LA-3685

C.35

**LOS ALAMOS SCIENTIFIC LABORATORY
of the
University of California**
LOS ALAMOS • NEW MEXICO

Error Bounds on Numerical Solutions of Dirichlet Problems for Quasilinear Elliptic Equations

UNITED STATES
ATOMIC ENERGY COMMISSION
CONTRACT W-7405-ENG. 36

This report expresses the opinions of the author or authors and does not necessarily reflect the opinions or views of the Los Alamos Scientific Laboratory.

Printed in the United States of America. Available from
Clearinghouse for Federal Scientific and Technical Information
National Bureau of Standards, U. S. Department of Commerce
Springfield, Virginia 22151
Price: Printed Copy $3.00; Microfiche $0.65

LA-3685
UC-32, MATHEMATICS
AND COMPUTERS
TID-4500

LOS ALAMOS SCIENTIFIC LABORATORY
of the
University of California
LOS ALAMOS • NEW MEXICO

Report written: August 1966

Report distributed: May 10, 1967

Error Bounds on Numerical Solutions of Dirichlet Problems for Quasilinear Elliptic Equations*

by

Thurman G. Frank

*Also presented as a dissertation to the faculty of the Graduate School of The University of Texas.

ABSTRACT

Let R be closed, bounded, simply connected region in the plane. Let P denote the Dirichlet problem $Au_{xx} + 2Bu_{xy} + Cu_{yy} = G$ on R in which A, B, C, G depend on x, y, u, u_x, u_y. It is assumed that A,B,C satisfy a uniform ellipticity condition and a condition (see L. Bers, F. John, and M. Scheichter, "Partial Differential Equations," Interscience Publ., 1964, pp. 262-264) which enables uniqueness of the solution of P to be established by means of a maximum principle; also it is assumed that R and the coefficient functions are such that $u \in C^4$ on R. Several finite difference analogues of P are studied which use, essentially, central differences except near the boundary. One such scheme uses the method of J. H. Bramble and B. E. Hubbard, "Contributions to Differential Equations," $\underline{2}$, 319-340, 1963, to treat the term $2Bu_{xy}$. It is shown that the solutions of the finite difference analogues converge, with decreasing mesh width h, to the solution of P. Moreover, the error is $0(h^p)$ with p either one or two depending on which particular combination of difference equations in the interior and at the boundary of R is used.

ACKNOWLEDGEMENTS

The research reported herein was presented as a dissertation to the faculty of the Graduate School of The University of Texas. It was accomplished during periods in which the author was employed by the Computation Center, The University of Texas, and by the Los Alamos Scientific Laboratory. Support was provided by the National Science Foundation and the Army Research Office (Durham) under grants awarded to The University of Texas and by the U. S. Atomic Energy Commission under contract with the Los Alamos Scientific Laboratory.

TABLE OF CONTENTS

LIST OF FIGURES

TABLE OF CONSTANTS

A large number of constants are used in this paper. These constants are listed below together with the number of the page on which they are defined.

Constant	Page
k_0, k_1	7
X, Y	16
Δ	21
k_0', η, I, J	26
$\bar{M}_n$, $\bar{N}_n$, $\bar{Q}_n$	42
k_2, k_3	46
k_4	47
k_5	50
k_6, k_7	52
k_8	53
K_0, K_1, h_1	54
σ_1	59
μ_1, h_2	62
σ_2, h_3	66
μ_2, h_4	74
K	81

CHAPTER I

INTRODUCTION

In this paper, we are concerned with the problem of extending the theoretical justification for using the method of finite differences to obtain numerical approximations to solutions of Dirichlet problems involving second order, quasilinear, uniformly elliptic partial differential equations in two independent variables on a closed, bounded, and simply connected region R.

The use of the method of finite differences results, through a process of "discretization," in numerical approximations to values of the solution of a given problem at a discrete set of points in the region associated with the problem. The points, called mesh points, at which numerical approximations are calculated are separated by a characteristic distance called the mesh width. The use of a finite difference method is theoretically justified if it can be shown that solutions of the resulting finite difference analogues corresponding respectively to successively smaller mesh widths converge to the solution of the continuous problem.

Until recently, the investigations of the convergence, with decreasing mesh width, of finite difference analogues of Dirichlet problems for elliptic partial differential equations have been concerned with linear equations. A brief review of the results of these investigations is helpful in placing similar studies for quasilinear equations in proper perspective.

[1] Some of the results contained in this dissertation were first announced in Abstract 66T-287, Notices, Amer. Math. Soc., 13, 496 (1966).

Studies of the convergence of solutions of finite difference analogues of linear elliptic partial differential equations can be roughly classified into two groups. The principal objective of the studies in the first of these two groups is to prove that the solutions of a finite difference analogue of a given continuous problem converge, as the mesh width is decreased to zero, to a solution of the given problem. One of the first[2] of these studies is reported in Courant, Friedrichs, and Lewy [1928].[3] In this, it is shown that solutions of a finite difference approximation to the Dirichlet problem for Laplace's equation converge, with decreasing mesh width, to the solution of the given problem. However, the proof given is nonconstructive in the sense that it provides no means by which the error in a solution of a finite difference analogue corresponding to a finite value of the mesh width can be estimated.

The studies included in the second group provide considerably more extensive results in that explicit estimates for the error, expressed either in terms of the boundary values and the shape of the region associated with the problem or by means of the solution of the continuous problem, are given. However, in order to get these "better" results, more conditions must be imposed on the coefficient functions, the boundary values, and the shape of the region. The usual requirement is that the solution of the continuous problem possess bounded partial derivatives up to fourth order. Of these studies, one of the earliest and best known is Gerschgorin [1930].[4] Gerschgorin establishes convergence, with decreasing mesh width, of solutions of

[2] See, also, Downing [1960].

[3] The use of brackets, [ ], indicates reference to the bibliography.

[4] See also Laasonen [1957], Rosenbloom [1952], Walsh and Young [1953], [1954], and Wasow [1952], [1957].

the finite difference analogue of the Dirichlet problem for Laplace's equation considered by Courant, et al. [1928], by showing that the error in a solution of the finite difference approximation corresponding to a finite value of the mesh width is majorized[5] by a function whose modulus is proportional to the product of the square of the mesh width and the maximum of the moduli of the fourth partial derivatives of the solution of the continuous problem. Gerschgorin also considers Dirichlet problems involving somewhat more general elliptic partial differential equations than Laplace's equation; however, fairly severe restrictions are placed on the coefficients in these equations.[6]

The investigations reported in this paper concerning finite difference analogues of Dirichlet problems for quasilinear elliptic partial differential equations are in the second of the two groups described above, whereas, other published results for quasilinear equations are, for the most part, in an intermediate position between these two groups.

Although the method of finite differences is widely used to obtain approximate solutions of Dirichlet problems for quasilinear elliptic partial differential equations, the theoretical justification for such procedures is very limited. The earliest published proof of the convergence, with decreasing mesh width, of such approximations is given in Bers [1953]. In this investigation, the use of the finite difference method for obtaining approximate solutions of the problem given by

[5] A function f is majorized by a function g in a region R if $|f| \leq g$ at every point in R.

[6] A more detailed discussion of Gerschgorin's results is given in Chapter XI where a comparison is made between these results and the results obtained in the sequel.

$$(1.1) \qquad \Delta u = F(x,y,u,\partial u/\partial x,\partial u/\partial y), \quad (x,y) \in R$$

$$(1.2) \qquad u = g(x,y) \qquad , \quad (x,y) \in S$$

where Δ denotes the Laplace operator $\partial^2/\partial x^2 + \partial^2/\partial y^2$, R is a simply connected, bounded region in the plane with boundary S, and g is a given continuous function on S is studied. It is shown that if the partial derivative of F with respect to u is nonnegative and if the partial derivatives of F with respect to $\partial u/\partial x$ and $\partial u/\partial y$ are uniformly bounded, then the Dirichlet problem for the finite difference equation obtained by replacing the derivatives of equation (1.1) with central divided differences has a unique solution and that the solution of this problem tends to a solution of the given problem as the mesh width is decreased.

Studies of the Dirichlet problem given by equations (1.1) and (1.2) where $F = F(x,y,u)$ are reported in Ablow and Perry [1959], Pohozaev [1960], Douglas [1961], Levinson [1963], McAllister [1964c], Parter [1964], and Greenspan and Parter [1965].

Ablow and Perry and Pohozaev consider the existence of a nonnegative solution of this problem where $F = u^2$ and g is nonnegative. The existence of a unique solution of the continuous problem is demonstrated. McAllister studies a discretized version of the same problem and proves convergence for an iteration scheme given by Ablow and Perry.

Douglas [1961] presents an algorithm for solving the nonlinear system of algebraic equations which arise from discretization of the problem given by equations (1.1) and (1.2) where $F = F(x,y,u)$ which utilizes the alternating direction implicit iteration method. The region R is taken to be the unit square in this study, and convergence of solutions of the discretized problem with decreasing mesh width is proved provided

$$\frac{\partial F}{\partial u}(x,y,u) \geqq A > -2\pi \text{ , } (x,y) \in R.$$

In Levinson [1963], the problem studied by Douglas [1961] is treated analytically for a more general region subject to certain smoothness conditions and the condition

$$\lim_{|u| \to \infty} \inf \frac{F(x,y,u)}{u} \geqq 0, \quad (x,y) \in R.$$

Levinson proves that the given problem has a bounded solution u of class C_2 in R and of class C in $R + S$.

The results reported in Parter [1964] and in Greenspan and Parter [1965] consist of extensions and applications of Levinson's results. The behavior of solutions of finite difference analogues of the problem considered by Levinson are studied, and convergence of these solutions, as the mesh width is decreased, to a solution of the continuous problem is established provided the continuous problem is assumed to have a unique solution and the finite difference equations are of "positive type."[7]

The solution of a finite difference analogue of a Dirichlet problem involving an elliptic partial differential equation containing a different type of nonlinearity from those listed above is reported in Young and Wheeler [1964]. The use of the Peaceman-Rachford method to solve the linear systems which arise together with the use of "natural iteration"[8] is investigated as a means of solving a finite difference analogue of the problem given by

[7] For a definition of finite difference equations of "positive type," see Forsythe and Wasow [1960] or Chapter III of the sequel.

[8] The method of "natural iteration" is described in Chapter IV.

$$(\partial/\partial x)(W\ \partial u/\partial x) + (\partial/\partial y)(W \partial u/\partial y) + 1 = 0, \quad (x,y) \in R$$

$$W = [(\partial u/\partial x)^2 + (\partial u/\partial y)^2]^{(n-1)/2}, \quad 0 < n \leq 1, \quad (x,y) \in R$$

$$\int_0^1 \int_0^1 u(x,y)\ dxdy = 1$$

$$u(x,y) = 0, \qquad\qquad (x,y) \in S$$

where R is the unit square with boundary S. However, no convergence proofs are given in this paper.

There are two closely related papers, McAllister [1964a], [1964b], in which the convergence, with decreasing mesh width, of solutions of finite difference approximations to Dirichlet problems for equations of the form[9]

$$A(x,y,u,\partial u/\partial x,\partial u/\partial y)\partial^2 u/\partial x^2 + 2B(---)\partial^2 u/\partial x\partial y + C(---)\partial^2 u/\partial y^2$$
$$- \gamma(---)u = 0$$

is studied. The coefficients are Lipschitz functions of their arguments, A and C satisfy relations of the form

$$K_0 \geq A(---), \quad C(---) \geq \mu > 0$$

uniformly in the arguments, and

$$|B| < \mu/2.$$

In McAllister [1964a], $\gamma = 0$ and the boundary values are required

[9] The notation used here of denoting the arguments of a function by (---) will be used frequently when several functions of the same set of arguments occur in an equation or series of terms.

to satisfy a three-point condition.[10] In McAllister [1964b], $B = 0$, and the arguments of the coefficients are (x,y,u), and if d is the diameter of the region of the problem, it is necessary that $d^2 < \mu/2$.

There are many more individual results which are concerned with the solutions of particular problems.

In this paper, we consider finite difference analogues of the Dirichlet problem for the following quasilinear partial differential equation

$$(1.3) \qquad A(x,y,u,\partial u/\partial x,\partial u/\partial y)\partial^2 u/\partial x^2 + 2B(---)\partial^2 u/\partial x\partial y + C(---)\partial^2 u/\partial y^2 = G(---)$$

which is assumed to be uniformly elliptic.[11]

The techniques and results presented include the following:

Two finite difference approximations to equation (1.3) are given constructively which agree with the differential equation to terms which are $O(h^2)$[12] where h is the mesh width. The existence of the first of these two approximations, which is of nonnegative type, is proved in Bramble and Hubbard [1963]. The second approximation presented is more convenient for practical use for some problems than the one due to Bramble and Hubbard but is not necessarily of nonnegative type.

[10] A three-point condition on the boundary values is defined in Chapter XI, part 3.

[11] Equation (1.3) is uniformly elliptic if there exist positive constants k_0, k_1 such that

$$k_1(\xi^2+\eta^2) \geq A\xi^2 + 2B\xi\eta + C\eta^2 \geq k_0(\xi^2+\eta^2)$$

for all real ξ and η and for all permissible values of the arguments of A, B, and C.

[12] Here, as usual, we say that $f(t) = O(g(t))$ as $t \to a$ if there exists a number M such that $|f(t)/g(t)| < M$ for all t sufficiently close to a.

Two methods for formulating finite difference approximations at mesh points near the boundary are considered. These include a linear interpolation scheme due to Collatz [1933] and asymmetric approximations to equation (1.3) which agree with the equation to terms which are $O(h)$.

Two finite difference analogues of the Dirichlet problem for equation (1.3) are analyzed. These two finite difference boundary value problems utilize the approximation due to Bramble and Hubbard at interior mesh points and differ according to the approximations used at mesh points near the boundary. They are denoted as problems P_1 and P_2.

The principal results obtained are the theorems, for sufficiently small mesh width, of the existence of solutions of each of the finite difference problems and the derivation of bounds for the errors in these solutions.

Error bounds are derived which are proportional to the product of h^p, $p \geq 1$, and the maximum of the moduli of the fourth partial derivatives of the solution of the continuous problem for each of the finite difference analogues considered.

Those aspects of the analysis which are believed to be new are:

(i) The partial differential equation studied is more general than previously reported investigations of finite difference approximations to quasilinear elliptic partial differential equations.

(ii) Convergence, with decreasing mesh width, of solutions of the finite difference analogues to the solution of the continuous problem is proved by means of error bounds which are $O(h^p)$.

(iii) The only restrictions placed on the region of the problem are that it be closed, bounded, and simply connected and that the boundary of the region be sufficiently smooth that the solution of the continuous problem has bounded and continuous fourth partial derivatives.

(iv) This is the first study in which the Brouwer Fixed Point Theorem is used to obtain error bounds directly.

The principal limitations of the study are:

(i) The smoothness requirements on the boundary of the region of the problem are frequently not met in practical applications.

(ii) It is required that the functions A, B, C, and G in equation (1.3) satisfy a condition which is sufficient to guarantee that the solution of the continuous problem satisfies a maximum principle. This condition involves the solution of the problem itself; thus, it is sometimes necessary to examine these functions after a solution is obtained in order to verify that all requirements are satisfied.

A brief outline is given below of the arguments and techniques which are used in this study.

First, the mixed derivative term is eliminated from equation (1.3) by the introduction of a third independent variable z. This is done in such a way that a finite difference analogue of the continuous problem which is of nonnegative type can be formulated. The variable z is specified by specifying the angle τ between the z and x axes at each mesh point.

The transformed equation has the form

$$\begin{aligned} A'(x,y,u,\partial u/\partial x,\partial u/\partial y)\partial^2 u/\partial x^2 + 2B'(---)\partial^2 u/\partial z^2 \\ + C'(---)\partial^2 u/\partial y^2 = G(---). \end{aligned} \tag{1.4}$$

The finite difference boundary value problems, P_1 and P_2, are formulated for the transformed equation (1.4). Next, finite difference equations are derived for the error E which is defined by

$$E = U - u$$

where U denotes a solution of one of the finite difference analogues of the continuous problem and u denotes the solution of the continuous problem. The functions U are replaced by $(u+E)$ in each of the finite difference equations comprising each of the problems P_1 and P_2. In each case the finite difference equation for the error has the form

$$\begin{aligned}&A'(x_i,y_j,(u_{i,j}+E_{i,j}),D_x(u_{i,j}+E_{i,j}),D_y(u_{i,j}+E_{i,j}))D_x^2(u_{i,j}+E_{i,j})\\ (1.5)\quad &+ 2B'(---)D_z^2(u_{i,j}+E_{i,j}) + C'(---)D_y^2(u_{i,j}+E_{i,j}) = G(---)\end{aligned}$$

where D_x, D_x^2, etc. denote applicable finite difference approximations to $\partial/\partial x$ and $\partial^2/\partial x^2$ respectively, etc.

The finite difference equations for the error are rewritten by representing the functions A', B', C', and G in equation (1.5) in terms of a definite integral. We illustrate the technique used by considering the function

$$A'(x_i,y_j,(u_{i,j}+E_{i,j}),D_x(u_{i,j}+E_{i,j}),D_y(u_{i,j}+E_{i,j})).$$

Assume that the first partial derivatives of A' are continuous and let

$$A(\theta) = A'(x_i,y_j,(u_{i,j}+\theta E_{i,j}),D_x(u_{i,j}+\theta E_{i,j}),D_y(u_{i,j}+\theta E_{i,j})).$$

Then

$$
\begin{aligned}
&A'(x_i,y_j,(u_{i,j}+E_{i,j}),D_x(u_{i,j}+E_{i,j}),D_y(u_{i,j}+E_{i,j}) \\
&= A'(x_i,y_j,u_{i,j},D_xu_{i,j},D_yu_{i,j}) + \int_0^1 (dA/d\theta)d\theta \\
(1.6)\qquad &= A'(x_i,y_j,u_{i,j},D_xu_{i,j},D_yu_{i,j}) + E_{i,j}\int_0^1 A'_r\,d\theta \\
&+ D_xE_{i,j}\int_0^1 A'_pd\theta + D_yE_{i,j}\int_0^1 A'_q\,d\theta
\end{aligned}
$$

where

$$A'_r = \frac{\partial}{\partial(u_{i,j}+\theta E_{i,j})}\ A'(x_i,y_j,(u_{i,j}+\theta E_{i,j}),D_x(u_{i,j}+\theta E_{i,j}),D_y(u_{i,j}+\theta E_{i,j})),$$

$$A'_p = \frac{\partial}{\partial D_x(u_{i,j}+\theta E_{i,j})}\,A'(---), \quad \text{and} \quad A'_q = \frac{\partial}{\partial D_y(u_{i,j}+\theta E_{i,j})}\ A'(---).$$

By using expansions such as equation (1.6) together with the linearity of the difference approximations to the derivatives and relations of the form

$$D_xu_{i,j} = \partial u_{i,j}/\partial x + O(h^P)$$

and

$$D_x^2u_{i,j} = \partial^2u_{i,j}/\partial x^2 + O(h^P),$$

the finite difference equation for the error is written in the form

$$
\begin{aligned}
&a_{i,j}D_x^2E_{i,j} + 2b_{i,j}D_z^2E_{i,j} + c_{i,j}D_y^2E_{i,j} + d_{i,j}D_xE_{i,j} \\
(1.7)\qquad &+ e_{i,j}D_yE_{i,j} + f_{i,j}E_{i,j} = g_{i,j}
\end{aligned}
$$

where the coefficients are functions of x_i, y_j, $u_{i,j}$, and $E_{i,j}$.

By using previously designated bounds on the coefficients in equation (1.3) and on the solution of the continuous problem, equation (1.7) is shown to be uniformly elliptic. Moreover, the function g is $O(h^p)$ where p is either one or two depending on the difference approximations to the derivatives which are used.

Next, we linearize the error equation (1.7) by replacing E, where it occurs in the coefficients, by a given function w. We then consider the boundary value problem for the linearized equation (1.7) with boundary values which are identically zero. We show that this problem has a unique solution which, for sufficiently small mesh width, is majorized by the function

$$H_{i,j} = \max_R |g_{i,j}| J_{i,j} \tag{1.8}$$

where J is a nonnegative, bounded function which depends on the ellipticity constants for equation (1.7) and the size of the region R. The method used to establish the estimate (1.8) and the resulting generality of the finite difference equations to which it applies are believed to be new.

We let W_p denote the set of functions defined on the mesh points in R such that if $w \in W_p$; then

$$\max_R |w_{i,j}| \leq Yh^p$$

where $Y = \max\limits_R |g_{i,j}| J_{i,j} h^{-p}$ and p is either one or two.

The Dirichlet problem for equation (1.7) is now considered as a transformation T:

$$Tw = s. \tag{1.9}$$

The function w in equation (1.9) is the function which is used to linearize equation (1.7), and the function s is the solution of the Dirichlet problem for the linearized equation. By virtue of the bound provided by equation (1.8), the transformation T transforms a function w from the set W_p into a function s which is also in the set W_p. Since the transformation T is continuous, the Brouwer Fixed Point Theorem can be applied to show that the transformation T has a fixed point in W_p, i.e., there exists a function $w^* \in W_p$ such that

$$Tw^* = w^*;$$

consequently

$$\max_R |w^*_{i,j}| \leq Yh^p.$$

Dirichlet problems for the error in the solutions of each of the problems P_1 and P_2 are formulated by using the appropriate difference quotients in equation (1.7). The results of the analysis described above are applied to these finite difference problems to establish the existence of the error functions E and to establish error bounds in terms of the mesh width.

The organization of the remaining chapters is as follows:

Chapter II consists of a description of the continuous problem studied. Consideration of sufficient conditions to establish uniqueness for the solution of the continuous problem leads naturally to sufficient conditions for the analysis of the finite difference analogues which follow.

Chapter III is devoted to the formulation of finite difference approximations. The transformation used to eliminate mixed derivative terms and the finite difference approximations used to replace the differential operators are described.

In Chapter IV, the finite difference analogues, problems P_1 and P_2, of the continuous problem are formulated. Brief discussions are given of some of the more commonly used methods for solving the sets of nonlinear simultaneous algebraic equations which result from the formulation of the finite difference problems.

Chapter V consists of the derivation of finite difference equations for the error in the solutions of the finite difference analogues of the continuous problem.

In Chapter VI, bounds are established, using majorant functions, for the solutions of the Dirichlet problems for the linearized error equations.

In Chapter VII, the Brouwer Fixed Point Theorem is applied to the Dirichlet problems for the error equations to establish bounds for the error in the solutions of the finite difference analogues of the continuous problem.

In Chapter VIII, the existence and uniqueness of solutions of the finite difference analogues of the continuous problem is proved. Use is made of the results obtained in Chapter VI to enable a fixed-point argument to be used.

Chapter IX consists of a further analysis of the error in the solution of a finite difference analogue which utilizes finite difference operators at mesh points near the boundary which have $O(h)$ accuracy. It is shown that the bound established in previous sections for the error in the solution of this problem can be improved from $O(h)$ to $O(h^2)$.

In Chapter X, a new finite difference operator is proposed which is more convenient for use for some problems than the finite difrerence operators which were described in Chapter III. This new operator differs from previous operators only in the manner in which terms containing partial derivatives with respect to z are treated. The new operator is not necessarily of nonnegative type, and this leads to some difficulty in establishing some of its properties. It is necessary to make an assumption, which has been verified by direct calculation for a number of cases, in order to show that the error analysis presented in previous chapters applies.

In Chapter XI, a comparison is made between the results achieved in this investigation and Gerschgorin's earlier results. Also, the applications of the results of this investigation are discussed with reference to specific problems.

CHAPTER II

DIRICHLET BOUNDARY VALUE PROBLEM FOR A QUASILINEAR ELLIPTIC PARTIAL DIFFERENTIAL EQUATION

In this chapter, we describe the Dirichlet problem which we study. The smoothness requirements which are placed on the coefficients in the differential equation, on the region of the problem, and on the boundary values are stated, and the existence and uniqueness of the solution of this problem are discussed.

Let R denote a simply connected, bounded region in the plane, and let S denote the boundary of R. We assume, without loss of generality, that R lies in the strip $0 \leq x \leq X$ and that $|y| \leq Y$. The boundary S is assumed to consist of a set of points with coordinates x,y which can be regarded as functions of arc length s. The functions $x(s)$, $y(s)$ are assumed to have fourth derivatives which are Hölder continuous.[1]

Let A, B, and C represent real-valued functions with Hölder continuous partial derivatives of second order of the five variables (x,y,r,p,q); $(x,y) \in R + S$, $-\infty < r,p,q < \infty$.

We consider the following quasilinear operator

$$(2.1) \qquad Lu = A(x,y,u,\partial u/\partial x,\partial u/\partial y)\partial^2 u/\partial x^2 + 2B(---)\partial^2 u/\partial x\partial y + C(---)\partial^2 u/\partial y^2.$$

The operator L is assumed to be uniformly elliptic, i.e., there exist constants k_0, $k_1 > 0$ such that

[1] A function $g(x,y)$ is said to be Hölder continuous in a region if for any two points (x_1,y_1) and (x_2,y_2) in this region, there exist positive constants K, α such that $\alpha \leq 1$ and such that $|g(x_1,y_1)-g(x_2,y_2)| \leq K|\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}|^\alpha$.

$$k_1[\xi^2+\eta^2] \geq A(x,y,r,p,q)\,\xi^2 + 2B(---)\,\xi\eta + C(---)\eta^2 \geq k_0[\xi^2+\eta^2] \tag{2.2}$$

for all real ξ and η and for all x,y,r,p,q such that $(x,y) \in R + S$, $-\infty < r,p,q < \infty$.

The Dirichlet problem which we study is

Problem P_0: Problem P_0 consists of finding a function u which has continuous derivatives up to second order in R, is continuous in $R + S$ and satisfies in $R + S$

$$Lu = G(x,y,u,\partial u/\partial x,\partial u/\partial y), \quad (x,y) \in R \tag{2.3}$$

$$u = \phi(x,y) \quad , \quad (x,y) \in S \tag{2.4}$$

where $G(---)$ and $\phi(x,y)$ are given functions with Hölder continuous derivatives of second and fourth order respectively.

The existence of the solution of problem P_0 can be established with weaker conditions on the coefficients and the functions G and ϕ than those indicated above. We have from Bers, John, and Schechter [1964], Part II, Chapter VII, the following

THEOREM 2.1. Let equation (2.3) be uniformly elliptic and let the coefficients A, B, and C be Hölder continuous in their five variables. Let the function G be bounded by a constant K and the function ϕ have Hölder continuous first partial derivatives. Then the solution of the Dirichlet problem for equation (2.3) exists.

In order to guarantee that the solution of problem P_0 is unique, a condition is placed on the coefficients and the function G in equation (2.3) which, for some problems, involves the solution of the given problem.

This condition is stated as follows:

Let $v = v(x,y)$ be an arbitrary function defined on $R + S$ which has continuous second-order partial derivatives in R and which is equal to zero on S, and let u be the solution of problem P_0. A sufficient condition that a solution of problem P_0 is unique, is that

$$(2.5)\qquad \partial^2u/\partial x^2\int_0^1 \partial A/\partial r\, d\theta + 2\partial^2u/\partial x\partial y\int_0^1 \partial B/\partial r\, d\theta + \partial^2u/\partial y^2\int_0^1 \partial C/\partial r\, d\theta$$

$$-\int_0^1 \partial G/\partial r\, d\theta \leq 0 \quad \text{for all} \quad (x,y) \in R$$

where

$$\int_0^1 \partial A/\partial r\, d\theta = \int_0^1 \partial A(x,y,(u+\theta v),\partial/\partial x(u+\theta v),\partial/\partial y(u+\theta v))/\partial r\, d\theta$$

etc. We now prove

<u>THEOREM 2.2</u>. Let the coefficients A, B, and C and the function G satisfy condition (2.5). Then the solution of problem P_0 is unique.

<u>Proof</u>:[2] We assume that problem P_0 has two distinct solutions u_1 and u_2 and show that this assumption leads to a contradiction. Let $A(x,y,u_1,\partial u_1/\partial x,\partial u_1/\partial y)$ be denoted by A_1, $A(x,y,u_2,\partial u_2/\partial x,\partial u_2/\partial y)$ by A_2, etc. Then, we have

$$(2.6)\qquad \left.\begin{aligned} A_i\partial^2u_i/\partial x^2 + 2B_i\partial^2u_i/\partial x\partial y + C_i\partial^2u_i/\partial y^2 = G_i, &\quad (x,y) \in R \\ u_i = \emptyset, &\quad (x,y) \in S \end{aligned}\right\} \quad i = 1,2$$

[2] The proof of Theorem 2.2 is given in a somewhat abbreviated form in Bers, John, and Schechter [1964], Part II, Chapter VII.

By subtracting equation (2.6), $i = 2$, from equation (2.6), $i = 1$, and denoting $u_1 - u_2$ by v, we obtain

$$
\begin{aligned}
&A_1\partial^2 v/\partial x^2 + 2B_1\partial^2 v/\partial x\partial y + C_1\partial^2 v/\partial y^2 + [A_1-A_2]\partial^2 u_2/\partial x^2 \\
(2.7)\quad &\qquad + 2[B_1-B_2]\partial^2 u_2/\partial x\partial y + [C_1-C_2]\partial^2 u_2/\partial y^2 = [G_1-G_2].
\end{aligned}
$$

The differences $[A_1-A_2]$, etc. can be evaluated by means of the technique illustrated by equation (1.6). Thus,

$$
\begin{aligned}
A_1-A_2 &= A(x,y,u_2+v,\ \partial(u_2+v)/\partial x,\ \partial(u_2+v)/\partial y) - A(x,y,u_2,\partial u_2/\partial x,\ \partial u_2/\partial y) \\
&= \int_0^1 dA(x,y,u_2+\theta v,\ \partial(u_2+\theta v)/\partial x,\ \partial(u_2+\theta v)/\partial y)/d\theta\ d\theta \\
&= v\int_0^1 \partial A/\partial r\ d\theta + \partial v/\partial x\int_0^1 \partial A/\partial p\ d\theta + \partial v/\partial y\int_0^1 \partial A/\partial q\ d\theta
\end{aligned}
$$

where

$$\partial A/\partial r = \partial A(x,y,u_2+\theta v,\partial(u_2+\theta v)/\partial x,\ \partial(u_2+\theta v)/\partial y)/\partial r, \text{ etc.}$$

We set

$$
\begin{aligned}
D = \partial^2 u_2/\partial x^2 \int_0^1 \partial A/\partial p\ d\theta + 2\partial^2 u_2/\partial x\partial y\int_0^1 \partial B/\partial p\ d\theta + \partial^2 u_2/\partial y^2\int_0^1 \partial C/\partial p\ d\theta \\
- \int_0^1 \partial G/\partial p\ d\theta
\end{aligned}
$$

$$
\begin{aligned}
E = \partial^2 u_2/\partial x^2 \int_0^1 \partial A/\partial q\ d\theta + 2\partial^2 u_2/\partial x\partial y\int_0^1 \partial B/\partial q\ d\theta + \partial^2 u_2/\partial y^2\int_0^1 \partial C/\partial q\ d\theta \\
- \int_0^1 \partial G/\partial q\ d\theta
\end{aligned}
$$

and

$$F = \partial^2 u_2/\partial x^2 \int_0^1 \partial A/\partial r \, d\theta + 2\partial^2 u_2/\partial x\partial y \int_0^1 \partial B/\partial r \, d\theta + \partial^2 u_2/\partial y^2 \int_0^1 \partial C/\partial r \, d\theta$$

$$- \int_0^1 \partial G/\partial r \, d\theta$$

Equation (2.7) can now be written in the form

$$\bar{L}v = A_1 \, \partial^2 v/\partial x^2 + 2B_1 \, \partial^2 v/\partial x\partial y + C_1 \, \partial^2 v/\partial y^2 + D \, \partial v/\partial x + E \, \partial v/\partial y + Fv = 0$$

where the coefficients depend on x,y, and the assumed solutions u_1 and u_2. Since $v = u_1 - u_2$ is zero on the boundary S, we can formulate a boundary value problem for v as follows:

$$(2.8) \qquad \bar{L}v = 0, \quad (x,y) \in R$$

$$(2.9) \qquad v = 0, \quad (x,y) \in S.$$

We now make use of a maximum principle as given in Courant and Hilbert [1962], p. 326.

Maximum Principle: Let v satisfy equation (2.8) in R, be continuous in $R + S$, and let $F \leq 0$, then v is less than or equal to the maximum of zero and the maximum of v on S.

By condition (2.5), $F \leq 0$. Therefore, by applying the maximum principle to both the solution v of the problem given by equations (2.8) and (2.9) and to the negative of the solution of this problem, we conclude that both $v \leq 0$ and $-v \leq 0$. Thus, $v \equiv 0$, and $u_1 = u_2$.

Various subsidiary conditions which insure that condition (2.5) is satisfied are obvious from its definition. This condition is satisfied, for instance, if the coefficients A, B, and C do not depend on u, and $\partial G/\partial r$ is nonnegative. If A, B, or C does depend on u, it is necessary to verify, after a solution is obtained, that condition (2.5) is satisfied.

In order to enable the error analysis of finite difference approximations to problem P_0 which follows to be carried out, we need a slightly stronger condition than condition (2.5). We let u and v be defined as above, and require that there exist a positive number Δ such that

$$\begin{aligned}(2.10)\qquad &\partial^2u/\partial x^2 \int_0^1 \partial A/\partial r \, d\theta + 2\partial^2u/\partial x\partial y \int_0^1 \partial B/\partial r \, d\theta \\ &+ \partial^2u/\partial y^2 \int_0^1 \partial C/\partial r \, d\theta - \int_0^1 \partial G/\partial r \, d\theta \leq \Delta\nu \quad \text{for all } (x,y) \in R\end{aligned}$$

where

$$\nu = \nu(x,y) = \max \left\{ \left|\frac{\partial}{\partial r} A\right|, \left|\frac{\partial}{\partial r} B\right|, \left|\frac{\partial}{\partial r} C\right|, \left|\frac{\partial}{\partial r} G\right| \right\}.$$

The error bounds, which are derived in the sequel, for solutions of finite difference approximations to problem P_0 depend on the partial derivatives up to fourth order of the solution of the continuous problem. We therefore assume that the solution of problem P_0 possesses bounded and continuous partial derivatives up to fourth order. The boundedness and continuity of partial derivatives of solutions of elliptic partial differential equations can be established by means of the a priori estimates of Schauder[3]. Sufficient conditions to insure the existence, by means of Schauder estimates, of bounded and continuous partial derivatives up to fourth order of the solution of problem P_0 are:

(i) the operator L is uniformly elliptic,

(ii) the functions A, B, C, and G have Hölder continuous second-order partial derivatives,

[3] Schauder estimates are discussed in Bers, John, and Schechter [1964] and in Courant and Hilbert [1962].

(iii) the function ϕ has Hölder continuous partial derivatives up to fourth order, and

(iv) the boundary S of R is sufficiently smooth, i.e., S consists of a set of points with coordinates x,y which can be regarded as functions of arc length s, and the functions $x(s)$, $y(s)$ have Hölder continuous derivatives of fourth order.

CHAPTER III

FINITE DIFFERENCE OPERATORS

In this chapter, finite difference analogues of equation (2.3) are given. We first describe a finite difference analogue of equation (2.3) which is applicable at points in R which are not near the boundary S. This finite difference analogue was first presented in Bramble and Hubbard [1963] for use with linear elliptic partial differential equations. Bramble and Hubbard [1963] proves the existence of such an approximation but does not provide a method for obtaining it in practice. A practical method for obtaining it is given here.

Two methods are given for formulating finite difference approximations near the boundary.

Theoretical estimates of the error in finite difference approximations to solutions of problems involving elliptic partial differential equations are not generally obtainable unless the finite difference operators are of nonnegative type[1] and are diagonally dominant. A finite difference operator L_h, when operating on an approximate solution $U(x_i,y_j)$ of problem P_0, can be written in the following form

$$L_h U(x_i,y_j) = \sum_{(m,n)} \sigma(x_i,y_j;x_m,y_n)U(x_m,y_n)$$

where the points (x_m,y_n) comprise a given set of points in $R + S$. If

[1] Exceptions to this rule are given in Bramble and Hubbard [1962] and in Rockoff [1964].

$$\sigma(x_i,y_j;x_i,y_j) < 0, \quad (x_i,y_j) \in R,$$

$$\sigma(x_i,y_j;x_m,y_n) \geq 0, \quad (x_m,y_n) \in R + S, \quad (x_m,y_n) \neq (x_i,y_j),$$

and

$$|\sigma(x_i,y_j;x_i,y_j)| \geq \sum_{\substack{(m,n) \\ (m,n)\neq(i,j)}} \sigma(x_i,y_j;x_m,y_n),$$

then L_h is said to be of nonnegative type and to be diagonally dominant.[2]

Elliptic partial differential operators are readily approximated by finite difference operators with the above properties provided the differential operators do not contain mixed derivative terms. Finite difference approximations, other than the one described below, which are of nonnegative type and are diagonally dominant have been formulated for differential operators containing mixed derivative terms.[3]; however, these approximations require that either the magnitude of the coefficient of the mixed derivative term be severely restricted or that unequal mesh widths be used.

The method of approximating differential operators containing mixed derivative terms which is presented below is an elaboration of a method which is given in Bramble and Hubbard [1963]. This method consists of transforming the differential operator, by means of the introduction of the directional derivative, into a form which is easily approximated by a finite difference operator with the desired properties. The transformation depends only on the requirements that the operator L be uniformly elliptic and that

[2] Forsythe and Wasow [1960], p. 181.

[3] See Greenspan [1960], Greenspan and Jain [1964], McAllister [1964a], and Pucci [1958].

the coefficients be continuous functions of their arguments. The resulting finite difference operator has an $O(h^2)$ truncation error.

Let (x,y) be a point in R and let z denote a line through the point (x,y) such that the angle between the line z and the x axis is equal to τ, $0 < \tau < \pi$, $\tau \neq \pi/2$ (see Figure 3.1). Let u be any function which has continuous partial derivatives of second order. Then the second directional derivative of u with respect to z exists and is given by

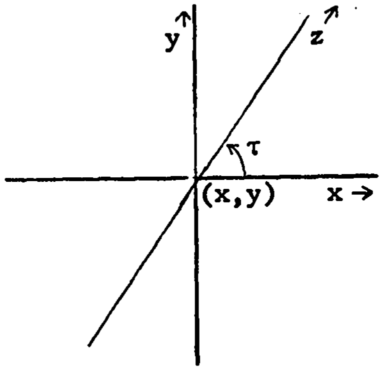


FIGURE 3.1

$$(3.1) \qquad \partial^2u/\partial z^2 = \cos^2\tau\ \partial^2u/\partial x^2 + 2\sin\tau\cos\tau\ \partial^2u/\partial x\partial y + \sin^2\tau\ \partial^2u/\partial y^2.$$

From equation (3.1), the mixed derivative term is given by

$$(3.2) \qquad 2\ \partial^2u/\partial x\partial y = (2/\sin 2\tau)\partial^2u/\partial z^2 - \cot\tau\ \partial^2u/\partial x^2 - \tan\tau\ \partial^2u/\partial y^2.$$

This expression for $\partial^2u/\partial x\partial y$ is substituted into equation (2.1) to obtain

$$(3.3) \qquad Lu = A'\ \partial^2u/\partial x^2 + 2B'\ \partial^2u/\partial z^2 + C'\ \partial^2u/\partial y^2$$

where

$$(3.4) \qquad \begin{aligned} A' &= A(x,y,u,\partial u/\partial x,\partial u/\partial y) - B(---)\cot\tau \\ B' &= B(---)/\sin 2\tau \\ C' &= C(---) - B(---)\tan\tau. \end{aligned}$$

The principal result of Bramble and Hubbard [1963] relating to the above procedure is summarized by the following theorem.

THEOREM 3.1. Let the coefficients in equation (2.1) be continuous functions of the indicated variables, and assume that condition (2.2) is satisfied. Let $\tan\tau = \gamma = \gamma(x,y)$. Then there exist constants k_0' and η, $0 < k_0'$, $1 \leq \eta < \infty$, such that $\gamma(x,y)$ can be specified at each point in R and

$$(3.5)\qquad \begin{aligned} k_0' &\leq A',\ C' \\ 0 &\leq B' \\ \gamma &= \pm\,\alpha/\beta \end{aligned}$$

where α and β are relatively prime integers and

$$1 \leq \alpha,\ \beta \leq \eta.$$

A proof of this theorem is sketched in Bramble and Hubbard [1963]. A complete proof is given in the Appendix of this paper.

The angle τ is specified at each point in R such that conditions (3.5) are satisfied. A method for doing this in practical applications is described later in this chapter.

The set of mesh points, at which numerical approximations are calculated, are the intersections of two families of straight lines called mesh lines. These two families of mesh lines are given by $x_i = ih$, $i = 0, 1, 2, \ldots, I$ and $y_j = jh$, $j = 0, \pm 1, \pm 2, \ldots, \pm J$ where I and J are positive integers such that $(I-1)h \leq X \leq Ih$ and $Jh \geq Y$.

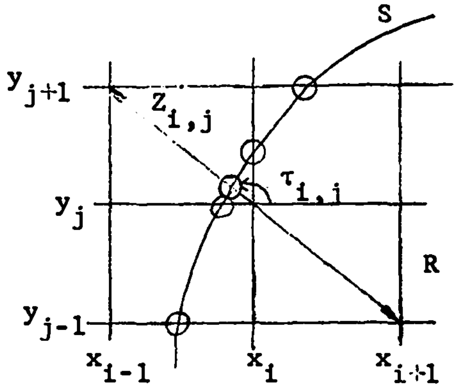
With each mesh point $(x_i,y_j) \in R$, there is associated a pair of points, either $(x_i+\beta h,\ y_j+\alpha h)$ and $(x_i-\beta h,\ y_j-\alpha h)$ or $(x_i-\beta h,\ y_j+\alpha h)$ and $(x_i+\beta h,\ y_j-\alpha h)$. Since α and β are relatively prime integers, these points will be mesh points (though not necessarily mesh points in R). These mesh points are called the diagonal neighbors of the mesh point (x_i,y_j). The distance between the mesh point (x_i,y_j) and either of its diagonal neighbors is given by

$$k_{i,j} = h(\alpha^2 + \beta^2)^{\frac{1}{2}}. \tag{3.6}$$

The four mesh points (x_i+h,y_j), (x_i-h,y_j), (x_i,y_j+h), and (x_i,y_j-h) are called the rectangular neighbors of the mesh point (x_i,y_j). The diagonal neighbors plus the rectangular neighbors are called the neighborhood $N(x_i,y_j)$ of the point (x_i,y_j).

A mesh point $(x_i,y_j) \in R$ is called a regular mesh point if each of the mesh points in $N(x_i,y_j)$ is in R. All mesh points in R that are not regular mesh points are called irregular. The disjoint sets of regular and irregular mesh points in R are denoted by R_h and R_b respectively.

For each mesh point $(x_i,y_j) \in R$, let the portion of the line z which connects the point (x_i,y_j) with its diagonal neighbors be denoted by $z_{i,j}$. The points on the boundary S which are at the intersections of the lines $z_{i,j}$ and the mesh lines x_i and y_j are called boundary mesh points (see Figure 3.2). The set of boundary mesh points is denoted by R_S. We assume that the mesh width h is sufficiently small that, for each mesh point $(x_i,y_j) \in R$, at least one of the mesh points on each line x_i, y_j, and $z_{i,j}$ which are in $N(x_i,y_j)$ is also in $R + S$.



O - Boundary mesh points

FIGURE 3.2

The finite difference approximation to the solution of problem P_0 is defined in R only at the mesh points (x_i,y_j) and is denoted by $U(x_i,y_j) = U_{i,j}$.

At regular mesh points, we use the usual central difference quotients:

$$\begin{aligned}\Delta_x U_{i,j} &= (U_{i+1,j} - U_{i,j})/h \\ \nabla_x U_{i,j} &= (U_{i,j} - U_{i-1,j})/h \\ (U_{i,j})_x &= (\Delta + \nabla)_x U_{i,j} \\ &= (U_{i+1,j} - U_{i-1,j})/2h\end{aligned} \tag{3.7}$$

and

$$\begin{aligned}(U_{i,j})_{xx} &= (\Delta\nabla)_x U_{i,j} \\ &= (U_{i+1,j} - 2U_{i,j} + U_{i-1,j})/h^2\end{aligned} \tag{3.8}$$

Similarly,

$$\begin{aligned}(U_{i,j})_y &= (\Delta + \nabla)_y U_{i,j} \\ &= (U_{i,j+1} - U_{i,j-1})/2h,\end{aligned} \tag{3.9}$$

$$\begin{aligned}(U_{i,j})_{yy} &= (\Delta\nabla)_y U_{i,j} \\ &= (U_{i,j+1} - 2U_{i,j} + U_{i,j-1})/h^2\end{aligned} \tag{3.10}$$

and

$$\begin{aligned}(U_{i,j})_{zz} &= (\Delta\nabla)_z U_{i,j} \\ &= (U_{i\pm\beta,j+\alpha} - 2U_{i,j} + U_{i\mp\beta,j-\alpha})/k^2.\end{aligned} \tag{3.11}$$

The finite difference operator L_h is defined at regular mesh points by the following finite difference equation.

$$L_h U_{i,j} = A'(x_i,y_j,U_{i,j},(\Delta+\nabla)_x U_{i,j}, (\Delta+\nabla)_y U_{i,j})(\Delta\nabla)_x U_{i,j} \quad (3.12)$$
$$+ 2B'(---)(\Delta\nabla)_z U_{i,j} + C'(---)(\Delta\nabla)_y U_{i,j}$$

The differences between the approximating difference quotients defined above and the corresponding exact derivatives can be estimated by means of Taylor's Theorem with remainder. We have

$$(3.13) \qquad (\partial u/\partial x)_{i,j} - (\Delta+\nabla)_x u_{i,j} = h^2\left[(\partial^3 u/\partial x^3)_{i\pm\theta,j}\right]/6, \quad 0 \leq \theta \leq 1,$$

and

$$(3.14) \qquad (\partial^2 u/\partial x^2)_{i,j} - (\Delta\nabla)_x u_{i,j} = h^2\left[(\partial^4 u/\partial x^4)_{i\pm\phi,j}\right]/12, \quad 0 \leq \phi \leq 1.$$

Similar relationships hold between the derivatives and difference quotients with respect to y and z.

Two alternate finite difference operators are defined at irregular mesh points. They are denoted by L_{b1} and L_{b2}.

The finite difference operator L_{b1} is an adaptation of a linear interpolation scheme originally given in Collatz [1933]. Consider the configuration of mesh points given in Figure 3.3a where τ is assumed to have been determined as indicated. The operator L_{b1} is defined at the point (x_i,y_j) by

$$(3.15) \qquad L_{b1}U_{i,j} = [\lambda/(\lambda+1)]U_{i+1,j} + [1/(\lambda+1)]U_{p,q} - U_{i,j}$$

where λh is the Euclidean distance between the point (x_i,y_j) and a boundary mesh point (x_p,y_q).

For the configuration given in Figure 3.3b, $\tau = 3\pi/4$, and a diagonal neighbor of the point (x_i,y_j) is not in $R + S$. For this configuration,

$$(3.16) \qquad L_{b1}U_{i,j} = [\lambda/(\lambda+1)]U_{i+1,j-1} + [1/(\lambda+1)]U_{p,q} - U_{i,j}.$$

The general case for which a diagonal neighbor is not in $R + S$ is illustrated by Figure 3.4. For this configuration, $\gamma_{i,j} = \alpha/\beta$, and the diagonal neighbor $(x_{i-\beta},y_{i-\alpha}) \notin R + S$; therefore, (x_i,y_j) is an irregular mesh point. Then there exists a point $(x_i - \lambda k \cos \tau,\ y_j - \lambda k \sin \tau) = (x_{i-\lambda\beta},\ y_{i-\lambda\alpha}) \in R_S$. In this case,

$$(3.17) \qquad L_{b1}U_{i,j} = [\lambda/(\lambda+1)]U_{i+\alpha,j+\beta} + [1/(\lambda+1)]U_{i-\lambda B,j-\lambda\alpha} - U_{i,j}.$$

A generalization of either (3.15), (3.16), or (3.17) is applicable to any mesh point in R_b. In case more than one mesh point in $N(x_i,y_j)$ is not in $R + S$, there is a choice regarding the precise definition of L_{b1}

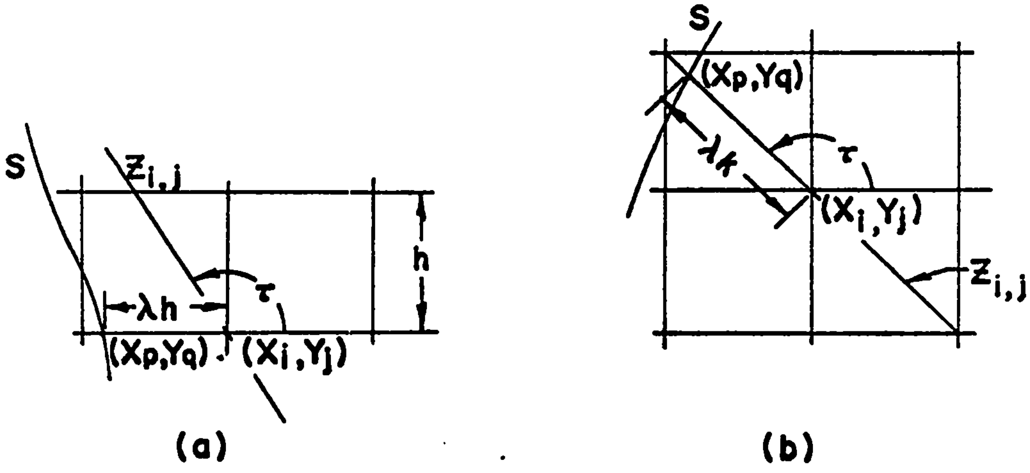


Figure 3.3

at (x_i, y_j). Insofar as the considerations here are concerned, the choice is arbitrary.

The operator L_{b2} utilizes formal approximations to the partial derivatives. Consider an irregular mesh point (x_i, y_j) and assume that the point $(x_{i-1}, y_j) \notin R$, i.e., the configuration given in Figure 3.3a. Then the term $\partial u/\partial x$ in equation (3.3) is approximated at the point (x_i, y_j) by

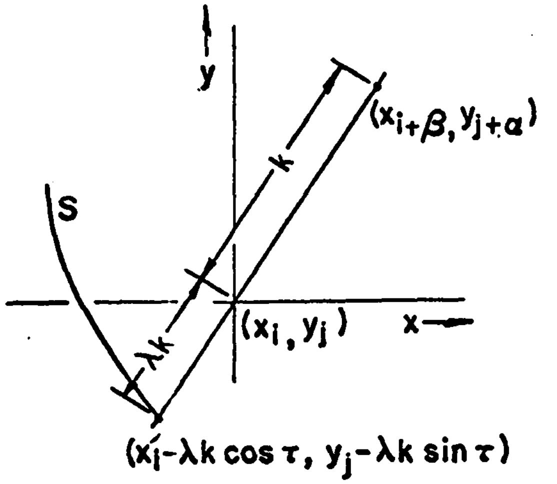

y
S
k
λk
$(x_{i+\beta}, y_{j+\alpha})$
(x_i, y_j)
x
$(x_i - \lambda k \cos\tau, y_j - \lambda k \sin\tau)$

Figure 3.4

$$(3.18) \qquad (\partial u/\partial x)_{i,j} \simeq (U_{i,j})_x = [1/(\lambda+1)h][U_{i+1,j} - U_{i-\lambda,j}]$$

and the term $\partial^2 u/\partial x^2$ by

$$(3.19) \qquad (\partial^2 u/\partial x^2)_{i,j} \simeq (U_{i,j})_{xx}$$

$$= (2/h)[(1/(\lambda+1))U_{i+1,j} - (1/\lambda)U_{i,j} + (1/\lambda(\lambda+1))U_{i-\lambda,j}].$$

Similar expressions are used when one or more of the mesh points (x_{i+1}, y_j), (x_i, y_{j+1}), (x_i, y_{j-1}) do not belong to $R + S$.

Next, assume that (x_i, y_j) is the irregular mesh point given in Figure 3.3b. In this case, the term $\partial^2u/\partial z^2$ is approximated by

$$(3.20)\qquad (\partial^2u/\partial z^2)_{i,j} \simeq (U_{i,j})_{zz}$$

$$= (2/h)[(1/(\lambda+1))U_{i+1,j-1} - (1/\lambda)U_{i,j}$$

$$+ (1/\lambda(\lambda+1))U_{i-\lambda,j+\lambda}],$$

and for the configuration given in Figure 3.4, $\partial^2u/\partial z^2$ is approximated by

$$(3.21)\qquad (\partial^2u/\partial z^2)_{i,j} \simeq (U_{i,j})_{zz}$$

$$= (2/k^2)[(1/(\lambda+1))U_{i+\beta,j+\alpha} - (1/\lambda)U_{i,j}$$

$$+ (1/\lambda(\lambda+1))U_{i-\lambda\beta,j-\lambda\alpha}].$$

Approximations such as those given by equations (3.18)-(3.21) are used to replace the partial derivatives in L to form the operator L_{b2}.

The differences between the finite difference quotients defined above and the corresponding exact derivatives are given below.

The for operator L_{b1} and the mesh point configurations given in Figure 3.3a,b, we have respectively

$$(3.22)\qquad \begin{aligned} &u(x_i,y_j) - [\lambda/(\lambda+1)]u(x_{i+1},y_j) - [1/(\lambda+1)]u(x_p,y_q) \\ &= -[\lambda h^2/2(\lambda+1)][\lambda(\partial^2u/\partial x^2)_{i-\theta,j} + (\partial^2u/\partial x^2)_{i+\phi,j}],\quad 0 \le \theta,\ \phi \le 1, \end{aligned}$$

and

$$u(x_i,y_j) - [\lambda/(\lambda+1)]u(x_{i+1},y_{j-1}) - [1/(\lambda+1)]u(x_p,y_q)$$

(3.23)

$$= -[\lambda h^2/(\lambda+1)][\lambda(\partial^2u/\partial z^2)_{i-\theta,j+\phi} + (\partial^2u/\partial z^2)_{i+\theta,j-\phi}],$$

$$0 \leq \phi,\ \phi \leq \sqrt{2}$$

For the configuration given in Figure 3.4, we have

$$u(x_i,y_j) - [\lambda/(\lambda+1)]u_{i+\alpha,j+\beta} - [1/(\lambda+1)]u_{i-\lambda\beta,j-\lambda\alpha}$$

(3.24)

$$= -[\lambda k^2/2(1+\lambda)][\lambda(\partial^2u/\partial z^2)_{i-\phi,j-\phi} + (\partial^2u/\partial z^2)_{i+\theta,j+\phi}],$$

$$0 \leq \theta \leq \beta,\ 0 \leq \phi \leq \alpha.$$

For the operator L_{b2} and the mesh point configuration given in Figure 3.3a,

$$(\partial u/\partial x)_{i,j} - [1/h(\lambda+1)][u(x_{i+1},y_j) - u(x_{i-\lambda},y_j)]$$

(3.25)

$$= [h/2(\lambda+1)][\lambda^2(\partial^2u/\partial x^2)_{i-\phi,j} - (\partial^2u/\partial x^2)_{i+\phi,j}]$$

and

$$(\partial^2u/\partial x^2)_{i,j} - 2\Big\{[1/(\lambda+1)]u(x_{i+1},y_j) + [1/\lambda(\lambda+1)]u(x_{i-\lambda},y_j)$$

(3.26)

$$- [1/\lambda]u(x_i,y_j)]\Big\}/h^2 = -[h/3(\lambda+1)][\partial^3u/\partial x^3)_{i+\phi,j} - \lambda^2(\partial^3u/\partial x^3)_{i-\theta,j}],$$

$$0 \leq \phi,\ \theta \leq 1.$$

For the mesh point configuration given in Figure 3.3b,

$$(\partial^2 u/\partial z^2)_{i,j} - [1/(\lambda+1)]u(x_{i+1},y_{j-1}) - [1/\lambda(\lambda+1)]u(x_p,y_q)$$

$$(3.27) \qquad + [1/\lambda]u(x_i,y_j)]/h^2 = -[\sqrt{2}\, h/3(\lambda+1)][(\partial^3 u/\partial z)_{i+\phi,j-\theta}$$

$$- \lambda^2(\partial^3 u/\partial z^2)_{i-\phi,j+\theta}], \quad 0 \le \phi, \theta \le \sqrt{2} ,$$

and a similar expression holds for the difference between $\partial^2 u/\partial z^2$ and the difference quotient given by equation (3.21).

We now describe a procedure by which the angle τ can be chosen at each mesh point in R. We assume that an iterative method is used to solve the finite difference analogue of problem P_0 and that the finite difference approximation to equation (2.3) is linearized in some manner so that the coefficients can be evaluated at each mesh point prior to each successive iteration.

We first require that the value of the angle τ corresponding to a mesh point (x_i,y_j) be chosen such that $\tan \tau$ has the same sign as the coefficient $B_{i,j}$. This insures that $B'_{i,j}$ is nonnegative. If

$$(3.28) \qquad \begin{aligned} &A_{i,j} - |B_{i,j}| > 0, \text{ and} \\ &C_{i,j} - |B_{i,j}| > 0, \end{aligned}$$

τ is chosen to be either $\pi/4$ or $3\pi/4$ depending on whether $B_{i,j}$ is positive or negative respectively.

If condition (3.28) is not satisfied, we resort to the following procedure. We know from Theorem 3.1 that a value of $\tau = \tau(x,y)$ exists such that condition (3.5) is satisfied at each mesh point. From condition (3.5), we have for such a value of τ

$$A'_{i,j} = A_{i,j} - B_{i,j} \cot \tau > 0$$

and

$$C'_{i,j} = C_{i,j} - B_{i,j} \tan \tau > 0$$

or

(3.29) $$B_{i,j}/A_{i,j} < \tan \tau < C_{i,j}/B_{i,j} \quad \text{if} \quad B_{i,j} > 0$$

and

(3.30) $$C_{i,j}/B_{i,j} < \tan \tau < B_{i,j}/A_{i,j} \quad \text{if} \quad B_{i,j} < 0.$$

Values of $A'_{i,j}$ and $C'_{i,j}$ are indicated schematically in Figure 3.5 as functions of $\gamma = \tan \tau$ for the case $B_{i,j} > 0$. For the case $B_{i,j} < 0$, the curves in Figure 3.5 are reflected about $\gamma = 0$. From condition (3.5), we know that the curves in Figure 3.5 intersect at a point such that $A'_{i,j}$, $C'_{i,j} \geq k'_0 > 0$. Therefore, if condition (3.28) is not satisfied, $\gamma_{i,j}$ can be chosen from

$$A_{i,j} - B_{i,j}/\gamma_{i,j} = C_{i,j} - B_{i,j}\,\gamma_{i,j}$$

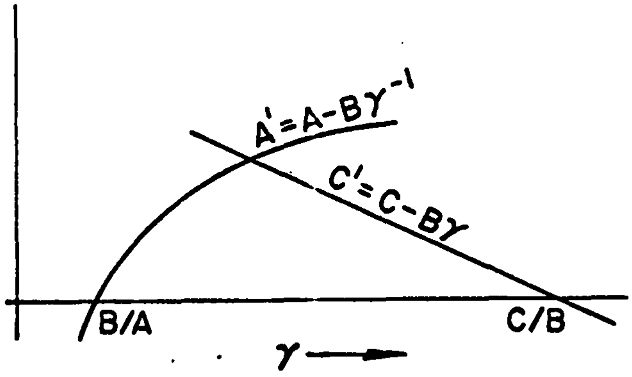


Figure 3.5

or

$$(3.31) \qquad \gamma_{i,j} = (1/2B_{i,j})[C_{i,j} - A_{i,j} + (C^2_{i,j} - 2C_{i,j}A_{i,j} + A^2_{i,j} + 4B^2_{i,j})^{1/2}].$$

According to Theorem 3.1, $\gamma_{i,j}$ can be chosen as the ratio of relatively prime integers α and β. The procedure outlined above will not, in general, result in such a choice. However, the value of $\gamma_{i,j}$ obtained from equation (3.31) can be approximated as closely as desired by a ratio of relatively prime integers.

We now show that the coefficients $A'_{i,j}$, $C'_{i,j}$, and $B'_{i,j}$ in the transformed operator (3.3) are bounded. From condition (2.2), the coefficients $A_{i,j}$, $C_{i,j}$, and $|B_{i,j}|$ are bounded by a constant k_1. Since $B_{i,j}\gamma_{i,j}$ is nonnegative, $A'_{i,j}$ and $C'_{i,j}$ are bounded by the same constant k_1. If condition (3.28) is satisfied, $B_{i,j}$ is also bounded by k_1. Suppose condition (3.28) is not satisfied; then two cases, corresponding to $|\gamma_{i,j}| < 1$ and $|\gamma_{i,j}| > 1$ respectively, must be considered. Assume $|\gamma_{i,j}| > 1$, then

$$\begin{aligned} B'_{i,j} &= B_{i,j}/\sin 2\tau \\ &= B_{i,j}/2 \sin \tau \cos \tau \\ &= B_{i,j}(1+\delta^2)/2\delta \end{aligned}$$

where δh is the distance between the mesh line $x = x_i$ and the intersection between the mesh line $y = y_{i+1}$ and the line $z_{i,j}$ (see Figure 3.6). Note that $0 < \delta < 1$. From conditions (3.29) and (3.30),

$$|\gamma_{i,j}| < C_{i,j}/|B_{i,j}|$$

or

$$1/\delta < C_{i,j}/|B_{i,j}|.$$

Therefore,

$$B'_{i,j} < |B_{i,j}|(1+\delta^2)C_{i,j}/2\ |B_{i,j}|$$

$$< (1+\delta^2)C_{i,j}/2.$$

Thus,

$$B'_{i,j} < k_1\ . \tag{3.32}$$

If $|\lambda| < 1$, a similar analysis gives the same result, i.e., that condition (3.32) is satisfied.

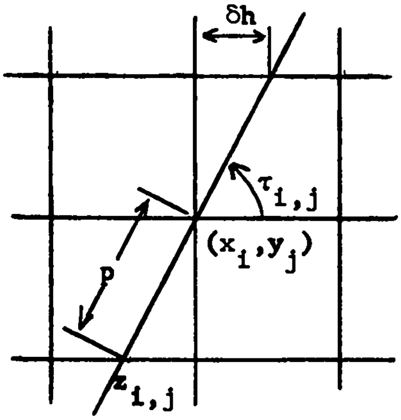


FIGURE 3.6

CHAPTER IV

THE FINITE DIFFERENCE PROBLEMS

In order to obtain an approximate solution $U_{i,j}$ of problem P_0, the function $\gamma = \gamma_{i,j}$ is first evaluated at each mesh point in $R_h + R_b$. Next, the operator L is transformed into the form given by equation (3.3). The transformed operator is then replaced by the finite difference operator L_h at each mesh point in R_h. At the mesh points in R_b, the approximate solution $U_{i,j}$ is required to satisfy one of the following equations:

$$L_{b1}U_{i,j} = 0,$$

or

$$L_{b2}U_{i,j} = G(x_i,y_j,U_{i,j},D_xU_{i,j},D_yU_{i,j})$$

where D_x and D_y denote applicable finite difference approximations to the partial derivatives with respect to x and y respectively. The value of $U(x,y)$ at each point in R_S is taken to be equal to $\emptyset(x,y)$.

We consider the following distinct discrete problems:

<u>Problem P_1</u>: Problem P_1 consists of finding a function $U_{i,j}$ which satisfies

$$(4.1) \quad L_hU_{i,j} = G(x_i,y_j,U_{i,j}, (\Delta+\nabla)_xU_{i,j},(\Delta+\nabla)_yU_{i,j}), \quad (x_i,y_j) \in R_h$$

$$(4.2) \quad L_{b1}U_{i,j} = 0 \quad , \ (x_i,y_j) \in R_b$$

$$(4.3) \quad U(x,y) = \emptyset(x,y) \quad , \ (x,y) \in R_S.$$

<u>Problem P_2</u>: For problem P_2, we require

$$(4.4)\qquad L_h U_{i,j} = G(x_i,y_j,U_{i,j},(\Delta+\nabla)_x U_{i,j},(\Delta+\nabla)_y U_{i,j}), \quad (x_i,y_j) \in R_h$$

$$(4.5)\qquad L_{b2} U_{i,j} = G(x_i,y_j,U_{i,j},D_x U_{i,j},D_y U_{i,j}) \quad , \quad (x_i,y_j) \in R_b$$

$$(4.6)\qquad U(x,y) = \phi(x,y) \quad , \quad (x,y) \in R_S.$$

Because the systems of equations comprising problems P_1 and P_2 are nonlinear, some iterative procedure is usually required to solve them. It is not our purpose to discuss such procedures in detail in this paper. We merely note some of the types of iterative methods which are used.

Usually, a method for solving a system of nonlinear algebraic equations involves a linearization of the system of equations in such a way that successive solutions of the linearized system converges to the solution of the nonlinear system. Frequently the form of the nonlinearity can be exploited to this end in a simple way for a particular problem. An example of the use of such a procedure for a continuous problem is given in Ablow and Perry [1959] where it is shown that the problem given by

$$(4.7)\qquad \Delta u = bu^2 \quad , \quad (x,y) \in R,$$

$$(4.8)\qquad u = \phi \quad , \quad (x,y) \in S,$$

where b is a nonnegative constant and ϕ is a given nonnegative function can be solved by forming successive iterants according to

$$\Delta u^{(n+1)} = bu^{(n+1)}u^{(n)} \quad , \quad (x,y) \in R$$

$$u^{(n+1)} = \phi \quad , \quad (x,y) \in S.$$

Here, $u^{(n)}$ and $u^{(n+1)}$ denote the n^{th} and $n+1^{st}$ iterants respectively. A discretized version of this problem and of the iteration scheme is presented in McAllister [1964c].

There are also several methods for solving general systems of nonlinear algebraic equations. One of these, the so-called "natural" method consists of requiring the current iterant to be the solution of the system of linear equations obtained by evaluating the coefficients which depend on U and other terms which contribute nonlinearities at the previous iterant (see, for instance, Young and Wheeler [1964]).

Another method which can be used to solve systems of nonlinear algebraic equations consists of a generalization of Newton's method. If we write the system of equations in vector form as

$$\vec{F}(\vec{U}) = 0 \tag{4.9}$$

and denote by $\vec{A}(\vec{U}) = (a_{i,j})$ the matrix with elements

$$a_{i,j} = \partial F_i(\vec{U})/\partial U_j$$

then successive iterants for Newton's method are obtained from

$$\vec{U}^{(n+1)} = \vec{U}^{(n)} - \left[\vec{A}(\vec{U}^{(n)})\right]^{-1}\vec{F}(\vec{U}^{(n)}).$$

The question of finding sufficient conditions for the convergence of the above procedure was settled in Kantorovich [1948]. The result of Kantorovich can be stated as follows:

Let J, B, C, and D be constants where

$$J = BCD,$$

and assume the following conditions are satisfied:

(i) for $\vec{U} = \vec{U}^{(0)}$, the matrix $\vec{A}(\vec{U}^{(0)})$ has an inverse and an estimate for its norm[1] is known

$$\|[\vec{A}(\vec{U}^{(0)})]^{-1}\| \leq B,$$

(ii) the vector $\vec{U}^{(0)}$ is a sufficiently close approximation to the solution of (4.9) that

$$\|[\vec{A}(\vec{U}^{(0)})]^{-1} \vec{F}(\vec{U}^{(0)})\| < C,$$

(iii) in the region defined by inequality (4.10) below, the components of the vector $\vec{F}(\vec{U})$ are twice continuously differentiable with respect to the components of $\vec{U}$ and satisfy

$$\sum_{j,k=1}^{N} |\partial^2 F_i/\partial U_j \partial U_k| \leq D, \; i = 1, 2, \ldots, N,$$

and

(iv) the constant J satisfies the inequality

$$J < 1/2 \; .$$

Then the system of equations (4.9) has a solution $\vec{U}^*$ which is located in the sphere

$$\|\vec{U} - \vec{U}^{(0)}\| \leq \left[[1-(1-2J)^{1/2}]/J\right]C.$$

Kantorovich also shows the convergence of the sequence $\vec{U}^{(n)}$ to be almost quadratic; for large n,

$$\left[\|\vec{U}^* - \vec{U}^{(n+1)}\| \Big/ \|\vec{U}^* - \vec{U}^{(n)}\|^p\right] < \infty$$

for any nonnegative p less than 2.

[1] The matrix norm used by Kantorovich is $\|\vec{V}\| = \max_{1\leq i\leq N} \sum_{j=1}^{N} |v_{i,j}|$ where $\vec{V} = (v_{i,j})$ is an $N \times N$ matrix.

CHAPTER V

THE ERROR EQUATIONS

The difference between the solution u of Problem P_0 and the solution U of a finite difference analogue of problem P_0 is defined as the error E, i.e.,

$$U_{i,j} = u_{i,j} + E_{i,j}. \tag{5.1}$$

By replacing $U_{i,j}$ by its equivalent $E_{i,j} + u_{i,j}$ in the finite difference equations comprising problems P_1 and P_2, finite difference equations are obtained for the error.

In order to simplify the notation, we define the following abbreviations for $n \leq 4$:

$\bar{M}_n > (M_n)_{i,j} = n^{th}$ partial derivative of u with respect to x evaluated at a point $(x_{i\pm\theta}, y_j)$, $0 \leq \theta \leq 1$,

$\bar{N}_n > (N_n)_{i,j} = n^{th}$ partial derivative of u with respect to y evaluated at a point $(x_i, y_{j\pm\theta})$, $0 \leq \theta \leq 1$, and

$\bar{Q}_n > (Q_n)_{i,j} = n^{th}$ partial derivative of u with respect to z evaluated at a point $(x_{i\pm\omega}, y_{j\pm\epsilon})$, $0 \leq \omega \leq \beta$, $0 \leq \epsilon \leq \alpha$, where $\gamma_{i,j} = \alpha.\beta$.

We consider first the finite difference equation (4.1) at mesh points in R_h

$$A'(x_i,y_j,(u_{i,j}+E_{i,j}),(\Delta+\nabla)_x(u_{i,j}+E_{i,j}),(\Delta+\nabla)_y(u_{i,j}+E_{i,j}))[(\Delta\nabla)_x(u_{i,j}+E_{i,j})] \tag{5.2}$$

$$+\ B'(---)[(\Delta\nabla)_z(u_{i,j}+E_{i,j})] + C'(---)[(\Delta\nabla)_y(u_{i,j}+E_{i,j})] = G(---)$$

The functions A', B', C', and G are expanded by using the definition of a definite integral. This is illustrated below for the function A':

$$A'(x_i,y_j,(u_{i,j}+E_{i,j}),\ (\Delta+\nabla)_x(u_{i,j}+E_{i,j}),\ (\Delta+\nabla)_y(u_{i,j}+E_{i,j}))$$

$$= A'(x_i,y_j,u_{i,j},(\Delta+\nabla)_x u_{i,j},(\Delta+\nabla)_y u_{i,j}) + E_{i,j}\left[\int_0^1 \bar{A}_r d\theta\right]_{i,j} \tag{5.3}$$

$$+\ (\Delta+\nabla)_x E_{i,j}\left[\int_0^1 \bar{A}_p d\theta\right]_{i,j} + (\Delta+\nabla)_y E_{i,j}\left[\int_0^1 \bar{A}_q d\theta\right]_{i,j}$$

where, with $\partial/\partial r$ denoting differentiation with respect to the third argument of $\bar{A}$,

$$\left[\int_0^1 \bar{A}_r d\theta\right]_{i,j} = \int_0^1 \partial A'(x_i,y_j,(u_{i,j}+\theta E_{i,j}),\ (\Delta+\nabla)_x(u_{i,j}+\theta E_{i,j}),$$

$$(\Delta+\nabla)_y(u_{i,j}+\theta E_{i,j}))/\partial r\ d\theta,$$

and similar definitions apply to the terms

$$\left[\int_0^1 \bar{A}_p d\theta\right]_{i,j} \text{ and } \left[\int_0^1 \bar{A}_q d\theta\right].$$

By making use of relationships such as equation (3.13), we have the following additional expansion.

$$A'(x_i,y_j,u_{i,j},(\Delta+\nabla)_x u_{i,j},(\Delta+\nabla)_y u_{i,j})$$

$$(5.4) \qquad = A'(x_i,y_j,u_{i,j},(\partial u/\partial x)_{i,j} + h^2(M_3)_{i,j}/6,\ (\partial u/\partial y)_{i,j} + h^2(N_3)_{i,j}/6)$$

$$= A'_{i,j} + h^2(M_3)_{i,j}\left[\int_0^1 \hat{A}_p d\theta\right]_{i,j}/6 + h^2(N_3)_{i,j}\left[\int_0^1 \hat{A}_q d\theta\right]_{i,j}/6$$

where

$$A'_{i,j} = A'(x_i,y_j,u_{i,j},(\partial u/\partial x)_{i,j},(\partial u/\partial y)_{i,j}),$$

$$\left[\int_0^1 \hat{A}_p d\theta\right]_{i,j} = \int_0^1 \partial A'(x_i,y_j,u_{i,j},(\partial u/\partial x)_{i,j} + \theta h^2(M_3)_{i,j}/6,$$

$$(\partial u/\partial y)_{i,j} + \theta h^2(N_3)_{i,j}/6)/\partial p\ d\theta,$$

and

$$\left[\int_0^1 \hat{A}_q d\theta\right]_{i,j}$$

is defined in a similar manner.

We use expansions such as (5.3) and (5.4) and the fact that the difference operator is linear with respect to second-order differences to write equation (5.2) in the following form:

$$a_{i,j}(\Delta\nabla)_x E_{i,j} + 2b_{i,j}(\Delta\nabla)_z E_{i,j} + c_{i,j}(\Delta\nabla)_y E_{i,j}$$

$$(5.5) \qquad + d_{i,j}(\Delta+\nabla)_x E_{i,j} + e_{i,j}(\Delta+\nabla)_y E_{i,j} + f_{i,j}E_{i,j} = g_{i,j}$$

$$- A'_{i,j}(\partial^2u/\partial x^2)_{i,j} - 2B'_{i,j}(\partial^2u/\partial z^2)_{i,j} - C'_{i,j}(\partial^2u/\partial y^2)_{i,j} + G_{i,j}$$

where

$$a_{i,j} = A'(x_i,y_j,(u_{i,j}+E_{i,j}),(\Delta+\nabla)_x(u_{i,j}+E_{i,j}),(\Delta+\nabla)_y(u_{i,j}+E_{i,j})),$$

$$b_{i,j} = B'(---),\quad c_{i,j} = C'(---),$$

$$d_{i,j} = \left[\int_0^1 \bar{A}_p d\theta\right]_{i,j}\left[(\Delta\nabla)_x u_{i,j}\right] + 2\left[\int_0^1 \bar{B}_p d\theta\right]_{i,j}\left[(\Delta\nabla)_z u_{i,j}\right]$$

$$+ \left[\int_0^1 \bar{C}_p d\theta\right]_{i,j}\left[(\Delta\nabla)_y u_{i,j}\right] - \left[\int_0^1 \bar{G}_p d\theta\right]_{i,j},$$

$$e_{i,j} = \left[\int_0^1 \bar{A}_q d\theta\right]_{i,j}\left[(\Delta\nabla)_x u_{i,j}\right] + 2\left[\int_0^1 \bar{B}_q d\theta\right]_{i,j}\left[(\Delta\nabla)_z u_{i,j}\right]$$

$$+ \left[\int_0^1 \bar{C}_q d\theta\right]_{i,j}\left[(\Delta\nabla)_y u_{i,j}\right] - \left[\int_0^1 \bar{G}_q d\theta\right]_{i,j},$$

$$f_{i,j} = \left[\int_0^1 \bar{A}_r d\theta\right]_{i,j}\left[(\Delta\nabla)_x u_{i,j}\right] + 2\left[\int_0^1 \bar{B}_r d\theta\right]_{i,j}\left[(\Delta\nabla)_z u_{i,j}\right]$$

$$+ \left[\int_0^1 \bar{C}_r d\theta\right]_{i,j}\left[(\Delta\nabla)_y u_{i,j}\right] - \left[\int_0^1 \bar{G}_r d\theta\right]_{i,j}$$

and

$$g_{i,j} = -h^2 \left\{(M_4)_{i,j}A'_{i,j}/12 + \left[(\Delta\nabla)_x u_{i,j}\right]\left[(M_3)_{i,j}\int_0^1 \hat{A}_p d\theta\right.\right.$$

$$\left. + (N_3)_{i,j} \int_0^1 \hat{A}_q d\theta\right]_{i,j}/6 + (Q_4)_{i,j}B'_{i,j}/6$$

$$+ \left[(\Delta\nabla)_z u_{i,j}\right]\left[(M_3)_{i,j}\int_0^1 \hat{B}_p d\theta + (N_3)_{i,j}\int_0^1 \hat{B}_q d\theta\right]_{i,j}/3$$

$$+ (N_4)_{i,j}C'_{i,j}/12 + \left[(\Delta\nabla)_y u_{i,j}\right]\left[(M_3)_{i,j}\int_0^1 \hat{C}_p d\theta\right.$$

$$\left.\left. + (N_3)_{i,j}\int_0^1 C_q d\theta\right]_{i,j}/6 - \left[(M_3)_{i,j}\int_0^1 \hat{G}_p d\theta + (N_3)_{i,j}\int_0^1 \hat{G}_q d\theta\right]_{i,j}/6\right\}$$

The last four terms in equation (5.5) sum to zero because of equations (2.3) and (3.3). The finite difference equation for the error in the solutions of problems P_1 and P_2 at mesh points in R_h is then

$$\begin{aligned} (5.6) \qquad & a_{i,j}(\Delta\nabla)_x E_{i,j} + 2b_{i,j}(\Delta\nabla)_z E_{i,j} + c_{i,j}(\Delta\nabla)_y E_{i,j} \\ & + d_{i,j}(\Delta+\nabla)_x E_{i,j} + e_{i,j}(\Delta+\nabla)_y E_{i,j} + f_{i,j} E_{i,j} = g_{i,j} \end{aligned}$$

which we write also as

$$(5.7) \qquad \tilde{L}_h E_{i,j} = g_{i,j}.$$

Since the coefficients a, b, and c depend on E, this is a nonlinear equation.

By using previously designated bounds and equations (3.13), (3.14), we find that the coefficients and the nonhomogeneous term in equation (5.6) are bounded as follows:

$$k_0' \leq a_{i,j},\ c_{i,j} \leq k_1$$

$$0 \leq b_{i,j} \leq k_1$$

$$|d_{i,j}| \leq k_2\{[\bar{M}_2 + h^2\bar{M}_4/12] + 2[\bar{Q}_2 + k^2\bar{Q}_4/12] + [\bar{N}_2 + h^2\bar{N}_4/12] + 1\}$$

where k_2 is a bound on the first partial derivatives of the functions A, B, C, and G. For h less than any designated value, say unity, a finite constant k_3 can be chosen such that

$$|d_{i,j}| \leq k_3.$$

Similarly, the coefficient $e_{i,j}$ is bounded in absolute value by k_3 for h less than unity. For the nonhomogeneous term, we have

$$|g_{i,j}| \le h^2 \{2k_2[\bar{M}_3 + \bar{N}_3][\bar{M}_2 + h^2\bar{M}_4/12 + 2\bar{Q}_2 + \eta^2\bar{Q}_4/3 + \bar{N}_2 + h^2\bar{N}_4/12 + 1]$$

$$+ k_1[\bar{M}_4 + 2\bar{Q}_4 + \bar{N}_4]\} /12.$$

For h less than unity, a finite constant k_4 can be chosen such that

$$|g_{i,j}| < h^2 k_4 .$$

Now, consider the coefficient $f_{i,j}$. Each term in the coefficient $f_{i,j}$ is bounded for finite h; however, we require also that $f_{i,j}$ be non-positive, i.e., that

$$\left[\int_0^1 \bar{A}_r d\theta\right]_{i,j} \left[(\Delta\nabla)_x u_{i,j}\right] + 2\left[\int_0^1 \bar{B}_r d\theta\right]_{i,j} \left[(\Delta\nabla)_z u_{i,j}\right]$$

$$+ \left[\int_0^1 \bar{C}_r d\theta\right]_{i,j} \left[(\Delta\nabla)_y u_{i,j}\right] - \left[\int_0^1 \bar{G}_r d\theta\right]_{i,j} \le 0 . \tag{5.8}$$

Inequality (5.8) is established by the use of condition (2.10). By substituting for $\partial^2 u/\partial x\partial y$ in condition (2.10) from equation (3.2), we obtain

$$\partial^2 u/\partial x^2 \int_0^1 (\partial A/\partial r - \cot\tau\ \partial B/\partial r)d\theta + 2\ \partial^2 u/\partial z^2 \int_0^1 (\cos 2\tau\ \partial B/\partial r)d\theta$$

$$+ \partial^2 u/\partial y^2 \int_0^1 (\partial C/\partial r - \tan\tau\ \partial B/\partial r)d\theta - \int_0^1 \partial G/\partial r\ d\theta \le \Delta v$$

From equations (3.4), this inequality can be written as

$$\partial^2 u/\partial x^2 \int_0^1 \partial A'/\partial r \, d\theta + 2\, \partial^2 u/\partial z^2 \int_0^1 \partial B'/\partial r \, d\theta$$

$$+ \partial^2 u/\partial y^2 \int_0^1 \partial C'/\partial r \, d\theta - \int_0^1 \partial G/\partial r \, d\theta \leq \Delta v \tag{5.9}$$

where

$$A' = A'(x,y,\ u+\theta v,\ \partial(u+\theta v)/\partial x,\ \partial(u+\theta v)/\partial y)$$

etc.

Inequality (5.8) can be written in the following form:

$$(\partial^2 u/\partial x^2 + h^2 M_4/12)_{i,j} \left[\int_0^1 \bar{A}_r d\theta\right]_{i,j} + 2(\partial^2 u/\partial z^2 + k^2 Q_4/12)_{i,j} \cdot$$

$$\left[\int_0^1 \bar{B}_r d\theta\right]_{i,j} + (\partial^2 u/\partial y^2 + h^2 N_4/12)_{i,j} \left[\int_0^1 \bar{C}_r d\theta\right]_{i,j} \tag{5.10}$$

$$- \left[\int_0^1 \bar{G}_r d\theta\right]_{i,j} \leq 0$$

where

$$\left[\bar{A}_r\right]_{i,j} = \left[\partial A'(x,y,u+\theta E,(\Delta+\nabla)_x(u+\theta E),(\Delta+\nabla)_y(u+\theta E))/\partial r\right]_{i,j}$$

etc. In order to be able to compare (5.10) with (5.9), we use relationships such as equation (3.13) to expand the quantities $\left[\int_0^1 \bar{A}_r d\theta\right]$, etc. as follows:

$$\left[\int_0^1 \bar{A}_r d\theta\right]_{i,j} = \left[\int_0^1 \partial A'(x,y,u+\theta E,\partial u/\partial x + \theta(\Delta+\nabla)_x E,\right.$$

$$\left.\partial u/\partial y + \theta(\Delta+\nabla)_y E)/\partial r\ d\theta\right]_{i,j} + \left[h^2 M_3 \int_0^1\int_0^1 \hat{A}_{rp} d\theta d\emptyset\right]_{i,j}/6 \tag{5.11}$$

$$+ \left[h^2 N_3 \int_0^1\int_0^1 \hat{A}_{rq} d\theta\ d\emptyset\right]_{i,j}/6$$

where

$$\left[\hat{A}_{rp}\right]_{i,j} = \left[\partial^2 A'(x,y,u+\theta E,\ \partial u/\partial x + \theta(\Delta+\nabla)_x E + \emptyset h^2 M_3/6,\ \partial u/\partial y + \theta(\Delta+\nabla)_y E\right.$$

$$\left. + \emptyset h^2 N_3/6)\partial r\partial p\right]_{i,j}$$

and $[\hat{A}_{rq}]_{i,j}$ is defined similarly.

We define a differentiable function v on $R + S$ which is equal to $E_{i,j}$ at the mesh points (x_i,y_j) and which has derivatives given by

$$(\partial v/\partial x)_{i,j} = (\Delta+\nabla)_x E_{i,j} \quad \text{and}$$

$$(\partial v/\partial y)_{i,j} = (\Delta+\nabla)_y E_{i,j}. \tag{5.12}$$

By using (5.11) and (5.12), inequality (5.10) can be written as

$$\left[\partial^2 u/\partial x^2 \int_0^1 \partial A'/\partial r\ d\theta + 2\ \partial^2 u/\partial z^2 \int_0^1 \partial B'/\partial r\ d\theta + \partial^2 u/\partial y^2 \int_0^1 \partial C'/\partial r\ d\theta\right.$$

$$\left. - \int_0^1 \partial G/\partial r\ d\theta\right]_{i,j} \leq -h^2 \left\{ M_4 \int_0^1 \bar{A}_r d\theta + 2\ \partial^2 u/\partial x^2 \left[M_3 \int_0^1\int_0^1 \hat{A}_{rp} d\theta d\emptyset\right.\right.$$

$$\left. + N_3 \int_0^1\int_0^1 \hat{A}_{rq} d\theta d\emptyset\right] + 4\eta^2 Q_4 \int_0^1 \bar{B}_r d\theta + 4\ \partial^2 u/\partial z^2 \left[M_3 \int_0^1\int_0^1 \hat{B}_{rp} d\theta d\emptyset\right.$$

$$+ N_3\int_0^1\int_0^1 \widehat{\tilde{B}}_{rq}d\theta d\phi\Big] + N_4\int_0^1 \bar{C}_r d\theta + 2\,\partial^2u/\partial y^2\Big[M_3\int_0^1\int_0^1 \widehat{\tilde{C}}_{rp}d\theta d\phi$$

$$+ N_3\int_0^1\int_0^1 \widehat{\tilde{C}}_{rq}d\theta d\phi\Big] - M_3\int_0^1\int_0^1 \widehat{\tilde{G}}_{rp}d\theta d\phi - N_3\int_0^1\int_0^1 \widehat{\tilde{G}}_{rq}d\theta d\phi\Big\}_{i,j} /12 \;. .$$

All quantities in the curly bracket are bounded; therefore, there exists a constant k_5 such that

$$\text{(5.13)}\qquad \Big[\partial^2u/\partial x^2\int_0^1 \partial A'/\partial r\; d\theta + 2\,\partial^2u/\partial z^2\int_0^1 \partial B'/\partial r\; d\theta$$

$$+ \partial^2u/\partial y^2\int_0^1 \partial C'/\partial r\; d\theta - \int_0^1 \partial G/\partial r\; d\theta\Big]_{i,j} \le h^2k_5 .$$

If $\nu(x_i,y_j)$ is nonzero, the validity of (5.13) is established at the mesh point (x_i,y_j) for $h < (\Delta\nu/k_5)^{1/2}$ by inequality (5.9), i.e., $f_{i,j}$ is nonpositive. If $\nu(x_i,y_j)$ is zero, $f_{i,j}$ is zero, and thus nonpositive, for all h.

We consider next the finite difference equation for the error in the solution of problem P_2 at irregular mesh points. This equation can take several forms depending on which of the six neighbors of an irregular mesh point (x_i,y_j) are not in $R + S$. Because of symmetry, it is only necessary to consider an irregular mesh point for which one rectangular neighbor and one diagonal neighbor are not in $R + S$ to illustrate the several possibilities.

Suppose for a mesh point (x_i,y_j) that $\gamma_{i,j} = 3$ and that both the diagonal neighbor (x_{i-1},y_{j-3}) and the rectangular neighbor (x_{i-1},y_j) are not in $R + S$ (see Figure 5.1); then (x_i,y_j) is an irregular mesh point. Equation (4.5) for $U_{i,j}$ is

$$L_{b2}U_{i,j} = A'(x_i,y_j,U_{i,j},D_xU_{i,j},D_yU_{i,j})2[U_{i+1,j}/(\lambda_1+1) - U_{i,j}/\lambda_1 + U_{i-\lambda_1,j}/\lambda_1(\lambda_1+1)]/h^2 + 2B'(---)[U_{i+1,j+3}/(\lambda_2+1)$$

$$(5.14) \quad - U_{i,j}/\lambda_2 + U_{i-\lambda_2,j-3\lambda_2}/\lambda_2(\lambda_2+1)]/5h^2 + C'(---)[U_{i,j+1}-2U_{i,j}+U_{i,j-1}]/h^2 = G(---).$$

where D_x and D_y denote applicable difference approximations to $\partial/\partial x$ and $\partial/\partial y$ respectively. The finite difference equation for the error is obtained from equation (5.14) by the same procedure that is used for regular mesh points. For the irregular mesh point considered above, the finite difference equation for the error has the form

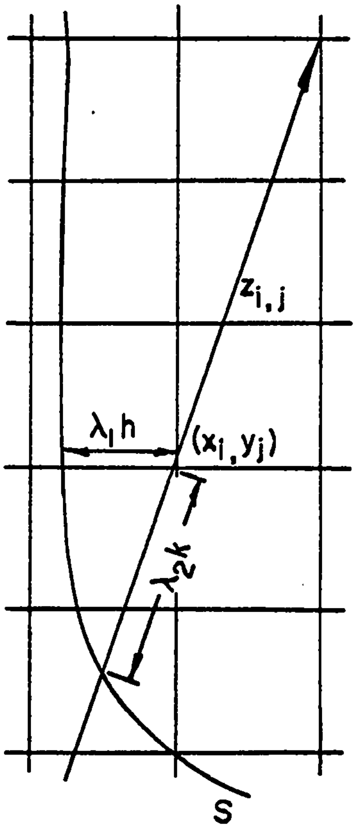


Figure 5.1

$$a_{i,j}2[E_{i+1,j}/(\lambda_1+1) - E_{i,j}/\lambda_1 + E_{i-\lambda,j}/\lambda_1(\lambda_1+1)]/h^2$$

$$+ 2b_{i,j}[E_{i+1,j+3}/(\lambda_2+1) - E_{i,j}/\lambda_2 + E_{i-\lambda_2,j-3\lambda_2}/\lambda_2(\lambda_2+1)]/5h^2$$

$$(5.15) \quad + c[E_{i,j+1} - 2E_{i,j} + E_{i,j-1}]/h^2 + d_{i,j}[E_{i+1,j}-E_{i-\lambda_1,j}]/(\lambda_1+1)h$$

$$+ e_{i,j}[E_{i,j+1}-E_{i,j-1}]/2h + f_{i,j}E_{i,j} = \bar{g}_{i,j}.$$

In equation (5.15), the coefficients and nonhomogeneous term are given by the same expressions as for a regular mesh point but where the difference quotients are those used in equation (5.14) rather than the symmetric difference quotients previously used. Also, the nonhomogeneous term $\bar{g}_{i,j}$ is proportional to h rather than h^2. Let k_6 be a constant such that $\bar{g}_{i,j} \leq hk_6$ for (x_i,y_j) an irregular mesh point.

The coefficients are bounded in the same way as for regular mesh points. In order to assure that $f_{i,j}$ is nonpositive at points where $\nu(x_i,y_j)$ is nonzero, it is necessary to specify that h be less than or equal to $(\Delta\nu/k_7)$ where k_7 is a constant which corresponds to k_5 for regular mesh points.

We denote the finite difference operator for the error in the solution of problem P_2 at irregular mesh points by

$$\tilde{L}_{b2}E_{i,j} = \bar{g}_{i,j}.$$

The error equation corresponding to equation (4.2) takes a somewhat different form. For the mesh point configuration given in Figure 3.5a, we have from equations (3.15) and (4.2):

$$L_{b1}(u_{i,j}+E_{i,j}) = \lambda(u_{i+1,j}+E_{i+1,j})/(\lambda+1) + (u_{p,q}+E_{p,q})/(\lambda+1)$$

$$- (u_{i,j}+E_{i,j}) = 0.$$

Since $E_{p,q} = 0$, we have

$$E_{i,j} = (\lambda E_{i+1,j} + \lambda u_{i+1,j} + u_{p,q})/(\lambda+1) - u_{i,j}$$

which by the use of equation (3.22) can be written as

(5.16) $$E_{i,j} = (\lambda E_{i+1,j} + \lambda h^2[\lambda(M_2)_{i-\theta,j} + (M_2)_{i+\theta,j}]/2)/(\lambda+1).$$

For the general case, the rightmost term in equation (5.16) will be denoted by $g'_{i,j}$.

We define the finite difference operator $\tilde{L}_{b1}$ by

$$\tilde{L}_{b1}E_{i,j} = \lambda E_{m,n}/(\lambda+1) - E_{i,j}$$

where

$$(x_i,y_j) \in R_b \quad \text{and} \quad (x_m,y_n) \in R_h.$$

The error equations derived above can be used together with zero boundary values to formulate boundary value problems for the error. The boundary value problems for the error functions associated with problems P_1 and P_2 are:

Problem $\tilde{P}_1$:

(5.17) $\qquad \tilde{L}_n E_{i,j} = g_{i,j}, \ |g_{i,j}| \leq h^2 k_4, \ (x_i,y_j) \in R_h$

(5.18) $\qquad \tilde{L}_{b1} E_{i,j} = g'_{i,j}, \ |g'_{i,j}| \leq h^2 k_8, \ (x_i,y_j) \in R_b$

(5.19) $\qquad E(x,y) = 0 \qquad\qquad , \ (x,y) \in R_S$

Problem P_2:

(5.20) $\qquad \tilde{L}_h E_{i,j} = g_{i,j}, \ |g_{i,j}| \leq h^2 k_4, \ (x_i,y_j) \in R_h$

(5.21) $\qquad \tilde{L}_{b2} E_{i,j} = \bar{g}_{i,j}, \ |\bar{g}_{i,j}| \leq hk_6, \ (x_i,y_j) \in R_b$

$\qquad E(x,y) = 0 \qquad\qquad , \ (x,y) \in R_S.$

CHAPTER VI

LINEAR DIFFERENCE OPERATORS

In order to derive error bounds for the solutions of problems P_1 and P_2, it is necessary to first establish some properties of linear difference operators.

Let a, b, c, d, e, and f be functions of x and y only which satisfy the following conditions for (x_i, y_j) a mesh point in R+S.

$$
\begin{aligned}
&0 < K_0 \leqq a_{i,j},\ c_{i,j} \leqq K_1 \\
&0 \leqq b_{i,j} \leqq K_1 \\
&|d_{i,j}|,\ |e_{i,j}| \leqq K_1 \\
&0 \leqq f_{i,j}
\end{aligned}
\tag{6.1}
$$

where K_0 and K_1 are finite constants. Also, let

$$h_1 \geqq 2K_0/K_1. \tag{6.2}$$

Let $v_{i,j}$ be an arbitrary function defined at the mesh points in $R_h + R_b + R_S$. At the mesh points in R_h, we define the finite difference operator $\bar{L}_h$ by

$$
\begin{aligned}
\bar{L}_h v_{i,j} = {} & a_{i,j}(\Delta\nabla)_x v_{i,j} + 2b_{i,j}(\Delta\nabla)_z v_{i,j} + c_{i,j}(\Delta\nabla)_y v_{i,j} \\
& + d_{i,j}(\Delta+\nabla)_x v_{i,j} + e_{i,j}(\Delta+\nabla)_y v_{i,j} + f_{i,j} v_{i,j}
\end{aligned}
\tag{6.3}
$$

At the mesh points in R_b, finite difference operators $\bar{L}_{b1}$ and $\bar{L}_{b2}$ are defined. The operator $\bar{L}_{b1}$ is the same as the operator L_{b1}

given by equations (3.15) and (3.16). The operator $\bar{L}_{b2}$ has the same form as the operator $\bar{L}_h$ but utilizes difference quotients such as those given by equations (3.17)-(3.21) rather than the symmetric difference quotients used above.

Just as the maximum of a function v, continuous on $R+S$, for which $Lv \geqq 0$ in R is less than or equal to the maximum of zero and the maximum of v on S, we prove

<u>LEMMA 6.1</u>. Let $v_{i,j}$ be an arbitrary function defined at the mesh points in $R_h + R_b + R_S$ such that

$$\bar{L}_h v_{i,j} \geqq 0, \qquad (x_i,y_j) \in R_h,$$
$$\bar{L}_{b2} v_{i,j} \geqq 0, \qquad (x_i,y_j) \in R_b.$$

If the mesh width h is less than h_1, then

$$v_{i,j} \leqq \max\ (0, \max_{R_S} v)$$

for all $(x_i,y_j) \in R_h + R_b$.

<u>Proof</u>: Define $\psi_{i,j}$ by

$$(6.4) \qquad \begin{aligned} \bar{L}_h v_{i,j} &= \psi_{i,j},\ (x_i,y_j) \in R_h, \\ \bar{L}_{b2} v_{i,j} &= \psi_{i,j},\ (x_i,y_j) \in R_b, \end{aligned}$$

then

$$\psi_{i,j} \geqq 0.$$

At the mesh points $(x_i,y_j) \in R_h$, equations (6.4) can be written as

$$v_{i,j} = (\mu_{i+1,j}/\mu_{i,j})v_{i+1,j}+(\mu_{i-1,j}/\mu_{i,j})v_{i-1,j} + (\mu_{i\pm\beta,j+\alpha}/\mu_{i,j})v_{i\pm\beta,j+\alpha}$$

$$\text{(6.5)} \qquad + (\mu_{i\mp\beta,j-\alpha}/\mu_{i,j})v_{i\mp\beta,j-\alpha} + (\mu_{i,j+1}/\mu_{i,j})v_{i,j+1} + (\mu_{i,j-1}/\mu_{i,j})v_{i,j-1}$$

$$- (h^2/\mu_{i,j})\psi_{i,j},$$

where

$$\mu_{i,j} = 2(a_{i,j}+2b_{i,j}/(\alpha^2+\beta^2) + c_{i,j} - h^2f_{i,j}/2),$$

$$\mu_{i+1,j} = (a_{i,j}+hd_{i,j}/2),\ \mu_{i-1,j} = (a_{i,j}-hd_{i,j}/2),$$

$$\mu_{i\pm\beta,j+\alpha} = 2b_{i,j}/(\alpha^2+\beta^2),\ \mu_{i\mp\beta,j-\alpha} = 2b_{i,j}/(\alpha^2+\beta^2),$$

$$\mu_{i,j+1} = (c_{i,j}+he_{i,j}/2),\ \mu_{i,j-1} = (c_{i,j}-he_{i,j}/2).$$

At the mesh points $(x_i,y_j) \in R_b$, the function $v_{i,j}$ is given by equations similar to (6.5). These equations involve values of v at one or more boundary mesh points, and values of the coefficients corresponding to these mesh points depend on which of the mesh points in $N(x_i,y_j)$ are not in $R + S$. By way of illustration, the value of the function $v_{i,j}$ is given below for the irregular mesh point illustrated in Figure 5.1.

$$v_{i,j} = (\mu_{i+1,j}/\mu_{i,j})v_{i+1,j}+(\mu_{i\pm\lambda_1,j}/\mu_{i,j})v_{i+\lambda_1,j}+(\mu_{i+1,j+3}/\mu_{i,j})v_{i+1,j+3}$$

$$\text{(6.6)} \qquad + (\mu_{i-\lambda_2,j-3\lambda_2}/\mu_{i,j})v_{i-\lambda_2,j-3\lambda_2} + (\mu_{i,j+1}/\mu_{i,j})v_{i+1,j}$$

$$+ (\mu_{i,j-1}/\mu_{i,j})v_{i,j-1} - (h^2/\mu_{i,j})\psi_{i,j}$$

where

$$\mu_{i,j} = 2(a_{i,j}/\lambda_1 + b_{i,j}/5\lambda_2 + c_{i,j} - h^2 f_{i,j}/2),$$

$$\mu_{i+1,j} = (2a_{i,j}+hd_{i,j})/(\lambda_1+1),\ \mu_{i+1,j+3} = 2b_{i,j}/5(\lambda_2+1),$$

$$\mu_{i-1,j} = (2a_{i,j}-h\lambda_1 d_{i,j})/\lambda_1(\lambda_1+1),\ \mu_{i-\lambda_2,j-3\lambda_2} = 2b_{i,j}/5\lambda_2(\lambda_2+1),$$

$$\mu_{i,j+1} = (c_{i,j}+he_{i,j}/2),\ \mu_{i,j-1} = (c_{i,j}-he_{i,j}/2).$$

Equation (6.6) is easily generalized to apply to any mesh point in R_b.

For $h < h_1$, the coefficients in both equations (6.5) and (6.6) satisfy

$$0 < \mu_{i,j}, \mu_{m,n},$$

and

$$\frac{1}{\mu_{i,j}} \sum_{(m,n)} \mu_{m,n} \leq 1$$

where the subscripts (m,n) take on all values of the subscripts included in equation (6.5) and (6.6) except (i,j).

Now, let $M = \max v_{i,j}$, $(x_i,y_j) \in R_h + R_b + R_S$. If $0 < M$ and $v_{i,j} = M$ at a point $(x_i,y_j) \in R_h + R_b$, then $v_{m,n} = M$ at each point $(x_m,y_n) \in R_h + R_b + R_S$ which is associated with (x_i,y_j) by the appropriate equation (6.5) or (6.6). If one of the points (x_m,y_n) is a point in R_S, the lemma is proved for the point (x_i,y_j). Otherwise, the same argument applies to each of the neighbors of the original point until a point which is associated with a point in R_S is reached.

If $v_{i,j}$ is nonpositive for all $(x_i,y_j) \in R_S$, then $v_{m,n}$ is nonpositive for all (x_m,y_n) in $R_h + R_b$. Therefore,

$$v_{i,j} \leq \max\,(0, \max_{R_S} v).$$

We consider next the boundary value problem given by

$$(6.7) \qquad \bar{L}_h v_{i,j} = t_{i,j} \quad , \quad (x_i,y_j) \in R_h$$

$$(6.8) \qquad v(x,y) = \phi(x,y) \ , \quad (x,y) \in R_S$$

where $t_{i,j}$ and $\phi_{i,j}$ are given functions, and all mesh points in R are assumed to be regular mesh points. We prove

LEMMA 6.2. Let $h < h_1$; then the boundary value problem given by equations (6.7) and (6.8) has a unique solution.

Proof: We first show that if a solution of the boundary value problem exists, it is unique. Suppose $v_{i,j}$ and $v'_{i,j}$ are two solutions of the given problem. Then $v''_{i,j} = (v_{i,j} - v'_{i,j})$ is a solution of the problem

$$\bar{L}_h v''_{i,j} = 0, \ (x_i,y_j) \in R_h$$

$$v''_{i,j} = 0, \ (x_i,y_j) \in R_S.$$

By Lemma 6.1, any solution of this problem is bounded above by zero. Similarly, by considering the function $-v''_{i,j}$, it is proved that any solution is bounded below by zero. Therefore, $v''_{i,j} \equiv 0$, and $v_{i,j} = v'_{i,j}$.

The determination of $v_{i,j}$ at any point (x_i,y_j) requires the solution of a set of linear algebraic equations with as many equations as unknowns. The uniqueness of the solution implies that the determinant of the matrix of coefficients is nonzero, i.e., the matrix of coefficients is nonsingular.

In this case, as is well known, the set of equations has one and only one solution.[1]

Next, we establish the existence of a bound on the solution of the boundary value problem given by equations (6.7) and (6.8).

Let σ_1 and the function $p_{i,j}$ be defined by

$$\sigma_1 = [1+(h^2+h[4K_0+K_1^2+h^2]^{1/2})/2K_0] \;/[1-hK_1/2K_0], \; h < h_1 \tag{6.9}$$

and

$$p_{i,j} = \sigma_1^I - \sigma_1^i, \; i = 0, 1, 2, \ldots, I. \tag{6.10}$$

We first establish some properties of the function $p_{i,j}$ by means of the following lemmas.

<u>LEMMA 6.3</u>. Let $h < h_1$; then if R contains only regular mesh points, the second difference quotient with respect to z of the function $p_{i,j}$ is nonpositive.

<u>Proof</u>: Let (x_i, y_j) be a regular mesh point and let $\gamma_{i,j} = \pm \alpha/\beta$. The second difference quotient with respect to z of the function $p_{i,j}$ is given by

$$(\Delta\nabla)_z p_{i,j} = (\Delta\nabla)_z(\sigma_1^I - \sigma_1^i) = -(\Delta\nabla)_z \sigma_1^i$$
$$= -[\sigma_1^{i+\beta} - 2\sigma_1^i + \sigma_1^{i-\beta}]/k^2 = -\sigma_1^i[\sigma_1^\beta - 2 + \sigma_1^{-\beta}]/k^2.$$

[1] See, for instance, Milne [1949], p. 8.

For $h < h_1$, $\sigma_1 \geqq 1$. Consider the function

$$F = F(\sigma_1) = \sigma_1^{\beta} - 2 + \sigma_1^{-\beta}, \quad \beta \geqq 1.$$

For $\sigma_1 = 1$, $F = 0$, and

$$dF/d\sigma_1 = \beta(\sigma_1^{\beta-1} - \sigma_1^{-\beta-1}).$$

For all β, $\sigma_1 \geqq 1$,

$$dF/d\sigma_1 \geqq 0;$$

therefore,

$$F \geqq 0.$$

always, and the value $(\Delta\nabla)_z p_{i,j}$ is nonpositive.

LEMMA 6.4. Let $h < h_1$; then if R contains only regular mesh points,

$$\bar{L}_h p_{i,j} \leqq -1.$$

Proof: By direct substitution of (6.10), we have

$$\bar{L}_h p_{i,j} = \bar{L}_h \sigma_1^I - \bar{L}_h \sigma_1^i.$$

$$= -a_{i,j}(\Delta\nabla)_x \sigma_1^i - 2b_{i,j}(\Delta\nabla)_z \sigma_1^i - d_{i,j}(\Delta+\nabla)_x \sigma_1^i + f_{i,j}(\sigma_1^I - \sigma_1^i).$$

By condition (6.1) and Lemma 6.3,

$$\begin{aligned} \bar{L}_n p_{i,j} &\leqq -a_{i,j}(\Delta\nabla)_x \sigma_1^i - d_{i,j}(\Delta+\nabla)_x \sigma_1^i \\ &\leqq -\sigma_1^i[a_{i,j}(\sigma_1 - 2 + \sigma_1^{-1})/h^2 + d_{i,j}(\sigma_1 - \sigma_1^{-1})/2h] \\ &\leqq -K_0(\sigma_1 - 2 + \sigma_1^{-1})/h^2 + K_1(\sigma_1 - \sigma_1^{-1})/2h . \end{aligned}$$

It is verified by direct calculation that σ_1 as defined by equation (6.9) is a solution of

$$(6.11) \qquad -K_0(\sigma_1-2+\sigma_1^{-1})/h^2 + K_1(\sigma_1-\sigma_1^{-1})/2h = -1$$

which proves the lemma.

Next, we prove

THEOREM 6.1. Assume that R contains only regular mesh points, and let $v_{i,j}$ be the solution of the boundary value problem given by equations (6.7) and (6.8). Then, for $h < h_1$,

$$\max_{R_h} |v_{i,j}| \leq p_{i,j} \max_{R_h} |t_{i,j}| + \max_{R_S} |\phi_{i,j}|.$$

Proof: Let

$$q_{i,j} = p_{i,j} \max_{R_h} |t_{i,j}| + \max_{R_S} |\phi_{i,j}|.$$

From condition 6.1 and Lemma 6.4,

$$\begin{aligned} L_h q_{i,j} &= \max_{R_h} |t_{i,j}| \bar{L}_h p_{i,j} + \max_{R_S} |\phi_{i,j}| f_{i,j} \\ &\leq -\max_{R_h} |t_{i,j}|. \end{aligned}$$

Also,

$$q_{i,j} \geq \max_{R_S} |\phi_{i,j}|$$

at mesh points in R_S. Hence

$$w_{i,j} = v_{i,j} - q_{i,j}$$

is nonpositive on R_S and

$$\bar{L}_h w_{i,j} = t_{i,j} - \bar{L}_h q_{i,j}$$

is nonnegative for $(x_i, y_j) \in R_h$. Therefore, by Lemma 6.1, $w_{i,j}$ is nonpositive in R_h, and

$$v_{i,j} \leq q_{i,j} = p_{i,j} \max_{R_h} |t_{i,j}| + \max_{R_S} |\phi_{i,j}|.$$

A similar argument holds when $v_{i,j}$ is replaced by $-v_{i,j}$, and we obtain

$$\max_{R_h} |v_{i,j}| \leq p_{i,j} \max_{R_h} |t_{i,j}| + \max_{R_S} |\phi_{i,j}|$$

which was to be proved.

COROLLARY 1. Let μ_1 be given by

$$\mu_1 = [K_1 + (K_1^2+4K_0)^{1/2}]/2K_0$$

then there exists an h_2 such that for $h < h_2$, the solution of the problem given by equations (6.7) and (6.8) is bounded as follows:

$$\max_{R_h} |v_{i,j}| \leq e^{\mu_1 X} \max_{R_h} |t_{i,j}| + \max_{R_S} |\phi_{i,j}|.$$

Proof: By virtue of Theorem 6.1, we need only show that there exists an $h_2 < h_1$ such that for $h < h_2$

$$p_{i,j} \leq e^{\mu_1 X},$$

i.e., that

$$\sigma_1^I - \sigma_1^i \leq e^{\mu_1 X}; \quad \sigma \geq 1,\ I \geq i \geq 0.$$

Since $\sigma_1^i \geqq 1$, it is sufficient to show that

$$\sigma_1^I \leqq e^{\mu_1 X} + 1.$$

Now,

$$\sigma_1^I = (\sigma_1^{1/h})^{Ih} = (\sigma_1^{1/h})^X$$

and

$$(\sigma_1^{1/h}) = \left\{(1+[h^2+h(4K_0+K_1^2+h^2)^{1/2}]/2K_0)/(1-hK_1/2K_0)\right\}^{1/h}$$

$$= \left[1+h(4K_0+K_1^2)^{1/2}/2K_0 + 0(h^2)\right]^{1/h} \left[1 + hK_1/2K_0 + 0(h^2)\right]^{1/h}$$

$$(\sigma_1^{1/h}) = [1+h\mu_1]^{1/h} + (1/h)[1+h\mu_1]^{(1/h)-1}\, 0(h^2) + \ldots \tag{6.12}$$

As h tends to zero, the first term on the right side of (6.12) tends to e^{μ_1} and the sum of the successive terms to zero. Therefore, there exists an $h_2 < h_1$ such that for $h < h_2$,

$$\sigma^I - 1 \leqq e^{\mu_1 X},$$

and by Theorem 6.1,

$$\max_{R_h} |v_{i,j}| \leqq e^{\mu_1 X} \max_{R_h} |t_{i,j}| + \max_{R_S} |\phi_{i,j}|.$$

We now return to the case of a general region R, i.e., a region containing both regular and irregular mesh points.

LEMMA 6.5. Let $v_{i,j}$ be an arbitrary function defined at the mesh points in $R_h + R_b + R_S$ such that

$$\bar{L}_h v_{i,j} \geq 0 , \quad (x_i,y_j) \in R_h$$

$$\bar{L}_{b1} v_{i,j} = 0 , \quad (x_i,y_j) \in R_b$$

$$v(x,y) = \phi(x,y), (x,y) \in R_S .$$

Then, for $h < h_1$,

(a) $$v_{i,j} \leq \max (0, \max_{R_b+R_S} v), (x_i,y_j) \in R_h$$

and

(b) $$v_{i,j} \leq \max (0, \max_{R_S} v), (x_i,y_j) \in R_h + R_b.$$

Proof: Part (a) of this lemma is proved by applying Lemma 6.1 directly. Part (b) is proved by the same argument that was used to prove Lemma 6.1. At a mesh point $(x_i,y_j) \in R_h + R_b$, $v_{i,j}$ is given as a weighted average with positive weights whose sum does not exceed unity of neighboring values $v_{m,n}$.

Consider next the boundary value problem given by

(6.13) $$\bar{L}_h v_{i,j} = t_{i,j} \quad , (x_i,y_j) \in R_h$$

(6.14) $$\bar{L}_{b1} v_{i,j} = t'_{i,j} \quad , (x_i,y_i) \in R_b$$

(6.15) $$v(x,y) = \phi(x,y) \quad , \quad (x,y) \in R_S$$

where $t_{i,j}$ and $t'_{i,j}$ are given functions. Then, we have

LEMMA 6.6. For $h < h_1$, the boundary value problem given by equations (6.13)-(6.15) has a unique solution.

Proof: This lemma is proved by the same argument that was used to prove Lemma 6.2.

Next, we prove

THEOREM 6.2. Let $v_{i,j}$ be the solution of the boundary value problem given by equations (6.13)-(6.15) where $\phi(x,y) = 0$, $(x,y) \in R_S$. For $h < h_2$

$$\max_{R_h+R_b} |v_{i,j}| \leq 2e^{\mu_1 X} \max_{R_h} |t_{i,j}| + 2 \max_{R_b} |t'_{i,j}|. \tag{6.16}$$

Proof: From Lemma 6.6, we know that the solution $v_{i,j}$ is unique. We apply Theorem 6.1 to $v_{i,j}$, $(x_i,y_j) \in R_h$ where the mesh points in R_b and those mesh points in R_S which are adjacent to mesh points in R_h are considered as boundary mesh points. Thus, for $h < h_2$,

$$\max_{R_h} |v_{i,j}| \leq e^{\mu_1 X} \max_{R_h} |t_{i,j}| + \max_{R_b} |v_{i,j}|. \tag{6.17}$$

Now, let (x_i,y_j) be an arbitrary mesh point in R_b. From the definition of $\bar{L}_{b1}$, $v_{i,j}$ is given by an equation of the form

$$v_{i,j} = v_{p,q}/(\lambda+1) + \lambda v_{m,n}/(\lambda+1) - t'_{i,j},\ 0 < \lambda < 1,$$

where (x_p,y_q) is a mesh point in R_S and (x_m,y_n) is a mesh point in R_h. Since $v_{p,q} = 0$ and $\lambda/(\lambda+1) \leq 1/2$, we have

$$\max_{R_b} |v_{i,j}| \leq 1/2 \max_{R_h} |v_{i,j}| + \max_{R_b} |t'_{i,j}|. \tag{6.18}$$

By substituting successively for $\max\limits_{R_b} |v_{i,j}|$ in (6.17) from (6.18) and for $\max\limits_{R_h} |v_{i,j}|$ in (6.18) from (6.17), we obtain (6.16).

Next, we establish a bound on the solution of the boundary value problem given by

(6.19) $$\bar{L}_h v_{i,j} = t_{i,j} \quad , (x_i,y_j) \in R_h$$

(6.20) $$L_{b2} v_{i,j} = \bar{t}_{i,j} \quad , (x_i,y_j) \in R_b$$

(6.21) $$v_{i,j} = \phi_{i,j} \quad , (x_i,y_j) \in R_S$$

We have

<u>LEMMA 6.7</u>. For $h < h_1$, the solution of the boundary value problem given by equations (6.19)-(6.21) exists and is unique.

<u>Proof</u>: This lemma is proved by the same argument that is used to prove Lemma 6.2.

Let h_3, σ_2, and the function $\tilde{p}_{i,j}$ be defined by

(6.22) $$h_3 = \min(h_3', h_3'')$$

where

(6.23) $$h_3' = K_0[1 - (2/3)^{1/\eta}]/4K_1,$$

and

(6.24) $$h_3'' = K_0[(3/2)^{1/\eta} - 1]/2\{1 + [2K_0 + 4K_3^2 + (h_3')^2]^{1/2}\},$$

(6.25) $$\sigma_2 = [1 + h^2/K_0 + h(2K_0 + 4K_1^2 + h^2)^{1/2}/K_0]/[1 - 2hK_1/K_0],$$

and

(6.26) $$\tilde{p}_{i,j} = \sigma_2^I - \sigma_2^{i\pm\lambda} \; ; \; i = 0, 1, 2, \ldots, I; \; 0 \leqq \lambda < 1.$$

Here, $\lambda = 0$ except when $\tilde{p}_{i,j}$ is evaluated at an irregular mesh point.

We prove

<u>LEMMA 6.8</u>. For $h < h_3$,

$$\sigma_2 \geqq 1$$

and

$$\sigma_2^\eta \leqq 3/2.$$

<u>Proof</u>: From equations (6.22) and (6.23)

$$h \leqq K_0[1 - (2/3)^{1/\eta}]/4K_1$$

$$< K_0/2K_1.$$

Therefore,

$$\sigma_2 \geqq 1.$$

From equations (6.22) and (6.24)

$$h \leqq K_0[(3/2)^{1/\eta}-1]/2\{1 + [2K_0+ 4K_1^2 + (h_3')^2]^{1/2}\}$$

or

$$h + h[2K_0+ 4K_1^2 + (h_3')^2] \leqq K_0[(3/2)^{1/\eta}-1]/2.$$

Obviously, h is less than unity; thus,

$$1 + h^2/K_0 + h[2K_0+ 4K_1^2 + h^2]^{1/2}/K_0 \leqq (3/2)^{1/\eta}[1 - (1-(2/3)^{1/\eta})/2],$$

and from equation (6.23)

$$1 + h^2/K_0 + h[2K_0 + 4K_1^2 + h^2]^{1/2}/K_0 \leq (3/2)^{1/\eta}[1 - 2K_1h/K_0].$$

Thus, from equation (6.25),

$$\sigma_2^{\eta} \leq 3/2.$$

<u>LEMMA 6.9</u>. For $h < h_3$, the second difference quotient with respect to z of the function $p_{i,j}$ is nonpositive.

<u>Proof</u>: The proof of this lemma for regular mesh points is the same as the proof of Lemma 6.3.

Let (x_i, y_j) be an irregular mesh point and let $\gamma_{i,j} = \pm\alpha/\beta$. We must consider two cases, i.e., either (but not both) of the diagonal neighbors $(x_i - \beta h,\ y_j \pm \alpha h)$ or $(x_i + \beta h,\ y_j \pm \alpha h)$ might not be in $R + S$ (since $\tilde{p}_{i,j}$ is independent of y, the sign of αh is immaterial).

If the diagonal neighbor $(x_i - \beta h,\ y_j \pm \alpha h) \notin R + S$, the second difference quotient with respect to z of $\tilde{p}_{i,j}$ is given by (see Figure 5.1)

$$(6.27) \qquad -[\sigma_2^{i+\beta}/(\lambda+1) - \sigma_2^{i}/\lambda + \sigma_2^{i-\lambda\beta}/\lambda(\lambda+1)]/k^2$$

where λk, $0 < \lambda < 1$, is the distance between the mesh point (x_i, y_j) and the point of intersection of S and $z_{i,j}$. We have

$$(6.28) \qquad [\sigma_2^{i+\beta}/(\lambda+1) - \sigma_2^{i}/\lambda + \sigma_2^{i-\lambda\beta}/\lambda(\lambda+1)]/k^2 = \sigma_2^{i}[\lambda\sigma_2^{\beta} - (\lambda+1) + \sigma_2^{-\lambda\beta}]/k^2\lambda(\lambda+1).$$

The coefficient of the square bracket on the right-hand side of (6.28) is nonnegative; thus, we need only show that

$$F = \lambda\sigma_2^{\beta} - (\lambda+1) + \sigma_2^{-\lambda\beta}$$

is nonnegative. Let

$$\sigma_2^{\beta} = (1+\epsilon),\ 0 \leqq \epsilon \leqq 1/2;$$

then

$$F = \lambda(1+\epsilon) - (\lambda+1) + (1+\epsilon)^{-\lambda}$$

$$\geqq \tilde{F} = \lambda(\lambda+1)\epsilon^2[1 - (\lambda+2)\epsilon/3]/2$$

and $\tilde{F}$ is nonnegative for $\epsilon \leqq 1/2$. By Lemma 6.7, $\sigma_2^{\beta} \leqq 3/2$ for $h < h_3$; therefore, (6.27) is nonpositive.

Now, suppose the point $(x_i+\beta h, y_j \pm \alpha h) \notin R + S$. Then, we must show that

$$-[\sigma_2^{i+\lambda\beta}/\lambda(\lambda+1) - \sigma_2^{i} + \sigma_2^{i-\beta}/(\lambda+1)]/k^2 \leqq 0 \tag{6.29}$$

or that

$$F = \sigma_2^{\lambda\beta} - (\lambda+1) + \lambda\sigma_2^{-\beta},\ 0 < \lambda < 1,\ \sigma_2 \geqq 1,$$

is nonnegative. Again, let $\sigma_2^{\beta} = 1 + \epsilon$, $0 \leqq \epsilon \leqq 1/2$; then

$$F = (1+\epsilon)^{\lambda} - (\lambda+1) + \lambda(1+\epsilon)^{-1}$$

$$F \geqq \tilde{F} = (1+\lambda-2\epsilon)\epsilon^2\lambda/2. \tag{6.30}$$

$$\tilde{F}|_{\lambda=0} = 0$$

$$\tilde{F}|_{\lambda=1} \geqq 0.$$

We need only show that F is positive for some pair of values of λ and ϵ in the interval $0 < \lambda < 1$, $0 \leqq \epsilon \leqq 1/2$ and that F has no zeros in this interval to complete the proof that F is nonnegative throughout the interval. Consider

$$\tilde{F}|_{(\lambda,\epsilon)=(1/2,1/4} = 1/64 > 0.$$

Suppose F has a zero in the interval of interest, then from (6.30),

$$1 + \lambda - 2\epsilon = 0$$

$$\lambda = 2\epsilon - 1,$$

but this relationship is not satisfied at any point in the interval, and we conclude that $\tilde{F}$ is nonnegative in the interval. Thus, (6.29) is nonpositive. This completes the proof of the lemma.

<u>LEMMA 6.10</u>. Let (x_i,y_j) be a mesh point in $R_h + R_b$. For $h < h_3$,

$$\bar{L}_h\hat{p}_{i,j} \leq -1, \quad (x_i,y_j) \in R_h$$

and

$$\bar{L}_{b2}\hat{p}_{i,j} \leq -1, \quad (x_i,y_j) \in R_b.$$

<u>Proof</u>: The proof of this lemma for mesh points in R_h is the same as the proof of Lemma 6.4 if h_3 and $\hat{p}_{i,j}$ are substituted for h_1 and $p_{i,j}$ respectively.

Suppose (x_i,y_j) is a mesh point in R_b. As in the case of regular mesh points, we can omit the difference quotient with respect to z by virtue of condition (6.1) and Lemma 6.9. Then, there are two cases that must be considered, i.e., either (but not both) of the rectangular neighbors (x_i-h,y_j) or (x_i+h,y_j) might not be in $R + S$.

First, we assume that the mesh point $(x_i-h,y_j) \notin R + S$. For this case

$$(6.31)\qquad \dot{\bar{L}}_{b2}\tilde{p}_{i,j} = -2a_{i,j}[\sigma_2^{i+1}/(\lambda+1) - \sigma_2^{i}/\lambda + \sigma_2^{i-\lambda}/\lambda(\lambda+1)]/h^2$$

$$- d_{i,j}[\sigma_2^{i+1}-\sigma_2^{i-\lambda}]/(\lambda+1)h + f(\sigma_2^{I}-\sigma_2^{i}),\ 0 < \lambda < 1$$

$$\leq \sigma_2^{i} \left\{-2K_0[\lambda\sigma_2-(1+\lambda)+\sigma_2^{-\lambda}]/h^2\lambda(\lambda+1) + K_1[\sigma_2-\sigma_2^{-\lambda}]/h(\lambda+1)\right\} .$$

Next, we show that

$$(6.32)\qquad 2[\lambda\sigma_2-(1+\lambda)+\sigma_2^{-\lambda}]/\lambda(\lambda+1) \geq [\sigma_2-2-\sigma_2^{-1}]/2.$$

Let $\sigma_2 = 1 + \epsilon,\ 0 \leq \epsilon \leq 1/2$, then

$$\lambda\sigma_2 - (1+\lambda) + \sigma_2^{-\lambda} = \lambda(1+\epsilon) - (1+\lambda) + (1+\epsilon)^{-\lambda}$$

$$\geq \epsilon^2[1 - (\lambda+2)\epsilon/3]/2.$$

Also,

$$\sigma_2 - 2 - \sigma_2^{-1} = 1 + \epsilon - 2 + (1+\epsilon)^{-1}$$

$$\leq \epsilon^2,$$

and we require only that

$$1 - (\lambda+2)\epsilon/3 \geq 1/2$$

which is true for $0 < \lambda < 1,\ 0 \leq \epsilon \leq 1/2$. This establishes (6.32). Now, we show that

$$[\sigma_2- \sigma_2^{-\lambda}]/(\lambda+1) \leq [\sigma_2- \sigma_2^{-1}].$$

This reduces to showing that

$$\sigma_2^{-\lambda} + \lambda\sigma_2 \geqq \lambda\sigma_2^{-1} + \sigma_2^{-1}$$

which is true for all λ, σ_2 in $0 < \lambda < 1$, $\sigma_2 \geqq 1$. From these results and (6.31), we have for the case when $(x_i - h, y_j) \notin R + S$,

$$\text{(6.33)} \qquad \bar{L}_{b2}\tilde{p}_{i,j} \leqq -K_0[\sigma_2 - 2 + \sigma_2^{-1}]/2h^2 + K_1[\sigma_2 - \sigma_2^{-1}]/h.$$

The quantity σ_2 is a solution of

$$-K_0(\sigma_2 - 2 + \sigma_2^{-1})/2h^2 + K_1(\sigma_2 - \sigma_2^{-1})/h = -1 ;$$

thus, the lemma is proved for the irregular mesh point being considered.

Now, suppose $(x_i + h, y_j) \notin R + S$. We must show that inequality (6.33) is satisfied for this configuration or that

$$\text{(6.34)} \qquad -2a_{i,j}[\sigma_2^{i+\lambda}/\lambda(\lambda+1) - \sigma_2^{i}/\lambda + \sigma_2^{i-1}/(\lambda+1)]/h^2 - d_{i,j}[\sigma_2^{i+\lambda} - \sigma_2^{i-1}]/(\lambda+1)h$$
$$\leqq -K_0[\sigma_2 - 2 + \sigma_2^{-1}]/2h^2 + K_1[\sigma_2 - \sigma_2^{-1}]/h.$$

First, we show that

$$\text{(6.35)} \qquad 2[\sigma_2^{\lambda} - (1+\lambda) + \lambda\sigma_2^{-1}]/\lambda(\lambda+1) \geqq [\sigma_2 - 2 + \sigma_2^{-1}]/2.$$

Let $\sigma_2 = 1 + \epsilon$, $0 \leqq \epsilon \leqq 1/2$. Then,

$$(1+\epsilon)^{\lambda+1} - (1+\lambda)(1+\epsilon) + \lambda - \epsilon^2\lambda(\lambda+1)/4 \geq 0.$$

This expression is greater than or equal to

$$1 + (1+\lambda)\epsilon + (1+\lambda)\lambda\epsilon^2/2 + (1+\lambda)\lambda(\lambda-1)\epsilon^3/6 - (1+\lambda)(1+\epsilon) + \lambda - \lambda(\lambda+1)\epsilon^2/4$$

and we need only show that

$$1/2 + (\lambda-1)\epsilon/3 \geq 0,$$

which is true for all λ,ϵ in $0 < \lambda < 1$, $0 \leqq \epsilon \leqq 1/2$.

In order to complete the proof of (6.34), we must also show that

$$[\sigma_2^{\lambda} - \sigma_2^{-1}]/(\lambda+1) \leqq \sigma_2 - \sigma_2^{-1}$$

or that

$$\sigma_2^{\lambda} + \lambda\sigma_2^{-1} \leqq \lambda\sigma_2 + \sigma_2$$

which is true for all λ,σ in $0 < \lambda < 1$, $\sigma_2 \geqq 1$.

This completes the proof of the lemma for all possible mesh point configurations.

The desired bound on the solution of the boundary value problem given by equations (6.19)-(6.21) is given by

<u>THEOREM 6.3</u>. Let $v_{i,j}$ be the solution of the boundary value problem given by equations (6.19)-(6.21). For $h < h_3$,

$$\max_{R_h+R_b} |v_{i,j}| \leqq \hat{p}_{i,j} \max_{R_h+R_b} \{|t_{i,j}|, |\bar{t}_{i,j}|\} + \max_{R_S} |\phi_{i,j}|.$$

<u>Proof</u>: This theorem is proved in the same way that Theorem 6.1 is proved. Use is made of Lemma 6.10, and other substitutions are obvious.

COROLLARY 1. Let μ_2 be defined by

$$\mu_2 = [2K_1 + 2(K_1^2 + K_0/2)^{1/2}]/K_0;$$

then there exists an h_4 such that for $h < h_4$, the solution of the problem given by equations (6.19)-(6.21) is bounded as follows:

$$\max_{R_h+R_b} |v_{i,j}| \leq e^{\mu_2 X} \max_{R_h+R_b} \{|t_{i,j}|, |\bar{t}_{i,j}|\} + \max_{R_S} |\phi_{i,j}|.$$

Proof: The proof of this corollary is the same as the proof of Corollary 1 to Theorem 6.1.

CHAPTER VII

ERROR BOUNDS FOR SOLUTIONS OF FINITE DIFFERENCE PROBLEMS

In this chapter, we use the results of Chapter VI to show that the boundary value problems for the error functions associated with problems P_1 and P_2, i.e., problems $\tilde{P}_1$ and $\hat{P}_2$ have solutions which are proportional to h^p, $p \geqq 1$. Let W_1 denote the set of functions defined at the mesh points in $R_h + R_b + R_S$ such that if $w \in W_1$, then

$$(7.1) \qquad \max_{R_h+R_b} |w_{i,j}| \leqq 4h^2 \max\{e^{\mu_1 X} k_4,\ k_\delta\}$$

$$w(x,y) = 0,\ (x,y) \in R_S$$

where

$$(7.2) \qquad \mu_1 = [k_1' + ((k_1')^2 + 4k_0')^{1/2}]/2k_0'$$

and

$$(7.3) \qquad k_1' = \max\ \{k_1, k_3\}.$$

Equation (5.17) which is satisfied by the function $E_{i,j}$ at points of R_h for problem $\tilde{P}_1$ is a nonlinear algebraic equation; we linearize it by replacing $E_{i,j}$ where it occurs in the arguments of the coefficients by a given function $w \in W_1$. Equation (5.17) is rewritten in the form

$$M_h s_{i,j} = g_{i,j},\ (x_i, y_j) \in R_h$$

where

$$M_h s_{i,j} = a_{i,j}(\Delta\nabla)_x s_{i,j} + 2b_{i,j}(\Delta\nabla)_z s_{i,j} + c_{i,j}(\Delta\nabla)_y s_{i,j}$$

$$+ d_{i,j}(\Delta+\nabla)_x s_{i,j} + e_{i,j}(\Delta+\nabla)_y s_{i,j} + f_{i,j} s_{i,j}$$

and where the arguments of $a_{i,j}$ $b_{i,j}$, and $c_{i,j}$ are

$$(x_i, y_j, (u_{i,j} + w_{i,j}), (\Delta+\nabla)_x(u_{i,j}+w_{i,j}), (\Delta+\nabla)_y(u_{i,j}+w_{i,j})),$$

and $E_{i,j}$ has been replaced by $s_{i,j}$ everywhere except in the coefficients. We also denote the difference operator $\tilde{L}_{b1}$ by M_{b1}.

Now, consider the boundary value problem $\bar{P}_1$ given by

(7.4) $$M_h s_{i,j} = g_{i,j} \quad , \quad (x_i, y_j) \in R_h$$

(7.5) $$M_{b1} s_{i,j} = g'_{i,j} \quad , \quad (x_i, y_j) \in R_b$$

(7.6) $$s_{i,j} = 0 \quad , \quad (x_i, y_j) \in R_S.$$

We have

LEMMA 7.1. For $h < h_2$, the solution s of problem $\bar{P}_1$ exists and is in the set W_1.

Proof: Since problem $\bar{P}_1$ is a linear difference equation problem, the results of Chapter VI can be applied to it. For $h < h_2$, we have from Lemma 6.6 that problem $\bar{P}_1$ has a unique solution s, and from Theorem 6.2,

$$\max_{R_h+R_b} |s_{i,j}| \leq 2e^{\mu_1 X} \max_{R_h} |g_{i,j}| + 2 \max_{R_b} |g'_{i,j}|.$$

From this and equations (5.17) and (5.18),

$$\max_{R_h+R_b} |s_{i,j}| \leq 2e^{\mu_1 X} h^2 k_4 + 2h^2 k_8$$

or

$$\max_{R_h+R_b} |s_{i,j}| \leq h^2 4 \max \{e^{\mu_1 X} k_4, k_8\}.$$

Also, $s(x,y) = 0$, $(x,y) \in R_S$. Hence

$$s \in W_1.$$

<u>THEOREM 7.1</u>. Let $h < h_2$, then problem $\tilde{P}_1$ for the error associated with problem P_1 has a solution $E = s^*$ where $s^* \in W_1$. Moreover, problem $\tilde{P}_1$ has no solution that is not in W_1.

<u>Proof</u>: In order to prove this theorem, we consider the problem $\bar{P}_1$ as a transformation T:

$$Tw = s.$$

By Lemma 7.1, for $h < h_2$, the transformation T takes a function w from the set W_1 into a function s which is also in the set W_1. The set W_1 is a closed n-cell[1], and the transformation T is continuous; therefore, we can apply the Brouwer Fixed Point Theorem to the transformation T. The

[1] A closed n-cell is defined as follows (see Lefschetz [1949], p. 30). Let G^n be an n-dimensional Euclidean space. Let the coordinates of a point in G^n be denoted by x_i, $i = 1, 2, 3, \ldots, n$. A closed n-cell is defined as the image of any continuous one-to-one mapping of the set

$$\sum_{i=1}^{n} x_i \leq 1.$$

The Brouwer Fixed Point Theorem states that[2] every continuous transformation of a closed n-cell into itself has a fixed point.

Thus, there exists a function $s^* \in W_1$ such that

$$Ts^* = s^*,$$

and we conclude that problem $\breve{P}_1$ for the error associated with problem P_1 has a solution $E = s^*$ where $s^* \in W_1$.

Now, suppose problem $\breve{P}_1$ has a solution $\bar{s}$ where $\bar{s}$ is not in W_1. This implies that $\bar{s}$ can be used to linearize the difference equation (5.17) for problem $\breve{P}_1$ and that $\bar{s}$ is the solution of the linearized problem. However, Theorem 6.2 applies to the linearized problem and states that, for $h < h_2$, any solution of this problem is bounded as follows:

$$\max_{R_h+R_b} |\bar{s}_{i,j}| \leq 2h^2[e^{\mu_1 X} k_4 + k_8].$$

Since any solution of problem $\breve{P}_1$ also has zero boundary values, we conclude that $\bar{s} \in W_1$ and thus have a contradiction. Therefore, problem $\breve{P}_1$ has no solution that is not in W_1.

Next, we consider problem $\breve{P}_2$. Let W_2 denote the set of functions defined at the mesh points in $R_h + R_b + R_S$ such that if $w \in W_2$, then

$$(7.7) \qquad \max_{R_h+R_b} |w_{i,j}| \leq he^{\mu_2 X} \max \{hk_4, k_6\}$$

$$(7.8) \qquad w(x,y) = 0, \quad (x,y) \in R_S$$

[2] Lefschetz [1949], p. 117.

where

$$(7.9) \qquad \mu_2 = [2k_1' + 2((k_1')^2 + k_0'/2)^{1/2}]/k_0'.$$

Equations (5.20) and (5.21) which are satisfied by $E_{i,j}$ at points of $R_h + R_b$ for problem $\hat{P}_2$ are both nonlinear algebraic equations. We linearize them in the manner given above for problem $\hat{P}_1$ by using a given function $w \in W_2$. Problem $\bar{P}_2$ consists of the linearized problem $\hat{P}_2$ and is given by

$$(7.10) \qquad M_h \; s_{i,j} = g_{i,j} \quad , \quad (x_i, y_j) \in R_h$$

$$(7.11) \qquad M_{b2} \; s_{i,j} = \bar{g}_{i,j} \quad , \quad (x_i, y_j) \in R_b$$

$$(7.12) \qquad s(x,y) = 0 \quad , \quad (x,y) \in R_S$$

where M_h and M_{b2} denote the linearized finite difference operators L_h and L_{b2} respectively.

Then, we have

<u>LEMMA 7.2</u>. For $h < h_4$, the solution s of problem $\bar{P}_2$ exists and is in the set W_2.

<u>Proof</u>: Lemma 6.7 and Corollary 1 to Theorem 6.3 apply to problem $\bar{P}_2$ and state that, for $h < h_4$, problem $\bar{P}_2$ has a unique solution s and

$$\max_{R_h+R_b} |s_{i,j}| \le e^{\mu_2 X} \max_{R_h+R_b} \{|g_{i,j}|, |\bar{g}_{i,j}|\}.$$

From this and equations (5.20) and (5.21),

$$\max_{R_h+R_b} |s_{i,j}| \leq he^{\mu_2 X} \max_{R_h+R_b} \{hk_4, k_6\}.$$

Since also, $s(x,y) = 0$, $(x,y) \in R_S$, $s \in W_2$.

<u>THEOREM 7.2</u>. Let $h < h_4$, then problem P_2 for the error associated with problem P_2 has a solution $E = s^*$ where $s^* \in W_2$. Moreover, problem $\tilde{P}_2$ has no solution which is not in W_2.

<u>Proof</u>: This theorem is proved by the same argument that is used to prove Theorem 7.1 by using Theorem 6.3 and Lemma 7.2 instead of Theorem 6.2 and Lemma 7.1.

From Theorems 7.1 and 7.2, we conclude that the error functions associated with problems P_1 and P_2 are bounded by quantities which are proportional to h^2 and h respectively. In Chapter IX, we show that the error for problem P_2 is actually bounded by a quantity which is proportional to h^2.

CHAPTER VIII

EXISTENCE AND UNIQUENESS

The existence of solutions of the discrete analogues, problems P_1 and P_2, of the given continuous problem, problem P_0, can be deduced from the existence of a solution of the continuous problem and the existence of solutions of problems P_1 and P_2. However, this reasoning depends on the assumption that the continuous problem satisfies all of the conditions given in Chapter II which insures the existence of bounded partial derivatives of fourth order of the solution of the continuous problem. Solutions of problems P_1 and P_2 can be shown to exist with fewer conditions than this. For sufficiently small mesh width, the same conditions which are used to establish the existence of the solution of problem P_0 are sufficient to establish the existence of solutions of problems P_1 and P_2.

We prove

THEOREM 8.1. Let equation (2.3) be uniformly elliptic and let the coefficients A, B, and C be Hölder continuous in their five variables. Let the function G be bounded by a constant K, and let the function ϕ have Hölder continuous first partial derivatives. Then, for $h < \min\{h_2, h_4\}$, the solutions of problems P_1 and P_2 exist.

Proof: First, we note that, by Theorem 2.1, the hypotheses of this theorem are sufficient to guarantee the existence of a solution of problem P_0 which has continuous second-order partial derivatives in R. Thus, the transformation (3.3) exists, and it makes sense to talk about problems P_1 and P_2.

In order to establish the existence of solutions of problems P_1 and P_2, we use a technique similar to that which was used to establish the existence of solutions of problems $\hat{P}_1$ and $\hat{P}_2$.

Let W_3 denote the set of functions defined at the mesh points in $R_h + R_b + R_S$ such that if $w \in W_3$, then

$$\max_{R_h+R_b} |w_{i,j}| \leq Ke^{\mu_1 X} + \max_{R_S} |\phi_{i,j}|$$

$$w(x,y) = \phi(x,y), \ (x,y) \in R_S$$

where

$$\mu_1 = [k_1 + (k_1^2+4k_0')^{1/2}]/2k!$$

Equation (4.1), problem P_1, is linearized by replacing $U_{i,j}$ where it occurs in the arguments of the functions A, B, C, and G by a given function $w \in W_3$. By Lemma 6.6, the linearized problem has a unique solution. This problem is considered as a transformation T_1 which takes a function w from the closed n-cell W_3 into a function U which, by Theorem 6.2, is also in W_3. Since this transformation is also continuous, the Brouwer Fixed Point Theorem can be applied; thus, the transformation T_1 has a fixed point in W_3, which is a solution of problem P_1.

The existence of a solution of problem P_2 is established in the same manner as for problem P_1. The solution of problem P_2 is in the closed n-cell W_4 where a function $w \in W_4$ if

$$\max_{R_h+R_b} |w_{i,j}| \leq Ke^{\mu_2 X} + \max_{R_S} |\phi_{i,j}|$$

$$w(x,y) = \phi(x,y), \ (x,y) \in R_S$$

and where

$$\mu_2 = [2k_1+2(k_1^2+ k_0'/2)^{1/2}]/k_0' .$$

Use is made of Lemma 6.7 and Corollary 1 to Theorem 6.3 to show that solutions of the linearized problem exist and are in W_4.

The solutions of problems P_1 and P_2 can be shown to be unique by a method similar to that used to show that the solution of problem P_0 is unique. We prove next

THEOREM 8.2. Let equation (2.3) be uniformly elliptic, let the functions A, B, C, and G have Hölder continuous first partial derivatives, and assume that condition (2.10) is satisfied. If $\nu(x,y)$ is nonzero for any $(x,y) \in R$, let $h < \min\{(\Delta\nu/k_5)^{1/2}, (\Delta\nu/k_7), h_1\}$. Otherwise, let $h < h_1$. Then the solutions of problems P_1 and P_2 are unique.

Proof: We consider first problem P_1. Suppose problem P_1 has two solutions U and $\bar{U}$. Let V denote the difference $U - \bar{U}$ and let $A'(x_i,y_j,U_{i,j},(\Delta+\nabla)_x U_{i,j},(\Delta+\nabla)_y U_{i,j})$ be denoted by $A'_{i,j}$, $A'(x_i,y_j,\bar{U}_{i,j},(\Delta+\nabla)_x \bar{U}_{i,j},(\Delta+\nabla)_y \bar{U}_{i,j})$ by $\bar{A}_{i,j}$, etc. Then at the mesh points in R_h, we have

$$(8.1) \qquad A'_{i,j}(\Delta\nabla)_x U_{i,j}+ 2B'_{i,j}(\Delta\nabla)_z U_{i,j}+ C'_{i,j}(\Delta\nabla)_y U_{i,j} = G_{i,j}$$

$$(8.2) \qquad \bar{A}_{i,j}(\Delta\nabla)_x \bar{U}_{i,j}+ 2\bar{B}_{i,j}(\Delta\nabla)_z \bar{U}_{i,j}+ \bar{C}_{i,j}(\Delta\nabla)_y \bar{U}_{i,j} = \bar{G}_{i,j}.$$

We subtract equation (8.2) from equation (8.1) to obtain

$$\begin{aligned} & A'_{i,j}(\Delta\nabla)_x V_{i,j} + 2B'_{i,j}(\Delta\nabla)_z V_{i,j} + C'_{i,j}(\Delta\nabla)_y V_{i,j} \\ (8.3) \qquad & + (A'_{i,j}- \bar{A}_{i,j})(\Delta\nabla)_x \bar{U}_{i,j} + 2(B'_{i,j}- \bar{B}_{i,j})(\Delta\nabla)_z \bar{U}_{i,j} \\ & + (C'_{i,j}- \bar{C}_{i,j})(\Delta\nabla)_y \bar{U}_{i,j} = (G_{i,j}- \bar{G}_{i,j}). \end{aligned}$$

The differences $[A'_{i,j} - \bar{A}_{i,j}]$, $[B'_{i,j} - \bar{B}_{i,j}]$, $[C'_{i,j} - \bar{C}_{i,j}]$, and $[G_{i,j} - \bar{G}_{i,j}]$ can be evaluated by means of equation (5.3) with $u_{i,j}$ and $E_{i,j}$ replaced by $U_{i,j}$ and $V_{i,j}$ respectively.

We have, for example,

$$
\begin{aligned}
[A'_{i,j} - \bar{A}_{i,j}] &= A'(x_i, y_j, (U+V)_{i,j}, (\Delta+\nabla)_x (U+V)_{i,j}, (\Delta+\nabla)_y (U+V)_{i,j}) \\
&\quad - A'(x_i, y_j, \bar{U}_{i,j}, (\Delta+\nabla)_x \bar{U}_{i,j}, (\Delta+\nabla)_y \bar{U}_{i,j}) \\
&= V_{i,j} \left[\int_0^1 \bar{A}_r d\theta\right]_{i,j} + (\Delta+\nabla)_x V_{i,j} \left[\int_0^1 \bar{A}_p d\theta\right]_{i,j} \\
&\quad + (\Delta+\nabla)_y V_{i,j} \left[\int_0^1 \bar{A}_q d\theta\right]_{i,j} .
\end{aligned}
\tag{8.4}
$$

By substituting expressions such as (8.4) for the differences $[A'_{i,j} - \bar{A}_{i,j}]$, etc. in equation (8.3) and rearranging, we obtain

$$
\begin{aligned}
&A'_{i,j} (\Delta\nabla)_x V_{i,j} + 2B'_{i,j} (\Delta\nabla)_z V_{i,j} + C'_{i,j} (\Delta\nabla)_y V_{i,j} + D_{i,j} (\Delta+\nabla)_x V_{i,j} \\
&\quad + E_{i,j} (\Delta+\nabla)_y V_{i,j} + F_{i,j} V_{i,j} = 0
\end{aligned}
\tag{8.5}
$$

where

$$
\begin{aligned}
D_{i,j} &= (\Delta\nabla)_x \bar{U}_{i,j} \left[\int_0^1 \bar{A}_p d\theta\right]_{i,j} + 2(\Delta\nabla)_z \bar{U}_{i,j} \left[\int_0^1 \bar{B}_p d\theta\right]_{i,j} \\
&\quad + (\Delta\nabla)_y \bar{U}_{i,j} \left[\int_0^1 \bar{C}_r d\theta\right]_{i,j} - \left[\int_0^1 \bar{G}_p d\theta\right]_{i,j} ,
\end{aligned}
$$

$$E_{i,j} = (\Delta\nabla)_x \bar{U}_{i,j} \left[\int_0^1 \bar{A}_q d\bar{o}\right]_{i,j} + 2(\Delta\nabla)_z \bar{U}_{i,j} \left[\int_0^1 \bar{B}_q d\theta\right]_{i,j}$$

$$+ (\Delta\nabla)_y \bar{U}_{i,j} \left[\int_0^1 \bar{C}_q d\theta\right]_{i,j} - \left[\int_0^1 \bar{G}_q d\theta\right]_{i,j} ,$$

$$F_{i,j} = (\Delta\nabla)_x \bar{U}_{i,j} \left[\int_0^1 \bar{A}_r d\theta\right]_{i,j} + 2(\Delta\nabla)_z \bar{U}_{i,j} \left[\int_0^1 \bar{B}_r d\theta\right]_{i,j}$$

$$+ (\Delta\nabla)_y \bar{U}_{i,j} \left[\bar{C}_r d\theta\right]_{i,j} - \left[\int_0^1 \bar{G}_r d\theta\right]_{i,j} .$$

The coefficients in equation (8.5) are bounded in the same way that the coefficients in the error equation, Chapter V, are bounded. Also, from Chapter V, $F_{i,j}$ is nonpositive for $h < (\Delta\nu/k_5)^{1/2}$.

At the mesh points in R_b, U and $\bar{U}$ satisfy equations of the form

$$(8.6) \qquad L_{b1}U_{i,j} = [\lambda/(\lambda+1)]U_{m,n} + [1/(\lambda+1)]U_{p,q} - U_{i,j}$$

and

$$(8.7) \qquad L_{b1}\bar{U}_{i,j} = [\lambda/(\lambda+1)]\bar{U}_{m,n} + [1/(\lambda+1)]\bar{U}_{p,q} - \bar{U}_{i,j}$$

where $0 < \lambda < 1$, (x_i,y_j) is a mesh point in R_b, (x_m,y_n) is a mesh point in R_h, and (x_p,y_q) is a point in R_S.

We subtract equation (8.7) from equation (8.6), note that $U_{p,q} = \bar{U}_{p,q}$, to obtain

$$(8.8) \qquad [\lambda/(\lambda+1)]V_{m,n} - V_{i,j} = 0.$$

We now formulate a boundary value problem for V as follows:

$$A'_{i,j}(\Delta\nabla)_x V_{i,j} + 2B'_{i,j}(\Delta\nabla)_z V_{i,j} + C'_{i,j}(\Delta\nabla)_y V_{i,j} + D_{i,j}(\Delta+\nabla)_x V_{i,j}$$

$$+ E_{i,j}(\Delta+\nabla)_y V_{i,j} + F_{i,j}V_{i,j} = 0, (x_i,y_j) \in R_h$$

$$[\lambda/(\lambda+1)]V_{m,n} - V_{i,j} = 0 , \quad (x_i,y_j) \in R_b$$

$$V(x,y) = 0 , \quad (x,y) \in R_S$$

By Lemma 6.5 the function V is bounded above by zero for $h < h_1$. Similarly, the function -V is bounded above by zero. Therefore, V = 0, and U = Ū.

The uniqueness of the solution of problem P_2 is established in the same manner as for problem P_1. Equation (8.5) applies at mesh points in R_h for problem P_2, and an equation of the same form as equation (8.5) but using asymmetric difference quotients applies at the mesh points in R_b. The solution of the finite difference problem for V corresponding to problem P_2 is shown to be identically zero by the use of Lemma 6.1.

CHAPTER IX

IMPROVEMENT OF ERROR BOUNDS

In Chapter VII, , solutions of problems $\breve{\hat{P}}_1$ and $\breve{\hat{P}}_2$ for the error in problems P_1 and P_2 respectively are shown to exist and to be bounded by quantities which are proportional to h^p. The exponent p is found to be equal to two for problem $\breve{\hat{P}}_1$ and to be not less than one for problem $\breve{\hat{P}}_2$. We now show that the solution of problem $\breve{\hat{P}}_2$ is actually bounded by a quantity which is proportional to h^2 in all cases. The methods used here are similar to methods presented in Bers [1953] and in Bramble and Hubbard [1963].

The solution of problem $\breve{\hat{P}}_2$ is a fixed point of the transformation consisting of the linear boundary value problem $\bar{P}_2$. Thus, we accomplish our objective by showing that the solution s of problem $\bar{P}_2$ is bounded by a quantity which is proportional to h^2.

We denote the finite difference operators M_h and M_{b2} (see equations (7.10), (7.11)), when applied at a general mesh point in $R_h + R_b$, by M_0. In this notation, problem $\bar{P}_2$ is given by

$$M_0 s_{i,j} = g_{i,j} \quad , \quad (x_i,y_j) \in R_h \tag{9.1}$$

$$M_0 s_{i,j} = \bar{g}_{i,j} \quad , \quad (x_i,y_j) \in R_b \tag{9.2}$$

$$s(x,y) = 0 \quad , \quad (x_i,y_j) \in R_S . \tag{9.3}$$

The operator M_0 is written in the following form[1]

$$M_0 s_{i,j} = \sum_{(m,n)} \sigma_{i,j;m,n} s_{m,n} \tag{9.4}$$

where the subscripts (m,n) take on all values corresponding to points in $R_h + R_b + R_s$. For a given mesh point (x_i, y_j), the coefficients $\sigma_{i,j;m,n}$ are equal respectively to h^{-2} times the coefficients $\mu_{m,n}$ which are defined in Lemma 6.1; otherwise, $\sigma_{i,j;m,n}$ is equal to zero. The coefficients $\sigma_{i,j;m,n}$ satisfy the following conditions for $h < h_5 = k_0'/k_1'$

$$\sigma_{i,j;i,j} < 0$$

$$\sigma_{i,j;m,n} \geq 0, \ (i,j) \neq (m,n) \tag{9.5}$$

$$|\sigma_{i,j;i,j}| \geq \sum_{\substack{(m,n) \\ (m,n)\neq(i,j)}} \sigma_{i,j;m,n}$$

We now define a function $G_{i,j;m,n}$ which is a discrete analogue of a Green's function for problem $\bar{P}_2$. The function $G_{i,j;m,n}$ is the solution of

$$\sum_{(m,n)} \sigma_{i,j;m,n} G_{m,n;p,q} = - \delta(x_i, y_i; x_p, y_q), \ (x_i, y_j) \in R_h + R_b \tag{9.6}$$

[1] This notation is also used in Chapter III in the definition of nonnegative difference operators.

$$(9.7) \qquad G_{i,j;p,q} = \delta(x_i,y_j;x_p,y_q), \quad (x_i,y_j) \in R_S$$

where δ is the Kronecker delta.

Next, we establish some properties of the function $G_{i,j;m,n}$.

LEMMA 9.1. For fixed $h < h_5$, the function $G_{i,j;m,n}$ exists and is unique.

Proof: If it exists, the function $G_{i,j;m,n}$ is unique. For, assume that $G'_{i,j;m,n}$ and $G''_{i,j;m,n}$ are two functions satisfying equations (9.6) and (9.7). Let $\bar{G}_{i,j;m,n} = G'_{i,j;m,n} - G''_{i,j;m,n}$. Then, $\bar{G}_{i,j;m,n}$ satisfies

$$M_0\bar{G}_{i,j;m,n} = 0, \; (x_i,y_j) \in R_h + R_b,$$

$$\bar{G}_{i,j;m,n} = 0, \; (x_i,y_j) \in R_S,$$

and by Lemma 6.1, $\bar{G}_{i,j;m,n} \leqq 0$. Similarly, $-\bar{G}_{i,j;m,n} \leqq 0$; therefore, $\bar{G}_{i,j;m,n} = 0$.

Since $G_{i,j;m,n}$ is the solution of a system of linear algebraic equations with an equal number of equations and unknowns, uniqueness implies existence.

LEMMA 9.2. Let $h < h_5$ and let $s_{i,j}$ be an arbitrary function defined on $R_h + R_b + R_S$. Then at each mesh point $(x_i,y_j) \in R_h + R_b + R_S$, $s_{i,j}$ is given by

$$(9.8) \qquad s_{i,j} = \sum_{(x_m,y_n)\in R_h+R_b} G_{i,j;m,n}[-M_0 s_{m,n}] + \sum_{(x_m,y_n)\in R_S} G_{i,m;m,n} s_{m,n}.$$

Proof: Let $w_{i,j}$ represent the right side of (9.8). Suppose $(x_i,y_j) \in R_S$. Then the first term on the right side of (9.8) is zero and $w_{i,j}$ is simply $s_{i,j}$. Now, let $(x_i,y_j) \in R_h + R_b$; then we consider

$$M_0 w_{i,j} = \sum_{(x_m,y_n)\in R_h+R_b+R_S} \sigma_{i,j;m,n} w_{m,n}$$

$$= \sum_{(x_m,y_n)\in R_h+R_b+R_S} \sigma_{i,j;m,n} \left\{ \sum_{(x_p,y_q)\ R_h+R_b} G_{m,n;p,q} \ [-M_0 s_{p,q}] + \sum_{(x_p,y_q)\in R_S} G_{m,n;p,q} s_{p,q} \right\}$$

$$= -M_0 s_{i,j} \sum_{(x_m,y_n)\in R_h+R_b+R_S} \sigma_{i,j;m,n} G_{m,n;i,j}$$

$$- \sum_{\substack{(x_p,y_q)\in R_h+R_b \\ (x_p,y_q)\neq(x_i,y_j)}} M_0 s_{p,q} \sum_{(x_m,y_n)\in R_h+R_b+R_S} \sigma_{i,j;m,n} \sigma_{i,j;m,n} G_{m,n;p,q}$$

$$+ \sum_{(x_p,y_q)\in R_S} s_{p,q} \sum_{(x_m,y_n)\in R_h+R_b+R_S} \sigma_{i,j;m,n} G_{m,n;p,q}$$

$$= M_0 s_{i,j}$$

Thus,

$$M_0 [s_{i,j} - w_{i,j}] = 0 \quad , \quad (x_i,y_j) \in R_h + R_b$$

$$[s_{i,j} - w_{i,j}] = 0 \quad , \quad (x_i,y_j) \in R_S$$

and by Lemma 6.1,

$$s_{i,j} = w_{i,j} \quad , \quad (x_i,y_j) \in R_h + R_b + R_S .$$

LEMMA 9.3. For $h < h_5$, the function $G_{i,j;m,n}$ is nonnegative.

Proof: Assume first that $(x_m,y_n) \in R_h + R_b$ and consider the function $-G_{i,j;m,n}$ which satisfies

$$M_0[-G_{i,j;m,n}] = \delta(x_i,y_j;x_m,y_n) \geq 0, \quad (x_i,y_j) \in R_h + R_b,$$

$$-G_{i,j;m,n} = 0 \qquad , \quad (x_i,y_j) \in R_S .$$

Now, let $(x_m,y_n) \in R_S$; then

$$M_0[-G_{i,j;m,n}] = 0 \qquad , \quad (x_i,y_j) \in R_h + R_b,$$

$$-G_{i,j;m,n} = -\delta(x_i,y_j;x_m,y_n), \quad (x_i,y_j) \in R_S .$$

By Lemma 6.1, $-G_{i,j;m,n} \leq 0$ for both cases above.

LEMMA 9.4. For $h < \min\{h_3,h_4\}$, the sum

$$\sum_{(x_m,y_n)\in R_h+R_b} G_{i,j;m,n} \leq e^{\mu_2 X}$$

where

$$\mu_2 = [2k_1' + 2((k_1')^2 + k_0'/2)^{1/2}]/k_0' .$$

Proof: Recall the function $\breve{p}_{i,j}$ which is defined in Chapter VI. By Lemma 9.2, $\breve{p}_{i,j}$ is given by

$$\breve{p}_{i,j} = \sum_{(x_m,y_n)\in R_h+R_b} G_{i,j;m,n}[-M_0\breve{p}_{m,n}] + \sum_{(x_m,y_n)\in R_S} G_{i,j;m,n}\breve{p}_{m,n}.$$

By Lemma 6.10, $M_0\breve{p}_{i,j} \leq -1$, and by Corollary 1 to Theorem 6.3, $0 \leq \breve{p}_{i,j} \leq e^{\mu_2 X}$ for $h < h_4$. From this and Lemma 9.3, we have

$$e^{\mu_2 X} \geq \sum_{(x_m,y_n)\in R_h+R_b} G_{i,j;m,n}.$$

We now prove the principal result of this chapter.

THEOREM 9.1. Let $h < \min\{h_i\}$, $i = 3, 4, 5$. Then the solution of problem $\bar{P}_2$ is bounded by a quantity which is proportional to the square of the mesh width.

Proof: From Lemma 9.2, the solution $s_{i,j}$ of problem $\bar{P}_2$ is given at the mesh points $(x_i,y_j) \in R_h + R_b$ by

$$(9.9)\qquad s_{i,j} = \sum_{(x_m,y_n)\in R_h} G_{i,j;m,n}[-M_0 s_{m,n}] + \sum_{(x_m,y_n)\in R_b} G_{i,j;m,n}[-M_0 s_{m,n}].$$

For points $(x_m,y_n) \in R_h$, we have from equations (9.1) and (5.20),

$$\begin{aligned}|M_0 s_{i,j}| &= |g_{i,j}| \\ &\leq h^2 k_4 .\end{aligned}$$

Thus, from Lemma 9.4,

$$\left| \sum_{(x_m,y_n)\in R_h} G_{i,j;m,n}[-M_0 s_{i,j}] \right|$$

$$\leq \max_{(x_m,y_n)\in R_h} \left|M_0 s_{i,j}\right| \sum_{(x_m,y_n)\in R_h} G_{i,j;m,n}$$

$$\leq h^2 k_4 e^{\mu_2 X} .$$

Next, we consider the term

$$(9.11) \qquad \sum_{(x_m,y_n)\in R_b} G_{i,j;m,n}[-M_0 s_{m,n}].$$

Let $\bar{R}_b$ be any subset of the mesh points in R_b. We show first that for any mesh point $(x_i,y_j) \in R_h + R_b$

$$(9.12) \qquad \frac{1}{\min\limits_{(x_m,y_n)\in\bar{R}_b} \left[- \sum\limits_{(x_p,y_q)\in R_h+R_b} \sigma_{m,n;p,q} \right]} \geq \sum_{(x_m,y_n)\in\bar{R}_b} G_{i,j;m,n}.$$

Let the function $s_{i,j}$ be defined by

$$(9.13) \qquad s_{i,j} = \begin{cases} 1, & (x_i,y_j) \in R_h + R_b \\ 0, & (x_i,y_j) \in R_S \end{cases} .$$

By Lemma 9.2, for $(x_i,y_j) \in R_h + R_b$,

$$(9.14) \qquad 1 = \sum_{(x_m,y_n)\in R_h+R_b} G_{i,j;m,n}[-M_0 s_{m,n}].$$

From conditions (9.5),

$$(9.15) \qquad -M_0 s_{m,n} \geqq 0, \ (x_m,y_n) \in R_h + R_b.$$

From Lemma 9.3, equation (9.14), and inequality (9.15), we have

$$(9.16) \qquad 1 \geqq \sum_{(x_m,y_n)\in\bar{R}_b} G_{i,j;m,n}[-M_0 s_{m,n}] \ .$$

Inequality (9.12) follows from Lemma 9.3 and inequalities (9.15) and (9.16).

Now let the subset $\bar{R}_b$ consist of all mesh points $(x_m,y_n) \in R_b$ such that at least one rectangular neighbor of (x_m,y_n) is not in $R + S$. Then, for $h < h_5$,

$$\min_{(x_m,y_n)\in\bar{R}_b} \left[-\sum_{(x_p,y_q)\in R_h+R_b} \sigma_{m,n;p,q} \right] \geqq k_0'/2h^2$$

so that

$$\frac{1}{\min\limits_{(x_m,h_n)\in\bar{R}_b} \left[-\sum\limits_{(x_p,y_q)\in R_h+R_b} \sigma_{m,n;p,q} \right]} \leqq \frac{2h^2}{k_0'} \ .$$

This statement follows from the difinitions of $\sigma_{m,n;p,q}$ and conditions (9.5).

Therefore, for all mesh points in $\bar{R}_b$,

$$\left| \sum_{(x_m,y_n)\in\bar{R}_b} G_{i,j;m,n}[-M_0 s_{m,n}] \right|$$

(9.17) $$\leq (2h^2/k_0') \max_{(x_i,y_j)\in\bar{R}_b} |\bar{g}_{i,j}|$$

$$\leq (2h^3/k_0')k_6 \; .$$

Now, let $\bar{R}_b$ consist of all mesh points in R_b which do not have rectangular neighbors which are not in $R + S$. For each such point (x_m,y_n), one diagonal neighbor is not in $R + S$, and from the definition of $\bar{g}$ at such a point, we have

$$\left|-M_0 s_{m,n}\right| = \left|-\bar{g}_{m,n}\right|$$

$$\leq (4/3)\eta h\bar{Q}\; B'_{m,n} + O(h^2).$$

Also, for the points under consideration,

$$-\sum_{(x_p,y_q)\in R_h+R_b} \sigma_{m,n;p,q} \geq 2b_{m,n}/h^2(\alpha^2+\beta^2) = 2B'_{m,n}/h^2(\alpha^2+\beta^2)$$

Thus, from (9.12)

$$h^2(\alpha^2+\beta^2)/2B'_{m,n} \geq \sum_{(x_m,y_n)\in\bar{R}_b} G_{i,j;m,n} \; .$$

Therefore,

$$\left|\sum_{(x_m,y_n)\in\bar{R}_b} G_{i,j;m,n}[-M_0 s_{m,n}]\right|$$

$$\leq (4/3)\eta\bar{Q}\ h \sum_{(x_m,y_n)\in\bar{R}_h} G_{i,j;m,n}B'_{m,n} + O(h^2)$$

$$\leq (4/3)\eta\bar{Q}\ h^3 \sum_{(x_m,y_n)\in\bar{R}_b} (\alpha^2+\beta^2) + O(h^2)$$

$$\leq (4/3)\eta^3\bar{Q}\ h^3 \left|\begin{matrix}\text{number of mesh}\\ \text{points in } \bar{R}_b\end{matrix}\right| + O(h^2)$$

$$\leq (4/3)\eta^3\bar{Q}\ h^3 \left|\begin{matrix}\text{number of mesh}\\ \text{points in } R_b\end{matrix}\right| + O(h^2).$$

Since the number of mesh points in R_b does not increase faster than h^{-1}, we have

$$(9.18) \qquad \sum_{(x_m,y_n)\in\bar{R}_b} G_{i,j;m,n}[-M_0 s_{m,n}] = O(h^2).$$

From inequalities (9.10), (9.17), and (9.18), we have the desired result, that the solution of problem $\bar{P}_2$ is $O(h^2)$ for $h < \min\{h_i\}$, $i = 3, 4, 5$.

CHAPTER X

ADDITIONAL FINITE DIFFERENCE OPERATORS

The finite difference approximations given in Chapter III can be used to approximate any uniformly elliptic partial differential equation; however, they might not be convenient for use for some problems since the difference quotient used to approximate $\partial^2 u/\partial z^2$ might involve values of the solution at mesh points which are far away from the mesh point of application. An approximation with accuracy and generality equal to the approximations given previously but which involves values of the solution only at nearby mesh points is given in this chapter for use with such problems.

Consider the mesh point configuration given in Figure 10.1 and assume that $\tau(x_i,y_j)$ has been determined to be between $\pi/4$ and $\pi/2$ as indicated. Let δh be the distance between the mesh line $x = x_i$ and the intersection between the mesh line $y = y_{j+1}$ and the line $z_{i,j}$.

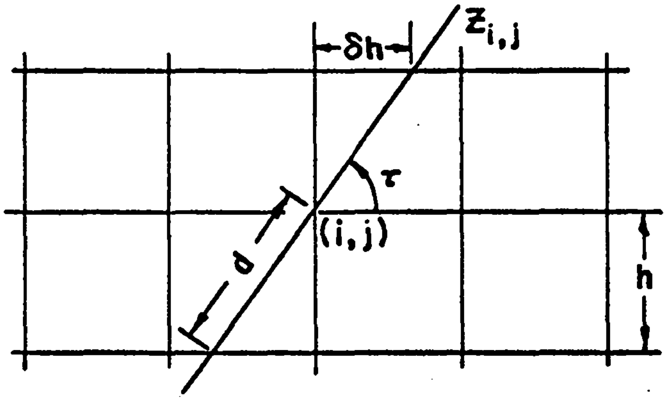


FIGURE 10.1

We consider the following approximation to $\partial^2u/\partial z^2$:

$$(10.1) \qquad (\partial^2u/\partial z^2)_{i,j} = \left[u_{i+\delta,j+1} - 2u_{i,j} + u_{i-\delta,j-1}\right]/d^2$$

$$+ d^2\left[(\partial^4u/\partial z^4)_{i\pm\phi,j\pm\theta}\right]/12, \quad 0 \leq \phi,\ \theta \leq 1,$$

where $d = h(1+\delta^2)^{1/2}$.

If $u_{i+\delta,j+1}$ and $u_{i-\delta,j-1}$ can be approximated to $O(h^p)$ in terms of values of u at mesh points near the point (x_i,y_j), say by expressions of the form

$$(10.2) \qquad u_{i+\delta,j+1} = \sum \ell_{i,j}u_{i,j} + O(h^p),$$

then these expressions can be substituted into equation (3.15) to obtain a difference quotient with a truncation error which is $O(h^{p-2})$.

In order to approximate $\partial^2u/\partial z^2$ to $O(h^2)$, $u_{i+\delta,j+1}$ and $u_{i-\delta,j-1}$ must be approximated to $O(h^4)$; thus, values of u at four neighboring mesh points must be used in the approximation (10.2). A possible approximation is

$$(10.3) \qquad u_{i+\delta,j+1} = -\left[\delta(1-\delta)(2-\delta)/6\right]u_{i-1,j+1} + \left[(1+\delta)(1-\delta)(2-\delta)/2\right]u_{i,j+1}$$

$$+ \left[\delta(1+\delta)2-\delta)/2\right]u_{i+1,j+1} - \left[\delta(1+\delta)1-\delta)/6\right]u_{i+2,j+1} + O(h^4).$$

A similar expression gives $u_{i-\delta,j-1}$ to $O(h^4)$ in terms of $u_{i-2,j-1}$, $u_{i-1,j-1}$, $u_{i,j-1}$, and $u_{i+1,j-1}$.

Obvious modifications of the above procedure yield approximations to $\partial^2u/\partial z^2$ which have truncation errors which are $O(h^2)$ and which depend on values of u at nearby mesh points for any value of τ, $0 < \tau < \pi$. If three-point interpolation formulas are used in place of (10.3), the truncation error of the resulting approximation to $\partial^2 u/\partial z^2$ is $O(h)$.

For mesh points near the boundary, special situations might arise which require special treatment. The variety of such special situations is significantly diminished if we consider only convex regions R. We, therefore, make this assumption for the remainder of this discussion.

If one or more mesh points in $N(x_i,y_j)$ are not in $R + S$, the linear interpolation procedure due to Collatz can be used to determine $U(x_i,y_j)$. By this procedure, the operator L_{b1} (see Chapter III) is obtained. Approximations which lead to operators similar to the operator L_{b2} can also be formulated. Suppose for a mesh point (x_i,y_j) that the point $(x_{i+\delta},y_{j+1})$ is not in $R + S$ (see Figure 10.2). Then $\partial^2 u/\partial z^2$ can be approximated by

$$(10.4) \quad (u_{zz})_{i,j} = 2\left[u_{p,q}/\lambda(\lambda+1) - u_{i,j}/\lambda + u_{i-\delta,j-1}/(\lambda+1)\right]/d^2 + O(h).$$

If the point $(x_{i+\delta},y_{j+1})$ is in $R + S$ but one or more of the points (x_p,y_q) which would normally be used in the interpolation formula for $u_{i+\delta,j+1}$ is not in $R + S$ and if $z_{i,j}$ intersects the boundary S at a point which is within a few mesh widths distance from the point (x_i,y_j), then the value of u equal to the boundary value at the point of intersection can be used in equation (10.4). Alternatively, for sufficiently small values of the mesh width, $u_{i+\delta,j+1}$ can always be approximated to the desired accuracy by interpolating between values of u which might be asymetrically located with respect to the point $(x_{i+\delta},y_{j+1})$. The above remarks regarding the point $(x_{i+\delta};y_{j+1})$ apply also to the point $(x_{i-\delta},y_{j-1})$.

One of the methods discussed above can be used at points near the boundary to formulate an approximation to $\partial^2 u/\partial z^2$ which has a truncation error which is not greater than $O(h)$ and such that the coefficients of the boundary values involved in the approximation are always positive.

We denote the finite difference approximations to $\partial^2u/\partial z^2$ which are discussed above at regular mesh points by $(\overline{\Delta\nabla})_z u_{i,j}$. Let $v_{i,j}$ be an arbitrary function defined on $R_h + R_b + R_S$. We define the finite difference operator L_{h1} by

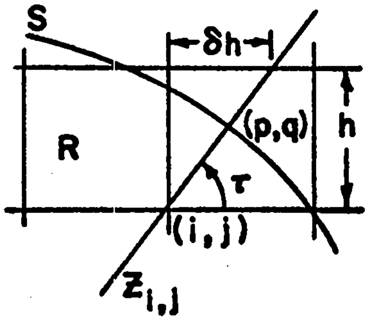


FIGURE 10.2

$$\begin{aligned} L_{h1}v_{i,j} &= a_{i,j}(\Delta\nabla)_x v_{i,j} + 2b_{i,j}(\overline{\Delta\nabla})_z v_{i,j} + c_{i,j}(\Delta\nabla)_y v_{i,j} \\ &\quad + d_{i,j}(\Delta+\nabla)_x v_{i,j} + e_{i,j}(\Delta+\nabla)_y v_{i,j} + f_{i,j}v_{i,j} \\ &= g_{i,j}, \quad (x_i,y_j) \in R_h, \end{aligned} \tag{10.5}$$

where the coefficients satisfy conditions (6.1). We replace the approximation to $\partial^2u/\partial z^2$ in the operator L_{b2} by one of the approximations discussed above for use at points near the boundary to form a difference operator which we denote by L_{b3}.

We formulate a boundary value problem as follows:

$$L_{h1}v_{i,j} = g_{i,j} \quad , \quad (x_i,y_j) \in R_h \tag{10.6}$$

$$L_{b3}v_{i,j} = g_{i,j} \quad , \quad (x_i,y_j) \in R_b \tag{10.7}$$

$$v(x,y) = \emptyset(x,y), \quad (x,y) \in R_S. \tag{10.8}$$

The finite difference operators L_{h1} and L_{b3} can be written in the following form.

$$\begin{aligned} L_{h1}v_{i,j} &= \sum_{(x_m,y_n)\in R_h+R_b+R_S} \sigma_{i,j;m,n}v_{m,n}, \quad (x_i,y_j) \in R_h \\ L_{b3}v_{i,j} &= \sum_{(x_m,y_n)\in R_h+R_b+R_S} \sigma_{i,j;m,n}v_{m,n}, \quad (x_i,y_j) \in R_b. \end{aligned} \tag{10.9}$$

The coefficients in (10.9) satisfy the following conditions:

$$\left.\begin{array}{ll} (10.10) & \sigma_{i,j;i,j} < 0 \\ (10.11) & \displaystyle\sum_{(m,n)\neq i,j} \sigma_{i,j;m,n} \leq |\sigma_{i,j;i,j}| \end{array}\right\} \begin{array}{l} (x_i,y_j) \in R_h+R_b \\ (x_m,y_n) \in R_h+R_b+R_S \end{array}$$

$$\sigma_{i,j;m,n} \geq 0, \ (x_i,y_j) \in R_h+R_b, \ (x_m,y_n) \in R_S. \tag{10.12}$$

We note that condition (10.11) is not the same as the condition of diagonal dominance given in Chapter II since not all of the $\sigma_{i,j;m,n}$ are nonnegative. However, the coefficients which are negative are only slightly negative[1],

[1] For example, the minimum value of the negative coefficients in equation (10.3) is $-\sqrt{3}/27 \approx -.064$.

and the finite difference operator seems to have many of the same characteristics as diagonally dominant operators.

In order to establish additional properties of the operators L_{h1} and L_{b3}, we assume that, corresponding to a given value of the mesh width $h < h_1$, there are N mesh points in $R_h + R_b$ and that there are $(M-N)$ points in R_S. We write the set of simultaneous algebraic equations comprising the problem given by equations (10.6), (10.7), and (10.8) in matrix form as follows:

$$\vec{A}\,\vec{V} = \vec{G} + \vec{B}\,\vec{\emptyset} \tag{10.13}$$

where the matrix $\vec{A}$ is an $N \times N$ matrix whose elements are the coefficients $\sigma_{i,j;m,n}$ where (x_i,y_j), $(x_m,y_n) \in R_h + R_b$, $\vec{V}$ is an N component vector with components $v_{i,j}$ which comprise the solution of the given problem, $\vec{g}$ is an N component vector with components $g_{i,j}$, $\vec{B}$ is an $N \times (M-N)$ matrix whose elements are the coefficients $\sigma_{i,j;m,n}$ where $(x_i,y_j) \in R_h + R_b$ and $(x_m,y_n) \in R_S$, and $\vec{\emptyset}$ is an M component vector with components equal to the boundary values at points in R_S.

Matrices such as the matrix $\vec{A}$ which occur in the formulation of finite difference approximations to elliptic partial differential equations are frequently of monotone type. A matrix $\vec{M}$ is said to be monotone[2] if $\vec{M}\,\vec{x} \geq 0$ implies that $\vec{x} \geq 0$ for any real vector $\vec{x}$. A necessary and sufficient condition for a matrix $\vec{M}$ to be monotonic is that all elements of the inverse matrix $\vec{M}^{-1}$ be nonnegative. It has been shown by direct computation that the negative of the matrix $\vec{A}$ in (10.13) is monotonic for a

[2] Collatz [1960], p. 43.

number of cases. This appears to be true for cases resulting from the use of either three-or four-point interpolation to determine values of the solution function at points between mesh points for use in approximating $\partial^2 u/\partial z^2$. We assume for the remainder of this discussion that $-\vec{A}$ is always monotonic for $h < h_1$; thus,

$$\vec{A}^{-1} \leq 0. \tag{10.14}$$

Next, we prove

LEMMA 10.1. For $h < h_1$, the matrix $\vec{A}$ in equation (10.13) is nonsingular.

Proof: The diagonal elements of the matrix $\vec{A}$ are the coefficients $\sigma_{i,j;i,j}$. The nondiagonal elements in any row of $\vec{A}$ are the coefficients $\sigma_{i,j;m,n}$ in equation (10.11). For $h < h_1$, the magnitude of the diagonal element in each row of $\vec{A}$ exceeds each of the nondiagonal elements. Thus, the rows of $\vec{A}$ are linearly independent, and $\vec{A}$ is nonsingular.

LEMMA 10.2. If $\vec{g} = \vec{\phi} = 0$, then equation (10.13) has only the trivial solution $\vec{V} = 0$.

Proof: The proof follows immediately from Lemma 10.1.

As in Chapter IX, we formulate a discrete analogue of a Green's function for the problem given by equations (10.6), (10.7), and (10.8). The Green's function $G_{i,j;m,n}$ is defined for each mesh point $(x_p,y_q) \in R_h+R_b+R_S$ by

$$\sum_{(x_m,y_n)\in R_h+R_b+R_S} \sigma_{m,n;p,q} = -\delta(x_i,y_j;x_p,y_q), \quad (x_i,y_j) \in R_h+ R_b \tag{10.15}$$

$$G_{i,j;p,q} = \delta(x_i,y_j;x_py_q), \quad (x_i,y_j) \in R_S \tag{10.16}$$

LEMMA 10.3. For $h < h_1$, the Green's function exists and is unique.

Proof: The existence and uniqueness of the solution of equations (10.15) and (10.16) follows from Lemma 10.1.

LEMMA 10.4. Let $v_{i,j}$ be an arbitrary function defined on $R_h + R_b + R_S$. Then for any mesh point $(x_i,y_j) \in R_h + R_b + R_S$, for $h < h_1$,

$$\begin{aligned} v_{i,j} = & \sum_{(x_m,y_n)\in R_h} G_{i,j;m,n}[-L_{h1}v_{m,n}] \\ & + \sum_{(x_m,y_n)\in R_b} G_{i,j;m,n}[-L_{b3}v_{m,n}] + \sum_{(x_m,y_n)\in R_S} G_{i,n;m,n}v_{m,n}. \end{aligned} \tag{10.17}$$

Proof: As in the proof of Lemma 9.2, we let $w_{i,j}$ represent the right side of (10.17). If (x_i,y_j) is a point in R_S, $v_{i,j} = w_{i,j}$. If (x_i,y_j) is a point in $R_h(R_b)$, $L_{h1}v_{i,j} = L_{h1}w_{i,j}(L_{b3}v_{i,j} = L_{b3}w_{i,j})$. Thus,

$$L_{h1}(v_{i,j}-w_{i,j}) = 0 \quad , \quad (x_i,y_j) \in R_h$$

$$L_{b3}(v_{i,j}-w_{i,j}) = 0 \quad , \quad (x_i,y_j) \in R_b$$

$$(v_{i,j}-w_{i,j}) = 0 \quad , \quad (x_i,y_j) \in R_S$$

and from Lemma 10.2, $v_{i,j} = w_{i,j}$, $(x_i,y_j) \in R_h + R_b + R_S$.

LEMMA 10.5. For $h < h_1$, the Green's function $G_{i,j;m,n}$ is nonnegative.

Proof: We extend the definition of $\sigma_{i,j;m,n}$ to include points in R_S as follows:

$$\sigma_{i,j;m,n} = \delta(x_i,y_j;x_m,y_n), \quad (x_i,y_j) \in R_S, \quad (x_m,y_n) \in R_h + R_b. \tag{10.18}$$

For convenience, we assume that the mesh points in $R_h + R_b$ are numbered from one to N, and we denote the coefficients $\sigma_{i,j;m,n}$ and the function $G_{i,j;m,n}$ by $\sigma_{i,j}$ and $G_{i,j}$, $1 \le i,m \le M$, respectively. Here the subscripts i and j take the place of (i,j) and (m,n). In this notation, the matrix $\vec{A}$ is given by

$$\vec{A} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2N} \\ \vdots & & & \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma_{NN} \end{bmatrix}$$

and the matrix $\vec{B}$ is given by

$$-\vec{B} = \begin{bmatrix} \sigma_{1,N+1} & \sigma_{1,N+2} & \cdots & \sigma_{1,M} \\ \sigma_{2,N+1} & & & \\ \vdots & & & \\ \sigma_{N,N+1} & & \cdots & \sigma_{N,M} \end{bmatrix}.$$

From (10.15), (10.16), and (10.18) the functions $G_{i,j}$ satisfy

$$(10.19)\qquad \begin{bmatrix} \vec{A} & -\vec{B} \\ 0 & \vec{I}_S \end{bmatrix} \begin{bmatrix} \vec{G}_{11} & \vec{G}_{12} \\ \vec{G}_{21} & \vec{I}_S \end{bmatrix} = \begin{bmatrix} -\vec{I}_{h+b} & 0 \\ 0 & \vec{I}_S \end{bmatrix}$$

where

$\vec{I}_S$ is the (M-N) x (M-N) identity matrix,

$\vec{G}_{11}$ is an N x N matrix,

$\vec{G}_{12}$ is an N x (M-N) matrix,

$\vec{G}_{21}$ is an (M-N) x N matrix,

$\vec{G}_{22}$ is an (M-N) x (M-N) matrix,

and $\vec{I}_{h+b}$ is the N x N identity matrix.

From (10.19), we have

$$(10.20)\qquad \vec{A}\,\vec{G}_{11} - \vec{B}\,\vec{G}_{21} = -\vec{I}_{h+b}$$

$$(10.21)\qquad \vec{A}\,\vec{G}_{12} - \vec{B} = 0$$

$$(10.22)\qquad \vec{G}_{21} = 0$$

From (10.14), (10.20), and (10.22),

$$\vec{G}_{11} = -I_{h+b}\vec{A}^{-1} \geq 0,$$

and from (10.12), (10.13), (10.14), and (10.21),

$$\vec{G}_{12} = \vec{A}^{-1}\vec{B} \geq 0.$$

Thus, $G_{i,j} = G_{i,j;m,n}$ is nonnegative.

Now, recall the function $\tilde{p}_{i,j}$ given by equation (6.26). We have

LEMMA 10.6. For $h < h_3$, the second difference quotients with respect to z in the operators L_{h1} and L_{b3} when applied to the function $\tilde{p}_{i,j}$ are nonpositive.

Proof: Let (x_i, y_j) be a regular mesh point. Then, two essentially different cases might arise. If $0 < \tan \tau < 1$,

$$\begin{aligned}(\overline{\Delta\nabla})_z \tilde{p}_{i,j} &= \left[-\sigma_2^{i+1} - 2\sigma_2^{i} + \sigma_2^{i-1}\right]/d^2 \\ &= -\sigma_2^{i-1}(\sigma_2^2 - 2\sigma_2 + 1)/d^2 \\ &\leqq 0\end{aligned}$$

for $\sigma_2 \geqq 1$. Now, suppose $\tan \tau > 1$ (see Figure 10.1). If we had exact values for $\tilde{p}$ at $(x_{i+\delta}, y_{j+1})$ and at $(x_{i-\delta}, y_{j-1})$, then the difference quotient with respect to z would be given by

$$(\overline{\Delta\nabla})_z \tilde{p}_{i,j} = -(\sigma_2^{i+\delta} - 2\sigma_2^{i} + \sigma_2^{i-\delta})/d^2, \ 0 < \delta < 1, \ \sigma_2 \geqq 1$$

$$= -\sigma_2^{i-\delta}(\sigma_2^{2\delta} - 2\sigma_2^{\delta} + 1)/d^2, \tag{10.22}$$

$$\leqq 0$$

for $0 < \delta < 1$, $\sigma \geqq 1$.

If we substitute interpolated values for $\sigma_2^{i+\delta}$ and $\sigma_2^{i-\delta}$ in (10.22) and if the interpolated values are sufficiently accurate, we would expect the inequality above to be satisfied. That this is true for three-point interpolation is readily verified as follows. We use the mesh point configuration given in Figure 10.1.as an example to obtain

$$(\overline{\Delta\nabla})_z p_{i,j} = -\sigma_2^i \left[\frac{\delta^2-\delta}{2}\,\sigma_2^{-1} + \frac{\delta+\delta^2}{2}\,\sigma_2 - 2 + \frac{\delta^2-\delta}{2}\,\sigma_2 + (1-\delta^2) \right.$$

$$\left. + \frac{\delta+\delta^2}{2}\,\sigma_2^{-1} \right] / d^2$$

$$= -\delta^2\sigma_2^i((1/\sigma_2) - 2 + \sigma_2)/d^2$$

$$\leq 0$$

for $0 < \delta < 1,\ \sigma_2 \geq 1.$

The use of four-point interpolation formulas for values of $\tilde{p}$ at $(x_{i+\delta}, y_{j+1})$ and $(x_{i-\delta}, y_{j-1})$ results in

$$(\overline{\Delta\nabla})_z \tilde{p}_{i,j} = -\sigma_2^i \left[\delta(\delta^2-1)\sigma_2^{-2}/6 + \delta(2-\delta)(2\delta+1)\sigma_2^{-1}/3 + (1-\delta^2)(2-\delta) \right.$$

$$\left. + \delta(2-\delta)(2\delta+1)\sigma_2/3 + \delta(\delta^2-1)\sigma_2^2/6 - 2 \right] / d^2 .$$

Since four-point interpolation is more accurate than three-point interpolation and since the inequality (10.22) is valid for three-point interpolation, the expression above should satisfy (10.22) also.

If $\tan \tau$ is negative, a similar analysis applies at mesh points in R_h.

If (x_i, y_j) is a mesh point in R_b, an extension of the proof of Lemma 6.9 applies. The detailing of this extension to cover all possible cases would be extremely long and is omitted.

Since all difference quotients in L_{h1} and L_{b3} except for the difference quotients with respect to z are the same as the difference quotients in the operators L_h and L_{b2} respectively, Lemma 10.6 can be used to prove Lemma 6.10 for the difference operators L_{h1} and L_{b3}, and we have for $h < h_3$,

$$(10.23) \qquad \begin{aligned} L_{h1}\tilde{p}_{i,j} &\leq -1, \quad (x_i,y_j) \in R_h \\ L_{b3}\tilde{p}_{i,j} &\leq -1, \quad (x_i,y_j) \in R_b. \end{aligned}$$

Next, we prove

LEMMA 10.7. For $h < \min\{h_1,h_3,h_4\}$, and for any mesh point $(x_i,y_j) \in R_h + R_b + R_S$,

$$e^{\mu_2 X} \geq \sum_{(x_m,y_n)\in R_h+R_b} G_{i,j;m,n}$$

where μ_2 is defined by Corollary 1 to Theorem 6.3.

<u>Proof:</u> From Lemma 10.4, we have for $(x_i,y_j) \in R_h + R_b + R_S$

$$\tilde{p}_{i,j} = \sum_{(x_m,y_n)\in R_h} G_{i,j;m,n}[-L_{h1}\tilde{p}_{m,n}]$$

$$+ \sum_{(x_m,y_n)\in R_b} G_{i,j;m,n}[-L_{b3}\tilde{p}_{m,n}]$$

$$+ \sum_{(x_m,y_n)\in R_S} G_{i,j;m,n}\tilde{p}_{m,n}.$$

Since $\tilde{p}_{i,j}$ is nonnegative, we have from Lemma 10.5 and inequalities (10.23),

$$\tilde{p}_{i,j} \geq \sum_{(x_m,y_n)\in R_h+R_b} G_{i,j;m,n},$$

and by Corollary 1 to Theorem 6.3,

$$e^{\mu_2 X} \geq \sum_{(x_m,y_n)\in R_h+R_b} G_{i,j;m,n}$$

for $(x_i,y_j) \in R_h+ R_b+ R_S$, $h < \min (h_1,h_2,h_3)$.

We consider next the following boundary value problem

$$\begin{aligned} L_{h1}v_{i,j} &= g_{i,j}\ , && (x_i,y_j) \in R_h \\ L_{b3}v_{i,j} &= g'_{i,j}\ , && (x_i,y_j) \in R_b \qquad (10.24) \\ v_{i,j} &= 0\ , && (x_i,y_j) \in R_S \end{aligned}$$

where $g_{i,j}$ and $g'_{i,j}$ are given functions. We prove

<u>THEOREM 10.1</u>. The solution $v_{i,j}$ of the problem given by equations (10.24) is bounded as follows:

$$|v_{i,j}| \leq e^{\mu_2 X} \{\max_{R_h} |g_{i,j}| + \max_{R_b} |g'_{i,j}|\}.$$

<u>Proof</u>: From Lemma 10.4, we have

$$v_{i,j} = - \sum_{(x_m,y_n)\in R_h} G_{i,j;m,n}g_{m,n} - \sum_{(x_m,y_n)\in R_S} G_{i,j;m,n}g'_{m,n},\ (x_i,y_j) \in R_h+ R_b.$$

Thus, by Lemma 10.5

$$|v_{i,j}| \le \max_{R_h} |g_{i,j}| \sum_{(x_m,y_n)\in R_h} G_{i,j;m,n} + \max_{R_b} |g'_{i,j}| \sum_{(x_m,y_n)\in R_b} G_{i,j;m,n},$$

and by Lemma 10.7

$$|v_{i,j}| \le e^{\mu_2 X} \{\max_{R_h} |g_{i,j}| + \max_{R_b} |g'_{i,j}|\}, \ (x_i,y_j) \in R_h + R_b.$$

The methods used in Chapters V, VII, and IX can be used together with Theorem 10.1 to show that the solution of a discrete analogue of problem P_0, which utilizes the difference quotients given above, converges to the solution of problem P_0 as the mesh width is decreased.

A boundary value problem is obtained, by the methods of Chapter V, for the error in the discrete analogue of the given problem. The boundary value problem for the error is of the form of the problem given by equations (10.24). The functions g and g' will be $O(h^p)$ where p depends on the exact form of the difference quotients used.

The Brouwer Fixed Point Theorem is used as in Chapter VII to show that the error is bounded by a quantity which is proportional to (h^p). If g is $O(h^2)$ and g' is $O(h)$, then the methods of Chapter IX are applicable and can be used to show that the overall error is $O(h^2)$.

CHAPTER XI

APPLICATIONS

1. Linear Elliptic Equations

The results which have been obtained in previous sections for Dirichlet problems for nonlinear elliptic partial differential equations are applicable to Dirichlet problems for general, linear, uniformly elliptic partial differential equations as special cases.

In order to illustrate this, we consider the problem given by

$$\begin{aligned}(11.1)\qquad & A(x,y)\partial^2u/\partial x^2 + C(x,y)\partial^2u/\partial y^2 + D(x,y)\partial u/\partial x + E(x,y)\partial u/\partial y \\ & + F(x,y)u = G(x,y) \quad , \quad (x,y) \in R\end{aligned}$$

$$(11.2)\qquad u(x,y) = \phi(x,y) \quad , \quad (x,y) \in S.$$

The region R with boundary S is assumed to satisfy the smoothness conditions given in Chapter II. The coefficients and the function ϕ are assumed to have Hölder continuous partial derivatives of second and fourth order respectively, and the function F is assumed to be nonpositive.

Equation (11.1) is uniformly elliptic if there exist positive constants k_0 and k_1 such that

$$\text{and} \qquad \left.\begin{aligned} & k_1 \geq A(x,y),\ C(x,y) \geq k_0 \\ & k_1 \geq |D(x,y)|,\ |E(x,y)| \end{aligned}\right\} (x,y) \in R + S.$$

The uniform ellipticity of equation (11.1), the condition that F is nonpositive, and the conditions on R are sufficient to guarantee that the given problem has a unique solution u.

We now make the additional assumption that, for $h < 2k_0/k_3$, a set of mesh lines can be superimposed over the region R in such a way that the four rectangular neighbors of each mesh point in R are in $R + S$, i.e., all mesh points in R are regular mesh points. This assumption is in no way essential but does simplify the following discussion.

Corresponding to a given set of mesh lines, an approximating finite difference boundary value problem is formulated by replacing the partial derivatives in equation (11.1) by central divided differences. We have

$$\begin{aligned} &A(x_i,y_j)(\Delta\nabla)_x U_{i,j} + C(x_i,y_j)(\Delta\nabla)_y U_{i,j} + D(x_i,y_j)(\Delta+\nabla)_x U_{i,j} \\ &+ E(x_i,y_j)(\Delta+\nabla)_y U_{i,j} + F(x_i,y_j)U_{i,j} = G(x_i,y_j), \quad (x_i,y_j) \in R_h \end{aligned} \tag{11.3}$$

$$U_{i,j} = \phi(x_i,y_j) \quad , \ (x_i,y_j) \in R_S \ . \tag{11.4}$$

The coefficients in equation (11.3) satisfy conditions (6.1), and by Lemma 6.2, the finite difference problem has a unique solution for $h < 2k_0/k_1$.

A finite difference equation for the error in the solution of the problem given by equations (11.3) and (11.4) is obtained by substituting for U in equation (11.3) from the equation

$$U_{i,j} = u_{i,j} + \epsilon_{i,j} \tag{11.5}$$

where $\epsilon_{i,j}$ is the error. We obtain

$$\begin{aligned} &A(x_i,y_j)(\Delta\nabla)_x(u_{i,j}+ \epsilon_{i,j}) + C(x_i,y_j)(\Delta\nabla)_y(u_{i,j}+ \epsilon_{i,j}) \\ &+ D(x_i,y_j)(\Delta+\nabla)_x(u_{i,j}+ \epsilon_{i,j}) + E(x_i,y_j)(\Delta+\nabla)_y(u_{i,j}+ \epsilon_{i,j}) \\ &+ F(x_i,y_j)(u_{i,j}+ \epsilon_{i,j}) = G(x_i,y_j), \quad (x_i,y_j) \in R_h. \end{aligned} \tag{11.6}$$

By using relationships such as equations (3.13) and (3.14) and the fact that the difference quotients are linear, equation (11.6) can be rewritten as

$$\begin{aligned} &A(x_i,y_j)(\Delta\nabla)_x\epsilon_{i,j} + C(x_i,y_j)(\Delta\nabla)_y\epsilon_{i,j} + D(x_i,y_j)(\Delta+\nabla)_x\epsilon_{i,j} \\ &+ E(x_i,y_j)(\Delta+\nabla)_y\epsilon_{i,j} + F(x_i,y_j)\epsilon_{i,j} = -h^2[A(x_i,y_j)(M_4)_{i,j} \\ (11.7)\quad &+C(x_i,y_j)(N_4)_{i,j} + 2D(x_i,y_j)(M_3)_{i,j} + E(x_i,y_j)(N_3)_{i,j}]/12 \\ &- A(x_i,y_j)(\partial^2u/\partial x^2)_{i,j} - C(x_i,y_j)(\partial^2u/\partial y^2)_{i,j} - D(x_i,y_j)(\partial u/\partial x)_{i,j} \\ &- E(x_i,y_j)(\partial u/\partial y)_{i,j} - F(x_i,y_j)u_{i,j} + G(x_i,y_j) \end{aligned}$$

where (M_i) and (N_i) have the same meaning as in Chapter V.

The last six terms in equation (11.7) sum to zero because of equation (11.1). The finite difference boundary value problem for the error is then given by

$$\begin{aligned} &A(x_i,y_j)(\Delta\nabla)_x\epsilon_{i,j} + C(x_i,y_j)(\Delta\nabla)_y\epsilon_{i,j} + D(x_i,y_j)(\Delta+\nabla)_x\epsilon_{i,j} \\ (11.8)\quad &+ E(x_i,y_j)(\Delta+\nabla)_y\epsilon_{i,j} + F(x_i,y_j)\epsilon_{i,j} = H(x_i,y_j),\ (x_i,y_j) \in R_h \end{aligned}$$

$$(11.9)\qquad \epsilon_{i,j} = 0\ ,\quad (x_i,y_j) \in R_S$$

where

$$\begin{aligned} H(x_i,y_j) = -h^2[&A(x_i,y_j)(M_4)_{i,j} + C(x_i,y_j)(N_4)_{i,j} + D(x_i,y_j)(M_3)_{i,j} \\ &+ E(x_i,y_j)(N_3)_{i,j}]/12. \end{aligned}$$

Lemma 6.2 applies to the finite difference problem for the error; thus, for $h < 2k_0/k_1$, the error is uniquely determined. Theorem 6.1 also applies to the finite difference problem for the error, and we have, for $h < 2k_0/k_1$,

$$\max_{(x_i,y_j)\epsilon R_h} |\epsilon_{i,j}| \leq p_{i,j} \max_{(x_i,y_j)\epsilon R_h} |H(x_i,y_j)|$$

$$\leq h^2(\sigma_1^{X/h})k_1[\bar{M}_4 + \bar{N}_4 + 2\bar{M}_3 + 2\bar{N}_3]/12$$

where

$$\sigma_1 = \left[\frac{1 + h^2/2k_0 + h/2k_0(\ k_0 + k_1^2 + h^2)^{1/2}}{1 - h\ k_1/2k_0}\right]$$

and X is the maximum distance across the region R. From Corollary 1 to Theorem 6.1, there exists an $\widetilde{h}$ such that, for $h < \widetilde{h}$,

$$\max_{(x_i,y_j)\epsilon R_h} |\epsilon_{i,j}| \leq h^2 e^{\mu_1 X} k_1[\bar{M}_4 + \bar{N}_4 + 2\bar{M}_3 + 2\bar{N}_3]/12$$

where

$$\mu = [k_1 + (k_1^2 + 4k_0)^{1/2}]/2k_0 \ .$$

Thus, we have an a priori bound, for h sufficiently small, for the error which is proportional to h^2.

We note that the restriction on the region R which enables the construction of mesh lines in such a way that all interior mesh points are regular mesh points is unnecessary. In case irregular mesh points exist, asymmetric difference approximations such as those given by equations (3.25) and (3.26) are used. For such a problem, the above analysis results in an error bound which is proportional to h; however, the techniques developed in Chapter IX are applicable, and an error bound which is proportional to

h^2 can be obtained.

In addition, equation (11.1) is easily generalized to include a mixed derivative term. If equation (11.1) contains a mixed derivative term, it is necessary to apply the transformation given in Chapter III before proceeding with the analysis given above.

As was noted in Chapter I, error bounds on solutions of finite difference approximations to Dirichlet problems for some linear elliptic partial differential equations are given in Gerschgorin [1930]. Gerschgorin's results are briefly summarized as follows.

Consider the Dirichlet problem given by

$$Lu = G(x,y) \quad , (x,y) \in R \tag{11.10}$$

$$u = \phi(x,y) \quad , (x,y) \in S \tag{11.11}$$

where L is a linear elliptic operator, and where the coefficients in L, the functions G and ϕ, and $R + S$ satisfy the smoothness conditions given at the beginning of this chapter.

Assume that an approximating finite difference boundary value problem is formulated by the use of the difference quotients given in Chapter III, and let the error in the solution of the finite difference problem be denoted by ϵ. Let the region R be included in the circular disc which is bounded by

$$(x-x_0)^2 + (y-y_0)^2 = r^2.$$

Then, if L is the Laplacian, i.e.,

$$L = \partial^2/\partial x^2 + \partial^2/\partial y^2$$

and $G(x,y) \equiv 0$, for h sufficiently small,

$$(11.12) \qquad \max |\epsilon| \leq h^2r^2 \max(\bar{M}_4,\bar{N}_4)/12 + h^2 \max(\bar{M}_2,\bar{N}_2).$$

The second term on the right side of equation (11.12) is present only when the region R contains irregular mesh points.

Gerschgorin considers two other special cases. Let L be defined by

$$(11.13) \qquad Lu = A\ \partial^2u/\partial x^2 + C\ \partial^2u/\partial y^2 + D\ \partial u/\partial x + E\ \partial u/\partial y + Fu$$

where $A > 0$, $C > 0$, and $F \leq 0$. If, in addition, the coefficients D and E are everywhere positive and

$$A + C + 1 + 2\sqrt{2}\ r^2F/2 > 0$$

$$D + E + \sqrt{2}\ rF > 0,$$

then, for h sufficiently small,

$$|\epsilon| \leq \frac{h^2}{24}\left[(1+2\sqrt{2})r^2\max\left\{\frac{2(\bar{M}_4,\bar{N}_4)(A+C)}{2(A+C)+(1+2\sqrt{2})rF}\right\} + 4\sqrt{2}\ r\ \max\left\{\frac{(\bar{M}_3,\bar{N}_3)(D+E)}{D+E+\sqrt{2}\ rF}\right\}\right].$$

For the second special case, L is defined as in equation (11.13), the coefficients D and E each have the same sign throughout R, and

$$|D| + |E| + \sqrt{2}\ rF > 0.$$

In this case, for h sufficiently small,

$$|\epsilon| \leq \frac{h^2}{24}\left[(1+2/\overline{2})r^2\ \max\left\{\frac{(\bar{M}_4,\bar{N}_4)(A+C)}{A+C+(1+2\sqrt{2}\ )r^2F}\right\} + 4\sqrt{2}\ r\ \max\left\{\frac{(\bar{M}_3,\bar{N}_3)(|D+E|)}{|D+E|+\sqrt{2}\ rF}\right\}\right].$$

By comparing the error bounds on solutions of finite difference approximations to Dirichlet problems for linear elliptic partial differential equations which are derived at the beginning of this chapter with Gerschgorin's results, we conclude that our results are much more general but that Gerschgorin's results are much sharper. For example, for the Dirichlet problem for Laplace's equation, our error bound increases exponentially with the diameter of the region R whereas Gerschgorin's error bound increases as the square of the radius of the region R. On the other hand, Gerschgorin's analysis applies only to special cases, whereas our analysis is applicable to the Dirichlet problem for any linear, uniformly elliptic partial differential equation which satisfies the given smoothness conditions and for which $F \leq 0$.

2. A Nonlinear Problem

The solutions of many of the problems associated with nuclear reactor design are obtainable only as finite difference approximations. We consider here a relatively simple problem in this field which might be encountered in the design of a research reactor containing a neutron irradiation facility.

We assume that the irradiation facility consists of a long pipe through the reactor core with an elliptic cross section and that we are interested in determining the thermal neutron flux distribution, at constant reactor power, inside the pipe in a plane perpendicular to its axis. The thermal neutron flux satisfies the diffusion equation

$$D(\partial^2 u/\partial x^2 + \partial^2 u/\partial y^2) - \Sigma u + W = 0$$

where D is the diffusion coefficient, Σ is the neutron absorption (removal) cross section, and W is the thermal neutron source. The diffusion coefficient is a positive constant. The neutron absorption cross section depends on the rate at which neutrons are absorbed which in turn depends on the neutron flux. We assume that Σ is given by

$$\Sigma\,(u) = \alpha\left(1 + \frac{1}{1+\zeta u}\right)$$

where α and ζ are positive constants. The thermal neutron source is related to the neutron scattering and absorption rates at thermal and higher neutron energies. We assume that W can be approximated by

$$W = \gamma - e^{u/\delta}$$

where γ and δ are positive constants.

The thermal neutron flux at the surface of the irradiation facility is given by a function ϕ of position only.

The thermal neutron flux distribution in the plane perpendicular to the axis of the irradiation facility is then the solution of the problem given by

$$(11.14) \qquad \partial^2u/\partial x^2 + \partial^2u/\partial y^2 = \frac{1}{D}\left[\alpha\left(1+\frac{1}{1+\zeta u}\right)\cdot u + e^{u/\delta} - \gamma\right], \quad (x,y) \in R$$

$$(11.15) \qquad u = \phi(x,y), \ (x,y) \in S$$

where R is the region subtended by the facility in a plane perpendicular to its axis and S is the boundary of R (see Figure 11.1).

Equation (11.14) is uniformly elliptic with k_0 and k_1 of condition (2.2) each equal to unity. Condition (2.10) is satisfied provided

$$\alpha\left(1+\frac{1}{1+\zeta u}\right) - \frac{\alpha\zeta u}{(1+\zeta u)^2} + \frac{1}{\delta}e^{u/\delta} \geqq 0$$

for all $(x,y) \in R$ and for all nonnegative u.[1] We assume that this is true.

The coefficients in equation (11.14) satisfy all necessary conditions given in Chapter II and we assume that the function $\phi(x,y)$ does also. In addition, the region R is sufficiently smooth.

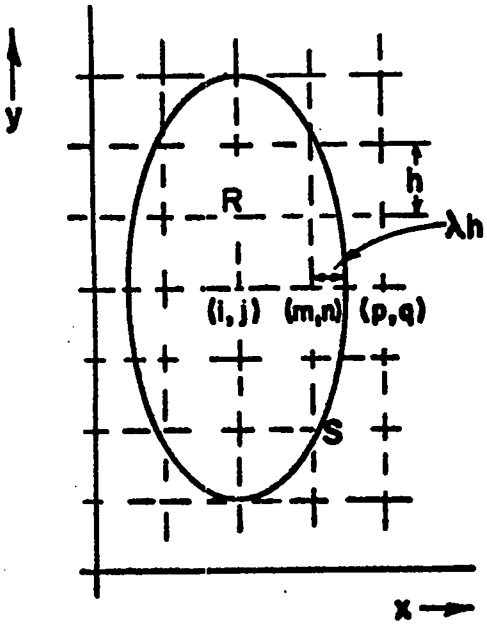


FIGURE 11.1

For purposes of illustration, a rather coarse set of mesh lines is superimposed over the region R in Figure 11.1. The interior mesh points on the major axis of the ellipse, except the ones nearest the boundary, are regular mesh points; all others are irregular. At the regular mesh points, the partial derivatives in equation (11.14) are approximated by

[1] From physical considerations, we know that u is nonnegative.

symmetric difference quotients to obtain finite difference equations for u. The finite difference equation at the point (x_i, y_j) is, for example,

$$(11.16) \qquad \begin{aligned} &(\Delta\nabla)_x U_{i,j} + (\Delta\nabla)_y U_{i,j} \\ &= \frac{1}{D}\left[\alpha\left(1 + \frac{1}{1+\zeta U_{i,j}}\right)U_{i,j} + \exp\left(\frac{U_{i,j}}{\delta}\right) - \gamma\right] . \end{aligned}$$

At irregular mesh points, the partial derivatives can be replaced by difference quotients such as the one given by equation (3.19); or, alternatively, the approximate solution U can be expressed at irregular mesh points in terms of its values at an adjacent regular mesh point and an adjacent boundary mesh point, i.e., by equation (3.15). We assume that the second alternative is used; then U at the point (x_m, y_n) is given by

$$(11.17) \qquad U_{m,n} = [\lambda/(\lambda+1)]U_{i,j} + [1/(\lambda+1)]U_{p,q}.$$

An equation such as (11.16) or (11.17) is applicable at each mesh point in R. These equations comprise a nonlinear system of algebraic equations with an equal number of equations and unknowns. This system can be solved by one of the methods mentioned in Chapter IV. The error analysis which has been presented in previous chapters can be applied to the difference E between the solution of this system of equations and the solution of the given problem.

The finite difference problem for the error is derived as follows. We first substitute for the approximate solution in equation (11.14) from

$$U = u + E$$

to obtain

$$(\Delta\nabla)_x(u_{i,j}+E_{i,j}) + (\Delta\nabla)_y(u_{i,j}+E_{i,j})$$
$$= \frac{1}{D}\left[\alpha\left(1+\frac{1}{1+\zeta(u_{i,j}+E_{i,j})}\right)(u_{i,j}+E_{i,j}) + \exp\frac{u_{i,j}+E_{i,j}}{\delta} - \gamma\right].$$

By using an expansion such as equation (5.3), the right-hand side of equation (11.18) can be written as

$$\frac{1}{D}\left[\alpha\left(1+\frac{1}{1+\zeta(u_{i,j}+E_{i,j})}\right)(u_{i,j}+E_{i,j}) + \exp\left(\frac{u_{i,j}+E_{i,j}}{\delta}\right) - \gamma\right]$$
$$= \frac{1}{D}\left[\alpha\left(1+\frac{1}{1+\zeta u_{i,j}}\right)u_{i,j} + \exp\left(\frac{u_{i,j}}{\delta}\right) - \gamma\right]$$
$$+ \frac{E_{i,j}}{D}\left\{\int_0^1\left[\alpha\left(1+\frac{1}{1+\zeta(u_{i,j}+\theta E_{i,j})}\right) - \frac{\zeta(u_{i,j}+\theta E_{i,j})}{(1+\zeta u_{i,j}+\zeta E_{i,j}\theta)^2}\right.\right.$$
$$\left.\left.+ \frac{1}{\delta}\exp\left(\frac{u_{i,j}+\theta E_{i,j}}{\delta}\right)\right]d\theta\right\}.$$

By using this expansion and relationships such as equation (3.14), we can write equation (11.18) as

$$(\Delta\nabla)_xE_{i,j} + (\Delta\nabla)_yE_{i,j} + f_{i,j}E_{i,j} = g_{i,j} - (\partial^2u/\partial x^2)_{i,j} - (\partial^2u/\partial y^2)_{i,j}$$
$$+ \frac{1}{D}\left[\alpha\left(1+\frac{1}{1+\zeta u_{i,j}}\right)u_{i,j} + \exp\left(\frac{u_{i,j}}{\delta}\right) - \gamma\right]$$

where

$$f_{i,j} = -\frac{1}{D}\left\{\int_0^1 \left[\alpha\left(1+\frac{1}{1+\zeta(u_{i,j}+\theta E_{i,j})}\right)\right.\right.$$

$$\left.\left. -\frac{\alpha\zeta(u_{i,j}+\theta E_{i,j})}{(1+\zeta u_{i,j}+\zeta E_{i,j}\theta)^2}+\frac{1}{\delta}\exp\left(\frac{u_{i,j}+\theta E_{i,j}}{\delta}\right)\right]d\theta\right\}$$

and

$$g_{i,j} = -h^2\left[(\partial^4 u/\partial x^4)_{i\pm\theta,j}+(\partial^4 u/\partial y^4)_{i,j\pm\phi}\right]/12,\ 0 \le \theta,\ \phi \le 1.$$

The last three terms in this equation sum to zero because of equation (11.14), and we have at regular mesh points

$$\tilde{L}_h E_{i,j} = (\Delta\nabla)_x E_{i,j} + (\Delta\nabla)_y E_{i,j} + f_{i,j}E_{i,j} = g_{i,j}.$$

The mesh point (x_m, y_n) is a typical irregular mesh point. At this mesh point, we have from equation (11.17)

$$u_{m,n} + E_{m,n} = [\lambda/(\lambda+1)](u_{i,j}+E_{i,j}) + [1/(\lambda+1)]u_{p,q},\ 0 < \lambda < 1.$$

By rearranging this equation and using relationships such as equation (3.22), we obtain

$$\tilde{L}_{b1}E_{m,n} = E_{m,n} - [\lambda/(\lambda+1)]E_{i,j} = g'_{m,n}$$

where $g'_{m,n}$ is the product of h^2 and a linear combination of second partial derivatives of u with respect to either x or y.

The finite difference problem for the error is then given by

$$\tilde{L}_h E_{i,j} = g_{i,j} \quad , \quad (x_i,y_j) \in R_h$$

$$\tilde{L}_{b1} E_{i,j} = g'_{i,j} \quad , \quad (x_i,y_j) \in R_b$$

$$E(x,y) = 0 \quad , \quad (x,y) \in R_S$$

where $g_{i,j}$ and $g'_{i,j}$ are $O(h^2)$. This finite difference problem is the same as problem P_1, Chapter V; and from Chapter VII, for h sufficiently small, its solution is bounded by a quantity which is $O(h^2)$.

3. The Problem of Minimal Surfaces

In this section, we discuss the numerical approximation of the solution of a Dirichlet problem for an equation which contains a mixed derivative term. The principal purpose of this discussion is to illustrate the use of the transformation given in Chapter III for equations containing mixed derivative terms.

Let R be a region in the x,y plane which is bounded by a Jordan curve S. Let ϕ be a closed curve in x,y,u space which has a one-to-one projection onto S. The problem of determining a function $u(x,y)$ which is continuous in $R+S$, has continuous derivatives up to second order in R, reduces to ϕ on S, and satisfies in R the partial differential equation

$$(11.19) \quad [1 + (\partial u/\partial y)^2]\partial^2 u/\partial x^2 - 2\partial u/\partial x \; \partial u/\partial y \; \partial^2 u/\partial x \partial y$$

$$+ [1 + (\partial u/\partial x)^2]\partial^2 u/\partial y^2 = 0$$

is known as the problem of minimal surfaces.[2]

[2] This problem is also known as Plateau's problem.

In order to state an existence theorem for the problem of minimal surfaces, we first describe what is meant by a three-point condition. Let S^* be the curve defined in x,y,u space by the equation $u = \phi(x,y)$. Let P_1^*, P_2^*, and P_3^* be three distinct points on S^* and denote by θ the positive acute angle between the x,y plane and the plane passing through P_1^*, P_2^*, and P_3^*. If, for all possible positions of the points P_1^*, P_2^*, and P_3^*, the quantity θ is less than or equal to some fixed finite constant Δ, then the boundary function ϕ satisfies a three-point condition with constant Δ.[3]

We now state the following[4]

THEOREM 11.1. Let there be given, on a convex Jordan curve S in the x,y plane, a function ϕ which satisfies a three-point condition with some constant Δ. Consider all functions $u(x,y)$ which satisfy, in the region R bounded by S, a Lipschitz condition and which reduce on S to the function ϕ. Then there exists in this class a function $u_0(x,y)$ which satisfies the partial differential equation (11.19).

We assume in the following that the region R is convex and that ϕ satisfies a three-point condition with some constant Δ.

Because of the three-point condition, the boundary function ϕ also satisfies a Lipschitz condition with some Lipschitz constant M. We then seek a solution of the given problem in the class of functions F which satisfy a Lipschitz condition with Lipschitz constant M and which reduce on S to the function ϕ.

[3] Rado [1951], p. 49. In addition to implying a restriction on the boundary function ϕ, the three-point condition implies that the curve S contains no arc which is a straight segment.

[4] Ibid., p. 61.

If a function u is in the class F, the first partial derivatives of u are bounded in absolute value by the Lipschitz constant M. Equation (11.19) is uniformly elliptic provided the coefficients are evaluated for a function $u \in F$ with k_1 and k_0 of condition (2.2) equal respectively to $(1+M^2)$ and unity. Condition (2.10) is also satisfied by equation (11.19).

In terms of the notation previously introduced, we write equation (11.19) as follows:

$$(11.20) \qquad A\, \partial^2u/\partial x^2 + 2B\, \partial^2u/\partial x\partial y + C\, \partial^2u/\partial y^2 = 0$$

where

$$A = [1 + (\partial u/\partial y)^2],\ B = -\partial u/\partial x\ \partial u/\partial y,\ \text{and}\ C = [1 + (\partial u/\partial x)^2].$$

We transform equation (11.20) into the form

$$(11.21) \qquad A'\, \partial^2u/\partial x^2 + 2B'\, \partial^2u/\partial z^2 + C'\, \partial^2u/\partial y^2 = 0$$

where

$$A' = A - B \cot \tau,\ B' = B(\sin 2\tau)^{-1},\ C' = C - B \tan \tau$$

and

$$(11.22) \qquad \tan \tau = (2B)^{-1}[C - A + (C^2-2AC+A^2+ 4B^2)^{1/2}],\ B \neq 0.$$

Here, τ is the angle between the z and x axes and depends on the values of the coefficients in equation (11.20) at each point in R. By equations (11.20) and (11.22),

$$\tan \tau = -(\partial u/\partial x)(\partial u/\partial y)^{-1},\ (\partial u/\partial y) \neq 0;$$

thus

$$A' = 1 + (\partial u/\partial y)^2 - (\partial u/\partial x)(\partial u/\partial y)^2(\partial u/\partial x)^{-1} = 1$$

$$B' = -\frac{(\partial u/\partial x)(\partial u/\partial y)}{\sin 2\tau} = -\frac{(\partial u/\partial x)(\partial u/\partial y)}{2\sin\tau\cos\tau}$$

$$= \frac{(\partial u/\partial x)(\partial u/\partial y)}{2(\partial u/\partial x)(\partial u/\partial y)}[(\partial u/\partial x)^2 + (\partial u/\partial y)^2]$$

$$= \frac{1}{2}[(\partial u/\partial x)^2 + (\partial u/\partial y)^2]$$

$$C' = 1 + (\partial u/\partial x)^2 - (\partial u/\partial x)^2(\partial u/\partial y)(\partial u/\partial y)^{-1} = 1$$

and the transformed Dirichlet problem is given by

$$\partial^2 u/\partial x^2 + [(\partial u/\partial x)^2 + (\partial u/\partial y)^2]\partial^2 u/\partial z^2 + \partial^2 u/\partial y^2 = 0, \quad (x,y) \in R$$

$$u(x,y) = \phi(x,y) \qquad , \quad (x,y) \in S.$$

In order to extend the discussion further, we assume that a finite difference approximation to the above problem is selected and that the system of nonlinear algebraic equations which result from the discretization are to be solved by the "natural iteration" method which is discussed in Chapter IV.

At each mesh point in R, we have a difference equation of the form

$$(11.23) \qquad A'_{i,j}\left(U^{(n)}_{i,j}\right) D^2_x U^{(n+1)}_{i,j} + 2B'\left(U^{(n)}_{i,j}\right) D^2_z U^{(n+1)}_{i,j} + C'\left(U^{(n)}_{i,j}\right) D^2_y U^{(n+1)}_{i,j} = 0$$

where D^2_x, D^2_z, and D^2_y denote appropriate difference quotients.

We consider first the case when D_z^2 is the difference quotient $(\Delta\nabla)_z$ (see Chapter III). The only new problem which we encounter is that of determining a method for selecting the direction z at each mesh point before solving for values of each successive iterant.

The optimum value of $\gamma_{i,j}$ at a mesh point (x_i,y_j), with regard to maximizing the coefficients $A'(x_i,y_j)$ and $B'(x_i,y_j)$, is given by

$$\gamma_{i,j} = \left(D_xU_{i,j}^{(n)}\right)\left(D_yU_{i,j}^{(n)}\right)^{-1}. \tag{11.24}$$

We wish to select an approximation $\tilde{\gamma}_{i,j} = \pm\,\alpha/\beta$ to $\gamma_{i,j}$, where α and β are positive integers, such that

$$A'_{i,j} = \left[1 + \left(D_xU_{i,j}^{(n)}\right)^2\right] - (\tilde{\gamma}_{i,j})^{-1}D_xU_{i,j}^{(n)}D_yU_{i,j}^{(n)} \geq 0 \tag{11.25}$$

and

$$C'_{i,j} = \left[1 + \left(D_yU_{i,j}^{(n)}\right)^2\right] - \tilde{\gamma}_{i,j}D_xU_{i,j}^{(n)}D_yU_{i,j}^{(n)} \geq 0. \tag{11.26}$$

Moreover, for convenience, the integers α and β should be as small as possible.

If $D_xU_{i,j}^{(n)}$ is equal to zero, the z direction is taken to be the x direction and equation (11.23) becomes

$$\left[1 + \left(D_yU_{i,j}^{(n)}\right)^2\right] D_x^2U_{i,j}^{(n+1)} + D_y^2U_{i,j}^{(n+1)} = 0.$$

Similarly, if $D_yU_{i,j}^{(n)}$ is equal to zero, the direction z is taken to be the y direction, and at this mesh point

$$D_x^2U_{i,j}^{(n+1)} + \left[1 + \left(D_xU_{i,j}^{(n)}\right)^2\right] D_y^2U_{i,j}^{(n+1)} = 0.$$

Now, suppose $\gamma_{i,j} \neq 0$. Let $n\bar{\gamma}_{i,j}$ denote the integer nearest $n\gamma_{i,j}$, n = 1, 2, 3, ..., Take successive values of $\tilde{\gamma}_{i,j}$ to be $n\tilde{\gamma}_{i,j}/n$, n = 1, 2, 3, ..., Each successive value of $\tilde{\gamma}_{i,j}$ is tested to determine whether or not equations (11.25) and (11.26) are satisfied. The first such value of $\tilde{\gamma}_{i,j}$ which satisfies these equations is used to determine the direction of the line $z_{i,j}$. By Theorem 3.1, a value of $\tilde{\gamma}_{i,j}$ can always be found such that equations (11.25) and (11.26) are satisfied.

The results of applying the above procedure to a specific example are indicated in Figure 11.2. Corresponding to each trial value of $\tilde{\gamma}_{i,j}$, there is a pair of mesh points $(x_{i\pm\beta}, y_{j+\alpha})$. The mesh point $(x_{i+\beta}, y_{j+\alpha})$ corresponding to successive values of $\tilde{\gamma}_{i,j}$ which would be selected for $\gamma_{i,j} = 0.3$ are given in Figure 11.2.

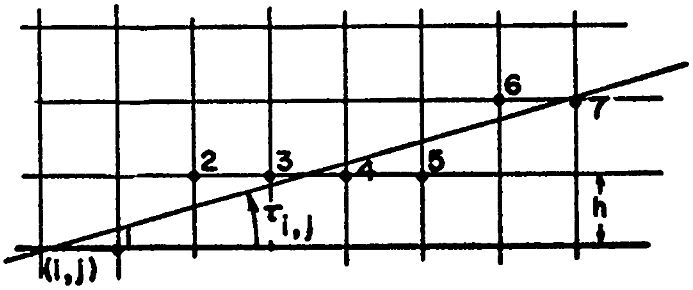


FIGURE 11.2

Next, we consider the case when D_x^2 stands for the difference quotient $(\Delta\nabla)_z$ (see Chapter X). Then $\gamma_{i,j}$ is calculated by equation (11.24) at each mesh point (x_i, y_j) in R. If either $D_x U_{i,j}^{(n)}$ or $D_y U_{i,j}^{(n)}$ is zero, the direction z is the same as in the previous case. Suppose $\gamma_{i,j}$ is positive; then τ is in the first quadrant (see Figure 11.3). If $\gamma_{i,j}$ is greater than one,

$$\delta_{i,j} = (\gamma_{i,j})^{-1},$$

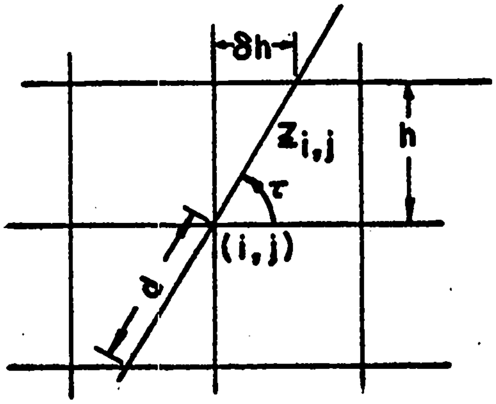
and if $\gamma_{i,j}$ is less than one,

$$\delta_{i,j} = \gamma_{i,j}.$$

Similar relationships hold if $\gamma_{i,j}$ is negative. The distance d is given in all cases by

$$d = h(1+\delta^2)^{1/2} .$$

For the configuration given in Figure 11.3, the term $(\overline{\Delta\nabla})_z U_{i,j}^{(n+1)}$ is computed from



$\gamma_{i,j} > 1$

FIGURE 11.3

$$(\overline{\Delta\nabla})_z U_{i,j}^{(n+1)} = \left[U_{i+\delta,j+1}^{(n+1)} - 2U_{i,j}^{(n+1)} + U_{i-\delta,j-1}^{(n+1)}\right] / d^2 .$$

where $U_{i+\delta,j+1}^{(n+1)}$ and $U_{i-\delta,j-1}^{(n+1)}$ are given by linear combinations of of $U_{m,n}^{(n+1)}$ at nearby points.

The special problems which might be encountered at mesh points near the boundary are not explored in detail here since it is not practical to attempt to generalize such situations any further than has been done in previous chapters.

The preceding discussion is predicated on the assumption that $U^{(0)}$ is taken from the class of functions F and that each successive iterant is in the class F in the sense that $|D_x U_{i,j}^{(n)}|$ and $|D_y U_{i,j}^{(n)}|$ are bounded at each mesh point in R by the constant M. If, for some iterant, this condition is not satisfied at all mesh points in R, there is no problem in proceeding as outlined above provided the first-order difference quotients are bounded in absolute value by some constant, say M'. Alternatively, it seems reasonable to expect that, since the solution of the given problem is known to be in the class F, convergence might be accelerated if the absolute values of the difference quotients are set equal to M for purposes of determining $\gamma_{i,j}$ and evaluating the coefficients when they would otherwise exceed M.

APPENDIX

ELIMINATION OF MIXED DERIVATIVE TERMS

The method of finite differences is more easily applied to obtain approximate solutions of problem P_0 if the mixed derivative term is eliminated. The method for doing this which is described here was introduced in Bramble and Hubbard [1963] as part of a study of linear elliptic difference equations. Bramble and Hubbard give a sketch of a proof of the validity of this procedure; a detailed proof is given here. This procedure can be used to eliminate the mixed derivative term from any uniformly elliptic partial differential equation in two independent variables provided the coefficients in the equation are continuous functions of their arguments.

Theorem 3.1 is proved here. We first prove the following lemma.

LEMMA A1. Let k_0 and k_1 be constants and let the coefficients in the equation

$$\text{(A1)} \qquad A(x,y)\ \partial^2u/\partial x^2 + 2B(x,y)\ \partial^2u/\partial x\partial y + C(x,y)\ \partial^2u/\partial y^2 = G(x,y), \qquad (x,y) \in R$$

satisfy the condition

$$\text{(A2)} \qquad k_1(\xi^2+\omega^2) \geq A\xi^2 + 2B\xi\omega + C\omega^2 \geq k_0(\xi^2+\omega^2)$$

for all real values of ξ and ω and for all $(x,y) \in R$. Then,

$$\text{(A3)} \qquad AC - B^2 \geq k_0^2$$

and

(A4) $$k_1^2 - k_0^2 \geq B^2.$$

Also, if

(A5) $$A - |B| \leq k_0^2/2k_1,$$

then

(A6) $$C - |B| > k_0^2/2k_1.$$

Proof: By alternately setting $(\xi,\omega) = (1,0)$ and $(\xi,\omega) = (0,1)$, we have that

(A7) $$k_1 \geq A, \; C \geq k_0.$$

Let $k_{0,i}$ be any positive number less than k_0. Then, by (A2),

$$A\xi^2 + 2B\xi\omega + C\omega^2 > k_{0,i}[\xi^2+\omega^2],$$

and for all nonzero ω,

(A8) $$[A-k_{0,i}](\xi/\omega)^2 + 2B(\xi/\omega) + [C-k_{0,i}] > 0.$$

Consider (A8) as a quadratic expression in the unknown (ξ/ω). Since this expression is nonzero, it has no real roots which implies that the discriminant is less than zero. Thus,

$$4B^2 - 4[A - k_{0,i}][C - k_{0,i}] < 0$$

or

$$AC - B^2 > -k_{0,i}^2 + k_{0,i}[A + C].$$

By (A7),

$$k_{0,i}[A + C] \geq 2k_{0,i}k_0 > 2k_{0,i}^2,$$

and thus

(A9) $$AC - B^2 > k_{0,i}^2.$$

Let $\{k_{0,i}\}$ be a sequence of positive numbers such that $k_{0,i}$ is less than k_0 for all i and such that the limit as i increases of $\{k_{0,i}\}$ is k_0. Inequality (A9) remains valid for all i and by passing to the limit

$$AC - B^2 \geq k_0^2$$

which proves the first statement of the lemma.

From inequalities (A3) and (A7), we have, respectively,

$$AC \geq k_0^2 + B^2$$

and

$$k_1^2 \geq AC.$$

Thus,

$$k_1^2 \geq k_0^2 + B^2$$

or

$$k_1^2 - k_0^2 \geq B^2$$

which establishes the second statement of the lemma.

Next, assume that (A5) is satisfied; then, because of (A7), $C > 0$, hence

$$AC \leq |B|C + C\, k_0^2/2k_1$$

or

(A10) $$AC - B^2 \leqq [C - |B|]|B| + C\, k_0^2/2k_1 .$$

By combining inequalities (A3), (A7), and (A10) we obtain

(A11) $$k_0^2 \leqq [C - |B|]|B| + k_0^2/2 .$$

We substitute $\xi = 1/\sqrt{2}$, $\omega = \pm 1/\sqrt{2}$ into inequality (A2) to obtain

$$k_1 \geqq A/2 \pm B + C/2 \geqq k_0$$

or

$$k_1 - k_0 \geqq \pm B + A/2 + C/2 - k_0 \geqq 0$$

from which, by (A7), we have

$$k_1 - k_0 \geqq \pm B.$$

This inequality together with (A11) gives us

$$[C - |B|][k_1 - k_0] \geqq k_0^2/2$$

or

$$C - |B| \geqq k_0^2/2\ [k_1 - k_0]$$

and finally

$$C - |B| \geqq k_0^2/2k_1 .$$

Because of symmetry, it is clear that if

$$C - |B| \leqq k_0^2/2k_1,$$

then

$$A - |B| > k_0^2/2k_1.$$

The principal result of this appendix is the following:

<u>THEOREM A1</u>. Let the coefficients in equation (A1) be continuous functions of their arguments and satisfy condition (A2). Then there exist constants k_0' and η, $k_0' > 0$, $1 \leqq \eta < \infty$, such that $\tan \tau = \gamma(x,y)$ can be specified at each point in R and

$$A' = A - \gamma^{-1}B \geqq k_0', \quad C' = C - \gamma B \geqq k_0'$$

$$B = B/\sin 2\tau \geqq 0$$

$$\gamma(x,y) = \pm \alpha/\beta$$

where α and β are relatively prime integers

$$1 \leqq \alpha,\ \beta \leqq \eta.$$

<u>Proof</u>: If $B = 0$, there is nothing to prove; thus, we assume that $|B| > 0$ throughout.

The angle τ is chosen at each point in R such that B' is positive; i.e., such that $\sin 2\tau$ has the same sign as B. We divide inequality (A3) by inequality (A4) to obtain

$$\text{(A12)} \qquad AC/B^2 \geqq 1 + k_0^2/(k_1^2-k_0^2)\ .$$

Since we choose τ such that $\sin\ 2\tau$ has the same sign as B, $\gamma = \tan \tau$ also has the same sign as B, and we can multiply (A12) by $(\gamma A)^{-1}B$ without changing the sense of the inequality to obtain

$$\text{(A13)} \qquad C/\gamma B \geqq B/\gamma A\ [1 + k_0^2/(k_1^2-k_0^2)].$$

Let R_1 and R_2 be subsets of R such that if $P \in R_1$, then

$$A - |B| \leq k_0^2/2k_1$$

and if $P \in R_2$, then

$$C - |B| \leq k_0^2/2 .$$

By Lemma A1, the sets R_1 and R_2 are disjoint and by the continuity of the coefficients in equation (A1), $R_1 \cup R_2$ is bounded and closed.

Let P_1 be a point in R_1. We choose relatively prime integers α and β such that

$$1 + k_0/4(k_1^2-k_0^2) < \alpha(P)A/\beta(P)|B| < 1 + 3k_0^2/4(k_1^2-k_0^2) \quad . \tag{A14}$$

Due to the continuity of the coefficients A and B, there exists an open set $S(P_1)$ containing P_1 such that (A14) holds for each point in $S(P_1)$. Corresponding to each point P in R_1, there exists such an open set $S(P)$. By the Heine-Borel Theorem, R_1 can be covered by a finite number of open sets $S(P)$. With each of the sets $S(P)$, there is associated a pair of relatively prime integers $\alpha(P)$ and $\beta(P)$ such that (A14) holds. Since there are only a finite number of such pairs of integers associated with any finite covering of R_1, there exists a constant η_1 such that

$$1 \leq \alpha(P),\ \beta(P) \leq \eta_1,\ P \in R_1.$$

We take $|\gamma(P)| = \alpha(P)/\beta(P)$ and sgn γ = sgn B. Then in R_1

$$A - \gamma^{-1}B > [k_0^2/4(k_1^2-k_0^2)]\gamma^{-1}B. \tag{A15}$$

By definition, in R_1,

$$|B| \geqq A - k_0^2/2k_1$$

$$\geqq (k_0/2)[2A/k_0 - k_0/k_1]$$

(A16) $$|B| \geqq k_0/2 \ .$$

Also,

(A17) $$|\gamma|^{-1} > 1/\eta_1 .$$

Therefore, from (A15), (A16), and (A17),

(A18) $$A - \gamma^{-1}B > [k_0^3/8(k_1^2-k_0^2)]/\eta_1 \ .$$

From (A14),

(A19) $$\gamma^{-1}B/A > [1 + 3k_0^2/4(k_1^2-k_0^2)]^{-1} \ .$$

By combining (A13), (A18), and (A19), we obtain

$$C/\gamma B > [1 + 3k_0^2/4(k_1^2-k_0^2)]^{-1}[1 + k_0^2/(k_1^2-k_0^2)]$$

$$> 4k_1^2/(4k_1^2-k_2^2)$$

$$> 1 + k_0^2/4k_1^2 + k_0^4/16k_1^4 + \ldots$$

$$> 1 + k_0^2/4k_1^2[1/(1-(k_0^2/4k_1^2))]$$

or

$$C - \gamma B > \gamma B k_0^2/(4k_1^2 - k_0^2)$$

and since

$$\gamma B > k_0/2\eta_1,$$

$$\text{(A20)} \qquad C - \gamma B > k_0^3/(k_1^2 - k_0/4)8\eta_1$$

Let

$$\lambda_1 = k_0^3[1/(k_1^2 - k_0/4)]/8\eta_1 \ .$$

Then, from (A15) and (A20), we have

$$\left.\begin{aligned} A - \gamma^{-1}B &= A' > \lambda_1 > 0 \\ C - \gamma B \quad &= C' > \lambda_1 > 0 \ . \end{aligned}\right\} \quad P \in R_1$$

In a manner exactly analogous to that given above, we can show there exists a constant η_2 such that for any point $Q \in R_2$, $|\gamma(Q)|$ can be chosen equal to $\alpha(Q)/\beta(Q)$ where $\alpha(Q)$ and $\beta(Q)$ are relatively prime integers and

$$1 \leqq \alpha(Q),\ \beta(Q) \leqq \eta_2.$$

We then let

$$\lambda_2 = k_0^3[1/(k_1^2 - k_0^2)]/8\eta_2$$

and obtain

$$\left.\begin{matrix} A' \\ C' \end{matrix}\right\} > \lambda_2,\ Q \in R_2.$$

The set $R_1 \cup R_2$ does not necessarily exhaust R. Consider the set $R - R_1 \cup R_2$. For each point in this set, we have

$$A - |B| > k_0^2/2k_1$$

and

$$C - |B| > k_0^2/2k_1,$$

and we take $|\gamma| \equiv 1$ for points in this set. Then,

$$\left.\begin{matrix} A' \\ C' \end{matrix}\right\} > k_0^2/2k_1 = \lambda_3$$

for such points.

Now, let

$$k_0' = \min[\lambda_1,\lambda_2,\lambda_3]$$

and

$$\eta = \max[\eta_1,\eta_2].$$

Then, for each point $(x,y) \in R$, the angle τ can be chosen such that

$$A',C' \geqq k_0' > 0,$$

$$B' \geqq 0,$$

and

$$\gamma(x,y) = \mp \alpha/\beta$$

where α and β are relatively prime integers and

$$1 \leqq \alpha,\ \beta \leqq \eta.$$

BIBLIOGRAPHY

Ablow, C. M., and C. L. Perry [1959]: Iterative solutions of the Dirichlet problem for $\nabla^2 u = u^2$, Jour. Soc. Indust. Appl. Math., vol. 7, no. 4, pp. 459-467.

Bers, L. [1953]: On mildly nonlinear partial difference equations of elliptic type, Jour. of Res. Nat. Bur. Standards, vol. 51, no. 5, pp. 229-236.

Bers, L., F. John, and M. Schechter [1964]: Partial Differential Equations, Interscience Publishers, New York, N. Y.

Bramble, J. H., and B. E. Hubbard [1962]: On the formulation of finite difference analogues of the Dirichlet problem for Poisson's Equation, Numerische Mathematik, vol. 4, pp. 313-327.

Bramble, J. H., and B. E. Hubbard [1963]: A theorem on error estimation for finite difference analogues of the Dirichlet problem for elliptic equations, Contributions to Differential Equations, vol. 2, pp. 319-340.

Collatz, L. [1933]: Bemerkungen zur Fehlerabschätzung für das Differenzenverfahren bei partiellen Differentialgleichungen, Z. Angew. Math. Mech., vol. 13, pp. 56-57.

Collatz, L. [1960]: The Numerical Treatment of Differential Equations, Springer-Verlag, Berlin.

Courant, R., K. Friedrichs, and H. Lewy [1928]: Über die partiellen Differenzengleicheingen der mathematischen Physik, Math. Am., vol. 100, pp. 32-74.

Courant, R., and D. Hilbert [1962]: Methods of Mathematical Physics, Interscience Publishers, New York, N. Y.

Douglas, J. [1961]: Alternating direction iteration for mildly nonlinear elliptic difference equations, Numer. Math., vol. 3, pp. 92-98.

Downing, A. C. [1960]: On the convergence of steady state multiregion diffusion calculations, Oak Ridge National Laboratory Report No. 2961, Union Carbide Corp., Oak Ridge, Tenn.

Epstein, B. [1962]: Partial Differential Equations,McGraw-Hill, Inc., New York, N. Y.

Forsythe, G., and W. R. Wasow [1960]: Finite-difference Methods for Partial Differential Equations, John Wiley and Sons, Inc., New York, N. Y.

Gerschgorin, S. [1930]: Fehlerabschätzung für das Differenzenverfabren zür Lösung partieller Differentialgleichungen, Z. Angew. Math., vol. 10, pp. 373-382.

Greenspan, D. [1960]: *On the approximate solution of elliptic differential equations with mixed partial derivatives*, Tech. Report No. 209, Math. Res. Ctr., U. S. Army, Univ. of Wis.

Greenspan, D. [1965]: *Introductory Numerical Analysis of Elliptic Boundary Value Problems*, Harper and Row, New York, N. Y.

Greenspan, D., and P. C. Jain [1964]: *On non-negative analogues of elliptic differential equations*, Tech. Report No. 490, Math. Res. Ctr., U. S. Army, Univ. of Wis.

Greenspan, D., and S. V. Parter [1965]: *Mildly nonlinear elliptic partial differential equations and their numerical solution II, Numer*. Math., vol. 7, pp. 129-146.

Gunn, J. E. [1964]: *On the two-stage method of Douglas for mildly non-linear elliptic difference equations*, Numer. Math. vol. 6, pp. 243-249.

Laasonen, P. [1957]: *On the degree of convergence of discrete approximations for the solutions of the Dirichlet problem*, Ann. Acad. Sci. Fenn. A. I., vol. 246, pp. 1-19.

Lefschetz, S. [1949]: *Introduction to Topology*, Princeton Univ. Press, Princeton, N. J.

Levinson, N. [1963]: *Dirichlet problem for $u = f(P,U)$*, Jour. of Math. and Mech., vol. 12, no. 4, pp. 567-575.

McAllister, G. T. [1964a]: *Quasilinear elliptic partial differential equations and difference equations*, Tech. Report No. 494, Math. Res. Ctr., U. S. Army, Univ. of Wis.

McAllister [1964b]: *Some nonlinear elliptic partial differential equations and difference equations*, Jour. Soc. Indust. Appl. Math., vol. 12, no. 4, pp. 772-777.

McAllister, G. T. [1964c]: *Difference methods for a nonlinear elliptic system of partial differential equations*, Tech. Report No. 504, Math. Res. Ctr., U. S. Army, Univ. of Wis.

Milne, W. E. [1949]: *Numerical Calculus*, Princeton Univ. Press, Princeton, N. J.

Motzkin, T. S., and W. Wasow [1953]: *On the approximation of linear elliptic differential equations by difference equations with positive coefficients*, J. Math. Phys., vol. 31, pp. 253-259.

Parter, S. V. [1964]: *Mildly nonlinear elliptic partial differential equations and their numerical solution I*, Tech. Report No. 470, Math. Res. Ctr., U. S. Army, Univ. of Wis.

Pohazaev, S. L. [1960]: *The Dirichlet problem for the equation $u = u^2$*, Soviet Mathematics, vol. 1, p. 1143.

Pucci, C. [1958]: Some topics in parabolic and elliptic equations, Inst. for Fluid Dynamics and Appl. Math. Lecture Series, vol. 36.

Rado, T. [1951]: On the problem of Plateau, Chelsea Pub. Co., New York, N. Y.

Rockoff, M. L. [1964]: Comparison of some iterative methods for solving large systems of linear equations, Nat. Bur. Stand. Report 8577.

Rosenbloom, P. C. [1952]: On the difference equation method for solving the Dirichlet problem, Nat. Bur. Standards Appl. Math. Ser., no. 18, pp. 231-237.

Schechter, S. [1962]: Iteration methods for nonlinear problems, Trans. Am. Math. Society, vol. 104, pp. 179-189.

Walsh, J. L., and David Young [1953]: On the accuracy of the numerical solution of the Dirichlet problem by finite differences, Jour. Res. Nat. Bur. Standards, vol. 51, pp. 343-363.

Walsh, J. L., and David Young [1954]: On the degree of convergence of solutions of difference equations to the solution of the Dirichlet problem, Jour. Math. Phys., vol. 33, pp. 80-93.

Wasow, W. [1952]: On the truncation error in the solution of Laplaces Equation by finite differences, Jour. Res. Nat. Bur. Standards, vol. 48, pp. 345-348.

Wasow, W. [1957]: The accuracy of difference approximations to plane Dirichlet problems with piecewise analytic boundary values, Quart. Appl. Math., vol. 15, pp. 53-63.