NASA Contractor Report 178372

ICASE REPORT NO. 87-61

ICASE

GALERKIN/RUNGE-KUTTA DISCRETIZATIONS FOR PARABOLIC EQUATIONS WITH TIME DEPENDENT COEFFICIENTS

Stephen L. Keeling

Contract No. NAS1-18107
September 1987

INSTITUTE FOR COMPUTER APPLICATIONS IN SCIENCE AND ENGINEERING
NASA Langley Research Center, Hampton, Virginia 23665

Operated by the Universities Space Research Association

National Aeronautics and
Space Administration

Langley Research Center
Hampton, Virginia 23665

Galerkin/Runge-Kutta Discretizations for Parabolic Equations with Time Dependent Coefficients

Stephen L. Keeling*

Abstract. A new class of fully discrete Galerkin/Runge-Kutta methods is constructed and analyzed for linear parabolic initial boundary value problems with time dependent coefficients. Unlike any classical counterpart, this class offers arbitrarily high order convergence while significantly avoiding what has been called *order reduction*. In support of this claim, error estimates are proved, and computational results are presented. Additionally, since the time stepping equations involve coefficient matrices changing at each time step, a preconditioned iterative technique is used to solve the linear systems only approximately. Nevertheless, the resulting algorithm is shown to preserve the original convergence rate while using only the order of work required by the base scheme applied to a linear parabolic problem with time independent coefficients. Furthermore, it is noted that special Runge-Kutta methods allow computations to be performed in parallel so that the final execution time can be reduced to that of a low order method.

*Supported by the National Aeronautics and Space Administration under NASA Contract No. NAS1-18107 while in residence at the Institute for Computer Applications in Science and Engineering (ICASE), NASA Langley Research Center, Hampton, VA 23665-5225.

1 Introduction.

In this paper, linear parabolic initial boundary value problems with time dependent coefficients are considered. Specifically, the goal is to construct and analyze fully discrete approximations to the unique solution $u(\mathbf{x},t)$ of:

$$\left\{\begin{array}{rcll} \partial_t u &=& -L(t)u & \text{in } \Omega\times[0,t^*] \\ u &=& 0 & \text{on } \partial\Omega\times[0,t^*] \\ u(\mathbf{x},0) &=& u^0(\mathbf{x}) & \text{in } \Omega, \end{array}\right. \tag{1.1}$$

where:

$$L(t)u \equiv -\sum_{i,j=1}^{N} \partial_{x_i}(\ell_{ij}(\mathbf{x},t)\partial_{x_j}u) + \ell_0(\mathbf{x},t)u.$$

Here, Ω is a bounded domain in $\mathbf{R}^N$ with $\partial\Omega$ sufficiently smooth. Also, $\ell_{ij}(\mathbf{x},t)$ and $\ell_0(\mathbf{x},t)$ are assumed to be smooth. Further, on $\bar{\Omega}\times[0,t^*]$, the matrix $\{\ell_{ij}\}_{i,j=1}^N$ is symmetric and uniformly positive definite and ℓ_0 is nonegative. Also, the initial data u^0 is assumed to be both sufficiently smooth and compatible, and precise hypotheses on the required smoothness of the solution u are made as needed.

Now, for $1\le p\le\infty$ and integers $s\ge 0$, let $W^{s,p}\equiv W^{s,p}(\Omega)$ represent the well-known Sobolev spaces consisting of functions with (distributional) derivatives of order $\le s$ in $L_p\equiv L_p(\Omega)$. Also, let $\|\cdot\|_{W^{s,p}}$ denote the usual norm. Then, in particular, take $H^s\equiv W^{s,2}$ and denote its norm by $\|\cdot\|_s$. In addition, let H_0^1 be the subspace of H^1 consisting of functions vanishing on $\partial\Omega$ in the sense of trace. Further, let the inner product on L_2 be denoted by $(\cdot,\cdot)$, and the associated norm by $\|\cdot\|$. Next, given Hilbert spaces H, H_1, and H_2, $\mathcal{B}(H_1,H_2)$ represents the Hilbert space of bounded linear operators from H_1 into H_2, and $\mathcal{B}(H)\equiv\mathcal{B}(H,H)$. Also, for $t_2>t_1$, $C^l([t_1,t_2],H)$ denotes the Banach space of operators, continuously differentiable to order $l\ge 0$, from $[t_1,t_2]$ into H. See Adams [1] for more details.

Now for each $t\in[0,t^*]$, let $L(t)$ be extended to be L_2-selfadjoint with domain $H^2\cap H_0^1$. Also, assume that for $l\ge 0$ and $m\ge 0$ sufficiently large, $L(t)\in C^l([0,t^*],\mathcal{B}(H^{m+2}\cap H_0^1,H^m))$ so that:

$$\|L^{(l)}(t)v\|_m \le c(l,m)\|v\|_{m+2} \qquad \forall v\in H^{m+2}\cap H_0^1 \tag{1.2}$$

where $L^{(l)}(t)\equiv D_t^l L(t)$. Note that here and throughout this work, c (sometimes with a subscript) is used to denote a general positive constant, not necessarily the same in any two places. Moreover, if in a given (in)equality, there is a crucial element upon which c is meant to depend, such dependence is indicated explicitly as in (1.2). Next, introducing the L_2-selfadjoint solution operator $T(t)$ for which $T(t)L(t)=I$ on $H^2\cap H_0^1$ and $L(t)T(t)=I$ on L_2, assume that for $l\ge 0$ and $m\ge 0$ sufficiently large, $T(t)\in C^l([0,t^*],\mathcal{B}(H^m,H^{m+2}\cap H_0^1))$ so that:

$$\|T^{(l)}(t)v\|_{m+2} \le c(l,m)\|v\|_m \qquad \forall v\in H^m \tag{1.3}$$

where $T^{(l)}(t)\equiv D_t^l T(t)$. Finally, assume that for sufficiently large $l\ge 0$ and $m\ge 0$, the solution u and the data u^0 satisfy:

$$\sup_{0\le t\le t^*}\|\partial_t^l u(t)\|_m \le c(m,l)\|u^0\|_{m+2l}. \tag{1.4}$$

For details connected with (1.2)-(1.4), see Sammon [16].

A rough description of the results now follows. For this, let h and k denote spatial and temporal discretization parameters respectively, and suppose that U_h^n is a fully discrete approximation to $u(nk)$ obtained according to the base scheme (1.38) described below. Now, in section 3, the error committed by (1.38), is shown to satisfy:

$$\max_{0\le n\le n^*} \|U_h^n - u^n\| \le c(h^r + k^\mu + hk^{\mu-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_\alpha \tag{1.5}$$

where $\alpha \equiv \max(r+1, 2\mu+2)$, $\mu \equiv \min(\nu, q+1)$, q is the number of Runge-Kutta stages, and r and ν represent respectively, optimal exponents, characteristic of the Galerkin method and the Runge-Kutta method upon which the fully discrete scheme is based. Note that under the mild condition that either $r \le 2\mu$ or $h^2 \le ck$, the above error is $O(h^r + k^\mu)$. Further, it is explained below that the methods which are most easily implemented have the property that $\nu \le q+1$ which makes the estimate optimal. It is also worth mentioning that *inverse properties* (associated with the use of a quasi-uniform triangulation of Ω) are never explicitly assumed, and as explained after Lemma 3.8, the constructions of section 2 are required for this.

Next, section 4 deals with (1.46), a variant of the base scheme which incorporates a preconditioned iterative method (PIM) for the time stepping equations (1.40). Specifically, these equations are solved only approximately at the nth time level with say, l_n outer iterations (4.5), and j_n inner (PIM) iterations (4.11), and it is shown that the above convergence rate can be preserved while keeping $\frac{1}{n^*}\sum_{n=0}^{n^*-1} l_n j_n$ bounded independently of h and k. Hence, the order of work is asymptotically as that for a linear parabolic problem with time independent coefficients. Additionally, in [14], semilinear and quasilinear problems are considered, and the latter are treated with methods such as those reported here to obtain comparable results.

It should also be mentioned that the discovery of the methods described below was fortuitous. Note that there are extrapolation options other than (1.35) which are apparently more natural. For example, D^l could be replaced by T^l in (1.35) since the latter is consistent with (1.39). This idea is considered together with (1.39) in a computational section. However, under rather general conditions, (1.5) is proved and demonstrated computationally only for (1.38) and (1.46). In fact, it has been reported by many authors ([7], [13], [8]) that unless the solution to the differential equation satisfies very restrictive conditions, a classical fully discrete scheme fashioned after (1.23) cannot be expected to offer optimal order convergence. Furthermore, with regard to efficiency, (1.39) requires the formation of q new stiffness matrices at every time step. On the other hand, (1.38) and (1.46) require only the formation of a single such matrix and, at the expense of at most $100q^{-1}\%$ more storage, the recall of $\mu - 1$ of its counterparts formed at previous time steps.

In [7], Crouzeix analyzes (1.39), and with Butcher's conditions $C(p-1)$ and $B(\nu)$, [5] he establishes the L_2 estimate:

$$\max_{0\le n\le n^*} \|U_h^n - u^n\| = O(h^r + k^{\min(p,\nu)}).$$

Since $O(h^r + k^\nu)$ has not generally been observed experimentally, this suboptimal phenomenon has been called *order reduction*. Note further that this L_2 estimate depends upon the assumption that the stages are computed exactly. On the other hand, in [13], Karakashian considers approximating the stages with a PIM, and proves that the above estimate holds while the order of work is kept optimal. Also, he constructs collocation type implicit Runge-Kutta methods (IRKM's) for which $p = \nu = q+1$. Nevertheless, such methods have limited stability for $q \ge 3$. In fact, there is a general trade-off among IRKM's in the sense that the more stable methods suffer more from order

reduction while those which do not suffer so, are not as stable. However, when (1.39) is modified as in (1.38), it is possible to achieve high order even for very stable methods. For example, in section 4, an algebraically stable IRKM is used for a problem of the form (1.1), and optimal order convergence is obtained with (1.46) but not with a counterpart based on (1.39).

Douglas, Dupont and Ewing [10] have analyzed Galerkin/Crank-Nicholson fully discrete approximations for a class of quasilinear parabolic problems, proving an optimal L_2 estimate for a method which is second order in time. Also, this rate was shown to be preserved by an algorithm in which the time stepping equations are solved only approximately with an optimal order of work. Then studying (1.1), Bramble and Sammon [3] have obtained similar results for some Galerkin/Obrechkoff fully discrete approximations, proving optimal L_2 estimates for methods up to fourth order in time. Finally, note that in [9], Dougalis and Karakashian analyze Galerkin/Runge-Kutta fully discrete approximations for the Korteweg-De Vries equation. In fact, they prove optimal L_2 estimates for some *modified* IRKM's which are up to fourth order. Hence, the spirit of their work is similar to that of the present study.

In the remainder of this section, there is a presentation of material relevant to the spatial and temporal discretizations considered here, which concludes with a precise definition of the schemes for which the above claims are made.

Spatial Discretizations

To make the following machinery more definite, consider the *Ordinary Galerkin Method* for the spatial approximation of the solution to (1.1). Let $D(t)(\cdot,\cdot)$ be a bilinear form defined by:

$$D(t)(v,w) \equiv \sum_{i,j=1}^{N} (\ell_{ij}(t)\partial_{x_i} v, \partial_{x_j} w) + (\ell_0(t)v, w) \qquad v, w \in H_0^1.$$

Next, let S_h represent an approximation subspace consisting of continuous, piecewise polynomials of degree $\leq r-1$, vanishing on $\partial\Omega$. Then, take $T_h(t)$: $L_2 \to S_h$ to be an approximation to the solution operator $T(t)$ defined by:

$$D(t)(T_h(t)w, \chi) \equiv (w, \chi) \qquad \forall w \in L_2, \quad \forall \chi \in S_h.$$

For more examples of Galerkin methods satisfying the assumptions enumerated below, see Bramble, Schatz, Thomée, and Wahlbin [4], and Sammon [16], [17].

Depending on the Galerkin method used for the spatial approximation, let H_E be a linear space equipped with a norm $\|\cdot\|_E$ and satisfying the following properties. Suppose $H^2 \cap H_0^1 \subset H_E$ and that:

$$\|v\|_1 \leq c\|v\|_E \qquad \forall v \in H_E, \tag{1.6}$$

$$\|v\|_E \leq c\|v\|_2 \qquad \forall v \in H^2. \tag{1.7}$$

For example, for the method described above, take $H_E = H_0^1$. Now let $\{S_h\}_{0<h<1}$ be a family of finite-dimensional subspaces of H_E satisfying the following for some integer $r \geq 2$:

$$\inf_{\chi \in S_h} \{\|v-\chi\| + h\|v-\chi\|_E\} \leq ch^s\|v\|_s \qquad \forall v \in H^s \cap H_0^1, \quad 2 \leq s \leq r. \tag{1.8}$$

Then suppose that for each $t \in [0, t^*]$, a corresponding family of operators $\{T_h(t)\}_{0<h<1}$ is given satisfying:

i. $T_h(t) : L_2 \to S_h$ is selfadjoint, positive semidefinite on L_2, and positive definite on S_h.

ii. For $0 < h < 1$, $T_h(t) \in C^l([0,t^*], \mathcal{B}(L_2, S_h))$ for $l \geq 0$ sufficiently large.

iii. For $2 \leq s \leq r$, $0 \leq t \leq t^*$, and $l \geq 0$ as large as required in the sequel:

$$\|[T^{(l)}(t) - T_h^{(l)}(t)]v\| + h\|[T^{(l)}(t) - T_h^{(l)}(t)]v\|_E \leq ch^s\|v\|_{s-2} \qquad \forall v \in H^{s-2}. \tag{1.9}$$

Hence, the restriction of $T_h(t)$ to S_h is invertible and its inverse is henceforth denoted by $L_h(t)$. Since $L_h(t)$ is also positive definite and selfadjoint on S_h, both $L_h(t)$ and $T_h(t)$ have square roots but it is also assumed that:

$$c\|T_h^{\frac{1}{2}}(t)w\| \leq c\|T_h^{\frac{1}{2}}(t)w\|_E \leq \|w\| \leq \|w\|_E \qquad \forall w \in L_2, \tag{1.10}$$

$$\|\chi\| \leq \|\chi\|_E \leq c\|L_h^{\frac{1}{2}}(t)\chi\| \leq c\|L_h^{\frac{1}{2}}(t)\chi\|_E \qquad \forall \chi \in S_h. \tag{1.11}$$

Also $L_h(t) \in C^l([0,t^*], \mathcal{B}(S_h))$ for $l \geq 0$ sufficiently large and in fact, Bales [2] has proved that for $0 \leq s, t \leq t^*$, and $l \geq 0$:

$$\|L_h^{\frac{1}{2}}(t)T_h^{(l)}(s)L_h^{\frac{1}{2}}(t)\chi\| \leq c(l)\|\chi\| \qquad \forall \chi \in S_h, \tag{1.12}$$

$$\|T_h^{\frac{1}{2}}(t)L_h^{(l)}(s)T_h^{\frac{1}{2}}(t)v\| \leq c(l)\|v\| \qquad \forall v \in L_2. \tag{1.13}$$

Then using the selfadjointness of these operators, the following are straightforward consequences of (1.12) and (1.13). For $0 \leq s, t \leq t^*$:

$$\|T_h^{\frac{1}{2}}(s)L_h^{\frac{1}{2}}(t)\chi\| \leq c\|\chi\| \qquad \forall \chi \in S_h, \tag{1.14}$$

$$\|L_h^{\frac{1}{2}}(s)T_h^{\frac{1}{2}}(t)v\| \leq c\|v\| \qquad \forall v \in L_2, \tag{1.15}$$

$$|(L_h^{(l)}(s)\chi, \chi)| \leq c(l)(L_h(t)\chi, \chi) \qquad \forall \chi \in S_h. \tag{1.16}$$

In addition to (1.16), assume that for $0 \leq t \leq t^*$ and $l \geq 0$:

$$|(L_h^{(l)}(t)\chi, \phi)| \leq c(l)\|\chi\|_E\|\phi\|_E \qquad \forall \chi, \phi \in S_h. \tag{1.17}$$

Next, defining the *elliptic projection* operator as $P_E(t) \equiv T_h(t)L(t)$, it follows from (1.9) and (1.2) that for $0 \leq t \leq t^*$:

$$\|[I - P_E(t)]v\| + h\|[I - P_E(t)]v\|_E \leq ch^s\|v\|_s \qquad \forall v \in H^s \cap H_0^1, \quad 2 \leq s \leq r. \tag{1.18}$$

In fact, with $\omega(t) \equiv P_E(t)u(t)$ and $\eta(t) \equiv u(t) - \omega(t)$, (1.2), (1.4), and (1.9) can be used [3] to show that:

$$\sup_{0 \leq t \leq t^*} \{\|\eta^{(l)}(t)\| + h\|\eta^{(l)}(t)\|_E\} \leq ch^s\|u^0\|_{s+2l} \qquad 2 \leq s \leq r, \quad 0 \leq 2l \leq \alpha - s. \tag{1.19}$$

Finally, it can be shown that $P_0 \equiv L_h(t)T_h(t)$ is for every $t \in [0,t^*]$, the orthogonal projection of L_2 onto S_h and that $T_h(t) = T_h(t)P_0$. Then, since $I - P_0$ is majorized by $I - P_E$ in L_2, it follows from (1.18) that:

$$\|(I - P_0)v\| \leq ch^s\|v\|_s \qquad \forall v \in H^s \cap H_0^1, \quad 2 \leq s \leq r. \tag{1.20}$$

Now, (1.1) has the following semidiscrete formulation. Find $u_h : [0, t^*] \to S_h$ satisfying:

$$\text{(1.21)} \qquad \begin{cases} \partial_t u_h &= -L_h(t) u_h \\ u_h(0) &= u_h^0. \end{cases}$$

where $u_h^0 \in S_h$ is a suitable approximation to u^0. In [17], Sammon analyzes approximations of the form (1.21), and with assumptions comparable to those described above, he proves an optimal L_2 estimate:

$$\sup_{0 \le t \le t^*} \|u(t) - u_h(t)\| \le c h^r \|u^0\|_r.$$

In the present paper, semidiscrete approximations are not analyzed. Instead, (1.21) serves only as a source of inspiration for fully discrete approximations, and u_h is not even mentioned in forthcoming proofs.

Temporal Discretizations

For the temporal approximation of the solution to (1.21), *Implicit Runge-Kutta Methods* (IRKM's) are now introduced. Given an integer $q \ge 1$, a *q-stage* IRKM is characterized by a set of constants $\{a_{ij}\}_{i,j=1}^q$, $\{b_j\}_{j=1}^q$, and $\{\tau_i\}_{i=1}^q$, and it is convenient to make the following definitions:

$$A \equiv \{a_{ij}\}_{i,j=1}^q, \qquad b^T \equiv \langle b_1, b_2, \ldots, b_q \rangle, \qquad B \equiv \operatorname*{diag}_{1 \le i \le q} \{b_i\},$$

$$M \equiv BA + A^T B - bb^T, \qquad \mathrm{T} \equiv \operatorname*{diag}_{1 \le i \le q} \{\tau_i\}, \qquad e^T \equiv \langle 1, 1, \ldots, 1 \rangle.$$

For the IRKM formulation used in this work, choose arbitrarily, $t_0 \in \mathbf{R}$, $\mathbf{y}_0 \in \mathbf{R}^n$, $\mathbf{F} : \mathbf{R}^{n+1} \to \mathbf{R}^n$ sufficiently smooth, and $k > 0$ sufficiently small, so that for $t_0 \le t \le t_0 + k$, smooth functions $\mathbf{y}, \hat{\mathbf{y}} : \mathbf{R} \to \mathbf{R}^n$ are well-defined by:

$$\text{(1.22)} \qquad \begin{cases} D_t \mathbf{y}(t) &= \mathbf{F}(t, \mathbf{y}(t)) \\ \mathbf{y}(t_0) &= \mathbf{y}_0, \end{cases}$$

$$\text{(1.23)} \qquad \begin{cases} \mathbf{y}^i(t) &= \mathbf{y}_0 + (t - t_0) \sum_{j=1}^q a_{ij} \mathbf{F}(t_0 + \tau_j (t - t_0), \mathbf{y}^j(t)), \quad 1 \le j \le q \\ \hat{\mathbf{y}}(t) &= \mathbf{y}_0 + (t - t_0) \sum_{i=1}^q b_i \mathbf{F}(t_0 + \tau_i (t - t_0), \mathbf{y}^i(t)). \end{cases}$$

The method is described as *explicit* if $a_{ij} = 0$, $i \le j$ and *implicit* if for any i, $a_{ii} \ne 0$. Also, it is said to have *order* ν if for every $\mathbf{y}$ and $\hat{\mathbf{y}}$ defined as above, $D_t^l \mathbf{y}(t_0) = D_t^l \hat{\mathbf{y}}(t_0)$, $0 \le l \le \nu$. Butcher [5] has developed simple conditions for the above parameters which guarantee a given order; however, only the following is explicitly required in this work:

$$\text{(1.24)} \qquad l! b^T A^{l-1} e = 1 \qquad 1 \le l \le \nu.$$

To see the roots of condition (1.24), let (1.22) have $n = 1$, $t_0 = 0$, $y_0 = 1$, and $F(y) = -y$, so that $y(t) = e^{-t}$. Then, from (1.23), $\hat{y}(t) = r(t)$ where $r(z)$ is a rational approximation to the exponential e^{-z} given by:

$$\text{(1.25)} \qquad r(z) \equiv 1 - z b^T (I + zA)^{-1} e.$$

Expanding this expression shows that $r(z)$ is a νth order approximation to the exponential if and only if (1.24) holds. Next, with regard to stability, an IRKM is said to be *A_0-stable* if:

$$|r(z)| \leq 1 \qquad \forall z \geq 0, \tag{1.26}$$

and *strongly A_0-stable* if:

$$\sup_{z \geq z_0} |r(z)| < 1 \qquad \forall z_0 > 0. \tag{1.27}$$

Also, a method is called *algebraically stable* if M and B are positive semidefinite. However, if an algebraically stable method is *irreducible* (not equivalent to a fewer stage method) then:

$$B \text{ is positive definite, and } M \text{ is positive semidefinite.} \tag{1.28}$$

One other notion of stability which is useful here is that of dissipativity:

$$-1 < -1 + \delta \leq r(z) \leq 1 \qquad \forall z \geq 0. \tag{1.29}$$

A_0-stability is required of all IRKM's considered in this work. However, in order for the approximations to decay with respect to the time step, strong A_0-stability must hold. In fact, to guarantee decay, both (1.27) and (1.28) are assumed. Then in section 4, the iterative scheme (1.46) described below requires at least (1.29) in addition to:

$$r(z) \leq 1 - c\frac{z}{(1+z)^3} \qquad \forall z \geq 0. \tag{1.30}$$

This growth condition is extremely mild and this author is unaware of any popular IRKM which fails to satisfy it. Also, requiring (1.29) and (1.30) improves on a related result of Karakashian [13] in which (1.27) is used. Next, note that the spectrum of A, $\sigma(A)$ is related to the poles of $r(z)$ and in addition to the above, it is assumed throughout this paper that:

$$\sigma(A) \subset \{x \in \mathbf{R} : x > 0\}. \tag{1.31}$$

Returning to the temporal discretization of (1.21), let a q-stage IRKM of order $\nu \geq 1$ be given. Assume also that there exists a $q \times q$ matrix D satisfying:

$$D[e; Ae; \ldots; A^{q-1}e] = [Ae; 2A^2e; \ldots; qA^qe]. \tag{1.32}$$

Again, this author is unaware of any well-known IRKM for which such a D fails to exist. In fact, the so-called collocation type methods are those for which $D = \mathrm{T}$. Now with $\mu \equiv \min(\nu, q+1)$, it follows from (1.32) and (1.24) that:

$$lAD^{l-1}e = D^le \qquad 1 \leq l \leq \mu - 1, \tag{1.33}$$

$$lb^TD^{l-1}e = 1 \qquad 1 \leq l \leq \mu. \tag{1.34}$$

Next, for $0 \leq n \leq n^* - 1$, $n^*k \equiv t^*$, let the real values $\{\delta_m^n\}_{m=0}^{\mu-1}$ be chosen distinctly, so that the $q \times q$ matrices $\{\Gamma_m^n\}_{m=0}^{\mu-1}$ are well-defined by:

$$\sum_{m=0}^{\mu-1} \Gamma_m^n(\delta_m^n)^l = D^l \qquad 0 \leq l \leq \mu - 1 \tag{1.35}$$

as the computation of their components involves the inversion of the $\mu \times \mu$ Vandermonde matrix $\{(\delta_m^n)^l\}_{m,l=0}^{\mu-1}$. In addition, assume that these parameters are bounded independently of n:

$$\max_{0\le m\le \mu-1}\{|\delta_m^n| + \max_{1\le i,j\le q}|(\Gamma_m^n)_{ij}|\} \le c. \tag{1.36}$$

Actually, it is clear below that the natural and computationally advantageous choice for (1.35) is:

$$\delta_m^n = \begin{cases} m/\mu, & 0 \le n \le \mu - 2 \\ -m, & \mu - 1 \le n \le n^* - 1. \end{cases}$$

In any case, define $t^n \equiv nk$ and $\tau_m^n \equiv t^n + \delta_m^n k$, and for $0 \le n \le n^* - 1$, $0 \le t \le t^*$, and $0 \le s \le k$, let the following be defined on $\mathbf{S}_h \equiv [S_h]^q$:

$$\mathcal{L}_h(t) \equiv \operatorname*{diag}_{q\times q}\{L_h(t)\}, \qquad \mathcal{L}_h^n \equiv \mathcal{L}_h(t^n), \qquad \bar{\mathcal{L}}_h^n(s) \equiv \sum_{m=0}^{\mu-1}\Gamma_m^n \mathcal{L}_h(t^n + \delta_m^n s), \qquad \bar{\mathcal{L}}_h^n \equiv \bar{\mathcal{L}}_h^n(k).$$

Now with:

$$U_h^0 = [I + kL_h^0]^{-1}P_0[I + kL^0]u^0 \tag{1.37}$$

suppose that for $0 \le n \le n^* - 1$, the approximation $U_h^n \in S_h$ is given, where $U_h^n \approx u^n$ and $u^n \equiv u(\mathbf{x}, t^n)$. Then, let $U_h^{n+1} \approx u^{n+1}$ be given by what is henceforth called the **base scheme**:

$$\begin{cases} \bar{U}_h^n &= eU_h^n - kA\bar{\mathcal{L}}_h^n\bar{U}_h^n \\ U_h^{n+1} &= (I - b^TA^{-1}e)U_h^n + b^TA^{-1}\bar{U}_h^n \end{cases} \tag{1.38}$$

where $\bar{U}_h^n \in \mathbf{S}_h$ is well-defined provided $[I + kA\bar{\mathcal{L}}_h^n]$ is invertible. Here, $A\bar{\mathcal{L}}_h^n$ for example, is understood in the sense of composition of operators defined on $\mathbf{S}_h$. Note that if the temporal discretization of (1.21) were accomplished as prescribed by (1.22) and (1.23), the following would result:

$$\begin{cases} \bar{U}_h^n &= eU_h^n - kA\tilde{\mathcal{L}}_h^n\bar{U}_h^n \\ U_h^{n+1} &= (I - b^TA^{-1}e)U_h^n + b^TA^{-1}\bar{U}_h^n \end{cases} \qquad \text{where} \quad \tilde{\mathcal{L}}_h^n \equiv \operatorname*{diag}_{1\le i\le q}\{L_h(t^n + k\tau_i)\}. \tag{1.39}$$

However, as discussed in the beginning of the Introduction, (1.38) is designed to improve upon (1.39) with the indicated modification.

Now, with regard to iterative approximations, note that an efficient method is needed for solving the time stepping equations:

$$[I + kA\bar{\mathcal{L}}_h^n]\bar{U}_h^n = eU_h^n. \tag{1.40}$$

According to (1.31), A can be transformed as follows:

$$SAS^{-1} = \Lambda \equiv \operatorname*{diag}_{1\le i\le q}\{\lambda_i\} + \operatorname*{subdiag}_{2\le i\le q}\{\theta_i\}; \qquad \lambda_i > 0, \quad 1 \le i \le q; \quad \theta_i = 0 \text{ or } 1, \quad 2 \le i \le q.$$

Then $\bar{V}_{l_n}^n \approx \bar{U}_h^n$ can be obtained by the (outer) iterations:

$$[I + k\Lambda\mathcal{L}_h^n](S\bar{V}_l^n) = \{SeU^n + kSA(\mathcal{L}_h^n - \bar{\mathcal{L}}_h^n)\bar{V}_{l-1}^n\} \equiv R_l^n \qquad 1 \le l \le l_n \tag{1.41}$$

where:

$$(1.42)\quad \tilde{V}_0^n \equiv \sum_{m=n-1-\mu_n}^{n-1} (-1)^{n-m-1} \binom{\mu_n+1}{n-m} \tilde{U}_h^m \qquad 1 \le n \le n^*-1, \qquad \tilde{V}_0^0 \equiv eU_h^0,$$

$$(1.43)\quad l_n \equiv \begin{cases} \mu, & n=0 \\ \mu+1-n, & 1 \le n \le \mu \\ 1, & \mu+1 \le n \le n^*-1, \end{cases} \qquad \mu_n \equiv \begin{cases} 0, & n=0 \\ n-1, & 1 \le n \le \mu \\ \mu, & \mu+1 \le n \le n^*-1, \end{cases}$$

and (1.41) is started with $\bar{V}_0^n \equiv \tilde{V}_0^n$ provided $\{\tilde{U}_h^m\}_{m=n-1-\mu_n}^{n-1}$ are computed as indicated below. Now consider the simple but important observation that if:

$$(1.44)\qquad \lambda_i \neq \lambda_j, \quad i \neq j \qquad \text{and} \qquad \theta_i = 0, \quad 2 \le i \le q,$$

then the block system above decouples into the following equations which can be solved in parallel:

$$[I + k\lambda_i L_h^n](S\bar{V}_l^n)_i = (R_l^n)_i \qquad 1 \le i \le q.$$

Then, to avoid having to factor new coefficient matrices at every time step, a preconditioned iterative method is used to approximate $\bar{V}_l^n$ with (inner) iterates, say $\{\tilde{V}_{l,j}^n\}_{0 \le j \le j_n}$. Further, it is shown that there exist integers $\{j_n\}_{n=0}^{n^*-1}$ such that:

$$(1.45)\qquad \frac{1}{n^*} \sum_{n=0}^{n^*-1} l_n j_n \le c$$

while the convergence order (1.5) is preserved for what is henceforth called the **iterative scheme**:

$$(1.46)\qquad \begin{cases} \tilde{U}_h^n &= \tilde{V}_{l_n,j_n}^n \\ U_h^{n+1} &= (I - b^T A^{-1} e) U_h^n + b^T A^{-1} \tilde{U}_h^n. \end{cases}$$

Finally, let the initial approximation for this scheme be given by (1.37) also.

2 The Product Space Operators.

In this section, the machinery elaborated between (1.2) and (1.20) is generalized to analogous operators defined on products of spaces on which their precursors are defined. Also, various technical lemmas are proved for later use. Now, in addition to $\mathbf{S}_h$, define the product spaces $\mathbf{L}_2 \equiv [L_2]^q$, $\mathbf{H}_0^1 \equiv [H_0^1]^q$, $\mathbf{H}_E \equiv [H_E]^q$, and $\mathbf{H}^m \equiv [H^m]^q$. Also, denote the natural product space norms by:

$$\|\mathbf{\Phi}\|_E \equiv \{\sum_{i=1}^q \|\phi_i\|_E^2\}^{\frac{1}{2}}, \qquad \|\mathbf{\Phi}\|_m \equiv \{\sum_{i=1}^q \|\phi_i\|_m^2\}^{\frac{1}{2}}, \qquad \text{and} \qquad \|\mathbf{\Phi}\| \equiv \|\mathbf{\Phi}\|_0.$$

Then, for $0 \le n \le n^*-1$, $0 \le t \le t^*$ and $0 \le s \le k$, let the following be defined on $\mathbf{H}^2 \cap \mathbf{H}_0^1$:

$$\mathcal{L}(t) \equiv \underset{q \times q}{\mathrm{diag}}\{L(t)\}, \qquad \mathcal{L}^n \equiv \mathcal{L}(t^n), \qquad \bar{\mathcal{L}}^n(s) \equiv \sum_{m=0}^{\mu-1} \Gamma_m^n \mathcal{L}(t^n + \delta_m^n s), \qquad \bar{\mathcal{L}}^n \equiv \bar{\mathcal{L}}^n(k).$$

The first step is to construct, for $0 \le n \le n^* - 1$ and $0 \le s \le k$, operators $\bar{\mathcal{T}}^n(s)$ ($\bar{\mathcal{T}}^n \equiv \bar{\mathcal{T}}^n(k)$) satisfying:

$$\text{(2.1)} \qquad \begin{cases} \bar{\mathcal{L}}^n(s)\bar{\mathcal{T}}^n(s) = I & \text{on } \mathbf{L}_2 \\ \bar{\mathcal{T}}^n(s)\bar{\mathcal{L}}^n(s) = I & \text{on } \mathbf{H}^2 \cap \mathbf{H}_0^1. \end{cases}$$

Note that with the following defined on $\mathbf{L}_2$ for $0 \le t \le t^*$ and $0 \le n \le n^* - 1$:

$$\mathcal{T}(t) \equiv \underset{q \times q}{\text{diag}}\{T(t)\}, \qquad \mathcal{T}^n \equiv \mathcal{T}(t^n)$$

$\bar{\mathcal{T}}^n(s)$ cannot be taken as a combination of such operators unless D is diagonal.

Lemma 2.1 *For $0 \le n \le n^* - 1$, $\bar{\mathcal{L}}^n(s) \in C^l([0,k], \mathcal{B}(\mathbf{H}^{m+2} \cap \mathbf{H}_0^1, \mathbf{H}^m))$ where $l, m \ge 0$ are as in (1.2). Also the following hold:*

$$\text{(2.2)} \qquad \|\partial_s^l \bar{\mathcal{L}}^n(s)\mathbf{v}\|_m \le c(l,m)\|\mathbf{v}\|_{m+2}$$

$$\text{(2.3)} \qquad \|[\bar{\mathcal{L}}^n(s) - \mathcal{L}^n]\mathbf{v}\|_m \le c(m)k\|\mathbf{v}\|_{m+2}$$

$$\forall v \in \mathbf{H}^{m+2} \cap \mathbf{H}_0^1, \quad 0 \le s \le k, \quad \textit{and} \quad 0 \le n \le n^* - 1.$$

Proof: The crucial observation is that by (1.35) with $l = 0$:

$$\bar{\mathcal{L}}^n(s) - \mathcal{L}^n = \sum_{i=0}^{\mu-1} \Gamma_i^n[\mathcal{L}(t^n + \delta_i^n s) - \mathcal{L}^n] = \sum_{i=0}^{\mu-1} \Gamma_i^n \int_{t^n}^{t^n + \delta_i^n s} \mathcal{L}^{(1)}(t)dt$$

and (2.3) follows with (1.36) and (1.2). Also, (2.2) follows using (1.36) and (1.2). ∎

Theorem 2.1 *Let $m, l \ge 0$ be as in (1.3). Then for k small enough, and $0 \le n \le n^* - 1$, there exist operators $\bar{\mathcal{T}}^n(s) \in C^l([0,k], \mathcal{B}(\mathbf{H}^m, \mathbf{H}^{m+2} \cap \mathbf{H}_0^1))$ satisfying (2.1) and:*

$$\text{(2.4)} \qquad \|\partial_s^l \bar{\mathcal{T}}^n(s)\mathbf{v}\|_{m+2} \le c(l,m)\|\mathbf{v}\|_m \qquad \forall \mathbf{v} \in \mathbf{H}^m.$$

Proof: Let $\mathbf{v} \in \mathbf{H}^m$ be chosen arbitrarily, and define $\mathcal{F}$: $\mathbf{H}^{m+2} \cap \mathbf{H}_0^1 \to \mathbf{H}^{m+2} \cap \mathbf{H}_0^1$ by:

$$\mathcal{F}\mathbf{u} \equiv \mathcal{T}^n\{\mathbf{v} + [\mathcal{L}^n - \bar{\mathcal{L}}^n(s)]\mathbf{u}\}.$$

By (1.3) and (2.3), with k small enough, there is an $\varepsilon \in (0,1)$ such that $\forall \mathbf{u}_1, \mathbf{u}_2 \in \mathbf{H}^{m+2} \cap \mathbf{H}_0^1$:

$$\|\mathcal{F}(\mathbf{u}_2 - \mathbf{u}_1)\|_{m+2} \le c(m)\|[\mathcal{L}^n - \bar{\mathcal{L}}^n(s)](\mathbf{u}_2 - \mathbf{u}_1)\|_m \le \varepsilon\|\mathbf{u}_2 - \mathbf{u}_1\|_{m+2}.$$

Hence $\mathcal{F}$ is a global contraction and has a unique fixed point. Thus, for every $\mathbf{v} \in \mathbf{H}^m$, there exists a unique element $\bar{\mathcal{T}}^n(s)\mathbf{v} \in \mathbf{H}^{m+2} \cap \mathbf{H}_0^1$ such that $\bar{\mathcal{L}}^n(s)\bar{\mathcal{T}}^n(s)\mathbf{v} = \mathbf{v}$. In particular, the first part of (2.1) holds. Also, if $\mathbf{u} \in \mathbf{H}^{m+2} \cap \mathbf{H}_0^1$ and $\mathbf{w} \equiv \bar{\mathcal{T}}^n(s)\bar{\mathcal{L}}^n(s)\mathbf{u} - \mathbf{u}$, then by the uniqueness, $\bar{\mathcal{L}}^n(s)\mathbf{w} = 0$ implies that $\mathbf{w} = 0$. So, (2.1) follows. Next, the following estimate is well-known:

$$\|\bar{\mathcal{T}}^n(s)\mathbf{v} - \mathcal{T}^n\mathbf{v}\|_{m+2} \le (1-\varepsilon)^{-1}\|\mathcal{F}\mathcal{T}^n\mathbf{v} - \mathcal{T}^n\mathbf{v}\|_{m+2}.$$

By (1.3) and (2.3):

$$\|\mathcal{F}\mathcal{T}^n\mathbf{v} - \mathcal{T}^n\mathbf{v}\|_{m+2} \le c(m)\|[\mathcal{L}^n - \bar{\mathcal{L}}^n(s)]\mathcal{T}^n\mathbf{v}\|_m \le c(m)k\|\mathcal{T}^n\mathbf{v}\|_{m+2}.$$

Then, for the case $l = 0$, (2.4) follows from the last two inequalities and (1.3). Now, by estimating difference quotients, it can be shown in a straightforward way that $\partial_s \bar{\mathcal{T}}^n(s) = -\bar{\mathcal{T}}^n(s)\partial_s \bar{\mathcal{L}}^n(s)\bar{\mathcal{T}}^n(s)$ for $0 \le s \le k$ and the smoothness of $\bar{\mathcal{T}}^n(s)$ follows inductively with Lemma 2.1. Finally, after differentiating the first line of (2.1):

$$\partial_s^l \bar{\mathcal{T}}^n(s) = -\sum_{i=0}^{l-1} \begin{pmatrix} l \\ i \end{pmatrix} \bar{\mathcal{T}}^n(s)\partial_s^{l-i}\bar{\mathcal{L}}^n(s)\partial_s^i \bar{\mathcal{T}}^n(s)$$

and (2.4) follows inductively using (2.2). ∎

Now with trivial modifications of the above, the following is obtained for the adjoints.

Lemma 2.2 *For $0 \le n \le n^* - 1$, $\bar{\mathcal{L}}^n(s)^* \in C^l([0,k], \mathcal{B}(\mathbf{H}^{m+2} \cap \mathbf{H}_0^1, \mathbf{H}^m))$ where $l, m \ge 0$ are as in (1.2), and:*

$$\|\partial_s^l \bar{\mathcal{L}}^n(s)^* \mathbf{v}\|_m \le c(l,m)\|\mathbf{v}\|_{m+2} \qquad \forall \mathbf{v} \in \mathbf{H}^{m+2} \cap \mathbf{H}_0^1. \tag{2.5}$$

Furthermore, with $m, l \ge 0$ as in (1.3), there exist operators $\bar{\mathcal{T}}^n(s)^* \in C^l([0,k], \mathcal{B}(\mathbf{H}^m, \mathbf{H}^{m+2} \cap \mathbf{H}_0^1))$ satisfying:

$$\|\partial_s^l \bar{\mathcal{T}}^n(s)^* \mathbf{v}\|_{m+2} \le c(l,m)\|\mathbf{v}\|_m \qquad \forall \mathbf{v} \in \mathbf{H}^m. \tag{2.6}$$

and $\bar{\mathcal{L}}^n(s)^* \bar{\mathcal{T}}^n(s)^* = I$ on $\mathbf{L}_2$, $\bar{\mathcal{T}}^n(s)^* \bar{\mathcal{L}}^n(s)^* = I$ on $\mathbf{H}^2 \cap \mathbf{H}_0^1$. ∎

The next step is to construct for $0 \le n \le n^* - 1$ and $0 \le s \le k$, operators $\bar{\mathcal{T}}_h^n(s)$ ($\bar{\mathcal{T}}_h^n \equiv \bar{\mathcal{T}}_h^n(k)$) satisfying:

$$\begin{cases} \bar{\mathcal{L}}_h^n(s)\bar{\mathcal{T}}_h^n(s) = \mathcal{P}_0 & \text{on } \mathbf{L}_2 \\ \bar{\mathcal{T}}_h^n(s)\bar{\mathcal{L}}_h^n(s) = I & \text{on } \mathbf{S}_h \end{cases} \tag{2.7}$$

where $\mathcal{P}_0 \equiv \underset{q \times q}{\operatorname{diag}}\{P_0\}$. Note that with the following defined on $\mathbf{L}_2$ for $0 \le t \le t^*$ and $0 \le n \le n^* - 1$:

$$\mathcal{T}_h(t) \equiv \underset{q \times q}{\operatorname{diag}}\{T_h(t)\}, \qquad \mathcal{T}_h^n \equiv \mathcal{T}_h(t^n)$$

$\bar{\mathcal{T}}_h^n(s)$ cannot be taken as a combination of such operators unless D is diagonal.

Now let $\{D_h(t)(\cdot,\cdot)\}_{0 \le t \le t^*}$ be a family of bilinear forms defined on $H_E \times H_E$ so that:

$$D_h(t)(\chi, \phi) = (L_h(t)\chi, \phi) \qquad \forall \chi, \phi \in S_h. \tag{2.8}$$

More specifically, with $D_h^{(l)}(t)(\cdot,\cdot) \equiv D_t^l D_h(\cdot,\cdot)$, assume that for $0 \le t \le t^*$, $l \ge 0$, and $2 \le m \le r$:

$$|D_h^{(l)}(t)(v,w) - (L^{(l)}(t)v, w)| \le c(l)h^{m-1}\|v\|_m \|w - u\|_E, \tag{2.9}$$
$$\forall v \in H^m \cap H_0^1, \quad \forall w \in H^2 \cap H_0^1 + S_h, \quad \forall u \in H^2 \cap H_0^1,$$

$$|D_h^{(l)}(t)(w,v)| \le c(l)\|w\|_E \|v\|_E \qquad \forall w, v \in H_E, \tag{2.10}$$

$$c\|\chi\|_E^2 \le D_h(t)(\chi, \chi) \qquad \forall \chi \in S_h. \tag{2.11}$$

For example, these assumptions are readily verified for the Ordinary Galerkin Method mentioned in the Introduction. For additional examples, see Sammon [16], [17]. Next, for $0 \le t \le t^*$, $0 \le n \le n^* - 1$ and $0 \le s \le k$, let the following be defined on $\mathbf{H}_E \times \mathbf{H}_E$:

$$\mathcal{D}_h(t)(\mathbf{w}, \mathbf{v}) \equiv \sum_{i=1}^{q} D_h(t)(w_i, v_i), \qquad \bar{\mathcal{D}}_h^n(s)(\mathbf{w}, \mathbf{v}) \equiv \sum_{m=0}^{\mu-1} \mathcal{D}_h(t^n + \delta_m^n s)(\Gamma_m^n \mathbf{w}, \mathbf{v}).$$

Lemma 2.3 *For $k > 0$ small enough, $l \geq 0$, $0 \leq n \leq n^* - 1$ and $0 \leq s \leq k$:*

$$|\partial_s^l \bar{\mathcal{D}}_h^n(s)(\mathbf{w},\mathbf{v})| \leq c(l)\|\mathbf{w}\|_E\|\mathbf{v}\|_E \qquad \forall \mathbf{w},\mathbf{v} \in \mathbf{H}_E \tag{2.12}$$

$$c\|\mathbf{X}\|_E^2 \leq \bar{\mathcal{D}}_h^n(s)(\mathbf{X},\mathbf{X}) \qquad \forall \mathbf{X} \in \mathbf{S}_h. \tag{2.13}$$

Proof: Using (2.10) and (1.36), (2.12) follows in a straightforward way. Then, by (2.12), (2.11) and (1.35) with $l = 0$:

$$\bar{\mathcal{D}}_h^n(s)(\mathbf{X},\mathbf{X}) = \mathcal{D}_h(t^n)(\mathbf{X},\mathbf{X}) + \int_0^s D_\tau \bar{\mathcal{D}}_h^n(\tau)(\mathbf{X},\mathbf{X})d\tau \geq c(1-k)\|\mathbf{X}\|_E^2 \qquad \forall \mathbf{X} \in \mathbf{S}_h$$

and (2.13) follows for k small enough. ∎

Discrete counterparts to Lemma 2.1 and Theorem 2.1 appear next.

Lemma 2.4 *For $0 \leq n \leq n^* - 1$, $\bar{\mathcal{L}}_h^n(s) \in C^l([0,k], \mathcal{B}(\mathbf{S}_h))$ where $l \geq 0$ is as in (1.13). Also, the following hold:*

$$\|\mathcal{T}_h^{\frac{1}{2}}(t)\partial_s^l \bar{\mathcal{L}}_h^n(s)\mathcal{T}_h^{\frac{1}{2}}(t)\mathbf{f}\| \leq c(l)\|\mathbf{f}\| \tag{2.14}$$

$$\|\mathcal{T}_h^{\frac{1}{2}}(t)[\mathcal{L}_h(t_2) - \mathcal{L}_h(t_1)]\mathcal{T}_h^{\frac{1}{2}}(t)\mathbf{f}\| \leq c|t_2 - t_1|\|\mathbf{f}\| \tag{2.15}$$

$$\|\mathcal{T}_h^{\frac{1}{2}}(t)[\bar{\mathcal{L}}_h^n(s) - \mathcal{L}_h^n]\mathcal{T}_h^{\frac{1}{2}}(t)\mathbf{f}\| \leq ck\|\mathbf{f}\| \tag{2.16}$$

$$\forall \mathbf{f} \in \mathbf{L}_2, \quad 0 \leq t, t_1, t_2 \leq t^*, \quad 0 \leq s \leq k, \quad 0 \leq n \leq n^* - 1.$$

Proof: The manipulations required are similar to those needed for Lemma 2.1, except that (1.13) is used instead of (1.2). ∎

Theorem 2.2 *Let $l \geq 0$ be as in (1.12). Then for k small enough, and $0 \leq n \leq n^* - 1$, there exist operators $\bar{\mathcal{T}}_h^n(s) \in C^l([0,k], \mathcal{B}(\mathbf{L}_2, \mathbf{S}_h))$ defined by:*

$$\bar{\mathcal{D}}_h^n(s)(\bar{\mathcal{T}}_h^n(s)\mathbf{f},\mathbf{X}) = (\mathbf{f},\mathbf{X}) \qquad \forall \mathbf{f} \in \mathbf{L}_2, \quad \forall \mathbf{X} \in \mathbf{S}_h, \quad 0 \leq s \leq k \tag{2.17}$$

and satisfying (2.7) in addition to:

$$\|\partial_s^l \bar{\mathcal{T}}_h^n(s)\mathbf{f}\|_E \leq c(l)\|\mathbf{f}\| \tag{2.18}$$

$$\|\mathcal{L}_h^{\theta_1}(t)\partial_s^l \bar{\mathcal{T}}_h^n(s)\mathcal{L}_h^{\theta_1}(t)\mathbf{X}\| \leq c(l)\|\mathbf{X}\| \tag{2.19}$$

$$\forall \mathbf{f} \in \mathbf{L}_2, \quad \forall \mathbf{X} \in \mathbf{S}_h, \quad \theta_1, \theta_2 = 0, \tfrac{1}{2} \quad 0 \leq t \leq t^*, \quad 0 \leq s \leq k, \quad 0 \leq n \leq n^* - 1.$$

Proof: That $\bar{\mathcal{T}}_h^n(s)$ is well-defined by (2.17) follows from Lemma 2.3 and the Lax-Milgram Lemma [6]. For (2.7), note first that by (2.8), for $\mathbf{X}, \mathbf{\Phi} \in \mathbf{S}_h$:

$$\bar{\mathcal{D}}_h^n(s)(\mathbf{X},\mathbf{\Phi}) \equiv \sum_{m=0}^{\mu-1} \mathcal{D}_h(t^n + \delta_m^n s)(\Gamma_m^n \mathbf{X},\mathbf{\Phi}) = \sum_{m=0}^{\mu-1} (\mathcal{L}_h(t^n + \delta_m^n s)\Gamma_m^n \mathbf{X},\mathbf{\Phi}) = (\bar{\mathcal{L}}_h^n(s)\mathbf{X},\mathbf{\Phi}).$$

Combining this with (2.17), it follows that $\forall \mathbf{X}, \mathbf{\Phi}, \mathbf{\Psi} \in \mathbf{S}_h$ and $\forall \mathbf{f} \in \mathbf{L}_2$:

$$(\bar{\mathcal{L}}_h^n(s)\bar{\mathcal{T}}_h^n(s)\mathbf{f},\mathbf{X}) = \bar{\mathcal{D}}_h^n(s)(\bar{\mathcal{T}}_h^n(s)\mathbf{f},\mathbf{X}) = (\mathbf{f},\mathbf{X}) = (\mathcal{P}_0\mathbf{f},\mathbf{X}),$$

$$\bar{\mathcal{D}}_h^n(s)(\bar{\mathcal{T}}_h^n(s)\bar{\mathcal{L}}_h^n(s)\Psi,\Phi) = (\bar{\mathcal{L}}_h^n(s)\Psi,\Phi) = \bar{\mathcal{D}}_h^n(s)(\Psi,\Phi).$$

Then (2.7) follows after setting $\mathbf{X} = [\bar{\mathcal{L}}_h^n(s)\bar{\mathcal{T}}_h^n(s) - \mathcal{P}_0]\mathbf{f}$, $\Phi = \bar{\mathcal{T}}_h^n(s)\bar{\mathcal{L}}_h^n(s)\Psi$ and using (2.13). Next, to obtain (2.19) for the case that $l = 0$, set $E \equiv (\mathcal{T}_h^n)^{\frac{1}{2}}[\mathcal{L}_h^n - \bar{\mathcal{L}}_h^n(s)](\mathcal{T}_h^n)^{\frac{1}{2}}$ so that by (2.16):

$$\|[I - E]\mathbf{X}\| \geq (1 - ck)\|\mathbf{X}\| \qquad \forall \mathbf{X} \in \mathbf{S}_h.$$

Hence, for k small enough:

$$\mathcal{L}_h^{\theta_1}(t)\bar{\mathcal{T}}_h^n(s)\mathcal{L}_h^{\theta_2}(t) = [\mathcal{L}_h^{\theta_1}(t)(\mathcal{T}_h^n)^{\frac{1}{2}}][I - E]^{-1}[(\mathcal{T}_h^n)^{\frac{1}{2}}\mathcal{L}_h^{\theta_2}(t)].$$

If $\theta_1 = \theta_2 = \frac{1}{2}$, the first case of (2.19) follows with (1.14) and (1.15). Otherwise, (1.10) is used. Now by estimating difference quotients, it can be shown in a straightforward manner that $\partial_s\bar{\mathcal{T}}_h^n(s) = -\bar{\mathcal{T}}_h^n(s)\partial_s\bar{\mathcal{L}}_h^n(s)\bar{\mathcal{T}}_h^n(s)$ for $0 \leq s \leq k$, and the smoothness of $\bar{\mathcal{T}}_h^n(s)$ follows inductively with Lemma 2.4. Next, after differentiating the first part of (2.7):

$$\mathcal{L}_h^{\theta_1}(t)\partial_s^l\bar{\mathcal{T}}_h^n(s)\mathcal{L}_h^{\theta_2}(t) = -\sum_{i=0}^{l-1}\binom{l}{i}[\mathcal{L}_h^{\theta_1}(t)\bar{\mathcal{T}}_h^n(s)\mathcal{L}_h^{\frac{1}{2}}(t)][\mathcal{T}_h^{\frac{1}{2}}(t)\partial_s^{l-i}\bar{\mathcal{L}}_h^n(s)\mathcal{T}_h^{\frac{1}{2}}(t)][\mathcal{L}_h^{\frac{1}{2}}(t)\partial_s^i\bar{\mathcal{T}}_h^n(s)\mathcal{L}_h^{\theta_2}(t)]$$

so (2.19) follows inductively using (2.14). Finally, by setting $\theta_1 = \frac{1}{2}$, $\theta_2 = 0$ and $\mathbf{X} = \mathcal{P}_0\mathbf{f}$ in (2.19), (2.18) follows with (1.11) since $\bar{\mathcal{T}}_h^n(s) = \bar{\mathcal{T}}_h^n(s)\bar{\mathcal{L}}_h^n(s)\bar{\mathcal{T}}_h^n(s) = \bar{\mathcal{T}}_h^n(s)\mathcal{P}_0$. ∎

Next, certain inequalities related to (2.9) are needed.

Lemma 2.5 *For $k > 0$ small enough, the following hold:*

$$|\partial_s^l\bar{\mathcal{D}}_h^n(s)(\mathbf{v},\mathbf{w}) - (\partial_s^l\bar{\mathcal{L}}^n(s)\mathbf{v},\mathbf{w})| \leq c(l)h^{m-1}\|\mathbf{v}\|_m\|\mathbf{w} - \mathbf{u}\|_E \tag{2.20}$$

$$|\partial_s^l\bar{\mathcal{D}}_h^n(s)(\mathbf{w},\mathbf{v}) - (\mathbf{w},\partial_s^l\bar{\mathcal{L}}^n(s)^*\mathbf{v})| \leq c(l)h^{m-1}\|\mathbf{v}\|_m\|\mathbf{w} - \mathbf{u}\|_E \tag{2.21}$$

$$|\sum_{i=0}^{l}\binom{l}{i}\partial_s^{l-i}\bar{\mathcal{D}}_h^n(s)([\partial_s^i\bar{\mathcal{T}}^n(s) - \partial_s^i\bar{\mathcal{T}}_h^n(s)]\mathbf{f},\mathbf{X})| \leq c(l)h^{m-1}\|\mathbf{f}\|_{m-2}\|\mathbf{X} - \mathbf{u}\|_E \tag{2.22}$$

$$\forall \mathbf{v} \in \mathbf{H}^m \cap \mathbf{H}_0^1,\ \ \forall \mathbf{w} \in \mathbf{H}^2 \cap \mathbf{H}_0^1 + \mathbf{S}_h,\ \ \forall \mathbf{u} \in \mathbf{H}^2 \cap \mathbf{H}_0^1,\ \ \forall \mathbf{f} \in \mathbf{H}^m,\ \ \forall \mathbf{X} \in \mathbf{S}_h\ \ 2 \leq m \leq r,\ \ l \geq 0,\ \ 0 \leq s \leq k,\ \ 0 \leq n \leq n^* - 1.$$

Proof: First, note that:

$$\partial_s^l\bar{\mathcal{D}}_h^n(s)(\mathbf{v},\mathbf{w}) - (\partial_s^l\bar{\mathcal{L}}^n(s)\mathbf{v},\mathbf{w}) = \sum_{i=0}^{\mu-1}(\delta_i^n)^l[\mathcal{D}_h^{(l)}(t^n + \delta_i^n s)(\Gamma_i^n\mathbf{v},\mathbf{w}) - (\mathcal{L}^{(l)}(t^n + \delta_i^n s)\Gamma_i^n\mathbf{v},\mathbf{w})]$$

so (2.20) follows with (2.9). Also, (2.21) follows similarly. For the remaining inequality, the key observation is that by (2.17), the left side of (2.22) is equal to:

$$|\partial_s^l[\bar{\mathcal{D}}_h^n(s)(\bar{\mathcal{T}}^n(s)\mathbf{f},\mathbf{X}) - (\mathbf{f},\mathbf{X})]| = |\partial_s^l[\bar{\mathcal{D}}_h^n(s)(\bar{\mathcal{T}}^n(s)\mathbf{f},\mathbf{X}) - (\bar{\mathcal{L}}^n(s)\bar{\mathcal{T}}^n(s)\mathbf{f},\mathbf{X})]|$$

and (2.22) follows using (2.20) and (2.4). ∎

The groundwork for a generalization of (1.9) is now complete.

Theorem 2.3 *For $k > 0$ small enough, the following holds for $0 \le n \le n^* - 1$, $0 \le s \le k$, $l \ge 0$, and $2 \le m \le r$:*

$$\|\partial_s^l[\bar{\mathcal{T}}^n(s) - \bar{\mathcal{T}}_h^n(s)]\mathbf{v}\| + h\|\partial_s^l[\bar{\mathcal{T}}^n(s) - \bar{\mathcal{T}}_h^n(s)]\mathbf{v}\|_E \le c(l)h^m\|\mathbf{v}\|_{m-2} \qquad \forall \mathbf{v} \in \mathbf{H}^{m-2}. \tag{2.23}$$

Proof: With $2 \le m \le r$, let $\mathbf{v} \in \mathbf{H}^{m-2}$ and define $\mathbf{X}_l \in \mathbf{S}_h$ to be the closest to $\partial_s^l \bar{\mathcal{T}}^n(s)\mathbf{v}$ in the norm $\|\cdot\|_E$. Then, define:

$$E^l \equiv \partial_s^l[\bar{\mathcal{T}}^n(s) - \bar{\mathcal{T}}_h^n(s)]\mathbf{v} = [\partial_s^l \bar{\mathcal{T}}^n(s)\mathbf{v} - \mathbf{X}_l] - [\partial_s^l \bar{\mathcal{T}}_h^n(s)\mathbf{v} - \mathbf{X}_l] \equiv E_0^l - E_h^l.$$

By (1.8) and (2.4):

$$\|E_0^l\| \le c(l)h^{m-1}\|\mathbf{v}\|_{m-2}.$$

Next, by (2.12) and (2.13):

$$c\|E_h^l\|_E^2 \le \bar{\mathcal{D}}_h^n(s)(E_h^l, E_h^l) = \bar{\mathcal{D}}_h^n(s)(E_0^l, E_h^l) - \bar{\mathcal{D}}_h^n(s)(E^l, E_h^l) \le c\|E_0^l\|_E\|E_h^l\|_E + |\bar{\mathcal{D}}_h^n(s)(E^l, E_h^l)|$$

By (2.22), with ill-defined sums understood to be zero:

$$\begin{aligned} |\bar{\mathcal{D}}_h^n(s)(E^l, E_h^l)| &= |\sum_{i=0}^{l} \binom{l}{i} \partial_s^{l-i} \bar{\mathcal{D}}_h^n(s)(E^i, E_h^l) - \sum_{i=0}^{l-1} \binom{l}{i} \partial_s^{l-i} \bar{\mathcal{D}}_h^n(s)(E^i, E_h^l)| \\ &\le c(l)h^{m-1}\|\mathbf{v}\|_{m-2}\|E_h^l\|_E + c\|E_h^l\|_E \sum_{i=0}^{l-1} \|E^i\|_E. \end{aligned}$$

So, the indicated estimate for $\|E^l\|_E$ follows inductively from the last three inequalities. Now, a duality argument is used to complete the proof. Define $\mathbf{X}_l^* \in \mathbf{S}_h$ to be the closest to $\bar{\mathcal{T}}^n(s)^* E^l$ in the norm $\|\cdot\|_E$. Then, with ill-defined sums understood to be zero:

$$\begin{aligned} \|E^l\|^2 &= \sum_{i=0}^{l} \binom{l}{i} [(E^i, \partial_s^{l-i} \bar{\mathcal{L}}^n(s)^* \bar{\mathcal{T}}^n(s)^* E^l) - \partial_s^{l-i} \bar{\mathcal{D}}_h^n(s)(E^i, \bar{\mathcal{T}}^n(s)^* E^l)] \\ &+ \sum_{i=0}^{l} \binom{l}{i} \partial_s^{l-i} \bar{\mathcal{D}}_h^n(s)(E^i, \bar{\mathcal{T}}^n(s)^* E^l - \mathbf{X}_l^*) + \sum_{i=0}^{l} \binom{l}{i} \partial_s^{l-i} \bar{\mathcal{D}}_h^n(s)(E^i, \mathbf{X}_l^*) \\ &- \sum_{i=0}^{l-1} \binom{l}{i} (E^i, \partial_s^{l-i} \bar{\mathcal{L}}^n(s)^* \bar{\mathcal{T}}^n(s)^* E^l) \equiv \sum_{j=1}^{4} F_j. \end{aligned}$$

By (2.21) and (2.6), F_1 satisfies:

$$|F_1| + |F_2| \le c(l)h\|E^l\| \sum_{i=0}^{l} \|E^i\|_E$$

while the case for F_2 follows with (2.12), (1.8), and (2.6). Then, by (2.22), (1.8), and (2.6):

$$|F_3| \le c(l)h^{m-1}\|\mathbf{v}\|_{m-2}\|\mathbf{X}_l^* - \bar{\mathcal{T}}^n(s)^* E^l\|_E \le c(l)h^m\|\mathbf{v}\|_{m-2}\|E^l\|.$$

Finally, by (2.5) and (2.6):

$$|F_4| \le c(l)\|E^l\| \sum_{i=0}^{l-1} \|E^i\|.$$

Now (2.23) follows inductively. ∎

3 The Base Scheme.

In this section, the base scheme (1.38) is analyzed for the approximation of the solution to (1.1) and (1.5) is established. That the stages are well-defined depends on the next lemma.

Lemma 3.1 *Provided (1.31) is satisfied, $[I + kA\mathcal{L}_h^n]$ is invertible, and for k small enough, $[I + kA\bar{\mathcal{L}}_h^n]$ is as well. Also the following hold:*

$$\|(k\mathcal{L}_h^n)^\theta [I + kA\mathcal{L}_h^n]^{-1}\mathbf{X}\| \le c\|\mathbf{X}\|, \tag{3.1}$$

$$\|(k\mathcal{L}_h^n)^{\theta_1} [I + kA\bar{\mathcal{L}}_h^n]^{-1}(k\mathcal{L}_h^n)^{\theta_2}\mathbf{X}\| \le c\|\mathbf{X}\|, \tag{3.2}$$

$$\|(k\mathcal{L}_h^n)^{\theta_1} [I + kA\mathcal{L}_h^m]^{-1}(k\mathcal{L}_h^n)^{\theta_2}\mathbf{X}\| \le c\|\mathbf{X}\|, \tag{3.3}$$

$$\forall \mathbf{X} \in \mathbf{S}_h, \quad 0 \le \theta \le 1, \quad \theta_1, \theta_2 = 0, \tfrac{1}{2}; \quad \theta_1 = -\theta_2 = \pm\tfrac{1}{2}, \quad 0 \le m \le n^*, \quad 0 \le n \le n^* - 1.$$

Proof: The invertibility of $[I + kA\mathcal{L}_h^n]$ and the estimate (3.1) involve a spectral argument after A is transformed to Jordan form, and the details are provided by Karakashian [13]. Now set:

$$E_1 \equiv [I + kA\mathcal{L}_h^n]^{-1} kA(\mathcal{L}_h^n - \bar{\mathcal{L}}_h^n) \qquad \text{and} \qquad E_2 \equiv kA(\mathcal{L}_h^n - \bar{\mathcal{L}}_h^n)[I + kA\mathcal{L}_h^n]^{-1}$$

so that:

$$(\mathcal{L}_h^n)^{\frac{1}{2}}[I + kA\bar{\mathcal{L}}_h^n](\mathcal{T}_h^n)^{\frac{1}{2}} = [I + kA\mathcal{L}_h^n][I - (\mathcal{L}_h^n)^{\frac{1}{2}} E_1 (\mathcal{T}_h^n)^{\frac{1}{2}}],$$

$$(\mathcal{T}_h^n)^{\frac{1}{2}}[I + kA\bar{\mathcal{L}}_h^n](\mathcal{L}_h^n)^{\frac{1}{2}} = [I - (\mathcal{T}_h^n)^{\frac{1}{2}} E_2 (\mathcal{L}_h^n)^{\frac{1}{2}}][I + kA\mathcal{L}_h^n].$$

By (3.1) and (2.16):

$$\|(\mathcal{L}_h^n)^{\frac{1}{2}} E_1 (\mathcal{T}_h^n)^{\frac{1}{2}}\mathbf{X}\| + \|(\mathcal{T}_h^n)^{\frac{1}{2}} E_2 (\mathcal{L}_h^n)^{\frac{1}{2}}\mathbf{X}\| \le ck\|\mathbf{X}\| \qquad \forall \mathbf{X} \in \mathbf{S}_h.$$

Hence, for k small enough, $[I + kA\bar{\mathcal{L}}_h^n]$ is invertible. Next, for $\theta_2 = 0, \pm\frac{1}{2}$:

$$(k\mathcal{L}_h^n)^{\frac{1}{2}}[I + kA\bar{\mathcal{L}}_h^n]^{-1}(k\mathcal{L}_h^n)^{\theta_2} = [I - (\mathcal{L}_h^n)^{\frac{1}{2}} E_1 (\mathcal{T}_h^n)^{\frac{1}{2}}]^{-1}(k\mathcal{L}_h^n)^{\theta_2 + \frac{1}{2}}[I + kA\mathcal{L}_h^n]^{-1}$$

and (3.2) follows for $\theta_1 = \frac{1}{2}$. For $\theta_1 = 0, \pm\frac{1}{2}$:

$$(k\mathcal{L}_h^n)^{\theta_1}[I + kA\bar{\mathcal{L}}_h^n]^{-1}(k\mathcal{L}_h^n)^{\frac{1}{2}} = (k\mathcal{L}_h^n)^{\theta_1 + \frac{1}{2}}[I + kA\mathcal{L}_h^n]^{-1}[I - (\mathcal{T}_h^n)^{\frac{1}{2}} E_2 (\mathcal{L}_h^n)^{\frac{1}{2}}]^{-1}$$

and (3.2) follows for $\theta_2 = \frac{1}{2}$. Now, for the case $\theta_1 = \theta_2 = 0$, with $\mathbf{X} \in \mathbf{S}_h$ chosen arbitrarily:

$$\|\mathbf{X}\|^2 \le \tfrac{1}{2}\|[I + kA\bar{\mathcal{L}}_h^n]\mathbf{X}\|^2 + \tfrac{1}{2}\|\mathbf{X}\|^2 + k|(A\bar{\mathcal{L}}_h^n\mathbf{X}, \mathbf{X})|.$$

Then with $\mathbf{\Psi} \equiv (\mathcal{L}_h^n)^{\frac{1}{2}}\mathbf{X}$, by (2.16):

$$|(A\bar{\mathcal{L}}_h^n\mathbf{X}, \mathbf{X})| \le |(A\mathbf{\Psi}, \mathbf{\Psi})| + |((\mathcal{T}_h^n)^{\frac{1}{2}}[\bar{\mathcal{L}}_h^n - \mathcal{L}_h^n](\mathcal{T}_h^n)^{\frac{1}{2}}\mathbf{\Psi}, A^T\mathbf{\Psi})| \le c(1 + k)(\mathcal{L}_h^n\mathbf{X}, \mathbf{X}).$$

Also by (3.2) with $\theta_1 = \frac{1}{2}$, $\theta_2 = 0$:

$$\|(k\mathcal{L}_h^n)^{\frac{1}{2}}\mathbf{X}\| \le c\|[I + kA\bar{\mathcal{L}}_h^n]\mathbf{X}\|$$

and the remaining case for (3.2) follows after combining the last three inequalities. Finally, using:

$$(k\mathcal{L}_h^n)^{\theta_1}[I+kA\mathcal{L}_h^m]^{-1}(k\mathcal{L}_h^n)^{\theta_2} = [(\mathcal{L}_h^n)^{\theta_1}(\mathcal{T}_h^m)^{\theta_1}](k\mathcal{L}_h^m)^{\theta_1}[I+kA\mathcal{L}_h^m]^{-1}(k\mathcal{L}_h^m)^{\theta_2}[(\mathcal{T}_h^m)^{\theta_2}(\mathcal{L}_h^n)^{\theta_2}]$$

(3.3) follows with (1.14), (1.15), and (3.1). ∎

Now, for the sequel, let the following be defined:

$$\xi^n \equiv U_h^n - \omega^n, \qquad \eta^n \equiv u^n - \omega^n,$$

$$r_h^n \equiv I - kb^T\mathcal{L}_h^n[I+kA\mathcal{L}_h^n]^{-1}e, \qquad \mathcal{R}_h^n \equiv I - kb^T\bar{\mathcal{L}}_h^n[I+kA\bar{\mathcal{L}}_h^n]^{-1}e,$$

$$\bar{u}^n(s) \equiv \sum_{m=0}^{\mu-1} D^l e\partial_t^l u^n \frac{s^l}{l!}, \quad (0\le s\le k) \qquad \bar{u}^n \equiv \bar{u}^n(k),$$

$$\bar{\omega}^n(s) \equiv \bar{\mathcal{T}}_h^n(s)\bar{\mathcal{L}}^n(s)\bar{u}^n(s), \quad (0\le s\le k) \qquad \bar{\omega}^n \equiv \bar{\omega}^n(k).$$

After some straightforward calculations, the following error equation is established:

$$\begin{aligned}
\xi^{n+1} = \mathcal{R}_h^n\xi^n \;&+\; kb^T\bar{\mathcal{L}}_h^n[I+kA\bar{\mathcal{L}}_h^n]^{-1}\{\bar{\omega}^n - e\omega^n - kA\sum_{m=0}^{\mu-1}\Gamma_m^n e\partial_t\omega(\tau_m^n)\}\\
&-\; kb^TA^{-1}[I+kA\bar{\mathcal{L}}_h^n]^{-1}AP_0\sum_{m=0}^{\mu-1}\Gamma_m^n\mathcal{L}(\tau_m^n)[\bar{u}^n - eu(\tau_m^n)]\\
&-\; kb^TA^{-1}[I+kA\bar{\mathcal{L}}_h^n]^{-1}A\sum_{m=0}^{\mu-1}\Gamma_m^n e\partial_t\{[P_E(\tau_m^n)-P_0]u(\tau_m^n)\}\\
&-\; \{\omega^{n+1} - \omega^n - kb^T\sum_{m=0}^{\mu-1}\Gamma_m^n e\partial_t\omega(\tau_m^n)\}\\
&\equiv\; \mathcal{R}_h^n\xi^n + \sum_{l=1}^{4}\psi_l^n \;\equiv\; \mathcal{R}_h^n\xi^n + \psi^n \qquad 0\le n\le n^*-1.
\end{aligned} \tag{3.4}$$

Now, stability is to be established in the following norms, which according to (1.11) are well-defined for $0\le n\le n^*$:

$$|||\chi|||_n \equiv \{(\chi,\chi) + k(L_h^n\chi,\chi)\}^{\frac{1}{2}} \qquad \chi\in S_h. \tag{3.5}$$

Also, from (1.16) with $l=0$, it follows that these norms are equivalent:

$$c_1|||\chi|||_m \le |||\chi|||_n \le c_2|||\chi|||_m \qquad \forall\chi\in S_h, \quad 0\le m,n\le n^*. \tag{3.6}$$

As in section 2, let the following be defined in the natural way for the product spaces:

$$(\mathbf{\Phi},\mathbf{\Psi}) \equiv \sum_{i=1}^{q}(\phi_i,\psi_i), \qquad \|\mathbf{\Phi}\| \equiv (\mathbf{\Phi},\mathbf{\Phi})^{\frac{1}{2}} \qquad \mathbf{\Phi},\mathbf{\Psi}\in\mathbf{L}_2,$$

$$|||\mathbf{X}|||_n \equiv \{\sum_{i=1}^{q}|||\chi_i|||_n^2\}^{\frac{1}{2}} \qquad \mathbf{X}\in\mathbf{S}_h.$$

Finally, (3.17) follows after using (3.7) and (1.26) in:

$$\|(kL_h^n)^{\frac{1}{2}}\mathcal{R}_h^n\xi^n\| \le \|[(kL_h^n)^{\frac{1}{2}}(\mathcal{R}_h^n - r_h^n)(kL_h^n)^{-\frac{1}{2}}](kL_h^n)^{\frac{1}{2}}\xi^n\| + \|r_h^n(kL_h^n)^{\frac{1}{2}}\xi^n\|.$$

Next, (3.15)-(3.17) are used to obtain (3.13). Suppose that ε_2 is small enough that $c_2 < 1$. Then, assume that $k_0 > 0$ is small enough that if $\theta \equiv (1 - c_1)/(1 + c_3k_0) > 0$, then $c_2 + \theta > 1$. Next, multiply (3.17) by θ and add the result to (3.16). With $c_5 \equiv c_2 + \theta - 1 > 0$, and $0 < k \le k_0$:

$$\|\mathcal{R}_h^n\xi^n\|^2 + (1 + c_5)\|(kL_h^n)^{\frac{1}{2}}\mathcal{R}_h^n\xi^n\|^2 \le |||\xi^n|||_n^2.$$

By (1.11), there is a $c_6 > 0$ such that:

$$c_6k\|\mathcal{R}_h^n\xi^n\|^2 \le \tfrac{1}{2}c_5\|(kL_h^n)^{\frac{1}{2}}\mathcal{R}_h^n\xi^n\|^2.$$

Also, by (2.15), with $\chi^n \equiv (kL_h^n)^{\frac{1}{2}}\mathcal{R}_h^n\xi^n$:

$$\|(kL_h^{n+1})^{\frac{1}{2}}\mathcal{R}_h^n\xi^n\|^2 = \|\chi^n\|^2 + ((T_h^n)^{\frac{1}{2}}[L_h^{n+1} - L_h^n](T_h^n)^{\frac{1}{2}}\chi^n, \chi^n) \le (1 + c_7k)\|(kL_h^n)^{\frac{1}{2}}\mathcal{R}_h^n\xi^n\|^2.$$

From the last three inequalities, it follows that:

$$(1 + c_6k)\|\mathcal{R}_h^n\xi^n\|^2 + (1 + \tfrac{1}{2}c_5)(1 + c_7k)^{-1}\|(k\mathcal{L}_h^{n+1})^{\frac{1}{2}}\mathcal{R}_h^n\xi^n\|^2 \le |||\xi^n|||_n^2.$$

So assume that k_0 above, is also small enough that $(1 + \frac{1}{2}c_5)(1 + c_6k_0)^{-1}(1 + c_7k_0)^{-1} \ge 1 + \varepsilon_1$, for some $\varepsilon_1 > 0$. Then (3.13) follows for some $\tilde{c} \in (-c_6, 0)$. ∎

The next two lemmas are useful in subsequent consistency estimates.

Lemma 3.3 *Let t_0, t_1, $t_2 \in [0, t^*]$ and $|t_2 - t_1| \le ck$. Then, for integers $m, l \ge 0$:*

$$\sup_{0\le t\le t^*} \|L_h^\theta(t_0)\partial_s^l\omega(t)\| \le c(l)\|u^0\|_{2(l+1)} \qquad \theta = 0, \tfrac{1}{2} \tag{3.18}$$

$$\|E\| \le c(l)(k^{m+1} + h^2k^m)\|u^0\|_{2l}, \qquad E \equiv \int_{t_1}^{t_2}(t_2 - t)^m\partial_t^l\omega(t)dt. \tag{3.19}$$

Also, there exist E_1 and E_2 such that $E = E_1 + E_2$ while:

$$\|[kL_h(t_0)]^{\frac{1}{2}}E_1\| \le c(l)hk^{m+\frac{1}{2}+i}\|u^0\|_{2(l+i)} \qquad i = 0, 1 \tag{3.20}$$

$$\|kL_h(t_0)E_2\| \le c(l)k^{m+1+i}\|u^0\|_{2(l+i)} \qquad i = 0, 1. \tag{3.21}$$

Furthermore, for $0 \le n \le n^*$:

$$|||E|||_n \le c(l)(k^{m+1} + hk^{m+\frac{1}{2}} + h^2k^m)\|u^0\|_{2l}. \tag{3.22}$$

Proof. By (1.6) or (1.17), (1.7), (1.19) and (1.4), for $0 \le t \le t^*$, and $\theta = 0, \frac{1}{2}$:

$$\|L_h^\theta(t_0)\partial_t^l\omega(t)\| \le c\|\partial_t^l\omega(t)\|_E \le c\|\partial_t^l\eta(t)\|_E + c\|\partial_t^lu(t)\|_2 \le c(l)(h + 1)\|u^0\|_{2(l+1)}$$

which gives (3.18). Next:

$$E = (t_2 - t_1)^m\partial_t^{l-1}\eta(t_1) - m\int_{t_1}^{t_2}(t_2 - t)^{m-1}\partial_t^{l-1}\eta(t)dt + \int_{t_1}^{t_2}(t_2 - t)^m\partial_t^lu(t)dt$$

and (3.19) follows with (1.19) and (1.4). Now, define:

$$\begin{aligned} E_1 &\equiv \int_{t_1}^{t_2}(t_2-t)^m\partial_t^l[\omega(t)-P_E(t_0)u(t)]dt \\ &= -(t_2-t_1)^m\partial_t^{l-1}[\omega(t_1)-P_E(t_0)u(t_1)]+m\int_{t_1}^{t_2}(t_2-t)^{m-1}\partial_t^{l-1}[\omega(t)-P_E(t_0)u(t)]dt, \\ E_2 &\equiv \int_{t_1}^{t_2}(t_2-t)^m P_E(t_0)\partial_t^l u(t)dt \\ &= -(t_2-t_1)^m P_E(t_0)\partial_t^{l-1}u(t_1)+m\int_{t_1}^{t_2}(t_2-t)^{m-1}P_E(t_0)\partial_t^{l-1}u(t)dt. \end{aligned}$$

By (1.17), (1.19), (1.18), and (1.4), for $i=0,\,1$:

$$\begin{aligned} \|[kL_h(t_0)]^{\frac{1}{2}}E_1\| &\le ck^{m+\frac{1}{2}+i}\sup_{0\le t\le t^*}\{\|\partial_t^{l-1+i}\eta(t)\|_E+\|[I-P_E(t_0)]\partial_t^{l-1+i}u(t)\|_E\} \\ &\le c(l)k^{m+\frac{1}{2}+i}h\|u^0\|_{2(l+i)}. \end{aligned}$$

By (1.2) and (1.4), for $i=0,\,1$:

$$\|kL_h(t_0)E_2\|\le ck^{m+1+i}\sup_{0\le t\le t^*}\|L(t_0)\partial_t^{l-1+i}u(t)\|\le c(l)k^{m+1+i}\|u^0\|_{2(l+i)}.$$

Now, since $E=E_1+E_2$, (3.20) and (3.21) are established. Finally:

$$(kL_h^nE,E)\le\tfrac{1}{2}\|(kL_h^n)^{\frac{1}{2}}E_1\|^2+\tfrac{1}{2}\|(kL_h^n)^{\frac{1}{2}}E\|^2+\tfrac{1}{2}\|kL_h^nE_2\|^2+\tfrac{1}{2}\|E\|^2$$

and (3.22) follows after combining this with (3.19)-(3.21). ∎

Lemma 3.4 *The following hold for* $0\le s\le k$, *and* $0\le t\le t^*$:

$$\|\mathcal{L}_h^\theta(t)\partial_s^l\bar\omega^n(s)\|\le c\|u^0\|_{2\mu}\qquad 0\le l\le\mu-1,\quad \theta=0,\tfrac{1}{2} \tag{3.23}$$

$$\|\partial_s^\mu\bar\omega^n(s)\|+h\|\mathcal{L}_h^{\frac{1}{2}}(t)\partial_s^\mu\bar\omega^n(s)\|\le ch^2\|u^0\|_{2\mu}. \tag{3.24}$$

Proof: From (1.17) or (1.6), (2.18), (2.2), and (1.4), it follows that for $0\le l\le\mu-1$ and $\theta=0,\,\frac{1}{2}$:

$$\begin{aligned} \|\mathcal{L}_h^\theta(t)\partial_s^l\bar\omega^n(s)\| &\le c(l)\sum_{i=0}^{l}\|\partial_s^{l-i}\bar{\mathcal{T}}_h^n(s)\partial_s^i[\bar{\mathcal{L}}^n(s)\bar u^n(s)]\|_E \le c(l)\sum_{i=0}^{l}\sum_{j=0}^{i}\|\partial_s^{i-j}\bar{\mathcal{L}}^n(s)\partial_s^j\bar u^n(s)\| \\ &\le c(l)\sum_{j=0}^{l}\sum_{m=j}^{\mu-1}\|\partial_t^m u^n s^{m-j}\|_2 \le c(l)\|u^0\|_{2\mu} \end{aligned}$$

which gives (3.23). Now, since $\partial_s^\mu\bar u^n(s)\equiv 0$, from (2.23), (2.2), and (1.4), it follows that:

$$\begin{aligned} \|\partial_s^\mu\bar\omega^n(s)\| &= \|\partial_s^\mu[\bar\omega^n(s)-\bar u^n(s)]\| \le c\sum_{i=0}^{\mu}\|\partial_s^{\mu-i}[\bar{\mathcal{T}}_h^n(s)-\bar{\mathcal{T}}^n(s)]\partial_s^i[\bar{\mathcal{L}}^n(s)\bar u^n(s)]\| \\ &\le ch^2\sum_{i=0}^{\mu}\sum_{j=0}^{i}\|\partial_s^{i-j}\bar{\mathcal{L}}^n(s)\partial_s^j\bar u^n(s)\| \le ch^2\sum_{j=0}^{\mu-1}\sum_{m=j}^{\mu-1}\|\partial_t^m u^n s^{m-j}\|_2 \le ch^2\|u^0\|_{2\mu}. \end{aligned}$$

The remaining component of (3.24) follows similarly, after using (1.17) first. ∎

The order of consistency is established in the next four lemmas.

Lemma 3.5 *ψ_1^n of (3.4) satisfies:*

$$|||\psi_1^n|||_n \le ck(k^\mu + hk^{m-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_{2(\mu+1)}. \tag{3.25}$$

Proof: First, it is proved that:

$$\partial_s^l \bar{\omega}^n(0) = D^l e \partial_t^l \omega^n \qquad 0 \le l \le \mu - 1. \tag{3.26}$$

Now, the result of differentiating $\bar{\mathcal{L}}_h^n(s)\bar{\omega}^n(s) = \mathcal{P}_0\bar{\mathcal{L}}^n(s)\bar{u}^n(s)$ is:

$$\sum_{i=0}^{l}[\sum_{m=0}^{\mu-1}\Gamma_m^n(\delta_m^n)^{l-i}]\begin{pmatrix} l \\ i \end{pmatrix}\mathcal{L}_h^{(l-i)}(t^n+\delta_m^n s)\partial_s^i\bar{\omega}^n(s) = \mathcal{P}_0\sum_{i=0}^{l}[\sum_{m=0}^{\mu-1}\Gamma_m^n(\delta_m^n)^{l-i}]\begin{pmatrix} l \\ i \end{pmatrix}\mathcal{L}^{(l-i)}(t^n+\delta_m^n s)\sum_{j=i}^{\mu-1}D^j e\partial_t^j u^n\frac{s^{j-i}}{(j-i)!}.$$

Letting $s \to 0$, with (1.35), it follows that for $0 \le l \le \mu - 1$:

$$\mathcal{L}_h^n\partial_s^l\bar{\omega}^n(0) = \mathcal{P}_0\sum_{i=0}^{l}D^{l-i}\begin{pmatrix} l \\ i \end{pmatrix}\mathcal{L}^{(l-i)}(t^n)D^i e\partial_t^i u^n - \sum_{i=0}^{l-1}D^{l-i}\begin{pmatrix} l \\ i \end{pmatrix}\mathcal{L}_h^{(l-i)}(t^n)\partial_s^i\bar{\omega}^n(0).$$

Then (3.26) follows inductively with:

$$L_h^n\partial_t^l\omega^n = P_0\sum_{i=0}^{l}\begin{pmatrix} l \\ i \end{pmatrix}L^{(l-i)}(t^n)\partial_t^i u^n - \sum_{i=0}^{l-1}\begin{pmatrix} l \\ i \end{pmatrix}L_h^{(l-i)}(t^n)\partial_t^i\omega^n$$

which results after differentiating $L_h(t)\omega(t) = P_0L(t)u(t)$. Therefore:

$$\bar{\omega}^n - e\omega^n = \sum_{l=1}^{\mu-1}D^l e\partial_t^l\omega^n\frac{k^l}{l!} + E, \qquad E \equiv \frac{1}{(\mu-1)!}\int_0^k (k-s)^{\mu-1}\partial_s^\mu\bar{\omega}^n(s)ds.$$

Next, by (1.35) and (1.33):

$$kA\sum_{m=0}^{\mu-1}\Gamma_m^n e\partial_t\omega(\tau_m^n) = kA\sum_{l=0}^{\mu-2}[\sum_{m=0}^{\mu-1}\Gamma_m^n(\delta_m^n)^l]e\partial_t^{l+1}\omega^n\frac{k^l}{l!} + F = \sum_{l=0}^{\mu-2}AD^l e\partial_t^{l+1}\omega^n\frac{k^{l+1}}{l!} + F$$
$$= \sum_{l=1}^{\mu-1}D^l e\partial_t^l\omega^n\frac{k^l}{l!} + F, \qquad F \equiv \frac{k}{(\mu-2)!}\sum_{m=0}^{\mu-1}A\Gamma_m^n e\int_{t^n}^{\tau_m^n}(\tau_m^n - t)^{\mu-2}\partial_t^\mu\omega(t)dt.$$

Now define:

$$\tilde{E} \equiv k^{\frac{1}{2}}b^TA^{-1}[I + kA\bar{\mathcal{L}}_h^n]^{-1}(k\mathcal{L}_h^n)^{\frac{1}{2}}A(\mathcal{T}_h^n)^{\frac{1}{2}}\bar{\mathcal{L}}_h^n(\mathcal{T}_h^n)^{\frac{1}{2}}[(\mathcal{L}_h^n)^{\frac{1}{2}}E].$$

so that by (3.2), (2.14), and (3.24):

$$|||\tilde{E}|||_n \le ck^{\frac{1}{2}}\|(\mathcal{L}_h^n)^{\frac{1}{2}}E\| \le ck^{\mu+\frac{1}{2}}\sup_{0\le s\le k}\|(\mathcal{L}_h^n)^{\frac{1}{2}}\partial_s^\mu\bar{\omega}^n(s)\| \le chk^{\mu+\frac{1}{2}}\|u^0\|_{2\mu}.$$

Next, by Lemma 3.3, let $F = F_1 + F_2$ where with (1.36):

$$\|(k\mathcal{L}_h^n)^{\frac{1}{2}}F_1\| \le chk^{\mu+\frac{1}{2}}\|u^0\|_{2(\mu+1)}, \qquad \|k\mathcal{L}_h^nF_2\| \le ck^{\mu+1}\|u^0\|_{2(\mu+1)},$$

and:

$$|||F|||_n \leq c(k^\mu + hk^{\mu-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_{2\mu}.$$

Then, define:

$$\begin{aligned}\tilde{F} \equiv\ & b^TA^{-1}[I+kA\bar{\mathcal{L}}_h^n]^{-1}(k\mathcal{L}_h^n)^{\frac{1}{2}}A(\mathcal{T}_h^n)^{\frac{1}{2}}[\bar{\mathcal{L}}_h^n-\mathcal{L}_h^n](\mathcal{T}_h^n)^{\frac{1}{2}}[(k\mathcal{L}_h^n)^{\frac{1}{2}}F] \\ & +b^TA^{-1}[I+kA\bar{\mathcal{L}}_h^n]^{-1}(k\mathcal{L}_h^n)^{\frac{1}{2}}A[(k\mathcal{L}_h^n)^{\frac{1}{2}}F_1]+b^TA^{-1}[I+kA\bar{\mathcal{L}}_h^n]^{-1}A[k\mathcal{L}_h^nF_2]\end{aligned}$$

so that by (3.2), (2.16), and the last three inequalities:

$$|||\tilde{F}|||_n \leq ck|||F|||_n + c\|(k\mathcal{L}_h^n)^{\frac{1}{2}}F_1\| + c\|k\mathcal{L}_h^nF_2\| \leq ck(k^\mu + hk^{\mu-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_{2(\mu+1)}.$$

Now since $\psi_1^n = \tilde{E} - \tilde{F}$, (3.25) follows. ∎

Lemma 3.6 *ψ_2^n of (3.4) satisfies:*

$$|||\psi_2^n|||_n \leq ck^{\mu+1}\|u^0\|_{2(\mu+1)}. \tag{3.27}$$

Proof: Define $\bar{v}(s) \equiv \bar{\mathcal{L}}^n(s)\bar{u}^n(s)$ so that with ill-defined terms understood to be zero:

$$\partial_s^l\bar{v}(s) = \sum_{i=0}^{l}\binom{l}{i}\sum_{m=0}^{\mu-1}\Gamma_m^n[(\delta_m^n)^{l-i}\mathcal{L}^{(l-i)}(t^n+\delta_m^ns)][\sum_{j=i}^{\mu-1}D^jed_t^ju^n\frac{s^{j-i}}{(j-i)!}].$$

Letting $s \to 0$, with (1.35) and (1.1), it follows that for $0 \leq l \leq \mu-1$:

$$\partial_s^l\bar{v}(0) = \sum_{i=0}^{l}\binom{l}{i}[\sum_{m=0}^{\mu-1}\Gamma_m^n(\delta_m^n)^{l-i}]D^ieL^{(l-i)}(t^n)\partial_t^iu^n = -D^ied_t^{l+1}u^n.$$

Hence:

$$\bar{\mathcal{L}}^n\bar{u}^n = -\sum_{l=0}^{\mu-1}D^ied_t^{l+1}u^n\frac{k^l}{l!}+E, \qquad E \equiv \frac{1}{(\mu-1)!}\int_0^k(k-s)^{\mu-1}\partial_s^\mu\bar{v}(s)ds.$$

Next, by (1.1) and (1.35):

$$\begin{aligned}\sum_{m=0}^{\mu-1}\Gamma_m^n\mathcal{L}(\tau_m^n)eu(\tau_m^n) &= -\sum_{m=0}^{\mu-1}\Gamma_m^ne\partial_tu(\tau_m^n) = -\sum_{l=0}^{\mu-1}[\sum_{m=0}^{\mu-1}\Gamma_m^n(\delta_m^n)^l]ed_t^{l+1}u^n\frac{k^l}{l!}-F \\ = -\sum_{l=0}^{\mu-1}D^ied_t^{l+1}u^n\frac{k^l}{l!}-F, &\qquad F \equiv \frac{1}{(\mu-1)!}\sum_{m=0}^{\mu-1}\Gamma_m^ne\int_{t^n}^{\tau_m^n}(\tau_m^n-t)^{\mu-1}\partial_t^{\mu+1}u(t)dt.\end{aligned}$$

Hence by (3.2):

$$|||\psi_2^n|||_n \leq ck(\|E\|+\|F\|).$$

Now, by (2.2) and (1.4):

$$\|E\| \leq ck^\mu\sum_{l=0}^{\mu}\sup_{0\leq s\leq k}\|\partial_s^{\mu-l}\bar{\mathcal{L}}^n(s)\partial_s^l\bar{u}^n(s)\| \leq ck^\mu\sum_{i=0}^{\mu-1}\|\partial_t^iu^n\|_2 \leq ck^\mu\|u^0\|_{2\mu}.$$

Also, by (1.36) and (1.4):

$$\|F\| \le ck^{\mu} \sup_{0\le t\le t^*} \|\partial_t^{\mu+1} u(t)\| \le ck\|u^0\|_{2(\mu+1)}$$

and (3.27) follows from the last three inequalities. ∎

Lemma 3.7 *ψ_3^n of (3.4) satisfies:*

$$|||\psi_3^n|||_n \le ckh^r \|u^0\|_{r+2}. \tag{3.28}$$

Proof: By (3.2) and (1.36):

$$|||\psi_3^n|||_n \le ck \sum_{m=0}^{\mu-1} \|\partial_t[(P_E(\tau_m^n) - P_0)u(\tau_m^n)]\| \le ck \sum_{m=0}^{\mu-1} \{\|\partial_t \eta(\tau_m^n)\| + \|[I - P_0]\partial_t u(\tau_m^n)\|\}$$

and (3.28) follows with (1.19), (1.20), and (1.4). ∎

Lemma 3.8 *ψ_4^n of (3.4) satisfies:*

$$|||\psi_4^n|||_n \le ck(k^{\mu} + hk^{\mu-\frac{1}{2}} + h^2 k^{\mu-1})\|u^0\|_{2(\mu+1)}. \tag{3.29}$$

Proof: First:

$$\omega^{n+1} - \omega^n = \sum_{l=1}^{\mu} \partial_t^l \omega^n \frac{k^l}{l!} + E, \qquad E \equiv \frac{1}{\mu!} \int_{t^n}^{t^{n+1}} (t^{n+1} - t)^{\mu} \partial_t^{\mu+1} \omega(t) dt.$$

By (1.35) and (1.34):

$$\begin{aligned} kb^T \sum_{m=0}^{\mu-1} \Gamma_m^n e \partial_t \omega(\tau_m^n) &= \sum_{l=0}^{\mu-1} b^T [\sum_{m=0}^{\mu-1} \Gamma_m^n (\delta_m^n)^l] e \partial_t^{l+1} \omega^n \frac{k^{l+1}}{l!} + F = \sum_{l=0}^{\mu-1} b^T D^l e \partial_t^{l+1} \omega^n \frac{k^{l+1}}{l!} + F \\ &= \sum_{l=1}^{\mu} \partial_t^l \omega^n \frac{k^l}{l!} + F, \qquad F \equiv \frac{k}{(\mu-1)!} \sum_{m=0}^{\mu-1} b^T \Gamma_m^n e \int_{t^n}^{\tau_m^n} (\tau_m^n - t)^{\mu-1} \partial_t^{\mu+1} \omega(t) dt. \end{aligned}$$

Hence $\psi_4^n = -E + F$ and (3.29) follows after applying Lemma 3.3 to E and F, and using (1.36) to obtain inequalities of the form (3.22). ∎

With the consistency complete, it is now appropriate to discuss the development of the techniques used. First, it is possible to construct an error equation alternative to (3.4) which circumvents the constructions of section 2. However, this requires inverse properties. For example, one option involves the following replacements:

$$\bar{\omega}^n \to \sum_{l=0}^{\mu-1} D^l e \partial_t^l \omega^n \frac{k^l}{l!}, \qquad \psi_2^n \to kb^T A^{-1}[I + kA\bar{\mathcal{L}}_h^n]^{-1} A \sum_{m=0}^{\mu-1} \Gamma_m^n \mathcal{L}_h(\tau_m^n)[\bar{\omega}^n - e\omega(\tau_m^n)].$$

Then in the proof of Lemma 3.6, $\bar{v}(s)$ is changed to $\bar{\mathcal{L}}_h^n(s)\bar{\omega}^n(s)$, and bounding derivatives of the latter involves bounding products of the form $L_h^{(i)}(s)T_h^{(j)}(t)$. This can be accomplished using inverse assumptions as demonstrated by Bales [2].

Also, the original idea for overcoming the suboptimal convergence rates mentioned in connection with (1.39), was to find $q \times q$ matrices $\{D_l\}_{l=0}^{\nu-1}$ with which the following would lead to optimal convergence estimates:

$$\sum_{m=0}^{\nu-1} \Gamma_m^n (\delta_m^n)^l \equiv D_l, \qquad 0 \le l \le \nu - 1; \qquad \tilde{\mathcal{L}}_h^n \equiv \sum_{m=0}^{\nu-1} \Gamma_m^n \mathcal{L}_h(t^n + \delta_m^n k) \approx \sum_{l=0}^{\nu-1} D_l \mathcal{L}_h^{(l)}(t^n) \frac{k^l}{l!}.$$

However, attempts to prove an optimal order of consistency have repeatedly led to the following conditions for the matrices $\{D_l\}_{l=0}^{\nu-1}$:

$$D_0 e = e; \qquad D_i D_j e = D_{i+j} e, \;\; 0 \le i, j, i+j \le \nu - 1; \qquad l A D_{l-1} e = D_l e, \;\; 1 \le l \le \nu - 1.$$

Consider for example, adapting the proof of Lemma 3.5. Unfortunately, even though the number of unknowns matches the number of constraints in the equations above, it is shown in [14] that they can be solved only if $\nu \le q + 1$.

Now, (1.5) is established in the following for (1.38).

Theorem 3.1 *Under the conditions of Lemma 3.1 and either Proposition 3.1 or 3.2, $\{U_h^n\}_{n=0}^{n^*}$ are well-defined by (1.37) and (1.38), and the following holds:*

$$\max_{0 \le n \le n^*} \|U_h^n - u^n\| \le c^*(h^r + k^\mu + hk^{\mu - \frac{1}{2}} + h^2 k^{\mu-1}) \|u^0\|_\alpha. \tag{3.30}$$

Also, unless $\tilde{c} < 0$ in (3.9), c^* depends exponentially on t^*.

Proof: Set $E \equiv [(h^r + k^\mu + hk^{\mu-\frac{1}{2}} + h^2 k^{\mu-1})\|u^0\|_\alpha]^2$. Then, combining (3.9), (3.25), (3.27), (3.28) and (3.29) for (3.4):

$$|||\xi^{n+1}|||_{n+1}^2 \le (1 + \tilde{c}k) |||\xi^n|||_n^2 + c_1 k E \qquad 0 \le n \le n^* - 1.$$

After dividing this by $(1 + \tilde{c}k)^{n+1}$ and summing, the result is:

$$|||\xi^n|||_n^2 \le (1 + \tilde{c}k)^n |||\xi^0|||_0^2 + c_1 |\tilde{c}|^{-1} |1 - (1 + \tilde{c}k)^n| E \qquad 0 \le n \le n^*.$$

Now, according to (1.37), $[I + kL_h^0]\xi^0 = [P_0 - P_E^0]u^0$. So with (1.20), (1.18) and a spectral argument, it follows that:

$$|||\xi^0|||_0^2 \le c \|[P_0 - P_E^0]u^0\| \|[I + kL_h]^{-1}[P_0 - P_E^0]u^0\| \le ch^r \|u^0\|_r. \tag{3.31}$$

Then, (3.30) follows with (1.19) and the last two inequalities. ∎

4 Iterative Approximations.

In this section, the iterative scheme (1.46) is analyzed for the approximation of the solution to (1.1), and (1.5) is established. First, a brief discussion of Preconditioned Iterative Methods (PIM's) is given. See Hageman and Young [11] for more information.

Let H be any finite-dimensional Hilbert space equipped with an inner product $(\cdot, \cdot)_H$ and an associated norm $\|\cdot\|_H$. Also, let Q: $H \to H$ be H-selfadjoint and positive definite, and suppose that an approximation is required for the solution x^* to:

$$Qx^* = b. \tag{4.1}$$

Then, suppose that Q_0: $H \to H$ is H-selfadjoint and positive definite, and that solving:

$$Q_0\tilde{x} = b \tag{4.2}$$

is relatively inexpensive. Furthermore, assume that Q and Q_0 are equivalent:

$$\rho_1(Q_0x, x)_H \le (Qx, x)_H \le \rho_2(Q_0x, x)_H \qquad \forall x \in H. \tag{4.3}$$

The operator Q_0 is called the preconditioner and the PIM's of interest in this work are those with the following properties:

i. If $\{x_j\}_{j=0}^J$ are given approximations to x^* of (4.1), then calculating x_{J+1} only involves computing Qx, Q_0x, $(Qx, x)_H$, and $(Q_0x, x)_H$ for certain $x \in H$, and solving equations of the form (4.2).

ii. There is a smooth decreasing function σ: $(0,1) \to (0,1)$ such that $\sigma(1) = 0$ and if (4.3) holds, then:

$$\|Q_0^{\frac{1}{2}}[x^* - x_j]\|_H \le c[\sigma(\rho_1/\rho_2)]^j \|Q_0^{\frac{1}{2}}[x^* - x_0]\|_H. \tag{4.4}$$

For example, the Preconditioned Conjugate Gradient Method satisfies the above properties, and it is popular for having $\sigma(s) = (1 - \sqrt{s})/(1 + \sqrt{s})$ as opposed to say $(1 - s)/(1 + s)$, which is offered by various other PIM's.

Now, the rough discussion prior to (1.46) is expanded with more details. First, suppose that for $0 \le n \le n^* - 1$, the approximations $\{U_h^m\}_{m=0}^n$ have been computed using methods described below, and recall that an efficient procedure is needed for computing $\bar{U}_h^n$ defined by (1.40). Next, let $\tilde{V}_0^n$ denote an initial approximation to $\bar{U}_h^n$ given as indicated in (1.42). Now, instead of actually computing $\{\bar{V}_l^n\}_{0 \le l \le l_n}$ as suggested by (1.41), proceed as follows. Let a sequence of positive integers $\{j_n\}_{n=0}^{n^*-1}$ be specified. Then, suppose in an inductive fashion, that for $l \ge 1$, $\tilde{V}_{l-1,j_n}^n$ has been computed from j_n PIM iterations as prescribed below, and let $\tilde{V}_l^n$ be defined by the outer iteration:

$$[I + k\Lambda\mathcal{L}_h^n](S\tilde{V}_l^n) = \{SeU^n + kSA(\mathcal{L}_h^n - \bar{\mathcal{L}}_h^n)\tilde{V}_{l-1,j_n}^n\} \qquad 1 \le l \le l_n \tag{4.5}$$

with the understanding that $\tilde{V}_{0,j_n}^n \equiv \tilde{V}_0^n$. Letting n and l be fixed, (4.5) can be written in the form:

$$[I + k\lambda_i L_h^n]\psi_i = \phi_i - k\theta_i L_h^n \psi_{i-1} \qquad 1 \le i \le q \tag{4.6}$$

where $\psi_0 \equiv 0$, $\psi_i \equiv (S\tilde{V}_l^n)_i$, $1 \le i \le q$, and according to (1.35) with $l = 0$:

$$\phi_i \equiv [SeU_h^n + k\Lambda S \sum_{m=1}^{\mu-1} \Gamma_m^n(\mathcal{L}_h^n - \mathcal{L}_h^{n-m})\tilde{V}_{l-1,j_n}^n]_i. \tag{4.7}$$

The natural preconditioning for (4.6) involves $[I + kL_h^0]$ which, according to (1.16) and (1.31), is equivalent to the operators of (4.6), i. e., for $1 \le i \le q$ and $0 \le n \le n^*$:

$$\rho_1([I + kL_h^0]\chi, \chi) \le ([I + \lambda_i L_h^n]\chi, \chi) \le \rho_2([I + kL_h^0]\chi, \chi) \qquad \forall \chi \in S_h. \tag{4.8}$$

Now, to cover the case that A is not diagonalizable, define $\tilde{\psi}_i$ with:

$$[I + k\lambda_i L_h^n]\tilde{\psi}_i = \phi_i - k\theta_i L_h^n \tilde{\psi}_{i-1}^{j_n} \qquad 1 \le i \le q \tag{4.9}$$

where $\tilde{\psi}_0^{j_n} \equiv 0$. Also, to obtain $\tilde{\psi}_i^{j_n}$ for $1 \le i \le q$, set $\tilde{\psi}_i^0 \equiv (S\tilde{V}_{l-1,j_n}^n)_i$ and let iterates $\{\tilde{\psi}_i^j\}_{0 \le j \le j_n}$ be given by a PIM with preconditioner $[I + kL_h^0]$. Then as (4.4) follows from (4.3), from (4.8) it follows that:

$$|||\tilde{\psi}_i - \tilde{\psi}_i^j|||_0 \le c[\sigma(\rho_1/\rho_2)]^j |||\tilde{\psi}_i - \tilde{\psi}_i^0|||_0. \tag{4.10}$$

Finally, take $\Psi \equiv \langle \psi_1, \psi_2, \ldots, \psi_q \rangle^T$ and $\tilde{\Psi}^j \equiv \langle \tilde{\psi}_1^j, \tilde{\psi}_2^j, \ldots, \tilde{\psi}_q^j \rangle^T$ so that $\tilde{V}_l^n = S^{-1}\Psi$ and inner iterates for (4.5) are defined by:

$$\tilde{V}_{l,0}^n = \tilde{V}_{l-1,j_n}^n \qquad \tilde{V}_{l,j}^n \equiv S^{-1}\tilde{\Psi}^j, \quad 1 \le j \le j_n. \tag{4.11}$$

Now the next objective is to show that:

$$|||\bar{U}_h^n - \tilde{V}_{l,j_n}^n|||_0 \le (ck + c[\sigma(\rho_1/\rho_2)]^{j_n}) |||\bar{U}_h^n - \tilde{V}_{l-1,j_n}^n|||_0 \qquad l \ge 1. \tag{4.12}$$

Then, given some $\varepsilon_0 > 0$, j_n is chosen so that:

$$|||\bar{U}_h^n - \tilde{V}_{l,j_n}^n|||_0 \le \beta_n |||\bar{U}_h^n - \tilde{V}_{l-1,j_n}^n|||_0 \qquad 1 \le l \le l_n \tag{4.13}$$

where:

$$\beta_n^2 \le \begin{cases} ck^2, & 0 \le n \le \mu \\ \varepsilon_0 t^{n+1}, & 0 \le n \le n^* - 1. \end{cases} \tag{4.14}$$

From the last three inequalities, it follows that the integers $\{j_n\}_{n=0}^{n^*-1}$ may be chosen to satisfy (1.45) as claimed in the Introduction. First, the outer iterations (1.41) and (4.5) are shown to be well-conceived.

Lemma 4.1 *With $\bar{U}_h^n$ given by (1.40), the following holds:*

$$|||\bar{U}_h^n - \bar{V}_2|||_0 \le ck|||\bar{U}_h^n - \bar{V}_1|||_0 \tag{4.15}$$

for every $\bar{V}_1, \bar{V}_2 \in \mathbf{S}_h$ satisfying:

$$[I + kA\mathcal{L}_h^n]\bar{V}_2 = eU_h^n + kA(\mathcal{L}_h^n - \bar{\mathcal{L}}_h^n)\bar{V}_1.$$

Proof. Since:

$$\bar{U}_h^n - \bar{V}_2 = [I + kA\mathcal{L}_h^n]^{-1}(k\mathcal{L}_h^n)^{\frac{1}{2}} A(\mathcal{T}_h^n)^{\frac{1}{2}} [\mathcal{L}_h^n - \bar{\mathcal{L}}_h^n](\mathcal{T}_h^n)^{\frac{1}{2}}(k\mathcal{L}_h^n)^{\frac{1}{2}}(\bar{U}_h^n - \bar{V}_1).$$

(4.15) follows with (3.2), (2.16), and (3.6). ∎

The next lemma shows that $\{\tilde{V}_{l,j}^n\}_{j \ge 0}$ converges to $\tilde{V}_l^n$ at a rate which reflects (4.10) whether or not the right sides of (4.6) and (4.9) are the same.

Lemma 4.2 *With $\tilde{V}_l^n$, $\tilde{V}_{l,j_n}^n$, and $\tilde{V}_{l,0}^n$ given by (4.5) and (4.11):*

$$|||\tilde{V}_l^n - \tilde{V}_{l,j_n}^n|||_0 \le c[\sigma(\rho_1/\rho_2)]^{j_n} |||\tilde{V}_l^n - \tilde{V}_{l,0}^n|||_0. \tag{4.16}$$

Proof. Letting $\sigma \equiv \sigma(\rho_1/\rho_2)$, with (4.10), it follows that:

$$\begin{aligned} |||\psi_i - \tilde{\psi}_i^j|||_0 &\leq |||\psi_i - \tilde{\psi}_i|||_0 + |||\tilde{\psi}_i - \tilde{\psi}_i^j|||_0 \leq |||\psi_i - \tilde{\psi}_i|||_0 + c\sigma^j |||\tilde{\psi}_i - \tilde{\psi}_i^0|||_0 \\ &\leq (1 + c\sigma^j)|||\psi_i - \tilde{\psi}_i|||_0 + c\sigma^j |||\psi_i - \tilde{\psi}_i^0|||_0. \end{aligned}$$

Subtracting (4.9) from (4.6), with (3.6), (1.31), and a spectral argument, it follows that:

$$\begin{aligned} |||\psi_i - \tilde{\psi}_i|||_0^2 &\leq c|||[I + k\lambda_i L_h^n]^{-1} k\theta_i L_h^n(\psi_{i-1} - \tilde{\psi}_{i-1}^{j_n})|||_n^2 \\ &\leq c(\{[I + kL_h^n](kL_h^n)[I + k\lambda_i L_h^n]^{-2}\}(kL_h^n)^{\frac{1}{2}}(\psi_{i-1} - \tilde{\psi}_{i-1}^{j_n}), (kL_h^n)^{\frac{1}{2}}(\psi_{i-1} - \tilde{\psi}_{i-1}^{j_n})) \\ &\leq c|||\psi_{i-1} - \tilde{\psi}_{i-1}^{j_n}|||_0^2 \qquad 1 \leq i \leq q \end{aligned}$$

Now, the last two inequalities give:

$$|||\psi_i - \tilde{\psi}_i^{j_n}|||_0 \leq c|||\psi_{i-1} - \tilde{\psi}_{i-1}^{j_n}|||_0 + c\sigma^{j_n}|||\psi_i - \tilde{\psi}_i^0|||_0$$

from which it follows recursively that:

$$|||\psi_i - \tilde{\psi}_i^{j_n}|||_0 \leq c\sigma^{j_n} \sum_{m=1}^{i} |||\psi_m - \tilde{\psi}_m^0|||_0.$$

Then (4.16) follows after recalling (4.11) and that $\tilde{V}_l^n = S^{-1}\Psi$. ∎

Finally, (4.12) is established in the following.

Lemma 4.3 *With $\bar{U}_h^n$ defined by (1.40), the sequence $\{\tilde{V}_{l,j_n}^n\}_{l\geq 0}$ satisfies (4.12).*

Proof. Applying Lemma 4.1 to (4.5):

$$|||\bar{U}_h^n - \tilde{V}_l^n|||_0 \leq ck|||\bar{U}_h^n - \tilde{V}_{l-1,j_n}^n|||_0.$$

Using this with (4.16) and (4.11):

$$\begin{aligned} |||\tilde{V}_l^n - \tilde{V}_{l,j_n}^n|||_0 &\leq c\sigma^{j_n}|||\tilde{V}_l^n - \tilde{V}_{l,0}^n|||_0 = c\sigma^{j_n}|||\tilde{V}_l^n - \tilde{V}_{l-1,j_n}^n|||_0 \\ &\leq c\sigma^{j_n}\{|||\bar{U}_h^n - \tilde{V}_l^n|||_0 + |||\bar{U}_h^n - \tilde{V}_{l-1,j_n}^n|||_0\} \leq c\sigma^{j_n}(ck+1)|||\bar{U}_h^n - \tilde{V}_{l-1,j_n}^n|||_0. \end{aligned}$$

Now, (4.12) follows after triangulating with $\tilde{V}_l^n$. ∎

The next objective is to show that the convergence rate (1.5) can be preserved even when $\{j_n\}_{n=0}^{n^*-1}$ are chosen so that (4.13) and (4.14) hold. So additional stability and consistency arguments follow. First, define:

$$\varsigma^n \equiv U_h^n - \omega^n \qquad \text{and} \qquad \psi^n \equiv \mathcal{R}_h^n \omega^n - \omega^{n+1}$$

where ψ^n is as in (3.4). Now, according to (1.46) and (1.38):

$$U_h^{n+1} - \mathcal{R}_h^n U_h^n = b^T A^{-1}(\tilde{U}_h^n - \bar{U}_h^n),$$

so it follows that:

$$\varsigma^{n+1} = \mathcal{R}_h^n \varsigma^n + \psi^n - b^T A^{-1}(\bar{U}_h^n - \tilde{U}_h^n). \tag{4.17}$$

By (4.13) and (1.46), $(\bar{U}_h^n - \tilde{U}_h^n)$ can be estimated in terms of $(\bar{U}_h^n - \tilde{V}_0^n)$. So, the error equation (4.17) is supplemented with the following one, which is obtained from (1.40) and (1.42) after some straightforward calculations:

$$\begin{aligned}
\bar{U}_h^n - \tilde{V}_0^n \;=\; & \\
& \sum_{m=n-1-\mu_n}^{n-1} (-1)^{n-m} \binom{\mu_n+1}{n-m} [I + kA\bar{\mathcal{L}}_h^n]^{-1} kA(\bar{\mathcal{L}}_h^n - \bar{\mathcal{L}}_h^m)[I + kA\bar{\mathcal{L}}_h^m]^{-1} e\varsigma^m \\
+ \; & [I + kA\bar{\mathcal{L}}_h^n]^{-1} \sum_{m=n-1-\mu_n}^{n-1} (-1)^{n-m-1} \binom{\mu_n}{n-m-1} [\varsigma^{m+1} - \varsigma^m] \\
- \; & \sum_{m=n-1-\mu_n}^{n-1} (-1)^{n-m} \binom{\mu_n+1}{n-m} (\bar{U}_h^m - \tilde{U}_h^m) \\
- \; & \sum_{m=n-1-\mu_n}^{n-1} (-1)^{n-m} \binom{\mu_n+1}{n-m} [I + kA\bar{\mathcal{L}}_h^n]^{-1} kA(\bar{\mathcal{L}}_h^n - \bar{\mathcal{L}}_h^m)[I + kA\bar{\mathcal{L}}_h^m]^{-1} \\
& \qquad \times \{\bar{\omega}^m - e\omega^m - kA \sum_{i=0}^{\mu-1} \Gamma_i^m e\partial_t \omega(\tau_i^m)\} \\
- \; & k \sum_{m=n-1-\mu_n}^{n} (-1)^{n-m} \binom{\mu_n+1}{n-m} [I + kA\bar{\mathcal{L}}_h^m]^{-1} AP_0 \sum_{i=0}^{\mu-1} \Gamma_i^m \mathcal{L}(\tau_i^m)[\bar{u}^m - eu(\tau_i^m)] \\
- \; & k \sum_{m=n-1-\mu_n}^{n} (-1)^{n-m} \binom{\mu_n+1}{n-m} [I + kA\bar{\mathcal{L}}_h^m]^{-1} AP_0 \sum_{i=0}^{\mu-1} \Gamma_i^m \partial_t \{[P_E(\tau_i^m) - P_0]u(\tau_i^m)\} \\
+ \; & \sum_{m=n-1-\mu_n}^{n} (-1)^{n-m} \binom{\mu_n+1}{n-m} \bar{\omega}^m \\
- \; & [I + kA\bar{\mathcal{L}}_h^n]^{-1} \sum_{m=n-1-\mu_n}^{n} (-1)^{n-m} \binom{\mu_n+1}{n-m} \{\bar{\omega}^m - e\omega^m - kA \sum_{i=0}^{\mu-1} \Gamma_i^m e\partial_t \omega(\tau_i^m)\} \\
\equiv \; & \sum_{l=1}^{8} \Theta_l^n \qquad 1 \le n \le n^* - 1.
\end{aligned} \tag{4.18}$$

Before analyzing these error equations, a few adjustments must be made in Propositions 3.1 and

3.2 for the following stronger stability inequality:

$$
\begin{aligned}
|||\varsigma^{n+1}|||_{n+1}^2 \;\le\; & (1+\tilde{c}k)|||\varsigma^n|||_n^2 - c_0|||[I-r_h^n]^{\frac{1}{2}}\varsigma^n|||_n^2 + ck^{-1}|||\bar{U}_h^n - \tilde{U}_h^n|||_n^2 \\
& + ck[(h^r + k^\mu + hk^{\mu-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_\alpha]^2 \qquad 0 \le n \le n^* - 1.
\end{aligned} \tag{4.19}
$$

Proposition 4.1 *Let (1.29) be satisfied. Then there are constants $c_0 > 0$ and $\tilde{c}$, such that (4.19) holds. In fact, $\tilde{c} < 0$ if (1.27) holds and $c(1)$ of (1.13) is small enough.*

Proof: By the same manipulations leading (3.10), for (4.17) it follows that:

$$
(1-c_3k)|||\varsigma^{n+1}|||_{n+1}^2 \le |||r_h^n\varsigma^n|||_n^2 + \varepsilon^{-1}c_1k\|\varsigma^n\|^2 + ck^{-1}|||\bar{U}_h^n - \tilde{U}_h^n|||_n^2 + ck^{-1}|||\psi^n|||_n^2.
$$

By (1.26), $I - r_h^n$ has a square root. Hence, taking $\chi^n \equiv [I + kL_h^n]^{\frac{1}{2}}\varsigma^n$, with (1.29) it follows that:

$$
|||r_h^n\varsigma^n|||_n^2 = \|\chi^n\|^2 - ([I+r_h^n][I-r_h^n]^{\frac{1}{2}}\chi^n, [I-r_h^n]^{\frac{1}{2}}\chi^n) \le |||\varsigma^n|||_n^2 - \delta([I-r_h^n]\chi^n, \chi^n).
$$

Using (3.12), there is a $\hat{c} \le 0$ such that:

$$
-([I-r_h^n]\chi^n, \chi^n) \le \tfrac{1}{2}(r_h^n\chi^n, \chi^n) + \tfrac{1}{2}(1+\hat{c}k)\|\chi^n\|^2 - \|\chi^n\|^2 = \tfrac{1}{2}\hat{c}k|||\varsigma^n|||_n^2 - \tfrac{1}{2}([I-r_h^n]\chi^n, \chi^n)
$$

where $\hat{c} < 0$ if (1.27) holds. Now with $\tilde{c} > \frac{1}{2}\delta\hat{c} + c_3 + \varepsilon^{-1}c_1$, and $c_0 = \frac{1}{2}\delta$, (4.19) follows after combining (3.25), (3.27), (3.28), (3.29) and the last three inequalities. ∎

As with Proposition 3.2, $\tilde{c} < 0$ is guaranteed for (4.19) by the following.

Proposition 4.2 *Let (1.27) and (1.28) be satisfied. Then, there are constants $c_0 > 0$ and $\tilde{c} < 0$ such that (4.19) holds.*

Proof: In the proof of Proposition 3.2, replace ξ^n with ς^n, and ψ^n with $\psi^n - b^TA^{-1}(\bar{U}_h^n - \tilde{U}_h^n)$. Then with $\varepsilon \equiv \frac{1}{2}(c_2 - c_1)$, redefine $\theta \equiv (1 - c_1 - 2\varepsilon)/(1 + c_3k_0)$ so that the last part of the proof is readily changed to give the following instead of (3.13):

$$
\|\mathcal{R}_h^n\varsigma^n\|^2 + (1+\varepsilon_1)\|(kL_h^{n+1})^{\frac{1}{2}}\mathcal{R}_h^n\varsigma^n\|^2 \le (1+\tilde{c}k)|||\varsigma^n|||_n^2 - \varepsilon\|(kL_h^n)^{\frac{1}{2}}\varsigma^n\|^2.
$$

As with (3.14), it follows that:

$$
|||\varsigma^{n+1}|||_{n+1}^2 \;\le\; (1+\tilde{c}k)|||\varsigma^n|||_n^2 - \varepsilon(kL_h^n\varsigma^n, \varsigma^n) + ck^{-1}|||\bar{U}_h^n - \tilde{U}_h^n|||_n^2 + ck^{-1}|||\psi^n|||_n^2.
$$

Next, since $\nu \ge 1$, $r(0) = 1 = -r'(0)$. So, there is a $c_0 > 0$ such that for all $z \ge 0$, $r(z)$ is greater that the linear function $1 - \frac{\varepsilon}{2c_0}z$, i. e., $-\varepsilon z \le -2c_0[1 - r(z)]$, $\forall z \ge 0$. Also, using (1.26), c_0 can be assumed small enough that $-\varepsilon \le -2c_0[1 - r(z)]$, $\forall z \ge 0$. After multiplying the latter by z and adding the result to former, it follows that:

$$
-\varepsilon(kL_h\chi, \chi) \le -c_0([I + kL_h^n][I - r_h^n]\chi, \chi) \qquad \forall \chi \in S_h.
$$

Hence, (4.19) follows with (3.25), (3.27), (3.28) and (3.29) and the last two inequalities. ∎

Now, estimation of the terms of (4.18) begins.

Lemma 4.4 $\{\Theta_l^n\}_{l=1}^3$ *of (4.18) satisfy:*

$$(4.20)\quad \sum_{l=1}^{3} |||\Theta_l^n|||_n \le c \sum_{m=n-1-\mu_n}^{n-1} \{|||\varsigma^{m+1}-\varsigma^m|||_m + |||\bar{U}_h^m - \tilde{U}_h^m|||_m\} + ck \sum_{m=n-1-\mu_n}^{n-1} |||\varsigma^m|||_m.$$

Proof: First, note that:

$$\Theta_1^n = \sum_{m=n-1-\mu_n}^{n-1} (-1)^{n-m} \begin{pmatrix} \mu_n+1 \\ n-m \end{pmatrix} [I + kA\bar{\mathcal{L}}_h^n]^{-1}(k\mathcal{L}_h^n)^{\frac{1}{2}} A$$

$$\times(\mathcal{T}_h^n)^{\frac{1}{2}}\{(\bar{\mathcal{L}}_h^n - \mathcal{L}_h^n) + (\mathcal{L}_h^n - \mathcal{L}_h^m) + (\mathcal{L}_h^m - \bar{\mathcal{L}}_h^m)\}(\mathcal{T}_h^n)^{\frac{1}{2}}[(\mathcal{L}_h^n)^{\frac{1}{2}}(\mathcal{T}_h^m)^{\frac{1}{2}}](k\mathcal{L}_h^m)^{\frac{1}{2}}[I + kA\bar{\mathcal{L}}_h^m]^{-1} e\varsigma^m.$$

So, Θ_1^n is estimated using (3.2), (2.16), (2.15), and (1.15). Also, estimates for Θ_2^n and Θ_3^n follow with (3.2) and (3.6). ∎

Lemma 4.5 $\{\Theta_l^n\}_{l=4}^6$ *of (4.18) satisfy:*

$$(4.21)\quad \sum_{l=4}^{6} |||\Theta_l^n|||_n \le ck(h^r + k^\mu)\|u^0\|_\alpha.$$

Proof: Recall E and F defined in the proof of Lemma 3.5. By (3.18), and (3.23):

$$\|E\| + \|F\| \le ck^\mu \|u^0\|_{2(\mu+1)}.$$

So Θ_4^n can be estimated using the techniques applied for the estimation of Θ_1^n. Also, the same manipulations used to prove Lemmas 3.6 and 3.7 give corresponding estimates for Θ_5^n and Θ_6^n. ∎

Lemma 4.6 Θ_7^n *of (4.18) satisfies:*

$$(4.22)\quad |||\Theta_7^n|||_n \le ck(k^{\mu_n} + hk^{\mu-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_{2(\mu+1)}.$$

Proof: By (3.26), for $n-1-\mu_n \le m \le n-1$:

$$\bar{\omega}^m \equiv E^m + F^m, \qquad F^m \equiv \frac{1}{(\mu-1)!}\int_0^k (k-s)^{\mu-1}\partial_s^\mu \bar{\omega}^m(s)ds$$

and taking $(m-n)^i|_{m-n=i=0} \equiv 1$:

$$E^m \equiv \sum_{l=0}^{\mu-1} D^l e \frac{k^l}{l!}\{\sum_{i=0}^{\mu-l}(m-n)^i \partial_t^{l+i}\omega^n \frac{k^i}{i!} + G_l^m\}, \qquad G_l^m \equiv \frac{1}{(\mu-l)!}\int_{t^n}^{t^m} (t^m - t)^{\mu-l}\partial_t^{\mu+1}\omega(t)dt.$$

Next, since:

$$\gamma_i^n \equiv \sum_{m=n-1-\mu_n}^{n} (-1)^{n-m} \begin{pmatrix} \mu_n+1 \\ n-m \end{pmatrix} (m-n)^i = 0 \qquad 0 \le i \le \mu_n$$

it follows after some re-indexing that:

$$\sum_{m=n-1-\mu_n}^{n}(-1)^{n-m}\begin{pmatrix}\mu_n+1\\ n-m\end{pmatrix}E^m=$$
$$\sum_{l=0}^{\mu-1}\frac{D^l e}{l!}\{\sum_{j=l+\mu_n+1}^{\mu}\frac{\gamma_{j-l}^n}{(j-l)!}\partial_t^j\omega^n k^j+\sum_{m=n-1-\mu_n}^{n}(-1)^{n-m}\begin{pmatrix}\mu_n+1\\ n-m\end{pmatrix}k^l G_l^m\}$$

where ill-defined sums are understood to be zero. Now, by Lemma 3.3:

$$k^j|||\partial_t^j\omega^n|||_n\le ck^{\mu_n+1}\|u^0\|_{2(\mu+1)}\qquad \mu_n+1\le j\le\mu$$

$$k^l|||G_l^m|||_n\le ck(k^\mu+hk^{\mu-\frac{1}{2}}+h^2k^{\mu-1})\|u^0\|_{2(\mu+1)}\qquad 0\le l\le\mu-1.$$

Hence:

$$|||\sum_{m=n-1-\mu_n}^{n}(-1)^{n-m}\begin{pmatrix}\mu_n+1\\ n-m\end{pmatrix}E^m|||_n\le ck(k^{\mu_n}+hk^{\mu-\frac{1}{2}}+h^2k^{\mu-1})\|u^0\|_{2(\mu+1)}.$$

Also, by (3.24):

$$|||F^m|||_n\le ck(hk^{\mu-\frac{1}{2}}+h^2k^{\mu-1})\|u^0\|_{2(\mu+1)}\qquad n-1-\mu_n\le m\le n.$$

Then (4.22) follows from the last two inequalities. ∎

Lemma 4.7 *Θ_8^n of (4.18) satisfies:*

(4.23) $$|||\Theta_8^n|||_n\le ck(k^\mu+h^2k^{\mu-1})\|u^0\|_{2(\mu+1)}.$$

Proof: As in the proof of Lemma 3.5:

$$\bar\omega^m-e\omega^m-kA\sum_{i=0}^{\mu-1}\Gamma_i^m e\partial_t\omega(\tau_i^m)=E^m-F^m,\qquad E^m\equiv\frac{1}{(\mu-1)!}\int_0^k(k-s)^{\mu-1}\partial_s^\mu\bar\omega^m(s)ds$$

and by (1.35):

$$F^m\equiv\frac{k^\mu}{(\mu-1)!}AD^{\mu-1}e\partial_t^\mu\omega^m+\frac{k}{(\mu-1)!}\sum_{i=0}^{\mu-1}A\Gamma_i^m e\int_{t^m}^{\tau_i^m}(\tau_i^m-t)^{\mu-1}\partial_t^{\mu+1}\omega(t)dt\equiv F_1^m+F_2^m.$$

First, using (3.2):

$$|||\Theta_8^n|||_n\le c\sum_{m=n-1-\mu_n}^{n}(\|E^m\|+\|F_2^m\|)+c\|\sum_{m=n-1-\mu_n}^{n}(-1)^{n-m}\begin{pmatrix}\mu_n+1\\ n-m\end{pmatrix}F_1^m\|.$$

Then, with (3.24):

$$\|E^m\|\le ch^2k^{\bar\mu}\|u^0\|_{2\mu}\qquad n-1-\mu_n\le m\le n.$$

Also, by Lemma 3.3:

$$\|F_2^m\|\le ck(k^\mu+h^2k^{\mu-1})\|u^0\|_{2(\mu+1)}\qquad n-1-\mu_n\le m\le n.$$

Finally, note that:

$$\sum_{m=n-1-\mu_n}^{n}(-1)^{n-m}\binom{\mu_n+1}{n-m}F_1^m=\frac{k^\mu}{(\mu-1)!}AD^{\mu-1}e$$
$$\times\{\sum_{m=n-\mu_n}^{n}(-1)^{n-m}\binom{\mu_n}{n-m}\int_{t^{m-1}}^{t^m}\partial_t^{\mu+1}u(t)dt-\sum_{m=n-1-\mu_n}^{n}(-1)^{n-m}\binom{\mu_n+1}{n-m}\partial_t^\mu\eta^m\}.$$

So, by (1.4) and (1.19):

$$\|\sum_{m=n-1-\mu_n}^{n}(-1)^{n-m}\binom{\mu_n+1}{n-m}F_1^m\|\le ck(h^2k^{\mu-1}+k^\mu)\|u^0\|_{2(\mu+1)}.$$

Now, (4.23) follows after combining the last four inequalities. ∎

Next, the above lemmas are combined with (4.13) for the estimation of the term $k^{-1}|||\bar{U}_h^n-\tilde{U}_h^n|||_n^2$, in (4.19).

Proposition 4.3 *With $\bar{U}_h^n$ defined by (1.40) and $\tilde{U}_h^n$ by (1.46), the following hold:*

$$k^{-1}|||\bar{U}_h^n-\tilde{U}_h^n|||_n^2\le ck^{-1}\beta_n^2\sum_{m=n-1-\mu_n}^{n-1}\{|||\bar{U}_h^m-\tilde{U}_h^m|||_m^2+|||\varsigma^{m+1}-\varsigma^m|||_m^2\}$$
$$+ck\beta_n^2\sum_{m=n-1-\mu_n}^{n-1}|||\varsigma^m|||_m^2+ck[(h^r+k^\mu+hk^{\mu-\frac{1}{2}}+h^2k^{\mu-1})\|u^0\|_\alpha]^2\qquad 1\le n\le n^*-1, \tag{4.24}$$

$$|||\bar{U}_h^0-\tilde{U}_h^0|||_0\le ck^\mu\|u^0\|_\alpha. \tag{4.25}$$

Proof. By (1.46), (4.13) and (3.6):

$$|||\bar{U}_h^n-\tilde{U}_h^n|||_n\le c\beta_n^{l_n}|||\bar{U}_h^n-\tilde{V}_0^n|||_n\qquad 0\le n\le n^*-1.$$

Then, combining (4.20), (4.21), (4.22), and (4.23) for (4.18) leads to:

$$|||\bar{U}_h^n-\tilde{V}_0^n|||_n\le\sum_{m=n-1-\mu_n}^{n-1}\{|||\bar{U}_h^m-\tilde{U}_h^m|||_m+|||\varsigma^{m+1}-\varsigma^m|||_m\}+ck\sum_{m=n-1-\mu_n}^{n-1}|||\varsigma^m|||_m$$
$$+ck(h^r+k^{\mu_n}+hk^{\mu-\frac{1}{2}}+h^2k^{\mu-1})\|u^0\|_\alpha\qquad 1\le n\le n^*-1.$$

From (4.14) and (1.43), it follows that:

$$\beta_n^{l_n}k^{\mu_n}\le ck^\mu\qquad 0\le n\le n^*-1.$$

Finally, (4.24) follows after combining the last three inequalities. Also, by (3.2), (3.31), and (3.18):

$$|||\bar{U}_h^0-\tilde{V}_0^0|||_0\le c|||U_h^0|||_0\le c|||\varsigma^0|||_0+c|||\omega^0|||_0\le c\|u^0\|_\alpha$$

and (4.25) follows after combining this with the two estimates above including the case $n=0$. ∎

Now, (4.24) demands an estimation of the differences $|||\varsigma^{n+1}-\varsigma^n|||_n$ and this is the content of the following.

Proposition 4.4 *If (1.29) and (1.30) are satisfied, then the following holds:*

$$c_1|||\varsigma^{n+1}-\varsigma^n|||_n^2+|||[I-r_h^{n+1}]^{\frac{1}{2}}\varsigma^{n+1}|||_{n+1}^2\le c_2k^2|||\varsigma^n|||_n^2+c|||\bar U_h^n-\tilde U_h^n|||_n^2$$

(4.26)

$$+(1+c_3k)|||[I-r_h^n]^{\frac{1}{2}}\varsigma^n|||_n^2+ck^2[(h^r+k^\mu+hk^{\mu-\frac{1}{2}}+h^2k^{\mu-1})\|u^0\|_\alpha]^2\qquad 0\le n\le n^*-1.$$

Proof: The following is established after some straightforward calculations:

$$([I+kL_h^{n+1}][I+r_h^{n+1}][\varsigma^{n+1}-\varsigma^n],[\varsigma^{n+1}-\varsigma^n])+([I+kL_h^{n+1}][I-r_h^{n+1}]\varsigma^{n+1},\varsigma^{n+1})$$

(4.27)

$$=2([I+kL_h^{n+1}][\varsigma^{n+1}-r_h^{n+1}\varsigma^n],[\varsigma^{n+1}-\varsigma^n])+([I+kL_h^{n+1}][I-r_h^{n+1}]\varsigma^n,\varsigma^n).$$

By (1.29) and (3.6), there is a $c_1>0$ such that:

$$([I+r_h^{n+1}][I+kL_h^{n+1}]^{\frac{1}{2}}[\varsigma^{n+1}-\varsigma^n],[I+kL_h^{n+1}]^{\frac{1}{2}}[\varsigma^{n+1}-\varsigma^n])\ge 2c_1|||\varsigma^{n+1}-\varsigma^n|||_n^2. \tag{4.28}$$

Next, using (3.6):

$$2|([I+kL_h^{n+1}][\varsigma^{n+1}-r_h^{n+1}\varsigma^n],[\varsigma^{n+1}-\varsigma^n])|\le c|||\varsigma^{n+1}-r_h^{n+1}\varsigma^n|||_n|||\varsigma^{n+1}-\varsigma^n|||_n$$

$$\le c|||\varsigma^{n+1}-\mathcal{R}_h^n\varsigma^n|||_n^2+c|||[\mathcal{R}_h^n-r_h^n]\varsigma^n|||_n^2+c|||[r_h^{n+1}-r_h^n]\varsigma^n|||_n^2+c_1|||\varsigma^{n+1}-\varsigma^n|||_n^2.$$

Combining (3.25), (3.27), (3.28) and (3.29) for (4.17), it follows that:

$$|||\varsigma^{n+1}-\mathcal{R}_h^n\varsigma^n|||_n\le ck(h^r+k^\mu+hk^{\mu-\frac{1}{2}}+h^2k^{\mu-1})\|u^0\|_\alpha+c|||\bar U_h^n-\tilde U_h^n|||_n.$$

Combining the last two inequalities with (3.7) and (3.8), the result is:

$$2|([I+kL_h^{n+1}][\varsigma^{n+1}-r_h^{n+1}\varsigma^n],[\varsigma^{n+1}-\varsigma^n])|\le ck^2|||\varsigma^n|||_n^2+c|||\bar U_h^n-\tilde U_h^n|||_n^2$$

(4.29)

$$+ck^2[(h^r+k^\mu+hk^{\mu-\frac{1}{2}}+h^2k^{\mu-1})\|u^0\|_\alpha]^2+c_1|||\varsigma^{n+1}-\varsigma^n|||_n^2.$$

For the last term in (4.27), suppose first that (1.27) holds. After some calculations:

$$(\{[I+kL_h^{n+1}][I-r_h^{n+1}]-[I+kL_h^n][I-r_h^n]\}\varsigma^n,\varsigma^n)=(E(kL_h^n)^{\frac{1}{2}}\varsigma^n,(kL_h^n)^{\frac{1}{2}}\varsigma^n) \tag{4.30}$$

where:

$$E\equiv(kL_h^n)^{-\frac{1}{2}}[r_h^n-r_h^{n+1}](kL_h^n)^{-\frac{1}{2}}+[(T_h^n)^{\frac{1}{2}}L_h^{n+1}(T_h^n)^{\frac{1}{2}}](kL_h^n)^{\frac{1}{2}}[r_h^n-r_h^{n+1}](kL_h^n)^{-\frac{1}{2}}$$

$$+(T_h^n)^{\frac{1}{2}}[L_h^{n+1}-L_h^n](T_h^n)^{\frac{1}{2}}[I-r_h^n].$$

By (3.8), (2.15), (1.13), and (1.26):

$$\|E\chi\|\le ck\|\chi\|\qquad\forall\chi\in S_h. \tag{4.31}$$

Next, define $r_\epsilon(z)\equiv[1+z]/[1+(1-\epsilon)z]$ with $\epsilon>0$ small enough that with (3.11) and (1.27):

$$|r_\epsilon(z)r(z)|\le\left\{\begin{array}{ll}(1+\epsilon z)(1-\theta z) & 0<z\le z_1\\ (1-\epsilon)^{-1}\sup\limits_{z\ge z_1}|r(z)| & z_1\le z\end{array}\right\}\le 1.$$

Hence:

$$\varepsilon(kL_h^n\chi,\chi) \le ([I+kL_h^n][I-r_h^n]\chi,\chi) \qquad \forall \chi \in S_h.$$

Combining this with (4.30) and (4.31):

$$|(\{[I+kL_h^{n+1}][I-r_h^{n+1}]-[I+kL_h^n][I-r_h^n]\}\varsigma^n,\varsigma^n)| \le c\varepsilon^{-1}k([I+kL_h^n][I-r_h^n]\varsigma^n,\varsigma^n). \tag{4.32}$$

Finally, for the case that (1.27) holds, (4.26) follows after combining (4.28), (4.29), and (4.32) for (4.27). Now, assume that (1.27) fails, so that $r(\infty) \equiv 1 - b^TA^{-1}e = 1$. Nevertheless, by (1.30) and (1.11), $[I-r_h^n]$ is positive and invertible, so define:

$$\tilde{r}_n \equiv [I+kL_h^n][I-r_h^n] \qquad \text{and} \qquad F^2 \equiv (kL_h^n)[I+kL_h^n]^{-2}\tilde{r}_n^{-1}.$$

Then, instead of (4.30), the following is used:

$$\begin{aligned}&(\{[I+kL_h^{n+1}][I-r_h^{n+1}]-[I+kL_h^n][I-r_h^n]\}\varsigma^n,\varsigma^n) = \\ &(F[(kL_h^n)^{-\frac{1}{2}}+(kL_h^n)^{\frac{1}{2}}][\tilde{r}_{n+1}-\tilde{r}_n][(kL_h^n)^{-\frac{1}{2}}+(kL_h^n)^{\frac{1}{2}}]F\tilde{r}_n^{\frac{1}{2}}\varsigma^n,\tilde{r}_n^{\frac{1}{2}}\varsigma^n)\end{aligned} \tag{4.33}$$

From (1.30), it follows that:

$$\|F\chi\| \le c\|\chi\| \qquad \forall \chi \in S_h. \tag{4.34}$$

Next, since $b^TA^{-1}e = 1 - r(\infty) = 0$:

$$r_h^n = I - b^TA^{-1}\{(kA\mathcal{L}_h^n)[I+kA\mathcal{L}_h^n]^{-1}\}e = I + b^TA^{-1}[I+kA\mathcal{L}_h^n]^{-1}e$$

and hence:

$$\tilde{r}_n = -b^TA^{-2}e + b^TA^{-1}(A^{-1}-I)[I+kA\mathcal{L}_h^n]^{-1}e.$$

Therefore:

$$\tilde{r}_{n+1}-\tilde{r}_n = b^T(I-A^{-1})[I+kA\mathcal{L}_h^{n+1}]^{-1}(k\mathcal{L}_h^n)^{\frac{1}{2}}[(\mathcal{T}_h^n)^{\frac{1}{2}}(\mathcal{L}_h^{n+1}-\mathcal{L}_h^n)(\mathcal{T}_h^n)^{\frac{1}{2}}](k\mathcal{L}_h^n)^{\frac{1}{2}}[I+kA\mathcal{L}_h^n]^{-1}e.$$

So by (3.3) and (2.15), for $\theta_1, \theta_2 = \pm\frac{1}{2}$:

$$\|(kL_h^n)^{\theta_1}[\tilde{r}_{n+1}-\tilde{r}_n](kL_h^n)^{\theta_2}\chi\| \le ck\|\chi\| \qquad \forall \chi \in S_h.$$

Combining this with (4.34) for (4.33) gives (4.32) and hence (4.26). ∎

Finally, the convergence result (1.5) is established for (1.46) as follows.

Theorem 4.1 *Let the conditions of either Proposition 4.1 or 4.2, in addition to those of Lemma 3.1 and Proposition 4.4 be satisfied. Then, $\{j_n\}_{n=0}^{n^*-1}$ can be chosen so that (4.13) and (4.14) hold and provided $\varepsilon_0 > 0$ is small enough, the approximations $\{U_h^n\}_{n=0}^{n^*}$ obtained by (1.37) and (1.46) satisfy:*

$$\max_{0\le n\le n^*}\|U_h^n - u^n\| \le c^*(h^r + k^\mu + hk^{\mu-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_\alpha. \tag{4.35}$$

Also, unless $\tilde{c} < 0$ in (4.19), c^* depends exponentially on t^*.

Proof. Add $c_1|||\bar{U}_h^n - \tilde{U}_h^n|||_n^2$ to both sides of (4.26) and multiply the resulting inequality by $\varepsilon k^{-1}t^{n+1}$. When this is added to (4.19), the result is:

$$|||\varsigma^{n+1}|||_{n+1}^2 + c_1\varepsilon k^{-1}t^{n+1}\{|||\varsigma^{n+1} - \varsigma^n|||_n^2 + |||\bar{U}_h^n - \tilde{U}_h^n|||_n^2\} + \varepsilon k^{-1}t^{n+1}|||[I - r_h^{n+1}]^{\frac{1}{2}}\varsigma^{n+1}|||_{n+1}^2$$

$$\leq (1 + \tilde{c}k + c_2\varepsilon k t^{n+1})|||\varsigma^n|||_n^2 + ck^{-1}|||\bar{U}_h^n - \tilde{U}_h^n|||_n^2 + ck[(h^r + k^\mu + hk^{\mu-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_\alpha]^2$$

$$+[\varepsilon k^{-1}t^{n+1}(1 + c_3k) - c_0]|||[I - r_h^n]^{\frac{1}{2}}\varsigma^n|||_n^2 \qquad 1 \leq n \leq n^* - 1.$$

Now, for the compression of this inequality and others below, let the following be defined:

$$Z^n \equiv |||\varsigma^n|||_n^2, \qquad D^{n+1} \equiv |||\varsigma^{n+1} - \varsigma^n|||_n^2 + |||\bar{U}_h^n - \tilde{U}_h^n|||_n^2,$$

$$S^n \equiv |||[I - r_h^n]^{\frac{1}{2}}\varsigma^n|||_n^2, \qquad E \equiv [(h^r + k^\mu + hk^{\mu-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_\alpha]^2.$$

With this notation, the following results after estimating $ck^{-1}|||\bar{U}_h^n - \tilde{U}_h^n|||_n^2$ with (4.24):

$$Z^{l+1} + c_1k^{-1}\varepsilon t^{l+1}D^{l+1} + \varepsilon k^{-1}t^{l+1}S^{l+1} \leq (1 + \tilde{c}k + c_2\varepsilon k t^*)Z^l$$
$$+c_4\beta_l^2 \sum_{m=l-1-\mu_l}^{l-1} [k^{-1}D^{m+1} + kZ^m] + ckE + [\varepsilon k^{-1}t^l + \varepsilon(1 + c_3t^*) - c_0]S^l \qquad 1 \leq l \leq n^* - 1.$$

Now, assume that $\varepsilon > 0$ is chosen small enough that $\varepsilon(1 + c_3t^*) \leq c_0$. In fact, if $\tilde{c} < 0$, suppose that for some $\hat{c} < 0$, $1 + \tilde{c}k + c_2\varepsilon k t^* \leq 1 + \hat{c}k$. Otherwise, if $\tilde{c} \geq 0$, take $\hat{c} > 0$ in the following. Now, after summing the last inequality over $1 \leq l \leq n \leq n^* - 1$, the result is:

$$(Z^{n+1} - Z^1) + \varepsilon k^{-1}(t^{n+1}S^{n+1} - t^1S^1) + c_1\varepsilon k^{-1}\sum_{l=1}^n t^{l+1}D^{l+1} \leq \hat{c}k\sum_{l=1}^n Z^l$$
$$+c_4\sum_{l=1}^n \beta_l^2 \sum_{m=l-1-\mu_l}^{l-1} [kZ^m + k^{-1}D^{m+1}] + ct^*E \qquad 1 \leq n \leq n^* - 1.$$

By (4.14) and (3.31), for $1 \leq n \leq n^* - 1$:

$$\hat{c}k\sum_{l=1}^n Z^l + c_4k\sum_{l=1}^n \beta_l^2 \sum_{m=l-1-\mu_l}^{l-1} Z^m \leq (\hat{c} + c_4\varepsilon_0(\mu+1)t^*)k\sum_{l=1}^n Z^l + ckZ^0 \leq \bar{c}k\sum_{l=1}^n Z^l + ckE$$

where $\bar{c} \leq 0$ if $\hat{c} < 0$ and $\varepsilon_0 > 0$ is small enough. Otherwise, take $\bar{c} > 0$ in the following. Next, since $(l+1)/(m+1) \leq \mu + 2$ if $0 \leq l - 1 - \mu_l \leq m \leq l - 1$, it follows using (4.14) that for $1 \leq n \leq n^* - 1$:

$$c_4k^{-1}\sum_{l=1}^n \beta_l^2 \sum_{m=l-1-\mu_l}^{l-1} D^{m+1} \leq c_4k^{-1}\varepsilon_0 \sum_{l=1}^n \sum_{m=l-1-\mu_l}^{l-1} \frac{l+1}{m+1} t^{m+1}D^{m+1} \leq c_5\varepsilon_0k^{-1}\sum_{l=1}^n t^{l+1}D^{l+1} + cD^1.$$

Combining the last three inequalities, for $1 \leq n \leq n^* - 1$:

$$Z^{n+1} + \varepsilon k^{-1}t^{n+1}S^{n+1} + (c_1\varepsilon - c_5\varepsilon_0)k^{-1}\sum_{l=1}^n t^{l+1}D^{l+1} \leq (Z^1 + \varepsilon S^1 + cD^1) + ct^*E + \bar{c}k\sum_{l=1}^n Z^l.$$

By (3.6), (4.26), (4.25), (1.26) and (3.31):

$$Z^1 + [S^1 + D^1] \leq [cD^1 + cZ^0] + [S^1 + D^1] \leq c(1+k^2)[Z^0 + E] + (1+ck)S^0 \leq c(Z^0 + E) \leq cE.$$

Now, assume that $\epsilon_0 > 0$ is chosen small enough so that:

$$|||\varsigma^{n+1}|||^2_{n+1} \leq ct^*[(h^r + k^\mu + hk^{\mu-\frac{1}{2}} + h^2k^{\mu-1})\|u^0\|_\alpha]^2 + \bar{c}k\sum_{l=0}^{n}|||\varsigma^l|||^2_l \qquad 0 \leq n \leq n^* - 1.$$

If $\bar{c} \leq 0$, ignore the last sum and (4.35) follows after (1.19). If $\bar{c} > 0$, then (4.35) follows with the discrete Gronwall Lemma and (3.31), but with c^* depending exponentially on t^*. ∎

5 Examples.

The principal aim of this section is to present some computational results showing the strength of methods analyzed in this work. However, it is appropriate to first indicate that the set of IRKM's which satisfy the many conditions imposed in foregoing proofs, is by no means vacuous. For example, in [15], it is explained that there exist q-stage methods of order $q+1$ and satisfying (1.26)-(1.32) and (1.44), provided $q = 1, 2, 3$, or 5. Furthermore, [15] gives explicit constructions of families of such methods for $q = 2$ and 3. On the other hand, it is shown in [15], that for every positive integer q, there exists a collocation type IRKM satisfying (1.27), (1.29)-(1.32), and (1.44).

As mentioned in the Introduction and more carefully in [15], the preferred methods in a parallel environment are those for which the eigenvalues of A are distinct. These have been referred to as ***multiply implicit*** (MIRK) methods. Further, they are called real if $\sigma(A) \subset \mathbf{R}$, and otherwise complex. While the latter case has not been studied here, it is discussed in [15]. By considering that discussion together with the results of Bramble and Sammon [3], it can be seen that complex MIRK's can be analyzed using quadratic preconditioning and hence inverse assumptions.

In contrast to MIRK's, there are the well-known methods for which the eigenvalues of A are identical and real. [12] As seen in (4.9), these so-called ***singly implicit*** (SIRK) methods offer a computational advantage on serial machines since at each time step, they require the formation of only a single new matrix with the dimension of S_h. A selection from this set of methods was made for the example considered below.

The following problem is of the class defined in the Introduction:

$$\begin{cases} \partial_t u &= -L(t)u & \text{in} & (-1,1)\times[0,.1] \\ u &= 0 & \text{on} & \{-1,1\}\times[0,.1] \\ u(x,0) &= 1-x^2 & \text{in} & (-1,1) \end{cases}$$

where:

$$L(t)u \equiv -\partial_x(\ell_1(x,t)\partial_x u) + \ell_0(x,t)u,$$

$$\ell_1(x,t) \equiv \frac{\frac{1}{18}\log(2)(3-x^2)}{(2+x^2)+t(1-x^2)}(2+x^2)^{t+1}, \qquad \ell_0(x,t) \equiv \log(2+x^2) - \tfrac{1}{3}\log(2)(2+x^2)^t.$$

The solution is given by:

$$u(x,t) = \frac{1-x^2}{(2+x^2)^t}.$$

k, h	CPU Time (sec)	L_2 error ($\times 10^9$)	Order
1/50	22	1.19	
1/60	30	.525	4.49
1/70	38	.266	4.42
1/80	48	.148	4.37
1/90	59	.0889	4.33
1/100	72	.0565	4.30

Table 1: Modified method

k, h	CPU Time (sec)	L_2 error ($\times 10^9$)	Order
1/50	22	28.5	
1/60	31	16.0	3.16
1/70	41	9.80	3.19
1/80	52	6.36	3.23
1/90	65	4.35	3.24
1/100	77	3.09	3.24

Table 2: Classical method

For the spatial discretization, the Ordinary Galerkin Method was used and S_h was constructed of smooth cubic splines defined on a uniform mesh. For the temporal discretization, the well-known three-stage *diagonally implicit* (DIRK) method was used as it satisfies (1.26)-(1.32). [8]

Now let (1.46) be identified as the *modified* method, and an analogue based on (1.39) as the *classical* method. In addition, let a *hybrid* method be given by (1.46), but with D^l replaced by T^l in (1.35). These three methods were tested on the ICASE SUN 3/180. Defining $E(h,k) \equiv \|U_h^{n^*} - u^{n^*}\|$, the L_2 errors $E(k) \equiv E(k,k)$ are reported in Tables 1 - 3, together with estimates of the convergence order obtained according to the formula: $\log(E(k_2)/E(k_1))/\log(k_2/k_1)$.

With regard to time consumption, recall that the computational burden for the classical method

k, h	CPU Time (sec)	L_2 error ($\times 10^9$)	Order
1/50	23	28.3	
1/60	30	15.8	3.21
1/70	38	9.61	3.21
1/80	49	6.25	3.22
1/90	59	4.28	3.22
1/100	71	3.04	3.23

Table 3: Hybrid method

is in forming q new stiffness matrices at each time step. On the other hand, with the constants $\{\delta_m^n\}_{0\le m\le \mu-1}^{0\le n\le n^*-1}$ chosen in the natural way as indicated in the Introduction, the burden for the modified method is in forming the terms ϕ_i of (4.7), for the right side of (4.9). Also, the initial steps are relatively expensive, but the effect of this diminishes as the number of time steps increases. Note that among the three methods tested, numbers for the modified method were obtained with greater speed and accuracy, as well as with fourth order convergence. On the other hand, the others suffer from suboptimal convergence as explained in the Introduction. However, no rigorous explanation can be offered for the identical accuracy obtained by the classical and hybrid methods. Further, this author is unaware of any proof of the better than second order convergence seen in Tables 2 and 3. In this connection, note that the above solution has no time derivatives which are even in the domain of $L(t)^2$, a condition considered necessary to escape order reduction in a general way. Nevertheless, only second order convergence is demonstrated for example, in Experiment 7.5.1 of Dekker and Verwer [8], where a stiff ordinary differential equation is considered. Further, the modified method has been applied to this problem to give not only fourth order convergence, but accuracy exceeding that reported for any method discussed in the Experiment.

References

[1] ADAMS, R. A., *Sobolev Spaces*, Academic Press, New York, London, Toronto, Sydney, San Francisco, 1975.

[2] BALES, L. A., *Semidiscrete and Single Step Fully Discrete Approximations for Second Order Hyperbolic Equations with Time-Dependent Coefficients*, Math. Comp., v. 43, 1984, pp. 383-414.

[3] BRAMBLE, J. H., AND SAMMON, P. H., *Efficient Higher Order Single Step Methods for Parabolic Problems: Part I*, Math. Comp., v. 35, 1980, pp. 655-677.

[4] BRAMBLE, J. H., SCHATZ, A. H., THOMÉE, V., AND WAHLBIN, L. B., *Some Convergence Estimates for Semidiscrete Galerkin Type Approximations for Parabolic Equations*, SIAM J. Numer. Anal., 14 (1977), pp. 218-241.

[5] BUTCHER, J. C., *Implicit Runge-Kutta Processes*, Math. Comp., 18 (1964), pp. 50-64.

[6] CIARLET, P. G., *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, New York, Oxford, 1978.

[7] CROUZEIX, M., *Sur l'approximation des équations différentielles opérationnelles linéaires par des méthodes de Runge-Kutta*, Thèse, Université de Paris VI, 1975.

[8] DEKKER, K., AND VERWER, J. G., *Stability of Runge-Kutta Methods for Stiff Nonlinear Differential Equations*, North-Holland, Amsterdam, New York, Oxford, 1984.

[9] DOUGALIS, V. A., AND KARAKASHIAN, O. A., *On Some High-Order Accurate Fully Discrete Galerkin Methods for the Korteweg-De Vries Equation*, Math. Comp., 45 (1985), pp. 329-345.

[10] DOUGLAS, J., JR., DUPONT, T., AND EWING, R., *Incomplete Iterations for Time-Stepping a Galerkin Method for a Quasilinear Parabolic Problem*, SIAM J. Numer. Anal., v. 16, 1979, pp. 503-522.

[11] HAGEMAN, L. A., AND YOUNG, D. M., *Applied Iterative Methods*, Academic Press, New York, London, Sydney, San Francisco, 1981.

[12] HAIRER, E., AND WANNER, G., *Algebraically Stable and Implementable Runge-Kutta Methods of High Order*, SIAM J. Numer. Anal., v. 18, 1981, pp. 1098-1108.

[13] KARAKASHIAN, O. A., *On Runge-Kutta Methods for Parabolic Problems with Time Dependent Coefficients*, Math. Comp., 47 (1986), pp. 77-106.

[14] KEELING, S. L., *Galerkin/Runge-Kutta Discretizations for Parabolic Partial Differential Equations*, Ph.D. Dissertation, University of Tennessee, 1986.

[15] KEELING, S. L., *On Implicit Runge-Kutta Methods for Parallel Computations*, ICASE Report No. 87-58, NASA Langley Research Center, Hampton, VA, 1987.

[16] SAMMON, P. H., *Approximations for Parabolic Equations with Time-Dependent Coefficients*, Ph. D. Thesis, Cornell University, 1978.

[17] SAMMON, P. H., *Convergence Estimates for Semidiscrete Parabolic Equation Approximations*, SIAM J. Numer. Anal., v. 19, 1982, pp. 68-92.

NASA
National Aeronautics and Space Administration

Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
NASA CR-178372 ICASE Report No. 87-61		

4. Title and Subtitle	5. Report Date
GALERKIN/RUNGE-KUTTA DISCRETIZATIONS FOR PARABOLIC EQUATIONS WITH TIME DEPENDENT COEFFICIENTS	September 1987
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Stephen L. Keeling	87-61
	10. Work Unit No. 505-90-21-01

9. Performing Organization Name and Address	11. Contract or Grant No.
Institute for Computer Applications in Science and Engineering Mail Stop 132C, NASA Langley Research Center Hampton, VA 23665-5225	NAS1-18107
	13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address	Contractor Report
National Aeronautics and Space Administration Langley Research Center Hampton, VA 23665-5225	14. Sponsoring Agency Code

15. Supplementary Notes

Langley Technical Monitor:
Richard W. Barnwell

Final Report

Submitted to Math. Comp.

16. Abstract

A new class of fully discrete Galerkin/Runge-Kutta methods is constructed and analyzed for linear parabolic initial boundary value problems with time dependent coefficients. Unlike any classical counterpart, this class offers arbitrarily high order convergence while significantly avoiding what has been called order reduction. In support of this claim, error estimates are proved, and computational results are presented. Additionally, since the time stepping equations involve coefficient matrices changing at each time step, a preconditioned iterative technique is used to solve the linear systems only approximately. Nevertheless, the resulting algorithm is shown to preserve the original convergence rate while using only the order of work required by the base scheme applied to a linear parabolic problem with time independent coefficients. Futhermore, it is noted that special Runge-Kutta methods allow computations to be performed in parallel so that the final execution time can be reduced to that

17. Key Words (Suggested by Author(s))	18. Distribution Statement
implicit Runge-Kutta methods, time dependent coefficients, error estimates, order reduction	64 - Numerical Analysis Unclassified - unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of pages	22. Price
Unclassified	Unclassified	40	A03

NASA FORM 1626 OCT 86