NUMERICAL METHODS FOR ORDINARY DIFFERENTIAL EQUATIONS

WITH APPLICATIONS TO PARTIAL DIFFERENTIAL EQUATIONS

A thesis submitted for the degree of

Doctor of Philosophy

by

Abdul Qayyum Masud Khaliq

Department of Mathematics and Statistics, Brunel University

Uxbridge, Middlesex, England. UB8 3PH

February 1983

ABSTRACT

The thesis develops a number of algorithms for the numerical solution of ordinary differential equations with applications to partial differential equations. A general introduction is given; the existence of a unique solution for first order initial value problems and well known methods for analysing stability are described.

A family of one-step methods is developed for first order ordinary differential equations. The methods are extrapolated and analysed for use in PECE mode and their theoretical properties, computer implementation and numerical behaviour, are discussed.

L_0-stable methods are developed for second order parabolic partial differential equations in one space dimension; second and third order accuracy is achieved by a splitting technique in two space dimensions. A number of two-time level difference schemes are developed for first order hyperbolic partial differential equations and the schemes are analysed for A_0-stability and L_0-stability. The schemes are seen to have the advantage that the oscillations which are present with Crank-Nicolson type schemes, do not arise.

A family of two-step methods is developed for second order periodic initial value problems. The methods are analysed, their error constants and periodicity intervals are calculated. A family of numerical methods is developed for the solution of fourth order parabolic partial differential equations with constant coefficients and variable coefficients and their stability analyses are discussed.

The algorithms developed are tested on a variety of problems from the literature.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor Dr. E.H. Twizell for his invaluable suggestions, continuous encouragement and constructive criticism during both the period of research and the writing of the thesis. He has always given patiently of his time and his endeavours have extended well beyond the bounds of mere supervision. I have learned a great deal from him and, for all that he has done, I am deeply grateful.

I am also indebted to Professor J.Ll. Morris, Mr. G.D. Smith and Professor J.R. Whiteman for many helpful discussions.

Finally, I wish to thank the British Government for its partial support in tuition fees under the Overseas Research Students Fees Support Scheme during the period 1981 - 1983.

"Occupying a unique place along the border between applied-mathematics and the concrete world of industry, the numerical solution of differential equations, probably more than any other branch of numerical analysis, is in a constant state of unrest and evolution. Being so widely and variously applied in the real world, its techniques are relentlessly put to the ruthless test of practical success and usefulness. Nor does it evolve solely through the cross influences of the practical necessities of engineering; unusual impetus is also given to this field by the outstanding advances in computer technology, which is gathering now to miniaturize hardware to lower the cost of the equipment, the arithmetic, the logic, the storage, and the output that is made more comprehensively grasped by directly presenting it to that most remarkable of the human senses-vision, through computer graphics, shifting thereby the engineer's or programmer's priorities in selecting the most appropriate solution algorithm".

Isaac Fried, 1979.

CONTENTS

CHAPTER 1

INTRODUCTION

Consider the first-order initial value problem

$$y' = f(x,y), \quad y(a) = \eta \ . \tag{1.1}$$

The following theorem outlined in Lambert (1973), with proof contained in Henrici (1962), states conditions on $f(x,y)$ which guarantee the existence of a unique solution of the initial value problem (1.1).

Theorem 1.1

Let $f(x,y)$ be defined and continuous for all points (x,y) in the region D defined by $a \le x \le b$, $-\infty < y < \infty$, a and b finite, and let there exist a constant L such that, for every x,y,y^* such that (x,y) and (x,y^*) are both in D,

$$|f(x,y) - f(x,y^*)| \le L \, | \, y-y^* \, | \ . \tag{1.2}$$

Then, if η is any given number, there exists a unique solution $y(x)$ of the initial value problem (1.1), where $y(x)$ is continuous and differentiable for all (x,y) in D.

The requirement (1.2) is known as a Lipschitz condition, and the constant L as a Lipschitz constant. This condition may be thought of as being intermediate between differentiability and continuity, in the sense that

$f(x,y)$ continuously differentiable with respect to y for all (x,y) in D

$\Rightarrow$ $f(x,y)$ satisfies a Lipschitz condition with respect to y for all (x,y) in D

$\Rightarrow$ $f(x,y)$ continuous with respect to y for all (x,y) in D.

In particular, if $f(x,y)$ possesses a continuous derivative with respect to y for all (x,y) in D, then, by the mean value theorem,

$$f(x,y) - f(x,y^*) = \frac{\partial f(x,\bar{y})}{\partial y} \, (y-y^*) \ ,$$

where $\bar{y}$ is a point in the interior of the interval whose end-points are y and y^*, and (x,y) and (x,y^*) are both in D. Clearly, (1.2) is then satisfied if L is chosen to be

$$L = \sup_{(x,y)\in D} \left| \frac{\partial f(x,y)}{\partial y} \right| . \tag{1.3}$$

In many areas such as control theory, chemical kinetics and biology, the dynamic behaviour is modelled, not with a single differential equation, but with a system of m simultaneous first-order equations in m dependant variables $y_1, y_2, \ldots y_m$. If each of these variables satisfies a given condition at the same value a of x then the initial value problem for a first-order system may be written as

$$\begin{aligned} y_1' &= f_1(x,y_1,y_2,\ldots,y_m) , \quad y_1(a) = \eta_1 , \\ y_2' &= f_2(x,y_1,y_2,\ldots,y_m) , \quad y_2(a) = \eta_2 , \\ &\vdots \qquad\qquad\qquad\qquad\qquad\quad \vdots \\ y_m' &= f_m(x,y_1,y_2,\ldots,y_m) , \quad y_m(a) = \eta_m . \end{aligned} \tag{1.4}$$

Introducing the vector notation

$$\underline{y} = (y_1,y_2,\ldots,y_m)^T , \quad \underline{f} = (f_1,f_2,\ldots,f_m)^T = \underline{f}(x,\underline{y}) , \quad \eta = (\eta_1,\eta_2,\ldots,\eta_m)^T,$$

T denoting transpose, the initial-value problem (1.4) may be written as

$$\underline{y}' = \underline{f}(x,\underline{y}) , \quad \underline{y}(a) = \underline{\eta} . \tag{1.5}$$

Theorem 1.1 readily generalises to give necessary conditions for the existence of a unique solution to (1.5); all that is required is that the region D now be defined by $a \le x \le b$, $-\infty < y_i < \infty$, $i = 1,2,\ldots,m$, and (1.2) be replaced by the condition

$$||\underline{f}(x,\underline{y}) - \underline{f}(x,y^*)|| \le L ||\underline{y} - \underline{y}^*|| , \tag{1.6}$$

where $(x,\underline{y})$ and $(x,\underline{y}^*)$ are in D, and $||.||$ denotes a vector norm. For the properties of vector and matrix norms see, for example, Mitchell and Griffiths (1980). In the case when each of the $f_i(x,y_1,y_2,\ldots,y_m)$, $i = 1,2,\ldots,m$, possesses a continuous derivative with respect to each of

the y_j, $j = 1,2,\ldots,m$, then

$$L = \sup_{(x,y)\in D} ||\partial \underline{f}/\partial \underline{y}|| \tag{1.7}$$

may be chosen analogously to (1.3), where $\partial \underline{f}/\partial \underline{y}$ is the Jacobian of $\underline{f}$ with respect to $\underline{y}$ - that is, the $m \times m$ matrix whose i,jth element is $\partial f_i (x,y_1,y_2,\ldots,y_m/\partial y_j$, and $||.||$ denotes a matrix norm subordinate to the vector norm employed in (1.6).

The first order system (1.5), namely $\underline{y}' = \underline{f}(x,\underline{y})$, where $\underline{y}$ and $\underline{f}$ are m-dimensional vectors, is said to be linear if

$$\underline{f}(x,y) = A(x)\underline{y} + \underline{\phi}\ (x)\ ,$$

where $A(x)$ is an $m \times m$ matrix and $\underline{\phi}(x)$ an m-dimensional vector; if, in addition, $A(x) = A$, a constant matrix, the system is said to be linear with constant coefficients. To find the general solution of the system

$$\underline{y}' = A\underline{y} + \underline{\phi}(x)\ , \tag{1.8}$$

let $\hat{\underline{y}}(x)$ be the general solution of the corresponding homogeneous system

$$\underline{y}' = A\underline{y}\ . \tag{1.9}$$

If $\underline{\Psi}(x)$ is any particular solution of (1.8), then

$$\underline{y}(x) = \hat{\underline{y}}(x) + \underline{\Psi}(x)$$

is the general solution of (1.8). A set of solutions $\underline{y}_k(x)$, $k = 1,2,\ldots,m$, of (1.9) is said to be linearly independent if

$$\sum_{k=1}^{m} a_k \underline{y}_k(x) \equiv \underline{0}\ ,$$

implies $a_k = 0$, $k = 1,2,\ldots,m$. The general solution of (1.9) may be written as a linear combination of the members of a set of m linearly independent solutions $\underline{y}_k(x)$, $k = 1,2,\ldots,m$. It can easily be seen that

$$\underline{y}(x) = \exp(\lambda_k x)\ \underline{c}_k\ , \tag{1.10}$$

where $\underline{c}_k$ is an m-dimensional vector, is a solution of (1.9) if

$$\lambda_k \underline{c}_k = A\underline{c}_k \quad ,$$

that is if λ_k is an eigenvalue of A and $\underline{c}_k$ is the corresponding eigenvector. Considering only the case where A possesses m distinct complex eigenvalues λ_k, $k = 1,2,\ldots,m$, the corresponding eigenvectors $\underline{c}_k$, $k = 1,2,\ldots,m$, are then linearly independent (Mitchell and Griffiths (1980), Chapter 1), and it follows that (1.10) forms a set of linearly independent solutions of (1.9), whose general solution is of the form

$$\sum_{k=1}^{m} N_k \exp(\lambda_k x)\underline{c}_k \quad ,$$

where the N_k, $k = 1,2,\ldots,m$ are arbitrary constants. The general solution of (1.8) is then

$$\underline{y}(x) = \sum_{k=1}^{m} N_k \exp(\lambda_k x)\underline{c}_k + \underline{\Psi}(x) \quad . \tag{1.11}$$

The solution of the initial value problem

$$\underline{y}' = A\underline{y} + \underline{\phi}(x) \quad , \quad \underline{y}(a) = \underline{\eta} \tag{1.12}$$

may now be found under the assumption that A has m distincit eigenvalues, and that the particular solution $\underline{\Psi}(x)$ of (1.8) is known. By (1.11), the general solution of (1.8) satisfies the initial conditions given in (1.12) if

$$\underline{\eta} - \underline{\Psi}(a) = \sum_{k=1}^{m} N_k \exp(\lambda_k a)\underline{c}_k \quad . \tag{1.13}$$

Since the vectors $\underline{c}_k$, $k = 1,2,\ldots,m$, form a basis of the m-dimensional vector space (Mitchell and Griffiths (1980), Chapter 1), $\underline{\eta} - \underline{\Psi}(a)$ may be expressed uniquely in the form

$$\underline{\eta} - \underline{\Psi}(a) = \sum_{k=1}^{m} n_k \underline{c}_k \quad . \tag{1.14}$$

On comparing (1.13) with (1.14), it is seen that (1.11) becomes a solution of (1.12) by choosing $n_k = N_k \exp(-\lambda_k a)$. The solution of (1.12) is thus

$$\underline{y}(x) = \sum_{k=1}^{m} N_k \exp\{(x-a)\lambda_k\}\underline{c}_k + \underline{\Psi}(x) \quad .$$

In Chapter 2 a family of one-step multiderivative methods based on

Padé approximants to the matrix exponential function is developed. The methods are extrapolated and analysed for use in PECE mode. Error constants, stability intervals and stability regions are calculated and the combinations compared with well known linear-multistep combinations and combinations using high accuracy Newton-Cotes quadrature formulas as correctors. A practical problem in applied chemistry is modelled mathematically and one of the fourth order methods developed is used to find the numerical solution. For the stability analyses of the methods, the definition of A-stability due to Dahlquist (1963) is used. Dahlquist associated a stability region with a multistep formula and introduced the concept of A-stability. These definitions are now quoted for completeness.

Definition 1.1

The stability region R associated with a multistep formula is defined as the set

$R = \{h\lambda$: the formula applied to $y' = \lambda y$, $y(x_0) = y_0$, with constant step size $h > 0$, produces a sequence $\{y_n\}$ satisfying $y_n \to 0$ as $n \to \infty\}$.

Definition 1.2

A formula is A-stable if the stability region associated with that formula contains the open left half-plane.

Dahlquist proved that an A-stable linear multistep formula must be implicit, that its maximum order is two, and , of those of second order, the one with the smallest truncation error coefficients is the trapezoidal rule.

Padé approximants to the exponential function (Padé (1892)), which are used extensively in the thesis are now defined.

Let $f(Z)$ be analytic in a region of the complex plane containing the origin $Z = 0$. A Padé approximation (Graves-Morris (1973)) $R_{m,k}(Z)$ to the function $f(Z)$ is defined by

(1.15) $$R_{m,k}(z) = \frac{P_k(z)}{Q_m(z)} ,$$

where $P_k(z)$ and $Q_m(z)$ are polynomials in z of degrees k and m respectively with leading coefficients unity. For each pair of non-negative integers m and k, $P_k(z)$ and $Q_m(z)$ are those polynomials for which the Taylor series expansion of $R_{m,k}(z)$ about the origin agrees with the Taylor series expansion of $f(z)$ for as many terms as possible. Since the ratio (1.15) contains essentially $m+k+1$ unknown coefficients, the requirement that

(1.16) $$Q_m(z)\, f(z) - P_k(z) = 0(|z|^{m+k+1}) , \quad |z| \to 0$$

gives rise to $m+k+1$ linear equations for these coefficients. The Padé Table is an infinite two-dimensional array of Padé approximations to the given function $f(z)$, where $R_{m,k}(z)$ occupies the intersection of the mth row and kth column.

For the function $f(z) = \exp(z)$, Varga (1962), the entries in the Padé Table are given explicitly by

$$P_k(z) = \sum_{j=0}^{m} \frac{(m+k-j)!\,m!}{(m+k)!\,j!\,(m-j)!}\,(z)^j$$

and

$$Q_m(z) = \sum_{j=0}^{k} \frac{(m+k-j)!\,k!}{(m+k)!\,j!\,(k-j)!}\,(-z)^j$$

and if

$$\exp(z) = \frac{P_k(z)}{Q_m(z)} + R^*_{m,k}(z) ,$$

then the remainder $R^*_{m,k}(z)$ is given by

$$R^*_{m,k}(z) = \frac{(-1)^{k+1} z^{(m+k+1)}}{(m+k)!\; Q_m(z)} \int_0^1 \exp(z(1-u))u^k(1-u)^m du .$$

The first twenty four entries of the Padé Table for $f(z) = \exp(z)$ are given in Appendix I.

Some properties of Padé approximants are given by Lambert (1973) as follows:

"Let $R_{m,k}(z)$ be the (m.k) Padé approximant to $\exp(-z)$, then $P_{m,k}(z)$ is

(i) A-acceptable if $m = k$

(ii) A(0)-acceptable if $m \geq k$

(iii) L-acceptable if $m = k+1$ or $m = k+2$"

The region of acceptability of $R_{m,k}(z)$ is that area of the complex plane within which the approximation $R_{m,k}(z)$ satisfies $|R_{m,k}(z)| < 1$.

In Chapters 3 and 4 several time discretizations are considered for the linear time-dependent partial differential equation

$$\frac{\partial u}{\partial t} = Du + f \quad , \tag{1.17}$$

where D is a differential operator involving only space-derivations, both D and f are independent of time t, and initial and boundary conditions are specified. A space-discretization and a finite-difference approximation may be used to reduce the problem (1.17) to the solution of a system of ordinary differential equations,

$$\frac{d\underline{U}}{dt} = A\underline{U} + \underline{s} \quad , \quad t > 0 \tag{1.18}$$

$$\underline{U}(0) = \underline{g} \tag{1.19}$$

where A is a square matrix, the vector $\underline{s}$ is the vector of frozen boundary values and the vector $\underline{U}$ is the computed solution of (1.17) for $t > 0$. The solution of the system of differential equations (1.18) subject to the specified initial conditions (1.19) is given by

$$\underline{U}(t) = -A^{-1}\underline{s} + \exp(tA)\ (\underline{g} + A^{-1}\underline{s}) \tag{1.20}$$

which may be written in step-wise fashion as

$$\underline{U}(t+\ell) = -A^{-1}\underline{s} + \exp(\ell A)\ (\underline{U}(t) + A^{-1}\underline{s}) \quad , \tag{1.21}$$

where ℓ is the time step.

The relationship between $\exp(z)$ and the matrix exponential function $\exp(\ell A)$ now follows in an obvious way. Formally the variable z is replaced by the matrix A in (1.15), such that

$$\exp(\ell A) \doteq \{Q_m(\ell A)\}^{-1} \cdot \{P_k(\ell A)\} \equiv R_{m,k}(\ell A)$$

is the (m,k) Padé approximation of $\exp(\ell A)$. The relationship between certain well-known numerical methods and the matrix Padé approximations may be shown, for example, by approximating the matrix exponential $\exp(\ell A)$ of equation (1.21) by the entry $R_{1,1}(\ell A)$ of the Padé Table to give

$$\underline{U}(t+\ell) = (I - \tfrac{1}{2}\ell A)^{-1} (I+\tfrac{1}{2}\ell A)(\underline{U}(t)-A^{-1}\underline{s})+A^{-1}\underline{s} \quad , \tag{1.22}$$

which, in implicit form, is

$$(I - \tfrac{\ell}{2}A)\underline{U}(t+\ell) = (I+\tfrac{\ell}{2}A)\underline{U}(t) + \ell\underline{s} \quad . \tag{1.23}$$

Equation (1.23) defines the Crank-Nicolson method applied to equation (1.18) if A is a tridiagonal matrix with the entry -2 on the diagonal and 1 on the super-and sub-diagonals. In a similar manner it can be shown that $R_{0,1}(\ell A)$ and $R_{1,0}(\ell A)$ approximations generate respectively the well known explicit and fully implicit methods for second order parabolic partial differential equations, see for example, Lawson and Morris (1978) and Smith and Twizell (1982). However, it is shown in Lawson and Morris (1978), that the (1,1) Padé approximant, (the Crank-Nicolson method) is an A-stable method and is less than satisfactory when a time discretization is used with time step which is too large relative to the spatial discretization.

In Chapter 3 a family of methods is developed for second order parabolic partial differential equations, which do not suffer from this feature. Second and third order accuracy is achieved in two space dimensions by a splitting technique. The methods are tested on two problems from the literature. The behaviours of the methods are also shown graphically. Stability of the methods is analysed by two well known methods; the von Neumann method and the Matrix method, which are now mentioned briefly. For full details, see for example, Smith (1978) and Mitchell and Griffiths (1980).

The von Neumann Method, developed by J. von Neumann and first discussed

in detail by O'Brien et al (1951), provides a simple necessary condition for numerical stability, and essentially depends on the uniform boundedness of the Fourier coefficients of the solution of the difference equation. It is assumed that there exist harmonic decompositions of the grid functions U_k at the initial time level and writes

$$U_k = \sum_j A_j \exp(i\beta_j x_k) \quad ,$$

where $i = \sqrt{-1}$, the frequencies β_j are, in general use, arbitrary, and a uniform grid is used. It is only necessary to consider the single term $\exp(i\beta x)$ where β is any real number and to use the superposition principle for linear problems. To investigate the growth of the grid functions as t increases for any value of β, it is necessary to find a solution of the difference equation which reduce to $\exp(i\beta x)$ when $t = 0$. Such a solution is

$$\exp(\alpha t)\ \exp(i\beta x)$$

where $\alpha = \alpha(\beta)$ is, in general, complex. The original grid function $\exp(i\beta x)$ will not grow with time if

$$|\exp(\alpha \ell)| \leq 1 \qquad (1.23)$$

where ℓ is the increment in t. This is the von Neumann necessary criterion for stability; this technique of analysing stability is called the von Neumann Method. The following points concerning the von Neumann method are worth mentioning:

(i) The method only applies rigorously if the coefficients of the linear difference equation are constant; though it is conventional to apply it locally when the coefficients are not constant.

(ii) For two level difference schemes with one dependent variable and any number of independent variables the von Neumann condition is sufficient as well as necessary for stability, otherwise the condition is only necessary.

(iii) Boundary conditions are neglected in the von Neumann analysis and hence, in theory, it only applies to pure initial value problems with periodic initial data.

It is noted that stability of a difference scheme is also related to the propagation of rounding errors which occur as a result of numerical calculations. Let

$$Z(x,t) = U(x,t) - \tilde{U}(x,t) \tag{1.24}$$

be the difference between the theoretical and numerical solutions of the difference equations. Since the error $Z(x,t)$ satisfies the original difference equation, the von Neumann analysis above may be applied using $Z(x,t)$ in place of $U(x,t)$. Thus the stability condition (1.23) ensures that the rounding errors introduced will not grow as the numerical solution is advanced with time.

The Matrix Method, unlike the von Neumann method, is applicable to initial-boundary value problems. A necessary and sufficient condition for stability, when the eigenvalues λ_s of A in (1.18) are distinct , is

$$\max_{1\le s\le -1} |\lambda_s| \le 1 \quad , \tag{1.25}$$

where $Mh = 1$ and h is the space discretization. This stability condition is identical to that obtained by the von Neumann method although their respective motivations are different. In general, the two methods produce similar stability requirements, except possibly for small differences, in most problems; see for example, Morton (1980).

In Chapter 4 a grid with step size h is superimposed on the space variable x in the first order linear hyperbolic partial differential equation

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0 \quad .$$

The space derivative is approximated by central difference, lower order backward difference, and higher order backward difference replacements, and the resulting linear systems of first order ordinary differential

equations are solved employing Padé approximants to the exponential matrix function.

A number of difference schemes for solving the first order hyperbolic equation are thus developed and each is extrapolated to give higher order accuracy. The schemes are tested on a number of problems from the literature.

In Chapter 5 the second order periodic initial value problem $y'' = f(x,y)$ is considered. Recently there has been considerable interest in the approximate solution of second order initial value problem, for the cases where it is known in advance that the required solution is periodic. The well-known class of Störmer-Cowell methods with step number greater than two, give numerical solutions which do not stay on the circular orbit but spiral inwards. This phenomenon is known as orbital instability. So Störmer-Cowell methods are often unsuitable for the integration of such problems. In Chapter 5 a family of two-step numerical methods is developed. The methods are analysed, and their periodicity intervals and intervals of absolute stability are calculated. The methods are also used in PECE mode and are tested on four problems from the literature.

In Chapter 6 a number of schemes are developed for fourth order parabolic partial differential equations in one and two space dimensions. The methods are analysed for stability and are tested on problems with constant coefficients, and variable coefficients in one and two space dimensions.

Most of the numerical results contained in this thesis were computed on a CDC 7600 computer. Unless otherwise stated, single precision arithmetic was used for the calculations.

Parts of the contents of Chapter 2, 4 have been published respectively in Twizell and Khaliq (1981) and Khaliq and Twizell (1982).

CHAPTER 2

ONE - STEP METHODS FOR FIRST ORDER ORDINARY DIFFERENTIAL EQUATIONS

2.1 Introduction

Consider a first order system of ordinary differential equations of order N given by

$$\underline{y}'(x) = \underline{f}(x,\underline{y}) \ , \ \underline{y},\underline{f} \in E^N \tag{2.0}$$

for which all solutions are assumed to be bounded. In the particular case of the linear initial value problem

$$\underline{y}'(x) = A\underline{y}(x) + \underline{B} \ , \ \underline{y}(x_0) = \underline{y}_0 \ , \tag{2.1}$$

where A is a square matrix of order N with constant coefficients, this means that the real part of the eigenvalues of A must be non-positive. Equations of the form (2.1), with $\underline{B} \neq \underline{0}$ a constant vector, arise in the numerical solution of first order hyperbolic partial differential equations and second order parabolic partial differential equations with inhomogenous boundary conditions. In such problems the eigenvalues of the matrix A are real or complex depending upon the finite difference approximation to the space derivative. Equations of the form (2.1) with $\underline{B} \equiv \underline{0}$ arise in the numerical solution of homogeneous second order parabolic partial differential equations when the space derivative is replaced by the usual central difference approximation. In this case the matrix A has negative real eigenvalues and was considered by Lawson and Morris (1978) and Gourlay and Morris (1980).

The methods to be considered in this chapter will be applied to the heat equation and first order hyperbolic partial differential equation in Chapters 3 and 4, respectively. Assuming that A is diagonalizable, and following Lambert (1973), it is therefore appropriate to consider the test equation (see also, for example, Hall and Watt (1976,p.34))

$$y' = \lambda y \qquad (\lambda < 0) \qquad y(x_0) = y_0$$

and to seek the solution in some interval $x_0 = a \leq x \leq b$.
In the case of a single equation of the form (2.0), λ takes the value of $\frac{\partial f}{\partial y}$, estimated at each step.

A family of one-step multiderivative methods based on Padé approximants to the exponential function, will be developed in section 2.2.

One-step multiderivative methods are known to give high accuracy when used to solve the problems for which higher derivatives are available, see, for example, Obrechkoff (1942), Ehle (1968), Thompson (1968), Barton,Willers and Zahar (1971), Gear (1971), Lambert (1973,p.202), Brown (1974, 1976) and others.

The first twenty four members of the family are given in Appendix II; the family is seen to contain five well-known methods. In section 2.3 the methods will be analysed, and in section 2.4 a practical problem in applied chemistry will be modelled. The methodswill be extrapolated to achieve higher accuracy in section 2.5. In section 2.6 the methods will be employed in appropriate predictor-corrector pairs. Stability regions, for the case λ complex, for certain predictor-corrector pairs, will be given in section 2.7. The predictor-corrector combinations will be tested on numerical examples in section 2.8 and finally conclusions will be drawn in section 2.9.

2.2 Derivation of the formulas

Suppose the independent variable x is incremented using a constant step size $h = (b - a)/N$ where N is a positive integer, then the solution of equation (2.1) will be computed at the points $x_i = ih$ $(i = 1,2...,N)$.

It is easy to show that the solution $y(x)$ satisfies the one-step relation

$$y(x + h) = e^{\lambda h}\, y(x) \ . \tag{2.3}$$

Using this relation, any numerical method will determine the solution y_{n+1} $(n = 0,1,...,N-1)$ whose accuracy will depend on the approximation to $e^{\lambda h}$ used in (2.3). Using the (m,k) Padé approximant to $e^{\lambda h}$ of the form

$$e^{\lambda h} \simeq R_{m,k}(\lambda h) = P_k(\lambda h)/Q_m(\lambda h) + O(h^{m+k+1}) \ ,$$

where P_k, Q_m are polynomials of degree k,m, respectively, defined by

$$P_k(\theta) = 1 + p_{1,k}\theta + p_{2,k}\theta^2 + ...+ p_{k,k}\theta^k \ ; \ P_0(\theta) \equiv 1 \tag{2.4}$$

and

$$Q_m(\theta) = 1 - q_{1,m}\theta + q_{2,m}\theta^2 - ....+ (-1)^m q_{m,m}\theta^m; \ Q_0(\theta) \equiv 1, \tag{2.5}$$

with $p_{1,k} > p_{2,k} > p_{k,k} > 0$ and $q_{1,m} > q_{2,m} > \ ... \ > q_{m,m} > 0$

depending on the chosen Padé approximant, equation (2.3) takes the form

$$(1 - q_{1,m}\lambda h + q_{2,m}\lambda^2 h^2 + ...+(-1)^m q_{m,m}\lambda^m h^m)y_{n+1} = (1 + p_{1,k}\lambda h + p_{2,k}\lambda^2 h^2 + ...+ p_{k,k}\lambda^k h^k)y_n \tag{2.6}$$

or

$$y_{n+1} - q_{1,m}hy'_{n+1} + q_{2,m}h^2y''_{n+1} + ...+(-1)^m q_{m,m}h^m y^{(m)}_{n+1} = y_n + p_{1,k}hy'_n + p_{2,k}h^2y''_n + ...+ p_{k,k}h^k y^{(k)}_n \ . \tag{2.7}$$

Equation (2.7) is a one-step multiderivative formula which is explicit if $m = 0$ (Taylor's series of order k) and implicit if $m \neq 0$; it is assumed that $y(x)$ is sufficiently often differentiable on a,b .

The non-zero coefficients of (2.7) for the family of algorithms yielded by the first twenty four entries of the Padé Table for the exponential function, are given in the Appendix II. It is seen that the methods based on the (0,1), (1,1) and (3,3) Padé approximants are, respectively, the Euler predictor, the Euler corrector or trapezoidal rule and Milne's starting procedure (Milne (1949)); the methods based on the (k,k) Padé approximants ($k \geq 1$) are one-step Obrechkoff methods and are given for $k = 2,3,4$ in, for example, Lambert (1973,p.47) and Lambert and Mitchell ((1962): Table I).

2.3 Analyses of the methods

With the multiderivative formula (2.7) may be associated the linear difference operator L defined by

$$(2.8) \qquad L[y(x);\mathbf{h}] = y(x+h) - y(x) + \sum_{i=1}^{m} (-1)^m q_{i,m} h^i y^{(i)}(x+h) - \sum_{i=1}^{k} p_{i,k} h^i y^{(i)}(x).$$

Expanding $y(x+h)$ and its derivatives as Taylor series about x, and collecting terms, gives

$$(2.9) \qquad L[y(x);h] = C_0 y(x) + C_1 h y'(x) + \ldots + C_t h^t y^{(t)}(x) + \ldots.$$

where the C_t are constants. The operator L and the associated multiderivative method (2.7) are of order s if, in (3.2), $C_0 = C_1 = \ldots. = C_s = 0$, $C_{s+1} \neq 0$; the term C_{s+1} in the principal part of the truncation error is known as the error constant. The error constants for the twenty four methods to be considered, are contained in Table 2.1.

The multiderivative formula (2.7) is said to be consistent with the differential equation if the order $s \geq 1$; the twenty four methods contained in Appendix II are clearly consistent.

Writing (2.7) in the form

$$(2.10) \qquad y_{n+1} - y_n = \sum_{i=1}^{k} p_{i,k} h^i y_n^{(i)} + \sum_{j=1}^{m} (-1)^{j+1} q_{j,m} h^j y_{n+1}^{(j)}$$

it is clear that the multiderivative methods are generated by the characteristic polynomials

(2.11) $\rho(r) = r-1$, $\sigma_{i,k}(r) = p_{i,k}$, $\gamma_{j,m}(r) = (-1)^{j+1} q_{j,m} r$

$(i = 1,\ldots,k\ ;\ j = 1,\ldots,m)$. The polynomial equation $\rho(r) = 0$ has only one zero, $r = 1$, and the twenty-four consistent multiderivative methods are therefore zero-stable and thus convergent.

The interval of absolute stability of equation (2.7), is determined by computing the interval of values of $\bar{h} = \lambda h$ for which the zero of the stability equation

$$\pi(r,\bar{h}) = 0 \tag{2.12}$$

is less than unity in modulus, where

$$\begin{aligned} \pi(r,\bar{h}) &= \rho(r) - \sum_{i=1}^{k} \bar{h}^i \sigma_{i,k}(r) - \sum_{j=1}^{m} \bar{h}^j \gamma_{j,m}(r) \quad , \\ &= (1 + \sum_{j=1}^{m} (-1)^j q_{j,m} \bar{h}^j) r - (1 + \sum_{i=1}^{k} p_{i,k} \bar{h}^i) \quad , \\ &= Q_m(\bar{h}) r - P_k(\bar{h}) \quad . \end{aligned} \tag{2.13}$$

The intervals of absolute stability for the multiderivative methods based on the first twenty four Padé approximants to the exponential function, are contained in Table 2.1 (the figures containing a decimal point have been truncated with two decimal places).

The formulas based on those (m,k) Padé approximants for which $m \geq k$ are seen to be unconditionally stable. This is verified by the following theorem whose proof is based on the properties of the coefficients $p_{i,k}, q_{j,m}$ $(i = 1,\ldots,k\ ;\ j = 1,\ldots,m)$:

Theorem 1

The multiderivative method (2.7) is absolutely stable if and only if $m \geq k$ for $m,k \leq 4$.

Proof:

Assume $m \geq k$; then the coefficients in the (m,k) Padé approximant satisfy $q_{i,m} \geq p_{i,m} \geq 0$ for all $i = 1,\ldots,m$ (m,k odd or even).

Table 2.1: Stability intervals and principal error terms of the one-step multiderivative formulas.

Method (Padé)	Stability interval	error constant
(0,1)	$\bar{h} \in (-2,0)$	$C_2 = 1/2$
(1,1)	$\bar{h} \in (-\infty,0)$	$C_3 = -1/12$
(1,0)	$\bar{h} \in (-\infty,0)$	$C_2 = -1/2$
(0,2)	$\bar{h} \in (-2,0)$	$C_3 = 1/6$
(1,2)	$\bar{h} \in (-6,0)$	$C_4 = -1/72$
(2,2)	$\bar{h} \in (-\infty,0)$	$C_5 = 1/720$
(2,1)	$\bar{h} \in (-\infty,0)$	$C_4 = 1/72$
(2,0)	$\bar{h} \in (-\infty,0)$	$C_3 = 1/6$
(0,3)	$\bar{h} \in (-2.51,0)$	$C_4 = 1/24$
(1,3)	$\bar{h} \in (-5.41,0)$	$C_5 = -1/480$
(2,3)	$\bar{h} \in (-11.84,0)$	$C_6 = 1/7200$
(3,3)	$\bar{h} \in (-\infty,0)$	$C_7 = -1/100800$
(3,2)	$\bar{h} \in (-\infty,0)$	$C_6 = -1/7200$
(3,1)	$\bar{h} \in (-\infty,0)$	$C_5 = -1/480$
(3,0)	$\bar{h} \in (-\infty,0)$	$C_4 = -1/24$
(0,4)	$\bar{h} \in (-2.78,0)$	$C_5 = 1/120$
(1,4)	$\bar{h} \in (-5.43,0)$	$C_6 = -1/3600$
(2,4)	$\bar{h} \in (-9.64,0)$	$C_7 = 1/75600$
(3,4)	$\bar{h} \in (-19.15,0)$	$C_8 = -1/1411200$
(4,4)	$\bar{h} \in (-\infty,0)$	$C_9 = 1/25401600$
(4,3)	$\bar{h} \in (-\infty,0)$	$C_8 = 1/1411200$
(4,2)	$\bar{h} \in (-\infty,0)$	$C_7 = 1/75600$
(4,1)	$\bar{h} \in (-\infty,0)$	$C_6 = 1/3600$
(4,0)	$\bar{h} \in (-\infty,0)$	$C_5 = 1/120$

The requirement $|r|<1$ leads to

(2.14)
$$-1 < \frac{1 + p_{1,k}\bar{h} + p_{2,k}\bar{h}^2 + \ldots + p_{k,k}\bar{h}^k}{1 - q_{1,m}\bar{h} + q_{2,m}\bar{h}^2 - \ldots + (-1)^m q_{m,m}\bar{h}^m} < 1.$$

The left hand side implies the requirement

$$2 + (p_{1,k} - q_{1,m})\bar{h} + (p_{2,k} + q_{2,m})\bar{h}^2 + \ldots + (p_{k,k} + (-1)^k q_{k,m}\bar{h}^k)$$
$$+ (-1)^{k+1} q_{k+1,m}\bar{h}^{k+1} + \ldots + (-1)^m q_{m,m}\bar{h}^m > 0$$

and, since $q_{i,m} \geq p_{i,k} \geq 0$ for $m \geq k$ (m,k odd or even), Padé (1892), this inequality is satisfied for $\bar{h} < 0$. The right hand side of (2.14) implies the requirement

$$(p_{1,k} + q_{1,m})\bar{h} + (p_{2,k} - q_{2,m})\bar{h}^2 + \ldots + (p_{k,k} - (-1)^k q_{k,m})\bar{h}^k$$
$$- (-1)^{k+1} q_{k+1,m}\bar{h}^{k+1} + \ldots - (-1)^m q_{m,m}\bar{h}^m < 0$$

and this inequality is also satisfied for $\bar{h} < 0$.

The multiderivative method given by (2.7) is thus absolutely stable if $m \geq k$ and $m,k \leq 4$.

If $m < k$ the method has only a finite interval of absolute stability as illustrated, for example, by the (0,1) method which is the Euler predictor formula. The hypothesis of the theorem is thus proved.

The methods based on the (k,k) Padé approximants, are optimal in that they have the smallest truncation errors ; they are absolutely stable. When used as correctors in PECE mode, however, they give smaller intervals of absolute stability, when used with the $(0,\ell)$ method as predictor ($\ell = 1,\ldots,k$), than the methods with $m < k$. This will be dealt with more fully in Section 2.6.

From Theorem 1 it is clear that one-step multiderivative methods of the form (2.7), based on the (m,k) Padé approximants with $m \geq k$, satisfy the definition of A_0-stability (Cryer (1973)). A_0-stability corresponds to "unconditional stability" for second order parabolic partial

differential equations, when the eigenvalues of the discretization matrix are real and negative. For example, using the (1,1) Padé approximant in (2.7), yields the trapezoidal rule, which is A_0-stable when applied to the test equation (2.3); it becomes the Crank-Nicholson method for second order parabolic partial differential equations, which is known to be unconditionally stable (Lawson and Morris (1978)).

The boundaries of stability regions for λ complex, can also be calculated from equation (2.12) by imposing $|r| = 1$, see for example, Hall and Watt (1976,p.38). The stability regions of the methods based on the (m,k) Padé approximants, for $4 \geq m \geq k$ are seen to contain the left half complex plane, thus satisfying the requirement of A-stability (Dahlquist (1963)). See also, Axelsson (1969), Ehle (1968). The amplification symbols for the (m,k) Padé approximants, for $m \geq k$, are shown in Figures 2.1-2.14. For the (m,k) Padé approximants with $m > k$, the amplification symbol approaches zero either monotonically or asymptotically by crossing the axis. For A-stable methods based on the (k,k) Padé approximants, the amplification symbol is

$$R_{k,k}(\bar{h}) = \frac{P(\bar{h})}{Q(\bar{h})} = \frac{P(\bar{h})}{P(-\bar{h})}$$

where $P(\bar{h})$ is defined in (2.4), $\bar{h} = \lambda h$, and is such that

$$R_{k,k}(\bar{h}) \to \pm 1 \quad \text{as} \quad \text{Re}\,(\bar{h}) \to -\infty .$$

The numerical methods of the form (2.7), applied to problems with rapidly decaying solutions, will thus not damp any oscillations. The trapezoidal rule (the (1,1) Padé approximant), is well known to have this property (Rosenbrock (1963)).

To overcome this difficulty a stronger stability property is defined which has been variously termed L-stability (Ehle (1969), Lambert (1973, p.237)), stiff A-stability (Axelsson (1969)), and strong A-stability

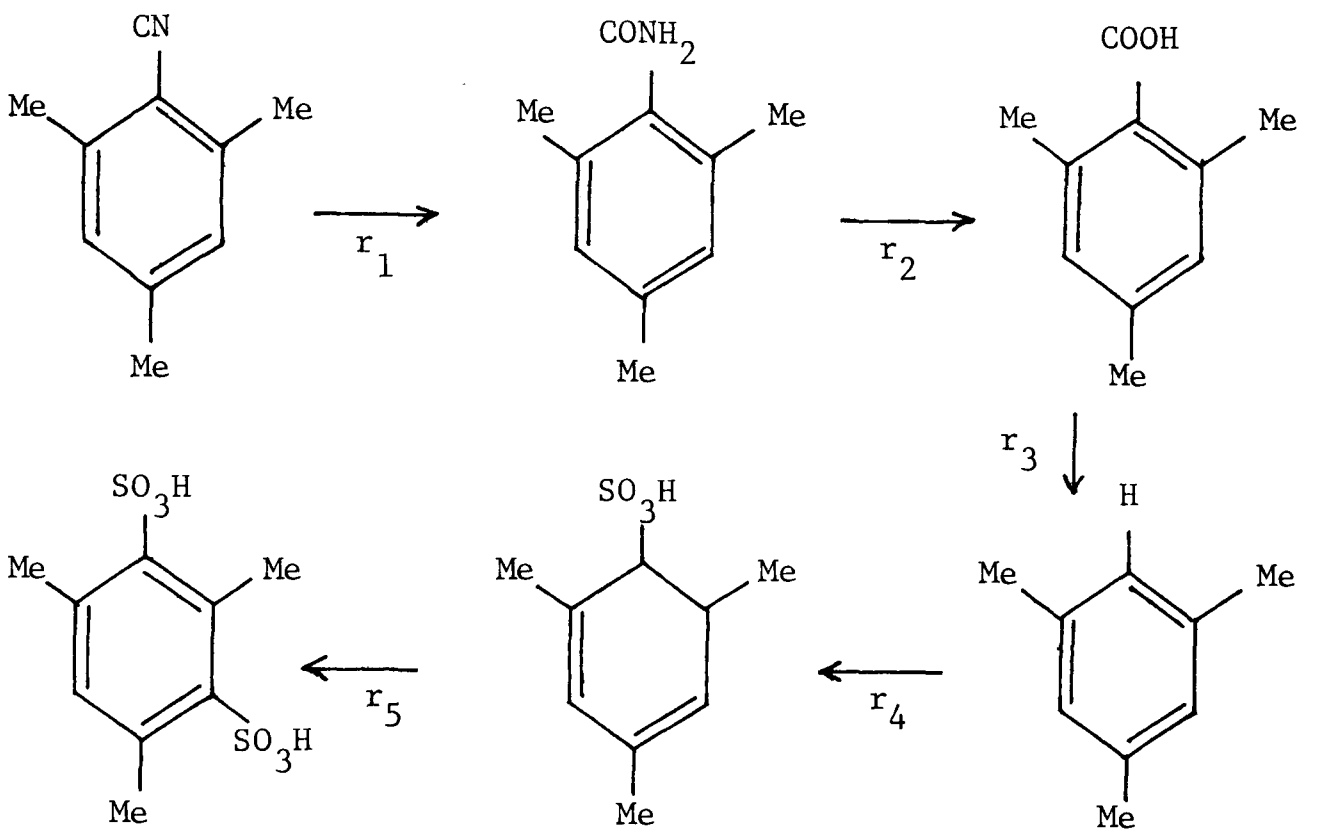
(Chipman (1971), Axelsson (1972)). Following Ehle (1969), Lambert (1973, p.236) has made the following definition of L-stability:

Definition: A one-step numerical method is said to be L-stable if it is A-stable and, in addition, when applied to the scalar test equation $y' = \lambda y$, λ a complex constant with $\mathrm{Re}\ \lambda < 0$, it yields $y_{n+1} = R(h\lambda)y_n$, where $|R(h\lambda)| \to 0$ as $\mathrm{Re}(h\lambda) \to -\infty$.

One-step multiderivative methods of the form (2.7) yielded by employing the (m,k) Padé approximants, for $m > k$, are thus L-stable; this is also clear from the corresponding Figures. It is noted that the amplification symbols for L-stable methods approach zero rapidly as soon as the degree of m increases compared to that of k, and hence oscillations will be damped quickly by employing higher order Padé approximants for which $m > k$. The behaviour of higher order (m,0) Padé approximants and corresponding (m,k) Padé approximants, for $m > k$, will be discussed in Chapter 3 for parabolic partial differential equations in which discontinuities exist between initial and boundary values.

2.4 Mathematical modelling of a Chemistry problem

Consider the sequence of first order reactions, described by the chain reaction below:



It is the reaction of Mesitonitrite in Sulphuric Acid. A discussion of the above reaction can be found in Gore et al (1983). The research for this problem was carried out at Brunel University by J. Al 'Kabi, E. F. Saad, D. N. Waters and G. F. Moxon, under Professor P. H. Gore, Department of Applied Chemistry.

The chemical reactions have been expressed in the form of the following initial value problem:

$$\frac{dy_1}{dt} = -r_1 y_1 \quad ; \quad y_1(0) = 1$$

$$\frac{dy_2}{dt} = r_1 y_1 - r_2 y_2 \quad ; \quad y_2(0) = 0$$

$$\frac{dy_3}{dt} = r_2 y_2 - r_3 y_3 \quad ; \quad y_3(0) = 0$$

$$\frac{dy_4}{dt} = r_3y_3 - r_4y_4 \quad ; \quad y_4(0) = 0$$

$$\frac{dy_5}{dt} = r_4y_4 - r_5y_5 \quad ; \quad y_5(0) = 0$$

$$\frac{dy_6}{dt} = r_5y_5 \quad ; \quad y_6(0) = 0 \quad .$$

This is a linear system which can also be written as

(2.15)
$$\frac{d\underline{y}}{dt} = A\underline{y} \quad ; \quad \underline{y}(0) = [1,0,0,0,0,0]^T$$

where the matrix A is of order 6 and is given by

$$A = \begin{bmatrix} -r_1 & & & & & \\ r_1 & -r_2 & & & 0 & \\ & r_2 & -r_3 & & & \\ & & r_3 & -r_4 & & \\ & 0 & & r_4 & -r_5 & \\ & & & & r_5 & 0 \end{bmatrix}$$

with $r_1 = 0.0006605$, $r_2 = 0.0009185$, $r_3 = 0.01694$, $r_4 = 1818.0$ $r_5 = 0.0004834$.

The theoretical solution of the problem is

$$y_1(t) = e^{-r_1t} \quad ,$$

$$y_2(t) = \frac{r_1}{r_2-r_1}\left[e^{-r_1t}-e^{-r_2t}\right] \quad ,$$

$$y_3(t) = r_1r_2\left[\frac{e^{-r_1t}}{(r_2-r_1)(r_3-r_1)} + \frac{e^{-r_2t}}{(r_1-r_2)(r_3-r_2)} + \frac{e^{-r_3t}}{(r_1-r_3)(r_2-r_3)}\right] \quad ,$$

$$y_4(t) = r_1r_2r_3\left[\frac{e^{-r_1t}}{(r_2-r_1)(r_3-r_1)(r_4-r_1)} + \frac{e^{-r_2t}}{(r_1-r_2)(r_3-r_2)(r_4-r_2)}\right.$$

$$+ \frac{e^{-r_3 t}}{(r_1-r_3)(r_2-r_3)(r_4-r_3)} + \frac{e^{-r_4 t}}{(r_1-r_4)(r_2-r_4)(r_3-r_4)} \Bigg] ,$$

$$y_5(t) = r_1 r_2 r_3 r_4 \Bigg[\frac{e^{-r_1 t}}{(r_2-r_1)(r_3-r_1)(r_4-r_1)(r_5-r_1)}$$

$$+ \frac{e^{-r_2 t}}{(r_1-r_2)(r_3-r_2)(r_4-r_2)(r_5-r_2)}$$

$$+ \frac{e^{-r_3 t}}{(r_1-r_3)(r_2-r_3)(r_4-r_3)(r_5-r_3)}$$

$$+ \frac{e^{-r_4 t}}{(r_1-r_4)(r_2-r_4)(r_3-r_4)(r_5-r_4)}$$

$$+ \frac{e^{-r_5 t}}{(r_1-r_5)(r_2-r_5)(r_3-r_5)(r_4-r_5)} \Bigg] ,$$

$$y_6(t) = 1 - \Bigg[\frac{r_2 r_3 r_4 r_5 \; e^{-r_1 t}}{(r_2-r_1)(r_3-r_1)(r_4-r_1)(r_5-r_1)}$$

$$+ \frac{r_1 r_3 r_4 r_5 \; e^{-r_2 t}}{(r_1-r_2)(r_3-r_2)(r_4-r_2)(r_5-r_2)}$$

$$+ \frac{r_1 r_2 r_4 r_5 \; e^{-r_3 t}}{(r_1-r_3)(r_2-r_3)(r_4-r_3)(r_5-r_3)}$$

$$+ \frac{r_1 r_2 r_3 r_5 \; e^{-r_4 t}}{(r_5-r_4)(r_2-r_4)(r_3-r_4)(r_5-r_4)}$$

$$+ \frac{r_1 r_2 r_3 r_4 \; e^{-r_5 t}}{(r_1-r_5)(r_2-r_5)(r_3-r_5)(r_4-r_5)} \Bigg] ,$$

and is such that the components of $\underline{y}(t)$ add up to unity at each time step.

The eigenvalues of A in (2.15) are widely spread with maximum modulus eigenvalue 1818.0 and minimum modulus eigenvalue zero, thus making the system highly stiff. The definition of a stiff system (Lambert (1973, p.231-232)) in which the stiffness ratio is a measure of computational effort is not valid for this type of problem where the minimum modulus eigenvalue is zero. However, in this case statements such as "Stiffness

occurs when stability rather than accuracy dictates the choice of step length", are preferred (Lambert (1980,p.21).

To compute the solution of the system (2.15), it can easily be shown that (2.15) satisfies the equation (2.3), which can be written as

$$\underline{y}(t+\ell) = \exp(\ell A)\underline{y}(t) , \tag{2.16}$$

where ℓ is a convenient time step.

The fourth order, A-stable method based on the (2,2) Padé approximant (Appendix II) may be used to determine the solution from (2.16). The numerical results were calculated using single precision arithmetic and the sum of the six components of $\underline{y}(t)$, $t = 0(500)5000$, was found to be unity to ten decimal places. The numerical results are given in the Table 2.2.

Table 2.2:

Computed solution of the modelled problem at time t = 0(500)5000.

Time	y_1	y_2	y_3	y_4	y_5	y_6	Sum
500	7.1875(-1)	2.2266(-1)	1.1661(-2)	1.0867(-7)	4.3459(-2)	3.4741(-3)	1.0000
1000	5.1660(-1)	3.0070(-1)	1.6177(-2)	1.5073(-7)	1.4105(-1)	2.5476(-2)	1.0000
1500	3.7130(-1)	3.0500(-1)	1.6656(-2)	1.5520(-7)	2.3566(-1)	7.1380(-2)	1.0000
2000	2.6687(-1)	2.7537(-1)	1.5174(-2)	1.4139(-7)	3.0525(-1)	1.3734(-1)	1.0000
2500	1.9182(-1)	2.3339(-1)	1.2936(-2)	1.2054(-7)	3.4532(-1)	2.1654(-1)	1.0000
3000	1.3787(-1)	1.9016(-1)	1.0581(-2)	9.8598(-8)	3.5924(11)	3.0215(-1)	1.0000
3500	9.9092(-2)	1.5083(-1)	8.4169(-3)	7.8428(-8)	3.5308(-1)	3.8858(-1)	1.0000
4000	7.1222(-2)	1.1736(-1)	6.5624(-3)	6.1148(-8)	3.3313(-1)	4.7172(-1)	1.0000
4500	5.1191(-2)	9.8001(-2)	5.0406(-3)	4.6968(-8)	3.0482(-1)	5.4894(-1)	1.0000
5000	3.6793(-2)	6.8259(-2)	3.8276(-3)	3.5666(-8)	2.7237(-1)	6.1875(-1)	1.0000

2.5 Extrapolation of the methods

Applying equation (2.3) over two single intervals h and replacing $e^{2\lambda h}$ by, for example, its $(1,1)$ Padé approximant, gives

$$(2.17)\quad y(x+2h) = (1+\tfrac{1}{2}\lambda h)(1-\tfrac{1}{2}\lambda h)^{-1}(1+\tfrac{1}{2}\lambda h)(1-\tfrac{1}{2}\lambda h)^{-1}y(x)$$

$$= (1+2\lambda h+2\lambda^2h^2+\tfrac{3}{2}\lambda^3h^3+\lambda^4h^4+\tfrac{5}{8}\lambda^5h^5)y(x)+0(h^6)$$

$$= y(x)+2hy'(x)+2h^2y''(x)+\tfrac{3}{2}h^3y'''(x)+h^4y^{(iv)}(x)+\tfrac{5}{8}h^5y^{(v)}(x)+0(h^6).$$

Alternatively, if equation (2.3) is written over a double interval $2h$, $y(x+2h)$ is given by

$$(2.18)\quad y(x+2h) = (1+\lambda h)(1-\lambda h)^{-1}y(x)$$

$$= (1+2\lambda h+2\lambda^2h^2+2\lambda^3h^3+2\lambda^4h^4+2\lambda^5h^5)y(x)+0(h^6)$$

$$= y(x)+2hy'(x)+2h^2y''(x)+2h^3y'''(x)+2h^4y^{(iv)}(x)+2h^5y^{(v)}(x)+0(h^6).$$

The Maclaurin expansion of $y(x+2h)$ about x produces

$$(2.19)\quad y(x+2h) = y(x)+2hy'(x)+2h^2y''(x)+\tfrac{4}{3}h^3y'''(x)+\tfrac{2}{3}h^4y^{(iv)}(x)+\tfrac{4}{15}h^5y^{(v)}(x)$$

$$+\ \tfrac{4}{45}h^6y^{(vi)}(x)+\tfrac{8}{315}h^7y^{(vii)}(x)+\tfrac{2}{315}h^8y^{(viii)}(x)+\tfrac{4}{2835}h^9y^{(ix)}(x)$$

$$+0(h^{10}),$$

and defining the values of $y(x+2h)$ yielded by (2.17) and (2.18) to be $y_{n+2}^{(1)}$ and $y_{n+2}^{(2)}$ respectively, it is seen that neither is $0(h^3)$ accurate. However, defining $y_{n+2}^{(E)}$ by

$$y_{n+2}^{(E)} = \tfrac{4}{3}y_{n+2}^{(1)} - \tfrac{1}{3}y_{n+2}^{(2)}$$

gives

$$(2.20)\quad y_{n+2}^{(E)} = y(x)+2hy'(x)+2h^2y''(x)+\tfrac{4}{3}h^3y'''(x)+\tfrac{2}{3}h^4y^{(iv)}(x)+0(h^5).$$

The error in $y_{n+2}^{(E)}$ defined by $y(x+2h) - y_{n+2}^{(E)}$, has principal part $E_5 = \frac{1}{10}$. The second order method based on the $(1,1)$ Padé approximant, has been extrapolated to give fourth order accuracy (see also Lindberg (1971)) by the Richardson technique.

Repeating the process for the (3,3) Padé method (Milne's method (1949)) leads to

$$y_{n+2}^{(1)} \equiv [(1+\tfrac{1}{2}\lambda h+\tfrac{1}{10}\lambda^2h^2+\tfrac{1}{120}\lambda^3h^3)(1-\tfrac{1}{2}\lambda h+\tfrac{1}{10}\lambda^2h^2-\tfrac{1}{120}\lambda^3h^3)^{-1}]^2 y(x)$$

$$= y(x)+2hy'(x)+2h^2y''(x)+\tfrac{4}{3}h^3y'''(x)+\tfrac{2}{3}h^4y^{(iv)}(x)+\tfrac{4}{15}h^5y^{(v)}(x)$$

$$+\tfrac{4}{45}h^6y^{(vi)}(x)+\tfrac{61}{2400}h^7y^{(vii)}(x)+\tfrac{23}{3600}h^8y^{(viii)}(x)+\tfrac{209}{144000}h^9y^{(ix)}(x)$$

$$+0(h^{10})$$

and

$$y_{n+2}^{(2)} \equiv (1+h+\tfrac{2}{5}\lambda^2h^2+\tfrac{1}{15}\lambda^3h^3)(1-\lambda h+\tfrac{2}{5}\lambda^2h^2-\tfrac{1}{15}\lambda^3h^3)^{-1}y(x)$$

$$= y(x)+2hy'(x)+2h^2y''(x)+\tfrac{4}{3}h^3y'''(x)+\tfrac{2}{3}h^4y^{(iv)}(x)+\tfrac{4}{15}h^5y^{(v)}(x)$$

$$+\tfrac{4}{45}h^6y^{(vi)}(x)+\tfrac{2}{75}h^7y^{(vii)}(x)+\tfrac{2}{225}h^8y^{(viii)}(x)+\tfrac{14}{3375}h^9y^{(ix)}(x)+0(h^{10}).$$

Defining $y_{n+2}^{(E)}$, in this case, by

$$y_{n+2}^{(E)} = \tfrac{64}{63}y_{n+2}^{(1)} - \tfrac{1}{63}y_{n+2}^{(2)} \qquad (2.21)$$

gives

$$y_{n+2}^{(E)} = y(x)+2hy'(x)+2h^2y''(x)+\tfrac{4}{3}h^3y'''(x)+\tfrac{2}{3}h^4y^{(iv)}(x)+\tfrac{4}{15}h^5y^{(v)}(x)$$

$$+\tfrac{4}{45}h^6y^{(vi)}(x)+\tfrac{8}{315}h^7y^{(vii)}(x)+\tfrac{2}{315}h^8y^{(viii)}(x)+\tfrac{599}{425250}h^9y^{(ix)}+0(h^{10}),$$

which, on comparison with equation (2.17), is seen to be eighth order accurate with $E_9 = \frac{1}{425250}$. It is clear that as m and k increase, the algebraic manipulation involved in the extrapolation procedure becomes tedious and difficult.

In the cases of the methods based on the (1,1) and (3,3) Padé approximants, the extrapolation procedure has produced two extra orders of accuracy. This phenomenon is a useful feature of multiderivative methods based on (m,m) Padé approximants, which is not evident in methods based on (m,k) Padé approximants ($m \neq k$) for which only a single extra order of accuracy is produced.

The extrapolating formulas connecting $y_{n+2}^{(E)}$, $y_{n+2}^{(1)}$ and $y_{n+2}^{(2)}$ satisfy one of the relations

$$(2.22)\qquad y_{n+2}^{(E)} = (2^{m+k}y_{n+2}^{(1)} - y_{n+2}^{(2)})/(2^{m+k} - 1) + 0(h^{m+k+2})$$

when $m \neq k$, or

$$(2.23)\qquad y_{n+2}^{(E)} = (2^{2m}y_{n+2}^{(1)} - y_{n+2}^{(2)})/(2^{2m} - 1) + 0(h^{2m+3})$$

when $m = k$. The extrapolation formulas for the twenty four multi-derivative methods outlined in Section 2.2 together with the error constantsofthe principal parts of their local truncation errors, defined for each method by

$$(2.24)\qquad y(x+2h) - y_{n+2}^{(E)} \quad ,$$

are contained in Table 2.3.

It is easy to see that $y_{n+2}^{(E)}$ may also be written in the form

$$(2.25)\qquad y_{n+2}^{(E)} = \frac{1}{2^{m+k}-1}\left[2^{m+k}\frac{P_k(\lambda h)}{Q_m(\lambda h)}^2 - \frac{P_k(2\lambda h)}{Q_m(2\lambda h)}\right]y_n + 0(h^{m+k+2}) \;;\; m \neq k$$

or

$$(2.26)\qquad y_{n+2}^{(E)} = \frac{1}{2^{2m}-1}\left[2^{2m}\frac{P_m(\lambda h)}{P_m(-\lambda h)}^2 - \frac{P_m(2\lambda h)}{P_m(-2\lambda h)}\right]y_n + 0(h^{2m+3}) \;;\; m = k.$$

Each of (2.23) and (2.24) is of the approximate form

$$(2.27)\qquad y_{n+2}^{(E)} \simeq S_{m,k}(\bar{h})y_n$$

and clearly the interval of absolute stability for each multiderivative method is the range of values of $\bar{h} = \lambda h$ for which

$$|S_{m,k}| < 1.$$

The intervals of absolute stability for equations (2.22) and (2.23), the extrapolated forms of equation (2.7), are thus determined by finding the range of values of $\bar{h}$ for which

$$(2.28)\qquad (-2^{m+k}+1)[Q_m(\bar{h})]^2Q_m(2\bar{h}) < 2^{m+k}[P_k(\bar{h})]^2Q_m(2\bar{h}) - P_k(2\bar{h})[Q_m(\bar{h})]^2 < (2^{m+k}-1)[Q_m(\bar{h})]^2Q_m(2\bar{h})$$

when $m \neq k$, or

$$(2.29)\qquad (-2^{2m}+1)[P_m(-\bar{h})]^2P_m(-2\bar{h}) < 2^{2m}[P_m(\bar{h})]^2P_m(-2\bar{h}) - P_m(2\bar{h})[P_m(-\bar{h})]^2$$

$$< (2^{2m}-1)[P_m(-\bar{h})]^2P_m(-2\bar{h})$$

when $m = k$.

Thus, for example, the interval of absolute stability for the extrapolated form of the method based on the (1,1) Padé approximant, is the interval of values of $\bar{h}$ for which

$$(2.30)\qquad -12 + 24\bar{h} - 15\bar{h}^2 + 3\bar{h}^3 < 12 - 9\bar{h}^2 - 5\bar{h}^3 < 12 - 24\bar{h} + 15\bar{h}^2 - 3\bar{h}^3 ,$$

where fractions have been cleared. The left hand side of (2.30) is satisfied for all $\bar{h} < 0$ while the right hand side is satisfied only for the interval $\bar{h} \in (-12.92,0)$, which is therefore the interval of absolute stability.

Clearly, as m and k increase, the algebraic manipulation involved in solving (2.28) or (2.29), becomes complicated. The interval of absolute stability of the extrapolated form of the multiderivative method based on the (3,3) Padé approximant, for example, is found by solving the inequality

$$\begin{aligned}(2.31)\qquad & -13608000 + 27216000\bar{h} - 25174800\bar{h}^2 + 14061600\bar{h}^3 - 5193720\bar{h}^4 \\ & +1315440\bar{h}^5 - 229257\bar{h}^6 + 26649\bar{h}^7 - 1840\bar{h}^8 + 63\bar{h}^9 \\ & < 13608000 - 2041200\bar{h}^2 + 204120\bar{h}^4 - 27783\bar{h}^6 - 8775\bar{h}^7 \\ & -1134\bar{h}^8 - 65\bar{h}^9 \\ & < 13608000 - 27216000\bar{h} + 25174800\bar{h}^2 - 14061600\bar{h}^3 + 5193720\bar{h}^4 \\ & -1315440\bar{h}^5 + 229257\bar{h}^6 - 26649\bar{h}^7 + 1840\bar{h}^8 - 63\bar{h}^9 ,\end{aligned}$$

where, again, fractions have been cleared. Both sides of (2.29) are satisfied for all $\bar{h} < 0$ and the interval of absolute stability is therefore $\bar{h} \in (-\infty,0)$.

The intervals of absolute stability for the extrapolated forms of all twenty four multiderivative methods derived in section 2.2, are also contained in Table 2.3. It must be noted that, whilst extrapolation has

improved accuracy, this has often been at the expense of a decreased interval of absolute stability. This is particularly so with the (0,1) and (1,1) Padé methods which are, of course, the Euler predictor formula and the Euler corrector formula (the trapezoidal rule) respectively. The extrapolated form of the (1,1) method does not satisfy Theorem 1 which, therefore, does not hold for the extrapolation formulas. However, it is seen from equation (2.25) that the extrapolation of L-stable methods based on (m,k) Padé approximants with $m > k$, satisfies the condition of L-stability. Thus, the extrapolation of L-stable methods of the form (2.7) based on (2.25), is L-stable since the degree of the denominator in $R_{m,k}(\bar{z})$ is greater than the degree of the numerator for $m > k$.

The amplification symbols for the extrapolated methods are also shown in Figures 2.1 - 2.14. It is seen that the amplification symbols of the extrapolated methods based on the (m,k) Padé approximants for $m > k$, approach zero faster than those of the methods themselves, thus damping oscillations more quickly.

Table 2.3: The extrapolating algorithms.

Method (Padé)	Extrapolating algorithm	Stability interval	error constant
(0,1)	$2y^{(1)}-y^{(2)}$	$\bar{h} \in (-1,0)$	$E_3 = 4/3$
(1,1)	$(4y^{(1)}-y^{(2)})/3$	$\bar{h} \in (-12.92,0)$	$E_5 = 1/10$
(1,0)	$2y^{(1)}-y^{(2)}$	$\bar{h} \in (-\infty,0)$	$E_3 = 4/3$
(0,2)	$(4y^{(1)}-y^{(2)})/3$	$\bar{h} \in (-2.57,0)$	$E_4 = 1/3$
(1,2)	$(8y^{(1)}-y^{(2)})/7$	$\bar{h} \in (-6.47,0)$	$E_5 = -8/945$
(2,2)	$(16y^{(1)}-y^{(2)})/15$	$\bar{h} \in (-\infty,0)$	$E_7 = -1/1890$
(2,1)	$(8y^{(1)}-y^{(2)})/7$	$\bar{h} \in (-\infty,0)$	$E_5 = -8/945$
(2,0)	$(4y^{(1)}-y^{(2)})/3$	$\bar{h} \in (-\infty,0)$	$E_4 = -1/3$
(0,3)	$(8y^{(1)}-y^{(2)})/7$	$\bar{h} \in (-2.02,0)$	$E_5 = 8/105$
(1,3)	$(16y^{(1)}-y^{(2)})/15$	$\bar{h} \in (-6.20,0)$	$E_6 = -1/540$
(2,3)	$(32y^{(1)}-y^{(2)})/31$	$\bar{h} \in (-11.44,0)$	$E_7 = 4/5425$
(3,3)	$(64y^{(1)}-y^{(2)})/63$	$\bar{h} \in (-\infty,0)$	$E_9 = 1/425250$
(3,2)	$(32y^{(1)}-y^{(2)})/31$	$\bar{h} \in (-\infty,0)$	$E_7 = 4/5425$
(3,1)	$(16y^{(1)}-y^{(2)})/15$	$\bar{h} \in (-\infty,0)$	$E_6 = 1/540$
(3,0)	$(8y^{(1)}-y^{(2)})/7$	$\bar{h} \in (-\infty,0)$	$E_5 = 8/105$
(0,4)	$(16y^{(1)}-y^{(2)})/15$	$\bar{h} \in (-3.23,0)$	$E_6 = 2/135$
(1,4)	$(32y^{(1)}-y^{(2)})/31$	$\bar{h} \in (-12.30,0)$	$E_7 = 8/27125$
(2,4)	$(64y^{(1)}-y^{(2)})/63$	$\bar{h} \in (-9.62,0)$	$E_8 = -1079/127575$
(3,4)	$(128y^{(1)}-y^{(2)})/127$	$\bar{h} \in (-7.98,0)$	$E_9 = 93341/88211025$
(4,4)	$(256y^{(1)}-y^{(2)})/255$	$\bar{h} \in (-\infty,0)$	$E_{11} = -1/144317250$
(4,3)	$(128y^{(1)}-y^{(2)})/127$	$\bar{h} \in (-\infty,0)$	$E_9 = 93341/88211025$
(4,2)	$(64y^{(1)}-y^{(2)})/63$	$\bar{h} \in (-\infty,0)$	$E_8 = 1079/127575$
(4,1)	$(32y^{(1)}-y^{(2)})/31$	$\bar{h} \in (-\infty,0)$	$E_7 = 8/27125$
(4,0)	$(16y^{(1)}-y^{(2)})/15$	$\bar{h} \in (-\infty,0)$	$E_6 = -2/135$

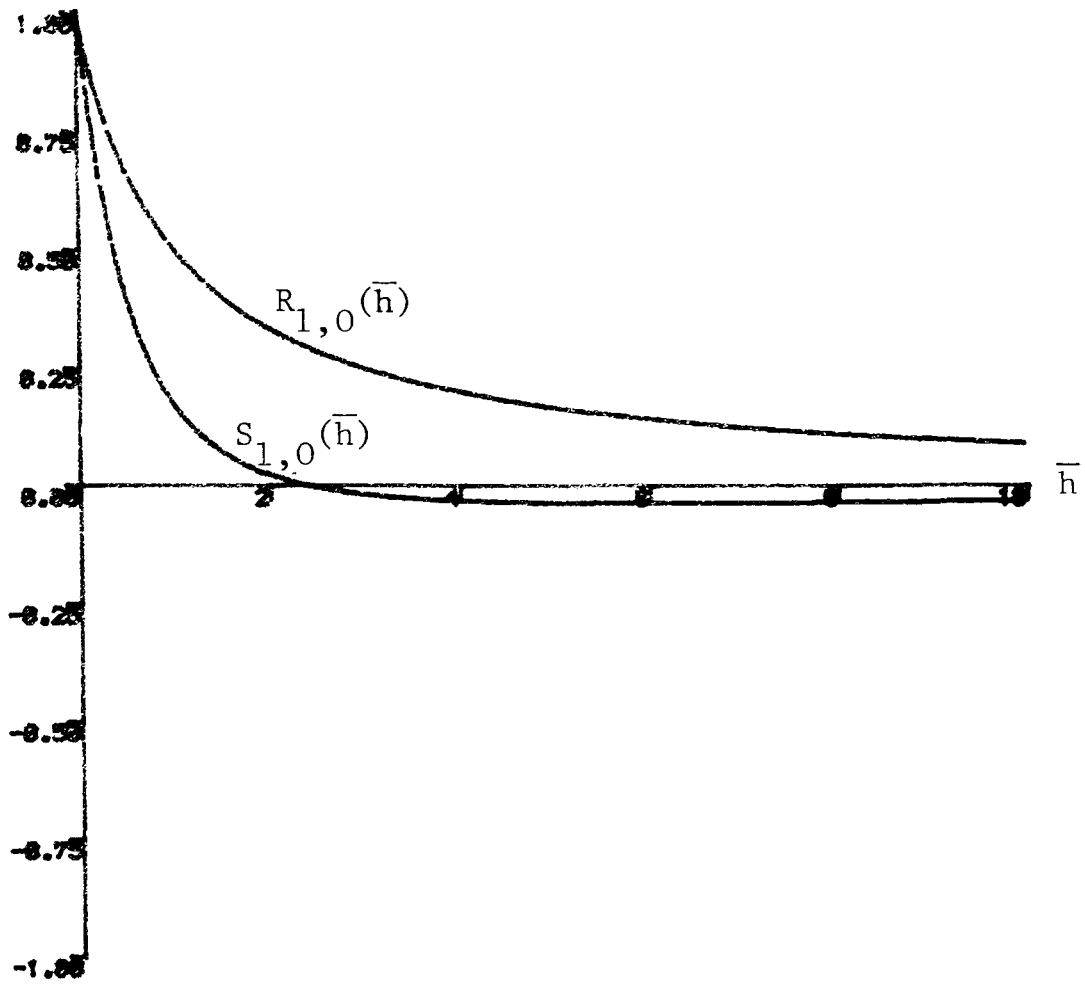


Figure 2.1: Amplification symbols $R_{1,0}(\bar{h})$ and $S_{1,0}(\bar{h})$.

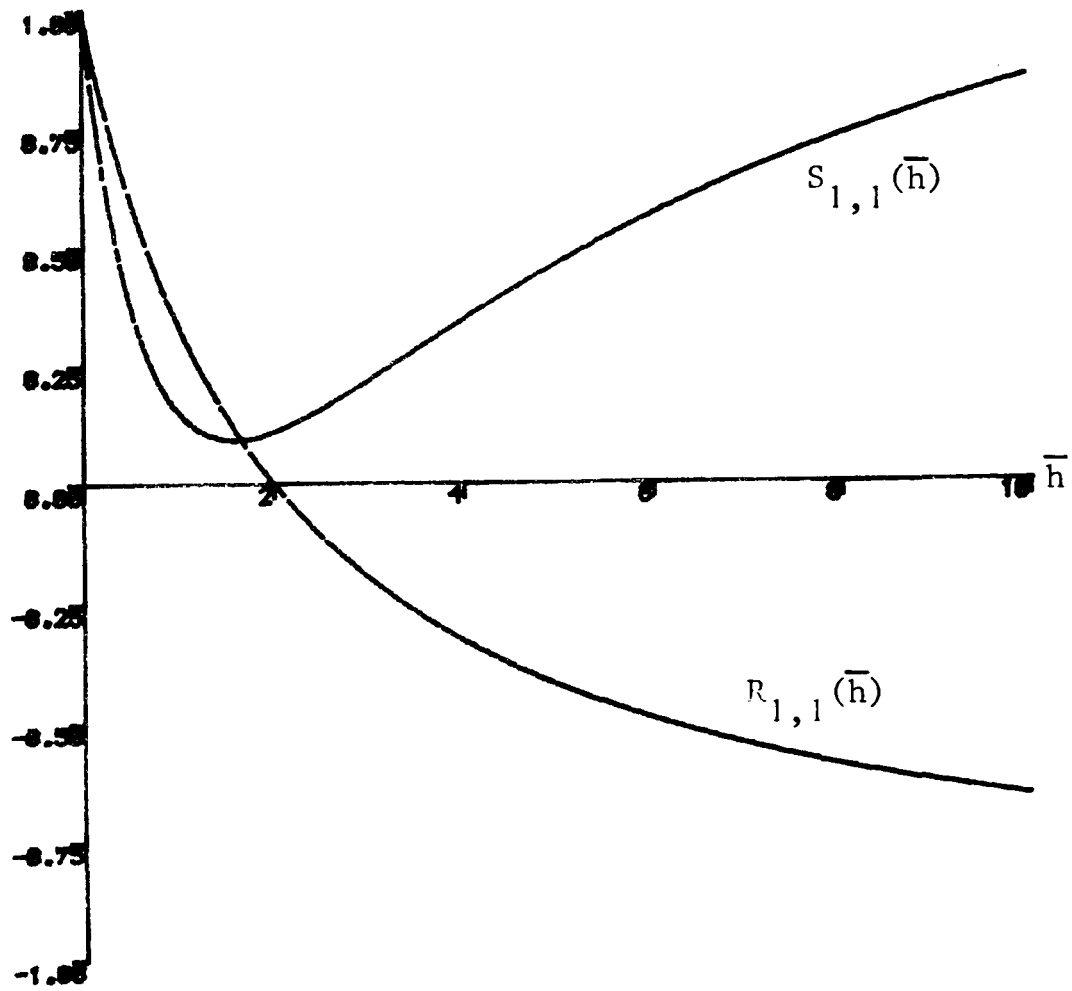


Figure 2.2: Amplification symbols $R_{1,1}(\bar{h})$ and $S_{1,1}(\bar{h})$.

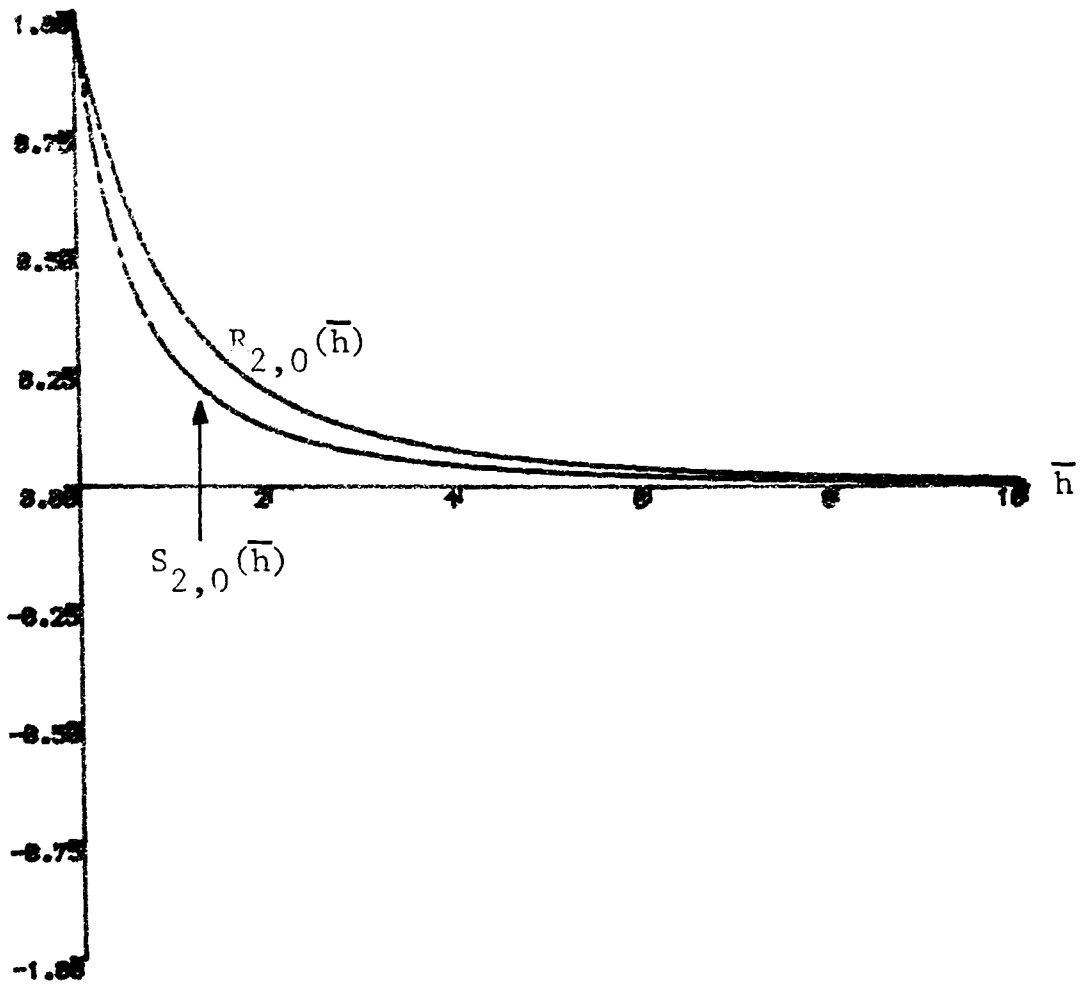


Figure 2.3: Amplification symbols $R_{2,0}(\bar{h})$ and $S_{2,0}(\bar{h})$.

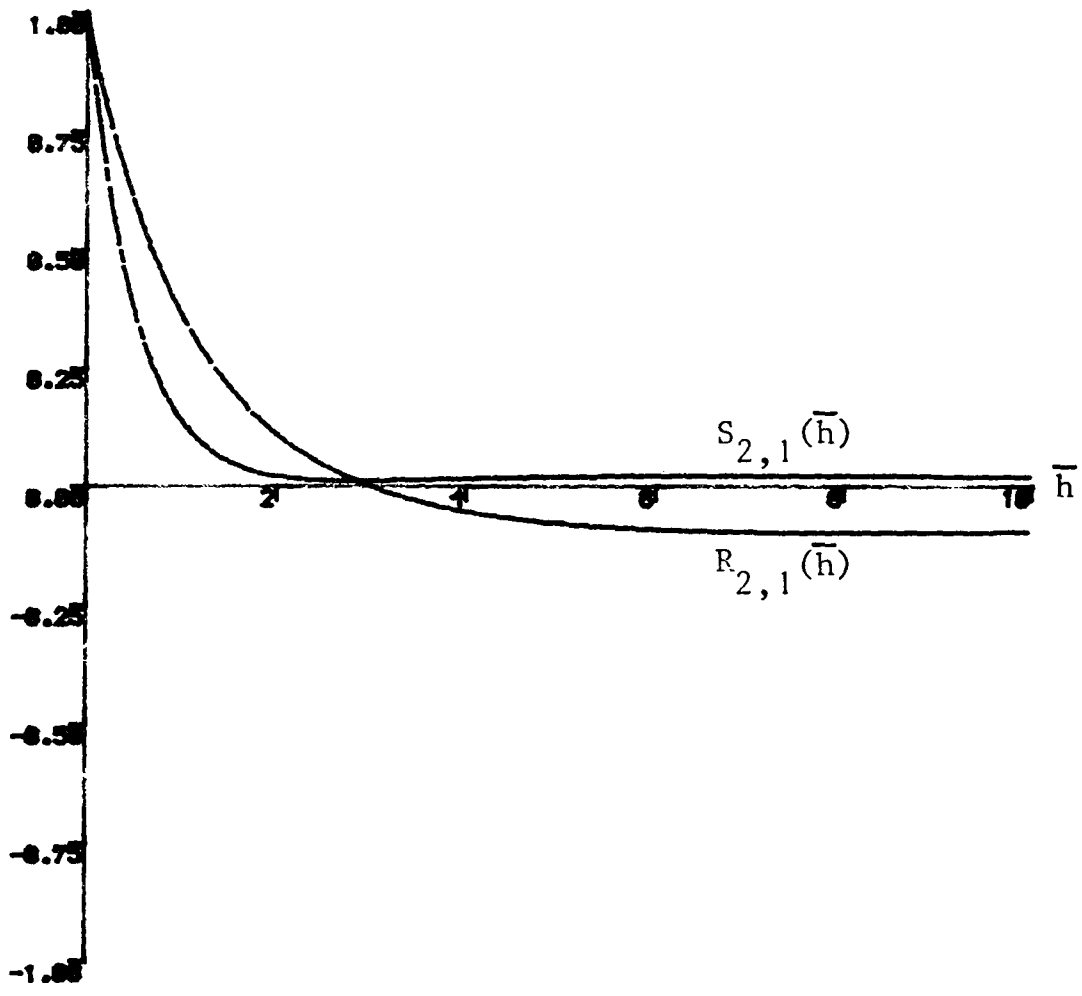


Figure 2.4: Amplification symbols $R_{2,1}(\bar{h})$ and $S_{2,1}(\bar{h})$.

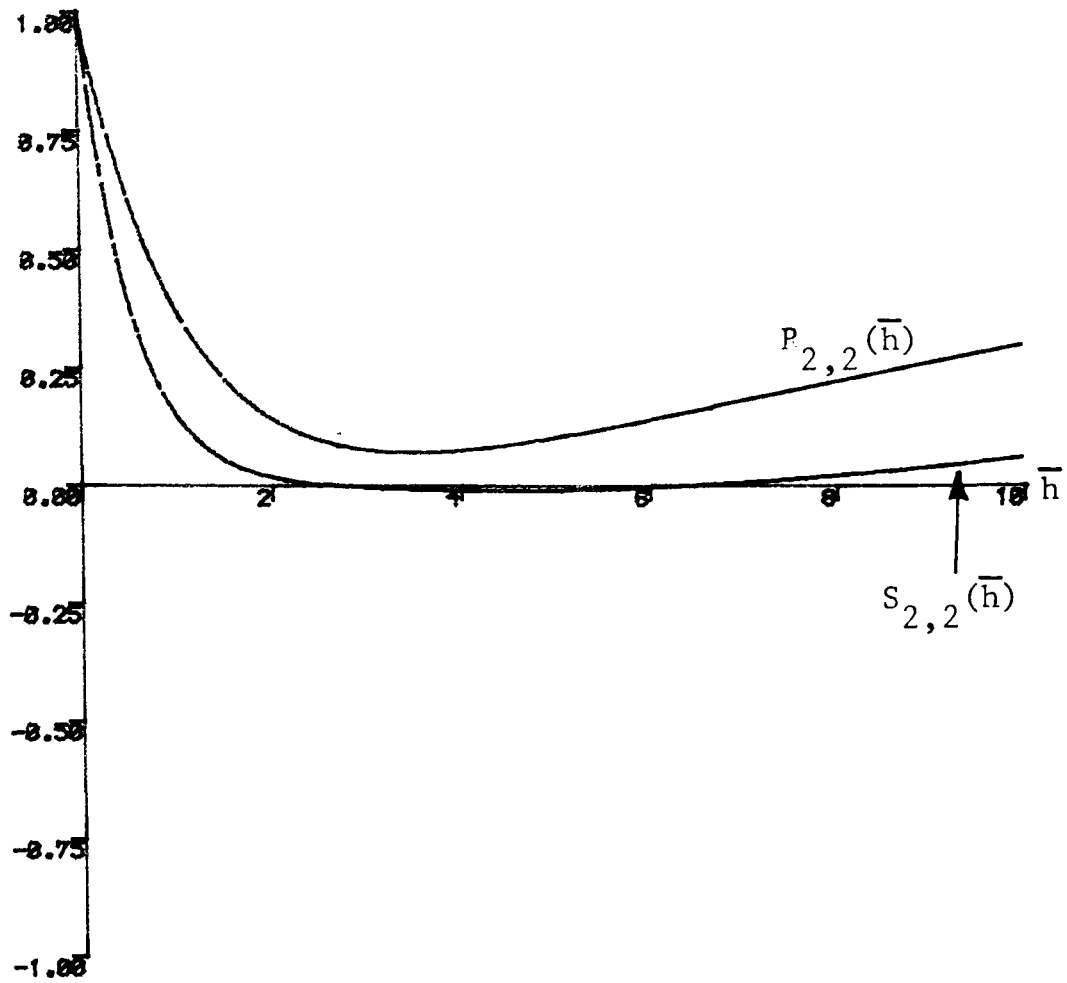


Figure 2.5: Amplification symbols $R_{2,2}(\bar{h})$ and $S_{2,2}(\bar{h})$.

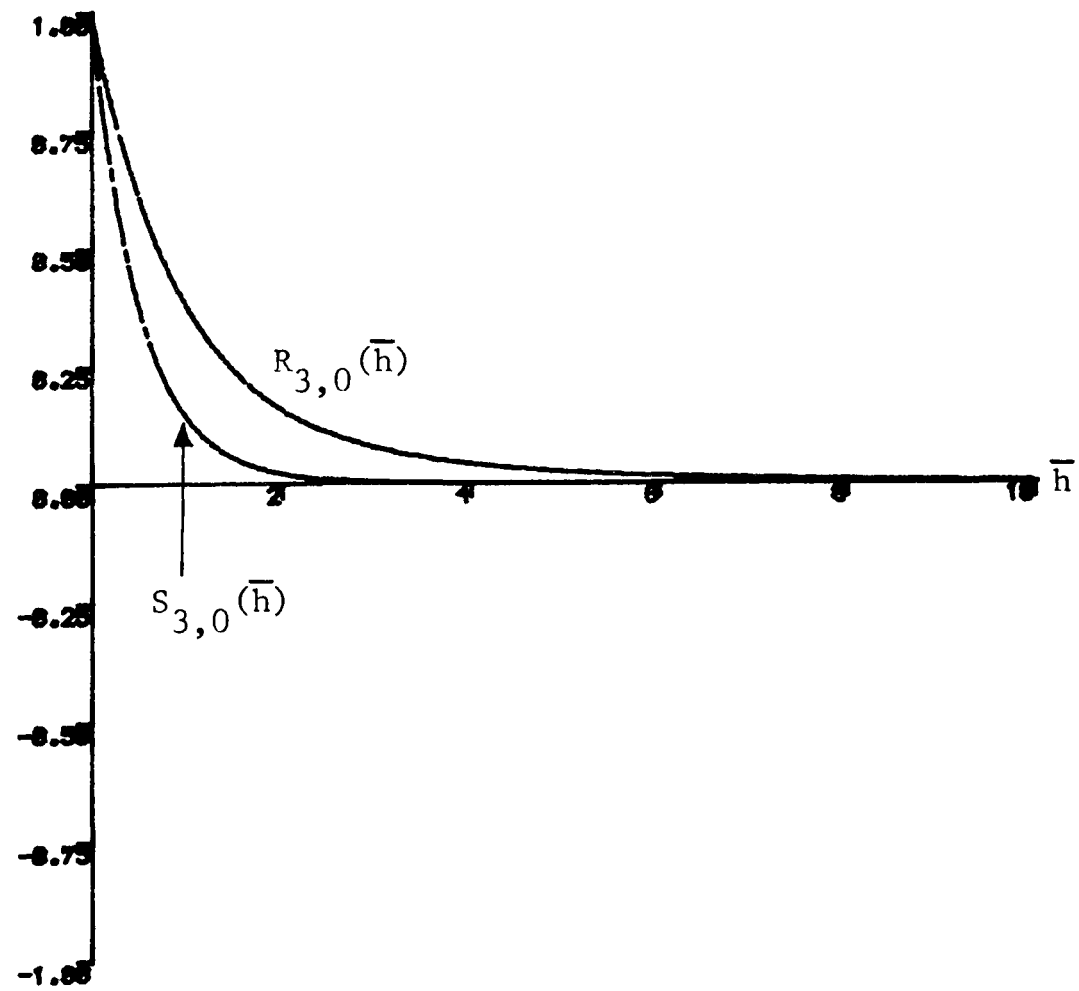


Figure 2.6: Amplification symbols $R_{3,0}(\bar{h})$ and $S_{3,0}(\bar{h})$.

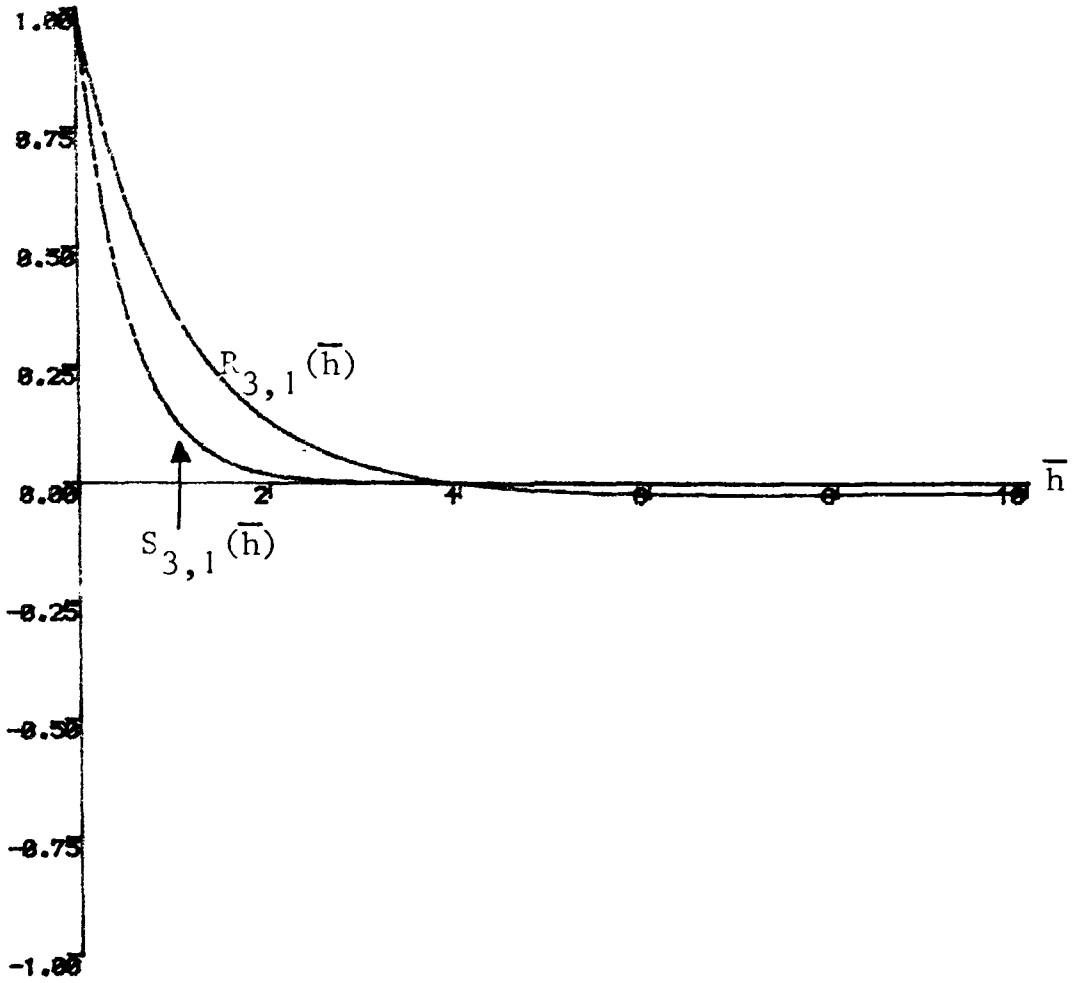


Figure 2.7: Amplification symbols $R_{3,1}(\bar{h})$ and $S_{3,1}(\bar{h})$.

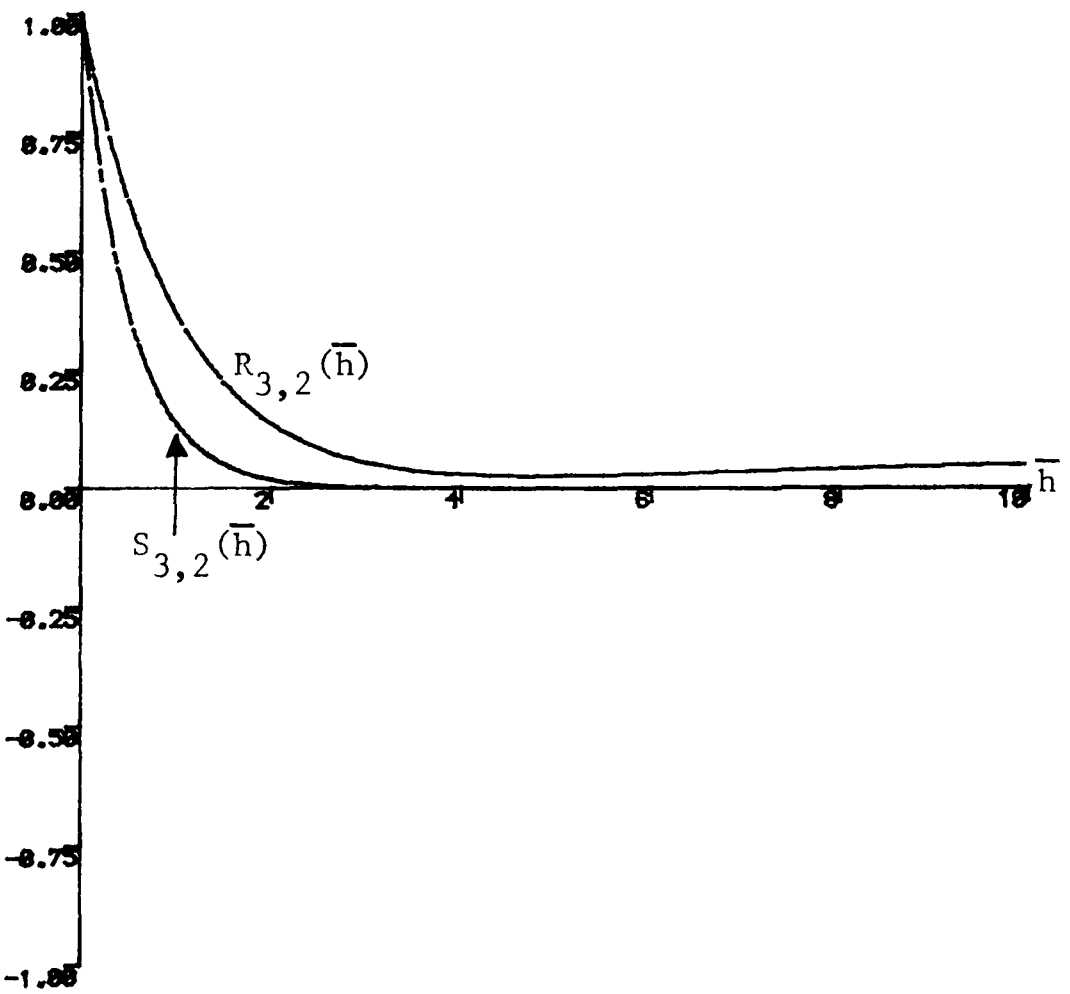


Figure 2.8: Amplification symbols $R_{3,2}(\bar{h})$ and $S_{3,2}(\bar{h})$.

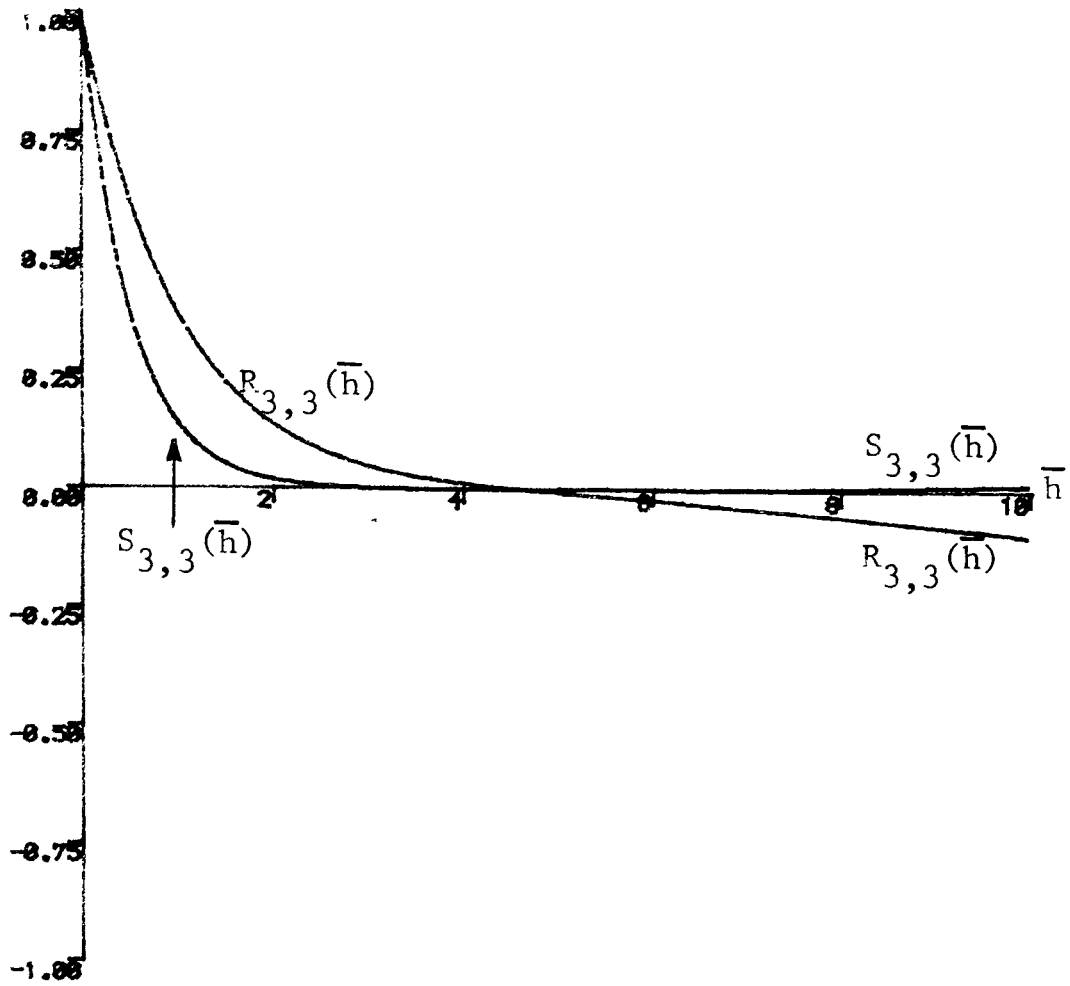


Figure 2.9: Amplification symbols $R_{3,3}(\bar{h})$ and $S_{3,3}(\bar{h})$.

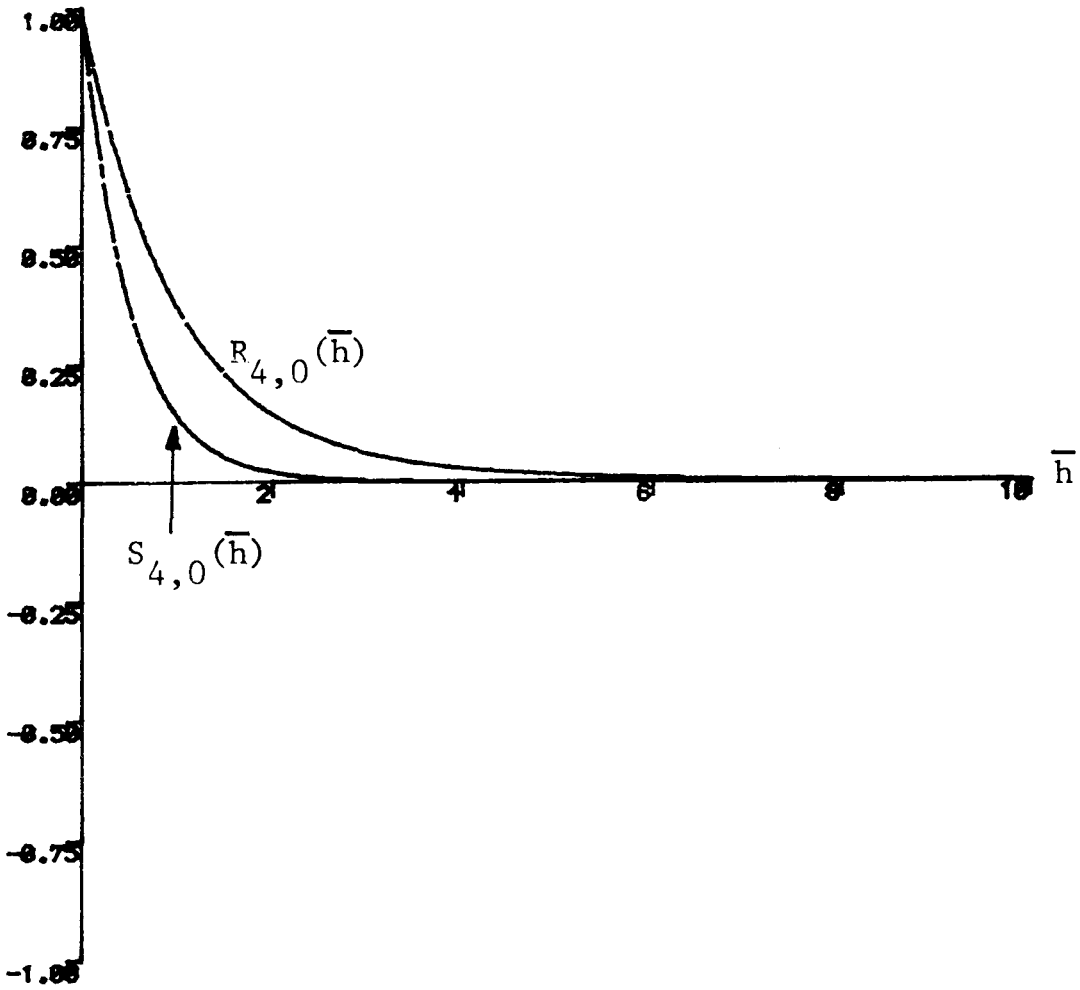


Figure 2.10: Amplification symbols $R_{4,0}(\bar{h})$ and $S_{4,0}(\bar{h})$.

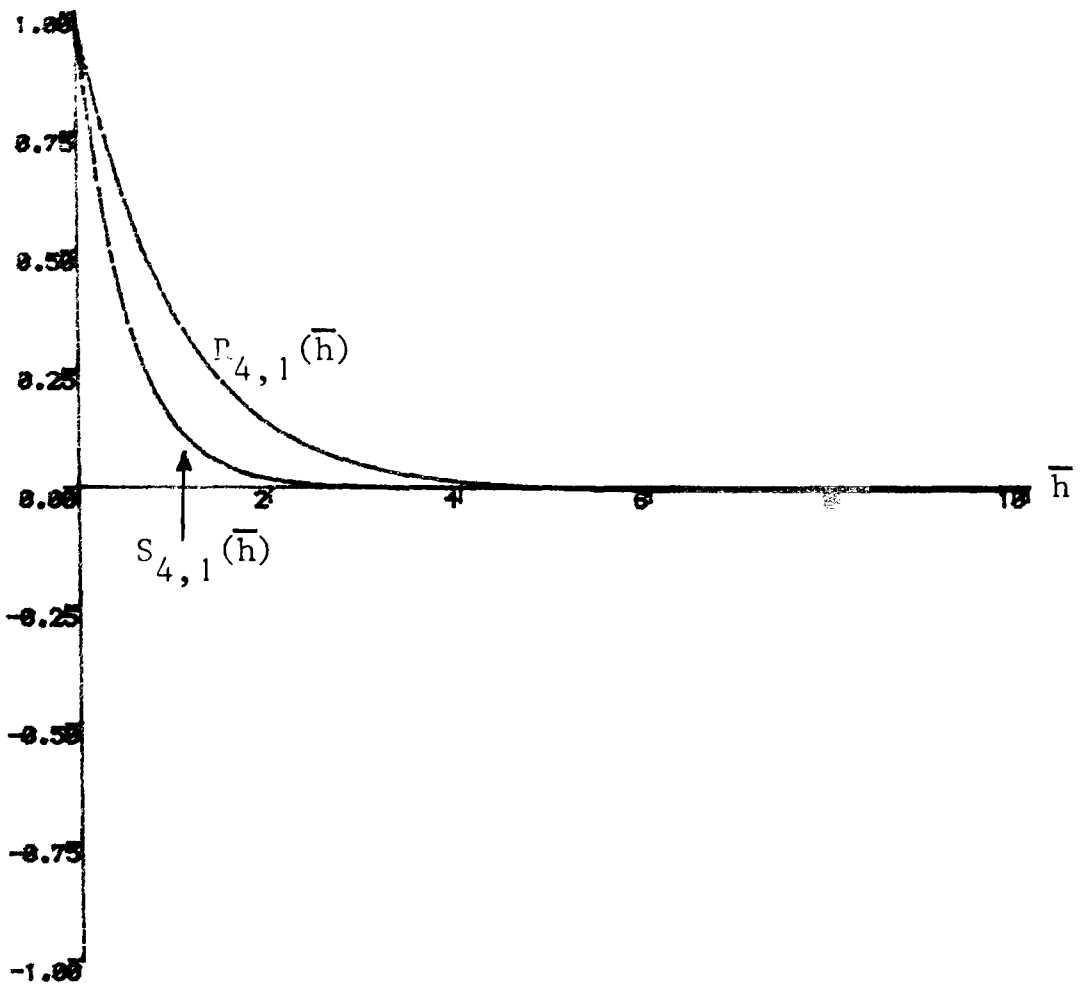


Figure 2.11: Amplification symbols $R_{4,1}(\bar{h})$ and $S_{4,1}(\bar{h})$.

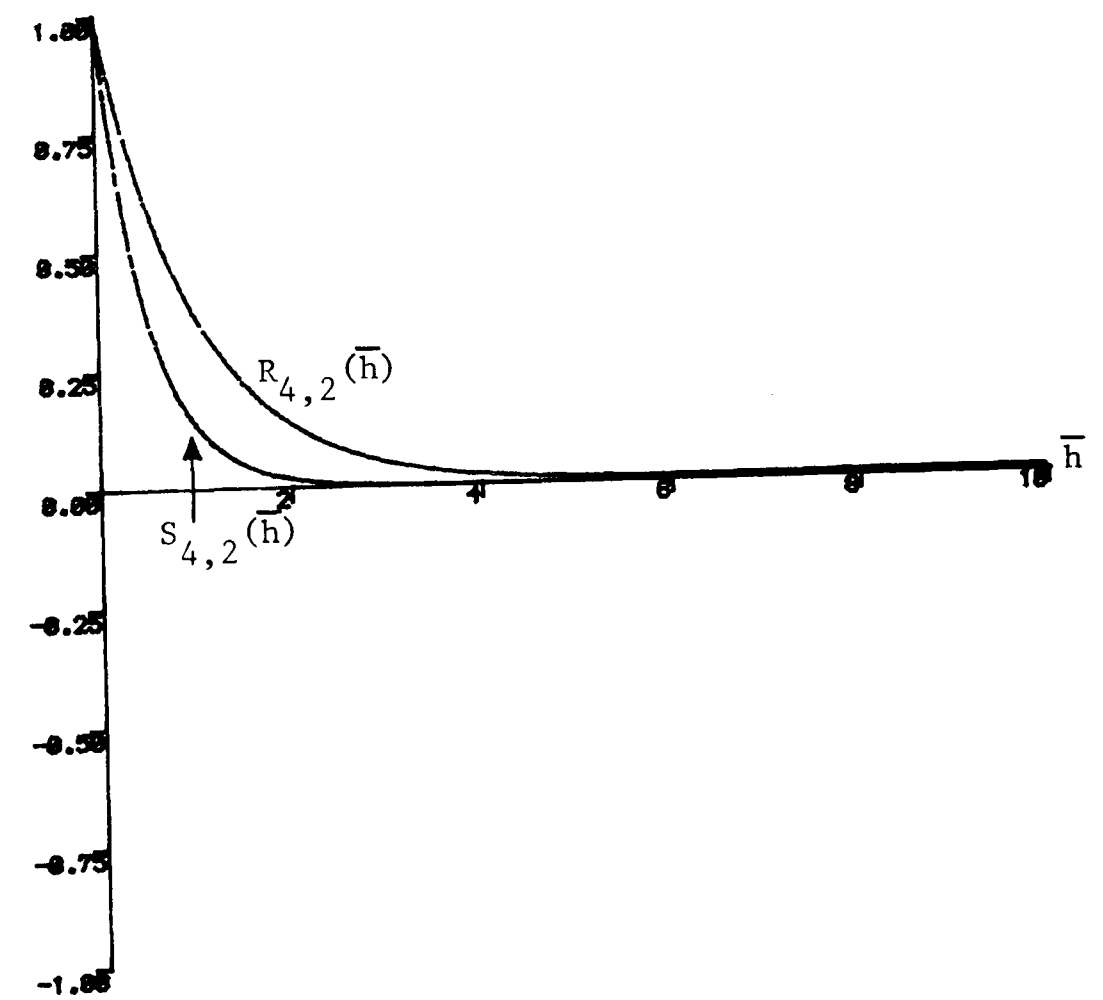


Figure 2.12: Amplification symbols $R_{4,2}(\bar{h})$ and $S_{4,2}(\bar{h})$.

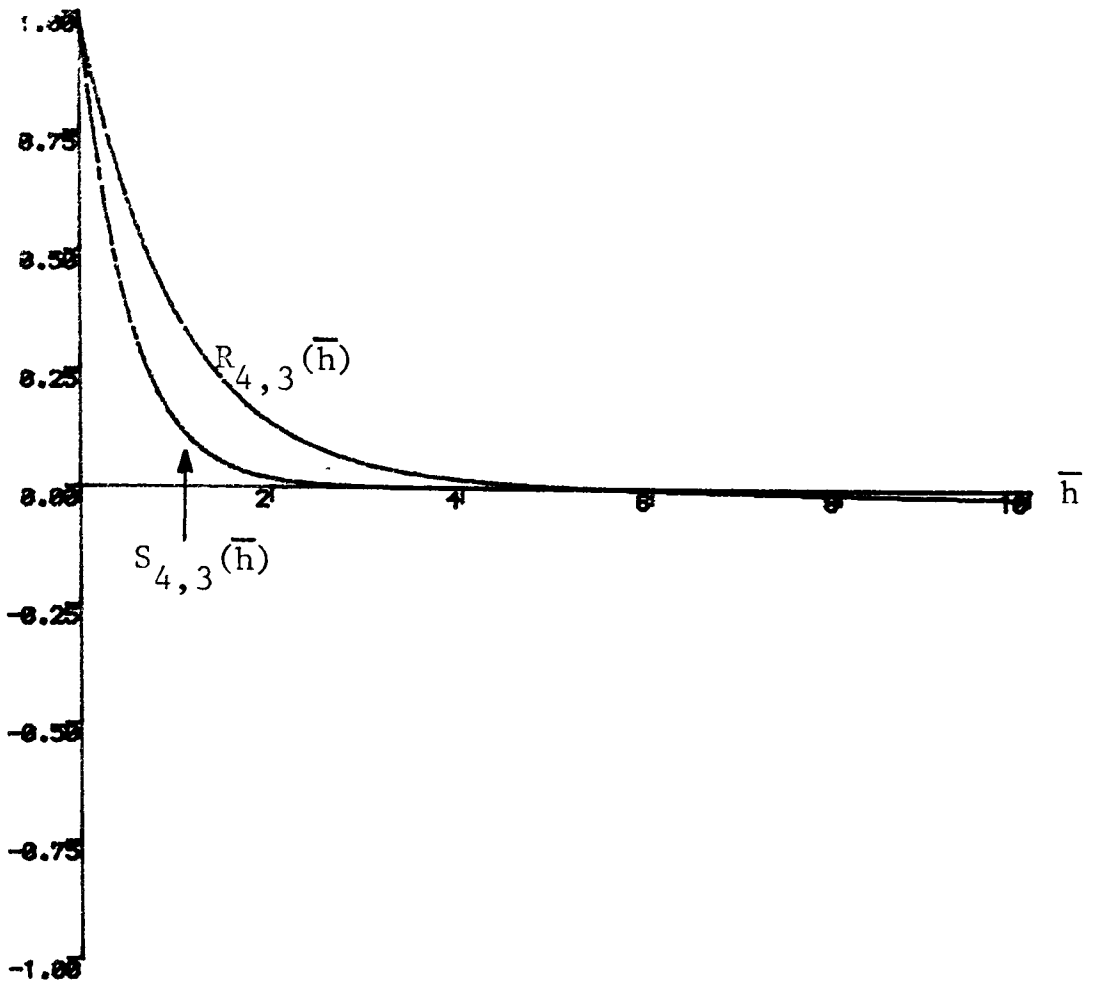


Figure 2.13: Amplification symbols $R_{4,3}(\bar{h})$ and $S_{4,3}(\bar{h})$.

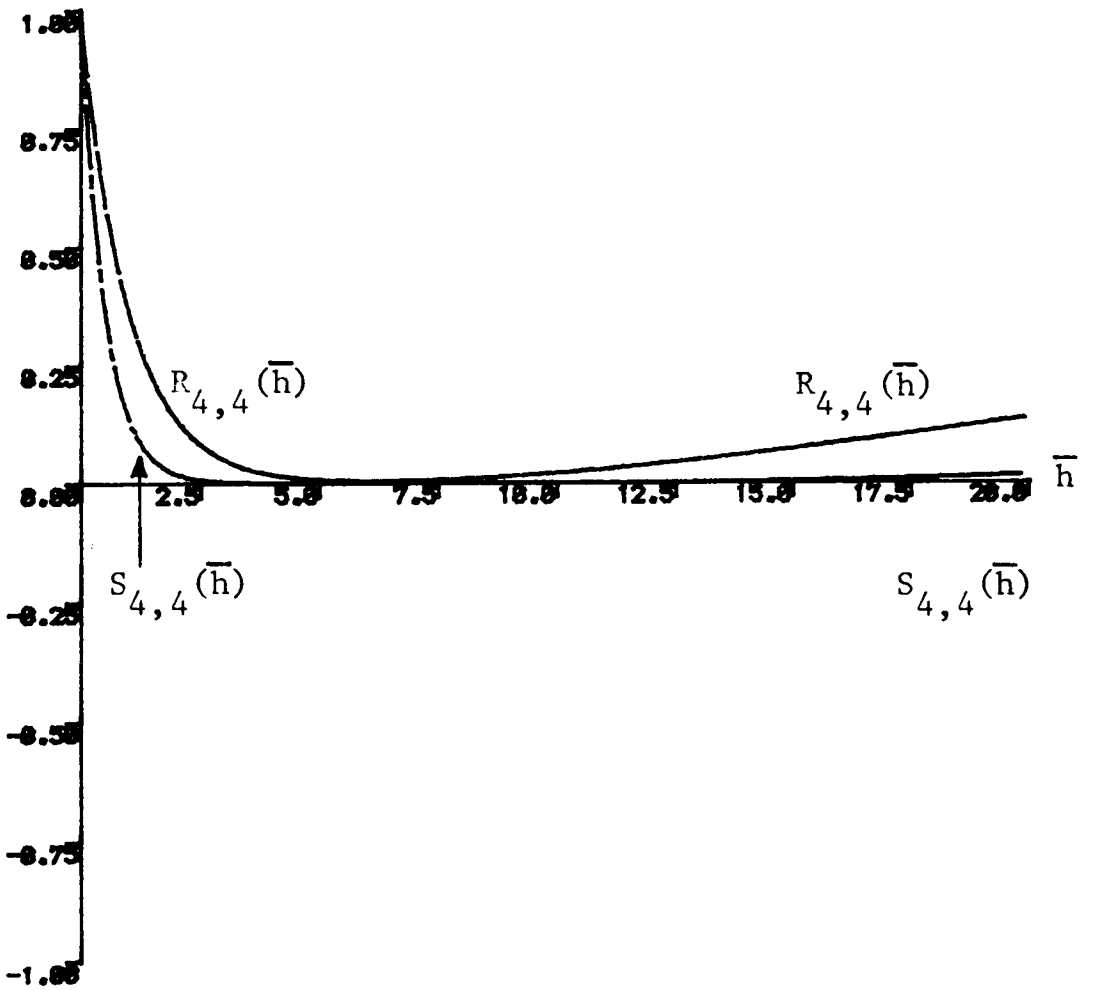


Figure 2.14: Amplification symbols $R_{4,4}(\bar{h})$ and $S_{4,4}(\bar{h})$.

2.6 Use in PECE mode

In this section the (0,1), (0,2), (0,3), (0,4) explicit formulas will be used as predictor formulas and all appropriate combinations of these four formulas with the twenty implicit formulas of section 2.2 as correctors, will be considered. Predictor-corrector methods for which the order of the predictor exceeds that of the corrector will not be constructed.

Using the general (0,k*) Padé approximant as predictor, the characteristic polynomials (from(2.11)), are

$$\rho^*(r) = r-1 \quad , \quad \sigma^*_{i,k^*}(r) = p_{i,k^*} \tag{2.32}$$

where the convention of associating an asterisk with the predictor has been adopted. Using the (m,k) Padé approximant ($m \neq 0$) as corrector, the characteristic polynomials (2.11) become

$$\rho(r) = r-1 \ , \ \sigma_{i,k}(r) = p_{i,k} \ (i=1,\ldots,k) \ , \ \gamma_{j,m}(r) = (-1)^{j+1} q_{j,m} r \ (j=1,\ldots,m).$$

This combination of predictor and corrector will be denoted by (0,k*);(m,k).

The stability polynomial for the (0,k*) ; (m,k) predictor-corrector combination in PECE mode is therefore

$$\begin{aligned} \pi_{PECE}(r,\bar{h}) &= \rho(r) - \sum_{i=1}^{k} \bar{h}^i \sigma_{i,k}(r) - \sum_{j=1}^{m} \bar{h}^j \gamma_{j,m}(r) \\ &\quad + \sum_{j=1}^{m} \bar{h}^j (-1)^{j+1} q_{j,m} \left[\rho^*(r) - \sum_{i=1}^{k^*} \bar{h}^i \sigma^*_{i,k}(r) \right] \\ &= r-1 - \sum_{i=1}^{k} \bar{h}^i p_{i,k} \\ &\quad + \sum_{j=1}^{m} (-1)^j q_{j,m} \bar{h}^j \left[1 + \sum_{i=1}^{k^*} p_{i,k^*} \bar{h}^i \right] \end{aligned} \tag{2.33}$$

and the interval of absolute stability is the range of values of $\bar{h}$ for which the zero r of

$$\pi_{PECE}(r,\bar{h}) = 0 \tag{2.34}$$

is less than unity in modulus.

Solving equation (2.34) for r gives

$$r = e^{\bar{h}} - T_{s+1}\bar{h}^{s+1} + O(\bar{h}^{s+2}) \tag{2.35}$$

where s is the order of the predictor-corrector combination $(0,k^*)$; (m,k). The term T_{s+1} is the error constant of the predictor-corrector combination.

The intervals of absolute stability and the error constants are contained in Tables 2.4, 2.5, 2.6 and 2.7 for the predictor-corrector combinations using, respectively, the $(0,1)$, $(0,2)$, $(0,3)$, $(0,4)$ Padé methods as predictors. All possible combinations of these explicit predictors with the other twenty implicit methods used as correctors, for which the order of the predictor does not exceed that of the corrector, are included in the tables.

It is easy to see that for all four predictors, using the $(1,4)$ method as corrector, gives the greatest interval of absolute stability as well as the smallest error modulus ; in the case of the $(0,3)$; $(1,3)$ combination, one derivative fewer is required in the corrector than in the $(0,3)$; $(1,4)$ combination for the same accuracy and the same interval of absolute stability.

For all four $(0,k)$ predictors, $k = 1,2,3,4$, it is seen that the $(0,k)$; $(k,0)$ predictor-corrector combination, gives the worst error in PECE mode and the smallest interval of absolute stability, except that the $(0,2)$; $(4,0)$ combination has a slightly smaller stability interval than the $(0,2)$; $(2,0)$ combination. This latter combination does, however, have a better principal error term and requires lower order derivatives.

The literature contains little on the size of stability intervals for one-step multiderivative methods used in PECE mode. They have been verified to be generally small, and examination of Tables 2.4, 2.5, 2.6 and 2.7, shows surprisingly that the greatest stability intervals in PECE mode arise with correctors based on $(1,k)$ formulas which themselves have poor stability intervals (Table 2.1). It can be deduced from Tables 2.4, 2.5,

Table 2.4 : Intervals of absolute stability and principal error terms of the correctors used with the (0,1) predictor.

Corrector	Stability interval	error constant
(1,1)	$\bar{h} \in (-2,0)$	$T_3 = 1/6$
(1,0)	$\bar{h} \in (-1,0)$	$T_3 = -1/2$
(1,2)	$\bar{h} \in (-2,0)$	$T_3 = 1/6$
(2,2)	$\bar{h} \in (-1.58,0)$	$T_3 = 1/4$
(2,1)	$\bar{h} \in (-1.37,0)$	$T_3 = 1/3$
(2,0)	$\bar{h} \in (-1,0)$	$T_3 = 2/3$
(1,3)	$\bar{h} \in (-2.53,0)$	$T_3 = 1/8$
(2,3)	$\bar{h} \in (-1.78,0)$	$T_3 = 1/5$
(3,3)	$\bar{h} \in (-1.54,0)$	$T_3 = 1/4$
(3,2)	$\bar{h} \in (-1.39,0)$	$T_3 = 3/10$
(3,1)	$\bar{h} \in (-1.22,0)$	$T_3 = 3/8$
(3,0)	$\bar{h} \in (-1.00,0)$	$T_3 = 1/2$
(1,4)	$\bar{h} \in (-2.61,0)$	$T_3 = 1/10$
(2,4)	$\bar{h} \in (-2.02,0)$	$T_3 = 1/6$
(3,4)	$\bar{h} \in (-1.67,0)$	$T_3 = 3/14$
(4,4)	$\bar{h} \in (-1.52,0)$	$T_3 = 1/4$
(4,3)	$\bar{h} \in (-1.41,0)$	$T_3 = 2/7$
(4,2)	$\bar{h} \in (-1.29,0)$	$T_3 = 1/3$
(4,1)	$\bar{h} \in (-1.16,0)$	$T_3 = 4/5$
(4,0)	$\bar{h} \in (-1.00,0)$	$T_3 = 1/2$

Table 2.5: Intervals of absolute stability and principal error terms of the correctors used with the (0,2) predictor.

Corrector	Stability interval	error constant
(1,1)	$\bar{h} \in (-2.0)$	$T_3 = -1/12$
(1,2)	$\bar{h} \in (-2.51,0)$	$T_4 = 1/24$
(2,2)	$\bar{h} \in (-2,0)$	$T_4 = 1/12$
(2,1)	$\bar{h} \in (-1.79,0)$	$T_4 = 1/8$
(2,0)	$\bar{h} \in (-1.61,0)$	$T_3 = 1/6$
(1,3)	$\bar{h} \in (-2.51,0)$	$T_4 = 1/24$
(2,3)	$\bar{h} \in (-2.13,0)$	$T_4 = 1/15$
(3,3)	$\bar{h} \in (-1.94,0)$	$T_4 = 1/12$
(3,2)	$\bar{h} \in (-1.82,0)$	$T_4 = 1/10$
(3,1)	$\bar{h} \in (-1.67,0)$	$T_4 = 1/8$
(3,0)	$\bar{h} \in (-1.50,0)$	$T_4 = 1/8$
(1,4)	$\bar{h} \in (-2.78,0)$	$T_4 = 1/30$
(2,4)	$\bar{h} \in (-2.26,0)$	$T_4 = 1/18$
(3,4)	$\bar{h} \in (-2.05,0)$	$T_4 = 1/14$
(4,4)	$\bar{h} \in (-1.92,0)$	$T_4 = 1/12$
(4,3)	$\bar{h} \in (-1.84,0)$	$T_4 = 2/21$
(4,2)	$\bar{h} \in (-1.74,0)$	$T_4 = 1/9$
(4,1)	$\bar{h} \in (-1.61,0)$	$T_4 = 2/15$
(4,0)	$\bar{h} \in (-1.47,0)$	$T_4 = 1/6$

Table 2.6 : Intervals of absolute stability and principal error terms of the correctors used with the (0,3) predictor.

Corrector	Stability interval	error constant
(1,2)	$\bar{h} \in (-2.38,0)$	$T_4 = -1/72$
(2,2)	$\bar{h} \in (-2.13,0)$	$T_5 = 1/45$
(2,1)	$\bar{h} \in (-2,0)$	$T_4 = 1/72$
(1,3)	$\bar{h} \in (-2.79,0)$	$T_5 = 1/120$
(2,3)	$\bar{h} \in (-2.28,0)$	$T_5 = 1/60$
(3,3)	$\bar{h} \in (-2.09,0)$	$T_5 = 1/48$
(3,2)	$\bar{h} \in (-1.97,0)$	$T_5 = 1/40$
(3,1)	$\bar{h} \in (-1.84,0)$	$T_5 = 7/240$
(3,0)	$\bar{h} \in (-1.59,0)$	$T_4 = 1/8$
(1,4)	$\bar{h} \in (-2.79,0)$	$T_5 = 1/120$
(2,4)	$\bar{h} \in (-2.40,0)$	$T_5 = 1/72$
(3,4)	$\bar{h} \in (-2.19,0)$	$T_5 = 17/1050$
(4,4)	$\bar{h} \in (-2.07,0)$	$T_5 = 1/48$
(4,3)	$\bar{h} \in (-1.99,0)$	$T_5 = 1/42$
(4,2)	$\bar{h} \in (-1.92,0)$	$T_5 = 1/36$
(4,1)	$\bar{h} \in (-1.76,0)$	$T_5 = 1/30$
(4,0)	$\bar{h} \in (-1.59,0)$	$T_5 = 1/12$

Table 2.7 : Intervals of absolute stability and principal error terms of the correctors used with the (0,4) predictor.

Corrector	Stability interval	error constant
(2,2)	$\bar{h} \in (-2.54,0)$	$T_5 = 1/720$
(1,3)	$\bar{h} \in (-2.92,0)$	$T_5 = -1/480$
(2,3)	$\bar{h} \in (-2.65,0)$	$T_6 = 1/248$
(3,3)	$\bar{h} \in (-2.48,0)$	$T_6 = 1/240$
(3,2)	$\bar{h} \in (-2.37,0)$	$T_6 = 7/1440$
(3,1)	$\bar{h} \in (-2.21,0)$	$T_5 = -1/480$
(1,4)	$\bar{h} \in (-3.21,0)$	$T_6 = 1/720$
(2,4)	$\bar{h} \in (-2.76,0)$	$T_6 = 1/360$
(3,4)	$\bar{h} \in (-2.57,0)$	$T_6 = 1/280$
(4,4)	$\bar{h} \in (-2.45,0)$	$T_6 = 1/240$
(4,3)	$\bar{h} \in (-2.37,0)$	$T_6 = 1/210$
(4,2)	$\bar{h} \in (-2.27,0)$	$T_6 = 1/80$
(4,1)	$\bar{h} \in (-2.15,0)$	$T_6 = 1/44$
(4,0)	$\bar{h} \in (-2,0)$	$T_5 = 1/120$

2.6 and 2.7, that as (m,k) correctors (m = 1,...,k), with increasing individual stability intervals, are used with a given predictor, the stability intervals in PECE mode decrease. It can also be deduced that the absolutely stable implicit methods of section 2.2, have inferior intervals of stability to those methods with finite stability intervals when used as correctors with any given (0,k) predictor.

Comparisons with the Milne-Simpson and Adams-Bashforth-Moulton combinations, show that the results of this section can give much bigger stability intervals than multi-step methods with the same order of accuracy. Comparisons with the results of Lawson and Ehle (1970), show that one-step multiderivative methods can also give comparable accuracy to that of one-step methods which use high accuracy Newton-Cotes quadrature formulas as correctors, but can simultaneously give bigger stability intervals. The use of a combination such as (0,4) ; (1,5) for instance, would give the same overall accuracy as the method of Lawson and Ehle (1970), but would have a stability interval bigger than $\bar{h} \in (-3.21,0)$, the stability interval for the (0,4) ; (1,4) combination which has accuracy one power fewer than the method of Lawson and Ehle (1970), the method of Lawson and Ehle (1970) has stability interval $\bar{h} \in (-2.07,0)$.

2.7 Stability Regions

Stability regions, for λ complex, associated with the (0,k*) ; (m,k) combinations in PECE mode will be plotted from equation (2.33), which is

$$\pi_{PECE}(r,\bar{h}) = r - 1 - \sum_{i=1}^{k} \bar{h}^i p_{i,k} + \sum_{j=1}^{m} (-1)^j q_{j,m} \bar{h}^j \left[1 + \sum_{i=1}^{k^*} p_{i,k^*} \bar{h}^i \right],$$

where $\bar{h} = \lambda h$ is complex. The stability region for the (0,k*);(m,k) combination in PECE mode is the region in the complex plan determined by solving the stability equation (2.34), namely

$$\pi_{PECE}(r,\bar{h}) = 0$$

for r. Writing $\bar{h} = u + iv$ $(i = +\sqrt{-1})$ and $r = \cos A + i \sin A$ (so that on the boundary of the region $|r| = 1$), equation (2.34) takes the form

$$(2.36) \qquad f_{k^*,m,k}(u,v) - \cos A + i\{g_{k^*,m,k}(u,v) - \sin A\} = 0$$

where A,u,v are real ; f,g are real valued functions and clearly change for each predictor-corrector combination. The stability region for the $(0,k^*)$; (m,k) combination, is found by solving the non-linear system

$$(2.37) \qquad f_{k^*,m,k}(u,v) - \cos A = 0 \ ,$$

$$g_{k^*,m,k}(u,v) - \sin A = 0 \ ,$$

for each of a series of values of A in the interval $0 \leq A < 360^{\circ}$.

It was found in section 2.5 that, for $k^* = 1,2,3,4$, the $(0,k^*);(k^*,0)$ combination gives the smallest interval of absolute stability when $\lambda < 0$ is real, and that the $(0,k^*);(m,k)$ combination gives the biggest stability interval when $m = 1$ and $k = 4$.

The stability regions, for λ complex, of these eight combinations will now be determined :

1. (a) the $(0,1)$; $(1,0)$ combination :

here, $r = 1 + \bar{h} + \bar{h}^2$,

$f_{1,1,0}(u,v) = 1+u+u^2-v^2$,

$g_{1,1,0}(u,v) = v + 2uv$;

(b) the $(0,1)$; $(1,4)$ combination :

here, $r = 1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{15}\bar{h}^3 + \frac{1}{120}\bar{h}^4$,

$f_{1,1,4}(u,v) = 1+u+\frac{1}{2}(u^2-v^2)+\frac{1}{15}(u^3-3uv^2)+\frac{1}{120}(u^4-6u^2v^2+v^4)$,

$g_{1,1,4}(u,v) = v + uv + \frac{1}{15}(3u^2v - v^3) + \frac{1}{30}(u^3v - uv^3)$.

The stability regions for these two combinations, in the second quarter-plane, are shown in Figure 2.15. The stability region for the Euler predictor-corrector combination in PECE mode is also shown in Figure 2.15. The error constants of all these combinations are of the same order as in

section 2.5.

2. (a) the (0,2) ; (2,0) combination :

here, $r = 1 + \bar{h} + \frac{1}{2}\bar{h}^2 - \frac{1}{4}\bar{h}^4$,

$f_{2,2,0}(u,v) = 1 + u + \frac{1}{2}(u^2 - v^2) - \frac{1}{4}(u^4 - 6u^2v^2 + v^4)$,

$g_{2,2,0}(u,v) = v + uv - u^3v + uv^3$;

(b) the (0,2) ; (1,4) combination :

here, $r = 1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{6}\bar{h}^3 + \frac{1}{120}\bar{h}^4$,

$f_{2,1,4}(u,v) = 1+u+\frac{1}{2}(u^2-v^2) + \frac{1}{6}(u^3-3uv^2) + \frac{1}{120}(u^4-6u^2v^2+v^4)$,

$g_{2,1,4}(u,v) = v + uv + \frac{1}{6}(3u^2v - v^3) + \frac{1}{30}(u^3v - uv^3)$.

The stability regions for these two combinations are shown in Figure 2.16.

3. (a) the (0,3) ; (3,0) combination :

here, $r = 1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{6}\bar{h}^3 + \frac{1}{12}\bar{h}^4 + \frac{1}{36}\bar{h}^6$,

$f_{3,3,0}(u,v) = 1 + u + \frac{1}{2}(u^2-v^2) + \frac{1}{6}(u^3-3uv^2) + \frac{1}{12}(u^4-6u^2v^2+v^4)$

$+ \frac{1}{36}(u^6-15u^4v^2+15u^2v^4-v^6)$,

$g_{3,3,0}(u,v) = v + uv + \frac{1}{6}(3u^2v-v^3) + \frac{1}{3}(u^3v-uv^3)$

$+ \frac{1}{18}(3u^5v-10u^3v^3+3uv^5)$;

(b) the (0,3) ; (1,4) combination :

here, $r = 1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{6}\bar{h}^3 + \frac{1}{24}\bar{h}^4$,

$f_{3,1,4}(u,v) = 1+u+ \frac{1}{2}(u^2-v^2) + \frac{1}{6}(u^3-3uv^2) + \frac{1}{24}(u^4-6u^2v^2+v^4)$,

$g_{3,1,4}(u,v) = v + uv + \frac{1}{6}(3u^2v - v^3) + \frac{1}{6}(u^3v - uv^3)$.

The stability regions for these two combinations are shown in Figure 2.17. The stability region for the fourth order Adams-Bashforth-Moulton combination in PECE mode, which has the same order error constant as the (0,3) ; (1,4) combination in section 2.5, is also shown in Figure 2.17.

4. (a) the (0,4) ; (4,0) combination :

here, $r = 1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{6}\bar{h}^3 + \frac{1}{24}\bar{h}^4 - \frac{1}{72}\bar{h}^6 - \frac{1}{576}\bar{h}^8$,

$$f_{4,4,0}(u,v) = 1 + u + \frac{1}{2}(u^2-v^2) + \frac{1}{6}(u^3-3uv^2) + \frac{1}{24}(u^4-6u^2v^2+v^4)$$
$$- \frac{1}{72}(u^6-15u^4v^2+15u^2v^4-v^6)$$
$$- \frac{1}{576}(u^8-28u^6v^2+70u^4v^4-28u^2v^6+v^8) ,$$

$$g_{4,4,0}(u,v) = v + uv + \frac{1}{6}(3u^2v-v^3) + \frac{1}{6}(u^3v-uv^3)$$
$$- \frac{1}{36}(3u^5v - 10u^3v^3+3uv^5)$$
$$- \frac{1}{72}(u^7v - 7u^5v^3+7u^3v^5-uv^7) .$$

(b) the (0,4) ; (1,4) combination :

here, $r = 1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{6}\bar{h}^3 + \frac{1}{24}\bar{h}^4 + \frac{1}{120}\bar{h}^5$,

$$f_{4,1,4}(u,v) = 1 + u + \frac{1}{2}(u^2-v^2) + \frac{1}{6}(u^3-3uv^2)$$
$$+ \frac{1}{24}(u^4-6u^2v^2+v^4) + \frac{1}{120}(u^5-10u^3v^2+5uv^4) ,$$

$$g_{4,1,4}(u,v) = v + uv + \frac{1}{6}(3u^2v - v^3) + \frac{1}{6}(u^3v - uv^3)$$
$$+ \frac{1}{120}(5u^4v - 10u^2v^3+v^5) .$$

The stability regions for these two combinations are shown in Figure 2.18. The stability region of the fourth order Adams-Bashforth-Moulton combination, which has the same order error constant in PECE mode as the (0,4) ; (4,0) combination, is also shown in Figure 2.18.

It is noted that the (0,3);(1,4) and (0,4);(1,4) combinations have the same stability regions as the fourth and fifth order Taylor series methods, respectively. The axes of all four figures are drawn to the same scale. The stability regions are, of course, applicable to the system of linear differential equations of the form

(2.38) $$\underline{y}'(x) = A\underline{y}(x) \; ; \quad \underline{y}(0) = \underline{y}_0 \quad ,$$

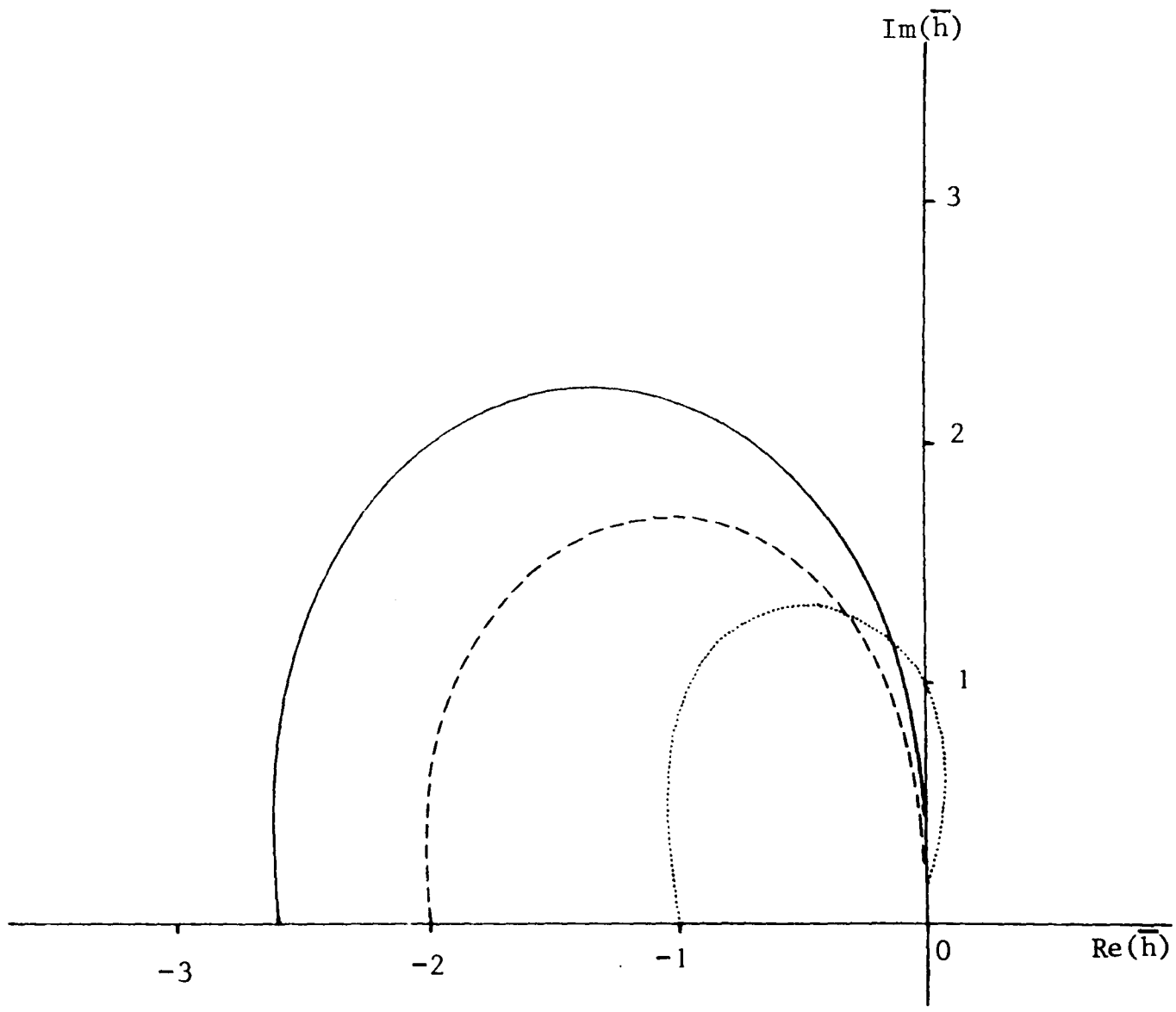


Figure 2.15 ——— (0,1);(1,4) combination

..... (0,1);(1,0) combination

- - - - - - - (0,1);(1,1) combination (Euler-modified Euler)

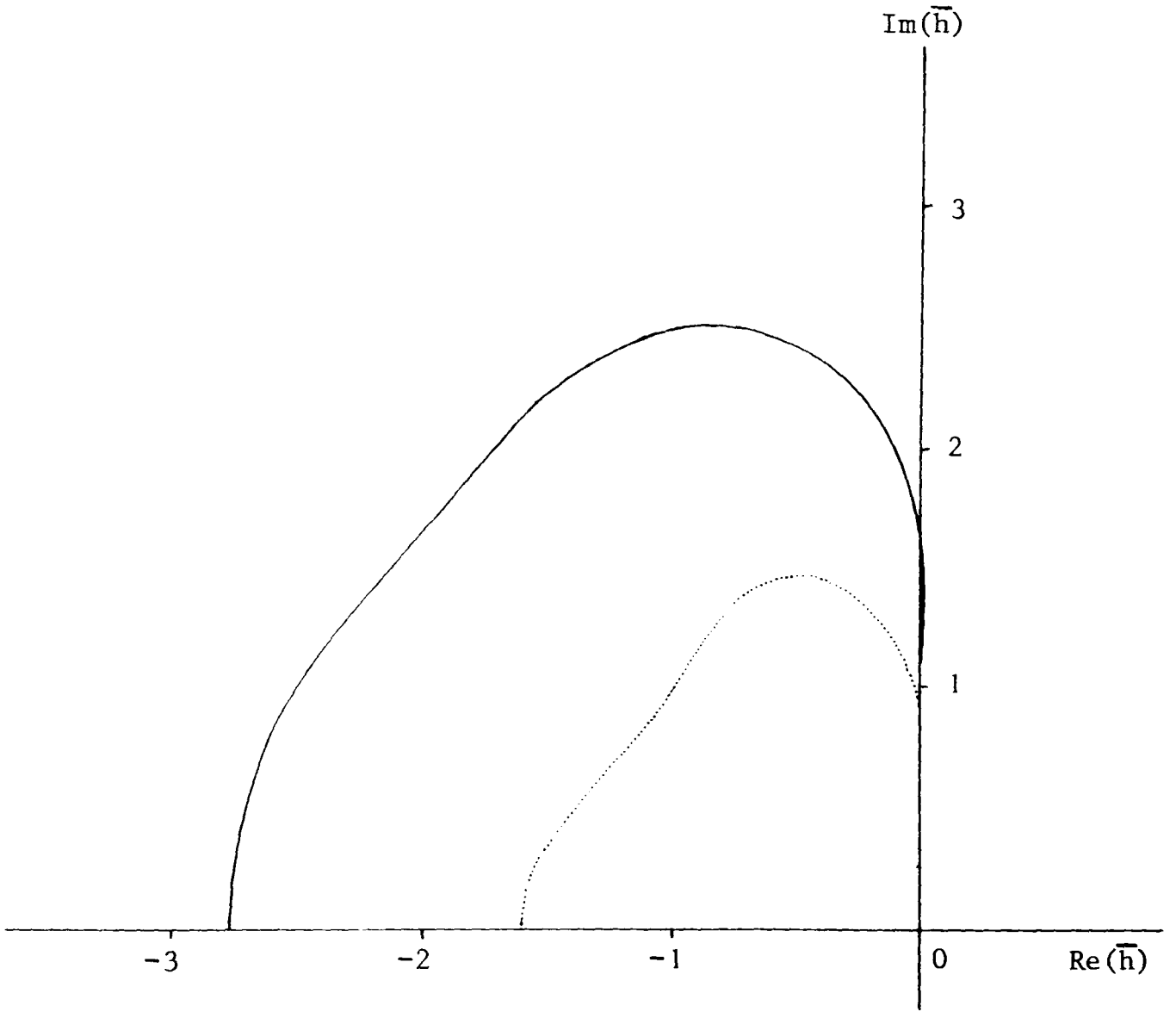


Figure 2.16 _____ (0,2);(1,4) combination

..... (0,2);(2,0) combination

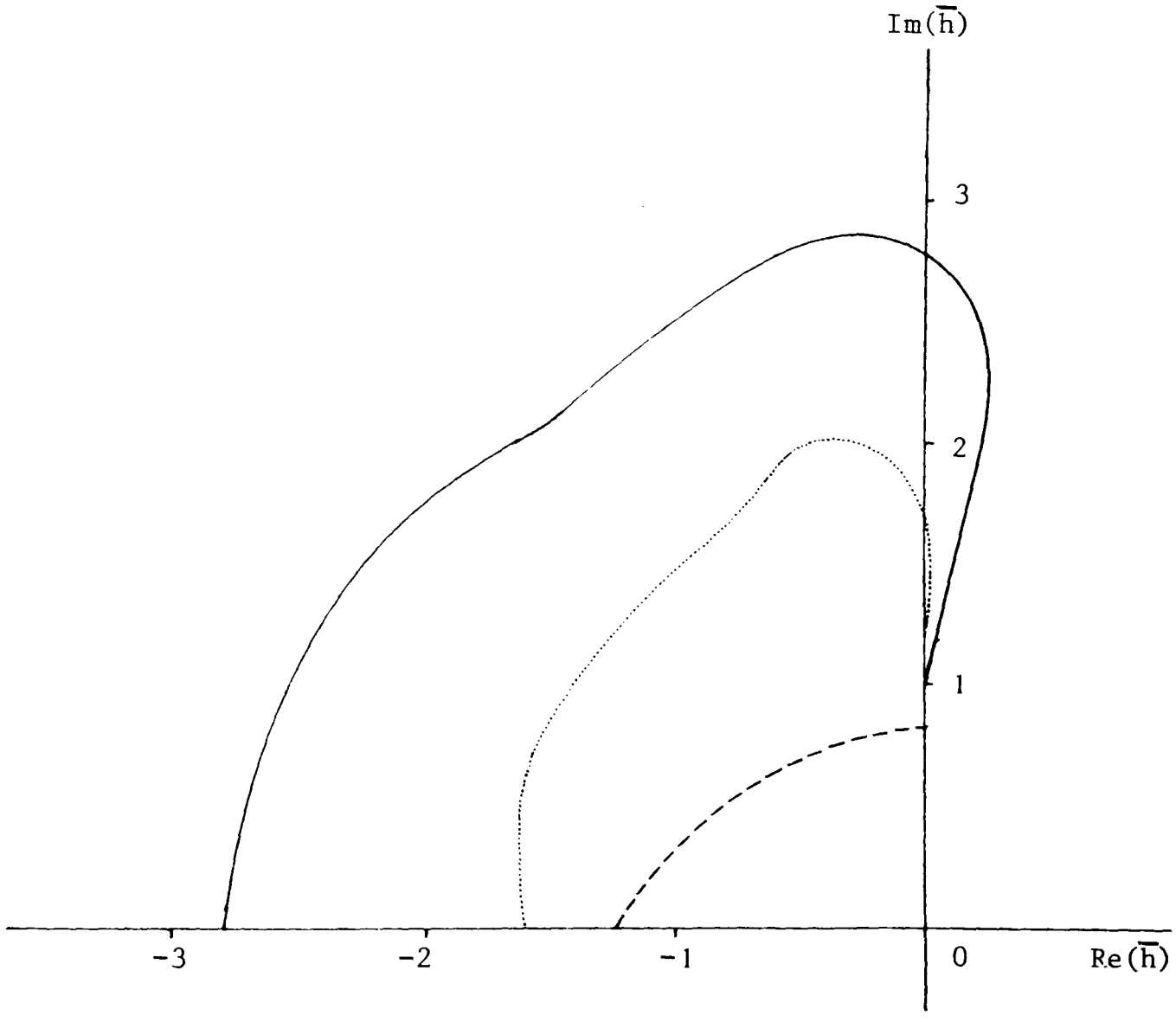


Figure 2.17 _____(0,3);(1,4) combination

.....(0,3);(3,0) combination

_ _ _ _ _ _ _fourth order Adams-Bashforth-Moulton

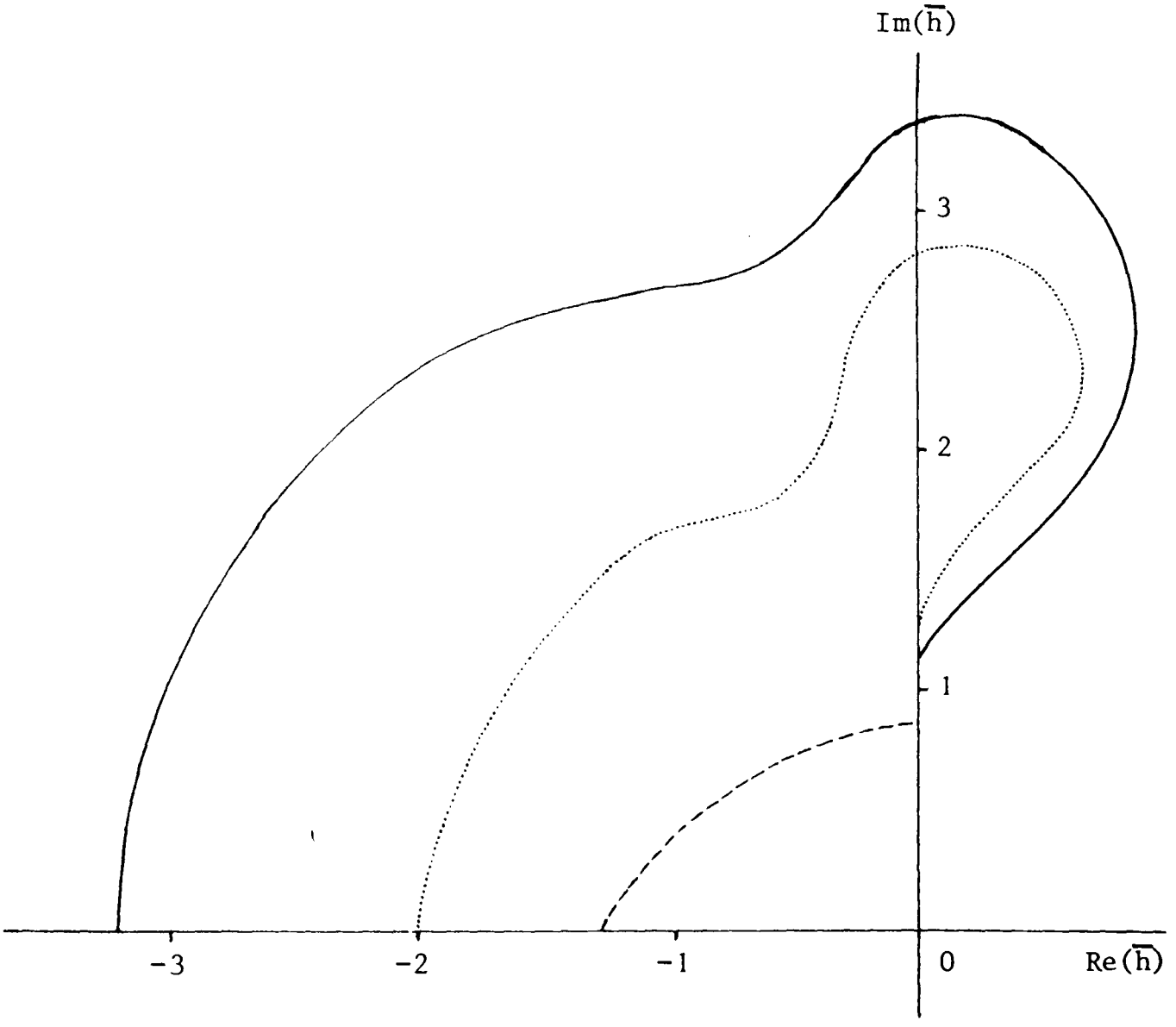


Figure 2.18 _____ (0,4);(1,4) combination

..... (0,4);(4,0) combination

_____ fourth order Adams-Bashforth-Moulton

where A is square matrix of order N; the real part of the eigenvalues $\lambda_j (j = 1,2,...N)$ of A must be non-positive. For non-linear systems the eigenvalues $\lambda_j (j = 1,2,...N)$ are those of the Jacobian matrix $\partial \underline{f}/\partial \underline{y}$; these eigenvalues are calculated at each point x_n.

2.8 Numerical examples

The $(0,k^*);(k^*,0)$ and $(0,k^*);(1,4)$ combinations $(k = 1,2,3,4)$ are tested on two problems, the first a system of the form (2.1) with complex eigenvalues, the second a system of the form (2.38) with negative real eigenvalues but a large stiffness ratio.

Problem 2.1

(Lambert (1973,p.229))

$$y_1' = 21y_1 + 19y_2 - 20y_3,$$

$$y_2' = 19y_1 - 21y_2 + 20y_3,$$

$$y_3' = 40y_1 - 40y_2 - 40y_3,$$

with initial conditions $\underline{y}(0) = (1,0,-1)^T$. The matrix of coefficients has eigenvalues $\lambda_1 = -2$, $\lambda_2 = -40 + 40i$, $\lambda_3 = -40 - 40i$ giving a moderate stiffness ratio of 20. The maximum steplength for each method is found by drawing the line $Im(\bar{h}) = -Re(\bar{h})$ in Figures 2.15-2.18 and estimating the point of intersection with the boundary of the stability region. The maximum steplengths for each of the predictor-corrector combinations follows in an abvious manner and are given in Table 2.8, truncated to three decimal places, together with the maximum steplengths which may be used with the Euler-modified Euler and Adams-Bashforth-Moulton combinations.

It was noted by Lambert (1973,p.229), that the theoretical solution of the problem, given by

$$y_1 = \frac{1}{2} e^{-2x} + \frac{1}{2} e^{-40x} (\cos 40x + \sin 40x),$$

$$y_2 = \frac{1}{2} e^{-2x} - \frac{1}{2} e^{-40x} (\cos 40x + \sin 40x),$$

$$y^3 = -e^{-40x} (\cos 40x - \sin 40x),$$

Table 2.8: Maximum steplengths which may be used each predictor-corrector combination for Problems 2.1 and 2.2 .

Combination	Maximum steplength	
	Problem 1	Problem 2
(0,1) ; (1,0)	0.025	0.00098
(0,1) ; (1,4)	0.050	0.00257
(0,1) ; (1,1) (Euler)	0.037	0.00197
(0,2) ; (2,0)	0.025	0.00159
(0,2) ; (1,4)	0.046	0.00274
(0,3) ; (3,0)	0.031	0.00157
(0,3) ; (1,4)	0.047	0.00275
(0,4) ; (4,0)	0.035	0.00197
(0,4) ; (1,4)	0.055	0.00317
A - B - M	0.016	0.00123

behaves as $\underline{y} = (\frac{1}{2}e^{-2x}, \frac{1}{2}e^{-2x}, 0)^T$ for $x > 0.1$ (approximately). The solution vector was therefore computed only for x in the interval $0 \leq x \leq 0.09$ using the step lengths $h = 0.01,\ 0.015,\ 0.03$.

The numerical results obtained were in keeping with the theory, and are given for $x = 0.09$ in Table 2.9. The results for the $(0,1);(1,0)$, $(0,2);(1.2)$ and $(0,3);(3,0)$ combinations, for which $h = 0.03$ exceeds the maximum steplength, display evidence of instability. For all other combinations, using all three values of h, the error was found to decay with increasing x.

Problem 2.2

$$y_1' = 0.01 - (0.01 + y_1 + y_2)(y_1^2 + 1001y_1 + 1001),$$

$$y_2' = 0.01 - (0.01 + y_1 + y_2)(1 + y_2^2),$$

with initial conditions $\underline{y}(0) = (0,0)^T$. This problem arises in reactor kinetics and has been discussed by Liniger and Willoughby (1967), Lambert (1973) and Cash (1980). The Jacobian matrix $\partial \underline{f}/\partial \underline{y}$ has eigenvalues -1012 and -0.01 at $x = 0$; it thus has an initial stiffness ratio $\simeq 10^5$ and may be classed initially as being very stiff. The maximum steplengths which may be used with the multiderivative predictor-corrector combinations are found by dividing the value of $Re(\bar{h})$, where the curves bounding the stability regions in Figures 2.15-2.18 cut the real axis, by -1012. These maximum values, truncated to five decimal places, are given in Table 2.8.

One of the main difficulties in the application of multiderivative methods to systems of non-linear equations, is in the calculation of the higher order derivatives. These were easily obtained for the present problem and were evaluated at each step of the following computations. The theoretical solution of the problem is not known and, following Cash (1980), was found approximately using the fourth order Runge-Kutta process.

The numerical experiments of Cash (1980,p.245) were repeated using

Table 2.9: Errors e_1, e_2, e_3 in y_1, y_2, y_3 at $x = 0.09$ for Problem 2.1 using the multiderivative predictor-corrector combinations with $h = 0.01, 0.015, 0.03$.

Combination		Errors in y_1, y_2, y_3		
		$h = 0.01$	$h = 0.015$	$h = 0.03$
(0,1) ; (1,0)	e_1	-0.262(-1)	-0.164(-1)	-0.313(+1)
	e_2	0.246(-1)	0.140(-1)	0.313(+1)
	e_3	0.630(-2)	0.167(-1)	-0.284(+1)
(0,1) ; (1,4)	e_1	-0.375(-2)	-0.855(-2)	-0.183(-1)
	e_2	0.375(-2)	0.853(-2)	0.182(-1)
	e_3	-0.104(-2)	0.716(-3)	0.124(-1)
(0,2) ; (2,0)	e_1	-0.275(-2)	-0.130(-1)	-0.189(+1)
	e_2	0.274(-2)	0.129(-1)	0.189(+1)
	e_3	-0.103(-1)	-0.412(-1)	0.878(+1)
(0,2) ; (1,4)	e_1	0.656(-3)	0.229(-2)	0.361(-2)
	e_2	-0.656(-3)	-0.229(-2)	-0.361(-2)
	e_3	-0.968(-3)	-0.486(-2)	0.111(-1)
(0,3) ; (3,0)	e_1	-0.686(-3)	-0.147(-2)	-0.125
	e_2	0.686(-3)	0.147(-2)	0.125
	e_3	0.199(-2)	0.104(-1)	0.163(+1)
(0,3) ; (1,4)	e_1	-0.145(-4)	0.237(-4)	0.534(-2)
	e_2	0.145(-4)	-0.237(-4)	-0.534(-2)
	e_3	0.232(-3)	0.139(-2)	0.391(-1)
(0,4) ; (4,0)	e_1	-0.960(-4)	-0.883(-3)	-0.180(-1)
	e_2	0.960(-4)	0.883(-3)	0.180(-1)
	e_3	0.623(-4)	0.245(-3)	0.116(-1)
(0,4) ; (1,4)	e_1	-0.695(-5)	-0.753(-4)	-0.491(-2)
	e_2	0.694(-5)	0.753(-4)	0.491(-2)
	e_3	-0.174(-4)	-0.133(-3)	-0.146(-2)

Theoretical solution is $\underset{\sim}{y}(0.09) \simeq (0.339, 0.436, 0.012)^T$

Table 2.10: Errors in y_1, y_2 for Problem2.2after ten steps of h = 0.001, 0.0001, 0.00001, 0.000001 using the (0,k*) ; (1,4) predictor-corrector combinations (k* = 1,2,3,4)

h	Theoretical solution (y_1,y_2)	Errors in y_1, y_2				
		(0,1) ; (1,4)	(0,2) ; (1,4)	(0,3) ; (1,4)	(0,4) ; (1,4)	Cash EBD
0.001	-0.1006914044(-1) 0.8978912350(-4)	0.241(-5) 0.135(-7)	0.246(-6) 0.728(-9)	0.101(-6) 0.823(-9)	0.149(-7) 0.572(-9)	0.815(-6) 0.628(-8)
0.0001	-0.6306050198(-2) 0.3670275606(-5)	0.394(-5) 0.392(-8)	0.135(-6) 0.132(-9)	0.455(-8) 0.353(-11)	0.650(-10) 0.662(-13)	0.835(-6) 0.819(-9)
0.00001	-0.9511426272(-3) 0.4835591013(-7)	0.929(-8) 0.920(-11)	0.318(-10) 0.326(-13)	0.104(-13) 0.800(-16)	0.141(-13) 0.379(-17)	0.231(-9) 0.222(-12)
0.000001	-0.9949622896(-4) 0.4983176581(-9)	0.101(-10) 0.100(-13)	0.348(-14) 0.345(-17)	0.120(-18) 0.638(-24)	0.105(-17) 0.921(-21)	0.300(-13) 0.246(-16)

the eight multiderivative predictor-corrector combinations discussed in section 2.5. The steplength h was given the values 0.001, 0.0001, 0.00001, 0.000001 and the solution was computed for ten steps in each case. Cash (1980) also used the value 0.01, but this value was greater than the maximum steplength for all eight predictor-corrector methods and was not used.

The numerical results obtained for Problem 2.2 using the $(0,k^*);(1,4)$ combinations $(k^* = 1,2,3,4)$ are summarized in Table 2.10. Comparison with the numerical results obtained using the extended backward differentiation formula of Cash (1980), show that the multiderivative methods developed in section 2.2 give smaller errors in PECE mode. For Problem 2.2 also, the numerical results were found to be in keeping with the theory.

Overall, the results obtained for the two problems, indicate strongly that multiderivative methods in PECE mode give very good numerical results for linear systems where the coefficient matrix has complex eigenvalues and for stiff systems of non-linear ordinary differential equations. They can readily be used to solve problems for which the higher derivatives can be obtained, or estimated, with reasonable ease.

2.9 Conclusions

A family of linear, one-step, multiderivative methods, based on Padé approximants to the exponential function, has been developed in this chapter. The family is seen to contain a number of well known methods including the Euler predictor, the Euler corrector (the trapezoidal rule) and a formula due to Milne (1949). It has been verified that, using comparable steplengths, much higher accuracy can be obtained using the family of one-step multiderivative methods than can be achieved using linear one-step methods. The family of multiderivative methods is therefore appropriate for use in problems which allow higher derivatives

to be found explicitly and which require high accuracy. Intervals of absolute stability have been calculated and it is seen that those members of the family which are fully implicit, in the sense that the highest derivative must be evaluated at the advanced point, are absolutely stable.

The family of multiderivative methods has been extrapolated to achieve higher accuracy and intervals of absolute stability are calculated for the extrapolation formulas. It is seen that, whilst extrapolation increases accuracy, stability intervals are sometimes shortened as a consequence; the most notable example of this is the trapezoidal rule.

Finally, the family of one-step multiderivative methods has been used in appropriate predictor-corrector pairs. Error constants, stability intervals and stability regions have been calculated for PECE mode. As with linear multistep (single derivative) methods used in PECE mode, the stability intervals are seen to be somewhat low. It is clear from Tables 2.4, 2.5, 2.6 and 2.7 however, that it is possible to achieve a bigger stability interval, with comparable accuracy, using one-step multi-derivative combinations in PECE mode than with some well known multi-step combinations, notably the Milne-Simpson and Adams-Bashforth-Moulton methods, or with one-step methods using high accuracy Newton-Cotes quadrature formulas as correctors.

CHAPTER 3

SECOND ORDER PARABOLIC EQUATIONS

3.1 Introduction

In recent papers, Lawson and Swayne (1976), Lawson and Morris (1978) and Gourlay and Morris (1980), attention has been devoted to the development of L_0-stable methods for the numerical solution of second order parabolic partial differential equations for which A_0-stable methods such as the Crank-Nicolson method, are unsatisfactory when time discretization is used with time steps which are too large relative to the space discretization, see for example, Smith et al (1973) and Wood and Lewis (1975).

Lawson and Morris (1978) developed a second order L_0-stable method as an extrapolation of a first order backward difference method in one and two space dimensions. This idea was developed further for one space variable by Gourlay and Morris (1980) who achieved third and fourth order accuracy in time by a novel multistage process. The second order method of Lawson and Morris (1978), was adapted and used in a practical problem involving a non-linear parabolic equation by Twizell and Smith (1981, 1982).

The extrapolation procedure of Lawson and Morris (1978) involved computing the solution of the parabolic equation at time $t + 2\ell$, in terms of the solution at time t, using a first order method with time step ℓ: second order accuracy was thus achieved. Gourlay and Morris (1980) extended the principal by computing the solution at time $t + 3\ell$, in terms of the solution at time t, using a time step ℓ, and thus achieved third order accuracy in time. These authors then went further, and achieved fourth order accuracy by computing the solution at time $t + 4\ell$ in terms of the solution at time t.

The multistage methods which evolved in this way involved a "spread" in time. In this chapter a family of methods will be developed which involves a similar "spread" in space, in that an increased number of points at each time level are used in the resulting finite difference schemes.

This concept of using a greater number of points at each time level was used by Twizell (1979) for second order hyperbolic equations and by Khaliq and Twizell (1982) for first order hyperbolic equations; the concept is discussed for second order parabolic equations in the text by Mitchell and Griffiths (1980).

The methods developed are applications of the methods for a system of first order ordinary differential equations discussed in Chapter 2. Following Lawson and Morris (1978) and Gourlay and Morris (1980), the space derivatives will be approximated by the usual second order central difference replacement. The principal part of the local truncation error of each finite difference scheme will, therefore, include the same component proportional to ℓh^2 , where h is the space step, encountered, though not stated explicitly, in Lawson and Morris (1978) and Gourlay and Morris (1980). This component notwithstanding, it was shown in Lawson and Morris (1978) and Gourlay and Morris (1980), that extrapolation in time leads to a worthwhile improvement in accuracy; this being demonstrated clearly by numerical experiments reported in those papers.

The family of multiderivative methods on which the finite difference schemes are based, uses Padé approximants to the matrix exponential function. Lawson and Morris (1978) used the (1,0) Padé approximant; in this chapter, four higher order Padé approximants are used to achieve higher order accuracy in time. The resulting finite difference schemes are implicit in nature and each requires one quindiagonal or sevendiagonal solver to determine the solution. This compares well with the multistage methods in Gourlay and Morris (1980) where, for problems with one space variable, five applications of a tridiagonal solver are needed to achieve third order accuracy in time and at least seven applications of a tridiagonal solver to achieve fourth order accuracy in time.

The methods developed in this Chapter will be tested on the model problems used in Lawson and Morris (1978) and Gourlay and Morris (1980).

For one space variable, the second order method will be seen to give results comparable overall to the best third order multistage methods in Gourlay and Morris (1980), and third order methods developed, to give results comparable overall to the best fourth order multistage method.

3.2 One-space dimension

Consider the constant coefficient heat equation in one space variable

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad ; \quad 0 < x < X \,, \quad t > 0 \tag{3.1}$$

with initial conditions

$$u(x,0) = g(x) \quad ; \quad 0 \le x \le X \tag{3.2}$$

and boundary conditions

$$u(0,t) = u(X,t) = 0 \quad ; \quad t > 0 \tag{3.3}$$

In (3.2), $g(x)$ is a given continuous function of x; it is not specified that $g(0) = 0$ or $g(X) = 0$, so that discontinuities between initial conditions and boundary conditions may occur.

The interval $0 \le x \le X$ is divided into $N+1$ subintervals each of width h so that $(N+1)h = X$, and the time variable t is discretized in steps of length ℓ. The open region $R = [0 < x < X] \times [t > 0]$ and its boundary ∂R have thus been covered by a rectangular mesh, the mesh points having co-ordinates $(mh, n\ell)$ with $m = 0,1,\ldots,N+1$ and $n = 0,1,2,\ldots$. The notation $u_m^n \equiv u(mh,n\ell)$ will be used to denote the solution of (3.1) while U_m^n will be used to denote the theoretical solution of an approximating finite difference scheme.

The space derivative in (3.1) is now replaced by

$$\frac{\partial^2 u}{\partial x^2} = \{u(x-h,t)-2u(x,t)+u(x+h,t)\}/h^2+0(h^2) \tag{3.4}$$

and (3.1) with (3.4) is applied to all N interior mesh points at time $t = n\ell$ $(n = 0,1,\ldots)$. This produces a system of ordinary differential equations of the form

(3.5) $$\frac{d\underline{U}}{dt} = A\,\underline{U}$$

where $\underline{U}(t) = \underline{U}(n\ell) = \underline{U}^n = (U_1^n, U_2^n, \ldots, U_N^n)^T$, T denoting transpose

In (3.5) the matrix A is given by

(3.6) $$A = h^{-2}\begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ 0 & & & 1 & -2 \end{bmatrix}$$

and has eigenvalues $\lambda_s = -4h^{-2}\sin^2[s\pi/2(N+1)]$ for $s = 1,2,\ldots,N$.

A practical difficulty with (3.5) is that the system is stiff because an acceptably small component of the local truncation error relating to the space discretization requires a large value of N; this, in turn, leads to a large range of eigenvalues of A and hence to a large stiffness ratio α given by

$$\alpha = \sin^2(N\pi/2(N+1))/\sin^2(\pi/2(N+1))$$

or $$\alpha = 4(N+1)^2/\pi^2 \quad \text{for large } N .$$

Solving (3.5) with initial vector $\underline{U}(0) = \underline{g}$ from (3.2), gives

$$\underline{U}(t) = \exp(tA)\underline{g}$$

which satisfies the recurrence relation

(3.7) $$\underline{U}(t+\ell) = \exp(\ell A)\underline{U}(t) \quad ; \quad t = 0, \ell, 2\ell, \ldots .$$

To obtain a numerical solution from (3.7), Padé approximants will be used. The $(0,1)$ Padé approximant gives a commonly used four point explicit scheme whose interval of absolute stability given in Table 2.1, is $(-2,0)$. It follows that for absolute stability, the choice of the (positive) steplength must satisfy

$-2 \le -4\ell/h^2 < 0$; that is the mesh ratio $r = \ell/h^2$ must satisfy

$$0 < r \le \tfrac{1}{2} .$$

The $(1,0)$ Padé approximant gives the L_0-stable fully implicit scheme developed and extrapolated by Lawson and Morris (1978); the $(1,1)$ Padé

approximant gives the A_0-stable Crank-Nicolson method which was also analysed in Lawson and Morris (1978).

Employing the (m,k) Padé approximants, for $m < k$, to the exponential function in (3.7), will yield explicit or semi-explicit methods, whose intervals of stability may be calculated from Table 2.1. In the following sections four higher order Padé approximants with the degree of the numerator less than or equal to the degree of denominator, will be used in (3.7) and the resulting algorithms analysed.

3.3 A second order method and its extrapolation

Using the (2,0) Padé approximant to $\exp(\ell A)$ in (3.7) gives

$$\underline{U}(t+\ell) = (I-\ell A+\tfrac{1}{2}\ell^2A^2)^{-1}\,\underline{U}(t) \tag{3.8}$$

suggesting the fully implicit scheme

$$(I-\ell A+\tfrac{1}{2}\ell^2A^2)\underline{U}(t+\ell) = \underline{U}(t)\ . \tag{3.9}$$

Following Gourlay and Morris (1980), a stability analysis verifies that (3.9) is L_0-stable, and using Taylor's theorem it is found that the principal part of the local truncation error at the mesh point $(mh,n\ell)$ for $m = 2,\ldots,N-1$ and $n = 0,1,2,\ldots$ is

$$(-\frac{1}{12}\ell h^2\,\frac{\partial^4 u}{\partial x^4} + \frac{1}{6}\ell^3\,\frac{\partial^3 u}{\partial t^3})^n_m\ , \tag{3.10}$$

though this accuracy is not attained at points adjacent to the boundaries. This phenomenon is seen to be present in all the subdiagonal Padé approximants for time dependent problems except the (1,0) Padé approximant. In a paper on hyperbolic equations, Oliger (1974,p.20) showed that using lower approximants near the boundaries, does not affect the stability or convergence properties of the scheme as a whole, and the numerical evidence to be reported in section 3.6 suggests that this is true for second order parabolic equations also.

In (3.10) the component $-\frac{1}{12}\ell h^2\partial^4 u/\partial x^4$ is due to the space discretization and the use of (3.4) in (3.1); this term will appear in the local truncation error of every finite difference scheme arising from the

use of Padé approximants to $\exp(\ell A)$ in (3.7). The term $\frac{1}{6}\ell^3\partial^3 u/\partial t^3$ in (3.10) relates to the use of the (2,0) Padé approximant in (3.7).

The principal part of the local truncation error may be written down for every finite difference scheme yielded by (3.7) in the form

$$(-\frac{1}{12}\ell h^2 \frac{\partial^4 u}{\partial x^4} + C_q \ell^q \frac{\partial^q u}{\partial t^q})_m^n \tag{3.11}$$

where the constants C_q are given in Table 2.1.

Applying (3.9) to the mesh point $(mh,n\ell)$ at time $t = n\ell$ leads to a linear system, the unknowns of which are the components of the N-vector $\underline{U}(t+\ell)$. This linear system is of the form

$$E\,\underline{U}(t+\ell) = \underline{\Phi}^n , \tag{3.12}$$

where E is a constant matrix of order N; it has the form

$$E = \begin{bmatrix} e_4 & e_2 & e_3 & & & & \\ e_2 & e_1 & e_2 & e_3 & & 0 & \\ e_3 & e_2 & e_1 & e_2 & e_3 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & e_3 & e_2 & e_1 & e_2 & e_3 \\ & 0 & & e_3 & e_2 & e_1 & e_2 \\ & & & & e_3 & e_2 & e_4 \end{bmatrix}$$

where

$$e_1 = 1+2r+3r^2 , \quad e_2 = -r-r^2 , \quad e_3 = \tfrac{1}{2}r^2 , \quad e_4 = 1+2r+\tfrac{5}{2}r^2 .$$

The vector $\underline{\Phi}^n = (\phi_1,\phi_2,\ldots\phi_N)^T$ is seen from (3.9) to be $\underline{U}^n$.

The solution $\underline{U}(t+\ell)$ of (3.9) is found from (3.11) by applying a quindiagonal solver. Gerschgorin's theorem fails to predict the positive definiteness of E. However, E is positive definite with spectral condition number $O(h^{-4})$, see, for example, Fried (1979,pp.19-21).

An alternative approach for finding $\underline{U}(t+\ell)$ is to write

$E = I - \ell A + \frac{1}{2}\ell^2 A^2$ in its complex factor form, namely

$$E = \frac{1}{2}\{(1+i)I - \ell A\}\{(1-i)I - \ell A)\}\ , \quad i = +\sqrt{-1}\ .$$

This suggests the complex splitting

$$\begin{aligned} &\{(1-i)I - \ell A\}\,\underline{U}^{*} = \underline{U}(t)\ , \\ &\tfrac{1}{2}\{(1+i) - \ell A\}\,\underline{U}(t+\ell) = \underline{U}^{*}\ . \end{aligned} \tag{3.13}$$

The solution $\underline{U}(t+\ell)$ is obtained from (3.13) by the application of two tridiagonal solvers using complex arithmetic. This is less efficient, in that it uses more CPU time, than using one quindiagonal solver with real arithmetic and (3.11) is therefore to be preferred to (3.13).

It has been noted already that (3.9) is L_0-stable; it may therefore be extrapolated to improve the accuracy in time.

First of all, $\underline{U}^{(1)} \equiv \underline{U}^{(1)}(t+2\ell)$ is computed by applying (3.7) over two single time steps with $\exp(\ell A)$ replaced by its (2,0) Padé approximant; secondly, $\underline{U}^{(2)} \equiv \underline{U}^{(2)}(t+2\ell)$ is computed by applying (3.7) over a double time step. In general, the extrapolated value $\underline{U}^{(E)} \equiv \underline{U}^{(E)}(t+2\ell)$ is determined from the formula

$$\underline{U}^{(E)} = \alpha\,\underline{U}^{(1)} - (\alpha-1)\,\underline{U}^{(2)} + O(\ell^{m+k+2});\ m \neq k\ , \tag{3.14}$$

where

$$\alpha = 2^{m+k}/(2^{m+k}-1)\ , \tag{3.15}$$

so that for the (2,0) Padé approximant, $\alpha = \frac{4}{3}$.

The principal part of the local truncation error of the extrapolated form of each finite difference method arising from using the (m,k) Padé approximant to $\exp(\ell A)$ in (3.7), will be of the form

$$\left(-\frac{1}{12}\ell h^2 \frac{\partial^4 u}{\partial x^4} + E_p \ell^p \frac{\partial^p u}{\partial t^p}\right)_m^n \quad ; \quad m = 2,\ldots,N-1 \tag{3.16}$$

where $p = m+k+2$. The constants E_p (Twizell and Khaliq (1981)), are contained in Table 2.3; for the (2,0) Padé approximant, $E_4 = -\frac{1}{3}$.

The amplification symbol of the extrapolated form of the method arising from using the (m,k) Padé approximant in (3.7), is

(3.17) $$S_{m,k}(z) = \alpha \{P_k(-z)/Q_m(-z)\}^2 - (\alpha-1)\{P_k(-2z)/Q_m(-2z)\}$$

where $z = -\ell\lambda$ and λ is an eigenvalue of A. Clearly, therefore,

(3.18) $$S_{2,0}(z) = \frac{4}{3}(1+z+\frac{1}{2}z^2)^{-2} - \frac{1}{3}(1+2z+2z^2)^{-1} .$$

Simplifying this expression for $S_{2,0}(z)$ gives

$$S_{2,0}(z) = \frac{1+2z+2z^2-\frac{1}{3}z^3-\frac{1}{12}z^4}{1+4z+8z^2+9z^3+\frac{25}{4}z^4+\frac{5}{2}z^5+\frac{1}{2}z^6}$$

which satisfies $|S_{2,0}(z)| \le 1$ and $\lim_{z\to\infty} S_{2,0}(z) = 0$, and thus verifies that the third order method, as an extrapolation of the second order method (3.9), is L_0-stable. The amplification symbols for the (2,0) method and its extrapolated form are plotted in Fig 2.3. It can be seen from Fig 2.3 that the asymptotic behaviour of the second order L_0-stable method (3.9) (and its extrapolated form), produces a growth factor which tends to zero monotonically, implying that no oscillations could appear and the method will behave smoothly, like the theoretical solution. This shows that the finite difference method based on the (2,0) Padé approximant is suitable for use with problems having discontinuities between initial conditions and boundary conditions.

3.4 Two third-order methods and their extrapolations

The extrapolation of (3.9) produces a scheme which is third order accurate in time. The same order of accuracy in time can be achieved when the (2,1) Padé approximant is used in (3.7) giving

(3.19) $$(I-\frac{2}{3}\ell A+\frac{1}{6}\ell^2A^2)\,\underline{U}(t+\ell) = (I+\frac{1}{3}\ell A)\,\underline{U}(t) .$$

Applying (3.19) to the mesh points $(mh,n\ell)$, with $m = 1,2,\ldots,N$, at time $t = n\ell$ $(n = 0,1,2,\ldots)$ leads to a linear system of the form (3.12). The elements of the matrix E in (3.12) are now given by

$$e_1 = 1+\frac{4}{3}r+r^2, \quad e_2 = -\frac{2}{3}r-\frac{2}{3}r^2, \quad e_3 = \frac{1}{6}r^2, \quad e_4 = 1+\frac{4}{3}r+\frac{5}{6}r^2$$

and the elements of the vector $\underline{\phi}^n$ by

$$\phi_1^n = (1-\tfrac{2}{3}r)\ U_1^n + \tfrac{1}{3}r\ U_2^n$$

$$\phi_m^n = \tfrac{1}{3}r\ U_{m-1}^n + (1-\tfrac{2}{3}r)\ U_m^n + \tfrac{1}{3}r\ U_{m+1}^n \quad ; \ m = 2,\ldots,N-1 \ ,$$

$$\phi_N^n = \tfrac{1}{3}r\ U_{N-1}^n + (1-\tfrac{2}{3}r)\ U_N^n \quad .$$

The solution $\underline{U}(t+\ell)$ of (3.19) is determined by applying a quin-diagonal solver. In view of the discussion of (3.13), it is not worthwhile to consider a complex splitting of (3.19). The principal part of the local truncation error of (3.19) at the interior mesh points $(mh,n\ell)$, $(m = 2,\ldots,N-1;\ \ n = 0,1,2,\ldots)$ is given by (3.11) with $q = 4$, and, from Table 2.1, $C_4 = 1/72$. A stability analysis shows that (3.19) is L_0-stable and the amplification symbol is shown in Fig 2.4. It is seen that the function $R_{2,1}(z)$ is negative for $z > 3$ and does in fact tend to zero more slowly than the extrapolated form of the method based on the (1,0) Padé approximant (see Lawson and Morris (1978)).

In view of its L_0-stability property, the method may be extrapolated to improve accuracy in time. The extrapolation formula (3.14) is used, and (3.15) yields $\alpha = \frac{8}{7}$. The principal part of the local truncation error of the extrapolated form of the method is given by (3.16) with $p = 5$; the value of E_5 is found from Table 2.3 to have the value $-8/945$.

The amplification symbol of the extrapolated form of (3.19) is

$$S_{2,1}(z) = \frac{8}{7}\left[\frac{1-\frac{1}{3}z}{1+\frac{2}{3}z+\frac{1}{6}z^2}\right]^2 - \frac{1}{7}\left[\frac{1-\frac{2}{3}z}{1+\frac{4}{3}z+\frac{2}{3}z^2}\right] \tag{3.20}$$

and it follows that the extrapolated form of this third order method is L_0-stable also. The amplification symbol $S_{2,1}(z)$ is also shown in Fig 2.4.

The second third-order method to be discussed is that based on the (3,0) Padé approximant to $\exp(\ell A)$ in (3.7). This approximant gives

$$\underline{U}(t+\ell) = (I-\ell A+\tfrac{1}{2}\ell^2A^2-\tfrac{1}{6}\ell^3A^3)^{-1}\underline{U}(t) \tag{3.21}$$

suggesting the fully implicit algorithm

$$(3.22) \qquad (I-\ell A+\frac{1}{2}\ell^2A^2-\frac{1}{6}\ell^3A^3)\ \underline{U}(t+\ell) = \underline{U}(t) \quad .$$

Writing (3.22) in the form

$$(3.23) \qquad F\ \underline{U}(t+\ell) = \underline{U}(t)$$

where F is a seven-diagonal, sparse matrix of the form

(3.24) F =

$$\begin{bmatrix}
f_5 & f_6 & f_3 & f_4 & & & \\
f_6 & f_1 & f_2 & f_3 & f_4 & & 0 \\
f_3 & f_2 & f_1 & f_2 & f_3 & f_4 & \\
f_4 & f_3 & f_2 & f_1 & f_2 & f_3 & f_4 \\
 & f_4 & f_3 & f_2 & f_1 & f_2 & f_3 \\
 & & f_4 & f_3 & f_2 & f_1 & f_2 \\
 & 0 & & f_4 & f_3 & f_2 & f_1 \\
 & & & & f_4 & f_3 & f_6
\end{bmatrix}
\begin{matrix} \\ \\ \\ \\ \\ \\ f_3 \\ f_5 \end{matrix}$$

with

$$f_1 = 1+2r+3r^2+\frac{10}{3}r^3, \quad f_2 = -r-2r^2-\frac{5}{2}r^3, \quad f_3 = r^2+\frac{1}{2}r^3,$$

$$f_4 = -\frac{1}{6}r^3, \quad f_5 = 1+2r+\frac{5}{2}r^2+\frac{7}{3}r^3, \quad f_6 = -r-2r^2-\frac{7}{3}r^3 \quad .$$

The solution $\underline{U}(t+\ell)$ can be computed using an LU decomposition algorithm. The principal part of the local truncation error is given by (3.11) with $q = 4$; the error constant is $C_4 = -\frac{1}{24}$ (from Table 2.1). It is easy to see that this third order method is L_0-stable and may be extrapolated to give fourth order accuracy using (3.14); from (3.15) it is seen that $\alpha = \frac{8}{7}$. The principal part of the local truncation error of the extrapolated form of the method is given by (3.16) with $p = 5$; from Table 2.3 it is seen that $E_5 = 8/105$ for the method.

The amplification symbol of the extrapolated form of the method is

$$(3.25) \qquad S_{3,0}(z) = \frac{8}{7}\ (1+z+\frac{1}{2}z^2+\frac{1}{6}z^3)^{-2} - \frac{1}{7}\ (1+2z+2z^2+\frac{4}{3}z^3)^{-1} \quad .$$

Simplifying (3.25) gives

$$S_{3,0}(z) = \frac{1+2z+2z^2+\frac{4}{3}z^3-\frac{1}{12}z^4-\frac{1}{42}z^5-\frac{1}{252}z^6}{1+4z+8z^2+16z^3+\frac{119}{12}z^4+\frac{20}{3}z^5+\frac{119}{36}z^6+\frac{7}{6}z^7+\frac{5}{18}z^8+\frac{1}{27}z^9} ,$$

from which it is found that $|S_{3,0}(z)| \leq 1$ and $\lim_{z\to\infty} S_{3,0}(z) = 0$. Thus the property of L_0-stability is retained by the extrapolated form of the method. The amplification symbols $R_{3,0}(z)$ and $S_{3,0}(z)$ are produced in Fig 2.6.

Like the third order method based on the (2,1) Padé approximant, the third order method based on the (3,0) Padé approximant loses accuracy at the mesh points $(h,n\ell)$ and $(Nh,n\ell)$; this can be seen by examining (3.24). The (3,0) method also loses accuracy at the mesh points $(2h,n\ell)$ and $((N-1)h,n\ell)$ for $n = 0,1,\ldots$ but the numerical results to be reported in section 3.6 suggest that this additional loss of accuracy does not affect convergence. The method based on the (3,0) Padé approximant does not use the mesh points with co-ordinates (0,0) or (X,0), where discontinuities between initial and boundary conditions may exist, whereas the method based on the (2,1) Padé approximant does use these points. It is clear that the components of the principal parts of the local truncation errors relating to the raw and extrapolated forms of the (3,0) method, are greater in modulus than those of the (2,1) method. Therefore, the method based on the (2,1) Padé approximant can be expected to give more accurate results than that based on the (3,0) approximant. However, the method based on the (2,1) approximant becomes overstable for larger values of r.

3.5 A fourth order method

The final algorithm to be discussed for diffusion problems with one space variable is that obtained by replacing $\exp(\ell A)$ in (3.7) by its (2,2) Padé approximant giving

$$\underline{U}(t+\ell) = (I-\tfrac{1}{2}\ell A+\tfrac{1}{12}\ell^2A^2)^{-1}(I+\tfrac{1}{2}\ell A+\tfrac{1}{12}\ell^2A^2)\,\underline{U}(t) \quad . \tag{3.26}$$

Written implicitly (3.26) becomes

(3.27) $$(I-\tfrac{1}{2}\ell A+\tfrac{1}{12}\ell^2A^2)\ \underline{U}(t+\ell) = (I+\tfrac{1}{2}\ell A+\tfrac{1}{12}\ell^2A^2)\ \underline{U}(t)$$

and, applying (3.27) to each of the N mesh points at time level $t = n\ell$ $(n = 0,1,2,\ldots)$, again leads to $\underline{U}(t+\ell)$ being determined from a linear system of the form (3.12). The elements of the matrix E in (3.12) are now given by

$$e_1 = 1+r+\tfrac{1}{2}r^2, \quad e_2 = -\tfrac{1}{2}r-\tfrac{1}{3}r^2, \quad e_3 = \tfrac{1}{12}r^2, \quad e_4 = 1+r+\tfrac{5}{12}r^2$$

while the elements of $\underline{\phi}^n$ become

$$\phi_1^n = (1-r+\tfrac{5}{12}r^2)\ U_1^n + r\ (\tfrac{1}{2}-\tfrac{1}{3}r)\ U_2^n + \tfrac{1}{12}r^2\ U_3^n$$

$$\phi_m^n = \tfrac{1}{12}r^2\ U_{m-2}^n + r\ (\tfrac{1}{2}-\tfrac{1}{3}r)\ U_{m-1}^n + (1-r+\tfrac{1}{2}r^2)\ U_m^n$$

$$+ r\ (\tfrac{1}{2}-\tfrac{1}{3}r)\ U_{m+1}^n + \tfrac{1}{12}r^2\ U_{m+2}^n \quad ; \quad m = 2,\ldots,N-1 .$$

The solution of (3.27) is computed using a quindiagonal solver. A complex splitting should not be considered for this method.

The principal part of the local truncation error of (3.27) at the mesh points $(mh,n\ell)$ $(m = 1,2,\ldots N;\ n = 0,1,2,\ldots)$ is given in (3.11) with $q = 5$ and $C_5 = 1/720$. This value of C_5 is much smaller in modulus than either of the values of E_5 relating to the extrapolated forms of the methods based on the $(2,1)$ and $(3,0)$ Padé approximants.

It may be expected, therefore, that the $(2,2)$ method will give good results, particularly near the centre of the interval $0 \le x \le X$, for problems which do not have discontinuities between initial and boundary conditions. The amplification symbol, given by

$$R_{2,2}(z) = (1-\tfrac{1}{2}z+\tfrac{1}{12}z^2)/(1+\tfrac{1}{2}z+\tfrac{1}{12}z^2) \quad ,$$

is always positive for $z \ge 0$ (it has a minimum value of $7-4\sqrt{3}$ when $z = 2\sqrt{3}$) and tends asymptotically to $+1$ as $z\to\infty$. The numerical method is therefore A_0-stable but is not L_0-stable and, like the numerical method based on the $(1,1)$ Padé approximant (Lawson and Morris (1978)), oscillations in the solution are induced. The symbol $R_{2,2}(z)$ and $S_{2,2}(z)$ are produced in Fig 2.5 where

$$S_{2,2}(z) = \tfrac{16}{15}\{R_{2,2}(z)\}^2 - \tfrac{1}{15}R_{2,2}(2z) \quad .$$

3.6 Numerical results

To illustrate the behaviour of some of the schemes discussed in earlier sections, the model problem (3.1) is solved with $X = 2$ and boundary conditions given by (3.3). The initial conditions are taken to be $g(x) = 1$ for $0 \le x \le 2$. This problem was discussed by Lawson and Morris (1978) and Gourlay and Morris (1980), and has theoretical solution given by

$$u(x,t) = \sum_{k=1}^{\infty} \{1-(-1)^k\} \frac{2}{k\pi} \sin(\tfrac{1}{2}k\pi x) \exp(-\tfrac{1}{4}k^2\pi^2 t) .$$

The methods based on the (2,0) and (2,1) Padé approximants will be denoted by P20 and P21, respectively. The method based on the (2,2) Padé approximant will be denoted by P22. These methods will be compared with the Crank-Nicolson method, which is based on the (1,1) Padé approximant and will be denoted by P11. The extrapolated form of the methods based on the (2,0) and (3,0) Padé approximants will be denoted by P20E and P30E.

All the methods are tested using $\ell = 0.025$, $h = 0.05$ (giving $r = 10$), $\ell = 0.1$, $h = 0.05$ (giving $r = 40$), and $\ell = 0.1$, $h = 0.025$ (giving $r = 160$). The maximum errors at time $t = 1.2$ are given in Table 3.1.

It is noted from Table 3.1 that, for $r = 40$, the second order method P20 gives results as accurate as the third order multistage method of Gourlay and Morris (1980, p.647). The third order methods P21, P20E and P30 give numerical results better than the fourth order multistage method of Gourlay and Morris (1980,p.653) for $\theta = \frac{1}{2}$ and comparable results for $\theta = 0$.

It is clear from Table 3.1 that the accuracies of the L_0-stable methods P20, P20E, P30, P30E and P21 increase as h is refined. The overstability of the method P21 is also apparent. Table 3.1 also shows that, in the case of the higher order method P30, extrapolation does not produce much improvement in accuracy, predicted by the theory,

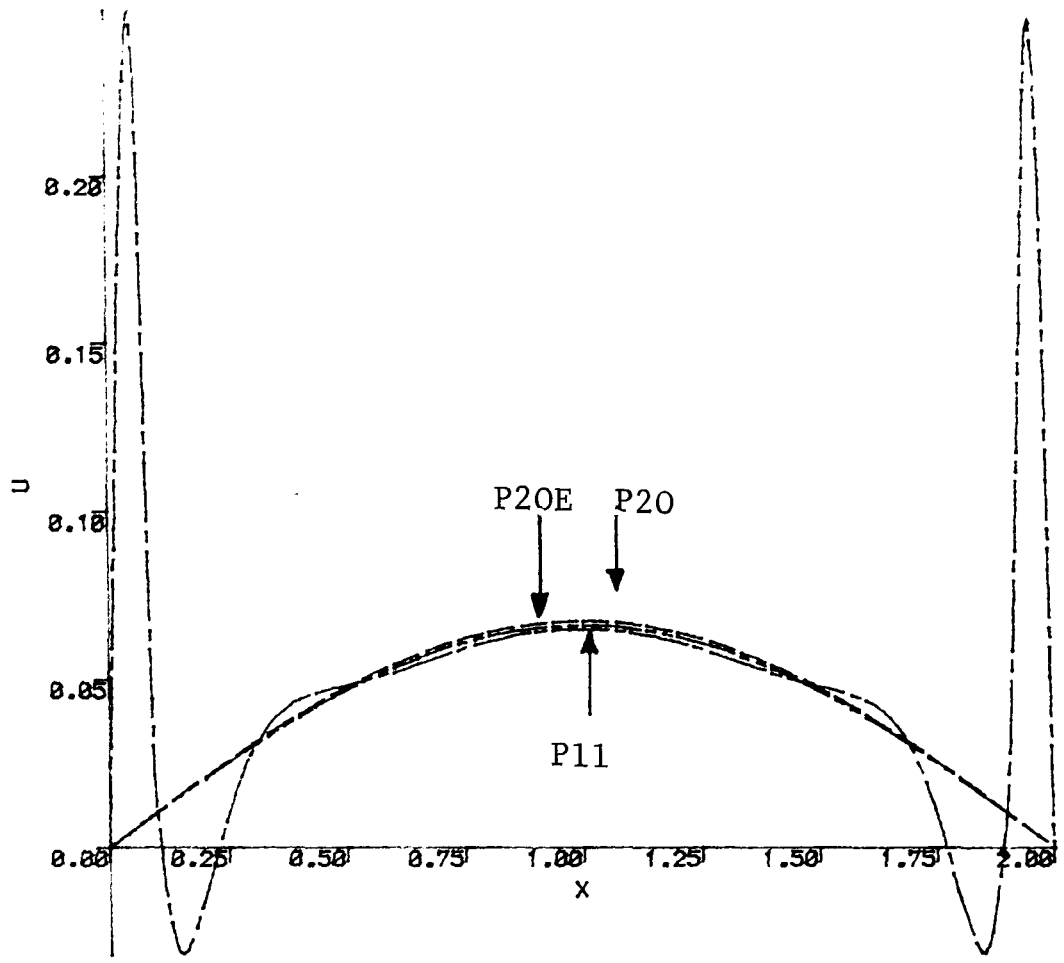


Figure 3.1: Numerical results at time $t = 1.2$ with $h = 0.05$, $\ell = 0.1$, $r = 40$.

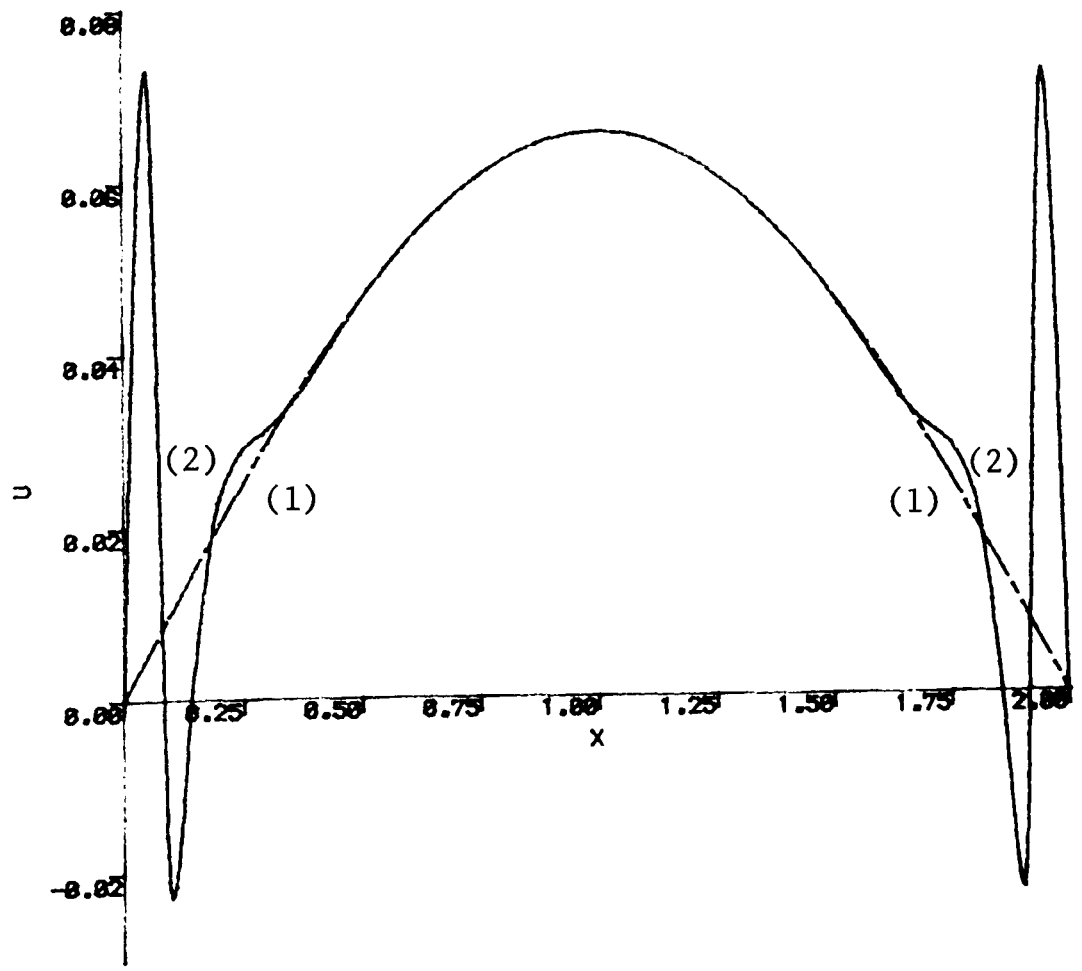


Figure 3.2: Numerical results at time $t = 1.2$ with $h = 0.05$, $\ell = 0.1$, $r = 40$.
(1) Theoretical solution, (2) P22.

for large values of h and small values of r. This is due to the dominance of the $O(\ell h^2)$ term in the principal part of the local truncation error; the accuracy of P30E is improved as h is refined. The A_0-stable method P22 gives poor results near the boundaries due to the discontinuities between boundary and initial conditions, and behaves in a similar way to the Crank-Nicolson method P11. The maximum errors given in Table 3.1 occur at the mid-point $x = 1$ for the L_0-stable methods, and near the boundaries for the A_0-stable methods. This is shown in Figures 3.1 and 3.2.

Table 3.1 Numerical results of model problem

Method	Order	Maximum errors r = 10	r = 40	r = 160
P11	2	0.28(-3)	0.24	0.52
P20	2	0.18(-3)	0.17(-2)	0.16(-2)
P20E	3	0.74(-4)	0.41(-3)	0.36(-3)
P21	3	0.67(-4)	0.28(-4)	-0.22(-4)
P30	3	0.69(-4)	0.17(-3)	0.12(-3)
P30E	4	0.67(-4)	0.87(-4)	0.37(-4)
P22	4	0.66(-4)	0.68(-1)	0.30

3.7 Two-space dimensions

Some of the difficulties encountered in implementing the methods developed for one-space dimension are magnified in the case of two-space dimensions. In particular, the square matrix A is now of order N^2 and is split into the form $A = B+C$, where B,C are block-diagonal and block-tridiagonal respectively, so that when the second power of A is required the matrices B^2, BC, CB, C^2 must be determined. Another difficulty is with the poor results given by A_0-stable methods, such as the Peaceman-Rachford method, when used to solve problems with discon-

tinuities between boundary and initial conditions.

The method which will be developed is based on the (2,0) Padé approximant to the matrix exponential function. This method will be seen to be second order accurate in time, the same as the Peaceman-Rachford method, and to be L_0-stable. In its extrapolated form the (2,0) method will be seen to be third order accurate in time and to retain the property of L_0-stability.

The constant coefficient heat equation in two space variables has the form

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad ; \quad 0 < x, y < X, \quad t > 0 \tag{3.28}$$

with homogeneous Dirichlet boundary conditions on the boundary $\partial\Omega$ of the square Ω defined by the lines $x = 0$, $y = 0$, $x = X$, $y = X$, and initial conditions

$$u(x,y,0) = g(x,y) \; . \tag{3.29}$$

It is assumed that $g(x,y)$, which is a given continuous function of x,y, does not necessarily have the value zero for $(x,y) \in \partial\Omega$, so that discontinuities between initial conditions and boundary conditions are permitted.

Both intervals $0 \le x \le X$ and $0 \le y \le X$ are divided into $N+1$ subintervals each of width h, so that $(N+1)h = X$ as before, and the time variable t is incremented in steps of ℓ. At each level $t = n\ell$ $(n = 0,1,2,...)$ the square Ω, together with its boundary $\partial\Omega$, have been superimposed by a square mesh with N^2 points within Ω and $N+2$ equally spaced points along each side of $\partial\Omega$.

The solution $u(x,y,t)$ of (3.28) is sought at each point $(kh,mh,n\ell)$ in $\Omega \times [\, t > 0 \,]$ where $k,m = 1,2,...N$ and $n = 0,1,2,..$. The theoretical solution of an approximating difference scheme at the mesh point $(kh,mh,n\ell)$ will be denoted by $U^n_{k,m}$; the vector $\underline{U}^n$ of such solutions will be ordered in the form

$$\underline{U}^n = (U^n_{1,1}, \quad U^n_{2,1}, \ldots U^n_{N,1}; U^n_{1,2}, U^n_{2,2}, \ldots, U^n_{N,2}; \ldots$$

(3.30)

$$\ldots; U^n_{1,N}, U^n_{2,N}, \ldots, U^n_{N,N})^T.$$

The space derivatives in (3.28) will be replaced by

$$\frac{\partial^2 u}{\partial x^2} = \{u(x-h,y,t) - 2u(x,y,t) + u(x+h,y,t)\}/h^2 + 0(h^2), \tag{3.31}$$

$$\frac{\partial^2 u}{\partial y^2} = \{u(x,y-h,t) - 2u(x,y,t) + u(x,y+h,t)\}/h^2 + 0(h^2) \tag{3.32}$$

and at each time level $t = n\ell$, (3.29) is applied to all N^2 interior mesh points of the square Ω with the space derivatives replaced by (3.31), (3.32). These N^2 applications result in a system of N^2 first order ordinary differential equations of the form (3.5), in which the matrix A is now of order N^2 and may be split into the constituent matrices B,C such that $A = B+C$.

The matrix B arises from the use of (3.31) in (3.29); it is block diagonal with tridiagonal blocks and has the form

$$B = h^{-2} \begin{bmatrix} B_1 & & & \\ & B_1 & & \\ & & B_1 & \\ & & & \ddots \\ & & & & B_1 \end{bmatrix} \tag{3.33}$$

where B_1 is the tridiagonal matrix of order N given by

$$B_1 = \begin{bmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & 0 \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 0 & & & 1 & -2 \end{bmatrix} \tag{3.34}$$

The matrix C arises from the use of (3.32) in (3.29); it is block tridiagonal with diagonal blocks and has the form

$$C = h^{-2} \begin{bmatrix} -2I & I & & \\ I & -2I & I & \\ & \ddots & \ddots & \ddots \\ & & I & -2I & I \end{bmatrix} \tag{3.35}$$

where I is the identity matrix of order N. The N^2 eigenvalues of the matrix A are real and negative and are given by

$$\lambda_{i,j} = -4h^{-2}\left[\sin^2\frac{i\pi}{2(N+1)} + \sin^2\frac{j\pi}{2(N+1)}\right] ; \quad i,j = 1,\ldots,N . \tag{3.36}$$

Solving the system of ordinary differential equations subject to the initial condition $\underline{U}(0) = \underline{g}$, gives

$$\underline{U}(t) = \exp\{t(B+C)\}\,\underline{g}$$

which satisfies the recurrence relation

$$\underline{U}(t+\ell) = \exp\{\ell(B+C)\}\,\underline{U}(t); \quad t = 0,\ell,2\ell,\ldots . \tag{3.37}$$

It is this recurrence relation which will be used in the development of the second order method.

3.8 Second order method and its extrapolation

The recurrence relation (3.37) may be written in the alternative forms

$$\underline{U}(t+\ell) = \exp(\ell B)\exp(\ell C)\,\underline{U}(t) + O(\ell^2) , \tag{3.38}$$

$$\underline{U}(t+\ell) = \exp(\ell C)\exp(\ell B)\,\underline{U}(t) + O(\ell^2) . \tag{3.39}$$

Using the (2,0) Padé approximant, equations (3.38), (3.39) may be written

$$\underline{U}^*(t+\ell) = (I-\ell B+\tfrac{1}{2}\ell^2B^2)^{-1}(I-\ell C+\tfrac{1}{2}\ell^2C^2)^{-1}\underline{U}(t) , \tag{3.40}$$

$$\underline{U}^+(t+\ell) = (I-\ell C+\tfrac{1}{2}\ell^2C^2)^{-1}(I-\ell B+\tfrac{1}{2}\ell^2B^2)^{-1}\underline{U}(t) , \tag{3.41}$$

respectively. Expanding the matrix inverses in (3.40), (3.41) confirms that each is only first order accurate in time when compared with the Maclaurin expansion of $\exp(\ell(B+C))$ given by

$$\exp(\ell(B+C)) = I+\ell(B+C) + \tfrac{1}{2}\ell^2(B^2+C^2+BC+CB) + \tfrac{1}{6}\ell^3(B^3+C^3+B^2C+BC^2+CB^2+C^2B+BCB+CBC)+\ldots . \tag{3.42}$$

Combining $\underline{U}^*$ and $\underline{U}^+$ by the linear relation

$$\underline{U}(t+\ell) = \tfrac{1}{2}(\underline{U}^* + \underline{U}^+) \tag{3.43}$$

gives

$$\underline{U}(t+\ell) = \{I+\ell(B+C) + \tfrac{1}{2}\ell^2(B^2+C^2+BC+CB) + O(\ell^3)\}\ \underline{U}(t) \tag{3.44}$$

which is second order accurate in time.

The splittings (3.40), (3.41) and the relation (3.43), are formalized by the following algorithm which requires four applications of a quindiagonal solver:

$$\begin{aligned}
&(I-\ell C+\tfrac{1}{2}\ell^2C^2)\ \underline{V}^{(1)} = \underline{U}(t)\ , \\
&(I-\ell B+\tfrac{1}{2}\ell^2B^2)\ \underline{U}^{*} = \underline{V}^{(1)}\ ; \\
&(I-\ell B+\tfrac{1}{2}\ell^2B^2)\ \underline{V}^{(2)} = \underline{U}(t)\ , \\
&(I-\ell C+\tfrac{1}{2}\ell^2C^2)\ \underline{U}^{+} = \underline{V}^{(2)}\ ; \\
&\underline{U}(t+\ell) = \tfrac{1}{2}(\underline{U}^{*}+\underline{U}^{+})\ .
\end{aligned} \tag{3.45}$$

In (3.45) $\underline{V}^{(1)}$ and $\underline{V}^{(2)}$ are intermediate vectors each of order N^2. Following Gourlay and Morris (1980, p.644), it is found that the second order algorithm (3.45) is L_0-stable.

The second order accuracy of the method may be extrapolated to third order by, first of all, considering (3.40), (3.41) over two single time steps to give

$$\underline{U}^{**}(t+2\ell) = \{(I-\ell B+\tfrac{1}{2}\ell^2B^2)^{-1}(I-\ell C+\tfrac{1}{2}\ell^2C^2)^{-1}\}^2\ \underline{U}(t) \tag{3.46}$$

$$\underline{U}^{++}(t+2\ell) = \{(I-\ell C+\tfrac{1}{2}\ell^2C^2)^{-1}(I-\ell B+\tfrac{1}{2}\ell^2B^2)^{-1}\}^2\ \underline{U}(t) \tag{3.47}$$

Expanding the matrix inverses in (3.46), (3.47) verifies that each is only first order accurate when compared with the Maclaurin expansion of $\exp\{2\ell(B+C)\}$ given by

$$\begin{aligned}
\exp\{2\ell(B+C)\} = {} & I+2\ell(B+C)+2\ell^2(B^2+C^2+BC+CB) \\
& + \tfrac{4}{3}\ell^3(B^3+C^3+B^2C+BC^2+C^2B+CB^2+BCB+CBC)+\ldots\ .
\end{aligned} \tag{3.48}$$

Substituting the expansions of (3.46), (3.47) in

$$\underline{U}^{(0)}(t+2\ell) = \tfrac{1}{2}(\underline{U}^{**}+\underline{U}^{++})\ ,$$

however, gives

$$\begin{aligned}
\underline{U}^{(0)}(t+2\ell) = {} & [\,I+2\ell(B+C)+2\ell^2(B^2+C^2+BC+CB) \\
& +\ell^3\{B^3+C^3+\tfrac{3}{2}(BC^2+B^2C+CB^2+C^2B)+BCB+CBC\}+O(\ell^4)]\underline{U}(t)
\end{aligned} \tag{3.49}$$

showing that $\underline{U}^{(0)}$ is second order accurate in time.

Writing (5.40), (5.41) over a double time step 2ℓ gives

$$\underline{U}^{(1)}(t+2\ell) = (I-2\ell B+2\ell^2B^2)^{-1}(I-2\ell C+2\ell^2C^2)^{-1}\underline{U}(t) \quad , \tag{3.50}$$

$$\underline{U}^{(2)}(t+2\ell) = (I-2\ell C+2\ell^2C^2)^{-1}(I-2\ell B+2\ell^2B^2)^{-1}\underline{U}(t) \quad . \tag{3.51}$$

Expanding the matrix inverses in (3.50), (3.51) gives

$$\underline{U}^{(1)}(t+2\ell) = \{I+2\ell(B+C)+2\ell^2(B^2+C^2+2BC) + 4\ell^3(B^2C+BC^2)+O(\ell^4)\}\underline{U}(t) \quad , \tag{3.52}$$

$$\underline{U}^{(2)}(t+2\ell) = \{I+2\ell(B+C)+2\ell^2(B^2+C^2+2CB) + 4\ell^3(C^2B+CB^2)+O(\ell^4)\}\underline{U}(t) \quad , \tag{3.53}$$

respectively, showing that each is first order accurate in time.

The linear combination of (3.49), (3.52), (3.53), defined by

$$\underline{U}^{(E)}(t+2\ell) = \frac{4}{3}\underline{U}^{(0)} - \frac{1}{6}(\underline{U}^{(1)} + \underline{U}^{(2)}) \quad ,$$

is third order accurate in time when compared with the Maclaurin expansion (3.48).

The principal part of the local truncation error of (3.43) when applied to the mesh points $(kh,mh,n\ell)$, with $k,m = 2,\ldots,N-1$ and $n = 0,1,2,\ldots$, is found to be

$$(\frac{1}{6}\ell^3 \frac{\partial^3 u}{\partial t^3} - \frac{1}{12}\ell h^2 (\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}))^n_{k,m} \quad ; \quad k,m \neq 1,N \tag{3.54}$$

which, following extrapolation, becomes

$$(-\frac{1}{3}\ell^4 \frac{\partial^4 u}{\partial t^4} - \frac{1}{12}\ell h^2 (\frac{\partial^4 u}{\partial x^4} + \frac{\partial^4 u}{\partial y^4}))^n_{k,m} \quad ; \quad k,m \neq 1,N \quad . \tag{3.55}$$

The implementation of the algorithm based on the (2,0) Padé approximant may be carried out by means of the following strategy:

(i) intermediate vectors $\underline{V}^{(1)}$, $\underline{V}^{(2)}$ are introduced and used to find the estimates $\underline{U}^*(t+\ell)$. $\underline{U}^+(t+\ell)$, as follows:

$$(I-\ell C+\frac{1}{2}\ell^2C^2)\,\underline{V}^{(1)} = \underline{U}(t) \quad , \tag{3.56}$$

$$(I-\ell B+\frac{1}{2}\ell^2B^2)\,\underline{U}^*(t+\ell) = \underline{V}^{(1)} \quad ; \tag{3.57}$$

(3.58) $$(I-\ell B+\tfrac{1}{2}\ell^2B^2)\ \underline{V}^{(2)} = \underline{U}(t) \quad ,$$

(3.59) $$(I-\ell C+\tfrac{1}{2}\ell^2C^2)\ \underline{U}^{+}(t+\ell) = \underline{V}^{(2)} \quad ;$$

(ii) intermediate vectors $\underline{V}^{(3)}$, $\underline{V}^{(4)}$ are introduced to extend the estimates $\underline{U}^{*}(t+\ell)$, $\underline{U}^{+}(t+\ell)$ over a second single time step as follows:

(3.60) $$(I-\ell C+\tfrac{1}{2}\ell^2C^2)\ \underline{V}^{(3)} = \underline{U}^{*}(t+\ell) \quad ,$$

(3.61) $$(I-\ell B+\tfrac{1}{2}\ell^2B^2)\ \underline{U}^{**}(t+2\ell) = \underline{V}^{(3)} \quad ;$$

(3.62) $$(I-\ell B+\tfrac{1}{2}\ell^2B^2)\ \underline{V}^{(4)} = \underline{U}^{+}(t+\ell) \quad ,$$

(3.63) $$(I-\ell C+\tfrac{1}{2}\ell^2C^2)\ \underline{U}^{++}(t+2\ell) = \underline{V}^{(4)} \quad ;$$

(iii) the second order estimate $\underline{U}^{(0)}$ is now calculated from

(3.64) $$\underline{U}^{(0)} = \frac{1}{2}\ (\underline{U}^{**} + \underline{U}^{++}) \quad ;$$

(iv) intermediate vectors $\underline{V}^{(5)}$, $\underline{V}^{(6)}$ are introduced and used with a double time step to find the estimates $\underline{U}^{(1)}(t+2\ell)$ and $\underline{U}^{(2)}(t+2\ell)$ as follows:

(3.65) $$(I-2\ell C+2\ell^2C^2)\ \underline{V}^{(5)} = \underline{U}(t) \quad ,$$

(3.66) $$(I-2\ell B+2\ell^2B^2)\ \underline{U}^{(1)}(t+2\ell) = \underline{V}^{(5)} \quad ;$$

(3.67) $$(I-2\ell B+2\ell^2B^2)\ \underline{V}^{(6)} = \underline{U}(t) \quad ,$$

(3.68) $$(I-2\ell C+2\ell^2C^2)\ \underline{U}^{(2)}(t+2\ell) = \underline{V}^{(6)} \quad ;$$

(v) the third order accurate estimate $\underline{U}^{(E)}(t+2\ell)$, given by

(3.69) $$\underline{U}^{(E)}(t+2\ell) = \frac{4}{3}\ \underline{U}^{(0)} - \frac{1}{6}(\underline{U}^{(1)} + \underline{U}^{(2)}) \quad ,$$

is now calculated.

In order to illustrate the behaviours of the L_0-stable methods in two space variables, the following model problem, which was introduced in the paper by Lawson and Morris (1978), is solved using the second order method (3.45), and third order method as an extrapolation of the second order method.

The problem is

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \quad ; \quad 0 < x,y < 2 \quad , \quad t > 0$$

subject to the initial conditions

$$u(x,y,0) = \sin\left(\tfrac{1}{2}\pi y\right) \quad ; \quad 0 \le x,y \le 2$$

and boundary conditions

$$u(x,y,t) = 0 \quad ; \quad x = 0,\ y = 0,\ x = 2,\ y = 2,\ t > 0 \ .$$

The initial distribution is shown in Fig 3.3 and the theoretical solution

$$u(x,y,t) = \sin \tfrac{1}{2}\pi y \sum_{k=1}^{\infty} [\{1-(-1)^k\} \frac{2}{k\pi} \sin\left(\tfrac{1}{2}k\pi x\right) \exp\left(-\tfrac{1}{4}\pi^2(k^2+1)t\right)]$$

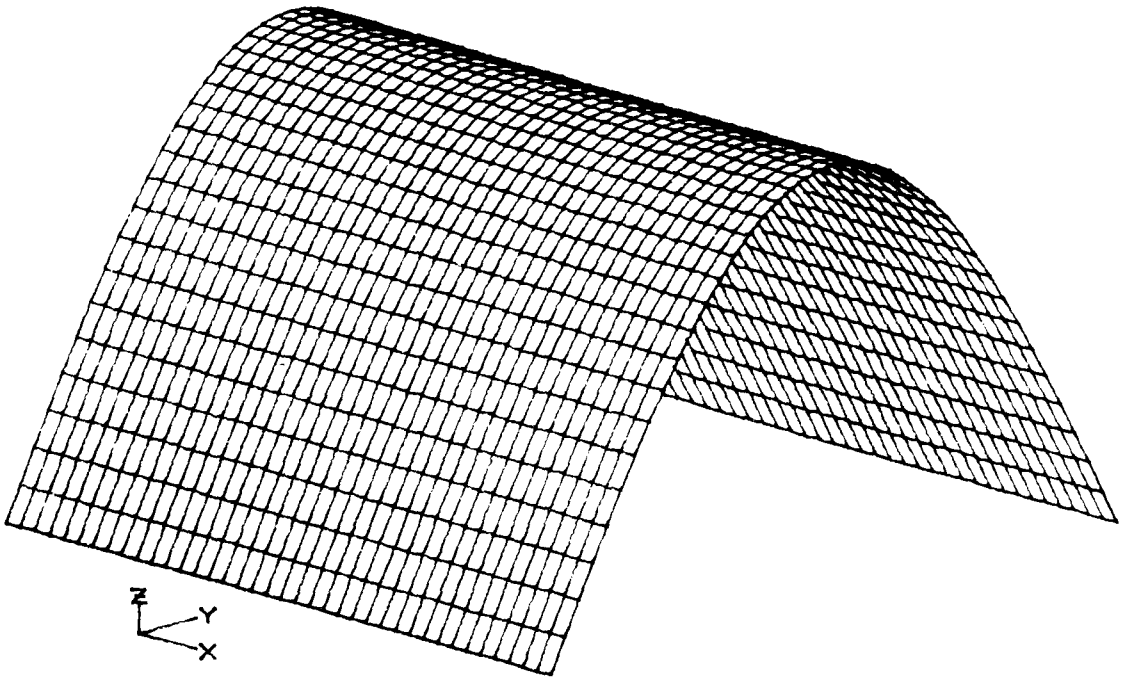
is depicted at time $t = 1.0$ in Fig 3.4.

The solution is computed at time $t = 1.0$ using $\ell = 0.025$, $h = 0.05$ (giving $r = 10$), $\ell = 0.1$, $h = 0.05$ (giving $r = 40$) and $\ell = 0.1$, $h = 0.025$ (giving $r = 160$). The maximum error found in each case is given in Table 3.2.

Table 3.2

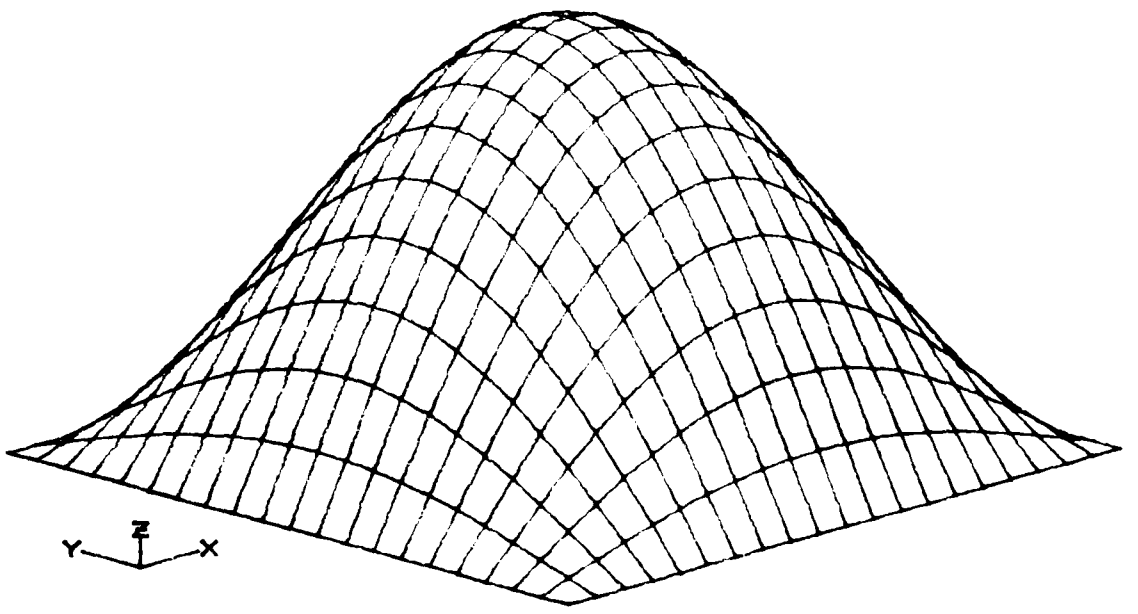
Method	Order	Maximum errors		
		r = 10	r = 40	r = 160
(2,0) Method (3.45)	2	0.46(-4)	0.34(-3)	0.33(-3)
Extrapolated (2,0)	3	0.80(-5)	0.35(-4)	0.25(-4)
Peaceman-Rachford	2	0.33(-3)	0.23(-1)	0.45(-1)

The distribution of the computed solution for the second order method is shown in Fig 3.5 and for the third order method is shown in Fig 3.6. It is seen that in each case the maximum errors occur at the point $x = 1$, $y = 1$. A comparison with the Lawson and Morris (1978) second order algorithm indicates that the second order method (3.45) gives higher accuracy at the expense of an increase in CPU time. However, the superior results justify this minimal increase in computer time.



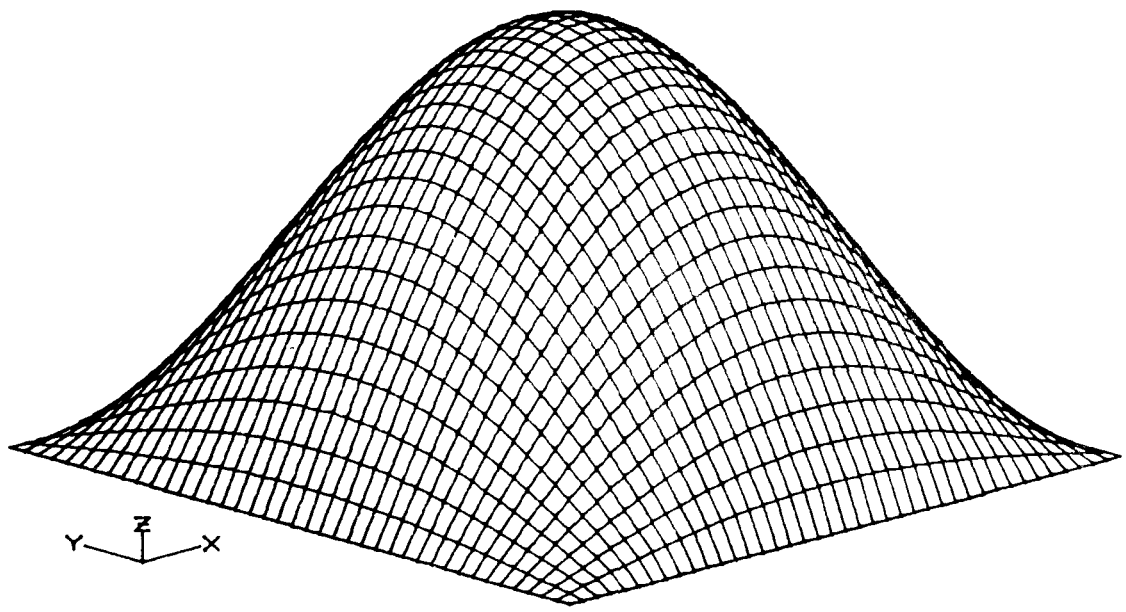
Z:0.000 TO 1.000

Figure 3.3: Initial distribution for two - space variable problem.



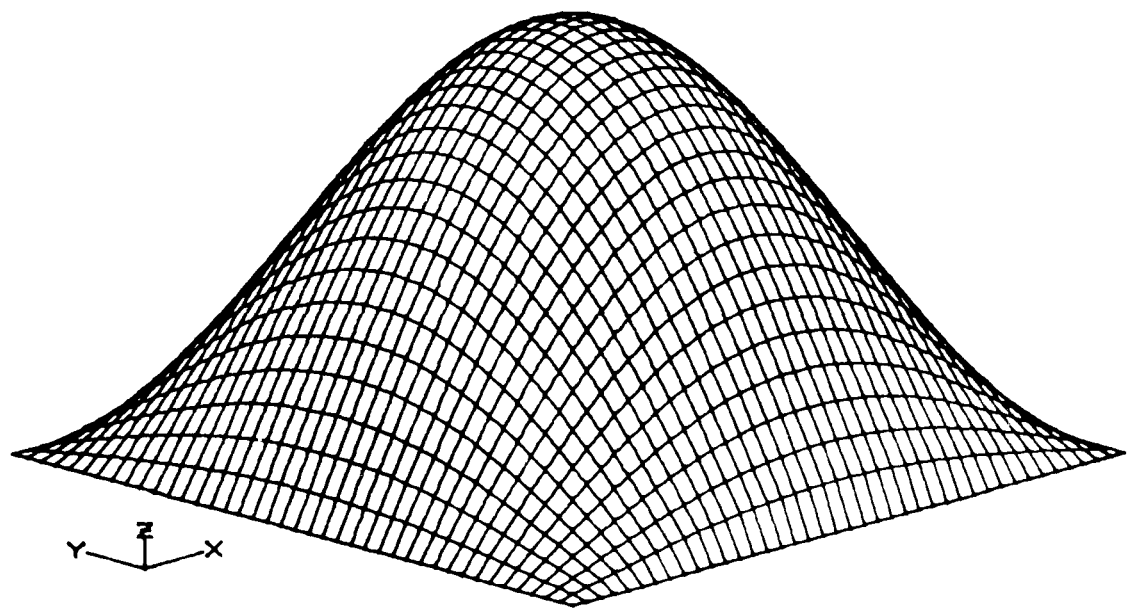
Z:0.00000 TO 0.00916

Figure 3.4: Theoretical solution at time t = 1.0 .



Z:0.00000 TO 0.00950

Figure 3.5: Numerical solution at time t = 1.0 with h = 0.05, ℓ = 0.1, r = 40 using P20.



Z:0.00000 TO 0.00919

Figure 3.6: Numerical solution at time t = 1.0 with h = 0.05, ℓ = 0.1, r = 40 using P20E.

CHAPTER 4

FIRST ORDER HYPERBOLIC EQUATIONS

4.1 Introduction

In recent years much attention has been devoted in the literature to the extrapolation in time of low order methods for the numerical solution of first order hyperbolic partial differential equations as well as for second order parabolic equations.

Essentially the same procedure may be followed for parabolic equations (Lawson and Morris (1978)), (Gourlay and Morris (1980)) and hyperbolic equations (Khaliq and Twizell (1982)): that is to say, the space derivatives in the differential equations are approximated by a suitable finite difference replacement, and the resulting system of first order ordinary differential equations solved using a stable numerical method. From this stage of the computation onwards, the accuracy in time can be controlled by a suitable choice of method for solving an ordinary differential equation; improvement in the accuracy in space, on the other hand, requires a different replacement of the space derivative in the partial differential equation.

From the point where the replacement of the space derivative has been chosen, accuracy in time can be varied by a multistage method (Gourlay and Morris (1980, 1981)) which involves a spread over three or more time increments, or by a method involving a similar spread over more than three mesh points at a give time level (Mitchell and Griffiths (1980), Twizell and Khaliq (1981), Khaliq and Twizell (1982)). The former type of methods is, in effect, an application of linear onestep methods for systems of ordinary differential equations, while the latter is an application of multiderivative methods (Twizell and Khaliq (1981)). A family of methods related to the latter type will be developed in this Chapter.

Both approaches have a weakness which is the other's strength: using a multistage method, seeking the solution at certain fixed times requires

the time interval to be divided into two or more subintervals depending upon the accuracy required, whereas the integration can be carried out without subdividing the time interval if an A-stable or L-stable multiderivative method is used. On the other hand, implicit multistage methods need only tridiagonal solvers to obtain the solution (five at each time level for third order accuracy in time and nine for fourth order accuracy (Gourlay and Morris (1980)), whereas the multiderivative methods in Chapter 3, based on central difference replacements of the space derivative, (Khaliq and Twizell (1982)), need quindiagonal or seven diagonal solvers.

The methods to be discussed in sections 4.3, 4.4, 4.5, are based on backward difference replacements of the space derivatives and can therefore be used explicitly so that here, too, they have an advantage over multistage formulations. The use of backward difference replacements has the advantage that the oscillations which are always present with central difference replacements (section 4.2), do not arise. Also the difficulties which arise in parabolic equations because of stiffness are not present in solving hyperbolic equations by multiderivative techniques. The methods will use function values at only two time levels as in Khaliq and Twizell (1982), unlike the methods developed by Oliger (1974), and depend on the theorems of Gustaffson (1972) for the establishment of stability. The methods are tested in section 4.6 on a number of problems from the literature and, finally, conclusions are drawn in section 4.7.

4.2 Central difference approximation in space

Consider the first order hyperbolic partial differential equation

$$\frac{\partial u}{\partial t} + a \frac{\partial u}{\partial x} = 0 \quad ; \quad a > 0,\ x > 0,\ t > 0 \ , \tag{4.1}$$

where a is a constant, with initial conditions

$$u(x,0) = g(x) \quad ; \quad x \geq 0 \tag{4.2}$$

and boundary conditions

$$u(0,t) = v(t) \quad ; \quad t > 0 \ ; \tag{4.3}$$

equations (4.1), (4.2), (4.3) form the initial-boundary value problem. Suppose that the solution of (4.1) is sought in some region $R = [0 < x < X] \times [t > 0]$ of the first quarter plane $x > 0,\ t > 0$ (out flow problem). The interval $0 \leq x \leq X$ is divided into N equal parts each of width h, so that $Nh = X$, and the time variable t is discretized in steps of length ℓ. The open region R and its boundary ∂R, consisting of the axes $t = 0$, $x = 0$ and the line $x = X$, have been covered by a rectangular mesh, the mesh points having co-ordinates $(mh, n\ell)$ where $m = 0,1,\ldots,N$ and $n = 0,1,2,\ldots$. The theoretical solution of a finite difference scheme approximating the differential equation at this point, will be denoted by $U_m^n \equiv U(mh, n\ell)$.

Replacing the space derivative in (4.1) by the central difference formula

$$\frac{\partial u}{\partial x} = [u(x+h,t) - u(x-h,t)]/2h + O(h^2) \ , \tag{4.4}$$

and applying (4.1) with (4.4), (4.2), (4.3) to all N interior mesh points at time level $t = n\ell$ $(n = 0,1,\ldots)$, leads to the system of first order ordinary differential equations

$$\frac{d\underline{U}}{dt} = -\frac{1}{2}aB\underline{U}(t) + \frac{1}{2}a\underline{w}_t \tag{4.5}$$

where $\underline{U}(t) = [U_1(t), U_2(t), \ldots, U_N(t)]^T$, T denoting transpose, is the vector of approximate solutions of (4.1) at time $t > 0$. In (4.5) B is a square matrix of order N given by

(4.6) $$hB = \begin{bmatrix} 0 & 1 & & & \\ -1 & 0 & 1 & & 0 \\ & -1 & 0 & 1 & \\ & & \ddots & \ddots & \ddots \\ 0 & & & -1 & 0 & 1 \\ & & & 0 & -1 & 0 \end{bmatrix}$$

and $\underline{w}_t = \frac{1}{h}\,[\,v_t, 0, \ldots, 0, -U^n_{N+1}\,]^T$ is the N-component vector whose first element is the numerical (frozen) value of the boundary condition at time $t = n\ell$ and whose last element is minus the value of the solution at the point $((N+1)h, t)$. This means that knowledge of the solution is required on some "boundary" beyond the part of the x-axis under consideration, thus overposing the problem (4.1). However, for periodic boundary conditions, central difference approximations have often been used in the literature, see for example, Mitchell (1969), Kreiss and Oliger (1972), Smith (1978), Mitchell and Griffiths (1980), together with further references therein.

The solution of (4.5) with (4.2) is

(4.7) $$\underline{U}(t) = B^{-1}\underline{w}_t + \exp\,(-\tfrac{1}{2}\,atB)\{\,\underline{g} - B^{-1}\underline{w}_t\,\}$$

where $\underline{g}$ is the vector of initial values. In equation (4.7), $\underline{U}(t)$ satisfies the recurrence relation

(4.8) $$\underline{U}(t+\ell) = B^{-1}\underline{w}_t + \exp\,(-\tfrac{1}{2}\,a\ell B)\{\underline{U}(t) - B^{-1}\underline{w}_t\,\}$$

It is clear from (4.6) that each eigenvalue $\lambda_s = 2i\cos\frac{s\pi}{N+1}$, $s = 1,2,\ldots, N$, $i = \sqrt{-1}$, is complex and hence $\exp\,(-\frac{1}{2}at\lambda_s)$ in (4.7) (the matrix B being diagonalizable, Morton (1980; p.678)) is an oscillatory function. Thus the solution of (4.5) exhibits oscillations.

Price et al (1966) first observed the possibility of oscillations occurring when a central difference replacement is used for the space discretization in the convection part of the diffusion-convection problem and emphasised the need for methods which do not introduce such oscillations.

If, for example, the (1,0) Padé approximant is used to replace the matrix exponential function, (4.8) becomes

$$(I + \tfrac{1}{2}a\ell B)\,\underline{U}(t+\ell) - \tfrac{1}{2}a\ell \underline{w}_{t+\ell} = \underline{U}(t) \quad . \tag{4.9}$$

Applying (4.9) to the mesh points $(mh,n\ell)$ gives the four point, implicit scheme,

$$U_m^{n+1} + \frac{1}{2}\,ap\,(U_{m+1}^{n+1} - U_{m-1}^{n+1}) = U_m^n \quad .$$

An analysis of the stability (Mitchell (1969,p.37), Khaliq and Twizell (1982)) of this second order algorithm, indicates that the method is stable for all positive ℓ. Employing the (1,1) Padé approximant in (4.8) leads to

$$(I + \tfrac{1}{4}\,a\ell B)\,\underline{U}(t+\ell) - \tfrac{1}{4}a\ell \underline{w}_{t+\ell} = (I - \tfrac{1}{4}a\ell B)\,\underline{U}(t) + \tfrac{1}{4}a\ell \underline{w}_t \quad . \tag{4.10}$$

Application of (4.10) to the mesh points $(mh,n\ell)$ gives a Crank-Nicholson type scheme

$$-\tfrac{1}{4}ap\,U_{m-1}^{+1} + U_m^{n+1} + \tfrac{1}{4}ap\,U_{m+1}^{n+1} = \tfrac{1}{4}ap\,U_{m-1}^n + U_m^n + \tfrac{1}{4}ap\,U_{m+1}^n \tag{4.11}$$

which is known to be unconditionally stable (Mitchell (1969,p.167), Gustafsson et al (1972,p.664)), where $p = \ell/h$.

The solution will remain oscillatory, though the schemes are stable. However, the oscillations may be reduced by making the coefficient matrices in (4.9) and (4.10) diagonally dominant, see for example, Hirsch and Rudy (1974), or by restricting the space-step h with respect to the convection parameter a. But for large values of a this may prove, computationally, prohibitively expensive, see,for example, Price et al (1966). Thus the oscillations will be present no matter whether an A-stable or L-stable method is used. Numerical results to support these observations are given in section 4.6.

4.3 Low order (one-sided) approximation in space

Replacing the space derivative in (4.1) by the low order backward

difference formula

$$\frac{\partial u}{\partial x} = [\, u(x,t) - u(x-h,t) \,]/h + 0(h) \,, \tag{4.12}$$

and applying (4.1) with (4.12), (4.2) and (4.3) to all N interior mesh points at time level $t = n\ell$ $(n = 0,\ldots)$, leads to the system of first order ordinary differential equations

$$\frac{d\underline{U}(t)}{dt} = -aC\underline{U}(t) + a\underline{c}_t \tag{4.13}$$

where $\underline{U}(t) = [U_1(t), U_2(t),\ldots, U_N(t)]^T$, T denoting transpose, is the vector of approximate solutions of (4.1) at time $t>0$. In (4.13) C is the lower bi-diagonal matrix of order N given by

$$hC = \begin{bmatrix} 1 & & & \\ -1 & 1 & & 0 \\ & -1 & 1 & \\ & & \ddots & \ddots \\ 0 & & & -1 \quad 1 \end{bmatrix} , \tag{4.14}$$

and $\underline{c}_t$ is the vector with N elements given by

$$h\underline{c}_t = [\, v_t, 0, 0, \ldots, 0 \,]^T \tag{4.15}$$

where v_t is the numerical (frozen) value of the boundary condition at time $t = n\ell$. The solution of the non-stiff system of differential equations (4.13) is

$$\underline{U}(t) = C^{-1}\underline{c}_t + \exp\,(-atC)\,\{\underline{g} - C^{-1}\underline{c}_t\}, \tag{4.16}$$

where $\underline{g}$ is the vector of initial values, and (4.16) satisfies the recurrence relation

$$\underline{U}(t+\ell) = C^{-1}\underline{c}_t + \exp\,(-a\ell C)\,\{\underline{U}(t) - C^{-1}\underline{c}_t\} \,. \tag{4.17}$$

Using the (m,k) Padé approximants to the exponential function (Appendix I) to replace the matrix exponential function in (4.17), leads to a family of implicit finite difference schemes which are unconditionally stable for $m \geq k$ (by Theorem 1) and which may be used explicitly because

of the nature of the initial and boundary conditions (4.2) and (4.3). The principal part of the local truncation error of such a method has the form

$$(4.18) \qquad \left(-\frac{1}{2}a\ell h\frac{\partial^2 u}{\partial x^2} + C_q \ell^2 \frac{\partial^q u}{\partial t^q}\right)_m^n$$

where the constant C_q $(q = m+k+1)$ are given in Table 2.1 (Chapter 2).

Using the (1,0) Padé approximant to the exponential matrix function in (4.17), gives

$$(4.19) \qquad \underline{U}(t+\ell) = C^{-1}\underline{c}_t + (I+a\ell C)^{-1}\{\underline{U}(t) - C^{-1}\underline{c}_t\} + O(\ell^2) .$$

Writing (4.19) in implicit form and replacing $\underline{c}_t$ with the N-component vector $\underline{c}_{t+\ell} = [v_{t+\ell},0,\ldots,0]^T$, where $v_{t+\ell}$ is the numerical value of the boundary condition at time $t+\ell$, gives

$$(4.20) \qquad (I+a\ell C)\,\underline{U}(t+\ell) - a\ell\underline{c}_{t+\ell} = \underline{U}(t) .$$

Applying (4.20) to the point $(mh,n\ell)$ gives the three point, implicit scheme

$$(4.21) \qquad (1+ap)\,U_m^{n+1} - apU_{m-1}^{n+1} = U_m^n .$$

This scheme appears in Mitchell and Griffiths (1980,p.170); it is first order accurate and can be shown to be unconditionally stable by the method of von Neumann.

Formula (4.21) gives the explicit algorithm

$$(4.22) \qquad U_m^{n+1} = \frac{ap}{1+ap}\,U_{m-1}^{n+1} + \frac{1}{1+ap}\,U_m^n$$

which, together with the initial and boundary conditions, enables $\underline{U}$ to be calculated at all grid points in the first quadrant of the (x,t) plane. Using the (1,1) Padé approximant to the matrix exponential function in (4.17), gives

$$(4.23) \qquad (I+\tfrac{1}{2}a\ell C)\,\underline{U}(t+\ell) = (I-\tfrac{1}{2}a\ell C)\,\underline{U}(t)+\tfrac{1}{2}a\ell(\underline{c}_{t+\ell}+\underline{c}_t) .$$

At the mesh points $(mh,n\ell)$ where $m = 1,2,\ldots,N$, $n = 1,2,\ldots$, (4.23) suggests the algorithm

for $m = 1$

$$(4.24)\qquad U_1^{n+1} = [\,(1-\tfrac{1}{2}ap)\,U_1^n + \tfrac{1}{2}ap(U_0^{n+1} + U_0^n)\,] \,/\, (1+\tfrac{1}{2}ap)\ ;$$

for $m = 2,\ldots,N$

$$(4.25)\qquad U_m^{n+1} = [\,\tfrac{1}{2}ap\,U_{m-1}^{n+1} + (1-\tfrac{1}{2}ap)\,U_m^n + \tfrac{1}{2}ap\,U_{m-1}^n\,] \,/\, (1+\tfrac{1}{2}ap)\ .$$

The principal part of the local truncation error of (4.23) at the mesh points $(mh,n\ell)$ is

$$(4.26)\qquad (-\tfrac{1}{2}a\ell h\,\frac{\partial^2 u}{\partial x^2} - \tfrac{1}{12}\ell^3\,\frac{\partial^3 u}{\partial t^3})_m^n\ ,$$

thus indicating a second order (in time), A_0-stable method. Employing the (2,0) and (2,1) Padé approximants in (4.17), gives second and third order L_0-stable methods respectively. Since the eigenvalues of the matrix C are $1/h$, (4.13) is a non-stiff system of differential equations and hence the solution (4.16) remains non-oscillatory.

Using the (2,2) Padé approximant to the exponential matrix function in (4.17), gives

$$(4.27)\qquad \begin{aligned}(I+\tfrac{a\ell}{2}C + \tfrac{a^2\ell^2}{12}C^2)\,\underline{U}(t+\ell) &= (I - \tfrac{a\ell}{2}C + \tfrac{a^2\ell^2}{12}C^2)\underline{U}(t)\\ &\quad + (\tfrac{a\ell}{2}I + \tfrac{a^2\ell^2}{12}C)\,\underline{c}_{t+\ell} + (\tfrac{a\ell}{2}I - \tfrac{a^2\ell^2}{12}C)\,\underline{c}_t\ .\end{aligned}$$

The implicit algorithm (4.27) may also be used explicitly at the mesh points $(mh,n\ell)$:

for $m = 1$

$$U_1^{n+1} = \left[\tfrac{1}{2}ap\,(U_0^{n+1}+U_0^n) + \frac{a^2p^2}{12}(U_0^{n+1} - U_0^n) + (1-\frac{ap}{2}+\frac{1}{12}a^2p^2)U_1^n\right] / (1+\frac{ap}{2}+\frac{a^2p^2}{12})\ ;$$

for $m = 2$

$$U_2^{n+1} = \left[\frac{ap}{2}(U_1^n+U_1^{n+1}) + \frac{a^2p^2}{6}(U_1^{n+1}-U_1^n) + \frac{a^2p^2}{12}(U_0^n-U_0^{n+1}) + (1-\frac{ap}{2}+\frac{a^2p^2}{12})U_2^n\right] /$$

$$(1 + \frac{ap}{2}+\frac{a^2p^2}{12})\ ;$$

for $m = 3,\ldots,N$

$$(4.28)\qquad U_m^{n+1} = \left[\frac{ap}{2}(U_{m-1}^{n+1}+U_{m-1}^n) + \frac{a^2p^2}{6}(U_{m-1}^{n+1}-U_{m-1}^n) + \frac{a^2p^2}{12}(U_{m-2}^n-U_{m-2}^{n+1}) + (1-\frac{ap}{2}+\frac{a^2p^2}{12})U_m^n\right] /$$

$$(1+\frac{ap}{2}+\frac{a^2p^2}{12})\ .$$

This scheme is fourth order accurate in time, and the principal part of its local truncation error is $(-\frac{1}{2}a\ell h\,\frac{\partial^2 u}{\partial x^2} + \frac{1}{720}\,\ell^5\,\frac{\partial^5 u}{\partial t^5})_m^n$ at $(mh, n\ell)$ in R.

The component of the local truncation error due to the chosen (M,K) Padé approximant, namely $(C_q \ell^q\,\frac{\partial^q u}{\partial t^q})_m^n$ can be improved by at least one power of ℓ by extrapolating in time; the other component $(-\frac{1}{2}a\ell h\,\frac{\partial^2 u}{\partial x^2})_m^n$, which is related to the space discretization, will not change. The extrapolating procedure, as in Chapter 2, determines $\underline{U}(t+2\ell)$ in terms of $\underline{U}(t)$: it first calculates $\underline{U}^{(1)} = \underline{U}^{(1)}(t+2\ell)$ by writing equation (4.17), in which the matrix exponential function has been replaced by an appropriate Padé approximant, over two single time steps, and then calculates $\underline{U}^{(2)} = \underline{U}^{(2)}(t+2\ell)$ by writing (4.17) over a double time step. The extrapolated value $\underline{U}^{(E)} = \underline{U}^{(E)}(t+2\ell)$ is then found from one of the formulas:

$$\underline{U}^{(E)} = (2^{M+K}\underline{U}^{(1)} - \underline{U}^{(2)})/(2^{M+K}-1) + O(\ell^{M+K+2}) \tag{4.29}$$

for $M \neq K$, or

$$\underline{U}^{(E)} = (2^{2m}\,\underline{U}^{(1)} - \underline{U}^{(2)})/(2^{2M} - 1) + O(\ell^{2M+3}) \tag{4.30}$$

for $M = K$.

The extrapolating formulas are contained in Table 2.3. The term $(-\frac{1}{2}a\ell h\ \partial^2 u/\partial x^2)_m^n$ will still be present in the principal part of the local truncation error of the extrapolated form of each finite difference method. There will also be a term of the form $(E_s \ell^s \partial^s u/\partial t^s)_m^n$ ($s = M+K+2$ for $M \neq K$, $s = 2M+3$ for $M = K$). The constants E_s are also contained in Table 2.3. Associated with each extrapolated method is the <u>amplification symbol</u>

$$S_{M,K}(\theta) = A(P_K(\theta)/Q_M(\theta))^2 - (A-1)\,P_K(2\theta)/Q_M(2\theta)\,, \tag{4.31}$$

where $\theta = a\ell\lambda$, λ an eigenvalue of C (actually, the eigenvalues of

the matrix C are all equal to $1/h$, but this will not be so in latter sections where the general form (4.31) will be needed). In (4.31) $A = 2^{M+K}/(2^{M+K}-1)$ and $P_K(\theta)$, $Q_M(\theta)$ are the polynomials of degree K,M respectively, which define the (M,K) Padé approximant $R_{M,K}(\theta) = P_K(\theta)/Q_M(\theta)$.

The extrapolated form of a method is A_0-stable, or stable in the conventional sense of perturbations in the initial conditions not being magnified as $t \to \infty$, if $|S_{M,K}(\theta)| \leq 1$. The extrapolated forms of the methods discussed in this section are therefore unconditionally stable, except the extrapolated form of the method based on the (1,1) Padé approximant which is stable only for $0 < ap \leq 6 + 4\sqrt{3}$, where $p = \ell/h$.

4.4 A higher order space replacement

Whereas extrapolation in time does, indeed, bring about some improvement in the principal parts of the local truncation errors of all finite difference schemes resulting from (4.17), the improvement of any one method may not be sufficient to justify its use for larger values of h. This is because the component of the local truncation error given by $(-\frac{1}{2}a\ell h\ \partial^2u/\partial x^2)_m^n$ is still present and tends to overshadow any improvement brought about by extrapolating in time.

This difficulty is partially removed by introducing a second order backward difference approximant to $\partial u/\partial x$ at the mesh points $(mh,n\ell)$ for $m = 2,3,\ldots,N$ and $n = 0,1,\ldots,$ whilst retaining the first order approximant (4.12) at the points $(h,n\ell)$ adjacent to the boundary $x = 0$. This mixture of approximants to $\partial u/\partial x$ is justified in the theorems of Gustafsson (1975), so that, provided a Padé approximant is chosen which would lead to unconditional stability if the lower order approximant (4.12) were used to every mesh point, the scheme resulting from the use of the mixture of approximants to $\partial u/\partial x$ will also be unconditionally stable and will have the convergence rate of the more accurate interior approximant (see also Oliger (1974)).

The schemes resulting from the use of different backward difference replacements of $\partial u/\partial x$ can all be used explicitly though some have the stability properties of implicit schemes.

Consider, then, the second order replacement

$$\frac{\partial u}{\partial x} = \{u(x-2h,t) - 4u(x-h,t) + 3u(x,t)\}/2h + O(h^2). \tag{4.32}$$

This replacement uses three mesh points at any time $t = n\ell$, so that it can only be used at mesh points $(mh,n\ell)$ for which $m = 2,3...$ and $n = 0,1,...$. At the mesh points $(h,n\ell)$, equation (4.12), written conveniently as

$$\frac{\partial u}{\partial x} = \{2u(x,t) - 2u(x-h,t)\}/2h + O(h) \tag{4.33}$$

is retained.

Applying (4.1) with (4.33) or (4.32), as appropriate, to the N mesh points at time level $t = n\ell$ leads to the first order system

$$\frac{d\underline{U}(t)}{dt} = -\frac{1}{2}\, a\, D\underline{U}(t) + \frac{1}{2}\, a\underline{d}_t \,. \tag{4.34}$$

In (4.34) D is the matrix of order N given by

$$hD = \begin{bmatrix} 2 & & & & 0 \\ -4 & 3 & & & \\ 1 & -4 & 3 & & \\ & \ddots & \ddots & \ddots & \\ 0 & & 1 & -4 & 3 \end{bmatrix} \tag{4.35}$$

and $\underline{d}_t$ is the N-component vector given by

$$h\underline{d}_t = [\,2v_t,\ -v_t, 0, \ldots, 0]^T \,. \tag{4.36}$$

One eigenvalue of the matrix D has the value $2/h$ and the other $N-1$ eigenvalues have the value $3/h$.

The solution of (4.34) with (4.2) is

$$\underline{U}(t) = D^{-1}\underline{d}_t + \exp\left(-\tfrac{1}{2}tD\right)\{\underline{g} - D^{-1}\underline{d}_t\} . \tag{4.37}$$

and it is easy to show that (4.38) satisfies the recurrence relation

$$\underline{U}(t+\ell) = D^{-1}\underline{d}_t + \exp\left(-\tfrac{1}{2}a\ell D\right)\{\underline{U}(t) - D^{-1}\underline{d}_t\} . \tag{4.38}$$

Only schemes based on Padé approximants for which $m \geq k$ will be considered. The amplification symbols of the extrapolated forms of such schemes may be obtained from (4.31) with $\theta = \frac{1}{2}a\ell\lambda$, λ now an eigenvalue of D.

Using the $(1,0)$ Padé approximant in (4.38), gives the L_0-stable scheme

$$(I + \tfrac{1}{2}a\ell D)\,\underline{U}(t+\ell) - \tfrac{1}{2}a\ell\underline{d}_{t+\ell} = \underline{U}(t) \tag{4.39}$$

which, from Table 2.1, is seen to be first order accurate in time. The principal part of the local truncation error at the mesh point $(h,n\ell)$ is, from (4.18),

$$\left(-\tfrac{1}{2}a\ell h\,\frac{\partial^2 u}{\partial x^2} - \tfrac{1}{2}\ell^2\,\frac{\partial^2 u}{\partial t^2}\right)_1^n$$

and at the mesh point $(mh,n\ell)$ is

$$\left(-\tfrac{1}{3}a\ell h^2\,\frac{\partial^3 u}{\partial x^3} - \tfrac{1}{2}\ell^2\,\frac{\partial^2 u}{\partial t^2}\right)_m^n$$

for $m = 2,3,\ldots,N$ and $n = 0,1,2,\ldots$. In view of its favourable stability properties, it is worthwhile to extrapolate (4.39) using (4.29). The extrapolated form can be used explicitly and is L_0-stable; its local truncation error is

$$\left(-\tfrac{1}{2}a\ell h\,\frac{\partial^2 u}{\partial x^2} + \tfrac{4}{3}\ell^3\,\frac{\partial^3 u}{\partial t^3}\right)_1^n$$

at the mesh point $(h,n\ell)$ adjacent to the boundary, and

$$\left(-\tfrac{1}{3}a\ell h^2\,\frac{\partial^3 u}{\partial x^3} + \tfrac{4}{3}\ell^3\,\frac{\partial^3 u}{\partial t^3}\right)_m^n$$

at the interior mesh points $(mh,n\ell)$ where $m = 2,\ldots,N$ and $n = 0,1,2,\ldots$.

Some improvement in accuracy may be achieved by using the $(1,1)$ Padé approximant to the matrix exponential function in (4.38) to give

$$(4.40)\qquad (I + \tfrac{1}{4}a\ell D)\,\underline{U}(t+\ell) - \tfrac{1}{4}a\ell \underline{d}_{t+\ell} = (I - \tfrac{1}{4}a\ell D)\,\underline{U}(t) + \tfrac{1}{4}a\ell \underline{d}_{t}$$

which is second order accurate in time and which is A_0-stable. The principal part of the local truncation error of (4.40) at the mesh points $(h,n\ell)$ is

$$(4.41)\qquad (-\tfrac{1}{2}a\ell h\,\frac{\partial^2 u}{\partial x^2} - \frac{1}{12}\ell^3\,\frac{\partial^3 u}{\partial t^3})_1^n$$

and at the mesh points $(mh,n\ell)$ away from the boundary is

$$(4.42)\qquad (-\tfrac{1}{3}a\ell h^2\,\frac{\partial^3 u}{\partial x^3} - \frac{1}{12}\ell^3\,\frac{\partial^3 u}{\partial t^3})_m^n$$

for $m = 2,3,\ldots,N$ and $n = 0,1,2,\ldots$. The expression in (4.41), (4.42), may be improved by extrapolation but, as noted in section (4.3), the extrapolated form is not A_0-stable.

4.5 Higher order time replacements

In view of the fact that all finite difference schemes resulting from the use of backward difference replacements of the space derivative in (4.1) can be used explicitly, it is worthwhile using higher entries from the Padé Table to approximate $\exp(-\frac{1}{2}a\ell D)$ in (4.38), even though it will be necessary to square the matrix D.

Using, first of all, the $(2,0)$ Padé approximant in (4.38) gives

$$(4.43)\qquad (I + \tfrac{1}{2}a\ell D + \tfrac{1}{8}a^2\ell^2 D^2)\,\underline{U}(t+\ell) - (\tfrac{1}{2}a\ell I + \tfrac{1}{8}a^2\ell^2 D)\underline{d}_{t+\ell} = \underline{U}(t)\ .$$

The matrix D^2 is given by

$$(4.44)\qquad h^2D^2 = \begin{bmatrix} 4 & & & & & & \\ -20 & 9 & & & & & \\ 21 & -24 & 9 & & & & \\ -8 & 22 & -24 & 9 & & & \\ 1 & -8 & 22 & -24 & 9 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ 0 & & 1 & -8 & 22 & -24 & 9 \end{bmatrix}$$

and has one eigenvalue equal to $4/h^2$ and $N-1$ eigenvalues equal to $9/h^2$.

Applying (4.43) to the mesh point $(jh,n\ell)$ in R leads to a linear system which may be written in matrix form as

$$E\underline{U}(t+\ell) = \underline{\phi}^n \ . \tag{4.45}$$

The matrix E is of order N and has the lower triangular form

$$E = \begin{bmatrix} e_1 & & & & & & \\ e_2 & e_4 & & & & & \\ e_3 & e_5 & e_4 & & & 0 & \\ e_7 & e_6 & e_5 & e_4 & & & \\ e_8 & e_7 & e_6 & e_5 & e_4 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ 0 & & e_8 & e_7 & e_6 & e_5 & e_4 \end{bmatrix} \tag{4.46}$$

where

$$e_1 = 1+ap+\frac{1}{2}a^2p^2 \ , \quad e_2 = 2ap- \frac{5}{2}a^2p^2 \ , \quad e_3 = \frac{1}{2}ap+\frac{21}{8}a^2p^2$$

$$e_4 = 1+\frac{3}{2}ap+\frac{9}{8}a^2p^2 \ , \quad e_5 = 2ap - 3a^2p^2 \ , \quad e_6 = \frac{1}{2}ap+\frac{11}{4}a^2p^2 \tag{4.47}$$

$$e_7 = -a^2p^2 \ , \quad e_8 = \frac{1}{8}a^2p^2 \ ,$$

and the vector $\underline{\phi}^n = (\phi_1^n \ , \ \phi_2^n \ , \ \ldots, \phi_N^n)^T$ has elements

$$\phi_1^n = U_1^n + ap(1+\frac{1}{2}ap)v_{t+\ell} \ , \ \phi_2^n = U_2^n - ap \ (\frac{1}{2} + \frac{11}{8}ap)v_{t+\ell} \ ,$$

$$\phi_3^n = U_3^n + \frac{3}{4}a^2p^2v_{t+\ell} \ , \ \phi_4^n = U_4^n - \frac{1}{8}a^2p^2v_{t+\ell} \ ,$$

$$\phi_j^n = U_j^n \quad (j = 5,6,\ldots,N) \ . \tag{4.48}$$

The finite difference scheme based on the $(2,0)$ Padé approximant is L_0-stable; the principal part of its local truncation error is

$$(- \frac{1}{3}a\ell h^2 \ \frac{\partial^3 u}{\partial x^3} + \frac{1}{6}\ell^3 \ \frac{\partial^3 u}{\partial t^3})_j^n \ , \quad j = 4,5,\ldots,N \tag{4.49}$$

which, on extrapolation, becomes

$$(- \frac{1}{3}a\ell h^2 \ \frac{\partial^3 u}{\partial x^3} - \frac{1}{3}\ell^4 \ \frac{\partial^4 u}{\partial t^4})_j^n \ , \quad j = 4,5,\ldots,N \ . \tag{4.50}$$

Expressions (4.49), (4.50) show that the loss of accuracy at the mesh points $(h,n\ell)$, $n = 0,1,\ldots$, experienced by the methods based on the lower order Padé approximants, has spread to the mesh points $(2h,n\ell)$, $(3h,n\ell)$. This is not a grave problem, however, for a space discretization involving a large value of N. Furthermore, the constant $C_3 = \frac{1}{6}$ in (4.49), is greater in modulus than its counterpart in (4.42) which relates to the A_0-stable method (4.40).

These observations indicate that the A_0-stable method (4.40) is to be preferred to the L_0-stable method (4.43). This is not so in the case of second order parabolic equations (Lawson and Morris (1978) and Chapter 3), for then the equivalent method based on the (1,1) Padé approximant (the Crank-Nicolson method), also requires a restriction on ℓ to ensure the decay of oscillations in U as $t \to \infty$.

Turning, next, to the (2,1) Padé approximant, (4.38) becomes

$$(4.51)\qquad (I+\tfrac{1}{3}a\ell D+\tfrac{1}{24}a^2\ell^2D^2)\,\underline{U}(t+\ell) - (\tfrac{1}{3}a\ell I+\tfrac{1}{24}a^2\ell^2 D)\underline{d}_{t+\ell}$$
$$= (I-\tfrac{1}{6}a\ell D)\,\underline{U}(t)+\tfrac{1}{6}a\ell\underline{d}_t\ .$$

Applying (4.51) to the mesh points $(jh,n\ell)$ requires the solution vector $\underline{U}(t+\ell)$ to be determined implicitly from a linear system of the form (4.45). The matrix E is still of the form (4.46) but its non-zero elements are now given by

$$(4.52)\qquad \begin{aligned} e_1 &= 1+\tfrac{2}{3}ap+\tfrac{1}{6}a^2p^2\ , \quad e_2 = -\tfrac{4}{3}ap-\tfrac{5}{6}a^2p^2\ , \quad e_3 = \tfrac{1}{3}ap+\tfrac{7}{8}a^2p^2\ , \\ e_4 &= 1+ap+\tfrac{3}{8}a^2p^2\ , \quad e_5 = -\tfrac{4}{3}ap-a^2p^2\ , \quad e_6 = \tfrac{1}{3}ap+\tfrac{11}{12}a^2p^2\ , \\ e_7 &= -\tfrac{1}{3}a^2p^2\ , \quad e_8 = \tfrac{1}{24}a^2p^2\ , \end{aligned}$$

while the elements of $\underline{\phi}^n$ are given by

$$(4.53)\qquad \begin{aligned} \phi_1^n &= (1-\tfrac{1}{3}ap)U_1^n + ap(\tfrac{2}{3}+\tfrac{1}{6}ap)v_{t+\ell}+\tfrac{1}{3}apv_t\ , \\ \phi_2^n &= \tfrac{2}{3}apU_1^n+(1-\tfrac{1}{2}ap)U_2^n - ap(\tfrac{1}{3}+\tfrac{11}{24}ap)v_{t+\ell}-\tfrac{1}{6}apv_t\ , \end{aligned}$$

$$\phi_3^n = -\frac{1}{6}ap\,U_1^n + \frac{2}{3}ap\,U_2^n + (1-\frac{1}{2}ap)U_3^n + \frac{1}{4}a^2p^2v_{t+\ell} ,$$

$$\phi_4^n = -\frac{1}{6}ap\,U_2^n + \frac{2}{3}ap\,U_3^n + (1-\frac{1}{2}ap)U_4^n - \frac{1}{24}a^2p^2v_{t+\ell} ,$$

$$\phi_j^n = -\frac{1}{6}ap\,U_{j-2}^n + \frac{2}{3}ap\,U_{j-1}^n + (1-\frac{1}{2}ap)U_j^n \quad , \quad j = 5,\ldots,N \quad .$$

The vector $\underline{U}(t+\ell)$ is found from (4.45) using forward substitution. The finite difference scheme based on the (2,1) Padé approximant, is L_0-stable ; the principal part of its local truncation error is

$$(4.54) \qquad (-\frac{1}{3}a\ell h^2\frac{\partial^3 u}{\partial x^3} + \frac{1}{72}\ell^4\frac{\partial^4 u}{\partial t^4})_j^n \quad , \quad j = 4,\ldots,N$$

which, following extrapolation using (4.29), becomes

$$(4.55) \qquad (-\frac{1}{3}a\ell h^2\frac{\partial^3 u}{\partial x^3} - \frac{8}{945}\ell^5\frac{\partial^5 u}{\partial t^5})_j^n \quad , \quad j = 4,\ldots,N \quad .$$

Expressions (4.54), (4.55) do indicate an improvement on (4.42), (4.50) and justify the use of (4.51) even though the three points near the boundary suffer greater error at each time step than the remaining N-3 points away from the boundary $x = 0$.

The final method to be considered is that obtained by replacing the exponential matrix function with its (2,2) Padé approximant in (4.38). The recurrence relation becomes

$$(4.56) \qquad (I+\frac{1}{4}a\ell D+\frac{1}{48}a^2\ell^2D^2)\underline{U}(t+\ell) - (\frac{1}{4}a\ell I+\frac{1}{48}a^2\ell^2D^2)\underline{d}_{t+\ell}$$
$$= (I-\frac{1}{4}a\ell D+\frac{1}{48}a^2\ell^2D^2)\underline{U}(t) + (\frac{1}{4}a\ell I-\frac{1}{48}a^2\ell^2D^2)\underline{d}_t ,$$

which gives rise to an A_0-stable method. Applying (4.56) to each mesh point $(jh,n\ell)$, $j = 1,2,\ldots,N$, at time $t = n\ell$, $n = 0,1,\ldots$, leads to the solution vector $\underline{U}(t+\ell)$ at the advanced time $t = (n+1)\ell$ being determined from a system of the form (4.45). The non-zero elements of E are arranged as in (4.46) and have the values

$$(4.57) \qquad e_1 = 1+\frac{1}{2}ap+\frac{1}{12}a^2p^2 \ , \ e_2 = -ap-\frac{5}{12}a^2p^2 \ , \ e_3 = \frac{1}{4}ap+\frac{7}{16}a^2p^2 \ ,$$
$$e_4 = 1+\frac{3}{4}ap+\frac{3}{16}a^2p^2 \ , \ e_5 = -ap-\frac{1}{2}a^2p^2 \ , \ e_6 = \frac{1}{4}ap+\frac{11}{24}a^2p^2 \ ,$$

$$e_7 = -\frac{1}{6}a^2p^2 \quad , \qquad e_8 = \frac{1}{48}a^2p^2 \quad .$$

The elements of the vector $\underline{\phi}^n$ are

$$\phi_1^n = (1-\tfrac{1}{2}ap+\tfrac{1}{12}a^2p^2)U_1^n+ap(\tfrac{1}{2}+\tfrac{1}{12}ap)v_{t+\ell}+ap(\tfrac{1}{2}-\tfrac{1}{12}ap)v_t \quad ,$$

$$\phi_2^n = ap(1-\tfrac{5}{12}ap)U_1^n+(1-\tfrac{3}{4}ap+\tfrac{3}{16}a^2p^2)U_2^n-ap(\tfrac{1}{4}+\tfrac{11}{48}ap)v_{t+\ell}$$
$$- ap(\tfrac{1}{4}-\tfrac{11}{48}ap)v_t \quad ,$$

$$\phi_3^n = ap(-\tfrac{1}{4}+\tfrac{7}{16}ap)U_1^n+ap(1-\tfrac{1}{2}ap)U_2^n+(1-\tfrac{3}{4}ap+\tfrac{3}{16}a^2p^2)U_3^n$$
$$+ \tfrac{1}{8}a^2p^2v_{t+\ell} - \tfrac{1}{8}a^2p^2v_t \quad ,$$

$$\phi_4^n = -\tfrac{1}{6}a^2p^2U_1^n+ap(-\tfrac{1}{4}+\tfrac{11}{24}ap)U_2^n+ap(1-\tfrac{1}{2}ap)U_3^n+(1-\tfrac{3}{4}ap+\tfrac{3}{16}a^2p^2)U_4^n$$
$$- \tfrac{1}{48}a^2p^2v_{t+\ell} + \tfrac{1}{48}a^2p^2v_t \quad ,$$

$$\phi_j^n = \tfrac{1}{48}a^2p^2U_{j-4}^n+ap(-\tfrac{1}{4}+\tfrac{11}{24}ap)U_{j-3}^n+ap(1-\tfrac{1}{2}ap)U_{j-2}^n+(1-\tfrac{3}{4}ap+\tfrac{3}{16}a^2p^2)U_{j-1}^n$$
$$+ (1-\tfrac{3}{4}ap+\tfrac{3}{16}a^2p^2)U_j^n \; ; \quad j = 5,\ldots,N \quad .$$

The local truncation error of (4.56) for $j = 4,\ldots,N$ and $n = 0,1,\ldots$ is

(4.59) $$(-\frac{1}{3}a\ell h^2\frac{\partial^3 u}{\partial x^3}+\frac{1}{720}\ell^5\frac{\partial^5 u}{\partial x^5})_j^n \quad ,$$

the time component in which may be improved by extrapolating, using (4.30), to give

(4.60) $$(-\frac{1}{3}a\ell h^2\frac{\partial^3 u}{\partial x^3}-\frac{1}{1890}\ell^7\frac{\partial^7 u}{\partial t^7})_j^n$$

In the event of an even higher order approximant to the space derivative being used in (4.1), instead of (4.32), the elegant methods of Gourlay and Morris (1980) for improving the accuracy in time of numerical methods for parabolic equations, can be used with the relations (4.17), (4.38).

Using a more accurate space replacement requires the matrix D to have increased band width. This band width would be increased still further

on squaring D and more than three points near the boundary would suffer loss of accuracy when solving (4.45) using the (2,0), (2,1), (2,2) Padé approximants, though stability would not be affected. It may, therefore be advisable to use the techniques of Gourlay and Morris (1980) with a space replacement (4.32), but the methods developed in this section and in sections (4.3), (4.4) can be implemented more quickly and are to be preferred for use with (4.32).

4.6 Numerical experiments

To discuss the behaviour of the methods developed in sections 4.3, 4.4, 4.5, the methods based on the (1,1), (2,0), (2,1), (2,2) Padé approximants without extrapolation,are tested on a number of problems from the literature. When these four Padé approximants are tested in conjunction with the matrix C given by (4.14), they will be named C11, C20, C21, C22, respectively, and when used in conjunction with the matrix D they will be named D11, D20, D21, D22, respectively.

The boundedness of the solution and the build-up of error may be examined with reference to two norms, as in Oliger (1974). Let $z_j^n = u(jh,n\ell) - U_j^n$, with $j = 0,1,\ldots N$ and $n = 0,1,\ldots$, so that $\underline{z}^n$ is the vector of such errors and has $N+1$ elements, and let $\underline{V}^n = (U_0^n, U_1^n, \ldots, U_N^n)^T$ be the vector (of order $N+1$) of solutions, including the boundary condition, at time $t = n\ell$. The norms are defined by

$$\| \underline{z}^n \|_\infty = \max_j | z_j^n | , \quad \| \underline{z}^n \|_2^2 = h\sum_{j=0}^{N} | z_j^n |^2, \quad \| \underline{V}^n \|_2^2 = h \sum_{j=0}^{N} |U_j^n|^2 .$$

The methods (4.9) and (4.10) based on the central difference approximation are also tested on the first two problems and their behaviour is shown graphically in Figures 4.1 - 4.4. The differential equation on which the methods are tested is

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

the initial and boundary conditions being different for each problem but $a = 1$ in each problem.

Problem 4.1 (Oliger (1974)).

Here the initial conditions are taken to be

$$g(x) = \sin 2k\pi x \quad ; \quad x \geq 0$$

and the boundary conditions to be

$$v(t) = -\sin 2k\pi t \quad ; \quad t > 0$$

where k is positive integer. The theoretical solution of this problem is

$$u(x,t) = \sin 2k\pi(x-t)$$

and the numerical solution will be calculated for $0 < x \leq 1$. The integer k gives the number of complete waves in the interval $0 \leq x \leq 1$. The scheme (4.9) produced results depicted in Fig. 4.1 at time $t = 1.0$ with $h = \frac{1}{80}$, $\ell = \frac{1}{20}$, $p = 4.0$ and $k = 2$. The solution computed using the Crank-Nicolson type scheme (4.10) at time $t = 1.0$ with $h = \frac{1}{80}$, $\ell = \frac{1}{20}$, $p = 4.0$ and $k = 2$, is shown in Fig. 4.2.

The solution was computed with $h = \frac{1}{640}$, $\ell = \frac{1}{80}$, $p = 8$ and $k = 2$, using the methods discussed in sections 4.3, 4.4, 4.5; the values of $\| \underline{V} \|_2$, $\| \underline{Z} \|_2$, $\| \underline{Z} \|_\infty$ at time $t = 0.5, 1.0, 2.0$ and 4.0 are given in Table 4.1. Choosing this small value of h has the effect of lessening the emphasis of the components $-\frac{1}{2}a\ell h\, \partial^2 u/\partial x^2$ and $-\frac{1}{3}a\ell h^2\, \partial^3 u/\partial x^3$ when the backward difference approximations (4.33) and (4.32) are used to replace the spatial derivative. The increased number of mesh points at each time level can be appreciably offset by using a large value of ℓ, and consequently of p. In the paper by Oliger (1974), for example, p was given the value $\frac{1}{4}$ compared with the value 8 in the present experiment.

Visual analysis of Table 4.1, and comparison with Table 3.1 in Oliger (1974), shows that errors for all eight formulations involving the matrices C and D show very little increase in magnitude after time

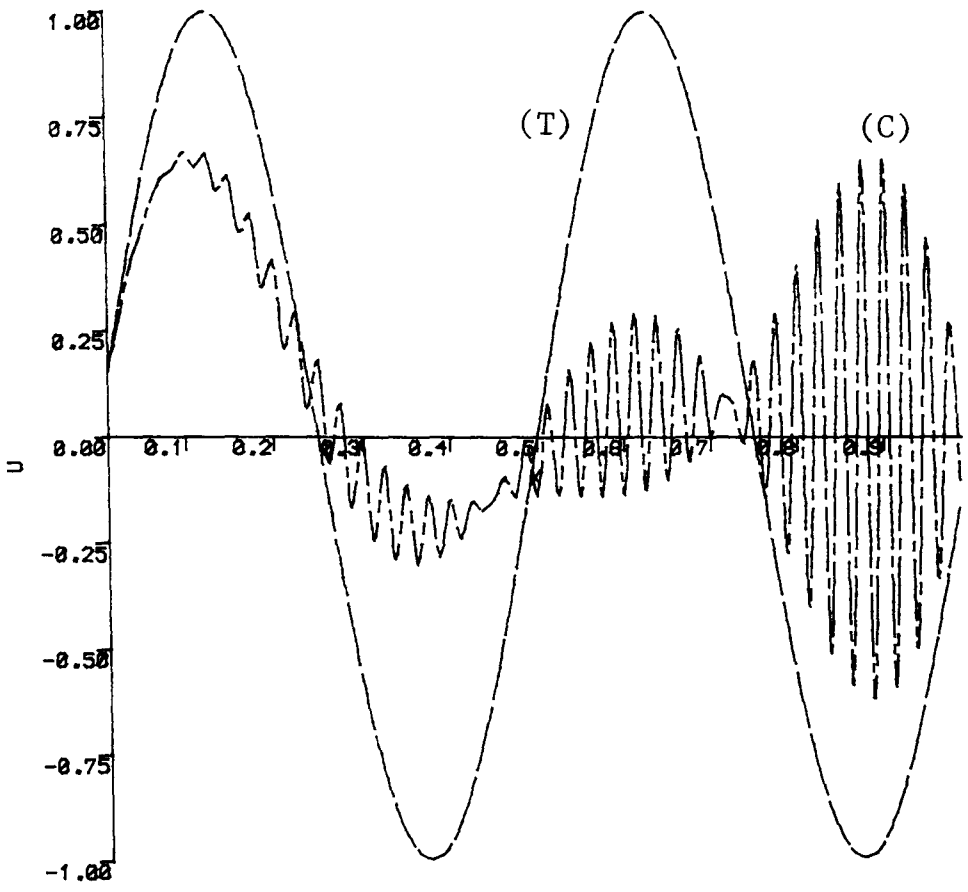


Figure 4.1: Numerical results at time t = 1.0 for Problem 4.1 using the backward difference scheme (4.9) with h= 1/80, ℓ = 0.05, p = 4.
Theoretical solution (T); computed solution (C).

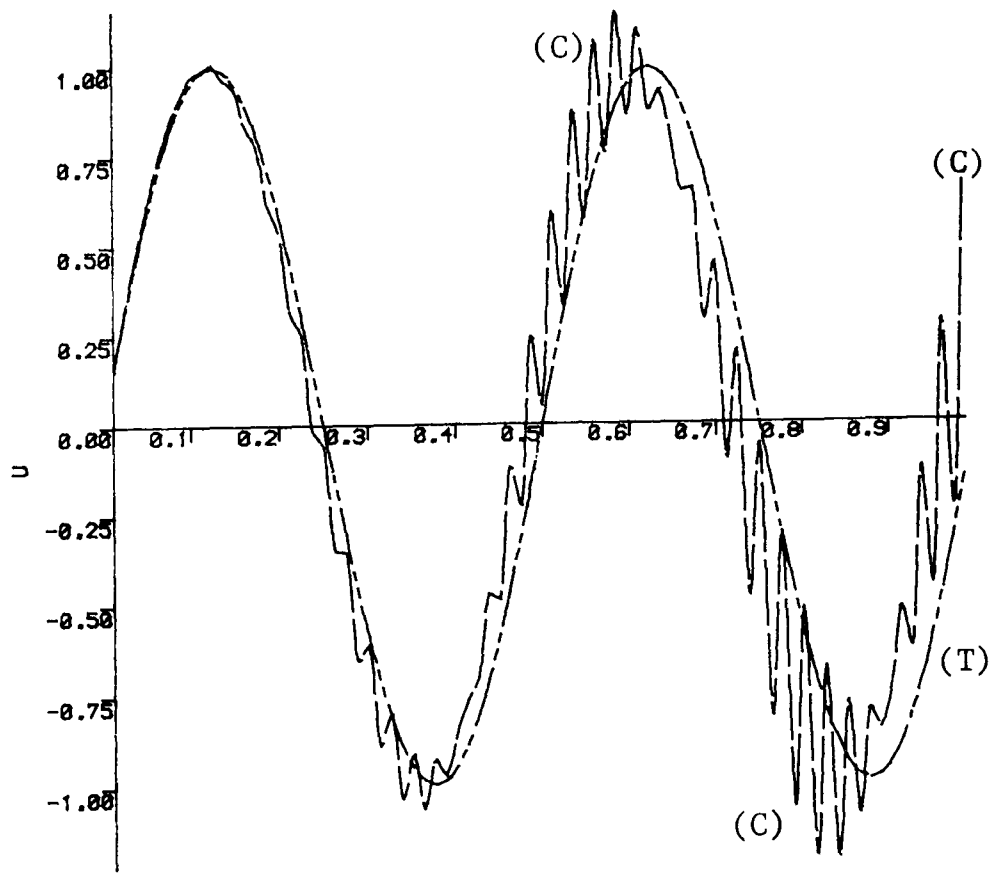


Figure 4.2: Numerical results at time t = 1.0 for Problem 4.1 using the Crank-Nicolson type scheme (4.10) with h = 1/80, ℓ = 0.05, p = 4.
Theoretical solution(T); computed solution(C).

$t = 1.0$. That is to say. the errors reach their maximum values very quickly, there being very little accumulation of errors after time $t = 1.0$. This observation contrasts with the results of Table 3.1 in Oliger (1974) where the errors, generally, show a gradual growth as time increases. The stagnation of errors experienced using these two-time level methods make them suitable for use with large values of t. The maximum error of each method was seen to be in keeping with the truncation errors given in sections 4.3, 4.4, 4.5. The methods are also seen to behave smoothly with the theoretical solution. The methods based on the $(2,1)$ and $(2,2)$ Padé approximants showed the greatest improvement when used with the matrix D (for any value of t), the corresponding improvements in the performance of the methods based on the $(1,1)$ and $(2,0)$ Padé approximants being less pronounced.

Problem 4.2 (Abarbanel et al (1975))

The boundary conditions and the initial conditions for this problem are the same as for Problem 4.1. The parameter k is given the value 4 and the solution computed with $h = 1/640$, $\ell = 1/80$, $p = 8$; the numerical results at time $t = 10.0$ are given in Table 4.2. The corresponding results for $k = 4$ are given in Table 4 of Abarbanel et al (1975) where the ratio p was given the value 0.9. In their Table 4 Abarbanel et al (1975) compare their results with earlier work by a number of authors Boris and Book (1973), Kreiss and Oliger (1972), Oliger (1974), and Richtmyer (1963). The results of this chapter show that the methods developed are very competitive with all methods tested in Abarbanel et al (1975) for $k = 4$. The growth of errors as a result of increasing the wave frequency was not pronounced as any of the methods tested in Abarbanel et al (1975). Allowing a factor of 3 for the faster CDC 7600 computer over the CDC 6600 computer used by Abarbanel et al (1975), the CPU times quoted in Table 4.2 are generally superior to the figures quoted in Abarbanel et al (1975). This observation is strengthened when it is further noted that the CPU times in Table 4.2 include the time taken to

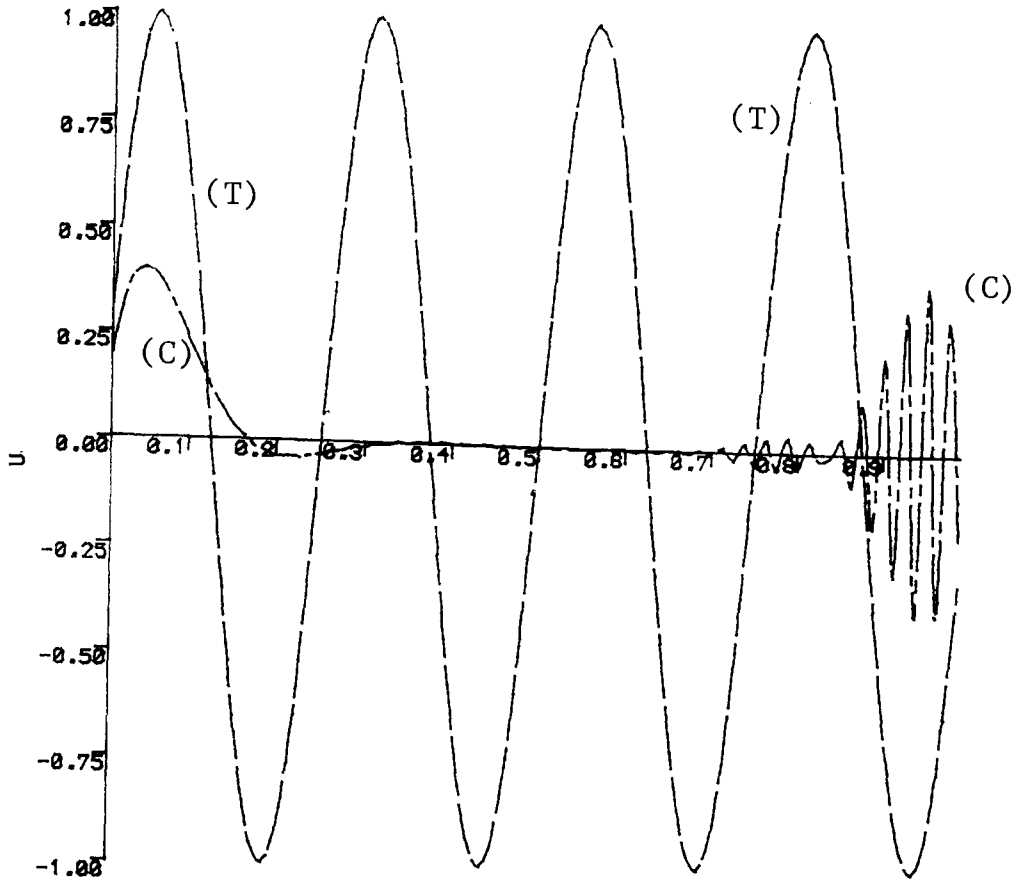


Figure 4.3: Numerical results at time t = 10.0 for Problem 4.2 using the backward difference scheme (4.9) with h = 1/80, ℓ = 0.05, p = 4.
Theoretical solution (T); computed solution (C).

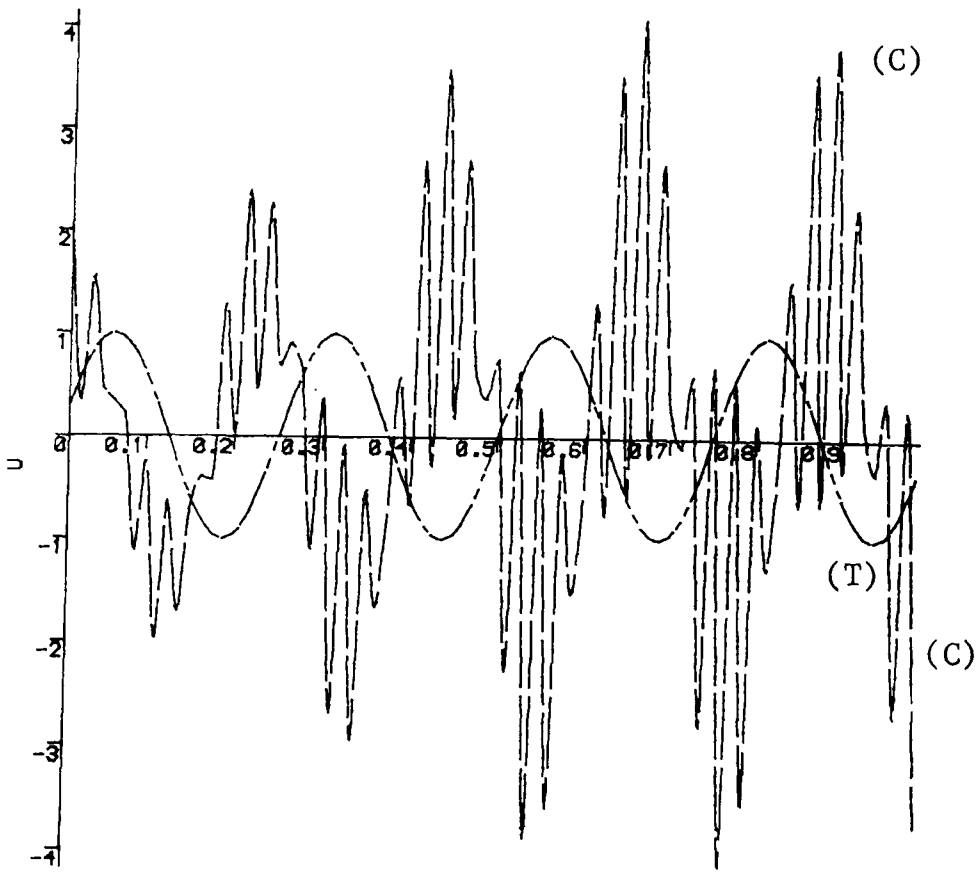


Figure 4.4: Numerical results at time t = 10.0 for Problem 4.2 using the Crank-Nicolson type scheme (4.10) with h = 1/80, ℓ = 0.05, p = 4.
Theoretical solution(T);computed solution(C).

compute $\| \underline{z} \|_\infty$ by 640 comparison statements in the computer program. It is confirmed again that the use of a small value of h in the methods which have higher accuracy in time, produces accuracy as high as do those methods, tested in Oliger (1974), Abarbanel et al (1975) with a larger value of h which have $0(h^4)$ error in space. Solutions computed using (4.9) and (4.10) for problem 4.2 are also shown in Figures 4.3 and 4.4 respectively.

Problem 4.3 (Khaliq and Twizell (1982))

The boundary condition for this problem is

$$u(0,t) = t \quad ; \quad t > 0$$

and the initial condition is

$$u(x,0) = 1 + x \quad ; \quad x \geq 0 \ ,$$

The theoretical solution of the problem is

$$u(x,t) = 1 + x-t \quad ; \quad x \geq t \ ,$$

$$u(x,t) = t - x \quad ; \quad x < t$$

so that there exists a discontinuity in the solution across the line $t = x$ in the (x,t) plane.

Problem 4.4 (Khaliq and Twizell (1982))

Here the initial condition is

$$u(x,0) = \exp(x) \quad ; \quad x \geq 0$$

and the boundary condition is

$$u(0,t) = \exp(t) \quad ; \quad t > 0 \ .$$

The theoretical solution of the problem is

$$u(x,t) = \exp(x-t) \quad ; \quad x \geq t \ ,$$

$$u(x,t) = \exp\ (t-x) \ ; \quad x < t$$

so that there exist discontinuities in the first derivatives across the line $t = x$ in the (x,t) plane.

Problems 4.3 and 4.4 were tested with $h = 1/80$, $\ell = 1/20$, $p = 4$ and the results are given at time $t = 1.0$ in Tables 4.3, 4.4 respectively. It is noted again that the methods based on the (2,1) and (2,2) Padé approximants, showed greater improvements than the improvements shown by the methods based on the (1,1) and (2,0) Padé approximants. Using the higher order space approximation, the highest accuracy was achieved by method D22 followed, in succession, by D21, D11, D20; this is in keeping with the local truncation errors of these methods and with the numerical results obtained for Problems 4.1 and 4.2. It was also found, as the computation proceeds, that, away from the boundary, the greatest errors were at those mesh points close to the line $t = x$ across which there were discontinuities.

Problem 4.5

The boundary condition for this problem is taken to be

$$u(0,t) = \exp(-t) \quad ; \quad t > 0$$

and the initial condition to be

$$u(x,0) = \exp(x) \quad ; \quad 0 \leq x \leq 1 \ .$$

The theoretical solution of the problem is

$$u(x,t) = \exp(x-t)$$

which decays as time increases. The problem was run with $h = 1/80$, $\ell = 1/20$ and $p = 4$; the numerical results at time $t = 2,4,8,10$ are given in Table 4.5.

The errors were found to behave in much the same way as in the other problems; that is, using the higher order space approximant, produced a more noticeable improvement in the methods based on the (2,1), (2,2) Padé approximants than in the other two methods. The two formulations based on the (1,1) Padé approximant, are seen to give good results at time $t = 10.0$, when, for $0 \leq x \leq 1$, the solution lies in the approximate interval $4.540 \times 10^{-5} < u < 1.234 \times 10^{-4}$. This is due to these formulations using fewer mesh points and thus experiencing smaller round off errors.

4.7 Conclusions

Two families of two-time level finite difference schemes, based on Padé approximants to the matrix exponential function, have been developed for the numerical solution of first order hyperbolic partial differential equations with initial and boundary conditions specified.

The oscillatory behaviour of the methods based on the usual central difference replacement of spatial derivative, was discussed. In order to obtain smooth solutions, the space derivative was replaced first of all by the usual first order backward difference approximant at each mesh point at a given time level, and the resulting system of first order ordinary differential equations was solved using the (1,1), (2,0), (2,1), (2,2) Padé approximants. Next, the space derivative at the mesh point adjacent to the boundary, at a given time level, was replaced by the same low order approximant, and by the usual second order backward difference approximant at all other mesh points. The resulting system of ordinary differential equations was solved using the same four Padé approximants.

All four numerical methods of each backward difference family were implicit in nature; those based on the (1,1) and (2,2) Padé approximants were seen to be A_0-stable and those based on the (2,0) and (2,1) Padé approximants were seen to be L_0-stable. The form of the given boundary conditions, however, meant that the backward difference methods were all used explicitly, obviating the need to solve a linear algebraic system. The CPU time for all eight backward difference methods were found to be fast.

The backward difference methods were tested on five problems from the literature; the results obtained were better than other results in the literature, even though the order of the methods, in many cases, was lower. It was found that the lower order (1,1) and (2,0) Padé approximants gave good results when the lower order replacement of the space derivative was used at each mesh point at a given time level, and

that the higher order (2,1) and (2,2) Padé approximants gave their best results when the higher order replacement of the space derivative was used at interior mesh points. This implies that lower order replacements in both space and time, or higher order replacements in both space and time, are most effective; this observation was also made by Abarbanel et al (1975;p.351). For problems with decaying solutions, the two backward difference formulations based on the (1,1) Padé approximant give very good results due to the smaller number of mesh points used, thus reducing round-off errors.

Table 4.1 Numerical results for Problem 4.1 at time t = 0.5,1.0,2.0,4.0

Method	$\|\underline{V}\|_2$	$\|\underline{Z}\|_2$	$\|\underline{Z}\|_\infty$	CPU (sec)	$\|\underline{V}\|_2$	$\|\underline{Z}\|_2$	$\|\underline{Z}\|_\infty$	CPU (sec)
	t = 0.5				t = 1.0			
C11	6.75(-1)	3.55(-2)	6.08(-2)	0.062	6.65(-1)	5.00(-2)	1.07(-1)	0.115
C20	6.70(-1)	5.76(-2)	1.01(-2)	0.070	6.58(-1)	8.55(-2)	1 56(-1)	0.123
C21	6.74(-1)	1.79(-1)	6.01(-2)	0.074	6.64(-1)	2.07(-1)	1.07(-1)	0.137
C22	6.75(-1)	1.71(-1)	5.98(-2)	0.078	6.64(-1)	1.97(-1)	1.04(-1)	0.145
D11	7.07(-1)	7.03(-3)	1.21(-2)	0.084	7.06(-1)	1.00(-2)	2.40(-2)	0.158
D20	7.03(-1)	4.67(-2)	9.11(-2)	0.088	7.00(-1)	7.20(-2)	1.17(-1)	0.169
D21	7.06(-1)	1.31(-2)	2.66(-3)	0.095	7.05(-1)	1.89(-2)	2.70(-2)	0.179
D22	7.06(-1)	1.23(-3)	2.42(-3)	0.119	7.06(-1)	1.75(-3)	2.71(-3)	0.227
	t = 2.0				t = 4.0			
C11	6.65(-1)	5.00(-2)	1.07(-1)	0.218	6.65(-1)	5.00(-2)	1.07(-1)	0.425
C20	6.59(-1)	8.61(-2)	1.56(-1)	0.249	6.59(-1)	8.61(-2)	1.56(-1)	0.487
C21	6.64(-1)	2.07(-1)	1.07(-1)	0.264	6.64(-1)	2.07(-1)	1.07(-1)	0.517
C22	6.64(-1)	1.97(-1)	1.04(-1)	0.279	6.64(-1)	1.97(-1)	1.04(-1)	0.547
D11	7.06(-1)	1.00(-2)	2.93(-2)	0.305	7.06(-1)	1.00(-2)	2.43(-1)	0.600
D20	7.00(-1)	7.30(-2)	1.27(-1)	0.312	7.00(-1)	7.30(-2)	1.27(-1)	0.689
D21	7.05(-1)	1.90(-2)	2.76(-2)	0.347	7.05(-1)	1.90(-2)	2.76(-2)	0.791
D22	7.05(-1)	1.75(-3)	2.71(-3)	0.445	7.06(-1)	1.75(-3)	2.71(-3)	0.877

Table 4.2(a) Numerical results for Problem 4.2 at time t = 10 using previously published methods.

Method	$\|\underline{V}\|_2$	$\|\underline{Z}\|_2$	CPU (sec)
Richtmeyer (p=0.9)	8.48(-1)	5.9(-1)	3.4
Abarbanel *et al* 3 level (p=0.9)	9.95(-1)	1.4(-2)	6.3
Abarbanel *et al* 2 level (p=0.9)	9.97(-1)	8.6(-3)	7.3
SHASTA (p=0.45)	4.55(-1)	3.2(-1)	15.2
Abarbanel *et al* 3 level (p=0.65)	1.00	1.0	1.4
Abarbanel *et al* 2 level (p=0.5)	1.00	6.5(-1)	1.8
Kreiss - Oliger (p=0.25)	1.00	1.2(-1)	3.6
Abarbanel *et al* 3 level (p=0.1)	1.00	2.8(-2)	8.8
Abarbanel et al 2 level (p=0.05)	1.00	5.0(-2)	17.4

Table 4.2 (b) Numerical results for Problem 4.2 at time t = 10

Method	$\|\underline{V}\|_2$	$\|\underline{z}\|_2$	$\|\underline{z}\|_\infty$	CPU (sec)
C11	5.61(-1)	1.86(-1)	4.00(-1)	1.049
C20	5.29(-1)	2.87(-1)	5.73(-1)	1.121
C21	5.54(-1)	3.82(-1)	3.85(-1)	1.278
C22	5.59(-1)	3.72(-1)	3.75(-1)	1.352
D11	7.04(-1)	8.05(-2)	1.94(-1)	1.483
D20	6.58(-1)	2.46(-1)	4.81(-1)	1.590
D21	6.96(-1)	4.17(-1)	6.47(-2)	1.697
D22	7.07(-1)	8.28(-3)	1.48(-2)	2.178

Table 4.3 Numerical results for Problem 4.3 at time t = 1.0

Method	$\|\underline{V}\|_2$	$\|\underline{z}\|_2$	$\|\underline{z}\|_\infty$	CPU (sec)
C11	1.78	2.01(-2)	9 64(-2)	0.007
C20	1.83	5.76(-2)	1.22(-1)	0.007
C21	1.80	1.59(-1)	1.01(-1)	0.008
C22	1.78	1.67(-2)	9.55(-2)	0 008
D11	1.76	1.75(-2)	4.00(-2)	0.008
D20	1.82	4.82(-2)	7.78(-2)	0.008
D21	1.79	1.63(-2)	3.72(-2)	0.009
D22	1.78	4.51(-3)	2.78(-3)	0.010

Table 4.4 Numerical results for Problem 4.4 at time t = 1.0

Method	$\|\underline{V}\|_2$	$\|\underline{z}\|_2$	$\|\underline{z}\|_\infty$	CPU (sec)
C11	5.97(-1)	1.40(-1)	5.76(-1)	0.009
C20	5.98(-1)	2.51(-1)	5.79(-1)	0.010
C21	5.99(-1)	2.38(-1)	5.50(-1)	0.011
C22	5.97(-1)	2.34(-1)	5.62(-1)	0.012
D11	5.83(-1)	9.04(-2)	5.40(-1)	0.012
D20	5.90(-1)	9.78(-2)	5.48(-1)	0.012
D21	5.82(-1)	8.53(-2)	5.43(-1)	0.013
D22	5.79(-1)	8.60(-2)	5.18(-1)	0.016

Table 4.5 : Numerical results for Problem 4.5 .

$\|Z\|_2$

Method	Time	2.0	4.0	8.0	10.0
C11		5.18(-4)	6.96(-5)	1.27(-6)	1.73(-7)
C20		1.80(-2)	2.44(-3)	4.47(-5)	6.61(-6)
C21		7.68(-2)	2.82(-2)	3.82(-3)	1.41(-3)
C22		4.07(-2)	1.50(-2)	2.03(-3)	7.46(-4)
D11		5.56(-4)	7.71(-5)	6.18(-6)	4.26(-6)
D20		1.90(-2)	2.57(-3)	4.71(-5)	6.37(-6)
D21		7.05(-3)	9.54(-4)	1.75(-5)	2.36(-6)
D22		8.80(-4)	1.51(-4)	1.04(-5)	2.74(-6)

$\|Z\|_\infty$

Method	Time	2.0	4.0	8.0	10.0
C11		1.09(-3)	1.48(-4)	2.82(-6)	4.01(-7)
C20		2.66(-2)	3.61(-3)	6.61(-5)	8.94(-6)
C21		8.55(-3)	1.16(-3)	2.11(-5)	2.86(-6)
C22		3.59(-3)	4.83(-4)	8.84(-6)	1.20(-6)
D11		1.20(-3)	1.84(-4)	1.41(-5)	8.94(-6)
D20		2.88(-2)	3.90(-3)	7.13(-5)	9.65(-6)
D21		1.08(-2)	1.45(-3)	2.66(-5)	8.60(-6)
D22		1.68(-3)	3.61(-4)	3.12(-5)	7.96(-6)

CHAPTER 5

SECOND ORDER PERIODIC INITIAL VALUE PROBLEMS

5.1 Introduction

Periodic initial value problems of the form $y'' = f(x,y)$ arise in the theory of orbital mechanics, and in recent years there has been considerable interest in the numerical solution of such problems.

Generally speaking, second order initial value problems can be divided into two distinct classes: (a) problems for which the solution period is known in advance; (b) problems for which this period is unknown initially. Numerov's methods applied to type (a) always stays on the orbit, whereas the Störmer-Cowell methods with step number greater than two spiral inwards. In the terminology of Stiefel and Bettis (1969), the former method is orbitally stable, the latter orbitally unstable. Modified numerical methods have been proposed by Gautschi (1969), Stiefel and Bettis (1969) and Jain et al (1979), which can be used to compute the solution for problems of type (a). For the numerical solution of problems of type (b), it is desirable that the method should be P-stable. Lambert and Watson (1976) have shown that certain linear multistep methods of arbitrary stepnumber possess a periodicity property when the product of steplength and angular frequency lies within the interval of periodicity and these authors developed symmetry conditions under which a linear multistep method possesses a non-vanishing interval of periodicity. However, Lambert and Watson have shown a P-stable linear multistep method cannot have order of accuracy greater than 2. Jain et al (1979) have derived higher order methods and claim that they are P-stable. It is noted that their concept of P-stability is considerably weaker than that given by Lambert and Watson (1976). Higher order P-stable methods are also proposed by Cash (1981) and Chawla (1981) whose methods need three function evaluation at each step. Cash (1981) has tested his fourth and sixth order P-stable methods on numerical examples and

achieved higher accuracy than Lambert and Watson (1976), by using the analytical solution to compute the solution with steplength h at y(h). However, for practical problems, it is usually necessary to use a starting procedure.

In section 5.2 of this chapter, a recurrence relation is developed which yields two-step multiderivative methods, employing Padé approximants to the exponential functions. The definition of P-stability given by Lambert and Watson (1976, p.199), is adapted. The methods are analysed in section 5.3. The interval of periodicity, principal part of local truncation errors and non-zero coefficients for the algorithms yielded by the first sixteen entries of the Padé Table (Appendix I) are given in Appendix III. The two-step multiderivative methods are given in Appendix IV. Fourth and sixth order methods based on the (2,2) and (3,3) Padé approximants are tested for comparison purposes on the problems discussed by Cash (1981), in section 5.4. The methods are analysed in PECE mode and tested on numerical examples discussed by Jain et al (1979) and Shampine and Gordon (1973) in section 5.5. Finally, conclusions are drawn in section 5.6.

5.2 Development of the methods

Consider the second order initial value problem

$$(5.1)\quad \underline{y}''(t) = \underline{f}(t,\underline{y}) \; ; \quad \underline{y}(t_0) = \underline{y}_0 \; , \quad \underline{y}'(t_0) = \underline{y}_0'$$

where $\underline{y} \in E^N$. A particular case of (1) is the linear problem

$$(5.2)\quad \underline{y}''(t) = A\,\underline{y}(t) + \underline{B} \; ; \quad \underline{y}(t_0) = \underline{y}_0 \; , \quad \underline{y}'(t_0) = \underline{y}_0' \; .$$

Equation (5.2) arises in the numerical solution of the simple wave equation $\partial^2 u/\partial t^2 = \partial^2 u/\partial x^2$ when the space derivative is approximated by a finite difference replacement such as

$$(5.3)\quad \frac{\partial^2 u}{\partial x^2} = h^{-2}\{u(x-h,t) - 2u(x,t) + u(x+h,t)\} + O(h^2) \; ,$$

where h is the increment in x arising from the space discretization. This leads to a system of ordinary differential equations of the form (5.2) in which the diagonalizable matrix A has real, negative eigenvalues, and $B = 0$ when the boundary conditions are zero.

It is therefore appropriate to consider the single test equation (Lambert and Watson (1976), Dahlquist (1978))

$$(5.4)\quad y''(t) = -\lambda^2 y(t) \; ; \quad y(t_0) = y_0 \; , \quad y(t_0) = y_0' \; ,$$

where $\lambda, y \in \mathbb{R}$, whose general solution

$$(5.5)\quad y(t) = a\cos\lambda t + b\sin\lambda t$$

is periodic with period $2\pi/\lambda$ for all a,b other than the trivial case $a = b = 0$.

The general solution (5.5) may be written in the alternate form

$$(5.6)\quad y(t) = a\exp(i\lambda t) + b\exp(-i\lambda t) \; , \; i = +\sqrt{-1}$$

which becomes

$$(5.7)\quad y(t) = -\tfrac{1}{2}i(iy_0 + \lambda^{-1}y_0')\exp\{i\lambda(t-t_0)\} - \tfrac{1}{2}i(iy_0 - \lambda^{-1}y_0')\exp\{-i\lambda(t-t_0)\}$$

when the initial conditions in (5.4) are introduced.

It may now be shown that $y(t)$ given in (5.7), satisfies the recurrence relation

$$y(t+\ell) - \{\exp(i\ell\lambda) + \exp(-i\ell\lambda)\}y(t) + y(t-\ell) = 0 , \tag{5.8}$$

where ℓ is a convenient increment in t. This recurrence relation may be used for $t = t_0 + \ell, \quad t_0 + 2\ell, \ldots$; for $t = t_0$ the initial conditions give $y(t_0) = y_0$ but the value $y(t_0+\ell)$ remains to be determined in terms of y_0 and y_0'. Equation (5.8) leads to a family of multiderivative methods for the solution of (5.1), the higher derivatives being easy to calculate because of the periodic properties of the problem.

Any numerical solution of (5.8) will determine $y(t)$ explicitly or implicitly depending on the approximations to $\exp(\pm i\ell\lambda)$. Using the (m,k) Padé approximant to $\exp(i\ell\lambda)$ of the form

$$\exp(i\ell\lambda) = P_k(i\ell\lambda)/Q_m(i\ell\lambda) + O(\ell^{m+k+1}) , \tag{5.9}$$

where P_k, Q_m are polynomials of degree k,m respectively, defined by

$$P_k(\theta) = 1 + p_1\theta + p_2\theta^2 + \ldots + p_k\theta^k \quad ; \quad P_0(\theta) \equiv 1 \tag{5.10}$$

and

$$Q_m(\theta) = 1 - q_1\theta + q_2\theta^2 - \ldots + (-1)^m q_m\theta^m \quad , \quad Q_0(\theta) \equiv 1 \tag{5.11}$$

with $p_1 > p_2 > \ldots p_k > 0$ and $q_1 > q_2 > \ldots > q_m > 0$ depending on the chosen Padé approximant, equation (5.8) takes the form

$$\begin{aligned} &Q_m(i\ell\lambda)Q_m(-i\ell\lambda)y(t+\ell) - \{Q_m(-i\ell\lambda)P_k(i\ell\lambda) + Q_m(i\ell\lambda)P_k(-i\ell\lambda)\}y(t) \\ &\qquad + Q_m(i\ell\lambda)Q_m(-i\ell\lambda)y(t-\ell) = 0. \end{aligned} \tag{5.12}$$

On substituting for the polynomials P_k, Q_m in (5.12), odd powers of $i\ell\lambda$ vanish and the recurrence relation takes the form

$$\begin{aligned} &\{1 - a_1\ell^2\lambda^2 + a_2\ell^4\lambda^4 - \ldots + (-1)^m a_m \ell^{2m}\lambda^{2m}\}y(t+\ell) \\ &- \{2 - b_1\ell^2\lambda^2 + b_2\ell^4\lambda^4 - \ldots + (-1)^s b_s\ell^{2s}\lambda^{2s}\}y(t) \\ &+ \{1 - a_1\ell^2\lambda^2 + a_2\ell^4\lambda^4 - \ldots + (-1)^m a_m\ell^{2m}\lambda^{2m}\}y(t-\ell) = 0 , \end{aligned} \tag{5.13}$$

where the a_j, b_j clearly depend on the Padé approximant beong used and $s = [\frac{1}{2}(m+k)]$.

For a single equation of the form (5.1), equation (5.13) yields the two-step multiderivative formula

$$\begin{aligned} & y_{n+1} + a_1\ell^2 y''_{n+1} + a_2\ell^4 y^{(iv)}_{n+1} + \ldots\ldots + a_m\ell^{2m} y^{(2m)}_{n+} \\ & \quad = 2y_n + b_1\ell^2 y''_n + b_2\ell^4 y^{(iv)}_n + \ldots\ldots + b_s\ell^{2s} y^{(2s)}_n \\ & \quad\quad - \{y_{n-1} + a_1\ell^2 y''_{n-1} + a_2\ell^4 y^{(iv)}_{n-1} + \ldots\ldots + a_m\ell^{2m} y^{(2m)}_{n-1}\} , \end{aligned} \tag{5.14}$$

$n = 1,2,\ldots,$ which is explicit of $m = 0$ and implicit of $m \neq 0$. It is assumed that $y(t)$ is sufficiently often differentiable. In (5.14), $y_j \equiv y(t_j) = y(t_0+j\ell)$, where $j = 0,1,2,\ldots$; the non-zero coefficients of (5.14) for the algorithms yielded by the first sixteen entries of the Padé table are given in Appendix III.

Initial value problems for which $f = f(t,y,y')$ may clearly be written in the form of a first order system $\underset{\sim}{u}' = \underset{\sim}{v}$, $\underset{\sim}{v}' = \underset{\sim}{f}(t,\underset{\sim}{u},\underset{\sim}{v})$ where $\underset{\sim}{u} = \underset{\sim}{y}$, $\underset{\sim}{v} = \underset{\sim}{y}'$. Multiderivative methods for first order systems were discussed in Chapter 2.

5.3 Analyses

With the multiderivative method (5.14), may be associated with the linear difference operator L defined by

$$\begin{aligned} L[y(t);\ \ell] &= y(t+\ell) - 2y(t) + y(t-\ell) + \sum_{j=1}^{m} a_j\ell^{2j} y^{(2j)}(t+\ell) \\ &\quad - \sum_{w=1}^{s} b_w\ell^{2w} y^{(2w)}(t) + \sum_{j=1}^{m} a_j\ell^{2j} y^{(2j)}(t-\ell) . \end{aligned} \tag{5.15}$$

Expanding $y(t+\ell)$ and $y(t-\ell)$ and their derivatives as Taylor series about t, and gathering terms, gives

$$L[y(t);\ \ell] = C_0 y(t) + C_1\ell y'(t) + C_2\ell^2 y''(t) + \ldots . \tag{5.16}$$

where the C_j are constants. The operator L and the associated multiderivative method (5.14), are of order p if, in (5.16), $C_0 = C_1 = \ldots$

$= C_{p+1} = 0$ and $C_{p+2} \neq 0$.

The term C_{p+2} is the error constant of the multiderivative method (5.14); the error constants for the first 16 methods of the family are contained in Appendix III. The multiderivative method (5.14), is consistent with the differential equation if $p \geq 1$; clearly the methods based on the $(0,1)$ and $(1,0)$ Padé approximants are inconsistent whilst all others are consistent.

Rearranging (14) in the form

$$(5.17) \quad y_{n+1} - 2y_n + y_{n-1} = \sum_{j=1}^{v} \ell^{2j}(-a_j y_{n+1}^{(2j)} + b_j y_n^{(2j)} - a_j y_{n-1}^{(2j)}),$$

where $v = \max(m,s)$ (clearly for $m > k$, $b_{s+1} = \ldots = b_v = 0$, and for $m < k$, $a_{m+1} = .. = a_v = 0$), it is seen that the multiderivative methods are generated by the characteristic polynomials

$$(5.18) \quad \rho(r) = r^2 - 2r + 1 , \quad \sigma_j(r) = - a_j r^2 + b_j - a_j$$

for $j = 1,\ldots, v$. The quadratic polynomial equation $\sigma(r) = 0$ has a double zero at $r = +1$ and the family of multiderivative methods is zero-stable; all except the methods based on the $(0,1)$ and $(1,0)$ Padé approximants are thus convergent.

It is easy to see from (5.14) and (5.17) that, for m less than, equal to, or greater than k, every member of the family of multiderivative methods, is symmetric with even stepnumber (two-steps) and even order p. The findings of Lambert and Watson (1976) on the periodicity of linear multistep method then carry over to multiderivative methods, as does the theory of weak stability (Lambert (1973, p.202)).

Writing $H = \ell\lambda$, equation (5.13) becomes

$$\{1 - a_1H^2+a_2H^4- \ldots + (-1)^m a_m H^{2m}\}y_{n+1} - \{2 - b_1H^2+b_2H^4- \ldots +(-1)^s b_s H^{2s}\}y_n$$

$$(5.19) \qquad + \{1 - a_1H^2+a_2H^4 - \ldots + (-1)^m a_m H^{2m}\}y_{n-1} = 0 .$$

The solution of (5.19) involves the nth powers of the zeros r_1 and r_2 of the periodicity polynomial

$$\Omega(r,H^2) = \{1 - a_1H^2 + a_2H^4 - \ldots + (-1)^m a_m H^{2m}\}r^2$$
$$- \{2 - b_1H^2 + b_2H^4 - \ldots + (-1)^s b_s H^{2s}\}r$$
$$+ \{1 - a_1H^2 + a_2H^4 - \ldots + (-1)^m a_m H^{2m}\} ,$$
$$= Q_m(-iH)Q_m(iH)r^2 - \{Q_m(-iH)P_k(iH) + Q_m(iH)P_k(-iH)\}r$$
$$+ Q_m(-iH)Q_m(iH) . \qquad (5.20)$$

The interval of periodicity of the multiderivative method (5.14), is determined by computing the values of H^2 for which the zeros of the periodicity equation (Lambert and Watson (1976,p.193))

$$\Omega(r,H^2) = 0 \qquad (5.21)$$

satisfy

$$r_1 = e^{i\theta(H)} , \qquad r_2 = e^{-i\theta(H)} , \qquad (5.22)$$

where $\theta(H) \in \mathbb{R}$; the multiderivative method is then orbitally stable. For each member of the family of multiderivative methods, the periodicity equation may be written down in terms of the associated Padé approximant. Those multiderivative methods which have interval of periodicity $H^2 \in (0,\infty)$ are said to be P-stable (Lambert and Watson (1976,p.199)). The intervals of periodicity for the consistent multiderivative methods based on the first sixteen entries of the Padé table are contained in Appendix III (those interval bounds occurring as integers or improper fractions, are exact, those occurring with one decimal place, have been rounded up or down depending on whether the number is a lower or upper bound of the interval). The consistent multiderivative formulas based on those (m,k) Padé approximants for which $m \geq k$ are seen to be P-stable.

In computing the solution at time $t = \ell$ the formula

$$y_1 = y_0 + \ell y_0' + \frac{1}{3}\ell^2 y_0'' + \frac{1}{6}\ell^2 y_1'' - \frac{1}{18}\ell^4 y_0^{(iv)} + \frac{1}{72}\ell^4 y_1^{(iv)} + O(\ell^5) , \qquad (5.23)$$

or the formula

$$y_1 = y_0 + \ell y_0' + \frac{1}{3}\ell^2 y_0'' + \frac{1}{6}\ell^2 y_1'' - \frac{1}{45}\ell^4 y_0^{(iv)} - \frac{7}{360}\ell^4 y_1^{(iv)}$$

$$+ \frac{1}{108}\ell^6 y_0^{(vi)} - \frac{11}{2160}\ell^6 y_1^{(vi)} + O(\ell^7) , \tag{5.24}$$

(Twizell (1981)) may be used in solving problems which are known to have outward-spiralling theoretical solutions. Otherwise, Taylor's series may be used to give y, to the necessary accuracy.

To investigate absolute stability, the family of multiderivative methods are applied to the single test equation (5.4). Writing $\bar{h} = -\ell^2 \lambda^2$ (Lambert (1973,p.258)), the stability polynomial for each of the methods from equation (5.13) takes the form

$$\pi(r,\bar{h}) = (1 + a_1\bar{h} + a_2\bar{h}^2 + \ldots + a_m\bar{h}^m)(r^2+1)$$

$$- (2 + b_1\bar{h} + b_2\bar{h}^2 + \ldots + b_s\bar{h}^s)r \tag{5.25}$$

and the interval of absolute stability (Dahlquist (1978,pp.133-134)), are found by solving the equation

$$\pi(r,\bar{h}) = 0 \tag{5.26}$$

in each case.

Methods based on the (m,k) Padé approximants for $m \geq k$ are found to be A-stable, whilst methods based on the (m,k) Padé approximants for $m < k$ have finite interval of absolute stability $\bar{h} \in [-\alpha, 0]$. The value of α is in fact the same as for the interval of periodicity $H^2 \in (0,\alpha)$. The analogy between P-stability and A-stability for two-step symmetric multistep methods, is thus obvious for the family of multiderivative methods developed in section (5.2), see also Dahlquist (1963, 1978), Lambert and Watson (1976), Chawla (1981) and Hairer (1979).

5.4 Numerical examples

The family of multiderivative methods developed in section 5.2, were tested on two problems well known in the literature. Numerical results for methods based on the $(2,2)$ and $(3,3)$ Padé approximants, are presented

in this section.

Problem 5.1

This is the almost periodic problem introduced by Stiefel and Bettis (1969) and considered by Lambert and Watson (1976) and Cash (1981). It is given by

$$z'' = -z + 0.001\, e^{it} \quad ; \quad z(0) = 1 \ , \quad z'(0) = 0.9995\, i, \quad z \quad C.$$

The theoretical solution satisfies

$$u(t) = \cos t + 0.0005\, t \sin t, \qquad u \in \mathbb{R} \ ,$$

$$v(t) = \sin t - 0.0005\, t \cos t, \qquad v \in \mathbb{R}$$

$$z(t) = u(t) + iv(t)$$

and represents the motion of the point $z(t)$ on a perturbation of a circular orbit. The distance of this point from the centre of the orbit at time t is given by

$$\gamma(t) = [u^2(t) + v^2(t)]^{\frac{1}{2}} = [1 + (0.0005t)^2]^{\frac{1}{2}}$$

so that the point spirals slowly outwards as time increases.

Following Lambert and Watson (1976), the differential equation is written in the form of the real linear system

$$\begin{aligned} u'' &= -u + 0.001 \cos t \ ; \quad u(0) = 1 \ , \ u'(0) = 0 \ , \\ v'' &= -v + 0.001 \sin t \ ; \quad v(0) = 0, \ v'(0) = 0.9995 \ , \end{aligned} \tag{5.27}$$

from which the higher derivatives, for use with the multiderivative methods developed in section 5.2 are easily determined.

The numerical solutions $U(t)$, $V(t)$ of the real system (5.27), were computed at $t=40\pi$ for $\ell = \pi/4, \ \pi/5, \ \pi/6, \ \pi/9, \ \pi/12,$ using the multiderivative methods based on the $(2,2)$, $(3,3)$ Padé approximants. The corresponding computed values $Z(t)$, $\Gamma(t)$ of $z(t)$, $\gamma(t)$ were then computed using

$$Z(t) = U(t) + iV(t) \qquad , \qquad \Gamma(t) = [U^2(t) + V^2(t)]^{\frac{1}{2}}.$$

The error moduli in the computed values $Z(t)$, $\Gamma(t)$ given by

$$E(z) \equiv |z(t) - Z(t)| = [\{u(t) - U(t)\}^2 + \{v(t) - V(t)\}^2]^{\frac{1}{2}}$$

$$E(\gamma) \equiv |\gamma(t) - \Gamma(t)| = |\{u^2(t) + v^2(t)\}^{\frac{1}{2}} - \{U^2(t) + V^2(t)\}^{\frac{1}{2}}|$$

were also calculated. The values of $\Gamma(t)$, $E(z)$, $E(\gamma)$ are given in Table 5.1.

It can be seen that for both methods tested, the path of the point $z(t)$ is an outward spiral for all steplengths, which is in keeping with the theoretical solution. The numerical solution obtained using the fourth order method based on the (2,2) Padé approximant, was found to be closer to the theoretical value $\gamma(40\pi)$ than the method due to Cash (1981), which is of comparable order, except for $\ell = \pi/4$ when the error modulus was 0.002339 compared to 0.002146 obtained by Cash.

The computed solution obtained using the sixth order multiderivative method based on the (3,3) Padé approximant, was found to be closer to the theoretical solution $\gamma(40\pi)$ than the sixth order methods of Lambert and Watson (1976) or Cash (1981) for all values of ℓ. Moreover, convergence to six decimal places was attained for higher values of ℓ using the (3,3) multiderivative method.

The approximate formulas (5.23), (5.24) were used with the (2,2), (3,3) Padé methods, respectively, whereas Cash (1981) used the theoretical solution.

Problem 5.2

This example was used by Lambert and Watson (1976) and Cash (1981) and is given by

$$y_1'' = -w^2 y_1 + \phi''(t) + w^2\phi(t) \; ; \; y_1(0) = a + \phi(0), \; y_1'(0) = \phi'(0)$$

$$y_2'' = -w^2 y_2 + \phi''(t) + w^2\phi(t) \; ; \; y_2(0) = \phi(0) \; , \; y_2'(0) = a\,w + \phi'(0) \; .$$

The theoretical solution of the problem is given by

Table 5.1

Computed results at $t = 40\pi$ for Problem 5.1

$\gamma(40\pi) = 1.001972$, $u(40\pi) = 1$, $v(40\pi) = -0.062832$

ℓ	(2,2) method			(3,3) method		
	Γ	$E(\gamma)$	$E(z)$	Γ	$E(\gamma)$	$E(z)$
$\pi/4$	1.004311	0.234(-2)	0.418(-2)	1.001981	0.908(-5)	0.813(-7)
$\pi/5$	1.002845	0.874(-3)	0.710(-3)	1.001974	0.236(-5)	0.567(-8)
$\pi/6$	1.002383	0.411(-3)	0.167(-3)	1.001972	0.792(-6)	0.642(-9)
$\pi/9$	1.002052	0.805(-4)	0.659(-5)	1.001972	0.699(-7)	0.501(-11)
$\pi/12$	1.001997	0.255(-4)	0.664(-6)	1.001972	0.125(-7)	0.159(-12)

Table 5.2

Error modulus in the computed solution at $t = 20\pi$ for Problem 5.2

w \ ℓ	π/32	π/8	π/2	π
(2,2) multiderivative method				
5	0.402(-15)	0.103(-12)	0.194(-10)	0.200(-9)
10	0.994(-16)	0.885(-13)	0.139(-11)	0.115(-11)
15	0.119(-15)	0.101(-14)	0.183(-12)	0.144(-10)
20	0.879(-16)	0.502(-14)	0.858(-12)	0.359(-11)
25	0.428(-16)	0.254(-14)	0.522(-12)	0.519(-11)
30	0.310(-15)	0.752(-15)	0.246(-12)	0.390(-11)
35	0.502(-15)	0.715(-15)	0.261(-12)	0.265(-11)
40	0.176(-15)	0.782(-15)	0.251(-13)	0.179(-11)
(3,3) multiderivative method				
5	0.185(-15)	0.953(-15)	0.340(-11)	0.182(-12)
10	0.166(-15)	0.116(-14)	0.231(-14)	0.264(-14)
15	0.118(-15)	0.795(-16)	0.946(-17)	0.576(-15)
20	0.423(-16)	0.885(-16)	0.380(-16)	0.189(-15)
25	0.319(-15)	0.480(-17)	0.102(-16)	0.119(-16)
30	0.290(-15)	0.600(-18)	0.522(-17)	0.256(-16)
35	0.138(-15)	0.491(-18)	0.261(-17)	0.137(-16)
40	0.271(-16)	0.261(-18)	0.183(-17)	0.309(-17)

$$y_1(t) = a \cos wt + \phi(t)$$

$$y_1(t) = a \sin wt + \phi(t)$$

and, following Lambert and Watson (1976) and Cash (1981), $\phi(t)$ is taken to be $e^{-0.05t}$. The parameter a was given the value zero, corresponding to the case when high frequency oscillations are not present in the theoretical solution. The results at $t = 20\pi$ for $w = 5(5)40$ and $\ell = \pi/32, \pi/8, \pi/2, \pi$ are given in Table 5.2.

Comparing Table 5.2 with Table 2 in Cash (1981), it is seen that, except in the isolated case $w = 5, \ell = \pi/8$, the fourth order multiderivative method tested in the present paper gives better results than the fourth order method of Cash; the sixth order method tested in the present paper always gives superior results to the sixth order method in Cash (1981) when applied to Problem 5.2 .

As with Problem5.1,formulas (5.23), (5.24) were used to compute $y(\ell)$.

5.5 Use in PECE mode

In common with texts and other papers, the convention of associating an asterisk with a predictor formula will be adopted. Using the general $(0,k^*)$ method as predictor and the general (m,k) method as corrector, the combination in PECE mode will be denoted by $(0,k^*);(m,k)$.

It is not necessary to choose a predictor formula for which $k^* = \max(m,k)$ and the existing theory relating to the order of the local truncation error of linear multistep methods used in PECE mode carries over to multiderivative methods used in PECE mode. In particular, if the order of the predictor is at least the order of the corrector, then the error constant of the predictor-corrector combination is that of the corrector alone. In addition, if the predictor and the corrector have the same order p, then Milne's device

$$C_{p+2}[y_{n+1}^{(c)} - y_{n+1}^{(P)}]/[C_{p+2}^{*} - C_{p+2}] \tag{5.28}$$

may be used to estimate the error constant of the predictor-corrector combination in PECE mode (provided $C_{p+2}^{*} \neq C_{p+2}$). In (5.26), the superscripts

(P) and (C) refer to the predictor and corrector, respectively.

The periodicity polynomial $\Omega_{PECE}(r,H^2)$ of the $(0,k^*);(m,k)$ combination in PECE mode may be shown to take the form

$$\Omega_{PECE}(r,H^2) = r^2 - [2 - 2\sum_{j=1}^{m}(-1)^j a_j H^{2j} + \sum_{j=1}^{s}(-1)^j b_j H^{2j}$$

$$(5.29) \qquad + \sum_{j=1}^{m}(-1)^j a_j H^{2j} \sum_{w=1}^{s^*}(-1)^w b_w^* H^{2w}]r + 1 \quad .$$

where $s^* = [\frac{1}{2}k^*]$.

The interval of periodicity of the $(0,k^*);(m,k)$ predictor-corrector combination is determined by computing the values of H for which the zeros of the periodicity equation

$$\Omega_{PECE}(r,H^2) = 0$$

satisfy (5.22).

It was found that the $(0,2);(1,2)$ combination, with error constant $C_4 = -\frac{1}{36}$ and periodicity interval $H^2 \in (0,9)$, has the smallest modulus error constant and the greatest interval of periodicity of the second order combinations.

Of the fourth order combinations, it was found that the $(0,4);(2,2)$ combination, for which $C_6 = \frac{1}{360}$ and $H^2 \in (0,15.89)$, is to be preferred to any fourth order combination when solving non-linear problems, because it requires no more than the second derivative of $\underline{f}(t,\underline{y})$. For linear problems the $(0,4);(1,3)$ combination which has $C_6 = \frac{-7}{2880}$ and $H^2 \in (0,4.88)$, may be used with small values of ℓ if higher accuracy is needed.

For non-linear problems of the form (5.1), the maximum steplength which may be used at any time t of the calculation, has the value $H^*/\Lambda(t)$, where $H^2 \in (0,H^{*2})$ is the periodicity interval of the predictor-corrector combination being used, and $\Lambda^2(t)$ is the largest modulus real part of the eigenvalues of the Jacobian $\partial\underline{f}/\partial\underline{y}$ at time t .

The $(0,4);(2,2)$ method was tested on the following problem which was

discussed in Shampine and Gordon (1973) and Jain et al (1979).

Problem 5.3

$$x'' = -\frac{x}{r^3} \; ; \; x(0) = 1, \; x'(0) = 0 \; ,$$

$$y'' = -\frac{y}{r^3} \; ; \; y(0) = 0, \; y'(0) = 1 \; ,$$

where $r = (x^2 + y^2)^{\frac{1}{2}}$. These equations are Newton's equations of motion for the two body problem and the initial conditions are such that the motion is circular. Clearly $x''(0) = -1$, $y''(0) = 0$ and, by successively differentiating the expressions for x'' and y'', it is easy to verify that $x(t)$ and its derivatives take the values $1,0,-1,0$ cyclically at $t = 0$, and that $y(t)$ and its derivatives take the values $0,1,0,-1$ cyclically at $t = 0$. Taylors series, with sufficiently small stepsizes, provides starting values for the following strategy where, for

$n = 1,2,\ldots,$ $\underline{w}_n = [x_n, y_n]^T = [x(n\ell), y(n\ell)]^T$:

P: $\underline{w}^{(P)}_{n+1}$ is calculated using, as predictor, the multiderivative method based on the $(0,4)$ Padé approximant;

E: (a) $\underline{w}'_{n+1}$ is evaluated using $\underline{w}'_{n+1} = \frac{1}{\ell} \sum_{m=1}^{6} \nabla \underline{w}^{(P)}_{n+1}/m + O(\ell^6)$,

where ∇ is the usual backward difference operator,

(b) $\underline{w}''_{n+1}$ is evaluated using $\underline{w}^{(P)}_{n+1}$ in the system of differential equations,

(c) $\underline{w}^{(iv)}_{n+1}$ is evaluated from the analytical expressions for $x^{(iv)}_{n+1}$, $y^{(iv)}_{n+1}$ which are easily determined (these contain x'_{n+1}, y'_{n+1}) ;

C: $\underline{w}^{(C)}_{n+1}$ is calculated using, as corrector, the multiderivative method based on the $(2,2)$ Padé approximant;

E: $\underline{w}'_{n+1}$, $\underline{w}''_{n+1}$, $\underline{w}^{(iv)}_{n+1}$ are re-evaluated as in (a), (b), (c) above using the corrected value $\underline{w}^{(C)}_{n+1}$ where appropriate.

The problem was tested using $\ell = \pi/18,\ \pi/15,\ \pi/10$ and the numerical solution at time $t = 12\pi$ determined using the $(0,4);(2,2)$ combination in PECE mode. Using the theoretical solution $x(t) = \cos t,\ y(t) = \sin t$ the error moduli for the three values of ℓ are easily found and are given in Table 5.3. Results are also tabulated using the $(0,4)$ method alone. Comparison with Table 1 in Jain et al (1979) shows that multi-derivative methods give accurate numerical results for non-linear as well as linear problems.

Problem 5.4

Changing the initial conditions in Problem 5.3 to

$$x(0) = 0.4\ ,\ x'(0) = 0\ ,$$

$$y(0) = 0\ \ ,\ y'(0) = 2\ ,$$

causes the orbit to become the eclipse (Shampine and Gordon (1973,p.245))

$$r^2 \equiv (x+0.6)^2 + y^2/0.64 = 1$$

and the period of revolution to be 2π. The problem was tested with $\ell = \pi/45,\ \pi/90,\ \pi/180,\ \pi/360,\ \pi/720$ and the value of r at time $t = 15\pi, 16\pi$ determined using the $(0,2);(2,2)$ combination in PECE mode. The values of x,y,r (theoretical values $-1.6,0,1$ and $0.4,0,1$, respectively) are given at time $t = 15\pi, 16\pi$ in Table 5.4. It is again clear that the multiderivative predictor-corrector combination used, gives accurate results. Unlike the method used and reported in Shampine and Gordon (1973,p.246), no step size or order changing was required to achieve the accuracy obtained using the multiderivative methods.

Table 5.3

Error moduli at $t = 12\pi$ for Problem 5.3

	(0,4) method	(0,4) ; (2,2) combination
ℓ	Error moduli	Error moduli
$\pi/18$	0.394(-7)	0.665(-8)
$\pi/15$	0.209(-6)	0.298(-7)
$\pi/10$	0.650(-5)	0.120(-5)

Table 5.4

Computed values of x,y,r at $t = 15\pi, 16\pi$ for Problem 5.4

	x	y	r
$t = 15\pi$			
$\pi/45$	-1.6003845	0.0244339	1.0017020
$\pi/90$	-1.5997948	0.0010838	0.9995914
$\pi/180$	-1.5999258	-0.0001281	0.9998517
$\pi/360$	-1.5999801	-0.0000584	0.9999603
$\pi/720$	-1.5999950	-0.0000163	0.9999899
$t = 16\pi$			
$\pi/45$	0.3999265	0.0057571	0.9999049
$\pi/90$	0.3995450	0.0252020	1.0000826
$\pi/180$	0.3999516	0.0081528	1.0000070
$\pi/360$	0.3999966	0.0021600	1.0000005
$\pi/720$	0.3999998	0.0005477	1.0000000

5.6 Conclusions

A family of two-step multiderivative methods, based on Padé approximants to the exponential function, has been developed for periodic initial value problems of second order ordinary differential equations. The method based on the (0,2) Padé approximant is seen to be the usual explicit method. The method based on the (1,1) Padé approximant, is found to be the formula, found to be unconditionally stable by Richtmyer and Morton (1967,p.263) in connection with the solution of second order hyperbolic equations, and discussed for second order ordinary differential equations by Dahlquist (1978). However, the topic of Dahlquist's paper was unconditional stability, not P-stability.

The methods based on the (m,k) Padé approximants, for $m \geq k$, are found to be P-stable, while the methods for $m < k$ are seen to have finite interval of periodicity. Following Dahlquist (1963, 1978), Lambert and Watson (1976) and Hairer (1979), it is concluded that for two-step multiderivative methods, the analogy between P-stability and A-stability is obvious. Numerical experiments have confirmed that the two-step multiderivative methods developed in this chapter give higher accuracy than the fourth and sixth order two-step Runge-Kutta type methods developed by Cash (1981). For non-linear problems, where higher derivatives cannot be calculated with ease, predictor-corrector combinations can be used. The application of the methods for fourth order parabolic partial differential equations in one and two space dimensions, will be discussed in Chapter 6.

CHAPTER 6

FOURTH ORDER PARABOLIC EQUATIONS

6.1 Introduction

The fourth order parabolic partial differential equation in one space variable given by

$$\frac{\partial^2 u}{\partial t^2} + \mu \frac{\partial^4 u}{\partial x^4} = 0 \quad ; \quad \mu > 0, 0 < x < X, t > 0 \tag{6.1}$$

arises in the study of the transverse vibrations of a uniform flexible beam (see, for example, Gorman (1975)). The term μ is the ratio of the flexural rigidity of the beam to its mass per unit length.

The initial conditions associated with (6.1) are of the form

$$u(x,0) = g_0(x) \quad ; \quad 0 \le x \le X \ , \tag{6.2}$$

$$\frac{\partial u}{\partial t}(x,0) = g_1(x) \quad ; \quad 0 \le x \le X \ , \tag{6.3}$$

and the boundary conditions are given by

$$u(0,t) = f_0 \ , \quad u(X,t) = f_1 \quad ; \quad t > 0 \ , \tag{6.4}$$

$$\frac{\partial^2 u}{\partial x^2}(0,t) = p_0 \ , \quad \frac{\partial^2 u}{\partial x^2}(X,t) = p_1 \ ; \quad t > 0 \ , \tag{6.5}$$

In (6.2), (6.3) the functions $g_0(x)$, $g_1(x)$ are continuous and in (6.4), (6.5) the terms f_0, f_1, p_0, p_1 are real constants.

To compute the solution of (6.1) with (6.2), (6.3), (6.4), (6.5), explicit and implicit finite difference schemes have been proposed by Albrecht (1957), Collatz (1951), Conte (1957), Conte and Royster (1956), and Crandall (1954). Evans (1965) derived finite difference methods by first writing (6.1) as two simultaneous second order parabolic partial differential equations (see also Dufort and Frankel (1953), and Richtmyer (1957)). Explicit and implicit finite difference methods based on the semi-explicit method of Lees (1961) and the high order method of Douglas (1956) for second order parabolic equations, have been formulated for the numerical solution of (6.1) with (6.2), (6.3) by Fairweather and Gourlay

(1967).

The explicit method of Collatz (1951) frequently needs a large number of time steps to compute the solution in view of the stability restriction on the method. The difference scheme given by Albrecht (1957) overcomes the stability problem but uses the value of the solution at four time levels to compute the solution at a fifth time level. The work of Fairweather and Gourlay (1967) gives superior numerical results to the methods of Evans (1965) and Richtmyer (1957), but more CPU time is required.

In this chapter a family of novel finite difference schemes is developed for the numerical solution of (6.1) with (6.2), (6.3), (6.4), (6.5); a related procedure was adopted by Lawson and Morris (1978) for second order parabolic equations, by Khaliq and Twizell (1982) for first order hyperbolic equations and by Twizell (1979) for second order hyperbolic equations. The methods developed and analysed are tested on problems discussed in the literature by Andrade and McKee (1977), and Fairweather and Gourlay (1967).

6.2 A recurrence relation

The interval $0 \le x \le X$ will be divided into $N+1$ subintervals each of width h so that $(N+1)h = X$ and the time variable t is discretized in steps of length ℓ. The open region $R = [\,0 < x < X\,] \times [\,t > 0\,]$ and its boundary ∂R consisting of the lines $x = 0$, $x = X$, $t = 0$ are thus covered by a rectangular mesh, the mesh points having co-ordinates $(mh, n\ell)$ where $m = 0,1,\ldots,N+1$ and $n = 0,1,2,\ldots$. The theoretical solution of a difference scheme approximating (6.1) will again be denoted by U_m^n at the mesh point $(mh, n\ell)$.

Superimposing this grid allows the space derivative in (6.1) to be approximated by the finite difference replacement

$$\frac{\partial^4 u}{\partial x^4} = h^{-4}\{u(x-2h,t)-4u(x-h,t)+6u(x,t)-4u(x+h,t)+u(x+2h,t)\}+O(h^4) \tag{6.6}$$

for $x = mh$ $(m=1,2,\ldots,N)$. For $m = 1$ and $m = N$, equation (6.6) introduces the points $(-h,t)$ and $(X+h,t)$ at time t. which are outside the region R. However, the values of u at mesh points outside the region R may be written in terms of function values in R and on ∂R by using (6.4), (6.5).

In (6.5) $\partial^2 u/\partial x^2$ is approximated by the usual second order replacement; this gives

$$\frac{\partial^2 u}{\partial x^2} = h^{-2}\{u(-h,t)-2u(0,t)+u(h,t)\}+0(h^2) \tag{6.7}$$

at the boundary points $(0,t)$, and

$$\frac{\partial^2 u}{\partial x^2} = h^{-2}\{u(X-h,t)-2u(X,t)+u(X+h,t)\}+0(h^2) \tag{6.8}$$

at the boundary (X,t). Using (6.4), (6.5), equations (6.7), (6.8)give

$$u(-h,t) = -u(h,t)+2f_0+h^2p_0+0(h^4) \tag{6.9}$$

$$u(X+h,t) = -u(X-h,t)+2f_1+h^2p_1+0(h^4) \tag{6.10}$$

respectively, and it is these expressions which will be used when (6.6) is used with $x = h$ and $x = Nh$.

Consider now the time level $t = n\ell$ and apply (6.1) with (6.6) to the N mesh points at this time level. This leads to the system of second order ordinary differential equations given by

$$\frac{d^2\underline{U}}{dt^2} = -\mu\, A\underline{U}(t)-\mu\underline{w} \quad , \tag{6.11}$$

where $\underline{U}(t) = [\,U_1(t),\ U_2(t),\ldots,U_N(t)]^T$, T denoting transpose, is the vector of computed solutions at the N mesh points at time t, A is the square matrix of order N given by

$$A = h^{-4}\begin{bmatrix} 5 & -4 & 1 & & & & \\ -4 & 6 & -4 & 1 & & 0 & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 6 & -4 \\ & 0 & & & 1 & -4 & 5 \end{bmatrix} \tag{6.12}$$

with eigenvalues

$$\lambda_s = 16\, h^{-4} \sin^4[\, s\pi/\{2(N+1)\}\,] \;;\quad s = 1,2,\ldots,N \tag{6.13}$$

and $\underline{w}$ is the vector of boundary values of order N given by

$$\underline{w} = h^{-4}[\, h^2p_0-2f_0, f_0, 0, \ldots, 0, f_1, h^2p_1-2f_1\,]^T \;. \tag{6.14}$$

Solving (6.11) subject to the initial conditions (6.2), (6.3) gives

$$\underline{U}(t) = -A^{-1}\underline{w} + \tfrac{1}{2}\exp(i\gamma tB)\,\{\underline{g}_0+(i\gamma tB)^{-1}\underline{g}_1+A^{-1}\underline{w}\} + \tfrac{1}{2}\exp(-i\gamma tB)\,\{\underline{g}_0-(i\gamma tB)^{-1}\underline{g}_1+A^{-1}\underline{w}\} \;, \tag{6.15}$$

in which $i = \sqrt{-1}$, $\gamma = \sqrt{\mu}$ and B is a matrix such that $B^2 = A$.

It is easy to show that (6.15) satisfies the recurrence relation

$$\underline{U}(t+\ell) - \{\exp(i\gamma\ell B)+\exp(-i\gamma\ell B)\}\underline{U}(t)+\underline{U}(t-\ell) = \{\exp(i\gamma\ell B)+\exp(-i\gamma\ell B)\}A^{-1}\underline{w}-2A^{-1}\underline{w} \tag{6.16}$$

with $t = \ell, 2\ell, \ldots$ and it is this relation which will be used in the development of the family of algorithms for solving (6.1) with (6.2), (6.3), (6.4), (6.5). It will not be necessary to compute γ, B or A^{-1} explicitly.

6.3 Solution at the first time step

It is clear that, using (6.15) with $t = \ell$, requires knowledge of $\underline{U}(\ell)$ which, unlike $\underline{U}(0)$, is not contained explicitly in the initial conditions. Writing $t = \ell$ in (6.15) and replacing the matrix exponential functions with their (0,3) Padé approximants leads to

$$\underline{U}(\ell) = (I- \tfrac{1}{2}\mu\ell^2A)\underline{g}_0+\ell(I- \tfrac{1}{6}\mu\ell^2A)\underline{g}_1- \tfrac{1}{2}\mu\ell^2\underline{w}+0(\ell^4) \;; \tag{6.17}$$

replacing the matrix exponential functions with their (0,5) Padé approximants leads to

$$\underline{U}(\ell) = (I- \tfrac{1}{2}\mu\ell^2A+ \tfrac{1}{24}\mu^2\ell^4A^2)\underline{g}_0+\ell(I- \tfrac{1}{6}\mu\ell^2A+\tfrac{1}{120}\mu^2\ell^4A^2)\underline{g}_1 - \tfrac{1}{2}\mu\ell^2(I- \tfrac{1}{12}\mu\ell^2A)\underline{w}+0(\ell^6) \tag{6.18}$$

and using the (0,7) Padé approximants leads to

$$(6.19)\qquad \underline{U}(\ell) = (I - \tfrac{1}{2}\mu\ell^2 A + \tfrac{1}{24}\mu^2\ell^4 A^2 - \tfrac{1}{720}\mu^3\ell^6 A^3)\underline{g}_0$$
$$+\ell(I - \tfrac{1}{6}\mu\ell^2 A + \tfrac{1}{120}\mu\ell^4 A^2 - \tfrac{1}{5040}\mu^3\ell^6 A^3)\underline{g}_1$$
$$- \tfrac{1}{2}\mu\ell^2(I - \tfrac{1}{12}\mu\ell^2 A + \tfrac{1}{360}\mu^2\ell^4 A^2)\underline{w} + O(\ell^8)\ .$$

In problems having time dependent boundary conditions f_0, f_1, p_0, p_1 in (6.4), (6.5) are functions of t and the vector $\underline{w}$ in (6.17), (6.18), (6.19) is evaluated using $t = \ell$ in the equation

$$(6.20)\qquad \underline{w}_t = h^{-4}\,[\,h^2 p_0(t) - 2f_0(t), f_0(t), 0, \ldots, 0, f_1(t), h^2 p_1(t) - 2f_1(t)\,]^T.$$

The complete algorithm for computing the numerical solution of (6.1) with (6.2), (6.3), (6.4), (6.5) may thus be listed as follows:

(i) the starting vector $\underline{U}(0) = \underline{g}_0$ is obtained from equation (6.2);

(ii) the starting vector $\underline{U}(\ell)$ is obtained using (6.17), (6.18) or (6.19) depending on the required accuracy;

(iii) $\underline{U}(t+\ell)$, with $t = \ell, 2\ell, \ldots,$ is obtained from the recurrence relation (6.16) in which the matrix exponential functions are replaced by suitable approximants. It is these matrix functions which will be replaced by Padé approximants in the next section.

6.4 Development and analyses of the methods

Using the (1,1) Padé approximants to the matrix exponential functions in (6.16) leads to a difference scheme written in matrix form as

$$(6.21)\qquad (I + \tfrac{1}{4}\mu\ell^2 A)\underline{U}(t+\ell) = (2I - \tfrac{1}{2}\mu\ell^2 A)\underline{U}(t) - (I + \tfrac{1}{4}\mu\ell^2 A)\underline{U}(t-\ell) - \mu\ell^2\underline{w} + O(\ell^4);$$

for problems with time dependent boundary conditions this becomes

$$(6.22)\qquad (I + \tfrac{1}{4}\mu\ell^2 A)\underline{U}(t+\ell) + \tfrac{1}{4}\mu\ell^2\underline{w}_{t+\ell}$$
$$= (2I - \tfrac{1}{2}\mu\ell^2 A)\underline{U}(t) - \tfrac{1}{2}\mu\ell^2\underline{w}_t - (I + \tfrac{1}{4}\mu\ell^2 A)\underline{U}(t-\ell) - \tfrac{1}{4}\mu\ell^2\underline{w}_{t-\ell} + O(\ell^4).$$

The principal part of the local truncation error of the method based on the (1,1) Padé approximant is given by

$$(6.23)\qquad \tfrac{1}{6}\mu h^2\ell^2\ \frac{\partial^6 u}{\partial x^6} - \tfrac{1}{6}\ell^4\ \frac{\partial^4 u}{\partial t^4}\ .$$

The component $\frac{1}{6}\mu h^2 \ell^2 \frac{\partial^6 u}{\partial x^6}$ of (6.23) is related to the space discretization and the use of (6.6) in (6.1); this component will be present in all methods derived by using Padé approximants to the matrix exponential function in (6.16). The other component of (6.23) is related only to the Padé approximant chosen for use in (6.16). The principal part of the local truncation error of any method arising from the use of the (m,k) Padé approximant in (6.16) will thus have the form

$$\frac{1}{6}\mu h^2 \ell^2 \frac{\partial^6 u}{\partial x^6} + C_q \ell^q \frac{\partial^q u}{\partial t^q} \quad , \tag{6.24}$$

where the C_q ($q = m+k+1$ for $m + k$ odd, and $q = m+k+2$ for $m + k$ even) are error constants and are given in Appendix III. All Padé approximants except the $(0,1)$, $(1,0)$ approximants lead to consistent methods.

Stability, in the conventional sense of a perturbation of the initial data not growing in magnitude as time increases, is analysed by recourse to the stability equation of the method.

Noting that the (m,k) Padé approximant to the matrix exponential function $\exp(i\gamma\ell B)$ has the form

$$\exp(i\gamma\ell B) = [Q_m(i\gamma\ell B)]^{-1} . P_k(i\gamma\ell B) + O(\ell^{m+k+1}) \tag{6.25}$$

where P_k, Q_m are polynomials of degrees k and m, respectively, with $P_0(i\gamma\ell B) \equiv I$ and $Q_0(i\gamma\ell B) \equiv I$ (I is the identity matrix of order N), the stability equation has the form

$$\begin{aligned} Q_m(i\gamma\ell\lambda^{\frac{1}{2}})Q_m(-i\gamma\ell\lambda^{\frac{1}{2}})\xi^2 - \{P_k(i\gamma\ell\lambda^{\frac{1}{2}})Q_m(-i\gamma\ell\lambda^{\frac{1}{2}}) + P_k(-i\gamma\ell\lambda^{\frac{1}{2}})Q_m(i\gamma\ell\lambda^{\frac{1}{2}})\}\xi \\ + Q_m(i\gamma\ell\lambda^{\frac{1}{2}})Q_m(-i\gamma\ell\lambda^{\frac{1}{2}}) = 0. \end{aligned} \tag{6.26}$$

In (6.26), λ is an eigenvalue of A and ξ is the amplification factor of the method. The von Neumann necessary condition for stability $|\xi| \leq 1$ hence requires,

$$|P_k(i\gamma\ell\lambda^{\frac{1}{2}})Q_m(-i\gamma\ell\lambda^{\frac{1}{2}}) + P_k(-i\gamma\ell\lambda^{\frac{1}{2}})Q_m(i\gamma\ell\lambda^{\frac{1}{2}})| \leq 2|Q_m(i\gamma\ell\lambda^{\frac{1}{2}})Q_m(i\gamma\ell\lambda^{\frac{1}{2}})|. \tag{6.27}$$

In the case of the method based on the (1,1) Padé approximant, the stability equation is

$$(6.28)\qquad (1+\tfrac{1}{4}\mu\ell^2\lambda)\xi^2-(2-\tfrac{1}{2}\mu\ell^2\lambda)\xi+(1+\tfrac{1}{4}\mu\ell^2\lambda) = 0$$

and it is found that $|\xi| \le 1$ for any $r = \ell/h^2 \ge 0$ since $\mu > 0$, $\lambda > 0$. The scheme is therefore unconditionally stable.

Using the (0,2) Padé approximant in (6.16) the resulting finite-difference method for problems with time dependent boundary conditions may be written in vector form as

$$(6.29)\qquad \underline{U}(t+\ell) = (2I-\mu\ell^2A)\underline{U}(t)-\mu\ell^2\underline{w}_t+\underline{U}(t-\ell).$$

This is the explicit scheme of Collatz (1951) for which $C_4 = \frac{1}{12}$. This method has an error constant which is the same order as that of (6.22) and, since it is explicit, it would appear to be a more desirable method to use. It is, however, stable only for $r \le \frac{1}{2\sqrt{\mu}}$ and may thus be used only with small time steps.

Turning now to the (1,2) Padé approximant, its use in (6.16) yields the method

$$(6.30)\qquad (I+\tfrac{1}{9}\mu\ell^2A)\underline{U}(t+\ell)+\tfrac{1}{9}\mu\ell^2\underline{w}_{t+\ell}$$

$$= (2I-\tfrac{7}{9}\mu\ell^2A)\underline{U}(t)-\tfrac{7}{9}\mu\ell^2\underline{w}_t-(I+\tfrac{1}{9}\mu\ell^2A)\underline{U}(t-\ell)-\tfrac{1}{9}\mu\ell^2\underline{w}_{t-\ell}.$$

The method is second order accurate with $C_4 = -1/36$, so that the method enjoys better accuracy than (6.22) or (6.29). Its stability equation

$$(1+\tfrac{1}{9}\mu\ell^2\lambda)\xi^2-(2-\tfrac{7}{9}\mu\ell^2\lambda)\xi+(1+\tfrac{1}{9}\mu\ell^2\lambda) = 0$$

yields the restriction $\mu\ell^2\lambda \le 36/5$ which, since $\lambda < 16h^{-4}$, leads to the stability condition $r \le 3\sqrt{5}\,/(10\sqrt{\mu})$. Thus it may be used with slightly bigger time steps than (6.29). However, the fact that it is implicit, does not make this method more attractive than (6.22) which, though implicit, is unconditionally stable.

A notable improvement in the accuracy in time is obtained by using the (2,2) Padé approximant to the matrix exponential functions in (6.16).

This approximant gives the method

$$(6.31)\qquad (I+\tfrac{1}{12}\mu\ell^2A+\tfrac{1}{144}\mu^2\ell^4A^2)\underline{U}(t+\ell)+(\tfrac{1}{12}\mu\ell^2I+\tfrac{1}{144}\mu^2\ell^4A)\underline{w}_{t+\ell}$$

$$=(2I-\tfrac{5}{6}\mu\ell^2A+\tfrac{1}{72}\mu^2\ell^4A^2)\underline{U}(t)-(\tfrac{5}{6}\mu\ell^2I+\tfrac{1}{72}\mu^2\ell^4A)\underline{w}_t$$

$$-(I+\tfrac{1}{12}\mu\ell^2A+\tfrac{1}{144}\mu^2\ell^4A^2)\underline{U}(t-\ell)-(\tfrac{1}{12}\mu\ell^2I+\tfrac{1}{144}\mu^2\ell^4A)\underline{w}_{t-\ell}$$

for which $C_6=\frac{1}{360}$ (from Appendix III). The stability equation is

$$(1+\tfrac{1}{12}\mu\ell^2\lambda+\tfrac{1}{144}\mu^2\ell^4\lambda^2)\xi^2-(2-\tfrac{5}{6}\mu\ell^2\lambda+\tfrac{1}{72}\mu^2\ell^4\lambda^2)\xi+(1+\tfrac{1}{12}\mu\ell^2\lambda+\tfrac{1}{144}\mu^2\ell^4\lambda^2)=$$

from which it is found that the method is unconditionally stable.

Squaring the matrix A involves an increase in the number of mesh points at each time level used in the computation. This notion of using a greater number of points at each time level was used by Khaliq and Twizell (1982) for first order hyperbolic partial differential equations and by Twizell (1979) for second order hyperbolic partial differential equations; Mitchell and Griffiths (1980) discussed the concept briefly for second order parabolic partial differential equations.

The same order of accuracy in time may be achieved by deleting the terms in A^2 from (6.31); this gives

$$(6.31a)\qquad (I+\tfrac{1}{12}\mu\ell^2A)\underline{U}(t+\ell)+\tfrac{1}{12}\mu\ell^2\underline{w}_{t+\ell}$$

$$=(2I-\tfrac{5}{6}\mu\ell^2A)\underline{U}(t)-\tfrac{5}{6}\mu\ell^2\underline{w}_t-(I+\tfrac{1}{12}\mu\ell^2A)\underline{U}(t-\ell)-\tfrac{1}{12}\mu\ell^2\underline{w}_{t-\ell}$$

for which $C_6 = 1/240$. Equation (6.31a) is, in fact, an application of Numerov's linear multi-step method for the numerical solution of a system of second order ordinary differential equations and the finite difference scheme resulting from it for the solution of (6.1) is stable only for $\mu\, r^2 \le 3/8$. Equation (6.31a) may be useful when very small time steps may be taken.

The (2,1) Padé approximant leads to the implicit method

$$(6.32)\qquad (I+\tfrac{1}{9}\mu\ell^2A+\tfrac{1}{36}\mu^2\ell^4A^2)\underline{U}(t+\ell)+(\tfrac{1}{9}\mu\ell^2I+\tfrac{1}{36}\mu^2\ell^4A)\underline{w}_{t+\ell}$$

$$= (2I - \tfrac{7}{9}\mu\ell^2 A)\underline{U}(t) - \tfrac{7}{9}\mu\ell\underline{w}_t$$

$$- (I + \tfrac{1}{9}\mu\ell^2 A + \tfrac{1}{36}\mu^2\ell^4 A^2)\underline{U}(t-\ell) - (\tfrac{1}{9}\mu\ell^2 I + \tfrac{1}{36}\mu^2\ell^4 A)\underline{w}_{t-\ell} .$$

This method has $C_4 = \frac{1}{36}$ and is found to be unconditionally stable. Its theoretical accuracy near the boundary is not second order in time; however, this does not diminish the overall accuracy of difference scheme, see, for example, Mitchell and Griffiths (1980,p.112-116, 121-125). It may be advisable to delete the terms in A^2 from (6.32), provided sufficiently small steps may be taken. The method then becomes identical to (6.30), which has error constants of the same magnitude as (6.32) and which is obviously more economical than (6.32) in relation to storage requirements.

Using the (2,0) Padé approximant to the matrix exponential functions in (6.16) gives the implicit scheme

(6.33) $$(I + \tfrac{1}{4}\mu^2\ell^4 A^2)\underline{U}(t+\ell) + \tfrac{1}{4}\mu^2\ell^4 A\underline{w}_{t-\ell}$$

$$= (2I - \mu\ell^2 A)\underline{U}(t) - \mu\ell^2\underline{w}_t - (I + \tfrac{1}{4}\mu^2\ell^4 A)\underline{U}(t-\ell) - \tfrac{1}{4}\mu^2\ell^4 A\underline{w}_{t-\ell}$$

which has error constant $C_4 = 7/12$. The method is unconditionally stable but its less favourable error constant and the fact that it requires A^2, suggest that the method based on the (1,1) Padé approximant, is to be preferred. It will be seen in section 6.5, however, to give generally better numerical results than (6.22) for the problems tested, when a higher order difference scheme is used for the first time step.

6.5 Numerical results and discussion

To examine the behaviour of the methods developed in section 6.4, the methods are tested on two problems from the literature. The methods based on the (1,1), (1,2), (2,0), (2,1), (2,2) Padé approximants, will be named T11, T12, T20, T21, T22, respectively.

Problem 6.1 (Fairweather and Gourlay (1967))

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial^4 u}{\partial x^4} = 0 \quad ; \quad 0 < x < 1 \quad , \quad t > 0$$

with initial conditions

$$u(x,0) = \frac{x^2}{12}(2x^2-x^3-1) \quad ; \quad 0 \le x \le 1$$

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

and boundary conditions

$$u(0,t) = u(1,t) = 0 \quad ; \quad t \ge 0$$

$$\frac{\partial^2 u}{\partial x^2}(0,t) = \frac{\partial^2 u}{\partial x^2}(1,t) = 0 \quad ; \quad t \ge 0 \ .$$

This problem was also considered by Evans (1965). The theoretical solution is given by

$$u(x,t) = \sum_{s=1}^{\infty} a_s \sin s\pi x \cos s^2\pi^2 t \ ,$$

where

$$a_s = \frac{4}{s^5\pi^5}\{\cos(s\pi) - 1\} \ .$$

In order to compare the numerical results with Fairweather and Gourlay (1967), the same mesh ratios have been chosen. In Table 6.1, the errors are shown for time $t = 0.02$, with $h = 0.05$, $\ell = 0.00125$. The errors for $t = 1.0$ with $h = 0.05$, $\ell = 0.005$ and with $h = 0.1$, $\ell = 0.02$ are quoted in Table 6.2 and Table 6.3 respectively.

Visual analyses of Tables 6.1, 6.2 and 6.3, and comparison with Tables I and III in Fairweather and Gourlay (1967, p.9), show that the numerical results for the second order methods are superior to those of Evan's method, Richtmyer's method and the semi-explicit method (Fairweather and Gourlay (1967, p.9). The fourth order method (in time) based on the (2,2) Padé approximant is seen to give better results than those of the higher order correct method of Douglas (Fairweather and Gourlay (1967, p.9), Tables I, III) for larger mesh ratios, especially when the time step is not too small relative to the space discretization. This is due to the fact that the component of the principal part of the local truncation error due to the chosen Padé approximant in (6.24), namely

$C_q \ \ell^q \ \partial^q u/\partial t^q$, is much smaller than the component due to the space discretization. Thus little improvement in numerical results can be expected for the fourth order method (in time) compared to second order methods (in time), for small values of the time step relative to the space discretization. This phenomenon was also observed in Chapter 3 for second order parabolic equations, when the (3,0) Padé approximant was employed to the exponential matrix function in (3.7). The present approach differs in detail to that of Fairweather and Gourlay (1967), in the manner in which the numerical solution of (6.1) is sought. However, following Fairweather and Gourlay (1967), the methods due to Gourlay and Morris (1980) and the methods developed in Chapter 3 may also be adopted to find the numerical solution of (6.1) by writing that equation as a system of two second order parabolic equations.

Table 6.1 Maximum errors at $t = 0.02$
$h = 0.05$, $\ell = 0.00125$ $(r = \frac{1}{2})$

Methods	x				
	0.1	0.2	0.3	0.4	0.5
T11	1.99(-6)	3.63(-6)	5.98(-6)	-7.73(-7)	-3.34(-6)
T20	1.80(-6)	3.94(-6)	3.76(-6)	-2.97(-8)	-1.63(-6)
T21	1.74(-6)	3.45(-6)	5.26(-6)	-4.30(-7)	-9.95(-7)
T22	1.67(-6)	2.70(-6)	4.90(-6)	-4.20(-8)	-2.86(-7)

Table 6.2 Maximum errors at $t = 1.0$
$h = 0.05$, $\ell = 0.005$ $(r = 2)$

Methods	x				
	0.1	0.2	0.3	0.4	0.5
T11	-1.75(-4)	-2.63(-4)	-2.36(-4)	-1.60(-4)	-1.16(-4)
T20	-1.74(-4)	-2.40(-4)	-1.79(-4)	-4.47(-5)	2.31(-5)
T21	-5.49(-5)	-1.10(-4)	-1.66(-4)	-1.23(-4)	-8.41(-5)
T22	-5.91(-5)	-1.29(-5)	-1.78(-5)	-2.61(-5)	-3.15(-5)

Table 6.3 Maximum errors at t = 1.0
h = 0.1, ℓ = 0.02 (r = 2)

Methods	x				
	0.1	0.2	0.3	0.4	0.5
T11	-2.93(-4)	-5.67(-4)	-9.29(-4)	-1.10(-3)	-1.07(-3)
T20	-8.43(-5)	-2.22(-5)	2.40(-4)	3.81(-4)	4.04(-4)
T21	-4.66(-4)	-4.81(-4)	-6.68(-4)	-7.23(-4)	-9.91(-5)
T22	-2.06(-4)	-4.45(-4)	-3.52(-4)	-2.31(-4)	-7.35(-5)

Problem 6.2 (Andrade and McKee (1977))

$$\frac{\partial^2 u}{\partial t^2} + a(x,t)\frac{\partial^4 u}{\partial x^4} = 0 \quad ; \quad a(x,t) > 0 \quad , \quad \tfrac{1}{2} < x < 1, t > 0$$

$$a(x,t) = \frac{1}{x} + \frac{x^4}{120} \ ,$$

with initial conditions

$$u(x,0) = 0 \quad ; \quad \tfrac{1}{2} \le x \le 1$$

$$\frac{\partial u}{\partial t}(x,0) = 1 + \frac{x^5}{120} \quad ; \quad \tfrac{1}{2} \le x \le 1$$

and boundary conditions

$$u(\tfrac{1}{2},t) = \{1+(\tfrac{1}{2})^5/120\} \sin t \quad ; \quad t > 0$$

$$u(1,t) = (1+1/120) \sin t \quad ; \quad t > 0$$

$$\frac{\partial^2 u}{\partial x^2}(\tfrac{1}{2},t) = \frac{1}{6}(\tfrac{1}{2})^3 \sin t \quad ; \quad t > 0$$

$$\frac{\partial^2 u}{\partial x^2}(1,t) = \frac{1}{6} \sin t \quad ; \quad t > 0 \ .$$

The theoretical solution is

$$u(x,t) = (1+\frac{x^5}{120})\sin t \ .$$

Table 6.4 Maximum absolute relative errors at $t = 0.01$

Value of r	No. of time steps	Methods				
		T11	T12	T20	T21	T22
0.05	80	3.45(-7)	3.47(-7)	3.45(-7)	3.31(-7)	9.91(-8)
0.1	40	3.41(-7)	3.52(-7)	3.43(-7)	3.25(-7)	8.07(-8)
0.25	16	3.22(-7)	3.90(-7)	3.28(-7)	3.19(-7)	6.91(-8)

In order to provide a comparison with Andrade and McKee (1977), the mesh ratios $r = 0.05, 0.1, 0.25$ are chosen. In Table 6.4, the maximum absolute relative errors are shown for time $t = 0.01$, where $a(x,t)$ is evaluated at $x = ih, i = 1,2,\ldots,N$. Following Mitchell and Griffiths (1980,p.26) it is verified that the methods T11, T12, T20, T21, T22, maintain the same order of accuracy for $\mu \equiv \mu(x,t)$. For stability analysis the stability criterion of the energy method due to Lees (1960), may be applied; however, for this problem stability of the methods will be verified by numerical experiments. It is seen from Table 6.4 that the methods T11, T12, T20, T21, T22, give superior results to that of Andrade and McKee (1977, p.13, Table 1). Method T12 is seen to have a better stability interval than the method developed by Andrade and McKee (1977) and the unconditional stability of the methods T11, T20, T21, T22, is an extra advantage. The relation (6.17) is used to calculate the numerical solution at the first time step for all the methods.

It is noticed that numerical calculations made by Andrade and McKee (1977, p.13) for the usual explicit method for this problem, are incorrect. The method gives better results than the method developed by Andrade and McKee (1977, p.13, Table 1). The maximum absolute relative errors for the same values of the mesh ratios using the usual explicit method (the $(0,2)$ Padé approximant) are seen to be slightly better than those of the method T11, which is in accordance with the local truncation errors of the methods as shown in Appendix III.

6.6 Two-space variables

The homogeneous partial differential equation

$$\frac{\partial^2 u}{\partial t^2} + \nabla^4 u = 0 \quad , \quad \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \quad , \tag{6.34}$$

with $0 < x,y < 1$, $t > 0$, subject to initial conditions of the form

$$\left.\begin{aligned} u(x,y,0) &= f_1(x,y) \\ \frac{\partial u}{\partial t}(x,y,0) &= f_2(x,y) \end{aligned}\right\} \quad (x,y) \in \Omega \tag{6.35}$$

and boundary conditions of the form

$$\left.\begin{aligned} u(0,y,t) &= u(1,y,t) = 0 \\ u(x,0,t) &= u(x,0,t) = 0 \\ \frac{\partial^2 u}{\partial x^2}(x,0,t) &= \frac{\partial^2 u}{\partial x^2}(x,1,t) = 0 \\ \frac{\partial^2 u}{\partial y^2}(0,y,t) &= \frac{\partial^2 u}{\partial y^2}(1,y,t) = 0 \end{aligned}\right\} \quad (x,y) \in \partial\Omega,\ t > 0 \ , \tag{6.36}$$

arises in the transverse vibration of a simply supported uniform plate. Superimpose a square grid over the unit square with mesh size $h = 1/N+1$ for some positive integer N. Let Ω be those grid points $(x,y) = (ih,jh)$ for $1 \le i,j \le N$ (that is, the interior of the square) and let $\partial\Omega$ be those points for which $i,j = 0$ or $N+1$ (the boundary of the square). Replacing the spatial derivatives in (6.34) with their central difference replacements, $\nabla^4 u$ becomes

$$\begin{aligned} \nabla^4 u = \frac{1}{h^4}[\ & u(x-2h,y,t)+u(x,y-2h,t)+u(x,y+2h,t)+u(x+2h,y,t) \\ & + 2\{u(x-h,y-h,t)+u(x-h,y+h,t)+u(x+h,y-h,t) \\ & + u(x+h,y+h,t)\} \\ & - 8\{u(x-h,y,t)+u(x,y-h,t)+u(x,y+h,t) \\ & + u(x+h,y,t)\}+20u(x,y,t)] + O(h^2) \ , \end{aligned}$$

and applying (6.34) with the boundary conditions (6.36) to each mesh point,

leads to a system of second order ordinary differential equations of the form

$$\frac{d^2\underline{U}}{dt^2} = - S\underline{U} \tag{6.37}$$

where S is an $N^2 \times N^2$ matrix and $\underline{U}$ is a vector of order N^2 of computed values such that

$$\underline{U} = (U_{1,1},U_{2,1},\ldots,U_{N,1};U_{1,2},U_{2,2},\ldots,U_{N,2};U_{1,N},U_{2,N},\ldots,U_{N,N})^T,$$

T denoting transpose. The matrix S is given by

$$S = h^{-4} \begin{bmatrix} B & C & D & & & & \\ C & B & C & D & & \text{O} & \\ D & C & B & C & D & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & D & C & B & C & D \\ & \text{O} & & D & C & B & C \\ & & & & D & C & B \end{bmatrix},$$

where B,C, and D are $N \times N$ matrices, such that

$$B = \begin{bmatrix} 19 & -8 & 1 & & & & \\ -8 & 20 & -8 & 1 & & \text{O} & \\ 1 & -8 & 20 & -8 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -8 & 20 & -8 & 1 \\ & & & 1 & -8 & 20 & -8 \\ & \text{O} & & & 1 & -8 & 19 \end{bmatrix},$$

$$C = \begin{bmatrix} -8 & 2 & & & & \\ 2 & -8 & 2 & & \text{O} & \\ & \ddots & \ddots & \ddots & \ddots & \\ & & 2 & -8 & 2 & \\ & \text{O} & & 2 & -8 & \end{bmatrix}, \qquad D = \begin{bmatrix} 1 & & & & \\ & 1 & & \text{O} & \\ & & \ddots & \ddots & \\ & & & 1 & \\ & \text{O} & & & 1 \end{bmatrix}.$$

The eigenvalues of S are given by

$$\lambda_{i,j} = \frac{16}{h^4}\{\sin^2\left(\frac{i\pi}{2(N+1)}\right)+ \sin^2\left(\frac{j\pi}{2(N+1)}\right)\}^2 \quad ; \quad i,j = 1,2,\ldots,N.$$

Solving (6.37) its solution is seen to satisfy the recurrence relation

$$\underline{U}(t+\ell) - \{\exp(i\ell E)+\exp(-i\ell E)\}\underline{U}(t)+\underline{U}(t-\ell) = \underline{0} \quad , \tag{6.38}$$

where $i = \sqrt{-1}$ and the matrix E is such that $E^2 = S$. The methods developed for one space dimension will be generalised for two space dimensions in this section. The principal part of the local truncation error includes $(\frac{1}{6}\ell^2h^2\nabla^6u)^n_{i,j}$, $i,j = 1,2,\ldots,N$ and $n = 0,1,2,\ldots$, which will always be present. However, the accuracy in time will depend upon the chosen Padé approximant. It is also assumed that u is sufficiently often differentiable with respect to both x and t.

Employing the $(1,1)$ Padé approximant to the matrix exponential function in (6.38) yields

$$(I+ \tfrac{1}{4}\ell^2S)\underline{U}(t+\ell) = (2I- \tfrac{1}{2}\ell^2S)\underline{U}(t)-(I+ \tfrac{1}{4}\ell^2S)\underline{U}(t-\ell) \tag{6.39}$$

A stability analyses shows that (6.39) is unconditionally stable; the scheme is seen to have the same order of accuracy as in the one space dimension case.

Applying the $(1,2)$ Padé approximant in (6.38) gives

$$(I+ \tfrac{\ell^2}{9} S)\underline{U}(t+\ell) = (2I- \tfrac{7}{9}\ell^2S)\underline{U}(t)-(I+\tfrac{\ell^2}{9} S)\underline{U}(t-\ell) \tag{6.40}$$

The method (6.40) is second order accurate with stability restriction $r \leqslant \frac{3\sqrt{5}}{20}$. To implement these methods on a computer, the direct method of matrix decomposition formulated by Buzbee and Door (1974), may be used to find the solution at $\underline{U}(t+\ell)$. Application of block Gaussian elimination to the matrices of the form S was described in Bauer and Reiss (1972) and Angle and Bellman (1972). Employing these Padé approximants in (6.37), which require squaring the matrix S, the difficulties encountered in implementing the methods in one space-dimension are magnified in the case of two-space variables. The methods based on the $(2,0)$,

(2,1), (2,2) Padé approximants are seen to have no linear factors, thus a complex splitting for each of the methods is suggested. However, it will cost more to use complex arithmetic than real arithmetic.

Employing the (2.0) Padé approximant in (6.38), yields the algorithm

$$
\begin{aligned}
&(I-\tfrac{1}{2}i\ell^2S)\underline{U}^* = (2I-\ell^2S)\underline{U}(t)\\
&(I+\tfrac{1}{2}i\ell^2S)\underline{U}(t+\ell) = \underline{U}^* - (I+\tfrac{1}{2}i\ell^2S)\underline{U}(t-\ell)\ ,
\end{aligned}
\tag{6.41}
$$

where $i = \sqrt{-1}$.

Application of the (2,1) Padé approximant in (6.38), suggests the algorithm,

$$
\begin{aligned}
&\{(1+2\sqrt{2}\ i)I+\tfrac{1}{2}\ell^2S\}\underline{U}^* = \underline{U}(t)\\
&\{(1-2\sqrt{2}\ i)I+\tfrac{1}{2}\ell^2S\}\underline{U}(t+\ell) = (18I+7\ell^2S)\underline{U}^*\\
&\qquad - \{(1-2\sqrt{2}\ i)I+\tfrac{1}{2}\ell^2S\}\underline{U}(t-\ell)
\end{aligned}
\tag{6.42}
$$

and the (2,2) Padé approximant in (6.38), yields

$$
\begin{aligned}
&\{(1-i\sqrt{3})I+\tfrac{1}{6}\ell^2S\}\underline{U}^* = \{2(\sqrt{21}+5)I-\tfrac{1}{3}\ell^2S\}\underline{U}(t)\\
&\{(1+i\sqrt{3})I+\tfrac{1}{6}\ell^2S\}\underline{U}(t+\ell) = \{(-\sqrt{21}+5)I-\tfrac{1}{6}\ell^2S\}\underline{U}^*\\
&\qquad - \{(1+i\sqrt{3})I+\tfrac{1}{6}\ell^2S\}\underline{U}(t-\ell)\ ,
\end{aligned}
\tag{6.43}
$$

where $\underline{U}^*$ is an intermediate vector. The algorithms are seen to have the same order of accuracy as in the one space dimension case and (6.41), (6.42), (6.43) are verified to be unconditionally stable. The algorithms (6.41), (6.42) and (6.43) were also tested on Problem 6.1 in conjunction with the matrix A and the same numerical results were found, as tabulated in Tables 6.1, 6.2 and 6.3. To examine the behaviour of the methods in two space dimensions, the methods are tested on a problem suggested by Andrade and McKee (1977). The methods based on the (1,1), (1,2), (2,0), (2,1), (2,2) Padé approximants used in conjunction with the matrix S will be named S11, S12, S20, S21, S22.

Problem 6.3 (Andrade and McKee (1977)

$$\frac{\partial^2 u}{\partial t^2} + a(x,y,t)\,\frac{\partial^4 u}{\partial x^4} + b(x,y,t)\,\frac{\partial^4 u}{\partial y^4} = 0 \quad ; \quad 0 < x,y < 1 \ , \ t > 0$$

where

$$a(x,y,t) = \frac{1}{2\pi^2}\left(1 - \frac{x^2}{2} + \frac{y^2}{8} + \frac{t^2}{8}\right) ,$$

$$b(x,y,t) = \frac{1}{2\pi^2}\left(1 + \frac{x^2}{2} - \frac{y^2}{8} - \frac{t^2}{8}\right) ,$$

with initial conditions

$$u(x,y,0) = 0 \quad ; \quad 0 \le x,y \le 1 \ ,$$

$$\frac{\partial u}{\partial t}(x,y,0) = \pi \sin \pi x \sin \pi y \quad ; \quad 0 \le x,y \le 1$$

and the homogeneous boundary conditions (6.36); the theoretical solution is given by

$$u(x,y,t) = \sin \pi t \sin \pi x \sin \pi y \ .$$

To compare the results with Andrade and McKee (1977, Table 2), the maximum absolute relative errors for the methods S11, S12, S20, S21, S22 at time $t = 0.05$ are tabulated in Table 6.5. It is seen from Table 6.5 that the second order methods S11, S12, S20, S21, do not give better results when compared with Andrade and McKee (1977, Table 2). However, the method S22 gives better results than those of Andrade and McKee (1977, Table 2).

Table 6.5 Maximum absolute relative errors at $t = 0.05$

Value of r	No. of time steps	Methods				
		S11	S12	S20	S21	S22
0.05	100	6.87(-5)	6.99(-5)	6.85(-5)	7.81(-6)	8.71(-7)
0.1	50	6.81(-5)	7.51(-5)	6.83(-5)	6.49(-6)	7.29(-7)
0.25	20	6.70(-5)	7.64(-5)	6.75(-5)	5.50(-6)	7.10(-7)

Appendix I: The first twenty four entries of the Padé table for $f(z) = e^z$

e^z	k = 0	k = 1	k = 2	k = 3	k = 4
m = 0	1	$1+z$	$1+z+\frac{1}{2}z^2$	$1+z+\frac{1}{2}z^2+\frac{1}{6}z^3$	$1+z+\frac{1}{2}z^2+\frac{1}{6}z^3+\frac{1}{24}z^4$
m = 1	$\frac{1}{1-z}$	$\frac{1+z/2}{1-z/2}$	$\frac{1+\frac{2}{3}z+\frac{1}{6}z^2}{1-\frac{1}{3}z}$	$\frac{1+\frac{3}{4}z+\frac{1}{4}z^2+\frac{1}{24}z^3}{1-\frac{1}{4}z}$	$\frac{1+\frac{4}{5}z+\frac{3}{10}z^2+\frac{1}{15}z^3+\frac{1}{120}z^4}{1-\frac{1}{5}z}$
m = 2	$\frac{1}{1-z+\frac{1}{2}z^2}$	$\frac{1+\frac{z}{3}}{1-\frac{2}{3}z+\frac{1}{6}z^2}$	$\frac{1+\frac{z}{2}+\frac{1}{12}z^2}{1-\frac{z}{2}+\frac{1}{12}z^2}$	$\frac{1+\frac{3}{5}z+\frac{3}{20}z^2+\frac{1}{60}z^3}{1-\frac{2}{5}z+\frac{1}{20}z^2}$	$\frac{1+\frac{2}{3}z+\frac{1}{5}z^2+\frac{1}{30}z^3+\frac{1}{360}z^4}{1-\frac{z}{3}+\frac{z^2}{30}}$
m = 3	$\frac{1}{1-z+\frac{1}{2}z^2-\frac{1}{6}z^3}$	$\frac{1+\frac{1}{4}z}{1-\frac{3}{4}z+\frac{1}{4}z^2-\frac{1}{24}z^3}$	$\frac{1+\frac{2}{5}z+\frac{1}{20}z^2}{1-\frac{3}{5}z+\frac{3}{20}z^2-\frac{1}{60}z^3}$	$\frac{1+\frac{z}{2}+\frac{1}{10}z^2+\frac{1}{120}z^3}{1-\frac{z}{2}+\frac{1}{10}z^2-\frac{1}{120}z^3}$	$\frac{1+\frac{4}{7}z+\frac{1}{7}z^2+\frac{2}{105}z^3+\frac{1}{840}z^4}{1-\frac{3}{7}z+\frac{1}{7}z^2-\frac{2}{105}z^3+\frac{1}{840}z^4}$
m = 4	$\frac{1}{1-z+\frac{1}{2}z^2-\frac{1}{6}z^3+\frac{1}{24}z^4}$	$\frac{1+\frac{1}{5}z}{1-\frac{4}{5}z+\frac{3}{10}z^2-\frac{1}{15}z^3+\frac{1}{360}z^4}$	$\frac{1+\frac{1}{3}z+\frac{1}{30}z^2}{1-\frac{2}{3}z+\frac{1}{5}z^2-\frac{1}{30}z^3+\frac{1}{360}z^4}$	$\frac{1+\frac{3}{7}z+\frac{1}{14}z^2+\frac{1}{210}z^3}{1-\frac{4}{7}z+\frac{1}{7}z^2-\frac{2}{105}z^3+\frac{1}{840}z^4}$	$\frac{1+\frac{z}{2}+\frac{3}{28}z^2+\frac{1}{84}z^3+\frac{1}{1680}z^4}{1-\frac{z}{2}+\frac{3}{28}z^2-\frac{1}{84}z^3+\frac{1}{1680}z^4}$

Appendix II: One-step multiderivative methods based on the first twenty-four entries of the Padé Table for the exponential function.

(0,1) : $y_{n+1} = y_n + hy'_n + O(h^2)$. (Euler's predictor).

(1,1) : $y_{n+1} = y_n + \frac{1}{2}h(y'_n+y'_{n+1}) + O(h^3)$. (Euler's corrector ; the trapezoidal rule).

(1,0) : $y_{n+1} = y_n + hy'_{n+1} + O(h^2)$.

(0,2) : $y_{n+1} = y_n + hy'_n + \frac{1}{2}h^2y''_n + O(h^3)$.

(1,2) : $y_{n+1} = y_n + \frac{2}{3}hy'_n + \frac{1}{3}hy'_{n+1} + \frac{1}{6}h^2y''_n + O(h^4)$.

(2,2) : $y_{n+1} = y_n + \frac{1}{2}h(y'_n+y'_{n+1}) + \frac{1}{12}h^2(y''_n-y''_{n+1}) + O(h^5)$.

(2,1) : $y_{n+1} = y_n + \frac{1}{3}hy'_n + \frac{2}{3}hy'_{n+1} - \frac{1}{6}h^2y''_{n+1} + O(h^4)$.

(2,0) : $y_{n+1} = y_n + hy'_{n+1} - \frac{1}{2}h^2y''_{n+1} + O(h^3)$.

(0,3) : $y_{n+1} = y_n + hy'_n + \frac{1}{2}h^2y''_n + \frac{1}{6}h^3y_n^{(iii)} + O(h^4)$.

(1,3) : $y_{n+1} = y_n + \frac{1}{4}h(3y'_n+y'_{n+1}) + \frac{1}{4}h^2y''_n + \frac{1}{24}h^3y_n^{(iii)} + O(h^5)$.

(2,3) : $y_{n+1} = y_n + \frac{1}{5}h(3y'_n+2y'_{n+1}) + \frac{1}{20}h^2(3y''_n-y''_{n+1}) + \frac{1}{60}h^3y_n^{(iii)} + O(h^5)$.

(3,3) : $y_{n+1} = y_n + \frac{1}{2}h(y'_n+y'_{n+1}) + \frac{1}{10}h^2(y''_n-y''_{n+1})$

$+ \frac{1}{120}h^3(y_n^{(iii)} + y_{n+1}^{(iii)}) + O(h^7)$. (Milne's starting procedure).

(3,2) : $y_{n+1} = y_n + \frac{1}{5}h(2y'_n+3y'_{n+1}) + \frac{1}{20}h^2(y''_n-3y''_{n+1}) + \frac{1}{60}h^3y_{n+1}^{(iii)} + O(h^6)$.

(3,1) : $y_{n+1} = y_n + \frac{1}{4}h(y'_n+3y'_{n+1}) - \frac{1}{4}h^2y''_{n+1} + \frac{1}{24}h^3y_{n+1}^{(iii)} + O(h^5)$.

(3,0) : $y_{n+1} = y_n + hy'_{n+1} - \frac{1}{2}h^2y''_{n+1} + \frac{1}{6}h^3y_{n+1}^{(iii)} + O(h^4)$.

(0,4) : $y_{n+1} = y_n + hy'_n + \frac{1}{2}h^2y''_n + \frac{1}{6}h^3y_n^{(iii)} + \frac{1}{24}h^4y_n^{(iv)} + O(h^5)$.

(1,4) : $y_{n+1} = y_n + \frac{1}{5}h(4y'_n+y'_{n+1}) + \frac{3}{10}h^2y''_n + \frac{1}{15}h^3y_n^{(iii)} + \frac{1}{120}h^4y_n^{(iv)} + O(h^6)$.

$$(2,4): \quad y_{n+1} = y_n + \frac{1}{3}h(2y'_n+y'_{n+1}) + \frac{1}{5}h^2(y''_n-\frac{1}{6}y''_{n+1}) + \frac{1}{30}h^3y_n^{(iii)} + \frac{1}{360}h^4y_n^{(iv)} + 0(h^7).$$

$$(3,4): \quad y_{n+1} = y_n + \frac{1}{7}h(4y'_n+3y'_{n+1}) + \frac{1}{14}h^2(2y''_n-y''_{n+1}) + \frac{1}{210}h^3(4y_n^{(iii)}+y_{n+1}^{(iii)}) + \frac{1}{840}h^4y_n^{(iv)} + 0(h^8).$$

$$(4,4): \quad y_{n+1} = y_n + \frac{1}{2}h(y'_n+y'_{n+1}) + \frac{3}{28}h^2(y''_n-y''_{n+1}) + \frac{1}{84}h^3(y_n^{(iii)}+y_{n+1}^{(iii)}) + \frac{1}{1680}h^4(y_n^{(iv)}-y_{n+1}^{(iv)}) + 0(h^9).$$

$$(4,3): \quad y_{n+1} = y_n + \frac{1}{7}h(y'_n+4y'_{n+1}) + \frac{1}{14}h^2(y''_n-2y''_{n+1}) + \frac{1}{210}h^3(y_n^{(iii)}+4y_{n+1}^{(iii)}) - \frac{1}{840}h^4y_{n+1}^{(iv)} + 0(h^8).$$

$$(4,2): \quad y_{n+1} = y_n + \frac{1}{3}h(y'_n+2y'_{n+1}) + \frac{1}{30}h^2(y''_n-6y''_{n+1}) + \frac{1}{30}h^3y_{n+1}^{(iii)} - \frac{1}{360}h^4y_{n+1}^{(iv)} + 0(h^7).$$

$$(4,1): \quad y_{n+1} = y_n + \frac{1}{5}h(y'_n+4y'_{n+1}) - \frac{3}{10}h^2y''_{n+1} + \frac{1}{15}h^3y_{n+1}^{(iii)} - \frac{1}{120}h^4y_{n+1}^{(iv)} + 0(h^6).$$

$$(4,0): \quad y_{n+1} = y_n + hy'_{n+1} - \frac{1}{2}h^2y''_{n+1} + \frac{1}{6}h^3y_{n+1}^{(iii)} - \frac{1}{24}h^4y_{n+1}^{(iv)} + 0(h^5).$$

Appendix III

The non-zero constants $a_j (j = 1,\ldots,m)$, $b_w (w = 1,\ldots,s)$ for the first sixteen entries of the Padé Table for the exponential function, together with the error constants and the intervals of periodicity.

(0,1) : All $a_j = 0$; all $b_w = 0$.

$C_2 = 1$ (method inconsistent).

(1,1) : $a_1 = -\frac{1}{4}$; $b_1 = \frac{1}{2}$.

$C_4 = -\frac{1}{6}$; $H^2 \in (0,\infty)$.

(1,0) : $a_1 = -1$; all $b_w = 0$.

$C_2 = -1$ (method inconsistent).

(0,2) : All $a_j = 0$; $b_1 = 1$.

$C_4 = \frac{1}{12}$; $H^2 \in (0,4)$.

(1,2) : $a_1 = -\frac{1}{9}$; $b_1 = \frac{7}{9}$.

$C_4 = -\frac{1}{36}$; $H^2 \in (0,\frac{36}{5})$.

(2,2) : $a_1 = -\frac{1}{12}$, $a_2 = \frac{1}{144}$; $b_1 = \frac{5}{6}$, $b_2 = \frac{1}{72}$.

$C_6 = \frac{1}{360}$; $H^2 \in (0,\infty)$.

(2,1) : $a_1 = -\frac{1}{9}$, $a_2 = \frac{1}{36}$; $b_1 = \frac{7}{9}$.

$C_4 = \frac{1}{36}$; $H^2 \in (0,\infty)$.

(2,0) : $a_2 = \frac{1}{4}$; $b_1 = 1$.

$C_4 = \frac{7}{12}$; $H^2 \in (0,\infty)$.

(0,3) : All $a_j = 0$; $b_1 = 1$.

$C_4 = \frac{1}{12}$; $H^2 \in (0,4)$.

(1,3) : $a_1 = -\frac{1}{16}$; $b_1 = \frac{7}{8}$, $b_2 = \frac{1}{48}$.

$C_6 = -\frac{7}{2880}$; $H^2 \in (0,6.5)$ and $(29.5,48)$.

(2,3) : $a_1 = -\frac{3}{50}$, $a_2 = \frac{1}{400}$, $b_1 = \frac{22}{25}$, $b_2 = \frac{17}{600}$.

$C_6 = \frac{1}{3600}$; $H^2 \in (0,8.2)$ and $(14.6,\frac{300}{7})$.

(3,3) : $a_1 = -\frac{1}{20}$, $a_2 = \frac{1}{600}$, $a_3 = -\frac{1}{14400}$;

$b_1 = \frac{9}{10}$, $b_2 = \frac{11}{330}$, $b_3 = \frac{1}{7200}$.

$C_8 = -\frac{1}{50400}$; $H^2 \in (0,\infty)$.

(3,2) : $a_1 = -\frac{3}{50}$, $a_2 = \frac{1}{400}$, $a_3 = -\frac{1}{3600}$;

$b_1 = \frac{22}{25}$, $b_2 = \frac{17}{600}$.

$C_6 = -\frac{1}{3600}$; $H^2 \in (0,\infty)$.

(3,1) : $a_1 = -\frac{1}{16}$, $a_3 = -\frac{1}{576}$; $b_1 = \frac{7}{8}$, $b_2 = \frac{1}{48}$.

$C_6 = \frac{-17}{2880}$; $H^2 \in (0,\omega)$.

(3,0) : $a_1 = -\frac{1}{12}$, $a_3 = -\frac{1}{36}$; $b_1 = 1$.

$C_4 = -\frac{1}{12}$; $H^2 \in (0,\infty)$.

(0,4) : All $a_j = 0$; $b_1 = 1$, $b_2 = \frac{1}{12}$.

$C_6 = \frac{1}{360}$; $H^2 \in (0,12)$.

Appendix IV

Two-step methods for second order equations.

(1,1): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{4}(f_{n+1} + 2f_n + f_{n-1})$

(0,2): $y_{n+1} - 2y_n + y_{n-1} = h^2 f_n$

(1,2): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{9}(f_{n+1} + 7f_n + f_{n-1})$

(2 2): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{12}(f_{n+1} + 10f_n + f_{n-1}) - \frac{h^4}{144}(f''_{n+1} - 2f''_n + f''_{n-1})$

(2,1): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{9}(f_{n+1} + 7f_n + f_{n-1}) - \frac{h^4}{36}(f''_{n+1} + f''_{n-1})$

(2,0): $y_{n+1} - 2y_n + y_{n-1} = h^2 f_n - \frac{h^2}{4}(f''_{n+1} + f''_{n-1})$

(0,3): $y_{n+1} - 2y_n + y_{n-1} = h^2 f_n$

(1,3): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{16}(f_{n+1} + 14f_n + f_{n-1}) + \frac{h^4}{48} f''_n$

(2,3): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{50}(3f_{n+1} + 44f_n + 3f_{n-1}) - \frac{h^4}{1200}(3f_{n+1} - 34f_n + 3f_{n-1})$

(3,3): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{20}(f_{n+1} + 18f_n + f_{n-1}) - \frac{h^4}{600}(f''_{n+1} - 22f''_n + f''_{n-1}) + \frac{h^6}{14400}(f^{iv}_{n+1} + 2f^{iv}_n + f^{iv}_{n-1})$

(3,2): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{50}(3f_{n+1} + 44f_n + 3f_{n-1}) \frac{-h^4}{1200}(3f''_{n+1} - 34f''_n + 3f''_{n-1}) \frac{+h^6}{3600}(f^{iv}_{n+1} + f^{iv}_{n-1})$

(3,1): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{16}(f_{n+1} + 14f_n + f_{n-1}) + \frac{h^4}{48} f''_n + \frac{h^6}{576}(f^{iv}_{n+1} + f^{iv}_{n-1})$

(3,0): $y_{n+1} - 2y_n + y_{n-1} = \frac{h^2}{12}(f_{n+1} + 12f_n + f_{n-1}) + \frac{h^6}{36}(f^{iv}_{n+1} + f^{iv}_{n-1})$

(0,4): $y_{n+1} - 2y_n + y_{n-1} = h^2 f_n + \frac{h^4}{12} f''_n$

REFERENCES

1. S. Abarbanel, D. Gottlieb and E. Turkel (1975), "Difference schemes with fourth order accuracy for hyperbolic equations", SIAM J. App. Math., 29, pp. 329-351.

2. J. Albrecht (1957), "Zum differenzenverfahren bei parabolischen differentialgleichungen", Z. Angew. Math and Mech., 37(5-6), pp. 202-212.

3. J. Al' Kabi, P. H. Gore, E. F. Saad, D. N. Waters and G. F. Moxon (1983), "Kinetic analysis of an extended single-path sequence of first order reactions: the reaction of mesitonitrate in sulphuric acid", to appear in Int. J. Chem. Kinetics.

4. C. Andrade and S. McKee (1977), "High accuracy A.D.I. methods for fourth order parabolic equations with variable coefficients", Int. J. Comp. App. Math., 3(1), pp. 11-14.

5. E. Angel and R. Bellman (1972), "Dynamic Programing and Partial Differential equations", Academic Press, New York.

6. O. Axelsson (1969), "A class of A-stable methods", BIT,9,pp. 185-199.

7. O. Axelsson (1972), "A note on a class of strongly A-stable methods", BIT, 12, pp. 1-4.

8. O. Axelsson (1975), "High-order methods for parabolic problems", J. Comp. and Appl. Math., 1, pp. 5-16.

9. D. Barton, I. M. Willers and R. V. M. Zahar (1971), "Taylor series methods for ordinary differential equations - an evaluation", from "Mathematical Software", ed. J. R. Rice, Academic Press.

10. L. Bauer and E. L. Reiss (1972), "Block five diagonal matrices and fast numerical solution of biharmonic equation", Math. Comp., 26 pp. 311-326.

11. J. P. Boris and D. L. Book (1973), "Flux corrected transport I. SHASTA A fluid transport algorithm that works", J. Comp. Phys., 11, pp. 38-64.

12. R. L. Brown (1974), "Multi-derivative numerical methods for the solution of stiff ordinary differential equations", Report UIUCDCS-R-74-672, Univ. of Illinois Dept. of Computer Science, U.S.A.

13. R. L. Brown (1976), "Numerical Integration of Linearized Stiff Ordinary Differential Equations", in "Numerical Methods for Differential Systems, Recent Developments in Algorithms, Software, and Applications", eds. L. Lapidus and W. E. Schiesser, Academic Press Inc. London.

14. R. Bulirsch and J. Stoer (1966), "Numerical treatment of ordinary differential equations by extrapolation methods," Numer. Math., 8, pp. 1-13.

15. B. L. Buzbee and F. W. Door (1974), "The direct solution of biharmonic equation on rectangular regions and the Poisson equation on irregular regions", SIAM J. Numer. Anal., 11(4), pp. 753-763.

16. B. L. Buzbee, F. W. Door, J. A. George and G. H. Golub (1971), "The direct solution of the discrete Poisson equation on irregular regions", SIAM J. Numer. Anal., 8, pp. 722-736.

17. J. R. Cash (1980), "On the integration of stiff system of O.D.Es using extended backward differentiation formulae", Numer. Math., 34, pp. 235-246.

18. J. R. Cash (1981), "High order P-stable formulae for the numerical integration of periodic initial value problems", Numer. Math., 37, pp. 355-370.

19. M. M. Chawla (1981), "Two-step fourth order P-stable methods for second order differential equations", BIT, 21, pp.190-193.

20. F. H. Chapman (1971), "A-stable Runge-Kutta processes", BIT, 11, pp. 384-388.

21. M. Ciment and S. H. Leventhal (1975), "High order compact implicit schemes for the wane equation", Math. Comp., 29(132), pp.985-994.

22. L. Collatz (1951), "Zur stabilitat des differenzenverfahrens bei der stabschwingungsgleichung", Z. Angew. Math and Mech., 31, pp. 392-393.

23. S. D. Conte (1957), "A stable implicit finite difference approximations to a fourth order parabolic equation", J. Assoc. Comp. Mach., 4, pp. 202-212.

24. S. D. Conte and W. C. Royster (1956), "Convergence of finite difference solutions to a solution of the equation of the vibrating rode", Proc. Amer. Math. Soc., 7, pp. 742-749.

25. S. H. Crandall (1954), "Numerical treatment of a fourth order partial differential equation", J. Assoc. Comp. Mach., 1, pp. 111-118.

26. J. Crank and P. Nicolson (1947), "A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type", Proc. Cambridge Philos. Soc., 43, pp.50-67. MR 8, 409.

27. C. W. Cryer (1973), "A new class of highly stable methods: A_0-stable methods", BIT, 13, pp. 153-159.

28. G. Dahlquist (1963), "A special stability problem for linear multistep methods", BIT, 3, pp. 27-43.

29. G. Dahlquist (1977), "On the relation of G-stability to other stability concepts for linear multistep methods", in "Topics in Numerical Analysis III, ed. J. J. Miller, Academic Press.

30. G. Dahlquist (1978), "On accuracy and unconditional stability of linear multistep methods for second order differential equations", BIT, 18, pp. 133-136.

31. J. Douglas, JR. (1956), "The solution of diffusion equation by a high order correct difference equation", J. Math. Phys, 35, pp. 145-151, MR 19, 884.

32. E. C. Du Fort and S. P. Frankel (1953), "Stability conditions in the numerical treatment of parabolic partial differential equations", MTAC, 7, pp. 135-152, MR 15, 474.

33. B. L. Ehle (1968), "High order A-stable methods for the numerical solution of system of differential equations", BIT, 8, pp. 276-278.

34. B. L. Ehle (1969), "On Padé approximations to the exponential function and A-stable methods for the numerical solution of initial-value problems", Re. Rept. CSRR 2010. University of Waterloo, Canada.

35. L. W. Ehrlich (1973), "Solving the biharmonic equation in a square: A direct versus a semidirect method", Comm. ACM, 16, pp. 711-714.

36. D. J. Evans (1965), "A stable explicit method for the finite difference solution of fourth order parabolic partial differential equation", Comp. J., 8, pp. 280-287.

37. G. Fairweather (1978), "A note on the efficient implementation of certain Padé methods for linear parabolic problems", BIT, 18, pp. 106-109.

38. G. Fairweather, A. R. Gourlay and A. R. Mitchell (1967), "Some high accuracy difference schemes with a splitting operator for equations of parabolic and elliptic type", Numer. Math., 10, pp. 56-66.

39. G. Fairweather and A. R. Gourlay (1967), "Some stable difference approximations to a fourth order parabolic partial differential equation", Math. Comp., 21, pp. 1-11.

40. I. Fried (1979), "Numerical Solution of Differential Equations", Academic Press, London.

41. W. Gautschi (1961), "Numerical integration of ordinary differential equations based on trigonometric polynomial", Numer. Math, 3, 381-397.

42. C. W. Gear (1971), "Numerical Initial Value Problems in Ordinary Differential Equations", Prentice Hall, Englewood Cliffs, New Jersey.

43. M. Goldberg and E. Tadmore (1978), "Scheme-independent stability criteria for difference approximations of hyperbolic initial-boundary value problems.I", Math. Comp., 32, pp. 1097-1107.

44. D. J. Gorman (1975), "Free Vibration Analysis of Beams and Shafts", John Wiley, New York.

45. A. R. Gourlay and J. Ll. Morris (1968), "Finite difference methods for nonlinear hyperbolic systems", Math. Comp., 22, pp. 28-39.

46. A. R. Gourlay and J. Ll. Morris (1980), "The extrapolation of first order methods for parabolic partial differential equations II", SIAM J. Numer. Anal., 17(5), pp. 641-655.

47. A. R. Gourlay and J. Ll. Morris (1981), "Linear Combinations of Generalized Crank-Nicolson schemes", IMA J. Numer. Anal., 1, pp. 347-357.

48. P. R. Graves-Morris. (ed) (1973), "Padé Approximants and their Applications", Academic Press, London.

49. B. Gustafsson (1975), "The convergence rate for difference approximations to mixed initial-boundary value problems", Maths. Comp., 29, pp. 396-406.

50. B. Gustafsson, H.-O. Kreiss and A. Sundström (1972), "Stability theory of difference approximations for mixed initial-boundary value problems. II", Maths. Comp., 26, pp. 649-686.

51. A. Hadjidimos (1971), "The numerical solution of model problem biharmonic equation by using extrapolated alternating direction implicit method", Ibid., 17, pp.301-317.

52. E. Hairer (1979), "Unconditionally stable methods for second order differential equations", Numer. Math., pp. 373-379.

53. C. A. Hall and T. A. Porsching (1980), "Padé approximants, fractional step methods and Navier-Stokes discretizations", SIAM J. Numer. Anal., 17(6), pp. 840-851.

54. G. Hall and J. M. Watt. (eds) (1976), "Modern Numerical Methods For Ordinary Differential Equations", Clarendon Press, Oxford.

55. P. Henrici (1962), "Dicrete Variable Methods in Ordinary Differential Equations", John Wiley and sons.

56. R. S. Hirsch and D. H. Rudy (1974), "The role of diagonal dominance and the cell Reynolds number in implicit difference methods for fluid mechanics problems", J. Comp. Phys, 16, pp. 304-310.

57. M. K. Jain, R. K. Jain, U. A. Krishnaiah (1979), "P-stable methods for periodic initial value problems of second order differential equations", BIT, 19, pp. 347-355.

58. A. Q. M. Khaliq and E. H. Twizell (1982), "The extrapolation of stable finite difference schemes for first order hyperbolic equations", Intern. J. Computer Math., 11, pp. 155-167.

59. H. -O. Kreiss and J. Oliger (1972), "Comparison of accurate methods for the integration of hyperbolic equations", Tellus, 24, pp.199-215.

60. J. D. Lambert (1973), "Computational Methods in Ordinary Differential Equations", John Wiley, Chichester.

61. J. D. Lambert (1980), "Stiffness", in "Computational Techniques for Ordinary Differential Equations" (eds. I. Gladwell and D. K. Sayers), Academic Press, London.

62. J. D. Lambert and A. R. Mitchell (1962), "On the solution of $y' = f(x,y)$ by a class of high accuracy difference formulae of low order", Z. Angew. Math. Phys, 13, pp. 223-232.

63. J. D. Lambert and I. A. Watson (1976), "Symmetric multistep methods for periodic initial value problems", J. Inst. Maths. Applics., 18, pp. 189-202.

64. J. D. Lawson and B. L. Ehle (1970),"Asymptotic error estimation for one-step methods based on quadrature", Aeq. Math, 5, pp. 236-246.

65. J. D. Lawson and J. Ll. Morris (1978), "The extrapolation of first order methods for parabolic partial differential equations I", SIAM J. Numer. Anal., 15(6), pp. 1212-1224.

66. J. D. Lawson and D. Swayne (1976), "A simple efficient algorithm for the solution of heat conduction problems", Proceedsings, Sixth Manitoba Conference on Numerical Mathematics, pp. 239-250.

67. M. Lees (1960), "A priori estimates for the solution of difference approximations to parabolic partial differential equations", Duke Math. J., 27, pp. 297-312, MR22, 12725.

68. M. Lees (1961), "Alternating direction and semi-explicit difference methods for parabolic partial differential equations", Numer. Math, 3, pp. 24-47, MR 27, 5376.

69. B. Lindberg (1971), "On smoothing and extrapolation for the trapezoidal rule", BIT, 11, pp. 29-52.

70. W. Liniger and R. A. Willoughby (1967), "Efficient numerical integration methods for stiff systems of differential equations", IBM Research Report RC-1970, Thomas J. Watson Research Centre, Yorktown Heights, N.Y.

71. W. E. Milne (1949), "A note on the numerical integration of differential equations", J. Res.Nat. Bur. Standards., 43, pp. 537-542.

72. A. R. Mitchell (1969), "Computational Methods in Partial Differential Equations", John Wiley, Chichester.

73. A. R. Mitchell and D. F. Griffiths (1980), "The Finite Difference Method in Partial Differential Equations", John Wiley, Chichester.

74. K. W. Morton (1980), "Stability of finite difference approximations to a diffusion-convection equation", Int. J. num. Meth. Engng, 12, pp.899-916.

75. S. P. Norsett (1975), "Numerical Solution of Ordinary Differential Equations", Ph. D. Thesis, University of Dundee, U.K.

76. S. P. Norsett (1978), "Restricted Padé approximations to the exponential function", SIAM J. Numer. Anal., 15, pp. 1008-1029.

77. N. Obrechkoff (1942), "Sur les quadrature mecaniques", Spisanic Bulgar. Akad. Nauk., 65, pp. 191-289 (in Bulgarian, French summary), (Reviewed in Math. Rev., 10, p.70.)

78. C. G. O'Brien, M. A. Hyman and S. Kaplan (1951), "A study of the numerical solution of partial differential equations", J. Math. Phys, 29, pp. 223-251.

79. J. Oliger (1974), "Fourth order difference methods for the initial boundary-value problem for hyperbolic equations", Math. Comp., 28, pp. 15-25.

80. M. H. Padé (1892), "Sur la répresentation approchée d'une fonction

par des fractions rationelles, Ann. de l'Ecole Normale Supérieure", Vol. 9 (Suppl.).

81. D. W. Peaceman and H. H. Rachford (1955), "The numerical solution of parabolic and elliptic differential equations", J. Soc. Indust. Appl. Math., 3, pp. 28-41.

82. H. S. Price, R. S. Varga and J. E. Warren (1966), "Application of oscillation matrices to diffusion - convection equation", J. Math. Phys., 45, pp. 301-311.

83. R. D. Richtmyer (1957), "Difference methods for initial value problems, Interscience Tracts in Pure and Applied Mathematics", Tract 4, Interscience, New York, MR 20, 438.

84. R. D. Richtmyer (1963), "A survey of finite difference methods for nonsteady fluid dynamics", NCAR Research Tech. notes 63-2.

85. R. D. Richtmyer and K. W. Morton (1967), "Difference Methods for Initial-Value Problems", John Wiley, New York.

86. H. H. Rosenbrock (1963), "Some general implicit processes for the numerical solution of differential equations", Comp. J, 5, pp. 329-330.

87. L. F. Shampine and M. K. Gordon (1973), "Computer Solution of Ordinary Differential Equations", W. H. Freeman and Company, San Francisco.

88. G. D. Smith (1978), "Numerical Solution of Partial Differential Equations: Finite Difference Methods", 2nd edn, Clarendon Press, Oxford.

89. I. M. Smith, J. L. Siemieniuch and I. Gladwell (1973), "A comparison of old and new methods for large systems of ordinary differential equations arising from parabolic partial differential equations", Numerical Analysis Research Report no. 13, Mathematics Department, University of Manchester, Manchester, England.

90. P. Smith and E. H. Twizell (1982), "The extrapolation of Padé approximants in the closed-loop simulation of human thermoregulation", Appl. Math. Modelling, 6, pp. 81-91.

91. E. Stiefel and D. G. Bettis (1969), "Stabilization of Cowell's methods", Numer. Math., 13, pp. 154-175.

92. G. J. Tee (1975), "A novel finite-difference approximation to the biharmonic operator", Comp. J, 6, pp. 177-192.

93. C. E. Thompson (1968), "Solution of linear differential equations", Comp. J, 10, pp. 417-418.

94. E. H. Twizell (1979), "An explicit difference method for the wave equation with extended stability range", BIT, 19, pp. 378-383.

95. E. H. Twizell (1981), "The numerical solution of the wave equation at the first time step", Brunel University Department of Mathematics Technical Report, TR/13/81.

96. E. H. Twizell and A. Q. M. Khaliq (1981), "One step multi-derivative methods for first order ordinary differential equations", BIT, 21, pp. 518-527.

97. E. H. Twizell and P. S. Smith (1981), "A numerical study of heat flow in elliptic cylinder with interior derivative boundary conditions", Brunel University Department of Mathematics Technical Report, TR/11/81.

98. E. H. Twizell and P. S. Smith (1982), "Numerical modelling of heat flow in the human torso I: finite difference methods", in J. Caldwell and A. O. Moscardini (eds.), "Numerical Modelling in Diffusion-Convection", Pentech Press, Plymouth, pp. 165-189.

99. R. S. Varga (1961), "On high order stable implicit methods for solving parabolic partial differential equations", J. Math. Phys, 40, pp. 220-231.

100. R. S. Varga (1962), "Matrix Iterative Analysis", Prentice Hall, Englewood Cliffs, N.J.

101. W. L. Wood and R. W. Lewis (1975), "A Comparison of Time-Marching Schemes for the Transient Heat Conduction Equation", Internat. J. Numer. Method. Engrg., 9, pp. 679-689.