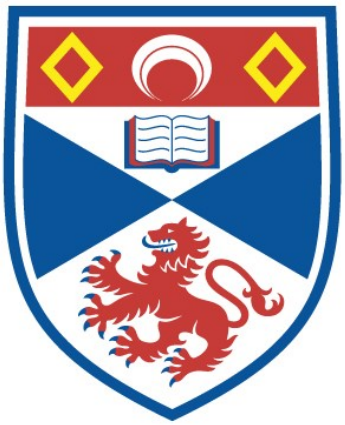
ALTERNATING DIRECTION METHODS FOR HYPERBOLIC SYSTEMS

Alexander R. Gourlay

A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



1966

ALTERNATING DIRECTION METHODS

FOR

HYPERBOLIC SYSTEMS

A thesis presented by

Alexander R. Gourlay, B.Sc.,

to

the University of St.Andrews,

in application for

the degree of

Doctor of Philosophy

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

DECLARATION.

I declare that the following thesis is a record of research work carried out by me, that the thesis is my own composition, and that it has not previously been presented in application for a higher degree.

PREFACE

In October 1960, I matriculated at the University of St.Andrews and read for a degree in Applied Mathematics in St.Salvator's College. In June 1964, I graduated with First Class Honours in Applied Mathematics. In July 1964, I was admitted, under Ordinance 16, as a full-time Research Student in the Department of Mathematics of St.Salvator's College under the supervision of Dr. A. R. Mitchell.

ACKNOWLEDGEMENTS

The Author is indebted to the Carnegie Trust for a Research Scholarship held during the period of research, and to Miss P. A. Cunningham for the expert typing of this thesis.

CERTIFICATE

I certify that Alexander R. Gourlay has spent nine terms in full-time research work under my direction, and is thus qualified to submit the accompanying thesis in application for the degree of Doctor of Philosophy.

Research Supervisor.

CONTENTS

CHAPTER I

INTRODUCTION.

1.1 The equations of mathematical physics

The mathematical formulation of many problems in science and technology leads either to a partial differential equation (p.d.e.) or to a set of such equations. A partial differential equation is an equation involving the rates of change of unknown quantities (such as temperature, pressure ...) with respect to two or more independent variables usually representing time, length or angle.

The exact solution to a p.d.e. in a region R with boundary ∂R is some function which satisfies the equation at every point in R and matches certain boundary conditions on ∂R. For example, the function representing the steady motion of an incompressible fluid through a straight uniform duct satisfies a partial differential equation (Laplace's equation) at every point and takes the value $p(x^2 + y^2)$ on the boundary.

For arbitrarily shaped regions and general boundary conditions it is not usually possible to determine an exact solution to a given partial differential equation. In an attempt to solve such problems approximate methods have been developed. Approximate methods fall into two classes:-

(a) analytical approximate methods where, for example, a truncated series may be obtained for the solution. These methods are usually valid only in certain areas of the region under consideration. However, in these areas, they may give extremely useful information regarding the behaviour of the solution.

(b) <u>numerical approximate methods</u> employing <u>finite difference techniques</u>; these are more frequently used and universally applicable. The concept of a finite difference method will be developed later.

Consider now the two dimensional second order partial differential equation

$$(1.1.1)\quad a\frac{\partial^2 u}{\partial x^2} + 2b\frac{\partial^2 u}{\partial x \partial y} + c\frac{\partial^2 u}{\partial y^2} + d\frac{\partial u}{\partial x} + e\frac{\partial u}{\partial y} + fu = g\ ,$$

where a, b, ... g are functions of the independent variables x,y and perhaps of the dependent variable u. Special cases of (1.1.1) occur frequently, as they are the mathematical forms of the conservation principles of physics. Equation (1.1.1) is said to be

(i) elliptic if $b^2-ac < 0$,

(ii) parabolic if $b^2-ac = 0$,

(iii) hyperbolic if $b^2-ac > 0$,

for all x,y,u in the region under consideration.

The simplest and best known elliptic equations are Laplace's equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

and Poisson's equation

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

If u is the electric potential and $g = -\rho/\epsilon$ where ρ is a charge density and ϵ the dielectric constant, then Poisson's equation is the partial differential equation formulation of Gauss's Law. This states

that the flux through any closed surface is equal to the total charge enclosed by the surface.

Those problems in which one of the two independent variables x,y becomes the time variable t are usually of parabolic or hyperbolic type. The simplest parabolic equation is the heat conduction equation

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

which governs the flow of heat in a bar or rod. The boundary conditions consist of either the temperature given at the two ends of the bar (two values of x), or some measure of diffusion from the ends. It is usual for the temperature distribution in the bar to be known at some instant in time. This is termed the <u>initial condition</u>.

Likewise in the case of the simplest hyperbolic equation, the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \tag{1.1.2}$$

the boundary conditions are given on two lines $x = a,b$ and the initial conditions, usually u and its time derivative $\frac{\partial u}{\partial t}$, are given at some instant in time. The majority of hyperbolic equations arise from vibration problems, or those in which discontinuities, such as shock waves, persist in time.

The wave equation can be written in a different manner as two first order equations. We introduce, as new dependent variables, the quantities

$$u_1 = \frac{\partial u}{\partial t}, \quad u_2 = \frac{\partial u}{\partial x} \tag{1.1.3}$$

and therefore (1.1.2) becomes

$$(1.1.4)\quad \frac{\partial u_1}{\partial t} = \frac{\partial u_2}{\partial x}.$$

From (1.1.3) it is easy to see that u_1 and u_2 are connected by the first order partial differential equation

$$(1.1.5)\quad \frac{\partial u_1}{\partial x} = \frac{\partial u_2}{\partial t}.$$

On introduction of the vector unknown $w = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$, the equations (1.1.4) and (1.1.5) can be written as the first order system of partial differential equations

$$(1.1.6)\quad \frac{\partial}{\partial t}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} - \begin{pmatrix} \cdot & 1 \\ 1 & \cdot \end{pmatrix} \frac{\partial}{\partial x}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = 0$$

or

$$\frac{\partial w}{\partial t} + A\frac{\partial w}{\partial x} = 0,$$

where A is the 2×2 matrix in (1.1.6).

This formulation of the wave equation is trivial in itself, but the principle involved forms the basis of the next section.

1.2 Classification of systems of partial differential equations.

The classification of systems of partial differential equations is most easily described when the equations are formulated in matrix notation.

Consider the system of partial differential equations

$$\frac{\partial u}{\partial t} + A\frac{\partial u}{\partial x} + B = 0 \qquad (1.2.1)$$

where u is an n-dimensional vector whose components u_i $(i=1,2,\ldots n)$ are unknown functions of the two independent variables x and t, A is an $n\times n$ matrix whose entries a_{ij} $(i,j=1,2,\ldots n)$ are functions of x,t and perhaps u, and B is an n-vector whose components are functions of x,t,u. A Characteristic of the system (1.2.1) in the x,t plane is defined as being a curve along which the values of u given on the curve together with equation (1.2.1) are not sufficient to determine the normal derivative of u to the curve. Let s denote the arc length along a curve in the x,t plane and let N be similarly defined along the normal to the curve. Then by the usual chain rules for partial differentiation we have

$$\frac{\partial u}{\partial x} = \frac{\partial u}{\partial s}\cdot\frac{\partial s}{\partial x} + \frac{\partial u}{\partial N}\cdot\frac{\partial N}{\partial x} = \frac{\partial u}{\partial s}\cdot\frac{\partial x}{\partial s} + \frac{\partial u}{\partial N}\cdot\frac{\partial x}{\partial N}$$

$$\frac{\partial u}{\partial y} = \frac{\partial u}{\partial s}\cdot\frac{\partial s}{\partial y} + \frac{\partial u}{\partial N}\cdot\frac{\partial N}{\partial y} = \frac{\partial u}{\partial s}\cdot\frac{\partial y}{\partial s} + \frac{\partial u}{\partial N}\cdot\frac{\partial y}{\partial N} \qquad (1.2.2)$$

Thus, substituting the above relations into (1.2.1), gives

$$(t_N I + x_N A)\frac{\partial u}{\partial N} + (t_s I + x_s A)\frac{\partial u}{\partial s} + B = 0 \qquad (1.2.3.)$$

where I denotes the unit $n\times n$ matrix and $t_N = \frac{\partial t}{\partial N}$ etc.

On a characteristic curve, it follows that the determinant of the matrix of coefficients of terms involving the derivative normal to the curve must vanish, and so

$$|t_N I + x_N A| = 0 \qquad (1.2.4)$$

Therefore the characteristic curves of the system (1.2.1) are given by the solutions of the ordinary differential equations

$$\frac{dx}{dt} = \lambda_i \quad (i = 1,2, \ldots n) \qquad (1.2.5)$$

where λ_i $(i = 1,2, \ldots n)$ are the eigenvalues of the matrix A. If z_i is a row eigenvector associated with the eigenvalue λ_i then multiplying (1.2.3) throughout on the left by z_i we obtain

$$z_i(t_s I + x_s A)\frac{\partial u}{\partial s} + z_i B = 0 \qquad (1.2.6)$$

Each derivative in (1.2.6) is directed along the tangent to the curve (1.2.5). If all the eigenvalues λ_i are real, and if the matrix A has full set of real, linearly-independent eigenvectors $z_1, \ldots z_n$ associated with them, then (1.2.1) can be transformed into the canonical form

$$z_i\left(\frac{\partial t_i}{\partial s} I + \frac{\partial x_i}{\partial s} A\right)\frac{\partial u}{\partial s} + z_i B = 0 \qquad (i = 1,2, \ldots n)$$

where each equation involves differentiation in one characteristic direction $\left(\frac{\partial t_i}{\partial s}, \frac{\partial x_i}{\partial s}\right)$ given by

$$\frac{\partial x_i}{\partial s} \Big/ \frac{\partial t_i}{\partial s} = \lambda_i$$

If the above reduction is possible, that is if the matrix A has n

real eigenvalues and a full set of n linearly independent eigenvectors, then the system (1.2.1) is termed hyperbolic.

If the matrix A has n complex eigenvalues then the characteristics are imaginary and the system is of elliptic type.

In the linear case, where the matrix A is independent of the solution vector u, and has n real distinct eigenvalues, the system (1.2.1) can in fact be reduced to the diagonal form

$$\frac{\partial U_i}{\partial t} + \lambda_i \frac{\partial U_i}{\partial x} + b_i = 0$$

where the components U_i are linear combinations of the u_i; and the b_i are known functions of x, t, $U = (U_1, \ldots U_n)$.

In the following we shall only be concerned with systems which are of hyperbolic type. It has been shown that a necessary condition for the system to be hyperbolic is that the matrix A has n real eigenvalues and a full set of linearly independent eigenvectors. The simplest sufficient condition is that the matrix A is symmetric, that is, $a_{ij} = a_{ji}$ where a_{ij} is the element in the i^{th} row and j^{th} column of the matrix A. When this occurs, the system is termed symmetric hyperbolic.

Let us now consider the extension of these ideas to the general first order system in M space dimensions,

$$(1.2.7) \qquad A_0 u_t = \sum_{i=1}^{M} A_i u_{x_i} + B$$

where the matrices $A_i (i = 0, \ldots M)$ and the vectors u and B are defined as before. Suppose for the moment that B = 0 and that all the matrices

are constant. Then by the method of separation of variables, solutions of the form

$$(1.2.8)\quad u = Ce^{i(\lambda_1 x_1 + \cdots + \lambda_M x_M)} e^{i\lambda t}$$

exist where $\lambda_1, \ldots \lambda_M$ and λ are parameters such that

$$(1.2.9)\quad \left|\lambda A_0 - \sum_{i=1}^{M} \lambda_i A_i\right| = 0$$

and C is a constant non-zero vector such that

$$(1.2.10)\quad \lambda A_0 C = \sum_{i=1}^{M} \lambda_i A_i C$$

If $\lambda_1, \ldots, \lambda_M$ are real, then (1.2.8) can be regarded as a general term in a Fourier representation of the solution of (1.2.7). If every root λ of (1.2.9) is real, in addition, then the solution of (1.2.7) is oscillatory, neither growing or decaying exponentially in time. Therefore we are led, in the multidimensional case, to say that (1.2.7) is of <u>hyperbolic</u> type if all the roots λ of (1.2.9) are real, and have associated with them a complete set of n linearly independent eigenvectors satisfying (1.2.10). This definition is also applied to the quasilinear system (1.2.7).

In practice, as FRIEDRICHS [6] has shown, the majority of systems of the form (1.2.7) are in fact of the <u>symmetric hyperbolic</u> type with the variable t <u>time-like</u>. This nomenclature implies that all the matrices $A_0, A_1, \ldots, A_M$ are symmetric and in addition, A_0 is positive definite. It follows that a matrix P exists such that

$$PA_0P' = I$$

and therefore the roots of (1.2.9) are the roots of

$$(1.2.11) \quad \left| \sum_{i=1}^{M} \lambda_i P A_i P' - \lambda I \right| = 0$$

For any real choice of λ_i, (i=1, ... M) the matrix $\sum_{i=1}^{M} \lambda_i P A_i P'$ is symmetric and thus λ must be real, and the necessary set of n linearly independent characteristic vectors exists.

In the next section the actual mechanism will be given for the reduction of a general second order hyperbolic partial differential equation to a first order symmetric hyperbolic system.

1.3 Reduction of second order hyperbolic partial differential equation to a first order system.

In this section it will be shown that equations expressible in the form

$$(1.3.1)\quad \frac{\partial^2 u}{\partial t^2} = \sum_{i=1}^{n} \sum_{j=1}^{n} g_{ij} \frac{\partial^2 u}{\partial x_i \partial x_j} \quad \text{and lower order terms.}$$

can be reduced to a first order symmetric hyperbolic system of the form

$$(1.3.2)\quad \frac{\partial U}{\partial t} = \sum_{i=1}^{n} A_i \frac{\partial U}{\partial x_i} + BU$$

where U is a vector whose components are linear combinations of u_t and the first partial derivatives of u with respect to the space variables. We shall restrict ourselves to the case when the coefficients g_{ij} are functions of x_i, $(i = 1, \ldots n)$ and t only. It will follow that A_i, $(i = 1, \ldots n)$ and B are likewise functions of x_i, $(i = 1, \ldots n)$ and t. Since the second order operator on the right of (1.3.1) must be of elliptic type, it follows that $z'Gz$ is a positive definite quadratic form for any real non-zero vector $z = (z_1, \ldots z_n)'$, where G is the matrix of coefficients g_{ij}.

The first step in the reduction is to formulate (1.3.1) in the form (1.2.7). This is accomplished by introducing the vector of unknowns

$$V = (u_{x_1}, \ldots, U_{x_n}, U_t) = (U_1, \ldots U_{n+1})$$

and writing (1.3.1) in the form

$$(1.3.3)\qquad \frac{\partial u_{n+1}}{\partial t} = \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij}\frac{\partial u_j}{\partial x_i} + \text{lower order terms.}$$

By the definition of the components of the vector V we also have the equations

$$(1.3.4)\qquad -\sum_{i=1}^{n} \sigma_{ij}\frac{\partial u_i}{\partial t} + \sum_{i=1}^{n} \sigma_{ij}\frac{\partial u_{n+1}}{\partial x_i} = 0 \quad (j = 1, \ldots n)$$

The combination of (1.3.3) and (1.3.4) as a matrix equation then gives us a system of the form

$$(1.3.5)\qquad A_0\frac{\partial V}{\partial t} = \sum_{i=1}^{n} \tilde{A}_i\frac{\partial V}{\partial x_i} + \tilde{B}V$$

where A_0, $\tilde{A}_i$, $(i = 1, \ldots n)$ are symmetric matrices. The matrix A_0 is given by

$$A_0 = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} & 0 \\ \sigma_{21} & & & & 0 \\ \vdots & & & & \vdots \\ \sigma_{n1} & & & \sigma_{nn} & 0 \\ 0 & & & 0 & 1 \end{bmatrix}$$

For any vector $Z = (z_1, \ldots z_n, z_{n+1})'$ the quadratic form $Z'A_0Z_0$ is guaranteed to be positive definite by the ellipticity of the right hand side of (1.3.1) and therefore the matrix A_0 is positive definite.

It follows that there exists a matrix P depending on $x_1, \ldots x_n, t$ such that

$$PA_0P' = I$$

We define a linear combination of the components of V by the relation

(1.3.6) $$V = P'U.$$

Introducing (1.3.6) into (1.3.5) and multiplying through on the left by the matrix P leads to the transformed system,

$$PA_0 \frac{\partial}{\partial t}(P'U) = \sum_{i=1}^{n} \tilde{PA}_i \frac{\partial}{\partial x_i}(P'U) + \tilde{PBP'U}$$

or

$$PA_0P' \frac{\partial U}{\partial t} + PA_0 \frac{\partial P'}{\partial t} U = \sum_{i=1}^{n} \tilde{PA}_i P' \frac{\partial U}{\partial x_i} + \sum_{i=1}^{n} \tilde{PA}_i \frac{\partial P'}{\partial x_i} U + \tilde{PBP}U$$

which is of the required form:

$$\frac{\partial U}{\partial t} = \sum_{i=1}^{n} A_i \frac{\partial U}{\partial x_i} + BU$$

where the matrices A_i, $(i = 1, \ldots n)$ are symmetric.

It should be noticed that no restrictions are required on the matrix B. In what follows we shall assume that the matrix B is identically zero, that is, we are interested in solving the equation

(1.3.7) $$\frac{\partial U}{\partial t} = \sum_{i=1}^{M} A_i \frac{\partial U}{\partial x_i},$$

particularly for the case M = 2. The extension to the case of more than two space dimensions will be mentioned in the concluding remarks.

1.4 Some examples

Before introducing the notation and concepts of numerical analysis required in the following chapters, we shall briefly indicate a few symmetric hyperbolic systems related to physical problems. The restriction to linear systems will be imposed except in Chapter V where the non-linear case will be considered in detail.

In section 1.1 it was shown how the one-dimensional wave equation can be written as a first order system. Indeed FRIEDRICHS has shown how Laplace's equation and the heat conduction equation can be similarly described [6]. But as we shall be mainly interested in the numerical solution of problems in two (or more) space dimensions we shall restrict ourselves to examples of this type.

The two-dimensional wave equation can be written as a first order system in a variety of ways:-

$$(a)\quad \frac{\partial U}{\partial t} = \begin{pmatrix} 1 & \cdot \\ \cdot & -1 \end{pmatrix} \frac{\partial U}{\partial x} + \begin{pmatrix} \cdot & 1 \\ 1 & \cdot \end{pmatrix} \frac{\partial U}{\partial y}, \quad U = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

u_1, u_2 being coupled solutions of the wave equation.

$$(b)\quad \frac{\partial U}{\partial t} = \begin{pmatrix} \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot \end{pmatrix} \frac{\partial U}{\partial x} + \begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 \\ \cdot & 1 & \cdot \end{pmatrix} \frac{\partial U}{\partial y}, \quad U = \begin{pmatrix} u_x \\ u_y \\ u_t \end{pmatrix}$$

and even

$$(c)\quad \frac{\partial U}{\partial t} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \frac{\partial U}{\partial x} + \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix} \frac{\partial U}{\partial y}, \quad U = \begin{pmatrix} u_x + \frac{1}{\sqrt{2}} u_t \\ u_y + \frac{1}{\sqrt{2}} u_t \end{pmatrix}$$

A three-dimensional example is furnished by the well-known Maxwell's equations

$$\frac{\partial \underline{B}}{\partial t} + \nabla \wedge \underline{E} = 0 \ , \ \frac{\partial \underline{E}}{\partial t} - \nabla \wedge \underline{B} = 0$$

governing the electric and magnetic intensity vectors, $\underline{E} = (E_1 , E_2, E_3)$ and $\underline{B} = (B_1, B_2, B_3)$, where $\nabla \wedge \underline{a}$ is the curl of the vector $\underline{a}$.

If $w' = (B_1, B_2, B_3, E_3, E_2, E_1)$ then the above equations can be written in the symmetric hyperbolic form

$$\frac{\partial w}{\partial t} + \begin{bmatrix} \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & -1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & -1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \frac{\partial w}{\partial x} + \begin{bmatrix} \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & -1 \\ 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & -1 & \cdot & \cdot & \cdot \end{bmatrix} \frac{\partial w}{\partial y} + \begin{bmatrix} \cdot & \cdot & \cdot & -1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \frac{\partial w}{\partial z} = 0$$

In the above examples all the matrices A_i (in the notation of (1.3.7)) are independent of x_i and t. However if we consider the polar form of the wave equation

$$(1.4.1) \quad \frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2}$$

then it is not difficult to show that an equivalent system is given by

$$\frac{\partial}{\partial t} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \frac{\partial}{\partial r} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} + \frac{1}{r} \begin{pmatrix} -\sin\theta & \cos\theta \\ \cos\theta & \sin\theta \end{pmatrix} \frac{\partial}{\partial \theta} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

where u_1, u_2 both satisfy (1.4.1), and the matrices are no longer constant.

1.5 Finite Differences

As stated in 1.1, it is usually not possible to obtain an exact solution of an arbitrary partial differential equation in a general region, and even though a so-called closed form solution is found it may not be in a form suitable for computation. For example consider the problem of solving Poisson's equation

$$(1.5.1)\quad \frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\phi}{\partial y^2} + 2 = 0$$

over the rectangle $-1 \leq x, y \leq 1$, with $\phi = 0$ on the perimeter. The closed form solution for this problem is given by

$$\phi = 1 - y^2 - \frac{32}{\pi^3} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^3} \operatorname{sech} \frac{(2n+1)\pi}{2} \cosh \frac{(2n+1)\pi x}{2} \cos \frac{(2n+1)\pi y}{2}$$

A numerical method will certainly provide an adequate numerical solution more easily and efficiently than the laborious evaluation of the above solution.

In a finite difference method the region of interest, $-1 \leq x, y \leq 1$ in the above example, is covered by a grid or rectangular mesh formed by two sets of lines parallel to the x-y axes respectively. (see Figure 1.5.1)

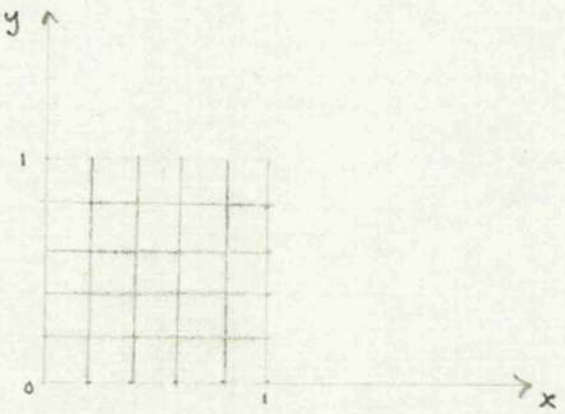


Figure 1.5.1

The grid is assumed to be uniform, the distance between parallel lines being h. The **nodes**, or **lattice points** are the intersections of the lines. The partial differential equation (1.5.1) is replaced by a set of simultaneous linear equations connecting the values of ϕ at the lattice points in such a way as to approximate the partial differential equations. For example, consider the following part of the grid shown in Figure 1.5.1

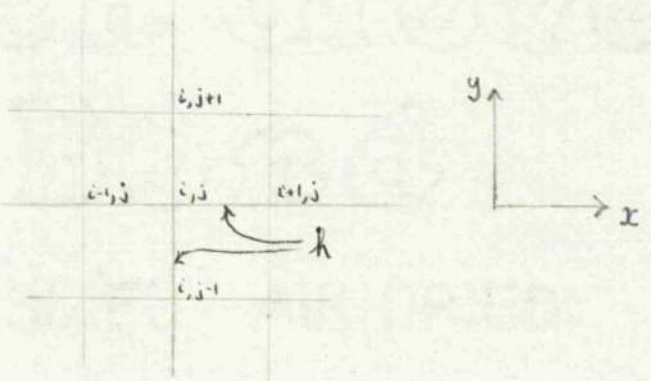


Figure 1.5.2

consisting of the points $(i+r, j+s)$, $(r = -1,0,1,\ s = -1,0,1)$. Let ϕ_{ij} denote the value of ϕ at the point (i,j). Then if we take the combination

$$\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j} = \delta x^2 \phi_{i,j}$$

and expand by Taylor's theorem about the point (i,j) we obtain

$$\delta^2_x \phi_{i,j} = \left[h^2 \frac{\partial^2}{\partial x^2} + \frac{h^4}{12} \frac{\partial^4}{\partial x^4} + \cdots \right] \phi_{i,j}$$

Therefore

$$\frac{1}{h^2} \delta^2_x \phi_{i,j} = \frac{\partial^2 \phi}{\partial x^2} i,j + O(h^2)$$

is an approximation to $\frac{\partial^2 \phi}{\partial x^2}$ at the point (i,j), and the error is second order in h. [Later this will in fact be referred to as having a fourth order truncation error]. Replacing $\frac{\partial^2 \phi}{\partial y^2}$ similarly, Poisson's equation (1.5.1) becomes

$$\frac{1}{h^2} [\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j} + \phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j+1}] + 2 = O(h^2)$$

or

$$(1.5.2) \quad \phi_{i+1,j} + \phi_{i-1,j} + \phi_{i,j+1} + \phi_{i,j-1} - 4\phi_{i,j} + 2h^2 = O(h^4)$$

The terms of order h^4 on the right hand side of equation (1.5.2) constitute the principal part of the truncation error of this formula. Neglecting the truncation error in (1.5.2) and scanning over the whole region with such a formula results in a large set of simultaneous equations requiring to be solved for the unknown functions $\phi_{i,j}$. Obviously the smaller we take h the better accuracy we obtain. But any decrease in h and increase in accuracy will be offset by the fact that a larger set of equations will have to be solved. However we shall not concern ourselves further with this problem but, having introduced the concept of a finite difference method will now turn to an important initial-boundary value problem.

Consider the solution of the equation

$$u_t + u_x = 0$$

in the region

$$0 \leqslant x \leqslant a; \quad t \geqslant 0$$

subject to the initial condition

$$u(x,0) = f(x); \quad 0 \leqslant x \leqslant a$$

and the boundary conditions

$$u(0,t) = g_0(t), \quad u(a,t) = g_a(t); \quad t \geqslant 0$$

The region is covered by a rectangular grid with spacing h in the x-direction and k in the t(time) - direction, as shown in Figure (1.5.3)

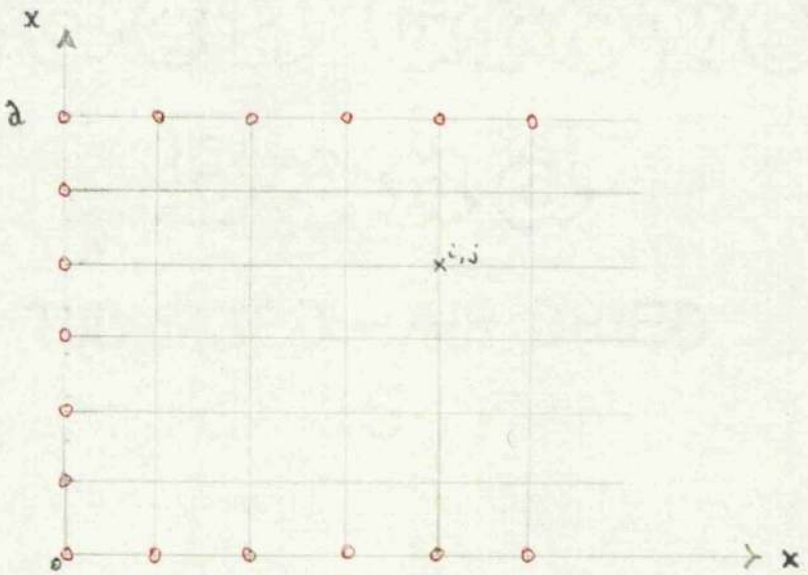


Figure (1.5.3)

The values of u at the points marked o are known from the initial and boundary conditions. If Nh = a, the points on the line x = a, known for all time may be denoted by u(Nh, jk). Suppose we have by some means calculated all the values u(ih, jk) for i=1,...N-1, j=1,...m.

The problem is now to introduce a replacement of the first order equation in such a way that the values $u(ih,(m+1)k)$, $(i=1,\ldots n-1)$ may be calculated from the known values $u(ih,jk)$, $(i=1,\ldots N-1, j=1,\ldots m)$. This will result in a step by step procedure for the numerical integration of the differential equation.

Perhaps the simplest replacement would be

$$\frac{u_{i,m+1}-u_{i,m}}{k} + \frac{1}{2h}(u_{i+1,m}-u_{i-1,m}) = 0\ (k+h^2)$$

which is usually written in the form

$$(1.5.3)\quad u_{i,m+1} = u_{i,m} - p/2\,(u_{i+1,m} - u_{i-1,m}) + 0(k^2+kh^2)$$

where

$$p = k/h$$

is the mesh ratio. An alternative scheme of comparable accuracy is given by

$$(1.5.4)\quad u_{i,m+1} = \frac{1}{2}(u_{i+1,m}+u_{i-1,m}) - p/2(u_{i+1,m}-u_{i-1,m}) + 0(k^2 + kh^2)$$

At first sight it is not obvious which of these schemes one should employ in an actual calculation. Both values of $u_{i,m+1}$ are obtained solely in terms of the values of u at the m^{th} time level. Such schemes are termed explicit. An implicit scheme is one involving more than one point at the $(m+1)^{st}$ level. We shall see in a later section the important difference between schemes (1.5.3) and (1.5.4) when we consider their stability with respect to growth of round-off error.

1.6. Notation and Concepts

In the preceding sections a certain amount of notation has already been introduced, such as the representation of the value $\phi(ih, jk)$ by $\phi_{i,j}$. In the following chapters we shall be interested in problems involving the independent variables x,y,t, with solution u(x,y,t). A square mesh in the space coordinate plane will be assumed, and therefore the following notation will be appropriate:

$$u(ih, jh, mk) = u^m_{i,j}$$

Sometimes this will be written simply as u_m. The formulation of the schemes will be more compact if we introduce the following difference operators:-

$$\Delta_x u^m_{i,j} = u^m_{i+1,j} - u^m_{i,j}$$

$$\nabla_x u^m_{i,j} = u^m_{i,j} - u^m_{i-1,j}$$

$$\Delta_y u^m_{i,j} = u^m_{i,j+1} - u^m_{i,j}$$

$$\nabla_y u^m_{i,j} = u^m_{i,j} - u^m_{i,j-1}$$

$$\Delta_t u_m = u_{m+1} - u_m$$

The variable u_m will be restricted to denote the value of the function, or vector function u, at the point (ih,jh,mk), and U_m will denote the vector of the totality of $u^m_{i,j}$ over the appropriate range of i,j. U_m therefore denotes the totality of values at u at nodes on the m^{th} time level.

We shall only be interested in <u>two level difference schemes</u>, that is schemes which connect points on two neighbouring time levels.

Such schemes can be written (for a two-space dimensional problem) in the form

$$(1.6.1) \qquad \sum_{i,j}{}' \tilde{A}_1^{i,j} u_{m+1}(x+ih, y+jh) = \sum_{i,j}{}' \tilde{A}_2^{i,j} u_m(x+ih, y+jh)$$

where the summations are over a finite range of i, j, not necessarily the same at the two time levels. If the notation introduced above is now employed, (1.6.1) can be put into the form

$$A_1(\Delta_x, \Delta_y, \nabla_x, \nabla_y)\, u_{m+1} = A_2(\Delta_x, \Delta_y, \nabla_x, \nabla_y) u_m$$

or briefly

$$(1.6.2) \qquad A_1 u_{m+1} = A_2 u_m$$

The inversion of the operator A_1, if possible, allows us to write (1.6.2) as

$$(1.6.3) \qquad u_{m+1} = C u_m$$

where C may now have an infinite expansion in terms of the difference operators. It should be noticed that the existence of A_1^{-1} is tantamount to assuming that (1.6.2) can be solved for u_{m+1}. Also, if in (1.6.1), Fourier series are substituted for $u_{m+1}(x,y)$ and $u_m(x,y)$, followed by a cancellation of a common factor, the remaining terms can be written as

$$(1.6.4) \qquad G_1 V_{m+1} = G_2 V_m$$

where

$$G_i = \sum_{r,s}{}' \tilde{A}_i \exp\left\{ i(\beta r h + \gamma s h) \right\} ,$$

and β, γ are real numbers. Equation (1.6.4) may be written in the form

(1.6.5) $V_{n+1} = G(\beta,\gamma)V_n$.

The quantity $G(\beta,\gamma)$ is referred to as the <u>amplification matrix</u> of the system (1.6.2). In general h will be some function of k so that the operator C in equation (1.6.3) can be regarded as involving k only.

If (1.6.2) is derived from a differential problem of the form

$$\frac{\partial u}{\partial t} = Au$$

where A is a linear differential operator in the space coordinates x,y, then scheme (1.6.2) is said to be <u>consistent</u> if

$$\lim_{k\to 0} \left|\left| \left\{ \frac{C(k)-I}{k} - A \right\} u \right|\right| = 0$$

uniformly in time. [||.|| denotes some norm]. This is simply a statement of the requirement that as k tends to zero (and hence h tends to zero as some function of k) the truncation error vanishes.

Let us now turn to the concept of <u>stability</u> with respect to sequences of calculations, each sequence with a fixed time increment k_i such that k_i tends to zero as i increases. The calculation can be regarded as applying the infinite set of operators

$$C(k_i)^m \qquad \begin{matrix} 0 \leq mk_i \leq T \\ i = 1,2,\ldots \end{matrix}$$

to the initial data. The requirement of stability is that no component of the inital data can be amplified more than a certain amount in any numerical procedure. The approximation is said to be stable if the above set of operators is uniformly bounded for the positive decreasing sequence $\{k_i\}$. This definition is a property solely of

the difference equation and is independent of the differential equation. In fact, to test for stability the matrices $G(\beta,\gamma)$ are used. We state, without proof, the two conditions for stability which shall be employed from time to time in this thesis.

A necessary condition for stability, the von NEUMANN condition [25], is that

(1.6.6) $$|\lambda_i| \leq 1 + 0(k) \qquad (i = 1,2, \ldots n)$$

where $\lambda_i, (i = 1,2, \ldots N)$ are the eigenvalues of $G(\beta,\gamma)$

[The quantity $o(k)$ must be included to allow for problems whose solutions have a genuine growth in time].

A sufficient condition due to LAX-RICHTMYER [17] and KREISS [14], [16] is that

(1.6.7) $$|\mu_i| \leq 1 + 0(k) \qquad (i = 1,2, \ldots n)$$

where $\mu_i, (i = 1, \ldots n)$ are the eigenvalues of G^*G, where G^* is the hermitian transpose of G. This sufficient condition can be used in the slightly different form

(1.6.8) $$||G^*G|| \leq 1 + 0(k)$$

where $||.||$ is some norm.

Operating on the initial data u_0, m times with $C(k)$ gives

$$u_m = [C(k)]^m u_0$$

which is an approximation to the exact solution $\bar{u}_T$ of the differential equation at $(ih, \beta h, mk)$ where $mk = T$. The operators $C(k_i)$ are said to form a convergent approximation to the solution of the initial value problem if

$$\lim_{i\to\infty} \| C(k_i)^{m_i} u_0 - \bar{u}_T \| = 0$$

where $\{k_i\}$ is a positive strictly decreasing sequence, such that

$$\lim_{i\to\infty} m_i k_i = T$$

LAX [17] has shown that for a properly posed linear initial value problem, a consistent finite difference replacement is convergent if and only if it is stable.

Now let us return and consider, in the light of these concepts, the schemes (1.5.3) and (1.5.4). It is a straightforward calculation to verify that these schemes are consistent. Let ρ_3, ρ_4 denote the amplification factors (for this problem G is a one-by-one matrix) of the schemes (1.5.3) and (1.5.4). Then

$$|\rho_3|^2 = 1 + \frac{k^2}{h^2} \sin^2\beta h.$$

(1.6.9)
$$|\rho_4|^2 = \cos^2\beta h + \frac{k^2}{h^2} \sin^2\beta h$$
$$= 1 - (1 - \frac{k^2}{h^2})\sin^2\beta h.$$

It is obvious that $|\rho_4| \leq 1$ if $|\frac{k}{h}| \leq 1$ for any choice of dependence of h on k. Therefore (1.5.4) is stable for $p \leq 1$. However with regard to (1.6.9) it does matter what the relation is between h and k. Suppose

$$ph = k^\alpha , \quad \alpha > 0$$

and thus

(1.6.10) $$|\rho_3|^2 = 1 + p^2 k^{2-2\alpha} \sin^2\beta h$$

or

$$|\rho_3| \leq 1 + 0(k)$$

if and only if

$a \leq \frac{1}{2}$.

Therefore k must tend to zero faster than h^2 in order that (1.5.3) be stable. However in practice (1.5.4) is a much more acceptable scheme and (1.5.3) would rarely be used.

Finally let us consider the <u>domains of dependence</u> of hyperbolic partial differential equations and their related difference schemes. In particular we shall concern ourselves with the solution of the wave equation

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

with appropriate initial and boundary conditions, in the region $0 \leq x \leq 1, \quad t \geq 0$.

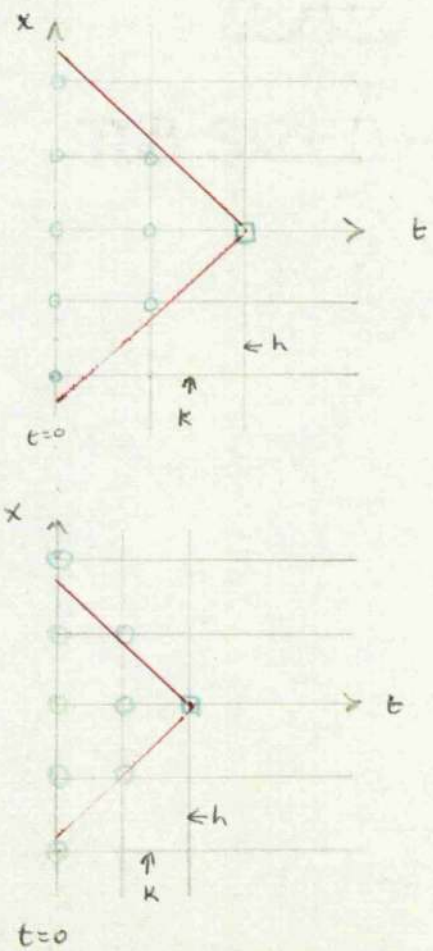


Figure 1.6.1

Suppose the value of u at the point marked □ (see Figure 1.6.1) has been obtained through some finite difference calculation which uses the points marked ○ . Let D_h denote the domain of dependence of the difference replacement, that is the interval between the extreme data points ○ on the line t = 0 employed in calculating the □ value. Similarly let D denote the corresponding domain of dependence for the differential equation. This consists of the interval on t = 0 between the characteristics which pass through the point □ and are inclined at 45° to the x,t axes. In Figure 1.6.1(a), k has been chosen so that $D_h \subset D$. Let $E = D \cap D_h$. Then any change in the initial data in E although affecting the final solution of the differential equation will in no way affect the solution of the difference scheme. This situation persists as the mesh size is reduced, the mesh ratio being kept constant. Therefore the difference solution cannot be expected to converge to the solution of the differential equation unless

$$D_h \supseteq D$$

as in Figure 1.6.1(b). This is referred to as the Courant-Friedrichs-Lewy Condition [3] and it is a necessary condition for the stability and convergence of a finite difference approximation to a hyperbolic equation (or system).

1.7 Existing Methods.

The problem which we shall consider in the following three chapters will be the numerical solution of the first order linear hyperbolic system in two space dimensions

$$(1.7.1)\quad \frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}$$

where u is an n-vector and A;B are nxn symmetric matrices which may be functions of x,y,t.

Although several authors have proposed schemes for the numerical integration of (1.7.1) only two schemes have any practical value. The most used scheme without doubt is the LAX-WENDROFF method [23] (see BURSTEIN [2]), which may be written, in difference operator notation, in the form

$$(1.7.2)\quad u_{m+1} = [I + \tfrac{1}{2}(pA+p^2A^2)\Delta_x + \tfrac{1}{2}(pA-p^2A^2)\nabla_x + \tfrac{1}{2}(pB+p^2B^2)\Delta_y + \tfrac{1}{2}(pB-p^2B^2)\nabla_y + \tfrac{1}{8}p^2(AB+BA)(\Delta_x+\nabla_x)(\Delta_y+\nabla_y)]u_m + O(h^3)$$

This is an explicit scheme involving nine points at the base level (see Figure 1.7.1)

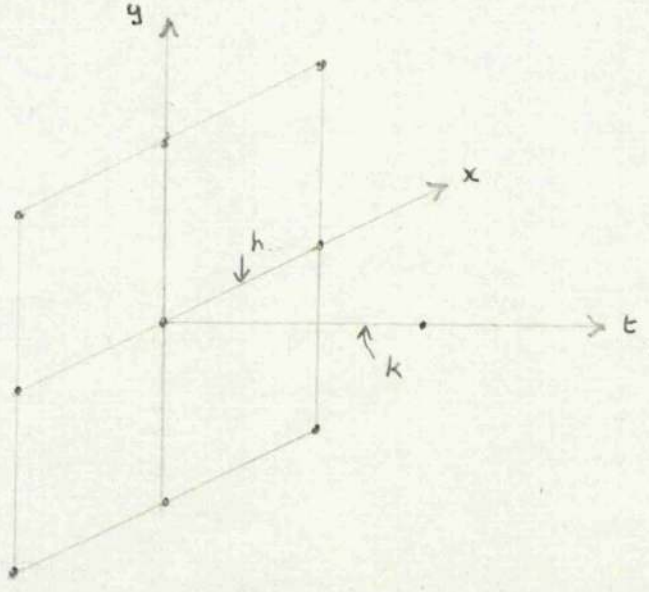


Figure 1.7.1

The worst feature of (1.7.2) is its poor stability characteristics. If we let

$$|\lambda_m| = \max\left\{|\lambda_A|, |\lambda_B|\right\}$$

where

$$|A - \lambda_A I| = 0, \quad |B - \lambda_B I| = 0$$

then the condition for (1.7.2) to be stable is

$$p|\lambda_m| \leq \frac{1}{2\sqrt{2}}$$

STRANG [34] has shown how to modify this method so that the stability condition is

$$p|\lambda_m| \leq 1$$

(the maximum allowable for an explicit scheme, by the Courant-Friedrichs-Lewy condition) but the complexity renders Strang's scheme valueless in practical calculations.

The second scheme is due to WENDROFF [40]. It is an eight-point explicit scheme based on four points at each of two neighbouring time levels (Figure 1.7.2)

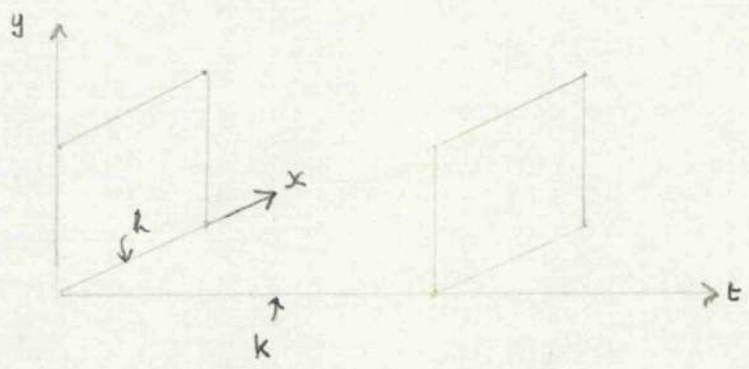


Figure 1.7.2

and is given by

$$(1.7.3)\quad [I+\tfrac{1}{2}(I-pA)\,\Delta_x + \tfrac{1}{2}(I-pB)\,\Delta_y+\tfrac{1}{4}(I-pA-pB)\,\Delta_y\Delta_x]u_{m+1}$$
$$= [I+\tfrac{1}{2}(I+pA)\,\Delta_x + \tfrac{1}{2}(I+pB)\,\Delta_y+\tfrac{1}{4}(I+pA+pB)\,\Delta_y\Delta_x]u_m+O(h^3)$$

If data is given on the planes, $x,y = 0$ this method is effectively explicit.

However although Wendroff demonstrated the convergence of this scheme for a restricted class of problems he was unable to prove anything with respect to its stability. Also, the scheme is only applicable if the matrices A,B are negative definite.

Other schemes have been proposed by ANUCINA [1], JOHANSON and KREISS [13] and KREISS [15].

1.8 Outlook

In the following chapters new implicit methods will be developed with a view to obtaining optimum stability characteristics. All the methods developed will be of the same accuracy as the methods of Lax and Wendroff, and Wendroff.

In Chapter II a class of formulae based on the same nodes as Wendroff's formula is investigated. In Chapter III a study of an eighteen-point scheme consisting of two nine-point operators is made. Chapter IV introduces a technique whereby all known and newly developed methods can be obtained by an algebraic procedure without direct resort to Taylor's theorem.

In Chapter V a fresh start is made to consider the problem when the matrices A and B are purely functions of the components of u in such a way that the system may be presented in the conservation form

$$\frac{\partial u}{\partial t} + \frac{\partial f(u)}{\partial x} + \frac{\partial g(u)}{\partial y} = 0$$

Finally in Chapter VI we consider briefly the modifications necessary to extend the schemes to the general N-space dimensional system

$$\frac{\partial u}{\partial t} = \sum_{i=1}^{N} A_i \frac{\partial u}{\partial x_i}$$

CHAPTER II

METHODS INVOLVING AN EIGHT POINT LATTICE

Gourlay and Mitchell [9]

Introduction.

A two-level difference scheme based on a total of eight nodes, four at each time level, will now be introduced for the numerical solution of the first order linear hyperbolic system (1.7.1). A general formula will be derived which will contain, as a special case, the formula of Wendroff mentioned in section 1.7. From the general formula a new scheme, which is one of a family of schemes, is obtained and cast in an alternating direction form.

An alternating direction method is one whereby a partial differential equation in q space variables is solved by a multi-step finite difference procedure of a particular type. The name arises from the fact that at each step in the procedure we solve along lines parallel to a space coordinate axis, this axis being different at each step. Although the finite difference scheme at each stage is implicit in nature, it is of a particular form which may be solved by a direct or non-iterative method.

Alternating direction methods were first introduced by PEACEMAN and RACHFORD in 1955 [27] for the numerical solution of the heat conduction equation in two space variables. Since then, they have been extended to obtain numerical solutions of the heat equation in an arbitrary number of space variables, Laplaces equation, the wave equation and the biharmonic equation.

The alternating direction methods introduced in this thesis are believed to be the first attempts at solving hyperbolic systems by such

procedures, although KREISS [15] and GARY [8] hinted at the possibility of their use.

2.1 The Scalar Case.

To introduce the techniques involved in replacing the partial differential system (1.7.1) by a finite difference scheme, consider first of all the replacement of the scalar form of (1.7.1), that is

$$\frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}$$

where A,B are scalar constants and u is a scalar function of x,y,t.

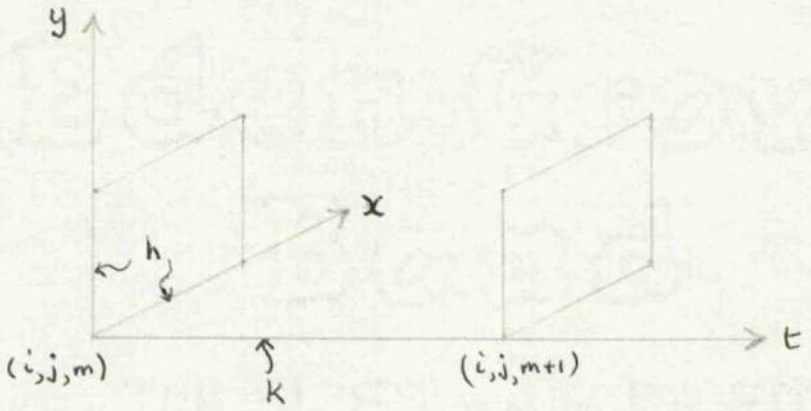


Figure 2.1.1

Assuming the origin to be a mesh point, the grid shown in Figure 2.1.1, contains the points $u(l_1h, l_2h, l_3k) = u^{l_3}_{l_1,l_2}$ $(l_1=i,i+1; l_2=j,j+1; l_3=m,m+1)$ where h,k are the mesh spacings in the space and time coordinate directions respectively. Introducing, as before, the forward difference operators Δ_x, Δ_y defined by

$$\Delta_x u^m_{i,j} = u^m_{i+1,j} - u^m_{i,j}$$

(2.1.1)

$$\Delta_y u^m_{i,j} = u^m_{i,j+1} - u^m_{i,j}$$

we obtain

$$u^m_{i+1,j} = (1 + \Delta_x)\, u^m_{i,j}$$

(2.1.2) $$u^m_{i,j+1} = (1 + \Delta_y)\, u^m_{i,j}$$

$$u^m_{i+1,j+1} = (1 + \Delta_y)(1 + \Delta_x) u^m_{i,j}$$

and similar results at the $(m+1)^{st}$ time level. Therefore a general formula through the eight points shown in Figure 2.1.1 can be written in the form

(2.1.3) $$[1 + a\Delta_x + b\Delta_y + c\Delta_y\Delta_x]u^{m+1}_{i,j} \; [1 + d\Delta_x + e\Delta_y + f\Delta_y\Delta_x]u^m_{i,j} = 0$$

where a,b,...f are arbitrary non-zero coefficients.

If the values of $u^{m+1}_{i,j}$, $\Delta_x u^m_{i,j}$... are now expanded, by Taylor series, in terms of u and its derivatives with respect to x,y,t at the point (i,j,m), the expressions

$$u^{m+1}_{i,j} = \left[u + k\frac{\partial u}{\partial t} + \frac{k^2}{2!}\frac{\partial^2 u}{\partial t^2} + \cdots \right]^m_{i,j}$$

(2.1.4) $$\Delta_x u^m_{i,j} = \left[u + h\frac{\partial u}{\partial x} + \frac{h^2}{2!}\frac{\partial^2 u}{\partial x^2} \cdots \right]^m_{i,j} - u^m_{i,j}$$

$$= \left[h\frac{\partial u}{\partial x} + \frac{h^2}{2!}\frac{\partial^2 u}{\partial x^2} + \cdots \right]^m_{i,j}$$

etc. are obtained. The derivatives with respect to t in (2.1.4) are now replaced by using the relations

$$\frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}$$

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial t}\left(A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}\right) = A\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial t}\right) + B\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial t}\right) = A^2\frac{\partial^2 u}{\partial x^2} + 2AB\frac{\partial^2 u}{\partial x \partial y} + B^2\frac{\partial^2 u}{\partial y^2}$$

$$p = k/h,$$

and the resulting expressions substituted in (2.1.3). On equating the coefficients of $h\frac{\partial u}{\partial x}$, $h\frac{\partial u}{\partial y}$, $\frac{h^2}{2}\frac{\partial^2 u}{\partial x^2}$, $h^2\frac{\partial^2 u}{\partial x \partial y}$, $\frac{h^2}{2}\frac{\partial^2 u}{\partial y^2}$ evaluated at (i,j,m) to zero we find that the following conditions on the unknowns a,b,...f hold:-

$$a = \tfrac{1}{2}(1 - pA) \qquad d = \tfrac{1}{2}(1 + pA)$$

(2.1.5) $$b = \tfrac{1}{2}(1 - pB) \qquad e = \tfrac{1}{2}(1 + pB)$$

$$f - c = \tfrac{1}{2}p\,(A + B)$$

The difference replacement (2.1.3) now takes the form

(2.1.6) $$[1 + \tfrac{1}{2}(1 - pA)\Delta_x + \tfrac{1}{2}(1 - pB)\Delta_y + c\Delta_y\Delta_x]u^{m+1}_{i,j}$$

$$= [1 + \tfrac{1}{2}(1 + pA)\Delta_x + \tfrac{1}{2}(1 + pB)\Delta_y + f\Delta_y\Delta_x]u^{m}_{i,j}$$

where

$$f - c = \tfrac{1}{2}p(A + B)$$

and the principal part of the truncation error is of order h^3.

The choice

$$c = \frac{1}{4}\left[1 - p(A + B)\right] \qquad f = \frac{1}{4}\left[1 + p(A + B)\right]$$

transforms (2.1.6) into Wendroff's scheme (1.7.3) for the scalar form of (1.7.1).

However if we make the choice

$$c = \frac{1}{4}(1 - pB)(1 - pA) \qquad f = \frac{1}{4}(1 + pB)(1 + pA)$$

then (2.1.6) can be written in the form

(2.1.7) $[1 + \frac{1}{2}(1-pB)\Delta_y][1 + \frac{1}{2}(1-pA)\Delta_x]u^{m+1}_{i,j} = [1+\frac{1}{2}(1+pB)\Delta_y][1+\frac{1}{2}(1+pA)\Delta_x]u^{m}_{i,j}$.

On introduction of an intermediate or auxiliary solution $v^{m+1}_{i,j}$ by means of the relation

(2.1.8a) $[1 + \frac{1}{2}(1 - pB)\Delta_y]v^{m+1}_{i,j} = [1 + \frac{1}{2}(1 + pA)\Delta_x]u^{m}_{i,j}$

equation (2.1.7) takes the form

$[1 + \frac{1}{2}(1 - pB)\Delta_y][1 + \frac{1}{2}(1 - pA)\Delta_x]w^{m+1}_{i,j} = [1 + \frac{1}{2}(1+pB)\Delta_y][1+\frac{1}{2}(1-pB)\Delta_y]v^{m+1}_{i,j}$

which is equivalent to

(2.1.8b) $[1 + \frac{1}{2}(1 \quad pA)\Delta_x]u^{m+1}_{i,j} = [1 + \frac{1}{2}(1 + pB)\Delta_y]v^{m+1}_{i,j}$

Equations (2.1.8a) and (2.1.8b) constitute an alternating direction factorisation of (2.1.7) as they require the solution of two point recurrence relations

(a) in the y-direction at the first step

(b) in the x-direction at the second step.

This process may be represented diagrammatically as in Figure 2.1.2

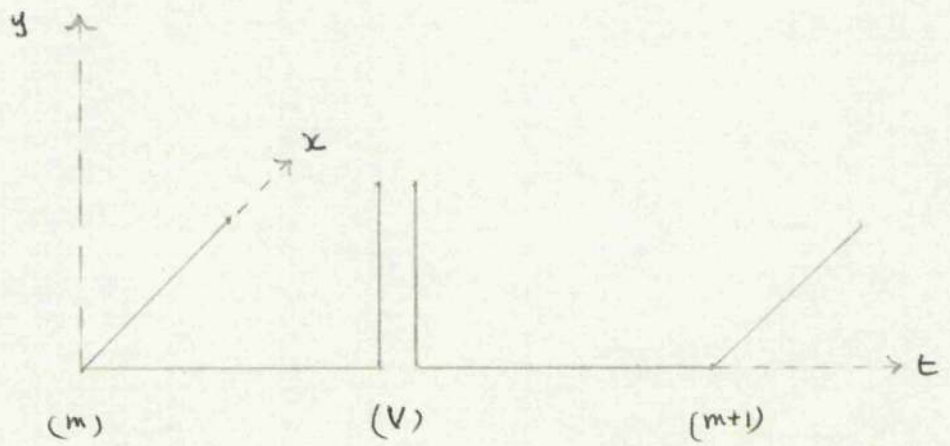


Figure 2.1.2.

We shall refer to this factorization as being of Peaceman-Rachford (P.R) type. In a later section it will be shown that this process requires modification near boundary planes.

The multiplication of both sides of (2.1.8b) by (1-pB) gives

$$(1-pB)[1+\tfrac{1}{2}(1-pA)\Delta_x]u^{m+1}_{i,j} = [1-pB+\tfrac{1}{2}(1+pB)(1-pB)\Delta_y]v^{m+1}_{ij}$$

$$= (1+pB)[1+\tfrac{1}{2}(1-pB)\Delta_y]v^{m+1}_{i,j} - 2pBv^{m+1}_{ij}$$

This is equivalent to the formula

$$(2.1.8c)\quad (1-pB)[1+\tfrac{1}{2}(1-pA)\Delta_x]u^{m+1}_{ij} = -2pBv^{m+1}_{ij} + (1+pB)[1+\tfrac{1}{2}(1+pA)\Delta_x]u^{m}_{ij}.$$

The two steps (2.1.8a) and (2.1.8c) constitute an alternative formulation which we shall refer to as the Douglas-Rachford (D.R.) type factorisation of (2.1.7).

Although the above schemes are of interest in themselves they are of no great practical value as they refer to the scalar form of (1.7.1). In the following sections we shall examine to what extent they carry through to the case when A and B are nxn matrices which may depend on x,y,t, and u is an n-vector function of x,y,t.

2.2 The Vector Case.

Let us now consider the case where A, B are nxn constant matrices and u is an n-vector function of x,y,t. Assume that $u^m_{i,j}$ now represents the vector u evaluated at the point (i,j,m).

Following (2.1.8a) and (2.1.8b) we introduce the pair of P.R. type formulae

$$[I+(eI+fB)\Delta_y]v^{m+1}_{i,j} = [I+(rI+sA)\Delta_x]u^m_{i,j}$$

$$[I+(eI+fA)\Delta_x]u^{m+1}_{i,j} = [I+(rI+sB)\Delta_y]v^{m+1}_{i,j} \qquad (2.2.1)$$

where I is the unit nxn matrix (or matrix operator), and r,s,e,f are scalar coefficients involving the mesh ratio p = k/h. Elimination of v^{m+1}_{ij} leads to the formula

$$[I+(eI+fB)\Delta_y][I+(eI+fA)\Delta_x]u^{m+1}_{ij}$$

$$= [I+(rI+sB)\Delta_y][I+(rI+sA)\Delta_x]u^m_{ij} \qquad (2.2.2)$$

which an evaluation of the products gives

$$[I+(eI+fB)\Delta_y+(eI+fA)\Delta_x+(eI+fB)(eI+fA)\Delta_y\Delta_x]u^{m+1}_{ij}$$

$$= [I+(rI+sB)\Delta_y+(rI+sA)\Delta_x+(rI+sB)(rI+sA)\Delta_y\Delta_x]u^m_{ij}$$

As in the scalar case the values of u, $\Delta_x u$ etc. at the points (i,j,m) and (i,j,m+1) are expanded as Taylor series in terms of u and its derivatives at the node (i,j,m). Derivatives with respect to time are replaced by the relations

$$\frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}$$

$$\frac{\partial^2 u}{\partial t^2} = A^2\frac{\partial^2 u}{\partial x^2} + (AB+BA)\frac{\partial^2 u}{\partial x\partial y} + B^2\frac{\partial^2 u}{\partial y^2} \qquad (2.2.3)$$

where the latter assumes the matrices A and B are constant.

If the coefficients of $h\frac{\partial u}{\partial x}$, $h\frac{\partial u}{\partial y}$, $\frac{h^2}{2}\frac{\partial^2 u}{\partial x^2}$, $h^2\frac{\partial^2 u}{\partial x \partial y}$, $\frac{h^2}{2}\frac{\partial^2 u}{\partial y^2}$ at the node (i,j,m) are in turn equated to zero, the equations

$$
\begin{aligned}
&(p + f - s)A + (e - r)I = 0\\
&(p + f - s)B + (e - r)I = 0\\
&(p^2 + 2pf)A^2 + (f + 2pe - s)A + (e - r)I = 0\\
&(p^2 + 2pf)B^2 + (f + 2pe - s)B + (e - r)I = 0\\
&(\tfrac{1}{2}p^2 + pf)(AB + BA) + (f^2 - s^2)BA + (pe + ef - rs)(A+B) + (e^2 - r^2)I = 0
\end{aligned}
\tag{2.2.4}
$$

are obtained. We seek a solution of equations (2.2.4) for the scalars e,f,r,s which is valid for all matrices A and B. Obviously the following set of equations must be satisfied:-

$$
\begin{aligned}
&p + f - s = 0 \qquad e - r = 0\\
&p^2 + 2pf = 0 \qquad f + 2pe - s = 0\\
&f^2 - s^2 = 0 \qquad pe + ef - rs = 0
\end{aligned}
$$

leading to the unique solution

$$
\begin{aligned}
&e = r = \tfrac{1}{2}\\
&-f = s = \tfrac{1}{2}p
\end{aligned}
$$

Substitution of the above values into (2.2.1) and (2.2.2) leads to the formulae

$$
\begin{aligned}
&[I + \tfrac{1}{2}(I - pB)\Delta_y]v_{ij}^{m+1} = [I + \tfrac{1}{2}(I + pA)\Delta_x]u_{ij}^{m}\\
&[I + \tfrac{1}{2}(I - pA)\Delta_x]u_{ij}^{m+1} = [I + \tfrac{1}{2}(I + pB)\Delta_y]v_{ij}^{m+1}
\end{aligned}
\tag{2.2.5}
$$

and

$$[I + \tfrac{1}{2}(I - pB)\Delta_y][I + \tfrac{1}{2}(I - pA)\Delta_x]u_{ij}^{m+1}$$

(2.2.6)

$$= [I + \tfrac{1}{2}(I + pB)\Delta_y][I + \tfrac{1}{2}(I + pA)\Delta_x]u_{ij}^{m}$$

respectively, where the truncation error is of order h^3.

The extension of the D.R. type factorization (2.1.8a) and (2.1.8c) follows a similar pattern, and is given by

$$[I + \tfrac{1}{2}(I - pB)\Delta_y]v_{ij}^{m+1} = [I + \tfrac{1}{2}(I + pA)\Delta_x]u_{ij}^{n}$$

(2.2.7)

$$(I + pB)[I + \tfrac{1}{2}(I - pA)\Delta_x]u_{ij}^{m+1} = -2pBv_{ij}^{m+1} + (I+pB)[I+\tfrac{1}{2}(I+pA)\Delta_x]u_{ij}^{m}$$

On elimination of v_{ij}^{m+1}, (2.2.7) reduces to (2.2.6).

It is interesting to compare (2.2.6) with Wendroff's formula. Recalling that the latter can be written in the form

$$[I + \tfrac{1}{2}(I - pA)\Delta_x + \tfrac{1}{2}(I - pB)\Delta_y + \frac{1}{4}(I - p(A+B)\Delta_y\Delta_x]u_{ij}^{m+1}$$

$$= [I + \tfrac{1}{2}(I + pA)\Delta_x + \tfrac{1}{2}(I + pB)\Delta_y + \frac{1}{4}(I + p(A+B)\Delta_y\Delta_x]u_{ij}^{m}$$

which is

$$[\{I+\tfrac{1}{2}(I-pB)\Delta_y\}\{I+\tfrac{1}{2}(I-pA)\Delta_x\} - \frac{1}{4}p^2BA\Delta_y\Delta_x]u_{ij}^{m+1}$$

$$= [\{I + \tfrac{1}{2}(I+pB)\Delta_y\}\{I+\tfrac{1}{2}(I+pA)\Delta_x\} - \frac{1}{4}p^2BA\Delta_y\Delta_x]u_{ij}^{n}$$

we see that

$$L_A u - L_W u = \frac{1}{4}p^2BA\Delta_y\Delta_x(u_{ij}^{m+1} - u_{ij}^{n})$$

where L_A, L_W are the difference operators in (2.2.6) and (2.2.8) respectively. Wendroff's scheme was devised initially as a centred difference scheme in order to lend itself to analysis by the Friedrichs energy method. In fact Wendroff's scheme was shown to be convergent for a particular class of problems for any number of

space dimensions, that is for the equation

$$\frac{\partial u}{\partial t} = \sum_{i=1}^{M} A_i \frac{\partial u}{\partial x_i}$$

where $A_i (i = 1, \ldots M)$ are negative definite symmetric matrices, but its stability could be demonstrated in one space dimension only [40]. It should be noted that the one-dimensional analogueof (2.2.6) coincides with Wendroff's one dimensional scheme to give

$$[I + \tfrac{1}{2}(I - pA)\Delta_x]u_i^{n+1} = [I + \tfrac{1}{2}(I + pA)\Delta_x]u_i^n$$

where of course no factorisation is possible.

2.3 The Family of Eight Point Schemes.

The scheme (2.2.6) is in fact only one of a family of possible eight point schemes connecting four points on each of two neighbouring t planes.

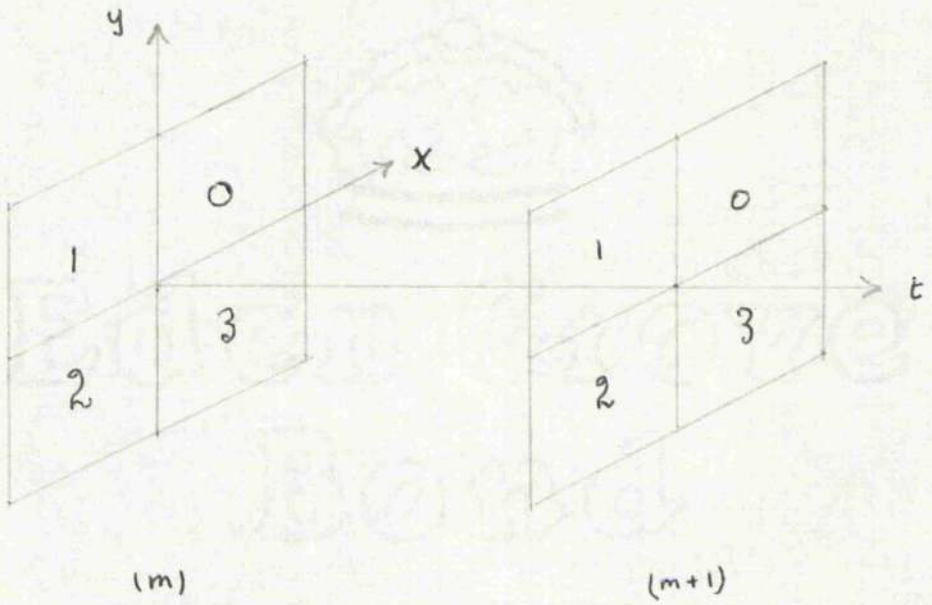


Figure 2.3.1

We define an (a,b) formula (a,b=0,1,2,3) as one which depends on the four points at the corners of square "a" at the time stage (m) together with the four points at the corners of the square "b" at time stage (m+1). (see Figure 2.3.1) The schemes fall into three distinct categories.

(i) (a,a) or "straight-through" type, consisting of

$$(0,0)\left[I+\tfrac{1}{2}(I-pB)\Delta_y\right]\left[I+\tfrac{1}{2}(I-pA)\Delta_x\right]u_{ij}^{n+1}=\left[I+\tfrac{1}{2}(I+pB)\Delta_y\right]\left[I+\tfrac{1}{2}(I+pA)\Delta_x\right]u_{ij}^{n}$$

$$(1,1)\left[I+\tfrac{1}{2}(I-pB)\Delta_y\right]\left[I-\tfrac{1}{2}(I+pA)\nabla_x\right]u_{m+1}$$

$$=\left[I+\tfrac{1}{2}(I+pB)\Delta_y\right]\left[I-\tfrac{1}{2}(I-pA)\nabla_x\right]u_m,$$

$$(2.3.1)\left[I-\tfrac{1}{2}(1+pB)\nabla_y\right]\left[I-\tfrac{1}{2}(I+pA)\nabla_x\right)u_{m+1}$$

$$((2,2))\qquad=\left[I-\tfrac{1}{2}(I-pB)\nabla_y\right]\left[I-\tfrac{1}{2}(I-pA)\nabla_x\right]u_m,$$

$$(3,3)\left[I-\tfrac{1}{2}(I+pB)\nabla_y\right]\left[I+\tfrac{1}{2}(I-pA)\Delta_x\right]u_{m+1}$$

$$=\left[I-\tfrac{1}{2}(I-pB)\nabla_y\right]\left[I+\tfrac{1}{2}(I+pA)\Delta_x\right]u_m,$$

(where of course (0,0) was the scheme derived in section (2.2))

(ii) (a,a±1) modulo 4 or "lateral" type, consisting of

$$(0,1)\left[I+\tfrac{1}{2}(I-pB)\Delta_y\right]\left[I-\tfrac{1}{2}pA\nabla_x\right]u_{m+1}=\left[I+\tfrac{1}{2}(I+pB)\Delta_y\right]\left[I+\tfrac{1}{2}pA\Delta_x\right]u_m,$$

$$(0,3)\left[I-\tfrac{1}{2}pB\nabla_y\right]\left[I+\tfrac{1}{2}(I-pA)\Delta_x\right]u_{m+1}=\left[I+\tfrac{1}{2}pB\Delta_y\right]\left[I+\tfrac{1}{2}(I+pA)\Delta_x\right]u_m,$$

$$(1,0)\left[I+\tfrac{1}{2}(I-pB)\Delta_y\right]\left[I-\tfrac{1}{2}pA\Delta_x\right]u_{m+1}=\left[I+\tfrac{1}{2}(I+pB)\Delta_y\right]\left[I+\tfrac{1}{2}pA\nabla_x\right]u_m,$$

$$(1,2)\left[I-\tfrac{1}{2}pB\nabla_y\right]\left[I+\tfrac{1}{2}(I-pA)\nabla_x\right]u_{m+1}=\left[I+\tfrac{1}{2}pB\Delta_y\right]\left[I+\tfrac{1}{2}(I+pA)\nabla_x\right]u_m,$$

$$(2,1)\left[I-\tfrac{1}{2}pB\Delta_y\right]\left[I-\tfrac{1}{2}(I+pA)\nabla_x\right]u_{m+1}=\left[I+\tfrac{1}{2}pB\nabla_y\right]\left[I-\tfrac{1}{2}(I-pA)\nabla_x\right]u_m,$$

$$(2,3)\left[I-\tfrac{1}{2}(I+pB)\nabla_y\right]\left[I-\tfrac{1}{2}pA\Delta_x\right]u_{m+1}=\left[I-\tfrac{1}{2}(I-pB)\nabla_y\right]\left[I+\tfrac{1}{2}pA\nabla_x\right]u_m,$$

$$(3,0)\left[I-\tfrac{1}{2}pB\Delta_y\right]\left[I+\tfrac{1}{2}(I-pA)\Delta_x\right]u_{m+1}=\left[I+\tfrac{1}{2}pB\nabla_y\right]\left[I+\tfrac{1}{2}(I+pA)\Delta_x\right]u_m,$$

$$(3,2)\left[I-\tfrac{1}{2}(I+pB)\nabla_y\right]\left[I-\tfrac{1}{2}A\Delta_x\right]u_{m+1}=\left[I-\tfrac{1}{2}(I-pB)\nabla_y\right]\left[I+\tfrac{1}{2}pA\nabla_x\right]u_m,$$

(iii) (a,a+2) modulo 4 or "diagonal" type, consisting of

$$(0,2)\left[I-\tfrac{1}{2}pB\nabla_y\right]\left[I-\tfrac{1}{2}pA\nabla_x\right]u_{ij}^{n+1}=\left[I+\tfrac{1}{2}pB\Delta_y\right]\left[I+\tfrac{1}{2}pA\Delta_x\right]u_{ij}^{n}$$

$$(1,3)\left[I-\tfrac{1}{2}pB\nabla_y\right]\left[I-\tfrac{1}{2}pA\Delta_x\right]u_{m+1}=\left[I+\tfrac{1}{2}pB\Delta_y\right]\left[I+\tfrac{1}{2}pA\nabla_x\right]u_m,$$

$$(2,0)\left[I-\tfrac{1}{2}pB\Delta_y\right]\left[I-\tfrac{1}{2}pA\Delta_x\right]u_{m+1}=\left[I+\tfrac{1}{2}pB\nabla_y\right]\left[I+\tfrac{1}{2}pA\nabla_x\right]u_m,$$

$$(3,1)\left[I-\tfrac{1}{2}pB\Delta_y\right]\left[I-\tfrac{1}{2}pA\nabla_x\right]u_{m+1}=\left[I+\tfrac{1}{2}pB\nabla_y\right]\left[I+\tfrac{1}{2}pA\Delta_x\right]u_m,$$

where Δ, ∇ are the forward and backward difference operators defined in section 1.6.

All the above schemes have overall truncation errors of $O(h^3)$, but, whereas each step of type (iii) is a consistent replacement of the differential equation, each step of types (i) and (ii) is not. Therefore the "straight-through" and "lateral" type schemes do not fall into the "general classification of alternating direction methods" proposed by DOUGLAS and GUNN [5].

All the schemes (a,b) may be derived by a procedure identical to that carried through in section 2.2 for the scheme (0,0), and their analogous P.R. and D.R. factorizations obtained.

2.4 Stability.

In this section we shall make an examination of the stability of formula (2.2.6), again for the case of A,B constant symmetric matrices. Making a Fourier transformation of the space variables, as described in the introduction, the result

$$B_2 A_2 \tilde{u}^{n+1}_{ij} = B_1 A_1 \tilde{u}^{n}_{ij}$$

is obtained, where

$$A_1, A_2 = I \cos \frac{\theta h}{2} \pm ipA \sin \frac{\theta h}{2}$$

(2.4.1)

$$B_1, B_2 = I \cos \frac{\gamma h}{2} \pm ipB \sin \frac{\gamma h}{2}$$

and β, γ are arbitrary real numbers.

The amplification matrix G of the method (2.2.6) is therefore given by

$$G = (B_2 A_2)^{-1}(B_1 A_1) = (\bar{B}_1 \bar{A}_1)^{-1}(B_1 A_1)$$

where $\bar{B}_1$ is the complex conjugate of B_1. The problem of stability is that of finding estimates for the bounds of powers of the amplification matrix G. The Lax-Richtmyer sufficient condition for stability requires that

(2.4.2) $$||G^*G|| \leq 1 + O(k)$$

where G^* is the hermitian transpose of G.

Here

(2.4.3) $$G^*G = (\bar{A}_1 \bar{B}_1)(A_1 B_1)^{-1}(\bar{B}_1 \bar{A}_1)^{-1}(B_1 A_1)$$

and thus if AB = BA,

$$G^*G = I$$

and the method is unconditionally stable by the criterion (2.4.2). However in actual problems the commutation condition will rarely be satisfied. A more general analysis is therefore required.

The spectral norm of A_1 will be given by the square root of the maximum modulus eigenvalue of the matrix

$$A_1^* A_1 = I \cos^2 \frac{\beta h}{2} + p^2 A^2 \sin^2 \frac{\beta h}{2}$$

since A is a symmetric matrix. If μ_M^2, μ_m^2 are the maximum and minimum eigenvalues of A^2 it follows that

$$||A_1|| = ||\bar{A}_1|| = (\cos^2 \frac{\beta h}{2} + p^2 \mu_M^2 \sin^2 \frac{\beta h}{2})^{\frac{1}{2}}$$

$$||A_1^{-1}|| = ||\bar{A}_1^{-1}|| = (\cos^2 \frac{\beta h}{2} + p^2 \mu_m^2 \sin^2 \frac{\beta h}{2})^{-\frac{1}{2}}$$

If ν_M^2, ν_m^2 are the maximum and the minimum eigenvalues of B^2, we likewise have

$$||B_1|| = ||\bar{B}_1|| = (\cos^2 \frac{\gamma h}{2} + p^2 \nu_M^2 \sin^2 \frac{\gamma h}{2})^{\frac{1}{2}}$$

$$||B_1^{-1}|| = ||\bar{B}^{-1}|| = (\cos^2 \frac{\gamma h}{2} + p^2 \nu_m^2 \sin^2 \frac{\gamma h}{2})^{-\frac{1}{2}}$$

A property of a norm (11.11) is

$$||XY|| \leq ||X|| \, ||Y||$$

and therefore, on taking norms through (2.4.3) and inserting the above values, the result

$$||G^* G|| \leq ||\bar{A}_1|| \, ||\bar{B}_1|| \, ||B_1^{-1}|| \, ||A_1^{-1}|| \, ||\bar{A}_1^{-1}|| \, ||\bar{B}_1^{-1}|| \, ||B_1|| \, ||A_1||$$

$$= \max_{\beta, \gamma} \left(\frac{\cos^2 \frac{\beta h}{2} + p^2 \mu_M^2 \sin^2 \frac{\beta h}{2}}{\cos^2 \frac{\beta h}{2} + p^2 \mu_m^2 \sin^2 \frac{\beta h}{2}} \right) \cdot \left(\frac{\cos^2 \frac{\gamma h}{2} + p^2 \nu_M^2 \sin^2 \frac{\gamma h}{2}}{\cos^2 \frac{\gamma h}{2} + p^2 \nu_m^2 \sin^2 \frac{\gamma h}{2}} \right)$$

$$= \max_{\beta,\gamma} \left(1 + \frac{p^2(\mu_M^2 - \mu_m^2)\sin^2\frac{\beta h}{2}}{(\cos^2\frac{\beta h}{2} + p^2\mu_m^2\sin^2\frac{\beta h}{2})} \right) \cdot \left(1 + \frac{p^2(\nu_M^2 - \nu_m^2)\sin^2\frac{\gamma h}{2}}{(\cos^2\frac{\gamma h}{2} + p^2\nu_m^2\sin^2\frac{\gamma h}{2})} \right)$$

is obtained. Further since

$$\sin z < z$$

for positive z (and we need only consider positive βh, γh).

We have.

$$||G^*G|| \leq \max_{\beta,\gamma} \left(1 + \frac{\beta^2 k^2(\mu_M^2 - \mu_m^2)}{(\cos^2\frac{\beta h}{2} + p^2\mu_m^2\sin^2\frac{\beta h}{2})} \right) \cdot \left(1 + \frac{\gamma^2 k^2(\nu_M^2 - \nu_m^2)}{(\cos^2\frac{\gamma h}{2} + p^2\nu_m^2\sin^2\frac{\gamma h}{2})} \right)$$

By elementary calculus it can be shown that the minimum value of $(\cos^2\frac{\beta h}{2} + p^2\mu_m^2\sin^2\frac{\beta h}{2})$ is either 1 or $p^2\mu_m^2$ depending on the sign of $(p^2\mu_m^2 - 1)$. In either case we obviously have the result

$$||G^*G|| \leq 1 + O(k^2).$$

By Lax and Richtmyer's condition (2.4.2) this is a (necessary and) sufficient condition for the stability of the scheme (2.2.6). Stability analyses for the remaining members of the family may be carried through in a similar manner to demonstrate their unconditional stability.

However we must not only ensure stability in the time direction but also guarantee that the two point recurrence relations inherent in the alternating direction formulation, are solved in such a manner as to prevent growth of round-off error in the x and y directions.

Consider solving an equation of the form

$$x = Qx + b$$

by a two point recurrence relation of the form

$$x_{n+1} = Qx_n + b$$

On subtraction of the above two equations, it can be seen that the error

$$e_n = x_n - x$$

obeys the two point recurrence relation

$$e_{n+1} = Q e_n .$$

Therefore in order that the errors do not grow, the condition

(2.4.4) $$||Q|| \leq 1$$

must not be imposed.

Consider now the first step of the P-R. formulation (2.2.5)

$$[I + \tfrac{1}{2}(I - pB)\Delta_y] v_{ij}^{m+1} = [I + \tfrac{1}{2}(I + pA)\Delta_x] u_{ij}^{m}$$

which may be written in the form.

$$\tfrac{1}{2}(I - pB) v_{ij+1}^{m+1} + \tfrac{1}{2}(I + pB) v_{ij}^{m+1} = [I + \tfrac{1}{2}(I + pA)\Delta_x] u_{ij}^{m} .$$

The two point recurrence relation may be solved in two directions. For calculation in the positive y-direction, condition (2.4.4) requires

$$||(I - pB)^{-1}(I + pB)|| \leq 1.$$

This is equivalent to requiring B to be negative definite. Reversing the calculation and solving in the negative y-direction will require the positive definiteness of B.

If we consider the first step of the factorized form of the (0,2) member of the family of eight point schemes, that is

$$[I - \tfrac{1}{2}pB\Delta_y] v_{ij}^{m+1} = [I + \tfrac{1}{2}pA\Delta_x] u_{ij}^{n}$$

or equivalently

$$[I - \tfrac{1}{2}pB] v_{ij}^{m+1} + \tfrac{1}{2}pB\, v_{i-1,j}^{m+1} = [I + \tfrac{1}{2}pA\Delta_x] u_{ij}^{n}$$

a similar analysis to the above shows that the condition

(a) is $p\lambda \leq 1$ for calculation in the positive y direction,

(b) is $p\lambda \geq 1$ for calculation in the negative y direction,

where λ is an eigenvalue of B. Similar criteria can be derived for the remaining members of the eight point family.

2.5 Matrices with Variable Coefficients.

The accuracy of the eight point schemes when A and B depend on x,y and t can be verified by direct substitution using Taylor expansions in terms of u and its derivatives at the point (i,j,m). This process will be demonstrated for the (2,0) member of the family, that is for the scheme

$$(2.5.1)\quad [I-p\tfrac{A}{2}\Delta_x - \tfrac{pB}{2}\Delta_y + \tfrac{p^2}{4}B\Delta_y A\Delta_x]u_{ij}^{m+1}=[I+\tfrac{1}{2}pAV_x+\tfrac{1}{2}pBV_y+\tfrac{1}{4}p^2BV_yAV_x]u_{ij}^{m}\ .$$

On expansion in terms of u and its derivatives at the point (i,j,m) up to and including terms of $0(h^2)$, the left hand side becomes

$$[u-\tfrac{1}{2}pA\frac{\partial u}{\partial x} - \frac{pA}{4}h^2\frac{\partial^2 u}{\partial x^2} - \frac{pB}{2}h\frac{\partial u}{\partial y} - \frac{pB}{4}h^2\frac{\partial^2 u}{\partial y^2} + \frac{p^2}{4}h^2B\frac{\partial}{\partial y}(A\frac{\partial u}{\partial x})]_{ij}^{m+1}$$

$$= [u + ph\frac{\partial u}{\partial t} + \frac{p^2h^2}{2!}\frac{\partial^2 u}{\partial t^2} - \frac{ph}{2}[A\frac{\partial u}{\partial x} + ph\frac{\partial}{\partial t}(A\frac{\partial u}{\partial x})] - \frac{pA}{4}h^2\frac{\partial^2 u}{\partial x^2}\ .$$

$$- \frac{ph}{2}[B\frac{\partial u}{\partial y} + ph\frac{\partial}{\partial t}(B\frac{\partial u}{\partial y})] - \frac{pB}{4}h^2\frac{\partial^2 u}{\partial y^2} + \frac{p^2h^2}{4}B\frac{\partial}{\partial y}(A\frac{\partial u}{\partial x})]_{ij}^{m}$$

$$= [u + \frac{ph}{2}[A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}] - \frac{ph^2}{4}(A\frac{\partial^2 u}{\partial x^2} + B\frac{\partial^2 u}{\partial y^2}) + \frac{p^2h^2}{4}B\frac{\partial}{\partial y}(A\frac{\partial u}{\partial x})]_{ij}^{m}$$

which is simply

$$[I + \tfrac{1}{2}pAV_x + \tfrac{1}{2}pBV_y + \tfrac{1}{4}p^2BV_yAV_x]u_{ij}^{m}$$

and thus the accuracy of (2.5.1) is demonstrated in the case of the matrices A and B depending on x,y and t. It should be noted that the matrices on the left hand side of equation (2.5.1) are evaluated at t = m + 1 whereas those on the right hand side are evaluated at t = m.

Applying the above technique to each member of the family of formulae in section 2.3, it can be verified that schemes of type (iii) are still satisfied up to and including terms of order h^3 whereas those

of types (i) and (ii) now have a truncation error of $O(h^2)$. However Wendroff showed that if the coefficient matrices A and B were evaluated not at the point (i,j,s) but at the point $(i+\frac{1}{2},j+\frac{1}{2},s)$ $(s = m,m+1)$ then his scheme had second order accuracy. Since the method developed in section 2.2 differs from Wendroffs formula (even in the case of variable coefficients) by a term of $O(h^3)$ it follows that the eight point schemes of "straight-through" type will have third order truncation error, if we evaluate the matrices, not at the point (i,j,s) but at the point $(i \pm \frac{1}{2}, j \pm \frac{1}{2}, s)$ $(s = m,m+1)$. The signs in $(i \pm \frac{1}{2}, j \pm \frac{1}{2}, s)$ are chosen so that this point is at the centre of the square of nodes at each time level. A similar modification is possible for the "lateral" schemes.

The stability of a difference scheme with slowly varying coefficients is governed by the local amplification matrix. According to LAX [20], the difference scheme is stable if the norm of the local amplification matrix does not exceed one. In the previous section, it was shown that

$$G^*G = I$$

provided

$$AB = BA, \tag{2.5.2}$$

and so (2.5.2) would appear to be the condition for the eight point methods with variable coefficients to be stable. This restrictive condition was also required by Lax and Wendroff in order that their high accuracy explicit scheme be stable for the case of variable coefficients [22]. It is hoped that even when (2.5.2) is not

satisfied, the methods given above can be used in the case of variable coefficients.

2.6 Boundary Modification for Factorised Form.

Consider the application of the factorised form of the (2,2) member of the family of schemes to a numerical problem. This formula is not valid at nodes which are influenced by values on the boundary planes $x = 0$, $y = 0$ (that is at the points o in the diagram Figure 2.6.1).

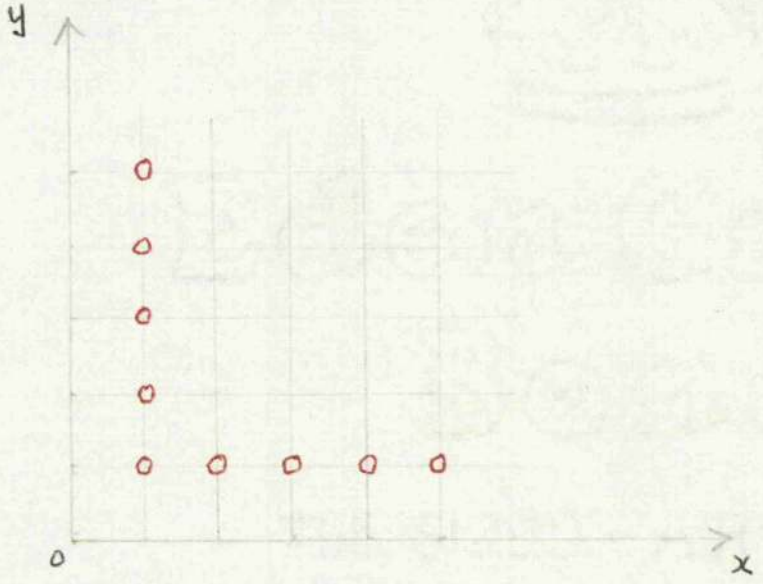


Figure 2.6.1

The totality of equations obtained by using the unsplit form over points on the $(q+1)\times(q+1)$ grid of the quadrant $x,y \geq 0$ of the x,y plane can be written in the form

$$[I - \tfrac{1}{2}(I_n + pB)H_y][I - \tfrac{1}{2}(I_n + pA)H_x]U^{m+1}$$
$$= [I - \tfrac{1}{2}(I_n - pB)H_y][I - \tfrac{1}{2}(I_n - pA)H_x]U^m + \tfrac{1}{2}(I_n + pA)k_1^{m+1} \quad (2.6.1)$$
$$+ \tfrac{1}{2}(I_n + pB)k_2^{m+1} - \tfrac{1}{4}(I_n + pB)(I_n + pA)k_3^{m+1} - \tfrac{1}{2}(I_n - pA)k_1^m$$
$$- \tfrac{1}{2}(I_n - pB)k_2^m + \tfrac{1}{4}(I_n - pB)(I_n - pA)k_3^m$$

where the matrix operators Hx, Hy are the q × q block matrices

$$Hx = \begin{bmatrix} H & & \\ & H & \\ & & \ddots \\ & & & \ddots \end{bmatrix} \quad , Hy = \begin{bmatrix} I_{nq} & & & \\ -I_{nq} & I_{nq} & & \\ & -I_{nq} & & \\ & & & \ddots \end{bmatrix}$$

H being the q × q block matrix

$$\begin{bmatrix} I_n & & & \\ -I_n & I_n & & \\ & -I_n & & \\ & & & \ddots \end{bmatrix}$$

and I the unit matrix of order nq^2. The quantities k_1^m, k_2^m , k_3^m are vectors involving boundary values picked up by ∇_x, ∇_y, $\nabla_y\nabla_x$ respectively at station m in time. U^m represents the totality of vector functions each of order n at the q^2 points on the plane t = mk scanned rowwise.

The totality of equations obtained by using the factorised form of (2,2) and scanning points by rows can be written in the form

$$[I - \tfrac{1}{2}(I_n + pB)Hy]V^{m+1} = [I - \tfrac{1}{2}(I_n - pA)Hx]U^m + C$$

(2.6.2)

$$[I - \tfrac{1}{2}(I_n + pA)Hx]U^{m+1} = [I - \tfrac{1}{2}(I_n - pB)Hy]V^{m+1} + D$$

where C and D depend on the boundary values of u on the planes x = 0, y = 0. Elimination of V^{m+1} gives

(2.6.3) $[I - \tfrac{1}{2}(I_n + pB)Hy][I - \tfrac{1}{2}(I_n + pA)Hx]U^{m+1}$

$= [I - \tfrac{1}{2}(I_n - pB)Hy][I - \tfrac{1}{2}(I_n - pA)Hx]U^m + [I - \tfrac{1}{2}(I_n - pB)Hy]C$

$+ [I - \tfrac{1}{2}(I_n + pB)Hy]D$

Comparison of (2.6.1) and (2.6.3) shows that if

$$C = a_1 k_1^{n+1} + a_2 k_2^{n+1} + a_3 k_3^{n+1} + a_4 k_1^{n} + a_5 k_2^{n} + a_6 k_3^{n}$$

(2.6.4)

$$D = b_1 k_1^{n+1} + b_2 k_2^{n+1} + b_3 k_3^{n+1} + b_4 k_1^{n} + b_5 k_2^{n} + b_6 k_3^{n}$$

then

$$b_1 = -Q(I_n + pB) \qquad b_2 = -Q(I_n + pA) \qquad b_3 = \tfrac{1}{2}Q(I_n + pB)(I_n + pA)$$

$$b_2 = Q(I_n - pB) \qquad b_5 = Q(I_n - pA) \qquad b_6 = \tfrac{1}{2}Q(I_n - pB)(I_n - pA)$$

where

$$Q = (4pB)^{-1}(I + pB)_i \qquad (i = 1,2,\ldots 6)$$

and

$$a_i = -(I - pB)^{-1}(I + pB)b_i \quad (i = 1,2,\ldots 6)$$

Similarly if we choose to use the D.R. type factorisation then (2,2) becomes

$$[I - \tfrac{1}{2}(I_n + pB)Hy]V^{n+1} = [I - \tfrac{1}{2}(I_n - pA)Hx]U^n + (2pB)^{-1}(I_n + pB)K$$

$$[I_n + pB][I - \tfrac{1}{2}(I_n + pA)Hx]U^{n+1} = 2pBV^{n+1} + (I_n - pB)[I - \tfrac{1}{2}(I_n - pA)Hx]U^n$$

where K represents the last six terms on the right hand side of (2.6.1).

The scheme (2,2) was chosen to demonstrate this modification because in a following section it is applied to a simple problem and compared with the Lax Wendroff method.

2.7 An Alternative Factorisation.

In Section (2.1) the P.R. and D.R. type factorisations were introduced. Both are designed to simplify the solution of an implicit scheme of the type

$$HU^{m+1} = GU^m \tag{2.7.1}$$

where H and G may each be written as the product

$$H = H_1H_2, \quad G = G_1G_2$$

of two simpler operators. The P.R. method for the solution of (2.7.1) then takes the form

$$H_1V^{m+1} = G_2U^m$$

$$H_2U^{m+1} = G_1V^{m+1}$$

whereas the D.R. method takes the form

$$H_1V^{m+1} = G_2U^m$$

$$\alpha H_2U^{m+1} = \beta V^{m+1} + \gamma G_2U^m$$

where

$$\beta + \gamma H_1 = \alpha H_2.$$

Both of these methods require modification at boundary planes (see section 2.6). In many cases this modification may be very complex or may not even exist (as in the case for the "lateral" and "diagonal" schemes developed in section 2.3). However a third type of factorization exists which requires straightforward modification and involves very little extra work.

For (2.7.1) this D'JAKONOV [4] type factorization (D-type) becomes

$$H_1V^{m+1} = G_1G_2U^m$$

$$H_2U^{m+1} = V^{m+1}$$

In the notation of section 2.7, the D-type splitting of the (2,2) member of the family of schemes takes the form

$$[I - \tfrac{1}{2}(I_n + pB)V_y]V^{m+1} = [I - \tfrac{1}{2}(I_n - pB)Hy][I - \tfrac{1}{2}(I_n - pA)Hx]U^m$$

$$[I - \tfrac{1}{2}(I_n + pA)V_x]U^{m+1} = V^{m+1}$$

where the boundary values for V^{m+1} in the first step are given by the second step. This is equivalent to the formulation

$$[I - \tfrac{1}{2}(I_n + pB)Hy]V^{m+1} = [I - \tfrac{1}{2}(I_n - pB)Hy][I - \tfrac{1}{2}(I_n - pA)Hx]U^m + C$$

$$[I - \tfrac{1}{2}(I_n + pA)Hx]U^{m+1} = V^{m+1}$$

where C is a contribution from the values of V^{m+1} and is given by

$$C \propto [I - \tfrac{1}{2}(I_n + pA)V_x]U^{m+1}$$

on a boundary at the V^{m+1} level, and

$$C = 0$$

elsewhere.

This procedure extends naturally to more than two space dimensions. For example for three space dimensions it takes the form

$$H_1V^{m+1} = GU^m$$

$$H_2W^{m+1} = V^{m+1}$$

$$H_3U^{m+1} = W^{m+1}$$

where on the respective boundaries

$$V^{m+1} = H_2H_3U^{m+1}$$

$$W^{m+1} = H_3U^{m+1}$$

2.8 Numerical Results.

The method given by (2.6.2) is now applied to a simple problem. Consider the system

$$\frac{\partial u}{\partial t} = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix} \frac{\partial u}{\partial x} + \begin{pmatrix} -1 & \cdot \\ \cdot & -\alpha \end{pmatrix} \frac{\partial u}{\partial y}$$

where the solution

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

is given by

$$u_1 = \sin (x-t) + \sin (y-t)$$

$$u_2 = \sin (x-t) + \cos (y-\alpha t)$$

Two cases are considered,

(i) $\alpha = 1$ where $AB = BA$

(ii) $\alpha = 2$ where $AB \neq BA$

In each case the solution given above will be used to obtain adequate boundary data, so that any difference approximation under consideration can be successfully employed to obtain a solution in a prescribed region. For example we shall choose the region to be the rectangular parallelipiped $0 \leq x, y \leq 1$, $0 \leq t \leq T$, and so the high accuracy explicit Lax Wendroff method (1.7.2) will require data on the five planes $t = 0$; $x,y = 0,1$. The method (2.6.2) will require starting data on the three planes $t = 0$, $x,y = 0$. The danger of overdetermining a well-posed problem in order to solve it by a particular finite difference scheme has been examined by PARTER [25] in the case of one space dimension.

The method of the present chapter with $h = \cdot 1$ requires one

hundred nodal values to be calculated at each half step. These appear in groups of ten and each group involves the step by step solution of a two point recurrence relation

(a) in the positive y direction at the first step

(b) in the positive x direction at the second step.

This two point recurrence must be solved in such a way that there is no significant error growth. This is accomplished easily in the present example as the matrices are both negative definite.

For comparison the method is solved by the most frequently used method, the Lax Wendroff method, which theoretically requires $p \leq 0{\cdot}12$ for stability in this problem. In fact problem (i) was found to be stable for $p \leq 0{\cdot}33$ and problem (ii) for $p \leq 0{\cdot}29$. BURSTEIN [2], using the Lax-Wendroff method to solve the equations of compressible flow in conservation form (see Chapter V)also observed that the Lax-Wendroff theoretical stability limit was too severe.

For $h = 0{\cdot}1$ the method (2.6.2) and the Lax Wendroff method (1.7.2) were allowed to run for 150 time steps in problems (i) and (ii) for $p = 0{\cdot}1, 0{\cdot}3$ and the results are quoted in Tables 2.1 and 2.2. Further results for the alternating direction method for higher values of p, but fewer time steps are shown in Table 2.3. The results are for the errors in u_1 correct to seven places of decimals. The results for u_2 are comparable and are not quoted.

Although the alternating direction method developed in this chapter is of comparable accuracy to the Lax-Wendroff method, it has advantages in its unconditional stability and its requirement of less

boundary data. However, it can only be used for problems where the two point recurrence relations can be solved without growth of round-off error. Further, it requires approximately 20% more computing time than the Lax Wendroff method.

Table 2.1 (Problem (1))

No. of Time Steps P	Method (2,2) 0.1	LAX-WENDROFF 0.3	Method (2,2) 0.3	LAX-WENDROFF 0.3
10	+0.000 1763	-0.000 2708	+0.000 4081	-0.000 6827
20	+0.000 3119	-0.000 8011	+0.000 6381	-0.001 1412
30	+0.000 4809	-0.000 7832	+0.000 6373	-0.001 1776
40	+0.000 6166	-0.000 9333	+0.000 5909	-0.001 1093
50	+0.000 6708	-0.001 0576	+0.000 4640	-0.000 8894
60	+0.000 6903	-0.001 0846	+0.000 3278	-0.000 6475
70	+0.000 6985	-0.001 1059	+0.000 1682	-0.000 3577
80	+0.000 6999	-0.001 1764	-0.000 0960	+0.000 1508
90	+0.000 6935	-0.001 2479	-0.000 2301	+0.000 4128
100	+0.000 6812	-0.001 3057	-0.000 3823	+0.000 6643
110	+0.000 6617	-0.001 3390	-0.000 5402	-0.000 9680
120	+0.000 6361	-0.001 3358	-0.000 6141	+0.001 1164
130	+0.000 6042	-0.001 3168	-0.000 6434	+0.001 1848
140	+0.000 5659	-0.001 2769	-0.000 6095	+0.001 1415
150	+0.000 5209	-0.001 2362	-0.000 5332	+0.001 0117

Table 2.2 (Problem (2))

P No. of Time Steps	Method (2,2) 0·1	LAX-WENDROFF 0·1	Method (2,2) 0·3	LAX-WENDROFF 0·3
10	+0.000 1650	-0.000 3004	+0.000 4041	-0.000 7461
20	+0.000 2840	-0.000 5758	+0.000 5155	-0.001 0186
30	+0.000 4331	-0.000 7759	+0.000 5325	-0.001 0647
40	+0.000 5047	-0.000 9233	+0.000 4789	-0.001 0811
50	+0.000 5484	-0.000 9913	+0.000 3600	-0.002 2701
60	+0.000 5777	-0.001 0234	+0.000 1895	-0.029 9503
70	+0.000 5987	-0.001 0377	-0.000 0566	+0.106 3796
80	+0.000 6113	-0.001 0941	-0.000 2342	unstable
90	+0.000 6142	-0.001 1386	-0.000 3734	thereafter
100	+0.000 6065	-0.001 1552	-0.000 4497	
110	+0.000 5894	-0.001 1458	-0.000 4662	
120	+0.000 5639	-0.001 1148	-0.000 4484	
130	+0.000 5266	-0.001 0624	-0.000 3958	
140	+0.000 4808	-0.000 9899	-0.000 3252	
150	+0.000 4282	-0.000 9009	-0.000 2506	

Table 2.3

P	No. of Time Steps	Problem (1)	Problem (2)
0•8	10	+0.000 2542	+0.000 0532
	20	+0.000 1680	−0.000 0781
	30	−0.000 0350	+0.000 1703
	40	−0.000 1960	−0.000 0972
	50	−0.000 2582	−0.000 4036
1•6	10	−0.000 7226	−0.001 2721
	20	+0.000 8349	+0.001 5666
	30	+0.000 6738	−0.000 5012
	40	−0.000 8743	−0.000 5369
	50	−0.000 6227	−0.000 8760
3.2	10	+0.004 7141	+0.007 6388
	20	−0.004 8478	−0.003 7688
	30	+0.005 0314	+0.006 6249
	40	−0.005 1987	−0.004 8655
	50	+0.005 3484	−0.005 4723

CHAPTER III

METHODS INVOLVING AN EIGHTEEN POINT LATTICE

Gourlay and Mitchell [11]

Introduction.

In this chapter an eighteen point implicit method will be developed for the numerical solution of the first order symmetric hyperbolic system,

$$\frac{\partial u}{\partial t} = A(x,y,t)\frac{\partial u}{\partial x} + B(x,y,t)\frac{\partial u}{\partial y} \quad .$$

This method is of comparable accuracy to the best existing methods and enjoys unconditional stability. It is shown that this scheme may be factorized in three ways, each of which requires the inversion of two block tridiagonal matrices.

3.1 Motivation

In the previous chapter a class of eight point methods was introduced for the numerical solution of the hyperbolic system

$$(3.1.1)\quad \frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}\ .$$

It was shown that these schemes could be written in a factorised form and that they enjoyed unconditional stability, provided certain two-point recurrence relations could be solved without growth of round-off error. This usually required the positive or negative definiteness of the matrices A and B.

In this chapter, we eliminate such restrictions by introducing a scheme based on eighteen points. This scheme requires boundary data on four planes. In fact we are now considering the initial-boundary value problem consisting of (3.1.1) in the region $0 \leqslant x,y \leqslant 1$, $t \geqslant 0$, subject to the initial condition

$$u(x,y,0) = u_0$$

and the boundary conditions

$$u(0,y,t),\ u(1,y,t),\ u(x,0,t),\ u(x,1,t)$$

given for $t \geqslant 0$. It is assumed that there is no discontinuity between the initial and boundary conditions.

The only scheme in use at the moment for the solution of the above problem is the Lax-Wendroff scheme (1.7.2). Although this method is accurate to second order in h, it suffers from the rather severe stability restriction

$$p \leqslant \frac{1}{2\sqrt{2}\,|\lambda m|}$$

where $p = k/h$ is the mesh ratio with h,k the mesh spacings is the space and time coordinates respectively and

$$|\lambda_m| = \max [|\lambda_A|, |\lambda_B|],$$

where

$$|A - \lambda_A I| = 0 \quad |B - \lambda_B I| = 0,$$

I being the $n \times n$ unit matrix. The eighteen point scheme to be introduced in this chapter, enjoys unconditional stability.

3.2 The Eighteen Point Replacement.

The region to be examined $0 \leq x,y \leq 1$, $0 \leq t \leq T$ is covered by a rectangular grid parallel to the coordinate axes, with h,k the space and time increments respectively. Let $p = k/h$ be the mesh ratio. As before, we introduce the notation

$$\Delta_x u_m = u^m_{i+1,j} - u^m_{i,j} \qquad i,j,m \text{ integers}$$

$$\nabla_x u_m = u^m_{i,j} - u^m_{i-1,j} \qquad 0 \leq i,j \leq q+1 \ (= 1/h)$$

$$n \geq 0$$

where Δ, ∇ are forward and backward difference operators respectively.

Consider the eighteen point central difference replacement of (3.1.1) in the form

$$[I_n + a_1(\Delta_x + \nabla_x) + b_1(\Delta_y + \nabla_y) + c_1(\Delta_y + \nabla_y)(\Delta_x + \nabla_x)]u_{m+1} \tag{3.2.1}$$

$$= [I_n + a_2(\Delta_x + \nabla_x) + b_2(\Delta_y + \nabla_y) + c_2(\Delta_y + \nabla_y)(\Delta_x + \nabla_x)]u_m$$

where a_i, b_i, c_i $(i = 1,2)$ are functions of p, A and B, and I_n is the nxn unit matrix. This formula contains two nine point operators at the adjacent time levels m, m+1. (see Figure 3.2.1)

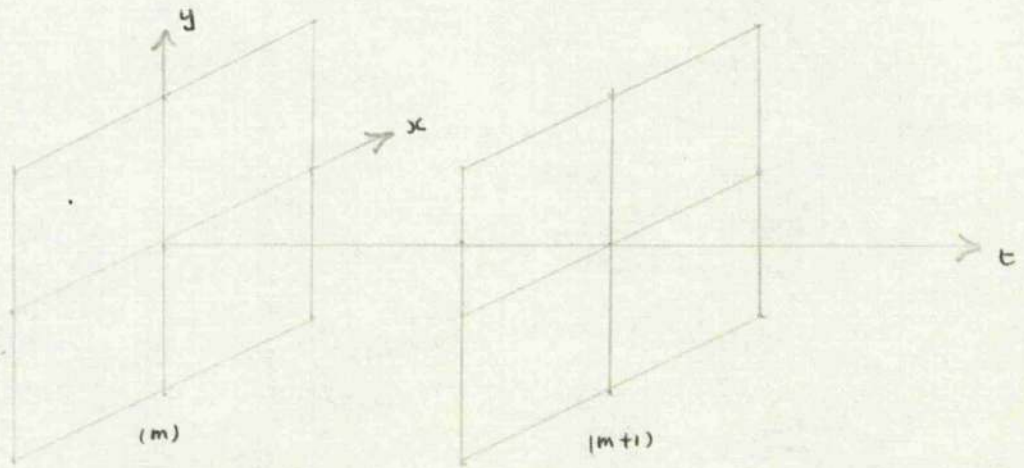


Figure 3.2.1

In order to obtain the values of the coefficients we make the assumption that the matrices A and B are constant. Therefore differentiation of (3.1.1) with respect to t gives the relation

$$(3.2.2)\quad \frac{\partial^2 u}{\partial t^2} = A^2 \frac{\partial^2 u}{\partial x^2} + (AB + BA)\frac{\partial^2 u}{\partial x \partial y} + B^2 \frac{\partial^2 u}{\partial y^2}.$$

The terms in (3.2.1) are now expanded by Taylor's theorem about the node (ih, jh, mk) and after elimination of $\frac{\partial u}{\partial t}$ and $\frac{\partial^2 u}{\partial t^2}$ by means of (3.1.1) and (3.2.2) the coefficients of $h\frac{\partial u}{\partial x}$, $h\frac{\partial u}{\partial y}$, $\frac{1}{2}h^2\frac{\partial^2 u}{\partial x^2}$, $h^2\frac{\partial^2 u}{\partial x \partial y}$, $\frac{1}{2}h^2\frac{\partial^2 u}{\partial y^2}$ are equated to zero. This leads to the following set of equations for the unknown matrix quantities a_i, b_i, c_i $(i = 1,2)$:

$$2a_1 + pA = 2a_2$$

$$2b_1 + pB = 2b_2$$

$$4pa_1A + p^2A^2 = 0$$

$$4pb_1B + p^2B^2 = 0$$

$$\tfrac{1}{2}p^2(AB + BA) + 2p(a_1B + b_1A) + 4(c_1 - c_2) = 0$$

The unique solution of this set of equations is given by

$$a_1 = -a_2 = -\frac{1}{4}pA$$

$$b_1 = -b_2 = -\frac{1}{4}pB$$

$$c_1 = c_2 = C$$

where C is an undetermined parameter. If these values are substituted into (3.2.1) the scheme

$$(3.2.3)\quad [I_n - \tfrac{1}{4}pA(\Delta_x + \nabla_x) - \tfrac{1}{4}pB(\Delta_y + \nabla_y) + c(\Delta_y + \nabla_y)(\Delta_x + \nabla_x)]u_{m+1}$$
$$= [I_n + \tfrac{1}{4}pA(\Delta_x + \nabla_x) + \tfrac{1}{4}pB(\Delta_y + \nabla_y) + c(\Delta_y + \nabla_y)(\Delta_x + \nabla_x)]u_m$$

is obtained. This scheme has truncation error of order h^3.

In the next section the choice of the parameter c will be discussed.

3.3 The Peaceman - Rachford Factorisation.

The choice

$$c = 0$$

in equation (3.2.3) leads to a scheme, illustrated in Figure 3.3.1, which involves only five points at each time level .

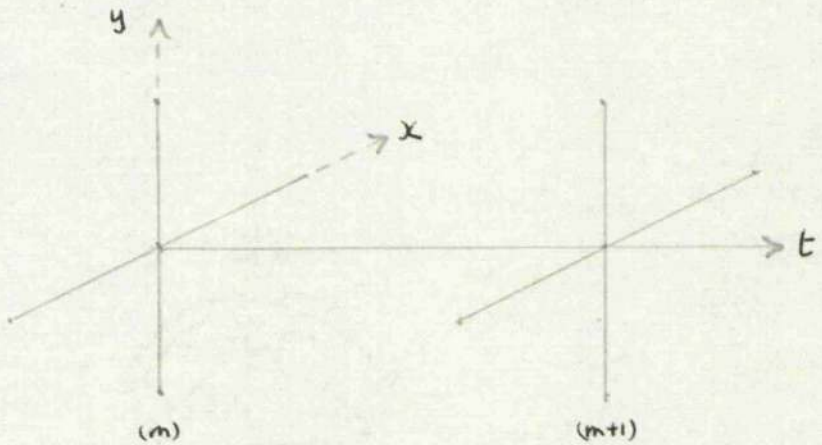


Figure 3.3.1

The actual solution of such a set of implicit equations would involve considerable difficulty and would require to be obtained by some iterative method. [VARGA [38]].

However if we make the choice

$$c = \frac{1}{16} p^2 BA$$

then each side of (3.2.3) can be written as the product of two factors in the form

$$(3.3.1)\quad [I_n - \tfrac{1}{4}pB(\Delta_y + \nabla_y)][I_n - \tfrac{1}{4}pA(\Delta_x + \nabla_x)]u_{m+1}$$
$$= [I_n + \tfrac{1}{4}pB(\Delta_y + \nabla_y)][I_n + \tfrac{1}{4}pA(\Delta_x + \nabla_x)]u_m$$

On introduction of the auxiliary solution v_{m+1} equation (3.3.1) can be written in the two step form

$$[I_n - \tfrac{1}{4}pB(\Delta_y + \nabla_y)]v_{m+1} = [I_n + \tfrac{1}{4}pA(\Delta_x + \nabla_x)]u_m$$

$$[I_n - \tfrac{1}{4}pA(\Delta_y + \nabla_x)]u_{m+1} = [I_n + \tfrac{1}{4}pB(\Delta_y + \nabla_y)]v_{m+1}$$

Each equation in (3.3.2) only requires the inversion of a block tridiagonal matrix to calculate v_{m+1} and u_{m+1} respectively.

As in Chapter II, let us now consider the totality of equations (3.3.2) over the q^2 internal points $(ih, jh)(1 \leq i, j \leq q)$. These equations can be written in the form

$$(3.3.3)\quad [I - \tfrac{1}{4}pBH_y][I - \tfrac{1}{4}pAH_x]U_{m+1} = [I + \tfrac{1}{4}pBH_y][I + \tfrac{1}{4}pAH_x]U_m + \tfrac{1}{4}pA(k_1^m + k_1^{m+1}) + \tfrac{1}{4}pB(k_2^m + k_2^{m+1}) + \tfrac{p^2}{16}BA(k_3^m - k_3^{m+1})$$

where the elements of U_m are the values of u at the q^2 internal mesh points at the m^{th} time level, the grid being scanned row-wise from the $y = 0$ plane, H_x, H_y are the $q \times q$ block matrices

$$H_x = \begin{bmatrix} M & & \\ & M & \\ & & M \\ & & & \ddots \end{bmatrix}, \quad H_y = \begin{bmatrix} 0 & I_{nq} & & \\ -I_{nq} & 0 & I_{nq} & \\ & -I_{nq} & 0 & \\ & & & \ddots \end{bmatrix}$$

with M the $q \times q$ block matrix

$$\begin{bmatrix} 0 & I_n & & \\ -I_n & 0 & I_n & \\ & -I_n & 0 & \ddots \\ & & \ddots & \ddots \end{bmatrix}$$

and the unit matrices I_n, I_{nq}, and I are of order n, nq, nq^2 respectively. The vectors k_1^m, k_2^m, k_3^m, each with q^2 components, constitute the

contributions from the boundary conditions picked up by $(\Delta_x + \nabla_x)u$, $(\Delta_y + \nabla_y)u$ and $(\Delta_y + \nabla_y)(\Delta_x + \nabla_x)u$ respectively at the m^{th} time level. Using (3.3.2) as a guide we now write (3.3.3) in the factorized form,

$$[I - \tfrac{1}{4}pBH_y]V_{m+1} = [I + \tfrac{1}{4}pAH_x]U_m + C$$

(3.3.4)

$$[I - \tfrac{1}{4}pAH_x]U_{m+1} = [I + \tfrac{1}{4}pBH_y]V_{m+1} + D$$

Where V_{m+1} is an intermediate solution and C and D are vectors to be obtained in terms of k_1^s, k_2^s, k_3^s $(s = m, m+1)$. This is accomplished by eliminating V_{m+1} from (3.3.4) to give

$$(3.3.5)\quad [I - \tfrac{1}{4}pBH_y][I - \tfrac{1}{4}pAH_x]U_{m+1} = [I + pBH_y][I + \tfrac{1}{4}pAH_x]U_m + [I + \tfrac{1}{4}pBH_y]C + [I - \tfrac{1}{4}pBH_y]D$$

The comparison of (3.3.5) and (3.3.3) shows that

$$(3.3.6)\quad C = D = \tfrac{1}{8}pA(k_1^m + k_1^{m+1}) + \tfrac{1}{8}pB(k_2^m + k_2^{m+1}) + \tfrac{1}{32}p^2BA(k_3^m - k_3^{m+1})$$

and so (3.3.4) and (3.3.6) constitute a Peaceman - Rachford type alternating direction scheme of second order accuracy for the hyperbolic system (3.1.1)

3.4 Alternative Factorisations.

An alternative D.R. type factorization of (3.2.3) with

$$c = \frac{1}{16}p^2BA$$

exists and can be obtained by a process similar to that given in section 2.1 for the eight point scheme. It is therefore given by the scheme

$$(3.4.1)\quad [I_n - \tfrac{1}{4}pB(\Delta_y + \nabla_y)]v_{m+1} = [I_n + \tfrac{1}{4}pA(\Delta_x + \Delta_x)]u_m$$

$$[I_n - \tfrac{1}{4}pA(\Delta_x + \nabla_x)]u_{m+1} = 2v_{m+1} - [I_n + \tfrac{1}{4}pA(\Delta_x + \nabla_x)]u_m$$

The elimination of v_{m+1} from (3.4.1) gives (3.3.1). If we consider the totality of equations (3.4.1) over the whole region the formulae

$$(3.4.2)\quad [I - \tfrac{1}{4}pBH_y]V_{m+1} = [I + \tfrac{1}{4}pAH_x]U_m + \tfrac{1}{2}K$$

$$[I - \tfrac{1}{4}pAH_x]U_{m+1} = 2V_{m+1} - [I + \tfrac{1}{4}pAH_x]U_m$$

are obtained where K denotes the last three terms in (3.3.3), and the notation is that of section 3.3.

A third factorisation, of the D'JAKONOV type [4], also exists for (3.3.1) and is given by

$$(3.4.3)\quad [I_n - \tfrac{1}{4}pB(\Delta_y + \nabla_y)]v_{m+1} = [I_n + \tfrac{1}{4}pB(\Delta_y + \nabla_y)][I_n + \tfrac{1}{4}pA(\Delta_x + \nabla_x)]u_m$$

$$[I_n - \tfrac{1}{4}pA(\Delta_x + \nabla_x)]u_{m+1} = v_{m+1}$$

This scheme need not be considered over the whole region as "boundary values" at the v_{m+1} level may be replaced by boundary values at the u_{m+1} level, using the second equation in (3.4.3)

For general regions it would appear that the third factorization is of most use. However all calculations carried out employed the Peaceman Rachford factorisation. These calculations will be described in a later section.

3.5 The Factorisations for Variable Matrices.

If the matrices A and B are functions of x and y only, it is an elementary exercise [see section 2.5] to show that the eighteen point scheme (3.3.1) still maintains second order accuracy.

If A and B are functions of x,y and t however, these procedures require modification. We write (3.3.2) in the form

$$[I_n - \tfrac{1}{4}pB_{m+\frac{1}{2}}(\Delta_y + \nabla_y)]v_{m+\frac{1}{2}} = [I_n + \tfrac{1}{4}pA_m(\Delta_x + \nabla_x)]u_m$$

(3.5.1)

$$[I_n - \tfrac{1}{4}pA_{m+\frac{1}{2}}(\Delta_x + \nabla_x)]u_{m+1} = [I_n + \tfrac{1}{4}pB_m(\Delta_y + \nabla_y)]v_{m+\frac{1}{2}}$$

where A_s, B_s $(s = m, m+\frac{1}{2})$ denote the values of the matrices at the points (ih, jh, sk).

As the expression

$$[B_m(\Delta_y + \nabla_y)B_{m+\frac{1}{2}}(\Delta_y + \nabla_y) - B_{m+\frac{1}{2}}(\Delta_y + \nabla_y)B_m(\Delta_y + \nabla_y)]v_{m+\frac{1}{2}}$$

may easily be shown to be of third order in h, it follows that (3.5.1), on elimination of $v_{m+\frac{1}{2}}$, is equivalent to

(3.5.2) $$[I_n - \tfrac{1}{4}pB_{m+\frac{1}{2}}(\Delta_y + \nabla_y)][I_n - \tfrac{1}{4}pA_{m+\frac{1}{2}}(\Delta_x + \nabla_x)]u_{m+1}$$
$$= [I_n + \tfrac{1}{4}pB_m(\Delta_y + \nabla_y)][I_n + \tfrac{1}{4}pA_m(\Delta_x + \nabla_x)]u_m$$

to second order accuracy.

The totality of the equations (3.5.2) over the whole x,y plane is given by

$$[I - \tfrac{1}{4}pB_{m+\frac{1}{2}}H_y][I - \tfrac{1}{4}pA_{m+\frac{1}{2}}H_x]U_{m+1} = [I + \tfrac{1}{4}pB_mH_y][I + \tfrac{1}{4}pA_mH_x]U_m$$
$$+ \tfrac{1}{4}p[A_mk_1^m + A_{m+\frac{1}{2}}k_1^{m+1}] + \tfrac{1}{4}p[B_mk_2^m + B_{m+\frac{1}{2}}k_2^{m+1}] + \tfrac{1}{16}p^2[k_3^m - k_3^{m+1}]$$

where k_1^s, k_2^s are defined as in section 3.3 but k_3^s is now the contribution from the boundary values picked up by

$$B_s(\Delta_y + V_y)A_s(\Delta_x + V_x) \qquad (s = m, m+1)$$

Equations (3.5.1) can now be written in the form

$$[I - \tfrac{1}{4}pB_{m+1}H_y]V_{m+1} = [I + \tfrac{1}{4}pA_mH_x]U_m + C$$

(3.5.4)

$$[I - \tfrac{1}{4}pA_{m+1}H_x]U_{m+1} = [I + \tfrac{1}{4}pB_mH_y]V_{m+1} + D.$$

If we assume

$$C = a_1k_1^{m+1} + a_2k_2^{m+1} + a_3k_3^{m+1} + a_4k_1^m + a_5k_2^m + a_6k_3^m$$

(3.5.5)

$$D = b_1k_1^{m+1} + b_2k_2^{m+1} + b_3k_3^{m+1} + b_4k_1^m + b_5k_2^m + b_6k_3^m,$$

then eliminating V_{m+1} from (3.5.4) and comparing the result with (3.5.3) shows that the coefficients a_i, b_i $(i = 1,\ldots 6)$ are given by the equations

$$B_ma_i = B_{m+1}b_i \qquad (i = 1,\ldots 6)$$

$$a_1 + b_1 = \tfrac{1}{4}pA_{m+1} \qquad a_4 + b_4 = \tfrac{1}{4}pA_m$$

$$a_2 + b_2 = \tfrac{1}{4}pB_{m+1} \qquad a_5 + b_5 = \tfrac{1}{4}pB_m$$

$$a_3 + b_3 = -\tfrac{1}{16}p^2 \qquad a_6 + b_6 = \tfrac{1}{16}p^2,$$

which have the unique solution

$$a_1 = QA_{m+1} \qquad a_4 = QA_m$$

(3.5.6) $$a_2 = QB_{m+1} \qquad a_5 = QB_m$$

$$a_3 = -\tfrac{1}{4}pQ \qquad a_6 = \tfrac{1}{4}pQ$$

where

$$Q = \tfrac{1}{4}p(B_{m+1}+B_m)^{-1}B_{m+1}\ .$$

Therefore the equations (3.5.4), (3.5.5) and (3.5.6) constitute a second order accurate alternating direction method for solving the differential equation (3.1.1) when the matrices are functions of x, y and t.

In the case of the Douglas - Rachford type formulation an analysis similar to the one above shows that (3.4.2) generalises to the scheme

$$[I - \tfrac{1}{4}pB_{m+1}H_y]V_{m+1} = [I + \tfrac{1}{4}A_mH_x]U_m + C \tag{3.5.7}$$

$$[I - \tfrac{1}{4}pA_{m+1}H_x]U_{m+1} = [2I + \tfrac{1}{4}p(B_m - B_{m+1})H_y]V_{m+1} - [I + \tfrac{1}{4}pA_mH_x]U_m + D$$

when A, B are functions of x,y and t. The coefficients of C and D, defined by (3.5.5) now satisfy the equations

$$(B_m - B_{m+1})a_i = B_{m+1}b_i \qquad (i = 1,\ldots 6)$$

$$2a_1 + b_1 = \tfrac{1}{4}pA_{m+1} \qquad 2a_4 + b_4 = \tfrac{1}{4}pA_m$$

$$2a_2 + b_2 = \tfrac{1}{4}pB_{m+1} \qquad 2a_5 + b_5 = \tfrac{1}{4}pB_m$$

$$2a_3 + b_3 = -\tfrac{1}{16}p^2 \qquad 2a_6 + b_6 = \tfrac{1}{16}p^2$$

which have the unique solution

$$a_1 = QA_{m+1} \qquad a_4 = QA_m$$

$$a_2 = QB_{m+1} \qquad a_5 = QB_m$$

$$a_3 = \tfrac{1}{4}pQ \qquad a_6 = \tfrac{1}{4}pQ$$

where

$$Q = (B_{m+1} + B_m)^{-1}B_{m+1}$$

and

$$b_i = B_{m+1}^{-1}(B_m - B_{m+1})a_i.$$

The extension of the D'JAKANOV [4] formulation is straightforward and leads to the scheme

$$[I_n - \tfrac{1}{4}pB_{m+1}(\Delta_y + \nabla_y)]v_{m+1} = [I_n + \tfrac{1}{4}pB_m(\Delta_y + \nabla_y)][I_n + \tfrac{1}{4}pA_m(\Delta_x + \nabla_x)]u_m$$

$$[I_n - \tfrac{1}{4}pA_{m+1}(\Delta_x + \nabla_x)]u_{m+1} = v_{m+1}$$

3.6 Stability

In this section the stability of the alternating direction method (3.3.1) will be examined for the case of constant matrices A and B. Away from the boundaries a Fourier transformations of the space variables is carried out in the usual way and leads to

(3.6.1) $B_2A_2\tilde{U}_{m+1} = B_1A_1\tilde{U}_m$

where

$$A_1A_2 = I \pm \tfrac{1}{2}ipA \sin \beta h$$

$$B_1,B_2 = I \pm \tfrac{1}{2}ipB \sin \gamma h$$

with β,γ arbitrary real numbers. From (3.6.1) the amplification matrix is given by

(3.6.2) $G = (B_2A_2)^{-1}(B_1A_1)$.

In [17], LAX and RICHTMYER show that a sufficient condition for stability is

(3.6.3) $||G^*G|| \leq 1 + O(k)$

as $k, h \to 0$ (but not independently), where G^* is the hermitian transpose of G and $||.||$ denotes the L_2 norm.

In this case

(3.6.4) $G^*G = (\bar{A}_1\bar{B}_1)(A_1B_1)^{-1}(\bar{B}_1\bar{A}_1)^{-1}(B_1A_1)$

where $\bar{A}_1$, $\bar{B}_1$ denote the complex conjugates of A_1, B_1 respectively. If we denote the maximum and minimum eigenvalues of A and B by a_M, a_m, b_M, b_m respectively, it can easily be shown that

$$||A_i|| = ||\bar{A}_i|| = (1 + p^2 a_M^2 \sin^2\beta h)^{\frac{1}{2}}$$

$$||A_i^{-1}|| = ||\bar{A}_i^{-1}|| = (1 + p^2 a_m^2 \sin^2\beta h)^{-\frac{1}{2}}$$

(3.6.5)

$$||B_i|| = ||\bar{B}_i|| = (1 + p^2 b_M^2 \sin^2\gamma h)^{\frac{1}{2}}$$

$$||B_i^{-1}|| = ||\bar{B}_i^{-1}|| = (1 + p^2 b_m^2 \sin^2\gamma h)^{-\frac{1}{2}}$$

Taking norms through (3.6.4) and expanding the right hand side the following is obtained:

$$||G^*G|| \leq ||\bar{A}_i|| \; ||\bar{B}_i|| \; ||B_i^{-1}|| \; ||A_i^{-1}|| \; ||\bar{A}_i^{-1}|| \; ||\bar{B}_i^{-1}|| \; ||B_i|| \; ||A_i||$$

$$= \max_{\beta,\gamma} \left\{ 1 + \frac{p^2(a_M^2 - a_m^2)\sin^2\beta h}{1 + p^2 a_m^2 \sin^2\beta h} \right\} \cdot \left\{ 1 + \frac{p^2(b_M^2 - b_m^2)\sin^2\gamma h}{1 + p^2 b_m^2 \sin^2\gamma h} \right\}$$

As the relation

$$\sin z \leq z$$

is always true for positive z (and we need only consider positive βh, γh) we have

$$||G^*G|| \leq \max_{\beta,\gamma} \left\{ 1 + \frac{k^2(a_M^2 - a_m^2)}{1 + p^2 a_m^2 \sin^2\beta h} \right\} \left\{ 1 + \frac{k^2(b_M^2 - b_m^2)}{1 + p^2 b_m^2 \sin^2\gamma h} \right\}$$

Since the denominator is always greater than or equal to one it follows that

$$||G^*G|| \leq 1 + O(k^2)$$

Therefore by the criterion (3.6.3) the method is stable.

If the condition

(3.6.6) $$AB = BA$$

is satisfied then it is easy to show that

$$G^*G = I$$

and hence

$$||G^*G|| = 1.$$

However the condition (3.6.6) is certainly not necessary for the best constant 1 to appear on the right hand side of (3.6.3). This may be shown by considering (3.1.1) with

$$A = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}, \quad B = \begin{bmatrix} e & f \\ f & -e \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

This represents a pair of hyperbolic equations

$$(3.6.7)\quad \frac{\partial^2 u_i}{\partial t^2} = (a^2+b^2)\frac{\partial^2 u_i}{\partial x^2} + 2(bf+ae)\frac{\partial^2 u_i}{\partial x \partial y} + (e^2+f^2)\frac{\partial^2 u_i}{\partial y^2}, \quad (i = 1,2).$$

If we set

$$s_\beta = \tfrac{1}{2}ip \sin \beta h, \qquad s_\gamma = \tfrac{1}{2}ip \sin \gamma h,$$

then

$$A_1, A_2 = \begin{bmatrix} 1 \pm as_\beta \pm bs_\beta \\ \pm bs_\beta \quad 1 \mp as_\beta \end{bmatrix},$$

$$B_1, B_2 = \begin{bmatrix} 1 \pm es_\gamma \pm fs_\gamma \\ \pm fs_\gamma \quad 1 \mp es_\gamma \end{bmatrix},$$

After a laborious but straightforward calculation, the result

$$G^*G = I$$

is obtained, where G is again given by (3.6.2). Thus the method (3.3.1) is unconditionally stable for the solution of (3.6.7) although A and B do not commute. In this case,

$$a_M = -a_m, \quad b_M = -b_m .$$

For the general case when the matrices depend on x,y and t no theoretical result concerning the stability of (3.3.1) has been obtained.

Nevertheless numerical calculations suggest that the outstanding stability properties of (3.3.1) are present also for the case of variable coefficients.

3.7 Numerical Results.

Each cycle of the alternating direction procedure advocated in this chapter requires the inversion of two block tridiagonal matrices. The procedure for carrying out such an inversion is discussed in the Appendix to this chapter.

The equation used to compare the numerical results obtained by the method (3.3.1) [in the factorised form (3.3.4) and (3.3.6)] and the Lax - Wendroff method (1.7.2) is

$$(3.7.1)\quad \frac{\partial u}{\partial t} = \begin{pmatrix} 1 & \cdot \\ \cdot & -1 \end{pmatrix} \frac{\partial u}{\partial x} + \begin{pmatrix} \cdot & 1 \\ 1 & \cdot \end{pmatrix} \frac{\partial u}{\partial y} \qquad 0 \leqslant x, y \leqslant 1,\ t \geqslant 0.$$

This is a particular case of (3.6.7) with $AB \neq BA$ and u_1, u_2 satisfying the wave equation. Appropriate initial and boundary conditions are chosen to give the theoretical solution,

$$(3.7.2)\quad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \cos(x + t) + \cos(y + t) \\ \cos(x - t) + \cos(y + t) \end{pmatrix}.$$

The theoretical stability condition for the Lax - Wendroff method is

$$(3.7.3)\quad p\,|\lambda| \leqslant \frac{1}{2\sqrt{2}}, \qquad |\lambda| = \max_{A,B} \left\{ \lambda_A, \lambda_B \right\}$$

where

$$|A - \lambda_A I| = 0 \qquad |B - \lambda_B I| = 0.$$

In this case this condition requires

$$p \leqslant 0{\cdot}357.$$

In fact the Lax - Wendroff method remained stable for $p \leqslant 0{\cdot}75$.

The results are shown in Table 3.1 for a range of values of p. Each entry in the table is the difference between the theoretical

solution for u_1, given by (3.7.2), and the value computed by the respective method, at the node $x = y = \frac{1}{2}$, $t = mk$. Values for u_2 behaved similarly and are not quoted.

The accuracy in the two methods is comparable, with the alternating direction method being stable for higher values of p.

As a second example we shall consider the polar form of (3.7.1) namely

$$\frac{\partial u}{\partial t} = \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \frac{\partial u}{\partial r} + \frac{1}{r}\begin{pmatrix} -\sin\theta & \cos\theta \\ \cos\theta & \sin\theta \end{pmatrix} \frac{\partial u}{\partial \theta} \quad . \tag{3.7.4}$$

in the sector $0 \leq r, \theta \leq 1$. For appropriate initial and boundary conditions the solution is given by

$$u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \cos(r\cos\theta + t) + \cos(r\sin\theta + t) \\ \cos(r\cos\theta - t) + \cos(r\sin\theta + t) \end{pmatrix}$$

The elements in the matrix B become infinite at $r = 0$. The eigenvalues of A and B are given by

$$\lambda_A = \pm 1, \quad \lambda_B = \pm \frac{1}{r}$$

and so the stability restriction on the Lax - Wendroff method becomes particularly severe near $r = 0$. The results, shown in Table 3.2(a) for a range of values of p are quoted at the node $r = 0 = \frac{1}{2}$. This time instability developed near $r = 0$ with the Lax Wendroff method for $p > 0{\cdot}2$ whereas the alternating direction method started to lose accuracy near $r = 0$ for $p > 0{\cdot}7$.

The problem given by (3.7.4) with the above solution was solved in the range $1 \leq r, \theta \leq 2$ and results were obtained similar in accuracy

and stability to those obtained from solving (3.7.1). These results are shown in Table 3.2(b) and are quoted at the node $r = 0 = \frac{1}{2}$.

All results quoted are for the difference between the theoretical solution given above and the computed solution for the first component. Values for u_2 were similar and are not quoted.

Table 3.1

h	p	No. of Time Steps	LAX-WENDROFF	A.D.I.
0·1	0·1	10	0.000 1547	+0.00008674
		30	-0.000 6087	+0.00018313
		50	-0.001 0093	+0.00006007
	0.4	10	-0.000 7312	+0.00016019
		30	-0.000 7208	-0.00095591
		50	-0.000 3697	-0.00074668
	0.7	10	-0.000 5372	-0.00033306
		30	-0.000 2560	-0.00061374
		50	+0.000 3957	-0.00075245
	1.0	10	$0(10^{18})$	-0.00089150
		30	-	+0.00005197
		50	-	-0.00003637
0·01	10.0	10	-0.000 6905	-0.00000083
		30	$0(10^{6})$	-0.00000056
		50	-	-0.00000052

Table 3.2

h	p	No. of Time Steps	(a) $0 \leq x, y \leq 1$		(b) $1 \leq x, y \leq 2$	
			LAX-WENDROFF	A.D.I.	LAX-WENDROFF	A.D.I.
0·1	0·1	10	-0.000 6492	-0.000 1061	-0.000 677[illegible]	-0.0002 9742
		30	-0.002 0029	-0.000 3502	-0.001 9548	-0.0008 4665
		50	-0.002 4867	-0.000 4385	-0.002 3687	-0.0007 8985
	0·3	10	-0.005 4269	-0.000 2790	-0.004 1947	-0.0008 3035
		30	$0(10^2)$	+0.000 4436	-0.001 5670	+0.0031 194
		50	-	+0.001 2525	+0.003 1099	+0.0044 085
	0.7	10	-	+0.002 0432	-0.011 6188	+0.0034 368
		30	-	-0.002 0432	+0.005 2006	+0.0040 563
		50	-	-0.004 4844	+0.009 4436	-0.0063 155
	1.0	10	-	+0.004 4406	-0.014 0796	+0.0091 3058
		30	-	+0.010 4198	$0(10^2)$	-0.0090 4150
		50	-	+0.082 7741	-	+0.0035 3400
0·01	10·0	10	$0(10^{39})$	-0.000 5122	-	-0.0006 0988
		30	-	+0.000 4186	-	+0.0000 6830
		50	-	+0.000 2398	-	+0.0003 6043

APPENDIX Inversion of a block tridiagonal matrix.

As observed in section 3.7, the application of the alternating direction method requires the inversion of a block tridiagonal matrix of the form

$$A = \begin{bmatrix} B_1 & C_1 & & & \\ A_2 & B_2 & C_2 & & \\ & \cdot & \cdot & \cdot & \\ & & & & C_{n-1} \\ & & & A_n & B_n \end{bmatrix}$$

where each submatrix A_i, B_i, C_i is of order N. We are concerned with finding the solution z of the system

$$Az = k.$$

If z and k are partitioned into components Z_i and K_i of order N, to correspond to the block formation of the matrix, then a direct method exists for obtaining the Z_i. The procedure can be formulated in two stages. The first step consists of calculating the quantities

$$W_1 = B_1^{-1}C_1 \ , \ G_1 = B_1^{-1}K_1$$

$$W_i = (B_i - A_iW_{i-1})^{-1}C_i \qquad 2 \leq i \leq N-1$$

$$G_i = (B_i - A_iW_{i-1})^{-1}(K_i - A_iG_{i-1}) \qquad 2 \leq i \leq N$$

The second step then obtains the solution components Z_i from the recurrence relation

$$Z_N = G_N \ , \ Z_i = G_i - W_i Z_{i+1} \qquad 1 \leq i \leq N-1$$

This process is commonly referred to as the direct method (VARGA [38])

and it is based on Gaussian elimination.

An alternative, more general method, is that due to SCHECHTER [31]. This is based on the fact that we require (just as in the direct method) to reduce A to the form

(3.9.1) $$A = LU$$

where L and U are partitioned square matrices.

Let us introduce the notation (as used by Schechter)

$$A = \begin{bmatrix} B_1 & C_1 & & & \\ A_2 & B_2 & C_2 & & \\ & & & C_{N-1} & \\ & & A_N & B_N & \end{bmatrix}$$

$$= [A_n, B_n, C_n]_1^N$$

The matrices L and U are restricted to be of the form

$$L = [M_n, I_n, 0]_1^N$$

$$U = [0, D_n, C_n]_1^N$$

The vectors s and k are partitioned as for the direct method.

Comparing the two sides of (3.9.1) we obtain the relations

$$D_1 = B_1$$

$$A_n = M_n D_{n-1}$$

$$B_n = M_n C_{n-1} + D_n \qquad 1 \leq n < N.$$

If the D_n are non-singular then the D_n and M_n are obtained recursively from

$$D_n = B_n - A_n D_{n-1}^{-1} C_{n-1}, \qquad D_1 = B_1$$

$$M_n = A_n D_{n-1}^{-1} \qquad 1 \leq n \leq N$$

To solve for Z let

$$Uz = y,$$

then

$$Ly = k$$

and the y and z may be obtained recursively from

$$y_n = K_n - M_n Y_{n-1}, \qquad 1 \leq n \leq N$$

where

$$y_1 = K_1,$$

and

$$Z_n = D_n^{-1}(y_n - C_n Z_{n+1}) \qquad 1 \leq n < N,$$

with

$$Z_N = D_N^{-1} y_N \ .$$

CHAPTER IV

GENERAL CLASSIFICATION OF METHODS

Gourlay and Mitchell [10]

Introduction

In this chapter an exact two-level difference replacement (and hence solution) will be obtained for the first order symmetric hyperbolic system

$$\frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}$$

Although this difference replacement cannot be used to calculate the solution of the above system, it is shown that all the previously known approximate methods together with those derived in this thesis, can be obtained as approximations to the exact difference replacement.

In the case of commuting matrices it will be shown that $0(h^4)$ accuracy can be obtained whereas in the more usual case of non-commuting matrices only $0(h^2)$ accuracy is possible.

4.1 The Exact Difference Replacement

Consider the hyperbolic system of equations

$$(4.1.1)\quad \frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}$$

where A and B are $n \times n$ real symmetric matrices whose entries may depend on x and y; and u is an n-component vector function of the space coordinates x,y,t. It is the purpose of this chapter to derive an exact difference replacement of (4.1.1) and to obtain approximations to this formula which can be used to solve (4.1.1) together with suitable initial and boundary conditions. We consider such problems as falling into three distinct types:-

(a) the Cauchy problem, where the solution is required in the region $-\infty < x < \infty$, $t > 0$, and initial data is given on the plane $t = 0$.

(b) the initial boundary value problem, where the solution is required in one of the octants of x,y,t space. We will consider for simplicity the octant $x,y,t \geq 0$ and allow initial and boundary data on the three quarter-planes $t=0 (x,y \geq 0)$, $x=0 (y,t \geq 0)$, $y=0 (x,t \geq 0)$.

(c) the mixed initial-boundary value problem where the solution is required in the region $0 \leq x,y \leq 1, t \geq 0$ and boundary data is given on the planes

$x=0,1$ $(0 \leq y \leq 1,\ t \geq 0)$, $y=0,1$ $(0 \leq x \leq 1,\ t \geq 0)$

in addition to initial data on the plane $t=0 (0 \leq x,y \leq 1)$

The region to be examined is covered by a rectangular grid parallel to the coordinate axes with h,k the space and time increments respectively.

The mesh ratio $\rho = k/h$ is assumed to be independent of x,y,t. If we choose the origin of the system to be a mesh point then, as before

$$u(ih, jh, mk) = u^m_{ij}$$

$$\Delta_x u_m = u^m_{i+1,j} - u^m_{ij}$$

$$\nabla_x u_m = u^m_{ij} - u^m_{i-1,j}$$

$$\Delta_y u_m = u^m_{i,j+1} - u^m_{i,j+1}$$

$$\nabla_y u_m = u^m_{i,j} - u^m_{i,j-1}$$

$$\Delta_t u^m_{i,j} = u^{m+1}_{ij} - u^m_{ij}$$

It is easy to show formally that

$$\Delta_s \nabla_s = \Delta_s - \nabla_s$$

$$I + \Delta_s = (I - \nabla_s)^{-1}$$

(4.1.2)

$$\Delta_s^2 + \nabla_s^2 = 2\Delta_s \nabla_s + \Delta_s^2 \nabla_s^2$$

$$\Delta_s^2 - \nabla_s^2 = \Delta_s \nabla_s (\Delta_s + \nabla_s)$$

for $s = x,y$ and I the unit operator. In what follows we shall make no distinction between the unit matrix of order n and its associated unit operator, referring to both as I.

Since

$$u^m_{i+1,j} = \left[1 + h\frac{\partial}{\partial x} + \frac{h^2}{2!}\frac{\partial^2}{\partial x^2} + \cdots\right] u^m_{i,j}$$

$$= e^{h\frac{\partial}{\partial x}} u^m_{ij}$$

by Taylor's theorem, we have

$$\Delta_x u^n_{i,j} = u^n_{i+1,j} - u^n_{ij}$$

$$= \left(e^{h\frac{\partial}{\partial x}} - I \right) u^n_{ij}$$

and therefore, formally we may write

$$e^{h\frac{\partial}{\partial x}} = I + \Delta_x$$

or

$$h\frac{\partial}{\partial x} = \log [I + \Delta_x].$$

Similarly the relations

$$h\frac{\partial}{\partial y} = \log (I + \Delta_y)$$

$$k\frac{\partial}{\partial t} = \log (I + \Delta_t)$$

are obtained. Substituting these expressions into (4.1.1) leads to the relation

$$\log(I + \Delta t) u^n_{ij} = p[A \log (I + \Delta_x) + B \log (I + \Delta_y)] u^n_{ij}$$

and therefore on equating the operators we have

$$(4.1.3) \quad w_{n+1} = \exp \left\{ p[A(x,y)\log(I + \Delta_x) + B(x,y)\log (I + \Delta_y)] \right\} w_n$$

Equation (4.1.3) may also be obtained in a slightly different manner. Since

$$w_{n+1} = e^{k\frac{\partial}{\partial t}} w_n$$

$$= e^{k[A(x,y)\frac{\partial}{\partial x} + B(x,y)\frac{\partial}{\partial y}]} w_n$$

we have, on replacing derivatives by differences the equation (4.1.3). Therefore (4.1.3) is equivalent to a Taylor expansion of u^{n+1}_{ij} about u^n_{ij} and is an exact difference representation of (4.1.1).

4.2 Constant Coefficient Matrices

We now consider the case where A and B are constant non-commuting matrices.

Defining

$$(I + \Delta_x)^{pA} = I + pA\Delta_x + \frac{1}{2!}pA(pA - I)\Delta_x^2 + \dots$$

and $(I + \Delta_y)^{pB}$ similarly, equation (4.1.3) reduces to

$$(4.2.1)\quad u_{n+1} = \tfrac{1}{2}\left\{(I + \Delta_x)^{pA}(I + \Delta_y)^{pB} + (I + \Delta_y)^{pB}(I + \Delta_x)^{pA}\right\} u_n.$$

If use is made of (4.1.2), another useful form of (4.2.1) may be obtained employing both forward and backward differences. This is given by

$$(4.2.2)\quad u_{n+1} = \tfrac{1}{2}\left\{(I + \Delta_x)^{\frac{1}{2}pA}(I - \nabla_x)^{-\frac{1}{2}pA}(I + \Delta_y)^{\frac{1}{2}pB}(I - \nabla_y)^{-\frac{1}{2}pB} + (I + \Delta_y)^{\frac{1}{2}pB}(I - \nabla_y)^{-\frac{1}{2}pB}(I + \Delta_x)^{\frac{1}{2}pA}(I-\nabla_x)^{-\frac{1}{2}pA}\right\} u_n$$

Obviously other more hybrid exact difference representations of (4.1.1) exist. A typical example is given by

$$(4.2.3)\quad u_{n+1} = \tfrac{1}{2}\left\{(I+\Delta_x)^{\frac{1}{2}pA}(I-\nabla_x)^{-\frac{1}{2}pA}(I+\Delta_y)^{pB} + (I+\Delta_y)^{pB}(I+\Delta_x)^{\frac{1}{2}pA}(I-\nabla_x)^{-\frac{1}{2}pA}\right\}$$

However, in general we shall confine our attention to the forms given by (4.2.1) and (4.2.2)

4.3 Approximations using forward differences

The direct determination of the right-hand side of (4.2.1) is obviously impracticable, and so approximations to (4.2.1), which can be used for calculation, are required. These are most simply described in terms of rational approximations involving the two operators Δ_x, Δ_y. All the schemes in this section are based on points in the positive quadrants $x \geqslant ih$, $y \geqslant jh$, at time levels $t = mk$, $(m+1)k$, for a reference point (ih, jh, mk).

If the right-hand side of (4.2.1) is expanded an array of explicit schemes can be obtained. Typical of this array are

$$u_{m+1} = [I + pA\Delta_x + pB\Delta_y]u_m \tag{4.3.1}$$

correct to first differences, and

$$u_{m+1} = [I + pA\Delta_x + pB\Delta_y + \tfrac{1}{2}pA(pA - I)\Delta_x^2 + \tfrac{1}{2}pB(pB - I)\Delta_y^2 + \tfrac{1}{2}p^2(AB + BA)\Delta_x\Delta_y]u_m \tag{4.3.2}$$

correct to second differences. Higher approximations can be obtained in a similar manner.

The simplest wholly implicit scheme is obtained by writing

$$u_{m+1} = [I + \alpha\Delta_x + \beta\Delta_y]^{-1}u_m,$$

expanding the right hand side and comparing with (4.2.1) up to first differences. This yields the scheme

$$[I - pA\Delta_x - pB\Delta_y]u_{m+1} = u_m \tag{4.3.3}$$

More complicated wholly implicit schemes, obtained by further expansion of (4.2.1) as an inverse operator, are less useful for calculation and are mainly of academic interest.

An implicit scheme of interest is obtained from the

approximation

(4.3.4) $u_{m+1} = [I + d\Delta_x + e\Delta_y + f\Delta_y\Delta_x]^{-1}[I + a\Delta_x + b\Delta_y + c\Delta_y\Delta_x]u_m$

If the right hand side of the above is expanded and the coefficients of Δ_x, Δ_y are equated to the corresponding coefficients in the expansion of (4.2.1), the equations

(4.3.5) $$a - d = pA$$
$$b - e = pB$$

are obtained. If, in addition, we require the right hand side of (4.3.4) to factorize into terms involving first differences only, the conditions

(4.3.6) $$ed = f$$
$$ba = c$$

must also be satisfied. Using (4.3.5) and (4.3.6) scheme (4.3.4) takes the form

(4.3.7) $[I + e\Delta_y][I + d\Delta_x]u_{m+1} = [I + (e + pB)\Delta_y][I + (d + pA)\Delta_x]u_m$

which can be written in the P.R. type factorized form as

(4.3.8) $$[I + e\Delta_y]v_{m+1} = [I + (d + pA)\Delta_x]u_m$$
$$[I + d\Delta_x]u_{m+1} = [I + (e + pB)\Delta_y]v_{m+1}$$

where v_{m+1} is an intermediate value, but is not in general an approximation to u at any time station.

In particular, if

(4.3.9) $d = \frac{1}{2}pA, \quad e = \frac{1}{2}pB$

equation (4.3.8) becomes a P.R. type scheme and v_{m+1} becomes an

approximation to $u_{m+\frac{1}{2}}$.

The coefficients of $\Delta x, \Delta y, \Delta z, \Delta y, \Delta x^2, \Delta y^2$ in the expansion of the right-hand side of (4.3.4) are now equated to the corresponding coefficients in the expansion of (4.2.1) and the results

$$a = \tfrac{1}{2}(I + pA) \qquad b = \tfrac{1}{2}(I + pB)$$

$$d = \tfrac{1}{2}(I - pA) \qquad e = \tfrac{1}{2}(I - pB)$$

(4.3.10)

$$c - f = \tfrac{1}{2}p(A + B)$$

obtained. Thus (4.3.4) becomes

(4.3.11) $$[I + \tfrac{1}{2}(I - pA)\Delta x + \tfrac{1}{2}(I - pB)\Delta y + f\Delta y\Delta x]u_{m+1}$$

$$= [I + \tfrac{1}{2}(I + pA)\Delta x + \tfrac{1}{2}(I + pB)\Delta y + c\Delta y\Delta x]u_m$$

where c,f satisfy (4.3.10). If

$$c = \frac{1}{4}[I + p(A + B)]$$

$$f = \frac{1}{4}[I - p(A + B)]$$

equation (4.3.10) is satisfied and (4.3.10) becomes WEIDROFF'S scheme (1.7.3). If, however,

$$c = \frac{1}{4}(I + pB)(I + pA) \qquad f = \frac{1}{4}(I - pB)(I - pA),$$

equation (4.3.10) is again satisfied and (4.3.11) becomes the scheme proposed in chapter II. This scheme can be factorised and rewritten in the form

(4.3.12) $$[I + \tfrac{1}{2}(I - pB)\Delta y]v_{m+1} = [I + \tfrac{1}{2}(I + pA)\Delta x]u_m$$

$$[I + \tfrac{1}{2}(I - pA)\Delta x]u_{m+1} = [I + \tfrac{1}{2}(I + pB)\Delta y]v_{m+1}$$

where v_{m+1} is not an approximation to u_{m+1}. It is fairly safe to assert that (4.3.12) constitutes the only practical method amongst the family of implicit schemes given by equations (4.3.10) and (4.3.11).

Computational details are given in chapter II.

Implicit formulae of higher order of accuracy than (4.3.11) can be derived but they will be of no practical value.

4.4 Approximations using forward and backward differences.

The approximations in this section are based on the exact difference representation (4.2.2), which involves backward as well as forward differences. Most of the schemes for a reference point (ih, jh, mk) involve points symmetrically placed about x = ih, y = jh at time levels t = mk, (m+1)k.

The direct expansion of (4.2.2) correct to first order in differences Δ_x, ∇_x, Δ_y, ∇_y produces the scheme

$$u_{m+1} = [I + \tfrac{1}{2}pA(\Delta_x + \nabla_x) + \tfrac{1}{2}pB(\Delta_y + \nabla_y)]u_m$$

$$= [G]u_m$$

The scheme proposed by GARABEDIAN [7], is a perturbation of this in the form

$$(4.4.1)\quad u_{m+1} = [G + \frac{1}{4}(\Delta_x\nabla_x + \Delta_y\nabla_y)]u_m.$$

If we expand (4.2.2) correct to second differences, we obtain

$$\begin{aligned} u_{m+1} = [I &+ \tfrac{1}{2}(\Delta_x + \nabla_x) + \tfrac{1}{2}pB(\Delta_y + \nabla_y) + \frac{1}{4}pA(\tfrac{1}{2}pA - I)\Delta_x^2 \\ &+ \frac{1}{4}pA(\tfrac{1}{2}pA + I)\nabla_x^2 + \frac{1}{4}pB(\tfrac{1}{2}pB - I)\Delta_y^2 + \frac{1}{4}pB(\tfrac{1}{2}pB + I)\nabla_y^2 \\ &+ \frac{1}{8}p^2(AB + BA)(\Delta_x + \nabla_x)(\Delta_y + \nabla_y) + \frac{1}{4}p^2A^2\Delta_x\nabla_x + \frac{1}{4}p^2B^2\Delta_y\nabla_y]u_m \end{aligned} \qquad (4.4.2)$$

which can be simplified, using relations (4.1.2), to produce

$$\begin{aligned} (4.4.3)\quad u_{m+1} &= [I + \tfrac{1}{2}pA(\Delta_x + \nabla_x) + \tfrac{1}{2}pB(\Delta_y + \nabla_y) + \tfrac{1}{2}p^2A^2\Delta_x\nabla_x + \tfrac{1}{2}p^2B^2\Delta_y\nabla_y \\ &\quad + \frac{1}{8}p^2(AB + BA)(\Delta_x + \nabla_x)(\Delta_y + \nabla_y)]u_m \\ &= [G_{LW}]u_m \end{aligned}$$

This is the Lax Wendroff scheme (1.7.2).

STRANG'S modification [34] of the Lax-Wendroff scheme may be obtained by expanding each of the binomial expressions in (4.2.2) to second order differences. Only those terms which cannot be represented on

a nine point lattice based at the reference point are neglected. It is then easy to show that Strang's scheme takes the form

$$u_{m+1} = [G_{LW} + \frac{1}{8}p^3 \left\{ [AB^2+B^2A](\Delta_x + \nabla_x)(\Delta_y\nabla_y) + [BA^2+A^2B](\Delta_y+\nabla_y)\Delta_x\nabla_x \right\}$$
$$+ \frac{1}{8}p^4[A^2B^2 + B^2A^2]\,\Delta_x\nabla_x\Delta_y\nabla_y]u_m$$

An implicit scheme of interest can be obtained from formula (4.2.2) by considering

$$(4.4.4)\quad u_{m+1} = [I + a\Delta_x + b\Delta_y + ba\Delta_y\Delta_x]^{-1}[I + c\nabla_x + d\nabla_y + dc\nabla_y\nabla_x]u_m$$

If the inverse operator is expanded, the right hand side agrees with (4.4.2) up to terms in second differences if

$$a = \tfrac{1}{2}pA = -c$$
$$b = \tfrac{1}{2}pB = -d.$$

If these values are substituted into (4.4.4), and the latter written as a product of factors, we obtain the scheme,

$$[I - \tfrac{1}{2}pB\Delta_y]v_{m+1} = [I + \tfrac{1}{2}pA\nabla_x]u_m$$

$$(4.4.5)\qquad [I - \tfrac{1}{2}pA\Delta_x]u_{m+1} = [I + \tfrac{1}{2}pB\nabla_y]v_{m+1}$$

where v_{m+1} is an approximation to $u_{m+\frac{1}{2}}$. Three similar schemes to (4.4.5) can be obtained by interchanging

(i) Δ_x and ∇_x,

(ii) Δ_y and ∇_y,

(iii) Δ_x and ∇_x, and Δ_y with ∇_y.

Although (4.4.5) is very similar in appearance to (4.3.8) together with (4.3.9), it is in fact an order of accuracy higher.

These are the "diagonal" schemes of chapter II.

Another implicit approximation to (4.2.2) can be obtained from the

form

$$u_{m+1} = [I + a(\Delta_x + \nabla_x) + b(\Delta_y + \nabla_y) + ba(\Delta_y + \nabla_y)(\Delta_x + \nabla_x)]^{-1}$$
$$[I + c(\Delta_x + \nabla_x) + d(\Delta_y + \nabla_y) + dc(\Delta_y + \nabla_y)(\Delta_x + \nabla_x)]u_m \quad (4.4.6)$$

This agrees with (4.4.2) to second order differences if

$$a = \frac{1}{4}pA = -c$$

$$b = \frac{1}{4}pB = -d$$

and then (4.4.6) can be written in factorized form as

$$[I - \tfrac{1}{4}pB(\Delta_y + \nabla_y)]v_{m+1} = [I + \tfrac{1}{4}pA(\Delta_x + \nabla_x)]u_m$$
$$[I - \tfrac{1}{4}pA(\Delta_x + \nabla_x)]u_{m+1} = [I + \tfrac{1}{4}pB(\Delta_y + \nabla_y)]v_{m+1} \quad (4.4.7)$$

where v_{m+1} is a first order approximation to $u_{m+\frac{1}{2}}$. Scheme (4.4.7) is the scheme studied in detail in chapter III.

If we consider the hybrid formula (4.2.3), an explicit formula accurate to second differences can be obtained by expanding (4.2.3) and simplifying through the use of (4.1.2). The result

$$u_{m+1} = [I + \tfrac{1}{2}pA(I + pA)\Delta_x + \tfrac{1}{2}pA(I-pA)\nabla_x + pB\Delta_y - \tfrac{1}{2}pB(I-pB)\Delta_y^2 + \tfrac{1}{4}p^2(AB+BA)$$
$$(\Delta_x + \nabla_x)\Delta_y]u_m \quad (4.4.8)$$

is obtained.

An implicit approximation to (4.2.3) is obtained by considering

$$u_{m+1} = [I + a\Delta_x + b\Delta_y + ba\Delta_y\Delta_x]^{-1}[I + c\nabla_x + d\Delta_y + dc\Delta_y\nabla_x]u_m \quad (4.4.9)$$

This agrees with (4.2.3) up to second differences if

$$a = -\tfrac{1}{2}pA = -c$$

$$b = -\tfrac{1}{2}pB = -d$$

and (4.4.9) may then be written as

(4.4.10) $[I + \frac{1}{2}(I - pB)\Delta_y]v_{m+1} = [I + \frac{1}{2}pA\nabla_x]u_m$

$$[I - \tfrac{1}{2}pA\Delta_x]u_{m+1} = [I + \tfrac{1}{2}(I + pB)\Delta_y]v_{m+1}$$

where v_{m+1} is not an approximation to u at any time level. This scheme is one of the "lateral" schemes described in chapter II. The remaining members of the "lateral" class of schemes can be obtained by considering th different hybrid exact solutions.

It might be expected that accuracy higher than second order could be obtained by considering a factorised scheme of a more general type:

$$u_{m+1} = \{[I + a\Delta_x + b\nabla_x][I + c\Delta_y + d\nabla_y]\}^{-1}[I+e\Delta_x+f\nabla_x][I+g\Delta_y+h\nabla_y]u_m$$

but, on expanding, it turns out that this increase in accuracy can only be achieved if the matrices A and B commute. This scheme will be considered later.

4.5 Summary of schemes and their accuracy.

At this juncture it is worth gathering together the schemes derived in this chapter which are likely to be of interest, and of practical value in solving the three types of problem mentioned in section 4.1.

The most significant explicit schemes are (4.3.1) and (4.4.1) which are accurate to order h, and (4.3.2), (4.4.3) and (4.4.8) which are accurate to order h^2. Although it is generally accepted that the most important explicit scheme is the Lax Wendroff method (4.4.3), it seems that formulae such as (4.3.2) and (4.4.8) might well be used at extremities of regions where the centred formula (4.4.3) is not applicable.

Implicit schemes of note are (4.3.3), (4.3.8), accurate to order h, and (4.3.11), (4.3.12), (4.4.5), (4.4.10), (4.4.7) accurate to order h^2. Except for schemes (4.3.3) and (4.3.11) (Wendroff's method) all are factorizable schemes. The accuracy of the factorized schemes (4.3.8), (4.4.5), (4.4.7) is order h at the initial step, whereas schemes (4.3.12), (4.4.10) are inconsistent with the differential system (4.1.1) over the first step. Care must be taken to incorporate the boundary conditions of a problem in such a way that the overall accuracy of a factorized scheme is maintained, particularly when the factorized scheme has no accuracy at the first step. This point is fully explained with regard to scheme (4.3.12) in chapter II, and with regard to scheme (4.4.7) in chapter III.

The original alternating direction schemes discussed in a general paper by DOUGLAS and GUNN [5] did not lose accuracy at the intermediate step, and so no boundary modification was required with these schemes.

Let us now consider which schemes are most suited for each problem.

For the Cauchy problem, only explicit schemes are applicable since no boundary data is given. It would appear that the Lax Wendroff method (4.4.3) is the most suitable.

For the initial-boundary value problem the schemes most suited are those based only on forward differences. In this respect the alternating direction method (4.3.12) would appear to be the best.

For the mixed initial-boundary value problem, nearly all the schemes are applicable, with the scheme (4.4.7) combining maximum stability with second order accuracy.

4.6 Variable Coefficients

If the matrices A and B depend on the space variables x and y, we should return to (4.1.3) and use it as a starting point for deriving the approximating difference schemes. It is simpler, however, to take the constant coefficient difference schemes obtained from (4.2.1) and (4.2.2) and attempt to modify them for the case of variable coefficients. In fact all the schemes accurate to first order maintain their accuracy.

Of the schemes accurate to second order, factorized schemes (4.4.5) and (4.4.7) maintain accuracy in the same form, whereas (4.3.2), the Lax Wendroff scheme (4.4.3), and Wendroff's scheme (4.3.11) retain their accuracy only if they are written in a modified form. This consists of replacing terms like $AB\Delta_x\Delta_y$ by $A\Delta_xB\Delta_y$ and so on, or by evaluating the matrices at some other x,y node. (This modification is explained in chapter II for the eight point alternating direction method).

4.7 Stability

We will now consider the stability of the schemes mentioned in section 4.5. It is assumed that the schemes have constant coefficients and that the boundary conditions, if any, are periodic. If these conditions are satisfied then the Fourier transformation of the space variables can be made in the usual way and the result

$$v_{m+1} = G(\beta,\gamma)v_m$$

obtained where G is the amplification matrix of the particular scheme under consideration and β,γ are arbitrary real numbers. Again the condition for stability, the Lax Richtmyer condition requires

$$||G^*G|| \leq 1 + O(k)$$

The only explicit schemes which are likely to be used extensively are the central schemes (4.4.1) and (4.4.3) which have the amplification matrices

$$\tfrac{1}{2}I(\cos\beta h + \cos\gamma h) + ip(A\sin\beta h + B\sin\gamma h)$$

and

$$I - p^2\left\{A^2(1-\cos\beta h) + B^2(1-\cos\gamma h) + \tfrac{1}{2}(AB+BA)\sin\beta h\sin\gamma h\right\} + ip(A\sin\beta h+B\sin\gamma$$

respectively. If A and B are constant, and if

$$|\lambda_m| = \max\left\{|\lambda_A|,|\lambda_B|\right\}$$

where

$$|A - \lambda_A I| = 0$$

$$|B - \lambda_B I| = 0$$

the former scheme is stable if

(4.7.1) $p \leq \dfrac{1}{2|\lambda_m|}$ (GARABEDIAN [7])

and the latter if

(4.7.2) $p \leq \dfrac{1}{\sqrt{2}|\lambda_m|}$ (LAX and WENDROFF [23])

The one quadrant schemes (4.3.1) and (4.3.2) are unlikely to be used as principal schemes in a calculation, but they may be employed at boundaries where the central schemes are not applicable. It can be seen from the amplification matrices, (for example the amplification matrix of (4.3.1) is

$$I - p[A(1-\cos\beta h)+B(1-\cos\gamma h)]+ip[A\sin\beta h+B\sin\gamma h],$$

that these schemes require A,B to be positive definite in order even to be conditionally stable.

When A and B are variable matrices, it appears that (4.7.1) and (4.7.2) still give a guide to the stability of the respective schemes although the Lax-Wendroff formula now requires some restrictive condition like the commutation of A and B [22]. Experimental results obtained by BURSTEIN [2], and also those given in chapters II and III suggest that (4.7.2) is in fact too severe a restriction on the mesh ratio in many cases.

The factorised implicit schemes (4.3.8) (with (4.3.9), (4.3.12), (4.4.5), (4.4.7), (4.4.10)) can all be written in the form

$$B_2A_2v_{m+1} = B_1A_1v_m$$

and so the amplification matrix for any of the above schemes is

$$G = A_2^{-1}B_2^{-1}B_1A_1$$

where the coefficient matrices are respectively

A_1, A_2	B_1, B_2
$(I \mp \frac{1}{2}pA) \pm \frac{1}{2}pAe^{i\beta h}$	$(I \mp \frac{1}{2}pB) \pm \frac{1}{2}pBe^{i\gamma h}$
$\frac{1}{2}[(I \mp pA)+(I \pm pA)e^{i\beta h}]$	$\frac{1}{2}[(I \mp pB) + (I \pm pB)e^{i\gamma h}]$
$(I + \frac{1}{2}pA) - \frac{1}{2}pAe^{\mp i\beta h}$	$(I + \frac{1}{2}pB) - \frac{1}{2}pBe^{\mp i\gamma h}$
$I \pm \frac{1}{2}pAi \sin\beta h$	$I \pm \frac{1}{2}pB\, i \sin\gamma h$
$(I + \frac{1}{2}pA) - \frac{1}{2}pAe^{\mp i\beta h}$	$\frac{1}{2}[(I \mp pB) + (I \pm pB)e^{i\gamma h}]$

If the matrices A and B commute, it can be shown for these implicit schemes that

$$G^*G = I,$$

and so all the implicit factorizable schemes described here are unconditionally stable. The scheme (4.3.8) also requires that A and B be positive definite. When the commutation condition

$$AB = BA$$

does not hold analyses similar to those given in chapters II and III will verify that the schemes are unconditionally stable.

4.8 Numerical solution of implicit schemes.

In order to obtain a numerical solution of (4.1.1) from an implicit scheme, it is desirable to use the latter in factorized form. For the implicit schemes (4.3.8), (4.3.12), (4.4.5), (4.4.10) the method consists of solving a two point formula in the y direction, followed by a similar procedure in the x direction. For the central formula (4.4.7), the individual steps in the y and x directions respectively require the solutions of a three point formula.

In fact, the solution of a two point vector formula involving matrix coefficients may impose conditions on the matrices A and B as shown in chapter II. It is quite easy to show that for methods (4.3.8) and (4.4.5), if the calculations are carried out in the negative y and x directions respectively, the eigenvalues of A and B must satisfy the conditions

$$\lambda_A \geq -\frac{1}{p}, \quad \lambda_B \geq -\frac{1}{p}$$

respectively, in order that round-off errors committed in the calculation do not grow. If the scheme (4.3.12) is used, it appears that the matrices must either be both positive definite, both negative definite, or one positive and one negative definite, whereas scheme (4.4.10) requires B either positive or negative definite and $\lambda_A \geq -\frac{1}{p}$, where the calculation at the second half step is performed in the negative x directio

Although the solution of a three point formula, as in the individual steps of (4.4.7) imposes no conditions on A and B, extra boundary values must be given in order that the solution can be carried out in the y and x directions respectively. (This method can

therefore only be used for the mixed-initial-boundary-value problem). These extra boundary values are also required by the Lax Wendroff scheme in a finite x,y region. The additional boundary information required could well restrict the use of central formulae in many problems.

4.9 Commuting Matrices.

Consider now the case where the matrices A,B in the differential system

$$\frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}$$

are constant and obey the commutation condition

$$AB = BA$$

Although this case is rarely encountered in practice, it will be of theoretical interest to examine approximations to the exact difference solutions (4.2.1) and (4.2.2). These equations now take the simpler forms

(4.9.1) $u_{m+1} = \left\{(I + \Delta_x)^{pA}(I + \Delta_y)^{pB}\right\} u_m$

and

(4.9.2) $u_{m+1} = \left\{(I + \Delta_x)^{\frac{1}{2}pA}(I - V_x)^{-\frac{1}{2}pA}(I + \Delta_y)^{\frac{1}{2}pB}(I - V_y)^{-\frac{1}{2}pB}\right\} u_m$

respectively.

From the form of the above two expressions it can be seen that it will be sufficient in fact to consider approximations to the quantities:

(4.9.3) $G(\Delta_x, pA) = (I + \Delta_x)^{pA}$

(4.9.4) $G(\Delta_x, V_x, pA) = (I + \Delta_x)^{\frac{1}{2}pA}(I - V_x)^{-\frac{1}{2}pA}$.

Consider the scalar function

$$f(z,q) = (1 + z)^q$$

and let us construct the table of Padé approximants to this function (WALL [39]). A Padé approximation is of the form

$$P_t^s(z,q) = \frac{N_s(z,q)}{N_t(z,q)}$$

where $N_s(z,q)$, $N_t(z,q)$ are polynomials in z of degrees s and t respectively

such that

$$f(z,q) = p_t^s(z,q) + O(z^{s+t+1})$$

We will assume that

$$N_s(0,q) = 1 = M_t(0,q)$$

By straightforward algebra, the Padé table for $(1+z)^q$ can be derived and the first few entries are given in Table 4.9.1

$t \backslash s$	0	1	2
0	1	$1-qz$	$1+qz+q\frac{(q-1)}{2}z^2$
1	$1-qz$	$\frac{1+\frac{1}{2}(1+q)z}{1+\frac{1}{2}(1-q)z}$	-
2	$1-qz+\frac{q}{2}(q+1)z^2$	-	$\frac{1+\frac{1}{2}(2+q)z+\frac{1}{12}(1+q)(2+q)z^2}{1+\frac{1}{2}(2-q)z+\frac{1}{12}(1-q)(2-q)z^2}$

Table 4.9.1

The various difference schemes, which are approximants to (4.9.1) are then given by

$$u_{m+1} = (M_t(\Delta x, pA))^{-1}(N_s(\Delta x, pA))(M_t(\Delta y, pB))^{-1}(N_s(\Delta y, pB))u_m.$$

This may be written in the form

$$M_t(\Delta x, pA)M_t(\Delta y, pB)u_{m+1} = N_s(\Delta x, pA)N_s(\Delta y, pB)u_m.$$

The latter scheme now allows a P.R. type factorization

$$M_t(\Delta y, pB)v_{m+1} = N_s(\Delta x, pA)u_m$$

$$M_t(\Delta x, pA)u_{m+1} = N_s(\Delta y, pB)v_{m+1}$$

(provided a boundary procedure exists), or a D-type factorized form

$$M_t(\Delta_y, pB)v_{m+1} = N_s(\Delta_x, pA)N_s(\Delta_y, pB)u_m$$

$$M_t(\Delta_x, pA)u_{m+1} = v_{m+1}$$

It should be noted that, in the commuting case, the above factorizations hold even for explicit and wholly implicit schemes. Thus the P_2^0 entry, which is difficult to use in the non-commuting case, can be applied in the form of a two step factorised scheme illustrated in Figure 4.9.1

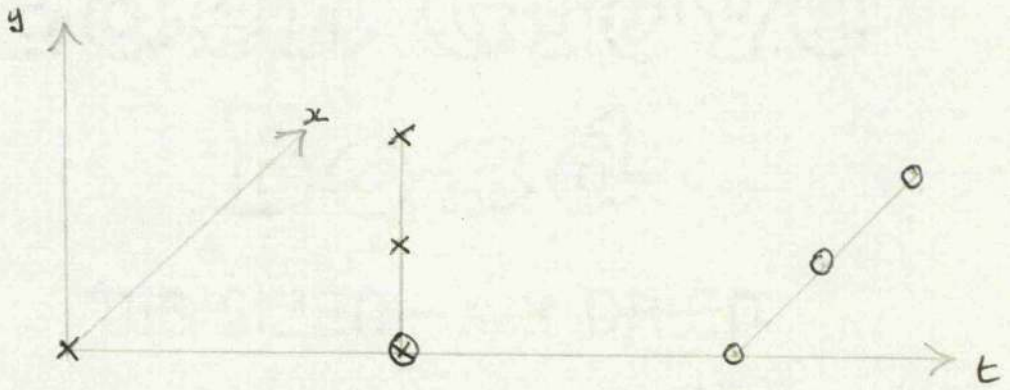


Figure 4.9.1

Apart from this scheme the only new scheme shown in the Padé table is the scheme P_2^2 which is accurate to order h^4. This scheme cannot be used in P.R. form, as a boundary procedure does not exist but may prove useful when used in the D-type form. It can be shown to be unconditionally stable in the Lax-Richtmyer sense.

Let us turn now to those schemes which arise from approximations to (4.9.2) Of course all these derived in section 4.4 carry through

and admit of factorisations. However we shall not repeat such schemes here.

Instead we will consider the previously mentioned, more general approximation to (4.9.4) in the form

$$[I + a\Delta_x + b\nabla_x]^{-1}[I + c\Delta_x + d\nabla_x]$$

It can be shown that a scheme accurate to third order is obtained when

$$a = -d = \frac{1}{6}(I - pA)(I - \frac{pA}{2})$$

$$b = -c = \frac{-1}{6}(I + pA)(I + \frac{pA}{2})$$

Once again we are able to demonstrate unconditional stability in the L-R sense.

Since the condition that the matrices commute is rather an artificial one we will not consider this case when the matrices A and B are functions of x,y and t. Almost certainly the accuracies demonstrated above will no longer be attained.

4.10 Comment

In a series of papers [32 - 37] STRANG considered the simple scalar equation

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0 \quad c \text{ const}(>0).$$

He was able to show that for the solution of an initial boundary value problem by a scheme which was not wholly explicit, stability restricted accuracy to order h^4 and in fact he presented the only two schemes which were stable, and attained this order of accuracy. They correspond to the P_2^2 and P_1^3 entries in the Padé table. Obviously in two space dimensions for the solution of a problem of type (b), we cannot hope, on the basis of Strang's results to obtain a scheme which will be more accurate than order h^4. What is found, however, is that the analogues in two space dimensions of Strang's "best" schemes in one space dimension, hold only for the case of commuting matrices. It would appear, therefore, that the greatest accuracy attainable by a scheme which can be used in a physical situation (where $AB \neq BA$) is restricted to order h^2. Since scheme (4.4.7) combines both this accuracy with unconditional stability it would appear to be the "best" attainable.

It should be noted that the above remarks apply only to two-level schemes. There remains to be considered the family of multi-level schemes. This will not be attempted in this thesis.

CHAPTER V

THE NONLINEAR CASE

Introduction

In this chapter the numerical solution of the hyperbolic system of conservation laws

$$\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = 0$$

where $f = f(u)$ and $g = g(u)$, will be considered.

New methods will be introduced which have satisfactory stability characteristics. The first is an explicit predictor-corrector method which is a generalisation of a two-step formulation of the Lax Wendroff method for nonlinear systems. The remaining methods are nonlinear analogues of schemes considered in chapters II and III. The smooth incorporation of boundary data will be considered in an effort to minimise the effects of nonlinear instability as reported by PHILLIPS [28] and RICHTMYER and MORTON [30].

5.1 The Differential system.

The first order non-linear partial differential equation written in the divergence form

$$\frac{\partial u}{\partial t} + \sum_{i=1}^{M} \frac{\partial f_i}{\partial x_i} = 0$$

is termed a conservation law. It expresses the fact that the quantity of u contained in any M dimensional hypervolume V of x-space, changes at a rate equal to the flux of the vector $(f_1,\ldots f_M)$ into v

$$\frac{d}{dt}\int_V u\,dx = \int_S f.\nu\,ds$$

where s is the bounding surface of v, and ν is the normal to the surface drawn inwards. A large number of physical laws are conservation laws, the quantities u and f depending on the variables which describe the state of the physical system, and perhaps on their derivatives. In particular if we consider theories which ignore such dissipative phenomena as viscosity and heat conduction, then the quantities u and f are functions of the state variables but not of their derivatives.

In this chapter we shall be concerned with the cases M = 1 and 2, although it is most likely that some of the methods may be generalised to an arbitrary number of space dimensions. The case M = 1 will be used to serve as an introduction to the case M = 2, although even in the former case some new schemes will be introduced.

If we choose the state variables as the components of u then the one dimensional system can be put in the form

(5.1.1) $\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} = 0 \qquad u(x,0) = \phi(x)$

where u is a vector of n components and f(u) is a vector valued function of the components of u. If we carry out the differentiation in (5.1.1), the quasilinear first order system of partial differential equations

(5.1.2) $\frac{\partial u}{\partial t} + A(u)\frac{\partial u}{\partial x} = 0,$

is obtained, where

$$A(u) = \begin{bmatrix} \frac{\partial f_1}{\partial u_1}, \ldots . \frac{\partial f_1}{\partial u_n} \\ \vdots \\ \frac{\partial f_n}{\partial u_1} \ldots . \frac{\partial f_n}{\partial u_n} \end{bmatrix}$$

is the Jacobian matrix, or gradient of f with respect to the components of u. If a general system of equations of the form (5.1.2) can be written in the form (5.1.1), then we say that it can be put into conservation form. The requirement that the matrix A has n real distinct eigenvalues μ_i (i = 1,...n) for all values of u, ensures that the system (5.1.2) is hyperbolic. The quantities μ_i(i = 1,...n) are the local sound speeds.

We are interested in the initial value problem consisting of (5.1.1) or (5.1.2) together with the initial data

$$u(x,0) = \phi(x)$$

prescribed on the line t = 0. In practice problems arising from physical phenomena are likely to introduce boundary conditions on x = a,b say.

The two dimensional analogues of (5.1.1) and (5.1.2) are given by

$$\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = 0 \tag{5.1.3}$$

and

$$\frac{\partial u}{\partial t} + A(u)\frac{\partial u}{\partial x} + B(u)\frac{\partial u}{\partial y} = 0 \tag{5.1.4}$$

respectively, where A and B are the Jacobian matrices of f and g with respect to the components of u. The assumption that A and B have real distinct eigenvalues, and that they may be simultaneously symmetrized ensures the hyperbolicity of (5.1.4).

5.2 Examples of Conservation Laws.

The simplest example of a conservation law in one space dimension is furnished by the scalar equation

$$(5.2.1)\quad \frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = 0 \qquad u(x,0) = \phi(x)$$

This may be written in the conservation form

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(\tfrac{1}{2}u^2) = 0$$

and it is a simple matter to verify that solutions of (5.2.1) satisfy the equation

$$(5.2.2)\quad u = \phi(x - ut)$$

The more general equation of the form (5.2.1), namely

$$\frac{\partial u}{\partial t} + a(u)\frac{\partial u}{\partial x} = 0 \qquad u(x,0) = \phi(x)$$

can likewise be shown to have a solution which satisfies

$$u = \phi(u - a(u)t)$$

A solution which satisfies the above equation is termed a <u>soft solution</u> of the equation

$$\frac{\partial u}{\partial t} + a(u)\frac{\partial u}{\partial x} = 0, \quad u(x,0) = \phi(x)$$

It is well-known that no smooth solution will in general exist for all time because of the non-linearity of (5.1.1) [21]. Instead we have to seek <u>weak solutions</u> defined by the requirement that

$$\iint (w_t u + w_x f)\,dxdt + \int w(x,0)\phi(x)\,dx = 0$$

be satisfied for all smooth functions w which vanish for $|x|+t$ large enough. The concepts of soft and weak solutions are discussed in LAX [18], and NOH and PROTTER [24].

Consider now the equations of one dimensional gas flow, in

the form

$$(5.2.5)\qquad \frac{\partial w}{\partial t} + \begin{vmatrix} u & \rho & 0 \\ 0 & u & \gamma p \\ 0 & \frac{1}{\rho} & u \end{vmatrix} \frac{\partial w}{\partial x} = 0$$

with

$$w = \begin{pmatrix} \rho \\ p \\ u \end{pmatrix},$$

where ρ, p, u denote density, pressure and velocity, and γ is the ratio of the specific heats. This system is not in conservation form, but a transformation to the new state variables

$$\rho = \rho, \quad e = \tfrac{1}{2}\rho u^2 + p/(\gamma-1), \quad m = \rho u$$

allows us to write the equations of one dimensional gas flow in the conservation form

$$(5.2.6)\qquad \frac{\partial W}{\partial t} + \frac{\partial f}{\partial x} = 0$$

where

$$W = \begin{pmatrix} \rho \\ e \\ m \end{pmatrix},$$

and

$$f = \begin{vmatrix} m & & \\ \frac{\gamma e m}{\rho} - \frac{\gamma-1}{2}\frac{m^2}{\rho^2} & & \\ (\gamma-1)e - \frac{\gamma-1}{2}\frac{m^2}{\rho} & & \end{vmatrix}$$

Notice that this formulation displays the laws of conservation of mass, energy and momentum.

In fact, Friedrichs has shown that even if we introduce magnetohydrodynamic terms into the compressible flow equations, the latter, now called the Landquist equations, can be written in conservation form. Numerical schemes for the integration of (5.2.6) have been developed in [24], [8].

In two space dimensions similar simple nonlinear systems exist. For example the equations

$$\left.\begin{aligned} &\frac{\partial u_i}{\partial t} + a(u_1)\frac{\partial u_i}{\partial x} + b(u_2)\frac{\partial u_i}{\partial y} = 0 \\ &u_i(x,y,0) = \phi_i(x,y) \end{aligned}\right\} \quad (i = 1,2)$$

have the soft solutions

$$u_i = \phi_i(x - a(u_1)t, y - b(u_2)t) \qquad (i = 1,2)$$

Also the equations of two dimensional gas flow may be written in the conservation form

$$\frac{\partial W}{\partial t} + \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = 0$$

where

$$W = \begin{bmatrix} \rho \\ m \\ n \\ e \end{bmatrix},$$

$$f = \begin{bmatrix} m \\ (\gamma-1)e - \frac{\gamma-3}{2}\frac{m^2}{\rho} - \frac{\gamma-1}{2}\frac{n^2}{\rho} \\ \frac{mn}{\rho} \\ \frac{\gamma e m}{\rho} - \frac{\gamma-1}{2}\left(\frac{m^3 + mn^2}{\rho^2}\right) \end{bmatrix}$$

$$g = \begin{bmatrix} n \\ \frac{mn}{\rho} \\ (\gamma-1)e - \frac{\gamma-3}{2}\frac{n^2}{\rho} - \frac{\gamma-1}{2}\frac{m^2}{\rho} \\ \frac{\gamma e n}{\rho} - \frac{\gamma-1}{2}\left(\frac{n^3 + nm^2}{\rho^2}\right) \end{bmatrix}$$

with $m = \rho u$, $n = \rho v$, $e = \frac{1}{2}\rho(u^2+v^2) + p/\gamma-1$, and u,v the two velocity

components. As before ρ and p denote the density and pressure respectively. These equations and their numerical integration are the subject of a paper by BURSTEIN [2].

5.3 Explicit One Dimensional Case.

Explicit schemes for the numerical solution of the system (5.1.1) have been proposed by LAX and WENDROFF [19], and RICHTMYER [29]. In the latter paper, (which is an excellent review of schemes for the solution of (5.1.1)) Richtmyer rewrites the LAX-WENDROFF scheme as a two-step process:

$$u_i^{*n+1} = \tfrac{1}{2}(u_{i+1}^n + u_{i-1}^n) - {}^p/_4(f_{i+1}^n - f_{i-1}^n)$$

(5.3.1)

$$u_i^{n+1} = u_i^n - {}^p/_2(f_{i+1}^{*n+1} - f_{i-1}^{*n+1})$$

where, as before, $u_i^{n+1} = u(ih,mk)$ and u_i^{*n+1} is an intermediate value. In fact, in the above formulation, u_i^{*n+1} is a first order approximation to $u_i^{n+\frac{1}{2}}$. The overall method is correct to second order.

If we use the approximation

$$u_i^{*n+1} = \tfrac{1}{2}(u_{i+1}^n + u_{i-1}^n) - {}^p/_2(f_{i+1}^n - f_{i-1}^n) \tag{5.3.2}$$

then we are predicting the value of u_i^{n+1}, correct to order h.

The scheme

$$u_i^{n+1} = au_i^n + (1-a)u_i^{*n+1} - {}^p/_2[b(f_{i+1}^{*n+1} - f_{i-1}^{*n+1}) + (1-b)(f_{i+1}^n - f_{i-1}^n)] \tag{5.3.3}$$

is a general corrector formula to be used in conjunction with (5.3.2). It may be regarded as a generalisation of the second step in (5.3.1). If u_i^{*n+1} is eliminated from (5.3.3) by using (5.3.2) then for the resulting formula to be accurate to order h^2 the conditions

$$a = 1 \qquad b = \tfrac{1}{2}$$

must be satisfied. These conditions are obtained by the usual Taylor expansions in terms of u and its derivative at the point (ih,mk). Thus equations (5.3.2) and (5.3.3) together with the above choice of a and b give the second order correct predictor-corrector scheme:

$$(5.3.4)\quad u_i^{*n+1} = \tfrac{1}{2}(u_{i+1}^n + u_{i-1}^n) - p/2(f_{i+1}^n - f_{i-1}^n)$$
$$u_i^{n+1} = u_i^n - p/4[(f_{i+1}^{*n+1} - f_{i-1}^n) + (f_{i+1}^n - f_{i-1}^n)]$$

To investigate the behaviour of a nonlinear finite difference scheme with respect to growth of round-off error, we assume that the concept of stability is a local one, and therefore consider the stability of the linearized version of (5.3.4). On elimination of u_i^{*n+1} it can be shown that (5.3.4) is equivalent in the linearized case to the scheme

$$(5.3.5)\quad u_i^{n+1} = u_i^n - \frac{pA}{8}[u_{i+2}^n - u_{i-2}^n + 2(u_{i+1}^n - u_{i-1}^n)] + \frac{p^2A^2}{8}(u_{i+2}^n - 2u_i^n + u_{i-2}^n)$$

where A is now a constant matrix. Taking Fourier transforms with respect to the space variables in the usual way we obtain the amplification matrix

$$G = 1 - \frac{pA}{4}i[\sin 2\beta h + 2\sin\beta h] - \frac{p^2A^2}{2}\sin^2\beta h,$$

and so

$$G^*G = 1 - (pA)^2\sin^2\beta h \sin^2\frac{\beta h}{2} - \frac{(pA)^2}{4}\sin^4\beta h(1 - p^2A^2)$$

(where we have assumed A is symmetric).

The eigenvalues of G^*G are less than or equal to one, and the method is stable, if

$$p|\lambda_A| \leq 1$$

By a similar anlysis it can be shown that the latter condition is that required for stability of the Lax-Wendroff scheme (5.3.1). However, the advantage of the formulation (5.3.4) will be demonstrated when we turn to the two space dimension case in section 5.8.

The actual implementation of schemes of the form (5.3.1) or (5.3.4) involves certain difficulties near boundary planes. Methods of incorporating the boundary data will be considered in section 5.5.

All the above methods are in the conservation difference form

$$u_{n+1} = u_n + D_x^h F$$

where D_x^h is a central difference operator and F is a function of the points such that if all the arguments are put equal, the function reduces to f. Lax has shown that the solution of a difference equation written in conservation form is a weak solution of the conservation law [18].

5.4 Implicit One Dimensional Case.

In a paper mainly concerned with the numerical solution of problems in fluid dynamics, [8], GARY introduced implicit predictor corrector schemes which are modifications of the basic scheme

$$u_i^{*m+1} = u_i - (P/2)A(u_i^m)(u_{i+1}^m - u_{i-1}^m)$$

(5.4.1)

$$u_i^{m+1} + (P/4)A\left(\frac{u_i^m + u_i^{*m+1}}{2}\right)\left(u_{i+1}^{m+1} - u_{i-1}^{m+1}\right)$$

$$= u_i^m - (P/4)A\left(\frac{u_i^m + u_i^{*m+1}}{2}\right)\left(u_{i+1}^m - u_{i-1}^m\right)$$

The above scheme is derived from the second order implicit scheme

(5.4.2) $u_i^{m+1} + P/4(f_{i+1}^{m+1} - f_{i-1}^{m+1}) = u_i^m - P/4(f_{i+1}^m - f_{i-1}^m)$

and has unconditional stability, as may easily be verified. Scheme (5.4.2) may not be used in general as it requires the solution of a nonlinear three point recurrence relation.

Considering (5.4.2), it would appear that a more obvious predictor-corrector version of it is

(5.4.3) $u_i^{m+1} + P/4\,[Q(u_{i+1}^{*m+1})u_{i+1}^{m+1} - Q(u_{i-1}^{*m+1})u_{i-1}^{m+1}]$

$$= u_i^m - P/4\,(f_{i+1}^m - f_{i-1}^m)$$

where u_{i+1}^{*m+1} is a predicted value of u_{i+1}^{m+1}, accurate to order h, and

$$f(n) = Q(u).u.$$

The overall scheme again enjoys unconditional stability and may be extended to two space dimensions in a slightly modified form (see section 5.7).

An interesting point is that u_{i+1}^{*m+1} may be obtained by at least two predictors:

$$(5.4.4)\quad u_i^{*n+1} = u_i^n - {}^p/_2[f_{i+1}^n - f_{i-1}^n]$$

or

$$(5.4.5)\quad u_i^{*n+1} = \tfrac{1}{2}(u_{i+1}^n + u_{i-1}^n) - {}^p/_2[f_{i+1}^n - f_{i-1}^n]$$

The former is always unstable whereas the latter is stable if the Courant-Freidrichs-Lewy condition is satisfied. In the solution of first order ordinary differential equations of the form

$$y' = f(x,y)$$

by predictor-corrector methods, it is well known that if the predictor is unstable, the corrector must be iterated, in order that the overall method be stable (see [12]). However we shall mainly be interested in the predictor (5.4.5).

Trouble may again be encountered near boundary planes but suitable smoothing (see section 5.5) may minimise its effect.

An important difference between (5.4.1) and (5.4.3) is that the latter is in conservation form whereas the former is not.

5.5 Smoothing of Boundary Data.

Consider the application of (5.3.4) to a problem with boundary data given on $x = 0, a$. We impose a grid on the region with $Nh = a$, where h is the mesh spacing, and so u takes the values u_0^0, u_0^1, u_0^2, ... and u_N^0, u_N^1, u_N^2, ... at nodes on $x = 0$ and $x = a$ respectively. We now illustrate the problem of smoothing of boundary data with respect to the calculation of u_{N-1}^{m+1} (see Figure 5.5.1).

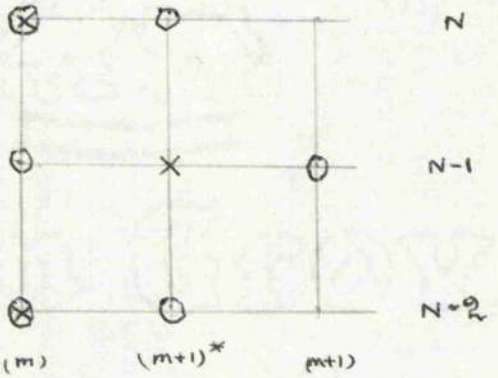


Figure 5.5.1

The predictor

$$(5.5.1)\quad u_{N-1}^{*m+1} = \tfrac{1}{2}(u_N^m + u_{N-2}^m) - \tfrac{1}{2}p(f_N^m - f_{N-2}^m)$$

calculates the approximate value u_{N-1}^{*m+1} in terms of the known boundary values u_N^m and f_N^m. The corrector

$$u_{N-1}^{m+1} = u_{N-1}^m - p/4\,[(f_N^{*m+1} - f_{N-2}^{*m+1}) + (f_N^m - f_{N-2}^m)]$$

however requires the boundary value f_N^{*m+1} (or u_N^{*m+1}) which of course is not known. This difficulty which is present in the calculation of $u_i^j (i = 1, N-1; j = 1, 2, 3..)$ using (5.3.4) may be overcome by several

procedures.

First, we may set $u_N^{*m+1} = u_N^{m+1}$. This however introduces, in the corrector formula an error which tends to spread through the region with time. This may be overcome by iterating the corrector until the iterates to the values u_i^{m+1} are "smooth" or convergent. In difference operator notation, this generates the scheme

$$u^*_{m+1} = \tfrac{1}{2}(\Delta_x - \nabla_x + 2)u_m - {}^p/_2(\Delta_x + \nabla_x)f_m$$

(5.5.2)

$$u_{m+1}^{(k+1)} = u_m - {}^p/_4[\Delta_x + \nabla_x](f_{m+1}^{(k)} + f_m) \qquad k = 1,\ldots K$$

where $f_{m+1}^{(1)} = f^*_{m+1}$. The value of K is chosen so that

$$||u_{m+1}^{(k)} - u_{m+1}^{(k-1)}|| < \epsilon$$

where ϵ is some preassigned small quantity.

The above iteration process may be shown to be convergent by the following analysis. We shall for ease only consider the case of a single scalar equation. The sequence of iterates $u_{m+1}^{(k)}; k=1,\ldots K$ is required to tend to the solution u_{m+1} of the equation

$$u_{m+1} = u_m - {}^p/_4(\Delta_x + \nabla_x)(f_{m+1} + f_m).$$

If we subtract the above equation from the corrector in (5.5.2) we have

$$(u_{m+1}^{(k+1)} - u_{m+1}) = - {}^p/_4(\Delta_x + \nabla_x)(f_{m+1}^{(k)} - f_{m+1})$$

If f is a slowly varying function of u, approximation to the above equation is given by

$$\left(u_{m+1}^{(k+1)} - u_{m+1}\right) = - \frac{p}{4}\tilde{A}(\Delta_x + \nabla_x)(u_{m+1}^{(k)} - u_{m+1})$$

where $\tilde{A}$ is regarded as a constant at the $(m+1)^{st}$ time level. Let

$$e_{k+1} = u_{m+1}^{(k+1)} - u_{m+1}$$

and therefore

$$e_{k+1} = -p/4\tilde{A}(\Delta_x + \nabla_x)e_k.$$

If the vector of errors e_k at points on the k^{th} level (assuming the error to be zero at the boundaries and the boundary conditions to be periodic) is denoted by E_k, the following matrix system is obtained.

$$E_{k+1} = +p/4\tilde{A} \begin{bmatrix} \cdot & +1 & & & \\ -1 & \cdot & +1 & & \\ & -1 & \cdot & & \\ & & & +1 & \\ & & & -1 & \cdot \end{bmatrix} E_k$$

$$= A_1 E_k.$$

It follows that the iterative process will be convergent if the eigenvalues of A_1, given by

$$\lambda_s = \pm i\frac{p}{2}\tilde{A}\cos\frac{s\pi}{N-1} \qquad s = 1,\ldots N-2$$

are all less than one in modulus. This requires

$$p|\tilde{A}| \leq 2.$$

This condition is satisfied as $p|A| \leq 1$ for stability, and therefore the process is convergent.

An alternative procedure would be to calculate u_N^{*n+1} to the required accuracy, by a direct process. To ensure compatability with the predictor-corrector scheme, this value should be obtained with the same principal part of truncation error as the predictor. Writing (5.5.1) in terms of difference operators for u_N^{*n+1}, we obtain

(5.5.3) $$u_N^{*n+1} = \tfrac{1}{2}(\Delta_x - \nabla_x + 2)u_N^n - \frac{p}{2}(\Delta_x + \nabla_x)f_N^n.$$

Using the relation

$$(1+\Delta_x) = (1-\nabla_x)^{-1}$$

or

$$\Delta_x = \nabla_x(1-\nabla_x)^{-1}$$
$$= \nabla_x + \nabla_x^2 + \ldots$$

and substituting in (5.5.3) up to second order terms, we obtain the formula

$$u_N^{*m+1} = \tfrac{1}{2}(2 + \nabla_x^2)u_N^m - {}^p/_2(2\nabla_x + \nabla_x^2)f_N^m$$

This has the same principal part of truncation error (of order h^2) as equation (5.5.3).

Therefore we may use the formula

$$(5.5.4)\quad u_N^{*m+1} = u_N^m + \tfrac{1}{2}(u_N^m - 2u_{N-1}^m + u_{N-2}^m) - p[f_N^m - f_{N-1}^m + \tfrac{1}{2}(f_N^m - 2f_{N-1}^m + f_{N-2}^m)]$$

to supply the missing predicted value. Revised stability requirements are not necessary as the formula (5.5.4) is not used more than once at any time level. Similar boundary predictors can be derived for the remaining schemes in sections 5.3 and 5.4.

A further method, perhaps the simplest of all would be to obtain the value of u_N^{*m+1} by an extrapolation of the values of u_i^{*m+1} $(i=1,\ldots N-1)$.

5.6 Explicit Two-Dimensional Case.

The two-step Lax Wendroff method for the first order system of conservation laws in two space variables

$$\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = 0$$

where $f = f(u)$, $g = g(u)$, is given in difference operator notation by

$$u^{*}_{m+\frac{1}{2}} = \tfrac{1}{4}[(\Delta_x - \nabla_x + 2)(\Delta_y - \nabla_y + 2)]u_m - {}^{k}/_{4}[(\Delta_x+\nabla_x)f_m+(\Delta_y+\nabla_y)g_m]$$

(5.6.1)

$$u_{m+1} = -{}^{k}/_{2}[(\Delta_x + \nabla_x)f^{*}_{m+\frac{1}{2}} + (\Delta_y + \nabla_y)g^{*}_{m+\frac{1}{2}}] + u_m$$

where $u_m = u^{n}_{ij}$ etc. (RICHTMYER [29]).

Similarly, the analogue in two space dimensions of the new predictor-corrector scheme introduced in section 5.3 is given by

$$u^{*}_{m+\frac{1}{2}} = \tfrac{1}{4}[(\Delta_x - \nabla_x + 2)(\Delta_y - \nabla_y + 2)]u_m - {}^{k}/_{2}[(\Delta_x+\nabla_x)f_m + (\Delta_y+\nabla_y)g_m]$$

(5.6.2)

$$u_{m+1} = u_m - {}^{k}/_{4}[(\Delta_x+\nabla_x)(f^{*}_{m+\frac{1}{2}}+f_m)+(\Delta_y+\nabla_y)(g^{*}_{m+\frac{1}{2}} + g_m)]$$

Both the above methods are again accurate to second order in h.

Following Richtmyer we will consider the stability of the above schemes for the system of fluid dynamics equations in the non-conservative Eulerian formulation

$$\frac{\partial w}{\partial t} + A(w)\frac{\partial w}{\partial x} + B(w)\frac{\partial w}{\partial y} = 0 \qquad (5.6.3)$$

where

$$w = \begin{bmatrix} \rho \\ u \\ v \\ p \end{bmatrix},$$

$$A(w) = \begin{bmatrix} u & \rho & \cdot & \cdot \\ \cdot & u & \cdot & 1/\rho \\ \cdot & \cdot & u & \cdot \\ \cdot & \rho c^2 & \cdot & u \end{bmatrix},$$

$$B(w) = \begin{bmatrix} v & \cdot & \rho & \cdot \\ \cdot & v & \cdot & \cdot \\ \cdot & \cdot & v & 1/\rho \\ \cdot & \cdot & \rho c^2 & v \end{bmatrix}$$

We make the usual Fourier transform of the space variables and introduce the notation

$$C = A\sin\alpha + B\sin\beta$$

It is easy to show that the eigenvalues of C are

$$\lambda = (\sin^2\alpha + \sin^2\beta)^{\frac{1}{2}} \begin{Bmatrix} u' \\ u' \\ u'+c \\ u'-c \end{Bmatrix}$$

where $u' = u\cos\theta + v\sin\beta$ with $\cos\theta = \dfrac{\sin\alpha}{\sqrt{\sin^2\alpha + \sin^2\beta}}$

Let $\mu = p\lambda$, and the eigenvalues of the amplification matrix G of the Lax Wendroff scheme are then given by

$$g = 1 - 2\mu^2 - i\mu(\cos\alpha + \cos\beta).$$

For the eigenvalues to have modulus less than or equal to 1, it follows after some simple manipulation, that the condition

(5.6.4) $p(|q|+c) \leq \sqrt{2}$

where $|q| = (u^2 + v^2)^{\frac{1}{2}}$ must be satisfied. (RICHTMYER [29]).

A similar analysis for the new predictor corrector scheme results in the eigenvalues of the amplification matrix being given by

$$g = 1 - \frac{\mu^2}{2} + \frac{i\mu}{2}(1 + \tfrac{1}{2}(\cos\alpha + \cos\beta).$$

After some manipulation we obtain

$$|g|^2 = 1 - \frac{\mu^2}{4}[4 - (c_1^2 s + c_2^2 s)^2 - \mu^2]$$

which leads to the stability condition

$$p^2(|q| + c)^2 \leq \frac{4 - (c_\alpha^2 s + c_\beta^2 s)^2}{s_\alpha^2 + s_\beta^2}$$

for all α, β.

It may be shown that the minimum of the quantity on the right is one, and therefore for stability we require

(5.6.5) $p(|q| + c) \leq 1$

A comparison of (5.6.4) and (5.6.5) demonstrates that the new predictor-corrector method has slightly poorer stability characteristics. However this scheme does have an interesting property which will be shown in a numerical example in section 5.8.

In initial boundary value probelms, similar procedure to those developed in section 5.5 for the one dimensional case can be applied to smooth the boundary data on the planes $x = a$; $y = b$. The procedure involving the iteration of the corrector is again convergent as a similar analysis shows.

5.7 Implicit two-dimensional case.

It is shown in section 5.4 that an implicit predictor corrector procedure can be developed for the solution of the one dimensional system of conservation laws. In this section an attempt is made to extend this scheme to two space variable problems.

The two-dimensional analogue of the scheme (5.4.2) is in fact the alternating direction procedure introduced in Chapter III for the linear case. The direct extension to the nonlinear case no longer holds as the boundary incorporation procedure given in section 3.5 is no longer valid. However the overall replacement

$$
\begin{aligned}
&[u_{m+1}+{}^{p}/_{4}[(\Delta_x+\nabla_x)f_{m+1}+(\Delta_y+\nabla_y)g_{m+1}]+{}^{p^2}/_{16}(\Delta_y+\nabla_y)B_{m+1}(\Delta_x+\nabla_x)f_{m+1}]\\
&=[u_m-{}^{p}/_{4}[(\Delta_x+\nabla_x)f_m+(\Delta_y+\nabla_y)g_m]+{}^{p^2}/_{16}(\Delta_y+\nabla_y)B_m(\Delta_x+\nabla_x)f_m]
\end{aligned}
\tag{5.7.1}
$$

still has second order accuracy, and unconditional stability. If we introduce a predicted value u^{**}_{m+1} (obtained by the predictor in (5.6.2)) and write the scheme in the split form (D'JAKONOV [4]):

$$
\begin{aligned}
&[I+{}^{p}/_{4}(\Delta_y+\nabla_y)B_1(u^{**}_{m+1})]u^{*}_{m+1}=H\\
&[I+{}^{p}/_{4}(\Delta_x+\nabla_x)A_1(u^{**}_{m+1})]u_{m+1}=u^{*}_{m+1}
\end{aligned}
\tag{5.7.2}
$$

where

$$f(u)=A_1(u).u$$

$$g(u)=B_1(u).u$$

and H is the expression on the right hand side of equation (5.7.1), then the accuracy is maintained. The introduction of u^{**}_{m+1} only introduces an error of third order. The stability analysis of the linearized form of (5.7.2) coincides with that given for the scheme introduced in chapter III, and therefore scheme (5.7.2) is

unconditionally stable.

An analogucous technique may be used to extend the schemes introduced in Chapter II to the nonlinear case. However this time predicted values may be required for the points $u^{n+1}_{i+\frac{1}{2},j+\frac{1}{2}}$ which lie on a grid which interlaces the main grid. The "diagonal" schemes, however, do not have this drawback; for example, one of the eight point diagonal schemes for the nonlinear case can be written in the form.

$$u^{**}_{m+1} = \tfrac{1}{2}(\Delta_x - \nabla_x + 2)u_m - p/2[(\Delta_x + \nabla_x)f_m + (\Delta_y + \nabla_y)g_m]$$

$$[I + \tfrac{1}{2}p\nabla_y B_1(u^{**}_{m+1})]u^{*}_{m+1} = [u_m - \tfrac{1}{2}p[\Delta_x f_m + \Delta_y g_m) + \tfrac{1}{4}p^2 \Delta_y B_m \Delta_x f_m]$$

$$[I + \tfrac{1}{2}p\nabla_x A_1(u^{**}_{m+1})]u_{m+1} = u^{*}_{m+1}$$

This scheme will only be useful if the matrices are positive or negative definite.

5.8 Numerical Results.

In order to compare and contrast the explicit methods (5.6.1), (5.6.2) with the alternating direction method (5.7.2), test runs were carried out on the problem

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{1}{4}u^2\right) + \frac{\partial}{\partial y}\left(\frac{1}{4}u^2\right) = 0$$

with initial data given by

$$u(x,y,0) = \frac{1}{4}(x + y)^2$$

The theoretical solution to this problem is given by

$$u(x,y,t) = \left\{\frac{1 - \sqrt{1 + (x + y)t}}{t}\right\}^2$$

The region under consideration $0 \leq x, y \leq 1$ was covered by a grid with mesh spacing $h = 0{\cdot}1$. The three methods were run for various values of the mesh ratio p and the results for the errors at the mid-point of the region are shown in Table 5.8.1.

The results tend to show that the three methods have comparable accuracy, the alternating direction method in general being slightly more accurate.

The effect of iterating the corrector in the predictor-corrector scheme seems to help to damp out the error, but only for small p. This process may be of some value in problems where the solution of the differential equation is not well-behaved.

Table 5.8.1.

p	No. of steps	Lax-Wendroff	Predictor-Corrector		A.D.I.
	No. of Iterations	-	0	2	-
•3	50	$•78 \times 10^{-3}$	$•60 \times 10^{-3}$	$•42 \times 10^{-3}$	$•49 \times 10^{-3}$
	100	$•80 \times 10^{-3}$	$•59 \times 10^{-3}$	$•44 \times 10^{-3}$	$•49 \times 10^{-3}$
•6	50	$•76 \times 10^{-3}$	$•59 \times 10^{-3}$	$•44 \times 10^{-3}$	$•47 \times 10^{-3}$
	100	$•72 \times 10^{-3}$	$•52 \times 10^{-3}$	$-•22 \times 10^{-3}$	$•44 \times 10^{-3}$
1•0	50	$•69 \times 10^{-3}$	$•53 \times 10^{-3}$	$•11 \times 10^{-2}$	$•44 \times 10^{-3}$
	100	$•68 \times 10^{-3}$	$•45 \times 10^{-3}$	$•18 \times 10^{-2}$	$•48 \times 10^{-3}$

CHAPTER VI

EXTENSIONS AND CONCLUDING REMARKS

Introduction.

Extensions of the schemes introduced in the preceeding chapters are now outlined to cover the case of more than two space dimensions.

6.1 Extensions to more than two space dimensions.

Let us now consider the extension of the finite difference schemes introduced in the preceeding chapters, to linear differential systems of the form

$$(6.1.1)\quad \frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y} + C\frac{\partial u}{\partial z}$$

where A,B and C are nxn matrix functions of x,y,z and t, and u is the required solution vector. To avoid tedious repetition, we shall restrict ourselves to a consideration of scheme (3.5.2)

By straightforward application of Taylor's theorem it can be verified that the formula

$$(6.1.2)\quad [I_n - \tfrac{1}{4}pC_{m+1}(\Delta_z + \nabla_z)][I_n - \tfrac{1}{4}pB_{m+1}(\Delta_y + \nabla_y)][I_n - \tfrac{1}{4}pA_{m+1}(\Delta_x + \nabla_x)]u_{m+1}$$
$$= [I_n + \tfrac{1}{4}pC_m(\Delta_z + \nabla_z)][I_n + \tfrac{1}{4}pB_m(\Delta_y + \nabla_y)][I_n + \tfrac{1}{4}pA_m(\Delta_x + \nabla_x)]u_m$$

is a finite difference replacement of (6.1.1) with truncation error of third order. An analysis parallel to that given in section 3.6 will verify the unconditional stability of (6.1.2). In order that equation (6.1.2) admit of a P.R. factorization of the form

$$[I_n - \tfrac{1}{4}pB_{m+1}(\Delta_y + \nabla_y)]u^{*}_{m+1} = [I_n + \tfrac{1}{4}pA_m(\Delta_x + \nabla_x)]u_m$$

$$(6.1.3)\quad [I_n - \tfrac{1}{4}pC_{m+1}(\Delta_z + \Delta_z)]u^{**}_{m+1} = [I_n + \tfrac{1}{4}pB_m(\Delta_y + \nabla_y)]u^{*}_{m+1}$$

$$[I - \tfrac{1}{4}pA_{m+1}(\Delta_x + \nabla_x)]u_{m+1} = [I_n + \tfrac{1}{4}pC_m(\Delta_z + \nabla_z)]u^{**}_{m+1}$$

where u^{*}_{m+1} and u^{**}_{m+1} are intermediate values, the commutation condition

$$BC = CB$$

must be satisfied. This may be verified by elimination of u^{*}_{m+1}, u^{**}_{m+1} from (6.1.3). Further, an attempt to write the totality of equations (6.1.3) for the region (c.f(3.5.4)) will show that there is no means of

incorporating the boundary data in a consistent manner. Since the D.R. factorization also suffers this inconsistency, we must turn to the third type of factorisations, the D'JAKONOV form. Formula (6.1.2) then takes the form.

$$[I_n - \tfrac{1}{4}pC_{m+1}(\Delta_z + \nabla_z)]u^{*}_{m+1} = Qu_m$$

$$(6.1.4)\quad [I_n - \tfrac{1}{4}pB_{m+1}(\Delta_y + \nabla_y)]u^{**}_{m+1} = u^{*}_{m+1}$$

$$[I_n - \tfrac{1}{4}pA_{m+1}(\Delta_x + \nabla_x)]u_{m+1} = u^{**}_{m+1}$$

where Q is the operator on the right hand side of equation (6.1.2). The boundary data may then be incorporated smoothly in a manner analoguous to that given in section 3.4. Thus the extensions of scheme (3.5.2) to more than two space dimensions appears to be achieved most easily when the D-factorisation is employed. Similar conclusions may be drawn for the extensions of the other schemes given in chapters II and III, and even for the nonlinear schemes of chapter V.

6.2 Concluding Remarks.

Since the methods developed in this thesis were specifically designed to solve the system

$$\frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y}$$

(or

$$\frac{\partial u}{\partial t} + \frac{\partial f}{\partial x} + \frac{\partial g}{\partial y} = 0$$

in the non-linear case).

they will require considerable modification to deal with the more complicated equation

$$\frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y} = S(x,y,t,u)$$

However, in passing, we may note that for the equation

$$\frac{\partial u}{\partial t} = A\frac{\partial u}{\partial x} + B\frac{\partial u}{\partial y} + S(x,y)$$

the changes are trivial. For if we employ the D'Jakonov factorization, the eighteen point scheme takes the form:

$$[I_n - \tfrac{1}{4}B_{m+1}(\Delta_y + \nabla_y)]u^*_{m+1} = [I_n + \tfrac{1}{4}B_m(\Delta_y + \nabla_y)][I_n + \tfrac{1}{4}A_m(\Delta_x + \nabla_x)]u_m + S_m$$

$$[I_n - \tfrac{1}{4}A_{m+1}(\Delta_x + \nabla_x)]u_{m+1} = u^*_{m+1}$$

which is again correct to second order in the space increment. In fact if $S = S(x,y,t)$ the extension is only slightly more complicated.

It would appear from experience in using the alternating direction methods developed in this thesis, that their accuracy compares favourably with the commonly-used Lax Wendroff method. The main advantage of the implicit methods is undoubtedly their unrestricted stability.

It is hoped in the future to test the alternating direction

method, and the two explicit two-step methods, on physical problems in the fields of hydrodynamics and elasticity, especially with regard to their behaviour in the neighbourhood of shocks and discontinuities.

The method of attack in this thesis for the numerical integration of a partial differential equation has been to reduce the original partial differential equation to a system of first order. In certain problems it may be more beneficial to reduce the problem to a second order system. For example the equation governing the vibrations of a flat plate

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial^4 u}{\partial x^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = 0$$

may be reduced to the second-order system

$$\frac{\partial \underline{w}}{\partial t} = A \nabla^2 \underline{w} \qquad (\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$$

where $\underline{w} = \begin{bmatrix} u \\ \frac{\partial u}{\partial t} \\ \nabla^2 u \end{bmatrix}$ and $A = \begin{bmatrix} \cdot & -1 \\ 1 & \cdot \end{bmatrix}$. The numerical solution of this system has been considered in FAIRWEATHER and GOURLAY [41].

REFERENCES

[1] ANUCINA, N.N. Difference schemes for solving the Cauchy problem for symmetric hyperbolic systems.
Soviet Mathematics 5, 60-64 (1964).

[2] BURSTEIN, S.Z. Numerical calculations of multidimensional shocked flows.
New York Univ., Courant Inst.Math.Sci.Res Rep NYU-10,433 (1964).

[3] COURANT, R.C., K.O. FRIEDRICHS and LEWY, H. (trs.Phyllis Fox) On the partial difference equations of mathematical physics.
New York Univ., Courant Inst.Math.Sci.Res.Rep. NYO-7689 (1956).

[4] D'JAKONOV, G.Ye, Difference schemes with a disintegrating operator for multidimensional problems.
U.S.S.R. Comput.Math. and Math.Phys. 4,581-607 (1963).

[5] DOUGLAS, Jim Jr, and GUNN, James E., A general formulation of alternating direction methods I. Parabolic and Hyperbolic Problems.
Num.Math. 6, 428-453 (1964).

[6] FRIEDRICHS, K.O., Symmetric hyperbolic partial differential equations.
Comm.Pure Appl.Math. 7,345-392 (1954).

[7] GARABEDIAN, P.R., Partial Differential Equations.
John Wiley and Sons, New York (1964).

[8] GARY, J., On certain finite difference schemes for hyperbolic systems.
Maths. of Comp. 18, 1-18 (1964).

[9] GOURLAY, A.R. and MITCHELL, A.R., Alternating Direction methods for hyperbolic systems.
Num.Math. 8 137-149 (1966).

[10] GOURLAY, A.R. and MITCHELL, A.R., Two level difference schemes for hyperbolic systems.
SIAM Journal on Num.Analy. (to appear).

[11] GOURLAY, A.R. and MITCHELL, A.R., A stable implicit method for hyperbolic systems in two space variables.
Num.Math. 8 367-375 (1966).

[12] GRAGG, W.B. and STETTER, H.J., Generalized multistep predictor-corrector methods.
Journ.Assoc.Comput.Mech. 11 188-209 (1964).

[13] JOHANNSON, O. and KREISS, H.-O., Uber das verfahren der zentralen Differenzen zur losung des Cauchyproblems fur Partielle Differentialgleichungen.
BIT. 3 97-107 (1963).

[14] KREISS, H.O., Uber die Stabilitats definition fur differenzgleichungen die partielle differenzgleichunge approximieren.
BIT. 2 153-181 (1962).

[15] KREISS, H.O., Uber implizite Differenzmethoden fur partielle Differentailgleichungen.
Num.Math. 5 24-47 (1963).

[16] KREISS, H.O., On difference approximations of dissipative type for hyperbolic differential equations.
Comm.Pure Appl.Math. 17 335-353 (1964).

[17] LAX, P.D. and RICHTMYER, R.D., Survey of the stability of linear finite difference equations.
Comm.Pure Appl.Math. 9 267-293 (1956).

[18] LAX, P.D., Hyperbolic systems of conservation Laws II.
Comm.Pure Appl.Math. 10 567-566 (1957).

[19] LAX, P.D. and WENDROFF, B., Systems of Conservation Laws.
Comm.Pure Appl.Math. 13 217-237 (1960).

[20] LAX, P.D., On the stability of difference approximations to solutions of hyperbolic equations with variable coefficients.
Comm.Pure Appl.Math. 14 497-520 (1961).

[21] LAX, P.D., Nonlinear hyperbolic systems of conservation laws.
Nonlinear Problems, The University of Wisconsin Press, Madison 3-12 (1963).

[22] LAX, P.D. and WENDROFF, B., Difference schemes with high order of accuracy for solving hyperbolic equations.
New York Univ., Courant Inst.Math.Sci.Res.Rep. NYO 9759 (1964).

[23] LAX, P.D. and WENDROFF, B., Difference schemes for hyperbolic equations with high order of accuracy.
Comm.Pure Appl.Math. 17 381-398 (1964).

[24] NOH, W.F. and PROTTER, M.H., Difference methods and the equations of hydrodynamics.
Journal of Mathematics and Mechanics 12 149-192 (1963).

[25] O'BRIEN, G.G., HYMAN, M.A. and KAPLAN, S., A study of the numerical solution of partial differential equations. J.Math. and Physics 29 223-251 (1951).

[26] PARTER, S.V., Stability, Convergence and Pseudo-Stability of finite difference equations for an over-determined problem. Num.Math. 4 277-292 (1962).

[27] PEACEMAN, D.W. and RACHFORD, H.H.Jr., The numerical solution of parabolic and elliptic differential equations. J.Soc.Ind.Appl.Math. 3 28-41 (1955).

[28] PHILLIPS, N.A., An example of non-linear computational instability. The Atmosphere and the Sea in Motion. Rockefeller Institute Press, New York 501-504 (1959).

[29] RICHTMYER, R.D., A survey of difference methods for non-steady fluid dynamics. N.C.A.R. Tech-Note 63-2 (1963).

[30] RICHTMYER, R.D. and MORTON, K.W., Stability studies for difference equations. New York University, Courant Inst.Math.Sci.Res.Rep. NYO-1480-5 (1964).

[31] SCHECHTER, S., Quasi-tridiagonal matrices and type-insensitive difference equations. Quart.Appl.Math. 18 258-295 (1960).

[32] STRANG, W.G., Polynomial approximation of Bernstein type. Trans.Amer.Math.Soc. 105 525-535 (1962).

[33] STRANG, W.G., Trigonometric polynomials and difference methods of maximum accuracy.
Jour.of Maths. and Physics 41 147-154 (1962).

[34] STRANG, W.G., Accurate partial difference methods I: Linear Cauchy Problems.
Arch.Rat.Mech.Anal. 12 392-402 (1963).

[35] STRANG, W.G., Accurate partial difference methods II: Non-linear problems.
Num.Math. 6 37-46 (1964).

[36] STRANG, W.G., Wiener-Hopf difference equations.
Jour.Math and Mech. 13 85-96 (1964).

[37] STRANG, W.G., Unbalanced Polynomials and difference methods for mixed problems.
SIAM Num.Anal.SerB 2 46-51 (1965).

[38] VARGA, R.S., Matrix Iterative Analysis.
Prentice-Hall, New York 1962.

[39] WALL, H.S., Analytic Theory of Continued Fractions.
Van Nostrand Co., New York, 1948.

[40] WENDROFF, B., On centred difference equations for hyperbolic systems.
J.Soc.Ind.Appl.Math. 8 549-555 (1960).

[41] FAIRWEATHER, G. and GOURLAY, A.R., Some stable difference approximations to a fourth-order parabolic partial differential equation.
Mathematics of Computation (to appear).