MATHEMATICS OF COMPUTATION
Volume 00, Number 0, Pages 000–000
S 0025-5718(XX)0000-0

# CONSERVATIVE, DISCONTINUOUS GALERKIN–METHODS FOR THE GENERALIZED KORTEWEG–DE VRIES EQUATION

J.L. BONA, H. CHEN, O. KARAKASHIAN, AND Y. XING

Abstract. We construct, analyze and numerically validate a class of conservative, discontinuous Galerkin schemes for the Generalized Korteweg–de Vries equation. Up to round-off error, these schemes preserve discrete versions of the first two invariants (the integral of the solution, usually identified with the mass, and the $L^2$–norm) of the continuous solution. Numerical evidence is provided indicating that these conservation properties impart the approximations with beneficial attributes, such as more faithful reproduction of the amplitude and phase of traveling–wave solutions. The numerical simulations also indicate that the discretization errors grow only linearly as a function of time.

## 1. Introduction

Considered here are the initial-boundary-value problems

$$\begin{cases} u_t + (u^{p+1})_x + \epsilon u_{xxx} = 0, & 0 < x < 1,\ t > 0, \\ u(x,0) = u^0(x), & 0 < x < 1, \end{cases} \tag{1.1}$$

for the Generalized Korteweg–de Vries equation, posed with periodic boundary conditions on the interval $[0,1]$, where $p$ is a non-negative integer and $\epsilon$ is a non-zero parameter. These evolution equations are among the simplest of a general class of models featuring nonlinear convection (the term $(p+1)u^p u_x$ in this case) and linear dispersion (the higher-order term $u_{xxx}$). This family of equations and others like them that feature nonlinearity and dispersion arise as mathematical models for the propagation of physical waves in a wide variety of situations (see *e.g.* [13, 35, 42, 15, 17, 23, 1]). The equations in (1.1) have also attracted attention because the mathematical theory pertaining to them is surprisingly interesting and subtle.

The initial-boundary-value problems (IBVP's henceforth) appearing in (1.1) are locally well posed in a wide range of function classes, including those that allow some of them to be justified as approximations of more complete models for physical phenomena (see [24, 30, 41, 20, 5] for theory in this direction). However, the resulting solutions do not always exist for all time. Singularity formation may occur and there are a few rigorous results in this direction as well (see [38, 37, 8]).

There remain puzzling, unresolved issues connected with singularity formation. A well-designed set of numerical simulations often provides helpful information in situations where rigorous results prove to be elusive. If smooth solutions form singularities during their evolution, they necessarily develop large values and large gradients (see [6]). Because of this attribute, solutions that form singularities in finite time are said to *blow up*. When singularities appear, they seem to form locally. Consequently, an idea that comes to the fore when designing a numerical scheme to investigate singularity formation is to implement time-dependent spatial refinements that are locally dense in regions where the solution is no longer of order one. Properly

1991 *Mathematics Subject Classification.* 65N12, 65N30, 35Q35, 35Q51, 35Q53, 35Q86, 76B15, 76B25 .

*Key words and phrases.* Discontinuous Galerkin methods, Korteweg–de Vries equations, error estimates, conservation laws.

The research of OK was supported in part by NSF Grant DMS-0811314. The research of YX was partially sponsored by the Office of Advanced Scientific Computing Research, US Department of Energy. The work was partly performed at the ORNL, which is managed by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725. JB and HC express thanks for the warm welcome they received during a visit to the Mathematics Department at the University of Tennessee at Knoxville to initiate this project. They also appreciate support from INRIA at the Université Bordeaux 1, France, during the concluding stage of the work. The authors also wish to thank the referees for a careful reading of the manuscript and valuable suggestions.

carried out, such a method has the possibility of retaining both stability and accuracy long enough for the structure of the singularity to become apparent.

The literature on numerical methods for the Korteweg–de Vries equation ((1.1) with $p = 1$) is vast, with finite difference, finite element, finite volume and spectral methods all having their proponents. The reader may consult [46, 47, 45, 40, 33, 7, 2, 48, 10, 32, 36] and the many references contained in these works for an introduction to the literature in this area. In the cases $p = 1, 2$, there are also an interesting class of non-standard methods based on the Inverse Scattering Transform (IST) (see [43, 44]). Certain of these schemes, *e.g.* [10, 43, 32, 44, 36], are conservative, meaning they preserve the discrete versions of continuous conservation laws for the equations. Experience shows that schemes preserving the discrete analogs of conservation laws appertaining to solutions of a partial differential equation often produce approximations that behave qualitatively like their continuous counterparts, in addition to featuring accuracy when the mesh size is sufficiently fine. Convergence results together with rigorous error bounds are available for some of the schemes mentioned above, but the extant analysis relies upon uniform spatial meshes. As indicated above, this is an assumption that probably should be avoided when tackling blowup issues. Indeed, previous numerical work described in [21] by two of the present authors and their collaborators has made it clear that capturing the blowup with uniform spatial and temporal grids is unlikely and that a successful approach to the simulation of blowing-up solutions of (1.1) will almost certainly require highly non-uniform meshes.

The Discontinuous Galerkin–method (DG–method henceforth) is a class of finite element approximations using discontinuous, piecewise polynomials as both the solution and test-function spaces (see [29] for a historical review). It combines advantages of both finite element and finite volume methods, including high order accuracy, high parallel efficiency, flexibility for hp-adaptivity and straightforward implementation on arbitrary meshes in geometries without any special symmetries. The DG-method has attracted considerable attention in the past two decades and has been applied successfully to produce good approximation of solutions to a wide range of partial differential equations, many of them arising in important applications areas. Particularly relevant for the present discussion is the fact that such schemes do not demand continuity at the spatial grid-points, and this allows a flexibility in making local refinements to an existing numerical grid not shared by continuous Galerkin methods.

The DG–method was originally introduced in the context of hyperbolic conservation laws. Later, the method was extended to deal with derivatives of order higher than one. Within the DG–framework, especially relevant to our development is the important body of work [27, 49, 50] on approximating solutions of evolution equations with higher-order derivatives using the Local Discontinuous Galerkin method (LDG–method) developed initially for the Korteweg–de Vries equation (KdV equation from now on) introduced by Yan and Shu [50]. The $L^2$–error estimates for the semi-discrete LDG methods for the KdV–equation were provided in [49]. Later, Cheng and Shu [27] proposed a new DG–method to solve directly time-dependent equations with higher-order spatial derivatives without the introduction of the auxiliary variables required by the LDG formulation. A key ingredient in this more recent method is a projection $\tilde{w}$ of the solution $u$ of (1.1) that is consistent with the dispersive term. This projection plays the role in approximating solutions of dispersive equations that the elliptic projection does in the context of parabolic or hyperbolic equations.

In the present paper, we construct a similar projection $w$ of the solution $u$ which, in addition to being consistent with the dispersive term, has the added advantage of being conservative. The new projection $w$ is used in the derivation of error estimates. Unlike $\tilde{w}$, the projection $w$ is global. Indeed, this is the way conservation is enforced. The non-locality of the projection leads to interesting analytical complications and necessitates the imposition of some additional assumptions on the mesh and the degree of the polynomials used in the approximation (see below). While one of these assumptions appears to be important in our context, the other does not. In any case, they are not particularly restrictive as far as simulating solutions of equations like (1.1) is concerned.

It is worthwhile commenting on the overarching intuition that guided this work. As mentioned already, the technical starting point was the methods introduced in [27]. A drawback of the ideas developed in this work is that the resulting schemes are dissipative (discussed in more detail presently). Dissipative schemes have an inherent problem with traveling waves possessing finite energy, as such methods constantly run down energy. In the case of nonlinear, dispersive wave equations, traveling waves subsist on a balance

being struck between nonlinear steepening and dispersive spreading. Dissipation destroys this balance. For the Generalized Korteweg–de Vries equations (GKdV–equations henceforth), it is known for a fact in case $p = 1, 2$ that arbitrary, finite-energy initial data resolves into traveling waves and a dispersive component. Numerical evidence (see again [21, 22]) indicates this to be true for other values of $p$ as well. Thus, dissipation introduced by the numerical scheme not only directly degrades accuracy, but it may result in the breakdown of the entire structure of a solution by destroying the traveling waves. In so far as this heuristic discussion has validity, it would appear wise to develop schemes that can integrate such traveling-wave solutions very accurately.

For the readers' benefit, the outline of the paper is sketched. Section 2 is devoted to notation and other preliminary material including the function spaces that are relevant to the analysis that follows. The discontinuous finite element spaces $V_h^q$ of degree $q \geq 2$ defined on a mesh $\mathcal{T}_h$ are then introduced. Based on these finite element spaces, conservative bilinear and multi-linear forms corresponding, respectively, to the dispersive and nonlinear terms in (1.1) can then be specified. These forms define operators which lead directly to a semi-discrete approximation (an approximation where the spatial variable is discrete, but the temporal variable remains continuous). The section is concluded by establishing existence and uniqueness of solutions to the semi-discrete approximations.

In Section 3, the projection $w$ which is consistent with the weak form for the dispersive term is constructed. It plays a central role in the subsequent development. As mentioned, the projection $w$ differs from the projection $\tilde{w}$ of [27] in that it maintains a conservative property that is enjoyed by the fully continuous problem. Propositions 3.1 and 3.2 constitute the technical core of the paper. Highlights of their content include the following.

(i) Existence of the projection $w$ is proved under the slightly unusual assumptions that the degree $q$ of the polynomials in the discontinuous finite element spaces is even and that the number of cells in the mesh $\mathcal{T}_h$ is odd.

(ii) Under the assumption that the number of adjacent cells of different length remains bounded as $h \downarrow 0$, it is shown that the projections $w$ are optimally close to solutions $u$ of the equation (1.1)

The technical requirements appearing in these lemmas stem from the global nature of the projection $w$. Section 5 contains numerical experiments designed in part to ascertain whether these conditions are essential or artifacts of the proofs. The evidence collected suggests somewhat surprisingly that the parity of $q$ has an effect on the convergence rates. The convergence rate for even $q$ appears to be $q + 1$ whilst, for odd values of $q$, the rate seems to be $q$. The weak regularity assumption on the mesh also has a bearing upon the numerical accuracy, whereas the requirement that the number of cells in $\mathcal{T}_h$ be odd was apparently not important as far as convergence rates and accuracy are concerned.

The principal convergence results are contained in Proposition 3.3 and Theorem 3.1. It is shown there that, under the above mentioned conditions, the semi-discrete approximation converges to the solution $u$ at the rate $O(h^q)$ as $h \downarrow 0$. This is suboptimal by one power of $h$, an effect owing to the treatment of the nonlinear term. This is the same rate obtained by Cheng and Shu for $q > 2$. When $q = 2$, we obtain here that the convergence is $O(h^2)$, whereas in [27] the error was only proved to be $O(h^{\frac{3}{2}})$ as $h \downarrow 0$.

In Section 4, some previously analyzed, semi-discrete dissipative schemes are reviewed and contrasted with the present method. There is also introduced a conservative, second order, temporal integration scheme. Applying this to a semi-discrete approximation yields a fully discrete numerical scheme.

A C–language program implementing the fully discrete scheme is used in Section 5 to effect the aforementioned numerical experiments. The outcomes of the experiments are discussed in some detail. A brief summary together with perspectives for future research concludes the paper.

## 2. The Numerical Approximation

Details of the numerical approximations are now set forth. This begins with a discussion of the spatial discretization which leads directly to a semi-discrete approximation of the continuous problem.

2.1. **The meshes.** Let $\mathcal{T}_h$ denote a partition of the real interval $[0,1]$ of the form $0 = x_0 < x_1 < \cdots < x_M = 1$. We will also say that $\mathcal{T}_h$ is a *mesh* on $[0,1]$. The points $x_m$ are called *nodes* while the intervals $I_m = [x_m, x_{m+1}]$ will be referred to as *cells*. The notation $x_m^- = x_m^+ = x_m$ will be useful in taking account, respectively, of left- and right-hand limits of discontinuous functions. The caveat followed throughout is that $x_0^- = x_M^-$ and $x_M^+ = x_0^+$ corresponding to the underlying spatial periodicity of the solutions being approximated.

The meshes $\mathcal{T}_h$ are taken to be *quasi-uniform*. This means that if $h_m = x_{m+1} - x_m$ and $h = h_{\max} = \max_m h_m$, then there is a positive constant $c$ such that, for all $m$,

$$0 < c \le \frac{h_m}{h}. \tag{2.1}$$

Additional constraints stemming from technical considerations are imposed on the mesh in Proposition 3.1.

2.2. **Function spaces.** In addition to the usual Sobolev spaces $W^{s,p} = W^{s,p}([0,1])$, repeated use will be made of the so-called *broken* Sobolev spaces $W^{s,p}(\mathcal{T}_h)$. These are the finite Cartesian products $\Pi_{I\in\mathcal{T}_h} W^{s,p}(I)$. Note that if $sp > 1$, the elements of $W^{s,p}(\mathcal{T}_h)$ are uniformly continuous when restricted to a given cell, but they may be discontinuous across nodes. For the purpose of quantifying these potential discontinuities, introduce the following notation: for $v \in W^{s,p}(\mathcal{T}_h)$, $s \ge 1$, let $v_m^+$ and $v_m^-$ denote the right-hand and left-hand limits, respectively, of $v$ at the node $x_m$. The *jump* $[v_m]$ (sometimes written $[v]_m$) of $v$ at $x_m$ is defined as $v_m^+ - v_m^-$. Similarly, the *average* $\{v_m\}$ (also denoted $\{v\}_m$) of $v$ at $x_m$ is $\frac{1}{2}(v_m^+ + v_m^-)$. These are all standard notations in the context of DG–methods. In all cases, the definitions are meant to adhere to the convention that $v_0^- = v_M^-$ and $v_M^+ = v_0^+$. Norms in the Sobolev classes $W^{s,p}$ will be denoted $\|\cdot\|_{W^{s,p}}$ or $\|\cdot\|_{W^{s,p}(I)}$ when the interval $I$ might be in doubt. In case the interval $I$ is clear from context, we will sometimes use an unadorned norm $\|\cdot\|$ to connote the $L^2(I)$–norm.

Use will also be made of the classes $L^p([0,T]; W^{s,r})$ of functions $u = u(x,t)$ which are measurable mappings from $[0,T]$ into $W^{s,r}$ and such that

$$\|u\|_{L^p([0,T];W^{s,r})} = \left( \int_0^T \|u(\cdot,\tau)\|_{W^{s,r}}^p \, d\tau \right)^{1/p} < \infty,$$

with the usual modification if $p = \infty$.

The following, basic embedding inequality (see [4]) will find frequent use in our development. For $v \in H^1(\mathcal{T}_h) = W^{1,2}(\mathcal{T}_h)$ and any cell $I \in \mathcal{T}_h$, there is a constant $c$ which is independent of the cell $I$ such that

$$\|v\|_{L^\infty(I)} \le c\Big( h_I^{-1/2} \|v\|_{L^2(I)} + h_I^{1/2} \|v_x\|_{L^2(I)} \Big), \tag{2.2}$$

where $h_I$ is the length of $I$. Indeed, the dependence of (2.2) on $h_I$ is easily ascertained by a simple scaling argument. Note that (2.2) may also be viewed as a trace inequality.

2.3. **The discontinuous polynomial spaces.** The spatial approximations will be sought in the space of discontinuous, piecewise polynomial functions $V_h^q$ subordinate to the mesh $\mathcal{T}_h$, *viz.*

$$V_h^q = \{ v \,|\, v\big|_{I_m} \in \mathcal{P}_q(I_m),\ m = 1, \cdots, M \}$$

where $\mathcal{P}_q$ is the space of polynomials of degree $q$ and $q \ge 2$. The spaces $V_h^q$ have well known, local approximation and inverse properties which are spelled out here for convenience (cf. [11], [25]). Let $q \ge 2$ be fixed and let $i, j$ be such that $0 \le j \le i \le q+1$. Then, for any cell $I$ and any $v$ in $H^j(I)$, there exists a $\chi \in \mathcal{P}_q(I)$ such that

$$|v - \chi|_{j,I} \le c h_I^{i-j} |v|_{i,I}, \tag{2.3}$$

where $|v|_{i,I}$ denotes the seminorm $\|v^{(i)}\|_{L^2(I)}$ on the Sobolev space $H^i(I)$ and the constant $c$ is independent of $h_I$. The above property continues to hold if the $L^p$–based Sobolev spaces replace the $L^2$–based classes $H^j$. In particular, it holds for the $L^\infty$ norm, which is to say, with $i, j$ as above, there is a $\chi \in \mathcal{P}_q(I)$ such that

$$|\partial_x^j (v - \chi)|_{L^\infty(I)} \le c h_I^{i-j} |\partial_x^i v|_{L^\infty(I)}. \tag{2.4}$$

The equally well-known inverse inequality

$$|\chi|_{j,I} \le c h_I^{-j} |\chi|_{0,I}, \tag{2.5}$$

for all $\chi \in \mathcal{P}_q(I)$ (see [25]) will also find frequent use.

2.4. **The weak formulation.** Multiplying the nonlinear term in the GKdV–equation (1.1) by $v \in H^1(\mathcal{T}_h)$, integrating the result over $[0,1]$ and integrating by parts cell by cell leads to the formula

$$\begin{aligned} \sum_{I\in\mathcal{T}_h} ((u^{p+1})_x, v)_I &= -\sum_{I\in\mathcal{T}_h} (u^{p+1}, v_x)_I + \sum_{m=0}^{M-1} \left[ (u_{m+1}^-)^{p+1} v_{m+1}^- - (u_m^+)^{p+1} v_m^+ \right] \\ &= -\sum_{I\in\mathcal{T}_h} (u^{p+1}, v_x)_I - \sum_{m=0}^{M-1} [u^{p+1} v]_m. \end{aligned} \tag{2.6}$$

Notice that the only way information (fluxes) can be transmitted between cells is through the jumps $[f(u)v]_m$ where $f(u) = u^{p+1}$. Information will be transmitted correctly if $u$ is smooth, e.g. if $u$ is the solution of the partial differential equation being studied here. However, if $u$ is a discontinuous approximation of a solution of (1.1), then the above formula need not feature correct transmission of information across the nodes and so $f(u) = u^{p+1}$ cannot be accurately reconstructed from its projection on the piecewise polynomial spaces. To counter this problem, it is standard to replace $f$ in the jump terms by a suitable function $\hat{f}$ which will insure the crucial requirement of consistency (correct transmission of information). Of course, it must also be the case that the choice of $\hat{f}$ will guarantee stability of the numerical scheme. In the present development, the choice of $\hat{f}$ is

$$\hat{f}(u_m^+, u_m^-) = \frac{1}{p+2} \sum_{j=0}^{p+1} (u_m^+)^{p+1-j} (u_m^-)^j. \tag{2.7}$$

This version of $\hat{f}$ leads directly to the nonlinear operator $\mathcal{N} : H^1(\mathcal{T}_h) \to V_h^q$ whose $L^2([0,1])$-inner product with any $v \in H^1(\mathcal{T}_h)$ is

$$(\mathcal{N}(u), v) = -\sum_{I\in\mathcal{T}_h} (u^{p+1}, v_x)_I - \sum_{m=0}^{M-1} \hat{f}(u_m^+, u_m^-)\, [v_m]. \tag{2.8}$$

The operator $\mathcal{N}$ is well defined by virtue of the Riesz Representation Theorem. The following, important consistency result obtains for this operator.

**Lemma 2.1.** *(i) The nonlinear term defined by* (2.8) *with the choice of* $\hat{f}$ *in* (2.7) *is consistent in the sense that for all 1-periodic functions* $u$ *in* $C^1([0,1])$*, there holds*

$$(\mathcal{N}(u), v) = ((u^{p+1})_x, v), \quad \forall v \in H^1(\mathcal{T}_h). \tag{2.9}$$

*(ii) The nonlinear term defined by* (2.8) *with the choice of* $\hat{f}$ *in* (2.7) *is conservative in the sense that*

$$(\mathcal{N}(v), v) = 0 \quad \forall v \in H^1(\mathcal{T}_h). \tag{2.10}$$

*Proof.* (i) For $u$ as specified above,

$$\hat{f}(u_m^+, u_m^-) = u^{p+1}(x_m), \quad m = 0, \cdots, M-1.$$

Thus,

$$\hat{f}(u_m^+, u_m^-)[v_m] = [u^{p+1} v]_m, \qquad m = 0, \cdots, M-1,$$

and (2.9) follows from (2.6).

(ii) To establish (2.10), it is suffices to notice that, on one hand,

$$\hat{f}(v_m^+, v_m^-)[v]_m = \frac{1}{p+2} [v^{p+2}]_m, \quad m = 0, \cdots, M-1,$$

and on the other, that

$$\sum_{I\in\mathcal{T}_h}(v^{p+1},v_x)_I=-\frac{1}{p+2}\sum_{m=0}^{M-1}[v^{p+2}]_m.$$

The proof of the Lemma is complete. □

To derive a bilinear form for the dispersive term, perform integration by parts twice to obtain

$$\sum_{I\in\mathcal{T}_h}(u_{xxx},v)_I=\sum_{I\in\mathcal{T}_h}(u_x,v_{xx})_I-\sum_{m=0}^{M-1}[u_{xx}v]_m+\sum_{m=0}^{M-1}[u_xv_x]_m. \tag{2.11}$$

There are many identities that can be used to express the jump terms appearing in (2.11). Indeed, for $\phi,\psi\in H^2(\mathcal{T}_h)$, we list three, among the possible ways of expressing $[\phi\psi]_m$, *viz.*

$$[\phi\psi]_m=\begin{cases}\phi_m^+[\psi]_m+[\phi]_m\psi_m^-,\\ \\ \phi_m^-[\psi]_m+[\phi]_m\psi_m^+,\\ \\ \{\phi\}_m[\psi]_m+[\phi]_m\{\psi\}_m.\end{cases} \tag{2.12}$$

These identities must be put into a context that ensures proper transmission of information (fluxes) across the nodes as well as stability and consistency with the IBVP (1.1). To this end, define the operator $\mathcal{D}:H^2(\mathcal{T}_h)\to V_h^q$ by

$$(\mathcal{D}(u),v)=\sum_{I\in\mathcal{T}_h}(u_x,v_{xx})_I-\sum_{m=0}^{M-1}\left(u_{xx}^+[v]_m-[u]_mv_{xx}^+\right)+\sum_{m=0}^{M-1}\{u_x\}_m[v_x]_m. \tag{2.13}$$

The next lemma delineates crucial properties of $\mathcal{D}$ that justify the particular form chosen in (2.13).

**Lemma 2.2.** *(i) The operator $\mathcal{D}$ defined by* (2.13) *is consistent in the sense that*

$$(\mathcal{D}(u),v)=(u_{xxx},v),\quad\forall v\in H^3(\mathcal{T}_h), \tag{2.14}$$

*is valid for all 1-periodic functions $u$ in $C^2([0,1])\cap H^3(\mathcal{T}_h)$.*

*(ii) The operator $\mathcal{D}$ defined by* (2.13) *is skew-adjoint, which is to say,*

$$(\mathcal{D}(v),v)=0\quad\forall v\in H^3(\mathcal{T}_h). \tag{2.15}$$

*Proof.* (i) With $u$ as specified, $[u]_m$, $[u_x]_m$ and $[u_{xx}]_m$ vanish. Thus, using the first identity in the display (2.12), we have $[u_{xx}v]_m=u_{xx}^+[v]_m+[u_{xx}]_mv_m^-=u_{xx}^+[v]_m-[u]_mv_{xx}^+$. Similarly, from the third identity in (2.12), one sees that $[u_xv_x]_m=\{u_x\}_m[v_x]_m+[u_x]_m\{v_x\}_m=\{u_x\}_m[v_x]_m$. The conclusion now follows from (2.11).

(ii) To establish (2.15), it is suffices to notice that the second sum on the right-hand side of (2.13) vanishes when $v=u$ and that

$$\sum_{I\in\mathcal{T}_h}(v_x,v_{xx})_I=\frac{1}{2}\sum_{I\in\mathcal{T}_h}\int_I\partial_x(v_x)^2\,dx=-\frac{1}{2}\sum_{m=0}^{M-1}[v_x^2]_m=-\sum_{m=0}^{M-1}\{v_x\}_m[v_x]_m.$$

The proof of the Lemma is complete. □

The semi-discrete approximation $u_h:[0,T]\to V_h^q$ of the solution $u$ of (1.1) is defined in terms of $\mathcal{N}$ and $\mathcal{D}$ by

$$\begin{aligned}(u_{ht},v)+(\mathcal{N}(u_h),v)\quad&+\quad\epsilon(\mathcal{D}(u_h),v)=0,\quad\forall v\in V_h^q,\ t\in[0,T],\\ u_h(0)\quad&=\quad Pu^0,\end{aligned} \tag{2.16}$$

where $P$ is a projection operator into $V_h^q$. Possible choices for $Pu^0$ are the $L^2$-projection of $u^0$ into $V_h^q$ or the Lagrange interpolant of $u^0$ in $V_h^q$. For both of these choices, the optimal estimate $\|u^0-u_h(0)\|=O(h^{q+1})$ as $h\downarrow 0$ obtains.

Expanding $u_h$ in terms of a basis for the finite-dimensional space $V_h^q$, it is readily seen that (2.16) is equivalent to a system whose independent variables are ordinary differential equations in the time-dependent coefficients of the expansion. It is immediate that this system has existence and uniqueness of a solution corresponding to given initial data $Pu^0$, at least locally in time. Global existence of the semi-discrete approximation will ensue as a byproduct of the following conservation law which implies appropriate *a priori* bounds.

**Theorem 2.1.** *The semi-discrete approximation $u_h$ satisfies*

$$\|u_h(t)\| = \|u_h(0)\| \tag{2.17}$$

*for all $t \geq 0$ for which the solution exists.*

*Proof.* Letting $v = u_h$ in (2.16) leads to

$$\frac{1}{2}\frac{d}{dt}\|u_h(t)\| + (\mathcal{N}(u_h), u_h) + \epsilon(\mathcal{D}(u_h), u_h) = 0.$$

The result follows at once from (2.10) and (2.15). □

Since all norms on $V_h^q$ are equivalent, the above result entails that $\|u_h\|_{L^\infty}$ is bounded for all $t > 0$ by a constant, which may of course depend on $h$. Since the restriction of $f$ to the space-time cylinder that contains $u_h$ is locally Lipschitzian, the existence of $u_h$ for all time follows.

## 3. Error estimates

For parabolic and hyperbolic equations, a centrally important tool used in deriving error estimates has been the so-called *Elliptic Projection* of the time-dependent solution $u$. Since the third derivative operator lacks the positivity property of elliptic operators, devising an appropriate projection for it turns out to be a little more subtle.

In an important contribution, Cheng and Shu constructed in [27] projection operators for a class of equations with third- and higher-order derivatives. One such projection, suitable for the GKdV–equation, is defined in the following manner. For $u \in H^3(\mathcal{T}_h)$, the projection $\tilde{w} \in V_h^q$ is specified by the conditions

$$\begin{array}{ll} (\tilde{w}, v)_I = (u, v)_I, & \forall v \in \mathcal{P}_{q-3}(I),\ I \in \mathcal{T}_h, \\ \tilde{w}(x_m^-) = u(x_m^-), & m = 1, \cdots, M, \\ \tilde{w}_x(x_m^+) = u_x(x_m^+), & m = 0, \cdots, M-1, \\ \tilde{w}_{xx}(x_m^+) = u_{xx}(x_m^+), & m = 0, \cdots, M-1. \end{array} \tag{3.1}$$

Note that for $q = 2$ the first condition is vacuous. Also, the definition is local to each cell. Hence, classical finite element approximation theory (see *e.g.* [25], [28]) can be brought to bear to show that $\tilde{w}$ is indeed well defined and that it is an optimal approximation to $u$ in the sense that

$$\|u - \tilde{w}\|_{W^{j,p}(I)} \leq c h_I^{q+1-j} |u|_{W^{q+1,p}(I)}, \quad I \in \mathcal{T}_h, \quad j = 0, 1, \quad p = 2, \infty. \tag{3.2}$$

where $h_I$ is the length of cell $I$. The projection $\tilde{w}$ defined by (3.1) is not consistent with the conservative approximation $\mathcal{D}$ defined by (2.13). That is to say, it is *not* the case that $(\mathcal{D}(\tilde{w}), v) = (\mathcal{D}(u), v), \quad \forall v \in V_h^q$.

This fact led us to define another projection $w$ of $u$ determined by the requirements

$$\begin{array}{ll} (w, v)_I = (u, v)_I, & \forall v \in \mathcal{P}_{q-3}(I),\ I \in \mathcal{T}_h, \\ w(x_m^-) = u(x_m^-), & m = 1, \cdots, M, \\ \{w_x\}_m = \{u_x\}_m = u_x(x_m), & m = 0, \cdots, M-1, \text{ or } m = 1, \cdots, M, \\ w_{xx}(x_m^+) = u_{xx}(x_m^+), & m = 0, \cdots, M-1. \end{array} \tag{3.3}$$

Note that the only difference between (3.3) and (3.1) is in the third equation. This seemingly minor change causes the construction of $w$ and the analysis of its properties to be more demanding. In fact, at present, we are able to show that $w$ exists only for even values of $q \geq 2$. Furthermore, the optimal approximation properties of $w$ require the imposition of some restrictions on the mesh $\mathcal{T}_h$ which will be spelled out later. Part of the difficulty resides in the fact that the averages $\{w_x\}_m$ force a coupling across cells which makes

the definition of $w$ a global one. However, it is straightforward to show that this new projection, when it exists, is consistent with the operator $\mathcal{D}$.

**Lemma 3.1.** *Let $u$ be smooth and 1-periodic. The projection $w$ defined by* (3.3) *satisfies*

$$(\mathcal{D}(w), v) = (\mathcal{D}(u), v), \quad \forall v \in V_h^q. \tag{3.4}$$

*Proof.* Integrating the term $\sum_{I\in\mathcal{T}_h}(w_x, v_{xx})_I$ in (2.13) by parts and using the second identity in (2.12) for the jumps $[wv_{xx}]_m$, it follows at once that

$$(\mathcal{D}(w), v) = -\sum_{I\in\mathcal{T}_h}(w, v_{xxx})_I - \sum_{m=0}^{M-1} w_{xx}^+(x_m)[v]_m + \sum_{m=0}^{M-1}\{w_x\}_m[v_x]_m - \sum_{m=0}^{M-1} w^-[v_{xx}]_m.$$

The conclusion of the lemma now follows from the definition of $w$. □

**Proposition 3.1.** *Suppose $u$ is sufficiently smooth and periodic. Further assume that $q \geq 2$ is even and that the number of cells in $\mathcal{T}_h$ is odd. Then, there exists a unique $w$ satisfying the conditions* (3.3). *The projection $w$ has the approximation properties*

$$\|u - w\|_{W^{j,p}(I)} \leq c h_I^{1-j}\left(\sum_{I\in\mathcal{T}_h^N} h_I^q \|u\|_{W^{q+1,\infty}(I)} + \sum_{I\in\mathcal{T}_h\setminus\mathcal{T}_h^N} h_I^{q+1}\|u\|_{W^{q+2,\infty}(I)}\right) \quad j = 0, 1, \quad p = 2, \infty, \tag{3.5}$$

*for a constant $c$ independent of $I$, where $\mathcal{T}_h^N$ is the set of cells whose length differs from at least one of its two immediate neighbors.*

*Proof.* We assume that $q \geq 4$. The case $q = 2$ falls to a similar, somewhat easier argument.

Let $\tilde{w}$ and $w$ be defined by (3.1) and (3.3), respectively, and let $e = w - \tilde{w}$. The quantity $e$ satisfies the conditions

$$\begin{array}{ll} (e, v)_I = 0, & \forall v \in \mathcal{P}_{q-3}(I), \forall I \in \mathcal{T}_h, \\ e(x_m^-) = 0, & m = 1, \cdots, M, \\ e_x(x_m^-) + e_x(x_m^+) = u_x(x_m) - \tilde{w}_x(x_m^-), & m = 0, \cdots, M-1, \text{ or } m = 1, \cdots, M, \\ e_{xx}(x_m^+) = 0, & m = 0, \cdots, M-1. \end{array} \tag{3.6}$$

For $\ell \geq 0$, let $P_\ell(t)$, be the usual Legendre polynomials that are orthogonal on $[-1, 1]$, normalized so that $P_\ell(1) = 1$. Given a cell $I_m = [x_m, x_{m+1}]$, consider the affine map

$$x = x(\xi) = \frac{h_m}{2}\xi + \frac{x_m + x_{m+1}}{2}, \quad -1 \leq \xi \leq 1, \tag{3.7}$$

that maps $[-1, 1]$ onto $I_m$. The family of rescaled Legendre polynomials $P_{m,\ell}(x)$ is defined by $P_{m,\ell}(x) = P_\ell(\xi)$ where $x$ and $\xi$ are related by (3.7). The polynomials $P_{m,\ell}$ are orthogonal with respect to the $L^2$-inner product on $I_m$.

Let $e_m$ denote the restriction of $e$ to $I_m$. The $e_m$'s can be expressed in terms of the rescaled Legendre polynomials thusly;

$$e_m(x) = \sum_{\ell=0}^{q} \alpha_{m,\ell} P_{m,\ell}(x) = \sum_{\ell=0}^{q} \alpha_{m,\ell} P_\ell(\xi), \quad m = 0, \cdots, M-1.$$

The first equation in (3.6) and the orthogonality of the Legendre polynomials imply that

$$\alpha_{m,\ell} = 0, \quad \ell = 0, \cdots, q-3, \quad m = 0, \cdots, M-1. \tag{3.8}$$

The second and fourth equations in display (3.6) my be used to solve for $\alpha_{m,q-2}$ and $\alpha_{m,q-1}$ in terms of $\alpha_{m,q}$ To accomplish this, use the identities

$$P_\ell'(\pm 1) = \frac{1}{2}(\pm 1)^{\ell-1}\ell(\ell+1), \qquad \ell = 1, \cdots, \tag{3.9}$$

$$P_\ell''(\pm 1) = \frac{1}{8}(\pm 1)^{\ell}(\ell-1)\ell(\ell+1)(\ell+2), \quad \ell = 2, \cdots, \tag{3.10}$$

which are easily proved by induction using the well-known recursion relations possessed by the Legendre polynomials.

From the second relation in (3.6), and taking account of the affine mapping defined in (3.7) and the normalization $P_\ell(1) = 1$, there follows immediately that

$$e(x_m^-) = e_{m-1}(x_m) = \alpha_{m-1,q-2} + \alpha_{m-1,q-1} + \alpha_{m-1,q} = 0, \quad m = 1, \cdots, M. \tag{3.11}$$

Similarly the fourth equation may be used to deduce that

$$\begin{aligned} e_{xx}(x_m^+) &= (q-3)(q-2)(q-1)q\,\alpha_{m,q-2} - (q-2)(q-1)q(q+1)\,\alpha_{m,q-1} \\ &\quad +(q-1)q(q+1)(q+2)\,\alpha_{m,q} = 0, \qquad m = 0, \cdots, M-1. \end{aligned} \tag{3.12}$$

Note that the factors $\frac{1}{8}$ and $h_m^{-2}$ which arises from taking the second derivative have been suppressed since they are of no importance when $e_{xx}$ is set equal to zero. The last two equations imply that

$$\alpha_{m,q-2} = -\frac{q(q+1)}{(q-2)(q-1)}\alpha_{m,q} \quad \text{and} \quad \alpha_{m,q-1} = \frac{2(2q-1)}{(q-2)(q-1)}\alpha_{m,q}, \quad m = 0, \cdots, M-1. \tag{3.13}$$

From the normalizations (3.9) and the third equation of (3.6), there obtains

$$\begin{aligned} &\tfrac{1}{h_{\ell-1}}\Big((q-2)(q-1)\,\alpha_{\ell-1,q-2} + (q-1)q\,\alpha_{\ell-1,q-1} + q(q+1)\,\alpha_{\ell-1,q}\Big) \\ &+\tfrac{1}{h_\ell}\Big( -(q-2)(q-1)\,\alpha_{\ell,q-2} + (q-1)q\,\alpha_{\ell,q-1} - q(q+1)\,\alpha_{\ell,q}\Big) = u_x(x_\ell) - \tilde{w}_x(x_\ell^-). \end{aligned} \tag{3.14}$$

for $m = 1, \cdots, M$ and $\ell \equiv m \bmod M$. Using the result of (3.13) in (3.14) leads to the system of equations

$$\hat{\alpha}_{\ell-1,q} + \hat{\alpha}_{\ell,q} = \frac{q-2}{2q(2q-1)}\Big(u_x(x_\ell) - \tilde{w}_x(x_\ell^-)\Big), \quad m = 1, \cdots, M, \ \ell \equiv m \bmod M, \tag{3.15}$$

where $\hat{\alpha}_{m,q} = \alpha_{m,q}/h_m$. The coefficient matrix of this system is an $M \times M$ circulant matrix with first row $[1, 1, 0, \cdots, 0]$. This matrix is invertible if and only if $M$ is odd, and in this case, its inverse is also circulant, having $\frac{1}{2}[1, -1, 1, -1, \cdots, -1, 1]$ as its first row. Thus, if $\eta_m = u_x(x_{m+1}) - \tilde{w}_x(x_{m+1}^-),\ m = 0, \cdots, M-1$, with $\eta_M := \eta_0$, then

$$\hat{\alpha}_{m,q} = \frac{q-2}{4q(2q-1)}\Big(\eta_m - \sum_{\ell \in \sigma_m}\big(\eta_\ell - \eta_{\ell+1}\big)\Big) \quad m = 0, \cdots, M-1, \tag{3.16}$$

where the index set $\sigma_m$ is such that each $\eta_\ell$ appears exactly once in the expression on the right-hand side of this formula.

It is clear from the inequalities in (3.2) that $|\eta_m| \le c h_m^q \|u^{q+1}\|_{L^\infty(I_m)}$. Proposition 3.2 (see below) also shows that $|\eta_\ell - \eta_{\ell+1}| \le c h_\ell^{q+1} \|u^{q+2}\|_{L^\infty(I_\ell \cup I_{\ell+1})}$ whenever $h_\ell = h_{\ell+1}$. It then follows that

$$|\alpha_{m,q}| \le c h_m \Big\{ h_m^q \|u^{q+1}\|_{L^\infty(I_m)} + \sum_{I \in \mathcal{T}_h^N} h_I^q \|u^{q+1}\|_{L^\infty(I)} + \sum_{I \in \mathcal{T}_h \setminus \mathcal{T}_h^N} h_I^{q+1} \|u^{q+2}\|_{L^\infty(I)} \Big\} \tag{3.17}$$

for $m = 0, \cdots, M-1$. Now in view of (3.8) and (3.13), all the $\alpha_{m,\ell}$'s satisfy (3.17). Finally, since $\|P_\ell\|_{L^\infty(0,1)} = \|P_{m,\ell}\|_{L^\infty(I_m)} \le c$ for some constant depending only on $\ell$, the estimate (3.5) for $p = \infty$, $j = 0$ follows from (3.2) and the triangle inequality. The case $p = 2$, $j = 0$ follows as a direct consequence. The remaining cases $p = 2, \infty$, $j = 1$ follow in turn from the bound $\|P'_{m,\ell}\|_{L^\infty(I_m)} \le c h_m^{-1}$. This concludes the proof. □

**Remark 3.1.** *Commentary is in order concerning the conditions imposed in the previous result.*

(i) *In contrast to the estimate* (3.2) *of the Cheng-Shu projection* (3.1)*, the bound* (3.5) *is not fully local due to the nonlocal nature of the projection. Also, it is suboptimal in terms of the regularity required in the proof.*

(ii) *For odd values of $q$, the left-hand side of* (3.15) *changes to $\hat{\alpha}_{\ell-1,q} - \hat{\alpha}_{\ell,q}$. The resulting circulant matrix is singular for all values of $M$. This is why $q$ is presumed to be even.*

(iii) *Notice that the proof of Propsition 3.1 depends upon the number $M$ of cells being odd. This is because of the periodicity required of the projection $w$. However, note also that the approximation $u_h$ can be determined whether or not $M$ is odd. Obviously, there is no problem connected with creating a mesh $\mathcal{T}_h$ with an odd number $M$ of cells. Moreover, this property is easily preserved in a process of repeated refinement or coarsening at later times in the temporal integration. Numerical experiments indicate that the convergence rates are the same, whether or not the mesh possesses an odd number of cells and so we have tentatively concluded this restriction is simply an artifact of our proof, which relies upon the projection.*

(iv) *For a uniform mesh, the parameter $\nu = \#\{\mathcal{T}_h^N\}$, the number of cells at least one of whose immediate neighbors have different lengths than does the cell in question, is zero and so the estimate* (3.5) *becomes optimal. On the other extreme, $\nu$ is bounded by the total number $M$ of cells in $\mathcal{T}_h$, in which case the estimate* (3.5) *becomes completely suboptimal. However it is possible, in fact straightforward, to achieve extremely local refinements while at the same time keeping $\nu$ quite small. This can be accomplished by implementing refinement in "patches", by which we mean a refinement wherein various subsets of contiguous cells are refined uniformly. This scheme of refinement is very well suited to the simulation of localized singularities.*

**Proposition 3.2.** *If $\tilde{w}$ is the projection of $u$ defined by* (3.1)*, then there are values $\zeta_{m,j}$, $j = 0, \cdots, q-2$ belonging to the cell $I_m$ such that*

$$\eta_m = u_x(x_{m+1}) - \tilde{w}_x(x_{m+1}^-) = h_m^q \sum_{j=0}^{q-2} \rho_j u^{(q+1)}(\zeta_{m,j}), \qquad m = 0, \cdots, M-1, \tag{3.18}$$

*where the constants $\rho_j$, $j = 0, \cdots, q-2$ depend only on $q$. Moreover, it transpires that*

$$|\eta_m - \eta_{m+1}| \le c h_m^{q+1} \|u^{q+2}\|_{L^\infty(I_m \cup I_{m+1})} \quad \text{whenever } h_m = h_{m+1}. \tag{3.19}$$

*Proof.* Consider the Legendre polynomial expansion

$$\tilde{w}_m(x) = \tilde{w}(x)|_{I_m} = \sum_{j=0}^{q} \alpha_{m,j} P_{m,j}(x),$$

of $\tilde{w}$ on $I_m$ and the Taylor expansion

$$u_m(x) = u(x)|_{I_m} = \sum_{j=0}^{q} \frac{u^{(j)}(x_m)}{j!}(x - x_m)^j + \frac{u^{(q+1)}(\xi_m(x))}{(q+1)!}(x - x_m)^{q+1},$$

of $u$ around $x_m$, where $\xi_m = \xi_m(x)$ lies in $I_m$. The first equation in (3.1) together with the preceding two formulas leads to

$$\begin{aligned} \int_{I_m} \tilde{w}(x)(x - x_m)^l dx &= \sum_{j=0}^{q} \alpha_{m,j} \int_{I_m} P_{m,j}(x)(x - x_m)^l dx \\ = \int_{I_m} u(x)(x - x_m)^l dx &= \sum_{j=0}^{q} \frac{u^{(j)}(x_m)}{j!(j+l+1)} h_m^{j+l+1} + \frac{u^{(q+1)}(\zeta_{m,l})}{(q+1)!(q+l+2)} h_m^{q+l+2}, \qquad l = 0, \cdots, q-3, \end{aligned} \tag{3.20}$$

where $\zeta_{m,l}$ is a point in cell $I_m$ obtained by utilizing the Mean Value Theorem for integrals. Using a change of variables related to the affine map (3.7) yields

$$\int_{I_m} P_{m,j}(x)(x - x_m)^l dx = \left(\frac{h_m}{2}\right)^{l+1} \int_{-1}^{1} P_j(\xi)(\xi+1)^l d\xi = \begin{cases} 0 & l < j \\ \dfrac{(l!)^2 h_m^{l+1}}{(l-j)!(l+j+1)!} & l \ge j \end{cases},$$

where the last equality is based on Rodrigues' formula that states $P_j(\xi) = \frac{1}{2^j j!} \frac{d^j}{d\xi^j}(\xi^2 - 1)^j$ and repeated use of integration by parts. Hence, the matrix whose elements are $\int_{I_m} P_{m,j}(x)(x - x_m)^l dx$ is lower triangular

and invertible. Consequently, (3.20) may be rewritten as

$$\alpha_{m,l} = \sum_{j=0}^{q} \beta_{j,l} h_m^j u^{(j)}(x_m) + h_m^{q+1} \sum_{j=0}^{q-3} \gamma_{j,l} u^{(q+1)}(\zeta_{m,j}), \qquad l = 0, \cdots, q-3. \tag{3.21}$$

where $\beta_{j,l}$ and $\gamma_{j,l}$ are constants that depend only on $q$.

The last three equations of (3.1), combined with the identities (3.9) for the Legendre polynomials, allow one to derive the formulas

$$\begin{aligned}
\tilde{w}(x_{m+1}^-) &= \sum_{j=0}^{q} \alpha_{m,j} = \sum_{j=0}^{q} \frac{u^{(j)}(x_m)}{j!} h_m^j + \frac{u^{(q+1)}(\zeta_{m,q-2})}{(q+1)!} h_m^{q+1} = u(x_{m+1}), \\
\tilde{w}_x(x_m^+) &= \sum_{j=0}^{q} \alpha_{m,j} P'_{m,j}(x_m) = \frac{2}{h_m} \sum_{j=0}^{q} \alpha_{m,j} P'_j(-1) = \frac{1}{h_m} \sum_{j=1}^{q} \alpha_{m,j} (-1)^{j-1} j(j+1) = u_x(x_m), \\
\tilde{w}_{xx}(x_m^+) &= \sum_{j=0}^{q} \alpha_{m,j} P''_{m,j}(x_m) = \frac{4}{h_m^2} \sum_{j=0}^{q} \alpha_{m,j} P''_j(-1) = \frac{1}{2h_m^2} \sum_{j=1}^{q} \alpha_{m,j} (-1)^j (j-1) j(j+1)(j+2) = u_{xx}(x_m).
\end{aligned}$$

These equations can be written as the linear system

$$\begin{bmatrix} 1 & 1 & 1 \\ -(q-2)(q-1) & (q-1)q & -q(q+1) \\ (q-3)(q-2)(q-1)q & -(q-2)(q-1)q(q+1) & (q-1)q(q+1)(q+2) \end{bmatrix} \begin{bmatrix} \alpha_{m,q-2} \\ \alpha_{m,q-1} \\ \alpha_{m,q} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=0}^{q} \frac{u^{(j)}(x_m)}{j!} h_m^j + \frac{u^{(q+1)}(\zeta_{m,q-2})}{(q+1)!} h_m^{q+1} \\ h_m u_x(x_m) \\ 2h_m^2 u_{xx}(x_m) \end{bmatrix} - \begin{bmatrix} \sum_{j=0}^{q-3} \alpha_{m,j} \\ \sum_{j=1}^{q-3} \alpha_{m,j} (-1)^{j-1} j(j+1) \\ \sum_{j=1}^{q-3} \alpha_{m,j} (-1)^j (j-1) j(j+1)(j+2) \end{bmatrix}$$

for the unknowns $\alpha_{m,q-2}$, $\alpha_{m,q-1}$ and $\alpha_{m,q}$. The determinant of this $3 \times 3$ matrix is $4q^2(q-1)^2(2q-1)$, hence it is invertible and we can therefore write $\alpha_{m,l}$ in the form

$$\alpha_{m,l} = \sum_{j=0}^{q} \beta_{j,l} h_m^j u^{(j)}(x_m) + h_m^{q+1} \sum_{j=0}^{q-2} \gamma_{j,l} u^{(q+1)}(\zeta_{m,j}), \qquad l = q-2, q-1, q.$$

It is then concluded that the equation

$$\begin{aligned}
\tilde{w}_x(x_{m+1}^-) &= \sum_{j=0}^{q} \alpha_{m,j} P'_{m,j}(x_{m+1}) = \frac{2}{h_m} \sum_{j=0}^{q} \alpha_{m,j} P'_j(1) = \frac{1}{h_m} \sum_{j=1}^{q} \alpha_{m,j} j(j+1) \\
&= \frac{1}{h_m} \sum_{j=0}^{q} \epsilon_j h_m^j u^{(j)}(x_m) + h_m^q \sum_{j=0}^{q-2} \rho_j u^{(q+1)}(\zeta_{m,j}),
\end{aligned} \tag{3.22}$$

holds, where $\epsilon_j$ and $\rho_j$ are constants that depend only on $q$. Since $\tilde{w}$ is an optimal approximation to $u$, (see (3.2)) it appears that

$$\tilde{w}_x(x_{m+1}^-) - u_x(x_{m+1}) = O(h_m^q) \quad \text{as} \quad h_m \downarrow 0.$$

For this relation to hold, the first term on the right-hand side of (3.22) must equal $u_x(x_{m+1})$. This establishes (3.18). Finally, when $h_m = h_{m+1}$, (3.18) allows the use of the Mean-Value Theorem to extract the additional factor of $h_m$. This concludes the proof. □

Attention is now turned to estimating the error $\|u_h(t) - u(t)\|$. The principal component of this task is the estimation of $u_h(t) - w(t)$.

**Proposition 3.3.** *Assume that the conditions of Proposition 3.1 hold and let $\zeta := u_h - w$, $\eta := w - u$ where $w$ is the projection of $u$ defined by (3.3). Suppose that for some $t^\star \in (0, T]$, it transpires that*

$$h^{-1} \|\zeta(t)\|_{L^\infty(0,1)} + \|\zeta_x(t)\|_{L^\infty(\mathcal{T}_h)} \leq 1, \quad \forall\ t \in [0, t^\star]. \tag{3.23}$$

*Then, for the same range $t \in [0, t^*]$, the inequality*

$$\|\zeta(t)\| \le Ce^{ct}\Big(\|\zeta(0)\| + h^q\Big), \tag{3.24}$$

*holds true, where the constants $C$ and $c$ depend on $p$ and $\|u\|_{L^\infty([0,t^*];W^{q+2,\infty}(0,1))}$.*

*Proof.* It follows from (1.1), (2.16), (2.9) and (2.14) that

$$(\zeta_t, v) + (\eta_t, v) + (\mathcal{N}(u_h) - \mathcal{N}(w), v) + (\mathcal{N}(w) - \mathcal{N}(u), v) + \epsilon(D(\zeta), v) = 0, \quad \forall v \in V_h^q. \tag{3.25}$$

In view of the skew-adjointness of $\mathcal{D}$ expressed in (2.15), if we set $v = \zeta$ in (3.25), there appears the differential equation

$$\frac{1}{2}\frac{d}{dt}\|\zeta\|^2 + (\eta_t, \zeta) + (\mathcal{N}(u_h) - \mathcal{N}(w), \zeta) + (\mathcal{N}(w) - \mathcal{N}(u), \zeta) = 0. \tag{3.26}$$

To begin, we observe that the mapping $u \to w := Pu$ defined by(3.3) is linear and thus commutes with the time differentiation operator, *viz.* $w_t = Pu_t$. Hence Proposition 3.1 implies that $\|w_t - u_t\| \le ch^q$, and so the bound

$$|(\eta_t, \zeta)| \le ch^q\|\zeta\| \tag{3.27}$$

emerges. The third and fourth terms on the left-hand side of (3.26) will be estimated separately.

Part I: Estimation of $(\mathcal{N}(u_h) - \mathcal{N}(w), \zeta)$.

In detail, the term $(\mathcal{N}(u_h) - \mathcal{N}(w), \zeta)$ is given by

$$\begin{aligned}(\mathcal{N}(u_h) - \mathcal{N}(w), \zeta) &= -\sum_{I\in\mathcal{T}_h}(u_h^{p+1} - w^{p+1}, \zeta_x)_I \\ &\quad -\frac{1}{p+2}\sum_{m=0}^{M-1}\sum_{j=0}^{p+1}\Big((u_h^+)_m^{p+1-j}(u_h^-)_m^j - (w^+)_m^{p+1-j}(w^-)_m^j\Big)[\zeta]_m \\ &:= \mathcal{E}_1 + \mathcal{E}_2.\end{aligned} \tag{3.28}$$

Since $u_h^{p+1} - w^{p+1} = \psi\zeta$ with $\psi := \sum_{j=0}^p u_h^{p-j}w^j$, integrating by parts yields

$$\mathcal{E}_1 = \frac{1}{2}\left(\sum_{I\in\mathcal{T}_h}(\psi_x, \zeta^2)_I + \sum_{m=0}^{M-1}[\psi\,\zeta^2]_m\right) := \mathcal{E}_1^{(1)} + \mathcal{E}_1^{(2)}. \tag{3.29}$$

Because $u_h = \zeta + w$, we can write $\psi = \sum_{j=0}^p\sum_{\ell=0}^{p-j}\binom{p-j}{\ell}\zeta^\ell w^{p-\ell}$. Hence, it follows from assumption (3.23) that

$$\big|\mathcal{E}_1^{(1)}\big| \le c\|\zeta\|^2. \tag{3.30}$$

for some constant $c$ depending only on $p$ and $u$ (through $w$).

It remains to obtain suitable bounds on the quantities $\mathcal{E}_1^{(2)}$ and $\mathcal{E}_2$. Both these terms contain powers of the form $\zeta^\ell$, $\ell = 2, \cdots, p+2$ with coefficients involving $w$. Using the trace/embedding inequality (2.2), the inverse inequality (2.5) and assumption (3.23) provides the inequalities

$$|\zeta_m^\pm|^\ell \le c|\zeta_m^\pm|^{\ell-2}h_m^{-1}\|\zeta\|_{\tilde{I}}^2 \le c\|\zeta\|_{\tilde{I}}^2, \quad \text{for } \ell \ge 3, \tag{3.31}$$

where $\tilde{I}$ is the union of the two cells to the right and left of the node $x_m$. To obtain a similar inequality for the quadratic powers of $\zeta$ requires a little more effort. The strategy is to combine parts of $\mathcal{E}_1^{(2)}$ and $\mathcal{E}_2$ to produce terms of the form $[w]\zeta^2$ with the jumps $[w]$ providing the needed extra degree of accuracy. What is

left after this sortie falls to the analysis leading to (3.35). In more detail, write

$$\begin{aligned}\mathcal{E}_1^{(2)} &= \frac{1}{2}\sum_{m=0}^{M-1}\sum_{j=0}^{p}\sum_{\ell=0}^{p-j}\binom{p-j}{\ell}\Big((w_m^+)^{p-\ell}(\zeta_m^+)^{\ell+2}-(w_m^-)^{p-\ell}(\zeta_m^-)^{\ell+2}\Big)\\ &= \frac{1}{2}(p+1)\sum_{m=0}^{M-1}\Big((w_m^+)^p(\zeta_m^+)^2-(w_m^-)^p(\zeta_m^-)^2\Big)+\mathcal{E}_1^{(4)} := \mathcal{E}_1^{(3)}+\mathcal{E}_1^{(4)},\end{aligned} \tag{3.32}$$

where $\mathcal{E}_1^{(4)}$ is an expression containing cubic and higher powers of $\zeta_m^+, \zeta_m^-$ with coefficients depending on $p$ and $w$. Just as in the argument leading to (3.31), it follows straightaway that

$$\big|\mathcal{E}_1^{(4)}\big| \le c\|\zeta\|^2. \tag{3.33}$$

Attention is turned now to $\mathcal{E}_2$. As in (3.32), write

$$\begin{aligned}\mathcal{E}_2 &= \frac{-1}{p+2}\sum_{m=0}^{M-1}\sum_{j=0}^{p+1}\sum_{\substack{\ell=0\\ k+\ell>0}}^{p+1-j}\sum_{k=0}^{j}\binom{p+1-j}{\ell}\binom{j}{k}(\zeta_m^+)^\ell(\zeta_m^-)^k(w_m^+)^{p+1-j-\ell}(w_m^-)^{j-k}\,[\zeta_m]\\ &= \frac{-1}{p+2}\sum_{m=0}^{M-1}\sum_{j=0}^{p}\Big((p+1-j)(w_m^+)^{p-j}(w_m^-)^j\zeta_m^+ + (j+1)(w_m^+)^{p-j}(w_m^-)^j\zeta_m^-\Big)[\zeta_m]+\mathcal{E}_2^{(2)}\\ &:= \mathcal{E}_2^{(1)}+\mathcal{E}_2^{(2)},\end{aligned} \tag{3.34}$$

where $\mathcal{E}_2^{(2)}$ is an expression containing cubic and higher powers of $\zeta_m^+, \zeta_m^-$ with coefficients depending on $p$ and $w$, and which therefore obeys the estimate

$$\big|\mathcal{E}_2^{(2)}\big| \le c\|\zeta\|^2. \tag{3.35}$$

What is left now is $\mathcal{E}_1^{(3)}$ and $\mathcal{E}_2^{(1)}$, which are estimated together. Indeed, noting that $\sum_{j=0}^{p}(p+1-j) = \sum_{j=0}^{p}(j+1) = \frac{1}{2}(p+1)(p+2)$, it follows that

$$\begin{aligned}\mathcal{E}_1^{(3)}+\mathcal{E}_2^{(1)} &= \frac{1}{p+2}\sum_{m=0}^{M-1}\sum_{j=0}^{p}\Big((p+1-j)\Big((w_m^+)^p-(w_m^+)^{p-j}(w_m^-)^j\Big)(\zeta_m^+)^2\\ &\quad +(p-2j)(w_m^+)^{p-j}(w_m^-)^j\zeta_m^+\zeta_m^- - (j+1)\Big((w_m^-)^p-(w_m^+)^{p-j}(w_m^-)^j\Big)(\zeta_m^-)^2\Big).\end{aligned} \tag{3.36}$$

In view of the range of the index $j$, it is clear that the $w$ terms in the first and third sums can be expressed, independently of each other, as $g_j(w^+, w^-)[w]$ for some functions $g_j$. On the other hand, in view of the range of values of $p-2j$ for $j = 0, \cdots, p$, the second sum is also seen to contain terms, each of which has the jump $[w]$ as a factor. On the other hand, (3.5) implies that $|[w]_m| = |[u-w]_m| \le ch_m^{q+1}$. We thus deduce that

$$\mathcal{E}_1^{(3)}+\mathcal{E}_2^{(1)} \le c\|\zeta\|^2. \tag{3.37}$$

Gathering together (3.30), (3.33), (3.35) and (3.37) leads to the conclusion

$$|(\mathcal{N}(u_h)-\mathcal{N}(w),\zeta)| \le c\|\zeta\|^2, \tag{3.38}$$

where the constant $c$ depends only on $p$ and $u$.

Part II: Estimation of $(\mathcal{N}(w)-\mathcal{N}(u),\zeta)$.

Note that this term also satisfies (3.28) with $u_h$ replaced by $u$. Defining $\eta = w - u$ and letting $\psi$ be given by $\psi = \sum_{j=0}^{p} u^{p-j} w^j$, it transpires after integration by parts that

$$
\begin{aligned}
(\mathcal{N}(w) - \mathcal{N}(u), \zeta) \quad &= \quad \sum_{I \in \mathcal{T}_h} \Big( (\psi\eta)_x, \zeta \Big)_I + \sum_{m=0}^{M-1} [\psi\, \eta\, \zeta]_m \\
&\quad -\frac{1}{p+2} \sum_{m=0}^{M-1} \sum_{j=0}^{p+1} \sum_{\substack{\ell=0 \\ k+\ell>0}}^{p+1-j} \sum_{k=0}^{j} \binom{p+1-j}{\ell} \binom{j}{k} (\eta_m^+)^\ell (\eta_m^-)^k (u_m^+)^{p+1-j-\ell} (u_m^-)^{j-k}\, [\zeta_m] \qquad (3.39) \\
&:= \quad \mathcal{E}_1 + \mathcal{E}_2 + \mathcal{E}_3.
\end{aligned}
$$

It follows from the estimate (3.5) that

$$
\big|\mathcal{E}_1\big| \le c h^q \|\zeta\|, \qquad (3.40)
$$

for some constant depending on $p$ and $u$. Similarly, applying the trace/embedding inequality (2.2), there obtains

$$
\big|\mathcal{E}_2\big| + \big|\mathcal{E}_3\big| \le c h^q \|\zeta\|. \qquad (3.41)
$$

Finally, estimate (3.24) of the theorem follows from combining the bounds (3.38), (3.40), (3.41) in (3.26) together with an application of Gronwall's Lemma. □

The groundwork has been laid for stating and proving the main convergence result for the semi-discrete approximation $u_h$ defined in (2.16)

**Theorem 3.1.** *Assume that the solution of* (1.1) *is sufficiently regular and that* $u_h(0)$ *is chosen to satisfy* $\|u^0 - u_h(0)\| = O(h^q)$ *as* $h \downarrow 0$. *Then, there exists* $h_0 > 0$ *depending on* $u, p$ *and* $T$ *and constants* $C$ *and* $c$ *such that for all* $h$ *in the range* $(0, h_0]$, *the inequality*

$$
\|(u - u_h)(t)\| \le C e^{ct} h^q, \quad \text{for} \quad 0 \le t \le T, \qquad (3.42)
$$

*holds, where as in Proposition 3.3 the constants* $C$ *and* $c$ *depend on* $p$ *and* $\|u\|_{L^\infty([0,T];W^{q+2,\infty}(0,1))}$.

*Proof.* To begin, note that by virtue of (3.5) and the triangle inequality, it is sufficient to prove the estimate

$$
\|\zeta(t)\| \le C e^{ct} h^q, \quad \text{for} \quad 0 \le t \le T \qquad (3.43)
$$

for suitable constants $C$ and $c$ and all $h$ sufficiently small. Another application of the triangle inequality yields that

$$
\|\zeta(0)\| \le \|u_h(0) - u^0\| + \|u^0 - w(0)\| \le c h^q, \qquad (3.44)
$$

holds for some constant $c$ depending only on $u^0$. By virtue of the embedding/trace and inverse inequalities (2.2) and (2.5), the inequality

$$
h^{-1} \|\zeta(0)\|_{L^\infty(0,1)} + \|\zeta_x(0)\|_{L^\infty(\mathcal{T}_h)} \le c h^{q-3/2} \qquad (3.45)
$$

is valid with a constant depending, as before, only upon $u_0$. Since $q \ge 2$, there exists $h_1$ depending only on $u^0$ through the constants in the above two inequalities such that

$$
h^{-1} \|\zeta(0)\|_{L^\infty(0,1)} + \|\zeta_x(0)\|_{L^\infty(\mathcal{T}_h)} \le \frac{1}{2}, \quad \text{for} \quad 0 < h \le h_1. \qquad (3.46)
$$

For $h \le h_1$, let $t^\star(h) := \sup\{t \ge 0 : \text{assumption (3.23) holds at } t\}$. In view of (3.46) and the fact that $\zeta$ is a continuous function of $t$, it is necessarily the case that $t^\star(h) > 0$. Consequently, the inequality (3.24) must hold for all $t \in [0, t^\star(h))$. Taking account of (3.44), it follows that (3.43) must also hold up to $t^\star(h)$. If $t^\star(h) \ge T$, the theorem is proved with $h_0 = h_1$. Otherwise, choose $h_0$ with $0 < h_0 \le h_1$ such that

$$
2\tilde{c} \max_{0 \le t \le T} \big\{ C e^{ct} \big\} h_0^{q-3/2} \le 1. \qquad (3.47)
$$

where the constants $C, c$ are those appearing in (3.43) and $\tilde{c}$ is a constant that depends on the constants in (2.1), (2.2) and (2.5). We will show that for all $h \le h_0$, $t^\star(h) \ge T$. Indeed, suppose to the contrary that

$t^\star(h) < T$ for some $h \le h_0$. Now with (3.43) holding up to $t = t^\star(h)$, the inequalities (2.1), (2.2) and (2.5) yield that

$$\begin{aligned}\max_{0\le t\le t^\star(h)} \Big(h^{-1}\|\zeta(t)\|_{L^\infty(0,1)} + \|\zeta_x(t)\|_{L^\infty(\mathcal{T}_h)}\Big) &\le 2\tilde{c} \max_{0\le t\le t^\star(h)} \big\{Ce^{ct}\big\} h_0^{q-3/2} \\ &< 2\tilde{c} \max_{0\le t\le T} \big\{Ce^{ct}\big\} h_0^{q-3/2} \le 1.\end{aligned}$$

which would contradict the assumption that $t^\star(h) < T$. □

## 4. Semi-conservative, dissipative and fully discrete schemes

There are semi-discrete schemes where one or both of the nonlinear and dispersive approximations $\mathcal{N}$ and $\mathcal{D}$ are defined in such a way that they are consistent, but dissipative, e.g. those outlined in [27]. A dispersive approximation $\mathcal{D}$ attuned to the projection $\tilde{w}$ introduced by (3.1) is defined by

$$(\mathcal{D}(u), v) = \sum_{I\in\mathcal{T}_h} (u_x, v_{xx})_I - \sum_{m=0}^{M-1} \Big(u^+_{xx}[v]_m - [u]_m v^+_{xx}\Big) + \sum_{m=0}^{M-1} u^+_x [v_x]_m. \tag{4.1}$$

This operator is dissipative since

$$(\mathcal{D}(v), v) = \frac{1}{2}\sum_{m=0}^{M-1} [v_x]^2_m \ge 0, \tag{4.2}$$

as follows from the same sort of calculation as that appearing in the proof of Lemma 2.2.

Dissipative counterparts of the nonlinear operator $\mathcal{N}$, e.g. those considered in [27], can be constructed using one of many monotone numerical fluxes developed in the context of hyperbolic conservation laws. Typical monotone numerical fluxes are exact or approximate Riemann solvers, including upwinding, Lax-Friedrichs-, Godunov-, Boltzmann- and Harten-Lax-Van Leer-type. As in [27], consider a continuous function $\hat{f}(u^+, u^-)$ which is nonincreasing in $u^+$, nondecreasing in $u^-$ and satisfies $\hat{f}(u, u) = u^{p+1}$. For instance, we could use the *upwind flux*

$$\hat{f}(u^+, u^-) = (u^-)^{p+1} \tag{4.3}$$

if the solution $u$ of (1.1) happens always to be positive, which is the case for the two exact solutions tested in this paper. It is easy to see that the corresponding operator $\mathcal{N}$ is dissipative in the sense that the multi-linear form $(\mathcal{N}(v), v) \ge 0, \forall v \in H^1(\mathcal{T}_h)$. Together with (4.2), the latter inequality implies global existence of the semi-discrete approximation $u_h$ defined by these schemes. In particular, we have $\|u_h(t)\| \le \|u_h(0)\|$ for all $t > 0$.

Error estimates which are $O(h^q)$ as $h \downarrow 0$ can also be established for these dissipative semi-discrete schemes by comparing $u_h$ to $\tilde{w}$. In this case, the existence and approximation properties of the projection are clear and the conditions on the mesh required by Proposition 3.1 are not needed.

4.1. **A conservative, fully discrete scheme.** Not just any time-stepping method employed in a fully discrete scheme will preserve the conservation properties of the semi-discrete approximations. A family of temporal integrators having arbitrarily high order in time and which does preserve the conservation laws up to round-off error is the implicit Runge-Kutta collocation type methods associated with the diagonal elements of the Padé table for $e^z$ (see e.g. [31]).

In this paper, we consider the first two members of this family of conservative schemes. Let $0 = t_0 < t_1 < \cdots < t_N = T$ be a partition of the interval $[0, T]$ and $\kappa_n = t_{n+1} - t_n$. The fully discrete second-order in time approximations $u^n$ to $u(\cdot, t_n)$ are constructed using the *midpoint rule* in the following manner. Let $u^0 = u_h(\cdot, 0)$, and for $n = 0, \cdots, N-1$, let $u^{n+1} \in V_h^q$ be defined as

$$u^{n+1} = 2u^{n,1} - u^n, \tag{4.4}$$

where $u^{n,1}$ is the solution of the equation

$$u^{n,1} - u^n + \frac{\kappa_n}{2}\Big(\mathcal{N}(u^{n,1}) + \epsilon\mathcal{D}(u^{n,1})\Big) = 0. \tag{4.5}$$

Some of the numerical experiments to be reported presently that evidenced very small spatial errors were conducted using fourth-order in time approximations constructed in the following manner: let

$$u^{n+1} = u^n + \sqrt{3}\left(u^{n,2} - u^{n,1}\right), \tag{4.6}$$

with $u^{n,1}$ and $u^{n,2}$ given as solutions of the coupled system of equations

$$u^{n,1} - u^n + \kappa_n\Big(a_{11}f^{n,1} + a_{12}f^{n,2}\Big) = 0, \tag{4.7}$$

$$u^{n,2} - u^n + \kappa_n\Big(a_{21}f^{n,1} + a_{22}f^{n,2}\Big) = 0, \tag{4.8}$$

where $f^{n,i} = \mathcal{N}(u^{n,i}) + \epsilon\mathcal{D}(u^{n,i})$, $i = 1, 2$ and $a_{11} = a_{22} = 1/4$, $a_{12} = 1/4 - \sqrt{3}/6$, $a_{21} = 1/4 + \sqrt{3}/6$.

Existence of $\{u^n\}_{n=0}^N$ can be established by using a variant of the Brouwer Fixed-Point Theorem (cf. [10]). The $L^2$-conservation property $\|u^n\| = \|u^0\|$ is equally straightforward. Uniqueness and convergence can be established under the CFL-type condition $\kappa_n h^{-1} \le c$ sufficiently small. In particular, assuming this CFL-condition to be valid, the convergence rates

$$\|u(\cdot, t_n) - u^n\| = O(h_{max}^q + \kappa^{2s}), \quad \kappa = \max_{0\le n\le N} \kappa_n,$$

can be rigorously proven for the fully discrete approximation. Here $s = 1$ for the Midpoint rule and $s = 2$ for the two-stage, fourth-order method. The arguments in favor of these assertions are very similar to those appearing already in [10, 32], and so we pass over the details.

## 5. Numerical experiments

Numerical experiments designed to gauge the performance of our conservative schemes are reported in this section. Interest is given particularly to two issues:

(1) Validation of the theoretical results, including a study of the convergence rates and in particular the dependence of these and other aspects of the approximations on the conditions specified in Proposition 3.1.

(2) Comparing the performance of the conservative methods to the dissipative methods of [27]. This includes not only a comparison of the convergence rates, but also a comparison of the errors as a function of time.

We have implemented and tested four classes of spatial approximation schemes corresponding to one of two choices for each of the operators $\mathcal{N}$ and $\mathcal{D}$. For each of $\mathcal{N}$ and $\mathcal{D}$, the spatial approximation was taken to be either the conservative discretization defined in Section 3 (indicated briefly by C) or the dissipative approximation as sketched in Section 4 (denoted by NC). The NC-NC method corresponds to the scheme considered in [27] with the difference that we have implemented it together with the conservative time-stepping method (4.4)-(4.5) so that comparisons between the various schemes, C-C, C-NC, NC-C and NC-NC are fair. The notation C-NC, say, connotes that the spatial approximation uses the conservative version for the nonlinear operator and the non-conservative approximation of the dispersive operator, and similarly for the other three integration methods (see Table 5.1). In fact, the outcome of the experiments using C-NC are not reported here since it turned out the approximations generated thereby were almost identical to those of the NC-NC scheme. This latter fact is part of the evidence supporting our view that conservative treatment of the dispersive term has a much larger effect on the resulting approximation than does using a conservative scheme for the nonlinearity. As mentioned already, the various spatial approximations were all implemented in a fully discrete version with the conservative, second order time stepping method (4.4)–(4.5). The nonlinear algebraic equations that arise in the simulation were solved using two different methods, *viz.* Newton's method and an explicit-implicit scheme where the nonlinear term was made explicit and the dispersive term implicit. There was little difference in accuracy or performance between the two schemes and consequently, we do not dwell further on this issue.

The numerical experiments reported here are only for the KdV-equation

$$u_t + uu_x + \epsilon u_{xxx} = 0 \tag{5.1}$$

| $\mathcal{N}$ | $\mathcal{D}$ | Designation |
|---|---|---|
| (2.8), (2.7) | (2.13) | C-C |
| (2.8), (2.7) | (4.1) | C-NC |
| (2.8), (4.3) | (2.13) | NC-C |
| (2.8), (4.3) | (4.1) | NC-NC |

TABLE 1. Definition of conservative, nonconservative and semi-conservative schemes.

itself, with $\epsilon = 1/24^2$. The computational domain was set to $[0,1]$ throughout and the domain was divided into $N$ cells. To check accuracy and convergence rates, two well-known solutions of (5.1) were used. The first is a so-called *cnoidal-wave* solution,

$$u(x,t) = a\,\mathrm{cn}^2\left(4K(x - vt - x_0)\right) \tag{5.2}$$

where $cn(z) = cn(z:m)$ is the Jacobi elliptic function with modulus $m \in (0,1)$ (see [3]) and the parameters have the values $a = 192m\epsilon K(m)^2$ and $v = 64\epsilon(2m-1)K(m)^2$ whilst $x_0$ is an arbitrary, real translation. Here, the function $K = K(m)$ is the complete elliptic integral of the first kind and the parameters are so organized that the solution $u$ has spatial period 1. The choice of parameters is a specialization of the general, cnoidal-wave solution which has three free parameters, though their range is restricted (see, e.g. [35], Section 3) It deserves note that the cnoidal waves comprise stable solutions of the time dependent problem [9], so numerical error will not set off instabilities of the continuous problem. Thus, any instability that manifests itself would be due solely to the numerical scheme. We used the value $m = 0.9$ in all the numerical experiments involving the cnoidal–wave solutions.

The classical *solitary-wave* solution

$$u(x,t) = A\,\mathrm{sech}^2\left(K(x - vt - x_0)\right) \tag{5.3}$$

was also used, with $A = 1$, $v = A/3$, $K = \frac{1}{2}\sqrt{\frac{A}{3\epsilon}}$ and $x_0 = 1/2$ so that the wave commences its evolution centered in the period domain. This traveling wave, too, is a stable solution of the KdV-equation (see [12] and [14] for the original proof of this fact). Of course, the latter is not periodic in space, but owing to its exponential decay, it can be treated as periodic by simply restricting it to the computational domain $[0,1]$ and imposing periodic boundary conditions across $x = 0$ and $x = 1$. This truncation and the resulting evolution that occurs when solving the periodic initial-value problem results in a solution of the KdV-equation (5.1) which is a high accuracy approximation of the solitary-wave over long time scales. Much of the numerical work on the KdV-equation has made use of this small trick to check for accuracy and convergence. Theory and sharp error estimates of the time scale over which such periodic approximations remain valid may be found in [16] and [26]. Another popular method of approximating solutions on the line or the half-line that decay rapidly to zero at infinity is to truncate it on a sufficiently long spatial domain as above and then use two-point boundary-value problems with homogeneous, Dirichlet conditions at the end-points (see e.g. [18]). A numerical scheme developed to simulate directly solutions on unbounded domains was put forward by Guo and Shen (see [34] and several, subsequent papers expanding on their original work).

5.1. **Convergence rates.** The results reported here begin with the case of a uniform mesh. Since the second order Crank-Nicholson time discretization is employed and our interest is in the effect of the various spatial discretizations, we determined the time-step by the relation $\kappa = Ch^2$. This relationship guarantees that the error will be dominated by the spatial discretization when $q = 2, 3$. For $q = 4$ and in view of the very small spatial errors, we used a two-stage implicit Runge-Kutta method of Gauss-Legendre type. This method is fourth-order accurate in time and can also be shown to have the conservative property $\|u^n\| = \|u^0\|, n = 1, 2, \cdots$.

Tables 2, 3 and 4 contain the numerical errors and the calculated rates of convergence for $q = 2, 3, 4$. The simulations of solutions of (5.1) that underlie the information in these tables were all made with the cnoidal-wave initial data with the value of the elliptic modulus $m = 0.9$ and the other parameters as specified below (5.2). It is worth pointing out that for the cnoidal-wave solutions, the value of $m$ carries with it the

TABLE 2. Cnoidal-wave problem, $q = 2$, uniform mesh.

| | $N$ | $\kappa$ | $L^2$ error | order | $L^\infty$ error | order |
|---|---|---|---|---|---|---|
| C-C method | 10 | 4.0E-02 | 1.3169E-00 | | 1.9388E-00 | |
| | 20 | 1.0E-02 | 1.2735E-00 | 0.0483 | 2.1475E-00 | -0.1475 |
| | 40 | 2.5E-03 | 1.7869E-01 | 2.8333 | 3.0294E-01 | 2.8256 |
| | 80 | 6.25E-04 | 1.2017E-02 | 3.8943 | 2.0728E-02 | 3.8694 |
| | 160 | 1.5625E-04 | 7.6271E-04 | 3.9778 | 1.3499E-03 | 3.9407 |
| | 320 | 3.90625E-05 | 4.8290E-05 | 3.9813 | 9.2342E-05 | 3.8697 |
| NC-C method | 10 | 4.0E-02 | 9.0693E-01 | | 1.5463E-00 | |
| | 20 | 1.0E-02 | 3.1383E-01 | 1.5310 | 6.0458E-01 | 1.3548 |
| | 40 | 2.5E-03 | 1.9160E-01 | 0.7119 | 3.4252E-01 | 0.8197 |
| | 80 | 6.25E-04 | 3.9244E-03 | 5.6095 | 7.6569E-03 | 5.4833 |
| | 160 | 1.5625E-04 | 5.4422E-04 | 2.8502 | 9.8365E-04 | 2.9605 |
| | 320 | 3.90625E-05 | 4.1574E-05 | 3.7104 | 7.8722E-05 | 3.6433 |
| NC-NC method | 10 | 4.0E-02 | 7.1270E-01 | | 1.2985E-00 | |
| | 20 | 1.0E-02 | 5.9638E-01 | 0.2571 | 1.1130E-00 | 0.2224 |
| | 40 | 2.5E-03 | 5.7218E-01 | 0.0598 | 1.0403E-00 | 0.0975 |
| | 80 | 6.25E-04 | 1.0466E-00 | -0.8712 | 1.6738E-00 | -0.6861 |
| | 160 | 1.5625E-04 | 2.0404E-01 | 2.3588 | 3.4832E-01 | 2.2646 |
| | 320 | 3.90625E-05 | 2.6643E-02 | 2.9370 | 4.5632E-02 | 2.9323 |

balance being struck between nonlinearity and dispersion. Values of $m > 0$ near to zero correspond to nearly linear behavior, (the Jacobi elliptic function cn is nearly a cosine) while values of $m < 1$ near to one are where nonlinear effects cannot be ignored (cn has a sharper crest and a wider trough). Starting with this initial data, the exact solution is as in (5.2) and it is compared directly to the output of the fully discrete schemes at the time $t = 10$ to determine the error. The $L^2$- and $L^\infty$-norms of this error are calculated numerically and reported in the tables. The computed convergence rates $r$ are simply

$$r = \frac{\log E(N) - \log E(2N)}{\log(2)}$$

where $E(M)$ is the $L^2$- or $L^\infty$-error made using $M$ cells in the spatial approximation.

For the C-C method, the convergence rate appears to be four for $q = 2$, three for $q = 3$ and five for $q = 4$. Note that the reported rate for $q = 4$ when $\kappa \sim h^2$ shows the second-order temporal convergence rate of the fully discrete scheme. As far as the assumptions in Propositions 3.1 are concerned, the parity $N$ of cells in $\mathcal{T}_h$ does not have any detectable effect on the accuracy achieved by the scheme. On the other hand, the parity of $q$ certainly does seem to matter. Indeed, it appears that the actual spatial convergence rate is $q+1$ when $q$ is even, but only $q$ when $q$ is odd. In the special case of $q = 2$ we observe fourth-order accuracy for the spatial error.

Next are reported simulations made when the mesh was far from uniform. Indeed, the mesh $\mathcal{T}_h$ was taken to be $2h, h, \cdots, 2h, h$. For such a mesh, the number of adjacent cells with differing lengths is maximal, which is to say, $\nu = M$. Again using data obtained from simulating the cnoidal-wave solution described above, the numerical error and orders of accuracy for $q = 2$ are determined and shown in Table 5.

The following points emerge from a study of the results reported in these tables.

(i) Care must be taken in comparing the results of tables 2 and 5 since, for the same number $N$ of cells, the maximum cell sizes are different by a factor of 4/3. Even taking this into account, there is a noticeable degradation in the errors and convergence rates for the C-C method whereas the NC-NC method seems to be immune to this effect. Thus, we tentatively conclude that accuracy and convergence rates both suffer in the presence of a non-uniform mesh.

TABLE 3. Cnoidal-wave problem, $q = 3$, uniform mesh

| | $N$ | $\kappa$ | $L^2$ error | order | $L^\infty$ error | order |
|---|---|---|---|---|---|---|
| C-C method | 10 | 4.0E-02 | 1.2083E-00 | | 2.1869E-00 | |
| | 20 | 1.0E-02 | 1.5809E-01 | 2.9342 | 3.5795E-01 | 2.6110 |
| | 40 | 2.5E-03 | 1.2153E-02 | 3.7014 | 3.3732E-02 | 3.4076 |
| | 80 | 6.25E-04 | 1.2048E-03 | 3.3344 | 3.3640E-03 | 3.3259 |
| | 160 | 1.5625E-04 | 1.3999E-04 | 3.1054 | 3.6877E-04 | 3.1894 |
| NC-C method | 10 | 4.0E-02 | 1.2644E-00 | | 1.9401E-00 | |
| | 20 | 1.0E-02 | 1.9830E-01 | 2.6727 | 3.3587E-01 | 2.5302 |
| | 40 | 2.5E-03 | 1.1657E-02 | 4.0884 | 2.0516E-02 | 4.0331 |
| | 80 | 6.25E-04 | 6.4542E-04 | 4.1748 | 1.2238E-03 | 4.0673 |
| | 160 | 1.5625E-04 | 3.7251E-05 | 4.1149 | 8.0597E-05 | 3.9245 |
| NC-NC method | 10 | 4.0E-02 | 9.7806E-01 | | 1.6220E-00 | |
| | 20 | 1.0E-02 | 7.4734E-01 | 0.3882 | 1.2326E-00 | 0.3961 |
| | 40 | 2.5E-03 | 3.6619E-02 | 4.3511 | 6.2686E-02 | 4.2974 |
| | 80 | 6.25E-04 | 1.3171E-03 | 4.7972 | 2.2584E-03 | 4.7948 |
| | 160 | 1.5625E-04 | 4.8798E-05 | 4.7544 | 8.3729E-05 | 4.7534 |

TABLE 4. Cnoidal-wave problem, $q = 4$, uniform mesh

| | $N$ | $\kappa$ | $L^2$ error | order | $L^\infty$ error | order |
|---|---|---|---|---|---|---|
| C-C method | 10 | 4.0E-02 | 7.6947E-02 | | 1.3825E-01 | |
| | 20 | 1.0E-02 | 8.2647E-03 | 3.2188 | 1.8905E-02 | 2.8704 |
| | 40 | 2.5E-03 | 3.8736E-06 | 11.0591 | 1.4713E-05 | 10.3274 |
| | 80 | 6.25E-04 | 5.3864E-08 | 6.1682 | 2.6274E-07 | 5.8073 |
| | 160 | 1.5625E-04 | 1.5628E-09 | 5.1071 | 7.3846E-09 | 5.1529 |
| NC-C method | 10 | 4.0E-02 | 2.2960E-01 | | 3.8102E-01 | |
| | 20 | 1.0E-02 | 6.3339E-02 | 1.8579 | 1.1280E-01 | 1.7561 |
| | 40 | 2.5E-03 | 3.3763E-06 | 14.1954 | 1.2967E-05 | 13.0866 |
| | 80 | 6.25E-04 | 5.3809E-08 | 5.9714 | 2.6235E-07 | 5.6272 |
| | 160 | 1.5625E-04 | 1.5628E-09 | 5.1056 | 7.3875E-09 | 5.1503 |
| NC-NC method | 10 | 4.0E-02 | 7.1739E-01 | | 1.1916E-00 | |
| | 20 | 1.0E-02 | 1.1308E-02 | 5.9873 | 1.9327E-02 | 5.9461 |
| | 40 | 2.5E-03 | 1.0106E-04 | 6.8059 | 1.7756E-04 | 6.7661 |
| | 80 | 6.25E-04 | 8.1893E-07 | 6.9472 | 1.5333E-06 | 6.8535 |
| | 160 | 1.5625E-04 | 6.8941E-09 | 6.8922 | 1.6311E-08 | 6.5546 |

(ii) Despite the observed reduction in the order of the C-C method, the errors are smaller than those of the NC-NC method in the range of meshes employed (in this respect, see also Figure 1) although this is expected to be reversed for larger values of $N$ on account of the apparent higher order of the NC-NC method.

5.2. **Comparison of the conservative and non-conservative methods.** In this subsection, further numerical results are presented with the aim of acquiring a deeper understanding of the performance of the conservative and non-conservative methods. A graphical approach is adopted to capture behavior that may not be revealed by simple tabulation of errors and convergence rates.

We start with the cnoidal-wave test problem with $q = 2$ and $\kappa = 0.000625$. Figure 1 shows the plots of the numerical solutions of three proposed methods, C-C, NC-C and NC-NC respectively, at time $t = 10$. The exact solution is also provided as a reference in the plot. The NC-NC method has a large phase error,

TABLE 5. Cnoidal-wave problem, $q = 2$, non-uniform mesh of type $2h, h, \cdots, 2h, h$.

| | $N$ | $\kappa$ | $L^2$ error | order | $L^\infty$ error | order |
|---|---|---|---|---|---|---|
| C-C method | 10 | 4.0E-02 | 1.3340E-00 | | 5.8547E-00 | |
| | 20 | 1.0E-02 | 9.1940E-01 | 0.5370 | 1.6786E-00 | 1.8023 |
| | 40 | 2.5E-03 | 6.1914E-01 | 0.5704 | 1.0938E-00 | 0.6179 |
| | 80 | 6.25E-04 | 2.3766E-01 | 1.3814 | 3.9930E-01 | 1.4538 |
| | 160 | 1.5625E-04 | 6.5006E-02 | 1.8703 | 1.1072E-01 | 1.8506 |
| | 320 | 3.90625E-05 | 1.6573E-02 | 1.9718 | 2.8665E-02 | 1.9496 |
| NC-C method | 10 | 4.0E-02 | 5.6937E-01 | | 1.0742E-00 | |
| | 20 | 1.0E-02 | 5.0389E-01 | 0.1763 | 9.5346E-01 | 0.1720 |
| | 40 | 2.5E-03 | 4.9450E-01 | 0.0271 | 8.8893E-01 | 0.1011 |
| | 80 | 6.25E-04 | 8.7717E-01 | -0.8269 | 1.4354E-00 | -0.6913 |
| | 160 | 1.5625E-04 | 1.8083E-01 | 2.2782 | 3.0860E-01 | 2.2176 |
| | 320 | 3.90625E-05 | 3.1798E-02 | 2.5076 | 5.4337E-02 | 2.5057 |
| NC-NC method | 10 | 4.0E-02 | 6.8821E-01 | | 1.2660E-00 | |
| | 20 | 1.0E-02 | 6.5336E-01 | 0.0750 | 1.2087E-00 | 0.0668 |
| | 40 | 2.5E-03 | 9.7878E-01 | -0.5831 | 1.6384E-00 | -0.4388 |
| | 80 | 6.25E-04 | 1.2109E-00 | -0.3070 | 1.8813E-00 | -0.1994 |
| | 160 | 1.5625E-04 | 3.2924E-01 | 1.8787 | 5.5988E-01 | 1.7485 |
| | 320 | 3.90625E-05 | 4.4494E-02 | 2.8875 | 7.6207E-02 | 2.8771 |

which makes the solution very inaccurate. On the other hand, both C-C and NC-C methods demonstrate quite a good approximation to the exact solution. We believe the large phase error of NC-NC method comes from its non-conservative aspect. Indeed, the change in the $L^2$-norm of the numerical solution from $t = 0$ to $t = 10$ was $3 \times 10^{-16}$ for the C-C method, $-3.06 \times 10^{-4}$ for the NC-C method and $-4.97 \times 10^{-2}$ for the NC-NC method. If the $L^2$-norm is not conserved, the magnitude of the wave decays as $t$ increases, thereby slowing its speed of propagation since larger amplitude waves travel faster in the KdV–context. This explains at least partially the large phase error of the NC-NC method. On the other hand, the better performance of NC-C method suggests that insisting upon a conservative version of the dispersive approximation plays a more important role in maintaining accuracy than does a conservative approach to approximating the nonlinear term.

We have also tried the nonuniform mesh of type $2h$, $h \cdots 2h$, $h$ with 80 cells. The comparison of the three numerical solutions are shown in Figure 1, bottom right. Again, better performance of the conservative method is observed.

Next, cubic polynomials were tested, i.e. the case $q = 3$. The same test as above was repeated with $N = 80$ and the same $\kappa$. The solutions at time $t = 10$ are shown in Figure 2, which indicates only small differences among these three methods. This fact can be observed from Table 3, where the $L^2$-errors are all of order $10^{-3}$. However, when we ran this test for much longer, out to $t = 200$, larger phase errors appeared again in the approximation made via the NC-NC method, as shown in the right graph of Figure 2.

The numerical tests conducted with the solitary-wave initial data were qualitatively entirely consistent with those conducted with the cnoidal-wave test problem. For that reason, we present only a small sampling of the solitary-wave tests. With $q = 2$ and $\kappa = 0.000625$, the numerical solutions of the three proposed methods at time $t = 25$, with the uniform meshes $N = 40$ and 80, are plotted in Figure 3. Again, one observes a large phase error in the NC-NC solution as well as a growing amplitude error.

5.3. **Time history of the $L^2$-error and the shape-error.** In this subsection, we investigate the longer time temporal evolution of the $L^2$-error of the three proposed methods. An interesting outcome of the longer-time experiments is that the $L^2$-error of the conservative method increases linearly with time, for $q = 2$ though we do not know how to prove such a result. The linear temporal growth of the error had been observed in an earlier work [22] where conservative, standard Galerkin methods using smooth splines were

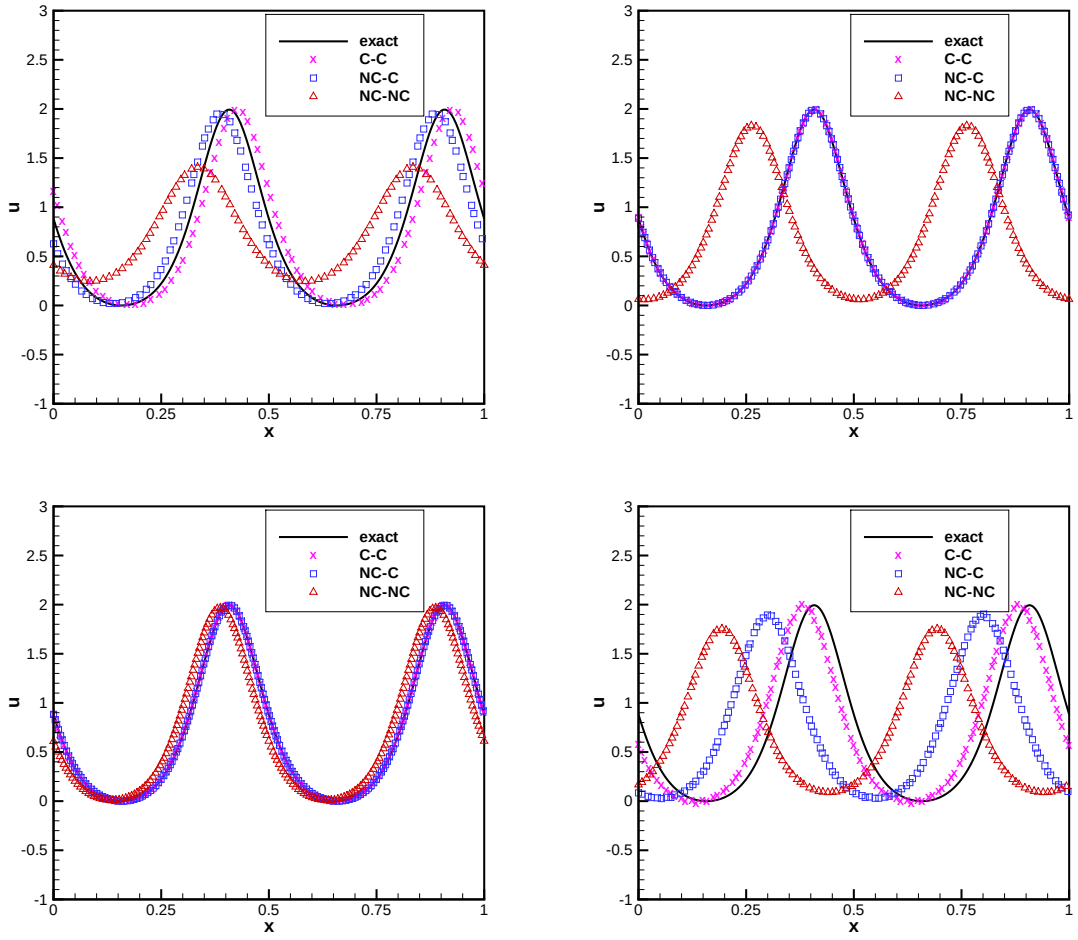


FIGURE 1. Numerical approximations of the cnoidal-wave problem using the C-C, NC-C and NC-NC methods; comparisons with the exact solution at time $t = 10$ with $q = 2$. Top left: 40 uniform cells; Top right: 80 uniform cells; Bottom left: 160 uniform cells; Bottom right: 80 nonuniform cells.

employed. Moreover, the shape error, defined below, is virtually constant in time for the fully conservative scheme.

Details of these simulations are now described. For the cnoidal-wave test problem with a uniform mesh, $N = 80$, $q = 2$ and $\kappa = 0.000625$, the time evolutions of the $L^2$-norm of the solution errors up to time $t = 10$ are shown in Figure 4, left. Observe that the C-C and NC-C methods have much smaller errors. Indeed, at time $t = 10$, the error of the NC-NC method is about 100 times larger than that of the C-C method. The C-C and NC-C methods show a linear and sublinear growth of the $L^2$-error, respectively. On the other hand, the left graph in Figure 4 shows that the same error for the NC-NC method is growing superlinearly. However, an examination of the same data in the logarithmic scale reveals that the error does not grow at an exponential rate.

The same simulation was made using cubic polynomials, the case where $q = 3$. The relevant time histories are plotted up to time $t = 200$ in Figure 5, left. As observed earlier, for small time, the differences between the errors for all three methods are small (as seen in Table 3). However, unlike the case $q = 2$ exhibited in

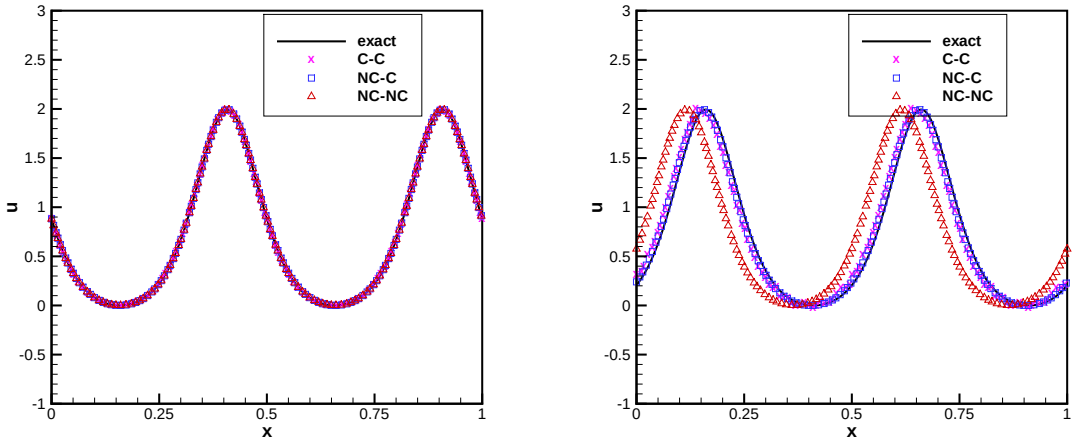


FIGURE 2. Numerical approximations of the cnoidal-wave problem using the C-C, NC-C and NC-NC methods; comparisons with $q = 3$ and 80 uniform spatial cells. Left: time $t = 10$; Right: $t = 200$.

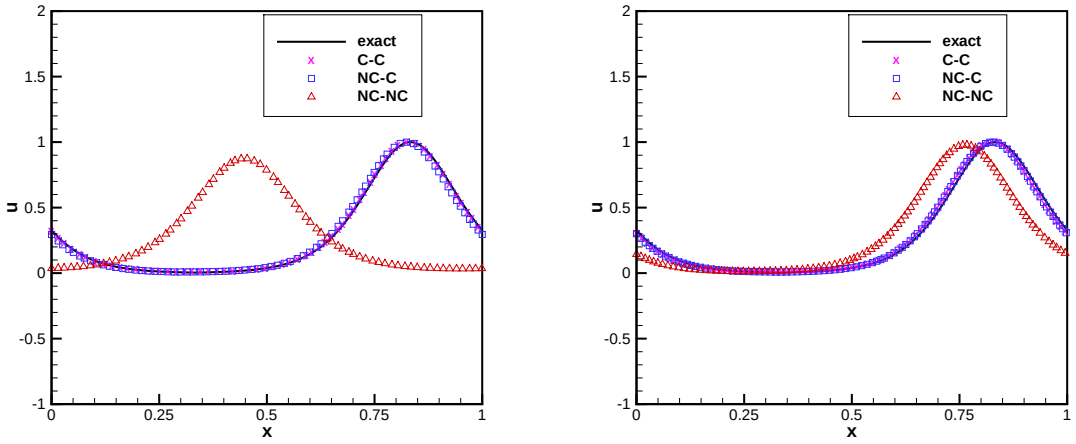


FIGURE 3. Numerical approximations of the solitary-wave problem using the C-C, NC-C and NC-NC methods; comparisons at time $t = 25$ with $q = 2$. Left: 40 uniform cells; Right: 80 uniform cells.

Figure 4 we see that now all three methods exhibit superlinear growth with the error of the NC-NC method growing at the fastest clip. Furthermore the difference between conservative and nonconservative methods becomes smaller as compared to the case $q = 2$. We feel that, as a general rule, these differences will become less pronounced as the degree of the polynomials increases and/or the mesh becomes nonuniform.

Attention is now turned to the solitary-wave test problem. Time histories of the $L^2$-errors up to time $t = 25$ for $q = 2$, are shown in Figure 5, right, and results similar to those outlined above are seen.

Finally, we consider the shape error $\hat{e} = \hat{e}(x, t)$ of an approximation $u_h$ of a true solution $u$ of (5.1). The shape error compares how good is the approximation, modulo the translation group on the period domain,

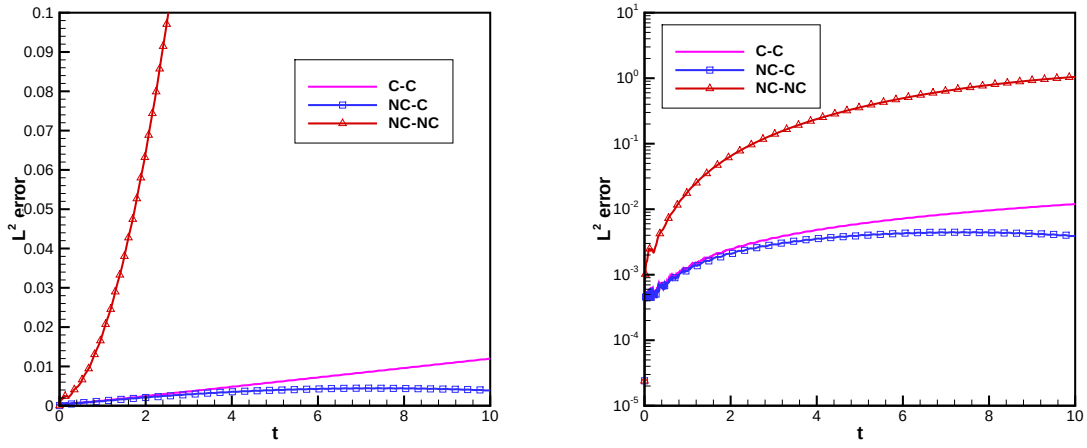


FIGURE 4. Time history of the $L^2$-error of the numerical approximations obtained from the C-C, NC-C and NC-NC methods for the cnoidal-wave problem with $q = 2$ and a uniform mesh with 80 cells. The right-hand graph shows the error in the logarithmic scale

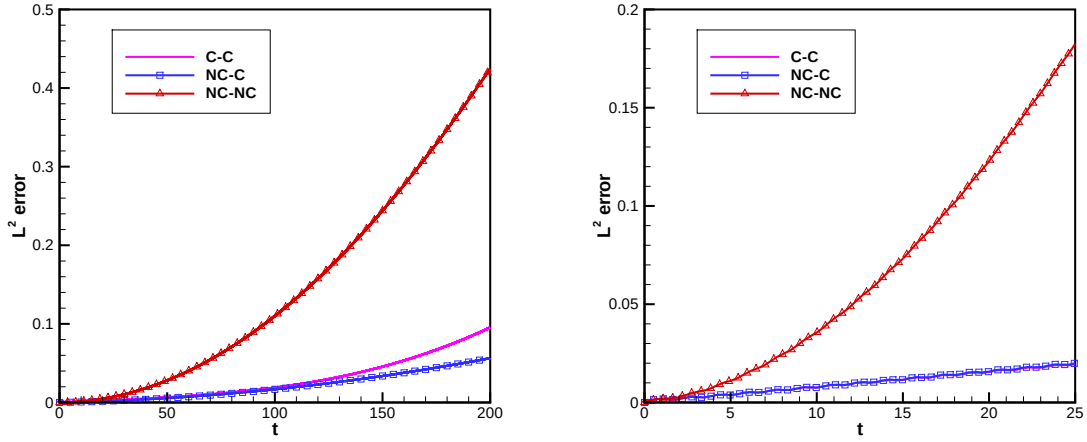


FIGURE 5. Time history of the $L^2$-error of the numerical approximations obtained using the C-C, NC-C and NC-NC methods. Left: the cnoidal-wave problem with $q = 3$ and a uniform spatial mesh with 80 cells; Right: the solitary-wave problem with $q = 2$ and a uniform mesh with 80 cells.

and is defined precisely to be

$$\hat{e}(x,t) = \min_{\xi\in[-0.5,0.5]} ||u_h(x,t) - u(x+\xi,t)||$$

where $u(x,t)$ is the exact solution and $u_h(x,t)$ is the numerically obtained approximation. Thus, the shape error is the minimization of the difference between the numerical approximation and the spatially shifted exact solution. Of course, the spatial $L^2$-norms appearing here are in fact a high-order Gauss-Legendre integration of the square of the discrete solution minus the exact solution evaluated at the nodes of the mesh. The concept of shape error was introduced in [22] with the purpose of providing a detailed analysis

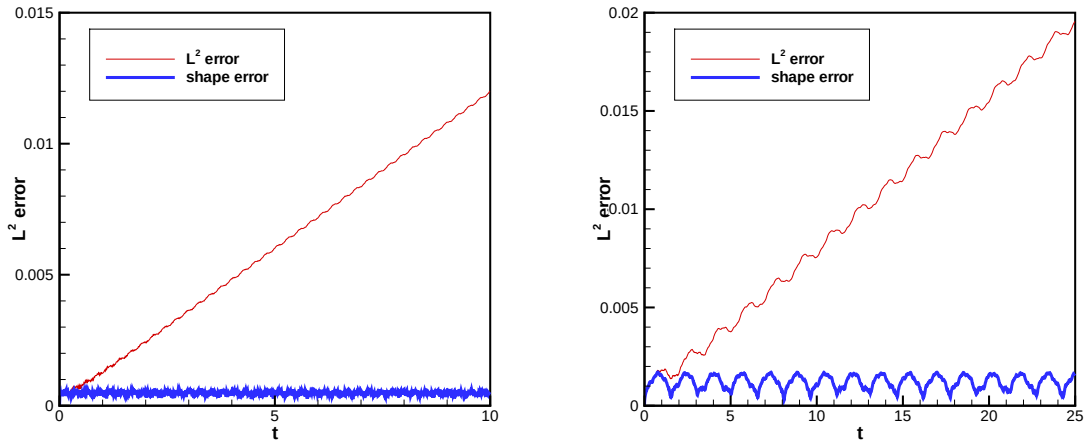


FIGURE 6. The time history of the $L^2$-error and shape error of the numerical approximations obtained using the C-C method with $q = 2$ and 80 uniform spatial cells. Left: the cnoidal-wave problem (5.1) and (5.2); Right: the solitary-wave problem (5.1) and (5.3).

of the error in terms of a shape and a phase component. The shape error and the absolute error, both in $L^2$-norm, of the C-C method are both given in Figure 6. Observe that the shape error is almost constant in time after an initial "settling down" period and shows a clearly visible periodic behavior. We believe this behavior of the shape error to be indicative of some very interesting phenomena, such as the existence of exact, discrete, traveling-wave solutions to the conservative numerical scheme when the space- and time-discretization lengths are constant and bear an appropriate relation to each other.

## 6. Summary

Constructed, analyzed and tested are conservative numerical schemes for the GKdV-equation. It is found that such schemes, in addition to possessing high accuracy, mimic very well the properties of the traveling-wave solutions considered here. As it is known that general initial data for the KdV-equation breaks up into solitary waves and a dispersive tail, the results displayed here indicate that the conservative scheme is likely to produce better approximations of general solutions than do the non-conservative ones.

Work in progress is aimed at broadening the range of initial data that are investigated numerically as well as considering higher powers of the nonlinearity. Similar theory and simulations are also being carried out for coupled systems of nonlinear, dispersive wave equations of the form investigated recently in [19] (see also [39]).

## References

[1] L. Abdelouhab, J.L. Bona, M. Felland, and J.-C. Saut. Non-local models for nonlinear, dispersive waves. *Physica D*, 40:360–392, 1989.
[2] K. Abe and O. Inoue. Fourier expansion solution of the Korteweg-de Vries equation. *J. Comp. Phys.*, 34:202–210, 1980.
[3] M. Abramowitz and I. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 55 of *Applied Mathematics Series*. National Bureau of Standards, 1965.
[4] R. Adams. *Sobolev Spaces*. Academic Press, New York, 1975.
[5] A.A. Alazman, J.P. Albert, J.L. Bona, M. Chen, and J. Wu. Comparisons between the BBM-equation and a Boussinesq system. *Advances Differential Eq.*, 11:121–166, 2006.
[6] J.P. Albert, J.L. Bona, and M. Felland. A criterion for the formation of singularities for the generalized Korteweg-de Vries equation. *Matemática Aplicada e Computacional*, 7:3–11, 1988.
[7] M.E. Alexander and J. Ll. Morris. Galerkin methods applied to some model equations for nonlinear dispersive waves. *J. Comp. Phys.*, 30:428–451, 1979.

[8] J. Angulo, J.L. Bona, F. Linares, and M. Scialom. Scaling, stability and singularities for nonlinear dispersive wave equations: The critical case. *Nonlinearity*, 15:759–786, 2002.
[9] J. Angulo, J.L. Bona, and M. Scialom. Stability of cnoidal waves. *Advances Differential Eq.*, 11:1321–1374, 2006.
[10] G. Baker, V.A. Dougalis, and O.A. Karakashian. Convergence of Galerkin approximations for the Korteweg-de Vries equation. *Math. Comp.*, 40:419–433, 1983.
[11] G. Baker, W. Jureidini, and O.A. Karakashian. Piecewise solenoidal vector fields and the Stokes problem. *SIAM J. Num. Anal.*, 27:1466–1485, 1990.
[12] T.B. Benjamin. The stability of solitary waves. *Proc. Royal Soc. London, Ser. A*, 328:153–183, 1972.
[13] T.B. Benjamin, J.L. Bona, and J.J. Mahony. Model equations for long waves in nonlinear dispersive systems. *Philos. Trans. Royal Soc. London, Ser. A*, 272:47–78, 1972.
[14] J.L. Bona. On the stabilty theory of solitary waves. *Proc. Royal Soc. London, Ser. A*, 349:363–374, 1975.
[15] J.L. Bona. Model equations for waves in nonlinear, dispersive systems. In *Proc. Int. Congress of Mathematicians, Helsinki, 1978*, volume 2, pages 887–894. Academia Scientiarum Fennica: Hungary, 1980.
[16] J.L. Bona. Convergence of periodic wave trains in the limit of large wavelength. *Appl. Sci. Research*, 37:21–30, 1981.
[17] J.L. Bona. On solitary waves and their role in the evolution of long waves. In H. Amann, N. Bazley, and K. Kirchgässner, editors, *Applications of Nonlinear Analysis in the Physical Sciences*, pages 183–205. Pitman: London, 1981.
[18] J.L. Bona, H. Chen, S.-M. Sun, and B.-Y. Zhang. Approximating initial-value problems by two-point boundary-value problems: The BBM-equation. *Bull. Iranian Math. Soc.*, 36:1–25, 2010.
[19] J.L. Bona, J. Cohen, and G. Wang. Glorbal well-posedness for a system of KdV-type with coupled quadratic nonlinearities. *Submitted*.
[20] J.L. Bona, T. Colin, and D. Lannes. Long wave approximations for water waves. *Arch. Rational Mech. Anal.*, 178:373–410, 2003.
[21] J.L. Bona, V.A. Dougalis, O.A. Karakashian, and W.R. McKinney. Fully discrete methods with grid refinement for the generalized Korteweg-de Vries equation. In M. Shearer, editor, *Proceedings of the workshop on viscous and numerical approximations of shock waves, N.C. State University*, pages 117–124. SIAM, Philadelphia, 1990.
[22] J.L. Bona, V.A. Dougalis, O.A. Karakashian, and W.R. McKinney. Conservative high order schemes for the Generalized Korteweg-de Vries equation. *Philos. Trans. Royal Soc. London, Ser. A*, 351:107–164, 1995.
[23] J.L. Bona, W.G. Pritchard, and L.R. Scott. An evaluation of a model equation for water waves. *Philos. Trans. Royal Soc. London, Ser. A*, 302:457–510, 1981.
[24] J.L. Bona, W.G. Pritchard, and L.R. Scott. A comparison of solutions of two models for long waves. In N. Lebovitz, editor, *Lectures in Applied Mathematics*, volume 20, pages 235–267. American Mathematical Society, Providence, 1983.
[25] S. Brenner and L.R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer Verlag, New York, third edition, 2002.
[26] H. Chen. Long-period limit of nonlinear dispersive waves: the BBM-equation. *Differential and Integral Eq.*, 19:463–480, 2006.
[27] Y. Cheng and C.-W. Shu. A discontinuous finite element method for time dependent partial differential equations with higher order derivatives. *Math. Comp.*, 77:699–730, 2008.
[28] P. Ciarlet. *The Finite Element Method for Elliptic Problems*. North Holland, Amsterdam, New York, Oxford, 1980.
[29] B. Cockburn, G.E. Karniadakis, and C.-W. Shu. *Discontinuous Galerkin methods, Theory, Computation and Applications*, volume 11 of *Springer Lecture Notes in Computational Science and Engineering*. Springer Verlag, Heidelberg, New York, 2000.
[30] W.L. Craig. An existence theory for water waves and the Boussinesq and Korteweg-de Vries scaling limits. *Comm. Partial Differential Equations*, 10:787–1003, 1987.
[31] K. Dekker and J.G. Verwer. *Stability of Runge-Kutta Methods for stiff nonlinear differential equations*. North Holland, 1984.
[32] V.A. Dougalis and O.A. Karakashian. On some high order accurate fully discrete Galerkin methods for the Kortweg-de Vries equation. *Math. Comp.*, 45:329–345, 1985.
[33] B. Fornberg and G.B. Whitham. A numerical and theoretical study of certain nonlinear wave phenomena. *Philos. Trans. Royal Soc. London, Ser. A*, 289:373–404, 1978.
[34] B.-Y. Guo and J. Shen. Laguerre-Galerkin method for nonlinear partial differential equations on a semi-infinite interval. *Numer. Math.*, 86:635–654, 2000.
[35] A. Jeffrey and T. Kakutani. Weak nonlinear dispersive waves: A discussion centered around the Korteweg–de Vries equation. *SIAM Rev.*, 14:582–643, 1972.
[36] O.A. Karakashian and W. McKinney. On optimal high order in time approximations for the Korteweg-de Vries equation. *Math. Comp.*, 55:473–496, 1990.
[37] Y. Martel and F. Merle. Stability of blow-up profile and lower bounds on the blow up rate for the critical generalized KdV equation. *Annals Math.*, 155:235–280, 2002.
[38] F. Merle. Existence of blow-up solutions in the energy space for the critical generalized KdV equation. *J. American Math. Soc.*, 14:666–678, 2001.
[39] T. Oh. Diophantine conditions in global well-posedness for coupled Kdv-type systems. *Elec. J. Differential Eqns.*, 2009:1–48, 2009.

[40] H. Schamel and K. Elsässer. The application of the spectral method to nonlinear wave propagation. *J. Comp. Phys.*, 22:501–516, 1976.
[41] G. Schneider and C.E. Wayne. On the validity of 2d-surface water wave models. *GAMM Mitt. Ges. Angew. Math. Mech.*, 25:127–151, 2002.
[42] A.C. Scott, F.Y.F. Chu, and D.W. McLaughlin. The soliton: A new concept in applied science. *Proc. IEEE*, 61:1443–1483, 1973.
[43] T. Taha and M. Ablowitz. Analytical and numerical aspects of certain nonlinear evolution equations. III. Numerical, Korteweg-de Vries equation. *J. Comp. Phys.*, 55:231–253, 1984.
[44] T. Taha and M. Ablowitz. Analytical and numerical aspects of certain nonlinear evolution equations. IV. Numerical, modified Korteweg-de Vries equation. *J. Comp. Phys.*, 77:540–548, 1988.
[45] F. Tappert. Numerical solution of the Korteweg-de Vries equation and its generalizations by the split-step Fourier method. In A. C. Newell, editor, *Nonlinear Wave Motion*, Lectures in Applied Mathematics, pages 215–216. Amer. Math. Soc., Providence, R.I., 1974.
[46] A.C. Vliegenthart. On finite-difference methods for the Korteweg-de Vries equation. *J. of Engrg. Math.*, 5:137–155, 1971.
[47] L.B. Wahlbin. A dissipative Galerkin method for the numerical solution of first order hyperbolic equations. In C. de Boor, editor, *Mathematical Aspects of Finite Elements in Partial Differential Equations*, pages 147–169. Academic Press, New York, 1974.
[48] R. Winther. A conservative finite element method for the Korteweg-de Vries equation. *Math. Comp.*, 34:23–43, 1980.
[49] Y. Xu and C.-W. Shu. Error estimates of the semi-discrete local discontinuous Galerkin method for nonlinear convection-diffusion and KdV equations. *Computer Methods in Appl. Mech. and Eng.*, 196:3805–3822, 2007.
[50] J. Yan and C.-W. Shu. A local discontinuous Galerkin method for KdV type equations. *SIAM J. Numer. Anal.*, 40:769–791, 2002.

Department of Mathematics, Statistics and Computer Science, The University of Illinois at Chicago, Chicago, IL, 60607, U.S.A.
*E-mail address*: bona@math.uic.edu

Department of Mathematical Sciences, The University of Memphis, Memphis, TN 38152, U.S.A.
*E-mail address*: hchen1@memphis.edu

Department of Mathematics, The University of Tennessee, Knoxville, TN 37996, U.S.A.
*E-mail address*: ohannes@math.utk.edu

Department of Mathematics, The University of Tennessee, Knoxville, TN 37996 and the Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, U.S.A.
*E-mail address*: xingy@math.utk.edu