A STUDY OF HOPSCOTCH METHODS FOR SOLVING PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS

BY

ALI DANAEE, B.Sc.,M.Sc.

A Doctoral Thesis

Submitted in partial fulfilment of the requirements

for the award of Doctor of Philosophy

of the Loughborough University of Technology

May, 1980.

Supervisor: PROFESSOR D.J. EVANS, Ph.D.,D.Sc.

Department of Computer Studies

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to my supervisor, Professor D.J. Evans, for his considerable guidance and encouragement throughout the course of this thesis.

I would also like to thank the University of Isfahan and the Ministry of Science and Higher Education of Iran for giving me the opportunity to continue my studies and for their financial support.

I am indebted to my family in general for their unfailing moral support and continued encouragement and to my wife in particular for displaying great understanding and coping admirably throughout my studies.

Finally, my thanks are due to Miss J.M. Briers for her expert typing of this thesis.

To My Parents

CONTENTS

CHAPTER I

INTRODUCTION AND MATHEMATICAL BACKGROUND

1.1 INTRODUCTION

In this introductory chapter we shall be concerned principally with the occurrences, nature and classification of certain types of partial differential equations leaving aside the discussion of particular numerical methods of solution to the remaining chapters.

For a large variety of physical and engineering problems, the dependent variable is expressed in terms of several independent variables. Such problems inherently give rise to the need for partial derivatives in the description of their behaviour. The study of the differential equations arising from these problems constitutes the field of "Partial Differential Equations".

The order of equation being equal (by analogy with the theory of Ordinary Differential Equations) to the order of the highest partial differential coefficient occuring in it.

In general, the solution of partial differential equations presents a much more difficult problem than the solution of ordinary differential equations and except for certain special types of linear partial differential equations, no general method of solution is available. The most important difference between the solution of partial differential equations and ordinary differential equations is that, the general solution of a linear ordinary differential equation contains arbitrary constants of integration whereas the general solution of a linear partial differential equation contains arbitrary functions. In most cases, the general solution of a partial differential equation is of little use, since it has to be made to satisfy other conditions called "boundary conditions" which arise from the physics of the problem (we shall be discussing "boundary conditions" later).

For example, the equations

$$3y^2 \frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} = 2u \qquad (1.1.1)$$

$$\frac{\partial^2 u}{\partial x^2} + a(x,y)\frac{\partial^2 u}{\partial y^2} = 0 \qquad (1.1.2)$$

(where a(x,y) is any given function) are typical partial differential equations of first and second order respectively, x and y being the independent variables and u≡u(x,y) is the dependent variable whose form is to be found. The equations (1.1.1) and (1.1.2) are linear in the sense that u and its partial derivatives occur only to first degree, and that products of u and its derivatives are absent. A typical non-linear equation in one independent variable is:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - u\frac{\partial u}{\partial x} \qquad \text{(Burger's equation)} \qquad (1.1.3)$$

A linear equation is said to be homogeneous if each term contains either the dependent variable or one of its derivatives. For example,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \qquad \text{(Laplace's equation)} \qquad (1.1.4)$$

is homogeneous, whereas

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x,y) \qquad \text{(Poisson's equation)} \qquad (1.1.5)$$

where f(x,y) is a given function, called an inhomogeneous equation.

As in the case of ordinary differential equations, if $u_1, u_2, \ldots, u_n$ are n different solutions of a linear homogeneous partial differential equation, in some given domain then

$$u = c_1u_1 + c_2u_2 + \ldots + c_nu_n \qquad (1.1.6)$$

is also a solution in the same domain where $c_1, c_2, \ldots, c_n$ are arbitrary constant.

1.2 CLASSIFICATION OF PARTIAL DIFFERENTIAL EQUATIONS

Partial differential equations are frequently classified in terms of their mathematical form (such as elliptic, parabolic,...) or in terms of the type of problems to which they apply (e.g. the wave equation, the diffusion equation,...).

In the mathematical sense, the second-order linear partial differential equation, in two independent variables such as

$$Lu \equiv A(x,y)\frac{\partial^2 u}{\partial x^2} + 2B(x,y)\frac{\partial^2 u}{\partial x \partial y} + C(x,y)\frac{\partial^2 u}{\partial y^2} + E(x,y,u,\frac{\partial u}{\partial x},\frac{\partial u}{\partial y}) = 0 \qquad (1.2.1)$$

can be classified according to the nature of the coefficients A,B and C. These coefficients are then constant or depend on the independent variables x and y only.

If $B^2-AC<0$, the equation is elliptic, if $B^2-AC=0$, the equation is parabolic and finally, $B^2-AC>0$ will lead to a hyperbolic equation. (This classification has been made as the equation (1.1.7) resembles that of a general conic section which gives, elliptic, parabolic and hyperbolic equations according to whether $\Delta=B^2-AC<,=$ or >0). This classification scheme is rather interesting since the values of A,B and C depend on the independent variables, thus it is possible for a partial differential equation to change its classification within the different regions of the domain for which the problem is defined. For example, the equation,

$$y\frac{\partial^2 u}{\partial x^2} + 2x\frac{\partial^2 u}{\partial x \partial y} + y\frac{\partial^2 u}{\partial y^2} = 0 \ , \qquad (1.2.4)$$

is elliptic in the region where $x^2-y^2<0$, parabolic along the lines $x^2-y^2=0$ and hyperbolic in the region where $x^2-y^2>0$.

A similar but more complicated classification can be carried out for linear equations in three or more independent variables. In the case of three independent variables, the terms elliptic, parabolic and hyperbolic

should be replaced by their three dimensional analogous (ellipsoidal, etc.) However the two-dimensional terms are often used for higher dimensional problems (e.g. the Laplace equation $\nabla^2 u=0$ in two or three variables is elliptic type, G. Stephenson (1970) p.14).

The second-order partial differential equations can also have constraints in the form of boundary values, initial values, or combinations of both. The elliptic class are equilibrium problems and are described in terms of a closed region having boundary conditions prescribed at every point on the region's boundary. Problems in the parabolic and hyperbolic class are "propagation" problems and can have prescribed boundary conditions on some part of the boundaries initial conditions along other parts, and can also have open-ended regions into which the solution propagates.

A list of the more familiar partial differential equations which frequently occur in physics and engineering is given in the following table (in this table, ∇^2 is known as the Laplace operator $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$).

Equation Type	Equation Form	Equation Application
Laplace's equation	$\nabla^2 u=0$	Steady-state flow of heat and fluids
Poisson's equation	$\nabla^2 u=-f$	Heat transfer with internal heating
The Diffusion equation	$\nabla^2 u=\frac{1}{K^2}\frac{\partial u}{\partial t}$	Non-equilibrium states of heat conduction
The wave equation	$\nabla^2 u=\frac{1}{C^2}\frac{\partial^2 u}{\partial t^2}$	Propagation of acoustic wave
The Biharmonic equation	$\nabla^4 u=F(x,y)$	Deformation of a plate

TABLE 1.1

1.3 WELL-POSEDNESS AND BOUNDARY CONDITIONS

In practical applications, it is very seldom that the general solution of an equation such as (1.2.1) is required; what is needed is a particular solution satisfying certain boundary conditions.

"The mathematical representation of a physical phenomenon by a partial differential equation and a set of boundary conditions is said to be well-posed or well-formulated provided two criteria are satisfied.

Firstly, the existing solution should be unique, since our experience of nature is such that a given set of circumstances leads to just one outcome.

Secondly, the solution obtained should be stable. In other words, a small change in the given boundary conditions should produce only a correspondingly small change in the solution. This is vital since, when the boundary conditions are arrived at by experiment, certain small observation errors in their values will always exist and these errors should not lead to large changes in the solution". (G. Stephenson, 1974, p.21).

To demonstrate the well-posedness condition, consider the Laplace's equation in two-dimensions,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \qquad (1.3.1)$$

with the given boundary conditions

$$u(x,0) = \frac{\sin nx}{n} \quad ,$$

$$\left(\frac{\partial u}{\partial y}\right)_{y=0} = 0 \qquad (1.3.2)$$

where n is a parameter. The solution can be found by separation of variables, to be,

$$u(x,y) = \frac{1}{n} \sin nx. \cosh ny \qquad (1.3.3)$$

As $n\to\infty$, the boundary conditions converge to $u(x,0)=0$, $\frac{\partial u}{\partial y} = 0$ which together with (1.3.1) implies, by Taylor's series, $u(x,y)=0$.

However, as $n\to\infty$, $u(x,y)$ given in (1.3.3) becomes infinitely large. Consequently, the problem defined by (1.3.1)-(1.3.2) is not well-posed

and could not be associated with a physical phenomenon.

Concerning the boundary conditions, there are four main types of such conditions which arise frequently in the description of physical phenomena, these are:

1. Dirichlet conditions, where u is specified at each point of the boundary of a region. For instance the problem of solving Laplace's equation $\nabla^2 u=0$ inside a region with prescribed values of u on the boundary is called the Dirichlet problem.
2. Neumann conditions, where values of the normal derivatives $\frac{\partial u}{\partial n}$ of the function are given on the boundary.
3. Mixed or Robin's conditions, where a combination of u and its derivatives is given on the boundary e.g. $\alpha\frac{\partial u}{\partial x} + \beta u = \gamma$ at $x = 0$.
4. Periodic conditions where the values of u and $\frac{\partial u}{\partial x}$ are such that

$$u(a) = u(b) \quad \text{and} \quad \frac{\partial u(a)}{\partial x} = \frac{\partial u(b)}{\partial x} .$$

In mathematical terminology the equilibrium problems (which are steady state problems) are known as "boundary value problems" (The typical physical examples are steady viscous flow, steady temperature distributions, etc.) The governing equations for equilibrium problems are usually "elliptic".

Propagation problems are initial value problems that have an unsteady state or transient nature. As an example we consider the diffusion equation,

$$Lu \equiv \frac{\partial u}{\partial t} - D(x,t)\frac{\partial^2 u}{\partial x^2} = H(x,t) \qquad (1.3.4)$$

for $0<x<1$ and $0<t\leq T$, with the initial condition

$$u = u_0(x) \qquad \text{for } t=0 \qquad (1.3.5)$$

and the boundary condition,

$$u = u(x,t) \qquad \text{for } x=0 \text{ and } x=1. \qquad (1.3.6)$$

In mathematical parlance such problems are known as "initial boundary value problems".

The governing equations for propagation problems are parabolic or hyperbolic.

1.4 CHARACTERISTIC OF PARTIAL DIFFERENTIAL EQUATIONS

There is a further important aspect of the classification of partial differential equations into hyperbolic, parabolic and elliptic types. This classification is due to the characteristics of the equation.

Here we shall study the characteristics of such equations and determine the specific directions for which integration of the partial differential equation reduces to the integration of an equation involving total differential only (G. Smith, 1975, p.98).

Let the 1^{st} and 2^{nd} derivatives in equation (1.2.1) be denoted as follows:

$$p = \frac{\partial u}{\partial x}, \quad q = \frac{\partial u}{\partial y}, \quad r = \frac{\partial^2 u}{\partial x^2}, \quad s = \frac{\partial^2 u}{\partial x \partial y} \text{ and } t = \frac{\partial^2 u}{\partial y^2}. \tag{1.4.1}$$

Let Γ be a curve in the x-y plane on which the values of u, p and q are such that they and the 2^{nd} order derivatives r,s and t satisfy the equation (1.2.1). (Γ is different from the initial values curve, since on the latter curve, the values of u are known). Therefore, the differentials of u, p and q in the directions tangential to Γ satisfy the following equations:

$$\begin{aligned} dp &= \frac{\partial p}{\partial x}dx + \frac{\partial p}{\partial y}dy = rdx + sdy, \\ dq &= \frac{\partial q}{\partial x}dx + \frac{\partial q}{\partial y}dy = sdx + rdy, \end{aligned} \tag{1.4.2}$$

where the partial differential equation (1.2.1) is written as

$$Ar + 2Bs + Ct + E = 0.$$

Thus it is easy to show by substitution using (1.4.2) that

$$A\frac{d}{dx}(dp - sdy) + 2Bs + C\frac{d}{dy}(dq - sdx) + E = 0$$

i.e.,
$$s\{A(\frac{dy}{dx})^2 - 2B(\frac{dy}{dx}) + C\} - \{A\frac{dp}{dx}\cdot\frac{dp}{dy} + C\frac{dq}{dx} + E\frac{dy}{dx}\} = 0 \tag{1.4.3}$$

Now if we choose $\frac{dy}{dx}$, the tangent to Γ at a point V(x,y) to satisfy

$$A(\frac{dy}{dx})^2 - 2B\frac{dy}{dx} + C = 0 \tag{1.4.4}$$

then along these two lines (i.e. the two roots of the first bracket of 1.4.3) the partial differential equation reduces to the ordinary differential equation

$$A\frac{dp}{dx}\cdot\frac{dq}{dy} + C\frac{dp}{dx} + E\frac{dy}{dx} = 0 \qquad (1.4.5)$$

giving the relationship between the total differential dp and dq with respect to x and y.

The roots of the equation (1.4.4) are called the characteristic directions of the differential equation which can be used for step-by-step integration.

1.5 APPROXIMATE SOLUTIONS

As mentioned before, the majority of partial differential equations cannot be integrated analytically. In these cases, it is necessary to employ some method of approximation. There exist many different approximate techniques, such as *finite differences* and *finite element* methods for solving partial differential equations.

Finite difference methods are still far and away the most widely used and understood for evolutionary problems. Although this is less true for parabolic equations, where finite element methods are increasingly important however, the finite difference methods remain as simple and flexible general purpose tools (Morton, K.W., 1977, p.700).

This thesis deals with finite difference methods where applied to solve parabolic partial differential equations, which shall be given in the next chapter.

1.6 MATHEMATICAL BACKGROUND IN MATRIX THEORY

Notations

A	square matrix of order n
$a_{i,j}$	number in the real field which is the element in the i^{th} row, and j^{th} columns of the matrix A. If $a_{i,j}$ are themselves matrices, then A is called a block matrix.
A^{-1}	inverse of A
A^T or A^*	transpose of A
A^H	conjugate transpose of A
$\|A\|$	determinant of A
$\rho(A)$	spectral radius of A
I	unit matrix of order n
O	null matrix
$\underline{x}$	column vector with elements x_i $(i=1,2,\ldots,n)$
$\overline{\underline{x}}$	complex conjugate of $\underline{x}$
$\|\|A\|\|$	norm of A
$\|\|\underline{x}\|\|$	norm of $\underline{x}$

Definitions

1.6.1 The matrix A is

1.6.2 *diagonal* if its only non-zero elements lie on the diagonal

1.6.3 *non-singular* if $\mathrm{Det}(A)\neq 0$

1.6.4 *symmetric* if $A=A^T$ $(a_{i,j}=a_{j,i},\ i,j=1,2,\ldots,n)$

1.6.5 *orthogonal* if $A^{-1}=A^T$

1.6.6 *diagonally dominant* if $|a_{i,i}| \geq \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{i,j}|$, for all $1\leq i\leq n$

A is said to be strictly diagonally dominant if the strict inequality holds for all $1\leq i\leq n$ in (1.6).

1.6.7 A is *Hermitian* if $A^H=A$. In terms of scalars A is Hermitian if and only if $a_{i,j}=\overline{a}_{j,i}$.

Since this implies that $a_{i,i}=\overline{a}_{i,i}$ ($a_{i,i}$ is real), hence the diagonal elements of a Hermitian matrix are real. A real symmetric matrix is always Hermitian, but a Hermitian matrix is symmetric only if it is real.

1.6.8 If A is real and $\underline{x}$ is complex, then A is *positive definite* if

$$(\underline{x},A\underline{x}) > 0 \qquad \text{for all } \underline{x}\neq 0$$

(N.B. if $\underline{x}$ and $\underline{y}$ are complex then $(\underline{x},\underline{y})=\sum_{i=1}^{n} x_i y_i^T$). A is non-negative or semi positive definite if $(\underline{x},A\underline{x})\geq 0$ for all $\underline{x}\neq 0$ with equality for at least one $\underline{x}\neq 0$.

1.6.9 A is a *band matrix* of *bandwidth* $w=p+q+1$ if $a_{i,j}=0$ for $j>i+p$ or $i>j+q$.

Many problems (e.g. boundary value problems) result in sparse linear systems, where the non-zero elements are located in a band centered along the principal diagonal. If p=q=1, then A is tridiagonal and a pentadiagonal matrix can be obtained when p=q=2.

Here we state an important theorem (without proof) which is sometimes used as a definition for positive (non-negative) definite matrices.

1.6.10 Two matrices A and B are called commutative if AB=BA. They possess the same set of eigenvectors.

Theorem 1.1

A real matrix is positive (non-negative) definite if and only if it is symmetric and all its eigenvalues are positive (non-negative, with at least one eigenvalue equal to zero).

If A is positive definite therefore, it can be written as $A=HJH^{-1}$ where J is a positive diagonal matrix (HJH^{-1} is the Jordan canonical form of A). It is known (Young, 1971, p.16) that, H can be taken to be an orthogonal

matrix (i.e. $H^T=H^{-1}$). If $J^{\frac{1}{2}}$ denotes the diagonal matrix whose elements are the positive square roots of the elements of J, then $A^{\frac{1}{2}}=HJ^{\frac{1}{2}}H^{-1}$ is positive definite by Theorem 1.1 (Hence $(A^{\frac{1}{2}})^2=(HJ^{\frac{1}{2}}H^{-1})^2=A$ is obtained).

Theorem 1.2

A real symmetric matrix A is positive (non-negative) definite if and only if it can be written in the form $A=P^TP$ where P is some non-singular (singular) matrix.

Proof:

i) Assume $A=P^TP$ ($|P|\neq 0$) then for any vector $\underline{V}\neq 0$

$$\underline{V}^TA\underline{V} = \underline{V}^TP^TP\underline{V} = (P\underline{V})^TP\underline{V} > 0 \rightarrow A \text{ is positive definite.}$$

ii) Let A be positive definite (and real). Since $A=A^{\frac{1}{2}}A^{\frac{1}{2}}$ and $A^{\frac{1}{2}}$ is symmetric therefore $A=(A^{\frac{1}{2}})^TA^{\frac{1}{2}}$. As $A^{\frac{1}{2}}$ is also positive definite $|A^{\frac{1}{2}}|\neq 0$, Thus putting $P=A^{\frac{1}{2}}$ gives the required condition. (The proof of non-negative can be done in the same manner).

1.7 VECTOR AND MATRIX NORMS

For the purpose of quantitatively discussing errors, it is convenient to associate with any vector or matrix a non-negative scalar that in some sense measures its magnitude. Such a quantity will be called a *norm*. The most common vector norms are special cases of L_p-norms which follow:

$$||\underline{x}||_p = (|x_1|^p+|x_2|^p+\ldots+|x_n|^p)^{1/p} \qquad 1\leq p\leq\infty$$

The Euclidean norm is obtained when p=2 and the maximum norm is the case when $p=\infty$, i.e.,

$$||x||_2 = (\sum_{i=1}^{n} |x_i|^2)^{1/2}$$

$$||x||_\infty = \sup_{1\leq i\leq n} |x_i|$$

For any vector norm, there exists a consistent matrix norm. Such a norm is given by the *matrix-bound norm* subordinate to the vector norm (G. Dahlquist 1974, p.175), i.e.

$$||A|| = \max_{\underline{x}\neq 0} \frac{||A\underline{x}||}{||\underline{x}||}$$

which is equivalent to

$$||A|| = \max ||A\underline{x}||$$
$$||\underline{x}|| = 1.$$

The matrix norm subrodinate to $||\underline{x}||_p$ is denoted by $||A||_p$ as follows:

$$||A||_1 = \max_j \sum_i |a_{i,j}|$$

$$||A||_2 = (\text{max eigenvalue of } A^TA)^{\frac{1}{2}}$$

and

$$||A||_\infty = \max_i \sum_j |a_{i,j}|.$$

The maximum vector norm $||\underline{x}||_\infty$ and its subordinate matrix norm $||A||_\infty$ are very often used as they have the advantage of being very simple to compute.

The vector norm must have the following properties:

i) $||\underline{x}||>0$ if $\underline{x}\neq 0$,

ii) $||\alpha\underline{x}||=|\alpha|\ ||\underline{x}||$, α is scalar,

iii) $||\underline{x}+\underline{y}||\leq||\underline{x}||+||\underline{y}||$

If a matrix and vector norm are related such that:

$$||A\underline{x}|| \leq ||A||.||\underline{x}|| \quad \text{for any } A \text{ and } \underline{x}$$

then the two norms are said to be consistent or compatible.

1.8 EIGENVALUES AND EIGENVECTORS OF A MATRIX

The eigenvalues λ_i and eigenvectors $\underline{x}_i$ of A satisfy:

$$(A-\lambda_i I)\underline{x}_i = 0 \qquad \underline{x}_i \neq 0$$

i.e. the eigenvalues of A are n-roots of the characteristic equation

$$p_A(\lambda) = |A-\lambda I| = 0$$

where $p_A(\lambda)$ is a polynomial of degree n.

The maximum eigenvalue λ_1, i=1,2,...,n is called the *spectral radius* of A and is denoted by $\rho(A)$.

Theorem 1.3 (Gerschgorin)

Let A have n eigenvalues λ_i, i=1,2,...,n. Then each λ_i lies in the union of the n discs

$$|z-a_{i,i}| \leq r_i, \quad r_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{i,j}|$$

in the z-plane.

From the Gerschgorin Theorem we obtain,

$$\rho(A) \leq \min(\max_i \sum_j |a_{i,j}|, \max_j \sum_i |a_{i,j}|)$$

Proof

For any λ the following inequality is satisfied:

$$|\lambda| \, ||\underline{x}|| = ||\lambda\underline{x}|| = ||A\underline{x}|| \leq ||A|| \cdot ||\underline{x}||$$

which indicates that:

$$\rho(A) \leq ||A|| \quad \text{for any norm}$$

(N.B. $\max_i \sum_j |a_{i,j}| = ||A||_1$ and $\max_j \sum_i |a_{i,j}| = ||A||_\infty$).

Theorem 1.4

If A is symmetric, diagonally dominant matrix with positive diagonal elements, it is positive definite.

Proof

Since A is symmetric, the eigenvalues of A are real. The application of Gerschgorin's theorem indicates that the eigenvalues of A are all positive since A is diagonally dominant with positive diagonal elements. Therefore according to theorem 1.1 A is positive definite.

In our new developments such matrices occur frequently.

1.9 CONVERGENCE OF SEQUENCES OF MATRICES

The matrix A is convergent to zero if the sequence of matrices A, A^2, A^2 converge to the null matrix 0. (Mitchell, 1976, p.15).

Theorem 1.5

$$\lim_{r\to\infty} A^r = 0 \qquad \text{if} \qquad ||A||<1.$$

Proof

$$||A^r|| = ||A\,A^{r-1}|\leq||A||.||A^{r-1}||\leq||A||^2||A^{r-2}||\leq..\leq||A||^r$$

Hence the result follows. This is a sufficient condition but not necessary. The following theorem states the necessary and sufficient condition.

Theorem 1.6

$\lim A^r=0$ if and only if $|\lambda_i|<1$ for all eigenvalues λ_i of A. For proof see Mitchell, 1976, p.15.

Theorem 1.7

Let λ be an eigenvalue of A with eigenvector $\underline{x}$. Then

1. $\alpha\lambda$ is an eigenvalue of αA with eigenvector $\underline{x}$,
2. $\lambda-\mu$ is an eigenvalue of $A-\mu I$ with eigenvector $\underline{x}$,
3. if A is non-singular, then $\lambda\neq0$ and λ^{-1} is an eigenvalue of A^{-1} with eigenvector $\underline{x}$.

Proof

The equation $A\underline{x}=\lambda\underline{x}$ implies that $\alpha A\underline{x}=\alpha\lambda\underline{x}$ and $(A-\mu I)\underline{x}=(\lambda-\mu)\underline{x}$. This indicates part 1 and 2. For part 3, note that $\lambda=0$ implies $A\underline{x}=0.x=0$. Hence, the homogeneous equation $A\underline{x}=0$ has a non-trivial solution and A is singular. Since A is assumed non-singular we must have $\lambda\neq0$. Then $A\underline{x}=\lambda\underline{x}$ implies that $A^{-1}\underline{x}=\lambda^{-1}\underline{x}$ (G.W. Stewart (1973), p.266).

1.10 THE EIGENVALUES OF SOME COMMON MATRICES

The eigenvalues of the (n×n) matrix

$$A = \begin{bmatrix} a & b & & & 0 \\ c & a & b & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & b \\ 0 & & & c & a \end{bmatrix} \tag{1.10.1}$$

where b and c are both real and have the same sign and a is real or complex, are given by

$$\lambda_i = a + 2\sqrt{bc}\ \cos\left(\frac{i\pi}{n+1}\right) \qquad i=1,2,\ldots,n. \tag{1.10.2}$$

If A is a (n×n) cyclic tridiagonal matrix, i.e.,

$$A = \begin{bmatrix} a & b & & & c \\ c & a & b & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & b \\ b & & & c & a \end{bmatrix} \tag{1.10.3}$$

then the eigenvalues are given by

$$\lambda_i = a+2\sqrt{bc}\ \cos\left(\frac{2i\pi}{n}\right) \qquad i=0,1,\ldots,n-1. \tag{1.10.4}$$

1.11 NUMERICAL SOLUTION OF A SYSTEM OF EQUATIONS WITH SOME SPECIAL MATRICES

Given the system of equations

$$\underline{Au}=\underline{f} \tag{1.11.1}$$

where

$$A = \begin{bmatrix} a_1 & b_1 & & & \\ c_2 & a_2 & b_2 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & b_{n-1} \\ 0 & & & c_n & a_n \end{bmatrix} \tag{1.11.2}$$

then the solution can be obtained by successive subtraction of a suitable multiple of each equation from each succeeding one, which changes the system to a simpler one.

Let

$$w_1 = \frac{b_1}{a_1} \quad ; \qquad w_i = \frac{b_i}{a_i - c_i w_{i-1}} \tag{1.11.3}$$

and

$$g_1 = \frac{f_1}{a_1} \quad ; \qquad g_i = \frac{f_i - c_i g_{i-1}}{a_i - c_i w_{i-1}} \quad . \qquad i=2,3,\ldots,n. \tag{1.11.4}$$

The components u_i of the solution vector $\underline{u}$ are then given recursively by:

$$u_n = g_n; \qquad u_i = g_i - w_i . u_{i+1} \quad i=n-1,n-2,\ldots,1. \tag{1.11.5}$$

This is an algorithm which is used very often.

Another algorithm which is frequently used is the system (1.11.1)

where

$$A = \begin{bmatrix} a_1 & b_1 & & & c_1 \\ c_2 & a_2 & b_2 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & b_{n-1} \\ b_n & 0 & & c_n & a_n \end{bmatrix} \tag{1.11.6}$$

Then the solution can be obtained in a similar manner:

Let

$$g_1 = \frac{b_1}{a_1} \; ; \qquad g_i = \frac{b_i}{a_i - c_i g_{i-1}}$$

$$h_1 = \frac{c_1}{a_1} \; ; \qquad h_i = \frac{c_i . h_{i-1}}{a_i - c_i g_{i-1}} \qquad (1.11.7)$$

$$k_1 = \frac{f_1}{a_1} \; ; \qquad k_i = \frac{f_i + c_i h_{i-1}}{a_i - c_i g_{i-1}} \; , \; i=2,3,\ldots,n-1.$$

$$G_1 = b_n \; ; \qquad G_i = g_{i-1} G_{i-1}$$

$$H_1 = c_n \; ; \qquad H_i = H_{i-1} - G_{i-1} h_{i-1}$$

$$F_1 = f_n \; ; \qquad F_i = F_{i-1} + G_{i-1} k_{i-1} \quad i=2,3,\ldots,n-1.$$

$$g_n = h_n = F_n = G_n = 0 \; ;$$

$$H_n = H_{n-1} - (G_{n-1} + c_n)(g_{n-1} + h_{n-1}) \; ;$$

and

$$f_n = F_{n-1} + (G_{n-1} + a_n) f_{n-1} \; .$$

The components u_i of the solution vector $\underline{u}$ are then given recursively by

$$u_n = \frac{k_n}{H_n} \; ; \quad u_i = k_i + g_i . u_{i+1} + h_i u_n, \; i=n-1,n-2,\ldots,1.$$

(Gane, 1974, p.70).

CHAPTER II

PARABOLIC EQUATIONS AND FINITE-DIFFERENCE APPROXIMATIONS

2.1 PARABOLIC EQUATIONS

The parabolic partial differential equations usually arise from mathematical descriptions of time-dependent or evolutionary processes; the solution of such equations can be thought of as evolving as time increases from a given initial state under the influence of certain boundary conditions.

The equations arising from diffusion in an isotropic median, heat conduction, boundary layer flow over a flat plate, and many others are of parabolic type. (Watt, 1978, Current Problems and Methods in P.D.E.s, Chapter 7).

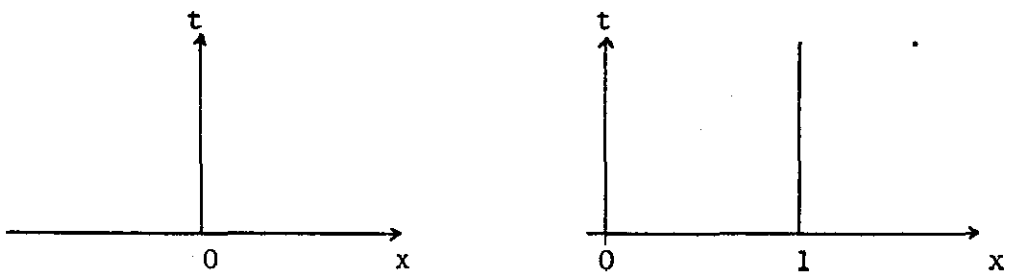
A linear parabolic equation is often written in the alternative form,

$$\frac{\partial u}{\partial t} = f(t,x,\frac{\partial u}{\partial x},\frac{\partial^2 u}{\partial x^2}) \quad . \tag{2.1.1}$$

As a typical example of a parabolic equation, consider the diffusion or heat conduction problem, which in general may be given in self-adjoint form,

$$a(x,t,u)\frac{\partial u}{\partial t} = g(x,t,u)+\frac{\partial}{\partial x}[k(x,t,u)\frac{\partial u}{\partial x}] \quad , \tag{2.1.2}$$

where $a(x,t,u)$ is the heat capacity, $g(x,t,u)$ is the source term and $k(x,t,u)$ is the conductivity. The domain of solution for a parabolic equation usually has one of the forms illustrated below.



$(t,x) \in (0,\infty)\times(-\infty,\infty)$ (a)

$(t,x) \in (0,\infty)\times(0,1)$ (b)

FIGURE 2.1.1

The case (a) leads to a purely initial value (Cauchy) problem, with initial values given as:

$$u(x,0) = f(x) \text{ for } t=0 \text{ and } -\infty<x<\infty \quad . \tag{2.1.3}$$

The second case (b) is an initial-boundary value problem, where the initial and boundary values are,

$$\begin{aligned} &u(x,0) = f(x) \text{ at } t=0, \text{ and } 0\leq x\leq 1 \text{ (initial condition)} \\ &a_1(x,t)u+b_1(x,t)\frac{\partial u}{\partial x} = c_1(x,t) \text{ at } x=0 \text{ and } t>0 \\ &a_2(x,t)u+b_2(x,t)\frac{\partial u}{\partial x} = c_2(x,t) \text{ at } x=1 \text{ and } t>0 \end{aligned} \tag{2.1.4}$$

The usual conditions satisfied by the coefficients are:

$$a_1\geq 0,\ b_1\leq 0 \text{ and } a_2\geq 0,\ b_2\geq 0 \ , \tag{2.1.5}$$

(Mitchell, A.R., 1976, p.18)

As a specific example, let us consider the case of diffusion or linear heat flow in one space dimension. If x denotes a coordinate along the length of a thin insolated bar in which heat can flow, and if u(x,t) denotes the temperature at position x, time t, the temperature satisfies the differential equation,

$$a(x,t)\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}[k(x,t)\frac{\partial u}{\partial x}] \tag{2.1.6}$$

where a(x,t) is the heat capacity of the material per unit volume, and k(x,t) is the thermal conductivity, (Richtmyer & Morton, 1967, p.4).

The equation (2.1.6) is linear with variable coefficients, but if we allowed a(x,t) and k(x,t) to vary with the temperature (a=a(x,t,u), k=k(x,t,u), then (2.1.6) becomes a non-linear equation. However, if a and k are constant coefficients, we obtain the simplest non-trivial member of the class (2.1.1) which is frequently used for analytical study, i.e.,

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad . \tag{2.1.7}$$

The analytical solution of (2.1.7) may be found by separation of variables. Assuming a solution of the form u(x,t)=X(x).T(t) leads to,

$$\frac{1}{T} \cdot \frac{\partial T}{\partial t} = \frac{1}{X} \cdot \frac{\partial^2 X}{\partial x^2}$$

A solution of this equation is $u(x,t)=e^{isx}.e^{-s^2t}$ which reduces to e^{isx} for t=0. (Stephenson, G., 1974, p.47).

This result indicates that, the solution has an exponential decay component in the t-direction. However, if a source term is present, the characteristics of the solution may be different. We shall be considering such a problem later when we are dealing with non-linear equations.

2.2 FINITE DIFFERENCE DISCRETIZATIONS

The finite-difference method for the solution of partial differential equations is based on the use of finite-difference approximations for derivatives. It consists of three steps:

i) The solution domain is divided into a grid of "node" points. This grid is uniformly spaced, and its shape reflects the nature of the problem and its boundary conditions.

ii) The governing partial differential equation is written in terms of the most convenient coordinate system available and is transformed into a partial differential equation by means of finite-difference approximations to the derivatives involved.

iii) Obtain the solution of difference equations at all the grid points by a suitable method.

Although this three-phase process may seem simple, considerable variation in grid types, grid sizes, partial differential equations, finite-difference approximations to these equations, the problem of consistency in approximating the continuous problem by a discrete problem, convergency of this approximate solution to the exact solution makes the topic of computer solution of partial differential equations an extremely diverse and interesting study.

2.3 INTRODUCTORY DISCRETIZATION THEORY

In general, to solve an initial boundary value problem, one associates a discrete problem to it, which can be solved by elementary algebraic manipulations (Van der Houwen, 1968, p.4). Hence we define the discrete analogue of an initial-boundary value problem.

Consider the equation (2.1.7) in a domain of the following shape,

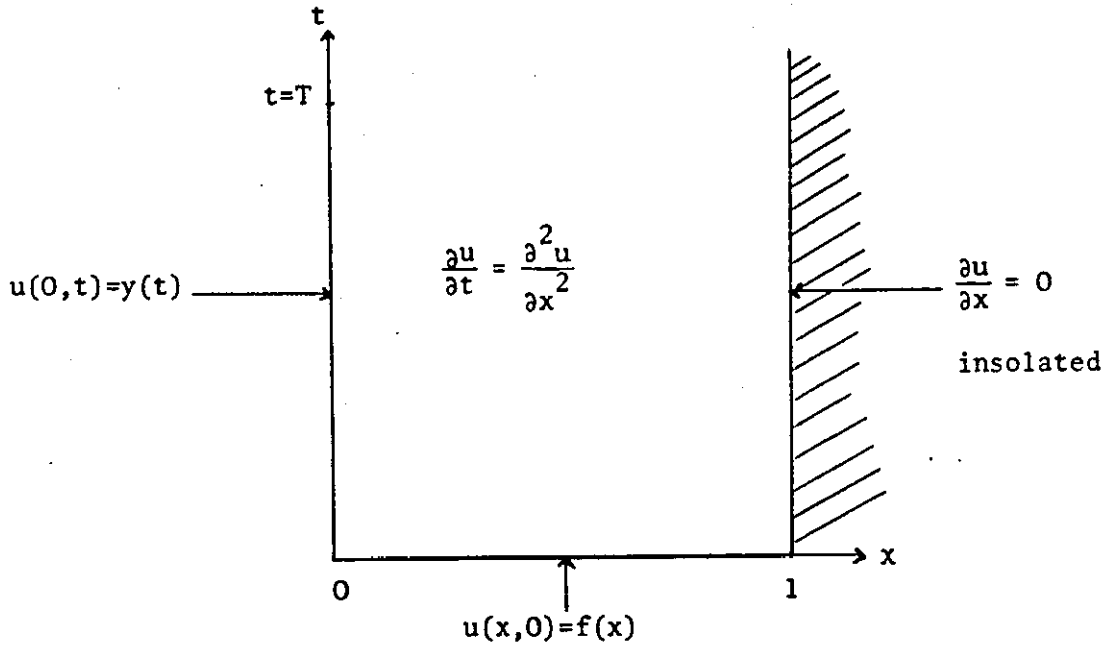


FIGURE 2.3.1

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

initial values: $u(x,0) = f(x) \quad t=0, \quad 0 \leq x \leq 1$, (2.3.1)

boundary values: $u(0,t) = g(t) \quad t>0, \quad x=0$

$\frac{\partial u(1,t)}{\partial x} = 0 \quad t>0, \quad x=1$ (2.3.2)

(N.B. $\frac{\partial u(1,t)}{\partial x} = 0$, means no heat flow at the boundary x=1).

First we replace the continuous interval [0,T] by a discrete set $\{t_j \mid 0=t_0<t_1<\ldots<t_M=T\}$ where we denote Δt_j and Δt by:

$$\Delta t_j = t_{j+1} - t_j \quad j=0,1,\ldots,M-1$$

$$\Delta T = \max_j \Delta t_j \quad 0 \leq j \leq M-1.$$

Together with the set of $\{t_j\}_{j=0}^{M}$, we take a finite set of points $\{x_i | 0=x_0<x_1<\ldots<x_N=1\}$ and in the same way as ΔT, we define

$$\Delta x_i = x_{i+1}-x_i \qquad i=0,1,\ldots,N-1,$$

$$\Delta x = \max_i \Delta x_i \qquad 0\leq i\leq N-1.$$

The set of nodes composed of the intersection of the set $\{t_j\}$ and $\{x_i\}$ creates a rectangular grid which we denote by $D^{(h)}$ and it is illustrated in Figure 2.3.2.

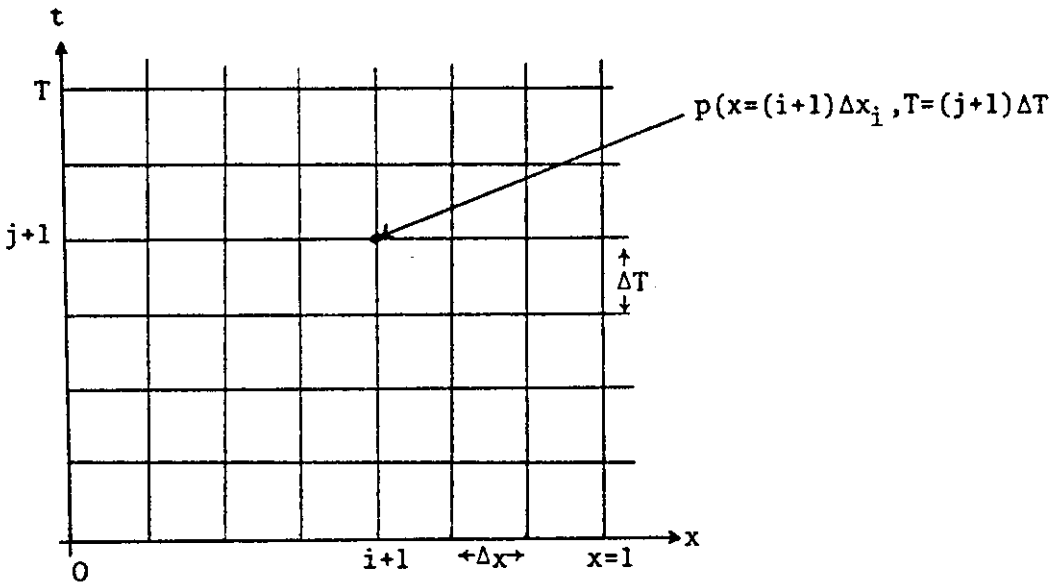


FIGURE 2.3.2

The rectangular grid $D^{(h)}$ is the most commonly used grid system for partial differential equations. However, in many engineering problems where other coordinates rather than Cartesian are imposed (e.g. cylindrical, polar, spherical), a grid of different type (e.g. circular, triangular, etc.) might be more applicable.

In our analysis, we concentrate attention on the first type of grid considered above (Figure 2.3.2), and without loss of generality, we denote Δx_i and Δt_j to be constant which are defined by h and k respectively.

The next step is to obtain an approximation form associated with the partial differential equation. Let $u_i^j=u(i\Delta x,j\Delta T)$ be the exact value at the point $(X=i\Delta x, T=j\Delta T)$, where the approximation to u_i^j is denoted by U_i^j. One of the simplest difference equations approximating the differential equation (2.1.7) is,

$$\frac{U_i^{j+1}-U_i^j}{\Delta T} = \frac{U_{i-1}^j-2U_i^j+U_{i+1}^j}{\Delta x^2} \qquad \begin{matrix} i=1,2,\ldots,N-1 \\ j=1,2,\ldots,M-1 \end{matrix} \tag{2.3.3}$$

where $U_i^0=f(i\Delta x)$, U_0^j and U_N^j are given.

To find (2.3.3) we use Taylor's expansion, assuming sufficient differentiability for the solution at the point $(X=i\Delta X,T=j\Delta T)$ with respect to x and t:

$$\begin{aligned} u(x+h,t) &= u(x,t)+h\frac{\partial u}{\partial x}+\frac{h^2}{2!}\frac{\partial^2 u}{\partial x^2}+\frac{h^3}{3!}\frac{\partial^3 u}{\partial x^3}+\frac{h^4}{4!}\frac{\partial^4 u}{\partial x^4}+O(h^5) \\ u(x-h,t) &= u(x,t)-h\frac{\partial u}{\partial x}+\frac{h^2}{2!}\frac{\partial^2 u}{\partial x^2}-\frac{h^3}{3!}\frac{\partial^3 u}{\partial x^3}+\frac{h^4}{4!}\frac{\partial^4 u}{\partial x^4}+O(h^5) \\ u(x,t+k) &= u(x,t)+k\frac{\partial u}{\partial t}+\frac{k^2}{2!}\frac{\partial^2 u}{\partial t^2}+\frac{k^3}{3!}\frac{\partial^3 u}{\partial t^3}+O(k^4) \ . \end{aligned} \tag{2.3.4}$$

Therefore, by some algebraic manipulation we obtain:

$$\begin{aligned} u(x+h,t)-2u(x,t)+u(x-h,t) &= h^2\frac{\partial^2 u}{\partial x^2}+\frac{h^4}{12}\frac{\partial^4 u}{\partial x^4}+O(h^6) \\ u(x,t+k)-u(x,t) &= k\frac{\partial u}{\partial t}+\frac{k^2}{2!}\frac{\partial^2 u}{\partial t^2}+\frac{k^3}{3!}\frac{\partial^3 u}{\partial t^3}+O(k^4) \end{aligned} \tag{2.3.5}$$

where as mentioned before, $h=\Delta x$ and $k=\Delta T$.

Using the result (2.3.4) and (2.3.5) in the equation (2.1.7), and the abbreviations $u_i^j, u_{i+1}^j, u_{i-1}^j, \ldots$ for $u(x,t)$, $u(x+h,t), u(x-h,t), \ldots$ we have,

$$\begin{aligned} u_i^{j+1} = (1-2\,k/h^2)u_i^j+\frac{k}{h^2}(u_{i+1}^j+u_{i-1}^j)+k\left(\frac{\partial u}{\partial t}-\frac{\partial^2 u}{\partial t^2}\right)+\frac{k^2}{2}\left(\frac{\partial^2 u}{\partial t^2}-\frac{h^2}{6k}\frac{\partial^4 u}{\partial x^4}\right) \\ +\frac{k^3}{3!}\left(\frac{\partial^3 u}{\partial t^3}-\frac{h^4}{60k^3}\frac{\partial^6 u}{\partial x^6}\right)+\ldots \ . \end{aligned} \tag{2.3.6}$$

By simplification of (2.3.6), bearing in mind that $\left(\frac{\partial u}{\partial t}-\frac{\partial^2 u}{\partial x^2}\right)$ is zero in the given domain, we obtain

$$u_i^{j+1} = (1-2p)u_i^j + p(u_{i-1}^j + u_{i+1}^j) + O(k^2+kh^2), \qquad (2.3.7)$$

where $p=k/h^2$. In formula (2.3.7), the term $O(k^2+kh^2)$ states the order of the local truncation error with the principal part $\frac{k^2}{2}(\frac{\partial^2 u}{\partial t^2} - \frac{h^2}{6k}\frac{\partial^4 u}{\partial x^4})$ stated in (2.3.6).

The formula (2.3.3) is obtained from the truncated form of (2.3.7) where U_i^j denotes the approximate value of u_i^j, and it is called an *explicit* finite-difference scheme, since it expressed only one unknown value of U_i^{j+1} at the advanced time directly, in terms of known values of the previous time level (Figure 2.3.3).

The equation (2.3.3) is sometimes called a forward step finite difference equation.

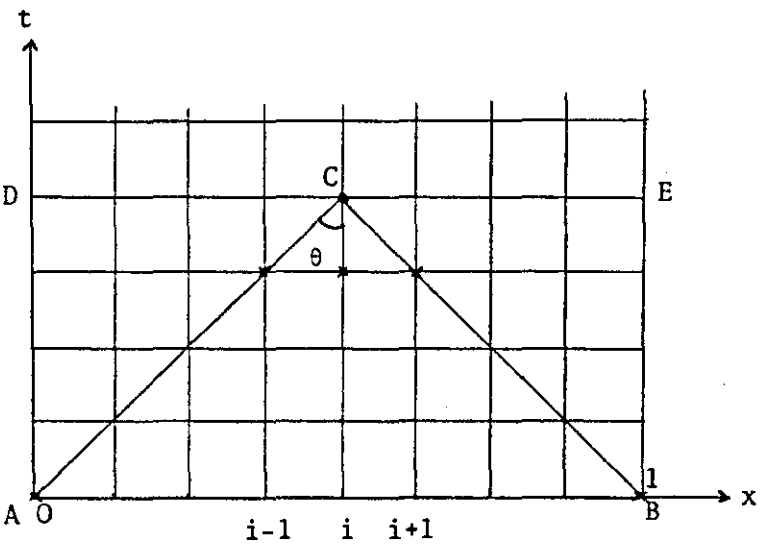


FIGURE 2.3.3

2.4 CONVERGENCY

By solving the purely algebraic system of equations (2.3.3) as an approximation of (2.1.7), two questions may arise:

1. Does the approximate solution approach the exact solution of the differential equation when h and k tend to zero?
2. Is the numerical difference scheme stable? In other words, what is the behaviour of the round-off errors when they are transmitted forward? Are they amplified or diminished during transmission?

To answer these questions, we shall consider the role of the *Local Truncation Error* (in abbreviation we call it L.T.E.).

Definition 2.4.1

The value $||u_i^j-U_i^j||$ which represents the difference between the theoretical solution of the differential and difference equation at a grid point (X=ih, T=jk), when $||.||$ is a suitable norm, is called the *discretization error*.

Definition 2.4.2

A difference scheme is said to be convergent, if the discretization error converges to zero as $h\to 0$ (h and k are related), (P.J. Van Der Houwen, 1968, p.8).

As an example of the convergence analysis for a difference formula, we turn to the explicit formula (2.3.3), introducing

$$z_i^j = u_i^j - U_i^j \tag{2.4.1}$$

The value of z_i^j satisfies a difference equation which can be obtained by subtracting (2.3.3) from (2.3.7) i.e.

$$z_i^{j+1} = (1-2p)z_i^j + p(z_{i-1}^j + z_{i+1}^j) + O(k^2+kh^2) \ . \tag{2.4.2}$$

If $\frac{\partial^2 u}{\partial t^2}$ and $\frac{\partial^4 u}{\partial x^4}$ in the principal part of L.T.E. remain bounded, we find by taking the modulus of (2.4.2) that,

$$|z_i^{j+1}| \leq p|z_{i+1}^j|+p|z_{i-1}^j|+(1-2p)|z_i^j|+A(k^2+k.h^2) \quad . \tag{2.4.3}$$

Let $||z^j||= \max_i |z_i^j|$ and suppose $0<p\leq 1/2$, then (2.4.3) becomes,

$$||z^{j+1}|| \leq ||z^j||+A(k^2+k.h^2) \quad .$$

Since the exact value u_i^j and the approximate value U_i^j have the same magnitude on the initial line, $||z^0||=0$, $i=1,2,\ldots,N-1$. Therefore,

$$||z^1|| \leq ||z^0||+A(k^2+kh^2) = A(k^2+kh^2)$$

$$||z^2|| \leq ||z^1||+A(k^2+kh^2) \leq 2A(k^2+kh^2)$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$||z^M|| \leq ||z^{M-1}||+A(k^2+kh^2) \leq MA(k^2+kh^2)$$

where k is denoted by T/M. Thus we obtain the result

$$\lim_{\substack{h\to 0\\ k\to 0}} ||z^M|| \leq \lim_{\substack{h\to 0\\ k\to 0}} TA(k+h^2) = 0 \tag{2.4.4}$$

Consequently, under the condition $0\leq p\leq 1/2$, the approximate solution (2.3.3) tends to the exact solution of the differential equation (2.1.7), upon unlimited refinement of the grid i.e. $h\to 0$ and $k\to 0$ (at a fixed point, which means k and h are related as $k=p.h^2$).

However, $0\leq p\leq 1/2$ is a severe restriction on the time step in the explicit formula, and it is a necessary condition for convergence. In a later section where the stability of a general formula is investigated we demonstrate how $p>1/2$ causes divergency.

2.5 CONSISTENCY

Here we concentrate on the condition for which a discrete problem is an approximation of the continuous problem. This is called the *consistency* problem. Let us first give a more general definition of an initial-boundary value problem.

Suppose we are required to find the solution of,

$$L(u) \equiv 0 \tag{2.5.1}$$

in a region $0=\{0<x<1\}\times\{0<t<T\}$, with initial condition

$$u = u_0(x) \quad \text{for } t=0 \tag{2.5.2}$$

and the boundary conditions

$$\begin{aligned} u(0,t) &= \phi_1(t) \quad x=0 \quad t>0 \\ u(1,t) &= \phi_2(t) \quad x=1 \quad t>0 \end{aligned} \tag{2.5.3}$$

where L is a linear operator (in (2.1.7), $L=\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}$. L could also be a non-linear operator). Let L_h denote an approximation to L on the grid net as shown in Figure 2.3.3.

Definition 2.5.1

The difference scheme $L_h(U)$ is said to be consistent with the initial-boundary value problem (2.5.1)-(2.5.3), if

$$||L_h(U)-L(u)|| \to 0 \text{ as } h\to 0 \tag{2.5.4}$$

at each point (X=ih,T=jk) in region, where $||.||$ is a suitable norm. The value $||L_h(U)-L(u)||$ is called the error of approximation.

In practice, consistency in the sense of definition (2.5.1) is easily verified. As an example, reconsider the initial-boundary value problem (2.1.7), and the finite difference scheme (2.3.3). Thus,

$$||L_h(U)-L(u)|| = ||(\frac{U_i^{j+1}-U_i^j}{k} - \frac{U_{i+1}^j-2U_i^j+U_{i-1}^j}{h^2})-(\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2})|| =$$

$$= ||\frac{k}{2}\frac{\partial^2 u}{\partial t^2} - \frac{h^2}{6}\frac{\partial^4 u}{\partial x^4}|| \to 0 \text{ as } h\to 0,\ k=0(h^2)\to 0,$$

provided $\frac{\partial^2 u}{\partial t^2}$ and $\frac{\partial^2 u}{\partial x^2}$ are bounded at every point of the region. Therefore, for consistency of the difference approximation to a parabolic equation, we require

$$\frac{\text{Local Truncation Error}}{k} \to 0 \text{ as } k,h\to 0.$$

As an example of an inconsistent difference replacement to (2.1.7), we examine the well-known Du-fort Frankel scheme which is obtained as follows:

$$\frac{U_i^{j+1}-U_i^{j-1}}{2k} = \frac{U_{i+1}^j-U_i^{j+1}-U_i^{j-1}+U_{i-1}^j}{h^2} \quad . \tag{2.5.5}$$

This is an explicit three level scheme which is stable (we show the stability proof later), and has

$$\text{L.T.E.} = \frac{1}{6}k^3\frac{\partial^3 u}{\partial t^3} - \frac{1}{12}k\,h^2\frac{\partial^4 u}{\partial x^4} + \frac{k^3}{h^2}\frac{\partial^2 u}{\partial t^2} + \ldots \quad . \tag{2.5.6}$$

Hence $\frac{\text{L.T.E.}}{k} \to (\frac{k}{h})^2 \cdot \frac{\partial^2 u}{\partial t^2}$ as $h\to 0$, $k=0(h)\to 0$, and if $\frac{k}{h}=\alpha$, the scheme (2.5.5) is not consistent with (2.1.7), but with the hyperbolic equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - \alpha^2\frac{\partial^2 u}{\partial t^2} \quad . \tag{2.5.7}$$

However, if $k\to 0$ faster than h (e.g. $k=0(h^2)$), then (2.2.5) is consistent with the diffusion equation (2.1.7).

The consistency of a difference scheme does not guarantee the convergence of the difference solution to the analytical solution.

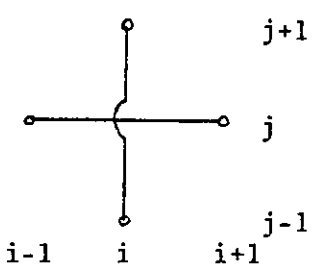


FIGURE 2.5.1

2.6 STABILITY

In the preceding sections we have discussed the convergency, and have given the conditions for which a finite difference scheme is convergent. In actual computations, however, one cannot construct the difference solution exactly, as one is faced with the phenomenon of round-off errors which give rise to a numerical solution $\tilde{U}_i^j$ instead of the exact solution U_i^j at any point (X=ih, T=jk), (P.J. Van Der Houwen, 1968, pp.8).

Definition 2.6.1:

The value $||U_i^j - \tilde{U}_i^j||$ i.e. the difference between the exact and numerical solution of a difference equation is called the numerical error.

In practice, one would of course, like this value to be small.

There are many definitions of stability (see P.J. Van Der Houwen, 1968, p.10) that are used in particular circumstances, but the basic idea is finding the conditions under which the numerical error, with increasing j, tends to zero uniformly for all $0 \leq i \leq N$ (or at any rate remain bounded). We shall be investigating these conditions below.

Together with the differential equation which is given in the operator form:

$$L(u) = f, \tag{2.6.1}$$

with some boundary conditions $\ell(u) = \phi$,

we consider the difference equation, which is applied to obtain the approximate solution, given in the symbolic form,

$$\begin{aligned} L_h(U^h) &= f^{(h)} \\ \ell_h(U^h) &= \phi^{(h)} \end{aligned} \tag{2.6.2}$$

where $U^{(h)}$ is the solution of the difference equation, L_h and ℓ_h are the difference operators associated with the difference scheme and the boundary condition respectively. The index (h) attached to $U^{(h)}$ and $f^{(h)}$ emphasizes

that these functions are not defined over the entire intervals but only at the points of the difference grid of step h (Godonov & Ryabenki, 1964, p.31). As we already assumed, k is defined in terms of h, so that the grid depends only on one parameter.

The basic idea of stability consists of considering a complete set of solutions $U=\{U_i^j\}_{i=1,2,\ldots,N-1}^{j=1,2,\ldots,M}$ originating from the discretization of the differential equation together with its boundary conditions and a perturbed set of equations with the solution $\tilde{U}=\{\tilde{U}_i^j\}_{i=1,2,\ldots,N-1}^{j=1,2,\ldots,M}$.

Suppose

$$L_h(U^{(h)}) = \alpha^{(h)} \quad , \qquad (2.6.3)$$

and

$$L_h(\tilde{U}^{(h)}) = \beta^{(h)} \quad ,$$

where $\alpha^{(h)}$ and $\beta^{(h)}$ are associated with the right-hand side of (2.6.2) and the discretization error of the boundary conditions and its perturbed form respectively. It is important to remember that, when h decreases, the number of equations increases.

We say the difference scheme is *stable* if in a suitable norm,

$$||U^{(h)}-\tilde{U}^{(h)}|| \leq C||\alpha^{(h)}-\beta^{(h)}||, \qquad (2.6.4)$$

for some constant C and for all values of h, $0<h<h_0$ (i.e., when the mesh is refined). In other words a difference scheme is stable, if small perturbations in the equations causes small perturbations in their solution uniformly (i.e. does not depend on the point considered) for all small h.

In this analysis, we considered the whole region of the problem. However, in an evolutionary problem, where step by step schemes are applied, it is more convenient to simplify the stability discussion to each step separately. To show this let the difference equation and the boundary condition (2.6.3) be reduced to the form given by,

$$\underline{U}^{j+1} = C(h)\underline{U}^j + k(h)\psi_h(\beta_h), \qquad (2.6.5)$$

$\underline{U}^0$ is given.

For equation (2.6.3) which now is given by (2.6.5) to be stable, we require,

$$||C(h)^m|| \leq K \ , \quad \text{for } m.k \leq T \tag{2.6.6}$$

(Godonov, Ryabenki, 1964, p.160) and the requirement usually is satisfied when

$$||\psi_h(\underline{\beta})-\psi_h(\underline{\alpha})|| \leq K_1||\underline{\beta}-\underline{\alpha}|| \ . \tag{2.6.7}$$

We shall apply this analysis to demonstrate the stability of a general form of two-step scheme in the next sections. We also consider multistep schemes (e.g. three level scheme) and define the amplification matrix which can be used for stability purposes.

2.7 A GENERAL FINITE-DIFFERENCE REPLACEMENT FOR PARABOLIC EQUATIONS

Before introducing the implicit finite-difference schemes, let us describe the domain of dependence for the solution of (2.1.7) when the explicit formula is applied for approximation.

As illustrated in Figure (2.3.3) the solution U_i^j at every point within the triangle ABC can be calculated provided the values on the initial line are known. Moreover, the knowledge of the boundary values on AB and BE is not required. Therefore, the explicit schemes act as a hyperbolic equation with two characteristics AC and BC. If we denote the slope of AC by $\tan\theta$ then,

$$\tan^{-1}\theta = \frac{h}{k} = \frac{1}{ph} \to \infty \quad \text{as } h\to 0,\ p=\text{const.} \tag{2.7.1}$$

i.e. $\theta\to\pi/2$ as $h\to 0$,

which simply means, the two characteristics AC and BE become the real characteristic DE of the parabolic equation. Thus, the explicit method for such initial boundary value problems requires small values of h and for large h does not give a good model for the parabolic equation. On the other hand, the requirement $p=k/h^2\leq 1/2$ for stability serves as a restriction which increases the computation involved.

Consequently, inspite of the simplicity of explicit method, it is desirable to establish a more efficient formula regarding the amount of computational work.

2.8 WEIGHTED AVERAGE FORMULA

We now consider a family of schemes as follows,

$$\frac{\partial u(x,t)}{\partial t} = \theta\left[\frac{\partial^2 u(x,t)}{\partial x^2}\right] + (1-\theta)\left[\frac{\partial^2 u(x,t)}{\partial x^2}\right] , \tag{2.8.1}$$

with finite-difference replacements:

$$\frac{U_i^{j+1} - U_i^j}{\Delta T} = \theta \frac{1}{h^2} \delta_x^2 U_i^j + (1-\theta) \frac{1}{h^2} \delta_x^2 U_i^{j+1}$$

where θ is a positive constant; $\theta=1$ gives the explicit scheme. Other values of θ give the implicit schemes. In particular, $\theta=0$ gives the fully implicit four-point scheme with backward time-step, and the six-points scheme with centred time-step, or the Crank-Nicolson form is obtained when $\theta=1/2$.

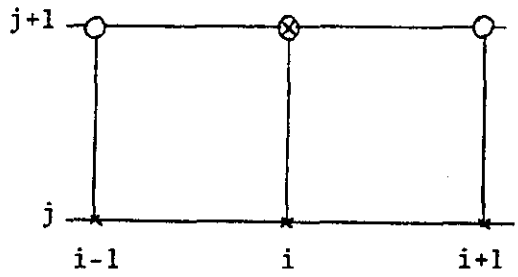


FIGURE 2.8.1

The L.T.E. of (2.8.1) can be easily verified by application of the Taylor's expansion and is of order $O(k+kh^2)$, with the principal part as,

$$\frac{h^2}{12} \frac{\partial^4 u}{\partial x^4} + \frac{k}{2} \cdot \frac{\partial}{\partial t}\left[\frac{\partial u}{\partial t} - 2\theta \frac{\partial^2 u}{\partial x^2}\right] + \theta \frac{k^2}{2}\left[\frac{\partial^4 u}{\partial^2 t \partial^2 x}\right] \tag{2.8.2}$$

In the case $\theta=1/2$, the second term vanishes and the L.T.E. becomes $O(k^2+h^2)$ which is one of the advantages of the Crank-Nicolson scheme.

2.9 STABILITY OF THE θ-METHOD

We can examine the stability of finite-difference schemes either by the Fourier method or by matrix analysis.

The Fourier method effectively ignores the boundary conditions and as they may affect the stability criterion, the matrix method is preferable (Fox, L., 1966, p.234). However, we shall consider first the Fourier analysis for the θ-formula, and obtain the result for θ=0,1/2 and 1. This method considers the growth of propagated errors of an initial line and expresses this error by Fourier series.

In the same manner, as finding the analytical solution of partial differential equations (separation of variables) we can determine the solution of the error propagation. (We use the complex exponential form rather than the sinusoidal form for simplicity).

Let $$A_n e^{\sqrt{-1}\, n\pi x} \equiv A_n\, e^{\sqrt{-1}\, i\, \beta_n h}$$

where $\beta_n=n\pi$ and A_n is the Fourier coefficient. Let E_i denote the error at each pivotal point on t=0, 0<x<1, therefore,

$$E_i = \sum_{n=0}^{N} A_n\, e^{\sqrt{-1}\, \beta_n ih} \quad i=0,1,\ldots,N. \tag{2.9.1}$$

The system (2.9.1) determines the (n+1) unknowns A_i uniquely. However, in the case of linear finite-difference schemes, A_n can always be ignored (Smith G.D., 1969, p.71) and we need to consider only the term $e^{\sqrt{-1}\beta_n ih}$.

To investigate the propagation of this error as t increases we need to find a solution of the finite-difference equation which reduces to $e^{\sqrt{-1}\, \beta_n ih}$ when t=0. Such a solution which is denoted by $E_{i,j}$ has the form,

$$E_{i,j} = e^{\sqrt{-1}\, \beta x}.\, e^{\alpha t} = e^{\sqrt{-1}\, \beta x}.e^{\alpha jk} = \xi^j\, e^{\sqrt{-1}\, \beta x} \tag{2.9.2}$$

where $\xi=e^{\alpha k}$, and α in general is a complex constant (Smith D.G., 1966, p.71). If $|\xi|\leq 1$ then the error is bounded as t increases.

Substituting (2.9.2) in the θ-formula, using central differences, gives

$$e^{\alpha jk}.e^{\sqrt{-1}\,\beta ih}(e^{jk}-1) = p.\theta.e^{-\beta ih}.e^{\alpha jk}(e^{\sqrt{-1}\,\beta h}.e^{\alpha k}-2e^{\alpha k}+e^{-\sqrt{-1}\,\beta h}.e^{\alpha k}) +$$

$$p(1-\theta)e^{\sqrt{-1}\,\beta ih}.e^{\alpha kj}(e^{\sqrt{-1}\,\beta h}-2+e^{-\sqrt{-1}\,\beta h})$$

$\therefore$

$$(e^{\alpha k}-1) = p.\theta.e^{\alpha k}(2\cos\beta h-2)+p(1-\theta)(2\cos\beta h-2)$$

or we can write,

$$e^{\alpha k}[1+2p\theta(1-\cos\beta h)] = [1+2p(1-\theta)(\cos\beta h-1)]$$

which leads to

$$e^{\alpha k} = \frac{1-4p(1-\theta)\sin^2\frac{\beta h}{2}}{1+4p\,\theta.\,\sin^2\frac{\beta h}{2}} \quad . \tag{2.9.3}$$

For stability we require $|e^{\alpha k}|\leq 1$. Therefore the stability of (2.8.1) depends on θ. If θ=0, we have the explicit (2.3.3), hence (2.9.3) becomes,

$$e^{\alpha k} = 1-4p\,\sin^2\frac{\beta h}{2} \quad ,$$

Thus stability is guaranteed if $p\leq 1/2$.

If θ=1, we obtain the implicit scheme and for p>0, the stability is unrestricted. For θ=1/2, we have,

$$e^{\alpha k} = \frac{1-2p\,\sin^2\frac{\beta h}{2}}{1+2p\,\sin^2\frac{\beta h}{2}}$$

which also implies that the Crank-Nicolson form is unconditionally stable.

2.10 STABILITY BY THE MATRIX METHOD

As we mentioned above, boundary conditions are ignored when the Fourier analysis is used for stability, while in matrix method, the effect of boundary values are included automatically.

We now consider the expanded form of the θ-formula (2.8.1), i.e.,

$$-p\theta U_{i-1}^{j+1}+(1+2p\theta)U_i^{j+1}-p\theta U_{i+1}^{j+1} = p(1-\theta)U_{i+1}^{j}+[1-2p(1-\theta)]U_i^{j}+p(1-\theta)U_{i-1}^{j}$$

$$i=0,1,\ldots,N. \qquad (2.10.1)$$

For i=0, we have

$$-p\theta U_{-1}^{j+1}+(1+2p\theta)U_0^{j+1}+p\theta U_1^{j+1} = p(1-\theta)U_1^{j}+[1-2p(1-\theta)]U_0^{j}+p(1-\theta)U_1^{j} \qquad (2.10.2)$$

and for i=N, (2.10.1) becomes,

$$-p\theta U_{N-1}^{j+1}+(1+2p\theta)U_N^{j+1}-p\theta U_{N+1}^{j+1}= p(1-\theta)U_{n+1}^{j}+[1-2p(1-\theta)]U_N^{j}+p(1-\theta)U_{N-1}^{j} \qquad (2.10.3)$$

Suppose the boundary values are of the mixed type, i.e.,

$$\begin{aligned} a_1u+b_1\frac{\partial u}{\partial x} &= c_1 \quad \text{at } x=0 \\ & \qquad\qquad t>0 \\ a_2u+b_2\frac{\partial u}{\partial x} &= c_2 \quad \text{at } x=1 \end{aligned} \qquad (2.10.4)$$

which can be replaced by

$$\begin{aligned} a_1U_0^{j+1}+b_1\frac{U_1^{j+1}-U_{-1}^{j+1}}{2h} = c_1 &\rightarrow U_{-1}^{j+1} = 2\frac{a_1}{b_1}h.U_0^{j+1}+U_1^{j+1}-2\frac{c_1}{b_1}.h \\ a_2U_N^{j+1}+b_1\frac{U_{N+1}^{j+1}-U_{N-1}^{j+1}}{2h} = c_2 &\rightarrow U_{N+1}^{j+1} = -2\frac{a_2}{b_2}h.U_{N-1}^{j+1}+U_{N-1}^{j+1}+2\frac{c_2}{b_2}.h \end{aligned} \qquad (2.10.5)$$

Therefore, for all the mesh points (2.10.1) can be written in the following matrix form,

$$\begin{bmatrix} 1+2p\theta(1-\frac{a_1}{b_1}h) & -2p\theta & & & \\ -p\theta & 1+2p\theta & -p\theta & & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & -p\theta \\ 0 & & & -2p\theta & 1+2p\theta(1+\frac{a_2}{b_2}h) \end{bmatrix} \begin{bmatrix} U_0 \\ U_1 \\ \vdots \\ U_{N-1} \\ U_N \end{bmatrix}^{j+1} =$$

$$\begin{bmatrix} 1-2p(1-\theta)(1-\frac{a_1}{b_1}h) & -2p(1-\theta) & & & \\ -p\theta & 1-2p(1-\theta) & -p(1-\theta) & & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & -p(1-\theta) \\ 0 & & & -p\theta & 1-2p(1-\theta)(1+\frac{a_2}{b_2}h) \end{bmatrix} \begin{bmatrix} U_0 \\ U_1 \\ \vdots \\ U_{N-1} \\ U_N \end{bmatrix}^{j} +2ph \begin{bmatrix} -\frac{c_1}{b_1} \\ 0 \\ \vdots \\ 0 \\ \frac{c_2}{b_2} \end{bmatrix}$$

(2.10.6)

The matrix equation (2.10.6) might be expressed in a more compact form as,

$$(I-p\theta B)\underline{U}^{j+1} = (I+p(1-\theta)B)\underline{U}^{j}+2ph.\underline{b} \qquad (2.10.7)$$

where

$$B = \begin{bmatrix} -2+\frac{a_1}{b_1} & 2 & & & \\ 1 & -2 & 1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ 0 & & & 2 & -2-\frac{a_2}{b_2} \end{bmatrix} \qquad (2.10.8)$$

and I is the unitary N×N matrix, $\underline{b}$ is the constant vector shown in (2.10.6).

It can be seen from (2.10.8) that diagonal dominancy of B requires the conditions (2.1.5) to be satisfied which was mentioned earlier.

In the case of the implicit form (i.e. $\theta \leq 1$), one is faced with a system of algebraic equations to solve. Fortunately, this system is of such a simple nature that a very elementary procedure i.e. the Gauss-elimimination method can be conveniently applied.

The formula (2.10.7) can also be represented by

$$\underline{U}^{n+1} = C\underline{U}^{n} + k.\ \psi(\underline{b}) \tag{2.10.9}$$

where $C=(I-p\theta B)^{-1}(I+p(1-\theta)B)$ and $\psi = \frac{2}{h}(I-p\theta B)^{-1}\underline{b}$ (as was mentioned in (2.6.5).

Consequently, for stability of (2.10.4) to be satisfied, we need the norm of C to not exceed the value unity. If the eigenvalues of C are denoted by λ_i and the eigenvalues of B by μ_i, then by the Theorem (1.7), we can write,

$$\lambda_i = \frac{1+p(1-\theta)\mu_i}{1-p\theta\mu_i}$$

For the matrix B and the Dirichlet boundary conditions, it is shown (Lowan, A.N., 1957, p.81) that, $\mu_i = -4\sin^2 \frac{i\pi h}{2N}$, $i=1,2,\ldots,N-1$, hence,

$$\lambda_i = \frac{1-4p(1-\theta)\sin^2 \frac{i\pi h}{2N}}{1+4p\theta\sin^2 \frac{i\pi h}{2N}} \tag{2.10.10}$$

which is an identical result to (2.9.3) obtained by Fourier analysis and hence, the stability is unrestricted for $\theta \leq 1/2$.

It is important to notice that, although the θ-formula is stable for $\theta \leq 1/2$, for any p, we still have to choose the step lengths h and k small enough to obtain a reasonable accuracy, i.e. to make the truncation error for the finite-difference method negligible (Walsh, J., 1966, p.109). Application of large time-step not only distrubs the accuracy, for some schemes (e.g. the Crank-Nicolson formula) but also causes a jump in the solution which is called the *noise effect*. This will be investigated later.

2.11 EQUATIONS OF INCREASED ACCURACY

The principal criteria of any finite-difference scheme, for the numerical approximation of an equation, are as follows:

1. Stability
2. Order of error of approximation
3. Simplicity (Saul,yev, 1964, p.83).

So far we have discussed the problem of stability and it has been shown that, stability almost always leads to convergency (or at least in some sense, e.g. Richtmyer & Lax stability gives convergence). However convergence may be of various degrees. Slow convergence requires more computational work and is impracticable whilst rapidly convergent processes need less arithmetical operations, hence are more desirable. Therefore, the speed or order of the convergency is important.

Let $L(u)\equiv 0$ denote a differential equation, with the corresponding finite-difference scheme $L_h(U_i^j)=0$. Then, for a sufficiently smooth function $v(x,t)$ satisfy the equation $L(v)=0$, the expression

$$L_h(v_i^j) = O(h^\sigma) \quad (\sigma>0) \tag{2.11.1}$$

denotes that the order of the error incurred by approximating the operator L by L_h at the node (ih,jk) is σ, for the class of function satisfying the equation $L(v)=0$ (Saul,yev, 1964,p.84).

As a general example, suppose $L(u)=0$ is the θ-formula with the finite-difference replacement (2.8.1). Therefore,

$$L_h(v_i^j) = \frac{h^2}{12}\frac{\partial^4 v_i^j}{\partial x^4} + \frac{k}{2}\frac{\partial}{\partial t}\left[\frac{\partial v_i^j}{\partial t} - 2\frac{\partial^2 v_i^j}{\partial x^2}\right] + \theta\frac{h^2}{2}\left[\frac{\partial^4 v_i^j}{\partial t^2 \partial x^2}\right] + \ldots = O(k+h^2) \tag{2.11.2}$$

If $k=O(h^2)$, then (2.11.2) indicates that the θ-formula has accuracy of $O(h^2)$. However, $\theta=1/2$ (i.e. the Crank-Nicolson scheme) gives rise to a more accurate form which is $O(k^2+h^2)$ for any k. Thus, as far as the accuracy is concerned, the Crank-Nicolson scheme is more desirable, simply

because one can apply a larger time-step and reduce the amount of computational work. But as was mentioned previously, large time-step causes the noise-effect which is an obvious drawback for the Crank-Nicolson scheme.

There are some alternative approaches to obtain accurate difference schemes, which can be expressed in the following classifications.

a) Implicit schemes with choice of certain parameters

b) Alternating methods and extrapolation

c) Multi-level schemes.

We shall explain these methods and present examples of finite-difference equations illustrating each of the aforesaid types.

In the first category we have:

i) Formulae with higher order central differences

One method for attaining a higher order of accuracy (i.e. larger σ) is based on the introduction of additional nodes in the approximation for the derivatives in the equation $L(u)\equiv 0$. As an example, consider the second derivative $\frac{\partial^2 u}{\partial x^2}$, which can be expressed as:

$$\frac{\partial u(x,t)}{\partial x^2} = \frac{\partial u_i^j}{\partial x^2} = \frac{1}{h^2}\left(\delta^2 u_i^j - \frac{1}{12}\delta^4 u_i^j + \frac{1}{90}\delta^6 u_i^j + \ldots\right) \qquad (2.11.3)$$

where $\delta^{2\ell} u_i^j$ ($\ell=1,2,\ldots$) are the standard central differences of even order.

Substitution of (2.11.3) in (2.1.7) results,

$$\frac{u_i^{j+1}-u_i^j}{k} = \frac{1}{h^2}\left(\delta^2 - \frac{1}{12}\delta^4 + \frac{1}{90}\delta^6 \ldots + 2\frac{(-1)^{\ell-1}[(\ell-1)!]^2}{(2\ell)!}\delta^{2\ell}\right)u_i^j \qquad (2.11.4)$$

The formula (2.11.4) reduces to the simple explicit scheme when $\ell=1$. For $\ell>1$, one needs more information to match up the central differences form near the boundaries. However, the use of (2.11.4) indicates that, the order of accuracy for $L(u)=0$ is $O(k+h^{2\ell})$ but it worsens the stability restriction in the case of the explicit scheme i.e. for stability, it is now required,

$$k \leq \frac{h^2}{2(1+\frac{4}{12}+\frac{16}{90}+\ldots)} \quad \text{(Saul,yev, 1964, p.89)} \qquad (2.11.5)$$

In the implicit case, one obtains,

$$U_i^{j+1} - \frac{k}{h^2}(\delta^2 - \frac{1}{12}\delta^4 + \frac{1}{90}\delta^6 \ldots)U_i^{j+1} = U_i^j \quad , \qquad (2.11.6)$$

which retains the unconditional stability of the method.

ii) Formulae with choice of parameters

Consider first, two unsymmetric explicit formulae of order O(h) which have been introduced by Saul,yev, 1964, p.31

$$U_i^{j+1} = \frac{1}{\omega+\alpha}\left[\alpha U_{i-1}^{j+1} + (1-\alpha)U_{i-1}^j + U_{i+1}^j - (2-\omega-\alpha)U_i^j\right] , \qquad (2.11.7)$$

and

$$U_i^{j+1} = \frac{1}{\omega+\alpha}\left[\alpha U_{i+1}^{j+1} + (1-\alpha)U_{i+1}^j + U_{i-1}^j - (2-\omega-\alpha)U_i^j\right] , \qquad (2.11.8)$$

where $\omega=1/p=h^2/k$, $0\leq\alpha\leq1$.

The template of (2.11.7) and (2.11.8) are shown in Figure (2.11.1) in (a) and (b) respectively.

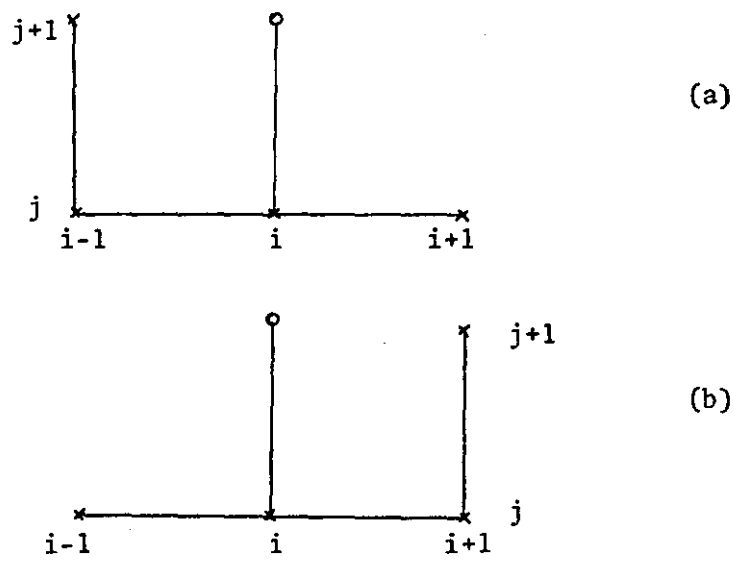


FIGURE 2.11.1

As can be seen (2.11.1a) can be applied from left to right and (2.11.1b) from the opposite direction.

A combination of (2.11.1a) and (2.11.1b) gives rise to an implicit equation of the form,

$$2h^2 L_h(U_i^j) \equiv -\alpha(U_{i-1}^{j+1}+U_{i+1}^{j+1})+2(\omega+\alpha)U_i^{j+1}-(2-\alpha)(U_{i-1}^{j}+U_{i+1}^{j})+2(2-\omega-\alpha)U_i^{j} = 0 \tag{2.11.9}$$

The equation (2.11.9) is equivalent to the θ-form for $\alpha=2\theta$.

Theorem 2.1

If the solution of (2.1.7) has derivatives up to eight order which are bounded in absolute magnitude throughout D, then the following relations hold in D:

$$L(u_i^j)-L_h(u_i^j) = \begin{cases} O(h^2) & \text{if} \quad \alpha\neq 1-\omega/6, \quad \alpha\geq 1-\omega/2 \\ O(h^4) & \text{if} \quad \alpha=1-\omega/6, \quad \alpha\neq 2\sqrt{5} \\ O(h^6) & \text{if} \quad \alpha=1-\omega/6, \quad \alpha=2\sqrt{5} \end{cases} \tag{2.11.10}$$

where u is the solution of (2.1.7) and $L(u_i^j)$ and $L_h(u_i^j)$ are the differential and difference expressions respectively.

Proof

In the proof, we shall consider two subjects:

i) Stability

ii) Accuracy.

For the investigation of the stability of (2.11.9), we use a matrix method. Hence we write (2.11.9) in the following form,

$$A_1\underline{U}^{j+1} = A_2\underline{U}^{j} \rightarrow \underline{U}^{j+1} = A_1^{-1}.A_2\underline{U}^{j} \tag{2.11.11}$$

where $A_1=-\alpha B+2\omega I$ and $A_2=(2-\alpha)B+2\omega I$, B is a triangular matrix as given before and I is the unitary matrix.

Since the eigenvalues of B are $\lambda_\ell(B)=-4\sin^2\frac{\ell\pi}{2N}$ $\ell=1,2,\ldots,N-1$ therefore,

$$\lambda_\ell(A_1) = 2\omega + 4\sin^2\frac{\ell\pi}{2N}$$

and $$\lambda_\ell(A_2) = 2\omega + 4(\alpha-2)\sin^2\frac{\ell\pi}{2N} \quad \ell=1,2,\ldots,N-1.$$

For the stability of (2.11.9) it is sufficient that,

$$\left|\frac{\lambda_\ell(A_2)}{\lambda_\ell(A_1)}\right| = \left|\frac{\omega+2(\alpha-2)\sin^2\frac{\ell\pi}{2N}}{\omega+2\alpha\sin^2\frac{\ell\pi}{2N}}\right| \leq 1, \tag{2.11.12}$$

or

$$-\omega-2\alpha\sin^2\frac{\ell\pi}{2N} \leq \omega+2(\alpha-2)\sin^2\frac{\ell\pi}{2N} \leq \omega+2\alpha\sin^2\frac{\ell\pi}{2N} \ . \tag{2.11.13}$$

The right hand side is always fulfilled and for the left hand side, we obtain

$$2(1-\alpha)\sin^2\frac{\ell\pi}{2N} \leq \omega \to 2(1-\alpha) \leq \omega \to k \leq \frac{h^2}{2(1-\alpha)} \tag{2.11.14}$$

which is the criterion for the stability of (2.11.9).

Thus, for any value of α, we have a specific value for ω to satisfy the stability condition.

For *accuracy* of (2.11.9), we can apply the Taylor series expansion in the usual manner to replace the derivatives involved and having done that, the following result can be obtained,

$$L_h(U_i^{j+1})-L(u_i^{j+1}) = -\frac{h^2(6-6\alpha-\omega)}{12\omega}\frac{\partial^2 u_i^{j+1}}{\partial x^2} - \frac{h^4(120+\omega^2-30\omega+15\alpha\omega-90\alpha)}{360\omega}$$

$$\frac{\partial^3 u_i^{j+1}}{\partial t^3} + O(h^6) \tag{2.11.15}$$

This is the formula to express the order of accuracy, which in general, is $O(h^2)$. However, if $6-6\alpha-\omega=0$, or $\alpha=1-\omega/6$, then the accuracy of (2.11.9) increases to $O(h^4)$. Finally if for such α, we choose ω in order the following equation holds,

$$120+\omega^2-30\omega+15\alpha\omega-90\alpha = 0$$

i.e. $\omega=2\sqrt{5}$, then the accuracy of $O(h^6)$ can be achieved.

Comment on formula (2.11.9)

1. For $\alpha=0$, and $\omega=6$, an explicit equation of the form

$$U_i^{j+1} = \frac{1}{6}(U_{i-1}^j + U_{i+1}^j) + \frac{2}{3} U_i^j \quad , \tag{2.11.16}$$

is obtainable with an accuracy of $O(h^4)$.

2. The stability restriction for the usual explicit equation is deduced from (2.11.14).

3. For $\alpha=1-\omega/6$, the equation coincides with the Douglas method having an accuracy of $O(h^4)$ if $\omega\neq 2\sqrt{5}$ and $O(h^6)$ where $\omega=2\sqrt{5}$.

4. The identity $\dfrac{\partial^{\ell} u}{\partial t^{\ell}} = \dfrac{\partial^{\ell+3} u}{\partial^{\ell-\beta} t \partial^{2\beta} x}$ $\ell=1,2,\ldots$ $\beta=1,2,\ldots$ has been applied to Theorem (2.1) whilst it can not be extended directly to the variable coefficient case or the multidimensional equations.

b: Alternating Methods and Extrapolation

As was mentioned earlier, the usual explicit and implicit finite-difference replacements for a parabolic equation are accurate of $O(k+h^2)$; i.e.,

$$\frac{u_i^{j+1}-u_i^j}{k} = \frac{u_{i+1}^j - 2u_i^j + u_{i-1}^j}{h^2} + O(k+h^2) \quad , \tag{2.11.17}$$

and

$$\frac{u_i^{j+1}-u_i^j}{k} = \frac{u_{i+1}^{j+1} - 2u_i^{j+1} + u_{i-1}^{j+1}}{h^2} + O(k+h^2) \quad . \tag{2.11.18}$$

If we use equations (2.11.17) at alternate levels (e.g. use (2.11.17) at odd levels and (2.11.18) at even levels), then the time-step k for even levels can be taken comparatively large while for the odd levels where (2.11.17) is applied, k should satisfy the stability condition $k/h^2 \leq 1/2$. However, these combinations can be shown by the following theorem to prove otherwise.

Theorem 2.2

The scheme

$$\frac{U_i^{2j+1}-U_i^{2j}}{k} = \frac{U_{i-1}^{2j+1}-2U_i^{2j+1}+U_{i+1}^{2j+1}}{h^2} \qquad (2.11.19)$$

$$\frac{U_i^{2j+2}-U_i^{2j+1}}{k} = \frac{U_{i-1}^{2j+1}-2U_i^{2j+1}+U_{i+1}^{2j+1}}{h^2} \qquad (2.11.20)$$

is absolutely stable, if the step k is constant or changes after an even number of steps (Saul,yev, 1964, p.24).

Proof

It is easy to show that the coefficient of stability for (2.11.19) is

$$(1+\frac{4k}{h^2}\sin^2\frac{\ell\pi h}{2})^{-1}$$

and for (2.11.20)

$$(1-\frac{4k}{h^2}\sin^2\frac{\ell\pi h}{2}) \qquad \ell=1,2,\ldots,N-1.$$

Accordingly, for any paired step, the coefficient of stability becomes:

$$\frac{1-4p\sin^2\frac{\ell\pi h}{2}}{1+4p\sin^2\frac{\ell\pi h}{2}} \quad , \qquad (2.11.21)$$

which has a modulus smaller than unity for any value of k, h and ℓ.

One can also apply (2.11.19) and (2.11.20) at alternate *nodes* rather than alternate levels. This is the Implicitly-Explicit method proposed by Saul,yev (1964) & P. Gordon (1965), and it was called "Hopscotch" later by A.R. Gourlay (1970), where in his paper, he investigates the full analysis of the method and shows that it is a second order, fast algorithm for solving partial differential equations. We shall be considering this method in detail later.

Here, we show different combinations of (2.11.17) and (2.11.18) which include some second order schemes. A typical example is obtained by taking

average of (2.11.17) and (2.11.18), i.e.,

$$\frac{U_i^{j+1}-U_i^j}{k} = \frac{1}{2}\left(\frac{U_{i+1}^j-2U_i^j+U_{i-1}^j}{h^2} + \frac{U_{i+1}^{j+1}-2U_i^{j+1}+U_{i-1}^{j+1}}{h^2}\right) \qquad (2.11.22)$$

which is the well-known Crank-Nicolson scheme and as mentioned earlier, is accurate to $O(k^2+h^2)$.

The summation of (2.11.19) and (2.11.20) also gives the following scheme,

$$\frac{U_i^{2j+2}-U_i^{2j}}{2k} = \frac{U_{i-1}^{2j+1}-2U_i^{2j+1}+U_{i+1}^{2j+1}}{h^2} \qquad (2.11.23)$$

which for convenience we shall write in the form (j odd),

$$\frac{U_i^{\ell+1}-U_i^{\ell-1}}{2h} = \frac{U_{i-1}^{\ell}-2U_i^{\ell}+U_{i+1}^{\ell}}{h^2} \qquad \ell=2j+1. \qquad (2.11.24)$$

This is the Richardson scheme and has an accuracy of $O(k^2+h^2)$ but surprisingly it is absolutely unstable. (See next section).

The extrapolation method also has been applied over implicit formula (2.11.18) in a recent paper by J.D. Lawson and J.Ll.Morris (1977), where the authors use a different analysis. Here we shall consider their method (but before, we pay attention to the Crank-Nicolson scheme and describe the *noise effect*).

First, let us consider the heat equation in one dimensional space,

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$$

where the initial and boundary values are provided in the usual way and a uniform grid is imposed in the region as before.

The replacement of the second order derivatives by the central-difference operator gives rise to the following differential equation

$$\frac{du}{dt} = \frac{\delta^2 x}{h^2}\, u(x,t) + O(h^2) \ . \qquad (2.11.25)$$

Equation (2.11.25) is then applied to all the interior mesh points to produce a system of ordinary differential equations

$$\frac{d}{dt}\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{N-1} \end{bmatrix} = \frac{1}{h^2}\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ 0 & & & 1 & -2 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{N+1} \end{bmatrix} + O(h^2) \qquad (2.11.26)$$

or

$$\frac{d\underline{u}}{dt} = A\underline{u}(t) \ , \qquad (2.11.27)$$

where $\underline{u}$ is the approximation vector which corresponds to the exact solution $\underline{u}(x,t)$. The solution of (2.11.27) is found to be,

$$\underline{u} = \exp(t.A)\underline{u}(0) \qquad (2.11.28)$$

where $\underline{u}(0)=\underline{f}(x)$ is the initial vector.

One can also write (2.11.28) in a step-wise fashion as

$$\underline{u}(t+k) = \exp(k.A)\underline{u}(t) \qquad t=K,2k,... \qquad (2.11.29)$$

The Padé approximation may be applied to obtain the value $\underline{u}(t+k)$ and in particular the (1,1) Padé approximation leads to the following replacement of (2.11.29),

$$\underline{u}(t+k) = \frac{I+\frac{1}{2}kA}{I-\frac{1}{2}kA}\underline{u}(t) \qquad (2.11.30)$$

which represents the familiar Crank-Nicolson scheme.

If we assume that the eigenvalues of A are λ_i ($\lambda_i<0$) and the corresponding eigenvectors are ω_i, then it is an easy matter to write the initial vector $f(x)=\sum_{i=1}^{N-1} \alpha_i\omega_i$ and the theoretical solution of (2.11.30) is

$$u(x,t) = \sum_{i=1}^{N-1} \alpha_i \left[\frac{1+k/2\ \lambda_i}{1-k/2\ \lambda_i}\right]^n \omega_i \qquad (2.11.31)$$

where

$$\lambda_i = -\frac{4}{h^2}\sin^2\left[\frac{i\pi}{2N}\right] \quad i=1,2,\ldots,N-1.$$

The growth factor $(1-k/2\ \lambda_i)^{-1}(1+k/2\ \lambda_i)$ has always a modulus smaller than unity for any value of k which indicates no restriction on the stability. However, if k takes large values then this factor becomes -1, while $\exp(n.k.\lambda_i)\to 0$. This is the origin of the Crank-Nicolson noise effect and the reason why the method may be called marginally stable.

This will also be the case if h is small for those λ_i corresponding to i=N-1,N-2,..., since for example, $\sin^2\left[\frac{(N-1)\pi}{2N}\right]\approx 1$ and $\lambda_{N-1}\approx -\frac{4}{h^2}$ (i.e. $k/2\ \lambda_i$ is large). Accordingly, one may expect to see components $\alpha_{N-1}\omega_{N-1}, \alpha_{N-2}\omega_{N-2}, \cdots$ of the initial condition preserved at subsequent solution steps, but with alternating sign (Lawson, J.D. and Morris, J.Ll. 1977).

There will be no oscillation if k is kept smaller than the critical value $\frac{2}{\max \lambda_i}$.

In contrast to proposing the (1,1) Padé approximation to exp(k,A), one may use the (1,0) Padé approximation to obtain fully implicit schemes,

$$(I-kA)\underline{u}(t+k) = \underline{u}(t) \qquad (2.11.32)$$

Thus, the growth factor in (2.11.31), in this case is always positive and hence there is no oscillatory behaviour. However, the equation (2.11.32) is only $O(k+h^2)$ i.e. first order accurate in time, and it requires more computational work to attain the same accuracy as that in the Crank-Nicolson scheme. If we apply (2.11.32) over a time-step 2k, we obtain

$$\underline{u}(t+2k) = (I-2kA)^{-1}\underline{u}(t) \qquad (2.11.33)$$

while by the use of (2.11.32) on two following time-step k, we have,

$$\underline{u}(t+2k) = (I-kA)^{-1}(I-kA)^{-1}\underline{u}(t) \qquad (2.11.34)$$

The expansion of the matrix inversion in (2.11.33) and (2.11.34) produce respectively,

$$\underline{u}(t+2k) = (I+2kA+4k^2A^2+8k^3A^3)\underline{u}(t)+O(k^4) \qquad (2.11.35)$$

$$\underline{u}(t+2k) = (I+2kA+3k^2+4A^3)\underline{u}(t)+O(k^4) \qquad (2.11.36)$$

On the other hand the Maclaurin expansion of exp(2kA) produces:

$$\underline{u}(t+2k) = (I+2kA+2k^2A^2+\tfrac{4}{3}k^3A^3)\underline{u}(t)+O(k^4) \ . \qquad (2.11.37)$$

By the combination of 2 times (2.11.36) and subtracting (2.11.35) we find

$$\underline{u}(t+2k) = (I+2kA+2k^2A^2)\underline{u}(t)+O(k^3) \qquad (2.11.38)$$

Equation (2.11.38) represents a second order approximation to the solution $\underline{u}(t+2k)$. It can be seen that the method is unconditionally stable and the growth factor tends to zero monotically.

c: Multi-level difference equations

The multi-level difference replacements are often used to construct a difference equation of higher accuracy than the minimum level scheme required by the differential equation.

For instance to improve the accuracy of the fully implicit scheme for the heat equation

$$\frac{u_i^{j+1}-u_i^j}{k} = \frac{u_{i+1}^{j+1}-2u_i^{j+1}+u_{i-1}^{j+1}}{h^2} + O(k) + O(h^2) \ , \qquad (2.11.39)$$

One can apply the following three level implicit scheme (Ritchmyer R.B. and Morton K.W., 1967 p.68),

$$\frac{\frac{3}{2}u_i^{j+1}-2u_i^j+1/2\ u_i^{j-1}}{k} = \frac{u_{i+1}^{j+1}-2u_i^{j+1}+u_{i-1}^{j+1}}{h^2} +O(k^2)+O(h^2). \qquad (2.11.40)$$

As we can see, one needs initial data on two levels (say t and t+k) to obtain the solution at t+2k. The extra initial information can be taken from a simple two level scheme at the start of the procedure.

A full discussion of multi-level schemes is given by Ritchmyer & Morton and also by Saul,yev. Here we only consider two well-known examples of three level schemes and refer the interested reader to the quoted references.

The first example is the explicit Richardson scheme,

$$\frac{u_i^{j+1}-u_i^{j-1}}{2k} = \frac{u_{i+1}^{j}-2u_i^{j}+u_{i-1}^{j}}{h^2} + O(k^2)+O(h^2) \qquad (2.11.41)$$

which can be written in the form

$$U_i^{j+1} = 2p\delta_x^2 U_i^j + U_i^{j-1} \quad . \qquad (2.11.42)$$

The stability of (2.11.42) can be examined either by Fourier analysis directly (Ames, W.1969, p.55) or by a splitting method (Mitchel, A.R., 1976, p.88). Here, we apply the last method and rewrite (2.11.42) as the two level system,

$$\begin{cases} U_i^{j+1} = 2p\delta_x^2 U_i^j + V_i^j \\ V_i^{j+1} = U_i^j \end{cases} \qquad (2.11.43)$$

which can be represented in vector form as,

$$\begin{bmatrix} U_i \\ V_i \end{bmatrix}^{j+1} = \begin{bmatrix} 2p\delta_x^2 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} U_i \\ V_i \end{bmatrix}^{j} \qquad (2.11.44)$$

If we introduce $W=\begin{bmatrix} U \\ V \end{bmatrix}$, and a typical Fourier term as

$$W_i^j = W_0^j \, e^{\sqrt{-1}\,\beta i h}$$

where W_0^j is a constant vector, and substitute into (2.11.44), the result,

$$W_i^{j+1} = \begin{bmatrix} -8p\sin^2\frac{\beta h}{2} & 1 \\ 1 & 0 \end{bmatrix} W_i^j$$

is obtained. The above matrix is called *amplification* matrix of the system with the eigenvalues

$$\lambda_i = -4p\sin^2\frac{\beta h}{2} \pm \left(1+16p^2\sin^2\frac{\beta h}{2}\right)^{1/2} \quad , \; i=1,2,.$$

where $p=k/h^2$.

Now for stability of the system (2.11.43), we require $|\lambda_i| \leq 1$, $i=1,2$. This condition violated by λ_2 for all values of p. Therefore the Richardson scheme is unconditionally unstable, if it is taken as a marching process.

If however, the initial-boundary value problem being considered is reduced to a pure boundary value problem, then the Richardson scheme becomes stable (Saul,yev, 1964, p.90). The last chapter of this thesis is concerned with this problem.

In contrast with (2.11.41) if we replace u_i^j by $\frac{1}{2}(u_i^{j+1}+u_i^{j-1})$ in $\delta_x^2 u_i^j$, then we obtain another well-known scheme which has been presented by Du-fort-Frankel,

$$\frac{u_i^{j+1}-u_i^{j-1}}{2k} = \frac{u_{i+1}^j - u_i^{j+1} - u_i^{j-1} + u_{i-1}^j}{h^2} + O[k^2+h^2+(\frac{k}{h})^2] \quad . \tag{2.11.45}$$

On designating $\underline{u}^j$ as the vector of mesh values along the j^{th} line, one can write the formula (2.11.45) in matrix form as

$$A\underline{U}^{j+1} = B\underline{U}^j + C\underline{U}^{j-1} \quad . \tag{2.11.46}$$

For the specific case under consideration, the matrices are:

$$A = (1+2p)I, \quad C = (1-2p)I \quad \text{and} \quad B = 2p\begin{bmatrix} 0 & 1 & & & \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & 1 \\ & & \ddots & \ddots & 0 \\ & & & 1 & 0 \end{bmatrix}$$

Now, we write (2.11.46) as

$$\underline{U}^{j+1} = A^{-1}B\underline{U}^j + A^{-1}C\underline{U}^{j-1} \quad . \tag{2.11.47}$$

Then it can be reduced to the two-level scheme

$$\underline{W}^{j+1} = M\underline{W}^j \quad \text{while} \quad \underline{W}^j = \begin{bmatrix} \underline{u}^j \\ \underline{v}^j \end{bmatrix} \quad \text{and} \quad M = \begin{bmatrix} A^{-1}B & A^{-1}C \\ I & 0 \end{bmatrix} \quad . \tag{2.11.48}$$

For stability, we require that the characteristic roots of M to be smaller than unity in modulus, and they can be found (upon using the definition of M) to be the zeros of the determinental equation

$$|\mu^2 A - \mu B - C| = 0 \quad . \tag{2.11.49}$$

Since the matrix $A^{-1}B$ and $A^{-1}C$ are commutative, (2.11.49) can be replaced by

$$\mu^2 - \lambda_i \mu - \eta_i = 0 \qquad (2.11.50)$$

where λ_i and η_i represent the eigenvalues of $A^{-1}B$ and $A^{-1}C$ respectively. Considering the definition of A,B and C, it is easy to show that μ is the root of

$$\mu^2 - \frac{4p}{1+2p} \cos \frac{i\pi}{N} \mu + \frac{1-2p}{1+2p} = 0 \quad i=1,2,\ldots,N-1. \qquad (2.11.51)$$

and clearly $|\mu| \leq 1$ for all p, thereby establishing stability.

However, this explicit second order scheme suffers from the term $\frac{k}{h}$ appearing in the L.T.E. (as seen before) which makes the finite-difference replacement inconsistent with the heat equation over large time step and one has to apply time-steps $k=0(h^{\alpha})$, $\alpha>1$.

Here we conclude that although it may appear in general that, concerning accuracy, three level formulae have an advantage over two-level schemes in the solution of parabolic equations it is possible that the introduction of an extra level may cause trouble in a particular sense (e.g. stability for Richardson, consistency for Du-Fort-Frankel).

2.12 NON-LINEAR EQUATIONS

Until now, we have considered only linear parabolic equations, but in actual fact, many scientific and engineering problems have a non-linear form. As an example, the heat conduction equation which has been used for analytical purposes has a more physically reasonable mode,

$$\rho c \frac{\partial u}{\partial t} = \frac{\partial}{\partial x}[K(u)\frac{\partial u}{\partial x}] \ , \qquad (2.12.1)$$

where the thermal conductivity term K(u) depends upon the temperature, ρ(density) and c (the specific heat) may also depend on u. Equation (2.12.1) is called a *self-adjoint* formula.

A more general case of a non-linear parabolic equation may be given as:

$$\frac{\partial u}{\partial t} = f(x,t,u,\frac{\partial u}{\partial x} , \frac{\partial^2 u}{\partial x^2}) \quad (\frac{\partial f}{\partial u_{xx}} > 0) \qquad (2.12.2)$$

in the region $D=\{0\leq x\leq 1\}\times\{0<t\leq T\}$, subject to the appropriate initial and (possibly-non-linear) boundary conditions.

Finite-Difference Replacement

One of the advantages of the finite-difference technique is that many of the methods and proofs, based on linear equations with constant coefficient carry over directly to non-linear equations (Mitchell, A.R., 1976, p.95). However, in this case both the numerical process and the analysis of stability and convergence become more complicated.

For non-linear problems stability depends not only on the form of the finite-difference approximation but also upon the values of the solution (i.e. we have local stability). The system may be stable for some values of t and not for others. As regards the approximation techniques, there exist two main difficulties.

If we use the explicit finite-difference scheme for a problem of type (2.12.2), we find it very easy to solve, but it suffers from the disadvantage

of restriction on the time-step to maintain stability. This limitation can be avoided by using an implicit-difference method (e.g. the Crank-Nicolson). In this case, we are faced with a non-linear system of equations to solve, and depending on f in (2.12.2), the algebraic problem of finding the solution may become difficult and one needs to use an iteration technique to evaluate the solution (Mitchell, A.R. 1976, pp.96).

Here, we investigate briefly some of the explicit and implicit methods and leave aside the iteration methods to the future.

The general θ-formula corresponding to (2.12.2) can easily be obtained. It is illustrated as follows:

$$U_i^{j+1}-U_i^j = k.f(i.h,j.k,\theta.U_i^j+(1-\theta)U_i^{j+1},\theta[\frac{U_{i+1}^j-U_{i-1}^j}{2h}, \frac{\delta_x^2 U_i^j}{h^2}], (1-\theta)[\frac{U_{i+1}^{j+1}-U_{i-1}^{j+1}}{2h}, \frac{\delta_x^2 U_i^{j+1}}{h^2}]) \qquad (2.12.3)$$

As was pointed out, for $\theta=1$, (2.12.3) becomes an explicit scheme, $\theta=0$ leads to the fully implicit scheme and $\theta=1/2$ is the usual Crank-Nicolson formula.

In the case of self-adjoint form, the formulation is rather different. We shall demonstrate some of the finite-difference replacements to (2.12.1).

The simplest difference approximation to (2.12.1) is

$$\rho(U_i^j).C(U_i^j)\ \frac{U_i^{j+1}-U_i^j}{k} = \frac{1}{h^2}\,\delta_x(K(U_i^j)\delta_x U_i^j)\ . \qquad (2.12.4)$$

or

$$\rho(U_i^j).c(U_i^j)\ \frac{U_i^{j+1}-U_i^j}{k} = \frac{1}{h^2}\{K(U_{i+\frac{1}{2}}^j).(U_{i+1}^j-U_i^j)-K(U_{i-\frac{1}{2}}^j).(U_i^j-U_{i-1}^j)\}. \qquad (2.12.5)$$

If we replace $K(u_{i+\frac{1}{2}}^j)$ and $K(u_{i-\frac{1}{2}}^j)$ by averaging over the interval, i.e.

$$K(\frac{U_{i+1}^j+U_i^j}{2}) \text{ and } K(\frac{U_i^j+U_{i-1}^j}{2})$$

respectively, then the latter expressions involve values of u at the grid, and the following result is obtainable:

$$\rho(U_i^j).c(U_i^j)\,\frac{U_i^{j+1}-U_i^j}{k} = \frac{1}{h^2}\{K(\frac{U_{i+1}^j+U_i^j}{2})(U_{i+1}^j-U_i^j)-K(\frac{U_i^j+U_{i-1}^j}{2})(U_i^j-U_{i-1}^j)\},$$

(2.12.5)

which is an explicit simulation for (2.12.1). In the same way, one can easily find the other difference scheme for the self-adjoint equation.

As we mentioned before, implicit replacement for non-linear partial differential equations gives rise to a system of non-linear difference equations. To establish such a system, we consider a non-linear parabolic equation of the form,

$$L(u) = f(x,t,u)\ , \tag{2.12.6}$$

where L is a *linear parabolic* partial differential operator. A simple example of (2.12.6) is denoted when $L \equiv \frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}$. If f(x,t,u) has polynomial form (2.12.6) is mildly non-linear, while an exponential form for f(x,t,u) gives a strongly non-linear problem.

The θ-form for (2.12.6) becomes:

$$U_i^{j+1}-U_i^j = k\{\theta[\frac{\delta_x^2}{h^2}U_i^j+f(ih,jk,U_i^j)]+(1-\theta)[\frac{\delta_x^2 U_i^{j+1}}{h^2}+f(ih,jk,U_i^{j+1})]\} \tag{2.12.7}$$

Denoting $p=k/h^2$ and expanding the central difference operators leads to the equation:

$$\begin{aligned}-\theta pU_{i-1}^{j+1}+(1+2p\theta)U_i^{j+1}-\theta pU_{i+1}^{j+1}-\theta k.f(U_i^{j+1}) &= p(1-\theta)U_{i-1}^j+[1-2p(1-\theta)]U_i^j \\ &\quad +p(1-\theta)U_{i+1}^j+k(1-\theta)f(u_i^j) \qquad (2.12.8)\\ &\qquad\qquad i=1,2,\ldots,N-1.\end{aligned}$$

Suppose the boundary conditions are given as:

$$\begin{aligned}\frac{\partial u}{\partial x} &= u \quad \text{at } x=0 \text{ for } t>0\\ \frac{\partial u}{\partial x} &= -u \quad \text{at } x=1 \text{ for } t>0\end{aligned} \tag{2.12.9}$$

with the given initial values $u(0,x)=1$, $0\leq x\leq 1$. By using central difference approximations to the boundary values we obtain:

$$u_{-1} = u_1 - 2hu_0 \qquad \text{and} \qquad u_{N+1} = u_{N-1} + 2ku_N \ . \tag{2.12.10}$$

By substituting (2.12.10) into (2.12.8) and arranging in matrix form, we find

$$[I-p(1-\theta)B]\underline{U}^{j+1} - k(1-\theta)f(\underline{U}^{j+1}) = [I+p\theta.B]\underline{U}^{j} + k\theta f(\underline{U}^{j}) \ , \tag{2.12.11}$$

$$B = \begin{bmatrix} 2(1+h) & -2 & & & \\ -1 & 2 & -1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ 0 & & & 2 & 2(1+h) \end{bmatrix}$$

To solve the non-linear system (2.12.11), one has to employ indirect (iterative) methods such as the Newton's or Secant iteration techniques.

2.13 PARABOLIC EQUATIONS IN SEVERAL SPACE VARIABLES

We now consider finite-difference methods of solution for the equation

$$\frac{\partial u}{\partial t} = L(u) \tag{2.13.1}$$

where $L \equiv \sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(a_i(x_1,x_2,\ldots,x_n,t)\frac{\partial}{\partial x_i}\right) - c(x_1,x_2,\ldots,x_n,t)$ is a second order elliptic operator with a strictly positive and c non-negative (Mitchel, A.R. 1976, p.45).

There are two categories of finite-difference methods for several space variable parabolic equations. Firstly the generalization of standard methods which are presented for one dimensional problems and secondly, splitting methods which have no single space variable analog.

i) Generalization of the standard methods

Consider the diffusion equation in two space variables which is a particular case of (2.13.1) namely,

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \phi(x,y) \quad . \tag{2.13.2}$$

Suppose the solution of (2.13.2) is required in the cylinder $R\times[0<t\leq T]$ where R is a closed connected region in the x-y plane, with continuous boundary ∂R. Appropriate initial and boundary data are provided as:

$$u(x,y,0) = f(x,y) \qquad t=0$$

$$u(x,y,t) = g(x,y,t) \qquad (x,y)\in\partial R \quad t>0;$$

where f and g are given for prescribed values of x,y,t.

In the same manner as the one dimensional case, the region R is covered by a rectilinear grid with sides, parallel to the axes, with Δx and Δy the grid spacings in the direction x and y and ΔT in the time direction.

The grid points are denoted by ($X=i\Delta x$, $Y=j\Delta y$, $T=n\Delta T$) where i,j and n are integers, and i=j=n=0 is the origin. The exact and approximate values of the solution at the grid points are denoted by $u^n_{i,j}$ and $U^n_{i,j}$ respectively.

If we substitute the expressions for $\frac{\partial^2}{\partial x^2}$ and $\frac{\partial^2}{\partial y^2}$ from the definition of central differences (where only second order terms are retained) into (2.13.2), we obtain the standard explicit replacement (for simplicity here we let $\phi(x,y)=0$):

$$U_{i,j}^{n+1} = [1+p(\delta_x^2+\delta_y^2)]U_i^j \qquad (2.13.3)$$

or

$$U_{i,j}^{n+1} = (1-4p)U_{i,j}^n+p(U_{i-1,j}^n+U_{i+1,j}^n+U_{i,j-1}^n+U_{i,j+1}^n) \qquad (2.13.4)$$

where we suppose $\Delta x=\Delta y=h$, $\Delta T=k$ and $p=k/h^2$.

Using Taylor series expansions and defining the difference between the exact solution of the differential and difference equations at the mesh point (ih,jh,nk) as

$$e_{i,j}^n = U_{i,j}^n-u_{i,j}^n \quad , \qquad (2.13.5)$$

then from (2.13.4) we obtain (Gane, C.R., 1976, p.23),

$$e_{i,j}^{n+1} = (1-4p)e_{i,j}^n+p(e_{i+1,j}^n+e_{i-1,j}^n+e_{i,j-1}^n+e_{i,j+1}^n) + \frac{h^2}{2}\left(\frac{\partial^2 u}{\partial t^2}\right)_{i,j}^n -$$

$$-k\,\frac{h^2}{12}\left(\frac{\partial^4 u}{\partial x^4}+\frac{\partial^4 u}{\partial y^4}\right)_{i,j}^n + \ldots \qquad (2.13.6)$$

Equation (2.13.6) shows that the L.T.E. of the formula (2.13.4) is $O(k^2+k.h^2)$ while the neighbouring points of (ih,jh,nk) are interior points of R, i.e. the partial derivatives of u are continuous and uniformly bounded for all x,y,t $R\times\{0<t\leq T\}$.

The stability restriction for (2.13.4) is now

$$\frac{\Delta T}{\Delta x^2}+\frac{\Delta T}{\Delta y^2} = \frac{2k}{h^2} \leq 1/2 \rightarrow p\leq 1/4 \quad . \qquad (2.13.7)$$

Such a severe restriction makes this method of doubtful practicability.

The backward difference scheme can also be generalized to permit an implicit finite difference approximation which has no stability restriction, but the resulting linear equations are no longer tridiagonal, which is not always easy to accomplish directly. Because of poor stability properties of the explicit scheme and the difficulty of handling the implicit methods, the splitting technique has been introduced.

ii) Splitting methods

In the numerical treatment of parabolic differential equations, splitting is referred to as a method of breaking down a complicated (multi-dimensional) process into a series of simple (one-dimensional processes). Well-known examples are the *Alternating Direction Implicit* (A.D.I. in abbreviation), the *Locally One Dimensional* (L.O.D.) and the *Hopscotch* methods.

The A.D.I. methods which were first introduced by Peaceman and Rachford (1955) and improved by Douglas and Rachford (1956) are a two-step process involving the solution of tridiagonal sets of equations along lines parallel to the x and y axes at the first and second steps respectively.

We illustrate this method with respect to the equation (2.13.2). The fully implicit scheme for (2.13.2) is,

$$\frac{U_{i,j}^{n+1}-U_{i,j}^{n}}{k} = \frac{1}{h^2}(\delta_x^2 U_{i,j}^{n+1} + \delta_y^2 U_{i,j}^{n+1}) \ . \qquad (2.13.8)$$

We can expand (2.13.8) by substituting the values δ_x^2 and δ_y^2 from their definition to obtain,

$$U_{i,j}^{n+1}+p(-U_{i+1,j}^{n+1}+2U_{i,j}^{n+1}-U_{i-1,j}^{n+1})+p(-U_{i,j+1}^{n+1}+2U_{i,j}^{n+1}-U_{j-1}^{n+1}) = U_{i,j}^{n} \qquad (2.13.9)$$

for i,j=1,2,...,N.

The compact form of (2.13.9) can be easily verified to be

$$(I+pA)\underline{U}^{n+1} = \underline{U}^{n} \qquad n\geq 0 \ , \qquad (2.13.10)$$

where A is a $N^2 \times N^2$ block tridiagonal matrix corresponding to the discretization which in this case becomes:

$$A = \begin{bmatrix} D & -I_N & & & \\ -I_N & D & -I_N & & 0 \\ & -I_N & \ddots & \ddots & \\ & & \ddots & \ddots & -I_N \\ & 0 & & -I_N & D \end{bmatrix}, \quad D = 4I_N - (L_N + L_N^T), \quad L_N = \begin{bmatrix} 0 & & & \\ 1 & \ddots & & 0 \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{bmatrix}$$

Hence at each time step, we are required to solve a large system of linear equations. The effort is to break-down this large system by splitting A=H+V where H and V arise from the representation of the respective bracketed terms in (2.13.9).

Defining

$$L = \begin{bmatrix} L_N & & & \\ & L_N & & 0 \\ & & \ddots & \\ 0 & & & L_N \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & & & \\ I_N & 0 & & 0 \\ & I_N & \ddots & \\ 0 & & I_N & 0 \end{bmatrix}$$

then $H=2I-(L+L^T)$, $V=2I-(B+B^T)$, and we now exploit this splitting of A to devise various iteration schemes for the solution of (2.13.2).

The novel observation of Peaceman-Rachford and Douglas was as follows, they noted that each of the finite difference equations

$$\begin{aligned} &\text{i)} \quad (I+pH)\underline{U}^{n+1} = (I-pV)\underline{U}^n \\ &\text{ii)} \quad (I+pV)\underline{U}^{n+1} = (I-pH)\underline{U}^n \ , \end{aligned} \tag{2.13.11}$$

gives rise to a computationally feasible method requiring only the solution of a set of tridiagonal matrix systems, and that each scheme, used on its own was conditionally stable. However, if they are used *alternately*, then the overall scheme is unconditionally stable (Gourlay, A.R. 1976, p.761) and is clearly an ADI method where p has the role of a constant acceleration parameter.

Douglas and Rachford (1956) proposed a new scheme introducing an intermediate value $\underline{U}^{*n+1}$ and considered the following scheme,

$$\begin{aligned}(I+pH)\underline{U}^{*n+1} &= (I-pV)\underline{U}^{n} \\ (I+pV)\underline{U}^{n+1} &= \underline{U}^{*n+1}+pV\underline{U}^{n} \quad ,\end{aligned} \tag{2.13.12}$$

which is again both unconditionally stable and computationally feasible involving the solution of a tridiagonal system along horizontal lines for $\underline{U}^{*n+1}$ and then only vertical lines for the solution of (2.13.) which is $\underline{U}^{n+1}$. It is important to note that $\underline{U}^{*n+1}$ has "no physical significance" and it is only the first estimate. Later Douglas and Rachford observed that, by the elimination of $\underline{U}^{*n+1}$ in the system (2.13.12),

$$(I+pH+pV)\underline{U}^{n+1} = \underline{U}^{n}-p^{2}HV(\underline{U}^{n+1}-\underline{U}^{n}) \quad , \qquad n>0 \tag{2.13.13}$$

thus the DPR procedure (2.13.12) is equivalent to a perturbation of the backward difference equation (2.13.10) by a term of $O(\Delta T^{2})$. (Gourlay, A.R. 1976, p.760). The importance here is that by perturbing an implicit scheme (which has the required accuracy and stability properties) in a suitable way we have obtained a scheme (2.13.13) with similar properties, but which is far simpler computationally. This idea was later generalized by Douglas and Gunn (1964) in order to realize in practice, schemes of the form

$$(I+A)\underline{U}^{n+1} = B\underline{U}^{n} \tag{2.13.14}$$

where $A = \sum_{i=1}^{q} A_i$ and each A_i, i=1,2,...,q is easily inverted.

They use,

$$\begin{aligned}(I+A_1)\underline{U}^{n+1}_{(1)} &= B\underline{U}^{n} - \sum_{j=2}^{q} A_j\underline{U}^{n} \\ (I+A_i)\underline{U}^{n+1}_{(i)} &= \underline{U}^{n+1}_{(i-1)} + A_i\underline{U}^{n} \quad i=2,3,\ldots,q \\ \underline{U}^{n+1} &= \underline{U}^{n+1}_{(q)}\end{aligned} \tag{2.13.15}$$

and show that it is equivalent to (2.13.14) with a perturbed right hand side (See Current Problem and Method in P.D.E.s, 1978, p.12-14, Edited by Wait-Gladwell).

Another important point of the DPR scheme is that, it can easily be generalized to more space dimensions.

A more general formulae of this splitting technique together with the stability analysis and convergence will be studied in the next chapter.

2.14 THE EXTRAPOLATION METHOD IN HIGHER SPACE DIMENSIONS

The extrapolation of a fully implicit scheme can also be extended to higher space dimensions. Here we study the case of two space dimensions.

Following the analysis for the one dimensional case, and considering the heat equation subject to the given initial and boundary conditions, the equation

$$\frac{d\underline{U}}{dt} = A\underline{U}(t) \quad , \tag{2.14.1}$$

arises when the spatial derivatives in equation (2.13.2) are replaced by finite differences. Here $\underline{U}(t)$ is a vector of unknowns of N^2 dimension and A is a matrix of order $N^2 \times N^2$.

The solution of (2.14.1) can be verified to be,

$$\underline{U}(t+\Delta T) = \exp(\Delta T.A)\underline{U}(t) \tag{2.14.2}$$

where $\underline{U}(0) = \underline{f}$, the vector of initial values

As was mentioned earlier the fully implicit method in two dimension results in a large system of equations whose coefficient matrix A has a bandwidth of 2N. Thus, A can be split into two components A_1 and A_2. Hence, equation (2.14.2) becomes,

$$\underline{U}(t+\Delta T) = \exp[(A_1+A_2)\Delta T]\underline{U}(t) \tag{2.14.3}$$

which may be approximated by

$$\underline{U}(t+\Delta T) = \exp(\Delta T.A_1)\exp(\Delta T.A_2)\underline{U}(t) \tag{2.14.4}$$

If now, the (1,0) Padé approximation is applied over these exponentials, then a split form of the totally implicit scheme can be obtained,

$$\begin{aligned} (I-\Delta T.A_1)\underline{U}^* &= \underline{U}(t) \\ (I-\Delta T.A_2)\underline{U}(t+\Delta T) &= \underline{U}^* \end{aligned} \tag{2.14.5}$$

where $\underline{U}^*$ is an intermediate vector. These two steps can be solved easily and only tridiagonal matrices are involved. Also, the algorithm is unconditionally stable with an accuracy of $O(\Delta T)$ with a growth factor decreasing monotonically.

Another proposed splitting to (2.14.3) is,

$$\begin{aligned}(I-\Delta T.A_1)\underline{U}^{*} &= \underline{U}(t)\\ (I-\Delta T.A_2)\underline{U}(t+\Delta T) &= \underline{U}^{*}\\ (I-\Delta T.A_2)\underline{U}^{**} &= \underline{U}(t+\Delta T)\\ (I-\Delta T.A_1)\tilde{\underline{U}}(t+2\Delta T) &= \underline{U}^{**}\end{aligned} \qquad (2.14.6)$$

and

which is a symmetrized application of (2.14.5). Eliminating the intermediate values in (2.14.6) results in the equations,

$$\tilde{\underline{U}}(t+2\Delta T) = [I+2\Delta T(A_1+A_2)+\Delta T^2(3A_1^2+3A_2^2+2A_1A_2+2A_2A_1)]\underline{U}(t)+O(\Delta T^3) \qquad (2.14.7)$$

The Maclaurin expansion of (2.14.3) over the increment $2\Delta T$ is

$$\underline{U}(t+2\Delta T) = [I+2\Delta T(A_1+A_2)+2\Delta T^2(A_1^2+A_2^2+A_1A_2+A_2A_1)]\underline{U}(t)+O(\Delta T^3)\ . \qquad (2.14.8)$$

Now, we define two new approximations called $\underline{U}^{(1)}$ and $\underline{U}^{(2)}$ which are given as follows,

$$\begin{aligned}\underline{U}^{(1)}(t+2\Delta T) &= (I-2\Delta TA_1)^{-1}(I-2\Delta TA_2)\underline{U}(t)\\ \underline{U}^{(2)}(t+2\Delta T) &= (I-2\Delta TA_2)^{-1}(I-2\Delta TA_1)\underline{U}(t)\end{aligned} \qquad (2.14.9)$$

and the expansion of these two new equations is simply verified to be,

$$\begin{aligned}\underline{U}^{(1)}(t+2\Delta T) &= [I+2\Delta T(A_1+A_2)+4\Delta T^2(A_1^2+A_1A_2+A_2^2)]\underline{U}(t)+O(\Delta T^3)\\ \text{and}\quad \underline{U}^{(2)}(t+2\Delta T) &= [I+2\Delta T(A_1+A_2)+4\Delta T^2(A_1^2+A_2A_1+A_2^2)]\underline{U}(t)+O(\Delta T^3)\end{aligned} \qquad (2.14.10)$$

The following linear combination of (2.14.7) and (2.14.10) is now proposed with accuracy of $O(\Delta T^2)$ i.e.,

$$\underline{U}(t+2\Delta T) = 2\tilde{\underline{U}} - \frac{1}{2}(\underline{U}^{(1)}+\underline{U}^{(2)})\ . \qquad (2.14.11)$$

Thus a second order method can be achieved by an extrapolation of a first order method. This novel algorithm requires four tridiagonal solutions per time step (Lawson, J.D. and Morris J.Ll., 1977), while the Peaceman-Rachford Technique, two tridiagonal solutions are required but it has a restriction on ΔT. Similar arguments are valid for the Crank-Nicolson method. However, for the extrapolated method one can apply a time-step *at least two times larger*

than that used for the Peaceman-Rachford method.

It is also worth noting that the Peaceman-Rachford method can be achieved by the same analysis as that for the extrapolated technique, i.e.,

$$\underline{U}(t+\Delta T) = \exp(\frac{\Delta T}{2} A_1)\exp(\Delta T\, A_2)\exp(\frac{\Delta T}{2} A_1)\underline{U}(t) \qquad (2.14.12)$$

Using the (1,0),(1,1) and (0,1) Padé approximations in (2.14.12) we have

$$\begin{aligned} (I - \frac{\Delta T}{2} A_2)\underline{U}^* &= (I + \frac{\Delta T}{2} A_1)\underline{U}(t) \\ (I - \frac{\Delta T}{2} A_1)\underline{U}(t+\Delta T) &= (I + \frac{\Delta T}{2} A_2)\underline{U}^* \ . \end{aligned} \qquad (2.14.13)$$

2.15 NON-UNIFORM GRID

The non-uniformity of a grid is due to two main reasons:

i) the behaviour of the solution of the differential equations,

ii) the irregularity of the boundary of the region involved.

We shall distinguish between these two cases:-

i) Consider the heat equation in one space dimensions and suppose the initial and boundary conditions are provided such that the solution has behaviour displayed in Figure 2.15.1 where the solution has an exponential decay after time t=a.

Obviously when $0<t\leq a$, one needs a very small time-step to provide the required accuracy, while after t=a, using the same time-step is not necessary.

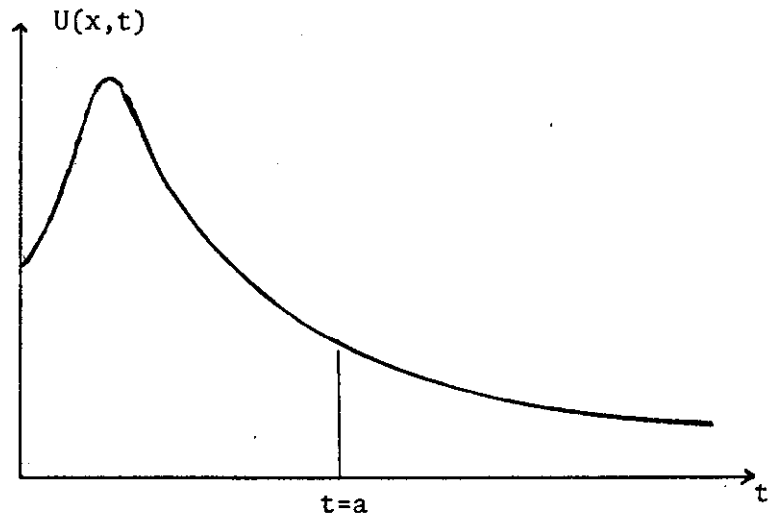


FIGURE 2.15.1

On the other hand, if one applies larger time-steps at t>a, not only is the computation speeded up, but the increased value of ΔT does not have any adverse effect on the solution (which becomes more and more smooth as the time increased). In fact, one may increase the time-step exponentially and still the decay solution compensates this growth. However, the usual

procedure is to increase the time step linearly (e.g. $\Delta T_2=2\Delta T_1$). Therefore, one needs a suitable test to change the time increment whenever necessary.

To find a suitable test, let us consider a differential equation of the form,

$$\frac{\partial u}{\partial t} = L(x,t,u) \tag{2.15.1}$$

By the application of Taylor's formula, one obtains,

$$u(t+\Delta T) = U(t) + \Delta T \frac{\partial u}{\partial t} + \frac{\Delta T^2}{2} \frac{\partial^2 U}{\partial t^2} + \ldots$$

or

$$u(t+k) = U(t) + k\ L(x,t,u) + \frac{k^2}{2} \frac{\partial^2 u}{\partial t^2} + \ldots . \tag{2.15.2}$$

If the term $\frac{k^2}{2} \frac{\partial^2 u}{\partial t^2}$ is small enough, then we can write

$$u(t+k) \simeq u(t) + k.L(x,t,u)$$

which may be used to integrate the solution for the required time-step. From (2.15.2) we can obtain a rough check on the accuracy of the approximation. We have,

$$\frac{k^2}{2} \frac{\partial^2 u}{\partial t^2} \simeq 1/2\ (U(t+k)-2U(t)+U(t-K)) . \tag{2.15.3}$$

To use this estimate, we need to store the back values u(t-h) and check the quantity (2.15.3) against some tolerance. If it is not small enough we reduce the time-step k, and if it is too small, we increase k and so on. (Danaee, A., 1978, p.45).

Although changing the step increment in the time-direction (parabolic direction) in some cases is desirable, it is undesirable in the direction of the spatial axes (elliptic direction). Since the employment of any non-uniformity of the grid results in an increase in the complexity of the net equations and also in view of assymetry, the difference equation has a larger error of approximation $O(h)$ rather than $O(h^2)$ (Saul,yev, 1964, p.148). However, in practice we may have a case when the use of a non-uniform grid is advantageous. As an example, reconsider the heat equation with the

solution illustrated in Figure 2.15.2.

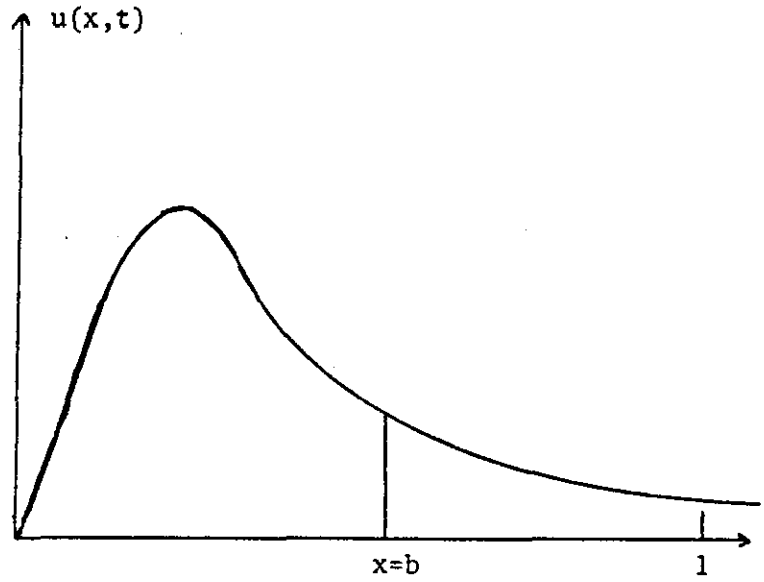


FIGURE 2.15.2

It can be seen from the behaviour of the solution that the increment Δx, should be significantly smaller in the interval $x\in[0,b]$ and larger in $x\in[b,1]$. In this case one needs to treat the finite difference approximation for two different increments near x=b.

To show this treatment, let [0,1] be divided as follows:

$$0 = x_0 < x_1 < \ldots < x_{n-1} < x_n = 1$$
$$h_i = x_i - x_{i-1} \quad 1 \leq i \leq n \tag{2.15.4}$$

and $$h = O(h_1) = \ldots = O(h_n)$$

We suppose the solution u(x,t) is sufficiently smooth and using Taylor series gives

$$u_{i+1}^j = u_i^j + h_{i+1} \frac{\partial u}{\partial x} + \frac{h_{i+1}^2}{2!} \frac{\partial^2 u}{\partial x^2} + \frac{h_{i+1}^3}{3!} \frac{\partial^3 u}{\partial x^3} + \ldots \tag{2.15.5}$$

$$u_{i-1}^j = u_i^j - h_i \frac{\partial u}{\partial x} + \frac{h_i^2}{2!} \frac{\partial^2 u}{\partial x^2} - \frac{h_i^3}{3!} \frac{\partial^3 u}{\partial x^3} + \ldots \quad . \tag{2.15.6}$$

Multiplication of equations (2.15.5) and (2.15.6) by h_i and h_{i+1} respectively results in the equation,

$$\frac{\partial^2 u_i}{\partial x^2} = \frac{2}{h_i(h_i+h_{i+1})} u^j_{i-1} + \frac{2}{h_{i+1}(h_i+h_{i+1})} u^j_{i+1} - \frac{2}{h_i h_{i+1}} u^j_i + O(h). \tag{2.15.7}$$

Accordingly, the simplest explicit finite difference scheme now becomes:

$$U^{j+1}_i = \frac{2k}{h_i+h_{i+1}} \left(\frac{U^j_{i-1}}{h_i} + \frac{U^j_{i+1}}{h_{i+1}}\right) + \left(1 - \frac{2k}{h_i h_{i+1}}\right) U^j_i + O(k+h) , \tag{2.15.8}$$

and the stability restriction

$$k \leq 1/2 \min_i h_i . h_{i+1} \tag{2.15.9}$$

has to be satisfied.

However, if for instance we choose $h_{i+1}=h_i(1\pm h_i)$ then $h_{i+1}-h_i=\pm h_i^2$ and accordingly instead of (2.15.7) we shall have,

$$\frac{\partial^2 u^j_i}{\partial x^2} = \frac{2}{h_i^2(2\pm h_i)} u^j_{i-1} + \frac{2}{h_i^2(1\pm h_i)(2\pm h_i)} u^j_{i+1} - \frac{2}{h_i^2(1\pm h_i)} u^j_i + O(h^2) \tag{2.15.10}$$

and consequently the order of accuracy of (2.15.8) increases to $O(k+h^2)$.

For two dimensional problems, the considerations expressed above, can also be applied. If for instance, the initial function f(x,y) varies particularly rapidly in some sub-domain D' of D, it is advisable and efficient to use a finer net in D'. Application of the notation illustrated in Figure (2.15.3) and using a relation of the type (2.15.5)-(2.15.7), one can easily obtain a general explicit formula for two dimensional problems, namely:

$$U^{n+1}_{ij} = \frac{2k}{h^{(1)}_i - h^{(1)}_{i+1}} \left(\frac{U^n_{i-1,j}}{h^{(1)}_i} + \frac{U^n_{i+1,j}}{h^{(1)}_{i+1}}\right) + \frac{2k}{h^{(2)}_j + h^{(2)}_{j+1}} \left(\frac{U^n_{i,j-1}}{h^{(2)}_j} + \frac{U^n_{i,j+1}}{h^{(2)}_{j+1}}\right) +$$

$$+ \left[1 - 2k\left(\frac{1}{h^{(1)}_i . h^{(1)}_{i+1}} + \frac{1}{h^{(2)}_j + h^{(2)}_{j+1}}\right)\right] U^n_{i,j} , \tag{2.15.11}$$

and the stability restriction analogous to (2.15.9) becomes:

$$k \leq 1/2 \quad \min_{i,j} \frac{h_i^{(1)}.h_{i+1}^{(1)}.h_j^{(2)}.h_{j+1}^{(2)}}{h_i^{(1)}.h_{i+1}^{(1)}+h_j^{(2)}.h_{j+1}^{(2)}} \tag{2.15.12}$$

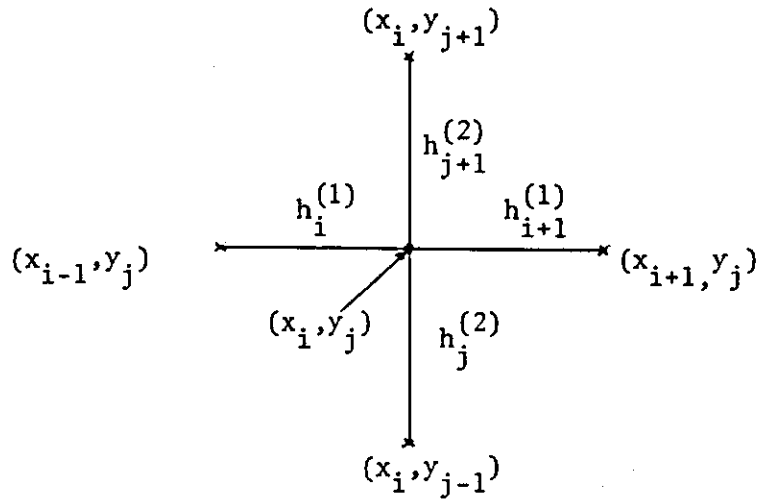


FIGURE 2.15.3

ii) The Irregularity of the Boundary

We shall now consider the non-uniformity which causes a complication not existing in one-dimensional problems.

Let the domain D of the solution be non-rectangular, as displayed in Figure (2.15.4). Internal grid points adjacent to the boundary (e.g. A,B,C... which are called quasi-boundary nodes) will normally require special treatment. Three different approaches might be indicated to construct a finite difference template at the quasi-boundary nodes. These approaches can be classified in the following manner:

1. Solution by transfer of boundary values
2. Solution by means of linear interpolation of boundary values
3. Solution without transfer of boundary values.

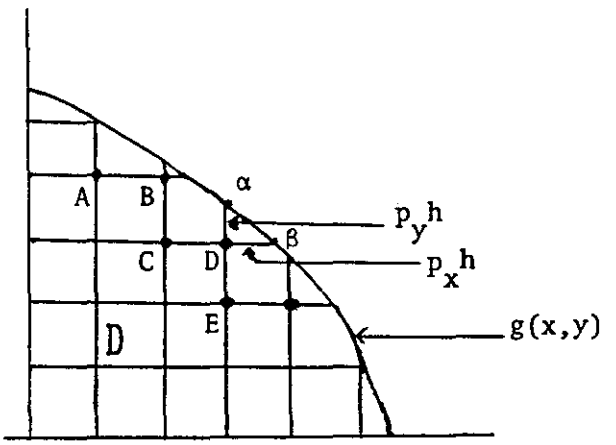


FIGURE 2.15.4

The first one which is the most simplest consists of taking the values $U_A, U_B, \ldots$ as being equal to the nearest boundary values of $g(x,y)$ to the points. Thus, if we are working with the standard explicit scheme, the stability criteria is $p \leq 1/4$ and the error of approximation at the regular points is $O(k+h^2)$ and at irregular points is $O(k+h)$.

The second method is to employ a linear interpolation for the irregular points using the interior points as well as the boundary values. For example, for the node D we may use either the following equations,

$$U_D = \frac{p_y . U_E + U_\alpha}{1+p_y} \quad , \qquad U_D = \frac{p_x . U_C + U_\beta}{1+p_x} \quad . \tag{2.15.13}$$

Finally, for the third method, the standard finite difference scheme can be applied at the interior points whilst at the quasi-boundary nodes, irregular equation of the type (2.15.11) might be used where $h_{i+1}^{(1)}$ and $h_{j+1}^{(2)}$ correspond to p_x and p_y in Figure (2.15.4) respectively.

2.16 ITERATIVE SOLUTION TECHNIQUES

With the application of the finite-difference methods to the solution of partial differential equations, one might have one of the following fundamental situations.

1. In the non-linear problems, either explicit or implicit schemes results in a non-linear or a set of non-linear equations to solve.
2. In the case of linear partial differential equations, by using an implicit finite-difference technique a set of (usually large) linear equations are obtained, which may be solved by direct or indirect methods.

We shall study both situations in this section. First we consider the iteration methods for determining the zeros of the equation

$$\underline{f}(\underline{x}) = 0 \quad , \tag{2.16.1}$$

where $\underline{f}$ and $\underline{x}$ are vectors of the same dimension N. If N=1, we have a single equation and for N>1, (2.16.1) is a system of N equations.

In general it is impossible to solve a system of equations of the form (2.16.1) directly, and some iterative methods of solution are necessary.

Factional Iteration: This method is based on the following principle.

Let us consider equation (2.16.1) where N=1. It is clear that any equation of this form can be written equivalently in the form

$$x = g(x) \quad . \tag{2.16.2}$$

If x_0 is some initial estimate of a root α of (2.16.2), a natural scheme suggested for [illegible]ining the solution of (2.16.2) is to form the iterative sequence,

$$x_{m+1} = g(x_m) \qquad m=0,1,... \tag{2.16.3}$$

An important [illegible]t concerning the convergence of this sequence and a proof of the e[illegible]nce of a unique root is contained in the following theorem.

Theorem 2.3

Let $g(x)$ satisfy the Lipschitz condition,

$$|g(x)-g(x')|\leq\lambda|x-x'| \tag{2.16.4}$$

for all values x,x' in the closed interval $I=[x_0-p,x_0+p]$ where the Lipschitz constant λ, satisfies $0\leq\lambda\leq 1$, and let the initial estimate x_0 be such that,

$$|x_0-g(x_0)|\leq(1-\lambda)f\ . \tag{2.16.5}$$

Then

i) all the iterates x_m, defined by (2.16.3), lie within the interval I, i.e.,

$$x_0-p\leq x_m\leq x_0+p\ . \tag{2.16.6}$$

ii) (Existance) the iterates converges to some point, say

$$\lim_{m\to\infty} x_m = \alpha \qquad \text{(in fact } |x_m-\alpha|\leq\lambda^m p)$$

which is a root of (2.16.2), and

iii) (Uniqueness) α is the only root in $[x_0-p,x_0+p]$.

For proof see E. Isaacson & H.B. Keller, 1966, p.86.

Collorary

If $|g'(x)|\leq\lambda<1$ for $|x-x_0|\leq p$ and (2.16.5) is satisfied, then the conclusion of theorem 2.3 is valid.

Proof

The mean value theorem implies $g(x_1)-g(x_2)=g'(\xi)(x_1-x_2)$, where λ may serve as the Lipschitz constant in (2.16.4).

Convergence Criterion

Definition 2.16.1

Let $x_0,x_1,\ldots$ be a sequence which converges to α, and $\epsilon_m=x_m-\alpha$. If there exists a number σ and a constant $c\neq 0$ such that,

$$\lim_{m\to\infty} \frac{|\varepsilon_{m+1}|}{|\varepsilon_m|^{\sigma}} = c \ , \qquad (2.16.7)$$

then σ is called the *order of convergence* of the sequence and c the *asymptotic error constant*. For σ=1,2,3 the convergence is said to be linear, quadratic and cubic respectively.

Newton-Raphson Method

Before proceeding to describe this method for a system of n equations with n unknowns, let us consider a single non-linear equation and derive the method from Taylors' formula. To solve the equation f(x)=0, we expand about x_m to obtain,

$$f(x) = f(x_m)+(x-x_m).f'(x_m)+\frac{1}{2}(x-x_m)^2.f''(\xi), \quad \xi\in(x_0,x). \qquad (2.16.8)$$

By neglecting the quadratic term, and rewriting the equation in iterative form i.e. $x\equiv x_{m+1}$, we have,

$$f'(x_m).(x_{m+1}-x_m)+f(x_m)\simeq f(x) = 0 \quad \text{for } m=0,1,... \qquad (2.16.9)$$

Thus, the Newton-Raphson method is defined by the following formulas

$$\begin{aligned} x_{m+1} &= x_m+\delta_m \qquad m=1,2,... \\ \delta_m &= -\frac{f(x_m)}{f'(x_m)} \end{aligned} \qquad (2.16.10)$$

By comparison with equation (2.16.3)

$$g(x_m) \equiv x_m+\delta_m = x_m - \frac{f(x_m)}{f'(x_m)} \ , \qquad (2.16.11)$$

and $g(x_m)$ is called the iteration function.

To study the convergence of this method, let ε_m be the error in the estimate x_m i.e.,

$$\varepsilon_m = x_m-x \ ,$$

then equation (2.16.8) can be written as

$$\frac{f(x_m)}{f'(x_m)} + x-x_m = x-x_{m+1} = -\frac{\frac{1}{2}(x-x_m)^2 f''(\xi)}{f'(x_m)}$$

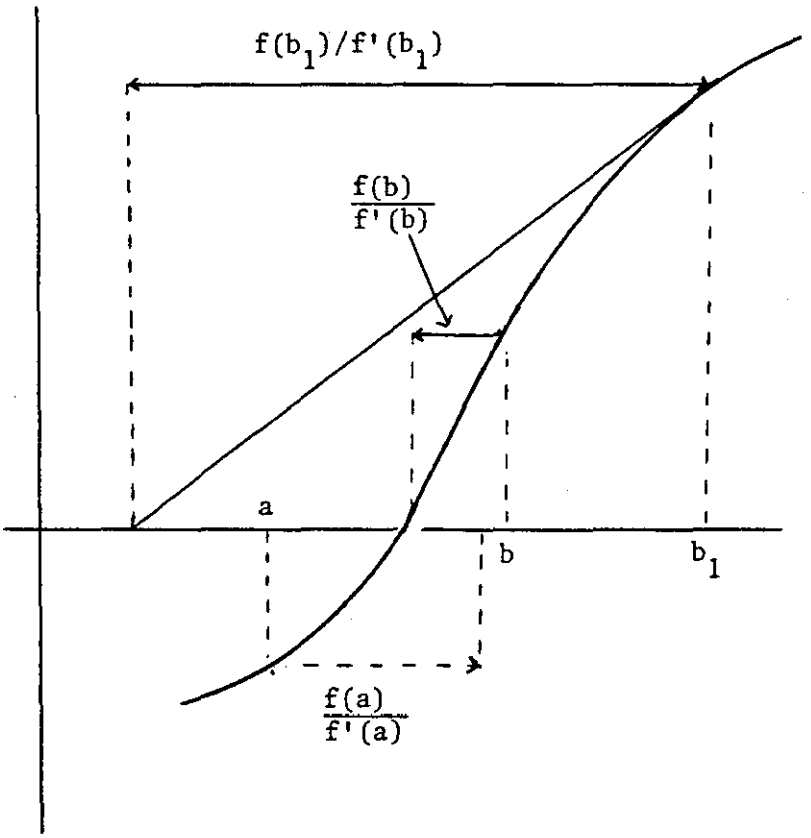


FIGURE 2.16.1

Thus we have

$$\varepsilon_{m+1} = \frac{1}{2}\varepsilon_m^2 \frac{f''(\xi)}{f'(x_m)} \qquad (2.16.12)$$

and as $x_m \to x$,

$$\frac{\varepsilon_{m+1}}{\varepsilon_m^2} \to \frac{1}{2}\frac{f''(x)}{f'(x)} \quad .$$

Since ε_{m+1} is approximately proportional to the square of ε_m provided $f(x)\neq 0$ the Newton-Raphson method is said to be quadratically convergent or to be of second order (Dalquist, G., 1976, p.223).

However, for the convergence of the above mentioned method, a good initial estimate must be provided. The following theorem indicates this fact.

Theorem 2.4

Suppose that $f'(x)\neq 0$, and $f''(x)$ does not change sign in interval $[a,b]$, and that $f(a).f(b)<0$. Then if

$$\left|\frac{f(a)}{f'(a)}\right| < b - a \quad \text{and} \quad \left|\frac{f(b)}{f'(b)}\right| < b - a,$$

the Newton-Raphson method converges from an arbitrary initial approximation $x_0 \in [a,b]$. That the theorem is true can be seen from Figure (2.16.1). While the iteration procedure converges from any point $x_0 \in [a,b]$, it may be diverged not from some points $x_0' \in [a_1,b_1]$. This is actually a drawback of the Newton-Raphson method.

Analogous to the single equation (2.16.8) from a Taylors' equation in N dimensions, we have,

$$\underline{f}(\underline{x}) = \underline{f}(\underline{x}_m) + f'(\underline{x}_m)(\underline{x}-\underline{x}_m) + O(|\underline{x}_1-\underline{x}_m|^2) \qquad (2.16.13)$$

where $\underline{x}_m$ is the iterate vector, $f'(\underline{x}_m)$ is the Jacobian matrix denoted by J, with elements

$$f'_{i,j}(\underline{x}) = \frac{\partial f_i(\underline{x})}{\partial x_j}, \quad 1\leq i,j\leq N \; . \qquad (2.16.14)$$

This leads to the Newton-Raphson method in N dimensions, i.e.,

$$J(\underline{x}_m)(\underline{x}_{m+1}-\underline{x}_m)+\underline{f}(\underline{x}_m) = \underline{0} , \qquad (2.16.15)$$

which is a *linear system* of equations for $\underline{x}_{m+1}$, and if $J(\underline{x}_m)$ is non-singular, it can be solved by a direct method (e.g. Gaussian elimination). However, if $J(\underline{x}_m)$ is a large sparse matrix an indirect or iterative method (e.g. S.O.R. which is defined later) is more favourable.

Inspite of rapid convergence each step of solution (2.16.15), requires the recalculation of N^2 entries of the (N×N) Jacobian matrix, and also the solution of a set of N linear equations. Therefore, the Newton-Raphson method is rather expensive in computational work. One may reduce the amount of computation by a modification and evaluate $J(\underline{x}_m)$ only occasionally, and not at every step (e.g. every 5^{th} step).

Practical Consideration of the Newton Method

Here we shall consider the practical application of the method, and illustrate the algorithmic procedure which can be applied to (2.16.15):

1. Calculate $\underline{f}_m=\underline{f}(\underline{x}_m)$ and $J_m=J(x_m)$,
2. Evalute $\underline{x}_{m+1}$ from the system $J(\underline{x}_m)(\underline{x}_{m+1}-x_m)+\underline{f}(\underline{x}_m)=0$ and,
3. Calculate $\underline{f}_{m+1}$.

These cases can now arise,

a) if $||\underline{f}_{m+1}||<<||\underline{f}_m||$ (e.g. $||\underline{f}_{m+1}||<||\underline{f}_m||/10$), we continue with the same Jacobian (i.e. modified Newton method).

b) if $||\underline{f}_{m+1}||<||\underline{f}_m||$ goto step 2.

c) if $||\underline{f}_{m+1}||>||\underline{f}_m||$ take $\underline{x}^*_{m+1}=\underline{x}_m-\lambda_m \underline{w}_m$ where $\underline{w}_m=\underline{x}_{m+1}-\underline{x}_m$ is found by solving the linear system (2.16.15), and $\lambda_m=\frac{1}{2^m}$ m=0,1,... until a reduction in $||\underline{f}_{m+1}||$ is obtained (J. Walsh, 1976).

This method which involves a parameter λ_m is called "damped Newton method" (where λ_m can also be chosen to be $1/10^m$). It can be noted that, only for $\lambda_m=1$, is the quadratic rate of convergence of Newton's method maintained.

It is also worthy of note that, for some nonlinear systems of equations the evaluation of the elements of the Jacobian matrix is either impossible or at least computationally expensive. In such cases we can use a functional iteration method, which does not use the Jacobian matrix at all, or use Newton's method with the Jacobian $J(\underline{x}_m)$ replaced by some approximation.

The Secant Method

The Secant method can be derived from Newton-Raphson's method by approximating the derivative $f'(x_m)$ by the difference quotient

$$f'(x_m) \approx \frac{f(x_m)-f(x_{m-1})}{x_m-x_{m-1}} \qquad (2.16.16)$$

which leads to the following method.

From given initial approximations x_0 and x_1, the sequence $x_2,x_3,\ldots$ is computed recursively, i.e.,

$$x_{m+1} = x_m + \delta_m$$

where

$$\delta_m = -f(x_m)\frac{x_m-x_{m-1}}{f(x_m)-f(x_{m-1})}, \quad f(x_m)\neq f(x_{m-1}) \qquad (2.16.17)$$

in this case the iteration function $g(x_m)\equiv\delta_m$.

The choice between the Secant method and Newton-Raphson's method depends on the amount of work required to compute $f'(x)$. Suppose the amount of work to compute $f'(x)$ is θ times the amount of work to compute a value of $f(x)$, then an asymptotic analysis can be used to motivate the rule:

If $\theta>0.44$, then use the Secant method, otherwise, use Newton-Raphson's method (Dahlquist, 1974, p.228).

In the case of a system of n-nonlinear equations, we let

$$\underline{\gamma}_m = \underline{f}(\underline{x}_{m+1}) - \underline{f}(\underline{x}_m)$$

and

$$\underline{\delta}_m = \underline{x}_{m+1} - \underline{x}_m \quad . \tag{2.16.18}$$

Then considering the truncated Taylor series (2.16.15), we obtain the system

$$\underline{J}(\underline{x}_m)\underline{\delta}_m \simeq \underline{\gamma}_m \quad .$$

When the Jacobian matrix $J(\underline{x}_m)$ is replaced by some approximation denoted by B_m, then we require

$$B_{m+1}\underline{\delta}_m = \underline{\gamma}_m \tag{2.16.19}$$

this equation is referred to as the quasi-Newton method.

where

$$B_{m+1} = B_m - \frac{(B_m \underline{\delta}_m - \underline{\gamma}_m)^T \underline{q}_m}{\underline{q}_m^T \underline{\delta}_m} \quad , \quad \underline{q}_m^T \underline{\delta}_m \neq 0$$

$\underline{q}_m$ is arbitrary with the best choice (perhaps) $\underline{q}_m = \underline{\delta}_m$.

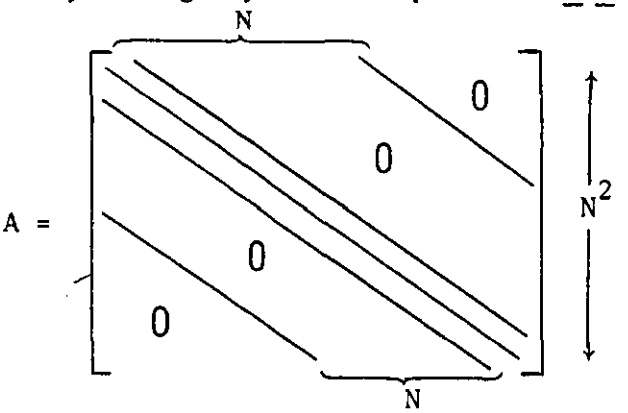
2.17 ITERATIVE METHODS FOR LINEAR EQUATIONS

Iterative methods are frequently used for large sparse systems of linear equations. Such systems appear in many engineering and physical processes resulting in a boundary-value problem involving partial differential equations. When finite-difference replacements are used to obtain an approximate solution of such problems appearance of an aforementioned system is inevitable.

As an example consider the Poisson equation

$$L(u) \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x,y) \qquad (2.17.1)$$

If the five-point operator L_h is applied to obtain an approximate solution of (2.17.1) a large system of equations $\underline{A}\underline{u}=\underline{b}$ is followed where,



and N is the number of mesh points. The non-zero elements are as indicated above.

If a direct method (e.g. Gaussian elimination) is used, then nearly all the zero elements within the band will be destroyed and we have to store nearly $2N^3$ elements instead of the $5N^2$ in the original matrix A.

Also the algorithm of a direct method is rather complicated and non-repetitive. Iteration methods on the other hand consist of repeated application of a simple algorithm, but an exact answer is given only as a limit of a sequence.

In some instances however, the right-hand side of (2.17.1) may depend not only on the mesh points but also on the solution of the equation. Hence,

in this case we are faced with a non-linear problem namely:

$$L(u) = f(x,y,u) \tag{2.17.2}$$

Thus, the algebraic problem, depending upon the complexity of f, may become complicated. In this case one may linearize equation (2.17.2) by introducing a sequence of functions $\{u_m\}$ which satisfy the boundary conditions and the linear partial differential equations:

$$L(u_{m+1}) = f(x,y,u_m) \tag{2.17.3}$$

This can be done for example, by a Newton linearization. We expand the function f about u_m:

$$f(u_{m+1}) = f(u_m)+(u_{m+1}-u_m)f'(u_m)+O[(u_{m+1}-u_m)^2] \tag{2.17.4}$$

The new linear partial differential equation now becomes,

$$L(u_{m+1})-f'(u_m)u_{m+1} = f(u_m)-u_m f'(u_m)\ , \tag{2.17.5}$$

and intuitively, we obtain a linear system $A'\underline{u}=\underline{b}'$ to solve.

For the indirect or iterative solution of a linear system, we may think of a *generalized linear method* as one which upon application to a linear system $A\underline{u}=\underline{b}$ reduces to a *feasible* iteration for the solution of the system. For instance, the Newton method does not provide a feasible iteration for the linear system $A\underline{u}=\underline{b}$, simply because we cannot obtain a convenient recurrence formula to evaluate

$$\underline{u}_{m+1} = \underline{u}_m - A^{-1}(A\underline{u}_m - \underline{b}) \quad \text{(Ortega \& Rheinboldt, 1970, p.214)}$$

In this section we shall examine some well-known methods such as the simultaneous and successive displacements methods.

Method of Simultaneous Displacements

The method of simultaneous displacements, which is also known as iteration by total steps or *Jacobi* method is the simplest of all iterative techniques.

If we assume that the diagonal elements $a_{i,i}$ of A are all non-zero, then the Jacobi method for $\underline{Au}=\underline{b}$ proceeds as follows:

We write the system of equations as

$$u_i = \frac{-\sum_{\substack{j=1 \\ j\neq i}}^{N} a_{i,j}u_j + b_i}{a_{i,i}} \quad i=1,2,\ldots,N. \qquad (2.17.6)$$

Then in the Jacobi method, we compute a sequence of approximations $\underline{u}_1,\underline{u}_2,\ldots$ by the formula

$$u_{i(m+1)} = \frac{-\sum_{\substack{j=1 \\ j\neq i}}^{N} a_{i,j}.u_{j(m)} + b_i}{a_{i,i}} \quad i=1,2,\ldots,N. \qquad (2.17.7)$$

The initial approximation is often taken to be $\underline{u}_{(0)}=0$. Then by taking the limit of (2.17.7) i.e., $\lim_{m\to\infty} \underline{u}_{(m)}=\underline{u}$, it can be shown that $\underline{u}$ is a solution of the original equation.

Inspite of its simplicity, it is seldom used because it is very slow to converge.

Method of Successive Displacements

In the Jacobi method, one does not use the new values $u_{i(m+1)}$ until every component of the vector $\underline{u}$ has been evaluated. In the method of successive displacements e.g. the Gauss-Seidel method, the new values are used in the calculation as soon as they have been computed.

Suppose that, the m^{th} iterative $\underline{u}_{(m)}=(u_{1(m)},u_{2(m)},\ldots,u_{N(m)})^T$ and the first i-1 components $u_{1(m+1)},u_{2(m+1)},u_{i-1(m+1)}$ of the (m+1)th iterate $\underline{u}_{(m+1)}$ have been determined. Then, to obtain the next component $u_{i(m+1)}$, the equation (2.17.7) becomes

$$u_{i(m+1)} = \frac{-\sum_{j=1}^{i-1} a_{i,j}u_{j(m+1)} - \sum_{j=i+1}^{N} a_{i,j}u_{j(m)} + b_i}{a_{i,i}} \quad i=1,2,\ldots,N \tag{2.17.8}$$

It is worthy to note that here only one approximation for each u_i needs to be stored at a time.

In order to write the compact form of the Gauss-Seidel iteration, we first write the matrix A in the split form

$$A = D-L-U \tag{2.17.9}$$

where D,L,U are the diagonal, strictly lower triangular and strictly upper triangular component matrices respectively. The assumption that the diagonal elements of A are non-zero then ensures that $(D-L)^{-1}$ exists and it is easy to verify that the Gauss-Seidel iteration may be written as (Ortega & Rheinboldt, 1970, p.215):

$$\underline{u}_{(m+1)} = (D-L)^{-1}[U\underline{u}_m+b] = \underline{u}_m-(D-L)^{-1}(A\underline{u}_m-b) \quad m=0,1,\ldots \tag{2.17.10}$$

By a simple modification of the Gauss-Seidel method, we can improve the rate of convergence of the method. This modification consists of introducing a parameter ω such that,

$$u_{i(m+1)} = u_{i(m)}+\omega(\tilde{u}_i-u_{i(m)}) \quad i=1,2,\ldots,N, \tag{2.17.11}$$

where $\tilde{u}_i$ is the new value from (2.17.9) before modification, and ω is called a *relaxation parameter*.

If we substitute the solution of (2.17.8) into (2.17.11), we obtain, after some rearrangement,

$$a_{i,i}u_{i(m+1)}+\omega\sum_{j=1}^{i-1} a_{i,j}u_{j(m+1)} = (1-\omega)a_{i,i}u_{i(m)}-\omega\sum_{j=i+1}^{N} a_{i,i}u_{j(m)}+ b_i \quad i=1,2,\ldots,N \tag{2.17.12}$$

which reduces to the Gauss-Seidel method for $\omega=1$. This modified method is called *The Successive Over-Relaxation* (in abbreviation SOR) method.

Here, ω should be chosen so that the rate of convergence is maximized. In its matrix form, the SOR method can be verified to be

$$\underline{u}_{m+1} = (D-\omega L)^{-1}[(1-\omega)D+\omega U]\underline{u}_m+\omega(D-\omega L)^{-1}\underline{b}$$
$$= \underline{u}_m - \omega(D-\omega L)^{-1}(A\underline{u}_m-\underline{b}) \qquad m=0,1,... \tag{2.17.13}$$

We shall now discuss the convergence of iterative methods. We can show that the Jacobi and Gauss-Seidel methods can be written in the standard form,

$$\underline{u}_{m+1} = B\underline{u}_m + C \qquad m=0,1,2,... \tag{2.17.14}$$

Such iteration methods are called *Stationary* since no variation occurs from iteration to iteration (i.e., the iteration matrix B does not change with m).

A comparison between (2.17.14) and (2.17.10) indicates that

$$B_{GS} = (D-L)^{-1}U \tag{2.17.15}$$

whilst the Jacobi iteration matrix is

$$B_J = D^{-1}(L+U) \tag{2.17.16}$$

A relation between the errors in successive approximations can be derived by subtracting from equation (2.16.2) the equation $\underline{u}=B\underline{u}+\underline{C}$, i.e.

$$\underline{u}_{m+1}-\underline{u} = B(\underline{u}_m-\underline{u}) = B[B(\underline{u}_{m-1}-\underline{u})]... = B^{m+1}(\underline{u}_0-\underline{u}) \ , \tag{2.17.17}$$

Now let $\lambda_1,\lambda_2,...,\lambda_N$ denote the eigenvalues of B with the corresponding eigenvectors, $\phi_1,\phi_2,...\phi_N$ which are linearly independent. Therefore, the components of the initial vector $\underline{e}_0$ may be represented by

$$\underline{e}_0 = \underline{u}_0-\underline{u} = \alpha_1\phi_1+\alpha_2\phi_2+ \ ...+\alpha_N\phi_N$$

and thus $$\underline{e}_m = \underline{u}_m-\underline{u} = \alpha_1\lambda_1^m\phi_1+\alpha_2\lambda_2^m\phi_2+ \ ...+\alpha_N\lambda_N^m\phi_N \ . \tag{2.17.18}$$

From equation (2.17.18), we indicate that the iteration (2.17.14) is convergent if $|\lambda_i| < 1$, i=1,2,...,N, and we can state the following theorem.

Theorem 2.5

A necessary and sufficient condition for a stationary iterative method, (2.17.14) to converge from an arbitrary initial approximation u_0 is that

$$\rho(B) = \max_{1\leq i\leq M} |\lambda_i(B)| < 1 , \qquad (2.17.19)$$

where $\rho(B)$ is the spectral radius of B.

If one wishes to reduce the amplitude of the error components $\alpha_j\phi_j$ in $\underline{e}_0$ by a factor of 10^{-1} then m iterations, where m is the smallest number such that,

$$|\lambda_j|^m \leq 10^{-k} \qquad \text{or } m \geq \frac{K}{-\log_{10}|\lambda_j|}$$

have to be performed.

Since the largest eigenvalue dominates, then *the asymptotic rate of convergence* for the iterative method (2.17.14) is denoted by

$$R = -\log_{10}(\rho(B)) \qquad (2.17.20)$$

However, it is often impractical to find the eigenvalues of B, and therefore it is difficult to apply theorem (2.5). Instead one can apply the following property of the norm-inequality on (2.17.17) and obtain

$$||\underline{u}^m-\underline{u}|| \leq ||B^m||.||\underline{u}_0'-\underline{u}|| \leq ||B||^m||\underline{u}_0-\underline{u}|| .$$

The sufficient condition for convergence now becomes $||B||<1$, for some consistent matrix norm. An estimate for the error in $\underline{u}_m$ can be found by using the relation

$$\underline{u}_m-\underline{u} = -B(\underline{u}_m-\underline{u}_{m-1})+B(\underline{u}_m-\underline{u}) . \qquad (2.17.21)$$

If the norm of B is denoted by $\beta<1$, then (2.17.21) becomes

$$||\underline{u}_m-\underline{u}|| \leq \frac{\beta}{1-\beta} ||\underline{u}_m-\underline{u}_{m-1}|| . \qquad (2.17.22)$$

We note that (2.17.22) is a rough estimate for checking the error but one should also consider the effect of round-off errors in the iterative process.

Now, we study the norm of B in iterative procedures to find out the condition for convergence of the various methods.

In Jacobi's method, the iteration matrix B_J is defined to have the elements,

$$b_{i,j} = \frac{a_{i,j}}{a_{i,i}} \quad i \neq j \quad \text{and} \quad b_{i,i} \neq 0$$

thus

$$||B_J||_\infty = \max_{1 \leq i \leq N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{|a_{i,j}|}{|a_{i,i}|} \quad , \tag{2.17.23}$$

Consequently, if A is strictly diagonally dominant, then $||B_J||_\infty < 1$ and the Jacobi method is convergent.

For the Gauss-Seidel method, we apply the subordinate matrix norm namely,

$$||B_{GS}||_\infty = \max_{\underline{x} \neq 0} \frac{||B_{GS}\underline{x}||_\infty}{||\underline{x}||_\infty} \tag{2.17.24}$$

Let $\underline{y}$ denote $B_{Gx}\underline{x}$, since without loss of generality (2.17.9) can be written as A=D-L-U, using $||\underline{y}||_\infty$ we obtain,

$$\underline{y} = B_{GS}\underline{x} = (D+L)^{-1}U\underline{x} \rightarrow \underline{y} = -\tilde{L}\underline{y} - \tilde{U}\underline{x}$$

$$\text{or} \quad ||y||_\infty \leq ||\tilde{L}||_\infty ||\underline{y}||_\infty + ||\tilde{U}||_\infty . ||\underline{x}||_\infty = s_i ||\underline{y}||_\infty + r_i ||\underline{x}||_\infty \tag{2.17.25}$$

where

$$s_i = \sum_{j=1}^{i-1} \frac{|a_{i,j}|}{|a_{i,i}|} \quad \text{and} \quad r_i = \sum_{j=i+1}^{N} \frac{|a_{i,j}|}{|a_{i,i}|} .$$

Therefore we have

$$(1-s_i)||\underline{y}||_\infty \leq r_i ||\underline{x}||_\infty \rightarrow ||\underline{y}||_\infty \leq \frac{r_i}{1-s_i} ||\underline{x}||_\infty$$

which results in

$$||B_{GS}||_\infty = \max_{1 \leq i \leq N} \frac{||\underline{y}||_\infty}{||\underline{x}||_\infty} = \max_{1 \leq i \leq N} \frac{r_i}{1-s_i} . \tag{2.17.26}$$

From equation (2.17.26) it follows that, the Gauss-Seidel iteration is convergent when A is strictly diagonally dominant.

In the case of S.O.R. the iteration matrix B_ω is given as follows,

$$B_\omega = (I-\omega\tilde{L})^{-1}[(1-\omega)I+\omega\tilde{U}]$$

which is the accelerated form of the Gauss-Seidel method. From the following theorem we can show that for convergence, the parameter ω must be in the range $0<\omega<2$.

Theorem 2.6

For the S.O.R. iteration matrix we have

$$\rho(B_\omega) \geq |\omega-1| \quad . \tag{2.17.27}$$

So the method can only converge for $0<\omega<2$.

Proof:

Since the determinant of a triangular matrix is the product of its diagonal elements and $(I-\omega L)^{-1}$ and $[(1-\omega)I+\omega U]$ are both triangular matrices, hence we obtain,

$$\det(B_\omega) = \det(I-\omega\tilde{L})^{-1}\det[(1-\omega)I+\omega\tilde{U}] = (1-\omega)^N.$$

On the other hand, if the eigenvalues of B_ω are denoted by $\lambda_1,\lambda_2,\ldots\lambda_N$, then from $\det(B_\omega)=\lambda_1\lambda_2\ldots\lambda_N$ it follows that

$$\max_i |\lambda_i| \geq |1-\omega|, \qquad 0<\omega<2.$$

Theorem 2.7

Let A be a symmetric matrix with positive diagonal elements. Then, the S.O.R. method converges if and only if A is positive definite and $0<\omega<2$.

Proof:

See Young, D.M., 1971, p.113.

Some Conclusive Remarks

1. In the finite-difference approximation of partial differential equations, we frequently have positive matrices.
2. In practice ω usually lies between 1 and 2. (The method is called successive under relaxation if $0<\omega<1$).
3. The optimum or best value for ω denoted by ω_{opt} for the maximum rate of convergence is given by (Young, D.M., 1971, p.169)

$$\omega_{opt} = \frac{2}{1+\sqrt{1-\mu^2}} \qquad (2.17.28)$$

where μ is the spectral radius of the Jacobi iteration matrix $D^{-1}(L+U)$ associated with the matrix A.

4. The methods which we have studied so far are *point* iterative methods. There are also some iteration techniques which correspond to the evaluation of a *group* of points simultaneously. These methods are called *'block iterative methods'*. We shall consider certain methods of block iteration as applied to the solution of partial differential equations in more detail later.

CHAPTER III

THE HOPSCOTCH TECHNIQUE FOR THE SOLUTION OF PARABOLIC PARTIAL DIFFERENTIAL EQUATIONS

3.1 INTRODUCTION

One of the most important methods for the determination of a more accurate numerical solution to the exact solution of a problem is the use of different numerical algorithms, which give errors of different signs. For instance, in the case of Laplace's equation in a convex-region, the finite difference method gives an approximation from below for the first eigenvalue whereas the variational Ritz method gives an approximation from above (Saul,yev, 1964, p.108). Therefore, a combination of these two solutions gives a more accurate approximation for the exact values.

In the application of finite-difference replacement for the approximate solution of partial differential equations, it would be desirable, if we can obtain the error of the algorithm being used, not only by the mean of the modulus of some estimate, but the precise values of the errors. This is almost impossible and there is no guarantee that, the finite-difference method will provide a solution with errors having the same sign.

However, in the case of parabolic equations we can deduce some interesting information.

In this chapter, we study some general principles for constructing an algorithm to provide a *bilateral approximation* to the solution.

The motivation of this study is based on two theorems which are given by Saul,yev (1964) and Gourlay (1970) where the *Hopscotch Technique* is introduced. We shall give full discussion of this technique and provide the background for the new development in Chapter VI. We also study the work of McGuire (1970), Gane (1974) and the recent development on the technique by Gourlay and McKee (1978).

3.2 THE HOPSCOTCH TECHNIQUE

To illustrate the essence of the hopscotch technique, we consider the following one dimensional model equation, namely,

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \tag{3.2.1}$$

where some boundary values together with the initial values $u(x,0)=f(x)$ are given

It can be shown (Weinberger, H.F., 1965, p.126) that by the application of Fourier transforms, the solution of (3.1) has the form,

$$u(x,t) = \sum_{k=1}^{\infty} c_k \, e^{-k^2\pi^2 t} \sin k\pi x \ , \tag{3.2.2}$$

with the Fourier coefficients c_k given by:

$$c_k = 2 \int_0^1 f(\xi) \sin k\pi\xi \, d\xi \ .$$

We assume that the series (32.2) together with its derivatives with respect to x and t are convergent in the region under consideration.

If the finite-difference explicit form is applied i.e.

$$U_i^{j+1} = (1-2p)U_i^j + p(U_{i+1}^j + U_{i-1}^j) \tag{3.2.3}$$

then for j=0 (i.e. in the first layer), we have,

$$U_i^{j+1} = (1-2p) \sum_{k=1}^{\infty} c_k \, e^{-k^2\pi^2 j\Delta T} \sin k\pi ih + p \left[\sum_{k=1}^{\infty} c_k \, e^{-k^2\pi^2 j\Delta T} \right.$$

$$\left. (\sin k\pi(i+1)h + \sin k\pi(i-1)h) \right] \ ;$$

$$= \sum_{k=1}^{\infty} c_k \, e^{-k^2\pi^2 j\Delta T} [(1-2p) + 2p\cos k\pi h] \sin k\pi ih \ , \tag{3.2.4}$$

and hence the error at the point $(x, t+\Delta T)$ is,

$$u(x,t+\Delta T) - U_i^{j+1} = \sum_{k=1}^{\infty} c_k \, e^{-k^2\pi^2 (j+1)\Delta T} \sin k\pi i \, h - \sum_{k=1}^{\infty} c_k \, e^{-k^2\pi^2 j\Delta T}$$

$$[(1-2p) + 2p\cos k\pi h] \sin k\pi ih$$

$$= \sum_{k=1}^{\infty} c_k \, e^{-k^2\pi^2 j\Delta T} \left[e^{-k^2\pi^2 \Delta T} - (1-2p) - 2p\cos k\pi h \right] \sin k\pi ih$$

$$= \sum_{k=1}^{\infty} c_k \, e^{-k^2\pi^2 j\Delta T} \left[1 - k^2\pi^2\Delta T + \frac{k^4\pi^4\Delta T^2}{2!} - \ldots - 1 + 2p - 2p \right.$$

$$\left. \left(1 - \frac{k^2\pi^2 h^2}{2!} + \frac{k^4\pi^4 h^4}{4!} + \ldots\right) \right] \sin k\pi ih$$

If we ignore the quantities of higher order of h, we obtain the error in the form,

$$u(i,(j+1)\Delta T)-U_i^{j+1} = \frac{\pi^4\Delta T(6\Delta T-h^2)}{12}\sum_{k=1}^{\infty} c_k\, k^4\, e^{-\pi^2k^2j\Delta T}\, \sin k\pi ih. \qquad (3.2.5)$$

Now let the following inequality hold everywhere in the region $D=[0<x<1]\times[0<t\leq T]$

$$\sum_{k=1}^{\infty} c_k\, k^4\, e^{-\pi^2k^2j\Delta T}\, \sin k\pi ih = \frac{\partial^2 u}{\partial t^2} \geq 0\ , \qquad (3.2.6)$$

then provided $p=\Delta T/\Delta x^2=k/h^2\leq 1/6$ we may write

$$U_i^j \leq u_i^j \quad \text{for } k=1\ . \qquad (3.2.7)$$

Since the coefficients on the right hand side of the formula (3.2.3) are positive, then provided the values of the k^{th} layer are positive, therefore the values on the $(k+1)^{th}$ layer also become positive. Hence, the inequality (3.2.7) is true for $k>1$.

If the inequality (3.2.6) is changed to be negative, then the approximate values obtained are over-estimated. But, since we know nothing about the size of $\frac{\partial^2 u}{\partial t^2}$ (as the exact solution is not known) the examination of (3.2.6) is not possible. However, for sufficiently large values of t (e.g. $t\geq t_1$) i.e. for such t that the principal contribution to the solution comes from the first harmonic, we may write (Saul,yev, 1964, p.112),

$$\text{sign}\ \frac{\partial^2 u}{\partial t^2} = \text{sign}\ c_1\, e^{-\pi^2 t}\, \sin\pi x,\ t\geq t_1\ , \qquad (3.2.8)$$

since the following inequality holds for the values of t:

$$|c_1 e^{-\pi^2}\, \sin\pi x| \geq \left|\sum_{k=2}^{\infty} c_k\, k^4\, e^{-\pi^2k^2t}\, \sin k\pi x\right|.$$

And since sin x is positive for $0<x<1$, the sign of $\frac{\partial^2 u}{\partial t^2}$ can be determined from the sign of c_1. Hence the following theorem can be established.

Theorem 3.1

If the following inequality holds:

$$\int_0^1 f(\xi)\sin\pi\xi d\xi > 0 \ , \tag{3.2.9}$$

then for sufficiently large values of t ($t>t_1$) and for sufficiently small h, the explicit method (3.2.3) for $p\leq 1/6$ gives an approximation to the solution of the problem (3.1), from below.

In the case of the implicit finite-difference formula i.e.,

$$(1+2p)U_i^{j+1}-p(U_{i-1}^{j+1}+U_{i+1}^{j+1}) = U_i^j \tag{3.2.10}$$

an analysis similar to the explicit model can not easily be applied (since the explicity plays an essential role). However, we can overcome this difficulty by the following consideration. Although using (3.2.10) we proceed to the $(k+1)^{th}$ layer from k^{th} layer, but for the analysis of the error we may suppose that the exact values at the $(k+1)^{th}$ layer are given and the error on the k^{th} layer is required. Therefore an explicit type analysis can be carried out and hence we have for the error

$$u(i\Delta x,j\Delta T)-U_i^j = \sum_{k=1}^{\infty} c_k\, e^{-\pi^2k^2j\Delta T}\sin k\pi ih[1+2p\,\cos k\pi h.e^{-k^2\pi^2\Delta T}-(1+2p)e^{-\pi^2k^2\Delta T}]$$

$$= \frac{\pi^2\Delta T(6\Delta T+h^2)}{12}\sum_{k=1}^{\infty} c_k k^4\, e^{-\pi^2k^2j\Delta T}\sin k\pi ih+O(h^6). \tag{3.2.11}$$

The equation (3.2.11) together with the same analysis as before results in the following theorem.

Theorem 3.2

If the condition of Theorem 3.1 holds, then the implicit equation (3.2.10) gives an approximation to the desired solution from above.

The conclusive remarks on Theorems 3.1 and 3.2 are that: if the initial function is positive (as is often the case), then the explicit finite difference scheme has a tendency to give an underestimated result

while the implicit scheme gives an overestimated result to the solution. Thus the implication is that, one can achieve a better accuracy from a combination of these schemes.

As a numerical experiment, we have solved the equation (3.2.1) where $f(x)=\sin x$ and the behaviour of the explicit and implicit result together with the exact solution is illustrated in Figure 3.2.1.

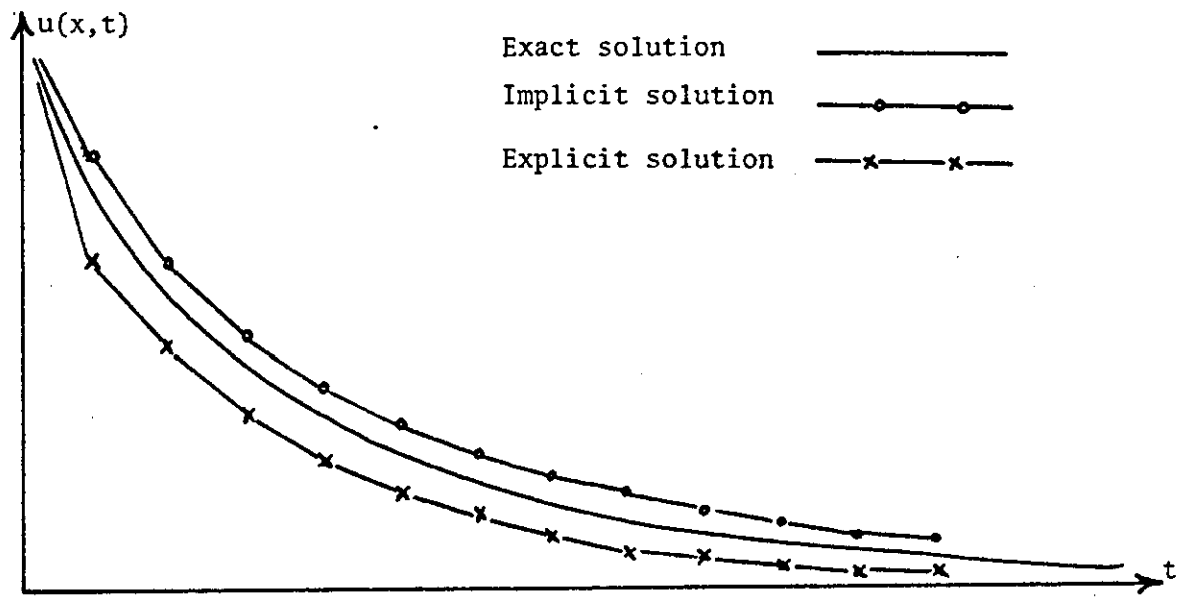


FIGURE 3.2.1

3.3 IMPLICIT-EXPLICIT SCHEMES

In the previous chapter, we briefly referred to a combination of the standard explicit and implicit schemes at alternate levels or alternate nodes. Here we shall consider the latter combination and investigate the accuracy, stability and convergence, in one and two dimensional problems. We also show some advantages of this combination compared with the explicit and implicit schemes in linear and non-linear problems.

Consider the problem (3.2.1) in the given region which is covered by a grid in the usual way. The algorithm consists of two steps:

At the first time step, every other point is evaluated by the explicit scheme (3.2.3), while the remaining points are obtained by the application of the implicit formula.

At the next time step the procedure is reversed, and this cycle is repeated.

This algorithm was first presented by Saul,yev (1964, p.68) where he reports according to the Theorems 3.1 and 3.2 that this scheme is more accurate than the standard explicit and implicit formulae. He also indicates that this algorithm is stable iff $p \leq 1$.

In a preliminary paper, P. Gordon (1965) reconsidered the scheme and claimed that the major difficulty was in fact the poor truncation error.

A full discussion of the scheme is given by Gourlay (1970) where the new name *Hopscotch* is given. Gourlay showed that the scheme may be regarded as an ADI scheme with a rather novel method of decomposing the problem into two simpler parts.

In this chapter we are primarily concerned with the Hopscotch method and its many variants.

3.4 THE HOPSCOTCH ALGORITHM

In order to develop the algorithm we restrict ourselves for the moment to the linear parabolic equation

$$\frac{\partial u}{\partial t} = L(u) + g(x,y,t) \ , \tag{3.4.1}$$

where L is a second-order linear, elliptic differential operator in the space variables x and y. Suppose the solution of (3.4.1) is required in the cylinder $R\times[0\leq t\leq T]$ where R is a closed region in x-y plane, with continuous boundary ∂R. Approximate initial and boundary values are given on t=0 and $\partial R\times[0\leq t\leq T]$ respectively.

To apply the finite-difference approximation we discretize the region $R\times[0\leq t\leq T]$ in the usual way by superimposing a rectilinear grid on the region where the mesh spacings in the space variables can be taken equal, namely, $\Delta x=\Delta y=h$, and the mesh spacing in the time direction ΔT denoted by k. We denote by $u^m_{i,j}$ and $U^m_{i,j}$ the exact and approximate values of the solution of (3.4.1) at the mesh point $(ih,jh,mk)=(x,y,t)$, $i,j=0,1,\ldots,N$ and $m=0,1,\ldots,$ respectively.

The algorithm of Saʋl,yev (1969) and Gordon (1965) consists of using alternately the simple explicit and implicit replacement of equation (3.4.1) namely,

$$U^{m+1}_{i,j} = U^m_{i,j} + k(L_h U^m_{i,j} + g^m_{i,j}) \tag{3.4.2}$$

$$U^{m+1}_{i,j} = U^m_{i,j} + h(L_h U^{m+1}_{i,j} + g^{m+1}_{i,j}) \tag{3.4.3}$$

where L_h is a finite difference approximation of the linear operator L and $g^m_{i,j}$ denotes the value of g(x,y,t) at the mesh points.

The Hopscotch formulation consists of replacing the two equations (3.4.2) and (3.4.3) by a single equation which defines the algorithm locally at all points. The following (odd-even) function is introduced by Gourlay (1970),

$$\theta_{i,j}^{m} = \begin{cases} 1 & \text{if } m+i+j \text{ is odd} \\ 0 & \text{if } m+i+j \text{ is even} \end{cases} \qquad (3.4.4)$$

Thus, the single Hopscotch formula can be obtained by the application of (3.4.4) as follows:-

$$U_{i,j}^{m+1} - k\theta_{i,j}^{m+1}[L_h U_{i,j}^{m+1} + g_{i,j}^{m+1}] = U_{i,j}^{m} + k\theta_{i,j}^{m}[L_h U_{i,j}^{m} + g_{i,j}^{m}] \ . \qquad (3.4.5)$$

This algorithm changes over at succeeding steps, hence we really only obtain answers at $m+1=2n$ where $n=0,1,\ldots$.

However, in contrast with the ADI (Peaceman-Rachford) method, the intermediate values obtained from the odd steps are meaningful approximations to the solution. This is why we regard (3.4.5) as advancing the calculation from $t=mk$ to $t=(m+1)k$. (Gourlay, A.R., 1970, p.377).

The novelty of the formula (3.4.5) appears if we write down two successive equations of the type (3.4.5) as follows:

$$\begin{aligned} U_{i,j}^{m+1} - k\theta_{i,j}^{m+1}[L_h U_{i,j}^{m+1} + g_{i,j}^{m+1}] &= U_{i,j}^{m} + k\theta_{i,j}^{m}[L_h U_{i,j}^{m} + g_{i,j}^{m}] \ , \\ U_{i,j}^{m+2} - k\theta_{i,j}^{m+2}[L_h U_{i,j}^{m+2} + g_{i,j}^{m+2}] &= U_{i,j}^{m+1} + h\theta_{i,j}^{m+1}[L_h U_{i,j}^{m+1} + g_{i,j}^{m+1}] \ , \end{aligned} \qquad (3.4.6)$$

which after some algebraic elimination reduces to the following equation,

$$U_{i,j}^{m+2} - k\theta_{i,j}^{m+2}[L_h U_{i,j}^{m+2} + g_{i,j}^{m+2}] = 2U_{i,j}^{m+1} - (U_{i,j}^{m} + k\theta_{i,j}^{m}[L_h U_{i,j}^{m} + g_{i,j}^{m}]) \qquad (3.4.7)$$

When $\theta_{i,j}^{m}$ is zero, equation (3.4.7) reduces to the explicit scheme,

$$U_{i,j}^{m+2} = 2U_{i,j}^{m+1} - U_{i,j}^{m} \ , \qquad (3.4.8)$$

and consequently, for half the points an extremely simple substitution attains the approximation required at the next time level.

Before giving the computational algorithm, let us describe the E-operator.

If $L_h U_{i,j}^{m+1}$ is a replacement which involves $U_{i,j}^{m+1}$ and its nearest neighbouring points along grid lines i.e. $U_{i\pm1,j}^{m+1}, U_{i,j\pm1}^{m+1}$, L_h is called an

E-operator. Thus, an E-operator can only be achieved in the replacement of first and second order differential terms. Such operators often occur in the solution of partial differential equations by the application of finite-difference approximations.

Thus, we conclude that, the Hopscotch process is explicit if L_h is an E-operator, otherwise the algorithm is said to be implicit. We shall consider implicit algorithms in forthcoming sections.

3.5 GENERAL SPLITTING FORMULA AND BLOCK HOPSCOTCH SCHEMES

A general splitting formula for equation (3.4.1) is given by,

$$U^{m+1}_{i,j}-k(\theta^{m+1}_{i,j}L^{(1)}_h+\eta^{m+1}_{i,j}L^{(2)}_h)U^{m+1}_{i,j} = U^{m}_{i,j}+k(\theta^{m}_{i,j}L^{(1)}_h+\eta^{m}_{i,j}L^{(2)}_h)U^{m}_{i,j}$$

$$+k(\theta^{m+1}_{i,j}g^{(1)m+1}_{i,j}+\eta^{m+1}_{i,j}g^{(2)m+1}_{i,j})+k(\theta^{m}_{i,j}g^{(1)m}_{i,j}+\eta^{m}_{i,j}g^{(2)m}_{i,j}) \quad , \qquad (3.5.1)$$

with the restrictions,

$$\theta^{m+1}_{i,j} + \theta^{m}_{i,j} = 1$$

$$\eta^{m+1}_{i,j} + \eta^{m}_{i,j} = 1$$

where $g^{(1)m}_{i,j} + g^{(2)m}_{i,j} = g^{m}_{i,j}$, $L_h \equiv L^{(1)}_h + L^{(2)}_h$ is the finite difference replacement for the differential operator $L \equiv L^{(1)} + L^{(2)}$, and $L^{(1)}$ and $L^{(2)}$ are one-dimensional operators, namely $L^{(1)} = \frac{\partial^2}{\partial x^2}$ and $L^{(2)} = \frac{\partial^2}{\partial y^2}$. $L^{(1)}_h$ and $L^{(2)}_h$ can be any constant E-operator.

The odd-even Hopscotch which was described earlier can be obtained by defining

$$\theta^{m}_{i,j} = \eta^{m}_{i,j} = \begin{cases} 1 \text{ if } i+j+m & \text{even} \\ 0 \text{ if } i+j+m & \text{odd} \end{cases} \qquad (3.5.2)$$

Other definitions can be given for $\theta^{m}_{i,j}$ and $\eta^{m}_{i,j}$ which lead to different types of splitting.

For instance, the values $\theta^{m}_{i,j}=\eta^{m}_{i,j}=\theta^{m+1}_{i,j}=1/2 \ \forall \ i,j,m$ gives the Crank-Nicolson scheme and

$$\theta^{m}_{i,j} = 1/2(1+(-1)^m) = \begin{cases} 1, & \text{if } m \text{ is even} \\ 0, & \text{if } m \text{ is odd} \end{cases}$$

$$\eta^{m}_{i,j} = 1-\theta^{m}_{i,j} \qquad \forall \ \ i,j,m$$

gives the Peaceman-Rachford method with a time step of 2k. (McGuire, R.G., 1970, p.7).

The general formulation (3.5.1) allows us to obtain a whole class of Hopsoctch algorithms. Thus, we consider other definitions of $\theta^{m}_{i,j}$ and $\eta^{m}_{i,j}$ to derive some new schemes.

Take

$$\theta^m_{i,j} = \eta^m_{i,j} = \begin{cases} 1, & \text{if } i+m \text{ odd} \\ 0, & \text{if } i+m \text{ even} \end{cases} \tag{3.5.3}$$

Hence, (3.5.1) becomes,

$$U^{m+1}_{i,j} - k\theta^{m+1}_{i,j}(L_h U^{m+1}_{i,j} + g^{m+1}_{i,j}) = U^m_{i,j} + k\theta^m_{i,j}(L_h U^m_{i,j} + g^m_{i,j}) \tag{3.5.4}$$

$j=1,2,\ldots,N-1.$

from which it is easily seen that this algorithm requires the solution of a tridiagonal system of linear finite difference equations to obtain the approximate values at points along alternating i-grid lines i.e. in the y direction. The above method is called the *Line Hopscotch* scheme. One can also define

$$\theta^m_i = \eta^m_i = \begin{cases} 1 \text{ if } j+m \text{ even} \\ 0 \text{ if } j+m \text{ odd} \end{cases}$$

to obtain in the same way a line hopscotch scheme in the x-direction.

A fast line hopsoctch algorithm can also be obtained by writing (3.5.4) with m replaced by m+1 and eliminating $k\theta^{m+1}_i[L_h U^{m+1}_i + g^{m+1}_{i,j}]$. From the resulting equation we then have

$$U^{m+2}_{i,j} - k\theta^{m+2}_i[L_h U^{m+2}_{i,j} + g^{m+2}_{i,j}] = 2U^{m+1}_{i,j} - U^m_{i,j} - k\theta^m_i[L_h U^m_{i,j} + g^m_{i,j}]$$

which reduces to

$$U^{m+2}_{i,j} = 2U^{m+1}_{i,j} - U^m_{i,j} \qquad \text{for } i+m = \text{odd}$$

When

$$\theta^m_{i,j} = \eta^{m+1}_{i,j} = \begin{cases} 1, & i+m \text{ even} \\ 0, & i+m \text{ odd} \end{cases}$$

we have the *A.D.I. Hopscotch* scheme,

$$U^{m+1}_{i,j} - k(\theta^{m+1}_{i,j}L^{(1)}_h + \theta^m_{i,j}L^{(2)}_h)U^{m+1}_{i,j} = U^m_{i,j} + k(\theta^m_{i,j}L^{(1)}_h + \theta^{m+1}_{i,j}L^{(2)}_h)U^m_{i,j}$$
$$+ k(\theta^{m+1}_{i,j}g^{(1)m+1}_{i,j} + \theta^m_{i,j}g^{(2)m+1}_{i,j}) + k(\theta^m_{i,j}g^{(1)m}_{i,j} + \theta^{m+1}_{i,j}g^{(2)m}_{i,j}) \tag{3.5.5}$$

or

$$(1-kL^{(2)}_h)U^{m+1}_{i,j} = (1+kL^{(1)}_h)U^m_{i,j} + k(g^{(2)m+1}_{i,j} + g^{(1)m}_{i,j}) \quad i+m \text{ even} \tag{3.5.6}$$

$$(1-kL^{(1)}_h)U^{m+1}_{i,j} = (1+kL^{(2)}_h)U^m_{i,j} + k(g^{(1)m+1}_{i,j} + g^{(2)m}_{i,j}) \quad i+m \text{ odd} \tag{3.5.7}$$

If we now substitute the standard finite difference replacements for $L_h^{(1)}$ and $L_h^{(2)}$ in (3.5.6) and (3.5.7) i.e.

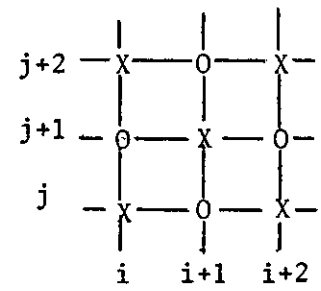
$$L_h^{(1)} \equiv \frac{\delta_x^2}{h^2} , \qquad L_h^{(2)} = \frac{\delta_y^2}{h^2}$$

then (3.5.6) requires the solution of a tridiagonal system of equations along alternate i-grid lines and therefore the equation (3.5.7) now becomes an explicit process (Morris, J.Ll. & Nicoll, I.E., p.324).

We note there that, we can not simplify equations (3.5.6) and (3.5.7) to obtain a fast formula as (3.4.8) for the A.D.I. hopsoctch method.

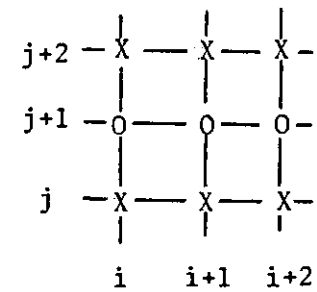
In the Figure 35.1 we demonstrate different hopscotch schemes where the symbols X and O correspond to explicit and implicit points respectively.

a) X Evaluated first with explicit formula (3.4.2).
O Evaluated later with implicit formula (3.4.4).
The points X and O are interchanged at the next time step. The process is explicit



(a) Odd-Even Hopscotch

b) X Evaluated first by explicit formula (3.4.2).
O Evaluated later by implicit equivalent to (3.5.4).
This requires the solution of a tridiagonal system of equations

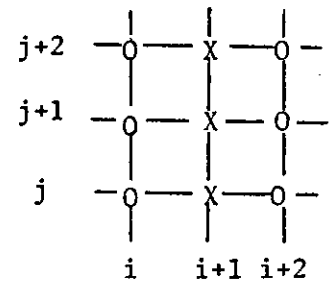


(b) Line-Hopscotch

c) O Evaluated first by implicit formula (3.5.6).

X Evaluated later by formula (3.5.7).

This scheme also requires the solution of a tridiagonal system of equations.



(c) ADI-Hopscotch

FIGURE 3.5.1

The processes (b) and (c) in Figure 3.5.1 are called *Block-Hopscotch* procedures. In the case of block procedures, the way in which we choose our block is critical for the efficiency of the method changes from one to another. Also one procedure might be more accurate due to the implicitness of the block as well as the cancellation of the L.T.E. terms. We shall describe these properties later in this chapter.

Here we describe another block called "Peripheral Hopscotch" which was first presented by Gane, C.R. (1974, p.73).

The peripherals consist of the X points, evaluated first using the explicit formula (3.4.2) and the peripherals consisting of the O points are evaluated later which require the solution of a cyclic tridiagonal system (Figure 3.5.2).

One can also derive the fast Hopscotch formula for the peripheral pattern.

As a comparison of the computational work involved in the aforesaid schemes, we notice that in the Peaceman-Rachaford method, the solution of a tridiagonal system is required for each line parallel to the x (or y) axes

in the first sweep and the same work is involved for each line parallel to the y (or x) axes in the second sweep of every cycle.

The Line and ADI Hopscotch methods, on the other hand, depending on whether m+i is odd or even, (where m is the level number and i is the line number), requires the solution of a tridiagonal system of equations for alternative lines in the first and second sweep in every two step process (where the neighbouring lines are evaluated explicitly).

The odd-even (point) Hopscotch is completely explicit and gives the faster method whereas the Crank-Nicolson scheme which is a fully implicit method is much slower.

Finally, as far as programming effort is concerned, the odd-even scheme is very easy to program while the Line and ADI Hopscotch methods are more complex and require more programming time. Another advantage of the hopscotch processes is the efficient use of storage and its economy in operation.

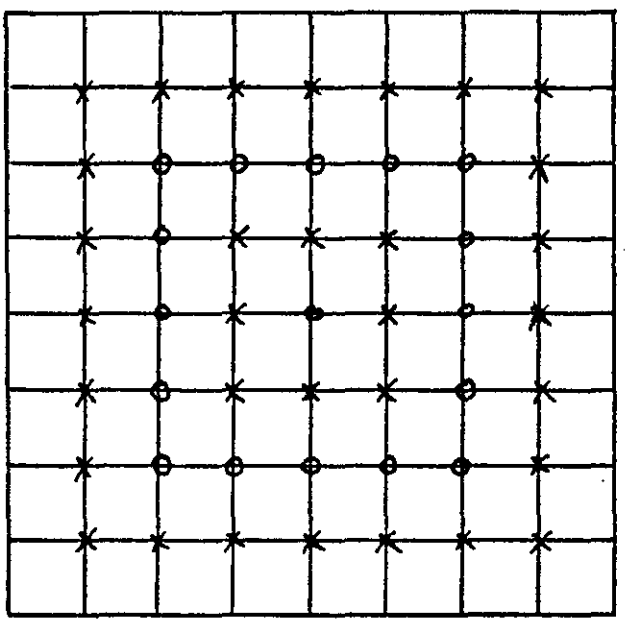


FIGURE 3.5.2

3.6 STABILITY OF HOPSCOTCH METHODS

For stability of the method (McGuire, 1970, pp.13), we consider the general formula (3.5.1) globally (over the entire mesh). First we introduce some notation.

Let,

$$\underline{U} = \{(u_{i,j})_j\}_i \tag{3.6.1}$$

denote the solution vector for the mesh points ordered along lines parallel to the y axes and the vector $\underline{g}$ has the same definition. Also let the diagonal matrices $I_1^{\theta}, I_1^{\eta}, I_2^{\theta}$ and I_2^{η} be defined by,

$$\begin{aligned} (I_1^{\theta}\underline{U}^{2m})_{i,j} &= \theta_{i,j}^{2m} U_{i,j}^{2m} , \\ (I_1^{\eta}\underline{U}^{2m})_{i,j} &= \eta_{i,j}^{2m} U_{i,j}^{2m} \end{aligned} \tag{3.6.2a}$$

$$\begin{aligned} I_2^{\theta} &= I - I_1^{\theta} \\ I_2^{\eta} &= I - I_1 . \end{aligned} \tag{3.6.2b}$$

where I is a block unitary matrix of $(N^2 \times N^2)$ order.

We introduce a coefficient matrix A and which is split into submatrices as follows:

$$\begin{aligned} (A\underline{U})_{i,j} &= -h^2 L_h U_{i,j} , \\ (H\underline{U})_{i,j} &= -h^2 L_h^{(1)} U_{i,j} \\ \text{and} \qquad (V\underline{U})_{i,j} &= -h^2 L_h^{(2)} U_{i,j} . \end{aligned} \tag{3.6.3}$$

We define also,

$$\begin{aligned} A_1 &= I_1^{\theta} H + I_1^{\eta} V , \\ A_2 &= I_2^{\theta} H + I_2^{\eta} V . \end{aligned} \tag{3.6.4}$$

The global form of the general formula (3.5.1) can be written as a two-step process:

$$\begin{aligned} (I+pA_1)\underline{U}^m &= (I-pA_2)\underline{U}^{m-1} + k(I_1^{\theta}\underline{g}^{(1)m} + I_1^{\eta}\underline{g}^{(2)m}) + k(I_2^{\theta}\underline{g}^{(1)m-1} + I_2^{\eta}\underline{g}^{(2)m-1}), \\ (I+pA_2)\underline{U}^{m+1} &= (I-pA_1)\underline{U}^{m} + k(I_2^{\theta}\underline{g}^{(1)m+1} + I_1^{\eta}\underline{g}^{(2)m+1}) + k(I_1^{\theta}\underline{g}^{(1)m} + I_2^{\eta}\underline{g}^{(1)m}), \end{aligned} \tag{3.6.5}$$

where $p = \Delta T/\Delta x^2 = k/h^2$ and m=1,3,5,... are odd integers with $\underline{U}^0$ given and the boundary values are absorbed in the g vector.

It can be seen that the class of methods given by (3.5.1) is equivalent to the Peaceman-Rachford procedure according to the splitting of the matrix A given by (3.6.4). Thus, we can rewrite the previously mentioned block method. For instance,

$$I_1^\theta = I_2^\theta = I_1^\eta = I_2^\eta = 1/2\ I,$$

leads to the splitting $A=1/2\ A_1 + 1/2\ A_2$ for the Crank-Nicolson scheme. Also,

$$I_1^\theta = I_2^\eta = I_1 \quad \text{with } (\text{diag } I_1)_{i,j} = \theta_{i,j}^m ,$$

$$I_2^\theta = I_1^\eta = I_1 = I - I_1 ,$$

where $\theta_{i,j}^m$ is a zero-one function gives,

$$A = (I_1H+I_2V)+(I_2H+I_1V)$$

for the ADI-Hopscotch method. The name ADI for this Hopscotch procedure is due to the splitting of A into H and V as in the normal ADI method and then H and V into I_1H, I_2H and I_1V, I_2V as in the true Hopscotch method.

For the stability of (3.6.5) we eliminate $\underline{U}^m$ to obtain

$$\underline{U}^{m+1} = T\underline{U}^{m-1} + \underline{d}^m \quad , \ m=1,3,5,... \tag{3.6.6}$$

where $\underline{d}^m$ is independent of the solution $\underline{u}$ and T is the amplification matrix with the following values:

$$\begin{aligned} T &= [I+p(I_2^\theta H+I_2^\eta V)]^{-1}[I+p(I_1^\theta H+I_1^\eta V)]^{-1}[I-p(I_1^\theta H+I_1^\eta V)][I-p(I_2^\theta H+I_2^\eta V)] \\ &= [I+p(I_2^\theta H+I_2^\eta V)]^{-1}\tilde{T}[I+p(I_2^\theta H+I_2^\eta V) , \end{aligned} \tag{3.6.7}$$

and

$$\tilde{T} = [I+p(I_1^\theta H+I_1^\eta V)]^{-1}[I-p(I_1^\theta H+I_1^\eta V)][I-p(I_2^\theta H+I_2^\eta V)][I+p(I_2^\theta H)I_2^\eta V)]^{-1}. \tag{3.6.8}$$

Then,

$$T^m = [I+p(I_2^\theta H+I_2^\eta V)]^{-1}\tilde{T}^m[I+p(I_2^\theta H+I_2^\eta V)] \tag{3.6.9}$$

Hence, applying the L_2-norm gives rise to:

$$||T^m|| \leq ||[I+p(I_2^{\theta}H+I_2^{\eta}V]^{-1}||.||T^m||.||[I+p(I_2^{\theta}H+I_2^{\eta}V)]||. \tag{3.6.10}$$

We now state Kellog's lemmas which can be used to prove the stability of the methods.

Lemma 3.1

If $\omega>0$ and $D+D^*$ is non-negative definite, then $\omega I+D$ has a bounded inverse and

$$||(\omega I+D)^{-1}|| \leq 1/\omega \ .$$

Lemma 3.2

If $\omega>0$ and $D+D^*$ is non-negative definite then the operator $(\omega I-D)(\omega I+D)^{-1}$ is bounded and has norm less than or equal to unity.

Therefore, if $(I_2^{\theta}H+I_2^{\eta}V)+(I_2^{\theta}H+I_2^{\eta}V)^*$ is non-negative definite according to Lemma 3.1, we have

$$||[I+p(I_2^{\theta}H+I_2^{\eta}V)]^{-1}|| \leq 1 \tag{3.6.11}$$

and also from Lemma 3.2 provided $(I_1^{\theta}H+I_1^{\eta}V)+(I_1^{\theta}H+I_1^{\eta}V)^*$ is non-negative definite then

$$||\tilde{T}|| \leq 1 \ ,$$

and $||\tilde{T}||^m \leq c$, where c is some constant (and p is fixed). Consequently the stability of the scheme under consideration is guaranteed, and we have:

Theorem 3.3

The Hopscotch process is stable for the solution of (3.4.1) if $I_1^{\theta}H+I_1^{\eta}V$ and $I_2^{\theta}H+I_2^{\eta}V$ both satisfy

$$(U,D_iU)+(D_iU,U) \geq 0$$

where $D_i=I_i^{\theta}H+I_i^{\eta}V$, i=1,2, for all real, appropriate ordered vectors $U\neq 0$.

Corollary

The above condition is satisfied if $(U,D_iU)+(D_iU,U)\geq 0$ for $D_i=I_i^{\theta}H$, $I_i^{\eta}V$, i=1,2, which can be used for stability of (3.5.1).

When I_1^θ and $I_1^\eta \in [0,1]$, in other words, the entries of I_1^θ and I_1^η are zero and one respectively, we have,

$$0 \leq I_i^\eta \text{ and } I_i^\theta \leq 1 \quad , i=1,2$$

Now if $I_1^\theta = I_1^\eta = I_1$ and $I_2^\theta = I_2^\eta = I_2 = I - I_1$ we obtain

$$I_1^\theta H + I_1^\eta V = I_1 A \quad \text{and} \quad I_2^\theta H + I_2^\eta V = I_2 A.$$

In this case, since A is positive definite, $A^{\frac{1}{2}}$ exists and we can write:

$$[I+pI_1A]^{-1}[I-pI_1A] = A^{-\frac{1}{2}}[I+pA^{\frac{1}{2}}I_1A^{\frac{1}{2}}]^{-1}[I-pA^{\frac{1}{2}}I_1A^{\frac{1}{2}}]A^{\frac{1}{2}} = A^{-\frac{1}{2}}A_1A^{\frac{1}{2}}$$

where $$A_1 = [I+pA^{\frac{1}{2}}I_1A^{\frac{1}{2}}]^{-1}[I-pA^{\frac{1}{2}}I_1A^{\frac{1}{2}}] \ . \tag{3.6.12}$$

In the same way, we obtain

$$[I+pI_2A]^{-1}[I-pI_2A] = A^{-\frac{1}{2}}A_2A^{\frac{1}{2}} \ ,$$

and

$$\tilde{T} = A^{-\frac{1}{2}}A_1A_2A^{\frac{1}{2}} \ . \tag{3.6.13}$$

Therefore

$$T^m = [I+pI_2A]^{-1}\tilde{T}^m[I+pI_2A] = [I+pI_2A]^{-1}A^{-\frac{1}{2}}(A_1A_2)^mA^{\frac{1}{2}}[I+pI_2A]$$

or

$$T^m = [A^{\frac{1}{2}}(I+pI_2A)]^{-1}(A_1A_2)^m[A^{\frac{1}{2}}(I+pI_2A)] \ . \tag{3.6.14}$$

Also, the matrices $A^{\frac{1}{2}}I_2A^{\frac{1}{2}}$ and $A^{\frac{1}{2}}I_1A^{\frac{1}{2}}$ are non-negative definite since, for instance,

$$A^{\frac{1}{2}}I_1A^{\frac{1}{2}} = (I_1A^{\frac{1}{2}})^*(I_1A^{\frac{1}{2}}) \ ,$$

when A is symmetric, since I_1 (and I_2) is diagonal with non-negative elements. Therefore, $A^{\frac{1}{2}}I_1A^{\frac{1}{2}}$ and $A^{\frac{1}{2}}I_2A^{\frac{1}{2}}$ are non-negative definite, symmetric and Kellog's 2nd lemma can be used to show that

$$||A_i|| \leq 1, \ i=1,2, \ ,$$

which implies that for stability, we require,

$$||T^m|| \leq ||[A^{\frac{1}{2}}(I+pI_2)]^{-1}||.||A^{\frac{1}{2}}(I+pI_2A)|| < c \ . \tag{3.6.15}$$

where c is a constant independent of m.

Hence we can write

$$[A^{\frac{1}{2}}(I+pI_2A)]^{-1} = \{[I+pA^{\frac{1}{2}}I_2A^{\frac{1}{2}}]A^{\frac{1}{2}}\}^{-1} = A^{-\frac{1}{2}}[I+pA^{\frac{1}{2}}I_2A^{\frac{1}{2}}]^{-1} ,$$

by the Kellog's 1st lemma we find that

$$||T^m|| \leq ||A^{-\frac{1}{2}}||.||A^{\frac{1}{2}}[I+pI_2A]|| \leq c \qquad (3.6.16)$$

since p is fixed and A is a positive definite matrix.

Therefore we may state:

Theorem 3.4

The two step process (3.5.1) is stable for the solution of (3.4.1) if the matrix A is positive definite whence $I_1^{\theta}=I_1^{\eta}=I_1$ and $I_2^{\theta}=I_2^{\eta}+I-I_1$ (Gourlay, 1970, p.380).

However, for a general class of Hopscotch process which can be derived from the formula (3.5.1), Gourlay and McGuire (1971) have applied the same analysis of Theorem 3.4 and shown that the general stability Theorem is as given earlier in Theorem 3.3. But in order to cast the Theorem 3.3 in a more workable form, the following lemma is given:

Lemma 3.3

If the N×N non-singular matrix A has n independent left eigenvectors, then:

i) the matrices I_1A and I_2A also have N independent left eigenvectors, and

ii) the null spaces of I_1A and I_2A are disjoint.

Proof: See Gourlay (1970), p.381.

Thus, the following theorem by McGuire is immediately obtainable:

Theorem 3.5

The algorithm (3.5.1) is stable for the solution of (3.4.1) in the case of $I_i^{\theta}, I_i^{\eta} \in \{0,1\}$ if the matrices H and V both have a full set of independent eigenvectors, are non-singular and if the eigenvalues of the

matrices $I_1^{\theta}H, I_1^{\eta}V, I_2^{\theta}H$ and $I_2^{\eta}V$ have non-negative real parts.

For the special case $I_1^{\theta}=I_1^{\eta}=I_1$ this theorem becomes the Theorem 3.4.

A simple 2×2 counter-example however, shows that, the Theorem 3.3 which indicates the sufficient condition for stability of the class (3.5.1) is not generally applicable in this particular context.

Let the positive-definite matrix A and matrix I_1 be as follows:

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \qquad I_1 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Therefore

$$D = I_1 A = \begin{bmatrix} 0 & 0 \\ -1 & 2 \end{bmatrix} \quad \text{and} \quad D^* = \begin{bmatrix} 0 & -1 \\ 0 & 2 \end{bmatrix} \rightarrow D+D^* = \begin{bmatrix} 0 & -1 \\ -1 & 4 \end{bmatrix}$$

which results in the eigenvalues of D+D* being $2\pm\sqrt{5}$, showing that D+D* is not non-negative definite.

To remedy this difficulty Gane, C.R.(1974, p.81) has applied a more suitable norm and proved the stability for the class of Hopscotch process with respect to this norm.

To describe his analysis we first give some definitions:

Definition 3.1

The M_2 or *mean 2nd power norm* of a column vector $\underline{U}$ is denoted by $||\underline{U}||_{M_2}$ and is given as follows:

$$||\underline{U}||_{M_2} = \{(1/N) \sum_{\ell=1}^{N} |U_\ell|^2\}^{1/2} = N^{-\frac{1}{2}} ||\underline{U}||_2 \,. \qquad (3.6.17)$$

Definition 3.2

The $A^{\frac{1}{2}}$-*vector norm* for a vector $\underline{U}$ for a positive,definite matrix A is defined to be:

$$||\underline{U}||_{A^{\frac{1}{2}}} = ||A^{\frac{1}{2}}\underline{U}||_{M_2} \,, \qquad (3.6.18)$$

and the $A^{\frac{1}{2}}$-*matrix norm* subordinate to (3.6.18) is:

$$||D||_{A^{\frac{1}{2}}} = ||A^{\frac{1}{2}}DA^{-\frac{1}{2}}||_2 \ . \tag{3.6.19}$$

Now reconsidering (3.6.13), we have

$$\tilde{T} = A^{-\frac{1}{2}}D_2D_1A^{\frac{1}{2}}$$

where

$$D_i = [I-pA^{\frac{1}{2}}I_iA^{\frac{1}{2}}][I+pA^{\frac{1}{2}}I_iA^{\frac{1}{2}}]^{-1} \ , \quad i=1,2, \ .$$

Since A is positive definite, then $A^{\frac{1}{2}}$ exists and is also positive definite, and as mentioned earlier,

$$A^{\frac{1}{2}}I_iA^{\frac{1}{2}} = (I_iA^{\frac{1}{2}})^*(I_iA^{\frac{1}{2}}), \quad i=1,2$$

and it can be shown that $A^{\frac{1}{2}}I_iA^{\frac{1}{2}}$, i=1,2, are non-negative definite. Therefore,

$$||\tilde{T}||_{A^{\frac{1}{2}}} = ||A^{\frac{1}{2}}(A^{-\frac{1}{2}}D_2D_1A^{\frac{1}{2}})A^{-\frac{1}{2}}||_2 = ||D_2D_1||_2 \leq ||D_2||_2||D_1||_2 \leq 1, \tag{3.6.20}$$

since by Kellog's 2nd lemma, $||D_i||\leq 1$, i=1,2,.

Also, considering lemma 3.1 and 3.2 we respectively obtain,

$$||[I+pI_1A]^{-1}||_{A^{\frac{1}{2}}} = ||(I+pA^{\frac{1}{2}}I_1A^{\frac{1}{2}})^{-1}||_2 \leq 1, \tag{3.6.21a}$$

and

$$||I+pI_1A||_{A^{\frac{1}{2}}} = ||I+pA^{\frac{1}{2}}I_1A^{\frac{1}{2}}||_2 = \rho(I+pI_1A) \ , \tag{3.6.21b}$$

since $(I+pA^{\frac{1}{2}}I_1A^{\frac{1}{2}})$ is *symmetric* and $I_1A=A^{-\frac{1}{2}}(A^{\frac{1}{2}}I_1A^{\frac{1}{2}})A^{-\frac{1}{2}}$.

Now because I_1A is simply composed of rows of A and *null* rows, it follows from Gerschgorin's Theorem and the structure of A resulting by the difference replacement of (3.4.1) that,

$$\rho(I_1A) \leq M \tag{3.6.22}$$

where M is assumed bounded {e.g. if $L=\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}$ then A has 4 on the diagonal and -1 off diagonals, hence by Theorem 3.1 $\rho(I_1A)\leq\min(\max \sum_{i=1}^{N-1}|a_{i,j}|, \max \sum_{j=1}^{N-1}|a_{i,j}|)=8$, i.e. M=8}. It then follows from (3.6.21b) that

$$||I+pI_1A||_{A^{\frac{1}{2}}} \leq 1+pM \ .$$

Thus, for equation (3.6.10) we obtain,

$$||T^m||_{A^{\frac{1}{2}}} \leq 1+pM,$$

which is independent of integer m, assuming $p=k/h^2$ *remains fixed.* Then the stability Theorem holds as follows:

Theorem 3.6

The odd-even, line and peripheral Hopsoctch processes are stable (in the $A^{\frac{1}{2}}$ norm) for the solution of equation (3.4.1) if the matrix A defined by:

$$(Au)_{i,j} = -h^2 L_h u_{i,j}$$

is *positive definite* (Gane, C.R., 1974, p.82).

3.7 CONVERGENCE OF THE HOPSCOTCH ALGORITHMS

The convergence of the Hopscotch processes have been discussed by McGuire (1970) and Gane (1974). Here, we follow the analysis of the second author and study the convergence with respect to the A -norm which can be useful for further development.

Hence, we consider the general 2-step formula (3.5.1) and define L_1, L_2 and $b_{i,j}$ as follows:

$$L_1 = \theta_{i,j}^{m+1} L_h^{(1)} + \eta_{i,j}^{m+1} L_h^{(2)}$$

$$L_2 = \theta_{i,j}^{m} L_h^{(1)} + \eta_{i,j}^{m} L_h^{(2)} \quad , \tag{3.7.1}$$

and $$b_{i,j}^{m} = (\theta_{i,j}^{m} g_{i,j}^{(1)m} + \eta_{i,j}^{m} g_{i,j}^{(2)m}) + (\theta_{i,j}^{m+1} g_{i,j}^{(1)m-1} + \eta_{i,j}^{m+1} g_{i,j}^{(2)m-1}).$$

In addition the following also holds:

$$L_h^{(1)} U_{i,j} = \frac{1}{h^2} [B_{i,j} U_{i-1,j} + C_{i,j}^{(1)} U_{i,j} + D_{i,j} U_{i+1,j}] = L^{(1)} U_{i,j} + O(h^{\sigma})$$

$$L_h^{(2)} U_{i,j} = \frac{1}{h^2} [A_{i,j} U_{i,j-1} + C_{i,j}^{(2)} U_{i,j} + E_{i,j} U_{i,j+1}] = L^{(2)} U_{i,j} + O(h^{\sigma}) \tag{3.7.2}$$

where $A_{i,j}, B_{i,j} \ldots, E_{i,j}$ are independent of $U_{i,j}$ and σ is given by (2.11.1).

Thus, we have,

$$L^{(1)} + L^{(2)} = \frac{1}{h^2} [A_{i,j} U_{i,j-1} + B_{i,j} U_{i-1,j} + C_{i,j} U_{i,j} + D_{i,j} U_{i+j} + E_{i,j} U_{i,j+1}] =$$

$$L_h U_{i,j} = L U_{i,j} + O(h^{\sigma}) \tag{3.7.3}$$

where

$$C_{i,j} = C_{i,j}^{(1)} + C_{i,j}^{(2)}$$

Therefore, the general two step process (3.5.1) becomes:

$$\begin{aligned} (1 - kL_1) U_{i,j}^{m} &= (1 + kL_2) U_{i,j}^{m-1} + k b_{i,j}^{m} \\ (1 - kL_2) U_{i,j}^{m+1} &= (1 + kL_1) U_{i,j}^{m} + k b_{i,j}^{m+1} \end{aligned} \quad , \tag{3.7.4}$$

where according to L_1 and L_2 we can have different methods (e.g. odd-even

or line Hopscotch). The elimination of $U^m_{i,j}$ in (3.7.4) results in the single formula, namely,

$$(1-kL_1)(1-kL_2)U^{m+1}_{i,j} = (1+kL_1)(1+kL_2)U^{m-1}_{i,j} + R^m_{i,j} \tag{3.7.5}$$

where $R^m_{i,j} = 2kg^m_{i,j}+\theta^{m+1}_{i,j}O(k^3)+\eta^{m+1}_{i,j}O(k^3)-2k^3L_1(\theta^{m+1}_{i,j}\frac{\partial g^{(1)m}_{i,j}}{\partial t}+\eta^{m+1}_{i,j}$

$$\frac{\partial g^{(2)m}_{i,j}}{\partial t})+\theta^{m+1}_{i,j}\,O(k^5/h^2)+\eta^{m+1}_{i,j}O(k^5/h^2).$$

(McGuire, 1970, p.23), and we assume enough differentiability for each $u, g, g^{(1)}$ and $g^{(2)}$.

Let $e^m_{i,j}$ denote the difference between the exact and approximate solutions and the mesh points i.e.,

$$e^m_{i,j} = u^m_{i,j} - U^m_{i,j}$$

then, the *error equation* is deduced from (3.7.5) to be:

$$(1-kL_1)(1-kL_2)e^{m+1}_{i,j} = (1+kL_1)(1+kL_2)e^{m-1}_{i,j} + d^m_{i,j} \tag{3.7.6}$$

where

$$d^m_{i,j} = 2k^3L_1L_2(\frac{\partial u}{\partial t})^m_{i,j}+2k^3L_1(\theta^{m+1}_{i,j}\frac{\partial g^{(1)m}_{i,j}}{\partial t} + \eta^{m+1}_{i,j}\frac{\partial g^{(2)m}_{i,j}}{\partial t})+O(k^3+kh^\sigma)$$

$$+\theta^{m+1}_{i,j}O(k^5/h^2)+\eta^{m+1}_{i,j}O(k^5/h^2).$$

Hence, the L.T.E. of (3.5.1) denoted by $d^m_{i,j}$ is,

$$d^m_{i,j} = O(kh^\sigma+k^3/h^2+k^3)\ , \tag{3.7.7}$$

where $\sigma=2$, or 1 depending on whether (i,j) is a regular, or irregular mesh point on the grid respectively. Here we suppose that the region under consideration contains only the regular mesh points so that $\sigma=2$.

We note that, there is a potential weakness in the Hopsoctch technique in that consistency relation requires k to be $O(h^2)$. This is similar to the weakness found in the Du-Fort-Frankel scheme to which the hopscotch algorithm can be related. This will be shown later.

The *global error* equation can be found for equation (3.5.1) by considering the whole region and written in the same manner as before,

$$(I-pI_2A)(I+pI_1A)\underline{e}^{m+1} = (I-pI_2A)(I-pI_1A)\underline{e}^{m-1}+\underline{d}^m \qquad (3.7.8)$$

where m is an odd integer and I_1, I_2, A are defined so that different hopscotch algorithms are denoted. The equation (3.7.8) can also be represented in the following form (replacing m by 2r+1),

$$\underline{e}^{2r+2} = T\underline{e}^{2r} + \underline{f}^{2r+1} \quad , \; r=0,1,2,\ldots \qquad (3.7.9)$$

where T is given in (3.6.7) and

$$\underline{f}^{2r+1} = (I+pI_1A)^{-1}(I+pI_2A)^{-1}\underline{d}^{2r+1} \; . \qquad (3.7.10)$$

Thus, from (3.7.9) we have

$$\underline{e}^{2r} = T^r\underline{e}^0+T^{r-1}\underline{f}+T^{r-1}\underline{f}^3+\ldots+T\underline{f}^{2r-3}+\underline{f}^{2r-1} \; . \qquad (3.7.11)$$

Now, if the processes are stable, we have $||T^r||_{A^{\frac{1}{2}}} \leq c$ (constant) for all r>0. Also from (3.7.10) we have

$$||\underline{f}^{2r+1}|| \leq ||(I+pA^{\frac{1}{2}}I_1A^{\frac{1}{2}})^{-1}||.||(I+pA^{\frac{1}{2}}I_2A^{\frac{1}{2}})^{-1}||.||\underline{d}^{2r+1}||$$

where $||.||$ denotes the $A^{\frac{1}{2}}$-norm.

The application of Lemma 3.1 indicates that

$$||\underline{f}^{2r+1}||_{A^{\frac{1}{2}}} \leq ||\underline{d}^{2r+1}||_{A^{\frac{1}{2}}},$$

Hence, we can write

$$||\underline{e}^{2r}||_{A^{\frac{1}{2}}} \leq ||T^r||_{A^{\frac{1}{2}}}||\underline{e}^0||_{A^{\frac{1}{2}}} + ||R||_{A^{\frac{1}{2}}}$$

while $$||R||_{A^{\frac{1}{2}}} = ||T^{r-1}\underline{f}+T^{r-2}\underline{f}^3+\ldots+\underline{f}^{2r-1}||_{A^{\frac{1}{2}}} \; .$$

Thus, $$||R||_{A^{\frac{1}{2}}} \leq r.c \max_{1\leq\ell\leq r} ||\underline{d}^{2\ell-1}||_{A^{\frac{1}{2}}} \; .$$

If by assumption, $||\underline{e}^0||=||\underline{u}^0-\underline{U}^0||=0$, then we obtain,

$$||\underline{e}^{2r}||_{A^{\frac{1}{2}}} \leq r.c \max_{1\leq\ell\leq r} ||\underline{d}^{2\ell-1}||_{A^{\frac{1}{2}}} \; . \qquad (3.7.12)$$

Since $d^{m}_{i,j}=O(kh^2+k^3/h^2+k^3)$ for all i,j,m and

$$||\underline{d}^{2\ell-1}||_{A^{\frac{1}{2}}} = ||A^{\frac{1}{2}}\underline{d}^{2\ell-1}||_{M_2} \leq ||A^{\frac{1}{2}}||_{M_2}.||\underline{d}^{2\ell-1}||_{M_2} =$$

$$= ||A^{\frac{1}{2}}||_2||\underline{d}^{2\ell-1}||_{M_2} , \qquad (3.7.13)$$

it follows from (3.7.7) that,

$$||\underline{d}^{2\ell-1}||_{M_2} = O(kh^2+k^3/h^2+k^3), \text{ for all } \ell\geq 1 \quad i,j \qquad (3.7.14)$$

Since $||A^{\frac{1}{2}}||_2=\rho(A^{\frac{1}{2}})=[\rho(A)]^{\frac{1}{2}}=(M)^{\frac{1}{2}}$, is bounded for a positive definite matrix A, ($\rho(A)\leq M$ was assumed for stability) therefore,

$$||\underline{d}^{2\ell-1}||_{A^{\frac{1}{2}}} \leq M^{\frac{1}{2}}.O(kh^2+k^3/h^2+k^3) . \qquad (3.7.15)$$

It now follows from (3.7.9) and (3.7.12) that

$$||\underline{e}^{2r}||_{A^{\frac{1}{2}}} \leq r.k.c.M^{1/2}.O(h^2+k^2/h^2+k^2) \qquad (3.7.16)$$

and the convergence of (3.7.6) can be achieved provided k↦0 faster than h which means, *k/h^2 must be kept constant*, which is an important matter for convergence. We now state the following theorem:

Theorem 3.7

The formula (3.5.1) for the solution of (3.4.1) is convergent in the $A^{\frac{1}{2}}$ norm with the rate of $O(k+h^2)$ if the mesh ratio $p=k/h^2$ is constant, and if the processes are stable and the region of the solution deals only with regular mesh points which guarantee the positive definiteness of the matrix A. (Gane, 1974, p.85).

The principal part of L.T.E.

In the expression for $d^{m}_{i,j}$, (3.7.5) the terms involving $O(k^3)$ and $O(kh^{\sigma})$ are the same for different methods deduced from the general formula (3.5.1). The only terms which are different are those involving $O(k^3/h^2)$. McGuire (1970). The magnitude of this term can be used for the comparison

of the accuracy of different methods. Here we shall look at this coefficient in the general formula (3.5.1) and compare different methods.

The terms involving $O(k^3/h^2)$ are as follows:

$$2k^3L_1L_2\frac{\partial u^m_{i,j}}{\partial t} \quad \text{and} \quad k^3L_1(\theta^{m+1}_{i,j}\frac{\partial g^{(1)m}_{i,j}}{\partial t}+\eta^{m+1}_{i,j}\frac{\partial g^{(2)m}_{i,j}}{\partial t}),$$

where $L_1, L_2, g^{(1)}$ and $g^{(2)}$ are as given earlier. Considering the definition of L_1 and L_2 which are given in (3.7.1)-(3.7.3) one can easily show that

$$2k^3L_1L_2\frac{\partial u^m_{i,j}}{\partial t} = 2k^3\{\theta^{m+1}_{i,j}L^{(1)}_h\theta^m_{i,j}L^{(1)}_h+\theta^{m+1}_{i,j}L^{(1)}_h\eta^m_{i,j}L^{(2)}_h+\eta^{m+1}_{i,j}L^{(2)}_h\theta^m_{i,j}L^{(1)}_h+$$

$$+\eta^{m+1}_{i,j}L^{(2)}_h\eta^m_{i,j}L^{(2)}_h\}\frac{\partial u^m_{i,j}}{\partial t} \qquad (3.7.17)$$

Also, for the second term one obtains,

$$k^3L_1(\theta^{m+1}_{i,j}\frac{\partial g^{(1)m}_{i,j}}{\partial t}+\eta^{m+1}_{i,j}\frac{\partial g^{(2)m}_{i,j}}{\partial t}) = k^3\{\theta^{m+1}_{i,j}L^{(1)}_h\theta^{m+1}_{i,j}\frac{\partial g^{(1)m}_{i,j}}{\partial t}+$$

$$+\theta^{m+1}_{i,j}L^{(1)}_h\eta^{m+1}_{i,j}\frac{\partial g^{(2)m}_{i,j}}{\partial t}+\eta^{m+1}_{i,j}L^{(2)}_h\theta^{m+1}_{i,j}\frac{\partial g^{(1)m}_{i,j}}{\partial t}+$$

$$+\eta^{m+1}_{i,j}L^{(2)}_h\eta^{m+1}_{i,j}\frac{\partial g^{(2)m}_{i,j}}{\partial t}\}. \qquad (3.7.18)$$

Now for different methods, we have different values for $\theta^m_{i,j}, \eta^m_{i,j}$ which results in different principal parts for the L.T.E.

For instance, the Crank-Nicolson method is achieved from (3.5.1), when $\theta^{m+1}_{i,j}=\theta^m_{i,j}=1/2$ and $\eta^{m+1}_{i,j}=\eta^m_{i,j}=0$. Therefore (3.7.17) becomes $k^3/2(L_h)^2.\frac{\partial u^m_{i,j}}{\partial t}$ and (3.7.18) will be $k^3/4\frac{\partial g^m_{i,j}}{\partial t}$.

For those methods with $\theta^m_{i,j}=\eta^m_{i,j}$, we have,

$$2k^3L_1L_2\frac{\partial u^m_{i,j}}{\partial t} = 2k^3\,\theta^{m+1}_{i,j}L_h\theta^m_{i,j}L_h\frac{\partial u^m_{i,j}}{\partial t} \qquad (3.7.19)$$

and

$$k^3L_1(\theta^{m+1}_{i,j}\frac{\partial g^{(1)m}_{i,j}}{\partial t}+\eta^{m+1}_{i,j}\frac{\partial g^{(2)m}_{i,j}}{\partial t}) = k^3\theta^{m+1}_{i,j}L_h\theta^{m+1}_{i,j}\frac{\partial g^m_{i,j}}{\partial t} \qquad (3.7.20)$$

Thus, for the odd-even hopscotch method (where $\theta^m_{i,j}$ is defined earlier), the $O(k^3/h^2)$ terms of the L.T.E. is

$$k^3/h^2\theta^{m+1}_{i,j}C_{i,j}(-2L\frac{\partial u^m_{i,j}}{\partial t}+2\frac{\partial g^m_{i,j}}{\partial t}) , \qquad (3.7.21)$$

whilst for the line-hopscotch method (3.7.19) becomes,

$$k^3/h^2\theta^{m+1}_{i,j}(B_{i,j}+D_{i,j})(2L\frac{\partial u^m_{i,j}}{\partial t}-2\frac{\partial g^m_{i,j}}{\partial t}) \qquad (3.7.22)$$

where $B_{i,j}$, $C_{i,j}$ and $D_{i,j}$ are given in (3.7.2). For the usual elliptic operator, where the matrix in (3.6.3) is diagonally dominant, (3.7.22) has a smaller magnitude than (3.7.21) which indicates that, the line hopscotch will be more accurate than the odd-even hopscotch process.

Compensation of L.T.E. and the role of mesh ratio

Although the principal part of the Local Truncation errors illustrated above, shows that the Crank-Nicolson scheme is more accurate than the hopscotch processes, the value of $p=k/h^2$ also plays a role regarding the accuracy of the hopscotch techniques.

To show this, following Danaee (1978), we consider a simple one-dimensional heat conduction equation (3.2.1) where the odd-even hopscotch is employed to evaluate the approximate solution.

We now consider two successive time-steps, and find the L.T.E. at the point where the solution is found explicitly (E) as well as where it is found implicitly (I).

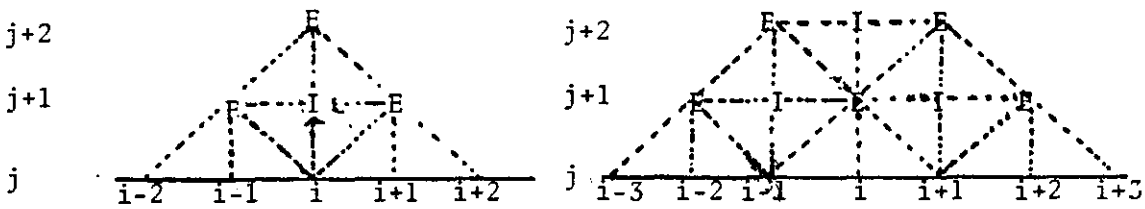


FIGURE 3.7.1

It has been shown that (Danaee, A., 1978, p.18), the principal part of L.T.E. at the explicit points (E) is

$$L.T.E. = 2(1+2p)k^2\left(\frac{\partial^2 u}{\partial t^2}\right)_{i,j} - \frac{kh^2}{6}(1+14p)\frac{\partial^4 u}{\partial x^4} + \ldots \tag{3.7.25}$$

while at implicit points (I), (3.7.21) becomes

$$L.T.E. = 2(1+2p)^2k^2\left(\frac{\partial^2 u}{\partial t^2}\right)_{i,j} - \frac{kh^2}{6}(52p^2+16p+1)\frac{\partial^4 u}{\partial x^4} + \ldots \tag{3.7.26}$$

However, for the simple heat equation (3.2.1), we can also obtain by differentiation,

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^4 u}{\partial x^4} \quad , \tag{3.7.27}$$

and substituting (3.7.25) in (3.7.23) and (3.7.24), the principal part of (3.7.23) and (3.7.24) become:

$$E_1 = \frac{kh^2}{6}(24p^2-2p-1)\frac{\partial^4 u}{\partial x^4} \tag{3.7.28}$$

$$E_2 = \frac{kh^2}{6}(48p^3-4p^2-4p-1)\frac{\partial^4 u}{\partial x^4} \quad . \tag{3.7.29}$$

The smallest truncation error occurs when (3.7.28) and (3.7.29) are minimized.

The graph of the two functions of p in E_1 and E_2 are displayed in Figure 3.7.2. As can be seen, at $p \simeq 0.33$ we obtain a more accurate result for the solution of the heat equation.

To compare the error behaviour of the hopscotch and Crank-Nicolson methods, we consider the previous heat example and draw the graph of the global error for both methods (Figure 3.7.3).

As the graph shows, for small time-steps (p small), the hopscotch procedure is more accurate than the Crank-Nicolson scheme while for large time-step (p large), the second method provides a much more accurate solution than the first. (We shall remember that in this case the noise effect appears).

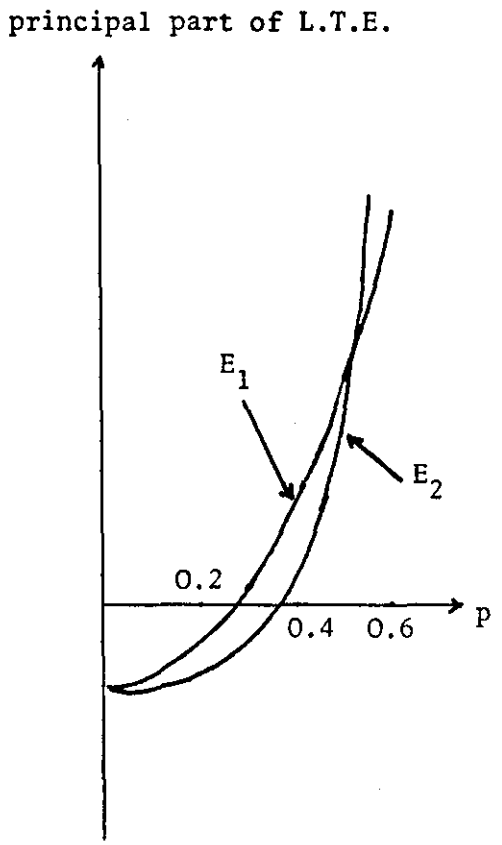


FIGURE 3.7.2

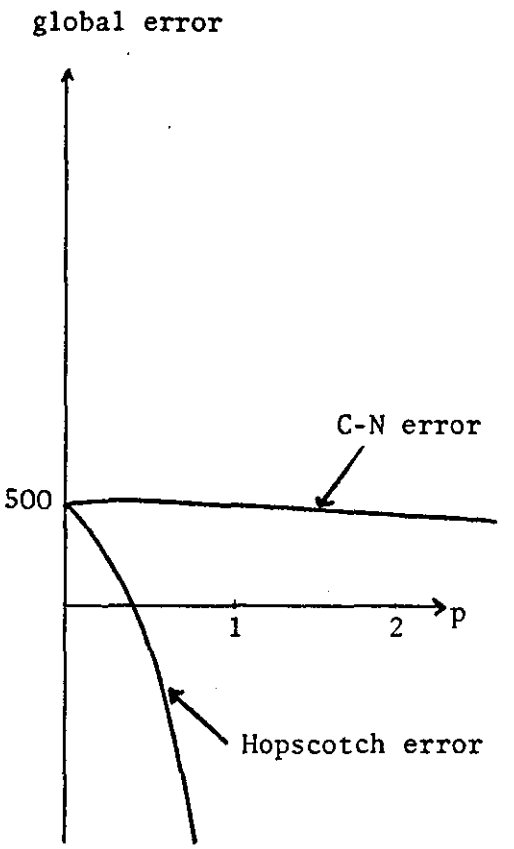


FIGURE 3.7.3

Here we conclude that, not only the stability of implicit formula in the hopscotch processes dominates the restruction on the stability of the explicit formula, but it compensates the L.T.E. in the process.

Therefore, different combinations of explicit and implicit formulae will give different L.T.E. In other words, the degree of implicitness of the scheme being used may also play a role. We shall be considering this matter later.

3.8 BLOCK ITERATIVE METHODS

In all iteration methods for the solution of a linear system of equations $\underline{Au}=\underline{b}$, considered so far, the value of each component of $\underline{u}_{(m)}$ is determined by an *explicit* linear formula, of which (2.17.8) is typical. By *explicit* we mean that the mth approximation to the component $u_{i(m)}$ of $\underline{u}_{(m)}$ can be determined by it at its proper step of the algorithm, without the necessity of simultaneously determining a group of components of $\underline{u}_{(m)}$.

The *implicit* method or *block iterative* methods are formulae by which a group of components of $\underline{u}_{(m)}$ are defined simultaneously in such a way that it is necessary to solve a linear sub-system for the whole subset of components at once before the solution can be evaluated. The advantage of implicit over explicit methods is that the rate of convergence of the former may be appreciably greater at the cost of some complication in the method.

Typical examples of implicit methods are ADI and SLOR (Successive Line Over Relaxation) which we describe here. To demonstrate the methods we begin with a model problem of *Dirichlet* kind, namely,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \ , \qquad (3.8.1)$$

subject to $u=f(x,y)$ on the boundary of the unit square $0\leq x,y\leq 1$. To obtain the finite difference approximation to (3.8.1), we consider a rectangular net which covers the square region with sides parallel to the coordinate axes with grid spacing h. If $N.h=1$, the number of internal grid points (nodes) is $(N-1)^2$, and if the five points formula is applied to solve (3.8.1) at all the internal nodes, a linear system of equation $\underline{Au}=\underline{b}$ can be obtained where A is a matrix of order $(N-1)^2$ given by a general form,

$$A = \begin{bmatrix} B_1 & C_1 & & & \\ A_2 & B_2 & C_2 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & C_{N-2} \\ 0 & & & A_{N-1} & B_{N-1} \end{bmatrix} \qquad (3.8.2)$$

with the submatrices of order (n-1) as follows:

$$B_i = \begin{bmatrix} 4 & -1 & & & \\ -1 & 4 & -1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -1 \\ 0 & & & -1 & 4 \end{bmatrix} \quad ; A_{i+1} = C_i = \begin{bmatrix} -1 & & & \\ & \ddots & & 0 \\ & & \ddots & \\ 0 & & & -1 \end{bmatrix} \qquad (3.8.3)$$

$1 \leq i \leq N-1.$

The vectors $\underline{u}$ and $\underline{b}$ are given by:

$$\underline{u} = \{u_{1,1}, u_{1,2}, \ldots, u_{1,N-1}, u_{2,1}, \ldots, u_{2,N-1}, \ldots, u_{N-1,1}, \ldots, u_{N-1,N-1}\}^T$$

and

$$\underline{b} = \{b_{1,1}, \ldots, b_{1,N-1}, b_{2,N-1}, \ldots, b_{2,N-1}, \ldots, b_{N-1,1}, \ldots, b_{N-1,N-1}\}^T \qquad (3.8.4)$$

respectively (Mitchel, A.R., 1976, p.102).

In an analogous manner to point SOR, we can define the line-SOR when we consider the group of nodes on one line and achieve the theoretical result as before, provided the coefficient A is *block consistently ordered* and possesses *property* A^{π} (or *block property A*). These properties are discussed in Varga, 1962, p.196.

For the SLOR method with *Dirichlet* boundary conditions, the B_i in (3.8.3) will be tridiagonal, and it will be necessary to solve subsystems of equations of the form

$$A_i\underline{\tilde{u}}_{i-1(m+1)}+B_i\underline{\tilde{u}}_{i(m+1)}+C_i\underline{\tilde{u}}_{i+1(m)} = \underline{b}_i,\ i=1,2,\ldots,N-1,$$

$$(A_1=C_{N-1}=0). \qquad (3.8.5)$$

or $$B_i\underline{\tilde{u}}_{i(m+1)} = \underline{d}\ ,\quad i=1,2,\ldots,N-1 \qquad (3.8.6)$$

where $\underline{\tilde{u}}_i$ is the column vector of values along the ith line and $\underline{d}_i$ $(1\leq i\leq N-1)$ are known since $\underline{\tilde{u}}_{i-1(m+1)}$ and $\underline{\tilde{u}}_{i+1(m)}$ are evaluated previously. The optimum value for the relaxation parameter now becomes

$$\omega_{opt} = \frac{2}{1+\sqrt{1-(\rho^L)^2}}\ , \qquad (3.8.7)$$

where ρ^L is the spectral radius of the Jacobi line method (Young, 1971, p.453). If the iterative matrix of line SOR is denoted by L_ω^L then it can be shown (Varga, 1962, p.204) that

$$\rho(L_J^L) = \frac{\cos(\frac{\pi}{N})}{2-\cos(\frac{\pi}{N})}\ , \qquad (3.8.8)$$

from which we can define the asymptotic rate of convergence to be

$$R_\infty(L^L_{\omega_{opt}}) \simeq \frac{2\sqrt{2}\pi}{N} \qquad N\to\infty \qquad (3.8.9)$$

so that,

$$\frac{R_\infty(L^L_{\omega_{opt}})}{R_\infty(L^{pt}_{\omega_{opt}})} \simeq \sqrt{2} \qquad N\to\infty \qquad (3.8.10)$$

for this model problem. In other words, for large N, the line successive over-relaxation iterative method yields an *increase* of approximately 40 per cent in the asymptotic rate of convergence of the point successive over-relaxation method. An important gain achieved here is that the criterion (3.8.10) is fixed and independent of the mesh spacing h=1/N (Varga, 1962, p.205)

Analogous to the line SOR method, one can apply a k-line SOR method for integer k>1. A comparison of the type (3.8.10) has been made by Varga where

where he shows that,

$$\frac{R_\infty(L_{\omega_{opt}}^{2L})}{R_\infty(L_{\omega_{opt}}^{L})} \simeq \sqrt{2} \quad N\to\infty \tag{3.8.11}$$

As shown in Cuthill and Varga (1959) and Varga (1960), it is possible in many cases to perform both single and double-line over-relaxation in approximately the same number of arithmetic operations per mesh point as required by the point SOR method. Therefore, the increase in rate of convergence of line-iterative methods over point-iterative methods results in corresponding decreases in total computational effort. However, for k-line SOR when k>1 some practical difficulties may arise which results in a little gain in comparison with the line-SOR (k=1) method which is the most significant iterative over-relaxation method for the model problem.

Symmetric SOR method

The *symmetric* SOR method (SSOR method) can be considered as two half iterations. The first half iteration is the same as the SOR method, while the second half iteration is the SOR with the equations taken in reverse order (Young, 1971, p.461).

Thus, if $\underline{u}_{(m+\frac{1}{2})}$ is determined from $\underline{u}_m$ by the forward SOR method, i.e.

$$\underline{u}_{(m+\frac{1}{2})} = L_\omega \underline{u}_{(m)} + (I-\omega L)^{-1}\omega\underline{c} \tag{3.8.12}$$

and $\underline{u}_{(m+1)}$ from $\underline{u}_{(m+\frac{1}{2})}$ by backward SOR method, then we obtain

$$\underline{u}_{(m+1)} = U_\omega \underline{u}_{(m+\frac{1}{2})} + (I-\omega U)^{-1}\omega\underline{b} , \tag{3.8.13}$$

where

$$\begin{aligned} L_\omega &= (I-\omega L)^{-1}(\omega U+(1-\omega)I) \\ U_\omega &= (I-\omega U)^{-1}(\omega L+(1-\omega)I) . \end{aligned} \tag{3.8.14}$$

Eliminating $\underline{u}_{(m+\frac{1}{2})}$ in (3.8.12) and (3.8.13) results in the following formula,

$$\underline{u}_{(m+1)} = U_\omega L_\omega \underline{u}_{(m)} + \omega(2-\omega)(I-\omega U)^{-1}(I-\omega L)^{-1}\underline{c} \qquad (3.8.15)$$

which indicates that

$$U_\omega L_\omega = I-\omega(2-\omega)(I-\omega U)^{-1}(I-\omega L)^{-1}D^{-1}A$$

where D is the diagonal of matrix A of the linear system to be solved, and $\underline{c}=D^{-1}\underline{b}$. Hence, $U_\omega L_\omega$ is non-singular if $0<\omega<2$ and if A is non-singular.

The analysis of convergence and strategies for obtaining the best value of ω for the SSOR method is given by Young (1971), p.462, and it is shown that SSOR method is convergent if and only if A is positive definite and $0<\omega<2$.

Alternating Direction Implicit Iterative Method

As it was shown previously the Peaceman-Rachford method for solving the heat conduction equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \qquad (3.8.16)$$

is given by

$$(1-\tfrac{1}{2}p.\delta_x^2)(1-\tfrac{1}{2}p\delta_y^2)U_{i,j}^{n+1} = (1+\tfrac{1}{2}p\delta_x^2)(1+\tfrac{1}{2}p\delta_y^2)U_{i,j}^{n} \ , \qquad (3.8.17)$$

after elimination of the intermediate values $U_{i,j}^{n+\frac{1}{2}}$. The equation (3.8.17) can be taken to represent an iteration procedure which converges if

$$U_{i,j}^{n} = U_{i,j}^{n+1} = U_{i,j} \ , \qquad (3.8.18)$$

for sufficiently large values of n. Substitution of (3.8.18) in (3.8.17) gives rise to the standard five point difference replacement for the Laplace equation. Therefore, the Peaceman-Rachford method (given in the previous chapter) applied to the heat equation (3.8.16) with the boundary conditions independent of the time, represents an iterative method for solving the Laplace equation in a square with Dirichlet boundary conditions (Mitchell, A.R., 1976, p.104). The parameter p in (3.8.17) is no longer the mesh ratio but is an *iteration parameter* which may be varied from iteration to iteration.

For the convergence analysis of the ADI method, we write the equation (3.8.17) for all the $(N-1)^2$ internal grid points of the unit square to obtain the matrix form, namely,

$$\left(\frac{2}{p_{m+1}} I + H\right)\underline{U}^*_{(m+1)} = \left(\frac{2}{p_{m+1}} I - V\right)\underline{U}_{(m)} + \underline{b}$$

and

$$\left(\frac{2}{p_{m+1}} I + V\right)\underline{U}_{(m+1)} = \left(\frac{2}{p_{m+1}} I - H\right)\underline{U}^*_{(m+1)} + \underline{b} \qquad (3.8.19)$$

respectively where p_{m+1} is the *variable iteration parameter*. In the equation (3.8.19), I is the unitary matrix of order $(N-1)^2$ and, H and V are commutative matrices of order $(N-1)^2$ given by:

$$H = \begin{bmatrix} B & & & \\ & B & & 0 \\ & & \ddots & \\ 0 & & & B \end{bmatrix}, V = \begin{bmatrix} 2J & -J & & \\ -J & 2J & -J & 0 \\ & \ddots & \ddots & \ddots \\ 0 & & -J & 2J \end{bmatrix} \quad \text{with } B = \begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & 0 \\ & \ddots & \ddots & \ddots \\ 0 & & -1 & 2 \end{bmatrix}$$

where B and J are of order (N-1), J is the unitary matrix of order N-1 $\underline{U}$ and $\underline{b}$ are vectors given by (3.8.4).

The two equations (3.8.19) can be given in a combined form as

$$\underline{U}_{(m+1)} = T_{m+1}\underline{U}_{(m)} + \underline{g} \qquad (3.8.20)$$

where

$$T_{m+1} = \left(\frac{2}{p_{m+1}} I + V\right)^{-1}\left(\frac{2}{p_{m+1}} I - H\right)\left(\frac{2}{p_{m+1}} I + H\right)^{-1}\left(\frac{2}{p_{m+1}} I - V\right)$$

and

$$\underline{g} = \left(\frac{2}{p_{m+1}} I + V\right)^{-1}\left[\left(\frac{2}{p_{m+1}} I - H\right)\left(\frac{2}{p_{m+1}} I + H\right)^{-1} + I\right]\underline{b} \; . \qquad (3.8.21)$$

For the convergence of (3.8.20) we require $\rho(T_{m+1}) \leq 1$.

If p_{m+1} is constant, then T_{m+1} is constant for all m, and the convergence of (3.8.20) is guaranteed if $\rho(T) \leq 1$.

Since H and V commute, they possess a common set of orthonormal eigenvectors $\alpha_{r,s}$ $(1 \leq r,s \leq N-1)$ with the corresponding eigenvalues given by,

$$\lambda_{r,s}(H) = 4 \sin^2 \frac{r\pi}{2N}$$

$$\mu_{r,s}(V) = 4 \sin^2 \frac{s\pi}{2N} \qquad 1 \leq r,s \leq N-1.$$

Hence, from the equation (3.8.20)

$$T_{m+1} \cdot \alpha_{r,s} = \frac{\left(\frac{2}{p_{m+1}} - 4 \sin^2 \frac{r\pi}{2N}\right)\left(\frac{2}{p_{m+1}} - 4 \sin^2 \frac{s\pi}{2N}\right)}{\left(\frac{2}{p_{m+1}} + 4 \sin^2 \frac{r\pi}{2N}\right)\left(\frac{2}{p_{m+1}} + 4 \sin^2 \frac{s\pi}{2N}\right)} \alpha_{r,s} \qquad (3.8.22)$$

for $1 \leq r,s \leq N-1$.

and therefore the process is convergent since $\rho(T_{m+1}) < 1$.

The variation of p_{m+1} with m, provides a substantial improvement in the convergence of the Peaceman-Rachford method for solving Laplace's equation in the considered square region.

The full analysis of the method together with the comparison of the rate of convergence of the method with variable and constant iterative parameters is given in Varga, 1962, p.p.209-217.

3.9 HOPSCOTCH METHODS FOR ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS

In a recent paper (1979), Gourlay and McKee have analysed the iterative Hopsoctch algorithm for the solution of the elliptic partial differential equations, in particular the Laplace equation.

Until then it was believed that the iterative hopscotch algorithm was simply symmetric successive over-relaxation for some ordering of the mesh points. The authors show that this belief is not true and they have studied the properties of the hopsoctch method for the solution of the Laplace equation in a separate way.

Here we shall consider their analysis where they prove the equivalence of the hopsoctch and the well-known Du-Fort-Frankel schemes.

We consider the Laplace equation on a unit square with grid points $(i\Delta x, j\Delta y)$ and suppose that the matrix equation $\underline{Au}=\underline{b}$ is obtained from the finite difference replacement of the partial differential equation. We shall assume that A has property A and the ordering chosen is the σ_1-ordering. The two step hopscotch process then can be written as follows:

$$\begin{aligned}(I+pI_1A)\underline{U}_{(m)} &= (I-pI_2A)\underline{U}_{(m-1)} + k_1 \\ (I+pI_2A)\underline{U}_{(m+1)} &= (I-pI_1A)\underline{U}_{(m)} + k_2\end{aligned} \tag{3.9.1}$$

where $\underline{k}_1$ and $\underline{k}_2$ are suitably defined.

To obtain the SSOR we first consider that, the σ_1-ordering means "relaxing" initially all mesh points when i+j is even and then all mesh points where i+j is odd. Therefore, the Jacobi matrix takes the form $\begin{bmatrix} 0 & R^T \\ R & 0 \end{bmatrix}$. Thus, the SSOR method with σ_1-ordering becomes:

$$\begin{aligned}\underline{U}_{(m)} &= \{I-\omega\begin{pmatrix}0&0\\R&0\end{pmatrix}\}^{-1}\{(1-\omega I)+\omega\begin{pmatrix}0&R^T\\0&0\end{pmatrix}\}\,\underline{U}_{(m-1)} + \underline{d}_1 \\ \underline{U}_{(m+1)} &= \{I-\begin{pmatrix}0&R^T\\0&0\end{pmatrix}\}^{-1}\{(1-\omega)I+\begin{pmatrix}0&0\\R&0\end{pmatrix}\}\underline{U}_{(m)} + \underline{d}_2\end{aligned} \tag{3.9.2}$$

on the other hand (3.9.1) with σ_1-ordering now becomes:

$$\underline{U}_{(m)} = \{I-p\begin{pmatrix}0 & 0\\ 0 & I\end{pmatrix}A\}^{-1}\{I+p\begin{pmatrix}I & 0\\ 0 & 0\end{pmatrix}A\}\underline{U}_{(m-1)} + \underline{b}_1$$

$$\underline{U}_{(m+1)} = \{I-p\begin{pmatrix}I & 0\\ 0 & 0\end{pmatrix}A\}^{-1}\{I+p\begin{pmatrix}0 & 0\\ 0 & I\end{pmatrix}A\}\ \underline{U}_{(m)} + \underline{b}_2\ , \qquad (3.9.3)$$

where I_1 and I_2 in (3.9.1) now are equivalent to

$$I_1 = \begin{pmatrix}0\\ I\end{pmatrix} \quad \text{and} \quad I_2 = \begin{pmatrix}I\\ 0\end{pmatrix}\ .$$

The equation (3.9.3) might be written in the following form when we apply the Jacobi method namely,

$$\underline{U}_{(m)} = \{\begin{pmatrix}I & 0\\ 0 & (1-p)I\end{pmatrix}-p\begin{pmatrix}0 & 0\\ R & 0\end{pmatrix}\}^{-1}\{\begin{pmatrix}(1+p)I & 0\\ 0 & I\end{pmatrix}+p\begin{pmatrix}0 & R^T\\ 0 & 0\end{pmatrix}\}\underline{U}_{(m-1)} + \underline{b}_1 \qquad (3.9.4a)$$

$$\underline{U}_{(m+1)} = \{\begin{pmatrix}(1-p)I & 0\\ 0 & I\end{pmatrix}-p\begin{pmatrix}0 & R^T\\ 0 & 0\end{pmatrix}\}^{-1}\{\begin{pmatrix}I & 0\\ 0 & (1+p)I\end{pmatrix}+p\begin{pmatrix}0 & 0\\ R & 0\end{pmatrix}\}\underline{U}_{(m)} + \underline{b}_2 \qquad (3.9.4b)$$

Let

$$D = \begin{pmatrix}I & 0\\ 0 & \frac{1}{1-p}I\end{pmatrix} \quad \text{and} \quad D' = \begin{pmatrix}\frac{1}{1-p}I & 0\\ 0 & I\end{pmatrix}$$

and premultiply (3.9.1a) and (3.9.1b) by D and D' respectively to obtain,

$$\underline{U}_{(m)} = \left\{\begin{pmatrix}I & 0\\ 0 & I\end{pmatrix} - \frac{p}{1-p}\begin{pmatrix}0 & 0\\ R & 0\end{pmatrix}\right\}^{-1}\left\{\begin{pmatrix}(1+p)I & 0\\ 0 & \frac{1}{1-p}I\end{pmatrix} + \frac{p}{1-p}\begin{pmatrix}0 & R^T\\ 0 & 0\end{pmatrix}\right\}\underline{U}_{(m-1)} + \underline{b}_1'$$

$$\underline{U}_{(m+1)} = \left\{\begin{pmatrix}I & 0\\ 0 & I\end{pmatrix} - \frac{p}{1-p}\begin{pmatrix}0 & R^T\\ 0 & 0\end{pmatrix}\right\}^{-1}\left\{\begin{pmatrix}\frac{1}{1-p}I & 0\\ 0 & (1+p)I\end{pmatrix} + \frac{p}{1-p}\begin{pmatrix}0 & 0\\ R & 0\end{pmatrix}\right\}\underline{U}_{(m)} + \underline{b}_2' \qquad (3.9.5a,$$

where $\underline{b}_1' = D\underline{b}_1$ and $\underline{b}_2' = D'\underline{b}_2$.

It is now clear that no choice of p as a function of ω can give (3.9.2). We thus see, similar as they are, OEH is not the same as SSOR with σ_1-ordering (Gourlay & McKee, 1979, p.105).

The Equivalence of Hopscotch and Du-Fort-Frankel Scheme

It is well known that the hopscotch scheme is a two-step implementation of the classical Du-Fort-Frankel method (Gourlay, 1977, p.779). Here, we shall give a formal proof which makes clear what the equivalence precisely is.

First we consider the implicit Du-Fort-Frankel scheme for solving the two dimensional heat equation namely:

$$(1+4p)U_{i,j}^{n+1} - 2p(U_{i+1,j}^{n}+U_{i-1,j}^{n}+U_{i,j+1}^{n}+U_{i,j-1}^{n})-(1-4p)U_{i,j}^{n-1} = 0, \tag{3.9.6}$$

The implicit analogous to (3.9.6) is given by,

$$(1+4p)U_{i,j}^{n+1}-p(U_{i,j-1}^{n+1}+U_{i,j+1}^{n+1})-2p(U_{i-1,j}^{n}+U_{i+1,j}^{n})-(1-4p)U_{i,j}^{n-1}-p(U_{i,j-1}^{n-1}+U_{i,j+1}^{n-1})=0 \tag{3.9.7}$$

The scheme (3.9.7) can be shown to have the L.T.E. $O(k^2+h^2+(\frac{k}{h})^2)$ like D.F.F. method.

Now we give the equivalence theorem.

Theorem 3.8

Odd-even hopscotch is equivalent to the Du-Fort-Frankel scheme in the following sense:

i) DFF must be started by using one step of OEM;

ii) DFF must only be employed on alternate grids, i.e. i+j+n even.

Proof

Consider an arbitrary point $(i\Delta x, j\Delta y, \Delta t)$ at which the solution has been calculated by the implicit 5-points formula, that is, by

$$U_{i,j}^{1} = \frac{1}{1+4p}[p(U_{i-1,j}^{1}+U_{i+1,j}^{1}+U_{i,j-1}^{1}+U_{i,j+1}^{1})+U_{i,j}^{0}] \tag{3.9.8}$$

To calculate $U_{i,j}^{2}$ we employ the explicit formula, i.e.,

$$U_{i,j}^{2} = (1-4p)U_{i,j}^{1}+p(U_{i+1,j}^{1}+U_{i-1,j}^{1}+U_{i,j-1}^{1}+U_{i,j+1}^{1}). \tag{3.9.9}$$

Since $U_{i\pm1,j}^{1}$ and $U_{i,j\pm1}^{1}$ can be calculated from (3.9.8), we can eliminate $U_{i,j}^{1}$ to obtain:

$$(1+4p)U_{i,j}^{2}-2p(U_{i-1,j}^{1}+U_{i+1,j}^{1}+U_{i,j+1}^{1}+U_{i,j-1}^{1})-(1-4p)U_{i,j}^{0} = 0 \tag{3.9.10}$$

This is, of course, the DFF scheme and so $U^2_{i,j}$ can be interpreted as having been calculated by DFF started in an odd-even hopscotch fashion.

Now suppose at the point $(i\Delta x, j\Delta y, x\Delta T)$, the implicit formula has been applied to obtain $U^n_{i,j}$ and assume that the values $U^n_{i\pm1,j}$ and $U^n_{i,j\pm1}$ have all been calculated using DFF. Then, we apply OEH and obtain $U^n_{i,j}$ from the implicit scheme and $U^{n+1}_{i,j}$ from the explicit scheme. Therefore, eliminating $U^n_{i,j}$ gives the DFF scheme and so by induction on n we have shown that OEH is equivalent to DFR subject to one provision. However, only at half of the mesh points the solution $U^n_{i,j}$ are calculated by the implicit formula and the other half being calculated by the explicit method. Thus, the equivalence is only true for half the points. The values $U^n_{i,j}$ associated with the points $(i\Delta x, j\Delta y, x\Delta T)$ calculated by the implicit method are infact calculated from the values,

$$U^{n-1}_{i,j}, U^n_{i\pm1,j} \quad \text{and} \quad U^n_{i,j\pm1}$$

which have been previously calculated by DFF. Thus, the values of $U^n_{i,j}$ are locally "filled-in" by the implicit formula (Gourlay & McKee, 1979, p.106).

In the same manner one can show the equivalence of the line hopscotch scheme with the implicit DFF method and prove the following theorem:

Theorem 3.9

Line hopscotch is equivalent to the implicit Du-Fort-Frankel scheme in the following sense:

i) Implicit DFF must be started by using one step of line hopscotch,

ii) Implicit DFF must only be employed on alternate lines, i.e. n+i even.

Proof: (see Gourlay & McKee (1979, p.106).

Optimum Parameter for Convergence

The equivalence of the hopscotch algorithm and the DFF scheme has been used to find the optimum value of the iteration parameter p for the former

method when it is applied to solve the equation (3.9.1), and the following theorem is obtained.

Theorem 3.10

Odd-even hopscotch applied to the model problem (3.9.1) converges for all $p>0$ and has an optimum value of p given by (for simplicity, N is taken to be an even integer)

$$\tilde{p} = \frac{1}{4 \sin \frac{\pi}{N}}$$

Proof

The equivalence of the DFF and OE hopscotch methods indicates that, we need to examine the convergence of DFF which requires $|\mu|<1$ in the equation,

$$(1+4p)\mu^2 - 4pA\mu - (1-4p) = 0 \qquad (3.9.11)$$

where $A=\cos\frac{i\pi}{N} + \sin\frac{j\pi}{N}$, $i,j=1,2,\ldots,N-1$.

It can be shown that

$$|\mu| = \frac{4p|A|+\{16p^2A^2+4(1-16p^2)\}}{2(1+4p)} \quad \text{when } p\leq\tilde{p} = \frac{1}{2}\frac{1}{(1-A^2)^{1/2}}$$

and that

$$|\mu| = \frac{16p^2-1}{(1+4p)^2} \qquad \text{when } p>\tilde{p}.$$

As we observe, $|\mu|\to 1$ as $p\to 0$ or $p\to\infty$.

Also since,

$$(1+4p)|\mu|^2 - 4p|A||\mu| - (1-4p) = 0 ,$$

we can show that for $p<\tilde{p}$

$$\frac{\partial|\mu|}{\partial p} = -\frac{(|\mu|^2-|A||\mu|+1)}{2|\mu|(1+4p)-4p|A|}$$

and hence we can deduce that for $0<p<\tilde{p}$ there is no turning point for and that $\frac{\partial|\mu|}{\partial p}<0$. Similarly we find $\frac{\partial|\mu|}{\partial p}>0$ for $\tilde{p}<p<\infty$. A graph of $|\mu|$ is shown in Figure 3.9.1.

Thus, the method converges for all $p>0$.

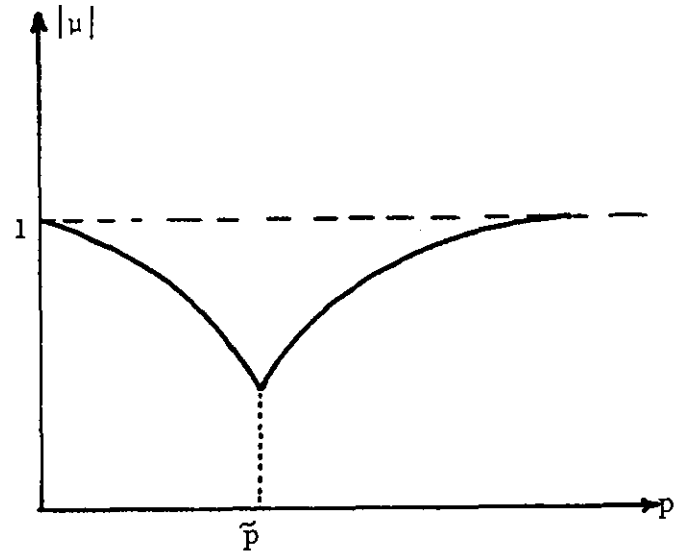


FIGURE 3.9.1

The optimum value of $|\mu|$ for convergence is

$$\min_{p} \max_{i,j} |\mu|$$

As the Figure 3.8 shows the smallest value for $|\mu|$ is obtained at $p=\tilde{p}$. However, to find the $\max_{i,j} |\mu|$ we consider that $|\mu|$ is a function of i and j and $|\mu|$ attains its maximum when p is as large as possible, i.e.

$$\tilde{p} = \frac{1}{4(1-A^2)^{\frac{1}{2}}} = \frac{1}{4(1-(\cos\frac{\pi}{N}+\sin\frac{\pi}{N})^2)^{\frac{1}{2}}} = \frac{1}{4\sin\frac{\pi}{N}} \quad .$$

The same analysis holds for line hopscotch and as given in the previous paper.

Theorem 3.11

Line hopscotch applied to the model problem (3.9.1) converges for all $p>0$ and has an optimum value for p given by

$$\tilde{p} = \frac{1}{4\sqrt{2}\sin\frac{\pi}{2N}} \qquad (3.9.12)$$

Proof: See Gourlay & McKee, 1979, p.108.

Convergence of Classes of Hopscotch Methods

It is shown (Gourlay & McKee 1979, 109) that the convergence of the hopscotch methods can be generalised since they behave similarly. The following lemma shows this similarly.

Lemma 3.3

Let $\sum_N^M$ be the set of N×N diagonal matrices which have precisely M of their diagonal elements equal to 1 and (N-M) of them equal to 0. Let I_1 and I_2 be any two members of the set $\sum_N^M$. Then there exists a permutation matrix $P(PP^T=P^TP=I)$ such that $I_1=P^TI_1P$ and similarly

$$\tilde{I}_2 = I-\tilde{I}_1 = P^TP-P^TI_1P = P^T(I-I_1)P = P^TI_2P \quad .$$

We now consider two hopscotch processed defined by I_1 and $\tilde{I}_1$ as follows:

$$\begin{aligned}(I+pI_2A)\underline{U}_{(m)} &= (I-pI_1A)\underline{U}_{(m-1)}+\underline{k}_1 \\ (I+pI_1A)\underline{U}_{(m+1)} &= (I-pI_2A)\underline{U}_{(m)}+\underline{k}_2 \ .\end{aligned} \tag{3.9.13}$$

and

$$\begin{aligned}(I+p\tilde{I}_2A)\underline{U}_{(m)} &= (I-p\tilde{I}_1A)\underline{U}_{(m-1)}+\tilde{\underline{k}}_1 \\ (I+p\tilde{I}_1A)\underline{U}_{(m+1)} &= (I-p\tilde{I}_2A)\underline{U}_{(m)}+\tilde{\underline{k}}_2 .\end{aligned} \tag{3.9.14}$$

Substituting for I_1 and $\tilde{I}_2$ in (3.9.14) from their respective relationships to I_1 and I_2, results in the following formulae;

$$\begin{aligned}P^T(I+pI_2PAP^T)P\underline{U}_{(m)} &= P^T(I-pI_1PAP^T)P\underline{U}_{(m-1)}+P^T\underline{k}_1P \\ P^T(I+pI_1PAP^T)P\underline{U}_{(m+1)} &= P^T(I-pI_2PAP^T)P\underline{U}_{(m)}+P^T\underline{k}_2P\end{aligned} \tag{3.9.15}$$

Since $P^{-1}=P^T$, (3.9.15) can be written in the form:

$$\begin{aligned}(I+pI_2PAP^T)P\underline{U}_{(m)} &= (I-pI_1PAP^T)P\underline{U}_{(m-1)}+P\underline{k}_1 \\ (I+pI_1PAP^T)P\underline{U}_{(m+1)} &= (I-pI_2PAP^T)P\underline{U}_{(m)}+P\underline{k}_2\end{aligned} \tag{3.9.16}$$

Introducing $\underline{V}_{(m)}=P\underline{U}_{(m)}$, $A'=PAP^T$, $\underline{d}_i=P\underline{k}_i$, i=1,2, (3.9.16) becomes:

$$(I+pI_2A')\underline{V}_{(m)} = (I-pI_1A')\underline{V}_{(m-1)}+\underline{d}_1$$
$$(I+pI_1A')\underline{V}_{(m+1)} = (I-pI_2A')\underline{V}_{(m)}+\underline{d}_2 \qquad (3.9.17)$$

which is of the form (3.9.13) with matrix A' instead of A. Therefore, the following theorem holds.

Theorem 3.12

If I_1 and $\tilde{I}_1 \in \sum_N^M$ define two hopscotch methods for a matrix A then the I_1 hopscotch method for A can also be regarded as an $\tilde{I}_1$ hopscotch method for $A'=PAP^T$, where P is such that $\tilde{I}_1=P^TIP$ (where A' is unitary).

Consequently if (3.9.13) is used to solve iteratively a linear system of equations $A\underline{u}=\underline{b}$ with A positive definite, and if the condition for convergence depends only on the properties of A (for instance on the eigenvalues of A which is the case for odd-even and line hopscotch), then the condition for convergence of (3.9.14) is similar to (3.9.13) when applied to solve the same system of equations.

If the above properties of convergence hold we say A has property Ω (Gourlay & McKee) and we state the next theorem.

Theorem 3.13

If the hopscotch process converges with property Ω for some $I_1 \in \sum_N^M$ then it will converge for all $I_1 \in \sum_N^M$.

It is important to note that, the result on the rate of convergence is different.

In the time-dependent problems, all the hopscotch methods in $\sum_N^M$ are either stable or unstable depending whether A has property Ω or not.

3.10 THE HOPSCOTCH METHODS FOR THE SOLUTION OF PARABOLIC AND ELLIPTIC EQUATIONS WITH MIXED DERIVATIVES

The hopscotch technique has been used by the previous author (1977) to solve

$$\frac{\partial u}{\partial t} = L(u) + g(x,y,t)$$

where

$$L \equiv a(x,y,t)\frac{\partial^2 u}{\partial x^2} + 2b(x,y,t)\frac{\partial^2 u}{\partial x \partial y} + c(x,y,t)\frac{\partial^2 u}{\partial y^2}, \qquad (3.10.1)$$

subject to $a > 0$, $c > 0$ and $ac-b^2 > 0$

in the region $R\times[0\leq t\leq T]$ where R is a closed region of the x,y plane with a continuous boundary ∂R.

The finite difference replacement L_h for (3.10.1) in the case of the O-E hopscotch is as follows:

$$L_h = \frac{1}{h^2}[a\delta_x^2 + c\delta_y^2 + b(\sigma^2+\sigma^4)] \qquad (3.10.2)$$

whilst for the line-hopscotch (3.10.2) becomes:

$$L_h = \frac{1}{h^2}[a\delta_x^2+c\delta_y^2+\tfrac{1}{2}bH_xH_y] \qquad (3.10.3)$$

where the symbols σ and H are defined to be

$$\begin{aligned}
\sigma^1 U^n_{i,j} &= U^n_{i+1,j+1} - U^n_{i,j+1} + U^n_{i,j} - U^n_{i+1,j} \\
\sigma^2 U^n_{i,j} &= U^n_{i,j+1} - U^n_{i-1,j+1} - U^n_{i,j} + U^n_{i-1,j} \\
\sigma^3 U^n_{i,j} &= U^n_{i,j} - U^n_{i-1,j} - U^n_{i,j-1} + U^n_{i-1,j-1} \qquad (3.10.4) \\
\sigma^4 U^n_{i,j} &= U^n_{i+1,j} - U^n_{i,j} - U^n_{i+1,j} + U^n_{i,j-1}
\end{aligned}$$

and

$$H_x U^n_{i,j} = U^n_{i+1,j} - U^n_{i-1,j}, \quad H_y U^n_{i,j} = U^n_{i,j+1} - U^n_{i,j-1}. \qquad (3.10.5)$$

The above formulations have been used for some examples and numerical comparisons made. In Chapter VI we shall consider this formulation in more detail and make a comparison with a new hopscotch strategy.

The hopscotch methods are also applied to solve the elliptic equation $L(u)=-g(x,y)$ where L has the same definition as (3.10.1).

3.11 NON-LINEAR EQUATIONS AND HOPSCOTCH TECHNIQUE

One of the advantages of the hopscotch techniques which compares with the other implicit methods is the simplicity of the former technique for non-linear equations.

Since the hopscotch algorithms are overall an explicit model, the problem of non-linearity can be treated more easily. This is more apparent in the O-E hopscotch case. While the application of the implicit methods leads to a large set of non-linear equations which must be solved by a suitable non-linear iterative method, the use of O-F hopscotch provides a single non-linear equation at every grid point which can be solved much more efficiently.

As an example Danaee (1978) has applied the O-E hopscotch to solve a real life (chemical-reaction) problem which consists of a system of five non-linear parabolic equations as follows:

$$\rho(T).CV(T)\frac{\partial T}{\partial t} = \rho(T).F_1 + \frac{\partial}{\partial z}\{D_1(T)\frac{\partial T}{\partial z}\}$$
$$\rho(T).\frac{\partial y_i}{\partial t} = \rho(T).F_i + \frac{\partial}{\partial z}\{D_i(T)\rho(T)\frac{\partial y_i}{\partial z}\} \quad i=2,3,4,5 \tag{3.11.1}$$

where the functions F_i and the coefficients $\rho(T)$, $CV(T)$ and $D_i(T)$ are some strong and rather complicated non-linear functions (Danaee, A., 1978, p.37). The solution of (3.11.1) is required for $t\in[0,8]$.

The standard explicit scheme together with the Crank-Nicolson methods have been applied to solve (3.11.1) and a comparison has been made with the solution obtained from the O-E hopscotch scheme.

As mentioned previously, for small time-steps (p small) the hopscotch methods gives a good approximation. Since in (3.11.1) one is bound to apply a very small time increment to reach the required accuracy (and to avoid the overflow which appears for large time steps because of the structure of the functions involved), then the hopscotch technique appears to be more efficient and economic.

The result presented in the aforementioned reference shows that the computation time for the Crank-Nicolson method with $DT=10^{-3}$ is ≈3100 sec while the comparable time for the hopscotch is ≈1740 sec. (by the CDC-7600). Although the hopscotch program was not designed to be extremely efficient the results suggest that the hopscotch method is nearly twice as fast as the Crank-Nicolson method in this case.

CHAPTER IV

HOPSCOTCH PROCEDURE FOR A FOURTH-ORDER PARABOLIC PARTIAL DIFFERENTIAL EQUATION

4.1 INTRODUCTION

The governing equation of the vibration of a thin beam which is clamped at its ends and set into vibration, is well-known to be a fourth-order parabolic equation, namely,

$$\frac{\partial^2 y}{\partial t^2} + \frac{\partial^4 y}{\partial x^4} = 0 \qquad (4.1.1)$$

where $y=y(x,t)$ denotes the displacement from the equilibrium position at a distance x along the beam from one end at time t.

Various finite-difference replacements have been applied to solve the equation (4.1.1) which can be seen in Evans, D.J. (1965), p.280, Richtmyer & Morton (1967), p.271 and Fairweather, G. & Gourlay, A.R., (1966), p.1.

The hopscotch technique was also applied by Orley, D.G. & McKee, S., (1973) p.335 for equation (4.1.1), where several computational schemes were discussed. However, the hopscotch schemes presented by the authors are proved to be conditionally stable with the stability range no better than that of the usual explicit scheme.

In this chapter we will show that, the hopscotch technique can be applied to the split form of (4.1.1) given in the book by Richtmyer & Morton, which results in an unconditional stable procedure.

4.2 THE HOPSCOTCH FORMULATION

Consider the equation (4.1.1) subject to the initial conditions:

$$y(0,t) = g_0(x) \ ,$$
$$\text{and} \qquad \frac{\partial y}{\partial t}(x,0) = g_1(x) \ , \tag{4.2.1}$$

for $0 \leq x \leq 1$, and the boundary conditions

$$y(0,t) = f_0(t), \quad y(1,t) = f_1(t)$$
$$\frac{\partial^2 y(0,t)}{\partial x^2} = p_0(t), \quad \frac{\partial^2 y(1,t)}{\partial x^2} = p_1(t) \ , \quad t \geq 0. \tag{4.2.2}$$

Following Richtmyer (1967), we introduce two new variables u and v such that:

$$u = \frac{\partial y}{\partial t} \ , \qquad v = \frac{\partial^2 y}{\partial x^2} \ .$$

The equation (4.1.1) can now be rewritten as a system of two second order parabolic equations:

$$\frac{\partial u}{\partial t} = - \frac{\partial^2 v}{\partial x^2}$$
$$\frac{\partial v}{\partial t} = \frac{\partial^2 u}{\partial x^2} \tag{4.2.3}$$

or as a second order system

$$\frac{\partial w}{\partial t} = A \frac{\partial^2 w}{\partial x^2} \tag{4.2.4}$$

where

$$w = \begin{bmatrix} u \\ v \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \ .$$

The system (4.2.3) is solved by different techniques including the Du Fort-Frankel (by Evans, D.J.) and the explicit-implicit methods (by Fairweather & Gourlay).

The hopscotch formulation for (4.2.3) is as follows:

$$W_i^{m+1} - k\theta_i^{m+1} A\delta_x^2 W_i^{m+1} = W_i^m + k\theta_i^m A\delta_x^2 W_i^m \ , \qquad 1 \leq i \leq N-1, \quad m > 0 \tag{4.2.5}$$

$$\theta_i^m = \begin{cases} 1 & i+m \text{ even} \\ 0 & i+m \text{ odd} \end{cases}$$

where W_i^m is denoted as the approximate solution to the equation (4.2.4) at the grid point p=(ih,mk), N is the number of mesh points in the x-direction and δ_x^2 is as given previously.

4.3 TRUNCATION ERROR OF THE DISCRETISED SOLUTION

To obtain the L.T.E., we consider the $(m+1)^{st}$ and $(m+2)^{nd}$ time levels and find the principal part of the L.T.E. for an explicit and an implicit point for both variables u and v (Danaee, 1978) as given in Figure 3.4. In the following table, we demonstrate these error terms.

	Principal part of L.E.T. for u-net	Principal part of L.T.E. for v-net
Explicit point	$2k^2[(1+4p^2)\frac{\partial^2 u}{\partial t^2}+\frac{\partial^4 u}{\partial x^4}]+$ $\frac{kh^2}{6}(1+28p^2)\frac{\partial^4 v}{\partial x^4}$	$2k^2[(1+4p^2)\frac{\partial^2 v}{\partial t^2}+\frac{\partial^4 u}{\partial x^4}]$ - $\frac{kh^2}{6}(1+28p^2)\frac{\partial^4 u}{\partial x^4}$
Implicit point	$2k^2[(1+4p^2)\frac{\partial^2 u}{\partial t^2}+(1+12p^2)\frac{\partial^4 u}{\partial x^4}]$ $+\frac{kh^2}{6}[(1+4p^2)^2+19^2p^4]\frac{\partial^4 v}{\partial x^4}$	$2k^2[(1+4p^2)\frac{\partial^2 v}{\partial t^2}+(1+12p^2)\frac{\partial^4 v}{\partial x^4}]$ $-\frac{kh^2}{6}[(1+4p^2)^2+19p^2p^4]\frac{\partial^4 u}{\partial x^4}$

TABLE (4.1)

where $p=k/h^2$.

As can be seen from Table (4.1) the local truncation error of the method is $O(k^2+kh^2)$ provided p is held constant. If p is not constant, the L.T.E. becomes $O(k^2+kh^2+k^4/h^4)$ which is of the Du Fort-Frankel type of inconsistency with the heat equation.

4.4 THE NUMERICAL STABILITY OF THE METHOD

The stability of the method could be discussed by two different analyses.

1. By finding the amplification matrix for the system (4.2.4) at one point (Richtmyer method), or
2. By considering two advanced time levels and finding the matrix form of the system (Gourlay, 1970).

Here we concentrate on the second method and also give the amplification matrix for the first method and show the stability of the scheme.

We consider the equations (4.2.3) and rewrite them in the form of (4.2.5) as in

$$U_i^{m+1}+p\theta_i^{m+1}\delta_x^2 V_i^{m+1} = U_i^m-p\theta_i^m\delta_x^2 V_i^m$$

and (4.4.1)

$$V_i^{m+1}-p\theta_i^{m+1}\delta_x^2 U_i^{m+1} = V_i^m+p\theta_i^m\delta_x^2 U_i^m \quad \text{for } i=1,2,\ldots,N-1.$$

These equations can be expressed in the matrix form:

$$\underline{U}^{m+1}+p\theta_1 T\underline{V}^{m+1} = \underline{U}^m-p\theta_0 T\underline{V}^m$$

and (4.4.2)

$$\underline{V}^{m+1}-p\theta_1 T\underline{U}^{m+1} = \underline{V}^m+p\theta_0 T\underline{U}^m$$

where the appropriate boundary conditions are considered and where

$$T = \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ & 0 & & 1 & -2 \end{bmatrix}, \quad \theta_r = \begin{bmatrix} \theta_1^{m+r} & & & \\ & \theta_2^{m+r} & & 0 \\ & & \ddots & \\ 0 & & & \theta_{N-1}^{m+r} \end{bmatrix}$$

$r=0,1,2.$

are $(N-1)\times(N-1)$ matrices.

If we let $\underline{W}=\begin{bmatrix}\underline{U}\\ \underline{V}\end{bmatrix}$, then (4.4.2) can be further simplified as follows,

$$\begin{bmatrix} I & p\theta_1 T \\ -p\theta_1 T & I \end{bmatrix} \underline{W}^{m+1} = \begin{bmatrix} I & -p\theta_0 T \\ p\theta_0 T & I \end{bmatrix} \underline{W}^m \qquad (4.4.3)$$

where I is the unitary matrix of order (N-1).

Therefore the two-step hopscotch process can be easily verified to be:

$$\underline{W}^{2m+2} = \begin{bmatrix} I & p\theta_2 T \\ -p\theta_2 T & I \end{bmatrix}^{-1} \begin{bmatrix} I & -p\theta_1 T \\ p\theta_1 T & I \end{bmatrix} \begin{bmatrix} I & p\theta_1 T \\ -p\theta_1 T & I \end{bmatrix}^{-1} \begin{bmatrix} I & -p\theta_0 T \\ p\theta_0 T & I \end{bmatrix} \underline{W}^{2n} \tag{4.4.4}$$

or in compact form:

$$\underline{W}^{2m+2} = M\underline{W}^{2m} \quad . \tag{4.4.5}$$

Since by definition $\theta_2=\theta_0$ and $\theta_i\theta_j=0$, $i\neq j$ and also since the inverse of the matrices in (4.4.4) can be easily obtained then in (4.4.5) M becomes:

$$M = \frac{1}{(1+p^2T^2)} \begin{bmatrix} I-p^2T^2 & -2pT \\ 2pT & I-p^2T^2 \end{bmatrix}$$

Given that the eigenvalues of T are τ_s, s=1,2,...,N-1, then the eigenvalue λ_s of M can be shown to be given by

$$\lambda_s = \frac{1-p^2\tau_s^2}{1+p^2\tau_s^2} \pm 2\sqrt{-1} \quad \frac{p\tau_s}{1+p^2\tau_s^2} \quad , \; s=1,2,\ldots,N-1,$$

the modulus of which are equal to unity. Thus, it follows immediately that the method is unconditionally stable.

Alternatively by considering a typical Fourier term for (4.2.4) we obtain

$$\underline{W}_i^m = \underline{W}_0^m \; e^{\sqrt{-1}\,.\beta mh}$$

and

$$\underline{W}_i^{m+1} = \begin{bmatrix} 1 & -4p\sin^2\frac{\beta h}{2} \\ 4p\sin^2\frac{\beta h}{2} & 1 \end{bmatrix} \underline{W}_i^m$$

Hence, if we let $d=-4p\sin^2\frac{\beta h}{2}$, for two time levels we obtain the following result:

$$\underline{W}_i^{2m+2} = G\,\underline{W}_i^{2m}$$

where G the amplification matrix has the form:

$$G = \begin{bmatrix} \frac{1-d^2}{1+d^2} & \frac{2d}{1+d^2} \\ \frac{-2d}{1+d^2} & \frac{1-d^2}{1+d^2} \end{bmatrix}$$

with eigenvalues $\mu_{1,2} = \frac{1-d^2}{1+d^2} \pm \frac{2\sqrt{-1}.d}{1+d^2}$ which possess unity modulus. Consequently the method is unconditionally stable.

4.5 EXTENSION TO THE $2n^{TH}$ ORDER PARABOLIC EQUATION

In this section we consider the general form of (4.1.1) i.e.,

$$\frac{\partial^n y}{\partial t^n} = (-1)^{n+1} \frac{\partial^{2n} y}{\partial x^{2n}} \quad . \tag{4.5.1}$$

Let us suppose the appropriate initial and boundary values for (4.5.1) are given and moreover, $y \in C^n_{2n}$ ($2n^{th}$ differentiable with respect to x and n^{th} differentiable with respect to t).

We introduce a sequence of new variables as follows:-

$$\begin{aligned} &u_0 = y, \quad u_1 = \frac{\partial^2 y}{\partial t}, \ldots, \quad u_r = \frac{\partial^r y}{\partial t^r}, \ldots, u_{n-1} = \frac{\partial^{n-1} y}{\partial t^{n-1}} \quad , \\ \text{and} \quad &v_0 = y, \quad v_1 = \frac{\partial^2 y}{\partial x^2}, \ldots, \quad v_r = \frac{\partial^{2r} y}{\partial x^{2r}}, \ldots, v_{n-1} = \frac{\partial^{2n-2} y}{\partial x^{2n-2}} \quad . \end{aligned} \tag{4.5.2}$$

Hence, the equation (4.5.1) now becomes:

$$\frac{\partial u_{n-1}}{\partial t} = (-1)^{n+1} \frac{\partial^2 v_{n-1}}{\partial x^2} \tag{4.5.3}$$

If we suppose that the following relation holds for the derivatives of y up to order n-1, i.e.

$$\frac{\partial v_r}{\partial t} = \frac{\partial^2 u_r}{\partial^2 x} \quad , \qquad \text{for } r=1,2,\ldots,n-1$$

then we obtain a new equation similar to the system (4.2.3) namely,

$$\begin{aligned} \frac{\partial u_{n-1}}{\partial t} &= (-1)^{n+1} \frac{\partial^2 v_{n-1}}{\partial x^2} \\ \frac{\partial v_{n-1}}{\partial t} &= \frac{\partial^2 u_{n-1}}{\partial x^2} \quad . \end{aligned} \tag{4.5.4}$$

Applying the hopscotch algorithm as in the previous section, we obtain the amplification matrix:

$$G = \begin{bmatrix} \frac{1-d^2}{1+d^2} & (-1)^n \frac{2d}{1+d^2} \\ \frac{-2d}{1+d^2} & \frac{1-d^2}{1+d^2} \end{bmatrix} \tag{4.5.5}$$

where d is defined earlier.

Therefore, the stability of the method is guaranteed when n is an even integer and for n odd we have instability.

Although, the assumption (4.5.3) may not be available in reality, nevertheless it shows that this splitting procedure can not give a suitable algorithm for equations of odd order.

The Two-Space Variable Case

The partial differential equation

$$\frac{\partial^2 y}{\partial t^2} + L^4 u = 0 \quad \text{where} \quad L^4 = \frac{\partial^4}{\partial x^4} + \frac{\partial^4}{\partial y^4} \tag{4.5.6}$$

arises in the study of transverse vibrations of a uniform plate and can assume the form

$$\frac{\partial w}{\partial t} = A\, L^2 w \tag{4.5.7}$$

where we have introduced the variables $u = \frac{\partial y}{\partial t}$ and $v = L^2 y$ where $w=\begin{bmatrix} u \\ v \end{bmatrix}$ and $A=\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Equation (4.5.7) can be solved by the hopscotch method and as this method is invariant with respect to the number of space dimensions, the analysis of stability would be exactly the same, except that in the amplification matrix G given in (4.5.5), $d=-4(\sin\frac{\alpha h}{2} + \sin^2 \frac{\beta h}{2})$ where $\Delta x=\Delta y$ and $p=\Delta T/\Delta x^2$.

The block hopscotch techniques can also be applied for higher-dimensional problems which have no analogies in the one-dimensional case.

4.6 NUMERICAL EXPERIMENT

In order to provide a comparison with previous methods, Fairweather and Gourlay, the vibration beam problem is solved together with the initial conditions,

$$y(x,0) = \frac{x}{12}(2x^2-x^3-1)$$

$$\frac{\partial y}{\partial t}(x,0) = 0, \qquad 0\leq x\leq 1$$

and the boundary conditions

$$y(0,t) = y(1,t) = 0$$

$$\frac{\partial^2}{\partial x^2} y(0,t) = \frac{\partial^2}{\partial x^2} y(1,t) = 0 \qquad t>0.$$

The same increment is chosen as in Fairweather & Gourlay, i.e. h=0.05 and k=0.00125 (for which p=1/2. In addition, the problem was tested for k=0.005 (i.e. p=2) to demonstrate the experimental stability and investigate the accuracy of the method.

The solution of (4.1.1) could be found by any suitable method (e.g. Numerov method) to solve the second order O.D.E. y''=f(x,y)). We have applied the simple Taylor expansion to find the solution y and the formula is given by

$$y_i^{m+1} = y_i^m + \frac{\Delta T}{2}(U_i^{m+1}+U_i^m) + O(\Delta T^3) \qquad (4.6.1)$$

since the terms $\frac{\partial y}{\partial t}$ and $\frac{\partial^2 y}{\partial t^2}$ can be substituted for u and $\frac{\partial u}{\partial t}$ respectively.

The theoretical solution of the problem with the given initial and boundary conditions is given by

$$y(x,t) = \sum_{s=1}^{\infty} d_s \sin(2s+1)\pi x.\cos(2s+1)^2\pi^2 t. \qquad (4.6.2)$$

where $d_s = -\frac{8}{(2s+1)^5\pi^5}$.

In Table 4.2 the difference between the theoretical solution y given by (6.2) and the computed solution obtained by the hopscotch method are presented. These are compared with results from other methods for t=0.02. It is worthy to note that, although we have not applied the same integration

procedure, these results show that using the simple Taylor expansion gives reasonable results. However since the hopscotch technique itself has a L.T.E. no better than $O(\Delta T^2)$, there is no need to apply a more accurate procedure for $y''=f(x,y)$. (In the case of higher order and n-nets, the same integration procedure can be applied as the terms of the Taylor series are computed i.e. u_i and v_i).

Similar results for the bending moment $v=\frac{\partial^2 y}{\partial x^2}$ are quoted in Table 4.3. In Tables 4.4 and 4.5, some results for different p in a large time domain are illustrated.

Method \ x	0.10	0.20	0.30	0.40	0.50
Exact solun.y	-0.00790415	-0.01504833	-0.02072927	-0.02439234	-0.02565928
Error of Evans method	0.00000844	0.00001416	0.00001740	0.00000140	-0.00001195
Error of Richtmyer	0.00022374	0.00036712	0.00040341	0.00036461	0.00033531
Error of semi explict	-0.00003006	-0.00006193	-0.00006690	-0.00005102	0.00001335
Error of H.O.C.M.	0.00000014	0.00000029	0.00000056	0.00000034	-0.00000017
Error of Hopscotch	-0.00000250	0.00000390	0.00001370	0.00000260	-0.00000980

TABLE 4.2

$\Delta x=0.05$, $\Delta T=0.00125$ $T=0.02$ $(p=1/2)$

Method \ x	0.10	0.20	0.30	0.40	0.50
Exact solun. y"	0.07626211	0.14770388	0.20316761	0.24183975	0.25569973
Error of Evans method	-0.00049570	0.00004763	-0.00237926	0.00043350	0.00312161
Error of Richtmyer	-0.00876844	-0.01113270	-0.00810363	0.00147851	0.00682881
Error of semi explicit	0.00033110	0.00272563	0.00169444	0.00250333	0.00255541
Error of H.O.C.M.	-0.00001717	-0.00008638	-0.00003195	0.00033881	-0.00023174
Error of Hopscotch	-0.00049570	0.00004770	-0.00237930	0.00043360	0.00312160

TABLE 4.3

Method \ x	0.10	0.20	0.30	0.40	0.50
Exact solun. y	-0.066497	-0.131790	-0.186191	-0.224557	-0.240327
Error of Evans Method	-0.003198	-0.002728	0.009803	0.012459	0.014032
Error of Richtmyer	0.002730	0.009479	0.017356	0.022981	0.022356
Error of semi explicit	-0.002736	0.005927	-0.004481	-0.002316	0.006511
Error of H.O.C.M.	-0.002586	-0.001907	0.000717	0.002196	-0.000665
Error of Hopscotch	-0.003188	-0.002727	0.009803	0.012459	0.014032

TABLE 4.4

$\Delta x=0.05$ $\Delta T=0.005$ (p=2) T=1.0

Method \ x	0.10	0.20	0.30	0.40	0.50
Exact solun. y''	0.007235	0.013805	0.019068	0.022483	0.023671
Error of Evans method	-0.000234	-0.000595	-0.000776	-0.000953	-0.001018
Error of Richtmyer	-0.000637	-0.001234	-0.001752	-0.002089	-0.002203
Error of semi explicit	-0.000198	-0.000412	-0.000673	-0.000980	-0.001304
Error of H.O.C.M.	0.000015	0.000006	-0.000018	-0.000034	-0.000035
Error of Hopscotch	-0.001741	-0.003282	-0.004554	-0.005369	-0.005644

TABLE 4.5

4.7 CONCLUSIVE REMARKS

Since it is well-known that the application of the method leads to a fast computational technique for the solution of P.D.E.'s, then we can conclude that for more realistic problems (i.e. non-linear equations, variable coefficients, etc.) and higher degree parabolic equations the application of the stable hopscotch methods presented in this chapter can lead to fruitful gains in computational efficiency.

CHAPTER V

THE APPLICATION OF SPLITTING METHODS TO THE NUMERICAL SOLUTION OF PARABOLIC EQUATIONS OF HIGHER ORDER

5.1 INTRODUCTION

In this chapter as a whole, we consider the splitting methods considered previously to solve a higher order parabolic equation.

To begin with, we consider the equation

$$\left(\frac{\partial}{\partial t} - \frac{\partial^2}{\partial x^2}\right)^2 u \equiv \frac{\partial^2 u}{\partial t^2} - 2\frac{\partial^3 u}{\partial t \partial^2 x} + \frac{\partial^4 u}{\partial x^4} = 0 \qquad (5.1.1)$$

subject to the initial conditions

$$\lim_{t\to 0} u = \begin{cases} f_1(x) & x\in[0,1] \\ 0 & x\notin[0,1], \end{cases} \qquad (5.1.2)$$

$$\lim_{t\to 0} \left(\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2}\right) = \begin{cases} f_2(x) & x\in[0,1] \\ 0 & x\notin[0,1] \end{cases} \qquad (5.1.3)$$

The equation (5.1.1)-(5.1.3) represents an initial value problem of *biparabolic* type.

If the initial function $f_1(x)$ and $f_2(x)$ are continuous in the region [0,1] the exact solution of (5.1.1)-(5.1.3) is given by Saul,yev (1964, p.184) as,

$$u(x,t) = \frac{1}{2\sqrt{\pi t}} \int_0^1 [f_1(\zeta) - t f_2(\zeta)] e^{-\frac{(x-\zeta)^2}{4t}} \, d\zeta. \qquad (5.1.4)$$

The standard simple implicit finite difference equation to solve the above problem is as follows:

$$\frac{\delta_t^2 U_i^j}{k^2} + \frac{\delta_x^4 U_i^j}{h^4} - \frac{\delta_x^2 U_i^{j+1} - \delta_x^2 U_i^{j-1}}{kh^2} = 0 , \qquad (5.1.5)$$

where k,h are the time and space increments respectively and δ is the central difference operator defined earlier.

The L.T.E. of (5.1.5) can be easily verified using the Taylor expansion which indicates that the equation (5.1.5) is of order $O(k^2+h^2)$ with the principal part as,

$$\frac{k^2}{3}\frac{\partial^3}{\partial t^3}\left[\frac{\partial^2 u}{\partial x^2} - \frac{1}{4}\frac{\partial^2 u}{\partial t^2}\right] - \frac{h^2}{6}\frac{\partial^6 u}{\partial x^6} . \qquad (5.1.6)$$

The stability of this scheme can be analysed in the same manner as given by N. Lowan (1957, p.17,63). The formula (5.1.5) can be expanded for the mesh points and rewritten in matrix form, namely,

$$-pU_{i-1}^{j+1}+(1+2p)U_i^{j+1}-pU_{i+1}^{j+1} = 2U_i^j-U_i^{j-1}-p^2(U_{i+2}^j-4U_{i+1}^j+6U_i^j-4U_{i-1}^j+U_{i-2}^j)$$
$$-p(U_{i+1}^{j-1}-2U_i^{j-1}+U_{i-1}^{j-1}) \qquad (5.1.7)$$

where $p=k/h^2$, $1\leq i\leq N-1$, $1\leq j\leq M$ and N is the number of mesh points of [0,1].

A compact form of (5.1.7) is as follows:

$$(I+pT)\underline{U}^{j+1} = (2I-p^2T^2)\underline{U}^j-(I-pT)\underline{U}^{j-1} \qquad (5.1.8)$$

where I is the unit matrix and

$$T = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & -1 \\ & 0 & & -1 & 2 \end{bmatrix} \quad \text{and} \quad \underline{U}^j = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_{N-1} \end{bmatrix}^j .$$

Since (I+pT) is non-singular then (5.1.8) can be expressed as:

$$\underline{U}^{j+1} = M\underline{U}^j+N\underline{U}^{j-1} \qquad (5.1.9)$$

where $M=(I+PT)^{-1}(2I-p^2T^2)$ and $N=(I+pT)^{-1}(I-pT)$.

According to Theorem (1.7) M and N have the same set of eigenvectors and if the eigenvalues of T,M and N are λ_s,μ_s and ν_s respectively. We have,

$$\mu_s = \frac{2-p^2\lambda_s^2}{1+p\lambda_s} \quad \text{and} \quad \nu_s = \frac{1-p\lambda_s}{1+p\lambda_2} \quad . \qquad (5.1.10)$$

Let us suppose $\underline{V}^j=\underline{U}^{j-1}$; therefore the three level formula (5.1.9) can be inverted to the following two level system:

$$\begin{bmatrix} \underline{U}^{j+1} \\ \underline{V}^{j+1} \end{bmatrix} = \begin{bmatrix} M & -N \\ I & 0 \end{bmatrix} \begin{bmatrix} \underline{U}^j \\ \underline{V}^j \end{bmatrix} \qquad (5.1.11)$$

or

$$\underline{W}^{j+1} = A\underline{W}^j \quad , \qquad (5.1.12)$$

where A is the amplification matrix. For the stability of (5.1.7) we require the spectral radius of A to be smaller than unity in modulus which can be obtained from the following quadratic formula

$$x^2 - \mu_s x + \nu_s = 0 \qquad (5.1.13)$$

where x are the eigenvalues of A. As the eigenvalues of T are known it can be shown that the roots of (5.1.13) are in the range of stability only if $p \leq 1/2$ which is no better than the usual explicit scheme.

However, by introducing a new variable ϕ, we can decompose the equation (5.1.1) into two second order parabolic equations, to give the following system.

Let

$$\phi = \frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} \quad ; \qquad (5.1.14)$$

then (5.1.1) becomes:

$$\begin{cases} \dfrac{\partial u}{\partial t} - \dfrac{\partial^2 u}{\partial x^2} = \phi & \text{(a)} \\ \dfrac{\partial \phi}{\partial t} - \dfrac{\partial^2 \phi}{\partial x^2} = 0 & \text{(b)} \end{cases} \qquad (5.1.15)$$

Let us denote $\underline{w} = \begin{Bmatrix} u \\ \phi \end{Bmatrix}$ and $\underline{g} = \begin{Bmatrix} \phi \\ 0 \end{Bmatrix}$, then the system (5.1.1)-(5.1.3) becomes,

$$\frac{\partial \underline{w}}{\partial t} = \frac{\partial^2 \underline{w}}{\partial x^2} + \underline{g} \qquad (5.1.16)$$

subject to the initial condition

$$\lim_{t \to 0} \underline{w} = \begin{cases} \begin{bmatrix} f_1(x) \\ f_2(x) \end{bmatrix} & x \in [0,1] \\ \begin{bmatrix} 0 \\ 0 \end{bmatrix} & \text{otherwise.} \end{cases} \qquad (5.1.17)$$

The system (5.1.16) subject to the given initial condition can now be solved by a suitable numerical method on the u and ϕ nets, provided some appropriate boundary conditions are given. Thus, the solution of (5.1.1) can be obtained by solving (5.1.15b) in the first step and (5.1.15a) in the second step. Therefore, the stability of the chosen numerical method can be

easily analysed by looking at each of the equations (5.1.15a)-(5.1.15b) separately which are of the simple second order or heat equation form with some constant right-hand side vector.

5.2 NUMERICAL EXAMPLE

Consider the equation (5.1.1)-(5.1.3) and let $f_1(x)=x(1-x)$, hence $f_2(x)=2$, with the exact solution,

$$u(x,t) = \frac{1}{2\sqrt{\pi t}} \int_0^1 [\zeta(1-\zeta)-2t]\, e^{-\frac{(x-\zeta)^2}{4t}}\, d\zeta \qquad (5.2.1)$$

From equation (5.1.15a) we obtain

$$\phi(x,t) = -\frac{1}{\sqrt{\pi t}} \int_0^1 e^{\frac{(x-\zeta)^2}{4t}}\, d\zeta \qquad (5.2.2)$$

The equations (5.2.1)-(5.2.2) can be simplified by changing the variable $y=\frac{x-\zeta}{2\sqrt{t}}$ and applying integration by parts, which results in,

$$u(x,t) = \frac{(1-2x)\sqrt{t}}{\sqrt{\pi}}\, e^{-y^2}\Big|_a^b + \frac{2t}{\sqrt{\pi}}\, y\, e^{-y^2}\Big|_a^b + \frac{[x(1-x)-4t]}{\sqrt{\pi}} \int_a^b e^{-y^2} dy, \qquad (5.2.3)$$

and

$$\phi(x,t) = -\frac{2}{\sqrt{\pi}} \int_a^b e^{-y^2}\, dy, \qquad (5.2.4)$$

where $a = \frac{x-1}{2\sqrt{t}}$ and $b = \frac{x}{2\sqrt{t}}$.

To solve the integral involved in the above equations, a recursive procedure using the Trapezium rule is applied and the accuracy of the summation is checked against a tolerance level of 10^{-6} and the result is used as the exact values of the solution for comparison with the numerical method.

The boundary values for the numerical method might be found from (5.2.1) by putting x=0 and x=1.

The hopscotch procedure is applied to evaluate the numerical solution of (5.1.1)-(5.1.2) on two nets with the given initial condition. The numerical and exact solution are displayed in the following tables.

In the following tables the 1st line corresponds to the analytical solution and the 2nd line shows the hopscotch solution obtained.

$\Delta x=1/20$, P=1, T=1

t	x=0.1	x=0.2	x=0.3	x=0.4	x=0.5
0.2	-1.11233&-1 -6.63018&-2	-1.18834&-1 -3.33454&-2	-1.24562&-1 -6.87645&-3	-1.28124&-1 1.02379&-2	-1.29333&-1 1.61545&-2
0.4	-2.44245&-1 -2.32284&-1	-2.53942&-1 -2.31177&-1	-2.61102&-1 -2.29757&-1	-2.65494&-1 -2.28637&-1	-2.66976&-1 -2.28218&-1
0.6	-3.41045&-1 -3.38691&-1	-3.50403&-1 -3.45920&-1	-3.57243&-1 -3.51067&-1	-3.61411&-1 -3.54147&-1	-3.62812&-1 -3.55171&-1
0.8	-4.19648&-1 -4.19255&-1	-4.28428&-1 -4.27675&-1	-4.34811&-1 -4.33771&-1	-4.3867&-1 -4.37463&-1	-4.39986&-1 -4.38697&-1
1	-4.87157&-1 -4.87109&-1	-4.95389&-1 -4.95294&-1	-5.01353&-1 -5.01220&-1	-5.04966&-1 -5.04808&-1	-5.06176&-1 -5.06009&-1

TABLE 1

$\Delta x=1/20$, P=2, T=1

t	x=0.05	x=0.15	x=0.25	x=0.35	x=0.45
0.1	-1.11233&-1 -6.58884&-2	-1.18834&-1 -3.25072&-2	-1.24562&-1 -5.66928&-3	-1.28124&-1 1.16926&-2	-1.29333&-1 1.76959&-2
0.3	-2.44245&-1 -2.34046&-1	-2.53942&-1 -2.34503&-1	-2.61102&-1 -2.34306&-1	-2.65494&-1 -2.33966&-1	-2.66976&-1 -2.33814&-1
0.5	-3.41045&-1 -3.39449&-1	-3.50403&-1 -3.47348&-1	-3.57243&-1 -3.53017&-1	-3.61411&-1 -3.56429&-1	-3.62812&-1 -3.57568&-1
0.7	-4.19648&-1 -4.19493&-1	-4.28428&-1 -4.28120&-1	-4.34811&-1 -4.34374&-1	-4.38687&-1 -4.38164&-1	-4.39986&-1 -4.39433&-1
0.9	-4.87157&-1 -4.87196&-1	-4.95389&-1 -4.95453&-1	-5.01353&-1 -5.01433&-1	-5.04966&-1 -5.05054&-1	-5.06176&-1 -5.06266&-1

5.3 THE TWO SPACE VARIABLE CASE

We now consider the equation

$$(\frac{\partial}{\partial t} - L)^2 W = 0 \tag{5.3.1}$$

where L is the second order elliptic differential operator in 2 space dimensions. Let us suppose the appropriate initial conditions are given, and let the new variable be:

$$\phi = \frac{\partial u}{\partial t} - Lu \tag{5.3.2}$$

Therefore, (5.3.1) can be expressed as a system of two-parabolic equations:

$$\frac{\partial u}{\partial t} - Lu \equiv \frac{\partial u}{\partial t} - (\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}) = \phi \tag{5.3.3}$$

and

$$\frac{\partial \phi}{\partial t} - L\phi \equiv \frac{\partial \phi}{\partial t} - (\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2}{\partial y^2}) = 0$$

The system (5.3.3) can be solved by any suitable finite-difference scheme provided the initial and boundary conditions are known.

5.4 THE n^{TH} ORDER FORM OR MULTI-PARABOLIC EQUATION

Let us now consider the general biparabolic equation as given earlier, i.e. when the equation has the following form,

$$(\frac{\partial}{\partial t} - L)^n u = 0 \tag{5.4.1}$$

where L is a general elliptic differential operator in several space dimensions and n is an integer ≥ 1.

For simplicity, let us concern ourselves with $L \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ in (5.4.1). Then we can obtain the following form for (5.4.1);

$$(\frac{\partial}{\partial t} - L)^n u \equiv \sum_{r=0}^{n} (-1)^r \frac{n!}{r!(n-r)!} \frac{\partial^{n-r}(L)^r}{\partial t^{n-r}} u = 0 \tag{5.4.2}$$

where $L^r \equiv \frac{\partial^r}{\partial x^r} + \frac{\partial^r}{\partial y^r}$.

The usual parabolic and biparabolic equation can be obtained from (5.4.2) when n=1 and n=2 respectively.

Let H be an operator of the form $H \equiv \frac{\partial}{\partial t} - L$ and let $Hu = \phi_1$, then

$$\begin{aligned} H^2 u &= H\phi_1 = \phi_2 \\ &\cdots\cdots\cdots\cdots\cdots \\ H^{i+1} u &= H\phi_i = \phi_{i+1} \\ &\cdots\cdots\cdots\cdots\cdots \\ H^{n-1} u &= H\phi_{n-2} = \phi_{n-1} \ . \end{aligned} \tag{5.4.3}$$

Therefore, (5.4.2) becomes

$$H^n u \equiv H\phi_{n-1} = 0 \ . \tag{5.4.4}$$

To evaluate the solution of (5.4.4) one has a system of n-parabolic equations to solve on n-different nets, $\phi_1, \phi_2, \ldots, \phi_{n-1}$ and u respectively. By solving the last equation $H\phi_{n-1}=0$ and substituting in the previous one, the solution of $H\phi_{n-2}=\phi_{n-1}$ can be obtained and so on. Finally, we solve $Hu=\phi_1$. Thus, we are always involved with a single equation to solve. Hence the stability analysis for any numerical method when applied to solve (5.4.2)

(5.4.2) is no more complicated than for the single parabolic equation. However, the investigation must be carried out to guarantee that the accumulated round-off errors from different nets remain bounded.

Problems of type (5.4.1) are called *polycalorique* equations by M. Nocilescu (1954) and have an analytical solution which can be obtained from the following theorem.

Theorem (5.1)

Let $f_0(x), f_1(x), \ldots, f_{n-1}(x)$ be given continuous functions such that

$$|f_i(x)| < Me^{\underline{K}x^2}, \quad i=0,1,2,\ldots,n-1, \tag{5.4.5}$$

where M and $\underline{K}$ are some constant scalars and suppose that

$$\lim_{t\to 0} u(x,t) = f_0(x) \tag{5.4.6}$$

and

$$\lim_{t\to 0} L^i u = f_i(x), \quad i=1,2,\ldots,n-1. \tag{5.4.7}$$

Then, the solution u(x,t) of (5.4.1) is given by,

$$u(x,t) = \frac{1}{2\sqrt{\pi t}} \int_{-\infty}^{\infty} F(x,t)\, e^{-\frac{(\xi-x)^2}{4t}}\, d\xi$$

where

$$F(x,t) = \sum_{p=0}^{n-1} \frac{(-1)^p}{p!}\, t^p\, f_p(x) .$$

For proof see Nicolescu, M. (1954), p.266.

5.5 ROUND-OFF ERROR IN THE COMPUTATIONAL PROCESS

Consider the equation (5.4.1) and let $\underline{\phi}_i^m$ and $\underline{\Phi}_i^m$ be the vectors of the exact and approximate solution of the difference equation applied to solve (5.4.1) respectively, at the time T=m.k, where k is the time increment, and m is the number of steps.

Hence from the equation (5.4.3), we have

$$\underline{\Phi}_i^{m+1} = C\,\underline{\Phi}_i^m + k.\underline{\Phi}_{i+1}^{m+1} \quad , \tag{5.5.1}$$

where i=n-1,n-2,...,2,1,0, ϕ_n=0 and ϕ_0=u, and C is the amplification matrix corresponding to the numerical method which has been used.

We can rewrite (5.5.1) in the following form:

$$\begin{aligned}
\underline{\Phi}_{n-1}^{m+1} &= C\underline{\Phi}_{n-1}^m \\
\underline{\Phi}_{n-2}^{m+1} &= C\underline{\Phi}_{n-2}^m + k.\underline{\Phi}_{n-1}^{m+1} = C\underline{\Phi}_{n-1}^m + k(C\underline{\Phi}_{n-1}^m) \\
&\vdots \\
\underline{\Phi}_2^{m+1} &= C\underline{\Phi}_2^m + k(C\underline{\Phi}_3^m + k(C\underline{\Phi}_4^m + \ldots + k(C\underline{\Phi}_{n-1}^m))\ldots) \\
\underline{\Phi}_1^{m+1} &= C\underline{\Phi}_1^m + k(C\underline{\Phi}_2^m + k(C\underline{\Phi}_3^m + \ldots + k(C\underline{\Phi}_{n-1}^m))\ldots)
\end{aligned} \tag{5.5.2}$$

and

$$\underline{U}^{m+1} = \underline{\Phi}_0^{m+1} = C\underline{\Phi}_0^m + k(C\underline{\Phi}_1^m + \ldots + k(C\underline{\Phi}_{n-1}^m))\ldots).$$

Let $\underline{\varepsilon}_i^m = \underline{\phi}_i^m - \underline{\Phi}_i^m$ be the vector of round-off errors which occur in computing the values $\underline{\Phi}_i^m$ after one time-step, then it is easy to show that the round-off error of the differential equation (5.4.1) has the form:

$$\underline{\varepsilon}_0^{m+1} = C\,\underline{\varepsilon}_0^m + k.C\underline{\varepsilon}_1^m + k^2.C\,\underline{\varepsilon}_2^m + \ldots + k^{n-2}.C\,\underline{\varepsilon}_{n-2}^m + k^{n-1}C\,\underline{\varepsilon}_{n-1}^m \ . \tag{5.5.3}$$

Thus, if $\underline{\varepsilon}$ is the maximum normed vector of errors $\underline{\varepsilon}_i$, we obtain the result:

$$|\underline{\varepsilon}^{m+1}| \leq ||C||\left(\frac{1-k^n}{1-k}\right)\underline{\varepsilon} \ , \tag{5.5.4}$$

which is a bound on the accumulated round-off error from the solution obtained from (n-1)-nets.

Consequently, if $||C||\leq 1$, the round-off error after each step has an upper bound and the process is stable (A.M. Lowan, 1957, p.10).

5.6 CONCLUSION

In this chapter, the splitting strategy has been clearly demonstrated to be an efficient approach in solving parabolic equations of higher order as opposed to the usual approach of treating the partial differential equations by explicit or implicit finite difference methods in situ.

In fact, for higher order equations we feel that it is the only sensible approach to consider. The result obtained from the example shows that good agreement with the theory for large times is obtainable.

CHAPTER VI

A NEW BLOCK HOPSCOTCH TECHNIQUE

6.1 INTRODUCTORY REMARKS

In Chapter III, we considered several hopscotch strategies and demonstrated the advantages of the block hopscotch techniques over the odd-even hopscotch scheme. As mentioned previously the basis of the hopscotch strategy is the division of the set of grid points into two disjoint subsets which include all the grid points in the region under consideration. The formula for creating this subdivision is under the user's control and may depend on the following factors:

i) the order of labelling of the points on the grid;

ii) the shape of the given region;

iii) the degree of implicitness required in the method (Gourlay, 1977, p.777).

Here we will study these three factors which lead to a logical development of the method and enable us to present a new pattern for block hopscotch schemes.

6.2 A NEW BLOCK ORDERING OF THE POINTS

Initially we introduce the basic concept of the labelling of the mesh points, and again consider the parabolic equation (3.4.1) consisting of:

$$\frac{\partial u}{\partial t} = Lu + g(x,y,t) \tag{6.2.1}$$

where the initial and boundary conditions are given as before and the solution of (6.2.1) is required in a cylinder $R\times\{0<t\leq T\}$. Suppose the square region R is covered by some orthogonal grid system in the usual way. Let the set of internal mesh points be denoted by R_h. Let the disjoint subsets B_ℓ (for all $1\leq\ell\leq N$) of R_h be such that:

$$\bigcup_{\ell=1}^{N} B_\ell = R_h \ .$$

Consider the hopscotch formulation:

$$U^{m+1}_{i,j}-k\theta^{m+1}_{i,j}(L_h U^{m+1}_{i,j}+g^{m+1}_{i,j}) = U^{m}_{i,j}+k\theta^{m}_{i,j}(L_h U^{m}_{i,j}+g^{m}_{i,j}) \tag{6.2.2}$$

where the mesh points $(i\Delta x, j\Delta y)$ are contained in the subset B_ℓ, $1\leq\ell\leq M$.

By replacing m with m+1 in (6.2.2) and eliminating $k\theta^{m+1}_{i,j}[L_h U^{m+1}_{i,j}+g^{m+1}_{i,j}]$ from the resulting equation we have:

$$U^{m+2}_{i,j}-k\theta^{m+2}_{i,j}[L_h U^{m+2}_{i,j}+g^{m+2}_{i,j}] = 2U^{m+1}_{i,j}-U^{m+1}_{i,j}-k\theta^{m}_{i,j}[L_h U^{m}_{i,j}+g^{m}_{i,j}] \tag{6.2.3}$$

which according to the definition of $\theta^{m}_{i,j}$ reduces to

$$U^{m+2}_{i,j} = 2U^{m+1}_{i,j}-U^{m}_{i,j} \tag{6.2.4}$$

when m+i+j is odd.

The function θ (which is called a zero-one function) is regarded as being a function of the time index $\underline{m}$ and of the space index vector $\underline{n}$, which is only defined at the mesh points and can be written as θ^{m}_{n} where n is a multi-index $(n_1,n_2,\ldots n_s)$, where s is the number of space dimensions. The fundamental relations which define the format of a hopscotch scheme is as given previously, are the following:

$$\theta_n^m+\theta_n^{m+1} = 1 \ , \qquad (6.2.5)$$

and

$$\theta_n^m\theta_n^{m+1} = 0 \ .$$

The precise definition of θ_n^m will take the form,

$$\theta_n^m = \tfrac{1}{2}[1-(-1)^m] \quad \text{if } (n_1,n_2,\ldots,n_\ell)\in B_\ell \qquad (6.2.6)$$

or

$$= \tfrac{1}{2}[1+(-1)^m] \quad \text{if } (n_1,n_2,\ldots,n_\ell)\in B_\ell .$$

The various choices of B_ℓ results in a different labelling of the mesh points. For instance, the choice of θ_n^m giving odd-even hopscotch corresponds to defining

$$B_\ell \equiv \{(n_1,n_2,\ldots,n_s): \sum_{i=1}^{s} |n_i| = 0 \bmod(2)\}. \qquad (6.2.7)$$

For example, the odd-even hopscotch for heat equation in two space dimensions can be written in the compact form:

$$U_{i,j}^{m+1}-U_{i,j}^m = p[\theta_{i,j}^{m+1}(\delta_x^2+\delta_y^2)U_{i,j}^{m+1}+\theta_{i,j}^m(\delta_x^2+\delta_y^2)U_{i,j}^m] \qquad (6.2.8)$$

where

$$\theta_{i,j}^m = \begin{cases} 1 & \text{if } m+i+j \text{ is even} \\ 0 & \text{" " " odd} \end{cases} \qquad (6.2.9)$$

whilst for the line hopscotch (6.2.9) becomes

$$\theta_{i,j}^m = \begin{cases} 1 & \text{if } m+i \text{ is even} \\ 0 & \text{" " " odd} \end{cases} \qquad (6.2.10)$$

Here we present a new subset B_ℓ as follows:

$$B_\ell \equiv \{(n_1,n_2,\ldots,n_s): \sum_{i=1}^{s} [\frac{n_i+1}{2}] = 0 \bmod(2)\} \ . \qquad (6.2.11)$$

In this case, the values of $\theta_{i,j}^m$ given in (6.2.9) become:

$$\theta_{i,j}^m = \begin{cases} 1 \text{ if } m+[\frac{i+1}{2}]+[\frac{j+1}{2}] \text{ is even} \\ 0 \text{ if } m+[\frac{i+1}{2}]+[\frac{j+1}{2}] \text{ is odd.} \end{cases} \qquad (6.2.12)$$

where [k] is the step function which obtains the largest integer $\leq k$.

In Figure (6.1) the new subdivision of R_h for a two dimensional problem is displayed. The symbols O and X correspond to the implicit and explicit points respectively. As can be seen from Figure (6.2.1)this new block strategy consists of groups of 4-points at which the solution is evaluated using the standard explicit and implicit formulae alternatively. As analogous to the well-known block hopsoctch schemes we call this new scheme *"Group hopscotch"*. An implicit molecule of this group hopscotch method is shown in Figure (6.2.1).

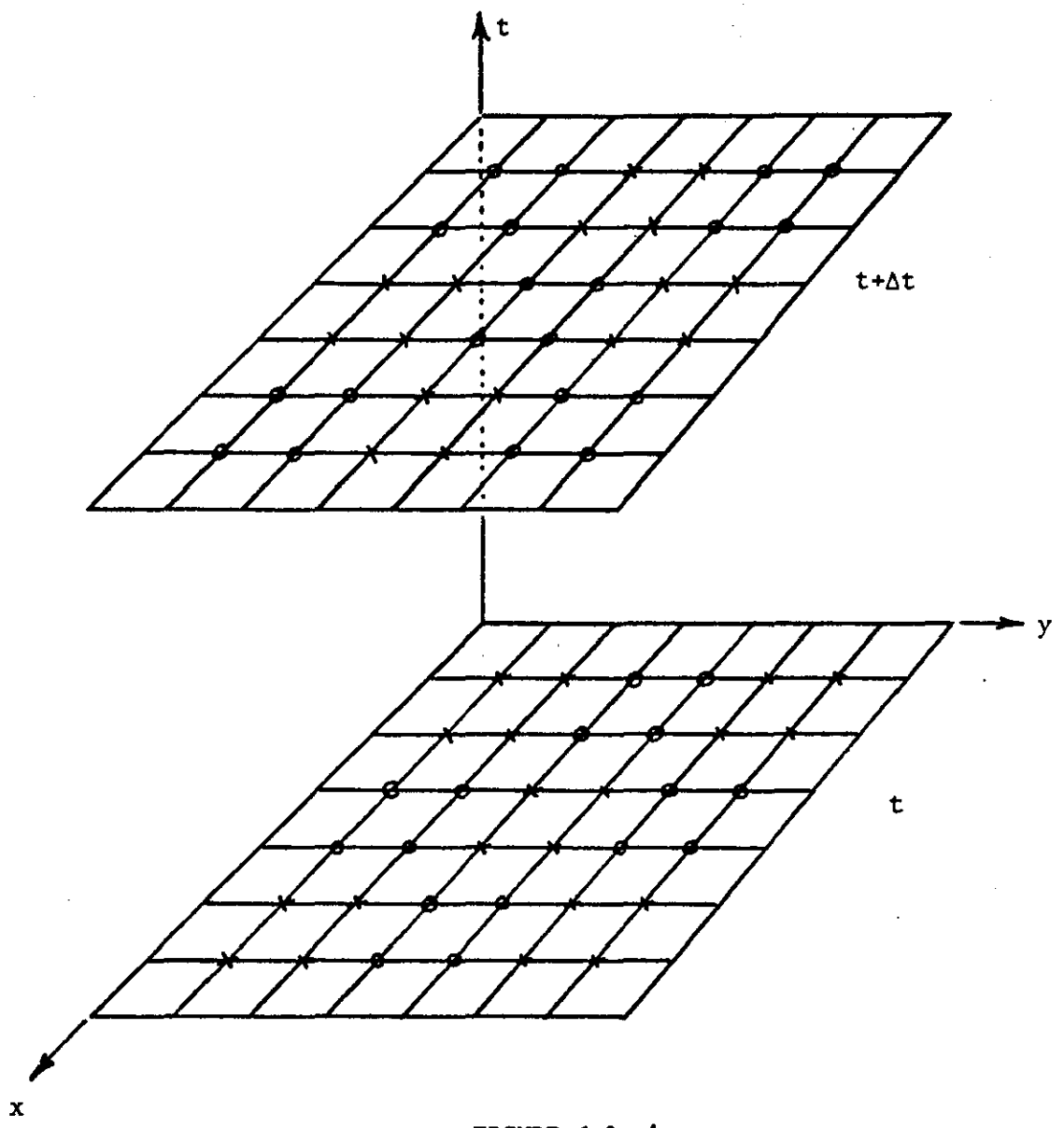


FIGURE 6.2.1

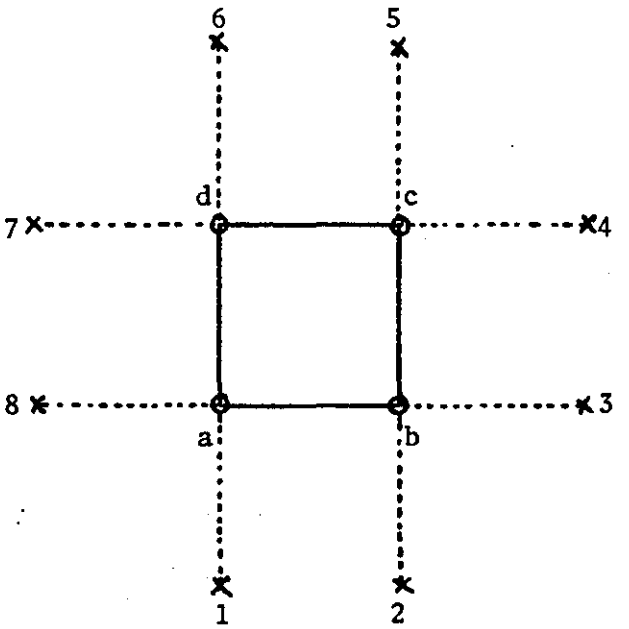


FIGURE 6.2.2

6.3 FORMULATION OF THE GROUP HOPSCOTCH METHOD

Application of the five point finite difference formula to (6.2.1) at explicit and implicit points can be easily verified (as given in Chapter III).

For simplification we let $\Delta x=\Delta y$ and $p=\Delta T/\Delta x^2$, then the implicit formulae for a group of four points displayed in Figure (6.2.2)results in the following:

$$\begin{aligned}
(1+4p)U_a^{m+1}-p(U_b^{m+1}+U_d^{m+1}) &= U_a^m+p(U_1^{m+1}+U_8^{m+1})+\Delta T.g_a^{m+1},\\
(1+4p)U_b^{m+1}-p(U_a^{m+1}+U_c^{m+1}) &= U_b^m+p(U_2^{m+1}+U_3^{m+1})+\Delta T.g_b^{m+1},\\
(1+4p)U_c^{m+1}-p(U_b^{m+1}+U_d^{m+1}) &= U_c^m+p(U_4^{m+1}+U_5^{m+1})+\Delta T.g_c^{m+1},\\
(1+4p)U_d^{m+1}-p(U_c^{m+1}+U_a^{m+1}) &= U_d^m+p(U_6^{m+1}+U_7^{m+1})+\Delta T.g_d^{m+1}\ .
\end{aligned} \tag{6.3.1}$$

Since the solution at the explicit points (i.e. $U_1^{m+1},U_2^{m+1},\ldots,U_8^{m+1}$) are found in the first step, then the right hand side of (6.3.1) is known and (6.3.1) can be solved for the values at points a,b,c and d. In practice, if possible, we solve the system (6.3.1) beforehand and therefore we have a set of explicit formulae to solve for this new block hopscotch method, namely:

$$\begin{aligned}
U_a^{m+1} &= \alpha A + \frac{\beta}{\frac{1}{2\beta}-2\beta}\left[\alpha(A+C)+\frac{\alpha}{2\beta}(B+D)\right],\\
U_b^{m+1} &= \alpha B + \frac{\beta}{\frac{1}{2\beta}-2\beta}\left[\alpha(B+D)+\frac{\alpha}{2\beta}(A+C)\right],\\
U_c^{m+1} &= \alpha C + \frac{\beta}{\frac{1}{2\beta}-2\beta}\left[\alpha(A+C)+\frac{\alpha}{2\beta}(B+D)\right],
\end{aligned} \tag{6.3.2}$$

and

$$U_d^{m+1} = D + \frac{\beta}{\frac{1}{2\beta}-2\beta}\left[\alpha(B+D)+\frac{\alpha}{2\beta}(A+C)\right],$$

where $\alpha=\frac{1}{1+4p}$, $\beta=\frac{p}{1+4p}$ and A,B,C and D are the right-hand side of (6.3.1) respectively.

6.4 THE MATRIX FORM AND STABILITY CONSIDERATIONS

To study the stability of the group hopscotch we first give the explicit and implicit formulation in matrix form.

If we write the standard implicit formula for all the groups (four points) of the region in matrix form, we obtain:

$$\begin{bmatrix} U_{i,j} \\ U_{i+1,j} \\ U_{i+1,j+1} \\ U_{i,j+1} \end{bmatrix}^{m+1} = \begin{bmatrix} U_{i,j} \\ U_{i+1,j} \\ U_{i+1,j+1} \\ U_{i,j+1} \end{bmatrix}^{m} - p \begin{bmatrix} 4 & -1 & 0 & -1 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ -1 & 0 & -1 & 4 \end{bmatrix} \begin{bmatrix} U_{i,j} \\ U_{i+1,j} \\ U_{i+1,j+1} \\ U_{i,j+1} \end{bmatrix}^{m+1} + p$$

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} U_{i-2,j} \\ U_{i-1,j} \\ U_{i-1,j+1} \\ U_{i-2,j+1} \end{bmatrix}^{m+1} + p \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} U_{i,j-2} \\ U_{i+1,j-2} \\ U_{i+1,j-1} \\ U_{i,j-1} \end{bmatrix}^{m+1} +$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} U_{i+2,j} \\ U_{i+3,j} \\ U_{i+3,j+1} \\ U_{i+2,j+1} \end{bmatrix}^{m+1} + p \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} U_{i,j+2} \\ U_{i+1,j+2} \\ U_{i+1,j+3} \\ U_{i,j+3} \end{bmatrix}^{m+1} + \Delta T \begin{bmatrix} g_{i,j} \\ g_{i+1,j} \\ g_{i+1,j+1} \\ g_{i,j+1} \end{bmatrix}^{m+1} . \tag{6.4.1}$$

Now let us introduce new subscripts r and s which are designed to correspond to a group of four points (i,j),(i+1,j),(i+1,j+1) and (i,j+1). This can be done if we let $r=[\frac{i+1}{2}]$ and $s=[\frac{j+1}{2}]$, hence

$$\tilde{U}^{m}_{r,s} = (U^{m}_{i,j}, U^{m}_{i+1,j}, U^{m}_{i+1,j+1}, U^{m}_{i,j+1})^{*} ,$$

and (6.4.1) can be rewritten as:

$$(I+pA)\tilde{U}^{m+1}_{r,s} + p(B\tilde{U}^{m+1}_{r-1,s} + B^{*}\tilde{U}^{m+1}_{r+1,s} + C\tilde{U}^{m+1}_{r,s-1} + C^{*}\tilde{U}^{m+1}_{r,s+1}) = \tilde{U}^{m}_{r,s} + \Delta T\, \tilde{g}^{m+1}_{r,s} \tag{6.4.2}$$

$$r,s = 1,2,\ldots,[\tfrac{N+1}{2}] .$$

where

$$A = \begin{bmatrix} 4 & -1 & 0 & -1 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ -1 & 0 & -1 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \text{ and } C = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and I is the (4×4) unitary matrix.

Equation (6.4.2) can be further simplified if we write in compact form: i.e.,

$$(I+pH)\tilde{\underline{U}}^{m+1} = \tilde{\underline{U}}^{m} + \Delta t.\tilde{\underline{g}}^{m+1} \tag{6.4.3}$$

where

$$\tilde{\underline{U}} = (\tilde{U}_{1,1},\tilde{U}_{1,2},\ldots,\tilde{U}_{1,L},\tilde{U}_{2,1},\tilde{U}_{2,2},\ldots,\tilde{U}_{2,L},\ldots,\tilde{U}_{L,1},\tilde{U}_{L,2},\ldots,\tilde{U}_{L,L})^{*},$$

$$L=[\tfrac{N}{2}]$$

and H is a block-tridiagonal matrix of order $(4.[\frac{N}{2}]^2)$ of the following form:

$$H = \begin{bmatrix} \alpha & \beta^* & & & \\ \beta & \alpha & \beta^* & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \beta^* \\ 0 & & & \beta & \alpha \end{bmatrix} \quad \alpha = \begin{bmatrix} A & B^* & & & \\ B & A & B^* & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & B^* \\ 0 & & & B & A \end{bmatrix}, \beta = \begin{bmatrix} C & & & \\ & \ddots & & 0 \\ & & \ddots & \\ 0 & & & C \end{bmatrix}$$

(6.4.4)

and I in (6.4.3) is a unitary matrix of order H.

Now if we write the explicit formula for all the groups in the region and simplify it by giving the compact matrix form analogous to (6.4.3) we have,

$$\tilde{\underline{U}}^{m+1} = (I-pH)\tilde{\underline{U}}^{m}+\Delta T.\tilde{\underline{g}}^{m} \; . \tag{6.4.5}$$

Next, let us define the diagonal matrix $I_1^{(B)}$ such that if $[I_1^{(B)}U^{2m}]_{i,j}$ denotes the component of the vector $I_1^{(B)}\underline{U}^{2m}$ corresponding to the spatial mesh point $(i\Delta x, j\Delta y)$, then

$$[I_1^{(B)}U^{2m}]_{i,j} = \theta_{i,j}^{2m}U_{i,j}^{2m} = \theta_{\ell}^{2m}U_{i,j}^{2m}, \text{ for all } i,j,\ell$$

where $(i\Delta x, j\Delta y)$ B_ℓ and θ_ℓ^{2m} is defined by (6.2.5). Hence, $I_1^{(B)}$ is a block diagonal matrix, whose N block diagonal elements are alternately (4×4) unit and null matrices, and whose order corresponds to the number of mesh points in the associated sets B_n. Let also $I_2^{(B)}=I-I_1^{(B)}$.

We now define the general two step block hopscotch process globally to be,

$$\begin{aligned}(I+pI_2^{(B)}H)\tilde{\underline{U}}^{2m+1} &= (I-pI_1^{(B)}H)\tilde{\underline{U}}^{2m}+\Delta T(I_2^{(B)}\tilde{\underline{g}}^{2m+1}+I_1^{(B)}\tilde{\underline{g}}^{2m})\\ (I+pI_1^{(B)}H)\tilde{\underline{U}}^{2m+2} &= (I-pI_2^{(B)}H)\tilde{\underline{U}}^{2m+1}+\Delta T(I_1^{(B)}\tilde{\underline{g}}^{2m+2}+I_2^{(B)}\tilde{\underline{g}}^{m+1}),\end{aligned} \tag{6.4.6}$$

where the data from the boundary of the plane region R have been absorbed in the $\tilde{\underline{g}}$ vectors.

Equation (6.4.6) can be further simplified in the form:

$$\tilde{\underline{U}}^{2m+2} = T\tilde{\underline{U}}^{2m} + \tilde{\underline{b}}^{2m+2} \tag{6.4.7}$$

where

$$T=(I+pI_1^{(B)}H)^{-1}(I-pI_2^{(B)}H)(I+pI_2^{(B)}H)^{-1}(I-pI_1^{(B)}H)$$

and

$$\begin{aligned}\tilde{\underline{b}}^{2m+2} = {}&\Delta T(I+pI_1^{(B)}H)^{-1}(I-pI_2^{(B)}H)(I+pI_2^{(B)}H)^{-1}(I_2^{(B)}\tilde{\underline{g}}^{2m+1}+I_1^{(B)}\tilde{\underline{g}}^{2m})+\\ &\Delta T(I+pI_1^{(B)}H)^{-1}(I_1^{(B)}\tilde{\underline{g}}^{2m+2}+I_2^{(B)}\tilde{\underline{g}}^{2m+1}) .\end{aligned}$$

In a similar manner as given previously, we can write

$$T = (I+pI_1^{(B)}H)^{-1}\tilde{T}(I+pI_1^{(B)}H) , \tag{6.4.8}$$

where

$$\tilde{T} = (I-pI_2^{(B)}H)(I+pI_2^{(B)}H)^{-1}(I-pI_1^{(B)}H)(I+pI_1^{(B)}H)^{-1} .$$

The stability of the block hopscotch which was given by Theorem (3.12) can now be applied by using the $A^{\frac{1}{2}}$ norm which requires the matrix H to be positive-definite. However it can be easily verified that H is positive-definite as it is a symmetric diagonally dominant matrix with positive diagonal elements (Theorem 1.4).

We can also deduce the stability of the new block hopscotch from the Theorem (3.13). We need to show that a unique permutation matrix exists such that $P^T I_1^{(B_1)} P = I_1^{(B_2)}$. For instance, we can find the matrix P to invert the group hopscotch scheme to the odd-even (point) hopscotch scheme. As an example, if we consider, $I_1^{(B_1)}$ to be:

$$I_1^{(B_1)} = \begin{bmatrix} 1 & & & & & & & \\ & 1 & & & & & & \\ & & 1 & & & 0 & & \\ & & & 1 & & & & \\ & & & & 0 & & & \\ & & & & & 0 & & \\ & & 0 & & & & 0 & \\ & & & & & & & 0 \end{bmatrix} \tag{6.4.9}$$

then P can be easily found to be:

$$P = \begin{bmatrix} 1 & & & & & & 1 & 0 \\ & 0 & & 0 & & & & \\ & & 1 & & & 0 & & \\ & & & 0 & 1 & & & \\ & 0 & & 1 & 0 & & 0 & \\ & & 0 & & & 1 & & \\ & 1 & & 0 & & & 0 & \\ 0 & & & & & & & 1 \end{bmatrix} \tag{6.4.10}$$

thus

$$P^T I_1^{(B_1)} P = \begin{bmatrix} 1 & & & & & & & \\ & 0 & & & & & & \\ & & 1 & & & 0 & & \\ & & & 0 & & & & \\ & & & & 1 & & & \\ & & & & & 0 & & \\ & & 0 & & & & 1 & \\ & & & & & & & 0 \end{bmatrix} = I_1^{(B_2)} \tag{6.4.11}$$

which corresponds to the point hopscotch. Therefore, as the point hopscotch is known to be stable, the new block hopsoctch is also stable.

6.5 LOCAL TRUNCATION ERRORS AND CONVERGENCE OF GROUP HOPSCOTCH

The L.T.E. of a general two-step process was considered in Chapter III, where it was shown that

$$d^m_{i,j} = O(kh^\sigma + k^2/h^2 + k^3)$$

and further we mentioned that, for all those methods in which $\theta^m_{i,j} = \eta^m_{i,j}$, the terms involving $O(k^3)$ and $O(kh^\sigma)$ are the same and only the term containing $O(k^3/h^2)$ differs from one scheme to another.

In the case of the group hopscotch scheme we deduce the magnitude of the latter term to compare the accuracy of the scheme with other hopscotch techniques and derive the convergence property for the method.

The terms involving $O(k^3/h^2)$ are as follows:

$$2k^3 L_1 L_2 \frac{\partial u^m_{i,j}}{\partial t} \quad \text{and} \quad k^3 L_1 \left(\theta^{m+1}_{i,j} \frac{\partial g^{(1)m}_{i,j}}{\partial t} + \eta^{m+1}_{i,j} \frac{\partial g^{(2)m}_{i,j}}{\partial t}\right)$$

where $L_1, L_2, g^{(1)}$ and $g^{(2)}$ are given in the general splitting formula in Chapter III. Thus for the hopscotch procedures ($\theta^m_{i,j} = \eta^m_{i,j}$) we obtain:

$$2k^3 L_1 L_2 \frac{\partial u^m_{i,j}}{\partial t} = 2k^3 \theta^{m+1}_{i,j} L_h \theta^m_{i,j} L_h \frac{\partial u^m_{i,j}}{\partial t}, \tag{6.5.1}$$

and for the second term, we have,

$$k^3 \theta^{m+1}_{i,j} L_h \theta^{m+1}_{i,j} \frac{\partial g^m_{i,j}}{\partial t} \quad . \tag{6.5.2}$$

Now we expand (6.5.1) and (6.5.2) at each corner of an implicit group shown in Figure 6.2. For instance at the corner a considering $\theta^m_{i-1,j} = \theta^m_{i,j-1} = \theta^{m+1}_{i,j} = 1$, we have,

$$2k^2 L_1 L_2 \frac{\partial u^m_{i,j}}{\partial t} = 2\frac{k^3}{h^2}(A_{i,j} + B_{i,j}) L \frac{\partial u^m_{i,j}}{\partial t} + O(k^3) + O\left(\frac{k^3}{h^2} h^\sigma\right), \tag{6.5.3}$$

and

$$k^3 \theta^{m+1}_{i,j} L_h \theta^{m+1}_{i,j} \frac{g^m_{i,j}}{t} = \frac{2k^3}{h^2}\left[(C_{i,j} + D_{i,j} + E_{i,j}) \frac{\partial g^m_{i,j}}{\partial t}\right] = -\frac{2k^3}{h^2} (A_{i,j} + B_{i,j}) \frac{\partial g^m_{i,j}}{\partial t} + O\left(\frac{k^3}{h^2} h^\sigma\right) \tag{6.5.4}$$

Therefore, the term containing $O(k^3/h^2)$ at the corner (a) is:

$$\frac{2k^3}{h^2}[(A_{i,j}+B_{i,j})(\frac{\partial u^m_{i,j}}{\partial t} - \frac{\partial g^m_{i,j}}{\partial t})]+O(k^3/h^2.h)+O(k^3). \qquad (6.5.5)$$

In the same way, the expression (6.5.5) for the other corners of Figure 6.2 can be found to be:

at corner b: $$\frac{2k^3}{h^2}[(A_{i,j}+D_{i,j})(\frac{\partial u^m_{i,j}}{\partial t} - \frac{\partial g^m_{i,j}}{\partial t})]+O(k^3/h^2.h)+O(k^3)$$

at corner c: $$\frac{2k^3}{h^2}[(D_{i,j}+E_{i,j})(\frac{\partial u^m_{i,j}}{\partial t} - \frac{\partial g^m_{i,j}}{\partial t})]+O(k^3/h^2.h)+O(k^3) \qquad (6.5.6)$$

and at corner d: $$\frac{2k^3}{h^2}[(E_{i,j}+B_{i,j})(\frac{\partial u^m_{i,j}}{\partial t} - \frac{\partial g^m_{i,j}}{\partial t})]+O(k^3/h^2.h)+O(k^3).$$

The comparison of (6.5.5) and (6.5.6) with the L.T.E. found for other hopscotch schemes (McGuire, 1971 and Gane 1974), we observe that, all three block hopscotch procedures have the same order of accuracy. The comparative accuracy of the methods also depends a great deal on the nature of the elliptic differential operator L, whose finite difference analogue is given by (3.7.3). In particular, if L is the self-adjoint second order elliptic operator, namely,

$$LU = \frac{\partial}{\partial x}(\alpha(x,y)\frac{\partial u}{\partial n}) + \frac{\partial}{\partial y}(\beta(x,y)\frac{\partial u}{\partial y})-\gamma(x,y)u(x,y,t) \qquad (6.5.7)$$

where α,β and γ are piecewise continuous functions in R, the closure of R, and satisfy

$$\gamma(x,y)>0,\ \beta(x,y)>0 \text{ and } \gamma(x,y)\geq 0 \text{ for all } (x,y)\in\bar{R}, \qquad (6.5.8)$$

we have from (3.7.1)-(3.7.3),

$$|C^{(1)}_{i,j}| \geq |B_{i,j}+D_{i,j}| \qquad (6.5.9)$$

and

$$|C^{(2)}_{i,j}| \geq |A_{i,j}+E_{i,j}| \qquad (6.5.10)$$

with $B_{i,j}$ and $D_{i,j}$ of opposite sign to $C^{(1)}_{i,j}$ and $A_{i,j}$, $E_{i,j}$ of opposite sign to $C^{(2)}_{i,j}$. If further, we suppose the coefficients $A_{i,j}, B_{i,j}, D_{i,j}$ and $E_{i,j}$ obtain the same values, then (6.5.8) and (6.5.9) can be rewritten:

$$|C^{(1)}_{i,j}| \geq |A_{i,j}+B_{i,j}| \tag{6.5.10}$$

$$|C^{(2)}_{i,j}| \geq |A_{i,j}+D_{i,j}| \quad . \tag{6.5.11}$$

Also we have,

$$|C_{i,j}| \geq |C^{(1)}_{i,j}|, |C^{(2)}_{i,j}|$$

Therefore, we conclude that, the new hopscotch scheme has the same accuracy as the other block schemes, particularly for the case $L \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ where $A_{i,j}=B_{i,j}=E_{i,j}=D_{i,j}=-1$.

However, in this new scheme, we require the number of mesh points in the x and y directions to be an odd number. It can be seen that an even number of mesh points leads to a half group (near the boundaries and therefore we have to apply the odd-even hopscotch scheme for these points, which disturbs the accuracy of the method. This is shown in the experimental results later.

The necessity of covering the region by complete groups of points is a disadvantage for the group hopscotch method, since specially in non-rectilinear regions we may not be able to do so and hence, we lose accuracy. In this case, the line-hopscotch procedure seems to be superior.

The convergence of the new scheme can be achieved following the analysis given in Chapter III.

Since the L.T.E. of the formula remains unchanged i.e. $d_{i,j}=O(kh^2+k^3/h^2+k^3)$ for all i,j and m, therefore,

$$||d^{2\ell-1}||_{A^{\frac{1}{2}}} \leq m^{\frac{1}{2}} \; O(kh^2+k^3/h^2+k^3)$$

and from (3.7.12), it follows that,

$$||e^{2r}||_{A^{\frac{1}{2}}} \leq r.k.c.m^{3/2} \; O(k^2+h^2/h^2+h^2).$$

Hence, provided $k \to 0$ faster than h, the convergence Theorem 3.7 also holds for the new hopscotch scheme.

6.6 EFFICIENCY OF THE GROUP HOPSCOTCH SCHEME

In section 3.5, we briefly mentioned the comparison of the work and programming involved in odd-even, line hopscotch methods together with the Peaceman-Rachford scheme. An investigation to show the number of arithmetic operations required for different methods, is given by Gourlay and McGuire (1971) who claimed that the odd-even hopscotch scheme is 3 times faster than the line hopscotch and the Peaceman-Rachford is about 4 2/3 slower than the odd-even hopscotch.

In the following table we give the number of arithmetic operations involved to evaluate the solution at the implicit points for three different hopscotch schemes, where n is the number of mesh points in the x and y directions.

Hopscotch Methods	Additions	Multips.	Divisions
Odd-Even	$5n^2/2$	$3n^2/2$	-
Group	$5n^2/2$	$4n^2/2$	-
Line	$5.5n^2/2$	$4.5n^2/2$	$n^2/2$

TABLE 6.1

The Table 6.1 shows that, the new scheme is faster than the line hopscotch but slightly slower than odd-even hopscotch.

As far as the storage requirement is concerned, all the methods given in Table 6.1, are similar.

6.7 HIGHER DIMENSIONAL PROBLEMS

The advantages of the new hopscotch strategy become apparent when we are concerned with problems involving 3 or more dimensions. In this section, we extend the new scheme to the three space dimensions for the heat conduction equation, i.e.

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} + g(x,y,z,t) \qquad (6.7.1)$$

where the appropriate initial and boundary conditions are given.

The simple square molecule in two dimension (Figure 6.2) now becomes a cubic molecule as shown in Figure 6.7.1. A molecule of implicit points is surrounded by explicit points and vice versa. The system of equations for the implicit molecules are as follows:

$$\begin{aligned}
&(1+6p)U_a^{m+1} - p(U_b^{m+1}+U_d^{m+1}+U_e^{m+1}) = U_a^m + p(U_1^m+U_{17}^m+U_{13}^m) + k.g_a^{m+1} \\
&(1+6p)U_b^{m+1} - p(U_a^{m+1}+U_c^{m+1}+U_f^{m+1}) = U_b^m + p(U_2^m+U_5^m+U_{18}^m) + k.g_b^{m+1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&(1+6p)U_h^{m+1} - p(U_g^{m+1}+U_e^{m+1}+U_d^{m+1}) = U_h^m + p(U_{11}^m+U_{15}^m+U_{23}^m) + k.g_e^{m+1}
\end{aligned} \qquad (6.7.2)$$

at the points a,b,c,...,g and h where without loss of generality we have denoted $\Delta x=\Delta y=\Delta z$ and $p=\Delta T/\Delta x^2$. The system (6.7.2) can be rewritten in a more compact matrix form as before. Hence, we obtain:

$$\begin{aligned}
&\text{Explicit form} \quad \tilde{\underline{U}}^{m+1} = (I-pH)\tilde{\underline{U}}^m + k.\tilde{\underline{g}}^m \ , \\
&\text{Implicit form} \quad (I+pH)\tilde{\underline{U}}^{m+1} = \tilde{\underline{U}} + k.\tilde{\underline{g}}^{m+1} \ ,
\end{aligned} \qquad (6.7.3)$$

and the hopscotch formula remains unchanged, namely,

$$(I+pI_2H)\tilde{\underline{U}}^{m+1} = (I-pI_1H)\tilde{\underline{U}}^m + k(I_2\tilde{\underline{g}}^{m+1} + I_1\tilde{\underline{g}}^m) \ , \qquad (6.7.4)$$

where

$$H = \begin{bmatrix} \alpha & \beta^* & \gamma^* & & & 0 \\ \beta & \gamma & & \ddots & & \\ \gamma & & \ddots & & \gamma^* & \\ & \ddots & & & \beta^* & \\ 0 & & & & \gamma & \beta & \alpha \end{bmatrix}, \alpha = \begin{bmatrix} a & b^* & & & 0 \\ b & \ddots & \ddots & & \\ & \ddots & & b^* & \\ 0 & & b & a \end{bmatrix}, \gamma = \begin{bmatrix} d & & & 0 \\ & d & & \\ & & \ddots & \\ 0 & & & d \end{bmatrix}$$

$$\beta = \begin{bmatrix} c & & & 0 \\ & \ddots & & \\ & & \ddots & \\ 0 & & & c \end{bmatrix}, a = \left[\begin{array}{cccc:ccc} 6 & -1 & 0 & -1 & -1 & & 0 \\ -1 & 6 & -1 & 0 & & \ddots & \\ 0 & -1 & 6 & -1 & & & -1 \\ -1 & 0 & -1 & 6 & & & \\ \hdashline -1 & & & & 6 & -1 & 0 & -1 \\ & \ddots & & 0 & -1 & 6 & -1 & 0 \\ & & \ddots & & 0 & -1 & 6 & -1 \\ 0 & & & -1 & -1 & 0 & -1 & 6 \end{array}\right],$$

$$b = \left[\begin{array}{cccc:ccc} 0 & -1 & 0 & 0 & & & \\ 0 & 0 & 0 & 0 & & & \\ 0 & 0 & 0 & 0 & & 0 & \\ 0 & 0 & -1 & 0 & & & \\ \hdashline & & & & 0 & -1 & 0 & 0 \\ & & & & 0 & 0 & 0 & 0 \\ & 0 & & & 0 & 0 & 0 & 0 \\ & & & & 0 & 0 & -1 & 0 \end{array}\right], c = \left[\begin{array}{cccc:ccc} 0 & 0 & 0 & -1 & & & \\ 0 & 0 & -1 & 0 & & & \\ 0 & 0 & 0 & 0 & & 0 & \\ 0 & 0 & 0 & 0 & & & \\ \hdashline & & & & 0 & 0 & 0 & -1 \\ & & & & 0 & 0 & -1 & 0 \\ & 0 & & & 0 & 0 & 0 & 0 \\ & & & & 0 & 0 & 0 & 0 \end{array}\right]$$

$$d=\left[\begin{array}{cc|cc} 0 & & -1 & \\ & 0 & -1 & 0 \\ & & 0 & -1 \\ & & & -1 \\ \hline 0 & & 0 & \end{array}\right]$$

and I_1 and I_2 in (6.7.4) are block-diagonal matrices whose diagonal elements are 8×8 null and unitary matrices alternatively.

It can be shown that H is a symmetric, diagonally dominant matrix with positive diagonal elements, hence it is positive definite. Therefore, according to the Theorem (3.6), the hopscotch algorithm is stable. In this case $\theta^m_{i,j,n}$ is denoted to be:

$$\theta^{m+1}_{i,j,n} = \begin{cases} 1 \text{ if } [\frac{i+1}{2}]+[\frac{j+1}{2}]+[\frac{n+1}{2}]+m \text{ is even} \\ 0 \text{ if } \quad " \quad " \quad " \quad " \text{ odd} \end{cases}$$

It is worthy to note that, the stability of the hopscotch technique is invariant with respect to the dimensions involved.

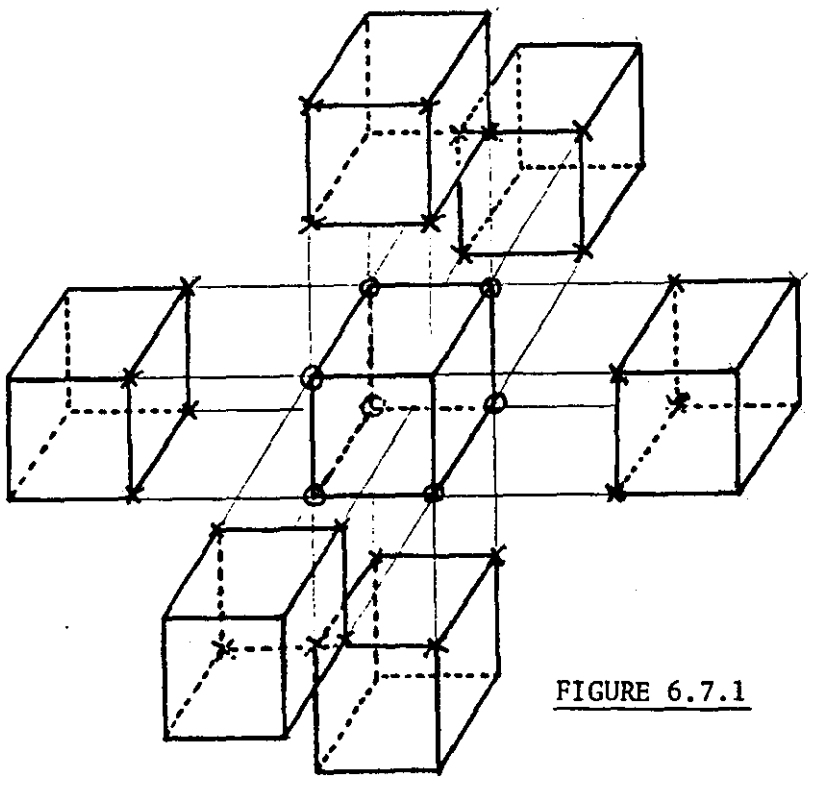


FIGURE 6.7.1

6.8 IMPLICITNESS RATIO

As far as the literature is concerned, the purpose of introducing the block hopscotch technique was to improve the accuracy over and above that of the point hopscotch. This improvement was shown to be due to the L.T.E. where the term $O(k^3/h^2)$ is smaller in magnitude for the block hopscotch schemes when compared with the point scheme.

However, it is not the only factor to consider for achieving higher accuracy. It can be seen from the experimental observations that, the ratio of explicit points neighbouring an implicit point plays an important role. In the following table, we demonstrate this ratio for different schemes.

$$\mu = \frac{\text{No. of implicit points}}{\text{No. of explicit neighbouring points}}$$

Hopscotch Scheme	1-Dim	2-Dims	3-Dims	n-Dims
Point	1/2	1/4	1/6	1/2 n
Line	-	1/2	1/4	1/2 (n-1)
Peripheral	-	1/2	1/4	1/2 (n-1)
Group	1	1/2	1/3	1/n

TABLE 6.2

As can be seen from Table 6.2, the new strategy for the hopscotch methods becomes more important when dealing with higher dimensional problems especially, when the amount of work involved is compared.

6.9 DEVELOPMENT OF THE GROUP HOPSCOTCH SCHEME

This idea of grouping the mesh points can be applied for a larger group than was suggested earlier as was displayed in Figure 6.2. For instance, a group of 9 points in two space dimensions as illustrated in Figure 6.9.1 can be treated in a similar implicit manner.

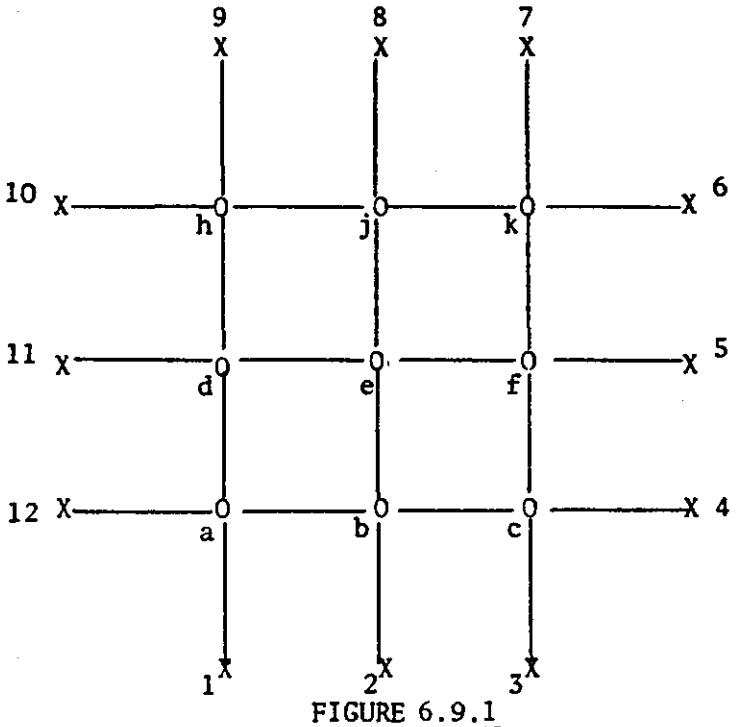


FIGURE 6.9.1

In this group all the points a,b,...,j,k are implicit and all the points 1,2,...,12 are evaluated by the explicit formula. The formulation for this new group can easily be verified when the equation (6.2.1) is considered and is as follows:

$$\begin{bmatrix} 1+4p & -p & & -p & & & & & \\ -p & 1+4p & -p & & -p & & & 0 & \\ & -p & 1+4p & & & -p & & & \\ -p & & & 1+4p & -p & & -p & & \\ & -p & & -p & 1+4p & -p & & -p & \\ & & -p & & -p & 1+4p & & & -p \\ & & & -p & & & 1+4p & -p & \\ & 0 & & & -p & & -p & 1+4p & -p \\ & & & & & -p & & -p & 1+4p \end{bmatrix} \begin{bmatrix} U_a \\ U_b \\ U_c \\ U_d \\ U_e \\ U_f \\ U_h \\ U_g \\ U_h \end{bmatrix}^{m+1} = \begin{bmatrix} \phi_a \\ \phi_b \\ \\ \\ \\ \\ \\ \phi_g \\ \phi_h \end{bmatrix} \qquad (6.9.1)$$

where

$$\begin{aligned}
\phi_a &= U_a^m + p(U_1^m + U_{12}^m) + \Delta T.g_a^{m+1} \\
\phi_b &= U_b^m + pU_2^m \qquad\quad + \Delta T.g_b^{m+1} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\phi_k &= U_k^m + p(U_6^m + U_7^m) + \Delta T.g_k^{m+1}
\end{aligned} \tag{6.9.2}$$

and the θ-function becomes

$$\theta_{i,j}^m = \begin{cases} 1 \text{ if } [\frac{i+1}{3}]+[\frac{j+1}{3}]+m \text{ is odd} \\ 0 \text{ " } \quad \text{ " } \quad \text{ " } \quad \text{ " even.} \end{cases} \tag{6.9.3}$$

The stability of the new group scheme can be verified by the application of the theorem (3.12), since the permutation matrix P can be found in the same fashion as (6.4.10) which shows the equivalence of the new group hopscotch scheme with the point hopscotch scheme and consequently the stability of the method can be deduced immediately. However, it is not difficult to investigate the stability directly.

The ratio of implicitness indicated in the Table 6.2, for the new scheme is given by the following formula:

$$\mu_n = \frac{3^n}{2n \times 3^{n-1}} \quad , \quad n=1,2,3,\ldots, \tag{6.9.4}$$

where n is the number of space dimensions. The ratio (6.9.4) shows an improvement for the latter scheme compared with the other hopscotch schemes given in Table 6.2.

The L.T.E. of the new group hopscotch scheme can be evaluated in the same way as given previously and the terms involving $O(k^3/h^2)$ are as given below.

At the four corners a,c,h and k shown in Figure 6.4 the principal parts of $O(k^3/h^2)$ are found respectively to be:

$$2\,k^3/h^2(A_{i,j}+B_{i,j})\left[L_h\frac{\partial u^m_{i,j}}{\partial t}-\frac{\partial g^m_{i,j}}{\partial t}\right],$$

$$2\,k^3/h^2(A_{i,j}+D_{i,j})\left[L_h\frac{\partial u^m_{i,j}}{\partial t}-\frac{\partial g^m_{i,j}}{\partial t}\right],$$

$$2\,k^3/h^2(B_{i,j}+E_{i,j})\left[L_h\frac{\partial u^m_{i,j}}{\partial t}-\frac{\partial g^m_{i,j}}{\partial t}\right], \qquad (6.9.5)$$

and $$2\,k^3/h^2(E_{i,j}+D_{i,j})\left[L_h\frac{\partial u^m_{i,j}}{\partial t}-\frac{\partial g^m_{i,j}}{\partial t}\right]$$

whereas at the points b,d,f and g we have

$$2\,k^3/h^2A_{i,j}\left(L_h\frac{\partial u^m_{i,j}}{\partial t}-\frac{\partial g^m_{i,j}}{\partial t}\right),$$

$$2\,k^3/h^2B_{i,j}\left(L_h\frac{\partial u^m_{i,j}}{\partial t}-\frac{\partial g^m_{i,j}}{\partial t}\right),$$

$$2\,k^3/h^2D_{i,j}\left(L_h\frac{\partial u^m_{i,j}}{\partial t}-\frac{\partial g^m_{i,j}}{\partial t}\right), \qquad (6.9.6)$$

and $$2\,k^3/h^2E_{i,j}\left(L_h\frac{\partial u^m_{i,j}}{\partial t}-\frac{\partial g^m_{i,j}}{\partial t}\right),$$

and finally at the centre point e, the term $O(k^3/h^2)$ vanishes. Thus generally speaking, the accuracy of the new scheme has the same order as the four-point group hopscotch scheme.

Although the 9 points group hopscotch provides a slightly more accurate approximation to the solution of the parabolic equation, the amount of work involved to solve the system (6.9.1), now becomes more critical. However, the numerical comparison is made for different hopscotch schemes including the latter which will be illustrated in the forthcoming sections.

From now on we call the scheme presented in section 6.2 as "Group 4" and the scheme given in 6.9 as "Group 9" hopscotch schemes.

6.10 EXTENSION OF THE GROUP STRATEGY TO PROBLEMS OF ONE SPACE DIMENSION

One of the advantages of the new block hopscotch (i.e., the group hopscotch) scheme is the application to one-dimensional linear problems. For instance, the scheme (6.3.1) for the heat equation in this case becomes:

$$\begin{aligned}(1+2p)U_i^{j+1}-pU_{i+1}^{j+1} &= U_i^j+pU_{i-1}^{j+1}+\Delta T.g_i^{j+1}\\ (1+2p)U_{i+1}^{j+1}-pU_i^{j+1} &= U_{i+1}^j+pU_{i+2}^{j+1}+\Delta T.g_{i+1}^{j+1}\end{aligned} \qquad (6.10.1)$$

where the mesh points selection for explicit and implicit points are displayed in Figure 6.10.1

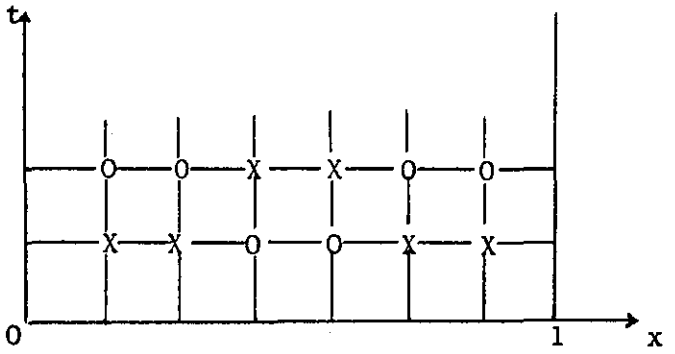


FIGURE 6.10.1

The system (6.10.1) can easily be solved beforehand to provide a total explicit scheme for the new technique.

The advantages of this scheme over the point-hopscotch scheme is due to the improvement of the L.T.E. of the group hopscotch scheme as was shown in the two-dimensional heat equation, as well as the implicitness ratio which in this case becomes $\mu_1=1$ whereas this ratio for the point-hopscotch was $\mu_1=1/2$.

One can also apply the analogy of the group 9-hopscotch scheme to the one-dimensional space problem and choose the explicit and implicit mesh points in a manner as shown in Figure 6.10.2.

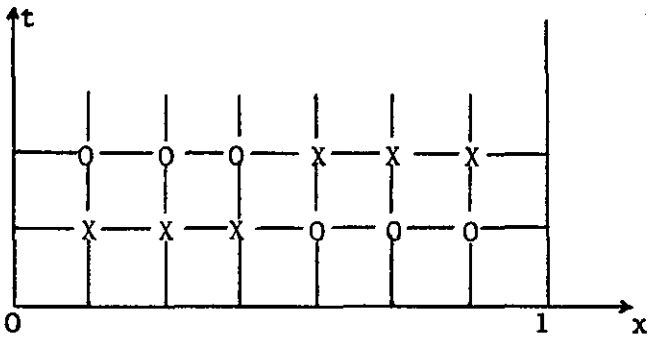
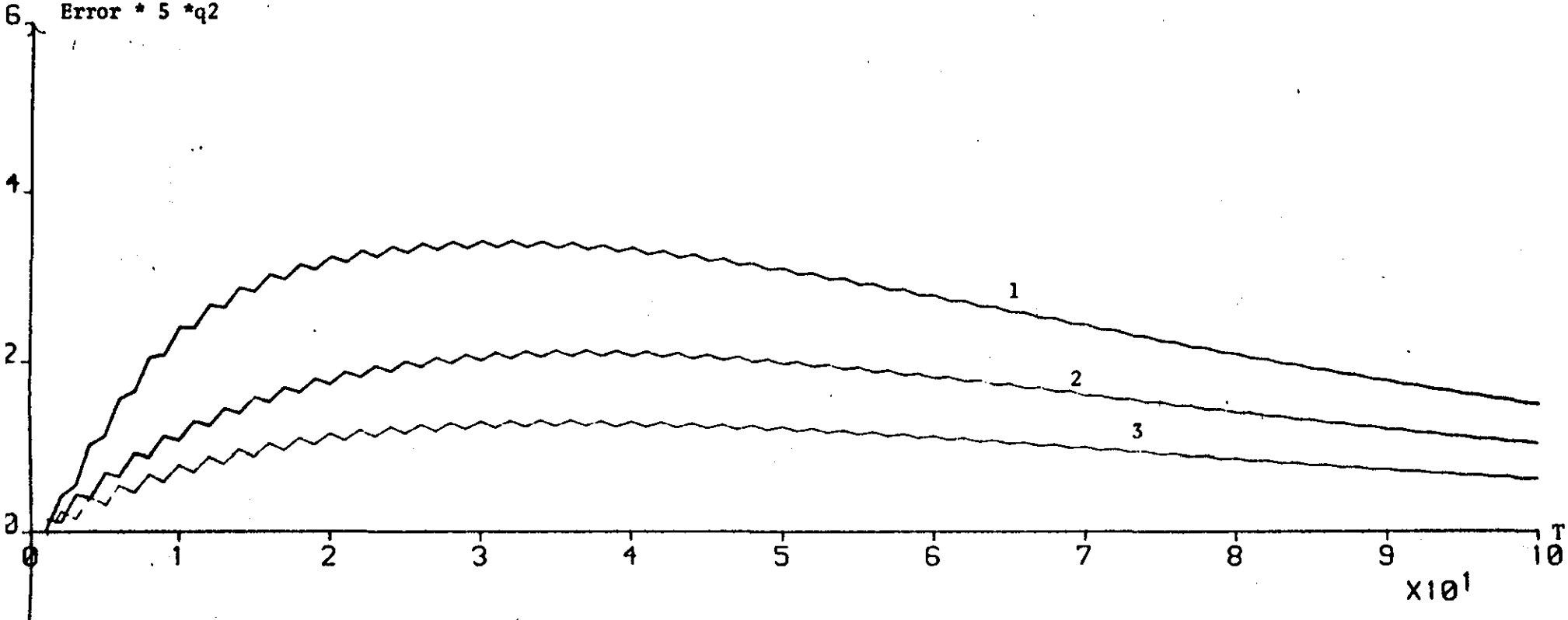


FIGURE 6.10.2

To compare the accuracy of the different hopscotch techniques in one-dimensional problems, we have solved the heat equation $\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}$ subject to the initial values $u(x,0)=\sin(x)$ and the boundary values $u(0,t)=u(1,t)=0$. The difference between the approximate solution obtained from three different hopscotch schemes with the exact solution $u(x,t)=\sin x.\exp(-t)$ are displayed in Figure 6.10.3, which indicates that as the group size increases so does the accuracy of the approximate solution.



N=20, h=1/N, P=1 ,

The error curves of point, 2-point and 3-point hopscotch schemes at the point (x=N/2.ht=mΔt) are illustrated by curves 1,2 and 3 respectively.

FIGURE 6.10.3

6.11 A COMPOSITE HOPSCOTCH SCHEME AND THE TREATMENT OF INCONSISTENCY

It was shown earlier that the principal part of the terms involving $O(k^3/h^2)$ in the L.T.E. for point-hopscotch and group 4-hopscotch schemes are respectively:

$$-\frac{2k^3}{h^2} C_{i,j}[L_h \frac{\partial u^m_{i,j}}{\partial t} - \frac{\partial g^m_{i,j}}{\partial t}]$$

and

$$\frac{2k^3}{h^2}(A_{i,j}+B_{i,j})[L_h \frac{\partial u^m_{i,j}}{\partial t} - \frac{\partial g^m_{i,j}}{\partial t}] ,$$

where the terms $A_{i,j}$,$B_{i,j}$ and $C_{i,j}$ are defined in (3.7.2)-(3.7.3).

For the parabolic equations with the second order elliptic partial differential operator given by

$$L \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \qquad (6.11.1)$$

We observe that,

$$A_{i,j}+B_{i,j}+C_{i,j}+D_{i,j}+E_{i,j} = 0 \qquad (6.11.2)$$

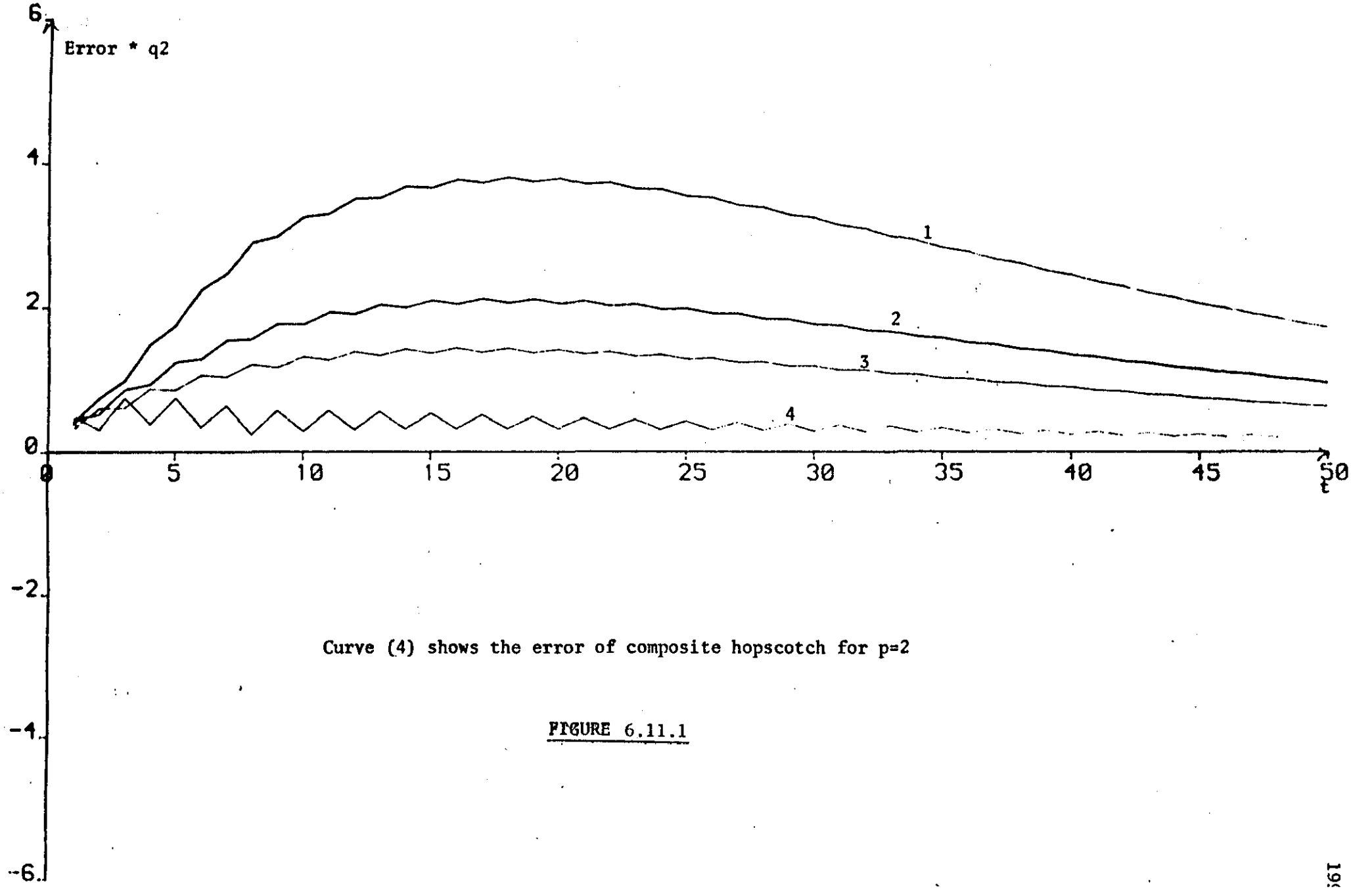
which in the case of $A_{i,j}=E_{i,j}$ and $B_{i,j}=D_{i,j}$, results in

$$2(A_{i,j}+B_{i,j}) = - C_{i,j} . \qquad (6.11.3)$$

Consequently, if we subtract the approximate solution of the point-hopscotch method from two times of the Group 4-hopscotch scheme, the terms $O(k^3/h^2)$ is eliminated and therefore not only do we obtain a more accurate result than the previous block hopscotch schemes but we also avoid the inconsistency phenomenon which is due to the term $O(k^3/h^2)$, provided (6.11.3) is satisfied in the region under consideration.

However, for the Dirichlet boundary conditions, the equation (6.11.3) is no longer true near the boundaries and although we have some improvement in the accuracy, the inconsistency term cannot always be eliminated completely whereas for Neumann boundary conditions satisfactory results might prevail.

The additional computational work can be compromised by applying a larger time step. The composite scheme can also be applied to one-dimensional problems. To show the improvement achieved for the case of the one-dimensional problem a comparison of the error curves with the previous scheme is given in Figure 6.11.1.



Curve (4) shows the error of composite hopscotch for p=2

FIGURE 6.11.1

6.12 COMPUTATIONAL EXPERIMENTS

In the following computational experiments we consider the same examples as given by Gourlay and McGuire (1971), Gourlay and Gane (1976).

Consider the equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + g(x,y,t) \qquad (6.12.1)$$

in the domain $0 \leq x,y \leq 1$ when $g(x,y,t)$ are given for two different cases as

$$\text{i)} \quad g(x,y,t) = \sin x.\sin y\, e^{-t} - 4 \qquad (6.12.2)$$

$$\text{ii)} \quad g(x,y,t) = \sin x.e^{-t}/(1+y)^2 - 2x - 6xy\ . \qquad (6.12.3)$$

The appropriate initial and boundary conditions are given to provide the exact solution of the problem as:

$$\text{i)} \quad u(x,y,t) = \sin x.\sin y\ .e^{-t} + x^2 + y^2 \qquad (6.12.4)$$

$$\text{and ii)} \quad u(x,y,t) = \sin x.\log(1+y)e^{-t} + x^3y + xy^2\ . \qquad (6.12.5)$$

Experiment (i)

The accuracy of the different hopscotch schemes are compared in the following tables. At first we choose N=19 (which is the number of mesh points in the x and y direction) which is suitable to cover the grid points by different groups exactly. The errors at the middle mesh point and also the maximum, minimum and average errors of different schemes are also given in the tables. Different tables present the errors for different mesh ratios and the time is chosen to be T=0.277 which corresponds to 100 steps when p=1.

We also examine the accuracy of different schemes when N=10, which the Group 4 scheme fails to cover the mesh points completely and so the mesh points near one of the boundary lines in each direction requires a

different procedure which as mentioned previously disturbs the accuracy obtained.

The errors at the middle point for different mesh ratios and for each scheme are given in the Tables 6.5 and 6.6 respectively. At this time T=1 is chosen for each problem.

1st Example

N=19, P=0.125, k=0.00277, T=0.2770 sec.

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	5.0900&-6	5.4358&-6	8.6148&-8	2.5233&-6
Group 4	5.6743&-6	6.0593&-6	9.6171&-8	2.8134&-6
Line	5.6579&-6	6.0366&-6	9.5996&-8	2.8084&-6
Peripheral	5.6741&-6	6.0574&-6	9.6260&-8	2.8140&-6
Group 9	5.8694&-6	6.2651&-6	9.9551&-8	2.9112&-6
Composite	6.2585&-6	6.6827&-6	1.0619&-7	3.1034&-6

(a)

P=0.25

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	1.6092&-6	1.7317&-6	2.6986&-8	7.9464&-7
Group 4	3.9478&-6	4.2272&-6	6.7107&-8	1.9556&-6
Line	3.8825&-6	4.1682&-6	6.6408&-8	1.9358&-6
Peripheral	3.9472&-6	4.2196&-6	6.7464&-8	1.9584&-6
Group 9	4.7281&-6	5.0501&-6	8.0625&-8	2.3470&-6
Composite	6.2865&-6	6.7391&-6	1.0723&-7	3.1169&-6

(b)

P=0.5

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-1.2321&-5	1.3259&-5	2.0988&-7	6.1230&-6
Group 4	-2.9598&-6	3.2800&-6	4.9200&-8	1.4758&-6
Line	3.2203&-6	3.4655&-6	5.1992&-8	1.5552&-6
Peripheral	-3.2314&-6	3.3384&-6	4.7762&-8	1.4647&-6
Group 9	1.6336&-7	6.6546&-7	4.4383&-10	1.0095&-7
Composite	6.4017&-6	7.0792&-6	1.1148&-7	3.1721&-6

(c)

TABLE 6.3 (continued)

P=1

Hospcotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-6.8171&-5	7.3038&-5	1.1609&-6	3.3851&-5
Group 4	-3.0624&-5	3.2753&-5	5.1536&-7	1.5215&-5
Line	-3.1666&-5	3.3901&-5	5.2654&-7	1.5534&-5
Peripheral	-3.0631&-5	3.3182&-5	5.0959&-7	1.5171&-5
Group 9	-1.8111&-5	1.9814&-5	2.9833&-7	8.9422&-6
Composite	6.9234&-6	8.5362&-6	1.3018&-7	3.4232&-6

(d)

P=2

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-2.9232&-4	3.1280&-4	4.9901&-6	1.4511&-4
Group 4	-1.4173&-4	1.5098&-4	2.3926&-6	7.0378&-5
Line	-1.4591&-4	1.5602&-4	2.4374&-6	7.1656&-5
Peripheral	-1.4175&-4	1.5298&-4	2.3694&-6	7.0198&-5
Group 9	-9.1427&-5	9.9040&-5	1.5174&-6	4.5172&-5
Composite	8.8637&-6	1.6315&-5	3.4233&-8	4.6090&-6

(e)

P=4

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-1.4851&-3	1.5096&-3	3.6866&-5	7.4650&-4
Group 4	-6.1383&-4	6.5104&-4	9.7994&-6	2.9563&-4
Line	-6.0632&-4	6.5785&-4	1.0171&-5	3.0046&-4
Peripheral	-6.1249&-4	6.5274&-4	9.3916&-6	2.9556&-4
Group 9	-3.9641&-4	4.2733&-4	6.2931&-6	1.9014&-4
Composite	2.5742&-4	3.4102&-4	4.7866&-6	1.5545&-4

(f)

TABLE 6.3

2nd Example

N=19, P=0.125, k=0.00277 T=0.2770 sec.

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-7.2296&-6	9.7601&-6	1.0552&-7	4.1932&-6
Group 4	-6.7372&-6	9.4470&-6	8.3266&-8	3.9486&-6
Line	-6.7509&-6	9.4448&-6	8.8406&-8	3.9527&-6
Peripheral	-6.7373&-6	9.4401&-6	8.3557&-8	3.9480&-6
Group 9	-6.5729&-6	9.3354&-6	6.4349&-8	3.8661&-6
Composite	-6.2448&-6	9.1339&-6	2.0198&-8	3.7039&-6

(a)

P=0.25

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-1.0162&-5	1.1882&-5	2.0872&-7	5.6508&-6
Group 4	-8.1920&-6	1.0452&-5	1.3176&-7	4.6719&-6
Line	-8.2468&-6	1.0469&-5	1.4023&-7	4.6885&-6
Peripheral	-8.1923&-6	1.0447&-5	1.3103&-7	4.6695&-6
Group 9	-7.5348&-6	9.9417&-6	1.1463&-7	4.3419&-6
Composite	-6.2215&-6	9.1351&-6	3.4623&-8	3.6938&-6

(b)

P=0.5

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-2.1900&-5	2.2689&-5	6.2169&-7	1.1484&-5
Group 4	-1.4013&-5	1.5029&-5	3.1368&-7	7.5657&-6
Line	-1.4232&-5	1.5262&-5	3.4754&-7	7.6323&-6
Peripheral	-1.4013&-5	1.5125&-5	3.1074&-7	7.5561&-6
Group 9	-1.1382&-5	1.2976&-5	2.4510&-7	6.2454&-6
Composite	-6.1251&-6	9.1377&-6	5.6625&-9	3.6582&-6

(c)

TABLE 6.4 (continued)

P=1

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-6.8955&-5	7.2363&-5	1.5064&-6	3.4865&-5
Group 4	-3.7322&-5	3.8970&-5	9.4648&-7	1.9153&-5
Line	-3.8199&-5	3.9336&-5	9.5629&-7	1.9420&-5
Peripheral	-3.7325&-5	3.8935&-5	9.4135&-7	1.9115&-5
Group 9	-2.6758&-5	2.8044&-5	7.5814&-7	1.3865&-5
Composite	-5.6900&.6	9.5149&-6	2.5921&-8	3.5873&-6

(d)

P=2

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-2.5771&-4	2.7350&-4	4.8200&-6	1.2863&-4
Group 4	-1.3092&-4	1.3716&-4	2.5705&-6	6.5668&-5
Line	-1.3444&-4	1.4071&-4	2.6099&-6	6.6737&-5
Peripheral	-1.3093&-4	1.3947&-4	2.5498&-6	6.5512&-5
Group 9	-8.8574&-5	9.2710&-5	1.8114&-6	4.4423&-5
Composite	-4.1347&-6	1.1825&-5	1.0194&-8	4.0284&-6

(e)

P=4

Hopscotch Schemes	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-1.2630&-3	1.2800&-3	3.2065&-5	6.3500&-4
Group 4	-5.2810&-4	5.5464&-4	8.9920&-6	2.5526&-4
Line	-5.2173&-4	5.6265&-4	9.2510&-6	2.5930&-4
Peripheral	-5.2683&-4	5.5535&-4	8.6510&-6	2.5520&-4
Group 9	-3.4531&-4	3.6724&-4	5.9956&-6	1.6683&-4
Composite	2.0681&-4	2.7761&-4	3.1624&-6	1.2556&-4

(f)

TABLE 6.4

1st Example

N=10, P=1, T=1

Hopscotch Methods	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-1.1118&-4	1.2376&-4	7.1198&-6	-6.3040&-5
Group 4	-5.1865&-5	6.2787&-5	3.1470&-6	3.1316&-5
Line	-5.0473&-5	6.1354&-5	3.2366&-6	2.9429&-5
Peripheral	-5.2835&-5	6.1237&-5	3.0225&-6	2.8028&-5
Group 9	-2.5200&-5	3.5082&-5	1.3491&-6	1.5253&-5
Composite	7.4475&-6	2.0742&-5	9.3314&-8	4.6368&-6

(a)

P=2

Hopscotch Methods	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-4.7417&-4	5.2699&-4	3.0417&-5	2.6900&-4
Group 4	-2.3854&-4	2.8451&-4	1.4557&-5	1.4279&-4
Line	-2.3297&-4	2.7995&-4	1.4929&-5	1.3526&-4
Peripheral	-2.4242&-4	2.7947&-4	1.4054&-5	1.2965&-4
Group 9	-1.3216&-4	1.7395&-4	8.2226&-6	7.8683&-5
Composite	-2.9075&-6	1.0104&-4	5.2946&-7	2.0209&-5

(b)

P=4

Hopscotch Methods	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-1.6761&-3	1.6764&-3	1.6872&-4	9.4098&-4
Group 4	-1.0695&-3	1.3427&-3	6.8409&-5	6.4286&-4
Line	-1.1282&-3	1.2411&-3	6.6486&-5	6.1084&-4
Peripheral	-1.1429&-3	1.1820&-3	7.0231&-5	5.9829&-4
Group 9	-5.9393&-4	7.1930&-4	3.7559&-5	3.2634&-4
Composite	-4.6278&-4	1.0190&-3	1.0622&-5	3.6882&-4

(c)

TABLE 6.5

2nd Example

N=10, P=1, T=1

Hopscotch Methods	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-1.1400&-4	1.2353&-4	8.9382&-6	6.5133&-5
Group 4	-6.3956&-4	7.0466&-5	5.4939&-6	3.8365&-5
Line	-6.2791&-5	7.0277&-5	5.5763&-6	3.6810&-5
Peripheral	-6.4798&-5	7.0151&-5	5.3607&-6	3.5633&-5
Group 9	-1.1493&-5	4.7287&-5	3.1762&-6	2.4873&-5
Composite	-1.3912&-5	2.5870&-5	8.3921&-7	1.1597&-5

(a)

P=2

Hopscotch Methods	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-4.2028&-4	4.5929&-4	2.9078&-5	2.3879&-4
Group 4	-2.2137&-4	2.5594&-4	1.5300&-5	1.3222&-4
Line	-2.1671&-4	2.5342&-4	1.5642&-5	1.2601&-4
Peripheral	-2.2474&-4	2.5288&-4	1.4880&-5	1.2130&-4
Group 9	-1.3175&-4	1.6343&-4	9.8353&-6	7.8374&-5
Composite	-2.2451&-5	9.0890&-5	4.8167&-7	2.5664&-5

(b)

P=4

Hopscotch Methods	Midpoint Error	Maximum Error	Minimum Error	Average Error
Point	-1.4796&-3	1.4796&-3	1.4582&-4	8.1741&-4
Group 4	-9.1865&-4	1.1385&-3	6.2841&-5	5.5116&-4
Line	-9.7116&-4	1.0643&-3	5.9830&-5	5.2553&-4
Peripheral	-9.8442&-4	1.0065&-3	6.3441&-5	5.1479&-4
Group 9	-5.2036&-4	6.2009&-4	3.5931&-5	2.8697&-4
Composite	-3.5774&-4	8.4642&-4	5.5152&-6	3.0352&-4

(c)

TABLE 6.6

6.13 EQUIVALENCE OF THE GROUP-HOPSCOTCH WITH THE DU-FORT-FRANKEL SCHEME

In the same manner as Gourlay & McKee (1977), we can prove the equivalence of the new block hopscotch with the implicit Du-Fort-Frankel scheme.

Let us consider the two-space dimensional heat equation and rewrite the Du-Fort-Frankel method for this equation, i.e.,

$$(1+4p)U_{i,j}^{m+1}-2p[U_{i+1,j}^{m}+U_{I-1,j}^{m}+U_{i,j+1}^{m}+U_{i,j-1}^{m}]-(1-4p)U_{i,j}^{m+1} = 0 \qquad (6.13.1)$$

If we apply the equation (6.13.1) for a group of four points as displayed in Figure (6.13.1) and give the matrix form, we obtain:

$$\begin{bmatrix} 1+4p & -p & 0 & -p \\ -p & 1+4p & -p & 0 \\ 0 & -p & 1+4p & -p \\ -p & 0 & -p & 1+4p \end{bmatrix} \begin{bmatrix} U_{i,j} \\ U_{i+1,j} \\ U_{i+1,j+1} \\ U_{i,j+1} \end{bmatrix}^{m+1} + 2p \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} U_{i-2,j} \\ U_{i-1,j} \\ U_{i-1,j+1} \\ U_{i-2,j+1} \end{bmatrix}^{m}$$

$$+2p \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} U_{i,j-2} \\ U_{i+1,j-2} \\ U_{i+1,j-1} \\ U_{i,j-1} \end{bmatrix}^{m} + 2p \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} U_{i,j+2} \\ U_{i+1,j+2} \\ U_{i+1,j+3} \\ U_{i,j+3} \end{bmatrix}^{m}$$

$$+2p \begin{bmatrix} 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} U_{i+2,j} \\ U_{i+3,j} \\ U_{i+3,j+1} \\ U_{i+2,j+1} \end{bmatrix}^{m} - \begin{bmatrix} 1-4p & p & 0 & p \\ p & 1-4p & p & 0 \\ 0 & p & 1-4p & p \\ p & 0 & p & 1-4p \end{bmatrix} \begin{bmatrix} U_{i,j} \\ U_{i+1,j} \\ U_{i+1,j+1} \\ U_{i,j+1} \end{bmatrix}^{m-1} = 0$$

(6.13.2)

The compact form of (6.13.2) can again be written as follows:

$$(I+pA)\tilde{U}_{r,s}^{m+1} = -2p(B\tilde{U}_{r-1,s}^{m}+C\tilde{U}_{r,s-1}^{m}+C^{*}\tilde{U}_{r,s+1}^{m}+B^{*}\tilde{U}_{r+1,s}^{m})+(I-pA)\tilde{U}_{r,s}^{m-1} \qquad (6.13.3)$$

where A,B,C and $\tilde{U}_{r,s}$ have the same definition as given in (6.4.2).

The usual explicit and implicit replacements in the same form

of (6.13.3) are as follows:

$$\tilde{U}^{m+1}_{r,s} = (I-pA)\tilde{U}^{m}_{r,s} - p(B\tilde{U}^{m}_{r-1,s} + C\tilde{U}^{m}_{r,s-1} + C^{*}\tilde{U}_{r,s+1} + B^{*}\tilde{U}_{r+1,s}) \quad (6.13.4)$$

$$(I+pA)\tilde{U}^{m}_{r,s} + p(B\tilde{U}_{r-1,s} + C\tilde{U}^{m}_{r,s-1} + C^{*}\tilde{U}^{m}_{r,s+1} + B^{*}\tilde{U}^{m}_{r+1,s}) = \tilde{U}^{m-1}_{r,s} \; . \quad (6.13.5)$$

Now, we multiply the equations (6.13.4) and (6.13.5) by (I+pA) and (I-pA) respectively and then add the new formulae to obtain the equation (6.13.3).

Hence the equivalence of the Group hopscotch and the GDFF can be observed which enables us to state the following theorem.

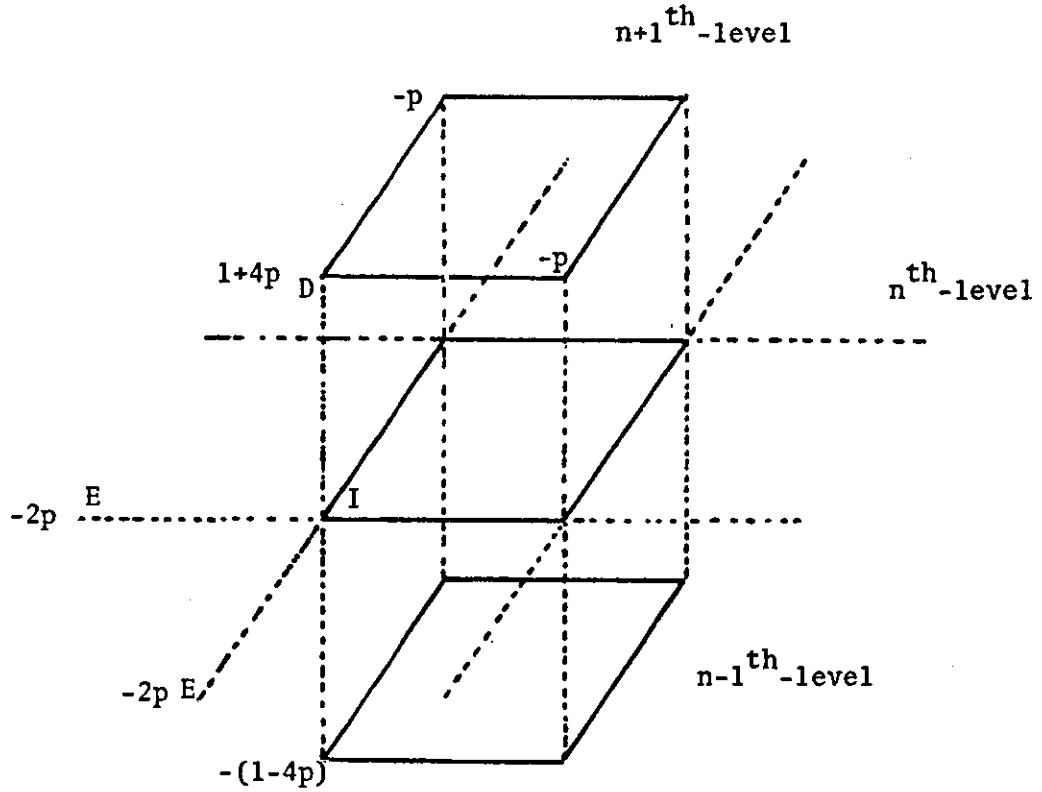


FIGURE 6.13.1

Theorem 6.1

The Group hopscotch is equivalent to the Group Du-Fort-Frankel scheme in the following sense:

i) GDFF must be started by using one step of Group hopscotch

ii) GDFF must only be employed on alternate groups.

Optimum Parameter for Convergence

In the same manner as §3.9 we shall prove the convergence of the new block hopscotch scheme by considering the equivalent DFF scheme and attempt to find the best value of p for which the convergence attains its optimum rate.

The related analysis can be done by two different approaches:

i) Considering the GDFF for a single point, i.e.,

$$(1+4p)U_{i,j}^{m+1}-p(U_{i+1,j}^{m+1}+U_{i,j+1}^{m+1})-2p(U_{i-1,j}^{m}+U_{i,j-1}^{m})-p(U_{i+1,j}^{m-1}+U_{i,j+1}^{m-1})$$
$$-(1-4p)U_{i,j}^{m-1} = 0 \qquad (6.13.6)$$

and substitute the related Fourier terms into (6.13.6). From which we obtain,

$$(1+4p-pC-\sqrt{-1}S)\mu^2-2p(C-\sqrt{-1}S)\mu-[(1-4p+pC)+\sqrt{-1}pS] = 0 \qquad (6.13.7)$$

where $C=\cos\alpha h+\cos\beta h$ and $S=\sin\alpha h+\sin\beta h$,

and μ represents the spectral radius of the iteration matrix. We were not able to solve (6.13.7) analytically due to its complexity to obtain a formula similar to those presented by Gourlay & McKee (1979) for point and line hopscotch schemes.

ii) By matrix analysis in which the whole region is considered by the following formula:

$$(I+pM)\tilde{\underline{U}}^{m+1}-2pN\tilde{\underline{U}}^{m}-(I-pM)\tilde{\underline{U}}^{m+1} = 0 \qquad (6.13.8)$$

where M and N are block matrices of order $(\frac{N-1}{2})\times(\frac{N-1}{2})$ represented below,

$$M=\begin{bmatrix} A & & & 0 \\ & A & & \\ & & A & \\ & & & A \\ 0 & & & \end{bmatrix} \qquad N=\begin{bmatrix} 0 & B^* & & C^* & \\ B & & B & & 0 \\ & B & 0 & C^* & \\ C & & 0 & B^* & C^* \\ & C & B & & B^* \\ & & B & 0 & C^* \\ 0 & & C & & 0 \\ & & & C & \end{bmatrix}$$

and A,B,C are 4×4 matrices as given in (6.13.3).

Since M and N are not commutative, therefore (6.13.8) can not be converted to the quadratic equation similar to (3.9.11) and the solution of (6.13.8) can not be verified in this way.

However, we simplified (6.13.8) to the following form

$$\tilde{\underline{U}}^{n+1}-2p(I+pM)^{-1}N\tilde{\underline{U}}^{n}-(I+pM)^{-1}(I-pM)\tilde{\underline{U}}^{n-1} = 0 \qquad (6.13.10)$$

and assume $A=(I+pM)^{-1}N$ and $B=(I+pM)^{-1}(I-pM)$ are commutative (without M and N to be such). Suppose λ and ν are the eigenvalues of A and B respectively. Then (6.13.10) can be written as,

$$\mu^2-2p\lambda\mu-\nu = 0 \ , \qquad (6.13.11)$$

which can be solved for μ.

The results deduced from (6.13.11) were not in agreement with the experiments obtained from the solution of the model problem and hencefore the assumption that $AB=BA$ did not help to overcome the difficulty to find an analytical form for the best value of p and the optimum value of μ.

However, we expect the rate of convergence for the new hopscotch scheme to be similar to the line-hopscotch and this is what we observed from the numerical experiments.

6.14 APPLICATION OF THE GROUP-HOPSCOTCH TECHNIQUE TO SOLVE PARABOLIC AND ELLIPTIC EQUATIONS WITH MIXED DERIVATIVES

The hopscotch strategy is applied to solve parabolic and elliptic equations with mixed derivatives by Gourlay & McKee (1977) as presented in section 3.10. Here we shall apply the grouping strategy for such problems and demonstrate the formulation involved together with some numerical results and make a comparison for different approximate solutions obtained.

To begin, we start with the equation (3.10.1) and write the finite-difference replacement for a group of four implicit points as shown in Figure 6.14.1.

In order to write a suitable finite difference replacement for the mixed derivative term in (3.10.1) we choose $(\sigma^2+\sigma^4)$ at the corners A and C and $(\sigma^1+\sigma^3)$ at the two other corners, where σ^i, i=1,2,3,4 are defined by (3.10.4).

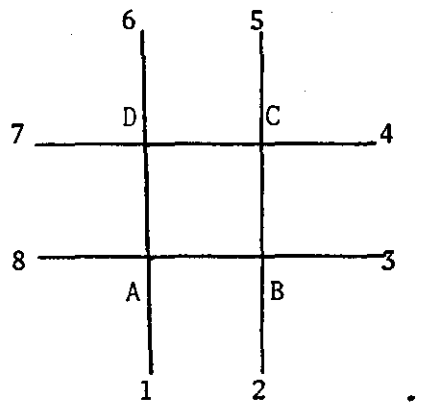


FIGURE 6.14.1

Therefore at the corner A and C we have

$$L_h \equiv \frac{1}{h^2}\left[a\delta_x^2+c\delta_y^2+b(\sigma^2+\sigma^4)\right] \qquad (6.14.1)$$

whereas at the corners B and D we have

$$L_h \equiv \frac{1}{h^2}\left[a\delta_x^2+c\delta_y^2+b(\sigma^1+\sigma^3)\right] \qquad (6.14.2)$$

where a,b and c are the coefficients given in (3.10.1).

The matrix formulation of the group's implicit scheme for the equation (3.10.1) leads to the same system as (6.4.2) with different coefficient matrices as follows:

$$A = \begin{bmatrix} 2(a+b+c) & -(a+b) & 0 & -(c+b) \\ -(a-b) & 2(a+b-c) & -(c-b) & 0 \\ 0 & -(c-b) & 2(a+b+c) & -(a+b) \\ -(c-b) & 0 & -(a-b) & 2(a+b-c) \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & a+b & -b & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & b & a-b & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 & -b & c+b \\ 0 & 0 & c-b & b \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \qquad (6.14.3)$$

The elliptic equation with mixed derivatives can also be solved by the Group-hopscotch scheme without any substantial difficulties.

Numerical Comparison

Here we solve some examples as given by Gourlay and McKee (1977) for the new scheme and compare the results to show the accuracy obtained.

Example 1 - Parabolic equation, constant coefficients

Here the problem consists of

$$\frac{\partial u}{\partial t} = a\frac{\partial^2 u}{\partial x^2} + 2b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} + g(x,y,t) \qquad (6.14.4)$$

where a=0.1, b=0.05 and c=0.15 together with the initial condition

$$u(x,y,0) = \sin(x+y) , \qquad (6.14.5)$$

and the boundary conditions,

$$\begin{aligned} u(0,y,t) &= e^{-(a+2b+c)t}\sin y, \\ u(1,y,t) &= e^{-(a+2b+c)t}\sin (1+y), \\ u(x,0,t) &= e^{-(a+2b+c)t}\sin x , \\ u(x,1,t) &= e^{-(a+2b+c)t}\sin (1+x), \end{aligned} \qquad (6.14.6)$$

and $g(x,y,t) = 0$.

The theoretical solution is given by

$$u(x,y,t) = e^{-(a+2b+c)t}\sin(x+y) \ .$$

The numerical calculation was carried out for N=11, and T=0.5 for different values of p. The results are shown in Table (6.14.1).

Constant coeff. h=1/11, T=0.5

p	No. of Steps	Errors of Point Hopscotch	Errors of Group 4 Hopscotch	Errors of Line Hopscotch
0.1	500	0.000134	0.000141	0.000120
0.5	100	0.000253	0.000133	0.000110
1	50	0.000382	0.000142	0.000079
5	10	0.000797	0.000653	0.000934

TABLE 6.14.1

Example 2 - Parabolic equation with variable coefficients

Here the problem consists of equation (6.14.4) together with,

$$a = x^2/2 + y^2, \quad b = -(x^2+y^2)/2 \text{ and } \quad c = x^2+y^2/2 \tag{6.14.7}$$

subject to the initial condition

$$u(x,y,0) = x^2y + xy^2$$

and the boundary conditions

$$\begin{aligned} u(0,y,t) &= 0 \\ u(1,y,t) &= (y+y^2)e^{-t} \\ u(x,0,t) &= 0 \\ u(x,1,t) &= (x+x^2)e^{-t} \end{aligned} \tag{6.14.8}$$

with g(x,y,t)=0. The theoretical solution is

$$u(x,y,t) = (x^2y+xy^2)e^{-t} \ .$$

The numerical results for N=11 and various values of p are presented in Table (6.14.2).

Variable coeff. h=1/11, T=0.5

p	No. of Steps	Errors of Point Hopscotch	Errors of Group 4 Hopscotch	Errors of Line Hopscotch
0.1	500	0.000043	0.000002	0.000002
0.5	100	0.000087	0.000037	0.000039
1	50	0.000238	0.000149	0.000157
5	10	0.054760	0.005103	0.005661

TABLE 6.14.2

Example 3 - Elliptic equation with constant coefficients

The group hopscotch iterative scheme is applied to solve the elliptic equation

$$Lu = -g(x,y,t) \qquad (6.14.9)$$

where L is as given earlier together with a=c=1 and b=-0.5. The initial guess $U^0_{i,j}=1$ was used and the required accuracy of 10^{-6} was achieved in 72 iterations for both the line and group hopscotch methods with the optimum value p=1.

6.15 CONCLUSIONS

We list below the advantages and disadvantages which have been observed in analysing the methods.

1. It is well-known that the hopscotch technique was established to convert complicated problems into some straightforward computational processes which are independent of dimensions.
   However, since the first hopscotch scheme (i.e. point-hopscotch) was not sufficiently accurate (i.e. in the case of long time steps, the accuracy dramatically decreases due to the L.T.E.) the block-hopscotch scheme was introduced to give a more accurate solution. However by using a block strategy, the explicitness of the method is sacrificed and therefore the scheme becomes less efficient.
   In this chapter we established a scheme which has the same order of accuracy as the other block hopscotch schemes. Moreover, it is also totally explicit and compatible with the point hopscotch method although it takes slightly more computational time.

2. By introducing this new strategy, it seems the choice of selecting the set B_ℓ given in §6.2, is now completed, and there would be no other alternative for a new strategy, except to make the group larger which obviously is not very useful.

3. One of the disadvantages of any hopscotch process is the Du-Fort Frankel type of inconsistency due to the term $O(k^3/h^2)$ in the L.T.E. The composite schemes were presented in §6.11 to tackle this inconsistency, and from the results displayed in the Tables (6.3)-(6.6), this difficulty seems to have been overcome to some extent.

4. This new strategy can be applied to one-dimensional problems, which have no analogy in the other block hopscotch methods.

5. As was shown in Chapter III, since the error of the hopscotch process changes the sign for different ratio $p=\Delta T/\Delta x^2$, therefore there exists a specific value of p for which, one obtains the most accurate solution from the scheme. This specific value of p increases for block hopscotch schemes which allows the user to apply longer time-steps and make the technique more efficient. This is shown in Figure 6. where the curves of global error at the mid-point of the one-dimensional problem for point, 2-point and 3-point hopscotch schemes are demonstrated.

6. Since the grouping strategy (particularly the Group-4 scheme) is totally explicit it is more suitable to be used if somebody has access to a parallel processing system.

7. The group-hopscotch like the other schemes can be used for open regions (r,θ) geometry and the mixed derivative case without any substantial difficulties.

8. Although the new block hopscotch scheme is more beneficial in the case of linear problems, for non-linear cases such as the chemical-reaction problems given in §3.1, the point hopscotch method seems to be the most straightforward scheme and easiest to apply. However, research towards a new scheme for the non-linear case is underway.

9. For higher dimensional problems (e.g. 3.1) the scheme analogous to Group 4 (which is shown in Figure 6.3) becomes more advantageous especially if one solves the system of equations corresponding to the implicit molecule beforehand.

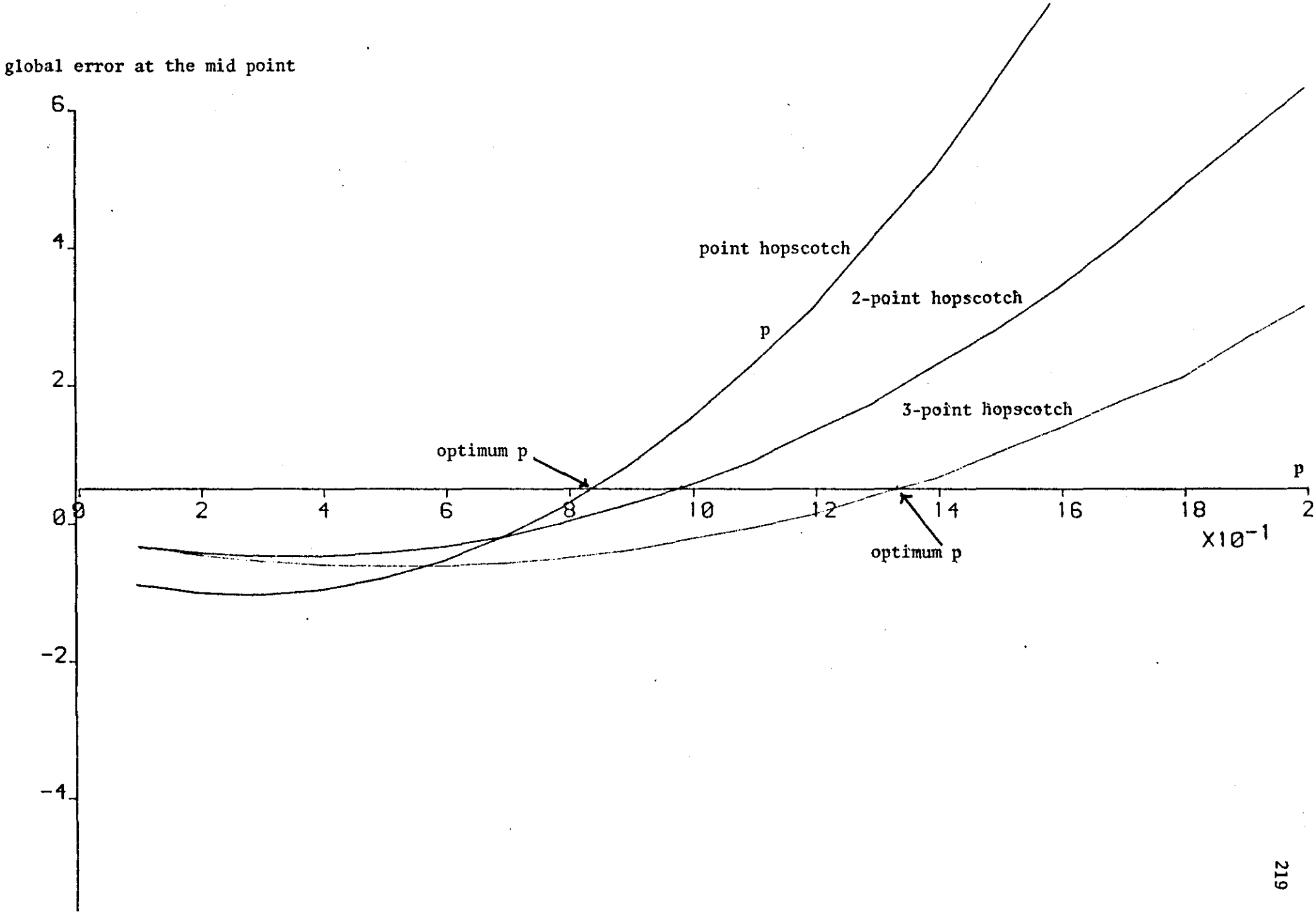
10. The results given in Tables 6.3,6.6 indicate that the composite hopscotch scheme is more accurate for p=2 and therefore is preferable.

Since in this case the solutions obtained are more than twice as accurate as the other results, henceforth the extra computational work required can be compromised by utilising longer time-steps.

11. In non-rectangular regions (e.g. the circle) one needs to use irregular groups and therefore the programming becomes more complicated as well as the accuracy being decreased. In this case the line hopscotch method seems to be superior.

12. Finally, for small time-step e.g. $\Delta T \leq \frac{1}{2}.\Delta x^2$ i.e. $p \leq \frac{1}{2}$, the point hopscotch method is recommended while for longer time-step in which $\frac{1}{2} < p \leq 5$ the composite hopscotch is the most accurate and efficient scheme.

For very long time-step (e.g. p>10) the hopscotch procedures are not recommended and one should apply some other technique (e.g. Morris, Extrapolation scheme).


global error at the mid point
6
4
2
0
-2
-4
point hopscotch
2-point hopscotch
p
3-point hopscotch
optimum p
optimum p
p
0
2
4
6
8
10
12
14
16
18
2
X10⁻¹

CHAPTER VII

NUMERICAL SOLUTION OF PARABOLIC EQUATIONS BY BOUNDARY VALUE TECHNIQUES

7.1 INTRODUCTION

The techniques that have been studied so far for the solution of parabolic equations were accomplished by means of *marching* or *step-by-step* procedures on an initial value problem.

In the case of elliptic equations however, *jury* or *boundary value* procedures are the natural ones to use. Here the consistent finite-difference equations constitute a system of algebraic equations in as many unknowns as there are interior mesh points in the region of interest. One obvious disadvantage which arises in this case is the computer storage requirements.

In previous time this difficulty has been treated by splitting the elliptic equations into a system of ordinary differential equations (e.g. the method of lines) and solving the problem. In recent years, the development of the high speed computer with large storage facility has led to the direct solution of such elliptic difference equations.

Recently, attention has been focussed on converting the parabolic equation to an equivalent elliptic form and to applying jury methods rather than the usual step-by-step procedures. This process can be described as the "Boundary Value Technique" and can be shown to be free from row-to-row error accumulation. This is an advantage if one is computing the solution for large times.

The motivation for this chapter can be found in the work of Greenspan, D. (1967), (1974), Carasso, A. (1968), and Carasso, A. & Parter, S.V. (1970).

7.2 THE BOUNDARY VALUE PROCEDURE

The problem to be considered here consists of a parabolic initial boundary value problem, which for simplicity is chosen in one space dimension to be

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(x,t,u,\frac{\partial u}{\partial x}) \qquad (7.2.1)$$

in a region $R=\{(x,t)\,|\,0\leq x\leq 1,\ t>0\}$, where the solution $u(x,t)$ of (7.2.1) attains a known steady-state value, $\tilde{u}(x)$ as $t\to\infty$ under some appropriate initial and boundary conditions which are given as follows:

$$\begin{aligned} u(x,0) &= f_0(x) \qquad 0\leq x\leq 1 \qquad t=0 \\ u(x,t) &= \phi_1(t) \qquad x=0, \qquad t>0 \qquad (7.2.2) \\ \text{and} \qquad u(x,t) &= \phi_2(t) \qquad x=1, \qquad t>0\ . \end{aligned}$$

Suppose for T sufficiently large, $u(x)$ is a reasonable approximation to the exact solution $u(x,T)$ at time $t=T$.

The alternative approach to the marching technique (as discussed in Chapter II and III) has been proposed by Greenspan (1967) as follows.

Consider a finite-difference approximation to the analytical problem (7.2.1)-(7.2.2) in the open rectangle $R_T=\{(x,t)\,|\,0<x<1,\ 0<t<T\}$. Interpret the finite-difference equations as a system of algebraic equations for the approximation to $u(x,t)$ at the interior mesh points of R_T and imagine solving this system, if possible, subject to the given initial and boundary conditions, and data $u=\tilde{u}(x)$ at $t=T_\infty$. Obviously one needs to be sure that the difference scheme is such that, the inclusion of the extra data at $t=T_\infty$ leads to a determinate system of algebraic equations. If the scheme is consistent with the differential equation, it is plausible that the solution of the finite difference equations would be an approximation to $u(x,t)$ at the mesh points (Carasso, A. 1968, p.3).

The finite-difference replacement selected by Greenspan, is the well-

known Richardson scheme which is described in Chapter II. As was shown previously, this scheme is unconditionally unstable, when used as a marching procedure. However, in a closed region as in the case here, the Richardson formula does not suffer from instability when used to solve a linear parabolic problem with time-independent coefficients.

At least one good reason why such a method may prove useful in practice, especially for large times is provided by its behaviour towards round-off error. The marching procedures tend to accumulate round-off error, whereas "*jury*" problems do not. (For proof, see Carasso, 1968, p.64).

Formulation of the Boundary Value Technique

The finite difference discretization of equation (7.2.1) by the scheme suggested by Greenspan will lead to the following formulation

$$\frac{U_{i,j+1}-U_{i,j-1}}{2\Delta T} = \frac{U_{i+1,j}-2U_{i,j}+U_{i-1,j}}{\Delta x^2} + f\left(x_i,t_j,U_{i,j},\frac{U_{i+1,j}-U_{i-1,j}}{2\Delta x}\right)$$

$$U_{i,0} = f_0(i,\Delta x) \tag{7.2.3}$$

$$\begin{aligned} U_{0,j} &= {}_1(j.\Delta t) \\ U_{N,j} &= {}_2(j,\Delta t) \end{aligned} \qquad i=1,2,\ldots,N-1, \quad j=1,2,\ldots,M-1, \tag{7.2.4}$$

where $\Delta x=1/N$ and $\Delta T=T/M$.

The approach suggested by Greenspan is in choosing M large and to solve the resulting $(M-1)\times(N-1)$ system of equations simultaneously. Indeed, the method was suggested for a general class of parabolic problems, linear and non-linear cases where several computational experiments were carried out.

The complete analysis of the method was given later by Carasso, A.(1968) where he discussed the convergence of the technique and evaluated the rate of convergence of the method for linear problems with time-independent coefficients to be $O(\Delta x^2+\Delta T^2)$ as $\Delta T\to 0$, $T\to\infty$ otherwise the rate of convergence reduced to $O(\Delta T^{3/2})$ which is also the case for mildly non-linear problems,

for sufficiently smooth exponentially decaying solutions.

Furthermore, he also considered the iterative solution of the problem and obtained the optimum value for the acceleration parameter. Finally, he gave the failure of the Boundary Value Technique to solve the parabolic equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \pi^2 u + \sin \pi x . \cos t \qquad 0<x<1,\ t>0$$

which had been attempted by Greenspan earlier.

The aim of this chapter is to apply various iterative procedures to solve some non-linear problems by this new strategy and compare the results with the Newton-iterative solution which was suggested by Greenspan. Specifically, we are interested in using the hopscotch techniques, block and point overrelaxation methods and compare the results obtained under these special circumstances.

7.3 ITERATIVE METHODS OF SOLUTION

We consider the equation (7.2.1)-(7.2.2) and rewrite the difference replacement in an iterative sequence as follows:

$$U^{n}_{i,j+1}-U^{n}_{i,j-1} = 2p(U^{n}_{i+1,j}-2U^{n+1}_{i,j}+U^{n}_{i-1,j})+2k.f(x,t,U^{n}_{i,j}, \frac{U^{n}_{i+1,j}-U^{n}_{i-1,j}}{2\Delta x}) \qquad (7.3.1)$$

for i=1,2,...,N-1, j=1,2,...,M-1.

where as before $p = \frac{\Delta t}{\Delta x^2}$

We now consider the case when f is a linear function and rewrite (7.3.1) using row-wise ordering of the mesh points in the region R such that for each point the totality of the difference equations produced in this way yields a (M-1)×(M-1) block linear system of the form

$$A\underline{U} = \underline{b} \qquad (7.3.2)$$

where

$$A = \begin{bmatrix} D_1 & F_1 & & & \\ E_1 & D_2 & F_2 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & F_{n-1} \\ 0 & & & E_{n-1} & D_n \end{bmatrix}_{(M-1)\times(M-1)} = D + E + F \qquad (7.3.3)$$

with

$$D_i = \begin{bmatrix} 4p & 1 & & & \\ -1 & 4p & 1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & & & 1 \\ 0 & & & -1 & 4p \end{bmatrix}_{(N-1)\times(N-1)} \qquad E_i = F_i = 2pI_{(N-1)\times(N-1)} \qquad (7.3.4)$$

where N=1/h and M=T/k, $p=k/h^2$,

and $\underline{b}$ is obtained by inserting the known boundary values when applied into equation (7.3.3). (Evans, D.J., 1980, internal report).

When f is non-linear, the equation (7.3.4) becomes a non-linear system of equations which can be solved by alternative techniques such as the Newton method.

From the block tridiagonal structure of A, we can consider the block iterative schemes for the solution of (7.3.2) in which each block of unknowns consist of all the points $U_{i,j}$ in a row of the grid. Then, the block simultaneous displacement method is given by

$$D\underline{U}^{(n+1)} = -(E+F)\underline{U}^{(n)} + \underline{b} \tag{7.3.5}$$

and the block successive overrelaxation method by

$$(D+\omega E)\underline{U}^{(n+1)} = -[\omega F+(\omega-1)D]\underline{U}^{(n)} + \omega\underline{b} \tag{7.3.6}$$

or

$$\underline{U}^{(n+1)} = -(D+\omega E)^{-1}[\omega F+(\omega-1)D]\underline{U}^{(n)} + \omega(D+\omega E)^{-1}\underline{b}$$

where the subscript n denotes the iteration cycle and ω the block overrelaxation parameter.

For the convergence of the block simultaneous displacement method we require $\rho(-D^{-1}(E+F))<1$. From the tridiagonal structure of A, it can be shown that

$$\rho(D^{-1}(E+F)) = \frac{2\rho\cos\frac{\pi}{N}(2\rho-i\cos\frac{\pi}{M})}{(4\rho^2+\cos^2\frac{\pi}{M})} \quad , i=\sqrt{-1} \quad , \tag{7.3.7}$$

and hence $|\rho|<1$. From the S.O.R. theory (Young, 1971) we can find the optimum overrelaxation parameter ω_{opt} so that $\rho(-(D+\omega E)^{-1}[\omega F+(\omega-1)D])$ is minimised. According to the theorem given by Young (1971), we have

$$\omega_{opt} = 2/\{1+\sqrt{1-(\gamma^2-\delta^2)}\} \tag{7.3.8}$$

where $\lambda=\gamma+i\delta$ are the eigenvalues of $(-D^{-1}(E+F))$ contained in an elliptical region.

7.4 THE HOPSCOTCH FORMULATION OF THE BOUNDARY VALUE TECHNIQUE

We now consider the equation (7.2.1) and the hopscotch algorithm to obtain the numerical solution.

The fully explicit and implicit form of equation (7.3.1) which leads to the hopscotch algorithm can be easily verified to be

$$U_{i,j}^{n+1} = (U_{i+1,j}^{n}+U_{i-1,j}^{n})/2-(U_{i,j+1}^{n}-U_{i,j-1}^{n})/4p + h^2.f(x_i,t_j,U_{i,j}^{n}, \frac{U_{i+1,j}^{n}-U_{i+1,j}^{n}}{2\Delta x}) \qquad (7.4.1)$$

$$U_{i,j}^{n+1} = (U_{i+1,j}^{n+1}+U_{i-1,j}^{n+1})/2-(U_{i,j+1}^{n+1}-U_{i,j-1}^{n+1})/4p + h^2.f(x_i,t_j,U_{i,j}^{n}, \frac{U_{i+1,j}^{n+1}-U_{i-1,j}^{n+1}}{2\Delta x}) \qquad (7.4.2)$$

or in the compact form as

$$\underline{U}^{n+1} = H\underline{U}^{n} + \underline{b} \qquad (7.4.3)$$

$$(I-H)\underline{U}^{n+1} = \underline{b} \qquad (7.4.4)$$

where

$$H = \begin{bmatrix} D & -E & & \\ E & \ddots & \ddots & 0 \\ & \ddots & \ddots & -E \\ 0 & & E & D \end{bmatrix}_{(M-1)\times(M-1)},$$

$$D = \begin{bmatrix} 0 & 1/2 & & \\ 1/2 & \ddots & \ddots & 0 \\ & \ddots & \ddots & 1/2 \\ 0 & & 1/2 & 0 \end{bmatrix}_{(N-1)\times(N-1)} \qquad E = \begin{bmatrix} 1/4p & & & \\ & 1/4p & & 0 \\ & & \ddots & \\ 0 & & & 1/4p \end{bmatrix}_{(N-1)\times(N-1)}$$

and

$$\underline{b} = h^2 \begin{bmatrix} f^n_{i,j} \\ \vdots \\ f^n_{i,j} \end{bmatrix} + \text{the boundary values.}$$

To write the hopscotch formulation, we rewrite the equations (7.4.3)-(7.4.4) over 2m steps and obtain the equations

$$(I-I_2H)\underline{U}^{2n+1} = I_1H\underline{U}^{2n}+(I_1\underline{b}^{2n}+I_2\underline{b}^{2n+1}) \tag{7.4.5}$$

$$(I-I_1H)\underline{U}^{2n+2} = I_2H\underline{U}^{2n+1}+(I_2\underline{b}^{2n+1}+I_1\underline{b}^{2n+2}) \tag{7.4.6}$$

and therefore

$$\underline{U}^{(2n+2)} = T\,\underline{U}^{2n}+\underline{g}^{2n+2} \tag{7.4.7}$$

where

$$T = (I-I_1H)^{-1}I_2H(I-I_2H)^{-1}I_1H)$$

I is the unitary matrix of order H, I_1 and I_2 are matrices with entries of 0 and 1 such that

$$I_1+I_2 = I \text{ and } I_1I_2 = 0$$

Let

$$\tilde{T} = (I-I_1H)T(I-I_1H)^{-1}$$

therefore $\tilde{T} = [I_2H(I-I_2H)^{-1}][I_1H(I-I_1H)^{-1}]$. (7.4.8)

For convergence of the hopscotch method we require $||\tilde{T}|| \leq 1$ for some suitable norm.

To investigate the convergence of the algorithm we define new matrices such that

$$C_1 = \tfrac{1}{2}I-I_1H \quad \text{and} \quad C_2 = \tfrac{1}{2}I-I_2H$$

hence

$$I_1H = \tfrac{1}{2}(I-2C_1) \quad \text{and} \quad I_2H = \tfrac{1}{2}(I-2C_2)$$

and $I-I_1H = \tfrac{1}{2}I+C_1 = \tfrac{1}{2}(I+2C_1)$, $I-I_2H = \tfrac{1}{2}I+C_2 = \tfrac{1}{2}(I_2+2C_2)$.

Therefore, $\tilde{T} = [(I-2C_2)(I+2C_2)^{-1}][(I-2C_1)(I+2C_1)^{-1}]$ (7.4.9)

By the application of Kellog's 2nd lemma we indicate that,

$$||(I-C_2)(I+C_2)^{-1}||_2 \leq 1, \quad ||(I-C_1)(I+C_1)^{-1}||_2 \leq 1$$

provided $(C_2+C_2^*)$ and $(C_1+C_1^*)$ are positive definite.

However, $(C_1+C_1^*)$ can be shown to be positive definite in the following manner,

$$C_1+C_1^* = \tfrac{1}{2}I-I_1H+\tfrac{1}{2}I-I_1H^* = I-I_1(H+H^*) \qquad (7.4.10)$$

and

$$H+H^* = \begin{bmatrix} 2D & & 0 \\ & \ddots & \\ 0 & & 2D \end{bmatrix} \text{ where D is given in (7.4.4)}$$

and therefore,

$$Q = I-I_1(H+H^*) = \begin{bmatrix} V & & 0 \\ & \ddots & \\ 0 & & V \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} 1 & -1 & & & 0 \\ 0 & 1 & 0 & & \\ & -1 & 1 & -1 & \ddots \\ & & \ddots & \ddots & -1 \\ 0 & & & 0 & 1 \end{bmatrix} \qquad (7.4.11)$$

To show that the block diagonal matrix Q is positive definite, we have to show that the matrix V is positive definite.

Let $V_2=\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$ and by the theorem (1.3) we have

$$\lambda_s(V_2) = 1 > 0 \ , \ s=1,2, \qquad (7.4.12)$$

and hence V_2 is positive definite. Now we apply the induction formula since if V_i is positive definite and so is V_{i+1}.

Define

$$V_{i+1} = \left[\begin{array}{c|c} V_i & \begin{matrix} 0 \\ 0 \end{matrix} \\ \hline \begin{matrix} & -1 \end{matrix} & 1 \end{array}\right] \quad \text{or} = \left[\begin{array}{c|c} V_i & -1 \\ \hline \begin{matrix} & 0 \end{matrix} & 1 \end{array}\right] \ , \qquad (7.4.13)$$

thus

$$\lambda\,(V_n) = (1-\lambda_n).\lambda_s(V_{n-1}) > 0,\ s=1,2,\ldots,n-1. \qquad (7.4.14)$$

Therefore, Q is positive definite and the condition $||T||_2 \leq 1$ is satisfied.

Consequently, the hopscotch iteration using the Boundary Value Technique for the equation (7.2.1) converges.

The convergence of the block hopscotch techniques can also be deduced by application of theorem (3.2).

7.5 NUMERICAL EXPERIMENTS

i) As a numerical example for the one-dimensional case, we consider the non-homogeneous, non-linear, Burger's equation,

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} - u\frac{\partial u}{\partial x} - xe^{-t}(1-e^{-t}) , \qquad (7.5.1)$$

subject to the initial and boundary conditions,

$$\begin{aligned} u(x,0) &= x , \qquad 0\leq x\leq 1, \quad t=0 \\ u(0,t) &= 0 , \qquad\qquad\qquad\qquad (7.5.2) \\ \text{and} \qquad u(1,t) &= e^{-t}, \qquad t>0. \end{aligned}$$

This example is solved by Greenspan (1974, p.144) where he applied the generalized Newton method to solve the set of non-linear equations obtained from the finite-difference replacement over the interior mesh points. The exact solution of (7.5.1)-(7.5.2) is known to be $u(x,t)=xe^{-t}$ and therefore the accuracy of any approximate solution can be easily deduced.

Here we have applied various iterative methods including the hopscotch schemes to solve this example by the Boundary Value technique. Similar to Greenspan, we have chosen $\Delta x=\Delta T=1/10$ and the boundary values at $T=10$ is set to zero.

The iterative sequence which is considered here has the following form:

$$U^n_{i,j+1}-U^n_{i,j-1} = \frac{2\Delta T}{\Delta x^2}(U^n_{i+1,j}-2U^{n+1}_{i,j}+U^n_{i-1,j})-\frac{\Delta T}{\Delta x}(U^n_{i+1,j}-U^n_{i-1,j})U^{n+1}_{i,j}-2\Delta T f_{i,j}$$

$$i=1,2,\ldots,N-1 \qquad (7.5.3)$$
$$j=1,2,\ldots,M-1$$

where $f_{i,j}=x_i e^{-t_j}(1-e^{-t_j})$. and n is the iteration number

The sequence (7.5.3) is applied in an appropriate form for block hopscotch as well as the block S.O.R. methods. The numerical results displayed in Table (7.1) show the iteration number, the accuracy together with the time taken by the CDC 7600 to obtain the solution by different methods.

$\Delta x=\Delta T=0.1$, $T=10$, $p=10$, $\varepsilon=5\times 10^{-5}$

Iterative Methods	Accel. parameter	No. of iteration	Maximum Error	Time in sec.	Time Ratio
Point Hopscotch	1	129	6.4485&-4	6.418	9.2
A.D. Hopscotch	1	126	7.1740&-4	7.882	11.9
Peripheral Hopscotch	1	126	5.6702&-4	11.885	18.0
Group 4 Hopscotch	1.1	72	2.6583&-4	3.665	5.5
Line Hopscotch	1	66	8.0439&-4	4.371	6.6
Group 9 Hopscotch	1.2	50	1.9911&-4	12.663	19.2
Peripheral S.O.R.	1.38	29	1.7638&-4	4.078	6.2
Point S.O.R.	1.50	23	1.4697&-4	1.159	1.8
Line (Column) S.O.R.	1.50	24	1.5588&-4	1.970	3.0
Group 4 S.O.R.	1.40	18	1.1181&-4	0.913	1.4
Group 9 S.O.R.	1.30	15	6.8420&-4	6.890	10.4
Line (Row) S.O.R.	0.78	7	8.0231&-5	0.661	1
Two Line (Row) S.O.R.	0.8	8	1.1897&-2	15.212	23.0
Newton	1.3	8	2.7529&-2	22.374	33.9

TABLE 7.1

Conclusive Remarks

The results given in Table (7.1) indicate that:

1. Generally speaking, the S.O.R. methods are faster than the hopscotch iteration schemes for solving problems by using boundary value techniques.
2. The group strategy is more beneficial in the case of the hopscotch techniques regarding the number of iterations involved. However, the computational time is increased significantly for the Group 9 schemes since a system of 9-equations has to be solved for every group calling two ALGOL procedures Gauss-band which carries out the Gaussian elimination for the system and Solve-band to solve the decomposed form of the equations.
3. The Group 4 hopscotch is the most economical method amongst the hopscotch procedures.
4. The Line (row) S.O.R. (SLOR) requires the same number of iterations as the Newton method, while it is ≈34 times faster when the computational times are compared and ≈5.5 times faster than the Group 4 hopscotch scheme.

ii) Two space dimensional example: Navier-Stokes Equation

State of the Problem

The 2nd example for the boundary-value technique considered here is the two dimensional steady-state, viscous, incompressible flow in a rectangular cavity.

Consider a square cavity DABC as shown in Figure 7.5.1 within which a steady fluid motion is generated by sliding an infinitely long plate lying on top of the cavity. Suppose that all the variables are normalised so

that the size of the cavity is the unit square and the sliding velocity is -1 (Greenspan, 1974, p.208) in the negative x-direction.

Let S be the square ABCD and denote its interior by R. On R the equation of motion to be satisfied are the two dimensional steady-state Navier-Stokes equations, namely:

$$\Delta\omega + R\left(\frac{\partial\psi}{\partial x} \cdot \frac{\partial\omega}{\partial y} - \frac{\partial\psi}{\partial y} \cdot \frac{\partial\omega}{\partial x}\right) = 0, \quad R \geq 0 \tag{7.5.4}$$

$$\Delta\psi = -\omega \tag{7.5.5}$$

where ψ is the stream function, ω is the vorticity, and R is a non-negative constant called the Reynolds number.

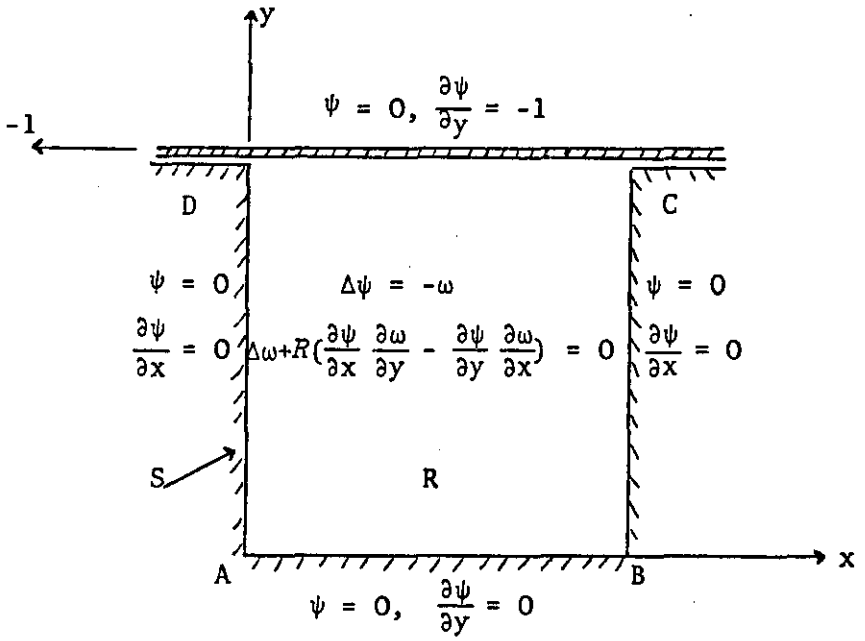


FIGURE 7.5.1: Cavity flow caused by a moving plate

If there is no fluid squeezed out of the cavity below the moving plate, the fluid motion forms closed paths within the cavity. The surfaces DA,AB,BC and CD are then segments of the bounding stream lines designated by ψ=0, along which can be specified that the velocities normal to these four surfaces are zero. On the other hand, we require that the tangential velocity vanishes on all the surfaces except the top plate

which is given as -1. Thus we obtain four additional boundary conditions $\frac{\partial\psi}{\partial x}=0$ on DA and CB, $\frac{\partial\psi}{\partial y}=0$ on AB and $\frac{\partial\psi}{\partial y}=-1$ on CD.

The vorticity equation (7.5.4) can also be regarded as the asymptotic time limit of the non-steady, time dependent equations and the system (7.5.3)-(7.5.4) now becomes:

$$\Delta\psi = -\omega \tag{7.5.6}$$

$$\frac{\partial\omega}{\partial t} = \frac{1}{R}\Delta\omega + \frac{\partial\psi}{\partial x}\cdot\frac{\partial\omega}{\partial y} - \frac{\partial\psi}{\partial y}\cdot\frac{\partial\omega}{\partial x}, \tag{7.5.7}$$

$$0<x<1,\ 0<y<1 \text{ and } t>0$$

which is elliptic-parabolic system. This new formulation can be solved by the marching procedures as well as by the boundary value technique which is the scope of this subsection. In many problems of practical concern, only the steady-state solution is the subject of interest.

Finite Difference Solution

Numerical solutions of the equations (7.5.6)-(7.5.7) have been the subject of many investigations. The study has been attempted to increase the understanding of optimum solution techniques. Because of its simplicity and economy a boundary value technique for problem (7.5.6)-(7.5.7) has been employed by Greenspan (1974) where the generalized Newton iteration scheme has been applied to solve the system of non-linear equations corresponding to the non-linear vorticity equation (7.5.7).

Other methods also have been attempted by Greenspan which are fully described in his book (1974).

The hopscotch technique for the solution of the non-steady parabolic equation together with the Bunemann Direct Method (BDM) and SOR methods for the stream-function equation have been applied by Smith, R.E. & Kidd,A.,

where the comparison of the computational time involved is made with the A.D.I. scheme.

By the experience gained from the one-dimensional Burger's equation, we now concentrate our attention on the boundary value technique when the relaxation schemes are used and when the Reynolds number is chosen to be the specific values 100 and 500. The method proceeds as follows.

Let D be the rectangular parallelpiped defined by $D=\{(x,y,t):0\leq x\leq 1, 0\leq y\leq 1\ 0\leq t\leq T\}$. Define R to be the interior and S to be the boundary of D. Using space grid size $\Delta x=\Delta y=h$ and time size $\Delta T=k$, we construct in the usual way the three dimensional sets of interior grid points, denoted by $R_{h,k}$ and boundary grid points, denoted by $S_{h,k}$.

We observe that (7.5.6)-(7.5.7) is a coupled system of partial differential equations in ψ and ω. But if ω is known (7.5.6) is a linear elliptic equation in ψ, while if ψ is known, (7.5.7) is a linear parabolic equation in ω. Thus, the initial guesses $\psi^{(0)}$ and $\omega^{(0)}$ can be applied and we construct a sequence of iterative solutions as follows:

Use $\omega^{(0)}$ in (7.5.6) to produce $\psi^{(1)}$ by solving the parabolic equation (7.5.7) in D. We shall remember that, the solution at t=T is chosen to be $\omega=0$ (the steady-state solution which is the essence of the technique). Therefore $\omega^{(1)}$ can be inserted in (7.5.5) to produce $\psi^{(2)}$ and then using $\psi^{(2)}$ we find $\omega^{(2)}$ and so on. In this way, we construct a sequence of discrete functions

$$\psi^{(1)},\psi^{(2)},\ldots,\psi^{(n)} \tag{7.5.8}$$

on $R_{h,\Delta t}$ and a sequence of discrete functions

$$\omega^{(1)},\omega^{(2)},\ldots,\omega^{(n)} \tag{7.5.9}$$

on $R_{h,\Delta T}+S_{h,\Delta T}$ which will both converge (Greenspan, 1974, p.231). For this purpose at each point of $R_{h,\Delta T}$ we obtain the difference equation

$$\psi(x+h,y,t_k)+\psi(x-h,y,t_k)+\psi(x,y+h,t_k)+\psi(x,y-h,t_k)-4\psi(x,y,t_k) =$$

$$-h^2\omega^{(1)}(x,y,t_k),\ k=1,2,\ldots,m-1. \qquad (7.5.10)$$

At the boundary grid points we set,

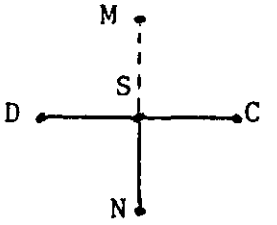
$$\psi(x,y,0) = 0, \quad \psi(x,y,T) = \psi(x,y,\infty) ,$$

$$\psi(0,y,t) = \psi(x,0,t) = \psi(x,1,t) = \psi(1,y,t) = 0. \qquad (7.5.11)$$

The iterative solution $\psi(x,y,t_k)$ can now be obtained at every step by inserting (7.5.11) whenever necessary. Here we apply the SOR and SLOR iterative techniques to solve the stream equation.

Next step is generating the sequence $\omega^{(n)}$ on $R_{h,\Delta T}$ which requires the values of ω on S_h. These boundary conditions on the vorticity equation can now be obtained by central differencing equation (7.5.6), applying the boundary conditions for the stream function and by enforcing the reflection condition at the boundary. They are as follows:

at the surface DC (Figure 7.5.1),



$$\frac{\psi_M-2\psi_S+\psi_N}{\Delta y^2} = -\omega_S ,$$

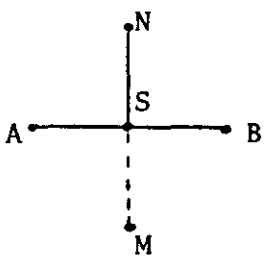
$$\psi_S = 0 ,$$

$$\frac{\psi_M-\psi_N}{2\Delta y} = -1.$$

Combining these conditions results in

$$\frac{2\psi_N-2\Delta y}{(\Delta y)^2} = -\omega_S . \qquad (7.5.12)$$

At the surface AB we have



$$\frac{\psi_M-2\psi_S+\psi_N}{\Delta y^2} = -\omega_S ,$$

$$\psi_S = 0 ,$$

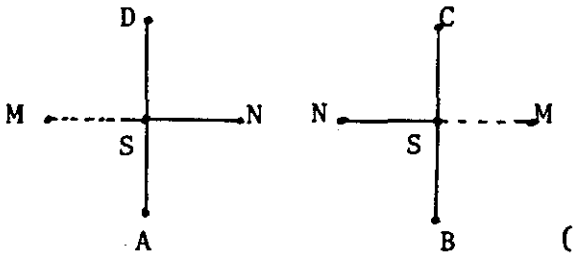
$$\frac{\psi_M-\psi_N}{2\Delta y} = 0 \text{ or } \psi_M = \psi_N ,$$

or
$$\frac{2\psi_N}{\Delta y^2} = -\omega_S \ . \tag{7.5.13}$$

At the surface AD and BC we can write

$$\frac{\psi_M - 2\psi_S + \psi_N}{\Delta x^2} = -\omega_S \ ,$$

$$\psi_S = 0 \text{ and } \psi_M = \psi_N \ ,$$



or
$$\frac{2\psi_N}{\Delta x^2} = -\omega_S \ . \tag{7.5.14}$$

Now we can obtain the finite-difference replacement of the vorticity equation and insert the approximate boundaries (7.5.12)-(7.5.14) to generate the iterative solution $\omega^{(n)}$.

We apply the Richardson formula to solve the vorticity equation, i.e.,

$$\frac{\omega_{i,j}^{k+1} - \omega_{i,j}^{k-1}}{2\Delta T} = \frac{1}{R}\left[\frac{\omega_{i+1,j}^{k} - 2\omega_{i,j}^{k} + \omega_{i-1,j}^{k}}{\Delta x^2} + \frac{\omega_{i,j+1}^{k} - 2\omega_{i,j}^{k} + \omega_{i,j-1}^{k}}{\Delta y^2}\right] +$$
$$\frac{\psi_{i+1,j}^{k(1)} - \psi_{i-1,j}^{k(1)}}{2\Delta x} \cdot \frac{\omega_{i,j+1}^{k} - \omega_{i,j}^{k}}{\Delta y} - \frac{\psi_{i,j+1}^{k(1)} - \psi_{i,j-1}^{k(1)}}{2\Delta y} \cdot \frac{\omega_{i,j}^{k} - \omega_{i-1,j}^{k}}{\Delta x} \ . \tag{7.5.15}$$

The equation (7.5.15) can be solved for $\omega_{i,j}^{k}$ at all the interior grid points of $R_{h,\Delta T} + S_{h,\Delta T}$ by any suitable method (i.e. SOR).

Here we shall apply the SOR methods for the solution of (7.5.15) and compare the efficiency of relaxation methods with some other techniques applied by Greenspan (1974) and Smith & Kidd (197).

"Upwind" Difference Equations

The presence of advective (or convective) terms in the vorticity equation causes some difficulties when one solves the iterative scheme such as (7.5.15).

If the terms $\frac{\partial\omega}{\partial y}$ and $\frac{\partial\omega}{\partial x}$ are replaced by central differences, some of the points $\omega(x\pm h,y,t)$ and $\omega(x,y\pm h,t)$ are outside $R_{h,\Delta T}$. To remedy this difficulty and avoid outside points, central differences must be replaced by one-sided differences. Now as can be seen from (7.15.15) that the terms $\omega^k_{i,j}$ which arise from $\frac{\partial\omega}{\partial y}$ and $\frac{\partial\omega}{\partial x}$ play a different role for the iteration scheme, since they may appear with different sign with the similar term obtained from $\Delta^2\omega$ and therefore the diagonal dominancy of the system may be reduced or even lost, which leads to the divergence of the iterative scheme. Therefore we choose the forward or backward difference replacement as follows:

if

$$A^k_{i,j} = \frac{\psi^k_{i+1,j}-\psi^k_{i-1,j}}{2\Delta x} \gtrless 0 \text{ then } \frac{\partial\omega}{\partial y} \to \frac{\omega^k_{i,j+1}-\omega^k_{i,j}}{\Delta y} \text{ else } \frac{\partial\omega}{\partial y} \to \frac{\omega^k_{i,j}-\omega^k_{i-1,j}}{\Delta y} \quad (7.5.16)$$

and if

$$B^k_{i,j} = \frac{\psi^k_{i,j+1}-\psi^k_{i,j-1}}{2\Delta y} \gtrless 0 \text{ then } \frac{\partial\omega}{\partial x} \to \frac{\omega^k_{i,j}-\omega^k_{i-1,j}}{\Delta x} \text{ else } \frac{\partial\omega}{\partial x} \to \frac{\omega^k_{i+1,j}-\omega^k_{i,j}}{\Delta x} \quad (7.5.17)$$

Thus, the term $\frac{\partial\psi}{\partial x} \cdot \frac{\partial\omega}{\partial y} - \frac{\partial\psi}{\partial y} \cdot \frac{\partial\omega}{\partial x}$ can assume four different forms.
It is said that in (7.5.15) one always employs "upwind differences" (Forsythe & Wasow, 1960, p.398).

Matrix Formulation

The line Gauss-Sidel form of (7.5.15) (from which the SLOR is obtained) is as follows:

$$T\underline{\omega}^{n+1}_j = \underline{F}^n_j \ , \ j=1,2,\ldots,N-1, \quad (7.5.18)$$

where

$$T = \begin{bmatrix} \alpha_{1,j} & -2p_1 & & & \\ -2p_1 & \alpha_{2j} & -2p_1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & -2p_1 \\ 0 & & & -2p_1 & \alpha_{N-1,j} \end{bmatrix}, \quad p_1=\frac{\Delta T}{\Delta x^2 R}, \quad \alpha_{i,j}=8p_1+\frac{\Delta T}{\Delta x}(|A^k_{i,j}|+|B^k_{i,j}|).$$

and

$$F^k_{i,j} = -(\omega^k_{i,j}-\omega^{k-1}_{i,j})+2p_1(\omega^k_{i,j+1}+\omega^k_{i,j-1})-\frac{\Delta T}{\Delta x}(A^k_{i,j},\omega^k_{i,j-1}-B^k_{i,j},\omega^k_{i-1,j}).$$

for j,i=1,2,...,N-1 and k=1,2,...,M-1.

Thus, the SLOR method can be found by splitting up the matrix T in the same manner as given in (7.3).

Numerical Results

We have solved the system (7.5.4)-(7.5.5) by the SOR and SLOR methods for values of R=100 and 500 with the space and time increments chosen to be h=1/20 and k=1/5 respectively. We have also chosen T_∞=5 and set the boundary values at this level equal to zero.

In the following table the number of iterations to obtain an accuracy of 10^{-3} together with the time taken (by the CDC 7600) for each case are given. The initial guesses for both systems were chosen to be zero.

Reynolds Number	Iterative Scheme used for		Accel.parameter		No.of Iterations	Time in Secs.
	Stream Func.	Vorticity Equ.	Stream Func.	Vorticity Equ.		
100	SOR	SOR	0.3	0.9	134	60.7
500	SOR	SOR	0.24	0.33	252	113.9
100	*SLOR	SLOR	0.2	1.6	73	47.2
500	*SLOR	SLOR	0.1	1	101	65.25

TABLE 7.2

As can be seen from the table, the values of the acceleration parameter were all less than one and consequently the iterative methods are of an under-relaxation type.

The graphical schematic for each case of the table 7.2 are presented in Figures (7.5.1)-(7.5.9).

*For the SLOR schemes the replacement given in p.250 is used for $\frac{\partial \omega}{\partial t}$.

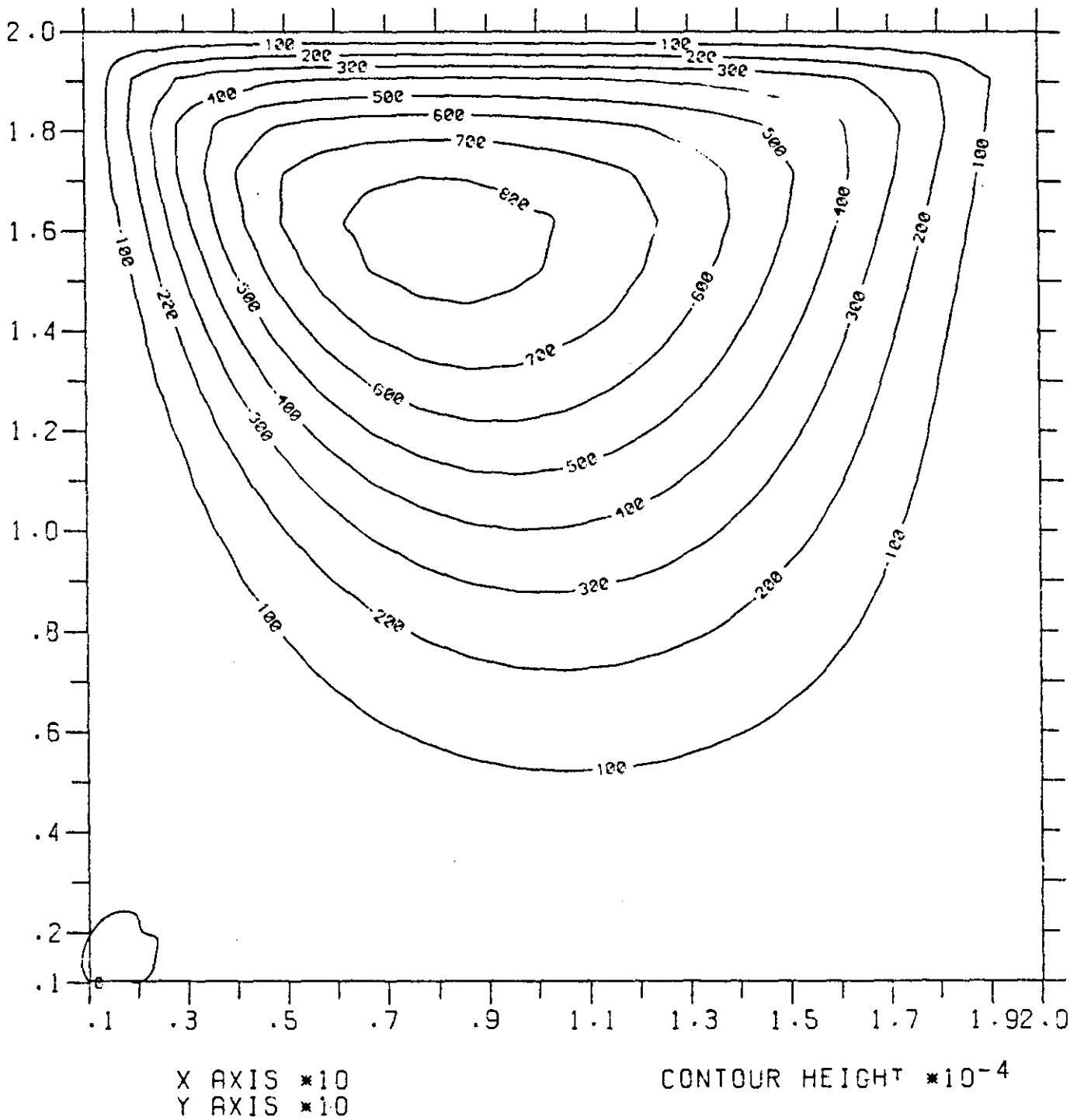


FIGURE 7.5.1

Stream Lines for R=100 by SOR Technique

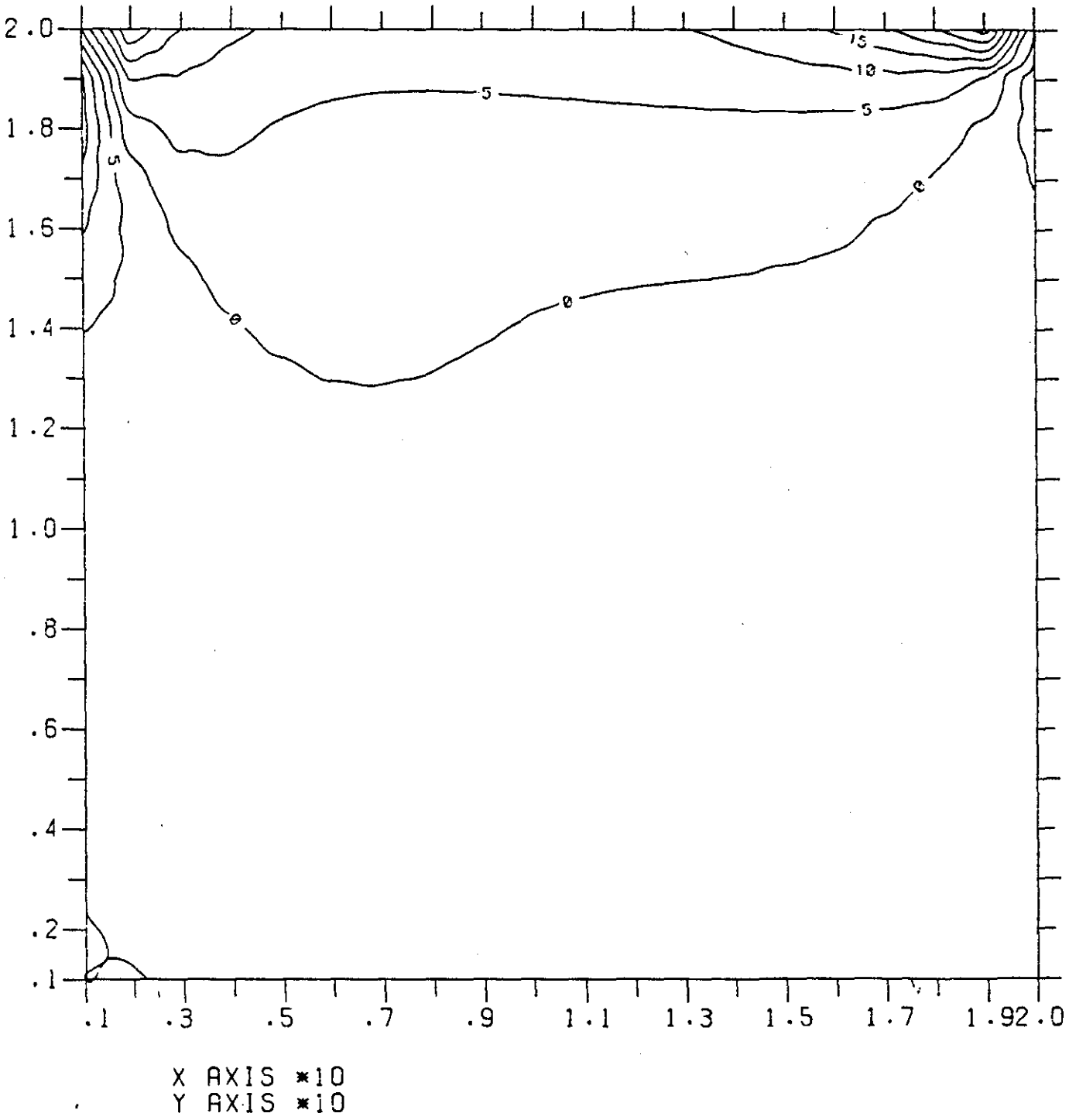


FIGURE 7.5.2

Vorticity curves for R=100 by SOR Technique

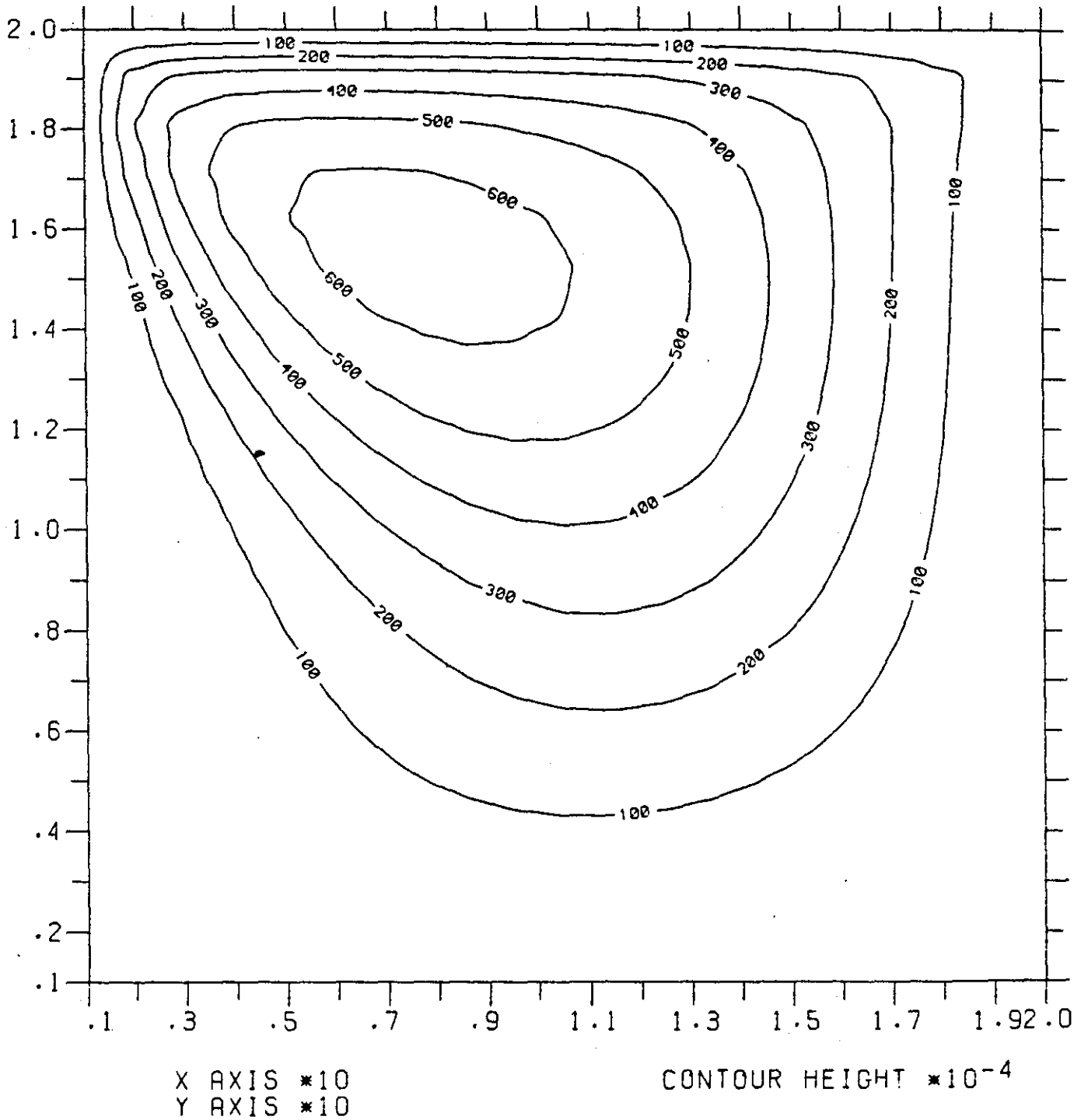


FIGURE 7.5.3

Stream Lines for R=500 by SOR Technique

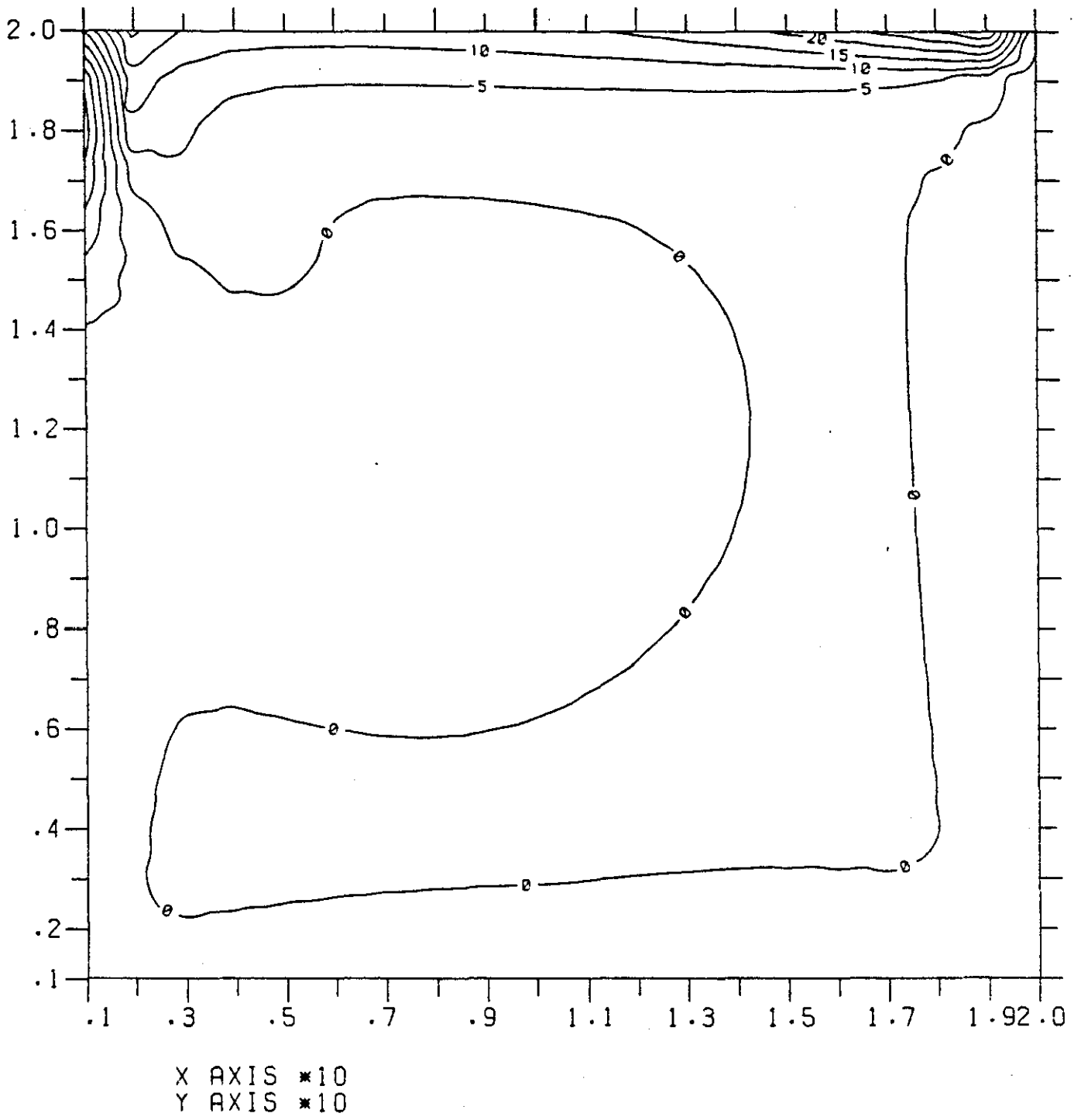


FIGURE 7.5.4

Vorticity curves for R=500 by SOR Technique

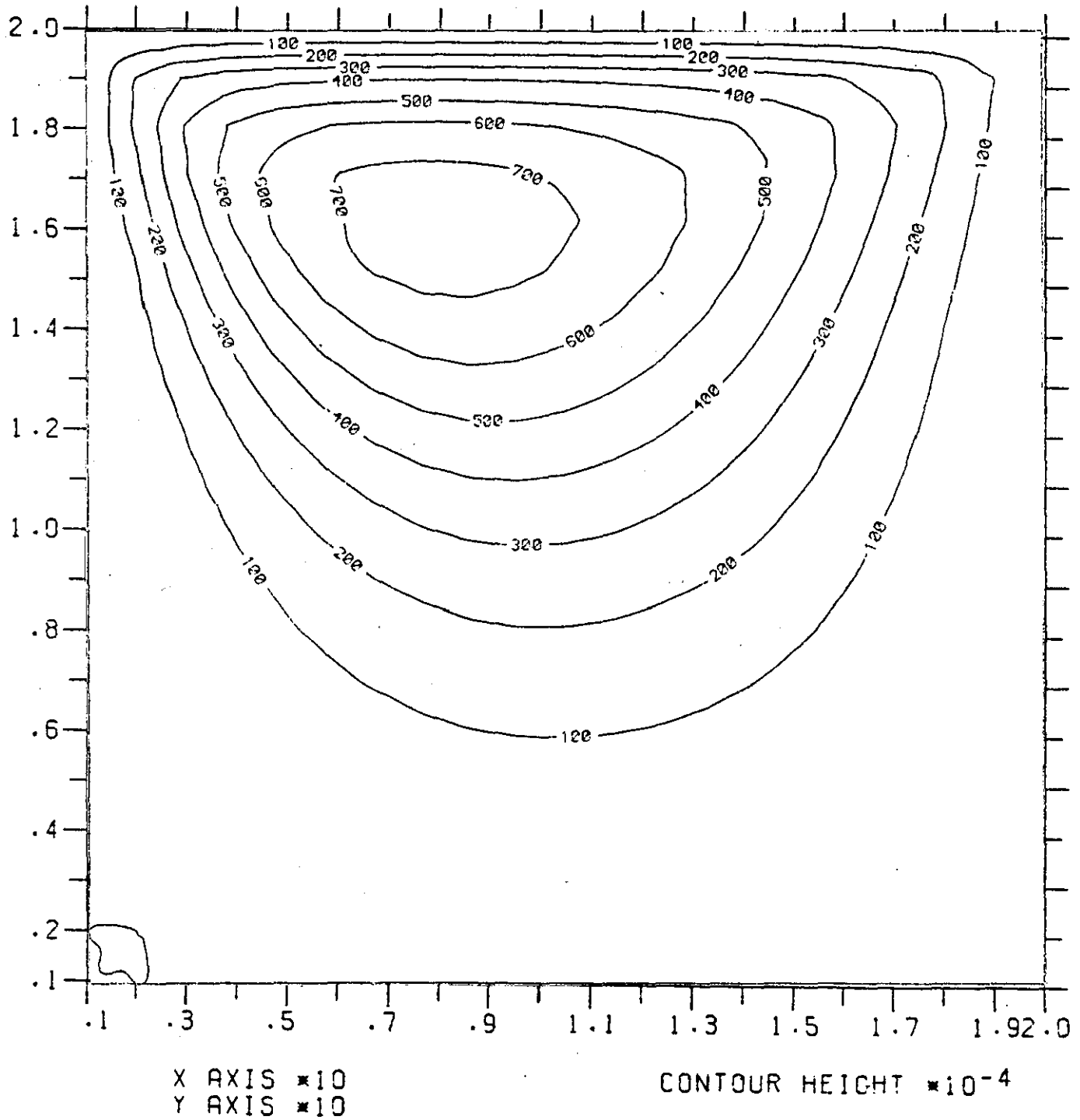


FIGURE 7.5.5

Stream Lines for R=100 by SLOR Technique

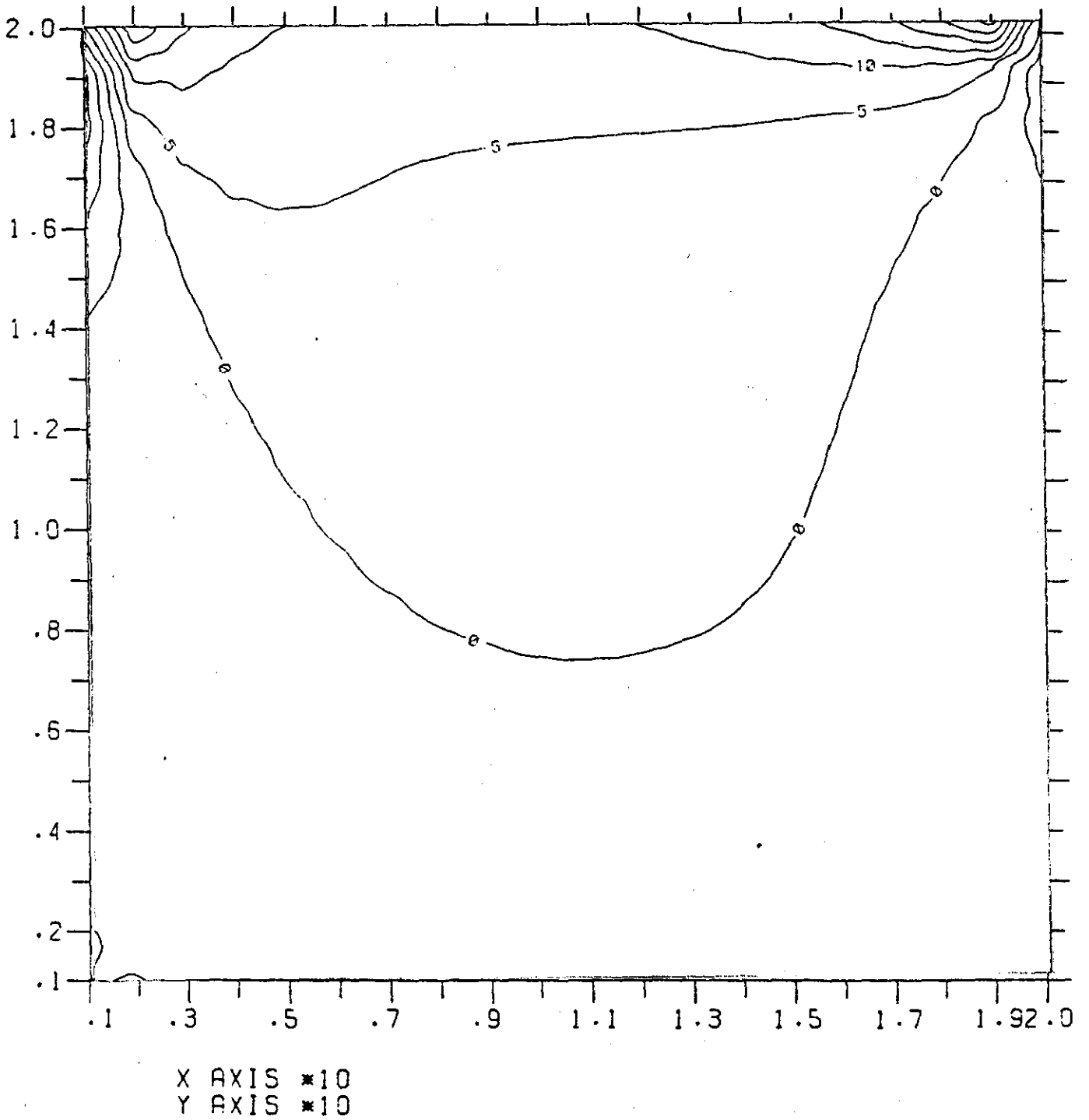


FIGURE 7.5.6

Vorticity Curves for R=100 by SLOR Technique

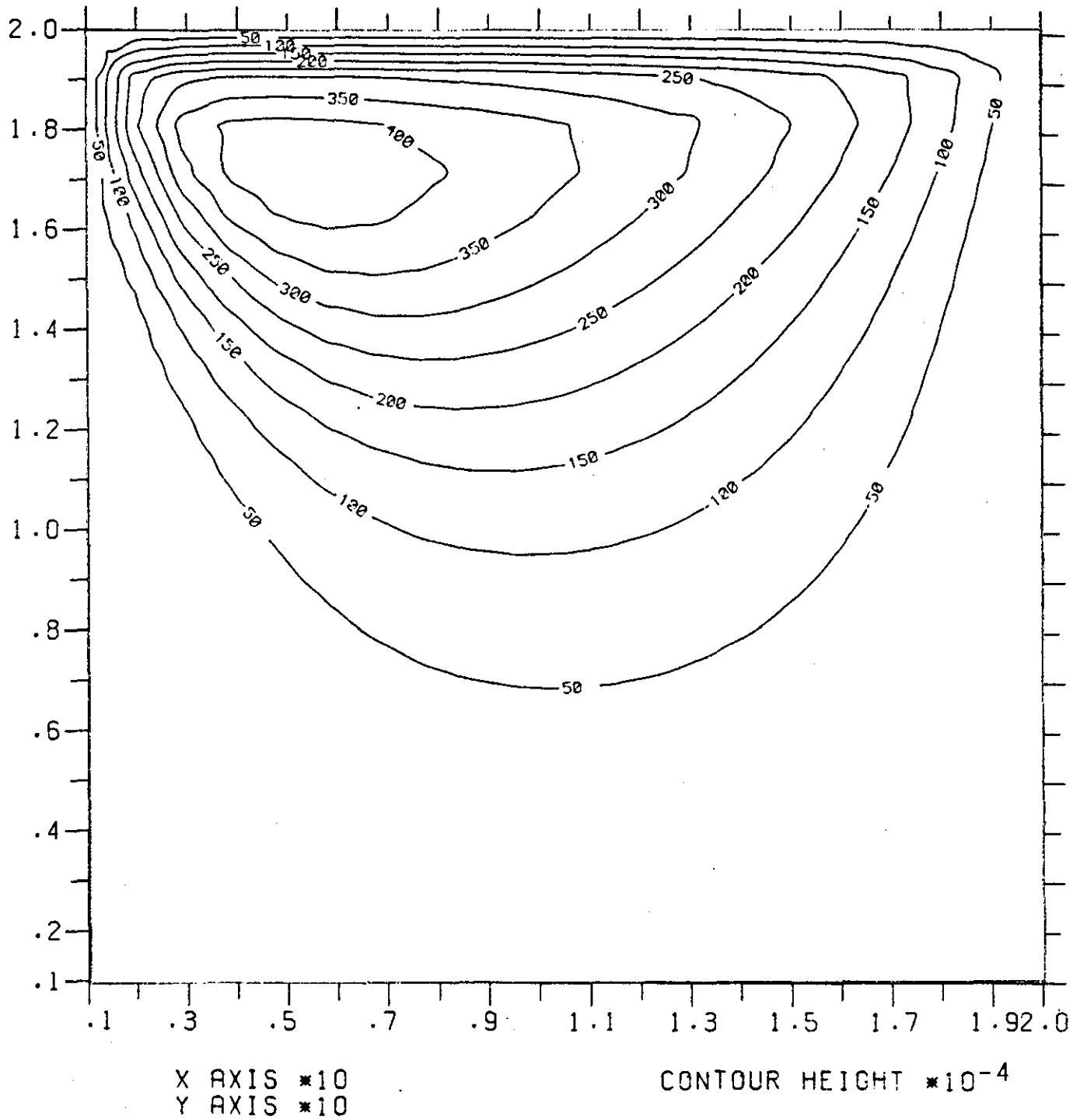


FIGURE 7.5.7

Stream Lines for R=500 by SLOR Technique

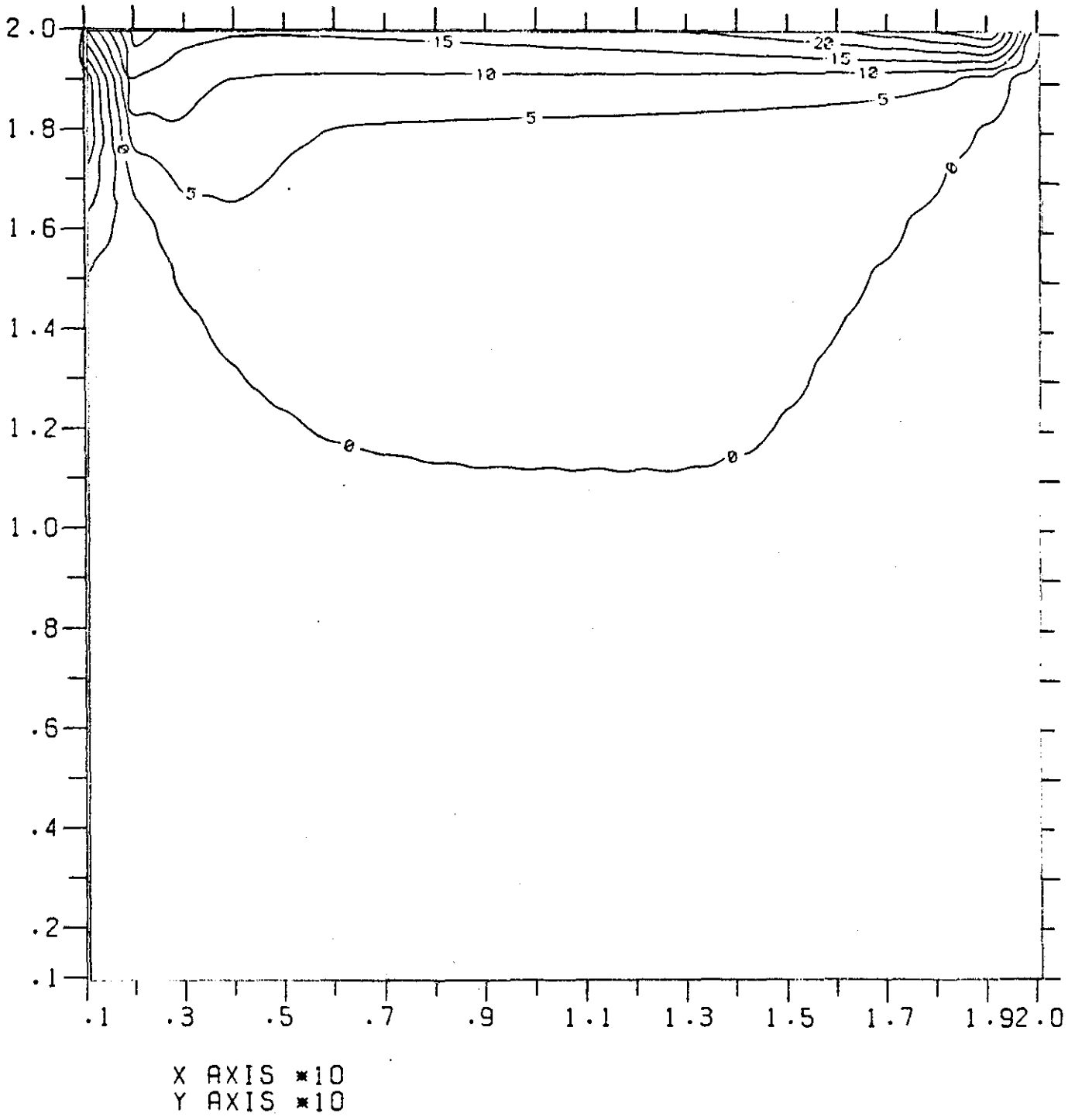


FIGURE 7.5.8

Vorticity Curves for R=500 by SLOR Technique

7.6 CONCLUSIVE REMARKS

From the experiences gained in carrying out the work in this chapter we can express our conclusions as follows.

1. The Boundary Value Technique discussed in this chapter is shown to be a very attractive method for the solution of parabolic equations which have steady-state decaying solutions. In contrast, we emphasize that the method is less attractive if the steady-state solution is not decaying. An attempt to solve a parabolic heat equation with an asymptotic solution of unity was also made but the convergence criterion was never satisfied for accuracies greater than 10^{-1}, where T_∞ was given the value 20. Obviously, one may obtain the required accuracy by putting $T_\infty \geq 100$, but the method will no longer be efficient.

2. For the Navier-Stokes equation, the results displayed in Table 7.2 show that the Boundary Value Technique is faster than the generalised Newton method suggested by Greenspan.

   A comparison was made with the results given by Smith, R.E. & Kidd, A. (1975) for $\Delta x=\Delta y=1/16$ and $\Delta T=2/10$ with $R=100$. The number of iterations for an accuracy of $\varepsilon=10^{-3}$ was found to be 123 by the SOR and 62 by the SLOR method. Although we could not obtain satisfactory convergence for $\varepsilon=10^{-4}$, the method seems to be compatible with the other iterative schemes suggested by Smith et al for the lower accuracy criteria (e.g. $\varepsilon=10^{-3}$).

   However, one might obtain a better accuracy if the term $\frac{\partial \omega}{\partial t}$ in Equ.(7.5.15) is replaced by,

$$\frac{1}{2\Delta T}[-3\omega_{i,j}^{k-1} + 4\omega_{i,j}^{k} - \omega_{i,j}^{k+1}] \ .$$

We also tried the Davidenko path method by solving the system (7.5.4)-(7.5.5) for small R and using the obtained solution as an initial guess to obtain the solutions for larger R. No improvement in the number of iterations and accuracy was achieved by this method. Finally, we conclude that, the Boundary Value Technique for a difficult problem such as the Navier-Stokes equation is quite compatible with the other techniques if the required accuracy is not too stringent. Further topics of research in this area are improving the efficiency and accuracy of the method and also the application of the group hopscotch methods discussed in Chapter VI.

REFERENCES

ACTON FORMAN S.: *Numerical methods that work*, (Harper international edition) 1970.

AFRIAT S.N.: *Composite matrices*, (Quart.J. of Mathematics, Oxford (2) 5, 81-98), 1954.

AMES W.F.: *Nonlinear partial differential equations in engineering*, (Academic Press) 1965.

AMES W.F.: *Numerical methods for partial differential equations*, (Nelson), 1969.

BALFOUR A. and BEVERIDGE W.T.: *Basic numerical analysis with Fortran*, (Heinemann Educational Books), 1972.

CALUS I.M. and FAIRLEY J.A.: *Fourier series and partial differential equations*, (Wiley Interscience), 1970.

CARASSO A. and PARTER S.V.: *An analysis of boundary-value techniques for parabolic problems*, (Mathematics of Computation, 24, No.110, 315-340), 1970.

CARASSO A.: *An analysis of numerical methods for parabolic problems over long times*, (Ph.D. Thesis, University of Wisconsin), 1968.

CHUEN YEN CHOW: *An introduction to computational fluid mechanics*, (John Wiley and Sons), 1979.

CIMENT M. and LEVENTHAL S.H. and WIENBERG B.C.: *The operator compact implicit method for parabolic equations*, (J. of Computational Physics, 28, 135-166), 1978.

CONTE S.D. and DAMES R.T.: *An alternating direction method for solving the Biharmonic equation*, (Mathematics of Computation 11, 198-204), 1957.

COURANT R. and HILBERT D.: *Methods of mathematical physics*, (Interscience Publishers), 1962.

CRANK J: *The mathematics of diffusion*, (Clarendon Press), Oxford, 1975.

DAHLQUIST G. and BJORCK A.: *Numerical methods*, (Prentice-Hall Inc.), 1974.

DANAEE A.: *The solution of non-linear parabolic equations*, (M.Sc. Thesis, Victoria University of Manchester) Jan. 1978.

DANAEE A. and EVANS D.J.: *A new block hopscotch technique*, (Internal Report, Computer Studies, Loughborough University), 1979.

DOUGLAS J.: *On incomplete iteration for implicit parabolic difference equations*, (J.Soc.Indust.Appli.Math. 9, No.3, 433-439), 1961.

EVANS D.J.: *A stable explicit method for the finite-difference solution of a fourth order parabolic partial differential equation*, (Computer Journal, 8, No.3, 280-287), 1965.

EVANS D.J.: *The numerical solution of the fourth boundary value problem for parabolic partial differential equations*, (J.Inst.Maths. Applics. 7, 61-75) 1971.

EVANS D.J.: *The solution of parabolic partial differential equations by boundary value methods*, (Internal Report), 1980.

EVANS D.J. and DANAEE A.: *Hopscotch procedures for a fourth-order parabolic partial differential equation*, (Internal Report, Computer Studies, Loughborough University), 1978.

EVANS D.J. and DANAEE A.: *A Splitting procedure for the numerical solution of parabolic equations of higher order*, (Internal Report, Computer Studies, Loughborough University), 1978.

EVANS, N.T.S. and GOURLAY, A.R.: *The solution of a two dimensional time dependent problem concerned with oxygen metabolism in tissues*, (I.Inst.Applics.Maths.), 1976.

FAIRWEATHER G. and GOURLAY A.R.: *Some stable difference approximations to a fourth-order parabolic partial differential equation*, (J. Mathematics of Computation, Vol.21, 1-11), 1967.

FAIRWEATHER G. and GOURLAY A.R. and MITCHELL A.R.: *Some high accuracy difference schemes with a splitting operator for equations of parabolic and elliptic type*, (Numerische Mathematik 10, 56-66), 1966.

FORSYTHE G.E. and WASOW W.R.: *Finite-difference methods for partial differential equations*, (John Wiley and Sons), 1960.

FREEMAN : *Non-linear equations arising from elliptic space operator, solution by S.O.R. - Newton methods*, (Paper presented in Summer School held in Manchester), 1978.

FOX L.: *Numerical solution of ordinary and partial differential equations*, (Pergamon Press), 1962.

FOX L.: *Finite-difference methods for elliptic boundary-value problems*, (A paper presented in The State of the Art of Numerical Analysis, Ed. Jacobs, D., 799-881), 1977.

GANE C.R.: *Computational techniques for the numerical solution of parabolic and elliptic partial differential equations*, (Doctoral Thesis, Loughborough University), 1974.

GANE C.R. and GOURLAY A.R.: *Block hopscotch procedures for second order parabolic differential equations*, (J.Inst.Maths.Applics. 19, 205-216), 1977.

GANE C.R. and STEPHENSON P.L.: *An explicit numerical method for solving non-linear transient equations of combined conduction and convection,*

GARY J.M.: *On the optimal time step and computational efficiency of difference schemes for P.D.E.,* (J. of Computational Physics, 16, 298-303), 1974.

GEKELER E.: *A convergence of finite difference approximations of parabolic initial-boundary value problems,* (SIAM J.Numer.Anal. 12, No.1, 2-12), 1975.

GERALD CURTIS F.: *Applied numerical analysis,* (Addison Wesley Publishing Company), 1973.

GORDON P.: *Non-symmetric difference equations,* (J.Soc.Indust.Appl.Math., 13, No.3, 667-672), 1965.

GODUNOV S.K. and RYABENKI V.S.: *The theory of difference schemes,* (North Holland Publishing Company), Amsterdam, 1964.

GOURLAY A.R.: *Hopsoctch: a fast second-order partial differential equation solver,* (J.Inst.Maths.Applics. 6, 375-390), 1970.

GOURLAY A.R.: *Recent developments of the hopscotch idea,* (Internal Report, Computer Studies, Loughborough University) 1974.

GOURLAY A.R.: *Recent developments of the hopscotch method,* (Advances in Computer Methods for P.D.E.s, 1-6), 1977.

GOURLAY A.R.: *Splitting methods for time dependent partial differential equations,* (paper presented in The State of the Art in Numerical Analysis, ed. Jacobs D.E.H. 757-), 1977.

GOURLAY A.R. and McGUIRE G.R.: *General hopscotch algorithm for the numerical solution of partial differential equations,* (J.Inst.Maths.Applics., Vol.7, 216-227), 1971.

GOURLAY A.R. and MacKEE S.: *The construction of hopscotch methods for parabolic and elliptic equations in two space dimensions with a mixed derivative,* (J. of Computational and Applied Maths. 3, No.3, 201-206), 1977.

GOURLAY A.R. and MacKEE S.: *Hopscotch methods for elliptic partial differential equations,* (J. of Computational and Applied Maths. 5, No.2, 103-110), 1979.

GOURLAY A.R. and MORRIS J.L.: *Hopscotch difference methods for linear hyperbolic systems,* (IBM J.Res.Dev. 16, 349-353), 1972.

GOURLAY A.R. and MORRIS J.L.: *Hopscotch methods of locally one dimensional type,* (Internal Report, Computer Studies, Loughborough University),

GREENSPAN D.: *Approximate solution of initial-boundary parabolic problems by boundary value techniques,* (Internal Report, Maths. Research Centre, United States Army, Madison, Wisconsin), 19

GREENSPAN D.: *Discrete numerical methods in physics and engineering,* (Academic Press), 1974.

GRIBBEN R.J.: *Elementary partial differential equations,* (Van Nostrand Reinhold Company), 1975.

HILDEBRAND FRANCIS B.: *Finite difference equations and simulations,* (Prentice-Hall Inc.) 1968.

HIRSCH and RUDY: *The role of diagonal dominance and cell Reynolds number in implicit difference methods for fluid mechanics problems,* (J. of Computational Physics 16), 1974.

ISAACSON E. and KELLER H.B.: *Analysis of numerical methods,* (John Wiley & Sons), 1966.

JENNINGS A.: *Matrix computation for engineers and scientists,* (John Wiley & Sons), 1977.

KEAST P. and MITCHELL A.R.: *Finite-difference solution of the third boundary problem in elliptic and parabolic equations,* (Numerische Mathematik 10, 67-75), 1967.

KREYSZIG E.: *Introductory functional analysis with applications,* (John Wiley and Sons), 1978.

LAWSON J.D. and MORRIS J.L.: *The extrapolation of first order methods for parabolic partial differential equations,* (Research Report, CS 77.20), 1977.

LOWAN N.A.: *The operator approach to problems of stability and convergence of solutions of difference equations and the convergence of various iteration procedures,* (Scripta Mathematica Studies, New York) 1957.

MacGUIRE G.R.: *Hopscotch methods for the solution of linear second order parabolic partial differential equations,* (M.Sc. Thesis, Dundee University), 1970.

MITCHELL A.R.: *Computational methods in partial differential equations,* (John Wiley and Sons), 1969.

MORRIS J.L. and NICOLL I.R.: *Hopscotch methods for an anisotripic thermal print head problem*, (J. of Computational Physics, 13, 316-337), 1973.

MORTON K.W.: *Initial value problems by finite difference and other methods*, (A paper presented in The State of the Art in Numerical Analysis, ed. Jacobs, D.E.H., 699-756), 1977.

NICOLESCU MIRON: *L'equation iterée de la chaleur*, (Academie de la republique populaire roumaine, institut de mathematiques, Vol.V, No.3-4, 243-349), 1954.

ORLEY D.G. and MacKEE S.: *A note on hopscotch as applied to the bending beam equation*, (J.Inst.Maths.Applics. 11, 335-338), 1973.

PEACEMAN D.W. and RACHFORD H.H.Jr.: *The numerical solution of parabolic and elliptic differential equations*, (J. of Soc.Indust.Appl. Math. 3, No.1, 28-41), 1955.

RICHTMYER R.D. and MORTON K.W.: *Difference methods for initial value problems*, (Interscience Publishers), 1967.

ROACHE P.J.: *Computational fluid dynamics*, (Hermosa Publishers), 1972.

SAUL'YEV V.K.: *Integration of equations of parabolic type by the method of nets*, (Pergamon Press), 1964.

SHOUP T.E.: *A practical guide to computer methods for engineers*, (Prentice-Hall Inc.), 1979.

SMITH G.D.: *Numerical solution of partial differential equations*, (Oxford University Press), 1969.

SMITH R.E. Jr. and KIDD A.: *Comparative study of two numerical techniques for the solution of viscous flow in a driven cavity*, NASA SP-378 (1975).

SOUTHWELL R.V.: *Relaxation methods in theoretical physics*, (Clarendon Press), Oxford, 1952.

STEPHENSON G.: *An introduction to partial differential equations for science students*, (Longman), 1970.

STEWART G.W.: *Introduction to matrix computations*, (Academic Press), 1973.

STRAWBRIDGE D.R. and HOOPER G.T.J.: *Numerical solution of the Navier-Stokes equations for axisymmetric flows*, (J.Mechanical Engineering Science, 10, No.5, 389-401), 1968.

VAN DER HOUWEN P.J.: *Finite difference methods*, (Mathematical Centre Tracts), Amsterdam, 1968.

VAN DER HOUWEN P.J. and VERWER J.G.: *Non-linear splitting methods for semi-discretized parabolic differential equations*, Afdeling Numerieke Wiskunde, 2e Boerhaavestraat, 49 Amsterdam.

WALSH J.: *Elliptic equations, origin and connection with parabolics*, (A paper presented in Summer School held in Manchester), 1978.

WATT J.: *Parabolic equations and their discretization*, (A paper presented in Summer School held in Manchester), 1978.

WEINBERGER H.F.: *A first course in partial differential equations*, (Xerox College Publishing), 1965.

WILLIAMS J.: *Methods for 2D and 3D-problems (elliptic and parabolic) alternating directions*, (A paper presented in Summer School held in Manchester), 1978.

WOOD W.L.: *Marginal stability: experiments with the diffusion equation*, (Finite Element symposium, ed. Crow J.E., Atlas Computer Laboratory, 139-154), 1974.