ON A CLASS OF CONSERVATIVE, HIGHLY ACCURATE GALERKIN METHODS FOR THE SCHRÖDINGER EQUATION

GEORGIOS D. AKRIVIS AND VASSILIOS A. DOUGALIS

Abstract. We construct and analyze efficient fully discrete Galerkin type methods that are of high order of accuracy and conservative in the L^2 sense for approximating the solution of a form of the linear Schrödinger equation with a time-dependent coefficient, found e.g. in underwater acoustics. The time stepping procedures are based on the class of implicit Runge–Kutta methods known as the $q-$stage Gauss–Legendre schemes. L^2 error estimates are proved that are of optimal order in space and of temporal order $q+2$. An iterative procedure at each time step for the efficient implementation of the two-stage scheme is proposed and analyzed.

Rèsumè. On construit et analyse des méthodes totalement discrètes du type Galerkin, qui sont L^2-conservatives et d'order arbitraire, pour approcher la solution d'une forme de l'équation linéaire de Schrödinger avec un coefficient qui dépend du temps, trouvée par exemple dans l'acoustique sous-marine. La procédure de discrétisation en temps est basée sur la classe des méthodes implicites de Runge–Kutta connues comme les schémas de Gauss–Legendre à q pas intermédiaires. On obtient des estimations dans L^2 pour les erreurs, qui sont d'order optimal en espace es d'order $q+2$ en temps. On propose et analyse aussi une procéduce itérative pour résoudre les systèmes linéaires à chaque pas de temps pour une application efficace de schéma Gauss–Legendre à deux pas intermédiaires.

Dedicated to Professor Dr. G. Hämmerlin on the occasion of his 60^{th} birthday, July 31, 1988.

1. Introduction

In this paper we shall study conservative numerical methods of high order of accuracy for approximating the solution of the following initial- and boundary-value problem for a partial differential equation of the Schrödinger type. Let Ω be a bounded domain in $\mathbb{R}^N$ with smooth boundary $\partial\Omega$ and let $0 < T < \infty$ be given. We seek a complex-valued function $u = u(x,t), (x,t) \in \bar{\Omega} \times [0,T]$, satisfying:

$$\begin{aligned} &u_t = \mathrm{i}L(t)u := \mathrm{i}\big(\alpha\Delta u + \beta(x,t)u\big) && \text{in} \quad \Omega \times [0,T], \\ &u = 0 && \text{on} \quad \partial\Omega \times [0,T], \\ &u(x,0) = u^0(x) && \text{in} \quad \bar{\Omega}, \end{aligned} \tag{1.1}$$

where α is a given nonzero real number, β is a given smooth real-valued function on $\bar{\Omega} \times [0,T]$ and u^0 is a given smooth complex-valued function on $\bar{\Omega}$. We shall assume

Work supported by the Institute of Applied and Computational Mathematics of the Research Center of Crete-FORTH.

that the data of (1.1) are smooth and compatible enough to ensure that the problem possesses a unique and smooth enough for our purposes solution; cf. e.g. [15, Chapter 5, Section 12] and [4] for relevant existence, uniqueness and regularity results. This form of the Schrödinger equation occurs, for example for $N = 1$, in underwater acoustics as “parabolic approximation” to the Helmholtz equation, cf. e.g. [20], posed with a variety of types of boundary conditions. For simplicity we consider here the case of homogeneous Dirichlet boundary conditions. In (1.1) the Laplacian Δ could have as well been replaced by a second-order, symmetric, uniformly positive definite elliptic operator on $\bar{\Omega}$ with space-dependent coefficients with no complications in the error estimates.

We shall discretize (1.1) in space by a Galerkin-finite element type method and in time by a class of implicit Runge–Kutta schemes of arbitrary order, known as the *Gauss–Legendre collocation type methods*. We shall estimate the error of the fully discrete approximations in L^2 and point out efficient ways for implementing the methods.

Many finite difference and spectral schemes, usually of second-order temporal accuracy, have been proposed in the literature for problems such as (1.1); see e.g. the survey [13], the collections of papers in [19] and [14] and their references. Galerkin-finite element methods have also been considered. For the linear Schrödinger equation with time-independent coefficients early error estimates for semidiscrete approximations may be found in [21], [23]. Semidiscrete and fully discrete Galerkin approximations with Runge–Kutta time stepping have been analyzed in [4] in the general case where the Laplacian in the right-hand side of the p.d.e. is replaced by a second-order elliptic operator with space- and time-dependent coefficients. For alternative approaches based on separating real and imaginary parts, cf. [10], [16]; in this paper we shall discretize (1.1) directly using complex arithmetic. Among the growing literature on Galerkin type methods for *nonlinear* Schrödinger equations, some error estimates are shown in [17], where second-order time stepping procedures coupled with finite difference or Galerkin type space discretizations are analyzed; for computations with such schemes cf. e.g. [18].

We next introduce notation that will be used in the sequel. For integral $s \geq 0$ let $H^s = H^s(\Omega)$ denote the usual, complex (Hilbert) Sobolev spaces with corresponding norm $\|\cdot\|_s$. For $f, g \in L^2 = H^0$, let

$$(f, g) = \int_\Omega f(x)\overline{g(x)}\,dx$$

be their L^2 inner product; here the overbar denotes complex conjugation. Let $\|\cdot\|$ denote the associated L^2 norm and $|\cdot|_\infty$ be the norm of $L^\infty = L^\infty(\Omega)$. As usual, H_0^1 will consist of the elements of H^1 that vanish on $\partial\Omega$ in the sence of trace. We shall discretize (1.1) in space by the *standard Galerkin method*, as follows. For $0 < h < 1$, let S_h be a family of finite-dimensional subspaces of H_0^1 in which approximations to the solution $u(\cdot, t)$ of (1.1) will be sought for given $t \in [0, T]$. We assume that S_h satisfies the *approximation property* that there exists an integer $r \geq 2$ and a constant $c > 0$

independent of h such that for $v \in H^s \cap H_0^1$

$$\inf_{\varphi\in S_h} \big(\|v-\varphi\| + h\|v-\varphi\|_1\big) \le ch^s\|v\|_s, \quad 1 \le s \le r \tag{1.2}$$

and the *inverse property* that for some $c > 0$ independent of h,

$$\|\varphi\|_1 \le ch^{-1}\|\varphi\|, \quad \forall \varphi \in S_h. \tag{1.3}$$

As in (1.2), (1.3), in the sequel the symbols c, C, c_i etc. will denote generic constants independent of the discretization parameter h and the time step. Such constants may also depend on the solution and the data of (1.1).

We define now the *semidiscrete approximation* of the solution $u(t) = u(\cdot, t)$ of (1.1) in S_h in the customary way as the map $u_h : [0, T] \to S_h$ satisfying (with $\beta(t) = \beta(\cdot, t)$):

$$\begin{aligned} &(u_{ht}, \varphi) = -\mathrm{i}\alpha a(u_h, \varphi) + \mathrm{i}(\beta(t)u_h, \varphi), \quad \forall \varphi \in S_h,\ 0 \le t \le T, \\ &u_h(0) = u_h^0, \end{aligned} \tag{1.4}$$

where, for $\varphi, \chi \in H^1$,

$$a(\varphi, \chi) = \sum_{j=1}^{N} (\partial_j\varphi, \partial_j\chi).$$

In addition, we shall henceforth assume that $u^0 \in H^r \cap H_0^1$ and that u_h^0 is an element of S_h such that

$$\|u^0 - u_h^0\| \le ch^r\|u^0\|_r; \tag{1.5}$$

for example $u_h^0 = Pu^0$, where $P : L^2 \to S_h$ is the L^2-projection operator onto S_h. If we introduce the linear operators $\mathit{\Delta}_h : S_h \to S_h$ and $B_h(t) : S_h \to S_h, L_h(t) : S_h \to S_h, 0 \le t \le T$, by

$$(\mathit{\Delta}_h\varphi, \chi) = -a(\varphi, \chi), \quad \forall \varphi, \chi \in S_h, \tag{1.6}$$

$$(B_h(t)\varphi, \chi) = (\beta(t)\varphi, \chi), \quad \forall \varphi, \chi \in S_h, \quad \text{i.e.,} \quad B_h(t)\varphi = P[\beta(\cdot, t)\varphi], \tag{1.7}$$

$$L_h(t) = \alpha\mathit{\Delta}_h + B_h(t), \quad 0 \le t \le T, \tag{1.8}$$

and note that, since α and β are real, $\mathit{\Delta}_h, B_h(t)$ and $L_h(t), 0 \le t \le T$, are Hermitian operators on $\{S_h, (\cdot, \cdot)\}$, we may write (1.4) as

$$\begin{aligned} &u_{ht} = \mathrm{i}L_h(t)u_h, \quad 0 \le t \le T, \\ &u_h(0) = u_h^0. \end{aligned} \tag{1.9}$$

The equations (1.4) or (1.9) represent an initial-value problem for a system of ordinary differential equations that obviously has a unique solution $u_h(t)$ for $0 \le t \le T$; they will only be used in the sequel in order to motivate the fully discrete approximations. Let us only remark here that it is not hard to show, by comparing e.g. u_h to the *elliptic projection* of u in the standard way, cf. [22], [4] that if u_h^0 satisfies (1.5), then for $t \in [0, T]$ there holds

$$\|(u-u_h)(t)\| \le ch^r\Big\{\|u^0\|_r + \int_0^t \big(\|u(s)\|_r + \|u_t(s)\|_r\big)\,ds\Big\},$$

i.e. that $u_h(t)$ satisfies an optimal-order L^2 error estimate if u is smooth enough.

We shall discretize (1.9) in time using the well-known class of implicit Runge–Kutta procedures of collocation type known as the $q-$stage ($q \geq 1$ integer) Gauss–Legendre methods, [5], [6], [9]. The methods are defined by constants $A = (a_{ij}) \in \mathbb{R}^{q\times q}, b = (b_1, \dots, b_q)^T \in \mathbb{R}^q, \tau = (\tau_1, \dots, \tau_q)^T \in \mathbb{R}^q$ that are constructed in the standard way. Specifically, the τ_i are the —distinct, in $(0,1)$— zeros of the shifted Legendre polynomials $(d/dx)^q\big(x^q(1-x)^q\big)$, the weights b_i are defined as the solution of the $q \times q$ linear system of equations represented by (3.1.1) below for $1 \leq \ell \leq q$, while the coefficients a_{ij} are defined for each $i, 1 \leq i \leq q$, as solutions of the $q\times q$ linear system of equations represented by (3.1.2). Let $k > 0$ be the (constant) time step, let $t^n = nk, n = 0, 1, \dots, M$, where $T = Mk$, and $t^{n,i} = t^n + \tau_i k, 1 \leq i \leq q$. Then, the $q-$stage Gauss–Legendre methods applied to the system of ordinary differential equations represented by (1.9) yield the following *full discretization* of (1.1): For $0 \leq n \leq M$, we seek $U^n \in S_h$, approximating $u(t^n)$, and $U^{n,m} \in S_h, 1 \leq m \leq q$, that satisfy

$$\begin{aligned}
&(a) \quad U^0 = u_h^0,\\
&\qquad \text{for } \ n = 0, 1, \dots, M-1:\\
&(b) \quad U^{n,m} = U^n + \mathrm{i}k\sum_{j=1}^{q} a_{mj}L_h^{n,j}U^{n,j}, \quad 1 \leq m \leq q,\\
&(c) \quad U^{n+1} = U^n + \mathrm{i}k\sum_{j=1}^{q} b_j L_h^{n,j}U^{n,j},
\end{aligned} \tag{1.10}$$

where $L_h^n = L_h(t^n), L_h^{n,j} = L_h(t^{n,j})$.

In Section 2 we shall show that, for each n, the system (1.10b) has a unique solution $\{U^{n,m}\}, 1 \leq m \leq q$, in $(S_h)^q$ and therefore that U^{n+1} is defined uniquely in S_h for $0 \leq n \leq M-1$, by (1.10c). We shall also verify that (1.10) is *unconditionally stable* in L^2 and, in particular, *conservative*, i.e. that it satisfies $\|U^n\| = \|U^0\|, 0 \leq n \leq M$; thus it mimicks the behavior of (1.1) and (1.9) for which $\|u(t)\| = \|u^0\|$ and $\|u_h(t)\| = \|u_h^0\|$ hold, respectively, for $0 \leq t \leq T$.

In Section 3 we shall estimate the error of the approximation U^n in the L^2 norm; specifically, we shall show in Theorem 3.1, which is the main result of the paper, that

$$\max_{0\leq n\leq M} \|U^n - u(t^n)\| \leq c\big(k^{\min(2q,q+2)} + h^r\big). \tag{1.11}$$

This error bound is of optimal order in space. As far as the temporal rate of convergence is concerned, it is well-known that the $q-$stage Gauss–Legendre methods have (classical) order $2q$ when applied to nonstiff ordinary differential equations. Therefore our proof certainly gives optimal-order temporal convergence, resp. $2, 4$, for the one- and two-stage, resp., methods that seem to enjoy current practical importance. For $q > 2$ our result — $O(k^{q+2})$ in time — shows the effect of "reduction of order due to stiffness". This result is no worse than analogous estimates proven in the literature for Runge–Kutta full discretizations of initial- and boundary-value problems for p.d.e.'s with time-dependent coefficients or nonlinear terms, posed with Dirichlet boundary conditions: In his thesis Brocéhn, [4], considers full discretizations of the Schrödinger equation with a general second-order elliptic operator with time-dependent coefficients using some semi-implicit Runge–Kutta methods. (The class of Gauss–Legendre schemes considered here is not semi-implicit with the exception of the one-stage scheme.) For his schemes, using different estimation techniques, he proves error bounds with temporal order of accuracy equal to $\min(p, q+1)$, where, in his notation, q is the order of the quadrature rule associated with the intermediate stages of the Runge–Kutta method and plays an analogous role to the q used here, and p is the classical order (equal to $2q$ for the Gauss–Legendre methods). For the special elliptic operator of the right-hand side of (1.1) he obtains a temporal rate of $q+2$, if $p = q-2$, and under the mesh condition that kh^{-2} remain bounded as $k, h \to 0$. In Theorem 2 of [8], the temporal discretization, by a class of Runge–Kutta schemes (disjoint from the Gauss–Legendre methods but suitable for parabolic problems) of an abstract semilinear parabolic equation is shown to have a rate of convergence exhibiting an analogous limitation to our $q+2$ result. In [3] the Gauss–Legendre methods are applied to a nonlinear p.d.e., the generalized Korteweg–de Vries equation, posed in one space dimension with periodic boundary conditions and discretized in space with smooth periodic splines on a uniform mesh. The exact analog to (1.11) is proved then by a different technique from the one at hand, with the details of the space discretization and the periodicity of the exact and discrete solutions playing a crucial role. (After the completion of the original version of the paper at hand, we learnt that Karakashian and McKinney, [12], proved the optimal order $2q$ for the Korteweg–de Vries equation; their remarkable proof again relies heavily on the periodic boundary conditions.)

Finally, in Section 4 we confine attention to the $2-$stage Gauss–Legendre method and devise a scheme that avoids solving the $2\dim S_h \times 2\dim S_h$ linear system represented by (1.10b) for $q = 2$. A suitable decoupling strategy and an iteration scheme enables us to produce stable and optimal-order accurate approximations to U^n by solving a number of linear systems of size $\dim S_h \times \dim S_h$ at each time step; these systems will have sparse matrices if S_h is furnished with a finite element basis with elements of small support.

Acknowledgement. The authors record their thanks to a referee for pointing out reference [4] to them.

2. Existence and stability of the fully discrete approximations

In this section we shall show that for each n the linear system represented by (1.10b) has a unique solution $U^{n,m}, 1 \le m \le q$, and that the resulting overall fully discrete scheme (1.10) is stable (conservative) in the L^2 norm. For this purpose we shall make use of the following well-known properties of the Gauss–Legendre methods, [6], [9]:

(2.1) For each $q \geq 1$ there exists a diagonal $q \times q$ matrix D with positive diagonal elements, such that the matrix DAD^{-1} is positive definite on $\mathbb{R}^q$. (See e.g. [9, Theorem 5.5.6, Cor. 5.1.4 and (5.1.23)].)

(2.2) $b_i > 0, b_i a_{ij} + b_j a_{ji} - b_i b_j = 0, \quad 1 \leq i, j \leq q,$
i.e. the Gauss–Legendre methods are *conservative* in the nonlinear context; cf. e.g. [7], [9, p.117].

We first examine the *existence of solutions* $U^{n,m}$ of the linear system (1.10b). On the product space $(S_h)^q$ we let $\mathbb{U}^n$ denote the vector $(U^{n,1}, \ldots, U^{n,q})^T$ and $\mathbb{L}_h^n : (S_h)^q \to (S_h)^q$ be the *diagonal operator* given, for $\mathbb{V} \in (S_h)^q$, by $\mathbb{L}_h^n \mathbb{V} = (L_h^{n,1} V_1, \ldots, L_h^{n,q} V_q)^T$. We write then the equations (1.10a,b) respectively as

$$\mathbb{U}^n = U^n \boldsymbol{e} + \mathrm{i}kA\mathbb{L}_h^n \mathbb{U}^n, \tag{2.3a}$$

$$U^{n+1} = U^n + \mathrm{i}kb^T \mathbb{L}_h^n \mathbb{U}^n. \tag{2.3b}$$

In (2.3a,b) and in the sequel, abusing notation a bit to avoid tensor products, for $\mathbb{V} \in (S_h)^q$ we let $b^T \mathbb{V} = \sum_{i=1}^q b_i V_i, A\mathbb{V} \in (S_h)^q : (A\mathbb{V})_i = \sum_{j=1}^q a_{ij} V_j, \boldsymbol{e} = (1, 1, \ldots, 1)^T \in \mathbb{R}^q$ and for $U \in S_h, U\boldsymbol{e} = (U, U, \ldots, U)^T \in (S_h)^q$. The existence of $\mathbb{U}^n$, solution of (2.3a), will follow from the following general lemma. (In the sequel we let $((\cdot, \cdot))$, resp. $|\!|\!| \cdot |\!|\!|$, denote the product inner product, resp. norm, in $(L^2)^q$.)

Lemma 2.1. *Suppose that $\mathbb{V} = (V_i)$ and $\mathbb{W} = (W_i)$ in $(S^h)^q$ satisfy the equation*

$$\mathbb{V} = \mathbb{W} + \mathrm{i}kA\mathbb{F}(\mathbb{V}), \tag{2.4}$$

where $\mathbb{F} : (S_h)^q \to (S_h)^q$ is a diagonal mapping such that $(\mathbb{F}(V))_i = F_i(V_i), 1 \leq i \leq q$, where $F_i : S_h \to S_h$ are given mappings with the property that $\mathrm{Im}(F_i(\varphi), \varphi) = 0$, for $\varphi \in S_h, 1 \leq i \leq q$. Then, there exists a constant c, depending only on the constants of the Gauss–Legendre method, such that

$$|\!|\!|\mathbb{V}|\!|\!| \leq c|\!|\!|\mathbb{W}|\!|\!|. \tag{2.5}$$

Proof. Let $D = \mathrm{diag}\,(d_1, \ldots, d_q), d_i > 0$, be the diagonal matrix mentioned in (2.1). Multiplying (2.4) by $D^2 A^{-1}$ on the left and taking the $(L^2)^q$ inner product with $\mathbb{V}$ we obtain

$$((D^2 A^{-1}\mathbb{V}, \mathbb{V})) = ((D^2 A^{-1}\mathbb{W}, \mathbb{V})) + \mathrm{i}k((D^2 \mathbb{F}(\mathbb{V}), \mathbb{V})). \tag{2.6}$$

By our hypotheses $\mathrm{Re}[\mathrm{i}k((D^2\mathbb{F}(\mathbb{V}), \mathbb{V}))] = 0$. Hence, taking real parts in (2.6) gives

$$\mathrm{Re}((D^2 A^{-1}\mathbb{V}, \mathbb{V})) = \mathrm{Re}((D^2 A^{-1}\mathbb{W}, \mathbb{V})). \tag{2.7}$$

Denote $DA^{-1}D^{-1} =: B = (b_{ij}) \in \mathbb{R}^{q\times q}$. By (2.1) there exists $\lambda > 0$ such that $\sum_{i,j=1}^q b_{ij}\xi_i\xi_j \geq \lambda \sum_{i=1}^q \xi_i^2$ for every $\xi \in \mathbb{R}^q$. Hence, putting $D\mathbb{V} = \mathbb{Y}$ we have

$$\begin{aligned}\operatorname{Re}((D^2A^{-1}\mathbb{V},\mathbb{V})) &= \operatorname{Re}((DA^{-1}D^{-1}\mathbb{Y},\mathbb{Y})) = \operatorname{Re}((B\mathbb{Y},\mathbb{Y}))\\ &\geq \lambda\big(\|\operatorname{Re}\mathbb{Y}\|^2 + \|\operatorname{Im}\mathbb{Y}\|^2\big) = \lambda\|\mathbb{Y}\|^2\\ &\geq \lambda \min_i d_i^2 \|\mathbb{V}\|^2.\end{aligned}$$

(2.5) follows then by (2.7) and the Cauchy–Schwarz inequality. ☐

Given now $U^n \in S_h$ apply Lemma 2.1 to the *linear* system (2.3a) where $\mathbb{F}(\mathbb{V}) = \mathbb{L}_h^n\mathbb{V}, F_i(V_i) = L_h^{n,i}V_i$ and $\operatorname{Im}(L_h^{n,i}\varphi,\varphi) = 0$ for $\varphi \in S_h$. By (2.5) we see that the homogeneous system ($U^n = 0$ in (2.3a)) has only the trivial solution. Hence, given $U^n \in S_h$, (2.3a) has a unique solution $\mathbb{U}^n = (U^{n,i})$ in $(S_h)^q$, which satisfies

$$\max_i \|U^{n,i}\| \leq c\|U^n\|, \tag{2.8}$$

for some constant c that depends only on the Gauss–Legendre method.

The *stability* (*conservativeness*) of the scheme (1.10) follows from the following result, stated again in slightly more general terms:

Lemma 2.2. *Suppose, given $U \in S_h$, that $\mathbb{V} \in (S_h)^q$ and $Y \in S_h$ satisfy the equations*

$$\mathbb{V} = U\boldsymbol{e} + \mathrm{i}kA\mathbb{F}(\mathbb{V}), \tag{2.9a}$$

$$Y = U + \mathrm{i}kb^T\mathbb{F}(\mathbb{V}), \tag{2.9b}$$

where $\mathbb{F}$ is a mapping that satisfies the hypotheses of Lemma 2.1. Then

$$\|Y\| = \|U\|. \tag{2.10}$$

Proof. (2.9b) gives

$$\begin{aligned}\|Y\|^2 &= \|U\|^2 + \mathrm{i}k(b^T\mathbb{F}(\mathbb{V}),U) - \mathrm{i}k(U,b^T\mathbb{F}(\mathbb{V})) + k^2(b^T\mathbb{F}(\mathbb{V}),b^T\mathbb{F}(\mathbb{V}))\\ &= \|U\|^2 - 2k\operatorname{Im}\sum_{j=1}^q b_j(W_j,U) + k^2\sum_{j,\ell=1}^q b_jb_\ell(W_j,W_\ell),\end{aligned}$$

where $W_j = F_j(V_j)$. Let $\mathbb{W} = (W_1,\dots,W_q)^T$. Replacing U in the right-hand side of the above equation by its expression $U = V_j - \mathrm{i}k(A\mathbb{W})_j$ that (2.9a) gives, we see, using the properties of $\mathbb{F}$ and (2.2) that

$$\begin{aligned}\|Y\|^2 &= \|U\|^2 - 2k^2\sum_{i=1}^q b_i\operatorname{Re}(W_i,(A\mathbb{W})_i) + k^2\sum_{i,j=1}^q b_ib_j(W_i,W_j)\\ &= \|U\|^2 - k^2\Big[\sum_{i,j=1}^q (b_ia_{ij} + b_ja_{ji} - b_ib_j)(W_i,W_j)\Big] = \|U\|^2.\end{aligned}$$ ☐

Applying this result to the scheme (1.10) we obtain

$$\|U^n\| = \|U^0\|, \quad 0 \leq n \leq M, \tag{2.11}$$

i.e. that our fully discrete method is *conservative* in the L^2 sense.

3. Consistency and convergence

In this section we shall study the consistency of the fully discrete scheme (1.10) and prove the error estimate (1.11). To this effect we shall first list some well-known algebraic properties of the Gauss–Legendre methods, [6], [9], that will be used in the sequel along with (2.1) and (2.2).

(3.1) The $q-$stage Gauss–Legendre method is *consistent of order* $2q$ (i.e. has accuracy of order $2q$ when applied to an ordinary differential equation $y' = f(t,y)$, where f and its partial derivatives of sufficiently high order with respect to y and t are smooth and bounded), and satisfies the following *order* (*simplifying*) *conditions*, [5]:

(3.1.1)
$$\sum_{i=1}^{q} b_i \tau_i^{\ell-1} = \ell^{-1}, \quad 1 \le \ell \le 2q,$$

(3.1.2)
$$\sum_{j=1}^{q} a_{ij} \tau_j^{\ell-1} = \tau_i^{\ell}/\ell, \quad 1 \le i, \ell \le q,$$

(3.1.3)
$$\sum_{i=1}^{q} b_i \tau_i^{\ell-1} a_{ij} = \ell^{-1} b_j (1 - \tau_j^{\ell}), \quad 1 \le j, \ell \le q.$$

(3.2) The $q-$stage Gauss–Legendre method corresponds to the $q-$th diagonal Padé rational approximation to the exponential, i.e. if $r(z) = 1 + zb^T(I - zA)^{-1}\boldsymbol{e}, \boldsymbol{e} = (1,\dots,1)^T \in \mathbb{R}^q$, then $r(z)$ is the $q-$th diagonal Padé approximant to $\exp(z)$.

For the purposes of the proof of convergence we shall compare the solution U^n of (1.10) to the *elliptic projection* $W = W(t) \in S_h, 0 \le t \le T$, of the solution $u(t)$ of (1.1), defined as usual by

(3.3)
$$(\Delta_h W, \varphi) = -a(W,\varphi) = (\Delta u, \varphi) \quad \forall \varphi \in S_h.$$

We shall denote the associated (time-independent) elliptic projection operator onto S_h (defined on $H^2 \cap H_0^1$) by P_I. In this notation, $W(t) = P_I u(t)$ and obviously $W^{(j)}(t) = P_I u^{(j)}(t)$; here and in the sequel $v^{(j)}(t) = (d/dt)^j v(t)$. By our assumptions on S_h there follows that

(3.4)
$$\|v - P_I v\| + h\|v - P_I v\|_1 \le ch^r \|v\|_r, \quad \forall v \in H^r \cap H_0^1.$$

Obviously (3.3) implies that $\|W^{(j)}(t)\|_1 \le c\|u^{(j)}(t)||_1 \le c_j, j \ge 0$. In addition, there exist constants c_j such that $\|L_h(t)W^{(j)}(s)\| \le c_j, t, s \in [0,T], j \ge 0$; this follows from the observation that for any $\varphi \in S_h$

$$\begin{aligned}\big|\big(L_h(t)W^{(j)}(s),\varphi\big)\big| &= \big|\alpha(\Delta_h W^{(j)}(s),\varphi)+(\beta(t)W^{(j)}(s),\varphi)\big|\\ &\le \big(|\alpha|\,\|\Delta u^{(j)}(s)\| + |\beta(t)|_\infty \|u^{(j)}(s)\|_1\big)\|\varphi\|\\ &\le c\|u^{(j)}(s)\|_2\,\|\varphi\|.\end{aligned}$$

In fact, if $L_h^{(j)}(t) : S_h \to S_h$ denotes the $j-$th time derivative of the operator $L_h(t)$, given by $(L_h^{(j)}(t)\varphi,\chi) = (\partial_t^j\beta(\cdot,t)\varphi,\chi), j \ge 1$, for $\chi,\varphi \in S_h$, i.e. by $L_h^{(j)} = B_h^{(j)}, j \ge 1$, we have $\|L_h^{(i)}(t)W^{(j)}(s)\| \le c_i\|u^{(j)}(s)\|_1, i \ge 1, j \ge 0$. Thus, we can generalize the previous estimate to

$$\|L_h^{(i)}(t)W^{(j)}(s)\| \le c_{ij}, \quad i,j \ge 0, \quad t,s \in [0,T], \tag{3.5}$$

and also note that

$$\|L_h^{(i)}(t)\| \le c_i, \quad i \ge 1, \tag{3.6}$$

where $\|\cdot\|$ denotes here the L^2 induced operator norm on S_h.

We shall also make use of the following property of the elliptic projection, namely that for constants c_{ij}

$$\|L_h(t)B^{(j)}(s)W^{(i)}(s)\| \le c_{ij}, \quad i,j \ge 0, \quad t,s \in [0,T], \tag{3.7}$$

which may be proved as follows. We have

$$L_h(t)B_h^{(j)}(s)W^{(i)}(s) = \alpha\Delta_h B_h^{(j)}(s)W^{(i)}(s) + B_h(t)B_h^{(j)}(s)W^{(i)}(s). \tag{3.8}$$

Since for $j \ge 0$, $B_h^{(j)}(t)\varphi = P(\partial_t^j\beta(t)\varphi)$, we have $\|B_h^{(j)}\varphi\| \le c_j\|\varphi\|$ for $\varphi \in S_h$. Hence in (3.8)

$$\|B_h(t)B_h^{(j)}(s)W^{(i)}(s)\| \le c_{ij}. \tag{3.9}$$

Also for $\varphi \in S_h$ we obtain, suppressing the dependence on s,

$$\begin{aligned}-(\Delta_h(t)B_h^{(j)}W^{(i)},\varphi) &= a(B_h^{(j)}W^{(i)},\varphi) = a(P[\beta^{(j)}W^{(i)}],\varphi)\\ &= a(P[\beta^{(j)}W^{(i)}] - \beta^{(j)}u^{(i)},\varphi) + a(\beta^{(j)}u^{(i)},\varphi).\end{aligned} \tag{3.10}$$

We obviously have

$$|a(\beta^{(j)}u^{(i)},\varphi)| = |(\Delta(\beta^{(j)}u^{(i)}),\varphi)| \le c_{ij}\|\varphi\|, \tag{3.11}$$

and by (1.3)

$$\begin{aligned}|a(P[\beta^{(j)}W^{(i)}] - \beta^{(j)}u^{(i)},\varphi)| &\le \|P[\beta^{(j)}W^{(i)}] - \beta^{(j)}u^{(i)}\|_1\,\|\varphi\|_1\\ &\le ch^{-1}\|P[\beta^{(j)}W^{(i)}] - \beta^{(j)}u^{(i)}\|_1\,\|\varphi\|.\end{aligned} \tag{3.12}$$

Now

$$\|P[\beta^{(j)}W^{(i)}] - \beta^{(j)}u^{(i)}\|_1 \le \|P[\beta^{(j)}(W^{(i)} - u^{(i)})]\|_1 + \|P[\beta^{(j)}u^{(i)}] - \beta^{(j)}u^{(i)}\|_1. \tag{3.13}$$

Since, as it may easily be seen from (1.3), (3.4), $\|Pv-v\|_1 \le ch^{r-1}\|v\|_r$ for $v \in H^r \cap H_0^1$, and since for β and u sufficiently smooth $\beta^{(j)}u^{(i)} \in H^r \cap H_0^1$, there follows

$$\begin{aligned}\|P[\beta^{(j)}(W^{(i)}-u^{(i)})]\|_1 &+ \|P[\beta^{(j)}u^{(i)}] - \beta^{(j)}u^{(i)}\|_1 \\ &\le ch^{-1}\|P[\beta^{(j)}(W^{(i)}-u^{(i)})]\| + c_{ij}h^{r-1} \\ &\le c_{ij}h^{r-1}.\end{aligned}$$

These estimates, when substituted in (3.13), yield, in conjunction with (3.10)–(3.12), that $\|\Delta_h B_h^{(j)} W^{(i)}\| \le c_{ij}$ since $r \ge 2$. Then (3.7) follows from (3.8) and (3.9).

We now embark upon the proof of the main result of this section. For a function v defined on $[0,T]$ we generally denote $v^n = v(t^n)$. We first define, for the purposes of the proof of consistency, $V^{n,m}$ for $0 \le n \le M-1, 1 \le m \le q$ and $V^n, 0 \le n \le M$ in S_h by

$$\begin{aligned}&(a) \quad V^0 = W^0, \\ &(b) \quad V^{n,m} = W^n + \mathrm{i}k\sum_{j=1}^{q} a_{mj}L_h^{n,j}V^{n,j}, \quad 1 \le m \le q, \\ &(c) \quad V^{n+1} = W^n + \mathrm{i}k\sum_{j=1}^{q} b_j L_h^{n,j}V^{n,j}.\end{aligned} \tag{3.14}$$

In Proposition 3.1 below we shall prove the *consistency* result, valid for u sufficiently smooth:

$$\max_{0\le n\le M}\|V^n - W^n\| \le ck\big(k^{\min(2q,q+2)} + h^r\big). \tag{3.15}$$

If this holds, then a simple stability calculation, as the following theorem shows, gives the error bound (1.11):

Theorem 3.1. *Let u be sufficiently smooth and suppose that* (3.15) *holds. Then*

$$\max_{0\le n\le M}\|U^n - u^n\| \le c\big(k^{\min(2q,q+2)} + h^r\big). \tag{3.16}$$

Proof. Let $V^{n,m}, V^n$ be defined by (3.14) and let $\varepsilon^{n,m} = U^{n,m} - V^{n,m}, \varepsilon^n = U^n - V^n, \zeta^n = U^n - W^n$. Then (1.10) and (3.14) yield

$$\begin{aligned}\varepsilon^{n,m} &= \zeta^n + \mathrm{i}k\sum_{j=1}^{q} a_{mj}L_h^{n,j}\varepsilon^{n,j}, \quad 1 \le m \le q, \\ \varepsilon^{n+1} &= \zeta^n + \mathrm{i}k\sum_{j=1}^{q} b_j L_h^{n,j}\varepsilon^{n,j}.\end{aligned}$$

The stability Lemma 2.2 gives then that $\|\varepsilon^{n+1}\| = \|\zeta^n\|$. Hence, for $0 \le n \le M-1$,

$$\|\zeta^{n+1}\| \le \|\varepsilon^{n+1}\| + \|V^{n+1} - W^{n+1}\| = \|\zeta^n\| + \|V^{n+1} - W^{n+1}\|.$$

Therefore, by (3.15) $\|\zeta^n\| \le \|\zeta^0\| + c(k^{\min(2q,q+2)} + h^r), 0 \le n \le M$, and the result follows from (3.4), (1.5) and the triangle inequality. □

Hence our task is to prove consistency:

Proposition 3.1. *If u is sufficiently smooth,* (3.15) *holds.*

Proof. We follow, up to a point, the technique of the consistency proof for Runge–Kutta discretizations of partial differential equations introduced in [11]. First define $\tau_{ij}, 1 \le i \le q, j \ge 0$ by

$$\begin{aligned} \tau_{i0} = 1,\ \tau_{ij} &= \sum_{\ell=1}^{q} a_{i\ell}\tau_{\ell,j-1}, \quad 1 \le i \le q,\ j \ge 0, \\ \iff \tau_{ij} &= (A^j \boldsymbol{e})_i, \quad 1 \le i \le q,\ j \ge 0. \end{aligned} \tag{3.17}$$

Note that by (3.1.2) we may infer that

$$\tau_{ij} = (\tau_i)^j/j!, \quad 1 \le i \le q,\ 0 \le j \le q. \tag{3.18}$$

Also define, for $1 \le m \le q, 0 \le n \le M-1$

$$\Lambda_m W^n = \sum_{j=0}^{2q} \tau_{mj} k^j W^{(j)n}, \tag{3.19}$$

$$e^{n,m} = V^{n,m} - \Lambda_m W^n. \tag{3.20}$$

We now make a preliminary observation. By (3.14) and (3.20) we have

$$V^{n+1} = W^n + \sum_{m,j=1}^{q} b_m (A^{-1})_{mj} (\Lambda_j W^n - W^n) + b^T A^{-1} e^n, \tag{3.21}$$

where $e^n = (e^{n,1}, \dots, e^{n,q})^T \in (S_h)^q$. Using (3.19) and (3.17) we can write

$$\begin{aligned} \sum_{m,j=1}^{q} b_m (A^{-1})_{mj} (\Lambda_j W^n - W^n) &= \sum_{m,j=1}^{q} b_m (A^{-1})_{mj} \Big[\sum_{\ell=1}^{2q} (A^\ell \boldsymbol{e})_j k^\ell W^{(\ell)n} \Big] \\ &= \sum_{\ell=1}^{2q} (b^T A^{\ell-1} \boldsymbol{e}) k^\ell W^{(\ell)n} = \sum_{\ell=1}^{2q} k^\ell W^{(\ell)n}/\ell! \end{aligned}$$

where in the last equality we used the identities

$$b^T A^{\ell-1} \boldsymbol{e} = 1/\ell!, \quad 1 \le \ell \le 2q, \tag{3.22}$$

that follow from the fact that the rational approximation $r(z)$, cf. (3.2), corresponding to the $q-$stage Gauss–Legendre method is an $O(z^{2q+1})$ approximation to $\exp(z)$ as $z \to 0$. We conclude therefore by (3.21) and Taylor's theorem that

$$\|V^{n+1} - W^{n+1}\| \le ck^{2q+1} + \|b^T A^{-1} e^n\|.$$

Hence, in order to prove (3.15) our preliminary observation is that it is sufficient to obtain an estimate of the form

$$\|b^T A^{-1} e^n\| \le ck\big(k^{\min(2q,q+2)} + h^r\big). \tag{3.23}$$

Note that (3.20) when substituted in (3.14b) yields for $0 \le n \le M-1$

$$e^{n,j} = E^{n,j} + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} L_h^{n,\ell} e^{n,\ell}, \quad 1 \le j \le q, \tag{3.24}$$

where we have put

$$E^{n,j} = -\Lambda_j W^n + W^n + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} L_h^{n,\ell} \Lambda_\ell W^n, \quad 1 \le j \le q. \tag{3.25}$$

The first main step of the proof consists of some long intermediate calculations, inevitable in any order estimation of Runge–Kutta methods, aiming towards transforming the right-hand side of (3.25) to a form more suitable for our purposes. To simplify notation a bit, in the sequel by $\varphi = O(k^\lambda + h^\mu)$ we shall mean that there exists a constant $c > 0$ independent of k and h such that $\|\varphi\| \le c(k^\lambda + h^\mu)$ for k, h sufficiently small.

First, using (3.25), (3.5), (3.17), (3.19) we have for $1 \le j \le q$

$$\begin{aligned}
E^{n,j} &= -\Lambda_j W^n + W^n + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} L_h^{n,\ell} \Big(\sum_{m=0}^{2q-1} \tau_{\ell m} k^m W^{(m)n} \Big) \\
&\quad + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} \tau_{\ell,2q} k^{2q} L_h^{n,\ell} W^{(2q)n} \\
&= -\Lambda_j W^n + W^n + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} L_h^{n,\ell} \Big(\sum_{m=0}^{2q-1} \tau_{\ell m} k^m W^{(m)n} \Big) \\
&\quad + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} (L_h^{n,\ell} - L_h^n) \Big(\sum_{m=0}^{2q-1} \tau_{\ell m} k^m W^{(m)n} \Big) + O(k^{2q+1}) \\
&= -\sum_{m=1}^{2q} \tau_{jm} k^m W^{(m)n} + \sum_{m=1}^{2q} \Big(\sum_{\ell=1}^{q} a_{j\ell} \tau_{\ell,m-1} \Big) \mathrm{i}k^m L_h^n W^{(m-1)n} \\
&\quad + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} (L_h^{n,\ell} - L_h^n) \Big(\sum_{m=0}^{2q-1} \tau_{\ell m} k^m W^{(m)n} \Big) + O(k^{2q+1}) \\
&= -\sum_{m=1}^{2q} \tau_{jm} k^m \big(W^{(m)n} - \mathrm{i} L_h^n W^{(m-1)n} \big) \\
&\quad + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} (L_h^{n,\ell} - L_h^n) \Big(\sum_{m=0}^{2q-1} \tau_{\ell m} k^m W^{(m)n} \Big) + O(k^{2q+1}) \\
&= I_1^{n,j} + I_2^{n,j} + O(k^{2q+1}),
\end{aligned} \tag{3.26}$$

where

$$I_1^{n,j} := -\sum_{m=1}^{2q} \tau_{jm} k^m \big(W^{(m)n} - \mathrm{i} L_h^n W^{(m-1)n} \big), \quad 1 \le j \le q, \tag{3.27}$$

$$I_2^{n,j} := \mathrm{i}\sum_{\ell=1}^{q} a_{j\ell}(L_h^{n,\ell} - L_h^n)\Big(\sum_{m=0}^{2q-1}\tau_{\ell m}k^{m+1}W^{(m)n}\Big), \quad 1 \le j \le q. \tag{3.28}$$

In $I_1^{n,j}$ we have, using (1.8)

$$W^{(m)n} - \mathrm{i}L_h^n W^{(m-1)n} = \partial_t^{m-1}\big(W_t^n - \mathrm{i}\alpha\Delta_h W^n\big) - \mathrm{i}B_h^n W^{(m-1)n}. \tag{3.29}$$

Introducing now for $0 \le t \le T$ $\psi(t) := W_t - u_t - \mathrm{i}\beta(t)\big(W(t) - u(t)\big)$, for which (3.4) shows that $\psi^{(j)}(t) = O(h^r)$, we can write using (3.3) and (1.1) that

$$W_t - \mathrm{i}\alpha\Delta_h W = P\psi + \mathrm{i}B_h(t)W. \tag{3.30}$$

Differentiating (3.30) with respect to t and using (3.29) we rewrite (3.27) as

$$\begin{aligned} I_1^{n,j} &= -\mathrm{i}\sum_{m=1}^{2q}\tau_{jm}k^m\big[\partial_t^{m-1}(B_h^n W^n) - B_h^n W^{(m-1)n}\big] + O(kh^r)\\ &= -\mathrm{i}\sum_{m=2}^{2q}\tau_{jm}k^m\Big[\sum_{\ell=0}^{m-2}\binom{m-1}{\ell}B_h^{(m-1-\ell)n}W^{(\ell)n}\Big] + O(kh^r). \end{aligned} \tag{3.31}$$

Turning now to the term $I_2^{n,j}$ we obtain by (3.6) and Taylor's theorem

$$I_2^{n,j} = \mathrm{i}\sum_{\ell=1}^{q} a_{j\ell}\Big(\sum_{\mu=1}^{2q-1}(\tau_\ell k)^\mu B_h^{(\mu)n}/\mu!\Big)\Big(\sum_{m=1}^{2q}\tau_{\ell,m-1}k^m W^{(m-1)n}\Big) + O(k^{2q+1}). \tag{3.32}$$

Use of (3.5) yields for $1 \le \ell \le q$ that

$$\begin{aligned} Z_\ell^n &:= \sum_{j=1}^{2q-1}\sum_{m=1}^{2q}(\tau_\ell)^j\tau_{\ell,m-1}k^{j+m}B_h^{(j)n}W^{(m-1)n}/j!\\ &= \sum_{\lambda=2}^{2q}k^\lambda\Big[\sum_{m=1}^{\lambda-1}(\tau_\ell)^{\lambda-m}\tau_{\ell,m-1}B_h^{(\lambda-m)n}W^{(m-1)n}/(\lambda-m)!\Big] + O(k^{2q+1}). \end{aligned}$$

Therefore (3.32) gives for $1 \le j \le q$

$$\begin{aligned} I_2^{n,j} = \mathrm{i}\sum_{\ell=1}^{q} a_{j\ell}\Big\{\sum_{\lambda=2}^{2q}k^\lambda\Big[\sum_{m=1}^{\lambda-1}(\tau_\ell)^{\lambda-m}\tau_{\ell,m-1}B_h^{(\lambda-m)n}W^{(m-1)n}/(\lambda-m)!\Big]\Big\}\\ + O(k^{2q+1}). \end{aligned} \tag{3.33}$$

Summarizing, we obtain by (3.26), (3.31) and (3.33) for $1 \le j \le q$

$$\begin{aligned} E^{n,j} = \mathrm{i}\sum_{\lambda=2}^{2q}k^\lambda\Big[\sum_{m=1}^{\lambda-1}(\delta_j^{m,\lambda} - \gamma_j^{m,\lambda})B_h^{(\lambda-m)n}W^{(m-1)n}/(\lambda-m)!\Big]\\ + O(k^{2q+1} + kh^r), \end{aligned} \tag{3.34}$$

where, for $1 \le j \le q, \lambda = 2,\dots,2q, m = 1,\dots,\lambda-1$:

$$\delta_j^{m,\lambda} = \sum_{\ell=1}^{q} a_{j\ell}(\tau_\ell)^{\lambda-m}\tau_{\ell,m-1}, \quad \gamma_j^{m,\lambda} = \tau_{j\lambda}(\lambda-1)!/(m-1)!. \tag{3.35}$$

We immediately observe, using (3.18), that for $\lambda = 1, \dots, q+1, 1 \le m \le \lambda - 1$

$$\delta_j^{m,\lambda} = \sum_{\ell=1}^{q} a_{j\ell}(\tau_\ell)^{\lambda-1}/(m-1)! = \big((\lambda-1)!/(m-1)!\big) \sum_{\ell=1}^{q} a_{j\ell}\tau_{\ell,\lambda-1} = \gamma_j^{m,\lambda}.$$

Therefore, we finally conclude that if $q \ge 2$, (3.34) yields for $1 \le j \le q$

$$\begin{aligned} E^{n,j} = \mathrm{i} \sum_{\lambda=q+2}^{2q} k^\lambda \Big[\sum_{m=1}^{\lambda-1} (\delta_j^{m,\lambda} - \gamma_j^{m,\lambda}) B_h^{(\lambda-m)n} W^{(m-1)n}/(\lambda-m)! \Big] \\ + O(k^{2q+1} + kh^r), \end{aligned} \tag{3.36}$$

while, simply, if $q = 1$

$$E^{n,1} = O\big(k(k^2 + h^r)\big). \tag{3.37}$$

Note that (3.37) used in conjunction with

$$\max_{1 \le j \le q} \|e^{n,j}\| \le c \max_{1 \le j \le q} \|E^{n,j}\|, \tag{3.38}$$

(which follows from the stability estimate of Lemma 2.1 applied to the equations (3.24)) gives the desired estimate (3.23) in the case $q = 1$. Therefore we henceforth concentrate on the cases $q \ge 2$ for which $E^{n,j}$ is given by (3.36). To this end let for $1 \le j \le q$,

$$\varphi^{n,j} = \mathrm{i} \sum_{\lambda=q+2}^{2q} k^\lambda \Big[\sum_{m=1}^{\lambda-1} (\delta_j^{m,\lambda} - \gamma_j^{m,\lambda}) B_h^{(\lambda-m)n} W^{(m-1)n}/(\lambda-m)! \Big]. \tag{3.39}$$

Then (3.36) is written as

$$E^{n,j} = \varphi^{n,j} + O(k^{2q+1} + kh^r), \quad 1 \le j \le q. \tag{3.40}$$

Obviously, (3.5) gives that $\varphi^{n,j} = O(k^{q+2})$ since $\delta_j^{m,\lambda} - \gamma_j^{m,\lambda}$ are not zero in general if $\lambda \ge q+2$. (3.40) then gives that $E^{n,j} = O(k^{q+2} + kh^r)$ and (3.38) implies in turn that $e^{n,j} = O(k^{q+2} + kh^r)$ thus yielding the estimate $b^T A^{-1} e^n = O(k(k^{q+1} + h^r))$. (This concludes essentially the application of the idea of the consistency proof of [11] to the case at hand).

The second step in the proof is the improvement of the $q+1$ exponent of k to the better value $q+2$. For this purpose we use the fact that we must actually estimate not the individual $e^{n,j}$ but their particular linear combination

$$b^T A^{-1} e^n = \sum_{i,j=1}^{q} b_i (A^{-1})_{ij} e^{n,j}.$$

First note that defining

$$\tilde{e}^{n,j} = e^{n,j} - \varphi^{n,j}, \quad \tilde{E}^{n,j} = E^{n,j} - \varphi^{n,j} \tag{3.41}$$

we may rewrite (3.24) as

$$\tilde{e}^{n,j} = \Big(\tilde{E}^{n,j} + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} L_h^{n,\ell} \varphi^{n,\ell} \Big) + \mathrm{i}k \sum_{\ell=1}^{q} a_{j\ell} L_h^{n,\ell} \tilde{e}^{n,\ell}. \tag{3.42}$$

Recall from (3.40), (3.41) that $\tilde{E}^{n,j} = O(k^{2q+1} + kh^r)$. Also, (3.7) and (3.39) yield that $\|L_h^{n,j}\varphi^{n,j}\| \le ck^{q+2}, 1 \le j \le q$. Hence, the stability estimate of Lemma 2.1, applied to (3.42), yields $\tilde{e}^{n,j} = O(k^{q+3} + kh^r)$. Therefore, it follows from (3.41) that

$$b^T A^{-1} e^n = O(k(k^{q+2} + h^r)) + b^T A^{-1}\varphi^n, \tag{3.43}$$

where $\varphi^n = (\varphi^{n,1}, \dots, \varphi^{n,q})^T \in (S_h)^q$. Hence, the desired estimate (3.23) will follow from the fact that actually

$$b^T A^{-1}\varphi^n = 0, \tag{3.44}$$

which is a consequence of (3.39) and the following cancellation property of the Gauss–Legendre methods that we state as a separate lemma:

Lemma 3.1. *Let $\delta_j^{m,\lambda}, \gamma_j^{m,\lambda}$ be defined for $1 \le j \le q, \lambda = 2, \dots, 2q, m = 1, \dots, \lambda - 1$ by (3.35) and denote $\delta^{m,\lambda} = \left(\delta_1^{m,\lambda}, \dots, \delta_q^{m,\lambda}\right)^T$ and $\gamma^{m,\lambda} = \left(\gamma_1^{m,\lambda}, \dots, \gamma_q^{m,\lambda}\right)^T \in \mathbb{R}^q$. Then*

$$b^T A^{-1}(\delta^{m,\lambda} - \gamma^{m,\lambda}) = 0. \tag{3.45}$$

Proof. For $\lambda \le q + 1, 1 \le m \le \lambda - 1$, we have already established by the simple calculation following (3.35) that $\delta^{m,\lambda} = \gamma^{m,\lambda}$. Hence we restrict attention to the interesting case $q + 2 \le \lambda \le 2q, 1 \le m \le \lambda - 1$. Define for the purposes of this lemma $T \in \mathbb{R}^{q\times q}$ as $T = \operatorname{diag}(\tau_1, \dots, \tau_q)$ — no confusion with the T of (1.1) will arise — and note that (3.1.1), (3.1.2) can be written equivalently as

$$b^T T^{\ell-1}\boldsymbol{e} = 1/\ell, \quad 1 \le \ell \le 2q, \tag{3.1.1$'$}$$

$$AT^{\ell-1}\boldsymbol{e} = T^\ell \boldsymbol{e}/\ell, \quad 1 \le \ell \le q, \tag{3.1.2$'$}$$

respectively; (3.1.2′) implies that

$$A^\ell \boldsymbol{e} = T^\ell \boldsymbol{e}/\ell!, \quad 0 \le \ell \le q. \tag{3.1.2$''$}$$

Now, (3.17) gives that $\delta^{m,\lambda} = AT^{\lambda-m}A^{m-1}\boldsymbol{e}$. On the other hand, using (3.22), since $\lambda \le 2q$, we have $b^T A^{-1}\gamma^{m,\lambda} = (\lambda - 1)! b^T A^{\lambda-1}\boldsymbol{e}/(m - 1)! = (\lambda(m - 1)!)^{-1}$. Hence, to show (3.45) it suffices to establish

$$b^T T^{\lambda-m}A^{m-1}\boldsymbol{e} = (\lambda(m - 1)!)^{-1}, \quad q + 2 \le \lambda \le 2q,\ 1 \le m \le \lambda - 1. \tag{3.46}$$

Obviously for each $\lambda, q + 2 \le \lambda \le 2q$, (3.46) holds for $1 \le m \le q + 1$; this follows from (3.1.2′′) and (3.1.1′) that yield

$$b^T T^{\lambda-m}A^{m-1}\boldsymbol{e} = b^T T^{\lambda-1}\boldsymbol{e}/(m - 1)! = (\lambda(m - 1)!)^{-1}.$$

Hence we focus on the case $q + 2 \le \lambda \le 2q, q + 2 \le m \le \lambda - 1$. Using (3.1.2′′) again gives, since $m - 1 \ge q + 1$,

$$b^T T^{\lambda-m}A^{m-1}\boldsymbol{e} = b^T T^{\lambda-m}A^{m-1-q}T^q\boldsymbol{e}/q!. \tag{3.47}$$

For integers $k \ge 1, 1 \le \ell \le q$ define

$$F(\ell, k) = b^T T^{\ell-1}A^k T^q\boldsymbol{e} \tag{3.48}$$

and note that by (3.1.3)

$$
\begin{aligned}
F(\ell,k) &= b^T T^{\ell-1} A(A^{k-1}T^q\boldsymbol{e}) = \sum_{i=1}^q b_i\tau_i^{\ell-1}\sum_{j=1}^q a_{ij}(A^{k-1}T^q\boldsymbol{e})_j\\
&= \ell^{-1}\sum_{j=1}^q b_j(1-\tau_j^\ell)(A^{k-1}T^q\boldsymbol{e})_j\\
&= \ell^{-1}\big(b^TA^{k-1}T^q\boldsymbol{e} - b^TT^\ell A^{k-1}T^q\boldsymbol{e}\big)\\
&= \ell^{-1}\big[F(1,k-1)-F(\ell+1,k-1)\big], \quad k\ge 1, 1\le \ell\le q.
\end{aligned}\tag{3.49}
$$

We can now calculate directly using (3.1.1), (3.1.3) for integer s, such that $1\le s\le q-1$, that

$$
\begin{aligned}
F(s,1) &= \sum_{i=1}^q b_i\tau_i^{s-1}\sum_{j=1}^q a_{ij}\tau_j^q = s^{-1}\sum_{j=1}^q b_j(1-\tau_j^s)\tau_j^q\\
&= s^{-1}\Big[\sum_{j=1}^q b_j\tau_j^q - \sum_{j=1}^q b_j\tau_j^{s+q}\Big] = \big[(q+1)(s+q+1)\big]^{-1}.
\end{aligned}\tag{3.50}
$$

We claim now that

$$
F(\ell,k) = \big[(q+1)\cdots(q+k)(q+\ell+k)\big]^{-1} \quad \text{for } \ k\ge 1, \ell\ge 1, \ell+k\le q. \tag{3.51}
$$

Indeed (3.50) shows that (3.51) holds for $k=1$. For the inductive step assume that (3.51) holds for all $k\le k'$ and $1\le \ell\le q-k$. Then, for $\ell\le q-k'-1$ we have by (3.49) and the inductive hypothesis that

$$
\begin{aligned}
F(\ell,k'+1) &= \ell^{-1}\big[F(1,k')-F(\ell+1,k')\big]\\
&= \Big\{\big[(q+1)\cdots(q+k')(q+k'+1)\big]^{-1} - \big[(q+1)\cdots(q+k')(q+k'+\ell+1)\big]^{-1}\Big\}/\ell\\
&= \big[(q+1)\cdots(q+k')(q+k'+1)(q+\ell+k'+1)\big]^{-1}.
\end{aligned}
$$

This completes the inductive step and shows the validity of (3.51). Using now (3.48), (3.51) and (3.47) we obtain with $k:=m-1-q\ge 1, \ell:=\lambda-m+1\ge 1$, since $k+\ell=\lambda-q\le q$ in the region of interest, that for $q+2\le\lambda\le 2q, q+2\le m\le\lambda-1$,

$$
b^TT^{\lambda-m}A^{m-1}\boldsymbol{e} = \big[q!(q+1)\cdots(m-1)\lambda\big]^{-1} = \big[\lambda(m-1)!\big]^{-1}.
$$

This identity completes the proof of the validity of (3.46). ☐

As a consequence of (3.45) and (3.39), (3.44) holds and the proof of Theorem 3.1 is now complete. ☐

4. Practical implementation of the two-stage scheme

In this section we shall study questions related to the efficient implementation of the fully discrete scheme generated by the two-stage Gauss–Legendre method which is given of course by the tableau, [9], [6]:

$$\begin{array}{c|c} A & \tau \\ \hline b^T & \end{array} = \begin{array}{cc|c} \frac{1}{4} & \frac{1}{4}-\frac{1}{2\sqrt{3}} & \frac{1}{2}-\frac{1}{2\sqrt{3}} \\ \frac{1}{4}+\frac{1}{2\sqrt{3}} & \frac{1}{4} & \frac{1}{2}+\frac{1}{2\sqrt{3}} \\ \hline \frac{1}{2} & \frac{1}{2} & \end{array}. \tag{4.1}$$

With these values of a_{ij}, b_i, τ_i the method is

$$\begin{aligned} &(a) \quad U^0 = u_h^0, \\ &\quad\ \ \text{for } n = 0,1,\ldots,M-1: \\ &(b) \quad U^{n,m} = U^n + \mathrm{i}k\sum_{j=1}^{2} a_{mj}L_h^{n,j}U^{n,j}, \quad m = 1,2, \\ &(c) \quad U^{n+1} = U^n + \mathrm{i}k\sum_{j=1}^{2} b_j L_h^{n,j}U^{n,j}. \end{aligned} \tag{4.2}$$

Note that eliminating the $L_h^{n,j}U^{n,j}$ from (4.2b and c) and using the particular values of the constants a_{ij} and b_i from (4.1), we may write (4.2c) simply as

$$U^{n+1} = U^n + \sqrt{3}(U^{n,2} - U^{n,1}). \tag{4.2c$'$}$$

In sections 2 and 3 we proved that for each $0 \le n \le M$ (4.2) has a unique solution, that the resulting scheme is L^2-conservative and that it satisfies the optimal in space and time estimate

$$\max_{0\le n\le M} \|U^n - u^n\| \le c\big(k^4 + h^r\big). \tag{4.3}$$

Let $d = \dim S_h$. To determine $U^{n,1}, U^{n,2}$ from (4.2b) one should solve, after choosing a basis for S_h, the $2d \times 2d$ linear system

$$\mathbb{J}^n\mathbb{U}^n = \mathbb{F}^n, \tag{4.4}$$

where

$$\begin{gathered} \mathbb{J}^n = \mathbb{J}^n(t^{n,1}, t^{n,2}) = \begin{pmatrix} I - \mathrm{i}ka_{11}L_h^{n,1} & -\mathrm{i}ka_{12}L_h^{n,2} \\ -\mathrm{i}ka_{21}L_h^{n,1} & I - \mathrm{i}ka_{22}L_h^{n,2} \end{pmatrix}, \\ \mathbb{U}^n = (U^{n,1}, U^{n,2})^T, \quad \mathbb{F}^n = (U^n, U^n)^T. \end{gathered} \tag{4.5}$$

With the aim of solving only (sparse) $d \times d$ systems of linear equations at each time step, we shall *uncouple* the two equations in (4.4) borrowing an idea from [3]. We first write (4.4) equivalently as

$$\mathbb{J}^{*n}\mathbb{U}^n = (\mathbb{J}^{*n} - \mathbb{J}^n)\mathbb{U}^n + \mathbb{F}^n, \tag{4.6}$$

where

$$\mathbb{J}^{*n} = \mathbb{J}^n(t^{*n}, t^{*n}), \quad t^{*n} = (t^{n,1} + t^{n,2})/2 = t^n + k/2, \tag{4.7}$$

the advantage being now that the operators L_h in the entries of $\mathbb{J}^{*n}$ are evaluated at the same point t^{*n} and that, in particular, they commute. This enables us to compute the solution $\mathbb{Y} = (Y_1, Y_2)^T \in (S_h)^2$ of systems of the form

$$\mathbb{J}^{*n}\mathbb{Y} = \mathbb{Z}, \tag{4.8}$$

given $\mathbb{Z} = (Z_1, Z_2)^T \in (S_h)^2$, "explicitly" as

$$\begin{aligned} \varDelta^{*n}Y_1 &= Z_1 + \mathrm{i}kL^{*n}(a_{12}Z_2 - a_{22}Z_1), \\ \varDelta^{*n}Y_2 &= Z_2 + \mathrm{i}kL^{*n}(a_{21}Z_1 - a_{11}Z_2), \end{aligned} \tag{4.9}$$

where $\varDelta^{*n} = I - k(\mathrm{i}L^{*n})/2 + k^2(\mathrm{i}L^{*n})^2/12, L^{*n} = L_h(t^{*n})$. It is easily seen that $\varDelta^{*n}$ is invertible since $\mathrm{i}L^{*n}$ is normal with purely imaginary eigenvalues and the polynomial $1 - z/2 + z^2/12$ has no roots on the imaginary axis. Systems with operator $\varDelta^{*n}$ like the ones in (4.9) can then in turn be solved by the complex analog of the procedure proposed for Padé diagonal methods in [2]. Consider e.g. the first equation of (4.9) and, putting $W_1 = a_{12}Z_2 - a_{22}Z_1, R_1 = W_1 + Z_1/2 - \mathrm{i}kL^{*n}Z_1/12, \varPhi_1 = Y_1 - Z_1$, rewrite it as

$$\varDelta^{*n}\varPhi_1 = \mathrm{i}kL^{*n}R_1. \tag{4.10}$$

Since $\varDelta^{*n} = (I - \mathrm{i}\mu kL^{*n})(I - \mathrm{i}\bar{\mu}kL^{*n})$, where $\mu = 1/4 - \mathrm{i}\sqrt{3}/12$, we may rewrite (4.10), letting $H^n = I - \mathrm{i}\mu kL^{*n}, K^n = I - \mathrm{i}\bar{\mu}kL^{*n}$, as $H^nK^n\varPhi_1 = \mathrm{i}(H^n - K^n)R_1/(2\,\mathrm{Im}\,\mu)$, from which

$$\varPhi_1 = \mathrm{i}[(K^n)^{-1} - (H^n)^{-1}]R1/(2\,\mathrm{Im}\,\mu). \tag{4.11}$$

To determine therefore $\varPhi_1$ (and hence Y_1) one must form R_1 in the right-hand side of (4.11) and solve two complex linear systems of size $d \times d$ with operators K^n and H^n noting that at the level of matrix-vector operations the corresponding matrices will be sparse if a finite element basis consisting of functions of small support is chosen for S_h. Similarly for $\varPhi_2$ with obvious parallelicity duly noted.

In order to take advantage of the fact that systems of the form (4.8) can be "easily" solved in the manner outlined above, we shall solve the original system (4.6) by the simplest iterative method that its form immediately suggests. Let $j_n, n = 0, 1, \ldots, M$, be given positive integers — in practice $j_n = 1$ or $j_n = 2$ — representing the number of iterations that will be performed at step n to solve (4.6). For $j_n, n = 0, 1, \ldots, M$ we shall compute U_j^n approximating U^n as follows:

Let $U_{j_0}^0 = U^0 = u_h^0$ (e.g. take $j_0 = 0$).
For $n = 0, 1, \ldots, M$:
Compute suitable *starting values* $U_0^{n,1}, U_0^{n,2}$.
Compute $U_{j_{n+1}}^{n,m}, m = 1, 2$ by the iterative scheme:
For $j = 0, 1, \ldots, j_{n+1} - 1$ solve

$$\mathbb{J}^{*n}\begin{pmatrix} U_{j+1}^{n,1} \\ U_{j+1}^{n,2} \end{pmatrix} = (\mathbb{J}^{*n} - \mathbb{J}^n)\begin{pmatrix} U_j^{n,1} \\ U_j^{n,2} \end{pmatrix} + \begin{pmatrix} U_{j_n}^n \\ U_{j_n}^{n,2} \end{pmatrix}. \tag{4.12}$$

Define $U^{n+1}_{j_{n+1}} = U^n_{j_n} + \sqrt{3}(U^{n,2}_{j_{n+1}} - U^{n,1}_{j_{n+1}})$. Note that when $j = j_{n+1} - 1$ in the inner (j) loop in (4.12) there are important savings in the number of operations: since only the *difference* $U^{n,2}_{j_{n+1}} - U^{n,1}_{j_{n+1}}$ is finally needed, we must solve directly only *one* system of the form $\Delta^{*n}(\Phi_2 - \Phi_1) = \mathrm{i}kL^{*n}(R_2 - R_1)$. Hence it is important to try to get by with only $j_n = 1$ iteration at each time step.

We shall analyze the convergence of the iterative scheme in (4.12) and at the same time demonstrate the attendant stability of the resulting new fully discrete method. To this end assume for the time being that we have available approximations $U^n_{j_n} \in S_h$ to U^n for $n = 0, 1, 2, 3, 4$ such that

$$\|U^n - U^n_{j_n}\| \le c_n\big(k^4 + h^r\big), \quad n = 0, 1, 2, 3, 4. \tag{4.13}$$

For $n \ge 3$, given $U^n_{j_n}$ approximating U^n, we shall compute the required starting values $U^{n,m}_0, m = 1, 2$, for the inner (j) loop in (4.12) using extrapolation from previous values, i.e. as

$$U^{n,m}_0 = \sum_{\lambda=0}^{3} \mu_{m\lambda} U^{n-\lambda}_{j_{n-\lambda}}, \quad m = 1, 2, \quad 3 \le n \le M - 1, \tag{4.14}$$

where the constants $\mu_{m\lambda}, m = 1, 2, \lambda = 0, 1, 2, 3$, will be computed by letting $p(t)$ be the cubic polynomial interpolating to the values $y^{n-\lambda} = y(t^{n-\lambda})$ of a smooth function $y(t)$ at the points $t^{n-\lambda}, \lambda = 0, 1, 2, 3$, and setting

$$p(t^n + \tau_m k) = \sum_{\lambda=0}^{3} \mu_{m\lambda} y^{n-\lambda}. \quad m = 1, 2.$$

Proposition 4.1. *Let u be sufficiently smooth, $U^n, 0 \le n \le M$, be the solution of* (4.2) *and suppose that there exist $U^n_{j_n} \in S_h, 0 \le n \le 4$, such that* (4.13) *holds. For $4 \le n \le M-1$ define $U^{n+1}_{j_{n+1}}$ by the scheme* (4.12) *with starting values $U^{n,m}_0$ as in* (4.14)*. Then if $j_{n+1} = 2$ for all n, there exists $k_0 > 0$ such that for $k \le k_0$*

$$\max_{0 \le n \le M} \|U^n - U^n_{j_n}\| \le c\big(k^4 + h^r\big). \tag{4.15}$$

If $j_{n+1} = 1$ for all n, then, given $0 < \varepsilon \le 1$, there exists $k_\varepsilon > 0$ such that for $k \le k_\varepsilon$ there holds

$$\max_{0 \le n \le M} \|U^n - U^n_{j_n}\| \le c\big(k^{4-\varepsilon} + h^r\big). \tag{4.16}$$

In (4.15) *and* (4.16) *c is a constant independent of k, h and ε.*

Proof. Given $U^n_{j_n}$, let $\tilde{U}^{n,m}, m = 1, 2$, denote the exact solution of the system

$$\tilde{U}^{n,m} = U^n_{j_n} + \mathrm{i}k \sum_{j=1}^{2} a_{mj} L^{n,j}_h \tilde{U}^{n,j}, \quad m = 1, 2, \tag{4.17}$$

which we can write using (4.5) as

$$\mathbb{J}^n \begin{pmatrix} \tilde{U}^{n,1} \\ \tilde{U}^{n,2} \end{pmatrix} = \begin{pmatrix} U^n_{j_n} \\ U^n_{j_n} \end{pmatrix}. \tag{4.18}$$

First we prove a preliminary estimate that implies the convergence of the sequence $U_j^{n,m}, j = 0,1,2,\dots$ to $\tilde U^{n,m}$. From (4.12) and (4.18) suppressing n we let $Y_{j,m} = U_j^{n,m} - \tilde U^{n,m}$, $\mathbb{Y}_j = (Y_{j,1}, Y_{j,2})^T$ and obtain $\mathbb{J}^{*n}\mathbb{Y}_{j+1} = (\mathbb{J}^{*n} - \mathbb{J}^n)\mathbb{Y}_j =: \mathbb{Z}_j = (Z_{j,1}, Z_{j,2})^T$, i.e.

$$\mathbb{Y}_{j+1} = \mathrm{i}kA\mathbb{L}^{*n}\mathbb{Y}_{j+1} + \mathbb{Z}_j, \tag{4.19}$$

where $\mathbb{L}^{*n} = \operatorname{diag}\,(L^{*n}, L^{*n})$ on $(S_h)^2$. There follows by Lemma 2.1 that $|||\mathbb{Y}_{j+1}||| \le c|||\mathbb{Z}_j|||$. On the other hand, using e.g. (3.6), we obtain $\|Z_{j,m}\| \le ck^2\big(\|Y_{j,1}\| + \|Y_{j,2}\|\big), m = 1,2$ and conclude therefore that

$$\max_{m=1,2}\|U_{j+1}^{n,m} - \tilde U^{n,m}\| \le \gamma k^2 \max_{m=1,2}\|U_j^{n,m} - \tilde U^{n,m}\|, \quad j = 0,1,\dots,j_{n+1}-1, \tag{4.20}$$

where γ is a constant independent of h, k and the choice of j_n.

We next estimate the difference $\|U_0^{n,m} - \tilde U^{n,m}\|$. (In what follows we let $n \ge 4$). Using (4.14) we have for $m = 1,2$

$$\begin{aligned}\tilde U^{n,m} - U_0^{n,m} &= (\tilde U^{n,m} - U^{n,m}) + (U^{n,m} - u^{n,m}) + \Big(u^{n,m} - \sum_{\lambda=0}^{3}\mu_{m\lambda}u^{n-\lambda}\Big)\\ &\quad + \Big[\sum_{\lambda=0}^{3}\mu_{m\lambda}\big(u^{n-\lambda} - U^{n-\lambda}\big)\Big] + \Big[\sum_{\lambda=0}^{3}\mu_{m\lambda}\big(U^{n-\lambda} - U_{j_{n-\lambda}}^{n-\lambda}\big)\Big].\end{aligned} \tag{4.21}$$

First, since $\tilde U^{n,m}$ satisfies (4.17) and $U^{n,m}$ (4.2b) we obtain, again by stability (Lemma 2.1), that

$$\max_{m=1,2}\|\tilde U^{n,m} - U^{n,m}\| \le c\|U^n - U_{j_n}^n\|. \tag{4.22}$$

Next, recalling the definition of $V^{n,m}$ from (3.14), write, with $W^{n,m} = W(t^{n,m})$,

$$U^{n,m} - u^{n,m} = (U^{n,m} - V^{n,m}) + (V^{n,m} - W^{n,m}) + (W^{n,m} - u^{n,m}) \tag{4.23}$$

and observe that by (3.14b), (4.2b), Lemma 2.1 and (4.3)

$$\max_{m=1,2}\|U^{n,m} - V^{n,m}\| \le c\|U^n - W^n\| \le c(k^4 + h^r). \tag{4.24}$$

Now from (3.19), (3.20) $V^{n,m} - W^{n,m} = e^{n,m} + (\varLambda_m W^n - W^{n,m})$. By (3.38) and (3.36) we have $\|e^{n,m}\| \le c(k^4 + kh^r), m = 1,2$, whereas (3.18) and Taylor's theorem yield for $m = 1,2$

$$\|\varLambda_m W^n - W^{n,m}\| = \Big\|\sum_{j=0}^{4}\tau_{mj}k^jW^{(j)n} - W^{n,m}\Big\| \le ck^3.$$

We conclude that

$$\|V^{n,m} - W^{n,m}\| \le c(k^3 + kh^r), \quad m = 1,2, \tag{4.25}$$

and, therefore, by (4.23)–(4.25) that for $k \le 1$,

$$\|U^{n,m} - u^{n,m}\| \le c(k^3 + h^r), \quad m = 1,2. \tag{4.26}$$

Finally, the definition of the extrapolation procedure yields

$$\big\|u^{n,m} - \sum_{\lambda=0}^{3} \mu_{m\lambda} u^{n-\lambda}\big\| \le ck^4. \tag{4.27}$$

We conclude, by (4.21), (4.22), (4.26) and (4.27) that there exist positive constants η and θ, independent of k, h and the choice of j_n such that

$$\max_{m=1,2} \|\tilde{U}^{n,m} - U_0^{n,m}\| \le \theta(k^3 + h^r) + \eta\|U^n - U^n_{j_n}\| + \sum_{\lambda=0}^{3} \mu_\lambda \|U^{n-\lambda} - U^{n-\lambda}_{j_{n-\lambda}}\|, \tag{4.28}$$

where $\mu_\lambda = \max_{m=1,2} |\mu_{m\lambda}|, 0 \le \lambda \le 3$.

We come now to the main part of the proof which is an induction step on n. First we treat the case $j_n = 2$ for all $n \ge 4$. We let $k \le k_0 = \big(2\sqrt{3}\,\gamma^2\big)^{-1}$ where γ is as in (4.20). Assume that for $4 \le m \le n$,

$$\|U^m - U^m_{j_m}\| \le c_m(k^4 + h^r), \tag{H1}$$

where c_m are positive constants satisfying

$$c_m = [1 + k^3(\eta + \mu_0)]c_{m-1} + k^3(\mu_1 c_{m-2} + \mu_2 c_{m-3} + \mu_3 c_{m-4}) + \theta k^2. \tag{H2}$$

Clearly, it may be arranged, by taking c_3 or c_4 large enough in (4.13), that (H1)–(H2) hold for $n = 4$. Define $\tilde{U}^{n+1} \in S_h$, conformal to the notation in (4.17) as

$$\tilde{U}^{n+1} = U^n_{j_n} + \sqrt{3}(\tilde{U}^{n,2} - \tilde{U}^{n,1}) \tag{4.29}$$

and split

$$U^{n+1} - U^{n+1}_{j_{n+1}} = (U^{n+1} - \tilde{U}^{n+1}) + (\tilde{U}^{n+1} - U^{n+1}_{j_{n+1}}). \tag{4.30}$$

Since the time stepping procedure is conservative in the L^2 sense, subtracting (4.17) from (4.2b) and (4.29) from (4.2c$'$) we obtain

$$\|U^{n+1} - \tilde{U}^{n+1}\| = \|U^n - U^n_{j_n}\|. \tag{4.31}$$

On the other hand, using (4.29) and (4.12) we have

$$\begin{aligned} \|\tilde{U}^{n+1} - U^{n+1}_{j_{n+1}}\| &= \|\sqrt{3}(\tilde{U}^{n,2} - U^{n,2}_{j_{n+1}}) - (\tilde{U}^{n,1} - U^{n,1}_{j_{n+1}})\| \\ &\le 2\sqrt{3} \max_{m=1,2} \|\tilde{U}^{n,m} - U^{n,m}_{j_{n+1}}\| \\ &\le 2\sqrt{3}\,\gamma^2 k^4 \max_{m=1,2} \|\tilde{U}^{n,m} - U_0^{n,m}\| \end{aligned} \tag{4.32}$$

where in the last inequality we used the fact that $j_{n+1} = 2$ and (4.20). We conclude therefore by (4.30)–(4.32) that

$$\|U^{n+1} - U^{n+1}_{j_{n+1}}\| \le \|U^n - U^n_{j_n}\| + 2\sqrt{3}\,\gamma^2 k^4 \max_{m=1,2} \|\tilde{U}^{n,m} - U_0^{n,m}\|. \tag{4.33}$$

Hence, (4.33) and (4.28) yield, if $k \le k_0 = \big(2\sqrt{3}\,\gamma^2\big)^{-1}$

$$\|U^{n+1} - U^{n+1}_{j_{n+1}}\| \le (1 + \eta k^3)\|U^n - U^n_{j_n}\| + k^3 \sum_{\lambda=0}^{3} \mu_\lambda \|U^{n-\lambda} - U^{n-\lambda}_{j_{n-\lambda}}\| + \theta k^2(k^4 + kh^r).$$

Therefore, using the induction hypothesis (H1) we obtain

$$\begin{aligned}&\|U^{n+1} - U^{n+1}_{j_{n+1}}\|\\ &\leq \big[\big(1 + k^3(\eta + \mu_0)\big)c_n + k^3(\mu_1 c_{n-1} + \mu_2 c_{n-2} + \mu_3 c_{n-3}) + \theta k^2\big](k^4 + kh^r).\end{aligned}$$

Hence, (H1)–(H2) hold for $m = n+1$ as well and the inductive step is complete. Clearly the constants c_n are uniformly bounded for $0 \leq n \leq M$ by a constant independent of k and h; (4.15) follows.

We examine now the consequences of taking $j_n = 1$ for all $n \geq 4$. Given $0 < \varepsilon \leq 1$ we shall assume that $k \leq k_\varepsilon = \big(2\sqrt{3}\,\gamma\big)^{-1/\varepsilon}$ for reasons that will become apparent below. Our induction hypotheses that replace (H1), (H2) are now that for $4 \leq m \leq n$,

$$\text{(H1}'\text{)} \qquad \|U^m - U^m_{j_m}\| \leq c_m(k^{4-\varepsilon} + h^r),$$

where c_m are positive constants given by

$$\text{(H2}'\text{)} \qquad c_m = [1 + k^{2-\varepsilon}(\eta + \mu_0)]c_{m-1} + k^{2-\varepsilon}(\mu_1 c_{m-2} + \mu_2 c_{m-3} + \mu_3 c_{m-4}) + \theta k.$$

The verification of the inductive step follows the lines of the previous proof: (4.29) to (4.31) still hold of course but since $j_{n+1} = 1$, (4.32) becomes $\|\tilde{U}^{n+1} - U^{n+1}_{j_{n+1}}\| \leq 2\sqrt{3}\,\gamma k^2 \max_{m=1,2} \|\tilde{U}^{n,m} - U^{n,m}_0\|$. Consequently we have $\|U^{n+1} - U^{n+1}_{j_{n+1}}\| \leq \|U^n - U^n_{j_n}\| + 2\sqrt{3}\,\gamma k^2 \max_{m=1,2} \|\tilde{U}^{n,m} - U^{n,m}_0\|$, and therefore for $k \leq k_\varepsilon = \big(2\sqrt{3}\,\gamma\big)^{-1/\varepsilon}$ we obtain by (H1$'$)

$$\begin{aligned}\|U^{n+1} - U^{n+1}_{j_{n+1}}\| &\leq \|U^n - U^n_{j_n}\| + 2\sqrt{3}\,\gamma k^2\Big[\theta(k^3 + h^r) + \eta\|U^n - U^n_{j_n}\|\\ &\quad + \sum_{\lambda=0}^{3} \mu_\lambda \|U^{n-\lambda} - U^{n-\lambda}_{j_{n-\lambda}}\|\Big]\\ &\leq \big[\big(1 + (\eta + \mu_0)k^{2-\varepsilon}\big)c_n + k^{2-\varepsilon}(\mu_1 c_{n-1} + \mu_2 c_{n-2} + \mu_3 c_{n-3})\big](k^{4-\varepsilon} + h^r)\\ &\quad + \theta k(k^{4-\varepsilon} + h^r),\end{aligned}$$

verifying that (H1$'$), (H2$'$) hold for $m = n+1$ as well. Obviously the constants c_n can be made uniformly bounded in n by a constant independent of ε, albeit larger than the corresponding constant in the $j_n = 2$ case; we conclude that (4.16) holds. □

It is easy to construct $U^n_{j_n}, 0 \leq n \leq 4$, such that (4.13) is valid. For $n = 0$ we already have stipulated that $U^0_{j_0} = U^0$. For $n = 1$ only the previous value U^0 is available. We set $U^{0,1}_0 = U^{0,2}_0 = U^0$ in (4.12) and generate the sequence $U^{0,m}_j$ taking $j_1 = 2$. This will suffice since the analog of (4.27) is an $O(k)$ bound implying that in the analog of (4.28) $\max_{m=1,2} \|\tilde{U}^{0,m} - U^{0,m}_0\| = O(k + h^r)$. Similarly for $n = 2, 3, 4$ it suffices to generate $U^n_{j_n}$ from (4.12) with $j_n = 1$, computing for each $n, U^{n-1,m}_0$ as the appropriate linear combination of the previous values $U^0_{j_0}, U^1_{j_1}, \ldots, U^{n-1}_{j_{n-1}}$ using the Lagrange interpolating polynomial of degree $n - 1$.

In practice we noticed that taking cubic polynomial extrapolation to generate the starting values for $n \geq 4$ and just $j_n = 1$ was generally sufficient to preserve the overall order of accuracy and stability of the scheme. We report in [1] these and other relevant numerical experiments that we performed with the method, including experiments in which the operator $\mathbb{J}^{*n}$ is not evaluated at every time step but rather every $m^* > 1$ time steps. Our experiments indicate that it is possible to take in many interesting examples m^* equal to, say, 20 and $j_n = 2$ (the m^* "large", $j_n = 1$ combination was unstable for some hard to integrate problems) and still preserve the overall order of accuracy and stability of the scheme.

Note added in proof: For more recent work related to error estimates for the Nonlinear Schrödinger Equation we refer the reader to [24], [25].

References

1. G.D. Akrivis and V.A. Dougalis, *On a conservative, high-order accurate finite element scheme for the 'parabolic' equation*, to appear in the Proceedings for the 2nd IMACS symposium on computational acoustics, Princeton University, 15-17 March 1989.
2. G.A. Baker, J.H. Bramble and V. Thomée, *Single step Galerkin approximations for parabolic problems*, Math. Comp. **31** (1977) 818–847.
3. J.L. Bona, V.A. Dougalis, O.A. Karakashian and W. McKinney, *Conservative high order schemes for the generalized Korteweg-de Vries equation*, to appear.
4. A. Brocéhn, *Galerkin methods for approximation of solutions of second order partial differential equations of Schrödinger type*, Ph.D. Thesis, University of Göteborg, 1980.
5. J.C. Butcher, *Implicit Runge-Kutta processes*, Math. Comp. **18** (1964) 50–64.
6. J.C. Butcher, *The numerical analysis of ordinary differential equations; Runge-Kutta methods and general linear methods*, John Wiley, Chichester, 1987.
7. M. Crouzeix, *Sur la B-stabilité des méthodes de Runge-Kutta*, Numer. Math. **32** (1979) 75–82.
8. M. Crouzeix and V. Thomée, *On the discretization in time of semilinear parabolic equations with nonsmooth initial data*, Math. Comp. **49** (1987) 359–377.
9. K. Dekker and J.G. Verwer, *Stability of Runge-Kutta methods for stiff nonlinear differential equations*, North Holland, Amsterdam, 1984.
10. E. Dendy, Jr., *An alternating direction method for Schrödinger's equation*, SIAM J. Numer. Anal. **14** (1977) 1028–1032.
11. V.A. Dougalis and O.A. Karakashian, *On some high order accurate fully discrete Galerkin methods for the Korteweg-de Vries equation*, Math. Comp. **45** (1985) 329–345.
12. O.A. Karakashian and W. McKinney, *On optimal high order in time approximations for the Korteweg-de Vries equation*, Math. Comp. (to appear).
13. D. Lee and S.T. McDaniel, *Ocean acoustic propagation by finite difference methods*, Comput. Math. Appl. **14** No. 5.
14. D. Lee, R.L. Sternberg and M.H. Schultz, eds., *Computational acoustics: wave propagation*, Proceedings of the 1st IMACS symposium on computational acoustics, New Haven, 6–8 August 1986, vols.1,2, North Holland, Amsterdam, 1988.
15. J.L. Lions and E. Magenes, *Problèmes aux limites non homogènes et applications*, vol. 2, Dunod, Paris, 1968.
16. A. Quarteroni, *Mixed approximations of evolution problems*, Comput. Meths. Appl. Mech. Engrg. **24** (1980) 137–163.
17. J.M. Sanz-Serna, *Methods for the numerical solution of the nonlinear Schrödinger equation*, Math. Comp. **43** (1984) 21–27.
18. J.M. Sanz-Serna and J.G. Verwer, *Conservative and nonconservative schemes for the solution of the nonlinear Schrödinger equation*, IMA J. Numer. Anal. **6** (1986) 25–42.
19. M.H. Schultz and D. Lee, eds., *Computational ocean acoustics*, Invited lectures from the workshop held at Yale University, 1-3 August 1984, Comput. Math. Appl. **11** (1985) Nos 7-8.

20. F.D. Tappert, *The parabolic approximation method*. In: Wave propagation and underwater acoustics, J.B. Keller and J.S. Papadakis, eds., pp. 224–287, Lecture Notes in Physics v. 70, Springer-Verlag, Berlin-Heidelberg, 1977.
21. V. Thomée, *Convergence estimates for semi-discrete Galerkin methods for initial-value problems*. In: Numerische, insbesondere approximations-theoretische Behandlung von Funktionalgleichungen, R.Ansorge and W. Tornig, eds., pp. 243–262, Lecture Notes in Mathematics v. 333, Springer-Verlag, Berlin-Heidelberg, 1973.
22. V. Thomée, *Galerkin finite element methods for parabolic problems*, Lecture Notes in Mathematics v. 1054, Springer-Verlag, Berlin-Heidelberg, 1984.
23. L.B. Wahlbin, *A dissipative Galerkin method for the numerical solution of first order hyperbolic equations*. In: Mathematical aspects of finite elements in partial differential equations, C. de Boor, ed., pp. 147–169, Academic Press, New York, 1974.
24. G.D. Akrivis, V.A. Dougalis and O.A. Karakashian, *On fully discrete Galerkin methods of second-order temporal accuracy for the Nonlinear Schrödinger Equation*, to appear in Numer. Math.
25. O. Karakashian, G.D. Akrivis and V.A. Dougalis, *On optimal-order error estimates for the Nonlinear Schrödinger Equation*, to appear.

Mathematics Department, University of Crete, Iraklion, Greece

Mathematics Department, University of Crete, Iraklion, Greece