Journal of Computational Physics 418 (2020) 109608

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

An energy-based discontinuous Galerkin method for semilinear wave equations

Daniel Appelö [a,1], Thomas Hagstrom [b,2], Qi Wang [c,3], Lu Zhang [b,*,2]

[a] Dept. of Applied Mathematics, University of Colorado, Boulder, Boulder, CO 80309, United States
[b] Dept. of Mathematics, Southern Methodist University, Dallas, TX 75275, United States
[c] Dept. of Mathematics, Southwestern University of Finance and Economics, 555 Liutai Ave, Wenjiang, Chengdu, Sichuan 611130, China

ARTICLE INFO

Article history:
Received 4 January 2020
Received in revised form 25 May 2020
Accepted 26 May 2020
Available online 30 May 2020

Keywords:
Discontinuous Galerkin
Semilinear wave equation
Upwind flux

ABSTRACT

We generalize the energy-based discontinuous Galerkin method proposed in [1] to second-order semilinear wave equations. A stability and convergence analysis is presented along with numerical experiments demonstrating optimal convergence for certain choices of the interelement fluxes. Applications to the sine-Gordon equation include simulations of breathers, kink, and anti-kink solitons.

1. Introduction

Discontinuous Galerkin methods have emerged as a method-of-choice for solving hyperbolic initial-boundary value problems in first-order Friedrichs form. Advantages include guaranteed stability on unstructured grids, local time evolution, and arbitrary order [2]. However, typical formulations of wave equations in physics arise as second-order systems. We believe it is advantageous to directly treat such second-order systems. First, first-order reformulations require more variables and thus may be less memory efficient while requiring additional initial and boundary conditions; the latter must be compatible with the original equations for the first-order reformulation to be equivalent.

The energy-based DG method was introduced in [1], where a formulation for linear problems whose energy takes the form of a simple sum of kinetic and potential energy is derived. Error estimates and experiments with the scalar wave equation were also shown. The extension of the method to the elastic wave equation is presented in [3]. There one has to account for additional symmetries of the potential energy in the analogue of equation (7) below, which directly lead to a multidimensional null space. An example with a more general energy form, namely the advective wave equation, is

[*] Corresponding author.
E-mail addresses: Daniel.Appelo@colorado.edu (D. Appelö), thagstrom@smu.edu (T. Hagstrom), qwang@swufe.edu.cn (Q. Wang), luzhang@smu.edu (L. Zhang).

[1] Supported in part by NSF Grant DMS-1913076. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[2] Supported in part by NSF Grant DMS-1418871. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the National Science Foundation.

[3] Supported in part by Fundamental Research Funds for the Central Universities Grant JBK2002002 and JBK1805001; in part by NSF-China Grant 11501460. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the Fundamental Research Funds for the Central Universities and National Science Foundation of China.

https://doi.org/10.1016/j.jcp.2020.109608
0021-9991/© 2020 Elsevier Inc. All rights reserved.

considered in [4]. Lastly, superconvergence results and improved error estimates for the method applied to the scalar wave equation are derived in [5,6].

The central aim of our development of the energy-based schemes is to provide a DG method for second-order wave equations which is as simple, reliable, flexible, and general as DG methods applied to Friedrichs systems. The essential ideas underpinning the formulation are:

i. Introduction of a variable which is **weakly** equal to the time derivative of the solution (see (7) below),
ii. Construction of numerical fluxes based on the energy flux at element boundaries.

Advantages of the proposed method are that we use the minimal number of variables required (compare with HDG [7] methods) and simple conservative or upwind fluxes can be chosen to be independent of the mesh (compare with IPDG [8,9]). The energy-DG method has similarities with LDG [10–12], which introduces weak gradients of the variables in each element. Although it seems clear that the method can be adapted to any second-order linear hyperbolic system, the formulation for nonlinear problems presented in [1] is both incomplete and inconvenient. In particular, the analogue of (7) proposed in [1] involves a nonlinear function of ϕ_u. Thus the equation would typically be overdetermined. Moreover, to guarantee the energy estimate the system must be satisfied for $\phi_u = u$, which directly leads to a nonlinear problem to calculate $\frac{\partial u}{\partial t}$. Our main result in this work is to show how all these potential issues can be avoided for semilinear problems.

The remainder of the paper is organized as follows. In Section 2 we introduce the semidiscretization, proposing a number of interelement fluxes and proving the basic energy estimate. In Section 3 we prove a suboptimal error estimate and present several numerical experiments in Section 4. The latter demonstrate optimal convergence for certain choices of flux: specifically an energy-conserving alternating flux as well as two energy-dissipating fluxes. We also present simulations of soliton solutions of the sine-Gordon equation. We summarize our results in Section 5 and point out areas for future research.

2. Problem formulation

We consider semilinear wave equations of the form

$$\frac{\partial^2 u}{\partial t^2} + \theta \frac{\partial u}{\partial t} = c^2 \Delta u + f(u), \quad \mathbf{x} \in \Omega \subset \mathbb{R}^d, \quad t \geq 0, \tag{1}$$

where $c > 0$ is the sound wave speed, which we take to be constant, and $\theta \geq 0$ is the dissipation coefficient, and $f(u)$ is a smooth function with $\lim_{u \to 0} \frac{f(u)}{u}$ bounded. The initial conditions are given by

$$u(\mathbf{x}, 0) = g_1(\mathbf{x}), \quad \frac{\partial u(\mathbf{x}, 0)}{\partial t} = g_2(\mathbf{x}), \quad \mathbf{x} \in \Omega \subset \mathbb{R}^d.$$

Note that when $\theta = 0$, (1) is the Euler-Lagrange equation derived from the Lagrangian

$$L = \frac{1}{2}\left(\frac{\partial u}{\partial t}\right)^2 - \frac{c^2}{2}|\nabla u|^2 - F(u),$$

where $F'(u) = -f(u)$. To derive an energy-based DG formulation for problem (1), we introduce a second scalar variable to produce a system which is first order in time,

$$\begin{cases} \frac{\partial u}{\partial t} - v = 0, \\ \frac{\partial v}{\partial t} + \theta v - c^2 \Delta u - f(u) = 0. \end{cases} \tag{2}$$

The energy takes the form

$$E = \frac{1}{2} \int_{\Omega} \left(v^2 + c^2 |\nabla u|^2 + 2F(u)\right) d\mathbf{x}. \tag{3}$$

Note that $F(u) > 0$ corresponds to a defocusing equation and $F(u) < 0$ gives a focusing equation. In the rest of analysis in this paper, we investigate the defocusing equation with $F(u) > 0$, although the method formulation applies in either case. The change of the energy is given by boundary contributions and a volume integral related to the dissipation:

$$\frac{dE}{dt} = -\theta \int_{\Omega} \left(\frac{\partial u}{\partial t}\right)^2 d\mathbf{x} + \int_{\partial\Omega} c^2 v \nabla u \cdot \mathbf{n} \, dS, \tag{4}$$

where $\mathbf{n}$ is the outward-pointing unit normal.

We note that in our error analysis we will make the stronger assumption $\frac{f(u)}{u} < 0$, which can be enforced after a transformation of variables if we only assume the ratio is bounded above. Then the defocusing assumption holds since

$$F(u) = -\int_0^u f(z)dz = \int_0^u \left(-\frac{f(z)}{z}\right) zdz > 0.$$

2.1. Semi-discrete DG formulation

We develop an energy-based DG scheme for problem (1) through the reformulation (2). Let the domain Ω be discretized by non-overlapping elements Ω_j; $\Omega = \cup_j \Omega_j$. Choose the components of the approximations, (u^h, v^h) to (u, v), restricted to Ω_j, to be polynomials or tensor-product polynomials of degree q and s respectively,[4]

$$U_h^q = \left\{u^h(\mathbf{x},t) : u^h(\mathbf{x},t) \in \Pi^q(\Omega_j),\ \ \mathbf{x}\in\Omega_j,\ \ t\ge 0\right\},\quad V_h^s = \left\{v^h(\mathbf{x},t) : v^h(\mathbf{x},t) \in \Pi^s(\Omega_j),\ \ \mathbf{x}\in\Omega_j,\ \ t\ge 0\right\}.$$

We seek an approximation to the system (2) which satisfies a discrete energy estimate analogous to (3). Consider a discrete energy in Ω_j,

$$E_j^h(t) = \frac{1}{2}\int_{\Omega_j} \left(v^h\right)^2 + c^2\left|\nabla u^h\right|^2 \, d\mathbf{x} + \sum_{\mathbf{k}} \omega_{\mathbf{k},j} F(u^h(\mathbf{x}_{\mathbf{k},j})), \tag{5}$$

and its time derivative

$$\frac{dE_j^h(t)}{dt} = \int_{\Omega_j} v^h \frac{\partial v^h}{\partial t} + c^2 \nabla u^h\cdot\nabla\frac{\partial u^h}{\partial t}\, d\mathbf{x} - \sum_{\mathbf{k}} \omega_{\mathbf{k},j} f(u^h(\mathbf{x}_{\mathbf{k},j}))\frac{\partial u^h}{\partial t}(\mathbf{x}_{\mathbf{k},j}), \tag{6}$$

where we have omitted t in $u^h(\mathbf{x}_{\mathbf{k},j})$ for simplicity. Note here we use a quadrature rule with nodes $\mathbf{x}_{\mathbf{k},j}$ in Ω_j and positive weights $\omega_{\mathbf{k},j}$ to approximate the integration of the nonlinear terms; in our experiments we use tensor-product Gauss rules with 16 nodes in each coordinate in a reference element. We did not try to determine the minimal number of nodes required to observe the convergence rates shown later. Minimal requirements for the error estimates are given in Assumption 1. To obtain a weak form which is compatible with the discrete energy (5) and (6), we choose $\phi_u \in U_h^q$, $\phi_v \in V_h^s$ and test the first equation of (2) with $-c^2\Delta\phi_u$, the second equation of (2), with ϕ_v and add terms which vanish for the continuous problem. This yields the following equations,

$$\begin{aligned}\int_{\Omega_j} -c^2\Delta\phi_u\left(\frac{\partial u^h}{\partial t} - v^h\right) d\mathbf{x} &= \int_{\partial\Omega_j} c^2\nabla\phi_u\cdot\mathbf{n}\left(v^* - \frac{\partial u^h}{\partial t}\right) dS \\ &\quad + \sum_{\mathbf{k}} \omega_{\mathbf{k},j}\phi_u(\mathbf{x}_{\mathbf{k},j})\frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})}\left(\frac{\partial u^h}{\partial t}(\mathbf{x}_{\mathbf{k},j}) - v^h(\mathbf{x}_{\mathbf{k},j})\right),\\ \int_{\Omega_j} \phi_v\frac{\partial v^h}{\partial t} - c^2\phi_v\Delta u^h + \theta\phi_v v^h\, d\mathbf{x} - \sum_{\mathbf{k}}\omega_{\mathbf{k},j}\phi_v(\mathbf{x}_{\mathbf{k},j}) f(u^h(\mathbf{x}_{\mathbf{k},j})) &= \int_{\partial\Omega_j} c^2\phi_v\left((\nabla u)^*\cdot\mathbf{n} - \nabla u^h\cdot\mathbf{n}\right) dS,\end{aligned}$$

where v^* and $(\nabla u)^*$ are numerical fluxes on both interelement and physical boundaries. In what follows, we apply integration by parts to obtain an alternative form,

$$\int_{\Omega_j} c^2\nabla\phi_u\cdot\nabla\left(\frac{\partial u^h}{\partial t} - v^h\right) d\mathbf{x} - \sum_{\mathbf{k}}\omega_{\mathbf{k},j}\phi_u(\mathbf{x}_{\mathbf{k},j})\frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})}\left(\frac{\partial u^h}{\partial t}(\mathbf{x}_{\mathbf{k},j}) - v^h(\mathbf{x}_{\mathbf{k},j})\right) = \int_{\partial\Omega_j} c^2\nabla\phi_u\cdot\mathbf{n}\left(v^* - v^h\right) dS, \tag{7}$$

and

$$\int_{\Omega_j} \phi_v\frac{\partial v^h}{\partial t} + c^2\nabla\phi_v\cdot\nabla u^h + \theta\phi_v v^h\, d\mathbf{x} - \sum_{\mathbf{k}}\omega_{\mathbf{k},j}\phi_v(\mathbf{x}_{\mathbf{k},j}) f(u^h(\mathbf{x}_{\mathbf{k},j})) = \int_{\partial\Omega_j} c^2\phi_v(\nabla u)^*\cdot\mathbf{n}\, dS. \tag{8}$$

Now by setting $\phi_u = u^h$ and $\phi_v = v^h$ and recalling (6) we arrive at

[4] For simplicity we abuse notation and let Π^r denote either the polynomials of degree r or the tensor-product polynomials of degree r in each coordinate on a reference element.

$$\frac{dE^h}{dt}=\sum_j \frac{dE_j^h}{dt}=-\sum_j \int_{\Omega_j} \theta\left(v^h\right)^2 \, d\mathbf{x}+\sum_j \int_{\partial\Omega_j} c^2\nabla u^h\cdot\mathbf{n}\left(v^*-v^h\right)+c^2 v^h(\nabla u)^*\cdot\mathbf{n}\, dS.$$

Note that if $\frac{f(u)}{u}=0$, then we need to impose the following additional equation to determine the mean value of $\frac{\partial u^h}{\partial t}$

$$\int_{\Omega_j} \tilde{\phi}_u\left(\frac{\partial u^h}{\partial t}-v^h\right) d\mathbf{x}=0.$$

Here, $\tilde{\phi}_u$ is an arbitrary constant function and this equation does not affect the energy.

The innovation here in comparison with the weak form proposed in [1] is the appearance of $\phi_u\frac{f(u^h)}{u^h}$ instead of $f(\phi_u)$ in (7). This exchange obviously yields an invertible linear system for computing $\frac{\partial u^h}{\partial t}$. The energy estimate still holds as the two terms are identical for the special choice $\phi_u=u^h$.

2.2. Fluxes

To complete the formulation of energy-based DG scheme proposed in Section 2.1, we must specify the numerical fluxes $v^*, (\nabla u)^*$ both at interelement and physical boundaries. Denote $'+'$ to be the trace of data from the outside of the element, $'-'$ to be the trace of data from the inside of the element. Here, we introduce the common notation for averages and jumps,

$$\{\{v^h\}\}\equiv\frac{1}{2}(v^h)^+ +\frac{1}{2}(v^h)^-,\quad [[v^h]]\equiv(v^h)^+\mathbf{n}^+ +(v^h)^-\mathbf{n}^-,$$

and

$$\{\{\nabla u^h\}\}\equiv\frac{1}{2}(\nabla u^h)^+ +\frac{1}{2}(\nabla u^h)^-,\quad [[\nabla u^h]]\equiv(\nabla u^h)^+\cdot\mathbf{n}^+ +(\nabla u^h)^-\cdot\mathbf{n}^-.$$

2.2.1. Interelement boundaries

To analyze the problem, we label two elements sharing one interelement boundary face, F_j, by 1 and 2. Then, besides the volume dissipation, if any, their net contribution to the discrete energy $E^h(t)$ is

$$\int_{F_j} J\, dS,$$

where

$$J=c^2\nabla u_1^h\cdot\mathbf{n}_1\left(v^*-v_1^h\right)+c^2 v_1^h(\nabla u)^*\cdot\mathbf{n}_1+c^2\nabla u_2^h\cdot\mathbf{n}_2\left(v^*-v_2^h\right)+c^2 v_2^h(\nabla u)^*\cdot\mathbf{n}_2. \tag{9}$$

We first introduce the so-called *central flux*,

$$v^*\equiv\frac{1}{2}\left(v_1^h+v_2^h\right),\quad (\nabla u)^*\equiv\frac{1}{2}\left(\nabla u_1^h+\nabla u_2^h\right). \tag{10}$$

Plug this back into (9) and use $\mathbf{n}_1=-\mathbf{n}_2$. Then we have

$$J=\frac{1}{2}\left(c^2\nabla u_1^h\cdot\mathbf{n}_1\left(v_2^h-v_1^h\right)+c^2 v_1^h\left(\nabla u_1^h+\nabla u_2^h\right)\cdot\mathbf{n}_1+c^2\nabla u_2^h\cdot\mathbf{n}_2\left(v_1^h-v_2^h\right)+c^2 v_2^h\left(\nabla u_1^h+\nabla u_2^h\right)\cdot\mathbf{n}_2\right)=0.$$

Second, we propose an *alternating flux*,

$$v^*\equiv v_1^h,\quad (\nabla u)^*\equiv\nabla u_2^h, \tag{11}$$

or

$$v^*\equiv v_2^h,\quad (\nabla u)^*\equiv\nabla u_1^h. \tag{12}$$

Using (11) as an example, we have

$$J=c^2\nabla u_1^h\cdot\mathbf{n}_1\left(v_1^h-v_1^h\right)+c^2 v_1^h\nabla u_2^h\cdot\mathbf{n}_1+c^2\nabla u_2^h\cdot\mathbf{n}_2\left(v_1^h-v_2^h\right)+c^2 v_2^h\nabla u_2^h\cdot\mathbf{n}_2=0.$$

If $\theta=0$, then it is clear that both the central flux (10) and the alternating flux (11) or (12) lead to an energy-conserving energy-based DG scheme since $J=0$. To develop an energy-dissipating scheme for $\theta=0$, we introduce a *Sommerfeld flux* which yields $J < 0$ in the presence of jumps. Let us denote a flux splitting parameter by $\xi > 0$ which has the same units as the wave speed c and note that,

$$v\nabla u \cdot \mathbf{n} = \frac{1}{4\xi}(v + \xi \nabla u \cdot \mathbf{n})^2 - \frac{1}{4\xi}(v - \xi \nabla u \cdot \mathbf{n})^2.$$

Then we enforce

$$\begin{cases} v^* - \xi(\nabla u)^* \cdot \mathbf{n}_1 = v_1^h - \xi\left(\nabla u_1^h\right) \cdot \mathbf{n}_1, \\ v^* - \xi(\nabla u)^* \cdot \mathbf{n}_2 = v_2^h - \xi\left(\nabla u_2^h\right) \cdot \mathbf{n}_2. \end{cases} \tag{13}$$

Solving system (13) yields

$$v^* = \{\{v^h\}\} - \frac{\xi}{2}[[\nabla u^h]], \quad (\nabla u)^* = \{\{\nabla u^h\}\} - \frac{1}{2\xi}[[v^h]]. \tag{14}$$

Plugging (14) into (9) we obtain

$$J = -\left(\frac{\xi c^2}{2}[[\nabla u^h]]^2 + \frac{c^2}{2\xi}\left|[[v^h]]\right|^2\right) < 0.$$

Thus we have an energy-dissipating scheme even when $\theta = 0$ if the *Sommerfeld flux* is used.

2.2.2. Physical boundaries

In this section, we focus on the boundary condition,

$$\gamma \frac{\partial u(\mathbf{x}, t)}{\partial t} + \eta c \nabla u(\mathbf{x}, t) \cdot \mathbf{n} = 0, \quad \mathbf{x} \in \partial\Omega, \tag{15}$$

where $\gamma^2 + \eta^2 = 1$ and $\gamma, \eta \geq 0$. Note that we have a homogeneous Dirichlet boundary condition if $\eta = 0$ and a homogeneous Neumann boundary condition when $\gamma = 0$. On the one hand, multiplying (15) by $\gamma c \nabla u \cdot \mathbf{n}$ gives

$$\gamma^2 \frac{\partial u}{\partial t} c \nabla u \cdot \mathbf{n} + \gamma\eta(c\nabla u \cdot n)^2 = 0, \tag{16}$$

on the other hand, multiplying (15) by $\eta\frac{\partial u}{\partial t}$ yields,

$$\eta\gamma\left(\frac{\partial u}{\partial t}\right)^2 + \eta^2 \frac{\partial u}{\partial t} c \nabla u \cdot \mathbf{n} = 0. \tag{17}$$

Combining (16) and (17), from (4) we have

$$\frac{dE}{dt} = -\int\limits_{\Omega} \theta \left(\frac{\partial u}{\partial t}\right)^2 d\mathbf{x} - \int\limits_{\partial\Omega} \gamma\eta c \left(\left(\frac{\partial u}{\partial t}\right)^2 + (c\nabla u \cdot \mathbf{n})^2\right) dS \leq 0.$$

The numerical fluxes v^* and $(\nabla u)^*$ are chosen to be consistent with the physical boundary condition (15),

$$\gamma v^* + \eta c (\nabla u)^* \cdot \mathbf{n} = 0. \tag{18}$$

By a similar analysis as in [1], the following family satisfies the consistency condition (18)

$$v^* = v^h - (\gamma - a\eta)\rho, \quad c(\nabla u)^* = c\nabla u^h - (\eta + a\gamma)\rho\mathbf{n},$$

with

$$\rho = \gamma v^h + \eta c\left(\nabla u^h\right) \cdot \mathbf{n}.$$

Then the contribution to the discrete energy from the physical boundaries, with element faces on physical boundaries denoted by B_j, is given by

$$\begin{aligned} \left.\frac{dE^h}{dt}\right|_{\partial\Omega} &= \sum_j \int\limits_{B_j} c^2 \nabla u^h \cdot \mathbf{n}(v^* - v^h) + c^2 v^h (\nabla u)^* \cdot \mathbf{n}\, dS \\ &= \sum_j \int\limits_{B_j} -c^2 \nabla u^h \cdot \mathbf{n}(\gamma - a\eta)\rho + c^2\left(v^* + (\gamma - a\eta)\rho\right)(\nabla u)^* \cdot \mathbf{n}\, dS \end{aligned}$$

$$
\begin{aligned}
&= \sum_j \int_{B_j} c^2 v^*(\nabla u)^* \cdot \mathbf{n} - c^2 \nabla u^h \cdot \mathbf{n}(\gamma - a\eta)\rho + c(\gamma - a\eta)\rho \left(c\nabla u^h \cdot \mathbf{n} - (\eta + a\gamma)\rho \right) dS \\
&= -\sum_j \int_{B_j} \gamma\eta c\Big((v^*)^2 + \big(c(\nabla u)^* \cdot \mathbf{n}\big)^2\Big) + c\rho^2\big((1-a^2)\gamma\eta + a(\gamma^2 - \eta^2)\big)\, dS
\end{aligned}
$$

which yields a nonincreasing contribution to the energy if

$$
b = (1-a^2)\gamma\eta + a(\gamma^2 - \eta^2) \geq 0.
$$

2.3. Stability of the scheme

We are now ready to establish the stability of the proposed energy-based DG scheme. To make the statement concise, we introduce a general formulation for the fluxes on the interelement boundaries,

$$
v^* \equiv \alpha v_1^h + (1-\alpha) v_2^h - \tau[[\nabla u^h]], \quad (\nabla u)^* \equiv (1-\alpha)\nabla u_1^h + \alpha \nabla u_2^h - \beta[[v^h]], \tag{19}
$$

with $0 \leq \alpha \leq 1$ and $\beta, \tau \geq 0$. Here τ has the same units as c and β units of c^{-1}. Note that the previous cases correspond to:

Central flux: $\alpha = 0.5, \tau = \beta = 0$.
Alternating flux: $\alpha = 0, \tau = \beta = 0$ or $\alpha = 1, \tau = \beta = 0$.
Sommerfeld flux: $\alpha = 0.5, \tau = \xi/2, \beta = 1/(2\xi)$.
Alternating flux with Sommerfeld flux: $\alpha = 0, \tau = \xi/2, \beta = 1/(2\xi)$ or $\alpha = 1, \tau = \xi/2, \beta = 1/(2\xi)$.

For the general flux formulation (19), we find that the contribution to the discrete energy from the interelement boundaries is the boundary integral of

$$
J = -c^2 \left(\beta|[[v^h]]|^2 + \tau[[\nabla u^h]]^2 \right) \leq 0.
$$

Theorem 1. *The discrete energy $E^h(t) = \sum_j E_j^h(t)$ with $E_j^h(t)$ defined in (5) satisfies*

$$
\begin{aligned}
\frac{dE^h}{dt} = &-\sum_j \int_{\Omega_j} \theta \left(v^h\right)^2 d\mathbf{x} - \sum_j \int_{F_j} c^2 \left(\beta|[[v^h]]|^2 + \tau[[\nabla u^h]]^2 \right) dS \\
&-\sum_j \int_{B_j} c\gamma\eta\Big((v^*)^2 + \big(c(\nabla u)^* \cdot \mathbf{n}\big)^2\Big) + bc\rho^2\, dS,
\end{aligned}
$$

where F_j is an interelement boundary and B_j is on the physical boundaries. If the flux parameters τ, β and b are non-negative, then $E^h(t) \leq E^h(0)$.

3. Error estimates

To analyze the numerical error of the scheme, we define the errors by

$$
e_u = u - u^h, \quad e_v = v - v^h, \tag{20}
$$

and compare (u^h, v^h) with an arbitrary polynomial $(\tilde{u}^h, \tilde{v}^h)$, $\tilde{u}^h \in U_h^q$, $\tilde{v}^h \in V_h^s$ with $q-2 \leq s \leq q$. To proceed, we denote the differences

$$
\tilde{e}_u = \tilde{u}^h - u^h, \quad \tilde{e}_v = \tilde{v}^h - v^h, \quad \delta_u = \tilde{u}^h - u, \quad \delta_v = \tilde{v}^h - v,
$$

and consider the numerical error energy

$$
\mathcal{E} = \sum_j \int_{\Omega_j} \frac{1}{2} c^2 |\nabla \tilde{e}_u|^2 + \frac{1}{2}\tilde{e}_v^2\, d\mathbf{x} - \sum_{\mathbf{k},j} \int_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})} \omega_{\mathbf{k},j} \frac{f(\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z)}{\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z} z dz. \tag{21}
$$

Here we assume:

$$
\frac{f(u)}{u} \leq -L < 0, \tag{22}
$$

which will guarantee the positivity of the numerical error energy $\mathcal{E}$. However, this restriction can be relaxed as we show in the remark below. The term $-\int_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})} \omega_{\mathbf{k},j} \frac{f(\tilde{u}^h(\mathbf{x}_{\mathbf{k},j})-z)}{\tilde{u}^h(\mathbf{x}_{\mathbf{k},j})-z} z dz$ is introduced to cancel terms involving $\frac{\partial \tilde{e}_u}{\partial t}$ in the calculation which follows. It also allows us to bound $\tilde{e}_u^2$ in the error analysis since

$$-\int_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})} \omega_{\mathbf{k},j} \frac{f(\tilde{u}^h(\mathbf{x}_{\mathbf{k},j})-z)}{\tilde{u}^h(\mathbf{x}_{\mathbf{k},j})-z} z dz \geq \min\left(-\frac{f(w)}{w}\right) \omega_{\mathbf{k},j} \int_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})} z\, dz = \frac{L}{2} \omega_{\mathbf{k},j} \tilde{e}_u^2(\mathbf{x}_{\mathbf{k},j}).$$

For the linear Klein-Gordon equation, where $f(u)/u$ is constant, it would reduce to the standard energy.

Since $\frac{\partial u}{\partial t} - v = 0$ both the continuous solution (u, v) and the numerical solution (u^h, v^h) satisfy (7). Therefore the error satisfies the following equation:

$$\int_{\Omega_j} c^2 \nabla \phi_u \cdot \nabla \left(\frac{\partial e_u}{\partial t} - e_v\right) d\mathbf{x} - \sum_{\mathbf{k}} \omega_{\mathbf{k},j} \phi_u(\mathbf{x}_{\mathbf{k},j}) \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} \left(\frac{\partial e_u}{\partial t}(\mathbf{x}_{\mathbf{k},j}) - e_v(\mathbf{x}_{\mathbf{k},j})\right) = \int_{\partial \Omega_j} c^2 \nabla \phi_u \cdot \mathbf{n} \left(e_v^* - e_v\right) dS. \tag{23}$$

Similarly, we note that (u, v) satisfies an equation analogous to (8) with the discrete quadrature replaced by an integral. Therefore we have:

$$\int_{\Omega_j} \phi_v \frac{\partial e_v}{\partial t} + c^2 \nabla \phi_v \cdot \nabla e_u + \theta \phi_v e_v \, d\mathbf{x} - \int_{\Omega_j} \phi_v f(u)\, d\mathbf{x} + \sum_{\mathbf{k}} \omega_{\mathbf{k},j} \phi_v(\mathbf{x}_{\mathbf{k},j}) f(u^h(\mathbf{x}_{\mathbf{k},j})) = \int_{\partial \Omega_j} c^2 \phi_v (\nabla e_u)^* \cdot \mathbf{n}\, dS. \tag{24}$$

Now, by using the relations $e_u = \tilde{e}_u - \delta_u$, $e_v = \tilde{e}_v - \delta_v$, choosing $\phi_u = \tilde{e}_u$, $\phi_v = \tilde{e}_v$, rearranging terms involving the integral of $f(u)$, and summing (23) and (24), we obtain

$$\begin{aligned}
\int_{\Omega_j} c^2 \nabla \tilde{e}_u \cdot \nabla \frac{\partial \tilde{e}_u}{\partial t} + \tilde{e}_v \frac{\partial \tilde{e}_v}{\partial t}\, d\mathbf{x} = & \int_{\Omega_j} c^2 \nabla \tilde{e}_u \cdot \nabla \left(\tilde{e}_v + \left(\frac{\partial \delta_u}{\partial t} - \delta_v\right)\right) + \tilde{e}_v \frac{\partial \delta_v}{\partial t} - c^2 \nabla \tilde{e}_v \cdot \nabla (\tilde{e}_u - \delta_u) - \theta \tilde{e}_v (\tilde{e}_v - \delta_v)\, d\mathbf{x} \\
& + \sum_{\mathbf{k}} \omega_{\mathbf{k},j} \tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \Big(f(u(\mathbf{x}_{\mathbf{k},j})) - f(u^h(\mathbf{x}_{\mathbf{k},j}))\Big) + \omega_{\mathbf{k},j} \tilde{e}_u(\mathbf{x}_{\mathbf{k},j}) \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} \left(\frac{\partial e_u(\mathbf{x}_{\mathbf{k},j})}{\partial t} - e_v(\mathbf{x}_{\mathbf{k},j})\right) \\
& + \left(\int_{\Omega_j} \tilde{e}_v f(u)\, d\mathbf{x} - \sum_{\mathbf{k}} \omega_{\mathbf{k},j} \tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) f(u(\mathbf{x}_{\mathbf{k},j}))\right) \\
& + \int_{\partial \Omega_j} c^2 \nabla \tilde{e}_u \cdot \mathbf{n} (\tilde{e}_v^* - \delta_v^* - (\tilde{e}_v - \delta_v)) + c^2 \tilde{e}_v \left((\nabla \tilde{e}_u)^* \cdot \mathbf{n} - (\nabla \delta_u)^* \cdot \mathbf{n}\right) dS.
\end{aligned} \tag{25}$$

An integration by parts in the volume integral $\int_{\Omega_j} c^2 \nabla \tilde{e}_u \cdot \nabla \delta_v \, d\mathbf{x}$ simplifies (25) to

$$\begin{aligned}
\int_{\Omega_j} c^2 \nabla \tilde{e}_u \cdot \nabla \frac{\partial \tilde{e}_u}{\partial t} + \tilde{e}_v \frac{\partial \tilde{e}_v}{\partial t}\, d\mathbf{x} = & \int_{\Omega_j} c^2 \nabla \tilde{e}_u \cdot \nabla \frac{\partial \delta_u}{\partial t} + c^2 \Delta \tilde{e}_u \delta_v + \tilde{e}_v \frac{\partial \delta_v}{\partial t} + c^2 \nabla \tilde{e}_v \cdot \nabla \delta_u - \theta \tilde{e}_v (\tilde{e}_v - \delta_v)\, d\mathbf{x} \\
& + \sum_{\mathbf{k}} \omega_{\mathbf{k},j} \tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \Big(f(u(\mathbf{x}_{\mathbf{k},j})) - f(u^h(\mathbf{x}_{\mathbf{k},j}))\Big) + \omega_{\mathbf{k},j} \tilde{e}_u(\mathbf{x}_{\mathbf{k},j}) \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} \left(\frac{\partial e_u(\mathbf{x}_{\mathbf{k},j})}{\partial t} - e_v(\mathbf{x}_{\mathbf{k},j})\right) \\
& + \left(\int_{\Omega_j} \tilde{e}_v f(u)\, d\mathbf{x} - \sum_{\mathbf{k}} \omega_{\mathbf{k},j} \tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) f(u(\mathbf{x}_{\mathbf{k},j}))\right) \\
& + \int_{\partial \Omega_j} c^2 \nabla \tilde{e}_u \cdot \mathbf{n} (\tilde{e}_v^* - \tilde{e}_v) + c^2 \tilde{e}_v (\nabla \tilde{e}_u)^* \cdot \mathbf{n} - c^2 \nabla \tilde{e}_u \cdot \mathbf{n} \delta_v^* - c^2 \tilde{e}_v (\nabla \delta_u)^* \cdot \mathbf{n}\, dS.
\end{aligned} \tag{26}$$

We now must choose suitable $(\tilde{u}^h, \tilde{v}^h)$ to achieve an acceptable error. Note that in what follows we will assume for simplicity that $(u^h, v^h) = (\tilde{u}^h, \tilde{v}^h)$ at $t = 0$, though we do not satisfy this condition in the numerical experiments. On Ω_j, we impose for all time t and $\forall \phi_u \in U_h^q, \forall \phi_v \in U_h^s$,

$$\int_{\Omega_j} \nabla\phi_u \cdot \nabla\delta_u \, d\mathbf{x} = 0, \quad \int_{\Omega_j} \delta_u \, d\mathbf{x} = 0, \quad \int_{\Omega_j} \phi_v \delta_v \, d\mathbf{x} = 0. \tag{27}$$

The solvability of both the H^1 projection equation for $\tilde{u}^h$ and the L^2 projection equation for $\tilde{v}$ follows from counting and uniqueness arguments. The systems to be satisfied by $\tilde{u}^h$ and $\tilde{v}^h$ on element Ω_j are linear, and the mass matrix determining $\tilde{v}^h$ from the third equation in (27) is obviously nonsingular. To show that $\tilde{u}^h$ is uniquely determined we first note that the number of linear equations matches the dimensionality of Π^q; that is, the dimension equals the number of independent equations represented by the first equation in (27) plus those represented by the second equation in (27). To establish uniqueness note that for zero data we could choose, $\phi_u = \delta_u$ implying that δ_u is constant, and then apply the second equation to show that the constant must be 0. In what follows, we also introduce fluxes $\delta_v^*, \nabla\delta_u^*$ built from $\delta_v, \nabla\delta_u$ according to the specification in Section 2.2.

Since $q-2 \le s \le q$, the first to the fourth volume term as well as the term involving $\theta\tilde{e}_v\delta_v$ on the right hand side of (26) all vanish. Substituting $e_u = \tilde{e}_u - \delta_u$, $e_v = \tilde{e}_v - \delta_v$, the equation (26) then yields

$$\begin{aligned}
\int_{\Omega_j} c^2 \nabla\tilde{e}_u \cdot \nabla\frac{\partial\tilde{e}_u}{\partial t} + \tilde{e}_v \frac{\partial\tilde{e}_v}{\partial t} \, d\mathbf{x} = & -\int_{\Omega_j} \theta\tilde{e}_v^2 \, d\mathbf{x} + \sum_{\mathbf{k}} \omega_{\mathbf{k},j}\tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \left(f(u(\mathbf{x}_{\mathbf{k},j})) - f(u^h(\mathbf{x}_{\mathbf{k},j})) \right) \\
+ \sum_{\mathbf{k}} \omega_{\mathbf{k},j}\tilde{e}_u(\mathbf{x}_{\mathbf{k},j}) \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} & \left(\frac{\partial\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})}{\partial t} - \tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \right) - \omega_{\mathbf{k},j}\tilde{e}_u(\mathbf{x}_{\mathbf{k},j}) \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}^j_{\mathbf{k}})} \left(\frac{\partial\delta_u(\mathbf{x}_{\mathbf{k},j})}{\partial t} - \delta_v(\mathbf{x}_{\mathbf{k},j}) \right) \\
& + \left(\int_{\Omega_j} \tilde{e}_v f(u) \, d\mathbf{x} - \sum_{\mathbf{k}} \omega_{\mathbf{k},j}\tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) f(u(\mathbf{x}_{\mathbf{k},j})) \right) \\
& + \int_{\partial\Omega_j} c^2 \nabla\tilde{e}_u \cdot \mathbf{n}(\tilde{e}_v^* - \tilde{e}_v) + c^2\tilde{e}_v(\nabla\tilde{e}_u)^* \cdot \mathbf{n} - c^2\nabla\tilde{e}_u \cdot \mathbf{n}\delta_v^* - c^2\tilde{e}_v(\nabla\delta_u)^* \cdot \mathbf{n} \, dS.
\end{aligned} \tag{28}$$

First, we rewrite the third term on the right-hand side of (28)

$$\begin{aligned}
\sum_{\mathbf{k}} \omega_{\mathbf{k},j}\tilde{e}_u(\mathbf{x}_{\mathbf{k},j}) \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} \left(\frac{\partial\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})}{\partial t} - \tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \right) = & -\sum_{\mathbf{k}} \omega_{\mathbf{k},j} \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} \tilde{e}_u(\mathbf{x}_{\mathbf{k},j})\tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \\
+ \frac{d}{dt} \sum_{\mathbf{k}} \int_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})} \omega_{\mathbf{k},j} \frac{f(\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z)}{\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z} z \, dz & - \sum_{\mathbf{k}} \int_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})} \omega_{\mathbf{k},j} \frac{d}{dt} \left(\frac{f(\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z)}{\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z} \right) z \, dz.
\end{aligned}$$

Then combining this with (21) and (28) gives

$$\begin{aligned}
\frac{d\mathcal{E}}{dt} = & \sum_j \int_{\Omega_j} c^2 \nabla\tilde{e}_u \cdot \nabla\frac{\partial\tilde{e}_u}{\partial t} + \tilde{e}_v \frac{\partial\tilde{e}_v}{\partial t} \, d\mathbf{x} - \frac{d}{dt} \sum_{\mathbf{k}} \int_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})} \omega_{\mathbf{k},j} \frac{f(\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z)}{\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z} z \, dz \\
= & -\sum_j \int_{\Omega_j} \theta\tilde{e}_v^2 \, d\mathbf{x} + \sum_{\mathbf{k},j} \omega_{\mathbf{k},j}\tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \left(f(u(\mathbf{x}_{\mathbf{k},j})) - f(u^h(\mathbf{x}_{\mathbf{k},j})) \right) - \omega_{\mathbf{k},j} \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} \tilde{e}_u(\mathbf{x}_{\mathbf{k},j})\tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \\
& - \sum_{\mathbf{k},j} \int_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})} \omega_{\mathbf{k},j} \frac{d}{dt} \left(\frac{f(\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z)}{\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z} \right) z + \omega_{\mathbf{k},j}\tilde{e}_u(\mathbf{x}_{\mathbf{k},j}) \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} \left(\frac{\partial\delta_u(\mathbf{x}_{\mathbf{k},j})}{\partial t} - \delta_v(\mathbf{x}_{\mathbf{k},j}) \right) dz \\
& + \sum_j \left(\int_{\Omega_j} \tilde{e}_v f(u) \, d\mathbf{x} - \sum_{\mathbf{k}} \omega_{\mathbf{k},j}\tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) f(u(\mathbf{x}_{\mathbf{k},j})) \right) \\
& + \sum_j \int_{\partial\Omega_j} c^2 \nabla\tilde{e}_u \cdot \mathbf{n}(\tilde{e}_v^* - \tilde{e}_v) + c^2\tilde{e}_v(\nabla\tilde{e}_u)^* \cdot \mathbf{n} - c^2\nabla\tilde{e}_u \cdot \mathbf{n}\delta_v^* - c^2\tilde{e}_v(\nabla\delta_u)^* \cdot \mathbf{n} \, dS.
\end{aligned} \tag{29}$$

Combining the contributions from neighboring elements we rewrite the boundary terms in (29) to obtain

$$\begin{aligned}
\frac{d\mathcal{E}}{dt} = & -\sum_j \int\limits_{\Omega_j} \theta \tilde{e}_v^2 \, d\mathbf{x} + \sum_{\mathbf{k},j} \omega_{\mathbf{k},j} \tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \left(f(u(\mathbf{x}_{\mathbf{k},j})) - f(u^h(\mathbf{x}_{\mathbf{k},j})) \right) - \omega_{\mathbf{k},j} \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} \tilde{e}_u(\mathbf{x}_{\mathbf{k},j}) \tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) \\
& - \sum_{\mathbf{k},j} \int\limits_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})} \omega_{\mathbf{k},j} \frac{d}{dt} \left(\frac{f(\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z)}{\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z} \right) z + \omega_{\mathbf{k},j} \tilde{e}_u(\mathbf{x}_{\mathbf{k},j}) \frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})} \left(\frac{\partial \delta_u(\mathbf{x}_{\mathbf{k},j})}{\partial t} - \delta_v(\mathbf{x}_{\mathbf{k},j}) \right) dz \\
& + \sum_j \left(\int\limits_{\Omega_j} \tilde{e}_v f(u) \, d\mathbf{x} - \sum_{\mathbf{k}} \omega_{\mathbf{k},j} \tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) f(u(\mathbf{x}_{\mathbf{k},j})) \right) \qquad (30) \\
& - \sum_j \int\limits_{B_j} c\gamma\eta \Big((\tilde{e}_v^*)^2 + \big(c(\nabla \tilde{e}_u)^* \cdot \mathbf{n} \big)^2 \Big) + bc(\gamma \tilde{e}_v + \eta c \nabla \tilde{e}_u \cdot \mathbf{n})^2 + c^2 \nabla \tilde{e}_u \cdot \mathbf{n} \delta_v^* + c^2 \tilde{e}_v (\nabla \delta_u)^* \cdot \mathbf{n} dS \\
& - \sum_j \int\limits_{F_j} c^2 \left(\beta |[[\tilde{e}_v]]|^2 + \tau [[\nabla \tilde{e}_u]]^2 \right) + c^2 [[\nabla \tilde{e}_u]] \delta_v^* + c^2 [[\tilde{e}_v]] \cdot (\nabla \delta_u)^* \, dS.
\end{aligned}$$

Here, again, F_j represents interelement boundaries and B_j represents physical boundaries.

In what follows, C is a constant independent of the element diameter h for a shape-regular mesh. It may differ from line to line. We denote Sobolev norms by $|| \cdot ||$ and the associated seminorms by $| \cdot |$. Here we will assume the solution is sufficiently smooth up to some time, T. We will also make assumptions on the boundedness of $f(w)$, $\frac{f(w)}{w}$, and $\frac{df}{dw}(w)$. Obviously these need only hold in some neighborhood of the actual values attained by u up to time T, but for simplicity we will assume they hold for all arguments, w. In addition, to simplify the statement of the main Theorem, we will assume:

$$s > \frac{d}{2} - 1, \qquad (31)$$

in which case functions in $H^{s+1}(\Omega)$ are continuous.

Lastly, we must assume that the quadrature formula is sufficiently accurate.

Assumption 1. The quadrature rule satisfies for $\phi \in \Pi^q$, all Ω_j, and all $g \in H^{s+1}(\Omega)$,

$$\sum_{\mathbf{k}} \omega_{\mathbf{k},j} \phi^2(\mathbf{x}_{\mathbf{k},j}) - \int\limits_{\Omega_j} \phi^2 \, d\mathbf{x} = 0, \qquad (32)$$

$$\sum_j \left| \sum_{\mathbf{k}} \omega_{\mathbf{k},j} \phi(\mathbf{x}_{\mathbf{k},j}) g(\mathbf{x}_{\mathbf{k},j}) - \int\limits_{\Omega_j} \phi \, g \, d\mathbf{x} \right| \leq C h^{s+1} \|\phi\|_{L^2(\Omega)} |g|_{H^{s+1}(\Omega)}. \qquad (33)$$

We then have the following error estimate.

Theorem 2. *Let $\bar{q} = min(q-1, s)$, $q - 2 \leq s \leq q$, with s satisfying (31). Suppose $\frac{f(w)}{w}$ is a smooth bounded function satisfying the upper bound (22) and that Assumption 1 holds. Then there exist numbers C, C_1, depending only on $s, q, \xi, \beta, \tau, b$, the bounds of $\frac{df(w)}{dw}$, $\frac{f(w)}{w}$, $\|u\|_{L^\infty([0,T], H^{\bar{q}+2}(\Omega))}$, $\|v\|_{L^\infty([0,T], H^{s+1}(\Omega))}$, and the shape regularity of the mesh, but independent of h, such that for $0 \leq t \leq T$:*

$$||\nabla e_u(\cdot, t)||^2_{L^2(\Omega)} + ||e_v(\cdot, t)||^2_{L^2(\Omega)} \leq C e^{C_1 t} h^{2\zeta}, \qquad (34)$$

where

$$\zeta = \begin{cases} \bar{q}, & \beta, \tau, b \geq 0, \\ \bar{q} + \frac{1}{2}, & \beta, \tau, b > 0. \end{cases}$$

Proof. Note that $\bar{q} + 2 \geq s + 1$, so that by (31) both u and v are continuous. Many terms in the error estimate will be of order $s+1$ but in the end these will be subsumed in the end by the order ζ terms. From the Bramble-Hilbert lemma (e.g., [13]), we have

$$\begin{aligned}
&||\delta_u||^2_{L^2(\Omega)} \le Ch^{2s+2}\,|u(\cdot,t)|^2_{H^{s+1}(\Omega)}\,, \quad ||\delta_v||^2_{L^2(\Omega)} \le Ch^{2s+2}\,|v(\cdot,t)|^2_{H^{s+1}(\Omega)}\,,\\
&\left|\left|\frac{\partial \delta_u}{\partial t}\right|\right|^2_{L^2(\Omega)} \le Ch^{2s+2}\left|\frac{\partial u(\cdot,t)}{\partial t}\right|^2_{H^{s+1}(\Omega)}.
\end{aligned}\tag{35}$$

We also note that since our approximations are piecewise polynomial and by assumption u and v are bounded we have

$$\|\tilde{u}^h\|_{L^\infty(\Omega)} \le C, \quad \|\frac{\partial \tilde{u}^h}{\partial t}\|_{L^\infty(\Omega)} \le C. \tag{36}$$

Now we estimate the nonlinear volume integrals in (30). By the mean value theorem for $f(w)$, the Cauchy-Schwarz inequality, (32)-(33), and (35) we obtain

$$\begin{aligned}
&\sum_{\mathbf{k},j}\omega_{\mathbf{k},j}\tilde{e}_v(\mathbf{x}_{\mathbf{k},j})\left(f(u(\mathbf{x}_{\mathbf{k},j})) - f(u^h(\mathbf{x}_{\mathbf{k},j}))\right)\\
&= \sum_{\mathbf{k},j}\omega_{\mathbf{k},j}\tilde{e}_v(\mathbf{x}_{\mathbf{k},j})\left.\frac{df(w)}{dw}\right|_{w=u(\mathbf{x}_{\mathbf{k},j})+\vartheta e_u}\left(u(\mathbf{x}_{\mathbf{k},j}) - u^h(\mathbf{x}_{\mathbf{k},j})\right)\\
&\le \sum_{\mathbf{k},j} C\omega_{\mathbf{k},j}\left|\tilde{e}_v(\mathbf{x}_{\mathbf{k},j})\right|\left|u(\mathbf{x}_{\mathbf{k},j}) - u^h(\mathbf{x}_{\mathbf{k},j})\right|\\
&\le \sum_{\mathbf{k},j} C\omega_{\mathbf{k},j}\left|\tilde{e}_v(\mathbf{x}_{\mathbf{k},j})\right|\left(\left|\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})\right| + \left|\delta_u(\mathbf{x}_{\mathbf{k},j})\right|\right)\\
&\le C\mathcal{E} + Ch^{s+1}\sqrt{\mathcal{E}}\,|u(\cdot,t)|_{H^{s+1}(\Omega)}
\end{aligned}\tag{37}$$

with $\vartheta \in [-1, 0]$ and

$$\begin{aligned}
&-\sum_{\mathbf{k},j}\omega_{\mathbf{k},j}\frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})}\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})\tilde{e}_v(\mathbf{x}_{\mathbf{k},j}) + \int_0^{\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})}\omega_{\mathbf{k},j}\frac{d}{dt}\left(\frac{f(\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z)}{\tilde{u}^h(\mathbf{x}_{\mathbf{k},j}) - z}\right)z\,dz\\
&+\omega_{\mathbf{k},j}\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})\frac{f(u^h(\mathbf{x}_{\mathbf{k},j}))}{u^h(\mathbf{x}_{\mathbf{k},j})}\left(\frac{\partial\delta_u(\mathbf{x}_{\mathbf{k},j})}{\partial t} - \delta_v(\mathbf{x}_{\mathbf{k},j})\right)\\
&+\sum_j\left(\int_{\Omega_j}\tilde{e}_v f(u)\,d\mathbf{x} - \sum_{\mathbf{k}}\omega_{\mathbf{k},j}\tilde{e}_v(\mathbf{x}_{\mathbf{k},j})f(u(\mathbf{x}_{\mathbf{k},j}))\right)\\
&\le \sum_{\mathbf{k},j}\omega_{\mathbf{k},j}\max\left|\frac{f(w)}{w}\right|\left|\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})\tilde{e}_v(\mathbf{x}_{\mathbf{k},j})\right| + \frac{\omega_{\mathbf{k},j}}{2}\max\left|\frac{df}{dw}\right|\max\left|\frac{d\tilde{u}^h}{dt}\right|\tilde{e}_u^2(\mathbf{x}_{\mathbf{k},j})\\
&+\omega_{\mathbf{k},j}\max\left|\frac{f(w)}{w}\right|\left(\left|\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})\frac{\partial\delta_u(\mathbf{x}_{\mathbf{k},j})}{\partial t}\right| + \left|\tilde{e}_u(\mathbf{x}_{\mathbf{k},j})\delta_v(\mathbf{x}_{\mathbf{k},j})\right|\right) + Ch^{s+1}\|\tilde{e}_v\|\,|f(u(\cdot,t))|_{H^{s+1}(\Omega)}\\
&\le C\mathcal{E} + Ch^{s+1}\sqrt{\mathcal{E}}\,|v(\cdot,t)|_{H^{s+1}(\Omega)}\,.
\end{aligned}\tag{38}$$

Then, using (37)-(38), (30) is simplified to

$$\begin{aligned}
\frac{d\mathcal{E}}{dt} &\le C\mathcal{E} + Ch^{s+1}\sqrt{\mathcal{E}}\left(|v(\cdot,t)|_{H^{s+1}(\Omega)} + |u(\cdot,t)|_{H^{s+1}(\Omega)}\right)\\
&\quad -\sum_j\int_{B_j} c\gamma\eta\left((\tilde{e}_v^*)^2 + \left(c(\nabla\tilde{e}_u)^*\cdot\mathbf{n}\right)^2\right) + bc(\gamma\tilde{e}_v + \eta c\nabla\tilde{e}_u\cdot\mathbf{n})^2 + c^2\nabla\tilde{e}_u\cdot\mathbf{n}\delta_v^* + c^2\tilde{e}_v(\nabla\delta_u)^*\cdot\mathbf{n}dS\\
&\quad -\sum_j\int_{F_j} c^2\left(\beta|[[\tilde{e}_v]]|^2 + \tau[[\nabla\tilde{e}_u]]^2\right) - c^2[[\nabla\tilde{e}_u]]\delta_v^* + c^2[[\tilde{e}_v]]\cdot(\nabla\delta_u)^*\,dS.
\end{aligned}$$

Now, we only need to consider the boundary integrals. We use the same analysis as in [1] and complete the estimates for the following cases:

Case I: $\beta = 0$ or $\tau = 0$,

$$\frac{d\mathcal{E}}{dt} \le C\mathcal{E} + Ch^{s+1}\sqrt{\mathcal{E}}\left(|v(\cdot,t)|_{H^{s+1}(\Omega)} + |u(\cdot,t)|_{H^{s+1}(\Omega)}\right) + Ch^{\bar{q}}\sqrt{\mathcal{E}}\left(|u(\cdot,t)|_{H^{\bar{q}+2}(\Omega)} + |v(\cdot,t)|_{H^{\bar{q}+1}(\Omega)}\right). \tag{39}$$

Then, combining a direct integration of (39) in time with the assumption $\tilde{e}_u = \tilde{e}_v = 0$ at $t = 0$, we obtain

$$\sqrt{\mathcal{E}}(t) \le C\left(e^{Ct}-1\right)\max_t\left(h^{\bar{q}}\left(|u(\cdot,t)|_{H^{\bar{q}+2}(\Omega)}+|v(\cdot,t)|_{H^{\bar{q}+1}(\Omega)}\right)+h^{s+1}\left(|v(\cdot,t)|_{H^{s+1}(\Omega)}+|u(\cdot,t)|_{H^{s+1}(\Omega)}\right)\right), \tag{40}$$

since $\tilde{e}_u = e_u + \delta_u$, $\tilde{e}_v = e_v + \delta_v$, then (34) follows from the triangle inequality and (40).

Case II: $\beta, \tau, b > 0$,

$$\frac{d\mathcal{E}}{dt} \le C\mathcal{E} + Ch^{s+1}\sqrt{\mathcal{E}}\left(|v(\cdot,t)|_{H^{s+1}(\Omega)}+|u(\cdot,t)|_{H^{s+1}(\Omega)}\right) + Ch^{\bar{q}+1/2}\sqrt{\mathcal{E}}\left(|u(\cdot,t)|_{H^{\bar{q}+2}(\Omega)}+|v(\cdot,t)|_{H^{\bar{q}+1}(\Omega)}\right). \tag{41}$$

Then again (34) with $\zeta = \bar{q} + \frac{1}{2}$ follows directly from an integration in time of (41) combined with the triangle inequality. □

Remark. If $\frac{f(u)}{u} \ge 0$ for some u we may introduce a new variable $u = e^{\chi t}w$, $\chi > 0$ and use the energy-based DG scheme to solve for w. Then so long as $\chi^2 + \chi\theta - \frac{f(u)}{u}$ is positive the hypotheses above are satisfied and so the energy and error estimates hold. This applies, for example, to the sine-Gordon equation. However, in our numerical experiments we solve for u rather than w.

Remark. For one-dimensional problems with flux parameters satisfying $\alpha(1-\alpha) = \beta\tau$, we can improve the error estimate to h^{s+1} with $q = s + 1$ by constructing $(\tilde{u}^h, \tilde{v}^h)$ to make the boundary term in (30) vanish as in [1].

Remark. We note that the error estimate appears to be overly conservative for the problems that we consider in the numerical experiments section. There we do not observe worse than linear growth of the error in time.

4. Numerical experiments

In this section we present numerical experiments to evaluate the performance of our scheme. In all cases we use a standard modal formulation and use the L^2 norm in space to evaluate the error in u. We also tabulate convergence of the linear part of the energy to verify that the theoretical convergence rates are attained. Note that we have not yet established error bounds in the L^2 norm beyond those which follow directly from convergence in the energy norm. As expected, the L^2 error experimentally is of higher order. We present the numerical experiments in both one and two dimensions. For two-dimensional problems, we consider a simple square domain and use the tensor-product Legendre polynomials to be the basis functions on a square reference element. All the numerical experiments are marched in time by the standard four-stage fourth-order explicit Runge-Kutta (RK4) scheme and the flux splitting parameter is chosen to be $\xi = c = 1$. In all experiments we choose the time step size sufficiently small to guarantee that the temporal error is dominated by the spatial error.

In considering the results note that optimal convergence orders are $q + 1$ for u and q for the energy. The theory only guarantees suboptimal convergence for the energy of order $q - 1/2$ for the energy-dissipating upwind fluxes and $q - 1$ for the energy-conservative fluxes. We will, nonetheless, experiment with $q = 1$ for both fluxes of both types.

4.1. Convergence in 1D

In this section we consider the sine-Gordon equation with a dissipating term, i.e., $\theta = 1$. Particularly, to investigate the order of convergence of our method, we solve the problem

$$\frac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} = c^2\frac{\partial^2 u}{\partial x^2} - \sin(u) + f(x,t), \quad x \in (-20, 20), \quad t \ge 0,$$

with a standing breather solution

$$\text{solution 1:} \quad u(x,t) = 4\arctan\frac{\sqrt{0.75}\cos(0.5t)}{0.5\cosh(\sqrt{0.75}x)}, \quad x \in (-20,20), \quad t \ge 0. \tag{42}$$

The initial conditions, Dirichlet boundary conditions and the external forcing $f(x,t)$ are chosen so that (42) is the exact solution.

As seen below, in our simulations we find the convergence rate for low degrees $q \le 3$ is not regular for some cases, so for comparison we also give the results for the manufactured solution

$$\text{solution 2:} \quad u(x,t) = e^{\sin(x-t)}, \quad x \in (-20,20), \quad t \ge 0. \tag{43}$$

The corresponding initial conditions, Dirichlet boundary conditions and external forcing are determined by the manufactured solution (43). For these two examples, we use the same space and time discretization, the only difference is the solution itself.

Table 1
L^2 errors in u for problem 1 (42) when the S.-flux is used. q is the degree of u^h, s is the degree of v^h and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	error \ N	400	800	1200	1600	2000	2400
1	0	$\|\|u - u^h\|\|_{L^2}$	8.80e-03(–)	6.70e-03(0.39)	2.95e-03(2.02)	2.23e-03(0.99)	1.78e-03(0.99)	1.49e-03(0.99)
	1		1.09e-02(–)	5.35e-03(1.03)	3.63e-03(0.96)	2.74e-03(0.97)	2.20e-03(0.98)	1.84e-03(0.98)
2	1	$\|\|u - u^h\|\|_{L^2}$	4.75e-05(–)	1.03e-03(-4.44)	1.86e-05(9.90)	2.55e-05(-1.09)	1.40e-06(13.03)	2.21e-06(-2.51)
	2		4.85e-05(–)	6.20e-05(-0.35)	1.85e-05(2.98)	2.40e-05(-0.89)	1.39e-06(12.76)	2.06e-05(-2.17)
q	s	error \ N	50	100	200	400	800	1600
3	2	$\|\|u - u^h\|\|_{L^2}$	1.91e-02(–)	4.23e-05(8.81)	2.49e-06(4.09)	1.70e-07(3.87)	8.77e-09(4.27)	5.41e-10(4.02)
	3		9.81e-03(–)	3.52e-05(8.12)	2.06e-06(4.09)	1.48e-07(3.80)	7.12e-09(4.38)	4.40e-10(4.02)
q	s	error \ N	80	100	120	140	160	180
4	3	$\|\|u - u^h\|\|_{L^2}$	5.14e-06(–)	1.66e-06(5.07)	6.58e-07(5.07)	3.02e-07(5.06)	1.54e-07(5.06)	8.47e-08(5.05)
	4		4.12e-06(–)	1.34e-06(5.05)	5.32e-07(5.05)	2.44e-07(5.05)	1.25e-07(5.04)	6.89e-08(5.04)
5	4	$\|\|u - u^h\|\|_{L^2}$	2.76e-07(–)	7.11e-08(6.08)	2.36e-08(6.04)	9.31e-09(6.04)	4.16e-09(6.04)	2.04e-09(6.03)
	5		2.22e-07(–)	5.72e-08(6.07)	1.91e-08(6.02)	7.54e-09(6.02)	3.37e-09(6.02)	1.66e-09(6.02)
6	5	$\|\|u - u^h\|\|_{L^2}$	1.45e-08(–)	3.14e-09(6.86)	8.71e-10(7.03)	2.95e-10(7.02)	1.16e-10(7.02)	5.06e-11(7.02)
	6		1.18e-08(–)	2.59e-09(6.80)	7.22e-10(7.01)	2.45e-10(7.01)	9.62e-11(7.01)	4.22e-11(7.01)

The discretization is performed on the computational domain $(-20, 20)$ with the element vertices $x_j = -20 + (j-1)h,\ j = 1, 2, \cdots, N+1,\ h = \frac{40}{N}$. We evolve the discretized problems until the final time $T = 2$ with the time step $\Delta t = 0.075h/(2\pi)$. We present the L^2 error for u. The degrees of the approximation space for u^h are set to be $q = (1, 2, 3, 4, 5, 6)$.

We test four different fluxes: the central flux denoted by C.-flux, the alternating flux with $\alpha = 0$ denoted by A.-flux, the Sommerfeld flux denoted by S.-flux, and the alternating flux with Sommerfeld flux with $\alpha = 0$, $\tau = \frac{\xi}{2}$, $\beta = \frac{1}{2\xi}$ denoted by A.S.-flux. Note that both the C.-flux and A.-flux are energy-conserving methods; both the S.-flux and A.S.-flux are energy-dissipating methods even when $\theta = 0$. We want to point out that $\alpha = 1$ has a similar performance to $\alpha = 0$ in the cases A.-flux and A.S.-flux; thus we only show the results for $\alpha = 0$ in the rest of the paper. We also consider two different approximation spaces: either u^h and v^h in the same space, i.e., $s = q$, or the degree of the approximation space of v^h one less than u^h, i.e., $s = q - 1$.

In experiments not shown, we observed that the convergence rate was somewhat irregular for all cases when L^2 projection was used to compute the initial conditions. One may use a special projection for the initial conditions to solve this problem; see for example the approach in [10] which discusses a projection for the local DG method with the alternating flux. But here, we adopt a simpler idea as in [4]: transform the problem into one with zero initial conditions,

$$u(x,t) = u_0(x) + \tilde{u}(x,t),$$

where $u_0(x) = u(x, 0)$. Then we get u by numerically solving for $\tilde{u}(x,t)$.

The L^2 error for u and the error in the linear part of the energy, that is the sum of the L^2 errors in ∇u and v, for both problems one and two are presented in Tables 1 through 16. A summary of the results is as follows; a clear takeaway is that the choice of $s = q - 1$ is more reliable for all but the central flux.

S.-flux: The L^2 error converges at the optimal rate, $q + 1$, for problem one with $q \geq 3$ and for problem two with $q \geq 2$. When $q = 1$ first order convergence is observed for $s = 0$, reduced somewhat when $s = 1$ for problem two. For problem one and $q = 2$ the observed convergence is nonmonotone, though the results are significantly more accurate than those obtained with $q = 1$. Convergence of the linear energy is optimal in all cases for problem one and when $s = q - 1$ for problem two, with some reduction of the convergence rate when $s = q$.

A.S.-flux: The observed L^2 convergence rates are similar to those observed for the S.-flux except there is some degradation when $s = q$. For the linear energy we observe optimal convergence in all cases for problem one and for problem two with $s = q - 1$. When $s = q$ there is a reduction in the rate to $q - 1/2$.

A.-flux: We observe optimal convergence in all cases when $s = q - 1$, except for problem one with $q = 2$ where, as for the two previous cases, convergence is nonmonotone, and problem two with $q = 1$, where the results seem more consistent with a rate of $1/2$. When $s = q$ and problem two we observe a reduction in the convergence rate. For the linear energy we observe optimal convergence only for problem one with $s = q - 1$; for problem two the rate appears to be reduced to $q - 1/2$. When $s = q$ the rate is only $q - 1$ which is still consistent with the theoretical results. Note that this means no convergence when $q = s = 1$. For problem two and $q = 5$ we observe an unexplained superconvergence of the energy errors.

C.-flux: Here the observed convergence of the L^2 error is optimal for odd values of q in all cases except problem two with $q = 1$. For even values of q it is typically suboptimal, and again nonmonotonic for problem one and $q = 2$. In contrast with the other three flux choices, we often see slightly better convergence rates when $s = q$. For both problems we observe optimal convergence of the error in the linear energy for q odd and suboptimal convergence for q even. Although the rates are typically the same, the errors are often smaller when $s = q - 1$.

Table 2
Linear energy errors for problem 1 (42) when the S.-flux is used. q is the degree of u^h, s is the degree of v^h, and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	7.85e-02(–)	3.95e-02(0.99)	2.66e-02(0.97)	2.00e-02(0.99)	1.60e-02(0.99)	1.34e-02(0.99)
	1		6.36e-01(–)	3.21e-02(0.99)	2.15e-02(0.99)	1.61e-02(0.99)	1.29e-02(0.99)	1.08e-02(1.00)
2	1	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.26e-03(–)	3.69e-04(1.77)	1.40e-04(2.38)	7.91e-05(2.00)	5.06e-05(2.00)	3.52e-05(2.00)
	2		1.14e-03(–)	2.85e-04(1.99)	1.27e-04(1.99)	1.17e-05(2.00)	4.60e-05(2.00)	3.19e-05(2.00)
q	s	N error	50	100	200	400	800	1600
3	2	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	9.67e-03(–)	1.17e-03(3.04)	1.49e-04(2.98)	1.87e-05(2.99)	2.34e-06(3.00)	2.93e-07(3.00)
	3		8.46e-03(–)	1.07e-03(2.98)	1.37e-04(2.97)	1.73e-05(2.99)	2.17e-06(2.99)	2.72e-07(3.00)
q	s	N error	80	100	120	140	160	180
4	3	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.63e-04(–)	6.74e-05(3.96)	3.26e-05(3.98)	1.77e-05(3.98)	1.04e-05(3.99)	6.48e-06(3.99)
	4		1.45e-04(–)	6.04e-05(3.93)	2.94e-05(3.96)	1.59e-05(3.96)	9.37e-06(3.97)	5.87e-06(3.98)
5	4	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.15e-05(–)	3.75e-06(5.01)	1.51e-06(4.98)	7.02e-07(4.98)	3.61e-07(4.99)	2.00e-07(4.99)
	5		1.02e-05(–)	3.33e-06(4.99)	1.35e-06(4.96)	6.28e-07(4.97)	3.23e-07(4.97)	1.80e-07(4.98)
6	5	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	7.53e-07(–)	2.07e-07(5.78)	6.98e-08(5.98)	2.78e-08(5.98)	1.25e-08(5.99)	6.16e-09(5.99)
	6		6.72e-07(–)	1.87e-07(5.74)	6.30e-08(5.96)	2.51e-08(5.97)	1.13e-08(5.98)	5.59e-09(5.98)

Table 3
L^2 errors in u for problem 2 (43) when the S.-flux is used. q is the degree of u^h, s is the degree of v^h and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\|\|u - u^h\|\|_{L^2}$	5.22e-01(–)	2.76e-01(0.92)	1.88e-01(0.95)	1.42e-01(0.98)	1.14e-01(0.98)	9.57e-02(0.98)
	1		6.65e-01(–)	4.18e-01(0.67)	3.23e-01(0.64)	2.71e-01(0.61)	2.38e-01(0.58)	2.14e-01(0.58)
2	1	$\|\|u - u^h\|\|_{L^2}$	8.25e-04(–)	1.04e-04(2.99)	3.09e-05(2.99)	1.30e-05(3.01)	6.68e-06(2.98)	3.87e-06(3.00)
	2		7.08e-04(–)	8.97e-05(2.98)	2.67e-05(2.99)	1.13e-05(2.99)	5.82e-06(2.97)	3.38e-06(2.98)
q	s	N error	50	100	200	400	800	1600
3	2	$\|\|u - u^h\|\|_{L^2}$	1.42e-02(–)	6.62e-04(4.42)	3.33e-05(4.31)	1.92e-06(4.12)	1.17e-07(4.03)	7.26e-09(4.01)
	3		1.09e-02(–)	4.81e-04(4.50)	2.24e-05(4.43)	1.21e-06(4.21)	7.19e-08(4.08)	4.41e-09(4.03)
q	s	N error	80	100	120	140	160	180
4	3	$\|\|u - u^h\|\|_{L^2}$	6.40e-05(–)	2.07e-05(5.06)	8.26e-06(5.03)	3.81e-06(5.01)	1.95e-06(5.00)	1.08e-06(5.00)
	4		4.35e-05(–)	1.39e-05(5.11)	5.50e-06(5.09)	2.52e-06(5.08)	1.28e-06(5.07)	7.05e-07(5.06)
5	4	$\|\|u - u^h\|\|_{L^2}$	3.41e-06(–)	8.90e-07(6.02)	2.97e-07(6.01)	1.18e-07(6.01)	5.28e-08(6.01)	2.60e-08(6.00)
	5		2.23e-06(–)	5.77e-07(6.06)	1.92e-07(6.04)	7.56e-08(6.03)	3.38e-08(6.02)	1.67e-08(6.02)
6	5	$\|\|u - u^h\|\|_{L^2}$	1.95e-07(–)	4.09e-08(6.99)	1.14e-08(7.00)	3.88e-09(7.00)	1.52e-09(7.01)	6.66e-10(7.01)
	6		1.31e-07(–)	2.76e-08(6.98)	7.71e-09(7.00)	2.62e-09(7.01)	1.03e-09(7.02)	4.49e-10(7.02)

Table 4
Linear energy errors for problem 2 (43) when the S.-flux is used. q is the degree of u^h, s is the degree of v^h, and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.23e+00(–)	6.59e-01(0.90)	4.50e-01(0.94)	3.42e-01(0.96)	2.76e-01(0.97)	2.31e-01(0.97)
	1		9.27e-01(–)	5.25e-01(0.82)	3.81e-01(0.79)	3.06e-01(0.76)	2.60e-01(0.73)	2.29e-01(0.71)
2	1	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.30e-02(–)	3.22e-03(2.01)	1.43e-03(2.00)	8.04e-04(2.00)	5.15e-04(2.00)	3.57e-04(2.00)
	2		2.49e-02(–)	8.15e-03(1.61)	4.31e-03(1.57)	2.76e-03(1.55)	1.96e-03(1.54)	1.48e-03(1.53)
q	s	N error	50	100	200	400	800	1600
3	2	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.06e-01(–)	1.33e-02(3.00)	1.68e-03(2.99)	2.11e-04(2.99)	2.64e-05(3.00)	3.30e-06(3.00)
	3		8.89e-02(–)	1.17e-02(2.93)	2.21e-03(2.40)	4.33e-04(2.36)	8.10e-05(2.42)	1.47e-05(2.46)
q	s	N error	80	100	120	140	160	180
4	3	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.99e-03(–)	8.20e-04(3.98)	3.96e-04(3.99)	2.14e-04(3.99)	1.26e-04(4.00)	7.84e-05(4.00)
	4		4.51e-03(–)	2.15e-03(3.32)	1.16e-03(3.39)	6.83e-04(3.43)	4.30e-04(3.46)	2.86e-04(3.48)
5	4	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.40e-04(–)	4.62e-05(4.98)	1.86e-05(4.98)	8.64e-06(4.99)	4.44e-06(4.99)	2.47e-06(4.99)
	5		2.11e-04(–)	6.26e-05(5.46)	2.27e-05(5.56)	9.55e-06(5.62)	4.51e-06(5.62)	2.34e-06(5.58)
6	5	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	9.90e-06(–)	2.63e-06(5.95)	8.84e-07(5.97)	3.52e-07(5.98)	1.58e-07(5.99)	7.80e-08(5.99)
	6		2.39e-05(–)	8.07e-06(4.87)	3.20e-06(5.07)	1.44e-06(5.19)	7.13e-07(5.26)	3.81e-07(5.31)

Table 5
L^2 errors in u for problem 1 (42) when the A.S.-flux is used. q is the degree of u^h, s is the degree of v^h and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\lVert u-u^h\rVert_{L^2}$	8.84e-03 (–)	4.32e-03(1.03)	2.92e-03(0.97)	2.20e-03(0.98)	1.77e-03(0.98)	1.48e-03(0.98)
	1		9.14e-02(–)	1.63e-02(2.49)	6.90e-03(2.12)	1.02e-02(-1.37)	4.34e-03(3.85)	3.64e-03(0.97)
2	1	$\lVert u-u^h\rVert_{L^2}$	4.32e-05(–)	3.64e-05(0.25)	6.45e-06(4.27)	1.07e-05(-1.75)	1.07e-06(10.33)	9.55e-07(0.60)
	2		4.69e-05(–)	5.12e-05(-0.13)	1.81e-05(2.56)	2.41e-05(-0.99)	1.39e-06(12.80)	2.03e-06(-2.09)
q	s	N error	50	100	200	400	800	1600
3	2	$\lVert u-u^h\rVert_{L^2}$	7.43e-02(–)	5.89e-05(10.30)	3.94e-06(3.90)	2.54e-07(3.95)	1.48e-08(4.10)	9.18e-10(4.01)
	3		1.34e-02(–)	4.22e-05(8.31)	2.54e-06(4.05)	2.16e-07(3.56)	7.42e-09(4.86)	4.48e-10(4.05)
q	s	N error	80	100	120	140	160	180
4	3	$\lVert u-u^h\rVert_{L^2}$	5.49e-06(–)	1.75e-06(5.13)	6.90e-07(5.10)	3.16e-07(5.07)	1.61e-07(5.05)	8.88e-08(5.04)
	4		4.26e-06(–)	1.36e-06(5.11)	5.38e-07(5.09)	2.46e-07(5.07)	1.25e-07(5.06)	6.91e-08(5.05)
5	4	$\lVert u-u^h\rVert_{L^2}$	3.70e-07(–)	1.01e-07(5.79)	3.51e-08(5.82)	1.42e-08(5.87)	6.46e-09(5.90)	3.21e-09(5.93)
	5		2.31e-07(–)	5.95e-08(6.07)	1.98e-08(6.04)	7.79e-09(6.04)	3.47e-09(6.04)	1.70e-09(6.04)
6	5	$\lVert u-u^h\rVert_{L^2}$	1.77e-08(–)	3.75e-09(6.95)	1.01e-09(7.20)	3.35e-10(7.15)	1.29e-10(7.12)	5.61e-11(7.09)
	6		1.22e-08(–)	2.63e-09(6.86)	7.29e-10(7.04)	2.47e-10(7.03)	9.67e-11(7.02)	4.23e-11(7.02)

Table 6
Linear energy errors for problem 1 (42) when the A.S.-flux is used. q is the degree of u^h, s is the degree of v^h, and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\left(\lVert\nabla e_u\rVert_{L^2}^2+\lVert e_v\rVert_{L^2}^2\right)^{1/2}$	8.69e-02(–)	4.42e-02(0.98)	2.96e-02(0.99)	2.23e-02(0.99)	1.78e-02(0.99)	1.49e-02(0.99)
	1		9.21e-02(–)	4.42e-02(1.06)	2.95e-02(0.99)	2.22e-02(0.99)	1.79e-02(0.96)	1.50e-02(0.98)
2	1	$\left(\lVert\nabla e_u\rVert_{L^2}^2+\lVert e_v\rVert_{L^2}^2\right)^{1/2}$	1.13e-03(–)	2.83e-04(2.00)	1.26e-04(2.00)	7.07e-05(2.00)	4.53e-05(1.99)	3.15e-05(2.00)
	2		1.24e-03(–)	3.13e-04(1.99)	1.40e-04(1.99)	7.88e-05(1.99)	5.05e-05(1.99)	3.51e-05(2.00)
q	s	N error	50	100	200	400	800	1600
3	2	$\left(\lVert\nabla e_u\rVert_{L^2}^2+\lVert e_v\rVert_{L^2}^2\right)^{1/2}$	1.56e-02(–)	1.30e-03(3.58)	1.76e-04(2.89)	2.26e-05(2.96)	2.85e-06(2.99)	3.58e-07(3.00)
	3		1.04e-02(–)	1.54e-03(2.75)	2.15e-04(2.85)	2.82e-05(2.93)	3.59e-06(2.97)	4.54e-07(2.99)
q	s	N error	80	100	120	140	160	180
4	3	$\left(\lVert\nabla e_u\rVert_{L^2}^2+\lVert e_v\rVert_{L^2}^2\right)^{1/2}$	1.56e-04(–)	6.21e-05(4.14)	2.93e-05(4.12)	1.56e-05(4.09)	9.08e-06(4.07)	5.63e-06(4.05)
	4		1.91e-04(–)	7.72e-05(4.05)	3.69e-05(4.06)	1.97e-05(4.05)	1.15e-05(4.04)	7.17e-06(4.03)
5	4	$\left(\lVert\nabla e_u\rVert_{L^2}^2+\lVert e_v\rVert_{L^2}^2\right)^{1/2}$	1.17e-05(–)	4.04e-06(4.77)	1.69e-06(4.78)	8.05e-07(4.82)	4.21e-07(4.85)	2.37e-07(4.87)
	5		1.49e-05(–)	5.18e-06(4.72)	2.18e-06(4.75)	1.04e-06(4.81)	5.44e-07(4.84)	3.06e-07(4.87)
6	5	$\left(\lVert\nabla e_u\rVert_{L^2}^2+\lVert e_v\rVert_{L^2}^2\right)^{1/2}$	7.80e-07(–)	2.04e-07(6.01)	6.57e-08(6.22)	2.54e-08(6.17)	1.12e-08(6.13)	5.46e-09(6.10)
	6		9.98e-07(–)	2.70e-07(5.86)	8.77e-08(6.16)	3.40e-08(6.14)	1.50e-08(6.13)	7.31e-09(6.11)

Table 7
L^2 errors in u for problem 2 (43) when the A.S.-flux is used. q is the degree of u^h, s is the degree of v^h and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\lVert u-u^h\rVert_{L^2}$	5.50e-01(–)	2.91e-01(0.92)	1.97e-01(0.96)	1.48e-01(0.99)	1.18e-01(1.02)	9.88e-02(1.00)
	1		9.92e-01(–)	5.85e-01(0.76)	4.29e-01(0.76)	3.47e-01(0.74)	2.95e-01(0.73)	2.59e-01(0.71)
2	1	$\lVert u-u^h\rVert_{L^2}$	6.88e-04(–)	8.64e-05(2.99)	2.56e-05(3.00)	1.08e-05(3.00)	5.55e-06(2.98)	3.21e-06(3.00)
	2		7.07e-04(–)	8.95e-05(2.98)	2.67e-05(2.98)	1.13e-05(2.99)	5.82e-06(2.97)	3.38e-06(2.98)
q	s	N error	50	100	200	400	800	1600
3	2	$\lVert u-u^h\rVert_{L^2}$	1.37e-02(–)	7.26e-04(4.24)	3.58e-05(4.34)	1.96e-06(4.19)	1.18e-07(4.06)	7.31e-09(4.01)
	3		1.37e-02(–)	7.12e-04(4.26)	3.03e-05(4.55)	1.40e-06(4.44)	7.54e-08(4.21)	4.47e-09(4.07)
q	s	N error	80	100	120	140	160	180
4	3	$\lVert u-u^h\rVert_{L^2}$	5.66e-05(–)	1.73e-05(5.32)	6.70e-06(5.19)	3.04e-06(5.14)	1.54e-06(5.11)	8.43e-07(5.09)
	4		4.58e-05(–)	1.43e-05(5.23)	5.58e-06(5.15)	2.54e-06(5.11)	1.29e-06(5.09)	7.08e-07(5.08)
5	4	$\lVert u-u^h\rVert_{L^2}$	2.95e-06(–)	8.09e-07(5.80)	2.80e-07(5.83)	1.13e-07(5.86)	5.16e-08(5.89)	2.57e-08(5.91)
	5		2.30e-06(–)	5.95e-07(6.06)	1.97e-07(6.05)	7.76e-08(6.05)	3.46e-08(6.04)	1.70e-08(6.04)
6	5	$\lVert u-u^h\rVert_{L^2}$	1.65e-07(–)	3.35e-08(7.13)	9.16e-09(7.11)	3.07e-09(7.09)	1.20e-09(7.07)	5.20e-10(7.06)
	6		1.36e-07(–)	2.82e-08(7.06)	7.81e-09(7.05)	2.64e-09(7.04)	1.03e-09(7.04)	4.50e-10(7.03)

Table 8
Linear energy errors for problem 2 (43) when the A.S.-flux is used. q is the degree of u^h, s is the degree of v^h, and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.36e+00(−)	7.23e-01(0.92)	4.92e-01(0.95)	3.73e-01(0.96)	3.00e-01(0.97)	2.51e-01(0.98)
	1		1.58e+00(−)	8.78e-01(0.84)	6.18e-01(0.87)	4.81e-01(0.87)	3.97e-01(0.86)	3.40e-01(0.85)
2	1	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.06e-02(−)	2.60e-03(2.03)	1.15e-03(2.01)	6.45e-04(2.01)	4.12e-04(2.01)	2.86e-04(2.01)
	2		2.59e-02(−)	8.35e-03(1.63)	4.39e-03(1.59)	2.80e-03(1.56)	1.98e-03(1.55)	1.49e-03(1.54)
q	s	N error	50	100	200	400	800	1600
3	2	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.08e-01(−)	1.54e-02(2.80)	2.08e-03(2.89)	2.68e-04(2.96)	3.37e-05(2.99)	4.22e-06(3.00)
	3		1.27e-01(−)	1.89e-02(2.75)	3.01e-03(2.66)	5.06e-04(2.57)	8.75e-05(2.53)	1.53e-05(2.51)
q	s	N error	80	100	120	140	160	180
4	3	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	2.07e-03(−)	8.18e-04(4.17)	3.84e-04(4.15)	2.03e-04(4.13)	1.17e-04(4.11)	7.24e-05(4.10)
	4		4.70e-03(−)	2.21e-03(3.38)	1.18e-03(3.43)	6.94e-04(3.46)	4.36e-04(3.48)	2.89e-04(3.49)
5	4	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.48e-04(−)	5.02e-05(4.85)	2.07e-05(4.87)	9.75e-06(4.88)	5.07e-06(4.89)	2.85e-06(4.91)
	5		2.37e-04(−)	7.34e-05(5.25)	2.80e-05(5.29)	1.24e-05(5.29)	6.12e-06(5.27)	3.31e-06(5.22)
6	5	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.08e-05(−)	2.72e-06(6.16)	8.85e-07(6.16)	3.44e-07(6.14)	1.52e-07(6.12)	7.40e-08(6.11)
	6		2.51e-05(−)	8.28e-06(4.97)	3.26e-06(5.12)	1.46e-06(5.22)	7.19e-07(5.28)	3.84e-07(5.33)

Table 9
L^2 errors in u for problem 1 (42) when the A.-flux is used. q is the degree of u^h, s is the degree of v^h, and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\|\|u - u^h\|\|_{L^2}$	1.78e-03(−)	4.87e-04(1.87)	1.98e-04(2.22)	1.12e-04(1.99)	7.12e-05(2.02)	4.92e-05(2.03)
	1		6.74e-01(−)	6.73e-01(0.00)	7.03e-01(-0.11)	6.96e-01(0.04)	6.91e-01(0.03)	6.87e-01(0.03)
2	1	$\|\|u - u^h\|\|_{L^2}$	3.62e-04(−)	2.16e-05(4.06)	6.18e-06(3.09)	8.19e-06(-0.98)	1.10e-06(8.99)	6.80e-07(2.65)
	2		3.23e-02(−)	2.58e-03(3.64)	2.86e-04(5.43)	2.98e-03(-8.14)	5.69e-04(7.41)	3.13e-04(3.27)
q	s	N error	50	100	200	400	800	1600
3	2	$\|\|u - u^h\|\|_{L^2}$	3.99e-03(−)	5.67e-05(6.14)	3.60e-06(3.98)	2.21e-07(4.03)	1.37e-08(4.00)	1.34e-09(3.36)
	3		1.95e-03(−)	9.72e-05(4.33)	9.94e-06(3.29)	2.02e-06(2.30)	4.31e-08(5.55)	3.18e-09(3.76)
q	s	N error	80	100	120	140	160	180
4	3	$\|\|u - u^h\|\|_{L^2}$	7.12e-06(−)	2.35e-06(4.96)	9.50e-07(4.98)	4.40e-07(4.99)	2.26e-07(4.99)	1.26e-07(4.99)
	4		8.64e-06(−)	2.88e-06(4.93)	1.16e-06(4.97)	5.39e-07(4.98)	2.77e-07(4.98)	1.54e-07(4.99)
5	4	$\|\|u - u^h\|\|_{L^2}$	3.88e-07(−)	1.01e-07(6.02)	3.41e-08(5.97)	1.36e-08(5.98)	6.10e-09(5.99)	3.01e-09(5.99)
	5		4.41e-07(−)	1.15e-07(6.02)	3.88e-08(5.97)	1.54e-08(5.98)	6.94e-09(5.98)	3.43e-09(5.99)
6	5	$\|\|u - u^h\|\|_{L^2}$	2.06e-08(−)	4.57e-09(6.74)	1.28e-09(6.97)	4.37e-10(6.98)	1.72e-10(6.97)	7.63e-11(6.92)
	6		2.25e-08(−)	5.02e-09(6.73)	1.41e-09(6.97)	4.80e-10(6.97)	1.90e-10(6.96)	8.41e-11(6.90)

Table 10
Linear energy errors for problem 1 (42) when the A.-flux is used. q is the degree of u^h, s is the degree of v^h, and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	7.49e-02(−)	3.74e-02(1.00)	2.50e-02(1.00)	1.87e-02(1.00)	1.50e-02(1.00)	1.25e-02(1.00)
	1		4.85e-01(−)	4.87e-01(-0.01)	4.86e-01(0.01)	4.85e-01(0.00)	4.85e-01(0.00)	4.85e-01(0.00)
2	1	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.27e-03(−)	3.17e-04(2.00)	1.41e-04(2.00)	7.92e-05(2.00)	5.08e-05(2.00)	3.52e-05(2.00)
	2		1.58e-02(−)	7.26e-03(1.13)	4.84e-03(1.00)	3.64e-03(1.00)	2.90e-03(1.01)	2.42e-03(1.00)
q	s	N error	50	100	200	400	800	1600
3	2	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	9.55e-03(−)	1.19e-03(3.01)	1.50e-04(2.99)	1.87e-05(3.00)	2.34e-06(3.00)	2.93e-07(3.00)
	3		2.66e-02(−)	6.45e-03(2.04)	1.61e-03(2.00)	4.04e-04(2.00)	1.01e-04(2.00)	2.52e-05(2.00)
q	s	N error	80	100	120	140	160	180
4	3	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.64e-04(−)	6.77e-05(3.96)	3.28e-05(3.98)	1.77e-05(3.99)	1.04e-05(3.99)	6.50e-06(3.99)
	4		1.03e-03(−)	5.32e-04(2.97)	3.09e-04(2.98)	1.95e-04(2.99)	1.31e-04(2.99)	9.18e-05(2.99)
5	4	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	1.15e-05(−)	3.75e-06(5.01)	1.51e-06(4.98)	7.03e-07(4.98)	3.61e-07(4.99)	2.01e-07(4.99)
	5		9.35e-05(−)	3.84e-05(3.99)	1.86e-05(3.98)	1.01e-05(3.98)	5.92e-06(3.99)	3.70e-06(3.99)
6	5	$\left(\|\nabla e_u\|_{L^2}^2 + \|e_v\|_{L^2}^2\right)^{1/2}$	7.52e-07(−)	2.07e-07(5.78)	6.97e-08(5.98)	2.77e-08(5.98)	1.25e-08(5.98)	6.16e-09(5.99)
	6		7.89e-06(−)	2.67e-06(4.85)	1.08e-06(4.95)	5.04e-07(4.97)	2.59e-07(4.98)	1.44e-07(4.98)

Table 11
L^2 errors in u for problem 2 (43) when the A.-flux is used. q is the degree of u^h, s is the degree of v^h and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\|\|u-u^h\|\|_{L^2}$	5.84e-01(–)	4.25e-01(0.46)	3.50e-01(0.48)	3.05e-01(0.48)	2.73e-01(0.50)	2.50e-01(0.49)
	1		4.31e+00(–)	4.34e+00(-0.01)	4.35e+00(-0.01)	4.36e+00(-0.01)	4.36e+00(0.00)	4.37e+00(-0.00)
2	1	$\|\|u-u^h\|\|_{L^2}$	7.73e-04(–)	9.79e-05(2.98)	2.92e-05(2.98)	1.24e-05(2.98)	6.36e-06(2.99)	3.69e-06(2.98)
	2		1.17e-02(–)	2.92e-03(2.00)	1.30e-03(2.00)	7.30e-04(2.01)	4.67e-04(2.00)	3.24e-04(2.00)
q	s	N error	50	100	200	400	800	1600
3	2	$\|\|u-u^h\|\|_{L^2}$	1.46e-02(–)	6.81e-04(4.42)	4.20e-05(4.02)	2.29e-06(4.20)	1.43e-07(4.00)	9.39e-09(3.93)
	3		3.71e-02(–)	2.74e-03(3.76)	1.78e-04(3.94)	1.12e-05(3.99)	7.03e-07(4.00)	4.39e-08(4.00)
q	s	N error	80	100	120	140	160	180
4	3	$\|\|u-u^h\|\|_{L^2}$	8.23e-05(–)	2.59e-05(5.18)	1.06e-05(4.92)	4.96e-06(4.90)	2.41e-06(5.41)	1.32e-06(5.08)
	4		9.51e-05(–)	3.05e-05(5.10)	1.21e-05(5.07)	5.55e-06(5.06)	2.83e-06 (5.04)	1.57e-06(5.03)
5	4	$\|\|u-u^h\|\|_{L^2}$	3.74e-06(–)	1.01e-06(5.88)	3.47e-07(5.84)	1.41e-07(5.86)	6.39e-08(5.91)	3.14e-08(6.03)
	5		4.19e-06(–)	1.10e-06(5.98)	3.69e-07(6.00)	1.47e-07(5.99)	6.57e-08(6.01)	3.24e-08(5.99)
6	5	$\|\|u-u^h\|\|_{L^2}$	2.18e-07(–)	4.60e-08(6.98)	1.29e-08(6.99)	4.35e-09(7.03)	1.74e-09(6.88)	7.89e-10(6.70)
	6		2.31e-07(–)	4.91e-08(6.93)	1.38e-08(6.97)	4.71e-09(6.97)	1.86e-09(6.97)	8.12e-10(7.03)

Table 12
Linear energy errors for problem 2 (43) when the A.-flux is used. q is the degree of u^h, s is the degree of v^h, and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	7.75e-01(–)	4.49e-01(0.79)	3.34e-01(0.73)	2.75e-01(0.69)	2.37e-01(0.66)	2.11e-01(0.64)
	1		1.05e+01(–)	1.05e+01(0.00)	1.05e+01(0.00)	1.05e+01(0.00)	1.05e+01(0.00)	1.05e+01(0.00)
2	1	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	3.14e-02(–)	1.03e-02(1.61)	5.40e-03(1.58)	3.44e-03(1.57)	2.43e-03(1.56)	1.83e-03(1.55)
	2		3.34e-01(–)	1.67e-01(1.00)	1.11e-01(1.00)	8.35e-02(1.00)	6.68e-02(1.00)	5.56e-02(1.00)
q	s	N error	50	100	200	400	800	1600
3	2	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	1.24e-01(–)	1.41e-02(3.13)	2.66e-03(2.41)	5.34e-04(2.31)	1.02e-04(2.39)	1.88e-05(2.44)
	3		5.24e-01(–)	1.35e-01(1.96)	3.40e-02(1.99)	8.52e-03(2.00)	2.13e-03(2.00)	5.32e-04(2.00)
q	s	N error	80	100	120	140	160	180
4	3	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	5.75e-03(–)	2.78e-03(3.26)	1.51e-03(3.35)	8.97e-04(3.38)	5.65e-04(3.46)	3.76e-04(3.46)
	4		2.06e-02(–)	1.06e-02(2.98)	6.15e-03(2.99)	3.87e-03(3.00)	2.60e-03(3.00)	1.82e-03(3.00)
5	4	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	3.04e-04(–)	9.07e-05(5.42)	3.29e-05(5.56)	1.38e-05(5.65)	6.43e-06(5.71)	3.27e-06(5.75)
	5		1.65e-03(–)	6.78e-04(3.98)	3.28e-04(3.98)	1.77e-04(3.99)	1.04e-04(3.99)	6.50e-05(3.99)
6	5	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	2.98e-05(–)	1.04e-05(4.72)	4.19e-06(4.98)	1.90e-06(5.12)	9.46e-07(5.23)	5.09e-07(5.27)
	6		1.25e-04(–)	4.14e-05(4.95)	1.67e-05(4.97)	7.76e-06(4.98)	3.98e-06(4.99)	2.21e-06(5.00)

Table 13
L^2 errors in u for problem 1 (42) when the C.-flux is used. q is the degree of u^h, s is the degree of v^h and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\|\|u-u^h\|\|_{L^2}$	2.32e-03(–)	5.80e-04(2.00)	2.58e-04(2.00)	1.45e-04(2.00)	9.28e-05(2.00)	6.44e-05(2.00)
	1		3.29e-03(–)	1.17e-03(1.50)	2.98e-04(3.36)	7.82e-04(-3.35)	2.52e-04(5.08)	6.50e-04(-5.21)
2	1	$\|\|u-u^h\|\|_{L^2}$	4.99e-03(–)	1.06e-03(2.23)	7.32e-05(6.60)	4.53e-05(1.67)	1.52e-05(4.89)	8.98e-06(2.89)
	2		2.85e-01(–)	4.11e-04(9.44)	3.85e-04(0.16)	3.17e-03(-7.32)	1.73e-03(2.72)	2.74e-04(10.10)
q	s	N error	50	100	200	400	800	1600
3	2	$\|\|u-u^h\|\|_{L^2}$	2.55e-03(–)	3.79e-05(6.07)	1.90e-06(4.32)	1.13e-07(4.07)	6.98e-09(4.02)	4.35e-10(4.00)
	3		2.07e-03(–)	5.44e-05(5.25)	2.20e-06(4.63)	1.19e-07(4.21)	7.04e-09(4.08)	4.36e-10(4.01)
q	s	N error	80	100	120	140	160	180
4	3	$\|\|u-u^h\|\|_{L^2}$	9.54e-06(–)	3.95e-06(3.95)	1.91e-06(3.99)	1.03e-06(3.98)	6.06e-07(3.99)	3.79e-07(3.99)
	4		9.93e-06(–)	3.58e-06(4.57)	1.51e-06(4.72)	7.23e-07(4.80)	3.79e-07(4.84)	2.13e-07(4.88)
5	4	$\|\|u-u^h\|\|_{L^2}$	5.82e-07(–)	9.49e-08(8.13)	2.53e-08(7.25)	9.13e-09(6.62)	3.88e-09(6.41)	1.84e-09(6.31)
	5		4.30e-07(–)	9.63e-08(6.70)	2.90e-08(6.58)	1.06e-08(6.54)	4.45e-09(6.49)	2.09e-09(6.43)
6	5	$\|\|u-u^h\|\|_{L^2}$	2.13e-08(–)	5.72e-09(5.89)	1.92e-09(5.98)	7.65e-10(5.98)	3.44e-10(5.98)	1.70e-10(5.99)
	6		2.17e-08(–)	5.31e-09(6.30)	1.61e-09(6.55)	5.76e-10(6.66)	2.34e-10(6.74)	1.05e-10(6.80)

Table 14
Linear energy errors for problem 1 (42) when the C.-flux is used. q is the degree of u^h, s is the degree of v^h, and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	6.37e-02(–)	3.18e-02(1.00)	2.12e-02(1.00)	1.59e-02(1.00)	1.27e-02(1.00)	1.06e-02(1.00)
	1		6.57e-02(–)	3.29e-02(1.00)	2.20e-02(1.00)	1.65e-02(1.00)	1.32e-02(1.00)	1.10e-02(1.00)
2	1	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	1.20e-02(–)	5.98e-03(1.01)	3.98e-03(1.00)	2.99e-03(1.00)	2.39e-03(1.00)	1.99e-03(1.00)
	2		9.23e-02(–)	9.18e-03(3.33)	6.12e-03(1.00)	4.60e-03(1.00)	3.67e-03(1.00)	3.06e-03(1.00)
q	s	N error	50	100	200	400	800	1600
3	2	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	1.30e-02(–)	9.40e-04(3.79)	1.08e-04(3.12)	1.33e-05(3.02)	1.66e-06(3.01)	2.07e-07(3.00)
	3		2.72e-02(–)	3.94e-03(2.78)	5.31e-04(2.89)	6.78e-05(2.97)	8.52e-06(2.99)	1.07e-06(3.00)
q	s	N error	80	100	120	140	160	180
4	3	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	2.07e-04(–)	1.04e-04(3.07)	5.97e-05(3.06)	3.74e-05(3.04)	2.49e-05(3.03)	1.75e-05(3.02)
	4		1.17e-03(–)	6.40e-04(2.70)	3.84e-04(2.80)	2.47e-04(2.86)	1.68e-04 (2.90)	1.19e-04(2.92)
5	4	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	1.72e-05(–)	3.54e-06(7.08)	1.19e-06(5.99)	5.14e-07(5.44)	2.54e-07(5.28)	1.38e-07(5.20)
	5		8.97e-05(–)	3.15e-05(4.69)	1.33e-05(4.75)	6.32e-06(4.80)	3.31e-06(4.85)	1.86e-06(4.88)
6	5	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	7.75e-07(–)	2.51e-07(5.04)	9.96e-08(5.08)	4.57e-08(5.06)	2.33e-08(5.04)	1.29e-08(5.03)
	6		7.09e-06(–)	2.79e-06(4.18)	1.23e-06(4.48)	6.01e-07(4.65)	3.19e-07(4.74)	1.81e-07(4.80)

Table 15
L^2 errors in u for problem 2 (43) when the C.-flux is used. q is the degree of u^h, s is the degree of v^h and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\|\|u-u^h\|\|_{L^2}$	1.01e-01(–)	4.67e-02(1.11)	3.07e-02(1.03)	2.29e-02(1.02)	1.83e-02(1.00)	1.52e-02(1.01)
	1		4.44e-01(–)	3.02e-01(0.56)	2.45e-01(0.52)	2.12e-01(0.50)	1.90e-01(0.49)	1.73e-01(0.50)
2	1	$\|\|u-u^h\|\|_{L^2}$	4.57e-03(–)	1.14e-03(2.00)	5.08e-04(1.99)	2.86e-04(2.00)	1.83e-04(2.00)	1.27e-04(2.00)
	2		1.85e-02(–)	4.66e-03(1.99)	2.07e-03(2.00)	1.17e-03(1.98)	7.46e-04(2.02)	5.18e-04(2.00)
q	s	N error	50	100	200	400	800	1600
3	2	$\|\|u-u^h\|\|_{L^2}$	4.08e-02(–)	5.12e-04(6.31)	2.38e-05(4.43)	1.42e-06(4.06)	8.84e-08(4.01)	5.51e-09(4.00)
	3		3.56e-02(–)	1.06e-03(5.06)	2.78e-05(5.26)	1.22e-06(4.52)	7.13e-08(4.09)	4.39e-09(4.02)
q	s	N error	80	100	120	140	160	180
4	3	$\|\|u-u^h\|\|_{L^2}$	1.19e-04(–)	5.01e-05(3.87)	2.43e-05(3.96)	1.23e-05(4.44)	7.07e-06(4.12)	4.50e-06(3.84)
	4		1.11e-04(–)	3.85e-05(4.75)	1.59e-05(4.84)	7.50e-06(4.89)	3.88e-06(4.93)	2.17e-06(4.95)
5	4	$\|\|u-u^h\|\|_{L^2}$	7.02e-06(–)	9.44e-07(8.99)	2.69e-07(6.88)	1.00e-07(6.41)	4.27e-08(6.39)	2.02e-08(6.35)
	5		4.14e-06(–)	9.47e-07(6.61)	2.84e-07(6.60)	1.04e-07(6.52)	4.39e-08(6.46)	2.06e-08(6.41)
6	5	$\|\|u-u^h\|\|_{L^2}$	2.24e-07(–)	5.73e-08(6.11)	1.91e-08(6.02)	7.77e-09(5.85)	3.47e-09(6.04)	1.70e-09(6.02)
	6		2.09e-07(–)	5.08e-08(6.34)	1.54e-08(6.54)	5.53e-09(6.66)	2.25e-09(6.73)	1.01e-09(6.79)

Table 16
Linear energy errors for problem 2 (43) when the C.-flux is used. q is the degree of u^h, s is the degree of v^h, and N is the number of the cells with uniform mesh size $h = 40/N$.

q	s	N error	400	800	1200	1600	2000	2400
1	0	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	3.27e-01(–)	1.59e-01(1.04)	1.05e-01(1.01)	7.87e-02(1.01)	6.29e-02(1.00)	5.24e-02(1.00)
	1		9.77e-01(–)	5.55e-01(0.82)	4.10e-01(0.75)	3.34e-01(0.71)	2.87e-01(0.68)	2.54e-01(0.66)
2	1	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	1.35e-01(–)	6.71e-02(1.00)	4.47e-02(1.00)	3.35e-02(1.00)	2.68e-02(1.00)	2.24e-02(1.00)
	2		4.22e-01(–)	2.11e-01(1.00)	1.41e-01(1.00)	1.06e-01(1.00)	8.44e-02(1.00)	7.04e-02(1.00)
q	s	N error	50	100	200	400	800	1600
3	2	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	3.07e-01(–)	1.05e-02(4.87)	1.18e-03(3.16)	1.45e-04(3.02)	1.80e-05(3.01)	2.24e-06(3.00)
	3		4.95e-01(–)	7.66e-02(2.69)	1.04e-02(2.88)	1.37e-03(2.93)	1.82e-04(2.91)	2.53e-05(2.85)
q	s	N error	80	100	120	140	160	180
4	3	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	2.88e-03(–)	1.43e-03(3.15)	8.39e-04(2.91)	5.02e-04(3.33)	3.36e-04(3.00)	2.32e-04(3.15)
	4		2.38e-02(–)	1.29e-02(2.75)	7.66e-03(2.84)	4.91e-03(2.89)	3.32e-03 (2.92)	2.35e-03(2.94)
5	4	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	2.24e-04(–)	4.63e-05(7.06)	1.63e-05(5.74)	7.28e-06(5.22)	3.63e-06(5.21)	1.97e-06(5.21)
	5		1.42e-03(–)	4.94e-04(4.73)	2.06e-04(4.80)	9.74e-05(4.85)	5.07e-05(4.89)	2.85e-05(4.91)
6	5	$\left(\|\nabla e_u\|_{L^2}^2+\|e_v\|_{L^2}^2\right)^{1/2}$	1.02e-05(–)	3.21e-06(5.19)	1.24e-06(5.21)	5.55e-07(5.22)	2.89e-07(4.89)	1.57e-07(5.18)
	6		1.26e-04(–)	4.63e-05(4.47)	1.98e-05(4.66)	9.51e-06(4.76)	4.99e-06(4.82)	2.82e-06(4.86)

4.2. Soliton solutions of the sine-Gordon equation in 1D

In this section, we consider the sine-Gordon equation without the dissipating term, i.e., $\theta = 0$,

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} - \sin(u), \quad x \in (-20, 20), \quad t \geq 0. \tag{44}$$

This equation appears in a number of physical applications and is famous for its soliton and multi-soliton solutions. Here, we focus on investigating these soliton solutions: breather soliton, kink soliton, anti-kink soliton and multi-soliton solutions: kink-kink collision, kink-antikink collision. In the numerical simulations, the number of elements is chosen to be $N = 120$. We impose no-flux conditions at the computational domain boundaries,

$$\frac{\partial u}{\partial x}(-20, t) = \frac{\partial u}{\partial x}(20, t) = 0, \quad t \geq 0.$$

4.2.1. Standing breather soliton

To numerically simulate the breather soliton solution of the sine-Gordon equation (44), we consider the initial conditions,

$$u(x, 0) = 4 \arctan \frac{\sqrt{0.75}}{0.5 \cosh(\sqrt{0.75}x)}, \quad \frac{\partial u}{\partial t}(x, 0) = 0, \quad x \in (-20, 20).$$

These conditions correspond to an exact standing breather soliton solution

$$u(x, t) = 4 \arctan \frac{\sqrt{0.75} \cos(0.5t)}{0.5 \cosh(\sqrt{0.75}x)}.$$

Time history of the numerical energy: we first study the numerical energy of the DG approximations to the standing breather solution. As above, we consider the four different fluxes: A.-flux, C.-flux, A.S.-flux and S.-flux; we also consider the cases where both u^h, v^h are in the same approximation space ($s = q$) and when the degree of the approximation space for v^h is one less than u^h ($s = q - 1$). The degree of the approximation space for u is fixed to be $q = 4$. We evolve the numerical solution until $T = 120$ with $h = 1/3$ and use the 4-stage Runge Kutta method with $\Delta t = 0.195h/(2\pi)$.

In Fig. 1, we present the numerical energy for the schemes with S.-flux, A.S.-flux, A.-flux and C.-flux. On the top panel, from the left to the right are the cases where u^h, v^h are in different approximation spaces, $s = q - 1$, and the same approximation space, $s = q$, respectively. Overall, we observe that the change of the numerical energy is not significant compared with the initial energy even for dissipating schemes. The A.S.-flux and S.-flux produce energy dissipating schemes and they have somewhat different performance depending on s, but even then the energy is conserved to around five digits. On the bottom panel, from the left to the right shows a close-up of the energy for the conservative schemes (A.-flux and C.-flux) with u^h, v^h in different spaces ($s = q - 1$) and in the same space ($s = q$), respectively. We notice that the numerical energy is conserved to around eight digits when $s = q - 1$ and seven digits when $s = q$.

The numerical standing breather soliton: the numerical standing breather solutions are shown in Fig. 2 and Fig. 3. In the simulation, u^h, v^h are chosen to be in the same approximation space with $q = s = 4$ and the S.-flux is used. Fig. 2 shows both exact and numerical breather solutions at several times, $t = 0, 45, 90, 120$ respectively. Fig. 3 presents the space-time plot of the breather solution from $t = 0$ to $t = 120$. We find that the numerical results match well with the analytic solution.

Time history of the L^2 error: the time history of the L^2 errors for the standing breather soliton solution with A.-flux, C.-flux, A.S.-flux and S.-flux are plotted in Fig. 4 for both $s = q$ and $s = q - 1$. Particularly, q is set to be 4 in the numerical simulation. The top panel is for energy-conserving schemes with the A.-flux and C.-flux from the left to right. The bottom panel is for energy-dissipating schemes with the A.S.-flux and S.-flux from the left to right. The error dynamics for all schemes except for the C.-flux are quite similar to each other and for the two values of s tested. For the C-flux., however, the errors display noticeably different patterns. Nonetheless, the peak errors for all eight experiments are comparable. Finally, considering that the standing breather solution is periodic in time, we note that the L^2 error grows linearly in time.

4.2.2. Kink soliton and antikink soliton

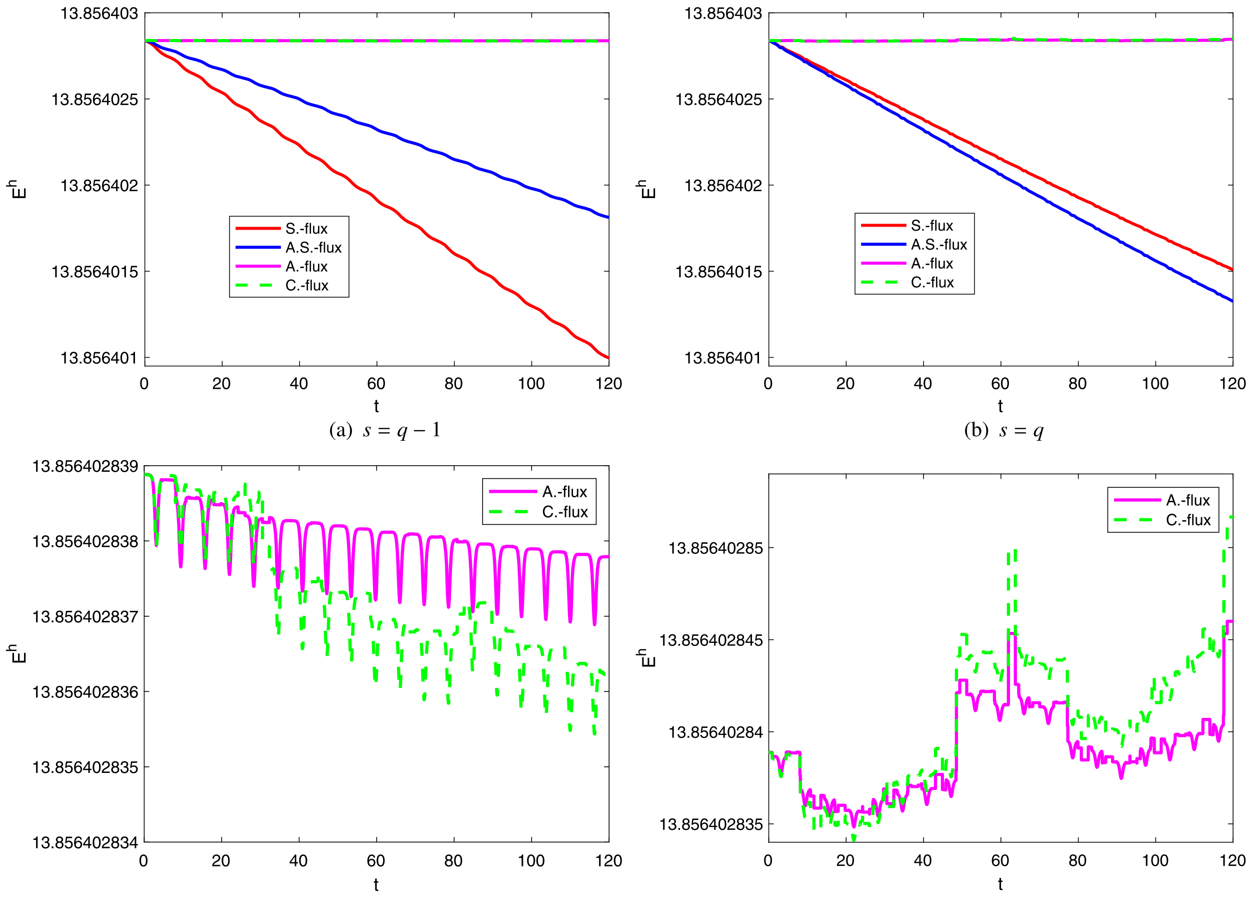
For the kink soliton solution, the sine-Gordon equation (44) is solved with the initial condition,

$$u(x, 0) = 4 \arctan \left(\exp \left(\frac{x}{\sqrt{1 - \mu^2}} \right) \right), \quad \frac{\partial u}{\partial t}(x, 0) = -\frac{2\mu}{\sqrt{1 - \mu^2}} \operatorname{sech} \left(\frac{x}{\sqrt{1 - \mu^2}} \right), \quad x \in (-20, 20).$$

The analytic kink solution

$$u(x, t) = 4 \arctan \left(\exp \left(\frac{x - \mu t}{\sqrt{1 - \mu^2}} \right) \right) \tag{45}$$

is a traveling wave increasing monotonically from 0 to 2π as x varies from $-\infty$ to ∞. In contrast with the kink soliton (45), for the antikink soliton solution we solve the sine-Gordon equation (44) with the initial conditions,



(a) $s = q - 1$

(b) $s = q$

(c) $s = q - 1$

(d) $s = q$

Fig. 1. Plots of the history of the numerical energy for the standing breather solution.

$$u(x,0) = 4\arctan\left(\exp\left(-\frac{x}{\sqrt{1-\mu^2}}\right)\right), \quad \frac{\partial u}{\partial t}(x,0) = \frac{2\mu}{\sqrt{1-\mu^2}}\text{sech}\left(\frac{x}{\sqrt{1-\mu^2}}\right), \quad x \in (-20, 20)$$

which leads to an analytic antikink solution

$$u(x,t) = 4\arctan\left(\exp\left(-\frac{x-\mu t}{\sqrt{1-\mu^2}}\right)\right). \tag{46}$$

Compared with the kink solution (45), the antikink soliton (46) is also a traveling wave solution, but the solution varies monotonically from 2π to 0 as x varies from $-\infty$ to ∞.

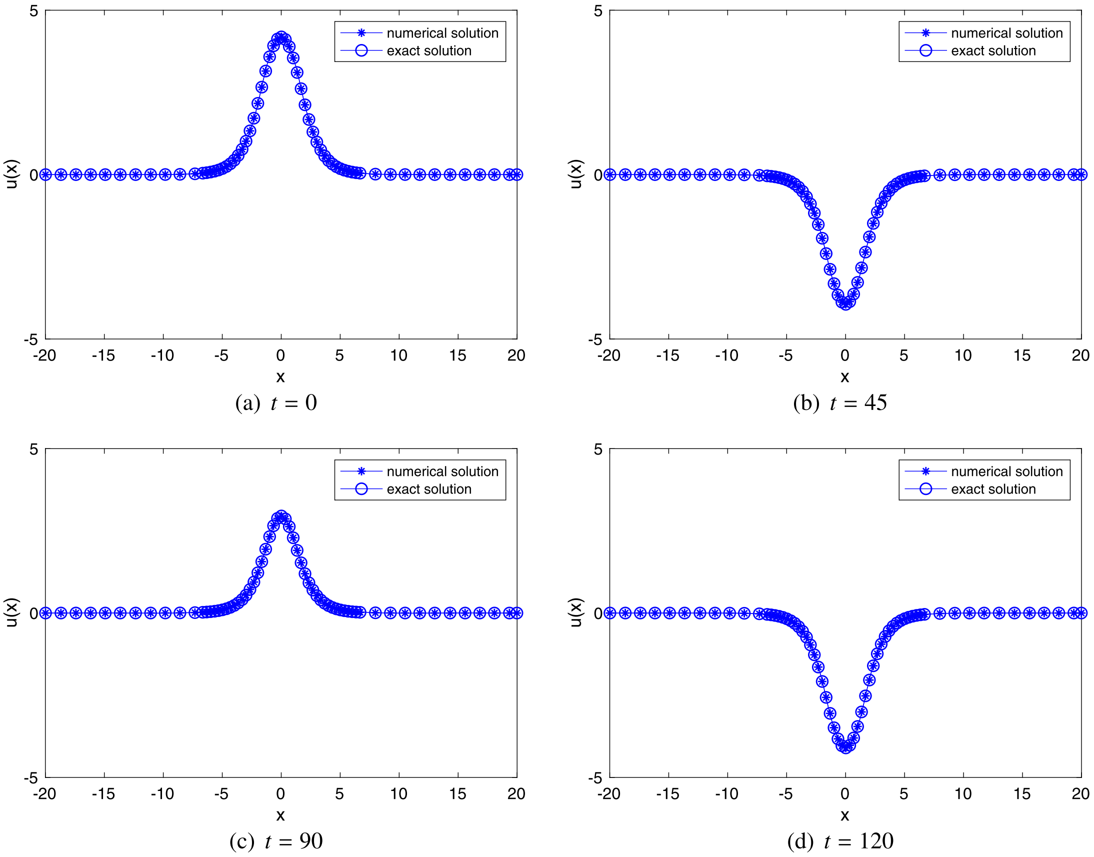
In the numerical simulation, the velocity for both the kink soliton and the antikink soliton is chosen to be $\mu = 0.2$. For the kink soliton, an energy-conserving scheme with the A.-flux is used, while the C.-flux is used in the simulation of the antikink soliton. We take u^h and v^h to be in different approximation spaces, i.e., $s = q - 1$, with $q = 4$. Finally, the problem is evolved with a 4-stage Runge Kutta time integrator until $T = 80$ with time step size $\Delta t = 0.01$.

The space-time plots of the kink and antikink solitons are shown in Fig. 5. From the left to the right are the kink soliton and the antikink soliton respectively. From the left graph, we see that the kink soliton increases monotonically from 0 to 2π and the antikink soliton decreases from 2π to 0 monotonically in the right graph. Both kink and antikink solitons move from the left to the right and keep their original shape.

4.2.3. *Kink-kink collision and kink-antikink collision*

To numerically simulate the kink-kink collision, we use the superposition of two kink solitons as the initial condition for (44), one moves from the left to the right and the other moves from the right to the left as follows,

$$u(x,0) = 4\arctan\left(\exp\left(\frac{x+10}{\sqrt{1-\mu^2}}\right)\right) + 4\arctan\left(\exp\left(\frac{x-10}{\sqrt{1-\mu^2}}\right)\right), \quad x \in (-20, 20),$$



(a) $t = 0$ (b) $t = 45$

(c) $t = 90$ (d) $t = 120$

Fig. 2. Plots of the standing breather with the degree of approximation space $q = s = 4$. The S.-flux is used in the simulation. From the top to the bottom, the left to the right, the numerical and exact breather solutions at $t = 0, 45, 90, 120$ are plotted.

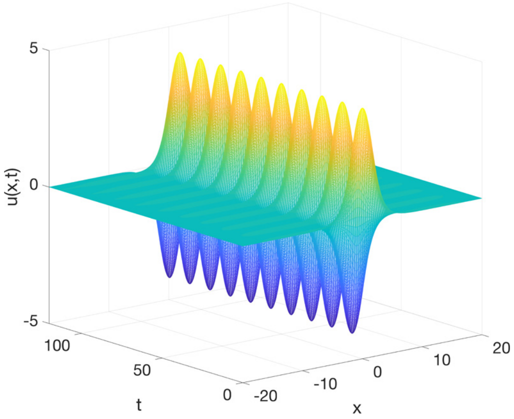
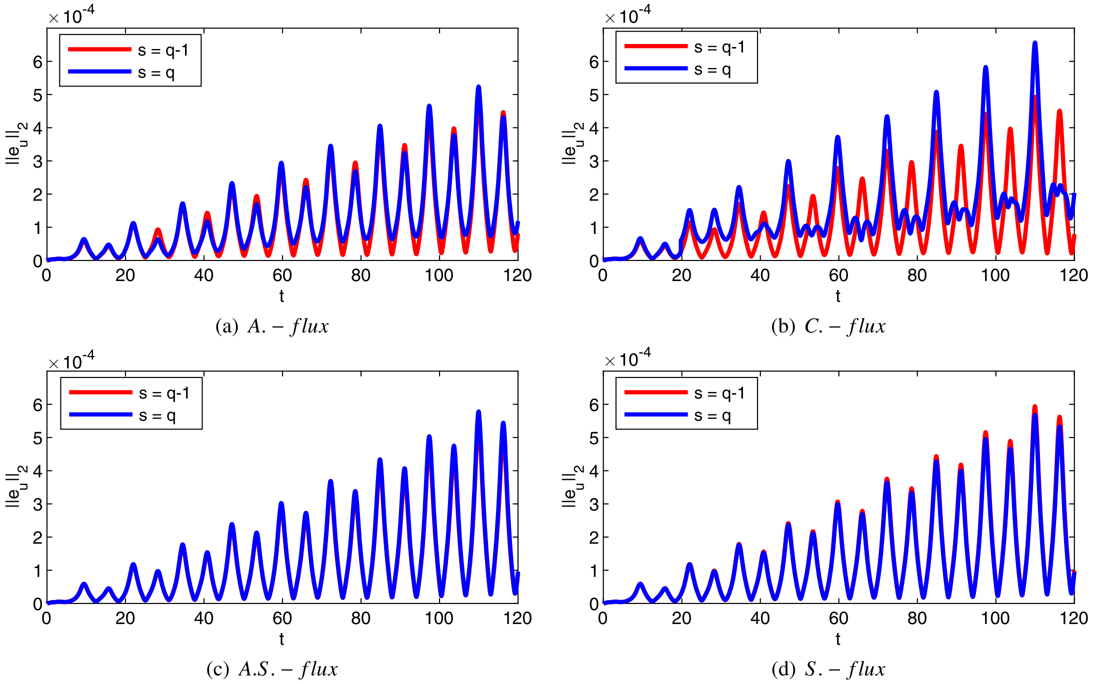


Fig. 3. Space-time plots of the standing breather with the degree of approximation space $q = s = 4$. The S.-flux is used in the simulation.

$$\frac{\partial u}{\partial t}(x,0) = -\frac{2\mu}{\sqrt{1-\mu^2}}\,\mathrm{sech}\left(\frac{x+10}{\sqrt{1-\mu^2}}\right) + \frac{2\mu}{\sqrt{1-\mu^2}}\,\mathrm{sech}\left(\frac{x-10}{\sqrt{1-\mu^2}}\right), \quad x \in (-20, 20).$$

Similarly, for the kink-antikink soliton collision we choose the superposition of a kink soliton and an antikink soliton as the initial conditions; the kink soliton moves from the left to the right and the antikink soliton moves from the right to the left as follows,

$$u(x,0) = 4\arctan\left(\exp\left(\frac{x+10}{\sqrt{1-\mu^2}}\right)\right) + 4\arctan\left(\exp\left(-\frac{x-10}{\sqrt{1-\mu^2}}\right)\right), \quad x \in (-20, 20),$$



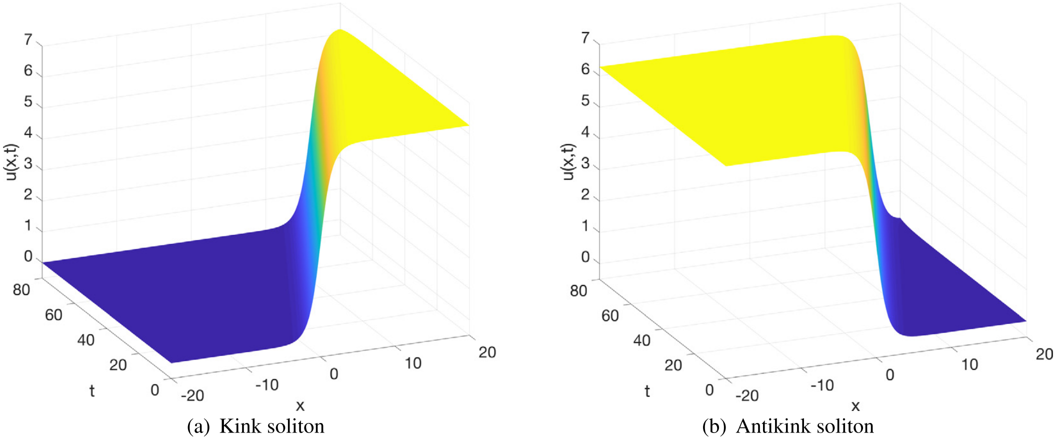
(a) $A.-flux$

(b) $C.-flux$

(c) $A.S.-flux$

(d) $S.-flux$

Fig. 4. Plots of the history of the L^2 errors for u, standing breather. The first row is for energy-conserving schemes, from the left to right the A.-flux and the C.-flux respectively. The second row is for energy-dissipating schemes, from the left to right the A.S.-flux and the S.-flux respectively. The degree of the approximation space for u is $q = 4$ and for v is s.



(a) Kink soliton

(b) Antikink soliton

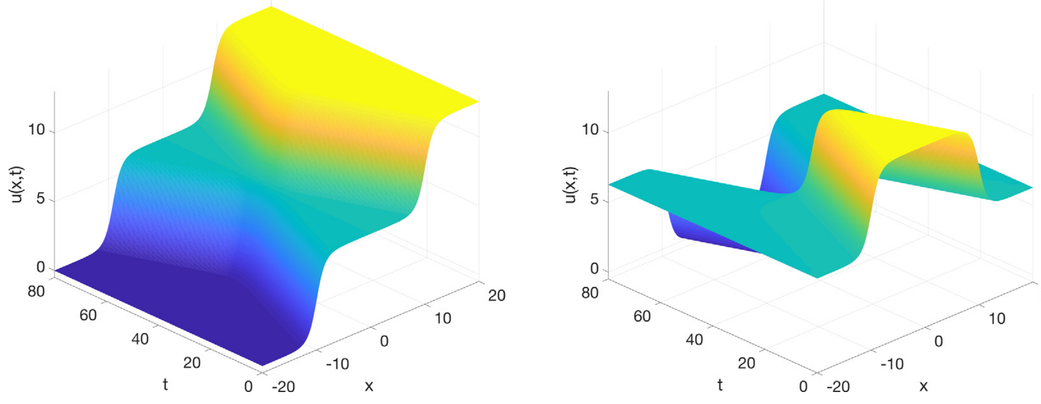
Fig. 5. From the left to the right, plots for the kink and antikink solitons respectively. The degree of approximation space for u is $q = 4$ and for v is $s = 3$. An energy-based DG scheme with the A.-flux is used in the simulation of kink soliton and the C.-flux is used for the simulation of antikink soliton.

$$\frac{\partial u}{\partial t}(x,0) = -\frac{2\mu}{\sqrt{1-\mu^2}}\text{sech}\left(\frac{x+10}{\sqrt{1-\mu^2}}\right) - \frac{2\mu}{\sqrt{1-\mu^2}}\text{sech}\left(-\frac{x-10}{\sqrt{1-\mu^2}}\right), \quad x \in (-20, 20).$$

Note that we simply use the superposition of two kink solitons (kink and antikink solitons) to be the initial conditions rather than the analytic solution of the corresponding collisions.

The parameter μ is chosen to be 0.2 in the numerical simulation. For the kink-kink collision soliton, an energy-dissipating scheme with the A.S.-flux is used, and the S.-flux is used in the simulation of kink-antikink collision soliton. Besides, u^h, v^h are assumed to be in different approximation spaces, i.e., $s = q - 1$, with $q = 4$. Finally, the problem is evolved with a 4-stage Runge-Kutta time integrator until $T = 80$ with time step size $\Delta t = 0.01$.

The plots of the kink-kink and the kink-antikink soliton collisions are shown in Fig. 6. In the left graph we observe that initially the two kinks move towards each other at the same speed. The kink with the profile from 0 to 2π moves from left to right and the kink with profile from 2π to 4π moves from right to left. After a certain time, they collide with each other and are immediately reflected, keeping their original shape while moving in the opposite direction. The space-time plot of



(a) Kink-kink collision

(b) Kink-antikink collision

Fig. 6. From the left to the right plots of the kink-kink and kink-antikink collisions respectively: the degree of approximation space for u is $q = 4$ and for v is $s = 3$. An energy-based DG scheme with A.S.-flux is used in the simulation of kink-kink collision and the S.-flux is used for the simulation of the kink-antikink collision.

Table 17
Linear regression estimates of the convergence rate for u with C.-flux, A.-flux, S.-flux and A.S.-flux for the 2D test problem. The approximation degrees for u^h, v^h are $q_x = q_y = s_x = s_y = q$.

Degree (q) of approx. of u	1	2	3	4	5	6
Rate fit with A.-flux	0.19	2.00	3.89	4.98	5.73	6.77
Rate fit with C.-flux	1.40	1.99	4.31	4.93	6.18	6.64
Rate fit with A.S.-flux	0.89	2.86	4.14	5.03	6.03	7.02
Rate fit with S.-flux	1.05	2.87	4.07	5.02	6.02	7.01

the kink-antikink collision is shown in the right graph. We see that the kink and antikink solitons move towards each other at the same speed. Here the kink with profile from 2π to 4π moves left to right and the antikink with profile from 4π to 2π moves right to left. After the collision, they move away from each other with their original velocity and direction but changed profiles.

4.3. Convergence in 2D

In this section we investigate the convergence rate of the proposed energy-based DG scheme in 2D. Specifically, we set $\theta = 0$ and $f(u) = -4u^3$, i.e.,

$$\frac{\partial^2 u}{\partial t^2} = c^2\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right) - 4u^3 + f_1(x, y, t), \quad (x, y) \in (0, 1) \times (0, 1), \quad t \geq 0. \tag{47}$$

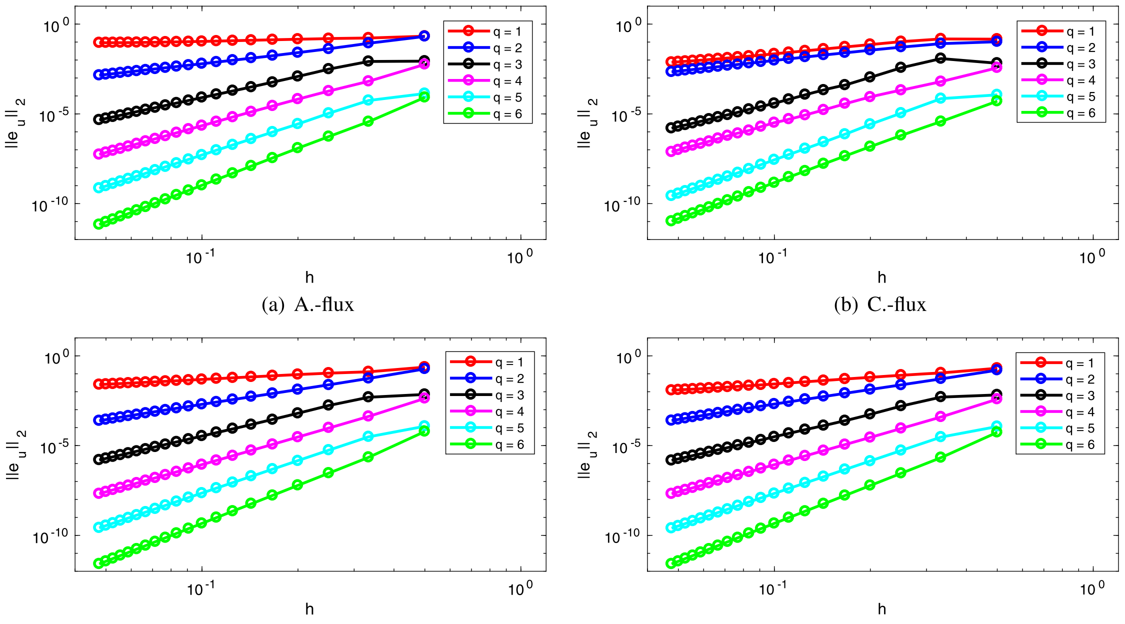
We construct a manufactured solution

$$u(x, y, t) = \cos(2\pi x)\cos(2\pi y)\sin(2\pi t), \quad (x, y) \in (0, 1) \times (0, 1), \quad t \geq 0, \tag{48}$$

to solve (47). The initial conditions, Neumann boundary conditions and external forcing $f_1(x, y, t)$ are determined by u in (48).

The discretization is performed with elements whose vertices are on the Cartesian grids defined by $x_i = ih, y_j = jh, i, j = 0, 1, \cdots, n$ with $h = 1/n$. We evolve the solution with the RK4 time integrator until the final time $T = 0.2$ with a time step size of $\Delta t = 0.075h/(2\pi)$. As in the 1D test, we use four different fluxes: C.-flux, A.-flux, A.S.-flux and S.-flux, but only consider the case where u^h and v^h are in the same approximation space, i.e., $q_x = s_x = q$ and $q_y = s_y = q$.

In Fig. 7, the L^2 errors for u are plotted against the mesh size $h_x = h_y = h$. Table 17 presents the linear regression estimates of the convergence rate for u based on the data in Fig. 7. Note that we only use the ten finest grids to compute the convergence rate here. From Table 17, we observe the optimal convergence rate of $q + 1$ for the A.S.-flux and the S.-flux when $q \geq 2$ and an order reduction by 1 compared with the optimal convergence rate for $q = 1$. For the A.-flux and C.-flux, we observe optimal convergence for $q \geq 3$, and an order reduction by 1 for $q = 2$. When $q = 1$, the A.-flux has an order reduction by 2 compared with the optimal rate and for the C.-flux an order reduction by $\frac{1}{2}$ compared with optimal. These observations are consistent with the results in 1D.



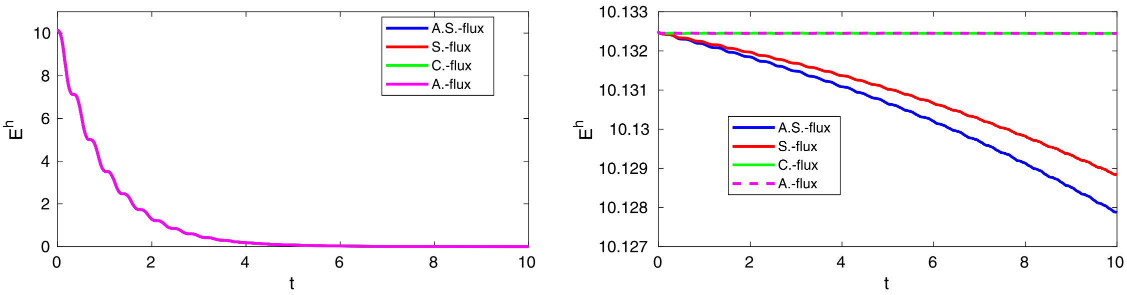
(a) A.-flux

(b) C.-flux

(c) A.S.-flux

(d) S.-flux

Fig. 7. The plot of L^2 errors for u for the 2D convergence test: from left to right, top to bottom A.-flux, C.-flux, A.S.-flux and S.-flux respectively. The approximation degrees for u^h, v^h are $q_x = q_y = s_x = s_y = q$.



(a) $\theta = 1$

(b) $\theta = 0$

Fig. 8. The plots of energy for DG solutions of (49) with four different fluxes. For the left graph, the dissipating term is considered, i.e., $\theta = 1$; while the right graph does not contain the dissipating term, i.e., $\theta = 0$.

4.4. Time history of the numerical energy in 2D

We compare the numerical energy for both cases with $\theta = 0$ and $\theta = 1$ in this section. Precisely, we consider,

$$\frac{\partial^2 u}{\partial t^2} + \theta \frac{\partial u}{\partial t} = c^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) - 4u^3, \quad (x, y) \in (0,1) \times (0,1), \quad t > 0, \tag{49}$$

with initial conditions

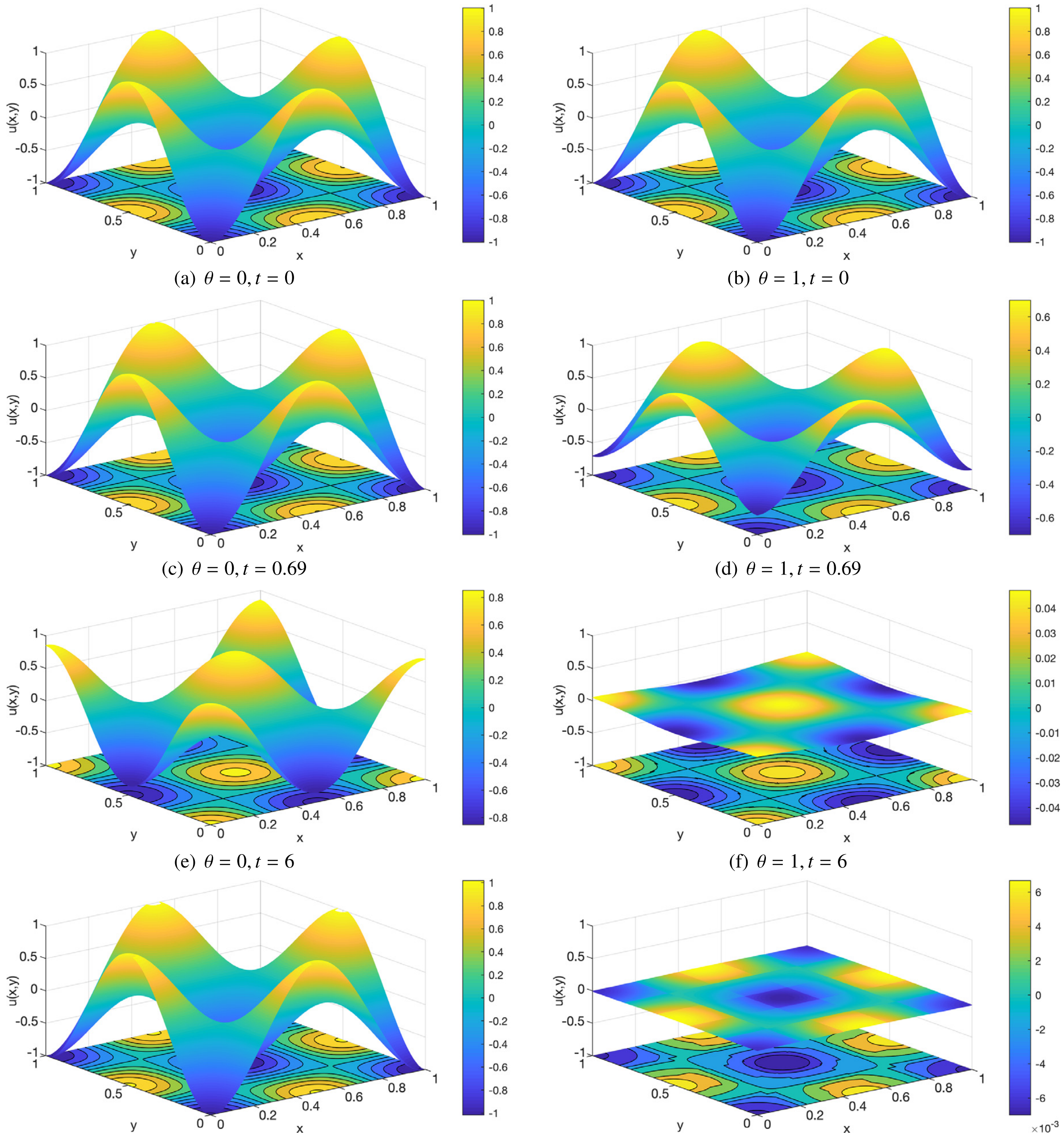
$$u(x, y, 0) = -\cos(2\pi x)\cos(2\pi y), \qquad \frac{\partial u}{\partial t}(x, y, 0) = \cos(2\pi x)\cos(2\pi y), \quad (x, y) \in (0,1) \times (0,1),$$

and flux free physical boundary conditions,

$$\frac{\partial u}{\partial x}(0, y, t) = \frac{\partial u}{\partial x}(1, y, t) = 0, \quad y \in (0,1), \quad t > 0; \quad \frac{\partial u}{\partial y}(x, 0, t) = \frac{\partial u}{\partial y}(x, 1, t) = 0, \quad x \in (0,1), \quad t > 0.$$

The space discretization is same as in Section 4.3 with $n = 5$. The degree of the approximation space is set to be $q_x = q_y = s_x = s_y = 4$. Finally, the problems are evolved with the RK4 time integrator until the final time $T = 10$ with time step size chosen to be $\Delta t = 0.075h/(2\pi)$.

In Fig. 8, the left graph shows the numerical energy evolution with four different fluxes for the problem with dissipating term, $\theta = 1$; the right graph presents the numerical energy evolution with four different fluxes for the problem without



(a) $\theta = 0, t = 0$

(b) $\theta = 1, t = 0$

(c) $\theta = 0, t = 0.69$

(d) $\theta = 1, t = 0.69$

(e) $\theta = 0, t = 6$

(f) $\theta = 1, t = 6$

(g) $\theta = 0, t = 10$

(h) $\theta = 1, t = 10$

Fig. 9. Plots of u for the focusing equation (50) at times $t = 0, 0.69, 6, 10$ with the S.-flux and $q_x = q_y = s_x = s_y = q = 4$. For the left column, $\theta = 0$; for the right column, $\theta = 1$.

dissipating term, $\theta = 0$. We observe that for the case without dissipating term both the A.-flux and C.-flux conserve the numerical energy; both S.-flux and A.S.-flux are energy dissipating but the total dissipation is small. For the case with dissipating term, $\theta = 1$, the numerical energy dissipates for all fluxes, the numerical energy evolution for schemes with A.S.-flux, S.-flux, C.-flux and A.-flux are on top of each other and the numerical energy dissipates very fast.

4.5. Focusing equation

Finally, we consider a focusing problem whose energy is indefinite. Specifically, we test the problem

$$\frac{\partial^2 u}{\partial t^2} + \theta \frac{\partial u}{\partial t} = c^2 \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) + 4u^3, \quad (x, y) \in (0,1) \times (0,1), \quad t > 0, \tag{50}$$

for both $\theta = 0$ and $\theta = 1$. We use the same initial data as in Section 4.4

$$u(x, y, 0) = -\cos(2\pi x)\cos(2\pi y), \quad \frac{\partial u}{\partial t}(x, y, 0) = \cos(2\pi x)\cos(2\pi y), \quad (x, y) \in (0, 1) \times (0, 1),$$

and periodic boundary conditions are imposed in both x and y directions with $u(0, y, t) = u(1, y, t)$, $y \in (0, 1)$ and $u(x, 0, t) = u(x, 1, t)$, $x \in (0, 1)$.

The space discretization is the same with the one in Section 4.3 and we set $n = 5$. The degree of the approximation space is set to be $q_x = q_y = s_x = s_y = 4$. Finally, the problems are evolved with the RK4 time integrator and the S.-flux with time step size $\Delta t = 0.075h/(2\pi)$.

Fig. 9 shows the time evolution of u. From the left column to the right column are for the problem (50) without ($\theta = 0$) and with ($\theta = 1$) the dissipating term respectively. On the left column, we observe that the solution u seems to be approximately periodic in time; it recovers its original shape around $t = 0.69$ at first. The right column is for $\theta = 1$, we note that the solution loses its energy as time goes by; at $t = 0.69$, it has a similar shape to the case $\theta = 0$, but the amplitude of the solution is smaller.

5. Conclusions and future work

In conclusion, we have demonstrated that the energy-based DG method for second-order wave equations can be generalized to semilinear problems. In particular we:

i. Modified the weak form proposed in [1] so that the time derivatives of the approximate solution can be computed via the solution of a linear system of equations in each element,
ii. Established the stability of the method by proving energy estimates for a wide choice of fluxes with mesh-independent parametrizations, including energy-conserving central or alternating fluxes as well as dissipative upwind fluxes,
iii. Derived suboptimal estimates of convergence in the energy norm,
iv. Observed, for polynomial degrees above 3, optimal convergence in the L^2 norm for the energy-conserving alternating flux as well as for dissipative methods based on Sommerfeld flux splitting.

Our main target for future work will be extensions to systems as well as to problems with more general nonlinearities. This will enable applications to a wider variety of problems of physical interest. Here again we plan to exploit the fact that the energy estimates only depend on the satisfaction of the weak form for certain test functions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Appelö, T. Hagstrom, A new discontinuous Galerkin formulation for wave equations in second-order form, SIAM J. Numer. Anal. 53 (2015) 2705–2726.
[2] J. Hesthaven, T. Warburton, Nodal Discontinuous Galerkin Methods, Texts in Applied Mathematics, vol. 54, Springer-Verlag, New York, 2008.
[3] D. Appelö, T. Hagstrom, An energy-based discontinuous Galerkin discretization of the elastic wave equation in second order form, Comput. Methods Appl. Mech. Eng. 338 (2018) 362–391.
[4] L. Zhang, T. Hagstrom, D. Appelö, An energy-based discontinuous Galerkin method for the wave equation with advection, SIAM J. Numer. Anal. 57 (2019) 2469–2492.
[5] Y. Du, L. Zhang, Z. Zhang, Convergence analysis of a discontinuous Galerkin method for wave equations in second-order form, SIAM J. Numer. Anal. 57 (2019) 238–265.
[6] Y. Du, J. Wang, Convergence analysis of an energy based discontinuous Galerkin method for the wave equation in second order form: hp version, Comput. Math. Appl. 79 (2020) 3223–3240.
[7] T. Bui-Thanh, From Godunov to a unified hybridized discontinuous Galerkin framework for partial differential equations, J. Comput. Phys. 295 (2015) 114–146.
[8] B. Riviére, M. Wheeler, Discontinuous finite element methods for acoustic and elastic wave problems, Contemp. Math. 329 (2003) 271–282.
[9] M. Grote, A. Schneebeli, D. Schötzau, Discontinuous Galerkin finite element method for the wave equation, SIAM J. Numer. Anal. 44 (2006) 2408–2431.
[10] C.-S. Chou, C.-W. Shu, Y. Xing, Optimal energy conserving local discontinuous Galerkin methods for second-order wave equation in heterogeneous media, J. Comput. Phys. 272 (2014) 88–107.
[11] P. Aursand, U. Koley, Local discontinuous Galerkin schemes for a nonlinear variational wave equation modelling liquid crystals, J. Comput. Appl. Math. 317 (2017) 478–499.
[12] N. Yi, H. Liu, An energy conserving local discontinuous Galerkin method for a nonlinear variational wave equation, Commun. Comput. Phys. 23 (2018) 747–772.
[13] P.G. Ciarlet, The Finite Element Method for Elliptic Problems, Classics in Applied Mathematics, vol. 40, SIAM, 2002.