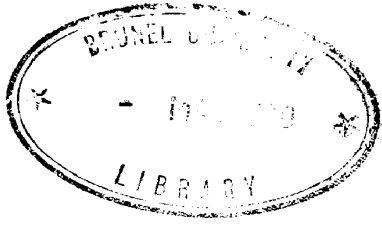


TR/92

January 1980

Explicit and implicit methods for
second order ordinary differential
equations.

E.H. Twizell



W9260323

0. Abstract

A family of explicit formulas is developed for solving a system of second order linear ordinary differential equations with constant coefficients and with initial conditions specified. A family of implicit formulas for solving the same system with specified boundary conditions is also developed.

Both families are based on Padé approximants to the exponential function and for each formula developed the order of the formula is seen to be one higher than the order of the Padé approximant used. In the case of the family of implicit formulas it is seen that the order of the formula is made arbitrarily high by using an appropriate Padé approximant.

It is shown that the families are readily applicable to the numerical solution of second order hyperbolic partial differential equations with constant coefficients.

The formulas developed are tested on four problems.

(1)

1. Explicit Methods For Initial Value Problems

Given the linear system of N second order ordinary differential equations

$$(0) \quad \underline{y}''(t) = A \underline{y}(t) ,$$

where A is a square matrix of order N which has constant elements, with initial conditions

$$(1) \quad \underline{y}(0) = \underline{g} , \underline{y}'(0) = \underline{\phi} ,$$

the solution may be shown to be of the form

$$(2) \quad \underline{y}(t) = \frac{1}{2} \exp(tB) (\underline{g} + B^{-1}\underline{\phi}) + \frac{1}{2} \exp(-tB) (\underline{g} - B^{-1}\underline{\phi}) ,$$

where B is a matrix such that $B^2 = A$.

The determination of the elements of B is a non-trivial exercise and an approximate method based on padé approximants to the exponential function is now developed which does not require the elements of B to be determined explicitly.

Suppose that a uniform discretization ℓ is superimposed on the independent variable t ; then it is easy to show that $\underline{y}(t)$ satisfies the recurrence relation

$$(3) \quad \begin{cases} \underline{y}(t + \ell) - \{\exp(\ell B) + \exp(-\ell B)\} \underline{y}(t) + \underline{y}(t - \ell) = \underline{0} \\ \underline{y}(0) = \underline{g} , \underline{y}(\ell) = \frac{1}{2} \{\exp(\ell B) + \exp(-\ell B)\} \underline{g} + \ell \underline{\phi} + \underline{0} \quad (\ell^3) \end{cases}$$

with $t = \ell, 2\ell, \dots$.

Any numerical solution of (3) will rely for its overall accuracy on the approximation to $\exp(\pm \ell B)$. For the (m, k) Padé approximant to $\exp(\ell B)$ of the form

$$R_{m,k}(\ell B) = (Q_m(\ell B))^{-1} P_k(\ell B) ,$$

(2)

where P_k, Q_m are matrix polynomials of degrees k, m respectively,
this leads, for $m+k$ even,

$$(4) \quad \exp(\ell B) + \exp(-\ell B) = 2 \left\{ I + \frac{\ell^2 B^2}{2!} + \frac{\ell^4 B^4}{4!} + \dots + \frac{\ell^{m+k} B^{m+k}}{(m+k)!} \right\} + O(\ell^{m+k+2})$$

$$= 2 \left\{ I + \frac{\ell^2 A}{2!} + \frac{\ell^4 A^2}{4!} + \dots + \frac{\ell^{m+k} A^{\frac{1}{2}(m+k)}}{(m+k)!} \right\} + O(\ell^{m+k+2})$$

since $B^2 = A$ and for $m+k$ odd, to

$$(5) \quad \exp(\ell B) + \exp(-\ell B) = 2 \left\{ I + \frac{\ell^2 B^2}{2!} + \frac{\ell^4 B^4}{4!} + \dots + \frac{\ell^{m+k-1} B^{m+k-1}}{(m+k-1)!} \right\} + O(\ell^{m+k+1})$$

$$= 2 \left\{ I + \frac{\ell^2 A}{2!} + \frac{\ell^4 A^2}{4!} + \dots + \frac{\ell^{m+k-1} A^{\frac{1}{2}(m+k-1)}}{(m+k-1)!} \right\} + O(\ell^{m+k+1})$$

In choosing which Padé approximant to use it must be noted in (3) that

$\underline{y}(\ell)$ is only second order accurate, (that is, its error is $O(\ell^3)$),

so that Padé approximants should be chosen for which $m+k \leq 2$. For

$m+k > 2$ a more accurate approximation to $\underline{y}(\ell)$ must be found otherwise the use of a higher order approximant is unjustified.

A table of fifteen Padé approximants to e^θ , where θ is some real scalar, together with the principal error term and the range of values of θ for which the approximation converges, is given in Table I.

These Padé approximants will be used in the derivation of fifteen allied methods (given Roman numerals) in this and in Section 2 of the paper.

The matrix analog in all fifteen cases is obvious, and a bound on $\ell \|B\|_S$, where $\|B\|_S$ is the spectral norm of B , is given by the modulus of the smaller bound on θ .

Referring to Table I it is clear that Methods I, II, III, IV, VII all have

$m+k \leq 2$ and may be used in (3) to give an explicit numerical solution to the system (0) with initial conditions (1).

Using $\underline{y}(t)$ to distinguish a computed solution from the theoretical solution $\underline{y}(t)$, defining the local discretization error to be $\underline{y}(t) - \underline{y}(t)$, and using E to denote the error predicted by (4) or (5), these five explicit algorithms with $t = \ell, 2\ell, \dots$ are as follows:

(3)

Method I : (0, 1) Padé approximant ; $E=O(\ell^2)$.

$$(6) \quad \begin{cases} \tilde{Y}(t+\ell) - 2\tilde{Y}(t) + \tilde{Y}(t-\ell) = 0, \\ \tilde{Y}(0) = \tilde{g}, \quad \tilde{Y}(\ell) = \tilde{g} + \ell\phi + O(\ell^2). \end{cases}$$

Local error = $\ell^2 \tilde{y}''(t) + \frac{1}{12} \ell^4 \tilde{y}^{(iv)}(t) + \dots$

Method II : (1,0) Padé approximant ; $E=O(\ell^2)$

Same explicit algorithm and local error as Method I

Method III : (1,1) Padé approximant ; $E=O(\ell^4)$.

$$(7) \quad \begin{cases} \tilde{Y}(t+\ell) - (2I + \ell^2 A) \tilde{Y}(t) + \tilde{Y}(t-\ell) = 0, \\ \tilde{Y}(0) = \tilde{g}(\ell) = \frac{1}{2} (2I + \ell^2 A) \tilde{g} + \ell\phi + O(\ell^3). \end{cases}$$

Local error = $\frac{1}{12} \ell^4 \tilde{y}^{(iv)}(t) + \frac{1}{360} \ell^6 \tilde{y}^{(iv)}(t) + \dots$

Method IV : (0,2) Padé approximant ; $E=O(\ell^4)$.

same explicit algorithm and local error as Method III.

Method VII : (2,0) Padé approximant ; $E=O(\ell^4)$.

Same explicit algorithm and local error as Method III.

Replacing $y(t)$ with $\tilde{y}(t)$, equations (6) and (7) may be written in the forms

$$(8) \quad L[\tilde{y}(t); (\ell)] = \sum_{j=-1}^1 [\alpha_j \tilde{y}(t+j\ell) - \ell^2 \beta_j \tilde{y}''(t+j\ell)]$$

or

$$(9) \quad L[\tilde{y}(t); \ell] = C_0 \tilde{y}(t) + C_1 \tilde{y}'(t) + C_2 \tilde{y}''(t) + \dots$$

assuming that $\tilde{y}(t)$ has as many derivatives as required on some closed interval.

Following Henrici [2] , equation (9) is said to have order r if

$$(10) \quad C_0 = C_1 = \dots = C_r = C_{r+1} = 0, \quad C_{r+2} \neq 0.$$

and to be consistent with (0) if $r \geq 1$.

$$(4)$$

Equation (7) is thus consistent, equation (6) is not. Defining the first characteristic polynomial of (8) to be

$$(11) \quad \alpha_{-1} + \alpha_0 \xi + \ell_1 \xi^2 = 0,$$

formulas (6) and (7) are seen to be zero-stable, for in both cases (11) has a double root at $\xi = 1$ and no other roots.

Equation (7) is therefore seen to be convergent while equation (6) is not convergent.

2. Implicit Methods For Boundary Value Problems

Given, as before, the linear system of N second order ordinary differential equations (0) with, now, the boundary conditions

$$(12) \quad \underline{y}(0) = \underline{g}_0, \quad \underline{y}(T) = \underline{g}_1,$$

the solution may be shown to be of the form

$$(13) \quad \underline{y}(t) = \exp(tB) \{ \exp(TB) - \exp(-TB) \}^{-1} \{ \underline{g}_1 - \exp(-TB) \underline{g}_0 \} \\ + \exp(-tB) \{ \exp(TB) - \exp(-TB) \}^{-1} \{ \exp(TB) \underline{g}_0 - \underline{g}_1 \}$$

It is easy to show that the solution $\underline{y}(t)$ as given by equation (13) satisfies the recurrence relation

$$(14) \quad \left\{ \begin{array}{l} \underline{y}(t+\ell) - \{ \exp(\ell B) + \exp(-\ell B) \} \underline{y}(t) + \underline{y}(t-\ell) = \underline{0}, \\ \underline{y}(0) = \underline{g}_0, \quad \underline{y}(T) = \underline{g}_1. \end{array} \right.$$

Any numerical solution of (14) will determine the vector $\underline{y}(t)$ implicitly and its accuracy will depend on the approximations to $\exp(\pm \ell B)$. Using (m, k) Padé approximants, expressions for $\{ \exp(\ell B) + \exp(-\ell B) \}$ are given in terms of powers of ℓ and the known matrix A in (4) and (5). The restriction $m+k \leq 2$ may be relaxed for the implicit relation (14).

(5)

Suppose that the interval $0 \leq t \leq T$ is discretized into $M+1$ subdivisions using a time step ℓ , then $(M+1)\ell = T$ and the solution of (0) with (14) will be computed at the M points $t_i = i\ell$ ($i = 1, 2, \dots, M$).

The implicit algorithms yielded by the fifteen methods of Table I are now derived. In each case E again denotes the error in $\{\exp(\ell B) + \exp(-B\ell)\}$ predicted by (4) or (5), the local discretization error is given by $\underline{y}(t) - \underline{Y}(t)$ and $t = t_1, t_2, \dots, t_M$

(clearly $t_0 = 0$ and $t_{M+1} = T$)

Method I : (0,1) Padé approximant ; $E = O(\ell^2)$.
No implicit algorithm

Method II : (1,0) Padé approximant ; $E = O(\ell^2)$.

$$(15) \quad \begin{cases} (I - \ell^2 A) \underline{Y}(t + \ell) - 2 \underline{Y}(t) + (I - \ell^2 A) \underline{Y}(t - \ell) = \underline{0}, \\ \underline{y}(0) = \underline{g}_0, \quad \underline{Y}(T) = \underline{g}_1. \\ \text{Local error} = -\ell^2 \underline{y}''(t) - \frac{11}{12} \ell^4 \underline{y}^{(iv)}(t) - \dots \end{cases}$$

Method III : (1,1) Padé approximant ; $E = O(\ell^4)$.

$$(16) \quad \begin{cases} (I - \frac{1}{4} \ell^2 A) \underline{Y}(t + \ell) - (2I + \frac{1}{2} \ell^2 A) \underline{Y}(t) + (I - \frac{1}{4} \ell^2 A) \underline{Y}(t - \ell) = \underline{0}, \\ \underline{Y}(0) = \underline{g}_0, \quad \underline{Y}(T) = \underline{g}_1. \\ \text{Local error} = -\frac{1}{6} \ell^4 \underline{y}^{(iv)}(t) - \frac{13}{720} \ell^6 \underline{y}^{(iv)}(t) - \dots \end{cases}$$

Method IV : (0,2) Padé approximant ; $E = O(\ell^4)$.

No implicit algorithm.

Method V : (1,2) Padé approximant ; $E = O(\ell^4)$.

$$(17) \quad \begin{cases} (I - \frac{1}{9} \ell^2 A) \underline{Y}(t + \ell) - (2I + \frac{7}{9} \ell^2 A) \underline{Y}(t) + (I - \frac{1}{9} \ell^2 A) \underline{Y}(t - \ell) = \underline{0}, \\ \underline{Y}(0) = \underline{g}_0, \quad \underline{Y}(T) = \underline{g}_1. \\ \text{Local error} = -\frac{1}{36} \ell^4 \underline{y}^{(iv)}(t) - \frac{7}{1080} \ell^6 \underline{y}^{(iv)}(t) - \dots \end{cases}$$

(6)

Method VI: (2,1) Padé approximant ; $E=0(\ell^4)$

$$(18) \left\{ \begin{array}{l} (I - \frac{1}{9}\ell^2 A + \frac{1}{36}\ell^4 A^2) \underline{Y}(t+\ell) - (2I + \frac{7}{9}\ell^2 A) \underline{Y}(t) + (I - \frac{1}{9}\ell^2 A + \frac{1}{36}\ell^4 A^2) \underline{Y}(t-\ell) = \underline{0}, \\ \underline{Y}(0) = \underline{g}_0, \quad \underline{Y}(T) = \underline{g}_1 \end{array} \right. .$$

$$\text{Local error} = \frac{1}{36}\ell^4 \underline{y}^{(iv)}(t) + \frac{23}{1080}\ell^6 \underline{y}^{(vi)}(t) + \dots$$

Method VII: (2,0) pade approximant t ; $E = 0(\ell^4)$.

$$(19) \left\{ \begin{array}{l} (I + \frac{1}{4}\ell^4 A^2) \underline{Y}(t+\ell) - (2I + \ell^2 A) \underline{Y}(t) + (I + \frac{1}{4}\ell^4 A^2) \underline{Y}(t-\ell) = \underline{0}, \\ \underline{Y}(0) = \underline{g}_0, \quad \underline{Y}(T) = \underline{g}_1 \end{array} \right. .$$

$$\text{Local error} = \frac{1}{3}\ell^4 \underline{y}^{(iv)}(t) + \frac{91}{360}\ell^6 \underline{y}^{(vi)}(t) + \dots$$

Method VIII : (2,2) Padé approximant ; $E=0(\ell^6)$.

$$(20) \left\{ \begin{array}{l} (I - \frac{1}{12}\ell^2 A + \frac{1}{144}\ell^4 A^2) \underline{Y}(t+\ell) - (2I + \frac{5}{6}\ell^2 A + \frac{1}{72}\ell^4 A^2) \underline{Y}(t) \\ \quad + (I - \frac{1}{12}\ell^2 A + \frac{1}{144}\ell^4 A^2) \underline{Y}(t-\ell) = \underline{0}, \\ \underline{Y}(0) = \underline{g}_0, \quad \underline{Y}(r) = \underline{g}_{-1} \end{array} \right. .$$

$$\text{Local error} = \frac{1}{360}\ell^6 \underline{y}^{(vi)}(t) + \frac{37}{75600}\ell^8 \underline{y}^{(viii)}(t) + \dots$$

Method IX: (0,3) padé approximant ; $E=0(\ell^4)$.
No implicit algorithm.Method X: (1,3) Padé approximant ; $E=0(\ell^6)$.

$$(21) \left\{ \begin{array}{l} (I - \frac{1}{16}\ell^2 A) \underline{Y}(t+\ell) - (2I + \frac{7}{8}\ell^2 A + \frac{1}{48}\ell^4 A^2) \underline{Y}(t) + (I - \frac{1}{16}\ell^2 A) \underline{Y}(t-\ell) = \underline{0}, \\ \underline{Y}(0) = \underline{g}_0, \quad \underline{Y}(T) = \underline{g}_1 \end{array} \right. .$$

$$\text{Local error} = -\frac{7}{2880}\ell^6 \underline{y}^{(vi)}(t) - \frac{1}{8060}\ell^8 \underline{y}^{(viii)}(t) - \dots$$

Method X^I : (2, 3) Padé approximant ; $E=0(\ell^6)$.

$$(22) \left\{ \begin{array}{l} (I - \frac{3}{50}\ell^2 A + \frac{1}{400}\ell^4 A^2) \underline{Y}(t+\ell) - (2I + \frac{22}{25}\ell^2 A + \frac{17}{600}\ell^4 A^2) \underline{Y}(t) \\ \quad + (I - \frac{3}{50}\ell^2 A + \frac{1}{400}\ell^4 A^2) \underline{Y}(t-\ell) = \underline{0}, \\ \underline{Y}(0) = \underline{g}_0, \quad \underline{Y}(T) = \underline{g}_1 \end{array} \right. .$$

$$\text{Local error} = \frac{1}{3600}\ell^6 \underline{y}^{(vi)}(t) + \frac{23}{252000}\ell^8 \underline{y}^{(viii)}(t) + \dots$$

(7)

Method XII : (3,2) Padé approximant ; $E=0(\ell^6)$.

$$(23) \left\{ \begin{array}{l} \left(I - \frac{3}{50} \ell^2 A + \frac{1}{400} \ell^4 A^2 - \frac{1}{3600} \ell^6 A^3 \right) \tilde{Y}(t+\ell) - \left(2I + \frac{22}{25} \ell^2 A + \frac{17}{600} \ell^4 A^2 \right) \tilde{Y}(t) \\ + \left(I - \frac{3}{50} \ell^2 A + \frac{1}{400} \ell^4 A^2 - \frac{1}{3600} \ell^6 A^3 \right) \tilde{Y}(t-\ell) = 0, \\ \tilde{Y}(0) = \underline{g}_0, \quad \tilde{Y}(T) = \underline{g}_1. \end{array} \right.$$

$$\text{Local error} = \frac{1}{3600} \ell^6 \tilde{y}^{(vi)}(t) + \frac{23}{252000} \ell^8 \tilde{y}^{(viii)}(t) - \dots$$

Method XIII : (3,1) Padé approximant ; $E=0(\ell^6)$.

$$(24) \left\{ \begin{array}{l} \left(I - \frac{1}{16} \ell^2 A - \frac{1}{576} \ell^6 A^3 \right) \tilde{Y}(t+\ell) - \left(2I + \frac{7}{8} \ell^2 A + \frac{1}{48} \ell^4 A^2 \right) \tilde{Y}(t) + \left(I - \frac{1}{16} \ell^2 A - \frac{1}{576} \ell^6 A^3 \right) \tilde{Y}(t-\ell) = \\ \tilde{Y}(0) = \underline{g}_0, \quad \tilde{Y}(T) = \underline{g}_1. \end{array} \right.$$

$$\text{Local error} = -\frac{17}{2880} \ell^6 \tilde{y}^{(vi)}(t) - \frac{5}{2688} \ell^8 \tilde{y}^{(viii)}(t) - \dots$$

Method XIV : (3,0) Padé approximant ; $E=0(\ell^4)$.

$$(25) \left\{ \begin{array}{l} \left(I - \frac{1}{12} \ell^4 A^2 - \frac{1}{36} \ell^6 A^3 \right) \tilde{Y}(t+\ell) - (2I + h^2 A) \tilde{Y}(t) + \left(I - \frac{1}{12} \ell^4 A^2 - \frac{1}{36} \ell^6 A^3 \right) \tilde{Y}(t-\ell) = 0, \\ \tilde{Y}(0) = \underline{g}_0, \quad \tilde{Y}(T) = \underline{g}_1. \end{array} \right.$$

$$\text{Local error} = -\frac{1}{12} \ell^4 \tilde{y}^{(iv)}(t) - \frac{49}{360} \ell^6 \tilde{y}^{(vi)}(t) - \dots$$

Method XV : (3,3) padé approximant ; $E=0(\ell^8)$

$$(26) \left\{ \begin{array}{l} \left(I - \frac{1}{20} \ell^2 A + \frac{1}{600} \ell^4 A^2 - \frac{1}{14400} \ell^6 A^3 \right) \tilde{Y}(t+\ell) \\ - \left(2I - \frac{9}{10} \ell^2 A + \frac{11}{300} \ell^4 A^2 + \frac{1}{7200} \ell^6 A^3 \right) \tilde{Y}(t) \\ + \left(I - \frac{1}{20} \ell^2 A + \frac{1}{600} \ell^4 A^2 - \frac{1}{14400} \ell^6 A^3 \right) \tilde{Y}(t-\ell) = 0, \\ \tilde{Y}(0) = \underline{g}_0, \quad \tilde{Y}(T) = \underline{g}_1. \end{array} \right.$$

$$\text{Local error} = -\frac{1}{50400} \ell^8 \tilde{y}^{(viii)}(t) - \frac{1}{324000} \ell^{10} \tilde{y}^{(x)}(t) - \dots$$

Examination of their local error expressions shows that algorithms (16) through (26) are all consistent but that (15) is not consistent. The first characteristic polynomials of all twelve algorithms have double roots at $\xi = 1$ and no other roots, thus satisfying the zero-stability criterion. Consequently only Method II based on the (1, 0) Padé approximant fails the convergence criterion (see Lambert [3]).

(8)

The results obtained are characterised with regard to the principal part of the local error and are summarised in Table II, Of the methods with $O(\ell^4)$ local error Method V based on the (1,2) Padé approximant has smallest principal error term and is the most economical in the sense that no powers of A higher than A itself are used. Of the methods with $O(\ell^6)$ local error, Methods XI and XII based respectively on the (2,3) and (3,2) Padé approximants both have the smallest modulus principal error term, but Method XI requires only A^2 whilst Method XII requires A^3 ; Method XI is thus the most economical of the methods with $O(\ell^6)$ local error.

Every one of the implicit schemes is of the form

$$(27) \quad \tilde{C}Y(t-\ell) + DY(t) + CY(t+\ell) = \tilde{0}$$

where C,D are square matrices of order N . More precisely, C and D are band matrices, the band width depending on the powers of the matrix A. Applying to the M points t_i ($i=1, \dots, M$) leads to the system of linear equations of order MN given by

$$(28) \quad \begin{bmatrix} -D & C & 0 & \dots & 0 \\ C & -D & C & & \\ 0 & C & -D & C & \dots \\ \vdots & & & & \ddots \\ 0 & & C & -D & C \\ 0 & & 0 & C & -D \end{bmatrix} \begin{bmatrix} \tilde{y}(t_1) \\ \tilde{y}(t_2) \\ \tilde{y}(t_3) \\ \vdots \\ \tilde{y}(t_{m-1}) \\ \tilde{y}(t_m) \end{bmatrix} + \begin{bmatrix} c\tilde{g}_0 \\ \tilde{0} \\ 0 \\ \vdots \\ 0 \\ c\tilde{g}_1 \end{bmatrix} = \begin{bmatrix} \tilde{0} \\ \tilde{0} \\ \tilde{0} \\ \vdots \\ \tilde{0} \\ \tilde{0} \end{bmatrix}$$

where $\tilde{0}$ is the zero matrix of order $N \times N$ and $\tilde{0} = (0, 0, \dots, 0)^T$ is the zero vector of order $N \times 1$.

System (28) is solved using any of the methods designed for block diagonal systems (see, for example Goult et al [I]).

3. Application to Hyperbolic Partial Differential Equations

Given the wave equation

$$(29) \quad \frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$$

over a region $R = \{[0 < x < 1] \times [t > 0]\}$ with boundary conditions

$$(30) \quad u(0, t) = u(1, t) = 0, \quad t > 0$$

and initial conditions

$$(31) \quad u(x, 0) = g, \quad \frac{\partial u}{\partial t}(x, 0) = \phi,$$

one approach is to replace the second order space derivative with the finite difference approximation

$$(32) \quad \frac{\partial^2 u}{\partial x^2} = \{u(x+h, t) - 2u(x, t) + u(x-h, t)\}/h^2 + O(h^2)$$

at every time step. If the space interval $0 \leq x \leq 1$ is divided into

N subintervals each of width h , and if $\underline{U} = (U_1, U_2, \dots, U_N)^T$

is the vector of computed values of u at a given time level, then (29) becomes

$$(33) \quad \frac{d^2 \underline{U}}{dt^2} = A \underline{U}(t)$$

where A is the matrix given by

$$(34) \quad A = \frac{1}{h^2} \begin{bmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & & \\ 0 & 1 & -2 & 1 & \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 1 & -2 \end{bmatrix}$$

It was shown in Twizell [4] that the solution of (33) satisfies the recurrence relation

(10)

$$(35) \begin{cases} \underline{U}(t+\ell) - \{\exp(\ell B) + \exp(-\ell B)\} \underline{U}(t) + \underline{U}(t-\ell) = \underline{0}, \\ \underline{U}(0) = \underline{g}, \quad \underline{U}(\ell) = \frac{1}{2} \{\exp(\ell B) + \exp(-\ell B)\} \underline{g} + \ell \underline{\phi} + O(\ell^3) \end{cases}$$

where B is a matrix such that $B^2 = A$.

The relation (35) is analogous to (5) and in solving numerically, the choice of Padé approximant to $\exp(\pm \ell B)$ controls the stability range of p , where $p = \ell/h$. If the (1,1) Padé approximant is used, relation (34) yields the well known five point explicit scheme which is stable for $0 \leq p \leq 1$; if the (2,2) Padé approximation is used, relation (34), with an improved approximation to $\underline{U}(\ell)$, yields the seven point explicit scheme developed in Twizell [4] which extends the stability range to $0 \leq p \leq \sqrt{3}$.

In simulating arterial blood pressure, where the heart beats every T seconds, the conditions (30),(31) are replaced by

$$(36) \begin{cases} u(0, t) = f_0(t), \quad u(L, t) = f_1(t), \\ u(x, 0) = u(x, T) = g(x) \end{cases}$$

Where $u(x, t)$ ($0 \leq x \leq L$; $0 \leq t \leq T$) is the arterial pressure and L is the length of the artery which is assumed to be a thin elastic cylindrical tube.

Replacement of the space derivative in (29) with (32) again leads to the linear system of second order ordinary differential equations (33) whose solution is now seen to satisfy the relation

$$(37) \begin{cases} \underline{U}(t+\ell) - \{\exp(\ell B) + \exp(-\ell B)\} \underline{U}(t) + \underline{U}(t-\ell) = \underline{0}, \\ \underline{U}(0) = \underline{U}(T) = \underline{g} \end{cases}$$

Relation (37) is analogous to (16) and may be solved by adapting any of the implicit algorithms (15) through (26).

4. Numerical Results

To examine the behaviour of the explicit and implicit formulas of sections 1 and 2, three problems were solved using the explicit methods I and III of section 1 and the twelve implicit methods given by the recurrence relations (15) through (26) of section 2.

Problem 1

$$y'' = y$$

Here $N=1$, and assuming the solution $y=e^t + e^{-t}$, initial conditions were specified as

$$y(0) = 2 ; y'(0) = 0 .$$

The theoretical solution for $t = 0.0(0.1)1.0$ is depicted in Fig:1 together with the computed solution using the explicit method based on the (0,1) or (1,0) Padé approximants. Fig. 1 shows clearly that this method is not consistent with the differential equation.

In Fig:2 the error modulus is graphed against t for $t=0.0(0.1)1.0$ using the explicit formula based on the (1,1), (0,2) or (2,0) Padé approximants. The theoretical solution for $t=1.0$ is $y= 3.086$.

Boundary conditions were specified as

$$y(0) = 2 ; y(1) = e + e^{-1}$$

and the error moduli for $t=0.0(0.1)1.0$ are given in Table III for the twelve implicit formulas given in equations (15) through (26). It was seen in section 2 that the method based on the (1,0) Padé approximant was not consistent with the differential equation, that the methods based on the (1,2) and (2,3) Padé approximants are the most economical and have the most favourable principal error terms of all the methods having $O(\ell^4)$ and $O(\ell^6)$ error terms respectively, and that the methods based on the (2,0) and (3,1) Padé approximants have the least favourable principal error terms of all the methods having $O(\ell^4)$ and $O(\ell^6)$ error terms. These findings are all substantiated by the numerical results to Problem 1.

(12)

Problem 2

$$\begin{aligned} \ddot{y}_1 &= -2y_1 + y_2, \\ \ddot{y}_2 &= y_1 - 2y_2 \end{aligned}$$

Here $N = 2$ and the matrix of coefficients is the 2×2 analog of the matrix on the right hand side of equation (34). The matrix of coefficients has negative eigenvalues $\lambda_1 = -1$ and $\lambda_2 = -3$ and associated eigenvectors $\underline{c}_1 = (1,1)^T$ and $\underline{c}_2 = (1,1)^T$. The theoretical solution was taken to be

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \left\{ -\frac{1}{2} \cos t + \left(\frac{1 + \cos 1}{2 \sin 1} \right) \sin t \right\} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \left\{ \frac{1}{2} \cos \sqrt{3}t + \left(\frac{1 - \cos \sqrt{3}}{2 \sin \sqrt{3}} \right) \sin \sqrt{3}t \right\} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

and initial conditions were specified as

$$\begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}; \quad \begin{bmatrix} y_1'(0) \\ y_2'(0) \end{bmatrix} = \left(\frac{1 + \cos 1}{2 \sin 1} \right) \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \left(\frac{\sqrt{3}(1 - \cos \sqrt{3})}{2 \sin \sqrt{3}} \right) \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

The theoretical solution for $t = 0.0(0.1)1.0$ is shown in Fig:3, together with the computed solution using the explicit method based on the (0,1) or (1,0) Padé approximants. As for Problem 1, it is clear that this method is not consistent with the system of differential equations.

In Fig:4 the error for each component of $\underline{y}(t)$ is graphed against t for $t = 0.0(0.1)1.0$ using the explicit method based on the (1,1), (0,2) or (2,0) Padé approximants. The theoretical solution for $t = 1.0$ is $\underline{y} = (0, 0)^T$.

Boundary conditions were specified as

$$\underline{y}(0) = (0, -1)^T; \quad \underline{y}(1) = (1, 0)^T$$

and the errors in each component of \underline{y} were computed for $t = 0.0(0.1)1.0$ using the eleven consistent implicit methods of section 2.

As for Problem 1 it was found that the greatest errors were experienced for $t = 0.5$; these errors are given in Table IV. The theoretical solution for $t = 0.5$ is $\underline{y} \approx (0.77, -0.77)^T$.

All numerical results obtained for Problem 2 are in agreement with the theoretical results developed in sections 1 and 2.

(13)

Problem 3

$$\begin{aligned} y_1'' &= 2y_1 + y_2, \\ y_2'' &= y_1 + 2y_2. \end{aligned}$$

Again $N = 2$, and the matrix of coefficients has positive eigenvalues $\lambda_1 = 1$ and $\lambda_2 = 3$ with associated eigenvectors $\underline{c}_1 = (1, -1)^T$ and

$\underline{c}_2 = (1, 1)^T$. The theoretical solution was taken to be

$$\begin{aligned} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} &= \left\{ \frac{1-e^{-1}}{2(e-e^{-1})} e^t + \frac{(e-1)}{2(e-e^{-1})} e^{-t} \right\} \begin{bmatrix} 1 \\ -1 \end{bmatrix} \\ &+ \left\{ \frac{(e^{-\sqrt{3}}+1)}{2(e^{\sqrt{3}}-e^{-\sqrt{3}})} e^{\sqrt{3}t} - \frac{(e^{-\sqrt{3}}+1)}{2(e^{\sqrt{3}}-e^{-\sqrt{3}})} e^{\sqrt{3}t} \right\} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \end{aligned}$$

and initial conditions were specified as

$$\begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}; \quad \begin{bmatrix} y_1'(0) \\ y_2'(0) \end{bmatrix} = \frac{2-e-e^{-1}}{2(e-e^{-1})} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \frac{3(2+e^3+e^{-3})}{2(e^3-e^{-3})} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

The theoretical solution for $t = 0.0(0.1)1.0$ is graphed in Fig:5, together with the computed solution using the explicit method based on the (0,1) or (1,0) Padé approximants. In Fig:6 the error moduli for each component of $\underline{y}(t)$ is graphed against t for $t = 0.0(0.1)1.0$ using the explicit method based on the (1,1), (0,2) or (2,0) Padé approximants.

Boundary conditions were specified as

$$\underline{y}(0) = (0, -1)^T; \quad \underline{y}(1) = (1, 0)^T$$

and the errors in each component of \underline{y} were computed for $t = 0.0(0.1)1.0$ using the eleven consistent implicit methods of section 2. These errors are given in Table V for $t = 0.5$; the solution at $t = 0.5$ is

$$\underline{y} \approx (0.44, -0.44)^T.$$

The numerical results obtained for Problem 3 are in agreement with the theoretical results of sections 1 and 2.

(14)

Problem 4

$$\begin{aligned} y_1'' &= 99y_1 + 14y_2, \\ y_2'' &= 7y_1 + 2y_2. \end{aligned}$$

As before $N=2$; the matrix of coefficients has positive eigenvalues $\lambda_1 = 1$, $\lambda_2 = 100$ and associated eigenvectors are $\underline{c}_1 = (1-7)^T$ and $\underline{c}_2 = (14,1)^T$. The theoretical solution was taken to be

$$\begin{aligned} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} &= \left\{ \frac{(1-14e^{-1})e^t + (14e-1)e^{-t}}{99(e-e^{-1})} \right\} \begin{bmatrix} 1 \\ -7 \end{bmatrix} \\ &+ \left\{ \frac{(7+e^{-10})e^{-10t} - (7+e^{10})e^{-10t}}{99(e^{10}-e^{-10})} \right\} \begin{bmatrix} 14 \\ 1 \end{bmatrix} \end{aligned}$$

and the initial conditions were specified as

$$\begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}; \quad \begin{bmatrix} y_1'(0) \\ y_2'(0) \end{bmatrix} = \frac{(2-14e-14e^{-1})}{99(e-e^{-1})} \begin{bmatrix} 1 \\ -7 \end{bmatrix} + \frac{10(14+e^{10}+e^{-10})}{99(e^{10}-e^{-10})} \begin{bmatrix} 14 \\ 1 \end{bmatrix}$$

The theoretical solution for $t = 0.0$ (0.1) 1.0 and the computed solution using the explicit method based on the (0, 1) or (1, 0) Padé approximants are graphed in Fig: 7. In Fig: 8 the error moduli for each component of $\underline{y}(t)$ is graphed against t for $t = 0.0$ (0.1) 1.0 using the explicit method based on the (1,1), (0,2) or (2,0) Padé approximants.

Boundary conditions were specified as

$$\underline{y}(0) = (0,-1)^T; \quad \underline{y}(1) = (1,0)^T$$

and the errors in each component of \underline{y} were computed for $t = 0.0$ (0.1)1.0 using the eleven implicit consistent methods of section 2. These errors are given in Table VI for $t = 0.5$ and are seen to be in agreement with the theoretical results of section 2; the solution at $t = 0.5$ is $\underline{y} = (0.07,-0.47)^T$.

5. Conclusions

Families of explicit and implicit formulas based on Padé approximants to the exponential function have been developed for solving a system of second order linear ordinary differential equations with constant coefficients.

For each formula developed the order of the principal error term was found to be one higher than the order of the principal error term of the Padé approximant used.

The formulas were tested on four problems. It is clear by studying the numerical results that, particularly for problems in which the matrix of coefficients has one or more large positive eigenvalues, the low order explicit methods have rapid error growth. This is easily explained by considering the theoretical solution of the differential equation given by

$$(38) \quad \tilde{y}(t) = \sum_{i=1}^N \{a_i \exp(\sqrt{\lambda_i} t)\} + b_i \exp(-\sqrt{\lambda_i} t)\} \underline{c}_i$$

Where, for $i=1,2,\dots,N$, λ_i are the eigenvalues of the matrix of coefficients (assumed distinct), \underline{c}_i are the associated eigenvectors, and a_i and b_i are constants. For any large positive eigenvalue the terms with positive exponents in (38) grow rapidly as t increases and the errors quickly swamp the computed solutions of the explicit formulas. This did not happen when the consistent implicit methods were used to solve the four problems.

Table I : Padé approximants to e^θ

Method	(m, k)	Padé approximant	Principal error term	Convergence range
I	(0,1)	$\frac{1+\theta}{1}$	θ^2	-
II	(1,0)	$\frac{1}{1-\theta}$	θ^2	$-1 < \theta < 1$
III	(1,1)	$\frac{2+\theta}{2-\theta}$	θ^3	$-2 < \theta < 2$
IV	(0,2)	$\frac{2+2\theta+\theta^2}{2}$	θ^3	-
V	(1,2)	$\frac{6+4\theta+\theta^2}{6-2\theta}$	θ^4	$-3 < \theta < 3$
VI	(2,1)	$\frac{6+2\theta}{6-4\theta+\theta^2}$	θ^4	$-1.16 < \theta < 3.16$
VII	(2,0)	$\frac{2}{2-2\theta+\theta^2}$	θ^3	$-0.73 < \theta < 2.73$
VIII	(2,2)	$\frac{12+6\theta+\theta^2}{12-6\theta+\theta^2}$	θ^5	$-1.58 < \theta < 7.58$
IX	(0,3)	$\frac{6+6\theta+3\theta^2+\theta^3}{6}$	θ^4	-
X	(1,3)	$\frac{24+18\theta+6\theta^2+\theta^3}{24+6\theta^2}$	θ^5	$-4 < \theta < 4$
XI	(2,3)	$\frac{60+36\theta+9\theta^2+\theta^3}{60-24\theta+3\theta^2}$	θ^6	$-2 < \theta < 10$
XII	(3,2)	$\frac{60+24\theta+3\theta^2}{60+36\theta+9\theta^2+\theta^3}$	θ^6	$-3.64 < \theta < 1.23$
XIII	(3,1)	$\frac{24+6\theta}{24+18\theta+6\theta^2-\theta^3}$	θ^5	$-0.97 < \theta < 2.63$
XIV	(3,0)	$\frac{6}{6-6\theta+3\theta^2-\theta^3}$	θ^4	$-0.70 < \theta < 1.60$
XV	(3,3)	$\frac{120+60\theta+12\theta^2+\theta^3}{120-60\theta^2+12\theta^2-\theta^3}$	θ^7	$-1.49 < \theta < 4.64$

Table II : Summary of the Implicit Methods of Section 2

Error	Method	Padé Approximant	Coefficient of Principal part of error	Highest Power of A
ℓ^2	II	(1,0)	-1	A
ℓ^4	III	(1,1)	$-\frac{1}{6}$	A
	V	(1,2)	$-\frac{1}{36}$	A
	VI	(2,1)	$\frac{1}{36}$	A ²
	VII	(2,0)	$\frac{1}{3}$	A ²
	XIV	(3,0)	$-\frac{1}{12}$	A ³
ℓ^6	VIII	(2,2)	$\frac{1}{360}$	A ²
	X	(1,3)	$-\frac{7}{2880}$	A ²
	XI	(2,3)	$\frac{1}{3600}$	A ²
	XII	(3,2)	$-\frac{1}{3600}$	A ³
	XIII	(3,1)	$-\frac{17}{2800}$	A ³
ℓ^8	XV	(3,3)	$-\frac{1}{50400}$	A ³

Table III : Error Moduli for $t=0.0(0.1)1.0$ for Problem 1 using the Implicit Methods of Section 2.

Error	Padé	t										
		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
ℓ^2	(1,0)	0.0	0.84(-1)	0.15	0.20	0.23	0.24	0.24	0.21	0.17	0.97(-1)	0.0
ℓ^4	(1,1)	0.0	0.15(-3)	0.27(-3)	0.36(-3)	0.41(-3)	0.44(-3)	0.43(-3)	0.38(-3)	0.30(-3)	0.17(-3)	0.0
	(1,2)	0.0	0.25(-4)	0.45(-4)	0.59(-4)	0.69(-4)	0.73(-4)	0.71(-4)	0.63(-4)	0.49(-4)	0.29(-4)	0.0
	(2,1)	0.0	0.25(-4)	0.45(-4)	0.60(-4)	0.69(-4)	0.73(-4)	0.71(-4)	0.64(-4)	0.50(-4)	0.29(-4)	0.0
	(2,0)	0.0	0.53(-3)	0.94(-3)	0.12(-2)	0.14(-2)	0.15(-2)	0.15(-2)	0.13(-2)	0.10(-2)	0.60(-3)	0.0
	(3,0)	0.0	0.76(-4)	0.14(-3)	0.18(-3)	0.21(-3)	0.22(-3)	0.22(-3)	0.19(-3)	0.15(-3)	0.87(-4)	0.0
ℓ^6	(2,2)	0.0	0.25(-7)	0.45(-7)	0.59(-7)	0.69(-7)	0.73(-7)	0.71(-7)	0.63(-7)	0.50(-7)	0.29(-7)	0.0
	(1,3)	0.0	0.22(-7)	0.39(-7)	0.52(-7)	0.60(-7)	0.64(-7)	0.62(-7)	0.56(-7)	0.43(-7)	0.25(-7)	0.0
	(2,3)	0.0	0.24(-8)	0.42(-8)	0.56(-8)	0.65(-8)	0.69(-8)	0.68(-8)	0.61(-8)	0.48(-8)	0.28(-8)	0.0
	(3,2)	0.0	0.27(-8)	0.47(-8)	0.62(-8)	0.72(-8)	0.76(-8)	0.74(-8)	0.66(-8)	0.51(-8)	0.29(-8)	0.0
	(3,1)	0.0	0.54(-7)	0.96(-7)	0.13(-6)	0.15(-6)	0.15(-6)	0.15(-6)	0.14(-6)	0.11(-6)	0.61(-7)	0.0
ℓ^8	(3,3)	0.0	0.18(-9)	0.29(-9)	0.35(-9)	0.39(-9)	0.44(-9)	0.43(-9)	0.37(-9)	0.29(-9)	0.17(-9)	0.0

Table IV: Errors for $t=0.5$ for Problem. 2 using the implicit methods of section 2

error	Padé	error in $y_1(0.5)$	error in $y_2(0.5)$
ℓ^4	(1,1)	-0.19(-2)	0.19(-2)
	(1,2)	-0.33(-3)	0.33(-3)
	(2,1)	0.31(-3)	-0.31(-3)
	(2,0)	0.69(-2)	0.69(-2)
	(3,0)	-0.94(-3)	0.94(-3)
ℓ^6	(2,2)	-0.98(-6)	0.98(-6)
	(1,3)	0.86(-6)	-0.86(-6)
	(2,3)	-0.98(-7)	0.98(-7)
	(3,2)	0.96(-7)	0.96(-7)
	(3,1)	0.21(-5)	-0.21(-5)
ℓ^8	(3,3)	0.58(-10)	-0.73(-11)

Table V: Errors for $t= 0.5$ for Problem 3 using the implicit methods of section 2.

error	Pade	error in $y_1(0.5)$	error in $y_2(0.5)$
ℓ^4	(1,1)	-0.86(-4)	0.86(-4)
	(1,2)	-0.14(-4)	0.14(-4)
	(2,1)	0.14(-4)	-0.14(-4)
	(2,0)	0.30(-3)	-0.30(-3)
	(3,0)	-0.43(-4)	0.43(-4)
ℓ^6	(2,2)	0.14(-7)	-0.14(-7)
	(1,3)	-0.12(-10)	0.12(-10)
	(2,3)	0.15(-8)	-0.15(-8)
	(3,2)	-0.13(-8)	0.13(-8)
	(3,1)	-0.30(-7)	0.30(-7)
ℓ^8	(3,3)	-0.19(-9)	0.20(-9)

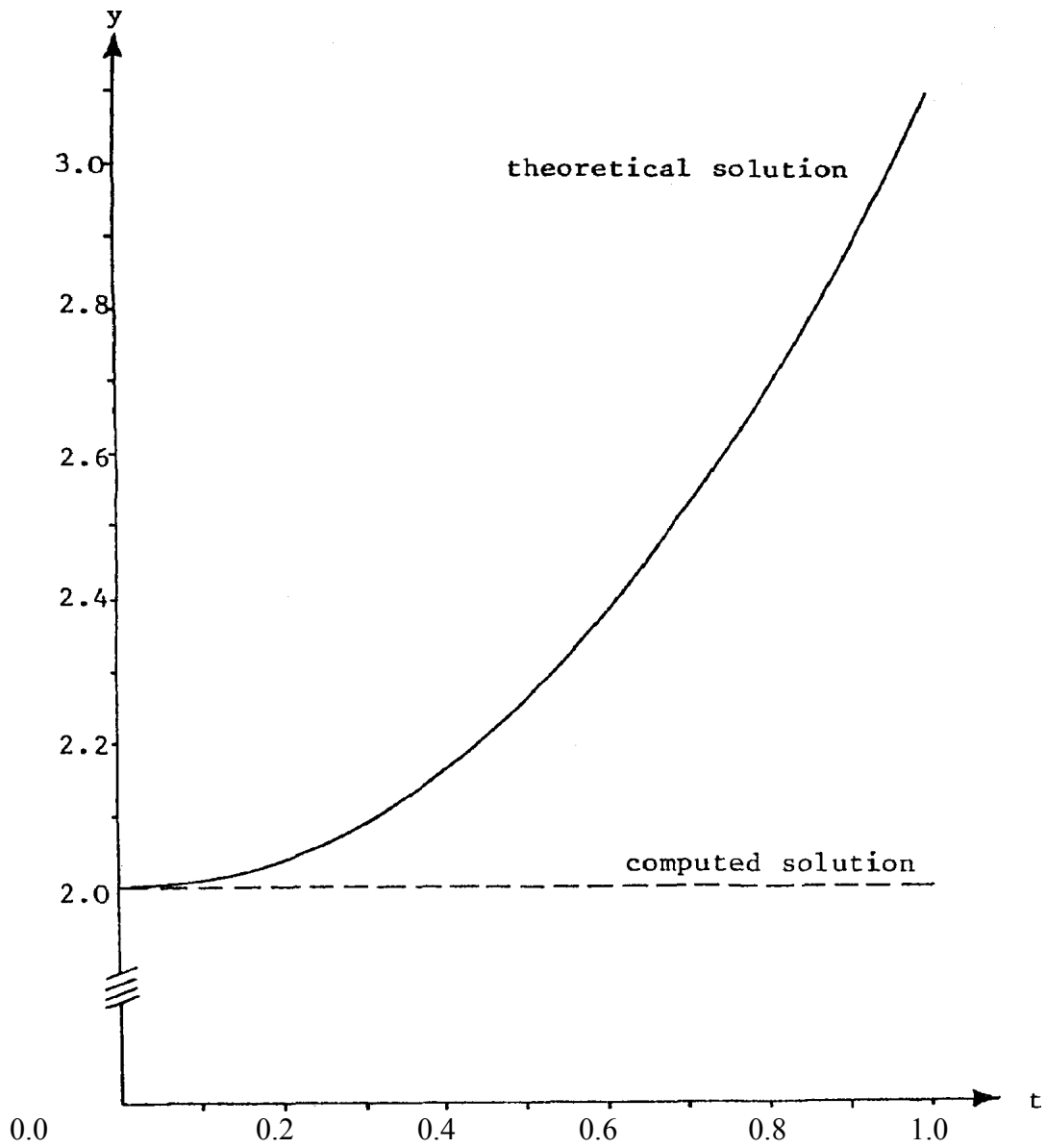
(21)

Table VI: Errors for $t= 0.5$ for Problem 4 using the implicit methods of section 2.

error	Padé	error in $y_1(0.5)$	error in $y_2(0.5)$
l^4	(1,1)	-0.22(-2)	-0.68(-4)
	(1,2)	-0.45(-3)	-0.17(-4)
	(2,1)	0.74(-3)	0.37(-4)
	(2,0)	0.32(-1)	0.20(-2)
	(3,0)	-0.25(-2)	-0.13(-3)
l^6	(2,2)	0.42(-4)	0.30(-5)
	(1,3)	-0.33(-4)	-0.23(-5)
	(2,3)	0.48(-5)	0.34(-6)
	(3,2)	-0.62(-5)	-0.44(-6)
	(3,1)	-0.10(-3)	-0.72(-5)
l^8	(3,3)	-0.29(-6)	-0.21(-7)

(22)

Figure 1: Theoretical and computed solutions to Problem 1 using the explicit method based on the (0,1) or (1,0) Padé approximants.



(23)

Figure 2: Error moduli for Problem 1 using the explicit method based on the (1,1), (0,2) or (2,0) Padé approximants

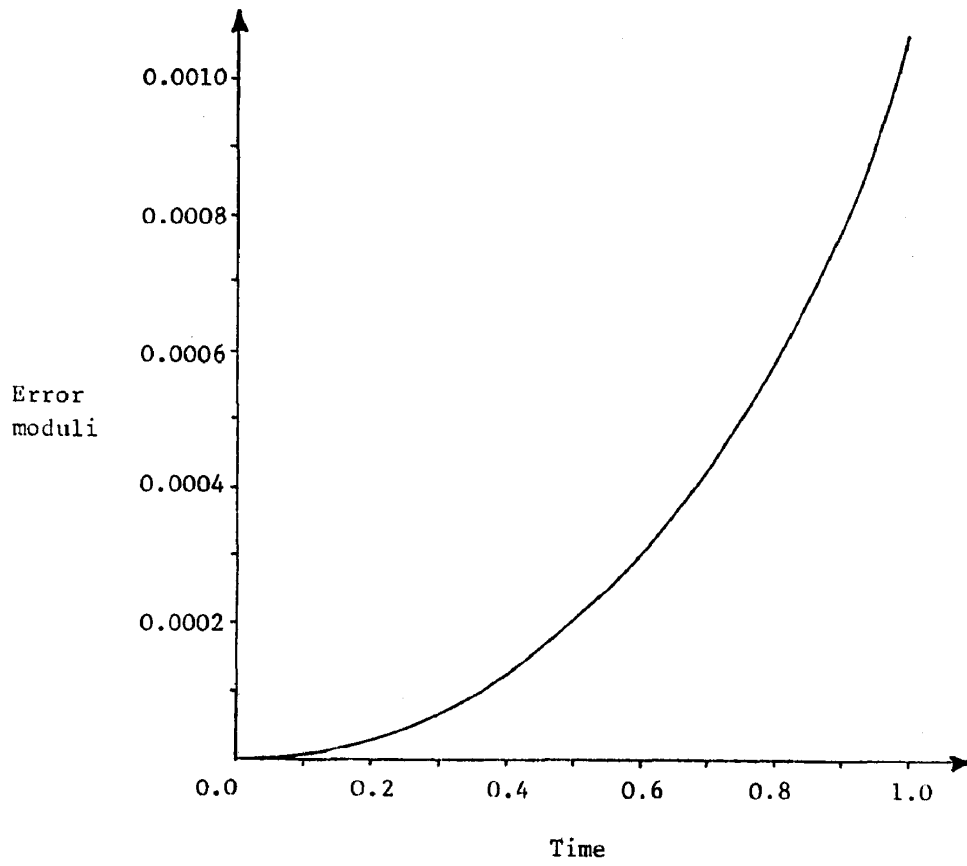
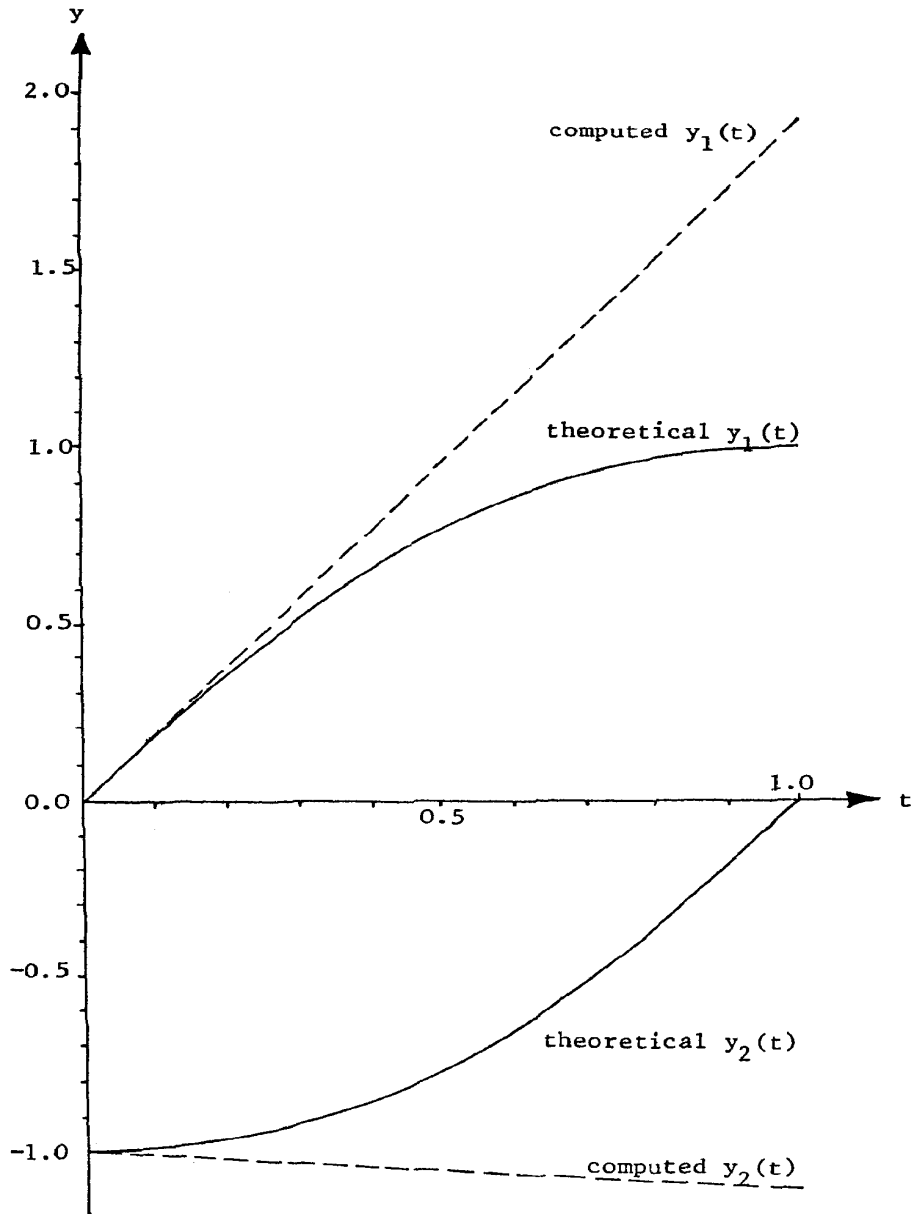


Figure 3: Theoretical and computed solutions to Problem 2 using the explicit method based on the (0,1) or (1,0) Padé approximants



(25)

Figure 4: Errors for Problem 2 using the explicit method based on the (1,1), (0,2) or (2,0) Padé approximants.

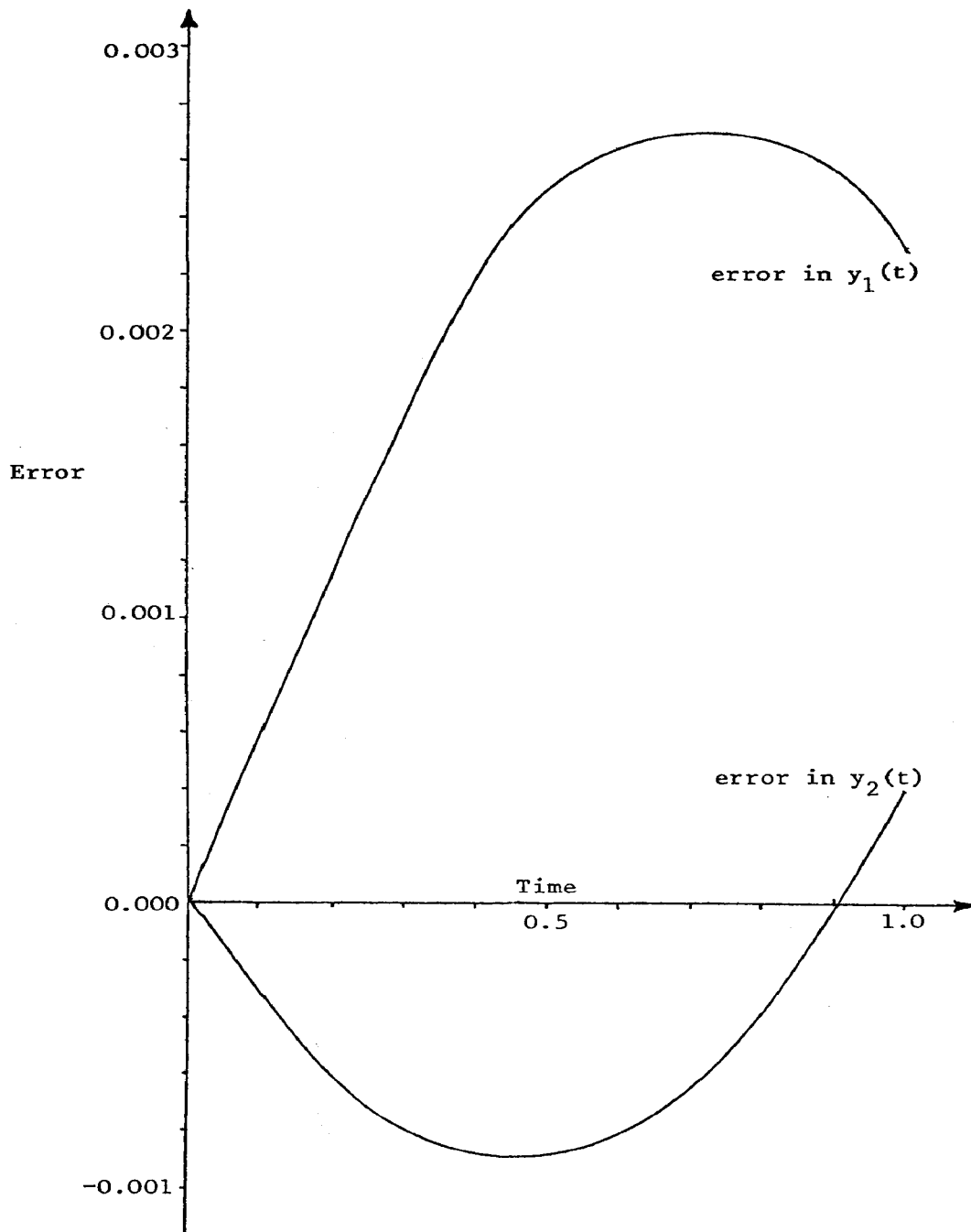
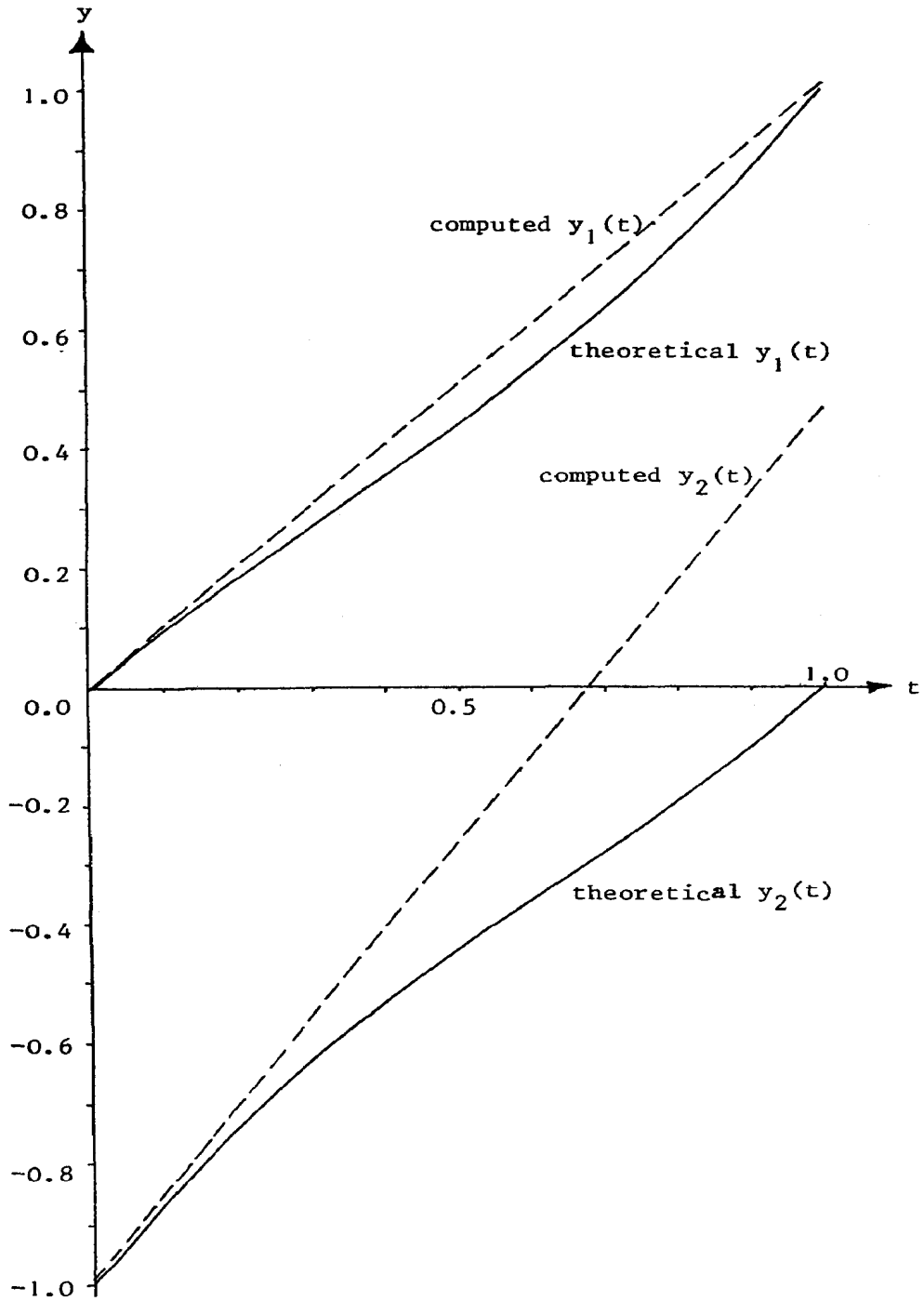


Figure 5: Theoretical and computed solutions to Problem 3 using the explicit method based on the (0,1) or (1,0) Padé approximant.



(27)

Figure 6: Error moduli for Problem 3 using the explicit method based on the (1, 1), (0, 2) or (2, 0) Pade approximants.

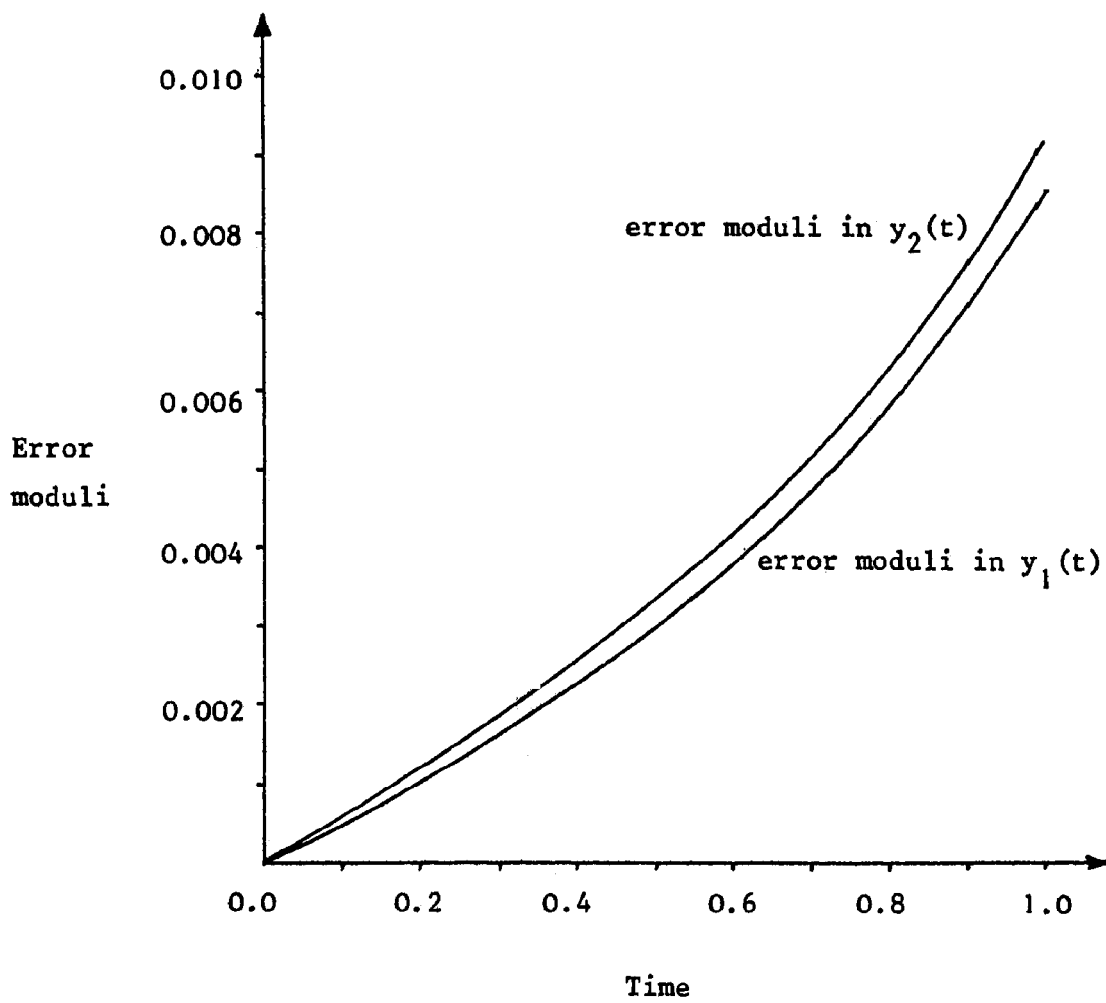
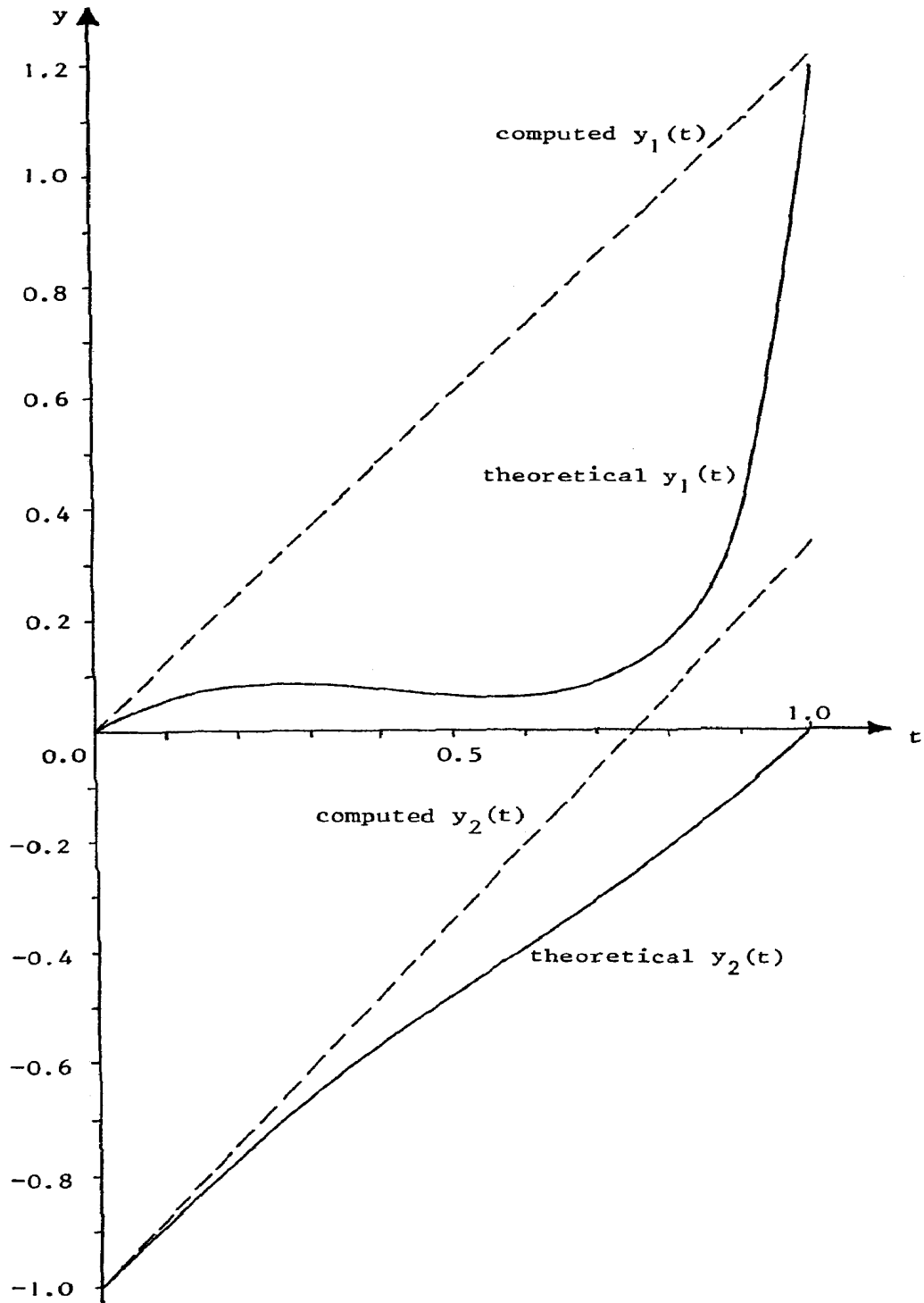
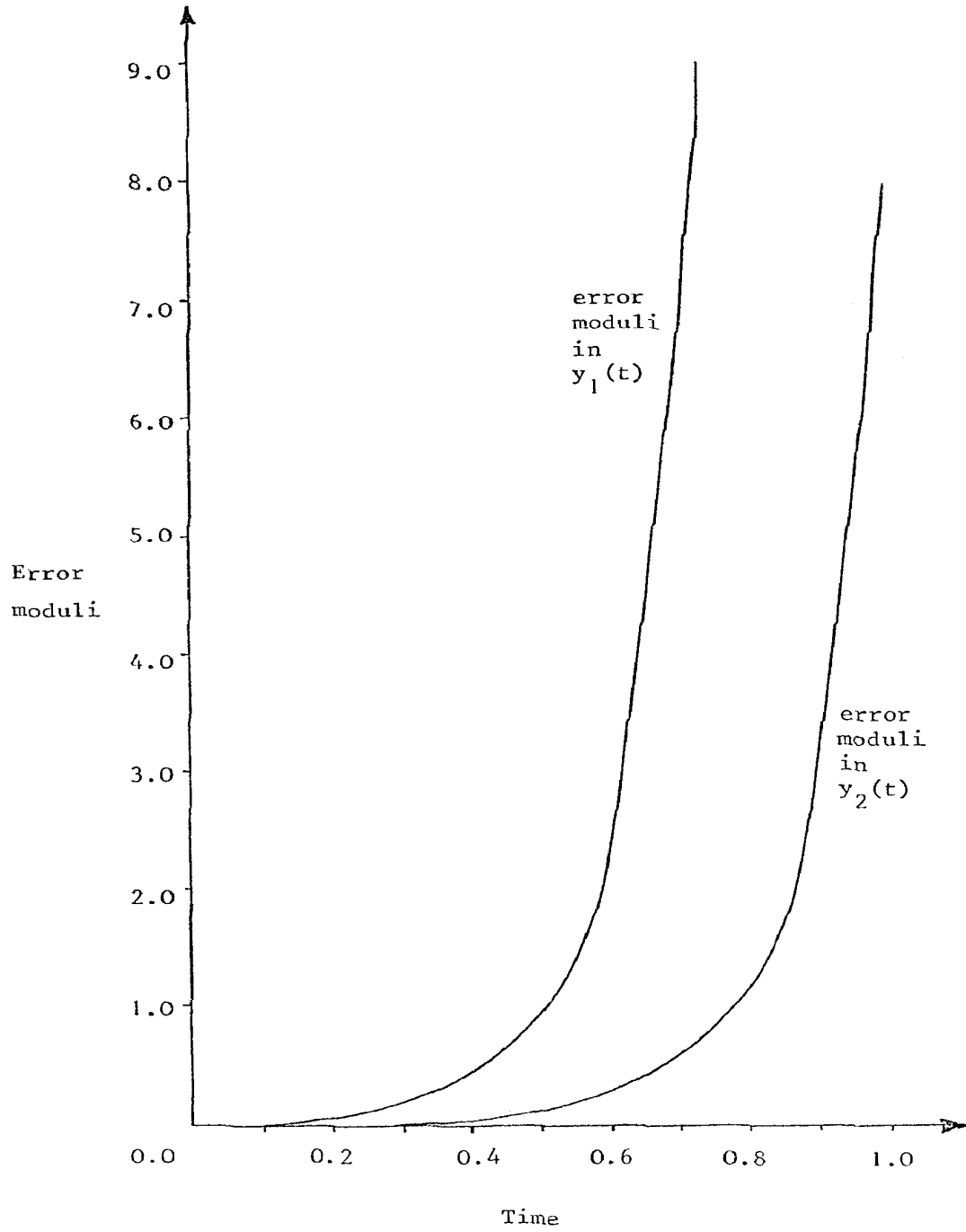


Figure 7: Theoretical and computed solutions to Problem 4 using the explicit method based on the (0,1) or (1,0) Padé approximants.



(29)

Figure 8: Error moduli for Problem A using the explicit method based on the (1, 1), (0, 2) or (2, 0) pade approximants.



References

1. R.J Goult et al, *Computational methods in linear algebra*, Stanley Thornes, 1974.
2. P. Henrici, *Discrete variable methods in ordinary differential equations*, Wiley, 1962.
3. J. D. Lambert, *Computational methods in ordinary differential equations*, Wiley, 1973.
4. E.H. Twizell, *An explicit method for the wave equation with extended stability range*, BIT, 19(1979), pp.373-378.

2 WEEK LOAN

BRUNEL UNIVERSITY LIBRARY
Uxbridge, Middlesex UB8 3PH
274000 Ext 257

**NOT TO BE
REMOVED
FROM THE LIBRARY**

XB 2356411 3

