungen homogener linearer partieller Differentialgleichungen 2 Ordnung mit konstanten Koeffizienten, Math. Annalen, 1936, pp 321–346, see also Courant-Hilbert, Methoden der mathematischen Physik, Vol. II, pp. 417–425.

[2] In particular, M. Mathisson, Le problème de M. Hadamard relatif á la diffusion des ondes, Acta Math. 71, 1939, pp. 249–282, J. Hadamard, Diffusion of waves, Ann. of Math., 1942, pp. 510–522, and P. Gunther, Zur Gultigkeit des Huygensschen Prinzips bei partiellen Differentialgleichungen vom normalen hyperbolischen Typus, Akademie Verlag, Berlin, 1952.

Also the coefficients $a_{11}, \cdots, c$ are supposed to be regular functions in $E_n(X)$, that is, having continuous derivatives up to a certain order, not specified here, but required by the operations which are to be carried out. L is to be hyperbolic, that is, for any point X of the domain under consideration the signature of the quadratic form $\sum_{i,k=1}^{n} a_{ik}(X)\xi_i\xi_k$ is to be $++\cdots+-$ or $-++\cdots+$. We have to consider surfaces S characteristic with respect to L, or, which is the same, with respect to the principal part

$$L' = \sum_{i,k=1}^{n} a_{ik} \frac{\partial^2}{\partial x_i \partial x_k}$$

of L. If $F(X) = 0$ is the equation of S, then $\sum_{i,k=1}^{n} a_{ik} F_{x_i} F_{x_k} = 0$ is to hold on S, as a consequence of $F(X)=0$. A point transformation $x_i=\varphi_i(y_1, \cdots, y_n) = \varphi_i(Y)$ of $E_n(X)$ into $E_n(Y)$, which takes the operator L into an operator M in $E_n(Y)$, takes S into a surface T, characteristic with respect to M. Together with the first order partial differential equation $\sum_{i,k=1}^{n} a_{ik} F_{x_i} F_{x_k} = 0$ we introduce the characteristic curves of the surfaces $F(X)=0$ which are called the bicharacteristics pertaining to L; they may be determined without reference to any special characteristic surface. We may also consider the characteristic curves belonging to all equations $\sum_{i,k=1}^{n} a_{ik} F_{x_i} F_{x_k} = d = \text{const}$. For $d \neq 0$ they are the geodesics belonging to a metric $d\sigma = \left(\sum_{i,k=1}^{n} h_{ik} dx_i dx_k \right)^{1/2}$, where the matrices (a_{ik}) and (h_{ik}) are inverse (and therefore the signature of $\sum_{i,k=1}^{n} h_{ik}\xi_i\xi_k$ is the same as that of $\sum_{i,k=1}^{n} a_{ik}\xi_i\xi_k$). Then the characteristic curves corresponding to $d = 0$, that is, the bicharacteristics, are by extension called the geodesics of zero length. We consider a fixed point $X = X_0 = (x_{10}, \cdots, x_{n0})$, and all geodesics of zero length going through X_0; these form a cone-like surface C of two sheets, the characteristic conoid with respect to L with vertex X_0. If the coefficients a_{ik} are constants, the bicharacteristics are straight lines and the characteristic conoids are strictly cones.

Let Γ be the square of the geodesic distance, introduced by Hadamard, between X_0 and any (perhaps not too distant) point X. On C we have $\Gamma=0$. As a function of X, Γ is regular in a neighborhood of X_0. The geodesics, and together with them Γ, are invariant under a point transformation of $E_n(X)$, but only the geodesics of length zero are invariant under a multiplication of the coefficients a_{ik} by a common non-constant factor. Since such a multiplication is quite natural with respect to the equation $L[u] = 0$ or $L[u] =$

$f(X)$, this non-invariant behaviour of Γ detracts considerably from the usefulness of this function.

Now let a Cauchy Problem be posed for $L[u] = f(X)$; that is, let it be required that the value of u at any point X_0 outside a given non-characteristic initial surface I be expressed as a functional of u and its first derivatives on I. Usually I is taken to be spacelike, or such that any characteristic conoid with vertex X_1 on I has no other points in common with I. It is unnecessary to give all the derivatives $u_{x_1}, \cdots, u_{x_n}$ on I, since it will suffice to know one linear combination of them, representing a derivative leading out of I. From this combination and from the inner derivatives of u on I, which are known if u is known on I, all derivatives $u_{x_1}, \cdots, u_{x_n}$ may be determined on I. The combination of derivatives which usually is supposed to be given is the so-called transversal derivative, which will be denoted by u_ν; its coefficients do not matter here.

It is found that for a determination of $u(X_0)$, nothing need be known about u and u_ν on I outside the part I^* cut out of I by the characteristic conoid C with vertex X_0. But there exist operators L for which one needs to know u and u_ν only on the edge s of I^* (i.e. on the intersection of C and I), and in an immediate neighborhood of s. More exactly, only the values of u, $u_{x_1}, \cdots, u_{x_n}$, and, for $n > 4$, the values of higher derivatives of u up to a certain order are needed on s; but these higher order derivatives of u are derivable from our knowledge of u and u_ν on I in certain sets of points in a neighborhood of s. Exactly in this case we say that Huygens' Principle is valid for the operator L or for an equation $L[u] = f(X)$, or that L obeys Huygens' Principle or is governed by it. It is found that the form of $f(X)$ is irrelevant; in investigations about Huygens' Principle it is often convenient to put $f(X) = 0$.

All these considerations are required to hold only for points X_0 within a certain region to which I belongs, possibly as a boundary surface.

The name "data" for u and u_ν on I recalls the fact that more often than not the process of the determination of u at X_0 is supposed to be carried out under the assumption that u and u_ν are arbitrarily prescribed on I. We have to determine u, at least on one side of I, in such a way that it satisfies the equation $L[u] = f(X)$, and moreover u and its transversal derivative are to assume, in an exactly defined sence, the prescribed values on I.

Hadamard has proved that Huygens' Principle can never hold if n is an odd number. The case $n = 2$ having a special character, there remain the cases $n = 2m$ with $m = 2, 3, 4 \cdots$.

The known solution formulae for Cauchy's Problem show that, if Huygens' Principle holds, $u(X_0)$ may be expressed as an integral over s whose integrand depends linearly on $u, u_{x_1}, \cdots, u_{x_n}$, and, for $n > 4$, on the

above-mentioned higher derivatives. If $f(x) \not\equiv 0$, we have to add a term relating to f and to C^*, the part of C cut off by I; but, as already mentioned, we may put $f = 0$.

It is well known that the wave equations

$$\sum_{i=1}^{2m-1} u_{x_i x_i} - \frac{\partial^2 u}{\partial t^2} = f(x_1, \cdots, x_{2m-1}, t) = f(X, t),$$

where t plays the role of x_{2m}, obey Huygens' Principle. It is also easily seen that if Huygens' Principle is valid for an operator L or an equation $L[u] = 0$, then it is also valid for equations $L_1[u_1] = 0$ into which $L[u] = 0$ is taken by any combination of the following transformations:

1) Point transformation of $E_n(X)$ (or a transformation of coordinates), the function u remaining unchanged, or $u_1 = u$;

2) $u_1 = \varphi(X)u$, without a change of coordinates, $\varphi(X)$ being a given non-vanishing function; and

3) u and the coordinates are unchanged, but the operator L is multiplied by a factor $\psi(X)$, $L_1 = \psi(X)L$, where $\psi(X)$ is a given non-vanishing function.

In these cases we shall say that L_1 is equivalent to L with respect to Huygens' Principle.

"Hadamard's Conjecture" says that the class of equations which obey Huygens' Principle, or are of "Huygens' type", is given by the wave equations and equations equivalent to wave equations. Nowadays it is known, since Karl L. Stellmacher in Göttingen published, in 1953, a paper containing a counterexample (Ein Beispiel einer Huygensschen Differentialgleichung, Nachr. der Akad. d. Wissensch. in Göttingen), that for $n > 4$ the class of equations of Huygens' type is larger than that mentioned above; but the actual extent of this class still remains unknown. And even for the case $n = 4$, which is of special importance because of certain physical consequences, no complete treatment seems to have been published as yet.

Now let us consider the wave equations themselves. For $m = 2$, that is, for the ordinary wave equation, we have Poisson's Formula

$$u(X_0, \pm r) \equiv u(x_{10}, x_{20}, x_{30}, \pm r) = M(u, r, X_0) + rM(u_r, r, X_0) \pm rM(u_t, r, X_0),$$

where u_r is the radial derivative of $u(x_1, x_2, x_3, 0)$ in $E_3(X)$ with pole at X_0, and where $M(U, r, X_0)$ means the mean value of $U(X)$ on the sphere with radius r and center X_0 for $U = u, u_r, u_t$ on $t = 0$. This sphere is the intersection of the characteristic cone $(x_1 - x_{10})^2 + (x_2 - x_{20})^2 + (x_3 - x_{30})^2 - (t \mp r)^2 = 0$ with the initial plane $t = 0$, and Poisson's Formula thus directly shows the validity of Huygens' Principle for initial surfaces of this kind. Now it may be concluded from Hadamard's theory that if Huygens' Principle holds for

one family of initial surfaces, it must hold quite generally (I am indebted to Stellmacher for this remark). But "how" it comes about that Huygens' Principle holds is not easily seen from a look at Poisson's Formula. Here Beltrami has given an explanation which may be found in Courant-Hilbert, Methoden der mathematischen Physik, Vol. II, p. 371. Also for $m > 2$, there exist solution formulae of Cauchy's Problem for wave equations which make the validity of Huygens' Principle evident, but an answer to the "how" is not at all obvious.

In 1932 I had used the mean value theorem for ultrahyperbolic equations to derive a formula of the kind just named, with the initial surface given by $t = \text{const}$. There was little or nothing new about this formula except the mode of its derivation, but it gave me a starting point from which I was able a little later to find a differential identity pertaining to the wave operator for any of the values $m = 2, 3, 4 \cdots$, which gave an answer to the question, and at the same time suggested that a similar identity must hold for any operator obeying Huygens' Principle. For the sake of simplicity I shall here consider a solution of Cauchy's Problem only on one, say the "upper", side (with respect to t) of the initial surface, so that we have to consider lower characteristic semicones $\overline{C}$ given by the equation

$$t - t_0 + r = 0, \tag{1}$$

$$\text{with } r = [(x_1 - x_{10})^2 + (x_2 - x_{20})^2 + \cdots + (x_{2m-1} - x_{2m-1,0})^2]^{1/2}.$$

For the points of $\overline{C}$, I introduce a set $\xi_1, \cdots, \xi_{2m-1}$ of parameters by the equations

$$\begin{aligned} x_i &= x_{i0} + \xi_i \qquad \text{for } i = 1, \cdots, 2m-1, \\ t &= t_0 - (\xi_1^2 + \xi_2^2 + \cdots + \xi_{2m-1}^2)^{1/2} \equiv t_0 - \varrho; \end{aligned} \tag{2}$$

on $\overline{C}$ we then have $r = \varrho$.

I shall denote by $\partial/\partial r$ the radial derivative of a function, say $U(x_1, \cdots, x_{2m-1}, t) = U(X, t)$, considered as a function in $E_{2m-1}(X)$, with t fixed; as pole I take X_0. Thus

$$\frac{\partial}{\partial r} = \sum_{i=1}^{2m-1} \frac{x_i - x_{i0}}{r} \frac{\partial}{\partial x_i}. \tag{3}$$

I write

$$\frac{\partial}{\partial (r^2)} = \frac{1}{2r} \frac{\partial}{\partial r}. \tag{4}$$

Furthermore, I write

$$\Delta_{2m-1} = \sum_{i=1}^{2m-1} \frac{\partial^2}{\partial x_i^2}; \tag{5}$$

for $m = 2$, I shall also write simply Δ.

With these notations we have the identity, valid on $\overline{C}$ for any regular function $u = u(X, t)$,

$$\begin{aligned}(6)\qquad &\frac{1}{r^{2m-2}}\left(\frac{\partial}{\partial(r^2)}\right)^{m-2}\left[r^{2m-3}\left(\underset{2m-1}{\Delta}-\frac{\partial^2}{\partial t^2}\right)[u]\right]\\ &=\sum_{i=1}^{2m-1}\frac{\partial}{\partial\xi_i}\left[\frac{x_i-x_{i0}}{r^{2m-1}}\left(\frac{\partial}{\partial r}+\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial(r^2)}\right)^{m-2}[r^{2m-3}u]\right.\\ &\left.+\frac{1}{r^{2m-4}}\left(\frac{\partial}{\partial x_i}-\frac{x_i-x_{i0}}{r}\frac{\partial}{\partial r}\right)\left(\frac{\partial}{\partial(r^2)}\right)^{m-2}[r^{2m-5}u]\right].\end{aligned}$$

For (6) I also write

$$(6')\qquad R\left(\underset{2m-1}{\Delta}-\frac{\partial^2}{\partial t^2}\right)[u]=\sum_{i=1}^{2m-1}\frac{\partial A_i[u]}{\partial\xi_i}$$

defining in this way differential operators $R, A_1, \cdots, A_{2m-1}$. The differential expressions $A_i[u]$ at first appear as functions in the $x_1, \cdots, x_{2m-1}, t$-space $E_{2m}(X, t)$. We then restrict our considerations of these functions to the semicone $\overline{C}$, that is, we substitute $x_1, \cdots, x_{2m-1}, t$ from (2). After that we differentiate with respect to the parameters $\xi_1, \xi_2, \cdots, \xi_{2m-1}$. This means that we have to treat $A_1[u], A_2[u], \cdots, A_{2m-1}[u]$ as composite functions and apply the pertinent rules of differentiation, or

$$(7)\qquad \frac{\partial}{\partial\xi_i}=\frac{\partial}{\partial x_i}-\frac{\partial\varrho}{\partial\xi_i}\frac{\partial}{\partial t}=\frac{\partial}{\partial x_i}-\frac{\xi_i}{\varrho}\frac{\partial}{\partial t}=\frac{\partial}{\partial x_i}-\frac{x_i-x_{i0}}{r}\frac{\partial}{\partial t}.$$

For $m = 2$, (6) reduces to

$$(6^*)\qquad \frac{1}{r}\left(\Delta-\frac{\partial^2}{\partial t^2}\right)=\sum_{i=1}^{3}\frac{\partial}{\partial\xi_i}\left[\frac{1}{r}u_{x_i}+\frac{x_i-x_{i0}}{r^2}u_t+\frac{x_i-x_{i0}}{r^3}u\right];$$

this identity is equivalent to Beltrami's identity mentioned above.

I shall sketch here a proof of (6). We make use of

$$(8)\qquad \underset{2m-1}{\Delta}=\frac{\partial^2}{\partial r^2}+\frac{2m-2}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}H_{2m-1},$$

where H_{2m-1} stands for the second Lamé-Beltrami differential operator on the unit sphere in $E_{2m-1}(X)$. Let $r, \alpha_1, \cdots, \alpha_{2m-1}$ be polar coordinates of $E_{2m-1}(X)$, with pole at X_0; then we have an equation of the form

$$(9)\qquad H_{2m-1}=\sum_{i,k=1}^{2m-1}\varphi_{ik}(\alpha_1, \cdots, \alpha_{2m-1})\frac{\partial^2}{\partial\alpha_i\partial\alpha_k}+\sum_{i=1}^{2m-1}\psi_i(\alpha_1, \cdots, \alpha_{2m-1})\frac{\partial}{\partial\alpha_i}.$$

A rather simple calculation shows that for the right hand member of (6) we may write

$$r^{-(2m-2)}\left(\frac{\partial^2}{\partial r^2}-\frac{\partial^2}{\partial t^2}\right)\left(\frac{\partial}{\partial(r^2)}\right)^{m-2}[r^{2m-3}u]$$
$$+\,r^{-(2m-4)}\left(\Delta_{2m-1}-\frac{\partial^2}{\partial r^2}-\frac{2m-2}{r}\frac{\partial}{\partial r}\right)\left(\frac{\partial}{\partial(r^2)}\right)^{m-2}[r^{2m-5}u].$$

If in this expression and in the left hand member of (6), Δ_{2m-1} is substituted from (8) and use is made of obvious commutability properties of certain differential operators, (6) may be reduced to the equation

$$(10)\quad \left(\frac{\partial}{\partial(r^2)}\right)^{m-2}\left[r^{2m-3}\left(\frac{\partial^2}{\partial r^2}+\frac{2m-2}{r}\frac{\partial}{\partial r}\right)[u]\right]=\frac{\partial^2}{\partial r^2}\left(\frac{\partial}{\partial(r^2)}\right)^{m-2}[r^{2m-3}u],$$

where a factor $r^{-(2m-2)}$ has been dropped on both sides. A direct calculation shows that (10) is satisfied by $u = r^k$ for any constant k. This shows that (10) is an identity, valid for any m times differentiable function u. Otherwise (10) would be an ordinary linear differential equation in u as a function of r, of an order at most equal to m, and thus could not have more than m linearly independent solutions.

Since t_0 does not occur explicitly in (6), this identity must be valid not only on $\overline{C}$, but on a family of characteristic semicones of vertices $(x_{10}, \cdots, x_{2m-1,0}, t_0)$ with an arbitrary value of t_0.

The most important peculiarity of the identity (6) or (6′) is that its right hand member is a divergence. With the help of Gauss' Divergence (or Integral) Theorem, it may then be integrated directly over a domain in the $\xi_1, \xi_2, \cdots, \xi_{2m-1}$-space $E_{2m-1}(\xi)$ or, as we may also put it, over any part C' of $\overline{C}$, with respect to $d\xi_1 d\xi_2 \cdots d\xi_{2m-1}$, whereby we get integrals over the boundary, or the edge, of C'. In particular, we may suppose C' to be the part C^* of $\overline{C}$ which is cut off by an initial surface I. Because of singularities at the vertex (positive powers of r, or ϱ, as denominators) we have at first to exclude the vertex by cutting it off by a surface, say $I_\delta : t = t_0 - \delta$, where δ is a small positive number. Thus we get a family of surfaces C_δ^*, bounded by the intersections s and s_δ of $\overline{C}$ with I and I_δ respectively, and we may write the identity

$$(11)\quad \begin{aligned}&\int_{C_\delta^*} R\left(\Delta_{2m-1}-\frac{\partial^2}{\partial t^2}\right)[u]\,d\xi_1 d\xi_2 \cdots d\xi_{2m-1}\\ &=\int_s \sum_{i=1}^{2m-1} A_i[u]\cdot k_i\,ds^\xi+\int_{s_\delta}\sum_{i=1}^{2m-1} A_i[u]\cdot k_i\,ds_\delta^\xi.\end{aligned}$$

On the right hand side of (11), s and s_δ are to be considered as surfaces in $E_{2m-1}(\xi)$ with the Euclidean elements ds^ξ and ds_δ^ξ; then s_δ simply is the sphere $\xi_1^2+\xi_2^2+\cdots+\xi_{2m-1}^2=\delta^2$; $k_1, k_2, \cdots, k_{2m-1}$ are the components of the outer normal of s or s_δ relative to the domain in $E_{2m-1}(\xi)$ which corresponds to C_δ^*.

We now let δ tend to zero; the limit of

$$\int_s \sum_{i=1}^{2m-1} A_i[u] \cdot k_i \, ds_\delta^\xi$$

exists and is found to be $-4\pi^{m-1}u(X_0, t)$, so that we may write the identity

$$4\pi^{m-1}u(X_0, t_0) = \int_s \sum_{i=1}^{2m-1} A_i[u] \cdot k_i \, ds^\xi - \int_{C^*} R\left(\Delta_{2m-1} - \frac{\partial^2}{\partial t^2}\right)[u] d\xi_1 d\xi_2 \cdots d\xi_{2m-1}. \tag{12}$$

The identity (12) yields a solution formula of Cauchy's Problem for the equation

$$\left(\Delta_{2m-1} - \frac{\partial^2}{\partial t^2}\right)[u] = f(X, t). \tag{13}$$

The first term on the right hand side depends on the values of u and its derivatives up to order $m-1$ on s. The second term must be replaced by $\int_{C^*} R[f] d\xi_1 d\xi_2 \cdots d\xi_{2m-1}$. From the form of (12) the validity of Huygens' Principle is evident. But already the differential identity (6) gives an answer to the "how".

Setting $L_{2m-1}[u]$ for $\left(\Delta_{2m-1} - \frac{\partial^2}{\partial t^2}\right)[u]$, we may write (6') a little more briefly, leaving out u:

$$RL_{2m-1} = \sum_{i=1}^{2m-1} \frac{\partial}{\partial \xi_i} A_i \,. \tag{6''}$$

I note in passing that (6'') is a special case of an equation of the type

$$P_1 Q_1 + P_2 Q_2 + \cdots + P_h Q_h = 0 \tag{14}$$

where $P_1, \cdots, Q_h$ are linear differential operators.

Alluding to the ability of the differential operator R to convert the differential expression $L_{2m-1}[u]$ into a divergence expression with respect to $\xi_1, \xi_2, \cdots, \xi_{2m-1}$, I have used for R the name "diversor", but I also say that

$$L_{2m-1}[u] = \left(\Delta_{2m-1} - \frac{\partial^2}{\partial t^2}\right)[u]$$

is "quasidivergent" on $\overline{C}$.

It was my guess that an identity of the type (6'') would be valid for any linear second order differential operator L governed by Huygens' Principle, $\xi_1, \xi_2, \cdots, \xi_{2m-1}$ being a set of parameters for the characteristic semiconoid considered. It is easily seen that, if such an identity (with L_{2m-1} replaced by L) exists with respect to a particular set $\xi_1, \xi_2, \cdots, \xi_{2m-1}$ of parameters, then it also exists with respect to any set $\eta_1, \eta_2, \cdots, \eta_{2m-1}$ of parameters.

For, let $a_1, a_2, \cdots, a_{2m-1}$ be any vector in $E_{2m-1}(\xi)$. Putting

$$b_j = \left|\frac{\partial(\xi_1, \xi_2, \cdots, \xi_{2m-1})}{\partial(\eta_1, \eta_2, \cdots, \eta_{2m-1})}\right| \sum_{i=1}^{2m-1} \frac{\partial \eta_j}{\partial \xi_i} a_i,$$

we have as a corollary of Gauss' Divergence Theorem

$$\sum_{i=1}^{2m-1} \frac{\partial a_i}{\partial \xi_i} d\xi_1 d\xi_2 \cdots d\xi_{2m-1} = \sum_{j=1}^{2m-1} \frac{\partial b_j}{\partial \eta_j} d\eta_1 d\eta_2 \cdots d\eta_{2m-1}, \tag{15}$$

or

$$\left|\frac{\partial(\xi_1, \xi_2, \cdots, \xi_{2m-1})}{\partial(\eta_1, \eta_2, \cdots, \eta_{2m-1})}\right| \sum_{i=1}^{2m-1} \frac{\partial a_i}{\partial \xi_i} = \sum_{j=1}^{2m-1} \frac{\partial b_j}{\partial \eta_j}. \tag{15'}$$

Identifying a_i with $A_i[u]$, and then writing $B_j[u]$ for b_j, we find

$$\left|\frac{\partial(\xi_1, \xi_2, \cdots, \xi_{2m-1})}{\partial(\eta_1, \eta_2, \cdots, \eta_{2m-1})}\right| RL[u] = \sum_{j=1}^{2m-1} \frac{\partial B_j[u]}{\partial \eta_j}. \tag{15''}$$

Therefore we may take the operator $\left|\frac{\partial(\xi_1, \xi_2, \cdots, \xi_{2m-1})}{\partial(\eta_1, \eta_2, \cdots, \eta_{2m-1})}\right| R$ as diversor for L with respect to $\eta_1, \eta_2, \cdots, \eta_{2m-1}$.

It is easily seen that a change of coordinates does not affect the validity of an equation of the type of (6'').

An investigation of Hadamard's solution formulae for Cauchy's Problem for the case of an even number of variables (Le Problème de Cauchy, p. 311, eq. 29) enabled me to confirm my guess, and thus I could prove the following

THEOREM. *Let*

$$L[u] \equiv \sum_{i,\,k=1}^{2m} a_{ik}(X) u_{x_i x_k} + 2 \sum_{i=1}^{2m} b_i(X) u_{x_i} + c(X) u$$

be hyperbolic and obey Huygens' Principle, and let $\overline{C}$ be any characteristic semiconoid, relative to L, for which $\xi_1, \xi_2, \cdots, \xi_{2m-1}$ is a set of parameters. Then there exists a linear differential operator R, of order $m-2$ (which, for $m = 2$, reduces to a mere factor), and $2m-1$ linear differential operators $A_1, A_2, \cdots, A_{2m-1}$, of orders at most $m-1$, such that on $\overline{C}$ the equation

$$RL[u] = \sum_{i=1}^{2m-1} \frac{\partial A_i[u]}{\partial \xi_i} \tag{16}$$

holds as an identity in u. A solution formula of Cauchy's Problem for $L[u] = f(X)$ may be obtained by an integration of (16) with respect to $d\xi_1 d\xi_2 \cdots d\xi_{2m-1}$ over the part C^* cut off $\overline{C}$ by an initial surface I.

The diversor R necessarily contains, for $m > 2$, differentiations leading out of $\overline{C}$. R is not uniquely determined, since, if we put $\hat{R} = \lambda R + \sum_{j=1}^{2m-1} \frac{\partial}{\partial \xi_j} D_j$,

where λ is a constant $\neq 0$ and where $D_1, \cdots, D_{2m-1}$ are differential operators of orders $\leqq m-3$, we have

$$\hat{R}L[u] = \sum_{i=1}^{2m-1} \frac{\partial(\lambda A_i + D_i L)[u]}{\partial \xi_i} \equiv \sum_{i=1}^{2m-1} \frac{\partial \hat{A}_i[u]}{\partial \xi_i}, \tag{17}$$

an identity of the same type as (16). We then say that R and $\hat{R}$ are equivalent.

I have made a study of identities of the type (16) without supposing the surface involved, say S, to be a characteristic semiconoid, and without any restrictions on the dimension number of the space in which the operators L and R are defined, or on the order of these operators. However, S must be a characteristic surface with respect to L.

If S is a characteristic semiconoid, and the order of L is 2, one might feel inclined to guess that (16), with $2m$ replaced by any integer $n > 2$, is essentially equivalent to the validity of Huygens' Principle. This, however, is not necessarily the case. For odd values, Hadamard's Theorem gives a negative answer; but there is also a more direct explanation. I have found, for the operator of cylindrical waves, the identity

$$v(u_{x_1 x_1} + u_{x_2 x_2} - u_{tt}) = \sum_{i=1}^{2} \frac{\partial}{\partial \xi_i} \left[v \left(u_{x_i} + \frac{x_i - x_{i0}}{r} u_t \right) - v(x_i - x_{i0}) u \right] \tag{18}$$

with $v = r^{-1/2}(\alpha \cos \varphi/2 + \beta \sin \varphi/2)$, where α and β are constants, $r \cos \varphi = \xi_1$, $r \sin \varphi = \xi_2$, ξ_1 and ξ_2 being parameters of a characteristic semicone $t - t_0 \pm r = 0$ as defined by (2). Writing $\xi = \xi_1 + i\xi_2$, we have

$$v = \alpha \mathscr{R}e(\zeta^{-1/2}) - \beta \mathscr{I}m(\zeta^{-1/2}),$$

and v is seen to have, as a function in the plane of the parameters ξ_1, ξ_2, a singularity of the same type as $\zeta^{-1/2}$. Singlevaluedness is then achieved by making a cut from $\zeta = 0$ to $\zeta = \infty$. This means that we have to make a cut issuing from the vertex of the semicone. Of course, it will be simplest to make such a cut along a bicharacteristic.

Because of the circumstances just described, and because of the exponent $-1/2$ of r in the expression for v, equation (18) does not imply Huygens' Principle. However, operators for which an equation of the type (16) holds, with $2m$ replaced by n, obey Huygens' Principle in an extended sense (which of course does not mean Hadamard's "major premise").

Even though (18) does not suffice to establish Huygens' Principle, I have used this equation to solve Cauchy's Problem for the cylindrical wave equation, making the cut along a bicharacteristic and introducing a family of characteristic semicones, say $\overline{C}(\mu)$, each having its vertex on the cut-line on which μ is a parameter. An equation (18) is then associated with $\overline{C}(\mu)$ for every μ and is integrated with respect to $d\xi_1 d\xi_2$ over the part $C^*(\mu)$ cut

off $\overline{C}(\mu)$ by a common initial surface; thereafter an Abel integral equation has to be solved.[3]

It is well known that

$$v_3(x_1, x_2, x_3, t) = \frac{F(r \pm t)}{r}$$

is a solution of $L_3[v_3] = 0$, and it is also known (Volterra, Leçons sur l'intégration des équations différentielles aux dérivées partielles, professées à Stockholm, Hermann, Paris, 1912, p. 58) that

$$v_2(x_1, x_2, t) = \frac{\alpha \cos \varphi/2 + \beta \sin \varphi/2}{r^{1/2}} F(r \pm t)$$

is a solution of $L_2[v_2] = 0$, φ being given by

$$x_1 - x_{10} = r \cos \varphi, \qquad x_2 - x_{20} = r \sin \varphi;$$

in both cases F is an arbitrary function.

We have in v_2 and v_3 examples of "functions of wave families" or, briefly, "wave functions" to which I shall assign the "order" zero, that is, functions v whose most general form, if the variables are denoted by $(x_1, x_2, \cdots, x_n) = X$, is given as $v = G(X)F(A(X))$, where G and A are given functions of X, and F is an arbitrary function of one variable. Wave functions of an order $k > 0$ are given by the equation

$$v = G_0(X)F(A(X)) + G_1(X)F'(A(X)) + \cdots + G_k(X)F^{(k)}(A(X)),$$

$F', \cdots, F^{(k)}$ denoting derivatives. The functions v have to satisfy a certain linear partial differential equation of second order (see Courant-Hilbert, Methoden der mathematischen Physik, Vol. II, pp. 407 and 452).

Our examples of diversors for L_2 and L_3 suggest a connection between the existence of such wave functions, at least those of zero order, and the validity of Huygens' Principle in the extended sense. I have been able to prove that such a connection between wave functions and diversors of any order actually exists, in such a way that the wave functions v are solutions, not of $L[v] = 0$, but of the adjoint equations $M[v] = 0$; we suppose here that a diversor R for L exists for the whole family of "wave surfaces" $A(X) = \text{const}$. The order of the wave function is the same as that of R, and its form can be completely determined if the identity $RL[u] = \sum_{i=1}^{n} \partial A_i[u]/\partial \xi_i$ is exactly known. A proof which I have found can, although it is short, not be given within the frame of this paper. I shall also mention

[3] I have been informed by A Douglis that in a still unpublished dissertation (New York University, 1953) Clifford S. Gardner has used a similar method to solve certain initial value problems for wave equations.

that, in the theory of ultrahyperbolic differential equations, the diversor concept may be of importance. Thus Hans Lewy[4] and I have found independently that the differential expression

$$\sum_{i=1}^{2m+1} \frac{\partial^2 u}{\partial x_i^2} - \sum_{i=1}^{2m+1} \frac{\partial^2 u}{\partial y_i^2}$$

is, for $m = 1$, quasidivergent on the surfaces

$$\left(\sum_{i=1}^{2m+1} (x_i - x_{i0})^2\right)^{\frac{1}{2}} + \left(\sum_{i=1}^{2m+1} (y_i - y_{i0})^2\right)^{\frac{1}{2}} = \text{constant},$$

and that the mean value theorem for solutions of

$$\sum_{i=1}^{2m+1} \frac{\partial^2 u}{\partial x_i^2} - \sum_{i=1}^{2m+1} \frac{\partial^2 u}{\partial y_i^2} = 0$$

may be obtained by an integration of the corresponding divergence identity. A similar statement can be proved also for $m > 1$.

In the preceding considerations the case of two variables has been excluded. Yet this is exactly the case for which, up to now, diversor theory has been applied most fully, only without such a name and with an appearance different from my presentation. I mean the Laplace Method, also named the Cascade Method, for equations of the type

$$L[u] = u_{xy} + au_x + bu_y + cu = f, \tag{19}$$

a, b, c, f being functions of x and y. It may be proved that the cases in which the associated Darboux series (see G. Darboux, Théorie des Surfaces, Vol. II, p. 29) breaks off at least in one direction, and in which the method thus leads to a finite integration formula, are exactly those in which L is quasidivergent on the one or the other family of characteristic lines $x = \text{const.}$ or $y = \text{const.}$

I have already mentioned Stellmacher's counterexample to Hadamard's Conjecture, for $m = 3, 4, 5 \cdots$. Also for $m = 2$, counterexamples are known, but only if the coefficients of the equation, and with them the solution, are taken to be complex valued. I shall here try to give a brief treatment of the equation

$$\overline{L}[u] \equiv \left(\Delta_3 - \frac{\partial^2}{\partial t^2}\right)[u] + 2\sum_{i=1}^{3} a_i u_{x_i} + 2bu_t + cu = 0, \tag{20}$$

to which I shall refer as the "special case", in contrast to the "general case" of non-constant coefficients, about which I only shall give some hints. In (20) $a_1, \cdots, c$ are supposed to be regular functions of (x_1, x_2, x_3, t).

[4] Cf. Hans Lewy, Extension of Huygens' Principle to the ultrahyperbolic equation, Annali di Matematica, 1955.

For the sake of simplicity we shall only consider lower sheets $\overline{C}$ of characteristic semicones, with the equation

$$t-t_0+r=0, \qquad r=\left(\sum_{i=1}^{3}(x_i-x_{i0})^2\right)^{1/2}.$$

Let us make a transformation of coordinates: $\xi_i=x_i-x_{i0}$, $i=1, 2, 3$, $\tau=t-t_0+r$. On $\overline{C}$ we have $\tau=0$, and there ξ_1, ξ_2, ξ_3 may be made to play again the role of parameters as in some preceding considerations. Then we have, with $\varrho=(\xi_1^2+\xi_2^2+\xi_3^2)^{1/2}=r$,

$$\begin{aligned}\overline{L}[u]=A[u]\equiv&\sum_{i=1}^{3}\left(\frac{\partial^2 u}{\partial\xi_i^2}+2\frac{\xi_i}{\varrho}\frac{\partial^2 u}{\partial\xi_i\partial\tau}+2a_i\frac{\partial u}{\partial\xi_i}\right)\\&+2\varrho^{-1}\left(\sum_{i=1}^{3}a_i\xi_i+b\varrho+1\right)\frac{\partial u}{\partial\tau}+cu.\end{aligned}\tag{21}$$

If Huygens' Principle is to be valid for $\overline{L}$, there must exist on $\overline{C}$ a diversor factor (or "weight function") V such that

$$v\cdot A[u]=\sum_{i=1}^{3}\frac{\partial A_i[u]}{\partial\xi_i}\tag{22}$$

holds on $\overline{C}$ as an identity in u, where $A_1[u]$, $A_2[u]$, $A_3[u]$ are first order linear differential expressions in u. We find

$$\begin{aligned}V\cdot A[u]=&\sum_{i=1}^{3}\frac{\partial}{\partial\xi_i}\left[V\frac{\partial u}{\partial\xi_i}-u\frac{\partial V}{\partial\xi_i}+\frac{2\xi_i V}{\varrho}\frac{\partial u}{\partial\tau}+2a_i Vu\right]\\&+2\left(-\sum_{i=1}^{3}\frac{\xi_i}{\varrho}\frac{\partial V}{\partial\xi_i}+\varrho^{-1}\left(\sum_{i=1}^{3}a_i\xi_i+b\varrho-1\right)V\right)\frac{\partial u}{\partial\tau}\\&+\left(\sum_{i=1}^{3}\left(\frac{\partial^2 V}{\partial\xi_i^2}-2a_i\frac{\partial V}{\partial\xi_i}-2\frac{\partial a_i}{\partial\xi_i}V\right)+cV\right)u.\end{aligned}\tag{23}$$

It is now easy to show that the vanishing of the coefficients of u and of $\partial u/\partial\tau$ in (23), i.e., the conditions

$$-\sum_{i=1}^{3}\frac{\xi_i}{\varrho}\frac{\partial V}{\partial\xi_i}+\varrho^{-1}\left(\sum_{i=1}^{3}a_i\xi_i+b\varrho-1\right)V=0,\tag{24}$$

$$\sum_{i=1}^{3}\left(\frac{\partial^2 V}{\partial\xi_i^2}-2a_i\frac{\partial V}{\partial\xi_i}-2\frac{\partial a_i}{\partial\xi_i}V\right)+cV=0,\tag{25}$$

are necessary and sufficient for the validity of (22). On $\overline{C}$, or in ξ_1, ξ_2, ξ_3-space $E_3(\xi)$, with $\tau=0$, we thus have for V two partial differential equations, one of first order and the other of second order; with our mode of procedure this latter equation is elliptic. Introducing the radial derivative $\partial/\partial\varrho$ in $E_3(\xi)$, with pole at (0, 0, 0), which is the same as the derivative along the bicharac-

teristics issuing from the vertex of $\overline{C}$, with $\varrho = r$ as parameter, we may write for (24)

$$-\frac{\partial V}{\partial \varrho}+\varrho^{-1}\left(\sum_{i=1}^{3} a_i\xi_i+b\varrho-1\right)V = 0, \tag{26}$$

which gives, integrated,

$$V = \frac{\lambda}{\varrho}\exp\left\{\int_0^{\bar{\varrho}}\varrho^{-1}\left(\sum_{i=1}^{3} a_i\xi_i+b\varrho\right)d\varrho\right\} \tag{27}$$

with the integral taken along a bicharacteristic. There, ξ_i/ϱ is constant, that is, independent of ϱ, and $\varrho^{-1}\left(\sum_{i=1}^{3} a_i\xi_i+b\varrho\right)$ is bounded. In the integral, ϱ varies from 0 at the vertex to $\bar{\varrho}$ at the other end of the bicharacteristic segment considered. The constant of integration λ is a function of the parameters which fix the individual bicharacteristics; for such parameters we may take the angular coordinates, say θ and φ, in a system ϱ, θ, φ of polar coordinates in $E_3(\xi)$ with pole at (0, 0, 0). We may also think of λ as a function, which we must assume to be regular and accordingly bounded, on the unit sphere U: $\xi_1^2+\xi_2^2+\xi_3^2 = 1$. Now λ actually must be a constant, which may be taken equal to unity. This may be concluded from Hadamard's solution formula of Cauchy's Problem for the equation $\overline{L}[u] = 0$, but it also results as a first consequence of the subsistence of (25) at the vertex. By (8), in fact, we may write

$$\sum_{i=1}^{3}\frac{\partial^2}{\partial\xi_i^2}+\frac{\partial^2}{\partial\varrho^2} = \frac{2}{\varrho}\frac{\partial}{\partial\varrho}+\frac{1}{\varrho^2}H,$$

with H meaning H_3. If we use polar coordinates ϱ, θ, φ in $E_3(\xi)$ and expand the left hand member of (25) with V substituted from (27) into a power series in ϱ, which also will contain negative powers of ϱ, we get as a first term $H[\lambda]/\varrho^3$, and (25) then implies $H[\lambda] = 0$. Now this equation has no solution regular on U except $\lambda = \text{const}$. This follows most easily from the fact that in a stereographic projection of U on a plane with the rectangular coordinates α, β, $H[\lambda] = 0$ goes over into $\partial^2\lambda/\partial\alpha^2+\partial^2\lambda/\partial\beta^2 = 0$; λ is regular and bounded in the whole plane, and thus must be a constant, according to a well-known theorem of potential theory.

Thus V is the product of $\bar{\varrho}^{-1}$ and the function

$$\exp\left\{\int_0^{\bar{\varrho}}\varrho^{-1}\left(\sum_{i=1}^{3} a_i\xi_i+b\varrho\right)d\varrho\right\}$$

which continuously takes on the value 1 at the vertex.

The problem of compatibility of (24) and (25), with $\lambda = 1$ as an additional condition, may be tackled in a variety of ways. Thus one may, after

introduction of polar coordinates, use (24) to eliminate from (25) all derivatives with respect to ϱ. More generally we shall consider as coordinates not ϱ, θ, φ, but σ, θ, φ, where $\sigma = g(\varrho, \theta, \varphi)$ with $g_\varrho > 0$. Then the bicharacteristics still are given by $\theta = \text{const.}$, $\varphi = \text{const.}$, and we have

$$\frac{\partial V}{\partial \varrho} = \frac{\partial V}{\partial \sigma} \cdot g_\varrho .$$

First transforming equation (25) using the coordinates σ, θ, φ, and then eliminating all derivatives with respect to σ, we obtain equation

$$\alpha_{11} V_{\theta\theta} + 2\alpha_{12} V_{\theta\varphi} + \alpha_{22} V_{\varphi\varphi} + \beta_1 V_\theta + \beta_2 V_\varphi + \gamma V = 0, \tag{28}$$

where $\alpha_{11}, \cdots, \gamma$ are functions of σ, θ, φ. This last equation we shall consider for fixed, small values of σ; geometrically, this means a restriction of variability of the points (x_1, x_2, x_3, t) to a closed manifold T_σ on $\overline{C}$ near to the vertex, or of (ξ_1, ξ_2, ξ_3) to a closed manifold in $E_3(\xi)$, containing $(0, 0, 0)$ in its interior.

Since (28) was obtained by an elimination process from an elliptic equation and a first order equation, both in a 3-space, it must be elliptic. If we can show that either $\gamma \geqq 0$ or $\gamma \leqq 0$, but not $\gamma \equiv 0$, holds on T_σ, for some value of σ, we can from known theorems about maxima and minima of solutions of elliptic equations conclude that there must exist a point P of T_σ where V vanishes. But then (24) tells us that V must vanish along the whole bicharacteristic going through P, and therefore also at the vertex, contradicting (27) (with $\lambda = 1$).

As for γ, it will depend on the function $g(\varrho, \theta, \varphi)$. We can also consider transformations of $\overline{L}[u]$ of the kind which do not affect the validity of Huygens' Principle, especially the transformation $\overline{L}_1[u_1] = F^{-1}L[Fu_1]$, where F is an arbitrary function, $\neq 0$, in $E_4(X, t)$. Then $\overline{L}_1$ has the same form as $\overline{L}$ and may replace it for our purposes. We may thus let γ depend on two arbitrary functions, which generally may be chosen so as to insure the non-vanishing of γ on T_σ for sufficiently small values of σ. The exceptional cases are the only ones in which Huygens' Principle can be valid, and they have to be singled out, essentially by algebraic methods. As a matter of fact, we do not need both functions g and F for this purpose, or at least only to a limited extent.

The method just described was my original method for tackling the problem. In principle it may be generalized to equations with non-constant coefficients; linear homogeneous equations of the types of (24) and (25), that is, one first order equation which is an ordinary differential equation along any bicharacteristic and one second order equation, may be derived without difficulty. Also a second order equation corresponding to (28) which

contains derivatives with respect to two variables of a closed manifold may be derived similarly as above; and it may be proved that this equation is elliptic. But the calculational difficulties are increased to a very high degree; my treatment of the general problem along these lines proved cumbersome and has not been fully explored. One of the reasons for the unwieldiness of the method is that the expression obtained for γ contains derivatives of the square-of-distance function Γ up to second order, and for a power series expansion third and fourth, and possibly still higher, derivatives of Γ at the vertex are needed, and the structure of these derivatives is far from simple.

Even for the special case (21) of $\bar{L}[u] = 0$, a different more direct procedure, which I at first had avoided, proved to be considerably easier. Let us instead of

$$V = \frac{1}{\bar{\varrho}} \exp\left\{ \int_0^{\bar{\varrho}} \frac{1}{\varrho} \left(\sum_{i=1}^{3} a_i \xi_i + b\varrho \right) d\varrho \right\}$$

consider the function

$$v = \log(\bar{\varrho} V) = \int_0^{\bar{\varrho}} \varrho^{-1} \left(\sum_{i=1}^{3} a_i \xi_i + b\varrho \right) d\varrho \tag{29}$$

for which (25) goes over into

$$\sum_{i=1}^{3} \left(\frac{\partial^2 v}{\partial \bar{\xi}_i^2} + \left(\frac{\partial v}{\partial \bar{\xi}_i} \right)^2 - \frac{2}{\bar{\varrho}} \frac{\partial v}{\partial \bar{\varrho}} - 2a_i \frac{\partial v}{\partial \bar{\xi}_i} + \frac{2a_i \bar{\xi}_i}{\bar{\varrho}^2} - 2 \frac{\partial a_i}{\partial \bar{\xi}_i} \right) + c = 0, \tag{30}$$

where the variables in v (and V) are denoted by $\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3$, and where $\bar{\varrho} = (\bar{\xi}_1^2 + \bar{\xi}_2^2 + \bar{\xi}_3^2)^{1/2}$. As the bicharacteristics are represented by the straight lines in $E_3(\xi)$ issuing from the origin (0, 0, 0), we may express ξ_1, ξ_2, ξ_3, along a bicharacteristic joining (0, 0, 0) and $(\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3)$, by putting

$$\xi_i = \mu \bar{\xi}_i, \qquad \varrho = \mu \bar{\varrho}, \qquad \text{and } x_i = x_{i0} + \mu \bar{\xi}_i, \qquad t = t_0 - \mu \bar{\varrho}.$$

We then may write

$$\begin{aligned} v(\bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3) = \int_0^1 \Big[& \sum_{i=1}^{3} \bar{\xi}_i a_i(x_{10} + \mu \bar{\xi}_1, \cdots, t_0 - \mu \bar{\varrho}) \\ & + \bar{\varrho} b(x_{10} + \mu \bar{\xi}_1, \cdots, t_0 - \mu \bar{\varrho}) \Big] d\mu. \end{aligned} \tag{31}$$

It is now seen that the differentiations of v in (30) may be carried out under the sign of integration in (31). In the following calculation the arguments of a_1, a_2, a_3, b and their derivatives are to be $x_{10} + \mu \bar{\xi}_1, \cdots, t_0 - \mu \bar{\varrho}$ in the integrands, and $x_{10} + \bar{\xi}_1, \cdots, t - \bar{\varrho}$ outside the integrals. We find

$$v_{\bar{\xi}_i} = \int_0^1 \left[a_i + \mu \sum_{j=1}^{3} \bar{\xi}_j \left(\frac{\partial a_j}{\partial x_i} - \frac{\bar{\xi}_i}{\bar{\varrho}} \frac{\partial a_j}{\partial t} \right) + \frac{\bar{\xi}_i}{\bar{\varrho}} b + \mu \bar{\varrho} \left(\frac{\partial b}{\partial x_i} - \frac{\bar{\xi}_i}{\bar{\varrho}} \frac{\partial b}{\partial t} \right) \right] d\mu.$$

Integrating by parts,

$$\int_0^1 \left(a_i+\frac{\bar\xi_i}{\bar\varrho}\,b\right)d\mu = \int_0^1 \left[\frac{d}{d\mu}\left(\mu\left(a_i+\frac{\bar\xi_i}{\bar\varrho}\,b\right)\right)-\mu\frac{d}{d\mu}\left(a_i+\frac{\bar\xi_i}{\bar\varrho}\,b\right)\right]d\mu,$$

we find that we may write

$$\bar\varrho v_{\bar\xi_i} = I_i+\bar\varrho a_i+\bar\xi_i b, \tag{32}$$

where

$$I_i = \int_0^{\bar\varrho}\left[\sum_{j=1}^3 \xi_j\left(\frac{\partial a_j}{\partial x_i}-\frac{\partial a_i}{\partial x_j}\right)+\varrho\left(\frac{\partial a_i}{\partial t}+\frac{\partial b}{\partial x_i}\right)-\frac{\xi_i}{\varrho}\sum_{j=1}^3 \xi_j\left(\frac{\partial a_j}{\partial t}+\frac{\partial b}{\partial x_j}\right)\right]d\varrho.$$

Similarly we find, with

$$\Box = \sum_{i=1}^3 \frac{\partial^2}{\partial x_i^2}-\frac{\partial^2}{\partial t^2}, \qquad D = \sum_{i=1}^3 \frac{\partial a_i}{\partial x_i}+\frac{\partial b}{\partial t},$$

$$\begin{aligned}\bar\varrho^2\sum_{i=1}^3 \frac{\partial^2 v}{\partial \bar\xi_i^2} = &\int_0^{\bar\varrho}\left[\varrho\sum_{i=1}^3 \xi_i\left(\Box a_i-\frac{\partial D}{\partial x_i}\right)+\varrho^2\left(\Box b+\frac{\partial D}{\partial t}\right)\right]d\varrho\\ &+\bar\varrho^2\left(\sum_{i=1}^3 \frac{\partial a_i}{\partial x_i}-\frac{\partial b}{\partial t}\right)-2\bar\varrho\sum_{i=1}^3 \bar\xi_i\frac{\partial a_i}{\partial t}+2\bar\varrho b.\end{aligned} \tag{33}$$

Making use of the equation $\sum_{i=1}^3 \bar\xi_i I_i = 0$, we finally get from (31) and (30)

$$\begin{aligned}&\sum_{i=1}^3\left[\int_0^{\varrho}\sum_{j=1}^3\left(\xi_j\left(\frac{\partial a_j}{\partial x_i}-\frac{\partial a_i}{\partial x_j}\right)+\varrho\left(\frac{\partial a_i}{\partial t}+\frac{\partial b}{\partial x_i}\right)-\frac{\xi_i}{\varrho}\sum_{j=1}^3 \xi_j\left(\frac{\partial a_j}{\partial t}+\frac{\partial b}{\partial x_j}\right)\right)d\varrho\right]^2\\ &+\int_0^{\bar\varrho}\left(\varrho\sum_{i=1}^3 \xi_i\left(\Box a_i-\frac{\partial D}{\partial x_i}\right)+\varrho^2\left(\Box b+\frac{\partial D}{\partial t}\right)\right)d\varrho\\ &+\bar\varrho^2\left(c-\sum_{i=1}^3 a_i^2+b^2-D\right) = 0.\end{aligned} \tag{34}$$

Here we obviously may write $O(\bar\varrho^4)$, $O(\bar\varrho^3)$, $O(\bar\varrho^2)$ for the first, second and third left hand term, respectively (the sum $\sum_{i=1}^3$ being considered as a single term). From this we conclude that at the vertex, we must have

$$c-\sum_{i=1}^3 a_i^2+b^2-\sum_{i=1}^3 \frac{\partial a_i}{\partial x_i}-\frac{\partial b}{\partial t} = 0. \tag{35}$$

Now we may take any point as vertex, so that (35) must be true generally, and we may cancel the last term of (34). If we thereafter divide (34) by $\bar\varrho^3$ and then let $\bar\varrho$ go to zero, $(\bar\xi_1, \bar\xi_2, \bar\xi_3)$ sliding along a fixed bicharacteristic with $\bar\xi_i/\bar\varrho_i = \eta_i = \text{const.}$, we get at the vertex

$$\sum_{i=1}^3 \eta_i\left(\Box a_i-\frac{\partial D}{\partial x_i}\right)+\left(\Box b+\frac{\partial D}{\partial t}\right) = 0. \tag{36}$$

Equation (36) has to hold for any unit vector (η_1, η_2, η_3), which implies that each term of the left hand side of (36) must vanish:

$$(37)\qquad \begin{aligned} &\sum_{j=1}^{3} \frac{\partial^2 a_i}{\partial x_j^2} - \frac{\partial^2 a_i}{\partial t^2} - \frac{\partial}{\partial x_i}\left(\sum_{j=1}^{3} \frac{\partial a_j}{\partial x_j} + \frac{\partial b}{\partial t}\right) = 0, \qquad i = 1, 2, 3, \\ &\sum_{j=1}^{3} \frac{\partial^2 b}{\partial x_j^2} - \frac{\partial^2 b}{\partial t^2} + \frac{\partial}{\partial t}\left(\sum_{j=1}^{3} \frac{\partial a_j}{\partial x_j} + \frac{\partial b}{\partial t}\right) = 0. \end{aligned}$$

(37) must hold for any point in the same way as (35), and the second term of (34) may be cancelled.

Because of the definiteness of the first term of (34), each individual integral must vanish. From this we conclude that the following equations must hold at the vertex, and therefore at any point:

$$(38)\qquad \begin{aligned} &\frac{\partial a_j}{\partial x_i} - \frac{\partial a_i}{\partial x_j} = 0, \qquad i, j = 1, 2, 3, \\ &\frac{\partial a_i}{\partial t} + \frac{\partial b}{\partial x_i} = 0, \qquad i = 1, 2, 3. \end{aligned}$$

The equations (37) are seen to be a consequence of (38) and thus need not be considered separately.

From (38) we conclude that there must exist a function $F \neq 0$ such that

$$(39)\qquad \begin{aligned} a_i &= \frac{\partial \log F}{\partial x_i} = \frac{1}{F}\frac{\partial F}{\partial x_i}, \\ -b &= \frac{\partial \log F}{\partial t} = \frac{1}{F}\frac{\partial F}{\partial t}. \end{aligned}$$

Then (35) gives

$$(40)\qquad c = \frac{1}{F}\left(\sum_{i=1}^{3} \frac{\partial^2 F}{\partial x_i^2} - \frac{\partial^2 F}{\partial t^2}\right).$$

Finally we find

$$(41)\qquad \begin{aligned} \bar{L}[u] &\equiv \sum_{i=1}^{3} \frac{\partial^2 u}{\partial x_i^2} - \frac{\partial^2 u}{\partial t^2} + 2\sum_{i=1}^{3} a_i \frac{\partial u}{\partial x_i} + 2b\frac{\partial u}{\partial t} + cu \\ &= \frac{1}{F}\left(\sum_{i=1}^{3} \frac{\partial^2 (Fu)}{\partial x_i^2} - \frac{\partial^2 (Fu)}{\partial t^2}\right), \end{aligned}$$

and thus $\bar{L}[u]$ must be equivalent to the ordinary wave equation, in full accordance with Hadamard's Conjecture.

A proof of this fact was, I think, first published by Mathisson, who used a method different from those described in this paper; namely operating not on the diversor factor V, but on another function, the so-called parametrix. However, both functions may be shown to be closely related.

Later Hadamard published a proof of the same fact. But already his

book Le Problème de Cauchy contains two equations, namely (39′) on p. 133, with an integral counterpart (41) on p. 141, and an unnumbered equation on p. 141, on which an investigation of Hadamard's Conjecture for a space of four variables may be based also in the general case of nonconstant coefficients.

Let $L[u] = 0$ be the equation of Huygens' type, in a 4-space, which is to be investigated. Then for Hadamard's equation E on p. 104, for which he writes $F(u) = 0$ and from which he derives his just mentioned equations (39′) and (41) with an object-function U, one has to take the adjoint of $L[u] = 0$. For $L[u] \equiv \overline{L}[u]$, (39′), considered at first on a characteristic semicone $\overline{C}$, is seen to be equivalent to equation (24) of this paper; also, (41) and our equation (27), with $\lambda = 1$, correspond to each other, giving $U = \text{const.}\ \varrho V$. Now Hadamard, by an extension, supposes (39′) and (41) to hold on any geodesic issuing from the vertex, and not only on the geodesics of length zero. In this way a regular function U is defined in a domain, say K, containing the vertex. In K we then may consider the expression $F(U)$; Hadamard's unnumbered equation on p. 141 then says that in case of validity of Huygens' Principle we must have $F(U) = 0$ valid on $\overline{C}$, but generally not in K. It may be proved that our equation (25) with $V = \text{const.}\ U/\varrho$ can be derived from this equation, together with Hadamard's equation (39′).

Even though the extension of Hadamard's equations from $\overline{C}$ to K contains a certain disadvantageous arbitrariness, it brings at the same time the advantage of more symmetry to the analytic problem, bought, as often is the case, at the price of an additional equation, namely that of $\overline{C}$, or $I = 0$.

In seminar talks given last fall at the Institute of Mathematical Sciences of New York University, I mentioned the existence of equations of the form $\overline{L}[u] = 0$ with complex coefficients which obey Huygens' Principle. In this case the definiteness of the first term of (34) becomes meaningless, and therefore we do not arrive at the system (38), but at another less stringent one, namely

$$
\begin{aligned}
&\frac{\partial a_3}{\partial x_2} - \frac{\partial a_2}{\partial x_3} = \varepsilon i \left(\frac{\partial a_1}{\partial t} + \frac{\partial b}{\partial x_1} \right), \\
&\frac{\partial a_1}{\partial x_3} - \frac{\partial a_3}{\partial x_1} = \varepsilon i \left(\frac{\partial a_2}{\partial t} + \frac{\partial b}{\partial x_2} \right), \\
&\frac{\partial a_2}{\partial x_1} - \frac{\partial a_1}{\partial x_2} = \varepsilon i \left(\frac{\partial a_3}{\partial t} + \frac{\partial b}{\partial x_3} \right), \qquad \varepsilon = \pm 1.
\end{aligned} \tag{42}
$$

Later my attention was called to Hadamard's paper in Ann. of Math., 1942, where, by way of an example, he has shown the existence of Huygens' type equations of the form $\overline{L}[u] = 0$, with complex coefficients.

A comparison of (42) with the equations of electrodynamics shows that the coefficients a_1, a_2, a_3, b (or rather a_1, a_2, a_3, ib) may be considered to form a four-vector potential, giving rise to a six-vector which is "self-dual" or "oppositely self-dual". In a manner well known in electrodynamics, this six-vector may be considered as an array containing the components of $H+iE$, or of $H-iE$, where H is the magnetic vector and E the electric vector of an electrodynamic field.

From (42) it follows that Maxwell's Equations for vacuum will be satisfied.

If we conversely take H and E satisfying Maxwell's Equations for vacuum to be given primarily, the vector $(a_1, a_2, a_3, -b)$ will be fixed except for an additive (complex) gradient vector, say $(\partial A/\partial x_1, \partial A/\partial x_2, \partial A/\partial x_3, \partial A/\partial t)$, and with a corresponding degree of indeterminacy, we get an equation $\overline{L}[u] = 0$, or rather an operator $\overline{L}$, of Huygens' type, which may be said to be associated with the electrodynamic process in question.

A connection of Maxwell's Equations with the equations of Huygens' type, namely the equation (37) and the equations of the general case corresponding to (37), has also been noted by Paul Günther in his paper, already cited. He does not, however, consider complex four-vectors, nor does he establish the existence of an associated operator $\overline{L}$, since he is mostly concerned with the general case, and from the point of view of real numbers. For this case he gives sets of necessary conditions for the validity of Huygens' Principle.

I have included in this paper more material than I was able to present within the time limits of my talk at the Conference; still many proofs and a fuller account of my investigations had to be left out.

Received January 20, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 327–338 (1956)

Multivalued Harmonic Functions in Three Variables*

STEFAN BERGMAN
Stanford University

1. Introduction

The basic idea of the method of the integral operators in the theory of partial differential equations is to introduce a mapping of solutions of a differential equation onto the space of analytic functions of one or more complex variables. Since, in contrast to the solutions of an equation, the analytic functions form an algebra (i.e., the sum *and* the product of two analytic functions yield again analytic functions) and also the theory of functions is more developed than the theory of solutions of differential equations, we can use properties of analytic functions in order to deduce various theorems about differential equations.

In the case of linear differential equations in two variables with analytic coefficients (and of elliptic type), the theory of integral operators has been developed to a large extent. These operators generalize the operator "$\mathfrak{Re}$" (= take the real part) for harmonic functions of two variables (see [2, 3, 4, 5, 12, 13, 15, 16]). It is the aim of this paper to indicate how to proceed in the case of equations in *three* variables. It seems that in this case the choice of analytic functions of *two* complex variables will serve us better than the choice of analytic functions of one complex variable (see [1, 6, 7, 8, 9, 10]).

In the present article we shall limit ourselves to the consideration of harmonic functions, but we wish to stress that similar procedures can be used for the study of other differential equations in three variables (see also Section 6).

In particular, using the mapping mentioned above we shall derive various properties of certain classes of multi-valued harmonic functions.

2. An Operator Transforming Analytic Functions of Two Variables into Harmonic Functions of Three Variables

Let $H(x, Z, Z^*) \equiv \tilde{H}(x, y, z)$, where $Z = (iy+z)/2$, $Z^* = (iy-z)/2$ denote solutions of the differential equation

*The work for this paper has been supported by the Office of Naval Research.

$$\frac{\partial^2 \tilde{H}}{\partial x^2}+\frac{\partial^2 \tilde{H}}{\partial y^2}+\frac{\partial^2 \tilde{H}}{\partial z^2}\equiv\frac{\partial^2 H}{\partial x^2}-\frac{\partial^2 H}{\partial Z \partial Z^*}=0. \tag{2.1}$$

Furthermore, let $\{\chi(\tau, \tau^*)\}$ be the algebra of analytic functions of the form

$$\chi(\tau, \tau^*)=2(\tau\tau^*)^{1/2}\chi_1(\tau, \tau^*)+\chi_2(\tau, \tau^*) \tag{2.2}$$

where χ_1 and χ_2 are analytic functions of two complex variables, τ and τ^*, and are regular at the origin.

Then the integral operator $\mathbf{C}_3(\chi)\equiv\mathbf{C}_3(\chi, L; \boldsymbol{X}_0)$,

$$\mathbf{C}_3(\chi)\equiv\frac{1}{\pi i}\int_L\int_{T=0}^1 u^{1/2}\frac{d[u^{1/2}\chi(u\zeta^{-1}T^2, u\zeta(1-T)^2)]}{du}dT\frac{d\zeta}{\zeta}, \tag{2.3}$$

$$\boldsymbol{X}\in N(\boldsymbol{X}_0),\qquad \boldsymbol{X}=(x, y, z),\qquad \boldsymbol{X}_0=(x_0, y_0, z_0),$$

where $u\equiv x+Z\zeta+Z^*\zeta^{-1}$ and L is a rectifiable (closed or open) curve in the ζ-plane, transforms functions $\chi(\tau, \tau^*)$ mentioned above into harmonic functions $H(\boldsymbol{X})\equiv\tilde{H}(x, Z, Z^*)$, provided that the differentiation under the integral sign is permitted for $\boldsymbol{X}\in N(\boldsymbol{X}_0)$. $N(\boldsymbol{X}_0)$ denotes a sufficiently small neighborhood of the point $\boldsymbol{X}_0$.

If L is the curve $[|\zeta|=1]$ and $\boldsymbol{X}_0=0$, then the inverse mapping is simple: if $H(x, Z, Z^*)$ is a solution of (2.1) which is analytic at the origin, then

$$\chi(Z, Z^*)=H(2(ZZ^*)^{1/2}, Z, Z^*). \tag{2.4}$$

Remark: Observe that $\chi(Z, Z^*)$ is the restriction of the values of $H(x, Z, Z^*)$ in the characteristic space $x^2-4(ZZ^*)\equiv x^2+y^2+z^2=0$.

The mappings (2.3) can be written in the form

$$\mathbf{C}_3(\chi)=\mathbf{B}_3(f, L; \boldsymbol{X}_0),\qquad \mathbf{B}_3(f, L; \boldsymbol{X}_0)\equiv\frac{1}{2\pi i}\int_L f(u, \zeta)\frac{d\zeta}{\zeta}, \tag{2.5}$$

$$f(u, \zeta)=2\int_0^1 u^{1/2}\frac{d[u^{1/2}\chi(u\zeta^{-1}T^2, u\zeta(1-T)^2)]}{du}dT \tag{2.6}$$

where $f(u, \zeta)$ is an analytic function of two variables u and ζ (see [1, 6, 7, 8 9, 10]). Expression (2.5) is a generalized and modified Whittaker representation for harmonic functions.

We call $f(u, \zeta)$ the $\mathbf{B}_3$-associate of H with respect to the curve L and the point $\boldsymbol{X}_0$. Similarly $\chi(Z, Z^*)$ is denoted as the $\mathbf{C}_3$-associate of H with respect to the curve L and the point $\boldsymbol{X}_0$.

There is a number of differences when we compare the operator $\mathscr{Re}[f(X+iY)]$ and the operator $\mathbf{C}_3(\chi, L; \boldsymbol{X}_0)$.

1. If we vary the point $\boldsymbol{X}_0$ then in general the resulting harmonic function will vary. Indeed, if the curve L is a closed curve the integral

$\frac{1}{2\pi i}\int_L f(u,\zeta)\frac{d\zeta}{\zeta}$ will yield the sum of residues and periods of the function $\zeta^{-1}f(x+Z\zeta+Z^*\zeta^{-1},\zeta)$, lying inside L. If $(\boldsymbol{x}_0, Z_0, Z_0^*)$ varies, some singularities of f (considered as a function of ζ for $\boldsymbol{X}\in N(\boldsymbol{X}_0)$) can slip in or out into the domain bounded by L and the periods can change so that $\mathbf{B}_3(f, L; \boldsymbol{X}_0)$ can have jumps if $\boldsymbol{X}_0$ (or L) varies. Consequently, the operator $\mathbf{B}_3$ may generate completely different harmonic functions for different values of $\boldsymbol{X}_0$.

2. In the case of the operator "$\mathscr{R}e$" (used for harmonic functions of two variables) the associate is again a function of two real variables X and Y, while in the case of three variables the associate $\chi(Z, Z^*)$ is a function of four real variables.

It is clear from the above remarks that one can study singularities and analytic continuation of harmonic functions by varying the path L in an admissible way. We wish to stress, however, that the singularity surfaces in the complex space may reduce to points when we consider x, y, z as real. For instance in

$$H(\boldsymbol{X}) = \frac{1}{(x-\alpha i)^2+y^2+z^2}, \qquad \alpha \text{ real},$$

the singularity line reduces to a point for $\alpha = 0$.

3. A Remark about Algebro-Logarithmic $\mathbf{B}_3$-Associates

Let us consider the application of our procedure to the study of harmonic functions H whose $\mathbf{B}_3$-associates are algebro-logarithmic functions. This class includes all harmonic functions H which are rational in x, y, z.

Indeed, if H is rational then, according to (2.4), the corresponding function χ is a rational function of $Z^{1/2}$ and $Z^{*1/2}$. When we substitute for Z and Z^* the expressions $u\zeta^{-1}T^2$ and $u\zeta(1-T)^2$, respectively, the function χ will be rational considered as a function of T. According to (2.6) we obtain $f(u,\zeta)$ by one integration and therefore the resulting function f must be algebro-logarithmic.

4. The Harmonic Functions with Algebraic $\mathbf{B}_3$-Associates

The simplest class of harmonic functions which can be investigated by using the method of integral operators is the class of harmonic functions with rational $\mathbf{B}_3$-associates. The theory of these functions is discussed in [1, 7, 8, 9, 10] and the papers mentioned there, and we shall not consider this case in the present paper.

The next class which can be investigated along similar lines is the class of functions H with algebraic $\mathbf{B}_3$-associates. In the following we shall show

that the classical theorems in the theory of integrals of algebraic functions of a complex variable can be interpreted not only as theorems on harmonic functions of two variables but also as results about harmonic functions of three variables.

Let

$$f(u, \zeta) = \frac{p_1(S, u, \zeta)}{p_2(u, \zeta)} \tag{4.1}$$

be a rational function of S, u and ζ where S, u, ζ are connected by an equation

$$a(S, u, \zeta) \equiv A_0(u, \zeta)S^n + A_1(u, \zeta)S^{n-1} + \cdots + A_n(u, \zeta) = 0. \tag{4.2}$$

$A_\nu(u, \zeta)$ are polynomials in u and ζ. If we replace in (4.1) u by $x + Z\zeta + Z^*\zeta^{-1}$ the resulting function

$$\frac{P_1(S, \zeta, \boldsymbol{X})}{P_2(\zeta, \boldsymbol{X})} = \frac{p_1(S, u, \zeta)}{p_2(u, \zeta)\zeta} \tag{4.3}$$

is, for every fixed X, an algebraic function of ζ. Substituting $x + \frac{iy+z}{2}\zeta + \frac{iy-z}{2}\zeta^{-1}$ for u into (4.2) and multiplying (4.2) by a conveniently chosen power of ζ we obtain the equation

$$\tilde{f}(\zeta, S; \boldsymbol{X}) = A_0(\zeta; \boldsymbol{X})S^n + A_1(\zeta; \boldsymbol{X})S^{n-1} + \cdots + A_n(\zeta; \boldsymbol{X}) = 0 \tag{4.4}$$

where the A_k are polynomials in ζ and x, y, z.

Let L be a *closed* curve in the ζ-plane.[1] We consider the class $P(S, L)$ of harmonic functions with $\{f(u, \zeta)\}$ as $\mathbf{B}_3$-associates, where $f(u, \zeta)$ are of the form (4.1) and S and u are connected by relation (4.2). We are interested in properties of this class when L varies.

According to Weierstrass ([18], p. 264; see also [1, 17]) every algebraic function $F(\zeta, S; \boldsymbol{X})$ defined on a Riemann surface $\mathfrak{R}(\boldsymbol{X})$ can be represented in the form

$$\begin{aligned} F(\zeta, S; \boldsymbol{X}) = &\sum_{\nu=1}^{n} C_\nu H(\zeta_\nu, S_\nu; \zeta, S) \\ &- \sum_{\alpha=1}^{\varrho} [g_\alpha^* H_\alpha(\zeta, S) - g_\alpha H_\alpha^*(\zeta, S)] + \frac{d}{d\zeta}\left[\sum_{\nu=1}^{r} F_\nu(\zeta, S)\right] \end{aligned} \tag{4.5}$$

where ϱ is the genus of the Riemann surface $\mathfrak{R}(\boldsymbol{X})$, n is the number of infinity points of F, r is a finite integer, $H_\alpha(\zeta, S)$, $H_\alpha^*(\zeta, S)$, $H(\zeta_\nu, S_\nu; \zeta, S)$ are Weierstrass integrands of the first, second and third kind, respectively, $F_\nu(\zeta, S)$ are conveniently chosen rational functions of ζ and S, (ζ_ν, S_ν), $\nu = 1, 2, \cdots, p$, are infinity points of $H(\zeta_\nu, S_\nu; \zeta, S)$ on the Riemann sur-

[1] Analogous results can be stated in the case of an open path L.

face $\mathfrak{R}(X)$ and $C_\nu(X)$, $g_\alpha^*(X)$, $g_\alpha(X)$, $\zeta_\nu(X)$, $S_\nu(X)$ are *algebraic* functions of x, y, z (because the expressions for these coefficients are computed from F by using only algebraic operations).

Let ϱ be the genus of the Riemann surface (4.4) and let $F(\zeta, S; X)$ be given by (4.3). S and ζ are connected by (4.4). We assume that if X varies in a neighborhood $N(X_0)$ of a fixed point X_0 then for $\zeta \in L$, ϱ is constant. The study of the period properties of the class $P(S, L)$ is, therefore, reduced to the study of the periods

$$\int_C H_\alpha(\zeta, S)d\zeta, \qquad \int_C H_\alpha^*(\zeta, S)d\zeta, \qquad \alpha = 1, 2, \cdots, \varrho,$$
$$\int_C H(\zeta_\nu, S_\nu; \zeta, S)d\zeta, \qquad \zeta_\nu \equiv \zeta_\nu(X).$$

C is a closed curve on the Riemann surface (4.4).

Each of the above integrals has at most 2ϱ different primitive values when for a fixed X, C varies in the ζ-plane.

We denote by[2]

$$2\omega_{\alpha\beta},\ 2\eta_{\alpha\beta},\ \Omega_\beta(\zeta_\nu, S_\nu), \qquad \alpha = 1, 2, \cdots, \varrho,$$
$$2\omega'_{\alpha\beta},\ 2\eta'_{\alpha\beta},\ \Omega'_\beta(\zeta_\nu, S_\nu), \qquad \beta = 1, 2, \cdots, 2\varrho,$$

the values of the above integrals taken over the ϱ cycles and ϱ conjugate cycles of the Riemann surface $\mathfrak{R}(X)$. Thus

$$\begin{aligned} H(X) &\equiv \frac{1}{2\pi i}\int_L F(\zeta, S)d\zeta \\ &= \sum_{\beta=1}^{\varrho}\Bigg[\sum_{\nu=1}^{n} C_\nu(X)\Omega_\beta(\zeta_\nu, S_\nu) + \sum_{\nu=1}^{n} C'_\nu(X)\Omega'_\beta(\zeta_\nu, S_\nu) \\ &\quad - \sum_{\alpha=1}^{\varrho}\left[g_\alpha^*(X)\omega_{\alpha\beta} - g_\alpha^{*\prime}(X)\omega'_{\alpha\beta} - g_\alpha(X)\eta_{\alpha\beta} - g'_\alpha(X)\eta'_{\alpha\beta}\right]\Bigg] \\ &\quad + \operatorname{Res}\sum_{\nu=1}^{r} F_\nu(\zeta, S) \end{aligned} \tag{4.6}$$

(see [18], p. 400).

Notice that C_ν, g_α^*, $g_\alpha^{*\prime}$, g_α, g'_α and $\sum_{\nu=1}^{r} F_\nu(\zeta, S)$ are *algebraic* functions of x, y, z.

According to Weierstrass (see [18], pp. 330—332) there exist the following relations among the various values of the periods of the integrals of the first and of the second kinds:

---

[2]Of course, all these quantities depend also on X. It should be noted that in this paper the primes do *not* represent derivatives.

$$\sum_{\alpha=1}^{\varrho}(\eta_{\alpha\beta}\omega_{\alpha\gamma}-\omega_{\alpha\beta}\eta_{\alpha\gamma})=0,\qquad \sum_{\gamma=1}^{\varrho}(\omega_{\alpha\gamma}\omega'_{\beta\gamma}-\omega_{\beta\gamma}\omega'_{\alpha\gamma})=0,$$

$$\sum_{\alpha=1}^{\varrho}(\eta'_{\alpha\beta}\omega'_{\alpha\gamma}-\omega'_{\alpha\beta}\eta'_{\alpha\gamma})=0,\qquad \sum_{\gamma=1}^{\varrho}(\eta_{\alpha\beta}\eta'_{\beta\gamma}-\eta_{\beta\gamma}\eta'_{\alpha\gamma})=0,$$

$$\sum_{\alpha=1}^{\varrho}(\eta_{\alpha\beta}\omega'_{\alpha\gamma}-\omega_{\alpha\beta}\eta'_{\alpha\gamma})=\begin{cases}0, & \beta\neq\gamma,\\ \dfrac{\pi i}{2}, & \beta=\gamma,\end{cases}\tag{4.7}$$

$$\sum_{\alpha=1}^{\varrho}(\eta_{\alpha\beta}\omega'_{\beta\gamma}-\omega_{\beta\gamma}\eta'_{\alpha\gamma})=\begin{cases}0, & \beta\neq\gamma,\\ \dfrac{\pi i}{2}, & \beta=\gamma.\end{cases}$$

We associate with the class $P(S, L)$ of our functions certain θ-functions defined as follows:

$$\theta(u_1, \cdots, u_\varrho ; \boldsymbol{X})$$

$$=\sum_{n_1=-\infty}^{\infty}\cdots\sum_{n_\varrho=-\infty}^{\infty}\exp\left\{\pi i\sum_{\alpha=1}^{\varrho}\sum_{\beta=1}^{\varrho}n_\alpha n_\beta\tau_{\alpha\beta}+\pi i\sum_{\alpha=1}^{\varrho}2n_\alpha v_\alpha+\sum_{\beta=1}^{\varrho}\sum_{\gamma=1}^{\varrho}2\varepsilon_{\beta\gamma}v_\beta v_\gamma\right\}\tag{4.8}$$

where

$$\tau_{\alpha\beta}=\left[\sum_{\gamma=1}^{\varrho}\frac{\partial\omega}{\partial\omega_{\gamma\alpha}}\omega'_{\gamma\beta}\right]\omega^{-1},\qquad \omega=|\omega_{\alpha\beta}|,$$

$$\varepsilon_{\beta\gamma}=\sum_{\alpha=1}^{\varrho}\omega_{\alpha\beta}\eta_{\alpha\gamma},\qquad v_\beta=\left[\sum_{\alpha=1}^{\varrho}\frac{\partial\omega}{\partial\omega_{\alpha\beta}}u_\alpha\right](2\omega)^{-1}\tag{4.9}$$

(see [18], pp. 531 ff. and 552 ff). These functions have the property that for $\boldsymbol{X}\in N(\boldsymbol{X}_0)$, where $N(\boldsymbol{X}_0)$ is a sufficiently small neighborhood of $\boldsymbol{X}_0$,

$$\theta(u_1+2\omega_{1\beta}, \cdots, u_\varrho+2\omega_{\varrho\beta}, \boldsymbol{X})=\theta(u_1, \cdots, u_\varrho, \boldsymbol{X})\exp\left\{\sum_{\alpha=1}^{\varrho}2\eta_{\alpha\beta}(u+\omega_{\alpha\beta}\right\},$$

$$\theta(u_1+2\omega'_{1\beta}, \cdots, u_\varrho+2\omega'_{\varrho\beta}, \boldsymbol{X})=\theta(u_1, \cdots, u_\varrho, \boldsymbol{X})\exp\left\{\sum_{\alpha=1}^{\varrho}2\eta'_{\alpha\beta}(u_\alpha+\omega'_{\alpha\beta}\right\}$$

hold.

We shall show that the functions $\eta_{\alpha\beta}$, $\eta'_{\alpha\beta}$, Ω_β and Ω'_β, $\alpha, \beta = 1, 2, \cdots, \varrho$, introduced in (4.6) can be expressed in terms of these θ-functions, their derivatives, the integrals $J_\nu(\zeta, S, \boldsymbol{X})$ of the first kind, $\omega_{\alpha\beta}$ and $\omega'_{\alpha\beta}$. By showing this we shall establish the following result:

If H is a function of the class $P(S, L)$ (see p. 330) *then, for $\boldsymbol{X}\in N(\boldsymbol{X}_0)$, H can be expressed in a closed form by a finite combination of integrals of the first kind $J_\nu(\zeta_\mu, S_\mu, \boldsymbol{X})$, the periods $\omega_\nu(\boldsymbol{X})$, $\omega'_\nu(\boldsymbol{X})$ of these integrals, θ-functions, their derivatives and algebro-logarithmic expressions. Here ζ_μ, S_μ are algebraic functions of $\boldsymbol{X}\equiv(x, y, z)$.*

In order to prove our statement we shall introduce the following notation (see [18], p. 516):

$$E_\beta(\zeta, S) = \exp\{\Omega_\beta(\zeta, S)\} \equiv \exp\left\{\int_c H(\zeta, S; \zeta', S')d\zeta'\right\},$$
$$E'_\beta(\zeta, S) = \exp\{\Omega'_\beta(\zeta, S)\} = \exp\left\{\int_{c'} H(\zeta, S; \zeta', S')d\zeta'\right\} \tag{4.11}$$

where c and c' are cycles and conjugate cycles on the Riemann surface (4.4) (for a fixed $\boldsymbol{X}$), respectively. The functions

$$2\pi i, \quad \log\frac{E'_\beta(\zeta_1, S_1)}{E'_\beta(\zeta_0, S_0)}, \quad \log\frac{E_\beta(\zeta_1, S_1)}{E_\beta(\zeta_0, S_0)}, \qquad \beta = 1, 2, \cdots, \varrho, \tag{4.12}$$

form a primitive system of periods of the integral

$$J(\zeta, S; \zeta_1, S_1, \zeta_0, S_0) = \int_{(a_0, b_0)}^{(\zeta, S)} \{H(\zeta_1, S_1; \zeta', S') - H(\zeta_0, S_0; \zeta', S')\}d\zeta' \tag{4.13}$$

where (ζ_0, S_0) and (ζ_1, S_1) are two arbitrary different points on the Riemann surface and ζ', S' are connected by the relation (4.4) (see [18], p. 399).

We note that

$$H(\zeta, S; \zeta', S') = \frac{F^*(\zeta, S; \zeta', S') - F^*(a_0, b_0; \zeta', S')}{\partial\tilde{f}/\partial S} \tag{4.14}$$

where

$$F^*(\zeta, S; \zeta', S') = \frac{f(\zeta', S)}{(\zeta - \zeta')(S - S')}$$

and ζ and S are connected by the equation (4.4).

Considered as a function of (ζ, S), F becomes infinite at the points (ζ', S'), $(a_1, b_1), \cdots, (a_\varrho, b_\varrho)$ on the Riemann surface (4.4) (see [18], pp. 61, 63). Therefore $H(\zeta, S; \zeta', S')$ is a rational function of ζ and S with a zero at (a_0, b_0) and poles at $(a_1, b_1), \cdots, (a_\varrho, b_\varrho)$.

With the above formulas and the notation

$$J_\alpha(\zeta, S) = \int_{(a_0, b_0)}^{(\zeta, S)} H_\alpha(\zeta', S')d\zeta', \tag{4.15}$$

$$J'_\alpha(\zeta, S) = \int_{(a_0, b_0)}^{(\zeta, S)} H'_\alpha(\zeta', S')d\zeta', \qquad \alpha = 1, 2, \cdots, \varrho, \tag{4.16}$$

there exists the following relation between the periods of the integrals of the first kind and those of the second kind (see [18], pp. 373, 486, 597, 599):

$$2\eta_{\alpha\beta} \equiv 2\eta_{\alpha\beta}(\boldsymbol{X}) = \frac{\theta_\alpha[J_1(\zeta_0, S_0) + 2\omega_{1\alpha}(\boldsymbol{X}) - \frac{1}{2}\sum_{\nu=1}^{2\varrho-2} J_1(p_\nu, q_\nu) - J_1(a_\alpha, b_\alpha), \cdots]}{\theta[J_1(\zeta_0, S_0) + 2\omega_{1\alpha}(\boldsymbol{X}) - \frac{1}{2}\sum_{\nu=1}^{2\varrho-2} J_1(p_\nu, q_\nu) - J_1(a_\alpha, b_\alpha), \cdots]}$$
$$- \frac{\theta_\alpha\left[J_1(\zeta_1, S_0) - \frac{1}{2}\sum_{\nu=1}^{2\varrho-2} J_1(p_\nu, q_\nu) - J_1(a_\alpha, b_\alpha), \cdots\right]}{\theta\left[J_1(\zeta_1, S_0) - \frac{1}{2}\sum_{\nu=1}^{2\varrho-2} J_1(p_\nu, q_\nu) - J_1(a_\alpha, b_\alpha), \cdots\right]}, \tag{4.17}$$

$$\theta_\alpha(u_1, \cdots, u_\varrho) = \frac{\partial\theta(u_1, \cdots, u_\varrho)}{\partial u_\alpha}, \qquad \alpha, \beta = 1, 2, \cdots, \varrho.$$

Here the (p_ν, q_ν) are zero points of the expressions

$$\frac{\sum_{\beta=1}^{\varrho} \overline{K}_\beta H_\beta(\zeta, S)}{\sum_{\beta=1}^{\varrho} \overline{K}'_\beta H_\beta(\zeta, S)}$$

where $\overline{K}_\beta$, $\overline{K}'_\beta$ are essentially arbitrary constants, and (ζ_0, S_0) is essentially an arbitrary point on the Riemann surface (4.4) (see [18], p. 593).

Following [18] pp. 306—307, we have

$$\begin{aligned}\int_{c_\beta} H(\zeta, S; \zeta', S')d\zeta' &\equiv \Omega_\beta(\zeta, S)\\ &= \sum_{\alpha=1}^{\varrho}\{2\omega_{\alpha\beta}J'_\alpha(\zeta, S) - 2\eta_{\alpha\beta}J_\alpha(\zeta, S) - 2\omega_{\alpha\beta}J'_\alpha(a_0, b_0)\\ &\quad + 2\eta_{\alpha\beta}J_\alpha(a_0, b_0)\}.\end{aligned} \tag{4.18}$$

The $\eta_{\alpha\beta}$ have been expressed in (4.17) in terms of periods of integrals of the first kind. In view of the relation

$$\begin{aligned}J'_\alpha(\zeta, S) - J'_\alpha(a_0, b_0)\\ = \frac{\theta_\alpha\left[J_1(\zeta, S) - \frac{1}{2}\sum_{\nu=1}^{2\varrho-2} J_1(p_\nu, q_\nu) - J_1(a_\alpha, b_\alpha), \cdots\right]}{\theta\left[J_1(\zeta, S) - \frac{1}{2}\sum_{\nu=1}^{2\varrho-2} J_1(p_\nu, q_\nu) - J_1(a_\alpha, b_\alpha), \cdots\right]}\\ - \frac{\theta_\alpha\left[J_1(a_0, b_0) - \frac{1}{2}\sum_{\nu=1}^{2\varrho-2} J_1(p_\nu, q_\nu) - J_1(a_\alpha, b_\alpha), \cdots\right]}{\theta\left[J_1(a_0, b_0) - \frac{1}{2}\sum_{\nu=1}^{2\varrho-2} J_1(p_\nu, q_\nu) - J_1(a_\alpha, b_\alpha), \cdots\right]}\end{aligned} \tag{4.19}$$

(see [18] p. 599), we have shown that $\Omega_\beta(\zeta_\nu, S_\nu)$ can be expressed in a closed form in terms of integrals and periods of the first kind.

5. A Class of Multivalued Harmonic Functions

Multivalued solutions of partial differential equations of two variables of elliptic type have been shown to be of considerable importance in physical applications, for instance in the theory of compressible fluids.

The theory of multivalued solutions in three variables also has various applications in physical problems. The first step in studying these solutions is to consider the harmonic equation. Using the above results and the theory of integral operators we shall obtain various results in this direction.

The elements H of the class $P(S, L)$ are, in general, multivalued har-

monic functions. In the previous sections we showed that they can be represented in terms of finitely many of them. We shall study the singularities of these functions and their developments in the neighborhood of their singular lines.

For the sake of brevity, we shall consider a simple case, when the equation (4.2) has the form

$$a(S, u, \zeta) \equiv S^2-[C_0u^2+(B_0+B_1\zeta+B_{-1}\zeta^{-1})u + (A_0+A_1\zeta+A_{-1}\zeta^{-1}+A_2\zeta^2+A_{-2}\zeta^{-2})] = 0; \tag{5.1}$$

C_0, B_0, B_1, B_{-1}, A_0, A_1, A_{-1}, A_2, A_{-2} are complex constants.[3]

By a convenient transformation (see e.g. [1], p. 553) our investigation can be reduced to the study of the periods of the integrals

$$\int \frac{\zeta_1^{\varkappa} d\zeta_1}{S_2}, \quad \varkappa = 0, 1, \qquad \int \frac{d\zeta_1}{(\zeta_1-\zeta_1^*(\boldsymbol{X}))S_2}, \tag{5.2}$$

where

$$S_2 = \sqrt{\zeta_1(1-\zeta_1)(k(\boldsymbol{X})-\zeta_1)} \tag{5.3}$$

and $\zeta_1^*(\boldsymbol{X})$ and $k(\boldsymbol{X})$ are algebraic functions of x, y, z.

The periods $\tilde{\omega} = \tilde{\omega}(\boldsymbol{X})$ of the first integral in (5.2) (i.e., for $\varkappa = 0$) satisfy the equation[4]

$$2k(k-1)\frac{d^2\tilde{\omega}^2}{dk^2}+2(2k-1)\frac{d\tilde{\omega}}{dk}+\frac{1}{2}\tilde{\omega} = 0 \tag{5.4}$$

while the periods of the third integral satisfy the equation

$$8k(k-1)(k-\zeta^*)\frac{d^3\tilde{\Omega}}{dk^3}+4[13k^2-(8\zeta^*+9)k+4\zeta^*]\frac{d^2\tilde{\Omega}}{dk^2} + 6(11k-9\zeta^*-4)\frac{d\tilde{\Omega}}{dk}+9\tilde{\Omega} = 0. \tag{5.5}$$

These equations are of the Fuchs type. The periods $\tilde{\omega}$ and $\tilde{\eta}$ of the first two integrals are singular on the lines

$$\mathscr{R}e\,[k(\boldsymbol{X})] = 0, \qquad \mathscr{I}m\,[k(\boldsymbol{X})] = 0, \tag{5.6}$$

$$\mathscr{R}e\,[k(\boldsymbol{X})] = 1, \qquad \mathscr{I}m\,[k(\boldsymbol{X})] = 0, \tag{5.7}$$

$$\mathscr{R}e\left[\frac{1}{k(\boldsymbol{X})}\right] = 0, \qquad \mathscr{I}m\left[\frac{1}{k(\boldsymbol{X})}\right] = 0 \tag{5.8}$$

while the third integral, in addition, becomes singular on

[3] Similar results hold also in the general case.

[4] $\tilde{\omega}$, $\tilde{\eta}$, and $\tilde{\Omega}$ are obtained from ω, η and Ω by a linear combination of the latter functions.

$$\mathscr{R}e[k(\boldsymbol{X})] = \mathscr{R}e[\zeta^*(\boldsymbol{X})], \quad \mathscr{I}m[k(\boldsymbol{X})] = \mathscr{I}m[(\zeta^*(\boldsymbol{X})]. \tag{5.9}$$

In the neighborhood of (5.6) two independent solutions $\tilde{\omega}_1(\boldsymbol{X})$ and $\tilde{\omega}_2(\boldsymbol{X})$ can be represented in the form

$$\begin{aligned} \tilde{\omega}_1(\boldsymbol{X}) &= -\pi v_{01}(k), \\ \tilde{\omega}_2(\boldsymbol{X}) &= v_{01}(k)\left\{\pi + i\log k - 4i\log 2 - i\int_0^k \frac{1+[v_{01}(k)]^2[k-1]}{[v_{01}(k)]^2 k[k-1]}\,dk\right\} \end{aligned} \tag{5.10}$$

where

$$v_{01}(k) = \left\{1+\sum_{b=1}^{\infty}\left(\frac{1\cdot 3\cdots(2b-1)}{2\cdot 4\cdots 2b}\right)^2 k^b\right\}, \quad k = k(\boldsymbol{X}).$$

One obtains similar representations for the development of functions in the neighborhood of the lines (5.6), (5.8), (5.9). For details see [1, 14].

6. The Generalization of the Method of Integral Operators to the Theory of Differential Equations and Systems of Differential Equations in Three and More Variables

The next step in the development of the approach based on the theory of integral operators consists in the generalization of our procedures for the study of harmonic vectors $\boldsymbol{q}$, satisfying the system

$$\operatorname{curl} \boldsymbol{q} = 0, \qquad \operatorname{div} \boldsymbol{q} = 0. \tag{6.1}$$

The vectors $\boldsymbol{q}$ can be considered as a generalization of analytic functions of a complex variable.

We obtain for $\boldsymbol{q}$ the representation

$$\begin{aligned} \boldsymbol{q} = \Big\{ &\mathbf{B}_3[f, L, \boldsymbol{X}_0], \mathbf{B}_3\left[\frac{i}{2}(\zeta+\zeta^{-1})f(u,\zeta), L, \boldsymbol{X}_0\right] + \mathscr{R}e[g(Z)], \\ &\mathbf{B}_3[\tfrac{1}{2}(\zeta-\zeta^{-1})f(u,\zeta), L, \boldsymbol{X}_0] - \mathscr{I}m[g(Z)]\Big\} \end{aligned} \tag{6.2}$$

in terms of an arbitrary function $f(u, \zeta)$ of complex variables u and ζ and in terms of an arbitrary function $g(Z)$ of the complex variable $Z = y+iz$ ([8], p. 488 ff.). See (2.5).

In the case of differential equations

$$T(\phi) \equiv \Delta_3\phi + C(r^2)\phi = 0 \tag{6.3}$$

where $C(r^2)$ is an entire function of $r^2 = x^2+y^2+z^2$ we can determine to every $C(r^2)$ a function $B(r^2, \sigma^2)$ defined for $0 \leqq r^2 < \infty$, $0 \leqq \sigma^2 \leqq 1$, such that the operator

$$\mathbf{p}_3[H(\boldsymbol{X})] \equiv H(\boldsymbol{X}) + \int_0^1 B^2(r^2, \sigma^2)H(\sigma^2\boldsymbol{X})\,d\sigma \tag{6.4}$$

yields solutions of (6.3). Here H is an arbitrary harmonic function of three variables which is regular at the origin (see [7], p. 500).

Combining (6.4) and (2.3) we obtain the representation

$$\phi(x, Z, Z^*) = \mathbf{p}_3[\mathbf{C}_3(\chi(Z, Z^*))] \tag{6.5}$$

of solutions of (6.3) in terms of analytic functions of two complex variables Z, Z^*.

The inverse operator, i.e., the operator yielding $\chi(Z, Z^*)$ in terms of ϕ,

$$\chi(Z, Z^*) = \phi(2(ZZ^*)^{1/2}, Z, Z^*) \tag{6.6}$$

is independent of the coefficient C of equation (6.3) (see [8], p. 502 and [7]).

Remark: The representation (6.4) can be generalized to the case of the differential equation

$$\Delta_3\phi + A(r^2)\left[x\frac{\partial\phi}{\partial x} + y\frac{\partial\phi}{\partial y} + z\frac{\partial\phi}{\partial z}\right] + C(r^2)\phi = 0 \tag{6.7}$$

where $A(r^2)$ and $C(r^2)$ are entire functions of r^2 (see [7], p. 424).

Operators yielding solutions of the differential equation

$$\frac{\partial^s\phi}{\partial x^s} + \frac{\partial^2\phi}{\partial y^2} + \frac{\partial^2\phi}{\partial z^2} + F(Z, \bar{Z})\phi = 0, \qquad s = 1 \text{ or } 2, \tag{6.8}$$

are discussed in [19].

Similar integral operators can be derived for certain systems of linear partial differential equations, e.g., in [11] functions $\psi = \psi(x_1, x_2, x_3, x_4)$ satisfying the system

$$\frac{\partial^2\psi}{\partial x_1^2} + \frac{\partial^2\psi}{\partial x_2^2} = F_1(x_1, x_2)\psi, \qquad \frac{\partial^2\psi}{\partial x_3^2} + \frac{\partial^2\psi}{\partial x_4^2} = F_2(x_3, x_4)\psi$$

are considered. Here $F_{\varkappa+1}$ ($\varkappa = 0$ or 1) is an entire function of two variables $x_{1+2\varkappa}$, $x_{2+2\varkappa}$. In [11] the following representation of ψ in terms of two functions of two complex variables f and g is derived:

$$\psi(x_1, x_2, x_3, x_4) = \mathscr{R}e\left\{\int_{t=-1}^{1}\int_{\tau=-1}^{1} E_1(z, \bar{z}, t)E_2(w, \bar{w}, \tau) \cdot [f(z(1-t^2), w(1-\tau^2)) + g(z(1-t^2), \bar{w}(1-\tau^2))]\frac{dt}{(1-t^2)^{1/2}}\frac{d\tau}{(1-\tau^2)^{1/2}}\right\}.$$

Here $z = x_1 + ix_2$, $w = x_3 + ix_4$, $\bar{z} = x_1 - ix_2$ and $\bar{w} = x_3 - ix_4$. E_1 and E_2 are generating functions which depend only on the equation, i.e., on F_1 and F_2, while f and g are two arbitrary analytic functions of *two* complex variables which are regular at the origin.

Bibliography

[1] Bergman, S., *Zur Theorie der algebraischen Potentialfunktionen des drei-dimensionalen Raumes*, Math. Ann., Vol. 99, 1928, pp. 629–659; Vol. 101, 1929, pp. 534–558.

[2] Bergman, S., *Zur Theorie der Funktionen, die eine lineare partielle Differentialgleichung befriedigen*, Mat. Sbornik, N. S., Vol. 2, 1937, pp. 1169–1198.

[3] Bergman, S., *Linear operators in the theory of partial differential equations*, Trans. Amer. Math. Soc., Vol. 53, 1943, pp. 130–155.

[4] Bergman, S., *Solutions of linear partial differential equations of the fourth order*, Duke Math. J., Vol. 11, 1944, pp. 617–649.

[5] Bergman, S., *Certain classes of analytic functions of two real variables and their properties*, Trans. Amer. Math. Soc., Vol. 57, 1945, pp. 299–331.

[6] Bergman, S., *A class of harmonic functions in three variables and their properties*, Trans. Amer. Math Soc , Vol. 59, 1946, pp. 419–458.

[7] Bergman, S., *Classes of solutions of linear partial differential equations in three variables*, Duke Math. J., Vol. 13, 1946, pp. 419–458.

[8] Bergman, S., *On solutions with algebraic character of linear partial differential equations*, Trans. Amer. Math. Soc., Vol , 69, 1950, pp. 461–507.

[9] Bergman, S., *The coefficient problem in the theory of linear partial differential equations*, Trans. Amer. Math. Soc., Vol. 73, 1952, pp. 1–34.

[10] Bergman, S., *Essential singularities of solutions of a class of linear partial differential equations in three variables*, J. Rat. Mech Anal , Vol. 3, 1954, pp 539–560.

[11] Bergman, S., and Schiffer, M., *Properties of solutions of a system of partial differential equations*, in *Studies in Mathematics and Mechanics Presented to Richard von Mises*, Academic Press, New York, 1954, pp. 79–87.

[12] Eichler, M. M. E., *On the differential equation* $u_{xx}+u_{yy}+N(x)u = 0$, Trans. Amer. Math. Soc., Vol. 65, 1949, pp. 259–278.

[13] Eichler, M. M. E , *Analytic functions in three-dimensional Riemannian spaces*, Duke Math. J., Vol. 16, 1949, pp. 339–349.

[14] Fuchs, L., *Die Periodicitaetsmoduln der hyperelliptischen Integrale als Funktionen eines Parameters aufgefasst*, Math. Werke, Vol. I, Berlin, 1904, pp. 241–281.

[15] Kreyszig, E., *On a class of partial differential equations*, J Rat Mech Anal., Vol. 4, 1955, pp. 907–923.

[16] Kreyszig, E , *On certain partial differential equations and their singularities*, J. Rat. Mech. Anal. (to appear).

[17] Weyl, H., *Die Idee der Riemannschen Flache*, B. G. Teubner, Leipzig-Berlin, 1923.

[18] Weierstrass, K., *Vorlesungen uber die Theorie der Abelschen Transcendenten*, Gesammelte Werke, Vol. 4, Mayer and Muller, Berlin, 1902.

[19] Bergman, S., *On singularities of solutions of certain differential equations in three variables* (to appear).

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 339–350 (1956)

Survey of Local Properties of Solutions of Elliptic Partial Differential Equations*

LIPMAN BERS

1. Introduction

The early stages of the theory of partial differential equations were dominated by the search for general solutions, that is, for explicit formulas involving arbitrary constants and functions which express all solutions of the equations considered. Later it was recognized that such general solutions can be found only in exceptional cases and that their knowledge does not necessarily lead to the particular solutions needed. From then on the primary aim was to find particular solutions satisfying given initial and boundary conditions. As far as linear elliptic equations are concerned this task is now nearing completion. Recently, however, the interest has shifted again to the study of properties of solutions, without regard to particular boundary conditions. This shift of interest has been brought about by the needs of the theory of nonlinear equations. In a certain sense, therefore, we are again interested in the general solution, but by a general solution we now understand not an explicit formula but rather the whole class of functions satisfying the given equation.

This talk is a report on one phase of the theory of classes of functions satisfying elliptic partial differential equations. We are concerned here with the local behavior of solutions, that is, with properties near a point.

In investigating such properties it is important to make as few hypotheses on the coefficients as possible. In fact, experience shows that the natural and therefore simplest proofs are usually found when no unnecessary hypotheses are made, even though at an early stage of an investigation it might be convenient to make additional assumptions. There is also a good, purely technical, reason for trying to develop a theory of linear elliptic equations under minimal assumptions on the coefficients. This reason was clearly pointed out by Schauder [43] many years ago. If one wants to study smooth nonlinear elliptic equations, even analytic nonlinear equations, it is convenient and sometimes indispensable to have precise information on solutions of linear equations of as general a nature as possible.

*The work for this paper was sponsored by the Office of Ordnance Research, United States Army, under Contract DA-30-069-ORD-835.

2. Smoothness of Solutions

A characteristic property of elliptic equations, not shared by equations of other types, is the fact that solutions are as smooth as the equation permits.

Consider the linear elliptic equation of order $2m \geqq 2$ for an unknown function of $n \geqq 2$ independent variables $\phi(x_1, \cdots, x_n) = \phi(x)$:

$$L\phi = \sum_{\nu=0}^{2m} \sum_{(\nu)} a_{i_1 \cdots i_n}(x) \frac{\partial^\nu \phi(x)}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}} = 0, \tag{2.1}$$

where

$$\sum_{(2m)} a_{i_1 \cdots i_n} t_1^{i_1} \cdots t_n^{i_n} > 0 \text{ for } t_1^2 + \cdots + t_n^2 > 0.$$

(Here and hereafter $\sum_{(\nu)}$ indicates summation over all non-negative $i_1, \cdots, i_n$ with $i_1 + \cdots + i_n = \nu$.) If the coefficients are real analytic, so are all solutions (John [35]). If the coefficients are Hölder continuous, all solutions have Hölder continuous derivatives up to order $2m$. This was proved by E. Hopf [34] for $m = 1$, but his proof is easily modified so as to cover the general case. If the coefficients have Hölder continuous derivatives up to the order k, then the solutions are $2m+k$ times Hölder continuously differentiable (Douglis and Nirenberg [20]).

By introducing new dependent and independent variables, an equation of the form (2.1) can be transformed, locally, into an equation of the same form not involving explicitly the unknown function or its first derivatives [10]. This permits us to weaken somewhat the hypotheses made above.

Similar results hold for elliptic systems, but in this talk I shall restrict myself to a single equation.

3. Zeros of Infinite Order and Unique Continuation

If the coefficients of the equation (2.1) are analytic, it is trivial to see that a solution which has a zero of infinite order, that is, which vanishes at a point x_0 faster than any power of $|x-x_0|$, vanishes identically. It seems highly likely that this statement, which implies that solutions of elliptic equations possess the very important unique continuation property, is true also for the most general non-analytic linear elliptic equations and systems. But positive results have been achieved only in special cases.

The answer is completely known for $m = 1, n = 2$. If the coefficients of the equation are, say, Hölder continuously differentiable, the unique continuation property follows from a well known theorem of Carleman [18]. (For other proofs of this theorem see [9] and Vekua [45], cf. also Gergen and Dressel [24, 25, 26].) But the unique continuation property also holds under more general conditions. Consider a second order elliptic equation

$$a_{11}\phi_{x_1x_1}+2a_{12}\phi_{x_1x_2}+a_{22}\phi_{x_2x_2}+a_1\phi_{x_1}+a_2\phi_{x_2}+a_0\phi=0 \tag{3.1}$$

with $a_{11}a_{22}-a_{12}^2\equiv 1$ and with measurable coefficients. We assume that the equation is uniformly elliptic, i.e., that the coefficients are uniformly bounded. By a solution we mean now a continuously differentiable function which possesses strong L_2 derivatives of the second order in the sense of Sobolev and Friedrichs. Nirenberg and the author showed [16, 17] that for such equations the unique continuation property holds (for equations involving only the highest derivatives this follows also from the work of Morrey [36]). The boundedness condition on the coefficients of the first derivatives and the unknown function may be weakened. It is, for instance, enough to assume that the coefficients belong to L_p for some $p>2$.

For the case $n>2$, $m=1$ there exists the following result (Müller [37], Heinz [32]): a solution of the equation

$$\Delta\phi+\sum_{i=1}^{n}a_i(x)\phi_{x_i}+a_0(x)\phi=0 \tag{3.2}$$

which vanishes at a point of infinite order vanishes identically.

The theorem on the non-existence of zeros of infinite order implies that the Cauchy problem for an elliptic equation cannot have more than one solution, but the converse implication has not been established. For elliptic systems of first order equations in $2n>2$ dependent and two independent variables the uniqueness of Cauchy's problem is known (Carleman [19], Douglis [20], Hartman and Wintner [28]) but the problem of zeros of infinite order is open.

4. Zeros of Finite Order

At a point x_0 at which the solution of the elliptic equation (2.1) with Hölder continuous coefficients vanishes, but not of infinite order, the asymptotic behavior of the solution may be completely described [10]. Without loss of generality set $x_0=0$. There exists a homogeneous polynomial $p_N(x)\not\equiv 0$ of degree N such that

$$\phi(x)=p_N(x)+O(|x|^{N+\varepsilon}) \tag{4.1}$$

where ε is a Hölder exponent of the leading coefficients, $0<\varepsilon<1$. This asymptotic formula may be formally differentiated $2m$ times. (This means that every derivative of ϕ of order $l\leqq 2m$ is "asymptotic" to the corresponding derivative of p_N.) It may be formally differentiated $2m+k$ times if the coefficients are k times Hölder continuously differentiable. The polynomial p_N satisfies the osculating equation

$$L_0\phi=\sum_{(2m)}a_{i_1\cdots i_n}(0)\frac{\partial^{2m}\phi}{\partial x_1^{i_1}\cdots\partial x_n^{i_n}}=0. \tag{4.2}$$

For the case $n = 2$, $m = 1$ stronger results are known. For instance, an asymptotic formula of the form (4.1), which may be formally differentiated once, holds if the leading coefficients are Hölder continuous and the others belong to L_p, $p > 2$ [12]. For uniformly elliptic equations with bounded measurable coefficients one may define the topological order of a zero and give some estimates [16, 17], but no asymptotic formula is in general true as was shown by an example of Hartman and Wintner [30].

(For earlier work on the asymptotic behavior of the solutions of second order elliptic equations in the plane, see Položii [39, 40], Bers [4, 5, 6, 9], Vekua [44], Hartman and Wintner [28, 30].)

There is no need to discuss in detail the asymptotic behavior of a solution of a nonlinear elliptic equation near a regular point. In fact, every solution of a quasi-linear elliptic equation is also a solution of a linear one, and the difference of two solutions of a general nonlinear elliptic equation may be written as a solution of some linear equation.

5. Removable Singularities

In this and the following sections we consider a (single-valued) solution $\phi(x)$ of the elliptic equation (2.1) defined in a deleted neighborhood of the origin.

If the coefficients of equation (2.1) are Hölder continuous and

$$\phi(x) = \begin{cases} o(|x|^{2m-n}), & n > 2m,\ n \text{ even or } n \leqq 2m,\ n \text{ odd}, \\ o(|x|^{2m-n} \log |x|), & n \leqq 2m,\ n \text{ even}, \end{cases} \tag{5.1}$$

then the singularity at the origin is removable, i.e., ϕ and its derivatives up to order $2m$ are continuous at the origin [10]. This is in a certain sense the best possible result.

For nonlinear equations the theory of singularities is as yet in an embryonic stage and only episodic results are known. One encounters here, as one would expect, a much greater variety of phenomena. Almost all results known refer to second order quasi-linear equations in the plane.

For the equation of minimal surfaces

$$(1+\phi_{x_2}^2)\phi_{x_1x_1} - 2\phi_{x_1}\phi_{x_2}\phi_{x_1x_2} + (1+\phi_{x_1}^2)\phi_{x_2x_2} = 0 \tag{5.2}$$

all isolated singularities of single-valued solutions are removable [4]. This property is shared by a wide class of quasi-linear equations of the form

$$A(\phi_{x_1}, \phi_{x_2})\phi_{x_1x_2} + 2B(\phi_{x_1}, \phi_{x_2})\phi_{x_1x_2} + C(\phi_{x_1}, \phi_{x_2})\phi_{x_2x_2} = 0. \tag{5.3}$$

An elliptic equation of this form can be always written as a conservation law

$$\frac{\partial}{\partial x_1} A_1(\phi_{x_1}, \phi_{x_2}) + \frac{\partial}{\partial x_2} A_2(\phi_{x_1}, \phi_{x_2}) = 0 \tag{5.4}$$

[7]. Finn [23], proving and extending a conjecture stated by the author, showed that if the functions A_1, A_2 are uniformly bounded for all values of their variables, then an isolated singularity of a single-valued solution of (5.4) is removable. This involves, of course, certain smoothness assumptions on the coefficients, but in the case of nonlinear equations one might, at the present stage of our knowledge, be as generous concerning smoothness as one wants.

There exist quasi-linear equations of the form (5.3) for which isolated singularities are removable even if one assumes only the single-valuedness of the derivatives rather than that of the solution. An example of such an equation is the potential equation of a subsonic gas flow

$$(5.5) \qquad (\varrho\phi_{x_1})_{x_1}+(\varrho\phi_{x_2})_{x_2}=0, \qquad \varrho=\left[1-\frac{\gamma-1}{2}(\phi_{x_1}^2+\phi_{x_2}^2)\right]^{1/(\gamma-1)}$$

where $\gamma > 1$. (For more details see [8].) On the other hand, there exist quasi-linear equations of the form (5.3) with single-valued solutions which remain bounded at a nonremovable isolated singular point. An example is given by the equation

$$(5.6) \qquad (1-\psi_{x_2}^2)\psi_{x_1x_1}+2\psi_{x_1}\psi_{x_2}\psi_{x_1x_2}+(1-\psi_{x_1}^2)\psi_{x_2x_2}=0,$$

which is "conjugate" to the equation of minimal surfaces. The function

$$\psi=\log\,[r+(1+r^2)^{1/2}], \quad r^2=x_1^2+x_2^2,$$

is a solution of (5.6) which is continuous and singular at the origin.

A general theory, which would bring out the underlying orderly pattern undoubtedly present here, is still a desideratum.

6. The Fundamental Singularity

For analytic elliptic equations (2.1) John [35] constructed the fundamental solution, that is a solution which has a singularity of lowest possible order. If the coefficients in (2.1) are constant and the coefficients of derivatives of order less than $2m$ vanish, this fundamental solution is of the form

$$(6.1) \qquad J_0(x)=q(x)\log|x|+\Omega(x)$$

where Ω is homogeneous of degree $2m-n$ and q is a homogeneous polynomial. The polynomial q vanishes identically if n is odd or if n is even and $2m < n$. For even n and $2m \geqq n$ the degree of q is $2m-n$. In the general case the fundamental solution of (2.1) is asymptotic to the solution J_0 of the osculating equation (4.2). Such a fundamental solution can also be constructed for nonanalytic equations with sufficiently smooth coefficients. For second order equations the theory of the fundamental solution is, of course, classical.

The following theorem holds for an elliptic equation (2.1) with Hölder continuous coefficients [10]. If

$$\phi(x) = O(|x|^{2m-n}) \tag{6.2}$$

then

$$\phi(x) = cJ_0(x)+o(|x|^{2m-n}) \tag{6.3}$$

where c is a constant, and this asymptotic relation may be formally differentiated $2m$ times. If ε, $0 < \varepsilon < 1$, is a Hölder exponent of the leading coefficients in (2.1), (6.2) may be replaced by the weaker condition $\phi(x) = O(|x|^{2m-n-\varepsilon+\delta})$, $\delta > 0$.

For equation (3.1) stronger results are known [11, 12]. Without loss of generality we assume that

$$a_{11} = a_{22},\ a_{12} = 0 \text{ at } x_1 = x_2 = 0. \tag{6.4}$$

If either $\phi > 0$, or $|\phi| \to +\infty$,

$$\phi(x) \sim c \log |x_1+ix_2| \tag{6.5}$$

and, assuming the coefficients to be Hölder continuous, this asymptotic relation may be differentiated twice. It may be differentiated once if the leading coefficients in (3.1) are Hölder continuous and the others are in L_p, $p > 2$.

7. Singularities of Higher Order

In the case of analytic elliptic equations (2.1) John [35] proved that if $\phi(x) = O(|x|^{-N})$ for some N, then the solution ϕ is asymptotic to a linear combination of the derivatives of a fundamental solution, up to order $N+2m-n$. It is likely that a similar result holds also for equations with Hölder continuous coefficients, but this has not been proved.

A complete theory of isolated singularities exists for the case $n = 2$, $m = 1$. If the coefficients of equation (3.1) are Hölder continuous, it follows from the theory of pseudo-analytic functions [5, 6, 11, 12, 13] that the following trichotomy holds: At the origin the complex gradient $\phi_{x_1}-i\phi_{x_2}$ has either an essential singularity, i.e. is not $O(|z|^{-N})$ for any N, or a removable singularity or a pole of integral order N. This means that, assuming (6.4), we have

$$\phi_{x_1}-i\phi_{x_2} \sim c(x_1+ix_2)^{-N} \tag{7.1}$$

where c is a non-vanishing complex constant and

$$\phi_{x_1x_1}-i\phi_{x_1x_2} \sim -\phi_{x_2x_2}-i\phi_{x_1x_2} \sim -Nc(x_1+ix_2)^{-N-1}. \tag{7.2}$$

Under the less restrictive assumption that only the leading coefficients in (3.1) are Hölder continuous and the others are in L_p, $p > 2$, the same

conclusion holds, except for the asymptotic formula (7.2). That the latter formula cannot be expected follows from the fact that without Hölder continuity of the coefficients one may not assert the continuity of the second derivatives of a solution.

For a uniformly elliptic equation (3.1) with bounded measurable coefficients it follows from the work of Nirenberg and the author [16, 17] that at an isolated non-removable singularity the complex gradient $\phi_{x_1} - i\phi_{x_2}$ either becomes infinite of finite order or has an essential singularity. In the latter case, if $a_0 \equiv 0$, the complex gradient comes arbitrarily close to every complex number.

8. Differentiators

If the coefficients of a linear elliptic equation are infinitely differentiable, so are all solutions. In some cases, however, solutions of elliptic equations with more general coefficients are still "infinitely differentiable" in some generalized sense.

This fact is observed easiest for equations which can be solved by separation of variables (Beltrami [2, 3], Bers and Gelbart [14, 15], Diaz [22], Protter [41]). The following example is typical of the equations treated in the references given above. Consider a linear elliptic equation of the form

$$\phi_{x_1 x_1} + \cdots + \phi_{x_n x_n} + a_1(x_1)\phi_{x_1} + \cdots + a_n(x_n)\phi_{x_n} = 0.$$

Introducing the variables

$$\sigma_j(x_j) = \int_0^{x_j} a_j(\xi)\,d\xi,$$
$$X_j(x_j) = \int_0^{x_j} e^{\sigma_j(\xi)}\,d\xi, \qquad j = 1, 2, \cdots, n,$$
$$Y_j(x_j) = \int_0^{x_j} e^{-\sigma_j(\xi)}\,d\xi,$$

we write this equation in the form

$$\phi_{X_1 Y_1} + \phi_{X_2 Y_2} + \cdots + \phi_{X_n Y_n} = 0.$$

It is formally clear that if ϕ is a solution, every function $\partial\phi/\partial X_j$ again satisfies a differential equation of the same form. For instance, $\psi = \phi_{X_1}$, satisfies the equation

$$\psi_{x_1 x_1} + \cdots + \psi_{x_n x_n} - a_1(x_1)\psi_{x_1} + a_2(x_2)\psi_{x_2} + \cdots + a_n(x_n)\psi_{x_n} = 0.$$

This procedure can be justified under mild restrictions on the coefficients.

It is natural to ask whether similar "differentiators" exist for equations which do not admit solutions by separation of variables. More precisely, is it possible to find for a given elliptic equation (2.1) differential operators of the form

$$M_k\phi = b_0^{(k)}(x)\phi + \sum_{j=1}^{n} b_j^{(k)}(x)\frac{\partial\phi}{\partial x_j}, \qquad k = 1, 2, \cdots, n,$$

such that whenever ϕ satisfies (2.1), the function $\psi = M_k\phi$ satisfies an equation of the same form as (2.1), with coefficients subject to the same smoothness requirements? If such differentiators can be found, we may talk about generalized infinite differentiability of solutions.

That differentiators may be found formally for second order equations in the plane follows from an observation by Markuševič [34]. A complete answer is contained in the theory of pseudoanalytic functions [5, 6, 9, 11, 12, 13]. This theory implies the existence of differentiators for every elliptic equation of the form (3.1) with Hölder continuous coefficients. These differentiators can be determined in every domain in which (3.1) has a positive solution.

Whether or not differentiators can be found for equations of higher order or for more independent variables is an interesting open question.

9 Asymptotic and Convergent Expansions

Closely connected with the problem on generalized infinite differentiability is the possibility of giving instead of an asymptotic formula an actual expansion for a solution of an elliptic equation near a regular point or a pole. Such expansions have been found for the cases in which differentiators exist.

Rather complete results have been obtained in the case $m = 1$, $n = 2$ [5, 6, 8, 9, 11, 13]. Consider an equation of the form (3.1) with Hölder continuous coefficients satisfying (6.4). There exist particular solutions of this equation, which are asymptotic, at the origin, to the harmonic functions $\log r$, $r^n \cos n\theta$, $r^n \sin n\theta$ (r, θ polar coordinates, $n = 0, \pm 1, \cdots$). Call these solutions $\Phi(x_1, x_2)$, $\Phi_n(x_1, x_2)$, $\tilde{\Phi}_n(x_1, x_2)$. Every solution of (3.1) which is regular near the origin admits the unique asymptotic expansion

$$\phi \sim a\Phi + \sum_{\nu=0}^{\infty} \alpha_\nu \Phi_\nu + \beta_\nu \tilde{\Phi}_\nu$$

with real constant coefficients a, α_ν, β_ν. This means that for $k = 1, 2, \cdots$

$$\phi(x_1, x_2) - a\Phi(x_1, x_2) - \sum_{\nu=0}^{k} \alpha_\nu \Phi_\nu(x_1, x_2) + \beta_\nu \tilde{\Phi}(x_1, x_2) = O(r^{k+1})$$

and this relation may be formally differentiated twice. Similarly, if ϕ is a single-valued solution and has at the origin a pole of order N, then

$$\phi \sim a\Phi + \sum_{\nu=-N}^{\infty} \alpha_\nu \Phi_\nu + \beta_\nu \tilde{\Phi}.$$

Under somewhat stronger regularity conditions on the coefficients

(Hölder continuous differentiability of the leading coefficients is sufficient) the particular solutions Φ_ν, $\tilde{\Phi}_\nu$ can be determined in such a way that the asymptotic expansions given above converge uniformly and absolutely on every compact subset of the largest disc around the origin in which ϕ is regular (Agmon and Bers [1]).

10. Branch Points

Until now we considered single-valued solutions. The theory of multiple-valued solutions has been developed only for second order equations in the plane and, as is well known, is very difficult even for the Laplace equation in more that two variables.

The theory of multiple-valued analytic functions of a complex variable is based on the concept of a Riemann surface: a multiple-valued analytic function may be considered as a single-valued analytic function defined on an appropriate covering surface. The same device works for linear elliptic equations of the form (3.1) since the theory of pseudoanalytic functions permits one to write every such equation in a conformally invariant form. One can, therefore, define branch points of finite order of a solution of equation (3.1) and describe precisely the behavior of a solution at such a point. There is no need to go into details; the description parallels exactly that of a harmonic function, in particular, we obtain asymptotic formulas and asymptotic and convergent expansions analogous to those given for single-valued solutions in §§4, 7 and 9 [6, 13].

Quite different phenomena are encountered in the theory of nonlinear equations. Up to now only the special case of the equation of minimal surfaces has been investigated. We consider a solution ϕ of (5.2) which is k-valued in the neighborhood of the origin, or, more generally, which has a k-valued gradient. If the gradient $\phi_{x_1}-i\phi_{x_2}$ is bounded, it attains a finite limit at the origin and

$$\phi(x, y) = \mathscr{R}e\{A+B(x_1+ix_2)^{1+p/k}\} + O(r^{1+(p+1)/k}) \tag{10.1}$$

where $p \geqq 0$ is an integer. If the gradient is unbounded, it becomes infinite at the origin and the solution behaves, in general, quite unlike a harmonic function. This behavior can, however, be described very precisely ([4], in particular see the figure on page 365).

Branch points of other quasi-linear elliptic equations have not yet been studied. Certainly, a great variety of new interesting types of branch points will be uncovered by a thorough investigation.

11. Concluding Remarks

We have given a catalogue of known results and open questions. Limitations of time do not permit us to discuss the methods used or the applications of local results to problems in the large (in particular to boundary value problems), to differential geometry, and to gas dynamics.

We are still far from a general theory of local behavior and it is likely that new methods will be needed to create such a theory. But I believe that the results achieved to date justify a continued interest in this field.

Note added on January 25, 1955: A paper by Hartman and Wintner [31] contains a new proof of the Müller-Heinz result (§3.). Another (unpublished) proof is due to Lax. The uniqueness of certain Cauchy problems has been established by Nirenberg (to appear) for a class of differential equations with constant leading coefficients. Concerning local behavior of solutions of second order elliptic equations with continuous but not Hölder continuous coefficients, see a forthcoming paper by Gilbarg and Serrin and the paper by L. Nirenberg in this issue.

Note added in proof, May 15, 1956: The unique continuation theorem (non-existence of zeros of infinite order) has been recently proved by Aronszajn (to appear) for second order elliptic equations with variable but sufficiently smooth (class C^4) leading coefficients. A strengthened version of the results reported in § 9 will be found in a forthcoming paper by the author.

Bibliography

[1] Agmon, S. and Bers, L., *The expansion theorem for pseudo-analytic functions*, Proc. Amer. Math. Soc., Vol 3, 1952, pp. 757–764.

[2] Beltrami, E., *Sulle funzioni potenziali di sistemi simmetrici intorno an un asse*, Opere matematice, Vol. 3, Milano, 1911, pp. 115–128.

[3] Beltrami, E., *Sulle teoria delle funzioni potenziali simmetriche*, ibid., pp. 349–377.

[4] Bers, L., *Isolated singularities of minimal surfaces*, Ann. Math , Vol. 53, 1951, pp. 364–386.

[5] Bers, L., *Partial differential equations and generalized analytic functions*, Proc. Nat. Acad. Sci., U.S.A., Vol. 36, 1950, pp 130–136 *Second note*, ibid , Vol. 37, 1951, pp. 42–47.

[6] Bers, L., *Theory of Pseudoanalytic Functions*, Lecture Notes, New York University, 1953.

[7] Bers, L., *Univalent solutions of linear elliptic systems*, Comm. Pure Appl. Math., Vol. 6 1953, pp. 513–526.

[8] Bers, L., *Nonlinear elliptic equations without nonlinear entire solutions*, J. Rat. Mech., Vol. 3, 1954, pp. 767–787.

[9] Bers, L., *Function-theoretical properties of solutions of partial differential equations of elliptic type*, Ann. Math. Study No. 33, Princeton University Press, 1954, pp. 69–94.

[10] Bers, L., *Local behavior of solutions of general linear elliptic equations*, Comm. Pure Appl. Math., Vol. 8, 1955, pp. 473–496.

[11] Bers, L., *Local theory of pseudoanalytic functions*, Lectures on Functions of a Complex Variable, Michigan University Press, 1955, pp. 213–244.

[12] Bers, L., *Remark on an application of pseudoanalytic functions*, to appear.

[13] Bers, L., *An outline of the theory of pseudoanalytic functions*, to appear.
[14] Bers, L., and Gelbart, A., *On a class of differential equations in mechanics of continua*, Quart. Appl. Math., Vol. 1, 1943, pp. 168–188.
[15] Bers, L , and Gelbart, A., *On a class of functions defined by partial differential equations*, Trans. Amer. Math. Soc., Vol. 56, 1944, pp. 67–93
[16] Bers, L. and Nirenberg, L., *On a representation theorem for linear elliptic systems with discontinuous coefficients and its applications*, Convegno Internazionale sulle Equazioni Derivate e Parziali, Agosto, 1954, pp. 111–140.
[17] Bers, L., and Nirenberg, L., *On linear and nonlinear elliptic boundary value problems in the plane*, Convegno Internazionale sulle Equazioni Derivate e Parziali, Agosto, 1954, pp. 141–167
[18] Carleman, T., *Sur les systèmes linéaires aux dérivées partielles du premier ordre à deux variables*, C. R Acad Sci. Paris, Vol. 197, 1933, pp. 471–474.
[19] Carleman, T., *Sur un problème d'unicité pour les systèmes d'équations aux dérivées partielles à deux variables indépendentes*, Ark. Mat. Astr Fys., Vol. 26B, No. 17, 1939, pp. 1–9.
[20] Douglis, A., *Uniqueness in Cauchy problems for elliptic systems of equations*, Comm Pure Appl. Math., Vol. 6, 1953, pp. 291–298
[21] Douglis, A. and Nirenberg, L., *Interior estimates for elliptic systems of partial differential equations*, Comm Pure Appl Math , Vol. 8, 1955, pp 503–538.
[22] Diaz, J. B., *On a class of partial differential equations of even order*, Amer. J. Math , Vol. 68, 1946, pp. 611-659
[23] Finn, R S , *Isolated singularities of solutions of nonlinear partial differential equations*, Trans. Amer. Math. Soc , Vol 75, 1953, pp 385–404.
[24] Gergen, J. J. and Dressel, F. G , *Mapping by p-regular functions*, Duke Math. J., Vol 18, No 1, 1951, pp. 185–210.
[25] Gergen, J. J. and Dressel, F. G , *Mapping for elliptic equations*, Trans. Amer. Math Soc., Vol. 77, 1954, pp 151-178.
[26] Gergen, J. J. and Dressel, F. G , *Uniqueness for p-regular mapping*, Duke Math. J , Vol., 19, No. 3, 1952, pp 435–444.
[27] Hartman, P. and Wintner, A., *On the third fundamental form of a surface*, Amer. J Math. Vol. 75, 1953, pp. 298–334.
[28] Hartman, P and Wintner, A , *On the local behavior of solutions of non-parabolic partial differential equations*, ibid , 1953, pp. 449–476.
[29] Hartman, P. and Wintner, A., *Umbilical points and W-surfaces*, Amer. J. Math , Vol 76, 1954, pp. 502-508.
[30] Hartman, P and Wintner, A , *On the local behavior of solutions of non-parabolic partial differential equations*, II· *The uniqueness of the Green singularity*, Amer J. Math , Vol 76, 1954, pp. 351–361.
[31] Hartman, P and Wintner, A., *On the local behavior of solutions of non-parabolic partial differential equations*, III. *Approximations by spherical harmonics*, Amer. J Math., Vol. 77, 1955, pp. 453–483.
[32] Heinz, E., *Über die Eindeutigkeit beim Cauchyschen Anfangswertproblem einer elliptischen Differentialgleichung zweiter Ordnung*, Nachr. Akad. Wiss. Gottingen. Math -Phys. Kl. IIa, No 1, 1955, pp. 1–12.
[33] Hopf, E., *Elementare Bemerkungen über Lösungen partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus*, Preuss Akad. Wiss. Sitzungsber., 19, 1927, pp. 147–152.
[34] Hopf, E., *Über den funktionalen, insbesondere den analytischen Character der Losungen elliptischer Differentialgleichungen zweiter Ordnung*, Math. Z., Vol. 34, 1931, pp. 191–233.
[35] John, F., *The fundamental solution of linear elliptic differential equations with analytic coefficients*, Comm. Pure Appl. Math., Vol. 3, 1950, pp 273-304.
[36] Morrey, C. B., *On the solutions of quasilinear elliptic partial differential equations*, Trans. Amer. Math. Soc., Vol. 43, 1938, pp. 126–166.
[37] Müller, C., *On the behavior of solutions of the differential equation* $\Delta U = F(x, U)$ *in the neighborhood of a point*, Comm. Pure Appl. Math., Vol. 7, 1954, pp. 505–515.

[38] Petrovsky, I. G., *On some problems in the theory of partial differential equations*, Uspekhi Matem Nauk, Vol. 1, 1946, pp. 44–70.

[39] Položii, G. N., *On p-analytic functions of a complex variable*, Doklady Akad. Nauk SSSR, Vol 58, 1947, pp. 1275-1278.

[40] Položii, G N., *Singular points and residues of p-analytic functions of a complex variable*, Doklady Akad Nauk SSSR, Vol. 60, 1949, pp. 769–772

[41] Protter, M H., *Generalized spherical harmonics*, Trans. Amer Math Soc., Vol. 63, 1948, pp. 314 341

[42] Protter, M H., *The periodicity problem in the theory of pseudoanalytic functions*, to appear.

[43] Schauder, J., *Équations du type elliptique, problèmes linéaires*, Enseignement Math., Vol 35, 1935, pp 126–139

[44] Vekua, I. N., *Systems of partial differential equations of first order of elliptic type and boundary value problems with applications to the theory of shells*, Mat. Sbornik, Vol. 31, 73, 1952, pp. 217–314.

[45] Vekua, I. N., *On a property of solutions of the system of generalized Cauchy-Riemann equations*, Bull. Acad. Sci. Georgian S S. R., Vol. 14, 1953, pp 449–453.

Received January 20, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 351–361 (1956)

On the Regularity Properties of Solutions of Elliptic Differential Equations

FELIX E. BROWDER
Brandeis University

During the past several years, there has been a great upsurge of interest in boundary value problems for elliptic partial differential equations and especially in the variational or Hilbert space approach [2—7, 11, 22]. To study the difficulties which arise in applying this technique, let us consider the single linear elliptic equation $Au = f$ under a set of homogeneous boundary conditions in a domain G of E^n (or of an n-manifold). We may consider a quadratic minimum problem or, more generally, a linear functional equation in a suitable Hilbert space H which corresponds to the original boundary value problem in two senses:

1) Every solution of the boundary value problem is also a solution of the functional equation.

2) Every solution of the functional equation which is sufficiently regular in the interior of G as well as on the boundary Γ of G is also a solution of the boundary value problem. Failing this regularity, these generalized solutions may be shown to satisfy the original differential conditions only in various weak or almost everywhere senses.

A simple example, treated in [9], is the Neumann problem for the Poisson equation $\Delta u = f$ with $\partial u/\partial n = 0$ on Γ. In this case, the Hilbert space H is the space of all functions in $L^2(G)$ with square-summable first derivatives (differentiation taken in the sense of Schwartz's theory of distributions [20]), the norm on H is the Dirichlet integral, and the functional equation which corresponds to the boundary value problem is the requirement that for all v in H

$$\sum_i \left(\frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i}\right) = -(f, v).$$

In order that the translation into a functional equation in the Hilbert space H should give us unambiguous information about boundary value problems in the usual sense, one must be able to prove that the solutions of the functional equation, the "generalized solutions" of our original problem, have all the regularity properties needed to reverse the translation procedure. These regularity properties fall into two classes, regularity in the interior of G and regularity at its boundary Γ. Regularity properties in the

interior may be derived by a number of distinct methods under various assumptions on the regularity of the coefficients of A: the fundamental solution method using the fundamental solutions constructed by Fritz John [2, 4, 13, 20], smoothing functions or Friedrichs' mollifiers [7, 10], function-space methods, estimation of difference quotients, and, for second-order equations, the Dirichlet growth estimates of Morrey [15, 16, 17, 19]. Except for the case $n = 2$ in the Dirichlet problem [4] and Morrey's results on the Dirichlet problem for second-order equations [16], the regularity properties on the boundary have been until very recently terra incognita. During the past year, the author has found a general method for deriving such properties for solutions of elliptic boundary value problems from smoothness conditions on the data of the problem. Independently, L. Nirenberg has obtained closely related results using a similar but slightly different method. It may be noted that the author's method depends upon a lemma on coercive forms due to N. Aronszajn [1] and yields sharper results than Nirenberg's method. This sharpness is reflected in the estimates in Section 4 which, in passing from L^2 estimates of the inhomogeneous term f to L^2 estimates of the $2m$-th derivatives of u (where $2m$ is the order of A), give in a certain sense a best possible result in the L^2 frame of reference.

We shall not attempt in the brief space of this note to do more than present the detailed statement of the results obtained, as well as suggestive remarks about the method of proof. The detailed presentation of the proofs will appear in another place. Beyond remarking at this point that the methods involved can be applied with purely formal alterations to systems of equations (elliptic and strongly elliptic for Sections 3 and 4, respectively), we shall not explicitly consider generalizations beyond the case of a single equation.

1. Notation and Remarks

Let G be a bounded domain in E^n with boundary Γ of class C^m, $m \geqq 1$ If α is the multi-index $(\alpha_1, \cdots, \alpha_r)$, $1 \leqq \alpha_i \leqq n$, D^α will denote the differential operator $\partial^r/\partial x_{\alpha_1} \cdots \partial x_{\alpha_r}$ while $|\alpha| = r$. We shall consider the following spaces of complex valued functions on G:

$L^p(G) = \{f|\ |f|^p$ summable with respect to Lebesgue n-measure on $G\}$, $p \geqq 1$,

$$||f||^p_{L^p(G)} = \int_G |f|^p\, dx,$$

$$(f, g) = \int_G f g^*\, dx, \qquad f, g \in L^2(G),$$

$$C^j(G) = \{f|\ D^\alpha f \text{ continuous in } G \text{ for } |\alpha| \leqq j\},$$

$$C^\infty(G) = \bigcap_j C^j(G),$$

$C^j(\bar{G}) = \bigcup_{G_1 \supset G} C^j(G_1)$,

$C_c^\infty(G) = \{f \mid f \in C^\infty(G),\ f = 0$ outside a compact subset of $G\}$,

and for $0 < \lambda < 1$,

$C^{j,\lambda}(G) = \{f \mid f \in C^j(G)$; there exists a $c_{\alpha,\lambda}(f)$ for each α with $|\alpha| = j$ such that $|D^\alpha f(x_1) - D^\alpha f(x)| \leqq c_{\alpha,\lambda}(f)\ |x_1 - x|^\lambda$ for every pair $x, x_1 \in G\}$,

$c^0_{\alpha,\lambda}(f) =$ the least such $c_{\alpha,\lambda}(f)$,

$$||f||_{C^{j,\lambda}(G)} = \sum_{|\alpha| \leqq j} \max_{x \in G} |D^\alpha f(x)| + \sum_{|\alpha| = j} c^0_{\alpha,\lambda}(f).$$

For $f \in C^j(\bar{G})$ we define $||f||_{H^{j,p}(G)} = \sum_{|\alpha| \leqq j} ||D^\alpha f||_{L^p(G)}$. Let $H^{j,p}(G)$ be the Banach space which is the completion of $C^j(\bar{G})$ with respect to this norm. It may be easily shown that $H^{j,p}(G)$ can be identified by continuity with a linear subset of $L^p(G)$. This is a special case of the following more general facts which may be deduced from a theorem of Sobolev [21].

(a) If $1/r \geqq 1/p - d/n > 0$, $p, d \geqq 1$, there exist a constant $c_{r,p,d}(G)$ such that for every non-negative integer j, $||f||_{H^{j,r}} \leqq c_{r,p,d}||f||_{H^{j+d,p}}$ for every f in $H^{j+d,p}(G)$. (Here, as in other places below, we have omitted the dependence of the norms on the domain G if no ambiguity arises.)

(b) If $1/p - d/n = 0$, then (a) is true for every $r \geqq 1$ with constants depending on r.

(c) If $1/p - d/n < 0$, there exists a $\xi(d, p, n)$, $0 < \xi(d, p, n) < 1$, such that for a suitable constant $c_{\xi,p,d}(G)||f||_{C^{j,\xi}} \leqq c_{\xi,p,d}(G)||f||_{H^{j+d,p}}$ for $j \geqq 0$, $f \in C^{j+d}(G)$.

If we designate by $T_{j,d,p,r}$ the identification mapping of $C^{j+d}(\bar{G})$ as a dense subset of $H^{j+d,p}(G)$ with $C^{j+d}(\bar{G})$ as a subset of $H^{j,r}(G)$, then (a) states that, for r satisfying the appropriate inequality, $T_{j,d,p,r}$ is a bounded linear transformation and hence can be extended by continuity to the whole of $H^{j+d,p}(G)$.

(d) If $1/r > 1/p - d/n$, $T_{j,d,p,r}$ is a compact linear transformation.

(e) If $1/r > 1/p - d/n$, then given an $\varepsilon > 0$, there exists a constant $c_{\varepsilon,r,p,d}(G)$ such that for all f in $C^{j+d}(\bar{G})$, $||f||_{H^{j,r}} \leqq \varepsilon||f||_{H^{j+d,p}} + c_\varepsilon||f||_{L^1(G)}$.

(f) If $j + \lambda < j_1 + \lambda_1$, then given an $\varepsilon > 0$ there exists a constant $c_{\varepsilon,j,j_1,\lambda,\lambda_1}$ such that for all f in $C^{j_1+\lambda_1}(G)$, $||f||_{C^{j,\lambda}} \leqq \varepsilon||f||_{C^{j_1,\lambda_1}} + c_\varepsilon||f||_{L^1(G)}$.

The most significant of the results (a)—(f) is (c); for, it is through this relation between integral norms of higher derivatives and Hölder norms of lower derivatives that we can derive the differentiability properties of the solutions of the functional equations which have replaced our differential equation with its boundary conditions. If we can show that our generalized solution u in H is the L^1 limit of a sequence $\{f_k\}$ of smooth functions whose

$H^{j,p}$ norms are uniformly bounded by a positive constant K, then by the reflexivity of the Banach spaces $H^{j,p}$ for $p > 1$, we may extract a subsequence $\{f_{k_i}\}$ converging weakly in $H^{j,p}$ to an element v whose $H^{j,p}$ norm is also bounded by K. But weak convergence in $H^{j,p}$ for $j \geqq 1$ may be shown to imply strong convergence in L^p and *a fortiori* in L^1 on the bounded domain G. Since limits in L^1 are unique, $u = v$ almost everywhere in G. Thus, under the assumptions we have made, after an inessential change, u would lie in $H^{j,p}$. But for j and p sufficiently large, the same sort of argument by (c) would identify u with an element of $C^{r,\lambda}$ and a differentiability result in the strict sense would follow.

Let us remark that for ease of formulation at one point below, we have used the $H^{-m,2}(G)$ norm which for a function f in $L^2(G)$ is defined to be $\sup_{g \in C^{m,2}(\overline{G})} |(f, g)| \, ||g||^{-1}_{H^{m,2}(G)}$.

In order to introduce an $H^{r,2}$ norm with respect to tangential derivatives on the boundary Γ of G, assumed compact and of class C^r, we choose a finite covering of Γ by coordinate neighborhoods $\{N_k', 1 \leqq k \leqq M\}$ with coordinate systems $\{y_i^{(k)}, 1 \leqq i \leqq n-1\}$ of class C^r. If α is the multi-index $(\alpha_1, \cdots, \alpha_s)$, $1 \leqq \alpha_i \leqq n-1$, the corresponding tangential differential operator $D^{\alpha}_{(t),k}$ acting on functions in N_k is defined to be $\partial^s / \partial y_{\alpha_1} \ldots \partial y_{\alpha_s}$. We may assume that there exists a covering of Γ by neighborhoods N_k such that for each k, $\overline{N}_k \subset N_k'$.

For $u \in C^r(\Gamma)$ we set

$$||u||^2_{H^{r,2}(\Gamma)} = \sum_{k=1}^{M} \sum_{|\alpha| \leqq r} \int_{N_k} |D^{\alpha}_{(t),k} u|^2 \, dS.$$

Completing $C^r(\Gamma)$ with respect to this norm we obtain a Hilbert space $H^{r,2}(\Gamma)$ which may be identified by continuity with a subset of $H^{0,2}(\Gamma)$. Complete analogues of properties (a)—(f) above may be obtained as before. In addition we have:

(g) Let R_r be the mapping from $C^r(\overline{G})$ into $C^r(\Gamma)$ obtained by assigning to each function its boundary value function. Then R is a completely continuous linear transformation from $H^{r,2}(G)$ into $H^{r-1,2}(\Gamma)$.

2. Elliptic Boundary Value Problems

Let $A = \sum_{|\zeta| \leqq 2m} c_\zeta(x) D^\zeta$ be a linear differential operator of order $2m$ on G. A is said to be elliptic at the point x in G if the characteristic form

$$c(x, \xi) = \sum_{|\zeta| = 2m} c_\zeta(x) \xi^{[\zeta]} \neq 0$$

for every real n-vector $\xi \neq 0$ ($\xi^{[\zeta]} = \xi_{\zeta_1} \cdots \xi_{\zeta_{2m}}$).

A Dirichlet form

$$a(f, g) = \sum_{|\alpha|,\ |\beta| \leqq m} (a_{\alpha\beta}(x) D^\alpha f, D^\beta g)$$

is said to be associated with the differential operator A if

$$Au = \sum_{|\alpha|,\ |\beta| \leqq m} (-1)^{m+|\beta|} D^\beta (a_{\alpha\beta} D^\alpha u).$$

We shall assume for such forms that their coefficients are bounded measurable in G.

The definition of an elliptic boundary value problem (which we might call a variational boundary value problem) involves:

(a) *A linear family V of functions on G, $C_c^\infty(G) \subset G \subset C^r(\bar{G})$* ($\Gamma$ is assumed to be of class C^r).

(b) *A symmetric boundary form*

$$\zeta(f, g) = \int_\Gamma \sum_{|\alpha|,\ |\beta| \leqq r} g_{\alpha\beta}(x) D^\alpha D^\beta g^* dS.$$

If $\zeta \not\equiv 0$, we must assume that there exists a set of integers $0 \leqq i_1 < i_2 < \cdots < i_j \leqq m-1$, $0 \leqq s_h$, $h = 1, \cdots, j$, as well as a constant $\psi > 0$ such that (if n is the exterior normal to Γ)

$$\psi^{-1} \left\{ \sum_{h=1}^{j} \left\| \frac{\partial^{i_h} f}{\partial n^{i_h}} \right\|^2_{H^{s_h, 2}(\Gamma)} \right\} \leqq \zeta(f, f) \leqq \psi \left\{ \sum_{h=1}^{j} \left\| \frac{\partial^{i_h} f}{\partial n^{i_h}} \right\|^2_{H^{s_h, 2}(\Gamma)} \right\}.$$

(c) *A second boundary form*

$$\phi = \phi_1 + \phi_2.$$

Here

$$\phi_1(f, g) = \int \sum_{|\alpha|,\ |\beta| \leqq m-1} j_{\alpha\beta}(x) D^\alpha f D^\beta g^* dS;$$

$\phi_2 \equiv 0$ if $\zeta \equiv 0$, while if $\zeta \not\equiv 0$ there must exist $\psi_1 \geqq 0$ such that for all $f, g \in V$

$$|\phi_2(f, g)|^2 \leqq \psi_1 \left\{ \sum_{h=1}^{j} \left\| \frac{\partial^{i_h} f}{\partial n^{i_h}} \right\|_{H^{s_h - 1, 2}(\Gamma)} \right\} \left\{ \sum_{h=1}^{j} \left\| \frac{\partial^{i_h} g}{\partial n^{i_h}} \right\|_{H^{s_h, 2}(\Gamma)} \right\}.$$

(In the last inequality, it is necessary to assume that for each h, $n_h + i_h \leqq r$, the differentiability class of the boundary.)

(d) *A Dirichlet form $a(f, g)$* associated with A for which for some $\psi_3 > 0$, $\psi_4 \geqq 0$ and all $f \in V$,

$$a(f, f) \geqq \psi_3 \|f\|^2_{H^{m, 2}(G)} - \psi_4 \|f\|_{H^{m, 2}(G)} \|f\|_{H^{m-1, 2}(G)}$$

(coerciveness of a on V).

From (d) combined with (b) it follows that there exists a constant k_0 such that the Hermitian symmetric form

$$[f, g] = \frac{a(f, g) + a(g, f)^*}{2} + \zeta(f, g) + k_0(f, g)$$

satisfies the inequality

$$[f, f] \geqq \frac{\psi_3}{2}||f||_{H^{m,2}(G)} \qquad \text{for } f \in V.$$

Let H be the completion of V with respect to the inner product $[\ ,\ ]$. Then H may be identified with a subset of $H^{m,2}(G)$. In fact, since every Cauchy sequence in the H norm from V is also a Cauchy sequence in $H^{m,2}(G)$, there exists a continuous linear transformation T from H into $H^{m,2}(G)$ which extends the identity map of V to itself. To show that T is one-to-one, it suffices to prove that $Tu = 0$ implies $u = 0$. But every element u of H is uniquely represented by the $(j+1)$-tuple $(Tu, v_1, \cdots, v_j)$ where v_h is the limit in $H^{s_h,2}(\Gamma)$ of the sequence $\{\partial^{t_h} f_k/\partial n^{t_h}\}$ for any Cauchy sequence $\{f_k\}$ from V converging to u in H. More precisely, there exists a $C \geqq 0$ such that

$$||u||_H^2 \leqq C\left\{||Tu||_{H^{m,2}(G)}^2 + \sum_{h=1}^{j}||v_h||_{H^{s_h,2}(\Gamma)}^2\right\}.$$

If $Tu = 0$, however,

$$||v_h||_{H^{0,2}(\Gamma)} = \lim_k \left\|\frac{\partial^{t_h} f_k}{\partial n^{t_h}}\right\|_{H^{0,2}(\Gamma)} \leqq c' \lim_k ||f_k||_{H^{m,2}(G)};$$

thus $||v_h||_{H^{0,2}(\Gamma)} = 0$. But the identification mapping of $H^{s_h,2}(\Gamma)$ into $H^{0,2}(\Gamma)$ is one-to-one and hence $||v||_{H^{s_h,2}(\Gamma)} = 0$ for each h. Thus $||u||_H = 0$ and T is one-to-one.

The forms $a(f, g)$, $\zeta(f, g)$, $\phi(f, g)$, and the L^2 inner product (f, g) as defined on the dense subset V of H may all be extended to bounded forms on H.

The elliptic boundary value problem $\boldsymbol{B}$ with homogeneous boundary conditions asks for a solution $u \in C^{2m}(G)$ of the equation $Au = f$, for a given function f which lies in H and for all $g \in V$ satisfies the integral condition

$$a(u, g)+\zeta(u, g)+\phi(u, g) = (-1)^m(Au, g). \tag{2.1}$$

We observe that if u had a sufficient number of derivatives continuous on the closed domain $\overline{G}$, the difference between the right and left hand sides of (2.1) could be expressed by integration by parts as an integral over the boundary of G. Under these conditions (2.1) would reduce to a family of linear conditions on the boundary derivatives. In general terms, our principal result is that with smooth coefficients and data, this situation always arises and the boundary conditions are assumed in the differential pointwise sense. We may consider some significant special cases to illustrate the general definition.

EXAMPLE 1. The Dirichlet problem: A any uniformly elliptic operator on G, $a(f, g)$ any one of its Dirichlet forms, $V = C_0^\infty(G)$, $\zeta = \phi = 0$.

EXAMPLE 2. The Neumann problem: A an elliptic differential operator with an associated form $a(f, g)$ coercive on $C^r(\overline{G})$, $V = C^r(\overline{G})$, $\zeta = \phi = 0$.

EXAMPLE 3. The mixed Dirichlet and Neumann problem: A a uniformly elliptic differential operator on G with an associated Dirichlet form $a(f, g)$ coercive on the space V of functions from $C^r(\bar{G})$ vanishing on neighborhoods of a subset Γ_1 of Γ, $\zeta = \phi = 0$.

EXAMPLE 4. A the Laplace operator, $V = C^r(\bar{G})$, $a(f, g)$ the ordinary Dirichlet integral, $\zeta = 0$, $\phi(f, g) = \int \xi(x) f g^* \, dS$. Here the boundary condition becomes $\partial u/\partial n + \xi u + 0$.

Equation (2.1) implies a functional equation for u in H; for all g in V we have

$$a(u, g) + \zeta(u, g) + \phi(u, g) = (-1)^m (f, g). \tag{2.2}$$

Equation (2.2) is our basic functional equation in H which will correspond to the boundary value problem $\boldsymbol{B}$. We may transform (2.2) into a more convenient form by writing it as

$$[u, g] + b(u, g) + \phi(u, g) - k_0(u, g) = (-1)^m (f, g), \qquad g \in V \tag{2.3}$$

where we have set $b(u, g) = \frac{1}{2}\{a(u, g) - a(g, u)^*\}$. It may be shown that there exists a bounded linear transformation V of H as well as a bounded linear transformation G from $H^{-m,2}$ into H such that $(-1)^m (f, g) = [Gf, g]$ for $g \in H$ while

$$b(u, g) + \phi(u, g) - k_0(u, g) = [Vu, g] \qquad \text{for } u, g \in H.$$

In terms of V and G, (2.2) becomes $(I + V)u = Gf$. It is easily shown that V is the sum of a skew-Hermitian and a compact transformation while G considered as an operator in H is compact. As a consequence, the theory of linear compact transformations on a Hilbert space is applicable to (2.2) to yield the familiar results concerning the Fredholm alternative, eigenvalues and eigenfunctions. In particular, if (2.2) has only the trivial solution $u = 0$ for $f = 0$, then a solution exists for every f in $H^{-m,2}(G)$; also there exists a $c_2 \geqq 0$ such that for each f and its corresponding solution u,

$$||u||_{H^{m,2}(G)} \leqq c_2 ||f||_{H^{-m,2}(G)} \leqq c_3 ||f||_{L^2(G)}.$$

If we add a large zero-order term to A of the correct sign, the new operator will satisfy the above uniqueness condition.

3. Regularity in the Interior

Our results on the regularity of the generalized solutions in the interior follow from the application of Korn's trick, a perturbation procedure on the properties of solutions of equations with constant coefficients which has been employed for second-order equations and systems by E. Hopf, Schauder and Morrey [12, 16, 18]. The principal technical refinement upon these latter proofs is the use of the important observation made by Nirenberg

that the theorems of Calderon and Zygmund [8] on singular integral operators in the L^p spaces yield precisely the technical information necessary to apply the Korn argument in the spaces $H^{2m,p}(G_1)$ for compact subdomains G_1 contained in the interior of G.

THEOREM 1. (i) *If* $a_{\alpha\beta} \in C^r(G)$, $r \geqq 0$, *then* $u \in H^{m+r,p}(G_1)$ *for every compact subdomain* G_1 *of* G *provided that* $f \in L^{p_1}(G)$, $p_1 = pn/[n+p(m-r)]$ *for* $r \leqq m$; $f \in H^{r-m,p}(G)$ *for* $r > m$.

(ii) *There exists* c_{G_1} *for every compact subset* G_1 *of* G, *such that*

$$||u||_{H^{m+r,p}(G_1)} \leqq c_{G^1} \begin{cases} ||f||_{L^{p_1}(G)} + ||u||_{L^1(G)}, & r \leqq m, \\ ||f||_{H^{r-m,p}(G)} + ||u||_{L^1(G)}, & r > m. \end{cases}$$

(iii) *If* $a_{\alpha\beta} \in C^{r,\lambda}(G)$, $0 < \lambda$, *then* $u \in C^{m+r,\lambda}(G_1)$ *for every compact subdomain* G_1 *of* G *provided* $f \in C^{0,\lambda}(G)$ *for* $r \leqq m$, $f \in C^{r-m,\lambda}(G)$ *for* $r > m$

(iv) *There exists a* c_{G_1} *for every compact* G_1 *in* G *such that if the assumptions of (iii) hold, then*

$$||u||_{C^{m+r,\lambda}(G_1)} \leqq c_{G_1}\{||f||_{C^{\max(r-m,0),\lambda}(G)} + ||u||_{L^1(G)}\}.$$

THEOREM 2. (i) *If* $a_{\alpha\beta} \in C^r(G)$, $f \in L^{p_1}(G)$, $p_1 > n/(2m-r)$ *for* $r \leqq m$, $f \in H^{r-m,p}(G)$, $p > n/m$ *for* $r > m$, *then* u *will have continuous* $(m+r-1)$*-st order derivatives in* G, *satisfying a Hölder condition on every compact subdomain of* G.

(ii) If $a_{\alpha\beta} \in C^0(G)$ *and if the solution* u *of the abstract boundary value problem* (2.2) *is unique for each* f, *then there exists a Green's function for the boundary value problem* $\boldsymbol{B}$, *i.e., a function* $g(x, z)$ *such that* $u(x) = \int_G g(x, z) f(z) dz$ *for* x *in* G *while* $\int_G |g(x, z)|^q dz < \infty$ *for each* x, *and* $q < n/(n-2m)$ *if* $n \geqq 2m$ *and all* q *if* $n < 2m$.

(iii) If $Au = \sum_{|\zeta| \leq 2m} c_\zeta(x) D^\zeta u$ *with* $c_\zeta \in C^{r,\lambda}(G)$ *and if* $f \in C^{r,\lambda}(G)$ *and* u *has continuous* $2m$*-th derivatives, then* $u \in C^{2m+r,\lambda}(G_1)$ *for each compact* G_1 *in* G *and there exists a constant* $c_{\lambda,r}(G_1)$, *such that for each such* f *and* u

$$||u||_{C^{2m+r,\lambda}(G_1)} \leqq c_{\lambda,r}(G_1)\{||f||_{C^{r,\lambda}(G)} + ||u||_{L^1(G)}\}.$$

4. Regularity at the Boundary

In order to obtain estimates over the whole domain G rather than compact subdomains, we restrict ourselves to the $H^{r,2}$ norms and apply a different method of estimation.

The following results are valid for any elliptic boundary value problem of type $\boldsymbol{B}$ (see Section 2) provided that the coefficients of the boundary forms involved have the same differentiability class as that required of the coefficients $a_{\alpha\beta}$ in the statements of Theorems 3 and 4 below; in addition,

we must assume that the form $a(f, g)$ in the boundary value problem is coercive on a neighborhood of each point of the boundary which is contained in the support of some function from V.

THEOREM 3. (i) *If* $a_{\alpha\beta} \in C^m(\bar{G})$, $f \in L^2(G)$, Γ *of class* C^{2m}, *then* $u \in H^{2m,2}(G)$ *and there exists a constant* $c \geqq 0$, *such that for all* $f \in L^2(G)$

$$||u||_{H^{2m,2}(G)} \leqq c\{||f||_{L^2(G)}+||u||_{L^1(G)}\}.$$

(ii) *If* $a_{\alpha\beta} \in C^{m+s}(\bar{G})$, $s > 0$, Γ *of class* C^{2m+s}, *and* $f \in H^{s,2}(G)$, *then* u *lies in* $H^{2m+s,2}(G)$ *and there exists a* $c_s(G) \geqq 0$ *such that for all* u *and* f *in the given boundary value problem*

$$||u||_{H^{2m+s,2}(G)} \leqq c_s\{||f||_{H^{s,2}(G)}+||u||_{L^1(G)}\}.$$

THEOREM 4. *If* $a_{\alpha\beta} \in C^{m+s}(\bar{G})$, *is of class* C^{2m+s}, *and* $f \in H^{s,2}(G)$ *for* $s+2m > n/2+t$ $(t \geqq 0)$, *then* $u \in C^t(\bar{G})$. *Furthermore, if the solution of the boundary value problem is unique for each* f *and if we set* $u = Gf$ *then, for* $r > n/4m$, G^r *is an operator of Hilbert-Schmidt type in* $L^2(G)$ *(finite double norm).*

The detailed development of the last conclusion and of its method of proof leads in a very simple way to the completeness of the eigenfunctions in non-selfadjoint problems as well as the asymptotic distribution of eigenvalues and eigenfunctions.

To bring out the main ideas of the proof of Theorem 3 which is the main result of our investigations, we may, without loss of generality, consider our solution u in the portion of the sphere $|x| < 1$ for which $x_n \geqq 0$, A having constant coefficients. u may be supposed to vanish on a neighborhood of the outer surface of the hemisphere and to satisfy a variational boundary value problem of type $\boldsymbol{B}$, as described in Section 2, on $x_n = 0$. Utilizing a technique introduced by Lichtenstein [14] which has been applied to second-order equations by Morrey, Schiffman, and Nirenberg [15, 16, 17, 19] (and to higher order equations by Nirenberg) one may estimate the L^2 norms of the tangential derivatives of u by taking tangential difference quotients and estimating their $H^{m,2}$ norms, using the fact that these difference quotients are themselves solutions of variational boundary value problems. Uniform estimates of this type for difference quotients, however, can be transformed without difficulty into L^2 estimates of the various tangential derivatives. To obtain these estimates for the non-tangential derivatives, we use a lemma of Aronszajn [1] on coercive forms. A simplified form of Aronszajn's lemma states that if on a smoothly bounded domain G we are given a finite set of differential operators $L_i = \sum_{|\alpha|=t} c_{i,\alpha} D^\alpha$, $i = 1, \cdots, s$, and if, setting

$$k_i(\zeta) = \sum_{|\alpha|=t} c_{i,\alpha} \zeta^{[\alpha]}, \qquad \zeta^{[\alpha]} = \zeta_{\alpha_1} \cdots \zeta_{\alpha_t},$$

we have

$$\sum_{i=1}^{s} |k_i(\zeta)|^2 > 0$$

for every complex n-vector $\zeta \neq 0$, then there exists a constant $k \geqq 0$ such that for $g \in H^{t,2}(G)$

$$||g||^2_{H^{t,2}(G)} \leqq k \left\{ \sum_{i=1}^{s} ||L_i g||^2_{L^2(G)} + ||g||^2_{L^2(G)} \right\}.$$

In particular, if we choose $L_1 = A$ and L_i, $i \geqq 2$, to be the various tangential derivatives, simultaneous vanishing of $k_{i(\zeta)}$ for $i \geqq 2$ implies $\zeta_i = 0$ for $i \geqq 2$. This in turn implies that $k_1(\zeta)$, which equals $a\zeta_1^{2m}$ if the other ζ vanish, will be zero only if $\zeta_1 = 0$. It follows, applying the lemma, that

$$||u||^2_{H^{2m,2}(G)} \leqq k\{||Au||^2_{L^2(G)} + \sum_{i \geqq 2} ||L_i u||^2_{L^2(G)} + ||u||^2_{L^2(G)}\},$$

and, applying the preceding estimate of the tangential derivatives, that

$$||u||^2_{H^{2m,2}(G)} \leqq k_1\{||Au||^2_{L^2(G)} + ||u||^2_{L^2(G)}\}.$$

Bibliography

[1] Aronszajn, N., *On coercive integro-differential forms*, Studies in Eigenvalue Problems No. 14, Lawrence, Kansas, 1955, pp 94–106.

[2] Browder, F. E., *The Dirichlet problem for linear elliptic equations of arbitrary even order with variable coefficients*, Proc Nat. Acad Sci., Vol 38, 1952, pp 230-235

[3] Browder, F. E., *The Dirichlet and vibration problems for linear elliptic differential equations of arbitrary order*, Proc Nat Acad. Sci., Vol. 38, 1952, pp. 741–747

[4] Browder, F. E., *Assumption of boundary values and the Green's function in the Dirichlet problem for the general linear elliptic equation*, Proc. Nat. Acad. Sci., Vol. 39, 1953, pp. 179–184.

[5] Browder, F. E., *Linear parabolic differential equations of arbitrary order, general boundary-value problems for elliptic equations*, Proc Nat Acad. Sci., Vol. 39, 1953, pp. 185–190

[6] Browder, F. E., *On the eigenfunctions and eigenvalues of the general linear elliptic differential operator*, Proc. Nat. Acad. Sci., Vol. 39, 1953, pp. 433–439.

[7] Browder, F. E., *Strongly elliptic systems of differential equations*, Ann. Math. Study No 33, Princeton University Press, 1954, pp. 15-51.

[8] Calderon, A. P. and Zygmund, A., *On the existence of certain singular integrals*, Acta Math., Vol. 88, 1952, pp 85–139

[9] Courant, R. and Hilbert, D., *Methoden der mathematischen Physik*, Vol 2, Springer, Berlin, 1937, Chapter VII.

[10] Friedrichs, K. O., *On the differentiability of solutions of linear elliptic differential equations*, Comm. Pure Appl Math., Vol. 6, 1953, pp. 299-325.

[11] Gårding, L., *Dirichlet's problem for linear elliptic partial differential equations*, Math. Scand., Vol. 1, 1953, pp. 55–72.

[12] Hopf, E., *Über den funktionalen, insbesondere den analytischen Character der Losungen elliptischer Differentialgleichungen zweiter Ordnung*, Math Z., Vol. 34, 1931, pp. 191–233

[13] John, F., *The fundamental solution of linear elliptic differential equations with analytic coefficients*, Comm. Pure Appl. Math., Vol. 3, 1950, pp. 273–304.

[14] Lichtenstein, L., *Über den analytischen Character der Losungen zwei-dimensionaler Variationsprobleme*, Bull. Acad Sci Cracovie, Ser. A, 1912, pp. 915–941
[15] Morrey, C. B. Jr, *Multiple integral problems in the calculus of variations and related topics*, Univ California Publ Math. N. S, Vol 1, 1943, pp 1–130.
[16] Morrey, C. B Jr., *Second-order elliptic systems of differential equations*, Ann Math Study No 33, Princeton University Press, 1954, pp 101–151.
[17] Nirenberg, L., *On non-linear elliptic partial differential equations and Holder continuity*, Comm. Pure Appl. Math, Vol 6, 1953, pp 103–156.
[18] Schauder, J., *Über lineare elliptische Differentialgleichungen zweiter Ordnung*, Math Z, Vol 38, 1934, pp. 257–282.
[19] Schiffmann, M., *Differentiability and analyticity of solutions of double integral variational problems*, Ann. Math., Vol. 48, 1947, pp 274–284
[20] Schwartz, L, *Théorie des distributions*, Hermann, Paris, Vol 1, 1950, Vol. 2, 1951.
[21] Sobolev, S, *On a theorem of functional analysis*, Mat. Sbornik, Vol 4, 1938, pp. 471-497
[22] Visik, M. I, *On strongly elliptic systems of differential equations*, Mat Sbornik, Vol 29, 1951, pp 615–676

Received January 20, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 363–371 (1956)

On the Calculus of Variations in Two Variables

LAMBERTO CESARI
Purdue University

Recent developments in the calculus of variations have been characterized by an enlargement of the class of the elements among which maxima, minima, stable and unstable extremals are being sought. In referring here to some of these developments of the calculus of variations for continuous parametric surfaces, I will emphasize 1) the concept of integral for continuous parametric surfaces defined with the use of topological considerations, 2) some of the topological and analytical processes used to assure equicontinuity of minimizing sequences, 3) some recent existence theorems for absolute maxima. Finally, I will mention some still unsolved problems concerning differentiability properties of the solutions.

1. The Concept of Integral over a Parametric Surface

An integral $I(S) = (S)\int f(p, t)\,dw$ over a continuous parametric surface S of a function $f(p, t)$ of the point $p = (x, y, z)$ on S and of the direction $t = (t_1, t_2, t_3)$ could be introduced as a Lebesgue-Tonelli integral for surfaces S of finite area given in an absolutely continuous representation. Yet it is preferable to introduce it as a Weierstrass integral for every representation of S [5a]. This also is in accordance with the direct method of the calculus of variations for curves and thus in line with L. Tonelli (1911), K. Menger (1937), G. Bouligand, N. Aronszajn, C. Pauc.

Let S: $p = p(w)$, $w \in A$, $w = (u, v)$, $p = (x, y, z)$, be any continuous surface in E_3, defined as a continuous mapping $p = p(w)$ from an admissible set A of the oriented w-plane E_2 into the oriented p-space E_3. For what concerns admissible sets let us mention that every open plane set $G \subset E_2$, every closed finitely connected Jordan region $R \subset E_2$, every set $A \subset R$ open in a region R, are admissible sets though the class of the admissible sets contains also other types. If A is a simple closed region, then S is said to be of the type of the disc; if in addition $p(w)$ is constant on the boundary curve A^* of A but not constant on A, then S is said to be of the type of the sphere.

If $\pi \subset A$ is any simple closed polygonal region in A, then the counterclockwise oriented boundary π^* of π has an image in E_3 which is an oriented closed curve C not necessarily simple. The projections of C on the yz, zx, xy oriented coordinate planes of E_3 (say the planes E_{21}, E_{22}, E_{23}) are oriented

closed plane curves C_1, C_2, C_3 not necessarily simple. Then the topological index $o(p; C_r)$, $p \in E_{2r}$, of any point $p \in E_{2r}$ not on C_r is defined and represents the number $\gtreqless 0$ of times C_r "links" the point p. We will put $o(p; C_r) = 0$ for every p on C_r, and thus, $o(p; C_r)$, $p \in E_{2r}$, can be thought of as a single-valued, integer-valued function of p in E_{2r}, $r = 1, 2, 3$. Also, $o(p; C_r)$ is finite everywhere, B-measurable in E_{2r}, and constant on each complementary component of C_r in E_{2r}.

A mild hypothesis in the large as to the finiteness of the Lebesgue area $L(S)$ of S is sufficient to assure that the functions $o(p; C_r)$, $p \in E_{2r}$, are L-integrable in E_{2r}, $r = 1, 2, 3$. Therefore, for each simple polygonal region $\pi \subset A$ we can define the real numbers

$$u_r = u_r(\pi) = (E_{2r}) \int o(p; C_r)\, dp, \qquad r = 1, 2, 3,$$

and therefore the vector $u = u(\pi) = (u_1, u_2, u_3)$, whose norm will be denoted by $|u|$. If $|u| > 0$ then the quotients

$$a_r = a_r(\pi) = \frac{u_r(\pi)}{|u(\pi)|}, \qquad r = 1, 2, 3, \quad a_1^2 + a_2^2 + a_3^2 = 1,$$

define a vector $a = a(\pi) = (a_1, a_2, a_3)$ of norm one. The numbers a_1, a_2, a_3 can be thought of as the direction cosines of the "average" normal to the piece of the surface S defined by $p = p(w)$ on π.

Now let $f(p, t) = f(x, y, z, t_1, t_2, t_3)$ be any continuous function of the point $p = (x, y, z)$ in a closed subset K of E_3 and of the vector $t = (t_1, t_2, t_3)$ satisfying the homogeneity condition $f(p, kt) = kf(p, t)$ for all $k \geqq 0$. For the sake of simplicity, we will suppose that K is compact. Let S: $p = p(w)$, $w \in A$, be any continuous surface of finite Lebesgue area, whose points $p(w)$, $w \in A$, are all in K.

Let $\gamma = [\pi]$ be any finite system of disjoint simple closed polygonal regions $\pi \subset A$ and, for every $\pi \in \gamma$, let $\bar{w}$ be any point of π, $\bar{p} = p(\bar{w})$ the image of $\bar{w}$ on S and, therefore, in K, $u = u(\pi) = (u_1, u_2, u_3)$ and $a = a(\pi) = (a_1, a_2, a_3)$ the vectors defined above, where we disregard all $\pi \in \gamma$ with $|u(\pi)| = 0$.

In various ways an index δ, or a system (δ) of indices, can be defined measuring the fineness of the system γ with respect to the surface S. In any case for every continuous surface S of finite Lebesgue area the following limit exists and is finite (L. Cesari [5a]):

$$(1) \qquad I(S) = (S) \int f(p, t)\, dw = \lim_{\delta \to 0} \sum_{\pi \in \gamma} f(\bar{p}, u) = \lim_{\delta \to 0} \sum_{\pi \in \gamma} f(\bar{p}, a) |u(\pi)|.$$

No matter what is the choice of δ, the approaching to zero of δ must assure that the maximum of the oscillations of $p = p(w)$ on the polygons π of γ

approaches zero, that the two-dimensional measures of the closed plane sets covered by the curves C_r on E_{2r} also approach zero, and finally that the system γ involves a part as large as we want of the area of the surface S.

The limit (1) is called the (Weierstrass) integral of the function $f(p, t)$ on the surface S and has a number of typical properties which we list below.

1) $I(S)$ is independent of the representation $p = p(w)$, $w \in A$, of the surface S (L. Cesari [5a]).

2) If $[S_n]$ is a sequence of surfaces S_n in K and $||S_n, S|| \to 0$, $L(S_n) \to L(S)$ as $n \to \infty$, where $||S_n, S||$ denotes the Fréchet distance and $L(S_n) < +\infty$, $L(S) < +\infty$, then $I(S_n) \to I(S)$ as $n \to \infty$ [5a].

3) If $f(p, t)$ is supposed to have continuous partial derivatives with respect to t_1, t_2, t_3, then a semiregularity condition can be introduced by means of the Weierstrass function $E(p, t, t')$. If $I(S)$ is non-negative and semiregular, if $[S_n]$ is a sequence of surfaces in K and $||S_n, S|| \to 0$ as $n \to \infty$, and $L(S_n) < +\infty$, $L(S) < +\infty$, then $I(S) \leqq \underline{\lim}\, I(S_n)$ as $n \to \infty$ (lower semicontinuity of $I(S)$) (L. Cesari [5b]).

4) Under a condition of regularity somewhat stronger than usual, $||S_n, S|| \to 0$, $I(S_n) \to I(S)$ as $n \to \infty$ implies $L(S_n) \to L(S)$ (V. E. Bononcini [2a]).

5) For linear integrals, that is, for integrals of the form

$$I(S) = (S) \iint [A(p)t_1 + B(p)t_2 + C(p)t_3]\, dw,$$

both Stokes and Gauss-Green theorems hold in the usual form (J. Cecconi [4]).

6) If S_n, S are surfaces of the type of the disc and $I(S)$ is a linear integral, if $||S_n, S|| \to 0$ and both the areas $L(S_n)$ and the lengths $l(0S_n)$ of the boundary curves $0S_n$ of the surfaces S_n are bounded from above, then $I(S_n) \to I(S)$ (continuity of linear integrals) (V. E. Bononcini [2b]).

7) If S is of the type of the sphere and $L(S) < +\infty$, then the linear integrals

$$V(S) = (S) \int z t_1 dw = (S) \int x t_2 dw = (S) \int y t_3 dw$$

have a common value $V(S) \gtreqless 0$ which can be thought of as the relative volume enclosed by S. If S_n, S are surfaces of the type of the sphere in K with $||S_n, S|| \to 0$, $L(S_n) < M$, then $V(S_n) \to V(S)$ (continuity of the volume in any class of surfaces of the type of the sphere whose areas are bounded from above) (Bononcini [2b]).

8) A cyclic additivity theorem for $I(S)$ analogous to the one for Lebesgue area was proved by J. Cecconi [4e] and, recently, also a strong additivity theorem has been proved by C. Neugebauer [12]. All these are properties of the integral $I(S)$ analytical in character and not involving differential properties

(in any proper or generalized sense) of the vector function $p(w)$. If we recall now that $L(S) < +\infty$ implies the existence almost everywhere in A^0 (A being any admissible set) of the generalized Jacobian vector $J = (J_1, J_2, J_3)$ (set theoretically and topologically defined) then we may add the following property (differential geometric in character):

9) If $L(S) < +\infty$ and the representation $p = p(w)$, $w \in A$, of S is such that the Lebesgue area is given by the classical area integral

$$L(S) = (A^0) \int |J| \, dw,$$

then the integral $I(S)$ is given by the Tonelli-Lebesgue integral

$$I(S) = (A^0) \int f(p, J) \, dw$$

(L. Cesari [5a]).

2. Smoothing Processes for Parametric Surfaces

(a) *Topological processes.* We shall restrict ourselves here to surfaces S: $p = p(w)$, $w \in A$, of the type of the disc; i.e., A is a closed simple region. The maximal continua $g \subset A$ of constancy for the vector $p(w)$ are of particular importance. Here even a single point $g = (w)$ is considered a continuum and thus the collection Γ of all the maximal continua g of constancy for $p(w)$ constitutes a subdivision of A into disjoint continua. A surface S is said to be *base* if no continuum $g \in \Gamma$ separates E_2; a surface S is said to be *open non-degenerate* if no continuum $g \in \Gamma$ separates A, or contains entirely the boundary A^* of A.

The operation of retraction of a surface S can be defined by abstract topological considerations (Whyburn) as well as by more elementary devices. The following theorems hold concerning the operation of retraction.

1) *Every surface S of the type of the disc possesses a base retraction S_0 having the same boundary curve $\theta S_0 = \theta S$* (C. B. Morrey [11h], E. J. McShane [9]).

2) *Given any simple closed continuous curve C in E_3 and any $\varepsilon > 0$, there is a $\delta = \delta(C, \varepsilon) > 0$ such that every surface S with $||\theta S, C|| < \delta$ has an open non-degenerate retraction S_0 with $||\theta S_0, C|| < \varepsilon$* (L. Cesari [5f]).

Both statements 1) and 2) can be used for the reduction of the surfaces S_n of a minimizing sequence to open non-degenerate surfaces.

(b) *An analytical smoothing process.* A minimizing sequence of surfaces, even all open non-degenerate, may not satisfy conditions of equicontinuity. This can be easily understood if we observe that each surface may contain thin cones of large diameter and very small area whose contribution to the value of the integral is then as small as we want. From here we conclude

that some further smoothing processes are needed beyond the statements 1) and 2) discussed in (a). For "saddle" surfaces, or, more generally, for those surfaces $p = p(w)$, $w \in A$, for which for every simple region $r \subset A$ the oscillation of $p(w)$ in r is equal to the oscillation of $p(w)$ on the boundary r^* of r, no further process of smoothing is needed (E. J. McShane [9]).

First we recall here the following lemma:

If $g(x)$, $0 \leqq x \leqq a$, is a real single-valued L-integrable function with

$$0 < \int_x^a g(\alpha)\, d\alpha \leqq k\, g^2(x)$$

almost everywhere in $0 \leqq x \leqq a$ and in particular for $x = 0$, then $a \leqq 2k\, g(0)$ (L. Cesari [5e]; another proof has been given by L. M. Graves).

Let us consider a polyhedral surface S of the type of the disc whose boundary curve θS has diameter D. Let p_0 be any point of θS and R the minimum radius of the sphere of center p_0 containing S. Suppose $R > D$ and let us consider for every $D \leqq \varrho \leqq R$ the intersection of the sphere $|p - p_0| = \varrho$ with the surface S. Such an intersection can be thought of as a finite sum of elementary continuous closed curves whose total length $l(\varrho)$ is a function of ϱ, $D \leqq \varrho \leqq R$. We shall suppose here, for the sake of simplicity, that $l(\varrho)$ is continuous in $D \leqq \varrho \leqq R$, and that $l(\varrho) > 0$ if $D \leqq \varrho < R$, $l(R) = 0$.

Let $k \geqq 1$ be any arbitrary constant. It cannot occur that

$$l(\varrho) > [k^{-1}a(S)] \qquad \text{for all } D \leqq \varrho \leqq D + [ka(S)]^{1/2} \leqq R,$$

where $a(S)$ is the area of S. Indeed, in such a case, we would have

$$a(S) = \int_{D_i}^{D+[ka(S)]^{1/2}} [k^{-1}a(S)]^{1/2} d\varrho < \int_D^{D+[ka(S)]^{1/2}} l(\varrho)\, d\varrho \leqq a(S),$$

a contradiction. Thus, there is a number r_1, $D < r_1 < D + [ka(S)]^{1/2}$, with

$$l(r_1) \leqq [k^{-1}a(S)]^{1/2}.$$

Let us now suppose that

$$kl^2(\varrho) \geqq \int_\varrho^R l(\alpha)\, d\alpha$$

for every $r_1 \leqq \varrho \leqq R$. Then by the lemma above, we have

$$R - r_1 \leqq 2kl(r_1) \leqq 2[ka(S)]^{1/2}$$

and thus

$$R \leqq D + 3[ka(S)]^{1/2}.$$

In the opposite case there is a number r_2, $r_1 \leqq r_2 \leqq R$, with

$$kl^2(\varrho) \geqq \int_\varrho^R l(\alpha) d\alpha \qquad \text{for all } r_1 \leqq \varrho \leqq r_2,$$

$$kl^2(r_2) \leqq \int_{r_2}^R l(\alpha)\, d\alpha.$$

Then by the lemma, we have

$$r_2 - r_1 \leqq 2kl(r_1) \leqq 2[ka(S)]^{1/2}.$$

On the other hand we can suppress all parts of S outside the sphere $|p-p_0| = r_2$ and replace all these parts by elementary surfaces spanning the various elementary curves which form the intersection of S with the sphere $|p-p_0| = r_2$. Thus we replace S by a surface S_0 which is contained in the sphere of center p_0 and radius

$$r_2 \leqq D + 3[ka(S)]^{1/2}$$

and whose area is less than $\frac{1}{4}l^2(r_2)$, plus the area a_1 of the part S_1 of S inside the sphere $|p-p_0| \leqq r_2$. The area of S has, therefore, been reduced by d, for which the inequalities

$$d > a(S-S_1) - \tfrac{1}{4}l^2(r_2) \geqq \int_{r_2}^{R} l(\alpha)\, d\alpha - \tfrac{1}{4}l^2(r_2) \geqq (k-\tfrac{1}{4})l^2(r_2)$$

hold. Finally, if we suppose that $0 < m \leqq f(p, t) \leqq M$ for all $p \in K$, $|t| = 1$, and $k = M/m$, then the value of the integral $I(S)$ has been reduced by a number d' for which the inequalities

$$d' > ma(S-S_1) - \tfrac{1}{4}Ml^2(r_2) \geqq \frac{3mk}{4}\, l^2(r_2)$$

hold.

3. Existence Theorems for Absolute Minima

By a consistent repeated use of the properties of the integral $I(S)$ (§1), of the processes described above (§2), together with the well-known cardinal representation theorem of C. B. Morrey [11i] of open non-degenerate surfaces, and the (ε, δ) process of L. C. Young [18] it is possible to obtain minimizing sequences of polyhedral surfaces for which equicontinuity holds and thus, the Ascoli theorem can be applied. The following theorem has been proved:

I. *Let K be any closed bounded convex set of E_3 and C a simple closed curve in K. Let W be the family of all oriented Fréchet surfaces S in K whose area is finite and whose boundary curve is C, and let us suppose that W is not empty. Then every positive definite and semiregular integral $I(S)$ has an absolute minimum in W* (L. Cesari [5e]).

This theorem holds also if $K = E_3$ provided $0 < m \leqq f(p, t) \leqq M < +\infty$, for all $p \in E_3$, $|t| = 1$, and some constants m, M.

By other methods, the same theorem has been obtained by A. G. Sigalov [16a]. An analogous theorem has been obtained also by J. M. Danskin [7] under slightly more restrictive conditions. A. G. Sigalov, J. M. Danskin, and L. Cesari have used the integral $I(S)$ defined in §1.

From a different view point W. H. Fleming and L. C. Young have obtained independent analogous results under the hypothesis that the given simple curve θS be rectifiable.

Using the processes discussed above J. Cecconi has recently proved a theorem analogous to I where W is the class of all surfaces S of the type of the sphere, contained in $K-K_0$ and not nulhomotop in $K-K_0$ where K_0 is a closed sphere contained in the interior of K. Various extensions have been announced from the results above by J. Cecconi [4] and L. Cesari [5].

Finally J. Cecconi, by using the same processes, has improved the previous results of A. G. Sigalov for non-parametric surfaces.

4. Remarks on the Differentiability Problem

The previous existence theorems have been obtained by direct methods and extend the existence theorem of E. J. McShane (1935) concerning integrals $(S)\int f(t)\,dw$ where f depends on t only. Also, the same existence theorems contain, as a particular case, the analogous existence theorem for Plateau's problem for surfaces of the type of the disc (surface of minimum area spanning a given simple curve C in E_3). (See T. Radó [15], J. Douglas [8], E. J. McShane [9], R. Courant [6], L. Tonelli [17], 1929-36.) For Plateau's problem, the more difficult case of surfaces of higher topological type has been solved by J. Douglas. Nevertheless, while for Plateau's problem the solution has been proved to be smooth, even analytic, and to satisfy the Euler partial differential equation (minimal surface), an analogous result has not yet been proved for the solutions of a general problem as in I (§3), even under convenient conditions of differentiability or analyticity of f.

The analogous problem (differentiability problem) for non-parametric surfaces has been recently solved independently by A. G. Sigalov [16b] and C. B. Morrey [11]. The results of C. B. Morrey concern essentially partial differential equations, or systems of such equations, and have been recently largely extended by the same author and by L. Nirenberg [13].

Bibliography

[1] Aronszajn, N., *Quelques recherches sur l'intégrale de Weierstrass*, Revue Sci. Math., Vol. 77, 1939, pp. 490–493; Vol. 78, 1940, pp. 165–167, 233–239.

[2] Bononcini, V. E., (a) *Sugli integrali regolari del calcolo delle variazioni per superficie in forma parametrica*, Rivista Mat. Univ. Parma, Vol. 3, 1952, pp. 131–151. (b) *Un teorema di continuità per integrali su superficie chiuse*, Ibid., Vol. 4, 1953, pp. 299–311.

[3] Bouligand, G., *Essai sur l'unicité des méthodes directes*, Mém. Soc. Roy. Sci. Liège, Sér. 3, Vol. 19, 1934, p. 88.

[4] Cecconi, J., (a) *Sul teorema di Gauss-Green*, Rend. Sem. Mat. Univ. Padova, Vol. 20, 1951, pp. 194–218. (b) *Sul teorema di Stokes*, Rivista Mat. Univ. Parma, Vol. 3, 1952,

pp. 233–264 (c) *Sul teorema di Gauss-Green per una particolare classe di superficie*, Rend Sem. Mat. Univ. Padova, Vol. 22, 1953, pp. 81–112. (d) *Un complemento al teorema di Stokes*, Ibid , pp 23–37. (e) *Sull' additività ciclica degli integrali sopra una superficie*, Rivista Mat. Univ Parma, Vol 4, 1953, pp. 43–67.

[5] Cesari, L., (a) *La nozione di integrale sopra una superficie in forma parametrica*, Ann. Scuola Norm. Super. Pisa, Ser. 2, Vol. 13, 1946, pp. 78–117. (b) *Condizioni sufficienti per la semicontinuità degli integrali sopra una superficie in forma parametrica*, Ibid , Ser. 2, Vol. 14, 1948, pp. 47–79. (c) *Condizioni necessarie per la semicontinuità degli integrali sopra una superficie in forma parametrica*, Ann Mat. Pura Appl., Ser. 4, Vol. 29, 1950, pp. 199–224. (d) *Area and representation of surfaces*, Bull. Amer. Math. Soc , Vol 56, 1950, pp. 218–232. (e) *An existence theorem of calculus of variations for integrals on parametric surfaces*, Amer J. Math , Vol. 74, 1952, pp 265–295. (f) *Surface Area*, Ann. of Math. Studies No. 35, Princeton University Press, 1956. (g) *Retraction, homotopy, integral*, Proc. International Congress Math , Amsterdam, 1954.

[6] Courant, R., (a) *On the problem of Plateau*, Proc. Nat. Acad Sci U.S.A. Vol. 22, 1936, pp. 367–372 (b) *Dirichlet's Principle, Conformal Mapping and Minimal Surfaces*, Interscience, New York, 1950

[7] Danskin, J. M , *On the existence of minimizing surfaces in parametric double integral problems of the calculus of variations*, Rivista Mat. Univ. Parma, Vol 3, 1952, pp 43–63.

[8] Douglas, J., *Solution of the problem of Plateau*, Trans Amer. Math. Soc., Vol. 33, 1931, pp. 263–321.

[9] McShane, E. J., (a) *Parametrization of saddle surfaces with applications to the problem of Plateau*, Trans. Amer. Math. Soc . Vol. 35, 1933, pp 716–733. (b) *Existence theorems for double integral problems of calculus of variations*, Ibid., Vol 38, 1935, pp. 549–563

[10] Menger, K., *Die metrische Methode in der Variationsrechnung*, Erg. Math. Kolloqu , No. 8, 1937, pp. 1–32.

[11] Morrey, C. B., (a) *Existence and differentiability theorem for the solutions of variational problems for multiple integrals*, Bull Amer Math. Soc., Vol. 47, 1940, pp 439–458. (b) *Multiple integral problems in the calculus of variations and related topics*, Univ. Calif. Publ. Math , N. S., Vol. 1, 1943, pp. 1–130. (c) *The problem of Plateau on a Riemannian manifold*, Ann of Math , Ser 2, Vol 49, 1948, pp. 807–851. (d) *The problem of Plateau on a Riemannian manifold*, Proc International Congress Math., Cambridge, Mass , 1950. (e) *Quasi-convexity and the lower semi-continuity of multiple integrals*, Pacific J. Math., Vol. 20, 1952, pp. 25–54 (f) *Second order elliptic systems of differential equations*, Proc. Nat. Acad. Sci. U.S A., Vol. 39, 1953, pp 201–206 (g) *Second order elliptic systems of differential equations*, Ann. of Math Studies No 33, Princeton University Press, 1954, pp. 101–159. (h) *The topology of (path) surfaces*, Amer. J. Math., Vol. 57, 1935, pp. 17–50. (i) *An analytic characterization of surfaces of finite Lebesgue area, I*, Amer. J. Math., Vol. 57, 1935, pp. 692–702.

[12] Neugebauer, C , (a) *A strong cyclic additivity theorem of a surface integral*, Rivista Mat. Univ. Parma (b) *A further extension of a cyclic additivity theorem of a surface integral*, ibid

[13] Nirenberg, L., *On a generalization of quasi-conformal mappings and its application to elliptic partial differential equations*, Ann. of Math. Studies No 33, Princeton University Press, 1954, pp. 95–100.

[14] Pauc, Ch. Y., (a) *Les méthodes directes en calcul des variations et en géométrie differentielle*, Actualités Sci. Ind., No 885, 1951. (b) *L'intégrale de Weierstrass-Bouligand-Menger et ses applications au calcul des variations*, Ann. Scuola Norm. Super. Pisa, Ser. 2, Vol. 8, 1939, pp. 51–68.

[15] Radó, T., (a) *On the Problem of Plateau*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 2, Springer, Berlin, 1933. (b) *Length and area*, Amer. Math. Soc. Colloquium Publication No. 30, 1948

[16] Sigalov, A. G., (a) *Two-dimensional problems in the calculus of variations*, Uspehi Matem. Nauk, N. S., Vol. 6, 1951, pp. 16–101 (translation 83, Amer. Math. Soc. 1953). (b) *On conditions of differentiability and analyticity of solutions of two-dimensional problems of the calculus of variations*, Doklady Akad. Nauk SSSR, N. S., Vol. 85, 1952, pp. 273–275.

[17] Tonelli, L., (a) *Sui massimi e minimi assoluti del calcolo delle variazioni*, Rend. Circ. Mat. Palermo, Vol. 32, 1911, pp. 297–337. (b) *Fondamenti di Calcolo della Variazioni*, Vol. 1 and 2, Zanichelli, Bologna, 1921–23. (c) *Sul problema di Plateau*, Rend. Accad. Lincei, Ser. 6, Vol. 24, 1936, pp. 333–339, 393–398. (d) *L'estremo assoluto degli integrali doppi*, Ann. Scuola Norm. Super. Pisa, Ser. 2, Vol. 2, 1933, pp. 89–130. (e) *Su l'esistenza della estremanti assolute per gli integrali doppi*, Ann. Scuola Norm. Super. Pisa, Ser. 2, Vol. 8, 1939, pp. 161–165.

[18] Young, L. C., (a) *A lemma in the theory of surfaces*, J. London Math. Soc., Vol. 19, 1944, pp. 209–219. (b) *Some applications of the Dirichlet integral to the theory of surfaces*, Trans. Amer. Math. Soc., Vol. 64, 1948, pp. 317–335.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 373–381 (1956)

Discontinuity of Solutions of Quasi-linear Differential Equations in Two Variables

Y. W. CHEN
Wayne University

This paper consists of two parts. The first part is concerned with hyperbolic differential equations of the following form:

$$(1) \qquad A(\varphi_x, \varphi_y)\varphi_{xx}+B(\varphi_x, \varphi_y)\varphi_{xy}+C(\varphi_x, \varphi_y)\varphi_{yy}+k(\varphi_x, \varphi_y)y^{-1}\varphi_y = 0.$$

A discontinuity of a second derivative of φ propagates along the backward characteristic to a point P on the axis $y = 0$, and the question is how this discontinuity will be reflected from the axis into the interior of the domain along the forward characteristic issued from P. The results of this part were published in two papers by the author [4]. We use this opportunity to demonstrate in a very simple manner both the quantitative and the qualitative features of the non-linear behavior of the solution, without going into the details of the convergence procedure which is indicated. The problem is essentially of local nature. Questions concerning the "linearized" theory and the "corrected linearized" theory of (1) have been studied by many authors in connection with gas-dynamic and ballistic problems (see papers by Goldstein [7], Lighthill [8], Meyer [10], Ward [11] and Whitham [12]).

In the second part of this paper we state an existence theorem for the equation of minimal surfaces and give an outline of the proof in §2.1. The problem is to find a solution $z(x, y)$ of

$$(2) \qquad (1+z_y^2)z_{xx}-2z_xz_yz_{xy}+(1+z_x^2)z_{yy} = 0.$$

in the exterior of a convex curve C with a finite number of protruding corners, such that it has a preassigned normal direction at infinity and $\partial z/\partial n = 0$ on C. It turns out that at the vertices of C, $z(x, y)$ will have finite jump discontinuities. A non-existence theorem for equation (2), which is simple to prove, is briefly mentioned in §2.2. The details of the proofs of both theorems are contained in a recent paper by the author [6]. Two open problems are formulated in §2.3; they are concerned with transversality and surfaces with discontinuous first partial derivatives. These questions are closely related to the free boundary theory of minimal surfaces which has been developed by Courant [3] and Lewy [9].

Although the two parts of the paper are entirely independent of each other, they have one point in common in their treatments. New parameters are introduced in each case to replace x, y as independent variables. In terms of the parameters, φ and z behave in a similar manner as in linear problems. It is after the transformation from the parameters back to the space variables, that significant "non-linear" behavior arises in the functions $\varphi(x, y)$ and $z(x, y)$.

§1.1 For simplicity of presentation we consider first the following type of equations:

$$-a(\varphi_x, \varphi_y)\varphi_{xx}+\varphi_{yy}+y^{-1}\varphi_y = 0. \tag{1.1}$$

The general equation (1) can be treated in the same manner and will be discussed briefly in §1.3. At a point $p(x = x_0, y = 0)$ let $\varphi_x = \varphi_x^0$, $\varphi_y = 0$, so that $a(\varphi_x^0, 0)$ is a positive constant $= A^2$ and (1.1) is hyperbolic with respect to the data.

We introduce the linear hyperbolic differential equation with constant coefficients for the second order terms:

$$-A^2\Phi_{xx}+\Phi_{yy}+y^{-1}\Phi_y = 0, \qquad A > 0. \tag{1.2}$$

Its two families of characteristics are given by $y-A^{-1}(x-x_0) = 2\tau$ and $y+A^{-1}(x-x_0) = 2\sigma$. Set $U = \Phi_x$, $V = \Phi_y$, then (1.1) is equivalent to

$$-AU_\sigma+V_\sigma+(\sigma+\tau)^{-1}V = 0, \quad AU_\tau+V_\tau+(\sigma+\tau)^{-1}V = 0.$$

By differentiating the last two equations, one obtains the following equations of second order:

$$\begin{aligned} E_1 &\equiv V_{\sigma\tau}+\tfrac{1}{2}(\sigma+\tau)^{-1}(V_\sigma+V_\tau)-(\sigma+\tau)^{-2}V = 0, \\ E_2 &\equiv U_{\sigma\tau}+\tfrac{1}{2}(\sigma+\tau)^{-1}(U_\sigma+U_\tau) = 0, \\ E_3 &\equiv y-(\sigma+\tau) = 0. \end{aligned} \tag{1.3}$$

When U and V have assigned values $U = f(\tau)$, $V = g(\tau)$ on the backward characteristic $\sigma = 0$, $E_1 = 0$ and $E_2 = 0$ give ordinary differential equations for U_σ and V_σ. The solutions of these differential equations depend on f, g and the initial conditions. If, however, for the initial data we have $\lim \tau^{1/2}U_\sigma = 0$ and $\lim \tau^{1/2}V_\sigma = 0$ as $\tau \to 0$, then the solutions will depend only on f and g, as is easily verified. In this case, and under the assumption that U and V are continuous across the characteristic $\sigma = 0$, there will be no jump discontinuity of U_σ and V_σ along the same line. Hence for the hyperbolic equations (1.3) and (1.4) we shall consider problems with such data assigned on the Mach lines $\sigma = 0$ and $\tau = 0$, that $\tau^{1/2}U_\sigma \to \text{const} \neq 0$, $\tau^{1/2}V_\sigma \to \text{const.} \neq 0$ will hold on $\sigma = 0$, as $\tau \to 0$, and that discontinuity of U_σ and V_σ will occur. A typical problem of this kind is given by the data $U = V = 0$ on $\sigma = 0$ and $U = \text{const.}\,\sigma^{1/2} + O(\sigma^{1/2+\epsilon})$, $V = \text{const.}\,\sigma^{1/2} +$

$O(\sigma^{\frac{1}{2}+\epsilon})$ with positive ε on $\tau = 0$. Since the differential equations are linear, we may find the solutions by superposition. The special solution with $U = V = 0$ on $\sigma = 0$ and $U = \text{const.}\ \sigma^{\frac{1}{2}}$, $V = \text{const.}\ \sigma^{\frac{1}{2}}$ on $\tau = 0$ is given by

$$V = V_0 = \text{const.}\ \frac{\sigma}{\sqrt{\sigma+\tau}}\, F\left(\frac{1}{2}, \frac{1}{2}, 2, \frac{\sigma}{\sigma+\tau}\right),$$

$$U = U_0 = \text{const.}\ \frac{\sigma}{\sigma+\tau}\, F\left(-\frac{1}{2}, \frac{3}{2}, 2, \frac{\sigma}{\sigma+\tau}\right),$$

where F stands for the hypergeometric function. If we know that on the left of the backward characteristic $U(\sigma, \tau) = V(\sigma, \tau) \equiv 0$, then the solutions U_0 and V_0 indeed have discontinuous derivatives with respect to σ along the characteristic $\sigma = 0$. This statement holds true also for the case when additional terms of the order $O(\sigma^{\frac{1}{2}+\epsilon})$ are present in the prescribed values on the forward characteristic $\tau = 0$.

By means of series expansions of the hypergeometric functions, it is easily verified that the following holds:

$$(1.4)\qquad \left(\log\frac{\tau}{\sigma+\tau}\right)^{-1}\frac{\partial U}{\partial \tau} \to \text{const.}\ \sigma^{-\frac{1}{2}}, \qquad \left(\log\frac{\tau}{\sigma+\tau}\right)^{-1}\frac{\partial V}{\partial \tau} \to \text{const.}\ \sigma^{-\frac{1}{2}}$$

as $\tau \to 0$, $\tau > 0$. If we make use of $y = (\sigma+\tau)$ to return to the x, y-plane, we find that

$$(1.5)\qquad \Phi_{xx}, \Phi_{xy}, \Phi_{yy} = \text{const.}\ y^{-\frac{1}{2}}\left(\log\frac{+y-A^{-1}(x-x_0)}{2y} + O(1)\right)$$

holds on the forward characteristic, as a consequence of the discontinuity of these derivatives on the backward Mach line.

§1.2 When the differential equation is non-linear, the Mach lines are not fixed in the x, y-plane; they depend on the solution φ. We shall now return to the x, y-plane by simply setting $y = \sigma+\tau$. Let $u = \varphi_x$, $v = \varphi_y$, $a(u, v) = M^2(u, v)$ in the neighborhood of $u = \varphi_x^0$, $v = 0$, and $M(\varphi_x^0, 0) = A$. The Mach lines are given by

$$(1.6)\qquad -M(u, v)y_\sigma + x_\sigma = 0, \qquad M(u, v)y_\tau + x_\tau = 0.$$

By eliminating x, we obtain

$$(1.7)\qquad 2My_{\sigma\tau} + M_\sigma y_\tau + M_\tau y_\sigma = 0.$$

Let the constants $M_u(\varphi_x^0, 0)$ and $M_v(\varphi_x^0, 0)$ be denoted respectively by M_1 and M_2. Then

$$(1.8)\qquad 2My_{\sigma\tau} + M_1(u_\sigma y_\tau + u_\tau y_\sigma) + M_2(v_\sigma y_\tau + v_\tau y_\sigma) = \cdots.$$

We substitute in (1.8) for u and v the special solution U_0 and V_0 obtained in

§1.1, and get

$$(1.9)\qquad y=(\sigma+\tau)+(2A)^{-1}\int_0^\sigma\int_0^\tau M_1(U_{0\sigma}+U_{0\tau})+M_2(V_{0\sigma}+V_{0\tau})d\sigma d\tau+\cdots.$$

By a simple computation it is found that

$$(1.10)\qquad \left(\log\frac{\sigma}{\sigma+\tau}\right)^{-1}y_\tau\to M_3\sigma^{1/2}$$

holds, as $\tau\to 0$, $\tau>0$, the terms not explicitly represented in (1.8) and (1.9) being taken care of in (1.10) by simple estimates. M_3 is a constant which depends on A, M_1, M_2 and the constant factors in U_0 and V_0. We assume that $M_3\neq 0$.

We obtain the second derivatives φ_{xx}, φ_{xy}, φ_{yy} as

$$(1.11)\qquad \varphi_{xx}=u_x=u_\sigma\sigma_x+u_\tau\tau_x=(u_\sigma y_\tau-u_\tau y_\sigma)/J=(u_\sigma/y_\sigma-u_\tau/y_\tau)/2M,$$

where the Jacobian $J=x_\sigma y_\tau-x_\tau y_\sigma=2My_\sigma y_\tau$ due to (1.6). From (1.4) and (1.10) we find $u_\tau/y_\tau\to$ const. σ^{-1}, as $\tau\to 0$, $\tau>0$.

Hence the second derivatives of φ are continuous on the forward characteristic, despite the discontinuity on the backward Mach line. This statement is valid, however, only when $J\neq 0$ between those two Mach lines issued from P, for otherwise the function $\varphi(x,y)$ cannot be formed. Thus the final result is an alternative which has to be tested in individual cases.

§1.3 The same consideration holds for the general equation (1) if $B(\varphi_x^0,0)=0$, $C(\varphi_x^0,0)=k(\varphi_x^0,0)\neq 0$. When, however, the constants do not satisfy these conditions, then the quantitative results will be different from the case treated.

A convergence scheme is obtained by adding supplementary terms on the right hand side of (1.3): $E_1=f_1$, $E_2=f_2$, and using (1.8) instead of E_3. Each time a higher approximation of U and V is obtained by the solution of linear differential equations $E_1=f_1$, $E_2=f_2$, with known terms in f_1 and f_2, the result is to be inserted into (1.8) to find y.

§2.1 THEOREM 1 (*Existence*). *Given in the x, y-plane a simple closed convex curve C with at most a finite number of protruding corners, and given any positive constant β, there exists a minimal surface $z=z(x,y)$ in the exterior region of C, satisfying the following two conditions*: 1) *at infinity*, $z=\beta x+O(1/\sqrt{x^2+y^2})$, *and* 2) $\partial z/\partial n=0$ *on C except at the corners.*

There are two stagnation points on C where $z_x=z_y=0$. At the vertex of a corner which is not a stagnation point, the function $z(x,y)$ has a finite jump discontinuity and there $z_x^2+z_y^2\to\infty$.

The solution is unique, when C is symmetric with respect to the x-axis (this axis is distinguished by the condition at infinity). The symmetric case

is easier to treat, because here the location of the stagnation points is known.

Our proof of Theorem 1 consists of two sections:

(I) Formulation of the problem in the parametric form and existence proof by variational method.[1]

(II) Proof that every solution obtained by (I) is capable of non-parametric representation, $z = z(x, y)$.

Remark: Because of the unboundedness of the first partial derivatives of $z(x, y)$, the differential equation (2) is not uniformly elliptic in the region exterior to C. For the problem with prescribed "local maximum speed" on the boundary curve and regularity at infinity (without assigning β) see Bers [1].

The following is an outline of the proof.

(I) Let Γ be a simple closed Jordan curve in the x, y-plane. The problem is formulated and solved for Γ, instead of restricting it to C. The domain of the parameter $w = u+iv = re^{i\theta}$ is the unit disk, and the surfaces are represented by harmonic functions

$$(2.1)\quad x = x(w) = \mathscr{R}e\,\varphi(w), \quad y = y(w) = \mathscr{R}e\,\psi(w), \quad z = z(w) = \mathscr{R}e\,\chi(w)$$

with the origin $w = 0$ corresponding to infinity in the x, y-plane. We require that the functions in (2.1) have simple poles at $w = 0$,

$$\varphi(w) = a(1+\beta^2)^{-\frac{1}{2}}w^{-1}+\cdots, \qquad \psi(w) = -ia\,w^{-1}+\cdots,$$
$$\chi(w) = a\,\beta(1+\beta^2)^{-\frac{1}{2}}w^{-1}+\cdots,$$

so that the residues are determined up to a common real factor a. Furthermore, we replace the condition $\partial z/\partial n = 0$ on the boundary curve by the "generalized" condition $\partial z/\partial r = 0$, which has a meaning for any Γ. An immediate consequence of this is

$$(2.2)\qquad z = z(w) = \mathscr{R}e[a\,\beta(1+\beta^2)^{-\frac{1}{2}}(w^{-1}+w)].$$

Hence $z(w)$ is determined except for the constant factor a.

The admissible functions $x(w)$, $y(w)$ for the variational problem

$$(2.3)\qquad \int_{r\to 1}\left(x\frac{\partial x}{\partial r}+y\frac{\partial y}{\partial r}\right)r\,d\theta = \min = m(\Gamma)$$

are those which satisfy the following two conditions:

1) The boundary mapping $\mathfrak{M}$: $x = x(e^{i\theta})$, $y = y(e^{i\theta})$ of the circle to Γ, is continuous and monotonic (and reverses the orientation of the plane). 2) For each pair $x(w)$, $y(w)$ the constant a is determined by the requirement that, in the series expansion of $w^2((\partial\varphi/\partial w)^2+(\partial\psi/\partial w)^2+(\partial\chi/\partial w)^2)$ at $w=0$,

[1] This part of the proof is carried out in an earlier paper [5] by the author.

the constant term vanishes. (Note that the first term in the expansion is of the power w^{-2}, but its coefficient is zero, because of the assumed behavior of φ, ψ and χ at $w = 0$.) By the mean value theorem of harmonic functions it is easily verified that a can be expressed by the boundary values of x and y: $a = (1/2\pi)(1+\beta^2)^{1/2} \int_{r=1} [x\, dv + (1+\beta^2)^{1/2} y\, du]$.

The minimum problem has at least one solution, and this solution is a minimal surface because of the first variation. The proof of these statements follows Courant's treatment of Plateau's problem [3].

(II) We consider any minimal surface solution[2] of the form (2.1) and (2.2) with boundary mapping $\mathfrak{M}$. The function

$$\zeta(w) = (-d\varphi + i d\psi)/d\chi = d\chi/(d\varphi + i d\psi) \tag{2.4}$$

is introduced. Note that $d\chi = 0$ at $w = \pm 1$ and nowhere else. It is then easily verified that $\zeta(0) = 1/[1+(1+\beta^2)^{1/2}]$, $\zeta'(0) = 0$ and $\zeta(w) \neq 0, \infty$ for $|w| < 1$.

LEMMA 1. *The non-vanishing of the Jacobian* $\partial(x, y)/\partial(u, v)$ in $|w| < 1$ *is equivalent to* $|\zeta(w)| < 1$ *for* $|w| < 1$.

LEMMA 2. *Let* Γ *be a polygon* C_p. *The boundary mapping cannot be a one-to-one point transformation. There exist stationary arcs on* $|w| = 1$ *which are mapped into single points on* C_p. *Along each stationary arc* $|\zeta(w)| = 1$.

LEMMA 3. *Let* Γ *be a convex polygon* C_k. *Then the following holds*: 1) $\zeta(\pm 1) = 0$, 2) *along each stationary arc,* arg ζ *is monotonically increasing,* 3) *if* $w = w_1$ *is an endpoint of a stationary arc,* $\zeta(w_1) = e^{-i\alpha}$, *then* $e^{i\alpha}\zeta - 1 = (w-w_1)^{1/2} P(w-w_1) + Q(w-w_1)$ *where* P *and* Q *are power series with* $P(0) \neq 0$, $Q(0) = 0$.

This is the main lemma. In the proof we make use of the level lines of the harmonic functions $px(w) + qy(w) \pm z^*(w)$, where $p+iq$ is any constant with $|p+iq| = 1$, and $z^*(w)$ is the harmonic conjugate of $z(w)$. Both the convexity of C_k and the behavior of these harmonic functions at $w = 0$ are essential for the proof. An obvious consequence of Lemma 3 is

LEMMA 4. $|\zeta(w)| < 1$ *for* $|w| < 1$.

From this lemma it follows by the argument principle that a stationary arc is mapped by $\mathfrak{M}$ into a vertex, and, conversely, any vertex which does not correspond to $w = \pm 1$, has a stationary arc as its antecedent. Since $z(w)$ varies on such an arc, we conclude the discontinuity of $z(x, y)$ at the vertex. For analytic behavior of $z(x, y)$ in the neighborhood of a corner, see [6].

---

[2] It is immaterial whether or not the solution is obtained from the minimum problem, so far as Lemmas 1–4 are concerned.

LEMMA 5. *Let C_k' be a convex polygon lying inside of C_k, then $m(C_k') > m(C_k)$, where m denotes the minimum value for the variational integral belonging to the curve.*

By virtue of the monotonicity of $m(C_k)$, the results obtained for polygons are extended to convex curves. Lemma 5 is proved by using the principle of subordination of conformal mappings.

Remark: In the case of symmetrical C_k, Theorem 1 reduces the problem of finding explicit solutions to that of a conformal mapping with a finite number of parameters.

§2.2 THEOREM 2 (*Non-existence*). *Given in the x, y-plane a polygon C_p and given any positive constant β, there exists no minimal surface solution $z = z(x, y)$ with* 1) $z = \beta x + O(1/\sqrt{x^2+y^2})$ *at* ∞ *and* 2) $z = 0$ *on* C_p.

The proof of this theorem is simple. Consider the solution $x'(w)$, $y'(w)$, $z'(w)$ of this problem in the parametric form. The boundary mapping $\mathfrak{M}'$ of $|w| = 1$ to C_p, given by $x = x'(e^{i\theta})$, $y = y'(e^{i\theta})$, must be a one-to-one point transformation. Since $\zeta(w)$, which is formed in the same way from x', y', z' as in (2.4), must become infinite at one of the vertices, it follows that the Jacobian $\partial(x', y')/\partial(u, v)$ cannot remain of the same sign in $|w| < 1$.

§2.3 In connection with Theorems 1 and 2 we would like to mention two open problems for minimal surfaces.

A. Let the minimal surface S be given by a solution $x(w)$, $y(w)$, $z(w)$ of Theorem 1, with the convex polygon C_k as boundary. Suppose C_k has N sides $t_1, t_2, \cdots, t_N$ and vertices $P_1, P_2, \cdots, P_N$. We consider the surface S^* which is conjugate to S. It is given by $x^*(w)$, $y^*(w)$, $z^*(w)$ where $\varphi = x + ix^*$, $\psi = y + iy^*$, $\chi = z + iz^*$. It is clear that S^* and S have the same $\zeta(w)$, hence S^* can be represented in the non-parametric form. Let the boundary of S^* be denoted by B^*. It is easily verified that 1) part of B^* consists of N straight line segments, $t_1', t_2', \cdots, t_N'$, 2) the remaining part of B^* consists of N convex arcs, $p_1', p_2', \cdots, p_N'$, such that $t_j' \perp t_j$ and along p_j', $|\zeta(w)| = 1$. The consecutive segments t_j' and t_{j+1}' have no point in common and are connected by p_j', unless p_j' reduces to a point, which case arises when and only when P_j is a stagnation point. Since C_k is convex, we may extend the segments t_j' so that they form a convex polygon C_k^*, which then contains the arcs p_j' in its interior.

Now, according to Theorem 2, the problem with given $\beta > 0$ and $z^* = 0$ on C_k^* has no non-parametric solution. Indeed, S^* satisfies the boundary condition only partially, it by-passes the corners by producing "free arcs" p_j'. There is some similarity between these facts and the occurence of cavities of flow around objects, except that, in the latter case, the free streamlines are outside of the body.

We may propose the following problem:

Q1. Let C^* be a simple closed convex curve, and let β be any positive constant. There exists a minimal surface $z = z(x, y)$ with: 1) $z = \beta x + O(1/\sqrt{x^2+y^2})$ at infinity. 2) $z = 0$ on B^*, which consists partly of arcs of C^* and partly of "free arcs"; each free arc is convex, is contained inside C^* and along it $|\zeta(w)| = 1$.[3] 3) $z(x, y)$ is defined in the exterior of B^*. 4) There are two stagnation points which lie on the non-free arcs of B^*. 5) The solution is unique.

The above statements hold true, when C^* is a triangle which is symmetric with respect to the y-axis. Indeed, in this case the boundary polygon C_k^* determines, by pure geometry, the corresponding polygon C_k of Theorem 1.

B. Another open problem is concerned with minimal surfaces with discontinuous first partial derivatives. First, let us reformulate Theorem 1 as a free boundary problem of minimal surfaces. Denote by $\overline{C}$ the cylinder which is produced by the displacement of C in the direction parallel to the z-axis. A solution of Theorem 1 is a minimal surface which lies in the exterior of $\overline{C}$, is regular at infinity and has a free boundary on $\overline{C}$. When C has N vertices, then $\overline{C}$ has N side faces and N edges. The minimal surface is transversal to each side face of $\overline{C}$, but in crossing from one side to the neighboring side, the surface will take a part of the common edge as free boundary. The surface is perpendicular to the edge when and only when the corresponding vertex of C is a stagnation point. The length of the free boundary segment along each edge is the amount of the jump discontinuity of $z(x, y)$ at the vertex.

Q2. Let S_1 and S_2 be two analytic surfaces which intersect along a curve g. We assume first that along g the angles formed by S_1 and S_2 are greater or equal $\pi/2$. Let γ be a Jordan curve which connects a point of S_1 to a point of S_2 without crossing either S_1 or S_2. Prove that there exists a minimal surface which is spanned by γ and a free boundary on the given surfaces and lies on the same side of S_1 and S_2 as γ does. The free boundary will contain in general a part of g. If however, the angles formed by S_1 and S_2 along g are less than $\pi/2$, then the minimal surface will be perpendicular to g wherever it crosses g.

Of particular interest is the case when S_1 and S_2 are themselves minimal surfaces and the points P_1 and P_2 lie on g. Better knowledge of the behavior of the free boundary in this case will make it possible to solve problems involving interface discontinuities of minimal surfaces. Lewy's theorem on the analyticity of trace is of fundamental importance to these questions.

[3] This means geometrically that the minimal surface is transversal to the x, y-plane along each free arc.

Bibliography

[1] Bers, L., *Boundary value problems for minimal surfaces with singularities at infinity*, Trans Amer. Math Soc., Vol. 70, 1951, pp 465–491

[2] Courant, R., and Hilbert, D, *Methoden der mathematischen Physik*, Vol. II, Springer, Berlin, 1937, Ch VII.

[3] Courant, R., *Dirichlet's Principle, Conformal Mapping and Minimal Surfaces*, Interscience, New York, 1950, Ch. VI.

[4] Chen, Y. W, *Supersonic flow through nozzles with rotational symmetry*, Comm. Pure Appl Math., Vol 5, 1952, pp. 57-86. *Flows through nozzles and related problems of cylindrical and spherical waves*, Ibid., Vol. 6, 1953, pp 179–229.

[5] Chen, Y. W, *Existence of minimal surfaces*, Trans Amer Math Soc, Vol 65, 1949, pp 338 347

[6] Chen, Y W, *Discontinuity and representations of minimal surface solutions*, Proc Conference on Differential Equations, University of Maryland (to appear)

[7] Goldstein, S., *Selected problems in gas dynamics*, International Congress of Mathematicians Cambridge, Vol 2, 1950, pp 280-291.

[8] Lighthill, M. J., *Mathematical methods in compressible flow theory*, Comm. Pure Appl Math Vol. 7, 1954, pp 1–10

[9] Lewy, H., *On minimal surfaces with partially free boundary*, Comm Pure Appl Math., Vol 4, 1951, pp 1–13.

[10] Meyer, R E., *The method of characteristics for problems of compressible flow involving two independent variables, I, II*, Quart J Mech Appl Math, Vol 1, 1948. pp. 196-219, 451–469. *Focusing effects in two-dimensional supersonic flow*, Philos Trans Roy. Soc London, Ser. A, Vol. 242, 1949, pp 153–171

[11] Ward, G. N., *The approximate external and internal flow past a quasi-cylindrical tube moving at supersonic speeds*, Quart. J. Mech Appl. Math, Vol 1, 1948, pp. 225–245.

[12] Whitham, G B, *The flow pattern of a supersonic projectile*, Comm. Pure Appl. Math., Vol. 5, 1952, pp. 301–348

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 383–390 (1956)

On Singular and Regular Cauchy Problems

J. B. DIAZ*

Institute for Fluid Dynamics and Applied Mathematics
University of Maryland

1. Singular Cauchy Problem for the Euler-Poisson-Darboux Equation

Consider the partial differential equation

$$\Delta u = u_{tt} + \frac{k}{t} u_t, \tag{1}$$

where $u = u(x_1, \cdots, x_m, t)$, the differential operator $\Delta = \partial^2/\partial x_1^2 + \cdots + \partial^2/\partial x_m^2$ is the Laplacian, and k is a parameter. This partial differential equation, which has been termed the Euler-Poisson-Darboux equation, has occurred, for special values of k and m, in many important and classical problems since the time of Euler. An exposition of the theory of Euler and Poisson (for an equation, with $m = 1$, which can be obtained from (1) by a change of variables) is to be found in Darboux [1].

When $k = 0$ in (1) this equation reduces to the wave equation, while if $k \neq 0$ then the coefficient k/t is infinite on the plane $t = 0$. "The" singular Cauchy problem for (1) consists in the determination of a solution of (1), for $t > 0$, meeting the following initial conditions on the "singular" plane $t = 0$:

$$u(x_1, \cdots, x_m, 0) = g(x_1, \cdots, x_m), \qquad u_t(x_1, \cdots, x_m, 0) = 0, \tag{2}$$

where g is a given function. For k any real number, this singular Cauchy problem was first solved by A. Weinstein [7], who employed what he termed the "method of recurrence" and a generalized method of descent. Let, for the time being, the solution of the Cauchy problem (1), (2) be denoted by u^k. Weinstein shows that for $k = -1, -3, \cdots$ the Cauchy problem for u^k can be reduced to a Cauchy problem for u^{k+2} by means of the following recurrence formulas:

$$u^{k+2}(x, t) = \frac{u_t^k(x, t)}{t}, \qquad u^{k+2}(x, 0) = \frac{\Delta u^k(x, 0)}{k+1}, \qquad u_t^{k+2}(x, 0) = 0, \tag{3}$$

*The research for this paper was supported by the United States Air Force through the Office of Scientific Research.

where x stands for $x_1, \cdots, x_m$. Moreover, for arbitrary solutions u^k of (1) there is the correspondence principle

$$u^k = t^{1-k} u^{2-k}, \tag{4}$$

which enables one to obtain from a solution u^k of (1) a solution of the same equation with parameter value $2-k$. Formula (4) is used in discussing the uniqueness of the solution. Weinstein first solves explicitly the Cauchy problem (1), (2) for $k \geqq m-1$, and then, by a process which consists essentially in a repeated application of (3), obtains a formula for a solution when $k < m-1$.

It is well known that for $k = m-1$ the Cauchy problem (1), (2) possesses a unique solution, which is given by a generalization of a formula of Poisson (see Ásgeirsson [2] and Courant-Hilbert [3]):

$$u(x_i, t) = \frac{1}{\omega_m} \int_{\sum_{j=1}^{m} \beta_j^2 = 1} g(x_i + t\beta_i) d\omega_m, \tag{5}$$

where $d\omega_m$ is the surface element of the unit sphere in m-dimensional Euclidean space and

$$\omega_m = \frac{2\pi^{m/2}}{\Gamma\left(\frac{m}{2}\right)}$$

is the surface area of this sphere. Using this formula and Hadamard's "method of descent" [6], an explicit formula for the solution of (1), (2)

$$u^k(x, t) = \frac{\omega_{k+1-m}}{\omega_{k+1}} \int_{\sum_{j=1}^{m} \beta_j^2 \leq 1} g(x_1+\beta_1 t, \cdots, x_m+\beta_m t)(1-\beta_1^2-\cdots-\beta_m^2)^{\frac{k-m-1}{2}} d\beta_1 \cdots d\beta_m, \tag{6}$$

may be obtained when k is any integer of the sequence $m, m+1, \cdots$, and this formula for u^k is then readily verified to be the solution of (1), (2) for any real $k > m-1$. This procedure may be called a generalized method of descent. By means of this explicit formula (6) and the recurrence relation (3), Weinstein obtains an explicit formula for a solution u^k of (1), (2) for all values of $k \neq -1, -3, -5, \cdots$. J. B. Diaz and H. F. Weinberger [8] obtained the solution of (1), (2), using again as a starting point the same explicit formula (6), for the solution for $k > m-1$, which was employed by Weinstein. The definite integral in (6) diverges for $k < m-1$; their method consists in finding the analytic continuation of this integral for $k < m-1$. Their final formula is again valid for $k \neq -1, -3, \cdots$.

The odd negative integer values of the parameter k play an exceptional

role in the solution of the Cauchy problem (1), (2). A solution of the Cauchy problem for these exceptional values was given by Diaz and Weinberger [8], after Weinstein [7] had pointed out the exceptional nature of the polyharmonic initial values in this problem. Their result is that if $g(x_1, \cdots, x_m)$ has continuous partial derivatives of order not less than the maximum of the two numbers $\frac{1}{2}(m+3-k)$ and $3-k$, then there exists a solution of (1), (2). The t-derivative of this solution of order $1-k$ becomes infinite like $\log t$ at $t = 0$ unless $g(x_1, \cdots, x_m)$ is polyharmonic of order $\frac{1}{2}(1-k)$, in which case the solution is just a polynomial in t. (Our attention was drawn to this behavior by the example $u = x^2+t^2 \log t$ for $k = -1$, $m = 1$, kindly communicated by Professor B. Friedman in a letter.) It should be noticed that for these exceptional values $k = -1, -3, -5, \cdots$, the solution does not depend continuously on the initial data, inasmuch as the slightest (no matter how smooth) deviation from the proper polyharmonic initial values produces a logarithmic infinity in t in certain derivatives of the solution. Another solution of the Cauchy problem (1), (2) by a different method and under weaker differentiability assumptions, has been given recently by E. K. Blum [9], who has also clarified the question of what is the "general solution" of (1) when k is a negative number.

2. Singular Cauchy Problem for a Generalization of the Euler-Poisson-Darboux Equation in Two Space Variables

The singular Cauchy problem in question consists in the determination of a function $u(x, y, t)$ which satisfies the following generalization of the Euler-Poisson-Darboux equation in two space variables:

$$u_{xx}+u_{yy}-u_{tt}-\frac{k}{t}u_t-h(x, y, t)u = 0, \qquad t > 0, \quad k > 0, \tag{7}$$

and the initial conditions

$$u(x, y, 0) = g(x, y), \qquad u_t(x, y, 0) = 0 \tag{8}$$

on the singular plane $t = 0$, where $h(x, y, t)$ and $g(x, y)$ are given functions. J. B. Diaz and G. S. S. Ludford [13] have shown that if $h(x, y, t)$ is continuous for $t \geqq 0$ and all real x, y, together with its "space" derivatives h_x, h_y, h_{xx}, h_{xy}, h_{yy}, while $g(x, y)$, which is supposed to be defined for all real x, y, is a twice continuously differentiable function (if $k \geqq 1$) or a thrice continuously differentiable function (if $0 < k < 1$), then there exists one and only one function $u(x, y, t)$ which is continuous for $t \geqq 0$ and all real x, y, together with its space derivatives u_x, u_y, u_{xx}, u_{xy}, u_{yy}, is twice continuously differentiable in x, y, t for $t > 0$, and solves the singular Cauchy problem (7), (8). It should be remarked that the analysis in [13] is restricted to the

case of two space variables (rather than m, as in equation (1)) since certain definite integrals appearing in the course of the argument, though they are improper, are still convergent if $m = 2$ but do not converge if $m > 2$. For $m > 2$ use would have to be made of Hadamard's concept ([6] p. 133) of the "finite part" of a divergent integral. The restriction $k > 0$ is motivated by the fact that already for the singular Cauchy problem (1), (2) the solution is not unique when $k < 0$ (see [7], [8]).

The approach in [13] consists in first finding an elementary solution in the sense of Hadamard [6] for the Euler-Poisson-Darboux equation (1) (the starting point is a formula for a fundamental solution for a corresponding elliptic equation, which was given by Diaz and Weinstein [10]). For the special case of two space variables in (1), one has that if ξ and η are real numbers, $\tau > 0$, and

$$r^2 = (t-\tau)^2-(x-\xi)^2-(y-\eta)^2, \quad r'^2 = (t+\tau)^2-(x-\xi)^2-(y-\eta)^2,$$

then the function

$$V(\xi, \eta, \tau; x, y, t) = \frac{2^k t^k}{r'^k r} F\left(\frac{k-1}{2}, \frac{k}{2}, \frac{1}{2}, \frac{r^2}{r'^2}\right), \tag{9}$$

where F is the hypergeometric function, is an elementary solution of equation (1) for $k > 0$ and $m = 2$. This elementary solution is then used to construct (cf. [6] p. 166) the solution of the following *regular* Cauchy problem:

$$\begin{aligned} &w_{xx}+w_{yy}-w_{tt}-\frac{k}{t}w_t = 0, \qquad t > t_0 > 0, \quad k > 0, \\ &w(x, y, t_0) = 0, \quad \frac{\partial w}{\partial t}(x, y, t_0) = F(x, y), \end{aligned} \tag{10}$$

where $F(x, y)$ is a given twice continuously differentiable function and t_0 is a given positive number. In the notation of (9) above, the solution w is furnished, for $t > t_0$, by the explicit formula

$$w(x, y, t) = \frac{1}{2\pi}\iint V(\xi, \eta, t_0 ; x, y, t)F(\xi, \eta)d\xi d\eta, \tag{11}$$

where the integration is to be performed over the circular disk $(\xi-x)^2+(\eta-y)^2 < (t-t_0)^2$. The next step is to construct the solution of the following *singular* Cauchy problem for the *non-homogeneous* Euler-Poisson-Darboux equation in two space variables:

$$\begin{aligned} &u_{xx}+u_{yy}-u_{tt}-\frac{k}{t}u_t = f(x, y, t), \qquad t > 0, \quad k > 0, \\ &u(x, y, 0) = 0, \quad \frac{\partial u}{\partial t}(x, y, 0) = 0, \end{aligned} \tag{12}$$

where $f(x, y, t)$ is a given function (such that $f, f_x, f_y, f_{xx}, f_{xy}, f_{yy}$ are continuous in the half space $t \geqq 0$). The solution of the *singular* Cauchy problem for a *non-homogeneous* equation (12) is given, in terms of the solution of the *regular* Cauchy problem for the corresponding *homogeneous* equation (10), by means of a Duhamel integral (cf. [3] p. 402; this is also known as the method of variation of parameters or "Stossmethode"). (The important difference between the usual application of Duhamel integrals to *regular* Cauchy problems, e.g. for the wave equation, is that there these integrals are proper integrals, whereas in the *singular* Cauchy problem under consideration here, the Duhamel integrals are improper, but convergent.) Thus, given the function $f(x, y, t)$ occurring in (12), let $w(x, y, t; t_0)$ denote, for each number $t_0 > 0$, the corresponding solution of (10) when $F(x, y)$ is replaced in the initial conditions of (10) by $f(x, y, t_0)$. Then the solution u of (12) is given explicitly, for $t > 0$, by the following Duhamel integral

$$u(x, y, t) = \int_0^t w(x, y, t; t_0)\, dt_0. \tag{13}$$

Once this explicit solution of (12) has been obtained, the original Cauchy problem (7), (8) may be solved by the Picard method of successive approximations, employing certain estimates for the Duhamel integral (13) and its space derivatives. The particular difficulty here is that the definite integrals involved are improper but convergent, and an important part in obtaining the required estimates is played by the behavior (cf. equation (9)) of the hypergeometric function $F(a, b, c, z)$ near $z = 1$. Still another important fact which is employed is the following (consider only the case $k \geqq m-1$, for definiteness): if the initial value function g occurring in (2) possesses continuous second order partial derivatives with respect to $x_1, \cdots, x_m$ then the corresponding solution of (1) has continuous second order partial derivatives with respect to $x_1, \cdots, x_m, t$ for *all* t (and not just for $t > 0$), and satisfies (1) for $t \lesseqgtr 0$ and its "limiting form"

$$\Delta u = (k+1)u_{tt}, \tag{14}$$

on the plane $t = 0$, the carrier of the Cauchy data. The Cauchy problem (1), (2), for *all* values of the time, has been considered by Diaz and Ludford [14].

3. A Regular Cauchy Problem for the Euler-Poisson-Darboux Equation

In a 1955 doctoral dissertation, written under the direction of Professor Weinstein at the University of Maryland, Miss R. M. Davis [12] has given an explicit formula for the solution of the following *regular* Cauchy problem for the Euler-Poisson-Darboux equation:

$$\Delta u = u_{tt} + \frac{k}{t} u_t, \qquad t > 0,$$

(15)

$$u(x_1, \cdots, x_m, t_0) = g(x_1, \cdots, x_m), \qquad \frac{\partial u}{\partial t}(x_1, \cdots, x_m, t_0) = 0,$$

where g is a given function and $t_0 > 0$. The carrier of the Cauchy data is the (regular) plane $t = t_0$. Her approach is based on a modification of the method of Marcel Riesz [5] for solving regular Cauchy problems for self-adjoint equations. A modification of the method of Riesz is needed because the differential equation in (15) is not self-adjoint. While, in principle, the extension of Riesz's method to not necessarily self-adjoint equations may be considered as a special case of Malmheden's [4] results concerning systems of equations, the main interest in Davis's thesis lies in that she has actually determined explicitly (for the first time) the "elementary solution in the sense of M. Riesz", V^α, for a nontrivial non-self-adjoint equation, and carried out the requisite analytic continuation in order to obtain an explicit formula for the solution of the Cauchy problem (15). The function ("of two points plus a parameter α")

$$V^\alpha(\xi_1, \cdots, \xi_m, \tau; x_1, \cdots, x_m, t)$$

determined by Davis satisfies the following conditions:

(16)

$$L_{(\xi,\tau)} V^{\alpha+2}(\xi, \tau; x, t) = V^\alpha(\xi, \tau; x, t),$$
$$M_{(x,t)} V^{\alpha+2}(\xi, \tau; x, t) = V^\alpha(\xi, \tau; x, t),$$

where ξ and x are abbreviations for $\xi_1, \cdots, \xi_m$ and $x_1, \cdots, x_m$, respectively, the operator (see (1))

$$L_{(\xi,\tau)} \equiv \sum_{i=1}^{m} \frac{\partial^2}{\partial \xi_i^2} - \frac{\partial^2}{\partial \tau^2} - \frac{k}{\tau} \frac{\partial}{\partial \tau},$$

and M is the formal Lagrange adjoint operator of L. In the present notation, her formula for V^α is

(17)

$$V^\alpha(\xi, \tau; x, t) = \frac{t^k r^{\alpha-m-1}}{2^{\alpha-k-1} \pi^{(m-1)/2} \Gamma\left(\frac{\alpha}{2}\right) \Gamma\left(\frac{\alpha+1-m}{2}\right)} F\left(\frac{\alpha+k-m-1}{2}, \frac{k}{2}, \frac{\alpha+1-m}{2}, \frac{r^2}{r'^2}\right),$$

where

$$r^2 = (t-\tau)^2 - \sum_{i=1}^{m} (x_i - \xi_i)^2, \qquad r'^2 = (t+\tau)^2 - \sum_{i=1}^{m} (x_i - \xi_i)^2.$$

It should be noticed that if in (17) one sets $m = 2$ and $\alpha = 2$ then V^α reduces to a constant multiple of the Hadamard elementary solution (9)

which was considered in Section 2. For the general relation between the Riesz and the Hadamard elementary solutions, of which the verification just mentioned is only a special case, see [5] pp. 182—184.

It is of interest to inquire whether the explicit formula for the solution of the *regular* Cauchy problem (15) with initial data given in the regular plane $t = t_0 > 0$ "converges" as $t_0 \to 0$ to a solution of the *singular* Cauchy problem (1), (2) with initial data given on the singular plane $t = 0$. While it is likely that this will be the case provided that $k > m - 1$, and it is perhaps true for $0 < k < m-1$, it is certainly not true in general for $k < 0$, even for k not an "exceptional value" (i.e. an odd negative integer) in view of the following simple example, which was kindly communicated by Miss Davis. (Of course, when $k = 0$ there is no doubt.) Consider $m = 1$ and $k = -2$ in (15); then the function

$$u(x, t) = x^2 - t^2 + \frac{2t^3}{3t_0} + \frac{t_0^2}{3}$$

is the unique solution of the Cauchy problem (15) with initial data

$$u(x, t_0) = x^2, \qquad \frac{\partial u}{\partial t}(x, t_0) = 0,$$

and yet the function u does not approach a solution of the corresponding singular initial value problem (1), (2) with initial data

$$u(x, 0) = x^2, \qquad \frac{\partial u}{\partial t}(x, 0) = 0.$$

Bibliography

[1] Darboux, G., *Leçons sur la Théorie Générale des Surfaces*, Second Edit., Vol II, Gauthier Villars, Paris, 1914—1915, pp. 54 ff.

[2] Ásgeirsson, L., *Über eine Mittelwerteigenschaft von Losungen homogener linearer partieller Differentialgleichungen zweiter Ordnung mit konstanten Koeffizienten*, Math. Ann., Vol. 113, 1937, pp. 321–346, spec. pp 326–327.

[3] Courant, R., and Hilbert, D., *Methoden der mathematischen Physik*, Vol. II, Springer, Berlin, 1937

[4] Malmheden, H., *A class of hyperbolic systems of linear differential equations*, Comm. Sém. Math Univ. Lund, Vol. 8, 1947, pp. 1–116.

[5] Riesz, M., *L'intégrale de Riemann-Liouville et le problème de Cauchy*, Acta Math., Vol. 81, 1949, pp. 1–223.

[6] Hadamard, J., *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Dover, New York, 1952

[7] Weinstein, A., *Sur le problème de Cauchy pour l'équation de Poisson et l'équation des ondes*, C. R. Acad. Sci. Paris, Vol. 234, 1952, pp. 2584–2585.

[8] Diaz, J. B., and Weinberger, H F., *A solution of the singular initial value problem for the Euler-Poisson-Darboux equation*, Proc. Amer. Math. Soc., Vol. 4, 1953, pp. 703–718.

[9] Blum, E. K., *The Euler-Poisson-Darboux equation in the exceptional cases*, Proc. Amer. Math. Soc., Vol. 5, 1954, pp. 511–520.

[10] Diaz, J. B., and Weinstein, A., *On the fundamental solutions of the singular Beltrami operator*, in *Studies in Mathematics and Mechanics presented to Richard von Mises*, Academic Press, New York, 1954, pp. 97–102.

[11] Weinstein, A., *On the wave equation and the equation of Euler-Poisson*, in *Proceedings of Symposia in Applied Mathematics*, Vol. V, *Wave Motion and Vibration Theory*, McGraw-Hill, New York, 1954, pp. 137–148.

[12] Davis, R. M., *On a regular Cauchy problem for the Euler-Poisson-Darboux equation*, Ph.D. Thesis, University of Maryland, June 1955.

[13] Diaz, J. B , and Ludford, G. S S , *On the singular Cauchy problem for a generalization of the Euler-Poisson-Darboux equation in two space variables*, Ann. Mat , Ser. IV, Vol. 38, 1955, pp. 33–50.

[14] Diaz, J. B., and Ludford, G. S. S., *The Euler-Poisson-Darboux equations and the method of descent*, Proceedings of the Conference on Differential Equations, University of Maryland, March 17–19, 1955 (to appear).

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 391–402 (1956)

A Criterion for the Validity of Huygens' Principle*

AVRON DOUGLIS

In this paper, we shall supply a proof of an analytical criterion for the validity of Huygens' Principle among linear, hyperbolic equations of second order in four independent variables. Our argument is based on an apparently new procedure of obtaining an integral formula for the solution of Cauchy's problem; the procedure stems from the integral equation theory of hyperbolic equations and succeeds in avoiding some troublesome questions that arise in the standard methods of Hadamard and of M. Riesz.

Cauchy's problem for a linear, hyperbolic[1] partial differential equation of second order

$$(1) \qquad Mu \equiv \sum_{i,j=0}^{n} \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i}\left(\sqrt{g}\, g^{ij} \frac{\partial u}{\partial x^j}\right) + \sum_{i=0}^{n} a^i \frac{\partial u}{\partial x^i} + ku = \psi$$

is the problem of determining a solution u with prescribed values and prescribed normal derivatives on a fixed initial manifold. The prescribed quantities are called the "Cauchy data".[2] "Huygens' Principle" is a rule of behavior of the solutions of problems of Cauchy for Lorentz's equation

$$(1a) \qquad \frac{\partial^2 u}{\partial x_0^2} - \sum_{1}^{n} \frac{\partial^2 u}{\partial x_p^2} = 0,$$

valid if n is odd and > 1, which states that the value of u at any point Q is determined by the data on the intersection of the initial manifold with the characteristic cone[3] issuing from Q. For a hyperbolic equation of second

---

*This paper represents results obtained under the auspices of the Office of Naval Research, Contract N6ori-201, T.O.I, and the National Science Foundation, Contract NSF-G1217.

[1]Equation (1) is *hyperbolic* in a region R if, at each point P of R, the quadratic form $\sum_{i,j=0}^{n} g^{ij}(P)t_i t_j$ in indeterminates t_i is of normal hyperbolic signature. We shall assume, more specifically, that the form can be reduced to a difference which consists of a single square minus three other squares.

[2]More generally, the term "Cauchy data" will cover not just the prescribed values of u and of its normal derivative but also the various derivatives of these quantities.

[3]The cones $(x_0-\xi_0)^2 - \sum_{p=1}^{n} (x_p-\xi_p)^2 = 0$ ($\xi_1 = \text{const.}$) are the characteristic cones for Lorentz's equation. More generally, the characteristic surfaces $\phi(x^i) = 0$ for a hyperbolic equation (1) are the surfaces on which the condition $\sum_{i,j=0}^{n} g^{ij} \frac{\partial \phi}{\partial x^i} \frac{\partial \phi}{\partial x^j} = 0$ is satisfied, a *characteristic conoid* being a characteristic surface of conoidal type.

order with variable coefficients or with lower order terms, Huygens' Principle is said to hold, if the value of a solution at any point Q is determined by the data on the intersection of any space-like[4] initial manifold with the characteristic conoid[5] that issues from Q. The question of principal interest concerning Huygens' Principle in this extended sense is the following, first raised by Hadamard [5, p. 236]. Does the validity of Huygens' Principle for the solutions of a given hyperbolic equation imply that this equation can be reduced to the same form as Lorentz's equation?[6]

This paper is concerned with the antecedent problem of translating Huygens' Principle into quantitative terms. Hadamard, indeed, has solved this problem in one way [5, p. 236], but Hadamard's precise condition is difficult to interpret, and alternative quantitative formulations of Huygens' Principle would seem to be desirable. We shall offer an alternative in Theorem 1. For this purpose, we first observe that any solution of a linear hyperbolic equation[7], say (1), is characterized by an integral equation in terms of its Cauchy data on a space-like initial manifold $\mathscr{S}$.[8] Let $C^Q_{\mathscr{S}}$ be the truncated portion of the characteristic conoid C^Q between its vertex Q and $\mathscr{S}$, $\mathscr{S}^Q$ the portion of $\mathscr{S}$ cut out by the conoid C^Q, $T^Q_{\mathscr{S}}$ the intersection of $\mathscr{S}$ with C^Q, and $D^Q_{\mathscr{S}}$ the region enclosed by $C^Q_{\mathscr{S}}$ and $\mathscr{S}^Q$. The integral equation is of the form

$$u(Q) = f(Q) + \int_{C^Q_{\mathscr{S}}} K(Q, P)\, u(P)\, dS_P\,, \tag{2}$$

where $K(Q, P)$, which depends only on the coefficients in (1), is continuous except possibly at Q and in the case $n = 3$ admits the estimate

$$K(Q, P) = O(\overline{PQ}^{-1}) \text{ as } P \to Q \text{ on } C^Q_{\mathscr{S}}\,,^9 \tag{3}$$

dS_P is the element of area of C^Q at P, and $f(Q)$ is a continuous function

---

[4] A manifold $\phi(x) = 0$ is *space-like* for Lorentz's equation, if $\left(\frac{\partial\phi}{\partial x_0}\right)^2 - \sum_{p=1}^{n}\left(\frac{\partial\phi}{\partial x_p}\right)^2 > 0$; it is space-like for equation (1), if $\sum_{i,j=0}^{n} g^{ij}\frac{\partial\phi}{\partial x^i}\frac{\partial\phi}{\partial x^j} > 0$.

[5] Cf. footnote 3.

[6] The question is significant only for odd n, since Hadamard [5] showed Huygens' Principle cannot possibly be valid, if n is even. The answer to the question has been shown by K. Stellmacher [8] to be in the negative for odd $n \geqq 5$ and by M. Mathisson [7], J. Hadamard [6], and L. Asgeirsson [unpublished manuscript dated 1939] to be in the affirmative for $n = 3$ in the special case of equations with $g^{ij} = \text{const}$. The general case for $n = 3$ still is undecided.

[7] We shall always assume without explicit mention each time that the quantities involved are suitably smooth. It is sufficient for all our conclusions that the g^{ij} in (1) be of class C^4, the a^i of class C^3, and k and f of class C^2, in the region considered. The initial manifold and the Cauchy data should be of class C^3.

[8] Cf. [2] and the bibliographical references cited there.

[9] This behavior of K can always be assured, in view of the estimates in [2, §4], for instance by requiring that $a^i(Q) = 0$ in (1). The latter conditions are easily brought about by changing the dependent variable say to $v = u \exp\{\tfrac{1}{2}g_{ij}(Q)\, a^i(Q)\, x^j\}$.

determined by the Cauchy data just on the surface of intersection $T_{\mathscr{S}}^{Q}$ of C^{Q} with $\mathscr{S}$. Plainly, Huygens' Principle will hold at Q — $u(Q)$ will be determined by the Cauchy data on $T_{\mathscr{S}}^{Q}$ — if

$$K(Q, P) = 0 \text{ for every point } P \text{ of } C_{\mathscr{S}}^{Q}. \tag{4}$$

Conversely, in the case $n = 3$ Huygens' Principle implies (4), as follows from

THEOREM 1. *Consider the set U of solutions of* (1) *which are of class C^2 in the interior of $C_{\mathscr{S}}^{Q}$ and of class C^1 on the boundary of this region. Assume condition* (3) *to hold. If the value at P of each member u of U is preserved, when the Cauchy data for u are altered in any part of $\mathscr{S}$ outside an arbitrarily narrow strip surrounding $T_{\mathscr{S}}^{Q}$, then condition* (4) *is valid.*

By a similar argument, we also shall demonstrate the uniqueness of the integral equation (2) (this fact is unrelated to Huygens' Principle) under conditions which are satisfied in the case $n = 3$. More precisely, we shall prove

THEOREM 2. *Suppose each member u of U satisfies two integral relations* (2) *and*

$$u(Q) = f_1(Q) + \int_{C_{\mathscr{S}}^{Q}} K_1(Q, P)\, u(P)\, dS_P, \tag{2a}$$

where $f(Q)$ and $f_1(Q)$ are determined by the Cauchy data for u on $T_{\mathscr{S}}^{Q}$, and K and K_1 are functions independent of u such that the difference $K-K_1$ is continuous for $P \neq Q$ and admits the estimate

$$K-K_1 = O(\overline{PQ}^{-1}) \quad \text{as} \quad P \to Q \text{ on } C_{\mathscr{S}}^{Q}. \tag{3a}$$

(*Huygens' Principle is not assumed.*) *Then $K \equiv K_1$.*

By contrast with this result, integral equations which, unlike (2), contain integrals over the interior of the characteristic conoid of the form

$$\int_{C_{\mathscr{S}}^{Q}} J(Q, P)\, u(P)\, dV_P,$$

dV_P being the volume element at P, are not unique. Furthermore, the integral equations (2) in the higher dimensional cases $n = 5, 7, \cdots$, in which cases condition (3a) need not hold, also are not unique, as can be seen from the construction described in [2, §4.3].

Theorems 1 and 2 are both proved by the same kind of argument, which we shall now describe in general terms as it applies to Theorem 1. We start with Green's formula

$$\int_{D_{\mathscr{S}}^{Q}} (vMu - uM^*v)\, dV = \int_{\mathscr{S}^{Q}+C_{\mathscr{S}}^{Q}} \left(v\frac{\partial u}{\partial \nu} - u\frac{\partial v}{\partial \nu}\right) dS,$$

where M^* is the formal adjoint to M and $\partial/\partial\nu$ represents differentiation in the direction of the "conormal" to the boundary of $D_{\mathcal{S}}^{Q}$. The "conormal" direction is an outward direction from any non-characteristic surface such as $\mathcal{S}$ but a tangential one to any characteristic surface such as C^Q. Consequently, by integration by parts the integral over $C_{\mathcal{S}}^{Q}$ on the right side of Green's formula can be transformed into an integral of the form $\int_{C_{\mathcal{S}}^{Q}} G[v]u\,dS$ plus an integral over $T_{\mathcal{S}}^{Q}$ plus a quantity evaluated at Q. The last quantity will be zero, if u and v are bounded; the coefficient $G[v]$ is a slightly singular differential expression in v of first order. Now, if it were possible, we should determine v in such a way as to satisfy

$$\text{(5)} \qquad \begin{aligned} &a)\ M^*v = 0 \text{ in } D_{\mathcal{S}}^{Q}, \\ &b)\ G[v] = K \text{ on } C_{\mathcal{S}}^{Q}. \end{aligned}$$

Then comparing the transformed Green's formula with the integral equation (2) we would infer an integral representation for the solutions of (1) of the form

$$\text{(6)} \qquad u(Q) = T_{\mathcal{S}}^{Q}[u] + \int_{\mathcal{S}^Q} \left(v\frac{\partial u}{\partial \nu} - u\frac{\partial v}{\partial \nu} \right) dS,$$

where $T_{\mathcal{S}}^{Q}[u]$ is a linear functional of u determined by the Cauchy data for u on $T_{\mathcal{S}}^{Q}$. In such a representation, the implications of Huygens' Principle would be clearly seen. Indeed, if Huygens' Principle is assumed, i.e., if $u(Q)$ is supposed unaffected by variations in u and $\partial u/\partial\nu$ at the points of $\mathcal{S}^Q$ outside an arbitrarily narrow boundary strip, then $v = \partial v/\partial\nu = 0$ on $\mathcal{S}^Q$ from the fundamental lemma of the calculus of variations applied to (6), and, if Huygens' Principle holds regardless of the particular initial surface, than v and all its first derivatives would have to be zero in $D_{\mathcal{S}}^{Q}$. Hence, $G[v] = 0$ and, finally, $K = 0$ on $C_{\mathcal{S}}^{Q}$, since $G = K$.

The program described is not quite feasible; because of the singularities in the kernel K and the differential expressions $G[v]$, a function v which strictly fulfills requirements (5) is denied us. A function which, in a generalized sense, satisfies these requirements can be obtained, however, by a procedure which involves approximating K by smooth functions and later utilizing estimates of Friedrichs-Lewy type, and this function will suit our purposes well enough.

Theorem 2 was suggested in principle by J. Hadamard in a conversation with the author. Theorem 1 essentially has been stated by M. Mathisson [7] and applied by him and later by P. Guenther [4], but the proof Mathisson offered is incorrect. Criterion (4), obtained here as the outcome of Theorem 1, can be deduced quite differently through Hadamard's theory of Cauchy's problem rather than through integral equations. This has been done by L. Asgeirsson in an unpublished paper.

I wish to thank P. D. Lax for several useful suggestions on the presentation of this work.

1. An Aspect of Green's Formula

Our first objective will be a particular kind of integral formula for a solution of Cauchy's problem. This will be described in §2, while here, in a preliminary step, we shall formulate Green's formula in a convenient way. For the time being, we do not restrict the number of space dimensions to three. Let [10]

$$M^*w = \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^i}\left(\sqrt{g}\, g^{ij}\frac{\partial w}{\partial x^j}\right) - \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^i}\left(\sqrt{g}\, a^i w\right) + kw$$

be the formal adjoint to the differential operator M defined in (1). Thus, if u and w are functions of class C^2, we have the identity

$$\sqrt{g}\,(wMu - uM^*w) = \frac{\partial}{\partial x^i}\left\{\sqrt{g}\, g^{ij}\left(w\frac{\partial u}{\partial x^j} - u\frac{\partial w}{\partial x^j}\right) + \sqrt{g}\, a^i uw\right\},$$

which we shall integrate over the $(n+1)$-dimensional region $D^Q_{\mathscr{S}}$ of x^i-space enclosed by the pieces $C^Q_{\mathscr{S}}$ and $\mathscr{S}^Q$ of C^Q and of $\mathscr{S}$, respectively. The result of the integration is *Green's formula*

$$\int_{D^Q_{\mathscr{S}}} (wMu - uM^*w)\sqrt{g}\, dx = \int_{C^Q_{\mathscr{S}}+\mathscr{S}^Q} n_i\left[\sqrt{g}\, g^{ij}\left(w\frac{\partial u}{\partial x^j} - u\frac{\partial w}{\partial x^j}\right) + \sqrt{g}\, a^i uw\right] dS$$

in which dx and dS represent the elements of Euclidean volume and area in x^i-space for $D^Q_{\mathscr{S}}$ and for its boundary, respectively, and the n_i are the components of the outward unit normal to the boundary. The left side of this formula is invariant under coordinate transformations; hence, the right side, too, is invariant.

Now we must recognize that the derivatives of u appearing in the integral over the characteristic cone $C^Q_{\mathscr{S}}$ can be eliminated by an integration by parts. These derivatives, in fact, are in the directions of the bicharacteristics which generate C^Q, as we readily see from the bicharacteristic equations for this characteristic surface. Let $\phi(x) = 0$ represent the equation of the surface C^Q. According to the bicharacteristic equations, $\dot{x}^i = g^{ij}\,\partial\phi/\partial x^j$, where the dot denotes d/ds, s being a parameter along a bicharacteristic curve $x^i = x^i(s)$ belonging to C^Q. On C^Q, moreover, the n_i are proportional to the $\partial\phi/\partial x^i$; hence, the combination $g^{ij}n_i\,\partial u/\partial x^j$ on C^Q is proportional to $\dot{x}^j\,\partial u/\partial x^j$ or to $\dot{u}$, as asserted.

In carrying out the integration by parts, it is useful to change to a

[10] From now on, the summation convention will be applied. The indices i, j, k, l will range from 0 to n, and p, q, r, s from 1 to n.

special coordinate system, the "normalizing coordinates" (y^i) described in [2, §2]. In this, the bicharacteristic curves emanating from Q are straight lines on which $s = -r = -\sqrt{\sum_1^n (y^p)^2}$, and the back characteristic conoid with vertex at Q is a right cone $\phi \equiv y^0 + r = \text{const.}$ whose unit normal (n_i) at each point is given by

$$n_i = \frac{\beta_i}{\sqrt{2}} \text{ on } C^Q,$$

where

$$\beta_0 = 1, \quad \beta_p = \frac{y^p - \xi^p}{r}, \quad r = \sqrt{\sum_1^n (y^p - \xi^p)^2}, \qquad p = 1, \cdots, n,$$

(ξ^i) being the coordinates of Q. The element of Euclidean area on C^Q is given by

$$dS = \sqrt{2}\, r^{n-1}\, dr\, d\omega,$$

where $d\omega$ is the element of surface area of the unit $(n-1)$-sphere. Hence, in Green's formula, the integral over $C^Q_{\mathscr{S}}$ can be written in normalizing coordinates as

$$\int_{C^Q_{\mathscr{S}}} \sqrt{h}\left[\beta_i h^{ij}\left(w\frac{\partial u}{\partial y^j} - u\frac{\partial w}{\partial y^j}\right) + \beta_i a^i uw\right] r^{n-1}\, dr\, d\omega,$$

or as

$$\int_{C^Q_{\mathscr{S}}} \sqrt{h}\left[-w\frac{du}{dr} + u\frac{dw}{dr} + \beta_i a^i uw\right] r^{n-1}\, dr\, d\omega,$$

since

$$\beta_i h^{ij}\frac{\partial}{\partial y^j} = \frac{\partial \phi}{\partial y^i} h^{ij}\frac{\partial}{\partial y^j} = \frac{d}{ds} = -\frac{d}{dr}$$

represents differentiation along a bicharacteristic. By integration by parts, this integral is further transformed into

$$\int_{C^Q_{\mathscr{S}}} uG(w) r^{n-1}\, dr\, d\omega - \int_{T^Q_{\mathscr{S}}} \sqrt{h}\, wur^{n-1}\, d\omega,$$

where

$$G(w) = r^{-(n-1)}\frac{d}{dr}(r^{n-1}\sqrt{h}\, w) + \sqrt{h}\frac{dw}{dr} + \sqrt{h}\, a^i \beta_i w,$$

and where $T^Q_{\mathscr{S}}$ is the $(n-1)$-dimensional locus of intersection of C^Q with $\mathscr{S}$. Green's formula, in normalizing coordinates, therefore becomes

$$\int_{D^Q_{\mathscr{S}}} \sqrt{h}\,(wMu - uM^*w)\, dV = \int_{C^Q_{\mathscr{S}}} uG(w)\, r^{n-1}\, dr\, d\omega$$
$$- \int_{\mathscr{S}^Q}\left(w\frac{\partial u}{\partial \nu} - u\frac{\partial w}{\partial \nu} + \nu_i a^i uw\right)\sqrt{h}\, dS - \int_{T^Q_{\mathscr{S}}} uw\sqrt{h}\, r^{n-1}\, d\omega, \tag{1.1}$$

where the ν_i ($\nu_0 > 0$, $\sum_1^n \nu_p^2 = 1$) are the components of the unit normal to $\mathcal{S}$ which points into the "future", and

$$\frac{\partial}{\partial \nu} = h^{ij}\nu_i \frac{\partial}{\partial y^j}$$

is differentiation in the "conormal" direction, an outward direction from a non-characteristic surface.

2. The Integral Formula for u

We desire to apply formula (1.1) to eliminate the integral over $C^Q_{\mathcal{S}}$ from (2) and, for this purpose, should like w to be such that $M^*w = 0$ in $D^Q_{\mathcal{S}}$ and that $G(w)/\sqrt{2} = K(Q, P)$ on $C^Q_{\mathcal{S}}$. Imposing the latter condition, in the case $n = 3$, leads to prescribing w on $C^Q_{\mathcal{S}}$ as a bounded function which has unbounded but square-integrable first inner derivatives over this truncated cone. Indeed, the equation

$$G(w) \equiv 2\sqrt{h}\frac{dw}{dr} + \sqrt{h}\left(\frac{n-1}{r} + \frac{1}{\sqrt{h}}\frac{d(\sqrt{h})}{dr} + a^i\beta_i\right)w = \sqrt{2}K$$

has a solution on C^Q given by

$$w = Wh^{-\frac{1}{4}} r^{-(n-1)/2} \exp\left\{-\tfrac{1}{2}\beta_i \int_0^r a^i(\tau - \varrho, \beta_p\varrho)\, d\varrho\right\},$$

where W is determined by quadrature from the condition

$$\frac{dW}{dr} = \frac{1}{\sqrt{2}} K h^{-\frac{1}{4}} r^{(n-1)/2} \exp\left\{\tfrac{1}{2}\beta_i \int_0^r a^i(\tau - \varrho, \beta_p\varrho)\, d\varrho\right\}.$$

In the case $n = 3$, in particular, and if condition (3) is valid, we can insure that $W = O(r)$ as $r \to 0$ and, thus, that the values of w to be prescribed on C^Q are bounded, and its inner derivatives are $O(r^{-1})$, as $r \to 0$. All these functions thus are square-integrable on any bounded part of the three-dimensional manifold C^Q as asserted.

We shall obtain a function w which, not in a strict, but in a generalized sense, in D^Q satisfies the requisite differential equation and on C^Q assumes boundary values $\bar{w}$ for which $G(\bar{w}) = K$. Let the norm $\|f\|_\pi$ of a function f relative to a subregion π of C^Q or of a space-like surface be defined as the square-root of the integral over π of the square of the function. Let R^Q be a part of the conical region $D^Q + C^Q$, truncated, if necessary, in which the coefficients in the differential equation are sufficiently well behaved; we shall suppose the lower boundary of R^Q, if it exists, to be a smooth, space-like surface, for instance a space-like plane. Our w will have "strong" derivatives w_i in R^Q in the following sense: if S is a portion of any sufficiently

smooth, space-like manifold which with part of C^Q encloses a bounded subregion of R^Q, then

$$||w||_S\,,\quad ||w_i||_S < \infty,$$

and, moreover, there exists a sequence of functions $w^{(\varkappa)}$ of class C^1 in R^Q such that

$$||w-w^{(\varkappa)}||_S \to 0, \qquad \left\|w_i - \frac{\partial w^{(\varkappa)}}{\partial y^i}\right\|_S \to 0 \qquad \text{as } \varkappa \to \infty.$$

This w, finally, will assume the requisite boundary values and also will satisfy the adjoint homogeneous differential equation in the sense that, for every function u of class C^2 in R^Q and for every sufficiently smooth space-like initial manifold $\mathscr{S}$ contained in R^Q, the integral identity

$$\begin{aligned}\int_{D^Q_{\mathscr{S}}} \sqrt{h}\, wMu\, dV &= \int_{C^Q_{\mathscr{S}}} u(P)K(Q,P)\,dC^Q_P \\ &\quad - \int_{\mathscr{S}^Q}\left(w\frac{\partial u}{\partial \nu} - uw_\nu + \nu_i a^i uw\right)\sqrt{h}\,dS - T^Q_{\mathscr{S}}[u],\end{aligned} \tag{2.1}$$

will be valid, where $w_\nu = h^{ij}\nu_i w_j$, and where $T^Q_{\mathscr{S}}[u]$ is a linear functional of u determined by the Cauchy data of u on $T^Q_{\mathscr{S}}$. Comparing (2.1) with (2) we thus will have the following integral representation of an arbitrary solution of the differential equation $Mu=F$ in terms of the values ϕ, ψ of u, $\partial u/\partial \nu$, respectively, on $\mathscr{S}$:

$$u(Q) = f_1(Q) + \int_{\mathscr{S}^Q}[w\psi + \phi(\nu_i a^i w - w_\nu)]\sqrt{h}\,dS + \int_{D^Q_{\mathscr{S}}}\sqrt{h}\,wF\,dV, \tag{2.2}$$

where $f_1(Q) = f(Q) + T^Q_{\mathscr{S}}[u]$ is a quantity that is determined by the Cauchy data prescribed on the intersection $T^Q_{\mathscr{S}}$ of C^Q with $\mathscr{S}$.

The first step in proving this result is to obtain a sequence of functions $(\bar{w}^{(\varkappa)})$ defined on C^Q such that

$$||\bar{w}^{(\varkappa)} - \bar{w}||_\pi \to 0, \quad ||D_p\bar{w}^{(\varkappa)} - D_p\bar{w}||_\pi \to 0 \quad \text{as } \varkappa \to \infty$$

on any bounded piece π of $C^Q \cap R^Q$, $\bar{w}$ again representing the agreed boundary value of w. The $\bar{w}^{(\varkappa)}$, in addition, are to be of such high differentiability, say of class C^3, on C^Q as to make possible the second step, which is to obtain a sequence of functions $w^{(\varkappa)}$ of class C^1 in R^Q and of class C^2 in the interior of this region satisfying the adjoint homogeneous equation $M^*w = 0$ in the interior of R^Q and fulfilling the boundary conditions $w^{(\varkappa)} = \bar{w}^{(\varkappa)}$ on C^Q. Each such $w^{(\varkappa)}$ might be obtained, for instance, as the limit of a sequence of solutions of suitable problems of Cauchy for space-like initial manifolds that tend to C^Q, convergence being assured by estimates easily obtainable from our integral relations. The third step consists in observing that inequalities of the type given by Friedrichs and Lewy, which are stated below, will

imply that, on any space-like manifold in R^Q, $(w^{(\varkappa)})$ and $(\partial w^{(\varkappa)}/\partial y^i)$ are Cauchy sequences; these sequences, therefore, converge in the mean on each space-like manifold S to functions w and w_i, respectively, of class L^2 in S. Finally, we let $\varkappa \to \infty$ in Green's identity (1.1) written for $w^{(\varkappa)}$ and thereby obtain (2.1). Obviously,

$$T^Q_{\mathcal{S}}[u] = \lim_{\varkappa\to\infty} \int_{T^Q_{\mathcal{S}}} u w^{(\varkappa)} \sqrt{h}\, r^{n-1}\, d\omega.$$

The requisite inequalities we now formulate in the following lemma. Both statement and proof are after Friedrichs and Lewy.[11]

LEMMA 1. *Let D be the region of x^i-space enclosed between a back characteristic conoid C and a space-like manifold S relative to a hyperbolic equation*

$$Nu \equiv a^{ij} \frac{\partial^2 u}{\partial x^i \partial x^j} + a^i \frac{\partial u}{\partial x^i} + cu = 0$$

which has bounded coefficients and continuously differentiable leading coefficients a^{ij} in the closure $\overline{D}$ of D. The boundary of D is assumed to be $C_S + S$, C_S being the part belonging to C. Then there exists a constant A such that the inequality

$$(2.3) \qquad \|u\|_S^2 + \sum_{i=0}^{n} \left\| \frac{\partial u}{\partial x^i} \right\|_S^2 \leqq A \left\{ \|u\|_{C_S}^2 + \sum_{p=1}^{n} \|D_p u\|_{C_S}^2 \right\}$$

holds for any solution of this equation which is of class C^2 in $\overline{D}$, the $D_p u$ here being n linearly independent inner derivatives of u relative to C, say

$$D_p u = (\alpha_0^2 + \alpha_p^2)^{-1/2} \left(\alpha_0 \frac{\partial u}{\partial x^p} - \alpha_p \frac{\partial u}{\partial x^0} \right) \quad (p \text{ here is a fixed index})$$

where (α_i) is the unit "forward"-pointing normal to C.

Let us assume that x^0 is a time-like variable in D and that

$$a^{00} > 0,$$

and let us also take S as a plane $x^0 = t_0 = \text{const}$. The lemma is based upon the identity

$$2u_0 Nu = (a^{00} u_0^2 - a^{pq} u_p u_q)_0 + 2(a^{0p} u_0^2 + a^{pq} u_0 u_q)_p + q(u_i, u),$$

a subscript j indicating partial differentiation with respect to x^j, in which $q(u_i, u)$ is a quadratic expression in the u_i and in u with coefficients that depend linearly on the coefficients in N and on the first derivatives of the a^{ij}. Integrating the two sides of this identity over D, and applying Gauss'

[11] Cf. [1, p. 379] and [3].

theorem, we obtain

$$0 = \int_{C_S+S} \{n_0(a^{00}u_0^2 - a^{pq}u_p u_q) + 2n_p(a^{0p}u_0^2 + a^{pq}u_0 u_q)\}\,dS + \int_D q\,dV$$
$$= \int_{C_S+S} n_0^{-1}\{u_0^2 a^{ij}n_i n_j - a^{pq}(n_p u_0 - n_0 u_p)(n_q u_0 - n_0 u_q)\}\,dS + \int_D q\,dV,$$

dV and dS here representing the elements of volume and area in x^i-space for D and for its boundary C_S+S, respectively, and n_i being the components of the outward unit normal to the boundary. Since

$$n_i = \begin{cases} -\delta_{i0} & \text{on } S, \\ +\alpha_i & \text{on } C, \end{cases}$$

and, since, furthermore, $a^{ij}\alpha_i\alpha_j = 0$ on the characteristic conoid C, it follows that we have

$$\int_S (u_0^2 a^{00} - a^{pq}u_p u_q)\,dS$$
$$= \int_{C_S} \alpha_0^{-1}\{-a^{pq}(\alpha_p u_0 - \alpha_0 u_p)(\alpha_q u_0 - \alpha_0 u_q)\}\,dS + \int_D q\,dV. \tag{2.4}$$

Now we recall that $a^{00} > 0$ and also, since the quadratic form $a^{pq}t_p t_q$ is negative definite, that

$$k_1 \sum t_p^2 \leqq -a^{pq}t_p t_q \leqq k_2 \sum t_p^2, \qquad k_1 > 0, \quad k_2 > 0. \tag{2.5}$$

Hence, if we set

$$k_3 = \min\,(a^{00}, k_1) \text{ on } S, \qquad k_4 = \max\left(\frac{k_2}{\alpha_0}\right) \text{ on } C_S,$$

we have from (2.4) the inequality

$$k_3 \sum ||u_i||_S^2 \leqq k_4 \sum ||D_p u||_{C_S}^2 + \left|\int_D q\,dV\right|.$$

However,

$$\left|\int_D q\,dV\right| \leqq k_5\left\{\sum \int_D u_i^2\,dV + \int_D u^2\,dV\right\},$$

where k_5 is a bound for the absolute values of the coefficients of the quadratic form q and, furthermore, there is a constant k_6 such that

$$\int_D u^2\,dV \leqq k_6\left\{||u||_{C_S}^2 + \sum \int_D u_i^2\,dV\right\}.$$

Hence, for a suitable constant k,

$$\sum ||u_i||_S^2 \leqq k\left(||u||_{C_S}^2 + \sum ||D_p u||_{C_S}^2 + \sum \int_D u_i^2\,dV\right).$$

Let us apply this inequality to a variable plane S_t: $x^0 = t$, denoting by D_t the region inside the cone cut off by S_t. Since

$$\int_{D_t} u_i^2\,dV = \int_0^t dt' \int_{S_{t'}} u_i^2\,dS = \int_0^t ||u_i||_{S_{t'}}^2\,dt',$$

the inequality implies that

$$\Phi(t) \leqq k \int_0^t \Phi(t')\, dt' + \Phi_0 ,$$

where

$$\Phi(t) = \sum ||u_i||^2_{S_t}$$

and

$$\Phi_0 = k\,(||u||^2_{C_S} + \sum ||D_p u||^2_{C_S}\,).$$

From this, it is seen that $\Phi \leqq \Phi_0 e^{kt}$ which together with the easily proved fact that

$$||u||^2_S \leqq k_7\,(\,||u||^2_{C_S} + \int_0^t ||u_0||^2_{S_t}\, dt)$$

($k_7 = \text{const.}$), implies the lemma.

3. Proof of Theorem 1

If Huygens' Principle is valid, then by (2.2) the integral expression

$$\int_{\mathscr{F}^Q} [w\psi + \phi(\nu_i a^i w - w_\nu)]\, \sqrt{h}\, dS$$

is determined by the values assumed by ϕ and by ψ on the boundary of the domain of integration alone. Taking $\phi \equiv 0$ we thus find, in particular, that the integral

$$\int_Q w\psi\, \sqrt{h}\, dS$$

is the same for all functions ψ with identical values on the boundary of $\mathscr{F}^Q$. Hence, $w = 0$ almost everywhere on $\mathscr{F}^Q$. Now it follows that the integral

$$\int_{\mathscr{F}^Q} \phi w_\nu\, \sqrt{h}\, dS$$

is the same for all functions ϕ with identical boundary values and that, therefore, $w_\nu = 0$ almost everywhere on $\mathscr{F}^Q$. Since $\mathscr{F}^Q$ is arbitrary, both w and $w_\nu = h^{ij}\nu_i w_j$, therefore, vanish almost everywhere on each member of any family of space-like surfaces that simply cover D^Q (or that simply cover any fixed part of D^Q to which we may desire to confine attention). This implies that w and the w_i vanish almost everywhere in D^Q (an $(n+1)$-dimensional region) from which fact, and from Fubini's theorem, we can further conclude that there is a space-like surface (an n-dimensional region) on which $w = w_i = 0$ almost everywhere. The conclusion that $K(Q, P)$ — a continuous function for $P \neq Q$ which is the limit in the mean of combinations of $w^{(\varkappa)}$ with its first derivatives—vanishes on C^Q, will follow at once, when we have established

LEMMA 2. *Suppose that w and the w_i satisfy the same conditions as in §2 and that, in addition, on the fixed space-like surface S, we have*

$$||w^{(\varkappa)}||_S \to 0, \qquad \left\| \frac{\partial w^{(\varkappa)}}{\partial y^i} \right\|_S \to 0$$

and

$$\int_{D_S^Q} (w^{(\varkappa)})^2 dV \to 0, \qquad \int_{D_S^Q} \left(\frac{\partial w^{(\varkappa)}}{\partial y^i}\right)^2 dV \to 0$$

as $\varkappa \to \infty$. Then

$$||w^{(\varkappa)}||_{C^Q} \to 0, \qquad ||D_\nu w^{(\varkappa)}||_{C^Q} \to 0 \qquad \text{as } \varkappa \to \infty.$$

The hypotheses at once imply that $||w^{(\varkappa)}||_{C^Q} \to 0$. The other limit relations follow like (2.3) from identity (2.4) and inequality (2.5).

4. Proof of Theorem 2

Subtracting (2a) from (2) gives us

$$f - f_1 + \int_{C_{\mathcal{S}}^Q} (K - K_1)\, u\, dS = 0. \tag{4.1}$$

We now repeat the arguments given in support of Theorem 1 with (4.1) in place of (2).

Bibliography

[1] Courant, R. and Hilbert, D., *Methoden der mathematischen Physik*, Vol. 2. Springer, Berlin, 1937. Distributed by Interscience, New York.

[2] Douglis, A., *The problem of Cauchy for linear hyperbolic equations of second order*, Comm Pure Appl. Math., Vol. 7, 1954, pp. 271–295

[3] Friedrichs, K and Lewy, H, *Über fortsetzbare Anfangsbedingungen bei hyperbolischen Differentialgleichungen in drei Veranderlichen*, Nachr. Ges. Wiss Gottingen, No 26, 1932, pp. 135-143.

[4] Guenther, P., *Zur Gultigkeit des Huygensschen Prinzips bei partiellen Differentialgleichungen vom normalen hyperbolischen Typus*, Berichte uber die Verhandlungen der Sachsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-naturwissenschaftliche Klasse, Band 100, Heft 2, 1952.

[5] Hadamard, J, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Yale University Press, New Haven, 1923.

[6] Hadamard, J., *The problem of diffusion of waves*, Annals of Mathematics, Second Series, Vol. 43, 1942, pp. 510–522.

[7] Mathisson, M, *Le Problème de M Hadamard relatif à la diffusion des ondes*, Acta Mathematica, Vol. 71, 1939, pp 249–282.

[8] Stellmacher, K. L, *Ein Beispiel einer Huygensschen Differentialgleichung*, Akademie der Wissenschaften, Gottingen, Mathematisch-Physikalische Klasse, IIa, Mathematisch-Physikalisch-Chemische Abteilung Nachrichten. 1953. Nr. 10, pp. 133–138

Received January 20, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 403–414 (1956)

Singularities of Generalized Axially Symmetric Potentials

ARTHUR ERDÉLYI
California Institute of Technology

§1. Let $a_0, a_1, a_2, \cdots$ be a sequence of real numbers such that

$$\limsup_{n\to\infty} |a_n|^{1/n} = 1, \tag{1}$$

and consider the two series

$$u = \sum_{n=0}^{\infty} a_n r^n P_n(\cos\theta), \tag{2}$$

$$f(z) = \sum_{n=0}^{\infty} a_n z^n. \tag{3}$$

In the first of these series r and θ are real variables, $r \geqq 0$, $0 \leqq \theta \leqq \pi$; in the second, z is a complex variable, $z = re^{i\theta}$, $r \geqq 0$, $-\pi \leqq \theta \leqq \pi$. Both series converge when $r < 1$ and diverge when $r > 1$; the first series defines a regular axially symmetric potential in the unit sphere in three dimensions, r and θ being two of the spherical polar coordinates, and the second series defines a regular analytic function in the unit circle of the complex z-plane. Each of these functions may possibly be continued analytically beyond the region of convergence of the series. Szegö [2] proved the following interesting theorem: *$z = e^{i\theta_0}$ is a singular point of $f(z)$ if and only if $r = 1$, $\theta = \theta_0$ is a singular circle of u.*

Szegö's theorem is of considerable importance in that the numerous known results on the behavior of power series on their circle of convergence may now be utilized in the theory of zonal harmonic expansions, and also in investigations of the Abel summability of series of Legendre polynomials. Szegö himself gave an interesting application of his theorem: If the a_n are real, and satisfy (1), if $N(n)$ denotes the number of non-zero coefficients in $a_0, a_1, \cdots, a_n$, and if

$$\lim_{n\to\infty} N(n)/n = 0, \tag{4}$$

then u is singular at every point of the unit sphere.

We shall look at Szegö's result from a slightly different point of view, comparing u with

$$v = \mathscr{R}e\, f(re^{i\theta}) = \sum_{n=0}^{\infty} a_n r^n \cos n\theta \tag{5}$$

rather than with $f(z)$. We then have a three-dimensional axially symmetric potential u and a two-dimensional symmetric potential v, and it is sufficient to consider these in the half-plane $r \geqq 0$, $0 \leqq \theta \leqq \pi$. Both potentials are regular in the semi-circle $r < 1$, and they are equal on the diameter of that semi-circle. Szegö's theorem now asserts that *the singularities of u on the boundary of the semi-circle of regularity coincide with those of v.*

This point of view throws new light on Szegö's result, and leads in a natural manner to some extensions of it. First, it is plausible to conjecture that the number of dimensions need not be three. Accordingly, we shall replace u by an axially symmetric potential in any number of dimensions. If use is made of the theory of generalized axially symmetric potentials developed in recent years by A. Weinstein and others, the "number of dimensions" need not be restricted to an integer. Such an extension emphasizes the fact that the essence of Szegö's theorem is the comparison of the singularities of related solutions of two similar partial differential equations in two independent variables, and it explains the necessity (mentioned by Szegö) of restricting the discussion to zonal harmonics rather than including arbitrary solid harmonics.

The semi-circular regions are the most natural regions to consider in connection with the series expansions on the right-hand sides of (2), (3), (5), but they appear unnaturally restrictive once one compares solutions of partial differential equations. Thus, secondly, it is tempting to try to replace the semi-circular regions by more general regions of regularity. The regions to be discussed here have about the same degree of generality and are maximal in the same sense as the Mittag-Leffler star of an analytic function with respect to a point.

Thirdly, in the comparison of u and v, as distinct from the comparison of u and f, the restriction to real values of the a_n, or, what is the same, to real values of u on the axis of symmetry, is no longer pertinent.

The results to be presented here are closely related to S. Bergman's researches on the singularities of solutions of partial differential equations, but this connection will not be followed up in the present lecture.

§2. Before discussing the comparison theorem, it will be useful to give a rapid survey of the salient points of generalized axially symmetric potential theory. This survey is based mainly on [3] and other papers by A. Weinstein.

Let us consider solutions of the n-dimensional Laplace equation

$$\frac{\partial^2 V}{\partial x_1^2} + \cdots + \frac{\partial^2 V}{\partial x_n^2} = 0 \tag{6}$$

which depend only on the two variables

$$x = x_1, \qquad y = (x_2^2 + \cdots + x_n^2)^{1/2}.$$

Such solutions $V = u(x, y)$ satisfy the partial differential equation

$$L(u) \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{k}{y}\frac{\partial u}{\partial y} = 0, \tag{7}$$

where

$$k = n-2;$$

it is clearly sufficient to consider such a solution in the half $y \geqq 0$ of the x, y-plane.

The partial differential equation (7) will from now on be considered with real, not necessarily integer, values of the parameter k, and will be called the differential equation of generalized axially symmetric potentials. Odd negative integer values of k will be excluded throughout the paper.

It may be remarked parenthetically that it is not necessary to interpret (7) as governing certain solutions of Laplace's equation in $k+2$ dimensions. If u is a solution of (7), and

$$U = e^{i(k-1)\phi/2} y^{(k-1)/2} u(x, y),$$

then $U(x, y, \phi)$ satisfies the partial differential equation

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{1}{y}\frac{\partial U}{\partial y} + \frac{1}{y^2}\frac{\partial^2 U}{\partial \phi^2} = 0,$$

which is the three-dimensional Laplace equation in cylindrical coordinates, so that (7) arises also in connection with certain, possibly multi-valued, three-dimensional potentials.

The introduction of non-integer values of k in (7) is justified by the resulting elegant theory (called *GASPT* by Weinstein) and also by the existence of important applications of (7) with non-integer values of k. Thus, the well-known Tricomi equation can be transformed into (7) with $k = 1/3$.

In either of the half-planes $y > 0$ and $y < 0$, (7) is an elliptic partial differential equation with analytic coefficients, and hence every twice continuously differentiable solution is an *analytic function* of x and y. The line $y = 0$, which will be called the *axis*, is a singular line of (7). This singular line is of a nature similar to that of a *regular singularity* of an ordinary linear differential equation with analytic coefficients. The *exponents* belonging to this singular line are 0 and $1-k$. Consequently, there are solutions of (7) which are *regular* on (some portion of) the axis. It is seen from the differential equation that $\partial\phi/\partial y = 0$ on the axis for regular solutions, and also that such solutions may be continued across the axis as even functions of y.

The following nomenclature will be used throughout this paper. The *axis* is the line $y = 0$. A *region* is an open connected set of points in the x, y-plane which is symmetric with respect to the axis. A region contains a portion of the axis. A *simple region* is a region whose intersection with each

line $x = \text{const}$ is either empty or else an interval. A simple region is simply-connected and convex in the y-direction. A *half-region* is that part of a region which lies in the half-plane $y \geqq 0$. A half-region is neither open nor closed. A *symmetric potential* is a solution of (7) which is regular in some region, and hence in some half-region, and is an even function of y. If k is not zero or a negative integer, it can be proved that a regular solution of (7) is an even function of y. If $k = 0, -2, -4, \cdots$, the assumption that u is an even function of y is part of the definition.

We now enumerate briefly some of the more important properties of symmetric potentials. If $u(x, y)$ is a symmetric potential, then its value on the axis,

$$g(x) = u(x, 0), \tag{8}$$

is an analytic function of x, regular on that portion of the axis which is contained in the region of regularity of $u(x, y)$. Conversely, given any analytic function $g(x)$, regular on some interval I_0, there exists a region containing I_0 and a symmetric potential regular in that region such that (8) holds. That this symmetric potential is uniquely determined follows from the *uniqueness theorem*: Two symmetric potentials which are equal on some segment (or some bounded infinite point-set) of the axis, are equal everywhere where they exist. Symmetric potentials can be continued analytically for both real and complex values of the variables x and y.

Negative odd integer values of k were excluded because, for such values of k, neither the existence theorem nor the uniqueness theorem holds. Note that the validity of the uniqueness theorem for $k = 0, -2, -4, \cdots$ is due to our definition of symmetric potentials which, in this case, includes the condition that u be an even function of y.

For the remainder of this section we assume $k > 0$.

A symmetric potential $u(x, y)$, with $k > 0$, can be expressed explicitly in terms of its values (8) on the axis in the form

$$u(x, y) = C_k \int_0^\pi g(x+iy\cos\alpha)\sin^{k-1}\alpha\, d\alpha, \qquad k > 0, \tag{9}$$

where

$$C_k^{-1} = \int_0^\pi \sin^{k-1}\alpha\, d\alpha = \Gamma\left(\frac{k}{2}\right)\Gamma\left(\frac{1}{2}\right) \Big/ \Gamma\left(\frac{k+1}{2}\right). \tag{10}$$

Contour integral solutions of (7) which are related to (9) have been obtained by A. G. Mackie [1]; they extend (9) to negative values of k.

Let C be a sufficiently smooth simple closed curve in the half-plane $y \geqq 0$, traversed counterclockwise, and bounding a closed domain D; let s be the arc-length measured along C and let $\partial/\partial n$ indicate differentiation in the direction of the exterior normal of C. For functions $f(x, y)$ and $g(x, y)$,

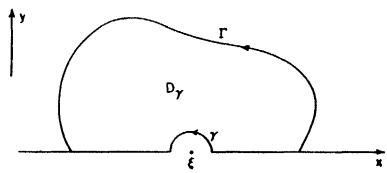
which are twice continuously differentiable in D, we have *Green's formulas* for (7),

$$\int_C y^k \frac{\partial f}{\partial n}\, ds = \iint_D y^k L(f)\, dx\, dy, \tag{11}$$

$$\int_C y^k \left(g \frac{\partial f}{\partial n} - f \frac{\partial g}{\partial n} \right) ds = \iint_D y^k [gL(f) - fL(g)]\, dx\, dy, \tag{12}$$

of which the first is a special case ($g = 1$) of the second. If C contains points of the axis, $k > 0$ must be assumed in (11) and (12).

We shall now apply these formulas to obtain a kind of converse to (9), expressing g as an integral involving u. We consider a finite number of continuously differentiable arcs, denoted in their totality by Γ, which, together with portions of the axis, bound a simply-connected half-region. The closure of this half-region will be denoted by D, and the boundary of D will be traversed counterclockwise, thus determining an orientation of Γ. In D, we fix a point $(\xi, 0)$ on the axis and a counterclockwise semi-circle γ with center



at $(\xi, 0)$, and denote the closed domain bounded by Γ, $-\gamma$ and portions of the axis by D_γ. If $u(x, y)$ is a symmetric potential regular in some half-region containing D, $k > 0$ and

$$\varrho = [(x-\xi)^2 + y^2]^{1/2} \geqq 0, \tag{13}$$

then we may apply (12) to the domain D_γ with $f = u$ and $g = \varrho^{-k}$. Clearly, these functions are twice continuously differentiable in D_γ; moreover, $L(u) = L(\varrho^{-k}) = 0$ so that the double integral vanishes, and $\partial u/\partial n = \partial \varrho^{-k}/\partial n = 0$ on the axis so that the integral extended over portions of the axis also vanishes. Thus,

$$\int_\Gamma - \int_\gamma y^k \left(\varrho^{-k} \frac{\partial u}{\partial n} - u \frac{\partial \varrho^{-k}}{\partial n} \right) ds = 0,$$

where $\partial/\partial n$ indicates differentiation in the direction of the exterior normal of D_γ. If the radius of γ is ε, we have on γ

$$x = \xi + \varepsilon \cos\theta; \qquad y = \varepsilon \sin\theta, \qquad 0 \leqq \theta \leqq \pi,$$

$$\varrho^{-k} = \varepsilon^{-k}, \qquad \frac{\partial \varrho^{-k}}{\partial n} = k\varepsilon^{-k-1},$$

and making $\varepsilon \to 0$ we obtain

$$\lim_{\varepsilon \to 0} \int_\gamma y^k \left(\varrho^{-k} \frac{\partial u}{\partial n} - u \frac{\partial \varrho^{-k}}{\partial n} \right) ds$$

$$= ku(\xi, 0) \int_0^\pi \sin^k \theta \, d\theta = k C_{k+1}^{-1} u(\xi, 0),$$

where C_{k+1} is defined by (10). We thus arrive at

$$g(\xi) = u(\xi, 0) = \frac{1}{k} C_{k+1} \int_\Gamma y^k \left(\varrho^{-k} \frac{\partial u}{\partial n} - u \frac{\partial \varrho^{-k}}{\partial n} \right) ds, \tag{14}$$

where $k > 0$, u is any symmetric potential regular in some half-region containing D and $\partial/\partial n$ indicates differentiation in the direction of the exterior normal.

In particular, let Γ be a semi-circle of radius a with center at $(\xi, 0)$. Then

$$\int_\Gamma \varrho^{-k} \frac{\partial u}{\partial n} y^k ds = a^{-k} \int_\Gamma \frac{\partial u}{\partial n} y^k ds = 0$$

by (11); also,

$$\frac{\partial \varrho^{-k}}{\partial n} = -ka^{-k-1}$$

on Γ, and hence we obtain the *mean value theorem*

$$u(\xi, 0) = C_{k+1} \int_0^\pi u(\xi + a \cos \theta, a \sin \theta) \sin^k \theta \, d\theta, \qquad k > 0. \tag{15}$$

§3. Alongside with symmetric potentials we shall consider *symmetric harmonic functions* $v(x, y)$. These are solutions of Laplace's equation

$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = 0 \tag{16}$$

which are even functions of y, $v(x, -y) = v(x, y)$, and are regular in some region; in short, they are symmetric potentials for $k = 0$. The relation

$$u(x, 0) = v(x, 0) = g(x) \tag{17}$$

sets up a correspondence between symmetric potentials (for a fixed k) and symmetric harmonic functions. For a fixed simply connected region of regularity, both u and v are uniquely determined by $g(x)$, and (17) represents a one-one correspondence. We shall represent explicitly each of the two functions u, v in terms of the other.

To express u in terms of v, we remark that

$$2v(x, y) = g(x + iy) + g(x - iy), \tag{18}$$

and hence

$$u(x, y) = 2C_k \int_0^{\pi/2} v(x, y\cos\alpha)\sin^{k-1}\alpha\, d\alpha \tag{19}$$

$$= 2C_k \int_0^1 v(x, yt)(1-t^2)^{k/2-1}dt, \qquad k>0,$$

by (9). In order to obtain a form valid also for negative k, we follow Mackie in introducing contour integrals, and re-writing (19) as

$$u(x, y) = \frac{\int_C v(x, yt)\,(1-t^2)^{k/2-1}\,dt}{\int_C (1-t^2)^{k/2-1}dt} \tag{20}$$

where the path of integration will be specified presently. Setting, for the moment,

$$\frac{\partial v(x, y)}{\partial y} = w(x, y),$$

the identity

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{k}{y}\frac{\partial}{\partial y}\right) v(x, yt)(1-t^2)^{k/2-1}$$

$$= -\frac{1}{y}\frac{d}{dt}\left[w(x, yt)(1-t^2)^{k/2}\right]$$

may be verified by explicit computation using (16). From this identity it follows that the right-hand side of (20) satisfies (7) provided that v satisfies (16) and C is a path of integration for which

$$\int_C (1-t^2)^{k/2-1}dt \neq 0, \qquad \int_C d\left[w(x, yt)(1-t^2)^{k/2}\right] = 0.$$

Moreover, under these circumstances clearly $u(x, y)$ is an even function of y and $u(x, 0) = v(x, 0)$. If $k > 0$, the interval from $t = 0$ to $t = 1$ may be taken as the path of integration, since $w(x, yt)$ is an odd function of t and $w(x, 0) = 0$. In general, we may take C to be a loop beginning and ending at $t = 0$ and encircling $t = 1$ counterclockwise, except that we must exclude odd negative integer and even positive integer values of k for which the denominator in (20) vanishes. There are many other possible paths, but for our purposes it will not be necessary to specify C in (20).

To express v in terms of u, we may proceed in either of two possible ways. Setting

$$z = x+iy, \qquad \bar{z} = x-iy, \qquad \zeta = \xi+i\eta, \qquad \bar{\xi} = \xi-i\eta,$$

we have from (13),

$$\varrho^2 = (x-\xi)^2+y^2 = (z-\xi)(\bar{z}-\xi),$$

and from (18) and (14),

$$v(\xi,\eta)=\frac{1}{2}[g(\zeta)+g(\bar{\zeta})]=\frac{1}{k}C_{k+1}\int_\Gamma y^k\left(h\frac{\partial u}{\partial n}-u\frac{\partial h}{\partial n}\right)ds, \qquad k>0, \tag{21}$$

where

$$2h(x,y,\xi,\eta)=[(z-\zeta)(\bar{z}-\zeta)]^{-k/2}+[(z-\bar{\zeta})(\bar{z}-\bar{\zeta})]^{-k/2}.$$

That branch of h must be taken for which both fractional powers are positive when (x, y) is on Γ and (ξ, η) is a point on the axis in $D-\Gamma$. This branch is an analytic function of the four real variables x, y, ξ, η which is single-valued and regular when (x, y) ranges over Γ and (ξ, η) ranges over $D-\Gamma$. Thus, (21) defines a symmetric harmonic function in the half-region $D-\Gamma$, and for this harmonic function $v(\xi, 0) = u(\xi, 0) = g(\xi)$ by (14).

There does not seem to be any obvious way of extending (21) to negative values of k. Instead, we shall follow Mackie in regarding (19) or (20) as an integral equation for v. Solution of that integral equation leads to

$$v(x,y)=\frac{\int_C u(x,yt)t^k(1-t^2)^{-k/2-1}\,dt}{\int_C t^k(1-t^2)^{-k/2-1}\,dt}. \tag{22}$$

A computation similar to that carried out in the proof of (20) shows that, $u(x, y)$ being a symmetric potential, the right-hand side of (22) is meaningful and satisfies Laplace's equation, provided that the path of integration C is chosen so that

$$\int_C t^k(1-t^2)^{-k/2-1}\,dt\neq 0, \qquad \int_C d\left[\frac{\partial u}{\partial y}(x,yt)t^k(1-t^2)^{-k/2}\right]=0.$$

Clearly, $v(x, y)$, as defined by (22), is an even function of y, and $v(x, 0) = u(x, 0)$. The following special contours may be used in (22): a path from $t = 0$ to $t = 1$ if $-1 < k < 0$, a loop starting and ending at $t = 0$ and encircling $t = 1$ if $k > -1$, and a loop starting and ending at $t = 1$ and encircling $t = 0$ if $k < 0$. For our purposes it will not be necessary to specify C in (22).

Equations (20) and (22) are particular cases of a relation connecting symmetric potentials of different "dimensions". Let k, m be real numbers, neither of them a negative odd integer. Let u_k be a regular solution of (7) which is an even function of y, and let u_m be a similar solution of the differential equation obtained by replacing k by m in (7). If we set, for the moment

$$\frac{\partial u_m(x,y)}{\partial y}=w(x,y)$$

the identity

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{k}{y}\frac{\partial}{\partial y}\right)u_m(x,yt)\,t^m(1-t^2)^{(m-k)/2-1}$$
$$=-\frac{1}{y}\frac{d}{dt}[w(x,yt)t^m(1-t^2)^{(m-k)/2}]$$

may be verified by explicit computation, and the relation

$$u_k(x, y) = \frac{\int_C u_m(x, yt)\, t^m (1-t^2)^{(m-k)/2-1}\, dt}{\int_C t^m (1-t^2)^{(m-k)/2-1}\, dt} \tag{23}$$

for paths of integration satisfying

$$\int_C t^m (1-t^2)^{(m-k)/2-1}\, dt \neq 0,$$

$$\int_C d[w(x, yt) t^m (1-t^2)^{(m-k)/2}] = 0$$

follows as before. The correspondence between symmetric potentials of different dimensions is established by the relation

$$u_k(x, 0) = u_m(x, 0)$$

which is clearly satisfied by (23).

§4. We are now ready to discuss singularities of symmetric potentials.

Let $u(x, y)$ be a symmetric potential, regular in some region containing an open interval I_0 of the axis. Then $g(x) = u(x, 0)$ is an analytic function of x, regular on I_0. Continuing $g(x)$ analytically along the axis, we obtain a unique maximal interval I such that $I_0 \subset I$ and $g(x)$ is regular on I. (I may be finite or infinite.) The extended function $g(x)$ on I determines uniquely a symmetric potential which is regular in some region containing I, and reduces to $g(x)$ on I. By the uniqueness theorem, this symmetric potential is an analytic continuation of the given symmetric potential and may be denoted by $u(x, y)$. For each x in I, there is a maximal interval J_x: $-\beta(x) < y < \beta(x)$ on the line through $(x, 0)$ parallel to the y-axis such that $u(x, y)$ is regular on J_x. The union of all the intervals J_x, where x runs over I, is a simple region R. Being simple, R is simply connected, and u is single-valued and regular in R. This region will be called the *simple region of regularity of* u with respect to I_0. We shall also write $I(u)$, $R(u)$ in order to indicate the dependence of the interval I and of the region R on u. I and R depend also on I_0 but this dependence will not be indicated in the notation.

The same construction may be repeated with a symmetric harmonic function $v(x, y)$, thus leading to $R(v)$, the simple region of regularity of the symmetric harmonic function v with respect to I_0.

THEOREM. *Let* $k \neq -1, -3, \cdots$, *and let* $u(x, y)$ *be a symmetric potential and* $v(x, y)$ *a symmetric harmonic function. If* $u(x, 0) = v(x, 0)$ *on some interval* I_0 *of the axis, then the simple regions of regularity of* u *and* v *with respect to* I_0 *coincide.*

COROLLARY. *For a fixed analytic function* $g(x)$, *regular on some interval* I_0 *of the axis, let* $u(x, y)$ *be the symmetric potential for which* $u(x, 0) = g(x)$.

Then the simple region of regularity of u with respect to I_0 is independent of k, $k \neq -1, -3, -5, \cdots$.

Proof of the theorem: Since $u(x, 0) = v(x, 0) = g(x)$ on I_0, clearly $I(u) = I(v) = I$. If (x, Y) is a point in $R(v)$, then v is regular at (x, y) for $0 \leqq y \leqq Y$. For $k > 0$, we see that $v(x, yt)$ is a regular function of the real variables x and y for $0 \leqq y \leqq Y$, $0 \leqq t \leqq 1$, and from (19), $u(x, y)$ is regular for $0 \leqq y \leqq Y$. Thus $R(u) \supset R(v)$ if $k > 0$. If $k < 0$, we remark that $v(x, y)$ may be continued to complex values of y, and $v(x, yt)$ is regular if $0 \leqq y \leqq Y$ and t is in a sufficiently small complex neighborhood of the segment $0 \leqq t \leqq 1$. Since C may always be chosen to lie within this neighborhood, it follows from (20) that $R(u) \supset R(v)$ also for negative values of k. By a similar argument it follows from (22), or, in the case of positive k from (21), that $R(u) \subset R(v)$. Hence $R(u) = R(v)$.

The corollary is a simple consequence of the theorem. Alternatively, it may be proved by means of (23).

Our theorem states that, broadly speaking, the *position of the singularities* of the symmetric potential $u(x, y)$ depends only on $g(x) = u(x, 0)$, and is independent of k. It is natural to ask whether the *nature of the singularities* is also independent of k. The following examples show that the answer to this question is in the negative.

The fundamental solution of (7) with a singularity at $(0, 0)$ is ([3] p. 31)

$$u_1(x, y) = (x^2+y^2)^{-k/2},$$

and the symmetric harmonic function determined by the condition $v_1(x, 0) = u_1(x, 0)$ for $x > 0$ is

$$v_1(x, y) = \tfrac{1}{2}\left[(x+iy)^{-k}+(x-iy)^{-k}\right].$$

In polar coordinates r, θ, where

$$x = r\cos\theta, \qquad y = r\sin\theta,$$

we have

$$u_1 = r^{-k}, \qquad v_1 = r^{-k}\cos k\theta.$$

Thus, $u_1 = v_1$ when $\theta = 0$, but u_1 and v_1 behave differently as $(x, y) \to (0, 0)$.

The second example will show that even the order of the singularity may depend on k. The fundamental solution of (7) with singularities at $(0, \pm b)$, $b > 0$, is ([3] p. 32)

$$u_2(x, y) = C_k\int_0^\pi (x^2+y^2+b^2-2yb\cos\alpha)^{-k/2}\sin^{k-1}\alpha\, d\alpha, \qquad k > 0.$$

Here

$$g_2(x) = u_2(x, 0) = (x^2+b^2)^{-k/2},$$

and the corresponding symmetric harmonic function is

$$v_2(x, y) = \tfrac{1}{2}\left[(z^2+b^2)^{-k/2}+(\bar{z}^2+b^2)^{-k/2}\right].$$

Thus,

$$v_2 = O(|z-ib|^{-k/2}) \text{ as } (x, y) \to (0, b),$$

while it is known that u_2 has a logarithmic singularity at $(0, b)$.

The third and last example shows the influence of k on the nature of the singularity. We set

$$g_3(x) = (b+ix)^{-s},$$

where $b > 0$, $s > 0$, and take that branch of $g_3(x)$ for which $g_3(0) > 0$. The corresponding symmetric potential and symmetric harmonic function are

$$u_3(x, y) = C_k \int_0^\pi (b+ix+y\cos\alpha)^{-s} \sin^{k-1}\alpha \, d\alpha,$$

$$v_3(x, y) = \tfrac{1}{2}\left[(b+y+ix)^{-s}+(b-y+ix)^{-s}\right].$$

They both have singularities at $(0, \pm b)$, and no other finite singularities. Now, the integral in u_3 is a known integral representation of Legendre functions, and its behavior as $(x, y) \to (0, b)$ may be determined by means of the known behavior of Legendre functions in the neighborhood of a singularity. It will be sufficient to give the results here. It turns out that

$$u_3(x, y) = \begin{cases} O(1) & \text{if } k > 2s, \\ O(\log|z-ib|) & \text{if } k = 2s, \\ O(|z-ib|^{-s+k/2}) & \text{if } k < 2s. \end{cases}$$

§5. Results of the nature of Szegö's theorem follow easily from the discussion of the preceding section. For the sake of simplicity we assume that $k \neq 0, -1, -2, \cdots$. Let us introduce polar coordinates r, θ by setting

$$x = r\cos\theta, \qquad y = r\sin\theta, \qquad -\pi \leqq \theta \leqq \pi,$$

with $0 \leqq \theta \leqq \pi$ in the half-plane. Symmetric product solutions of (7), obtained by separation of the variables r and θ, are of the form

$$r^n C_n^{k/2}(\cos\theta), \qquad n = 0, 1, 2, \cdots,$$

where $C_n^{k/2}(\cos\theta)$ is the Gegenbauer (ultraspherical) polynomial defined by the generating function

$$(1-2t\cos\theta+t^2)^{-k/2} = \sum_{n=0}^{\infty} t^n C_n^{k/2}(\cos\theta).$$

In particular, $C_n^{k/2}(1) \neq 0$ for $k \neq 0, -1, -2, \cdots$, $n = 0, 1, 2, \cdots$, and

$$C_n^{k/2}(-1) = (-1)^n C_n^{k/2}(1).$$

Symmetric product solutions of the two-dimensional Laplace equation are of the form

$$r^n \cos n\theta, \qquad n = 0, 1, 2, \cdots.$$

Suppose now that u is a symmetric potential and v a symmetric harmonic function such that $u(x, 0) = v(x, 0) = g(x)$ on some interval I_0 containing the origin. Then $g(x)$ is regular in some neighborhood of the origin, and possesses a power series expansion

$$g(x) = u(x, 0) = v(x, 0) = \sum_{n=0}^{\infty} a_n x^n.$$

We now claim that

$$u(x, y) = \sum_{n=0}^{\infty} a_n r^n \frac{C_n^{k/2}(\cos\theta)}{C_n^{k/2}(1)}, \tag{24}$$

$$v(x, y) = \sum_{n=0}^{\infty} a_n r^n \cos n\theta. \tag{25}$$

Indeed, the series for u and v have the same radius of convergence (in the real r, θ-plane) as the series for g (in the complex x-plane). Within their circle of convergence, (24) and (25) represent, respectively, a symmetric potential and a symmetric harmonic function, and both u and v reduce to g when $y = 0$.

The infinite series in (24) and (25) converge in the largest open circular disk with center at the origin and contained in the common simple region of regularity of u and v with respect to an interval containing the origin; on the boundary of the circle of convergence u is regular (singular) wherever v is regular (singular).

If, in particular, $g(x)$ is a real function of the real variable x, i.e., if the a_n are all real, than the singularities of

$$g(x+iy) = \sum_{n=0}^{\infty} a_n r^n e^{ni\theta} \tag{26}$$

are symmetrically situated with respect to the axis and are, in fact, identical with the singularities of v. In this case we may compare (24) and (26) instead of (24) and (25).

Bibliography

[1] Mackie, A. G. *Contour integral solutions of a class of differential equations*, J. Rat. Mech. Anal. Vol. 4, 1955, pp 733–750

[2] Szegö, G., *On the singularities of zonal harmonic expansions*, J. Rat. Mech. Anal. Vol. 3, 1954, pp. 561–564.

[3] Weinstein, A., *Generalized axially symmetric potential theory*, Bull. Amer. Math. Soc. Vol. 59, 1953, pp. 20–38.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 415–423 (1956)

Growth Properties of Solutions of Non-linear Elliptic Equations*

ROBERT FINN
University of Southern California

§1. We shall consider non-linear elliptic second order partial differential equations of the form

$$a(x, y, \varphi, p, q)\varphi_{xx}+2b(x, y, \varphi, p, q)\varphi_{xy}+c(x, y, \varphi, p, q)\varphi_{yy} = D(x, y, \varphi, p, q) \tag{1}$$

where $p = \varphi_x$, $q = \varphi_y$, and the coefficients are assumed normalized so that $a > 0$ and $ac-b^2 = 1$.

It is known that if $D = 0$ and if there is a fixed constant $K < \infty$ such that $a+c < K$ for every solution of (1) throughout the domain of definition (i.e. if the equation is uniformly elliptic), then the solutions exhibit a behavior which patterns itself very closely on that of harmonic functions. In particular, there is an *a priori* estimate on the magnitude of the gradient, and also of the higher derivatives, of a bounded solution in a compact subdomain of the region of regularity, and the first boundary value problem (Dirichlet problem) will in general be correctly set in the sense of Hadamard [8]. Now there are some equations of the form (1) for which similar behavior of the solutions is to be expected on the basis of physical considerations, but which do not fall into the category of uniformly elliptic equations. For such equations the usual methods of proof break down, and it is therefore desirable to discuss the behavior of solutions by methods which take account of the non-linearity of the equation in a natural manner. We shall approach this question by a study of the consequences of abstracting essential features of the non-linearities which occur in the minimal surface equation

$$\alpha\varphi_{xx}+2\beta\varphi_{xy}+\gamma\varphi_{yy} = 0, \tag{2}$$

where

$$\alpha = \frac{1+q^2}{(1+p^2+q^2)^{1/2}}, \quad \beta = \frac{-pq}{(1+p^2+q^2)^{1/2}}, \quad \gamma = \frac{1+p^2}{(1+p^2+q^2)^{1/2}},$$

the equation whose solutions represent surfaces of constant mean curvature H,

$$\alpha\varphi_{xx}+2\beta\varphi_{xy}+\gamma\varphi_{yy} = 2H(1+p^2+q^2), \tag{3}$$

*This investigation was supported by the Office of Naval Research.

and the equation of polytropic flow of an ideal gas,

$$\left(1-\frac{\varphi_x^2}{c^2}\right)\varphi_{xx}-2\frac{\varphi_x\varphi_y}{c^2}\varphi_{xy}+\left(1-\frac{\varphi_y^2}{c^2}\right)\varphi_{yy}=0 \tag{4}$$

where $c^2=1-\frac{\gamma-1}{2}(\varphi_x^2+\varphi_y^2)$, and γ is the ratio of specific heats for the gas.

§2. The results of the present study suggest the formulation of a general principle. Many of the properties of the solutions of an equation of form (1) can be obtained as consequences of simple requirements on the *conformal structure* of certain Riemannian metrics which are defined over the x, y-plane in a natural manner by every solution of the equation. These requirements can be characterized *a priori* by simple conditions on the coefficients. For example, the coefficients of (1) induce a metric over the x, y-plane,

$$R_1: \quad ds_1^2=c\,dx^2-2b\,dx\,dy+a\,dy^2.$$

The requirement of uniform ellipticity is equivalent to the requirement that this metric should be quasi-conformal with respect to the Euclidean metric of the plane. By this is meant the following: Denote the Euclidean metric of the plane by

$$R_0: \quad ds_0^2=dx^2+dy^2.$$

Let $Q_M(x, y)=\max ds_1/ds_0$ at the point (x, y) for all choices of the ratio dy/dx, and let $Q_m(x, y)=\min ds_1/ds_0$ among all such ratios. We shall say that the metrics R_1 and R_0 are quasi-conformally related in a region G provided there is a fixed constant $K<\infty$ such that $Q_M/Q_m<K$ for all (x, y) in G. In this case we shall write $R_1\sim R_0$. (Clearly $Q_M/Q_m=1$ at a point if and only if the metrics coincide up to a multiplicative factor.) Thus, uniform ellipticity of (1) is equivalent to the condition $R_1\sim R_0$.

One may also consider the metric induced over the plane by the surface S_φ represented in x, y, φ-space by a solution $\varphi(x, y)$ of (1)

$$R_2: \quad ds_2^2=\gamma dx^2-2\beta\,dx\,dy+\alpha\,dy^2$$

where α, β, γ are the coefficients of the second order terms in the equation (3) of surfaces of constant mean curvature. Thus, R_2 and R_1 are conformally related for every solution of (3). The condition that R_2 and R_1 should be in quasi-conformal relation, $R_2\sim R_1$, is simply expressed. It is

$$a\gamma+c\alpha-2b\beta\leqq K \tag{5}$$

for some constant $K<\infty$.

Finally, one may consider equations (1) which have a self-adjoint form,[1]

$$[\mathscr{A}(x, y, \varphi, p, q)]_x+[\mathscr{B}(x, y, \varphi, p, q)]_y=0. \tag{6}$$

[1]Conditions under which (1) may be written in the form (6) have been given by Bers [5].

Equation (6) is an integrability condition. We may conclude the existence of a conjugate function $\psi(x, y)$, with $\psi_y = \mathscr{A}$, $\psi_x = -\mathscr{B}$, and we may consider the metric

$$R_3: \quad ds_3^2 = d\varphi^2 + d\psi^2.$$

The function $\psi(x, y)$ defines a surface S_ψ and the associated metric

$$R_4: \quad ds_4^2 = G\,dx^2 - 2F\,dx\,dy + E\,dy^2$$

where E, F, G are expressed in terms of the functions $\mathscr{A}$ and $\mathscr{B}$

$$E = \frac{1+\mathscr{A}^2}{(1+\mathscr{A}^2+\mathscr{B}^2)^{1/2}}, \quad F = \frac{-\mathscr{A}\mathscr{B}}{(1+\mathscr{A}^2+\mathscr{B}^2)^{1/2}}, \quad G = \frac{1+\mathscr{B}^2}{(1+\mathscr{A}^2+\mathscr{B}^2)^{1/2}}.$$

The condition $R_0 \sim R_4$ is particularly simple. It is

$$\mathscr{A}^2 + \mathscr{B}^2 < K^2 < \infty. \tag{7}$$

It should be noted that for a given equation (1) the functions $\mathscr{A}$ and $\mathscr{B}$ are not uniquely determined. For example, to any two functions $\mathscr{A}$, $\mathscr{B}$, the functions $\bar{\mathscr{A}} = \mathscr{A} + q$, $\bar{\mathscr{B}} = \mathscr{B} - p$, give rise to the same equation. Hypotheses introduced in this paper on $\mathscr{A}$ and $\mathscr{B}$ should be understood to mean that it is possible to choose these functions to have the specified properties. We may note also that when these hypotheses are satisfied for a particular solution $\varphi(x, y)$ then the appropriate theorems are valid for this solution. The main interest of the theorems lies of course in the *a priori* estimates which result from properties of $\mathscr{A}$ and $\mathscr{B}$ which are independent of the solution considered.

In the cases of equations (2), (3), and (4) the forms for $\mathscr{A}$ and $\mathscr{B}$ arising from the considerations of classical mechanics are of special interest:

$$\left[\frac{p}{(1+p^2+q^2)^{1/2}}\right]_x + \left[\frac{q}{(1+p^2+q^2)^{1/2}}\right]_y = 0, \tag{2a}$$

$$\left[\frac{p}{(1+p^2+q^2)^{1/2}}\right]_x + \left[\frac{q}{(1+p^2+q^2)^{1/2}}\right]_y = 2H, \tag{3a}$$

$$(\varrho p)_x + (\varrho q)_y = 0 \tag{4a}$$

where

$$\varrho = \left[1 - \frac{\gamma-1}{2}(p^2+q^2)\right]^{\frac{1}{\gamma-1}}.$$

Note that $R_0 \sim R_4$ for any solution of (2a), (3a) or (4a), $R_1 \sim R_2$ for any solution of (2a) or (3a).

§3. We shall now state some of the consequences of imposing relations of quasi-conformality among the metrics introduced in §2. No attempt will be made to present an exhaustive discussion; rather we shall try to emphasize

the sense of the main contributions. For a more complete presentation, the reader is referred to the references cited.

THEOREM 1. *Suppose* i) $D = 0$ *and* ii) $R_1 \sim R_2$ (*i.e.*, $a\gamma+c\alpha-2b\beta \leqq K < \infty$). *Let* $\varphi(x, y)$ *be a solution of* (1) *in the disc* $x^2+y^2 < R^2$, *and denote by* $\mathfrak{A}$ *the area of the surface* S_φ *represented by* $\varphi(x, y)$. *Then, setting* $\omega(x, y) = (p^2+q^2)^{1/2}$,

$$\omega(0, 0) \leqq \frac{40}{\pi}\frac{\mathfrak{A}}{R^2}\exp\left\{\frac{256\,\pi K^2\,\mathfrak{A}}{R^2}\right\}.$$

In particular, we obtain an *a priori* estimate for the inclination of a minimal surface $\varphi(x, y)$ in terms of the area of the surface and the distance to the boundary of the region of regularity [11].

THEOREM 2. *Suppose* i) $D = 0$, ii) $R_1 \sim R_2$, iii) *additional conditions.*[2] *Then any function* $\varphi(x, y)$ *which is a solution of* (1) *throughout the* x, y*-plane is necessarily a linear function* $\varphi = Ax+By+C$ [12].

This result generalizes a celebrated theorem of S. N. Bernstein [3] on solutions of the minimal surface equation.[3] Another result of this type has been obtained independently by Bers [7].

THEOREM 3. *Suppose* i) $R_0 \sim R_4$ (*i.e.*, $\mathscr{A}^2+\mathscr{B}^2 < K^2 < \infty$), ii) *additional conditions.*[4] *Let* $\varphi(x, y)$ *be a solution of* (6) *in the disc* $x^2+y^2 < R^2$, *and suppose* $|\varphi| < M < \infty$. *Denote by* $\mathfrak{A}$ *the area of the surface* S_φ *defined by* $\varphi(x, y)$ *in* $x^2+y^2 < R^2$. *Then there are constants* λ_1 *and* λ_2, *depending only on the equation and not on the solution considered, such that*

$$\mathfrak{A} < \lambda_1 R^2+\lambda_2 MR.$$

We remark that under weaker conditions it is possible to find a bound for the area in any concentric subcircle. Such conditions are formulated in Lemma 6 of [11].

THEOREM 4. *Under the hypotheses of Theorems* 1 *and* 3, *there exists a bound* (*explicitly known*) *for the magnitude of the gradient of a solution* $\varphi(x, y)$ *of* (6), *which depends only on a bound for* $|\varphi(x, y)|$, *on the distance to the boundary of the domain of definition, and on the equation considered* [11].

This result is an immediate consequence of Theorems 1 and 3.

The effect of Theorem 4 is that for all solutions $\varphi(x, y)$ of (6) which are bounded in magnitude throughout a region G of regularity, equation (6) will

---

[2] For a detailed statement, see [12]. Essentially the equation is assumed to arise from a suitable variational problem. This restriction is believed by the author to be unnecessary.

[3] For the equation (3) with $H \neq 0$ the conclusion of Theorem 2 is almost a triviality.

[4] It is sufficient that the function φ does not occur explicitly in $\mathscr{A}$ or $\mathscr{B}$. However, much more general conditions can be considered. See [11] where the extensions are not given but are evident from the method of proof.

remain uniformly elliptic in any compact subdomain G_1 of G. According to a theorem of Morrey [18], the first derivatives of any such solution will satisfy a fixed Hölder condition[5] in G_1, and by a theorem of E. Hopf [14] the second derivatives will similarly satisfy a Hölder condition. Using a theorem of McShane [17] on the conformal representation of surfaces, we obtain the following result:

THEOREM 5. *Under the hypotheses of Theorems* 1 *and* 3, *the first boundary value problem (Dirichlet problem) for an equation* (6) *can be solved for arbitrary continuous prescribed data* φ_c *on any convex plane curve* Γ [11].

Previous results of this type [1, 16, 19] had required that the boundary curve satisfy a three point condition (see, e.g. [20]). Theorem 5 thus gives sufficient conditions under which the three point condition becomes superfluous. Convexity of the curve Γ is known to be necessary.

Theorem 5 derives particular interest from Theorem 6, which we state in a qualitative manner.

THEOREM 6. *Theorem* 5 *cannot be extended to a significantly broader class of equations.*

Clearly the region of validity for the conclusion of Theorem 5 cannot be sharply defined. We may, however, give an example of an equation (6) which satisfies all the hypotheses of Theorem 5 except the condition $R_2 \sim R_1$, and for which the first boundary value problem cannot in general be solved for continuous data on a convex curve. Since this example is particularly simple we include it in this report.

Let N be any integer, $N \geqq 2$. Let $\mathscr{A}(p) = \frac{1}{2}\left(1 - \frac{1}{3p^3}\right)$ for $p \geqq 1$, $\mathscr{A}(p) = \frac{1}{2}\left(-1 - \frac{1}{3p^3}\right)$ for $p \leqq -1$. In the interval $-1 \leqq p \leqq 1$ we may define $\mathscr{A}(p)$ so that $\mathscr{A}_p(p) > 0$ for all p, $\mathscr{A}(0) = 0$, and $\mathscr{A}(p)$ is N times continuously differentiable for all p. Now define $\mathscr{B}(q) = \frac{1}{2}q$ for $|q| \leqq 1$, and $\mathscr{B}(q)$ N times differentiable with $\mathscr{B}_q > 0$ and $|\mathscr{B}| < \frac{1}{2}\sqrt{2}$ for all q. Clearly $\mathscr{A}^2 + \mathscr{B}^2 < 1$ for all (p, q).

We easily verify that the function

$$\varphi^0(x,y) = \begin{cases} \frac{1}{2}[x(64+y^2)]^{1/2}, & x \geqq 0, \\ -\frac{1}{2}[|x|(64+y^2)]^{1/2}, & x \leqq 0, \end{cases}$$

is a continuous function throughout the disc $x^2+y^2 \leqq 1$, and a solution of the elliptic partial differential equation

$$\mathscr{A}_x + \mathscr{B}_y = 0 \tag{8}$$

[5] A function $f(x, y)$ is said to satisfy a Hölder condition with constant K and exponent α in a region G, if

$$|f(x_1, y_1) - f(x_2, y_2)| \leqq K[(x_1-x_2)^2 + (y_1-y_2)^2]^{\alpha/2}$$

for any two points (x_1, y_1), (x_2, y_2) of G.

except on the line $x = 0$, where φ_x^0 becomes unbounded. On the circumference Γ: $x^2+y^2 = 1$, $\varphi^0(x, y)$ defines a continuous function of arc φ_c. Let us choose φ_c as prescribed data.

If there should exist in the unit disc a solution $\varphi(x, y)$ which assumes the values φ_c on Γ, then the function $\Phi(x, y) = -\varphi(-x, y)$ would be another such solution. For, in the range of values (p, q) assumed by $\varphi(x, y)$ in the disc (8) takes the form

$$\frac{1}{p^2}\varphi_{xx}+p^2\varphi_{yy} = 0$$

and φ_c is antisymmetric with respect to reflection in the line $x = 0$. By general uniqueness theorems for quasi-linear elliptic equations (see e.g. [9]) we find $\Phi(x, y) \equiv \varphi(x, y)$, and hence $\varphi(x, y) = 0$ on the line $x = 0$. Consequently $\varphi(x, y)$ assumes, on the semi-circle for which (say) $x \geqq 0$, the same boundary values as $\varphi^0(x, y)$. Another application of the uniqueness theorem shows that $\varphi(x, y) \equiv \varphi^0(x, y)$. Therefore $\varphi(x, y)$ cannot be a solution throughout the disc; that is, the boundary value problem has no solution for the continuous prescribed data φ_c.

We remark that an assertion of S. N. Bernstein [2], that the first boundary value problem can be solved for continuous data on a circle whenever it can be solved for analytic data, is evidently incorrect.

Our next theorem can be regarded as an analogue of Schwarz's lemma.

THEOREM 7. *Suppose* i) $R_0 \sim R_4$ (*i.e.*, $\mathscr{A}^2+\mathscr{B}^2 < K^2 < \infty$), ii) *additional conditions.*[6] *Then to every equation* (6) *and positive real number* R *there corresponds a positive increasing function* $f_R(\varrho)$, *with* $f_R(\varrho) < 1$ *for* $0 \leqq \varrho < R$ *such that if* $\varphi(x, y)$ *is a solution of* (6) *in* $x^2+y^2 < R^2$ *and* $\psi(x, y)$ *is its conjugate function, the relations* a) $\varphi(0, 0)+i\psi(0, 0) = 0$ *and* b) $|\varphi+i\psi| < 1$, *imply the relation* $|\varphi(x, y)+i\psi(x, y)| < f_R\ (\sqrt{x^2+y^2})$.

A suitable function $f_R(\varrho)$ can be found explicitly [13]. The author does not know whether this function can be chosen so that $f_R(0) = 0$. If so, it must have an infinite slope at this point, for the counterexample given above shows that the conditions of the theorem are not enough to impose an *a priori* bound on the gradient of a solution. If the condition $R_1 \sim R_2$ is also imposed, a more striking analogy with classical function theory is obtained [11].

As a final result, we include a theorem which was first obtained by quite different considerations, but which can be viewed within the framework of this paper.

[6] It is sufficient that $\mathscr{A} = \mathscr{A}(x, y, p, q)$, $\mathscr{B} = \mathscr{B}(x, y, p, q)$, and that $\mathscr{A}(x, y, 0, 0)$ and $\mathscr{B}(x, y, 0, 0)$ are constant for all (x, y) in the region considered. These conditions include the cases of principal interest, but are stronger than necessary (see [13])

THEOREM 8. *Suppose* $\mathscr{A} = \mathscr{A}(p, q)$, $\mathscr{B} = \mathscr{B}(p, q)$, *and that* $\mathscr{A}^2+\mathscr{B}^2 < K^2 < \infty$ *for all* (p, q). *Let* H *denote any constant, and suppose that the domain of values* (p, q) *for which the equation*

$$\mathscr{A}_x+\mathscr{B}_y = 2H \tag{6a}$$

is of elliptic type is a convex set in the p, q*-plane. Then a single valued solution* $\varphi(x, y)$, *for which* (6a) *is elliptic, admits no isolated singularities* [10].

This theorem was first proved for the minimal surface equation (2) by Bers [4], who also conjectured that a stronger result would be true. The present result includes the minimal surface equation, the equation (3) for surfaces of constant mean curvature[7], and subsonic solutions (4) of the equation of gas-dynamics. In particular, it yields the result that a subsonic flow of an ideal gas cannot become sonic at an isolated point. The method of proof given by this author differes essentially from that of Bers and seems also to be simpler.

It is of interest to note that the equation considered in the counter-example given above satisfies the hypotheses of Theorem 8. Thus we find that under hypotheses sufficient to impose both, a striking regularity on the local behavior of a solution and an *a priori* estimate on the growth of a solution in a circle (Theorem 7), it is nevertheless impossible to find a bound on the gradient of a solution interior to its domain of definition.

§4. We include here the statements of a few of the principal lemmas and brief outlines of their application to the theorems stated in §3. For a more complete discussion the reader must be referred to the cited literature.

LEMMA 1. *Let* $\varphi(x, y)$ *be of class* C'' *in the unit disc* $x^2+y^2 \leqq 1$, *and denote by* $\mathfrak{A}$ *the area of the surface* S_φ *defined by* $\varphi(x, y)$. *Suppose that* S_φ *is mapped conformally onto the unit disc* $|w| \leqq 1$ of the $w = u+iv$ *plane. Then,* setting $z = x+iy$,

$$\iint\limits_{|w|\leqq 1} \{|z_u|^2+|z_v|^2\}\,du\,dv \leqq 2\mathfrak{A}.$$

This is immediate, since the area of a surface expressed in isothermal coordinates is an integral of a sum of squares, of which we have discarded those involving φ.

If we now apply a device due to Lebesgue [15], we find that, after a suitable normalization of the mapping, it is possible to estimate the oscillation of the mapping function $z(w)$ near the boundary of the disc $|w| \leqq 1$ (see [11], p. 403).

[7] The theorem is stated in [10] only for the case $H = 0$. However, the extension to the more general case is immediate.

LEMMA 2. *Subject to the restrictions indicated in footnote* 4, *for any elliptic equation* (6) *there are constants* k_0 and k_1 *such that*

$$p\mathscr{A}+q\mathscr{B} \geqq k_0 \frac{p^2+q^2}{(1+p^2+q^2)^{1/2}} - k_1$$

for all (p, q).

A special case of this lemma is proved in [11], p. 404. Theorem 3 follows immediately from Lemma 2 and a simple Green's identity.

LEMMA 3. *If* $\mathscr{A}$ *and* $\mathscr{B}$ *are suitably restricted functions* (*see footnote* 6) *and if* $R_0 \sim R_4$, *then* $R_2 \sim R_3$, *that is,* $\varphi+i\psi$ *is quasi-conformal as a function of* $u+iv$.

This result can be obtained as a consequence of Lemma 2. Together with Lemma 1 it implies Theorem 7 (see [13]).

LEMMA 4. *If* $D = 0$ *and if* $a\gamma+c\alpha-2b\beta \leqq K < \infty$, *then the* (*bounded*) *function* $(p-iq)/[1+(1+p^2+q^2)^{1/2}]$ *is quasi-conformal in* $u+iv$ *for any solution of* (1).

A proof of Lemma 4 appears in [11], p. 406. Lemmas 1 and 4 imply Theorem 1; together with Lemma 2 they yield Theorem 4. Theorem 2 is based on Lemma 4 but requires some further discussion that cannot be given here.

The proof of Theorem 8 given by the author is based on considerations other than those of this paper. An approach which is closer to the spirit of the present context has been given by Bers [6]. However, an extension of Theorem 8 to the case where H is a function of the unknown φ, obtained recently by Howard Jenkins and the author, does not seem amenable to the method of Bers.

Bibliography

[1] Bernstein, S. N., *Sur la généralisation du problème de Dirichlet*, Math. Ann., Vol. 69, 1910, pp. 82–136.

[2] Bernstein, S. N., *Sur les équations du calcul des variations*, Ann. Sci. École Norm. Sup., Vol. 29, 1912, pp. 431–485.

[3] Bernstein, S. N., *Sur un théorème de géométrie et ses applications aux équations aux dérivées partielles du type elliptique*, Comm. Inst. Sci Math Méc. Univ Kharkoff, Vol. 15, 1915–1917, pp. 38–45.

[4] Bers, L., *Isolated singularities of minimal surfaces*, Ann Math., Vol. 53, 1951, pp. 364–386.

[5] Bers, L., *Univalent solutions of linear elliptic systems*, Comm. Pure Appl. Math., Vol. 6, 1953, pp. 513–526.

[6] Bers, L., *Function-theoretical properties of solutions of partial differential equations of elliptic type*, Ann. Math. Studies No 33, Princeton University Press, 1954, pp. 69–94.

[7] Bers, L., *Non-linear elliptic equations without non-linear entire solutions*, J. Rat. Mech. Anal., Vol. 3, 1954, pp. 767–787.

[8] Bers, L. and Nirenberg, L., *On linear and nonlinear elliptic boundary value problems in the plane*, Convegno Internazionale sulle Equazione Derivate e Parziali, Trieste, August 1954, pp. 141–167.

[9] Courant, R., and Hilbert, D., *Methoden der mathematischen Physik*, Vol. II, Springer, Berlin, 1937, pp. 274–277.

[10] Finn, R., *Isolated singularities of solutions of non-linear partial differential equations*, Trans. Amer Math. Soc., Vol. 75, 1953, pp. 385–404.

[11] Finn, R., *On equations of minimal surface type*, Ann. of Math., Vol. 60, 1954, pp. 397–416.

[12] Finn, R., *On a problem of type, with application to elliptic partial differential equations*, J. Rat. Mech. Anal., Vol. 3, 1954, pp. 789–799.

[13] Finn, R., *On the growth of solutions of non-linear elliptic partial differential equations*, Duke Math. J. (to appear)

[14] Hopf, E., *Über den funktionalen, insbesondere den analytischen Charakter der Lösungen elliptischer Differentialgleichungen zweiter Ordnung*, Math. Z., Vol. 34, 1931, pp. 191–233

[15] Lebesgue, H., *Sur le problème de Dirichlet*, Rend. Circ. Mat. Palermo, Vol. 24, 1907, pp. 371–402

[16] Leray, J., and Schauder, J., *Topologie et équations fonctionelles*, Ann. Sci. École Norm. Sup., Vol. 51, 1934, pp. 45–78.

[17] McShane, E. J., *Parametrization of saddle-surfaces, with application to the problem of Plateau*, Trans. Amer. Math. Soc., Vol. 35, 1933, pp. 716–733.

[18] Morrey, C. B., *On the solutions of quasi-linear elliptic equations*, Trans. Amer. Math. Soc., 43, 1938, pp. 126–166

[19] Nirenberg, L., *On nonlinear elliptic partial differential equations and Holder continuity*, Comm. Pure Appl. Math., Vol. 6, 1953, pp. 103–156

[20] Radó, T., *On the Problem of Plateau*, Springer, Berlin, 1933, Chap. IV.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 425–434 (1956)

Difference Methods on a Digital Computer for Laplacian Boundary Value and Eigenvalue Problems*

GEORGE E. FORSYTHE
University of California, Los Angeles

1. Finite Difference Approximations

We begin with

DEFINITION 1. *A simple closed curve* Γ: $x(s)+iy(s)$ $(0 \leqq s \leqq s_t)$ *is said to be piecewise analytic if* $x(s)$ *and* $y(s)$ *are real analytic functions of the arc length* s *of* Γ *in each of a finite number* t *of closed intervals* $0 = s_0 \leqq s \leqq s_1$, $s_1 \leqq s \leqq s_2, \cdots, s_{t-1} \leqq s \leqq s_t$.

Let the open region R of the x, y-plane have a boundary C which is the union of a finite number of piecewise analytic simple closed curves. For real-valued functions $u = u(x, y)$ defined on $R \cup C$, denote by Δu the Laplacian $u_{xx}+u_{yy}$. In numerical computations R, C, u, and Δu are frequently replaced by corresponding quantities associated with a finite difference network. These may be introduced as follows (see [1] for a similar treatment):

For a given *mesh constant* $h > 0$ let a *net* consist of the lines $x = \mu h$, $y = \nu h$, $\mu, \nu = 0, \pm 1, \pm 2, \cdots$. The points $(\mu h, \nu h)$ are called *nodes*. The M nodes which are within R form the *net region* R_h, assumed to be connectable by line segments of the net within R. A point $(\mu h, \nu h)$ of R_h is said to be a *regular interior point* if each of the four *neighboring nodes* $(\mu h \pm h, \nu h \pm h)$ is in $R \cup C$. All other points of R_h are called *irregular interior points*. Certain points P of R_h have some number p $(1 \leqq p \leqq 4)$ of neighboring nodes $P_1, \cdots, P_p$ not in the open region R. For any such point P, let S_k denote the closed segment of the net joining P to P_k. For each $k = 1, \cdots, p$, the point of $S_k \cap C$ closest to P will be called a *boundary point* of the net. The set of boundary points will be denoted by C_h. Some points of C_h may be nodes, but each irregular interior point of R_h has one or more *neighboring points* in C_h which are not nodes. (The definition of C_h given in [2], although intended to be the same, is not precise.)

*The research leading to this paper was sponsored jointly by the Office of Naval Research, project NR-044-144, and the Office of Ordnance Research, project TB2-0001(1210).

Corresponding to u, defined on $R \cup C$, is a real-valued net-function $u = u_h$ defined on $R_h \cup C_h$. Corresponding to Δu is $\Delta^{(h)}u$ defined as follows. For each point $P : (x, y)$ of R_h, let the neighboring points of $R_h \cup C_h$ be denoted by $(x+h_E, y)$, $(x, y+h_N)$, $(x-h_W, y)$, and $(x, y-h_S)$, where each of the h's is positive ($E \sim$ "east"; $N \sim$ "north", etc.). Let $b_E = h_E/h$, $b_N = h_N/h$, etc. For regular interior points $b_E = b_N = \cdots = 1$. Let $u(x, y) = u_P$, $u(x+h_E, y) = u_E$, $u(x, y+h_N) = u_N$, etc.

Define the *residual* r_P at P by

$$(1) \quad r_P = \frac{u_E}{b_E(b_E+b_W)B} + \frac{u_W}{b_W(b_E+b_W)B} + \frac{u_N}{b_N(b_N+b_S)B} + \frac{u_S}{b_S(b_N+b_S)B} - 4u_P,$$

where

$$B = \frac{1}{4b_N b_S} + \frac{1}{4b_E b_S}.$$

Then

$$(2) \quad \Delta^{(h)}u = r_P/h^2.$$

(The definition of $\Delta^{(h)}u$ in [2] differs by the factor $2B$.) In case P is a regular interior point, (1) reduces to the usual formula

$$(3) \quad r_P = u_E+u_W+u_N+u_S-4u_P$$

When there are irregular interior points, the operator $\Delta^{(h)}$ is not self-adjoint over R_h. For treating the eigenvalue problem it is often more convenient to deal with the self-adjoint part of $\Delta^{(h)}$, denoted here and in [2] by Δ_h.

2. Asymptotic Inequalities for All Eigenvalues

For the operator Δ, the eigenvalue problem

$$(4) \quad \begin{aligned} \Delta u+\lambda u &= 0 \quad \text{in } R, \\ u &= 0 \quad \text{on } C \end{aligned}$$

is known to have an infinite sequence of linearly independent eigenfunctions $u^{(k)}$ and corresponding eigenvalues $\lambda^{(k)}$, numbered so that

$$(5) \quad 0 < \lambda^{(1)} < \lambda^{(2)} \leqq \lambda^{(3)} \leqq \cdots.$$

Similarly, the eigenvalue problem

$$(6) \quad \begin{aligned} \Delta_h u+\lambda_h u &= 0 \quad \text{on } R_h, \\ u &= 0 \quad \text{on } C_h \end{aligned}$$

has a finite sequence of linearly independent eigenfunctions $u_h^{(k)}$ and eigenvalues $\lambda_h^{(k)}$, where

$$0 < \lambda_h^{(1)} < \lambda_h^{(2)} \leqq \lambda_h^{(3)} \leqq \cdots \leqq \lambda_h^{(M)}.$$

(Recall that M is the number of points of R_h.)

For rectangles whose sides are lines of the net, it is well known that

$$\lambda_h^{(k)} < \lambda^{(k)}, \qquad k = 1, \cdots, M; \tag{7}$$

see also below. Since for general regions R close lower bounds for the $\lambda^{(k)}$ are not easily found, the author has been investigating [3, 5a, 2, 4] whether (7) might also hold for other regions.

Suppose one deals with a polygonal region R which is a finite union of squares and half-squares of a net of mesh constant $2h$. Then R_{2h} has no irregular interior points, and neither does R_h, obtained by refining the net. At any point (x, y) of R_{2h},

$$\Delta_{2h} u = \Delta_h u + \tfrac{1}{4} h^2 L_h u, \tag{8}$$

where $L_h u = h^{-4}\{u_{xx\bar{x}\bar{x}} + u_{yy\bar{y}\bar{y}}\}$. Here

$$u_{xx\bar{x}\bar{x}} = u(x-2h, y) - 4u(x-h, y) + 6u(x, y) - 4u(x+h, y) + u(x+2h, y),$$

and $u_{yy\bar{y}\bar{y}}$ is defined analogously. Following the idea [3] of putting the fundamental eigenfunction $u_h = u_h^{(1)}$ of the problem for R_h into the minimization problem for R_{2h}, one finds that

$$\lambda_{2h}^{(1)} \leqq \lambda_h^{(1)} - \tfrac{1}{4} h^2 \left(\sum_{(2h)} u_h L_h u_h\right) \left(\sum_{(2h)} u_h^2\right)^{-1}, \tag{9}$$

where $\sum_{(2h)}$ denotes a sum over the points of R_{2h}. Hence whenever

$$\sum_{(2h)} u_h L_h u_h > 0 \tag{10}$$

it follows that

$$\lambda_{2h}^{(1)} < \lambda_h^{(1)}. \tag{11}$$

If (11) should hold also for h replaced by $h/2$, $h/4$, $h/8$, $\cdots$, it would then follow that $\lambda_h^{(1)} < \lambda^{(1)}$, since the methods of [5] show that $\lambda_h^{(1)} \to \lambda^{(1)}$, as $h \to 0$.

Recent computation with SWAC (see Section 4 below) for the L-shaped region of [3] shows that (11) fails for $h = 1/16$. Hence (10) is not true for all polygonal regions of the type considered here. I conjecture, however, that (10) holds for any *convex* R which is the union of squares and half-squares of the net.

Lacking a proof of this conjecture, the author proved [3] an asymptotic form of (7) for $k = 1$, and in [5a] and [2] generalized it to include certain regions with curved boundaries (the case $k = 1$ of Theorem 2 below). Thus, under the hypotheses of Theorem 2 below, the desired inequality (7) has already been proved for the first eigenvalue, though only for sufficiently small h. Less delicate inequalities of type $\lambda_h^{(1)} < \lambda^{(1)}$, valid for all h, have since been found by Weinberger [6, 7] and Hersch [8] without such stringent hypotheses on R, by allowing R_h to correspond to a region a little larger than

R. Assuming $x(s)$, $y(s)$ defining Γ to be seven times continuously differentiable, Saul'ev [8a] has recently found inequalities of the type $|\lambda_h^{(k)}-\lambda^{(k)}| = O(h^{2-\epsilon})$, as $h \to 0$.

During the 1955 Berkeley Conference for Partial Differential Equations, at which this paper was presented, the author completed an extension of the results of [2] to all eigenvalues of Δ and Δ_h. This has been abstracted in [4], and will now be summarized. Let τ be the angle between the x-axis and the positive direction of the tangent to C. For any function v, let v_n denote the inner normal derivative of v on C. For any two functions v, w, define an inner product (v, w) by the formula

$$12(v, w) = \iint_R (v_{xx}w_{xx}+v_{yy}w_{yy})\,dx\,dy+\int_C v_n w_n \sin^2 2\tau\,d\tau.$$

Let the infinite symmetric matrix D have elements $d_{ij} = (u^{(i)}, u^{(j)})$, $i, j = 1, 2, \cdots$, where $u^{(k)}$ is an eigenfunction of (4) belonging to $\lambda^{(k)}$, normalized so that $\iint_R (u^{(k)})^2\,dx\,dy = 1$. Corresponding to an m-fold multiple eigenvalue of Δ, we are to take m linearly independent functions $u^{(k)}$.

Define a sequence of real numbers $\gamma^{(k)}$, $k = 1, 2, \cdots$, as follows. If $\lambda^{(k)}$ is a simple eigenvalue of (4), take $\gamma^{(k)} = d_{kk}$. If $\lambda^{(k)} = \lambda^{(k+1)} = \cdots = \lambda^{(k+m-1)}$ is an m-fold multiple eigenvalue of (4), then define $\gamma^{(k)} \geqq \gamma^{(k+1)} \geqq \cdots \geqq \gamma^{(k+m-1)}$ to be the m (real) eigenvalues of (d_{ij}), $i, j = k, \cdots, k+m-1$, the corresponding m-rowed principal minor of D.

DEFINITION 2. *A region R bounded by a finite number of piecewise analytic curves is said to have strictly convex corners if, at any point $x(s_j)+iy(s_j)$ where distinct analytic curves meet (see Definition 1), the interior angle is less than π (equality not permitted).*

THEOREM 1. *Let R, an open region with strictly convex corners, be bounded by C, the union of a finite number of piecewise analytic, simple closed curves. Define h, $\lambda^{(k)}$, $\lambda_h^{(k)}$, $\gamma^{(k)}$ as above. Then for $k = 1, 2, \cdots$ one has $-\infty < \gamma^{(k)} < \infty$ and, as $h \to 0$,*

$$\lambda_h^{(k)} \leqq \lambda^{(k)}-\gamma^{(k)}h^2+o(h^2). \tag{12}$$

A proof of Theorem 1 will appear in a later paper. If R is also convex, then C is a simple closed curve with $d\tau \geqq 0$, and D is therefore a Gramian matrix. Hence all $\gamma^{(k)} > 0$, and we have the following corollary to Theorem 1:

THEOREM 2. *Let R, a convex open region with strictly convex corners, be bounded by C, a piecewise analytic, simple closed curve. Then, with the above notation, $0 < \gamma^{(k)} < \infty$, and for each $k = 1, 2, \cdots$ there exists $h_0(k) > 0$ such that $\lambda_h^{(k)} < \lambda^{(k)}$ if $h \leqq h_0(k)$.*

The case $k = 1$ of Theorems 1 and 2 is the result proved in [2]. (In Theorem 1 of [2] the assumption that R be simply connected was never used. Note: The proof of Lemma 1 in [2] is incorrect.)

As an asymptotic result Theorem 1 is the best possible, in the following sense. For a rectangle R of sides π/p, π/q, the relation (12) is actually an equality for all k, up to the term $o(h^2)$. For, corresponding to the eigenfunction $u^{(m,n)} = \sin mpx \sin nqy$ of both (4) and (6) (it is convenient here to use a double superscript m, n instead of the single superscript k), one has

$$\lambda^{(m,n)} = (mp)^2+(nq)^2;$$

and, as $h \to 0$ over values of h for which R is a union of squares of the net,

$$\begin{aligned} \lambda_h^{(m,n)} &= \frac{\sin^2 (mph/2) + \sin^2 (nqh/2)}{(h/2)^2} \\ &= \lambda^{(m,n)} - \frac{(mp)^4+(nq)^4}{12} h^2+o(h^2) \\ &= \lambda^{(m,n)} - (u^{(m,n)}, u^{(m,n)})h^2/12+o(h^2). \end{aligned} \tag{13}$$

Since $(u^{(m,n)}, u^{(\mu,\nu)}) = 0$ unless $|m-\mu|+|n-\nu| = 0$, one has $\gamma^{(m,n)} = (u^{(m,n)}, u^{(m,n)})/12$, even for multiple eigenvalues. Hence (13) shows that (12) is an equality for the rectangle R, up to the term $o(h^2)$.

3. Solving the Difference Equations on a Machine

Let us first consider solving the *Dirichlet difference problem*

$$\Delta^{(h)}u = 0 \qquad \text{on } R_h, \tag{14a}$$

$$u = \text{prescribed function on } C_h \tag{14b}$$

on an automatic digital computer. Assume there are no irregular interior points of R_h, so that (14a) is equivalent to

$$r_P = u_E+u_N+u_W+u_S-4u_P = 0 \qquad \text{on } R_h. \tag{14c}$$

To solve the system of M equations (14c) in the unknown values of u at the M points of R_h, iterative methods are usually preferable to elimination methods, because of their simplicity. The *successive overrelaxation* process of [10] and [9] has proved very successful. In it the points of R_h are repeatedly scanned in a cyclic order. In its turn, at each point P of R_h the value u_P is immediately replaced by a new value u'_P,

$$u'_P = u_P+\tfrac{1}{4}\beta r_P, \tag{15}$$

where β $(1 \leqq \beta < 2)$ is a parameter to be discussed below. Using the new value of u_P, one goes on to the next point of R_h, and so on. For $\beta = 1$ this

is the *Liebmann process* [11], in which $u'_P = \frac{1}{4}(u_E+u_N+u_W+u_S)$. Any reasonable initial values $u_\beta^{(0)}$ of the function u over R_h will suffice, and one may take all $u_\beta^{(0)} = 0$.

Let $u_\beta^{(k)}$ be the result of k sweeps through R_h with successive overrelaxation. Let $u^{(\infty)}$ be the exact solution of (14); define the *error* $e_\beta^{(k)} = u_\beta^{(k)} - u^{(\infty)}$. Since the solution process is linear, one knows that

$$e_\beta^{(k)} = H_\beta^k e_\beta^{(0)},$$

where H_β is a certain linear transformation of R_h into itself. The convergence of $e_\beta^{(k)}$ to 0, as $k \to \infty$, depends on the maximum modulus of the eigenvalues $\lambda_i(H_\beta)$ of H_β. For the Dirichlet problem it is known [9] that $|\lambda_i(H_\beta)| < 1$ (all i, $0 < \beta < 2$), so that successive overrelaxation ($1 \leqq \beta < 2$) always converges. The ultimate speed of convergence depends on $\max_i |\lambda_i(H_\beta)|$. Now, for any single step (15), to choose $\beta = 1$ is the best *tactic*, in the sense that the negative definite error function $\sum_{R_h} e\Delta^{(h)}e$ is brought as close to the desired value 0 as is possible in the one step, by choosing $\beta = 1$ in (15). However, hand computers have long observed [12] that certain choices $\beta > 1$ apparently represent a better *strategy* than $\beta = 1$, in the sense that $e_\beta^{(k)} \to 0$ faster asymptotically, as $k \to \infty$. (Related ideas of strategy and tactics for other linear processes will be found in [13].) In [10] and [9] the matter of a best strategy is settled by showing how to pick the unique β_0 for which $\max_i |\lambda_i(H_\beta)|$ is minimized. Indeed,

$$\beta_0 = \frac{2}{1+\sqrt{1-\sigma}}, \tag{16}$$

where

$$\sigma = \max_i |\lambda_i(H_1)|. \tag{17}$$

The number σ is easily approximated by setting $\beta = 1$ and estimating the asymptotic value of $||e_1^{(k+1)}||/||e_1^{(k)}||$, as $k \to \infty$, where $||\cdot||$ denotes some convenient norm function. When $\beta = \beta_0$, all the $|\lambda_i(H_\beta)|$, $i = 1, \cdots, M$, turn out to be equal, so that the manner of approach of $||e_{\beta_0}^{(k)}||$ to 0 is utterly unpredictable.

When R has a curvilinear boundary C, the theory of [9] still holds, and the successive overrelaxation method is just as attractive. One simply replaces the formula (3) for r_P at irregular interior points P of R_h by the formula (1).

For the eigenvalue problem (6) successive overrelaxation is readily adaptable to getting $\lambda_h^{(1)}$ and its associated eigenvector $u_h^{(1)}$. One uses a trial value λ for $\lambda_h^{(1)}$, and solves the system

$$\begin{aligned} (\Delta_h+\lambda I)u &= 0 \quad \text{on } R_h, \\ u &= 0 \quad \text{on } C_h \end{aligned} \tag{18}$$

by successive overrelaxation starting with some nonzero vector $u_\beta^{(0)}$. For the residual, assuming that R_h has only regular interior points (so that $\Delta_h \equiv \Delta^{(h)}$), one uses

$$r_P = u_E + u_N + u_W + u_S - 4u_P + \lambda u_P \tag{19}$$

in (15). If λ were exactly equal to $\lambda_h^{(1)}$, the iterates $u_\beta^{(k)}$ would stay roughly in the affine subspace S (of M-dimensional space) which is orthogonal to $u_h^{(1)}$ and passes through $u_\beta^{(0)}$. Apart from an exceptional set of $u_\beta^{(0)}$ of zero measure, one would always have $u_\beta^{(k)} \to \alpha u_h^{(1)}$, as $k \to \infty$, where α is some nonzero number depending on $u_\beta^{(0)}$. If λ is near $\lambda_h^{(1)}$ but not equal to it, one finds that $u_\beta^{(k)}$ moves asymptotically into the line $\alpha u_h^{(1)}$ while converging to 0 or ∞. An improved value of λ is then obtainable from the Rayleigh quotient $\varrho_h(u_\beta^{(k)})$, where

$$\varrho_h(v) = -\Big(\sum_{R_h} v \Delta_h v\Big)\Big(\sum_{R_h} v^2\Big)^{-1}. \tag{20}$$

One then uses the better λ in (19), and alternates a few steps of successive overrelaxation with the Rayleigh quotient calculation, until a satisfactory $\alpha u_h^{(1)}$ is obtained.

For getting $\lambda_h^{(k)}$ and $u_h^{(k)}$ $(k > 1)$, the idea of solving (18) with λ near $\lambda_h^{(k)}$ is undoubtedly still a good one, but successive overrelaxation no longer converges, because $\Delta_h + \lambda_h^{(k)} I$ is not semidefinite for $k > 1$. A simple iteration which can solve indefinite systems must be substituted—perhaps that of Kaczmarz [14].

4. SWAC Codes

SWAC is an electronic digital computer built by the National Bureau of Standards, but now owned by the Department of Defense and loaned to the University of California, Los Angeles, for research in numerical analysis. It is a three or four address machine with a Williams tube memory of 256 words; each word consists of 36 binary digits and a sign. Add time is .000064 second; multiplication time is slightly less than six add times. An auxiliary magnetic drum memory holds 4096 words in 128 channels of 32 words each (increased to 8192 words in late 1955). One can transfer a block of up to 32 words between the drum and the Williams tube memory in .017 second. For substantial problems, computing time is measured in hours, while coding time is measured in weeks.

In connection with the author's graduate seminar on numerical methods for elliptic partial differential equations, held in the spring of 1955, a few general codes were prepared following the methods of Section 3 above.

In all these codes the net region R_h is represented by a subset of the

32 by 128 rectangular array of cells on SWAC's magnetic drum. Parameters indicate the smallest rectangle in which R_h is contained. In all the codes the data in successive drum channels are brought to the Williams tube memory for the relaxation process, and the improved values are sent back when they are no longer needed.

The author's basic Code 60 solves Dirichlet's problem for the operator $\Delta^{(h)} = \Delta_h$ for a region with no irregular interior points. Here C_h consists solely of nodes of the net, and can thus also be represented by cells of the drum. Points of R_h have a 0 in the most significant binary digit. Points not in R_h are tagged with a 1 in the same digit, and can be an arbitrary subset of the 32 by 128 rectangular array. Code 60 assumes that the values of u have been placed in the drum cells representing C_h, in digits less significant than the first (which is assumed to be tagged with a 1, as stated above). It performs successive overrelaxation at all points of R_h, leaving all other values of u unchanged. The overrelaxation factor β defined in Section 3 can be modified at will by the SWAC operator. (It is fun for the operator to participate in the solution. Having the operator manage β, while SWAC manages all other numbers, seems a satisfactory division of labor!) The error measure used, denoted by $||e_\beta^{(k)}||$, is the first power norm of $r^{(k)}$; it is displayed at the end of each sweep through R_h. It is easily accumulated during the algorithm (15), since

$$||e_\beta^{(k)}|| = \sum_{R_h} |r_P^{(k)}|.$$

The ratio $||e_\beta^{(k)}|| / ||e_\beta^{(k-1)}||$ is also displayed, and is useful, for example, in approximating the σ of (17).

When R_h is a rectangle of 30 by 68 interior points, a relaxation sweep through the 2040 points takes SWAC 8.5 seconds. When $\beta = 1$, to reduce $e_1^{(k)}$ by a factor of 10^{-6} would take about 2300 sweeps—over five hours on SWAC. Using the optimal value $\beta = \beta_0 = 1.858+$, only 90 sweeps would be needed—about 13 minutes on SWAC. Including the time necessary to estimate that 1.875 is a pretty good value for β, and running with this β, a reduction of 10^{-6} in $e_\beta^{(k)}$ is found in practice to take 20 or 30 minutes (see [15]).

Code 61 is a modification of Code 60 in which boundary conditions corresponding to $\partial u/\partial n = u_n = 0$ are permitted at any point of C_h. Suppose, for example, that it is desired to have $\partial u/\partial x = 0$ for $x = 0$, where R is a region with $x > 0$. Then values of $u(x, y)$ for the lines $x = -h$, $x = 0$, $x = h$, $x = 2h$, $\cdots$ are all stored on the drum for appropriate values of y. When the code comes to $u(-h, y)$ for any y, the value is simply replaced by $u(h, y)$. The values $u(0, y)$, $u(h, y)$, $u(2h, y)$, $\cdots$ are all relaxed according to (15). Again the relaxation routine is notified of the different types of interior and boundary points by suitable identification digits.

Mr. Kenneth Ralston of U.C.L.A. has prepared a SWAC code analogous to Code 60, dealing with curvilinear boundaries. Note in this case that it is no longer possible to map C_h onto the drum in a simple manner. For example, the point of C_h "north" of the point $(\mu h, \nu h)$ of R_h may not be the same as the point of C_h "west" of $(\mu h+h, \nu h+h)$. Identification digits distinguish: (a) points not in R_h, to be left alone, (b) regular interior points, to be treated by (15), with r_P as in (3), and (c) irregular interior points, to be treated by (15), with r_P as in (1).

To deal with irregular interior points Ralston stores the coefficients of u_E, u_W, u_N, u_S in formula (1) in four cells of certain special channels of the drum. Since one meets the irregular interior points of R_h in a fixed order, all these special coefficients can be stored once and for all in the proper sequence. (Such of the u_E, u_W, u_N, u_S as are fixed boundary values are actually multiplied into the coefficients stored in the special channels.) Ralston can also use the special channels to simulate a general mixed boundary condition $au_n+bu+c = 0$ on a straight boundary segment parallel to the x or y axis.

The author's Code 70 performs successive overrelaxation for the eigenvalue problem (6) over regions with no irregular interior points, following the method described in Section 3. The quantities β and $4-\lambda$ are both parameters of the code. Auxiliary to Code 70 are three other codes: Code 80 doubles the values of u over R_h, while Code 82 halves these values of u. Code 81 forms the Rayleigh quotient (20).

With Codes 70—82 the author found the fundamental eigenvalue for the L-shaped region of [3] for various h, and found the counter-example mentioned after (11). Code 61 has been adapted for some soil engineering calculations by Mr. D. Isherwood and Mrs. Louise Straus of U.C.L.A. Code 60 has been modified by Miss Rita Powers at U.C.L.A. to study the influence of the order in which the points of R_h are relaxed on the rate of convergence of successive overrelaxation.

To make Ralston's code more available for problems, a code is badly needed to tag the drum cells with the identification digits, to precompute the coefficients (1), and to store the latter on the special channels. Ralston and the author have outlined such a code, assuming C to be a piecewise quadratic function of x, y. Another needed code is one to get $\lambda_h^{(k)}$, for $k > 1$.

Using the ORACLE computer at Oak Ridge National Laboratories, W. C. Sangren [unpublished] has also computed $\lambda_h^{(1)}$ for the L-shaped region of [3] and for other simple regions with rectilinear boundaries.

Bibliography

[1] Kantorovič, L. V., and Krylov, V. I., *Približennye Metody Vysšego Analiza*, 4th Edit., Gostchizdat, Moscow-Leningrad, 1952.

[2] Forsythe, G. E., *Asymptotic lower bounds for the fundamental frequency of convex membranes*, Pacific J. Math., Vol. 5, 1955, pp. 691–702.

[3] Forsythe, G. E., *Asymptotic lower bounds for the frequencies of certain polygonal membranes*, Pacific J. Math., Vol. 4, 1954, pp. 467–480.

[4] Forsythe, G. E., *Asymptotic lower bounds for all the eigenvalues of a convex membrane*, Bull. Amer. Math. Soc. Abstract, Vol. 61, 1955, p. 551.

[5] Courant, R., Friedrichs, K. O., and Lewy, H., *Über die partiellen Differenzengleichungen der mathematischen Physik*, Math. Ann., Vol. 100, 1928, pp. 32–74.

[5a] Forsythe, G. E., *Arbitrarily close lower bounds for the fundamental frequency of a convex membrane*, Bull. Amer. Math. Soc. Abstract, Vol. 59, 1953, p. 551.

[6] Weinberger, H. F., *Lower bounds for minima by finite difference methods*, Bull. Amer. Math. Soc. Abstract, Vol. 61, 1955, p. 301.

[7] Weinberger, H. F., *Upper and lower bounds for eigenvalues by finite difference methods*, Comm. Pure Appl. Math., Vol. 9, 1956 (this issue) pp. 000–000.

[8] Hersch, J., *Équations différentielles et fonctions de cellules*, C. R. Acad. Sci. Paris, Vol. 240, 1955, pp. 1602–1604.

[8a] Saul'ev, V. K., *K voprosu rešeniya zadači o sobstvennyh značeniyah metodom konečnyh raznostei*, Vyčisl. Mat. Vyčisl. Tehn., Vol. 2, 1955, pp. 116–144.

[9] Young, D., *Iterative methods for solving partial difference equations of elliptic type*, Trans. Amer. Math. Soc , Vol. 76, 1954, pp. 92–111 (condensation of his Harvard dissertation 1950).

[10] Frankel, S. P., *Convergence rates of iterative treatments of partial differential equations*, Math. Tables and Other Aids to Computation, Vol. 4, 1950, pp 65–75.

[11] Liebmann, H., *Die angenäherte Ermittlung harmonischer Funktionen und konformer Abbildung*, S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss., Vol. 47, 1918, pp. 385–416.

[12] Fox, L., *A short account of relaxation methods*, Quart J. Mech. Appl. Math., Vol. 1, 1948, pp. 253–280.

[13] Stiefel, E., *Relaxationsmethoden bester Strategie zur Lösung linearer Gleichungssysteme*, Comment. Math. Helv., Vol 29, 1955, pp. 157–179.

[14] Kaczmarz, S., *Angenaherte Auflosung von Systemen linearer Gleichungen*, Bull. Int. Acad. Polon. Sci. Cl. Sci. Math. Nat Ser. A Sci. Math., 1937, pp. 355–357.

[15] Forsythe, G E , *Relaxation methods*, Section 125 of I. S. Sokolnikoff, *Mathematical Theory of Elasticity*, McGraw-Hill, New York, 1956.

Received January 20, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 435–445 (1956)

On the Local Behavior of Solutions of $\Delta u = g(x, u, \nabla u)$

PHILIP HARTMAN

The Johns Hopkins University

1. Introduction

This note is a resumé of a paper "On the local behavior of solutions of non-parabolic partial differential equations. III.", by Wintner and myself, which will appear shortly in the American Journal of Mathematics.[1] The results of the paper concern the asymptotic behavior, near a zero or an isolated singularity, of functions $u = u(x)$ subject to a differential inequality of the form

$$|\Delta u| \leqq \text{const.}\ (|\nabla u| + |u|), \tag{1}$$

where Δu is the Euclidean Laplacian and ∇u the gradient of u. For example, (1) holds if u is the solution of a non-linear partial differential equation

$$\Delta u = g(x, u, \nabla u), \tag{2}$$

where

$$|g(x, u, \nabla u)| \leqq \text{const.}\ (|\nabla u| + |u|); \tag{3}$$

in particular, if u is the solution of a linear equation

$$\Delta u + \lambda \cdot \nabla u + \mu u = 0. \tag{4}$$

where $\mu = \mu(x)$ is a continuous scalar, $\lambda = \lambda(x)$ is a continuous vector and a dot denotes scalar multiplication.

For the sake of simplicity, the results will be stated and the proofs indicated for the case (4). It will also be assumed that the position vector x is 3-dimensional; it will be clear that analogous results hold if x is a vector of any number ($\geqq 2$) of dimensions.

The theorems to be obtained are known in the case when the number of independent variables in (4) is 2, that is, when x is a binary vector ([4, 5]; for earlier results and references to the literature, see Bers[1]). In this case,

---

[1]Note added May 15, 1956: This paper has appeared in Amer. J. Math., Vol. 77, 1955, pp. 453–474. For other results of the type discussed here, see L. Bers, Local behavior of solutions of general linear elliptic equations, Comm. Pure Appl. Math., Vol. 8, 1955, pp. 473–496, and a forthcoming paper of N. Aronszajn (Abstract 184, Bull. Amer. Math. Soc., Vol. 62, 1956, p. 154).

Δu in (4) can be replaced by a more general linear combination of second order derivatives with sufficiently smooth coefficients, since this more general situation can be reduced to (4) by a conformal mapping. The problem of replacing Δ by a more general operator, when x is not a binary vector, will be left open.

In the plane, the results are obtained by applications of a device of Carleman [3]. This device depends on the use of the Cauchy-Riemann equations and cannot be applied in space. The methods to be employed below are a refinement (and at the same time, a simplification) of a method used by C. Müller [7] to obtain an extension of Carleman's uniqueness theorem from the plane to space. This method depends on the use of Fourier expansions for $u = u(x)$ on spheres $|x| = \text{const.}$ in terms of spherical harmonics. The desired asymptotic formulae are first proved in the sense of the L^2-mean.

2. A Zero of $u(x)$

For the sake of simplicity, it will be supposed that $\lambda(x)$ *and* $\mu(x)$ *are continuous vector and scalar functions, respectively, on the closed sphere* $|x| \leqq 1$. This assumption will not be repeated below. By a C^1-solution $u = u(x)$ of (4) one means that $u(x)$ is of class C^1 and that $u(x)$ is the sum of a harmonic function and of the Newtonian potential of $(\lambda \cdot \nabla u + \mu u)/4\pi$.

(I) *Let* $u = u(x)$ *be a* C^1*-solution of* (4) *on* $|x| \leqq 1$ *satisfying*

$$u(x) = o(|x|^n) \quad \text{as} \quad x \to 0, \qquad n = 0, 1, \cdots. \tag{5}$$

Then $u(x) \equiv 0$ *on* $|x| \leqq 1$.

This can be considered to be an extension of Müller's [7] uniqueness theorem. Müller assumes an inequality of the form $|\Delta u| \leqq \text{const.}\,|u|$, instead of the more general inequality (1) implied by (4). Actually, Müller assumes an integrated form of the inequality $|\Delta u| \leqq \text{const.}\,|u|$ but it will be clear, from the proof of (I), that (1) can be replaced by an integrated form of (1). The proof of (I) below avoids Müller's procedure of extending the domain of definition of $u(x)$ from $|x| \leqq 1$ to the entire x-space.

Theorem (I) and its proof contain the uniqueness theorem of Heinz [6], in which (5) is replaced by the assumption that $u(x) \equiv 0$ on some small sphere $|x| \leqq \varepsilon$. Lemmas 1, 2 and 3 below give a differentiated (hence a sharper) form of the main inequalities of Heinz ([6] Satz 1, p. 7).

The main result on the behavior of solutions of (4) near a zero is the following:

(II) *Let* $u = u(x) \not\equiv 0$ *be a* C^1*-solution of* (4) *on* $|x| \leqq 1$ *satisfying*

$$u(0) = 0. \tag{6}$$

Then there exist a positive integer N *and a spherical harmonic* $S_N(e) \not\equiv 0$ *of*

order N satisfying

$$u(x) = r^N S_N + o(r^N), \qquad \nabla u(x) = \nabla(r^N S_N) + o(r^{N-1}), \tag{7}$$

as $r = |x| \to 0$, *uniformly in* e.

In (II) and in the sequel $x = re$, when $r = |x|$ and e is a unit vector. It will be clear from the proof that the remainder terms $o(r^N)$, $o(r^{N-1})$ in (7) can be improved to $O(r^{N+1}|\log r|)$, $O(r^N|\log r|)$, respectively.

An immediate consequence of the second of the asymptotic relations (7) is the following

COROLLARY. *Under the assumptions of (II), the zeros of* $\nabla u(x)$ *cannot cluster at* $x = 0$.

It may be mentioned that if one assumes that λ and μ are of class C^1 (so that u is of class C^2), then the asymptotic formula for u in (7) can be "differentiated" twice.

3. Isolated Singularities

The next three theorems deal with a C^1-solution $u = u(x)$ of (4) on a punctured sphere $0 < |x| \leqq 1$. The first theorem concerns "removable" singularities:

(III) *Let* $u = u(x)$ *be a* C^1*-solution of* (4) *on* $0 < |x| \leqq 1$ *and let*

$$u(x) = o\left(\frac{1}{r}\right) \text{ as } r = |x| \to 0. \tag{8}$$

Then $u(x)$ *can be defined at* $x = 0$ *so as to be of class* C^1 *on* $|x| \leqq 1$ (*that is, at* $x = 0$ *also*).

Note that there is no assumption on ∇u in (III).

The next theorem concerns the uniqueness of the Green singularity:

(IV) *Let* $u = u(x)$ *be a* C^1*-solution of* (4) *on* $0 < |x| \leqq 1$ *and let*

$$u(x) \to \infty \text{ as } r = |x| \to 0. \tag{9}$$

Then there exists a constant $c > 0$ *satisfying, as* $r \to 0$,

$$u(x) = \frac{c}{r} + O(|\log r|), \qquad \nabla u(x) = \nabla\left(\frac{c}{r}\right) + O\left(\frac{1}{r}\right). \tag{10}$$

Theorems (III) and (IV) are known under more severe smoothness assumptions on λ and μ (Bôcher [2]). Actually, Bôcher assumes that λ and μ are analytic, but his proof is applicable if, for example, λ and μ are of class C^2.

In contrast to the proofs of (I), (II) above and of (V) below, the proof of (IV) will not depend on Fourier analysis, but on a certain type of Green's identity (which can be considered to be a generalization of Carleman's planar identity [3]).

Theorem (III) will depend on (IV) and a corresponding existence

theorem which states that on a sufficiently small punctured sphere $0 < |x| \leqq \varepsilon$, the equation (4) has a C^1-solution $u = G(x)$ satisfying $G(x) \sim 1/r$ as $r = |x| \to 0$ (Lichtenstein; cf. [5] for references and indications of a proof of the corresponding theorem in the plane). Theorem (IV) and the existence theorem just mentioned furnish estimates for ∇u as $x \to 0$ in (III). The proof of (III) then follows by a variant of the proof of (II). The details will not be given in this resumé.

The last theorem deals with "pole-like" singularities:

(V) *Let $u = u(x)$ be a C^1-solution of* (4) *on $0 < |x| \leqq 1$ satisfying, as $r \to 0$,*

$$(11) \qquad u(x) = o(r^{-N-2}), \qquad \nabla u(x) = o(r^{-N-3}) \text{ for some integer } N \geqq 0.$$

Then there exists a spherical harmonic $S_N(e)$ of order N (possibly $S_N(e) \equiv 0$) satisfying, as $r \to 0$,

$$u = r^{-N-1} S_N + O(r^{-N}|\log r|), \quad \nabla u = \nabla(r^{-N-1} S_N) + O(r^{-N-1}|\log r|),$$

uniformly in e.

The proof of this theorem is also similar to that of (II) and will not be given in this note.

4. Preliminaries

When it is convenient, a function $u(x)$ of x will be considered as a function $u(r, e)$ of $r = |x|$ and $e = x/|x|$ if $x \neq 0$. For a fixed r, the function $u(r, e)$ is a function on the unit sphere

$$(12) \qquad B\colon |e| = 1.$$

The symbol $||\cdots||$ will refer to the L^2-norm on (12), so that $||u(r, e)||$ is a function of r. The element of area on B will be denoted by $d\sigma$ and geographical coordinates on B by (θ, φ) (so that $x = (r\cos\varphi \sin\theta, r \sin\varphi \sin\theta, r\cos\theta)$ and, for example, $d\sigma = \sin\theta\, d\theta\, d\varphi$).

Let $S_{nj}(e)$ be the spherical harmonic

$$(13) \qquad S_{nj} = c_{nj} P_{nj}(\theta) \cos j\varphi \text{ or } S_{n\,j+n} = c_{nj} P_{nj}(\theta) \sin j\varphi, \quad n = 0, 1, \cdots,$$

for $j = 0, 1, \cdots, n$ or $j = 1, \cdots, n$, respectively, where the $P_{nj}(\theta)$ are the Legendre functions and c_{nj} are normalizing factors. Thus S_{nj}, where $j = 0, 1, \cdots, 2n$ and $n = 0, 1, \cdots$, is a complete orthonormal sequence on B.

Let $f(x)$ be a continuous function on $|x| \leqq 1$ and let $u = u(x)$ be a C^1-solution of the Poisson equation

$$(14) \qquad \Delta u = f.$$

Let $F_{nj}(r)$ and $U_{nj}(r)$ denote the Fourier constants

$$(15_1) \qquad F_{nj}(r) = \int_B f(r, e) S_{nj}(e) d\sigma,$$

$$(15_2) \qquad U_{nj}(r) = \int_B u(r, e) S_{nj}(e) d\sigma$$

of f and u, respectively. Thus u has the Fourier series

$$(16) \qquad u(r, e) \sim \sum_{n=0}^{\infty} \sum_{j=0}^{2n} U_{nj}(r) S_{nj}(e);$$

$u^N(r, e)$ will denote the partial sum

$$(17) \qquad u^N(r, e) = \sum_{n=0}^{N} \sum_{j=0}^{2n} U_{nj}(r) S_{nj}(e).$$

The Parseval relation for f can be written as

$$(18) \qquad \Phi(r) = ||f(r, e)||^2 = \sum_{n=0}^{\infty} \sum_{j=0}^{2n} |F_{nj}(r)|^2.$$

The main lemmas given below will deal with estimates for U_{nj} and for $||u(r, e) - u^N(r, e)||^2$ in terms of $\Phi(r)$. The starting point for the proof of these lemmas will be the fact that U_{nj} satisfies the singular, inhomogeneous differential equation

$$(19) \qquad (r^2 U')' - n(n+1)U = r^2 F,$$

where $' = d/dr$, $U = U_{nj}$ and $F = F_{nj}$. For a given F, the general solution of (19) is

$$(20) \qquad (2n+1)U = r^n \left(A - \int_r^1 s^{1-n} F(s)\, ds\right) - r^{-1-n} \left(B - \int_r^1 s^{n+2} F(s)\, ds\right),$$

where $A = A_{nj}$, $B = B_{nj}$ are integration constants.

The continuity of $u(x)$ implies that $U_{nj}(r)$ is continuous at $r = 0$ in which case (20) reduces to

$$(21) \qquad (2n+1)U = r^n \left(A - \int_r^1 s^{1-n} F(s)\, ds\right) - r^{-1-n} \int_0^r s^{n+2} F(s)\, ds.$$

Since $u(r, e)$ is of class C^1, the partial derivative $u_r = \partial u/\partial r$ has the Fourier coefficient

$$(22) \qquad U_{nj}'(r) = \int_B u_r(r, e) S_{nj}(e)\, d\sigma$$

obtained by differentiating (21):

$$(23) \qquad (2n+1)U' = nr^{n-1} \left(A - \int_r^1 s^{1-n} F(s)\, ds\right) + (n+1) r^{-2-n} \int_0^n s^{n+2} F(s)\, ds.$$

On putting $r = 1$ in (21) and (23), it follows that the integration constant A is given by

$$(24) \qquad A = (n+1)U(1) + U'(1), \qquad A = A_{nj},\ U = U_{nj},$$

so that Bessel's inequality implies the estimate

$$\sum_{n=0}^{\infty}\sum_{j=0}^{2n}\frac{A_{nj}^2}{(n+1)^2}\leqq(||u||^2+||u_r||^2)_{r=1}<\infty. \tag{25}$$

In addition to the Fourier constants U, U' of u, u_r, the Fourier constant

$$V_{nj}(r)=\int_B u_\varphi(r,e)S_{nj}(e)\,d\sigma \tag{26}$$

of $u_\varphi=\partial u/\partial\varphi$ will be needed. Since $d\sigma=\sin\theta\,d\theta\,d\varphi$, a partial integration shows that

$$V_{nj}(r)=-\int_B u(r,e)\,S_{nj\,\varphi}(e)\,d\sigma.$$

It is clear from the definition (13) of S_{nj} that $S_{nj\,\varphi}=-j\,S_{n\,j+n}$ if $j=0,\cdots,n$ and that $S_{nj\,\varphi}=(j-n)S_{n\,j-n}$ if $j=n+1,\cdots,2n$; thus

$$V_{nj}=\begin{cases}-j\,U_{n\,j+n} & \text{for } 0\leqq j\leqq n,\\ (j-n)U_{n\,j-n} & \text{for } n+1\leqq j\leqq 2n.\end{cases} \tag{27}$$

5. Estimates for Large N

In view of (21),

$$\tfrac{1}{3}(2n+1)^2U^2\leqq A^2r^{2n}+r^{2n}\left(\int_r^1 s^{1-n}F(s)\,ds\right)^2+r^{-2(1+n)}\left(\int_0^r s^{n+2}F(s)ds\right)^2. \tag{28}$$

If Schwarz's inequality is applied to each of the last two integrals, one obtains, after some simple calculation,

$$\tfrac{1}{3}(2n+1)^2U^2\leqq A^2r^{2n}+(2N-2\alpha-1)^{-1}r^{2\alpha+3}\int_0^1 s^{-2\alpha}F^2(s)ds\ (\leqq\infty) \tag{29}$$

if $j=0,1,\cdots,2n$ and $n=N+1,N+2,\cdots$, where $N>\alpha+\frac{1}{2}$. If the inequalities (29) are summed for $j=0,1,\cdots,2n$ and $n=N+1,N+2,\cdots$, then Parseval's relation leads to the following

LEMMA 1. *Let $f(x)$ be a continuous function on $|x|\leqq 1$ and let $u(x)$ be a C^1-solution of the Poisson equation* (14) *on $|x|\leqq 1$. Let α be a (non-negative) number and N an integer satisfying*

$$N>\alpha+\tfrac{1}{2}. \tag{30}$$

Then there exists a constant C (depending only on bounds for $|u|$ and $|\nabla u|$ on $|x|=1$, and independent of N, α) such that

$$||u-u^N||^2\leqq C\chi(r)\ \text{ for }\ 0<r\leqq 1, \tag{31}$$

where

$$\chi(r)=r^{2N+2}+(2N-2\alpha-1)^{-1}r^{2\alpha+3}\int_0^1 s^{-2\alpha}\Phi(s)\,ds\ (\leqq\infty). \tag{32}$$

A modification of the arguments leading to this result yields

LEMMA 2. *Under the conditions of Lemma* 1 *on* u, f, N *and* α, *there exists a constant* C (*depending only on the bounds for* $|u|$ *and* $|\nabla u|$ *on* $|x| = 1$, *and independent of* N, α) *such that*

$$r^2||u_r - u_r^N||^2, \ ||u_\varphi - u_\varphi^N||^2 \leqq C\chi^*(r) \text{ for } 0 < r < 1, \tag{33}$$

where

$$\chi^*(r) = N^2 r^{2N+2}(1-r^2)^{-3} + (2N-2\alpha-1)^{-1} r^{2\alpha+3}\int_0^1 s^{-2\alpha}\Phi(s)ds \ (\leqq \infty). \tag{34}$$

6. The Main Lemmas

Let N be an integer and γ a number (not necessarily an integer) subject to the inequalities

$$2 \leqq N \leqq \gamma + 2. \tag{35}$$

The estimates to be obtained for small n will be based on assumptions that, as $r \to 0$,

$$\Phi(r) = o(r^{2\gamma}), \tag{36}$$

$$||u(r, e)|| = o(r^{N-1}). \tag{37}$$

In view of Schwarz's inequalities and (15), (16), these assumptions imply that, as $r \to 0$,

$$F_{nj}(r) = o(r^\gamma), \tag{38}$$

$$U_{nj}(r) = o(r^{N-1}). \tag{39}$$

In view of (38),

$$r^{-1-n}\int_0^r s^{n+2}F(s)\,ds = o(r^{2+\gamma}) \text{ as } r \to 0; \tag{40}$$

hence, in view of (21), (39) and $\gamma + 2 \geqq N$,

$$I = o(1) \text{ if } n < N, \text{ where } I = I_{nj}(r) = A - \int_r^1 s^{1-n}F(s)\,ds. \tag{41}$$

Thus, if $n < N$, the relation (21) becomes

$$(2n+1)U = r^n\int_0^r s^{1-n}F(s)\,ds - r^{-1-n}\int_0^r s^{n+2}F(s)\,ds. \tag{42}$$

Arguments, similar to those used to obtain (29), give

$$(2n+1)^2U^2 \leqq (3-2n+2\beta)^{-1}r^{2\beta+3}\int_0^r s^{-2\beta}F^2(s)\,ds \tag{43}$$

if $n < N < \beta + 3/2$.

Summing the inequalities (43) for $n = 0, 1, \cdots, N-1$ and $j = 0, 1, \cdots, 2n$ and applying Parseval's relation yield the part of the following lemma dealing with u^{N-1}; the parts involving u_r^{N-1}, u_φ^{N-1} are proved similarly.

LEMMA 3. *Let the conditions on* u, f *of Lemma* 1 *hold. In addition, let there*

exist an integer N and a number γ satisfying (35)—(37). *Finally, let β be a number satisfying*

$$N-1 < \beta + \tfrac{3}{2} < \gamma+2. \tag{44}$$

Then, for $0 < r \leqq 1$,

$$||u^{N-1}||^2, \quad r^2||u_r^{N-1}||^2, \quad ||u_\varphi^{N-1}||^2 \leqq \chi^{**}(r), \tag{45}$$

where

$$\chi^{**}(r) = (5-2N+2\beta)^{-1} r^{2\beta+3} \int_0^r s^{-2\beta} \Phi(s)\, ds. \tag{46}$$

In order to complete the set of estimates furnished by Lemmas 1—3, note that (21) and (35), (36) imply, for $r \to 0$,

$$U_{Nj}(r) = O(|\log r|) = o(r^\varepsilon) \tag{47}$$

for every $\varepsilon > 0$. If the last part of assumption (35) is strengthened to

$$2 \leqq N < \gamma+2, \tag{48}$$

then (47) can be improved to

$$\lim_{r\to 0} r^{-N} U_{Nj}(r) = B_{Nj} \text{ exists for } j = 0, \cdots, 2N. \tag{49}$$

In this case, define a spherical harmonic $S_N(e)$ of order N by

$$S_N(e) = \sum_{j=0}^{2N} B_{Nj} S_{Nj}(e). \tag{50}$$

Lemmas 1, 3 and the relations (47) or (49), (50) are used to obtain the following

LEMMA 4. *Let u, f, N and γ satisfy the conditions of Lemma* 3. *Then, for every $\varepsilon > 0$,*

$$||u|| = o(r^{N-\varepsilon}) \text{ as } r \to 0. \tag{51}$$

If (35) *is strengthened to* (48), *then* (49) *holds and* (51) *can be strengthened to*

$$||u - r^N S_N|| = o(r^N) \text{ as } r \to 0, \tag{52}$$

where $S_N = S_N(e)$ is the spherical harmonic (50).

Also, Lemmas 2, 3 and variants of (47), (49), (50) imply that the asymptotic relations (51), (52) can be "differentiated".

LEMMA 5. *Under the assumptions of Lemma* 4 *leading to* (51),

$$||\nabla u|| = o(r^{N-1-\varepsilon}) \text{ as } r \to 0; \tag{53}$$

under the assumptions leading to (52),

$$||\nabla u - \nabla(r^N S_N)|| = o(r^{N-1}) \text{ as } r \to 0. \tag{54}$$

Actually, (53) and (54) are first derived for the cases when ∇ is replaced

by $\partial/\partial r$ or by $\partial/\partial\varphi$. Note that

$$ru_r = u_x x+u_y y+u_z z, \qquad u_\varphi = u_y x-u_x y. \tag{55}$$

In these formulae, (x, y, z) represent the components of the vector x. The asymptotic formulae derived for u_φ have analogues for

$$u_\psi = u_x z-u_z x. \tag{56}$$

From (55), (56), we obtain

$$r^2 u_x = xru_r-yu_\varphi+zu_\psi . \tag{57}$$

This relation and the asymptotic formulae for u_r , u_φ , u_ψ lead to corresponding formulae for u_x . The partial derivatives u_y , u_z are treated in a similar way to obtain Lemma 5.

7. Proof of (II)

The Lemmas 1—5 are made applicable to solutions of (4) by writing (4) as a Poisson equation (14), where

$$f(x) = -(\lambda(x) \cdot \nabla u(x)+\mu(x)u(x)). \tag{58}$$

An outline of the proof of (II) will now be given; the detailed calculations will be omitted.

The assumption (6) and the fact that u is of class C^1 imply that, for $N = 1$, there exists a spherical harmonic $S_1(e)$ (of order 1) satisfying (7), as $r \to 0$, uniformly in e. If $S_1 \not\equiv 0$, then the proof of (II) is complete. If $S_1(e) \equiv 0$, then, for $N = 2$,

$$u(x) = o(r^{N-1}), \qquad \nabla u = o(r^{N-2}). \tag{59}$$

Assume (59) for a fixed $N \geqq 2$. The first relation in (59) gives (37); the relations (58) and (59) imply (35), (36) with $\gamma = N-2$. Applications of the first parts of Lemmas 4 and 5 imply (36) with $\gamma = N-1-\varepsilon$, where $\varepsilon > 0$ is arbitrary. Then applications of the second parts of Lemmas 4 and 5 imply the existence of an $S_N(e)$ (possibly $S_N \equiv 0$) satisfying (52) and (54).

The relations (52), (54) and the representation of u as a solution of the Poisson equation (14) in terms of the Green and Poisson kernels for $|x| < R$ and $|x| = R$, respectively, permit the improvement of (52), (54) to (7).

The existence of an N for which (7) holds and for which, in addition, $S_N(e) \not\equiv 0$ then follows from (I), which remains to be proved.

8. Proof of (I)

The reasoning just completed shows that assumption (5) of (I) implies that, as $r \to 0$,

$$u(x) = o(r^n), \qquad \nabla u(x) = o(r^{n-1}) \text{ for } n = 0, 1, \cdots. \tag{60}$$

Hence, Lemmas 1—3 and the arguments involving (55)—(57) can be used to

show that, for every N,

$$\Phi(r) \leqq C\left\{\frac{N^2 r^{2N-2}}{(1-r^2)^3} + r^{2N-1}\int_0^1 s^{2-2N}\Phi(s)\,ds\right\}, \tag{61}$$

since $\Phi(r) \leqq \text{const.}\{||\nabla u||^2+||u||^2\}$ by (58). The constant C in (61) depends on the bounds of u, ∇u, λ and μ, but is independent of N. If (61) is multiplied by r^{2-2N}, a quadrature gives

$$(1-\tfrac{1}{2}Cr^2)\int_0^r s^{2-2N}\Phi(s)\,ds \leqq C\left\{\frac{N^2 r}{\varepsilon^3} + \tfrac{1}{2}r^2\int_r^1 s^{2-2N}\Phi(s)\,ds\right\} \tag{62}$$

for $0 \leqq r \leqq (1-\varepsilon)^{\frac{1}{2}}$.

It follows that $\Phi(r) \equiv 0$ on the sphere $|x| \leqq r$ if $1-\frac{1}{2}Cr^2 > 0$ and hence $f \equiv 0$ there. But then u is harmonic, and thus (60) implies that $u \equiv 0$ on this sphere. In view of the (uniform) estimate for the radius of the sphere involved, assertion (I) follows.

9. Proof of (IV)

The proof of (IV) is analogous to the proof of the corresponding planar theorem [5], pp. 351—352. A (local) standard type of existence theorem and a variation of constants show that there is no loss of generality in supposing that $\mu = 0$, so that (4) reduces to

$$\Delta u + \lambda \cdot \nabla u = 0. \tag{63}$$

This implies that (II) is applicable to the function $u(x)-u(x_0)$ for every choice of x_0. Hence, the Corollary of (II) implies that $\nabla u(x)$ has only a finite number of zeros in $|x| \leqq \text{const.} < 1$. In particular, (9) shows that there exists a sequence of constants satisfying $C_n \to \infty$ as $n \to \infty$ and such that the locus

$$L = L_C : u(x) = C \tag{64}$$

is a closed Jordan C^1-surface surrounding $x = 0$. Furthermore, the inward normal derivative u_p on L satisfies

$$u_p \geqq 0 \quad \text{on } L. \tag{65}$$

In what follows, T will denote a volume with a boundary S, p will denote the outward unit normal vector on S.

Let ξ be a 3-dimensional vector and let $x-\xi$ have the components (X, Y, Z). Set

$$v = v(x, \xi) = \frac{X}{\varrho^3}, \qquad \varrho = |x-\xi| \tag{66}$$

(i.e., v is the first component of $-\nabla_\xi(1/\varrho)$). Let $w = w(x, \xi)$ denote the vector potential

$$w = \left(0, \frac{Z}{\varrho^3}, -\frac{Y}{\varrho^3}\right), \tag{67}$$

so that

$$\nabla \times w = \nabla v, \qquad \nabla = \nabla_\xi. \tag{68}$$

If T has a smooth boundary S and if the points $\xi = 0$ and $\xi = x$ are not in $T+S$, then Green's identity gives

$$\int_S \nabla u(\xi) \cdot (vp + w \times p)\, d_\xi S = \int_T v\, \Delta u(\xi)\, d_\xi T. \tag{69}$$

Since (68) holds, terms of the type $\nabla v \cdot \nabla u$ and $\nabla u \cdot (\nabla \times w)$ are eliminated from the volume integral in (69).

Choose $T = T(C, \varepsilon)$ to be the region in the ξ-space bounded by $B : |\xi| = 1$, $L : u(\xi) = C$ and $|x-\xi| = \varepsilon$. It is readily verified that the contribution of $|x-\xi| = \varepsilon$ to the left side of (69) satisfies

$$\int_{|x-\xi|=\varepsilon} \nabla u \cdot (vp + w \times p) ds \to 4\pi u_x(x) \quad \text{as } \varepsilon \to 0,$$

where u_x is the first component of ∇u.

Since L is a level surface of u, it follows that $\nabla u \cdot (w \times p) = 0$ on L. Hence, if $T = T(C, 0)$, (69) reduces to

$$4\pi u_x(x) = -\int_T v\, \Delta u\, dT - \int_L v\, u_p\, dS + \int_B \nabla u \cdot (vp + w \times p) dS.$$

This relation, the corresponding relations for u_y, u_z and Green's identity

$$0 = \int_T \Delta u\, dT - \int_L u_p\, dS - \int_B \nabla u \cdot p\, dS$$

are the analogues of (22) and (28) in [5]. The remainder of the proof of (IV) now follows from the arguments of [5], pp. 356—357.

Bibliography

[1] Bers, L., *Theory of pseudo-analytic functions*, Lecture Notes, New York University, 1953.

[2] Bôcher, M., *Singular points of functions which satisfy partial differential equations of the elliptic type*, Bull Amer. Math. Soc., Vol. 9, 1903, pp. 455–465.

[3] Carleman, T., *Sur les systèmes linéaires aux dérivées partielles du premier ordre à deux variables*, C. R. Acad. Sci Paris, Vol 197, 1933, pp. 471–474.

[4] Hartman, P., and Wintner, A., *On the local behavior of solutions of non-parabolic partial differential equations*, Amer. J. Math., Vol. 75, 1953, pp. 449–476.

[5] Hartman, P., and Wintner, A., *On the local behavior of solutions of non-parabolic partial differential equations, II. The uniqueness of the Green singularity*, Ibid., Vol. 76, 1954, pp. 351–361.

[6] Heinz, E., *Über die Eindeutigkeit beim Cauchyschen Anfangswertproblem einer elliptischen Differentialgleichung zweiter Ordnung*, Nachr. Akad. Wiss. Gottingen Math.-Phys. Kl. IIa, 1955, No. 1.

[7] Muller, C., *On the behavior of the solutions of the differential equations $\Delta U = F(x, U)$ in the neighborhood of a point*, Comm. Pure Appl. Math., Vol. 7, 1954, pp. 505–516.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 447–466 (1956)

The Scope and Limitations of the Method of Wiener and Hopf*

ALBERT E. HEINS
Carnegie Institute of Technology

1. Introduction

We shall discuss here a chapter in boundary value problems which arise in the solution of certain elliptic partial differential equations. These boundary value problems may be formulated as integral equations of the Wiener-Hopf type—an integral equation which was first discussed seriously in 1931 and the application of which was first made in 1944.

An integral equation of the Wiener-Hopf type has the general form

$$\lambda f(x)+\int_0^\infty f(y)K(x-y)dy = g(x)$$

where $g(x)$ and $K(x)$ are known functions of x, and $f(x)$ is unknown. The homogeneous equation was discussed by Wiener and Hopf in 1931. The non-homogeneous case appears to be tractable once the homogeneous case is known. The kernel $K(x)$ of the integral equation as well as the function $g(x)$ are subject to certain order conditions at infinity which we shall discuss later. It was more than a decade after the fundamental work of Wiener[1] and Hopf that this equation found some interesting and important applications to some boundary value problems in the study of acoustics, hydrodynamics and electromagnetic theory. These problems distinguish themselves by two outstanding features. First they involve the solution of the partial differential equation

$$\frac{\partial}{\partial\alpha}p\frac{\partial u}{\partial\alpha}+\frac{\partial^2 u}{\partial\beta^2}+k^2u = 0, \qquad p = p(\alpha), \tag{1.1}$$

where k is a constant not necessarily real,[2] but, for the present, not zero. The

*This paper is based on research conducted in part under a contract between the Office of Ordnance Research of the Department of the Army and the Carnegie Institute of Technology.

[1]Professor Wiener informed the writer in 1944 that he had been aware of one such application before this date

[2]Actually k may depend on α, but we shall not discuss this aspect here. Equation (1.1) may be generalized to the form

$$\frac{\partial}{\partial\alpha}p\frac{\partial u}{\partial\alpha}+\frac{\partial}{\partial\gamma}q\frac{\partial u}{\partial\gamma}+\frac{\partial^2 u}{\partial\beta^2}+k^2u = 0$$

where now $k = k(\alpha,\gamma)$, $p = p(\alpha,\gamma)$ and $q = q(\alpha,\gamma)$.

boundary conditions under which this equation is to be solved are of the Sturm-Liouville type for the variable α and, for the variable β, are appropriately prescribed at infinity by the particular boundary-value problem at hand. In practice the most common boundary conditions are the vanishing of u, of the α-derivative of u, or a linear combination of such terms.

The second feature is the geometry. Instead of having to solve equation (1.1) over a closed surface of an infinite cylinder whose generators are in the direction of the β-axis, we prescribe boundary conditions on a *semi-infinite cylinder* or a *half-plane*. Thus, the elementary methods available for solution for the infinite cylindrical surfaces are not of any use here because the boundary conditions are assigned on part of a coordinate surface, that is, we are concerned with mixed boundary value problems. We are thus led to formulate such a problem in terms of an integral equation. The particular surfaces we consider, coupled with the equation (1.1) and the Sturm-Liouville boundary conditions in α, permit, through an application of Green's Theorem with an appropriate Green's function, to formulate such problems as a Wiener-Hopf integral equation. For a kernel K with appropriate properties we have the methods of the Fourier transform in the complex domain and elementary aspects of analytic continuation to help us solve these integral equations. Accordingly we are in a position to find u at any point in space, at least as an integral representation.

In order to keep the computational details down to a minimum, we shall concentrate the major portion of this talk on two simple problems in hydrodynamics and acoustic diffraction. This will serve to indicate the major features of the method, in so far as the formulation as well as the solution of the problem is concerned. We shall also summarize other applications to problems in acoustics and electrodynamics, indicating those features which do not show up in this problem.

In the more elementary problems (geometrically) we note that other methods are available to solve these problems. One such example is the diffraction of a half-plane by the use of multi-valued solutions of the wave equation—a procedure made familiar by Sommerfeld at the turn of the last century. No one appears to have been able to study any of the more general problems which we have described by Sommerfeld's method. Another method is the method of analytic continuation used by Lewy, Stoker and others at New York University. Here again one observes contact with the method we discuss, but no overlap. There are several problems[3] which ultimately can also be reduced to integral equations of the Wiener-Hopf type but which are what we may term accidents in the sense that neither the geo-

[3]Such cases, which involve a change in variables to provide the correct form of the kernel and appropriate limits of integration, are not difficult to recognize as a rule.

metries nor the partial differential equations are appropriate. Such problems as these we will not discuss, but rather we will confine ourselves to the broad outlines regarding the geometry and the particular partial differential equations which we described earlier in this section.

The main tools which we employ are the complex Fourier transform and analytic continuation. In employing the Fourier transform we are limiting the growth at infinity of the various functions which require transforms. That is, at every stage, the unknown and known functions are dominated by the Fourier transforms of $K(x)$ in appropriate regions of the complex plane. Since solutions of equation (1.1) for $|k| \neq 0$ are of this type, the methods will apply here. A second point to be made is that boundary conditions are supplied on a half-line. On occasion, we shall find it necessary to mutilate a function so that these methods will be applicable. The procedure of analytic continuation will be described in Section 3.

2. Water Waves over a Channel of Infinite Depth

A simple physical example which leads to an integral equation of the Wiener-Hopf type is the study of water waves over a channel of infinite depth with a dock. This involves the solution for $y < 0$ of the partial differential equation[4]

$$\phi_{xx}+\phi_{yy}-k^2\phi = 0, \qquad 0 < k < 1, \tag{2.1}$$

subject to the boundary conditions

$$\phi_y = \beta\phi, \qquad y = 0,\ x > 0 \text{ (free surface)}, \tag{2.2}$$

and

$$\phi_y = 0, \qquad y = 0,\ x < 0 \text{ (dock)}. \tag{2.3}$$

It is required to find $\phi(x, y)$ which vanishes at infinity save on the line $y = 0$, $x > 0$, while on this line $\phi(x, 0)$ will be bounded for $x \to \infty$. As we shall see, this formulation will produce a standing wave solution for $x \to \infty$, $y = 0$ if we assume that $\phi(x, y)$ is source free and $\beta^2-k^2 > 0$. In order to produce a traveling wave solution we shall have to find a second solution which is not source free, but this is elementary in this particular case. We observe that had only the boundary condition (2.2) or (2.3) been given on the entire surface $y = 0$, the problem would be trivial, but with the two conditions (2.2) and (2.3) we are compelled to employ more subtle methods.

To this end we note that a Green's function of the form

---

[4]Actually, the three-dimensional linear theory of gravity waves leads to the solution of $\phi_{xx}+\phi_{yy}+\phi_{zz} = 0$, but we have assumed that the propagation normal to the surface wave is not normal to the edge of the dock and hence a factor $\exp\{ikz\}$ may be suppressed from $\phi(x, y, z)$. See T. R. Greene and A. E. Heins [B3].

$$G(x, y, x', y') = \frac{1}{2\pi} K_0 [k\sqrt{(x-x')^2+(y-y')^2}] \tag{2.4}$$

is associated with equation (2.1). Here K_0 is the imaginary Bessel function of the second kind. G satisfies equation (2.1) save at the point $P(x, y) = P'(x', y')$ where it possesses a logarithmic singularity. Furthermore for $r = \sqrt{x^2+y^2} \to \infty$, G vanishes exponentially. Such a function may be used to provide the required representation for the $\phi(x, y)$. Indeed, since ϕ is source free we have

$$\phi(x, y) = \int_{-\infty}^{\infty} [G\phi_{y'} - \phi G_{y'}]dx', \qquad y' = 0. \tag{2.5}$$

In its present form, equation (2.5) brings in boundary conditions on the entire line $y' = 0$. That is,

$$\phi(x, y) = -\int_{-\infty}^{0} \phi G_{y'} dx' + \int_0^{\infty} [\beta G - G_{y'}]\phi dx', \qquad y' = 0. \tag{2.5a}$$

We note, however, that if we replace G by G', where

$$G' = G(x, y, x', y') + G(x, y, x', -y'),$$

we can supply a second possible representation of the form

$$\phi(x, y) = \int_0^{\infty} G' \phi_{y'} dx', \qquad y' = 0 \tag{2.6}$$

in view of the fact that $\partial G'/\partial y' = 0$ at $y' = 0$. Thus (2.6) has incorporated into it the boundary conditions (2.3). If we now employ the boundary condition (2.2), we are left with

$$\phi(x, y) = \beta \int_0^{\infty} G'(x, y, x', 0)\phi(x', 0)dx' \tag{2.7}$$

which yields for $y = 0$ the equation

$$\phi(x, 0) = \beta \int_0^{\infty} G'(x, 0, x', 0)\phi(x', 0)dx'. \tag{2.8}$$

Since $G'(x, 0, x', 0)$ is a function of $x-x'$ and the limits of integration run from zero to infinity, equation (2.8) is an integral equation of the Wiener-Hopf type and of the second kind.

3. The Solution of Equation (2.8)

The solution of equation (2.8) depends on an application of the complex form of the Fourier transform theorem. We observe in this equation a convolution kernel with limits of integration which do not permit a direct use of the convolution theorem. Nevertheless, it is still possible to apply this theorem if we observe that only one part of $\phi(x, 0)$ appears under the integral sign, namely $\phi(x, 0)$ for $x > 0$. Accordingly let us write

$$\phi_1(x, 0) \equiv \begin{cases} \phi(x, 0), & x > 0, \\ 0, & x < 0, \end{cases}$$

and

$$\phi_2(x, 0) \equiv \begin{cases} \phi(x, 0), & x < 0, \\ 0, & x > 0. \end{cases}$$

Then equation (2.8) becomes

$$\begin{aligned} \phi_1+\phi_2 &= \beta \int_0^\infty G'(x, 0, x', 0)\phi_1(x', 0)\,dx' \\ &= \beta \int_{-\infty}^\infty G'(x, 0, x', 0)\phi_1(x', 0)\,dx'. \end{aligned} \tag{3.1}$$

The last equation looks somewhat odd in that our definitions of ϕ_i have cast this apparently from an equation in one unknown to one in two unknowns. We shall see that this split is essential for the application of the Fourier transform theorem.

Suppose now we apply the Fourier transform theorem to equation (3.1). If we write

$$\Phi_1(w) = \int_0^\infty e^{-iwx}\phi_1(x, 0)\,dx, \tag{3.2}$$

$$\Phi_2(w) = \int_{-\infty}^0 e^{-iwx}\phi_2(x, 0)\,dx \tag{3.3}$$

and

$$g'(w) = \int_{-\infty}^\infty e^{-iwx}G'(x, 0, 0, 0)dx \tag{3.4}$$

we would have formally

$$\Phi_1+\Phi_2 = \beta g'\Phi_2\,. \tag{3.5}$$

Equation (3.5) will make sense if there is at least one w for which Φ_1, Φ_2 and g' exist. As we shall see presently, there is actually a strip in the complex w-plane for which all three functions exist. Let us then examine the integrals (3.2), (3.3) and (3.4). In the first place, ϕ is bounded for $x \to \infty$, and for the present let us assume that it is integrable for finite positive x. Then Φ_1 exists and defines a function of w, analytic in the lower half-plane $\mathcal{I}m\ w < 0$ since this inequality defines the ordinate of convergence of the integral. The function $\phi_2(x, 0)$ vanishes exponentially of the order $O(e^{kx})$ as $x \to -\infty$, as one can see readily from equation (3.1), and if we again assume that $\phi_2(x, 0)$ is integrable for finite and negative x, we have that $\Phi_2(w)$ is analytic in the upper half-plane $\mathcal{I}m\ w > -k$. The function $g'(w)$ can be calculated explicitly, and is indeed $1/\sqrt{k^2+w^2}$.[5] Hence equation (3.5) is defined in the strip $-k < \mathcal{I}m\ w < 0$. Upon rearranging this equation we obtain

[5] Cf. [D1] p. 416.

$$\Phi_2(w) = \Phi_1(w)\left[\frac{\beta}{\sqrt{k^2+w^2}} - 1\right]. \tag{3.6}$$

The method of Wiener and Hopf asks whether this equation (3.6) may be rewritten so that each side is analytic in an upper or lower half-plane and both sides are regular in a common strip. Suppose then that it is possible to write $1-\beta/\sqrt{k^2+w^2}$ as $K_-(w)/K_+(w)$ where $K_-(w)$ is analytic in an appropriate lower half-plane and $K_+(w)$ is analytic in an appropriate upper half-plane with both functions analytic in the common strip. Then

$$K_+(w)\Phi_2(w) = -K_-(w)\Phi_1(w), \tag{3.7}$$

and in this form we can determine the functions $\Phi_1(w)$ and $\Phi_2(w)$ but for a polynomial factor. For, now the left side of equation (3.7) analytic in the upper half-plane $\mathcal{I}m\ w > -k$ and the right side is analytic in the lower half-plane $\mathcal{I}m\ w < 0$. Both sides are analytic in a common strip and hence the left side is the analytic continuation of the right side. Therefore, each side is analytic everywhere, that is

$$K_+(w)\Phi_2(w) = -K_-(w)\Phi_1(w) = E(w)$$

where $E(w)$ is an entire function. Before we can determine the entire function $E(w)$ we require the explicit form of the factors K_- and K_+.

The determination of K_- and K_+ can be carried out explicitly by applying the Cauchy integral theorem to $\log[1-\beta/\sqrt{k^2+w^2}]$. That is, we decompose $\log K(w)$ into $\log K_-(w) - \log K_+(w)$. Rather than working with the logarithm directly, we shall use its derivative which will simplify the calculation somewhat. We have

$$\frac{K'(w)}{K(w)} = \frac{K_-'(w)}{K_-(w)} - \frac{K_+'(w)}{K_+(w)} = \frac{1}{2\pi i}\int \frac{K'(t)}{K(t)(t-w)}\,dt$$

where the path of integration in the integral is a rectangle, two of whose sides are parallel to the imaginary axis while the other two sides are parallel to the real axis. The sides parallel to the imaginary axis may be permitted to recede to infinity in view of the order condition on $K'(t)/K(t)$ for $|t| \to \infty$ inside the strip; the sides parallel to the real axis run from $-\infty+i\gamma$ to $\infty+i\gamma$ and $-\infty+i\gamma_1$ to $\infty+i\gamma_1$ respectively, where $-k < \gamma < \mathcal{I}m\ w < \gamma_1 < 0$. The integral over the upper path provides us with a function analytic in the lower half-plane $\mathcal{I}m\ w < 0$ while the integral over the lower path gives us a function analytic in the upper half-plane $\mathcal{I}m\ w > -k$. We find as a result of this calculation that[6]

[6] We shall presently make an application of Liouville's theorem on polynomials. This requires that $K_-(w)$ and $K_+(w)$ be of algebraic order for $|w| \to \infty$, $\mathcal{I}m\ w$ suitably limited For this reason we have added $1/\beta\pi i$ to K'_-/K_- and K'_+/K_+ Otherwise $K_-(w)$ and $K_+(w)$ would be $O(\exp\{w/\beta\pi i\})$, $|w|\to\infty$, $\mathcal{I}m\ w$ suitably limited. More subtle aspects of this notion are observed in many of the applications; see, for example, Carlson and Heins [C2], Levine and Schwinger [C12]

$$\frac{K_-'(w)}{K_-(w)} = -\frac{1}{2(w-ik)} - \frac{1}{\pi\beta i}\frac{d}{dw}\left[(w^2+k^2)f(w)-w\right] + \frac{(w^2+k^2)f(w)}{2\beta\pi i(w+\sigma)} + \frac{(w^2+k^2)f(w)}{2\beta\pi i(w-\sigma)} + \frac{\beta f(\sigma)}{2\pi i(w+\sigma)} + \frac{\beta f(-\sigma)}{2\pi i(w-\sigma)} + \frac{w}{w^2-\sigma^2}.$$

and

$$\frac{K_+'(w)}{K_+(w)} = \frac{1}{2(w+ik)} + \frac{1}{\pi\beta i}\frac{d}{dw}\left[(w^2+k^2)f(-w)+w\right] - \frac{1}{2\beta\pi i(w+\sigma)}\left[(w^2+k^2)f(-w)-\beta^2 f(\sigma)\right] - \frac{1}{2\beta\pi i(w-\sigma)}\left[(w^2+k^2)f(-w)-\beta^2 f(-\sigma)\right]$$

where

$$f(w) = (w^2+k^2)^{-\frac{1}{2}}\log\left[\frac{1+t}{1-t}\right],$$

$$t = \left(\frac{w+ik}{w-ik}\right)^{\frac{1}{2}} \text{ and } \sigma^2 = \beta^2-k^2 > 0.$$

The branches of the radicals have been chosen so that

$$-\frac{3\pi}{2} < \arg(w-ik) < \frac{\pi}{2}$$

and

$$-\frac{\pi}{2} < \arg(w+ik) < \frac{3\pi}{2}.$$

This enables us to find K_- and K_+ up to arbitrary constants of integration. The quotient of the constants can be determined by studying the asymptotic form of K_- and K_+ in the appropriate half-planes of analyticity and comparing it with the asymptotic form of K in its strip of regularity. We observe that

$$K_-(w) = cw\left[1-\frac{2\beta}{\pi i w}\log w\right], \quad |w| \to \infty, \quad \mathcal{I}m\, w < 0$$

and

$$K_+(w) = cw\left[1+\frac{2\beta}{\pi i w}\log(-iw)\right], \quad |w| \to \infty, \quad \mathcal{I}m\, w > -k$$

where c is a constant and $\log -i = -i\pi/2$.

With this information we can determine the entire function $E(w)$. We note that $\Phi_1(w)$ and $\Phi_2(w) \to 0$ for $|w| \to \infty$ and $\mathcal{I}m\, w$ suitably limited, as a

consequence of the Riemann-Lebesgue lemma. In view of this fact, we observe that $E(w)=o(w)$, $|w|\to\infty$ and therefore $E(w)=\alpha$, a constant by Liouville's theorem. This then provides us with Φ_1 and Φ_2 in terms of K_- and K_+. We have therefore $\Phi_1 = \alpha/cw$, $|w|\to\infty$, $\mathcal{I}m\ w<0$ while $\Phi_2(w) = -\alpha/cw$, $|w|\to\infty$, $\mathcal{I}m\ w > -k$. Using the Abelian theorem which relates the transform at infinity to the function at the origin, we see that

$$\phi_1 = -\frac{i}{c}, \qquad x\to 0^+,$$

and

$$\phi_2 = -\frac{i}{c}, \qquad x\to 0^-.$$

It follows that $\phi_1=\phi_2$ at $x=0$ and it is indeed permissible to apply Green's theorem in the neighborhood of the origin.

The determination of $\phi(x,y)$ is possible since we can write the integral representation (2.7) as a complex Fourier integral in virtue of the convolution theorem. Here we have

$$\phi(x,y) = \frac{\beta}{2\pi}\int_C \frac{\exp\{-|y|\sqrt{k^2+w^2}+iwx\}}{\sqrt{k^2+w^2}}\,\Phi_1(w)\,dw$$

where C is a path of integration in the strip $-k<\mathcal{I}m\ w<0$. But $\Phi_1(w)$ is simply α/K_-, so that

$$\phi(x,y) = \frac{\alpha\beta}{2\pi}\int_C \frac{\exp\{y\sqrt{k^2+w^2}+iwx\}}{K_-(w)\sqrt{k^2+w^2}}\,dw. \tag{3.8}$$

The absolute sign on y has been removes since y is always negative. From equation (3.8) it is possible to write representations for ϕ if $x\lesseqgtr 0$. Consider the first case, $x<0$. Here the path C is closed in the lower half-plane taking into account the presence of the branch cut from $-ik$ to $-i\infty$. Since $K_-(w)\sqrt{k^2+w^2}=O(w^2)$, $|w|\to\infty$, $\mathcal{I}m\ w<0$, there are no contributions from the closing path at infinity and we have for $x<0$

$$\phi(x,y) = -\frac{\alpha\beta}{\pi}\int_k^\infty \frac{e^{vx}}{K_-(-iv)}\,\frac{\cos y\sqrt{v^2-k^2}}{\sqrt{v^2-k^2}}\,dv, \tag{3.9}$$

an integral which clearly converges for $x<0$. The y-derivative of ϕ vanishes for $y=0$, $x<0$, thus verifying one boundary condition.

For $x<0$, the path C is closed in the upper half-plane. Here the singularities are two poles at $\pm\sigma$ and there is a branch cut from ik to $i\infty$. The integral is evaluated in the same fashion as it was for the case $x<0$, save for the fact that we now have two simple poles to account for. In order to bring these singularities into prominence, we replace $K_-(w)$ by $K_+(w)[1-\beta/\sqrt{k^2+w^2}]$. Then

$$\phi(x, y) = \frac{\alpha\beta}{2\pi}\int_C \frac{\exp\{y\sqrt{k^2+w^2}+iwx\}}{(\sqrt{k^2+w^2}-\beta)K_+(w)}\,dw$$

$$= -\frac{i\alpha\beta^2}{\sigma}e^{\beta y}\left[\frac{e^{i\sigma x}}{K_+(\sigma)} - \frac{e^{-i\sigma x}}{K_+(-\sigma)}\right] \tag{3.10}$$

$$+ \frac{\alpha\beta}{\pi}\int_k^\infty \frac{[\sqrt{v^2-k^2}\cos y\sqrt{v^2-k^2}+\beta\sin y\sqrt{v^2-k^2}]}{(v^2+\sigma^2)K_+(iv)}e^{-vx}\,dv.$$

Clearly, the integral in equation (3.10) vanishes with $x \to \infty$, $y \leqq 0$ or $y \to -\infty$, $x > 0$, since in the first case the integral is absolutely convergent while in the second case the Riemann-Lebesgue lemma may be applied. It is permissible to differentiate (3.10) to demonstrate for $x > 0$, $y = 0$ that the boundary condition as well as the differential equation are satisfied. Since we have produced a solution which satisfies the conditions permitting the application of Green's theorem and the complex Fourier transform, the method is justified. From the original work of Wiener, there are no other solutions to the integral equation with this order at infinity.

An examination of $K_+(\sigma)$ and $K_+(-\sigma)$ will reveal that they are the same in magnitude, but opposite in phase. This tells us that $\phi(x, y)$ describes a standing wave solution as $x \to \infty$. In order to form a traveling wave solution at infinity, one has only to differentiate ϕ with respect to x, thus not affecting the boundary conditions but introducing a logarithmic singularity at the origin. This produces now a second solution which, at infinity, is out of phase with the first solution and, therefore, an appropriate linear combination of the two will produce the desired traveling wave solution.

In conclusion we note that we study the case $k = 0$ (normal incidence) from the representation (3.9) and (3.10). The methods we have employed for the case $k \neq 0$ are inapplicable because Green's function is simply the logarithm of the distance between the integration coordinates and the observation point, and the Fourier transform of this does not exist. Yet, when we pass to the limit $k = 0$, we produce two new representations which we can show satisfy the partial differential equation and the appropriate boundary conditions; but we shall not pursue this matter here.

4. A Second Formulation of the Problem in Section 2

We were successful in formulating the integral equation in Section 2, because amongst other things we could provide a Green's function which eliminated one of the boundary conditions completely. Let us now examine the representation (2.5) using the boundary condition $\partial G^2/\partial y' = \beta G^2$ for $y = 0$. If such a Green's function could be determined, the integration in (2.5) would now extend from $-\infty$ to 0 and we would have

$$\int_{-\infty}^{0} [G^2\phi_{y'} - \phi G^2_{y'}]dx'$$

for the contribution on the line $y' = 0, x' < 0$. But $\phi_{y'} = 0$ for $y = 0, x < 0$ and $G^2_{y'} = \beta G^2$ for $-\infty < x < \infty$, $y = 0$ so that we are left with

$$-\beta \int_{-\infty}^{0} \phi(x', 0)G^2(x, y, x', 0)\,dx'. \tag{4.1}$$

However, an application of Green's theorem implies a closed path and we are required to discuss the line integral on a path which recedes to infinity in the lower half-plane. Before we can do this, however, we require a knowledge of ϕ and G^2 for $|x| \to \infty$, $y \to -\infty$. For $y \to -\infty$ or $x \to -\infty$, ϕ vanishes exponentially, while for $x \to \infty$, ϕ is asymptotic to $\exp\{\beta y\} \cdot [\varrho_1 \exp\{i\sigma x\} + \varrho_2 \exp\{-i\sigma x\}]$ where ϱ_1 and ϱ_2 are constants. The Green's function which satisfies the boundary conditions $\partial G/\partial y' = \beta G$ on $y' = 0$ is subject to a measure of arbitrariness. Indeed, if we require that the Green's function shall describe outward-going waves on the surface $y' = 0$, that is, G^2 be asymptotic to $\exp\{i\sigma|x|\}$, $|x| \to \infty$, $y = 0$, we shall encounter a situation for which the bilateral Fourier transform is inapplicable. Such a Green's function which obeys the requirement of outward-going waves is asymptotic to

$$\frac{i\beta}{2\sigma} \exp\{\beta(y+y')\} \exp\{i\sigma|x-x'|\}$$

for $|x'| << |x|$. It is, of course, not possible to take the bilateral Fourier transform of such a function. Yet we can modify G^2, with this asymptotic form in mind, so that it can be useful. We can obviously add any solution of the homogeneous equation (2.1), which satisfies the boundary condition for the free surface, without changing any of the major characteristics of the Green's function save the conditions at infinity. Let us then add

$$-\frac{i\beta}{2\sigma} \exp\{\beta(y+y')\} \exp\{i\sigma(x'-x)\}$$

to G^2. Clearly, the modified G^2 will then vanish exponentially for $x' >> x$ and will be asymptotic to

$$-\frac{\beta}{\sigma} \exp\{\beta(y+y')\} \sin \sigma(x-x')$$

for $x >> x'$. We shall denote the modified G^2 by G^{*2}. As a result of the order conditions on ϕ and G^{*2}, we find

$$\phi(x, y) = -\beta \int_{-\infty}^{0} \phi(x', 0)G^{*2}(x, y, x', 0)\,dx'. \tag{4.2}$$

In virtue of the $(x-x')$-dependence of G^{*2}, (4.2) becomes an integral equa-

tion of the Wiener-Hopf type for $y = 0$. (The change of limits is a minor point which can be overcome by replacing x by $-x$).

The bilateral Fourier transform of $G^{*2}(x, 0, x', 0)$ is

$$\frac{1}{\sqrt{k^2+w^2}-\beta},$$

and, in view of the fact that G^{*2} is bounded for $x >> x'$ and exponentially decreasing of the order $\exp\{k(x-x')\}$ for $x << x'$, and is otherwise integrable, this transform exists in the strip $-k < \mathcal{I}m\, w < 0$. We now define ϕ_1 and ϕ_2 as we did in Section 3. Then in this strip of analyticity we have

$$\frac{\beta\Phi_2(w)}{\beta-\sqrt{k^2+w^2}} = \Phi_1+\Phi_2$$

where Φ_1 and Φ_2 are the unilateral Fourier transforms of ϕ_1 and ϕ_2. Upon simplifying this we get

$$\frac{\Phi_2(w)\sqrt{k^2+w^2}}{\beta-\sqrt{k^2+w^2}} = \Phi_1,$$

precisely the same equation which we encountered in Section 3.

5. Acoustic and Electromagnetic Diffraction Problems

Time does not permit us here to pursue in any detail the many problems in acoustic and electromagnetic diffraction which may be formulated as integral equations of the Wiener-Hopf type. Still, a few general comments and one example will serve to show us what is different in these problems from the hydrodynamic problems. Here we are concerned with the solution of the partial differential equation

$$\frac{\partial}{\partial\alpha}p\frac{\partial u}{\partial\alpha}+\frac{\partial^2 u}{\partial\beta^2}+k^2u = 0, \qquad p \equiv p(\alpha), \tag{5.1}$$

The boundary conditions are still of the type which we described in Section 1. The change in sign before k^2 is crucial, since now equation (5.1) may admit solutions which are bounded everywhere at infinity (the so-called plane-wave solutions) as well as solutions which are bounded at infinity in ducts or tubes (the so called wave-guide modes). In the case of the wave-guide modes, the change in sign before k^2 now permits the possibility of a finite series of terms bounded at infinity whose character depends on the geometric character of the wave-guide, the number k, and the boundary condition—essentially the Sturm-Liouville aspect of this problem. This is in direct contrast to the hydrodynamical problem which permits at most one bounded term at infinity. It is not a difficult task to construct the Green's functions for a wave

guide of a regular shape, that is, one for which the Sturm-Liouville problem for the cross-section is known. By imposing boundary conditions on Green's function as we did in Section 3 we can eliminate all boundary contributions along the β-axis save for $\beta > 0$ or $\beta < 0$. Since the β-dependence of Green's function will appear in the form $\beta - \beta'$ we will again have a Wiener-Hopf integral equation. The presence of these bounded terms at infinity is felt in the Fourier transform through the existence of real roots. By adjusting the parameter k we can take into account as many terms as we desire.

For problems involving free space propagation by plane waves, the formulation of the problem is complicated by the presence of the bounded plane-wave terms. In addition to imposing a condition involving the nature of the scattered field at infinity, we are required to describe the incident and reflected plane-waves with care. Their contribution in the formulation of the integral equation is found by evaluating the closing path in Green's theorem with the aid of such an asymptotic device as Kelvin's method of stationary phase. The scattered field will make no contribution, if the appropriate Green's function is chosen.

As a simple example, consider a perfectly conducting plane $-\infty < x < \infty$, $y = 0$, $-\infty < z < \infty$ above which there is a second perfectly conducting semi-infinite plane $-\infty < x < 0$, $y = a$, $-\infty < z < \infty$, parallel to the first one. For a two-dimensional electromagnetic field, which is independent of the z-coordinate, every field component satisfies a partial differential equation of the form

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + k^2 \phi = 0. \tag{5.2}$$

There are scalar problems in electromagnetic theory for which there is one component of the electric field which satisfies a differential equation of the form (5.2) and for which the normal derivative vanishes on the perfect conductors.[7] Outside of the duct formed by these two planes, ϕ is of the order

$$\frac{f(\theta) \exp\{ikr\}}{\sqrt{r}} \tag{5.3}$$

for $r = \sqrt{x^2+y^2} \to \infty$, $x = r \cos\theta$, $y = r \sin\theta$, $0 \leqq \theta \leqq \pi$. That is, there is only an outward going radiated field in free space and this satisfies Sommerfield's radiation condition. Inside the duct, $x \to -\infty$, $0 \leqq y \leqq a$, ϕ is asymptotic to

$$\sum_{n=0}^{M} [A_n \exp\{i\Gamma_n x\} + B_n \exp\{-i\Gamma_n x\}] \cos\frac{n\pi y}{a} + O(e^{Lx}), \qquad L > 0.$$

These are the so-called wave-guide (or parallel plate) modes which remain

[7] This problem is also equivalent to a problem in acoustic diffraction.

bounded as $x \to -\infty$. Here $\Gamma_n^2 = k^2 - n^2\pi^2/a^2$ if $n \leqq M$. Such terms as $\exp\{\pm i\Gamma_n x\}$ describe traveling waves far down the duct. A particular form of excitation of the structure could be obtained by assuming $A_0 = 1, A_n = 0$, $n = 1, \cdots, M$. This would correspond to the lowest mode exciting the duct, causing a series of reflected modes in the duct and a radiated field outside of the duct of the form (5.3).

This problem can be formulated with the aid of two Green's functions which are appropriate for two different regions in the x, y-plane. For example, a Green's function which can be used in conjunction with equation (5.2) and satisfies the boundary conditions $G_{y'} = 0$ at $y' = 0$ and $y' = a$ and, furthermore, describes outward-going radiation from the source point is

$$G^1(x, y, x', y') = \sum_{n=1}^{\infty} \frac{i \cos\frac{n\pi y}{a} \cos\frac{n\pi y'}{a} e^{i\Gamma_n|x-x'|}}{a\Gamma_n} - \frac{ie^{ik|x-x'|}}{2ak}.$$

The positive root Γ_n is understood if Γ_n^2 is positive and otherwise $\Gamma_n = i|\Gamma_n|$. A second Green's function, which satisfies the boundary condition $G^2_{y'} = 0$, $y' = a$ and the radiation condition for $r \to \infty$, $y' > a$, is simply

$$G^2(x, y, x', y') = \frac{i}{4}\{H_0^{(1)}[k\sqrt{(x-x')^2+(y-y')^2}] + H_0^{(1)}[k\sqrt{(x-x')^2+(y+y'-2a)^2}]\}.$$

For the region $-\infty < x < \infty$, $y > a$ we have the following representation for ϕ:

$$\phi = -\int_0^{\infty} G^2(x, y, x', y') \frac{\partial\phi}{\partial y'} dx', \qquad y' = a. \tag{5.4}$$

This representation incorporates the boundary condition on the half-plane and the condition at infinity. In order to attain this condition at infinity we employed the Hankel function $H_0^{(1)}$ rather than $H_0^{(2)}$ which does not have the correct behavior at infinity.

Similarly, for $-\infty < x < \infty$, $0 \leqq y \leqq a$, we have

$$\phi = e^{ikx} + \int_0^{\infty} G^1(x, y, x', y') \frac{\partial\phi}{\partial y'} dx', \qquad y' = a, \tag{5.5}$$

when we employ the boundary conditions on ϕ and G^2 at infinity. The continuity of the representations (5.2) and (5.3) for $y = a$, $x > 0$ gives us an integral equation of the Wiener-Hopf type and of the first kind. That is,

$$e^{ikx} + \int_0^{\infty} [G^1+G^2]\phi_{y'} dx' = \begin{cases} 0, & y = y' = a, \quad x > 0, \\ I(x), & y = y' = a, \quad x < 0, \end{cases} \tag{5.6}$$

where $I(x)$ is the discontinuity of ϕ on the barrier.

From the representation (5.5) it is possible to find the amplitudes of the reflected propogating modes in the duct. Indeed, since we have prescribed the asymptotic form of ϕ, we have from (5.5)

$$B_0 = -\frac{i}{2ak}\int_0^\infty e^{ikx'}\phi_{y'}(x', y')\,dx', \qquad y' = a,$$

$$B_n = \frac{i(-)^n}{a\Gamma_n}\int_0^\infty e^{i\Gamma_n x'}\phi_{y'}(x', y')dx', \qquad y' = a, \qquad n = 1, \cdots, M.$$

It is clear that these amplitudes are the Fourier transforms of $\phi_y(x, a)$ evaluated at $w = -k$ or $-\Gamma_n$. The other behavior at infinity, the angular dependence $f(\theta)$ in (5.3), may be determined from equation (5.4). For, now when $r \to \infty$, $0 \leqq \theta < \pi$,

$$\phi(x, y) = \sqrt{\frac{2}{\pi kr}}\, e^{i(kr-\pi/4)}\int_0^\infty e^{-ikx'\cos\theta}\phi_{y'}(x', a)dx'$$

or

$$f(\theta) = \sqrt{\frac{2}{\pi k}}\, e^{-i\pi/4}\int_0^\infty e^{-ikx'\cos\theta'}\phi_{y'}(x', a)\,dx',$$

again a factor which depends on the Fourier transform of $\phi_y(x, a)$.

In order to simplify our presentation we shall assume that k has a small positive imaginary part.[8] As equation (5.6) stands, it is impossible to apply the complex Fourier transform since the bilateral transform of $\exp\{ikx\}$ does not exist. We can overcome this difficulty by mutilating $\exp\{ikx\}$. We shall rewrite equation (5.6) as

$$(5.7) \qquad \phi_0(x) + \int_0^\infty [G^1 + G^2)]\,\phi_{y'}\,dx' = \begin{cases} 0, & y = y' = a, \quad x > 0, \\ J(x), & y = y' = a, \quad x < 0, \end{cases}$$

where now

$$\phi_0(x) \equiv \begin{cases} \exp\{ikx\}, & x > 0, \\ 0, & x < 0, \end{cases}$$

$$\phi_y(x, a) \equiv 0, \qquad x < 0,$$

$$J(x) \equiv 0, \qquad x > 0.$$

Equation (5.7) now coincides with equation (5.6) for $x > 0$ and, therefore, still expresses the condition of continuity of the ϕ's on the surface $y = a$, $x > 0$. If we accept for the present that ϕ_y and J are integrable for finite x (a fact which we may justify by the solution of the problem) we have the following information: $J(x) = O(e^{-ikx})$, $x \to -\infty$, $\phi_y = O(e^{ikx}/\sqrt{x})$, $x \to \infty$.

[8] A more delicate application of the method of analytic continuation which we employ enables us to avoid this assumption.

This would provide us with the further information that the unilateral Fourier transform of $J(x)$,

$$\int_{-\infty}^{0} J(x)\ \exp\ \{-iwx\}\, dx = \Phi_1(w),$$

is analytic in the upper half-plane $\mathscr{I}m\ w > -\mathscr{I}m\ k$. We also have that

$$\int_0^{\infty} \phi_y(x, a) \exp\{-iwx\}\, dx = \Phi_2(w)$$

is analytic in the lower half-plane $\mathscr{I}m\ w < \mathscr{I}m\ k$. The Fourier transforms of $G^1(x, a, 0, a)$ and $G^2(x, a, 0, a)$ are

$$g^1(w) = \int_{-\infty}^{\infty} G^1(x, a, 0, a) \exp\{-iwx\}\, dx = -\frac{\cos \sigma a}{\sigma \sin \sigma a}$$

and

$$g^2(w) = \int_{-\infty}^{\infty} G^2(x, a, 0, a) \exp\{-iwx\}\, dx = \frac{i}{\sigma},$$

respectively, where $\sigma = \sqrt{k^2 - w^2}$, and we take that branch of σ which reduces to k when $w = 0$. Both g^1 and g^2 are analytic in the strip $-\mathscr{I}m\ k < \mathscr{I}m\ w < \mathscr{I}m\ k$. Finally, the transform of $\phi_0(x)$ is

$$\Phi_0(w) = \int_0^{\infty} \exp\{-iwx\}\, \phi_0(x)\, dx = \frac{1}{i(w-k)}$$

which is analytic in the lower half-plane $\mathscr{I}m\ w < \mathscr{I}m\ k$. We again observe the common strip of analyticity of all the transforms which we employ. We have then

$$\begin{aligned} \Phi_1(w) &= \frac{1}{i(w-k)} + \Phi_2(w)\left[\frac{i}{\sigma} - \frac{\cos \sigma a}{\sigma \sin \sigma a}\right] \\ &= \frac{1}{i(w-k)} - \frac{\Phi_2(w) \exp\{-i\sigma a\}}{\sigma \sin \sigma a}. \end{aligned} \tag{5.8}$$

The factoring of this last equation brings in a number of new features which we have not discussed before. Let us assume that we can factor

$$\frac{\exp\{-i\sigma a\}}{\sigma \sin \sigma a} = \frac{K_-(w)}{K_+(w)}$$

where $K_-(w)$ is analytic and free of zeros in the lower half-plane while $K_+(w)$ is analytic and free of zeros in the upper half-plane.[9] Then equation (5.8) becomes

$$K_+(w)\Phi_1(w) = \frac{K_+(w)}{i(w-k)} - K_-(w)\Phi_1(w). \tag{5.9}$$

[9] $K_-(w)$ and $K_+(w)$ are zero-free because we actually decompose $\log K(w)$ additively (see Section 3).

The first term in equation (5.9) is analytic in the upper half-plane $\mathcal{I}m\, w > -\mathcal{I}m\, k$ and the last term is analytic in the lower half-plane $\mathcal{I}m\, w < \mathcal{I}m\, k$. The center term is still analytic in the strip. However, this term may be decomposed into two terms, each of which is regular in the approriate half-plane, for,

$$\frac{K_+(w)}{i(w-k)} = \frac{K_+(w)-K_+(k)}{i(w-k)} + \frac{K_+(k)}{i(w-k)}.$$

The first term no longer contains a pole at $w = k$ and is, therefore, analytic in the upper half-plane $\mathcal{I}m\, w > -\mathcal{I}m\, k$ while the second terms only contains a pole at $w = k$ and is analytic in the lower half-plane $\mathcal{I}m\, w < \mathcal{I}m\, k$. This enables us then to write equation (5.9) as

$$(5.10)\quad K_+(w)\Phi_1(w) - \frac{K_+(w)-K_+(k)}{i(w-k)} = -K_-(w)\Phi_2(w) + \frac{K_+(k)}{i(w-k)}.$$

In view of the fact that the left side is analytic in the upper half-plane $\mathcal{I}m\, w > -\mathcal{I}m\, k$, the right side is analytic in the lower half-plane $\mathcal{I}m\, w < \mathcal{I}m\, k$ and both sides are analytic in a common strip, the left side is the analytic continuation of the right side. It is therefore analytic everywhere and an entire function $E(w)$.

The success we had in Section 3 in determining $E(w)$ depended on Liouville's theorem on polynomials. The same is true here and we have to show that we can determine $K_-(w)$ and $K_+(w)$ so that they are of algebraic growth as $|w| \to \infty$, $\mathcal{I}m\, w$ suitably limited. In the first place, the factoring of $\exp\{-i\,\sigma a\}$ multiplicatively is equivalent to factoring σ additively. This can be done by Cauchy's theorem, as in Section 3, considering the function $1/\sigma$ to insure convergence of the integrals. We have then, by direct computation,

$$\frac{1}{\sigma} = \frac{2}{\pi\sigma}\arctan\sqrt{\frac{k+w}{k-w}} + \frac{2}{\pi\sigma}\arctan\sqrt{\frac{k-w}{k+w}}$$

where the first term is analytic in the lower half-plane $\mathcal{I}m\, w < \mathcal{I}u\, k$ and the second term is analytic in the upper half-plane $\mathcal{I}m\, w > -\mathcal{I}m\, k$. It follows then that

$$\exp\{-i\,\sigma a\} = \exp\{L_-(w)\}\exp\{L_+(w)\}$$

where

$$L_-(w) = -\frac{2ia\sigma}{\pi}\arctan\sqrt{\frac{k+w}{k-w}}$$

and

$$L_+(w) = -\frac{2ia\sigma}{\pi}\arctan\sqrt{\frac{k-w}{k+w}}.$$

For $|w| \to \infty$, $\mathcal{I}m\, w < \mathcal{I}m\, k$, $L_-(w)$ is asymptotic to

$$-\frac{iaw}{\pi} \log\left[-\frac{2w}{k}\right]$$

while for $|w| \to \infty$, $\mathcal{I}m\, w > -\mathcal{I}m\, k$, $L_+(w)$ is asymptotic to

$$\frac{iaw}{\pi} \log \frac{2w}{k}.$$

We now turn to the factoring of $\sigma \sin \sigma a$ which is accomplished by the Weierstrass product representation and states that

$$\sigma \sin \sigma a = a(k^2 - w^2) \prod_{n=1}^{\infty} \left[1 - \frac{a^2(k^2 - w^2)}{n^2\pi^2}\right].$$

The factor

$$M_-(w) = (k-w) \prod_{n=1}^{\infty} \left[\sqrt{1 - \frac{a^2k^2}{n^2\pi^2}} + \frac{iaw}{n\pi}\right] e^{-iaw/n\pi}$$

is free of zeros in the lower half-plane $\mathcal{I}m\, w < \mathcal{I}m\, k$ while

$$M_+(w) = M_-(-w)$$

is free of zeros in the upper half-plane $\mathcal{I}m\, w > -\mathcal{I}m\, k$. The exponential factor $\exp\{-iaw/n\pi\}$ has been inserted into all the products to render them absolutely convergent. Hence

$$aM_-(w)M_+(w) = \sigma \sin \sigma a$$

and

$$K_-(w) = \frac{\exp\{L_-(w)\}}{M_-(w)},$$

while

$$K_+(w) = aM_+(w) \exp\{-L_+(w)\}.$$

This provides us with the basic factors $K_-(w)$ and $K_+(w)$. For $|w| \to \infty$, $\mathcal{I}m\, w < \mathcal{I}m\, k$,

$$M_-(w) = \frac{\exp\{-iaw\gamma/\pi\}}{\Gamma(iaw/\pi)}$$

(γ is the Euler-Mascheroni constant) up to a constant of proportionality since the products in this limit can be shown to be independent of ak. Similarly, for $|w| \to \infty$, $\mathcal{I}m\, w > -\mathcal{I}m\, k$,

$$M_+(w) = \frac{\exp\{iaw\gamma/\pi\}}{\Gamma(-iaw/\pi)}.$$

Upon examining the asymptotic forms of $K_-(w)$ and $K_+(w)$ we find that for $|w| \to \infty$, $\mathcal{I}m\, w < \mathcal{I}m\, k$,

$$K_-(w) = O\left[\frac{1}{\sqrt{w}} \exp\left\{\frac{iaw}{\pi}\left(\gamma - 1 - \log\frac{2\pi}{ak}\right) + \frac{aw}{2}\right\}\right],$$

while for $|w| \to \infty$, $\mathscr{I}m\, w > -\mathscr{I}m\, k$, we have

$$\frac{1}{K_+(w)} = O\left[\frac{1}{\sqrt{w}} \exp\left\{\frac{iaw}{\pi}\left(1 - \gamma + \log\frac{2\pi}{ak}\right) - \frac{aw}{2}\right\}\right].$$

The exponential factors in K_- and K_+ prevent the application of Liouville's theorem which we employed in Section 3. However, suppose we redefine K_- as $e^{A(w)}K_-$ and K_+ as $e^{A(w)}K_+$. Then, if we choose

$$A(w) = -\frac{iaw}{\pi}\left(\gamma - 1 - \log\frac{2\pi}{ak}\right) - \frac{aw}{2},$$

the factors $e^A K_-$ and $e^A K_+$ will both be of algebraic order for $|w| \to \infty$, $\mathscr{I}m\, w$ suitably limited. Upon using these modified forms of K_- and K_+ in equation (5.10), we find that the entire function of separation is identically zero. This may be seen since Φ_1 and Φ_2 go to zero with $|w| \to \infty$, $\mathscr{I}m\, w$ properly limited, as a result of the Riemann-Lebesgue lemma. This then produces Φ_1 and Φ_2, and we can show directly from these functions that $\phi_y(x, a)$ and $J(x)$ possess properties which we require in order to use them in the integral equation and the Fourier transform theorem. Having $\Phi_2(w)$, we can determine the amplitudes of the reflected waves and the radiated field. Indeed,

$$B_0 = -\frac{i}{2ak}\Phi_2(-k),$$

$$B_n = \frac{i(-)^n}{a\Gamma_n}\Phi_2(-\Gamma_n),$$

while

$$f(\theta) = \sqrt{\frac{2}{\pi k}}\, e^{-i\pi/4}\Phi_2(k\cos\theta).$$

These expressions can be simplified once we make use of the precise forms of K_- and K_+ but we shall not pursue this matter.

We can determine $\phi(x, y)$ for any point in the plane $-\infty < x < \infty$, $y > 0$ by making use of the representations (5.4) and (5.5). For $0 \leqq y \leqq a$ we have, with the aid of the convolution theorem,

$$\phi(x, y) = \exp\{ikx\} - \frac{1}{2\pi}\int_C \frac{\cos\sigma y}{\sigma \sin\sigma a}\Phi_2(w)\exp\{iwx\}\, dw$$

where C is a path in the strip of analyticity $-\mathscr{I}m\, k < \mathscr{I}m\, w < \mathscr{I}m\, k$. The closing of the path depends on the sign of x. Similarly, for $y \geqq a$ we have

$$\phi(x, y) = \frac{i}{2\pi}\int_C \frac{\Phi_2(w)}{\sigma}\exp\{iwx+i\sigma(y-a)\}\,dw$$

where C is the same path as we have just described. It is possible to show from these representations that ϕ is continuous along the line $x > 0$, $y = a$, that ϕ_y vanishes along the lines $x < 0$, $y = a$ and $-\infty < x < \infty$, $y = 0$, and that the two forms of ϕ satisfy the original differential equation as well as the conditions at infinity. This is accomplished by reducing the integrals over C to real integrals (i.e., integrals over a real variable). In this final form we can also verify that the results are valid for the case k real.

We shall not discuss any problems with more complicated cross-sections; for, this will take us into the realm of the various special functions such as Bessel functions. The method of formulation and solution, however, is still valid, and it is always possible to produce an integral representation for the solution. If we are only interested in the physical parameters such as the reflection coefficient or the amplitude of the radiated wave, the results ultimately depend on geometric parameters, the wave-number and the characteristic numbers of the Sturm-Liouville problem. This, of course, is of great advantage from the computational point of view.

The comments we have made here hold for any number of obstacles extending to infinity in one direction and generated by lines moving parallel to themselves. In general, we shall obtain systems of integral equations of the Wiener-Hopf type for which we do not have a method of solution as we did in the case of a single unknown. While in the case of a single variable we have Cauchy's integral theorem to help us carry out the factoring in the complex domain, there does not appear to be in general any simple parallel to this for the case of systems. This is one of the unsolved problems in this subject about which we would like to have more information.

Bibliography

The following publications represent most of the pertinent information on this subject.

A. Theory of the Wiener-Hopf Integral Equation

[1] Wiener, N., and Hopf, E., *Über eine Klasse singulärer Integralgleichungen*, Sitz. Berlin Akad. Wiss., 1931, pp. 696–706.

[2] Paley, R. E. A. C., and Wiener, N., *The Fourier transform in the complex domain*, Amer. Math. Soc Coll. Publication, New York, Vol. XIX, 1934, Chap. IV.

[3] Titchmarsh, E. C., *Introduction to the Theory of Fourier Integrals*, Oxford University Press, Oxford, 1937.

[4] Smithies, F., *Singular integral equations*, Proc. London Math. Soc., 2nd Series, Vol. 46, 1940, pp. 490–466.

B. Hydrodynamical Applications

[1] Greene, T. R., and Heins, A. E., *Water waves over a channel of infinite depth*, Quart. Appl. Math., Vol. 11, 1953, pp. 201–214.

[2] Heins, A. E., *Water waves over a channel of finite depth with a dock*, Amer. J. Math., Vol. 70, 1948, pp. 730–748.

[3] Heins, A. E., *Water waves over a channel of finite depth with a submerged plane barrier*, Canadian J. Math., Vol 2, 1950, pp. 210–222.

[4] Heins, A. E., *On gravity waves*, Proc. Symposia Appl. Math., Vol. IV, McGraw-Hill, New York, 1954, pp. 75–88.

C. Acoustical and Electromagnetic Applications

[1] Blass, J , *The Extension of the Wiener-Hopf Technique to Radiation Problems Involving Boundaries of Elliptical Cross-Section*, Dissertation, Polytechnic Institute of Brooklyn, 1951.

[2] Carlson, J. F., and Heins, A. E., *The reflection of an electromagnetic plane wave by an infinite set of plates, I*, Quart. Appl. Math., Vol. IV, 1947, pp. 313–329.

[3] Chester, W., *The propagation of sound waves in an open-ended channel*, Philos. Mag. (7), Vol. 41, 1950, pp. 11–33.

[4] Copson, E. T., *On an integral equation arising in the theory of diffraction*, Quart. J. Math., Oxford Ser , Vol. 17, 1946, pp 19–34.

[5] Heins, A. E , *The radiation and transmission properties of a pair of semi-infinite parallel plates*, Quart. Appl Math., Vol. 6, 1948, pp 157–166, pp. 215–220.

[6] Heins, A. E., *Some remarks on the coupling of two ducts*, J. Math. Physics, Vol. 30, 1951, pp. 164–169.

[7] Heins, A. E., and Carlson, J. F., *The reflection of an electromagnetic plane wave by an infinite set of plates, II*, Quart J. Math., Vol 5, 1947, pp. 82–88.

[8] Heins, A. E., and Feshbach, H., *The coupling of two acoustical ducts*, J. Math. Phys. Mass Inst. Tech , Vol. 26, 1947, pp. 143–155.

[9] Heins, A. E., and Feshbach, H., *On the coupling of two half-planes*, Proc. Symposia Appl. Math., Vol V, McGraw-Hill, New York, 1954, pp. 75–88.

[10] Iijima, T , *On the Electromagnetic Fields in Case of Existence of a Semi-infinite Hollow Conductive Circular Cylinder, II*, Electrotechnical Laboratory, Agency of Industrial Science and Technology, Tokyo, Rep. No 531, 1952, 214 pp.

[11] Jones, D. S , *A simplifying technique in the solution of a class of diffraction problems*, Quart. J. Math., Oxford Ser. (2), Vol. 3, 1952, pp. 189–196.

[12] Levine, H., and Schwinger, J. S., *On the radiation of sound from an unflanged circular pipe*, Physical Rev. (2), Vol. 73, 1948, pp. 383–406.

[13] Vajnshsteyn, L. A , *A rigorous solution of the problem of the plane wave-guide with an open end*, Izvestia Akad. Nauk SSSR. Ser. Fiz , Vol. 12, 1948, pp. 144–165.

[14] Vajnshsteyn, L. A , *On the theory of diffraction by two parallel half-planes*, Izvestiya Akad. Nauk SSSR Ser. Fiz , Vol. 12, 1948, pp. 166–180

[15] Vajnshsteyn, L. A , *The theory of sound vibration in open tubes*, Akad. Nauk SSSR. Zurnal Tehn. Fiz , Vol. 19, 1949, pp 911–930.

[16] Vajnshsteyn, L. A , *The radiation of unsymmetrical electromagnetic waves from the open end of a circular wave-guide*, Doklady Akad. Nauk SSSR. (N S.), Vol. 74, 1950, pp. 485–488.

D. Other References

[1] Watson, G. N., *Bessel Functions*, The University Press, Cambridge, 1952.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 467–470 (1956)

On the Existence Problem for Surfaces of Constant Mean Curvature*

ERHARD HEINZ
University of Goettingen

1. Statement of the Problem

We shall be concerned with the extension of Plateau's problem to surfaces of constant mean curvature. This problem can be stated analytically as follows. Given a Jordan curve Γ in three-dimensional space and a real number H, determine a vector $\mathfrak{x}(u, v) = (x(u, v), y(u, v), z(u, v))$ with the following properties:

(A) The functions $x(u, v)$, $y(u, v)$ and $z(u, v)$ belong to the class $C^{(2)}$ and satisfy the differential equations

$$\Delta x = 2H(y_u z_v - y_v z_u), \quad \Delta y = 2H(z_u x_v - z_v x_u), \quad \Delta z = 2H(x_u y_v - x_v y_u),$$
$$x_u^2 + y_u^2 + z_u^2 = x_v^2 + y_v^2 + z_v^2, \quad x_u x_v + y_u y_v + z_u z_v = 0$$

or, in abbreviated form,

$$\Delta \mathfrak{x} = 2H(\mathfrak{x}_u \times \mathfrak{x}_v), \qquad \mathfrak{x}_u^2 = \mathfrak{x}_v^2, \qquad \mathfrak{x}_u \mathfrak{x}_v = 0.$$

(B) The functions $x(u, v)$, $y(u, v)$ and $z(u, v)$ are continuous for $u^2+v^2 \leqq 1$, and the equation $\mathfrak{x} = \mathfrak{x}(u, v)$ maps the unit circle $u^2+v^2 = 1$ in a one-to-one and continuous way onto Γ.

For $H = 0$ this is precisely Plateau's problem which has been investigated in the literature a great number of times and was finally solved by J. Douglas and T. Radó.[1] It is obvious that we can always assume Γ to be contained in the unit sphere. Then we have

THEOREM 1. *Let Γ be rectifiable and contained in the unit sphere $x^2+y^2+z^2 \leqq 1$. Then there exists a vector $\mathfrak{x}(u, v)$ satisfying conditions* (A) *and* (B), *provided* $|H| < \frac{1}{8}(\sqrt{17}-1)$.

2. The Functional $E[\mathfrak{x}]$

In view of the relation between our existence problem and Plateau's problem, we may expect to establish the existence proof with the help of

*This is a summary of the results obtained in [3].

[1] For a simplified presentation see T. Radó [5] and R. Courant [2], Chapter III.

a similar variational principle. Let us denote by $\mathfrak{D}^*$ the set of all vector functions with the following properties:

(1) $\mathfrak{x}(u, v)$ belongs to $C^{(2)}$ for $u^2+v^2 < 1$,

(2) $\mathfrak{x}(u, v)$ is continuous for $u^2+v^2 \leqq 1$ and satisfies the inequality $|\mathfrak{x}(u, v)| \leqq 1$,

(3) The transformation $\mathfrak{x} = \mathfrak{x}(u, v)$ maps the unit circle $u^2+v^2 = 1$ monotonically in the weak sense onto Γ and carries three fixed distinct points A, B, C of the unit circle into three fixed distinct points A^*, B^*, C^* of Γ.

For $\mathfrak{x} \in \mathfrak{D}^*$ we set

$$E[\mathfrak{x}] = \iint\limits_{u^2+v^2<1} [x_u^2+y_u^2+z_u^2+x_v^2+y_v^2+z_v^2 + \frac{4H}{3}\{x(y_u z_v - y_v z_u)+y(z_u x_v - z_v x_u)+z(x_u y_v - x_v y_u)\}]\,du\,dv,$$

where we assume $|H| < \frac{3}{2}$. Then the integrand of the functional $E[\mathfrak{x}]$ is a positive definite quadratic form in the quantities $x_u, \cdots, z_v$, but the functional is not quadratic in its argument $\mathfrak{x}$ if $H \neq 0$. With these definitions we have

THEOREM 2. *For $|H| < \frac{1}{8}(\sqrt{17}-1)$ there exists a function $\mathfrak{x}^*(u, v)$ in $\mathfrak{D}^*$ such that the inequality $E[\mathfrak{x}^*] \leqq E[\mathfrak{x}]$ holds for every $\mathfrak{x} \in \mathfrak{D}^*$. Furthermore, the vector $\mathfrak{x}^*(u, v)$ satisfies conditions* (A) *and* (B) *of our existence problem.*

The solution of the variational problem $E[\mathfrak{x}] = \text{minimum}$ ($\mathfrak{x} \in \mathfrak{D}^*$) is obtained with the aid of certain existence and compactness theorems for the equation $\Delta\mathfrak{x} = 2H(\mathfrak{x}_u \times \mathfrak{x}_v)$ described in the following section.

3. The Boundary Value Problem for the Equation $\Delta\mathfrak{x} = 2H(\mathfrak{x}_u \times \mathfrak{x}_v)$

Let us denote by $\overline{\mathfrak{M}}_H$ the set of vector functions $\mathfrak{x}(u, v)$ with the following properties:

(1) $\mathfrak{x}(u, v)$ belongs to $C^{(2)}$ for $u^2+v^2 < 1$ and satisfies the equation $\Delta\mathfrak{x} = 2H(\mathfrak{x}_u \times \mathfrak{x}_v)$,

(2) $\mathfrak{x}(u, v)$ is continuous for $u^2+v^2 \leqq 1$ and satisfies the inequality $|\mathfrak{x}(u, v)| \leqq 1$.

Then we have the following results:

THEOREM 3. *Let the function $\bar{\mathfrak{x}}(\varphi)$ be continuous for $-\infty < \varphi < \infty$ and $\bar{\mathfrak{x}}(\varphi+2\pi) = \bar{\mathfrak{x}}(\varphi)$. Furthermore, let $|\bar{\mathfrak{x}}(\varphi)| \leqq 1$. Then there exists a function $\mathfrak{x}(u, v) \in \overline{\mathfrak{M}}_H$ such that $\mathfrak{x}(\cos\varphi, \sin\varphi) = \bar{\mathfrak{x}}(\varphi)$, provided that $|H| < \frac{1}{8}(\sqrt{17}-1)$.*

THEOREM 4. *Let* $|H| < \frac{1}{2}$ *and* $\{\mathfrak{x}^{(n)}(u,v)\}$ *be a sequence of functions belonging to* $\mathfrak{M}_H$, *whose boundary values* $\bar{\mathfrak{x}}^{(n)}(\varphi) = \mathfrak{x}^{(n)}(\cos\varphi, \sin\varphi)$ *converge to a continuous function* $\bar{\mathfrak{x}}(\varphi)$ *uniformly in* $0 \leqq \varphi \leqq 2\pi$. *Then there exists a subsequence* $\{\mathfrak{x}^{(n_k)}(u,v)\}$ *of functions which, together with their first derivatives, converge to a function* $\mathfrak{x}(u,v) \in \mathfrak{M}_H$ *uniformly in every disc* $u^2+v^2 \leqq R^2 < 1$. *Furthermore, we have* $\mathfrak{x}(\cos\varphi, \sin\varphi) = \bar{\mathfrak{x}}(\varphi)$.

It should be noted that corresponding results hold for elliptic systems of the form

$$\Delta x_m = Q_m\left(\frac{\partial x_1}{\partial u}, \cdots, \frac{\partial x_n}{\partial u}; \frac{\partial x_1}{\partial v}, \cdots, \frac{\partial x_n}{\partial v}; x_1, \cdots, x_n; u, v\right), \quad m = 1, \cdots, n,$$

where Q_m is a homogeneous quadratic form in the quantities $\partial x_1/\partial u, \cdots, \partial x_n/\partial v$. Equations of the form

$$\Delta u = au_x^2 + 2bu_xu_y + cu_y^2 + 2du_x + 2eu_y + f, \quad a = a(x,y,u), \cdots, f = f(x,y,u),$$

and

$$\sum_{i,j=1}^{n} a_{ij}(x_1, \cdots, x_n) \frac{\partial^2 u}{\partial x_i \partial x_j} = f\left(x_1, \cdots, x_n; u; \frac{\partial u}{\partial x_1}, \cdots, \frac{\partial u}{\partial x_n}\right)$$

(with $|f(x_1, \cdots, x_n; u; p_1, \cdots, p_n)| \leqq B(p_1^2 + \cdots + p_n^2) + C$), involving one single unknown function u, have been investigated by S. Bernstein [1] pp. 84—90, and by M. Nagumo [4], respectively.

4. Some Open Questions

Consider the vector $\mathfrak{x}(u,v) = (x(u,v), y(u,v), z(u,v))$, where

$$x(u,v) = \frac{(1+\lambda^2)u}{1+\lambda^2(u^2+v^2)},$$

$$y(u,v) = \frac{(1+\lambda^2)v}{1+\lambda^2(u^2+v^2)},$$

$$z(u,v) = \frac{\lambda(u^2+v^2-1)}{1+\lambda^2(u^2+v^2)},$$

and λ is a real parameter. For $u^2+v^2 < \infty$, $\mathfrak{x}(u,v)$ satisfies the equations

$$\Delta\mathfrak{x} = 2H(\mathfrak{x}_u \times \mathfrak{x}_v), \qquad \mathfrak{x}_u^2 = \mathfrak{x}_v^2, \qquad \mathfrak{x}_u\mathfrak{x}_v = 0,$$

where $H = \dfrac{2\lambda}{1+\lambda^2}$. Furthermore, we have

$$\mathfrak{x}(\cos\varphi, \sin\varphi) = (\cos\varphi, \sin\varphi, 0) \qquad \text{for } 0 \leqq \varphi \leqq 2\pi.$$

Thus for every real value H in the interval $-1 < H < 1$ there exist two distinct real solutions of the boundary value problem

$$\Delta \mathfrak{x} = 2H(\mathfrak{x}_u \times \mathfrak{x}_v), \qquad \mathfrak{x}(\cos\varphi, \sin\varphi) = (\cos\varphi, \sin\varphi, 0),$$

while for $|H| > 1$ there are no real solutions of the form considered above. This example raises the following questions about the boundary value problem for the equation $\Delta \mathfrak{x} = 2H(\mathfrak{x}_u \times \mathfrak{x}_v)$:

(1) What is the maximum number of distinct solutions?

(2) Is there, in general, a finite number H^*, depending on the boundary values, such that there exists no solution for $|H| > H^*$?

Bibliography

[1] Bernstein, S., *Sur la généralisation du problème de Dirichlet*, Math. Ann., Vol. 69, 1910, pp. 82–136.

[2] Courant, R., *Dirichlet's Principle, Conformal Mapping and Minimal Surfaces*, Interscience, New York, 1950

[3] Heinz, E., *Über die Existenz einer Fläche konstanter mittlerer Krummung bei vorgegebener Berandung*, Math. Ann., Vol. 127, 1954, pp. 258–287.

[4] Nagumo, M., *On principally linear elliptic differential equations of second order*, Osaka Math J., Vol. 6, 1954, pp. 207–229

[5] Radó, T., *On the Problem of Plateau*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 2, Springer, Berlin, 1933.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 471–478 (1956)

On the Reflection Principle for Polyharmonic Functions*

ALFRED HUBER

Institute for Fluid Dynamics and Applied Mathematics
University of Maryland

In the following let G denote a region in the n-dimensional space, the boundary of which contains an open subset S of the hyperplane $x_1 = 0$. The classical reflection principle of H. A. Schwarz states: *If the function $w(x_1, x_2, \cdots, x_n)$ is harmonic in G and assumes the boundary value 0 on S, then it can be continued analytically across S into the reflected domain G^* by putting*

$$w(-x_1, x_2, \cdots, x_n) = -w(x_1, x_2, \cdots, x_n). \tag{1}$$

In the present article we are concerned with an extension of this theorem to functions which are harmonic of higher order. (A function $w(x_1, x_2, \cdots, x_n)$ is said to be harmonic of order p ($p =$ positive integer) if it is of class C^{2p} and satisfies the differential equation $\Delta^p w = 0$).

The case $p = 2$ has been treated by H. Poritsky [5] and R. J. Duffin [1]. Duffin proved the following theorem: *Suppose that w is biharmonic in G and that w/x_1 assumes the boundary value 0 on S. Then w can be continued analytically across S into G^* by the relation*

$$w(-x_1, x_2, \cdots, x_n) = -w + 2x_1 \frac{\partial w}{\partial x_1} - x_1^2 \Delta w. \tag{2}$$

This continuation formula was even previously published by H. Poritsky [5]. Limiting himself to the case $n = 2$ he demonstrated the validity of (2) under the assumption that w is biharmonic on $G \cup S$ and that $w = \partial w/\partial x_1 = 0$ on S.

The reflection principles (1) and (2) are special cases of the following statement:

THEOREM. *Suppose that the function $w(x_1, x_2, \cdots, x_n)$ is harmonic of order p ($p =$ positive integer) in G and that w/x_1^{p-1} assumes the boundary value 0 on S. Then w can be continued analytically across S into the reflected domain G^* by the relation*

*) This research was supported in part by the United States Air Force under Contract No. AF18 (600)-573 monitored by the Office of Scientific Research, Air Research and Development Command.

$$(3) \qquad w(-x_1, x_2, \cdots, x_n) = \sum_{k=0}^{p-1} \frac{(-x_1)^{p+k}}{(k!)^2} \Delta^k \left(\frac{w(x_1, x_2, \cdots, x_n)}{x_1^{p-k}} \right).$$

Remarks. In the boundary condition for w, x_1^{p-1} cannot be replaced by any smaller power of x_1. This is clearly indicated by the example $x_1^{p-1} \log (x_1^2+x_2^2)$, a function which is harmonic of order p in $x_1 > 0$, but cannot be continued analytically into the origin.

By a generalized Kelvin transformation (see e.g. M. Nicolesco [4] p. 15) with respect to a hypersphere which is tangential to $x_1 = 0$, one arrives at an analogous reflection principle for spherical boundaries. (For the case $p = 2$ see R. J. Duffin [1]).

The reflection principle (3) is contained in a forthcoming paper [2].

Proof: We proceed in two steps,

(**A**) *the function w can be continued analytically beyond S,*

(**B**) *the continuation can be extended to the whole of G^* and is given by* (3).

In [2] we pointed out that Duffin's method [1, report No. 6] of proving statement (**A**) in the special case $p = 2$ can be generalized. One proceeds by complete induction with respect to p. However, we gave no details of this extension, because it was rather involved. Meanwhile, Duffin has considerably simplified his reasoning [1, report No. 13]. By making use of the ideas underlying this new version and applying some relations due to M. Nicolesco [3], we recently found a relatively simple proof of (**A**) for general p, which we present in this article.

On the other hand, [2] contains a detailed proof of statement (**B**). Therefore, we shall now limit ourselves to an abbreviated account of this part of the demonstration. Here our method is different from those used by the above mentioned authors in the case $p = 2$.

Proof of (**A**): The statement is classical for harmonic functions ($p = 1$). Let now p be an arbitrary positive integer $\geqq 2$. We shall demonstrate (**A**) for the order p under the hypothesis that it is valid for the order $p-1$. This will complete the proof.

It is sufficient to show that w can be continued across the interior of an arbitrarily prescribed subcompact K of S. We consider a closed δ-neighborhood K_δ of K, where $\delta > 0$ is small enough so that K_δ is contained in S. Furthermore, we introduce the cylinders $C(\eta) = K \times [0 < x_1 < \eta]$ and $C_\delta(\eta) = K_\delta \times [0 < x_1 < \eta]$, choosing $\eta > 0$ so small that the conditions $\eta < \delta$ and $C_\delta(2\eta) \subset G$ are fulfilled. (We make here the irrelevant assumption that G is a subregion of $x_1 > 0$.)

Writing x for $(x_1, x_2, \cdots, x_n)$ we have, by hypothesis,

$$(4) \qquad w(x) = o(x_1^{p-1}) \qquad \text{for } x_1 \to 0.$$

This holds uniformly in $(x_2, x_3, \cdots, x_n)$ if we limit ourselves to points x lying in $C_\delta(\eta)$.

It is well known that w admits the decomposition

$$w(x) = \sum_{\nu=0}^{p-1} x_1^\nu h_\nu(x), \tag{5}$$

where the functions $h_1, h_2, \cdots, h_{p-1}$ are harmonic in $C_\delta(\eta)$. This yields the representation

$$w(x) = u(x) + x_1^{p-1} h(x), \tag{6}$$

where h is harmonic and u is harmonic of order $p-1$. We state that

$$u(x) = o(x_1^{p-2}) \qquad \text{for } x_1 \to 0, \tag{7}$$

uniformly in $x_2, x_3, \cdots, x_n$ for $x \in C(\eta)$. To prove (7) we first discuss the behavior of h in the neighborhood of K. Since, by differentiation of (6),

$$\Delta^{p-1} w = 2^{p-1}(p-1)! \frac{\partial^{p-1} h}{\partial x_1^{p-1}}, \tag{8}$$

this reduces essentially to an investigation of the growth of $\Delta^{p-1} w$. At this point it is convenient to introduce the iterated means of M. Nicolesco [3]. They are defined as follows:

$$(x_1 - x_{10})^2 + \cdots + (x_n - x_{n0})^2 = R^2,$$

$$\mu_0(w; x_0; R) = \text{arithmetical mean of } w \text{ over the sphere}$$

$$\mu_s(w; x_0; R) = \frac{n}{R^n} \int_0^R r^{n-1} \mu_{s-1}(w; x_0; r)\, dr, \qquad s = 1, 2, 3, \cdots.$$

M. Nicolesco proved that the relations

$$\mu_s(w; x_0; R) = w(x_0) + \sum_{\nu=1}^{p-1} A_{s\nu} R^{2\nu} \Delta^\nu w(x_0), \qquad s = 0, 1, 2, \cdots, \tag{9}$$

where

$$A_{s\nu} = \frac{1}{2^\nu \nu!\, n(n+2)\cdots(n+2\nu-2)} \cdot \left(\frac{n}{n+2\nu}\right)^s,$$

are satisfied by any function $w(x)$ which is harmonic of order p in a region containing $(x_1 - x_{10})^2 + \cdots + (x_n - x_{n0})^2 \leqq R^2$ in its interior.

With every point $x \in C(\eta)$ we now associate the radius $R(x) = \frac{1}{2} x_1$. It is easy to verify that, by (4),

$$\mu_s(w; x; \tfrac{1}{2} x_1) = o(x_1^{p-1}), \qquad s = 0, 1, 2, \cdots, \tag{10}$$

for $x_1 \to 0$, uniformly in $x_2, x_3, \cdots, x_n$ if $x \in C(\eta)$. We infer from (9) that

$$\mu_s(w; x; \tfrac{1}{2} x_1) = w(x) + \sum_{\nu=0}^{p-1} \frac{A_{s\nu}}{2^{2\nu}} x_1^{2\nu} \Delta^\nu(x), \quad s = 0, 1, 2, \cdots, p-1. \tag{11}$$

This may be considered as a system of linear equations for $w, x_1^2 \Delta w, \cdots, x_1^{2p-2} \Delta^{p-1} w$. Its determinant is easily seen to be different from zero. Hence, we have the representation

$$x_1^{2p-2}\Delta^{p-1}w = \sum_{s=0}^{p-1} B_s \mu_s(w; x; \tfrac{1}{2}x_1). \tag{12}$$

The coefficients $B_0, B_1, \cdots, B_{p-1}$ are independent of x. From (8), (10) and (12) we conclude that

$$\frac{\partial^{p-1}h}{\partial x_1^{p-1}} = \frac{1}{2^{p-1}(p-1)!}\Delta^{p-1}w = o(x_1^{1-p}) \qquad \text{for } x_1 \to 0.$$[1]

An easy argument yields

$$h(x) = o(\log x_1) \qquad \text{for } x_1 \to 0. \tag{13}$$

Estimate (7) is now an immediate consequence of (4), (6) and (13).

By applying the induction hypothesis we conclude from (7) that u can be continued analytically across the interior of K. We define

$$v(x) = x_1^{p-2}h(x) + \frac{1}{p-1}\frac{\partial u}{\partial x_1}.$$

It follows from (9) that

$$w = u + x_1\left(v - \frac{1}{p-1}\frac{\partial u}{\partial x_1}\right). \tag{14}$$

Hence

$$v = \frac{w}{x_1} + \frac{1}{p-1}\frac{\partial u}{\partial x_1} - \frac{u}{x_1}. \tag{15}$$

We are now going to prove that

$$v(x) = o(x_1^{p-2}) \qquad \text{for } x_1 \to 0, \tag{16}$$

when x varies on $C(\eta)$. For this purpose we consider the Taylor development of u around an arbitrary point x_0 in S. We conclude from (7) that the factor x_1^{p-1} must be contained in each term of this series. This implies

$$\frac{1}{p-1}\frac{\partial u}{\partial x_1} - \frac{u}{x_1} = O(x_1^{p-2}) \qquad \text{for } x \to x_0. \tag{17}$$

Obviously, (17) also holds for $x_1 \to 0$, uniformly in $x_2, x_3, \cdots, x_n$, if x is allowed to vary on $C(\eta)$. (16) is now a consequence of (15), (4) and (17).

Making again use of the induction hypothesis we infer from (16) that v can be continued analytically across the interior of K. It follows immediately from (14) that the same is true for w.

Proof of (**B**). We introduce the notation $L[w]$ for the right-hand side of (3).

---

[1] This estimate can also easily be obtained as a consequence of an inequality of M. Nicolesco (relation (26) in this paper). Professor L. Nirenberg kindly drew our attention to this result of M. Nicolesco which later led us to the application reproduced at the end of this paper.

If the differentiations and multiplications are carried out, then $L[w]$ is seen to be a polynomial in x_1, w and some derivatives of w. Therefore, $L[w]$ is analytic on $G \cup S$.

Moreover, each term $x_1^{p+k}\Delta^k(w/x_1^{p-k})$ is harmonic of order p. For, we have, applying the decomposition (5),

$$\Delta^p\{x_1^{p+k}\Delta^k[w^{(p)}/x_1^{p-k}]\} = \Delta^p\{x_1^{p+k}\Delta^k[(f^{(p-k)}+x_1^{p-k}g^{(k)})/x_1^{p-k}]\}$$
$$= \Delta^p\{x_1^{p+k}\Delta^k[f^{(p-k)}/x_1^{p-k}]\} = \Delta^p\{x_1^{p+k}(h^{(p)}/x_1^{p+k})\} = 0.$$

Here the superscripts indicate the order of harmonicity.

Now we shall prove that the Taylor developments of the functions $w(-x_1, x_2, \cdots, x_n)$ and $L[w(x_1, x_2, \cdots, x_n)]$ are identical at an arbitrary point of S. This will complete the proof of (**B**). For, then we can immediately draw the conclusion that $w(-x_1, x_2, \cdots, x_n)$, being identical with $L[w(x_1, x_2, \cdots, x_n)]$, can be continued into the whole of G. Therefore, (3) yields in fact a continuation of $w(x_1, x_2, \cdots, x_n)$ into the reflected domain G^*.

Let

$$w = \sum a_{\nu_1\nu_2\cdots\nu_n} x_1^{\nu_1} x_2^{\nu_2} \cdots x_n^{\nu_n} \tag{18}$$

be the Taylor development of w at an arbitrary point of S, which we may assume to be the origin of the coordinate system without losing generality. It remains to prove that the development of $L[w]$ at the same point is given by

$$L[w] = \sum b_{\nu_1\nu_2\cdots\nu_n} x_1^{\nu_1} x_2^{\nu_2} \cdots x_n^{\nu_n}, \tag{19}$$

where

$$b_{\nu_1\nu_2\cdots\nu_n} = (-1)^{\nu_1} a_{\nu_1\nu_2\cdots\nu_n}. \tag{20}$$

The application of the operator L to the power series (18) may be carried out term by term.

From the boundary condition we infer $\partial^k w/\partial x_1^k = 0$ $(k = 0, 1, \cdots, p-1)$ on S. Since this holds at all points of S it follows that $a_{\nu_1\nu_2\cdots\nu_n} = 0$ for $\nu_1 \leqq p-1$ and arbitrary $\nu_2, \nu_3, \cdots, \nu_n$. On the other hand, it is easy to see that the operator L never decreases the number of factors x_1 in a term. Therefore, we have also $b_{\nu_1\nu_2\cdots\nu_n} = 0$ for $\nu_1 \leqq p-1$ and arbitrary $\nu_2, \nu_3, \cdots, \nu_n$. This proves (20) for $\nu_1 \leqq p-1$.

It can be verified (see [2] for details) that

$$L[a_{\nu_1\nu_2\cdots\nu_n} x_1^{\nu_1} x_2^{\nu_2} \cdots x_n^{\nu_n}] = (-1)^{\nu_1} a_{\nu_1\nu_2\cdots\nu_n} x_1^{\nu_1} x_2^{\nu_2} \cdots x_n^{\nu_n} \tag{21}$$

for $p \leqq \nu_1 \leqq 2p-1$ and arbitrary $\nu_2, \nu_3, \cdots, \nu_n$. Since L never decreases the number of factors x_1, terms of order $\geqq 2p$ in x_1 cannot contribute to the

coefficients $b_{\nu_1 \nu_2 \cdots \nu_n}$, where $p \leqq \nu_1 \leqq 2p-1$, and (20) is thus demonstrated for $p \leqq \nu_1 \leqq 2p-1$.

Finally, we observe that both functions $w(-x_1, x_2, \cdots, x_n)$ and $L[w(x_1, x_2, \cdots, x_n)]$ are harmonic of order p. Using the differential equation (and possibly differentiating it) any derivative of order $\nu_1 \geqq 2p$ in x_1 can be expressed as a linear combination of derivatives of the orders ν_1-2, $\nu_1-4, \cdots, \nu_1-2p$ in x_1. If the equality of these last has already been shown, it follows that the first ones have also to be equal. By induction we thus are able to assert the validity of (20) in the remaining range $\nu_1 \geqq 2p$. This completes the proof of (**B**).

The analytic continuation of w is, of course, unique. If the extended function, defined in G^*, is reflected back across S into G, then the original function w is obtained. Therefore, we have the following identity:

$$\sum_{k=0}^{p-1} \frac{x_1^{p+k}}{(k!)^2} \Delta^k \left(\frac{L[w(x_1, x_2, \cdots, x_n)]}{(-x_1)^{p-k}} \right) \equiv w(x_1, x_2, \cdots, x_n).$$

This is equivalent to

$$L^2[w] \equiv w. \tag{22}$$

We have proved this identity only for those functions w which satisfy the hypotheses of the above theorem with a non-empty set S. However, the following more general statement is true:[2]

Suppose that the function $w(x)$ is defined and harmonic of order p in an arbitrary region G. Then (22) *holds throughout G.*

We briefly indicate the demonstration. Let us first make the additional assumption that G contains points lying on the hyperplane $x_1 = 0$. The functions w and $L^2[w]$ are both analytic in G. We consider the Taylor development of w around an arbitrary point on $x_1 = 0$. By applying the operator L^2 term by term we obtain from it the Taylor development of $L^2[w]$. By arguments which are quite similar to those used in the proof of statement (**B**) one verifies the equality of all corresponding coefficients in these two series. This proves (22) under the special hypothesis introduced above.

In order to treat the general case we consider the operator

$$L_t[w] \equiv \sum_{k=0}^{p-1} \frac{(-x_1+t)^{p+k}}{(k!)^2} \Delta^k \left(\frac{w(x_1, x_2, \cdots, x_n)}{(x_1-t)^{p-k}} \right).$$

We have

$$L_t^2[w(x)]-w(x) \equiv L_*^2[v(x^*)]-v(x^*),$$

---

[2] See R. J. Duffin [1] and H. Poritsky [5] for the special case $p = 2$.

where $x^* = (x_1-t, x_2, \cdots, x_n)$, if we define $v(x_1^*, x_2^*, \cdots, x_n^*) \equiv w(x_1^*+t, x_2^*, \cdots, x_n^*)$. (By L_* we understand the operator L, but with $x_1^*, x_2^*, \cdots, x_n^*$ taking the place of $x_1, x_2, \cdots, x_n$.)

Consider a fixed value of t which has been chosen such that the hyperplane $x_1 = t$ intersects with G. The function $v(x^*)$ is harmonic of order p in a region intersecting with $x_1^* = 0$. By the previous result we have $L_*^2[v(x^*)]-v(x^*) \equiv 0$.

Let now $x \in G$ be fixed. Then $L_t^2[w(x)]-w(x)$ is a polynomial in t. It follows from the above remarks that

$$L_t^2[w(x)]-w(x) \equiv 0 \tag{23}$$

for a non-empty open set of values t. We infer that (23) holds for all values of t. Setting $t = 0$ we obtain (22).

We conclude with the following application:

THEOREM. *Let $w(x)$ be harmonic of order p throughout the half-space $x_1 > 0$ and suppose that w/x_1^{p-1} assumes the boundary value 0 on $x_1 = 0$. Then, if for some positive integer q*

$$w(x) = o(|x|^q), \qquad |x| = \sqrt{x_1^2+x_2^2+\cdots+x_n^2}, \tag{24}$$

at infinity, w is a polynomial of the form

$$w(x) = \sum_{\substack{\nu_1 \geqq p \\ \nu_1+\nu_2+\cdots+\nu_n \leqq q-1}} a_{\nu_1\nu_2\cdots\nu_n} x_1^{\nu_1} x_2^{\nu_2} \cdots x_n^{\nu_n}. \tag{25}$$

The proof is based on the following estimate of M. Nicolesco ([4] p. 19): Let M denote the least upper bound of $w(x)$ on $(x_1-x_{10})^2+(x_2-x_{20})^2+\cdots+(x_n-x_{n0})^2 \leqq R^2$, $w(x)$ being harmonic of order p in a region which contains this sphere in its interior. Then there exists a constant $C(n, p; \nu_1, \nu_2, \cdots, \nu_n)$ such that

$$R^{\nu_1+\nu_2+\cdots+\nu_n} \frac{\partial^{\nu_1+\nu_2+\cdots+\nu_n} w(x_0)}{\partial x_1^{\nu_1} \partial x_2^{\nu_2} \cdots \partial x_n^{\nu_n}} \leqq CM. \tag{26}$$

For any given $\varepsilon > 0$ there exists a number $\varrho(\varepsilon)$ such that $|w(x)| < \varepsilon|x|^q$ in $|x| > \varrho$. The least upper bound of $|w|$ within a sphere of radius $x_1/2$ around the point x is therefore not greater than $\varepsilon 2^q|x|^q$ for sufficiently large $|x|$. Making use of (26) we conclude that

$$x_1^{\nu_1+\nu_2+\cdots+\nu_n} \frac{\partial^{\nu_1+\nu_2+\cdots+\nu_n} w(x)}{\partial x_1^{\nu_1} \partial x_2^{\nu_2} \cdots \partial x_n^{\nu_n}} \leqq \varepsilon\, 2^{q+\nu_1+\nu_2+\cdots+\nu_n} C|x|^q. \tag{27}$$

The function $w(x)$ can be reflected with respect to the hyperplane $x_1 = 0$. It is easy to verify that the reflection formula is a finite linear combination of terms of the form (27). Therefore, the growth condition (24) is also fulfilled by the extended function.

Hence $w(x)$ is now defined and harmonic of order p in the whole space, satisfying (24) at infinity. It follows from a theorem of M. Nicolesco [3] that w is a polynomial of order $\leqq q-1$. From the condition that $w/x_1^{p-1} \to 0$ at $x_1 = 0$ we infer that only terms of order $\geqq p$ in x_1 can occur.

If we choose $q = p$ we arrive at the following

COROLLARY. *Suppose that $w(x)$ is harmonic of order p throughout the half-space $x_1 > 0$ and that w/x_1^{p-1} assumes the boundary value 0 on $x_1 = 0$. Then $w = o(|x|^p)$ implies $w \equiv 0$.*

This is a result of the Phragmén-Lindelöf type which answers a question raised by Professor A. Weinstein.

Bibliography

[1] Duffin, R. J., *Continuation of a biharmonic function by reflection*, Bull. Amer. Math. Soc., Vol. 59, 1953, p. 347; Technical Reports No. 6, March 1953, and No. 13, February 1954, Carnegie Institute of Technology.

[2] Huber, A., *The reflection principle for polyharmonic functions*, Pacific J. Math, Vol 5, 1955, pp. 433–439

[3] Nicolesco, M., *Sur les fonctions de n variables, harmoniques d'ordre p*, Bull Soc. Math. France, Vol. 60, 1932, pp. 129–151.

[4] Nicolesco, M., *Les fonctions polyharmoniques*, Actualités scientifiques et industrielles, 331, Hermann, Paris 1936.

[5] Poritsky, H., *Application of analytic functions to twodimensional biharmonic analysis*, Trans. Amer. Math. Soc., Vol. 59, 1946, pp. 248–279.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 479–486 (1956)

On Linear Differential Equations in Banach Spaces*

TOSIO KATO

New York University and University of Tokyo, Japan

1. Introduction

This paper is concerned with the solution of a Cauchy problem in an abstract linear space. Consider a linear differential equation

$$\frac{du(t)}{dt} = A(t)u(t)+f(t), \qquad a \leqq t \leqq b, \tag{1.1}$$

and the associated homogeneous equation

$$\frac{du(t)}{dt} = A(t)u(t). \tag{1.1a}$$

Here t is a real variable; the unknown function $u(t)$ and the given function $f(t)$ take values in a Banach space X; and $A(t)$ is a given function whose values are linear operators in X. We wish to find the solution of (1.1) or (1.1a) when the initial value $u(a)$ is prescribed. The problem is rather trivial if the operators $A(t)$ are bounded[1] and, say, strongly continuous in t; for, then the solution can be obtained by a standard iteration procedure. But we are mainly interested in the case where the $A(t)$ are unbounded, so that their domains are proper subsets of X. In such a case the existence of the solution is by no means obvious.

The Schroedinger equation in quantum mechanics is expressed exactly in the form (1.1a) with X a Hilbert space and with the operators $iA(t)$ self-adjoint. It is also obvious that linear partial differential equations of parabolic type which are of the first order in the time variable t can be written in the form (1.1). Our study is primarily intended for application to such problems. It is to be noted, however, that a large number of hyperbolic differential equations can also be reduced to the form (1.1) (see e.g., [1]), at least formally, by introducing a suitable space X as the set of "column-vector functions", in which the corresponding "matrix-functions" $A(t)$

*The research reported in this article was supported by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

[1]Bounded linear operators are understood to be defined everywhere in X, unless otherwise stated.

operate. We hope that our results may be applicable at least to some of these problems.

In this paper we shall present some uniqueness and existence theorems for the solution of (1.1). Only a sketch of the proof will be given, however; a complete proof will be published elsewhere.[2]

The formal differential equation (1.1) admits various interpretations. The derivative $du(t)/dt$ may be taken either in the strong or weak sense, and the equality may be required for all t or, e.g., for almost all t (cf. Lions [4]). Again, it may be regarded as an operational equation $[d/dt-A(t)]u(t) = f(t)$ in some linear space consisting of vector-valued functions $u(t)$, $a \leqq t \leqq b$, and in this interpretation the operator $d/dt-A(t)$ may be taken in some extended sense, either strong or weak.[3]

However, in the present paper we take (1.1) in the simplest sense. We require that $du(t)/dt$ exist as a strong derivative in X, that $A(t)u(t)$ make sense, that (1.1) hold for all t in the open interval $a < t < b$, and that $u(t)$ be strongly continuous in the closed interval. In particular, $u(t)$ must remain in the instantaneous domain $\mathfrak{D}[A(t)]$ of $A(t)$—a rather severe requirement in view of the fact that $\mathfrak{D}[A(t)]$ is in general of first category in X. However, this is a customary requirement in mixed initial- and boundary-value problems for differential equations.

2. Semi-Group Theory

It is instructive first to consider the case in which $A(t) = A$ is independent of t. Then the solution of (1.1) is formally given by

$$u(t) = \exp\{(t-a)A\}u(a) + \int_a^t \exp\{(t-s)A\}f(s)ds. \tag{2.1}$$

The question of defining the exponential function $\exp\{tA\}$ is exactly the problem of generating a one-parameter semi-group; this has been solved by Hille [2], Yosida [6], Phillips [5] and others.

In the Hille-Yosida theorem the "infinitesimal generator" A is assumed to satisfy the following conditions: 1) that A be a closed linear operator in X with a dense domain, and 2) that all positive real numbers belong to the resolvent set of A, and that

$$||(\lambda I-A)^{-1}|| \leqq \lambda^{-1}, \qquad \lambda > 0. \tag{2.2}$$

In what follows, the class of operators A satisfying these conditions will be denoted by (S). The Hille-Yosida theorem states that for each $A \in (S)$ there

[2] The reader is also referred to an earlier paper [3] by the author, where the same problem is treated under more restrictive assumptions than those used in the present paper.

[3] Cf. Friedrichs [1], where reference to other related papers may be found.

exists a unique operator-valued function $\exp\{tA\}$ defined for $t \geqq 0$ with these properties:

(2.3) $\exp\{tA\}$ is bounded and strongly continuous in t with $\exp\{0A\} = I$;

(2.4) $\exp\{(t+s)A\} = \exp\{tA\}\exp\{sA\}$;

(2.5) $\begin{cases} \exp\{tA\} \text{ maps the domain } \mathfrak{D}(A) \text{ into itself and for each } u \in \mathfrak{D}(A), \\ \exp\{tA\}u \text{ is strongly differentiable, with } (d/dt)\exp\{tA\}u = \\ A\exp\{tA\}u = \exp\{tA\}Au; \end{cases}$

(2.6) $\exp\{tA\}$ is permutable with the resolvent $(\lambda I - A)^{-1}$.

According to Phillips [5], one may replace the inequality (2.2) by the weaker condition[4]

$$\|(\lambda I - A)^{-n}\| \leqq k\lambda^{-n}, \qquad n = 1, 2, \cdots, \qquad \lambda > 0, \tag{2.7}$$

where k is independent of n and λ. The class of operators A satisfying these conditions will be denoted by (S''). Of course $(S'') \supset (S)$. The function $\exp\{tA\}$ can still be defined for each $A \in (S'')$ so as to retain the properties (2.3)—(2.6). Furthermore, Phillips [5] has proved that (2.1) is a solution of (1.1) if $f(t)$ is strongly continuously differentiable. It may be remarked that (2.1) is also a solution of (1.1) if $f(t) \in \mathfrak{D}(A)$ for all t and if $Af(t)$ and $f(t)$ are strongly continuous in t.

It is natural to attempt to use the condition $A \in (S'')$ in the general case where $A(t)$ actually depends on t. But it appears that the class (S'') is too wide to work with, therefore we introduce a class (S') of infinitesimal generators lying between (S) and (S''). A linear operator A belongs by definition to (S') if there exists a bounded linear operator Q with a bounded inverse such that

$$\bar{A} = QAQ^{-1}$$

belongs to (S). This implies that Q maps $\mathfrak{D}(A)$ onto $\mathfrak{D}(\bar{A})$ in a one-to-one fashion. It is easily seen that $(S) \subset (S') \subset (S'')$. The fact that one can define the function $\exp\{tA\}$ for $A \in (S')$ with the properties (2.3)—(2.6) can be established without relying on Phillips' results. We need only set

$$\exp\{tA\} = Q^{-1}\exp\{t\bar{A}\}Q, \tag{2.8}$$

where $\exp\{t\bar{A}\}$ is defined by the Hille-Yosida theorem. Note also that since $(\lambda I - A)^{-1} = Q^{-1}(\lambda I - \bar{A})^{-1}Q$, all positive real numbers belong to the resolvent set of A.

---

[4] Actually the condition given by Phillips is more general than (2.7) in that the right-hand side of his equation contains $(\lambda-\omega)^{-n}$ instead of λ^{-n}. The difference is not essential, however, since either form can be reduced to the other by using $A+\omega I$ instead of A. We have chosen the form (2.7) for convenience of comparison with the case (2.2).

3. Uniqueness Theorem

After these preliminaries we return to the general case with time-dependent $A(t)$ and introduce the following assumption.

ASSUMPTION 1. For each t, $a \leqq t \leqq b$, the operator $A(t)$ belongs to the class (S'), so that there is a bounded linear operator $Q(t)$ with a bounded inverse $Q(t)^{-1}$ such that $\tilde{A}(t) = Q(t)A(t)Q(t)^{-1} \in (S)$. Moreover $Q(t)$ is strongly continuously differentiable, that is, the strong derivative[5] $\dot{Q}(t) = dQ(t)/dt$ exists and is strongly continuous.

THEOREM 1. *Under Assumption* 1, *the solution of* (1.1) *is uniquely determined by the initial value* $u(a)$.

This is an immediate consequence of

LEMMA 1. *Let* $u(t)$ *be strongly continuous for* $a \leqq t \leqq b$. *Furthermore, let the strong right-hand derivative* $D^+u(t)$ *exist, and let* $u(t) \in \mathfrak{D}[A(t)]$ *and*

$$D^+u(t) = A(t)u(t)$$

for $a < t < b$. *Then there exists a constant* c *depending only on* $A(t)$ *such that* $||u(t)|| \leqq c||u(a)||$ *for* $a \leqq t \leqq b$.

This lemma can be reduced to a corresponding theorem, proved in the earlier paper [3], for the case where $A(t) \in (S)$, if we introduce a new unknown

$$y(t) = \exp\{-ht\}Q(t)u(t)$$

with a suitable constant h.

4. Existence Theorem

In order to prove the existence of the solution of (1.1) for a given initial value $u(a)$, we need some additional assumptions.[6] In general, it cannot be expected that a solution exists for all $u(a) \in X$, since this is not the case even if $A(t) = A$ is independent of t (see (2.5)). Thus we seek a theorem which will ensure the existence of a solution for each $u(a)$ belonging to $\mathfrak{D}[A(a)]$. Essentially this implies that we require the validity of the differential equation (1.1) not only in the open interval $a < t < b$ but also at the initial time $t = a$.[7]

In addition to Assumption 1, we make the following assumption:

---

[5] The results would be the same if we assume that $\dot{Q}(t)$ is a weak derivative.

[6] The second example of Section 5 shows that the change of $A(t)$ must be sufficiently smooth in some sense.

[7] This is an essential restriction of our theory. There are many problems in which the initial value may be arbitrary and yet the differential equation is satisfied in the open interval. This is usually the case for classical parabolic equations However, the condition $u(a) \in \mathfrak{D}[A(a)]$ is indispensable in the case of the Schroedinger equation

ASSUMPTION 2. The operator $A(t)$ changes smoothly with t in the following sense. There exists a bounded operator-valued function $R(t)$ with a bounded inverse $R(t)^{-1}$ and with a strongly continuous second derivative[8] $\ddot{R}(t)$ such that the operator $\bar{A}(t) = R(t)A(t)R(t)^{-1}$ has a domain $\mathfrak{D}$ independent of t. It then follows that the operator $B(t) = [I-\bar{A}(t)][I-\bar{A}(a)]^{-1}$ is bounded for each t. It is further assumed that $B(t)$ is strongly continuously differentiable.

Let us examine this assumption and its immediate consequences. The operator $R(t)$ maps $\mathfrak{D}[A(t)]$ onto $\mathfrak{D}$ in a one-to-one fashion. Since by Assumption 1 all positive real numbers λ belong to the resolvent set of $A(t)$, the same is true for $\bar{A}(t)$, and, in particular, $[I-\bar{A}(t)]^{-1}$ exists. $B(t)$ and $B(t)^{-1}$ are continuous in t in the uniform topology, so that there exists a constant M such that $||B(t)|| \leqq M$, $||B(t)^{-1}|| \leqq M$. Moreover, $B(t)$ is of bounded variation in the sense that

$$\sum_{j=1}^{n} ||B(t_j)-B(t_{j-1})|| \leqq \int_a^b ||\dot{B}(t)||dt \equiv N \tag{4.1}$$

for any partition $a = t_0 < t_1 < \cdots < t_n = b$ of the interval $a \leqq t \leqq b$. Again, it follows that

$$[I-\bar{\bar{A}}(t)]^{-1} = [I-\bar{\bar{A}}(a)]^{-1}B(t)^{-1}$$

and

$$[I-A(t)]^{-1} = R(t)^{-1}[I-\bar{A}(t)]^{-1}R(t)$$

are continuous in t in the uniform topology.

We can now state our main theorem.

THEOREM 2. *Under Assumptions* 1 *and* 2, *there exists a unique operator-valued function* $U(t, s)$, *defined in the triangle* $a \leqq s \leqq t \leqq b$, *with the following properties*:

$$\begin{cases} U(t, s) \text{ is a bounded linear operator and is strongly continuous in } (s,t), \\ \text{with } U(t, t) = I; \end{cases} \tag{4.2}$$

$$U(t, r) = U(t, s)U(s, r), \qquad r \leqq s \leqq t; \tag{4.3}$$

$$\begin{cases} U(t, s) \text{ maps } \mathfrak{D}[A(s)] \text{ into } \mathfrak{D}[A(t)], \text{ and for each } u_0 \in \mathfrak{D}[A(s)], \\ u(t) = U(t, s)u_0 \text{ is strongly differentiable for } t \geqq s \text{ and satisfies } (1,1a); \\ \text{moreover } A(t)u(t) \text{ is strongly continuous in } t. \end{cases} \tag{4.4}$$

In particular $u(t) = U(t, a)u(a)$ *satisfies the homogeneous equation* (1.1a) *for each* $u(a) \in \mathfrak{D}[A(a)]$.

The solution of the inhomogeneous equation (1.1) can be expressed in terms of the operator $U(t, s)$ formally:

---

[8] Again, the result would be the same if we assume that $\ddot{R}(t)$ is a weak derivative.

$$u(t) = U(t, a)u(a) + \int_a^t U(t, s)f(s)ds. \tag{4.5}$$

As to the validity of this expression as the solution of (1.1), we have

THEOREM 3. (4.5) *is the solution of* (1.1) *if* $u(a) \in \mathfrak{D}[A(a)]$ *and if the function* $f(s)$ *satisfies one of the following conditions*:

(4.6) $f(t)$ *is strongly continuously differentiable,*

or

(4.7) $f(t) \in \mathfrak{D}[A(t)]$, *and* $A(t)f(t)$ *and* $f(t)$ *are strongly continuous in* t.

The proofs of these theorems, which are rather complicated, will be sketched here; they are like those given in the earlier paper [3]. The problem can, without loss of generality, be reduced to the case $R(t) = I$. It is then similar to the problem treated in the earlier paper [3], except that now we have $A(t) \in (S')$ instead of $A(t) \in (S)$. The operator $U(t, s)$ is constructed by a Riemann product integration method, and the properties (4.2) and (4.3) are proved. The proof of (4.4) is the most complicated step; it is effected by showing that the operator

$$W(t, s) = [I - A(t)]U(t, s)[I - A(s)]^{-1}$$

is bounded. To do this, a limiting procedure based on the approximation of $U(t, s)$ by a finite "Riemann product" must be used. To prove Theorem 3 we must first show that

$$\int_a^t U(t, s)f(s)ds$$

belongs to $\mathfrak{D}[A(t)]$. This requires some calculation in the case where condition (4.6) is assumed to hold.

5. Examples

Rather than discuss the applications of our results in detail, we shall illustrate them by some simple but non-trivial examples.

First, consider the Schroedinger equation

$$-\frac{1}{i}\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\left[p(x,t)\frac{\partial u}{\partial x}\right] + q(x,t)u, \quad 0 \leqq x \leqq 1, \quad a \leqq t \leqq b. \tag{5.1}$$

Here the unknown $u = u(x, t)$ is a function of two variables x and t. We assume that the coefficients p, q are smooth real-valued functions of x, t and that $p > 0$; furthermore, the time-dependent homogeneous boundary conditions

$$u_x(0, t) - \alpha(t)u(0, t) = 0, \qquad u_x(1, t) + \beta(t)u(1, t) = 0 \tag{5.2}$$

are imposed at the endpoints of the interval $[0, 1]$. In (5.2), $\alpha(t)$, $\beta(t)$ are smooth real-valued functions.

Equation (5.1) may be written in the form (1.1a) where $A(t)$ is, for each t, a second-order differential operator with the boundary conditions (5.2). If we take as X the Hilbert space $L^2(0, 1)$, then $A(t)$ can be written in the form $A(t) = -iH(t)$ with the operator $H(t)$ self-adjoint. Thus Assumption 1 is satisfied simply with $Q(t) = I$.

It can be shown that Assumption 2 is also satisfied. The domain of $H(t)$ consists of all functions of $L^2(0, 1)$ whose second derivatives also belong to $L^2(0, 1)$ and which satisfy the boundary conditions (5.2). Thus it is easy to find the operator $R(t)$ required in Assumption 2. We can take as $R(t)$ a multiplicative operator

$$R(t) = \exp\{r(x, t)\}, \tag{5.3}$$

where $r(x, t)$ is any smooth real-valued function such that

$$r_x(0, t)+\alpha(t) = 0, \qquad r_x(1, t)-\beta(t) = 0. \tag{5.4}$$

For instance, we may take

$$r(x, t) = -\alpha(t)x+\tfrac{1}{2}[\alpha(t)+\beta(t)]x^2.$$

Then $R(t)$ maps the domain of $H(t)$ onto the set $\mathfrak{D}$ of all functions $u(x) \in L^2(0, 1)$, with $u_{xx} \in L^2(0, 1)$, which satisfy the boundary conditions $u_x(0) = u_x(1) = 0$. (Thus $\mathfrak{D}$ is independent of t.) It is easy to see that the remaining conditions of Assumption 2 are satisfied.

Thus (5.1), interpreted as a differential equation in $X = L^2(0, 1)$, has one and only one solution with a given initial value $u(a) \in \mathfrak{D}[H(a)]$.

It is true that Assumption 2 is still rather restrictive, and much would be gained if it could be replaced by a weaker assumption. A natural question is whether it is sufficient to require that the operator $[I-A(t)]^{-1}$, which is bounded by Assumption 1, be, say, strongly continuously differentiable. But the following example shows that this is not the case. Consider the Schroedinger equation

$$\frac{du(t)}{dt} = -iH(t)u(t) \tag{5.5}$$

in the Hilbert space $X = L^2(-\infty, \infty)$, where $H(t)$ is a multiplicative operator

$$H(t) = -(x-t)^{-2}. \tag{5.6}$$

Then the solution of (5.5), if it exists, could not be other than

$$u(t) = u(x, t) = \exp\left\{\frac{i}{x-t} - \frac{i}{x-a}\right\} u(x, a).$$

But this does not satisfy (1.1) as an operational equation in our sense, unless $u(x, a)$ were to vanish identically in the interval $a \leqq x \leqq b$, because other-

wise $u(t)$ does not belong to the domain of $H(t)$ for any t, $a \leqq t \leqq b$. Such $u(x, a)$ are very special functions and do not form a dense set in X. Nevertheless, the operator $H(t)$ is self-adjoint and changes rather smoothly in the sense that $[I-iH(t)]^{-1}$ is not only continuously differentiable but also analytic in t.

It is interesting to note that, even in this example, we can find a bounded operator function $R(t)$ with the property that $\tilde{H}(t)=R(t)H(t)R(t)^{-1}$ has a constant domain. We need only take as $R(t)$ the unitary operator which maps $u(x)$ into $u(x+t)$. But then $\dot{R}(t)$ is not a bounded operator, so that the Assumption 2 is not satisfied.

Bibliography

[1] Friedrichs, K. O., *Symmetric hyperbolic linear differential equations*, Comm. Pure Appl Math., Vol. 7, 1954, pp. 345–392.

[2] Hille, E., *Functional analysis and semi-groups*, Amer. Math. Soc. Colloquium Publication, New York, 1948.

[3] Kato, T., *Integration of the equation of evolution in a Banach space*, J. Math. Soc. Japan, Vol. 5, 1953, pp. 208–304.

[4] Lions, J. L., *Sur certains problèmes mixtes*, C. R. Acad. Sci. Paris, Vol. 240, 1955, pp. 390–392.

[5] Phillips, R. S., *Perturbation theory for semi-groups of linear operators*, Trans. Amer. Math. Soc., Vol. 74, 1953 pp. 199–221.

[6] Yosida, K., *On the differentiability and the representation of one-parameter semi-groups of linear operators*, J. Math Soc. Japan, Vol. 1, 1949, pp. 15–21.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 487–497 (1956)

An Unsymmetric Operator Arising in the Theory of Neutron Diffusion

JOSEPH LEHNER

Los Alamos Scientific Laboratory
University of California, Los Alamos, New Mexico

§1. This paper is an account of joint work by Dr. G. M. Wing of the Los Alamos Scientific Laboratory and the author, performed under the auspices of the Atomic Energy Commission. It is concerned with the physical problem of the transport of neutrons in a material body. Details of the proofs may be found in the papers listed at the end of this report.

Suppose we have a material body which in the presence of neutrons is subject to nuclear processes such as fission, absorption, and scattering. Given a certain initial distribution of neutrons, what are the subsequent distributions in space and time?

We have considered two geometries, the infinite slab and the sphere. Our results are complete only for the former. The same method of analysis is used in both problems, but important differences arise. Our principal results are contained in Theorems 1—3 (§3) and Theorem 4 (§4).

When enough simplifying assumptions are made, the problem leads to the linear Boltzmann equation. We assume that scattering is isotropic, that neutron energies are not changed by a nuclear process, that the various cross-sections are independent of space and time, and that neutrons do not interact with each other.

Now let us consider an infinite slab of width $2a$. Since the behavior is assumed to be stratified in planes, we define a position coordinate x extending from $-a$ to a. The neutron density at time t will depend on x and on μ, the cosine of the angle between the direction of the neutron beam and the positive x-axis. If the slab is surrounded by vacuum, we get the following equations:

$$\frac{\partial u}{\partial t} = Au, \quad u = u(x, \mu, t), \qquad |x| \leqq a, \ |\mu| \leqq 1,$$

$$u(\pm a, \mu, t) = 0, \qquad \mu \gtrless 0, \quad t > 0, \tag{1.1}$$

$$A \cdot = -\mu \frac{\partial}{\partial x} + \frac{c}{2} \int_{-1}^{1} \cdot \, d\mu'.$$

In an initial value problem we would have the further condition

$$u(x, \mu, 0) = f(x, \mu). \tag{1.2}$$

Here $c > 0$ is a physical parameter which measures the "multiplication" of neutrons, and u differs by an elementary factor from the neutron density.

The problem, as set in (1.1)—(1.2), is that of solving an operator equation, $u_t = Au$. The boundary condition will be met by a suitable definition of the domain $\mathscr{D}(A)$ of A. Let H be the Hilbert space of functions $f(x, \mu, t)$, $|x| \leqq a$, $|\mu| \leqq 1$, $t > 0$, which, for fixed t, satisfy

$$\int_{-1}^{1}\int_{-a}^{a} |f(x, \mu, t)|^2 \, dx \, d\mu < \infty.$$

We then define the domain of A to be the linear manifold in H of functions g such that

$$\begin{aligned} &g \text{ is absolutely continuous in } x \text{ for each fixed } \mu, \\ &Ag \in H, \qquad g(\pm a, \mu, t) = 0, \qquad \mu \lesseqgtr 0. \end{aligned} \tag{1.3}$$

Thus if u satisfying $u_t = Au$ also lies in $\mathscr{D}(A)$, it will solve (1.1). Incidentally the inner product in H is

$$(f, g) = \int_{-1}^{1}\int_{-a}^{a} f(x, \mu)\,\overline{g(x, \mu)}\, dx \, d\mu.$$

From the theory of semi-groups we take the notion of the exponential function of an operator. If $T(t) = \exp\{tA\}$ exists for $t > 0$, and $T(t)$ is continuous at $t = 0$ (with $T(0) = 1$), then $T(t)f$ solves the above problem for any f in $\mathscr{D}(A)$:

$$u(x, \mu, t) = T(t)[f], \qquad t > 0. \tag{1.4}$$

For, it is known that

$$\frac{dT}{dt} f = ATf = TAf, \qquad f \in \mathscr{D}(A), \tag{1.5}$$

where dT/dt is the strong derivative of T. Hence, when $f \in \mathscr{D}(A)$, $u = Tf$ is in $\mathscr{D}(A)$ and satisfies the boundary condition. It also fulfills (1.2) in the form

$$\|u(x, \mu, t) - f(x, \mu)\| \to 0, \quad t \to 0. \tag{1.6}$$

We must, therefore, establish the existence of the semi-group of operators $T(t)$, which is conveniently done by the Hille-Yosida theorem. This theorem states that an operator A generates a semi-group $T(t) = \exp\{tA\}$ continuous for $t \geqq 0$ provided (i) A is closed, (ii) $\mathscr{D}(A)$ is dense in H, (iii) $\|R(\lambda, A)\| \leqq 1/(\lambda - k)$, $\lambda > k$ for some $k > 0$, where $R(\lambda, A)$ is the resolvent of A. Only the third condition presents any difficulty.

We note that A is a perturbation of a differential operator by a bounded one:

$$A = -D + cJ, \quad D = \mu\frac{\partial}{\partial x}, \quad J = \frac{1}{2}\int_{-1}^{1} \cdot \, d\mu'. \tag{1.7}$$

Clearly $||J|| = 1$. Let $g \in \mathcal{D}(A)$. Then

$$\mathcal{R}e(Dg, g) = \int_{-1}^{1}\int_{-a}^{a} dx\, d\mu \left(\mu \frac{\partial g}{\partial x}\bar{g} + \mu \frac{\partial \bar{g}}{\partial x} g\right) = \int_{-1}^{1} d\mu\, [\mu g \bar{g}]_{-a}^{a}$$
$$= \int_{1}^{0} \mu |g(a, \mu)|^2 d\mu - \int_{-1}^{0} \mu |g(a, \mu|^2 d\mu \geqq 0,$$

where we have used the boundary condition (1.1). Now

$$\mathcal{R}e((\lambda - A)g, g) = \lambda ||g||^2 + \mathcal{R}e(Dg, g) - c\,\mathcal{R}e(Jg, g) \geqq (\lambda - c)||g||^2.$$

Setting $(\lambda - A)g = f$, we have $g = R(\lambda, A)f$, and f is an arbitrary element of H since λ is in the resolvent set of A. Then

$$||f||\, ||g|| \geqq |(f, g)| \geqq \mathcal{R}e(f, g) = \mathcal{R}e((\lambda - A)g, g) \geqq (\lambda - c)||g||^2$$

and this implies (iii) with $k = c$.

This very simple argument establishes the existence of the semi-group $T(t)$ and thus solves our problem in principle. However, we want to have many more details about the solution $u = Tf$.

For this purpose we need a more concrete representation of u. It is known that the resolvent of A is the Laplace transform of $T(t)$, and we can use the inversion formula to express $T(t)$:

$$T(t)f = \lim_{\omega \to \infty} \frac{1}{2\pi i} \int_{\gamma - i\omega}^{\gamma + i\omega} e^{\lambda t} R(\lambda, A) f\, d\lambda, \qquad t > 0, \quad f \in \mathcal{D}(A), \tag{1.8}$$

where $\gamma > c$. The integral is to be considered as the strong limit in H of Riemann sums. However, we now fix the point (x, μ); then $R(\lambda, A)f$ becomes a complex-valued continuous function of λ, and the integral becomes an ordinary Riemann integral.

In order to treat the above integral we must know the singularities of $R(\lambda, A)$, i.e., the spectrum of A.

§2. The operator A is an awkward one to handle. It is not bounded or self-adjoint, it is not compact nor is its resolvent compact. About the only hold we have on A is its connection with a certain integral operator which is more amenable to treatment.

For the time being let us consider only the half-plane $\mathcal{R}e\, \lambda \geqq 0$. Suppose λ is an eigenvalue of A,

$$A\psi = \lambda\psi. \tag{2.1}$$

Integrating this equation with respect to x and using the boundary condition (1.1), we obtain

$$\mu\psi(x, \mu) = \frac{c}{2} \int_{-a}^{x} e^{-\lambda(x - x')/\mu} \varphi(x')\, dx', \qquad \mu > 0, \tag{2.2}$$

and a similar equation when $\mu < 0$, where

$$\varphi(x) = \int_{-1}^{1} \psi(x, \mu')\, d\mu'. \tag{2.3}$$

Solving (2.2) for ψ and integrating over μ, we get an integral equation for φ:

$$\frac{2}{c}\varphi = \mathscr{L}_\lambda \varphi, \tag{2.4}$$

where $\mathscr{L}_\lambda$ is the integral operator

$$\mathscr{L}_\lambda f = \int_{-a}^{a} E(\lambda|x-y|)\, f(y)\, dy, \tag{2.5}$$

and E is the exponential integral

$$E(u) = \int_1^\infty e^{-ut} t^{-1}\, dt, \qquad \mathscr{R}e\, u \geqq 0, \quad u \neq 0. \tag{2.6}$$

I.e., if λ is an eigenvalue of A, then $2/c$ is an eigenvalue of $\mathscr{L}_\lambda$, and the converse is also true. Therefore, we can determine the point spectrum of A in the closed half-plane $\mathscr{R}e\, \lambda \geqq 0$ by discussing the integral equation

$$\varrho\varphi = \mathscr{L}_\lambda \varphi \tag{2.7}$$

and finding those λ for which an eigenvalue ϱ equals $2/c$. Note that the spectral variable λ has left its linear position in (2.1) and is in (2.7) a parameter of the kernel.

Now it is possible to show that there are only a finite number of λ which satisfy the above conditions and these are all real. Suppose λ could be complex, i.e., suppose

$$\frac{2}{c}\varphi = \mathscr{L}_\lambda \varphi \tag{2.8}$$

could be satisfied by a non-real λ. Let φ be a normalized eigensolution. Then $2/c = (\mathscr{L}_\lambda \varphi, \varphi)$, and the right member is real since the left is. The right member is easily expressed by means of Fourier transforms:

$$(\mathscr{L}_\lambda \varphi, \varphi) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Delta(\omega, \lambda)\, |\Phi(\omega)|^2\, d\omega \tag{2.9}$$

where Δ is the transform of the kernel E and Φ is the transform of φ. Since $\mathscr{I}m\, \{\Delta(\omega, \lambda)\}$ is continuous and never zero when λ is not real, it follows that $(\mathscr{L}_\lambda \varphi, \varphi)$ cannot be real. Thus all the solutions λ of (2.8) are real. We write $\lambda = \beta > 0$.

Consider (2.7) with $\lambda = \beta$. The eigenvalues ϱ are continuous, decreasing functions of β. The kernel of (2.7) is not degenerate, hence there are infinitely many "eigenvalue curves" $\varrho_\nu(\beta)$ in the β, ϱ-plane. These are all positive ($\varrho_\nu > 0$), for, $\mathscr{L}_\beta$ is a positive operator. We have to show that the intersections of the curves ϱ_ν with the horizontal line $\varrho = 2/c$ are finite in number. There is always one intersection at least since $\varrho_1 \to \infty$ as $\beta \to 0$. But the

other intercepts $\varrho_\nu(0)$, $\nu = 2, 3, \cdots$, are at finite positions on the ϱ-axis and cluster only at $\varrho = 0$. This is proved by approximating the kernel $E(\beta|x-y|)$ near $\beta = 0$ by $-\log \beta b|x-y|$, where b is a certain constant, and applying the *minimax* principle.

The point spectrum of A in the closed right half-plane is therefore a finite, non-empty set $\beta_1 > \beta_2 > \cdots > \beta_m > 0$, $m \geqq 1$. This is the entire point spectrum. For, suppose λ ($\mathscr{R}e\,\lambda < 0$) were an eigenvalue of A with eigenfunction ψ. Between ψ and $\varphi(x) = \int_{-1}^{1} \psi(x, \mu')d\mu'$ there is the relation (2.2). Let

$$F_a(z) = F(z) = \frac{c}{2}\int_{-a}^{a} e^{-z\lambda(x-x')}\varphi(x')\,dx',$$

i.e., $zF(z) = \psi(x, 1/z)$ for $z > 0$. $F(z)$ is an entire function of z of order one, and so is

$$G(z) = z^{-1}\int_0^z F^2(t)\,dt. \tag{2.10}$$

Since $\int_{-1}^{1} |\psi|^2 d\mu < \infty$ by hypothesis, we get from the preceding that $G(r) \to 0$, $r \to \infty$. Also, by direct examination of the integral, we see that $G(z) = O(1)$ on certain lines (depending on λ) which make with the real axis angles smaller than π. Then the Phragmén-Lindelöf theorem shows that G is bounded in the whole plane and therefore identically zero (since $G(r) \to 0$). Hence, $F(z)$ vanishes identically. The definition of $F(z)$ then yields the vanishing of $\varphi(x')$ almost everywhere in $(-a, x)$, but x was an arbitrary point (almost everywhere) in $(-a, a)$. Hence, $\varphi(x)$ is a null function and the proof is complete.

The discussion of the remaining parts of the spectrum and the resolvent set is much simpler. First of all, there can be no residual spectrum. To see this, introduce the adjoint operator

$$A^* = D + cJ \tag{2.11}$$

with domain defined as for A except that the boundary condition is now

$$u(\pm a, \mu, t) = 0, \qquad \mu \gtrless 0, \qquad t > 0.$$

Note that A^* differs from A only in the sign of the differential operator D (and in its domain). The mapping $f^*(x, \mu) = \overline{f(-x, \mu)}$ carries $\mathscr{D}(A)$ onto $\mathscr{D}(A^*)$ in a one-one manner, and carries the point spectrum into its complex conjugate. Hence, the point spectra of A and A^* are identical. By a known theorem (Stone, Linear Transformations in Hilbert Space, p. 143), if λ is in the residual spectrum of A, then $\bar{\lambda}$ is in the point spectrum of A^*. Thus, λ would be in both the point and residual spectra of A. Hence there is no residual spectrum.

From what has been said, a point λ in the left half-plane is either in the continuous spectrum of A or in the resolvent set. But we can show by an example that $(\lambda-A)^{-1}$ is unbounded, and thus λ must be in the spectrum. Namely, let ($\mathscr{R}e\,\lambda < 0$)

$$f_\delta = e^{-\lambda(x+a)/\mu}\cdot\left(\frac{x+a}{a}\right) = \begin{cases} 1/\delta \text{ for } \delta^2 \leqq \mu < \delta, \\ 0 \text{ otherwise} \end{cases} \tag{2.12}$$

where $0 < \delta < \frac{1}{2}$. Then for $\delta \to 0$ it is easily verified that

$$||f_\delta|| > C > 0, \qquad ||(\lambda-A)f_\delta|| \to 0.$$

Therefore, the entire left half-plane is in the continuous spectrum, and, in fact, the entire closed half-plane $\mathscr{R}e\,\lambda \leqq 0$, since the spectrum is a closed point set.

The right half-plane, on the other hand, is the whole resolvent set (except for the point spectrum). Suppose λ is in the right half-plane, but not in the point spectrum. We can obtain an explicit expression for the resolvent by integrating the equation $(\lambda-A)h = g$, $g \in H$. We get for $\mu > 0$, $\mathscr{R}e\,\lambda > 0$

$$\mu h(x,\mu) = \int_{-a}^{x} e^{-\lambda(x,x')/\mu}\left[\frac{c}{2}\xi(x') + g(x',\mu)\right]dx', \tag{2.13}$$

where ξ is the solution of the integral equation

$$\frac{2}{c}\xi = \mathscr{L}_\lambda \xi + G_\lambda.$$

Here G_λ is a sort of transform of g,

$$G_\lambda(x) = \int_0^1 \frac{d\mu}{\mu}\int_{-a}^{x} e^{-\lambda(x-x')/\mu} g(x',\mu)dx' - \int_{-1}^{0}\frac{d\mu}{\mu}\int_{x}^{a} e^{-\lambda(x-x')/\mu} g(x',\mu)\,dx'. \tag{2.14}$$

A similar equation holds for $\mu < 0$. From (2.14) we can estimate

$$\int_{-a}^{a} |G_\lambda(x)|^2 dx \leqq \beta^{-2}||g||^2,$$

and then the integral equation gives a bound on ξ:

$$\int_{-a}^{a} |\xi(x)|^2 dx \leqq c_1||g||^2,$$

where c_1 depends only on λ. Then we get from (2.13)

$$||h|| = ||\,R(\lambda, A)\,g|| \leqq c_2||g||,$$

c_2 being a function of λ alone. Hence, the operator $(\lambda-A)^{-1}$ is bounded and λ is in the resolvent set of A. By more detailed methods one can show that

$$c_2 = \frac{b_1}{\beta^2}, \qquad |\tau| > b_2 \tag{2.15}$$

where b_1, b_2 depend only on the physical parameter c.

Putting our results together, we have found the spectral decomposition of A:

(2.16)
Point spectrum: $\beta_1 > \beta_2 > \cdots > \beta_m$, $m \geqq 1$.
Resolvent set: the remainder of the half-plane $\mathscr{R}e\,\lambda > 0$.
Continuous spectrum: $\mathscr{R}e\,\lambda \leqq 0$.
Residual spectrum: empty.

§3. We now know that we must shift the path of integration in (1.8) to the left, crossing the singularities $\beta_1, \cdots, \beta_m$ of the resolvent $R(\lambda, A)$. These singularities are poles or isolated essential singularities, and we wish to show that they are simple poles.

In order to do this, we shall use the biorthonormality of the eigenfunctions $\psi_1, \psi_2, \cdots, \psi_1^*, \psi_2^*, \cdots$. But it is not clear in advance that the system of eigenfunctions can be normalized. What we must prove is that if $A\psi = \lambda\psi$, then $(\psi, \psi^*) \neq 0$. The proof is not based on general considerations but on a detailed calculation specific to the function ψ and the relation (2.2) between ψ and φ.

Now let

$$R(\lambda, A)f = \sum_1^\infty B_n f \cdot (\lambda-\beta_j)^{-n} + A_0 f + A_1 f \cdot (\lambda-\beta_j) + \cdots.$$

If $B_k \neq 0$, $k > 1$, then $B_{k-1} \neq 0$, for $B_k = (A-\lambda)B_{k-1}$, by the operational calculus (Taylor, Acta Math., Vol. 84, p. 213). Hence, $B_2 \neq 0$. Therefore, in order to show that the resolvent has only simple poles it is sufficient to prove $B_2 = 0$.

Now $B_2 \neq 0$ implies the existence of an element of order 2, i.e., a $g \in \mathscr{D}(A)$ such that

$$(\beta_j-A)^2 g = 0, \qquad (\beta_j-A)\,g \neq 0$$

(Taylor, p. 217). It follows that $(\beta_j-A)g$ is an eigenfunction; expand it in the appropriate eigenfunctions and use the biorthonormal relations referred to above. All the coefficients of the expansion turn out to be zero, and thus $(\beta_j-A)g = 0$, a contradiction. Hence, $B_2 = 0$, as required.

A simple calculation now gives the residue:

$$\operatorname*{Res}_{\lambda=\beta_j} R(\lambda, A)f = (f, \psi_j^*)\psi_j, \qquad j = 1, 2, \cdots, m.$$

(The above discussion has assumed the simplicity of the eigenvalues; the case of multiple eigenvalues is handled in the same way.)

From (1.8) we now get

$$T(t)f - \sum_{j=1}^m e^{\beta_j t}(f, \psi_j^*)\psi_j = \frac{1}{2\pi i}\int_R e^{\lambda t} R(\lambda, A)f\,d\lambda, \tag{3.1}$$

where R consists of the horizontal lines $\beta \pm i\omega$, $\delta \leqq \beta \leqq \gamma$, and the vertical

line $\delta+i\tau$, $|\tau| \leqq \omega$, and $0 < \delta < \beta_m$. Using the identity $\lambda R(\lambda, A)f = f+R(\lambda, A)\cdot Af$, $f \in \mathcal{D}(A)$, and the estimate (2.15), we have

$$||R(\lambda, A)f|| \leqq \frac{||f|| + c_2(\beta)\,||Af||}{|\lambda|}.$$

Hence the contribution to the integral in (3.1) from the horizontal lines goes to 0 as $\omega\to\infty$, while the integral on the vertical line is estimated as follows:

$$\left\|\frac{1}{2\pi i}\int_R\right\| = \left\|\frac{1}{2\pi i}\int_{\mathrm{Re}\,\lambda=\delta} e^{\lambda t}\cdot f\cdot\frac{d\lambda}{\lambda}\right\| + \left\|\frac{1}{2\pi i}\int_{\mathrm{Re}\,\lambda=\delta} e^{\lambda t}R(\lambda, A)\cdot Af\frac{d\lambda}{\lambda}\right\|$$

$$\leqq ||f|| + \frac{1}{2\pi}e^{\delta t}\int_{-\infty}^{\infty} c_3(f)\,\frac{d\tau}{\delta^2+\tau^2} = ||f|| + \frac{c_3(f)\,e^{\delta t}}{2\delta}.$$

Choosing $\delta = 1/t$, we obtain

THEOREM 1. *The solution of the problem* (1.1), (1.2) *is given by*

$$u(x, \mu, t) = \sum_{j=1}^{m} e^{\beta_j t}(f, \psi_j^*)\,\psi_j + \zeta(x, \mu, t),$$

provided $f \in \mathcal{D}(A^2)$, *where*

$$||\zeta(x, \mu, t)|| \leqq ||f|| + c_4(f)\cdot t,$$

c_4 *depending only on the initial distribution* $f(x, \mu)$.

This result is satisfactory in that it exhibits an error term ζ of much lower magnitude than the principal terms. More precise results can be obtained only by a very detailed investigation of the resolvent and apparently by further restricting the class of initial distributions f. Our best result in this direction is

THEOREM 2. *When* f *is such that*

$$f \in \mathcal{D}(A), \qquad \left\|\frac{\partial f}{\partial x}\right\| < \infty,$$

$$Af \in \mathcal{D}(A), \qquad \left\|\frac{\partial (Af)}{\partial x}\right\| < \infty$$

we have

$$\lim_{t\to\infty} \zeta(x, \mu, t) = 0$$

for almost all fixed (x, μ) *in* $|x| \leqq a$, $|\mu| \leqq 1$. Note that the result is given for ζ itself instead of $||\zeta||$.

We shall not prove Theorem 2 here, but refer the reader to [2].

It is interesting to observe that the order of magnitude of ζ cannot be significantly lowered (unless f is an eigenfunction, in which case of course $\zeta = 0$). This is shown by [2], Theorem 3.

THEOREM 3. *Let* $f = \sum_{j=1}^{m} (f, \psi_j^*)\psi_j + g$. *Then for each* $\varepsilon > 0$

$$||\zeta[g]|| > C_\varepsilon e^{-\varepsilon t}.$$

This result is obtained by using the spectral radius theory of Gelfand (Riesz-Nagy, Leçons d'Analyse Fonctionelle, p. 414).

§4. We now take up the problem of the sphere, where, as previously mentioned, our results are not complete. On the assumption of the usual symmetries, the natural coordinates would be r, the radial distance from the center of the sphere, and θ, the angle between the direction of the neutron beam and the radius vector. However, a simplification results if we introduce the coordinates

$$x = r\cos\theta, \qquad y = r\sin\theta, \tag{4.1}$$

a procedure originally suggested by J. von Neumann. The equations are now

$$\begin{aligned} &\frac{\partial u}{\partial t} = Bu, \quad u = u(x, y, t), \quad x^2+y^2 \leqq a^2, && y \geqq 0, \quad t > 0, \\ &u(x, y, t) = 0, \quad x^2+y^2 = a^2, && x \leqq 0, \\ &u(x, y, 0) = f(x, y) \end{aligned} \tag{4.2}$$

where again u is essentially the neutron density, and

$$\begin{aligned} Bf &= -\frac{\partial f}{\partial x} + \frac{c}{2r}\int_{-r}^{r} f(x', \sqrt{r^2-x'^2}, t)\,dx', \\ r &= \sqrt{x^2+y^2}, \qquad a = \text{radius of sphere.} \end{aligned} \tag{4.3}$$

In order to define the domain $\mathscr{D}(B)$ of B, we first define the Hilbert space I of functions $g(x, y, t)$ over $x^2+y^2 \leqq a^2$, $t > 0$ for which $||g|| < \infty$, where the norm is given by

$$||g|| = \left[\int_0^a y\,dy \int_{-Y}^{Y} |g(x, y)|^2\,dx\right]^{1/2} \tag{4.4}$$

and

$$Y = \sqrt{a^2-y^2}, \qquad 0 \leqq y \leqq a.$$

Then the domain, $\mathscr{D}(B)$ of B is the class of all functions $f(x, y, t) \in I$ for which

$$\begin{aligned} &f \text{ is absolutely continuous in } -Y \leqq x < Y \text{ for} \\ &\text{fixed } y \text{ in } 0 < y < a \quad (\text{fixed } t > 0), \quad Af \in I; \\ &f(x, y, t) = 0, \quad x^2+y^2 = a^2, \quad x \leqq 0, \quad t > 0. \end{aligned} \tag{4.5}$$

Under the above norm B possesses an adjoint

$$B^* = D+cK, \qquad D = \frac{\partial}{\partial x}, \qquad K = \frac{c}{2r}\int_{-r}^{r} \cdots. \tag{4.6}$$

The domain of B^* is defined in an obvious way, and $\mathscr{D}(B)$, $\mathscr{D}(B^*)$ are isometric under the mapping

$$f^*(x, y) = \overline{f(-x, y)}.$$

The general procedure is the same as in the slab case. The problem is solved by the semi-group of operators $U(t)$, $t \geqq 0$, where

$$U(0) = 1, \qquad U(t) = e^{tB}. \tag{4.7}$$

To prove the existence of U we work slightly differently, using the perturbation theory of R. S. Phillips (Trans. Amer. Math. Soc., Vol. 74, 1953, pp. 199—221). We show exactly as before that $-D$ generates a semi-group of operators; then so does B, since K is a bounded operator (Phillips, p. 205).

Also, $R(\lambda, -D)$ is compact since we have

$$R(\lambda, -D)f = \int_{-y}^{x} e^{-\lambda(x-x')} f(x', y)\, dx'. \tag{4.8}$$

Hence, by Phillips' Theorem 4.3 (p. 209), $R(\lambda, B)$ is compact and B has a pure point spectrum $\{\lambda_1, \lambda_2, \cdots; \beta_1 \geqq \beta_2 \geqq \cdots; \lambda_n = \beta_n + i\tau_n\}$ consisting of isolated points. The spectrum of B is either finite, or, if infinite, accumulates only at the infinite point of the plane.

By methods similar to those in P. D. Lax [3] we can prove, on the assumption of an infinite spectrum,

THEOREM 4. *For $f \in \mathscr{D}(A)$, we have*

$$\left\| T(t)f - \sum_{j=1}^{n} (f, \psi_j^*)\psi_j e^{\lambda_j t} \right\| \leqq C_n e^{(\beta_{n+1}+\varepsilon)t}, \qquad t > 0, \tag{4.9}$$

for each $\varepsilon > 0$ and $n = 1, 2, 3, \cdots$.

If the spectrum is finite, (4.9) becomes

$$\left\| T(t)f - \sum_{j=1}^{N} (f, \psi_j^*)\psi_j e^{\lambda_j t} \right\| \leqq C_s e^{-st}, \qquad t > 0,$$

where N is the number of eigenvalues and $s > 0$ is arbitrary.

Formula (4.9) is an *asymptotic expansion* of the solution. To get a better result, we would, just as before, express the solution as an integral:

$$T(t)f = \lim_{\omega \to \infty} \frac{1}{2\pi i} \int_{\gamma - i\omega}^{\gamma + i\omega} e^{\lambda t} R(\lambda, B)f\, d\lambda, \qquad t > 0,\ f \in \mathscr{D}(B),$$

where $\gamma > \mathscr{Re}\, \lambda_1$. We would again shift the line of integration to the left, collecting the residues at the poles of the resolvent (points of the spectrum). If the remaining integral $\to 0$ as $\gamma \to -\infty$, *we would obtain a series representation for $T(t)f$.* This is the point at which a new difficulty arises, one which we have not yet surmounted.

In general we cannot expect a series representation for all $t > 0$. For, $||R(\lambda, -D)|| = O(e^{-2a\lambda})$ for $\mathscr{R}e\,\lambda < 0$, and $||R(\lambda, B)||$ would very likely be no smaller. Now if we had an estimate

$$||R(\lambda, B)|| \leqq Ce^{-b\beta}(1+|\tau|)^{-2}, \qquad \beta < 0, \tag{4.10}$$

for some $b > 0$, then we could assert

$$\lim_{\gamma\to-\infty} \left\| \frac{1}{2\pi i}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\lambda t} R(\lambda, B) f\, d\lambda \right\| \leqq \lim_{\gamma\to-\infty} \frac{C}{2\pi}\int_{-\infty}^{\infty} e^{\gamma t-\gamma b}\frac{d\tau}{(1+|\tau|)^2} = 0,$$

provided $t > b$. *So for $t > b$, we would get a series representation*

$$T(t)f = \sum_{j=1}^{\infty} e^{\lambda_j t}(f, \psi_j^*)\psi_j, \qquad t > b. \tag{4.11}$$

The great disparity between the mathematical structure of the solutions in the slab and sphere geometries can perhaps be explained on physical grounds. If we normalize the neutron velocity at unity, the time required for a neutron to traverse a diameter of the sphere is $2a$ provided it does not suffer collisions. That is, as soon as $t > 2a$, every neutron in the initial distribution of neutrons has either left the sphere or collided at least once. Since collision is a smoothing process, we might expect the neutron distribution $T(t)f$ to be representable in a simple way (as by a series of eigenfunctions) for $t > 2a$, *but not necessarily for $t \leqq 2a$*. Our method would give this result if we could prove (4.10) with $b = 2a$.

On the other hand, this situation does not obtain in the slab. A neutron traveling parallel to the sides of the slab may remain indefinitely in the slab without colliding. A beam of neutrons of this sort would be damped at a slower rate ($\beta = 0$) than a beam represented by any eigenfunction, $\beta = \beta_1, \cdots, \beta_m > 0$, and, therefore, could not be included in any linear combination of the finite set of eigenfunctions. Hence, an additional term in the solution is needed and this is in fact the residual integral term ζ of Theorem 1.

We have not yet succeeded in establishing (4.10). A sizable quantity of information about $R(\lambda, B)$ has been gathered, and we hope that this will lead in time to the required estimate.

Bibliography

[1] Lehner, J., and Wing, G. M., *On the spectrum of an unsymmetric operator arising in the transport theory of neutrons*, Comm. Pure Appl. Math., Vol. 8, 1955, pp. 217–234.

[2] Lehner, J., and Wing, G. M., *Solution of the linearized Boltzmann equation for the slab geometry*, Duke Math. J., Vol. 23, 1956, pp. 125–142.

[3] Lax, P. D., *Spectral theory of functions on semigroups and the separation of variables*, Studies in Eigenvalue Problems, Technical Report 14, Conference on Partial Differential Equations, Univ. of Kansas, Summer 1954.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 499–508 (1956)

A Variational Method in the Theory of Harmonic Integrals*

CHARLES B. MORREY, JR.
University of California, Berkeley

1. Introduction

This paper is a report of work done by the writer and Dr. James Eells, Jr. on this subject at the Institute for Advanced Study during the year 1954—55. This work is contained in parts I and II of a paper having the title above. Part I, by the writer and J. Eells, has appeared in the Annals of Mathematics [11] and part II, by the writer as sole author, in the American Journal of Mathematics [12]. A brief report on part I has appeared in the Proceedings of the National Academy of Sciences [10]. Parts I and II also appeared as Air Force OSR-Technical Notes. Part I considers the case of compact manifolds without boundary and part II considers the case of those with boundary. In this report, references are kept to a minimum; additional references are found in the papers referred to above. Some of the results have been obtained independently and by different methods by K. O. Friedrichs [5].

2. Notation and Terminology

A function f on an open set G in E^n is said to be of *class* C^k_μ, $0 < \mu \leqq 1$, if it possesses partial derivatives of order $\leqq k$ which satisfy uniform Hölder conditions for $\mu<1$ and uniform Lipschitz conditions for $\mu=1$. We say that f is *of class* $\mathfrak{L}_2$ on G if f is square-integrable on G. The function f is *of class* $\mathfrak{P}_2$ on G if f is in $\mathfrak{L}_2$ on G and there are functions $g_1, \cdots, g_n$ in $\mathfrak{L}_2$ on G such that, if R is any cell interior to G, there is a sequence of Lipschitz functions $\{f_j\}$ on $\overline{R}$ such that

$$\lim_{j\to\infty}\int_R \left[(f_j-f)^2+\sum_{k=1}^{n}\{(\partial f_j/\partial x^k)-g_k\}^2\right]dx = 0. \tag{1}$$

Functions of class $\mathfrak{P}_2$ have been studied in detail elsewhere (see Calkin [1] and Morrey [8]) and have been used extensively in potential theory and the calculus of variations.

*The research reported on herein was partially supported by the United States Air Force, through the Office of Scientific Research of the Air Force Research and Development Command.

Riemannian n-manifolds $\mathfrak{M}$ of class C_μ^k are defined using coordinate transformations of class C_μ^k; in case $\mathfrak{M}$ has a boundary, we assume that each boundary point is in a neighborhood which is the image, under a coordinate system of class C_μ^k, of that part of the ball $B(0, R)$, $|x| < R$, for which $x^n \leqq 0$ — the part where $x^n = 0$ corresponds to a subset of the boundary. On a Riemannian manifold $\mathfrak{M}$ of class C_μ^k, $k \geqq 1$, the metric tensor g_{ij} is assumed to be of class C_μ^{k-1}. Exterior differential forms of degree r (called r-forms, $r = 0, \cdots, n$) on $\mathfrak{M}$ are of class $\mathfrak{L}_2$, $\mathfrak{P}_2$, C_μ^k, C^∞, or analytic if their components in every admissible coordinate system on $\mathfrak{M}$ have these respective properties. We do not assume that $\mathfrak{M}$ is orientable and parallel theories are developed for the "even" and "odd" forms (see [3]).

Let $\mathfrak{L}_2^r$ denote the real Hilbert space obtained by completing the linear space of all Lipschits r-forms on $\mathfrak{M}$ (r-forms on $\mathfrak{M}$ whose components satisfy a Lipschitz condition) with respect to the invariant inner product

$$(\omega, \eta) = \int_{\mathfrak{M}} \omega \wedge *\eta, \tag{2}$$

where $\omega \wedge *\eta$ denotes the exterior product of ω with the adjoint $(n-r)$-form $*\eta$ of η; for notation see de Rham and Kodaira [3]. Let $\mathfrak{P}_2^r$ denote the real Hilbert space obtained by completing the linear space of all Lipschitz r-forms on $\mathfrak{M}$ with respect to any of the inner products given in equation (3) below.

Let $\mathfrak{U}$ denote a covering of $\mathfrak{M}$ by co-ordinate patches $\mathfrak{U}_1, \cdots, \mathfrak{U}_Q$, whose domains in E^n are $G_1, \cdots, G_Q$. Then we define the inner product

$$((\omega, \eta))_{\mathfrak{U}} = (\omega, \eta) + \sum_{q=1}^{Q} \int_{G_q} \sum_{(i)\alpha} \omega_{(i)x^\alpha} \eta_{(i)x^\alpha} \, dx, \tag{3}$$

where $\omega_{(i)x^\alpha} = \partial \omega_{(i)} / \partial x^\alpha$; here and later $\alpha = 1, \cdots, n$ and $(i) = i_1 \cdots i_r$, with $i_1 < \cdots < i_r$. Any two such inner products are topologically equivalent.

We let d denote the exterior differential operator on $\mathfrak{M}$ and δ its dual. It is easy to see that d is a bounded linear transformation from the whole of $\mathfrak{P}_2^{r-1}$ into $\mathfrak{L}_2^r$. Similarly, δ is a bounded linear transformation from $\mathfrak{P}_2^{r+1}$ into $\mathfrak{L}_2^r$. We let $\mathfrak{H}^r$ denote the subspace of $\mathfrak{P}_2^r$ of harmonic fields on $\mathfrak{M}$; i.e., $\omega \in \mathfrak{H}^r$ means that ω is an r-form in $\mathfrak{P}_2^r$ such that $d\omega = \delta\omega = 0$. An r-form ω in $\mathfrak{P}_2^r$ is said to be harmonic on $\mathfrak{M}$ if and only if $d\omega$ and $\delta\omega$ are also in $\mathfrak{P}_2^r$ on interior subsets of $\mathfrak{M}$ with

$$\Delta\omega \equiv \delta(d\omega) + d(\delta\omega) = 0.$$

If $\mathfrak{M}$ is compact and without boundary, then any harmonic form is a harmonic field; this is not true in case $\mathfrak{M}$ has a boundary.

3. Preliminary Results

The following convergence theorem is essentially due to Rellich [13]:

A. *Weak convergence in $\mathfrak{P}_2^r$ implies strong convergence in $\mathfrak{L}_2^r$.*

If we define the (weak) Dirichlet integral of an r-form by

$$D(\omega) = (d\omega, d\omega) + (\delta\omega, \delta\omega) \tag{4}$$

then the lower semicontinuity of the norm and the boundedness of d and δ imply

B. *The Dirichlet integral is lower semicontinuous with respect to weak convergence in $\mathfrak{P}_2^r$.*

If $D_0(\omega)$ denotes the (weak) Dirichlet integral in the case of a domain G in E^n, then an inspection of equation (4) shows that we may write

$$\begin{aligned} D_0(\omega) &= \int_G \sum_{(i)\alpha} (\omega_{(i)x^\alpha})^2\, dx \\ &\quad + 2 \sum_{(i)(j)} \sum_{\alpha,\beta} \int_G (\omega_{(i)x^\alpha}\omega_{(j)x^\beta} - \omega_{(i)x^\beta}\omega_{(j)x^\alpha})\, dx. \end{aligned} \tag{5}$$

The second integral vanishes for any ω of class $\mathfrak{P}_2$ with support interior to G. This fact allows us to give a simple proof of the following lemma, essentially due to Gaffney [6]:

With each point p of a manifold $\mathfrak{M}$ (without boundary), each $\varepsilon > 0$, and $r = 0, \cdots, n$, we may associate a constant L and a co-ordinate patch $\mathfrak{U}$ containing p with domain B_R, $|x| < R$, such that

$$D(\omega) \geqq (1-\varepsilon) \int_{B_R} \sum_{(i)\alpha} (\omega_{(i)x^\alpha})^2 dx - L\cdot(\omega, \omega) \tag{6}$$

for every $\omega \in \mathfrak{P}_2^r$ with support in $\mathfrak{U}$.

By using a partition of unity, we may easily show

C. *With each finite covering $\mathfrak{U}$ of a compact manifold $\mathfrak{M}$ without boundary by co-ordinate patches and each $r = 0, \cdots, n$, we can associate a pair of constants $K_{\mathfrak{U}}$ and $L_{\mathfrak{U}}$ such that*

$$D(\omega) \geqq K_{\mathfrak{U}}\cdot((\omega, \omega))_{\mathfrak{U}} - L_{\mathfrak{U}}\cdot(\omega, \omega), \qquad K_{\mathfrak{U}} > 0 \tag{7}$$

for all $\omega \in \mathfrak{P}_2^r$.

4. The Variational Method for a Compact Manifold without Boundary

Using **A**, **B**, and **C** of Section 3, we apply our variational method to this case as follows:

LEMMA 1. *Suppose that $\mathfrak{M}$ is any closed linear manifold in $\mathfrak{L}_2^r$ such that $\mathfrak{M} \cap \mathfrak{P}_2^r$ is not empty. Then there is a form ω in $\mathfrak{M} \cap \mathfrak{P}_2^r$ which minimizes $D(\omega)$ among all such forms with $(\omega, \omega) = 1$.*

For, let $\{\omega_k\}$ be a minimizing sequence, and take a co-ordinate covering $\mathfrak{U}$ of $\mathfrak{M}$. From relation (7) we see that $((\omega_k, \omega_k))_{\mathfrak{U}}$ is uniformly bounded. Hence a subsequence converges weakly in $\mathfrak{P}_2^r$ to some form ω in $\mathfrak{P}_2^r$. The lemma now follows from **A** and **B**.

THEOREM 1. *For each $r = 0, \cdots, n$ the linear manifold $\mathfrak{H}^r$ of harmonic fields on $\mathfrak{M}$ is finite dimensional.*

We choose linear manifolds $\mathfrak{M}_1, \mathfrak{M}_2, \cdots$ in $\mathfrak{L}_2^r$ and the minimizing forms ω_k in $\mathfrak{M}_k$, $k = 1, 2, \cdots$, where $\mathfrak{M}_1 = \mathfrak{L}_2^r$, $\mathfrak{M}_2$ is orthogonal to ω_1, $\mathfrak{M}_3$ is orthogonal to ω_1 and ω_2, etc. Then $0 \leqq D(\omega_1) \leqq D(\omega_2) \leqq \cdots$. Clearly ω_k is a harmonic field if and only if $D(\omega_k) = 0$. Suppose that this held for all k; then relation (7) would imply that $((\omega_k, \omega_k))_{\mathfrak{U}}$ is uniformly bounded, whence a subsequence $\{\omega_j\}$ would converge weakly in $\mathfrak{P}_2^r$, and hence strongly in $\mathfrak{L}_2^r$, to some ω_0. This is impossible, since the ω_j form an orthonormal set in $\mathfrak{L}_2^r$.

THEOREM 2. *For any co-ordinate covering $\mathfrak{U}$ of $\mathfrak{M}$ there is a constant $\lambda_{\mathfrak{U}} > 0$ such that, if ω is in $\mathfrak{P}_2^r$ and ω is orthogonal to $\mathfrak{H}^r$, then*

$$D(\omega) \geqq \lambda_{\mathfrak{U}} \cdot ((\omega, \omega))_{\mathfrak{U}}.$$

From Theorem 1 it follows that $\mathfrak{H}^r$ is spanned by a finite number $\omega_1, \cdots, \omega_k$ of the above sequence. For any ω in $\mathfrak{P}_2^r$ and orthogonal to $\mathfrak{H}^r$, it is easy to see from the homogeneity of $D(\omega)$ that

$$D(\omega) \geqq D(\omega_{k+1}) \cdot (\omega, \omega) = h \cdot (\omega, \omega), \qquad h > 0. \tag{8}$$

The result now follows from relation (7).

THEOREM 3. *If ω_0 is any form in $\mathfrak{L}_2^r$ which is orthogonal to $\mathfrak{H}^r$, then there is a unique form Ω_0 in $\mathfrak{P}_2^r$ which is orthogonal to $\mathfrak{H}^r$ and satisfies*

$$(d\Omega_0, d\zeta) + (\delta\Omega_0, \delta\zeta) = (\omega_0, \zeta) \tag{9}$$

for all $\zeta \in \mathfrak{P}_2^r$. Ω_0 is called the potential of ω_0.

By setting $\zeta = \Omega_1 - \Omega_0$, it is easy to see that Ω_0 is unique. Ω_0 is found by minimizing

$$I(\omega) = D(\omega) - 2(\omega, \omega_0) \tag{10}$$

among all ω in $\mathfrak{P}_2^r$ which are orthogonal to $\mathfrak{H}^r$. To do this, it is sufficient to select minimizing sequences $\{\omega_k\}$ for which $I(\omega_k) \leqq 0$; clearly, by Theorem 2, $((\omega_k, \omega_k))_{\mathfrak{U}}$ is uniformly bounded for any such sequence.

All the results above hold on any manifold of class C_1^1 or better. In case $\mathfrak{M}$, Ω_0, ω_0 are sufficiently differentiable, equation (9) implies that $\Delta\Omega_0 = \omega_0$. However, one sees easily that the *second* derivatives of the g_{ij}, as well as those of the components of Ω_0, enter into the formulas for $\Delta\Omega_0$ if $r > 0$; in our most general case none of these derivatives is at hand.

5. The Differentiability of Potentials and Harmonic Fields on a Manifold without Boundary

We say that ω is of class $\mathfrak{L}_{2\lambda}$, $0 \leqq \lambda < n/2$, if, for each coordinate patch $\mathfrak{U}$ with domain B_R, there is a constant $L = L(\mathfrak{U}, \omega)$ such that

$$\int_{B_r} \sum_{(i)} \omega_{(i)}^2 \, dx \leqq L^2 r^{2\lambda}, \qquad 0 \leqq r \leqq R; \tag{11}$$

$\mathfrak{P}_{2\lambda}$ is defined similarly. The importance of the spaces $\mathfrak{P}_{2\lambda}$ arises from the fact that if $\omega \in \mathfrak{P}_{2\lambda}$ with $\lambda = n/2-1+\mu$, $0 < \mu < 1$, then $\omega \in C_\mu^0$. We may now state the following results concerning differentiability:

THEOREM 4. *Suppose that $\omega \in \mathfrak{L}_2^r \ominus \mathfrak{H}^r$ and Ω is its potential.*

(i) *If $\mathfrak{M}$ is of class C_1^1, then Ω, $d\Omega$, and $\delta\Omega \in \mathfrak{P}_2$.*

(ii) *If $\mathfrak{M}$ is of class C_1^1, and $\omega \in \mathfrak{L}_{2\lambda}$, then Ω, $d\Omega$, and $\delta\Omega \in \mathfrak{P}_{2\lambda}$ and hence in C_μ^0 if $\lambda = n/2-1+\mu$, $0 < \mu < 1$.*

(iii) *If $\mathfrak{M}$ is of class C_1^1 and $\omega \in \mathfrak{P}_2$, then $d\Omega$ and $\delta\Omega$ are the potentials of $d\omega$ and $\delta\omega$, respectively.*

(iv) *If $\mathfrak{M}$ is of class C_μ^k and $\omega \in C_\mu^{k-2}$, $k \geqq 2$, $0 < \mu < 1$, then Ω, $d\Omega$ and $\delta\Omega \in C_\mu^{k-1}$. If $k \geqq 3$ and $\omega \in C_\mu^{k-3}$, then $\Omega \in C_\mu^{k-1}$.*

(v) *If $\mathfrak{M}$ and ω are of class C^∞ or analytic, then so is Ω.*

In all cases, if we set $\alpha = d\Omega$ and $\beta = \delta\Omega$, we have

$$\delta\alpha + d\beta = \delta(d\Omega) + d(\delta\Omega) = \omega, \qquad d\alpha = \delta\beta = 0. \tag{12}$$

THEOREM 5. *Suppose that H is a harmonic field.*

(i) *If $\mathfrak{M} \in C_1^1$, then $H \in \mathfrak{P}_{2\lambda}$ with $\lambda = n/2-1+\mu$ for any μ, $0 < \mu < 1$.*

(ii) *If $\mathfrak{M} \in C_\mu^k$, $k \geqq 2$, $0 < \mu < 1$, then $H \in C_\mu^{k-1}$.*

(iii) *If $\mathfrak{M} \in C^\infty$ or is analytic, then so is H.*

In both Theorems 4 and 5, 0-forms have an additional degree of differentiability (except in the second part of Theorem 4(iv)). It should be observed that we can form $\Delta\Omega$ as indicated in (12) even though the individual components of Ω do not have the necessary second derivatives (if $r > 0$).

The results for Ω and H follow from a simplification of the writer's previous work on the solutions of second order elliptic systems [9]. The long papers referred to contain a new short proof in the analytic case. To get the differentiability of $d\Omega$ and $\delta\Omega$, we select a coordinate patch and approximate to ω, Ω, and the g_{ij} by smooth functions. At each step, we see that (12) and integration by parts imply that α and β satisfy

$$\begin{aligned} (d\alpha, d\zeta) + (\delta\alpha - \omega, \delta\zeta) &= 0, \\ (d\beta - \omega, d\zeta) + (\delta\beta, \delta\zeta) &= 0. \end{aligned} \qquad \zeta \in \mathfrak{P}_2, \tag{13}$$

Equations (13) are equivalent (using (5)) to those of the type previously considered. Since the $\mathfrak{L}_2$ norms of α and β over the patch are uniformly

bounded, the $\mathfrak{P}_2$ norms over interior domains are (using a theorem originally stated by Shiffman [14]) also and an approximation theorem shows that (13) holds in the limit.

6. The Kodaira Decomposition on Manifolds without Boundary

The following theorem complements the well-known orthogonal decomposition theorem of Kodaira [7].

THEOREM 6. *If ω is any form in $\mathfrak{L}_2$, then there exists a harmonic field H and forms α, β, and Ω in $\mathfrak{P}_2$ such that*

$$\omega = H + \delta\alpha + d\beta, \quad d\alpha = \delta\beta = 0, \qquad \alpha = d\Omega, \quad \beta = \delta\Omega, \tag{14}$$

where Ω is the potential of $\omega - H$. If the first equation of (14) holds for a harmonic field H_1 and forms α_1 and β_1 in $\mathfrak{P}_2$, then $H_1 = H$, $\delta\alpha_1 = \delta\alpha$, and $d\beta_1 = d\beta$.

The sets $\mathfrak{C}^r$ of all forms $\delta\alpha$ for α in $\mathfrak{P}_2^{r+1}$ and $\mathfrak{D}^r$ of all forms $d\beta$ for β in $\mathfrak{P}_2^{r-1}$ are closed linear manifolds in $\mathfrak{L}_2^r$ and

$$\mathfrak{L}_2^r = \mathfrak{H}^r \oplus \mathfrak{C}^r \oplus \mathfrak{D}^r. \tag{15}$$

If $\mathfrak{M} \in C_1^1$ and $\omega \in \mathfrak{L}_{2\lambda}$ or $\mathfrak{P}_{2\lambda}$, $0 \leqq \lambda < n/2$, then $\delta\alpha$ and $d\beta$ have the same properties.

If $\mathfrak{M} \in C_\mu^k$ and $\omega \in C_\sigma^l$ with $k \geqq 2$, $0 < \mu < 1$, $0 < \sigma < 1$, and either $l < k-1$ or $l = k-1$ and $\sigma \leqq \mu$, then $\delta\alpha$ and $d\beta$ have the same differentiability properties as ω.

If $\mathfrak{M}$ and $\omega \in C^\infty$ or are analytic, so are $\delta\alpha$ and $d\beta$.

This theorem follows immediately from Theorems 4 and 5, equation (14) and the following well-known relations

$$(\omega, d\eta) = (\delta\omega, \eta), \qquad (\delta\omega, d\eta) = 0 \tag{16}$$

for ω and η in $\mathfrak{P}_2$.

7. The Variational Method on Compact Manifolds with Boundary

If ω is an r-form in $\mathfrak{P}_2$ on a Riemannian manifold $\mathfrak{M}$ of class at least C_1^1, we define the tangential and normal components of ω along $\mathfrak{B}$ (the boundary of $\mathfrak{M}$) by the conditions that

$$t\omega = \sum_{i_1 < \cdots < i_r < n} \omega_{i_1 \cdots i_r} dx^{i_1} \wedge \cdots \wedge dx^{i_r}, \qquad n\omega = \omega - t\omega \tag{17}$$

in any boundary coordinate system as described in Section 2; $t\omega$ and $n\omega$ are invariantly defined only along $\mathfrak{B}$. The relations (16) hold if

(18) $$\text{either } n\omega = 0 \text{ or } t\eta = 0.$$

We also have

(19) $$t*\omega = *n\omega, \qquad n*\omega = *t\omega.$$

The manifolds $\mathfrak{P}_2^+$ of all ω in $\mathfrak{P}_2$ (for any given r) with $n\omega = 0$ and $\mathfrak{P}_2^-$ of all ω in $\mathfrak{P}_2$ with $t\omega = 0$ are closed linear manifolds in $\mathfrak{P}_2$. Theorems **A** and **B** of Section 3 carry over to the present case and the formula (5) enables us to prove the Gaffney lemma and Theorem **C** for each of the manifolds $\mathfrak{P}_2^+$ and $\mathfrak{P}_2^-$. Thus the variational method and the lemma and theorems of Section 4 carry over without change to $\mathfrak{P}_2^+$ and $\mathfrak{P}_2^-$. If $\omega \in \mathfrak{L}_2$, we denote its potentials in $\mathfrak{P}_2^+$ and $\mathfrak{P}_2^-$ by Ω^+ and Ω^-, respectively. The differentiability results of Section 5 are shown to hold for Ω^+ and Ω^- right up to the boundary from which it follows that $n\delta\Omega^+ = td\Omega^- = 0$. The fact that (9) holds for all ζ in $\mathfrak{P}_2^+$ in the case of Ω^+, and $t\zeta$ is thus arbitrary, leads to a natural boundary condition from which we conclude that $nd\Omega^+ = 0$; correspondingly we find that $t\delta\Omega^- = 0$. Theorem 4(iii) is replaced by

THEOREM 7. *Suppose $\omega \in \mathfrak{P}_2$, H^+ and H^- are its respective projections on $\mathfrak{H}^+ = \mathfrak{H} \cap \mathfrak{P}_2^+$ and $\mathfrak{H}^- = \mathfrak{H} \cap \mathfrak{P}_2^-$, Ω^+ and Ω^- are the respective plus and minus potentials of $\omega - H^+$ and $\omega - H^-$, and $\alpha^\pm = d\Omega^\pm$ and $\beta^\pm = \delta\Omega^\pm$. Then*

(i) *α^+ is the plus potential of $d\omega$, β^- is the minus potential of $\delta\omega$, $d\omega \in \mathfrak{L}_2 \ominus \mathfrak{H}^+$, and $\delta\omega \in \mathfrak{L}_2 \ominus \mathfrak{H}^-$.*

(ii) *If $\omega \in \mathfrak{P}_2^+$, then $\delta\omega \in \mathfrak{L}_2 \ominus \mathfrak{H}^+$ and β^+ is its plus potential.*

(iii) *If $\omega \in \mathfrak{P}_2^-$, then $d\omega \in \mathfrak{L}_2 \ominus \mathfrak{H}^-$ and α^- is its minus potential.*

For, we have just seen that α^+ and $\beta^+ \in \mathfrak{P}_2^+$ and α^- and $\beta^- \in \mathfrak{P}_2^-$. Clearly

(20) $$\delta\alpha^\pm + d\beta^\pm = \omega - H^\pm, \qquad d\alpha^\pm = \delta\beta^\pm = 0.$$

It remains only to show that (20) implies $(9^\pm)$ which requires the vanishing of a certain boundary integral.

8. A Kodaira Decomposition

From Theorem 7, we obtain immediately

THEOREM 8. *Suppose η^+, η^-, and $\eta \in \mathfrak{P}_2^+$, $\mathfrak{P}_2^-$, and $\mathfrak{P}_2$, respectively. Then there are unique forms α_1^+ and α_2^+ in $\mathfrak{P}_2^+ \cap (\mathfrak{L}_2 \ominus \mathfrak{H}^+)$ and β_1^- and β_2^- in $\mathfrak{P}_2^- \cap (\mathfrak{L}_2 \ominus \mathfrak{H}^-)$ such that*

(21) $$\begin{aligned} &\delta\alpha_1^+ = \delta\eta^+, \quad d\alpha_1^+ = 0, \quad d\beta_1^- = d\eta^-, \quad \delta\beta_1^- = 0, \\ &d\alpha_2^+ = d\eta, \quad \delta\alpha_2^+ = 0, \quad \delta\beta_2^- = \delta\eta, \quad d\beta_2^- = 0. \end{aligned}$$

This is proved by letting $\Omega_1^\pm$ and $\Omega_2^\pm$ be the respective potentials of $\eta^\pm - H^\pm$ and $\eta - H^\pm$ and using Theorem 7 with

$$\alpha_1^+ = d(\delta\Omega_1^+), \quad \alpha_2^+ = \delta(d\Omega_2^+), \quad \beta_1^- = \delta(d\Omega_1^-), \quad \beta_2^- = d(\delta\Omega_2^-).$$

From this theorem, we easily obtain

THEOREM 9. *The sets $\mathfrak{H}^r$ of all harmonic r-fields, $\mathfrak{E}^r$ of all r-forms of the form $\delta\alpha$ for α in $\mathfrak{P}_2^{(r+1)+}$, and $\mathfrak{D}^r$ of all r-forms of the form $d\beta$ for β in $\mathfrak{P}_2^{(r-1)-}$ are closed linear manifolds in $\mathfrak{L}_2^r$ and (15) holds. In the decomposition (14), the forms α and β are uniquely determined by the conditions that $\alpha \in \mathfrak{P}_2^+$, $\beta \in \mathfrak{P}_2^-$, and $d\alpha = \delta\beta = 0$. The differentiability results of Theorem 6 hold right up to the boundary.*

The first two sentences follow immediately from Theorem 8 (for $\mathfrak{E}^r$ and $\mathfrak{D}^r$) and the boundedness and approximation theorems mentioned in Section 5 (for $\mathfrak{H}^r$). The differentiability results follow from those for the potentials as before.

From the theorems above, we easily show that Theorem 2 holds (even though $\mathfrak{H}^r$ is not finite dimensional) and obtain the following theorem:

THEOREM 10. *If $\mathfrak{M}$ is of class C_1^1 at least and $\omega \in \mathfrak{L}_2 \ominus \mathfrak{H}$, there is a unique form Ω in $\mathfrak{P}_2 \cap (\mathfrak{L}_2 \ominus \mathfrak{H})$ such that (9) holds for every ζ in $\mathfrak{P}_2$. In fact if Ω^+ and Ω^- are the respective potentials of ω, we have*

$$\begin{aligned} &\Omega = \Gamma^+ + \Delta^-, \qquad d\Omega = d\Omega^+, \qquad \delta\Omega = \delta\Omega^-, \\ &nd\Omega = t\delta\Omega = 0, \qquad \delta\Gamma^+ = d\Delta^- = 0, \end{aligned} \tag{22}$$

where Γ^+ and Δ^- are the respective projections of Ω^+ on $\mathfrak{E}$ and Ω^- on $\mathfrak{D}$. The differentiability properties of Ω are the same as those of $\Omega^\pm$ in terms of those of $\mathfrak{M}$ and ω.

The proof of the existence of Ω is just like that of Theorem 3. Since (9) now holds for all ζ in $\mathfrak{P}_2$ we easily find from Theorem 8 and the definitions of $\Omega^\pm$ that $d\Omega = d\Omega^+$, $\delta\Omega = \delta\Omega^-$. The formula (22) for Ω and the differentiability results then follow from Theorems 8 and 9.

9. Boundary Values

The boundary value theorems for harmonic forms recently obtained by Conner [2] follow easily from the preceding results. These are roughly that there exists a harmonic form K with

(i) nK and ndK given, or

(ii) tK and $t\delta K$ given, or

(iii) ndK and $t\delta K$ given.

Since

$$n\Omega^+ = nd\Omega^+ = t\Omega^- = t\delta\Omega^- = nd\Omega = t\delta\Omega = 0,$$

these results are reduced to showing the existence of *some* form ω with the given boundary values; the desired K may then be found by subtracting off the proper potential of $\Delta\omega$. The best differentiability results are obtained but are too complicated to state here. We also obtain the following result of

Duff and Spencer [4] concerning the Dirichlet problem (by minimizing the Dirichlet integral in $\mathfrak{P}_2 \cap (\mathfrak{L}_2 \ominus \mathfrak{H})$).

THEOREM 11. *If ω is any form in $\mathfrak{P}_2$, there is a harmonic form K with $nK = n\omega$ and $tK = t\omega$. The differentiability properties of K are roughly the same as those of ω (as in Theorem 6). Any two solutions differ by a harmonic field in $\mathfrak{H}^+ \cap \mathfrak{H}^-$. If $\mathfrak{M}$ and $\mathfrak{B}$ are analytic any such harmonic field vanishes identically.*

The last sentence follows from the Cauchy-Kowalewsky theorem. If the uniqueness of the solutions of elliptic equations with given Cauchy data were known in the non-analytic case, the uniqueness of the solution of the Dirichlet problem would be established in general.

For harmonic fields the following results extending those of Duff and Spencer [4] are obtained:

THEOREM 12. (a) *If ω is any closed r-form in $\mathfrak{P}_2$, there is a unique harmonic field H such that*

$$\omega = H + d\beta, \qquad tH = t\omega, \qquad \beta \text{ and } d\beta \in \mathfrak{P}_2^-, \qquad 0 \leqq r \leqq n-1.$$

(b) *The dual of* (a).

In both cases, the differentiability of H at $\mathfrak{B}$ are roughly the same as those of ω (as in Theorem 6).

This follows immediately from the decomposition.

THEOREM 13. *If η is any form in $\mathfrak{P}_2$, there is a form ω in $\mathfrak{P}_2$ such that $t\omega = t\eta$ and $d\omega \in \mathfrak{H}$; $d\omega$ is uniquely determined. If also, $\eta = \delta\chi$ for some χ in $\mathfrak{P}_2$, then there is a unique form ω of the form $\delta\xi$ with ξ in $\mathfrak{P}_2$ which satisfies the given conditions.*

In both cases, let H^- be the projection of $d\eta$ on $\mathfrak{H}^-$, let α be the minus potential of $d\eta - H^-$ and define $\omega = \eta - \delta\alpha$. This ω is easily seen to have the desired properties. In the first case if ω_1 also satisfies the condition, the difference $\omega_2 \in \mathfrak{P}_2^-$ so that $d\omega_2 \in \mathfrak{L}_2 \ominus \mathfrak{H}$ so $d\omega_2 = 0$. In the second case, we also have $\delta\omega_2 = 0$ so $\omega_2 \in \mathfrak{H}^-$. But from Theorem 7(i), $\omega_2 \in \mathfrak{L}_2 \ominus \mathfrak{H}^-$.

Bibliography

[1] Calkin, J. W., *Functions of several variables and absolute continuity, I*, Duke Math. J., Vol 6, 1940, pp. 170–185.

[2] Conner, P. E., *The Green's and Neumann's problems for differential forms on Riemannian manifolds*, Proc. Nat. Acad Sci. U S A., Vol 40, 1954, pp. 1151–1155.

[3] de Rham, G., and Kodaira, K., *Harmonic Integrals*, Mimeographed Notes, Institute for Advanced Study, 1950

[4] Duff, G. F. D., and Spencer, D C., *Harmonic tensors on Riemannian manifolds with boundary*, Ann. of Math., Vol. 56, 1952, pp. 128–156.

[5] Friedrichs, K. O., *Differential forms on Riemannian manifolds*, Comm. Pure Appl. Math., Vol. 8, 1955, pp. 551–590.

[6] Gaffney, M., *The harmonic operator for exterior differential forms*, Proc. Nat. Acad. Sci. U.S.A., Vol. 37, 1951, pp. 48–50

[7] Kodaira, K., *Harmonic fields in Riemannian manifolds*, Ann. of Math., Vol. 50, 1949, pp. 587–665.

[8] Morrey, C. B., Jr., *Functions of several variables and absolute continuity, II*, Duke Math. J., Vol. 6, 1940, pp. 186–215.

[9] Morrey, C. B., Jr., *Second order elliptic systems of differential equations*, Ann. of Math. Studies, No. 33, Princeton Univ. Press, 1954, pp. 101–159.

[10] Morrey, C. B., Jr., and Eells, J., Jr., *A variational method in the theory of harmonic integrals* Proc. Nat. Acad. Sci. U.S.A., Vol. 41, 1955, pp. 391–395

[11] Morrey, C. B., Jr., and Eells, J., Jr., *A variational method in the theory of harmonic integrals, I*, Ann. of Math., Vol. 63, No. 1, 1956, pp. 91–128.

[12] Morrey, C. B., Jr., *A variational method in the theory of harmonic integrals, II*, Amer J. Math., Vol. 78, No 1, 1956, pp. 137–170.

[13] Rellich, F., *Ein Satz uber mittlere Konvergenz*, Nachr. Akad. Wiss. Göttingen, Math-Phys. Kl., 1930, pp. 30–35.

[14] Shiffman, M., *Differentiability and analyticity of solutions of double integral variational problems*, Ann. of Math., Vol. 48, 1947, pp. 274–284.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 509–530 (1956)

Estimates and Existence of Solutions of Elliptic Equations*

LOUIS NIRENBERG

Introduction

This paper is a survey of some *a priori* estimates for solutions of linear elliptic partial differential equations, and their use in solving the Dirichlet problem. For simplicity we shall consider only the case of zero Dirichlet data. In §1 we describe estimates, under various conditions on the coefficients, for the classical situation—a single equation of second order; and in §2, we discuss the extensions of these results to equations of arbitrary order and to systems, made within the last few years.

Notation: We consider functions defined in a fixed bounded domain $\mathscr{D}$ of Euclidean n-space with boundary $\dot{\mathscr{D}}$ and closure $\bar{\mathscr{D}}$. All other domains will be subdomains of $\mathscr{D}$.

Differentiation $\partial/\partial x_i$ is denoted by D_i, and $D^j u$ is used to denote any derivative of u of order j. A function f is said to satisfy a Hölder condition with exponent α, $0 \leqq \alpha < 1$ and coefficient H if for every pair of points P, Q

$$|f(P)-f(Q)| \leqq H|P-Q|^\alpha$$

where $|P-Q|$ is the distance from P to Q. For j a non-negative integer, and $0 \leqq \alpha < 1$, we introduce in a domain $\mathscr{A}$, $C^{\mathscr{A}}_{j+\alpha}$, the class of functions with Hölder continuous (exponent α) derivatives in $\bar{\mathscr{A}}$ up to order j. For functions in $C^{\mathscr{A}}_{j+\alpha}$ we define a norm

$$|u|^{\mathscr{A}}_{j+\alpha} = \max_{\bar{\mathscr{A}}} |u| + \cdots + \max_{\bar{\mathscr{A}}} |D^j u| + \text{l.u.b.}_{P,Q \in \mathscr{A}} \frac{|D^j u(P)-D^j u(Q)|}{|P-Q|^\alpha}.$$

Here the max. and l.u.b. are also taken with respect to all derivatives up to order j. Under this norm $C^{\mathscr{A}}_{j+\alpha}$ is a Banach space.

A domain $\mathscr{D}$ is said to be of class $L_{j+\alpha}$, for $j > 0$ an integer, and $0 \leqq \alpha < 1$, if $\mathscr{D}$ is covered by a finite number of subdomains $\mathscr{A}_i$ such that each $\mathscr{A}_i$ admits a one-to-one mapping into n-space with non-vanishing Jacobian which takes $\dot{\mathscr{D}} \cap \bar{\mathscr{A}}_i$ into an $(n-1)$-dimensional plane, and which, together with its inverse, has Hölder continuous (exponent α) derivatives up to order j.

*Work sponsored by the Office of Ordnance Research, United States Army.

In addition to the pointwise norms $|u|_a^{\mathscr{A}}$ for $a > 0$ we introduce integral norms:

$$||u||_j^{\mathscr{A}} = \left\{ \sum_{i=0}^{j} \int_{\mathscr{A}} |D^i u|^2 \, dx \right\}^{\frac{1}{2}}.$$

Clearly these norms come from related scalar products. We introduce corresponding Hilbert spaces:

$H_j^{\mathscr{A}}$, the completion of functions in $C_j^{\mathscr{A}}$ under this norm,

$\mathring{H}_j^{\mathscr{A}}$, the completion, under this norm, of functions in $C_j^{\mathscr{A}}$ which vanish in a whole neighborhood of the boundary $\dot{\mathscr{A}}$.

We shall omit the superscript $\mathscr{A}$ in the spaces and norms whenever $\mathscr{A}$ is the given domain $\mathscr{D}$.

A function u in H_j is said to have derivatives up to order j in a generalized (L^2) sense (Friedrichs [14], Sobolev [34]). If $u \in \mathring{H}_j$ then its derivatives up to order $j-1$ are said to vanish in a generalized sense on the boundary.

In many applications it is useful to have relations between pointwise estimates for a function and integral estimates for its derivatives. The best known relations of this kind are due to Sobolev [34] (see also Hörmander [21]).[1] Others are due to Friedrichs and Morrey. We formulate here a useful result overlapping these and, since its proof is not easily available, sketch the proof in the appendix.

LEMMA 1. *Let $\mathscr{D}$ be a bounded domain of class L_2 in n-space. Let ϱ be a positive integer, and k and p constants with*

$$p \geqq 1, \quad 0 \leqq k < n, \quad \text{and} \quad \varrho > \frac{n-k}{p}, \quad \frac{n-k}{p} \neq \text{integer}.$$

Set $s = \varrho - [(n-k)/p] - 1$, and

$$\alpha = 1 + \left[\frac{n-k}{p}\right] - \frac{n-k}{p}.$$

Then

$$(1) \qquad (|u|_{s+\alpha})^p \leqq K\left[\text{l.u.b.}_{P \in \mathscr{D}} \int_{\mathscr{D}} |P-Q|^{-k} |D^\varrho u(Q)|^p \, dQ + \int_{\mathscr{D}} |u|^p \, dx\right]$$

where K is a constant depending only on the domain $\mathscr{D}$ and the constants n, ϱ, k, and p. Here the l.u.b. on the right is also taken with respect to all derivatives $D^\varrho u$.

The appendix also contains the extension of the lemma to functions having generalized derivatives.

An analogue of the lemma was derived by D. Greco [18].

[1] A number of new estimates are derived in this very interesting paper which is concerned mainly with general equations with constant coefficients. The paper also contains a complete characterization of the class of equations with constant coefficients whose solutions are all infinitely differentiable.

1. Equations of Second Order

We consider a single second order linear elliptic equation[2]

$$Lu \equiv -a_{ij}D_iD_ju+a_iD_iu+au = f \tag{1.1}$$

for $u(x_1, \cdots, x_n)$, and shall assume throughout that the coefficients satisfy

$$\begin{gathered} |a_{ij}|,\ |a_i|,\ |a| \leqq M, \\ a_{ij}\xi_i\xi_j \geqq m\xi_i^2 \end{gathered} \tag{1.2}$$

for all real $\xi_1, \cdots, \xi_n$ and for some positive constants m, M. The Dirichlet problem we consider is that of finding a solution of (1.1) which vanishes on the boundary $\dot{\mathcal{D}}$. Uniqueness is assured if

$$a \geqq 0; \tag{1.3}$$

otherwise the well known alternative theorem holds.

Under various assumptions regarding the smoothness of the coefficients of (1.1) there exists a variety of *a priori* estimates for solutions which lead to different existence proofs. We shall describe several of these estimates below together with some existence proofs in which they may be employed. For applications to nonlinear equations it is useful to have estimates for linear equations which make minimal requirements on the coefficients; this is illustrated in §1.4.

1.1. *Estimate from the maximum principle.* Under conditions (1.2) and (1.3) it follows from the well-known maximum principle (see [20, 25, 11]) that the solution of the Dirichlet problem (with zero boundary values) satisfies

$$|u| \leqq \text{const. max } |f|$$

where the constant depends only on m and M.

1.2. *Square integral estimates.* If the coefficients a_{ij} are once continuously differentiable then multiplication of (1.1) by $\bar{u}$ and integration by parts leads to the inequality

$$\begin{aligned} \int_{\mathcal{D}} \bar{u}Lu\,dx &= \int_{\mathcal{D}} (a_{ij}D_iu\overline{D_ju}+\bar{u}(D_ja_{ij}+a_i)D_iu+a|u|^2)\,dx \\ &\geqq \text{const. } ||u||_1^2-\text{const. } ||u||_0^2 \end{aligned}$$

for functions u vanishing in a whole neighborhood of the boundary. The constants are positive and depend only on m, M and on a bound for the derivatives of the a_{ij}.

By means of a variational argument, or with the aid of the projection

[2]The summation convention is used.

theorem in Hilbert space, the inequality leads easily (as for example in [11]) to the existence of a "generalized solution" in $\mathring{H}_1$ of the Dirichlet problem. This method, together with the problem of showing that the generalized solution is indeed a regular solution of the problem and has continuous second derivatives, is discussed in §2.2 for a more general situation.

1.3. *Schauder's estimates.* In contrast to the simple integral estimate of §1.2, which requires some differentiability of the coefficients, Schauder ([31, 32]) obtained pointwise estimates under the assumption that for some positive $\alpha < 1$ the function f and the coefficients in (1.1) belong to C_α. The estimates are of two kinds:

Interior estimates. Let u belong to $C_{2+\alpha}$ and be a solution of (1.1). Then for any compact subdomain $\mathscr{A}$ of $\mathscr{D}$

$$|u|^{\mathscr{A}}_{2+\alpha} \leqq \text{const. } (|f|_\alpha + |u|_0)$$

where the constant depends on the diameter of $\mathscr{D}$, the distance from $\mathscr{A}$ to $\dot{\mathscr{D}}$, the constant m and the maximum of the $|\ |_\alpha$ norms of the coefficients of (1.1).

Boundary estimates. Let u belong to $C_{2+\alpha}$, be a solution of (1.1), and vanish on the boundary of the domain $\mathscr{D}$ which is assumed to be of class $L_{2+\alpha}$. Then

$$|u|_{2+\alpha} \leqq \text{const. } (|f|_\alpha + |u|_0),$$

where the constant depends only on the domain $\mathscr{D}$, m and the maximum of the $|\ |_\alpha$ norms of the coefficients of (1.1).

The estimates clearly yield compactness of solutions of equations of the form (1.1). They also furnish easily the existence of solutions of the Dirichlet problem for equation (1.1) (with $a \geqq 0$) in a domain of class $L_{2+\alpha}$, provided f and the coefficients are merely in C_α. The solution may be found, for instance, in the following way: approximate the coefficients and the inhomogeneous term f of (1.1) by infinitely differentiable sequences with uniformly bounded $|\ |_\alpha$ norms and solve the resulting approximating equations by, say, the procedures of §1.2; then, assuming that the solutions obtained by this procedure have been shown to be sufficiently regular, use the compactness to find a subsequence which converges to a solution of (1.1).

Another procedure, used by Schauder, is the continuity method: For $0 \leqq t \leqq 1$ set $L_t = tL + (1-t)\Delta$. The boundary estimates, together with the estimate of (1.1), yield the uniform inequality

$$|u|_{2+\alpha} \leqq \text{const. } |L_t u|_\alpha$$

for functions u vanishing on $\dot{\mathscr{D}}$. The linear operator L_t considered as a mapping of $C_{2+\alpha}$ into C_α is clearly bounded and, by the preceding inequality, is one-to-one and has a bounded inverse. Let T be the set of t values for which

L_t maps $C_{2+\alpha}$ *onto* C_α. A standard iteration argument shows that T is an open set in the interval; from the compactness it follows easily that T is also closed. Since T contains $t = 0$, i.e., for every f in C_α there exists a solution in $C_{2+\alpha}$ of the Dirichlet problem for $\Delta u = f$ (see [22]), it follows that T is the whole interval. Thus the original problem, for $t = 1$, is solved.

With the aid of the interior estimates it is possible to solve the Dirichlet problem $Lu = f$ with prescribed continuous boundary values for u and in a much wider class of domains (see [11]).

1.4. *More refined estimates for* $n = 2$. In connection with nonlinear equations in two dimensions, *a priori* estimates for linear equations have been derived which depend only on m and M. A typical estimate is the following: There exist positive constants $\alpha < 1$, K such that the solution of the Dirichlet problem for (1.1) in the unit circle $\mathscr{D}$ satisfies

$$(1.4)\quad |u|_{1+\alpha} + \text{l.u.b.}_{P \in \mathscr{D}} \left\{\int |P-Q|^{-2\alpha} |D^2 u(Q)|^2 \, dQ\right\}^{\frac{1}{2}} \leqq K(\max |u| + \max |f|).$$

With the aid of (1.4) it is possible to solve equations (1.1) whose coefficients are merely bounded measurable (see, for earlier work, [33]). But the chief interest of such estimates is in the solution of nonlinear equations. To illustrate this we find a solution of the quasi-linear equation

$$-a_{ij}(x_1, x_2, u, D_1 u, D_2 u) D_i D_j u + a_i(x_1, x_2, u, D_1 u, D_2 u) D_i u = f(x_1, x_2, u, D_1 u, D_2 u)$$

in the unit circle which vanishes on the boundary. We shall make the (rather strong) hypotheses that, for all values of the five arguments $|f|$ is uniformly bounded, and the coefficients satisfy (1.2). In addition, we assume that f and the coefficients satisfy a Hölder condition in all arguments in every bounded set of the arguments.

We observe first, with the aid of the maximum principle in §1.1, that if $a \equiv 0$ in (1.1) then from (1.4) the inequality

$$(1.5)\qquad |u|_{1+\alpha} \leqq \text{const.} \max |f| \leqq C$$

follows for the solution of the Dirichlet problem, where the constant C depends only on m and M. Now let S denote the set of functions u in C_1 satisfying $|u|_{1+\sigma} \leqq C$; S is a closed convex compact subset of C_1. If in the coefficients of the quasi-linear equation and in f, we insert for u, $D_1 u$, $D_2 u$ the values of a function of S and its derivatives, we obtain a linear elliptic equation with Hölder continuous coefficients. By the Schauder existence theory of §1.3 this has a unique solution U vanishing on the boundary; we have thus defined a (nonlinear) transformation $Au = U$. We observe, by (1.5), that $|U|_{1+\alpha} \leqq C$, so that A maps the set S into itself. Furthermore, with the aid of the boundary estimates of §1.3 it is easily seen that the

mapping is continuous. By Schauder's fixed point theorem we conclude that A has a "fixed point"; i.e., there is a function u in S with $Au = u$. Clearly, u is a solution of the quasi-linear equation.

Estimate (1.4) is a special case of the estimates contained in [5—7] where different boundary value problems are treated for linear and nonlinear equations. This work is an extension of the basic work in this field of C. B. Morrey [26]. (See also [2], and [28] for its bibliography on nonlinear problems.) The proofs of the estimates involve quasi-conformal mappings and function-theoretic arguments, and are, therefore, restricted to two dimensions.

2. Equations of Higher Order and Systems

Up to the present the only approach which has led to a general existence theory for equations with variable coefficients of order greater than two is a Hilbert space approach based on L^2 estimates analogous to §1.2. For second order equations this is described in [11]. The extension to equations of higher order and to systems was carried out by L. Gårding [17], M. I. Vishik [37] and F. E. Browder [9] (references to the earlier work of these authors may be found in the papers quoted; see also the bibliography of [30]).

In treating systems Vishik [36] introduced the concept of strongly elliptic systems, for which the Dirichlet problem is well posed. That this is not the case in general is shown by an example, due to Bicadze [8], of a second order system for two functions $u(x, y)$, $v(x, y)$. Setting $x+iy = z$, and $w = u+iv$ we write this system in the complex form (as suggested by Paul Berg)

$$\frac{\partial^2 w}{\partial \bar{z}^2} \equiv \frac{1}{4}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)^2 w = 0,$$

for which the general solution is clearly

$$w = f(z)+\bar{z}g(z)$$

with $f(z)$ and $g(z)$ complex analytic. We observe that all the solutions of the form

$$w = f(z)(1-z\bar{z}) \qquad \text{in } |z| \leqq 1$$

with arbitrary analytic f, vanish on the boundary of the unit circle. Thus the problem of finding a solution of the equation with prescribed boundary values is not well posed.

We begin our discussion of more general equations by defining ellipticity of a system and strong ellipticity, and then describe the known generalizations of the results of §1 to these equations.

2.1. *Elliptic systems.* Consider a system of linear differential equations (complex coefficients) for functions

$$u_1(x), \cdots, u_N(x), \qquad x = (x_1, \cdots, x_n),$$
$$l_{ij}(x, D)u_j(x) = f_i(x), \qquad i = 1, \cdots, N, \tag{2.1}$$

where l_{ij} is a polynomial in D: $(D_1, \cdots, D_n)$ with coefficients depending on x. Here summation convention is employed.

Let $s_1, \cdots, s_N, t_1, \cdots, t_N$ be integers (some may be negative) such that the order of the differential operator l_{ij} does not exceed s_i+t_j, and denote by $l'_{ij}(x, D)$ the sum of terms in l_{ij} with order exactly s_i+t_j. The characteristic matrix of the system associated with the integers $s_1, \cdots, t_N$ is defined as $l'_{ij}(x, \xi)$ for scalars ξ: $(\xi_1, \cdots, \xi_n)$. Its determinant is a homogeneous form in ξ of degree $\sum (s_i+t_i)$.

ELLIPTICITY: *The system is called elliptic if there exist integers $s_1, \ldots, t_N$ for which the associated characteristic determinant is different from zero for all real $\xi \neq 0$.*

This rather general definition, which was suggested by the hyperbolic systems of Leray, has been employed by Douglis and Nirenberg [12].

STRONG ELLIPTICITY: *The system* (2.1) *is termed strongly elliptic if there exist integers $s_1, \cdots, t_N$, with $s_i - t_i$, $i = 1, \cdots, N$, such that the quadratic form*

$$l'_{ij}(x, \xi)v_i\bar{v}_j,$$

is different from zero for all real $\xi \neq 0$ and all complex $v_1, \cdots, v_N \not\equiv 0$.

This is a generalization of Vishik's strongly elliptic systems for which $s_1 = s_2 = \cdots = t_N$.

Dirichlet's problem for a strongly elliptic system: Find a solution $u_1, \cdots, u_N$ of (2.1) with u_i and its normal derivatives up to order s_i-1, $i = 1, \cdots, N$, prescribed on the boundary.

An example of a strongly elliptic system, with $n = 2$, is the following:

$$(D_1^2+D_2^2)u_1-D_2^3u_2 \ + \text{lower order terms} = f_1,$$
$$D_2^3u_1+(D_1^2+D_2^2)^2u_2 + \text{lower order terms} = f_2.$$

Here $s_1 = t_1 = 1$, $s_2 = t_2 = 2$. Dirichlet's problem is: find solutions u_1, u_2 with $u_1, u_2, \partial u_2/\partial \nu$ prescribed on the boundary, where $\partial/\partial\nu$ denotes differentiation along the normal.

An elliptic equation for one function is clearly strongly elliptic.

In attempting to generalize the estimates of §1 the following question arises: Is there an extension of the result of §1.1 to equations of order greater than two? No general answer is known. Consider for example a single elliptic equation of order $2s$:

$$Lu = f.$$

It would be interesting to know all operators L for which a generalized maximum principle of the form

$$|u|_{s-1} \leqq \text{const.}\ (|Lu|_0 + |u|^{\dot{\mathscr{D}}}_{s-1})$$

is possible, where the constant is independent of u. Here

$$|u|^{\dot{\mathscr{D}}}_j = \max_{\dot{\mathscr{D}}} |u| + \cdots + \max_{\mathscr{D}} |D^j u|.$$

Such an inequality has been established by C. Miranda [23] in the case $n = 2$ and $L = (D_1^2 + D_2^2)^2$, which is the biharmonic operator. His argument is ingenious but intricate; using the estimate and the Hahn-Banach theorem he solves the general Dirichlet problem for $(D_1^2 + D_2^2)^2 u = 0$.

2.2. *The L^2 existence theory.* We consider for simplicity a single elliptic equation with complex coefficients of order $\varrho = 2s$, which we write in the convenient form

$$Lu \equiv \sum_{\sigma,\tau=0}^{s} (-1)^\sigma D^\sigma a^{\sigma,\tau} D^\tau u = f. \tag{2.2}$$

By partial integration this may be written in integral form: for every function ζ in C_∞ which vanishes near the boundary

$$B[\zeta, u] \equiv \sum_{\sigma,\tau \leq s} \int D^\sigma \zeta \cdot \overline{a^{\sigma,\tau} D^\tau u}\, dx = \int \zeta f dx \equiv (\zeta, f).$$

Here $(\ ,\ )$ is the L^2 scalar product.

Our Dirichlet problem is: find a solution u with $u = Du = \cdots = D^{s-1}u = 0$ on the boundary; the existence proof is based on an inequality generalizing that of §1.2 which, for general equations with variable coefficients is due to Gårding [17]. Assume that the equation is uniformly elliptic, i.e., for some positive m and all real ξ: $(\xi_1, \cdots, \xi_n)$ the inequality

$$\mathscr{R}e\, \xi^s a^{s,s} \xi^s \geqq m \sum_j \xi_j^{2s}$$

holds throughout $\mathscr{D}$.

GÅRDING'S INEQUALITY. *For all functions u in C_∞ which vanish near the boundary the inequality*

$$\mathscr{R}e\, B[u, u] \equiv \mathscr{R}e\, (u, Lu) \geqq c_1 ||u||_s^2 - c_2 ||u||_0^2 \tag{2.3}$$

holds, where $c_1 > 0$, $c_2 \geqq 0$ are fixed constants depending on m, the diameter of $\mathscr{D}$, on a bound for the coefficients $a^{\sigma,\tau}$, and on the modulus of continuity of the leading coefficients $a^{s,s}$.

In the appropriate L^2 formulation of the Dirichlet problem we seek, for given f in L^2, a function u in $\mathring{H}_s$ which is a "weak" solution of $Lu = f$, i.e., which satisfies

$$(L^*\zeta, u) = (\zeta, f)$$

for all ζ in C_∞ vanishing near the boundary; here L^* is the formal adjoint of L. Since u is in H_s this equation may also be written as

$$B[\zeta, u] = (\zeta, f).$$

For simplicity we shall assume that the operator is symmetric, i.e. $B[u, u]$ is real, and that the constant c_2 in (2.3) vanishes — which can be ensured by considering $L+c_2$ in place of L. From the definition of $\mathring{H}_s$ it follows easily that we also have

$$B[u, u] \geqq c_1 ||u||_s^2 \tag{2.4}$$

for all u in $\mathring{H}_s$. Since, obviously, $B[u, u] \leqq$ const. $||u||_s^2$, we may employ $B[u, v]$ as an alternative scalar product in the Hilbert space $\mathring{H}_s$; clearly the norms $||u||_s$ and $B[u, u]^{\frac{1}{2}}$ are equivalent.

Consider now the generalized Dirichlet problem for $Lu = f$. Using (2.4) we see that the linear functional (ζ, f) defined for all ζ in $\mathring{H}_s$ satisfies

$$|(\zeta, f)| \leqq ||f||_0 \, ||\zeta||_0 \leqq ||f||_0 \, ||\zeta||_s \leqq c_1^{-\frac{1}{2}} ||f||_0 (B[\zeta, \zeta])^{\frac{1}{2}}$$

and is therefore a bounded linear functional with respect to the norm $(B[\zeta, \zeta])^{1/2}$. By the well known representation theorem there exists therefore a function u in $\mathring{H}_s$ such that

$$(\zeta, f) = B[\zeta, u];$$

u is the desired solution of Dirichlet's problem.

To complete this existence theory it remains to show, under suitable assumptions on the smoothness of the coefficients and the domain, that the generalized solution is $2s$ times continuously differentiable in $\mathcal{D}$ and satisfies the Dirichlet data in the classical sense. The differentiability of the solution has been proved by various methods (for bibliography see for example [30], where an essentially self-contained exposition of the theory is given, including differentiability of the solution and assumption of boundary data). Friedrichs [16] gave a proof entirely within the L^2 framework, and based only on Gårding's inequality. The proof in [30] is a modification of his; it first applies a difference quotient operation to the equation, and then uses Gårding's inequality to estimate the $||\ ||_s^{\mathcal{A}}$ norms in compact subdomains $\mathcal{A}$ of the difference quotients of the solution. The desired differentiability then follows easily. In case of a smooth boundary this procedure is also used in [30] to show that the solution is regular in the full domain and that the boundary values are assumed. By a local transformation of variables the boundary (locally) is made to lie in a plane; then the difference quotient operation is applied in directions parallel to the plane. This procedure yields bounds on the L^2 norms for derivatives parallel to the plane of arbitrarily high order. The normal derivatives are then estimated with the aid of Lemma 2 of the appendix (for the case $p = 2$). Browder has also recently shown the regular-

ity of the solution at the boundary, using an inequality of N. Aronszajn [1] in place of Lemma 2. Previously the regularity at the boundary had been established only for second order equations in two dimensions [11] and then by Morrey [27] for second order systems.

Browder's method yields the useful inequality for any function u belonging both to H_{2s} and $\mathring{H}_s$

$$||u||_{2s} \leqq \text{const.}\ (||Lu||_0+||u||_0)$$

where the constant depends only on the operator L and on the domain $\mathscr{D}$.*

S. Agmon in an unpublished work has employed another approach to boundary value problems of arbitrary order and to the problem of regularity at the boundary.

2.3. *Schauder-type estimates.* Recently a number of simplified proofs of Schauder's results of §1.3 have been published. One appears in the book by Miranda [25]. Morrey's extensive paper on second order systems [27] contains a derivation of the estimates for such systems. Douglis and Nirenberg [12] have given another derivation of Schauder's results and have extended the interior estimates to the general elliptic systems of §2.1. The result may be roughly formulated as follows. We may assume that $\max s_i = 0$ and that for some positive number m, and $\varrho = \sum (s_i+t_i)$,

$$\det |l'_{ij}(x, \xi)| \geqq m \sum_i \xi_i^{\varrho} \qquad \text{for all real } \xi\text{: } (\xi_1, \cdots, \xi_n).$$

Then, *for any compact subdomain* $\mathscr{A}$ *of* $\mathscr{D}$,

$$\sum_i |u_j|^{\mathscr{A}}_{t_j+\alpha} \leqq \text{const.} \sum_j (|l_{ij}u_j|_{-s_i}+|u_i|_0)$$

where the constant depends only on the diameter of $\mathscr{D}$, *the distance from* $\mathscr{A}$ *to* $\mathscr{D}$, n, m *and the maximum of the* $|\ |_{-s_i}$ *norms of the coefficients of* l_{ij}, $i = 1, \cdots, N$.

This estimate yields the compactness (in every compact subdomain) of solutions of equations of the form (2.1). In addition this estimate together with related estimates in [12] yields differentiability theorems for solutions of nonlinear elliptic equations and of regular variational problems in the calculus of variations.

Similar boundary estimates for strongly elliptic equations would lead to existence theorems for Dirichlet's problem under mild requirements on the coefficients; we can get the solutions, as in §1.3, by a compactness argument applied to solutions of approximating equations, which, in turn, can be obtained, for example, by the procedure of §2.2.

2.4. *More refined estimates.* For dimension $n > 2$ no estimates of the form (1.4), depending only on bounds of the coefficients of the equation, are known. Estimates of the form (1.4) have been proved, but with the constant

*Added in proof: A similar result has been announced by O. V. Guseva, Dokl. Akad. Nauk SSSR (N.S.), Vol. 102, 1955, pp. 1069–1072 (Russian).

K depending also on the modulus of continuity of the coefficients a_{ij} (see Nirenberg [29], and Miranda [24] where a similar result is established for second order equations in the integral form (2.2)). Analogous estimates were established by Morrey [27] for general second order elliptic systems.

With the aid of a result of A. P. Calderon and A. Zygmund [10] it is easy to derive similar estimates for solutions of the most general elliptic systems (2.1) in compact subdomains as will be shown in a forthcoming paper by the author (see also Vekua [35]). In their derivation certain L^p integral norms for derivatives of the solution are used in place of the singular L^2 integrals occurring in (1.4). We formulate a typical estimate for a single elliptic equation $Lu = f$ of order ϱ, with bounded coefficients, whose characteristic polynomial $l(x, \xi)$ satisfies

$$l(x, \xi) \geqq m \sum_i \xi_i^\varrho \tag{2.5}$$

for some positive constant m: *For every $p > 1$ and every compact subdomain $\mathscr{A}$ of $\mathscr{D}$ the inequality*

$$\sum_{j=0}^{\varrho} \int_{\mathscr{A}} |D^j u|^p \, dx \leqq K_1 \cdot \int_{\mathscr{D}} (|Lu|^p + |u|^p) dx \tag{2.6}$$

holds, where the constant K_1 depends only on the diameter of $\mathscr{D}$, the distance from $\mathscr{A}$ to $\dot{\mathscr{D}}$, n, ϱ, p, m, a bound on the coefficients, and also on the modulus of continuity of the leading coefficients.[3] For $n/p \neq$ integer we infer by Lemma 1 that

$$|u|^{\mathscr{A}}_{\varrho - n/p} \leqq K_2 \left\{ \int_{\mathscr{D}} (|Lu|^p + |u|^p) dx \right\}^{\frac{1}{p}}$$

where the constant has the same dependence as above.

In these estimates the boundedness of the lower order coefficients is not necessary. For instance, if $p > n$ then, as far as these coefficients are concerned, K_2 depends on their L^p norms.

These inequalities, in a more precise form, may be used to study the behaviour of solutions of elliptic equations near isolated points. To illustrate this we state

THEOREM 1. *Let u be in $C_{\varrho-1}$ and have derivatives of order ϱ in L^p, $p > n$, in a neighborhood of the origin; assume that u satisfies $Lu = 0$ where L is an elliptic operator of order ϱ with bounded leading coefficients which are continuous at the origin, and with remaining coefficients in L^p. Let L_0 represent the elliptic operator with constant leading coefficients equal to those of L at the origin, and with remaining coefficients zero. Then*

(a) *there exists a polynomial $P(x)$ of degree $\varrho - 1$ satisfying $L_0 P(x) = 0$ such that*

$$D^j(u(x) - P(x)) = o(|x|^{\varrho - j - n/p}), \qquad j < \varrho,$$

[3] This dependence may be weakened.

(b) *if* $u = O(|x|^k)$, $k > \varrho - n/p$, $k \neq$ *integer, then* $u = o(|x|^k)$,

(c) *if* $u = O(|x|^k)$, $k > \varrho - n/p$, $k =$ *integer, and if the leading coefficients satisfy a Dini condition at the origin then there exists a homogeneous polynomial* $P_k(x)$ *satisfying* $L_0 P_k = 0$ *such that*

$$D^j(u(x) - P_k(x)) = o(|x|^{k-j}), \qquad j < \varrho;$$

(d) *if the leading coefficients satisfy a Hölder condition at the origin, and if* u *vanishes at the origin but not of infinite order, then there exists a homogeneous polynomial* $P_N(x) \not\equiv 0$ *of degree* N *satisfying* $L_0 P_N = 0$ *such that*

$$D^j(u(x) - P_N(x)) = O(|x|^{k-j+\alpha}), \qquad j < \varrho.$$

Here $\alpha > 0$ *is a constant depending only on the equation.*

The theorem is proved with the aid of techniques developed by Bers [3] (in fact (d) was known to Bers). In a paper [4] in this issue Bers gives a survey of the theory of local behavior of solutions of elliptic equations; references may be found there. We further state

THEOREM 2. *Consider a function* u *defined in a punctured sphere* $0 < |x| \leqq 1$. *Let* L *be an elliptic operator of order* ϱ *in the sphere with bounded leading coefficients satisfying* (2.5) *and with remaining coefficients in* L^p, $p > n$. *Then there exists a constant* ε *depending only on* m, p, n *and* ϱ, *such that if in some neighborhood of the origin the leading coefficients differ by less than* $\varepsilon(m, p, n, \varrho)$ *from constant values*[4] *then, for any real* k,

$$(2.7) \qquad Lu = O(|x|^{k-\varrho}), \quad u(x) = \begin{Bmatrix} O(|x|^k) \\ \text{or} \\ o(|x|^k) \end{Bmatrix} \Rightarrow D^j u = \begin{Bmatrix} O(|x|^{k-j}) \\ \text{or} \\ o(|x|^{k-j}) \end{Bmatrix}, \qquad j < \varrho.$$

THEOREM 3. *Let* u *be defined outside a sphere and let* L *be an elliptic operator of order* ϱ, *satisfying* (2.5), *outside the sphere. Assume, for simplicity, that* L *contains only terms of order* ϱ *whose coefficients satisfy the conditions of the preceding theorem near infinity*[5] *(for some* $p > n$ *and the same* $\varepsilon(m, p, n, \varrho)$*). Then for any real* k, (2.7) *holds near infinity.*

The estimates also yield a Liouville-type theorem (see [12] for the corresponding statement for operators with constant coefficients): *Let* u *satisfy* $Lu = 0$ *in the whole space and assume that* $u = O(|x|^{\varrho-\varepsilon})$ *for some* $\varepsilon > 0$. *Assume that* L *contains only terms of order* ϱ, *satisfying* (2.5), *which differ from constant values by less than* $\varepsilon(m, p, n, \varrho)$ *for some* $p > n/\varepsilon$, *then* $u(x)$ *is a polynomial of degree* $\varrho - 1$.

For $\varrho = 2$ a related result was indicated in [29].

The estimates described here, together with analogous estimates for

[4] For instance if their are continuous at the origin.

[5] For instance if they are continuous at infinity.

equations in integral form (2.2), yield new differentiability theorems for linear equations and, with the aid of the differentiability theorems resulting from §2.3, furnish strong differentiability theorems for general nonlinear elliptic equations and for regular variational problems in the calculus of variations. We may assert, for instance: *A solution u in C_ϱ of a nonlinear elliptic differential equation of order ϱ is differentiable more than ϱ times provided the differential operator has sufficiently many derivatives with respect to all arguments.*

The estimates are, however, not strong enough to furnish existence theorems for nonlinear equations, except for special second order quasi-linear equations with uniformly bounded coefficients whose leading coefficients are "close to constants". See also a forthcoming paper by H. O. Cordes.

Appendix

1. *Proof of Lemma* 1. For the proof of Lemma 1 we shall need a few additional lemmas. We first state

LEMMA 2. *If $p \geqq 1$, $0 \leqq j < \varrho$, then for every $\varepsilon > 0$,*

$$\int |D^j u|^p \, dx \leqq \varepsilon \int |D^\varrho u|^p \, dx + \text{const.} \int |u|^p \, dx$$

in a domain of class L_2, where summation over all derivatives $D^j u$, $D^\varrho u$ is understood. The constant depends only on the domain and on n, p, ϱ and ε.

For $n = 1$ the lemma is due to I. Halperin and H. R. Pitt [19]. For $n > 1$ and $p = 2$ proofs were given by Ehrling [13], and Nirenberg [30]; the proof in [30], as mentioned there, extends easily to arbitrary $p \geqq 1$. (Sobolev [34] has established much stronger results; one can estimate, in terms of the right hand side, the integrals of higher powers of $|D^j u|$, for $j < \varrho$.)

Definitions: A domain $\mathcal{D}$ is said to have the *cone property* if there exists a finite right spherical cone C (i.e. the intersection of a half-infinite spherical cone with a sphere about the vertex) such that every point in $\overline{\mathcal{D}}$ is the vertex of a cone C_P lying in $\overline{\mathcal{D}}$ and congruent to C. The domain $\mathcal{D}$ is said to have the *strong cone property* if it has the cone property and if there exist positive constants d and λ, such that any two points P, Q in $\overline{\mathcal{D}}$, with $|P-Q| \leqq d$, are vertices of cones C_P, C_Q congruent to C with the property that the volume of the set $\{\mathcal{D} \cap C_P \cap C_Q \cap$ sphere about P of radius $|P-Q| \cap$ sphere about Q of radius $|P-Q|\}$ is not less than $\lambda |P-Q|^n$.

It is not difficult to see that a domain of class L_2 has the strong cone property. We shall denote by θC_P the cone obtained by contracting a cone C_P by θ, $0 < \theta < 1$, about the vertex. We set

$$I_P[u] = \int_{\mathcal{D}} |P-Q|^{-k} |D^\varrho u(Q)|^p \, dQ,$$

where summation over all derivatives of order ϱ is implied.

Lemma 1 follows immediately, with the aid of Lemma 2, from the following

LEMMA 3. *Under the assumptions of Lemma* 1, *but replacing "$\mathcal{D}$ is of class L_2"* by *"$\mathcal{D}$ has the strong cone property"*, we *have*

$$(|u|_{s+\alpha})^p \leqq \overline{K} \left[\text{l.u.b.}_{P \in \mathcal{D}} I_P[u] + \sum_{0 \leqq j < \varrho} \int |D^j u(Q)|^p \, dQ \right] = \overline{K} I$$

where $\overline{K}$ depends only on $\mathcal{D}$ and the constants n, ϱ, k, p.

The last equation serves to define I.

The proof of Lemma 3 is based on the following Lemmas 4 and 5.

LEMMA 4. *Let C_P be a finite right spherical cone with vertex at P. Let $\varkappa$ and p be non-negative with*

$$p \geqq 1, \qquad p+\varkappa < n.$$

Set $\beta = p+\varkappa$. Then for every function u with continuous first derivatives in C_P we have

$$\int_{\frac{1}{2}C_P} |P-Q|^{-\beta} |u(Q)|^p \, dQ \leqq \text{const.} \left[\int_{C_P} (|P-Q|^{-\varkappa} |Du(Q)|^p + |u(Q)|^p) dQ \right]$$

where on the right hand side the sum over all first derivatives Du is taken. The constant depends only on C_P, n, $\varkappa$ and p.

Proof: We may take P to be the origin, and denote $|Q|$ by r. Let $\zeta(Q)$ be a continuously differentiable function depending on r alone, with $\zeta \geqq 0$, $\zeta \equiv 1$ in $\frac{1}{2}C_P$ and $\zeta \equiv 0$ outside of $\frac{3}{4}C_P$. Then

$$\int_{\frac{1}{2}C_P} |Q|^{-\varkappa-p} |u(Q)|^p \, dQ \leqq \int_{C_P} \zeta r^{-\varkappa-p} |u|^p r^{n-1} \, dr \, d\omega \tag{2}$$

where $d\omega$ is the element of solid angle. By partial integration we find

$$\int_{C_P} \zeta r^{-\varkappa-p} |u|^p r^{n-1} \, dr \, d\omega \leqq \frac{1}{n-\varkappa-p} \int_{C_P} r^{n-\varkappa-p} (|\zeta_r| |u|^p + p\zeta |u|^{p-1} |u_r|) dr \, d\omega$$

$$\leqq \text{const.} \int_{C_P} |u|^p \, dQ + \frac{p}{n-\varkappa-p} \int_{C_P} \zeta r^{-\varkappa} |u_r| |u|^{p-1} r^{1-p} \, dQ$$

since $\zeta_r \equiv 0$ in $\frac{1}{2}C_P$.

In case $p = 1$ the desired inequality follows. For $p > 1$ we find, with the aid of the inequality

$$ab \leqq \varepsilon \frac{a^p}{p} + \varepsilon^{-q/p} \frac{p-1}{p} b^{p/(p-1)}$$

which holds for arbitrary positive a, b, ε, that (by choosing ε large)

$$\frac{p}{n-\varkappa-r}\int_{C_P}\zeta r^{-\varkappa}|u_r|\left|\frac{u}{r}\right|^{p-1}dQ \leqq \tfrac{1}{2}\int_{C_P}\zeta r^{-\varkappa-p}|u|^p\,dQ+\text{const.}\int_{C_P}\zeta r^{-\varkappa}|u_r|^p\,dQ.$$

Inserting this into the previous estimate we find

$$\int_{C_P}\zeta r^{-\varkappa-p}|u|^p\,dQ \leqq \text{const.}\int_{C_P}(|u|^p+r^{-\varkappa}|u_r|^p)\,dQ$$

which, by (2), proves the lemma.

In addition, we have the following well known

LEMMA 5. *Let C_P, C_Q be two finite right spherical cones, with vertices at P, Q, congruent to a given cone C. Denote by G the set of points R satisfying*

$$R \in C_P \cap C_Q, \qquad |R-P| < |P-Q|, \qquad |R-Q| < |P-Q|,$$

and assume that the volume of G is not less than $\lambda|P-Q|^n$ for some positive λ. Let β, p be constants satisfying

$$p \geqq 1, \qquad \beta < n$$

and

$$\alpha = 1-\frac{n-\beta}{p} > 0.$$

Then

$$\frac{|u(P)-u(Q)|}{|P-Q|^\alpha} \leqq \text{const.}\left[\left\{\int_{C_P}|P-R|^{-\beta}|Du(R)|^p\,dR\right\}^{\frac{1}{p}} + \left\{\int_{C_Q}|Q-R|^{-\beta}|Du(R)|^p\,dR\right\}^{\frac{1}{p}}\right]$$

where the constant depends on C, n, β, p and λ.

Proof: $|u(P)-u(Q)| \leqq |u(P)-u(R)|+|u(R)-u(Q)|$. Integration with respect to R over G yields the inequality

$$(3)\quad \lambda|P-Q|^n|u(P)-u(Q)| \leqq \int_{G_P}|u(P)-u(R)|\,dR+\int_{G_P}|u(Q)-u(R)|\,dR,$$

where G_P denotes the set of points R satisfying $R \in C_P$, $|R-P| < |P-Q|$ (similarly for G_Q). If we set $|R-P| = r$ and $d\omega$ for the element of solid angle about P we find

$$\int_{G_P}|u(P)-u(R)|\,dR \leqq \int_{G_P} r^{n-1}\,dr\,d\omega\left(\int_0^r|u_r|\,dr\right) \leqq \text{const.}\,|P-Q|^n\int_{G_P}|u_r|\,dr\,d\omega$$
$$= \text{const.}\,|P-Q|^n\int_{G_P}|u_r|\,r^{\frac{n-1-\beta}{p}}\cdot r^{\frac{1}{p}-1+\alpha}\,dr\,d\omega = I_1.$$

In case $p = 1$ we have therefore

$$\int_{G_P}|u(P)-u(R)|\,dR \leqq I_1 \leqq \text{const.}\,|P-Q|^{n+\alpha}\int_{G_P} r^{-\beta}|u_r|\,dR.$$

Since a similar estimate holds for $\int_{G_Q}|u(Q)-u(R)|\,dR$ we obtain the desired result from (3).

Suppose now $p>1$. Then, by Hölder's inequality,

$$\begin{aligned}\int_{G_P}|u(P)-u(R)|\,dR &\leqq I_1\\ &\leqq \text{const.}|P-Q|^n\left\{\int_{G_P}|u_r|^p r^{-\beta}dR\right\}^{\frac{1}{p}}\left\{\int_{G_P} r^{\frac{p\alpha}{p-1}-1}\,dr\,d\omega\right\}^{\frac{p-1}{p}}\\ &= \text{const.}|P-Q|^{n+\alpha}\left\{\int_{G_P}|u_r|^p r^{-\beta}dR\right\}^{\frac{1}{p}}.\end{aligned}$$

A similar inequality holds for $\int_{G_Q}|u(Q)-u(R)|dR$, and from (3) the desired result follows.

Proof of Lemma 3: Let C be the finite cone related to the strong cone property of $\mathscr{D}$. For arbitrary P in $\mathscr{D}$ let C_P be a cone in $\overline{\mathscr{D}}$ congruent to C. With α as defined in Lemma 1, set $\beta=n+p(\alpha-1)$.

By repeated application of Lemma 4 (with each application the exponent of $|P-Q|$ is increased by p) we find, for $\theta=2^{1+s-\varrho}$, that

$$\begin{aligned}&\int_{\theta C_P}|P-R|^{-\beta}|D^{s+1}u(R)|^p\,dR\\ &\qquad\leqq \text{const.}\left[\int_{C_P}|P-R|^{-k}|D^\varrho u(R)|^p\,dR+\sum_{s<j\leqq\varrho}\int_{\mathscr{D}}|D^j u(R)|^p\,dR\right]\end{aligned}$$

where the constant depends only on C, n, ϱ, k and p. Since such an inequality holds for every point P it follows from Lemma 5 that for any pair of points P, Q sufficiently close, and any derivative D^s of order s,

$$(4)\qquad \frac{|D^s u(P)-D^s u(Q)|}{|P-Q|^\alpha}\leqq K_1\left[I_P[u]+I_Q[u]+\sum_{s<j\leqq\varrho}\int_{\mathscr{D}}|D^j u(R)|^p\,dR\right]^{\frac{1}{p}}.$$

Here K_1 depends only on $\mathscr{D}$, and on the constants n, ϱ, k and p. Since $\mathscr{D}$ is connected the estimate

$$|D^s u|_\alpha\leqq \text{const.}\,I^{\frac{1}{p}},$$

and hence Lemma 3, follows easily.

2. *Additional Remarks.* In practice it may occur that one has a bound for integrals $I_P[u]$ (see (1)) only for points P lying in a compact subdomain $\mathscr{A}$ of $\mathscr{D}$. In this case the analogue of Lemma 1, in fact, a simpler form of it, holds in $\mathscr{A}$:

LEMMA 1'. *Let $\mathscr{D}$ be a bounded domain in n space, $\mathscr{A}$ a compact subdomain, and let k, p, ϱ, α be as in Lemma 1. Then*

$$(|u|^{\mathscr{A}}_{s+\alpha})^p\leqq \overline{K}\left[\text{l.u.b.}_{P\in\mathscr{A}} I_P[u]+\int_{\mathscr{D}}|u|^p\,dx\right]$$

where $\bar{K}$ depends only on the domains $\mathscr{A}$, $\mathscr{D}$ and on n, ϱ, k, p.

The derivation is similar to that of Lemma 1. We observe first that by applying Lemma 2 to spheres covering $\mathscr{A}$ and lying in $\mathscr{D}$ we may derive the analogue of Lemma 2:

$$\int_{\mathscr{A}} |D^j u|^p \, dx \leqq \varepsilon \int_{\mathscr{D}} |D^\varrho u|^p \, dx + \text{const.} \int_{\mathscr{D}} |u|^p \, dx.$$

Furthermore, following the proof of Lemma 3 we may derive the analogue of (4): for P, Q in $\mathscr{A}$, and sufficiently close together,

$$\frac{|D^s u(P) - D^s u(Q)|}{|P-Q|^\alpha} \leqq K_1 \left[I_P[u] + I_Q[u] + \int_{\mathscr{D}} |u|^p \, dx \right]^{\frac{1}{p}}. \tag{5}$$

In the derivation of (5), via analogues of Lemmas 4 and 5, we may use spheres in place of cones.

In conclusion we show that Lemma 1′ also holds for functions having generalized derivatives in L^p in the sense of Friedrichs and Sobolev ([14, 34]).

Definition: A function u is said to have generalized derivatives up to order ϱ in L^p in a domain $\mathscr{D}$ if there exists a sequence $\{u_n\}$ of functions in C_ϱ converging to u in L^p whose derivatives $\{D^j u_n\}$, $j \leqq \varrho$, also converge in L^p. The limits of $D^j u_n$ are functions in L^p which are called the generalized derivatives $D^j u$ of u. It is easily seen that they are defined uniquely, independently of the sequence $\{u_n\}$.

THEOREM. *Let $\mathscr{D}$ be a bounded domain, $\mathscr{A}$ a compact subdomain, and let k, p, ϱ, α, s be as in Lemma 1. Let u have generalized derivatives up to order ϱ in L^p in $\mathscr{D}$ and assume that for each point P in $\mathscr{A}$ the integrals*

$$I_P[u] = \int_{\mathscr{D}} |P-R|^{-k} |D^\varrho u(R)|^p \, dR$$

exist and are uniformly bounded. Then u is almost everywhere equal to a function $\tilde{u}$ in $C^{\mathscr{A}}_{s+\alpha}$ with

$$(|\tilde{u}|^{\mathscr{A}}_{s+\alpha})^p \leqq \bar{\bar{K}} \left[\text{l.u.b.}_{P \in \mathscr{A}} I_P[u] + \int_{\mathscr{D}} |u|^p \, dx \right] \equiv \bar{\bar{K}} \bar{I} \tag{6}$$

where $\bar{\bar{K}}$ depends only on the domains $\mathscr{A}$, $\mathscr{D}$, and on n, ϱ, k, p.

Let $\mathscr{B}$ be a subdomain of $\mathscr{D}$ with $\bar{\mathscr{A}} \subset \mathscr{B} \subset \bar{\mathscr{B}} \subset \mathscr{D}$, and set

$$[u]^{\mathscr{B}} = \text{l.u.b.}_{P \in \mathscr{A}} \int_{\mathscr{B}} |P-R|^{-k} |D^\varrho u(R)|^p \, dR.$$

In order to prove the theorem it suffices to show that there exists a sequence $\{u_n\}$ of functions in $C^{\mathscr{B}}_\varrho$ which approximate u in the L^p norm in $\mathscr{B}$, whose derivatives up to order ϱ approximate in L^p those of u in $\mathscr{B}$, and for which the norms $[u_n]^{\mathscr{B}}$ are uniformly bounded by a fixed constant times the norms

$[u]^{\mathscr{D}}$. For, from Lemma 1′ applied to the sequence $\{u_n\}$ in the domain $\mathscr{B}$, it follows that a subsequence converges, in the norm $|\ |_s^{\mathscr{A}}$, to a function $\tilde{u}$ in $C_{s+\alpha}^{\mathscr{A}}$ which satisfies (5) for some $\overline{\overline{K}}$; clearly $u = \tilde{u}$ almost everywhere.

We describe briefly how such approximating functions $\{u_n\}$ may be constructed with the aid of Friedrichs' mollifiers [15]. Let $j(x) \geqq 0$ be a C_∞ function with support in $|x| < 1$ satisfying

$$\int j(x)\,dx = 1.$$

For ε sufficiently small the smoothing operator, or mollifier, J_ε is defined for functions in $\mathscr{D}$, as a convolution

$$J_\varepsilon u = \int j_\varepsilon(x-y)u(y)dy, \qquad j_\varepsilon(x-y) = \varepsilon^{-n} j\left(\frac{x-y}{\varepsilon}\right),$$

where $J_\varepsilon u$ is defined for x in $\mathscr{B}$. The operator J_ε clearly commutes with differentiation and has in addition the following easily established properties [15] : (a) $J_\varepsilon u \to u$ in the L^p norm in $\mathscr{B}$, (b) the L^p norm of $J_\varepsilon u$ is not greater than that of u in $\mathscr{D}$.

We claim that $J_{1/n}u$ is a suitable approximating sequence. To prove this we must show that there exist positive constants ε_0, C depending only on the function $j(x)$ such that for $\varepsilon < \varepsilon_0$ we have

$$[J_\varepsilon u]^{\mathscr{B}} \leqq C[u]^{\mathscr{D}}.$$

We show, in fact, that for suitable ε_0, C, and arbitrary P in $\mathscr{A}$, we have

$$(7)\quad \int_{\mathscr{B}} |P-R|^{-k}|\,(J_\varepsilon f)(R)|^p\,dR \leqq C\int_{\mathscr{D}} |P-R|^{-k}|f(R)|^p\,dR, \qquad \varepsilon < \varepsilon_0,$$

for any arbitrary measurable function f for which the right hand side is finite. Since, by Hölder's inequality,

$$|J_\varepsilon f|^p \leqq \int j_\varepsilon(x-y)\,|f|^p\,dy \cdot \left[\int j_\varepsilon(x-y)\,dy\right]^{p-1} = J_\varepsilon(|f|^p),$$

we may assume, in the proof of (7), that $p = 1$ and $f \geqq 0$. Thus, assuming as a final simplification that P is the origin, it suffices to prove the inequality

$$(7)'\qquad \int_{\mathscr{B}} |R|^{-k}\,(J_\varepsilon f)\,(R)\,dR \leqq C\int_{\mathscr{D}} |R|^{-k} f(R)\,dR.$$

We may choose $\varepsilon_0 = \frac{1}{2}$ times the distance from $\mathscr{B}$ to $\mathscr{D}$. The integral on the left of (7)′ may be expressed as

$$(8)\qquad \begin{aligned}&\int_{\mathscr{B}} dx \int_{\mathscr{D}} \left[j_\varepsilon(x-y)\,|x|^{-k} f(y)\right] dy \\ &\qquad = \int_{\substack{|y|<2\varepsilon \\ y\in\mathscr{D}}} dy \int_{\mathscr{B}} [\quad]\,dx + \int_{\substack{|y|>2\varepsilon \\ y\in\mathscr{D}}} dy \int_{\mathscr{B}} [\quad]\,dx = I_1 + I_2.\end{aligned}$$

Since $j_\varepsilon(x-y) \equiv 0$ for $|x-y| > \varepsilon$ we may assume, in I_2, that $|y| \leqq 2|x|$ so

that

$$I_2 \leqq 2^k \int_{\substack{|y|>2\varepsilon \\ y\in\mathscr{D}}} dy \int_{\mathscr{B}} j_\varepsilon(x-y)|y|^{-k} f(y)\,dx \leqq 2^k \int_{\substack{|y|>2\varepsilon \\ y\in\mathscr{D}}} |y|^{-k} f(y)\,dy.$$

In estimating I_1 we observe that

$$\begin{aligned}\int j_\varepsilon(x-y)|x|^{-k}\,dx &\leqq \varepsilon^{-n}(\max j)\int_{|x-y|<\varepsilon}|x|^{-k}\,dx \\ &\leqq \varepsilon^{-n}(\max j)\int_{|x|<\varepsilon}|x|^{-k}\,dx \\ &\leqq \frac{\max j}{n-k}\,\varepsilon^{-k}\omega = C_1\varepsilon^{-k}\end{aligned}$$

where ω is the surface area of the unit sphere. Thus

$$I_1 \leqq C_1\varepsilon^{-k}\int_{|y|<2\varepsilon} f\,dy \leqq 2^k C_1 \int_{|y|<2\varepsilon}|y|^{-k} f\,dy.$$

Inserting these estimates for I_1 and I_2 into (8) we obtain (7)′ with $C = 2^k C_1$. This completes the proof of the theorem.

Remark: In the proof of (7) the constant C is greater than unity and therefore the constant $\overline{\overline{K}}$ in (5) is greater than $\overline{K}$ of Lemma 1′. For any *fixed* point P and *fixed* function f, however, we can, by choosing ε sufficiently small, achieve that C in (7) is as close to unity as desired. To see this we write the integral on the left of (7)′ as a sum

$$\int_{\substack{|R|>\eta \\ R\in\mathscr{B}}} + \int_{\substack{|R|<\eta \\ R\in\mathscr{B}}} \tag{9}$$

where η is so small that

$$\int_{|R|<\eta}|R|^{-k} f\,dR < \delta\int_{\mathscr{D}}|R|^{-k} f\,dR$$

for arbitrary given $\delta > 0$. By our preceding analysis we easily recognize that the second integral in (9) is not greater than

$$2^k C_1\,\delta\int_{\mathscr{D}}|R|^{-k} f\,dR.$$

We now choose $\varepsilon(\delta)$ so small that the first integral is not greater than $(1+\delta)\int|R|^{-k} f\,dR$. The desired result follows.

As a consequence of the Remark we may infer the following

Lemma. The theorem holds with $\overline{\overline{K}}$ equal to $\overline{K}$ of Lemma 1′.

Since the constant $\overline{K}$ depended on K_1 of (5), it suffices to show that $\tilde{u}$ satisfies (5) with the same constant K_1. Applying inequality (5) to $J_\varepsilon u$ and using the Remark at the points P and Q we find

$$\frac{|D^s J_\varepsilon u(P) - D^s J_\varepsilon u(Q)|}{|P-Q|^\alpha} \leqq K_1(1+\delta)\left[I_P[u]+I_Q[u]+\int_{\mathscr{D}}|u|^p\,dx\right]^{\frac{1}{p}}$$

for arbitrary $\delta > 0$, provided ε is sufficiently small. Letting $\delta \to 0$ and, correspondingly, $\varepsilon \to 0$, we find that $\tilde{u}$ satisfies (5).

A similar argument may be used to extend Lemma 1 to functions with generalized derivatives. A modified mollifier of the type employed by Friedrichs in [15], §4, is then to be used. We merely state the extension.

Theorem. *Under the conditions of Lemma* 1 *let u have generalized derivatives up to order* ϱ *in* L^p *in* $\mathscr{D}$ *and assume that for all* P *in* $\mathscr{D}$ *the integrals* $I_P[u]$ *exist and are uniformly bounded. Then u is equal to a function* $\tilde{u}$ *in* $C_{s+\alpha}$ *almost everywhere, and*

$$(|\tilde{u}|_{s+\alpha})^p \leqq K(\text{l.u.b.}_{P\in\mathscr{D}} I_P[u]+\int |u|^p\,dx).$$

Bibliography

[1] Aronszajn, N., *On coercive integro-differential quadratic forms*, Conference on Partial Differential Equations, University of Kansas, 1954, Technical Report No. 14, pp. 94–106.

[2] Bers, L., *Existence and uniqueness of a subsonic flow past a given profile*, Comm. Pure Appl. Math., Vol. 7, No. 3, 1954, pp. 441–504.

[3] Bers, L., *Local behaviour of solutions of general linear elliptic equations*, Comm. Pure Appl. Math., Vol. 8, 1955, pp. 473–496.

[4] Bers, L., *Survey of local properties of solutions of elliptic equations*. This issue.

[5] Bers, L., and Nirenberg, L., *Boundary value problems for nonlinear elliptic equations in two independent variables*, Proc. Int. Congress, Amsterdam, September 1954, pp. 84–85.

[6] Bers, L., and Nirenberg, L., *On a representation theorem for linear elliptic systems with discontinuous coefficients and its applications*, Convegno Internazionale sulle Equazioni Derivate e Parziali, 1954, pp. 111–140.

[7] Bers, L., and Nirenberg, L., *On linear and nonlinear elliptic boundary value problems in the plane*. In the same volume as [6], pp. 141–167.

[8] Bicadze, A. V., *On the uniqueness of the solutions of the Dirichlet problem for elliptic partial differential equations*, Uspekhi Mat. Nauk (N.S.) 3, 1948, pp. 211–212 (Russian).

[9] Browder, F. E., *Strongly elliptic systems of differential equations*, Contributions to the Theory of Partial Differential Equations, Ann. Math. Studies No. 33, Princeton, 1954, pp. 15–51.

[10] Calderon, A. P., and Zygmund, A., *On the existence of singular integrals*, Acta Math., Vol. 88, 1952, pp. 85–139.

[11] Courant, R., and Hilbert, D., *Methods of Mathematical physics*, Vol. 2, Interscience Publishers, New York-London (to appear).

[12] Douglis, A., and Nirenberg, L., *Interior estimates for elliptic systems of partial differential equations*, Comm. Pure Appl. Math., Vol. 8, 1955, pp. 503–538.

[13] Ehrling, G., *On a type of eigenvalue problems for certain elliptic differential operators*, Math. Scandinavica, Vol. 2, 1954, pp. 267–285.
[14] Friedrichs, K. O., *On differential operators in Hilbert spaces*, Amer. J. Math., Vol. 61, 1939, pp. 523–544.
[15] Friedrichs, K. O., *The identity of weak and strong extensions of differential operators*, Trans. A.M.S., Vol. 55, No. 1, 1944, pp. 132–151.
[16] Friedrichs, K. O., *On the differentiability of the solutions of linear elliptic differential equations*, Comm. Pure Appl. Math., Vol. 6, 1953, pp. 299–326.
[17] Gårding, L., *Dirichlet's problem for linear elliptic partial differential equations*, Math. Scandinavica, Vol. 1, 1953, pp. 55–72.
[18] Greco, D., *Criteri di compattezza per insiemi di funzioni in n variabili indipendenti*, Ricerche di Mat. Napoli, Vol. 1, 1952, pp. 124–144.
[19] Halperin, I., and Pitt, H. R., *Integral inequalities connected with differential operators*, Duke Math. J., Vol. 4, 1938, pp. 613–625.
[20] Hopf, E., *Elementare Bemerkungen über die Lösungen partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus*, Sitz Preuss Akad. Wiss., Vol 19, 1927, pp. 147–152.
[21] Hörmander, L., *On the theory of general partial differential operators*, Acta Math., Vol. 94, 1955, pp. 161–248.
[22] Kellogg, O. D., *On derivatives of harmonic functions on the boundary*, Trans A.M.S., Vol. 33, 1931, pp 486–510.
[23] Miranda, C., *Formule di maggiorazione e teorema di esistenza per le funzioni biarmoniche di due variabili*, Giorn. Mat. Battaglini, Vol 78, 1948-9, pp. 97–118.
[24] Miranda, C., *Sui sistemi di tipo ellittico di equazioni lineari a derivate parziali del primo ordine, in n variabili indipendenti*, Atti della Accad Naz. dei Lincei, Ser. 8 (1a), Vol. 3, 1952, pp. 85–121.
[25] Miranda, C., *Equazioni alle derivate parziali di tipo ellittico*, Berlin, Springer-Verlag, 1955.
[26] Morrey, C. B., *On the solutions of quasilinear elliptic partial differential equations*, Trans. A.M.S., Vol. 43, 1938, pp 126–166.
[27] Morrey, C. B., *Second order elliptic systems of differential equations*, Contributions to the Theory of Partial Differential Equations, Ann. Math. Studies No 33, Princeton, 1954, pp. 101–159.
[28] Nirenberg, L., *On nonlinear elliptic partial differential equations and Hölder continuity*, Comm. Pure Appl. Math., Vol. 6, 1953, pp. 103–156
[29] Nirenberg, L., *On a generalization of quasi-conformal mappings and its application to elliptic partial differential equations*, Contributions to the Theory of Partial Differential Equations, Ann. Math. Studies No. 33, Princeton, 1954, pp. 95–100.
[30] Nirenberg, L., *Remarks on strongly elliptic partial differential equations*, Comm. Pure Appl. Math., Vol. 8, 1955, pp. 648–674.
[31] Schauder, J., *Über lineare elliptische Differentialgleichungen zweiter Ordnung*, Math. Z., Vol. 38, 1934, pp. 257–282.
[32] Schauder, J., *Numerische Abschätzungen in elliptischen linearen Differentialgleichungen*, Studia Math., Vol. 5, 1934, pp. 34–42.
[33] Schauder, J., *Sur les équations linéaires du type elliptique à coefficients continus*, C. R. Acad. Sci. Paris, Vol. 199, 1934, pp. 1366–1368.
[34] Sobolev, S. L., *On a theorem of functional analysis*, Mat. Sbornik, N.S. 4, 1938, pp. 471–497.
[35] Vekua, N., *On a method of solving boundary value problems for partial differential equations*, Doklady Akad. Nauk. Vol. 101, 1955, pp. 593–596.
[36] Vishik, M. I., *On strongly elliptic systems of differential equations*, Mat. Sbornik, N S., Vol. 29, 1951, pp. 615–676.
[37] Vishik, M. I., *On general boundary problems for elliptic differential equations*, Trudy Moskov. Mat. Obsč., Vol. 1, 1952,.pp. 187–246 (Russian).

Received January 20, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 531–500 (1956)

New Isoperimetric Inequalities for Eigenvalues and Other Physical Quantities*

L. E. PAYNE

Institute for Fluid Dynamics and Applied Mathematics
University of Maryland

1. Introduction

As in the book of Pólya and Szegö [17], we shall in this paper understand the term "isoperimetric" inequality to mean any inequality connecting two or more physical quantities associated with the same region. The first such inequality to be discovered (the one from which the name was derived) was the classical inequality between the area of a region and its perimeter. A number of known and conjectured inequalities appear in [17]. In particular the authors give some which involve the eigenvalues of the classical membrane and plate problems. The first such inequality was the conjecture of Rayleigh [18] that of all membranes of given area the circle has the minimum principal frequency. This conjecture was proven independently by G. Faber [7] and E. Krahn [10]. Additional inequalities for eigenvalues have been given by Szegö [22], Weinstock [28] and Payne [11]. Furthermore, inequalities relating the eigenvalues of a given problem to those of the corresponding finite difference problem have been obtained by Pólya [15], Forsythe [8], Weinberger [25] and Hersch [9]. It should be noted that the result of Szegö [22] and those of Payne [11] are the only cases in which it is known that for a given region, the eigenvalues of one problem are not greater than eigenvalues with lower index of another problem.

There exist also in the literature many methods for obtaining arbitrarily close estimates for eigenvalues such as the Rayleigh-Ritz method [19], the methods of Weinstein [26], Aronszajn (see the bibliography of [1]) and others. We shall not be concerned with such methods in this paper. In fact, isoperimetric inequalities are in general not improvable. They are, however, simple and often yield information which is not easily obtainable otherwise.

*The research for this paper was supported in part by the United States Air Force under Contract No. AF 18(600)-573 — monitored by the Office of Scientific Research, Air Research and Development Command.

We shall in the second section of this paper derive new inequalities for the eigenvalues of certain classical plate and membrane problems. For such problems we suppose that the displacement satisfies a certain differential equation in a plane domain D with boundary C in the x, y-plane. In the last section we consider the problem of Stekloff [21] in an infinite three-dimensional region D^* which lies exterior to the closed surface C^*. In this case we obtain isoperimetric inequalities for the first and second eigenvalues. In order to establish some of the eigenvalue relationships obtained in this paper we shall be required to derive inequalities for certain other physical quantities.

The eigenvalue problems which we consider are the following:

$$
\begin{aligned}
&\textbf{A.}\ \Delta u+\lambda u = 0 \text{ in } D, \qquad u = 0 \text{ on } C,\\
&\textbf{B.}\ \Delta^2 w+\Lambda\Delta w = 0 \text{ in } D, \qquad w = \frac{\partial w}{\partial n} = 0 \text{ on } C,\\
&\textbf{C.}\ \Delta^2\varphi-\Omega\varphi = 0 \text{ in } D, \qquad \varphi = \frac{\partial \varphi}{\partial n} = 0 \text{ on } C,\\
&\textbf{D.}\ \Delta^2\psi-a\Delta\psi-\Gamma\psi = 0 \text{ in } D, \qquad (a = \text{const.} > 0),\ \psi = \frac{\partial \psi}{\partial n} = 0 \text{ on } C,\\
&\textbf{E.}\ \Delta^2\zeta+\mu\Delta\zeta+b\zeta = 0 \text{ in } D \qquad (b = \text{const.} > 0),\ \zeta = \frac{\partial \zeta}{\partial n} = 0 \text{ on } C,\\
&\textbf{F.}\ \Delta\chi = 0 \text{ in } D, \qquad \frac{\partial \chi}{\partial n} = k\chi \text{ on } C^*,\ \chi = O(R^{-1}) \text{ as } R \to \infty,
\end{aligned}
\tag{1}
$$

where Δ denotes the Laplace operator, $\vec{n}$ denotes the outward normal from D (or D^*) on C (or C^*), $R^2 = x^2+y^2+z^2$, and λ, Λ, Ω, Γ, μ and k are constants (eigenvalues) to be determined. In (1), **A** is the classical membrane problem, **B** and **E** occur in the treatment of the buckling problem for a clamped plate, **C** and **D** are well-known vibration problems for a clamped plate and **F** is the problem of Stekloff. In **D** the plate is subjected to all around lateral tension (see [27]) and in **E** the plate is elastically supported.

For each of the problems in (1) it is known that infinitely many eigenvalues exist. Following the usual convention we order them as follows:

$$\lambda_1 \leqq \lambda_2 \leqq \lambda_3 \leqq \cdots, \tag{2}$$

the other eigenvalues being similarly ordered. The corresponding eigenfunctions are designated as $u_1, u_2, u_3, \cdots$, etc.

2. Membrane and Plate Problems

The following inequalities will be established in this section:

(3)
$$\text{a)}\ P \geqq 2\pi v_{\max}^2,$$
$$\text{b)}\ \lambda_1^2 P > 8\pi,$$
$$\text{c)}\ \Gamma_n \leqq \Omega_n + a\sqrt{\Omega_n},$$
$$\text{d)}\ \Gamma_n \geqq \Omega_{n-m} + a\lambda_{m+1},$$
$$\text{e)}\ \sum_{i=1}^{n} \Gamma_i \geqq \sum_{i=1}^{n} \Omega_i + a\sum_{i=1}^{n} \lambda_i,$$
$$\text{f)}\ \mu_n \leqq \Lambda_n + \frac{b}{\lambda_1},$$
$$\text{g)}\ \mu_n \geqq \max\,[\Lambda_n, 2\sqrt{b}],$$

where P is the torsional rigidity of a beam with cross section D and v is the stress function ($\Delta v = -2$ in D, $v = 0$ on C, and $P = \mathscr{D}(v)$, where $\mathscr{D}$ denotes the Dirichlet integral).

In order to prove (3b) we find it necessary first to establish (3a). This inequality follows directly from results given in [17]. In the conformal mapping estimates for the torsional rigidity of a simply connected region D, Pólya and Szegö obtained the following relationship between P, $v_{\max}$ and the coefficients a_i of the mapping function which maps D onto the unit circle:

(4)
$$P \geqq 2\pi\left[\sum_{n=1}^{\infty} \frac{n}{n+1}|a_n|^2\right]^2, \qquad \text{[17, § 5.7 (11)]},$$
$$2v_{\max} = \sum_{n=1}^{\infty} |a_n|^2, \qquad \text{[17 § 5.11]}.$$

It follows then from these expressions that

(5)
$$P \geqq 2\pi\left[\tfrac{1}{2}\sum_{n=1}^{\infty}|a_n|^2\right]^2 = 2\pi v_{\max}^2$$

with equality holding only if $a_2 = a_3 = \cdots = 0$, that is, if C is a circle. This is actually a sharper inequality for $v_{\max}$ than the one given in [17, §5.11, (12)]. For convex C, (5) can be obtained rather easily from the results of Payne and Weinberger [12].

The inequality (3b) now follows from (5) with the observation that u_1 and v are positive in D and satisfy the integral relation

(6)
$$2\iint_D u_1\,dA = \lambda_1 \iint_D v u_1\,dA \leqq \lambda_1 v_{\max} \iint_D u_1\,dA.$$

This inequality (3b) is still far from the conjectured inequality of Pólya and Szegö [17]

(7)
$$\lambda_1^2 P \geqq \tfrac{1}{2}(2.4048)^4\pi$$

(we have obtained the value 2 instead of 2.4048). The discrepancy is not unexpected since we have replaced v by $v_{\max}$ in (6). However, the inequality is still of interest; for, as was pointed out in [17], nothing was previously known about the lower bound for the dimensionless quantity $\lambda_1^2 P$ for non-convex figures. In fact, all that was known for convex C was that this quantity is positive.

The third and fourth of inequalities (3) follow from the maximum-minimum definition of the eigenvalues Γ_i, [6]. If we let

$$\mathfrak{A}(f) = \iint_D (\Delta f)^2 dA + a\mathscr{D}(f), \tag{8}$$

then

$$\Gamma_n = \max_{g_i} \min_{f} \frac{\mathfrak{A}(f)}{\iint_D f^2 dA}, \qquad i = 1, 2, \cdots, n-1, \tag{9}$$

for any sufficiently smooth functions f and g_i (all independent) which vanish together with their first derivatives on C and satisfy the condition

$$\iint_D f g_i dA = 0, \qquad i = 1, 2, \cdots, n-1. \tag{10}$$

By Schwarz's inequality

$$\Gamma_n \leqq \max_{g_i} \min_{f} \left[\frac{\iint_D (\Delta f)^2 dA}{\iint_D f^2 dA} + a \sqrt{\frac{\iint_D (\Delta f)^2 dA}{\iint_D f^2 dA}} \right] \tag{11}$$

$$= \Omega_n + a\sqrt{\Omega_n}$$

which is inequality (3c). On the other hand, from (9) we have

$$\begin{aligned} \Gamma_n &\geqq \max_{g_i} \left[\min_{f} \frac{\iint_D (\Delta f)^2 dA}{\iint_D f^2 dA} + a \min_{f} \frac{\mathscr{D}(f)}{\iint_D f^2 dA} \right] \\ &\geqq \max_{g_i'} \left[\min_{f} \frac{\iint_D (\Delta f)^2 dA}{\iint_D f^2 dA} + a \min_{f} \frac{\mathscr{D}(f)}{\iint_D f^2 dA} \right], \end{aligned} \tag{12}$$

where in g_i' we admit m functions whose normal derivatives are not required to vanish on C; hence $\{g_i'\} \supset \{g_i\}$. We choose for the functions g_i'

$$g_i' = \begin{cases} u_i, & i = 1, 2, \cdots, m, \\ \varphi_i, & i = m+1, m+2, \cdots, n-1. \end{cases} \tag{13}$$

It follows then from the variational definitions of Ω_i and λ_i that

$$\Gamma_n \geqq \Omega_{n-m}+a\lambda_{m+1}, \qquad 0 \leqq m \leqq n-1. \tag{14}$$

The fifth of inequalities (3) can be derived from a minimum principle for $\sum_{i=1}^{n} \Gamma_i$ given in Courant-Hilbert [6], i.e.,

$$\sum_{i=1}^{n} \Gamma_i = \min_{q_i} \sum_{i=1}^{n} \mathfrak{A}(q_i), \tag{15}$$

where the functions q_i are sufficiently smooth, vanish with their normal derivatives on C and satisfy the condition

$$\iint_D q_i q_j \, dA = \begin{cases} 0 \text{ for } i \neq j \\ 1 \text{ for } i = j. \end{cases} \tag{16}$$

We now make an interesting observation, namely

$$\begin{aligned} \sum_{i=1}^{n} \Gamma_i &\geqq \min_{q_i} \sum_{i=1}^{n} \iint_D (\Delta q_i)^2 \, dA + \min_{q_i} \sum_{i=1}^{n} \mathfrak{D}(q_i) \\ &\geqq \min_{q_i} \sum_{i=1}^{n} \iint_D (\Delta q_i)^2 \, dA + \min_{q_i'} \sum_{i=1}^{n} \mathfrak{D}(q_i') \\ &= \sum_{i=1}^{n} \Omega_i + a \sum_{i=1}^{n} \lambda_i. \end{aligned} \tag{17}$$

In (17) the functions q_i' satisfy the same conditions as the q_i except that their normal derivatives are not required to vanish on C. A similar application of the minimum principle (15) was given in a thesis by Boyce [5], which to my knowledge was the first use ever made of this principle. It is obvious that inequalities of type (17) are derivable from quite general positive definite operators $\mathfrak{A}(q)$.[1] In particular similar inequalities give considerable improvement in the auxiliary problem estimates of Aronszajn and Donoghue [2] for the higher eigenvalues of the clamped plate problems which they considered.

To establish (3*f*) we use the following variational definition of μ_n:

$$\mu_n = \min_f \frac{\iint_D (\Delta f)^2 \, dA + b \iint_D f^2 \, dA}{\mathfrak{D}(f)}, \tag{18}$$

where f is any sufficiently smooth function which vanishes with its normal derivative on C and satisfies

$$\mathfrak{D}(f, \zeta_i) = 0, \qquad i = 1, 2, \cdots, n-1. \tag{19}$$

[1] Similar inequalities for the eigenvalues of Hermitian matrices were given by Ky Fan, National Bureau of Standards Applied Mathematics Series No. 39, Washington D.C., 1954, pp. 131–139.

Thus if we choose for f a linear combination of the first n eigenfunctions w_i, the constants so chosen that (19) is satisfied, then we obtain

$$\mu_n \leqq \Lambda_n + \frac{b}{\lambda_1}, \tag{20}$$

the inequality (3f).

On the other hand, it is clear from (18) that

$$\mu_n \geqq \Lambda_n . \tag{21}$$

Likewise,

$$\begin{aligned} \mu_n &= \frac{\iint_D (\Delta\zeta_n)^2 dA + b\iint_D \zeta_n^2 dA}{\mathscr{D}(\zeta_n)} \\ &\geqq \frac{2\sqrt{b\iint_D (\Delta\zeta_n)^2 dA \iint_D \zeta_n^2 dA}}{\mathscr{D}(\zeta_n)} \\ &\geqq 2\sqrt{b}, \end{aligned} \tag{22}$$

by Schwarz's inequality. Hence

$$\mu_n \geqq \max\,[\Lambda_n , 2\sqrt{b}), \tag{23}$$

which is inequality (3g). From (11), (14), (20) and (21) it is apparent that for large n the ratios Γ_n/Ω_n and μ_n/Λ_n approach 1.

3. The Problem of Stekloff

We turn now to the eigenvalue problem (1**F**). No isoperimetric inequalities for the eigenvalues of this problem appear in the literature. R. Weinstock [28] has treated in a recent paper the corresponding two-dimensional interior Stekloff problem for simply-connected D and analytic C and has proved that in this case the first non-trivial eigenvalue is not greater than $2\pi/L$, where L is the perimeter of C. His proof involves a mapping of D onto the unit circle and hence can not be employed in our problem.

In this section we shall establish inequalities for k_1 and k_2. It is necessary also to give estimates for certain other physical quantities. We show in particular that

$$\begin{aligned} &\text{a)}\ k_1 \leqq \frac{4\pi\mathscr{C}}{S}, \\ &\text{b)}\ k_2 \leqq \max\left[\frac{4\pi\mathscr{C}}{S}, \frac{S}{3W_{\text{ave}}}\right], \\ &\text{c)}\ W_{\text{ave}} \geqq \tfrac{1}{2}V \quad (C^* \text{ axially symmetric, not ring-shaped}), \\ &\text{d)}\ \mathscr{P}_{\text{ave}} \geqq 2V \quad (C^* \text{ axially symmetric}), \\ &\text{e)}\ k_2 \leqq \frac{2S}{3V} \quad (C^* \text{ convex, axially symmetric}), \end{aligned} \tag{24}$$

where $\mathscr{C}$ is the electrostatic capacity of a condenser C with respect to an infinitely large sphere (see [17]), S is the area of C, V its volume, W_{ave} and $\mathscr{P}_{ave}$ are the average virtual mass and average polarization of C (see [20]) and $h = R\partial R/\partial n$.

If C is a sphere the solution to the Stekloff problem is known. The eigenfunctions are the spherical harmonics and the eigenvalues are

$$k_1 = \frac{1}{\alpha}, \qquad k_{3n-1} = k_{3n} = k_{3n+1} = \frac{n+1}{\alpha}, \quad n \geqq 1, \tag{25}$$

where α is the radius of C.

The variational definition of the eigenvalues k_i is as follows (see [6]):

$$k_i = \min_{U \in N} \frac{\mathscr{D}(U)}{\iint_{C^*} U^2 dS} \tag{26}$$

under the condition

$$\iint_{C^*} U\chi_i dS = 0, \qquad i = 1, 2, \cdots, n-1. \tag{27}$$

In (26) N is the class of functions continuous in D^*+C^* and $O(1/R)$ at infinity. Inequality (24a) follows directly from (26) if we choose U to be the electrostatic potential h_0 of C^*.

In order to prove (24b) we make use of the flow potential Φ_j defining the irrotational motion of C^* at unit velocity in the x_j-direction in an incompressible (unit density) inviscid fluid. It is well known that Φ_j is a harmonic function, $O(1/R^2)$ at infinity, which satisfies on C^* the condition

$$\frac{\partial \Phi_j}{\partial n} = -\frac{\partial x_j}{\partial n}, \tag{28}$$

that is,

$$\begin{aligned} \frac{\partial \Phi_1}{\partial n} &= -\frac{\partial x}{\partial n} \text{ on } C^*, \\ \frac{\partial \Phi_2}{\partial n} &= -\frac{\partial y}{\partial n} \text{ on } C^*, \\ \frac{\partial \Phi_3}{\partial n} &= -\frac{\partial z}{\partial n} \text{ on } C^*. \end{aligned} \tag{29}$$

We note first that $\mathscr{D}(h_0, \Phi_j) = 0$, $j = 1, 2, 3$, and that with the proper choice of the origin also $\iint_{C^*} h_0 \Phi_j dS = 0$, $j = 1, 2, 3$. It follows then from (26) that

$$k_2 \leqq \max\left[\frac{4\pi\mathscr{C}}{S}, \frac{\mathscr{D}(\Phi_j)}{\iint_{C^*}\Phi_j^2\,dS}\right], \qquad j = 1, 2, 3. \tag{30}$$

But

$$\frac{\mathscr{D}(\Phi_j)}{\iint_{C^*}\Phi_j^2\,dS} \leqq \frac{\iint_{C^*}\left(\frac{\partial\Phi_j}{\partial n}\right)^2 dS}{\mathscr{D}(\Phi_j)} = \frac{\iint_{C^*}\left(\frac{\partial x_j}{\partial n}\right)^2 dS}{W_j}, \qquad j = 1, 2, 3, \tag{31}$$

where W_j is the virtual mass associated with motion in the x_j-direction. However, since k_2 is either less than $4\pi\mathscr{C}/S$ or each of three other quantities, it must be less than $4\pi\mathscr{C}/S$ or the quotient formed by dividing the sum of the numerators of the other three quantities by the sum of the denominators, that is

$$\begin{aligned} k_2 &\leqq \max\left[\frac{4\pi\mathscr{C}}{S}, \frac{\iint_{C^*}\left[\left(\frac{\partial x}{\partial n}\right)^2+\left(\frac{\partial y}{\partial n}\right)^2+\left(\frac{\partial z}{\partial n}\right)^2\right]dS}{W_1+W_2+W_3}\right] \\ &= \max\left[\frac{4\pi\mathscr{C}}{S}, \frac{S}{3W_{\text{ave}}}\right]. \end{aligned} \tag{32}$$

This establishes (24b).

Inequalities (24a) and (24b) are inconvenient in that they relate two quantities neither of which can usually be evaluated for a given region. One would prefer to have inequalities for k_1 and k_2 which involve only the geometry of C^*. These can of course be obtained from inequalities for $\mathscr{C}$ and W_{ave}. In particular it is known (see [17] and [23]) that

$$\begin{aligned} &4\pi\mathscr{C} \leqq \mathscr{M} \leqq \frac{S^2}{3V}, && C^* \text{ convex}, \\ &4\pi\mathscr{C} \leqq \iint_{C^*}\frac{dS}{h}, && C^* \text{ star-shaped}, \end{aligned} \tag{33}$$

where $\mathscr{M}$ is the Minkowski number for C^* and $h = R\,\partial R/\partial n$. On the other hand, Pólya [16] has conjectured that

$$W_{\text{ave}} \geqq \tfrac{1}{2}V. \tag{34}$$

Although we are unable to prove this inequality for general C^* we can show that it is valid if C^* is axially symmetric but not ring-shaped. This will establish (24c).

Let us introduce the polarization potentials v_1, v_2, and v_3 (see Schiffer and Szegö [20]) brought about by uniform fields in the x, y, and z-directions

respectively. The potential v_j is $O(1/R^2)$ at infinity and assumes on C^* the value

$$v_j = x_j + \text{constant}, \tag{35}$$

where the constant is determined in such a way that v_j is $O(1/R^2)$ at infinity. The polarization $\mathscr{P}_j$ for a uniform field in the x_j-direction is defined as

$$\mathscr{P}_j = \mathscr{D}(v_j). \tag{36}$$

For C^* axially symmetric we choose the x-axis to lie along the axis of symmetry and let y and z be any two orthogonal directions in a plane perpendicular to x. Then obviously $\mathscr{P}_2 = \mathscr{P}_3$. On the other hand, the flow potential admits in this case a stream function Ψ_1, $O(1/R)$ at infinity, which satisfies the Stokes-Beltrami relations

$$r\frac{\partial \Phi_1}{\partial x} = \frac{\partial \Psi_1}{\partial r}, \quad r\frac{\partial \Phi_1}{\partial r} = -\frac{\partial \Psi_1}{\partial x} \tag{37}$$

where $r^2 = y^2 + z^2$. The condition (28) reduces to the following condition on Ψ_1 when C^* is not ring-shaped:

$$\Psi_1 = -\tfrac{1}{2}r^2 \qquad \text{on } C^*. \tag{38}$$

One can then easily check that the function v_2 (a similar expression holds for v_3) may be expressed as

$$v_2 = -\frac{2\Psi_1 y}{r^2} = -\frac{2\Psi_1 \sin\theta}{r}, \tag{39}$$

the constant in (35) vanishing in this case. The expression for $\mathscr{P}_2$ reduces to

$$\begin{aligned}\mathscr{P}_2 = 4\iiint_{D^*} \left\{ \frac{1}{r}\left[\left(\frac{\partial \Psi_1}{\partial x}\right)^2 + \left(\frac{\partial \Psi_1}{\partial r}\right)^2\right] - \frac{2\Psi_1}{r^2}\frac{\partial \Psi_1}{\partial r} + \frac{\Psi_1^2}{r^3}\right\} \sin^2\theta \, d\theta \, dr \, dx \\ + 4\iiint_{D^*} \frac{\Psi_1^2}{r^3}\cos^2\theta \, d\theta \, dr \, dx \\ = 4\pi \iint_{D_1^*} \frac{1}{r}\left[\left(\frac{\partial \Psi_1}{\partial x}\right)^2 + \left(\frac{\partial \Psi_1}{\partial r}\right)^2\right] dr\, dx - 4\pi \iint_{D_1^*} \frac{\partial}{\partial r}\left[\frac{\Psi_1^2}{r^2}\right] dr\, dx,\end{aligned} \tag{40}$$

where D_1^* is the upper half of the meridian section of D^* ($r > 0$). But the first term in the last equation is recognized as $2W_1$, and the last expression can be transformed into a boundary integral and evaluated, i.e.,

$$\begin{aligned}4\pi \iint_{D_1^*} \frac{\partial}{\partial r}\left[\frac{\Psi_1^2}{r^2}\right] dr\, dx &= 4\pi \int_{C_1^*} \frac{\Psi_1^2}{r^2}\frac{\partial r}{\partial n}\, ds \\ &= \pi \int_{C_1^*} r^2 \frac{\partial r}{\partial n}\, ds = -V,\end{aligned} \tag{41}$$

where V is the volume of C^*. We have denoted the boundary of D_1^* by C_1^*

and made use of the fact that $\Psi_1 = O(r^2)$ as $r \to 0$. Thus we have shown that for C^* axially symmetric

$$\mathscr{P}_2 = \mathscr{P}_3 = 2W_1 + V. \tag{42}$$

We have then

$$\begin{aligned} 6W_{\text{ave}} &= 2[W_1 + 2W_2] = \mathscr{P}_2 - V + 4W_2 \\ &\geqq 4\sqrt{\mathscr{P}_2 W_2} - V \geqq 3V, \end{aligned} \tag{43}$$

the latter reduction following from the inequality of Schiffer and Szegö [20]

$$\mathscr{P}_i W_i \geqq V^2. \tag{44}$$

Hence for a convex axially symmetric body

$$k_1 \leqq \frac{S}{3V} \tag{45}$$

and

$$k_2 \leqq \frac{2S}{3V} \tag{46}$$

which is (24e). On the basis of Pólya's conjecture and various known bounds for $\mathscr{C}$, one is led to conjecture that (45) and (46) are true for any C^*.

Inequality (24d) follows directly from (42) just as did (24c), i.e.,

$$\begin{aligned} 3\mathscr{P}_{\text{ave}} &= \mathscr{P}_1 + 2\mathscr{P}_2 = \mathscr{P}_1 + 4W_1 + 2V \\ &\geqq 4V + 2V = 6V. \end{aligned} \tag{47}$$

This inequality still holds even if C^* is ring-shaped; for, it was demonstrated by the work of Weiss and Payne [29] that for such boundaries C^*, (42) is replaced by

$$\mathscr{P}_2 = \mathscr{P}_3 = 2W_1 + V + Q \tag{48}$$

where Q is a positive quantity depending on the circulation about C^*.

It is apparent that (42) yields also the results

$$\begin{aligned} &\text{a)}\ W_2(W_2 + 2W_1) \geqq \frac{3V^2}{4}, \\ &\text{b)}\ \mathscr{P}_2(\mathscr{P}_2 + 2\mathscr{P}_1) \geqq 12V^2, \\ &\text{c)}\ \mathscr{P}_1\mathscr{P}_2^2 \geqq 8V^3. \end{aligned} \tag{49}$$

These quantities as well as $\mathscr{P}_{\text{ave}}$ and W_{ave}, are recognized as the invariants of the polarization and virtual mass tensors. It follows as before that b) and c) are valid even for ring-shaped domains.

We mention in passing that in a manner similar to that used in obtaining (32) one derives the alternate inequality

$$k_2 \leqq \max\left[\frac{4\pi\mathscr{C}}{S},\ \frac{\mathscr{P}_{\text{ave}}}{J_0}\right], \tag{50}$$

where J_0 is the polar moment of inertia of C^*. For a star-shaped region one obtains the following upper bound for $\mathscr{P}_j$ (see [12]):

$$\mathscr{P}_j + V = \tfrac{1}{3}\iint_{C^*} h\left[\frac{\partial}{\partial n}(x_j - v_j)\right]^2 dS \geqq \frac{\left[\iint_{C^*} x_j \frac{\partial}{\partial n}(x_j - v_j)\, dS\right]^2}{3\iint_{C^*} \frac{x_j^3}{h}\, dS} \tag{51}$$

$$= \frac{(\mathscr{P}_j + V)^2}{3\iint_{C^*} \frac{x_j^2}{h}\, dS}$$

or

$$\mathscr{P}_j + V \leqq 3\iint_{C^*} \frac{x_j^2}{h}\, dS. \tag{52}$$

From this it follows that

$$\mathscr{P}_{\text{ave}} \leqq \iint_{C^*} \frac{R^2}{h}\, dS - V. \tag{53}$$

This inequality was proved in a different way by Payne and Weinberger [13].

We can, by choosing principal directions, obtain upper bounds for k_3 and k_4 from (30) and (31). Also in the two dimensional exterior Stekloff problem (χ is $O(1/R)$ at infinity) one can easily show that the first eigenfunction is not less than the perimeter of C^* divided by twice its area. This follows directly from the fact that in two dimensions the average virtual mass is greater than the area of C^* (see [20]).

Bibliography

[1] Aronszajn, N., *Approximation methods for eigenvalues of completely continuous symmetric operators*, Symposium on Spectral Theory and Differential Problems, Stillwater, Oklahoma, 1951, pp. 179–202.

[2] Aronszajn, N., and Donoghue, F., *Variational approximation methods applied to eigenvalues of a clamped rectangular plate*, Studies in Eigenvalue Problems, Technical Report 12, University of Kansas, 1954

[3] Aronszajn, N., and Weinstein, A., *On a unified theory of eigenvalues of plates and membranes*, Amer. J. Math. Vol. 64, 1942, pp. 623 645

[4] Aronszajn, N., and Zeichner, A., *A new type of auxiliary problem for approximation of eigenvalues by Weinstein's method*, Studies in Eigenvalue Problems, Technical Report 5, Stillwater, Oklahoma, 1951.

[5] Boyce, W. E., *Vibration of Rotating Beams*, Doctoral dissertation, Carnegie Inst Tech. 1955

[6] Courant, R., and Hilbert, D., *Methoden der mathematischen Physik*, Vol. I, Springer, Berlin, 1931, English Edition, *Methods of Mathematical Physics*, Interscience, New York, 1953.

[7] Faber, G., *Beweis, dass unter allen homogenen Membranen von gleicher Spannung die kreisformige den tiefsten Grundton gibt*, S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss., 1923, pp. 169–172.

[8] Forsythe, G. E., *Asymptotic lower bounds for the frequencies of certain polygonal membranes*, Pacific J. Math., Vol. 4, 1954, pp. 467–480.

[9] Hersch, J., *Équations différentielles et fonctions de cellules*, C. R. Acad. Sci. Paris, Vol. 240, 1955, pp. 1602–1604.

[10] Krahn, E., *Über eine von Rayleigh formulierte Minimaleigenschaft des Kreises*, Math. Ann., Vol. 94, 1925, pp. 97–100.

[11] Payne, L. E., *Inequalities for eigenvalues of membranes and plates*, J. Rat. Mech. Anal., Vol. 4, 1955, pp. 517–529.

[12] Payne, L. E., and Weinberger, H. F., *New bounds in harmonic and biharmonic problems*, J. Math. Phys., Vol. 33, 1955, pp. 291–307.

[13] Payne, L. E., and Weinberger, H. F., *Upper and lower bounds for polarization*, Bull. Amer. Math. Soc. Abstract 310t, Vol. 59, 1953, p. 244.

[14] Pólya, G., *Sur la fréquence fondamentale des membranes vibrantes et la résistance élastique des tiges à la torsion*, C. R. Acad. Sci. Paris, Vol. 225, 1947, pp. 346–348.

[15] Pólya, G., *Sur une interprétation de la méthode des différences finies qui peut fournir des bornes supérieures ou inférieures*, C. R. Acad. Sci. Paris, Vol. 235, 1952, pp. 995–997.

[16] Pólya, G., *A minimum problem about the motion of a solid through a fluid*, Proc. Nat. Acad. Sci. U.S.A., Vol. 33, 1947, pp. 218–221.

[17] Pólya, G., and Szegö, G., *Isoperimetric inequalities in mathematical physics*, Annals of Math. Studies No. 27, Princeton University Press, 1951.

[18] Lord Rayleigh, *The Theory of Sound*, 2nd Edit., 1894–96, Reprinted, Macmillan, London, 1929, 1937.

[19] Ritz, W., *Über eine neue Methode zur Lösung gewisser Variationsprobleme der Mathematischen Physik*, J. Reine Angew. Math., Vol. 135, 1909, pp. 1–61.

[20] Schiffer, M., and Szegö, G., *Virtual mass and polarization*, Trans. Amer. Math. Soc., Vol. 67, 1949, pp. 130–205.

[21] Stekloff, M. W., *Sur les problèmes fondamentaux de la physique mathématique*, Ann. Sci. École Norm. Sup., Vol. 19, 1902, pp. 455–490.

[22] Szegö, G., *Inequalities for certain eigenvalues of a membrane of given area*, J. Rat. Mech. Anal., Vol. 3, 1954, pp. 343–356.

[23] Szegö, G., *Über einige neue Extremaleigenschaften der Kugel*, Math. Z., Vol. 33, 1931, pp. 419–425.

[24] Szego, G., *On membranes and plates*, Proc. Nat. Acad. Sci. U.S.A., Vol. 36, 1950, pp. 210–216.

[25] Weinberger, H. F., *Lower bounds for minima by finite difference methods*, Bull. Amer. Math. Soc., Abstract 504, Vol. 61, 1955, p. 301 (see also paper in this volume).

[26] Weinstein, A., *Étude des spectres des équations aux dérivées partielles de la théorie des plaques élastiques*, Mémorial des Sciences Mathématiques, Gauthier-Villars, Paris, No. 88, 1937.

[27] Weinstein, A., and Wei Zang Chien, *On the vibrations of a clamped plate under tension*, Quart. Appl. Math., Vol. 1, 1943, pp. 61–68.

[28] Weinstock, R., *Inequalities for a classical eigenvalue problem*, J. Rat. Mech. Anal., Vol. 3, 1954, pp. 745–753.

[29] Weiss, G., and Payne, L. E., *Torsion of a shaft with a toroidal cavity*, J. Appl. Phys., Vol. 25, 1954, pp. 1321–1328.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 543–550 (1956)

Remarks on Courant's Nodal Line Theorem

ÅKE PLEIJEL
University of Lund, Sweden
Institute for Fluid Dynamics and Applied Mathematics,
University of Maryland

Introduction

According to a well-known theorem by Sturm, a vibrating string is divided into exactly n nodal intervals by the zeros of its n-th eigenfunction ω_n. Let us also mention that in generalizing this theorem Sturm conjectured, and Liouville and Rayleigh proved, that a linear combination of ω_m, ω_{m+1}, $\cdots$, ω_n with constant coefficients has at least $m-1$ and at most $n-1$ zeros in the open interval covered by the string.

Even a superficial study of vibrating membranes shows that the possibility of generalizing theorems of this kind to higher dimensional problems is strongly limited. The intrinsic interest of Courant's nodal line theorem is the fact that it carries over one half of Sturm's nodal theorem for the string to the theory of membranes. Indeed, the theorem states that n is an upper bound for the number of nodal domains of the n-th eigenfunction of a membrane. In examples it is easily seen that this upper bound is occasionally attained.

The main point of the following lecture is the observation that Courant's nodal line theorem can be sharpened with the help of a theorem due to Faber and Krahn. Thus it will be proved that for certain membrane problems, in particular for all membranes with fixed boundaries, the maximal division by nodal lines occurs only for a *finite* number of eigenfunctions. This obviously gives a stronger emphasis to the difference between string problems and similar problems in several dimensions.

For the convenience of my audience I shall first recall a few facts about vibrating membranes and the nodal line theorem. Discussion of regularity questions will be omitted.

1. Vibrating Membranes

Let V be that connected domain of the x,y-plane which is covered by a vibrating membrane and let S be its boundary. We consider eigenvalue problems of the type

$$\begin{aligned} &\Delta u+\zeta u = 0 \text{ in } V, \\ &u = 0 \text{ on } S_0 \subset S, \\ &\frac{\partial u}{\partial \nu} = 0 \text{ on } S_1 = S-S_0, \end{aligned} \tag{1}$$

where Δ is the Laplace operator and ν is the normal of S. Provided V is bounded and S is sufficiently regular, the problem has a discrete spectrum of eigenvalues $\zeta_1 \leqq \zeta_2 \leqq \cdots \to +\infty$ and a corresponding set of eigenfunctions $\omega_1, \omega_2, \cdots$. If the boundary of the membrane is fixed, i.e. if $S = S_0$, we write $\zeta_n = \lambda_n$, $\omega_n = \varphi_n$, and in the case of a free membrane, $S = S_1$, we denote the eigenvalues and eigenfunctions by $\zeta_n = \mu_n$, $\omega_n = \psi_n$.

Let K be the class of square integrable functions defined in V which are continuous, have piecewise continuous first order derivatives and which, if S_0 is non-empty, also satisfy the boundary condition $u = 0$ on S_0. If u belongs to the subclass K_n of $K = K_1$ which is defined by the additional conditions of orthogonality

$$\int_V u\omega_i \, dV = 0, \qquad i = 1, 2, \cdots, n-1,$$

then, provided u is not identically zero,

$$\int_V \operatorname{grad}^2 u \, dV \Big/ \int_V u^2 \, dV \geqq \zeta_n . \tag{2}$$

Equality holds if and only if u is an eigenfunction with the eigenvalue $\zeta = \zeta_n$. This, in particular, requires that u is twice continuously differentiable in V.

We also recall that a solution of the membrane equation $\Delta u + \zeta u = 0$, which is regular in this way, cannot vanish in all points of a subdomain of V without vanishing identically. This follows from the fact that the solution of the Cauchy problem for the membrane equation with $u = \partial u/\partial \nu = 0$ on an arbitrary curve is unique and hence identically zero (see [6], p. 212).

2. Proof of the Nodal Line Theorem

In this section we recall Horst Herrmann's proof of the nodal line theorem in a slightly modified form (see [3]).

Let U be a twice continuously differentiable function defined in a subregion of V. A connected domain Ω is called *nodal domain* of U with respect to problem (1) if $U \neq 0$ in Ω and if Ω is bounded by lines $U = 0$ and perhaps also by arcs of S on which U then satisfies the boundary condition of problem (1).

Let us first prove

THEOREM 1. *If $-\Delta U/U \leqq t$ and $t < \zeta_n$, then the number N of nodal domains of U is less than n.*

Assume $N \geqq n$ and let $\Omega_1, \Omega_2, \cdots, \Omega_n$ be nodal domains of U bounded by $\Xi_1, \Xi_2, \cdots, \Xi_n$. Put

$$U_i = \begin{cases} U \text{ in } \Omega_i, \\ 0 \text{ in } V - \Omega_i, \end{cases}$$

and choose the constants $a_1, a_2, \cdots, a_n$ so that

$$F_n = \sum_{i=1}^{n} a_i U_i$$

belongs to K_n and is not identically zero. Hence, according to (2),

$$\int_V \operatorname{grad}^2 F_n \, dV \geqq \zeta_n \int_V F_n^2 \, dV. \tag{3}$$

On the other hand it follows by Green's theorem that

$$\int_V \operatorname{grad}^2 F_n \, dV = \sum_{i=1}^{n} a_i^2 \left(- \int_{\Omega_i} U_i \Delta U_i \, dV + \int_{\Xi_i} U_i \frac{\partial U_i}{\partial \nu} \, dS \right).$$

The integrals over Ξ_i vanish since either $U_i = 0$ or $\partial U_i / \partial \nu = 0$ on Ξ_i. From the assumption of the theorem it is clear that

$$- \int_{\Omega_i} U_i \Delta U_i \, dV \leqq t \int_{\Omega_i} U_i^2 \, dV.$$

Thus

$$\int_V \operatorname{grad}^2 F_n \, dV \leqq t \int_V F_n^2 \, dV \tag{4}$$

which, together with (3), contradicts the assumption $t < \zeta_n$.

If $\zeta_n < \zeta_{n+1}$ and $t = \zeta_n$ Theorem 1 shows that $N \leqq n$. Courant's nodal line theorem states that this also holds when $\zeta_n = \zeta_{n+1}$.

THEOREM 2. *If $-\Delta U / U \leqq \zeta_n$, then the number N of nodal domains of U is less than or equal to n.*

Suppose $N > n$ and choose F_n as above. Since (4) is valid with $t = \zeta_n$, equality must hold in (3) so that F_n is an eigenfunction (see the end of Section 1). But F_n vanishes identically in Ω_{n+1} which is impossible since F_n is not identically zero in V.

3. Consequences of the Nodal Line Theorem

For $U = \omega_n$ we obtain the nodal line theorem as it was stated by Courant in [1], p. 392. *The number of nodal domains of an eigenfunction belonging to the n-th eigenvalue is less than or equal to n.*

If ω is an eigenfunction belonging to the first eigenvalue, the domain V itself is the only nodal domain of this function which is consequently $\neq 0$ in V (it is known from Weber's mean value theorem that an eigenfunction which vanishes at a point of V necessarily assumes both positive and negative values, see [6], p. 218). The possibility of orthogonalizing eigenfunctions belonging to the same eigenvalue shows that the first eigenvalue is always simple. Because of the orthogonality of eigenfunctions belonging to different eigenvalues it also follows that $N \geqq 2$ if $n > 1$. *Thus the first eigenfunction is the only one not changing its sign in V.*

G. Pólya has proved that $\mu_2 < \lambda_1$ [7]. By applying Theorem 1 to the function $U = \psi_2$ it follows that the number of nodal domains of ψ_2 with respect to the problem of the membrane with fixed boundary is less than 1. This means that ψ_2 can have no closed nodal line. Hence, using the theorem quoted at the end of Section 1, it follows that the nodal lines of ψ_2 consist of a transverse cutting V into two simply connected regions.

It was recently shown by L. E. Payne [5] that if V is convex, $\mu_3 < \lambda_1$ and that for $n > 3$, $\mu_n \leqq \lambda_{n-2}$. It follows that the nodal lines of ψ_3 consist of one or two transverses not cutting each other, and that ψ_n can have at most $n-2$ "interior" nodal domains.

These properties of the nodal lines of a free membrane are also easily deduced from Courant's nodal line theorem as stated above, by observing that λ_n depends monotonically on V.

Let us also mention the interesting application of the nodal line theorem made by L. E. Payne and H. F. Weinberger in proving a conjecture by A. Weinstein [5].

4. Example of Maximal Subdivision by the Nodal Lines

The maximal case $N = n$ in Courant's theorem holds in a trivial way when $n = 1, 2$. That it can also occur for larger values of n is usually exemplified by reference to the case of the square membrane with fixed boundary. If V is the square $0 \leqq x \leqq \pi$, $0 \leqq y \leqq \pi$, a complete system of eigenfunctions vanishing on the boundary is given by

$$\sin kx \sin ly, \tag{5}$$

where k and l are positive integers. The corresponding eigenvalues are $\lambda = k^2+l^2$, the first ones being $\lambda_1 = 2$, $\lambda_2 = \lambda_3 = 5$, $\lambda_4 = 8$, $\lambda_5 = \lambda_6 = 10$, $\cdots$. The number of nodal domains of (5) is $N = kl$ which for $\lambda = \lambda_4$ equals 4.

The general eigenfunction with eigenvalue λ is

$$\sum_{k^2+l^2=\lambda} C_{kl} \sin kx \sin ly, \tag{6}$$

where the C_{kl} are constants. It should be observed that even if the number of nodal domains for each term in (6) is less than a certain value it is quite possible that the sum itself has a larger number of nodal domains. This, for instance, occurs when $\lambda = 10$.

5. Sharpening of the Nodal Line Theorem for Membranes with Fixed Boundaries

In his paper ([3], p. 236) H. Herrmann remarked that a possibility of sharpening the nodal line theorem would be to show that for sufficiently

large values of n the maximal case $N = n$ never occurs. However, he considered that such a large deviation, from what holds for the vibrating string, is improbable.

In 1923 and 1924 G. Faber [2] and E. Krahn [4] proved Rayleigh's conjecture that for membranes with fixed boundaries and with given area V, the circular one has the smallest first eigenvalue. Later G. Pólya and G. Szegö [8], p. 6, showed that the Faber-Krahn theorem can be deduced by Steiner's symmetrization. The theorem can be expressed by the inequality

$$\lambda_1 \geqq \frac{\pi j^2}{V}, \tag{7}$$

where V is the area of the membrane and j is the smallest positive zero of the Bessel function J_0. The theorem holds even if the region of the membrane is multiply connected.

Let $\Omega_1, \Omega_2, \cdots, \Omega_N$ be nodal domains of the eigenfunction φ of a membrane with fixed boundary and assume that this eigenfunction belongs to the n-th eigenvalue $\lambda = \lambda_n$. In each Ω_i the function $\varphi \neq 0$, so that λ_n is the first eigenvalue of a membrane covering Ω_i and being fixed along the boundary of this domain. Hence, according to (7), the value λ_n satisfies the inequalities

$$\frac{\Omega_i}{\pi j^2} \geqq \frac{1}{\lambda_n} \tag{8}$$

in which Ω_i denotes the area of the i-th nodal domain. By adding these inequalities we find

$$\frac{V}{\pi j^2} \geqq \frac{N}{\lambda_n}, \tag{9}$$

where V is the area of the membrane. Thus a necessary condition that the maximal case $N = n$ should occur is that

$$\frac{V}{\pi j^2} \geqq \frac{n}{\lambda_n}. \tag{10}$$

Let us assume $N = n$ for infinitely many values of n. According to Weyl's asymptotic law for the eigenvalues, the right hand side of (10) tends to $V/4\pi$ as n tends to infinity. It then follows that

$$\frac{V}{\pi j^2} \geqq \frac{V}{4\pi}$$

which can be true only if $2 \geqq j$. But $j = 2.4048\ldots$. Hence it is proved that *for the eigenfunctions φ_n of a membrane with fixed boundary the maximum n of the number of nodal domains is attained only for a finite number of eigenvalues.*

From (9) it follows that

$$\limsup_{n\to\infty} \frac{N}{n} \leqq \left(\frac{2}{j}\right)^2 = 0.691\ldots$$

which shows that for every $\varepsilon > 0$ the inequality

$$N \geqq \left[\left(\frac{2}{j}\right)^2 + \varepsilon\right] n$$

is valid only for a finite number of values of n.

6. Square Membrane with Fixed Boundary

If the eigenfunctions and eigenvalues are explicitly known, the necessary condition (10) for maximal subdivision by nodal lines can be used to isolate those values of n for which the number of nodal domains equals n. Let us for instance consider the square membrane with fixed boundary of Section 4. For this membrane the number $m(\lambda)$ of eigenvalues $k^2+l^2 < \lambda$ coincides with the number of lattice points (k, l) in $x^2+y^2 < \lambda$, $x \geqq 1$, $y \geqq 1$. Covering this region by squares $k \leqq x \leqq k+1$, $l \leqq y \leqq l+1$ with corners (k, l) in the region, one finds that

$$m(\lambda) > \frac{\pi\lambda}{4} - 2\sqrt{\lambda} + 1.$$

In considering cases of maximal subdivision $N = n$ we may assume $\lambda_{n-1} < \lambda_n$ so that $m(\lambda_n) = n-1$. Hence

$$n > \frac{\pi\lambda_n}{4} - 2\sqrt{\lambda_n} + 2.$$

If this result is introduced in (10) it follows, since $V = \pi^2$, that the condition

$$\frac{\pi}{4} - \frac{2}{\sqrt{\lambda_n}} + \frac{2}{\lambda_n} < \frac{\pi}{j^2}$$

is necessary for maximal subdivision. This inequality can only be fulfilled if $\lambda_n < 51$. Now $\lambda_{33} = 50$, $\lambda_{34} = 52$ and it is easy to investigate whether relation (10), which for the square has the form

$$\frac{n}{\lambda_n} \leqq \frac{\pi}{j^2} = 0.545\ldots,$$

holds for the different eigenvalues $\lambda_n \leqq 50$. This investigation leaves us with the only possible values $\lambda = 2, 5, 8, 10, 13, 17$. For the last three of these remaining values the nodal lines of the corresponding eigenfunctions (6) are easily constructed (see [1], p. 259 and [6], p. 80). *In this way one*

finds that maximal subdivision by nodal lines at the square membrane with fixed boundary only occurs for the first, second and fourth eigenvalues.

By considering the particular eigenfunctions $\sin kx \sin ky$ it is seen that for a square membrane with fixed boundary

$$\limsup_{n\to\infty} \frac{N}{n} \geqq \frac{2}{\pi} = 0.636\ldots .$$

Thus, for every $\varepsilon > 0$ the inequality

$$N \geqq \left(\frac{2}{\pi} - \varepsilon\right) n$$

holds for infinitely many values of n.

7. On the Problem of Sharpening the Nodal Line Theorem for Free Membranes

In the case of a free membrane the inequality (8) cannot be applied to nodal domains adjacent to S and the reasoning of Section 5 is not valid. In spite of this it seems highly probable that the result of that section is also true for free membranes. This belief is supported by the fact that the theorem holds for the free rectangular membrane as well as in other cases when the method of separating variables can be applied.

Consider, for instance, the free square membrane $0 \leqq x \leqq \pi$, $0 \leqq y \leqq \pi$. Here the eigenvalues are $\mu = k^2+l^2$, where k and l are non-negative integers, the corresponding eigenfunctions being

$$\psi = \sum_{k^2+l^2=\mu} C_{kl} \cos kx \cos ly.$$

Since $\psi(x, 0)$, $\psi(x, \pi)$, $\psi(0, y)$, $\psi(\pi, y)$ are polynomials of degree $\leqq \sqrt{\mu}$ in $\cos x$ and $\cos y$, the number of "nodal points" on the boundary is $\leqq 4\sqrt{\mu}$ which implies that the number of nodal domains adjacent to the boundary has the same upper bound. Consequently, if N is the number of nodal domains of an eigenfunction belonging to the eigenvalue $\mu = \mu_n$, the number of "inner" nodal domains is $\geqq N-4\sqrt{\mu_n}$. For these inner nodal domains relations (8) are valid, and by adding these relations we find

$$\frac{V}{\pi j^2} \geqq \frac{N-4\sqrt{\mu_n}}{\mu_n}.$$

Assuming that $N = n$ for an infinite number of values of n we obtain by transition to the limit the same contradiction as in Section 5.

While the method of Section 5 carries over to any number of dimensions greater than 2 this is not true for our treatment of the special problem of this section. In order to treat, for instance, the case of the free three-dimen-

sional membrane $0 \leqq x \leqq \pi$, $0 \leqq y \leqq \pi$, $0 \leqq z \leqq \pi$, it would be necessary to use, in a special case, the theorem quoted in [1], p. 394. This theorem which generalizes part of the Liouville-Rayleigh theorem for the string asserts that a linear combination, with constant coefficients, of the n first eigenfunctions can have at most n nodal domains. However, as far as I have been able to find, there is no proof of this assertion in the literature.

Bibliography

[1] Courant, R., and Hilbert, D., *Methoden der mathematischen Physik*, Vol. I, Springer, Berlin, 1931. English Edit., *Methods of Mathematical Physics*, Vol. I, Interscience, New York, 1953.

[2] Faber, G., *Beweis, dass unter allen homogenen Membranen von gleicher Fläche und gleicher Spannung die kreisförmige den tiefsten Grundton gibt*, S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss., 1923, pp. 169–172.

[3] Herrmann, H., *Beziehungen zwischen den Eigenwerten und Eigenfunktionen verschiedener Eigenwertprobleme*, Math. Z., Vol. 40, 1935, pp. 221–241.

[4] Krahn, E., *Über eine von Rayleigh formulierte Minimaleigenschaft des Kreises*, Math. Ann., Vol. 94, 1924, pp. 97–100.

[5] Payne, L. E., *Inequalities for eigenvalues of membranes and plates*, J. Rat. Mech. Anal., Vol. 4, 1955 (to appear).

[6] Pockels, F., *Über die partielle Differentialgleichung $\Delta u + k^2 u = 0$ und deren Auftreten in der mathematischen Physik*, Z. Math. Physik, Vol. 37, 1892, pp. 100–105.

[7] Pólya, G., *Remarks on the foregoing paper*, J. Math. Physics, Vol. 31, 1952, pp. 55–57.

[8] Pólya, G., and Szegö, G., *Isoperimetric inequalities in mathematical physics*, Annals of Math. Studies No. 27, Princeton University Press, 1951.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 551–568 (1956)

Remarks on the Application of Gradient Mappings to the Calculus of Variations and the Connected Boundary Value Problems in Partial Differential Equations

E. H. ROTHE
University of Michigan

1. Introduction

Let

$$j(y) = \int_D f(t, y, y_1, \cdots, y_n)\, dt. \tag{1.1}$$

Here D denotes a bounded open domain in the n-dimensional real Euclidean space of points t with coordinates $t_1, \cdots, t_n$ while $y_j = \partial y/\partial t_j$. We consider the minimum problem for (1.1) among functions $y = y(t)$ satisfying the boundary condition

$$y = 0 \text{ on } \dot{D} \tag{1.2}$$

where $\dot{D}$ denotes the boundary of D.

The aim of the present paper is to start a treatment of such problems which is based on the theory of gradient mappings. The facts used from this theory are essentially the follwing:[1] Let H be a real Hilbert space, V a closed bounded convex set in H, and $I(x)$ a scalar (i.e., a real valued function) defined for points x in V. Then, if $I(x)$ is weakly continuous[2], it reaches a maximum and a minimum in V since V is weakly compact. Moreover, since the norm $||x||$ and its square (x, x) are weakly lower semi-continuous,[3] the scalar

$$i(x) = x + I(x) \tag{1.3}$$

is weakly lower semi-continuous and therefore reaches a minimum in some point x_0 of V. Finally, if x_0 is an interior point of V, then grad $i(x_0) = 0$; i.e., we have with $G(x) = \text{grad}\, I(x)$ [4]

[1] Cf. [13].

[2] This means continuous in the "bounded weak topology". The neighborhoods of this topology are the intersections of the weak neighborhoods with some sphere containing V, cf. [13], p. 10.

[3] Cf. [13], p. 11.

[4] We recall the definition of grad I in a Hilbert space H: the Fréchet differential (cf. Def. 3.1) $dI = d(x, h)$ of I at x is a linear bounded functional of h. Therefore, there exists a uniquely determined element $G = G(x)$ in H such that $dI = (h, G)$ where the parenthesis denotes the scalar product. By definition $G = \text{grad}\, I$.

$$g(x) = x+G(x) = 0 \text{ for } x = x_0 . \tag{1.4}$$

For the application of this theory one has to be sure that $I(x)$ is weakly continuous. It is known that this is the case if $g(x)-x = G(x)$ is completely continuous[5]. However in the direct application to the variational problem in question this assumption is not satisfied. Indeed the first variation of the integral (1.1) belonging to the increment η of y is of the form

$$\int_D \eta(t)\, \tilde{E}(y)\, dt = (\eta, \tilde{E}) \tag{1.5}$$

where the parenthesis at the right denotes the usual scalar product in the space $L^2(D)$ of functions which are square integrable over D and where $\tilde{E}(y)$ denotes the Euler expression

$$\tilde{E}(y) = f_y - \sum_{i=1}^{n} \frac{d}{dt_i} f_{y_i}. \tag{1.6}$$

This would suggest to consider $\tilde{E}(y)$ as the gradient of $j(y)$ (cf. footnote 4). However $E(y)-y$, being a differential operator of order two, is not completely continuous in $L^2(D)$.

The present paper deals only with the special case that f is of the form

$$f(t, y, y_1, \cdots, y_n) \equiv \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik}(t)\, y_i y_k + \sum_{i=1}^{n} b_i(t)\, y_i + c(t, y) \tag{1.7}$$

where the quadratic form with the matrix (a_{ik}) is positive definite in the closure $\overline{D}$ of D. In this case, to overcome the difficulty just mentioned, we will proceed as follows: Let

$$L(y) = -2 \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{\partial}{\partial t_i} \left(a_{ik} \frac{\partial y}{\partial t_k} \right). \tag{1.8}$$

We consider the boundary value problem

$$L(y) = x, \qquad y = 0 \text{ on } \dot{D} \tag{1.9}$$

for $x = x(t)$ belonging to a linear dense subset $\tilde{L}$ of L^2 of sufficiently smooth functions. If

$$y = M(x) \tag{1.10}$$

is the solution of the problem (1.9), then M turns out to be a bounded linear positive definite symmetric operator on $\tilde{L}$ (Lemma 2.2). This space can therefore be re-normed[6] by using the norm which is based on the scalar product

$$[x, \xi] = (x, M(\xi)). \tag{1.11}$$

---

[5] Cf. [13], p. 11.

[6] The device of re-norming the Hilbert space was first used by K. O. Friedrichs [3]

This re-normed space will be denoted by $\tilde{H}$ and its completion by H. We now introduce x instead of y as the "independent variable" in (1.1); i.e., we set

$$i(x) = j(M(x)). \tag{1.12}$$

It will be shown that the gradient $g(x)$ of $i(x)$ (with its definition based on the scalar product (1.11)) is $\tilde{E}(M(x))$ (cf. (1.6)) and has the desired property that $g(x)-x$ is completely continuous (Theorems 3.1 and 3.2). It can now be concluded immediately that any closed solid sphere V of H contains at least one point x_0 for which $i(x_0)$ is a minimum in V. The point $y_0=M(x_0)$ gives then a minimum for the scalar (1.1). While x_0 may be an "abstract" element of H, y_0 turns out to be in $L^2(D)$ and is actually of class $\mathfrak{B}_2$ in the sense of Calkin and Morrey.[7] Moreover, a sufficient condition is given for x_0 to be an interior point of V provided that the radius of this sphere is large enough (Theorem 3.4). In this case x_0 satisfies (1.4), and $y_0 = M(x_0)$ satisfies the (generalized) Euler equation $\tilde{E}(y_0) = 0$. $g(x)$ will be called the Euler operator.

Under further conditions on the coefficients of (1.7) it will be shown in Section 4 that the Euler operator has a Fréchet differential

$$l_2(x, \xi) = \xi + L(x, \xi)$$

with completely continuous L. The linear form l_2 in ξ will be called the Jacobi operator (at x). It will be seen that for a critical point x_0 of $i(x)$ (i.e., a point x_0 satisfying (1.4)) to be isolated it is sufficient that the Jacobi operator at x_0 is non-singular. In this case it follows easily that $d^2i(x, \xi, \zeta) = [l_2(x_0, \xi), \zeta)]$ is the second Fréchet differential of $i(x)$ at x_0, and that for $\xi = \zeta$ this differential is a non-degenerate quadratic form which in its normal form contains at most a finite number of negative squares.[8] Earlier results[9] show that for the topologically defined Morse type numbers m^r of dimension r the relations $m^r = \delta_s^r$ hold, a result which is classical for non-degenerate critical points of a function of a finite number of variables.

2. The Operator M

The object of this section is the discussion of the various spaces referred to in the introduction and of the operator (1.10).

LEMMA 2.1. *Let D be the domain of the introduction. We assume that its boundary $\dot{D}$ has Hölder continuous derivatives up to and including order three.*

---

[7] Cf. [5], p. 4.

[8] For a treatment of question concerning the index form by using directly forms in Hilbert space (instead of going to the limit in the dimension) see [4].

[9] Cf. [11], p. 466.

Let $a_{ik}(t)$, $i, k = 1, 2, \cdots, n$, be functions defined in the closure $\overline{D}$ of D and have there Hölder continuous derivatives up to and including order two while their matrix is positive definite for every t in $\overline{D}$. Finally let $x = x(t)$ have Hölder continuous derivatives of order one in $\overline{D}$. Then the boundary value problem (1.9) *has one and only one solution $y(t)$ which has continuous derivatives up to and including order three in $\overline{D}$.*

Proof: The uniqueness is classical. A proof for the existence of a solution with continuous second order derivatives was given by Schauder [14]. The extension of Schauder's results to derivatives of order three has been effected by Morrey.[10]

LEMMA 2.2. *As in the introduction let $L^2 = L^2(D)$ be the Hilbert space of functions $y = y(t)$ which are square integrable over $\overline{D}$. Let $\tilde{L}$ be the subset of elements of L^2 which satisfy the assumptions made about $x(t)$ in Lemma* 2.1. *For $x \subset \tilde{L}$ denote the unique solution $y = y(t)$ of* (1.9) *by*

$$y = M(x) \tag{2.1}$$

and the derivatives $y_j = \partial y/\partial t_j$ by

$$y_j = M_j(x), \qquad j = 1, 2, \cdots, n. \tag{2.2}$$

Then a) M and M_j are linear bounded operators with domain $\tilde{L}$; b) M is symmetric and positive definite.

Proof: a) The linearity is obvious. The boundedness follows from the fact that there exists a constant μ such that for the solution y of (1.9) and its derivatives the inequalities

$$\int_D y^2 dt \leqq \mu \int_D x^2 dt, \qquad \int_D \left(\frac{\partial y}{\partial t_j}\right)^2 dt \leqq \mu \int_D x^2 dt \tag{2.3}$$

hold[11]. b) Let $x = x(t)$, $\xi = \xi(t)$ be elements of $\tilde{L}$, and $y = M(x)$, $\eta = M(\xi)$. With the usual notation for the scalar product in L^2 we have from Green's theorem

$$(\xi, M(x)) = (L(\eta), y) = 2 \int_D \sum_{i,k=1}^{n} a_{ik} \frac{\partial y}{\partial t_i} \frac{\partial \eta}{\partial t_k} dt. \tag{2.4}$$

This obviously implies the symmetry of M. Moreover (2.4) with $\xi = x$ together with the positive definiteness assumption about the matrix of the a_{ik} implies the existence of a constant m such that the inequality

$$(x, M(x)) \geqq m \int_D \sum_{i=1}^{n} \left(\frac{\partial y}{\partial t_i}\right)^2 dt \tag{2.5}$$

holds. Thus $(x, M(x)) \geqq 0$. But here the equality sign can, by (2.5), hold

[10] Cf. [6], Theorem 7.5, p. 149.
[11] Cf. [2].

only if y is constant, and, therefore on account of the boundary condition, if $y = 0$. But then $x = L(y) = 0$. Thus $(x, M(x)) > 0$ unless $x = 0$.

Lemma 2.2 allows us to re-norm $\tilde{L}$ by using the scalar product (1.11).

DEFINITION 2.1. The linear space which as a point set coincides with $\tilde{L}$ and whose scalar product is defined by (1.11) is called $\tilde{H}$. The Hilbert space obtained from $\tilde{H}$ by completion is called H. The norm of an element y of $\tilde{L}$ or L^2 is $||y|| = +\sqrt{(y, y)}$. The norm of an element x in $\tilde{H}$ or H is $+\sqrt{[x, x]}$ where $[x, x]$ is given by (1.11).

M and M_j are bounded operators on $\tilde{L}$. We claim they are also bounded on $\tilde{H}$. This is an immediate consequence of the following

LEMMA 2.3. *Let* $||M||$ *be the norm of* M *in* $\tilde{L}$. *Then*

$$[M(x), M(x)] \leqq ||M||^2 [x, x], \tag{2.6}$$

$$[M_j(x), M_j(x)] \leqq ||M|| \frac{[x, x]}{m}. \tag{2.7}$$

Proof: Since M is a bounded linear symmetric positive definite operator on $\tilde{L}$, there exists a bounded symmetric operator $\sqrt{M}$ with the properties

$$\sqrt{M}\sqrt{M} = M, \qquad ||\sqrt{M}|| = \sqrt{||M||}.^{12}$$

We have by (1.11)

$$[M(x), M(x)] = (M(x), M^2(x)) = (M\sqrt{M(x)}, M\sqrt{M(x)})$$
$$\leqq ||M||^2 (\sqrt{M(x)}, \sqrt{M(x)}) = ||M||^2 (x, M(x))$$

which proves (2.6). In a similar way we obtain

$$[M_j(x), M_j(x)] \leqq ||M_j(x)||^2 \, ||M||.$$

If we combine this inequality with the inequality

$$||M_j(x)||^2 = \int_D \left(\frac{\partial y}{\partial t_j}\right)^2 dt \leqq \frac{[x, x]}{m} \tag{2.8}$$

derived from (2.5) we obtain (2.7).

DEFINITION 2.2. By the preceding lemma M and M_j as operators on $\tilde{H}$ can be extended to H. From now on M and M_j denote these extended operators mapping H into itself.

LEMMA 2.4. *A Cauchy sequence of* $\tilde{L}$ *is a Cauchy sequence of* $\tilde{H}$.

Proof: The lemma follows from the fact that, as vector spaces, $\tilde{L}$ and $\tilde{H}$ are identical and from the inequality

$$[x, x] = (x, M(x)) = (\sqrt{M(x)}, \sqrt{M(x)}) \leqq ||M||(x, x) = ||M|| \, ||x||^2. \tag{2.9}$$

[12] Cf. [7], p. 15.

LEMMA 2.5. *With proper identification, L^2 as a vector space is a subspace of H as a vector space.*

Proof: An element x of L^2 if determined by a Cauchy sequence of elements x_n in $\tilde{L}$. By the preceding lemma $\{x_n\}$ is also a Cauchy sequence in $\tilde{H}$ and therefore determines an element x' of H. We identify x' with x.

LEMMA 2.6. *For each $x \subset H$ the elements $M(x)$, $M_j(x)$ of H are in L^2 (cf. the identification of the preceding lemma).*

Proof: Let $\{x_n\}$ be a sequence of elements of $\tilde{H}$ converging to x in H. The elements $y_n = M(x_n)$ are then in $\tilde{L}$ since $\tilde{H}$ and $\tilde{L}$ are identical as point sets and M maps $\tilde{L}$ into itself. We claim that the y_n form a Cauchy sequence in L^2. Indeed

$$(2.10)\qquad \begin{aligned}(y_n-y_m,\ y_n-y_m) &= (M(x_n-x_m),\ M(x_n-x_m)) \\ &\leqq ||M||\,(\sqrt{M(x_n-x_m)},\ \sqrt{M(x_n-x_m)}) = ||M||\,[x_n-x_m,\ x_n-x_m].\end{aligned}$$

Since $\{x_n\}$ is a Cauchy sequence in $\tilde{H}$, this inequality shows that the y_n form a Cauchy sequence in L^2. By Lemma 2.4 they form a Cauchy sequence also in H and, by the identification of Lemma 2.5, the limit

$$(2.11)\qquad y = \lim_{n\to\infty} y_n = \lim_{n\to\infty} M(x_n)$$

is the same whether the limit is taken in H or L^2. Finally since M is a continuous map $H \to H$, the limit (2.11) equals $M(x)$, which proves the part of the lemma concerning M. The part concerning M_j is proved correspondingly if instead of (2.10) one uses the inequality obtained by replacing x by $x_n - x_m$ in (2.8).

DEFINITION 2.3. By the preceding lemma we may consider M and M_j as mappings $H \to L^2$. Considered in this way they will be denoted by M^2, M_j^2, respectively, while the notation M, M_j will always be used in the sense of Definition 2.2.

LEMMA 2.7. *For $x \subset H$ the element $y = M(x)$ of H is a function of class $\mathfrak{P}_2$ in the sense of Calkin and Morrey.*[13]

Proof: By Lemma 2.6, y is in L^2. According to the definition of the class $\mathfrak{P}_2$ we have therefore only to show that y possesses "generalized derivatives" $\partial y/\partial t_j$, $j = 1, 2, \cdots, n$, in the following sense which we explain for $j = 1$: denote by C_n an n-cell of points $t = (t_1, t_2, \cdots, t_n)$ of D given by the inequalities

$$(2.12)\qquad a_i \leqq t_i \leqq b_i, \qquad i = 1, 2, \cdots, n,$$

and by C' the $(n-1)$-cell of points $t' = (t_2, t_3, \cdots, t_n)$ given by the last

[13]For the definition of the class $\mathfrak{P}_2$ see [5], p. 4.

$n-1$ of the inequalities (2.12). We then have to exhibit a summable function $\eta(t)$ such that for almost all cells

$$\int_{C'} [y(b_1, t') - y(a_1, t')]\, dt' = \int_C \eta(t)\, dt. \tag{2.13}$$

We choose for η the function $\eta = M_1(x)$ which is summable by Lemma 2.6. To show that (2.13) is true with this η we consider a sequence $\{x^n\}$ of elements in $\tilde{H}$ converging to x. $y^n = M(x^n)$ and $\eta^n = M_1(x^n)$ are then elements of $\tilde{L}$ and we know (cf. Lemma 2.6 and its proof) that

$$\lim_{n\to\infty} y^n = \lim_{n\to\infty} M(x^n) = y \text{ in } L^2, \tag{2.14}$$

$$\lim_{n\to\infty} \eta^n = \lim_{n\to\infty} M_1(x^n) = M_1(x) = \eta \text{ in } L^2. \tag{2.15}$$

Since for each n, η^n is the derivative with respect to t_1 of y^n we have for each cell C

$$\int_{C'} [y^n(b_1, t') - y^n(a_1, t')]\, dt' = \int_C \eta^n(t)\, dt. \tag{2.16}$$

Because of (2.15) it is clear that for $n \to \infty$ the right member of (2.16) approaches the right member of (2.13). To prove (2.13) for almost all cells C it will therefore be sufficient to show that for almost all a_1, b_1 a subsequence of the left numbers of (2.16) tends to the left member of (2.13). In other words we have to prove: if

$$\delta_n(t_1) = \int_{C'} [y^n(t_1, t') - y(t_1, t')]\, dt', \tag{2.17}$$

then for some subsequence $\{n_i\}$

$$\lim_{i\to\infty} \delta_{n_i}(t_1) = 0 \text{ for almost all } t_1.$$

But it is well known that this relation is a consequence of

$$\lim_{n\to\infty} \int_{a_1}^{b_1} \delta_n^2(t_1)\, dt_1 = 0,$$

and this latter relation is, because of (2.14), a consequence of the inequality

$$\prod_{j=2}^{n} \frac{1}{(b_j - a_j)} \int_{a_1}^{b_1} \delta_n(t_1)^2\, dt_1 \leqq \int_{a_1}^{b_1} \int_{C'} [y^n(t_1, t') - y(t_1, t')]^2\, dt'\, dt_1$$

$$\leqq \int_D [y^n(t) - y(t)]^2\, dt$$

which in turn follows from (2.17) by using Schwarz's inequality.

We conclude this section with the following theorem which is basic for all that follows.

THEOREM 2.1. *The mappings M, M_j (cf. Definition 2.2) and M^2 (cf. Definition 2.3) are completely continuous.*

Proof: The maps in question are extensions to H of maps defined in the dense subset $\tilde{H}$ of H. It is easy to see that it is sufficient to prove the theorem for the restrictions of the maps to $\tilde{H}$. We first prove the part of the theorem concerning M^2. We have to show: Let V_R be the sphere in $\tilde{H}$ defined by

$$[x, x] \leqq R^2, \qquad x \subset \tilde{H}, \tag{2.18}$$

and $\{x^n\}$ a sequence of points of V_R. Then the sequence $\{y^n\} = \{M(x^n)\}$ contains a subsequence which is a Cauchy sequence in the L^2 metric. The proof will be based on Rellich's lemma.[14] We first note the inequality

$$||M(x)||^2 \leqq ||M||\,[x, x] \tag{2.19}$$

which is proved like (2.10). For $x \subset V_R$, (2.19) implies

$$\int_D y^2(t)\,dt \leqq ||M||\,R^2; \tag{2.20}$$

on the other hand (2.8) shows that for $x \subset V_R$

$$\int_D \left(\frac{\partial y}{\partial t_j}\right)^2 dt \leqq \frac{R^2}{m}. \tag{2.21}$$

But by Rellich's lemma the inequalities (2.20), (2.21) imply the existence of a Cauchy subsequence of the sequence $\{y^n\} = \{M(x^n)\}$. This proves the part of the theorem concerning M^2. It also proves the part concerning M since, by Lemma 2.4, y_n is a Cauchy sequence also in H.

To prove the part concerning M_j we note first that $\partial y/\partial t_i$ satisfies the assumptions made about x. Therefore we can, in (2.8), replace x by $y_i = \partial y/\partial t_i = M_i(x)$ and consequently y by $M(y_i)$. We thus obtain

$$\int_D \left(\frac{\partial M(y_i)}{\partial t_j}\right)^2 dt \leqq \frac{[y_i, y_i]}{m}. \tag{2.22}$$

But by (2.9) and (2.21) the right member of (2.22) is not greater than $||M||\,(y_i, y_i)/m \leqq ||M||\,R^2/m^2$. Thus

$$\int_D \left(\frac{\partial M(y_i)}{\partial t_j}\right)^2 dt \leqq ||M||\left(\frac{R}{m}\right)^2. \tag{2.23}$$

On the other hand we obtain from (2.21) the inequality

$$\int_D (M(y_i))^2\,dt \leqq ||M||^2 \int_D \left(\frac{\partial y}{\partial t_i}\right)^2 dt \leqq ||M||^2 \frac{R^2}{m}. \tag{2.24}$$

Again, by Rellich's lemma the inequalities (2.23), (2.24) allow us to

[14] Cf [8].

conclude that the sequence $\{M(y_i^n)\}$ contains an L^2 Cauchy subsequence if $\{y_i^n\} = \{M_i(x^n)\}$. We claim that for the corresponding subsequence of the integers n, the y_i^n form a Cauchy sequence in H. This follows from the following inequalities obtained by application of the Schwarz inequality and of (2.21):

$$\begin{aligned}[y_i^n - y_i^m\,,\, y_i^n - y_i^m] &= (y_i^n - y_i^m\,,\, M(y_i^n - y_i^m)) \\ &\leqq (\|y_i^n\| + \|y_i^m\|)\,\|M(y_i^n) - M(y_i^m)\| \\ &\leqq \|M(y_i^n) - M(y_i^m)\|\,\frac{2R}{\sqrt{m}}\,.\end{aligned}$$

3. The Scalar (1.1)

The object of this section is to investigate (1.1) in the form (1.12). Particularly it will be shown that $i(x)$ has a gradient which satisfies the assumptions discussed in the introduction and allows immediately to conclude the existence of a minimum of (1.12).

We consider first the differential expression

$$\lambda(y) = \gamma(t, y) + \sum_{i=1}^{n} \beta_i(t)\, y_i + \sum_{i=1}^{n}\sum_{k=1}^{n} a_{ik}\, y_i y_k\,, \qquad y_i = \frac{\partial y}{\partial t_i}\,. \tag{3.1}$$

LEMMA 3.1. *Let a_{ik} satisfy the same assumptions as in Lemma 2.1. Let $\beta_i(t)$ be continuous in $\overline{D}$, and $\gamma(t, y)$ be continuous for $t \subset \overline{D}$ and satisfy a Lipschitz condition for all y. For $x \subset \tilde{L}$ (cf. Lemma 2.2) we define $\Lambda(x)$ by the integral*

$$\Lambda(x) = \int_D \lambda(y)\, dt \tag{3.2}$$

where y in $\lambda(y)$ is replaced by the operator $M(x)$ defined in (1.10). Then:

a) *In each sphere V_R in H (Definition 2.1) defined by $[x, x] \leqq R^2$, Λ satisfies a Lipschitz condition; i.e., there exists a constant C_R such that*

$$|\Lambda(x+\xi) - \Lambda(x)|^2 \leqq C_R^2[\xi, \xi] \quad \text{for} \quad x,\, x+\xi \subset V_R\,. \tag{3.3}$$

b) *Λ can be extended to all of H in such a way that (3.3) holds.*

Proof: a) Denote by λ_0, λ_1, λ_2 the first, second and third term of (3.1), respectively, and by Λ_i, $i = 0, 1, 2$, the integral in (3.2) with λ replaced by λ_i such that

$$\Lambda = \Lambda_0 + \Lambda_1 + \Lambda_2\,. \tag{3.4}$$

If then L_0 denotes a Lipschitz constant for γ, and $\mu(D)$ the measure of D we have by (2.19) with $\eta = M(\xi)$, $\eta_i = \partial M/\partial t_i$.

$$\begin{aligned}|\Lambda_0(x+\xi) - \Lambda_0(x)|^2 &= \left|\int_D [\gamma(t, y+\eta) - \gamma(t, y)]\, dt\right|^2 \leqq L_0^2 \left(\int_D |\eta|\, dt\right)^2 \\ &\leqq L_0^2\, \mu(D)\, \|\eta\|^2 \leqq L_0^2 \mu(D)\, \|M\|\, [\xi, \xi].\end{aligned} \tag{3.5}$$

Moreover, if B denotes an upper bound for $\beta_i(t)$, $t \subset \overline{D}$, $i = 1, 2, \cdots, n$, we have by (2.7)

$$\left(\int_D \beta_i(t)\,\eta_i\,dt\right)^2 \leqq B^2\mu(D)\,||\eta_i||^2 \leqq \frac{B^2}{m}\mu(D)\,[\xi,\xi]\,||M||$$

and, therefore,

$$\Lambda_1(x+\xi)-\Lambda_1(x) = \sum_{i=1}^{n}\int_D \beta_i(t)\,\eta_i(t)\,dt \leqq \frac{nB}{\sqrt{m}}\sqrt{\mu(D)}\,\sqrt{[\xi,\xi]}\,\sqrt{||M||}. \tag{3.6}$$

Finally we note that by (2.4) and (1.11) $\Lambda_2 = [x, x]/2$ and consequently

$$\begin{aligned}|\Lambda_2(x+\xi)-\Lambda_2(x)| = |[\xi, x+\tfrac{1}{2}\xi]| &\leqq \sqrt{[\xi,\xi]}\,\sqrt{[x+\tfrac{1}{2}\xi, x+\tfrac{1}{2}\xi]} \\ &\leqq \sqrt{[\xi,\xi]}\,R.\end{aligned} \tag{3.7}$$

Formulas (3.4)—(3.7) obviously prove (3.3) with a suitable C_R.

b) Let x be an arbitrary element in H and R a positive number greater than $\sqrt{[x, x]}$. Then there exists a sequence of points x_n in the intersection $V_R \cap \tilde{H}$ such that $x_n \to x$ in H. Replacing, in (3.3), x by x_m and ξ by $x_n - x_m$ we see from this inequality that the $\Lambda(x_n)$ form a Cauchy sequence in H since the x_n form a Cauchy sequence. We define $\Lambda(x) = \lim_{n\to\infty}\Lambda(x_n)$.

We recall the definition of a Fréchet differential.

DEFINITION 3.1. The scalar $i(x)$ defined in H (or a subset of H) is said to have a Fréchet differential at the point x_0 if there exists a bounded linear functional in ξ, $l_0(\xi) = l(x_0, \xi)$ such that the following is true: If $R(x_0, \xi)$ is defined by the equation

$$i(x_0+\xi)-i(x_0) = l(x_0, \xi)+R(x_0, \xi), \tag{3.8}$$

then

$$\lim_{\xi\to 0}\frac{R(x_0, \xi)}{\sqrt{[\xi,\xi]}} = 0. \tag{3.9}$$

If $i(x)$ has a Fréchet differential for all x_0 of a subset H_1 of H and if the limit (3.9) is uniform in H_1, then $i(x)$ is said to have a uniform Fréchet differential in H_1.

We consider now the scalar $j(y)$ defined by (1.1) and (1.7) under the assumption that the coefficients a_{ik}, b_i, c satisfy the assumptions made in Lemma 3.1 for a_{ik}, β_i, γ respectively. The scalar $i(x)$ defined by (1.12) for $x \in \tilde{H}$ can then, by Lemma 3.1, be extended to H, and the extended scalar which we again denote by $i(x)$ satisfies a Lipschitz condition of the form (3.3). We will now prove the existence of the Fréchet differential di of $i(x)$ under the additional condition that the b_i have continuous first derivatives

and that $\partial c/\partial y$ exists and satisfies the assumptions made concerning γ in Lemma 3.1.

THEOREM 3.1. *Let the coefficients of* (1.7) *satisfy the assumptions just made. Let*

$$\tilde{E}(y) = -2\sum_{i,k=1}^{n} \frac{\partial}{\partial t_i}\left(a_{ik}\frac{\partial y}{\partial t_k}\right) - \sum_{i=1}^{n}\frac{\partial b_i}{\partial t_i} + \frac{\partial c}{\partial y}. \tag{3.10}$$

Moreover for $x \in \tilde{H}$, $\xi \in \tilde{H}$ *let*

$$E(x) = \tilde{E}(M(x)), \tag{3.11}$$

and

$$l(x_0, \xi) = \int_D E(x_0)\eta(t)\,dt, \qquad \eta = M(\xi). \tag{3.12}$$

Then for fixed $x_0 \in \tilde{H}$, $l(x_0, \xi)$ *is a bounded linear functional of* ξ *for* ξ *in* $\tilde{H}$ *and can therefore be extended to* H. *It can also be extended to* H *as a function of* x_0. *Moreover* $l(x_0, \xi)$ *is the Fréchet differential of* $i(x)$,

$$di(x_0, \xi) = l(x_0, \xi) = (E(x_0), \eta), \tag{3.13}$$

and is uniform (*Definition* 3.1) *in each sphere* V_R.

Proof: That for fixed $x_0 \in \tilde{H}$, $l(x_0, \xi)$ is linear in ξ is clear. To prove the boundedness we note first that $E(x_0) = \tilde{E}(M(x_0))$ is in L^2, and that therefore, by (3.12) and (2.19)

$$|l(x_0, \xi)| \leqq ||\eta||\,||E(x_0)|| \leqq \sqrt{||M||}\,||E(x_0)||\sqrt{[\xi, \xi]}. \tag{3.14}$$

Thus $l(x_0, \xi)$ is bounded (in ξ) and can be extended to H. To prove that $l(x_0, \xi)$ can also be extended as a function of x_0 we note first that, by (3.10), (3.11), (1.8) and (1.9),

$$E(x_0) = x_0 - \sum_{i=1}^{n}\frac{\partial b_i}{\partial t_i} + \frac{\partial c}{\partial y}, \qquad y = M(x_0) \tag{3.15}$$

and claim that E can be extended. This is obvious for the first term in (3.15), but also for the second since this term is independent of x_0. To prove that also the last term in (3.15) can be extended to H we set $c_1(t, y) = \partial c/\partial y$ and note that by assumption there exists a constant K such that

$$|c_1(t, y+\zeta) - c_1(t, y)| \leqq K|\zeta|. \tag{3.16}$$

Moreover $c_1(t, M(x))$ is, for $x \in \tilde{H}$, in L^2 and can therefore, by Lemma 2.5, be considered as element of H. Let now x_0 be an arbitrary element of H, and $\{x_n\}$ a sequence of elements of $\tilde{H}$ converging (in the H metric) to x_0. With the abbreviation $c_n = c_1(t, M(x_n))$ we have then from (2.19) and (3.16)

$$\begin{aligned}[c_n - c_m, c_n - c_m] &\leqq ||M||\,||c_n - c_m||^2 \leqq ||M||\,K^2\,||M(x^n - x^m)||^2 \\ &\leqq ||M||^2 K^2 [x_n - x_m, x_n - x_m].\end{aligned} \tag{3.17}$$

This shows that the $c_n = c_1(t, M(x^n))$ form a Cauchy sequence in H, and $c_1(t, x_0)$ can be defined in the obvious fashion.

To prove that the extended $l(x_0, \xi)$ is a bounded linear functional in ξ we note that, by (3.12), (3.15) and (1.11), the equality

$$l(x_0, \xi) = [E(x_0), \xi] \tag{3.18}$$

with

$$E(x_0) = x_0 - \sum_{i=1}^{n} \frac{\partial b_i}{\partial t_i} + \frac{\partial c}{\partial y} \tag{3.19}$$

holds, first for x_0 in $\tilde{H}$, then by continuity for all x_0 in H. But (3.18) sets in evidence that $l(x_0, \xi)$ is, for fixed x_0, a bounded linear functional in ξ.

We now prove that $l(x_0, \xi)$ is the Fréchet differential at x_0 of $i(x)$. We have from the definition (1.1), (1.12) and (1.7) for $x \in \tilde{H}$, $\xi \in \tilde{H}$

$$i(x+\xi) - i(x) = I_0 + I_1 + A_2(x+\xi) - A_2(x) \tag{3.20}$$

where

$$I_1 = \int_D \sum_{i=1}^{n} b_i \eta_i \, dt, \qquad I_0 = \int_D \big(c(t, y+\eta) - c(t, y)\big)\, dt, \tag{3.21}$$

$$y = M(x), \quad \eta = M(\xi), \quad \eta_i = \frac{\partial \eta}{\partial t_i},$$

while A_2 is defined in the beginning of the proof of Lemma 3.1. We see from (3.7) that

$$A_2(x+\xi) - A_2(x) = [\xi, x] + \tfrac{1}{2}[\xi, \xi]. \tag{3.22}$$

Moreover we obtain from (3.21)

$$I_1 = -\int_D \eta(t) \sum_{i=1}^{n} \frac{\partial b_i}{\partial t_i} \tag{3.23}$$

and

$$I_2 = \int_D \eta \frac{\partial c}{\partial y}\, dt + R_2 \tag{3.24}$$

where the arguments of $\partial c/\partial y$ are t and $y = M(x)$ while

$$R_2 = \int_D \eta(t) \left[\frac{\partial c(t, y^*)}{\partial y} - \frac{\partial c(t, y)}{\partial y}\right] dt, \qquad y^* = y + \vartheta\eta, \qquad 0 < \vartheta < 1$$

and therefore, by (3.16) and (2.19)

$$|R_2| \leqq K \, ||\eta||^2 \leqq K \, ||M||\, [\xi, \xi]. \tag{3.25}$$

(3.20)—(3.24) together with (3.12) and (3.19) show that $i(x_0+\xi)-i(x_0) = l(x_0, \xi)+R$ where $R = \frac{1}{2}[\xi, \xi]+R_2$ and therefore by (3.25)

$$|R| \leqq [\xi, \xi](\tfrac{1}{2}+K\,||M||)$$

which proves that l is the uniform Fréchet differential of $i(x)$.

THEOREM 3.2. *Under the assumptions of Theorem* 3.1 *the gradient* $g(x)$ *of* $i(x)$ *is of the form*

$$g(x) = x+G(x) \tag{3.26}$$

where $G(x)$ *is completely continuous.*

Proof: By definition of the gradient, (3.13) and (3.18) show that

$$g(x) = E(x); \tag{3.27}$$

therefore we see from (3.19) that

$$G(x) = g(x)-x = -\sum_{i=1}^{n} \frac{\partial b_i}{\partial t_i}+\frac{\partial c}{\partial y}. \tag{3.28}$$

Since the b_i do not depend on x it is sufficient to show that if V_R is the sphere defined by

$$[x, x] \leqq R^2, \tag{3.29}$$

then every sequence of points in V_R contains a subsequence $\{x_n\}$ for which (with the abbreviations used in (3.17)) the c_n as elements of H form a Cauchy sequence. To see this we note that by Theorem 2.1 the mapping M^2, i.e., M considered as the mapping $H \to L^2$ is completely continuous. Therefore every sequence of points in V_R contains a subsequence $\{x_n\}$ such that the $M(x_n)$ form a Cauchy sequence in L^2, i.e., such that $||M(x_n-x_m)|| = ||M(x_n)-M(x_m)|| \to 0$ as $m, n \to \infty$. But (3.17) shows then that the c_n form a Cauchy sequence in H.

THEOREM 3.3. *Under the assumptions of Theorem* 3.1 *the scalar* $i(x)$ *reaches a minimum in every sphere* (3.29).

Proof: If we write

$$i(x) = \tfrac{1}{2}[x, x]+I(x) \tag{3.30}$$

it follows from (2.4) and the definition of $i(x)$ that $I(x)$ is the extension of the integral

$$\int_D \sum_{i=1}^{n} [b_i(t)y_i+c(t, y)]\,dt, \qquad y = M(x),\ y_i = \frac{\partial y}{\partial t_i}.$$

Since x is the gradient of $\frac{1}{2}[x, x]$, comparison with (3.26) shows that $G(x)$ is the gradient of $I(x)$. Thus by Theorem 3.2, $I(x)$ has a completely continu-

ous gradient and this fact, as has been recalled in the introduction, implies the weak continuity[15] of $I(x)$. Since $[x, x]$ is weakly lower semi-continuous[16] it follows now that $i(x)$ is weakly lower semi-continuous and therefore reaches a minimum in the weakly compact sphere V_R.

COROLLARY TO THEOREM 3.3. *Let Δ be the intersection of $\mathfrak{B}_2$[17] with the range of $M(x)$ (x varies over H).*[18] *Then for every positive number R_1 there exists an element y_0 such that $j(y) \geqq j(y_0)$ for all y in Δ satisfying the inequality*

$$\int_D \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ik} \frac{\partial y}{\partial t_i} \frac{\partial y}{\partial t_k} dt \leqq R_1$$

(which is also satisfied by y_0).

The corollary is an immediate consequence of Theorem 3.3, Lemma 2.7 and the equalities (2.4) and (1.11).

THEOREM 3.4. *Sufficient for the existence of a critical point, i.e., a point x_0 satisfying* (1.4) *is that the coefficient c in* (1.7) *is non negative and that the assumptions of Theorem* 3.1 *are satisfied.*

Proof: The theorem will be proved if we can show that for some R the point x_0 of V_R which by Theorem 3.3 minimizes $i(x)$ in V_R is an interior point of this sphere. We will show that this is true for large enough R; to see this we note that for suitable m, M

$$(3.31) \qquad m \sum_{i=1}^{n} \int_D y_i^2 dt \leqq \int_D \sum_{i,k=1}^{n} a_{ik} y_i y_k dt \leqq M \sum_{i=1}^{n} \int_D y_i^2 dt, \qquad y_i = \frac{\partial y}{\partial t_i}.$$

Moreover we obtain from Schwarz's inequality

$$\left| \int_D \sum_{i=1}^{n} b_i y_i dt \right| \leqq B \sqrt{\sum_{i=1}^{n} \int_D y_i^2 dt}, \qquad B = \sqrt{\sum_{i=1}^{n} \int_D b_i^2 dt}.$$

From these inequalities, the assumption $c \geqq 0$ and the definition of $i(x)$ (cf. (1.1), (1.12) and (1.7)) we obtain the inequality

$$i(x) \geqq m \sum_{i=1}^{n} \int_D y_i^2 dt - B \left[\sum_{i=1}^{n} \int_D y_i^2 dt \right]^{1/2}.$$

Applying again (3.31) and noting that the middle member of this inequality

---

[15] Cf. footnote 2.

[16] Cf. footnote 3.

[17] See footnote 7 and Lemma 2.7.

[18] The range of $M(x)$ can be shown to consist of all functions y in L^2 to which there exists a sequence of elements y^m in L^2 which are 3 times differentiable and vanish on the boundary D and for which $||y^m - y|| \to 0$ as $m \to \infty$ while the $\partial y^m / \partial t_i$ form a Cauchy sequence in L^2. By Lemma 2.7 the range of $M(x)$ is contained in $\mathfrak{B}_2$.

equals $[x, x]/2$ (cf. (2.4)) we see that

$$i(x) \geqq \frac{1}{\sqrt{2M}} \sqrt{[x, x]} \left[\frac{m}{\sqrt{2M}} \sqrt{[x, x]} - B \right].$$

If the radius R of V_R is greater than $\sqrt{2M}$ and such that $mR/\sqrt{2M} - B > i(0)$, then $i(x) > i(0)$ for x on the boundary of V_R, which shows that the absolute minimum in V_R can not be taken on the boundary.

4. The Second Differential of $i(x)$

THEOREM 4.1. *In addition to the assumptions of Theorem* 3.1 *we assume that* $c_2(t, y) = \partial^2 c/\partial y^2$ *exists and satisfies the assumptions made about* γ *in Lemma* 3.1. *Then* $g(x)$ *has a Fréchet differential*

$$l_2(x, \xi) = \xi + L(x, \xi) \tag{4.1}$$

where L *is the differential of* G *and is given by the extension of*

$$\eta c_2(t, y), \qquad \eta = M(\xi), \quad y = M(x). \tag{4.2}$$

L *is completely continuous in* x *and* ξ.

Proof: Since the b_i do not depend on $y = M(x)$ we see from (3.28) that with $c_1(t, y) = \partial c/\partial y$

$$G(x+\xi) - G(x) = c_1(t, y+\eta) - c_1(t, y) = \eta c_2(t, y) + R \tag{4.3}$$

where

$$R = (c_2(t, y+\vartheta\eta) - c_2(t, y))\,\eta \qquad 0 < \vartheta < 1.$$

Since c_2 satisfies a Lipschitz condition in y there exists a constant K such that $|R| < K|\eta|^2$. We therefore see from (2.9) and (2.19) that

$$[R, R] \leqq ||M||\,(R, R) \leqq ||M||\,K^2\,||\eta||^2 \leqq M^2K^2[\xi, \xi]$$

which together with (4.3) proves that $L(x, \xi)$ is the Fréchet differential of G. It now follows immediately that (4.1) is the Fréchet differential of g since ξ is the differential of x. That L can be extended to H follows from the proof given for Lemma 3.1 since c_2 satisfies the assumptions made in this lemma concerning γ. Finally, the complete continuity of $L(x, \xi)$ in ξ follows from Theorem 2.1 while the proof, for the complete continuity in x, i.e., of $c_2(t, M(x))$, is identical with the proof for the complete continuity of c_1 given in Theorem 3.2.

COROLLARY TO THEOREM 4.1. *Under the assumptions of Theorem* 4.1, $i(x)$ *has a second differential*

$$d^2 i(x, \xi, \zeta) = [l_2(x, \xi), \zeta]. \tag{4.4}$$

The corollary is an immediate consequence of Theorem 4.1 and [10], Lemma 2.3.

THEOREM 4.2. *Under the assumptions of Theorem* 4.1 *let* x_0 *be a critical point of* $i(x)$, *i.e., a solution of* (1.4). *We assume moreover that the linear map* $l_2(x_0, \xi)$ in ξ *is non-singular.*[19] *Then the following statements hold:*

a) x_0 *is an isolated critical point;*

b) *for* $x = x_0$ *the bilinear form* (4.4) *in* ξ *and* ζ *is non-degenerate;*[20]

c) *there exists an ortho-normal base* e_ν *of* H, *a sequence of positive numbers* p_ν, *and a non negative finite integer* s *such that*

$$d^2 i(x_0, \xi, \xi) = -\sum_{\nu=1}^{s} p_\nu \xi_\nu^2 + \sum_{\nu=s+1}^{\infty} p_\nu \xi_\nu^2, \qquad \xi_\nu = [e_\nu, \xi].^{21}$$

Proof: On account of the complete continuity of G the statements follow from the discussion in [10], pp. 78—81.

THEOREM 4.3. *Let the assumptions of Theorem* 4.1 *be satisfied. Let* x_0 *be a critical point of* $i(x)$, *and* $l_2(x_0, \xi)$ *non-singular. Moreover we assume that* $c(y, t)$ *has derivatives with respect to* y *up to the fourth order and that all these derivatives satisfy the assumptions about* γ *in Lemma* 3.1. *Let* m^r *be the Morse type number of dimension* r *of the critical point* x_0.[22] *Then*

$$m^r = \begin{cases} 1 \text{ if } r = s, \\ 0 \text{ if } r \neq s, \end{cases} \qquad r = 0, 1, 2, \cdots. \tag{4.5}$$

Proof: Under certain assumptions (4.5) was proved in [11], Theorem 7.2. If one uses Theorem 7.1 of the same paper it is a matter of routine to check that these assumptions are satisfied in the present case. We therefore omit the details.

We finally note the following two connections a) and b) between the operator $g(x)$ and the operator $l_2(x, \xi)$ which are the analogues of well-known connections between the Euler and the Jacobi expression in the calculus of variations.[23]

THEOREM 4.4. *Let the assumptions of Theorem* 4.1 *be satisfied.* a) *Let* $x = x(\alpha)$ *be a solution of* (1.4) *depending on the real parameter* α *and let*

[19] I.e., $l_2(x_0, \xi) = 0$ only if $\xi = 0$.

[20] A symmetric bilinear form $B(\xi, \zeta)$ is non-degenerate if "$B(\xi, \zeta_0) = 0$ for all ξ" implies $\zeta_0 = 0$.

[21] If $s = 0$ the symbol $\sum_0^s$ is supposed to denote 0.

[22] If U is a small enough neighborhood of x_0, then for $r = 0, 1, 2$, the number m^r is defined as the r-th relative Betti number of the set $\{i(x) \leqq i(x_0)\} \cap U$ modulo the set $\{i(x) \leqq i(x_0)\} \cap U - x_0$. (Singular homology theory is used).

[23] See e.g., [1], p. 62 and p. 74.

$x(0) = x_0$. *We assume moreover that the derivative* $x'(\alpha)$ *of* x *with respect to* α *exists at* $\alpha = 0$. *Then*

$$l_2(x_0, x'(0)) = 0. \tag{4.6}$$

b) Furthermore,

$$l_2(x, \xi) = \tfrac{1}{2} \operatorname{grad}_\xi d^2i(x, \xi, \xi)$$

where the index ξ *indicates that the gradient operation is to be taken with respect to* ξ.

Proof: a) Since l_2 is the differential of g we have

$$0 = g(x(\alpha)) - g(x_0) = l_2(x_0, x(\alpha) - x_0) + r(x_0, x(\alpha) - x_0) \tag{4.7}$$

with

$$\lim_{x \to x_0} \frac{[r(x_0, x - x_0), r(x_0, x - x_0)]}{[x - x_0, x - x_0]} = 0. \tag{4.8}$$

On account of (4.8) and the fact that

$$\lim_{\alpha \to 0} \frac{x(\alpha) - x_0}{\alpha} = x'(0)$$

exists, it is now easy to see that (4.6) is obtained by dividing (4.7) by α and letting $\alpha \to 0$.

b) $d^2i(x, \xi, \xi)$ is a homogeneous form of degree 2 in ξ. Therefore we have by Euler's theorem on homogeneous forms[24]

$$2d^2i(x_0, \xi, \xi) = [\operatorname{grad}_\xi d^2i(x_0, \xi, \xi), \xi].$$

Comparison of this equality with (4.4) yields the assertion b).

Bibliography

[1] Bolza, O., *Vorlesungen uber Variationsrechnung*, Teubner, Leipzig und Berlin, 1909.

[2] Caccioppoli, R., *Limitazione integrali per le soluzioni di un'equazione lineare ellitica a derivate parziali*, Giorn. Mat Battaglini, Ser. 4, Vol. 80, 1951, pp. 186–212.

[3] Friedrichs, K. O., *Spektraltheorie halbbeschrankter Operatoren und Anwendung auf die Spektralzerlegung von Differentialoperatoren*, Math. Ann., Vol. 109, 1934, pp. 465–487.

[4] Hestenes, M. R., *Applications of the theory of quadratic forms in Hilbert space to the calculus of variations*, Pacific J. Math., Vol. 1, 1951, pp. 525–581.

[5] Morrey, C. B. Jr., *Multiple integral problems in the calculus of variations and related topics*, Univ. California Publ Math, N.S., Vol. 1, No. 1, 1943, pp. 1–130.

[6] Morrey, C. B. Jr., *Second order elliptic systems of differential equations*, Ann. of Math Studies No. 33, Princeton University Press, pp. 101–159.

[7] von Sz. Nagy, B., *Spektraldarstellung linearer Transformationen des Hilbertschen Raumes*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 5, Springer, Berlin, 1942.

[24] The proof of Euler's theorem in the Hilbert space case is the same as in the classical case of a function of a finite number of variables. For details see [9] Lemma 3.1.

[8] Rellich, F., *Ein Satz über mittlere Konvergenz*, Nachr. Akad. Wiss. Göttingen, Math.-Phys. Kl., 1930, pp. 30–35.
[9] Rothe, E. H., *Completely continuous scalars and variational methods*, Ann. of Math., Vol. 49, 1948, pp. 265–278.
[10] Rothe, E. H., *Critical points and gradient fields of scalars in Hilbert space*, Acta Math., Vol. 85, 1951, pp. 73–98.
[11] Rothe, E. H., *Leray-Schauder index and Morse type numbers in Hilbert space*, Ann. of Math., Vol. 55, 1952, pp. 433–467
[12] Rothe, E. H., *Correction to the paper "Leray-Schauder index and Morse type numbers in Hilbert space"*, Ann. of Math., Vol. 58, 1953, pp. 593–594.
[13] Rothe, E. H., *Gradient mappings*, Bull. Amer. Math. Soc., Vol. 59, 1953, pp. 5–19.
[14] Schauder, J., *Über lineare elliptische Differentialgleichungen zweiter Ordnung*, Math. Z., Vol. 38, 1934, pp. 257–282.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 569–576 (1956)

Mean Values and Continuity of Riesz Potentials*

KENNAN T. SMITH
University of Kansas

1. Introduction

This paper contains results of N. Aronszajn and the author concerning the determination of the values of certain types of functions at a point in terms of the values of the functions at neighboring points. An example will explain the problem.

If u is a function with finite Dirichlet integral over a domain D with boundary of class $C^{0,1}$, then u is known to have boundary values in various senses. One method of obtaining the boundary values is to extend u to a function with finite Dirichlet integral over a domain containing $\overline{D}$. The values of the extension, defined and well determined on the boundary except on a set of points of capacity 0, are the boundary values of u. Here, more explicit means of determining u at a point x_0 on the boundary directly in terms of the values of u inside D and near x_0 will be obtained.

The functions to which the theorems apply are those which belong to the best possible complete classes of admissible functions in eigenvalue and boundary value problems for linear elliptic differential operators. If D is a bounded domain in R^n with a smooth boundary whose closure is contained in an open sphere S, then a universal space containing the complete classes of admissible functions in problems of order $2m$ concerning D is the completion of the space $C_0^{2m}(S)$ of functions $2m$ times continuously differentiable and vanishing outside a compact subset of S with respect to the norm

$$||u||_m^2 = \sum_{|k| \leq m} \int \left| \frac{\partial^{|k|} u}{\partial_{x_1}^{k_1} \cdots \partial_{x_n}^{k_n}} \right|^2. \tag{1.1}$$

The complete space is called $P_0^m(S)$.[1] Some of its properties and a more detailed discussion of its role in differential problems can be found in the Proceedings of the Conference on Partial Differential Equations held at the University of Kansas during June and July, 1954 [9]. One important prop-

*The research for this paper was done under Contract N58304, Office of Naval Research.

[1]The actual space for problems concerning D is the space of restrictions to $\overline{D}$ of all functions in $P_0^m(S)$. This space is independent of S and can be obtained by a construction independent of S, namely by the completion of the space of functions $2m$ times differentiable on $\overline{D}$ with respect to the norm in (1.1), integration being taken over D only.

erty is that each function in $P_0^m(S)$ is differentiable k times in the ordinary sense except at a set of points of capacity 0 of order $2m-2k$, for any $k \leqq m$.[2]

Similar spaces (in which the functions are defined less precisely, outside sets of measure 0 rather than outside sets of $2m$-capacity 0) have been used in differential problems by a number of authors. Particular attention is called to the articles in these proceedings by P. Lax, F. Browder, A. N. Milgram, L. Nirenberg (where spaces H_m are used) and C. B. Morrey (where a space $\mathfrak{B}_2$ is used). Also related are the spaces of (B.L.)-functions considered by O. Nikodym [7] and (B.L.)-functions of order k considered by J. Deny [3].

2. Statement of the Problem

The subject of this paper is the determination of the value of a function $u \in P_0^m(S)$ at a point $x_0 \in S$ in terms of the values of u at points near x_0. When $2m > n$ there is no problem to be solved, since the functions in $P_0^m(S)$ are continuous. When $2m \leqq n$, however, the functions are not continuous; indeed, they are not defined everywhere—only up to sets of capacity 0 of order $2m$.

The determination of the value $u(x_0)$ is achieved in two ways:

a) by limits of mean values,

b) by continuity on the complement of a small set.

The theorems falling under a) generalize a very simple mean value theorem of O. Frostman [5]. The theorems falling under b) generalize similar theorems on the special case $m = 1$, due to N. Wiener [10], O. D. Kellogg and F. Vasilesco [6], M. Brelot [2], and J. Deny [3]. The basic result, which is the starting point, asserts that the functions in $P_0^m(S)$ are Riesz potentials.

If $u \in P_0^m(S)$, then there is a unique function g, square integrable over R^n, such that, except on a set of $2m$-capacity 0,

$$u(x) = K_m g(x) = \int K_m(x-y) g(y)\, dy. \tag{2.1}$$

The norms $||u||_m$ and $||g||_0 = ||g||_{L^2}$ are equivalent. The kernel is the kernel of order m of M. Riesz:

$$K_m(x) = \frac{\Gamma\left(\frac{n-m}{2}\right)}{\pi^{\frac{n}{2}} 2^m \Gamma\left(\frac{m}{2}\right)} |x|^{m-n}. \tag{2.2}$$

Each potential of the type (2.1) *which vanishes outside S belongs to $P_0^m(S)$.*

[2] The general theory of functional completion is developed in [1]. A complete account of the properties of the spaces $P_0^m(S)$ and their application to differential problems will be given in a forthcoming paper. The first to obtain similar differentiability properties for Newtonian potentials (which are a special case of the functions in $P_0^1(S)$) was G. C. Evans [4].

The restriction of a potential of type (2.1) to a hyperplane of dimension $n-k$ is a potential in the hyperplane of order $m-k/2$. Consequently, it is important to have theorems about the potentials for non-integral as well as integral values of m. In the proofs it is necessary to make use of potentials of measures and their properties. If $0 < 2m < n$ and μ is a positive Borel measure,

$$K_{2m}\mu(x) = \int K_{2m}(x-y)d\mu(y). \tag{2.3}$$

The measure μ is said to be of finite $2m$-energy if

$$||\mu||_{2m}^2 = \int K_{2m}\mu(x)d\mu(x) < \infty. \tag{2.4}$$

If μ is of finite $2m$-energy then, by the composition formula of M. Riesz [8], $g = K_m\mu$ is square integrable and

$$K_{2m}\mu(x) = K_m g(x) \text{ and } ||g||_{L^2} = ||\mu||_{2m}. \tag{2.5}$$

The theorems are proved in general by considering first potentials of the form $K_{2m}\mu$ and afterward potentials of the more general form $K_m g$. (Not every $K_m g$ is a linear combination of the $K_{2m}\mu$, and treatment of the former is essentially more difficult.) This plan of proof would appear to exclude the case $2m = n$. That it does not is due to the fact that a kernel can be defined to fill the place of the non-existent K_n. Using this kernel, which is equivalent locally to $\log 1/|x-y|$, there exists a theory of potential, of capacity, and of balayage quite similar to the well-known theory using Riesz kernels.[3] In the framework of this theory the statements below, and even their proofs, remain essentially correct.

3. The Mean Value Theorems

For each point x let $\{\varphi_x^k\}$ be a sequence of positive Borel measures such that the total mass of φ_x^k is $1 (|\varphi_x^k| = 1)$ and such that the support of φ_x^k is contained in a sphere $S(x, \varrho_k(x))$ with $\varrho_k(x) \to 0$. We are interested in the validity of the statements

$$K_{2m}\mu(x) = \lim_{k\to\infty} \int K_{2m}\mu(y)d\varphi_x^k(y), \tag{3.1}$$

$$K_m g(x) = \lim_{k\to\infty} \int K_m g(y)d\varphi_x^k(y). \tag{3.2}$$

In general, the potentials on the left sides of (3.1) and (3.2) are defined only up to a set of $2m$-capacity 0. If, however, the measure or the square integrable function g is non-negative, then the corresponding potential is defined everywhere, provided $+\infty$ is admitted as a value. In this case it makes sense

[3] The details will be given in the forthcoming paper mentioned above.

to ask whether (3.1) or (3.2) holds at an individual point x chosen in advance. Putting the question in this form underlines the difference between (3.1) and (3.2). Under appropriate hypotheses (3.1) holds at an individual point chosen in advance; under the same hypotheses (3.2) holds, in general, only outside a set of $2m$-capacity 0.[4]

THEOREM 1. *In order that* (3.1) *hold at a fixed point x for every positive measure μ it is necessary and sufficient that there exist a constant M (independent of k and y) such that*

$$K_{2m}\varphi_x^k(y) \leqq MK_{2m}(x-y) \text{ for all } y. \tag{3.3}$$

The necessity of (3.3) is proved by a simple explicit construction. Frostman [5] verified (3.3) in the case when φ_x^k is the normalized Lebesgue measure in the sphere $S(x, \varrho_k(x))$ and noticed its sufficiency. Although Frostman's special case has been very useful, much finer measures than the normalized Lebesgue measures satisfy (3.3); we shall see that sets which are comparatively rare at x can support measures which do.

If A is a set, $\gamma_{2m}(A)$ denotes its outer capacity of order $2m$, and μ_A denotes its capacitary distribution. The support of μ_A is contained in $\bar{A}$; if A is closed, $|\mu_A| = ||\mu_A||_{2m}^2 = \gamma_{2m}(A)$, $K_{2m}\mu_A(x) \geqq 1$ on A except on a set of γ_{2m}-capacity 0, $K_{2m}\mu_A(x) = 1$ for almost every μ_A, and there exists a constant M depending only on m such that $K_{2m}\mu_A(x) \leqq M$.

For each point x let $\{A_k(x)\}$ be a sequence of closed sets such that $A_k(x) \subset S(x, \varrho_k(x))$ with $\varrho_k(x) \to 0$. The problem is to see whether each $A_k(x)$ supports a measure φ_x^k such that (3.1) holds.

THEOREM 2. *If for a fixed point x there is a constant $c > 0$ (independent of k) such that*

$$\gamma_{2m}[A_k(x)] \geqq c\gamma_{2m}[S(x, \varrho_k(x))], \tag{3.4}$$

then each $A_k(x)$ supports a measure φ_x^k for which (3.1) *holds.*[5] *One choice for φ_x^k is the normalized capacitary distribution for $A_k(x)$.*

This theorem is a simple consequence of the fact mentioned above that there exists a uniform bound for all capacitary potentials. In view of the fact that the point mass at x (φ_x^k = point mass at x) satisfies (3.3), it is not possible to give a condition which is nontrivial and strictly necessary in Theorem 2. Condition (3.4) is essentially necessary, however. If (3.1) holds for the normalized capacitary distributions of the sets $A_k(x)$, and if the part

[4] Application of the theorems to potentials of measures or functions which are not non-negative is immediate. In the latter case the statement of each theorem is the same, in the former, exception must be made of a set of points of $2m$-capacity 0.

[5] If $0 \leqq k < 2m$ and if Π is a hyperplane of dimension $n-k$ passing through x, then $\gamma_{2m}[S(x, \varrho) \cap \Pi]$ is of the order of ϱ^{n-2m}. Thus $A_k(x) = S(x, \varrho_k(x)) \cap \Pi$ satisfies (3.4). This gives an indication of how small the sets $A_k(x)$ can be.

of $A_k(x)$ outside $S(x, \frac{1}{2}\varrho_k(x))$ has non-zero $2m$-capacity, then (3.4) holds. More generally, if (3.1) holds for measures φ_x^k supported by the sets $A_k(x)$, and if there is an $\varepsilon > 0$ such that the φ_x^k measure of the part of $A_k(x)$ outside $S(x, \frac{1}{2}\varrho_k(x))$ is $\geqq \varepsilon|\varphi_x^k| = \varepsilon$, then (3.4) holds. What is needed is a supplementary condition to ensure that the φ_x^k do not concentrate too quickly to the point mass at x.

Now we turn our attention to (3.2) and give one of our principal theorems.

Observe first that if the φ_x^k are the normalized capacitary distributions of sets $A_k(x)$ satisfying (3.4), then

$$K_{2m}\varphi_x^k(y) \leqq \frac{M(x)}{\gamma_{2m}[S(x, \varrho_k(x))]} \tag{3.5}$$

where $M(x)$ is a bound independent of k. This follows from the existence of a uniform bound for all capacitary potentials. Observe also that (3.5) implies (3.3).

THEOREM 3. *If $2m \leqq n$, and if the φ_x^k satisfy* (3.5), *then* (3.2) *holds for any square integrable g except on a set of γ_{2m}-capacity* 0.

We assume that $K_m g$ is finite for at least one x. This is always true if $2m < n$; if $2m = n$ it is true if and only if $(1+|x|)^{-n/2}g(x)$ is integrable.

Proof: It is sufficient to prove the theorem when x is interior to some fixed sphere S.

Let μ be the positive measure of finite $2m$-energy which is the result of sweeping $K_m g$ on S. Then the support of μ is contained in S, and $K_{2m}\mu(x) \geqq K_m g(x)$ on S except for a set of γ_{2m}-capacity 0. For each $\varepsilon > 0$ choose an open set B_ε such that $\gamma_{2m}(B_\varepsilon) < \varepsilon$, and, on $S - B_\varepsilon$, $K_{2m}\mu(x)$ and $K_m g(x)$ are uniformly continuous. It will be shown that (3.2) holds at any point x_0 at which the capacitary potential for one of the sets B_ε is $\leqq \frac{1}{2}$. Since the capacity of the set where the capacitary potential for B_ε exceeds $\frac{1}{2}$ is at most 4ε, this will prove the theorem.

We fix x_0 and ε with the property indicated, and for simplicity of notation omit henceforth all subscripts, etc., which refer to x_0.

Since no measures with supports in the sets $S(x_0, \varrho) \cap B_\varepsilon$ have the property (3.1) as $\varrho \to 0$,[6] it follows from Theorem 2 that

$$\lim_{\varrho \to 0} \frac{\gamma_{2m}[S(x_0, \varrho) \cap B_\varepsilon]}{\gamma_{2m}[S(x_0, \varrho)]} = 0. \tag{3.6}$$

If θ^k denotes the restriction of φ^k to B_ε, if $\nu^k = \varphi^k - \theta^k$ and if β^k is the capacitary distribution for $S(x_0, \varrho_k(x_0)) \cap B_\varepsilon$, then from (3.5) and from the fact that $K_{2m}\beta^k(x) \geqq 1$ on $S(x_0, \varrho_k(x_0)) \cap B_\varepsilon$ we obtain

[6] Relation (3.1) does not hold for $K_{2m}\mu_B$.

$$|\theta^k| \leqq \int K_{2m}\beta^k(x)\,d\theta^k \leqq \int K_{2m}\varphi^k\,d\beta^k \leqq M\frac{\gamma_{2m}[S(x_0,\varrho_k)\cap B_\varepsilon]}{\gamma_{2m}[S(x_0,\varrho_k)]},$$

so that from (3.6) it follows that $|\theta^k| \to 0$.

By using the uniform continuity of $K_{2m}\mu$ and $K_m g$ on a set containing the support of each ν^k and the fact that $|\nu^k| = |\varphi^k| - |\theta^k|$ converges to 1, we can prove easily that (3.1) and (3.2) hold at the point x_0 for the measures ν^k. On the other hand, by Theorem 1, (3.1) holds at x_0 for the measures φ^k. Therefore,

$$\int K_m g(x)\,d\theta^k \leqq \int K_{2m}\mu(x)\,d\theta^k \to 0,$$

and the proof is complete.

4. The Continuity Theorems

In Section 3 an "in the large" continuity theorem was used; namely, that for any potential $K_m g$ and any $\varepsilon > 0$ there exists a set B such that $\gamma_{2m}(B) < \varepsilon$ and, on $R^n - B$, $K_m g$ is continuous. In this section the nature of the sets whose removal ensures continuity at a single point will be discussed.

A set A is said to be $2m$-thin at a point x_0 if there exists a positive measure μ such that[7]

$$K_{2m}\mu(x_0) < \liminf_{\substack{x\in A\\ x\to x_0}} K_{2m}\mu(x). \tag{4.1}$$

The following theorem was proved by Deny [3] in the case $m = 1$. The proof (which is straightforward) is the same in the general case.

THEOREM 4. *For each potential $K_m g$ and each x except in a set of γ_{2m}-capacity 0 there exists a set $A(x)$, $2m$-thin at x, such that the restriction of $K_m g$ to $R^n - A(x)$ is continuous at x.*

By an argument of Frostman [5] it can be proved that the process of balayage takes continuous functions into continuous functions in certain cases.

THEOREM 5. *If A is a closed set which is $2m$-thin at none of its points, and if $K_m g$ is a potential which is continuous on A, then the result of sweeping $K_m g$ on A is continuous on R^n.*

Theorem 4 is the best general theorem on continuity at a point which can be expected. It is useless, however, without a geometric characterization of the property of being thin at a point. For $m = 1$ a geometric characterization has been obtained by Wiener [10] (see also Brelot[2]), in the form of the convergence or divergence of a series, and by Kellogg and Vasilesco [6] in the

[7] The original definition for $m = 1$ was given by M. Brelot [2].

form of the convergence or divergence of an integral. The main result of this section is that the theorems of Wiener, and Kellogg and Vasilesco hold for all m. The theorem of Kellogg and Vasilesco is an easy consequence of the theorem of Wiener. The proof of half of the theorem of Wiener, the half which asserts that if the series converges, the set is thin, is the same as the proof for $m = 1$. The proof of the other half requires a new line of argument, for, the basis of the proof when $m = 1$ is the maximum principle, which is not valid in general. The argument will be sketched.

If x_0 and A are a given point and set, respectively, and if ϱ is a given number > 1, let S_k denote the set of points x such that $\varrho^k \leqq K_{2m}(x-x_0) < \varrho^{k+1}$, and let $A_k = A \cap S_k$.

THEOREM 6. *In order that A be $2m$-thin at x_0 it is necessary and sufficient that any of the conditions below be satisfied:*

a) *There is a μ of finite $2m$-energy such that $K_{2m}\mu(x_0) < \infty$ while* $\lim_{\substack{x \in A \\ x \to x_0}} K_{2m}\mu(x) = \infty$.

b) $\sum_{k=1}^{\infty} \varrho^k \gamma_{2m}(A_k)$ *converges for some* $\varrho > 1$.

c) $\sum_{k=1}^{\infty} \varrho^k \gamma_{2m}(A_k)$ *converges for all* $\varrho > 1$.

d) $\int_0 \gamma_{2m}[S(x_0, \varrho) \cap A]\, \varrho^{2m-1-n}\, d\varrho < \infty$.

Proof: To simplify the expressions it is assumed that the constant is dropped from K_{2m}, i.e., that $K_{2m}(x) = |x|^{2m-n}$. It is assumed also that the part of the theorem asserting that b) implies a) is known.

Suppose that A is $2m$-thin at x_0 and open. (The latter is no restriction, since by the lower semi-continuity of potentials a larger set than A must be $2m$-thin at x_0 and open.) Let ν be a measure such that (4.1) holds, and let ξ_0 be such that $K_{2m}\nu(x_0) < \xi_0 < \liminf_{\substack{x \in A \\ x \to x_0}} K_{2m}\nu(x)$. Choose $\theta < 1$ so that $\xi_1 = \xi_0 - \theta^{2m-n} K_{2m}\nu(x_0) > 0$. We shall prove that b) holds for any ϱ which is so large that $\theta < \varrho^{1/(n-2m)} - 1$, $0 < 1 - (1/\varrho)^{1/(n-2m)}$.

For such ϱ and for $x \in S_k$ and $y \notin S_{k-1} \cup S_k \cup S_{k+1}$, $K_{2m}(y-x) \leqq \theta^{2m-n} K_{2m}(y-x_0)$. Therefore, if ν_k denotes the restriction of ν to $S_{k-1} \cup S_k \cup S_{k+1}$, and if $\varphi_k = \nu - \nu_k$, then for $x \in S_k$, $K_{2m}\varphi_k(x) \leqq \theta^{2m-n} K_{2m}\varphi_k(x_0) \leqq \theta^{2m-n} K_{2m}\nu(x_0)$. It follows that if k is large and $x \in A \cap S_k$, then $K_{2m}\nu_k(x) \geqq \xi_1$. Thus, if k is large and C is a compact subset of $A_k = A \cap S_k$, then $\gamma_{2m}(C) = |\mu_C| \leqq (1/\xi_1) \int K_{2m}\nu_k\, d\mu_C = (1/\xi_1) \int K_{2m}\mu_C\, d\nu_k \leqq (M/\xi_1)\,|\nu_k|$, where M is the uniform bound for all capacitary potentials. Since A is open, A_k is capacitable, and it follows that $\gamma_{2m}(A_k) \leqq (M/\xi_1)\,|\nu_k|$ for large k. From the

finiteness of $K_{2m}(x_0)$ it can be proved easily that $\sum_{k=1}^{\infty} \varrho^k |\nu_k|$ converges, and b) is established.

By using the part of the theorem which is assumed known, a) is established also. If ν is replaced by a measure for which a) holds, then θ can be taken arbitrarily small. Hence ϱ can be taken arbitarily close to 1, and c) is established.

Bibliography

[1] Aronszajn, N., and Smith, K. T., *Functional spaces and functional completion*, Technical Report 10, University of Kansas, 1954, Ann. Inst. Fourier Grenoble (to appear 1956).

[2] Brelot, M., *Points irréguliers et transformations continues en théorie du potentiel*, J. Math. Pures Appl. Vol. 19, 1940, pp. 319–337.

[3] Deny, J., *Les potentiels d'énergie finie*, Acta Math., Vol. 82, 1950, pp. 107–183.

[4] Evans, G. C., *On potentials of positive mass*, Trans. Amer. Math. Soc., Vol. 37, 1935, pp. 226–253.

[5] Frostman, O., *Potentiels d'équilibre et capacité des ensembles*, Thesis, Lund, 1935

[6] Kellogg, O. D., and Vasilesco, F., *A contribution to the theory of the capacity*, Amer. J. Math , Vol. 51, 1929, pp. 515–526.

[7] Nikodym, O., *Sur une classe de fonctions considérées dans l'étude du problème de Dirichlet*, Fund. Math., Vol. 21, 1933, pp. 129–150.

[8] Riesz, M., *Intégrales de Riemann-Liouville et potentiels*, Acta Univ. Szeged Sect. Sci. Math., Vol 9, 1938, pp 1-42

[9] Smith, K. T., *Functional spaces, functional completion, and differential problems*, Conference on Partial Differential Equations, University of Kansas, Summer 1954, Technical Report 14, Office of Naval Research.

[10] Wiener, N., *The Dirichlet problem*, J. Math. Physics, Vol. 3, No. 3, 1924, pp. 127–146.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 577–595 (1956)

On Radiation Conditions *

J. J. STOKER

1. Introduction

In linear wave-propagation problems which are time-independent, either by virtue of the fact that the motion considered is a steady motion or that a simple harmonic time dependence has been factored out, and which take place in unbounded domains, it is in general not possible to characterize uniquely the solutions having the desired physical characteristics by imposing only boundedness conditions at infinity. It is, in fact, necessary to impose sharper conditions in most cases. In the simplest case in which the medium is such as to include a full neighborhood of the point at infinity which is in addition made up of homogeneous matter, the correct radiation condition for simple harmonic waves is not difficult to guess. It is simply that the wave at infinity behaves like an outgoing spherical wave from an oscillatory point source, and such a condition is what is commonly called the radiation, or Sommerfeld, condition. Among other things this condition precludes the possibility that there might be an incoming wave generated at infinity—which, if not ruled out, would make a unique solution of the steady-state problem impossible.

If the refracting or reflecting obstacles to the propagation of waves happen to extend to infinity—for example, if a rigid reflecting wall should happen to go to infinity—it is by no means clear *a priori* what conditions should be imposed at infinity in order to ensure the uniqueness of a solution having appropriate properties otherwise. One point of view to be presented here is that the difficulty arises because the steady-state problem is an unnatural problem in mechanics and that, in principle at least, one should rather formulate and solve an appropriate initial-value problem and then find the solution of the time-independent problem by making a passage to the limit in allowing the time to tend to infinity.[1]

The time-independent problems are unnatural because one makes a hypothesis about the motion that holds for all time, while Newtonian me-

*This paper represents results obtained under Contract N6 ori-201, Task Order No. 1, sponsored by the Office of Naval Research.

[1]The formulation of the usual radiation condition is doubtlessly motivated by an instinctive consideration of the same sort of hypothesis combined with the feeling that a homogeneous medium at infinity will have no power to reflect anything back to the finite region.

chanics is basically concerned with the prediction—in a unique way, furthermore—of the motion of a mechanical system from given initial conditions. Of course, in mechanics of continua that are unbounded it is necessary to impose conditions at infinity (in the space variables, that is) not derivable directly from Newton's laws, but for the initial-value problem it should suffice in general to impose only boundedness conditions at infinity.

The desirability, on general grounds, of attacking wave propagation problems in principle as initial value problems is fairly obvious. However, it is clear that such a program may not always be feasible, or even if feasible, still quite a bit more difficult to carry out than it would be to solve the corresponding time-independent problem. Two interesting special cases will be discussed in Sections 2 and 3 below in which the program can be carried out without difficulty. In the problem of Section 3, in addition, which refers to waves created on a running stream, it will be seen that new insights into the physical phenomena are gained which could not be obtained if the time-independent problem alone were considered.

There are many problems in which little doubt exists regarding the appropriate conditions to be imposed at ∞ in time-independent prolems. Perhaps the essential point in mind in using the word "appropriate" is that conditions leading to a unique solution should be found. This is, however, not always a simple matter in time-independent problems, as is well known. For example, if eigenfunctions exist—which may well be the case in some types of unbounded domains—it is clear that unique solutions of radiation problems would not be determined by conditions of the type of Sommerfeld's radiation condition. (This would remain true in such cases also if the approach recommended above were to be used, i.e., if the time-independent solution were to be sought as the limit as $t \to \infty$ of the solution of an appropriate initial-value problem, since free vibrations would occur and would have amplitudes fixed by the initial conditions, and no limit independent of these conditions could be expected.) Thus uniqueness theorems for time-independent radiation problems are decidedly interesting. In Section 4 a recently formulated uniqueness theorem for a very old problem, i.e. the problem of diffraction around a slit first solved more than fifty years ago by Sommerfeld, is given; at the same time a new and quite elementary solution is given for the problem which was motivated by the insight furnished by the uniqueness theorem.

2. Oscillatory Waves in an Infinite Ocean

The problem in question concerns the propagation of surface-gravity waves in an infinite ocean. We consider only irrotational incompressible

flow and assume further that the motion is a small oscillation in the neighborhood of the rest position of equilibrium; a linearized theory can then be obtained. We restrict ourselves to two-dimensional motion in an x, y-plane. with the y-axis taken vertically upward and the x-axis in the originally undisturbed horizontal free surface. Under these assumptions there exists a velocity potential $\phi(x, y, t)$ which is a harmonic function in the lower half-plane:

$$\phi_{xx}+\phi_{yy} = 0, \qquad y < 0, \quad t > 0. \tag{2.1}$$

(Here and in what follows subscripts denote partial derivatives.) The free-surface boundary conditions (cf. Lamb, *Hydrodynamics*, p. 364) are

$$-\phi_y+\eta_t = 0, \tag{2.2}$$

$$\phi_t+g\eta = -\frac{1}{\varrho}p, \qquad \text{for } y = 0, \quad t > 0. \tag{2.3}$$

In these equations $\eta = \eta(x, t)$ represents the vertical displacement of the free surface measured from the x-axis, and $p = p(x, t)$ represents the pressure applied on the free surface. We suppose that ϕ and its first and second derivatives tend to zero at ∞ for any given time t—in fact that they tend to zero in such a way that Fourier transforms exist—but we do not, in accordance with our discussion above, make any assumptions about the behavior of our functions as $t \to \infty$. At the time $t = 0$ we prescribe the following initial conditions:

$$\phi(x, 0, 0) = \phi_t(x, 0, 0) = p(x, 0) = 0 \tag{2.4}$$

which imply that the free surface is initially at rest in its horizontal equilibrium position.

In what follows we consider only the special case in which the surface pressure $p(x, t)$ is given by

$$p(x, t) = \delta(x)e^{i\omega t}, \qquad t > 0. \tag{2.5}$$

Here $\delta(x)$ is the Dirac δ-function; in other words, an oscillatory pressure point is applied at the origin to the free surface and maintained there for all time. By inserting this expression for p in (2.3) and eliminating the quantity η by use of (2.2) the free-surface condition is obtained in the form

$$g\phi_y+\phi_{tt} = -\frac{i\omega}{\varrho}\delta(x)e^{i\omega t}, \qquad t > 0. \tag{2.6}$$

Our problem now consists in finding a solution $\phi(x, y, t)$ of (2.1) which behaves properly at ∞ and which satisfies the free-surface condition (2.6) and

the initial conditions (2.4).[2] This problem is readily solved by standard methods, e.g. by using Fourier transforms.[3]

In the author's paper[4] the solution of the problem is given in the following form

$$(2.7) \qquad \phi(x, y, t) = -\frac{i\omega}{\varrho\pi}\int_0^\infty e^{sy}\cos sx \int_0^t \frac{e^{i\omega(t-\tau)}}{\sqrt{gs}}\sin(\sqrt{gs}\,\tau)d\tau\, ds.$$

This solution is also seen to be the unique solution of our problem. Our object is to study the behavior of this solution as $t \to \infty$.

Since y is negative (we do not discuss here the limit as $y \to 0$, that is, the behavior on the free surface), the integral with respect to s converges well and there is no singularity on the positive real axis of the complex s-plane. However, the passage to the limit $t \to \infty$ is more readily carried out by writing the solution in a different form in which a singularity—a pole, in fact—then appears on the real axis of the s-plane. (It seems, indeed, likely that such an occurrence would be the rule in any considerations of the present kind since the limit function as $t \to \infty$ would not usually be a function having a Fourier transform and one could expect that it would somehow appear as a contribution in the form of a residue at a pole.) It is convenient—and of course legitimate by Cauchy's theorem—to replace the path of integration along the s-axis in the s-plane by a path L indicated in Figure 1. The path

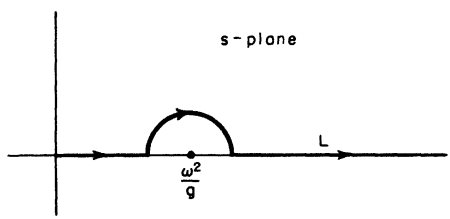


Figure 1.

[2]The steady-state problem is formulated as follows: One sets $\phi = \psi(x, y)e^{i\omega t}$. The time factor $e^{i\omega t}$ can be canceled from (2.1) and (2.6), and the harmonic function then satisfies the free-surface condition $g\psi_y - \omega^2\psi = -\frac{i\omega}{\varrho}\delta(x)$. In addition one would postulate that the motion should have the character of an outgoing progressing wave at $x = -\infty$ and $x = +\infty$. This problem was solved long ago by H. Lamb, *On deep water waves*, Proc. London Math. Soc., Ser. 2, Vol. 2, 1904, pp. 371–400 (see pp. 388–389). It is a curious fact that this problem, which involves the determination of a *harmonic* function and not a solution of a wave equation, nevertheless has not a unique solution of the desired type if only boundedness conditions are imposed at ∞

[3]Solutions of problems of this type can be found in the book by I. N. Sneddon, *Fourier Transforms*, McGraw-Hill, New York, 1951, p. 278.

[4]Stoker, J. J., *Some remarks on radiation conditions*, in *Proceedings of Symposia in Applied Mathematics*, Vol. 5, *Wave Motion and Vibration Theory*, McGraw-Hill, New York, 1954, pp. 97–102.

L lies on the positive real axis except for a semicircle in the upper half-plane centered at the point $s = \omega^2/g$.

We now replace $\sin \sqrt{gs}\,\tau$ in (2.7) by exponentials and carry out the integration with respect to τ to obtain

$$\phi(x, y, t) = -\frac{i\omega e^{i\omega t}}{\pi\varrho}\int_L e^{sy}\cos sx \cdot\left[-\frac{1}{2\sqrt{gs}}\frac{e^{i(\sqrt{gs}-\omega)t}}{\sqrt{gs}-\omega} - \frac{1}{2\sqrt{gs}}\frac{e^{-i(\sqrt{gs}+\omega)t}}{\sqrt{gs}+\omega} + \frac{1}{gs-\omega^2}\right] ds. \tag{2.8}$$

We wish now to consider the three items in the bracket separately, and as we see, two of them do indeed have a singularity at $s = \omega^2/g$ which is bypassed through our choice of the path L. The first two items are obviously the result of the initial conditions and hence could be expected to provide transients which die out as $t \to \infty$. This is in fact the case, as can be seen easily in the following way: That branch of $\sqrt{s}$ is taken which is positive on the positive real axis, and we operate always in the right half-plane. If, in addition, s is in the upper half-plane, it follows that $i(\sqrt{gs}\pm\omega)$ has its real part negative (ω being real). Consider now the contribution furnished by the first item in the square brackets. Since the exponential has a negative real part on the semicircular portion of the path L, it is clear that as $t \to +\infty$ this part of the path makes a contribution which tends to zero as $t \to \infty$. The remaining portions of L, which lie on the real axis, make contributions which die out like $1/t$: this can be seen easily by integration by parts, for example, or by application of known results about Fourier transforms. The second item in the square brackets has no singularity on the real axis, so that the path L can be taken entirely on the real axis; thus, in accordance with the remarks just made concerning the similar situation for the first item, it is clear that this contribution also dies out like $1/t$. Thus for large t we obtain the following asymptotic representation for ϕ:

$$\phi(x, y, t) \simeq -\frac{i\omega}{\pi\varrho}e^{i\omega t}\int_L \frac{e^{sy}\cos sx}{gs-\omega^2}ds. \tag{2.9}$$

Actually, the right-hand side is the solution of the time-independent problem formulated in a footnote above—as obtained, for example, in the paper of Lamb cited previously—when the condition at ∞ is the radiation condition stating that ϕ behaves like an outgoing progressing wave. The solution furnished by (2.9) is a little more awkward to obtain directly through use of the radiation condition—at least by methods known to the author—than obtaining solution (2.7) of the initial-value problem. In particular, the asymptotic behavior of an integral representation must be investigated in this case also before the radiation condition can be used. Thus

we have verified in this special case that the radiation condition can be replaced by boundedness conditions (in the space variables, that is) if one treats an appropriate initial-value problem and afterwards allows t to tend to ∞.

Even though not strictly necessary—since (2.9) is known to furnish the desired time-independent solution—it is perhaps of interest to show directly that the right-hand side of (2.9) has the behavior one expects for an outgoing progressing wave when $x \to +\infty$. The procedure is the same as that used above in discussing (2.8): The factor cos sx is replaced by exponentials to obtain

$$\phi(x, y, t) \simeq \frac{\omega}{\varrho} e^{i\omega t} \left(\frac{1}{2\pi i} \int_L \frac{e^{sy} e^{isx}}{gs - \omega^2} ds + \frac{1}{2\pi i} \int_L \frac{e^{sy} e^{-isx}}{gs - \omega^2} ds \right). \tag{2.10}$$

By the same argument as above one sees that the first integral makes a contribution that tends to zero as $x \to +\infty$. The second integral is treated by deforming the path L over the pole $s = \omega^2/g$ into a path M which consists of the positive real axis except for a semicircle in the lower half-plane. The contribution of the second integral then consists of the residue at the pole plus the integral over the path M. But the contribution of the latter integral is, once more, seen to tend to zero as $x \to +\infty$ because of the factor e^{-isx}. Thus $\phi(x, y, t)$ behaves for large x as follows:

$$\phi(x, y, t) \simeq -\frac{\omega}{g\varrho} e^{(\omega^2/g)y} e^{-i[(\omega^2/g)x - \omega t]}. \tag{2.11}$$

This represents a progressing wave in the positive x-direction which, in addition, has the wavelength $2\pi g/\omega^2$ appropriate to a progressing sine wave with the frequency ω in water of infinite depth.

3. Waves on a Running Stream

When a disturbance is initiated at the surface of a stream flowing with uniform velocity over a horizontal bed—by immersing a solid object in the stream, for example—the effect is in general the creation of waves on the surface of the water. The purpose of the discussion to follow is to investigate such occurrences mathematically in terms of the classical hydrodynamical theory of irrotational flow in an incompressible perfect fluid.

Special problems of this type have, of course, been studied; in the book of Lamb (p. 406), for example, the case of a steady flow past an obstacle is treated in considerable detail. This special case of a steady flow of a stream past an obstacle involves a number of pecularities, not to say obscurities, that deserve study and analysis. Lamb, following Rayleigh, postulates the existence of friction forces of an unnatural and artificial type from the point

of view of mechanics (which, however, do not destroy the existence of an irrotational flow) that have the effect of making the solution of the steady state problem unique without postulating any but boundedness conditions at ∞. It would seem more reasonable to achieve this end simply by postulating the behavior at ∞ in terms of conditions of the type which are now usually called radiation conditions. However, in this instance it is not entirely obvious what these conditions should be—for example, it is not sufficient to postulate that the disturbance should yield a wave having the character of an outgoing progressing wave at ∞; instead, it is necessary to ensure the uniqueness of the solution and, in some cases, to require that the amplitude of the disturbance dies out on going to infinity on the upstream side. The phrase "in some cases" is used deliberately, since there are also cases in which the solution is uniquely determined by boundedness conditions at ∞ alone. In fact, the steady state problem contains an essential dimensionless parameter, the ratio gh/U^2 (with h the undisturbed depth, U the undisturbed velocity of the liquid), and the character of the motion changes radically when this parameter passes through the "critical value" unity: For $gh/U^2 < 1$ no radiation condition at ∞ is necessary, while for $gh/U^2 > 1$ such a condition must be imposed. At the critical value itself, i.e. for $gh/U^2 = 1$,the representation of the steady state solution as given by Lamb yields an unbounded amplitude in the waves at ∞.

The point of view taken here is that these curious features of the steady state solution can all be understood, without introducing mechanically artificial forces and without having to make guesses about the conditions to be imposed at ∞, simply by studying the manner in which the motion builds up in the time starting at the moment when the disturbance is initiated in the uniform stream. In other words, the steady motion is to be found by studying what happens in an appropriately formulated initial value problem when the time tends to infinity.[5] It turns out, in fact, that as $t \to \infty$ the wave amplitude dies out both upstream and downstream if $gh/U^2 < 1$, and that it dies out upstream but not in general downstream if $gh/U^2 > 1$. In other words, the conditions at ∞ which we know to be necessary for the uniqueness of the steady state solution come out automatically. In addition, a new insight is gained in the transition case when $gh/U^2 = 1$; here, it is found, the steady state does not exist in the sense that the amplitude of the disturbance tends to become ∞ everywhere as $t \to \infty$.

We consider irrotational flows of an incompressible perfect fluid under gravity and having a free surface. Only two-dimensional flows are treated.

---

[5]The details are carried out in a paper by the author: Stoker, J. J., *Unsteady waves on a running stream*, Comm. Pure Appl Math., Vol. 6, 1953, pp. 471–482.

Figure 2 indicates the general situation: The x-axis is taken along the undisturbed free surface of a flow with uniform velocity U in the positive x-direction in water of undisturbed depth h, and the y-axis is taken positive upward. There is a rigid bottom surface at $y = -h$. It follows that a velocity potential $\Phi(x, y, t)$ exists and that it is a harmonic function of (x, y)

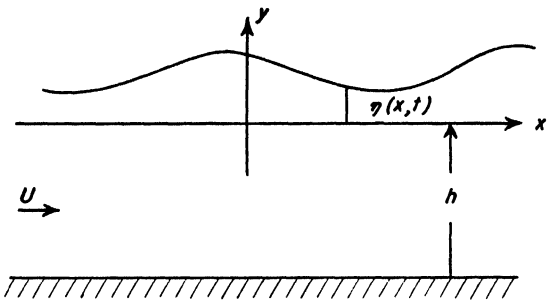


Figure 2. Stream with a free surface.

throughout the interior of the fluid. At the free surface $y = \eta(x, t)$ the boundary conditions are the dynamical condition (from Bernoulli's law)

$$\frac{p}{\varrho} + g\eta + \Phi_t + \frac{1}{2}(\Phi_x^2 + \Phi_y^2) = 0, \tag{3.1}$$

and the kinematic condition (stating that a particle on the free surface always remains on it)

$$\eta_t + \Phi_x \eta_x - \Phi_y = 0. \tag{3.2}$$

At the bottom $y = -h$ the boundary condition is, of course,

$$\Phi_y = 0. \tag{3.3}$$

We study here a linearized version of the problem in which only small disturbances of the uniform flow are assumed to occur. It is then natural to set

$$\Phi(x, y, t) = Ux + \phi(x, y, t) \tag{3.4}$$

and to regard $\phi(x, y, t)$ as the potential of a small disturbance considering it and its derivatives to be small of the same order. It is clear that $\phi(x, y, t)$ is a harmonic function. At the same time the free surface elevation $\eta(x, t)$ and its derivatives are also considered to be small of this same order. The velocity U of the undisturbed stream is, however, to be considered as a finite and not as a small quantity of the order of ϕ. Insertion of (3.4) in (3.1) and (3.2) yields the boundary conditions

$$\frac{p}{\varrho}+g\eta+\phi_t+U\phi_x = 0, \qquad \text{at } y = 0, \tag{3.5}$$

$$\eta_t+U\eta_x-\phi_y = 0, \tag{3.6}$$

when all but the lowest order terms in ϕ and η are rejected and an unessential constant in (3.5) is omitted. At the same time, these conditions are now to be satisfied at $y = 0$ rather than at the disturbed position of the free surface —this stipulation is also consistent with the linearization. The prescribed surface pressure $p = p(x, t)$ must evidently be assumed to be of the same order as ϕ and η. At the bottom the condition

$$\phi_y = 0, \qquad y = -h \tag{3.7}$$

must hold.

At the initial instant $t = 0$ we suppose the flow to be the undisturbed steady flow given by $\Phi = Ux$, $\eta = 0$, which is clearly a possible motion compatible with the full nonlinear theory. This leads to the condition

$$\phi(x, y, 0) = \eta(x, 0) = p(x, 0) = 0 \tag{3.8}$$

at the time $t = 0$. As a consequence we have from (3.5) the condition

$$\phi_t(x, y, 0) = 0. \tag{3.9}$$

Once the surface pressure $p(x, t)$ is given, the problem is completely formulated: A velocity potential $\phi(x, y, t)$ in the region $-h < y < 0$, $-\infty < x < \infty$, and the corresponding surface elevation $\eta(x, t)$ are to be determined subject to the boundary conditions (3.5), (3.6), (3.7), the initial conditions (3.8), (3.9) and the boundedness conditions at ∞. The conditions at ∞ are not formulated here explicitly: rather, they are implicitly used in the course of employing Fourier transforms with respect to the space variable x. This problem could be shown to have a unique solution.

The problem formulated above is solved by applying the Fourier transform in x to $\phi(x, y, t)$ to yield $\bar{\phi}(s, y, t)$:

$$\bar{\phi}(s, y, t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \phi(x, y, t)\, e^{-isx}\, dx. \tag{3.10}$$

The transform $\bar{\phi}$ of ϕ is now determined in the usual way through use of the conditions imposed on ϕ, after which $\bar{\phi}$ is inverted to obtain ϕ in the following form:

$$\phi(x, y, t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} A(s, t)\cosh s(y+h)\, e^{isx}\, ds, \tag{3.11}$$

with $A(s, t)$ defined as

$$A(s,t) = -\frac{isU\bar{p}}{\varrho \cosh sh}\left[\frac{1}{s^2U^2 - gs\tanh sh}\right.$$

$$+\frac{1}{2\sqrt{gs\tanh sh}}\,\frac{\exp\{-it(sU+\sqrt{gs\tanh sh})\}}{sU+\sqrt{gs\tanh sh}} \tag{3.12}$$

$$\left.-\frac{1}{2\sqrt{gs\tanh sh}}\,\frac{\exp\{-it(sU-\sqrt{gs\tanh sh})\}}{sU-\sqrt{gs\tanh sh}}\right].$$

The path of integration is the real axis. One finds easily that the integrand behaves for large s like $e^{|s|y}/s$, since the denominators of the terms in the square brackets in (3.12) behave like s^2, the ratio $\cosh s(y+h)/\cosh sh$ behaves like $e^{|s|y}$ for large s, and $\bar{p}(s)$ tends to zero at ∞ in general. Since y is negative it is clear that the integral converges uniformly. (We omit a discussion of the behavior on the free surface corresponding to $y = 0$, although such a discussion would not present any real difficulties.) Upon examining the function $A(s, t)$ in (3.12) it might seem that it has singularities at zeros of the denominators (and such zeros can occur, as we shall see), but in reality one can easily verify that the function has no singularities when the three terms in the square brackets are taken together—or, as one might also put it, any singularities in the individual terms cancel each other. Thus the solution given by (3.11) is a regular harmonic function in the strip $-h \leqq y < 0$ for all time t, or, in other words a motion exists no matter what values are given to the parameters. In addition, the fact that the integral converges ensures that ϕ (and also its derivatives) tends to zero for any given time when $|x| \to \infty$—this is the content of the so-called Riemann-Lebesgue theorem. This means that the amplitude of the disturbance dies out at infinity at any given time t—a not unexpected result since an appreciable time must elapse before the effects of a disturbance are felt at a distance from the seat of the disturbance.

However, we know from the discussion above (and from everyday observations of streams, for that matter) that as $t \to \infty$ it may happen that a disturbance exists downstream as a wave with non-vanishing amplitude. Our main interest here is to study such a passage to the limit. It is clear that one cannot accomplish such a purpose simply by letting $t \to \infty$ in (3.11), since for one thing, the transform $\bar{\phi}$ of ϕ cannot exist if ϕ does not tend to zero at ∞. What must be done is to consider the contributions of the separate items in the brackets in (3.12), and to avoid any singularities caused by zeros in their denominators by regarding $A(s, t)$ as an analytic function in the neighborhood of the real axis of a complex s-plane and deforming the path of integration in (3.11) by Cauchy's integral theorem in such a way as to avoid such singularities. One can then study the limit situation as $t \to \infty$.

One can, in fact, proceed in much the same fashion as in the preceding section, except that the calculation is more complicated. The results of special interest have already been stated above.

4. Sommerfeld's and Other Diffraction Problems

In the two preceding sections we have discussed time-independent solutions of wave propagation problems which result as limits as $t \to \infty$ of the solution of time-dependent initial value problems. Here we deal with some problems which are time-independent at the outset, and interest ourselves particularly in the question of the formulation of conditions at ∞ which will ensure the uniqueness of the solutions.

We are concerned here with boundary problems for the differential equation (the so-called reduced wave equation) $\nabla^2 f + f = 0$ in two dimensions and in unbounded domains. Such a differential equation arises in dealing with steady simple harmonic solutions of the linear wave equation. It also occurs in various problems concerning gravity waves in water, and the basic differential equation is then Laplace's equation in three dimensions. In general, the physical problems are wave propagation problems in which a primary, or incoming, wave is prescribed and it is desired to calculate the "secondary", that is the reflected, refracted, and diffracted waves which result.

One of the main difficulties to be overcome, as has been stated above, is the formulation of appropriate conditions at infinity which will determine the solution uniquely. For general domains it is not known how to formulate these conditions, and, in fact, it would seem to be a very difficult task to do so since such a formulation would almost certainly require consideration of many special cases. In one special case, however, the appropriate condition to be imposed at infinity has been known for a long time. This is the case in which any reflecting or refracting obstacles lie in a bounded domain of the plane, or, stated otherwise, it is the case in which a full neighborhood of the point at infinity is made up entirely of the homogeneous medium in which the waves propagate. In this case, the condition at ∞ which determines the "secondary" waves uniquely is Sommerfeld's radiation condition, which states, roughly speaking, that these waves behave like a cylindrical outgoing progressing wave at ∞. However, if the reflecting or refracting obstacles extend to infinity, the Sommerfeld condition may not be appropriate at all. Consider, for example, the case in which the entire x-axis is a reflecting barrier, and the primary wave is an incoming plane wave from infinity. It is clear on physical grounds that the secondary wave will be the reflected plane wave, which certainly does not behave at ∞ like a cylindrical wave since, for example, its amplitude does not even tend to zero at ∞. Another

case is that of Sommerfeld's classical diffraction problem in which an incoming plane wave is reflected from a barrier consisting of the positive half of the x-axis. In this case, the secondary wave has both a reflected component which has a non-zero amplitude at ∞, and a diffracted part which dies out at ∞. Recently[6] a uniqueness theorem has been derived which is applicable to Sommerfeld's diffraction problem as a special case.

The uniqueness theorem is formulated in the following way: We assume that $f(x, y)$ is a complex-valued solution[7] of the equation

$$\nabla^2 f + f = 0 \tag{4.1}$$

in a domain D with boundary Γ, part of which may extend to infinity. It is supposed that any circle C in the x, y-plane cuts out of D a domain in which the application of Green's formula is legitimate, and, in addition, that the

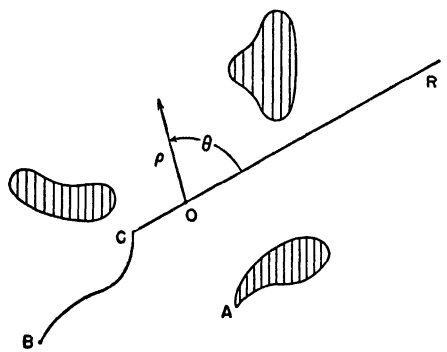


Figure 3. The domain D.

boundary curve Γ outside a sufficiently large circle consists of a single half-ray R going to ∞ (cf. Figure 3).[8] On the boundary Γ the condition

$$f_n = 0 \tag{4.2}$$

is imposed, i.e. the normal derivative of f vanishes, corresponding to a reflecting barrier. (We could also replace this condition on part, or all, of Γ

[6] Peters, A. S., and Stoker, J. J., *A uniqueness theorem and a new solution for Sommerfeld's and other diffraction problems*, Comm. Pure Appl. Math., Vol. 7, 1954, pp. 565–586. In proving the theorem use was made of the methods devised in a paper by F. Rellich, *Über das asymptotische Verhalten der Lösungen von $\Delta u + u = 0$ in unendlichen Gebieten*. Jber. Deutsch Math. Verein, Vol. 53, Abt. 1, 1943, pp 57–65.

[7] It is natural to consider such complex solutions, since, for example, a plane wave is obtained by taking $f(\varrho, \theta) = \exp\{\pm i\varrho \cos(\theta + \alpha)\}$ in polar coordinates ϱ, θ.

[8] Our theorem also holds if D is the more general domain in which the ray R is replaced at ∞ by a sector, and the proof given in the following section holds with insignificant modifications for this case also.

by the condition $f = 0$.) We now write the solutions of f in D which satisfy (4.2) in the form

$$f = g+h, \tag{4.3}$$

in order to formulate the conditions at ∞ in a convenient way. What we have in mind is to separate the solution into a part h which satisfies a radiation condition and a part g which contains, roughly speaking, the prescribed incoming wave together with any secondary reflected or refracted waves which also do not satisfy a radiation condition. More precisely, we require h to satisfy the following radiation condition:

$$\lim_{\varrho\to\infty} \int_C \left| \frac{\partial h}{\partial \varrho} + ih \right|^2 ds = 0. \tag{4.4}$$

Here C is taken to be a circle, with its center 0 (cf. figure above) on the half-ray R going to infinity, and with radius ϱ so large that all obstacle curves except a part of R lie in its interior. This condition clearly follows from the well-known Sommerfeld radiation condition, which requires that

$$\lim_{\varrho\to\infty} \varrho^{1/2} \left(\frac{\partial h}{\partial \varrho} + ih \right) \to 0 \tag{4.4}_1$$

uniformly in θ, and, incidentally, this is a condition independent of the particular point from which ϱ is measured.

The behavior of the function g at infinity is prescribed as follows:

$$g \sim g_1+g_2 \qquad \text{at } \infty, \tag{4.5}$$

where the function g_1 is prescribed once for all[9] while g_2 satisfies the same radiation condition as h, i.e. the condition (4.4). (That the behavior of g at ∞ is fixed only within an additive function satisfying the radiation condition is natural and inevitable.)

Finally, we prescribe regularity conditions at re-entrant angular points (such as A, B, C in the figure) of the boundary of D; these conditions are that

$$f(\varrho, \theta) \sim c_1, \qquad f_\varrho(\varrho, \theta) \sim \frac{c_2}{\varrho^k}, \qquad k < 1, \tag{4.6}$$

with ϱ, θ polar coordinates centered at the particular singular point, and c_1 and c_2 constants. (These conditions on f mean physically that the velocity of the medium may be infinite at a corner, but not as strongly as it would be for a source or sink.) At other boundary points we require continuity of f and its normal derivative.

[9] How the function g_1 should be chosen is a matter for later discussion.

We now state our theorem as follows:

UNIQUENESS THEOREM. *A solution f of* (4.1) *in D is uniquely determined if it* 1) *satisfies the boundary condition* (4.2), 2) *admits of a decomposition of the form* (4.3) *with h a function satisfying* (4.4), *g a function behaving as prescribed by* (4.5) *at* ∞ *and* 3) *satisfies the regularity conditions at the boundary Γ of D.*

We proceed to discuss the implications of the theorem. The theorem is at first sight somewhat unsatisfactory since it involves the assumption that every solution considered can be decomposed according to (4.3), with $g(\varrho, \theta)$ a certain function the behavior of which at ∞, in so far as the leading term g_1 (cf. (4.5)) in its asymptotic development is concerned, is not given *a priori*. However, it is not difficult in some instances at least to guess how the function $g_1(\varrho, \theta)$ should be defined. For example, suppose the domain D consisted of the exterior of bounded obstacles only. In such a case it seems clear that $g_1(\varrho, 0)$ should be defined as the function describing the incoming wave—either as a plane wave from infinity, say, or a wave originating from an oscillatory source—since bounded obstacles give rise only to reflected and diffracted components which die out at ∞ and which could be expected to satisfy the radiation condition. Even if there is a half-ray in the boundary that goes to ∞ (as was postulated above), it still would seem appropriate to take $g_1(\varrho,\theta)$ as the function describing the incoming wave, provided that it arises from an oscillatory point source, since such a source would hardly lead to reflected or refracted secondary waves which would violate the radiation condition. However, if the incoming wave is a plane wave and an obstacle extends to ∞, one expects an outgoing reflected wave to occur which would in general not satisfy the radiation condition; in this case the function $g_1(\varrho, \theta)$ should be taken as the sum of the incoming plane wave and an outgoing reflected wave. For example, one might consider the

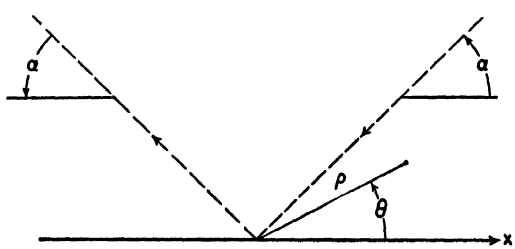


Figure 4. Infinite straight line barrier.

case in which the entire x-axis is a reflecting barrier, as in Figure 4. In this case one would in an altogether natural way define $g_1(\varrho, 0)$ as the sum of the incoming and of the reflected waves as follows:

$$g_1(\varrho, \theta) = \exp\{i\varrho \cos(\theta-\alpha)\}+\exp\{i\varrho \cos(\theta+\alpha)\}, \tag{4.7}$$

with α the angle of incidence of the incoming plane wave. If we were then to set $f \equiv g_1+h$ (i.e. we set $g \equiv g_1$ everywhere) and prescribe that h should satisfy the radiation condition, it is clear that we would have a unique solution by taking $h \equiv 0$. Our uniqueness theorem does not apply directly here since there are two reflecting half-rays going to ∞, but it could be easily modified so that it would apply to this case. Thus we have—for the first time, it seems—a uniqueness theorem for this particularly simple problem of the reflection of a plane wave by a rigid plane. A less trivial example is the classical Sommerfeld diffraction problem in which a plane wave coming from infinity at angle α to the x-axis is reflected and diffracted by a rigid half-plane barrier along the positive x-axis, as indicated in

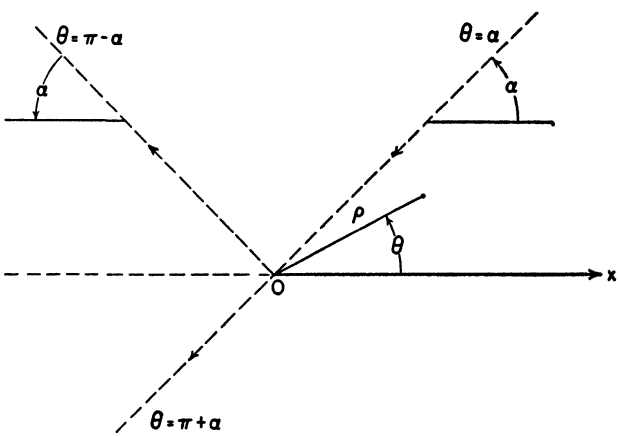


Figure 5. Sommerfeld's diffraction problem.

Figure 5. In this case it seems plausible to define the function $g_1(\varrho, \theta)$ as follows:

$$(4.8)\quad g_1(\varrho,\theta) = \begin{cases} \exp\{i\varrho\cos(\theta-\alpha)\}+\exp\{i\varrho\cos(\theta+\alpha)\}, & 0<\theta<\pi-\alpha, \\ \exp\{i\varrho\cos(\theta-\alpha)\}, & \pi-\alpha<\theta<\pi+\alpha, \\ 0, & \pi+\alpha<\theta<2\pi. \end{cases}$$

This function is, of course, discontinuous, corresponding to the division of the plane into the regions in which a) the incoming wave and its reflection from the barrier coexist, b) the region in which only the wave transmitted past the edge of the barrier exists, and c) the region in the shadow created by the barrier. Again we would be inclined to take $g \equiv g_1$ (cf. (4.3) and (4.5)) and set $f \equiv g_1+h$, with h satisfying the radiation condition. Of course, the function $h(\varrho, \theta)$ in (4.3) representing the diffracted wave would then also be discontinuous in that case since the sum g_1+h is everywhere continuous.

It has been shown in the paper cited above that the well-known solution given by Sommerfeld can be decomposed in this way and that h then satisfies the radiation condition. The uniqueness theorem is thus shown to be applicable in at least one important special case.

One might hazard a guess regarding the right way to determine the function g in cases involving unbounded domains: it seems highly plausible that it would always be correctly given by the methods of geometrical optics. By this we mean, from the mathematical point of view, that g would be the lowest order term in an asymptotic expansion of the solution f with respect to the frequency of the motion which is valid for large frequencies; the methods of geometrical optics would thus be available for determining g. However, to prove a theorem of such generality would seem to be a very difficult task since it would probably require some sort of representation for the solution of wave propagation problems when more or less arbitrary domains and boundary data are prescribed.

Once having observed that there is some point in decomposing the solution of Sommerfeld's problem in the way indicated above into the sum of two discontinuous functions, one of which satisfies the radiation condition, it was observed that the latter fact opens the way to a new solution of the diffraction problem which is entirely elementary, straightforward, and which can be written down in a few lines. In other words, once the reluctance to work with discontinuous functions is overcome, the solution of the problem is reduced to something quite trivial. Consequently it seems worthwhile to give this new solution in spite of the fact that the problem was solved long ago by Sommerfeld[10], and afterwards by many others.

We set

$$f(\varrho, \theta) = g(\varrho, \theta)+h(\varrho, \theta) \tag{4.9}$$

with $g(\varrho, \theta)$ defined by

$$g(\varrho, \theta) = \begin{cases} \exp\{i\varrho \cos(\theta-\alpha)\} + \exp\{i\varrho \cos(\theta+\alpha)\}, & 0<\theta<\pi-\alpha, \\ \exp\{i\varrho \cos(\theta-\alpha)\}, & \pi-\alpha<\theta<\pi+\alpha, \\ 0, & \pi+\alpha<\theta<2\pi. \end{cases} \tag{4.10}$$

In addition we have

$$f_\theta = 0 \qquad \text{for } \theta = 0, \quad \theta = 2\pi \tag{4.11}$$

and we also require

$$\lim_{\varrho\to\infty} \varrho^{1/2}\left(\frac{\partial h}{\partial \varrho} + ih\right) = 0 \text{ uniformly in } \theta, \tag{4.12}$$

[10]Sommerfeld, A., *Mathematische Theorie der Diffraktion*, Math. Ann., Vol. 47, 1896, pp. 317–374.

since the validity of the radiation condition in this strong form can be proved.

The desired solution will now be found by developing $f(\varrho, \theta)$ into a Fourier series in θ for fixed ϱ, and determining the coefficients of the series through use of the radiation condition (4.12); afterwards, the series can be summed to yield a convenient integral representation of the solution. That such a process will be successful can be seen very easily: The Fourier series for $f(\varrho, \theta)$ will, because of (4.11) and the fact that f is a solution of the reduced wave equation, be of the form $\sum c_n J_{n/2}(\varrho) \cos(n\theta/2)$; the Fourier coefficients for $g(\varrho, \theta)$ as defined by (4.10) are given in terms of integrals of the form

$$I_n = \int_0^{2\pi} \exp\{i\varrho \cos(\theta \pm \alpha)\} \cos\frac{n\theta}{2}\, d\theta.$$

Since $J_{n/2}(\varrho)$ as well as its derivatives behave like $1/\varrho^{1/2}$ for large values of ϱ and the integrals I_n—by a straightforward application of the method of stationary phase, for example—also behave in this way, it is clear that the limit relation (4.12) when used in connection with (4.9) will serve to determine the coefficients c_n.

We proceed to carry out this program. The finite Fourier transform $\bar{f}$ of f is introduced by the formula.

$$\bar{f}(\varrho, n) = \int_0^{2\pi} f(\varrho, \theta) \cos\frac{n\theta}{2}\, d\theta. \tag{4.13}$$

Since $f_\theta = 0$ for $\theta = 0, 2\pi$, we find for $f_{\theta\theta}$ the transform

$$\bar{f}_{\theta\theta} = -\frac{n^2}{4}\bar{f}, \tag{4.14}$$

by using two integrations by parts. Since f is a solution of

$$\varrho^2 f_{\varrho\varrho} + \varrho f_\varrho + f_{\theta\theta} + \varrho^2 f = 0, \tag{4.15}$$

it follows that $\bar{f}$ is a solution of

$$\varrho^2 \bar{f}_{\varrho\varrho} + \varrho \bar{f}_\varrho + \left(\varrho^2 - \frac{n^2}{4}\bar{f}\right) = 0, \tag{4.16}$$

and solutions of this equation are given by

$$\bar{f}(\varrho, n) = a_n J_{n/2}(\varrho). \tag{4.17}$$

The transform of $g(\varrho, \theta)$ is, of course,

$$\bar{g}(\varrho, n) = \int_0^{2\pi} g(\varrho, \theta) \cos\frac{n\theta}{2}\, d\theta, \tag{4.18}$$

and we have, in view of (4.9), the relation

$$(4.19)\qquad \int_0^{2\pi} h(\varrho,\theta)\cos\frac{n\theta}{2}\,d\theta = a_n J_{n/2}(\varrho) - \int_0^{2\pi} g(\varrho,\theta)\cos\frac{n\theta}{2}\,d\theta$$

or

$$(4.20)\qquad \bar{h}(\varrho,n) = a_n J_{n/2}(\varrho) - \bar{g}(\varrho,n).$$

We must next apply the operation $\varrho^{1/2}(\partial/\partial\varrho + i)$ to both sides of (4.19) and then make the passage to the limit, with the result

$$(4.21)\qquad 0 = \lim_{\varrho\to\infty} \varrho^{1/2}\left(\frac{\partial}{\partial\varrho}+i\right)\left[a_n J_{n/2}(\varrho) - \int_0^{2\pi} g(\varrho,\theta)\cos\frac{n\theta}{2}\,d\theta\right].$$

Since the functions $J_{n/2}(\varrho)$ behave asymptotically in the following way:

$$J_{n/2}(\varrho) \sim \left(\frac{2}{\pi\varrho}\right)^{1/2}\cos\left(\varrho - \frac{\pi}{4} - \frac{n\pi}{4}\right),$$

and since these expansions can be differentiated, we have

$$(4.22)\qquad \left(\frac{\partial}{\partial\varrho}+i\right)J_{n/2}(\varrho) \sim \left(\frac{2}{\pi\varrho}\right)^{1/2}\exp\left\{i\left(\varrho + \frac{\pi}{4} - \frac{n\pi}{4}\right)\right\}$$

as an easy calculation shows. The behavior of the integral in (4.21) can be found readily by the well-known method of stationary phase, which states that

$$\int_a^b \psi(\theta)\exp\{i\varrho\phi(\theta)\}\,d\theta \sim \left(\frac{2\pi}{\varrho|\phi''(\alpha)|}\right)^{1/2}\psi(\alpha)\exp\left\{i\varrho\left(\phi(\alpha)\pm\frac{\pi}{4}\right)\right\}$$

in which α is a simple zero of the derivative $\phi'(\theta)$ in the range $a<\theta<b$, and the ambiguous sign in the exponential is to be taken the same as the sign of $\phi''(\alpha)$. In the present case, in which $g(\varrho,\theta)$ is defined by (4.10), one sees at once that the term $\exp\{i\varrho\cos(\theta-\alpha)\}$ in the first line of (4.10) has a point of stationary phase in the appropriate interval; this term leads to the dominant term in the asymptotic expansion for large ϱ. The point of stationary phase is at $\theta=\alpha$, and we find from the above formula

$$\int_0^{\pi-\alpha}\exp\{i\varrho\cos(\theta-\alpha)\}\cos\frac{n\theta}{2}\,d\theta \sim \left(\frac{2\pi}{\varrho}\right)^{1/2}\cos\frac{n\alpha}{2}\exp\left\{i\left(\varrho-\frac{\pi}{4}\right)\right\}.$$

The right-hand side yields the behavior at ∞ of the integral in (4.21). This expansion can also be differentiated, so that we find

$$(4.23)\qquad \left(\frac{\partial}{\partial\varrho}+i\right)\int_0^{2\pi} g(\varrho,\theta)\cos\frac{n\theta}{2}\,d\theta \sim 2\left(\frac{2\pi}{\varrho}\right)^{1/2}\cos\frac{n\alpha}{2}\exp\left\{i\left(\varrho+\frac{\pi}{4}\right)\right\}.$$

Use of (4.22) and (4.23) in (4.21) furnishes, finally, the coefficients a_n:

$$(4.24)\qquad a_n = 2\pi\cos\frac{n\alpha}{2}\exp\left\{i\frac{n\pi}{4}\right\}.$$

The Fourier series for $f(\varrho, \theta)$ is

$$f(\varrho, \theta) = \frac{1}{2\pi} \bar{f}(\varrho, 0) + \frac{1}{\pi} \sum_{1}^{\infty} \bar{f}(\varrho, n) \cos \frac{n\theta}{2}$$

or, from (4.17) and (4.24),

$$f(\varrho, \theta) = J_0(\varrho) + 2 \sum_{1}^{\infty} \exp\left\{\frac{in\pi}{4}\right\} J_{n/2}(\varrho) \cos \frac{n\alpha}{2} \cos \frac{n\theta}{2}.$$

In the paper by Peters and the author cited above, it is shown that this series representation of the solution can, with no great difficulty, be transformed into integral representations of various kinds, including the well-known form given by Sommerfeld.

Received January 5, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 597–612 (1956)

A Direct Approach to the Problem of Stability in the Numerical Solution of Partial Differential Equations

JOHN TODD
National Bureau of Standards

Introduction

We shall give a direct arithmetic approach to the problem of stability of some partial differential equations, in particular the equation of diffusion (in one and two dimensions), the wave equation and the equation of the vibrating bar. We do not attempt here to indicate the scope of the method. The results obtained by the usual analytic approach[1] are recovered, and the operator theoretical treatment is motivated. We make considerable use of properties of special matrices and simple facts about the characteristic roots of a rational function $R(A)$ of a matrix A; these are summarized in §2.

1. What is Meant by the Numerical Solution of a Partial Differential Equation?

For simplicity take the case of

$$u_{xx} = u_t \qquad 0 < x < 1, \quad t > 0 \tag{1}$$

with

$$u(0, t) = 0, \qquad u(1, t) = 0, \quad t \geq 0, \qquad u(x, 0) = f(x). \tag{2}$$

What is required is something like this: values $U(m, n) = U(mh, nk)$ at points of a sufficiently dense net (i.e. h, k sufficiently small) such that

$$|u(mh, nk) - U(mh, nk)| \tag{3}$$

is sufficiently small over the region of interest. Indeed we should want more:

$$|u(x, t) - U(mh, nk)| \tag{4}$$

small for

$$|x - mh| \leq \tfrac{1}{2}h, \qquad |t - nk| \leq \tfrac{1}{2}k. \tag{5}$$

Various alternatives are possible. For instance, instead of requiring the

[1]Some of this work was carried out during a stay at the Institute of Mathematical Sciences, New York University and was stimulated by a series of lectures given by Professor R. D. Richtmyer, Director of the Atomic Energy Commission Computing Facility. See P. D. Lax and R. D. Richtmyer, Survey of the stability of linear finite difference equations, Comm. Pure Appl. Math., Vol. 9, No. 2, pp. 267–294.

absolute error to be small we might require a small relative error; i.e., we tolerate errors whose growth does not exceed that of the solution. Again, the region of interest might be specified in advance, or be specified implicitly, e.g., the region from $t = 0$ to $t = t_0$, if for $t \geqq t_0$

$$|u(x, t)|$$

is less than an assigned quantity.

To obtain these values by the most economical procedure is the task. For simplicity let us assume that we are concerned with direct evaluations of the U's; for some values of the parameters direct calculations at a coarse mesh followed by interpolation could be more efficient.

In the light of this statement of the problem we shall begin by discussing two methods based on different discretizations of (1). For definiteness let us suppose $h^{-1} = M+1$ for an integer M; it will be convenient to write $\theta = \theta_M = \frac{1}{2}\pi/(M+1)$.

A) The discretization indicated[2] by $\begin{matrix} & \times & \\ \times & \times & \times \end{matrix}$ leads to

$$\frac{U(m, n+1)-U(m, n)}{k} = \frac{U(m-1, n)-2U(m, n)+U(m+1, n)}{h^2}. \tag{6}$$

B) The discretization indicated by $\begin{matrix} \times & \times & \times \\ \times & \times & \times \end{matrix}$ leads to

$$\frac{U(m, n+1)-U(m, n)}{k} = \frac{\delta_m^2 U(m, n+1)+\delta_m^2 U(m, n)}{2h^2}, \tag{7}$$

where $\delta_m^2 g(m) = g(m+1)-2g(m)+g(m-1)$.

Our first remark is to the effect that both difference equations (6), (7) are *compatible* with the differential equation (1).

Using relations of the form

$$U(m, n+1) = U(m, n)+kU_t(m, n)+O(k^2)$$

in the difference equation (6), it reduces to

$$u_t+O(k) = u_{xx}+O(h).$$

In the same way we find that the difference equation (7) can be reduced to

$$u_t+O(k) = u_{xx}+O(h^2).$$

When there is compatibility one may hope for *convergence*, in the sense that the truncation error given by (4) can be made arbitrarily small by choosing h, k appropriately. In fact, in both these cases convergence can be established.[3]

---

[2] This is the pattern of lattice points (mh, nk) which occur as arguments of the function U in the approximating difference equation.

[3] A general treatment of (stability and) convergence which covers the case A) has been given by F. John [24]; his treatment does not cover the implicit case B). See also §4 below.

2. Results from Matrix Theory

We shall write

$$T_n = \begin{pmatrix} -2 & 1 & & & & \\ 1 & -2 & 1 & & & \\ & \cdot & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & 1 & -2 & 1 & \\ & & & 1 & -2 & \end{pmatrix} = \begin{pmatrix} & & \cdots & & & \\ \cdots, & 1, & -2, & 1, & \cdots & \\ & & \cdots & & & \end{pmatrix}.$$

The characteristic values of T_n are [39, 40]

$$\lambda_i = 2\cos\frac{i\pi}{n+1} - 2 = -4\sin^2\frac{i\pi}{2(n+1)}, \quad i = 1, 2, \cdots, n, \tag{8}$$

and the characteristic vector v_i corresponding to λ_i is

$$v_i = \left(\sin\frac{i\pi}{n+1}, \sin\frac{2i\pi}{n+1}, \cdots, \sin\frac{ki\pi}{n+1}, \cdots, \sin\frac{ni\pi}{n+1}\right). \tag{9}$$

We also use the fact that

$$T_n^2 = \begin{pmatrix} 5 & -4 & 1 & & & & \\ -4 & 6 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 6 & -4 \\ & & & & 1 & -4 & 5 \end{pmatrix};$$

thus T_n^2 has characteristic roots

$$\lambda_i^2 = 16\sin^4\frac{i\pi}{2(n+1)}, \qquad i = 1, 2, \cdots, n,$$

and the same characteristic vectors v_i as before.

We also need results about the characteristic roots of partitioned matrices [1, 47]. In particular, the characteristic roots of the $n^2 \times n^2$ matrix

$$\begin{pmatrix} T_n & I_n & & & \\ I_n & T_n & I_n & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & I_n & T_n & I_n \\ & & & I_n & T_n \end{pmatrix},$$

where I_n is the $n \times n$ unit matrix, are

$$\lambda_{i,j} = -2-2\cos\frac{i\pi}{n+1} - 2\cos\frac{i\pi}{n+1}, \quad i = 1, 2, \cdots, n, \; j = 1, 2, \cdots, n.$$

3. Introduction and Propagation of Errors

We are now able to begin our proper investigations. In the numerical solution of the problems under consideration we proceed line by line from the data $f(x)$ or, rather, $\{U(m, 0)\}$ to the solutions $\{U(m, 1)\}, \{U(m, 2)\}, \cdots$. In the first place we would like to know the effect of a line of errors—or what is the same, of a small change in the data—on the later computed values. We then have to examine the cumulative effect of all earlier errors.

How are these errors introduced? Consider the two cases in turn. Assume we are using a calculating machine with a fixed length accumulator working in multiples of 2ε, which might be 10^{-11} or 2^{-42} for instance.

A) Suppose, e.g., that $r = k/h^2 = 1/6$—a value which has certain advantages from the convergence point of view (cf. [33]). Then the (explicit) recurrence relation is

$$U(m, n+1) = \tfrac{1}{6}U(m-1, n) + \tfrac{2}{3}U(m, n) + \tfrac{1}{6}U(m+1, n).$$

Unless $U(m+1, n)$ and $U(m-1, n)$ are multiples of 6 and $U(m, n)$ is a multiple of 3 the individual terms on the right cannot be obtained exactly and, in general (unless the rounding errors compensate), $U(m, n+1)$ will not be obtained exactly. The maximum absolute error is clearly 3ε, where we have assumed that "rounded multiplication" is carried out, not e.g., "chopped". These errors occur for $m = 1, 2, \cdots, M$.

B) Somewhat similar considerations apply in the implicit case. The details depend on the methods used to solve the system of *simultaneous* linear equations for the new line of values (cf. footnote 3).

How are these errors propagated? We discuss A) and B) separately. We have, from (6), where $r = k/h^2$,

$$U(m, n+1) = rU(m-1, n) + (1-2r)U(m, n) + rU(m+1, n).$$

In matrix notation, writing

$$\chi_n = (U(1, n), U(2, n), \cdots, U(M, n))$$

for the vector which gives the values of U on the n-th line or row, we have

$$\chi_{n+1} = A\chi_n \tag{10}$$

where

$$A = \begin{pmatrix} & \cdots & \\ \cdots, & r, & 1-2r, & r, & \cdots \\ & \cdots & \end{pmatrix} = I + rT_n.$$

Hence

$$\chi_n = A^n \chi_0 .$$

The matrix A has characteristic values

$$\alpha_k = 1+r[2\cos 2k\theta - 2] = 1-4r\sin^2 k\theta; \tag{11}$$

the characteristic vectors are those of T_M.

B) We now have, from (7),

$$B\chi_{n+1} = (-B+4)\chi_n \tag{12}$$

where

$$B = \begin{pmatrix} \cdots \\ \cdots, -r, 2+2r, -r, \cdots \\ \cdots \end{pmatrix} = 2I - rT_n .$$

Hence

$$\chi_n = (4B^{-1}-I)^n \chi_0 .$$

The characteristic values of B are $2+4r\sin^2 k\theta$ and those of $4B^{-1}-I$ are

$$\beta_k = \frac{4}{2+4r\sin^2 k\theta} - 1 = \frac{1-2r\sin^2 k\theta}{1+2r\sin^2 k\theta}; \tag{13}$$

the characteristic vectors are again those of T_M.

Although our immediate interest is to find bounds for the propagated error, we have effectively obtained explicit solutions to the difference equations. We express the data $\{f(mh)\}$ in terms of the characteristic vectors of A, thus

$$\chi_0 = \{f(mh)\} = \sum_{i=1}^{M} c_i v_i \tag{14}$$

where the v_i are given by (9); here we have

$$c_i = \frac{2}{M+1} \sum_{s=1}^{M} f(sh) \sin \frac{is\pi}{M+1} . \tag{15}$$

Then, in case A), we have

$$\chi_n = A^n \chi_0 = \sum_{i=1}^{M} c_i A^n v_i = \sum_{i=1}^{M} c_i \alpha_i^n v_i \tag{16}$$

where the α_i are given by (11), while in case B) we have

$$\chi_n = \sum_{i=1}^{M} c_i \beta_i^n v_i \tag{17}$$

where the β_i are given by (13). The factors α_i, β_i indicate the amplification in the individual terms of (14) as we pass from row to row.

4. Convergence Considerations

From these explicit expressions for the solution of the difference equation we can, by passage to the limit, obtain the explicit solution to the

differential equation. We discuss A) only. A formal solution to the original differential equation (1) is

$$U(x,t) = \sum_{i=1}^{\infty} a_i \sin i\pi x \exp\{-i^2\pi^2 t\} \tag{18}$$

where

$$a_i = 2\int_0^1 f(y) \sin i\pi y\, dy. \tag{19}$$

This is to be compared with

$$U(m,n) = \sum_{i=1}^{M} c_i \sin 2mi\theta\, [1-4r\sin^2 i\theta]^n \tag{20}$$

where the c_i are given by (15). The fact that $c_i = c_i(M) \to a_i$ as $M \to \infty$ is evident if f is integrable in the sense of Riemann. To relate the individual terms of the two summations (18), (20) we only have to handle the exponential factors: for $i \ll M$ we have

$$\left[1-4r\sin^2\frac{i\pi}{2(M+1)}\right]^n \doteq \exp\left\{-4rn\frac{i^2\pi^2}{4(M+1)^2}\right\} = \exp\{i^2\pi^2 t\}$$

since $t = nk = nrh^2 = nr(M+1)^{-2}$. However, these formal considerations require justification. For a summary of results in this field we refer to Juncosa and Young [25]. The results available in this case are of the following form:

If $0 < r < \frac{1}{2}$ and if f is continuous except at a finite number of points where it has a finite jump then

$$U(m,n) \to u(x,t).$$

Estimates of the rate of convergence are available under various assumptions on f.

A different, more abstract, treatment is possible and more appropriate. The assumptions of compatibility and stability, i.e. uniform boundedness of $U(m,n)$, together with differentiability properties of the data, are sufficient to ensure the existence of subsequences $\{h_j\}$, $\{k_j\}$ such that the corresponding sequence $\{U_j\}$ converges to the solution u (see e.g. [4, 24, 28, 29]).

5. Estimates for the Generated Error

We are at present interested in the solution (16) to (6) when we regard the data χ_0 as a perturbation brought on by rounding errors. In these circumstances we study the growth of the solution $E = E(m,n)$ on the assumption[4] that

[4] This is the natural assumption when we are dealing with an explicit case such as A); if we deal with an implicit case such as B), the appropriate assumption may depend on the method of solution of the system of equations. If a relaxational solution is used, then it would be natural to assume that the residuals were uniformly small (cf. A. R. Mitchell [34, 35]).

(22) $$|f(mh)| \leqq \varepsilon.$$

From (15), using (22) we obtain

$$|c_i| \leqq \frac{2}{M+1} \cdot \varepsilon \cdot 1 \cdot M < 2\varepsilon.$$

Then from (16) we have

$$|E(m, n)| < 2\varepsilon \sum_{i=1}^{M} \alpha_i^n.$$

If we let $\bar{\alpha} = \max_{i=1,2,\cdots,M} |\alpha_i|$ then

$$|E(m, n)| < 2M\varepsilon\, \bar{\alpha}^{-n}.$$

Let us consider the case when $0 < r \leqq \frac{1}{2}$. Then $\bar{\alpha} < 1$ and we have

$$|E(m, n)| < 2M\varepsilon,$$

an estimate for the error caused at (m, n) by the initial error. If we consider the total error E^* caused by errors at $\chi_0, \chi_1, \cdots, \chi_{n-1}$ we obtain

$$|E^*(m, n)| < 2Mn\varepsilon.$$

In general we shall only be concerned with a bounded range of t. Then, in view of $r = k/h^2$ satisfying $0 < r \leqq \frac{1}{2}$, we may assume $n \sim h^{-2} \sim M^2$ and therefore,

$$|E^*(m, n)| = O(M^3\varepsilon).$$

This indicates the loss of accuracy due to the necessarily approximate character of our computation. For instance, with $M = 100$, and therefore 10^4 time steps, we would magnify an error by 10^6. This, of course, may not be very realistic; it is, however, quite tolerable. Better results can be obtained with more care; for instance, if $r = 1/6$ then $\bar{\alpha} < 1/3$.

In these circumstances, i.e. $0 < r \leqq \frac{1}{2}$, we have "stability" which we are defining as uniform boundedness of the solution of the discrete problem. It is usual to contrast this behavior with that when $\frac{1}{2} < r$, but this is unreasonable in so far as there is no convergence here apart from cases of special data. In practice, however, there may be no good estimates for the critical mesh shape (corresponding to $r = \frac{1}{2}$ in the present example), and the computer may be forced to resort to experiment; what can happen may be surmised from an estimate of $E(m, n)$, say for $r = 3/4$, which gives $\bar{\alpha} = 2$. Then

$$|E(m, n)| < 2M\varepsilon\, 2^{n+1},$$

and this "instability" will rapidly become apparent.

In case B) we see that all the characteristic roots β_k given by (13) are in absolute value less than unity for all r. This discretization is stable for all shapes of meshes. Results concerning convergence have been given by Juncosa and Young [25*a*].

6. Two More Discretizations

The first of these is due to DuFort and Frankel [16] and the second to L. F. Richardson [49]; both are improper in so far as they require extra data for starting—these additional data can be obtained, for instance, by using one of the methods (6), (7).

C) The discretization indicated by $\begin{matrix} & \times & \\ \times & \cdot & \times \\ & \times & \end{matrix}$ leads to

$$\frac{U(m, n+1)-U(m, n-1)}{2k} = \frac{U(m-1, n)-U(m, n+1)-U(m, n-1)+U(m+1, n)}{h^2}.$$

This can be written as

$$(1+2r)U(m, n+1) = (1-2r)U(m, n-1)+2r[U(m-1, n)+U(m+1, n)]$$

or, in matrix notation, as

$$\chi_{n+1} = \alpha C\chi_n+\beta I\chi_{n-1} \tag{23}$$

where

$$C = \begin{pmatrix} & \cdots & \\ \cdots, & 1, 0, 1, & \cdots \\ & \cdots & \end{pmatrix} = 2I+T_M$$

and $\alpha = 2r/(1+2r)$, $\beta = 1-2\alpha = (1-2r)/(1+2r)$.

It is clear that we can write

$$\chi_n = P_0^{(n)}(C)\chi_0+P_1^{(n)}(C)\chi_1$$

where P_0, P_1 are polynomials (of degree $< n$). Expansion of χ_0, χ_1 in terms of the characteristic vectors of C being carried out we see that the behavior of χ_n depends on the magnitude of the characteristic values of $P_0^{(n)}(C)$ and $P_1^{(n)}(C)$ which are $P_0^{(n)}(\lambda)$ and $P_1^{(n)}(\lambda)$, respectively, where λ is any characteristic root of C. Both polynomials satisfy a recurrence relation of the form

$$P^{(n+1)}(x) = xP^{(n)}(x)+P^{(n-1)}(x)$$

with initial conditions

$$P_0^{(0)} = 1, \qquad P_0^{(1)} = 0, \qquad P_1^{(0)} = 0, \qquad P_1^{(1)} = 1.$$

The recurrence relation can be solved in the usual way giving

$$P_0^{(n)} = a_0\mu_1^n+b_0\mu_2^n, \qquad P_1^{(n)} = a_1\mu_1^n+b_1\mu_2^n$$

where μ_1, μ_2 are the roots of the quadratic equation

$$\mu^2-\alpha x\mu-\beta = 0. \tag{24}$$

We have stability if $|\mu_1| \leqq 1$, $|\mu_2| \leqq 1$ whenever x is replaced by any

characteristic root $\gamma_i = 2\cos 2i\theta$, $i = 1, 2, \cdots, M$, of C. This we show to be the case for all values of r.

Consider $\mu^2 - 2(\alpha \cos \phi)\mu - 1 + 2\alpha = 0$ for $0 < \phi < \pi$. If this has complex roots they are equal in absolute value and $|\mu_i| = \sqrt{(1-2r)/(1+2r)} < 1$. If the roots are real then they are

$$\mu = \alpha \cos \phi \pm \sqrt{\alpha^2 \cos^2 \phi + 1 - 2\alpha}.$$

Since $\alpha > 0$, $0 < \phi < \pi$

$$\sqrt{1 - 2\alpha + \alpha^2 \cos^2 \phi} = \sqrt{(1-\alpha \cos \phi)^2 - 2\alpha(1-\cos \phi)} < 1 - \alpha \cos \phi;$$

it follows that $|\mu_i| < 1$.

An alternative approach to this and similar cases (§§8, 9 below) is the following: We can consider, in place of (23), the relation

$$L_{n+1} = HL_n$$

where L_n is the $2M$-dimensional vector

$$(\chi_n, \chi_{n-1})$$

and H is the $2M \times 2M$ matrix

$$H = \begin{pmatrix} \alpha C, & \beta I \\ I, & 0 \end{pmatrix}.$$

The characteristic roots of H are [1, 47] those of

$$\begin{pmatrix} \alpha\gamma_i, & \beta \\ 1, & 0 \end{pmatrix},$$

i.e., the roots of the quadratics

$$\mu^2 - \alpha\gamma_i\mu - \beta = 0.$$

The rest of the discussion is the same as before.

The convergence of this method has been discussed by M. L. Juncosa and D. M. Young [25b].

D) The discretization of Richardson is indicated by $\begin{matrix} & \times & \\ \times & \times & \times \\ & \times & \end{matrix}$ and leads to

$$U(m, n+1) = U(m, n-1) + 2rU(m-1, n) - 4rU(m, n) + 2rU(m+1, n)$$

which can be expressed as

$$\chi_{n+1} = 2rT\chi_n + \chi_{n-1}.$$

This can be discussed in either of the two ways used in connection with the DuFort-Frankel case. We are lead to the quadratic

$$\mu^2 - 2r\chi\mu - 1 = 0$$

where x runs through the characteristic roots (8) of T. Since

$$4r^2x^2+4 > 0$$

the roots of this quadratic are always real and different; since their product is -1 it follows that one of them is always greater in absolute value than unity. This process is, therefore, always unstable.

7. Effect of Lower Order Terms

We now consider the case

$$u_t = u_{xx}+au_x+bu$$

where a, b are constants. We shall show that the same conditions for stability hold for this equation as for (1). In the first place, there is no loss of generality in assuming $a = 0$, for, if not, a change of variable $u \to e^{-\frac{1}{2}ax}u$ will effect this.

Next we note that the matrix for

$$u_t = u_{xx}+bu$$

differs from the corresponding matrix for (1) by an additional term bk in each diagonal element. The change in the characteristic roots is therefore $O(k)$. Since as we have remarked, we can consider $n = O(k^{-1})$ the amplification is of amount at most

$$(1+O(k))^{O(k^{-1})}$$

which is bounded. The same conditions therefore apply.

8. The Wave Equation

The methods used for the equation (1) can be employed to discuss

$$u_{xx} = u_{tt}\,.$$

We confine our attention to the simple discretization indicated by $\times\times\times$ which leads to

$$\frac{U(m,n+1)-2U(m,n)+U(m,n-1)}{k^2} = \frac{U(m-1,n)-2U(m,n)+U(m+1,n)}{h^2}.$$

Writing $\gamma^2 = k^2/h^2$, this yields

$$\chi_{n+1} = (2I+\gamma^2 C)\chi_n-\chi_{n-1}\,.$$

As before, the question of stability reduces to a discussion of the magnitude of the roots of

$$\mu^2-(2+\gamma^2x)\mu+1 = 0$$

where x is a characteristic root of C. This equation must have equal or complex roots and, therefore, $-2 \leqq 2+\gamma^2 x \leqq 2$, i.e. $-4 \leqq \gamma^2 x \leqq 0$. Since $x=-4\sin^2 r\theta$ varies from 0 to -4 (nearly) we see that we must have $\gamma^2 \leqq 1$.

We note that this condition was proved sufficient for *convergence* in the classical paper of Courant, Friedrichs and Lewy [4].

9. Transverse Vibrations of a Bar

The transverse vibrations of a bar are governed by the equation

$$u_{xxxx}+u_{tt} = 0$$

where $u(x, 0)$, $u_t(x, 0)$ are given and $u(0, t)=u_{xx}(0, t)=u(1, t)=u_{xx}(1, t)=0$, $t \geqq 0$.

E) With a straightforward discretization indicated by $\begin{matrix} & \times & \\ \times & \times & \times \\ & \times & \end{matrix}$ we replace u_{xxxx} by

$$h^{-4}\{U(m+2, n)-4U(m+1, n)+6U(m, n)-4U(m-1, n)+U(m-2, n)\}$$

and u_{tt} by

$$k^{-2}\{U(m, n+1)-2U(m, n)+U(m, n-1)\},$$

and obtain a relation of the following type for χ_n, the n-th line of values of U:

$$\chi_{n+1} = -r^2 D\chi_n-\chi_{n-1}$$

where

$$D = \begin{pmatrix} & \cdots & \\ \cdots, & 1, -4, 6+2r^{-2}, -4, & 1, \cdots \\ & \cdots & \end{pmatrix}.$$

Similar arguments to those used in case C) are applicable. We also require the fact that the characteristic roots of D are

$$16\sin^4 i\theta-2r^{-2}.$$

Corresponding to the equation (24) we have

$$\mu^2+r^2 x\mu+1 = 0.$$

The product of the roots is unity and, therefore, we shall have instability unless the roots are equal or complex. To insure this we require

$$r^4[16\sin^4 i\theta-2r^{-2}]^2 \leqq 4, \qquad i = 1, 2, \cdots, n,$$

for which it is sufficient that

$$r^2 \leqq \tfrac{1}{4}.$$

F) We now consider the discretization $\begin{matrix} \times\times\times\times\times \\ \times\times\times\times\times \\ \times\times\times\times\times \end{matrix}$ which leads to

$$U(m, n+2)-2U(m, n+1)+U(m, n)$$
$$= -\tfrac{1}{4}r^2[\delta^4 U(m, n+2)+2\delta^4 U(m, n+1)+\delta^4 U(m, n)]$$

where $\delta^4 = \delta_m^4$ has its usual significance. Two vectors χ_n , χ_{n-1} being given, the above equation is an implicit equation for the succeeding line. As in case B), we can write this as

$$[I+r^2E]\chi_{n+2} = (2I-r^2E)\chi_{n+1}-(I+rE)\chi_n ,$$

where

$$E = \begin{pmatrix} & \cdots & \\ \cdots, & 1, -4, 6, -4, & 1, \cdots \\ & \cdots & \end{pmatrix}.$$

This gives

$$\chi_{n+2} = \{(2I-\tfrac{1}{4}r^2E)(I+\tfrac{1}{4}r^2E)^{-1}\}\chi_{n+1}-\chi_n .$$

It follows that

$$\chi_n = R_0(E)\chi_0+R_1(E)\chi_1$$

where R_0 , R_1 are certain rational functions. To determine the behavior of χ_n we need the characteristic roots of $R_0(E)$, $R_1(E)$: these are $R_0(\lambda_i)$, $R_1(\lambda_i)$, where the λ_i are the characteristic roots of E.

We are led to consider the roots of

$$\mu^2-\frac{1-\frac{1}{4}r^2\lambda}{1+\frac{1}{4}r^2\lambda}\mu+1 = 0$$

where λ is a root of E. In order not to have unstability we require that the roots of this quadratic be equal or complex. It is then necessary that

$$\left(\frac{2-\frac{1}{4}r^2\lambda}{1+\frac{1}{4}r^2\lambda}\right)^2 \leqq 4$$

for all λ; this is certainly the case, no matter what the value of $r = k/h^2$.

10. Two-dimensional Cases

We now discuss the equation

$$u_t = u_{xx}+u_{yy}$$

for $u = u(x, y, t)$ where $u(x, y, 0)$ is given and $u(x, y, t) = 0$ for all $t > 0$, and where (x, y) is a point on the boundary of the unit square. We take an explicit discretization on a lattice with sides g, h, k in t, x, y:

$$\frac{1}{g}\{U(m, n, \tau+1)-U(m, n, \tau)\}$$
$$= \frac{1}{h^2}\delta_m^2 U(m, n, \tau)+\frac{1}{k^2}\delta_n^2 U(m, n, \tau). \tag{25}$$

Then, for the M^2-dimensional vector χ, which gives the values of U for $t = \tau g$, we have

$$\chi_{\tau+1} = F\chi_\tau$$

where

$$F = \begin{pmatrix} & & & & \cdots & & & & \\ \cdots;\cdots,\gamma,\cdots;\cdots,\alpha,-2\alpha-2\gamma+1,\alpha,\cdots;\cdots,\gamma,\cdots;\cdots \\ & & & & \cdots & & & & \end{pmatrix}$$

and $\alpha = g/h^2$, $\gamma = g/k^2$.

The characteristic roots of F are found to be

$$1-4\alpha\cos^2 i\theta-4\gamma\sin^2 j\theta, \qquad i, j = 1, 2, \ldots, n.$$

In order that these should not exceed unity in absolute value it is sufficient that $\alpha+\gamma \leqq \frac{1}{2}$.

If, instead of using the differences in the $t(=\tau g)$-plane only, we took a mean among the differences in this plane and the next one, say

$$\frac{1}{g}\{U(m, n, \tau+1)-U(m, n, \tau)\} = \theta V_{\tau+1}+(1-\theta)V_\tau,$$

where V_τ is the expression on the right of (25), we would have an implicit scheme which could be expressed in the following form:

$$[I-\theta(F-I)]\chi_{\tau+1} = [I+(1-\theta)(F-I)]\chi_\tau.$$

We may write this as

$$\chi_{\tau+1} = \frac{I+(1-\theta)(F-I)}{I-\theta(F-I)}\chi_\tau$$

and thus we shall have stability if the characteristic roots of the matrix

$$\frac{I+(1-\theta)(F-I)}{I-\theta(F-I)}$$

lie between ± 1. The characteristic values of $F-I$ are

$$-4\alpha\cos^2 i\theta-4\gamma\cos^2 j\theta, \qquad i = 1, 2, \cdots, M, \quad j = 1, 2, \cdots, M.$$

No restrictions on α, γ are required if $\theta \geqq \frac{1}{2}$ since then

$$\frac{1-(1-\theta)(4\alpha\cos^2 i\theta+4\gamma\cos^2 j\theta)}{1+\theta(4\alpha\cos^2 i\theta+4\gamma\cos^2 j\theta)}$$

is always less than unity. However, for $\theta < \frac{1}{2}$ we see that we need

$$\alpha+\gamma \leqq \frac{1}{2-4\theta}.$$

11. Practical Significance of Stability

It is now possible to discuss the practical significance of stability; we shall, however, only indicate some of the factors which influence the choice of method of handling a particular problem. By use of a complicated discretization, a small truncation error may be obtained; this permits the use of larger values of h, k, the allowable error being fixed. With a complicated implicit method, stability for a „squarer" mesh may be possible, so that fewer time steps are necessary. In both cases, for some machines, difficulties of storage are additional to the complication of arithmetic.

Bibliography

[1] Afriat, S. N., *Composite matrices*, Quart. J. Math , Oxford Ser., Vol. 5, 1954, pp. 81–98.

[2] Bruce, G. H., Peaceman, D. W., Rachford, H. H. Jr , and Rice, J. D., *Calculations of unsteady-state gas flow through porous media*, Petroleum Trans. Amer. Inst. Mech. Engrs., Vol 198, 1953, pp. 79–92.

[3] Collatz, L , *Numerische Behandlung von Differentialgleichungen*, Springer, Berlin, 1951.

[4] Courant, R., Friedrichs, K. O., and Lewy, H., *Über die partiellen Differenzengleichungen der mathematischen Physik*, Math. Ann , Vol. 100, 1928, pp. 32–74.

[5] Courant, R., Isaacson, E., and Rees, M., *On the solution of non-linear hyperbolic differential equations by finite differences*, Comm. Pure Appl. Math., Vol. 5, 1952, pp 243–255.

[6] Crandall, S. H., *On a stability criterion for partial difference equations*, J. Math. Physics, Vol. 32, 1953, pp. 80-81.

[7] Crandall, S. H., *Numerical treatment of a fourth order parabolic partial differential equation*, J. Assoc. Comput. Mach., Vol. 1, 1954, pp. 111–118.

[8] Crandall, S. H., *Implicit vs. explicit recurrence formula for the linear diffusion equation*, J. Assoc. Comput. Mach., Vol 2, 1954, pp 42–49.

[9] Crank, J., and Nicolson, P., *A practical method for numerical evaluation of solution of partial differential equations of heat-conduction type*, Proc Cambridge Philos. Soc., Vol. 43, 1947, pp. 50–67.

[10] Dahlquist, G., *Convergence and stability for a hyperbolic difference equation with analytic initial values*, Math. Scand., Vol 2, 1954, pp. 91–102

[11] Douglas, J., *On the numerical integration of* $u_{xx}+u_{yy} = u_t$ *by implicit methods*, J. Soc. Indust. Appl. Math., Vol. 3, 1955, pp 42–65

[12] Douglas, J., *On the numerical solution of quasi-linear parabolic differential equations*, Pacific J. Math , to appear

[13] Douglas, J., *On the relation between stability and convergence in the numerical solution of linear parabolic and hyperbolic differential equations*, Abstract, Bull. Amer. Math. Soc , Vol. 61, 1955, p. 550

[14] Douglas, J., *The solution of the diffusion equation by a high order correct difference equation*, Abstract, Bull Amer Math Soc., Vol. 61, 1955, p. 226.

[15] Douglas, J., and Gallie, T. M., *Variable time steps in the solution of the heat-flow equation by a difference equation*, Proc. Amer. Math. Soc., Vol. 6, 1955, pp. 787–793.

[16] DuFort, E. C., and Frankel, S. P., *Stability conditions in the numerical treatment of parabolic differential equations*, Math. Tables and Other Aids to Computation, Vol. 7, 1953, pp. 135–152.

[17] Eddy, R. P., *Stability in the numerical solution of initial value problems in partial differential equations*, U. S. Naval Ordnance Lab , White Oak, Maryland, Memo 10232, 1949.

[18] Frankel, S. P., *Stability problems in partial differential equations*, Proc. Numerical Analysis Symposium, Santa Monica, 1953, Prentice-Hall, New York.

[19] Frankel, S. P., *Convergence rates of iterative treatments of partial differential equations*, Math. Tables and Other Aids to Computation, Vol. 4, 1950, pp. 65–75.

[20] Hildebrand, F. B., *On the convergence of numerical solutions of the heat-flow equation*, J. Math. Physics, Vol. 31, 1952, pp. 35–41.

[21] Hildebrand, F. B., *Advanced Calculus for Engineers*, McGraw-Hill, New York, 1956.

[22] Hildebrand, F. B., *Methods of Applied Mathematics*, Prentice-Hall, New York, 1952.

[23] Hyman, M. A., *On the numerical solution of partial differential equations*, Dissertation, University of Delft, 1953.

[24] John, F., *On integration of parabolic equations by difference methods*, Comm. Pure Appl. Math., Vol 5, 1952, pp. 155–211.

[25] Juncosa, M. L., and Young, D. M., *On the order of convergence of solutions of a difference equation to a solution of the diffusion equation*, J. Soc. Indust. Appl Math., Vol 1, 1953, pp. 111–135.

[25a] Juncosa, M L., and Young, D. M., *On the accuracy of finite difference methods for solving the diffusion equation*, Bull Amer. Math Soc. Abstract, Vol. 61, 1955, pp. 145-146.

[25b] Juncosa, M. L and Young, D. M., *On the Frankel-Nicolson procedure for solving parabolic partial differential equations*, Rand Report, No. P-465, 1954

[26] Laasonen, P., *Über eine Methode zur Losung der Warmeleitungsgleichung*, Acta Math., Vol. 81, 1949, pp. 309–317.

[27] Lax, P. D., *Weak solutions of non-linear hyperbolic equations and their numerical computation*, Comm. Pure Appl. Math., Vol 7, 1954, pp. 159–163

[28] Lax, P. D., *Initial value problems for non-linear hyperbolic equations*, Proc Conference on Partial Differential Equations, Lawrence, Kansas, 1954.

[29] Lax, P. D., *Approximation to solutions of differential equations—an operator-theoretical approach*, Berkeley Proceedings, 1955, to appear as a report at the University of Kansas

[30] Leutert, W. W., *On the convergence of unstable approximate solutions of the heat equation to the exact solution*, J. Math. Physics, Vol. 30, 1952, pp. 245–251.

[31] Leutert, W. W., *On the convergence of approximate solutions of the heat equation to the exact solution*, Proc. Amer. Math. Soc., Vol 2, 1951, pp 433–439.

[32] Lewy, H., *On the convergence of solutions of difference equations*, in *Studies and Essays Presented to Richard Courant*, Interscience, New York, 1948, pp. 211–214.

[33] Milne, W. E., *Numerical Solution of Differential Equations*, Wiley, New York, 1953.

[34] Mitchell, A. R., *Round-off errors in the solution of the heat conduction equations by relaxational solutions*, Appl. Sci. Research, Ser. A, Vol. 4, 1953, pp. 109–119

[35] Mitchell, A. R., *Round-off errors in relaxational solutions of Poisson's equation*, Appl Sci. Research, Ser. B, Vol 3, 1954, pp. 456–464.

[36] Mitchell, A R., and Rutherford, D. E., *On the theory of relaxation*, Proc. Glasgow Math. Assoc., Vol. 1, 1953, pp. 101–110.

[37] O'Brien, G. G., Hyman, M. A., and Kaplan, S., *A study of the numerical solution of partial differential equations*, J. Math. Physics, Vol 29, 1951, pp. 223–251

[38] Richtmyer, R. D., Unpublished lectures at New York University, 1954–1955.

[39] Rutherford, D. E., *Some continuant determinants arising in physics and chemistry*, Proc. Roy Soc. Edinburgh, Sect A, Vol. 62, 1947, pp 229–236.

[40] Rutherford, D. E., *Some continuant determinants arising in physics and chemistry, II*, Proc. Roy Soc Edinburgh, Sect A, Vol. 63, 1952, pp. 232–261.

[41] Schoenberg, I. J., *Partial differential and difference equations*, unpublished lecture notes, National Bureau of Standards, 1951.

[42] Thomas, L. H., *Stability of partial differential equations*, U. S. Naval Ordnance Lab. Report No 1132, 1950.

[43] Thomas, L. H., Unpublished lecture notes.

[44] Thomas, L. H., I. B. M. Seminar, November 1949.

[45] Todd, J., *Solution of differential equations by recurrence relations*, Math. Tables and Other Aids to Computation, Vol. 4, 1950, pp. 39–44.

[46] von Neumann, J., and Richtmyer, R. D., *A method for numerical calculation of hydrodynamic shocks*, J. Appl. Physics, Vol. 21, 1950, pp. 232–237.

[47] Williamson, J., *The latent roots of a matrix of special type*, Bull. Amer. Math. Soc., Vol. 37, 1931, pp. 585–590.

[48] Evans, G. W., Brousseau, R., and Keirstead, R., *Stability considerations for various difference equations for the linear heat conduction equation*, J. Math. Physics, Vol. 34, 1956, pp. 267–285.

[49] Richardson, L. F., *The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam*, Philos. Trans. Roy. Soc. London, Ser. A, Vol. 210, 1910, pp. 307–357.

[50] Blanch, G., *On the numerical solution of parabolic partial differential equations*, J.Research Nat. Bur. Standards, Vol. 50, No. 6, 1953, pp. 343–356.

[51] Peaceman, D. W., and Rachford, H. H. Jr., *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Indust. Appl. Math., Vol. 3, 1955, pp. 28–41.

Received January 20, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 613–623 (1956)

Upper and Lower Bounds for Eigenvalues by Finite Difference Methods*

H. F. WEINBERGER

Institute for Fluid Dynamics and Applied Mathematics, University of Maryland

1. Introduction

We are concerned with estimating the minimum of the ratio of two quadratic functionals corresponding to certain differential operators on a finite region R. It was pointed out by Courant [1] that if the functionals contain no derivatives of order higher than the first, an upper bound may be found by means of a finite difference minimum problem. This is done by substituting in the minimum principle a broken linear function taking on the values of the solution of the finite difference problem at the mesh points.[1]

This idea was extended in the case of the torsion problem by Synge [7]. He obtained both upper and lower bounds for the torsional rigidity by means of two kinds of pyramid functions. His method, however cannot be extended to eigenvalue problems.

The finite difference problems for a square mesh giving upper bounds for torsional rigidity and membrane frequencies were explicitly formulated by Polya [5].

In this paper we present a method of finding a lower bound for the minimum by means of a finite difference method [8]. While the upper bound is obtained from a finite difference problem on a grid that is smaller than or equal to R, the lower bound is given by a problem on a grid that is slightly larger. A very similar method has meanwhile been independently developed by Hersch [3]. It is described and improved upon in Section 3.

Forsythe [2] has shown that for convex domains there exist finite difference problems giving a lower bound for the lowest membrane frequency, provided the mesh size is below a critical size. However, there seems to be no way of estimating this critical size.

Our result holds for very general domains and always gives a lower bound.

---

*The research for this paper was supported by the United States Air Force through the Office of Scientific Research.

[1]See also L. Collatz, *Konvergenz des Differenzverfahrens bei Eigenwertproblemen partieller Differentialgleichungen*, Deutsche Math., Vol. 3, 1938, pp. 200–212.

While the method here presented works for higher order operators, such as the biharmonic operator, we shall restrict our attention to the membrane problem

$$\begin{aligned} &\Delta u+\lambda u = 0 \quad \text{in } R, \\ &u = 0 \text{ on the boundary } \dot{R}. \end{aligned} \tag{1.1}$$

The minimum problem for the lowest eigenvalue is

$$\lambda = \min_{v=0 \text{ on } \dot{R}} \frac{D(v, v)}{\iint_R v^2 \, dx \, dy}. \tag{1.2}$$

Here D is the Dirichlet integral.

In this case we obtain a finite difference problem giving a lower bound for λ. An upper bound for λ is given by a different finite difference equation on a smaller domain [5]. In Section 4 we derive another upper bound in terms of the finite difference equation used for the lower bound but on the smaller domain. This permits us to estimate the difference between the two bounds in the case of certain domains. In this case, then, one can find both upper and lower bounds for λ with prescribed accuracy by solving a single finite difference problem with a determined mesh size.

2. A Lower Bound by a Finite Difference Method

We begin with a basic lower bound for the membrane problem (1.1). Let R_h be a grid consisting of the union of squares of side h with their sides adjacent and parallel to the x and y axes. We require that R_h contain not only the region R on which the membrane problem is to be solved, but also all its left and downward translates of distance at most h:

$$R_h \supset \{x, y | (x+\alpha, y+\beta) \in R \quad \text{for some } 0 \leqq \alpha \leqq h, \ 0 \leqq \beta \leqq h\}. \tag{2.1}$$

We now consider functions $v(x_i, y_j)$ defined only at the mesh points (x_i, y_j) of the finite difference grid and vanishing at points on the boundary of or outside R_h. Let

$$\lambda^{(h)} = \min \frac{D^*(v, v)}{\sum h^2 v^2(x_i, y_j)} \tag{2.2}$$

in this class of functions. Here

$$D^*(v, v) = \sum \{[v(x_i+h, y_j)-v(x_i, y_j)]^2+[v(x_i, y_j+h)-v(x_i, y_j)]^2\}, \tag{2.3}$$

and sums are taken over all mesh points. The minimizing function w of (2.2) satisfies the finite difference equation

$$\Delta_h w+\lambda^{(h)} w = 0 \tag{2.4}$$

where

$$\Delta_h w(x_i, y_j) = \frac{1}{h^2}[w(x_i+h, y_j)+w(x_i-h, y_j)+w(x_i, y_j+h) + w(x_i, y_j-h)-4w(x_i, y_j)].$$

We now continue the eigenfunction $u(x, y)$ of the membrane problem (1.1) as zero outside R, and let

$$v(x_i, y_j) = u(x_i+\alpha, y_j+\beta), \quad 0 \leqq \alpha \leqq h, \ 0 \leqq \beta \leqq h \tag{2.5}$$

Because of (2.1) this function is admissible in the minimum principle (2.2). Thus we have

$$h^2 \lambda^{(h)} \sum u^2(x_i+\alpha, y_j+\beta) \leqq \sum [u(x_i+\alpha+h, y_j+\beta)-u(x_i+\alpha, y_j+\beta)]^2 + [u(x_i+\alpha, y_j+\beta+h)-u(x_i+\alpha, y_j+\beta)]^2. \tag{2.6}$$

Integrating this inequality with respect to α and β from 0 to h gives

$$h^2 \lambda^{(h)} \iint_R u^2 \, dxdy \leqq \iint_R \{[u(x+h)-u(x)]^2+[u(y+h)-u(y)]^2\} dxdy. \tag{2.7}$$

Now, by Schwarz's inequality,

$$\begin{aligned} \iint_R [u(x+h, y)-u(x, y)]^2 \, dxdy &= \iint \left[\int_0^h u_x(x+\gamma, y) d\gamma\right]^2 dxdy \\ &\leqq \iint h \int_0^h u_x^2(x+\gamma, y) d\gamma dxdy \\ &= h^2 \iint u_x^2(x, y) \, dxdy, \end{aligned} \tag{2.8}$$

with a similar result for the second term on the right-hand side of (2.7). Thus,

$$\lambda^{(h)} \leqq \frac{D(u)}{\iint_R u^2 \, dxdy} \tag{2.9}$$

or

$$\lambda^{(h)} \leqq \lambda. \tag{2.10}$$

In other words, $\lambda^{(h)}$ is a lower bound for λ.

3. An Improved Lower Bound

Let $R_{h/2}$ be a refinement of the mesh R_h. That is, each square is divided into four squares and the squares not needed for $R_{h/2}$ to satisfy (2.1) are eliminated. Then we also have

$$R_h \supset \{x, y | (x+\alpha, y+\beta) \in R_{h/2} \text{ for any } \alpha = 0 \text{ or } h/2, \quad \beta = 0 \text{ or } h/2.\} \tag{3.1}$$

We use in the minimum principle (2.2) the function

$$v(x_i, y_j) = q(x_i+\alpha, y_j+\beta), \quad \alpha = 0 \text{ or } h/2, \quad \beta = 0 \text{ or } h/2, \tag{3.2}$$

where q is now the minimizing function for the corresponding problem on $R_{h/2}$. A derivation similar to that of (2.9) but with integrals replaced by sums yields

$$\lambda^{(h)} \leqq \lambda^{(h/2)}. \tag{3.3}$$

A similar process was independently discovered at about the same time by J. Hersch [3]. He obtained the lower bound

$$\lambda \geqq \frac{8}{h^2}\left(\sin^{-1}\sqrt{\frac{h^2\lambda^{(h)}}{8}}\right)^2 \tag{3.4}$$

which is an improvement of (2.10) by a term of $O(h^2)$.

The reason for the improvement is that some of the terms which are neglected in the Schwarz inequality used to derive (3.3) can be estimated. In fact, we shall now show that an improvement of $O(h)$ can be obtained.

We have, using the substitution (3.2) and summing numerator and denominator of (2.2) with respect to α and β,

$$\lambda^{(h)} \leqq \frac{\sum_{\alpha,\beta=0,\,h/2} D^*(q(x_i+\alpha, y_j+\beta),\, q(x_i+\alpha, y_j+\beta))}{h^2\sum' q^2(x_i, y_j)}. \tag{3.5}$$

Here $\sum'$ indicates summation over the mesh points of $R_{h/2}$. Now it is easily verified that

$$\begin{aligned}
&\sum_{\alpha,\beta=0,\,h/2} D^*(q, q) \\
&\quad = \sum' \{[q(x_i+h, y_j)-q(x_i, y_j)]^2+[q(x_i, y_j+h)-q(x_i, y_j)]^2\} \\
&\quad = 4D'(q, q)-\sum'\left[q\left(x_i+\frac{h}{2}, y_j\right)-2q(x_i, y_j)+q\left(x_i-\frac{h}{2}, y_j\right)\right]^2 \\
&\qquad -\sum'\left[q\left(x_i, y_j+\frac{h}{2}\right)-2q(x_i, y_j)+q\left(x_i, y_j-\frac{h}{2}\right)\right]^2
\end{aligned} \tag{3.6}$$

where D' denotes the Dirichlet sum analogous to D^*, but with h replaced by $h/2$. We note that when (x_i, y_j) is an interior point of $R_{h/2}$,

$$\begin{aligned}
&\left[q\left(x_i+\frac{h}{2}, y_j\right)-2q(x_i, y_j)+q\left(x_i-\frac{h}{2}, y_j\right)\right]^2 \\
&\qquad +\left[q\left(x_i, y_j+\frac{h}{2}\right)-2q(x_i, y_j)+q\left(x_i, y_j-\frac{h}{2}\right)\right]^2 \\
&\qquad \geqq \frac{1}{2}\left(\frac{h}{2}\right)^4[\Delta_{h/2}q]^2 = \frac{h^4}{32}[\lambda^{(h/2)}q(x_i, y_j)]^2.
\end{aligned} \tag{3.7}$$

We are left with the last two sums in (3.4) extended over boundary points of $R_{h/2}$. To estimate these sums, we use the following identity:

$$
\begin{aligned}
&-h\sum{}'' \Delta_{h/2} q \left\{ x_i \left[q\left(x_i+\frac{h}{2}, y_j\right) - q\left(x_i-\frac{h}{2}, y_j\right)\right]\right. \\
&\left. + y_j \left[q\left(x_i, y_j+\frac{h}{2}\right) - q\left(x_i, y_j-\frac{h}{2}\right)\right] + \left(\frac{h}{2}\right)^3 \Delta_{h/2} q \right\} \\
&= \sum_{\alpha,\beta=0,\,h/2} D^*(q(x_i+\alpha, y_j+\beta), q(x_i+\alpha, y_j+\beta)),
\end{aligned} \tag{3.8}
$$

where $\sum''$ denotes summation over the boundary points of $R_{h/2}$. This is the finite difference analogue of

$$
2\lambda \iint_R u^2\, dxdy = \oint_{\dot R} r \frac{\partial r}{\partial n}\left(\frac{\partial u}{\partial n}\right)^2 ds \quad \text{when} \quad \Delta u+\lambda u = 0, \quad u = 0 \text{ on } \dot R, \tag{3.9}
$$

which is a special case of an identity due to Rellich [6]. The latter was used extensively by Payne and the author [4] for the estimation of quadratic functionals.

From the identity (3.8) it follows that

$$
\begin{aligned}
&\sum{}'' \left\{ \left[q\left((x_i+\frac{h}{2}, y_j\right) - 2q(x_i, y_j) + q\left(x_i-\frac{h}{2}, y_j\right)\right]^2\right. \\
&\left. + \left[q\left(x_i, y_j+\frac{h}{2}\right) - 2q(x_i, y_j) + q\left(x_i, y_j-\frac{h}{2}\right)\right]^2 \right\} \geqq \frac{h}{4\varrho} \sum_{\alpha,\beta=0,\,h/2} D^*(q,q)
\end{aligned} \tag{3.10}
$$

where ϱ is defined as follows. At boundary points of $R_{h/2}$ where the normal is in the x or y direction define a function ξ by

$$
\xi = -\mathbf{r}\cdot\mathbf{n} - \tfrac{1}{2}h \tag{3.11}
$$

where $\mathbf{r}$ is the radius vector and $\mathbf{n}$ the unit inward normal. At reentrant corners define two numbers ξ_1 and ξ_2 by (3.11), letting $\mathbf{n}$ be an inward unit vector in the x and y directions, respectively, and put

$$
\xi = \sup\left(\xi_1, \xi_2, \tfrac{3}{2}\xi_1+\tfrac{1}{2}\xi_2, \tfrac{3}{2}\xi_2+\tfrac{1}{2}\xi_1\right). \tag{3.12}
$$

At non-reentrant right-angle corners, let $\xi = 0$. Then

$$
\varrho = \sup \xi. \tag{3.13}
$$

It is easily seen that ϱ is at most equal to the side of the smallest square containing $R_{h/2}$.

Inserting (3.7) and (3.10) in (3.6) yields the inequality

$$
\begin{aligned}
\left(1+\frac{h}{4\varrho}\right) \sum_{\alpha,\beta=0,\,h/2} D^*(q, q) &\leqq 4D'(q, q) - \frac{h^4\lambda^{(h/2)}}{32}\sum{}' q^2 \\
&= h^2\lambda^{(2/h)}\left(1-\frac{h^2\lambda^{(h/2)}}{32}\right)\sum{}' q^2,
\end{aligned} \tag{3.14}
$$

since q is the minimizing function for $R_{h/2}$. Inequality (3.14), together with

(3.5), gives

$$\left(1+\frac{h}{4\varrho}\right)\lambda^{(h)} \leqq \lambda^{(h/2)}\left(1-\frac{\lambda^{(h/2)}h^2}{32}\right). \tag{3.15}$$

This is a recursion relation which gives an inequality between $\lambda^{(h)}$ and $\lambda^{(h/2^n)}$. Since we know that $\lambda^{(h/2^n)}$ is a lower bound for λ, each of these inequalities gives a lower bound for λ. It is easy to see that the bounds improve as $n \to \infty$ so that the best bound is given by the limiting inequality. If we take $\varrho = \infty$ in (3.12), the recursion may be solved explicitly. The limiting inequality is precisely that given by Hersch, namely (3.4). However, it is clear that for small h the correcting term $h/4\varrho$ is more important than the term in h^2.

It might be remarked that since Hersch apparently did not prove that $\lambda \leqq \lambda^{(h)}$, he seems to have assumed that $\lambda^{(h/2^n)} \to \lambda$ as $n \to \infty$. This is, however, not the case for certain R, for example, when R contains a cut.

The author has been unable to solve the recursive inequality (3.12) explicitly. However, replacing $\lambda^{h/2}$ by λ only in the $O(h^2)$ term before going to the limit yields

$$\lambda \geqq \lambda^{(h)}\prod_{n=0}^{\infty}\frac{1+\frac{h}{4\varrho}2^{-n}}{1-\frac{\lambda h^2}{32}4^{-n}} \geqq \lambda^{(h)}\frac{1+\frac{h}{2\varrho}+\frac{h^2}{12\varrho^2}}{1-\frac{\lambda h^2}{24}+\frac{\lambda^2 h^4}{576}}. \tag{3.16}$$

4. The Connection Between Upper and Lower Bounds

It is not difficult to establish the convergence of the lower bounds $\lambda^{(h)}$ to λ as $h \to 0$, provided R is sufficiently regular and the boundary of R_h approaches that of R. The same is true of the known lower bounds. Thus for a large class of regions R it is possible to find arbitrarily close upper and lower bounds for λ by finite difference methods. However, one does not know *a priori* how small to take the mesh size h in order to obtain a given accuracy. This means that in practice one must in general solve many finite difference problems with smaller and smaller mesh size before λ is approximated to a desired accuracy. We shall now attempt to remedy this situation at least for certain regions R.

The upper bound, as explicitly formulated by Polya [5], comes from a finite difference problem on a finite difference domain $\overline{R_h}$ whose boundary points are contained in R. (It is not necessary to consider points all of whose neighbors are on the boundary of or exterior to $\overline{R_h}$, since these do not enter the difference equation.) The upper bound $\bar{\mu}^{(h)}$ is the lowest eigenvalue of the equation

$$(4.1)\quad \Delta_h u(x_i, y_j)+\frac{\bar{\mu}}{12}[6u(x_i, y_j)+u(x_i+h, y_j)+u(x_i-h, y_j)+u(x_i, y_j+h)$$
$$+u(x_i, y_j-h)+u(x_i+h, y_j+h)+u(x_i-h, y_j-h)] = 0$$

at interior points of $\overline{R_h}$, with $u = 0$ at boundary points. We first relate $\bar{\mu}^{(h)}$ to the lowest eigenvalue $\bar{\lambda}^{(h)}$ of the problem

$$(4.2)\quad \Delta_h v(x_i, y_j)+\bar{\lambda} v(x_i, y_j) = 0$$

in $\overline{R_h}$ with $v = 0$ at boundary points. The variational principle for $\bar{\mu}^{(h)}$ is

$$\frac{1}{\bar{\mu}^{(h)}} \geqq \frac{\frac{h^2}{12}\sum w(x_i, y_j)[6w(x_i, y_j)+w(x_i+h, y_j)+w(x_i-h, y_j)+w(x_i, y_j+h)}{+w(x_i, y_j-h)+w(x_i+h, y_j+h)+w(x_i-h, y_j-h)]}{\sum\{[w(x_i+h, y_j)-w(x_i, y_j)]^2+[w(x_i, y_j+h)-w(x_i, y_j)]^2\}}$$

$$(4.3)\quad = \frac{h^2\sum w^2-h^2\sum\{[w(x_i+h, y_j)-w(x_i, y_j)]^2+[w(x_i, y_j+h)}{-w(x_i, y_j)]^2\}-h^2\sum[w(x_i+h, y_j+h)-w(x_i, y_j)]^2}{\sum\{[w(x_i+h, y_j)-w(x_i, y_j)]^2+[w(x_i, y_j+h)-w(x_i, y_j)]^2\}}$$
$$\geqq \frac{h^2\sum w^2}{\sum[w(x_i+h, y_j)-w(x_i, y_j)]^2+[w(x_i, y_j+h)-w(x_i, y_j)]^2}-3h^2$$

for any function w vanishing on the boundary of and exterior to $\overline{R_h}$. Letting $w = v$, we find

$$(4.4)\quad \bar{\mu}^{(h)} \leqq \frac{\bar{\lambda}^{(h)}}{1-3h^2\bar{\lambda}^{(h)}}.$$

Since $\bar{\mu}^{(h)}$ is an upper bound for λ, we have

$$(4.5)\quad \lambda \leqq \frac{\bar{\lambda}^{(h)}}{1-3h^2\bar{\lambda}^{(h)}}.$$

Thus, an upper bound for λ can be obtained from the simpler problem (4.2). Furthermore, the problems (2.4) for $\bar{\lambda}^{(h)}$ and (4.2) for $\lambda^{(h)}$ differ only in their domains. It is clear that

$$(4.6)\quad R_h \supset \overline{R_h},$$

and that they are close to each other for small h. We shall give an estimate for $\lambda^{(h)}-\bar{\lambda}^{(h)}$ under certain conditions.

Let R undergo a homothetic transformation (stretching) of ratio $1 : (1+\alpha)$. It is easily seen that the lowest eigenvalue of the new domain is $\lambda/(1+\alpha)^2$. Suppose that α is so chosen that all the mesh points on the boundary of R_h having at least one neighboring mesh point interior to R_h are contained in the new region $(1+\alpha)R$. Then R_h serves as an $\overline{R_h}$ to $(1+\alpha)R$

and we have by (4.5)

$$\lambda \leqq (1+\alpha)^2 \frac{\lambda^{(h)}}{1-3h^2\lambda^{(h)}}. \tag{4.7}$$

This, together with (2.10) or one of its improvements, gives upper and lower bounds for λ in terms of the single number $\lambda^{(h)}$. However, these bounds are close to each other only if α is small.

In order to estimate α, we assume that R is convex. We further suppose that R_h is minimal with respect to the condition (2.1). Geometric considerations then show that one can take

$$\alpha = \frac{3h}{\bar{r}} \tag{4.8}$$

where $\bar{r}$ is the radius of the largest circle interior to R. Thus, for a convex region, the upper bound (4.7) differs from the lower bound $\lambda^{(h)}$ only by terms of $O(h)$. The difference can be estimated *a priori* (using some crude bound for λ) and the appropriate mesh size h for a given error can be determined. A single finite difference problem is then solved to given both the upper and the lower bound.

In order to treat non-convex regions we turn to the question of the variation in $\lambda^{(h)}$ when R_h is decreased while h is kept fixed.

5. A Bound for the Variation of the Finite Difference Problem

Consider two finite difference domains R_h and R_h' with

$$R_h' \subset R_h. \tag{5.1}$$

Let the lowest eigenvalues of the finite difference membrane problem on these domains be denoted by $\lambda^{(h)}$ and $\lambda^{(h)\prime}$, respectively. Let w be the eigenfunction on R_h corresponding to $\lambda^{(h)}$, normalized by

$$h^2 \sum w^2 = 1. \tag{5.2}$$

To obtain an upper bound for $\lambda^{(h)\prime}$ we insert in the minimum principle the function

$$\bar{w}(x_i, y_j) = \begin{cases} w(x_i, y_j) & \text{at interior points of } R_h', \\ 0 & \text{elsewhere.} \end{cases} \tag{5.3}$$

We have

$$\lambda^{(h)\prime} \leqq \frac{D^*(\bar{w}, \bar{w})}{h^2 \sum \bar{w}^2} \leqq \frac{D^*(w, w) + \sum'' w^2}{h^2(\sum w^2 - \sum (w-\bar{w})^2)} = \frac{\lambda^{(h)} + \sum'' w^2}{1 - h^2 \sum (w-\bar{w})^2}, \tag{5.4}$$

where $\sum''$ denotes summation over boundary points of R_h'. We estimate the two sums on the right for a particular class of variations. Let R_h' consist of all mesh points in R_h such that the length of any path consisting of grid lines to a boundary point is at least kh (k an integer). Let S_t be the sum of w^2

over points whose minimum distance along grid lines to the boundary of R_h is exactly l. Then (5.4) becomes

$$\lambda^{(h)\prime} \leqq \frac{\lambda^{(h)}+S_k}{1-h^2(S_1+\cdots+S_k)}. \tag{5.5}$$

Since w satisfies the difference equation (2.4) and $S_0 = 0$,

$$S_k \leqq 16S_{k-1} \leqq \cdots \leqq 16^{k-1}S_1, \tag{5.6}$$

so that

$$\lambda^{(h)\prime} \leqq \frac{\lambda^{(h)}+16^{k-1}S_1}{1-\dfrac{16^{k-1}-1}{15}h^2S_1}. \tag{5.7}$$

Thus, if S_1 is small, this upper bound for $\lambda^{(h)\prime}$ is close to $\lambda^{(h)}$, which, because of (5.1), is a lower bound for $\lambda^{(h)\prime}$. We shall now estimate S_1 and show it to be small under certain assumptions on R_h. We begin with the identities (3.8) and (3.14) with $\frac{1}{2}h$ replaced by h and q by w:

$$\begin{aligned} &-2h\sum{}''\Delta_h w(x_i, y_j)\{x_i[w(x_i+h, y_j)-w(x_i, y_j)] \\ &\qquad +y_j[w(x_i, y_j+h)-w(x_i, y_j)]+h^3\Delta_h w(x_i, y_j)\} \\ &= \sum[w(x_i+2h, y_j)-w(x_i, y_j)]^2+[w(x_i, y_j+2h)-w(x_i, y_j)]^2 \\ &\leqq \frac{4\lambda^{(h)}\left(1-\lambda^{(h)}\dfrac{h^2}{8}\right)}{1+\dfrac{h}{2\varrho}}, \end{aligned} \tag{5.8}$$

because of the normalization (5.2). We again define

$$\xi = -\mathbf{r}\cdot\mathbf{n}-\tfrac{1}{2}h \tag{5.9}$$

at boundary points where R_h has a unique inward normal $\mathbf{n}$ but not at all at non-reentrant corners. At reentrant corners define ξ_1 and ξ_2 by using inward vectors in the x and y direction, respectively, and let ξ be the smaller of the two. We now let

$$\sigma = \inf \xi \tag{5.10}$$

and assume that

$$\sigma > 0. \tag{5.11}$$

This implies that R_h is star-shaped. The maximum and minimum values of x must be attained for $y = 0$, and the maximum and minimum values of y for $x = 0$. Furthermore, $\sigma+\frac{1}{2}h$ is the smallest absolute value of the x-coordinate of a point whose y-coordinate does not have the maximum or minimum value or of the y-coordinate of a point whose x-coordinate

does not have the maximum or minimum value. From (5.8)

$$S_1 \leqq \frac{2h\lambda^{(h)}\left(1-\lambda^{(h)}\frac{h^2}{8}\right)}{\sigma\left(1+\frac{h}{2\varrho}\right)} \tag{5.12}$$

follows. This can be made small, provided $h/\sigma \to 0$ as $h \to 0$.

Applying these results in order to find an upper bound for λ, we seek a k such that

$$R_h' \subset \overline{R_h} \tag{5.13}$$

so that

$$\bar{\lambda}^{(h)} \leqq \lambda^{(h)'} \tag{5.14}$$

and hence

$$\lambda \leqq \frac{\lambda^{(h)'}}{1-3h^2\lambda^{(h)'}}. \tag{5.15}$$

For this purpose, we assume that (a) the maximum and minimum of y on the boundary of R occur at $x = 0$ and that the maximum and minimum of x occur at $y = 0$, (b) the curvature of the boundary of R be bounded in the neighborhood of these four vertices, (c) the intersection of R with any line parallel to the x or y axis is simply connected, and (d) the points of maximum and minimum x and y all lie on mesh lines. The last condition requires that the x and y diameters of R be multiples of h. This can be done by choosing h properly and stretching R in either the x or y direction by a factor of at most $1+h/d$ where d is the larger of the diameters in the x and y directions. This stretching reduces λ by a factor not smaller than $(1+h/d)^{-2}$.

Let R_h be the minimal region satisfying (2.1). Then conditions (a), (b) and (d) imply that

$$\sigma \geqq \left(\frac{2h}{c}\right)^{1/2} \tag{5.16}$$

where c is a bound for $|d^2y/dx^2|$ when y is within h of its maximum or minimum value and for $|d^2x/dy^2|$ when x is within h of its maximum or minimum value on the boundary of R.

It follows from condition (c) that if we take

$$k = 2, \tag{5.17}$$

(5.13) is satisfied. Thus, using (5.14), (5.7), (5.12) and (5.16), we have

$$\bar{\lambda}^{(h)} \leqq \lambda^{(h)}\left\{1+16(2c)^{1/2}h^{1/2}+\frac{h}{2\varrho}\right\}\left\{1-(2c)^{1/2}\lambda^{(h)}h^{5/2}+\frac{h}{2\varrho}\right\}^{-1}. \tag{5.18}$$

Recalling the contraction performed to satisfy condition (d), we have

$$\lambda \leqq \left(1+\frac{h}{d}\right)^2 \frac{\bar{\lambda}^{(h)}}{1-3\,\bar{\lambda}^{(h)}\,h^2}. \tag{5.19}$$

This bound is arbitrarily close to $\lambda^{(h)}$. Hence, we again find arbitrarily close upper and lower bounds for λ by choosing h sufficiently small and solving a single finite difference problem for h.

Bibliography

[1] Courant, R., *Variational methods for the solutions of problems of equilibrium and vibrations*, Bull. Amer. Math. Soc., Vol. 49, 1943, pp. 1–23.

[2] Forsythe, G. E., *Asymptotic lower bounds for the fundamental frequency of convex membranes*, Pacific J. Math., Vol. 5, 1955, pp. 691–702.

[3] Hersch, J., *Équations différentielles et fonctions de cellules*, C. R. Acad. Sci. Paris, Vol. 240, 1955, pp. 1602–1604.

[4] Payne, L. E., and Weinberger, H. F., *New bounds in harmonic and bi-harmonic problems*, J. Math. Physics, Vol. 33, 1955, pp. 291–307.

[5] Polya, G., *Sur une interprétation de la méthode des différences finies qui peut fournir des bornes supérieures ou inférieures*, C. R. Acad. Sci. Paris, Vol. 235, 1952, pp. 995–997.

[6] Rellich, F., *Darstellung der Eigenwerte von $\Delta u + \lambda u$ durch ein Randintegral*, Math. Z., Vol. 46, 1940, pp. 635–643.

[7] Synge, J. L., *Approximations in boundary value problems by the method of the hypercircle in function space*, Rend. Circ. Mat. Palermo, Ser. V, Vol. 10, 1951, pp. 24–47.

[8] Weinberger, H. F., *Lower bounds for minima by finite difference methods*, Bull. Amer. Math. Soc., Vol. 61, 1955, p. 301.

Received January 20, 1956.

COMMUNICATIONS ON PURE AND APPLIED MATHEMATICS, VOL. IX, 625–632 (1956)

Some New Methods in Two-Dimensional Variational Problems with Special Reference to Minimal Surfaces*

L. C. YOUNG
University of Wisconsin

1. Introduction

1. Apart from developments which were largely formal, the first real advance in parametric two-dimensional variational problems was made some twenty or more years ago in the form of an existence theorem for the *problem of least area*, subject to the artificial condition that only surfaces would be admitted whose characteristic was less than a preassigned bound. In the still more special case of surfaces of the type of the disc, a similar existence theorem has been obtained during the last few years for more general regular problems, by Sigalov [1], Danskin [2] and Cesari [3]. In a Memoir [4] to appear shortly all these results are extended to non-regular problems, except for minor differences concerning the interpretation of boundaries (see §2 below).

However, when the artificial condition concerning the characteristic is removed, it is now known, as the result of an example due to Fleming [5] which will be published soon, that no general existence theorem is true even for the problem of least area, as long as we retain the traditional notion of surface. Thus the simplest problem of the subject, that of least area, already demands the admission of *generalized surfaces*.

We are concerned with problems of the type

$$L(f_0) = \min,$$

where f_0 is an element of the space of continuous functions $f = f(x, j)$, positively homogeneous in j, with arguments x, j which are vectors[1] in 3-space, and where

$$L(f) = \iint_{\text{unit square}} f[x(u, v), j(u, v)]\, du\, dv, \quad j(u, v) = [x_u, x_v] \text{ vector product.}$$

The minimum of $L(f_0)$ is taken for all admissible $x(u, v)$ and the latter are

*The work reported here was sponsored in part by the Office of Ordnance Research, United States Army, under Contract DA-11-022-ORD-1511.

[1] j is really a bivector.

sufficiently smooth functions (which we take here to have, say, continuous partial derivatives) subject to certain boundary conditions.

In the problem of least area f_0 is the particular function $f_0(x, j) = |j|$. However, it is peculiar to our approach, that we concern ourselves with the *linear functional* $L(f)$ and not merely with the number $L(f_0)$. This linear functional is, for variational purposes, the very essence of the notion of surface: in fact when two $x(u, v)$ define the same functional $L(f)$, they are absolutely equivalent for variational purposes, and vice-versa. Therefore we by-pass all questions of change of parameter by *identifying* the notion of a parametric surface with the corresponding linear functional $L(f)$ derived by surface integration. Our problem is now that of the minimum of the number $L(f_0)$ when $L(f)$ is an admissible surface. This is a problem in the conjugate space to that of the functions f, and since each problem treats one f_0 at a time, the most appropriate topology is the *weak topology*.

For later purposes, the integral derived from that which defines $L(f)$ in terms of $x(u, v)$ by changing the range of integration to a closed subset of the unit square, will be termed *elementary subsurface*. Further we term *linear* an $f(x, j)$ of the form of a scalar product $(p(x) \cdot j)$ and we term *exact* an f of this form for which $p(x)$ is the *curl* of a vector-function $q(x)$; such f are, of course, still assumed to belong to our space of functions.

2. Our admissible surfaces $L(f)$ were to be subject to boundary conditions. However, these conditions cannot really be properly formulated in terms of $x(u, v)$, and this is why previous work contains unnatural restrictions (absence of multiple boundary-points) and what seems to us a misleading generality (inclusion of "heavy" boundary-curves). (A problem with a boundary-curve which is "heavy", i.e. of positive, or infinite, two-dimensional measure, would here come under a different heading; we would regard the true boundary as variable in a certain abstract set. This difference has to be borne in mind in comparing results based on our methods with the results of Sigalov, Danskin, Cesari and others.) The unsatisfactory nature of the traditional notion of boundary is evidenced by the fact that when C is "heavy" and $L_1(f)$, $L_2(f)$ are surfaces with the boundaries C, $-C$, the surface $L_1(f)+L_2(f)$ does not generally have in the traditional sense the boundary $C+(-C) = 0$. Similarly the traditional notion of boundary does not in general satisfy the following formulation of *Stokes' Theorem*: if f_0 is exact, $L(f_0)$ depends only on the boundary of $L(f)$.

We shall ensure very simply (cf. [6]; also [7]) the universal validity of this formulation of Stokes' Theorem. We define the *boundary* of $L(f)$ as the *restriction of* $L(f)$ *to the class of exact* f. It is convenient also to define the *track* of $L(f)$ as the restriction of $L(f)$ to the class of linear f.

Admissible surfaces $L(f)$ are now linear functionals with a given restriction to the class of exact f and *are derived from sufficiently smooth parametric representations* $x(u, v)$. Clearly this last condition is now out of place: as stated before, we require *generalized* surfaces, and these are simply the weak limits of sequences $\{L_n(f)\}$ of elementary surfaces. It turns out that a generalized surface is then merely *any non-negative linear functional* $L(f)$, i.e., one for which $L(f) \geqq 0$ whenever $f \geqq 0$. Its boundary will, as before, be defined as its restriction to the class of exact f, and its track as its restriction to the class of linear f. One may wonder whether this generalization of the notion of surface has gone too far. We shall see that this is not the case (§3). Historically more and more general curves and surfaces were gradually admitted in the calculus of variations as they began to appear in everyday analysis. The present extension is in keeping with the fact that linear functionals occur so much in analysis today. If by any chance our solutions turn out to be elementary, certainly no harm is done; but actually they are even more important in the opposite case, when they may be regarded as *hidden solutions* of the original problem. Thus the engineer's solution to the problem of the fastest road up a mountain, when the gradient is restricted to be not too steep, is an approximate one, consisting of a zig-zag which tends in the limit to a hidden solution; this is of course a problem for curves, but the case of surfaces is quite similar.

2. General Characterization

3. We denote by $A(\lambda)$ and by $B(\lambda)$, the class of elementary surfaces and that of generalized surfaces, respectively, with the boundary λ. We write further $A^*(\lambda)$ for the sequential weak closure of $A(\lambda)$. The original problem was that of the minimum of $L(f_0)$ when $L(f) \in A(\lambda)$. However, to understand this problem properly and to include hidden solutions, we must of course replace in it $A(\lambda)$ by $A^*(\lambda)$. Instead we shall replace $A(\lambda)$ by the much wider class $B(\lambda)$, and we have to justify this.

Given a set E of linear functionals $L(f)$ and given a measure α on E we term *centre of gravity* of E, for this measure, the integral

$$\int L(f)\, d\alpha$$

if it exists; the quantity $\alpha(E)$ is termed the corresponding mass.

The following result has been proved jointly with Fleming [8]: *The relation* $A(\lambda) \neq 0$ *implies that* $B(\lambda)$ *is the set of the centres of gravity of* $A^*(\lambda)$, *with unit mass if* $\lambda \neq 0$, *otherwise with arbitrary mass.*

From this result it follows that the minimum of $L(f_0)$ for $L(f) \in A^*(\lambda)$, where $A(\lambda) \neq 0$, is identical with that for $L(f) \in B(\lambda)$ and that the solutions of either problem determine in simple manner those of the other.

4. Henceforth we concern ourselves with $B(\lambda)$. It would be useful to characterize in some way the union of those sets $B(\lambda)$ for which $A(\lambda) \neq 0$, i.e., to characterize the generalized surfaces with elementary boundaries. This would commit us more deeply as to the precise definition of elementary surface, and we prefer, as a first contribution, to replace the notion of elementary boundary by a much wider one, which has to be considered in any case if we wish our admissible boundaries to constitute a complete linear space.

By a σ-polyhedron, we mean a countable sum

$$\sum a_k L_k(f)$$

where the a_k are positive reals and the $L_k(f)$ are bounded triangles. We denote by Λ the class of boundaries of σ-polyhedra. It is easily seen that Λ includes all conceivable elementary boundaries.

We denote by $\mathscr{E}$ the class of generalized surfaces with the same tracks as elementary subsurfaces (end of §1). The following *characterization* has been obtained jointly with Fleming [8]:

$$\sum_{\lambda \in \Lambda} B(\lambda) = (\text{set of centres of gravity of } \mathscr{E}).$$

3. Homology Principle

5. In deriving differential equations and the like, we shall follow the natural approach adopted for curves by Carathéodory. This leads to conditions both necessary and sufficient which involve, however, in the general case, the use of certain simple types of *discontinuous* integrands $f(x, j)$; for such integrands we denote by $\bar{f}(x, j)$, for some $h > 0$, the mean value of $f(x+h\xi, j)$, over the unit cube of centre 0, for the vector ξ. For the relevant f it so happens that $\bar{f}$ is continuous and that the limit

$$(L, f) = \lim_{h \to 0} L(\bar{f})$$

exists for all surfaces L which we have occasion to use. Of course, $(L, f) = L(f)$ when f is continuous.

We shall make a few minor changes of notation. We shall allow f to be on occasion a vector-function, which we suppose continuous and positively homogeneous, and we extend the symbol $L(f)$ in the obvious way to this case, as a vector with the appropriate components. Furthermore, given, in the subspace of our original space of functions $f(x, j)$ consisting of functions $\psi(j)$ independent of x, a non-negative linear functional $M(\psi)$, we write for each x, $M_x(f)$ for the value of $M(\psi)$ when ψ is derived by fixing x in $f(x, j)$, and we term $M_x(f)$ a generalized surface concentrated at the point x. When we substitute for $f(x, j)$ the vector j, the value obtained for $M_x(f)$ is a vector

termed *oriented area* of this concentrated surface, and the direction of the oriented area is termed *orientation*. The latter is either a unit vector or 0.

It follows from general theorems on linear functionals that every generalized surface $L(f)$ has the form

$$\int M_x(f)\, d\nu$$

where ν is a measure in x-space, vanishing outside some sphere. Hence, denoting by $\theta(x)$ the orientation of $M_x(f)$, there exists a finite measure μ, vanishing outside some sphere, such that, for linear f, $L(f)$ becomes

$$\int f[x, \theta(x)]\, d\mu.$$

In this last expression we shall allow $f(x, j)$ to be an arbitrary integrable function and we denote the result by [track L, f]. We observe that μ and $\theta(x)$ are determined in a virtually unique manner by the track of L and not merely by L itself.

6. We term a function $\varphi(x, j)$, which need not be continuous[2], *homologous to* 0, if it satisfies the following conditions:

(a) φ is bounded in the set $|x| < a$, $|j| = 1$, where a is any positive number,

(b) $\varphi(x, tj) = t\varphi(x, j)$ for $t \geqq 0$,

(c) for each linear function $x(u, v)$ whose constant Jacobian $[x_u \times x_v]$ is $j \neq 0$, the composite function $\varphi[x(u, v), j]$ is defined for almost all (u, v) and is the derivative of its integral with respect to (u, v),

(d) we have for every polyhedron $L(f)$ (polyhedron in the classical sense: finite sum of triangles) whose boundary vanishes

$$[\text{track } L, \varphi] = 0.$$

We shall be concerned with functions of the form $f+\varphi$, where f is continuous and φ is homologous to 0; we call such a function *homologous to a continuous function*, namely to f. By combining the characterization at the end of §4 with some recent results[3] we find:

(i) The quantity $(L, f+\varphi)$ exists whenever $L \in B(\lambda)$, $\lambda \in \Lambda$ and $f+\varphi$ is homologous to a continuous function.

(ii) If L is translated through a vector ξ, $(L, f+\varphi)$ varies continuously with ξ.

(iii) If φ is homologous to 0, $\lambda \in \Lambda$ and L', L'' both belong to $B(\lambda)$, then $(L', \varphi) = [\text{track } L'', \varphi]$.

---

[2] φ is not everywhere defined.

[3] See [9]. The relevant statements are (4.1), (4.2), (4.3), (6.1) and (5.1).

From this last statement it follows that, for each φ homologous to 0, the quantity (L, φ) is a linear functional $g(\lambda)$ for $\lambda \in \Lambda$, where L is chosen arbitrarily in $B(\lambda)$, and that this quantity can also be written as [track L, φ]. Conversely, it turns out[3] that all linear functionals $g(\lambda)$ $(\lambda \in \Lambda)$ can be so expressed and that the classical Hahn-Banach Theorem yields the following *homology principle*:

If $\lambda_0 \in \Lambda$, $L_0 \in B(\lambda_0)$ and $L_0(f_0) \leqq L(f_0)$ for all $L \in B(\lambda_0)$, then there exists a non-negative $f_0+\varphi$ homologous to f_0 such that $(L_0, f_0+\varphi) = 0$.

Evidently the converse also holds: If for a non-negative $f_0+\varphi$ homologous to f_0 we have $(L_0, f_0+\varphi) = 0$ where $L_0 \in B(\lambda_0)$, $\lambda_0 \in \Lambda$, then $L_0(f_0) \leqq L(f_0)$ for all $L \in B(\lambda_0)$.

We observe further that while φ is not everywhere defined, it can be shown [8], by a somewhat cumbersome method, that any φ homologous to 0 can be replaced by a scalar product $(p(x), j)$, where $p(x)$ is everywhere defined but in general discontinuous, such that, whenever $x(u, v)$ is Lipschitzian and $j(u, v)$ is its Jacobian, the relation $\varphi[x(u, v), j(u, v)] = (p[x(u, v)] \cdot j(u, v))$ holds almost everywhere in (u, v). A vector-function $p(x)$ thus associated with a φ homologous to 0 will be termed a *field*.

7. For simplicity we now limit ourselves to the problem of least area, $f_0(x, j) = |j|$, and we suppose as in the homology principle that L_0 is a minimizing generalized surface with boundary in Λ. By our characterization theorem (end of §4), L_0 is a centre of gravity of $\mathscr{E}$,

$$L_0(f) = \int L(f)\, d\alpha.$$

A similar formula holds for $(L_0, f_0+\varphi)$; it follows that $(L, f_0+\varphi) = 0$ almost everywhere in $\mathscr{E}$ relative to the measure α, and hence that L_0 is a centre of gravity of the subset of $\mathscr{E}$ consisting of generalized surfaces $L(f)$ for which $(L, f_0+\varphi) = 0$. Since each of these is a minimizing generalized surface by the converse of the homology principle, it follows easily, from the fact that f_0 is regular, that every such $L(f)$ reduces to the elementary subsurface with the same track, i.e., that

$$L(f) = \iint f[x(u, v), j(u, v)]\, du\, dv$$

where $x(u, v)$ has continuous partial derivatives and Jacobian $j(u, v)$, and where integration is over a closed set. It follows that

$$\begin{aligned} 0 = (L, f_0+\varphi) &= L(f_0)+[\text{track } L, \varphi] \\ &= \iint \{f_0[x(u, v), j(u, v)]+\varphi[x(u, v), j(u, v)]\}\, du\, dv, \end{aligned}$$

and, therefore, that the non-negative function $f_0+\varphi$ vanishes for almost all

(u, v). We now write $\varphi = -(p(x) \cdot j)$ for all relevant (u, v) and, by decomposing the given closed set into a countable sum of closed subsets together with a set of plane measure 0 and a set in which $j(u, v) = 0$, the last two of which contribute nothing to the expression for $L(f)$ as an integral, we express each of our $L(f)$ as a countable sum of corresponding elementary subsurfaces and, therefore, as their centre of gravity. By composition we express $L_0(f)$ as centre of gravity, with a new measure α, of the set of these new elementary subsurfaces which we continue to denote by the same symbol $L(f)$.

Each of these $L(f)$ has the form given above where, in addition, by suitable choice of our decompositions, the following conditions are satisfied:

$$j(u, v) \neq 0$$

and

$$|j(u, v)| = f_0[x(u, v), j(u, v)] = -\varphi[x(u, v), j(u, v)] = (p[x(u, v)] \cdot j(u, v)).$$

Since, furthermore, $f_0+\varphi$ is non-negative, we find that $|q(x)| \leqq 1$ and our equations reduce, by the converse of Schwarz's inequality, to

$$\frac{j(u, v)}{|j(u, v|} = p[x(u, v)],$$

a relation which implies, in particular, that $p[x(u, v)]$ is continuous and of absolute value unity.

Hence, a necessary and sufficient condition in order that a generalized surface $L_0(f)$ with boundary in A be a solution of the problem of least area is that $L_0(f)$ be, for some field $p(x)$, where $|p(x)| \leqq 1$, the centre of gravity of elementary subsurfaces satisfying the conditions

$$j(u, v) \neq 0, \qquad \frac{j(u, v)}{|j(u, v)|} = p[x(u, v)]$$

at every (u, v) of the corresponding closed sets. We can thus regard $L_0(f)$ as made up by piecing together parts of certain minimal surfaces in a slightly more general sense than usual[4], since $p(x)$ is not differentiable (in fact discontinuous) in general.

Bibliography

[1] Sigalov, A. G., *Two-dimensional problems of the calculus of variations*, Uspehi Matem. Nauk, N. S, Vol. 6, 1951, pp. 16–101.

[2] Danskin, J. M., *On the existence of minimizing surfaces in parametric double integral problems of the calculus of variations*, Rivista Mat. Univ. Parma, Vol. 3, 1952, pp. 43–63.

[4] The usual equations require interpreting derivatives "à la Laurent Schwartz".

[3] Cesari, L., *An existence theorem of calculus of variations for integrals on parametric surfaces*, Amer. J. Math., Vol. 74 1952, pp. 265–295.
[4] Young, L. C., *On Generalized Surfaces of Finite Topological Type*, Mem. Amer. Math. Soc., No. 17, 1955.
[5] Fleming, W. H., *An example in the problem of least area*, submitted to Proc. Amer. Math. Soc.
[6] Fleming, W. H., and Young, L. C., *A generalized notion of boundary*, Trans. Amer. Math. Soc., Vol. 76, 1954, pp. 457–484.
[7] de Rham, G., and Kodaira, K., *Harmonic Integrals*, Lectures, Institute for Advanced Study. Princeton, 1950.
[8] Fleming, W. H., and Young, L. C., *Representations of generalized surfaces as mixtures*, submitted to Rend. Circ. Mat. Palermo
[9] Young, L. C , *A variational algorithm*, Rivista Mat. Univ. Parma, Vol. 5, 1954, pp. 255–268.

Received January 5, 1956.