Comput. Methods Appl. Mech. Engrg. 199 (2010) 1424–1436

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

journal homepage: www.elsevier.com/locate/cma

A local discontinuous Galerkin method for a doubly nonlinear diffusion equation arising in shallow water modeling

Mauricio Santillana *, Clint Dawson

Institute for Computational Engineering and Sciences, University of Texas at Austin, United States

ARTICLE INFO

Article history:
Received 14 May 2009
Received in revised form 16 November 2009
Accepted 18 November 2009
Available online 26 November 2009

Keywords:
Discontinuous Galerkin
Nonlinear diffusion
Doubly nonlinear
Shallow water equations
Diffusive wave approximation

ABSTRACT

In this paper, we study a local discontinuous Galerkin (LDG) method to approximate solutions of a doubly nonlinear diffusion equation, known in the literature as the diffusive wave approximation of the shallow water equations (DSW). This equation arises in shallow water flow models when special assumptions are used to simplify the shallow water equations and contains as particular cases: the Porous Medium equation and the parabolic p-Laplacian. Continuous in time *a priori* error estimates are established between the approximate solutions obtained using the proposed LDG method and weak solutions to the DSW equation under physically consistent assumptions. The results of numerical experiments in 2D are presented to verify the numerical accuracy of the method, and to show the qualitative properties of water flow captured by the DSW equation, when used as a model to simulate an idealized dam break problem with vegetation.

1. Introduction

In this paper, we study a numerical scheme based on the local discontinuous Galerkin (LDG) method as a means to approximate solutions to a doubly nonlinear diffusion equation, known in the literature as the diffusive wave approximation of the shallow water equations (DSW). This equation arises in shallow water flow models when special assumptions are used to simplify the shallow water equations (SWE), and it gives rise to the following initial/boundary-value problem (IBVP):

$$\begin{cases} \frac{\partial u}{\partial t} - \nabla \cdot \left(\frac{(u-z)^{\alpha}}{|\nabla u|^{1-\gamma}} \nabla u \right) = f & \text{on } \Omega \times (0,T], \\ u = u_0 & \text{on } \Omega \times \{t=0\}, \\ \left(\frac{(u-z)^{\alpha}}{|\nabla u|^{1-\gamma}} \nabla u \right) \cdot n = B_N & \text{on } \partial\Omega \cap \Gamma_N \times (0,T], \\ u = B_D & \text{on } \partial\Omega \cap \Gamma_D \times (0,T], \end{cases} \tag{1}$$

where Ω is an open, bounded subset of $\mathbb{R}^2$, Γ_N and Γ_D are subsets of $\partial\Omega \in C^1$ such that $\partial\Omega = \Gamma_N + \Gamma_D f : \Omega \times (0,T] \to \mathbb{R}, u_0 : \Omega \to \mathbb{R}, B_N : \partial\Omega \cap \Gamma_N \times (0,T] \to \mathbb{R}$, and $B_D : \partial\Omega \cap \Gamma_D \times (0,T] \to \mathbb{R}$ are given, $z : \overline{\Omega} \to \mathbb{R}^+$ is a positive time independent function, n is the outward normal to $\Gamma_N, 0 < \gamma \leqslant 1, 1 < \alpha < 2$ and $u : \Omega \times (0,T] \to \mathbb{R}$ is the unknown. Here, $|\cdot| : \mathbb{R}^d \to \mathbb{R}$ refers to the Euclidean norm in $\mathbb{R}^d$ $(d = 1, 2$, in our work).

The DSW equation has been successfully applied as a model to simulate *overland flow* and *shallow water flow in vegetated areas*, where water flow is driven mainly by gravitational forces and dominated by shear stresses. See for example [29,20,33,17,18] and [25]. In these water flow regimes, the solution $u(x,t)$ of the IBVP (1) represents the time evolution of the water height with respect to a given datum. The time independent function $z(x)$ represents the *bathymetry* or *topography* over which the water flows, the – frequently time dependent – function f represents sources and sinks (e.g. rainfall or infiltration) and the boundary conditions, B_N and B_D, simulate lateral inflow/outflow and the presence of specified water elevation, respectively. A detailed mathematical formulation and derivation of the IBVP (1), in the context of shallow water modeling can be found in [2].

The use of a single equation to describe the time evolution of water flow *in lieu* of the full shallow water system of equations becomes advantageous, both from the conceptual and computational points of view. Indeed, numerically solving the DSW equation is considerably cheaper than numerically solving the SWE [20]. However, determining convergence and stability of numerical schemes to approximate solutions of the DSW equation is not a simple task [25]. Difficulties to analyze numerical schemes aimed at approximating solutions of the DSW equation arise from the fact that – to the best of our knowledge – existence, uniqueness, and regularity of solutions to the DSW equation for general non-zero *bathymetries*, $z(x)$, have not been studied. Note that the DSW equation contains as particular cases two complicated nonlinear diffusion equations: the Porous Medium equation (PME), when $z = 0$ and $\gamma = 1$, and the p-Laplacian for $1 < p < 2$, when $\alpha = 0$

* Corresponding author. Present address: Harvard University Center for the Environment. Tel.: +1 512 698 1564.
E-mail address: msantill@fas.harvard.edu (M. Santillana).

0045-7825/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cma.2009.11.016

and $p = \gamma + 1$, this case is not considered in our work, recall that we consider only $1 < \alpha < 2$.

The motivation for the present work emanates from our previous work contained in [2] and [25]. Particularly from [25], where we studied *numerically* some qualitative properties of solutions to the DSW for a collection of non-zero bathymetries $z(x)$ in 1D, using the continuous Galerkin method. Our findings indicated that characteristics such as: the existence of compactly supported solutions, as well as the finite speed of propagation of disturbances (found analytically for solutions for the DSW equation, for flat bathymetries in 1D in [16] ($z = 0$)), persisted for non-zero bathymetries. The property of finite speed of propagation – as opposed to the infinite speed of propagation in the heat equation, for example – can be understood as a consequence of the *advection–diffusion* nature of certain types of nonlinear diffusion equations such as the PME (see [27]) and the DSW equation, and gives rise the presence of free boundaries (locations where the solution goes from $u = 0$ to $u > 0$) and oftentimes traveling sharp fronts. Discontinuous Galerkin methods are well known to be able to capture sharp fronts in solutions to hyperbolic systems – as well as to be locally mass conservative – thus, making them suitable methods to solve our problem. Furthermore, the LDG approach outlined here fits into an overall discontinuous Galerkin framework being developed by our group for the approximation of shallow water systems [21].

The work presented in this paper is organized in the following way. In Sections 1.1–1.4 and 1.5, we introduce the DSW equation, and present a brief introduction to DG methods, the notation used in such methods, and all the preliminary information needed to carefully set up and study our particular LDG method. The numerical method is constructed in Section 2, and the details of the continuous in time error analysis are presented in Section 2.2. In Section 3 the results of 2D numerical experiments are shown.

1.1. The DSW equation

For completeness in our presentation, we mention some of the key characteristics of the DSW equation that make it an interesting problem to be studied, as well as the context in which we intend to approach our convergence analysis.

The DSW is a doubly nonlinear diffusion equation, since the product of two nonlinearities involving u and ∇u, namely $(u-z)^\alpha$ and $\nabla u/|\nabla u|^{1-\gamma}$, appear inside the divergence term. Also, when written in the form

$$\frac{\partial u}{\partial t} - \nabla \cdot (a(u, \nabla u)\nabla u) = f \tag{2}$$

with the diffusion coefficient a given by

$$a(u, \nabla u) = \frac{(u-z)^\alpha}{|\nabla u|^{1-\gamma}}, \tag{3}$$

one immediately notices that the nonlinearity involving the gradient of u inside the divergence, $\nabla u/|\nabla u|^{1-\gamma}$, is at best γ-Hölder continuous w.r.t. ∇u, since it scales as $|\nabla u|^\gamma$ and $0 < \gamma \leqslant 1$. As a consequence, the familiar *coercivity* and *continuity* conditions

$$\begin{aligned} \mu\|u\|_V^2 &\leqslant (a(u)\nabla u, \nabla u) \quad \text{and} \quad (a(u)\nabla u, \nabla w) \\ &\leqslant M\|u\|_V\|w\|_V \quad \text{for } u, w \in V, \end{aligned} \tag{4}$$

commonly assumed in the numerical analysis of nonlinear diffusion equations (see [28,15] or [26]) will not hold.[1] This fact motivates the need for further assumptions or properties on the type of solutions to be approximated if one is to produce a meaningful numerical method. To this end, we follow the strategy we presented in [25], for the convergence of the continuous Galerkin method, to restrict our analysis to the approximation of solutions satisfying physically consistent properties based on shallow water modeling theory. Even though the DSW is a degenerate diffusion equation, we will assume that (1) the solution u does not vanish (*i.e.* $u > \epsilon$, for a small $\epsilon > 0$), which corresponds to a wet condition throughout the domain, and (2) that the gradient of u is bounded. The latter assumption is consistent with the derivation of the DSW from the SWE, for general and smooth bathymetries. With these assumptions, we proved in [25] that the continuous Galerkin (CG) method converges to the (assumed to be unique and regular) solution of the DSW equation, for finite elements of polynomial order k with order $\mathcal{O}(h^{k\gamma})$ (here h represents the diameter of the spatial triangulation). We also found that, in theory, we had to use polynomials of degree $k > 4$ in order to ensure the boundedness needed on the discrete solution for the proof to succeed. In our numerical experiments in [25], however, we found that we could achieve convergence for our method even for solutions of the DSW that vanished in large regions of the domain, as well as for solutions with unbounded gradient. Moreover, we found that for nondegenerate solutions we could achieve – optimal – convergence rates $\mathcal{O}(h^2)$ with piecewise linear elements. These results show the gap between our *conservative* theoretical convergence analysis and the actual numerically achievable convergence rates. We do not address this gap in this work. Instead, we focus on extending the results obtained for the CG method to the LDG method.

1.2. DG methods

The LDG method is one of many *discontinuous Galerkin* (DG) methods. These methods are characterized by the fact that continuity across elements is not enforced in the linear space where the basis functions live, and thus, the approximate solutions produced are discontinuous or "broken". This major difference with the continuous Galerkin finite element method gives rise to very interesting properties that characterize all DG methods. These can be summarized as follows: (1) They can easily handle various shapes in different elements across the domain, as well as local spaces of different types (orders). This is the case since continuity is not enforced *strongly* across elements. (2) The previous property makes these methods suitable to handle structured and unstructured meshes in domains with general geometries. (3) Their high degree of locality makes them highly parallelizable. (4) They are element-wise conservative (This statement is meaningful when modeling nonlinear conservation laws). (5) They are ideally suited for *hp*-refinement (or *hp*-adaptivity). A good reference that offers a review on the development of discontinuous Galerkin methods is the book by Cockburn et al. [13].

The LDG method was introduced by Cockburn and Shu in [14] as an extension, to general convection–diffusion problems, from the numerical techniques introduced by Bassi and Rebay in [4] to solve the compressible Navier–Stokes equations. One of the basic ideas in the LDG method is to rewrite, say the parabolic equation at hand, as a degenerate first order system of equations, and solve for u and $\nabla u (= \boldsymbol{q})$ as *independent unknowns*. Even though this strategy is also utilized in methods based on a *mixed* formulation, in the LDG method one further discretizes the resulting first order system using particular DG techniques. It is particular to the LDG method studied in this work that the approximation to u, and the approximation to each of the components of $\boldsymbol{q}$ belong to the same approximation spaces. This choice makes the coding of the method simpler than the standard mixed methods. Also, the so-called *numerical flux*, $\widehat{U}$, (introduced to properly define the values of the solution u and the fluxes across all element boundaries) does not depend on $\boldsymbol{q}$, making it possible for the local variable $\boldsymbol{q}$ to be solved in terms of u. The particular numerical fluxes, $\widehat{U}$ and $\widehat{\boldsymbol{Q}}$, used in our method are introduced in Section 2. Examples of other consistent

[1] In (4), $(\cdot,\cdot)$ represents the appropriate duality pairing.

numerical fluxes, in the context of elliptic problems, can be found in [9].

Works addressing the properties of the LDG method in the context of convection–diffusion problems include for example: [14,12,10] and [1]. The applicability of the LDG method has been explored for example, for elliptic problems in [3], and in [19]; for nonlinear diffusion problems in [7], and in [22]; for a class of nonlinear problems in fluid mechanics in [8]; for Richard's equation (another nonlinear parabolic equation) in [23]; for nonlinear second-order elliptic and hyperbolic systems in [24]; for nonlinear convection–diffusion and KdV equations in [30]; and for PDE's with higher order derivatives in [31] and [32].

1.3. Regularized problem

In [25] we used a strategy that consisted of constructing a regularized numerical scheme to approximate the possibly degenerate diffusion coefficient $a(u,\nabla u)$ in (3) with nondegenerate diffusion coefficients a_ϵ in (1), such that $0<\epsilon\leqslant a_\epsilon(u)$ and with the property that $a(u)=\lim_{\epsilon\to 0}a_\epsilon(u)$, for a small parameter ϵ. We will use a similar strategy in our study.

We present the nondegenerate problem that we will approximate numerically along with some properties and results that will be used in the analysis carried out in the next sections. We begin by introducing the nondegenerate version of the IBVP (1), obtained by replacing the function $(s-z)^\alpha$ with a sequence of bounded Lipschitz functions $\{\beta_\epsilon(s)\}$, with the properties that (i) $\{\beta_\epsilon(s)\}$ converges uniformly to $(s-z)^\alpha$ as $\epsilon\to 0$, and (ii) for small $\epsilon>0$ the following holds $\beta_\epsilon(s)\geqslant\epsilon$ for all $t\in[0,T]$. To this end, the bathymetry $z(x)$ will be assumed to be a smooth and bounded time independent function defined in Ω. The nondegenerate IBVP is given by

$$\begin{cases} \frac{\partial u}{\partial t}-\nabla\cdot\left(\beta_\epsilon(u)\frac{\nabla u}{|\nabla u|^{1-\gamma}}\right)=f & \text{on } \Omega\times(0,T],\\ u=u_0 & \text{on } \Omega\times\{t=0\},\\ \left(\frac{\beta_\epsilon(u)}{|\nabla u|^{1-\gamma}}\nabla u\right)\cdot n=B_N & \text{on } \partial\Omega\cap\Gamma_N\times(0,T],\\ u=B_D & \text{on } \partial\Omega\cap\Gamma_D\times(0,T]. \end{cases} \tag{5}$$

In the next section we develop a numerical scheme to approximate this nondegenerate problem as explained in Section 1.1. The fact that solutions to the nondegenerate problem (5) are close to the original solution to problem (1) as $\epsilon\to 0$ will be understood as in [2] for $z=0$, and will be assumed for the general case $z\neq 0$.

Remark 1.1. For intuition purposes one could choose for example the following sequence $\beta_\epsilon(u)=(u-z)^\alpha+\epsilon$.

1.4. Previous results

For completeness, we present some essential results needed in the subsequent sections. For proofs of the next two lemmas see Section 1.5 in [25] and the references therein.

Lemma 1.1. *Let u_1 and u_2 be non negative $L^\infty(\Omega)$ functions, then for $\alpha\geqslant 1$*

$$|u_1^\alpha-u_2^\alpha|\leqslant\alpha\Big(\max(\|u_1\|_{L^\infty(\Omega)},\|u_2\|_{L^\infty(\Omega)})\Big)^{\alpha-1}|u_1-u_2|. \tag{6}$$

Lemma 1.2 (Coercivity and continuity). *Let η_1 and η_2 be bounded vector valued functions in $\mathbb{R}^n(n\geqslant 1)$, then the following estimates hold true*

$$\gamma\mathscr{A}_0|\eta_1-\eta_2|^2\leqslant\left(\frac{\eta_1}{|\eta_1|^{1-\gamma}}-\frac{\eta_2}{|\eta_2|^{1-\gamma}}\right)(\eta_1-\eta_2) \tag{7}$$

and

$$\left|\frac{\eta_1}{|\eta_1|^{1-\gamma}}-\frac{\eta_2}{|\eta_2|^{1-\gamma}}\right|\leqslant\mathscr{A}_0|\eta_1-\eta_2|\leqslant\frac{2}{\gamma}|\eta_1-\eta_2|^\gamma, \tag{8}$$

where

$$\mathscr{A}_0:=\int_0^1|\lambda\eta_1+(1-\lambda)\eta_2)|^{\gamma-1}d\lambda.$$

1.5. DG notation

Let $\{\mathscr{T}_h\}$ denote a family of regular finite element partitions of Ω such that no individual element Ω_e crosses $\partial\Omega$. For the error analysis described below, we will assume that $\mathscr{T}_h$ is a locally quasi-uniform finite element mesh. Let h_e denote the element diameter with h being the maximal element diameter. We will also assume each element Ω_e is Lipschitz and affinely equivalent to one of several reference elements [6]. Let $\mathscr{P}^k(\Omega_e)$ denote the space of (possibly) discontinuous piecewise polynomials of degree at most $k, k\geqslant 1$, defined on Ω_e, and let

$$\mathscr{M}=\{v: v|_{\Omega_e}\in\mathscr{P}^k(\Omega_e)\}.$$

We will assume that $\mathscr{P}^k(\Omega_e)$ is chosen such that the usual space of continuous, piecewise polynomials of order k defined on the triangulation $\mathscr{T}_h$ are contained in $\mathscr{M}$.

We will denote by ε_i the set of all interior element faces, with ε_D the set of all element faces along the Dirichlet boundary Γ_D, and ε_N the set of all element faces along the Newmann boundary Γ_N. Note that if ε is an interior face in the finite element mesh, then ε has two elements adjacent to it, we will denote them by Ω_e^- and Ω_e^+. Also, if v and $\boldsymbol{w}$ are smooth real valued and vector valued functions, respectively, defined on these elements, we will denote their traces on ε, from the interior of the element Ω_e^-, as v^- and $\boldsymbol{w}^-$; and from the exterior of the element Ω_e^+, as v^+ and $\boldsymbol{w}^+$. We will denote by $\boldsymbol{n}^-$ the outward normal vector to the element Ω_e^- at ε and by $\boldsymbol{n}^+$ the outward normal vector to the element Ω_e^+ at ε. The previous definition implies naturally that $\boldsymbol{n}^+=-\boldsymbol{n}^-$. We will define the average $\{\cdot\}$ and the jump $[\![\cdot]\!]$ on the face ε as:

$$\{v\}=\frac{(v^-+v^+)}{2},\quad \{\boldsymbol{w}\}=\frac{(\boldsymbol{w}^-+\boldsymbol{w}^+)}{2}, \tag{9}$$

$$[\![v]\!]=v^-\boldsymbol{n}^-+v^+\boldsymbol{n}^+,\quad [\![\boldsymbol{w}]\!]=\boldsymbol{w}^-\cdot\boldsymbol{n}^-+\boldsymbol{w}^+\cdot\boldsymbol{n}^+. \tag{10}$$

We will also denote by $(\cdot,\cdot)_E$, the usual L^2 inner product over a d-dimensional domain E, and by $\langle\cdot,\cdot\rangle_{\partial E}$, the $(d-1)$-dimensional integral over the surface ∂E. To simplify notation, we will omit the dependence on the domain and denote with $(\cdot,\cdot)$ the integrals over the whole domain $E=\Omega$, $(\cdot,\cdot)_\Omega$.

Throughout the paper, C will be a generic positive constant with different values and the explicit dependence with respect to parameters will be written inside parenthesis.

We refer the reader to Chapter 4 in [6] and [11] for proofs of the following lemmas.

Lemma 1.3 (Interpolation error). *Let $u\in H^{k+1}(\Omega)$, then there exists an "interpolant" $\hat{u}\in\mathscr{M}$ which satisfies*

$$\|\hat{u}-u\|_{H^s(\Omega)}\leqslant Ch^{k+1-s}\|u\|_{H^{k+1}(\Omega)}.$$

Lemma 1.4 (Inverse inequalities). *Let $v\in\mathscr{M}$ then, there exists a constant K_0 independent of h and v such that*

$$\|v\|_{L^\infty(\Omega)}\leqslant K_0h^{-1}\|v\|_{L^2(\Omega)}\quad\text{and}\quad\|\nabla v\|_{L^\infty(\Omega)}\leqslant K_0h^{-1}\|\nabla v\|_{L^2(\Omega)}.$$

The following trace theorem is well known. See Chapter 4 in [6]:

Theorem 1.1 (Trace inequality). *Suppose that region R has a Lipschitz boundary. Then there exists a constant $C=C(R)$ such that for $v\in H^1(R)$,*

$\|v\|_{L^2(\partial R)} \leqslant C\,\|v\|_{L^2(R)}^{1/2}\|v\|_{H^1(R)}^{1/2}$.

By the trace inequality and inverse inequality, for any $v \in \mathscr{M}$

$$\|v\|_{L^2(\partial\Omega_e)} \leqslant C(\Omega_e, K_0)h_e^{-\frac{1}{2}}\|v\|_{L^2(\Omega_e)}. \tag{11}$$

2. The local discontinuous Galerkin method

In this section, we study the approximation properties of numerical solutions to the DSW equation, through the regularized initial/boundary-value problem (5), obtained using the LDG method. In order to formulate the LDG method it is appropriate to rewrite the nonlinear degenerate parabolic IBVP (1) as a degenerate first order system of equations where $u, \nabla u$, and $a(u,\nabla u)\nabla u$, are now considered as *independent unknowns*:

$$\begin{cases} u_t - \nabla\cdot\boldsymbol{q} = f & \text{on } \Omega\times(0,T],\\ \tilde{\boldsymbol{q}} = \nabla u & \text{on } \Omega\times(0,T],\\ \boldsymbol{q} = a(u,\tilde{\boldsymbol{q}})\tilde{\boldsymbol{q}} & \text{on } \Omega\times(0,T], \end{cases} \tag{12}$$

where

$$a(u,\tilde{\boldsymbol{q}}) = \frac{(u-z)^{\alpha}}{|\tilde{\boldsymbol{q}}|^{1-\gamma}}, \tag{13}$$

with initial and boundary conditions given as before by

$$\begin{cases} u = u_0 & \text{on } \Omega\times\{0\},\\ \boldsymbol{q}\cdot n = B_N & \text{on } \partial\Omega\cap\Gamma_N\times(0,T],\\ u = B_D & \text{on } \partial\Omega\cap\Gamma_D\times(0,T], \end{cases} \tag{14}$$

where $\partial\Omega = \Gamma = \Gamma_N + \Gamma_D$. Moreover, assuming $u, \boldsymbol{q}$ and $\tilde{\boldsymbol{q}}$ are smooth enough, we multiply each equation in (12) by test functions $w \in \mathscr{M}$, $\boldsymbol{v} \in (\mathscr{M})^d$ and $\tilde{\boldsymbol{v}} \in (\mathscr{M})^d$ respectively (where d is the spatial dimension), and integrate by parts over an element Ω_e to obtain the local weak form of (12):

$$\begin{cases} (u_t,w)_{\Omega_e} + (\boldsymbol{q},\nabla w)_{\Omega_e} - \langle \boldsymbol{q}\cdot\boldsymbol{n}_e, w\rangle_{\partial\Omega_e} = (f,w)_{\Omega_e} & \forall w\in\mathscr{M},\\ (\tilde{\boldsymbol{q}},\tilde{\boldsymbol{v}})_{\Omega_e} + (u,\nabla\cdot\tilde{\boldsymbol{v}})_{\Omega_e} - \langle u, \boldsymbol{v}\cdot\boldsymbol{n}_e\rangle_{\partial\Omega_e} = 0 & \forall\tilde{\boldsymbol{v}}\in(\mathscr{M})^d,\\ (\boldsymbol{q},\boldsymbol{v})_{\Omega_e} - (a(u,\tilde{\boldsymbol{q}})\tilde{\boldsymbol{q}},\boldsymbol{v})_{\Omega_e} = 0 & \forall\boldsymbol{v}\in(\mathscr{M})^d, \end{cases} \tag{15}$$

where $\boldsymbol{n}_e$ represents the outward normal vector to the faces of the element Ω_e. The discontinuous Galerkin method consists of finding approximations $(U,\mathbf{Q},\tilde{\mathbf{Q}})$ to the solution $(u,\boldsymbol{q},\tilde{\boldsymbol{q}})$ of (15), where $U \in \mathscr{M}$ and $\mathbf{Q},\tilde{\mathbf{Q}} \in \mathscr{M}^{\boldsymbol{d}}$, satisfying for all $t\in[0,T]$

$$\begin{cases} (U_t,w)_{\Omega_e} + (\mathbf{Q},\nabla w)_{\Omega_e} - \langle \widehat{\mathbf{Q}}\cdot\boldsymbol{n}_e, w\rangle_{\partial\Omega_e} = (f,w)_{\Omega_e} & \forall w\in\mathscr{M},\\ (\tilde{\mathbf{Q}},\tilde{\boldsymbol{v}})_{\Omega_e} + (U,\nabla\cdot\tilde{\boldsymbol{v}})_{\Omega_e} - \langle \widehat{U}, \boldsymbol{v}\cdot\boldsymbol{n}_e\rangle_{\partial\Omega_e} = 0, & \forall\tilde{\boldsymbol{v}}\in(\mathscr{M})^d,\\ (\mathbf{Q},\boldsymbol{v})_{\Omega_e} - (a(U,\tilde{\mathbf{Q}})\tilde{\mathbf{Q}},\boldsymbol{v})_{\Omega_e} = 0 & \forall\boldsymbol{v}\in(\mathscr{M})^d, \end{cases} \tag{16}$$

for every element Ω_e in the domain Ω.

By construction, the approximants $(U,\mathbf{Q},\tilde{\mathbf{Q}})$ may be discontinuous across element boundaries. As a consequence, at a given face ε the functions $(U,\mathbf{Q},\tilde{\mathbf{Q}})$ may be *multi-valued*. This is why the *numerical fluxes* $\widehat{\mathbf{Q}}$ and $\widehat{U}$ are introduced in (16). This issue is clearly explained in the context of elliptic problems in [9] and in [3], and in the context of nonlinear diffusion problems in [7].

For the LDG method that we will analyze and implement, the numerical fluxes are chosen in the following simple way:

$$\widehat{U} = \begin{cases} \{U\} & \text{if } \varepsilon\in\varepsilon_i,\\ B_D & \text{if } \varepsilon\in\Gamma_D,\\ U & \text{if } \varepsilon\in\Gamma_N, \end{cases} \tag{17}$$

and

$$\widehat{\mathbf{Q}} = \begin{cases} \{\mathbf{Q}\} - \sigma[\![U]\!] & \text{if } \varepsilon\in\varepsilon_i,\\ \mathbf{Q} - \sigma(U\boldsymbol{n} - B_D\boldsymbol{n}) & \text{if } \varepsilon\in\Gamma_D,\\ B_N & \text{if } \varepsilon\in\Gamma_N. \end{cases} \tag{18}$$

Note that the numerical flux $\widehat{U}$ does not depend on $\mathbf{Q}$. This makes it possible for the local variable $\mathbf{Q}$ to be solved in terms of U by using the second and third equations of (16). This is a particular property that distinguishes the LDG method (hence the name "local"). The penalty parameter σ appearing in the definition of the numerical fluxes will be chosen in order to enhance the stability and thus, the accuracy of the method.

Remark 2.1. The fluxes defined in (17) and (18) are both *consistent* and *conservative* as defined in [3] and [9].

The resulting LDG formulation is obtained in two steps. First, by summing over all elements Ω_e to find

$$\begin{cases} (U_t,w) + (\mathbf{Q},\nabla w) - \langle\widehat{\mathbf{Q}},[\![w]\!]\rangle_{\varepsilon_i} - \langle\widehat{\mathbf{Q}}\cdot\boldsymbol{n},w\rangle_{\partial\Omega} = (f,w) & \forall w\in\mathscr{M},\\ (\tilde{\mathbf{Q}},\tilde{\boldsymbol{v}}) + (U,\nabla\cdot\tilde{\boldsymbol{v}}) - \langle\widehat{U},[\![\boldsymbol{v}]\!]\rangle_{\varepsilon_i} - \langle\widehat{U},\boldsymbol{v}\cdot\boldsymbol{n}\rangle_{\partial\Omega} = 0 & \forall\tilde{\boldsymbol{v}}\in(\mathscr{M})^d,\\ (a(U,\tilde{\mathbf{Q}})\tilde{\mathbf{Q}},\boldsymbol{v}) - (\mathbf{Q},\boldsymbol{v}) = 0 & \forall\boldsymbol{v}\in(\mathscr{M})^d, \end{cases} \tag{19}$$

where we have denoted with $(\cdot,\cdot) = \sum_e(\cdot,\cdot)_{\Omega_e}$ the sum of all element integrals. And second, by substituting the values of the *numerical fluxes* (17) and (18) in (19)

$$\begin{cases} (U_t,w) + (\mathbf{Q},\nabla w) - \langle\{\mathbf{Q}\},[\![w]\!]\rangle_{\varepsilon_i} + \langle\sigma[\![U]\!],[\![w]\!]\rangle_{\varepsilon_i}\\ \quad -\langle B_N,w\rangle_{\Gamma_N} - \langle\mathbf{Q}\cdot\boldsymbol{n},w\rangle_{\Gamma_D} + \langle\sigma(U-B_D),w\rangle_{\Gamma_D} = (f,w) & \forall w\in\mathscr{M},\\ (\tilde{\mathbf{Q}},\tilde{\boldsymbol{v}}) + (U,\nabla\cdot\tilde{\boldsymbol{v}}) - \langle\{U\},[\![\tilde{\boldsymbol{v}}]\!]\rangle_{\varepsilon_i} - \langle U,\tilde{\boldsymbol{v}}\cdot\boldsymbol{n}\rangle_{\Gamma_N} = \langle B_D,\tilde{\boldsymbol{v}}\cdot\boldsymbol{n}\rangle_{\Gamma_D} & \forall\tilde{\boldsymbol{v}}\in(\mathscr{M})^d,\\ (a(U,\tilde{\mathbf{Q}})\tilde{\mathbf{Q}},\boldsymbol{v}) - (\mathbf{Q},\boldsymbol{v}) = 0 & \forall\boldsymbol{v}\in(\mathscr{M})^d, \end{cases} \tag{20}$$

where, for simplicity, we have denoted with

$$\langle\cdot,\cdot\rangle_{\varepsilon_i} := \sum_e\langle\cdot,\cdot\rangle_{\partial\Omega_e\setminus\Gamma},\quad \langle\cdot,\cdot\rangle_{\Gamma_N} = \sum_e\langle\cdot,\cdot\rangle_{\partial\Omega_e\cap\Gamma_N},\quad \text{and}$$
$$\langle\cdot,\cdot\rangle_{\Gamma_D} = \sum_e\langle\cdot,\cdot\rangle_{\partial\Omega_e\cap\Gamma_D}$$

the sum of the boundary integrals in all interior element boundaries ε_i, in all element boundaries along the Newman boundary Γ_N, and in all element boundaries along the Dirichlet boundary Γ_D, respectively. In order to enforce the initial condition we set

$$(U_0,w) = (u_0,w)\quad \forall w\in\mathscr{M},\quad t=0. \tag{21}$$

Note that using integration by parts for some of the terms in the second equation of (20), the following expression holds,

$$\begin{aligned}(U,\nabla\cdot\tilde{\boldsymbol{v}}) - \langle\{U\},[\![\tilde{\boldsymbol{v}}]\!]\rangle_{\varepsilon_i} - \langle U,\tilde{\boldsymbol{v}}\cdot\boldsymbol{n}\rangle_{\Gamma_N}\\ = -(\nabla U,\boldsymbol{v}) + \langle[\![U]\!],\{\tilde{\boldsymbol{v}}\}\rangle_{\varepsilon_i} + \langle U,\tilde{\boldsymbol{v}}\cdot\boldsymbol{n}\rangle_{\Gamma_D}.\end{aligned} \tag{22}$$

Based on the previous observation we will rewrite the LDG formulation for the IBVP (5) as,

$$\begin{cases} (U_t,w) + (\mathbf{Q},\nabla w) - \langle\{\mathbf{Q}\},[\![w]\!]\rangle_{\varepsilon_i} + \langle\sigma[\![U]\!],[\![w]\!]\rangle_{\varepsilon_i}\\ \quad -\langle B_N,w\rangle_{\Gamma_N} - \langle\mathbf{Q}\cdot\boldsymbol{n},w\rangle_{\Gamma_D} + \langle\sigma(U-B_D),w\rangle_{\Gamma_D} = (f,w) & \forall w\in\mathscr{M},\\ (\tilde{\mathbf{Q}},\tilde{\boldsymbol{v}}) - (\nabla U,\tilde{\boldsymbol{v}}) + \langle[\![U]\!],\{\tilde{\boldsymbol{v}}\}\rangle_{\varepsilon_i} + \langle U,\tilde{\boldsymbol{v}}\cdot\boldsymbol{n}\rangle_{\Gamma_D} = \langle B_D,\tilde{\boldsymbol{v}}\cdot\boldsymbol{n}\rangle_{\Gamma_D} & \forall\tilde{\boldsymbol{v}}\in(\mathscr{M})^d,\\ (a(U,\tilde{\mathbf{Q}})\tilde{\mathbf{Q}},\boldsymbol{v}) - (\mathbf{Q},\boldsymbol{v}) = 0 & \forall\boldsymbol{v}\in(\mathscr{M})^d \end{cases} \tag{23}$$

Remark 2.2. It is clear that any continuous classical solution of problem (12)–(14) will satisfy problem (23) since all terms involving jumps across elements $[\![\cdot]\!]$, will be zero and all boundary terms will satisfy strongly the boundary conditions.

Remark 2.3. As mentioned in Section 1.3, the diffusion coefficient $a(u,\tilde{\boldsymbol{q}})$ in (13) will be approximated by the family of Lipschitz non-degenerate diffusion coefficients of the form

$$a_\epsilon(u,\tilde{\boldsymbol{q}}) = \frac{\beta_\epsilon(u)}{|\tilde{\boldsymbol{q}}|^{1-\gamma}}, \tag{24}$$

and we will denote with $\beta(\cdot)$ any member of the family $\{\beta_\epsilon(\cdot)\}$ in the subsequent analysis to simplify the notation. Furthermore, note that any solution of the IBVP (5) will also be a solution (12)–(14) with the regularized diffusion coefficient (24).

Remark 2.4. It is not difficult to see that, for a given $\sigma > 0$, the system of nonlinear ordinary differential equations arising from (23) will have at least one solution. Indeed, the fact that the right hand side of this system is – at least – locally Hölder continuous with respect to U and each component of $\widetilde{\mathbf{Q}}$ ensures existence of at least one solution. See [25] for a more detailed argument.

2.1. Stability analysis

Even though the proof of Theorem 2.1 can be established as a Corollary of Theorem 2.2, we present it here for clarity. Indeed, many of the mathematical manipulations presented in the proof of Theorem 2.1 can be easily followed and will be used in the more elaborate setting of the proof of Theorem 2.2.

Theorem 2.1 (Stability). *Let U and $\widetilde{\mathbf{Q}}$ be solutions of (23) and (21) with $B_N = 0$, and $B_D = 0$. Then*

$$\|U(t)\|^2_{L^2(\Omega)} + \|\widetilde{\mathbf{Q}}\|^{1+\gamma}_{L^{1+\gamma}(0,T,L^{1+\gamma}(\Omega)} + \int_0^T \left(\|\sigma^{\frac{1}{2}}[\![U]\!]\|^2_{L^2(\varepsilon_i)} + \|\sigma^{\frac{1}{2}}U\|^2_{L^2(\Gamma_D)}\right) \leqslant C\left(\epsilon, \|u_0\|^2_{L^2(\Omega)}, \|f\|^2_{L^2(0,T;L^2(\Omega))}\right). \tag{25}$$

Proof. Note that choosing $w = U$, $\tilde{\boldsymbol{v}} = \mathbf{Q}$, and $\boldsymbol{v} = \widetilde{\mathbf{Q}}$ and adding all terms on the left hand side in (23) we obtain, after several cancellations

$$\frac{1}{2}\frac{\partial}{\partial t}\|U(t)\|^2_{L^2(\Omega)} + \langle \sigma[\![U]\!], [\![U]\!]\rangle_{\varepsilon_i} + \langle \sigma U, U\rangle_{\Gamma_D} + \left(a(U,\widetilde{\mathbf{Q}})\widetilde{\mathbf{Q}}, \widetilde{\mathbf{Q}}\right) = (f,U). \tag{26}$$

From the observation that

$$\epsilon\|\widetilde{\mathbf{Q}}\|^{1+\gamma}_{L^{1+\gamma}(\Omega)} \leqslant \int_\Omega \beta(U)|\widetilde{\mathbf{Q}}|^{1+\gamma} = (a(U,\widetilde{\mathbf{Q}})\widetilde{\mathbf{Q}}, \widetilde{\mathbf{Q}}).$$

Eq. (26) leads to

$$\frac{1}{2}\frac{\partial}{\partial t}\|U(t)\|^2_{L^2(\Omega)} + \|\sigma^{\frac{1}{2}}[\![U]\!]\|^2_{L^2(\varepsilon_i)} + \|\sigma^{\frac{1}{2}}U\|^2_{L^2(\Gamma_D)} + \epsilon\|\widetilde{\mathbf{Q}}\|^{1+\gamma}_{L^{1+\gamma}(\Omega)} \leqslant (f,U). \tag{27}$$

Furthermore, since

$$(f,U) \leqslant \frac{1}{2}\|U(t)\|^2_{L^2(\Omega)} + \frac{1}{2}\|f\|^2_{L^2(\Omega)}.$$

Eq. (27) implies

$$\frac{1}{2}\frac{\partial}{\partial t}\|U(t)\|^2_{L^2(\Omega)} + \|\sigma^{\frac{1}{2}}[\![U]\!]\|^2_{L^2(\varepsilon_i)} + \frac{1}{2}\|\sigma^{\frac{1}{2}}U\|^2_{L^2(\Gamma_D)} + \epsilon\|\widetilde{\mathbf{Q}}\|^{1+\gamma}_{L^{1+\gamma}(\Omega)} \leqslant \frac{1}{2}\|U(t)\|^2_{L^2(\Omega)} + \frac{1}{2}\|f\|^2_{L^2(\Omega)}. \tag{28}$$

Since the second, third, and fourth terms of the left hand side of the previous equation are nonnegative we obtain

$$\frac{1}{2}\frac{\partial}{\partial t}\|U(t)\|^2_{L^2(\Omega)} \leqslant \frac{1}{2}\|U(t)\|^2_{L^2(\Omega)} + \frac{1}{2}\|f\|^2_{L^2(\Omega)}$$

which, by Gronwall's Lemma, leads to

$$\|U(t)\|^2_{L^2(\Omega)} \leqslant C(\|U_0\|^2_{L^2(\Omega)}, \|f\|^2_{L^2(0,T;L^2(\Omega))}) \quad \text{for all } t \in [0,T].$$

Integrating (28) in time from 0 to T, the following must also hold:

$$\|\widetilde{\mathbf{Q}}\|^{1+\gamma}_{L^{1+\gamma}(0,T,L^{1+\gamma}(\Omega)} \leqslant C(\epsilon, \|U_0\|^2_{L^2(\Omega)}, \|f\|^2_{L^2(0,T;L^2(\Omega))}). \tag{29}$$

Likewise for the second and third terms of (28). Finally, by choosing $w = U_0$ in the first equation of (21) we obtain

$$(U_0,U_0) = (u_0,U_0) \leqslant \frac{1}{2}\|U_0\|^2_{L^2(\Omega)} + \frac{1}{2}\|u_0\|^2_{L^2(\Omega)}$$

which implies

$$\|U_0\|^2_{L^2(\Omega)} \leqslant \|u_0\|^2_{L^2(\Omega)}$$

Thus, the result of the Theorem follows at once. □

2.2. Continuous in time a priori error analysis

In this section, we will study how close (possibly nonunique) solutions to the LDG approximation problem (23), U, are to the true weak solution u of problem (12)–(14) with the regularized diffusion coefficient (24).

The analysis requires that comparison functions be carefully chosen.

We define $\hat{u} \in \mathscr{M}$ and $\hat{\boldsymbol{q}} \in \mathscr{M}^d$ to be L^2 projections, given by

$$(u - \hat{u}, v) = 0 \quad \forall\, v \in \mathscr{M} \tag{30}$$

and

$$(\boldsymbol{q} - \hat{\boldsymbol{q}}, \boldsymbol{v}) = 0 \quad \forall \boldsymbol{v} \in \mathscr{M}^d. \tag{31}$$

Furthermore, define $\hat{\tilde{\boldsymbol{q}}} \in \mathscr{M}^d$ by

$$(\hat{\tilde{\boldsymbol{q}}}, \boldsymbol{v}) = (\nabla\hat{u}, \boldsymbol{v}) - \langle [\![\hat{u}]\!], \{\boldsymbol{v}\}\rangle_{\varepsilon_i} - \langle \hat{u} - B_D, \boldsymbol{v}\cdot\boldsymbol{n}\rangle_{\Gamma_D} \quad \forall \boldsymbol{v} \in \mathscr{M}^d. \tag{32}$$

We note that, defining the projection $\pi\tilde{q} \in \mathscr{M}^d$ by

$$(\pi\tilde{q}, \boldsymbol{v}) = (\nabla u, \boldsymbol{v}) \quad \forall \boldsymbol{v} \in \mathscr{M}^d$$

we have

$$(\hat{\tilde{\boldsymbol{q}}} - \pi\tilde{q}, \boldsymbol{v}) = (\nabla(\hat{u} - u), \boldsymbol{v}) - \langle [\![\hat{u} - u]\!], \{\tilde{\boldsymbol{v}}\}\rangle_{\varepsilon_i} - \langle \hat{u} - u, \boldsymbol{v}\cdot\boldsymbol{n}\rangle_{\Gamma_D}. \tag{33}$$

Choosing $\boldsymbol{v} = \hat{\tilde{\boldsymbol{q}}} - \pi\tilde{q}$ and applying the trace theorem and (11) it is easily seen that

$$\|\hat{\tilde{\boldsymbol{q}}} - \pi\tilde{q}\|_{L^2(\Omega)} \leqslant C\sum_e \left[\|\nabla(\hat{u} - u)\|_{L^2(\Omega_e)} + h_e^{-1/2}\|\hat{u} - u\|^{1/2}_{L^2(\Omega_e)}\|\hat{u} - u\|^{1/2}_{H^1(\Omega_e)}\right]. \tag{34}$$

We make the following assumptions on the boundedness of the solution and approximations, namely that there exist positive constants $K_1, \overline{K}_1, K_2$ and $\overline{K}_2$ such that

$$\|u\|_{L^\infty(0,T;L^\infty(\Omega))} \leqslant K_1, \tag{35}$$

$$\|U\|_{L^\infty(0,T;L^\infty(\Omega))} \leqslant \overline{K}_1, \tag{36}$$

$$\|\hat{\tilde{\boldsymbol{q}}}\|_{L^\infty(0,T;L^\infty(\Omega))} \leqslant K_2, \tag{37}$$

$$\|\widetilde{\mathbf{Q}}\|_{L^\infty(0,T;L^\infty(\Omega))} \leqslant \overline{K}_2. \tag{38}$$

We will show in Lemmas 2.1 and 2.2 that the constants in our error estimate are independent of $\overline{K}_1$ and $\overline{K}_2$ provided we choose the maximum diameter of the mesh, h, small enough and the finite element space is of high enough order. We note that (37) holds if ∇u is bounded and $\nabla u - \hat{\tilde{\boldsymbol{q}}}$ is small (or merely bounded).

Theorem 2.2. *Let $(u,\tilde{\boldsymbol{q}},\boldsymbol{q})$ be the solution of problem (12)–(14) and let $(U,\widetilde{\mathbf{Q}},\mathbf{Q})$ be the solution of problem (23). Let $\chi_u = u - \hat{u}, \vec{\chi}_q =$*

$\tilde{\boldsymbol{q}} - \hat{\tilde{\boldsymbol{q}}}$ and $\chi_q = \boldsymbol{q} - \hat{\boldsymbol{q}}$. Furthermore assume (35)–(38) hold. Then for all $t \in [0,T]$, there exists a constant $C = C(\epsilon, \gamma, K_1, \bar{K}_1, K_2, \bar{K}_2, T)$ such that

$$\begin{aligned}
&\|U(t) - u(t)\|^2_{L^2(\Omega)} + \|\tilde{\boldsymbol{Q}} - \tilde{\boldsymbol{q}}\|^2_{L^2(0,T;L^2(\Omega))} \\
&\quad + \int_0^T \left(\|\sigma^{\frac{1}{2}}[\![U(t) - u(t)]\!]\|^2_{L^2(\varepsilon_i)} + \|\sigma^{\frac{1}{2}}(U(t) - u(t))\|^2_{L^2(\Gamma_D)} \right) \\
&\leqslant \|\chi_u(t)\|^2_{L^2(\Omega)} + \|\tilde{\chi}_q\|^2_{L^2(0,T;L^2(\Omega))} + C\left(\|U(0) - u(0)\|^2_{L^2(\Omega)} + \|\chi_u\|^2_{L^2(0,T;L^2(\Omega))} \right) \\
&\quad + C \int_0^T \left(\|\sigma^{-\frac{1}{2}}\{\chi_q\}\|^2_{L^2(\varepsilon_i)} + \|\sigma^{\frac{1}{2}}\chi_u\|^2_{L^2(\Gamma_D)} + \|\sigma^{-\frac{1}{2}}\chi_q\|^2_{L^2(\Gamma_D)} + \|\sigma^{\frac{1}{2}}\{\chi_u\}\|^2_{L^2(\varepsilon_i)} \right) \\
&\quad + C \int_0^T \int_\Omega |\tilde{\chi}_q|^{2\gamma}. \qquad (39)
\end{aligned}$$

Proof. Since the solution u of (12)–(14) satisfies the weak form (23), the following three equations hold:

$$\begin{aligned}
&(U_t - \hat{u}_t, w) + (\boldsymbol{Q} - \hat{\boldsymbol{q}}, \nabla w) - \langle \{\boldsymbol{Q} - \hat{\boldsymbol{q}}\}, [\![w]\!] \rangle_{\varepsilon_i} + \langle \sigma [\![U - \hat{u}]\!], [\![w]\!] \rangle_{\varepsilon_i} \\
&\quad - \langle (\boldsymbol{Q} - \hat{\boldsymbol{q}}) \cdot \boldsymbol{n}, w \rangle_{\Gamma_D} + \langle \sigma(U - \hat{u}), w \rangle_{\Gamma_D} \\
&\quad = (u_t - \hat{u}_t, w) + (\boldsymbol{q} - \hat{\boldsymbol{q}}, \nabla w) - \langle \{\boldsymbol{q} - \hat{\boldsymbol{q}}\}, [\![w]\!] \rangle_{\varepsilon_i} + \langle \sigma [\![u - \hat{u}]\!], [\![w]\!] \rangle_{\varepsilon_i} + \\
&\quad - \langle (\boldsymbol{q} - \hat{\boldsymbol{q}}) \cdot \boldsymbol{n}, w \rangle_{\Gamma_D} + \langle \sigma(u - \hat{u}), w \rangle_{\Gamma_D}, \qquad (40)
\end{aligned}$$

$$\begin{aligned}
&(\tilde{\boldsymbol{Q}} - \hat{\tilde{\boldsymbol{q}}}, \tilde{\boldsymbol{v}}) - (\nabla(U - \hat{u}), \tilde{\boldsymbol{v}}) + \langle [\![U - \hat{u}]\!], \{\tilde{\boldsymbol{v}}\} \rangle_{\varepsilon_i} \\
&\quad + \langle U - \hat{u}, \tilde{\boldsymbol{v}} \cdot \boldsymbol{n} \rangle_{\Gamma_D} = (\tilde{\boldsymbol{q}} - \hat{\tilde{\boldsymbol{q}}}, \tilde{\boldsymbol{v}}) - (\nabla(u - \hat{u}), \tilde{\boldsymbol{v}}) \\
&\quad + \langle [\![u - \hat{u}]\!], \{\tilde{\boldsymbol{v}}\} \rangle_{\varepsilon_i} + \langle B_D - \hat{u}, \tilde{\boldsymbol{v}} \cdot \boldsymbol{n} \rangle_{\Gamma_D}, \qquad (41)
\end{aligned}$$

and

$$\begin{aligned}
&\left(\beta(U) \left(\frac{\tilde{\boldsymbol{Q}}}{|\tilde{\boldsymbol{Q}}|^{1-\gamma}} - \frac{\hat{\tilde{\boldsymbol{q}}}}{|\hat{\tilde{\boldsymbol{q}}}|^{1-\gamma}} \right), \boldsymbol{v} \right) - (\boldsymbol{Q} - \hat{\boldsymbol{q}}, \boldsymbol{v}) \\
&\quad = \left(\beta(u) \left(\frac{\tilde{\boldsymbol{q}}}{|\tilde{\boldsymbol{q}}|^{1-\gamma}} - \frac{\hat{\tilde{\boldsymbol{q}}}}{|\hat{\tilde{\boldsymbol{q}}}|^{1-\gamma}} \right), \boldsymbol{v} \right) - (\boldsymbol{q} - \hat{\boldsymbol{q}}, \boldsymbol{v}) \\
&\quad - \left((\beta(U) - \beta(u)) \frac{\hat{\tilde{\boldsymbol{q}}}}{|\hat{\tilde{\boldsymbol{q}}}|^{1-\gamma}}, \boldsymbol{v} \right). \qquad (42)
\end{aligned}$$

Note that the first and second terms on the right hand side of (40) and the second term on the right hand side of (42) are zero by definition of $\hat{u}$ and $\hat{\boldsymbol{q}}$. Furthermore, by (32), the entire right hand side of (41) vanishes.

To simplify notation, let $\xi_u = U - \hat{u}$, $\tilde{\xi}_q = \tilde{\boldsymbol{Q}} - \hat{\tilde{\boldsymbol{q}}}$ and $\xi_q = \boldsymbol{Q} - \hat{\boldsymbol{q}}$. Now, choosing $w = \xi_u$, $\tilde{\boldsymbol{v}} = \tilde{\xi}_q$, and $\boldsymbol{v} = \xi_q$, and adding Eqs. (40)–(42) we obtain, after multiple cancellations:

$$\begin{aligned}
&\frac{1}{2}\frac{\partial}{\partial t}\|\xi_u(t)\|^2_{L^2(\Omega)} + \langle \sigma [\![\xi_u]\!], [\![\xi_u]\!] \rangle_{\varepsilon_i} + \langle \sigma \xi_u, \xi_u \rangle_{\Gamma_D} \\
&\quad + \left(\beta(U) \left(\frac{\tilde{\boldsymbol{Q}}}{|\tilde{\boldsymbol{Q}}|^{1-\gamma}} - \frac{\hat{\tilde{\boldsymbol{q}}}}{|\hat{\tilde{\boldsymbol{q}}}|^{1-\gamma}} \right), \tilde{\xi}_q \right) \\
&\quad = -\langle \{\chi_q\}, [\![\xi_u]\!] \rangle_{\varepsilon_i} + \langle \sigma [\![\chi_u]\!], [\![\xi_u]\!] \rangle_{\varepsilon_i} - \langle \xi_u, \chi_q \cdot \boldsymbol{n} \rangle_{\Gamma_D} \\
&\quad + \langle \sigma \chi_u, \xi_u \rangle_{\Gamma_D} + \left(\beta(u) \left(\frac{\tilde{\boldsymbol{q}}}{|\tilde{\boldsymbol{q}}|^{1-\gamma}} - \frac{\hat{\tilde{\boldsymbol{q}}}}{|\hat{\tilde{\boldsymbol{q}}}|^{1-\gamma}} \right), \tilde{\xi}_q \right) \\
&\quad - \left((\beta(U) - \beta(u)) \frac{\hat{\tilde{\boldsymbol{q}}}}{|\hat{\tilde{\boldsymbol{q}}}|^{1-\gamma}}, \tilde{\xi}_q \right). \qquad (43)
\end{aligned}$$

Furthermore, from the result of Lemma 1.2 and provided $\beta(U) \geqslant \epsilon > 0$

$$\gamma \epsilon \mathscr{A} |\tilde{\xi}_q|^2_{L^2(\Omega)} \leqslant \left(\beta(U) \left(\frac{\tilde{\boldsymbol{Q}}}{|\tilde{\boldsymbol{Q}}|^{1-\gamma}} - \frac{\hat{\tilde{\boldsymbol{q}}}}{|\hat{\tilde{\boldsymbol{q}}}|^{1-\gamma}} \right), \tilde{\boldsymbol{q}} - \hat{\tilde{\boldsymbol{q}}} \right), \qquad (44)$$

where $\mathscr{A} := \inf_{(0,T)\times\Omega}(\mathscr{A}_0) = 1/(\|\tilde{\boldsymbol{Q}}\|_{L^\infty(0,T,L^\infty(\Omega))} + \|\hat{\tilde{\boldsymbol{q}}}\|_{L^\infty(0,T,L^\infty(\Omega))})^{1-\gamma}$.

Using (44), we can establish that

$$\frac{1}{2}\frac{\partial}{\partial t}\|\xi_u(t)\|^2_{L^2(\Omega)} + \|\sigma^{\frac{1}{2}}[\![\xi_u]\!]\|^2_{L^2(\varepsilon_i)} + \|\sigma^{\frac{1}{2}}\xi_u\|^2_{L^2(\Gamma_D)} + \gamma\epsilon\mathscr{A}\|\tilde{\xi}_q\|^2_{L^2(\Omega)} \leqslant \sum_{i=1}^{6} T_i, \qquad (45)$$

where $T_i, i = 1, \ldots, 6$, are the terms arising from the right hand side of (43). We now proceed to bound the terms $T_1 - T_6$.

For T_1, we multiply and divide by $\sigma^{\frac{1}{2}}$ to get

$$T_1 = \langle \sigma^{-\frac{1}{2}}\{\chi_q\}, \sigma^{\frac{1}{2}}[\![\xi_u]\!] \rangle_{\varepsilon_i} \leqslant \frac{\epsilon_1}{2}\|\sigma^{\frac{1}{2}}[\![\xi_u]\!]\|^2_{L^2(\varepsilon_i)} + \frac{1}{2\epsilon_1}\|\sigma^{-\frac{1}{2}}\{\chi_q\}\|^2_{L^2(\varepsilon_i)}. \qquad (46)$$

For term T_2 we have the following inequality

$$T_2 = \langle \sigma [\![\chi_u]\!], [\![\xi_u]\!] \rangle_{\varepsilon_i} \leqslant \frac{1}{2}\|\sigma^{\frac{1}{2}}[\![\chi_u]\!]\|^2_{L^2(\varepsilon_i)} + \frac{1}{2}\|\sigma^{\frac{1}{2}}[\![\xi_u]\!]\|^2_{L^2(\varepsilon_i)}. \qquad (47)$$

The terms T_3 and T_4 are handled identically to T_1 and T_2 with ε_i replaced by Γ_D.

As for terms T_5 and T_6, note that

$$\begin{aligned}
T_5 &= \int_\Omega \beta(u) \left(\frac{\tilde{\boldsymbol{q}}}{|\tilde{\boldsymbol{q}}|^{1-\gamma}} - \frac{\hat{\tilde{\boldsymbol{q}}}}{|\hat{\tilde{\boldsymbol{q}}}|^{1-\gamma}} \right) \tilde{\xi}_q \leqslant \frac{2}{\gamma} C_1 \int_\Omega |\tilde{\chi}_q|^\gamma |\tilde{\xi}_q| \\
&\leqslant \frac{2}{\gamma} C_1 \left(\frac{1}{2\epsilon_2} \int_\Omega |\tilde{\chi}_q|^{2\gamma} + \frac{\epsilon_2}{2}\|\tilde{\xi}_q\|^2_{L^2(\Omega)} \right), \qquad (48)
\end{aligned}$$

where $C_1 = C(K_1)$, and

$$\begin{aligned}
T_6 &= \left((\beta(U) - \beta(u)) \frac{\hat{\tilde{\boldsymbol{q}}}}{|\hat{\tilde{\boldsymbol{q}}}|^{1-\gamma}}, \tilde{\xi}_q \right) \leqslant C_2 \int_\Omega |u - U| \hat{\tilde{\boldsymbol{q}}}|^\gamma |\tilde{\xi}_q| \\
&\leqslant C_2 \|\hat{\tilde{\boldsymbol{q}}}\|^\gamma_{L^\infty(\Omega)} \|u - U\|_{L^2(\Omega)} \|\tilde{\xi}_q\|_{L^2(\Omega)} \\
&\leqslant C_2 \|\hat{\tilde{\boldsymbol{q}}}\|^\gamma_{L^\infty(\Omega)} \left(\frac{1}{2\epsilon_3} (\|\chi_u\|^2_{L^2(\Omega)} + \|\xi_u\|^2_{L^2(\Omega)}) + \frac{\epsilon_3}{2}\|\tilde{\xi}_q\|^2_{L^2(\Omega)} \right), \qquad (49)
\end{aligned}$$

where $C_2 = C(K_1, \bar{K}_1)$. From (45)–(49) and choosing ϵ_1, ϵ_2 and ϵ_3 small enough so that for $\bar{\epsilon}$ and ϵ^* small positive numbers, $0 < \bar{\epsilon} \leqslant \gamma\epsilon\mathscr{A} - \left(\frac{1}{\gamma} C_1 \epsilon_2 + \frac{1}{2} C_2 \|\hat{\tilde{\boldsymbol{q}}}\|^\gamma_{L^\infty(\Omega)} \epsilon_3 \right)$ and $0 < \epsilon^* \leqslant \frac{1}{2}(1 - \epsilon_1)$, we obtain

$$\begin{aligned}
&\frac{1}{2}\frac{\partial}{\partial t}\|\xi_u(t)\|^2_{L^2(\Omega)} + \epsilon^*\|\sigma^{\frac{1}{2}}[\![\xi_u]\!]\|^2_{L^2(\varepsilon_i)} + \frac{1}{2}\|\sigma^{\frac{1}{2}}\xi_u\|^2_{L^2(\Gamma_D)} + \bar{\epsilon}\|\tilde{\xi}_q\|^2_{L^2(\Omega)} \\
&\quad \leqslant C\Big(\|\xi_u\|^2_{L^2(\Omega)} + \|\chi_u\|^2_{L^2(\Omega)} + \|\sigma^{1/2}[\![\chi_u]\!]\|^2_{\varepsilon_i} + \|\sigma^{-\frac{1}{2}}\chi_q \cdot \boldsymbol{n}\|^2_{L^2(\Gamma_D)} \\
&\qquad + \|\sigma^{-\frac{1}{2}}\{\chi_q\}\|^2_{L^2(\varepsilon_i)} + \|\sigma^{\frac{1}{2}}\chi_u\|^2_{L^2(\Gamma_D)} + \int_\Omega |\tilde{\chi}_q|^{2\gamma} \Big). \qquad (50)
\end{aligned}$$

Since the second, third, and fourth terms of the left hand side in the previous inequality are nonnegative, we can use Gronwall's Lemma to find that for all $t \in [0,T]$,

$$\begin{aligned}
\|\xi_u(t)\|^2_{L^2(\Omega)} \leqslant C\Big[&\|\xi_u(0)\|^2_{L^2(\Omega)} + \|\chi_u\|^2_{L^2(0,T,L^2(\Omega))} + \int_0^T \int_\Omega |\tilde{\chi}_q|^{2\gamma} \\
&+ \int_0^T \left(\|\sigma^{-\frac{1}{2}}\{\chi_q\}\|^2_{L^2(\varepsilon_i)} + \|\sigma^{\frac{1}{2}}\chi_u\|^2_{L^2(\Gamma_D)} \right. \\
&\left. + \|\sigma^{1/2}[\![\chi_u]\!]\|^2_{\varepsilon_i} + \|\sigma^{-\frac{1}{2}}\chi_q \cdot \boldsymbol{n}\|^2_{L^2(\Gamma_D)} \right) \Big]. \qquad (51)
\end{aligned}$$

Likewise, integrating (50) again in time from 0 to T, we can establish the boundedness of the three remaining terms of the left hand side of (50):

$$\int_0^T \left(\|\sigma^{\frac{1}{2}}[\![\xi_u]\!]\|^2_{L^2(\varepsilon_i)} + \|\sigma^{\frac{1}{2}}\xi_u\|^2_{L^2(\Gamma_D)} \right) \quad \text{and} \quad \|\tilde{\xi}_q\|^2_{L^2(0,T,L^2(\Omega))}. \qquad (52)$$

The result of the Theorem follows immediately from the triangle inequality. □

Corollary 2.1. *If $u(t), \tilde{\boldsymbol{q}}(t), \boldsymbol{q}(t)$ are sufficiently smooth for $0 < t \leqslant T$, and the approximations $U, \tilde{\boldsymbol{Q}}, \boldsymbol{Q}$ are constructed with piecewise polynomials of degree at most k and satisfy the assumptions of Theorem 2.2, then for all $t \in [0,T]$ and $\sigma = \mathscr{O}(\frac{1}{h})$, there exists a constant*

$$C = C(\epsilon, \gamma, \Omega, T, K_1, \overline{K}_1, K_2, \overline{K}_2, \|u\|_{L^\infty(0,T;H^{k+1}(\Omega))}, \|\boldsymbol{q}\|_{L^\infty(0,T;H^k(\Omega))}) \quad (53)$$

such that

$$\|U(t)-u(t)\|_{L^2(\Omega)} + \|\tilde{\mathbf{Q}}(t)-\tilde{\boldsymbol{q}}(t)\|_{L^2(0,T;L^2(\Omega))} + \int_0^T \frac{1}{h}\left(\|[\![U(t)-u(t)]\!]\|_{L^2(\varepsilon_i)} + \|(U(t)-u(t))\|_{L^2(\Gamma_D)}\right) \leqslant Ch^{k\gamma}. \quad (54)$$

Proof. The corollary follows from Theorem 2.2 and the following bounds.

$$\|\chi_u(t)\|^2_{L^2(\Omega)} \leqslant Ch^{2(k+1)}\|u\|^2_{H^{k+1}(\Omega)}, \quad (55)$$

$$\|\tilde{\boldsymbol{\chi}}_q\|^2_{L^2(\Omega)^2} \leqslant Ch^{2k}\|u\|^2_{H^{k+1}(\Omega)}, \quad (56)$$

$$\|\chi_u\|^2_{L^2(0,T,L^2(\Omega))} \leqslant Ch^{2(k+1)}\int_0^T \|u\|^2_{H^{k+1}(\Omega)}. \quad (57)$$

Using the trace inequality,

$$\|\sigma^{\frac{1}{2}}[\![\chi_u]\!]\|^2_{L^2(\varepsilon_i)} + \|\sigma^{\frac{1}{2}}\chi_u\|^2_{L^2(\Gamma_D)} \leqslant Ch^{-1}\sum_e \|\chi_u\|_{L^2(\Omega_e)}\|\chi_u\|_{H^1(\Omega_e)} \leqslant Ch^{2k}\|u\|^2_{H^{k+1}(\Omega)}. \quad (58)$$

Note also that

$$\int_\Omega |\tilde{\boldsymbol{\chi}}_q|^{2\gamma} \leqslant \left(\int_\Omega |\tilde{\boldsymbol{\chi}}_q|^2\right)^{\frac{2\gamma}{2}} |\Omega|^{1-\gamma} = \|\tilde{\boldsymbol{\chi}}_q\|^{2\gamma}_{L^2(\Omega)}|\Omega|^{1-\gamma} \leqslant Ch^{2k\gamma}\|u\|^{2\gamma}_{H^{k+1}(\Omega)} \quad (59)$$

and

$$\|\sigma^{-\frac{1}{2}}\{\boldsymbol{\chi}_q\}\|^2_{L^2(\varepsilon_i)} + \|\sigma^{-\frac{1}{2}}\{\boldsymbol{\chi}_q\}\|^2_{L^2(\Gamma_D)} \leqslant Ch\sum_e \|\boldsymbol{q}-\hat{\boldsymbol{q}}\|_{L^2(\Omega_e)}\|\boldsymbol{q}-\hat{\boldsymbol{q}}\|_{H^1(\Omega_e)} \leqslant Ch^{2k}\|\boldsymbol{q}\|_{H^k(\Omega)}. \quad (60)$$

The result of the Corollary follows immediately. □

Lemma 2.1 (Boundedness of the approximation). *Under the assumptions of Theorem 2.2, choosing $\gamma \geqslant 1/2$, and provided h is sufficiently small and $k \geqslant 3$, if $\|u\|_{L^\infty(0,T,L^\infty(\Omega))} \leqslant K_1$, then $\|U\|_{L^\infty(0,T,L^\infty(\Omega))} \leqslant K_1(1+\epsilon_{k_1})$ for a small parameter ϵ_{k_1}.*

Proof. Clearly

$$\|U\|_{L^\infty(0,T,L^\infty(\Omega))} \leqslant \|U-u\|_{L^\infty(0,T,L^\infty(\Omega))} + \|u\|_{L^\infty(0,T,L^\infty(\Omega))}. \quad (61)$$

From Corollary 2.1 and Lemma 1.4 we obtain

$$\|U\|_{L^\infty(0,T;L^\infty(\Omega))} \leqslant \|U-u\|_{L^\infty(0,T;L^\infty(\Omega))} + K_1 \leqslant Ch^{-1}\|U-\hat{u}\|_{L^\infty(0,T;L^2(\Omega))} + \|\hat{u}-u\|_{L^\infty(0,T;L^\infty(\Omega))} + K_1 \leqslant C(h^{3\gamma-1}+h^3) + K_1$$

Thus, we can choose a sufficiently small h so that $Ch^{1/2} \leqslant \epsilon_{k_1}K_1$, which implies

$$\|U\|_{L^\infty(0,T,L^\infty(\Omega))} \leqslant K_1(1+\epsilon_{k_1}) \qquad \square$$

Lemma 2.2 (Boundedness of the gradient of the approximation). *Under the assumptions of Theorem 2.2, choosing $\gamma \geqslant 1/2 + \lambda/4$ with $2 \geqslant \lambda > 0$, and provided h is sufficiently small and $k \geqslant 4$, if $\|\nabla u\|_{L^\infty(0,T,L^\infty(\Omega))} \leqslant K_3$, then $\|\tilde{\mathbf{Q}}\|_{L^\infty(0,T,L^\infty(\Omega))} \leqslant K_3(1+\epsilon_{k_2})$ for a small parameter ϵ_{k_2}.*

Proof. Returning to (41), for any $0 < t < T$,

$$(\tilde{\mathbf{Q}}-\hat{\tilde{\boldsymbol{q}}}, \tilde{\boldsymbol{v}}) = (\nabla(U-\hat{u}), \tilde{\boldsymbol{v}}) - \langle[\![U-\hat{u}]\!], \{\tilde{\boldsymbol{v}}\}\rangle_{\varepsilon_i} - \langle U-\hat{u}, \tilde{\boldsymbol{v}}\cdot\boldsymbol{n}\rangle_{\Gamma_D}. \quad (62)$$

Setting $\tilde{\boldsymbol{v}} = \tilde{\boldsymbol{\chi}}_q$, and using trace and inverse inequalities

$$\|\tilde{\boldsymbol{\chi}}_q\|^2_{L^2(\Omega)} = (\nabla\xi_u, \tilde{\boldsymbol{\chi}}_q) - \langle[\![\xi_u]\!], \{\tilde{\boldsymbol{\chi}}_q\}\rangle_{\varepsilon_i} - \langle\xi_u, \tilde{\boldsymbol{\chi}}_q\cdot\boldsymbol{n}\rangle_{\Gamma_D} \leqslant \|\nabla\xi_u\|_{L^2(\Omega)}\|\tilde{\boldsymbol{\chi}}_q\|_{L^2(\Omega)} + \|\xi_u\|_{L^2(\varepsilon_i)}\|\tilde{\boldsymbol{\chi}}_q\|_{L^2(\varepsilon_i)} + \|\xi_u\|_{L^2(\Gamma_D)}\|\tilde{\boldsymbol{\chi}}_q\|_{L^2(\Gamma_D)} \leqslant Ch^{-1}\|\xi_u\|_{L^2(\Omega)}\|\tilde{\boldsymbol{\chi}}_q\|_{L^2(\Omega)}.$$

Therefore,

$$\|\tilde{\boldsymbol{\chi}}_q\|_{L^2(\Omega)} \leqslant Ch^{k\gamma-1}. \quad (63)$$

Now following the argument used in the proof of Lemma 2.1, we see that the result follows if $k \geqslant 4$ and h is small enough so that $Ch^{4\gamma-2} \leqslant Ch^\lambda \leqslant \epsilon_{k_2}K_2$. □

3. Numerical experiments: 2D

In this section, we investigate numerically, the order of accuracy of the proposed LDG method. We also present results of some numerical experiments aimed at solving two ideal 2D problems: a dam break event, and flow in a channel with vegetation resulting from a dam break event. The main motivation to show the latter results is to provide the reader with convincing evidence that the DSW equation captures the physics of the aforementioned ideal problems. In fact, the setting of the simulated flow in a channel with vegetation was inspired by an actual experiment shown in [5].

The 2D LDG finite element formulation on unstructured triangular elements was coded in order to carry out the numerical experiments. A second-order backward difference formula (BDF) time integrator was used to solve the problem forward in time. Picard iteration was used to linearize the resulting nonlinear system, and the conjugate gradient method was used to solve the resulting linear systems.

Table 1
Convergence rates to approximate Barenblatt solutions for $\alpha = 5/3$ and $\gamma = 1/2$ using $t_0 = 2$ and $t_f = 2.1$ and $\Omega = [-2, 2]$.

dt	h	$\|U-u\|_{L^2(\Omega)}$	Conv. rate
1/100	1/2	2.81×10^{-3}	–
1/400	1/4	6.76×10^{-4}	2.06
1/800	1/8	1.58×10^{-4}	2.10
1/1600	1/16	3.68×10^{-5}	2.10

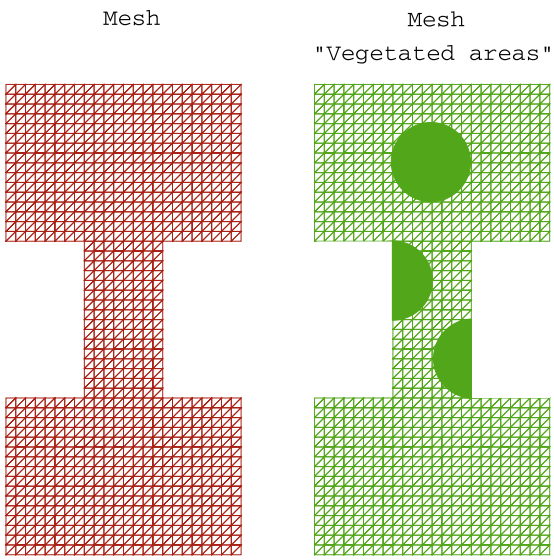

Fig. 1. Mesh for the dam break simulation (left) without vegetation (right) with vegetation.

3.1. Numerical convergence

In order to verify the accuracy of the implemented 2D LDG scheme, we chose to reproduce an analytic Barenblatt solution to the DSW equation for a flat bathymetry ($z(x,y)=0$). The explicit expression for such solution $u(x,t)$ in 1D (spatially) is presented in [16] and used in [25] to numerically investigate the convergence rates of a one dimensional CG scheme. We extended this analytic solutions to 2D (spatially) by simply by setting $u(x,y,t)=u(x,t)$ where

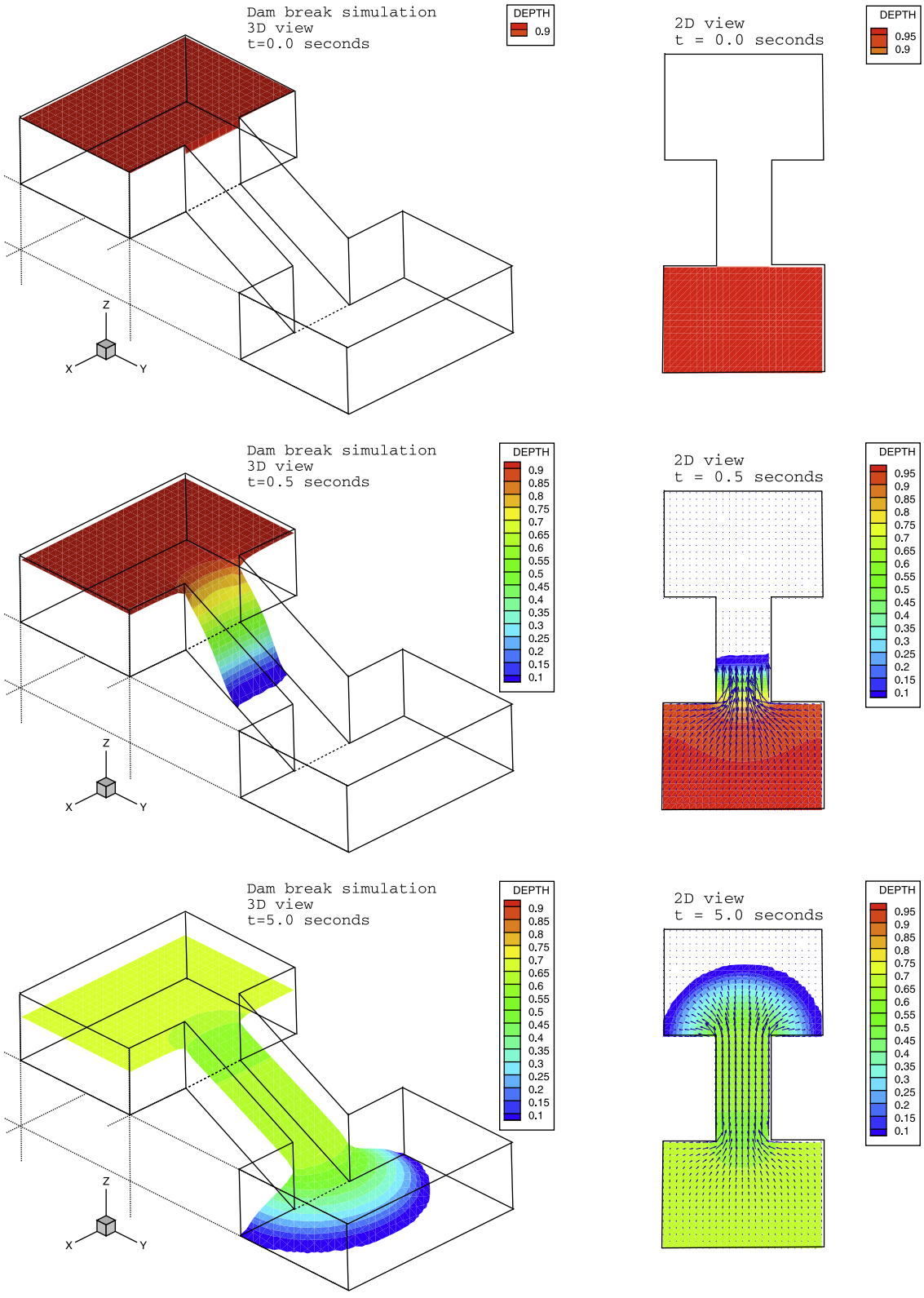

Fig. 2. Dam break simulation. Figures showing evolution of water depth (meters) at times = 0.0, 0.5, and 5 s. (left) 3D views, (right) 2D views.

$$u(x,t) = t^{-\frac{1}{\gamma(m+1)}}[C - k(m,\gamma)|\Phi|^{\frac{\gamma+1}{\gamma}}]_+^{\frac{\gamma}{m\gamma-1}}, \tag{64}$$

$$M = \int_{-\infty}^{\infty} u(x,t)dx,$$

where $[s(x)]_+$ denotes the positive part of $s(x)$, $m = 1 + \alpha/\gamma$, C is a positive function related to the initial mass M, given by

$$k(m,\gamma) = \frac{m\gamma - 1}{m(\gamma+1)}\left(\frac{1}{\gamma(m+1)}\right)^{\frac{1}{\gamma}}, \quad \text{and} \quad \Phi = x t^{-\frac{1}{\gamma(m+1)}}.$$

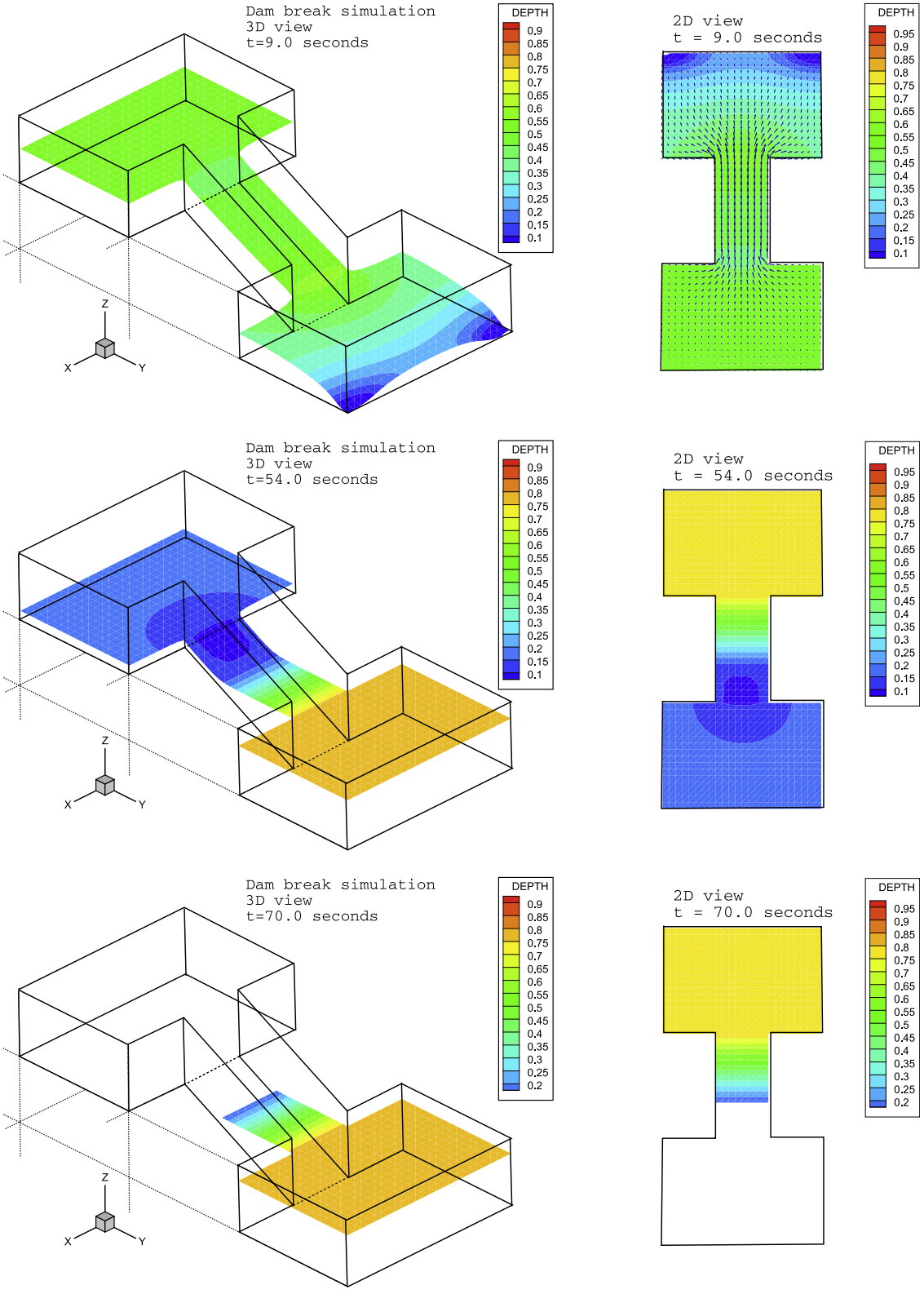

Fig. 3. Dam break simulation. Figures showing evolution of water depth (meters) at times = 9.0, 54.0, and 70.0 s. (left) 3D views, (right) 2D views.

We used our 2D code to reproduce this solution on the domain $\Omega = [-2,2] \times [-0.5,0.5]$, for the time interval $t \in [2,2.1]$, for $\alpha = 5/3$ and $\gamma = 1/2$. We prescribed the appropriate Dirichlet boundary conditions ($U(-2,y,t) = u(-2,t)$ and $U(2,y,t) = u(2,t)$) on the boundaries $x = -2$ and $x = 2$, and zero-Newmann boundary conditions on the boundaries $y = -0.5$ and $y = 0.5$. We restricted our error analysis to a numerical domain Ω such that u is nondegenerate ($u > 0$) everywhere for our simulation time, $t \in [t_i, t_f]$. The results are shown in Table 1. Note that the function given by (64) is Lipschitz continuous and compactly supported, in particular, its gradient is **bounded** and continuous in our cylinder $\Omega \times [t_0, t_f]$.

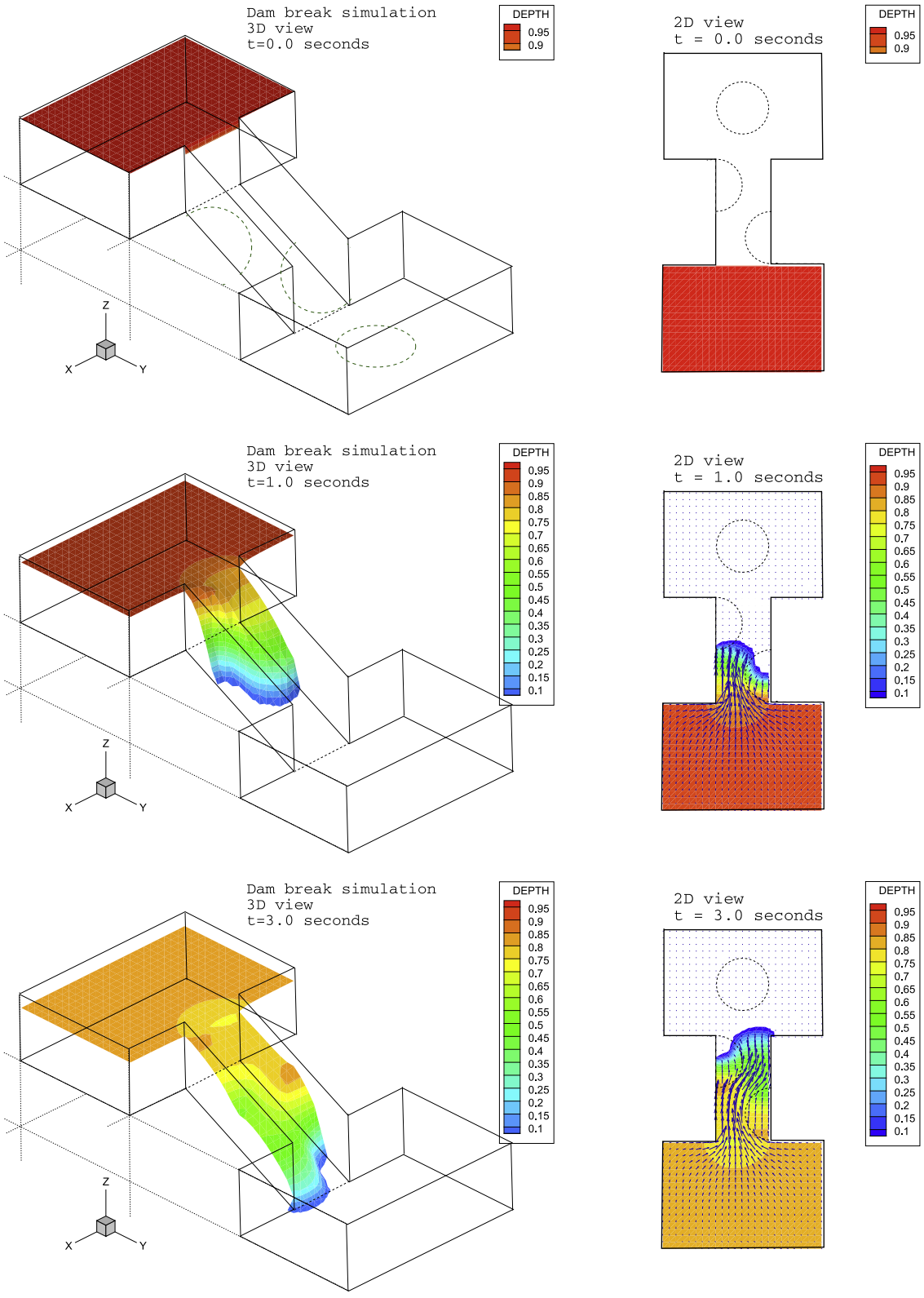

Fig. 4. Dam break simulation with vegetation. Figures showing evolution of water depth (meters) at times = 0, 1.0, and 3.0 s. (left) 3D views, (right) 2D views.

Remark 3.1. Assuming that the BDF integrator gives rise to order Δt^2 errors, where Δt is the time step, we chose to push the limits in our investigation to see if we could observe optimal convergence (h^2 for piecewise linear elements), despite the fact that Corollary 2.1 suggests convergence results of the type, $\|u(t_n) - U^n\|_{L^2(\Omega)} \leqslant C(u, t_n)(\Delta t^2 + h^{1/2})$ for $\gamma = 1/2$, when approximating nondegenerate solutions $u \in H^2(\Omega)$, using piecewise linear basis functions ($k = 1$). The previous motivation lead us to chose the time step much smaller than the grid diameter in our convergence experiments.

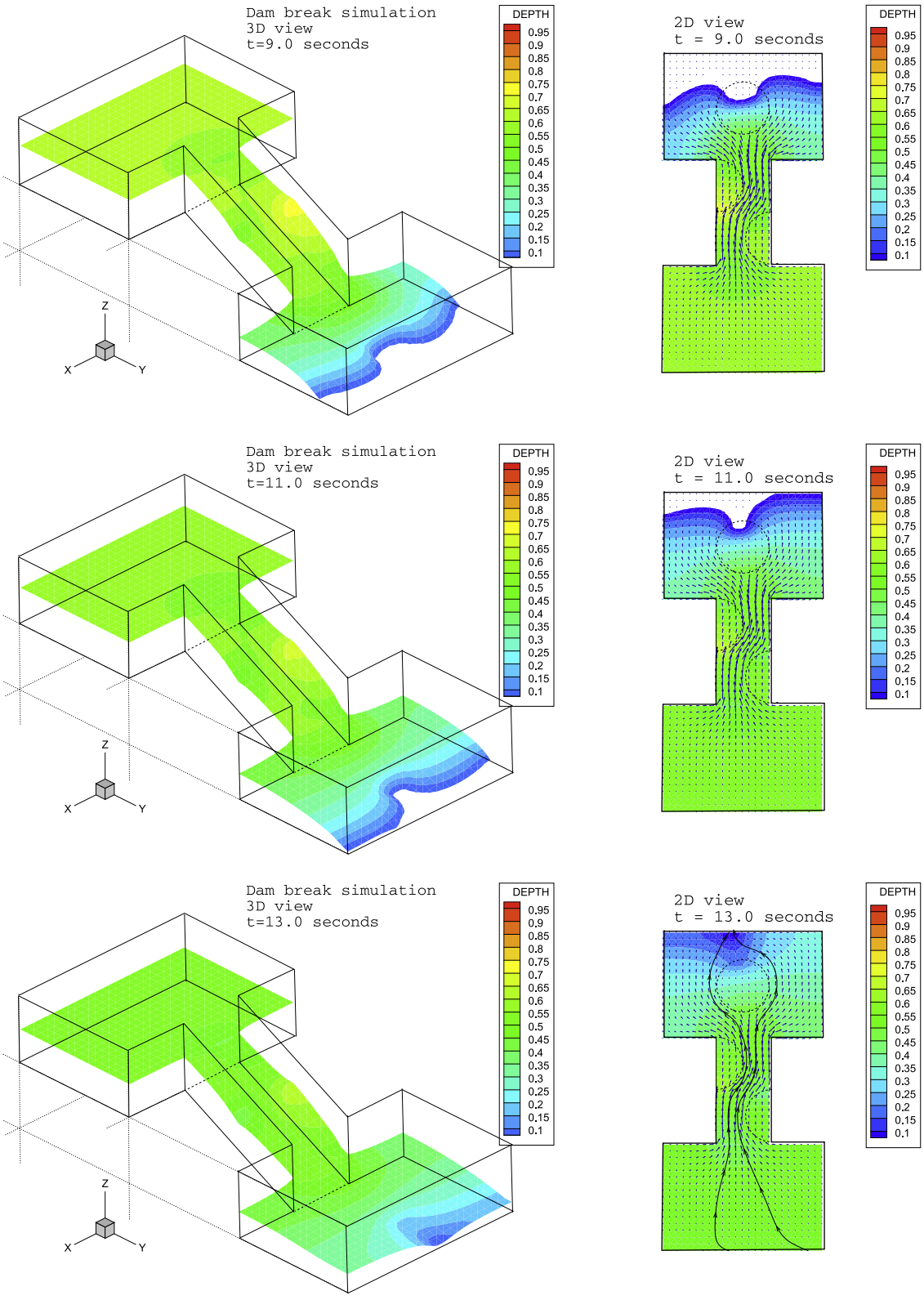

Fig. 5. Dam break simulation with vegetation. Figures showing evolution of water depth (meters) at times = 9.0, 11.0, and 13.0 s. (left) 3D views, (right) 2D views.

The convergence rates shown in Table 1 show that the error estimates obtained in Corollary 2.1 for $\gamma = 1/2$ are very conservative. Corollary 2.1, loosely speaking, suggests that the error decreases as $\mathcal{O}(h^{1/2})$ for piecewise linear basis functions, yet in practice, we observe $\mathcal{O}(h^2)$ convergence. This is not necessarily surprising since removing the degeneracy in the IBVP (1) gives rise to a presumably well-behaved parabolic problem, where optimal convergence rates – such as the ones observed in the numerical experiments – could, in principle, be achieved.

3.2. The Dam break problem

In this section, we present the results of 2D simulations of the evolution of water depth profiles in an ideal dam break problem. This problem consists of simulating the water flow resulting from removing an ideal dam that keeps water on a confined area of the domain. The set up is as follows, a channel was designed to connect two reservoirs, one completely filled with water (up hill) and the other completely empty (down hill). The channel is considered to be dry at the beginning as well. When the ideal dam is removed from the upper reservoir, water is expected to flow down hill, flooding first the channel with a well defined front, and later flooding the lower reservoir; first with a well defined and radially symmetric front, and later filling it gradually. This process is expected to continue until all the water is transferred fully to the lower reservoir.

The units used in this ideal setting were meters for the water depth and height, and seconds for the time. This numerical experiment was computed in a domain with a uniform friction coefficient $c_f = 1$ (this value was chosen for simplicity and without any physical meaning) and with *zero* Newmann boundary conditions on $\partial\Omega$. The mesh of the computational domain is shown in Fig. 1 (left), the initial condition and water bed of this problem are presented at the top left of Fig. 2. The – wet condition – parameter, introduced in Sections 1.1 and 1.3, was chosen to be $\epsilon = .01$ to provide stability in the code. Recall that the typical depth in the domain is $\mathcal{O}(1)$. The mesh radius is of the order $h \sim 0.125$ m (in a domain with characteristic lengths of order $L \sim 6$ m and $W \sim 3$ m, respectively), and the time step was comparable in size, *i.e.* $\Delta t = 0.125$ s. The experiment was run from $t = 0.0$ s to $t = 70.0$ s. 3D and 2D views of the numerically simulated evolution of the water depth are presented in Figs. 2 and 3.

As discussed before, the main features of the phenomenon are captured, these include: (1) The down-hill flow of water, (2) the appearance of a flooding wave with a well defined front propagating in the direction of lowest potential energy points (lowest points in space), see Figs. 2 and 3, (3) the radial symmetry of the water flow both, at the entrance of the channel (uphill) as well as at the exit of the channel (down hill) throughout the event, (4) the radial symmetry in the flooding front when reaching the lowest reservoir, see upper views of Fig. 3, (this is a consequence of the previous observation), and (5) the eventually gradual transfer of water from the upper part to the lowest one.

Some of the characteristics of the phenomenon that are *not* captured are mostly related to two factors: the diffusive nature of the DSW equation, and the vertical integration utilized to derive it. Related to the first factor, the physical interaction of the water flow with the walls is not captured. For example, when water flows in a confined channel, ripples form as a consequence of momentum transfers between the water and the walls (as well as friction). Also, when water frontally hits a wall (as it happens in the lower views of Fig. 3) water sloshes and forms reflecting waves. These features are not present in the experiments we show. Another obvious characteristic *not* captured with the DSW equation as a model, and related to the second factor, is the vertical velocity profile of the water flow.

3.3. The Dam Break problem with vegetation

In this section, we present the results of 2D simulations of the evolution of water depth profiles in an ideal dam break problem with vegetation in some regions of the domain. This problem was inspired by the experimental setting shown in [5]. The numerical implementation was set up similarly to the one presented in Section 3.2. The main difference consists of including three islands of vegetation in different locations of the domain. These vegetated regions, considered to have the same vegetation density, modify the water flow lines in the experimental setting of [5]. It is observed, as intuition would suggest, that water flows more rapidly away from them. Their inclusion in the numerical simulations is done only by assigning a higher value of the friction coefficient c_f inside these areas. Throughout the domain $c_f = 1$ and in the vegetated regions $c_f = 5$. The bathymetry remained the same as well as all the remaining computational variables presented in the dam break problem in Section 3.2. Again, we chose to simulate the water flow resulting from removing an ideal dam that keeps water on a confined (uphill) area of the domain. The mesh for this problem and the location of the islands of vegetation are shown in Fig. 1 (right). This experiment was run from $t = 0.0$ to $t = 70.0$ as well. However, since the most relevant features of this event take place before $t = 20.0$, only views for $t \in [0, 20]$ are presented. 3D and 2D views of the numerically simulated evolution of the water depth are presented in Figs. 4 and 5.

Figs. 4 and 5 show very good agreement with the expected features of the phenomenon. In particular, they clearly display the fact that, as expected, water flows more rapidly away from the vegetated areas. Also, the flooding front propagates throughout the domain in a way that qualitatively captures the expected dynamics. Again, the limitations of the DSW equation as a model appear as described in the previous section.

4. Conclusions

In this study, we prove that nondegenerate approximate solutions to the DSW equation, obtained using the LDG method, converge to true solutions of such equation, provided the true solution is sufficiently smooth. We show that for discontinuous finite elements of polynomial order k (with $k \geqslant 4$, see Corollary 2.1), one can ensure convergence $\mathcal{O}(h^{k\gamma})$. Numerical experiments in 2D show that the theoretical convergence rate obtained in Corollary 2.1 – for a nondegenerate true solution $u \in C^{\infty}(\Omega, t)$ of the DSW equation, and $\gamma = 1/2$ – is conservative. Indeed, in this case we observe numerical convergence rates $\mathcal{O}(h^2)$ for piecewise linear finite elements ($k = 1$).

We also present numerical experiments aimed at showing the qualitative characteristics of water flow captured by the DSW equation when used as a model to simulate an idealized dam break problem with vegetation. The numerical experiments show very good agreement with the expected features of the phenomenon.

Acknowledgements

This work was supported in part by the National Science Foundation, Project Nos. DMS-0411413, and DMS-0620697, and Centro de Investigación en Geografía y Geomática,"Ing. Jorge L. Tamayo", A.C. The conclusion of this manuscript was possible due to the generous Henson Environmental Fellowship awarded to Mauricio Santillana at the Harvard University Center for the Environment.

References

[1] V. Aizinger, C. Dawson, B. Cockburn, P. Castillo, The local discontinuous Galerkin method for contaminant transport, Adv. Water Resources 24 (1) (2000) 73–87.

[2] R. Alonso, M. Santillana, C. Dawson, On the diffusive wave approximation of the shallow water equations, Eur. J. Appl. Math. 19 (5) (2008) 575–606. October.
[3] Douglas N. Arnold, Franco Brezzi, Bernardo Cockburn, L. Donatella Marini, Unified analysis of discontinuous Galerkin methods for elliptic problems, SIAM J. Numer. Anal. 39 (5) (2002) 1749–1779.
[4] F. Bassi, S. Rebay, A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier–Stokes equations, J. Comput. Phys. 131 (2) (1997) 267–279.
[5] S.J. Bennett, T. Pirim, B.D. Barkdoll, Using simulated emergent vegetation to alter stream flow direction within a straight experimental channel, Geomorphology 44 (2002) 115–126.
[6] Susanne C. Brenner, Ridgway Scott, The Mathematical Theory of Finite Element Methods, Springer-Verlag, New York, 1994.
[7] R. Bustinza, G.N. Gatica, A local discontinuous Galerkin method for nonlinear diffusion problems with mixed boundary conditions, SIAM J. Sci. Comput. 26 (1) (2004) 152–177.
[8] R. Bustinza, G.N. Gatica, A mixed local discontinuous Galerkin method for a class of nonlinear problems in fluid mechanics, J. Comput. Phys. 207 (2) (2005) 427–456.
[9] P. Castillo, B. Cockburn, I. Perugia, D. Schötzau, An a priori error analysis of the local discontinuous Galerkin method for elliptic problems, SIAM J. Numer. Anal. 38 (5) (2000) 1676–1706.
[10] P. Castillo, B. Cockburn, D. Schötzau, C. Schwab, Optimal a priori error estimates for the hp-version of the local discontinuous Galerkin method for convection–diffusion problems, Math. Comput. (2002).
[11] P.G. Ciarlet, Finite Element Method for Elliptic Problems, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2002.
[12] B. Cockburn, C. Dawson, Some extensions of the local discontinuous Galerkin method for convection–diffusion equations, in: The Proceedings of the Conference on the Mathematics of Finite Elements and Applications, Elsevier, 2000, pp. 225–238.
[13] B. Cockburn, G.E. Karniadakis, C.W. Shu (Eds.), Discontinuous Galerkin Methods, Springer, 2000.
[14] B. Cockburn, C.W. Shu, The local discontinuous Galerkin method for time-dependent convection–diffusion systems, SIAM J. Numer. Anal. 35 (6) (1998) 2440–2463.
[15] J. Douglas, T.F. Dupont, Galerkin methods for parabolic equations, SIAM J. Numer. Anal. 7 (1970) 575–626.
[16] J.R. Esteban, J.L. Vázquez, Homogeneous diffusion in $\mathbb{R}$ with power-like nonlinear diffusivity, Archive Rat. Mech. Anal. 103 (1988) 39–80.
[17] K. Feng, F.J. Molz, A 2-d diffusion based wetland flow model, J. Hydrol. 196 (1997) 230–250.
[18] P. Di Giammarco, E. Todini, P. Lamberti, A conservative finite elements approach to overland flow: the control volume finite element formulation, J. Hydrol. 175 (1996) 267–291.
[19] P. Houston, J. Robson, E. Süli, Discontinuous Galerkin finite element approximation of quasilinear elliptic boundary value problems i: the scalar case, IMA J. Numer. Anal. 25 (4) (2005) 726–749.
[20] T.V. Hromadka, C.E. Berenbrock, J.R. Freckleton, G.L. Guymon, A two-dimensional dam-break flood plain model, Adv. Water Resources (1985) 8.
[21] E.J. Kubatko, J.J. Westerink, C. Dawson, *hp* Discontinuous Galerkin methods for advection-dominated problems in shallow water, Comput. Methods Appl. Mech. Engrg. 196 (2006) 437–451.
[22] A. Lasis, E. Süli, *hp*-Version discontinuous Galerkin finite element method for semilinear parabolic problems, SIAM J. Numer. Anal. 45 (4) (2007) 1544–1569.
[23] H. Lia, M.W. Farthinga, C.N. Dawson, C.T. Miller, Local discontinuous Galerkin approximations to richards equation, Adv. Water Resources 30 (2007) 555–575.
[24] C. Ortner, E. Süli, Discontinuous Galerkin finite element approximation of nonlinear second-order elliptic and hyperbolic systems, SIAM J. Numer. Anal. 45 (4) (2007) 1370–1397.
[25] M. Santillana, C. Dawson, A numerical approach to study the properties of solutions of the diffusive wave approximation of the shallow water equations, Comput. Geosci., February 2009.
[26] Vidar Thomée, Galerkin finite element methods for parabolic problems, Springer Series in Computational Mathematics, vol. 25, Springer, 1997.
[27] J.L. Vázquez, The Porous Medium Equation. Mathematical Theory, Oxford University Press, USA, 2006.
[28] M.F. Wheeler, A priori L^2 error estimates for Galerkin approximations to parabolic partial differential equations, SIAM J. Numer. Anal. 10 (4) (1973) 723–759.
[29] Th. Xanthopoulos, Ch. Koutitas, Numerical simulation of a two dimensional flood wave propagation due to dam failure, J. Hydraul. Res. 14 (4) (1976) 321–331.
[30] Yan Xu, Chi-Wang Shu, Error estimates of the semi-discrete local discontinuous Galerkin method for nonlinear convection–diffusion and kdv equations, Comput. Methods Appl. Mech. Engrg. 196 (37–40) (2007) 3805–3822. Special Issue Honoring the 80th Birthday of Professor Ivo Babuska.
[31] J. Yan, C.W. Shu, A local discontinuous Galerkin method for kdv type equations, SIAM J. Numer. Anal. 40 (2) (2002) 769–791.
[32] J. Yan, C.W. Shu, Local discontinuous Galerkin methods for partial differential equations with higher order derivatives, J. Sci. Comput. 17 (1–4) (2002) 27–47.
[33] W. Zhang, T.W. Cundy, Modeling of two-dimensional overland flow, Water Resources Res. 25 (1989) 2019–2035.