SHEFFIELD POLYTECHNIC

Declaration to be signed by each person depositing a thesis

I consent/~~do not consent~~ to this thesis being consulted, borrowed or photo-copied.

Signed ....S. J. Wisher....

Permanent address 15 Storthwood Court
Storth Lane
Sheffield S10 3HP.

I, the supervisor, consent/~~do not consent~~.

Signed ....Graham J. Leggett....

Dept. ....Mathematics & Statistics....

Without the consent of the author and College supervisor, no thesis may be consulted, borrowed, or photo-copied for five years after the date of its deposit.

ON THE USE OF SPLINES IN THE NUMERICAL SOLUTION OF HYPERBOLIC PARTIAL DIFFERENTIAL EQUATIONS

Being a thesis submitted for the degree of

Master of Philosophy

at

Sheffield City Polytechnic

by

STEPHEN JOHN WISHER

March 1977

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my Director of Studies, Dr G F Raggett, of Sheffield City Polytechnic for many helpful discussions during the preparation of this thesis.

My thanks are also due to Dr D F Mayers of Oxford University for some helpful suggestions on the numerical schemes, to Dr J A R Stone of Sheffield City Polytechnic for help in deriving the practical case studies, to the Polytechnic Computer Unit for the running of my programs and to Mrs Christine Barker for her expert typing of this thesis.

I am also indebted to the Sheffield Education Authority for the provision of a Research Assistantship.

CONTENTS

CHAPTER 1

Introduction

1.1 Spline Development

A spline is a simple mechanical device used by draftsmen for drawing smooth curves. It consists of a slender bar made of wood or some other flexible material to which weights are attached. The spline is placed on a sheet of graph paper and constrained by use of the weights so that it takes the shape of the curve we wish to draw.

The term "spline function" was first used by Schoenberg (1946) in a paper describing the use of generalised splines and other piecewise polynomials to approximate smooth functions of one variable. The properties of spline functions were in fact employed in a few isolated instances before this date although no reference was made to the name itself. Schoenberg's early paper was an important contribution to the use of spline functions but it was not until the 1960's that further work in the field was published.

In recent years spline functions have been used to solve a wide range of numerical problems. For example Birkhoff and De Boor (1965), Curtis and Powell (1967) and Schumaker (1969) have described methods of interpolating and approximating functions whilst the numerical solution of ordinary differential equations has been considered by Bickley (1968), Fyfe (1969) and Albasiny and Hoskins (1969). Although El Tom (1974, 1976) has proposed several spline function methods for solving integral equations, there has not until recently been significant work on the numerical solution of partial differential equations. In 1972 a method using cubic splines for solving moving boundary problems in heat flow was discussed by Crank and Gupta (1972). In their paper the authors used the spline function not to obtain general solutions to the problem but to determine the position of the boundary at each time level.

The general numerical solution of the well known one-dimensional heat conduction equation was considered by Papamichael and Whiteman (1973). A similar method has since been used by Raggett and Wilson (1974) to obtain solutions to the one-dimensional wave equation. These last two papers form the basis of the work described in this thesis.

1.2 Finite Differences

As will be seen in the following chapters we compare our methods using cubic splines with a well known fully implicit finite difference approximation. It is therefore useful for us to briefly describe the development of finite difference methods for the numerical solution of partial differential equations.

The basic types of partial differential equations, parabolic, hyperbolic and elliptic, were originally approximated using simple finite difference methods by Courant, Friedrichs and Lewy (1928). They also showed that the convergence of these methods depends on the mesh ratio satisfying certain conditions. This fact was also proved by von Neumann, who also applied the technique of stability analysis to a wide variety of problems during the Second World War. Some of von Neumann's work is discussed in a study of the numerical solution of partial differential equations by O'Brien, Hyman and Kaplan (1951). The finite difference approximations we have used in this thesis are in fact a natural extension of the scheme devised by von Neumann for the numerical solution of the wave equation.

Since von Neumann's work on finite differences there have been many types of schemes proposed, possibly the most well known being that of Crank and Nicolson (1947). The majority of these schemes use rectangular grids which have constant step lengths in both the time and space directions. There are, in fact, very few instances where unequal step lengths have been used for solving partial differential equations. Murray and Landis (1959) deformed the grid in the space direction by compressing or stretching it as an aid to solving moving boundary problems in heat flow. They considered the problem where the boundary between a solid region and a liquid region is moving with time. They then approximated the heat conduction equation, which describes these practical circumstances, by finite difference approximations with a constant step length on the solid side of the boundary and a different constant step length on the liquid side. This meant a change in the size of the step length from one side of the boundary to the other.

Moving boundary problems of this type are often referred to as "Stefan problems" and have also been solved using an alternative variable step length method due to Douglas and Gallie (1955). In this paper the authors introduced a variable time step length while keeping the size of the space mesh length fixed. This method is particularly useful in solving parabolic equations where, with the increase of time, the solution becomes smoother and smoother. It is therefore advisable to use a small time step length at the beginning of the computations and then increase it as time increases.

1.3 Present Work

As suggested in section 1.1 a great number of numerical problems can now be solved with the aid of spline functions. In chapter 2 we define spline functions in general and derive some expressions which results from the definition of the cubic spline, this being the degree of spline function we will use throughout this thesis. The cubic spline definition and these resulting expressions are given in detail in Ahlberg, Nilson and Walsh (1967).

It was indicated by Raggett (1974) that the cubic spline method for solving the one-dimensional wave equation (see Raggett and Wilson (1974)) could be extended to obtain solutions to more general hyperbolic partial differential equations. This work is shown in chapter 3 where, in addition to the splines schemes, we have derived a well known implicit finite difference approximation from which we will draw comparisons.

In chapter 4 both the splines scheme and the difference approximation are developed to cover equations having variable coefficients. The methods of solution in both this chapter and the previous one require the Knots for the splines schemes and the mesh points for the difference schemes to be equally spaced. The extension of the methods to cover arbitrarily spaced knots and mesh points is given in chapter 5.

The truncation errors and stability analysis for each of the constant coefficient and variable coefficient schemes, both using constant step lengths and variable step lengths, are given in their respective chapters. Numerical procedures are also described for obtaining spline solutions at points intermediate to the knots.

These additional solutions are found directly from the spline function itself. This is found to be a major advantage of the splines schemes over the more well known finite difference methods.

In chapter 6 several case studies are developed where equations with both constant and variable coefficients are discussed. The methods based on both uniformly and arbitrarily spaced knots are also employed.

CHAPTER 2

Spline Functions

2.1 Definition of a spline function

Consider an interval $a \leqslant x \leqslant b$ and subdivide it into m sub-intervals by inserting knots at the points $x_o, x_1, \ldots, x_m$ where

$$a = x_o < x_1 < \ldots < x_m = b \qquad (2.1.1)$$

Then a spline function s(x) of degree n with knots $x_o, x_1, \ldots, x_m$ is a function possessing the following two properties.

(i) In each interval $x_{i-1} \leqslant x \leqslant x_i$ $(i = 1,2, \ldots, m)$, s(x) is a polynomial of degree n or less.

(ii) S(x) and its derivatives of orders 1,2, ..., n-1 are continuous.

Thus a spline function is a piece-wise polynomial function satisfying certain conditions regarding continuity of the function and its derivatives. As explained by Greville (1969) when $n = o$ condition (ii) is not operative and a spline function of degree 0 is a step function. For $n > 0$, a spline function of degree n could equally well be defined as a function whose n^{th} derivative is a step function.

2.2 Cubic Splines

Let f(x) be a function with continuous derivatives in the range $a \leqslant x \leqslant b$. Then S(x) is a cubic spline interpolating to the function f(x) at the knots $x_o, x_1, \ldots, x_m$ if

(i) In each interval $x_{i-1} \leqslant x \leqslant x_i$ $(i = 1,2, \ldots, m)$, S(x) is a cubic polynomial.

(ii) $S'(x)$ and $S''(x)$ are continuous

(iii) $S(x_i) = f(x_i)$ $(i = 0,1, \ldots , m)$.

Thus a cubic spline consists of a set of cubic polynomial arcs joined smoothly end to end. The smoothness consists of continuity up to the second derivative, but the third derivative will, in general, have a discontinuity at each of the points $x = x_i$ $(i = 0, 1, \ldots , m)$.

If we represent $S''(x_i)$, the second derivative of the cubic spline at the points $x = x_i$, by M_i then from Ahlberg, Nilson and Walsh (1967) we have from the linearity of the second derivative on the interval $\left[x_{i-1}, x_i\right]$

$$S''(x) = M_{i-1}\left(\frac{x_i - x}{h_i}\right) + M_i \left(\frac{x - x_{i-1}}{h_i}\right) \tag{2.2.1}$$

where $h_i = x_i - x_{i-1}$. Integrating twice and evaluating the constants of integration, we obtain

$$S(x) = M_{i-1}\frac{(x_i - x)^3}{6h_i} + M_i \frac{(x - x_{i-1})^3}{6h_i} + \left(f(x_{i-1}) - \frac{h_i^2 M_{i-1}}{6}\right)\frac{(x_i - x)}{h_i} + \left(f(x_i) - \frac{h_i^2 M_i}{6}\right)\frac{(x - x_{i-1})}{h_i} \quad (i = 1,2, \ldots,m) \tag{2.2.2}$$

$$S'(x) = -M_{i-1}\frac{(x_i - x)^2}{2h_i} + M_i \frac{(x - x_{i-1})^2}{2h_i} + \frac{f(x_i) - f(x_{i-1})}{h_i} - \left(\frac{M_i - M_{i-1}}{6}\right)h_i \quad (i = 1,2, \ldots,m) \tag{2.2.3}$$

From (2.2.3) we have the expressions for the one-sided limits of the derivatives

$$S'(x_i -) = \frac{h_i}{6} M_{i-1} + \frac{h_i}{3} M_i + \frac{f(x_i) - f(x_{i-1})}{h_i} \quad (i = 1,2, \ldots, m) \tag{2.2.4}$$

$$S'(x_i +) = -\frac{h_{i+1}}{3} M_i - \frac{h_{i+1}}{6} M_{i+1} + \frac{f(x_{i+1})-f(x_i)}{h_{i+1}} \quad (i = 0,1, \ldots, (m-1)) \tag{2.2.5}$$

From the continuity of $S'(x)$ at the points $x = x_i$ $(i = 0,1, \ldots, m)$ we can equate (2.2.4) and (2.2.5) thereby obtaining the expression

$$\frac{h_i}{6} M_{i-1} + \left(\frac{h_i + h_{i+1}}{3}\right) M_i + \frac{h_{i+1}}{6} M_{i+1} = \frac{f(x_{i+1}) - f(x_i)}{h_{i+1}} - \frac{f(x_i) - f(x_{i-1})}{h_i}$$

$$(i = 1,2, \ldots, (m-1)) \tag{2.2.6}$$

Given the function values $f(x_i)$ $(i = 0,1, \ldots,m)$ we require two additional conditions in order to express (2.2.6) in tri-diagonal form. As shown by Curtis (1970) there are three choices available for these two extra conditions, namely

(i) $S''(x_i) = 0 \quad (i = 0,m)$ (2.2.7)

(ii) $S'(x_i) = f'(x_i) \quad (i = 0,m)$ (2.2.8)

(iii) $\left[S(x) - f(x)\right]_{x=\frac{1}{2}(x_{i-1}+x_i)} = \left[S(x) - f(x)\right]_{x=\frac{1}{2}(x_i+x_{i+1})}$

$(i = 1, m-1)$ (2.2.9)

Each of these conditions can easily be applied in practice although the second used by Birkhoff and de Boor (1965), and the third, used by Curtis and Powell (1967), are most accurate and therefore most popular. However in the solution of partial differential equations these additional "end conditions" will be given by the boundary conditions associated with the problem.

Once the choice of "end conditions" has been decided upon, substitution of these into (2.2.6) gives a system of (m-1) equations which are linearly independent, tri-diagonal and diagonally dominant. They can therefore easily be solved for the values M_i $(i = 1, 2, \ldots, (m-1))$, the satisfaction of the "end conditions" then giving M_o and M_m. The spline function S(x) can then be obtained from (2.2.2).

2.3 Multiple Knots

The condition (2.1.1) can be relaxed to the form (see for example Cox (1975))

$$a = x_o \leqslant x_1 \leqslant \cdots \leqslant x_m = b \tag{2.3.1}$$

If in general the points x_r to x_{r+k} are such that

$$x_r = x_{r+1} = \ldots = x_{r+k-1} = x_{r+k} \tag{2.3.2}$$

then these coincident knots can be regarded as a single knot of multiplicity k. In a case such as this the spline function S(x) of degree n has n - k - 1 continuous derivatives instead of the original n -1 as stated in section 2.1. For example, at a knot of multiplicity 2 the cubic spline function is itself continuous but has no continuous derivatives. Similarly, for a cubic spline which has a multiple knot of degree 3, the function S(x) has a jump discontinuity and it is therefore of no use to consider knots of multiplicity which are greater than the degree of the spline function itself.

2.4 Cubic B-Splines

The use of B-splines or fundamental splines, for the case of equally spaced knots, was first introduced by Schoenberg (1946). These B-splines can be shown to be non-zero only over a small number of intervals between successive knots.

For the particular case of the cubic B-spline we can define this as being a cubic spline which is zero everywhere except over four adjacent intervals between knots. Using the notation employed in section 2.2 this means that the cubic spline function with knots x_{i-4}, x_{i-3}, x_{i-2}, x_{i-1} and x_i is zero everywhere in the range $a \leqslant x \leqslant b$ except within the interval $x_{i-4} \leqslant x \leqslant x_i$.

These cubic B-splines have been shown to be useful for fitting cubic spline surfaces to general sets of data by Hayes and Halliday (1974).

CHAPTER 3

Hyperbolic Partial Differential Equations with Constant Coefficients

3.1 Simple Initial and Boundary Conditions

Suppose that $u(x,t)$ satisfies the second order hyperbolic partial differential equation

$$\frac{\partial^2 u}{\partial t^2} = a \frac{\partial^2 u}{\partial x^2} + b \frac{\partial u}{\partial x} + cu \quad (0 \leqslant x \leqslant 1, \ t > 0) \tag{3.1.1}$$

where a,b,c are constants and $a > 0$. We consider (3.1.1) to be subject to the following boundary conditions

$$u(0,t) = f_1(t) \quad ; \quad u(1,t) = f_2(t) \tag{3.1.2}$$

and initial conditions

$$u(x,0) = g_1(x) \quad ; \quad \frac{\partial u}{\partial t}(x,0) = g_2(x) \tag{3.1.3}$$

where $f_1(t)$, $f_2(t)$, $g_1(x)$ and $g_2(x)$ are known functions. These conditions are only of a simple form but, as is shown in section 3.9, they can be of a more general nature and the following methods of solution will still apply.

3.2 Cubic Spline Finite Difference Scheme

To obtain solutions to (3.1.1) we will initially consider the interval $0 \leqslant x \leqslant 1$ subdivided by equally spaced knots where the step length between successive knots is h, so that $x_i = ih$ $(i = 0,1, \ldots, N)$. We now replace the time derivative in (3.1.1) by a finite difference approximation and the space derivative by a cubic spline thus obtaining, at the point (ih,jk)

$$\frac{U_{i,j-1} - 2U_{i,j} + U_{i,j+1}}{k^2} = a\left\{\theta M_{i,j-1} + (1-2\theta)M_{i,j} + \theta M_{i,j+1}\right\}$$

$$+ \; b\left\{\theta L_{i,j-1} + (1-2\theta)L_{i,j} + \theta L_{i,j+1}\right\}$$

$$+ \; c\left\{\theta U_{i,j-1} + (1-2\theta)U_{i,j} + \theta U_{i,j+1}\right\}$$

(3.2.1)

$(i = 0,1, \ldots , N \; ; \; j = 1,2, \ldots \; ; \; Nh = 1)$

where $L_{i,j} = S'_j(x_i)$; $M_{i,j} = S''_j\,(x_i)$; $S_j(x)$ denoting the cubic spline interpolating the values $U_{i,j}$ on the j^{th} time level.

With constant step length h the spline function (2.2.2), for the j^{th} time line within the interval $\left[x_{i-1}, x_i\right]$, becomes

$$S_j(x) = M_{i-1,j}\frac{(x_i-x)^3}{6h} + M_{i,j}\frac{(x-x_{i-1})^3}{6h} + \left(U_{i-1,j} - \frac{h^2 M_{i-1,j}}{6}\right)\frac{(x_i-x)}{h}$$

$$+ \left(U_{i,j} - \frac{h^2}{6} M_{i,j}\right)\frac{(x-x_{i-1})}{h} \quad (i = 1,2,\ldots,N) \qquad (3.2.2)$$

and thus the continuity of the ~~second~~ first derivative of the cubic spline with equally spaced knots on the j^{th} time line gives

$$\frac{1}{6} M_{i-1,j} + \frac{2}{3} M_{i,j} + \frac{1}{6} M_{i+1,j} = \frac{U_{i-1,j} - 2U_{i,j} + U_{i+1,j}}{h^2} \qquad (3.2.3)$$

$(i = 1,2, \ldots, (N-1))$

This equation also holds on the $(j-1)^{th}$ time line

$$\frac{1}{6} M_{i-1,j-1} + \frac{2}{3} M_{i,j-1} + \frac{1}{6} M_{i+1,j-1} = \frac{U_{i-1,j-1} - 2U_{i,j-1} + U_{i+1,j-1}}{h^2}$$

$(i = 1,2, \ldots, (N-1))$ (3.2.4)

and on the $(j+1)^{th}$ time line

$$\frac{1}{6} M_{i-1,j+1} + \frac{2}{3} M_{i,j+1} + \frac{1}{6} M_{i+1,j+1} = \frac{U_{i-1,j+1} - 2U_{i,j+1} + U_{i+1,j+1}}{h^2}$$

$$(i = 1,2, \ldots,(N-1)) \qquad (3.2.5)$$

We now wish to form a relationship similar to (3.2.3) incorporating the first derivative of the cubic spline. Writing $L_{i,j} = s_j'(x_i)$, then (2.2.4) and (2.2.5) respectively become, on the j^{th} time line

$$L_{i,j} = S_j'(x_i-) = \frac{h}{6} M_{i-1,j} + \frac{h}{3} M_{i,j} + \frac{U_{i,j} - U_{i-1,j}}{h}$$

$$(i = 1,2, \ldots,N) \qquad (3.2.6)$$

$$L_{i,j} = S_j'(x_i+) = -\frac{h}{3} M_{i,j} - \frac{h}{6} M_{i+1,j} + \frac{U_{i+1,j} - U_{i,j}}{h}$$

$$(i = 0,1,\ldots, (N-1)) \qquad (3.2.7)$$

From (3.2.6) we have

$$L_{i+1,j} = \frac{h}{6} M_{i,j} + \frac{h}{3} M_{i+1,j} + \frac{u_{i+1,j} - U_{i,j}}{h}$$

$$(i = 0,1, \ldots, (N-1)) \qquad (3.2.8)$$

and from (3.2.7)

$$L_{i-1,j} = -\frac{h}{3} M_{i-1,j} - \frac{h}{6} M_{i,j} + \frac{U_{i,j} - U_{i-1,j}}{h}$$

$$(i = 1,2, \ldots, N) \qquad (3.2.9)$$

To obtain the required relationship, equations (3.2.6) and (3.2.7) are now added together and the result added to half the sum of (3.2.8) and (3.2.9), thus giving on the j^{th} time line

$$\frac{1}{6} L_{i-1,j} + \frac{2}{3} L_{i,j} + \frac{1}{6} L_{i+1,j} = \frac{U_{i+1,j} - U_{i-1,j}}{2h} \qquad (3.2.10)$$

$$(i = 1,2, \ldots, (N-1))$$

As with the expression for the $M_{i,j}$ values, (3.2.10) also holds on the $(j-1)^{th}$ time line

$$\frac{1}{6} L_{i-1,j-1} + \frac{2}{3} L_{i,j-1} + \frac{1}{6} L_{i+1,j-1} = \frac{U_{i+1,j-1} - U_{i-1,j-1}}{2h}$$

$$(i = 1,2, \ldots, (N-1)) \qquad (3.2.11)$$

and on the $(j+1)^{th}$ time line

$$\frac{1}{6} L_{i-1,j+1} + \frac{2}{3} L_{i,j+1} + \frac{1}{6} L_{i+1,j+1} = \frac{U_{i+1,j+1} - U_{i-1,j+1}}{2h}$$

$$(i = 1,2, \ldots, (N-1)) \qquad (3.2.12)$$

To obtain the finite difference scheme incorporating splines we now combine the above relationships for the $M_{i,j}$ and $L_{i,j}$ values by performing the following operations:-

(i) Equation (3.2.3) is multiplied throughout by $a(1-2\theta)$

(ii) Equations (3.2.4) and (3.2.5) are summed and the resulting equation is multiplied throughout by $a\theta$.

(iii) Equation (3.2.10) is multiplied throughout by $b(1-2\theta)$

(iv) Equations (3.2.11) and (3.2.12) are summed and the resulting equation is multiplied throughout by $b\theta$.

The expressions produced in (i) - (iv) are now summed together giving the following result -

$$a(1-2\theta)\left(\frac{1}{6}M_{i-1,j}+\frac{2}{3}M_{i,j}+\frac{1}{6}M_{i+1,j}\right)$$

$$+\ a\theta\left(\frac{1}{6}M_{i-1,j-1}+\frac{2}{3}M_{i,j-1}+\frac{1}{6}M_{i+1,j-1}+\frac{1}{6}M_{i-1,j+1}+\frac{2}{3}M_{i,j+1}\right.$$

$$\left.+\ \frac{1}{6}M_{i+1,j+1}\right)$$

$$+\ b(1-2\theta)\left(\frac{1}{6}L_{i-1,j}+\frac{2}{3}L_{i,j}+\frac{1}{6}L_{i+1,j}\right)$$

$$+\ b\theta\left(\frac{1}{6}L_{i-1,j-1}+\frac{2}{3}L_{i,j-1}+\frac{1}{6}L_{i+1,j-1}+\frac{1}{6}L_{i-1,j+1}+\frac{2}{3}L_{i,j+1}\right.$$

$$\left.+\ \frac{1}{6}L_{i+1,j+1}\right)$$

$$=\ a(1-2\theta)\left(\frac{U_{i-1,j}-2U_{i,j}+U_{i+1,j}}{h^2}\right)$$

$$+\ a\theta\left(\frac{U_{i-1,j-1}-2U_{i,j-1}+U_{i+1,j-1}}{h^2}+\frac{U_{i-1,j+1}-2U_{i,j+1}+U_{i+1,j+1}}{h^2}\right)$$

$$+\ b(1-2\theta)\left(\frac{U_{i+1,j}-U_{i-1,j}}{2h}\right)+b\theta\left(\frac{U_{i+1,j-1}-U_{i-1,j-1}}{2h}\right.$$

$$\left.+\ \frac{U_{i+1,j+1}-U_{i-1,j+1}}{2h}\right) \qquad (3.2.13)$$

We now eliminate the M_{ij} and $L_{i,j}$ values from the left hand side of (3.2.13) using (3.2.1). The following three time level finite difference scheme incorporating splines then results,

$$(1-\beta_1\theta)U_{i-1,j+1}+4(1+\beta_2\theta)U_{i,j+1}+(1-\beta_3\theta)U_{i+1,j+1}$$

$$=\{2+\beta_1(1-2\theta)\}U_{i-1,j}+4\{2-\beta_2(1-2\theta)\}U_{i,j}+\{2+\beta_3(1-2\theta)\}U_{i+1,j}$$

$$-\ (1-\beta_1\theta)U_{i-1,j-1}-4(1+\beta_2\theta)U_{i,j-1}-(1-\beta_3\theta)U_{i+1,j-1} \qquad (3.2.14)$$

$$(i = 1,2,\ \ldots,\ (N-1))$$

where $\beta_1 = 6ar^2 - 3brk + ck^2$; $\beta_2 = 3ar^2 - ck^2$; $\beta_3 = 6ar^2 + 3brk + ck^2$ and the mesh ratio $r = k/h$.

This difference scheme is implicit in nature and has three unknown values on the advanced time line $t = (j+1)k$. It can therefore be expressed in the form of a tridiagonal system of equations and solved using an algorithm based on the Gaussian elimination process. We will describe the solution of these tri-diagonal systems in the [redacted] appendi[redacted].

3.3 Truncation Error

To obtain the truncation error associated with (3.2.14) we first rearrange the scheme into the form

$$
\begin{aligned}
&(1-\beta_1\theta)\ (U_{i-1,j+1} + U_{i-1,j-1}) + 4\ (1 + \beta_2\theta)\ (U_{i,j+1} + U_{i,j-1}) \\
&\qquad + (1 - \beta_3\theta)\ (U_{i+1,j+1} + U_{i+1,j-1}) \\
&= \{2 + \beta_1(1-2\theta)\}U_{i-1,j} + 4\{2 - \beta_2(1-2\theta)\}U_{i,j} \\
&\qquad + \{2 + \beta_3(1-2\theta)\}U_{i+1,j}
\end{aligned}
\qquad (3.3.1)
$$

We now expand each term of (3.3.1) about the mesh point (ih, jk) using Taylor series approximations. The following expression, to fourth order, thus results

$$(1-\beta_1\theta)\left[2\left\{U_{i,j}-h\left(\frac{\partial u}{\partial x}\right)_{i,j}+\frac{h^2}{2!}\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j}-\frac{h^3}{3!}\left(\frac{\partial^3 u}{\partial x^3}\right)_{i,j}+\frac{h^4}{4!}\left(\frac{\partial^4 u}{\partial x^4}\right)_{i,j}\right\}\right.$$

$$\left.+k^2\left\{\left(\frac{\partial^2 u}{\partial t^2}\right)_{i,j}-h\left(\frac{\partial^3 u}{\partial x\partial t^2}\right)_{i,j}+\frac{h^2}{2!}\left(\frac{\partial^4 u}{\partial x^2\partial t^2}\right)_{i,j}\right\}+\frac{k^4}{12}\left(\frac{\partial^4 u}{\partial t^4}\right)_{i,j}\right]$$

$$+\ 4(1+\beta_2\theta)\left[2U_{i,j}+k^2\left(\frac{\partial^2 u}{\partial t^2}\right)_{i,j}+\frac{k^4}{12}\left(\frac{\partial^4 u}{\partial t^4}\right)_{i,j}\right]$$

$$+\ (1-\beta_3\theta)\left[2\left\{U_{i,j}+h\left(\frac{\partial u}{\partial x}\right)_{i,j}+\frac{h^2}{2!}\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j}+\frac{h^3}{3!}\left(\frac{\partial^3 u}{\partial x^3}\right)_{i,j}+\frac{h^4}{4!}\left(\frac{\partial^4 u}{\partial x^4}\right)_{i,j}\right\}\right.$$

$$\left.+k^2\left\{\left(\frac{\partial^2 u}{\partial t^2}\right)_{i,j}+h\left(\frac{\partial^3 u}{\partial x\partial t^2}\right)_{i,j}+\frac{h^2}{2!}\left(\frac{\partial^4 u}{\partial x^2\partial t^2}\right)_{i,j}\right\}+\frac{k^4}{12}\left(\frac{\partial^4 u}{\partial t^4}\right)_{i,j}\right]$$

$$=\{2+\beta_1(1-2\theta)\}\left[U_{i,j}-h\left(\frac{\partial u}{\partial x}\right)_{i,j}+\frac{h^2}{2!}\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j}-\frac{h^3}{3!}\left(\frac{\partial^3 u}{\partial x^3}\right)_{i,j}+\frac{h^4}{4!}\left(\frac{\partial^4 u}{\partial x^4}\right)_{i,j}\right]$$

$$+\ 4\{2-\beta_2(1-2\theta)\}\ U_{i,j}$$

$$+\{2+\beta_3(1-2\theta)\}\left[U_{i,j}+h\left(\frac{\partial u}{\partial x}\right)_{i,j}+\frac{h^2}{2!}\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j}+\frac{h^3}{3!}\left(\frac{\partial^3 u}{\partial x^3}\right)_{i,j}+\frac{h^4}{4!}\left(\frac{\partial^4 u}{\partial x^4}\right)_{i,j}\right]$$

(3.3.2)

Since the partial derivatives with respect to t in (3.3.2) are of even order only we can use (3.1.1) to replace them by partial derivatives with respect to x. After appropriate rearrangement we obtain the following truncation error for (3.2.14)

$$k^2h^2\Big[c^2r^2f(\theta)u + 2bcr^2f(\theta)\frac{\partial u}{\partial x}$$

$$+\left\{r^2(2ac+b^2)f(\theta) - \frac{1}{6}c^2k^2\theta\right\}\frac{\partial^2 u}{\partial x^2}$$

$$+\left\{2abr^2f(\theta) - \frac{1}{6}bck^2\theta\right\}\frac{\partial^3 u}{\partial x^3}$$

$$+\left\{a^2r^2f(\theta) + \frac{1}{12}a - \frac{1}{72}ch^2 - \frac{1}{6}ack^2\theta\right\}\frac{\partial^4 u}{\partial x^4}\Big] \qquad (3.3.3)$$

$$\text{where}\quad f(\theta) = \frac{1}{12}(1 - ck^2\theta) - \theta \qquad (3.3.4)$$

3.4 Stability Analysis

To examine the stability of (3.2.14) we use the well known von Neumann method. Initially we replace $U_{i,j}$ by $U_{m,n}$ in (3.2.14) so as to avoid any conflict between variables and then look at solutions of (3.2.14) which have the form (see Mitchell (1969))

$$U_{m,n} = e^{im\gamma}e^{in\lambda} \qquad (i^2 = -1) \qquad (3.4.1)$$

where γ is an arbitrary real number and λ is a complex parameter to be determined. Substituting (3.4.1) into (3.2.14) and dividing by $e^{im\gamma}e^{in\lambda}$, we obtain

$$(1-\beta_1\theta)e^{-i\gamma}(e^{i\lambda}+e^{-i\lambda})+4(1+\beta_2\theta)(e^{i\lambda}+e^{-i\lambda})+(1-\beta_3\theta)e^{i\gamma}(e^{i\lambda}+e^{-i\lambda})$$

$$= \{2+\beta_1(1-2\theta)\}e^{-i\gamma}+4\{2-\beta_2(1-2\theta)\} + \{2+\beta_3(1-2\theta)\}e^{i\gamma}$$

Due to the expressions

$$2\text{Cos}\ \lambda = e^{i\lambda} + e^{-i\lambda}$$

and $\text{Cos}\ \lambda = 1 - 2\text{Sin}^2\frac{\lambda}{2}$

we can replace $(e^{i\lambda}+e^{-i\lambda})$ in terms of $\text{Sin}^2\frac{\lambda}{2}$, the result being

$$2(1-2\text{Sin}^2\frac{\lambda}{2})\left[(1-\beta_1\theta)e^{-i\gamma}+4(1+\beta_2\theta) + (1-\beta_3\theta)e^{i\gamma}\right]$$

$$= \{2+\beta_1(1-2\theta)\}e^{-i\gamma}+4\{2-\beta_2(1-2\theta)\} + \{2+\beta_3(1-2\theta)\}e^{i\gamma} \qquad (3.4.2)$$

If we now replace the factors β_1, β_2 and β_3 by their full expressions as given in (3.2.14) and let h and k tend to zero in such a way as r remains fixed then we obtain

$$\text{Sin}^2\frac{\lambda}{2} = \frac{3ar^2\text{Sin}^2\gamma/2}{3-2(1-6ar^2\theta)\text{Sin}^2\gamma/2} \qquad (3.4.3)$$

For stability we require $U_{m,n}$ to be bounded. Therefore λ must be real and hence the condition

$$0 \leqslant \text{Sin}^2\frac{\lambda}{2} \leqslant 1 \qquad (3.4.4)$$

must be satisfied. Substituting (3.4.3) into (3.4.4) and noting that $\text{Sin}^2\gamma/2\ \varepsilon[0,1]$ we have

$$r^2 \leqslant \{3a(1-4\theta)\}^{-1} \quad .$$

Hence the following conditions governing the stability of (3.2.14) are obtained

(a) if $\theta \geqslant \frac{1}{4}$ it is unconditionally stable

(b) if $\theta < \frac{1}{4}$ it is stable when

$$r \leqslant \{3a(1-4\theta)\}^{-\frac{1}{2}} \quad . \tag{3.4.5}$$

It should be noted that the stability condition (3.4.5) does not depend on the coefficients b or c. Fox (1962) indicated that when solving the general parabolic partial differential equation

$$\frac{\partial u}{\partial t} = a \frac{\partial^2 u}{\partial x^2} + b \frac{\partial u}{\partial x} + cu + d \tag{3.4.6}$$

where a, b, c and d are functions of x and t only, then it is reasonable to assume that the presence of the lower order terms, u and $\partial u/\partial x$, have no great effect on the stability condition for a particular finite difference scheme. This assumption is borne out in two examples given by Richtmyer and Morton (1967). In the hyperbolic equation (3.1.1) it appears that this conjecture is also true since the quantities u and $\partial u/\partial x$ are also effectively multiplied by h^2 and h respectively as was the case in parabolic equations. However, although the stability condition is practically unaffected by these lower order terms, it must be noted that they may require a smaller value of k to be used. For example (see Richtmyer and Morton (1967), page 195) if a large value of c is present in (3.1.1) then a small value of k, and hence k^2, must be used to counteract this.

3.5 Comparison with a well known finite difference scheme

As a comparison to the method incorporating cubic splines, which was derived in section 3.2, we now consider the well known implicit finite difference scheme as an approximation to (3.1.1). This method was described by Mitchell (1969) for the solution of the wave equation and is easily extended to our more general case, giving at (ih, jk)

$$\begin{aligned} u_{i,j-1}-2u_{i,j}+u_{i,j+1} = ar^2\Big\{&\alpha(u_{i-1,j-1}-2u_{i,j-1}+u_{i+1,j-1}) \\ &+(1-2\alpha)(u_{i-1,j}-2u_{i,j}+u_{i+1,j}) \\ &+\alpha(u_{i-1,j+1}-2u_{i,j+1}+u_{i+1,j+1})\Big\} \\ +\frac{bhr^2}{2}\Big\{&\alpha(u_{i+1,j-1}-u_{i-1,j-1}) \\ &+(1-2\alpha)(u_{i+1,j}-u_{i-1,j}) \\ &+\alpha(u_{i+1,j+1}-u_{i-1,j+1})\Big\} \\ +ck^2\{&\alpha u_{i,j-1}+(1-2\alpha)u_{i,j}+\alpha u_{i,j+1}\} \end{aligned} \tag{3.5.1}$$

$(i = 1,2, \ldots, (N-1) \; ; \; j = 1,2, \ldots \; ; \; Nh = 1)$

where $r = k/h$. Rearranging this into a form similar to (3.2.14) we have

$$\alpha\left(-ar^2+\frac{bhr^2}{2}\right)u_{i-1,j+1}+\{1+\alpha(2ar^2-ck^2)\}u_{i,j+1}+\alpha\left(-ar^2-\frac{bhr^2}{2}\right)u_{i+1,j+1}$$

$$= \left\{ \left(ar^2 - \frac{bhr^2}{2} \right)(1-2\alpha) \right\} u_{i-1,j} + \{2+(-2ar^2+ck^2)(1-2\alpha)\} u_{i,j}$$

$$+ \left\{ \left(ar^2 + \frac{bhr^2}{2} \right)(1-2\alpha) \right\} u_{i+1,j}$$

$$-\alpha\left(-ar^2 + \frac{bhr^2}{2} \right) u_{i-1,j-1} - \{1+\alpha(2ar^2-ck^2)\} u_{i,j-1} - \alpha\left(-ar^2 - \frac{bhr^2}{2} \right) u_{i+1,j-1}$$

$$(i = 1,2, \ldots, (N-1)). \qquad (3.5.2)$$

The choice of $\alpha=0$ in (3.5.2) gives rise to the usual explicit approximation to (3.1.1). Also, as shown by Raggett and Wilson (1974), when solving the wave equation ($a=1$, $b=c=0$ in (3.1.1)) the approximation (3.2.14) reduces to (3.5.2) when $\theta = \alpha + \frac{1}{6r^2}$.

3.6 Truncation Error and Stability conditions for (3.5.2)

The truncation error for the scheme (3.5.2) is obtained in the same manner as that of (3.3.3) and is seen to be, at (ih,jk)

$$k^2h^2\left[c^2r^2f(\alpha)u + 2bcr^2f(\alpha)\frac{\partial u}{\partial x} + r^2(2ac+b^2)f(\alpha)\frac{\partial^2 u}{\partial x^2} \right.$$

$$\left. + \left\{ 2abr^2f(\alpha) - \frac{1}{6}b \right\}\frac{\partial^3 u}{\partial x^3} + \left\{ a^2r^2f(\alpha) - \frac{1}{12}a \right\}\frac{\partial^4 u}{\partial x^4} \right] \qquad (3.6.1)$$

where the function f is given by (3.3.4)

The nature of the two truncation errors (3.3.3) and (3.6.1) are very similar, there being a number of like terms when the parameters α and θ are chosen equal. The two truncation errors are $O(k^2h^2)$ and both may be considerably simplified by choosing the function f to equal zero. Thus from (3.3.4), θ and α are chosen such that

$$\theta = \alpha = \frac{1}{12 + ck^2} \qquad (3.6.2)$$

Methods of choosing α and θ in order to increase the accuracy of the two schemes will be discussed in the later chapter covering case studies.

A detailed von Neumann stability analysis of the scheme (3.5.2) gives the following conditions for stability:

(a) if $\alpha \geqslant \frac{1}{4}$ it is unconditionally stable

(b) if $\alpha < \frac{1}{4}$ it is stable when

$$r \leqslant \{a(1-4\alpha)\}^{-\frac{1}{2}} . \qquad (3.6.3)$$

In view of the fact that the two truncation errors (3.3.4) and (3.6.1) are of the same order and that the stability conditions are very alike for both schemes (3.2.14) and (3.5.2), then both schemes seem equally viable for obtaining the solution to (3.1.1). Both schemes are three time level and require the evaluation of a tri-diagonal system of equations. It is therefore acceptable to assume that the computing times for both schemes will be very comparable. Relative advantages of the two methods will be discussed in the later case studies although it is clear that an additional advantage of using the cubic spline scheme is that it is possible to obtain a spline function from (3.2.2). This can then be used to obtain solutions at points intermediate to the mesh points on any particular time line. The procedure for obtaining these intermediate solutions is included in the following section.

3.7 Numerical Procedure

The method for obtaining cubic spline solutions at points intermediate to mesh points is similar to that given by Raggett and Wilson (1974) for solving the wave equation. However for the general hyperbolic partial differential equation (3.1.1), the procedure is considerably more complicated by the presence of the $L_{i,j}$'s in (3.2.1). The steps in the numerical procedure are as follows:

(i) Evaluation of the $M_{i,o}$, $i = 0,1, \ldots, N$.

Using the function value initial condition given in (3.1.3) equation (3.2.3) with $j = 0$ gives

$$\frac{1}{6} M_{i-1,o} + \frac{2}{3} M_{i,o} + \frac{1}{6} M_{i+1,o} = \frac{g_1(x_{i-1}) - 2g_1(x_i) + g_1(x_{i+1})}{h^2} \qquad (3.7.1)$$

$$(i = 1,2, \ldots, (N-1))$$

which is a tri-diagonal system with (N-1) equations and (N+1) unknowns. As indicated in section 2.2 we require two "end conditions" in order to obtain the $M_{i,o}$ values from (3.7.1). These additional conditions are found by putting $\theta = j = 0$ in (3.2.1) and by making use of central difference approximations to represent the derivative initial condition in (3.1.3). Thus on the boundaries $x = 0$ and $x = 1$ we have

$$k^2(aM_{o,o} + bL_{o,o}) = 2(f_1(k) - f_1(0) - kg_2(0)) - ck^2 f_1(0) \qquad (3.7.2)$$

and

$$k^2(aM_{N,o}+bL_{N,o}) = 2(f_2(k)-f_2(0)-kg_2(1))-ck^2f_2(0). \quad (3.7.3)$$

Similarly, from (3.2.7)

$$L_{o,o} = -\frac{h}{3}M_{o,o} - \frac{h}{6}M_{1,o} + \frac{g_1(h)-g_1(0)}{h} \quad (3.7.4)$$

and from (3.2.6)

$$L_{N,o} = \frac{h}{3}M_{N,o} + \frac{h}{6}M_{N-1,o} + \frac{g_1(1)-g_1(x_{N-1})}{h} \quad (3.7.5)$$

Equations (3.7.1) - (3.7.5) are (N+3) in number and are solvable for $M_{i,o}$, $i = 0, 1, \ldots, N$ along with $L_{o,o}$ and $L_{N,o}$.

(ii) Evaluation of the $L_{i,o}$, $i = 1,2, \ldots, (N-1)$

Averaging (3.2.6) and (3.2.7) gives

$$L_{i,j} = \frac{h}{12}(M_{i-1,j} - M_{i+1,j}) + \frac{U_{i+1,j} - U_{i-1,j}}{2h} \quad (3.7.6)$$

$$(i = 1,2, \ldots, (N-1))$$

Having obtained $M_{i,o}$ $(i = 0,1, \ldots,N)$ in the previous step, we now put $j = 0$ in (3.7.6) giving the required values of the $L_{i,o}$'s.

(iii) Cubic Spline solution on the initial time line.

Using the $M_{i,o}$ values found in step (i) equation (3.2.2) is used directly to evaluate spline solutions at any point on $t = 0$.

(iv) Mesh point solutions on time line $t = k$.

Putting $j = 0$ in (3.2.14) and using a central difference approximation to represent the derivative initial condition in (3.1.3) gives the set of (N-1) equations

$$2(1-\beta_1\theta)u_{i-1,1} + 8(1+\beta_2\theta)u_{i,1} + 2(1-\beta_3\theta)u_{i+1,1}$$

$$= \{2 + \beta_1(1-2\theta)\}g_1(x_{i-1}) + 4\{2-\beta_2(1-2\theta)\}g_1(x_i)$$

$$+ \{2+\beta_3(1-2\theta)\}g_1(x_{i+1}) + 2k(1-\beta_1\theta)g_2(x_{i-1})$$

$$+ 8k(1-\beta_2\theta)g_2(x_i) + 2k(1-\beta_3\theta)g_2(x_{i+1}) \quad . \tag{3.7.7}$$

$$(i = 1,2, \ldots,(N-1))$$

Since $u_{o,1} = f_1(k)$ and $u_{N,1} = f_2(k)$ then these with (3.7.7) give an easily solvable tri-diagonal system for the unknown mesh points values $u_{i,1}$, $i = 1,2, \ldots,(N-1)$.

(v) Evaluation of the $M_{i,1}$, $i = 0,1, \ldots, N$.

These values are evaluated in a similar manner to the correspon-ding $M_{i,o}$ values found in step (i). From (3.2.3) with $j = 1$ we have

$$\frac{1}{6} M_{i-1,1} + \frac{2}{3} M_{i,1} + \frac{1}{6} M_{i+1,1} = \frac{u_{i-1,1} - 2u_{i,1} + u_{i+1,1}}{h^2} \tag{3.7.8}$$

$$(i = 1,2, \ldots,(N-1))$$

As in step (i) we require additional conditions in order to obtain the required values. These are obtained in a similar manner, giving the following results

$$k^2(aM_{o,1}+bL_{o,1}) = f_1(0)-2f_1(k)+f_1(2k)-ck^2f_1(k) \tag{3.7.9}$$

$$k^2(aM_{N,1}+bL_{N,1}) = f_2(0)-2f_2(k)+f_2(2k)-ck^2f_2(k) \tag{3.7.10}$$

$$L_{o,1} = -\frac{h}{3}M_{o,1} - \frac{h}{6}M_{1,1} + \frac{u_{1,1} - f_1(k)}{h} \tag{3.7.11}$$

$$L_{N,1} = \frac{h}{3}M_{N,1} + \frac{h}{6}M_{N-1,1} + \frac{f_2(k) - u_{N-1,1}}{h} \tag{3.7.12}$$

Again we have (N+3) equations which are solved for $M_{i,1}$, $(i = 0,1, \ldots,N)$, $L_{o,1}$ and $L_{N,1}$.

(vi) Evaluation of the $L_{i,1}$, $i = 1,2, \ldots,(N-1)$.

These values are evaluated in a similar manner to the corresponding $L_{i,o}$ values found in step (ii). Thus equation (3.7.6) is used with $j = 1$, the $M_{i,1}$ values required being known from step (v).

(vii) Cubic spline solution on $t = k$.

As in step (iii) equation (3.2.2) is used directly, with $j = 1$ in this case.

(viii) The General scheme.

A full spline solution over two successive time lines is now known. Suppose the scheme is fully developed and that solutions are known along the $(j-1)^{th}$ and j^{th} time lines. Then equation (3.2.14) along with the boundary conditions (3.1.2) constitute a tri-diagonal system of linear equations which is solvable for the mesh point values $u_{i,j+1}$, $i = 1,2, \ldots,(N-1)$.

For the evaluation of solutions at points intermediate to mesh points on the $(j+1)^{th}$ time line, the following procedure is adopted. Using the known values of $L_{i,j}$, $M_{i,j}$, $L_{i,j-1}$ and

$M_{i,j-1}$, $i = 0,1, \ldots, N$ (3.2.1) gives (N+1) equations in the 2(N+1) unknowns $L_{i,j+1}$, $M_{i,j+1}$, $i = 0,1, \ldots, N$. A further two equations are obtained from (3.2.7) with i = 0 and from (3.2.6) with i = N. The remaining (N-1) equations are obtained from (3.7.6) now evaluated at the $(j+1)^{th}$ time line. Equation (3.2.2) then give any required intermediate values on this new time line.

3.8 Evaluation of Truncation Errors

To estimate the accuracies of the finite difference formulae (3.2.14) and (3.5.2) we will obtain numerical values of their respective truncation errors. The derivative terms within the truncation errors are evaluated using relevant forward and backward difference approximations on the boundaries and central difference approximations for the internal mesh points. For example

$$\left(\frac{\partial^4 u}{\partial x^4}\right)_{i,j} \simeq \frac{1}{h^4}\left(U_{i+2,j}-4U_{i+1,j}+6U_{i,j}-4U_{i-1,j}+U_{i-2,j}\right) \qquad (3.8.1)$$

and using mean central differences, to avoid having to evaluate solutions at half-way points between mesh points, we have

$$\left(\frac{\partial^3 u}{\partial x^3}\right)_{i,j} \simeq \frac{1}{2h^3}\left(U_{i+2,j}-2U_{i+1,j}+2U_{i-1,j}-U_{i-2,j}\right). \qquad (3.8.2)$$

3.9 Derivative Boundary Conditions

The boundary conditions given in (3.1.2) can be generalised to the form

$$\alpha_1(t)U(0,t) + \alpha_2(t)\frac{\partial u}{\partial x}(0,t) = f_1(t) \qquad (3.9.1)$$

$$\alpha_3(t)U(1,t) + \alpha_4(t)\,\frac{\partial u}{\partial x}(1,t) = f_2(t) \tag{3.9.2}$$

where $\alpha_1(t)$, $\alpha_2(t)$, $\alpha_3(t)$, $\alpha_4(t)$, $f_1(t)$, $f_2(t)$ are known functions which are continuous and bounded as $t \to \infty$. The implementation of (3.9.1) and (3.9.2) is reasonably simple and, for convenience, we will illustrate the procedure for (3.9.1) only. It is usual to represent the derivative in (3.9.1) by the central difference approximation giving (see for example Fox (1962) p 248)

$$\alpha_1(t)U_{o,j} + \alpha_2(t)\left(\frac{U_{1,j} - U_{-1,j}}{2h}\right) = f_1(t) \ . \tag{3.9.3}$$

Rearranging to make $U_{-1,j}$ the subject of the equation we have

$$U_{-1,j} = U_{1,j} - \frac{2h}{\alpha_2(t)}\left(f_1(t) - \alpha_1(t)U_{o,j} \right) \tag{3.9.4}$$

The finite difference scheme (3.2.14) on this boundary is of the form

$$(1-\beta_1\theta)U_{-1,j+1} + 4(1+\beta_2\theta)U_{o,j+1} + (1-\beta_3\theta)U_{1,j+1}$$

$$= \{2+\beta_1(1-2\theta)\}U_{-1,j} + 4\{2-\beta_2(1-2\theta)\}U_{o,j} + \{2+\beta_3(1-2\theta)\}U_{1,j}$$

$$-(1-\beta_1\theta)U_{-1,j-1} - 4\,(1+\beta_2\theta)U_{o,j-1} - (1-\beta_3\theta)U_{1,j-1} \tag{3.9.5}$$

By using (3.9.4) we can eliminate the external value $U_{-1,j}$ from (3.9.5) giving

$$\left\{4(1+\beta_2\theta) + \frac{2h\alpha_1(t)}{\alpha_2(t)}(1-\beta_1\theta)\right\}U_{o,j+1} + \left\{2-(\beta_1+\beta_3)\,\theta\right\}U_{1,j+1}$$

$$=\left\{4\left[2-\beta_2(1-2\theta)\right]+\frac{2h\alpha_1(t)\left[2+\beta_1(1-2\theta)\right]}{\alpha_2(t)}\right\}U_{o,j}+\left\{4+(\beta_1+\beta_3)(1-2\theta)\right\}U_{1,j}$$

$$-\ \{4(1+\beta_2\theta)\ +\ \frac{2h\alpha_1(t)(1-\beta_1\theta)}{\alpha_2(t)}\}\ U_{o,j-1}\ -\{2-(\beta_1+\beta_3)\theta\}U_{1,j-1}$$

$$-\ \frac{2hf_1(t)}{\alpha_2(t)}\ \beta_1 \qquad (3.9.6)$$

Applying this procedure to the other boundary condition (3.9.2) we can obtain an easily sovable tri-diagonal system of equations on each time level. This method applies identically to the implicit scheme (3.5.2) and has been used in a later case study.

When the equation (3.1.1), or the equivalent equation with variable coefficients, has derivative boundary conditions, then the truncation error for the scheme under consideration will also be affected. This is because we have used the central difference formula

$$\left(\frac{\partial u}{\partial x}\right)_{i,j} = \frac{U_{i+1,j} - U_{i-1,j}}{2h} - \frac{h^2}{6}\left(\frac{\partial^3 u}{\partial x^3}\right)_{i,j} + \ldots \quad i = O,N$$

$$(3.9.7)$$

to eliminate the external values $U_{-1,j}$ and $U_{N+1,j}$. The third and higher order derivative terms in (3.9.7) must therefore be included in the truncation error and due account has been taken of this in the later case studies.

It must also be noted that the von Neumann stability analysis given in sections 3.4 and 3.6 does not take into account the boundary conditions. A more rigorous analysis of stability can be performed by employing the matrix method (see for example Ames 1969). In this case it is extremely complicated since our finite difference schemes are three time level and we shall not attempt to employ it in this thesis.

CHAPTER 4

Hyperbolic Partial Differential Equations with Variable Coefficients

4.1 Cubic Spline Finite Difference Scheme

In this case it is required to solve the equation

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x}\left[a(x,t)\frac{\partial u}{\partial x} \right] + b(x,t)\frac{\partial u}{\partial x} + c(x,t)U \qquad (4.1.1)$$

$$(0 \leq x \leq 1, \quad t > 0)$$

where $a(x,t)$, $b(x,t)$ and $c(x,t)$ are variable coefficients and $a(x,t)>0$ at all points of the solution domain. For ease we will rewrite (4.1.1) in the form

$$\frac{\partial^2 u}{\partial t^2} = a(x,t)\frac{\partial^2 u}{\partial x^2} + (a'(x,t)+b(x,t))\frac{\partial u}{\partial x} + c(x,t)U \qquad (4.1.2)$$

where the prime denotes partial differentiation with respect to x. We will consider the initial and boundary conditions associated with (4.1.2) to be the same as those given in section 3.1 for the constant coefficients case. This is only for convenience since the method described for derivative boundary conditions in section 3.9 applies equally well here.

If we again replace the time derivative in (4.1.2) by a finite difference approximation and the space derivatives by a cubic spline we obtain at (ih, jk)

$$\frac{U_{i,j-1}-2U_{i,j}+U_{i,j+1}}{k^2} = \theta a_{i,j-1}M_{i,j-1}+(1-2\theta)a_{i,j}M_{i,j}+\theta a_{i,j+1}M_{i,j+1}$$

$$+\theta(a'_{i,j-1}+b_{i,j-1})L_{i,j-1}+(1-2\theta)(a'_{i,j}+b_{i,j})L_{i,j}$$

$$+\ \theta(a'_{i,j+1}+b_{i,j+1})L_{i,j+1}$$

$$+\ \theta c_{i,j-1}U_{i,j-1}+(1-2\theta)c_{i,j}U_{i,j}+\theta c_{i,j+1}U_{i,j+1} \qquad (4.1.3)$$

$(i = 0,1, \ldots,N \ ; \ j = 1,2, \ldots \ ; \ Nh = 1)$

where the prime again denotes partial differentiation with respect to x and the $M_{i,j}$'s and $L_{i,j}$'s are as given for (3.2.1).

To derive the finite difference scheme incorporating splines for (4.1.1) we perform the same analysis as for the constant coefficient case given in section 3.2. We again make use of the continuity relationships (3.2.3) and (3.2.10) by taking combinations in the following manner -

(i) Equation (3.2.3) is multiplied throughout by $a_{i,j}(1-2\theta)$.

(ii) Equations (3.2.4) and (3.2.5) are multiplied by $a_{i,j-1}$ and $a_{i,j+1}$ respectively. The resulting equations are added together and multiplied throughout by θ.

(iii) Equation (3.2.10) is multiplied throughout by $(a'_{i,j} + b_{i,j})\ (1-2\theta)$.

(iv) Equations (3.2.11) and (3.2.12) are multiplied by $(a'_{i,j-1}+b_{i,j-1})$ and $(a'_{i,j+1} + b_{i,j+1})$ respectively. The resulting equations are added together and multiplied throughout by θ.

These expressions are now added together giving the following equation

$$\theta\rho_{j-1} + (1-2\theta)\rho_j + \theta\rho_{j+1} = \theta\mu_{j-1} + (1-2\theta)\mu_j + \theta\mu_{j+1} \tag{4.1.4}$$

where $$\rho_j = a_{i,j}\left(\frac{1}{6}M_{i-1,j} + \frac{2}{3}M_{i,j} + \frac{1}{6}M_{i+1,j}\right)$$

$$+ (a'_{i,j}+b_{i,j})\left(\frac{1}{6}L_{i-1,j} + \frac{2}{3}L_{i,j} + \frac{1}{6}L_{i+1,j}\right)$$

and $$\mu_j = a_{i,j}\left(\frac{U_{i-1,j}-2U_{i,j}+U_{i+1,j}}{h^2}\right) + \left(a'_{i,j}+b_{i,j}\right)\left(\frac{U_{i+1,j}-U_{i-1,j}}{2h}\right)$$

In order to eliminate the $M_{i,j}$'s and $L_{i,j}$'s from (4.1.4) we now use (4.1.3) directly and with i replaced by both (i-1) and (i+1). Following this procedure for the constant coefficients case we immediately obtained the required finite difference scheme. In this case however, use must also be made of Taylor Series expansions such that (4.1.3) can be employed to replace the M's and L's in (4.1.4) by mesh point values. We can therefore write (4.1.3), with i replaced by (i-1), in the form

$$\frac{U_{i-1,j-1}-2U_{i-1,j}+U_{i-1,j+1}}{k^2} = \theta\left(a_{i,j-1} - ha'_{i,j-1} + \frac{h^2}{2!}a''_{i,j-1} - \frac{h^3}{3!}a'''_{i,j-1} + \frac{h^4}{4!}a^{iv}_{i,j-1}\right)M_{i-1,j-1}$$

$$+(1-2\theta)\left(a_{i,j} - ha'_{i,j} + \frac{h^2}{2!}a''_{i,j} - \frac{h^3}{3!}a'''_{i,j} + \frac{h^4}{4!}a^{iv}_{i,j}\right)M_{i-1,j}$$

$$+\theta\left(a_{i,j+1} - ha'_{i,j+1} + \frac{h^2}{2!}a''_{i,j+1} - \frac{h^3}{3!}a'''_{i,j+1} + \frac{h^4}{4!}a^{iv}_{i,j+1}\right)M_{i-1,j+1}$$

$$+\theta\Big(a'_{i,j-1} - ha''_{i,j-1} + \frac{h^2}{2!}a'''_{i,j-1} - \frac{h^3}{3!}a^{iv}_{i,j-1}$$

$$+b_{i,j-1} - hb'_{i,j-1} + \frac{h^2 b''}{2!}{}_{i,j-1} - \frac{h^3 b'''}{3!}{}_{i,j-1} + \frac{h^4 b^{iv}}{4!}{}_{i,j-1}\Big)L_{i-1,j-1}$$

$$+(1-2\theta)\Big(a'_{i,j} - ha''_{i,j} + \frac{h^2}{2!}a'''_{i,j} - \frac{h^3}{3!}a^{iv}_{i,j}$$

$$+b_{i,j} - hb'_{i,j} + \frac{h^2}{2!}b''_{i,j} - \frac{h^3}{3!}b'''_{i,j} + \frac{h^4}{4!}b^{iv}_{i,j}\Big)L_{i-1,j}$$

$$+\theta\Big(a'_{i,j+1} - ha''_{i,j+1} + \frac{h^2}{2!}a'''_{i,j+1} - \frac{h^3}{3!}a^{iv}_{i,j+1}$$

$$+b_{i,j+1} - hb'_{i,j+1} + \frac{h^2 b''}{2!}{}_{i,j+1} - \frac{h^3 b'''}{3!}{}_{i,j+1} + \frac{h^4 b^{iv}}{4!}{}_{i,j+1}\Big)L_{i-1,j+1}$$

$$+\theta c_{i-1,j-1}U_{i-1,j-1} + (1-2\theta)c_{i-1,j}U_{i-1,j} + \theta c_{i-1,j+1}U_{i-1,j+1}$$

(4.1.5)

where again the primes denote partial derivatives with respect to x,which are shown to fourth order only. By expanding (4.1.3) with i replaced by (i+1) in a similar manner we can now use the results to express (4.1.4) in the form of the following finite difference scheme at (ih,jk)

$$(\phi_{i-1,j+1} - \theta\psi_{i,j+1})U_{i-1,j+1} + 4(\phi_{i,j+1} + 3\theta\gamma_{i,j+1})U_{i,j+1}$$

$$+ (\phi_{i+1,j+1} - \theta\chi_{i,j+1})U_{i+1,j+1}$$

$$=\{2\phi_{i-1,j} + k^2 c_{i-1,j} + (1-2\theta)\psi_{i,j}\}U_{i-1,j} + 4\{2\phi_{i,j} + k^2 c_{i,j} - 3(1-2\theta)\gamma_{i,j}\}U_{i,j}$$

$$+\{2\phi_{i+1,j} + k^2 c_{i+1,j} + (1-2\theta)\chi_{i,j}\}U_{i+1,j}$$

$$-(\phi_{i-1,j-1} - \theta\psi_{i,j-1})U_{i-1,j-1} - 4(\theta_{i,j-1} + 3\theta\gamma_{i,j-1})U_{i,j-1}$$

$$-(\phi_{i+1,j-1} - \theta\chi_{i,j-1})U_{i+1,j-1}$$

(4.1.6)

$$(i = 1,2, \ldots, (N-1))$$

where

$$\phi_{i,j} = 1-k^2\theta c_{i,j} \; ; \; \gamma_{i,j} = r^2 a_{i,j} \; ; \; \psi_{i,j} = 6\gamma_{i,j} - \beta_{i,j} \; ;$$

$$\chi_{i,j} = 6\gamma_{i,j} + \beta_{i,j}$$

and where $\beta_{i,j} = 3rk(a'_{i,j} + b_{i,j})$.

In deriving (4.1.6) we have neglected some lower order terms from (4.1.5) and its equivalent expansion. These additional terms are given in section 4.2 and have been included in the estimate of the numerical value of the truncation error for case study 5 where (4.1.6) has been employed.

4.2 Truncation Error and Stability conditions for (4.1.6)

The truncation error for the scheme (4.1.6) is obtained in the same manner as for the cases with constant coefficients. It is, however, much more complex and in fact involves some derivatives of the variable coefficients and also some odd order derivatives of u with respect to t. Thus (4.1.1) cannot be used to eliminate these latter derivatives and we obtain the truncation error at (ih, jk)

$$k^2h^2\left[\left\{-\frac{r^2\theta}{6}\left(6\frac{\partial^2 c}{\partial t^2}+h^2\frac{\partial^4 c}{\partial x^2\partial t^2}+\frac{k^2}{2}\frac{\partial^4 c}{\partial t^4}\right)-\frac{1}{6}\left(\frac{\partial^2 c}{\partial x^2}+\frac{h^2}{12}\frac{\partial^4 c}{\partial x^4}\right)\right\}U\right.$$

$$+\left\{-r^2\theta\left(\frac{\partial^2 a'}{\partial t^2}+\frac{\partial^2 b}{\partial t^2}+\frac{k^2}{12}\left(\frac{\partial^4 a'}{\partial t^4}+\frac{\partial^4 b}{\partial t^4}\right)\right)-\frac{k^2\theta}{3}\frac{\partial^3 c}{\partial x\partial t^2}-\frac{1}{6}\left(2\frac{\partial c}{\partial x}+\frac{k^2}{2}\frac{\partial^3 c}{\partial x^3}\right)\right\}\frac{\partial u}{\partial x}$$

$$+\left\{-r^2\theta\left(\frac{\partial^2 a}{\partial t^2}+\frac{k^2}{12}\frac{\partial^4 a}{\partial t^4}\right)-\frac{k^2\theta}{12}\left(2\frac{\partial^2 c}{\partial t^2}+h^2\frac{\partial^4 c}{\partial x^2\partial t^2}+\frac{k^2}{6}\frac{\partial^4 c}{\partial t^4}\right)\right.$$

$$\left.-\frac{1}{12}\left(2c+h^2\frac{\partial^2 c}{\partial x^2}+\frac{h^4}{12}\frac{\partial^4 c}{\partial x^4}\right)\right\}\frac{\partial^2 u}{\partial x^2}+\left\{-\frac{k^2\theta}{6}\left(\frac{\partial^2 a'}{\partial t^2}+\frac{\partial^2 b}{\partial t^2}+\frac{k^2}{12}\left(\frac{\partial^4 a'}{\partial t^4}\right.\right.\right.$$

$$\left.\left.\left.+\frac{\partial^4 b}{\partial t^4}\right)\right)-\frac{1}{6}(a'+b)-\frac{h^2k^2\theta}{3}\frac{\partial^3 c}{\partial x\partial t^2}-\frac{h^2}{36}\left(2\frac{\partial c}{\partial x}+\frac{h^2}{3}\frac{\partial^3 c}{\partial x^3}\right)\right\}\frac{\partial^3 u}{\partial x^3}$$

$$+\left\{-\frac{k^2\theta}{12}\left(\frac{\partial^2 a}{\partial t^2}+\frac{k^2}{12}\frac{\partial^4 a}{\partial t^4}\right)-\frac{a}{12}-\frac{k^2h^2\theta}{144}\left(2\frac{\partial^2 c}{\partial t^2}+h^2\frac{\partial^4 c}{\partial x^2\partial t^2}+\frac{k^2}{6}\frac{\partial^4 c}{\partial t^4}\right)\right.$$

$$\left.-\frac{h^2}{144}\left(2c+h^2\frac{\partial^2 c}{\partial x^2}+\frac{h^4}{12}\frac{\partial^4 c}{\partial x^4}\right)\right\}\frac{\partial^4 u}{\partial x^4}+\left\{-\frac{r^2\theta}{6}\left(12\frac{\partial c}{\partial t}+\frac{k^2}{2}\frac{\partial^3 c}{\partial t^3}\right.\right.$$

$$\left.\left.+2h^2\frac{\partial^3 c}{\partial t\partial x^2}\right)\right\}\frac{\partial u}{\partial t}+\left\{-\frac{r^2\theta}{12}\left(12c+8k^2\frac{\partial^2 c}{\partial t^2}+\frac{2k^4}{3}\frac{\partial^4 c}{\partial t^4}+k^2h^2\frac{\partial^4 c}{\partial x^2\partial t^2}\right)\right\}\frac{\partial^2 u}{\partial t^2}$$

$$+\left\{-\frac{k^2r^2\theta}{36}\left(12\frac{\partial c}{\partial t}+2h^2\frac{\partial^3 c}{\partial t\partial x^2}+2k^2\frac{\partial^3 c}{\partial t^3}\right)\right\}\frac{\partial^3 u}{\partial t^3}+\left\{\frac{r^2}{12}-\frac{k^2}{144}r^2\theta\left(12c+8k^2\frac{\partial^2 c}{\partial t^2}\right.\right.$$

$$\left.\left.+\frac{2k^4}{3}\frac{\partial^4 c}{\partial t^4}+k^2h^2\frac{\partial^4 c}{\partial x^2\partial t^2}\right)\right\}\frac{\partial^4 u}{\partial t^4}+\left\{-r^2\theta\left(2\left(\frac{\partial a'}{\partial t}+\frac{\partial b}{\partial t}\right)+\frac{k^3}{3}\left(\frac{\partial^3 a'}{\partial t^3}+\frac{\partial^3 b}{\partial t^3}\right)\right.\right.$$

$$-\frac{k^2\theta}{3}\left(2\frac{\partial^2 c}{\partial x\partial t}+\frac{1}{3}(h^2+k^2)\frac{\partial^4 c}{\partial t\partial x^3}\right)\Bigg\}\frac{\partial^2 u}{\partial x\partial t}+\Bigg\{-\frac{r^2\theta}{2}\left(2(a'+b)+k^2\left(\frac{\partial^2 a'}{\partial t^2}+\frac{\partial^2 b}{\partial t^2}\right)\right.$$

$$\left.+\frac{k^4}{12}\left(\frac{\partial^4 a'}{\partial t^4}+\frac{\partial^4 b}{\partial t^4}\right)\right)-\frac{k^2\theta}{6}\left(2\frac{\partial c}{\partial x}+k^2\frac{\partial^3 c}{\partial x\partial t}+\frac{h^3}{3}\frac{\partial^3 c}{\partial x^3}\right)\Bigg\}\frac{\partial^3 u}{\partial t^2\partial x}$$

$$+\Bigg\{-r^2\theta\left(2\frac{\partial a}{\partial t}+\frac{k^3}{3}\frac{\partial^3 a}{\partial t^3}\right)+\frac{k^2}{6}h\,\theta\left(2\frac{\partial^2 c}{\partial t\partial x}+\frac{1}{3}(h^2+k^2)\frac{\partial^4 c}{\partial x\partial t^3}\right)\Bigg\}\frac{\partial^3 u}{\partial t\partial x^2}$$

$$+\Bigg\{-\frac{k^2r^2}{6}\theta\left(2\left(\frac{\partial a'}{\partial t}+\frac{\partial b}{\partial t}\right)+\frac{k^2}{2}\left(\frac{\partial^3 a'}{\partial t^3}+\frac{\partial^3 b}{\partial t^3}\right)\right)-\frac{k^4\theta}{18}\left(2\frac{\partial^2 c}{\partial t\partial x}\right.$$

$$\left.+\frac{1}{3}(h^2+k^2)\frac{\partial^4 c}{\partial x\partial t^3}\right)\Bigg\}\frac{\partial^4 u}{\partial t^3\partial x}+\Bigg\{\frac{1}{6}-\frac{r^2\theta}{2}\left(2a+k^2\frac{\partial^2 a}{\partial t^2}+\frac{k^4}{12}\frac{\partial^4 a}{\partial t^4}\right)$$

$$-\frac{k^4}{12}\theta\left(2c+2k^2\frac{\partial^2 c}{\partial t^2}+\frac{k^2}{6}\frac{\partial^4 c}{\partial t^4}+\frac{k^2h^2}{2}\frac{\partial^4 c}{\partial x^2\partial t^2}\right)\Bigg\}\frac{\partial^4 u}{\partial t^2\partial x^2}$$

$$+\Bigg\{-\frac{k^2\theta}{6}\left(2\left(\frac{\partial a'}{\partial t}+\frac{\partial b}{\partial t}\right)+\frac{k^2}{3}\left(\frac{\partial^3 a'}{\partial t^3}+\frac{\partial^3 b}{\partial t^3}\right)\right)-\frac{k^2}{18}h^2\theta\left(2\frac{\partial^2 c}{\partial t\partial x}\right.$$

$$\left.+\frac{1}{3}(h^2+k^2)\frac{\partial^4 c}{\partial t\partial x^3}\right)\Bigg\}\frac{\partial^4 u}{\partial t\partial x^3}+\ldots\Bigg] \qquad (4.2.1)$$

In addition to this expression from the Taylor series expansion of each term in the scheme (4.1.6), we also have some terms resulting from the derivation of the scheme itself. These terms result in the following expression and must be added to (4.2.1) to give the full truncation error for (4.1.6)

$$-k^2\Big[\theta B_{j-1} M_{i-1,j-1} + (1-2\theta)B_j M_{i-1,j} + \theta B_{j+1} M_{i-1,j+1}$$

$$+ \theta C_{j-1} L_{i-1,j-1} + (1-2\theta)C_j L_{i-1,j} + \theta C_{j+1} L_{i-1,j+1}$$

$$+ \theta D_{j-1} M_{i+1,j-1} + (1-2\theta)D_j M_{i+1,j} + \theta D_{j+1} M_{i+1,j+1}$$

$$+ \theta E_{j-1} L_{i+1,j-1} + (1-2\theta)E_j L_{i+1,j} + \theta E_{j+1} L_{i+1,j+1}\Big] \tag{4.2.2}$$

where

$$B_j = -ha'_{i,j} + \frac{h^2}{2} a''_{i,j} - \frac{h^3}{6} a'''_{i,j} + \frac{h^4}{24} a^{iv}_{i,j}$$

$$C_j = -ha''_{i,j} + \frac{h^2}{2} a'''_{i,j} - \frac{h^3}{6} a^{iv}_{i,j} - hb'_{i,j} + \frac{h^2}{2} b''_{i,j} - \frac{h^3}{6} b'''_{i,j}$$

$$+ \frac{h^4}{24} b^{iv}_{i,j}$$

$$D_j = ha'_{i,j} + \frac{h^2}{2} a''_{i,j} + \frac{h^3}{6} a'''_{i,j} + \frac{h^4}{24} a^{iv}_{i,j}$$

$$E_j = ha''_{i,j} + \frac{h^2}{2} a'''_{i,j} + \frac{h^3}{6} a^{iv}_{i,j} + hb'_{i,j} + \frac{h^2}{2} b''_{i,j} + \frac{h^3}{6} b'''_{i,j}$$

$$+ \frac{h^4}{24} b^{iv}_{i,j}$$

The von Neumann stability condition for (4.1.6) is obtained by applying the method locally. This is because the method, which is based on Fourier series, is only applicable to difference schemes with constant coefficients. If we therefore derive the stability condition by considering the coefficients to be constant then it is reasonable to assume that the scheme with variable coefficients will be stable if the condition obtained is satisfied at every point in the solution domain. Since the scheme (4.1.6) reduces to (3.2.14) when the variable coefficients are considered to be constant then the stability conditions are naturally the same. Thus for the splines scheme (4.1.6)

(i) if $\theta \geqslant \frac{1}{4}$ it is unconditionally stable

(ii) if $\theta < \frac{1}{4}$ it is stable when $r \leqslant \{3a(x,t)(1-4\theta)\}^{-\frac{1}{2}}$

(4.2.3)

is satisfied independently at each point of the solution domain.

4.3 Comparable Finite Difference Scheme

Again as a comparison to the scheme incorporating splines we consider the well known finite difference approximation. The method is analogous to the constant coefficients case described in section 3.5 in that, at (ih,jk), we consider as an approximation to (4.1.2) the expression

$$\frac{U_{i,j-1}-2U_{i,j}+U_{i,j+1}}{k^2} = \frac{1}{h^2}\Big\{\alpha a_{i,j-1}\delta_x^2 U_{i,j-1} + (1-2\alpha)a_{i,j}\,\delta_x^2\,U_{i,j} + \alpha a_{i,j+1}\,\delta_x^2\,U_{i,j+1}\Big\}$$

$$+\frac{1}{2h}\Big\{\alpha(a'_{i,j-1}+b_{i,j-1})\delta_x U_{i,j-1}+(1-2\alpha)(a'_{i,j}+b_{i,j})\delta_x U_{i,j} + \alpha(a'_{i,j+1}+b_{i,j+1})\,\delta_x\,U_{i,j+1}\Big\}$$

contd. ...

$$+\alpha c_{i,j-1}U_{i,j-1}+(1-2\alpha)c_{i,j}U_{i,j} + \alpha c_{i,j+1}\,U_{i,j+1}$$

(4.3.1)

$(i = 0,1,\ldots,N\ ;\ j = 1,2,\ \ldots\ ;\ Nh = 1)$

where $\delta_x^2\, U_{i,j} = U_{i-1,j} - 2U_{i,j} + U_{i+1,j}$

and $\delta_x\, U_{i,j} = U_{i+1,j} - U_{i-1,j}$.

Collecting like terms in (4.3.1) and letting $r = k/h$ we have the tri-diagonal system

$$\left\{-\alpha r^2 a_{i,j+1}+\frac{\alpha h r^2}{2}(a'_{i,j+1}+b_{i,j+1})\right\}U_{i-1,j+1}+\left\{1+2\alpha r^2 a_{i,j+1}-\alpha k^2 c_{i,j+1}\right\}U_{i,j+1}$$

$$+\left\{-\alpha r^2 a_{i,j+1} - \alpha\,\frac{hr^2}{2}(a'_{i,j+1} + b_{i,j+1})\right\}U_{i+1,j+1}$$

$$=\left\{(1-2\alpha)r^2 a_{i,j}-(1-2\alpha)\frac{hr^2}{2}(a'_{i,j}+b_{i,j})\right\}U_{i-1,j}+\left\{2-2(1-2\alpha)r^2 a_{i,j}\right.$$

$$\left.+(1-2\alpha)k^2 c_{i,j}\right\}U_{i,j} + \left\{(1-2\alpha)r^2 a_{i,j}+(1-2\alpha)\frac{hr^2}{2}(a'_{i,j}+b_{i,j})\right\}U_{i+1,j}$$

$$-\left\{-\alpha r^2 a_{i,j-1}+\frac{\alpha h r^2}{2}(a'_{i,j-1}+b_{i,j-1})\right\}U_{i-1,j-1}-\left\{1+2\alpha r^2 a_{i,j-1}-\alpha k^2 c_{i,j-1}\right\}$$

$$U_{i,j-1} -\left\{-\alpha r^2 a_{i,j-1} - \frac{\alpha h r^2}{2}(a'_{i,j-1}+b_{i,j-1})\right\}U_{i+1,j-1}$$

(4.3.2)

$(i = 1,2,\ \ldots,\ (N-1))$

4.4 Truncation Error and Stability Conditions for (4.3.2)

In the usual manner we expand each term of (4.3.2) about the mesh point (ih,jk) and thus obtain the expression for the truncation error

$$k^2h^2\Bigg[-\alpha r^2\left(\frac{\partial^2 c}{\partial t^2}+\frac{k^2}{12}\frac{\partial^4 c}{\partial t^4}\right)u - \alpha r^2\left(\frac{\partial^2 a'}{\partial t^2}+\frac{\partial^2 b}{\partial t^2}+\frac{k^2}{12}\left[\frac{\partial^4 a'}{\partial t^4}+\frac{\partial^4 b}{\partial t^4}\right]\right)\frac{\partial u}{\partial x}$$

$$-\alpha r^2\left(\frac{\partial^2 a}{\partial t^2}+\frac{k^2}{12}\frac{\partial^4 a}{\partial t^4}\right)\frac{\partial^2 u}{\partial x^2}+\left\{-\frac{1}{6}(a'+b)-\frac{k^2}{6}\alpha\left(\frac{\partial^2 a'}{\partial t^2}+\frac{\partial^2 b}{\partial t^2}\right.\right.$$

$$\left.\left.+\frac{k^2}{12}\left[\frac{\partial^4 a'}{\partial t^4}+\frac{\partial^4 b}{\partial t^4}\right]\right)\right\}\frac{\partial^3 u}{\partial x^3}+\left\{-\frac{a}{12}-\frac{k^2}{12}\alpha\left(\frac{\partial^2 a}{\partial t^2}+\frac{k^2}{12}\frac{\partial^4 a}{\partial t^4}\right)\right\}\frac{\partial^4 u}{\partial x^4}$$

$$-\alpha r^2\left(2\frac{\partial c}{\partial t}+\frac{k^3}{3}\frac{\partial^3 c}{\partial t^3}\right)\frac{\partial u}{\partial t}-\alpha\frac{r^2}{2}\left(2c+k^2\frac{\partial^2 c}{\partial t^2}+\frac{k^4}{12}\frac{\partial^4 c}{\partial t^4}\right)\frac{\partial^2 u}{\partial t^2}$$

$$-\alpha\frac{k^2r^2}{6}\left(2\frac{\partial c}{\partial t}+\frac{k^3}{3}\frac{\partial^3 c}{\partial t^3}\right)\frac{\partial^3 u}{\partial t^3}+\left\{\frac{r^2}{12}-\alpha\frac{k^2r^2}{24}\left(2c+k^2\frac{\partial^2 c}{\partial t^2}+\frac{k^4}{12}\frac{\partial^4 c}{\partial t^4}\right)\right\}\frac{\partial^4 u}{\partial t^4}$$

$$-\alpha r^2\left(2\left[\frac{\partial a'}{\partial t}+\frac{\partial b}{\partial t}\right]+\frac{k^3}{3}\left[\frac{\partial^3 a'}{\partial t^3}+\frac{\partial^3 b}{\partial t^3}\right]\right)\frac{\partial^2 u}{\partial t\partial x}-\alpha\frac{r^2}{2}\left(2(a'+b)\right.$$

$$\left.+k^2\left[\frac{\partial^2 a'}{\partial t^2}+\frac{\partial^2 b}{\partial t^2}\right]+\frac{k^2}{12}\left[\frac{\partial^4 a'}{\partial t^4}+\frac{\partial^4 b}{\partial t^4}\right]\right)\frac{\partial^3 u}{\partial t^2\partial x}-\alpha\frac{k^2r^2}{6}\left(2\left[\frac{\partial a'}{\partial t}+\frac{\partial b}{\partial t}\right]\right.$$

$$\left.+\frac{k^3}{3}\left[\frac{\partial^3 a'}{\partial t^3}+\frac{\partial^3 b}{\partial t^3}\right]\right)\frac{\partial^4 u}{\partial t^3\partial x}-\alpha r^2\left(2\frac{\partial a}{\partial t}+\frac{k^3}{3}\frac{\partial^3 a}{\partial t^3}\right)\frac{\partial^3 u}{\partial t\partial x^3}-\alpha\frac{r^2}{2}\left(2a\right.$$

$$\left.+k^2\frac{\partial^2 a}{\partial t^2}+\frac{k^4}{12}\frac{\partial^4 a}{\partial t^4}\right)\frac{\partial^4 u}{\partial t^2\partial x^2}-\alpha\frac{k^2}{6}\left(2\left[\frac{\partial a'}{\partial t}+\frac{\partial b}{\partial t}\right]\right.$$

$$+\frac{k^3}{3}\left[\frac{\partial^3 a'}{\partial t^3}+\frac{\partial^3 b}{\partial t^3}\right]\Bigg)\frac{\partial^4 u}{\partial t\partial x^3}+\ldots\Bigg] \qquad (4.4.1)$$

This truncation error (4.4.1), for the fully implicit finite difference scheme (4.3.2), is slightly simpler than that for the scheme incorporating splines. This is due to the fact that there are no terms such as $c_{i-1,j}$, $c_{i-1,j-1}$ etc. in the finite difference scheme.

The von Neumann stability condition for (4.3.2) is found by again applying the method locally. Hence by considering the coefficients to be constant we can obtain the following conditions governing stability.

(i) if $\alpha \geqslant \frac{1}{4}$ the scheme is unconditionally stable.

(ii) If $\alpha < \frac{1}{4}$ the scheme is stable provided

$$r \leqslant \{a(x,t)(1-4\alpha)\}^{-\frac{1}{2}} \qquad (4.4.2)$$

is satisfied independently at every point in the solution domain.

The stability condition (4.4.2) can also be rewritten in the alternative form

$$r \leqslant \{\max a(x,t)(1-4\alpha)\}^{-\frac{1}{2}} . \qquad (4.4.3)$$

This form is used by Saul'yev (1964) when considering parabolic equations with variable coefficients and is equally applicable to the stability condition (4.2.3).

Although our method of obtaining the conditions (4.2.3) and (4.4.2) is not a rigorous proof of stability it is supported by numerical evidence in the case study 5 where the coefficients are varying throughout the range $0 \leq x \leq 1$. A more thorough analysis of stability can be performed using the energy method. It has, in fact, been shown by Lees (1960) that the finite difference approximation

$$\frac{1}{k^2}\delta_t^2 U_{i,j} = a(x,t)\left\{\frac{1}{h^2}\left(\alpha\delta_x^2 U_{i,j-1} + (1-2\alpha)\delta_x^2 U_{i,j} + \alpha\delta_x^2 U_{i,j+1}\right)\right\}$$

$$+ b(x,t)\left\{\frac{1}{2h}\delta_x U_{i,j}\right\} + c(x,t)\left\{\frac{1}{2k}\delta_t U_{i,j}\right\}$$

$$+ d(x,t) U_{i,j} + e(x,t) \qquad (4.4.4)$$

of the equation

$$\frac{\partial^2 u}{\partial t^2} = a(x,t)\frac{\partial^2 u}{\partial x^2} + b(x,t)\frac{\partial u}{\partial x} + c(x,t)\frac{\partial u}{\partial t} + d(x,t)U + e(x,t) \qquad (4.4.5)$$

has the same condition governing unconditional stability as the equivalent scheme for constant coefficients. The stability conditions for (4.4.4) with constant a,b,c,d and e are the same as (3.6.3). Unfortunately the energy method used in the Lees paper gives no information about the conditional stability of the scheme although this has since been remedied by Friberg (1961). In his paper, Friberg modified the method used by Lees to examine the conditional stability of (4.4.4) with $b(x,t) = c(x,t) = d(x,t) = e(x,t) = 0$. By constructing a system of first order equations equivalent to

$$\frac{\partial^2 u}{\partial t^2} = a \frac{\partial^2 u}{\partial x^2} \quad \text{where } a = a(x,t) \tag{4.4.6}$$

he showed that (4.4.4) (with b=c=d=e=0) is stable for $\alpha<\frac{1}{4}$ when

$$r^2 \leqslant a(1-4\alpha) \tag{4.4.7}$$

which is identical to the condition (3.6.3) governing stability of the scheme for constant coefficients.

These results deribed by Lees and Friberg indicate that the change from constant coefficients to variable coefficients will not greatly affect the stability conditions for a particular shceme provided that they are satisfied at every point in the solution domain.

4.5 Numerical Procedure and Evaluation of Truncation Errors

The numerical procedure for finding cubic spline solutions at points intermediate to mesh points for the equation with variable coefficients (4.1.1) is almost identical to that given in section 3.7. The only major difference is in deriving the "end conditions", since the continuity expression (3.2.3) will be unchanged and thus used in the same manner. As an example, the "end condition" (3.7.2) on the boundary x=0 will become

$$k^2\{a_{o,o}M_{o,o}+(a'_{o,o}+b_{o,o})L_{o,o}\} = 2(f_1(k)-f_2(0)-kg_2(0)) - k^2c_{o,o}f_1(0). \tag{4.5.1}$$

Similarly equations (3.7.3), (3.7.9) and (3.7.10) are affected by the variable coefficients and with the exception of replacing

(3.2.1) by (4.1.3) in the general scheme, the remaining procedure is unchanged.

The estimates of the truncation errors (4.4.1), for the implicit scheme, and (4.2.1) plus (4.2.2) for the splines scheme are obtained using the method outlined in section 3.8. In this case, however we have to evaluate the derivatives of the variable coefficients with respect to both x and t and also some mixed derivatives of the coefficient $c(x,t)$. This is easily done using difference approximations similar to (3.8.1) provided the variables $a(x,t)$, $b(x,t)$ and $c(x,t)$ are sufficiently differentiable. Assuming this is so, then it can be seen that the truncation errors (4.2.1) and (4.4.1) are both $O(k^2h^2)$ as was the case for the equivalent schemes with constant coefficients. The expression (4.2.2) which is added to (4.2.1) to give the full truncation error for the scheme (4.1.6), is $O(k^2h)$ although in practice it should be noted that the numerical values of these truncation errors will greatly depend on the actual values of a, b and c.

When obtaining an estimate of the truncation error for the splines scheme (4.1.6) we must first use the numerical procedure to find values for the M's and L's since these are required in (4.2.2) This might be thought of as a disadvantage in that it requires more calculations to obtain the full truncation error but since we will use the numerical procedure to find solutions at points intermediate to mesh points then little extra work is required.

4.6 Singularities in the variable coefficients

It is sometimes the case that one or more of the variable coefficients $a(x,t)$, $b(x,t)$ or $c(x,t)$ will have a singularity at a point in the range $0 \leqslant x \leqslant 1$. This problem is best discussed using an example due to Collatz (1960), where the equation under consideration is given by (4.1.1) with $a(x,t) = 1$, $b(x,t) = 1/x$ and $c(x,t) = 0$. In this case finite difference approximations require special consideration at the singular point $x = 0$. One method is to employ L' Hospital's rule, which in this case gives

$$\lim_{x\to 0} \frac{\partial u/\partial x}{x} = \lim_{x\to 0} \frac{\partial^2 u/\partial x^2}{1} = \left(\frac{\partial^2 u}{\partial x^2}\right)_{x=0} \qquad (4.6.1)$$

whereby the differential equation becomes at $x=0$

$$2\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial t^2} \qquad (4.6.2)$$

A finite difference scheme can now be used to approximate (4.6.2) and hence the problem has been removed. In a similar manner other singularities occuring in (4.1.1) can be dealt with and the schemes (4.1.6) and (4.3.2) modified to give the required solutions. The equation under consideration in case study 5 does in fact have a singularity at $x=2$ since the coefficient $b(x,t) = 2(x-2)^{-1}$. Fortunately this poses no problem in this instance since the equation (4.1.1) has only been considered in the range $0 \leqslant x \leqslant 1$.

Using L'Hospital's rule is only one method of dealing with singularities of this form. We will return to this problem in the next chapter where the alternative method of mesh refinement is discussed.

CHAPTER 5

Unequal Step Lengths

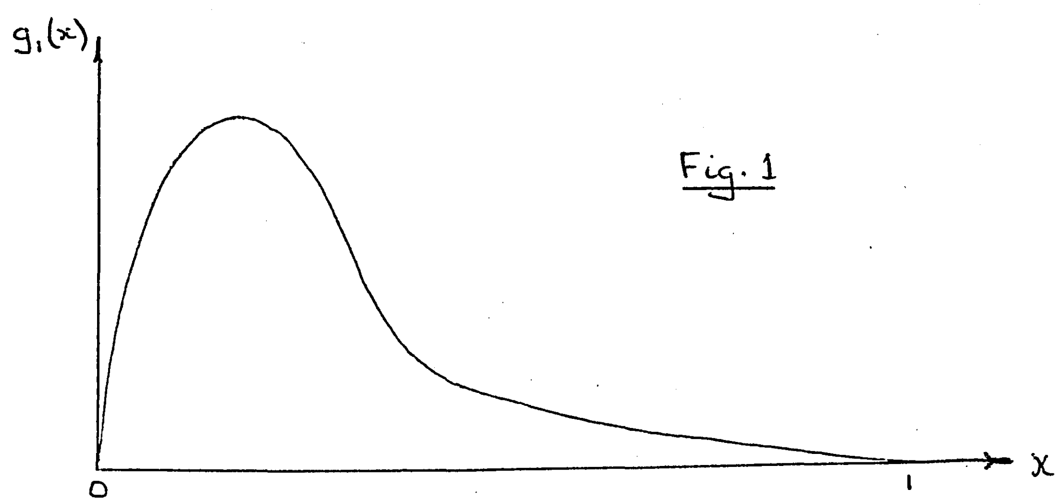
5.1 General Considerations

As indicated in section 2.2 the knots x_i $(i = 0,1,\ldots,N)$ need not necessarily be equally spaced in the range $a \leqslant x \leqslant b$. The distance between successive knots given by

$$h_i = x_i - x_{i-1} \qquad (i = 1,2,\ldots,N) \qquad (5.1.1)$$

is chosen as the constant value h in chapters 3 and 4 simply because it is usual in deriving finite difference approximations to have a constant step length. As will be shown in this chapter both the well known finite difference scheme and the scheme incorporating splines can be generalised to the case of non-uniform step lengths.

Before deriving these latter schemes it is convenient to first consider the relative merits of employing a variable mesh; the main advantage of this over the constant step length schemes being obtained when the initial function $g_1(x)$, as given in (3.1.3) is of the form illustrated in Fig. 1.



Here, as suggested by Saul'yev (1964), we would use a smaller step length h in the first third of the interval $[0,1]$ than in the remaining two thirds. This is because the function is varying rapidly in the first third of the interval and hence more solution values are required. The alternative to using variable step lengths in this case, is to employ a very small constant step length throughout the range $[0,1]$ and thus obtain the required solutions. This is very uneconomical with regard to computing time since the number of arithmetical operations is greatly increased and thus a variable step length scheme is desirable.

The major disadvantage of using a non-uniform mesh to solve partial differential equation is that the schemes produced are rather more complex and are also likely to have a larger error. The reason for this is that there are errors associated with all finite difference approximations which usually depend on the size of the step length h. If we therefore increase h in one portion of the range so as to decrease it in another than the order of the error will still depend on the larger step length used. This is also the case in the finite difference schemes incorporating splines since the spline function is, by definition, a piecewise polynomial in each of the intervals between successive knots. If we therefore increase the distance between knots in one part of the range then we would naturally expect the truncation error to be large in that vicinity. We have found that this can be counteracted to some extent in our schemes by choosing the parameters α and θ to eliminate some terms in the truncation errors. This technique is discussed in section 5.5 where we use varying values of α and θ.

An additional advantage of employing schemes with unequal mesh lengths may occur when the equation (4.1.1) has a singularity in one of the variable coefficients. As indicated by Ames (1969), a very common method of dealing with this type of problem is to attempt to diminish the effect of the singularity by refining the mesh in the region where it occurs. Thus by reducing the size of the step length around the singularity we minimize the 'area of infection' it causes. If we were solving a problem of this type using either of the constant step length schemes (4.1.6) or (4.3.2) we would have to reduce the size of the mesh throughout the whole range $0 \leqslant x \leqslant 1$. This is again very uneconomical with regards to computing time, particularly when a large amount of solutions are not required in the x direction. It would therefore be useful if we had a scheme with variable step lengths where we could use small mesh widths around the singularity and larger mesh widths in the remainder of the range.

5.2 Cubic Spline Finite Difference Approximation to (3.1.1)

Recalling the results of Ahlberg, Nilson and Walsh (1967), outlined in section 2.2, we can show that equation (2.2.6) becomes on the j^{th} time line

$$\frac{h_i}{6} M_{i-1,j} + \frac{h_i + h_{i+1}}{3} M_{i,j} + \frac{h_{i+1}}{6} M_{i+1,j}$$

$$= \frac{h_i U_{i+1,j} - (h_i + h_{i+1}) U_{i,j} + h_{i+1} U_{i-1,j}}{h_i \, h_{i+1}}$$

$$(i = 1,2,\ldots,(N-1)) \qquad (5.2.1)$$

where h_i is as given in (5.1.1). Similarly (5.2.1) also holds on the $(j-1)^{th}$ and $(j+1)^{th}$ time lines, these three equations being equivalent to (3.2.3),(3.2.4) and (3.2.5) for the case of constant step lengths.

To obtain the expressions corresponding to (3.2.10) we first multiply (2.2.4) and (2.2.5) by h_i and h_{i+1} respectively thus obtaining on the j^{th} time line

$$h_i L_{ij} = \frac{h_i^2}{6} M_{i-1,j} + \frac{h_i^2}{3} M_{i,j} + U_{i,j} - U_{i-1,j} \tag{5.2.2}$$

$$h_{i+1} L_{i,j} = -\frac{h_{i+1}^2}{3} M_{i,j} - \frac{h_{i+1}^2}{6} M_{i+1,j} + U_{i+1,j} - U_{i,j} \tag{5.2.3}$$

These equations are now added together giving

$$(h_i+h_{i+1})L_{i,j} = \frac{h_i^2}{6} M_{i-1,j} + \frac{(h_i^2-h_{i+1}^2)}{3} M_{i,j} - \frac{h_{i+1}^2}{6} M_{i+1,j} + U_{i+1,j} - U_{i-1,j} \tag{5.2.4}$$

Replacing i by (i+1) and (i-1) in (5.2.2) and (5.2.3) respectively

$$h_{i+1}L_{i+1,j} = \frac{h_{i+1}^2}{6} M_{i,j} + \frac{h_{i+1}^2}{3} M_{i+1,j} + U_{i+1,j} - U_{i,j} \tag{5.2.5}$$

$$h_i L_{i-1,j} = \frac{-h_i^2}{3} M_{i-1,j} - \frac{h_i^2}{6} M_{i,j} + U_{i,j} - U_{i-1,j} \tag{5.2.6}$$

Adding these equations together we obtain the expression

$$h_i L_{i-1,j} + h_{i+1} L_{i+1,j} = \frac{-h_i^2}{3} M_{i-1,j} + \frac{(h_{i+1}^2 - h_i^2)}{6} M_{i,j} + \frac{h_{i+1}^2}{3} M_{i+1,j}$$

$$+ U_{i+1,j} - U_{i-1,j} . \tag{5.2.7}$$

The final expression corresponding to the constant step lengths equation (3.2.10) is obtained by adding (5.2.4) to half of (5.2.7) thus giving the required result

$$\frac{h_i}{6} L_{i-1,j} + \frac{h_i+h_{i+1}}{3} L_{i,j} + \frac{h_{i+1}}{6} L_{i+1,j}$$

$$= \frac{(h_i^2 - h_{i+1}^2)}{12} M_{i,j} + \frac{U_{i+1,j} - U_{i-1,j}}{2} \qquad (5.2.8)$$

As in all previous cases (5.2.8) again holds on the $(j-1)^{th}$ and $(j+1)^{th}$ time lines. Thus by following the steps given in section 3.2 we take combinations of (5.2.1) and (5.2.8) in such a way as to eliminate the majority of the M's and L's using (3.2.1). Unfortunately in this case, we have no way of eliminating the terms in $M_{i,j}$, $M_{i,j-1}$ and $M_{i,j+1}$ resulting from (5.2.8) and they therefore appear in the following finite difference scheme

$$(\phi_i-\gamma_i\theta)U_{i-1,j+1}+2(1+\chi_i\theta)U_{i,j+1}+(\beta_i-\psi_i\theta)U_{i+1,j+1}$$

$$+ \frac{bk^2}{2}(h_{i+1}-h_i)(\theta M_{i,j-1}+(1-2\theta)M_{i,j}+\theta M_{i,j+1})$$

$$=\{2\phi_i+\gamma_i(1-2\theta)\}U_{i-1,j}+2\{2-\chi_i(1-2\theta)\}U_{i,j}+\{2\beta_i+\psi_i(1-2\theta)\}U_{i+1,j}$$

$$-(\phi_i-\gamma_i\theta)U_{i-1,j-1} -2(1+\chi_i\theta)U_{i,j-1} -(\beta_i-\psi_i\theta)U_{i+1,j-1} \qquad (5.2.9)$$

$$(i = 1,2, \ldots, (N-1))$$

where

$$\gamma_i = 6ar_i - 3bh_{i+1}s_i + ch_ih_{i+1}s_i,$$

$$\chi_i = 6a\nu_i - ck^2,$$

$$\psi_i = 6as_i + 3bh_ir_i + ch_ih_{i+1}r_i,$$

$$\phi_i = \frac{h_i}{h_i + h_{i+1}}, \quad \beta_i = \frac{h_{i+1}}{h_i + h_{i+1}}$$

and the mesh ratios r_i, s_i and ν_i are given by

$$r_i = \frac{k^2}{h_i(h_i+h_{i+1})}, \quad s_i = \frac{k^2}{h_{i+1}(h_i+h_{i+1})}, \quad \nu_i = \frac{k^2}{2h_i h_{i+1}}$$

Due to the presence of the M's in (5.2.9) we cannot calculate solutions of (3.1.1) in the usual manner common to most finite difference approximations. Assuming the scheme (5.2.9) is fully developed in that mesh point solutions on the $(j-1)^{th}$ and j^{th} time lines are known, then we must obviously determine the M values on both these and also the advanced time line (j+1) before further solutions can be found. To overcome this problem we use a numerical procedure, similar to that given in section 3.7, at each stage of the computation. Thus by following the steps given below (5.2.9) can be used to obtain the required solutions to (3.1.1):

(i) Putting $\theta=0$ in (5.2.9) and by making use of the derivative initial condition in (3.1.3) we have

$$2\phi_i U_{i-1,1}+4U_{i,1}+2\beta_i U_{i+1,1}+\frac{bk^2}{2}(h_{i+1}-h_i)M_{i,o}$$

$$= (2\phi_i+\gamma_i)U_{i-1,o}+2(2-\chi_i)U_{i,o}+(2\beta_i+\psi_i)U_{i+1,o}$$

$$+2k\phi_i g_2(x_{i-1})+4kg_2(x_i)+2k\beta_i g_2(x_{i+1}) \qquad (5.2.10)$$

(ii) To evaluate $M_{i,o}$ $(i = 0,1,\ldots,N)$ we follow the procedure given in step (i) of section 3.7, where (3.7.2) and (3.7.3) will be unchanged and where the equations (3.7.1), (3.7.4) and (3.7.5) will be replaced respectively by

$$\frac{h_i}{6} M_{i-1,o} + \frac{(h_i+h_{i+1})}{3} M_{i,o} + \frac{h_{i+1}}{6} M_{i+1,o}$$

$$= \frac{h_i g_1(x_{i+1}) - (h_i+h_{i+1}) g_1(x_i) + h_{i+1} g_1(x_{i-1})}{h_i h_{i+1}} \qquad (5.2.11)$$

$$(i = 1,2,\ldots,(N-1))$$

$$L_{o,o} = -\frac{h_1}{3} M_{o,o} - \frac{h_1}{6} M_{1,o} + \frac{g_1(h)-g_1(0)}{h_i} \qquad (5.2.12)$$

$$L_{N,O} = \frac{h_N}{3} M_{N,O} + \frac{h_N}{6} M_{N-1,O} + \frac{g_1(1)-g_1(x_{N-1})}{h_N} \qquad (5.2.13)$$

As in section 3.7 the values $L_{o,o}$ and $L_{N,O}$ will also result from these equations.

(iii) The remaining $L_{i,o}$ values $(i = 1,2, \ldots,(N-1))$ are again obtained by averaging (2.2.4) and (2.2.5), thus giving on the j^{th} time line

$$L_{i,j} = \frac{h_i}{12} M_{i-1,j} + \frac{(h_i-h_{i+1})}{6} M_{i,j} - \frac{h_{i+1}}{12} M_{i+1,j}$$

$$+ \frac{h_i U_{i+1,j} - (h_i-h_{i+1}) - h_{i+1} U_{i-1,j}}{2h_i h_{i+1}} \qquad (5.2.14)$$

$$(i = 1,2,\ldots(N-1))$$

(iv) Having found $M_{i,o}$ $(i = 0,1,\ldots,N)$ we can now proceed with the solution of the tri-diagonal system (5.2.10) in the usual manner. This results in our obtaining the required mesh point solutions $U_{i,1}$ $(i = 0,1, \ldots,N)$.

These solutions are now used in finding $M_{i,1}$ $(i = 0,1,\dots,N)$ by the method given in step (v) of the section 3.7. Again only the equations (3.7.8), (3.7.11) and (3.7.12) are affected by the change from constant to variable step lengths and hence the required $M_{i,1}$ values are easily found.

Using equation (5.2.14) the values $L_{i,1}$ $(i = 0,1,\dots,N)$ are also found at this stage.

(v) Since we have now determined $M_{i,o}$, $M_{i,1}$, $L_{i,o}$ and $L_{i,1}$ $(i = 0,1,\dots,N)$ we can hence develop the general scheme for obtaining $M_{i,j+1}$ and subsequently $U_{i,j+1}$ for $j = 1,2, \dots$. We do this using a simple iterative process, beginning with an initial approximation $M^{(0)}_{i,j+1}$, to $M_{i,j+1}$, given by

$$M^{(o)}_{i,j+1} = 2M_{i,j} - M_{i,j-1} \quad (i = 0,1,\dots,N;\ j = 1,2,\dots). \tag{5.2.15}$$

Substituting (5.2.15) into the scheme (5.2.9) we thereby reduce it to an ordinary finite difference formula which is easily solvable for initial approximations to $U_{i,j+1}$, which we denote by $U^{(0)}_{i,j+1}$.

(vi) Suppose now that the iterative process is fully developed and that we wish to obtain improved approximations to $M_{i,j+1}$. This is done by reconsidering (5.2.1) which, on denoting the improved values by $M^{(n+1)}_{i,j+1}$, can be written in the form

$$\frac{h_i}{6} M^{(n+1)}_{i-1,j+1} + \frac{(h_i+h_{i+1})}{3} M^{(n+1)}_{i,j+1} + \frac{h_{i+1}}{6} M^{(n+1)}_{i+1,j+1}$$

$$= \frac{h_i U^{(n)}_{i+1,j+1} - (h_i+h_{i+1}) U^{(n)}_{i,j+1} + h_{i+1} U^{(n)}_{i-1,j+1}}{h_i\ h_{i+1}} . \tag{5.2.16}$$

$$(i = 1,2,\dots,(N-1))$$

As the values on the right-hand-side of (5.2.16) are known then this system is (N-1) equations in (N+1) unknowns. The two additional equations required to express (5.2.16) in tri-diagonal form are obtained by firstly letting i=O and i=N in (3.2.1). This introduces two additional unknowns $L_{o,j+1}$ and $L_{N,j+1}$ which we can eliminate using the continuity expressions (see (2.2.4) and (2.2.5))

$$L_{i,j} = \frac{-h_{i+1}}{3}M_{i,j}\frac{-h_{i+1}}{6}M_{i+1,j}+\frac{U_{i+1,j}-U_{i,j}}{h_{i+1}} \qquad (5.2.17)$$

$$(i = 0,1,\ldots,(N-1))$$

$$L_{i,j} = \frac{h_i}{6} M_{i-1,j} + \frac{h_i}{3} M_{i,j} + \frac{U_{i,j}-U_{i-1,j}}{h_i} \qquad (5.2.18)$$

$$(i = 1,2,\ldots,N)$$

Thus by substituting for $L_{o,j+1}$ and $L_{N,j+1}$ the two additional equations become

$$\left(a\theta-\frac{bh_1\theta}{3}\right)M^{(n+1)}_{0,j+1} - \frac{bh_1\theta}{6} M^{(n+1)}_{1,j+1}$$

$$= \frac{U_{0,j-1}-2U_{0,j}+U^{(n)}_{0,j+1}}{k^2} - a\{\theta M_{0,j-1} + (1-2\theta)M_{0,j}\}$$

$$-b\left\{\theta L_{0,j-1}+(1-2\theta)L_{0,j}+\theta\left(\frac{U^{(n)}_{1,j+1}-U^{(n)}_{0,j+1}}{h_1}\right)\right\}$$

$$- c\{\theta U_{0,j-1}+(1-2\theta)U_{0,j} + \theta U^{(n)}_{0,j+1}\} \qquad (5.2.19)$$

and

$$(a\theta+\frac{bh_N\theta}{3})M^{(n+1)}_{N,j+1} + \frac{bh_N\theta}{6}M^{(n+1)}_{N-1,j+1}$$

contd. ...

$$= \frac{U_{N,j-1}-2U_{N,j}+U^{(n)}_{N,j+1}}{k^2} - a\left\{\theta M_{N,j-1}+(1-2\theta)M_{N,j}\right.$$

$$-b\{\theta L_{N,j-1}+(1-2\theta)L_{N,j}+\theta\left(\frac{U^{(n)}_{N,j+1}-U^{(n)}_{N-1,j+1}}{h_N}\right)\Bigg\}$$

$$-c\{\theta U_{N,j-1}+(1-2\theta)U_{N,j}+\theta U^{(n)}_{N,j+1}\} \qquad (5.2.20)$$

Equations (5.2.16), (5.2.19) and (5.2.20) form an easily solvable tri-diagonal system from which we obtain the improved approximations $M^{(n+1)}_{i,j+1}$.

(vii) The values $M^{(n+1)}_{i,j+1}$ as obtained from (vi) are now used in the finite difference scheme (5.2.9), which in iterative notation, is

$$(\phi_i-\gamma_i\theta)U^{(n+1)}_{i-1,j+1}+2(1+\chi_i\theta)U^{(n+1)}_{i,j+1}+(\beta_i-\psi_i\theta)U^{(n+1)}_{i+1,j+1}$$

$$+\frac{bk^2}{2}(h_{i+1}-h_i)(\theta M_{i,j-1}+(1-2\theta)M_{i,j}+\theta M^{(n+1)}_{i,j+1})$$

$$=\{2\phi_i+\gamma_i(1-2\theta)\}U_{i-1,j}+2\{2-\chi_i(1-2\theta)\}U_{i,j}+\{2\beta_i+\psi_i(1-2\theta)\}U_{i+1,j}$$

$$-(\phi_i-\gamma_i\theta)U_{i-1,j-1}-2(1+\chi_i\theta)U_{i,j-1}-(\beta_i-\psi_i\theta)U_{i+1,j-1}\ .$$

$$(i = 1,2,\ \ldots,(N-1)) \qquad (5.2.21)$$

Thus by solving (5.2.21) we obtain improved approximations to $U_{i,j+1}$ which we have denoted by $U^{(n+1)}_{i,j+1}$.

(viii) A test is now performed on these $U^{(n+1)}_{i,j+1}$ values by examining numerical values of the inequality

$$\left| U_{i,j+1}^{(n+1)} - U_{i,j+1}^{(n)} \right| \leq \epsilon \quad (i = 1, 2, \ldots, (N-1)) \quad (5.2.22)$$

for some fixed tolerance ϵ.

If (5.2.22) is satisfied for all i then $U_{i,j+1}^{(n+1)}$ are taken as the required mesh point solutions. However, if (5.2.22) is violated then the $U_{i,j+1}^{(n+1)}$ values must be re-employed in (5.2.16) before proceeding as before.

This iterative process might seem a rather uneconomical method of solution when compared with the simpler finite difference scheme described in the next section. Fortunately the M and L values which we have found in the process of evaluating the mesh point solutions are also of further use. As mentioned earlier, a major advantage of the splines schemes is that by making use of the spline function (2.2.2) we can easily obtain solutions at points intermediate to those at mesh points. This is done in the constant step length case by following the numerical procedure described in section 3.7. In this variable step length case however the numerical procedure is incorporated into finding the mesh point solutions and thus there is little extra computation required when intermediate solutions are to be obtained.

5.3 Finite Difference Approximation to (3.1.1)

Expanding $U_{i+1,j}$ and $U_{i-1,j}$ by the well known Taylor series expansions we find

$$U_{i+1,j} = U_{i,j} + h_{i+1}\left(\frac{\partial u}{\partial x}\right)_{i,j} + \frac{h_{i+1}^2}{2!}\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j} + \ldots \quad (5.3.1)$$

$$U_{i-1,j} = U_{i,j} - h_i\left(\frac{\partial u}{\partial x}\right)_{i,j} + \frac{h_i^2}{2!}\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j} - \ldots \quad (5.3.2)$$

where h_i is as given in (5.1.1). Multiplying the equations (5.3.1) and (5.3.2) by h_i and h_{i+1} respectively we obtain, on addition of the results, the following expression

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j} = \frac{2(h_i U_{i+1,j} - (h_i + h_{i+1}) U_{i,j} + h_{i+1} U_{i-1,j})}{h_i h_{i+1} (h_i + h_{i+1})} + O(h). \quad (5.3.3)$$

Here we are assuming that the mesh lengths are such that

$$h = O(h_1) = O(h_2) = \dots = O(h_N) . \quad (5.3.4)$$

As was mentioned earlier, the use of variable step lengths in approximating derivatives is likely to yield errors of a larger order of magnitude than for the comparable approximations with constant step lengths. This is borne out by the error of $O(h)$ in (5.3.3) when we know that for the equivalent approximation using constant steps the error is $O(h^2)$.

This problem also has to be considered for the similar finite difference approximation to $\partial u|\partial x$. In this case the constant step length approximation has an error $O(h^2)$ and by multiplying (5.3.1) and (5.3.2) respectively by h_i^2 and h_{i+1}^2 we find that the error can be kept of the same order. We thus obtain the following expression

$$\left(\frac{\partial u}{\partial x}\right)_{i,j} = \frac{h_i^2 U_{i+1,j} - (h_i^2 - h_{i+1}^2) U_{i,j} - h_{i+1}^2 U_{i-1,j}}{h_i h_{i+1} (h_i + h_{i+1})} + O(h^2). \quad (5.3.5)$$

The approximations (5.3.3) and (5.3.5) are now used to replace the derivatives in (3.1.1). After suitable re-arrangement of the terms these substitutions result in the finite difference scheme

$$(-2a\alpha r_i+b\alpha h_{i+1}r_i)U_{i-1,j+1}+(1+4a\alpha\nu_i+2b\alpha\nu_i(h_i-h_{i+1})-\alpha ck^2)U_{i,j+1}$$

$$+(-2a\alpha s_i-b\alpha h_i s_i)U_{i+1,j+1}$$

$$=\{2a(1-2\alpha)r_i-b(1-2\alpha)h_{i+1}r_i\}U_{i-1,j}$$

$$+\{2-4a(1-2\alpha)\nu_i-2b(1-2\alpha)\nu_i(h_i-h_{i+1})+ck^2(1-2\alpha)\}U_{i,j}$$

$$+\{2a(1-2\alpha)s_i+b(1-2\alpha)h_i s_i\}U_{i+1,j}$$

$$-(-2a\alpha r_i+b\alpha h_{i+1}r_i)U_{i-1,j-1}-(1+4a\alpha\nu_i+2b\alpha\nu_i(h_i-h_{i+1})-\alpha ck^2)U_{i,j-1}$$

$$-(-2a\alpha s_i-b\alpha h_i s_i)U_{i+1,j-1} \quad (i=1,2,\ldots,(N-1)) \qquad (5.3.6)$$

where

$$r_i=\frac{k^2}{h_i(h_i+h_{i+1})},\quad s_i=\frac{k^2}{h_{i+1}(h_i+h_{i+1})},\quad \nu_i=\frac{k^2}{2h_i\,h_{i+1}}$$

(5.3.7)

5.4 Truncation Errors and Stability Conditions

Using the Taylor series expansions (5.3.1) and (5.3.2) we can easily obtain the following truncation errors

$$k^2\Bigg[c^2k^2f(\theta)u+\{2bck^2f(\theta)+\frac{1}{3}c^2k^2\theta(h_i-h_{i+1})\}\frac{\partial u}{\partial x}$$

$$+\left\{k^2(2ac+b^2)f(\theta)+\frac{1}{6}b(h_i-h_{i+1})(1+ck^2\theta)-\frac{1}{3}c^2k^2\theta\frac{h_i^3+h_{i+1}^3}{h_i+h_{i+1}}\right\}\frac{\partial^2 u}{\partial x^2}$$

$$+\left\{2abk^2f(\theta)+\frac{1}{6}k^2\theta(h_i-h_{i+1})(2ac+3b^2)-\frac{1}{18}c\left(\frac{h_i^4-h_{i+1}^4}{h_i+h_{i+1}}\right)\right.$$

$$\left.-\frac{1}{6}bck^2\theta\left(\frac{h_i^3+h_{i+1}^3}{h_i+h_{i+1}}\right)\right\}\frac{\partial^3 u}{\partial x^3}+\left\{a^2k^2f(\theta)+\tfrac{1}{2}abk^2\theta(h_i-h_{i+1})\right.$$

$$\left.+\frac{1}{12}a\left(\frac{h_i^3+h_{i+1}^3}{h_i+h_{i+1}}\right)(1-2ck^2\theta)+\frac{1}{24}b\left(\frac{h_i^4-h_{i+1}^4}{h_i+h_{i+1}}\right)-\frac{1}{72}c\left(\frac{h_i^5+h_{i+1}^5}{h_i+h_{i+1}}\right)\right\}\frac{\partial^4 u}{\partial x^4}\Bigg]$$

(5.4.1)

for the scheme (5.2.9) incorporating splines and

$$k^2\Bigg[c^2k^2f(\alpha)U+2bck^2f(\alpha)\ \frac{\partial u}{\partial x} + k^2(2ac+b^2)f(\alpha)\ \frac{\partial^2 u}{\partial x^2}$$

$$+ \{2abk^2f(\alpha)+\frac{1}{3}\ a\ (h_i-h_{i+1})-\frac{1}{6}\ bh_ih_{i+1}\}\ \frac{\partial^3 u}{\partial x^3}$$

$$+\left\{a^2k^2f(\alpha)+\frac{1}{24}\ bh_ih_{i+1}(h_i-h_{i+1})-\frac{1}{12}\ a\left(\frac{h_i^3+h_{i+1}^3}{h_i+h_{i+1}}\right)\right\}\ \frac{\partial^4 u}{\partial x^4}\Bigg] \tag{5.4.2}$$

for the fully implicit finite difference scheme (5.3.6). In both these expressions $f(\theta)$ and $f(\alpha)$ are as given by (3.3.4).

Due to the variable step lengths in the x direction the usual von Neumann stability analysis cannot be employed on either of the schemes (5.2.9) or (5.3.6). This is because the procedure depends on letting the step lengths tend to zero in such a way that the mesh ratio remains fixed. Since in this case, at each mesh point, the mesh ratios depends on two different step lengths h_i and h_{i+1} then we would have to let h_i and h_{i+1} tend to zero at the same rate.

As the two schemes (5.2.9) and (5.3.6) naturally reduce to the equivalent approximations for constant step lengths when $h_i = h_{i+1}$, then it is very likely that the stability conditions will be similar to (3.4.5) and (3.6.3). We have in fact, found by numerical examples, that for values of α and θ greater than or equal to ½ then the schemes (5.2.9) and (5.3.6) are always stable. In addition we have discovered that values of α,θ and r which gave stable solutions for the constant step length case do likewise for problems employing a variable mesh when the mesh ratio r is taken as $k^2|\min\{h_ih_{i+1}\}$. Using the notation as given in

Saul'yev (1964) we can therefore express the stability conditions in the following algebraic form

(a) for the scheme incorporating splines (5.2.9)

(i) if $\theta \geq \frac{1}{4}$ it is unconditionally stable.

(ii) if $\theta < \frac{1}{4}$ it is stable provided

$$\frac{k^2}{\min\{h_i h_{i+1}\}} \leq \frac{1}{3a(1-4\theta)} \tag{5.4.3}$$

(b) for the finite difference scheme (5.3.6)

(i) if $\alpha \geq \frac{1}{4}$ it is unconditionally stable

(ii) if $\alpha < \frac{1}{4}$ it is stable provided

$$\frac{k^2}{\min\{h_i h_{i+1}\}} \qquad \frac{1}{a(1-4\alpha)} \;. \tag{5.4.4}$$

In the following chapter on numerical case studies we use values of α, θ, k, h_i and h_{i+1} which either satisfy or violate the above stability conditions.

5.5. Difference Schemes with variable α and θ.

In certain simpler problems with constant coefficients the truncation errors for the schemes (3.2.14) and (3.5.2) can be reduced to simpler forms by choosing α and θ to eliminate some of the terms. This unfortunately is not so simple when the mesh lengths are varying in the x direction since we require a different value of α or θ to eliminate these terms at each mesh point. If we are therefore to simplify the truncation error and reduce it a lower order of magnitude we must derive difference schemes comparable to (5.2.9) and (5.3.6) with variable values of α and θ.

For the fully implicit finite difference scheme this is simply done by replacing α by α_i in the initial approximation to (3.1.1) as follows

$$\frac{U_{i,j-1}-2U_{i,j}+U_{i,j+1}}{k^2} = a\left\{\alpha_i\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j-1} +(1-2\alpha_i)\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j} +\alpha_i\left(\frac{\partial^2 u}{\partial x^2}\right)_{i,j+1}\right\}$$

$$+b\left\{\alpha_i\left(\frac{\partial u}{\partial x}\right)_{i,j-1} +(1-2\alpha_i)\left(\frac{\partial u}{\partial x}\right)_{i,j} +\alpha_i\left(\frac{\partial u}{\partial x}\right)_{i,j+1}\right\}$$

$$+c\left\{\alpha_i U_{i,j-1}+(1-2\alpha_i)U_{i,j}+\alpha_i U_{i,j+1}\right\}$$

$$(i = 1,2, \ldots, (N-1)) \quad (5.5.1)$$

where $\partial^2 u|\partial x^2$ and $\partial u|\partial x$ are respectively given by (5.3.3) and (5.3.5). As would be expected this results in the scheme (5.3.6) with α replaced by α_i.

If we adopt this procedure for the finite difference scheme incorporating splines we are unable to eliminate the $M_{i,j}$ and $L_{i,j}$ values which result from following the steps in section 3.2. However since the nine-point formula (5.2.9) is applied independently at each of the mesh points (ih,jk) we can use a different value of the parameter θ at each of these points. This simply means we replace θ by θ_i in (5.2.9) and thus the method of obtaining solutions will be as given in the iterative procedure described in steps (i) - (viii) of this section.

The truncation errors for these new schemes will be the same as (5.4.1) and (5.4.2) again with θ and α replaced by θ_i and α_i,; respectively. To illustrate the usefulness of the variable θ and α schemes consider solving the wave equation using (5.2.9). The truncation error for this scheme becomes, on setting a=1, b=c=0,

$$k^2\left\{k^2\left(\frac{1}{12}-\theta\right) + \frac{1}{12}\left(\frac{h_i^3+h_{i+1}^3}{h_i+h_{i+1}}\right)\right\}\frac{\partial^4 u}{\partial x^4}. \quad (5.5.2)$$

With the exception of letting $\theta = 1/12$ we cannot simplify this expression any further. If we now consider using the variable θ extension of (5.2.9) then the truncation error (5.5.2) will be unchanged except for θ being replaced by θ_i. This whole expression ((5.2.9) with $\theta=\theta_i$) can now be eliminated by choosing

$$\theta_i = \frac{1}{12}\left[1 + \frac{h_i^3+h_{i+1}^3}{k^2(h_i+h_{i+1})}\right] \tag{5.5.3}$$

By using values of θ_i given by (5.5.3) we will have no contribution to the truncation error from the 4^{th} order term and we therefore need only consider lower order terms. This process is equal ly applicable to the ordinary finite difference approximation and is discussed again in the following chapter.

Unfortunately it is sometimes the case that the value of α_i or θ_i which eliminates a derivative term also gives unstable solutions when it is used in the particular scheme concerned. We must therefore ensure that the parameters α_i and θ_i always satisfy the stability conditions (5.4.4) and (5.4.3), respectively, where α and θ are again replaced by α_i and θ_i.

5.6 Equations with Variable Coefficients

The use of unequal step lengths can also be adapted to the solution of equations with varying coefficients. The method of obtaining both the finite difference scheme incorporatint splines and the fully implicit finite difference approximation is simply a combination of the procedures described in chapter 4 and sections 2 and 3 of this chapter.

With the exception of letting $\theta = 1/12$ we cannot simplify this expression any further. If we now consider using the variable θ extension of (5.2.9) then the truncation error (5.5.2) will be unchanged except for θ being replaced by θ_i. This whole expression ((5.2.9) with $\theta=\theta_i$) can now be eliminated by choosing

$$\theta_i = \frac{1}{12}\left[1 + \frac{h_i^3 + h_{i+1}^3}{k^2(h_i+h_{i+1})}\right] \qquad (5.5.3)$$

By using values of θ_i given by (5.5.3) we will have no contribution to the truncation error from the n^{th} order term and we therefore need only consider lower order terms. This process is equally applicable to the ordinary finite difference approximation and is discussed again in the following chapter.

Unfortunately it is sometimes the case that the value of α_i or θ_i which eliminates derivative terms also gives unstable solutions when it is used in the particular scheme concerned. We must therefore ensure that the parameters α_i and θ_i always satisfy the stability condition (5.4.4) and (5.4.6) respectively, where α and θ are again replaced by α_i and θ_i.

5.6 Equations with Variable Coefficients

The use of unequal step lengths can also be adapted to the solution of equations with varying coefficients. The method of obtaining both the finite difference scheme incorporating splines and the fully implicit finite difference approximation is simply a combination of the procedures described in chapter 4 and sections 2 and 3 of this chapter.

For the scheme incorporating splines we thus obtain the tri-diagonal approximation to (4.1.1) at (ih,jk)

$$\{\beta_i-(6s_ia_{i,j+1}+3h_ir_i\rho_{j+1}+h_ih_{i+1}r_ic_{i+1,j+1})\theta\}\ U_{i+1,j+1}$$

$$+2\{1+(6\nu_ia_{i,j+1}-k^2c_{i,j+1})\theta\}\quad U_{i,j+1}$$

$$+\{\phi_i-(6r_ia_{i,j+1}-3h_{i+1}s_i\rho_{j+1}+h_ih_{i+1}s_ic_{i-1,j+1})\theta\}U_{i-1,j+1}$$

$$=\{2\beta_i+(6s_ia_{i,j}+3h_ir_i\rho_j+h_ih_{i+1}r_ic_{i+1,j})(1-2\theta)\}\ U_{i+1,j}$$

$$+\ 2\{2\ -(6\nu_ia_{i,j}-k^2c_{i,j})(1-2\theta)\}\ U_{i,j}$$

$$+\{2\phi_i+(6r_ia_{i,j}-3h_{i+1}s_i\rho_j+h_ih_{i+1}s_ic_{i-1,j})(1-2\theta)\}\ U_{i-1,j}$$

$$-\{\beta_i-(6s_ia_{i,j-1}+3h_ir_i\rho_{j-1}+h_ih_{i+1}r_ic_{i+1,j-1})\theta\}U_{i+1,j-1}$$

$$-2\{\ 1+(6\nu_ia_{i,j-1}-k^2c_{i,j-1})\theta\}U_{i,j-1}$$

$$-\{\phi_i-(6r_ia_{i,j-1}-3h_{i+1}s_i\rho_{j-1}+h_ih_{i+1}s_ic_{i-1,j-1})\theta\}U_{i-1,j-1}$$

$$-\ \frac{k^2}{2}(h_{i+1}-h_i)\Big[\theta\rho_{j-1}M_{i,j-1}+(1-2\theta)\rho_jM_{i,j}+\theta\rho_{j+1}M_{i,j+1}\Big] \qquad (5.6.1)$$

$$(i\ =\ 1,2,\ \ldots,(N-1))$$

where β_i, ϕ_i, r_i, s_i and ν_i are as given for (5.2.9) and

$$\rho_j\ =\ a'_{i,j}\ +\ b_{i,j} \qquad (5.6.2)$$

Due to the $M_{i,j+1}$ term in (5.6.1) we must again use the iterative procedure described in section 5.2.

A similar approximation to (4.1.1) can also be derived for the fully implicit finite scheme. This is again easier to use in practice since no iterative procedure is necessary.

As was found previously, the von Neumann method for examining stability can not be applied directly to schemes having either variable coefficients or unequal step lengths. However, as was the case for these earlier equations, the von Neumann stability analysis can be applied locally to give the following conditions:

For the scheme incorporating splines

(i) provided $\theta \geq \frac{1}{4}$ it is unconditionally stable and

(ii) if $\theta < \frac{1}{4}$ it is stable when

$$\frac{k^2}{\min\{h_i h_{i+1}\}} \leq \frac{1}{3a(x,t)(1-4\theta)} . \tag{5.6.3}$$

Similarly for the implicit finite difference approximation

(i) if $\alpha \geq \frac{1}{4}$ it is unconditionally stable and

(ii) if $\alpha < \frac{1}{4}$ it is stable provided

$$\frac{k^2}{\min\{h_i h_{i+1}\}} \leq \frac{1}{a(x,t)(1-4\alpha)} . \tag{5.6.4}$$

The conditions (5.6.3) and (5.6.4) must again be satisfied independently at each point of the solution domain.

CHAPTER 6

Case Studies

6.1 Case Study 1

As an illustration of the constant coefficient methods we shall consider the well-known wave equation ((3.1.1) with a=1, b=c=0)

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} \quad (0 \leq x \leq 1,\ t>0) \tag{6.1.1}$$

together with the initial conditions

$$u(x,0) = \sin \pi x \quad ; \quad \frac{\partial u}{\partial t}(x,0) = 0 \tag{6.1.2}$$

and the boundary conditions

$$U(0,t) = U(1,t) = 0 \tag{6.1.3}$$

(a) Uniform step lengths

Numerical solutions to this problem have been found using both the finite difference scheme incorporating splines (3.2.14) and the fully implicit finite difference scheme (3.5.2). The truncation errors for these approximations were given to fourth order in chapter 3 and it is not a difficult task to show that, for the simpler wave equation, they take the form

$$k^2h^2\left[\frac{1}{12}(1+r^2-12r^2\theta)\frac{\partial^4 u}{\partial x^4}+\frac{1}{360}h^2((r^4-1)+5(1-6r^2\theta)(r^2+1))\frac{\partial^6 u}{\partial x^6}\right] \tag{6.1.4}$$

for the splines scheme and

$$k^2h^2\left[\frac{1}{12}\left((r^2-1)-\alpha r^2\right)\frac{\partial^4 u}{\partial x^4}+\frac{1}{360}h^2\left((r^4-1)-30\alpha r^2(r^2+1)\right)\frac{\partial^6 u}{\partial x^6}\right] \tag{6.1.5}$$

for the finite difference scheme, where terms to 6^{th} order are now given.

These truncation errors are both $O(k^2h^2+k^4)$ but as mentioned in section 3.6 they can be reduced to a lower order of magnitude by optimally choosing values of the parameters α, θ and r. Thus for the expression (6.1.4), if we choose θ such that

$$\theta = \frac{1}{12r^2}(1+r^2) \qquad (6.1.6)$$

then the 4^{th} order term is eliminated and the truncation error is $O(k^2h^4+k^6)$. Using this value of θ, if we now choose r such that the 6^{th} order term in (6.1.4) equals zero then it can be shown, by evaluating terms up to the 8^{th} order derivative, that the truncation error is now $O(k^2h^6+k^8)$. This optimal value of r is simply r=1.

The truncation error (6.1.5) for the finite difference scheme can also be reduced to lower order of magnitude using the same method. By choosing

$$\alpha = \frac{1}{12r^2}(r^2-1) \qquad (6.1.7)$$

the truncation error becomes $O(k^2h^4+k^6)$ and with this value of α and r=1 it is reduced further still to $O(k^2h^6+k^8)$.

As can be seen, although the expressions for α and θ given by (6.1.6) and (6.1.7) are different, the resulting values of r which eliminated the 6^{th} order terms are the same. It should also be noted that the truncation errors using these optimal values of r, θ and α are also of the same orders.

This procedure of reducing the order of a truncation error has also been used by Crandall (1955) where he employed a finite difference scheme analogous to (3.5.2) in the numerical solution of the heat conduction equation.

If we now substitute the value r=1 into (6.1.6) and (6.1.7) we see that the values $\theta=1/6$ and $\alpha=0$, together with this value of r, reduce respectively the scheme incorporating splines and the finite difference scheme to $O(k^2h^6+k^8)$. As these values of θ and α are less than ¼ we do not have unconditional stability and must therefore check that the conditions (3.4.5) and (3.6.3) are satisfied. In this case with r=1 the conditions are satisfied but in using this value of r we place severe restrictions on the choices of h and k. For example, if we require a small mesh length in the space direction then we must use an identically small value for the mesh length in the time direction.

To overcome this restriction on the mesh ratio r it is advisable to just eliminate the 4^{th} order terms in the truncation errors (6.1.4) and (6.1.5) using values of θ and α from (6.1.6) and (6.1.7). In doing this in practise we find we have a slightly larger error but are able to use a wider range of values of h and k.

Once numerical solutions to (6.1.1) have been found using both of the schemes (3.2.14) and (3.5.2) we will compare their respective accuracies in two ways. Firstly by obtaining numerical values of the truncation errors (6.1.4) and (6.1.5) at each of the mesh points and then by comparing the maximum absolute values out to a certain point in time. This error we denote by TE_{SC} for the splines scheme when approximating an equation with constant coefficients and TE_{DC} for the ordinary difference scheme when approximating equations of the same form. The second method of examining the accuracy of either scheme is done by finding the maximum absolute value of the difference between the numerical solutions and the analytic solution to the same point in time. For this error, when considering equations of the form given by (3.1.1), we denote the error for the splines scheme by E_{SC} and similarly by E_{DC} for the finite difference approximation. It should be noted that for some problems the analytic solution will be unknown and therefore only the errors TE_{SC} and TE_{DC} can be examined.

It is easily shown by either separation of variables or Laplace Transforms that the analytic solution to the wave equation (6.1.1), with the prescribed initial and boundary conditions (6.1.2) and (6.1.3), is given by

$$U(x,t) = \sin\pi x \cos\pi t \, . \qquad (6.1.8)$$

The numerical solutions to (6.1.1) have been evaluated for a range of values of the parameters α and θ and for various mesh lengths h and k. The results given in tables 1 and 2 have been found when the space step length h has been taken as 0.1 and where computations for both schemes have been taken out to t=0.5. The truncation errors TE_{SC} and TE_{DC} have been evaluated to 4^{th} order using the expressions given in section 3.8.

It can be seen from tables 1 and 2 that for k = 0.05 the values of $\theta = 5/12$ and $\alpha = -1/4$ reduce their respective truncation errors to zero when derivative terms upto and including 4^{th} order are evaluated. This is also the case when $\theta = 17/12$ and $\alpha = -15/12$ are used with k = 0.025. The reason for this is simply because θ and α have been chosen such that the expressions (6.1.6) and (6.1.7) are satisfied.

As an illustration of the effect of eliminating both 4^{th} and 6^{th} order terms from the truncation errors we have used $h=k=0.1$ and $\theta = 1/6$ in (3.2.14) and the same values of h and k with $\alpha=0$ in (3.5.2). On examining the same errors as given in tables 1 and 2 we find that $E_{SC} = 2.20\times10^{-6}$ and $E_{DC} = 5.40\times10^{-6}$ where errors to $t = 0.5$ have again been considered. The truncation errors TE_{SC} and TE_{DC} are $O(k^2h^6+k^8)$ with these values of θ, α and r and will therefore be of negligable size.

(b) Variable step lengths

For the second part of this case study we again consider the wave equation (6.1.1) which, together with the initial and boundary conditions (6.1.2) and (6.1.3), is now to be solved by varying the size af the step lengths in the space direction. Initially these numerical solutions have been found by using both (5.2.9) and (5.3.6), these having constant values for the parameters θ and α at each of the mesh points.

In this particular problem the use of variable step lengths is not very advantageous since neither the function value initial condition nor the shape of the plotted numerical solutions are of the form illustrated by Fig. 1. It does however illustrate the usefulness of schemes such as (5.2.9) which allow the mesh points to be unequally spaced.

For simplicity we have initially arranged the step lengths as if they were to be used on two different problems. Firstly if the "peak", as shown in fig. 1, was of the form

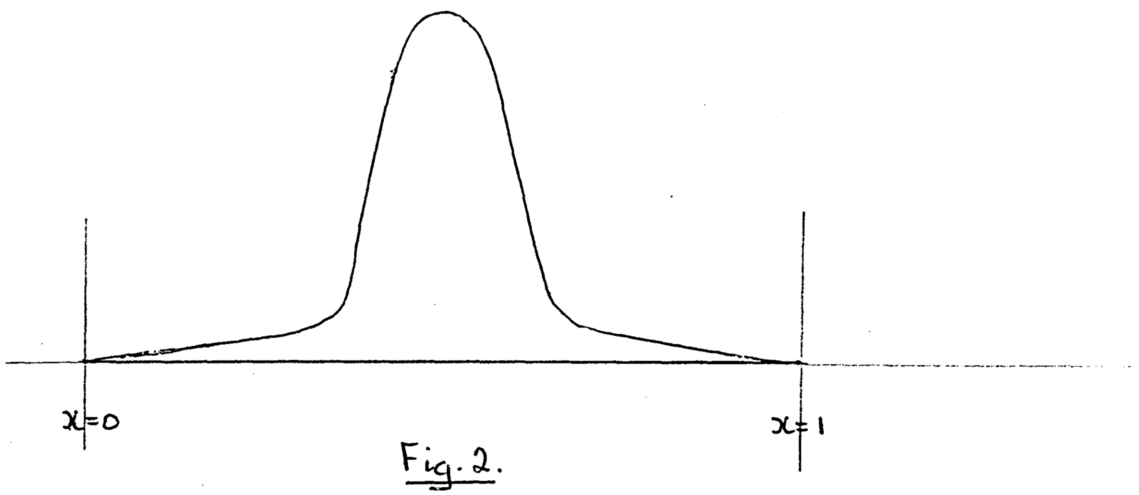


Fig. 2.

then the step lengths near the centre of the range $[0,1]$ have to be smaller than those in the remainder. For our first numerical computations we use the following x values as internal mesh points

0.15, 0.30, 0.40, 0.45, 0.50, 0.55, 0.60, 0.70, 0.85.

For future reference we will call these values mesh set A.

As a comparison the second variable mesh has been chosen as if the solutions are varying rapidly near each of the boundaries x=0 and x=1. The smaller step lengths have thus been arranged near these boundaries and are given by the following internal mesh values

0.075, 0.15, 0.25, 0.35, 0.50, 0.65, 0.75, 0.85, 0.925

which we refer to as mesh set B. As can be seen both this mesh and mesh set A have nine internal mesh points. The reason for this is that in part (a) of this case study numerical solutions are derived to (6.1.1) using $h = 0.1$, in which case there were again nine constantly spaced mesh points. Thus the solutions obtained here can be suitably compared with those obtained with uniform step lengths having the same number of internal mesh points.

Using both mesh sets A and B the numerical solutions to (6.1.1) are evaluated by the scheme (5.2.9) which incorporated cubic splines and also by the more common finite difference approximation (5.3.6). To examine their respective accuracies we obtain numerical values for the same errors as given in tables 1 and 2. We again denote the maximum absolute value of the truncation errors by TE_{SC} for the splines schemes and TE_{DC} for the difference scheme. Similarly E_{SC} and E_{DC} are used to represent the maximum absolute difference between the numerical solutions and the analytic solutions which are again found from (6.1.8). The results for mesh set A are given in tables 3 and 4 and those for mesh set B in tables 5 and 6 where the time step length k has again been taken as 0.05 and 0.025.

As was indicated in section 5.5 the truncation errors for both variable step length schemes can be considerably simplified by choosing values of the parameters θ and α which eliminate the major contributing terms. Since these expressions for the truncation errors vary with the size of the step lengths then it was shown that only by making use of schemes with variable θ and α can any significant simplification be made. For the splines scheme with variable θ, the choice of θ_i as given by (5.5.3) results in the elimination of the 4^{th} order term of the truncation error when considering the wave equation. Similarly for the variable step length finite difference scheme with variable parameter α the 4^{th} order term can again be eliminated by choosing

α_i such that

$$\alpha_i = \frac{1}{12}\left[1 - \frac{(h_i^3 + h_{i+1}^3)}{k^2(h_i+h_{i+1})}\right] \qquad (6.1.9)$$

It should be noted that, whereas for the splines truncation error the choice of θ_i subject to (5.5.3) meant no terms upto and including 4^{th} order had to be considered the values of α_i given by (6.1.9) still leave a 3^{rd} order term in the expression for the truncation error which must be taken into account. We can therefore reasonably expect that the errors for the splines scheme will be smaller than those for the ordinary finite difference approximation when the step lengths are unequally spaced, and when values of θ_i and α_i are obtained from (5.5.3) and (6.1.9) respectively. This is borne out by numerical results when both mesh set A and mesh set B are used in the solution of the wave equation. Using the previously adopted notation for errors we find that these results are as given in table 7 where the truncation errors have been evaluated to 4^{th} order only and K has been taken as equal to 0.05.

The choice of mesh points using variable step length schemes is obviously not limited to our mesh set A and mesh set B arrangements. We can in fact arrange the mesh points in any way that may be helpful for solving a particular problem, although we must remember that the order of the truncation error is dependent on the size of the spacing between these mesh points. As a further example on the use of our variable step length schemes we have used as mesh points both the zeros and extrema of the shifted Chebyshev polynomials. These points have been shown to be advantageous in the solution of ordinary differential equations by several authors, for example Picken (1970), Kizner (1966). The extreme values of the Chebyshev polynomial have in fact been used by El-Gendi (1975) for solving parabolic partial differential equations where his method was one of transforming the parabolic equation to a system of ordinary differential equations.

Within the range $0 \leq x \leq 1$ the zeros and extrema of the shifted Chebyshev polynomial

$$T_n^*(x) = \text{Cos}\,(n\text{Cos}^{-1}(2x-1)) \tag{6.1.10}$$

are respectively given by

$$x_i = \frac{1}{2}\left[1 + \text{Cos}\left(\frac{2i+1}{2n}\right)\pi\right] \quad (i = 0,1,\ \ldots,(n-1)) \tag{6.1.11}$$

and

$$x_i = \frac{1}{2}\left[1 + \text{Cos}\left(\frac{i\pi}{n}\right)\right] \quad (i = 1,2,\ \ldots,n) \tag{6.1.12}$$

where n denotes the number of internal mesh points. In order to compare the results produced using the above points with the results of earlier calculations we again use nine internal mesh points for both the splines and finite difference schemes. Thus using a time step length k of 0.05 we solve the wave equation by employing both (5.2.9) and (5.3.6) on a mesh with space step lengths given by (6.1.11) and (6.1.12). Some of the maximum absolute errors E are given in table 8 and as can be seen they are of similar magnitude to those of tables 3 and 5 where mesh sets A and B are used.

We can again hope to improve on these results by employing the variable parameter extensions to (5.2.9) and (5.3.6), and then by using values of the variable parameters θ_i and α_i which are obtained from (5.5.3) and (6.1.9), respectively. In doing this terms up to and including 4^{th} order are again eliminated from the splines and finite difference truncation errors with the exception of a 3^{rd} order term remaining in the latter. Results obtained by employing this procedure are given in table 8 where truncation errors have been quoted to 4^{th} order terms.

Due to time considerations this is the only case study we have computationally considered where variable step lengths have been used. Obviously the theory derived in chapter 5 can easily be extended in principal to solve more complex problems.

6.2 Case Study 2

As a second example of solving equations with constant coefficients we consider the general hyperbolic partial differential equation (3.1.1) with a=b=c=1 so that

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial u}{\partial x} + U . \qquad (0\leqslant x\leqslant 1,\ t>0) \qquad (6.2.1)$$

For simplicity we again consider (6.2.1) to have the function value initial and boundary conditions given by (6.1.2) and (6.1.3) respectively.

Numerical solutions to (6.2.1) are found using both the scheme (3.2.14) which incorporates splines and the ordinary finite difference approximation (3.5.2). Using the notation described in the previous case study we again let TE_{SC} and TE_{DC} denote the numerical values of the truncation errors for the splines and difference schemes respectively since (6.2.1) has constant coefficients. In order to examine the effect of the lower order terms in (6.2.1) we again use a space step length h of 1/10 thus enabling us to compare the results produced here with those of the previous case study. Similarly the time step length K has also been taken as 0.05 and 0.025 and the absolute maximum values for the errors obtained out to t=0.5. These ℓ_∞ norm values for the truncation errors are given in tables 9 and 10 where various values of the parameters θ and α are given. The truncation errors shown are again evaluated to 4^{th} order derivative terms.

In section 3.6 it was suggested that the truncation errors (3.3.4) and (3.6.1) could be considerably simplified by choosing values of the parameters θ and α such that they reduce the function f to zero. This was done by obtaining values of θ and α from (3.6.2) which in this case, with a=b=c=1, becomes

$$\theta = \alpha = \frac{1}{12+k^2} \qquad (6.2.2)$$

Using k = 0.05 we here solve (6.2.1) using both the splines and finite difference approximations with θ and α being found from (6.2.2). The maximum absolute values of the truncation errors out to t=0.5 are given in table 9.

As with case study 1 we also use both schemes (3.2.14) and (3.5.2) with values of θ, α and r that violate the stability conditions (3.4.5) and (3.6.3). This has been done purely to test the conditions and will be used in a later case study with variable coefficients where the von Neumann stability analysis is not as stringent as that for equations with constant coefficients.

6.3 Case Study 3

As a more practical application of the use of the constant coefficient theory described in chapter 3 we consider the equation

$$\frac{1}{c_1^2}\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - \mu^2 u , \qquad (6.3.1)$$

which is a dispersive wave equation. The parameters c_1 and μ are constant with dimensions of velocity and 1/length respectively.

Equation (6.3.1) occurs in several areas of mathematical physics. It is, for example, the one-dimensional form of the Klein-Gordon equation which is of some interest in the field of quantum mechanics. In this context $u(x,t)$ is the wave function, c_1 is the velocity of light and

$$\mu = Mc_1/h , \; h = \text{Planck's constant}/2\pi , \qquad (6.3.2)$$

where M is the mass of the particle.

The equation has also been derived by Richtmyer (1960) and by Briscoe and Kovitz (1968) in work on the development of corrugations in the shape of plane shock fronts. The governing equation for pressure fluctuations $p(x,t)$ in the gas behind the advancing shock front is given here by

$$\frac{1}{c_1^2}\frac{\partial^2 p}{\partial t^2} = \frac{\partial^2 p}{\partial x^2} + \frac{\partial^2 p}{\partial y^2} , \qquad (6.3.3)$$

where the constant c_1 is now the sonic speed. With the assumption that pressure variations are of the form

$$p(x,y,t) = u(x,t)e^{i\mu y} \qquad (6.3.4)$$

equation (6.3.3) reduces to equation (6.3.1).

A simple physical situation which can be modelled by (6.3.1) is that of a vibrating string embedded in an elastic medium. Such a system will exhibit all the features of the more complex systems mentioned previously.

If the displacement of the string from its equilibrium position is given by $u(x,t)$ then the restoring force on an element of length dx will consist of two parts. Firstly a force due to the tension T in the string and secondly a force $vudx$ due to the elasticity of the medium. The constant v will depend on the elastic properties of the medium. The differential equation for $u(x,t)$ is now given by (6.3.1) with

$$\mu^2 = {}^{v}/T \tag{6.3.5}$$

and

$$c_1 = T/\rho \tag{6.3.6}$$

where ρ is the mass per unit length of the string.

Now suppose that the string is embedded in an elastic medium with end fixed at $x=0$ and $x=\ell$. Let the string in an equilibrium position be subject to an initial velocity V_o acting perpendicular to the string. The displacement $u(x,t)$ now satisfies (6.3.1) subject to the boundary conditions

$$u(0,t) = 0, \quad u(\ell,t) = 0 \tag{6.3.7}$$

and the initial conditions

$$u(x,0) = 0 \quad , \quad \frac{\partial u}{\partial t}(x,0) = V_o \tag{6.3.8}$$

The equation (6.3.1) and the above conditions may be expressed in a nondimensional form by introducing the nondimensional primed variables

$$x' = x/\ell, \quad t' = c_1 t|\ell , \quad u' = c_1 u/(V_o \ell) \ . \tag{6.3.9}$$

Equation (6.3.1) thus becomes on dropping the primes

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - \lambda^2 u \tag{6.3.10}$$

subject to

$$u(0,t) = 0, \quad u(1,t) = 0 \tag{6.3.11}$$

and

$$u(x,0) = 0 \ , \ \frac{\partial u}{\partial t}(x,0) = 1 \tag{6.3.12}$$

where

$$\lambda^2 = v/\rho \tag{6.3.13}$$

The equation (6.3.10) is solved numerically by both the cubic spline technique (3.2.14) and the well known finite difference approximation (3.5.2). The value λ is taken to be unity without any loss of generality. As with the two previous case studies various values of θ, α, k and h are used in the computations. Tables 11 and 12 show the maximum absolute values for the truncation errors (3.3.3) and (3.6.1), each of which is evaluated in the program at every mesh point.

From the tables we see that values of the parameters θ and α in the vicinity of $5/12$ and $-\frac{1}{3}$, respectively, best minimise the numerical values of the truncation errors when k=0.05 is used. For a time step length of 0.025 the corresponding values of θ and α are 17/12 and -15/12 respectively. At first sight these results are hardly surprising in view of the fact that the above values of θ and α reduce the leading term of the truncation errors (3.3.3) and (3.6.1) to zero when solving the one-dimensional wave equation, this being very similar in nature to (6.3.10).

The procedure for reducing terms in the truncation errors to zero for the wave equation is discussed earlier in section 6.1. It is equally applicable in this case study and it is easily shown that

$$\theta = \frac{a(1+ar^2) - ch^2/6}{ack^2(2+ar^2)+12a^2r^2} \tag{6.3.14}$$

and

$$\alpha = \frac{(ar^2-1)}{ar^2(12+ck^2)} \tag{6.3.15}$$

reduce the 4^{th} order terms of the truncation errors (3.3.3) and (3.6.1), respectively, to zero. Thus by choosing values of θ and α from the above we find that for h=0.1 and k=0.05

$$TE_{SC} = 3.96 \times 10^{-5} \tag{6.3.16}$$

$$TE_{DC} = 5.24 \times 10^{-5} \tag{6.3.17}$$

where the notation TE_{SC} and TE_{DC} is as used in the previous case studies. It should be noted that the values of θ and α from (6.3.14) and (6.4.15) only eliminate the 4^{th} order terms from the truncation errors and that the terms in u, $\partial u|\partial x$, $\partial^2 u|\partial x^2$ and $\partial^3 u|\partial x^3$ still have to be considered when solving equations other than those with b=c=0.

We can also hope to improve the accuracy of the results for both splines and finite difference schemes by again choosing θ and α such that they reduce the function f in (3.3.3) and (3.6.1) to zero. For this particular case study with $\lambda=1$ we thus choose θ and α from the following expression

$$\theta = \alpha = \frac{1}{12-k^2} \qquad (6.3.18)$$

Results for both schemes using values of θ and α from (6.3.18) are given in table 11 where a time step length of k = 0.05 has been employed.

The use of (6.3.18) means that the truncation errors for this case study are of higher order than they would normally be. This would lead one to suppose that these values of θ and α would produce the best ℓ_∞ approximation of the truncation errors. However, as can be seen from table 11, this is not the case. This anomaly indicates the possible pitfalls of depending completely on truncation errors to indicate accuracy. In this case study the analytic solution is not easily derivable so no other choice of examining accuracy prevails.

6.4 Case Study 4

It was suggested in section 3.4 that the lower order terms $b\partial u|\partial x$ and cu in equation (3.1.1) have no great effect on the stability conditions (3.4.5) and (3.6.3). In this case study we test the effect of large values of the coefficient c on both the unconditional stability condition and the accuracy of the schemes (3.2.14) and (3.5.2). Setting a=1 and b=0 equation (3.1.1) becomes

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + cu \qquad (6.4.1)$$

which we consider together with the simple initial and boundary conditions given by (6.1.2) and (6.1.3) respectively. The analytical solution to this problem is given by

$$U(x,t) = \operatorname{Sin}\pi x \operatorname{Cos} (\pi^2 - C)^{\frac{1}{2}} t. \qquad (6.4.2)$$

Again using a constant space step length of h=0.1 the numerical solutions to (6.4.1) are found from the tri-diagonal splines scheme (3.2.14). As with the previous case studies we compare the results using this method with those obtained from the ordinary finite difference approximation (3.5.2). The values of c have been chosen to be negative to avoid any occurence of negative square roots in (6.4.2). Tables 13, 14 and 15 show maximum absolute numerical values of both the truncation errors and the errors between the numerical solutions and the analytic solutions. The notation used is again as given in case study 1. All results shown have been found using a value of 1/3 for the parameters θ and α.

6.5 Case Study 5

This case study is included to illustrate a practical application of the theory for variable coefficients.

A certain type of cutting device, known as the velocity transformer, has a symmetric cutting tool of varying cross-sectional area. A velocity which is varying sinusoidally in time is applied to one end of the cutting tool which is so shaped that the amplitude of the velocity is greatly increased at the other cutting edge.

If it is assumed that the velocity transmitted through the tool is dependent only on time and the distance x measured down the tool then, for small amplitude velocity variations, the differential equation for the velocity U(x,t) is

$$\rho A(x) \frac{\partial^2 u}{\partial x^2} = \frac{\partial}{\partial x}\left(EA(x) \frac{\partial u}{\partial x} \right) \tag{6.5.1}$$

where E and ρ are taken as constants, and are respectively the Young's modulus and the density of the material comprising the cutting tool, whose cross-sectional area at any point is A(x).

Consider a velocity of the form

$$U = U_o \operatorname{Sin} wt \tag{6.5.2}$$

applied to one end of the cutting tool at x=O. The other end of the cutting tool, x=ℓ, is a free end and therefore

$$\frac{\partial u}{\partial x} (\ell, t) = 0 \tag{6.5.3}$$

To specify the initial conditions for the problem suppose that

$$U(x,0) = \frac{\partial u}{\partial t}(x,0) = 0 , \qquad 0 \leqslant x \leqslant \ell. \tag{6.5.4}$$

Suppose that the ends of the cutting tool at x=O and x=ℓ have, respectively, radii r_o, r_1 and cross-sectional areas A_o, A_1. Fig. 3 illustrates such a tool in the form of a truncated cone. At any general point x the cone will have radius r(x) and cross-sectional area A(x) given by

$$A(x) = \pi r_o^2 \left\{ 1 - \left(1 - \frac{r_1}{r_o} \right) \frac{x}{\ell} \right\}^2 . \tag{6.5.5}$$

The problem can be nondimensional by introducing new variables defined by

$$x' = \frac{x}{\ell}, \quad t' = \frac{1}{\ell}\sqrt{\frac{E}{\rho}}\, t, \quad u' = \frac{u}{u_o}, \quad A' = \frac{A}{A_o}, \quad w' = \ell\sqrt{\frac{\rho}{E}}\; w.$$

In these variables the problem becomes (on dropping the primes)

$$A(x)\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x}\left(A(x)\frac{\partial u}{\partial x}\right) \qquad (6.5.6)$$

with

$$A(x) = \left\{1 - \left(1 - \frac{r_1}{r_o}\right)x\right\}^2 \qquad (6.5.7)$$

subject to

$$u(0,t) = \sin wt, \qquad (6.5.8)$$

$$\frac{\partial u}{\partial x}(1,t) = 0 \qquad (6.5.9)$$

and

$$U(x,0) = \frac{\partial u}{\partial t}(x,0) = 0 \qquad (6.5.10)$$

Rearranging (6.5.6) gives

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{1}{A}\frac{dA}{dx}\frac{\partial u}{\partial x} \qquad (6.5.11)$$

and by using (6.5.7), equation (6.5.11) becomes

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - \frac{2R}{1-Rx}\frac{\partial u}{\partial x} \qquad (6.5.12)$$

with

$$R = 1 - \frac{r_1}{r_o} \qquad (6.5.13)$$

Using Laplace Transforms the analytical solution to (6.5.12) can be obtained in the form of a series of step functions. Provided $x+t<4$ the first three terms of this solution are given by

$$U(x,t) = \frac{1}{1-Rx}\Bigg\{ \sin\left[w(t-x)\right] H(t-x)$$
$$- \left[A_1 \sin\left[w(t-x-2)-\phi\right] + \frac{2\alpha}{\alpha^2+w^2} e^{t-x-2} \right] H(t-x-2)$$
$$+ \left[A_1 \sin\left[w(t+x-2)-\phi\right] + \frac{2\alpha}{\alpha^2+w^2} e^{t+x-2} \right] H(t+x-2) \Bigg\} \qquad (6.5.14)$$

where $\alpha = \frac{R}{1-R}$, $A_1 = \sqrt{\frac{(w^2-\alpha^2)^2+4\alpha^2}{(w^2+\alpha^2)^2}}$, $\phi = \arctan\left(\frac{2\alpha}{w^2-\alpha^2}\right)$

and the function ~~$H(x,t)$~~ $H(t-x)$ is the Heaviside unit function given by

$$\text{~~}H(x,t)\text{~~ } H(t-x) = 1 \quad \text{when } x<t$$
$$= 0 \text{ elsewhere .} \qquad (6.5.15)$$

For simplicity the parameters r_o, r_1 and w have been taken as 1, 0.5 and 1 respectively. The equation (6.5.12) thus becomes

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} + \frac{2}{x-2}\frac{\partial u}{\partial x} . \qquad (6.5.16)$$

Although the analytical solution to this equation can be found directly from (6.5.14), it is sufficient in practice to test the accuracy of numerical solutions by simply using the first term of the expression (6.5.14). This is only possible when $x+t<2$ and thus for the equation (6.5.16) we can obtain theoretical solutions from the expression

$$U(x,t) = \frac{2}{2-x} \sin(t-x) \qquad (6.5.17)$$

when $x+t<2$ and $x<t$.

Equation (6.5.16), subject to (6.5.8), (6.5.9) and (6.5.10), is solved numerically both using the splines scheme (4.1.6) and the finite difference approximation (4.3.2). As for the previous case studies computations are performed using both methods with $h=0.1$ and with a range of values of K, θ and α. Numerical values of both truncation errors and differences between numerical solutions and analytic solutions are examined at all mesh points. The maximum absolute values of the truncation errors are denoted by TE_{SV} and TE_{DV} for the variable coefficient spline and difference schemes, respectively. Similarly, we represent the maximum absolute difference between the numerical solutions and the analytic solutions by E_{SV} and E_{DV}. Tables 16 and 17 give results for $k=0.05$and $k=0.025$, respectively, where errors have been examined out to $t=0.5$ in both cases.

For equations with variable coefficients it is illustrated in section 4.2 that the stability conditions for both the splines scheme and the finite difference approximation can be obtained by considering the equation to have constant coefficients. To test the validity of the conditions obtained by making this assumption we have chosen values of θ, α and k which both satisfy and violate (4.2.3) and (4.4.2). With $h=0.1$ and $k=0.05$ the stability conditions (4.2.3) and (4.4.2) are satisfied provided $\theta \geqslant -1/12$ and $\alpha \geqslant -3/4$ respectively. As can be seen from table 16 stable numerical solutions are in fact obtained when these values of θ and α are used.

As with all previous case studies the collocated spline function (3.2.2) is used to obtain solutions at points intermediate to those found at the mesh points. For this particular case study these additional solutions are found at points mid-way between mesh points. To test the accuracy of obtaining additional solutions using the spline function we again compare the results with those found from the analytic solution (6.5.17). Table 18 gives the errors between the spline solutions and the analytic solutions when $\theta=1/3$ and $k=0.05$. All results shown are for the time line $t=0.5$ and only solutions up to $x=0.45$ are given due to the nature of the Heaviside function present in the analytic solution.

CHAPTER 7

Conclusions and Extensions

Conclusions

The computational results obtained from the case studies cited in chapter 6 are best examined initially in a piecemeal fashion prior to obtaining more general conclusions.

In case study 1 we considered the well-known one-dimensional wave equation and obtained solutions by initially employing equally spaced knots. As can be seen from tables 1 and 2, for all positive values of the parameters α and θ, the scheme incorporating splines always has smaller errors than does the well-known finite difference scheme. The cases where negative values of α and θ are used will be discussed with respect to all case studies later. In an attempt to increase the accuracy of both the splines and difference schemes a method of eliminating terms from the truncation errors was also employed. By equating the fourth order terms of the truncation errors to zero we found that errors were significantly reduced in size for both schemes with the splines scheme still maintaining a smaller error than that of the difference approximation. Similarly, by reducing both the fourth and sixth order terms to zero the ℓ_∞ norm errors between the numerical solutions and the analytic solution were again decreased. However this approach has required that values of the mesh lengths h and k be equal thus placing a severe restriction on both spline and finite difference schemes.

To illustrate the theory described in chapter 5, case study 1 was also solved by considering the knots to be unequally spaced. We initially employed the two mesh sets A and B, as previously defined,and obtained results using both a scheme incorporating splines and a finite difference approximation. For mesh set A the splines scheme always gives a smaller maximum absolute value for both the truncation errors and the difference between the numerical solutions and the analytic solution. However, as can be seen from tables 5 and 6, results obtained using mesh set B show that the errors associated with the finite differnce approximation are sometimes smaller than those associated with the splines scheme.

As a further study of the methods employing arbitrarily spaced knots we considered them to be located respectively at the zeros and at the extrema of the shifted Chebyshev polynomials. Here we again found that the splines scheme was more accurate than the well known finite difference approximation although there was no reduction in the size of the errors when they were compared with results obtained using either mesh sets A or B.

It should be noted that all the numerical solutions obtained using unequally spaced knots produce larger errors than do the corresponding solutions with the same number of equally spaced knots with a fixed knot length h. This is hardly surprising since the spline approximation is a cubic between each pair of adjacent knots and there will always be at least one knot interval of length greater than h when unequal knots are employed. This will give rise to a larger error at some points of the x range so that the ℓ_∞ norm is likely to be increased for the unequally spaced knots case, even though some of the errors may be reduced over knot interval lengths smaller than h.

In section 5.5 variable parameter extensions of both the splines and difference schemes were derived in order to remove some terms from their respective truncation errors. Whereas the truncation error for the splines scheme only contains a fourth order derivative term when solving the wave equation, the truncation error for the comparable finite difference scheme contains both a third and fourth order derivative term. We can therefore choose values of θ_i which eliminate the fourth order derivative term from the splines truncation error, but unfortunately we cannot choose values of α_i which eliminate both the third and fourth order terms from the truncation error associated with the finite difference scheme. As can be seen from tables 7 and 8 the use of the θ_i values which eliminate the fourth order term from the truncation error does in fact result in more accurate solutions, than with fixed θ values, for all the unequal knot spacings previously described. Since the parameter α_i can only be chosen to eliminate either the third or fourth order derivative term from the finite difference scheme's truncation error, the splines scheme is obviously more accurate when these optimal values of θ_i and α_i are employed. This is borne out by results shown in tables 7 and 8.

As mentioned in the appendix, for the one-dimensional wave equation, the Evans algorithm can be employed to solve the symmetric tri-diagonal systems resulting from both the spline and difference schemes. Unfortunately in this particular case no significant improvement in accuracy is gained when results are compared with those found from a more well known method of solution.

The equation considered in case study 2 is a second example of an equation with constant coefficients. Here we again found that for positive values of θ and α the truncation error for the splines scheme was always smaller in absolute magnitude than that of the finite difference approximation when a time step length of $k=0.05$ was used. On halving k we found that, with the single exception of when $\theta=\alpha=\frac{1}{4}$, the same conclusions held. For the values of $\theta=\alpha=\frac{1}{4}$ with $k=0.025$, the truncation error for the finite difference method was slightly less than that of the splines scheme although as can be seen from table 10 the difference is negligible.

In chapter 3 a method for eliminating the function f from the truncation errors of both schemes was discussed. It was indicated that this did not suggest that the truncation errors would be reduced in magnitude but only that they would be considerably simplified. From table 9 it can be seen that the truncation errors for both schemes do in fact remain of the same order of magnitude when the function f is removed.

In section 6.3 the schemes for solving equations with constant coefficients were employed on a problem of a more practical nature. The time step length k was again initially taken to be 0.05 and results obtained show that the scheme incorporating splines always has a smaller absolute value for the truncation error than does the finite difference approximation. When the time step length is reduced to 0.025, only values of θ and α equal to

1 and 17/12 give better results for the splines scheme, the remainder resulting in smaller errors for the difference approximation. As with case study 2 the function f, as given by (3.3.4) can be eliminated from the truncation errors by optimally choosing the parameters θ and α. If this is done in this case study, again no improvement in the accuracy of either scheme is achieved.

The accuracy of both the splines and difference schemes has been shown to be improved by at least an order of magnitude when choosing the parameters θ and α such that the fourth order derivative terms are eliminated from their respective truncation errors. By doing this the maximum absolute numerical value of the truncation error for the splines scheme is seen to be smaller in magnitude than that of the finite difference approximation.

The final example of obtaining solutions to equations with constant coefficients was discussed in case study 4. Here the values of a and b in equation (3.1.1) were taken to be unity and zero, respectively, whilst the constant coefficient c was allowed to take various negative values. As can be seen from tables 13, 14 and 15 the scheme incorporating splines has smaller errors than the well-known finite difference approximation when the time step length k is taken to be either 0.05, 0.025 or 0.0125. It should also be noted that the larger the constant c is made the less accurate the resulting solutions are found to be. However, since c is effectively multiplied by k^2 in both the splines and difference schemes, then as k is decreased in size the accuracy of the results from both methods is significantly increased.

All computational results shown for this case study have been found using a value for the parameters θ and α equal to 1/3. From the stability conditions (3.4.5) and (3.6.3), both schemes should be unconditionally stable since values of c have no effect on the stability. This is confirmed from tables 13,14 and 15 for the computations performed.

To illustrate the procedure for finding solutions to equations with variable coefficients a problem with engineering applications was discussed in case study 5. As is shown in tables 16 and 17 the scheme (4.1.6), which incorporates splines, almost always yields smaller maximum errors than the finite ddifference approximation (4.3.2),when values of θ and α are chosen such that both methods are unconditionally stable. There is, in fact, only one value of θ and α which gives more accurate solutions for the finite difference scheme, this being when $\theta=\alpha=\frac{1}{2}$ is used with a time step length of k=0.025. As can be seen from table 17 these values of θ, α and k give almost equal maximum absolute errors from the analytic solution and thus no significant difference in the accuracies of the methods is observed.

As was suggested in section 4.4 the stability analysis for both the scheme incorporating splines and the finite difference approximation is not very rigorous for equations having variable coefficients. We have therefore endeavoured to add numerical support to these stability analyses by using some values of θ, α and r in the computations which violate the stability conditions. These solutions were always found to be numerically unstable, all others being stable.

In all computations performed, we have found that for values of the parameters θ, $\alpha \geq 0$ the spline schemes nearly always give an improvement over the more well-known finite difference approximations. In a vast number of cases a significant improvement in accuracy has been observed when the ℓ_∞ norm values of both the truncation errors and the difference between the numerical solutions and the analytic solutions have been examined.

Conversely, for the values θ, $\alpha < 0$ we have found that the finite difference methods always yield more accurate solutions than do the spline schemes. However, negative values of θ and α are liable to violate the stability conditions as has been seen from the tables of results. They are therefore of limited applicability particularly as we have found that they give no significant improvement in accuracy over results obtained when positive values of θ and α are used. In 75% of the solution sets we can, in fact, choose a positive value of θ which gives more accurate solutions than do any negative values of the parameter α.

An additional advantage of the schemes incorporating splines is that, by using the spline function (2.2.2), it is possible to obtain solution values at any points intermediate to the mesh points on any particular time line. As was illustrated in section 6.5, these additional solutions have always been found to be in error from the analytic solutions by an amount of the same order of magnitude as were the solutions found at the mesh points.

Extensions

The spline techniques employed in this thesis may also be of use in obtaining solutions to the following more general problems

(i) Systems of first order equations

The second order one-dimensional wave equation

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

can be expressed in the form of the two simultaneous first order equations

$$\frac{\partial p}{\partial x} = \frac{\partial q}{\partial t} , \quad \frac{\partial q}{\partial x} = \frac{\partial p}{\partial t}$$

where

$$p = \frac{\partial u}{\partial x} \text{ and } q = \frac{\partial u}{\partial t} .$$

Discussions with Professor Morton of Reading University have indicated that the above system of equations may be solved using a spline technique similar to that employed in this thesis. He has suggested that even by using linear splines, instead of cubic splines, superconvergence may also be obtained.

This may also be the case if spline techniques are applied to more general systems of first order equations, for example, the equations of unsteady one-dimensional isentropic flow of compressible fluid

$$\frac{\partial u}{\partial t} + U \frac{\partial u}{\partial x} + \frac{a^2}{\rho} \frac{\partial \rho}{\partial x} = 0$$

$$\rho \frac{\partial u}{\partial x} + \frac{\partial \rho}{\partial t} + U \frac{\partial \rho}{\partial x} = 0$$

where ρ is the density of the fluid, a is the velocity of sound, x is the displacement and u is the velocity in the x direction.

(ii) Problems in two and three space dimensions

The obvious major extension of the work described in this thesis is to partial differential equations in two or three space dimensions. Here it may be advisable to initially consider parabolic equations rather than hyperbolic in order to avoid any problems caused by the characteristics in higher space dimensions. In the present one-dimensional work we have been careful to choose values of h and k so as not to violate the characteristics.

For equations in two dimensions various space regions might be considered. Many finite difference methods are available for obtaining solutions in a rectangular space region and the spline techniques should also be readily applicable to cases of this type. Problems, however, will be encountered when circular or elliptic regions are considered. For curved boundaries of this form a transformation might be available such that each point inside the region can be mapped onto a point inside a rectangular region. Some methods of transforming irregular shaped regions onto rectangular regions are discussed in Ahlberg, Nilson and Walsh (1967).

TABLE 1

Errors for case study 1 with k=0.05

$\theta=\alpha$	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
-15/12	Unstable	Unstable	Unstable	Unstable
-1/4	Unstable	Unstable	0	3.72×10^{-5}
1/4	9.98×10^{-5}	3.25×10^{-3}	2.99×10^{-4}	9.55×10^{-3}
1/3	4.99×10^{-5}	1.63×10^{-3}	3.49×10^{-4}	1.12×10^{-2}
5/12	0	1.11×10^{-5}	3.99×10^{-4}	1.28×10^{-2}
2/3	1.50×10^{-4}	4.83×10^{-3}	5.49×10^{-4}	1.74×10^{-2}
1	3.49×10^{-4}	1.12×10^{-2}	7.49×10^{-4}	2.37×10^{-2}
17/12	5.99×10^{-4}	1.90×10^{-2}	9.98×10^{-4}	3.13×10^{-2}

TABLE 2

Errors for case study 1 with k=0.025

$\theta=\alpha$	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
-15/12	Unstable	Unstable	0	2.31×10^{-4}
-1/4	6.24×10^{-5}	8.15×10^{-3}	3.74×10^{-5}	4.56×10^{-3}
1/4	4.37×10^{-5}	5.72×10^{-3}	5.62×10^{-5}	7.02×10^{-3}
1/3	4.06×10^{-5}	5.15×10^{-3}	5.93×10^{-5}	7.54×10^{-3}
5/12	3.74×10^{-5}	4.75×10^{-3}	6.24×10^{-5}	7.79×10^{-3}
2/3	2.81×10^{-5}	3.63×10^{-3}	7.18×10^{-5}	8.95×10^{-3}
1	1.56×10^{-5}	2.01×10^{-3}	8.42×10^{-5}	1.07×10^{-2}
17/12	0	2.53×10^{-5}	9.98×10^{-5}	1.25×10^{-2}

TABLE 3

Errors for case study 1 using mesh A with k=0.05

$\theta=\alpha$	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
-1/4	Unstable	Unstable	Unstable	Unstable
1/4	3.73×10^{-3}	3.67×10^{-3}	2.02×10^{-2}	2.40×10^{-2}
1/3	7.46×10^{-3}	2.02×10^{-3}	2.23×10^{-2}	2.54×10^{-2}
5/12	1.12×10^{-2}	1.23×10^{-3}	2.44×10^{-2}	2.69×10^{-2}
2/3	2.24×10^{-2}	4.47×10^{-3}	3.07×10^{-2}	3.14×10^{-2}
1	3.73×10^{-2}	1.10×10^{-2}	4.48×10^{-2}	3.73×10^{-2}
17/12	5.59×10^{-2}	1.89×10^{-2}	6.34×10^{-2}	4.45×10^{-2}

TABLE 4

Errors for case study 1 using mesh A with k=0.025

$\theta=\alpha$	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
-1/4	Unstable	Unstable	Unstable	Unstable
1/4	1.32×10^{-3}	6.25×10^{-3}	4.25×10^{-3}	2.15×10^{-2}
1/3	1.19×10^{-3}	5.84×10^{-3}	4.38×10^{-3}	2.19×10^{-2}
5/12	1.05×10^{-3}	5.43×10^{-3}	4.51×10^{-3}	2.24×10^{-2}
2/3	6.99×10^{-4}	4.20×10^{-3}	4.91×10^{-3}	2.35×10^{-2}
1	1.63×10^{-3}	1.44×10^{-3}	5.43×10^{-3}	2.49×10^{-2}
17/12	2.80×10^{-3}	1.27×10^{-3}	6.09×10^{-3}	2.67×10^{-2}

TABLE 5

Errors for case study 1 using mesh B with k=0.05

$\theta=\alpha$	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
-1/4	Unstable	Unstable	2.62×10^{-3}	7.17×10^{-3}
1/4	1.54×10^{-3}	7.60×10^{-3}	3.29×10^{-3}	4.19×10^{-3}
1/3	1.32×10^{-3}	5.99×10^{-3}	3.41×10^{-3}	5.22×10^{-3}
5/12	1.10×10^{-3}	4.38×10^{-3}	3.52×10^{-3}	6.46×10^{-3}
2/3	9.24×10^{-4}	8.99×10^{-4}	3.85×10^{-3}	1.11×10^{-2}
1	2.16×10^{-3}	6.79×10^{-3}	4.39×10^{-3}	1.75×10^{-2}
17/12	3.70×10^{-3}	1.46×10^{-2}	6.16×10^{-3}	2.55×10^{-2}

TABLE 6

Errors for case study 1 using mesh B with k=0.025

$\theta=\alpha$	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
-1/4	5.49×10^{-4}	1.27×10^{-2}	7.39×10^{-4}	2.74×10^{-3}
1/4	4.67×10^{-4}	1.01×10^{-2}	7.81×10^{-4}	2.86×10^{-3}
1/3	4.53×10^{-4}	9.74×10^{-3}	7.89×10^{-4}	2.96×10^{-3}
5/12	4.39×10^{-4}	9.31×10^{-3}	7.96×10^{-4}	3.28×10^{-3}
2/3	3.98×10^{-4}	8.11×10^{-3}	8.17×10^{-4}	3.90×10^{-3}
1	3.43×10^{-4}	6.45×10^{-3}	8.45×10^{-4}	4.87×10^{-3}
17/12	2.75×10^{-4}	4.49×10^{-3}	8.80×10^{-4}	6.39×10^{-3}

TABLE 7

Errors for case study 1 using variable parameter schemes

Mesh	θ/α	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
A	(5.5.3)	0	3.07×10^{-4}	-	-
A	(6.1.9)	-	-	1.01×10^{-3}	1.35×10^{-2}
B	(5.5.3)	0	1.49×10^{-4}	-	-
B	(6.1.9)	-	-	1.08×10^{-3}	1.15×10^{-2}

TABLE 8

Errors for case study 1 using zeros and extrema of Shifted Chebyshev Polynomial

Mesh	θ/α	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
(6.1.12)	1/3	-	8.83×10^{-3}	-	9.04×10^{-3}
(6.1.12)	5/12	-	7.21×10^{-3}	-	1.02×10^{-2}
(6.1.11)	(5.5.3)	0	2.05×10^{-4}	-	-
(6.1.11)	(6.1.9)	-	-	Unstable	Unstable
(6.1.12)	(5.5.3)	0	1.55×10^{-4}	-	-
(6.1.12)	(6.1.9)	-	-	2.97×10^{-3}	1.25×10^{-2}

TABLE 9

Errors for case study 2 with k=0.05

$\theta=\alpha$	TE_{SC}	TE_{DC}
-15/12	Unstable	Unstable
-1/4	Unstable	$6.02x10^{-5}$
1/4	$4.14x10^{-4}$	$8.78x10^{-4}$
1/3	$2.53x10^{-4}$	$1.01x10^{-3}$
5/12	$1.42x10^{-4}$	$1.05x10^{-3}$
2/3	$4.58x10^{-4}$	$1.43x10^{-3}$
1	$9.21x10^{-4}$	$2.02x10^{-3}$
17/12	$1.50x10^{-3}$	$2.67x10^{-3}$
(6.2.2)	$7.11x10^{-4}$	$5.61x10^{-4}$

TABLE 10

Errors for case study 2 with k=0.025

$\theta=\alpha$	TE_{SC}	TE_{DC}
-15/12	Unstable	$1.50x10^{-5}$
-1/4	$2.15x10^{-4}$	$1.07x10^{-4}$
1/4	$1.64x10^{-4}$	$1.62x10^{-4}$
1/3	$1.55x10^{-4}$	$1.72x10^{-4}$
5/12	$1.44x10^{-4}$	$1.82x10^{-4}$
2/3	$1.18x10^{-4}$	$2.11x10^{-4}$
1	$7.45x10^{-5}$	$2.43x10^{-4}$
17/12	$3.54x10^{-5}$	$2.62x10^{-4}$

TABLE 11

Errors for case study 3 with h=0.1 and k=0.05

$\theta=\alpha$	TE_{SC}	TE_{DC}
-15/12	Unstable	Unstable
-1/4	Unstable	5.19×10^{-5}
1/4	2.21×10^{-3}	2.68×10^{-3}
1/3	9.73×10^{-4}	3.01×10^{-3}
5/12	4.01×10^{-5}	3.55×10^{-3}
2/3	1.58×10^{-3}	4.12×10^{-3}
1	3.03×10^{-3}	4.83×10^{-3}
17/12	4.52×10^{-3}	5.95×10^{-3}
(6.3.18)	4.33×10^{-3}	1.98×10^{-3}

TABLE 12

Errors for case study 3 with h=0.1 and k=0.025

$\theta=\alpha$	TE_{SC}	TE_{DC}
-15/12	Unstable	1.27×10^{-5}
-1/4	1.28×10^{-3}	3.64×10^{-4}
1/4	1.02×10^{-3}	5.22×10^{-4}
1/3	9.35×10^{-4}	5.45×10^{-4}
5/12	8.29×10^{-4}	5.66×10^{-4}
2/3	6.31×10^{-4}	5.97×10^{-4}
1	2.98×10^{-4}	7.38×10^{-4}
17/12	1.02×10^{-5}	8.60×10^{-4}

TABLE 13

Errors for case study 4 using k=0.05

c	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
-10	4.01×10^{-4}	7.53×10^{-3}	8.10×10^{-4}	1.49×10^{-2}
-20	1.15×10^{-3}	1.42×10^{-2}	1.58×10^{-3}	1.90×10^{-2}
-30	2.20×10^{-3}	2.02×10^{-2}	2.65×10^{-3}	2.38×10^{-2}
-40	3.55×10^{-3}	2.63×10^{-2}	4.03×10^{-3}	2.92×10^{-2}
-50	5.19×10^{-3}	4.33×10^{-2}	5.71×10^{-3}	4.63×10^{-2}
-100	1.77×10^{-2}	1.52×10^{-1}	1.85×10^{-2}	1.55×10^{-1}

TABLE 14

Errors for case study 4 using k=0.025

c	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
-10	1.29×10^{-5}	9.24×10^{-4}	8.81×10^{-5}	6.37×10^{-3}
-20	3.37×10^{-5}	1.77×10^{-3}	1.37×10^{-4}	6.65×10^{-3}
-30	9.94×10^{-5}	3.79×10^{-3}	2.04×10^{-4}	7.40×10^{-3}
-40	1.84×10^{-4}	5.56×10^{-3}	2.91×10^{-4}	8.46×10^{-3}
-50	2.88×10^{-4}	1.01×10^{-2}	3.98×10^{-4}	1.34×10^{-2}
-100	1.09×10^{-3}	3.91×10^{-2}	1.22×10^{-3}	4.26×10^{-2}

TABLE 15

Errors for case study 4 using k=0.0125

c	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
-10	1.03×10^{-5}	3.37×10^{-3}	1.49×10^{-5}	4.34×10^{-3}
-20	7.60×10^{-6}	1.60×10^{-3}	1.79×10^{-5}	3.59×10^{-3}
-30	3.60×10^{-6}	5.26×10^{-4}	2.22×10^{-5}	2.89×10^{-3}
-40	1.50×10^{-6}	2.49×10^{-4}	2.78×10^{-5}	2.97×10^{-3}
-50	7.90×10^{-6}	1.24×10^{-3}	3.43×10^{-5}	4.47×10^{-3}
-100	5.74×10^{-5}	8.40×10^{-3}	8.60×10^{-5}	1.20×10^{-2}

TABLE 16

Errors for case study 5 using k=0.05

$\theta=\alpha$	TE_{SV}	E_{SV}	TE_{DV}	E_{DV}
1	2.63×10^{-3}	2.56×10^{-2}	3.79×10^{-3}	3.14×10^{-2}
2/3	1.81×10^{-3}	1.85×10^{-2}	3.39×10^{-3}	3.53×10^{-2}
5/12	7.32×10^{-4}	6.55×10^{-3}	2.88×10^{-3}	2.90×10^{-2}
1/3	5.72×10^{-4}	2.94×10^{-3}	2.55×10^{-3}	2.72×10^{-2}
1/4	4.54×10^{-4}	6.97×10^{-3}	2.42×10^{-3}	2.60×10^{-2}
0	1.97×10^{-3}	1.73×10^{-2}	1.73×10^{-3}	2.01×10^{-2}
-1/12	3.33×10^{-3}	2.13×10^{-2}	1.38×10^{-3}	1.69×10^{-2}
-3/4	Unstable	Unstable	5.07×10^{-3}	2.00×10^{-2}
-1	Unstable	Unstable	Unstable	Unstable

TABLE 17

Errors for case study 5 using k=0.025

$\theta=\alpha$	TE_{SV}	E_{SV}	TE_{DV}	E_{DV}
1	2.18×10^{-4}	1.31×10^{-2}	6.61×10^{-4}	2.54×10^{-2}
2/3	1.81×10^{-4}	1.62×10^{-2}	6.05×10^{-4}	2.36×10^{-2}
5/12	1.49×10^{-4}	1.94×10^{-2}	5.68×10^{-4}	2.25×10^{-2}
1/3	1.37×10^{-4}	2.06×10^{-2}	5.57×10^{-4}	2.20×10^{-2}
1/4	1.25×10^{-4}	2.17×10^{-2}	5.45×10^{-4}	2.16×10^{-2}
0	2.12×10^{-4}	2.48×10^{-2}	5.03×10^{-4}	2.00×10^{-2}
-1/12	2.30×10^{-4}	2.56×10^{-2}	4.88×10^{-4}	1.94×10^{-2}
-3/4	4.97×10^{-4}	2.98×10^{-2}	3.56×10^{-4}	1.32×10^{-2}
-1	6.74×10^{-4}	2.95×10^{-2}	3.01×10^{-4}	1.00×10^{-2}

TABLE 18

Comparison between intermediate solutions found using (3.2.2) and mesh point solution found using (4.1.6), for case study 5.

x values	Intermediate Solutions (3.2.2)	Mesh point Solutions (4.1.6)
0.05	1.31×10^{-3}	
0.10		6.46×10^{-4}
0.15	3.72×10^{-4}	
0.20		1.73×10^{-3}
0.25	1.48×10^{-3}	
0.30		2.94×10^{-3}
0.35	1.30×10^{-3}	
0.40		1.02×10^{-3}
0.45	3.78×10^{-3}	

TABLE 19

Comparison between the Evans algorithm and the more well known method. Errors for case study 1

	Evans algorithm.		Well known method, (AP2) to (AP7)	
$\theta=\alpha$	TE_{SC}	E_{SC}	TE_{DC}	E_{DC}
1/4	9.98×10^{-5}	3.21×10^{-3}	9.98×10^{-5}	3.25×10^{-3}
1/3	4.99×10^{-5}	1.61×10^{-3}	4.99×10^{-5}	1.63×10^{-3}
5/12	0	2.41×10^{-5}	0	1.11×10^{-5}

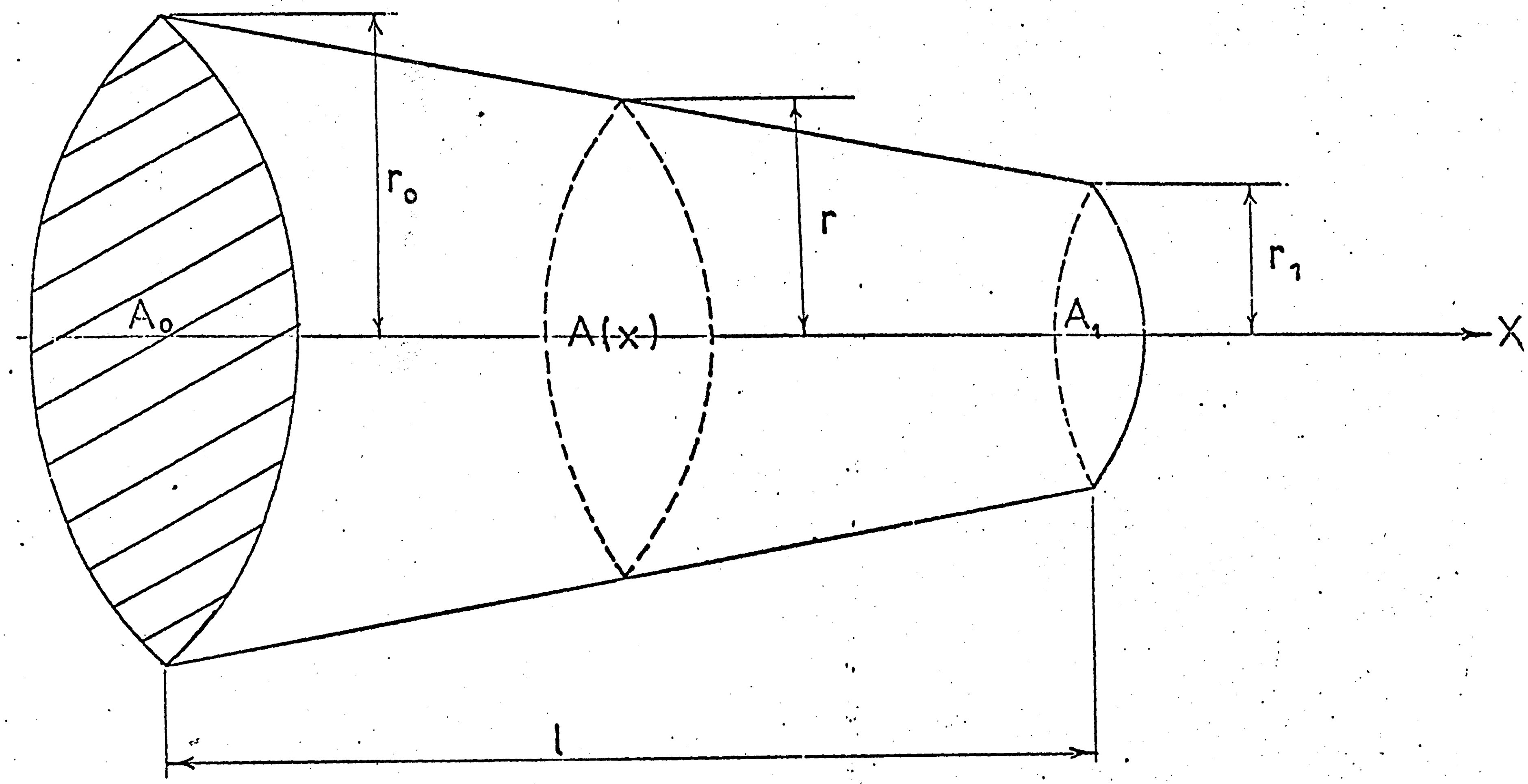


Fig. 3

Truncated cone cutting tool

APPENDIX

The Solution of Tri-diagonal Systems

The difference schemes derived in this thesis are all implicit in nature and have three unknown solution values on the advanced time line $t = (j+1)k$. They can therefore be expressed in the tri-diagonal form

$$\begin{bmatrix} b_1 & c_1 & & & & \bigcirc \\ a_2 & b_2 & c_2 & & & \\ & a_3 & b_3 & c_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & a_{n-1} & b_{n-1} & c_{n-1} & \\ \bigcirc & & & a_n & b_n & \end{bmatrix} \begin{bmatrix} U_{1,j+1} \\ U_{2,j+1} \\ U_{3,j+1} \\ \vdots \\ U_{n-1,j+1} \\ U_{n,j+1} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \\ d_{n-1} \\ d_n \end{bmatrix} \quad \text{(AP1)}$$

An efficient method for solving systems such as (AP1) has been described by numerous authors, see for example, Todd (1962), Richtmyer and Morton (1967) and Mitchell (1969). The method which is based on the Gaussian Elimination process takes advantage of the sparsity of the matrix in (AP1). A simple derivation of the method is as follows.

Letting

$$U_{i,j+1} = \alpha_i + \beta_i U_{i+1,j+1} \qquad (i = 1,2,\ldots,(n-1)) \qquad \text{(AP2)}$$

then by rearranging the first line of (AP1) to the form

$$U_{1,j+1} = (d_1 - c_1 U_{2,j+1})/b_1 \qquad \text{(AP3)}$$

we can easily see that

$$\alpha_1 = \frac{d_1}{b_1} \text{ and } \beta_1 = \frac{-c_1}{b_1} \quad . \tag{AP4}$$

Similarly for the i^{th} row of (AP1), which becomes on substitution from (AP2)

$$U_{i,j+1} = \frac{d_i - a_i\alpha_{i-1} - c_i U_{i+1,j+1}}{a_i \beta_{i-1} + b_i} \tag{AP5}$$

we can show that

$$\alpha_i = \frac{d_i - a_i\alpha_{i-1}}{a_i\beta_{i-1} + b_i} , \quad \beta_i = \frac{-c_i}{a_i\beta_{i-1} + b_i} \quad . \tag{AP6}$$

$$(i = 2,3, \ldots, (n-1))$$

By similar reasoning it can be shown that from the n^{th} line of (AP1) we have

$$U_{n,j+1} = \frac{d_n - a_n\alpha_{n-1}}{a_n\beta_{n-1} + b_n} \tag{AP7}$$

Using the above expressions the values $U_{i,j+1}$ $(i = 1,2,\ldots,n)$ are found by firstly evaluating the α_i and β_i $(i = 1,2,\ldots,(n-1))$ from (AP4) and (AP6). We then determine the value $U_{n,j+1}$ from (AP7) and subsequently the remaining solutions $U_{i,j+1}$ $(i = (n-1),(n-2),\ldots,1)$ are easily evaluated from the expression (AP2).

Provided the matrix in the system (AP1) is diagonally dominant it has been shown by Douglas (1959) that the above method does not introduce large rounding errors into the solutions

$U_{i,j+1}$. However, when solving partial differential equations using either splines or differences, care must be taken in choosing values of the mesh lengths h and k if the round-off errors are to be controlled. This is because whilst the truncation errors will diminish as h and k are made smaller the rounding errors will become larger unless more significant digits are considered. This fact is borne out by some of the results shown in the earlier case studies.

It has recently been shown by Evans (1972) that an improvement on the above method is available for certain symmetric tri-diagonal systems. When solving equations of the form

$$cx_{i-1} + bx_i + cx_{i+1} = d_i \qquad (i = 1,2,\ldots,n) \qquad \text{(AP8)}$$

the Evans algorithm required 5n multiplications and 4n additions as opposed to the 6n multiplications and 4n additions required by the method described earlier.

In the numerical case studies discussed in chapter 6 we have solved all the tri-diagonal systems using the method given by equations (AP2) to (AP7) above. For case study 1 the wave equation does in fact result in a system of the form given by (AP8) when either the splines scheme (3.2.14) or the difference approximation (3.5.2) are used. We have therefore applied the Evans algorithm to case study 1 and obtained, along with the solutions, numerical values of both the truncation errors and the difference between the analytic solution and the computed solutions. Some results using the spline technique (3.2.14) are given in table 19 where the notation used is that of the previous case studies and where the step lengths h and k have been chosen as 0.1 and 0.05 respectively. The comparative results from table 1 of case study 1 are also shown.

REFERENCES

E L ALBASINY and W D HOSKINS, Cubic Spline Solutions to Two-Point Boundary Value Problems, Comput. J. 12 (1969) 151-153

J H AHLBERG, E N NILSON and J L WALSH, The Theory of Splines and their Application, Academic Press, London (1967)

W F AMES, Numerical Methods for Partial Differential Equations, Nelson and Sons (1969)

W G BICKLEY, Piecewise Cubic Interpolation and Two-Point Boundary Value Problems, Comput. J. 11 (1968) 206-208

G BIRKHOFF and C R DE BOOR, Piecewise Polynomial Interpolation and Approximation, In Approximation of Functions edited by H L Garabedian, Elsevier, Amsterdam (1965) 164-190

M G BRISCOE and A A KOVITZ, Shock Reflection from a Wavy Wall, J. Fluid Mech. 31 (1968) 529-546.

L COLLATZ, The Numerical Treatment of Differential Equations, 3rd ed., Springer-Verlag, Berlin (1960)

R COURANT, K FRIEDRICHS and H LEWY, Uber die partiellen Differenzengleichungen der mathematischen Physik, Math. Ann. 100 (1928) 32-74, English translation by Phyllis Fox, On the partial difference equations of mathematical physics, IBM Journal 11 (1967) 215-234

M G COX, An Algorithm for Spline Interpolation, J. Inst. Math. Appl. 15 (1975) 95-108

S H CRANDALL, An Optimum Implicit Recurrence Formula for the Heat Conduction Equation, Qtly. Apl. Maths. 13 (1955) 318-320

J CRANK and R S GUPTA, A Method for Solving Moving Boundary Problems in Heat Flow Using Cubic Splines or Polynomials, J. Inst. Math. Appl. 10 (1972) 296-304

J CRANK and P NICOLSON, A Practical Method for the Numerical Evaluation of Solutions of Partial Differential Equations of the Heat Conduction Type, Proc. Camb. Phil. Soc. 43 (1947) 50-67

A R CURTIS, The Approximation of a Function of One Variable by Cubic Splines, In Numerical Approximations to Functions and Data edited by J G Hayes, Athlone Press, London (1970) 28-42

A R CURTIS and M J D POWELL, Using Cubic Splines to Approximate Functions of One Variable to Prescribed Accuracy, AERE Harwell report no. AERE-R5602, HMSO (1967)

J DOUGLAS Jr, Round-off Error in the Numerical Solution of the Heat Equation, J. Assoc. Comp. Machinery 6 (1959) 48-58

J DOUGLAS Jr and T M GALLIE Jr, Variable Time Steps in the Solution of the Heat Flow Equation by a Difference Equation, Proc. Amer. Soc. 6 (1955) 787-793

S E EL-GENDI, On Chebyshev Solution of Parabolic Partial Differential Equations, J. Inst. Math. Appl. 16 (1975) 283-289

M E A EL-TOM, Spline Function Approximations to the Solution of Singular Volterra Integral Equations of the Second Kind, J. Inst. Math. Appl. 14 (1974) 303-309

M E A EL-TOM, Application of Spline Functions to Systems of Volterra Integral Equations of the First and Second Kinds, J. Inst, Math. Appl. 17 (1976) 295-310

D J EVANS, An Algorithm for the Solution of Certain Tri-Diagonal Systems of Linear Equations, Comput, J. 15 (1972) 356-359

L FOX,Numerical Solution of Ordinary and Partial Differential Equations, Pergamon Press, London (1962).

J FRIBERG, Conditionally Stable Difference Approximations for the Wave-Operator, Nordisk Tidj Inform Behand. 1 (1961) 69-86.

D J FYFE,The Use of Cubic Splines in the Solution of Two-Point Boundary Value Problems, Comput. J. 12 (1969) 188-192

T N E GREVILLE, Introduction to Spline Functions, In Theory and Applications of Spline Functions edited by T N E Greville, Academic Press, New York (1969) 1-35

J G HAYES and J HALLIDAY, The Lease-Squares Fitting of Cubic Spline Surfaces to General Data Sets, J. Inst. Math. Appl. 14 (1974) 89-103

W KIZNER,Error Curves for Lanczos' "Selected Points" Method, Comput. J.8 (1966) 372-382

M LEES,Von Neumann Difference Approximation to Hyperbolic Equations, Pacific J. Math. 10 (1960) 213-222

A R MITCHELL, Computational Methods in Partial Differential Equations, Wiley, Aberdeen (1969).

W D MURRAY and F LANDIS, Numerical and Machine Solutions of Transient Heat-Conduction Problems Involving Melting or Freezing, J Heat Transfer 81 (1959) 106-112

G G OBRIEN,M A HYMAN and S KAPLAN, A Study of the Numerical Solution of Partial Differential Equations, J. Math. Phys. 29 (1951) 223-251

N PAPAMICHAEL and J R WHITEMAN, A Cubic Spline Technique for the One Dimensional Heat Conduction Equation, J. Inst. Math. Appl.11 (1973) 111-113

S M PICKEN, Algorithms for the Solution of Differential Equations in Chebyshev Series by the Selected Points Method, Techn. Rep. Maths 94, National Physical Laboratory (1970)

G F RAGGETT, On the Solution of Hyperbolic Partial Differential Equations Using Cubic Splines, Paper presented at the International Congress of Mathematicians, Vancouver, Canada (1974)

G F RAGGETT, J A R STONE and S J WISHER, The Cubic Spline Solution of Practical Problems Modelled by Hyperbolic Partial Differential Equations, Computer Methods in Applied Mechanics and Engineering 8 (1976) 139-151

G F RAGGETT and P D WILSON, A Fully Implicit Finite Difference Approximation to the One-Dimensional Wave Equation Using a Cubic Spline Technique, J. Inst. Math. Appl. 14 (1974) 75-77

R D RICHTMYER, Taylor Stability in Compressible Flow Comm. Pure Appl. Math. 13 (1960) 297-319

R D RICHTMYER and K W MORTON, Difference Methods for Initial-Value Problems, Interscience publishers, New York, (1967)

V K SAUL'YEV, Integration of Equations of Parabolic Type by the Method of Nets, Pergamon press, London (1964)

I J SCHOENBERG, Contributions to the Problem of Approximation of Equidistant Data by Analysic Functions, Quart. Appl. Math. 4 (1946), Part A 45-99, Part B 112-141

L L SCHUMAKER, Some Algorithms for the Computation of Interpolating and Approximating Spline Functions, In Theory and Applications of Spline Functions edited by T N E Greville, Academic Press, New York (1969) 87-102

J TODD, Survey of Numerical Analysis, McGraw-Hill (1962)