CONVERGENCE AND STABILITY

OF

FINITE DIFFERENCE SCHEMES

FOR SOME ELLIPTIC EQUATIONS

A Thesis

Submitted to the Faculty of Graduate Studies

in Partial Fulfilment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

in the

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF SASKATCHEWAN

by

MURLI MANOHAR GUPTA

Saskatoon, Saskatchewan

August 1971

CONVERGENCE AND STABILITY OF FINITE DIFFERENCE SCHEMES FOR SOME ELLIPTIC EQUATIONS

Murli Manohar Gupta

ABSTRACT

The problem of convergence and stability of finite difference schemes used for solving boundary value problems for some elliptic partial differential equations has been studied in this thesis. Generally a boundary value problem is first replaced by a discretized problem whose solution is then found by numerical computation.

The difference between the solution of the discretized problem and the exact solution of the boundary value problem is called the discretization error. This error is a measure of the accuracy of the numerical solution, provided the roundoff error is negligible. Estimates of the discretization error are obtained for a class of elliptic partial differential equations of order 2m ($m \geq 1$) with constant coefficients in a general n-dimensional domain. This result can be used to define finite difference approximations with an arbitrary order of accuracy.

The numerical solution of a discretized problem is usually obtained by solving the resulting system of algebraic equations by some iterative procedure. Such a procedure must be stable in order to yield a numerical solution. The stability of such an iteration scheme is studied in a general setting and several sufficient conditions of stability are obtained.

When a higher order differential equation is solved numerically, roundoff error can accumulate during the computations. In order to reduce this error the differential equation is sometimes replaced by several lower order differential equations. The method of splitting is analyzed for the two-dimensional biharmonic equation and the convergence of the discrete solution to the exact solution is discussed. An iterative procedure is presented for obtaining the numerical solution. It is shown that this method is also applicable to non-rectangular domains.

The accuracy of numerical solutions of a nonselfadjoint elliptic differential equation is discussed when it is solved with a finite non-zero mesh size. This equation contains a parameter which may take large values. Some extensions to the two-dimensional Navier-Stokes equations are also presented.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor R. Manohar, my thesis advisor, for his valued assistance and encouragement throughout the preparation of this thesis. Sincere gratitude is also expressed to Professor A.A. Samarskii whose private communication led to the ideas which developed in the results of Chapter 2.

I wish to thank Dr. G.L. Saini and Mr. John W. Stephenson for their helpful comments and careful reading of the manuscripts of this thesis. Thanks are also due to Dr. E.K.H. Hoehn and Dr. K.H.V. Booth for reading the preliminary manuscript.

I wish to express my thanks to all my friends, too numerous to name, for their help and encouragement. Special appreciation goes to Miss Elsie Chitrinia who also helped me type the rough draft. Thanks are due to Miss Mary-Ann Wiest for skilfully typing the final manuscript.

This research was made possible by the financial support of the University of Saskatchewan in the form of a Graduate Fellowship.

PREFACE

Several attempts have been made in the last decade to solve the Navier-Stokes equations numerically (see [37] and other references given there). In 1968, we tried to develop a computationally efficient method to solve these equations under certain boundary conditions describing some problems of fluid flow [37,38]. The primary concern in the above study was to obtain the numerical solutions for various values of a parameter, the Reynolds number, while the questions of convergence and stability of the finite difference schemes remained unanswered. We also found that these questions were not answered even for equations simpler than Navier-Stokes equations. In this thesis we discuss the convergence and stability of finite difference equations obtained by discretizing various partial differential equations of elliptic type. Although the equations treated may appear unrelated, they are all connected with the problem of Navier-Stokes equations in some way or other. We do not attempt to study any finite difference scheme for the Navier-Stokes equations, but hope that the results obtained in this thesis will give further insight into the methods of solving them numerically. An outline of the contents of the thesis follows.

In Chapter 1, we obtain error estimates for a general class of difference methods for the Dirichlet problem for an elliptic differential equation of order 2m with constant coefficients. There are very few results in the literature which deal with differential equations of order higher than two. Courant, Friedrichs and Lewy [1] constructed a difference analogue of the Dirichlet problem for the biharmonic equation and proved the convergence of the discrete solution to the exact solution as the mesh width h tends to zero. Saul'ev [2] considered the Dirichlet problem for a class of elliptic differential equations of order 2m ($m \geq 1$) and defined a difference approximation. He proved the existence of a solution of this discrete system and proved its convergence to a weak solution of the boundary value problem as $h \to 0$. Littman [3] also proved similar results for another class of differential equations with a strongly elliptic operator.

In 1964, Thomée [4] studied a general class of difference methods for the Dirichlet problem for elliptic equations with constant coefficients of order 2m ($m \geq 1$). He defined a finite difference approximation of the Dirichlet problem and proved that the solution of this approximation converges to the exact solution as $h \to 0$, with a discretization error of order $h^{\frac{1}{2}}$. In Chapter 1, we consider a modification of the difference approximation given by Thomée and prove that the order of discretization error is increased considerably. The method of our proof is essentially the same as that used by Thomée. Later on we show that some of the known results for the biharmonic equation and a general fourth order differential equation follow from our results as special cases. We also apply these results in Chapter 3 to the problem of the biharmonic equation.

Once a differential equation is discretized, a system of difference equations is obtained which is generally solved by some iterative method. These iterative methods can also be interpreted as resulting from the discretization of a corresponding time dependent Cauchy problem. In order to obtain the numerical solution of the difference equations by an iterative method, it is essential to know a priori that the iterative scheme is stable. In Chapter 2, we consider the stability of two-layer iteration schemes since any multi-layer iteration scheme can be reduced to a two-layer (two-level) scheme by introducing new variables [11]. In order to study the stability of these iteration schemes in general, we first put them in the form of an operator equation using linear operators in a real Hilbert space. It may be noted that the difference equations discussed in Chapter 1 can be put into this form and solved by iterative methods discussed in Chapter 2.

In the last few years, the problem of stability of two-layer iteration schemes has been studied by various authors. In particular, Samarskii [13] has reported various necessary and sufficient conditions of stability for such schemes with selfadjoint operators. He extended some of his results to the case of nonselfadjoint operators but the conditions of stability given by him involve the inverses of certain operators. These inverses are generally not known a priori and therefore such conditions are not useful in practice. In Chapter 2, we obtain several sufficient conditions of stability in the case of nonselfadjoint operators. These conditions are expressed in terms of the norms of operators known a priori. In some cases we show that these conditions are necessary, too. Finally, we demonstrate the application of some of the results obtained in this chapter by a numerical example.

In Chapter 3 we study some boundary value problems of the biharmonic equation in two dimensions. Such problems often occur in physics and engineering, for example, the motion of a viscous fluid at low Reynolds number, bending of an elastic plate, etc. Several authors have tried to solve the biharmonic equation numerically using finite difference methods, but most of these attempts have been restricted to problems in some simple domains such as a rectangle. One approach is to discretize the biharmonic equation directly and solve the resulting system of algebraic equations by some iterative method. Usually the convergence of these iterative methods is very slow and sometimes it can not be guaranteed even for simple domains. In order to overcome this difficulty, another approach which is frequently adopted is to replace the biharmonic equation by two simultaneous Poisson equations which can be more efficiently solved. However, the simultaneous solution of two coupled Poisson equations introduces some new problems of convergence, about which very little is known. We study this problem in Chapter 3 and define some finite difference schemes for solving the Poisson equations. We show the convergence of the discrete solution to the exact solution of the biharmonic boundary value problems as the mesh size tends to zero. We also obtain some estimates of the discretization error for these difference approximations.

In Chapter 4, we consider some of the problems associated with the computation of numerical solutions of biharmonic boundary value problems using the discretization schemes suggested in Chapter 3. The first boundary value problem is first replaced by two coupled Poisson equations which are discretized and solved, usually by an iterative method. These iterations are called "inner iterations". The solution of the whole system, with a properly chosen approximation for the boundary conditions, is obtained by an iterative process called "outer iterations". We show that these outer iterations diverge as $h \to 0$, irrespective of the inner iterations or the approximation used on the boundary. A modification of the basic outer scheme using a relaxation parameter makes it convergent. We study the spectral radii of the matrices governing the outer iterations and show the suitability of some of the approximations to be used on the boundary. Finally, we solve some boundary value problems using the methods developed in Chapters 3 and 4. We also compare our results with those obtained by using some of the existing methods.

In Chapter 5 we study the problem of solving a particular second order elliptic differential equation in a rectangular domain. This equation is nonselfadjoint and contains a parameter λ which may take large values. This equation has a similar character as the Navier-Stokes equations but is easier to analyze because of its linearity. It is discretized using two different finite difference schemes both of which are convergent and stable, in the sense of

Chapters 1 and 2, when the mesh size h is allowed to decrease. However, for actual computations, one has to use a finite nonzero mesh size and one of the finite difference schemes gives divergent results especially when used for large values of λ. This result was observed by Burns [35] for one-dimensional case who also gave some numerical computations for the discretization error. Boughner [36] obtained some more estimates which confirmed the above observation for the two-dimensional case. These results were contrary to the expectations. In this chapter we obtain the exact solution of the difference equations and study their asymptotic behaviour for large values of λ. The behaviour of the solutions as observed by Burns and Boughner can be easily explained by these asymptotic solutions of the difference equations. We have also obtained some stability conditions for these equations which confirm the above results. These results also explain, in part, why certain finite difference schemes used for solving the Navier-Stokes equations remain stable and convergent even for large values of the Reynolds number.

CONTENTS

		Page
	ABSTRACT	iii
	ACKNOWLEDGEMENTS	v
	PREFACE	vi
Chapter 1.	DISCRETIZATION ERROR FOR DIRICHLET PROBLEMS	1
1.1	Introduction	1
1.2	Mathematical Preliminaries	4
1.3	Finite Difference Approximation	14
1.4	Applications	17
Chapter 2.	STABILITY AND CONVERGENCE OF ITERATION SCHEMES	25
2.1	Preliminaries	25
2.2	Selfadjoint Operator	28
2.3	Nonselfadjoint Operator	32
2.4	Time-Dependent Operators	41
2.5	Stability With Respect to the Right Hand Side	46
2.6	A Numerical Example	48
Chapter 3.	THE BIHARMONIC EQUATION: DISCRETIZATION ERROR	52
3.1	Introduction	52
3.2	Discretization Error: The Second Boundary Value Problem	57

CONTENTS - (Continued)

		Page
3.3	Discretization Error: The First Boundary Value Problem	72
3.4	Examples of the Boundary Operator	76
Chapter 4.	THE BIHARMONIC EQUATION: NUMERICAL SOLUTION	80
4.1	The Outer Iteration Scheme	80
4.2	$\gamma_h \to \gamma$ as $h \to 0$	85
4.3	The Modified Iterations	88
4.4	Numerical Examples	92
4.5	Conclusions	99
Chapter 5.	COMPUTATIONAL PROBLEM ASSOCIATED WITH A NONSELFADJOINT EQUATION	102
5.1	Introduction	102
5.2	Asymptotic Expansion of the Solutions: One Dimensional Case	105
5.3	Two Dimensional Case	107
5.4	Stability	110
Chapter 6.	EXTENSION OF THE NUMERICAL PROCEDURE TO NAVIER-STOKES EQUATIONS	113
	REFERENCES	118

CHAPTER 1

DISCRETIZATION ERROR FOR DIRICHLET PROBLEMS

1.1 Introduction

In this chapter we study a class of finite difference schemes for the first boundary value problem for elliptic differential equations of order 2m ($m \geq 1$) with constant coefficients. For second order equations, one can apply the maximum principle to both the differential equation and its discrete analogue and thus obtain estimates of discretization error. However, the situation is quite different for higher order equations where the maximum principle can not be applied in general.

In 1928 Courant, Friedrichs and Lewy [1] constructed a difference analogue of the Dirichlet problem for the biharmonic equation and proved the convergence of the solution of the discrete problem to the exact solution when the mesh width is refined. Saul'ev [2] considered the problem of finding a weak solution of the Dirichlet problem for a class of elliptic equations of order 2m, $m \geq 1$. He defined a difference approximation and proved the existence of its solution. He also proved that this discrete solution converges in the mean to the weak solution (whose existence is not assumed) of the boundary value problem as the mesh size $h \to 0$. Littman [3] considered a generalized Dirichlet problem for a class of differential equations with strongly elliptic operators and vanishing Dirichlet data. He discretized the differential equations by replacing the derivatives with central differences and proved the existence of a unique discrete solution. He also proved the convergence of this discrete solution and its difference quotients to the exact solution and the corresponding derivatives as $h \to 0$. The estimates of discretization error were not obtained in any of these studies.

In 1964, Thomée [4] discussed a class of difference schemes for the Dirichlet problem for elliptic differential equations of order 2m with constants coefficients. He divided the set of grid points into two subsets and called them the sets of "interior" and "boundary" grid points. At each of the interior grid points he defined a difference operator consistent with the differential operator, whereas at the boundary grid points the approximate solution was obtained by interpolation using the boundary data. Thomée proved that the difference between the solution of this finite difference approximation and the exact solution is of order $h^{1/2}$ as $h \to 0$.

Zlámal [5] showed by means of an example that the estimates of Thomée as such can not be improved. Zlámal discussed the Dirichlet problem for a fourth order elliptic differential equation with variable coefficients. He prescribed a difference operator at each of the interior and boundary grid points and proved the discretization error to be of order $h^{3/2}$, as $h \to 0$. He also defined another difference analogue for the case of a rectangular domain and proved the discretization error to be of order h^2. Bramble [6] discussed the Dirichlet problem for the biharmonic equation and constructed a difference approximation with discretization error of order h^2 in a special norm and of order $h^2 (\log h^{-1})^{\frac{1}{2}}$ in the maximum norm.

The results of Zlámal cited above suggest that it is possible to obtain better error estimates than those of Thomée by prescribing a suitable difference operator for the boundary grid points. Thomée has also noted that his error estimate is of low order because of the crudeness of approximations at the boundary. We study, in this chapter, the Dirichlet problem for elliptic differential equations with constant coefficients of order 2m, $m \geq 1$. We consider a class of finite difference schemes and define a discrete operator at each of the interior as well as boundary grid points, thus introducing a modification of the difference approximation considered by Thomée. We prove that the order of discretization error is considerably improved by this modification. The method of our proof is essentially similar to that of Thomée with the difference that we use different sets and norms to include the discretization on the boundary grid points.

The estimates given by Zlámal for the fourth order differential equations follow from our results in the case of constant coefficients. For the same equations we define another difference approximation which is shown to have a discretization error of order h^2 for general domains. The results of Bramble are shown to follow from our theorems. Finally, we give an example of the Dirichlet problem for a second order elliptic differential equation.

1.2 Mathematical Preliminaries

Let D be a bounded domain in the n-dimensional vector space R^n with piecewise smooth boundary $\dot{D}$. Let x be a vector in D with components $x_1, x_2, \ldots x_n$ and m be a natural number. We shall consider the differential operator

$$(1.1) \qquad Lu \equiv L(D)u = \sum_{|\beta|=|\gamma|=m} a_{\beta\gamma} D^{\beta+\gamma} u \ , \ a_{\beta\gamma} = a_{\gamma\beta} \ ;$$

where $\beta = (\beta_1, \ldots, \beta_n) \ , \ \gamma = (\gamma_1, \ldots, \gamma_n) \ , \ |\beta| = \sum_{j=1}^{n} \beta_j$

and $D^\beta = (i^{-1} \ \partial/\partial x_1)^{\beta_1} \ \ldots \ (i^{-1} \ \partial/\partial x_n)^{\beta_n} \ , \ i = \sqrt{-1}$.

In (1.1) β and γ are multi-indices and β_j, γ_j are non-negative integers. The coefficients $a_{\beta\gamma}$ are real valued constants. We assume that L(D) is an elliptic operator so the characteristic polynomial L (ξ) is positive definite:

$$(1.2) \qquad L(\xi) = \sum_{|\beta|=|\gamma|=m} a_{\beta\gamma} \ \xi^{\beta+\gamma} \ \geq \ c|\xi|^{2m} \ ,$$

where $\xi^\beta = \xi_1^{\beta_1} \ldots \xi_n^{\beta_n} \ , \ |\xi|^2 = \sum_{j=1}^{n} |\xi_j|^2$

and c is a positive constant.

Let C^k (M) be the set of complex valued functions which are k times continuously differentiable on M. Let $L^2(R^n)$ be the set of complex valued functions, square integrable in R^n and $L^2(D)$ be the subset of elements of $L^2(R^n)$ which vanish outside D. For u, $v \in L^2(R^n)$ we define

$$(u,v) = \int_{R^n} u(x) \overline{v(x)} \, dx ,$$

$$||u|| = (u,u)^{\frac{1}{2}} .$$

We shall consider the Dirichlet problem

$$(1.3) \quad \begin{aligned} Lu(x) &= F(x) \quad , x \in D \\ D^\beta u(x) &= D^\beta f(x) \; , x \in \dot{D} \; , |\beta| \le m-1 . \end{aligned}$$

Under the assumptions of sufficient regularity of F, f and $\dot{D}$, the problem (1.3) has a unique solution in $C^k(\overline{D})$ where $k \ge 2m$, $\overline{D} = D \cup \dot{D}$ [4].

We now introduce mesh points in R^n of the form $\xi h = (\xi_1 h, \ldots, \xi_n h)$, $h > 0$ and ξ_j are integers. A grid function is a complex valued function defined on a set of grid points and for such a function we write $u_\xi = u(\xi h)$. If a grid function u_ξ is defined only for a subset M of the grid points, its definition can be extended to all grid points by setting $u_\xi = 0$, $\xi \notin M$.

We consider a class of difference analogues of Lu of the form

$$(1.4) \qquad L_h u_\xi = h^{-2m} \sum_\alpha c_\alpha u_{\xi+\alpha} \; , \; \alpha = (\alpha_1, \ldots, \alpha_n) \; ,$$

where α_j are integers and c_α are complex numbers. The constants c_α are defined for all α but only a finite number of them are non-zero. By a neighbor of the grid point ξ we mean a grid point $\xi+\alpha$ for which $c_\alpha \neq 0$. Let D_h be the set of all grid points in D and $\dot{D}_h$ be the set of all points of $\dot{D}$ that lie on the grid lines. We divide D_h into two subsets of interior and boundary grid points. The grid point ξ belongs to the interior set ($\xi \in D_h'$) if all the neighbors of ξ are in $\overline{D}$. The grid points of D_h, that do not belong to D_h', form the set D_h^* of boundary grid points. Thus $D_h = D_h' \cup D_h^*$ and let $\overline{D}_h = D_h \cup \dot{D}_h$.

We quote the following definitions from Thomée [4]:

Definition:

Let W be a neighborhood of origin in R^n. The operator L_h is said to be consistent with L if for $u \in C^{2m}(W)$,

$$L_h u_0 = Lu(0) + o(1) \text{ , as } h \to 0 .$$

Definition:

The truncation error of L_h is said to be of order N if for $u \in C^{2m+N}(W)$,

$$L_h u_0 - Lu(0) = O(h^N) \text{ , as } h \to 0.$$

Definition:

The characteristic polynomial of the operator L_h in (1.4) is the trigonometric polynomial

$$p(\theta) = \sum_\alpha c_\alpha \, e^{i\langle\alpha,\theta\rangle} , \tag{1.5}$$

where $\theta = (\theta_1, \ldots, \theta_n)$ and $\langle\alpha,\theta\rangle = \sum_{j=1}^{n} \alpha_j \theta_j$.

We shall study the characteristic polynomial $p(\theta)$ for real θ and because of its periodicity, it is sufficient to study $p(\theta)$ on the set

$$S = \{\theta \;/\; |\theta_j| \le \pi \;,\; j = 1(1)n\} \;.$$

The definitions of consistency and truncation error of the difference operator L_h can be converted into the properties of its characteristic polynomial $p(\theta)$.

Lemma 1.1 [4]: L_h is consistent with L if and only if

$$p(\theta) \;=\; L(\theta) + o\,(\,|\theta|^{2m}\,)\;,\; \theta \to 0\;.$$

The truncation error is of order N if and only if

$$p(\theta) \;=\; L(\theta) + O\,(\,|\theta|^{2m+N}\,)\;,\; \theta \to 0\;.$$

The discrete inner product and norm can be defined as:

$$(u,v)_h \;=\; h^n \sum_{\xi} u_\xi . \bar{v}_\xi \;,\; ||u||_h \;=\; (u,u)_h^{\frac{1}{2}}\;.$$

The sum will always be finite as we shall consider functions which vanish outside D_h. Let $e_j = (0, 0, ..,1, ..,0)$ be the multi-index with 1 in the j-th position and 0 elsewhere.

We define the first difference quotients

$$\partial_j u_\xi \;=\; (ih)^{-1}(u_{\xi+e_j} - u_\xi)$$

$$\bar{\partial}_j u_\xi \;=\; (ih)^{-1}(u_\xi - u_{\xi-e_j})\;.$$

For grid functions which vanish outside D_h, we have by partial summation

(1.6) $$(\partial_j u, v)_h = (u, \bar{\partial}_j v)_h \cdot$$

The discrete Sobelev norm of order m is defined as

(1.7) $$||u||_{h,m} = \left(\sum_{|\beta| \le m} ||\partial^\beta u||_h^2 \right)^{\frac{1}{2}} ,$$

where

$$\partial^\beta = \partial_1^{\beta_1} \ . \ . \ . \ \partial_n^{\beta_n} , \ \bar{\partial}^\beta = \bar{\partial}_1^{\beta_1} \ . \ . \ . \ \bar{\partial}_n^{\beta_n} .$$

The maximum norm over the region D_h can be defined as

(1.8) $$|u|_{h,D_h} = \sup_{\xi h \, \varepsilon \, D_h} |u_\xi| .$$

We require the following three lemmas. These lemmas are similar to those given by Thomée (See [4] lemmas 3.1, 3.2 and 3.3). However, the grid function u defined in these lemmas is assumed to vanish outside D_h while in the lemmas given by Thomée, they were assumed to vanish outside D_h'. The proofs of these lemmas follow on the same lines as given by Thomée.

<u>Lemma 1.2</u>: Let u be a grid function vanishing outside D_h. There exist constants C independent of u and h such that

(1.9) $$||u||_h \le C \ ||\partial_j u||_h, \ j=1 \ , \ . \ . \ . \ . \ , \ n \ ;$$

and for fixed $m \ge 1$

(1.10) $$||u||_h \le C \ ||u||_{h,m} \ .$$

Lemma 1.3: Let u be a grid function vanishing outside D_h and L_h be a difference operator of the type (1.4) with characteristic polynomial $p(\theta)$. Then,

$$(1.11) \qquad (L_h u,u)_h = h^{n-2m} (2\pi)^{-n} \int_S p(\theta) \; |\hat{u}(\theta)|^2 \, d\theta \; ,$$

where

$$(1.12) \qquad \hat{u}(\theta) = \sum_{\xi} u_\xi \, . \, e^{-i<\xi,\theta>} \; .$$

Lemma 1.4: For a grid function u vanishing outside D_h there exists a constant C, independent of u and h, such that

$$(1.13) \qquad ||u||^2_{h,m} \leq C \sum_{j=1}^{n} ||\partial_j^m u||^2_h \; .$$

We define the following norm for functions vanishing outside D_h:

$$(1.14) \qquad |||u|||_{h,m} = \left(h^n \sum_{D_h'} |u_\xi|^2 + \sum_{D_h^*} |h^{-m} u_\xi|^2 \right)^{\frac{1}{2}} \; .$$

Definition: (Property D_m^*)

The domain $\bar{D}$ has property D_m^* if there is a natural number N such that for all sufficiently small h, the following is valid: for any point $\xi \epsilon D_h^*$ consider all half rays through ξ. At least one of them contains m consecutive grid points outside D_h and within the distance Nh from ξ.

The above property is satisfied if $\dot{D}$ is sufficiently regular.

For m = 1 it is always satisfied. The following lemma connects the norms defined in (1.7) and (1.14).

Lemma 1.5: Let the domain $\bar{D}$ have property D_m^*. Then, for grid functions u vanishing outside D_h there exists a constant C independent of u and h such that

$$(1.15) \qquad |||u|||_{h,m} \le C\, ||u||_{h,m} .$$

Proof: Let ξ be a grid point of D_h^* and let the half ray in the negative direction of the x_1 - axis contain m consecutive grid points outside D_h within the distance Nh from ξ. Let $\xi - (N_1 + 1)e_1$, where $N_1 + m \le N$, be the first of the m consecutive grid points of the half ray for which u vanishes. Then, we have the following representation:

$$h^{-m} u_\xi = \sum_{j=0}^{N_1} \begin{pmatrix} m+j-1 \\ j \end{pmatrix} \bar{\partial}_1^m u_{\xi - je_1} .$$

Using Schwarz' inequality, we get

$$|h^{-m} u_\xi|^2 \le \left(\sum_{j=0}^{N_1} \begin{pmatrix} m+j-1 \\ j \end{pmatrix}^2 \right) \left(\sum_{j=0}^{N_1} \left(\bar{\partial}_1^m u_{\xi - je_1} \right)^2 \right)$$

$$\le C \sum_{j=0}^{N_1} \left(\bar{\partial}_1^m u_{\xi - je_1} \right)^2 .$$

Similar estimates hold for all other points of D_h^*, and by definition of the norms (1.7) and (1.14) we get

$$|||u|||^2_{h,m} = h^n \sum_{D_h'} |u_\xi|^2 + \sum_{D_h^*} |h^{-m} u_\xi|^2$$

$$\leq C \; ||u||^2_{h,m} ,$$

which proves the lemma.

The following lemma due to Thomée [4] makes it possible to compare difference operators in terms of their characteristic polynomials.

Lemma 1.6: Let $p_j(\theta) = \sum_\alpha c_\alpha^{(j)} e^{i<\alpha,\theta>}$, $j=1,2$ be two trigonometric polynomials such that for any trigonometric polynomial $t(\theta) = \sum_\alpha c_\alpha e^{i<\alpha,\theta>}$,

$$\int_S p_1(\theta) \; |t(\theta)|^2 \; d\theta \leq \int_S p_2(\theta) \; |t(\theta)|^2 \; d\theta .$$

Then $p_1(\theta_0) \leq p_2(\theta_0)$ for all real θ_0.

Definition:

The difference operator L_h of (1.4) is said to be elliptic [4] if its characteristic polynomial is positive definite:

$$p(\theta) > 0 \quad \text{if} \quad \theta \neq 0, \; \theta \varepsilon S.$$

In particular if L_h is elliptic, then its characteristic polynomial $p(\theta)$ is real valued for real θ and $c_\alpha = \bar{c}_{-\alpha}$.

Lemma 1.7 [4]: Let L_h be a difference operator consistent with L. L_h is elliptic if and only if there is a positive constant C such that, for all real θ,

$$(1.16) \qquad p(\theta) \ge C \sum_{j=1}^{n} (1 - \text{Cos } \theta_j)^n .$$

The following theorem relates the inner product $(L_h u,u)_h$ with the norm $||u||_{h,m}$.

Theorem 1.1: Let L_h be consistent with L. Then, L_h is elliptic if and only if there is a constant C independent of u and h such that

$$(1.17) \qquad ||u||_{h,m}^2 \le C \, (L_h u,u)_h$$

where u vanishes outside D_h.

Proof: Let L_h be elliptic, then by lemma 1.7

$$(1.18) \qquad \sum_{j=1}^{n} (1 - \text{Cos } \theta_j)^n \le C \, p(\theta) ,$$

where $p(\theta)$ is the characteristic polynomial of L_h. Now,

$$\sum_{j=1}^{n} ||\partial_j^m u||_h^2 = \sum_{j=1}^{n} (Q_h u,u)_h \, , \; Q_h = \sum_{j=1}^{n} \bar{\partial}_j^m \, \partial_j^m .$$

The characteristic polynomial of Q_h is

$$q(\theta) = \sum_{j=1}^{n} \left((e^{i\theta_j} - 1)/i \right)^m \left((1 - e^{-i\theta_j})/i \right)^m$$

$$= 2^m \sum_{j=1}^{n} (1 - \text{Cos } \theta_j)^m$$

so that $q(\theta) \le c_1 p(\theta)$, where c_1 is another positive constant independent of u or h.

From lemma 1.3, we get

$$(1.19) \qquad \sum_{j=1}^{n} ||\partial_j^m u||_h^2 = (Q_h u,u)_h \le c_1 (L_h u,u)_h$$

which together with lemma 1.4 gives the inequality (1.17). Conversely, if (1.17) is satisfied then obviously (1.19) is also satisfied. By lemmas 1.3 and 1.6, the inequality (1.18) follows. Finally, from lemma 1.7 it follows that L_h is elliptic.

We need to define a new difference operator

$$(1.20) \qquad L_{h,m} u_\xi = \begin{cases} L_h u_\xi & , \xi \in D_h' \\ h^m L_h u_\xi & , \xi \in D_h^* \\ 0 & , \xi \notin D_h \end{cases}$$

The following theorem relates the norm of u with the norm of $L_{h,m}$ u.

<u>Theorem 1.2</u>: Let the domain $\bar{D}$ have the property D_m^* and let L_h be an elliptic difference operator consistent with L. Then, for the grid functions u vanishing outside $\bar{D}_h$

$$(1.21) \qquad ||u||_{h,m} \le C ||L_{h,m} u||_h .$$

<u>Proof</u>: If u = 0 outside D_h, then

$$\begin{aligned} (L_h u,u)_h &= h^n \sum_\xi L_h u_\xi . \bar{u}_\xi \\ &= h^n \left(\sum_{D_h'} L_h u_\xi . \bar{u}_\xi + \sum_{D_h^*} L_h u_\xi . \bar{u}_\xi \right) \\ &= h^n \left(\sum_{D_h'} L_{h,m} u_\xi . \bar{u}_\xi + \sum_{D_h^*} h^{-m} L_{h,m} u_\xi . \bar{u}_\xi \right) . \end{aligned}$$

By Schwarz' inequality,

$$|(L_h u,u)_h| \le ||L_{h,m}\, u||_h \cdot |||u|||_{h,m}$$

From theorem 1.1 and lemma 1.5 we get

$$||u||^2_{h,m} \le C\ |(L_h u,u)_h| \le C\ ||L_{h,m}\, u||_h \cdot |||u|||_{h,m}$$

$$\le C\ ||L_{h,m}\, u||_h \cdot ||u||_{h,m}$$

Thus,

$$||u||_{h,m} \le C\ ||L_{h,m}\, u||_h$$

for the grid functions u vanishing outside D_h.

1.3 Finite Difference Approximation

We can now formulate the finite difference analogue of the Dirichlet problem (1.3). For the interior mesh points D_h' which have all their neighbors in $\bar{D}_h$, we define

$$(1.22) \qquad L_h\, u_\xi = F_\xi \quad , \quad \xi \varepsilon D_h' \ .$$

We assume that the truncation error of L_h is of order k. The mesh points in D_h^* have at least one neighbor outside $\bar{D}_h$. The values of u_ξ at these external points can be extrapolated using the boundary data and the values of u_ξ inside D_h. We formally form the operator $L_h u_\xi$ at the points of D_h^* by inserting these extrapolated values. In this way, we get an expression of the form $L_h u_\xi \equiv \bar{L}_h u_\xi - \ell_h(f)$ where the operator $\bar{L}_h u_\xi$ contains the terms with u_ξ, $\xi \varepsilon D_h$ and $\ell_h(f)$ is a linear function of the boundary data. Thus, on the grid points of D_h^* we prescribe

$$L_h u_\xi \equiv \bar{L}_h u_\xi - \ell_h(f) = F_\xi \ , \ \xi \in D_h^* \ ,$$

or,

$$\bar{L}_h u_\xi = F_\xi + \ell_h(f) \ , \ \xi \in D_h^* \ , \text{ and} \tag{1.23}$$

$$u_\xi = f_\xi \ , \ \xi \in \dot{D}_h \ . \tag{1.24}$$

We assume that the truncation error of the difference operator L_h thus formed on D_h^* is of order ℓ. Let $e_h = u - u_h$ denote the error function, then $e_{h,\xi}$ satisfies

$$\begin{aligned} L_h\, e_{h,\xi} &= O(h^k) \ , \ \xi \in D_h' \ ; \\ L_h\, e_{h,\xi} &= O(h^\ell) \ , \ \xi \in D_h^* \ ; \\ e_{h,\xi} &= 0 \qquad , \ \xi \notin D_h \ . \end{aligned} \tag{1.25}$$

The operator $L_{h,m}$ can be formed as follows:

$$L_{h,m}\, e_{h,\xi} = \begin{cases} L_h\, e_{h,\xi} = O(h^k) & , \ \xi \in D_h' \\ h^m\, L_h\, e_{h,\xi} = O(h^{\ell+m}) & , \ \xi \in D_h^* \\ 0 & , \ \xi \notin D_h \ . \end{cases} \tag{1.26}$$

The following theorems prove the existence and uniqueness of the discrete solution and give estimates of discretization error.

<u>Theorem 1.3</u>: Let the difference operator L_h be consistent with L and

elliptic. The discrete Dirichlet problem (1.22) - (1.24) has exactly one solution for arbitrary f and F.

Proof: The difference equations (1.22), (1.23) form a linear system of equations with the same number of equations as the number of unknowns. The discrete Dirichlet problem has a unique solution if the homogeneous form of this linear system has only the trivial solution $u_\xi = 0$. We thus want to find u_ξ such that $L_h u_\xi = 0$ in D_h. From Theorem 1.1, $||u||_{h,m} = 0$, which proves the result.

Theorem 1.4: Let the domain $\bar{D}$ have the property D_m^* and let L_h be an elliptic difference operator, consistent with L. Let the truncation error of L_h be of order k at the points of D_h' and of order ℓ at the points of D_h^*. The error function e_h satisfies the following estimate

$$(1.27) \qquad ||e||_{h,m} = O(h^\alpha) \ , \ \alpha = \min\,(k,\ \ell + m + \tfrac{1}{2}) \text{ as } h \to 0 \ .$$

Proof: Since the error function $e_{h,\xi}$ is only defined in $\bar{D}_h$ and is zero on the boundary $\dot{D}_h$, we have $e_{h,\xi} = 0$, $\xi \notin D_h$. From theorem 1.2, $||e_h||_{h,m} \leq C\ ||L_{h,m}\, e_h||_h$. The operator $L_{h,m}$ is defined in (1.26) and with that definition

$$||L_{h,m}\, e_h||_h^2 = h^n \sum_\xi (L_{h,m}\, e_{h,\xi})^2$$

$$= h^n \sum_{D_h'} (L_{h,m}\, e_{h,\xi})^2 + h^n \sum_{D_h^*} (L_{h,m}\, e_{h,\xi})^2$$

$$= h^n \sum_{D_h'} (O(h^k))^2 + h^n \sum_{D_h^*} (O(h^{\ell+m}))^2$$

The number of mesh points in the set D_h' is $O(h^{-n})$ and the number of mesh points in D_h^* is $O(h^{-n+1})$, hence

$$||L_{h,m}\, e_h||_h^2 \leq O\,(h^{2k}) + O\,(h^{2\ell + 2m + 1}).$$

It follows that

$$||e_h||_{h,m} \leq C\, ||L_{h,m}\, e_h||_h \leq O\,(h^{\alpha}),\ \alpha = \min\,(k,\ \ell+m+\tfrac{1}{2}) \text{ as } h \to 0.$$

Theorem 1.5: If the conditions of theorem 1.4 are satisfied, then

(1.28) $$|e_h|_{h,D_h} = O\,(h^{\alpha}),\ \alpha = \min\,(k,\ \ell+m+\tfrac{1}{2}) \quad \text{as} \quad h \to 0\ .$$

Proof: From the discrete Sobolev inequality [40]

$$\max_{\xi \varepsilon D_h} |u_{\xi}| \leq C\, ||u||_{h,p}\ ,\ p = \left[\frac{n}{2}\right] + 1\ .$$

If $m \geqq p$, then for the error function we have

$$|e_h|_{h,D_h} \leq C\, ||u||_{h,m}$$

and from theorem 1.4 the estimate (1.28) follows.

1.4 Applications

Zlámal [5] has shown by means of the following example that the error estimate of Thomée's approximation can not be improved. In fact, Thomée himself has stated that the reason for such a large error estimate is the crude approximation at the boundary. The example consists of the Dirichlet problem

$$(1.29)\qquad \begin{aligned} &Lu \equiv \Delta\Delta u = F && \text{in } D,\\ &u = 0,\ u_n = 0 && \text{on } \dot{D}, \end{aligned}$$

where D is the square $0 < x,y < 1$ and

$F(x,y) = 24\ [x^2(1-x)^2 + y^2(1-y)^2] + 8(6x^2-6x+1).\ (6y^2-6y+1).$

The exact solution is $u(x,y) = [x(1-x)y(1-y)]^2$.

The set D_h^* consists of the mesh points inside the square D at a distance h from the boundary $\dot{D}$. The difference approximation of Thomée [4] is as follows:

$$L_hU = \Delta_h\Delta_h\, U(P) = F(P)\ ,\ P \in D_h'$$
$$U(P) = 0\ ,\ P \in D_h^* \cup \dot{D}_h\ .$$

The discretization error $e = u-U$ for this approximation satisfies

$$||e||_{h,2}^2 \geq h^2\ \Sigma\ (e_{x\bar{x}}^2 + e_{y\bar{y}}^2) \geq h^{-2} \sum_{D_h^*} e^2 \geq Ch\ .$$

Thus, for the above example $||e||_{h,2} \geq C_1h^{\frac{1}{2}}$.

As seen from theorem 1.4, this order can be significantly improved by defining the difference operator at every mesh point of D_h.

Zlámal [5] considered the following Dirichlet problem

$$(1.30)\qquad \begin{aligned} Lu(P) &= F(P)\ ,\ P \in D\\ D^\nu u(P) &= D^\nu f(P)\ ,\ P \in \dot{D}\ ,\ \nu = 0,1\ ; \end{aligned}$$

where $Lu \equiv (a(x,y)\ u_{xx})_{xx} + 2(b(x,y)\ u_{xy})_{xy} + (c(x,y)\ u_{yy})_{yy}$ and L is assumed to be a uniformly elliptic operator. The difference operator defined by Zlámal is

$$L_h U = (aU_{x\bar{x}})_{x\bar{x}} + (bU_{xy})_{\bar{x}\bar{y}} + (bU_{\bar{x}\bar{y}})_{xy} + (cU_{y\bar{y}})_{y\bar{y}}$$

which has a truncation error of order h^2. The difference approximation is defined as

$$L_h U(P) \equiv \begin{cases} L_h\, U(P) = F\ (P)\ , & P \in D_h' \\ \bar{L}_h\, U(P) - \ell_h(f) = F\ (P)\ , & P \in D_h^* \\ U(P) = f\ (P)\ , & P \in \dot{D}_h \end{cases} .$$

The operator $L_h\, U(P)$ is formed formally at the points of D_h^* using third order extrapolation formulae for the neighbors of P outside $\bar{D}_h$. The local truncation error of L_h at D_h^* is of order h^{-1}.

In case when L has constant coefficients, theorem 1.4 can be applied with $k = 2$, $\ell = -1$ and $m = 2$. The discretization error satisfies the estimate

$$||u - U||_{h,2} = ||e||_{h,2} = O\ (h^{3/2})\ , \quad h \to 0.$$

which is the same as that obtained by Zlámal [5].

In the case of a rectangular domain, Zlámal [5] prescribed a higher order difference approximation for (1.30). He used a fourth order extrapolation formula for the external neighbors of mesh points in D_h^*, resulting in a truncation error of O(1) at points of D_h^*. In the case of constant coefficients this result follows from theorem 1.4 with $k = 2$, $\ell = 0$, and $m = 2$ yielding $||e||_{h,2} = O(h^2)$, $h \to 0$. The error estimate of $O(h^{3/2})$ can also be improved in the case of a general domain by taking a higher order extrapolation formula.

We consider (1.30) with

$$\text{(1.31)} \qquad Lu = au_{xxxx} + 2bu_{xxyy} + cu_{yyyy} .$$

At the mesh points of D_h' we define the difference operator

$$L_h U = aU_{\bar{x}x\bar{x}x} + 2bU_{\bar{x}x\bar{y}y} + cU_{\bar{y}y\bar{y}y} .$$

For the mesh points in D_h^*, at least one neighbor lies outside $\bar{D}_h$. If $(x,y) \varepsilon D_h^*$ and $(x-2h,y) \notin \bar{D}_h$ with the boundary $\dot{D}$ intersecting in $(x-\alpha h,y)$, $0 < \alpha \le 2$, we define

$$\text{(1.32a)} \qquad U(x-2h,y) = \frac{3(\alpha-2)^2}{\alpha^2} U(x,y) - \frac{2(\alpha-2)^2}{(\alpha+1)^2} U(x+h,y)$$

$$+ \frac{6(3\alpha^2 - 2\alpha - 2)}{\alpha^2(\alpha+1)^2} f(x-\alpha h,y) + \frac{6(\alpha-2)}{\alpha(\alpha+1)} h f_x(x-\alpha h,y) .$$

If $(x-h,y)$ also lies outside $\bar{D}_h$, then with $0 < \alpha \le 1$, we define

$$\text{(1.32b)} \qquad U(x-h,y) = \frac{2(1-\alpha)^2}{\alpha^2} U(x,y) - \frac{(1-\alpha)^2}{(1+\alpha)^2} U(x+h,y)$$

$$+ \frac{2(3\alpha^2 - 1)}{\alpha^2(1+\alpha)^2} f(x-\alpha h,y) - \frac{2(1-\alpha)}{\alpha(1+\alpha)} h f_x(x-\alpha h,y) .$$

The above approximations represent extrapolations of order h^4. Similar extrapolation formulae can be used for all the mesh points that lie outside $\bar{D}_h$. At the mesh points of D_h^*, we formally form the difference operator $L_h U$ replacing the unknown functional values with their extrapolation formulae. The local truncation error of L_h is of order h^2 at the points of D_h' and of order 1 at points of D_h^*. From theorem 1.4 we get the following result with k = 2, ℓ = 0, m = 2

$$||e_h||_{h,2} = O\ (h^2)\ ,\ h \to 0\ .$$

Using theorem 1.5, we get

$$|e_h|_{\bar{D}_h} = O\ (h^2),\ h \to 0\ . \tag{1.33}$$

The above result is true for the fourth order elliptic differential equations of the type (1.31) and for any two-dimensional bounded domain. This result holds for the biharmonic equation since the biharmonic equation is a special case of (1.31) with a = c = b = 1. We also note that an $O(h^2)$ approximation in the maximum norm for the Dirichlet problem of the biharmonic equation in a general domain is the best which has so far been obtained.

Bramble [6] also discussed the Dirichlet problem for the biharmonic equation in regions with a piecewise smooth boundary. He defined the usual 13-point difference operator at the points of D_h' and constructed an approximate operator with truncation error of order h^{-1} at the points of D_h^*. He obtained the following results:

$$||e||_{(2p)} = (h^2 \sum_\xi e_\xi^{2p})^{\frac{1}{2p}} \leq O\ (h^2)\ ,\ p \geq 1\ ,\ h \to 0\ ;$$

$$||e|| + ||\delta e|| \leq O\ (h^2)\ ,\ h \to 0 \tag{1.34a}$$

where

$$||e||^2 = h^2 \sum_\xi e_{h,\xi}^2 ,$$

$$||\delta e||^2 = h^2 \sum_\xi (\partial_1 e_\xi^2 + \partial_2 e_\xi^2).$$

The grid function e_ξ is assumed to vanish outside D_h. Bramble also obtained the following result

(1.34b) $$|e|_{h,D_h} \leq Ch^2|\log h^{-1}|^{\frac{1}{2}} , \ h \text{ small}.$$

We can apply theorem 1.4 with $k = 2$, $\ell = -1$, $m = 2$ and obtain the estimate

(1.35a) $$||e||_{h,2} \leq O(h^{3/2}) , \ h \to 0 .$$

Using theorem 1.5 we obtain the following result

(1.35b) $$|e|_{h,D_h} \leq O(h^{3/2}) , \ h \to 0 .$$

We note that the results (1.34) obtained by Bramble follow from (1.35) on applying a lemma given by Bramble (lemma 3.3 of [6]).

Finally, we consider the following Dirichlet problem with constant coefficients and homogeneous boundary conditions:

(1.36) $$Pu \equiv -\sum_{j,k=1}^{n} a_{jk} \partial^2 u/\partial x_j \partial x_k = F(x), \ x = (x_1, \ldots , x_n) \in D ,$$

$$a_{jk} = a_{kj} \text{ and the matrix } (a_{jk}) \text{ is positive definite.}$$

The operator P can be approximated by the following difference operator [7]:

$$P_h u \equiv \sum_j a_{jj}\, \partial_j\, \bar{\partial}_j u + \tfrac{1}{4} \sum_{j \neq k} a_{jk}\, (\partial_j + \bar{\partial}_j)\, (\partial_k + \bar{\partial}_k)\, u\ .$$

Thomée [4] showed that P_h is an elliptic difference operator and that his results can be applied to this example giving a discretization error of order $h^{\frac{1}{2}}$ in general, i.e., $||e||_{h,1} = O(h^{\frac{1}{2}})$, $h \to 0$. He also showed that for certain domains made up of rectangular regions (so that $D_h^* \subset \dot{D}_h$), the error estimate $||e||_{h,1} = O(h^2)$ is valid. Bramble [8] also discussed this problem and tried to estimate the discretization error in $||\cdot||_{h,0}$ norm in such a way that the approximations near the boundary are not so important. He proved the estimate $||e||_{h,0} = O(h)$, $h \to 0$. Bramble also stated that to his knowledge no second order approximations have been proved in general for this example.

We shall show that by defining a difference operator at each mesh point of the domain D_h, a better estimate can be obtained in general. We note that all the neighbors of a grid point ξ are the points $\xi + \alpha$, where $\alpha_j = 0, \pm 1$, $j = 1, \ldots, n$. If the point ξ is in the set D_h^*, then at least one of its neighbors lies outside $\bar{D}_h$. We assume that $\xi - e_1$ is such a neighbor and the point $\xi - \beta e_1$, $0 < \beta \le 1$, lies on the boundary $\dot{D}$. We use the following extrapolation formula for $u_{\xi - e_1}$:

$$(1.37)\quad u_{\xi-e_1} = \frac{2\ (\beta-1)}{\beta}\, u_\xi + \frac{2}{\beta\ (\beta+1)}\, u_{\xi-\beta e_1} + \frac{1-\beta}{1+\beta}\, u_{\xi+e_1}\ ,\ 0 < \beta \le 1\ .$$

Similar extrapolations can be used for all those points that lie outside $\bar{D}_h$ and are the neighbors of some points in D_h^*. The difference operator P_h can now be formally formed at all mesh points of D_h. The truncation error of P_h is of order h^2 at the points of D_h'. Since the extrapolation (1.37) incurs an error of order h^3, the truncation error of P_h at the points of D_h^* is of order h. The theorem 1.4 can now be applied with $k = 2$, $\ell = 1$, $m = 1$ and we obtain the following estimate

$$(1.38) \qquad ||e||_{h,1} = ||u-u_h||_{h,1} \leq O(h^2), \ h \to 0 .$$

We have thus proved the second order error estimate for the Dirichlet problem (1.36) in a general domain. The result (1.38) is the best obtainable for the second order difference operator P_h.

CHAPTER 2

STABILITY AND CONVERGENCE OF ITERATION SCHEMES

2.1 Preliminaries

In this chapter we consider a class of iteration schemes which can be written in the form

$$(2.1)\qquad B\,u^{(n+1)} = B\,u^{(n)} - \tau\,(A\,u^{(n)} - f)\;,$$

where A and B are certain linear operators, τ an iteration parameter and n the iteration number. The scheme (2.1) can be used to obtain the approximate solution at the (n+1)th level using the approximation at the nth level and is called a two-level (or two-layer) iteration scheme. When B is an identity operator, (2.1) is called an explicit scheme.

The iteration scheme (2.1) can be used to obtain the solution u of a system of algebraic equations written in the matrix form:

$$(2.2)\qquad A\,u = f.$$

The difference equations obtained by the discretization of elliptic equations can be put in the form (2.2). The iteration scheme (2.1) can also arise from the discretization of the abstract Cauchy problem

$$(2.3)\qquad \frac{\partial u}{\partial t} + L\,u = f\;;\; u(0) = u_0\;,\; 0 \le t \le t_0\;,$$

where L is a differential operator in the space variables.

The iteration number n in (2.1) can be considered as representing the time level t and the following notation can be introduced

$$u^{(n)} = y(t) \equiv y \ , \ u^{(n+1)} = y(t+\tau) \ , \ t = n\tau \ , \tag{2.4}$$

$$\frac{y(t+\tau) - y(t)}{\tau} = y_t \ .$$

It may be noted that y_t represents a difference quotient throughout this chapter. The equation (2.1) reduces to the form

$$B(t)\, y_t + A(t)\, y = f(t),\ y(0) = y_0,\ 0 \le t < t_0 \ , \tag{2.5}$$

which is called the canonical form of the two-level iteration scheme (2.1) [9]. We shall use the same notation as used by Samarskii [9] and consider the iteration scheme (2.5) as an operator equation in a real Hilbert space H where an inner product (,) and a norm $||\cdot||$ are defined related by $||y|| = \sqrt{(y,y)}$. The space H depends upon a parameter h which is a vector in some normed space. The vectors y and f belong to H; A(t) and B(t) are linear operators dependent upon the parameters t, τ and h and map H into itself for each $t \in \omega_\tau$, where ω_τ is the set given by

$$\omega_\tau = \{t = j\tau \ , \ j = 1 \ , \ 2 \ , \ . \ . \ . \ , \ j_0 \ ; \ j_0 = t_0/\tau\} \ .$$

We shall need the following definitions:

<u>Definitions</u> [10]:

Let u, v be arbitrary vectors of H and A, B be linear operators mapping H into itself. Then,

$$A \geq B \text{ if } (Au,u) \geq (Bu,u) \;\forall u \in H ;$$

$$A \text{ is selfadjoint } (A = A^*) \text{ if } (Au,v) = (u,Av) ;$$

$$\text{positive } (A > 0) \text{ if } (Au,u) > 0 , u \neq 0 ;$$

$$\text{positive definite } (A \geq \delta E) \text{ if } (Au,u) \geq \delta(u,u) , \delta > 0 ,$$

where E is the identity operator.

If $B = B^* > 0$, then the square root $B^{\frac{1}{2}}$ exists and $B^{\frac{1}{2}} = (B^{\frac{1}{2}})^* > 0$.

Definition [9]: A positive operator $A = A(t)$, dependent upon $t \in \omega_\tau$, is called Lipschitz continuous in t if

$$(2.6) \qquad | (A(t) u,u) - (A(t-\tau) u,u) | \leq \tau c_2 (A(t-\tau) u,u) ,$$

where c_2 is a positive operator.

Definition [10]: The norm of the operator A is defined as

$$||A|| = \sup_{x \neq 0} \frac{||Ax||}{||x||} , x \in H .$$

If A is selfadjoint, then

$$||A|| = \sup_{x \neq 0} \frac{|(Ax,x)|}{(x,x)} , x \in H .$$

Definition [11]: If D(t) is a positive linear operator on H, then the following energy norm can be defined

$$||y||_{d(t)} = (D(t) y,y)^{\frac{1}{2}} ,$$

where the lower case letter d(t) relates to the operator D(t).

We shall now consider the following special cases of the operator A when it is i) selfadjoint, ii) nonselfadjoint, and iii) A is time dependent. Most of the results for the selfadjoint cases are due to Samarskii. We have briefly reproduced them in order to motivate our study of the nonselfadjoint cases.

2.2 Selfadjoint Operator

When A(t) is a selfadjoint positive operator, Lipschitz continuous in $t \in \omega_\tau$; and B(t) is a positive operator then the equation (2.5) is said to belong to the initial family of two layer schemes (IF - 2) according to Samarskii [9]. For such schemes Samarskii gave some conditions of stability which are sufficient. Before we quote these results, we give the following definitions:

Definition [12]: The initial value problem (2.5) is said to be properly posed if there exists τ_0 such that for $\tau \le \tau_0$, a solution of (2.5) exists for arbitrary $y_0 \in H$ and $f(t) \in H$, $t \in \omega_\tau$.

Definition [12]: The iteration scheme (2.5) is stable with respect to the initial data and with respect to the right hand side if

$$||y(t)||_{(1)} \le M_1||y(0)||_{(1)} + M_2 \max_{\omega_\tau} ||f(t)||_{(2)} ,$$

where $||\cdot||_{(1)}$ and $||\cdot||_{(2)}$ are certain norms on H and M_1, M_2 are positive constants, independent of τ and h. For $f(t) \equiv 0$, the above condition defines the stability with respect to the initial data.

Samarskii [9] proved the following results:

Theorem 2.1: If the scheme (2.5) belongs to IF - 2 and the condition

$$B \geq 0.5\,\tau\,(1 - c_1\,\tau)\,A \tag{2.7}$$

is satisfied with a positive constant c_1 independent of τ and h, then the scheme (2.5) is stable for $\tau \leq \tau_0$, $\tau_0 < 1/4c_1$.
Moreover, the solution y of (2.5) satisfies the following estimate:

$$||y(t+\tau)||_{a(t)} \leq M_1||y(0)||_{a(0)} + M_2 \max_{0\leq t'\leq t} ||f(t')||_{a^{-1}(t')}$$

$$+ M_3 \max_{0\leq t'\leq t} ||f_{\bar{t}}(t')||_{a^{-1}(t')} ,$$

where M_i are positive constants depending only upon c_1, c_2 and t_0.

Theorem 2.2: If (2.5) belongs to IF - 2 and the condition

$$B \geq \varepsilon\, E + 0.5\,\tau\, A\ , \quad \varepsilon > 0 \tag{2.8}$$

is satisfied, then the solution of (2.5) satisfies the following estimate

$$||y(t+\tau)||_{a(t)} \leq M_1||y(0)||_{a(0)} + \frac{1}{\sqrt{\varepsilon}} M_2 \max_{0\leq t'\leq t} ||f(t')|| \ .$$

Remark: The results of theorem 2.2 hold if $A = A_0 + A_1$, where $A_0(t)$ is a selfadjoint, positive operator, Lipschitz continuous in t and $A_1(t)$ is a nonselfadjoint operator satisfying

$$||A_1(t)y||^2 \leq c_3(A_0(t)y,y) ,$$

where c_3 is a positive constant independent of τ and h.

Further, Samarskii [9] proved that the condition (2.7) is also necessary for the stability of (2.5) in the case when A and B do not depend upon t and B is selfadjoint. In this case the operators A and B are called "fixed" operators. If the operator B is also positive, then $B^{\frac{1}{2}}$ exists and the equation (2.5) can be written as

$$B^{\frac{1}{2}} B^{\frac{1}{2}} y_t + A y = f ,$$

which on putting $x = B^{\frac{1}{2}} y$ becomes

$$B^{\frac{1}{2}} x_t + A B^{-\frac{1}{2}} x = f \text{ ; or,}$$

(2.9) $$x_t + Cx = \phi , \; C = B^{-\frac{1}{2}} A B^{-\frac{1}{2}} , \; B^{-\frac{1}{2}} f = \phi;$$

and $$||x|| = (B^{\frac{1}{2}}y , B^{\frac{1}{2}}y)^{\frac{1}{2}} = (By,y)^{\frac{1}{2}} = ||y||_b .$$

Thus, the norm of x is equivalent to the B-norm of y and the stability of (2.9) is equivalent to the stability of (2.5) in B-norm. First of all, we shall discuss the stability of (2.9) with respect to the initial data. The homogeneous form of the equation (2.9) can be written as

(2.10) $$x(t+\tau) = S\,x(t) , \; S = E - \tau C ,$$

$$0 \leq t = n\tau < t_0 , \; x(0) = x_0 \in H .$$

The initial value problem (2.10) is stable [11] if the powers $S(\tau)^n$ of the transition operator $S(\tau)$ remain uniformly bounded for all

$\tau \in (0, \bar{\tau})$. Thus (2.10) is stable if there exists a constant c_0, independent of τ and the space parameter h, such that for $0 < \tau < \bar{\tau}$ the following bound is valid:

$$(2.11) \qquad ||x(t)|| \le \exp\,(c_0 t)\,||x(0)|| \ , \ 0 \le t = n\tau < t_0 \ ;$$

where x(0) is an arbitrary initial vector of H.

Samarskii [13] proved the following lemmas. Lemmas 2.1 is also applicable to nonselfadjoint operators.

Lemma 2.1: If the transition operator S is independent of t, then the condition

$$(2.12) \qquad ||S|| \le \rho \ , \ \rho = e^{c_0 \tau}$$

with c_0 independent of τ and h, is necessary and sufficient for the stability of (2.10).

Lemma 2.2: If $C = C^*$, then the conditions

$$(2.13) \qquad (1-\rho)\, E \ \le \ \tau C \ \le \ (1+\rho)\, E \ , \ \rho > 0$$

are necessary and sufficient for the bound (2.12).

Lemma 2.3: If A and B are positive selfadjoint operators, then the conditions

$$(2.14) \qquad (1-\rho)\, B \ \le \ \tau A \ \le \ (1+\rho)\, B \ , \ \rho = e^{c_0 \tau}$$

with any constant c_0 are necessary and sufficient for the stability of (2.5) in B-norm with respect to the initial data.

2.3 Nonselfadjoint Operator

When A is a positive nonselfadjoint operator, C is also nonselfadjoint and positive. In this case C^{-1} exists and Samarskii [13] proved the following two results:

(2.15a) For $\rho \geq 1$, $C^{-1} \geq (\tau / 1+\rho) E \Rightarrow ||S|| \leq \rho$;

(2.15b) For $\rho \leq 1$, $||S|| \leq \rho \Rightarrow C^{-1} \geq (\tau / 1+\rho) E$.

We also note that for $\rho \leq 1$

(2.15c) $$C^{-1} \geq (\tau / 1+\rho) E \Leftrightarrow ||S|| \leq 1 .$$

The conditions (2.15a) and (2.15c) are sufficient conditions for the stability of the explicit scheme (2.10), whereas (2.15b) is a necessary condition. Each of these conditions involves the operator C^{-1} which requires a knowledge of A^{-1}. However, the operator A^{-1} is rarely known a priori and the conditions (2.15) can not be utilized in practice. We now derive some conditions which can be directly checked in terms of the known operators.

If the system (2.10) is stable, then from lemma 2.1,

$$||S|| \leq \rho \Rightarrow \frac{||Sx||}{||x||} \leq \rho \ , \ \rho > 0 \ , \ 0 \neq x \in H .$$

Since $$||Sx||^2 = ||(E-\tau C)\ x||^2$$

$$= ||x||^2 - 2\ \tau(Cx,x) + \tau^2||Cx||^2 ,$$

and $(Cx,x) \le ||Cx||\cdot||x||$,

therefore

$$||x||^2-2\tau||Cx||\cdot||x|| + \tau^2||Cx||^2 \le ||Sx||^2 \le \rho^2||x||^2 ,$$

or,

$$\Big|\, ||x|| - \tau||Cx|| \,\Big| \le \rho||x|| ,$$

$$\frac{1-\rho}{\tau}\, ||x|| \le ||Cx|| \le \frac{1+\rho}{\tau}\, ||x||.$$

Since this is true for all $x \in H$, it follows that

$$||C|| = \sup_{x\neq 0} \frac{||Cx||}{||x||} \le \frac{1+\rho}{\tau} , \quad \rho > 0 . \tag{2.16}$$

The inequality (2.16) provides a necessary condition for the stability of (2.10).

A sufficient condition for stability can be obtained as follows:

$$||Sx||^2 = ||x||^2 - 2\,\tau\,(Cx,x) + \tau^2||Cx||^2$$

$$\le ||x||^2 + 2\,\tau\,||Cx||\cdot||x|| + \tau^2||Cx||^2$$

$$\le (\,1+\tau||C||)^2||x||^2 ;$$

$$||S\,|| = \sup_{x\neq 0} \frac{||Sx||}{||x||} \le 1+\tau||C|| .$$

For the stability, from lemma 2.1, it is sufficient that

$$1 + \tau||C|| \le \rho ,\ \rho = e^{c_0\tau} ; \quad \text{or,}$$

$$\tau \le \frac{\rho - 1}{||C||} \, , \quad \rho = e^{c_0 \tau} \, . \tag{2.17}$$

This is a general sufficient condition for stability and gives the stability range which depends upon ρ and the norm of the operator C. The range given by (2.17) becomes meaningless when $\rho \le 1$. However, this range can be improved if more information is available about the operator C. We shall consider several possibilities.

When $C = C^* > 0$, $S = E - \tau C = S^*$; and

$$||S|| = \sup_{x \ne 0} \frac{|(Sx,x)|}{(x,x)} = \sup_{x \ne 0} \left| 1 - \tau \frac{(Cx,x)}{(x,x)} \right|$$

It follows that

$$||S|| \le \rho \quad \text{if and only if} \quad \tau \le \frac{1+\rho}{||C||} \, , \quad \rho \ge 1 \, . \tag{2.18}$$

In the case when C is nonselfadjoint and positive definite operator $(C \ge \delta E \, , \, \delta > 0)$, then

$$\frac{||Sx||^2}{||x||^2} = 1 - 2\tau \frac{(Cx,x)}{(x,x)} + \tau^2 \frac{||Cx||^2}{||x||^2}$$

$$\le 1 - 2\tau\delta + \tau^2 ||C||^2 \, .$$

For the stability it is sufficient from lemma 2.1 that

$$||S||^2 \le 1 - 2\tau\delta + \tau^2 ||C||^2 \le \rho^2$$

which leads to the condition

$$(2.19)\qquad 0 < \tau \le \tilde{\tau}\ ,\ \tilde{\tau} = \frac{\delta + (\delta^2 + ||C||^2\ (\rho^2 - 1))^{\frac{1}{2}}}{||C||^2}\ .$$

If the operator C is sum of a selfadjoint and a nonselfadjoint operator, then the stability range can be further improved. We shall use lemma 2.1 with $\rho = 1$ as the stability criterion in the following theorems:

Theorem 2.3: Let C_0 and C_1 be linear operators mapping H into itself and

$$(2.20)\qquad C = C_0 + C_1\ ,\ C_0 = C_0^* > 0\ ;\ C_1 \ge \delta E\ ,\ \delta > 0\ .$$

The corresponding iteration scheme (2.10) is stable provided

$$(2.21)\qquad 0 < \tau \le \tau_0\tilde{\theta}\ ,\quad \text{where } \tilde{\theta} = \frac{2\ (1 - \rho_0 + b)}{1 - \rho_0{}^2 + 2b + a^2}\ ,$$

$$a = \tau_0||C_1||\ ,\ b = \tau_0\delta \quad \text{and } \rho_0 = ||E - \tau_0\ C_0||\ ;$$

and τ_0 is any real number satisfying the conditions

$$(2.22)\qquad \rho_0 \le 1 + b\quad ,\quad (1 - \rho_0)^2 < a^2\ .$$

Proof: $\quad S = E - \tau C = E - \tau C_0 - \tau C_1$

$$= (\theta E - \tau C_0) + \{(1 - \theta)\ E - \tau C_1\}\ ,\ 0 \le \theta \le 1\ .$$

By the triangle inequality

$$||S|| \le \theta||\ E - \tau/\theta\ C_0\ || + ||(1 - \theta)\ E - \tau C_1||\ .$$

We rescale τ in terms of τ_0 and θ by writing $\tau = \tau_0\theta$, $0 \le \theta \le 1$. Moreover,

$$(2.23) \quad ||\{(1-\theta)E-\tau C_1\}x||^2 = (1-\theta)^2||x||^2-2\tau(1-\theta)(C_1x,x)+\tau^2||C_1x||^2$$

$$\le \{(1-\theta)^2-2\tau\delta(1-\theta)+\tau^2||C_1||^2\}||x||^2 \;;$$

This is true for all $x \in H$ so that

$$||(1-\theta)E-\tau C_1|| \le \{(1-\theta)^2-2\tau\delta(1-\theta)+\tau^2||C_1||^2\}^{\frac{1}{2}} ,$$

and $||S|| \le f(\theta)$, where

$$(2.24) \quad f(\theta) = \theta||E-\tau_0C_0|| + \{(1-\theta)^2-2\tau_0\theta\delta(1-\theta) + \theta^2\tau_0^2||C_1||^2\}^{\frac{1}{2}}$$

$$= \theta\rho_0 + \{(1-\theta)^2 -2b\,\theta\,(1-\theta) + a^2\,\theta^2\}^{\frac{1}{2}} \;;$$

$$a = \tau_0||C_1|| \;,\; b = \tau_0\delta \;,\; \rho_0 = ||E - \tau_0C_0|| \;.$$

The value of $f(\theta)$ is less than unity if

$$0 \le \theta \le \tilde{\theta} \;,\quad \tilde{\theta} = \frac{2\,(1-\rho_0 +b)}{1-\rho_0^2 +2b +a^2} \;.$$

The value of $\tilde{\theta}$ lies in the interval $[0,1]$ if $\rho_0 \le 1+b$, and $(1-\rho_0)^2 < a^2$.

Thus, $||S|| \le f(\theta) \le 1$ if $0 < \tau = \tau_0\theta \le \tau_0\tilde{\theta}$,

where τ_0 is any real number satisfying (2.22).

Theorem 2.4: Let $C = C_0 + C_1$ such that $C > 0$ and

(2.25) $$C_0 = \frac{1}{2}(C + C^*) \; , \; C_1 = \frac{1}{2}(C - C^*) \; .$$

A sufficient condition for the stability of the scheme (2.10) is

(2.26) $$0 < \tau \le \tau_0 \tilde{\theta} \; , \; \tilde{\theta} = \frac{2\,(1 - \rho_0)}{1 - \rho_0{}^2 + a^2} \; ;$$

$$\rho_0 = ||E - \tau_0 C_0|| \; , \; a = \tau_0 ||C_1||$$

and τ_0 is any positive real number such that $\rho_0 < 1$.

<u>Proof</u>: For any $x \in H$, $(C_1 x, x) = 0$ since $C_1 = (C - C^*)/2$.

From (2.23) we get

$$||(1-\theta)E - \tau C_1|| \le \{(1-\theta)^2 + \tau^2 ||C_1||^2\}^{\frac{1}{2}} \; ,$$

$$||S|| \le \theta \, \rho_0 + \{(1-\theta)^2 + \theta^2 \, a^2\}^{\frac{1}{2}} \; , \; \theta \ge 0 \; .$$

Proceeding as in theorem 2.3, we find that

$$||S|| \le 1 \quad \text{if} \quad 0 < \tau = \tau_0 \theta < \tau_0 \tilde{\theta} \; , \; \tilde{\theta} = \frac{2\,(1 - \rho_0)}{1 - \rho_0{}^2 + a^2} \; .$$

This holds for any positive real number τ_0 for which $\rho_0 < 1$.

<u>Theorem 2.5</u>: Let the operators C_0 and C_1 in theorem 2.3 satisfy the following conditions

(2.27) $$\gamma_1 E \le C_0 \le \gamma_2 E \; , \; C_1 \ge \delta E \; , \; ||C_1|| \le \gamma_3 \; ;$$

$$0 \le \gamma_1 \le \gamma_2 \; , \; 0 \le \delta \le \gamma_3 \; .$$

In this case the stability range is largest when $\tau_0 = 2/(\gamma_1 + \gamma_2)$ and the stability condition becomes

$$(2.28) \qquad 0 < \tau \le \frac{2\,(\gamma_1+\delta)}{(\gamma_1+\gamma_2)\,(\gamma_1+\delta) + (\gamma_3{}^2-\gamma_1{}^2)}$$

<u>Proof</u>: If $\gamma_1 E \le C_0 \le \gamma_2 E$, then

$(1-\tau_0\gamma_2)\ E \le E-\tau_0 C_0 \le (1-\tau_0\gamma_1)\ E$, and

$\rho_0 = ||E-\tau_0 C_0|| = \max\ (|1-\tau_0\gamma_1|\ ,\ |1-\tau_0\gamma_2|)$.

The stability range (2.21) is

$$0 < \tau \le \tau_0\tilde{\theta}\ ,\ \tau_0\tilde{\theta} = \frac{2\tau_0\ (1-\rho_0 + b)}{1-\rho_0{}^2 + 2b + a^2}\ ,\ a = \tau_0\gamma_3\ ,\ b = \tau_0\delta\ .$$

We wish to choose τ_0 such that $\tau_0\tilde{\theta}$ is a maximum. There are two cases:

Case 1: $\rho_0 = |1-\tau_0\gamma_1| = 1-\tau_0\gamma_1$ if either $\tau_0 < \gamma_2^{-1} \le \gamma_1^{-1}$;

or, $\gamma_2^{-1} < \tau_0 < \gamma_1^{-1}$, $\tau_0 < 2/(\gamma_1+\gamma_2)$.

In this case $\tau_0\tilde{\theta}$ is an increasing function of τ_0. Its maximum is attained when τ_0 has its maximum value which is $2/(\gamma_1+\gamma_2)$.

Case 2: $\rho_0 = |1-\tau_0\gamma_2| = \tau_0\gamma_2-1$ if either $\gamma_2^{-1} < \gamma_1^{-1} < \tau_0$;

or, $\gamma_2^{-1} < \tau_0 < \gamma_1^{-1}$, $\tau_0 > 2/(\gamma_1+\gamma_2)$.

If $\gamma_3 > \gamma_2$, then $\tau_0\tilde{\theta}$ is a decreasing function of τ_0 and the minimum permissible value of τ_0, which is $2/(\gamma_1+\gamma_2)$, gives the maximum of $\tau_0\tilde{\theta}$.

If $\gamma_3 < \gamma_2$, then we should have $\tau_0 \in \left(\frac{2}{\gamma_2+\gamma_3}, \frac{2}{\gamma_2-\delta}\right)$ in order to satisfy (2.22). In this case $\tau_0 = 2/(\gamma_2+\gamma_3)$ gives the maximum of $\tau_0\tilde{\theta}$. However, $\tau_0 = 2/(\gamma_1+\gamma_2)$ gives a larger stability range. Substituting this value of τ_0 in (2.21) we get (2.28).

Corollary 1: In the case of theorem 2.4, if $(C_1x,x) = 0$ for all $x \in H$, then the stability condition in this case is obtained by putting $\delta = 0$ in (2.28).

Corollary 2: When $C = C_0 + C_1$, $||C|| \le \gamma_2+\gamma_3$, a necessary condition of stability is obtained from (2.16) as follows:

(2.29) $$\tau \le \frac{2}{\gamma_2+\gamma_3} .$$

When $\gamma_3 \le \gamma_1$, (2.29) becomes a necessary and sufficient condition for the stability of (2.10). In the case when C is selfadjoint, $\gamma_3 = 0$ and this condition becomes $\tau < 2/\gamma_2$ which is equivalent to (2.18) with $\rho = 1$. This condition is the same as given in lemma 2.2.

Corollary 3: The function $f(\theta)$ of (2.24) assumes a minimum in the interval $(0,\tilde{\theta})$. This minimum is achieved for $\theta = \hat{\theta}$ given by

(2.30) $$\hat{\theta} = \frac{(1+b)\ (d-\rho_0^2)^{\frac{1}{2}} - \rho_0\ (a^2-b^2)^{\frac{1}{2}}}{d\ (d-\rho_0^2)^{\frac{1}{2}}} , \quad d = a^2+2b+1 .$$

The iteration parameter $\hat{\tau}$ for the optimal convergence of the iteration scheme is given by $\hat{\tau} = \tau_0 \hat{\theta}$, $\tau_0 = 2/(\gamma_1+\gamma_2)$. The norm of the transition operator S of the iteration scheme (2.10) satisfies

$$(2.31) \qquad ||S|| \le \hat{\rho} \ , \ \hat{\rho} = \frac{\rho_0(1+b) + (a^2-b^2)^{\frac{1}{2}} (d-\rho_0{}^2)^{\frac{1}{2}}}{d} < 1 \ .$$

When $\delta = 0$, $b = 0$ and we obtain the values of $\hat{\tau}$ and $\hat{\rho}$ for the optimum convergence of the iteration scheme (2.10) as given by Samarskii [14].

One can verify the conditions of theorem 2.5 in terms of the norms of A and B by using the following lemma:

Lemma 2.4: Let $A = A_0 + A_1$ in the iteration scheme (2.5) and the following conditions be satisfied

$$(2.32) \qquad \gamma_1 B \le A_0 \le \gamma_2 B \ , \ \alpha_1 E \le B \le \alpha_2 E \ ,$$

$$||A_1|| \le \beta \ , \ A_1 \ge \alpha E \ .$$

Then, $\gamma_1 E \le C_0 \le \gamma_2 E$, $C_1 \ge \delta E$ and $||C_1|| \le \gamma_3$,

with $\delta = \alpha/\alpha_2$, $\gamma_3 = \beta/\alpha_1$.

Proof: Since $A_0 = A_0^*$ and $C_0 = B^{-\frac{1}{2}} A_0 B^{-\frac{1}{2}}$, from lemma 2.3 the following inequalities are equivalent:

$$\gamma_1 B \le A_0 \le \gamma_2 B \quad \text{and} \quad \gamma_1 E \le C_0 \le \gamma_2 E \ .$$

Moreover, since $B = B^* > 0$, B^{-1} exists and the following inequalities are equivalent:

$$\alpha_1 E \leq B \leq \alpha_2 E \quad , \quad \alpha_2^{-1} E \leq B^{-1} \leq \alpha_1^{-1} E \; .$$

Now,

$$(C_1 x, x) = (B^{-\frac{1}{2}} A_1 B^{-\frac{1}{2}} x, x) = (A_1 B^{-\frac{1}{2}} x, B^{-\frac{1}{2}} x)$$

$$\geq \alpha (B^{-\frac{1}{2}} x, B^{-\frac{1}{2}} x) = \alpha (B^{-1} x, x)$$

$$\geq \alpha/\alpha_2 (x,x) \; ;$$

$$||C_1 x||^2 = (B^{-1} A_1 B^{-\frac{1}{2}} x, A_1 B^{-\frac{1}{2}} x)$$

$$\leq \alpha_1^{-1} ||A_1 B^{-\frac{1}{2}} x||^2$$

$$\leq \alpha_1^{-1} ||A_1||^2 ||B^{-\frac{1}{2}} x||^2$$

$$\leq \beta^2 / \alpha_1^2 \; ||x||^2$$

so that $||C_1|| \leq \beta/\alpha_1$.

2.4 Time-Dependent Operators

In order to study the case when A and B depend upon t, we assume that B(t) is a selfadjoint and positive operator and satisfies a Lipschitz condition in t:

$$(2.33)\qquad |\,(\{B(t) - B(t-\tau)\}\, y,y\,)\,| \le c_1\, \tau\, (\,B(t-\tau)\, y,y\,)\ ,$$

$$0 < t = n\tau < t_0\ ,\ y \in H\ ,$$

where c_1 is a positive constant independent of τ and h.

Since $B(t) = B^*(t) > 0$, the square root $B^{\frac{1}{2}}(t)$ exists for each $t \in \omega_\tau$.

We put $B^{\frac{1}{2}}(t-\tau)\, y = z$ in (2.33) and obtain

$$|\,(\,B^{-\frac{1}{2}}(t-\tau)\, B(t)\, B^{-\frac{1}{2}}(t-\tau)\, z,z\,) - (z,z)\,| \le c_1\, \tau\, (z,z)$$

$$|\,(\,B^{-\frac{1}{2}}(t-\tau)\, B(t)\, B^{-\frac{1}{2}}(t-\tau)\, z,z\,)\,| \le (1 + c_1\tau)\, (z,z)$$

$$||B^{\frac{1}{2}}(t)\, B^{-\frac{1}{2}}(t-\tau)\, z||^2 \le e^{c_1\tau}\, ||z||^2\ ,\ z \in H$$

$$(2.34)\qquad ||B^{\frac{1}{2}}(t)\, B^{-\frac{1}{2}}(t-\tau)|| \le e^{c_1\,\tau/2}\ ,\ t \in \omega_\tau\ .$$

We wish to consider the stability of the initial value problem

$$(2.5')\qquad B(t)\,\frac{y(t+\tau) - y(t)}{\tau} + A(t)\, y(t) = 0\ ,$$

$$y(0) = y_0\ ,\ t \in \omega_\tau\ .$$

The operator equation (2.5′) can be written as

$$(2.35)\qquad x(t+\tau) = S(t)\, \bar{x}(t)\ ,\ S(t) = E - \tau C(t)\ ,$$

where $x(t+\tau) = B^{\frac{1}{2}}(t)\, y(t+\tau)$,

$\bar{x}(t) = B^{\frac{1}{2}}(t)\, y(t)$, and $C(t) = B^{-\frac{1}{2}}(t)\, A(t)\, B^{-\frac{1}{2}}(t)$.

Using the estimate (2.34) we get

$$||\bar{x}(t)|| = ||B^{\frac{1}{2}}(t)\ y(t)||$$

$$= ||B^{\frac{1}{2}}(t)\ B^{-\frac{1}{2}}(t-\tau)\ x(t)||$$

$$\leq ||B^{\frac{1}{2}}(t)\ B^{-\frac{1}{2}}(t-\tau)||\cdot||x(t)||$$

$$\leq \exp\ (c_1\ \tau/2)\ ||x(t)||\ ,\ t\ \varepsilon\ \omega_\tau\ .$$

The following lemma is a counterpart of lemma 2.1 but gives a sufficient condition of stability:

Lemma 2.5: The initial value problem (2.5′) with operators depending upon t is stable if

$$(2.36) \qquad ||S(t)|| \leq e^{c_0\tau} = \rho\ ,\ t\ \varepsilon\ \omega_\tau$$

where c_0 is a real number.

Proof:

$$||x(t+\tau)|| = ||S(t)\ \bar{x}(t)||$$

$$\leq ||S(t)||\cdot||\bar{x}(t)||$$

$$\leq \exp\ (\ (c_0 + c_1/2)\ \tau\ )\ ||x(t)||\ ,\ t\ \varepsilon\ \omega_\tau\ .$$

This demonstrates the stability of (2.35) and hence that of (2.5′).

When A(t) is selfadjoint, Samarskii [13] obtained the following conditions, the first of which is sufficient for the stability and the second is sufficient for the instability of (2.5′).

<u>Theorem 2.6</u>: If $B(t) = B^*(t) > 0$, $A(t) = A^*(t)$ and (2.33) is satisfied for $B(t)$, then the following conditions are sufficient for the stability of (2.5′) in the $B(t)$-norm:

$$(2.14') \qquad \frac{1-\rho}{\tau} B(t) \leq A(t) \leq \frac{1+\rho}{\tau} B(t) \ , \ \rho = e^{c_0 \tau} \ , \ t \in \omega_\tau \ .$$

<u>Theorem 2.7</u>: If $A(t) = A^*(t) > 0$, $B(t) = B^*(t) > 0$ and $A(t)$ satisfies the condition (2.33), then the condition

$$(2.37) \qquad A(t) \geq \frac{1+\rho}{\tau} B(t) \ , \ t \in \omega_\tau$$

$$\text{with} \quad \rho = \exp\,(c_0 \tau^\gamma) \ , \ 0 \leq \gamma < 1$$

is sufficient for the instability of (2.5′) in $A(t)$-norm.

In the case when $A(t)$ is a nonselfadjoint operator, one can obtain a sufficient condition of the form (2.15) if one follows the analysis of Samarskii [13]. As this condition would involve the inverses of $A(t)$ and $B(t)$, it can not be used in practice. However, our analysis of theorems 2.3 - 2.5 is valid in this case and the stability results (2.17) - (2.28) hold. We rewrite the theorems 2.3 and 2.5 for the present case whereas the other results can be extended to the case of time-dependent operators in a similar manner.

<u>Theorem 2.3′</u>: Let $C_0(t)$ and $C_1(t)$ be linear operators on H such that

$$C(t) = C_0(t) + C_1(t) \ , \ C_0(t) = C_0^*(t) \ ; \ C_1(t) \geq \delta E \ , \ \delta > 0 \ .$$

Then the corresponding initial value problem (2.5´) is stable in $B(t)$-norm ($||S(t)|| \leq 1$, $t \in \omega_\tau$) provided

(2.21´) $$0 < \tau \leq \tau_0 \tilde{\theta} \ , \ \tilde{\theta} = \frac{2\ (1 - \rho_0 + \tau_0\delta)}{1 - \rho_0{}^2 + 2\tau_0\delta + a^2} \ ,$$

where $a = \tau_0 \ ||C_1(t)||$, $\rho_0 = ||E - \tau_0 C_0(t)||$, $t \in \omega_\tau$;

and τ_0 is a real number satisfying the conditions

$$\rho_0 \leq 1 + \tau_0\delta \ , \quad (1 - \rho_0)^2 < a^2 \ .$$

Theorem 2.5´: Let the operators $C_0(t)$ and $C_1(t)$ satisfy the conditions

(2.27´) $$\gamma_1 \ E \leq C_0(t) = C_0^*(t) \leq \gamma_2 \ E \ ,$$

$$C_1(t) \geq \delta \ E \ , \ ||C_1(t)|| \leq \gamma_3 \ , \ t \in \omega_\tau \ ;$$

$$0 \leq \gamma_1 \leq \gamma_2 \ , \quad 0 \leq \delta \leq \gamma_3 \ .$$

In this case $||S(t)|| \leq 1$ and the stability range is largest when $\tau_0 = 2/(\gamma_1 + \gamma_2)$ and the stability condition is given by (2.28).

The proofs of these theorems follow their counterparts in the case of fixed operators. These theorems assure the uniform boundedness of $||S(t)||$ for each $t \in \omega_\tau$ which, from lemma 2.5, ensures the stability of the initial value problem (2.5´).

2.5 Stability With Respect to the Right Hand Side

We shall now discuss the stability of the iteration scheme (2.5) with respect to the right hand side $f(t)$. Samarskii and Gulin [15] showed that under certain conditions an iteration scheme, which is stable with respect to the initial data, is also stable with respect to the right hand side. When $A(t)$ is nonselfadjoint, their results involve the inverses of $A(t)$ and $B(t)$. In the following theorem, we prove that the iteration scheme (2.5′) is stable with respect to the right hand side if the sufficient conditions of stability with respect to the initial data are satisfied.

Theorem 2.8: Let $B(t)$ be a positive selfadjoint operator satisfying a Lipschitz condition in t. Let $A(t)$ be a nonselfadjoint operator and let

$$(2.36) \qquad ||S(t)|| \le \rho = e^{c_0\tau}, \quad t \in \omega_\tau .$$

Then, the solution of (2.5) satisfies the following estimate

$$(2.38) \qquad ||y(t+\tau)||_{b(t)} \le \tilde{\rho}^{n+1} ||y(0)||_{b(0)} + \sum_{n'=0}^{n} \tau\, \tilde{\rho}^{n-n'} ||f(t')||_{b^{-1}(t')},$$

where $\tilde{\rho} = \exp(\tilde{c}_0\tau)$, $\tilde{c}_0 = c_0 + c_1/2$, $t' = n'\tau$.

Proof: From (2.5) we get

$$B^{\frac{1}{2}}(t)\, B^{\frac{1}{2}}(t) \cdot \frac{y(t+\tau) - y(t)}{\tau} + A(t)\, y(t) = f(t) .$$

Writing $x(t+\tau) = B^{\frac{1}{2}}(t)\ y(t+\tau)$ and $\bar{x}(t) = B^{\frac{1}{2}}(t)\ y(t)$ we get

$$x(t+\tau) = S(t)\ \bar{x}(t) + \tau B^{-\frac{1}{2}}(t)\ f(t)\ ,$$

$$\text{where } S(t) = E - \tau C(t)\ ,\ C(t) = B^{-\frac{1}{2}}(t)\ A(t)\ B^{-\frac{1}{2}}(t)\ .$$

It follows that

$$||x(t+\tau)|| \le ||S(t)||\cdot||\bar{x}(t)|| + \tau||B^{-\frac{1}{2}}(t)\ f(t)||$$

$$\le e^{c_0\tau}\cdot e^{c_1\tau/2}\ ||x(t)|| + \tau||f(t)||_{b^{-1}(t)}$$

$$= \tilde{\rho}\ ||x(t)|| + \tau||f(t)||_{b^{-1}(t)}\ ,\ t > 0\ ;$$

$$||x(\tau)|| \le ||S(0)||\cdot||\bar{x}(0)|| + \tau||f(0)||_{b^{-1}(0)}$$

$$\le \exp\ (c_0\tau)||y(0)||_{b(0)} + \tau||f(0)||_{b^{-1}(0)}$$

$$\le \tilde{\rho}\ ||y(0)||_{b(0)} + \tau||f(0)||_{b^{-1}(0)}\ .$$

Using the above inequalities for $t' = \tau, 2\tau, \ldots$ we get

$$||x(t+\tau)|| \le \tilde{\rho}^{n}||x(\tau)|| + \tau \sum_{n'=1}^{n} \tilde{\rho}^{n-n'}||f(t')||_{b^{-1}(t')}$$

$$\le \tilde{\rho}^{n+1}||y(0)||_{b(0)} + \tau \sum_{n'=0}^{n} \tilde{\rho}^{n-n'}||f(t')||_{b^{-1}(t')}\ ;$$

$$t' = n'\tau\ ,\quad t = n\tau\ .$$

This inequality gives the estimate (2.38) since

$$\| x(t+\tau) \| = \| y(t+\tau) \|_{b(t)} .$$

2.6 A Numerical Example

We consider the boundary value problem

$$\text{(2.39)} \qquad L\,u \equiv u_{xx} + \lambda\, u_x = 0 \;,\; \lambda > 0 \;,\; 0 < x < 1 \;;$$

$$u(0) = 0 \;,\; u(1) = 1 \;.$$

For the numerical solution of (2.39) we introduce the mesh points $x_i = ih$, $i = 0,1, \ldots ,N$, $Nh = 1$. Let u_i denote the approximate value of $u(x_i)$. We replace u_{xx} in (2.39) by its three point difference analogue $h^{-2}(u_{i-1} - 2u_i + u_{i+1})$, whereas the first derivative u_x can be replaced either by the central differences $(2h)^{-1}(u_{i+1} - u_{i-1})$ or by the forward differences $h^{-1}(u_{i+1} - u_i)$.

The differential equation (2.39) is discretized at each x_i, $i = 1\ (1)\ N\text{-}1$. If y denotes the vector $(u_1, u_2, \ldots ,u_{N-1})^T$, then the difference analogue of (2.39) can be written as

$$\text{(2.40)} \qquad (A_0 + A_1)\, y = f \;,$$

where

$$A_0 = -h^{-2} \begin{pmatrix} -2 & 1 & & \bigcirc \\ 1 & -2 & 1 & \\ & \ddots & \ddots & 1 \\ \bigcirc & & 1 & -2 \end{pmatrix} = A_0^* > 0 \;;$$

$$A_1 = -\frac{\lambda}{2h}\begin{pmatrix} 0 & 1 & & & \\ -1 & 0 & 1 & & \text{\Large 0} \\ & \ddots & \ddots & \ddots & \\ \text{\Large 0} & & & & 1 \\ & & & -1 & 0 \end{pmatrix}, \quad f = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ h^{-2} + (2h)^{-1}\lambda \end{pmatrix}$$

for central differences;

$$A_1 = -\frac{\lambda}{h}\begin{pmatrix} -1 & 1 & & \text{\Large 0} \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \text{\Large 0} & & & -1 \end{pmatrix}, \quad f = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ h^{-2} + \lambda h^{-1} \end{pmatrix}$$

for forward differences .

Let the Hilbert space H consist of the (N-1) - dimensional vectors $(u_1, \ldots\ldots, u_{N-1})^T$ with the inner product and norm defined by

$$(u,v) = \sum_{i=1}^{N-1} u_i v_i \ , \quad ||u|| = \sqrt{(u,u)} \ .$$

The eigenvalues of A_0 are $4\,h^{-2}\sin^2\frac{k\pi h}{2}$, $k = 1, \ldots, N-1$, so that $\gamma_1 E \leq A_0 \leq \gamma_2 E$ with $\gamma_1 = 4h^{-2}\sin^2\frac{\pi h}{2}$ and $\gamma_2 = 4h^{-2}\cos^2\frac{\pi h}{2}$.

In the case when central differences are used, $A_1 = -A_1^*$ and $||A_1|| \leq \lambda/h$. In the case of forward differences $||A_1|| \leq 2\,\lambda/h$ and $A_1 \geq \delta E$, $\delta = \frac{\lambda}{h}\left(1 - \cos\frac{\pi h}{2-h}\right)$.

The system of difference equations (2.40) can be solved using the following explicit scheme

$$\frac{y^{(n+1)} - y^{(n)}}{\tau} + A_0\, y^{(n)} + A_1\, y^{(n)} = f \,, \tag{2.41}$$

where $y^{(n)}$ is the n-th approximation and τ is an iteration parameter. The equation (2.41) can be written in the operator form

$$y_t + A_0\, y + A_1\, y = f \,,$$

where $\quad y \equiv y^{(n)} \quad$ and $\quad y_t \equiv \dfrac{y^{(n+1)} - y^{(n)}}{\tau}$.

We obtain the following stability conditions from theorem 2.5: For the central difference approximation,

$$\tau < \frac{h^2\,(1 - \cos \pi h)}{\sin^2 \pi h + \lambda^2 h^2/4} \;; \tag{2.42a}$$

For the forward difference approximation,

$$\tau < \frac{h^2\,(1 + \lambda h \kappa - \cos \pi h)}{\sin^2 \pi h + \lambda^2 h^2 + 2\,\lambda h \kappa} \,, \tag{2.42b}$$

$$\kappa = 0.5\,\left(1 - \cos \frac{\pi h}{2-h} \right) .$$

For the numerical computations, the conditions (2.42) are indeed sufficient for stability. However, the stability ranges obtained here are conservative and we give a comparison of the stability ranges of τ in the following table for the case when $h = 0.1$:

Parameter λ	Forward Differences		Central Differences	
	Theoretical	Numerical	Theoretical	Numerical
1.0	0.0046	0.0047	0.005	0.005
10.0	0.0005	0.003	0.0014	0.005
100.0	0.000011	0.0008	0.00002	0.0001

CHAPTER 3

THE BIHARMONIC EQUATION: DISCRETIZATION ERROR

3.1 Introduction

The biharmonic equation in two-dimensions can be written as

$$\Delta\Delta u(P) = F(P) , P \varepsilon D , \tag{3.1}$$

where Δ is the Laplace operator and D is a bounded region with boundary $\dot{D}$. If the function u and its normal derivative u_n are prescribed on $\dot{D}$, i.e.,

$$u(P) = f(P) , u_n(P) = g(P) , P \varepsilon \dot{D} , \tag{3.2a}$$

then (3.1) and (3.2a) constitute the first boundary value problem of the biharmonic equation. If u and its second derivatives are prescribed on the boundary $\dot{D}$,

$$u(P) = f(P) , \Delta u(P) = g(P) , P \varepsilon \dot{D} , \tag{3.2b}$$

then (3.1) and (3.2b) constitute the second boundary value problem.

In order to solve these boundary value problems, the domain D is first covered by a square grid of size h. The biharmonic equation is replaced by its finite difference approximation at each grid point of D and the boundary conditions are replaced by some appropriate finite difference analogues. This leads to a system of linear algebraic equations which must be solved by an iterative or direct method.

The iterative methods frequently considered are those of the alternating direction or the relaxation type. Conte and Dames [16,17] used an alternating direction method for solving the biharmonic boundary value problems in the unit square. They proved the convergence of the iterations and also gave a method for determining the optimum iteration parameters. However, their analysis does not hold for arbitrary domains and for general boundary conditions. Fairweather, Gourlay, and Mitchell [18] and Hadjidimos [19] discussed some improvements to the convergence of the iterative scheme given by Conte and Dames as applied to the unit square.

The iterative methods of relaxation type include the Jacobi and successive over relaxation (S.O.R.) methods. Windsor [20] proved that the point Jacobi and line Jacobi methods diverge for the second boundary value problem (3.1) and (3.2b). He also proved that certain extrapolated procedures converge but the rate of convergence is of the order h^4, where h is the mesh size. Parter [21] discussed a two line - SOR method for the first boundary value problem (3.1), (3.2a) in a rectangular domain. The normal derivative was discretized using second order finite differences whereas the thirteen point difference operator was used to discretize the biharmonic equation. The matrix of the resulting algebraic system must be symmetric and positive definite for this method to be applicable. The rate of convergence strongly depends upon the choice of the relaxation factor. This makes the method unsatisfactory since in general it is difficult to find the optimum relaxation factor.

Viznyuk and Molchanov have given several explicit iterative methods for solving the first boundary value problem (3.1) and (3.2b) [22,23]. Some of these methods are applicable to general domains. The convergence of these iterative schemes was proved and the estimates of the optimum iteration parameters were given. However, these results do not have any practical value since the optimum convergence itself is very slow. The values of the optimum parameters are of the order 10^{+5} and 10^{-6} which introduces large round off errors.

Direct methods of solving the difference equations related to the biharmonic equation also exist in the literature [24,25] and are recommended for solving large linear sparse systems, especially when the system is ill-conditioned. For this reason Ehrlich [26] advocated the use of direct methods and has done a comparative study of some iterative and direct methods used for solving the biharmonic equation. He found that a direct method is competitive with an iterative method so far as the arithmetic operations are concerned. He also gave a rounding error analysis which indicated that the iterative methods are not necessarily more accurate than the direct methods.

An alternative approach is to reduce the biharmonic equation to two coupled Poisson equations. The two second order boundary value problems are solved by the existing methods. This procedure is generally used to obtain the numerical solution of problems in fluid dynamics (see [38] and other references given there). However, this reduction introduces some new problems of convergence related to the simultaneous solution of the Poisson equations. This will be discussed in Chapter 4, while in this chapter we shall discuss the discretization error and convergence of the discrete solution to the exact solution, as $h \to 0$.

The biharmonic equation (3.1) can be reduced to the following equations:

$$(3.3a) \qquad \Delta u(P) = v(P) \ , \ P \in D \cup \dot{D} \ ;$$

$$(3.3b) \qquad \Delta v(P) = F(P) \ , \ P \in D \ .$$

Wood [27] used this technique to solve the second boundary value problem in a rectangular domain. She discretized the equations (3.3) and defined an iterative procedure to solve the resulting algebraic system. The convergence of the discrete solution to the exact solution was proved in the mean as the mesh size $h \to 0$. However, the analysis of Wood is not applicable to general domains since the error functions can not always be expanded in Fourier series as done by Wood.

The first boundary value problem for the biharmonic equation in a rectangular domain was studied by Smith [28] who split the biharmonic equation (3.1) into two Poisson equations. These equations were discretized and an "inner-outer" iteration method was devised to solve the two coupled algebraic systems. The outer iteration scheme involved the solution of two discrete Poisson equations at each step, m, say. Boundary conditions were given at each stage but the conditions on v varied with m. The Poisson difference equations can be solved using various direct and iterative techniques and the term "inner iterations" is used to identify this procedure.

Smith proved that the spectral radius of the basic outer iteration scheme was given by $2(h\sigma_h)^{-1}$ where $\sigma_h \to \sigma$ as $h \to 0$, and σ is a constant. This showed that the basic outer iteration scheme was divergent for small values of h. The convergence was achieved by introducing a relaxation factor in the outer iterations and the spectral radius of the modified iteration scheme was shown to be $\leq 1 - \sigma h$, $h \to 0$. We note here that Smith's analysis is valid only for rectangular domains and for one particular approximation of the boundary conditions for (3.3b). Ehrlich [41] has shown that the rate of convergence of Smith's method can be improved by using two parameters in the outer iterations.

In the present chapter, we consider the biharmonic boundary value problems in a general two-dimensional domain. In the first part we study the second boundary value problem (3.1), (3.2b) and reduce the biharmonic equation to two Poisson equations (3.3) with Dirichlet data known on the boundary. We discretize the second order equations and estimate their discretization errors using the maximum principle. We define two difference approximations and prove the overall discretization error to be of order h^2. Further, we combine the difference approximations for the Poisson equations to obtain an equivalent difference analogue of the biharmonic equation and show a relation between the local truncation error of the biharmonic operator and the overall discretization error. The results of Chapter 1 can not be applied for this purpose since there we dealt only with the

Dirichlet problems.

In the second part of this chapter we study the first boundary value problem (3.1), (3.2a) in a general two-dimensional domain and again reduce the biharmonic equation to two Poisson equations (3.3). The boundary conditions for (3.3b) are not known a priori and have to be approximated from the known data. We discretize the Poisson equations and define difference approximations at every grid point of the region of integration. In order to obtain the error estimates we combine the two difference systems and get an equivalent difference analogue of the biharmonic equation. We now apply the results of Chapter 1 to estimate the discretization error. This also provides a criterion for defining the missing boundary conditions. Finally, we give some examples of the possible definitions of these missing conditions and estimate the corresponding discretization error.

3.2 Discretization Error: The Second Boundary Value Problem

We consider the boundary value problem

$$\begin{aligned} \Delta\Delta u(P) &= F(P) \ , \quad P \in D \ , \\ u(P) &= f(P) \ , \quad P \in \dot{D} \ , \\ \Delta u(P) &= g(P) \ , \quad P \in \dot{D} \ , \end{aligned} \tag{3.4}$$

where D is a bounded two-dimensional domain with the boundary $\dot{D}$. The equations (3.4) can be reduced to two second order Dirichlet problems by writing $\Delta u \equiv v$:

(3.5a) $$\Delta u(P) = v(P) , \quad P \in D ,$$
$$u(P) = f(P) , \quad P \in \dot{D} ;$$

(3.5b) $$\Delta v(P) = F(P) , \quad P \in D ,$$
$$v(P) = g(P) , \quad P \in \dot{D} .$$

The domain D is covered by a square grid of size h ($h > 0$). Let D_h be the set of all mesh points in D and $\dot{D}_h$ be the set of all points of $\dot{D}$ lying on the grid lines; $\bar{D}_h = D_h \cup \dot{D}_h$.
For any grid function V(x,y), we define the difference analogue Δ_h of the Laplace operator as

$$\Delta_h V(x,y) = V_{x\bar{x}}(x,y) + V_{y\bar{y}}(x,y) .$$

This is the usual five point operator. Similarly the difference analogue of the biharmonic operator is the usual thirteen point operator

$$\Delta_h^2 V(x,y) = \Delta_h (\Delta_h V(x,y)) .$$

We give some names for the neighborhoods of a point P_0 relative to the operators Δ_h and Δ_h^2. Let $\phi(P) = \delta(P, P_0)$, the Kronecker delta, and define

$$N_1 (P_0) = \{P / \Delta_h \phi(P) \neq 0\} ;$$

$$N_2 (P_0) = \{P / \Delta_h^2 \phi(P) \neq 0\} .$$

The set $N_1(P_0)$ consists of five mesh points (the Laplace neighbors) and $N_2(P_0)$ consists of thirteen points (the biharmonic neighbors of P_0).

The maximum norm is defined as

$$(3.6) \qquad |V|_{Q_h} = \max_{P \varepsilon Q_h} |V(P)| ,$$

for any grid function V defined on the set Q_h.

The set $\bar{D}_h$ of all mesh points can be divided into the following subsets:

D_h' consists of those points P such that $N_2(P) \subset \bar{D}_h$,

$D_{h,1}$ consists of the points P of $D_h - D_h'$ such that $N_1(P) \subset \bar{D}_h$,

$D_{h,2}$ consists of the remaining mesh points of D_h .

Let $D_h'' = D_h' \cup D_{h,1}$, $D_h^* = D_{h,1} \cup D_{h,2}$, $D_h = D_h' \cup D_{h,1} \cup D_{h,2}$.

The Dirichlet problems (3.5) can be discretized and a finite difference approximation defined at each mesh point of D_h. Let U, V denote the discrete grid functions defined on $\bar{D}_h$ and correspond to the continuous functions u, v. We define the discrete analogue of (3.5) as

$$(3.7a) \qquad \begin{aligned} \Delta_h U(P) &= V(P) , \quad P \varepsilon D_h'' , \\ \Delta_h^* U(P) &= V(P) , \quad P \varepsilon D_{h,2} , \\ U(P) &= f(P) , \quad P \varepsilon \dot{D}_h ; \end{aligned}$$

$$
\text{(3.7b)} \qquad
\begin{aligned}
\Delta_h V(P) &= F(P) , \quad P \in D_h'' , \\
\Delta_h^* V(P) &= F(P) , \quad P \in D_{h,2} , \\
V(P) &= g(P) , \quad P \in \dot{D}_h .
\end{aligned}
$$

The operator Δ_h approximates the Laplace operator with a truncation error of order h^2, and the operator Δ_h^* is a suitable approximation of Δ for the mesh points of $D_{h,2}$. We study two difference approximations for Δ_h^*. If (x,y) is a mesh point of $D_{h,2}$, then at least one of its Laplace neighbors lies outside $\bar{D}_h$. If $(x-\alpha h,y)$ and $(x,y-\beta h)$ lie in $\dot{D}_h$ with $0 < \alpha , \beta \le 1$, then Δ_h^* can be defined by the Shortley-Weller formula

$$
\begin{aligned}
\text{(3.8)} \quad \Delta_h^* V(x,y) \equiv 2h^{-2} [& (\alpha+1)^{-1} V(x+h,y) + \alpha^{-1} (\alpha+1)^{-1} V(x-\alpha h,y) \\
& + (\beta+1)^{-1} V(x,y+h) + \beta^{-1} (\beta+1)^{-1} V(x,y-\beta h) \\
& - (\alpha^{-1}+\beta^{-1}) V(x,y)] ,
\end{aligned}
$$

with the truncation error given by

$$
|\Delta_h^* v - \Delta v| \le 2/3 \, h \, M_3 .
$$

We also study another approximation defined by

$$
\begin{aligned}
\text{(3.9)} \quad \Delta_h^* V(x,y) = h^{-2} [& V(x+h,y) + \alpha^{-1} V(x-\alpha h,y) + V(x,y+h) \\
& + \beta^{-1} V(x,y-\beta h) - (2+\alpha^{-1}+\beta^{-1}) V(x,y)] .
\end{aligned}
$$

The truncation error of this operator is given by

$$|\Delta_h^* v - \Delta v| \leq C\, M_2 \ ,$$

where C is a constant and M_i is defined by

$$M_i = \sup_{j+k=i} \left| \frac{\partial^i v}{\partial x^j \, \partial y^k} \right| .$$

We shall show that the error of these discretizations is of the order h^2 , as $h \to 0$. We introduce the discrete analogue of the Green's function [29] defined by

$$\Delta_{h,P}\, G_h(P,Q) = -h^{-2}\, \delta(P,Q) \ , \quad P \in D_h'' \ ,$$

$$(3.10) \qquad \Delta^*_{h,P}\, G_h(P,Q) = -h^{-2}\, \delta(P,Q) \ , \quad P \in D_{h,2} \ ,$$

$$G_h(P,Q) = \delta(P,Q) \ , \quad P \in \dot{D}_h \ ,$$

$$\text{for} \quad Q \in \bar{D}_h \ .$$

We shall first prove some lemmas:

Lemma 3.1 (Maximum Principle):

If $\quad \Delta_h\, V(P) \geq 0 \ , \ P \in D_h''$, and,

$$\Delta_h^*\, V(P) \geq 0 \ , \ P \in D_{h,2}$$

for any mesh function V defined on $\bar{D}_h$, then V(P) attains its maximum on $\dot{D}_h$.

The proof of this lemma follows from Collatz [30] using the fact that both Δ_h and Δ_h^* are positive operators.

Lemma 3.2: If $V(P)$ is a mesh function defined on $\bar{D}_h$, then for any $P \varepsilon \bar{D}_h$

$$(3.11) \qquad V(P) = h^2 \sum_{Q \varepsilon D_h''} G_h(P,Q) \; (-\Delta_h V(Q)) + h^2 \sum_{Q \varepsilon D_{h,2}} G_h(P,Q) \; (-\Delta_h^* V(Q)) + \sum_{Q \varepsilon \dot{D}_h} G_h(P,Q) \; V(Q).$$

Proof: If we denote the right hand side of (3.11) by $Z(P)$ and use the definition (3.10) of $G_h(P,Q)$, we get

$$Z(P) = V(P) \; , \quad P \varepsilon \dot{D}_h$$

$$\Delta_h^* Z(P) = \Delta_h^* V(P) \; , \quad P \varepsilon D_{h,2}$$

$$\Delta_h Z(P) = \Delta_h V(P) \; , \quad P \varepsilon D_h'' \; .$$

Thus, for $W(P) = Z(P) - V(P)$ we have

$$W(P) = 0 \; , \quad P \varepsilon \dot{D}_h$$

$$\Delta_h^* W(P) = 0 \; , \quad P \varepsilon D_{h,2}$$

$$\Delta_h W(P) = 0 \; , \quad P \varepsilon D_h'' \; .$$

From lemma 3.1, $W(P) \equiv 0$ which proves (3.11).

Lemma 3.3: $G_h(P,Q) \geq 0 \; , \; Q \varepsilon \bar{D}_h$.

Proof: For an arbitrary but fixed $Q \varepsilon \bar{D}_h$, the function $-G_h(P,Q)$

satisfies the following inequalities

$$\Delta_{h,P}\,(-G_h(P,Q)) \geq 0\ ,\quad P \in D_h''$$

$$\Delta^*_{h,P}\,(-G_h(P,Q)) \geq 0\ ,\quad P \in D_{h,2}\ .$$

Thus by lemma 3.1, $G_h(P,Q) \geq 0\ ,\ Q \in \bar{D}_h$.

Lemma 3.4: $$\sum_{Q\varepsilon D_{h,2}} G_h(P,Q) \leq 1\ ,\quad P \in \bar{D}_h\ .$$

Proof: Let Z(Q) be the mesh function defined by

$$Z(Q) = \begin{cases} 1 & Q \in D_h \\ 0 & Q \in \dot{D}_h\ . \end{cases}$$

Then,

$$\Delta_h\, Z(Q) = 0\ ,\ Q \in D_h''$$

$$-\Delta^*_h\, Z(Q) \geq h^{-2}\ ,\ Q \in D_{h,2}$$

for both approximations (3.8) and (3.9).

Applying (3.11) to the function Z(Q) for $P \varepsilon D_h$, we get

$$1 = h^2 \sum_{Q\varepsilon D_{h,2}} G_h(P,Q)\ (-\Delta^*_h\, Z(Q) \geq \sum_{Q\varepsilon D_{h,2}} G_h(P,Q)\ .$$

If $P \varepsilon \dot{D}_h$, then the inequality is trivial.

Lemma 3.5: If d is the diameter of the smallest circumscribed circle containing the domain D, then the discrete Green's function defined by (3.10) with Δ_h^* defined by (3.8) satisfies

$$\text{(3.12)} \qquad h^2 \sum_{Q \varepsilon D_h} G_h(P,Q) \leq d^2/16 \ , \quad P \ \varepsilon \ \bar{D}_h \ .$$

Proof: Let 0 be the centre of the circumscribed circle of diameter d and let

$$W(P) \ = \ r(P)^2/4 \ = \ (x(P)^2 + y(P)^2)/4 \ , \quad P \ \varepsilon \ \bar{D}_h \ ,$$

x(P) and y(P) are cartesian coordinates of P relative to 0.

Then, we have

$$\Delta_h \, W(P) \ = \ 1 \ , \quad P \ \varepsilon \ D_h''$$

$$\Delta_h^* \, W(P) \ = \ 1 \ , \quad P \ \varepsilon \ D_{h,2}$$

$$W(P) \ \leq d^2/16 \ , \quad P \ \varepsilon \ \dot{D}_h \ .$$

If $V(P) = h^2 \sum_{Q \varepsilon D_h} G_h(P,Q)$, then

$$\Delta_h \, V(P) \ = \ -1 \ , \quad P \ \varepsilon \ D_h''$$

$$\Delta_h^* \, V(P) \ = \ -1 \ , \quad P \ \varepsilon \ D_{h,2}$$

$$V(P) \ = \ 0 \ , \quad P \ \varepsilon \ \dot{D}_h \ .$$

By the maximum principle, the function $Z(P) = V(P) + W(P)$ attains its maximum on $\dot{D}_h$, i.e.,

$$V(P) + W(P) \le d^2/16 \ , \quad P \in \bar{D}_h \ .$$

Since $W \ge 0$, (3.12) follows.

<u>Lemma 3.6</u>: If in lemma 3.5 Δ_h^* is defined by (3.9), then

$$h^2 \sum_{Q \in D_h} G_h(P,Q) \le d^2/16 + 0(h^2) \ , \quad P \in \bar{D}_h \ . \tag{3.13}$$

<u>Proof</u>: Proceeding as in the proof of lemma 3.5, we have for the mesh function $W(P) = r(P)^2/4$

$$\Delta_h W(P) = 1 \ , \quad P \in D_h''$$

$$\Delta_h^* W(P) = (\alpha+\beta+2)/4 \ , \quad P \in D_{h,2}$$

$$W(P) \le d^2/16 \ , \quad P \in \dot{D}_h \ .$$

If $V(P) = h^2 \sum_{Q \in D_h} G_h(P,Q)$, then for $Z(P) = V(P) + W(P)$ we have

$$\Delta_h Z(P) = 0 \ , \quad P \in D_h''$$

$$\Delta_h^* Z(P) = (\alpha+\beta-2)/4 \ , \quad P \in D_{h,2}$$

$$Z(P) \le d^2/16 \ , \quad P \in \dot{D}_h \ .$$

Since $\Delta_h Z(P) = 0$ in D_h'', $Z(P)$ attains its maximum on $D_{h,2}$ which is the boundary of D_h''. Thus,

$$\max_{P\varepsilon D_h''} Z(P) \le Z(Q) , \quad Q \varepsilon D_{h,2} \tag{3.14}$$

$$\le \max_{Q\varepsilon D_{h,2}} Z(Q) .$$

Since $\Delta_h^* Z(x,y) \equiv h^{-2} [Z(x+h,y) + Z(x,y+h) + \alpha^{-1} Z(x-\alpha h,y)$

$$+ \beta^{-1} Z(x,y-\beta h) - (2+\alpha^{-1}+\beta^{-1}) Z(x,y)]$$

$$= (\alpha+\beta-2)/4 , \quad (x,y) \varepsilon D_{h,2}$$

and $\quad Z(x-\alpha h,y) , Z(x,y-\beta h) \le d^2/16 ,$

we have, $h^2(\alpha+\beta-2)/4 \le Z(x+h,y) + Z(x,y+h) + d^2(\alpha^{-1}+\beta^{-1})/16$

$$-(2+\alpha^{-1}+\beta^{-1}) Z(x,y) .$$

Using (3.14) we get

$$(2+\alpha^{-1}+\beta^{-1}) Z(x,y) \le 2 \max_{D_{h,2}} Z(Q) + d^2(\alpha^{-1}+\beta^{-1})/16$$

$$- h^2 (\alpha+\beta-2)/4 .$$

As this is true for all $(x,y) \varepsilon D_{h,2}$, we get

$$\max_{D_{h,2}} Z(Q) \le d^2/16 + 0(h^2) .$$

From (3.14) we obtain

$$\max_{\bar{D}_h} Z(Q) \le d^2/16 + 0(h^2) .$$

Since $W(P) \geq 0$, (3.13) follows.

Now we are in a position to estimate the discretization error. Let $\varepsilon_1(P)$ and $\varepsilon_2(P)$ denote the error functions defined by

$$\varepsilon_1(P) = v(P) - V(P)$$

$$\varepsilon_2(P) = u(P) - U(P) , \quad P \in \bar{D}_h .$$

Theorem 3.1: If U is the solution of the discrete system (3.7), (3.8) and u is the solution of (3.4), then

$$(3.15) \qquad |\varepsilon_2|_{\bar{D}_h} = \sup_{\bar{D}_h} |u-U| \leq O(h^2) , \quad h \to 0 .$$

Proof: The discrete functions are defined on $\bar{D}_h$ and $\varepsilon_1(P) = \varepsilon_2(P) = 0$, $P \in \dot{D}_h$. From lemma 3.2 we get for $\varepsilon_1(P)$,

$$(3.16) \qquad \varepsilon_1(P) = h^2 \sum_{Q \in D_h''} G_h(P,Q) \, (-\Delta_h \varepsilon_1(Q))$$

$$+ \; h^2 \sum_{Q \in D_{h,2}} G_h(P,Q) \, (-\Delta_h^* \varepsilon_1(Q)) , \quad P \in \bar{D}_h .$$

Now, $|-\Delta_h \varepsilon_1(Q)| = |-\Delta_h v(Q) + \Delta_h V(Q)|$

$$= |\Delta v(Q) - \Delta_h v(Q)| \leq h^2/6 \, M_4 , \quad Q \in D_h'' ;$$

$$|-\Delta_h^* \varepsilon_1(Q)| = |\Delta v(Q) - \Delta_h^* v(Q)| \leq 2h/3 \, M_3 , \quad Q \in D_{h,2} .$$

From (3.16), it follows that

$$|\varepsilon_1(P)| \le (h^2 \sum_{D_h''} G_h(P,Q)) \max_{D_h''} |-\Delta_h \varepsilon_1(Q)|$$

$$+ h^2 (\sum_{D_{h,2}} G_h(P,Q)) \max_{D_{h,2}} |-\Delta_h^* \varepsilon_1(Q)|$$

$$\le d^2/16 \, . h^2/6 \, M_4 + h^2 . 1 . 2h/3 \, M_3 ,$$

using lemmas 3.4 and 3.5.

$$|\varepsilon_1|_{\bar{D}_h} = \max_{\bar{D}_h} |\varepsilon_1(P)| \le d^2/96 \, h^2 \, M_4 , \quad h \to 0 .$$

The equation (3.16) is also satisfied by $\varepsilon_2(P)$ and

$$|-\Delta_h \varepsilon_2(Q)| = |-\Delta_h u(Q) + \Delta_h U(Q)|$$

$$= |-\Delta_h u(Q) + V(Q)|$$

$$\le |\Delta u(Q) - \Delta_h u(Q)| + |\varepsilon_1(Q)|$$

$$\le h^2/6 \, M_4 + |\varepsilon_1|_{\bar{D}_h} , \quad Q \in D_h'' ;$$

$$|-\Delta_h^* \varepsilon_2(Q)| \le 2/3 \, h \, M_3 + |\varepsilon_1|_{\bar{D}_h} , \quad Q \in D_{h,2} .$$

Thus,

$$|\varepsilon_2(P)| \le d^2/16 \, (h^2/6 \, M_4 + |\varepsilon_1|_{\bar{D}_h})$$

$$+ h^2 \, (2/3 \, h \, M_3 + |\varepsilon_1|_{\bar{D}_h})$$

$$\le d^2/96 \, h^2 \, M_4 \, (1+d^2/16) + 0(h^3) ,$$

$$|\varepsilon_2|_{\bar{D}_h} = \sup_{\bar{D}_h} |u(P) - U(P)| \le O(h^2) \ , \quad h \to 0 \ .$$

Theorem 3.2: The error bound in theorem 3.1 is valid when the approximation (3.9) is used instead of (3.8).

Proof: The proof follows on the similar lines as in theorem 3.1. Here we have

$$|-\Delta_h^* \varepsilon_1(Q)| \le C \, M_2 \ , \quad Q \in D_{h,2} \ ,$$

and using lemmas 3.4 and 3.6 we get

$$|\varepsilon_1(P)| \le (d^2/16 + O(h^2)) \, h^2/6 \, M_4 + h^2.1.C \, M_2 \ ,$$

$$|\varepsilon_1|_{\bar{D}_h} \le h^2 \, (C \, M_2 + d^2/96 \, M_4) \ , \quad h \to 0 \ .$$

Similarly,

$$|\varepsilon_2|_{\bar{D}_h} \le h^2 \, (1+d^2/16) \, (C \, M_2 + d^2/96 \, M_4) \ , \quad h \to 0$$

which proves the theorem.

Thus, we have shown that the discretization error for the biharmonic equation with boundary conditions (3.4) and the difference approximations defined by (3.7) - (3.9) is of order h^2, as $h \to 0$. We shall now combine the discrete systems for the second order Dirichlet problems and obtain an equivalent difference approximation for the biharmonic boundary value problem (3.4).

If $(x,y) \in D_h'$, then $N_2(x,y) \subset \bar{D}_h$. Thus, for the mesh points

P in D_h' we have from (3.7)

$$\Delta_h U(P) = V(P) , \Delta_h V(P) = F(P)$$

which can be combined to give

$$(3.17) \qquad \Delta_h \Delta_h U(P) = F(P) , \quad P \in D_h' .$$

If $(x,y) \in D_{h,1}$, then $\Delta_h V(x,y) = F(x,y)$. At least one element of $N_1(x,y)$ lies in $D_{h,2}$. Assuming $(x-h,y)$ is such a mesh point, we get from (3.7a)

$$h^{-2} [\Delta_h U(x+h,y) + \Delta_h U(x,y+h) + \Delta_h^* U(x-h,y) + \Delta_h U(x,y-h) - 4\Delta_h U(x,y)] = F(x,y)$$

which is the same as

$$(3.18) \qquad \bar{\Delta}_h^2 U(x,y) \equiv \Delta_h^2 U(x,y) + h^{-2} [\Delta_h^* U(x-h,y) - \Delta_h U(x-h,y)] = F(x,y) , \quad (x,y) \in D_{h,1} .$$

Similar expressions are valid for other mesh points of $D_{h,1}$.

If $(x,y) \in D_{h,2}$, then $\Delta_h^* V(x,y) = F(x,y)$. If the operator Δ_h^* is defined in (3.8) and if we assume that the points $(x,y+h)$ and $(x+h,y)$ belong to D_h'', then from (3.7a) and (3.8)

$$(3.19) \qquad \bar{\Delta}_h^2 U(x,y) \equiv 2h^{-2} [\alpha^{-1} (\alpha+1)^{-1} g(x-\alpha h,y) + \beta^{-1}(\beta+1)^{-1} g(x,y-\beta h) - (\alpha^{-1}+\beta^{-1}) \Delta_h^* U(x,y) +$$

$$+ \quad (\alpha+1)^{-1} \Delta_h U(x+h,y) + (\beta+1)^{-1} \Delta_h U(x,y+h)]$$

$$= \quad F(x,y) \;,\; (x,y) \in D_{h,2} \;.$$

In case the operator Δ_h^* is defined by (3.9),

$$\text{(3.20)} \qquad \bar{\Delta}_h^2 \, U(x,y) \equiv h^{-2} [\Delta_h U(x+h,y) + \Delta_h U(x,y+h) + \alpha^{-1} g(x-\alpha h,y) + \beta^{-1} g(x,y-\beta h) - (2+\alpha^{-1}+\beta^{-1}) \, \Delta_h^* U(x,y)] = F(x,y) \;,\; (x,y) \in D_{h,2}.$$

Similar expressions are valid for all mesh points of $D_{h,2}$. On the boundary $\dot{D}_h$, the function U(x,y) is defined by

$$\text{(3.21)} \qquad U(x,y) = f(x,y) \;,\; (x,y) \in \dot{D}_h \;.$$

We have thus defined two difference approximations for the second biharmonic boundary value problem (3.4) in a general domain. Both of these approximations have a discretization error of order h^2, $h \to 0$, as proved in theorems 3.1 and 3.2. On the set of interior grid points D_h', the biharmonic operator is replaced by its thirteen point discrete analogue as in (3.17) which has a truncation error of order h^2. However, on the boundary grid points D_h^* ($D_h^* = D_{h,1} \cup D_{h,2}$) the biharmonic operator is replaced by an approximation $\bar{\Delta}_h^2$ as in (3.19) and (3.20). The truncation error of this operator is of order h^{-1} if Δ_h^* is defined by (3.8) and is of order h^{-2} if Δ_h^* is defined by (3.9).

It follows that in order to obtain a discretization error bound of order h^2, it is sufficient to approximate the biharmonic operator by its second order finite difference analogue at the interior grid points whereas at the boundary grid points, the biharmonic operator can be replaced by difference analogues with a truncation error of order as low as h^{-2}.

3.3 Discretization Error: The First Boundary Value Problem

We consider the boundary value problem

$$(3.22)\qquad \begin{aligned} \Delta\Delta u(P) &= F(P)\ , \quad P \in D\ , \\ u(P) &= f(P)\ , \quad P \in \dot{D}\ , \\ u_n(P) &= g(P)\ , \quad P \in \dot{D}\ , \end{aligned}$$

where u_n is the outward normal derivative on $\dot{D}$ and D is a bounded two-dimensional domain with boundary $\dot{D}$. The equations (3.22) can be split into two second order boundary value problems by writing $\Delta u \equiv v$:

$$(3.23a)\qquad \begin{aligned} \Delta u(P) &= v(P)\ , \quad P \in D \\ u(P) &= f(P)\ , \quad P \in \dot{D}\ ; \end{aligned}$$

$$(3.23b)\qquad \begin{aligned} \Delta v(P) &= F(P)\ , \quad P \in D \\ v(P) &= \Delta u(P)\ , \quad P \in \dot{D}\ . \end{aligned}$$

In this case the values of v are undefined on the boundary $\dot{D}$ and have to be approximated from the known data. We define the difference approximations for (3.23) as follows:

$$\Delta_h U(P) = V(P) , \quad P \in D_h''$$

(3.24a) $$\Delta_h^* U(P) = V(P) , \quad P \in D_{h,2}$$

$$U(P) = f(P) , \quad P \in \dot{D}_h ;$$

$$\Delta_h V(P) = F(P) , \quad P \in D_h''$$

(3.24b) $$\Delta_h^* V(P) = F(P) , \quad P \in D_{h,2}$$

$$V(P) = \bar{\Delta}_h U(P) , \quad P \in \dot{D}_h .$$

In the above definition Δ_h is the five point finite difference analogue of the Laplace operator, Δ_h^* is the first order Shortley-Weller formula defined in (3.8) and $\bar{\Delta}_h$ is an operator that approximates Δ on the boundary $\dot{D}$. The selection of $\bar{\Delta}_h$ will be discussed later.

Since the boundary condition for (3.24b) is undefined, we can not apply the Green's function technique to estimate the discretization error. On the other hand, the results of Chapter 1 are applicable to the Dirichlet problem (3.22). In order to do so, we first combine the difference approximations in (3.24) and obtain an equivalent difference analogue of the Dirichlet problem (3.22).

If $(x,y) \in D_h'$, then $N_2\ (x,y) \subset \bar{D}_h$ and we get on combining $\Delta_h U(P) = V(P)$, $\Delta_h V(P) = F(P)$

(3.25) $$\Delta_h \Delta_h U(x,y) = F(x,y) , \quad (x,y) \in D_h' .$$

If $(x,y) \in D_{h,1}$, then $\Delta_h V(x,y) = F(x,y)$. There is at least one member of $N_1(x,y)$ that lies in $D_{h,2}$. Assuming $(x-h,y)$ is such an element, we get

$$(3.26)\quad \bar{\Delta}_h^2 U(x,y) \equiv \Delta_h^2 U(x,y) + h^{-2}\,[\Delta_h^* U(x-h,y) - \Delta_h U(x-h,y)]$$

$$= F(x,y)\ ,\ (x,y) \varepsilon D_{h,1}\ .$$

Similar expressions are valid for other mesh points of $D_{h,1}$. If $(x,y) \varepsilon D_{h,2}$, then from (3.24b) and (3.8) we get

$$\Delta_h^* V(x,y) \equiv 2\,h^{-2}\,[(\alpha+1)^{-1} V(x+h,y) + \alpha^{-1}(\alpha+1)^{-1} V(x-\alpha h,y)$$

$$+ (\beta+1)^{-1} V(x,y+h) + \beta^{-1}(\beta+1)^{-1} V(x,y-\beta h)$$

$$- (\alpha^{-1}+\beta^{-1}) V(x,y)] = F(x,y)\ ,\ 0 < \alpha\ ,\ \beta \le 1\ .$$

From (3.24) it follows that

$$(3.27)\quad \bar{\Delta}_h^2 U(x,y) \equiv 2\,h^{-2}\,[(\alpha+1)^{-1}\,\Delta_h^* U(x+h,y) + (\beta+1)^{-1}\,\Delta_h^* U(x,y+h)$$

$$+ \alpha^{-1}(\alpha+1)^{-1}\,\bar{\Delta}_h U(x-\alpha h,y) + \beta^{-1}(\beta+1)^{-1} \bar{\Delta}_h U(x,y-\beta h)$$

$$- (\alpha^{-1}+\beta^{-1}) \Delta_h^* U(x,y)] = F(x,y)\ ,\ (x,y) \varepsilon D_{h,2}\ ;$$

$$0 < \alpha\ ,\ \beta \le 1\ .$$

The local truncation error of the operator Δ_h^2 at the mesh points of D_h' is of the order h^2. Since the operator Δ_h^* approximates Δ with an error of order h, the operator $\bar{\Delta}_h^2$ of (3.26) approximates the biharmonic operator with an error of order h^{-1} at the points of $D_{h,1}$. If the operator $\bar{\Delta}_h^2$ of (3.27) also has the truncation error of order h^{-1}, then theorem 1.4 can be applied to obtain estimates for the discretization error.

From the definition of Δ_h^2 and $\bar{\Delta}_h^2$ in (3.27), it follows that

$$(\Delta_h^2 - \bar{\Delta}_h^2)\, u(x,y) = h^{-2}\, [\Delta_h u(x+h,y) - 2(\alpha+1)^{-1}\, \Delta_h^* u(x+h,y)$$

$$+ \Delta_h u(x,y+h) - 2(\beta+1)^{-1}\, \Delta_h^* u(x,y+h) - 4\Delta_h u(x,y)$$

$$+ 2(\alpha^{-1}+\beta^{-1})\, \Delta_h^* u(x,y) + \Delta_h u(x-h,y) + \Delta_h u(x,y-h)$$

$$- 2\alpha^{-1}(\alpha+1)^{-1}\, \bar{\Delta}_h u(x-\alpha h,y) - 2\beta^{-1}(\beta+1)^{-1}\, \bar{\Delta}_h u(x,y-\beta h)]$$

Expanding in Taylor series about (x,y), we get for a sufficiently smooth function u

$$(3.28)\quad (\bar{\Delta}_h^2 - \Delta_h^2) u(x,y) = h^{-2}\, [2\, \alpha^{-1}\, (\alpha+1)^{-1} \left(\Delta u(x,y) - \bar{\Delta}_h u(x-\alpha h,y) \right)$$

$$+ 2\beta^{-1}(\beta+1)^{-1} \left(\Delta u(x,y) - \bar{\Delta}_h u(x,y-\beta h) \right)] + O(h^{-1}) \ .$$

The truncation error of $\bar{\Delta}_h^2$ satisfies

$$|\bar{\Delta}_h^2\, u - \Delta^2\, u| \le |\bar{\Delta}_h^2\, u - \Delta_h^2\, u| + |\Delta_h^2\, u - \Delta^2\, u|$$

$$\le |\bar{\Delta}_h^2\, u - \Delta_h^2\, u| + O(h^2) \ .$$

This truncation error is of order h^{-1} if the operator $\bar{\Delta}_h$ is chosen such that it approximates Δ with an error of order at least h. From (3.28), the operator $\bar{\Delta}_h$ must be chosen such that

$$(3.29)\qquad |\bar{\Delta}_h u(x-\alpha h,y) - \Delta u(x,y)| \le O(h) \ , \ h \to 0 \ , \ 0 < \alpha \le 1 \ .$$

With the above provisions, the biharmonic operator Δ^2 is approximated with an error of order h^2 at the interior mesh points (D_h') of D_h and with an error of order h^{-1} at the boundary mesh points D_h^* $(D_h^* = D_{h,1} \cup D_{h,2})$. We can apply theorem 1.5 with k = 2, $\ell = -1$ and m = 2 and obtain the estimate

$$\max_{\bar{D}_h} | u(P) - U(P) | \leq C h^{3/2} , h \to 0 . \tag{3.30}$$

3.4 <u>Examples of the Boundary Operator</u>

If D is a rectangular domain, the operator Δ_h^* reduces to Δ_h and the operator $\bar{\Delta}_h^2$ of (3.26) reduces to Δ_h^2. The operator $\bar{\Delta}_h^2$ of (3.27) takes the following form:

$$\bar{\Delta}_h^2 U(x,y) \equiv \Delta_h^2 U(x,y) + h^{-2} [\bar{\Delta}_h U(x-h,y) - \Delta_h U(x-h,y) + \bar{\Delta}_h U(x,y-h) - \Delta_h U(x,y-h)] = F(x,y),\ (x,y) \in D_{h,2} . \tag{3.31}$$

The set of boundary mesh points in this case is $D_{h,2}$ and the truncation error of $\bar{\Delta}_h^2$ depends upon that of $\bar{\Delta}_h$. If the operator $\bar{\Delta}_h$ approximates Δ with an error of order h^p, then the truncation error of $\bar{\Delta}_h^2$ of (3.31) is of order h^{p-2}. Applying theorem 1.5 with k = 2, $\ell = p-2$ and m = 2 we get

$$\max_{\bar{D}_h} | u(P) - U(P) | \leq O(h^q) , h \to 0 , \quad q = \min (2, p+1/2) . \tag{3.32}$$

We now give several examples of $\bar{\Delta}_h$. If (x,y) is a mesh point of $D_{h,2}$ and (x-h,y) lies on the left boundary of the rectangle D, then $\bar{\Delta}_h$ can be approximated as follows:

(3.33) $$\bar{\Delta}_h u(x-h,y) = h^{-2} [f(x-h,y+h) + f(x-h,y-h) - 2 f(x-h,y)] .$$

For a smooth function u,

$$\bar{\Delta}_h u(x-h,y) = \Delta u(x-h,y) + 0(1) .$$

In this example p = 0 and (3.32) gives the error estimate of order $h^{\frac{1}{2}}$.

Another example of $\bar{\Delta}_h$ is that used by Smith [28]

(3.34) $$\bar{\Delta}_h u(x-h,y) = h^{-2} [2 u(x,y) - 2h g(x-h,y) + f(x-h,y+h) + f(x-h,y-h) - 4 f(x-h,y)] .$$

The truncation errof of $\bar{\Delta}_h$ is of order h and with p = 1, (3.32) gives the error estimate of order $h^{3/2}$.

A modification of (3.34) is given by

(3.35) $$\bar{\Delta}_h u(x-h,y) = h^{-2} [0.5 u(x+h,y) - h g(x-h,y) + f(x-h,y+h) + f(x-h,y-h) - 2.5 f(x-h,y)] .$$

This modification also has a truncation error of order h. In this approximation we have used the function value at (x+h,y) instead of at (x,y) as in (3.34).

A higher order approximation is given by

$$(3.36) \qquad \bar{\Delta}_h u(x-h,y) = h^{-2} [4\ u(x,y) - 0.5\ u(x+h,y) + f(x-h,y+h) + f(x-h,y-h) - 5.5\ f(x-h,y) - 3\ h\ g(x-h,y)]\ .$$

The truncation error of this approximation is $O(h^2)$ and from (3.32) we get the discretization error estimate of order h^2.

A modification of (3.36) which uses the function values at the points (x+h,y) and (x+2h,y) instead of at (x,y) and (x+h,y) is given by

$$(3.37) \qquad \bar{\Delta}_h u(x-h,y) = h^{-2} [1.5\ u(x+h,y) - 4/9\ u(x+2h,y) + f(x-h,y+h) + f(x-h,y-h) - 55/18\ f(x-h,y) - 5/3\ h\ g(x-h,y)]\ .$$

The truncation error of this approximation is also $O(h^2)$. The approximations (3.35) and (3.37) are given here for they have some computational advantage over the approximations (3.34) and (3.36), respectively, as will be seen in chapter 4.

In the case of a general domain we can construct a difference operator $\bar{\Delta}_h$ which satisfies (3.29). As an example let

$$(3.38) \qquad \bar{\Delta}_h u(x-\alpha h,y) = h^{-2} [u(x+h,y) + u(x,y+h) + u(x,y-h) + \frac{2\alpha-1}{\alpha^2}\ f(x-\alpha h,y) + (1-\alpha^{-1})\ h\ g(x-\alpha h,y) + (\ (1-\alpha^{-1})^2 - 4\ )\ u(x,y)]\ .$$

With this approximation we obtain a truncation error of order h at the mesh points of D_h^*. Consequently the discretization error is of order $h^{3/2}$ from (3.30).

Error bounds can be obtained in a similar manner when other approximations for Δ_h, Δ_h^* and $\bar{\Delta}_h$ are used in (3.24).

CHAPTER 4

THE BIHARMONIC EQUATION: NUMERICAL SOLUTION

4.1 The Outer Iteration Scheme

In section 3.3 we presented a difference approximation for solving the first boundary value problem. The problem is

$$(4.1)\qquad \begin{aligned} \Delta\Delta u(P) &= F(P)\,, \quad P \in D \\ u(P) &= f(P)\,, \quad P \in \dot{D} \\ u_n(P) &= g(P)\,, \quad P \in \dot{D}\,. \end{aligned}$$

The biharmonic equation is replaced by two simultaneous Poisson equations by defining an intermediate variable $v = \Delta u$ (see equations 3.23). The following difference approximations are defined for obtaining the numerical solution of these Poisson equations:

$$(4.2a)\qquad \begin{aligned} \Delta_h U(P) &= V(P)\,, \quad P \in D_h'' \\ \Delta_h^* U(P) &= V(P)\,, \quad P \in D_{h,2} \end{aligned}$$

$$(4.2b)\qquad \begin{aligned} \Delta_h V(P) &= F(P)\,, \quad P \in D_h'' \\ \Delta_h^* V(P) &= F(P)\,, \quad P \in D_{h,2} \end{aligned}$$

The boundary condition for the discrete Poisson equation in U (4.2a) is known

$$(4.3)\qquad U(P) = f(P)\,, \quad P \in \dot{D}_h\,;$$

while for the discrete Poisson equation in V it is defined by

(4.4) $$V(P) = \bar{\Delta}_h U(P) \ , \quad P \in \dot{D}_h \ ,$$

which involves the unknown solution U(P). Several examples of the operator $\bar{\Delta}_h U$ were given in section 3.4.

In order to obtain the solution of the biharmonic boundary value problem using (4.2) - (4.4) we define the following iterative procedure. We start with some initial approximation U_0 to the discrete solution U and obtain successive approximations $U_1, U_2, \ldots, U_m$ using the following iterative scheme:

(4.5a) $$\begin{aligned} \Delta_h V_{m+1}(P) &= F(P) \ , \quad P \in D_h'' \\ \Delta_h^* V_{m+1}(P) &= F(P) \ , \quad P \in D_{h,2} \\ V_{m+1}(P) &= \bar{\Delta}_h U_m(P) \ , \quad P \in \dot{D}_h \ ; \end{aligned}$$

(4.5b) $$\begin{aligned} \Delta_h U_{m+1}(P) &= V_{m+1}(P) \ , \quad P \in D_h'' \\ \Delta_h^* U_{m+1}(P) &= V_{m+1}(P) \ , \quad P \in D_{h,2} \\ U_{m+1}(P) &= f(P) \ , \quad P \in \dot{D}_h \end{aligned}$$

In the case of a rectangular domain $D = \{0 \le x=ih \le Mh=a, 0 \le y=jh \le Nh=b\}$, the operator Δ_h^* at the mesh points of $D_{h,2}$ reduces to the discrete Laplace operator Δ_h. If we denote by U the vector $U = (U^1, U^2, \ldots, U^{N-1})^T$ where $U^j = (U_{1,j}, U_{2,j}, \ldots, U_{M-1,j})^T$, then the iteration scheme (4.5) can be written in the matrix form

$$(4.6) \qquad L^2 U_{m+1} + h^2 (\Phi (U_m) + \Phi' (U_m)) = h^4 F + L D ,$$

where F is a constant vector arising from the values of F(x,y) and D arises from the values of f(x,y) and g(x,y). Moreover,

$$L = \begin{pmatrix} A & I & & & 0 \\ I & A & I & & \\ & \ddots & \ddots & \ddots & \\ & & & & I \\ 0 & & & I & A \end{pmatrix} , \quad A = \begin{pmatrix} -4 & 1 & & & 0 \\ 1 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ 0 & & & 1 & -4 \end{pmatrix}$$

$$\Phi (U) = \begin{pmatrix} \phi^1 \\ \phi^2 \\ \vdots \\ \phi^{N-1} \end{pmatrix} , \quad \phi^j = \begin{pmatrix} (\bar{\Delta}_h U)_{0,j} \\ 0 \\ \vdots \\ 0 \\ (\bar{\Delta}_h U)_{M,j} \end{pmatrix} , \; j = 1(1)N-1$$

$$\Phi' (U) = \begin{pmatrix} V^0 \\ 0 \\ \vdots \\ 0 \\ V^N \end{pmatrix} , \quad V^j = \begin{pmatrix} (\bar{\Delta}_h U)_{1,j} \\ (\bar{\Delta}_h U)_{2,j} \\ \vdots \\ (\bar{\Delta}_h U)_{M-1,j} \end{pmatrix} , \; j = 0 , N$$

L is a block tridiagonal matrix of order (N-1) with blocks of order (M-1), A is a tridiagonal matrix of order (M-1) and I is the identity matrix of order (M-1)

The iteration scheme (4.6) can be written as

(4.7) $$U_{m+1} = - L^{-2}M\, U_m + L^{-2}\, (h^4\, F + L\, D)$$

where

(4.8) $$MU = h^2\, (\Phi\, (U) + \Phi'\, (U))\ .$$

If U is the solution of (4.7) and $E_m = U_m - U$ is the error at the m-th iteration, then

(4.9) $$E_m = H\, E_{m-1} = (H)^m\, E_0\ ,\ H = -L^{-2}\, M\ .$$

The iteration matrix is H and its spectral radius $\rho(H)$ determines the convergence of the outer iterations defined in (4.5).

An eigenvalue λ of $L^{-2}M$ satisfies

$$L^{-2}M\, u = \lambda\, u \Rightarrow Mu = \lambda\, L^2 u\ .$$

For the real inner product we have

$$\lambda = \frac{(MU,U)}{(LU,LU)}\ ,\quad LU \neq 0$$

Thus,

(4.10) $$\rho(L^{-2}M) = \sup_{U\neq 0} \frac{(MU,U)}{(LU,LU)}\ ,$$

where M is defined in (4.8) and

$$(MU,U) = h^2 \sum_{P\varepsilon \dot{D}_h} \bar{\Delta}_h U(P)\, .\, U(P^-)\ ,$$

$$(LU,LU) = h^4 \sum_{P\varepsilon D_h} (\Delta_h U(P))^2\ .$$

The point P^- is the closest neighbor of P in D_h and U is a mesh function which is defined on D_h and vanishes on $\dot{D}_h$.

When the approximation (3.34) is used to define the boundary condition (4.4), Smith [28] proved that

$$\rho(L^{-2}M) = 2(h\sigma_h)^{-1} \text{ where } \sigma_h \to \sigma \text{ as } h \to 0 \text{ , } 0 < \sigma < \infty \text{ ;}$$

$$\rho(H) \sim 2 \ (\sigma h)^{-1} \text{ as } h \to 0 \text{ .} \tag{4.11}$$

This shows that as $h \to 0$, the outer iteration scheme defined by (4.5) and (3.34) is divergent.

If we use a different approximation for defining V in (4.4) (e.g. (3.35) - (3.37), the spectral radius of H will differ from (4.11). Let $\tilde{M}$ represent the matrix operator of (4.8) corresponding to this approximation. The iteration matrix in this case is $H = -L^{-2}\tilde{M}$ and

$$\rho(H) = \rho(L^{-2}\tilde{M}) = \sup_{U\neq 0} \frac{(\tilde{M}U,U)}{(LU,LU)}$$

$$= \sup_{U\neq 0} \frac{(MU,U)}{(LU,LU)} \cdot \frac{(\tilde{M}U,U)}{(MU,U)}$$

$$\leq \sup_{U\neq 0} \frac{(MU,U)}{(LU,LU)} \cdot \sup_{U\neq 0} \frac{(\tilde{M}U,U)}{(MU,U)} \text{ ;}$$

$$\rho(H) \leq 2 \ (h\sigma_h)^{-1} \cdot \gamma_h \tag{4.12}$$

where $$\gamma_h = \sup_{U\neq 0} \frac{(\tilde{M}U,U)}{(MU,U)} \text{ .}$$

We shall show that $\rho(H)$ in this case is bounded by $2\,\gamma(h\sigma)^{-1}$, as $h \to 0$, where γ is a finite constant and depends upon the approximation used in (4.4).

4.2 $\gamma_h \to \gamma$ as $h \to 0$

(4.13)
$$\gamma_h = \sup_{U \neq 0} \frac{(\tilde{M}U,U)}{(MU,U)}$$

where U is a mesh function that vanishes on the boundary $\dot{D}_h$ of the rectangular domain, M is the operator defined in (4.8) and corresponds to the approximation (3.34) while $\tilde{M}$ corresponds to the other approximation used in (4.4). If $\tilde{M}$ is defined by (3.35), then

(4.14)
$$(MU,U) = \sum_{P\varepsilon\dot{D}_h} 2\,(U(P^-))^2$$

(4.15)
$$(\tilde{M}U,U) = \sum_{P\varepsilon\dot{D}_h} 0.5\,U(P^-)\,U(P^{--}) \;.$$

The point P^- is the closest neighbor of P in D_h and P^{--} is the next mesh point of D_h on the same grid line. As an example, if P is the point (x,y) on the left boundary ($x = 0$, $0 < y < b$), then P^- is the point (x+h,y) and P^{--} is the point (x+2h,y).

By lemma 12 of [28], a mesh function U can be extended to a continuous function U(x,y;h) by the formula

(4.16)
$$U(x,y;h) = \sum_{p=1}^{M-1} \sum_{q=1}^{N-1} a_{pq}\, u_{pq} \;,$$

where $a_{pq} = [U , u_{pq}]_h$, $u_{pq} = \sin p \pi x/a \cdot \sin q \pi y/b$

and $$[U,V]_h = h^2 \sum_{P \epsilon \bar{D}_h} U(P) \cdot V(P) .$$

Let the four sides of the rectangle D be denoted by B_1, B_2, B_3, and B_4 where

$$B_1 = \{0 < x = ih < a , y = 0\}$$

$$B_2 = \{0 < x = ih < a , y = b\}$$

$$B_3 = \{0 < y = jh < b , x = 0\}$$

$$B_4 = \{0 < y = jh < b , x = a\} .$$

We can write

$$(MU,U) = I_1 + I_2 + I_3 + I_4 ,$$

$$(\tilde{M}U,U) = \tilde{I}_1 + \tilde{I}_2 + \tilde{I}_3 + \tilde{I}_4$$

where $$I_i = \sum_{P \epsilon B_i} 2 (U(P^-))^2$$

$$\tilde{I}_i = \sum_{P \epsilon B_i} 0.5\, U(P^-) \cdot U(P^{--}) , \quad i = 1,2,3,4 .$$

For the function u_{pq}, we have

$$I_1 = \sum_{x=0}^{a} 2 \sin^2 (p \pi x/a) \sin^2 (q \pi h/b)$$

$$\tilde{I}_1 = \sum_{x=0}^{a} \sin^2 p\,\pi\, x/a\ (0.5 \sin (q\,\pi\, h/b) \cdot \sin (2q\,\pi\, h/b)) .$$

Thus,

(4.17a) $$\tilde{I}_1/I_1 = 0.5 \cos q\,\pi\, h/b = \tilde{I}_2/I_2 .$$

Similarly

(4.17b) $$\tilde{I}_3/I_3 = 0.5 \cos p\,\pi\, h/a = \tilde{I}_4/I_4 .$$

It follows that

$$- 0.5 \le \tilde{I}_i/I_i \le 0.5 , \quad i = 1,2,3,4$$

and

$$- 0.5 \le \frac{(\tilde{M}u_{pq} , u_{pq})}{(Mu_{pq} , u_{pq})} \le 0.5 .$$

This bound is valid for any u_{pq} and thus holds for $U(x,y;h)$ given in (4.16). Consequently

$$\sup_{U \ne 0} \left| \frac{(\tilde{M}U,U)}{(MU,U)} \right| = |\gamma_h| \le 0.5 .$$

As $h \to 0$, $\tilde{I}_i/I_i \to 0.5$ and

(4.18) $$\gamma_h \to \gamma = 0.5 .$$

If we use the second order difference approximation defined by (3.36), the corresponding ratios are given by

$$\tilde{I}_1/I_1 = \tilde{I}_2/I_2 = 2 - 0.5 \cos q \pi h/b$$

$$\tilde{I}_3/I_3 = \tilde{I}_4/I_4 = 2 - 0.5 \cos p \pi h/a$$

In this case, $3/2 \le \frac{(\tilde{M}U,U)}{(MU,U)} \le 5/2$

so that $3/2 \le \gamma_h \le 5/2$.

Moreover,

(4.19) $$\gamma_h \to \gamma = 3/2 \text{ , as } h \to 0 .$$

Similar results can be proved for other difference approximations used in (4.4). In the case of the approximation (3.37) we have $-13/6 \le \gamma_h \le 31/18$ and $\gamma_h \to \gamma = 5/6$ as $h \to 0$.

4.3 The Modified Iterations

We have seen that the spectral radius $\rho(H)$ of the iteration scheme (4.5) is bounded by $2\gamma (h\sigma)^{-1}$, as $h \to 0$, where γ and σ are finite constants. Therefore, the basic outer iteration scheme (4.5) is divergent, as $h \to 0$, irrespective of the boundary approximation used in (4.4). However, the scheme (4.5) can be modified such that the iterations converge. The modified iteration scheme is defined as follows:

We start with an initial approximation U_0 and put $\bar{U}_0 = U_0$. The successive approximations $U_1, U_2, \ldots, U_m$ and $\bar{U}_1, \bar{U}_2, \ldots, \bar{U}_m$ are computed using the following procedure:

(4.20a) $$\Delta_h V_{m+1}(P) = F(P), \quad P \in D_h$$

$$V_{m+1}(P) = \bar{\Delta}_h \bar{U}_m(P), \quad P \in \dot{D}_h ;$$

(4.20b) $$\Delta_h U_{m+1}(P) = V_{m+1}(P), \quad P \in D_h$$

$$U_{m+1}(P) = f(P), \quad P \in \dot{D}_h ;$$

(4.20c) $$\bar{U}_{m+1}(P) = \omega \bar{U}_m(P) + (1-\omega) U_{m+1}(P), \quad P \in D_h .$$

This procedure can be written in the matrix form

(4.21) $$\bar{U}_{m+1} = \bar{H} \bar{U}_m + (1-\omega) L^{-2} D ,$$

(4.22) $$\bar{H} = \omega I + (1-\omega) H , \; H = -L^{-2} M ;$$

(4.23) $$U_{m+1} = H \bar{U}_m + L^{-2} D .$$

If $E_m = U_m - U$, then

(4.24) $$E_{m+1} = (\bar{H})^m H E_0 .$$

The convergence of outer iterations is now governed by the spectral radius of $\bar{H}$. If λ is the largest eigenvalue of $L^{-2}M$, then the corresponding eigenvalue μ of $\bar{H}$ from (4.22) is given by

(4.25) $$\mu = \omega - (1-\omega)\,\lambda\ .$$

All the eigenvalues of $\bar{H}$ lie in the interval [-1,1] if ω is selected such that $\mu \geq -1$, for which

(4.26) $$1 > \omega > (\lambda-1)\ /\ (\lambda+1)\ .$$

The iteration scheme (4.20) converges for any value of ω satisfying (4.26). The convergence of the modified iterations is optimum if ω is chosen such that it satisfies (4.26) and for which

(4.27) $$\rho\ (\bar{H}\ (\omega)) = \max_i \left|\ \mu^i(\omega)\ \right|$$

is minimum, where μ^i are the eigenvalues of $\bar{H}(\omega)$.

We now give some examples which demonstrate the usefulness of the above process.

We consider the square domain $0 \leq x,y \leq 1$. For $h = 0.1$, the largest eigenvalue of $L^{-2}M$ has the following approximate values:

For the first order approximation (3.34) $\lambda_1 \approx 4.85$

For the improved approximation (3.35) $\lambda_2 \approx 2.06$

For the second order approximation (3.36) $\lambda_3 \approx 7.6$

For the improved approximation (3.37) $\lambda_4 \approx 3.84$

Here the subscripts identify the approximation used in (4.4). The corresponding minimum values of ω such that $\rho(\bar{H}) < 1$ are

(4.28) $$\omega_1 \geq 0.65\ ,\ \omega_2 \geq 0.34\ ,\ \omega_3 \geq 0.76\ ,\ \omega_4 \geq 0.58\ .$$

For $h = 0.05$, the extreme eigenvalues of $L^{-2}M$ have the following approximate values

$$\lambda_1 \approx 10.5 \ , \ \lambda_2 \approx 4.85 \ , \ \lambda_3 \approx 16.25 \ , \ \lambda_4 \approx 8.7 \ .$$

The corresponding minimum values of ω are

$$(4.29) \quad \omega_1 \geq 0.826 \ , \ \omega_2 \geq 0.655 \ , \ \omega_3 \geq 0.942 \ , \ \omega_4 \geq 0.793 \ .$$

For $h = 0.025$ the extreme eigenvalue of $L^{-2}M$, when (3.35) is used, is approximately 10.67 and the corresponding range of ω is $1 > \omega \geq 0.8285$.

We note here that the use of the approximation (3.35) over (3.34) in (4.4) results in a reduction of the spectral radius of H by half. This also gives a larger range for ω. Similarly, the use of (3.37) over (3.36) results in a larger range for ω.

A larger range for ω is desirable since in practice it is neither convenient nor economical to find the eigenvalue spectrum of H in order to find an optimum value of ω. Ordinarily, one tries some values of ω for the modified iterations (4.20) and selects the one for which the convergence is reasonably fast. If the range of ω is small, this search becomes difficult. Thus, we choose the boundary approximation which gives a larger range for ω.

The rate of convergence of (4.20) depends upon $\rho(\bar{H})$ and the smaller $\rho(\bar{H})$ is, the faster is the convergence. From (4.22), the eigenvalue spectrum of $\bar{H}(\omega)$ is obtained by mapping the eigenvalue spectrum of H into the interval [-1,1] and the zero eigenvalues of H are mapped into ω. Consequently, $\rho(\bar{H}(\omega)) \geq \omega$ and we should choose the smallest value of ω for which $\rho(\bar{H}(\omega))$ is a minimum. This is possible if the range of ω is large.

In view of these considerations, it is advantageous to choose (3.35) over (3.34) for the first order approximation and (3.37) over (3.36) for the second order approximation. It may be noted that the range of ω becomes smaller if one uses a higher order approximation. An optimum choice of ω depends upon the desired accuracy and the available computing time.

4.4 Numerical Examples

In this chapter and in chapter 3 we have discussed a method for solving biharmonic boundary value problems numerically. In order to study its effectiveness we solved several boundary value problems. We also solved these problems using certain other known methods. The results of these computations are given in the present section.

First of all we considered the boundary value problem

$$\begin{aligned} \Delta\Delta u(x,y) &= 8 \ , \ (x,y) \in D \\ u(x,y) &= 0 \ , \ (x,y) \in \dot{D} \\ u_n(x,y) &= (1-2x)\ y(1-y) \\ &+ (1-2y)\ x(1-x) \ , \ (x,y) \in \dot{D} \end{aligned} \tag{4.30}$$

where $D = \{(x,y) \ / \ 0 \leq x,y \leq 1\}$. The exact solution of this boundary value problem is $u(x,y) = xy(1-x)(1-y)$.

The biharmonic equation was replaced by two Poisson equations which were discretized as in equations (4.2) - (4.4). The systems of algebraic equations corresponding to the discrete Poisson equations (4.2a) and (4.2b) were then solved by using the direct method of odd/even reduction as given by Buzbee, Golub and Nielson [31]. Thus, the inner iterations consisted of the direct solution of discrete Poisson equations while the procedure (4.20) was used for outer iterations. The smallest values of ω, the iteration parameter, needed to make these outer iterations convergent are given in (4.28) and (4.29).

The results of these computations are given in the following table. This table contains the order of the discretization error corresponding to each boundary approximation used in (4.4) as well as values of the optimum convergence factor ω_0, the number N of outer iterations needed for convergence and the maximum error ε of the discrete solution after N outer iterations.

As expected the use of the discretization (3.35) in place of (3.34) and of the discretization (3.37) in place of (3.36) led to a smaller value of ω_0 for a particular order of accuracy and resulted in a faster rate of convergence.

We also solved the problem (4.30) by discretizing the biharmonic operator using the thirteen point formula and solving the resulting algebraic system by the alternating direction method as proposed by Conte and Dames [17]. The mesh size was chosen to be $h = 0.05$. We used the iteration parameters given in [17] and found that iterations diverged. A discrete solution was obtained with another choice of iteration parameters but the maximum discretization

Step Size h	Boundary Approximation	Order of Discretization Error	ω_0	N	ε
0.05	(3.33)	$h^{1/2}$	*	*	0.03
0.1	(3.34)	$h^{3/2}$	0.82	6	0.0001
	(3.35)	$h^{3/2}$	0.67	3	0.0001
	(3.36)	h^2	0.88	5	0.0001
	(3.37)	h^2	0.80	5	$9\text{x}10^{-5}$
				20	$7\text{x}10^{-8}$
0.05	(3.34)	$h^{3/2}$	0.91	16	0.000011
	(3.35)	$h^{3/2}$	0.82	6	0.000011
	(3.36)	h^2	0.942	11	$9\text{x}10^{-5}$
	(3.37)	h^2	0.88	8	$7\text{x}10^{-5}$
0.025	(3.35)	$h^{3/2}$	0.88	8	0.0007

*No iterations were required in this case.

error was found to be 0.06248 at the point x = y = 0.5.

Next we considered the following boundary value problem which was also discussed by Greenspan [32]:

$$\text{(4.31)} \qquad \begin{aligned} \Delta\Delta u(x,y) &= 0 \ , \quad (x,y) \in D \\ u(x,y) &= x^3-2y^2 \ , \quad (x,y) \in \dot{D} \ ; \end{aligned}$$

$$u_x(0,y) = u_x(1,y) = 3x^2 \ , \ 0 \le y \le 1$$

$$u_y(x,0) = u_y(x,1) = -4y \ , \ 0 \le x \le 1$$

where $D = \{(x,y) \ / \ 0 \le x,y \le 1\}$. The exact solution is

$$u(x,y) = x^3-2y^2.$$

Greenspan solved this problem using a variational method. He triangulated the square domain and replaced the differential equation by a system of 361 linear algebraic equation which was solved by successive overrelaxation with zero initial data. The mesh size was 0.05. Greenspan obtained a solution after 1101 iterations when the maximum error was 0.00035 at $x = 0.65$, $y = 0.5$. The computing time on CDC 3600 was 24 minutes. He also tried to compute with $h = 0.01$ but could not obtain any significant result even though the computations were carried out for three hours.

We applied the method of splitting to this problem with $h = 0.05$ and used the improved first order approximation (3.35) with $\omega = 0.73$ to define the boundary condition in (4.4). The outer iterations converged to the solution of the discrete problem in 26 iterations. The computing time on IBM 360/50 was 4 minutes and the maximum error of this solution was 0.00085 at $x = 0.2$, $y = 0.5$.

When the improved second order approximation (3.37) was used with $\omega = 0.845$, the convergence was achieved in 32 iterations. The maximum error was 0.00035 at $x = 0.05$, $y = 0.95$ and computing time was 6 minutes.

The method (4.20) is also applicable to non-rectangular domains. In order to demonstrate this we chose the following boundary value problem. It was possible to introduce some transformation to transform the domain D into a rectangular domain but this was not done. The analysis of sections 4.1 - 4.3 is no longer applicable in the case of non-rectangular domains as the difference equations can not, in general, be written in a matrix form of the type (4.6). The example we considered was

$$\text{(4.32)} \qquad \begin{aligned} \Delta\Delta u(x,y) &= 64 && , (x,y) \in D \\ u(x,y) &= (1-x^2-y^2)^2 && , (x,y) \in \dot{D} \\ u_n(x,y) &= 0 && , (x,y) \in \dot{D} \end{aligned}$$

where $D = \{(x,y) \,/\, x \geq 0 \,,\, y \geq 0 \,,\, x^2+y^2 \leq 1\}$ and u_n is the outward normal derivative. The exact solution is $u(x,y) = (1-x^2-y^2)^2$.

The biharmonic equation was split into two Poisson equations and difference approximations were defined as in equations (4.2) - (4.4). The boundary condition in (4.4) was defined by formulas of the type (3.38) with a truncation error of order h. The overall discretization error of this finite difference approximation would be of order $h^{3/2}$ from (3.30).

We used the method of successive overrelaxation to solve the inner iterations for the discrete Poisson equations. With h = 0.1, the extreme eigenvalue of the basic outer iteration matrix was ≈ -16.0 which showed divergence of the iteration scheme (4.5). A modification to this scheme was introduced as in (4.20) and it was found that $\omega \geq 0.88$ for the convergence of outer iterations. With

$\omega = 0.9$ the iterations converged to a discrete solution in 26 iterations with a maximum error of 0.0033 at $x = y = 0.5$. Zlámal [5] computed the solution of this problem without splitting the biharmonic equation. He used a difference approximation with discretization error of order $h^{3/2}$ and found that the maximum error between the discrete solution and the exact solution was 0.00956 at $x = 0.4$, $y = 0.2$.

We carried out some partial computations with $h = 0.05$. In this case the extreme eigenvalue of the basic outer iteration matrix was ≈ -29.0 and for convergence $\omega \geq 0.933$. The iterative procedure was found to be convergent with $\omega = 0.95$ and after 20 outer iterations a solution was obtained with maximum error 0.069 at $x = y = 0.5$.

It may be noted here that one could use the non-iterative solution obtained with the $O(h^{1/2})$ approximation (3.33) as a starting approximation for the outer iterations. In many cases the discrete Poisson equations could be solved by using direct methods that have been published recently [33,34] and deal with the direct solution of discrete Poisson equations on non-rectangular domains.

Finally, we solved two examples to demonstrate the application of the method of splitting to second biharmonic boundary value problems (see section 3.2). We first considered the boundary value problem

$$(4.33)\qquad \begin{aligned} \Delta\Delta u(x,y) &= 8 \quad &, (x,y) \in D \\ u(x,y) &= 0 \quad &, (x,y) \in \dot{D} \\ \Delta u(x,y) &= -2x(1-x) \\ &\quad -2y(1-y) &, (x,y) \in \dot{D} \end{aligned}$$

where D is the square $0 \le x,y \le 1$ and the exact solution is $u(x,y) = xy\,(1-x)\,(1-y)$.

The problem (4.33) was replaced by two Poisson equations with well defined boundary conditions. The difference approximations were defined as in (3.7) and discrete Poisson equations were solved by the direct method of odd/even reduction [31]. The computations were carried out for h = 0.05 and the discrete solution was found to be close to the exact solution, the maximum error being smaller than 10^{-7}.

For the sake of comparison, the problem (4.33) was solved by discretizing the biharmonic equation using the thirteen point formula with h = 0.05. We used the alternating direction method proposed by Conte and Dames [16] as well as its modifications [18,19] for solving the resulting system of algebraic equations. The iterative scheme diverged in each case. However, we were able to obtain the discrete solution with a maximum error of 0.06248 at x = y = 0.5.

The second example we considered was

$$(4.34)\qquad \begin{aligned} \Delta\Delta u(x,y) &= 0 &&, (x,y) \in D \\ u(x,y) &= (1-x^2-y^2)^2 &&, (x,y) \in \dot{D} \\ \Delta u(x,y) &= -8 + 16\,(x^2+y^2) &&, (x,y) \in \dot{D} \end{aligned}$$

where $D = \{(x,y) \,/\, x \ge 0 \,,\, y \ge 0 \,,\, x^2+y^2 \le 1\}$ and the exact solution is $u(x,y) = (1-x^2-y^2)^2$.

We defined finite difference approximations as in (3.7) and solved the systems of algebraic equations corresponding to each

Poisson equation by successive overrelaxation. The maximum error with $h = 0.1$ and 0.05 was 0.0019 and 0.00049, respectively, and occurred at $x = y = 0.4$.

4.5 Conclusions

In a recent paper Wood [42] has discussed the effects of a periodicity condition on the solution of elliptic difference equations. She used the usual thirteen point discrete analogue for solving the biharmonic boundary value problems. This reduced the problem to the solution of a matrix equation and several methods were analyzed for solving it. Wood found that the alternating direction method of Douglas and Rachford as used by Conte and Dames [16,17] with a cycle of two parameters was superior to the Peaceman and Rachford type alternating direction method [7]. She also discussed the k-line successive overrelaxation method and found it to be inferior to the Douglas-Rachford ADI method with two parameters. Thus, according to Wood, the method used by Conte and Dames for solving the biharmonic boundary value problems is the best available. However, from the numerical examples (4.30) and (4.33) we conclude that the splitting method is much better. It is applicable to non-rectangular domains too.

Spijker [43] has studied the effect of splitting of difference formulas on the roundoff error. He considered second order differential equations and tried to obtain a procedure for reducing the accumulated roundoff error in the numerical solution. He showed that by an appropriate choice of splitting schemes, the roundoff error is substantially reduced. For an initial value problem for a second order ordinary differential equation, Spijker compared a fourth order multistep difference scheme with several split schemes and found that the roundoff error was reduced by as much as a factor of 100. In case of a second order hyperbolic equation it was found that the accumulated roundoff error of the split equations behaved like $O(h^{-1})$ whereas for the original difference equation it behaved like $O(h^{-2})$.

In light of the above results, we conjecture that similar results are also valid for higher order differential equations. In particular, if the biharmonic equation is solved without splitting, one would expect its solution to have a large amount of accumulated roundoff error.

In conclusion, we can say that splitting of the biharmonic equation into two Poisson equations and solving them by the method discussed in chapters 3 and 4 has definite advantages over solving the discretized biharmonic equation directly. In case of the second biharmonic boundary value problems these advantages are much more pronounced. In fact, higher order differential equations should be split into several lower order equations whenever the boundary conditions for the split equations are known. In case of the first biharmonic boundary value problem, one has to select an appropriate difference approximation for the missing boundary condition and find a corresponding value of the iteration parameter for the convergence of outer iterations. This appears to be a better procedure than solving the discretized biharmonic equation directly which leads to an ill-conditioned system of linear algebraic equations.

CHAPTER 5

COMPUTATIONAL PROBLEM ASSOCIATED WITH A NONSELFADJOINT EQUATION

5.1 Introduction

In this chapter we consider the Dirichlet problem for the following equation

$$(5.1) \qquad \Lambda u \equiv u_{xx} + u_{yy} + \lambda_1 u_x + \lambda_2 u_y = 0$$

in a rectangular domain $D = \{0 < x < a \; , \; 0 < y < b\}$; the parameters λ_1 and λ_2 are assumed to be positive. We are interested in considering those cases for which the parameters assume large values. In order to obtain the numerical solution we discretize (5.1) over the mesh

$$D_h = \{(ih,jh) \; ; \; i=1(1)N-1 \; , \; j=1(1)M-1 \; , \; Nh=a \; , \; Mh=b\} \; .$$

The equations (5.1) can be discretized using two different finite difference approximations. In both these cases we replace the second derivatives by their three point central difference analogues while the first derivatives are replaced by using central differences in one case and forward differences in the other. The resulting difference equations are

$$(5.2) \qquad h^2 \Lambda_1 u_{i,j} \equiv (1+ \frac{\lambda_1 h}{2}) u_{i+1,j} + (1- \frac{\lambda_1 h}{2}) u_{i-1,j} +$$

$$+ \left(1+\frac{\lambda_2 h}{2}\right) u_{i,j+1} + \left(1-\frac{\lambda_2 h}{2}\right) u_{i,j-1} - 4u_{i,j} = 0$$

(Central differences) ;

(5.3) $$h^2 \Lambda_2 u_{i,j} \equiv (1+\lambda_1 h)\, u_{i+1,j} + u_{i-1,j} + (1+\lambda_2 h)\, u_{i,j+1}$$

$$+ u_{i,j-1} - (4+\lambda_1 h+\lambda_2 h)\, u_{i,j} = 0 .$$

(Forward differences)

The operators Λ_1 and Λ_2 approximate the difference operator Λ with truncation errors of order h^2 and h, respectively. If we define the difference approximation (5.2) or (5.3) at every mesh point of D_h and apply theorem 1.5, we can conclude that the discrete solutions of these equations converge to the exact solution with errors of order h^2 and h, respectively, as the mesh size h tends to zero. From this analysis it is clear that (5.2) produces a better approximation to the solution than (5.3) does, at least in principle ($h \to 0$). However, in actual computations this expectation is not realized particularly for large values of λ_1 or λ_2. This discrepancy was observed by Burns [35] and Boughner [36]. Similar behaviour is also observed when one tries to solve the Navier-Stokes equations for large values of the Reynolds number [38].

Burns and Boughner analyzed the one-dimensional case of (5.1) given by

$$(5.1') \qquad u_{xx} + \lambda\, u_x = 0 \;,\; 0 < x < a \;,\; u_0 \text{ and } u_a \text{ given.}$$

The corresponding finite difference equations are

$$(5.2') \qquad \Lambda_1'\, u_n \equiv \left(1+\frac{\lambda h}{2}\right) u_{n+1} - 2\, u_n + \left(1-\frac{\lambda h}{2}\right) u_{n-1} = 0 \;,\; n=1(1)N-1$$

(Central differences)

$$(5.3') \qquad \Lambda_2'\, u_n \equiv (1+\lambda h)\, u_{n+1} - (2+\lambda h)\, u_n + u_{n-1} = 0 \;,\; n=1(1)N-1$$

(Forward differences) .

For the difference equation (5.2′) which is obtained by using central differences, they proved that for a fixed h and large λ the discretization error satisfies the following bound for odd values of N

$$(5.4) \qquad ||u-u_h||_h = ||e||_h \geq \frac{\sqrt{h}}{\cos \pi h} \left(0.5 - (\lambda h)^{-1}\right) + 0(e^{-\lambda h})$$

where the norm $||\cdot||_h$ is given by

$$(5.5) \qquad ||e||_h^2 = \sum_{n=1}^{N-1} |e_n|^2$$

This shows that $||e||_h$ does not approach zero when λ takes large values. The following estimates were also proved by Burns and Boughner:

$$(5.6) \qquad ||e||_h \leq \begin{cases} c\sqrt{h}\,(\lambda h/2 - 1), & \text{central differences} \\[2ex] \dfrac{c\sqrt{h}}{1+\lambda h}, & \text{forward differences} \end{cases}$$

from which it follows that for forward differences $||e||_h \to 0$ as $\lambda \to \infty$ whereas for central differences this is not true. Moreover, Burns obtained some numerical results indicating that the discretization error in the case of central differences can be reduced by choosing N odd.

5.2 Asymptotic Expansion of the Solutions: One Dimensional Case

In order to understand this behaviour completely, we first write down the explicit solutions of the difference equations (5.2´) and (5.3´) and then study their asymptotic behaviour for large values of λ with a fixed value of h. The solutions of (5.2´) and (5.3´) with boundary conditions $u = u_0$ at $x = 0$ and $u = u_a$ at $x = a$ are given by

(5.7) $$u_n = u_a \frac{(1 - \zeta_n)}{(1 - \zeta_N)} + u_0 \frac{(\zeta_n - \zeta_N)}{(1 - \zeta_N)}, \quad n=1(1)N-1$$

where $\zeta_n = \left(\frac{2 - \lambda h}{2 + \lambda h}\right)^n$ for central differences

and $\zeta_n = (1 + \lambda h)^{-n}$ for forward differences .

The exact solution of (5.1´) is

(5.8) $$u(x_n) = \frac{(1-e^{-\lambda x_n})\, u_a - (e^{-\lambda a} - e^{-\lambda x_n})\, u_0}{1 - e^{-\lambda a}}$$

When $\lambda h \leq 2$, one can obtain the following estimates using Taylor's series

(5.9) $$|e_n| = |u_n - u(x_n)| \leq 0(h^q) \ , \ h \to 0$$

where q=2 for central differences and q=1 for forward differences.

When $\lambda h \geq 2$, we consider the special case when $a = 1$, $u_0 = 0$ and $u_a = 1$. The solution of (5.2´) can be written as

(5.10) $$u_n = \frac{1 - (-1)^n \exp\{-n\ (4y + 0(y^3))\}}{1 - (-1)^N \exp\{-N\ (4y + 0(y^3))\}} \ , \ y = (\lambda h)^{-1}.$$

For a fixed value of h and $\lambda h \to \infty$, we obtain the following asymptotic values for the solution (5.10)

(5.11a) If N is even, $$u_n \to \begin{cases} n/N & , \ n \text{ even} \\ \infty & , \ n \text{ odd} \ , \end{cases}$$

(5.11b) If N is odd, $$u_n \to \begin{cases} 0 & , \ n \text{ even} \\ 1 & , \ n \text{ odd} \ , \end{cases}$$

while the exact solution $u(x_n)$ approaches 1 everywhere $(0 < x_n < 1)$. In a similar manner we find that the solution of (5.3´) is given by

(5.12) $$u_n = 1 - y^n + 0\ (y^{n+1}) \ , \ y = (\lambda h)^{-1}$$

which converges to the exact solution as λh increases.

Dorr [39] has studied a more general problem of finding the asymptotic behaviour of solutions to finite difference equations for certain singular perturbation problems in ordinary differential equations. He obtained the above estimates to illustrate the advantage of one-sided differences over the central differences and his results appeared after we had completed our study. Whereas Dorr was interested in the ordinary differential equations, we want to extend these results to partial differential equations of the type (5.1) and are including them in this chapter for the sake of continuity and completeness. Further, these results have a direct relevance in the study of Navier-Stokes equations where similar behaviours are observed [38].

5.3 Two Dimensional Case

For the two-dimensional case, we consider the following boundary value problem

$$(5.13)\quad \begin{aligned} u_{xx} + u_{yy} + \lambda_1 u_x + \lambda_2 u_y &= 0 \ , \ 0 < x < 1 \ , \ 0 < y < 1 \\ u(0,y) = u(1,y) &= 0 \ , \ 0 \le y \le 1 \\ u(x,0) &= f(x) \ , \ 0 \le x \le 1 \\ u(x,1) &= 0 \ , \ 0 \le x \le 1 \ . \end{aligned}$$

The exact solution of this boundary value problem is

$$(5.14) \qquad u(x,y) = e^{-0.5(\lambda_1 x+\lambda_2 y)} \sum_{n=1}^{\infty} f_n \cdot \frac{\sinh \gamma_n (1-y)}{\sinh \gamma_n} \sin n \pi x$$

where $f_n = 2\int_0^1 \exp(\lambda_1 x/2)\, f(x) \sin n \pi x\, dx$

and $\gamma_n = 0.5\ (\lambda_1^2 + \lambda_2^2 + 4\, n^2\, \pi^2)^{1/2}$.

In the case of the central difference approximation (5.2), Boughner [36] obtained the following estimate for a fixed value of h with large values of λ_1 and λ_2 and odd values of N

$$(5.15) \qquad ||e||_h \geq (2 \cos \pi h)^{-1} (0.5 - (\lambda_2 h)^{-1})\ ||f||_h + 0\ (\lambda_1 h \exp(-0.5\ \lambda_2 h))$$

where the norm $||\cdot||_h$ is defined by

$$(5.16) \qquad ||e||_h^2 = h^2 \sum_{i,j=1}^{N-1} e_{i,j}^2 .$$

As in the one-dimensional case discretization error for the central difference approximation does not approach zero for large values of λ_i. For the forward difference equations, the following estimate was obtained

$$(5.17) \qquad ||e||_h \leq 0\ (\lambda_1^{-1}) + 0\ (h \exp(-0.5\ \lambda_2 h)) ,$$

which shows the convergence for large values of λ_i.

The explicit solution of the difference equations can be obtained by separation of variables. It is given by

$$u_{ij} = \sum_{n=1}^{N-1} (P_n q_{1n}^j + Q_n q_{2n}^j) p^i \sin (i n\pi/N) , \tag{5.18}$$

$$i,j = 1(1)N-1$$

where
$$P_n = \frac{2 q_{1n}^N}{N(q_{1n}^N - q_{2n}^N)} \sum_{k=1}^{N-1} p^{-k} f(x_k) \sin (kn\pi/N) ,$$

$$Q_n = - \left(\frac{q_{2n}}{q_{1n}} \right)^N P_n .$$

In the case of central differences, q_{1n} and q_{2n} are the solutions of

$$(1 + \frac{\lambda_2 h}{2}) q^2 - (2 - \mu_n) q + (1 - \frac{\lambda_2 h}{2}) = 0 ,$$

$$\mu_n = - 2 + p (1+ \frac{\lambda_1 h}{2}) \cos n\pi h ,$$

$$p^2 = \frac{2 - \lambda_1 h}{2 + \lambda_1 h} .$$

In the case of forward differences, q_{1n} and q_{2n} are the solutions of

$$(1+\lambda_2 h) q^2 - (2+\lambda_2 h-\mu_n) q + 1 = 0 ,$$

$$\mu_n = - (2+\lambda_1 h) + p^{-1} \cos n \pi h ,$$

$$p = (1+\lambda_1 h)^{-\frac{1}{2}} .$$

For a fixed value of h, the value of p for central differences is real when $\lambda_1 h < 2$, zero when $\lambda_1 h = 2$ and is complex when $\lambda_1 h > 2$. We can expect the solution of (5.2) to be identically zero when $\lambda_1 h = 2$. We can also expect this solution to have an oscillatory nature when $\lambda_1 h > 2$.

Numerical solutions obtained by us show the same behaviour as discussed above. As $\lambda_i \to \infty$, the exact solution (5.14) converges to zero everywhere in the rectangular domain except for some grid points near the boundary y = 0. The solution of (5.3) converges to the exact solution as λ_i increases whereas that of (5.2) does not. On the other hand for $\lambda_i h < 2$, the solution of (5.2) is more accurate than that of (5.3).

5.4 Stability

We need to consider the problem of stability if we solve the difference equations by some iterative method. We study the regions of stability for the difference equations (5.2´) and (5.3´) when λh is large.

We consider the following iteration scheme

$$u_n^{m+1} = u_n^m + \tau \Lambda_i' u_n^m , \quad n=1(1)N-1 , \; i=1 , 2 \tag{5.19}$$

where τ is an iteration parameter, m is the iteration number and

Λ_1' , Λ_2' are the operators defined in (5.2′) and (5.3′). The spectrum of (5.19) will give the required conditions of stability. The spectrum can be obtained by the method of Godunov and Ryabenkii (see [11]). For the finite difference equation (5.2′) using central differences, the spectrum is enclosed by the ellipse

$$\zeta_1 = 1 - 2r + 2r \cos\theta + ir\,\lambda h \sin\theta \ , \ r = \tau/h^2 \ , \tag{5.20}$$

while for the equation (5.3′) using forward differences, it is enclosed by the ellipse

$$\zeta_2 = 1 - 2r - \lambda rh + (2r + \lambda rh)\cos\theta + ir\,\lambda h \sin\theta \ . \tag{5.21}$$

For the stability of (5.19), its spectrum must lie inside the unit circle $|\zeta| \leq 1$. It follows that the central difference iteration scheme is stable if

$$\begin{aligned} r &\leq 0.5 \ , & \lambda h &\leq 2 \\ r &\leq 2\,(\lambda h)^{-2} \ , & \lambda h &> 2 \end{aligned} \tag{5.22}$$

The forward difference iteration scheme is stable if

$$r \leq (2 + \lambda h)^{-1} \ . \tag{5.23}$$

Consequently, for large values of λh the stability range for central difference scheme is much smaller than that for the forward difference scheme.

We have carried out some numerical computations which confirm the above results. The actual stability ranges for several values of λ and h are given in the following table:

Mesh Size h	Parameter λ	Stability range for τ (τ < ..)	
		Central differences	Forward differences
0.1	10.0	0.005	0.003
	100.0	0.0001	0.001
	1000.0	0.000002	0.0001
0.05	10.0	0.001	0.001
	100.0	0.0002	0.0003
	1000.0	0.000001	0.00004

We have thus shown that the use of central differences to discretize u_x and u_y in (5.1) leads to large discretization errors and small stability ranges when the parameters λ_1 and λ_2 take large values. Even though forward differences are less accurate than central differences (as $h \to 0$), they give smaller discretization errors and better stability ranges. The non-central differences are also used for discretizing the Navier-Stokes equations to ensure that these equations are of "positive" type which guarantees the convergence of the iterations. It can be said on the basis of the results of this chapter that this also reduces the discretization error for large Reynolds numbers.

CHAPTER 6

EXTENSION OF THE NUMERICAL PROCEDURE TO NAVIER-STOKES EQUATIONS

In chapters 3 and 4 we discussed a method for solving biharmonic boundary value problems numerically. In chapter 5 we considered a nonselfadjoint elliptic differential equation and discussed the behaviour of its numerical solution in actual computation when the mesh size has a finite non-zero value. We were particularly interested in the solution for large values of the parameter appearing in the differential equation.

In this chapter we consider the Navier-Stokes equations in two dimensions which can be written in a nondimensional form as

$$\Delta\Delta\psi + R\left(\frac{\partial\psi}{\partial x} \cdot \frac{\partial(\Delta\psi)}{\partial y} - \frac{\partial\psi}{\partial y} \cdot \frac{\partial(\Delta\psi)}{\partial x} \right) = 0 , \qquad (x,y) \in D . \tag{6.1}$$

This is a fourth order nonlinear differential equation which describes the flow of a viscous fluid in two dimensions. The values of ψ along with its first or second derivatives are prescribed on the boundary of the domain D. The Reynolds number R is a positive parameter which depends on the geometry of the flow and may take large values.

We note that for R = 0 the equation (6.1) reduces to the biharmonic equation while the terms multiplied by R make it nonlinear and nonselfadjoint. It was found in chapter 4 that the method of splitting the biharmonic equation into two Poisson equations is not only convenient and efficient but it also helps in reducing the roundoff error. A similar splitting can be applied for the Navier-Stokes equations by introducing the vorticity ω and writing (6.1) as

$$\omega = -\Delta\psi , \tag{6.2a}$$

$$\Delta\omega + R(\psi_x \omega_y - \psi_y \omega_x) = 0 . \tag{6.2b}$$

If the values of ψ and its second derivatives are prescribed on the boundary of the domain D, then the boundary conditions for (6.2a) and (6.2b) are well defined. However, if the values of ψ and its first derivatives are given as boundary conditions for (6.1), then ω is undefined on the boundary. In this case boundary conditions for (6.2b) are approximated from the known data as in section 3.3. A method of inner-outer iterations can be applied to obtain the numerical solution of the equations (6.2).

We first discretize the equations (6.2) and replace the Laplace operator in (6.2a) and (6.2b) by its five point discrete analogue. The equation (6.2b) contains ω_x and ω_y multiplied by the Reynolds number. We are often interested in solving equations (6.2) for large values of R and use the results of chapter 5 to replace the first derivatives of ω by non-central difference approximations. The central differences may be used to discretize ω_x and ω_y when R takes small values ($Rh < 2/M$, where $M = \max\{|\psi_x| , |\psi_y|\}$). As seen in chapter 5, use of central differences with large values of R leads to a divergent solution. Several authors have been able to obtain numerical solutions of (6.2) using central differences for ω_x and ω_y but these results have been restricted to small values of the Reynolds number [37,38]. For large values of the Reynolds number, non-central differences are used which makes the resulting coefficient matrix for (6.2b) diagonally dominant [7] and the method of successive over-relaxation can be used to solve this matrix equation. However, at this point we emphasize the fact that use of non-central differences is necessary in order to obtain a numerical solution which approximates the exact solution.

Once the differential equations (6.2) are discretized and a suitable approximation for ω on the boundary is defined, the two coupled systems of algebraic equations can be solved by an iterative procedure as discussed in chapter 4. In order to make this procedure convergent, we need an iteration parameter which can be obtained from section 4.3 if an estimate of the spectral radius of the basic outer iteration matrix is known. We also know from section 4.3 that the convergence of the outer iteration scheme can be accelerated by using a boundary approximation for ω which uses more inner points of the discretized region. This reduces the spectral radius of the basic outer iterations. Consequently, a smaller value of the iteration parameter is needed which in turn gives a faster rate of convergence.

When the Navier-Stokes equations (6.2) are discretized using non-central differences for ω_x, ω_y and the discrete Laplace operator, the local truncation error at a mesh point is of order h, the mesh size. If the boundary conditions for ω are chosen judiciously, one can expect the overall discretization error to be of order h. This will also prove the convergence of the discrete solution to the exact solution as the mesh size h tends to zero.

If we are solving the non-stationary Navier-Stokes equations which can be written as

(6.3a) $$\omega = - \Delta \psi$$

(6.3b) $$\omega_t + \omega_x \psi_y - \omega_y \psi_x = \frac{1}{R} \Delta \omega ,$$

we discretize the time derivative ω_t and solve the equations step by step in the time-direction. The time derivative must be discretized in such a way that this procedure is stable. Sufficient conditions for the stability of operator-difference schemes are given in chapter 2 and are applicable to the cases of selfadjoint, nonselfadjoint as well as time dependent operators. These results can be utilized to ascertain the stability of the time-iterative procedure for solving (6.3) numerically.

The results of chapter 1 are not applicable to elliptic differential equations with variable coefficients as some of the lemmas used there are only valid for constant operators. These lemmas can possibly be modified and extended to the case of variable coefficients. In [5], Zlámal proved discretization error estimates for a class of linear fourth order differential equations with variable coefficients in two dimensions. It is not clear as to how these results can be generalized. There is also a need for obtaining discretization error estimates for quasi-linear and nonlinear elliptic differential equations.

We have extended numerical procedures developed for the biharmonic equation and for a linear nonselfadjoint differential equation to solve the Navier-Stokes equations which are nonlinear. In order to handle the nonlinearity these equations are solved iteratively by linearizing the equations at each step of the outer iterations. Results of the previous chapter are applicable as we solve linear equations at each step. We have not proved the convergence of this iterative procedure to the solution of the nonlinear equations nor have we obtained any error estimates. In this sense some of the extensions in this chapter may be considered as heuristic but the method has been found useful for solving certain boundary value problems of Navier-Stokes equations numerically [37,38]. These equations represent flow of fluids and for $R > 1000$ such flows are physically unstable and are never observed. However solutions of the Navier-Stokes equations do exist mathematically and have been obtained for Reynolds numbers as high as 100000 [37]. It is therefore not clear whether numerical solutions for large values of R have any physical significance nor is it clear as to how accurate these solutions are.

REFERENCES

1. R. Courant, K.O. Friedrichs and H. Lewy, On partial difference equations of mathematical physics, Math. Ann. 100(1928) 32-74
2. V.K. Saul'ev, On a class of elliptic equations solvable by the method of finite differences (Russian), Vyčisl. Mat. 1(1957) 81-86, MR 20 #2541
3. W. Littman, Solution of Dirichlet problems by the method of finite differences, C.R. Acad. Sci., Paris 247(1958) 2270-2272, MR 21 #6470
4. V. Thomée, Elliptic difference operators and Dirichlet problem, Contributions to Diff. Eqn. 3(1964) 301-324
5. M. Zlámal, Discretization error estimates for elliptic boundary value problems of the fourth order, SIAM J. Numer. Anal. 4(1967) 626-639
6. J.H. Bramble, A second order finite difference analogue of the first biharmonic boundary value problem, Numer. Math. 9(1966) 236-249
7. G.E. Forsythe and W. Wasow, Finite difference methods for partial differential equations, J. Wiley and Sons, New York. 1967
8. J.H. Bramble, Error estimates in elliptic boundary value problems, Numer. Analysis of Partial Diff. Equations: C.I.M.E., II ciclo, Ispra (1967) published by Edizioni Cremonese, Rome (1968) 69-107

9. A.A. Samarskii, Classes of difference schemes, Ž. Vyčisl. Mat. i Mat. Fiz. 7(1967) 1096-1133

10. L.V. Kantorovich and G.P. Akilov, Functional analysis in normed spaces, Pergamon Press, New York. 1964

11. R.D. Richtmyer and K.W. Morton, Difference methods for initial value problems, Wiley Interscience, New York. 1967

12. V.S. Ryabenkii and A.F. Filippov, On the stability of difference equations (Russian), Gostekhizdat, Moscow. 1956

13. A.A. Samarskii, Necessary and sufficient conditions for the stability of two-layer difference schemes, Soviet Math. Dokl. 9(1968) 946-950

14. A.A. Samarskii, Two layer iteration schemes for nonselfadjoint equations, Soviet Math. Dokl. 10(1969) 554-558

15. A.A. Samarskii and A.V. Gulin, On the stability of difference schemes with respect to right hand members, Soviet Math. Dokl. 11(1970) 630-634

16. S.D. Conte and R.T. Dames, An alternating direction method for solving the biharmonic equation, Math. Tables and Aids Comp. 12(1958) 198-205

17. S.D. Conte and R.T. Dames, On an alternating direction method for solving the plate problem with mixed boundary conditions, J.A.C.M. 7(1960) 264-273

18. G. Fairweather, A.R. Gourlay and A.R. Mitchell, Some high accuracy difference schemes with a splitting operator for equations of parabolic and elliptic type, Numer. Math. 10(1967) 56-66

19. A. Hadjidimos, On some high accuracy difference schemes for solving elliptic equations, Numer. Math. 13(1969) 396-403

20. E.S. Windsor, Iterative solutions of biharmonic difference equations, M.S. thesis, New York University. 1957

21. S.V. Parter, On two-line iterative methods for the Laplace and biharmonic difference equations, Numer. Math. 1(1959) 240-252

22. G.I. Viznyuk, On explicit methods for solution of a biharmonic equation in an arbitrary domain, Dopovidi Akad. Nauk Ukrain. RSR ser. A 7(1968) 579-583

23. G.I. Viznyuk and I.M. Molchanov, On explicit methods with regularization for solution of biharmonic equation, Dopovidi Akad. Nauk Ukrain. RSR ser. A 8(1969) 6-11

24. A.F. Cornock, The numerical solution of Poisson and biharmonic equations by matrices, Proc. Cambridge Philos. Soc. 50(1954) 524-535

25. W.G. Bickley and J. McNamee, Matrix and other direct methods for the solution of systems of linear difference equations, Philos. Trans. Roy. Soc. London ser. A 252(1959) 69-131

26. L. Ehrlich, Iterative versus a directive method for solving fourth order elliptic difference equations, Proc. 21 st National Conf. of Assoc. Comp. Mach. 1966, 29-35

27. W.L. Wood, On an explicit numerical method for solving the biharmonic equation, Numer. Math. 11(1968) 413-426

28. J. Smith, The coupled equation approach to the numerical solution of the biharmonic equation by finite differences. I, SIAM J. Numer. Anal. 5(1968) 323-339

29. J.H. Bramble and B.E. Hubbard, On the formulation of finite difference analogues of the Dirichlet problem for Poisson equations, Numer. Math. 4(1962) 313-327

30. L. Collatz, Numerical treatment of differential equations, Springer Verlag, Berlin. 1960

31. B.L. Buzbee, G.H. Golub and C.W. Nielson, The method of odd/even reduction and factorization with application to Poisson's equation, Rep. LA-4141, Los Alamos Scientific Laboratory, Los Alamos, New Mexico. June 1969

32. D. Greenspan, A numerical approach to biharmonic problems, Comput. J. 10(1967) 198-201

33. J.A. George, The use of direct methods for the solution of the discrete Poisson equation on non-rectangular regions, Rep. STAN-CS-70-159, Computer Science Department, Stanford University, Stanford. June 1970

34. B.L. Buzbee, F.W. Dorr, J.A. George and G.H. Golub, The direct solution of the discrete Poisson equation on irregular regions, Rep. STAN-CS-71-195, Computer Science Department, Stanford University, Stanford. Dec. 1970

35. I.F. Burns, A numerical solution to a nonselfadjoint elliptic partial differential equation, TR no. 8, Univ. Computing Center, University of Tennessee, Knoxville, Tennessee. 1967

36. R.T. Boughner, The discretization error in finite difference solutions to the linearized Navier-Stokes equations for incompressible fluid flow at large Reynolds numbers, Rep. ORNL-TM-2165, Oak Ridge National Laboratory, Oak Ridge, Tennessee. 1968

37. M.M. Gupta, Numerical solution of two-dimensional Navier-Stokes equations, M.Sc. thesis, University of Saskatchewan, Saskatoon. 1969

38. M.M. Gupta and R. Manohar, Numerical solution of a separated viscous flow problem by a nonuniform finite difference net, to appear in Int. J. Num. Methods Engg. 3(1971)

39. F.W. Dorr, The numerical solution of singular perturbations of boundary value problems, SIAM J. Numer. Anal. 7(1970) 281-313

40. S.L. Sobolev, On estimates for certain sums for functions defined on a grid, Izv. Akad. Nauk SSSR, ser. Mat. 4(1940) 5-16

41. L.W. Ehrlich, Solving the biharmonic equation as coupled finite difference equations, SIAM J. Numer. Anal. 8(1971) 278-287

42. W.L. Wood, Periodicity effects on solution of difference equations, SIAM J. Numer. Anal. 8(1971) 439-464

43. M.N. Spijker, Reduction of roundoff error by splitting of difference formulas, SIAM J. Numer. Anal. 8(1971) 345-357