TR/65 JUNE 1976

FINITE ELEMENT MULTISTEP MULTIDERIVATIVE SCHEMES FOR PARABOLIC EQUATIONS

BY

PHILIP MOORE

ABSTRACT

The linear, homogeneous, parabolic equation is solved by applying finite element discretizations in space and A_0 —stable, linear multistep, multiderivative (L.M.S.D.) methods in time. Such schemes are unconditionally stable. An error analysis establishes an optimal bound in the L_2 —norm. Methods typifying the class of L.M.S.D. schemes are derived and their implementation examined.

1. The Linear Parabolic Problem

We shall consider the initial boundary value problem

$$\frac{\partial u}{\partial t} = \sum_{i,j=1}^{N} \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u}{\partial x_j}\right) - a(x)u \equiv Lu, (x,t) \in \Omega \times (0,\infty) \qquad (1{\cdot}1a)$$

$$u(x,0) = g(x). \quad , \quad x \in \Omega \qquad (1{\cdot}1b)$$

$$u(x,t) = 0 \quad , \quad (x\ ,\ t) \in \Gamma \times (0,\infty) \qquad (1 \cdot 1c)$$

where $x = (x_1,\dots,x_N)$ is a point of a bounded domain Ω, with boundary Γ, lying in the N-'dimensional Euclidean space. Without loss of generality the boundary value is taken to be homogeneous Dirichlet. Non-homogeneous Dirichlet and Newmann boundary conditions apply with only minor adjustments.

For simplicity we allow

$$\{a_{ij}(x)\}_{i\,j=1}^{N}, a(x) \in C^{\infty}(\overline{\Omega}) \ , \Gamma \in C^{\infty}$$

where Ω is the closure of Ω. We also assume that

$$a(x) \geq 0 \qquad (1{\cdot}2i)$$

and the matrix $a_{ij}(x)$ is uniformly positive definite

$$\text{i.e.} \quad a_{ij}(x) = a_{ji}(x) \quad 1 \leq i, j \leq N, \quad x \in \overline{\Omega}$$

$$\text{and} \sum_{i,j=1}^{N} a_{ij}\ \xi_i\ \xi_j > \gamma \sum_{i=1}^{N} \xi_i^2 \quad \text{for some positive constant } \gamma \qquad (1{\cdot}2ii)$$

Before we can formulate the weak form of the problem (1.1) it is necessary to introduce Sobolev spaces. The Sobolev space $H^m(\Omega)$ is defined to be the space of real functions which, together with their first m generalised derivatives, are in $L_2(\Omega)$ the space of square integrable functions over Ω. The space $H^m(\Omega)$ is a Hilbert space, the inner product $(\cdot,\cdot)_m$ being given by

$$(u,v)_m = \sum_{|j| \leq m} \int_\Omega D^j u \, D^j v \, dx$$

where $j = (j_1, \ldots, j_N)$, $|j| = j_1 + \ldots + j_N$ and $D^j u = \dfrac{\partial^{|j|} u}{\partial x_1^{j_1} \ldots \partial x_N^{j_N}}$.

The associated norm, $\|\cdot\|$, is defined to be

$$\| v \|_m = (v,v)_m^{\frac{1}{2}}$$

The norm and inner product on $L_2(\Omega)$ are denoted respectively by $\|\cdot\|$ and $(\cdot,\cdot)$ where

$$\| v \| = \left(\int_\Omega v^2 dx \right)^{\frac{1}{2}} , \quad (u,v) = \int_\Omega uv \, dx$$

Further we denote by $H_0^1(\Omega)$ the space of all real functions v, where $v \in H^1(\Omega)$ and $v|_\Gamma = 0$ in the generalised sense. To formulate the weak problem associated with (1·1) we multiply the equation by an arbitrary function $v \in H_0^1(\Omega)$ and integrate over (Ω). Using Green's theorem we get

$$\int_\Omega \frac{\partial u}{\partial t} v \, dx + \sum_{i,j=1}^{N} \int_\Omega a_{ij}(x) \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_i} dx + \int_\Omega a(x) \, u \, v \, dx = 0 \qquad (1.3)$$

We adopt the notation

$$a(u,v) = \sum_{i,j=1}^{N} a_{ij}(x) \frac{\partial u}{\partial x_j} \frac{\partial v}{\partial x_i} dx + \int_\Omega a(x) \, uv \, dx$$

and consequently rewrite (1·3) as

$$\left(\frac{\partial u}{\partial t} , v \right) + a(u,v) = 0 \qquad \forall \, v \in H_0^1(\Omega) \, , \, t > 0 \qquad (1.4)$$

The weak solution of the problem (1·1) is the function $u(x,t) \in H_0^1(\Omega)$ which satisfies (1·4) for all $t > 0$ and the initial identity (1·1b).

We determine the asymptotic behaviour of u(x,t) by employing the 'energy method'. Denoting $\alpha(t) \equiv \| u(\cdot, t) \|$ we have by applying (1·2) to the expression (1·4) with $v = u(x,t)$

$$\alpha(t) \frac{d}{dt} \alpha(t) + \gamma \alpha(t)^2 \leq \left(\frac{\partial u}{\partial t}, u \right) + a(u,u) = 0$$

Cancelling throughout by $\alpha(t)$, multiplying by $e^{-\gamma t}$, and integrating from 0 to T we achieve

$$\| u(x,T) \| \leq e^{-\gamma T} \| g(x) \| \qquad (1 \cdot 5)$$

2. The Galerkin Procedure

Let V^o be a finite dimensional subspace of $H_0^1(\Omega)$ The Galerkin method is to find an approximation, U(x,t), to u(x,t) of the form

$$U(x,t) = \sum_{i=1}^{d} C_i(t) V_i(x) \qquad (2 \cdot 1)$$

where $\{V_i(x)\}_{i=1}^{d}$ is a basis of V^o. The continuous-time Galerkin solution to (1·1) is the function (2-1) where the coefficients $\{C_i(t)\}_{i=1}^{d}$ are determined by the discrete analogue to (1·4), namely

$$\left(\frac{\partial U}{\partial t}, V \right) + a(U,V) = 0 \text{ for any } V \in V^0, \; t > 0 \qquad (2 \cdot 2)$$

Substituting $\{V_i(x)\}_{i=1}^{d}$ in turn for V in (2·2), and assembling in Matrix form, we see that

$$M \frac{d}{dt} \underline{C} + K \underline{C} = \underline{0} \qquad (2 \cdot 3)$$

where M and K are constant, positive-definite matrices. The elements of M and K are

$$M_{ij} = (V_i, V_j) \quad \text{and} \quad K_{ij} = a(V_i, V_j), \; 1 \leq i, j \leq d.$$

An appropriate initial condition is derived from a discretized form of the identity(1·1b) . Let $\bar{g}(x) \in V^0$ be an approximation to g(x) and define $U(x,0) = \bar{g}(x)$. This yields an initial condition for $\underline{C}(0)$, say

$$\underline{C}(0) = \underline{a} \tag{2·4}$$

The equations (2·3) and (2-4) define the continuous time Galerkin solution.

Applying the energy method to (2·2) we have, by the previously described manipulations, that

$$\| U(x, T) \| \le e^{-\gamma T} \| g(x) \| \tag{2·5}$$

The expressions (1·5) and (2-5) will be influential in our choice of time discretization schemes to approximate U(x,t). Any method that preserves the asymptotic behaviour of the true solution is 'well-posed'. This concept of 'strong stability' or well-posedness is investigated by Crouzeix [2] and Nassif [10]. Following them we define a 'k-step' approximation method to be strongly stable if U^n , the approximant to $U(x,n\Delta_t)$, satisfies

$$\| U^n \| \le C e^{-\alpha n \Delta_t} \sum_{j=0}^{k-1} \| U^j \| \tag{2·6}$$

where α is some positive constant. Throughout this, and the following pages, we use C and c as generic constants.

We now impose a necessary property on the subspace V^0, namely $V^0 \equiv V_h^p$, where V_h^p has the property that for any $\tilde{v} \in H^{p+1}(\Omega) \cap H_0^1(\Omega)$ there exists an element $V \in V_h^p$ such that whenever h is sufficiently small

$$\| \tilde{v} - v \| + h \| \tilde{v} - v \|_1 \le ch^{s+1} \| \tilde{v} \|_{s+1} , \quad s = 1,2, \ldots, p \tag{2.7}$$

Any function $\phi \in V_h^p$ can be expressed as $f = \underline{x}^T \underline{v}$ where $\underline{x}$ is a vector of constants and $\underline{V} = (V_1, V_2, \ldots, V_d)$. We assume that the space V_h^p exhibits the following properties :

(Pi) $\quad ch^{-N} \| \phi \|^2 \leq \| \underline{x} \|_E^2 \leq ch^{-N} \| \phi \|^2$

(pii) $\quad a(\phi, \phi) \leq Ch^{-2} \| \phi \|^2$

for any $\phi \in V_h^p$, where $\| \underline{x} \|_E$ is the Euclidean norm on $\mathbb{R}^d$.
The above properties are satisfied by the finite element subspaces used in practise.

Let $\Lambda_{[k]}$, $\underline{x}_{[k]}$ and $\Lambda_{[m]}$, $\underline{x}_{[m]}$ be respectively an eigenvalue, eigenvector of the matrices M and K. We derive bounds on these eigenvalues by utilising (Pi), (Pii) and (1·2). Now,

$$\Lambda_{[m]} \| \underline{x}_{[m]} \|_E^2 = \Lambda_{[m]} \underline{x}_{[m]}^T \underline{x}_{[m]} \underline{x}_{[m]}^T M \underline{x}_{[m]} = \| \phi_{[m]} \|^2$$

$$\text{i.e.} \quad \frac{1}{C} h^N \leq \Lambda_{[m]} \leq \frac{1}{C} h^N .$$

Similarly,

$$\Lambda_{[k]} \| \underline{x}_{[k]} \|_E^2 = \underline{x}_{[k]}^T k \underline{x}_{[k]} = a(\phi_{[k]}, \phi_{[k]})$$

$$\text{i.e.} \quad \frac{\gamma}{C} h^N \leq \Lambda_{[k]} \leq \frac{C}{C} h^{N-2} .$$

It is important to see that the eigenvalues of $S = M^{-1} K$ are positive and unbounded with respects to h. The largest eigenvalue of S, Λ_{max}, is of magnitude $\Lambda_{max}, \sim Ch^{-2}$ whereas the smallest eigenvalue is bounded from above. Consequently the system of differential equations (2·3) is a stiff system.

3. A_0 - stable, linear multistep, multiderivative methods

Most classical methods for solving initial value problems of first order ordinary differential equations require, for reasons of stability, a condition of the form $|\Lambda_{max} \Delta_t| < C$, where Δ_t is the time increment and C a constant usually between one and ten. For the stiff system (2·3) this condition requires $h^{-2} \Delta_t$ to be small which imposes a severe limitation on the step length Δ_t. As we will be required to solve a linear algebraic equation at each time interval this restriction is prohibitive. We are thus lead to consider only methods where the region of absolute stability is unbounded. Since the eigenvalues A of the matrix S are real the classes of A_0 -stable methods are sufficient. Zlamal [15] employed the class of A_0 - stable, linear multistep methods to solve the system (2·3). Other authors, including Nassif [10], Makinson [8], have studied various one-step methods for the solution of stiff systems. Following Obrechkoff (see [7,pp199]), Enright [4], Genin [5] amongst others, we shall consider multistep formulae that incorporate the higher derivatives. We refer to such schemes as A -stable, linear multistep, multiderivative methods (L.M.S.D's). This follows the terminology of Genin but we note that the title 'Obrechkoff methods' is also used, e.g. [7].

A L.M.S.D, method is of the type

$$\sum_{j=0}^{k} \alpha_j y_{n+j} = \sum_{j=0}^{k} \sum_{r=1}^{m} \beta_{rj} \Delta_t^r y_{n+j}^r \qquad (3\text{-}1)$$

where $\alpha_k > 0$ and $y_n^r \equiv \left.\frac{d^r}{dt^r} y \right|_{t = n\Delta_t}$

Analogous to linear multistep methods (cf.[6, pp 221]) the method (3·1) is said to be of order q if, for Δ_t sufficiently small

$$L[y(t), \Delta_t] = \sum_{j=0}^{k} \left\{ \alpha_j y(t + j\Delta_t) - \sum_{r=1}^{m} \beta_{rj} \Delta_t^r y^r (t + j\Delta_t) \right\}$$

$$= C_{q+1} \Delta_t^{q+1} y^{q+1}(t) + 0(\Delta_t^{q+2}) \tag{3.2}$$

for any sufficiently differentiable function y(t). Expanding $L[y(t), \Delta_t]$ by Taylor's theorem with integral form of the remainder we have (cf.[6, pp 247])

$$L[y(t), \Delta_t] = \leq \Delta_t^{q+1} \int_0^k G(s) y^{q+1}(t + s\Delta_t)\, ds$$

$$\leq G \Delta_t^{q+1} \sup_{t \leq s \leq t + k\Delta_t} \left(\left| y^{q+1}(s) \right| \right) \tag{3.3}$$

where G(s) is the kernel function and $G = \int_0^k G(s) ds$.

The concept of A_0 -stability was introduced by Cryer [3]. A multistep method is A_0 -stable if, applied to the equation $y^n = \lambda y$, $y(0) = 1$, for any real $\lambda > 0$, it gives approximate values y of $y(n\Delta_t)$ such that $y^n \rightarrow 0$ as $n \rightarrow \infty$. Considering (3·1), this is equivalent to the roots of $P(\xi, \tau)$ being of modulus less than one for $\tau > 0$, where

$$p(\xi, \tau) = \rho(\xi) + \sum_{r=1}^{m} \tau^r \sigma_r(\xi), \rho(\xi) = \sum_{j=0}^{k} \alpha_j \xi^j \text{ and } \sigma_r(\xi) = \sum_{j=0}^{k} \beta_{rj} (-1)^{r-1} \xi^j \tag{3-4}$$

$$r = 1, 2, \ldots, m.$$

In addition we require that the L.M.S.D. methods satisfy the conditions of zero-stability and consistency, ([7 pp.30]). Zero-stability dictates that the roots of $p(\xi)$ with modulus equal to one are simple. The consistency condition is maintained by

$$\sum_{j=0}^{k} \alpha_j = 0 \quad \text{and} \quad \sum_{j=0}^{k} j\alpha_j = \sum_{j=0}^{k} \beta_{1j} .$$

We shall always assume that the characteristic polynomials $\rho(\xi)$ and $[\sigma_r(\xi\}_{r=1}^{m}$ have no common factor. Similarly, the polynomials, $\{\mu_j(\tau)\}_{j=0}^{k}$, where

$$\mu_j(\tau) = \alpha_j + \sum_{r=1}^{m} (-1)^{r-1} \beta_{rj} \tau^r$$

shall have no common factor. These assumptions are compatible with the L.M.S.D.scheme being irreducible to an equivalent scheme with a lower value for k or m.

The following two results, although required in the later analysis, are of interest in themselves.

<u>Lemma 1</u> Let the L.M.S.D. scheme (3·1) be A_0 - stable, then there exists a positive constant μ, such that

$$\mu_k \ (\tau) > u \ , \text{ for all } \tau \geq 0$$

Proof: Since a > 0 by definition the expression $\mu_k(\tau)$ is not identically zero. Let us assume that $\mu_k(\tau)$ has a root at $\tau = \bar{\tau}$. The function

$$f(\xi,,\tau) \equiv \frac{P(\xi,,\tau)}{\mu_k(\tau)} = \sum_{j=0}^{k} \frac{\mu_j(\tau)}{\mu_k(\tau)} \xi^j$$

is well defined except at the zeros of $\mu_k(\tau)$. As $\tau \to \bar{\tau}$ at least one of the coefficients of $f(\xi,\tau)$ must become unbounded since $\tau = \bar{\tau}$ may not be a root of all $\{\mu_j(\tau)\}_{j=0}^{k-1}$. Consequently, as $\tau \to \bar{\tau}$, at least one of the roots of $f(\xi,\tau)$, and hence of $p(\xi,\tau)$, must become unbounded and have modulus greater than one. This contradicts the assumption of A_0 - stability and we deduce that $\mu_k(\tau)$ must be bounded away from zero, $\tau > 0$. Since $\mu_k(0) = \alpha_k > 0$ the proof is complete.

Lemma 2. Let the L.M.S.D. scheme (3·1) be A_0-stable, then

$$\beta_{mk} \neq 0$$

Proof: Trivially, if $\{\beta_{mj}\}_{j=0}^{k-1}$ are all zero then $\beta_{mk} \neq 0$ otherwise the scheme will incorporate only the first (m - 1) derivatives. Let us assume that at least one $\beta_{ms} \neq 0$, $0 \leq s \leq k-1$, and further that $\beta_{mk}=0$. Using the function $f(\xi, \tau)$ of lemma 1 it is obvious that the coefficient of ξ^s must become unbounded as $\tau \to \infty$. Once again (cf. lemma 1) this comprises a contradiction in the initial assumption of A_0 -stability and we deduce that $\beta_{mk} \neq 0$.

Corollary. Every A_0 -stable L.M.S.D. scheme (3·1) must be implicit.

Finally, we investigate the approximate solution of (2·2) by the L.M.S.D. method (3·1). Let us again denote U^n to be an approximant to $U(x, n\Delta_t)$. Assuming that $\{U^j\}_{j=0}^{k-1}$ are given, the recurrence relationship for U^{n+k}, $n \geq 0$, is given by the system of difference equations

$$\left(\sum_{j=0}^{k} \alpha_j U^{n+j}, V\right) - \left(\sum_{j=0}^{k}\sum_{r=1}^{m} \beta_{rj} \Delta_t^r U_{(r)}^{n+j}, V\right) = 0 \qquad (3·5)$$

$$\left(U_{(r)}^{n+j}, V\right) + a\left(U_{(r-1)}^{n+j}, V\right) = 0 \qquad r = 1,2,\ldots,m \qquad (3·6)$$

The computational aspects of (3·5) and (3·6) will be investigated in chapter 6. The implementation procedures described there are equivalent to the solution of the linear system of equations

$$A\underline{U}^{n+k} \equiv \left[\alpha_k I + \sum_{r=1}^{m} (-1)^{r-1} \beta_{rk} \Delta_t^r (M^{-1}K)^r\right] \underline{U}^{n+k} = \underline{\tilde{U}}$$

for some predetermined vector $\underset{\sim}{\tilde{U}}$. The condition number of the matrix A where

$$\text{Cond}\ (A) = \frac{\max\ (\Lambda[a])}{\min\ (\Lambda[a])} \quad ,$$

and ($\Lambda_{[a]}$) is the set of eigenvalues of A is readily seen by (Pi) and (Pii) of chapter 2 to satisfy

$$\text{Cond}(A) = 0(h^{-2m}\ \Delta^m{}_t\ ) .$$

Hence, by lemma 1, the matrix A is positive definite and, if we exclude the unrealistic case when $\Delta_t^{h-2} \longrightarrow 0$, the condition number of A does not grow too fast for small m.

4. Theorems

The analysis of chapter 5 will prove the following theorems.

Theorem 1

Let the L.M.S.D.method $(3 \cdot 1)$ of order q be consistent, zero-stable and A_0 - stable. Let the roots of the polynomial $p(\xi)$ with modulus equal to one be real, the modulus of the roots of the polynomial $\sigma_m(\xi)$ be less than one, and $\sigma_1(-1) \neq 0$. Further, let $g(x) \in L_2(\Omega)$. Then for any $t_o > 0$ there exists a positive constant $C(t_0)$ such that for $n\ \Delta_t \geq t_0$, and h, Δ_t sufficiently small

$$\| u(x, n\Delta_t) - U^n \| \leq C(t_0) \left\{ \Delta_t^q + h^{p+1}) \| g \| + \sum_{j=0}^{k-1} \| u(x, j\Delta_t) - U^j \| \right\}$$

and

$$\| U^n \| \leq C\, e^{-\alpha n \Delta t \lambda 1} \sum_{j=0}^{k-1} \| U^j \|$$

Corollary

If in addition we assume that U^0 is the projection of g(x) onto V_h^p by the L_2 - inner product and $\{U^j\}_{j=1}^{k-1}$ are the values derived from a weakly A_0 - stable Padé scheme of order q^{-1}, then

$$\| u(x,\ n\Delta_t) - U^n \| \leq C(t_0) \left\{ \Delta_t^q + h^{p+1} \right\} \| g \|$$

and

$$\| U^n \| \leq C\ e^{-\alpha n \Delta t \lambda 1} \quad \| g \|$$

Theorem 2

Let us further restrict w=1 to be the only root of $\rho(\xi)$ with modulus equal to one, then with the assumptions of theorem 1

$$\| u(x,n\Delta_t) - U^n \| < C(t_0,\beta)\, e^{-\beta n\Delta_t \lambda_1} \left\{ (\Delta_t^q + h^{p+1}) \| g \| + \sum_{j=0}^{k-1} \| u(x,j\Delta_t) - U^j \| \right\}$$

for some for arbitrary positive constant β, $0 < \beta < 1$.

Corollary

If the initial values are defined to be exactly those described in the corollary to theorem 1, then

$$\| u(x, n\Delta_t) - U^n \| < C(t_0,\beta)\, e^{-\beta n\Delta_t \lambda_1} \left\{ (\Delta_t^q + h^{p+1} \right\} \| g \|$$

5. Proof of Theorems

Let $\{\lambda_i\}_{i=1}^{\infty}$ and $\{\psi_i\}_{i=1}^{\infty}$ be respectively the eigenvalues (in increasing order) and the corresponding orthononnal eigenfunctions of the continuous eigenvalue problem

$$a(\psi,v) = \lambda\ (\psi,v) \qquad \forall\ v \in H_0^1(\Omega) \tag{5·1}$$

The eigenvalues are well-known to be positive and distinct.

Further let $\{\Lambda_i\}_{i=1}^{d}$ and $\{\psi_i\}_{i=1}^{d}$ be the eigenvalues (in increasing order) and the corresponding orthonormal eigenfunctions of the discrete eigenvalue problem

$$a(\psi,V) = \Lambda\ (\psi,V) \qquad \forall\ v \in v_h^p \tag{5·2}$$

Strang and Fix [12, Theorem 6·1, 6·2] have proved results for eigenvalues and eigenfunctions using subspaces, S_h , on a regular mesh. The only property of S_h utilised in the proof is the approximation property

$$\| u - Pu \|_s \leq Ch^{k-S} \| u \|_k \qquad s = 0, \text{ or } 1$$

where Pu is the Ritz approximation of u (i.e.$a(u - Pu,V) = 0,\ \forall\ V \in S_h$)

A well-known consequence of (2 .7) is that

$$\| u - Pu \| + h \| u - P u \|_1 \le Ch^{P+1} \| u \|_{p+1} .$$

Hence, for k = p+l, all conditions are satisfied and the theorems yield for h sufficiently small

$$0 \le \Lambda_i - \lambda_i \le Ch^{2p} \lambda_i^{p+1} , , \quad i = 1, 2, \dots, d \tag{5.3}$$

$$\|\psi_i - \psi_i\| \le Ch^{p+1} \lambda_i^{\frac{1}{2}(p+1)} , \quad i = 1, 2, \dots, d \tag{5.4}$$

We adopt the following notations

$$v_i = (v, \psi_i) , \quad \bar{v}_i = (v, \psi_i) , \quad v \in H_0^1(\Omega) \tag{5.5}$$

$$V_i = (v, \psi_i) , \quad \bar{V}_i = (V, \psi_i) , \quad V \in V_h^P$$

We bound the error $u(x, n\Delta_t) - U^n$ by using the relationship

$$u(x, n\Delta_t) - U^n = e_1 + e_2$$

where $e_1 = u(x, n\Delta_t) - U(x, n\Delta_t)$ and $e_2 = U(x, n\Delta_t) - U^n$ and proving bounds on e_1 and e_2 .

The solution u(x,t) of (1·1) can be expressed as

$$u(x, t) = \sum_{i=1}^{\infty} g_i e^{-\lambda_i t} \Psi_i \tag{5.6}$$

where $\{g_i\}_{i=1}^{\infty}$ are the Fourier coefficients of g(x),

Similarly, the solution U(x,t) of the continuous Galerkin problem (2·2) can be expressed by

$$U(x, t) = \sum_{i=1}^{d} U_i^o e^{-\Lambda_i t} \Psi_i \tag{5.7}$$

where $\{U_i^0\}_{i=1}^d$ are the coefficients of $\bar{g}(x) \in V_h^P$ with respect to the basis $\{ \Psi_i \}_{i=1}^d$.

a/

Let $U^n = \sum_{i=1}^{d} U_i^n \psi_i$ Using (5.7) we can write $e_2 = \sum_{i=1}^{d} \in_i^n \psi_i$ and hence $\| e_2 \|^2 = \sum_{i=1}^{d} |\in_i^n|^2$, where

$$\in_i^n = U_i^o \ e^{-\Lambda_i n \Delta t} - U_i^n \tag{5.8}$$

Also let $U_{(r)}^n \equiv \sum_{i=1}^{d} U_{i,r}^n \Psi_i$ be the discreate approximation to $\frac{\partial^r}{\partial t^r} U(x, \ t) |_{t=n\Delta t}$. Substituting $U_{(r)}^n$ into (3.6) and (5.2) with $V = \psi_i$ gives us the relationship

$$U_{i,r}^n + \Lambda_i U_{i,r-1}^n = 0, \quad r = 1, \ldots\ldots\ldots, m \qquad \text{where} \quad U_{i,o}^n \equiv U_i^n$$

Consequently, we can construct the recurrence relationship

$$U_{i,r}^n = (-1)^r \ \Lambda_i^r \ U_i^n \qquad r = 1, \ldots, m \tag{5.9}$$

Combining (5.9) and (3.5) with V= ψ_1 yields

$$\sum_{j=0}^{k} (\alpha_j + \sum_{r=1}^{m} (-1)^{r-1} \beta_{rj} \Delta_t^r \ \Lambda_i^r) U_i^{n+j} = 0 \tag{5.10}$$

Derfine $\delta_j(\tau) = \mu_j(\tau) / \mu_k(\tau)$ where $\mu_j(\tau) = \alpha_j + \sum_{r=1}^{m} (-1)^{r-1} \beta_{rj} \tau^r$ and subsequently rewrite (5.10) as

$$\sum_{j=0}^{k} \delta_j(\Delta_t \ \Lambda_i) U_i^{n+j} = 0 \tag{5.11}$$

The expressions (5.8) and (5.11) combine to give

$$\sum_{j=0}^{k} \delta_j(\Delta_t \ \Lambda_i) \in_i^{n+j} = \sum_{j=0}^{k} \delta_j(\Delta_t \ \Lambda_i) U_i^o \ e^{-\Lambda i(n+j)\Delta)} \equiv d_i^n \tag{5.12}$$

We conclude this sub-section by bounding d_i^n we see from (3·2) and (3·3) that

$$\sum_{j=0}^{k} \mu_j(\Delta_t \Lambda_i) U_i^o e^{-\Lambda i(n+j)\Delta_t} \equiv L\left[U_i^o e^{-\Lambda i t}, \Delta_t\right]_{t=n\Delta_t}$$

$$\leq G \Delta_t^{q+1} \Lambda_i^{q+1} |U_i^o| e^{-n\Delta_t \Lambda_i}$$

By lemma 1 a positive supremum of $\mu_k(\tau)^{-1}$, $\tau > 0$,must exist from which we conclude that

$$d_i^n \leq C \Delta_t^{q+1} \Lambda_i^{q+1} |U_i^o| e^{-n\Delta t \Lambda_i} \tag{5.13}$$

Alternatively, by lemmas 1 and 2, $\delta_j(\tau)$ j=0,l,...,k, are bounded for any $\tau > 0$, thus

$$d_i^n \leq C |U_i^o| e^{-n\Delta t \Lambda_i} \tag{5.14}$$

b) This section uses a method employed by Henrici [7,pp242] and adapted by Zlamal [14]. Define $\hat{p}(\xi, \tau)$ by

$$\hat{p}(\xi, \tau) = \delta_k(\tau) + d_{k-1}-(\tau)\xi + \ldots + \delta_0(\tau)\xi^k$$

Note that $\hat{P}(\xi, \tau) = \mu_k(\tau)^{-1} \xi^k p(\frac{1}{\xi}, \tau)$ and hence the roots of $p(\xi,T)$ are the reciprocals of the roots of $p(\xi, \tau)$. It is intuitively obvious that the roots of $p(\xi, \tau)$ approach the roots of $\rho(\xi)$ and $\sigma_m(\xi)$ as, respectively, $\tau \to 0$ and $\tau \to \infty$.

The essential roots of $p(\xi)$ (i.e.those of modulus one) are by assumption real, and by zero—stability single. The consistency condition dictates that $w = 1$ is always an essential root. Let us assume the most general situation when these essential roots are $w_1 = 1$, $w_2 = -1$. Any other root $\{w_i\}_{i=3}^{k}$ of $\rho(\xi)$ has modulus less than one, say $|w_i| \leq 1 - \theta, 0 < \theta \leq 1$. We employ a theorem from complex analysis eg. [1 , Theorem 11,pp 131] , to show that for each sufficiently shall $\in > 0$, there exists a $\tau_\in > 0$, such that the equation $p(\xi, \tau) = 0, \tau < \tau_\in$, has the same number of roots in the disc $|\xi - \xi_0| < \in$ as the equation $p(\xi) = 0$. Furthermore, if ξ_0 is a root of $\rho(\xi)$ of multiplicity p then the p roots of $p(\xi, \tau)$ that approach it are distinct for τ sufficiently small. Hence no complications arise from a root of multuplicity greater than one.

We denote $w_{1,2}$ to be correspondingly w_1 or w_2. Selecting $\in < \frac{\theta}{2}$ we have that, for $\tau < \tau_{\theta/2}$, the equation $p(\xi, \tau)$ has only one root in the disc $|\xi - w_{1,2}| < \frac{\theta}{2}$ Let this root be $\xi w_{1,2}(\tau)$.

Rearranging the above we deduce that for any $0 < \in < \theta/2$ there exists a $\tau_\in$ such that $|\xi w_{1,2}(\tau) - W_{1,2}| < \in$ whenever $\tau < \tau_\in$. This is a definition for $\xi_{w_{1,2}}(\tau)$ to tend continuously to $w_{1,2}$ as τ tends to zero. Thus $\xi_{w_{1,2}}(\tau)$ can be expressed as an analytic function of τ,

$$\text{i.e.} \quad \xi_{w_i}(\tau) = w_i + a_1^i \tau + a_2^i \tau^2 + \ldots\ldots \qquad i = 1, 2$$

Corresponding expressions hold for the other roots $(\xi_{w_i}(\tau))_{i=3}^k$ of $p(\xi, \tau)$.

Remembering that $|w_i| < 1 - \theta$, $i = 3, 4, \ldots, k$, we deduce that for τ sufficiently small, say $\tau < \tau_1, |\xi_{W_i}(\tau), | < 1 - \frac{\theta}{2}, i = 3, \ldots, k$.

Expanding $p(\xi_{w_{1,2}}(\tau), \tau)$ about the point $w_{1,2}$ we see that

$$p(\xi_{w_i}(\tau), \tau) = \rho(w_i) + \tau a_1^i \rho'(w_i) + \tau\sigma_1(w_i) + 0(\tau^2) = 0 \quad i = 1,2$$

and by comparing coefficients that

$$a_1^i = \frac{\sigma_1(w_i)}{\rho'(w_i)} \qquad i = 1,2 \qquad (5.15)$$

We know that $\sigma_1(1) = \rho'(1)$ by the consistency condition, $\sigma_1(-1) \neq 0$ by assumption and $\rho'(w_{1,2}) \neq 0$ by zero-stability.

Thus,

$$\xi_{w_i}(\tau) \mid w_i + a_1^i \tau + 0(\tau^2) \text{ where } a_1^i \text{ is real and non} - \text{zero}$$

and

$$|\xi_{w_i}(\tau)| = |1 + \frac{a_1^i \tau}{w_i} + 0(\tau^2)| \quad i = 1,2.$$

But as $|\xi_{w_i}(\tau)| < 1$ for $\tau < 0$ we must have $a_1^i / w_i < 0$. Consequntly, for τ sufficiently small, say $\tau < \tau_2$

$$|\xi_{w_i}(\tau) < 1 - \hat{\alpha}\tau, i = 1,2, \text{ for same } \hat{\alpha}, \hat{\alpha} \geq \frac{1}{2} \min \left\{ |a_1^1| = 1, |a_2^1| \right\}$$

Thus, we have shown that for $\tau < \hat{\tau}$, $\hat{\tau} = \min(\tau_1, \tau_2)$

$$|\xi_{w_i}(\tau)| < 1 - \alpha\tau, \quad \alpha > 0, \qquad i = 1, 2, \ldots, k$$

and hence, for $\tau < \hat{\tau}$ all roots $\hat{\xi}(\tau)$ of $\hat{P}(\xi, \tau)$ satisfy

$$|\hat{\xi}(\tau)| > \frac{1}{1 - \alpha\tau}.$$

Therefore, $\frac{1}{\hat{p}(\xi,\tau)}$ is holmorphic for $|\xi| \leq \frac{1}{1-\alpha\tau}$, $\tau < \hat{\tau}$, and the function can be expressed by a Taylor series expansion

i.e. $$\frac{1}{\hat{p}(\xi,\tau)} = \gamma_0(\tau) + \gamma_1(\tau)\xi + \gamma_2(\tau)\xi^2 + \ldots \qquad \tau < \hat{\tau}$$

where, by Cauchy's estimate, eg [1. pp 122]

$$|\gamma_\ell(\tau)| \leq C(1 - \alpha\tau)^\ell \qquad \ell = 0,1,\ldots \quad \text{whenever} \quad \tau < \hat{\tau}.$$

Similarly, let the roots of $\sigma_m(\xi)$ be $\{z_i\}_{i=1}^k$ These roots are by assumption less than one in modulus, say $|z_i| \leq 1 \;\theta\; 0 < \theta \leq 1$. Applying the aforementioned theorem we prove that the equatio $P(\xi, T) = 0$ has the same number of roots in the dis c $|\xi - z_i| < \frac{\theta}{2}$ as the equation $\sigma_m(\xi) = 0$, ehenever $\tau > C$. Repeating the above argument we have that, for $\tau > C$, the roots $\xi_i(\tau)$ of $p(\xi, \tau)$ satisfy

$$|\xi_i(\tau)| < 1 - \frac{\theta}{2}, \quad i = 1, 2, \ldots, k.$$

Tills leaves a finite interval $|\hat{\tau}, C|$ where the roots $\xi_i(\tau)$ of $p(\xi, \tau)$ are known to be of modulus less than one. The roots are continuous functions of τ over a finite intervals, hence

$$|\xi_i(\tau)| < 1 - \hat{\theta}, \quad 0 < \hat{\theta} < 1, \quad \text{whenever} \quad \hat{\tau} \le \tau \le C.$$

and we conclude that there exists a constant $\alpha, 0 < \alpha < 1$, such that

$$|\xi_i(\tau)| < 1 - \bar{\alpha} \qquad \text{whenever} \quad \tau \ge \hat{\tau}.$$

An identical argument shows that

$$|\gamma_\ell(\tau)| \le C(1-\bar{\alpha})^\ell \qquad \text{whenever} \quad \tau \ge \hat{\tau}, \; \ell = 0, 1, \ldots$$

Summarising, we have proved that,

$$|\gamma_\ell(\tau)| \le \begin{cases} C(1-\alpha\tau)^\ell \le C e^{-\alpha\ell\tau} & \tau < \hat{\tau} \\ C(1-\bar{\alpha})^\ell \le C e^{-\bar{\alpha}\ell}, & \tau \ge \hat{\tau} \end{cases}$$

Making $\hat{\tau}$ smaller if necessary we achieve $\bar{\alpha} = \alpha\hat{\tau}$. Denoting by i_* the smallest integer such that $\Delta_t \Lambda_i > \hat{\tau}$ we see that

$$|\gamma_\ell(\Delta_t \Delta_i)| \le \begin{cases} C e^{-\alpha_\ell \Delta_t \Delta_i} & i < i_* \\ C e^{-\alpha_\ell \tau} & i \ge i_* \end{cases} \tag{5.16}$$

c) We now assume that $w = 1$ is the only essential root of $p(\xi)$. The value a_1 of (5·15) is now equal to -1 by the consistency relationship. Thus for $\Delta_t \Delta_i$ sufficiently small

$$\xi w(\Delta_t \Delta_i) = 1 - \Delta_t \Delta_i + 0(\Delta_t^2 \Delta_i^2) = e^{-\Delta_t \Delta_i} + g$$

where g is an analytic function of $\Delta_t \Delta_i$ and $g = 0(\Delta_t^2 \Delta_i^2)$ at $\Delta_t \Delta_i = 0$

Expanding $p(\xi_w(\Delta_t \Lambda_i))$ about the point $e^{-\Delta_t \Delta_i}$ and equating to zero we have by (3·4) that

$$p(\xi_w(\Delta_t\Lambda_i)\ ,\ \Delta_t \Lambda_i) = \rho(e^{-\Delta_t \Lambda_i}) + \sum_{r=1}^{m} (\Delta_t \Lambda_i)^r \sigma_r (e^{-\Delta_t \Lambda_i}) + g\rho'(e^{-\Delta_t \Lambda_i})$$

$$+ 0(g^2) + o(\Delta_t \Delta_i g) = 0.$$

By substituting $y(t) = e^{-\Delta it}$ it into (3·2) and letting $t = 0$ we deduce

$$\rho(e^{-\Delta_{t+}\Lambda_i}) + \sum_{r=1}^{m} (\Delta_t\Lambda_i)^r \sigma_r (e - \Delta_t\Lambda_i) = C_{q+1} \Delta_t^{q+1} (-\Lambda_i)^{q+1} + 0\left((\Delta_t\Lambda_i)^{q+2}\right) .$$

Consequently, by combining the above expressions

$$g\rho'(e - \Delta_t\Lambda_i) = - C_{q+1} (-\Delta_t\Lambda_i)^{q+1} + 0(\Delta_t\Lambda_i) + 0\left((\Delta_t\Lambda_i)^{q+2}\right) + 0(g^2)$$

and thus, using $\rho'(e - \Delta_t\Lambda_i) = \rho'(1) + 0(\Delta_t\Lambda_i)$

$$g = \frac{(-1)^q}{\rho'(1)} C_{q+1} (\Delta_t\Lambda_i)^{q+1} + 0\left((\Delta_t\Lambda_i)^{q+2}\right) \equiv C(\Delta_t\Lambda_i)^{q+1} + 0\left((\Delta_t\Lambda_i)^{q+2}\right)$$

With the above expression of g we have established the bound,

$$\xi_w(\Delta_t\Lambda_i) \le e^{-\Delta_t\Lambda_i} \left[+C(\Delta_t\Lambda_i)^{q+1} \right] < 1$$

whenever $\Delta_t \Delta_i$ is sufficiently small. Utilising a previous result, we realise that the other roots $\{\xi_{w_i}\}^k{}_{j=2}$ of $p(\xi, \Delta_t\Lambda_i)$ satisfy $|\xi_{w_i}| < 1 - \frac{\theta}{2}$, given $\Delta_t \Delta_i$ sufficiently small. Therefore, we can select a value $\hat{\tau} > 0$ such that, for $0 < \Delta_t \Lambda_i < \hat{\tau}$

$$|\xi_{w_i}(\Delta_t\Lambda_i) \le e^{-\Delta_t\Lambda_i} \left[1 + C(\Delta_t\Lambda_i)^{q+1} \right] < 1 \qquad j-1,2,\dots,k.$$

Extending the argument as before we easily achieve

$$\left|\gamma_\ell (\Delta_t\Lambda_i)\right| \le C e^{-\ell\Delta_t\Lambda_i} \left[1 + c(\Delta_t\Lambda_i)^{q+1}\right]^\ell ,\Delta_t\Lambda_i < \hat{\tau}$$

Hence, for $\Delta_t\Lambda_i < \hat{\tau}$ and β, $0 < \beta < 1$

$$e^{-\ell\Delta t\Lambda i}\left[1 + C(\Delta_t\Lambda_i)^{q+1}\right]^{\ell} \le e^{-\frac{(1+\beta)}{2}\ell\Delta_t\Lambda_i}\left(e^{-\frac{(1-\beta)}{2}\Delta_t\Lambda_i}\left[1 + c(\Delta_t\Lambda_i)q + 1\right]\right)^{\ell}$$

and since $! + cx^{q+1} \le e^{\frac{1-\beta}{2}x}$ whenever $x < \tau_\beta < \hat{\tau}$ we have

$$\left|\gamma_\ell(\Delta_t\Lambda_i)\right| \le C\, e^{-\frac{(1+\beta)}{2}\ell\Delta_t\Lambda_i} \quad , \quad \Delta_t\Lambda_i < \tau_\beta$$

For $\Delta_t\Lambda_i \ge \tau_\beta$ we recall from a previous result that

$$|\gamma_\ell(\Delta_t\Lambda_i)| \le C\, e^{-\alpha\ell}, \quad 0 < \bar{\alpha} < 1$$

Making τ_β smaller if necessary we achieve $\bar{\alpha} = \frac{(1+\beta)}{2}\tau_\beta$.

Denoting by $i_*(\beta)$ the smallest integer such that $\Delta_t\Lambda_i > \tau_\beta$ see that

$$|\gamma_\ell(\Delta_t\Lambda_i)| \le \begin{cases} C\, e^{-\frac{(1+\beta)}{2}\ell\Delta_t\Lambda_i} & i < i_*(\beta) \\ C\, e^{-\frac{(1+\beta)}{2}\tau_\beta\ell} & i \ge i_*(\beta) \end{cases}$$

for some β, $0 < \beta < 1$.

By comparing coefficients in the expansion of

$$\frac{1}{\hat{p}(\xi,\tau)} = \frac{1}{\delta_k(\tau) + \xi\delta_{k-1}(\tau) + \ldots + \xi^k\delta_0(\tau)} = \gamma_0(\tau) + \xi\gamma_1(\tau) + \ldots.$$

we establish

$$\delta_k(\tau)\gamma_\ell(\tau) + \delta_{k-1}(\tau)\gamma_{\ell-1}(\tau) + \ldots + \delta_0(\tau) = \begin{cases} 1 & \ell = 0 \\ 0 & \ell > 0 \end{cases} \tag{5.18}$$

where $\gamma_\ell = 0$ for $\ell < 0$.

d) Henceforth, the following inequalities will be used extensively:

$$x\, e^{-\alpha x} \le (e\alpha)^{-1} < (2\alpha)^{-1} \tag{5.19}$$

$$x_p\, e^{-\alpha x} < \left(\frac{2\alpha}{p}\right)^{-p}$$

for any $x \geq 0$, $\alpha > 0$ and p a positive integer.

If we rewrite (5-12) with $n \equiv n - k - \ell$, multiply this by $Y_\ell(\Delta_t \Lambda_i)$, sum for $\ell = 0,1,\ldots, n-k$ and then apply (5·18) we prove

$$\epsilon_i^n = -\left[\delta_{k-1}(\Delta_t\Delta_i)\gamma_{n-k}(\Delta_t\Delta_i) + \ldots\ldots + \delta_0(\Delta_t\Delta_i)\gamma_{n-2k+1}(\Delta_t\Delta_i)\right]\epsilon_i^{k-1}$$
$$-\left[\delta_{k-2}(\Delta_t\Lambda_i)\gamma_{n-k}(\Delta_t\Lambda_i) + \ldots\ldots + \delta_0(\Delta_t\Delta_i)\gamma_{n-2k+2}(\Delta_t\Delta_i)\right]\epsilon_i^{k-2}$$
$$-\ldots\ldots - \delta_0(\Delta_t\Delta_i)\,\gamma_{n-k}(\Delta_t\Delta_i)\,\epsilon_i^0 + \sum_{\ell=0}^{n-k} d_i^{n-k-\ell}\,\gamma_\ell(\Delta_t\Delta_i) \tag{5.20}$$

Using (5·13), (5·16) , (5·20) and the inequalities (5·19) a bound on ϵ_i^n can be constructed as follows: for $i < i_*$

$$\left|\epsilon_i^n\right| \leq C\, e^{-\alpha(n-2k+1)\Delta_t\Lambda_i} \sum_{j=1}^{k-1} \left|\epsilon_i^j\right| + C\Delta_t^{q+1} \sum_{\ell=0}^{n-k} \Lambda_i^{q+1}\left|U_i^0\right| e^{-(n-k-\ell)\Delta_t\Lambda_i}\, e^{-\alpha\ell\Delta_t\Lambda_i} \tag{5.21}$$

Note the for $n\,\Delta_t \geq t_0$ and $(2k-1)\,\Delta_t \leq t_{0/2}$

$$e^{-\alpha(n-2k-1)\Delta_t\Lambda_i} \leq e^{-\frac{1}{2}\alpha t_0\Lambda_i} \leq C(t_0)\,\Lambda_i^{-s} \tag{5.22}$$

where s will be determined later. For $a - 1 \geq 0$

$$\Delta_t\, e^{-(n-k)\Delta_t\Lambda_i} \sum_{\ell=0}^{n-k} e^{-(\alpha-1)\ell\Delta_t\Lambda_i} \leq (n-k+1)\,\Delta_t\, e^{-(n-k)\Delta_t\Lambda_i}$$
$$\leq (n-k)\,\Delta_t\, e^{-(n-k)\Delta_t\Lambda_i} \leq \frac{C}{\Lambda_i}\, e^{-\frac{(n-k)}{2}\Delta_t\Lambda_i}$$
$$\leq \frac{C}{\Lambda_i}\; e - t_0\,\Lambda\;\; i/4 \leq C(t_0)\,\Lambda_i^{-(q+1)}$$

For $\alpha - 1 < 0$

$$S = \sum_{\ell=0}^{n=k} e^{-(\alpha-1)\ell\Delta_t\Lambda_i} \leq \frac{e^{-(\alpha-1)(n-k+1)\Delta_t\Lambda_i}}{e^{-(\alpha-1)\Delta_t\Lambda_i - 1}} \qquad \text{Hence,}$$

$$S\Delta_t \; e^{-(n-k)\,\Delta_t\Lambda_i} \le \frac{C\,\Delta_t\, e^{-\alpha(n-k)\Delta_t\Lambda_i}}{e\,(1-\alpha)\Delta_t\Lambda_i - 1} \le \frac{C\; e^{-\alpha(n-k)\,\Delta_t\Lambda_i}}{(1-\alpha)\Lambda_i}$$

$$\le \frac{C\; e^{-\alpha to \Lambda_{i/2}}}{\Lambda_i} \qquad \le \; C(t_0)\,\Lambda_i^{-\,(q+1)}$$

Thus, we have shown that

$$\Delta_t\, e^{-(n-k)\Delta_t\Lambda_i} \sum_{\ell=0}^{n=k} e^{-(\alpha-1)\,\ell\,\Delta_t\Lambda_i} \le C(t_0)\Lambda_i^{-(q+1)} \tag{5.23}$$

Collecting together (5·21)—(5·23), we conclude that whenever $i < i_*$,

$$\left|\epsilon_i^n\right| \le C\,(t_0)\;\;\Lambda_i^{-s} \sum_{j=1}^{k=1} \left|\epsilon_i^j\right| + C\,(t_0)\;\Delta_t^q \left|U_0^i\right| \tag{5.24}$$

For $i > i_*$, using (5·14) , (5.16) and (5.20)

$$\left|\epsilon_i^n\right| \le C\,e^{-\alpha\hat{\tau}(n-2k+1)} \sum_{j=1}^{k-1} \left|\epsilon_i^j\right| + C\left|U_i^0\right| \sum_{\ell=0}^{n-k} e^{-\Lambda_i(n-k-\ell)\Delta_t}\, e^{-\alpha\hat{\tau}\ell} \tag{5.25}$$

But $$e^{-\alpha\hat{\tau}(n-2k+1)} \le Ce^{-\alpha\hat{\tau}n} \le Cn^{-q} \le C(t_o)\Delta_t^q \tag{5.26}$$

as $n\Delta_t \ge t_o$. Also,

$$S = \sum_{\ell=0}^{n-k} e^{-\wedge_i(n-k-\ell)\Delta_t}\;\; e^{-\alpha\hat{\tau}\ell} \le \sum_{\ell=o}^{n-k} e^{-\hat{\tau}(n-k-\ell+a\ell)}$$

$$\le e^{-\hat{\tau}(n-k)} \sum_{\ell=o}^{n-k} e^{-\hat{\tau}(\alpha-1)\,\ell}$$

For $\alpha - 1 \ge 0$

$$S \le (n-k-1)e^{-\hat{\tau}(n-k)} \le 2(n-k)e^{-\hat{\tau}(n-k)} \le Ce^{-\hat{\tau}(n-k)}/2$$
$$\le Ce^{-\hat{\tau}n/2} \le Cn^{-q} \le C(t_o)\Delta_t^q$$

Similarly, for $a - 1 < 0$

$$S \le \frac{e^{-\hat{\tau}(n-k)}\, e^{\hat{\tau}(1-\alpha)(n-k+1)}}{e^{\hat{\tau}(1-\alpha)} - 1} \le \frac{Ce^{-\hat{\tau}\alpha(n-k+1)}}{\hat{\tau}(1-\alpha)} \le Ce^{-\hat{\tau}\alpha n} \le C(t_o)\Delta_t^q .$$

Combining we have proved that

$$\sum_{\ell=0}^{n-k} e^{-\Lambda_i (n-k-\ell)} \Delta_t e^{-\alpha \hat{\tau} \ell} \le C(t_0) \Delta_t^q \tag{5.27}$$

and the expressions (5.25) - (5.27) yield, for $i \ge i_*$

$$|\epsilon_i^n| \le C(t_0) \Delta_t^q \left\{ \sum_{j=1}^{k-1} |\epsilon_i^j| + |U_i^0| \right\} \tag{5.28}$$

From the bounds (5.24) and (5.28) we achieve

$$\sum_{i=1}^{d} |\epsilon_i^n|^2 \le C(t_0) \left\{ \Delta_t^{2q} \sum_{i=1}^{d} |U_i^0|^2 + \sum_{i<i_*} \Lambda_i^{-2s} \sum_{j=1}^{k-1} |\epsilon_i^j|^2 \right.$$

$$\left. + \Delta_t^{2q} \sum_{i \ge i_*} \sum_{j=1}^{k-1} |\epsilon_i^j|^2 \right\}$$

Using $|\epsilon_i^j| \le |U_i^0| + |U_i^j|$ we prove that

$$\| e_2 \|^2 = \sum_{i=1}^{d} |\epsilon_i^n|^2 \le C(t_0) \left\{ \Delta_t^{2q} \sum_{j=0}^{k-1} \| U^j \|^2 + \sum_{i<i_*} \Lambda_i^{-2s} \sum_{j=1}^{k-1} |\epsilon_i^j|^2 \right\} \tag{5.29}$$

Mihlin [9] has proved that $\Lambda_i \ge \lambda_i \ge c i^{\frac{2}{N}}$, c a positive constant.. Thus for any $s > N$

$$\sum_{i=1}^{d} \Lambda_i^{-s} \le \sum_{i=1}^{\infty} \lambda_i^{-s} \le C.$$

We use this result frequently in the following analysis. Let

$e_3 = \sum_{i>i_*} \sum_{j=1}^{k-1} \Lambda_i^{-2s} |\epsilon_i^j|^2$. We can write ϵ_i^j as

$$\epsilon_i^j = U_i^o e^{-j\Delta_t \Lambda_i} - U_i^j = e^{-j\Delta_t \Lambda_i} (U_i^o - \overline{U}_i^o) + e^{-j\Delta_t \Lambda_i} (\overline{U}_i^o - u_i^o)$$

$$+ (e^{-j\Delta_t \Lambda_i} - e^{-j\Delta_t \lambda_i}) u_i^o + (u_i^j - \overline{U}_i^j) + (\overline{U}_i^j - U_i^j)$$

from which

$$|e_3| \leq C\sum_{j=o}^{k-1}\sum_{i<i_*}\Lambda_i^{-2s}|u_i^j-\overline{U}_i^j|^2+\sum_{j=o}^{k-1}\sum_{i<i_*}\Lambda_i^{-2s}|\overline{U}_i^j-U_i^j|^2$$

$$+C\sum_{j=1}^{k-1}\sum_{i<i_*}\Lambda_i^{-2s}|e^{-j\Delta_t\Lambda_i}-e^{-j\Delta_t\lambda_i}|^2|u_i^o|^2$$

(5.30)

The expression (5.30) can be investigated by using (5.3) and (5.4), whence

$$e_4^j=\sum_{i<i_*}\Lambda_i^{-2s}|u_i^j-\overline{U}_i^j|^2\leq\lambda_1^{-2s}\sum_{i=1}^{\infty}|u_i^j-\overline{U}_i^j|^2\leq c\,\|u^j-U^j\|^2$$

$$e_5^j=\sum_{i<i_*}\Lambda_i^{-2s}|\overline{U}_i^j-U_i^j|^2 .\quad \text{Now,}\quad U_i^j-\overline{U}_i^j=\int_\Omega U^j(\psi_i-\psi_i)dx$$

i .e . $|\overline{U}_i^j-U_i^j|^2\leq C\,\|U^j\|^2\,h^{2(p+1)}\lambda_i^{p+1}$, but as $\lambda_i\geq ci^{\frac{2}{N}}$ the series

$\sum\lambda_i^{-(2s-p-1)}$ is c onvergent if we select $2s=p+1+N$.

Thus

$$e_5^j\leq C\,\|U^j\|^2\,h^{2(p+1)}$$

$$e_6^j=\sum_{i<i*}|e^{-j\Delta t\,i}-e^{-j\Delta t\lambda i}|^2\,\Lambda_i^{-2s}\,|u_i^0|^2$$

$$\leq\sum_{i<i_*}\Lambda_i^{-2s}|j\Delta_t\Lambda_i-j\Delta_t\lambda_i|^2|u_i^0|^2\leq j^2\Delta_t^2h^{4p}\sum_{i=1}^{\infty}\lambda_i^{-2(s-p+1)}|u_i^o|^2$$

and selecting $s=p+1+\frac{N}{2}$ we have

$e_6^j\leq C\Delta_t^2h^{4p}\,\|g\|^2$.

Substituting the above bounds in (5.30) we establish a bound on $|e_3|$, namely

$$|e_3|\leq C\left\{\sum_{j=0}^{k-1}\|u_j-U^j\|^2+h^{2(p+1)}\sum_{j=0}^{k-1}\|U^j\|^2+\Delta_t^2h^{4p}\,\|g\|^2\right\} \tag{5.31}$$

The desired result is obtained by substituting (5·31) into (5·29) and using the inequality

$$\| U^j \| \le \| U^j - u^j \| + \| u^j \| \le \| U^j - u^j \| + \| g \|$$

i.e. $$\| e_2 \| \le C(t_o)\left\{ \sum_{j=0}^{k-1} \| U^j - u^j \| + (h^{p+1} + \Delta_t^q) \| g \| \right\} \qquad (5.32)$$

e) We now extend the analysis of section d to the situation when w =1 is the only essential root of (ξ). Using (5·13), (5·17),(5·20) and the inequalities (5·19) a bound on $\in_i^n$ is constructed as follows; for $i < i_*(\beta)$

$$|\in_i^n| \le Ce^{-\frac{(1+\beta)}{2}(n-2k+1)\Delta_t \Lambda_i} \sum_{j=1}^{k-1} |\in_i^j|$$

$$+ C\Delta_t^{q+1} \sum_{\ell=o}^{n-k} \wedge_i^{q+1} \, |U_i^0| \, e^{-(n-k-\ell)\Delta_t \wedge_i} \, e^{-\frac{(1+\beta)\ell\Delta_t\Lambda_i}{2}} \qquad (5.33)$$

Now for $n\Delta_t \ge t_0$ and $(2k-1)\Delta_t \le t_{0/2}$

$$e^{-\frac{(1+\beta)}{2}(n-2k-1)\Delta_t\Lambda_i} \le Ce^{-\beta n \Delta_t \wedge_i} e^{-\frac{(1-\beta)}{2} t_0 \Lambda i/2}$$

$$\le C(t_o, \beta) e^{-\beta n \Delta_t \lambda_1} \Lambda_i^{-s} \qquad (5.34)$$

where as before, s will be determined later. Also, define

$$S = \Delta_t e^{-(n-k)\Delta_t\Lambda_i} \sum_{\ell=o}^{n-k} e^{\frac{(1-\beta)}{2}(n-k+1)\Delta_t\Lambda_i} \quad \text{and hence}$$

$$S \le \Delta_t e^{-(n-k)\Delta_t\Lambda_i} \frac{e^{\frac{(1-\beta)}{2}(n-k+1)\Delta_t\Lambda_i}}{e^{\frac{(1-\beta)}{2}\Delta_t\Lambda_i} - 1}$$

$$\le \frac{Ce^{\frac{(1-\beta)}{2}(n-k+1)\Delta_t\Lambda_i}}{(1-\beta)\wedge_i} \le \frac{C(\beta)}{\Lambda_i} e^{-\beta n \Delta_t \Lambda_i} e^{-\frac{(1-\beta)(n-k+1)\Delta_t\Lambda_i}{2}}$$

$$\le \frac{C(\beta)}{\Lambda_i} e^{-\beta n \Delta_t \lambda_1} e^{-\frac{(1-\beta)}{2} t_0 - \Lambda i/2} \le C(t_o, \beta) e^{-\beta n \Delta_t \lambda_1} \Lambda_i^{-(q+1)}$$

(5.35)

From (5·33) - (5·35) we have whenever $i < i_*(\beta)$

$$|\in_i^n| \leq C(t_0, \beta) e^{-Bn\Delta_t \lambda_1} \left\{ \Lambda_i^{-s} \sum_{j=1}^{k-1} |\in_i^j| + \Delta_t^q \; | U_i^0 | \right\} \tag{5.36}$$

Similarly, for $i \geq i_*(\beta)$, using (5·14), (5·18) and (5·20) we have

$$|\in_i^n| \leq Ce^{-(\frac{1+\beta}{2})\tau(n-2k+1)} \sum_{j=1}^{k-1} |\in_i^j|$$

$$+ C\,|U_i^0| \sum_{\ell=0}^{n-k} e^{-\Lambda_i (n-k-\ell)\Delta_t} e^{-(\frac{1+\beta}{2})\tau\ell}$$

(5.37)

where, for simplicity, we denote $\tau \equiv \tau_\beta$. But

$$e^{-(\frac{1+\beta}{2})\tau(n-2k+1)} \leq Ce^{-\beta n\tau} e^{-(\frac{1-\beta}{2})\tau n} \leq C(\beta)e^{-\beta n\tau} n^{-q}$$

$$\leq C(t_0, \beta)e^{-\beta n\tau} \Delta_t^q \tag{5.38}$$

Also, let $S = \sum_{\ell=0}^{n-k} e^{-\Lambda_i (n-k-\ell)\Delta_t} e^{-(\frac{1+\beta}{2})\tau\ell}$, and thus

$$S \leq e^{-\tau(n-k)} \sum_{\ell=0}^{n-k} e^{\tau(\frac{1-\beta}{2})\ell} \leq e^{-\tau(n-k)} \frac{e^{\tau(\frac{1-\beta}{2})(n-k-1)}}{e^{\tau(\frac{1-\beta}{2})} - 1}$$

$$\leq \frac{Ce^{-(\frac{1-\beta}{2})\tau n}}{\tau(1-\beta)} \leq C(t_0, \beta)e^{-\beta n\tau} \Delta_t^q \tag{5.39}$$

The expressions (5·37) -(5·39) yield that, for $i \geq i_*(\beta)$

$$|\in_i^n| \leq C(t_0, \beta)\Delta_t^q \, e^{-\beta n \Delta_t \lambda_1} \left\{ \sum_{j=1}^{k-1} |\in_i^j| + |U_i^0| \right\} \tag{5.40}$$

where we take Δ_t, sufficiently small to allow $\Delta_t\lambda_1 < \tau$.

Following a course identical to section d we arrive at the result

$$\| e_2 \| \le C(t_0, \beta) e^{-\beta n \Delta_t \lambda_1} \left\{ \sum_{j=0}^{k-1} \| U^j - u^j \| + (h^{p+1} + \Delta_t^q) \| g \| \right\}$$

(5.41)

f) The error $e_1 = u(x, n\Delta_t) - U(x, n\Delta_t)$ will now be bounded. From (5·6) and (5·7) we have

$$e_1 = \sum_{i=1}^{\infty} g_i e^{-n\Delta_t \lambda_i} \psi_i - \sum_{i=1}^{d} U_i^0 e^{-n\Delta_t \Lambda_i} \psi_i$$

$$= \sum_{i>d} g_i e^{-n\Delta_t \lambda_i} \psi_i + \sum_{i=1}^{d} (e^{-n\Delta_t \lambda_i} - e^{-n\Delta_t \Lambda_i}) g_i \psi_i$$

$$+ \sum_{i=1}^{d} e^{-n\Delta_t \Lambda_i} (g_i - \bar{g}_i) \psi_i + \sum_{i=1}^{d} e^{-n\Delta_t \Lambda_i} \bar{g}_i (\psi_i - \psi_i)$$

$$+ \sum_{i=1}^{d} e - n\Delta_t \Lambda_i (\bar{g}_i - U_i^0) \psi_i \qquad (5.42)$$

Zlainal [15] uses a technique from Thome'e [13] to show that $\lambda_{d+1} \ge ch^{-2}$. Hence, using (5·3) and (5·4) we have for some β, $0 \le \beta < 1$

$$e7 \equiv \sum_{i>d} e^{-n\Delta_t \lambda_i} g_i \psi_i \le e^{-\beta n \Delta_t \lambda_1} \sum_{i>d} e^{-(1-\beta) n \Delta_t \lambda_{d+1}} g_i \psi_i$$

$$\le e^{-\beta n \Delta_t \lambda_1} e^{(1-\beta) t_0 \lambda_{d+1}} \sum_{i=1}^{\infty} g_i \psi_i \le C(t_0, \beta) e^{-\beta n \Delta_t \lambda_i} \lambda_{d+1}^{-(\frac{p+1}{2})} \sum_{\ell=1}^{\infty} g_i \psi_i$$

i .e. $\| e_7 \| \le C(t_0, \beta) h^{P+1} e^{-\beta n \Delta_t \lambda_1} \| g \|$.

Let $e_8 \equiv \sum_{i=1}^{d} (e^{-n\Delta_t \lambda_i} - e^{-n\Delta_t \Lambda_i}) g_i \psi_i$. By the mean—value theorem

$$e_8 \le \sum_{i=1}^{d} n\Delta_t \,|\lambda_i - \Lambda_i|\, e^{-n\Delta_t \Lambda_i} g_i \psi_i$$

$$\le Cn\,\Delta_t h^{2p} e^{-\beta n\Delta_t \lambda_i} \sum_{i=1}^{d} \lambda_i^{p+1} e^{-(1-\beta)n\Delta_t \Lambda_i} g_i \psi_i$$

$$\le C(\beta)\, h^{2p} e^{-\beta n\Delta_t \lambda_1} \sum_{i=1}^{d} \lambda_i^{p} e^{-(\frac{1-\beta}{2})t_o \lambda_i} g_i \psi_i$$

$$\le C(t_o, \beta) h^{2p} e^{-\beta n\Delta_t \lambda_1} \sum_{i=1}^{\infty} g_i \psi_i$$

i .e. $\quad \| e_8 \| \le C(t_o, \beta) h^{2p} e^{-\beta n\Delta_t \lambda_1} \| g \|$

Let $e_9 \equiv \sum_{i=1}^{d} e^{-n\Delta_t \Lambda_i} (g_i - \bar{g}_i)\psi_i$. However $g_i - \bar{g}_i = \int_\Omega g(x)(\psi_i - \psi_i)dx$

and thus by Cauchy's inequality

$$\| e_9 \| \le C \| g \| h^{p+1} e^{-\beta n\Delta_t \lambda_1} \sum_{i=1}^{d} e^{-(1-\beta)n\Delta_t \lambda_i} \lambda_i^{\frac{p+1}{2}}$$

$$\le C(t_o, \beta) \| g \| h^{p+1} e^{-\beta n\Delta_t \lambda_1} \sum_{i=1}^{\infty} \lambda_i^{-N} \quad \text{by (5.19) and } n\Delta_t \ge t_o$$

$$\le C(t_o, \beta) \| g \| h^{p+1} e^{-\beta n\Delta_t \lambda_1} \quad \text{as } \lambda_i \ge c i^{\frac{2}{N}}$$

Let $e_{10} = \sum_{i=1}^{d} e^{-n\Delta_t \Lambda_i} \bar{g}_i (\psi_i - \psi_i)$. Thus

$$\| e_{10} \| \le Ch^{p+1} e^{-\beta n\Delta_t \lambda_i} \| g \| \sum_{i=1}^{d} e^{-(1-\beta)n\Delta_t \lambda_i} \lambda_i^{\frac{p+1}{2}}$$

$$\le C(t_o, \beta) h^{p+1} e^{-\beta n\Delta_t \lambda_1} \| g \| \sum_{i=1}^{\infty} \lambda_i^{-N} \le C(t_o, \beta) h^{p+1} e^{-\beta n\Delta_t \lambda_1} \| g \|$$

Finally $e_{11} \equiv \sum_{i=1}^{d} e^{-n\Delta_t \Lambda_i} (\bar{g}_i - U_i^o)\psi_i$. If U° is the orthogonal projection of $g(x)$ with respect to the L_2 -inner product then $e_{11} = 0$, otherwise

$$e_{11} = \sum_{i=1}^{d} e^{-n\Delta_t \Lambda_i} \Big((\bar{g}_i - g_i) + (g_i - U_i^o)\Big)\psi_i \quad \text{and}$$

$$\| e_{11} \| \le C(t_o, \beta)\ e^{-\beta n \Delta_t \lambda_1} \left\{ \| g \|\, h^{p+1} + \| g - U^o \| \right\} \qquad \text{(cf, eg)}$$

Using (5·42) and the above bounds we conclude that

$$\| e_1 \| \le \begin{cases} c\,(t_o,\beta) e^{-\beta n \Delta_t \lambda_1} (h^{p+1} \|g\| + \|g-u^\circ\|) \\ c(t_o,\beta) e^{-\beta n \Delta_t \lambda_1} (h^{p+1} \|g\| \text{ . If } U^\circ \text{ is the } L_2\text{-inner product projection of } g(x) \text{ onto } V_h^p \end{cases} \qquad (5{\cdot}43)$$

for some arbitary $\beta,\ 0 \le \beta < 1$

g) Returning to (5·11) we have

$$\sum_{j=o}^{k} \delta_j (\Delta_t \Lambda_i) U_i^{n+j} = o.$$

Rewrite the above with n - n - k - ℓ, multiply this by $\gamma_\ell(\Delta_t \Lambda_i)$, sum for ℓ=0,1...,n-k and apply (5·18) to achieve the expression (5.20) with $\in_i^j$ replaced by U_i^j, and $d_i \equiv 0$. Let us assume that Δ_t is sufficiently small so that $\Delta_t \lambda_1 < \hat{\tau}$ Using the remodelled expression of (5·20) and (5·16) we obtain

$$| U_i^n | \le Ce^{-\alpha n \Delta_t \lambda_1} \sum_{j=o}^{k-1} | U_i^j |$$

from which it follows that (5·44)

$$\| U^n \| = \left(\sum_{i=1}^{d} | U_i^n |^2 \right)^{\frac{1}{2}} \le C\, e^{-\alpha n \Delta_t \lambda_1} \sum_{j=0}^{k-1} \| U^j \|$$

which is the desired asymptotic result.

h) Initial Approximants U^j to $u(x, j\Delta_t)$, $j=0,1,\ldots,k-1$.

This section is concerned with the estimate $\| e_2 \|$ under the assumption that U° is the orthogonal projection of g(x) onto V_h^p with respect to the L_2-inner product and $\{U^j\}_{j=1}^{k-1}$ are the approximate solutions of (2.2) at time $t=j\Delta_t$ obtained by a weakly A_0-stable Padé scheme of order q -1 .

Other viable methods for deriving these approximants include the weakly A_0-stable Runge-Kutta schemes. Such schemes have been thoroughly investigated by Crouzeix [2] and we refer the reader to his thesis for an account of these schemes.

A difference method derived from a Pade approximation of order q - 1 is a one-step method of the type

$$y_{n+1} - y_n = \sum_{s=o}^{1} \sum_{r=1}^{\tilde{m}} \tilde{\beta}_{rs} \Delta_t^r y_{n+s}^r \qquad (5.45)$$

where

$$R(\tau(\equiv \frac{1 + \sum_{r=1}^{\tilde{m}} (-1)^r \tilde{\beta}_{ro} \tau^r}{1 + \sum_{r=1}^{\tilde{m}} (-1)^{r-1} \tilde{\beta}_{rl} \tau^r} \text{ is an approximation}$$

$e^{-\tau}$, such that

$$| e^{-\tau} - R(t) | \leq C \tau^q \text{ as } \tau \to 0 \qquad (5\cdot46)$$

We note that any Pade scheme is a one-step, multiderivative method and satisfies (see (3·2)) the relation

$$y_{n+1} - y_n - \sum_{s=o}^{1} \sum_{r=1}^{\tilde{m}} \tilde{\beta}_{rs} \Delta_t^r y_{n+s}^r = \tilde{C}_q \Delta_t^q y^q (n\Delta_t) + O(\Delta_t^{q+1})$$

$$\leq \tilde{G} \Delta_t^q \sup_{o\leq s\leq 1} \left\{ | y^q (n + s) \Delta_t) | \right\} \qquad (5.47)$$

A Padé scheme is said to be weakly A —stable (see | 2 |) if $|R(\tau)| \leq 1$, for any $\tau \geq 0$. The inequality (5·46) is stated to hold for small τ. However, as $|e^{-\tau} - R(\tau)| \geq 2\tau \geq 0$,(5·46) is satisfied a fortiori for any $\tau \geq 0$. Applying the scheme (5·45) to the system of differential equations (2·2) we see immediately from an obvious adaptation of (5·10) that

$$\left(1 + \sum_{r=1}^{\tilde{m}} \Delta_t^r \beta_{rl} (-1)^{r-1} \Lambda_i^r\right) U_i^{j+1} - \left(1 + \sum_{r=1}^{\tilde{m}} \Delta_t^r \tilde{\beta}_{ro} (-1)^r \Lambda_i^r\right) U_i^j = 0$$

$$\text{or} \quad U_i^{j+1} = R(\Delta_t \Lambda_i) U_i^j, \quad j = 0,1,\ldots\ldots, k-2. \tag{5.48}$$

The recurrence equation (5·48) yields

$$U_i^{j+1} = \left[R(\Delta_t \Lambda_i)\right]^{j+1} U_i^o \tag{5.49}$$

It is easily derived from (5·8) and (5·49) that

$$\in_i^{j+1} = U_i^o \left(e^{-\Lambda_i (j+1) \Delta_t} - \left[R(\Delta_t \Lambda_i)\right]^{j+1} \right)$$

and by using the definition of weak A_0—stability, and (5·46)

$$|\in_i^{j+1}| \leq | U_i^o | \, | e^{-\Lambda_i (j+1) \Delta_t} - \left[R(\Delta_t \Lambda_i)\right]^{j+1} |$$

$$\leq (j+1) | U_i^o | \left| e - \Lambda_i \Delta_t - R(\Delta_t \Lambda_i) \right| \leq C | U_i^o | \Delta_t^q \Lambda_i^q$$

$$j = 0,1,\ldots,k-2 \tag{5·50}$$

Consequently, returning to (5·29) we note

$$\sum_{j=1}^{k-1} \sum_{i<i_*} \Lambda_i^{-2s} |\in_i^j|^2 \leq C \Delta_t^{2q} \sum_{i<i_*} \Lambda_i^{-2(s-q)} | U_i^o |^2$$

$$\leq C \Delta_t^{2q} \| U^o \|^2 \text{ by selecting } s = q + \frac{N}{2} \tag{5·51}$$

The initial approximant U° to $g(x)$ is defined to be the projection of $g(x)$ onto V_h^p by the L_2 - inner product, and is thus well known to satisfy,

$$\| U^\circ \| \leq \| g \|$$

Using the definition of weak A_0 -stability, namely $| R(\tau) | \leq 1$, for $\tau \geq 0$ we have by (5·49)

$$\| U^j \|^2 = \sum_{i=1}^{d} | U_i^j |^2 \leq \sum_{i=1}^{d} | U_i^o |^2 = \| U^o \|^2 \leq \| g \|^2 \quad j = 1,2,\ldots\ldots\ldots, k-1 \qquad (5.52)$$

The expression (5·29) can now be reformulated by (5·51)and (5·52) to read

$$||e_2|| \leq C(t_o)\, \Delta_t^q\, ||g|| \qquad (5\cdot 53)$$

We are able to deduce immediately the corresponding result when $w=1$ is the only essential root of $\rho(\xi)$

i.e. $$\| e_2 \| \leq C(t_o,\beta)\, e^{-\beta n t \Delta \lambda 1}\, \Delta_t^q\, \| g \| \qquad (5\cdot 54)$$

The theorems can now be established. Theorem 1 is determined from the relation $\| u(x,n\Delta_t) - U^n \| \leq \| e_1 \| + \| e_2 \|$ and the bounds (5·32), (5·43) with $\beta = 0$, and (5·44). Its corollary follows immediatelt by using (5·53) instead of (5·32).Theorem 2 and its corollary follow from the bounds (5·41), (5·43), and (5·54).

6. <u>Practical Examples of L.M.S.D.Schemes</u>

To illustrate the multistep, multiderivative methods we select k = m = 2 and derive a family of fifth-order, A_0 -stable methods. Any fifth order method with k = m = 2 may be expressed as

$$(\alpha-1)y_n + (1-2\alpha)y_{n+1} + \alpha y_{n+2} = \Delta_t\Big\{(\frac{7}{15}-\beta)y'_n + \frac{8}{15}y'_{n+1}$$

$$+ \beta y'_n\Big\} + \Delta_t^2\Big\{(\frac{5}{72}+\frac{\alpha}{12}-\frac{\beta}{3})y''_n + (-\frac{19}{180}+\frac{5\alpha}{6}-\frac{4\beta}{3})y''_{n+1}$$

$$+ (\frac{1}{360}+\frac{\alpha}{12}-\frac{\beta}{3})y''_{n+2}\Big\} \qquad (6.1)$$

We test for A_o -stability by employing the Routh-Hurwitz criterion e.g. [7,pp.80], For simplicity we define, as before,

$$\mu_2(\tau) = \alpha + \beta\tau + (\frac{\beta}{3} - \frac{\alpha}{12} - \frac{1}{360})\tau^2$$

$$\mu_1(\tau) = (1-2\alpha) + \frac{8}{15}\tau + (\frac{4\beta}{3} - \frac{5\alpha}{6} + \frac{19}{180})\tau^2$$

$$\mu_0(\tau) = (\alpha-1) + (\frac{7}{15} - \beta)\tau + (\frac{\beta}{3} - \frac{\alpha}{12} - \frac{5}{72})\tau^2$$

for any $\tau > 0$. By (3·4) we require the roots of the polynomial

$p(\xi,\tau) = \sum_{j=o}^{2} \mu_j(\tau)\xi^j$ to be less than one in modulus, for all $\tau > 0$.

By the Rough-Hurwitz criterion this requirement is satisfied if,

$\mu_2(\tau) > \mu_1(\tau) - \mu_0(\tau)$

$$\text{i. e. } (4\alpha - 2) - \frac{\tau}{15} + (\frac{2\alpha}{3} - \frac{2\beta}{3} - \frac{8}{45})\tau^2 > 0 \qquad (6.2i)$$

$\mu_2(\tau) > \mu_0(\tau)$

$$\text{i. e. } 1 + (2\beta - \frac{7}{15})\tau + \tau^2/15 > 0 \qquad (6.2ii)$$

$\mu_2(t) + \mu_1(\tau) + \mu_0(\tau) > 0$

$$\text{i. e. } \tau + (2\beta - \alpha + \frac{1}{30})\tau^2 > 0 \qquad (6.2iii)$$

for all $\tau > 0$. Note that, by lemmas 1 and 2

$$\mu_2(\tau) > 0, \ (\frac{\beta}{3} - \frac{\alpha}{12} - \frac{1}{360}) > 0. \qquad (6.2iv)$$

The inequalities (6·2) are satisfied if

$$\alpha > \frac{1}{2}, \quad 2\beta - \alpha \geq -\frac{1}{30}$$

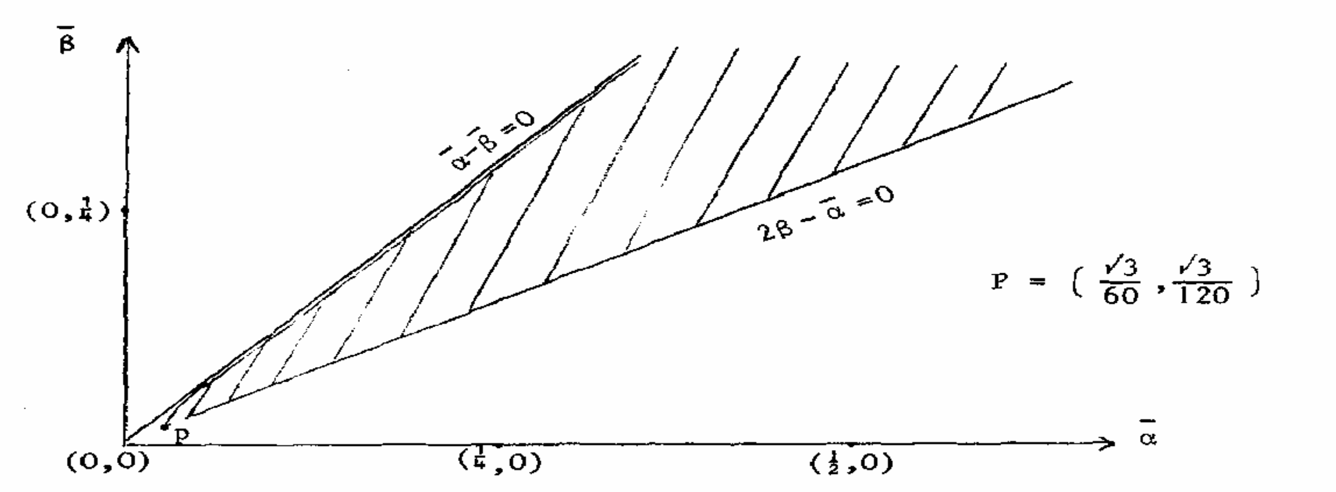
and $$(\frac{1}{15})^2 < 4(4\alpha - 2)(\frac{2\alpha}{3} - \frac{2\beta}{3} - \frac{8}{45})$$

The region defined is best seen if we change the basis and let

$$\bar{\alpha} = \alpha - \frac{1}{2}, \qquad \bar{\beta} = \beta - \frac{7}{30}$$

from which we deduce that

$$\bar{\alpha} > 0, \ 2\bar{\beta} - \bar{\alpha} > 0 \quad \text{and} \quad (\frac{1}{15})^2 < \frac{32}{3}\bar{\alpha}(\bar{\alpha} - \bar{\beta})$$



(Diagram I)

The shaded area of Diagram 1 contains the permissible values for $\bar{\alpha}$ and $\bar{\beta}$ We note that the error constant of the L.M.S.D. scheme (7·5) is given by

$$C_6 = -\frac{11}{21600} - \frac{\alpha}{240} + \frac{\beta}{90}$$

The selection of particular values from the admissible range of the parameters α and β is now considered. Any scheme proposed to solve the stiff system of equations (2·3) should exhibit certain characteristics, of which, the principle is related to the nature of the analytic solution.

Let us apply the scheme (3·1) to the scalar test equation $y = -\lambda y, \lambda > 0$. By the definition of A_0 -stability we know that the approximate solution $Y_n \to 0$ as $n \to \infty$. For $\lambda >> 0$ the solution Y_n approaches the solution of the difference equation

$$\sum_{j=0}^{k} \beta_{mj} \overline{Y}_{n+j} = 0 \qquad \text{as } \lambda \to \infty .$$

Without loss of generality we shall assume that the roots $\{\xi_i\}_{i=1}^{k}$ of the equation $\sigma_m(\xi) = 0$ (see 3·4) are real and distinct, then

$$Y_n = \sum_{i=1}^{k} a_i \xi_i^n \text{ as } \lambda \to \infty$$

where $\{a_i\}_{i=1}^{k}$ are constants determined by the initial values $\{Y_i\}_{i=o}^{k-1}$

By assumption we know that $|\xi_i| < 1$, $i = 1,2, \ldots,k$, and hence $Y_n \to 0$ as $n \to \infty$. This convergence has previously been referred to as stability at ∞. However, the rate of convergence may be increased by allowing the roots $\{\xi_i\}_{i=1}^{k}$ of $\sigma_m(\xi) = 0$ to be equal, or close to zero. Consequently, given a very stiff system of equations it is desirable to use a multistep scheme where the roots of $\sigma_m(\xi)$ are equal, or close to zero.

Equally, we desire that the normalised error constant, $\tilde{C}_{q+1}$, is small

$$\text{i.e. } \tilde{C}_{q+1} \equiv \frac{C_{q+1}}{\sum_{j=0}^{k} \beta_{lj}} \quad \text{where } C_{q+1} \text{ is defined by (3.2)}$$

Consequently, we advance the following possibilities:

$$\alpha = \frac{11}{20}, \quad \beta = \frac{79}{300}, \quad \tilde{C}_6 = \frac{1}{8000} \quad |\xi_1| \sim .86 \qquad (6.3i)$$

$$\alpha = \frac{3}{5}, \quad \beta = \frac{7}{24}, \quad \tilde{C}_6 = \frac{1}{4320} \quad |\xi_1| \sim .77 \qquad (6.3ii)$$

$$\alpha = \frac{2}{3}, \quad \beta = \frac{1}{3}, \quad \tilde{C}_6 = \frac{1}{2400} \quad |\xi_1| \sim .55 \qquad (6.3iii)$$

$$\alpha = \frac{23}{30}, \quad \beta = \frac{2}{5}, \quad \tilde{C}_6 = \frac{1}{1350} \quad \xi_1 = 0 \qquad (6.3iv)$$

where ξ_1 is the largest root in modulus of $\sigma_2\ (\xi)$.

Higher order A_0 -stable L.M. S.D. methods may be obtained by-allowing either or both of m and k to be greater than two. Without reference to the general class of such schemes we note the following particular examples;

k = 2, m = 3.

$$\frac{9}{10} y_{n+2} - \frac{4}{5} y_{n+1} - \frac{1}{10} y_n = \Delta_t \left\{ \frac{23}{40} y'_{n+2} + \frac{2}{5} y'_{n+1} + \frac{1}{40} y'_n \right\}$$
$$- \frac{3}{20} \Delta_t^2 y''_{n+2} + \frac{1}{60} \Delta_t^3 y''_{n+2} \qquad , \ q = 6, \ \tilde{C}_7 = - \frac{1}{12600} \qquad (6.4i)$$

$$\frac{15}{14} y_{n+2} - \frac{8}{7} y_{n+1} + \frac{1}{14} y_n = \Delta_t \left\{ \frac{39}{70} y'_{n+2} + \frac{16}{35} y'_{n+1} - \frac{1}{70} y'_n \right\}$$
$$- \frac{4}{35} \left\{ y''_{n+2} - y''_{n+1} \right\} + \frac{1}{105} \Delta_t^3 y''_{n+2}, \ q = 7, \ \tilde{C}_8 = - \frac{1}{176400} \qquad (6.4ii)$$

k = 3, m = 2

$$\sum_{j=o}^{3} \alpha_j y_{n+1} = \Delta_t \sum_{j=0}^{3} \beta_j y'_{n+j} - c \Delta_t^2 y''_{n+3}$$

where

$$\alpha_3 = \frac{11}{60} + \frac{39}{4}c \qquad \beta_3 = \frac{1}{20} + \frac{67}{12}c \qquad (6.4ii)$$

$$\alpha_2 = \frac{9}{20} - \frac{63}{4}c \qquad \beta_2 = \frac{9}{20} + \frac{9}{4}c$$

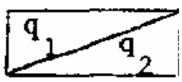
$$\alpha_1 = -\frac{9}{20} + \frac{9}{4}c \qquad \beta_1 = \frac{9}{20} - \frac{27}{4}c$$

$$\alpha_0 = -\frac{11}{60} + \frac{15}{4}c \qquad \beta_0 = \frac{1}{20} - \frac{13}{12}c$$

This is a sixth order method with error constant $\tilde{C}_7 = \frac{1}{80}(c - \frac{1}{35})$ unless $c = \frac{1}{35}$ which yields a seventh order scheme with error constant $\tilde{C}_8 = \frac{1}{19600}$. A_0 - stability is ensured by the condition

$$c > {}^{384}/17{,}275$$

From the relevant theory, e.g. Cryer [3], or by direct evaluations we have established the following table concerning maximum orders of A_0 -stable L.M.S.D. schemes. The diagram expresses for $1 \leq m + k \leq 5$:



q_1 = maximum order of A_0 -stable L.M.S.D.scheme for specific values of m and k.

q_2 = as q_1 but with added stipulation of stability as ∞.

m \ k	1	2	3	4
1	2 / 1	2 / 2	3 / 3	4 / 4
2	4 / 3	5 / 5	7 / 7	
3	6 / 5	8 / 8		
4	8 / 7			

b. Implementation

Any scheme proposed to solve the linear parabolic equation should be efficient in terms of computer storage and operations. For any finite element space V_h^p the matrices M and K are banded matrices, thus an efficient method of solution should preserve and utilise this characteristic. Remembering the definition of the matrices M and K we have immediately from (3·5) and (3·6)

$$\sum_{j=o}^{k} \alpha_j M \underline{U}^{n+j} - \sum_{j=o}^{k} \sum_{r=1}^{m} \Delta_t^r \beta_{rj} M \underline{U}_{(r)}^{n+j} = \underline{o}$$

where

$$M \underline{U}_{(r)}^{n+j} = -K \underline{U}_{(r-1)}^{n+j} \qquad r = 1,2,\ldots\ldots, m.$$

On combining these two equations we achieve

$$\sum_{j=o}^{k} \alpha_j \underline{U}^{n+j} - \sum_{j=0}^{k} \sum_{r=1}^{m} \Delta_t^r \beta_{rj} (-1)^r (M^{-1}k)^r \underline{U}^{n+j} = \underline{o} \qquad (6.5)$$

The equation (6-5) is obviously impractical as it entails full matrices M^{-1} $K,(M^{-1}K)^2$,... $(M^{-1}$ $K)^m$. However, by the use of complex arithmetic the sparseness of the matrices M and K is utilised. We illustrate this mode of implementation by reference to the family of equations (6·1). Equation (6·5) can be seen to be

$$\sum_{j=o}^{2} \tilde{u}_j(\Delta_t M^{-1}k)\underline{U}_{n+j} = 0$$

where $\tilde{\mu}_2(x) = \frac{\alpha}{\gamma} + \frac{\beta}{\gamma}x + x^2$ $\qquad$ $,\gamma = \frac{\beta}{3} - \frac{1}{360} - \frac{\alpha}{12}$

$$\tilde{\mu}_1(x) = \frac{(1-2\alpha)}{\gamma}\left[1 - \frac{8x}{15(2\alpha-1} + \frac{\left(-\frac{19}{360} + \frac{5\alpha}{6} - \frac{4\beta}{3}\right)x^2}{(2\alpha-1)}\right]$$

$$\mu_1(x) = \frac{\alpha-1}{\gamma}\left[1 - \frac{(\frac{7}{15} - \beta}{1-\alpha}x + \frac{\left(\frac{5}{72} + \frac{\alpha}{12} - \frac{\beta}{3}\right)x^2}{1-\alpha}\right], \quad \alpha \neq 1$$

(6.6)

The roots of $\tilde{\mu}_2(x)$ are readily seen to be complex whenever α and β are permissible. Thus let

$\tilde{\mu}_2(x) = (Z_2 - x)(\bar{Z}_2 - x)$ and further let

$Z_1^{(1)}, Z_1^{(2)}$ and $Z_o^{(1)}, Z_o^{(2)}$ be respectively the roots of $\gamma\tilde{\mu}_1(x)/_{1-2\alpha}$ and $\gamma\tilde{\mu}_o(x)/_{\alpha-1}$.

Consequently a simple manipulations shows that (6·6) is equivalent to

$$M\,\underline{U}^{n,1} = (M - Z_o^{(1)}\Delta_t k)\,\underline{U}^n$$

$$M\,\underline{U}^{n,2} = (M - Z_o^{(1)}\Delta_t k)\,\underline{U}^{n+1}$$

$$(Z_2 M - \Delta_t k)\,\underline{U}^{n,3} = \left(\frac{2\alpha - 1}{\gamma}\right)(M - Z_o^{(2)}\Delta_t k)\,\underline{U}^{n,1} + \left(\frac{1-\alpha}{\gamma}\right)(M - Z_1^{(2)}\Delta_t k)\,\underline{U}^{n,2}$$

$$\underline{U}^{n+2} = \frac{I_m\,\underline{U}^{n,3}}{I_m\,\bar{Z}_2}$$

Although three intermedi ate steps are necessary at each time interval it is necessary to invert only two matrices- For the particular example (6·3iv) only one intermediate step exists at each time interval, requiring the inversion of only one matrix.

The use of complex arithmetic, and the extra storage necessary, may be prohibitive. However, A—stable L.M.S.D. methods of arbitrary order have been investigated by several authors with the intention of simplifying the implementation. Of particular interest is the family of one—step Hermite formulae suggested by Makinson [8] and investigated fully by Norsett [11]. Norsett derived a family of A(o)—stable, one—step methods of order m + 1 where the coefficient matrix, $G_m\ (M^{-1}\ K)$, of $\underline{U}^{n+1}$ is given by

$$G_m\,(M^{-1}k) \equiv (I + \frac{\Delta_t}{\gamma} M^{-1}k)^m, \text{ for a specified parameter } \gamma.$$

Continuing with the construction of L.M.S.D. methods with k = m =2 we now establish a family of fourth order, A_0 -stable methods where the coefficient matrix of $\underline{U}^{n+2}$ has the same characteristics as $G_2\ (M^{-1}\ K)$. The family of fourth order schemes with the above properties is given by

$$\alpha y_{n+2} + (1-2\alpha)y_{n+1} + (\alpha-1)y_n = \Delta_t\{\beta\alpha y'_{n+2} + (\frac{1}{2} - \alpha + 4\beta\alpha - 3\beta\alpha)y'_{n+1}$$

$$+ (\alpha + \frac{1}{2} - 5\beta\alpha + 3\beta^2\alpha)y'_n\} + \Delta_t^2\left\{-\frac{\beta^2\alpha}{4}y''_{n+2} + (\frac{3\alpha}{2} - 4\beta\alpha + 2\beta^2\alpha - \frac{1}{12})y''_{n+1}\right.$$

$$\left. + (\frac{\alpha}{2} + \frac{1}{12} - 2\beta\alpha + \frac{5}{4}\beta^2\alpha)y''_n\right\} \qquad (6.7)$$

Applying the Routh-Hurwitz criterion we deduce that (6·7) is

A_o – stable if for any $\alpha > \frac{1}{2}$

$$1 - \frac{\sqrt{12}}{6} < \beta < \min\left\{1 - \frac{1}{2\alpha}\sqrt{\frac{2\alpha}{3}},\ \frac{2}{3} - \frac{1}{6\alpha}\sqrt{4\alpha^2 - 2\alpha}\right\}$$

and $\alpha^2(3\beta^2 + 1 - 4\beta)^2 < (4\alpha - 2)\{\beta^2\alpha - 2\beta\alpha + \alpha - \frac{1}{6}\}$

Alternatively, A_o -stability is ensured by $\alpha > \frac{1}{2}$ and

$$1 + \frac{1}{2\alpha}\sqrt{\frac{2\alpha}{3}} < \beta < 1 + \frac{\sqrt{12}}{6}$$

The normalised error constant of the scheme (6·7) is expressed by

$$\tilde{C}_5 = \alpha(\frac{\beta}{6} - \frac{1}{24} - \frac{\beta^2}{8} -) - \frac{1}{720}$$

As before,we require that the choices of values for α and β yield a balance between the stability of infinity and the error constant. However, the A -stability requirement on β forces the modulus of the roots of $\sigma_2(\xi)$ to be extremely close to one for small values of $\tilde{C}_5$. . The one important exception is when

$$\alpha = \frac{5 + 16\sqrt{10}}{90} \quad , \beta = \frac{12 - 2\sqrt{10}}{13}$$

$$\text{i.e} \left(\frac{5 + 16\sqrt{10}}{90}\right) y_{n+2} + \left(\frac{40 - 16\sqrt{10}}{45}\right) y_{n+1} + \left(\frac{16\sqrt{10} - 85}{90}\right) y_n$$

$$= \Delta_t \left\{ \left(\frac{7\sqrt{10} - 10}{45}\right) y'_{n+2} + \left(\frac{40 - 4\sqrt{10}}{45}\right) y'_{n+1} + \left(\frac{5 - \sqrt{10}}{15}\right) y'_n \right\}$$

$$- \Delta_t^2 \left(\frac{2\sqrt{10} - 5}{45}\right) y''_{n+2}$$

and $\tilde{C}_5 = \dfrac{4 - \sqrt{10}}{270}$.

The scheme (6·8) has roots equal to zero at infinity. Its implementation is readily seen to be expressed by

$$(M + \frac{6 - \sqrt{10}}{13}\Delta_t k)\,\underline{U}^{n,1} = \left(\frac{112 - 88\sqrt{10}}{169}\right)\left(\frac{2 - \sqrt{10}}{3} M + \Delta_t k\right)\underline{U}^{n+1}$$

$$+ \left(\frac{74 - 34\sqrt{10}}{169}\right)\left(- \frac{(53 + \sqrt{10}}{18} M + \Delta_t k\right)\underline{U}^{n}$$

$$(M + \frac{6 - \sqrt{10}}{13}\Delta_t K)\underline{U}^{n+2} = M\underline{U}^{n,1}$$

and requires the inversion of only one matrix. The scheme (6·4iii) can be manipulated to exhibit the same characteristic i.e. the polynomial $\mu_2(\tau)$ having a double root. Given

$$C = 105(4\sqrt{2}-3)/_{1127}$$

the scheme (6·4iii) yields a sixth order method with the property.

We conclude this chapter with the following remarks

(1) We conjecture that the maximum order of an A_0 —stable L.M.S.D. scheme which is stable at infinity is

$$q = m(k+l) - 1$$

Thus it is advisable to select $m > 1$ for the derivation of high order schemes.

(2) A clear advantage in increasing m rather than k results from the error constant decreasing more rapidly for m increasing than with k increasing, particularly if considered in conjunction with the rate of convergence of infinity.

(3) With respect to the system of equations (2·3), maximum order, A_0 —stable L.M.S.D. schemes, with $m > 1$, invariably require complex arithmetic for their implementation. Ease of implementation, as characterised by (6·7), may only be obtained by relaxing the stipulation of maximum order. However, once this relaxation is operative we can derive high order A_0 —stable L.M.S.D's that are simple to implement. We conjecture that schemes of order $q = mk$ can possess this property.

Note that the number of intermediate step evaluations at each time interval increases with m.

With regard to the above remarks we advance the merits of the classes of L.M.S.D. schemes where $m = k-l$, k or $k + l$, for $k \geq 2$. Such schemes incorporate a balance of high order, low error constant, and ease of implementation.

References

1. Ahlfors, L.V., "Complex Analysis", McGraw-Hill, New York, 2nd edition (1966).

2. Crouzeix, M., Ph.D. Thesis, Université Paris VI.

3. Cryer, C.W., "A New Class of Highly Stable Methods: A_0 - Stable Methods". BIT 13, 153-159 (1973).

4. Enright, W.H., "Second Derivative Multistep Methods for Stiff Ordinary Differential Equations". SLAM J. Numer. Anal., V.11, 321-331 (1974).

5. Genin, Y. "An Algebraic Approach to A - Stable Linear Multistep - Multiderivative Integration Formulas". BIT 14, 382-406 (1974)

6. Henrici, P. "Discrete Variable Methods in Ordinary Differential Equations". Wiley, New York-London-Sydney (1962).

7. Lambert, J.D. "Computational Methods in Ordinary Differential Equations". Wiley, London (1973).

8. Makinson, G.J. "Stable High Order Implicit Methods for the Numerical Solution of Systems of Differential Equations". The Computer Journal, Vol.II, No.3, 305-310 (1968).

9. Mihlin, S.G. "Mathematical Physics, An Advanced Course". Amsterdam, North-Holland (1970).

10. Nassif, N.R. "On the discretization of the Time Variable in Parabolic Partial Differential Equations". Finite Elements Symposium. Brunel University (1975).

11. Norsett, S.P. "One-step Methods of Hermite Type for Numerical Integration of Stiff Systems". BIT 14, 63-77 (1974).

12. Strang, G. and Fix, G.J. "An Analysis of the Finite Element Method". Englewood Cliffs, N.Y., Prentice-Hall (1973).

13. Thomée, V. "Some Convergence Results for Galerkin Methods for Parabolic Boundary Value Problems", to appear.

14. Zlámal, M. "Finite Element Multistep Discretizations of Parabolic Boundary Value Problems". Maths. Comp., V.29, No.130, 1-10 (1975).

15. Zlámal, M. "Finite Element Multistep Methods for Parabolic Equations", to appear.

16. Zlámal, M. "Finite Element Methods in Heat Conduction Problems". Finite Elements Symposium, Brunel University (1975).