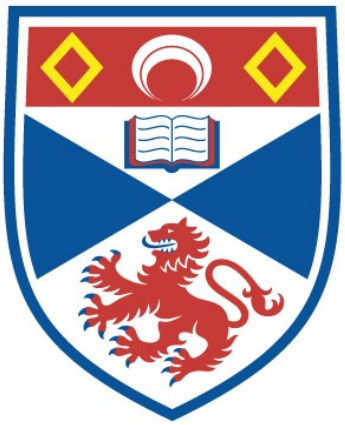
ALTERNATING DIRECTION IMPLICIT METHODS FOR PARTIAL DIFFERENTIAL EQUATIONS

Graeme Fairweather

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



1966

ALTERNATING DIRECTION IMPLICIT METHODS

FOR

PARTIAL DIFFERENTIAL EQUATIONS.

A thesis presented by

Graeme Fairweather, 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 1963, I graduated with First Class Honours in Applied Mathematics. In July 1963, 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.

ACKNOWLEDGEMENT.

The Author is indebted to the Carnegie Trust for a Research Scholarship held during the period of research.

CERTIFICATE.

I certify that Graeme Fairweather 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.

continued overleaf.

continued overleaf.

CHAPTER V. A HIGH ACCURACY ADI METHOD FOR THE WAVE EQUATION.

CHAPTER I.

INTRODUCTION.

I.1. Introductory Remarks.

Partial differential equations occur in most branches of applied mathematics, physics and engineering. Many of the linear problems which arise involve the solution of an equation obtained by suitably specialising the form

$$\nabla^2 u = \lambda \frac{\partial^2 u}{\partial t^2} + \mu \frac{\partial u}{\partial t}$$

where λ and μ are certain physical constants. Here the operator ∇^2 is the Laplacian operator in the space of one, two or three dimensions. For example,

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

in three dimensional rectangular coordinates, where the unknown function u is a function of the space coordinates x, y, z and the time coordinate t.

In particular, Laplace's equation,

(1.1) $$\nabla^2 u = 0,$$

is satisfied by such physical functions as the velocity potential of an ideal incompressible fluid, the gravitational potential in free space, the electrostatic potential in the steady flow of electric currents in solid conductors and the steady-state temperature distribution in solids.

The wave equation,

(1.2) $$\nabla^2 u = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2}$$

arises in the study of propagation of waves with velocity c, independent of the wave length. In particular, it is satisfied by the components of the electric or magnetic vector in electromagnetic theory, by suitably chosen components of displacements in the theory of elastic vibrations

and by the velocity potential in the theory of sound (acoustics) for a perfect gas.

The diffusion equation,

$$\nabla^2 u = \frac{1}{\beta^2}\frac{\partial u}{\partial t} \tag{1.3}$$

is satisfied, for example, by the temperature at a point of a homogeneous body and by the concentration of a substance in the theory of diffusion, where β is a suitably prescribed constant.

Differential equations of higher order involving the operator ∇^2 are also frequently encountered. In particular, the biharmonic equation in two dimensions,

$$\nabla^4 u \equiv \nabla^2\nabla^2 u \equiv \frac{\partial^4 u}{\partial x^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = 0, \tag{1.4}$$

is involved in many problems in the theory of elasticity. The same equation also arises in the discussion of the slow motion of a viscous fluid.

Solutions of these equations for regions of arbitrary shape are, of course, not known, but even for those problems for which analytic solutions in series form are available, the series often do not lend themselves readily to numerical calculations. Direct numerical solutions of these equations are therefore of considerable importance.

There are many numerical methods for solving partial differential equations. Of these, only one stands out as being universally applicable to both linear and non-linear problems— the method of finite differences. It is the major purpose of this thesis to attempt to improve a class of finite difference methods, the class of alternating

direction implicit (ADI) methods, for solving several types of linear partial differential equations. In Chapters II and III, high accuracy ADI methods for equation (1.3) in two and three space variables are derived and, as a result, families of iterative techniques for solving (1.1) are obtained. ADI methods for the numerical solution of (1.4), and (1.2) in two and three space variables, are discussed in Chapters IV and V respectively. All numerical calculations were carried out on the IBM 1620 computer of the University of St. Andrews.

In the remainder of this Chapter, several fundamental concepts of the solution of partial differential equations by finite difference methods are explained.

I.2. Finite Difference Methods for the Solution of Partial Differential Equations.

Every partial differential equation is either of elliptic, parabolic or hyperbolic type, for example, (1.1) and (1.4) are elliptic partial differential equations while (1.2) is hyperbolic and (1.3) is parabolic, (see Sneddon [39]). Problems of parabolic or hyperbolic type lead to initial value problems since the conditions at time $t = 0$ are specified as well as possibly conditions along the boundaries of the region, and the dependent variable is then calculated for $t > 0$. On the other hand, elliptic equations lead to boundary value problems. In this case, we must find a function which satisfies the differential equation and also conditions on the boundary of a closed region.

The method of solution by finite differences depends on the type of the partial differential equation but the main approach is to cover the domain of the independent variables by a rectangular network of planes, not necessarily equally spaced, but parallel to the principal planes. The finite set of points of intersection of the planes constitute the mesh points, or nodes, and we seek to determine approximate values of the desired solution at these points. The values at the mesh points are required to satisfy difference equations obtained either by replacing partial derivatives by partial difference operators or by certain other more sophisticated techniques.

The method is best described by means of an elementary example. Consider the parabolic equation

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

subject to appropriate initial and boundary conditions. This equation may be written in operator form

$$(1.5) \qquad (D_x^2 + D_y^2 - D_t)u = 0.$$

In this example and throughout the remainder of the thesis, the domain of the variables x, y will be square or rectangular and will be covered by a square mesh where $\Delta x = \Delta y = h$. The natural and simplest finite difference representations of D_x and D_y are

$$D_x^2 = h^{-2}\delta_x^2, \quad D_y^2 = h^{-2}\delta_y^2,$$

where δ_x and δ_y, the central difference operators in the x and y directions respectively, are such that

$$\delta_x u(x,y,t) = u(x+\tfrac{1}{2}h,y,t) - u(x-\tfrac{1}{2}h,y,t)$$

and

$$\delta_y u(x,y,t) = u(x,y+\tfrac{1}{2}h,t) - u(x,y-\tfrac{1}{2}h,t).$$

There are two equally natural representations of D_t:

$$D_t = k^{-1}\Delta_t, \quad D_t = k^{-1}\nabla_t,$$

where k is the mesh size in the t-direction and Δ_t and ∇_t are the forward and backward difference time operators respectively where

$$\Delta_t u(x,y,t) = u(x,y,t+k) - u(x,y,t)$$

and

$$\nabla_t u(x,y,t) = u(x,y,t) - u(x,y,t-k).$$

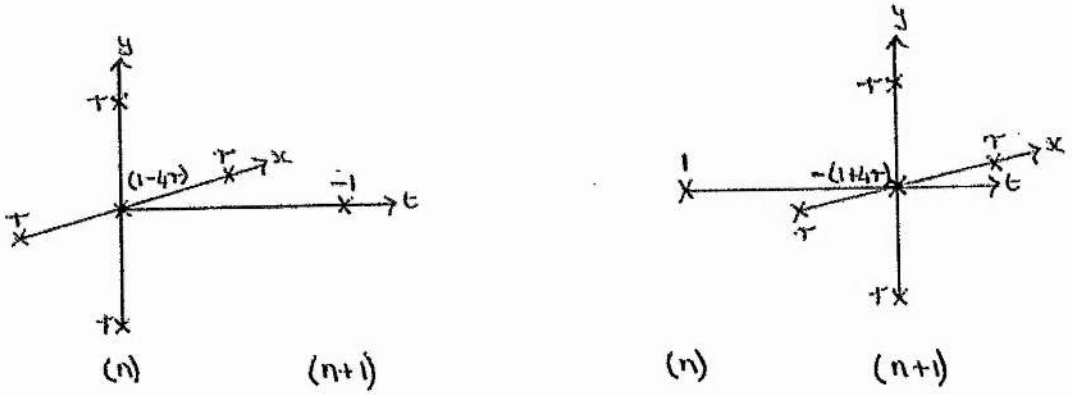
If applied to (1.5) with $u_{i,j,n} = u(x_i,y_j,t_n)$ where $x_i = ih$, $y_j = jh$ and $t_n = nk$, these forward and backward time operators lead to somewhat different difference equations which can be written as

$$(1.6) \quad r(u_{i,j+1,n} + u_{i+1,j,n} + u_{i,j-1,n} + u_{i-1,j,n}) + (1-4r)u_{i,j,n} - u_{i,j,n+1} = 0,$$

and

(1.7) $$r(u_{i,j+1,n} + u_{i+1,j,n} + u_{i-1,j,n} + u_{i,j-1,n}) - (1+4r)u_{i,j,n} + u_{i,j,n+1} = 0$$

respectively, where $r = k/h^2$ is the mesh ratio. The coefficients in the two approximations can be represented schematically in the following way:



These schemes show the relative locations of the points to which the coefficients are applied. The first method, (1.6), the forward difference method, is an example of an explicit method while the second method, (1.7), the backward difference method, is an example of an implicit method, a difference method being termed explicit or implicit according to whether each of the difference equations at $t = nk$ contains one or several of the unknown values $u(ih, jh, (n+1)k)$. It is obvious that every step of the calculation is much easier for an explicit method than for an implicit one - an explicit method requires at each step the solution of a number of equations with only a single unknown in each of them, while an implicit method demands the solution of a system of simultaneous linear equations. Unfortunately, most of the explicit methods are only conditionally stable, and this imposes an upper limit on the allowable mesh ratios. The concept of stability is discussed in the next section.

We see that, in problems involving parabolic - and also hyperbolic - partial differential equations, we can construct numerical solutions

step-by-step in time using a so-called "marching" process. For elliptic problems, on the other hand, the solution cannot be found at any specific point without being found at all points. This fact is borne out by the following example.

Consider Laplace's equation in two dimensions

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

with appropriate boundary conditions. In operator notation this equation may be written as

$$(D_x^2 + D_y^2)u = 0.$$

The most common finite difference replacement is obtained by putting $D_x^2 = h^{-2}\delta_x^2$ and $D_y^2 = h^{-2}\delta_y^2$ yielding the formula

$$u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{i,j} = 0 \tag{1.8}$$

at the node $x = ih$, $y = jh$. The application of this formula to each node in turn together with the boundary conditions yields a set of linear equations for the unknown nodal values $u_{i,j}$. Thus, the solution of an elliptic equation by finite difference methods reduces to the solution of a system of simultaneous linear equations. However, since there is one equation for each mesh point and since there may be several hundred mesh points, great care must be taken in the choice of method for solving these equations, in case the computing time, even for a very fast digital computer, becomes excessive. Iterative methods are indicated because of the large number of zero elements in the matrix of the coefficients, and are nearly always used. With such methods, we take an initial approximation to the solution and successively modify it according to a given rule, until convergence has been achieved,

that is, until the difference between successive estimates of the exact solution is less than a prescribed amount. The problem of stability does not arise in the solution of a boundary value problem involving a linear elliptic partial differential equation. However, the convergence factor of the iterative method is of critical importance, and is discussed more fully in I.4.

One iterative method for the solution of the equations arising from the application of (1.8) to each node is the method of Altman or Von Mises (Martin and Tee [28]) which may be written in the form

$$(1.9)\quad \alpha[u^{(n)}_{i,j+1} + u^{(n)}_{i-1,j} + u^{(n)}_{i+1,j} + u^{(n)}_{i,j-1}] + (1-4\alpha)u^{(n)}_{i,j} - u^{(n+1)}_{i,j} = 0,$$

where $u^{(n)}_{i,j}$ denotes the n th estimate of $u_{i,j}$ - the exact solution of the difference equation at the node (ih, jh) - and α is an iteration parameter which is chosen in order to accelerate the convergence of the method. It is easily seen that there is a great similarity between (1.6) and (1.9). In fact, if, in (1.6), r is replaced by α and each time step is regarded as an iteration, we obtain (1.9). This connection between finite difference methods for the solution of parabolic equations and iterative methods for the solution of systems of equations arising from elliptic difference equations is employed in chapters II and III.

Just as (1.6) had its limitations for solving the parabolic equation so (1.9) is not one of the better iterative methods for the solution of linear equations.

I.3. Stability and Convergence of Initial Value Problems.

When discussing the numerical solution of partial differential equations of parabolic and hyperbolic types by means of difference methods, we must consider three distinct solutions, namely 1) the theoretical solution of the differential equation, 2) the theoretical solution of the difference equation, and 3) the numerical solution of the difference equation. The difference between 1) and 2) is due to truncation error, the error introduced by discretizing the problem. This error may be defined as follows: Let the differential equation be rewritten as the differential operator equation

$$Du = 0$$

and the difference method be written as

$$\Delta u = 0$$

then, for any sufficiently smooth function v, we define the truncation error as the difference

$$Dv - \Delta v.$$

As an example, consider the difference equation (1.6) as an approximation to the differential equation (1.3) with $\beta = 1$, and two space variables, and let u(x,y,t) be any function with continuous partial derivatives of sufficiently high order. By expanding $u_{i,j+1,n}$ $u_{i,j-1,n}$, etc., in terms of $u_{i,j,n}$ and its derivatives, where $u_{i,j,n} = u(ih, jh, nk)$, we find that the difference between the finite difference expression and the corresponding differential expression is

$$(1.10a)\quad \left[r(u_{i,j+1,n} + u_{i,j-1,n} + u_{i+1,j,n} + u_{i-1,j,n}) + (1-4r)u_{i,j,n} - u_{i,j,n+1}\right] + \Delta t\left\{\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2}\right\}_{i,j,n} = -\tfrac{1}{2}(\Delta t)^2\left(\frac{\partial^2 u}{\partial t^2}\right)_{i,j,n} + \frac{1}{12}\Delta t.h^2\left\{\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}\right\}_{i,j,n}$$

The right hand side of this equation is the principal part of the truncation error of the difference scheme (1.6) and is a measure of the accuracy of the scheme. The fact that the coefficients of Δt and of h^2 on the right hand side of this equation are bounded - from the assumed continuity of the partial derivatives occurring - is expressed as follows:

$$(1.10b)\qquad (u_{i,j+1,n} + u_{i,j-1,n} + u_{i+1,j,n} + u_{i-1,j,n} - 4u_{i,j,n})/h^2$$
$$+ (u_{i,j,n} - u_{i,j,n+1})/\Delta t + \left[\frac{\partial u}{\partial t} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2}\right]_{i,j,n} = O(\Delta t) + O(h^2)$$
$$\text{as } \Delta t,\ h \to 0.$$

This equation is to be interpreted as meaning that there exist two positive constants A and B such that the absolute value of the left hand side of (1.10b) is less than or equal to $A\Delta t + Bh^2$ for all sufficiently small Δt and h. Such a difference scheme is said to be second order correct in space and first order correct in time.

The estimate (1.10b) of the truncation error holds in particular if u is the exact solution of the differential equation and, in this case, the second term on the loft hand side of (1.10b) vanishes. Furthermore, using

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^4 u}{\partial x^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4}$$

from the differential equation, and $\Delta t = rh^2$, the right hand side of (1.10a) becomes

$$-h^4\left[\tfrac{1}{2}r\left(r - \tfrac{1}{6}\right)\left(\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}\right) + r^2\frac{\partial^4 u}{\partial x^2 \partial y^2}\right]_{i,j,n}$$

and is said to be of order h^4, [see Lowen [27], section VI].

This error gives rise to the question of whether the theoretical

solution of the difference equation tends to that of the differential equation as the mesh is made finer and finer, that is, the problem of convergence. It was found some years ago by Courant, Friedrichs and Lewy [9] that certain conditions must hold between the increments of the independent variables for convergence to occur.

Round-off error causes the difference between 2) and 3) and can give rise to the phenomenon of instability. Roughly speaking, if a difference equation is unstable and if an error such as a rounding error is made at any stage of the computation, this error will increase exponentially as the number of time steps increases, and, eventually, the calculated values of the solution of the difference equation will bear no resemblance to the theoretical solution. Von Neumann [31] proposed a criterion for investigating this problem for linear equations with constant coefficients, and his criterion showed that stability occurred when precisely the same ratios as called for by Courant, Friedrichs and Lewy held. Douglas [17] has since shown that, for "wide classes" of difference analogues of linear parabolic and hyperbolic differential equations, stability in the sense of Von Neumann implies convergence.

Various aspects of stability have been considered by several authors. Lax and Richtmyer [24] give a definition of stability in terms of uniform boundedness of a certain set of operators. They then show that under suitable conditions, for linear initial value problems, stability is necessary and sufficient for convergence in a certain uniform sense for arbitrary initial data. Todd [43] and Lowan [27] carry out stability

analyses by means of the direct estimation of the eigenvalues of the matrices induced by difference operators.

Since the difference equations with which we are concerned in this thesis are linear with constant coefficients, Von Neumann's method is used to examine stability. Using operator techniques similar to those of Todd [43] and Lowan [27], we now give a definition of stability in the sense of Von Neumann. We shall consider for the moment the numerical solution of a linear partial differential equation in the region $0 \leq x,y \leq 1$, $t \geq 0$ for which the solution is specified on the boundaries and sufficient initial conditions (that is, value of the function for the parabolic case, the function and its t-derivative for the hyperbolic case, or more if of higher order) are given to ensure the existence of the solution of the problem. Let $(N+1)h = 1$, $x_i = ih$, $y_j = jh$, $t_n = n\Delta t$ and denote $f(x_i,y_j,t_n)$ by $f_{i,j,n}$. Let $u_{i,j,n}$ be the solution of the difference analogue of the differential equation. Further, let $\underline{U}_n$ represent the column vector $(u_{1,1,n}, u_{1,2,n}, \ldots, u_{N,N,n})$. Then, any partial difference equation may be written in the following matrix form:

$$A_1\underline{U}_{n+1} = A_0\underline{U}_n + A_{-1}\underline{U}_{n-1} + \cdots + A_{-q}\underline{U}_{n-q} + \underline{b}_n, \qquad n \geq q, \tag{1.11}$$

where A_1 is non-singular and $\underline{b}_n$ contains the boundary conditions. (A_1 is the unit matrix for explicit equations.) For (1.11) to be practicable, it is necessary that $\underline{U}_0, \underline{U}_1, \ldots, \underline{U}_q$ be known initially. We shall assume that these starting values can be obtained by some procedure, and, moreover, that they are close to the values of the solution of the differential equation. We now perform a stability analysis on (1.11). Let an error be introduced by round-off or some

other source into the vector $\underline{U}_m$ and call it $\underline{U}^*_m$. Then, step ahead using (1.11) and call the perturbed solution $\underline{U}^*_n$. The resulting equations become

$$(1.12)\quad \begin{cases} A_1\underline{U}^*_{m+1} = A_0\underline{U}^*_m + A_{-1}\underline{U}_{m-1} + \cdots + A_{-q}\underline{U}_{m-q} + \underline{b}_m \\ A_1\underline{U}^*_{m+2} = A_0\underline{U}^*_{m+1} + A_{-1}\underline{U}^*_m + \cdots + A_{-q}\underline{U}_{m-q+1} + \underline{b}_{m+1} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ A_1\underline{U}^*_{n+1} = A_0\underline{U}^*_n + A_{-1}\underline{U}^*_{n-1} + \cdots + A_{-q}\underline{U}^*_{n-q} + \underline{b}_n, \qquad n \geqslant m+q . \end{cases}$$

Let

$$(1.13)\qquad \underline{\varepsilon}_n = \underline{U}^*_n - \underline{U}_n .$$

Then

$$(1.14)\quad \begin{cases} A_1\underline{\varepsilon}_{n+1} = A_0\underline{\varepsilon}_n + A_{-1}\underline{\varepsilon}_{n-1} + \cdots + A_{-q}\underline{\varepsilon}_{n-q}, & n \geqslant m+q \\ \underline{\varepsilon}_m, \underline{\varepsilon}_{m+1}, \cdots, \underline{\varepsilon}_{m+q} \text{ determined by (1.12).} \end{cases}$$

We note that (1.14) is nothing more than the homogeneous equation corresponding to (1.11). As A_i is time-independent, m may be taken to be zero.

Von Neumann's technique is to apply the method of separation of variables to (1.14), (see O'Brien et al [31]). For this method to be applicable, it is necessary that the matrices A_i, $i = 1, 0, \cdots, -q$, possess a common set of eigenfunctions φ_{ps}, $p,s = 1, \cdots, N$. This is equivalent to requiring that the matrices A_i commute. We allow the identity matrix to have any complete orthonormal set as eigenfunctions. Consequently, we must restrict ourselves to such difference equations. Let

$$(1.15)\qquad A_i\,\varphi_{ps} = \lambda^{(i)}_{ps}\,\varphi_{ps}, \quad p,s = 1, \cdots, N;\ i = 1, 0, \cdots, -q.$$

Assume a solution of (1.14) of the form

$$\underline{\varepsilon}_n = \sum_{ps=1}^{N} a_n^{ps} \underline{\varphi}_{ps}.$$

Since the error equation is linear, we need only consider one term in this expansion, and, in fact, we may take

(1.16) $$\underline{\varepsilon}_n = a_n \underline{\varphi}_{ps}. \quad \ldots$$

Then

(1.17) $$\lambda_{ps}^{(1)} a_{n+1} - \lambda_{ps}^{(0)} a_n - \ldots - \lambda_{ps}^{(-q)} a_{n-q} = 0.$$

It is well-known (Milne Thomson [29]) that the general solution of (1.17) is

(1.18) $$a_n = \sum_{j=1}^{q+1} c_j \rho_{ps,j}^n$$

where the $\rho_{ps,j}$ are the distinct roots of

(1.19) $$\lambda_{ps}^{(1)} \rho^{q+1} - \lambda_{ps}^{(0)} \rho^q - \ldots - \lambda_{ps}^{(-q)} = 0.$$

If a root ρ is repeated t times, a term

$$\rho^n (d_0 + d_1 n + \ldots + d_{t-1} n^{t-1})$$

replaces the corresponding t terms above. Note that a_n grows if $\max |\rho_{ps,j}| > 1$, decreases if $\max |\rho_{ps,j}| < 1$, remains bounded if $\max |\rho_{ps,j}| = 1$ and no multiple root has modulus one, and grows slowly if a multiple root has modulus one. Thus, only in the first case does the error grow significantly in a reasonable number of time steps. Hence, the following definition of stability will be adopted. Equation (1.11) is <u>stable</u> if and only if

(1.20) $$\max |\rho_{ps,j}| \leq 1, \quad j = 1, \ldots, q+1; \; p,s = 1, \ldots, N.$$

In this thesis, we shall consider only the case where the roots are distinct.

Let us now apply this analysis to the forward and backward difference formulae derived previously for the solution of (1.3). For each of

these formulae, the eigenfunctions (not normalised) may be taken to be

$$\varphi_{ps} = [\sin \pi p x_i \sin \pi s y_j], \qquad p,s,i,j = 1, \ldots, N, \tag{1.21}$$

(see Rutherford [38]). It may readily be seen that the eigenvalues of the A_1 matrix are, respectively,

$$\begin{cases} \lambda_{ps} = 1 \\ \lambda_{ps} = 1/r + 4[\sin^2 \pi p/2(N+1) + \sin^2 \pi s/2(N+1)]. \end{cases} \tag{1.22}$$

Moreover, the <u>stability ratios</u> are, respectively,

$$\rho = 1 - 4r[\sin^2 \pi p/2(N+1) + \sin^2 \pi s/2(N+1)] \tag{1.23a}$$

$$\rho = 1/[1 + 4r(\sin^2 \pi p/2(N+1) + \sin^2 \pi s/2(N+1)]. \tag{1.23b}$$

Thus, from (1.20), the forward difference formula is stable if $0 \leq r \leq \frac{1}{4}$ while the backward difference formula is <u>unconditionally stable</u>, that is, stability is guaranteed for all positive values of r.

A simpler method of examining stability of difference formulae based on the above analysis is the following. Instead of considering the error vector $\underline{\varepsilon}_n$, we shall cosider one of its components, $\varepsilon_{i,j,n}$, and, from (1.16), we put

$$\varepsilon_{i,j,n} = a_n \sin \pi p x_i \sin \pi s y_j, \qquad p,s = 1, \ldots, N, \tag{1.24}$$

where, in general, the difference equation applied to the node $(ih, jh, n\Delta t)$ does not involve boundary points. The error component $\varepsilon_{i,j,n}$ then satisfies the same equation as $u_{i,j,n}$. For example, considering once again (1.6), the error equation is found by replacing u by ε. Using (1.24), we find that

$$\frac{a_{n+1}}{a_n} = 1 - 4r[\sin^2 \pi p/2(N+1) + \sin^2 \pi s/2(N+1)]$$

which is seen to be (1.23a). We then require $|\rho| = \left|\frac{a_{n+1}}{a_n}\right| \leq 1$ as before. This is the procedure used in the remainder of this thesis.

I.4. Convergence of Iterative Procedures.

The method discussed in I.3 for examining the stability of a finite difference replacement of a partial differential equation may be used to examine the convergence of an iterative procedure for the solution of an elliptic difference equation. As an illustration, consider the iterative formula (1.9) for the solution of (1.8). It may be written in the matrix form

$$\underline{U}^{(n+1)} = A\underline{U}^{(n)} + \underline{b} \tag{1.25}$$

where the vector $\underline{U}^{(n)} = (u_{1,1}^{(n)}, \ldots\ldots, u_{N,N}^{(n)})$ and the vector $\underline{b}$ is determined by the given boundary data. We define the error vector,

$$\underline{\varepsilon}_n = \underline{U}^{(n)} - \underline{U},$$

as the discrepancy between the approximate solution $\underline{U}^{(n)}$ and the exact solution $\underline{U}$ of the system of equations

$$(I - A)\underline{U} = \underline{b}.$$

The object of the iterative procedure is thus to reduce the components of the error vector to zero. For (1.25), the error equation is

$$\underline{\varepsilon}_{n+1} = A\underline{\varepsilon}_n. \tag{1.26}$$

As in I.3, we consider

$$\underline{\varepsilon}_n = a_n \varphi_{ps}$$

where φ_{ps}, $p,s, = 1, \ldots, N$, are the eigenfunctions of A. As before,

$$\varphi_{ps} = [\sin \pi p x_i \sin \pi s y_j] \qquad p,s = 1, \ldots, N.$$

Thus, for the procedure to converge, we require that

$$\max |\lambda_{ps}| \leq 1 \qquad p,s = 1, \ldots, N, \tag{1.27}$$

that is, the maximum modulus eigenvalue of A, which we shall call the convergence factor of the iterative procedure, or spectral radius of A

must be of modulus less than or equal to one. For this example,

$$\lambda_{ps} = 1 - 4\alpha[\sin^2 \pi p/2(N+1) + \sin^2 \pi s/2(N+1)]$$

and so, for convergence, we require

(1.28) $$\alpha \leq \tfrac{1}{4} .$$

In order to obtain the best, that is , the most rapidly convergent, iterative formula, we minimise $\max |\lambda_{ps}|$ with respect to α so that the components of the error vector are reduced to zero as quickly as possible. A simple calculation shows that the required value of α is

$$\alpha = \tfrac{1}{4} .$$

The <u>optimum convergence factor</u> is then given by

$$\min_{\alpha} \left\{ \max_{p,s} |\lambda_{ps}| \right\} = \cos \pi/(N+1).$$

In chapters II and III, the convergence of iterative schemes is investigated by considering only one component of the error vector $\underline{\varepsilon}_n$, as in I.3, and optimum convergence factors are found as in the above example.

CHAPTER II.

IMPROVED ALTERNATING DIRECTION IMPLICIT METHODS FOR PARABOLIC AND ELLIPTIC EQUATIONS IN TWO SPACE VARIABLES.

II.1. Introduction.

Alternating direction implicit (ADI) methods are usually n-step procedures for the numerical solution of partial differential equations in n space variables. The difference equations used at each step are implicit, but the systems of equations arising are of a particularly simple form and may be solved by a direct non-iterative method. The first ADI method was introduced by Peaceman and Rachford [34] in 1955 for the numerical solution of the diffusion equation in two space variables, and the term "alternating direction method" arises from the fact that in the first step of this method we solve along horizontal mesh lines and in the second step we solve along vertical mesh lines, or vice versa. This will become apparent later in this chapter.

In 1956, Douglas and Rachford [12] formulated an ADI method for the solution of the diffusion equation in two space variables which could also be extended to three space variables. Since then, variants of these methods have been used to solve various partial differential equations , (see [6], [7], [25]), several of which will be discussed in this thesis. In this chapter, we now consider the diffusion equation and Laplace's equation in two space variables, and derive new ADI methods for the solution of these equations.

II.2. The ADI Methods of Peaceman, Douglas and Rachford.

Consider the diffusion equation in two space variables

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\partial u}{\partial t} \quad , \tag{2.1}$$

where the temperature u is a function of the space coordinates x,y and the time t. The initial and boundary conditions are $u(x,y,0) = f(x,y)$ over the unit square $0 < x,y < 1$ and $u(x,y,t) = 0$ for $t \geq 0$ at points on the boundary of the unit square, respectively. The assignment $u = 0$ on the boundary involves no essential loss of generality from an arbitrary, sufficiently smooth specification of boundary values. Other boundary conditions are considered in II.7 .

The ADI method of Peaceman and Rachford [34], (see also Douglas [14]), which is referred to as the P.R. method, may be written in the form

$$(-\tfrac{1}{2}r\delta_x^2 + 1)u_{m+\frac{1}{2}} = (\tfrac{1}{2}r\delta_y^2 + 1)u_m \tag{2.2a}$$

$$(-\tfrac{1}{2}r\delta_y^2 + 1)u_{m+1} = (\tfrac{1}{2}r\delta_x^2 + 1)u_{m+\frac{1}{2}} \tag{2.2b}$$

where δ_x, δ_y are the usual central difference operators in the x and y directions respectively, u_m, $u_{m+\frac{1}{2}}$ and u_{m+1} are the values of u at the nodes $(i\Delta x, j\Delta y, m\Delta t)$, $(i\Delta x, j\Delta y, (m+\frac{1}{2})\Delta t)$ and $(i\Delta x, j\Delta y, (m+1)\Delta t)$ respectively, $(i,j = 1, 2, \ldots, N-1;\ m = 1, 2, \ldots)$, Δx, Δy and Δt are the mesh lengths in the x, y and t directions respectively and $r = \Delta t/h^2$, where $\Delta x = \Delta y = h$, is the mesh ratio. Elimination of $u_{m+\frac{1}{2}}$ from (2.2) leads to

$$(-\tfrac{1}{2}r\delta_x^2 + 1)(-\tfrac{1}{2}r\delta_y^2 + 1)u_{m+1} = (\tfrac{1}{2}r\delta_x^2 + 1)(\tfrac{1}{2}r\delta_y^2 + 1)u_m, \tag{2.3}$$

where $(-\frac{1}{2}r\delta_x^2 + 1)(-\frac{1}{2}r\delta_y^2 + 1)$ and $(\frac{1}{2}r\delta_x^2 + 1)(\frac{1}{2}r\delta_y^2 + 1)$ are both nine-

point operators.

An alternative form of the P.R. method, given by Douglas [11], is

$$(-\tfrac{1}{2}r\delta_x^2 + 1)u^*_{m+1} = (\tfrac{1}{2}r\delta_x^2 + r\delta_y^2 + 1)u_m \tag{2.4a}$$

$$(-\tfrac{1}{2}r\delta_y^2 + 1)u_{m+1} = u^*_{m+1} - \tfrac{1}{2}r\delta_y^2 u_m \tag{2.4b}$$

where u^*_{m+1} denotes an approximation to u_{m+1}. Elimination of u^*_{m+1} again yields (2.3). The value of this formulation will become evident when alternating direction methods for three space variables are considered in Chapter III. However, at present, we consider the P.R. method in the more convenient form (2.2).

The second ADI method, <u>the D.R. method</u>, was formulated by Douglas and Rachford [12] and is given by

$$(-r\delta_x^2 + 1)u^*_{m+1} = (r\delta_y^2 + 1)u_m \tag{2.5a}$$

$$(-r\delta_y^2 + 1)u_{m+1} = u^*_{m+1} - r\delta_y^2 u_m \tag{2.5b}$$

u^*_{m+1} again denoting an approximation to u_{m+1}. Formulae (2.5) lead to

$$(-r\delta_y^2 + 1)(-r\delta_x^2 + 1)u_{m+1} = (1 + r^2\delta_x^2\delta_y^2)u_m \tag{2.6}$$

after elimination of u^*_{m+1}, where $(-r\delta_y^2 + 1)(-r\delta_x^2 + 1)$ and $(1 + r^2\delta_x^2\delta_y^2)$ are, once more, nine-point operators.

A <u>generalised formula</u> which involves nine-point operators on each of two neighbouring levels of time is

$$[1 + A(\delta_x^2 + \delta_y^2) + B\delta_x^2\delta_y^2]u_{m+1} = [1 + C(\delta_x^2 + \delta_y^2) + D\delta_x^2\delta_y^2]u_m, \tag{2.7}$$

where A, B, C and D are functions of r. Formulae (2.3) and (2.6) can

both be expressed in the form of (2.7). In fact, the coefficients are

$$A = -\tfrac{1}{2}r, \quad B = \tfrac{1}{4}r^2, \quad C = \tfrac{1}{2}r, \quad D = \tfrac{1}{4}r^2,$$

for the P.R. method, and

$$A = -r, \quad B = r^2, \quad C = 0, \quad D = r^2,$$

for the D.R. method.

Also, expanding (2.3) and (2.6) as Taylor series in terms of u_m and its derivatives and replacing derivatives with respect to t using the relations $\frac{\partial u}{\partial t} = \nabla^2 u$, $\frac{\partial^2 u}{\partial t^2} = \nabla^4 u$, etc., from (2.1), where $\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$, it is easily shown that the principal parts of the truncation errors are

$$\tfrac{1}{12}rh^4\left(\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}\right)$$

for the P.R. method and

$$\tfrac{1}{2}r(r + \quad)h^4\left(\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}\right) + r^2h^4\frac{\partial^4 u}{\partial x^2 \partial y^2}$$

for the D.R. method.

In the first section of this chapter, values of A, B, C and D are found which eliminate the terms of order h^4 in formula (2.7) and the resulting formula is rewritten as a pair of P.R. or D.R. type formulae.

II.3. Generalized P.R. and D.R. Formulae.

(Mitchell and Fairweather [19])

In its present form, formula (2.7) is of little use as a means of solving (2.1) since, if Nh = 1, it requires at each time step the solution of $(N-1)^2$ linear equations in $(N-1)^2$ unknowns. However, if (2.7) can be written as a pair of P.R. or D.R. type formulae, that is, a pair of formulae which utilises the same points as the P.R. or D.R. formulae, the numerical calculation involves the solution of (N-1) equations in (N-1) unknowns first along lines in the x direction and next along lines in the y direction. At every stage of this scheme, the matrices arising are tridiagonal, or Jacobi, matrices. Such systems of equations can be solved directly by the algorithm outlined in Appendix II.A at the end of this chapter.

Considering first the points used by the P.R. formulae, we may write (2.7) as

$$(A\delta_x^2 + 1)u_{m+\frac{1}{2}} = (C\delta_y^2 + 1)u_m \tag{2.8a}$$

$$(A\delta_y^2 + 1)u_{m+1} = (C\delta_x^2 + 1)u_{m+\frac{1}{2}} \tag{2.8b}$$

provided $B = A^2$ and $D = C^2$. Next using the same points as the D.R. formulae, (2.7) can be written as

$$(A\delta_x^2 + 1)u^*_{m+1} = [(C - \tfrac{D}{A})\delta_y^2 + (1 - \tfrac{C}{A})]u_m \tag{2.9a}$$

$$(A\delta_y^2 + 1)u_{m+1} = (\tfrac{D}{A}\,\delta_y^2 + \tfrac{C}{A})u_m + u^*_{m+1} \tag{2.9b}$$

provided $B = A^2$.

Formulae (2.8) and (2.9) are the generalized P.R. and D.R. type formulae respectively and formula (2.7) cannot in general be represented by (2.8) or (2.9). If, however, $B = A^2$, (2.7) can be written as the D.R. type formulae (2.9), and if, in addition, $D = 0^2$, (2.7) can be written as the P.R. type formulae (2.8).

II.4. The Optimum Formula.

We now expand the terms u_{m+1}, $(\delta_x^2 + \delta_y^2)u_{m+1}$, etc., in (2.7) as Taylor series in terms of u_m and its derivatives, replacing time derivatives by $\frac{\partial u}{\partial t} = \nabla^2 u$, $\frac{\partial^2 u}{\partial t^2} = \nabla^4 u$, etc., from (2.1) where $\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. The expansions up to and including terms involving h^6 are

$$
\begin{aligned}
u_m &= u \\
u_{m+1} &= u + rA_1 + \tfrac{1}{2}r^2A_2 + r^2A_3 + \tfrac{1}{6}r^3A_4 + \tfrac{1}{2}r^3A_5 \\
(\delta_x^2 + \delta_y^2)u_{m+1} &= A_1 + (r + \tfrac{1}{12})A_2 + 2rA_3 + (\tfrac{1}{2}r^2 + \tfrac{1}{12}r + \tfrac{1}{360})A_4 \\
&\quad + (\tfrac{3}{2}r^2 + \tfrac{1}{12}r)A_5 \\
\delta_x^2\delta_y^2 u_{m+1} &= A_3 + (r + \tfrac{1}{12})A_5 \\
(\delta_x^2 + \delta_y^2)u_m &= A_1 + \tfrac{1}{12}A_2 + \tfrac{1}{360}A_4 \\
\delta_x^2\delta_y^2 u_m &= A_3 + \tfrac{1}{12}A_5
\end{aligned}
$$

where

$$A_1 = h^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right), \quad A_2 = h^4\left(\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}\right), \quad A_3 = h^4\frac{\partial^4 u}{\partial x^2 \partial y^2},$$

$$A_4 = h^6\left(\frac{\partial^6 u}{\partial x^6} + \frac{\partial^6 u}{\partial y^6}\right), \quad A_5 = h^6\frac{\partial^4}{\partial x^2 \partial y^2}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right).$$

If these expressions are substituted into formula (2.7), values of A, B, C and D can be found which will eliminate A_1, A_2, A_3 and either A_4 or A_5. However, a formula of type (2.7) is only of use as a means of solving (2.1) if it can at least be written in the form (2.9). This will be possible only if $B = A^2$. Hence, the coefficients in the optimum formula are chosen so that they eliminate A_1, A_2 and A_3 and satisfy $B = A^2$. These coefficients are

$$A = -\tfrac{1}{3}(r - \tfrac{1}{6}), \qquad B = \tfrac{1}{4}(r - \tfrac{1}{6})^2,$$
$$C = \tfrac{1}{3}(r + \tfrac{1}{6}), \qquad D = \tfrac{1}{4}(r + \tfrac{1}{6})^2.$$

The formula of type (2.7) with minimum truncation error which can be written in the form (2.9) is thus

$$(2.10)\quad [1 - \tfrac{1}{3}(r - \tfrac{1}{6})(\delta_x^2 + \delta_y^2) + \tfrac{1}{4}(r - \tfrac{1}{6})^2\delta_x^2\delta_y^2]u_{m+1}$$
$$= [1 + \tfrac{1}{3}(r + \tfrac{1}{6})(\delta_x^2 + \delta_y^2) + \tfrac{1}{4}(r + \tfrac{1}{6})^2\delta_x^2\delta_y^2]u_m .$$

The principal part of the truncation error is in fact

$$-\tfrac{1}{12}r(r^2 - \tfrac{1}{20})h^6\left(\frac{\partial^6 u}{\partial x^6} + \frac{\partial^6 u}{\partial y^6}\right),$$

which is an order h^2 better than the original D.R. or P.R. formula. It also follows that if r is approximately equal to $1/2\sqrt{5}$, the truncation error may even be of order h^8. When $r = \frac{1}{6}$, (2.10) degenerates into the well-known explicit formula

$$u_{m+1} = [1 + \tfrac{1}{6}(\delta_x^2 + \delta_y^2) + \tfrac{1}{36}\delta_x^2\delta_y^2]u_m$$

and so the method of alternating directions is inapplicable.

Formula (2.10) can now be written as the D.R. type formulae

$$(2.11a)\qquad [-\tfrac{1}{3}(r - \tfrac{1}{6})\delta_x^2 + 1]u^*_{m+1} = \frac{2r}{(r - \frac{1}{6})}[\tfrac{1}{3}(r + \tfrac{1}{6})\delta_y^2 + 1]u_m$$

$$(2.11b)\qquad [-\tfrac{1}{3}(r - \tfrac{1}{6})\delta_y^2 + 1]u_{m+1} = -\frac{r + \frac{1}{6}}{r - \frac{1}{6}}[\tfrac{1}{3}(r + \tfrac{1}{6})\delta_y^2 + 1]u_m + u^*_{m+1}$$

and, since the optimum coefficients satisfy $D = C^2$ as well as $B = A^2$, (2.10) can also be written as the P.R. formulae

$$(2.12a)\qquad [-\tfrac{1}{3}(r - \tfrac{1}{6})\delta_x^2 + 1]u_{m+\frac{1}{2}} = [\tfrac{1}{3}(r + \tfrac{1}{6})\delta_y^2 + 1]u_m$$

$$(2.12b)\qquad [-\tfrac{1}{3}(r - \tfrac{1}{6})\delta_y^2 + 1]u_{m+1} = [\tfrac{1}{3}(r + \tfrac{1}{6})\delta_x^2 + 1]u_{m+\frac{1}{2}} .$$

II.5. Alternative Derivation of Optimum Formula.

Originally, it was not recognised by Peaceman, Douglas and Rachford that the P.R. and D.R. methods were intimately associated with the Crank-Nicolson and backward difference analogues of the diffusion equation respectively. The connections between the P.R. and D.R. methods and these simpler formulae become apparent when equations (2.3) and (2.6) are rewritten in the forms

$$(2.13)\quad u_{m+1} = u_m + \tfrac{1}{2}r(\delta_x^2 + \delta_y^2)(u_{m+1} + u_m) - \tfrac{1}{4}r^2\delta_x^2\delta_y^2(u_{m+1} - u_m)$$

and

$$(2.14)\quad u_{m+1} = u_m + r(\delta_x^2 + \delta_y^2)u_{m+1} - r^2\delta_x^2\delta_y^2(u_{m+1} - u_m)$$

respectively. From formula (2.13), we see that the P.R. method is a perturbation of the Crank-Nicolson formula and is thus second-order correct in both space and time, and, from (2.14), the D.R. method is clearly a perturbation of the backward difference formula and is thus second-order correct in space but only first-order correct in time. (See Richtmyer [37] Table I page 93).

Consider the formula

$$(2.15)\quad u_{m+1} = u_m + (\delta_x^2 + \delta_y^2)[\tfrac{1}{2}(r + \tfrac{1}{6})u_m + \tfrac{1}{2}(r - \tfrac{1}{6})u_{m+1}]$$

which is a two-dimensional analogue of the optimum formula for (2.1) in one space variable, (see Richtmyer [37] page 95 12), and is second - order correct in both space and time. By considering a perturbation of formula (2.15), we can not only obtain a formula which is fourth-order correct in space and second-order correct in time but also write this

formula as a pair of P.R. or D.R. type formulae. Suppose this perturbation is of the form

$$(2.16)\quad u_{m+1} - u_m - (\delta_x^2 + \delta_y^2)[\tfrac{1}{12}(r + \tfrac{1}{6})u_m + \tfrac{1}{12}(r - \tfrac{1}{6})u_{m+1}] + \delta_x^2\delta_y^2(pu_{m+1} - qu_m) = 0$$

where p and q are parameters which are functions of r. Writing (2.16) in the form (2.7), we find that

$$A = -\tfrac{1}{12}(r - \tfrac{1}{6}), \quad B = p, \quad C = \tfrac{1}{12}(r + \tfrac{1}{6}), \quad D = q.$$

Hence, if $p = \frac{1}{4}(r - \frac{1}{6})^2$ and $q = \frac{1}{4}(r + \frac{1}{6})^2$, formula (2.16) can be written as a pair of P.R. or D.R. type formulae and, in fact, (2.16) is then the optimum formula (2.10) derived in II.3, which, as we have seen, is fourth-order correct in space and second-order correct in time.

II.6. Stability of The P.R., D.R. and Optimum Formulae.

The stability of formulae of type (2.7) is analyzed by the procedure outlined in I.3 . We assume that there exists an error $\varepsilon_{i,j,m}$ at each mesh point $(i\Delta x, j\Delta y, m\Delta t)$, $(i,j = 1, 2, \ldots, N-1;$ $m = 0, 1, 2, \ldots)$, where $\Delta x = \Delta y = 1/N$. If the error is now expanded in the form

$$\varepsilon_{i,j,m} = a_m \sin \pi p x_i \sin \pi q y_j \qquad (p,q = 1, 2, \ldots, N-1),$$

where $x_i = i\Delta x$, $y_j = j\Delta y$, and substituted into equation (2.7) with u replaced by ε and $\Delta x = \Delta y = h = 1/N$, it follows that

$$(2.17) \qquad \frac{a_{m+1}}{a_m} = \frac{1 - 4C(\sin^2 \pi p/2N + \sin^2 \pi q/2N) + 16D\sin^2 \pi p/2N \sin^2 \pi q/2N}{(-4A\sin^2 \pi p/2N + 1)(-4A\sin^2 \pi q/2N + 1)}$$

where use has been made of

$$\delta_x^2 \sin \pi p x_i = -4\sin^2 \pi p/2N \sin \pi p x_i , \quad (p = 1, 2, \ldots, N-1)$$

and $B = A^2$, the condition necessary for (2.7) to be written in D.R. form. Thus, for the D.R. method

$$\frac{a_{m+1}}{a_m} = \frac{1 + r^2 S_p^2 S_q^2}{1 + r(S_p^2 + S_q^2) + r^2 S_p^2 S_q^2}$$

where $S_p^2 = 4\sin^2 \pi p/2N$ and $S_q^2 = 4\sin^2 \pi q/2N$. This ratio clearly has an absolute value less than or equal to unity for all p,q in the range $1 \leq p,q \leq N-1$, and $r > 0$ and hence the D.R. method is unconditionally stable. It is interesting to note that if equation (2.5a) were used for every time step, then

$$\frac{a_{m+1}}{a_m} = \frac{1 - rS_p^2}{1 - rS_q^2} .$$

For some values of p,q and r, this ratio has an absolute value considerably greater than unity. Hence, such a procedure is highly unstable. A similar situation arises when formula (2.2a) is used for every time step. Formula (2.12a) cannot be used at every time step as it is inconsistent.

If $D = C^2$ in addition to $B = A^2$, equation (2.7) can then be written in the P.R. form and the numerator of (2.17) factorises to yield

$$\frac{a_{m+1}}{a_m} = \frac{(-CS_p^2 + 1)(-CS_q^2 + 1)}{(-AS_p^2 + 1)(-AS_q^2 + 1)} \quad . \tag{2.18}$$

When $A = -\frac{1}{8}r$, $C = \frac{1}{8}r$, this ratio again has an absolute value less than unity for $1 \leq p,q \leq N-1$ and $r > 0$. Hence, the P.R. method is unconditionally stable.

For the optimum formula (2.10), equation (2.18) becomes

$$\frac{a_{m+1}}{a_m} = \frac{[\frac{1}{8}(r + \frac{1}{6})S_p^2 - 1][\frac{1}{8}(r + \frac{1}{6})S_q^2 - 1]}{[\frac{1}{8}(r - \frac{1}{6})S_p^2 - 1][\frac{1}{8}(r - \frac{1}{6})S_q^2 - 1]}$$

$$= \lambda \cdot \mu \qquad \text{say.}$$

In order to prove that $-1 \leq \frac{a_{m+1}}{a_m} \leq 1$, it is sufficient to show that

$$\vee(s) = \frac{\frac{1}{8}(r + \frac{1}{6})S_s^2 - 1}{\frac{1}{8}(r - \frac{1}{6})S_s^2 + 1}$$

has an absolute value less than unity, since $\lambda = \vee(p)$ and $\mu = \vee(q)$. For this to be true, we require

$$rS_s^2 \geqslant 0$$

and

$$S_s^2 \leq 12.$$

These inequalities are satisfied for all s in the range $1 \leq s \leq N-1$ and $r > 0$. Hence, formula (2.10) is unconditionally stable.

II.7. Modified ADI Method for Non-Zero Boundary Conditions.

When the boundary conditions are non-zero, the ADI method given by (2.12) requires modification. If we consider the totality of the difference equations (2.10), we obtain a system of $(N-1)^2$ linear equations in the $(N-1)^2$ unknowns $u(ih, jh, (m+1)\Delta t)$, $(i,j = 1, 2, \ldots, N-1)$, where m is fixed $(m = 1, 2, \ldots)$, which may be written in the matrix form

$$(2.10)' \quad [I + \tfrac{1}{2}(r - \tfrac{1}{6})(H + V) + \tfrac{1}{4}(r - \tfrac{1}{6})^2 HV]\underline{u}_{m+1}$$
$$= [I - \tfrac{1}{2}(r + \tfrac{1}{6})(H + V) + \tfrac{1}{4}(r + \tfrac{1}{6})^2 HV]\underline{u}_m + \underline{c},$$

where H and V are the matrices of order $(N-1)^2$ given by

$$\begin{bmatrix} \overline{H} & & & \\ & \overline{H} & & \\ & & \ddots & \\ & & & \overline{H} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2I & -I & & \\ -I & 2I & -I & \\ & \ddots & \ddots & \ddots \\ & & -I & 2I \end{bmatrix}$$

respectively, where $\overline{H}$ is the matrix of order $(N-1)$ given by

$$\begin{bmatrix} 2 & -1 & & \\ -1 & 2 & -1 & \\ & \ddots & \ddots & \ddots \\ & & -1 & 2 \end{bmatrix}$$

and I is the unit matrix of order $(N-1)$, and

$$\underline{c} = -\tfrac{1}{2}(r - \tfrac{1}{6})\underline{k}'_{m+1} - \tfrac{1}{4}(r - \tfrac{1}{6})^2 \underline{k}''_{m+1} - \tfrac{1}{2}(r + \tfrac{1}{6})\underline{k}'_m + \tfrac{1}{4}(r + \tfrac{1}{6})^2 \underline{k}''_m,$$

where $\underline{k}'_n$ and $\underline{k}''_n$, $(n = m, m+1)$, are constant vectors whose ℓth components are obtained from the boundary values occurring in $(\delta_x^2 + \delta_y^2)u(ih, jh, n\Delta t)$ and $\delta_x^2\delta_y^2 u(ih, jh, n\Delta t)$ respectively, where $\ell = (N-1)(j-1) + i$. The modified ADI method takes the form

$$(2.12a)' \qquad [I + \tfrac{1}{12}(r - \tfrac{1}{6})H]\underline{u}_{m+\frac{1}{2}} = [I - \tfrac{1}{12}(r + \tfrac{1}{6})V]\underline{u}_m + \underline{a}$$

$$(2.12b)' \qquad [I + \tfrac{1}{12}(r - \tfrac{1}{6})V]\underline{u}_{m+1} = [I - \tfrac{1}{12}(r + \tfrac{1}{6})H]\underline{u}_m + \underline{b}$$

where

$$\underline{a} = -\frac{(r-\frac{1}{6})^2}{4r}\underline{k}'_m - \frac{(r-\frac{1}{6})^3}{8r}\underline{k}''_m - \frac{(r-\frac{1}{6})(r+\frac{1}{6})}{4r}\underline{k}'_{m+1} + \frac{(r-\frac{1}{6})(r+\frac{1}{6})^2}{8r}\underline{k}''_{m+1}$$

and

$$\underline{b} = -\frac{(r-\frac{1}{6})(r+\frac{1}{6})}{4r}\underline{k}'_m - \frac{(r-\frac{1}{6})^2(r+\frac{1}{6})}{8r}\underline{k}''_m - \frac{(r+\frac{1}{6})^2}{4r}\underline{k}'_{m+1} + \frac{(r+\frac{1}{6})^3}{8r}\underline{k}''_{m+1}.$$

This modification means that, if, in (2.12a) a boundary value $\begin{cases} u_{m+\frac{1}{2}}(\theta h, jh) \\ u_m(ih, \theta h) \end{cases}$ is involved on the $\begin{cases} \text{left} \\ \text{right} \end{cases}$ hand side, the term involving this boundary value is replaced on the right hand side by

$$(*) \qquad \begin{cases} a_1 D_y^- u_m(\theta h, jh) + b_1 D_y^+ u_{m+1}(\theta h, jh) \\ a_1 D_x^- u_m(ih, \theta h) + b_1 D_x^+ u_{m+1}(ih, \theta h) \end{cases}$$

and if, in (2.12b), a boundary value $\begin{cases} u_{m+1}(ih, \theta h) \\ u_{m+\frac{1}{2}}(\theta h, jh) \end{cases}$ is involved on the $\begin{cases} \text{left} \\ \text{right} \end{cases}$ hand side then the term involving this boundary value is replaced on the right hand side by

$$(**) \qquad \begin{cases} a_2 D_x^- u_m(ih, \theta h) + b_2 D_x^+ u_{m+1}(ih, \theta h) \\ a_2 D_y^- u_m(\theta h, jh) + b_2 D_y^+ u_{m+1}(\theta h, jh) \end{cases}$$

where $i, j = 2, \ldots, N-2$, $\theta = 0, N$,

$$a_1 = -\frac{(r-\frac{1}{6})^2}{4r}, \quad b_1 = a_2 = -\frac{(r-\frac{1}{6})(r+\frac{1}{6})}{4r}, \quad b_2 = -\frac{(r+\frac{1}{6})^2}{4r}$$

and

$$D_x^+ = [1 + (r+\tfrac{1}{6})\delta_x^2], \qquad D_x^- = [1 - (r-\tfrac{1}{6})\delta_x^2].$$

When boundary values occur on both sides of (2.12a) or (2.12b), they are replaced as prescribed by (*) or (**) and the coefficient of the boundary value common to these replacements is halved.

The ADI method (2.11) may be modified in a similar manner to take the form

$$(2.11a)' \qquad [I + \tfrac{1}{2}(r-\tfrac{1}{6})H]\underline{u}^*_{m+1} = \frac{2r}{(r-\frac{1}{6})}[I - \tfrac{1}{2}(r+\tfrac{1}{6})V]\underline{u}_m + \underline{c}$$

$$(2.11b)' \qquad [I + \tfrac{1}{2}(r-\tfrac{1}{6})V]\underline{u}_{m+1} = -\frac{(r+\frac{1}{6})}{(r-\frac{1}{6})}[I - \tfrac{1}{2}(r+\tfrac{1}{6})V]\underline{u}_m + \underline{u}^*_{m+1}.$$

The ADI methods formulated in Chapters III and V of this thesis are also modified in this way, namely, the non-zero boundary values occur only in the first step, the boundary values in the succeeding steps being considered zero.

II.8. Generalized Iterative Procedure for Laplace's Equation.

In this section, and later in Chapter III, we make use of the heuristic analogy between iterative methods for solving elliptic partial differential equations and numerical methods for solving parabolic partial differential equations, (see Varga [44], Chapter 8), in order to obtain a generalized iterative procedure for the solution of Laplace's equation.

It is an old idea that, provided $u(x,y,t) = u(x,y,0)$ for all $t > 0$ where (x,y) is a point on the boundary of the unit square, the steady-state solution of equation (2.1) is the solution of Laplace's equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \tag{2.19}$$

subject to the boundary condition $u(x,y) = g(x,y)$. We can now regard finite difference formulae for the solution of (2.1) as iterative procedures for the solution of (2.19), considering r, previously equal to $\Delta t/h^2$, as an iteration parameter. There is no certainty, however, that the difference equation which yields the most accurate solution of (2.1) will also provide the best iterative technique for solving (2.19). In order to examine this point, we return to difference formulae of type (2.7) and, again, for convenience, consider zero boundary values. The modification required when the boundary values are non-zero is stated later.

As shown in II.2, formulae of type (2.7) can be written in P.R. form if $B = A^2$ and $D = C^2$. In addition, these formulae represent (2.1) with a principal truncation error of order h^4 if the additional

condition

$$r + A - C = 0$$

is satisfied. These three relations between the four coefficients enable us to obtain the values

$$A = A, \quad B = A^2, \quad C = (r + A), \quad D = (r + A)^2,$$

for the coefficients of formulae (2.7). When these values are substituted into (2.7), we obtain the iterative formula

$$(2.20) \qquad (A\delta_x^2 + 1)(A\delta_y^2 + 1)u_{m+1} = [(r + A)\delta_x^2 + 1][(r + A)\delta_y^2 + 1]u_m$$

for solving (2.19) where u_m and u_{m+1} are now the m-th and (m+1)th estimates respectively of the value of u at the node $(i\,\Delta x, j\,\Delta y)$ where $i,j = 1, 2, \ldots, N-1$. If we put

$$A = -\tfrac{1}{2}(r + f),$$

where f is a parameter, then (2.20) becomes

$$(2.21) \quad [-\tfrac{1}{2}(r + f)\delta_x^2 + 1][-\tfrac{1}{2}(r + f)\delta_y^2 + 1]u_{m+1} = [\tfrac{1}{2}(r - f)\delta_x^2 + 1][\tfrac{1}{2}(r - f)\delta_y^2 + 1]u_m$$

which can be written as the P.R. type formulae

$$(2.22a) \qquad [-\tfrac{1}{2}(r + f)\delta_x^2 + 1]u_{m+\frac{1}{2}} = [\tfrac{1}{2}(r - f)\delta_y^2 + 1]u_m$$

$$(2.22b) \qquad [-\tfrac{1}{2}(r + f)\delta_y^2 + 1]u_{m+1} = [\tfrac{1}{2}(r - f)\delta_x^2 + 1]u_{m+\frac{1}{2}}\,.$$

Comparing formulae (2.3) and (2.21), we see that, when $f = 0$ in (2.21), we obtain the P.R. formula (2.3). In addition, $f = -\frac{1}{c}$ yields the optimum formula (2.10).

We note that, because of the lack of consistency at each half step for values of f other than $f = 0$, the ADI method defined by (2.22) is not included in the general formulation of ADI methods given recently by

Douglas and Gunn [16], unless, of course, f = 0.

When the boundary values are non-zero, (2.22) must be modified in the same manner as (2.12). In II.7, we put $\underline{k}'_m = \underline{k}'_{m+1} = \underline{k}'$ and $\underline{k}''_m = \underline{k}''_{m+1} = \underline{k}''$, where $\underline{k}'$ and $\underline{k}''$ are constant vectors, since the boundary values are independent of the number of iterations, and replace $-\frac{1}{6}$ by f. We then proceed as before with (*) and (**) simplified to give

$$\frac{r + f}{2}[1 - f\delta_y^2]u(\Theta h, jh)$$

$$\frac{r + f}{2}[1 - f\delta_x^2]u(ih, \Theta h)$$

and

$$\frac{r - f}{2}[1 - f\delta_x^2]u(ih, \Theta h)$$

$$\frac{r - f}{2}[1 - f\delta_y^2]u(\Theta h, jh)$$

respectively.

We now determine the value (or values) of the parameter f which will make (2.22) the best iteration scheme for solving (2.19).

II.9. Optimum Convergence Factor.

Using the method described in I.3 and I.4 for the analysis of convergence, we find that, for the iterative formulae (2.22)

$$\frac{a_{m+1}}{a_m} = \frac{[r - (f + [2\sin^2 \pi p/2N]^{-1})][r - (f + [2\sin^2 \pi q/2N]^{-1})]}{[r + (f + [2\sin^2 \pi p/2N]^{-1})][r + (f + [2\sin^2 \pi q/2N]^{-1})]} .$$

This means that the error at the node $(i\Delta x, j\Delta y)$ is reduced by the factor a_{m+1}/a_m after each iteration, since r (> 0) is constant and thus independent of m. As we have seen in I.4, for the iterative method to converge, we require that this factor of reduction, a_{m+1}/a_m, be less than or equal to one in absolute value, for all p,q in the range $1 \leq p,q \leq N-1$ and $r > 0$. In order that this condition be satisfied, we require

$$-(2\cos^2 \pi/2N)^{-1} \leq r < \infty . \tag{2.23}$$

However, although the iterative method may converge, it will not be an effective method unless its convergence is reasonably rapid. Thus, to obtain the best, that is, the most rapidly convergent, method, the convergence factor, which, in this case, is given by

$$\lambda = \left\{ \max_{1 \leq s \leq N-1} \left| \frac{r - (f + [2\sin^2 \pi s/2N]^{-1})}{r + (f + [2\sin^2 \pi s/2N]^{-1})} \right| \right\}^2 ,$$

must be minimised as a function of r (> 0).

To effect this minimisation, we consider the function

$$G(x,r) = \frac{r - x}{r + x} , \qquad r > 0,$$

where $0 < a \leq x \leq b$.

Now $$\frac{dG}{dx} = \frac{-2r}{(r + x)} < 0 \qquad \text{for all } r > 0$$

so that the maximum of $|G(x,r)|$ occurs at one of the end points of the interval. Hence,

$$\max_{a \le x \le b} |G(x,r)| = \max \left\{ \left| \frac{r-a}{r+a} \right|, \left| \frac{r-b}{r+b} \right| \right\}.$$

From Figure I, it is easily seen that

$$\max_{a \le x \le b} |G(x,r)| = \begin{cases} \dfrac{b-r}{b+r}, & 0 < r \le r^* \\ \dfrac{r-a}{r+a}, & r \ge r^*, \end{cases}$$

where r^* is given by

$$\frac{b-r^*}{b+r^*} = \frac{r^*-a}{r^*+a};$$

that is $r^* = \sqrt{ab}$.

Thus, it follows that

$$\begin{aligned} \min_{r>0} \left\{ \max_{a \le x \le b} |G(x,r)| \right\} &= G(a,\sqrt{ab}) \\ &= \frac{\sqrt{ab}-a}{\sqrt{ab}+a} \\ &= \frac{1-(a/b)^{\frac{1}{2}}}{1+(a/b)^{\frac{1}{2}}}. \end{aligned}$$

Thus, with $a = f + (2\cos^2 \pi/2N)^{-1}$ and $b = f + (2\sin^2 \pi/2N)^{-1}$,

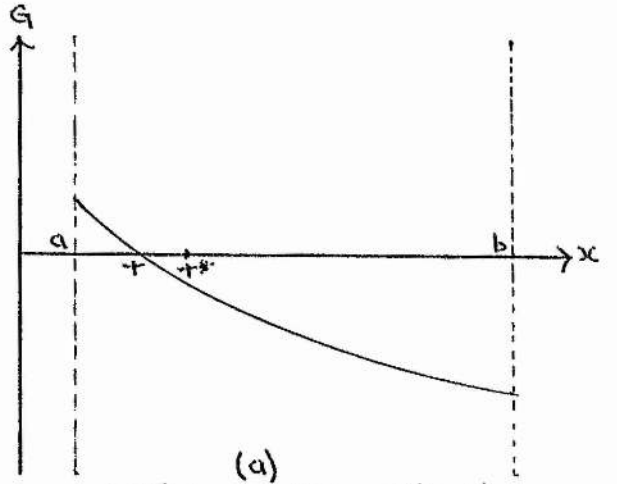
$$(2.24) \qquad \min_{r>0} \lambda = \frac{1 + f\sin^2 \pi/N - \sin^2 \pi/N \, (1 + 2f + f^2 \sin^2 \pi/N)^{\frac{1}{2}}}{1 + f\sin^2 \pi/N + \sin^2 \pi/N \, (1 + 2f + f^2 \sin^2 \pi/N)^{\frac{1}{2}}}$$

For each permissible value of f, (2.24) gives the optimum convergence factor, and the value of r necessary to achieve this optimum convergence is

$$(2.25) \qquad r^* = (1 + 2f + f^2 \sin^2 \pi/N)^{\frac{1}{2}} / \sin^2 \pi/N .$$

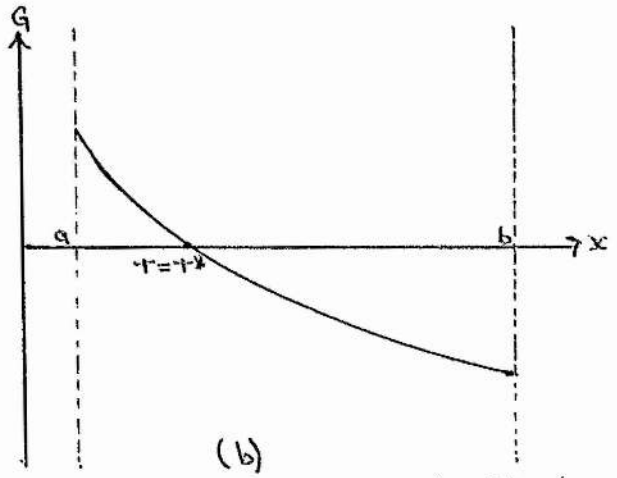
It should be emphasised that the above analysis depends on r being

FIGURE I.



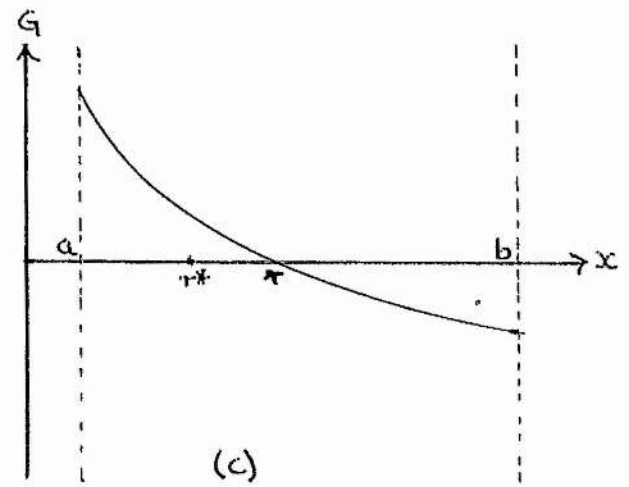
(a)

$\tau < \tau^*$: $\max|G| = G(b,\tau)$



(b)

$\tau = \tau^*$: $\max|G| = G(b,\tau) = G(a,\tau)$



(c)

$\tau > \tau^*$: $\max|G| = G(a,\tau)$

kept constant during the iterations. When r is allowed to vary and take the value r_i $(1 \leq i \leq m)$, for each of m successive iterations, the situation is more complicated. This case will be discussed in the next section of this chapter.

Returning to (2.24), it can be seen that the optimum convergence factor is unity when $f = -(2\cos^2 \pi/2N)^{-1}$, and tends to zero as f tends to infinity. For N = 10, 30 and 100, the optimum convergence factors and corresponding iteration parameters for various values of f within the permissible range given by (2.23) are shown in Tables 2.1 and 2.2 respectively. From Table 2.1, we see that the best convergence factor for a given value of N is obtained when f is positive and as large as possible. In particular, the original P.R. method (f = 0) has a better convergence factor than the optimum formula $(f = -\frac{1}{6})$ derived previously to solve (2.1).

The substantial improvement in convergence of the alternating direction method which arises from the choice of large positive values of the parameter f may, however, be accompanied by a certain loss of accuracy. The iteration procedure described by (2.21) reduces to

$$(\delta_x^2 + \delta_y^2 - f\delta_x^2\delta_y^2)u = 0, \tag{2.26}$$

when $u_{m+1} = u_m = u$, for m sufficiently large. However, equation (2.19) can be replaced by

$$(\delta_x^2 + \delta_y^2 + \tfrac{1}{6}\delta_x^2\delta_y^2)u = 0,$$

the well-known nine-point approximation to Laplace's equation which is

N \ f	$-\frac{1}{6}$	0	1	2	10
10	0.59121	0.52786	0.33664	0.24849	0.06742
30	0.84243	0.81073	0.69575	0.62637	0.38608
100	0.94998	0.93909	0.89688	0.86894	0.74998

Table 2.1 - Convergence Factors.

N \ f	$-\frac{1}{6}$	0	1	2	10
10	2.64749	3.23607	5.69354	7.50737	17.88616
30	7.81301	9.56677	16.60028	21.48524	44.96650
100	25.99470	31.83662	55.15163	71.21605	146.23423

Table 2.2 - Optimum Parameters.

correct to fourth-order differences, and so it follows from (2.26) that equation (2.21) is most accurate when $f = -\frac{1}{6}$, a result obtained previously, and that there is a loss of accuracy when f is large. In fact, as f tends to infinity, equation (2.26) degenerates into

$$(\delta_x^2 \delta_y^2)u = 0, \tag{2.27}$$

which is no longer a difference approximation to Laplace's equation. Accordingly a balance is required between the rate of convergence (an optimum when f is infinite) and the accuracy of the process (an optimum when $f = -\frac{1}{6}$). However, it will be shown in II.11, by means of a numerical example, that, when $f = -\frac{1}{6}$, the spectacular improvement in accuracy more than compensates for the slower convergence.

II.10. Methods of Selecting Sequences of Iteration Parameters.

If the iteration parameter r in (2.22) is allowed to vary and take the value r_i $(1 \leq i \leq m)$ for each of m successive iterations, then after the m-th iteration the initial error at the node $(s\Delta x, t\Delta y)$ where $s,t = 1, 2, \ldots, N-1$, will have been reduced by the factor

$$\frac{a_m}{a_o} = \prod_{i=1}^{m} \frac{(r_i - \mu)}{(r_i + \mu)} \cdot \frac{(r_i - \nu)}{(r_i + \nu)} \tag{2.28}$$

where $\mu = f + (2\sin^2 p\pi/2N)^{-1}$ and $\nu = f + (2\sin^2 q\pi/2N)^{-1}$ $1 \leq p,q \leq N-1$. Equation (2.28) follows immediately from II.9 . In order to obtain the optimum values of $r_1, r_2, r_3, \ldots, r_m$, we must minimise

$$\phi_m(a,b,\underline{r}) = \max_{a\leq\mu,\nu\leq b} \prod_{i=1}^{m} \left| \frac{(r_i - \mu)}{(r_i + \mu)} \cdot \frac{(r_i - \nu)}{(r_i + \nu)} \right| \tag{2.29}$$

$$= \left\{ \max_{a\leq\gamma\leq b} \prod_{i=1}^{m} \left| \frac{r_i - \gamma}{r_i + \gamma} \right| \right\}^2$$

where $\underline{r} = (r_1, r_2, \ldots, r_m)$ and $a = f + (2\cos^2 \pi/2N)^{-1}$, $b = f + (2\sin^2 \pi/2N)^{-1}$. The problem of minimising ϕ_m is equivalent to the problem of determining the minimax of the rational functions involved over certain domains.

In this section, methods of obtaining the optimum, or "good", that is, near optimum, parameters for the model problem - the Dirichlet problem for a unit square - will be briefly summarised.

One choice of parameters, based on an idea of Peaceman and Rachford [34], was presented by Young [52] and Young and Ehrlich [50]. These parameters are given explicitly by the formula

$$r_i = a\left(\frac{b}{a}\right)^{\frac{2i-1}{2m}}, \qquad i = 1, 2, \ldots, m, \tag{2.30}$$

where m is the smallest integer such that

$$\delta^{2m} \leq a/b \tag{2.31}$$

where $\delta = \sqrt{2} - 1 \doteq 0.414$. In practice, a number of cycles of parameters is used until the error at every node is less than or equal to some preassigned quantity. It may be shown (Birkhoff et al [3]) that, after one cycle of Young's parameters,

$$\phi_m(a,b,\underline{r}) \leq \left[\frac{1 - (a/b)^{1/2m}}{1 + (a/b)^{1/2m}}\right]^2.$$

A similar choice of parameters was presented by Wachspress [45] and Wachspress and Habetler [46]. This parameter sequence is given by

$$r_i = a\left(\frac{b}{a}\right)^{\frac{i-1}{m-1}}, \quad m \geq 2, \quad i = 1, 2, \ldots, m, \tag{2.32}$$

where m is the smallest integer such that

$$\delta^{2(m-1)} \leq a/b \tag{2.33}$$

where, once more, $\delta = \sqrt{2} - 1 \doteq 0.414$. After one cycle of these parameters

$$\phi_m(a,b,\underline{r}) \leq \left[\frac{1 - (a/b)^{1/2m-2}}{1 + (a/b)^{1/2m-2}}\right]^4.$$

Neither of these choices results in optimum values of the parameters but they do give convergence rates not far from the optimum. Young and Frank [51] have shown that the Wachspress parameters are a better choice than those of Young. Numerical experiments tend to confirm this superiority. (See Birkhoff et al [3]).

Wachspress [46] has devised an algorithm for calculating exact optimum parameters when the number of parameters is a power of 2. An independent proof of this algorithm was simultaneously given by Gastinel (unpublished). In experiments carried out by Young and Frank [51] and Birkhoff et al [3], there was very little, if any, gain by using the optimum parameters instead of the Wachspress parameters. In a later paper, Wachspress [47], this algorithm is generalized to the case in which the region under consideration is rectangular.

Recently, De Boor and Rice [10] have shown that the use of simple programming methods gives the optimum parameters for all $m \geqslant 1$. Also, they devise a new parameter sequence which appears to be relatively easy to use and gives nearly optimum results.

The most recent method of choosing the parameters is that devised by W.B.Jordan, who has obtained an exact expression for the optimum parameters in terms of Jacobian elliptic functions. From this expression, Jordan obtains other estimates of the parameters. His analysis is summarised in the Appendix of Wachspress [47].

It was originally believed (Peaceman and Rachford [34]) that, for an alternating direction method of type (2.22) to be convergent, the parameter r had to be the same for both steps of the iteration. However, Pearcy [35] has shown that (2.22) is still convergent when the parameter is changed on the half-step, but he has been unable to decide whether or not this leads to more rapid convergence.

II.11. ADI Methods and Non-Rectangular Regions.

It should be emphasised that the theory developed in the previous sections of this chapter for ADI methods only applies when the operators b_x and b_y commute, that is, if and only if the region under consideration is rectangular, (Birkhoff and Varga [2]). However, the P.R. method has been used with considerable success for the solution of the five-point Laplace difference equation in cases where the theory is known to be inapplicable. (See Young and Ehrlich [50], Price and Varga [36]). We now investigate the possibility of using (2.22) with $f \neq 0$ for the Dirichlet problem on non-rectangular regions.

We observe that a difference approximation to Laplace's equation of type (2.26), with $f \neq 0$, that is, a difference approximation involving nine points, can only be employed when the boundaries of the region under consideration are parallel to the coordinate axes, otherwise it would involve a point outside the region when applied to the nodes nearest to the bounderies. Also, for a region with boundaries parallel to the coordinate axes other than a square or rectangle, formulae (2.22) do not directly yield the solution of Laplace's equation. This may be seen by considering the totality of difference equations of type (2.26) which gives rise to a system of linear equations of the form

$$A\underline{u} = \underline{g} ,$$

where $\underline{u}$ is the vector of the unknowns $u_{i,j}$ and $\underline{g}$ is a constant vector arising from the boundary values. Writing (2.22) in matrix form, we obtain

$$[I + \tfrac{1}{2}(r + f)H]\underline{u}_{m+\frac{1}{2}} = [I - \tfrac{1}{2}(r - f)V]\underline{u}_m + \tfrac{1}{2}(r + f)\underline{g}$$

$$[I + \tfrac{1}{2}(r + f)V]\underline{u}_{m+1} = [I - \tfrac{1}{2}(r - f)H]\underline{u}_{m+\frac{1}{2}} + \tfrac{1}{2}(r - f)\underline{g} \ ,$$

where H and V are matrices which are such that if $[H\underline{u}](x_o,y_o)$ denotes the component of the vector $H\underline{u}$ corresponding to the mesh point (x_o,y_o), then

$$[H\underline{u}](x_o,y_o) = -u(x_o-h,y_o) + 2u(x_o,y_o) - u(x_o+h,y_o)$$

$$[V\underline{u}](x_o,y_o) = -u(x_o,y_o-h) + 2u(x_o,y_o) - u(x_o,y_o+h).$$

This method clearly converges to the solution of

$$(H + V + fHV)\underline{u} = \underline{g} \ ,$$

and only in the case of a square, or rectangle, is this the original set of equations, that is, A may be written in the form

$$A = H + V + fHV$$

if and only if the region is a square or rectangle in which case, if A is an N x N matrix, H and V take the forms

$$H = \begin{bmatrix} \overline{H} & & & & \\ & \overline{H} & & & \\ & & \overline{H} & & \\ & & & \ddots & \\ & & & & \overline{H} \end{bmatrix} \qquad V = \begin{bmatrix} 2I & -I & & & \\ -I & 2I & -I & & \\ & -I & 2I & -I & \\ & & & \ddots & \\ & & & -I & 2I \end{bmatrix}$$

respectively, where $\overline{H}$ is the N x N matrix

$$\overline{H} = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & -1 & 2 \end{bmatrix}$$

and I is the NxN unit matrix.

As an example of a problem involving a region on which H and V (or δ_x^2 and δ_y^2) do not commute, consider the Dirichlet problem for the region shown in Figure II, (see Varga [44] page 218, example 3), where the circled, numbered nodes are the unknowns.

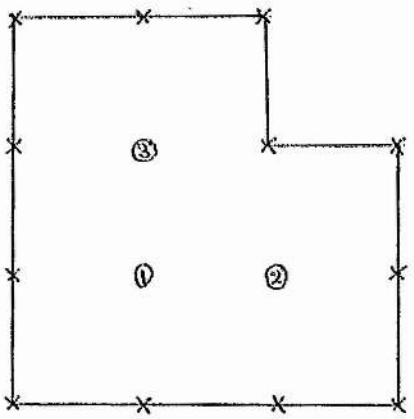


Figure II.

In this case,

$$A = \begin{bmatrix} 4(1+f) & -(1+2f) & -(1+2f) \\ -(1+2f) & 4(1+f) & f \\ -(1+2f) & f & 4(1+f) \end{bmatrix} \qquad H = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

$$V = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 2 \end{bmatrix} \qquad H + V + fHV = \begin{bmatrix} 4(1+f) & -(1+2f) & -(1+2f) \\ -(1+2f) & 4(1+f) & f \\ -(1+2f) & 0 & 4(1+f) \end{bmatrix}$$

from which it is obvious that

$$A \neq H + V + fHV$$

and hence the ADI method (2.22) does not yield the solution of the problem.

However, the ADI method (2.22) can be used on regions whose

boundaries are parallel to the axes in conjunction with the numerical alternating procedure of Miller [26], which is a numerical analogue of the Schwarz alternating procedure, (Kantorovich and Krylov [22]). This procedure enables one to solve the Dirichlet problem for Laplace's equation on the union of two overlapping plane regions, provided that the Dirichlet problem can be solved on each separately, and that their boundaries intersect at non-zero angles. The numerical alternating procedure may be described as follows: Consider the Dirichlet problem on the union R of two closed regions R_1, R_2 and suppose that it is possible to solve the Dirichlet problem on R_1 and R_2. Let $\beta_i = \partial R_i \cap R_j^0 \neq 0$, $i \neq j$, $\alpha_i = \partial R_i - \beta_i$, $i = 1, 2$ where ∂R_i denotes the boundary and R_i^0 the interior of R_i, and suppose that $\partial R = \alpha_1 \cup \alpha_2$. (See Figure III).

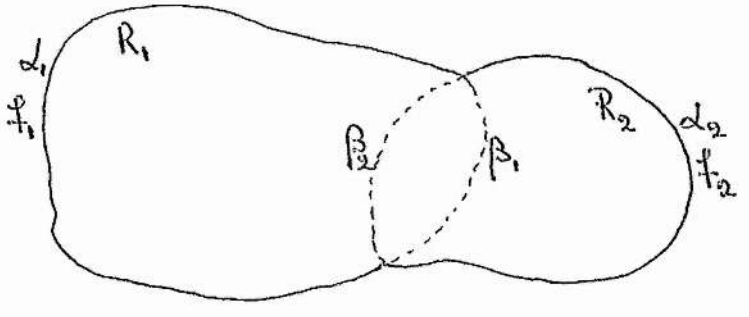


Figure III.

Let us consider a fixed Dirichlet problem, with given boundary function f on ∂R, and let u denote the solution of the problem. Let f_i be the boundary function on ∂R_i which is the restriction of f to the external arc α_i; that is, $f_i = f$ on α_i, $f_i = 0$ on β_i. The numerical alternating procedure may be written as follows:

Let $u_1^o(x,y)$ be the solution of Laplace's equation on R_1 for boundary values f_1 on α_1 and arbitrary values g_1^o on β_1;

let $u_1^n(x,y)$ be the solution of Laplace's equation on R_1 for boundary values f_1 on α_1 and $u_2^{n-1}(x,y)$ on β_1, $n = 1, 2, \ldots$;

let $u_2^n(x,y)$ be the solution of Laplace's equation on R_2 for boundary values f_2 on α_2 and $u_1^n(x,y)$ on β_2, $n = 0, 1, 2, \ldots$.

Miller [26] shows that u_1^n converges on R_1 and u_2^n converges on R_2 to the solution u of Laplace's equation on R for boundary values f on ∂R. Thus, if the region under consideration is the union of two overlapping rectangles R_1 and R_2, we may employ the ADI method (2.22) to solve Laplace's equation (2.19) on $R_1 \cup R_2$. This method may extended to cover the case of three or more overlapping regions, and, also, there is no need to restrict the region to two dimensions.

The solution of Laplace's equation on an L-shaped region by direct application of the ADI method (2.22) with $f = 0$ and in conjunction with the numerical alternating procedure with $f = -\frac{1}{6}$ is considered in II.13.

II.12. Footnote.

Originally, the formulae occurring in this chapter and in the paper by Mitchell and Fairweather [30] were not expressed in terms of central differences but in terms of a network of nodal points, which were numbered, as illustrated in Figure IV, for ease of reference. In

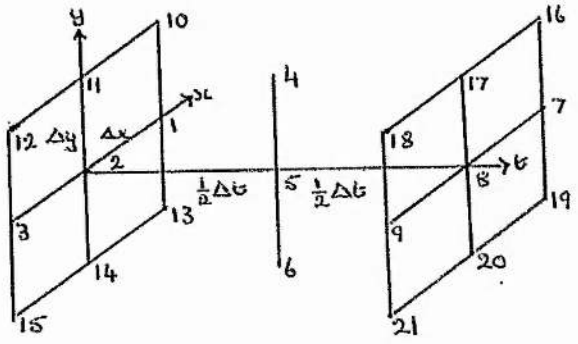


Figure IV.

particular, the generalized formula (2.7) was written in the form

(2.34) $k[(u_8 + aW + bX) + (cu_2 + dY + eZ)] = 0$

where

$$W = u_{16} + u_{18} + u_{19} + u_{21}$$

$$X = u_7 + u_9 + u_{17} + u_{20}$$

$$Y = u_{10} + u_{12} + u_{13} + u_{15}$$

$$Z = u_1 + u_3 + u_{11} + u_{14}$$

and k, a, b, c, d and e are functions of r. In order that this formula be written as a pair of P.R. or D.R. type formulae, the conditions $b^2 = a$ and $e = cd$ or simply $b^2 = a$ must be satisfied respectively. The optimum formula (2.10) was obtained by expanding u_8, W, Y, X, Z as Taylor series in terms of u and its derivatives at the node 2 and,

with k = 1, determining values of a, b, c, d and e which eliminated the terms of order h^4 in formula (2.34). The optimum formula then appeared as

$$u_8 + \frac{(r-\frac{1}{6})^2}{4(r+\frac{5}{6})^2} W - \frac{(r-\frac{1}{6})}{2(r+\frac{5}{6})} X - \frac{(r-\frac{5}{6})^2}{(r+\frac{5}{6})^2} u_2 - \frac{(r+\frac{1}{6})^2}{4(r+\frac{5}{6})^2} Y + \frac{(r-\frac{5}{6})(r+\frac{1}{6})}{2(r+\frac{5}{6})^2} Z = 0.$$

The family of iterative methods for the solution of Laplace's equation was obtained by a similar procedure. (See Mitchell and Fairweather [30]). However, we were unable to explain the behaviour of these iterative methods when the parameter f became infinite. (This parameter f was introduced in the following manner: the coefficients a, c, d and e of the iterative formula were obtained in terms of the coefficient b which was then put equal to $-(r+f)/2(r+f+1)$.)

The paper by Fairweather and Mitchell [19] was also written in terms of numbered nodes. However, when this paper was submitted to "The Computer Journal", the referee suggested that it be written in terms of central differences in order to simplify several formulae occurring in it. This suggestion was also adopted to a certain extent in the final version of the paper by Mitchell and Fairweather [30] and, as a result, the poor accuracy of our iterative formulae for large f became apparent.

All further research was carried out in terms of central differences, which have considerably simplified the formulation and analysis of ADI methods.

II.13. Numerical Experiments.

The iteration procedure (2.22) is now used to solve Laplace's equation (2.19) in two regions, one on which the matrices H and V commute, namely the unit square $0 \leq x,y \leq 1$, and on an L-shaped region where the matrices do not commute.

The purpose of the first set of experiments (A) is to compare, for various values of N and f, the accuracy and the rate of convergence of the iterative procedure (2.22) for the solution of (2.19) in the unit square. It will be shown that, particularly in the case of variable iteration parameters, (2.22) with $f = -\frac{1}{6}$ is a far superior method for solving (2.19) than the P.R. method, which is generally recognised as being the best of the existing methods. In the second set of experiments (B), with the aid of the numerical alternating procedure described in II.11, (2.22) with $f = -\frac{1}{6}$ is used to solve (2.19) in an L-shaped region obtained by removing a $6/11 \times 6/11$ square from a corner of the unit square. The accuracy and rate of convergence of this method are compared with those of the P.R. method which is used directly to solve the same problem. The greater accuracy of (2.22) with $f = -\frac{1}{6}$ as compared with the P.R. method will again be demonstrated.

(A) Square Region. The problem consists of equation (2.19) together with the boundary conditions

(2.35) $u(0,y) = u(1,y) = 0, \quad 0 \leq y \leq 1; \quad u(x,0) = u(x,1) = \sin \pi x, \quad 0 \leq x \leq 1.$

The theoretical solution of this problem is

(2.36) $$u(x,y) = \operatorname{sech} \tfrac{1}{2}\pi \cosh \pi(y - \tfrac{1}{2}) \sin \pi x.$$

Each experiment is started with $u_0(x,y) = 0$, for all (x,y) inside the unit square.

(i) Constant Iteration Parameters. The first numerical experiment is carried out to test the theoretical predictions of II.9 , where the iteration parameter r is held constant during each calculation and takes the value r* given by (2.25). The results for N = 11 (one hundred internal points) and a range of values of f are shown in Table 2.3 . Each calculation is continued until the error, that is, the difference between the theoretical and the computed solutions of (2.19) settles down in the eighth decimal place, that is, until

(2.37) $$\left|u_{m+1} - u_m\right| < 10^{-8} ,$$

and the errors are quoted at a node nearest to the centre of the square. In addition, the decrease in error as the iterations proceed is shown in Table 2.4 for three representative values of f, namely f = 1, 0 (Peaceman-Rachford method), and $-\frac{1}{6}$.

The outstanding feature of both tables is the spectacular accuracy of the ADI method (2.22) with $f = -\frac{1}{6}$. Although the convergence is slowest at this value of f, it is seen from Table 2.4 that, after twelve iterations, the accuracy of the calculation with $f = -\frac{1}{6}$ has attained the accuracy of the Peaceman-Rachford calculations and for further iterations, its accuracy is much superior.

The constant parameter case was not pursued for increased values of

Table 2.3 .

f	r	Number of Iterations	Error
$-\frac{1}{6}$	2.9029,1488	46	-0.0000,0005
$-\frac{3}{20}$	2.9734,8178	44	-0.0003,8457
$-\frac{1}{12}$	3.2412,7533	42	-0.0019,1787
0	3.5494,6553	39	-0.0038,2376
$\frac{1}{12}$	3.8347,6652	36	-0.0057,1781
$\frac{1}{6}$	4.1019,5708	34	-0.0076,0012
1	6.2286,5288	24	-0.0257,9964

Table 2.4 .

M \ f	1	0	$-\frac{1}{6}$
2	0.0644,6629	0.1995,4099	0.2471,2585
6	-0.0240,7639	0.0152,5348	0.0351,1361
10	-0.0257,6738	-0.0019,4536	0.0052,6092
14	-0.0257,9903	-0.0036,3734	0.0007,9000
18	-0.0257,9963	-0.0038,0528	0.0001,1825
22	-0.0257,9964	-0.0038,2193	0.0000,1762
26	-0.0257,9964	-0.0038,2358	0.0000,0259
30	-0.0257,9964	-0.0038,2374	0.0000,0034
34	-0.0257,9964	-0.0038,2376	0.0000,0001
38	-0.0257,9964	-0.0038,2376	-0.0000,0004
42	-0.0257,9964	-0.0038,2376	-0.0000,0005
46	-0.0257,9964	-0.0038,2376	-0.0000,0005

M = Number of Iterations.

N since the use of a sequence of iteration parameters greatly improves the convergence rate of the iterative procedure. In any case, it is clear that the pattern established in the present section for a constant parameter with N = 11 will persist for higher values of N.

(ii) Variable Iteration Parameters. It is shown by Birkhoff et al [3] that, in general, the Peaceman-Rachford method with Wachspress parameters, that is, the sequence of parameters given by (2.32), is superior to all variants of the method of successive overrelaxation for the numerical solution of the Dirichlet problem. This is particularly so when N is large. We now compare the Peaceman-Rachford method with the optimum ADI method given by (2.22) with $f = -\frac{1}{6}$, using Wachspress parameters in both cases.

The number of parameters required for a given value of N is calculated from (2.33) both for $f = 0$ and $f = -\frac{1}{6}$. The results are shown in Table 2.5. There is no significant difference in the number of parameters required in the two cases. For the values of N quoted, only N = 21 and N = 51 require one more parameter in the optimum case than in the Peaceman-Rachford case. The values of the Wachspress parameters themselves are shown in Table 2.7 for various values of N for $f = 0$ and $f = -\frac{1}{6}$.

Calculations are carried out for $f = 0$ and $-\frac{1}{6}$ with N = 11 and 15, and four iterations are required for convergence in each case. The errors at the centre nodes after each iteration are shown in Table 2.6, and the spectacular accuracy of the ADI method (2.22) with $f = -\frac{1}{6}$ is

Table 2.5.

No. of Wachspress Parameters

N \ f	0	$-\frac{1}{6}$
11	4	4
15	4	4
21	4	5
31	5	5
41	5	5
51	5	6
61	6	6
71	6	6
81	6	6
91	6	6
101	6	6
501	8	8
1001	9	9

Table 2.6.

N \ f	0	$-\frac{1}{6}$
11	0.3981,4530	0.3984,9918
	0.3697,7270	0.3871;0369
	0.1449,7010	0.1898;8553
	-0.0038,2376	-0.0000;0005
15	0.3985,1816	0.3985;2878
	0.3913,9998	0.3963;1102
	0.2160,1536	0.2557;2028
	-0.0020,7587	-0.0000,0001

Table 2.7 .

Wachspress Iteration Parameters.

N \ f	0	$-\frac{1}{6}$
11	0.5103,3610	0.3436,6943
	1.8594,9096	1.4253,9918
	6.7753,5187	5.9119,6836
	24.6870,7504	24.5204,0837
15	0.5055,2345	0.3388,5678
	2.2698,3608	1.7363,9906
	10.1917,2471	8.8978,0531
	45.7615,6548	45.5948,9882
31	0.5012,8597	0.3346,1930
	2.2259,7612	1.6435,2645
	9.8845,1701	8.0723,9507
	43.8925,0890	39.6486,2390
	194.9060,6725	194.7394,0058
51	0.5004,7462	0.3338,0795
	2.8512,7119	1.4560,3384
	16.2440,7538	6.3510,6062
	92.5446,8171	27.7026,3296
	527.2394,9560	120.8358,6026
		527.0728,2893
101	0.5001,2096	0.3334,5429
	2.6444,8828	1.9120,6939
	13.9832,5374	10.9640,4948
	73.9392,1410	62.8692,5651
	390.9681,8835	360.5003,2616
	2,067.3214,6340	2,067.1547,9673

again demonstrated. On this occasion, however, unlike the constant iteration parameter case, the number of iterations required for convergence is the same for the optimum method as for the Peaceman-Rachford method, and so the considerable improvement in accuracy is achieved without additional computation. This is true for almost all values of N.

(B) L-Shaped Region. We now consider the solution of (2.19) in the L-shaped region shown in Figure V.

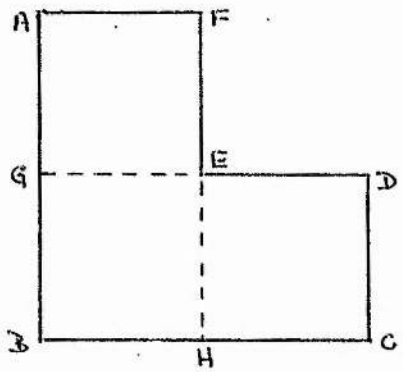


Figure V .

For convenience, the boundary values at nodes on AB, BC, CD and AF are taken from (2.35) and those at nodes on EF and ED from (2.36). Thus, the theoretical solution of the problem is again given by (2.36).

This problem is solved first by direct application of the Peaceman-Rachford method to the L-shaped region, using Rachspress parameters. Although the theory on which the determination of these parameters is based does not apply in this case, since the matrices corresponding to H and V no longer commute, it has been shown by Young and Ehrlich [50] and Price and Varga [36] that their use produces reasonably rapid convergence of the iteration procedure. For this problem, the Peaceman-

Rachford method converges after three sequences of the parameters given in Table 2.7 for N = 11, f = 0. Calculations are again carried out until (2.37) is satisfied and the maximum error, which occurs at the node x = 6/11, y = 2/11 is -0.001,696,02 , the theoretical solution at this point being 0.608,528,42.

The problem is next solved using the ADI method (2.22) with $f = -\frac{1}{6}$ together with the numerical alternating procedure. The L-shaped region is divided into two overlapping rectangles DCDG (region R_1) and ABHF (region R_2) and, initially we place $u_0(x,y) = 0$, for all (x,y) inside the region. By means of (2.22) with $f = -\frac{1}{6}$ and one sequence of the appropriate parameters, a solution is obtained in R_1 with the values of u along GE equal to zero. This calculation gives a first estimate of the function u along EH which enables a solution to be obtained in R_2, again using (2.22) with $f = -\frac{1}{6}$ and one sequence of the Wachspress parameters. A new estimate of the values along GE is thus obtained. This procedure is continued until (2.37) is satisfied. After seven applications of the alternating procedure, the maximum error is -2×10^{-8} again at the node x = 6/11, y = 2/11.

In principle, there is no limit to the number of overlapping regions to which the alternating procedure can be extended, and so theoretically the ADI method can be used to solve Laplace's equation to h^4 accuracy in any region bounded by lines parallel to the coordinate axes. In practice, however, the method becomes rather tedious if a large number of overlapping regions occurs. It should be noted that Saul'ev [53]

has also devised a technique for using ADI methods in non-rectangular regions. This method involves the solution of a related problem in the smallest rectangle which encloses the original region.

II.14. Concluding Remarks.

There have been only two previous attempts to solve the nine-point Laplace difference equation by means of an ADI procedure. The first was by Samarskii and Andreev [54] who considered alternating direction methods for solving the iterative formula

$$(1 - r\delta_x^2)(1 - r\delta_y^2)u_{m+1} = [(1 - r\delta_x^2)(1 - r\delta_y^2) + r(\delta_x^2 + \delta_y^2 + \tfrac{1}{6}\delta_x^2\delta_y^2)]u_m.$$

It is not possible to factorize the right hand side of this formula, and so it cannot be split into Peaceman-Rachford form like (2.22). As a result, the examination of the convergence of the procedure is considerably more difficult and the method more complicated than the method (2.22). The second attempt was by Cannon and Douglas [55] who proposed a three level alternating direction iterative method. The presence of the extra level, of course, adds undue complication to the numerical procedure.

Appendix IIA. Solution of Tridiagonal Systems of Equations.

The use of either (2.8) or (2.9) leads to (N-1) sets of (N-1) linear simultaneous equations of the form

$$
\begin{aligned}
& b_1 u_1 + c_1 u_2 && = d_1, \\
& a_j u_{j-1} + b_j u_j + c_j u_{j+1} && = d_j, \qquad 2 \le j \le N-2 \\
& a_{N-1} u_{N-2} + b_{N-1} u_{N-1} && = d_{N-1},
\end{aligned}
\tag{A2.1}
$$

at each time step.

Several authors, for example, Evans and Forrington [18], Bakes [1] and Ting [42], have developed direct non-iterative methods for the solution of equations of this form. The most common method is that formulated by L.H.Thomas [41]. Using this method, the solution of equations (A2.1) may be obtained in the following manner.

Let

$$
\begin{aligned}
& w_o = 0 \\
& w_j = c_j/(b_j - a_j w_{j-1}), \qquad 1 \le j \le N-2
\end{aligned}
\tag{A2.2}
$$

and

$$
\begin{aligned}
& g_o = 0 \\
& g_j = (d_j - a_j g_{j-1})/(b_j - a_j w_{j-1}), \qquad 1 \le j \le N-1.
\end{aligned}
\tag{A2.3}
$$

The solution of (A2.1) is

$$
\begin{aligned}
& u_{N-1} = g_{N-1} \\
& u_j = g_j - w_j u_{j+1}, \qquad 1 \le j \le N-2.
\end{aligned}
\tag{A2.4}
$$

Thus, w and g are computed in order of increasing j, and u is computed

in order of decreasing j.

While this method is equivalent to plain Gaussian elimination, it avoids the error growth associated with the back substitution in the elimination method and also minimises the storage problems in machine computation. A proof, by matrix algebra, of the method employing equations (A2.1) to (A2.4) is given by Bruce et al [5].

CHAPTER III.

IMPROVED ADI METHODS FOR THREE SPACE VARIABLES.

III.1. Introduction.

Having obtained an efficient ADI method for the solution of the diffusion equation in two space variables, Douglas and Rachford [12] then made the obvious generalization of this method to three space variables. The resulting formula, when applied to the diffusion equation in a cube, is again a perturbation of the standard backward difference implicit formula and is thus second-order correct in space and first-order correct in time.

In chapter II, we saw that a more accurate alternating direction method is that of Peaceman and Rachford [34]. Unfortunately, the straight-forward generalization of the P.R. method (2.2) to three space variables is not stable for large values of the mesh ratio, (see Douglas et al [13]). However, Douglas [11] has derived an unconditionally stable method which is a modification of the Crank-Nicolson difference formula and consequently is second-order correct in both space and time. The method is a direct extension of formulae (2.4).

In the first part of this chapter, an optimum finite difference formula for the solution of the diffusion equation in three space variables is derived employing techniques similar to those developed in chapter II. This formula is then rewritten as a set of alternating direction formulae of either Douglas-Rachford or Douglas type. The remainder of the chapter is devoted to the formulation of a generalized iterative procedure for the solution of Laplace's equation in three space variables.

III.2. The Douglas-Rachford and Douglas Methods.

Consider the diffusion equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \frac{\partial u}{\partial t} \tag{3.1}$$

for $u \equiv u(x,y,z,t)$ where x,y,z are space coordinates and t is the time, subject to the initial and boundary conditions $u(x,y,z,0) = f(x,y,z)$ over the cube $0 < x,y,z < 1$ and $u(x,y,z,t) = 0$, for all $t \geq 0$, where (x,y,z) is a point on the surface of the unit cube. When the boundary values are non-zero, the ADI methods formulated in this chapter must be modified in the same way as (2.11) in II.7.

The alternating direction method of Douglas and Rachford (D.R. method) [12] for solving this problem takes the form, [c.f. (2.5)],

$$\begin{aligned}
(-r\delta_x^2 + 1)u^*_{m+1} &= [r(\delta_y^2 + \delta_z^2) + 1]u_m \\
(-r\delta_y^2 + 1)u^{**}_{m+1} &= -r\delta_y^2 u_m + u^*_{m+1} \\
(-r\delta_z^2 + 1)u_{m+1} &= -r\delta_z^2 u_m + u^{**}_{m+1}
\end{aligned} \tag{3.2}$$

where $\delta_x, \delta_y, \delta_z$ are the usual central difference operators in the x,y,z directions respectively, u_m, u_{m+1} are the values of u at the nodes $(i\Delta x, j\Delta y, k\Delta z)$ at times $m\Delta t$ and $(m+1)\Delta t$ respectively, $(i,j,k = 1, 2, \ldots, N-1;\ m = 1, 2, \ldots)$, u^*_{m+1}, u^{**}_{m+1} denote approximations to u_{m+1}, and $r = \Delta t/h^2$ where $\Delta x = \Delta y = \Delta z = h$. Elimination of u^*_{m+1} and u^{**}_{m+1} leads to

$$\begin{aligned}
(1 - r\delta_x^2)(1 - r\delta_y^2)(1 - r\delta_z^2)u_{m+1} = \\
[1 + r^2(\delta_x^2\delta_y^2 + \delta_y^2\delta_z^2 + \delta_z^2\delta_x^2) + r^3\delta_x^2\delta_y^2\delta_z^2]u_m,
\end{aligned} \tag{3.3}$$

where $(1 - r\delta_x^2)(1 - r\delta_y^2)(1 - r\delta_z^2)$ and $[1 + r^2(\delta_x^2\delta_y^2 + \delta_y^2\delta_z^2 + \delta_z^2\delta_x^2) + r^3\delta_x^2\delta_y^2\delta_z^2]$

are both twenty-seven point operators. When this formula is written in the form

$$u_{m+1} = u_m + r(\delta_x^2 + \delta_y^2 + \delta_z^2)u_{m+1} - r^2(\delta_x^2\delta_y^2 + \delta_y^2\delta_z^2 + \delta_z^2\delta_x^2)(u_{m+1} - u_m) + r^3\delta_x^2\delta_y^2\delta_z^2(u_{m+1} - u_m),$$

it is seen to be a perturbation of the standard backward difference formula in three space variables.

The Douglas method [11] is given by [c.f. (2.4)]

$$\begin{aligned}
(-\tfrac{1}{3}r\delta_x^2 + 1)u^*_{m+1} &= [\tfrac{1}{3}r(\delta_x^2 + 2\delta_y^2 + 2\delta_z^2) + 1]u_m \\
(-\tfrac{1}{3}r\delta_y^2 + 1)u^{**}_{m+1} &= -\tfrac{1}{3}r\delta_y^2 u_m + u^*_{m+1} \qquad (3.4a) \\
(-\tfrac{1}{3}r\delta_z^2 + 1)u_{m+1} &= -\tfrac{1}{3}r\delta_z^2 u_m + u^{**}_{m+1}.
\end{aligned}$$

On elimination of u^*_{m+1}, u^{**}_{m+1}, formulae (3.4a) lead to

$$(1 - \tfrac{1}{3}r\delta_x^2)(1 - \tfrac{1}{3}r\delta_y^2)(1 - \tfrac{1}{3}r\delta_z^2)u_{m+1} = [1 + \tfrac{1}{3}r(\delta_x^2 + \delta_y^2 + \delta_z^2) + \tfrac{1}{4}r^2(\delta_x^2\delta_y^2 + \delta_y^2\delta_z^2 + \delta_z^2\delta_x^2) - \tfrac{1}{8}r^3\delta_x^2\delta_y^2\delta_z^2]u_m,$$

where $(1 - \tfrac{1}{3}r\delta_x^2)(1 - \tfrac{1}{3}r\delta_y^2)(1 - \tfrac{1}{3}r\delta_z^2)$

and $[1 + \tfrac{1}{3}r(\delta_x^2 + \delta_y^2 + \delta_z^2) + \tfrac{1}{4}r^2(\delta_x^2\delta_y^2 + \delta_y^2\delta_z^2 + \delta_z^2\delta_x^2) - \tfrac{1}{8}r^3\delta_x^2\delta_y^2\delta_z^2]$

are again twenty-seven point operators. This formula, which may easily be shown to be a perturbation of the Crank-Nicolson difference formula, in three space variables, was obtained independently by Brian [4] from the alternating direction formulae

$$\begin{aligned}
(-\tfrac{1}{3}r\delta_x^2 + 1)u^*_{m+1} &= [\tfrac{1}{3}r(\delta_y^2 + \delta_z^2) + 1]u_m \\
(-\tfrac{1}{3}r\delta_y^2 + 1)u^{**}_{m+1} &= -\tfrac{1}{3}r\delta_y^2 u_m + u^*_{m+1} \qquad (3.4b) \\
(-\tfrac{1}{3}r\delta_z^2 + 1)u_{m+1} &= -(\tfrac{1}{3}r\delta_z^2 + 1)u_m + 2u^{**}_{m+1}
\end{aligned}$$

which are based on the Douglas-Rachford formulae (3.2). The coefficient of u_{m+1} in the third equation is arbitrary and may be taken to be unity.

Just as (2.3) and (2.6) were special cases of a generalized formula involving nine points on each of two neighbouring levels of time so formulae (3.3) and (3.5) are special cases of a generalized formula involving twenty-seven point operators on each of two neighbouring levels of time. This generalized formula may be written in the form

$$(3.6)\quad [1 + A\sum\delta_x^2 + B\sum\delta_x^2\delta_y^2 + C\delta_x^2\delta_y^2\delta_z^2]u_{m+1} = [1 + D\sum\delta_x^2 + E\sum\delta_x^2\delta_y^2 + F\delta_x^2\delta_y^2\delta_z^2]u_m,$$

where $\sum\delta_x^2 = \delta_x^2 + \delta_y^2 + \delta_z^2$, $\sum\delta_x^2\delta_y^2 = \delta_x^2\delta_y^2 + \delta_y^2\delta_z^2 + \delta_z^2\delta_x^2$ and A, B, C, D, E and F are functions of r. For the D.R. method, the coefficients in (3.6) are

$$A = -r,\quad B = r^2,\quad C = -r^3,\quad D = 0,\quad E = r^2,\quad F = -r^3,$$

and for the Douglas method

$$A = -\tfrac{1}{2}r,\quad B = \tfrac{1}{4}r^2,\quad C = -\tfrac{1}{8}r^3,\quad D = \tfrac{1}{2}r,\quad E = \tfrac{1}{4}r^2,\quad F = -\tfrac{1}{8}r^3.$$

Also, by expanding (3.3) and (3.5) as Taylor series in terms of u_m and its derivatives and replacing derivatives with respect to t using the relations $\frac{\partial u}{\partial t} = \nabla^2 u$, $\frac{\partial^2 u}{\partial t^2} = \nabla^4 u$, . . etc., from (3.1) where $\nabla^2 \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$, we find that the principal parts of the truncation errors are

$$-[\tfrac{1}{12}rh^4(\tfrac{\partial^4}{\partial x^4} + \tfrac{\partial^4}{\partial y^4} + \tfrac{\partial^4}{\partial z^4}) + \tfrac{1}{6}r^2h^4\nabla^4] \text{ and } -\tfrac{1}{12}rh^4(\tfrac{\partial^4}{\partial x^4} + \tfrac{\partial^4}{\partial y^4} + \tfrac{\partial^4}{\partial z^4})$$

for the D.R. and Douglas methods respectively.

III.3. Generalized D.R. and Douglas Formulae.

(Fairweather and Mitchell [20])

Since it requires at each time step the solution of (N-1) linear simultaneous equations in (N-1) unknowns, where Nh = 1, formula (3.6) is a practicable method of solving (3.1) only if it can be written as a set of D.R. or Douglas type alternating direction formulae, that is, a set of formulae which utilises the same points as the D.R. or Douglas formulae. The tridiagonal systems of equations arising from the use of either set of formulae can then be solved directly by the algorithm outlined in Appendix IIA.

Provided

$$B = A^2, \quad C = A^3, \quad F = EA, \tag{3.7}$$

we can write (3.6) as a set of D.R. type formulae

$$\begin{aligned}
(1 + A\delta_x^2)u^*_{m+1} &= (D - \tfrac{E}{A})(\delta_y^2 + \delta_z^2)u_m + (1 - \tfrac{D}{A})\, u_m \\
(1 + A\delta_y^2)u^{**}_{m+1} &= \tfrac{E}{A}\, \delta_y^2 u_m + u^*_{m+1} \qquad (3.8) \\
(1 + A\delta_z^2)u_{m+1} &= (\tfrac{E}{A}\, \delta_z^2 + \tfrac{D}{A})u_m + u^{**}_{m+1}
\end{aligned}$$

or as the Douglas type formulae

$$\begin{aligned}
(1 + A\delta_x^2)u^*_{m+1} &= (D - \tfrac{E}{A})(\delta_y^2 + \delta_z^2)u_m + (1 + D\delta_x^2)u_m \\
(1 + A\delta_y^2)u^{**}_{m+1} &= \tfrac{E}{A}\, \delta_y^2 u_m + u^*_{m+1} \qquad (3.9) \\
(1 + A\delta_z^2)u_{m+1} &= \tfrac{E}{A}\, \delta_z^2 u_m + u^{**}_{m+1}
\end{aligned}$$

Formulae (3.8) and (3.9) are the generalized D.R. and Douglas type formulae respectively. Formula (3.6) cannot be represented by (3.8) or

(3.9) unless the conditions (3.7) are satisfied.

III.4. The Optimum Formula.

The optimum formula of type (3.6) is obtained in the same manner as that in II.4 . We expand the terms u_{m+1}, $\sum\delta_x^2 u_{m+1}$, etc., in (3.6) as Taylor series in terms of u_m and its derivatives, replacing the time derivatives by using $\frac{\partial u}{\partial t} = \nabla^2 u$, $\frac{\partial^2 u}{\partial t^2} = \nabla^4 u$, etc., from (3.1), then determine the values of A, B, C, D, E and F which satisfy (3.7) and also eliminate terms up to and including order h^4 in this expansion.

The expansions up to and including terms involving h^6 are

$$u_m = u$$

$$u_{m+1} = u + rA_1 + \tfrac{1}{2}r^2A_2 + r^2A_3 + \tfrac{1}{6}r^3A_4 + \tfrac{1}{2}r^3A_5 + r^3A_6$$

$$\sum\delta_x^2 u_{m+1} = A_1 + (r + \tfrac{1}{12})A_2 + 2rA_3 + (\tfrac{1}{2}r^2 + \tfrac{1}{12}r + \tfrac{1}{360})A_4 + (\tfrac{3}{2}r^2 + \tfrac{1}{12}r)A_5 + 3r^2A_6$$

$$\sum\delta_x^2\delta_y^2 u_{m+1} = A_3 + (r + \tfrac{1}{12})A_5 + 3rA_6$$

$$\delta_x^2\delta_y^2\delta_z^2 u_{m+1} = A_6$$

$$\sum\delta_x^2 u_m = A_1 + \tfrac{1}{12}A_2 + \tfrac{1}{360}A_4$$

$$\sum\delta_x^2\delta_y^2 u_m = A_3 + \tfrac{1}{12}A_5$$

$$\delta_x^2\delta_y^2\delta_z^2 u_m = A_6$$

which are the expansions in II.4 with additional A_6 terms, where, now,

$$A_1 = h^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right), \qquad A_2 = h^4\left(\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4} + \frac{\partial^4 u}{\partial z^4}\right),$$

$$A_3 = h^4\left(\frac{\partial^4 u}{\partial x^2\partial y^2} + \frac{\partial^4 u}{\partial y^2\partial z^2} + \frac{\partial^4 u}{\partial z^2\partial x^2}\right), \qquad A_4 = h^6\left(\frac{\partial^6 u}{\partial x^6} + \frac{\partial^6 u}{\partial y^6} + \frac{\partial^6 u}{\partial z^6}\right),$$

$$A_5 = h^6\left[\frac{\partial^4}{\partial x^2\partial y^2}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) + \frac{\partial^4}{\partial y^2\partial z^2}\left(\frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\right) + \frac{\partial^4}{\partial z^2\partial x^2}\left(\frac{\partial^2 u}{\partial z^2} + \frac{\partial^2 u}{\partial x^2}\right)\right],$$

$$A_6 = h^6 \frac{\partial^6 u}{\partial x^2 \partial y^2 \partial z^2}.$$

When these expressions have been substituted into formula (3.6), values of the coefficients A, B, C, D, E and F are determined which eliminate A_1, A_2, A_3, and A_4 and satisfy the conditions (3.7) which enable one to write the resulting formula as a set of D.R. or Douglas type formulae.

The required coefficients are

$$A = -\tfrac{1}{3}(r - \tfrac{1}{6}), \quad B = \tfrac{1}{4}(r - \tfrac{1}{6})^2, \quad C = -\tfrac{1}{8}(r - \tfrac{1}{6})^3,$$
$$D = \tfrac{1}{3}(r + \tfrac{1}{6}), \quad E = \tfrac{1}{4}(r + \tfrac{1}{6})^2, \quad F = -\tfrac{1}{8}(r + \tfrac{1}{6})^2(r - \tfrac{1}{6}).$$

The optimum formula of type (3.6) which can be written in the form (3.8) or (3.9) is thus

$$(3.10)\quad [1 - \tfrac{1}{3}(r - \tfrac{1}{6})\textstyle\sum \delta_x^2 + \tfrac{1}{4}(r - \tfrac{1}{6})^2 \sum \delta_x^2\delta_y^2 - \tfrac{1}{8}(r - \tfrac{1}{6})^3 \delta_x^2\delta_y^2\delta_z^2]u_{m+1}$$
$$= [1 + \tfrac{1}{3}(r + \tfrac{1}{6})\textstyle\sum \delta_x^2 + \tfrac{1}{4}(r + \tfrac{1}{6})^2 \sum \delta_x^2\delta_y^2 - \tfrac{1}{8}(r + \tfrac{1}{6})^2(r - \tfrac{1}{6})\delta_x^2\delta_y^2\delta_z^2]u_m,$$

and the principal part of the truncation error is

$$-\tfrac{1}{12}r(r^2 - \tfrac{1}{20})A_4 + \tfrac{1}{4}r(r + \tfrac{1}{6})^2 A_6.$$

which may be shown to be $O(h^4) + O(\Delta t^2)$. When this formula is written in the form

$$u_{m+1} = u_m + \tfrac{1}{3}\textstyle\sum \delta_x^2[(r - \tfrac{1}{6})u_{m+1} + (r + \tfrac{1}{6})u_m] - \sum\limits_{x\,y} \delta_x^2\delta_y^2[(r - \tfrac{1}{6})^2 u_{m+1}$$
$$- (r + \tfrac{1}{6})^2 u_m] + \tfrac{1}{8}(r - \tfrac{1}{6})\delta_x^2\delta_y^2\delta_z^2[(r - \tfrac{1}{6})^2 u_{m+1} - (r + \tfrac{1}{6})^2 u_m]$$

it is seen to be a perturbation of a three-dimensional analogue of the formula (2.15).

Formula (3.10) can now be written as the D.R. type formulae

$$[1 - \tfrac{1}{12}(r - \tfrac{1}{6})\delta_x^2]u^*_{m+1} = \frac{r(r+\frac{1}{6})}{(r-\frac{1}{6})}(\delta_y^2 + \delta_z^2)u_m + \frac{2r}{(r-\frac{1}{6})}u_m$$

$$(3.11)\qquad [1 - \tfrac{1}{12}(r - \tfrac{1}{6})\delta_y^2]u^{**}_{m+1} = -\frac{(r+\frac{1}{6})}{2(r-\frac{1}{6})}\delta_y^2 u_m + u^*_{m+1}$$

$$[1 - \tfrac{1}{12}(r - \tfrac{1}{6})\delta_z^2]u_{m+1} = -\frac{(r+\frac{1}{6})}{(r-\frac{1}{6})}[1 + \tfrac{1}{12}(r + \tfrac{1}{6})\delta_z^2]u_m + u^{**}_{m+1}$$

or as the Douglas formulae

$$[1 - \tfrac{1}{12}(r - \tfrac{1}{6})\delta_x^2]u^*_{m+1} = \frac{r(r+\frac{1}{6})}{(r-\frac{1}{6})}(\delta_y^2 + \delta_z^2)u_m + [1 + \tfrac{1}{12}(r + \tfrac{1}{6})\delta_x^2]u_m$$

$$(3.12)\qquad [1 - \tfrac{1}{12}(r - \tfrac{1}{6})\delta_y^2]u^{**}_{m+1} = -\frac{(r+\frac{1}{6})}{2(r-\frac{1}{6})}\delta_y^2 u_m + u^*_{m+1}$$

$$[1 - \tfrac{1}{12}(r - \tfrac{1}{6})\delta_z^2]u_{m+1} = -\frac{(r+\frac{1}{6})}{2(r-\frac{1}{6})}\delta_z^2 u_m + u^{**}_{m+1}$$

provided $r \neq \frac{1}{6}$.

When $r = \frac{1}{6}$, equation (3.6), like (2.10), becomes explicit, and the method of alternating directions is, of course, no longer applicable.

III.5. Stability of The Optimum Formula.

The stability of formula (3.10) is analyzed by the usual procedure of assuming that there exists an error $\varepsilon_{i,j,k,m}$ at each mesh point $i\Delta x, j\Delta y, k\Delta z, m\Delta t$ ($i,j,k = 1, 2, \ldots, N-1$; $m = 0, 1, 2, \ldots$). If the error is now expanded in the form

$$\varepsilon_{i,j,k,m} = a_m \sin \pi p x_i \sin \pi q y_j \sin \pi s z_k \qquad (p,q,s = 1, 2, \ldots, N-1)$$

where $x_i = i\Delta x$, $y_j = j\Delta y$, $z_k = k\Delta z$, and substituted into equation (3.10) with u replaced by ε and $\Delta x = \Delta y = \Delta z = 1/N$, we find that

$$\frac{a_{m+1}}{a_m} = \frac{1 - (r+\frac{1}{6})(X+Y+Z) + (r+\frac{1}{6})^2(XY+YZ+ZX) + (r-\frac{1}{6})(r+\frac{1}{6})^2 XYZ}{[1 + (r-\frac{1}{6})X][1 + (r-\frac{1}{6})Y][1 + (r-\frac{1}{6})Z]} , \tag{3.13}$$

where $X = 2\sin^2 \pi p/2N$, $Y = 2\sin^2 \pi q/2N$ and $Z = 2\sin^2 \pi s/2N$, ($p,q,s = 1, 2, \ldots, N-1$).

For stability, we require that $-1 \leq \frac{a_{m+1}}{a_m} \leq 1$ for all p,q,s in the range $1 \leq p,q,s \leq N-1$. It may be shown after some manipulation that this condition is satisfied provided $r \leq \frac{11}{12}$. While an unconditionally stable formula is most desirable, a high accuracy formula with conditional stability also has its advantages. It can be employed to make calculations for small t, where a cruder mesh can be used to obtain the same accuracy as the D.R. or Douglas method. The method can also provide very accurate starting values for a multi-level scheme.

III.6. Generalized Iterative Procedure for Laplace's Equation.

The remainder of this chapter is devoted to the derivation of a family of iteration techniques for solving Laplace's equation in three space variables. This family of techniques contains the Douglas method and the optimum formula derived in III.4 as special cases. Use is once more made of the fact that, provided $u(x,y,z,t) = u(x,y,z,0)$ for all $t > 0$, where x,y,z is a point on the surface of the unit cube, the steady-state solution of equation (3.1) is the solution of Laplace's equation,

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0, \tag{3.14}$$

subject to the boundary condition $u(x,y,z) = f(x,y,z)$ for x,y,z a point on the surface of the unit cube.

Let us return to difference formulae of type (3.6) where r is now considered to be an iteration parameter and boundary values are taken to be zero. Modifications must again be made when the boundary values are non-zero. As shown previously, formulae of type (3.6) can be written in D.R. or Douglas form if $B = A^2$, $C = A^3$ and $F = EA$. If, in addition, the conditions

$$\begin{aligned} A - D &= -r \\ 2rA + B - E &= -r^2 \end{aligned} \tag{3.15}$$

are satisfied, then these formulae represent (3.1) and are second order correct in both space and time. From these relations between the coefficients, we obtain the values

$$A = A, \quad B = A^2, \quad C = A^3, \quad D = (r + A), \quad E = (r + A)^2, \quad F = A(r + A)^2.$$

If these values of the coefficients are substituted into (3.6), we obtain the iterative formula

$$(3.16)\quad (1 + A\delta_x^2)(1 + A\delta_y^2)(1 + A\delta_z^2)u_{m+1}$$
$$= [1 + (r + A)\sum\delta_x^2 + (r + A)^2 \sum\delta_x^2\delta_y^2 + A(r + A)^2\delta_x^2\delta_y^2\delta_z^2]u_m$$

for solving (3.14). If, as in II.6, we put $A = -\frac{1}{3}(r + f)$, where f is a parameter, formula (3.16) becomes

$$(3.17)\quad [1 - \tfrac{1}{3}(r + f)\delta_x^2][1 - \tfrac{1}{3}(r + f)\delta_y^2][1 - \tfrac{1}{3}(r + f)\delta_z^2]u_{m+1}$$
$$= [1 + \tfrac{1}{3}(r - f)\sum\delta_x^2 + \tfrac{1}{4}(r - f)^2\sum\delta_y^2\delta_z^2 - \tfrac{1}{8}(r + f)(r - f)^2\delta_x^2\delta_y^2\delta_z^2]u_m ,$$

which can be written as the Douglas formulae

$$[1 - \tfrac{1}{3}(r + f)\delta_x^2]u^*_{m+1} = \frac{r(r - f)}{(r + f)}[\delta_y^2 + \delta_z^2]u_m + [1 + \tfrac{1}{3}(r - f)\delta_x^2]u_m$$

$$(3.18a)\quad [1 - \tfrac{1}{3}(r + f)\delta_y^2]u^{**}_{m+1} = -\frac{(r - f)^2}{2(r + f)}\delta_y^2 u_m + u^*_{m+1}$$

$$[1 - \tfrac{1}{3}(r + f)\delta_z^2]u_{m+1} = -\frac{(r - f)^2}{2(r + f)}\delta_z^2 u_m + u^{**}_{m+1}$$

or as the Douglas-Rachford formulae

$$[1 - \tfrac{1}{3}(r + f)\delta_x^2]u^*_{m+1} = \frac{r(r - f)}{(r + f)}[\delta_y^2 + \delta_z^2]u_m + \frac{2r}{(r + f)}u_m$$

$$(3.18b)\quad [1 - \tfrac{1}{3}(r + f)\delta_y^2]u^{**}_{m+1} = -\frac{(r - f)^2}{2(r + f)}\delta_y^2 u_m + u^*_{m+1}$$

$$[1 - \tfrac{1}{3}(r + f)\delta_z^2]u_{m+1} = -\frac{(r - f)}{(r + f)}[1 + \tfrac{1}{3}(r - f)\delta_z^2]u_m + u^{**}_{m+1}$$

When f takes the value 0 in (3.17), (3.18a) and (3.18b), we obtain the methods of Douglas [11] and Brian [4] respectively, and when $f = -\frac{1}{6}$, we obtain the optimum formula derived in III.4. We now determine the value (or values) of $f \geqslant -\frac{1}{6}$ which will make (3.18) the best iteration scheme for solving (3.14).

III.7. Optimum Convergence Factor.

In the manner of II.9, the convergence factor of scheme (3.17) is obtained from

$$(3.19) \qquad \lambda \equiv \frac{a_{m+1}}{a_m}$$

$$= \frac{1 - (r-f)(X + Y + Z) + (r-f)^2(XY + YZ + ZX) + (r+f)(r-f)^2XYZ}{[1 + (r+f)X][1 + (r+f)Y][1 + (r+f)Z]}$$

where $X = 2\sin^2 \pi p/2N$, $Y = 2\sin^2 \pi q/2N$ and $Z = 2\sin^2 \pi s/2N$, $(p,q,s = 1, 2, \ldots, N-1)$ by determining its maximum modulus value as a function of X,Y,Z in the cube $\mu \leq X,Y,Z \leq \nu$, where $\mu = 2\sin^2 \pi/2N$ and $\nu = 2\cos^2 \pi/2N$. In order to obtain the optimum convergence factor, we minimise the resulting function with respect to r (> 0) for each f $(\geq -\frac{1}{6})$, noting that $|\lambda| < 1$ for all r when $f \geq 0$ and for

$$r < -\frac{8f^2 + 12f + 3}{8f}$$

when $-\frac{1}{6} \leq f < 0$. In order to carry out this minimax procedure, we consider two cases.

Case 1: $0 < r \leq f$.

In this case, $\lambda > 0$ and $\frac{\partial\lambda}{\partial X}, \frac{\partial\lambda}{\partial Y}, \frac{\partial\lambda}{\partial Z} < 0$ for $\mu \leq X,Y,Z \leq \nu$. It is easily shown that, since λ is a positive decreasing function of X,Y,Z,

$$\hat{\lambda}(\mu,\nu) \equiv \max_{\mu \leq X,Y,Z \leq \nu} |\lambda(X,Y,Z)| = \lambda(\mu,\mu,\mu)$$

where

$$\lambda(\mu,\mu,\mu) = \frac{1 + 3(f-r)\mu + 3(f-r)^2\mu^2 + (f+r)(f-r)^2\mu^3}{[1 + (f+r)\mu]^3} .$$

This function is now minimised with respect to r (> 0) for a given value of f and the result

(3.20) $$\min_{0<r\le f} \hat{\lambda}(\mu,\nu) = \frac{1}{[1+2f\mu]^3} = \frac{1}{[1+4f\sin^2\pi/2N]^3}$$

is obtained. The value of r necessary to give this optimum convergence is $r^* = f$, since $\hat{\lambda}(\mu,\nu)$ is a positive decreasing function of r in the range $0 < r \le f$.

For N = 10, 30 and 100, the convergence factors for various values of f are shown in Table 3.1 .

N \ f	1	2	10
10	0.75566	0.58486	0.12905
30	0.96784	0.93704	0.73206
100	0.99705	0.99410	0.97097

Table 3.1 .

It appears that the best convergence factor for a given value of N is obtained when f is positive and as large as possible.

Case 2: $r > f$.

In order to facilitate the examination of this case, we make the transformation

$$X' = (r+f)X, \quad Y' = (r+f)Y, \quad Z' = (r+f)Z.$$

Formula (3.19) then takes the form

(3.21) $$\lambda = \frac{1 - R(X'+Y'+Z') + R^2(X'Y' + Y'Z' + Z'X' + X'Y'Z')}{[1+X'][1+Y'][1+Z']}$$

where $R = \frac{r-f}{r+f}$ and $0 < a \le X', Y', Z' \le b$ where $a = (r+f)\mu$ and $b = (r+f)\nu$. In the following analysis, we make the assumption that $r+f > 0$. This is not a severe restriction but merely excludes r in the range $0 < r \le \frac{1}{2}$

when $-\frac{1}{6} \leq f \leq 0$. Case 2 has now become $R > 0$.

Returning to (3.21), we find that

$$\frac{\partial \lambda}{\partial X'} = \frac{R(R+1)[(Y'+Z') - 1/R]}{[1+X']^2[1+Y'][1+Z']} . \tag{3.22}$$

Similar results hold by symmetry for $\frac{\partial \lambda}{\partial Y'}$ and $\frac{\partial \lambda}{\partial Z'}$.

When $1/R$ is in the range $0 < 1/R \leq 2a$, $\frac{\partial \lambda}{\partial X'}$, $\frac{\partial \lambda}{\partial Y'}$, $\frac{\partial \lambda}{\partial Z'}$ are positive, and it is easily seen that

$$\max_{a \leq X',Y',Z' \leq b} \lambda(X',Y',Z') = \lambda(b,b,b),$$

$$\min_{a \leq X',Y',Z' \leq b} \lambda(X',Y',Z') = \lambda(a,a,a).$$

Now $$\lambda(a,a,a) = \frac{1 - 3aR + (a^3 + 3a^2)R^2}{(1+a)^3}$$

$$= \frac{(1 - \frac{3}{2}aR)^2 + (a^3 + \frac{3}{4}a^2)R^2}{(1+a)^3}$$

$$> 0.$$

Hence, $$\max_{a \leq X',Y',Z' \leq b} |\lambda(X',Y',Z')| = \max_{a \leq X',Y',Z' \leq b} \lambda(X',Y',Z')$$

$$= \lambda(b,b,b), \qquad 0 < 1/R \leq 2a.$$

By a similar argument, it may be shown that for $1/R$ in the range $2b \leq 1/R < \infty$,

$$\max_{a \leq X',Y',Z' \leq b} |\lambda(X',Y',Z')| = \lambda(a,a,a).$$

When $1/R$ is in the range $2a \leq 1/R \leq 2b$, λ has a stationary point when

$$X' = Y' = Z' = 1/2R.$$

However, this point is neither a maximum nor a minimum; so the maximum modulus value of λ occurs on the surface of the cube $a \leq X',Y',Z' \leq b$ which lies wholly in the positive octant. A systematic search, in fact,

reveals that the maximum modulus λ for R in the range $2a \leq 1/R \leq 2b$ is located at one of the corners of the cube, and, by symmetry, it is sufficient to consider the four corners (a,a,a), (a,a,b), (a,b,b) and (b,b,b). The values of λ at these four points are respectively

$$\lambda(a,a,a) = \frac{1 - 3aR + a^2(3 + a)R^2}{(1 + a)^3},$$

$$\lambda(a,a,b) = \frac{1 - (2a + b)R + a(a + 2b + ab)R^2}{(1 + a)^2(1 + b)},$$

$$\lambda(a,b,b) = \frac{1 - (a + 2b)R + b(2a + b + ab)R^2}{(1 + a)(1 + b)^2},$$

$$\lambda(b,b,b) = \frac{1 - 3bR + b^2(3 + b)R^2}{(1 + b)^3},$$

For R in the stated range, $\lambda(a,a,a)$, $\lambda(a,b,b)$ and $\lambda(b,b,b)$ are positive and since $\lambda(a,b,b)$ is less than either $\lambda(a,a,a)$ or $\lambda(b,b,b)$, for every permissible R, $\lambda(a,b,b)$ can be discarded in the search for the maximum. Finally, $\lambda(a,a,b)$ can either be positive or negative depending on R. When it is positive, is is less than either $\lambda(a,a,a)$ or $\lambda(b,b,b)$. Hence, for $R > 0$,

$$\max_{a \leq X', Y', Z' \leq b} |\lambda(X',Y',Z')| = \max[\lambda(a,a,a), \lambda(b,b,b), -\lambda(a,a,b)].$$

Thus, for the case $r > |f|$, that is, $r > f$ together with $r+f > 0$, the convergence factor is given by

$$\hat{\lambda}(\mu,\nu) \equiv \max_{a \leq X', Y', Z' \leq b} |\lambda(X',Y',Z')| = \max[\lambda(\mu,\mu,\mu), \lambda(\nu,\nu,\nu), -\lambda(\mu,\mu,\nu)].$$

In order to obtain the optimum convergence factor for a given value of f, we must now minimise $\hat{\lambda}(\mu,\nu)$ as a function of r. The value of r which minimises $\hat{\lambda}(\mu,\nu)$ is taken to be r^* given by

$$\lambda(\mu,\mu,\mu) = \lambda(\nu,\nu,\nu).$$

After some manipulation, it may be shown that r* satisfies the equation

$$6r^3\sin^2\pi/N + 3(3 + 2f)r^2\sin^2\pi/N - 2f(4 - \sin^2\pi/N + 3f\sin^2\pi/N)r$$
$$- [3 + 12f + (8 + 7\sin^2\pi/N)f^2 + 6f^3\sin^2\pi/N] = 0.$$

Values of r* were calculated for N = 10, 30 and 100 and for various $f \geqslant -\frac{1}{6}$, and are shown in Table 3.2(b). The corresponding values of $\lambda(\mu,\mu,\mu)$ $[\equiv \lambda(\nu,\nu,\nu)]$ and $-\lambda(\mu,\mu,\nu)$ are shown in Table 3.2(a) and Table 3.3 respectively. On comparing Tables 3.2(a) and 3.3 , it is seen that for each N and f and appropriate r*, $\lambda(\mu,\mu,\mu) > -\lambda(\mu,\mu,\nu)$. Thus, certainly for the pairs of values of N and f shown in the Tables

$$\min_{r>|f|} \hat{\lambda}(\mu,\nu) = \lambda(\mu,\mu,\mu) \quad [\equiv \lambda(\nu,\nu,\nu)].$$

Table 3.2(a) gives the optimum convergence factors for the values of N and f quoted and Table 3.2(b) gives the appropriate values of r*. From Table 3.2(a), it appears once again that the best convergence factor for a given value of N is obtained when f is positive and as large as possible. It can also be seen from this table that the original Douglas formula (f = 0) has a better convergence factor than the optimum Douglas type formula ($f = -\frac{1}{6}$) derived previously to solve (3.1). Comparing Tables 3.1 and 3.2(a), we see that, for given N and f (> 0), the best convergence factor is obtained by choosing r* ($> |f|$) as given by Table 3.2(b).

The substantial improvement in convergence of the alternating direction method which arises from the choice of large positive values of the parameter f may, again, be accompanied by a loss of accuracy.

N \ f	$-\frac{1}{6}$	0	1	2	10
10	0.83141	0.67193	0.33189	0.22013	0.04470
30	0.97258	0.90193	0.69278	0.59817	0.32241
100	0.99731	0.97803	0.89644	0.85717	0.70910

Table 3.2(a) .

N \ f	$-\frac{1}{6}$	0	1	2	10
10	0.62375	1.35440	3.93958	5.65159	15.38427
30	0.84502	3.14033	11.22797	15.80561	36.30478
100	0.90890	7.50273	36.94483	52.10808	116.68160

Table 3.2(b) .

N \ f	$-\frac{1}{6}$	0	1	2	10
10	0.25153	0.34016	0.19220	0.11386	0.00388
30	0.42458	0.67596	0.59321	0.50730	0.23646
100	0.46236	0.86213	0.85656	0.81860	0.65922

Table 3.3 .

The iteration procedure described by (3.17) converges if $u_{m+1} = u_m = u$ for m sufficiently large, and so (3.17) reduces to

$$(3.23)\quad [(\delta_x^2 + \delta_y^2 + \delta_z^2) - f(\delta_x^2\delta_y^2 + \delta_y^2\delta_z^2 + \delta_z^2\delta_x^2) + \frac{f(r+f)}{2}\delta_x^2\delta_y^2\delta_z^2]u = 0.$$

For sufficiently large f, this equation degenerates into

$$(\delta_x^2\delta_y^2\delta_z^2)u = 0,$$

which is no longer a difference approximation to Laplace's equation. Consequently, a balance is required between the rate of convergence (an optimum when f is infinite) and the accuracy of the process (an optimum when $f = -\frac{1}{6}$).

III.8. Sequences of Iteration Parameters.

The above analysis is dependent on constant r throughout the iteration procedure. No attempt will be made in this thesis to obtain a sequence of iteration parameters. Because of the form of the convergence rate, this would be an exceedingly difficult task. Douglas [11] has obtained a "reasonably good" parameter sequence for his method, but, in order to simplify the analysis, he relaxes the bounds on λ.

CHAPTER IV.

A GENERALIZED ADI METHOD FOR THE BIHARMONIC EQUATION.

IV.1. Introduction.

In the problem of determining the elastic buckling loads of flat plates under partial edge compression, the plate stress u is given by the biharmonic equation

$$\nabla^4 u \equiv \frac{\partial^4 u}{\partial x^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4} = 0. \tag{4.1}$$

It is convenient to consider a square homogeneous plate ($0 \leq x \leq 1$, $0 \leq y \leq 1$) supported along its edges and buckled by moments along two opposite plate edges. This leads to the boundary value problem consisting of (4.1) together with the boundary conditions

$$\begin{aligned}
&\text{(a)} \quad u = \frac{\partial^2 u}{\partial x^2} = 0 && \text{at } x = 0,\ 1 && (0 \leq y \leq 1)\\
&\text{(b)} \quad u = 0 && \text{at } y = 0,\ 1 && (0 \leq x \leq 1)\\
&\text{(c)} \quad \frac{\partial^2 u}{\partial y^2} = f_1(x) && \text{at } y = 0 && (0 \leq x \leq 1)\\
&\text{(d)} \quad \frac{\partial^2 u}{\partial y^2} = f_2(x) && \text{at } y = 1 && (0 \leq x \leq 1)
\end{aligned} \tag{4.2i}$$

Several attempts have been made to solve this boundary value problem using finite differences. Originally, relaxation methods were used to be followed by more sophisticated techniques such as an ADI method by Conte and Dames [7], a "two-line" iterative method by Parter [32] and over-relaxation by White [49] and Tee [40].

In this chapter, we shall consider ADI methods of the Douglas-Rachford type as a means of solving the biharmonic equation, and obtain a generalized version of the method used by Conte and Dames.

IV.2. The Conte-Dames (C.D.) Method.

In the usual way, a square mesh is superimposed over the region $0 \leq x \leq 1$, $0 \leq y \leq 1$, with the mesh size $h = 1/N$, where $N(>3)$ is an integer, and the boundary conditions (4.2(i)) become

$$
\begin{aligned}
&\text{(a)} \quad u_{i,j} = 0 \\
&\qquad\quad u_{i+1,j} = -u_{i-1,j} \qquad (i = 0,N) \quad (j = 1, 2, \ldots, N-1), \\
&\text{(b)} \quad u_{i,0} = u_{i,N} = 0 \qquad\qquad\qquad\quad (i = 1, 2, \ldots, N-1), \\
&\text{(c)} \quad u_{i,1} = -u_{i,-1} + h^2 f(ih) \qquad\quad (i = 1, 2, \ldots, N-1), \\
&\text{(d)} \quad u_{i,N+1} = -u_{i,N-1} + h^2 g(ih) \qquad (i = 1, 2, \ldots, N-1),
\end{aligned}
\tag{4.2(ii)}
$$

where $u_{i,j}$ denotes the value of u at the node (ih, jh).

The Conte-Dames (C.D.) method for solving (4.1) is a double-sweep iterative process given by

$$
\begin{aligned}
u_{i,j}^{(n+\frac{1}{2})} &= u_{i,j}^{(n)} - r(\delta_y^4 u_{i,j}^{(n+\frac{1}{2})} + 2\delta_x^2\delta_y^2 u_{i,j}^{(n)} + \delta_x^4 u_{i,j}^{(n)}) \\
u_{i,j}^{(n+1)} &= u_{i,j}^{(n+\frac{1}{2})} - r(\delta_x^4 u_{i,j}^{(n+1)} - \delta_x^4 u_{i,j}^{(n)}), \qquad (n \geqslant 0)
\end{aligned}
\tag{4.3}
$$

where $u_{i,j}^{(n)}$ is the n-th approximation to u at the node (ih, jh), $(i,j = 1, 2, \ldots, N-1)$, δ_x, δ_y denote the usual central difference operators in the x and y directions respectively and r is an iteration parameter chosen to accelerate the covergence of the iterative procedure. If $u_{i,j}^{(n+\frac{1}{2})}$ is eliminated, equations (4.3) become

$$
\begin{aligned}
u^{(n+1)} = u^{(n)} &- r(\delta_x^4 u^{(n+1)} + 2\delta_x^2\delta_y^2 u^{(n)} + \delta_y^4 u^{(n+1)}) \\
&- r^2\delta_x^4\delta_y^4(u^{(n+1)} - u^{(n)}),
\end{aligned}
\tag{4.4}
$$

where the lower suffices i, j have been omitted for convenience.

The C.D. formulae (4.3) constitute a convergent iterative method for solving the biharmonic equation which is similar to the Douglas-

Rachford ADI method for solving Laplace's equation.

IV.3. The Generalized C.D. Formulae. (Fairweather and Mitchell [19])

Consider the generalized form of (4.3)

$$(a_o + a_1\delta_y^4)u^{(n+\frac{1}{2})} = (b_o + b_1\delta_x^4 + b_2\delta_y^4 + b_3\delta_x^2\delta_y^2)u^{(n)}$$

$$u^{(n+\frac{1}{2})} = (c_o + c_1\delta_x^4)u^{(n+1)} - (d_o + d_1\delta_x^4)u^{(n)}. \tag{4.5}$$

This reduces to the C.D. formulae (4.3) if the coefficients take the values

$$a_o = b_o = c_o = 1, \quad a_1 = -b_1 = -\tfrac{1}{6}b_3 = c_1 = d_1 = r, \quad b_2 = d_o = 0.$$

If $u^{(n+\frac{1}{2})}$ is eliminated in (4.5), we obtain the formula

$$(a_o + a_1\delta_y^4)(c_o + c_1\delta_x^4)u^{(n+1)}$$
$$= [(a_o + a_1\delta_y^4)(d_o + d_1\delta_x^4) + (b_o + b_1\delta_x^4 + b_2\delta_y^4 + b_3\delta_x^2\delta_y^2)]u^{(n)}. \tag{4.6}$$

The coefficients in (4.6) are now adjusted so that if the process converges, that is, if $u^{(n+1)} = u^{(n)} = u$ for n sufficiently large, then (4.6) reduces to

$$[\alpha_o(\delta_x^4 + 2\delta_x^2\delta_y^2 + \delta_y^4) + \alpha_1\delta_x^4\delta_y^4]u = 0, \tag{4.7}$$

which is a fourth-order finite difference replacement of the biharmonic equation for all values of the parameters α_o and α_1. This occurs when the coefficients in (4.5) and (4.6) satisfy the relations

$$b_o = a_o(c_o - d_o), \quad b_1 = \frac{a_o}{a_1}\alpha_1 - \alpha_o, \quad b_2 = a_1(c_o - d_o) - \alpha_o,$$
$$b_3 = -2\alpha_o, \qquad d_1 - c_1 = -\frac{1}{a_1}\alpha_1.$$

Using these relations between the coefficients, (4.6) becomes

$$(a_o + a_1\delta_y^4)(c_o + c_1\delta_x^4)u^{(n+1)}$$
$$= [(a_o + a_1\delta_y^4)(c_o + c_1\delta_x^4) - \{\alpha_o(\delta_x^4 + 2\delta_x^2\delta_y^2 + \delta_y^4) + \alpha_1\delta_x^4\delta_y^4\}]u^{(n)}$$

In order to simplify the analysis which follows and yet retain the essential character of the generalization, we take

(4.9) $$\alpha_0 = r, \qquad \alpha_1 = r^2(1 - \gamma),$$

together with

$$a_0 = c_0 = 1, \quad a_1 = c_1 = r, \quad d_0 = 0,$$

where r and γ are adjustable parameters. With these simplifications, (4.8) becomes

$$(1 + r\delta_y^4)(1 + r\delta_x^4)u^{(n+1)}$$
$$= [(1 + r\delta_y^4)(1 + r\delta_x^4) - \{r(\delta_x^4 + 2\delta_x^2\delta_y^2 + \delta_y^4) + (1 - \gamma)r^2\delta_x^4\delta_y^4\}]u^{(n)}$$

which reduces to

(4.10) $$(1 + r\delta_y^4)(1 + r\delta_x^4)u^{(n+1)} = (1 - 2r\delta_x^2\delta_y^2 + \gamma r^2\delta_x^4\delta_y^4)u^{(n)}.$$

Formulae (4.5) may then be written in the form

(4.11)
$$(1 + r\delta_y^4)u^{(n+\frac{1}{2})} = (1 - \gamma r\delta_x^4 - 2r\delta_x^2\delta_y^2)u^{(n)}$$
$$u^{(n+\frac{1}{2})} - (1 + r\delta_x^4)u^{(n+1)} = -\gamma r\,\delta_x^4 u^{(n)}.$$

Clearly the generalized method described by (4.10), or (4.11), converges to the solution of a difference approximation of the biharmonic equation given by

(4.12) $$[(\delta_x^4 + 2\delta_x^2\delta_y^2 + \delta_y^4) + (1 - \gamma)r\delta_x^4\delta_y^4]u = 0$$

which agrees with the standard difference form

$$(\delta_x^4 + 2\delta_x^2\delta_y^2 + \delta_y^4)u = 0,$$

except for a term involving eighth differences. This term will, in general, have a negligble effect on the accuracy whereas its inclusion can be used to accelerate the convergence of the ADI method. When $\gamma = 1$, the term disappears, and (4.10) and (4.11) become (4.4) and (4.3) respectively.

At every stage of the iteration scheme (4.11), the matrices of the coefficients of the linear systems involved are all of order N-1 with

at most five non-zero elements in any row (either on the main diagonal or two adjacent diagonals). Such "quidiagonal" systems of equations can be solved efficiently by an extension of the algorithm for solving tridiagonal systems described in Appendix IIA. This extension, which was formulated by Conte and Dames [7] is described in the Appendix at the end of this chapter.

IV.4. Convergence of the Iterative Procedure.

As in I.3, the error $\varepsilon_{i,j}^{(n)}$ is defined by

$$\varepsilon_{i,j}^{(n)} = u_{i,j}^{(n)} - u_{i,j},$$

where $u_{i,j}$ is the solution of (4.12). The error growth is governed by equation (4.10) with u replaced by ε, together with homogeneous boundary conditions. The error is expanded in the usual form

$$\varepsilon_{i,j}^{(n)} = a_n \sin \pi p x_i \sin \pi q y_j \qquad (p,q = 1, 2, \ldots, N-1)$$

where $x_i = ih$, $y_j = jh$, $(i,j = 1, 2, \ldots, N-1)$, and substituted into (4.9) with u replaced by ε. Simple manipulation yields

$$\lambda = \frac{a_{n+1}}{a_n} = \frac{R^2 - 2RS_p^2S_q^2 + \gamma S_p^4 S_q^4}{R^2 + R(S_p^4 + S_q^4) + S_p^4 S_q^4}, \tag{4.13}$$

where $S_p^2 = 4\sin^2 \pi p/2N$, $S_q^2 = 4\sin^2 \pi q/2N$ and $R = 1/r$.

To facilitate the examination of the convergence factor, we introduce $\bar{\lambda}$, where

$$\bar{\lambda} = \frac{R^2 - 2RS_p^2S_q^2 + \gamma S_p^4 S_q^4}{R + 2RS_p^2S_q^2 + S_p^4 S_q^4} \tag{4.14}$$

Since $$S_p^4 + S_q^4 \geqslant 2S_p^2S_q^2,$$

it follows that $|\bar{\lambda}| \geqslant |\lambda|$ for all p and q. If we put $z = S_p^2 S_q^2$ then (4.13) becomes

$$\bar{\lambda}(R,\gamma,z) = \frac{R^2 - 2Rz + \gamma z^2}{R^2 + 2Rz + z^2}. \tag{4.15}$$

In the analysis which follows, the convergence of the iterative procedure will always be based on $\bar{\lambda}$. Since $|\lambda| \leq |\bar{\lambda}|$, the actual convergence factor will generally be better than the figure quoted.

IV.5. Optimum Convergence Factor.

We now examine (4.15) with $R > 0$, $0 < z_1 \leq z \leq z_2$ where $z_1 = 16 \sin^4 \pi/2N$, $z_2 = 16 \cos^4 \pi/2N$, and $\gamma \leq 1$. A typical graph of $\bar{\lambda}$ against z, for given R and γ, is illustrated in Figure VI.

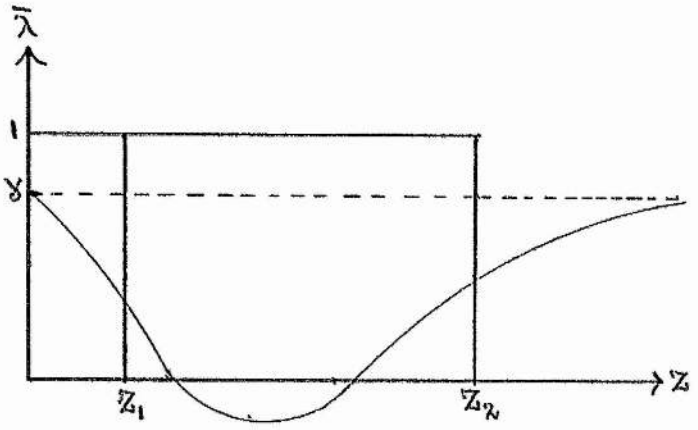


Fig. VI

This curve has a minimum at

$$z = \frac{2}{\gamma + 1} R,$$

where the value of $\bar{\lambda}$ is

(4.16c) $$- \frac{1 - \gamma}{3 + \gamma} \qquad (\equiv G_3)$$

and $\bar{\lambda}$ tends to γ as z tends to infinity. The values of $\bar{\lambda}$ at the limits of z are

(4.16a) $$\frac{R^2 - 2Rz_1 + \gamma z_1^2}{R^2 + 2Rz_1 + z_1^2} \qquad (\equiv G_1)$$

and

(4.16b) $$\frac{R^2 - 2Rz_2 + \gamma z_2^2}{R^2 + 2Rz_2 + z_2^2} \qquad (\equiv G_2)$$

respectively. Depending on the values of R and γ, the minimum of the curve may lie inside or outside the permissible range of values of

z. If the minimum lies outside, the maximum modulus value of $\bar{\lambda}$ is either $|G_1|$ or $|G_2|$. If the minimum lies inside, as shown in Figure VI, the maximum modulus value is $|G_1|$, $|G_2|$, or $-G_3$.

To simplify the determination of the convergence factor (the maximum modulus value of $\bar{\lambda}$) for given values of γ and R, we consider two cases. If γ_c is given by

$$(\gamma_c + 3)(1 - \gamma_c) = \frac{32 \cos^2 \pi/N}{(1 + \cos^2 \pi/N)^2}$$

then

Case I: $\gamma_c \leq \gamma \leq 1$.

It may easily be shown that the convergence factor in this case is

$$|G_2| \quad \text{if} \quad 0 < R \leq R_c$$
$$|G_1| \quad \text{if} \quad R_c \leq R$$

where

$$R_c = \tfrac{1}{8}\left[\left\{(1 - \gamma)^2(z_1 + z_2)^2 + 32(1 + \gamma)z_1 z_2\right\}^{\frac{1}{2}} - (1 - \gamma)(z_1 + z_2)\right].$$

Let us now consider

Case II: $\gamma \leq \gamma_c$

When the minimum lies within the permissible range of values of z, that is, when $z_1 \leq \frac{2}{\gamma + 1} R \leq z_2$, the convergence factor is given by

$$|G_2| \quad \text{if} \quad \frac{\gamma + 1}{2} z_1 \leq R \leq R_1$$
$$-G_3 \quad \text{if} \quad R_1 \leq R \leq R_2$$
$$|G_1| \quad \text{if} \quad R_2 \leq R \leq \frac{\gamma + 1}{2} z_2$$

where

$$R_1 = \frac{2 - (\gamma + 3)[\tfrac{1}{8}(1 - \gamma)]^{1/2}}{\gamma + 1} z_2$$

and

$$R_2 = \frac{2 + (\gamma + 3)[\tfrac{1}{3}(1 - \gamma)]^{\frac{1}{3}}}{\gamma + 1} z_1$$

provided $\bar{\gamma}_c \leq \gamma \leq \gamma_c$, where $\bar{\gamma}_c$ is given by

$$(\bar{\gamma}_c + 3)^2(1 - \bar{\gamma}_c) = 8.$$

We note that this is an extremely small range of values of γ for, as $M \to \infty$, $\gamma_c \to \bar{\gamma}_c$.

When $\gamma < \bar{\gamma}_c$, the convergence factor is

$$-G_3 \text{ if } \frac{\gamma + 1}{2} z_1 \leq R \leq R_2$$

$$|G_1| \text{ if } R_2 \leq R \leq \frac{\gamma + 1}{2} z_2.$$

If the minimum lies outside the permissible range of values of z, then either

$$R > \frac{\gamma + 1}{2} z_2,$$

when the convergence factor is $|G_1|$ for all $\gamma \leq \gamma_c$; or

$$0 \leq R < \frac{\gamma + 1}{2} z_1,$$

when the convergence factor is

$$|G_2| \text{ if } \bar{\gamma}_c \leq \gamma \leq \gamma_c$$

or

$$\left.\begin{aligned} &|G_2| \text{ if } 0 \leq R \leq \bar{R}_c \\ &|G_1| \text{ if } \bar{R}_c \leq R \leq \frac{\gamma + 1}{2} z_1 \end{aligned}\right\} \quad \text{if } \gamma \leq \gamma_c$$

where $\bar{R}_c$ satisfies the polynomial in R given by

$$G_2 = -G_1.$$

The determination of $\bar{R}_c$ is not pursued, since it is of little practical importance.

In order to obtain the optimum convergence factor for a given value of N, it is necessary to find the values of γ and R which minimise the maximum value of $\bar{\lambda}$. Clearly the minimum occurs where $\gamma = \gamma_c$ and $R = R_1 = R_2 = R_c$. This leads to an optimum convergence factor of

$$\frac{1 - \gamma_c}{3 + \gamma_c}.$$

Thus, for any value of N, and γ in the range $0 < \gamma_c \leq \gamma \leq 1$, the convergence factor is worst when $\gamma = 1$ and best when $\gamma = \gamma_c$.

The case N=10 is illustrated in Table 4.1. As γ decreases from 1 (the C.D. value) to 0.23969 (γ_c correct to five decimal places), the convergence factor improves from 0.90451 to 0.23469, both values correct to five places. It should be noted, however, that a decrease in γ leads to an increase in $(1-\gamma)$, a decrease in R and hence an increase in r and so the eighth-order difference term in (4.12), $(1 - \gamma)rb_x^4 b_y^4$, becomes relatively more important. This will have little effect on the accuracy of the limiting solution of the iterative process, except possibly in circumstances when N is large and γ has a value close to γ_c.

γ	$1 - \gamma$	R_c	CONVERGENCE FACTOR
1	0	0.38197	0.90451
0.875	0.125	0.20177	0.82662
0.750	0.250	0.11915	0.72305
0.500	0.500	0.05582	0.48909
0.375	0.625	0.04141	0.36757
0.250	0.750	0.03157	0.24484
0.23969	0.76031	0.03090	0.23469

Table 4.1.

IV.6. Concluding Remarks.

Conte and Dames, of course, did not advocate that R should be kept constant during iterations. In fact, they obtained a set of iteration parameters R_k (k = 1, 2, . . , m), similar to Wachspress parameters for Laplace's equation, which, for N = 10, reduced the error by approximately 10^{-6} for 18 double sweeps over the grid. In comparison, it can be seen from Table4.1 that the optimum convergence factor for the method formulated in this chapter is 0.23469 when N = 10, which reduces the error by approximately 10^{-6} after 10 double sweeps, or by approximately 10^{-11} after 18 double sweeps. Alternatively, an error reduction of approximately 10^{-6} after 18 double sweeps can be obtained by choosing $\gamma = 0.5$, which from Table4.1 leads to a convergence factor of 0.48909.

It is felt that much experimental work could be carried out with advantage on the biharmonic equation. In particular, numerical experiments with variable parameters similar to those carried out in chapter II for Laplace's equation, could determine the relative merits of the ADI method and Parter's two-line iterative method for solving the biharmonic equation.

Finally, it should be pointed out that the method proposed in this chapter can be used for boundary conditions other than those described by (4.2), (see Conte and Dames [8]). If the region departs from the rectangular, however, the method will require further justification and the convergence factors quoted will require modification.

Appendix. Solution of Quidiagonal Systems of Equations.

The quidiagonal systems arising from the iterative procedure (4.11) may be written in the general form

$$(A4.1)\qquad \begin{bmatrix} c_1 & d_1 & e_1 & & & & & \\ b_2 & c_2 & d_2 & e_2 & & & & \\ a_3 & b_3 & c_3 & d_3 & e_3 & & & \\ & a_4 & b_4 & c_4 & d_4 & e_4 & & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ & & & a_{N-2} & b_{N-2} & c_{N-2} & d_{N-2} & \\ & & & & a_{N-1} & b_{N-1} & c_{N-1} & \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ \cdot \\ u_{N-2} \\ u_{N-1} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ \cdot \\ f_{N-2} \\ f_{N-1} \end{bmatrix}$$

Such systems may be solved by the following elimination scheme:

Let

$$w_1 = c_1$$
$$\beta_1 = d_1/w_1, \quad \beta_0 = 0, \quad \beta_{N-1} = 0,$$
$$\gamma_1 = e_1/w_1, \quad \gamma_0 = 0, \quad \gamma_{N-1} = \gamma_{N-2} = 0,$$

and define recursively

$$\delta_j = b_j - a_j \beta_{j-2},$$
$$w_j = c_j - a_j \gamma_{j-2} - \delta_j \beta_{j-1}, \qquad \text{for } 2 \le j \le N-1$$
$$\beta_j = (d_j - \delta_j \gamma_{j-1})/w_j$$
$$\gamma_j = e_j/w_j.$$

Next form

$$h_0 = 0,$$
$$h_1 = f_1/w_1,$$
$$h_j = (f_j - a_j h_{j-2} - \delta_j h_{j-1})/w_j, \qquad 2 \le j \le N-1.$$

The values of u are then obtained successively from the formulae

$$u_{N-1} = h_{N-1}$$

$$u_j = h_j - \beta_j u_{j+1} - \gamma_j u_{j+2}, \quad j = N-2, N-3, \ldots, 1.$$

For one complete iteration that is, a double sweep over the grid, one must solve 2(N-1) systems of the form (A4.1). Each system is of order N-1, and requires twice the work necessary for the solution of a tridiagonal system.

CHAPTER V.

A HIGH ACCURACY ADI METHOD FOR THE WAVE EQUATION.

V.1. Introduction

In 1962, Konovalov [23] formulated an ADI method of Douglas-Rachford type for the numerical solution of the wave equation in two space variables. In the same year, Lees [25] derived two ADI methods, also of Douglas-Rachford type, each based on the Von Neumann difference equation [31] in two space variables and of a higher accuracy than the method of Konovalov. In this chapter, The methods of Lees will be generalized and an ADI scheme of increased accuracy obtained. The results are then extended to the wave equation in three space variables.

V.2. The Methods of Konovalov and Lees.

Consider the wave equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \tag{5.1}$$

for $u \equiv u(x,y,t)$ where x, y are the space coordinates and t is the time, subject to the initial conditions $u(x,y,0) = f(x,y)$, $u_t(x,y,0) = g(x,y)$ over the unit square $0 < x,y < 1$, and the boundary condition $u(x,y,t) = 0$ for $t \geq 0$ at points on the boundary of the unit square. Other boundary conditions can also be considered, as described in previous chapters of this dissertation.

Konovalov's method is based on the backward difference equation

$$u_{m+1} - 2u_m + u_{m-1} = p^2(\delta_x^2 + \delta_y^2)u \tag{5.2}$$

and is given by

$$\begin{aligned} &u^*_{m+1} - 2u_m + u_{m-1} - p^2(\delta_x^2 u^*_{m+1} + \delta_y^2 u_m) = 0 \\ &u^*_{m+1} = u_{m+1} - p^2\delta_y^2(u_{m+1} - u_m), \end{aligned} \tag{5.3}$$

where $p = \Delta t/h$.

Elimination of u^*_{m+1} from (5.3) leads to

$$u_{m+1} - 2u_m + u_{m-1} - p^2(\delta_x^2 + \delta_y^2)u_{m+1} + p^4\delta_x^2\delta_y^2(u_{m+1} - u_m) = 0$$

which is seen to be a perturbation of (5.2). The principal part of the truncation error of this formula is

$$-p^3 h^3 \frac{\partial}{\partial t}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right).$$

Lees proposed two ADI methods for the solution of (5.1) each of which is a perturbation of the two-dimensional analogue of the standard Von Neumann difference equation

$$u_{m+1} - 2u_m + u_{m-1} = p^2\delta_x^2[\eta u_{m+1} + (1 - 2\eta)u_m + \eta u_{m-1}]$$

where η is a parameter. The first method is given by

$$u^*_{m+1} - 2u_m + u_{m-1} - p^2\delta_x^2[\eta u^*_{m+1} + (1 - 2\eta)u_m + \eta u_{m-1}]$$

$$- p^2\delta_y^2[(1 - 2\eta)u_m + 2\eta u_{m-1}] = 0 \qquad (5.4)$$

$$u^*_{m+1} = u_{m+1} - p^2\eta\,\delta_y^2(u_{m+1} - u_{m-1})$$

in the usual notation. If u^*_{m+1} is eliminated from (5.4), we obtain the formula

$$u_{m+1} - 2u_m + u_{m-1} - p^2(\delta_x^2 + \delta_y^2)[\eta u_{m+1} + (1 - 2\eta)u_m + \eta u_{m-1}]$$

$$+ p^4\eta^2\,\delta_x^2\delta_y^2[u_{m+1} - u_{m-1}] = 0. \qquad (5.5)$$

which has a truncation error of

$$-h^4[(\eta - \tfrac{1}{12})p^4\nabla^4 u + \tfrac{1}{12}p^2\left(\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}\right)] + O(h^5) + \text{higher order terms.}$$

The second method suggested by Lees,

$$u^*_{m+1} - 2u_m + u_{m-1} - p^2\delta_x^2[\eta u^*_{m+1} + (1 - 2\eta)u_m + \eta u_{m-1}] - p^2\delta_y^2 u_m = 0 \qquad (5.6)$$

$$u^*_{m+1} = u_{m+1} - \eta p^2\delta_y^2(u_{m+1} - 2u_m + u_{m-1})$$

yields

$$u_{m+1} - 2u_m + u_{m-1} - p^2(\delta_x^2 + \delta_y^2)[\eta u_{m+1} + (1 - 2\eta)u_m + \eta u_{m-1}]$$

$$+ p^4\eta^2\,\delta_x^2\delta_y^2(u_{m+1} - 2u_m + u_{m-1}) = 0, \qquad (5.7)$$

on elimination of u^*_{m+1}. This method has a truncation error

$$-h^4[(\eta - \tfrac{1}{12})p^4\nabla^4 u + \tfrac{1}{12}p^2\left(\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}\right)] + O(h^6) + \text{higher order terms.}$$

Thus the principal parts of the truncation errors are the same for both methods and are an order h higher than that of Konovalov's method.

V.3. Generalized ADI Method and An ADI Method of Increased Accuracy.

Formulae (5.5) and (5.7) are particular cases of the generalized formula

$$(5.8)\quad u_{m+1} - 2u_m + u_{m-1} + (\delta_x^2 + \delta_y^2)(au_{m+1} + bu_m + cu_{m-1}) + \delta_x\delta_y(du_{m+1} + eu_m + fu_{m-1}) = 0$$

which can be factorised into

$$u^*_{m+1} - 2u_m + u_{m-1} + \delta_x^2(au^*_{m+1} + bu_m + cu_{m-1}) + \delta_y^2[(b - \tfrac{e}{a})u_m + (c - \tfrac{f}{a})u_{m-1}] = 0 \qquad (5.9)$$

$$u^*_{m+1} = u_{m+1} + \delta_y^2(au_{m+1} + \tfrac{e}{a}u_m + \tfrac{f}{a}u_{m-1})$$

provided $d = a^2$.

We now expand the terms in (5.8) as Taylor series in terms of u and its derivatives at the node $(ih, jh, m\Delta t)$, the even derivatives with respect to t being replaced by $\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}$, $\frac{\partial^4 u}{\partial t^4} = \frac{\partial^4 u}{\partial x^4} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} + \frac{\partial^4 u}{\partial y^4}$, etc., from (5.1). The expansions up to and including terms involving h^6 are

$$u_{m+1} - 2u_m + u_{m-1} = p^2A_1 + \tfrac{1}{12}p^4A_3 + \tfrac{1}{6}p^4A_4 + \tfrac{1}{720}p^6A_7 + \tfrac{1}{240}p^6A_8$$

$$(\delta_x^2 + \delta_y^2)u_{m+1} = A_1 + pA_2 + \tfrac{1}{12}(p^2 + \tfrac{1}{6})A_3 + p^2A_4 + \tfrac{1}{6}p(p^2 + \tfrac{1}{2})A_5 + \tfrac{1}{3}p^3A_6 + (\tfrac{1}{24}p^4 + \tfrac{1}{24}p^2 + \tfrac{1}{360})A_7 + \tfrac{1}{8}p^2(p^2 + \tfrac{1}{3})A_8$$

$$\delta_x^2\delta_y^2u_{m+1} = A_4 + pA_6 + \tfrac{1}{12}(p^2 + \tfrac{1}{180})A_8$$

$$(\delta_x^2 + \delta_y^2)u_{m-1} = A_1 - pA_2 + \tfrac{1}{12}(p^2 + \tfrac{1}{6})A_3 + p^2A_4 - \tfrac{1}{6}p(p^2 + \tfrac{1}{2})A_5 - \tfrac{1}{3}p^3A_6 + (\tfrac{1}{24}p^4 + \tfrac{1}{24}p^2 + \tfrac{1}{360})A_7 + \tfrac{1}{8}p^2(p^2 + \tfrac{1}{3})A_8$$

$$\delta_x^2\delta_y^2u_{m-1} = A_4 - pA_6 + \tfrac{1}{12}(p^2 + \tfrac{1}{180})A_8$$

$$(\delta_x^2 + \delta_y^2)u_m = A_1 + \tfrac{1}{12}A_3 + \tfrac{1}{360}A_7$$

$$\delta_x^2\delta_y^2u_m = A_4 + \tfrac{1}{360}A_8$$

where

$$A_1 = h^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right), \quad A_2 = h^3\frac{\partial}{\partial t}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right), \quad A_3 = h^4\left(\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}\right),$$

$$A_4 = h^4\frac{\partial^4 u}{\partial x^2 \partial y^2}, \qquad A_5 = h^5\frac{\partial}{\partial t}\left(\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}\right), \quad A_6 = h^5\frac{\partial^5 u}{\partial t \partial x^2 \partial y^2},$$

$$A_7 = h^6\left(\frac{\partial^6 u}{\partial x^6} + \frac{\partial^6 u}{\partial y^6}\right), \quad A_8 = h^6\frac{\partial^4}{\partial x^2 \partial y^2}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right).$$

If these expansions are substituted into (5.8) and the derivatives A_i $(i = 1, 2, \ldots, 6)$ eliminated, the coefficients become

$$a = c = \tfrac{1}{12}(1 - p^2), \quad b = -\tfrac{1}{6}(1 + 5p^2), \quad d = f = \tfrac{1}{144}(1 - p^2)^2,$$

$$e = -\tfrac{1}{72}(1 + 10p^2 + p^4)$$

and so we obtain the formula

$$u_{m+1} - 2u_m + u_{m-1} + (\delta_x^2 + \delta_y^2)\left[\tfrac{1}{12}(1 - p^2)(u_{m+1} + u_{m-1}) - \tfrac{1}{6}(1 + 5p^2)u_m\right]$$

$$(5.10) \qquad + \delta_x^2\delta_y^2\left[\tfrac{1}{144}(1 - p^2)^2(u_{m+1} + u_{m-1}) - \tfrac{1}{72}(1 + 10p^2 + p^4)u_m\right] = 0,$$

which, since $d = a^2$, can be written the ADI form

$$u^*_{m+1} - 2u_m + u_{m-1} + \tfrac{1}{12}(1 - p^2)\delta_x^2\left[u_{m+1} - \frac{2(1 + 5p^2)}{(1 - p^2)}u_m + u_{m-1}\right]$$

$$(5.11) \qquad + p^2\frac{(1 + p^2)}{(1 - p^2)}\delta_y^2 u_m = 0$$

$$u^*_{m+1} = u_{m+1} + \tfrac{1}{12}(1 - p^2)\delta_y^2\left[u_{m+1} - \frac{2(1 + 10p^2 + p^4)}{(1 - p^2)^2}u_m + u_{m-1}\right]$$

which is an order h^2 higher than either of Lees' original formulae.

If $p = 1$, the ADI method degenerates into the explicit formula

$$u_{m+1} - 2u_m + u_{m-1} - (\delta_x^2 + \delta_y^2)u_m - \delta_x^2\delta_y^2 u_m = 0.$$

V.4. Stability.

The stability of the difference equation (5.10) is examined using the method employed in II.6. When we substitute

$$\varepsilon_{i,j,m} = a_m \sin \pi s x_i \sin \pi q y_j \qquad (s,q = 1, 2, \ldots, N-1)$$

into (5.10) with u replaced by ε, we obtain

(5.12) $$a_{m+1} - 2\alpha a_m + a_{m-1} = 0$$

where

$$\alpha = \frac{1 - \frac{1}{12}(1 + 5p^2)(S_s^2 + S_q^2) + \frac{1}{144}(1 + 10p^2 + p^4)S_s^2 S_q^2}{1 - \frac{1}{12}(1 - p^2)(S_s^2 + S_q^2) + \frac{1}{144}(1 - p^2)S_s^2 S_q^2}$$

and $S_s^2 = 4\sin^2 \pi s/2N$, $S_q^2 = 4\sin^2 \pi q/2N$. The difference equation (5.10) will be stable provided the roots of the equation

$$\rho^2 - 2\alpha\rho + 1 = 0$$

(c.f.(1.19)) are at most equal to one in absolute value. These roots are equal to one in absolute value if $|\alpha| \leq 1$. This requirement leads to the inequalities

$$-(S_s^2 + S_q^2) + \tfrac{1}{6}S_s^2 S_q^2 \leq 0$$

$$2 - \tfrac{1}{6}(1 + 2p^2)(S_s^2 + S_q^2) + \tfrac{1}{72}(1 + 4p^2 + p^4)S_s^2 S_q^2 \geq 0$$

where $0 \leq S_s^2, S_q^2 \leq 4$. The first inequality is satisfied for all S_s^2, S_q^2 in the permissible range, while the second is satisfied provided

(5.13) $$p \leq \sqrt{3} - 1.$$

This conditional stability is not a serious drawback in solving the wave equation because the upper limit placed on the time increment is proportional to the mesh size in the distance coordinate and not to the square of the mesh size as in the heat conduction equation. Thus calculations can be carried out which can reach reasonable values of the

time without using a prohibitive number of time steps. In addition, there is no asymptotic solution of the wave equation as the time tends to infinity and so there is no particular significance attached to the solution of the wave equation for large values of the time.

Lees used energy methods to prove that (5.5) and (5.7) are unconditionally stable if $\eta > \frac{1}{4}$. Using the Von Neumann method for stability analysis, it may be shown that the second method of Lees is unconditionally stable for $\eta > \frac{1}{8}$. When $\eta \leq \frac{1}{8}$, for stability, we require

(5.14) $$p \leq \sqrt{2}(1 + \sqrt{1 - 8\eta})^{-\frac{1}{2}}.$$

V.5. The Three-Space Variable Case.

The logical extension of the first method of Lees (5.4) to three space dimensions for the solution of

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

is

$$u^*_{m+1} - 2u_m + u_{m-1} - p^2\delta_x^2[\eta u^*_{m+1} + (1 - 2\eta)u_m + \eta u_{m-1}]$$
$$- p^2(\delta_y^2 + \delta_z^2)[(1 - 2\eta)u_m + 2\eta u_{m-1}] = 0$$

$$u^*_{m+1} = u^{**}_{m+1} - p^2\eta\,\delta_y^2(u^{**}_{m+1} - u_{m-1})$$

$$u^{**}_{m+1} = u_{m+1} - p^2\eta\,\delta_z^2(u_{m+1} - u_{m-1}),$$

with obvious extension to the original notation. If u^*_{m+1} and u^{**}_{m+1} are eliminated from these equations, we obtain

$$(5.15)\quad u_{m+1} - 2u_m + u_{m-1} - p^2\textstyle\sum \delta_x^2[\eta(u_{m+1} + u_{m-1}) + (1 - 2\eta)u_m]$$
$$+ \eta^2 p^4 \textstyle\sum \delta_x^2\delta_y^2(u_{m+1} - u_{m-1}) - \eta^3 p^6 \delta_x^2\delta_y^2\delta_z^2(u_{m+1} - u_{m-1}) = 0$$

where $\sum\delta_x^2 = \delta_x^2 + \delta_y^2 + \delta_z^2$ and $\sum\delta_x^2\delta_y^2 = \delta_x^2\delta_y^2 + \delta_y^2\delta_z^2 + \delta_z^2\delta_x^2$.

The second method of Lees becomes

$$u^*_{m+1} - 2u_m + u_{m-1} - p^2\delta_x^2[\eta u^*_{m+1} + (1 - 2\eta)u_m + \eta u_{m-1}]$$
$$+ p^2(\delta_y^2 + \delta_z^2)u_m = 0$$

$$u^*_{m+1} = u^{**}_{m+1} - \eta p^2\delta_y^2(u^{**}_{m+1} - 2u_m + u_{m-1})$$

$$u^{**}_{m+1} = u^{***}_{m+1} - \eta p^2\delta_z^2(u_{m+1} - 2u_m + u_{m-1})$$

which, on elimination of u^*_{m+1} and u^{**}_{m+1}, leads to

$$(5.16)\quad u_{m+1} - 2u_m + u_{m-1} - p^2\textstyle\sum \delta_x^2[\eta(u_{m+1} + u_{m-1}) + (1 - 2\eta)u_m]$$
$$+ \eta^2 p^4 \textstyle\sum \delta_x^2\delta_y^2(u_{m+1} - 2u_m + u_{m-1}) - \eta^3 p^6 \delta_x^2\delta_y^2\delta_z^2(u_{m+1} - 2u_m + u_{m-1}) = 0.$$

A generalized form of (5.15) and (5.16) is

$$(5.17)\quad u_{m+1} - 2u_m + u_{m-1} + \sum\delta_x^2(Au_{m+1} + Bu_m + Cu_{m-1}) + \sum\delta_x^2\delta_y^2(Du_{m+1} + Eu_m + Fu_{m-1}) + \delta_x^2\delta_y^2\delta_z^2(Gu_{m+1} + Hu_m + Iu_{m-1}) = 0$$

which may be factorised into

$$u^*_{m+1} - 2u_m + u_{m-1} + \delta_x^2(Au^*_{m+1} + Bu_m + Cu_{m-1}) + (\delta_y^2 + \delta_z^2)[(B - \tfrac{E}{A})u_m + (C - \tfrac{F}{A})u_{m-1}] = 0$$

$$(5.18)\quad u^*_{m+1} = u^{**}_{m+1} + \delta_y^2(Au^{**}_{m+1} + \tfrac{E}{A}u_m + \tfrac{F}{A}u_{m-1})$$

$$u^{**}_{m+1} = u_{m+1} + \delta_z^2(Au_{m+1} + \tfrac{E}{A}u_m + \tfrac{F}{A}u_{m-1})$$

provided

$$D = A^2, \quad G = A^3, \quad H = EA, \quad I = AF.$$

By expanding the terms u_{m+1}, $\sum\delta_x^2u_{m+1}$, etc., in (5.17) as Taylor series, we obtain the optimum formula as in V.3, the coefficients in this case being

$$A = C = \tfrac{1}{12}(1 - p^2), \quad B = -\tfrac{1}{6}(1 + 5p^2), \quad D = F = \tfrac{1}{144}(1 - p^2)^2,$$

$$E = -\tfrac{1}{72}(1 + 10p^2 + p^4).$$

Substitution of these values into (5.17) yields the formula

$$(5.19)\quad u_{m+1} - 2u_m + u_{m-1} + \tfrac{1}{12}\sum\delta_x^2[(1 - p^2)(u_{m+1} + u_{m-1}) - 2(1 + 5p^2)u_m] + \tfrac{1}{144}[\sum\delta_x^2\delta_y^2 + \tfrac{1}{12}(1 - p^2)\delta_x^2\delta_y^2\delta_z^2][(1 - p^2)^2(u_{m+1} + u_{m-1}) - 2(1 + 10p^2 + p^4)u_m] = 0$$

which factorises into

$$u^*_{m+1} - 2u_m + u_{m-1} + \tfrac{1}{12}(1 - p^2)\delta_x^2[u_{m+1} - \frac{2(1 + 5p^2)}{(1 - p^2)}u_m + u_{m-1}] + \frac{p^2(1 + p^2)}{(1 - p^2)^2}(\delta_y^2 + \delta_z^2)u_m = 0$$

$$(5.20)\quad u^{*}_{m+1} = u^{**}_{m+1} + \tfrac{1}{12}(1 - p^2)\delta_y^2\left[u^{**}_{m+1} - \frac{2(1 + 10p^2 + p^4)}{(1 - p^2)^2}\, u_m + u_{m-1}\right]$$

$$u^{**}_{m+1} = u_{m+1} + \tfrac{1}{12}(1 - p^2)\delta_z^2\left[u_{m+1} - \frac{2(1 + 10p^2 + p^4)}{(1 - p^2)^2}\, u_m + u_{m-1}\right].$$

Formulae (5.20) constitute an ADI method for solving the wave equation in three space dimensions which is an order h^2 more accurate than the three dimensional analogues of the methods of Lees.

The stability requirement of (5.19) is exactly the same as in the two dimensional case, namely $p \leq \sqrt{3} - 1$.

V.6. Numerical Experiments.

The ADI method (5.11) is now used to solve a model problem consisting of the wave equation (5.1) together with the boundary conditions $u = \sin \pi x \sin \pi y$, $\frac{\partial u}{\partial t} = 0$ for $0 \leq x,y \leq 1$, $t = 0$, and $u = 0$ on the boundary of the unit square for $t > 0$. The theoretical solution of this problem is

$$u = \sin \pi x \sin \pi y \cos 2\pi t \tag{5.21}$$

within the open region $0 \leq x,y \leq 1$, $t > 0$.

To start any calculation, it is necessary to know the value of u at the nodal points on the plane $t = \Delta t$ as well as those on the plane $t = 0$. If u and $\frac{\partial u}{\partial t}$ are given at $t = 0$ as functions of x and y, then $u(ih, jh, \Delta t)$, $(i,j = 1, 2, \ldots, N-1)$, may be calculated from the Taylor series expansions involving u and its derivatives at $t = 0$, replacing second and higher derivatives with respect to t using (5.1). However, if u and $\frac{\partial u}{\partial t}$ are given numerically at $t = 0$, some other method of determining $u(ih, jh, \Delta t)$ must be employed. In all the calculations carried out, $u(ih, jh, \Delta t)$ was obtained from the known theoretical solution, which is equivalent to the Taylor series expansion for this problem.

The two methods of Lees are of comparable accuracy, and we compare the ADI method (5.11) with the second method of Lees. In all experiments carried out, the mesh size is taken to be $\frac{1}{11}$ in the x and y directions. The mesh size in the t direction is then given by $\Delta t = p/11$, where p,

the mesh ratio, must satisfy (5.13) for stability. The error in any calculation, which is defined as the difference between the computed solution and the theoretical solution given by (5.21), is always quoted at one of the four nodes nearest to the centre of the unit square. In Table 5.1, the errors using the second method of Lees are shown at every fifth time step for $p = 0.66$ and $\eta = \frac{1}{4}, \frac{1}{2}, 1$. The minimum error occurs when $\eta = \frac{1}{4}$. Also shown for comparison in Table 5.1 are the errors using the method (5.11) with $p = 0.66$. The values obtained by this method appear to be three places of decimals better than the corresponding values obtained by the method of Lees ($\eta = \frac{1}{4}$). It should be noted that the method of Lees can be made more accurate by choosing $\eta < \frac{1}{4}$, since the error is almost proportional to η . However, the gain in accuracy is obtained at the expense of a loss in unconditional stability. Lees' method will, of course, never achieve the accuracy of the method (5.11).

Calculations are now carried out using the method (5.11) for a range of time increments 0.03, 0.05, 0.06, 0.075. The errors are given in Table 5.2 at time intervals of 0.3 up to $t = 3.0$. The errors are of comparable order, with computations based on $\Delta t = 0.06$ giving greatest accuracy. The calculations with $\Delta t = 0.075$ are, of course, theoretically unstable and the instability becomes apparent beyond $t = 1.8$.

In conclusion, it should be emphasised that the theory of the ADI methods presented in this chapter only holds if the region under consideration is rectangular since, as we have seen in chapter II, the operators δ_x and δ_y commute only on a rectangle. This does not mean

τ \ h	$\frac{1}{4}$	$\frac{1}{2}$	1	HIGH ACCURACY METHOD $\Delta\tau = 0.06$
0.3	-0.009,353	-0.018,065	-0.034,928	-0.000,008
0.6	-0.010,148	-0.020,010	-0.040,224	-0.000,008
0.9	0.025,128	0.047,913	0.090,206	0.000,021
1.2	0.037,914	0.074,418	0.148,135	0.000,031
1.5	-0.019,930	-0.035,643	-0.057,948	-0.000,018
1.8	-0.069,506	-0.134,848	-0.261,908	-0.000,058
2.1	-0.011,155	-0.027,693	-0.076,422	-0.000,007
2.4	0.087,278	0.165,541	0.306,223	0.000,075
2.7	0.062,022	0.128,138	0.276,733	0.000,048
3.0	-0.076,253	-0.137,287	-0.224,734	-0.000,068

Table 5.1.

τ \ $\Delta\tau$	0.03	0.05	0.06	0.075 [UNSTABLE]
0.3	-0.000,015	-0.000,012	-0.000,008	0.000,001
0.6	-0.000,015	-0.000,012	-0.000,008	0.000,001
0.9	0.000,039	0.000,031	0.000,021	-0.000,003
1.2	0.000,056	0.000,045	0.000,031	-0.000,005
1.5	-0.000,032	-0.000,026	-0.000,018	0.000,002
1.8	-0.000,103	-0.000,082	-0.000,058	-0.000,004
2.1	-0.000,011	-0.000,009	-0.000,007	-0.000,186
2.4	0.000,132	0.000,106	0.000,075	-0.002,775
2.7	0.000,085	0.000,068	0.000,048	-0.040,980
3.0	-0.000,121	-0.000,097	-0.000,068	-0.609,823

Table 5.2.

that the ADI methods fail for problems involving non-rectangular regions. However, much experimental work is necessary to determine whether or not the ADI methods attain the same accuracy in such cases.

CONCLUDING REMARKS.

In this thesis, new ADI methods have been developed for linear partial differential equations with constant coefficients. The extension of these methods to linear equations with variable coefficients requires investigation. Also, the ADI methods formulated have employed the minimum number of levels of time, namely two for the diffusion equation and Laplace's equation, and three for the wave equation. It may be possible to formulate for these equations ADI methods of the type described in this thesis which involve more time levels. This has in fact been done by Douglas, Gannon and Gunn [16,55] for their class of ADI methods. In general, however, higher level methods require increased computational effort, and, in particular, Gannon and Douglas [55] have been unable to obtain a parameter sequence for a three-level method of solving Laplace's equation.

The ADI method (2.22) of Chapter II for solving Laplace's equation can only be used when the boundaries of the region under consideration are parallel to the coordinate axes. An extension of the method is required to cover the case of a region with a slanting boundary.

Finally, parameter sequences for the generalized iterative procedures formulated in Chapters III and IV for solving Laplace's equation in three space variables and the biharmonic equation respectively have yet to be determined. Because of the complicated nature of the convergence factors of these methods, the determination of parameter

sequences in these cases may prove to prove to be a formidable task.

References.

[1] BAKES, M.D.: An alternative method of solution of certain tridiagonal systems of linear equations. Computer J. 7, 135-136 (1964).

[2] BIRKHOFF, G. and VARGA, R.S.: Implicit alternating direction methods. Trans. Amer. Math. Soc. 92, 13-24 (1959).

[3] BIRKHOFF, G., VARGA, R.S. and YOUNG, D.M.: Alternating direction implicit methods. Advances in Computers Vol 3, F.L.Alt(editor), Academic Press 1962, pp 189-273.

[4] BRIAN, P.L.T.: A finite difference method of high-order accuracy for the solution of three-dimensional transient heat conduction problems. A.I.Ch.E. Journal 7, 367-370 (1961).

[5] BRUCE, G.H., PEACEMAN, D.W., RACHFORD, H.H. and RICE, J.D.: Calculations of unsteady-state gas flow through porous media. Trans. A.I.M.E. 198, 79 (1953).

[6] CONTE, S.D.: Numerical solution of vibration problems in two space variables. Pacific J. 7, 1535-1544 (1957).

[7] CONTE, S.D. and DAMES, R.T.: An alternating direction method for solving the biharmonic equation. M.T.A.C. 12, 198-205 (1958).

[8] CONTE, S.D. and DAMES, R.T.: On an alternating direction method for solving the plate problem with mixed boundary conditions. J. Assoc. Comp. Mach. 7, 264-273 (1960).

[9] COURANT, R., FRIEDRICHS, K.O. and LEWY, H.: Uber die partiellen differenzengleichungen der mathematishen physik. Math. Ann. 100, 32-74 (1928).

[10] DE BOOR, C.M. and RICE, J.R.: Tchebycheff approximation by

$$\alpha \prod [(x - r_j)/(x + r_j)]$$

and application to ADI iteration. J. Soc. Indust. Appl. Math. 11, 159-169 (1963).

[11] DOUGLAS, J.: Alternating direction methods for three space variables. Numer. Math. 4, 41-63 (1962).

[12] DOUGLAS, J., and RACHFORD, H.H.: On the numerical solution of heat conduction problems in two and three space variables. Trans. Amer. Math. Soc. 82, 421-439 (1956).

[13] DOUGLAS, J., KELLOGG, R.B. and VARGA, R.S.: Alternating direction iteration methods for n space variables. Math. Comp. 17, 279-282 (1963).

[14] DOUGLAS, J.: On the numerical integration of $u_{xx} + u_{yy} = u_t$ by implicit methods. J. Soc. Indust. Appl. Math. 3, 42-65 (1955).

[15] DOUGLAS, J.: A survey of numerical methods for parabolic differential equations. Advances in Computers Vol 2, F.L.Alt (editor), Academic Press 1961, pp 1-54.

[16] DOUGLAS, J. and GUNN, J.E.: A general formulation of alternating direction methods. Part 1. Parabolic and hyperbolic problems. Numer. Math. 6, 428-453 (1964).

[17] DOUGLAS, J.: On the relation between stability and convergence in the numerical solution of linear parabolic and hyperbolic equations. J. Soc. Indust. Appl. Math. 4, 20-37 (1956).

[18] EVANS, D.J. and FORRINGTON, C.V.D.: Note on the solution of certain tridiagonal systems of linear equations. Computer J. 5, 327-328 (1962).

[19] FAIRWEATHER, G. and MITCHELL, A.R.: A generalised alternating direction method of Douglas-Rachford type for solving the biharmonic equation. Computer J. 7, 242-245 (1964).

[20] FAIRWEATHER, G. and MITCHELL, A.R.: A new alternating direction method for parabolic equations in three space variables. J. Soc. Indust. Appl. Math. (To appear).

[21] FAIRWEATHER, G. and MITCHELL, A.R.: A high accuracy alternating direction method for the wave equation. J. Inst. Math. Appl. (To appear).

[22] KANTOROVICH, L.V. and KRYLOV, V.I.: Approximate methods of higher analysis. Translated by C.D.Benster. New York: Interscience 1958, pp 616-639.

[23] KONOVALOV, A.N.:The method of fractional steps for solving the Cauchy problem for the multi-dimensional wave equation. Dokl. Akad. Nauk. 147, 25-27 (1962).

[24] LAX, P.D. and RICHTMYER, R.D.: Survey of the stability of linear finite difference equations. Comm. Pure Appl. Math. 9, 267-293 (1956).

[25] LEES, M.: Alternating direction methods for hyperbolic differential equations. J. Soc. Indust. Appl. Math. 10, 610-616 (1962).

[26] MILLER, K.: Numerical analogs to the Schwarz alternating procedure. Numer. Math. 7, 91-103 (1965).

[27] LOWAN, A.N.: The operator approach to problems of stability and convergence of solutions of difference equations and the convergence of various iteration procedures. Scripta Mathematica, New York, 1957.

[28] MARTIN, D.W. and G.J. TEE: Iterative methods for linear equations with symmetric positive definite matrix. Computer J. 4, 242-254 (1961).

[29] MILNE-THOMSON, L.M.: Calculus of finite differences. Macmillan and Co., London, 1951.

[30] MITCHELL, A.R. and FAIRWEATHER, G.: Improved forms of the alternating direction methods of Douglas, Peaceman and Rachford for solving parabolic and elliptic equations. Numer. Math. 6, 285-292 (1964).

[31] O'BRIEN, G.G., HYMAN, M.A. and KAPLAN, S.: A study of the numerical solution of partial differential equations. J. Math. Phys. 29, 223-251 (1951).

[32] PARTER, S.V.: On 'two-line' iterative methods for the Laplace and biharmonic difference equations. Numer. Math. 1, 240-252 (1959).

[33] PARTER, S.V.: Some computational results on 'two-line' iterative methods for the biharmonic difference equation. J. Assoc. Comp. Mach. 8, 359-365 (1961).

[34] PEACEMAN, D.W. and RACHFORD, H.H.: The numerical solution of parabolic and elliptic differential equations. J. Soc. Indust. Appl. Math. 3, 28-41 (1955).

[35] PEARCY, C.M.: On convergence of alternating direction procedures. Numer. Math. 4, 172-176 (1962).

[36] PRICE, H.S. and VARGA, R.S.: Recent numerical experiments comparing successive overrelaxation iterative methods with alternating direction implicit methods. Gulf Research and Development Company, Pittsburgh, Pa., 1962.

[37] RICHTMYER, R.D.: Difference methods for initial value problems. Interscience Tracts in Pure and Applied Mathematics, Vol 4, Interscience Publishers, New York, 1957.

[38] RUTHERFORD, D.E.: Some continuant determinants arising in physics and chemistry. Proc. Royal Soc. Edinburgh 62, 229-236 (1947); 63, 232-261 (1952).

[39] SNEDDON, I.N.: Elements of partial differential equations. M'Graw Hill, New York, 1957.

[40] TEE, G.J.: A novel finite difference approximation to the biharmonic operator. Computer J. 6, 177-192 (1963).

[41] THOMAS, L.H.: Elliptic problems in linear difference equations over a network. Watson Scientific Computing Laboratory, Columbia University, New York, 1949.

[42] TING, T.C.T.: A method of solving a system of linear equations whose coefficients form a tridiagonal matrix. Quart. Appl. Math. 22, 105-116 (1964).

[43] TODD, J.: A direct approach to the problem of stability in the numerical solution of partial differential equations. Comm. Pure Appl. Math. 9, 597-612 (1956).

[44] VARGA, R.S.: Matrix iterative analysis. Prentice-Hall, New Jersey, 1962.

[45] WACHSPRESS, E.L.: CURE: a generalised two-space-dimension multi-group coding for the IBM-704. Report KAPL 1724, Knolls Atomic Power Laboratory, General Electric Company, Schnectady, New York, 1957.

[46] WACHSPRESS, E.L. and HABETLER, G.J.: An alternating-direction-implicit iteration technique. J. Soc. Indust. Appl. Math. 10, 343-363 (1962).

[47] VARGA, R.S.: Extended application of alternating direction implicit iteration model problem theory. J. Soc. Indust. Appl. Math. 11, 994-1016 (1963).

[48] VARGA, R.S.: Optimum alternating direction implicit parameters for a model problem. J. Soc. Indust. Appl. Math. 10, 343-363 (1962).

[49] WHITE, R.N.: Optimum solution techniques for finite difference equations. Proc. A.S.C.E. 89, 115-136 (1963).

[50] YOUNG, D.M. and EHRLICH, L.: Numerical experiments involving boundary problems in differential equations. Boundary Problems in Differential Equations, University of Wisconsin Press, Madison, Wisconsin, 1960, pp143-162.

[51] YOUNG, D.M. and FRANK, T.G.: A survey of computer methods for solving elliptic and parabolic partial differential equations. I.C.C. Bull. 2, 1-61 (1963).

[52] YOUNG, D.M.: Notes on the alternating direction implicit iteration method. Report NN-28, Digital Computing Center, The Ramo Wooldridge Corporation, 1956.

[53] SAUL'EV, V.K.: Solution of certain boundary value problems on high-speed computers by the fictitious-domain method. Siburek. Mat. Z. 4, 912-925 (1963).

[54] SAMARSKII, A.A. and ANDREEV, V.B.: On a difference scheme of increased accuracy for equations of elliptic type with some spatial variables. Z. Vycisl. Mat. i Mat. Fiz. 3, 1006-1013 (1963).

[55] CANNON, J.R. and DOUGLAS, J.: Three-level alternating direction iterative methods. Contributions to Differential Equations, 3, 189-198 (1964).