

Discontinuous Galerkin finite element methods for hyperbolic nonconservative partial differential equations

S. Rhebergen*, O. Bokhove and J.J.W. van der Vegt

Department of Applied Mathematics, University of Twente, P.O. Box 217, 7500 AE, Enschede, The Netherlands

Abstract

We present space- and space-time discontinuous Galerkin finite element (DGFEM) formulations for systems containing nonconservative products, such as occur in dispersed multiphase flow equations. The main criterium we pose on the formulation is that if the system of nonconservative partial differential equations can be transformed into conservative form, then the formulation must reduce to that for conservative systems. Standard DGFEM formulations cannot be applied to nonconservative systems of partial differential equations. We therefore introduce the theory of weak solutions for nonconservative products into the DGFEM formulation leading to the new question how to define the path connecting left and right states across a discontinuity. The effect of different paths on the numerical solution is investigated and found to be small. We also introduce a new numerical flux that is able to deal with nonconservative products. Our scheme is applied to two different systems of partial differential equations. First, we consider the shallow water equations, where topography leads to nonconservative products, in which the known, possibly discontinuous, topography is formally taken as an unknown in the system. Second, we consider a simplification of a depth-averaged two-phase flow model which contains more intrinsic nonconservative products.

Key words: nonconservative products, discontinuous Galerkin finite element methods, numerical fluxes, arbitrary Lagrangian Eulerian (ALE) formulation, two-phase flows

PACS: 02.60.Cb, 02.70.Dh, 47.55.-t, 47.85.Dh

1991 MSC: 35L60, 35L65, 35L67, 65M60, 76M10

* Corresponding author.

Email addresses: rhebergens@math.utwente.nl (S. Rhebergen),
o.bokhove@math.utwente.nl (O. Bokhove),
j.j.w.vandervegt@math.utwente.nl (J.J.W. van der Vegt).

1 Introduction

Systems of equations containing nonconservative products cannot be transformed into divergence form, i.e., equations of the form $\partial_t u + g(u)\partial_x u = 0$ cannot be written as $\partial_t u + \partial_x f(u) = 0$. This causes problems once the solution becomes discontinuous, because the weak solution in the classical sense of distributions then does not exist. Consequently, no classical Rankine-Hugoniot shock conditions can be defined. To overcome these problems we use the theory of Dal Maso, LeFloch and Murat (DLM) [3] for nonconservative products. In this theory a definition is given for nonconservative products $g(u)\partial_x u$, where $g : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a smooth function, but $u :]a, b[\rightarrow \mathbb{R}^m$ may admit discontinuities. Using this theory, a notion of a weak solution can be given to the Riemann problem for nonconservative hyperbolic partial differential equations. A problem with this theory is, however, the introduction of a path in phase space connecting the left and right state across a discontinuity. It is possible to derive an expression for this the path by constructing entropy solutions to the hyperbolic equations (see LeFloch [9]), but that construction can be a very difficult as well as costly job. In this article we will investigate therefore also the influence of this path in phase space and propose a new discontinuous Galerkin finite element method (DGFEM) suitable for hyperbolic partial differential equations in nonconservative form.

We are particularly interested in solving dispersed two-phase two-fluid models. The use of a DG method for these problems is of interest because it can deal efficiently with unstructured and deforming grids, local mesh refinement (h -adaptation), adjustment of the polynomial order in each element (p -refinement), and parallel computation. These benefits stem from the very compact stencil used in DG methods. Dispersed two-phase two-fluid models contain, however, nonconservative products which are introduced in the governing equations in the modeling procedure [4,5]. This poses serious problems and at present there is no literature available how to deal with nonconservative products in a DGFEM context, which motivated the research discussed in this article.

Over the years several authors have been developing numerical methods suitable for nonconservative hyperbolic partial differential equations with non-smooth solutions. Toumi [15] introduced a generalized Roe solver based on the DLM theory, which was later applied by Toumi and Kumbaro [16] to shock tube problems and two-fluid problems. The work by Toumi [15] was also used by Parés [10], Castro, Gallardo and Parés [2] and Parés and Castro [11] to develop numerical schemes in the finite volume context. An alternative approach is followed by Saurel and Abgrall [13] in which the DLM theory is not used. They apply the criterium in multi-fluid flows, where the phases are separated by well-defined interfaces, that if pressure and velocity are uniform in both

fluids, these variables must remain uniform during their temporal evolution (in the absence of surface tension). Using this criterium they construct a Godunov scheme for the conservative part of the system. The nonconservative part is then adjusted to meet the criterium above. They also use this criterium for dispersed two phase flows, where the interfaces are not well-defined; in this case their approach therefore seems less valid.

Here we will use the DLM theory in a DGFEM context. This work differs from the previously mentioned work in that we do not formulate a weak formulation based on generalized Roe solvers. Instead, we present and use a new numerical flux in the context of the DLM theory.

The outline of this article is as follows. We first summarize the main theory of weak solutions for partial differential equations in nonconservative form as proposed by Dal Maso, LeFloch and Murat [3] in Section 2, but in space-time. Using this theory we derive the space-time DGFEM formulation in Section 3 and state the space DGFEM formulation as a special case in Appendix A. In DGFEM methods, the numerical flux plays an essential role. In Section 4 we derive therefore the numerical flux for systems with nonconservative products (NCP flux) which can also be applied to moving grids. In Section 5 we apply DGFEM to two nonconservative hyperbolic systems and show numerical results using a linear path in phase space. The effect of different paths in phase space on the numerical solution is investigated in Section 6. Finally, conclusions are drawn in Section 7.

2 Nonconservative hyperbolic partial differential equations

The main topic of this article is the derivation of a formulation for DGFEM suitable for nonlinear hyperbolic partial differential equations in nonconservative form and the numerical investigation of these systems. We use the DLM theory to overcome the absence of a weak solution in the classical sense of distributions for these types of equations. In an article by Dal Maso, LeFloch and Murat [3], a definition was given for nonconservative products of the form $g(u)\partial_x u$, where $g : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is a smooth function, but $u :]a, b[\rightarrow \mathbb{R}^m$ may admit discontinuities. They assumed u to be a function of bounded variation (BV), viz. a Lebesgue integrable function whose first derivative is a bounded Borel measure, and the product $g(u)\partial_x u$ is defined as a Borel measure on $]a, b[$. Such a definition is necessary when g is not the differential of a smooth function f , i.e., there is no f such that $g(u)\partial_x u$ admits a conservative form $\partial_x f$. The following example, given by LeFloch [9], illustrates the DLM theory.

Consider the function $u(x)$ composed of two constant vectors u_L and u_R in

\mathbb{R}^m with $u_L \neq u_R$:

$$u(x) = u_L + H(x - x_d)(u_R - u_L), \quad x \in]a, b[, \quad (1)$$

where $x_d \in]a, b[$ and $H : \mathbb{R} \rightarrow \mathbb{R}$ is the Heaviside function with $H(x) = 0$ if $x < 0$ and $H(x) = 1$ if $x > 0$. Consider any smooth function $g : \mathbb{R}^m \rightarrow \mathbb{R}^m$. We see immediately that $g(u)\partial_x u$ is not defined at $x = x_d$ since here $|\partial_x u| \rightarrow \infty$. Dal Maso, LeFloch and Murat [3] introduce therefore a smooth regularization u^ε of the discontinuous function u . They show that in this particular case, if the total variation of u^ε remains uniformly bounded with respect to ε :

$$g(u) \frac{du}{dx} \equiv \lim_{\varepsilon \rightarrow 0} g(u^\varepsilon) \frac{du^\varepsilon}{dx}$$

gives a sense to the nonconservative product as a bounded measure. This limit, however, depends on how we choose u^ε . Introduce a Lipschitz continuous path $\phi : [0, 1] \rightarrow \mathbb{R}^m$, satisfying $\phi(0) = u_L$ and $\phi(1) = u_R$, connecting u_L and u_R in \mathbb{R}^m . The following regularization u^ε for u then emerges:

$$u^\varepsilon(x) = \begin{cases} u_L, & \text{if } x \in]a, x_d - \varepsilon[\\ \phi\left(\frac{x - x_d + \varepsilon}{2\varepsilon}\right), & \text{if } x \in]x_d - \varepsilon, x_d + \varepsilon[\\ u_R, & \text{if } x \in]x_d + \varepsilon, b[\end{cases} \quad \varepsilon > 0. \quad (2)$$

Using this regularization, LeFloch [9] states that when ε tends to zero, then:

$$g(u^\varepsilon) \frac{du^\varepsilon}{dx} \rightarrow C \delta_{x_d},$$

vaguely in the sense of measures on $]a, b[$, where δ_{x_d} is the Dirac measure at x_d and the scalar C is given by:

$$C = \int_0^1 g(\phi(\tau)) \frac{d\phi}{d\tau}(\tau) d\tau.$$

We see that the limit of $g(u^\varepsilon)\partial_x u^\varepsilon$ depends on ϕ . There is one exception, namely if an $f : \mathbb{R}^m \rightarrow \mathbb{R}$ exists with $g = \partial_u f$. In this case:

$$C = f(u_R) - f(u_L).$$

We are, however, interested in the case when such a function f does not exist. We see then that the definition of the nonconservative product $g(u)\partial_x u$ must depend on the path ϕ chosen in the regularization. In Section 6, we will investigate the effect of different paths ϕ on the numerical solution. For now, assume that the path ϕ is given. In Dal Maso, LeFloch and Murat [3] it is assumed that the path belongs to a fixed family of paths in \mathbb{R}^m . These paths are Lipschitz continuous maps $\phi : [0, 1] \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ which satisfy the following properties:

- (H1) $\phi(0; u_L, u_R) = u_L$, $\phi(1; u_L, u_R) = u_R$,
(H2) $\phi(\tau; u_L, u_L) = u_L$,
(H3) $\left| \frac{\partial \phi}{\partial \tau}(\tau; u_L, u_R) \right| \leq K |u_L - u_R|$.

Dal Maso, LeFloch and Murat [3] consider functions $u :]a, b[\rightarrow \mathbb{R}^m$ of bounded variation, viz. $u \in BV(]a, b[, \mathbb{R}^m)$. These are functions of $L^1(]a, b[, \mathbb{R}^m)$ whose first order derivative is a bounded Borel measure on the interval $]a, b[$. Since u is BV, u admits a countable set of discontinuity points and at each such point x_d , a left trace $u_L = \lim_{\varepsilon \downarrow 0} u(x_d - \varepsilon)$ and a right trace $u_R = \lim_{\varepsilon \downarrow 0} u(x_d + \varepsilon)$ exist. For more on Borel measures, BV functions and related topics, see, e.g., [19].

Based on the family of paths satisfying (H1)-(H3), the following theorem is given by Dal Maso, LeFloch and Murat [3]:

Theorem 1 *Let $u :]a, b[\rightarrow \mathbb{R}^m$ be a function of bounded variation and $g : \mathbb{R}^m \rightarrow \mathbb{R}^m$ be a continuous function. Then, there exists a unique real-valued bounded Borel measure μ on $]a, b[$ characterized by the two following properties:*

- (1) *If u is continuous on a Borel set $B \subset]a, b[$, then:*

$$\mu(B) = \int_B g(u) \frac{du}{dx} d\lambda,$$

where λ is the Borel measure.

- (2) *If u is discontinuous at a point x_d of $]a, b[$, then:*

$$\mu(\{x_d\}) = \int_0^1 g(\phi(\tau; u_L, u_R)) \frac{\partial \phi}{\partial \tau}(\tau; u_L, u_R) d\tau.$$

By definition, this measure μ is the nonconservative product of $g(u)$ by $\partial_x u$ and is denoted by:

$$\mu = \left[g(u) \frac{du}{dx} \right]_{\phi}.$$

In this article we will derive a space-time DGFEM weak formulation for non-linear hyperbolic systems of partial differential equations in nonconservative form in multi-dimensions:

$$U_{i,0} + G_{ikr} U_{r,k} = 0, \quad \bar{x} \in \mathbb{R}^q, \quad t > 0, \quad (3)$$

with $U \in \mathbb{R}^m$, $G \in \mathbb{R}^m \times \mathbb{R}^q \times \mathbb{R}^m$; we use the comma notation to denote partial differentiation and the summation convention on repeated indices. Here, $(\cdot)_{,0}$ denotes partial differentiation with respect to time and $(\cdot)_{,k}$ ($k = 1, \dots, q$) partial differentiation with respect to the spatial coordinates. In a space-time context, space and time variables are, however, not explicitly distinguished. A point at time $t = x_0$ with position $\bar{x} = (x_1, x_2, \dots, x_q)$ has Cartesian coordinates

$x = (x_0, \bar{x}) \in \mathbb{R}^{q+1}$. We can write (3) then as:

$$T_{ikr}U_{r,k} = 0, \quad x \in \mathbb{R}^{q+1}, \quad x_0 > 0, \quad (4)$$

with $U \in \mathbb{R}^m$ and $T \in \mathbb{R}^m \times \mathbb{R}^{q+1} \times \mathbb{R}^m$ given by:

$$T_{ikr} = \begin{cases} \delta_{ir}, & \text{if } k = 0, \\ G_{ikr}, & \text{otherwise,} \end{cases} \quad (5)$$

where δ represents the Kronecker delta symbol. Dal Maso, LeFloch and Murat [3] give a similar theorem to Theorem 1 for the nonconservative term $T_{ikr}U_{r,k}$ in multi-dimensions. As before, assume a given family of Lipschitz continuous paths $\phi : [0, 1] \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ that satisfy, for some $K > 0$ and for all $U^L, U^R \in \mathbb{R}^m$ and $\tau \in [0, 1]$, the properties:

- (H1) $\phi_r(0; U^L, U^R) = U_r^L$, $\phi_r(1; U^L, U^R) = U_r^R$,
- (H2) $\phi_r(\tau; U^L, U^L) = U_r^L$,
- (H3) $\left| \frac{\partial \phi_r}{\partial \tau}(\tau; U^L, U^R) \right| \leq K |U_r^L - U_r^R|$,
- (H4) $\phi_r(\tau; U^L, U^R) = \phi_r(1 - \tau; U^R, U^L)$.

Note that we have added an extra property, H4, that does not have to be satisfied in the one dimensional case. Let $\Omega \subset \mathbb{R}^{q+1}$ with $\Omega = \Omega_u \cup S_u \cup I_u$ where Ω_u is the set of points of approximate continuity, S_u the set of points of approximate jump and I_u contains the irregular points. The DLM theorem then states:

Theorem 2 *Let $U : \Omega \rightarrow \mathbb{R}^m$ be a bounded function of bounded variation defined on an open subset Ω of \mathbb{R}^{q+1} and $T : \mathbb{R}^m \rightarrow \mathbb{R}^m$ be a locally bounded Borel function. Then there exists a unique family of real-valued bounded Borel measures μ_i on Ω , $i = 1, 2, \dots, m$ such that*

- (1) *if B is a Borel subset of Ω_u , then:*

$$\mu_i(B) = \int_B T_{ikr}U_{r,k} d\lambda, \quad (6)$$

where λ is the Borel measure;

- (2) *if B is a Borel subset of S_u , then:*

$$\mu_i(B) = \int_{B \cap S_u} \int_0^1 T_{ikr}(\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau}(\tau; U^L, U^R) d\tau n_k^L dH^q, \quad (7)$$

with U^L and U^R the left and right traces at the discontinuity, where H^q denotes the q -dimensional Hausdorff measure and where we choose n^L the outward normal with respect to the left state;

- (3) *if B is a Borel subset of I_u , then:*

$$\mu_i(B) = 0. \quad (8)$$

The measure μ_i is the nonconservative product of T_{ikr} by $U_{r,k}$ and is denoted by:

$$\mu_i = \left[T_{ikr} U_{r,k} \right]_{\phi}. \quad (9)$$

In particular, a piecewise \mathcal{C}^1 function U is a weak solution of (4) if and only if the following two conditions are satisfied [2]:

- (1) U is a classical solution in the domains where it is \mathcal{C}^1 .
- (2) Along a discontinuity U satisfies the generalized Rankine-Hugoniot jump condition:

$$-\sigma(U_i^R - U_i^L) + \int_0^1 G_{ikr}(\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau}(\tau; U^L, U^R) d\tau \bar{n}_k^L = 0, \quad (10)$$

where σ is the speed of propagation of the discontinuity, U^L and U^R are the left and right limits of the solution at the discontinuity and \bar{n}^L is the space component of the space-time normal n^L (see e.g. LeFloch [9]).

Note that when $G(U)$ is the Jacobian of some flux function $F(U)$ the jump condition (10) is independent of the family of paths and reduces to the usual Rankine-Hugoniot condition:

$$F_{ik}(U^R) \bar{n}_k^L - F_{ik}(U^L) \bar{n}_k^L = \sigma(U_i^R - U_i^L). \quad (11)$$

3 Space-time DGFEM discretization

In this section we will introduce the formulation for space-time DGFEM for systems of hyperbolic partial differential equations containing nonconservative products. We will start by introducing space-time elements, function spaces, trace operators and basis functions, after which we derive the space-time DG formulation. In Appendix A we also give the formulation for space DGFEM.

3.1 Space-time elements

In the space-time DGFEM method, the space and time variables are not distinguished. A point at time $t = x_0$ with position vector $\bar{x} = (x_1, x_2, \dots, x_q)$ has Cartesian coordinates (x_0, \bar{x}) in the open domain $\mathcal{E} \subset \mathbb{R}^{q+1}$. At time t , the flow domain $\Omega(t)$ is defined as:

$$\Omega(t) := \{\bar{x} \in \mathbb{R}^q : (t, \bar{x}) \in \mathcal{E}\}.$$

By taking t_0 and T as the initial and final time of the evolution of the space-time flow domain, the space-time domain boundary $\partial\mathcal{E}$ consists of the hyper-surfaces:

$$\begin{aligned}\Omega(t_0) &:= \{x \in \mathcal{E} : x_0 = t_0\}, \\ \Omega(T) &:= \{x \in \mathcal{E} : x_0 = T\}, \\ \mathcal{Q} &:= \{x \in \partial\mathcal{E} : t_0 < x_0 < T\}.\end{aligned}$$

The time interval $[t_0, T]$ is partitioned using the time levels $t_0 < t_1 < \dots < T$, where the n -th time interval is defined as $I_n = (t_n, t_{n+1})$ with length $\Delta t_n = t_{n+1} - t_n$. The space-time domain \mathcal{E} is then divided into N_t space-time slabs $\mathcal{E}^n = \mathcal{E} \cap I_n$. Each space-time slab \mathcal{E}^n is bounded by $\Omega(t_n)$, $\Omega(t_{n+1})$ and $\mathcal{Q}^n = \partial\mathcal{E}^n / (\Omega(t_n) \cup \Omega(t_{n+1}))$.

The flow domain $\Omega(t_n)$ is approximated by $\Omega_h(t_n)$, where $\Omega_h(t) \rightarrow \Omega(t)$ as $h \rightarrow 0$, with h the radius of the smallest sphere completely containing the largest space-time element. The domain $\Omega_h(t_n)$ is divided into N_n non-overlapping spatial elements $K_j(t_n)$. Similarly, $\Omega(t_{n+1})$ is approximated by $\Omega_h(t_{n+1})$. We can relate each element $K_j^n = K_j(t_n)$ to a master element $\hat{K} \subset \mathbb{R}^q$ through the mapping F_K^n :

$$F_K^n : \hat{K} \rightarrow K_j^n : \bar{\xi} \mapsto \bar{x} = \sum_i x_i(K_j^n) \chi_i(\bar{\xi})$$

with x_i the spatial coordinates of the vertices of the spatial element K_j^n and χ_i the standard Lagrangian shape functions defined on element \hat{K} . The space-time elements \mathcal{K}_j^n are constructed by connecting K_j^n with K_j^{n+1} using linear interpolation in time, resulting in the mapping $G_{\mathcal{K}}^n$ from the master element $\hat{\mathcal{K}} \subset \mathbb{R}^{q+1}$ to the space-time element \mathcal{K}^n :

$$\begin{aligned}G_{\mathcal{K}}^n : \hat{\mathcal{K}} \rightarrow \mathcal{K}^n : \xi \mapsto (t, \bar{x}) &= \left(\frac{1}{2}(t_{n+1} + t_n) + \frac{1}{2}(t_{n+1} - t_n)\xi_0, \right. \\ &\quad \left. \frac{1}{2}(1 - \xi_0)F_K^n(\bar{\xi}) + \frac{1}{2}(1 + \xi_0)F_K^{n+1}(\bar{\xi}) \right).\end{aligned}$$

The tessellation \mathcal{T}_h^n of the space-time slab \mathcal{E}_h^n consists of all space-time elements \mathcal{K}_j^n ; thus the tessellation \mathcal{T}_h of the discrete flow domain $\mathcal{E}_h := \cup_{n=0}^{N_t-1} \mathcal{E}_h^n$ then is defined as $\mathcal{T}_h := \cup_{n=0}^{N_t-1} \mathcal{T}_h^n$.

The element boundary $\partial\mathcal{K}_j^n$, which is the union of open faces of \mathcal{K}_j^n , consists of three parts: $K_j(t_n^+) = \lim_{\epsilon \downarrow 0} K_j(t_n + \epsilon)$, $K_j(t_{n+1}^-) = \lim_{\epsilon \downarrow 0} K_j(t_{n+1} - \epsilon)$ and $\mathcal{Q}_j^n = \partial\mathcal{K}_j^n / (K_j(t_n^+) \cup K_j(t_{n+1}^-))$. Define the grid velocity $v \in \mathbb{R}^q$ as $v = \Delta\bar{x} / \Delta t$. The outward space-time normal vector at an element boundary point on $\partial\mathcal{K}_j^n$ is given by:

$$n = \begin{cases} (1, \bar{0}) & \text{at } K_j(t_{n+1}^-), \\ (-1, \bar{0}) & \text{at } K_j(t_n^+), \\ (-v_k \bar{n}_k, \bar{n}) & \text{at } \mathcal{Q}_j^n, \end{cases} \quad (12)$$

where $\bar{0} \in \mathbb{R}^q$. Note that since the space-time normal vector n has length one, the space component of the space-time normal \bar{n} has a length $|\bar{n}| = 1/\sqrt{1 + v \cdot v}$. It can be convenient to split the element boundaries into separate faces. In addition to the faces $K_j(t_n^+)$ and $K_j(t_{n+1}^-)$, we also define therefore interior and boundary faces. An interior face is shared by two neighboring elements \mathcal{K}_i^n and \mathcal{K}_j^n , such that $\mathcal{S}_{ij}^n = \mathcal{Q}_i^n \cap \mathcal{Q}_j^n$, and a boundary face is defined as $\mathcal{S}_{Bj}^n = \partial\mathcal{E}^n \cap \mathcal{Q}_j^n$. The set of interior faces in time slab I^n is denoted by \mathcal{S}_I^n and the set of all boundary faces by \mathcal{S}_B^n . The total set of faces is denoted by $\mathcal{S}_{I,B}^n = \mathcal{S}_I^n \cup \mathcal{S}_B^n$.

3.2 Function spaces and trace operators

We consider approximations of $U(x, t)$ and functions $V(x, t)$ in the finite element space V_h , which is defined as:

$$V_h = \left\{ V \in (L^2(\mathcal{E}_h))^m : V|_{\mathcal{K}} \circ G_{\mathcal{K}} \in (P^p(\hat{\mathcal{K}}))^m, \forall \mathcal{K} \in \mathcal{T}_h \right\},$$

where $L^2(\mathcal{E}_h)$ is the space of square integrable functions on \mathcal{E}_h and $P^p(\hat{\mathcal{K}})$ denotes the space of polynomials of degree at most p on the reference element $\hat{\mathcal{K}}$. Here m denotes the dimension of U .

We now introduce some operators as defined in Klaij et al. [8]. The trace of a function $f \in V_h$ at the element boundary $\partial\mathcal{K}^L$ is defined as:

$$f^L = \lim_{\epsilon \downarrow 0} f(x - \epsilon n^L),$$

with n^L the unit outward space-time normal at $\partial\mathcal{K}^L$. When only the space components of the outward normal vector are considered we will use the notation \bar{n}^L . A function $f \in V_h$ has a double valued trace at element boundaries $\partial\mathcal{K}$. The traces of a function f at an internal face $\mathcal{S} = \bar{\mathcal{K}}^L \cap \bar{\mathcal{K}}^R$ are denoted by f^L and f^R . The jump of f at an internal face $\mathcal{S} \in \mathcal{S}_I^n$ in the direction k of a Cartesian coordinate system is defined as:

$$[[f]]_k = f^L \bar{n}_k^L + f^R \bar{n}_k^R,$$

with $\bar{n}_k^R = -\bar{n}_k^L$. The average of f at $\mathcal{S} \in \mathcal{S}_I^n$ is defined as:

$$\{\{f\}\} = \frac{1}{2}(f^L + f^R).$$

The jump operator satisfies the following product rule at $\mathcal{S} \in \mathcal{S}_I^n$ for $\forall g \in V_h$ and $\forall f \in V_h$, which can be proven by direct verification:

$$[[g_i f_{ik}]]_k = \{\{g_i\}\} [[f_{ik}]]_k + [[g_i]]_k \{\{f_{ik}\}\}. \quad (13)$$

Consequently, we can relate element boundary integrals to face integrals:

$$\sum_{\mathcal{K} \in \mathcal{T}_h^n} \int_{\mathcal{Q}} g_i^L f_{ik}^L \bar{n}_k^L d\mathcal{Q} = \sum_{\mathcal{S} \in \mathcal{S}_I^n} \int_{\mathcal{S}} [g_i f_{ik}]_k d\mathcal{S} + \sum_{\mathcal{S} \in \mathcal{S}_B^n} \int_{\mathcal{S}} g_i^L f_{ik}^L \bar{n}_k^L d\mathcal{S}. \quad (14)$$

3.3 Basis functions

Polynomial approximations for the trial function U and the test functions V in each element $\mathcal{K} \in \mathcal{T}_h^n$ are introduced as:

$$U(t, \bar{x})|_{\mathcal{K}} = \hat{U}_m \psi_m(t, \bar{x}) \text{ and } V(t, \bar{x})|_{\mathcal{K}} = \hat{V}_l \psi_l(t, \bar{x}),$$

with ψ_m the basis functions, $\bar{x} \in \mathbb{R}^q$, and expansion coefficients \hat{U}_m and \hat{V}_l , respectively, for $m, l = 0, 1, 2, \dots, N$, where N depends on the polynomial degree of the basis functions in V_h and the space dimension q . The basis functions are defined such that the test and trial functions can be split into an element mean at time t_{n+1} and a fluctuating part. The basis functions ψ_m are given by:

$$\psi_m = \begin{cases} 1, & \text{for } m = 0 \\ \varphi_m(t, \bar{x}) - \frac{1}{|K_j(t_{n+1}^-)|} \int_{K_j(t_{n+1}^-)} \varphi_m(t, \bar{x}) dK & \text{for } m = 1, 2, \dots, N \end{cases}$$

where the functions $\varphi_m(x)$ in element \mathcal{K} are related to the basis functions $\hat{\varphi}_m(\xi)$, with $\hat{\varphi}_m(\xi) \in P^p(\hat{\mathcal{K}})$ and ξ the local coordinates in the master element $\hat{\mathcal{K}}$, through the mapping $G_{\mathcal{K}}$:

$$\varphi_m = \hat{\varphi}_m \circ G_{\mathcal{K}}^{-1}.$$

3.4 Weak formulation

In this section we derive a space-time DGFEM weak formulation for equations containing nonconservative products. Before discussing the space-time DGFEM weak formulation for equations containing nonconservative products, we first introduce as a reference the space-time DGFEM weak formulation for equations in conservative form (see, e.g., van der Vegt and van der Ven [18]).

Consider partial differential equations in conservative form:

$$U_{i,0} + F_{ik,k} = 0, \quad \bar{x} \in \mathbb{R}^q, \quad x_0 > 0, \quad (15)$$

where $U \in \mathbb{R}^m$ and $F \in \mathbb{R}^m \times \mathbb{R}^q$. Using the approach discussed in van der Vegt and van der Ven [18], the space-time DG formulation for (15) can be

stated as:

Find a $U \in V_h$ such that for all $V \in V_h$:

$$\begin{aligned}
0 = & - \sum_{\mathcal{K} \in \mathcal{T}_h^n} \int_{\mathcal{K}} \left(V_{i,0} U_i + V_{i,k} F_{ik} \right) d\mathcal{K} \\
& + \sum_{\mathcal{K} \in \mathcal{T}_h^n} \left(\int_{K(t_{n+1}^-)} V_i^L U_i^L dK - \int_{K(t_n^+)} V_i^L U_i^L dK \right) \\
& + \sum_{\mathcal{S} \in \mathcal{S}_I^n} \int_{\mathcal{S}} \llbracket V_i \rrbracket_k \{ F_{ik} - v_k U_i \} d\mathcal{S} + \sum_{\mathcal{S} \in \mathcal{S}_B^n} \int_{\mathcal{S}} V_i^L (F_{ik}^L - v_k U_i^L) \bar{n}_k^L d\mathcal{S}.
\end{aligned} \tag{16}$$

Note that at this point no numerical fluxes have been introduced yet into the DG formulation. We continue now with equations containing nonconservative products. Let $U \in V_h$. We know that the numerical solution is continuous on an element and discontinuous across a face, so, using Theorem 2, U is a weak solution to (4) if:

$$0 = \int_{\mathcal{E}_h} V_i d\mu_i \tag{17a}$$

$$\begin{aligned}
= & \sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\mathcal{K}} V_i (U_{i,0} + G_{ikr} U_{r,k}) d\mathcal{K} \\
& + \sum_{\mathcal{K} \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} \widehat{V}_i \left(\int_0^1 \delta_{ir} \frac{\partial \phi_r}{\partial \tau} (\tau; U_L, U_R) d\tau n_0^L \right) dK \right. \\
& \quad \left. + \int_{K(t_n^+)} \widehat{V}_i \left(\int_0^1 \delta_{ir} \frac{\partial \phi_r}{\partial \tau} (\tau; U_L, U_R) d\tau n_0^L \right) dK \right) \\
& + \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i \left(\int_0^1 G_{ikr} (\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau} (\tau; U^L, U^R) d\tau \bar{n}_k^L \right. \\
& \quad \left. + \int_0^1 \frac{\partial \phi_i}{\partial \tau} (\tau; U^L, U^R) d\tau n_0^L \right) d\mathcal{S}
\end{aligned} \tag{17b}$$

$$\begin{aligned}
= & \sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\mathcal{K}} V_i (U_{i,0} + G_{ikr} U_{r,k}) d\mathcal{K} \\
& + \sum_{\mathcal{K} \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} \widehat{V}_i (U_i^R - U_i^L) n_0^L dK + \int_{K(t_n^+)} \widehat{V}_i (U_i^R - U_i^L) n_0^L dK \right) \\
& + \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i \left(\int_0^1 G_{ikr} (\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau} (\tau; U^L, U^R) d\tau \bar{n}_k^L \right. \\
& \quad \left. - v_k \delta_{ir} \int_0^1 \frac{\partial \phi_r}{\partial \tau} (\tau; U^L, U^R) d\tau \bar{n}_k^L \right) d\mathcal{S}
\end{aligned} \tag{17c}$$

$$\begin{aligned}
&= \sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\mathcal{K}} V_i (U_{i,0} + G_{ikr} U_{r,k}) d\mathcal{K} \\
&\quad + \sum_{\mathcal{K} \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} \widehat{V}_i (U_i^R - U_i^L) dK - \int_{K(t_n^+)} \widehat{V}_i (U_i^R - U_i^L) dK \right) \\
&\quad + \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i \left(\int_0^1 G_{ikr} (\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau} (\tau; U^L, U^R) d\tau \bar{n}_k^L \right) d\mathcal{S} \\
&\quad + \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i [[v_k U_i]]_k d\mathcal{S},
\end{aligned} \tag{17d}$$

where $V \in V_h$ is an arbitrary test function. Furthermore, \widehat{V} is the value (numerical flux) of the test function V on a face \mathcal{S} and δ represents the Kronecker delta symbol. In (17d) we used the definition of n_0^L as given in (12). The crucial point in obtaining the DG formulation is the choice of the numerical flux for the test function V . We choose the numerical flux for V such that if there exists an F , with $G_{ikr} = \partial F_{ik} / \partial U_r$, then the DG formulation for the system containing nonconservative products reduces to the conservative space-time DGFEM weak formulation given by (16).

Theorem 3 *If the numerical flux \widehat{V} for the test function V in (17d) is defined as:*

$$\widehat{V} = \begin{cases} \{\{V\}\} & \text{at } \mathcal{S} \in \mathcal{S}_I, \\ 0 & \text{at } K(t_n) \subset \Omega_h(t_n) \quad \forall n, \end{cases} \tag{18}$$

then the DG formulation (17d) will reduce to the conservative space-time DGFEM formulation (16) when there exists an F such that $G_{ikr} = \partial F_{ik} / \partial U_r$.

Proof Assume there is an F , such that $G_{ikr} = \partial F_{ik} / \partial U_r$. We immediately see that:

$$\int_0^1 G_{ikr} (\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau} (\tau; U^L, U^R) d\tau \bar{n}_k^L = -[[F_{ik}]]_k. \tag{19}$$

Integrating by parts the volume integral in (17d) and using (19) we obtain:

$$\begin{aligned}
0 &= - \sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\mathcal{K}} (V_{i,0} U_i + V_{i,k} F_{ik}) d\mathcal{K} + \sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\partial \mathcal{K}} V_i^L (U_i^L n_0^L + F_{ik}^L \bar{n}_k^L) d(\partial \mathcal{K}) \\
&\quad + \sum_{\mathcal{K} \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} \widehat{V}_i (U_i^R - U_i^L) dK - \int_{K(t_n^+)} \widehat{V}_i (U_i^R - U_i^L) dK \right) \\
&\quad - \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i [[F_{ik} - v_k U_i]]_k d\mathcal{S}.
\end{aligned} \tag{20}$$

Using the definition of the normal vector (12), the element boundary integral in (20) becomes:

$$\begin{aligned}
\sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\partial \mathcal{K}} V_i^L (U_i^L n_0^L + F_{ik}^L \bar{n}_k^L) d(\partial \mathcal{K}) &= \sum_{\mathcal{K} \in \mathcal{K}_h} \int_{\mathcal{Q}} V_i^L (F_{ik}^L - v_k U_i^L) \bar{n}_k^L d\mathcal{Q} \\
&\quad + \sum_{\mathcal{K} \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} V_i^L U_i^L dK - \int_{K(t_n^+)} V_i^L U_i^L dK \right).
\end{aligned} \tag{21}$$

We will now use relations (13) and (14) to write the element boundary integrals as face integrals:

$$\begin{aligned}
& \sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\mathcal{Q}} V_i^L (F_{ik}^L - v_k U_i^L) \bar{n}_k^L d\mathcal{Q} = \\
& \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \llbracket V_i (F_{ik} - v_k U_i) \rrbracket_k d\mathcal{S} + \sum_{\mathcal{S} \in \mathcal{S}_B} \int_{\mathcal{S}} V_i^L (F_{ik}^L - v_k U_i^L) \bar{n}_k^L d\mathcal{S} \\
& = \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \left(\{\!\{ V_i \}\!\} \llbracket F_{ik} - v_k U_i \rrbracket_k + \llbracket V_i \rrbracket_k \{\!\{ F_{ik} - v_k U_i \}\!\} \right) d\mathcal{S} \\
& \quad + \sum_{\mathcal{S} \in \mathcal{S}_B} \int_{\mathcal{S}} V_i^L (F_{ik}^L - v_k U_i^L) \bar{n}_k^L d\mathcal{S}.
\end{aligned} \tag{22}$$

Combining (20), (21) and (22) we obtain:

$$\begin{aligned}
0 & = - \sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\mathcal{K}} (V_{i,0} U_i + V_{i,k} F_{ik}) d\mathcal{K} \\
& \quad + \sum_{K \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} V_i^L U_i^L dK - \int_{K(t_n^+)} V_i^L U_i^L dK \right) \\
& \quad + \sum_{K \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} \widehat{V}_i (U_i^R - U_i^L) dK - \int_{K(t_n^+)} \widehat{V}_i (U_i^R - U_i^L) dK \right) \\
& \quad + \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \left(\{\!\{ V_i \}\!\} \llbracket F_{ik} - v_k U_i \rrbracket_k + \llbracket V_i \rrbracket_k \{\!\{ F_{ik} - v_k U_i \}\!\} \right) d\mathcal{S} \\
& \quad - \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i \llbracket F_{ik} - v_k U_i \rrbracket_k d\mathcal{S} + \sum_{\mathcal{S} \in \mathcal{S}_B} \int_{\mathcal{S}} V_i^L (F_{ik}^L - v_k U_i^L) \bar{n}_k^L d\mathcal{S}.
\end{aligned} \tag{23}$$

The term $\{\!\{ V_i \}\!\} \llbracket F_{ik} - v_k U_i \rrbracket_k$ is set to zero in the space-time DG formulation for conservative systems by arguing that the formulation must be conservative. For a general nonconservative system we can not use this argument. Instead, we note that by taking $\widehat{V} = \{\!\{ V \}\!\}$ on the faces $\mathcal{S} \in \mathcal{S}_I$, the contribution $\int_{\mathcal{S}} \{\!\{ V_i \}\!\} \llbracket F_{ik} - v_k U_i \rrbracket_k d\mathcal{S}$ cancels with $-\int_{\mathcal{S}} \widehat{V}_i \llbracket F_{ik} - v_k U_i \rrbracket_k d\mathcal{S}$. Furthermore, taking $\widehat{V} = 0$ on the time faces $K(t_n) \subset \Omega_h(t_n) \forall n$, we obtain the space-time DGFEM weak formulation for conservative systems given by (16). \square

Theorem 3 allows us to finalize the derivation of the DGFEM formulation for hyperbolic nonconservative partial differential equations. First, we start with

the volume integral of (17d) and integrate by parts in time only, to obtain:

$$\begin{aligned}
0 &= \sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\mathcal{K}} \left(-V_{i,0} U_i + V_i G_{ikr} U_{r,k} \right) d\mathcal{K} \\
&+ \sum_{\mathcal{K} \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} V_i^L U_i^L dK - \int_{K(t_n^+)} V_i^L U_i^L dK \right) \\
&+ \sum_{\mathcal{K} \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} \widehat{V}_i (U_i^R - U_i^L) dK - \int_{K(t_n^+)} \widehat{V}_i (U_i^R - U_i^L) dK \right) \\
&- \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \left(\{V_i\} \llbracket v_k U_i \rrbracket_k + \llbracket V_i \rrbracket_k \{v_k U_i\} \right) d\mathcal{S} - \sum_{\mathcal{S} \in \mathcal{S}_B} \int_{\mathcal{S}} V_i^L v_k U_i^L \bar{n}_k^L d\mathcal{S} \\
&+ \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i \left(\int_0^1 G_{ikr} (\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau} (\tau; U^L, U^R) d\tau \bar{n}_k^L \right) d\mathcal{S} \\
&+ \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i \llbracket v_k U_i \rrbracket_k d\mathcal{S},
\end{aligned} \tag{24}$$

where we used relation (12) for the time component of the space-time normal vector and relations (13) and (14) to write the element boundary integrals as face integrals. For the numerical flux for the test function V in (24) we use (18), and thus obtain:

$$\begin{aligned}
0 &= \sum_{\mathcal{K} \in \mathcal{T}_h} \int_{\mathcal{K}} \left(-V_{i,0} U_i + V_i G_{ikr} U_{r,k} \right) d\mathcal{K} \\
&+ \sum_{\mathcal{K} \in \mathcal{T}_h} \left(\int_{K(t_{n+1}^-)} V_i^L U_i^L dK - \int_{K(t_n^+)} V_i^L U_i^L dK \right) \\
&- \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \llbracket V_i \rrbracket_k \{v_k U_i\} d\mathcal{S} - \sum_{\mathcal{S} \in \mathcal{S}_B} \int_{\mathcal{S}} V_i^L (v_k U_i^L) \bar{n}_k^L d\mathcal{S} \\
&+ \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \{V_i\} \left(\int_0^1 G_{ikr} (\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau} (\tau; U^L, U^R) d\tau \bar{n}_k^L \right) d\mathcal{S}.
\end{aligned} \tag{25}$$

Theorem 3 states that the weak formulation given by (25) can be reduced to the space-time DGFEM formulation (16), when an F exists such that $G_{ikr} = \partial F_{ik} / \partial U_r$. However, this formulation is generally numerically unstable. Problematic in the conservative space-time DGFEM formulation are the interior $\llbracket V_i \rrbracket_k \{F_{ik} - v_k U_i\}$ and boundary $V_i^L (F_{ik}^L - v_k U_i^L) \bar{n}_k^L$ flux terms, see (16). Generally, a stabilizing term is added to these flux terms, together forming an upwind numerical flux. Furthermore, the following upwind flux is introduced in the conservative space-time DGFEM formulation at the time faces, a formulation naturally ensuring causality in time:

$$\widehat{U} = \begin{cases} U^L & \text{at } K(t_{n+1}^-) \\ U^R & \text{at } K(t_n^+) \end{cases}. \tag{26}$$

It replaces the traces of U taken from the interior of $\mathcal{K} \in \mathcal{T}_h^n$. In (25), we also introduce the upwind flux (26) at the time faces. We also need a stabilizing term in (25). To understand how we add our stabilizing term, consider again the conservative space-time formulation. As mentioned above, a stabilizing term is added to $\llbracket V_i \rrbracket_k \{F_{ik} - v_k U_i\}$. Denote this stabilizing term as F_{ik}^{stab} , then $\{F_{ik} - v_k U_i\} + F_{ik}^{stab} = \{F_{ik}\} + \tilde{F}_{ik} = \hat{F}_{ik}$, where $\tilde{F} = \{-v_k U_i\} + F_{ik}^{stab}$ and \hat{F}_{ik} is the space-time numerical flux. In the nonconservative space-time formulation (25) we do not have a counterpart for $\{F_{ik}\}$. This term is hidden in the second and last term of (25), as shown in the proof of Theorem 3. We do have $\{-v_k U_i\}$. We can add a stabilizing term to $\{-v_k U_i\}$ such that $\{-v_k U_i\} + H_{ik}^{stab} = \tilde{H}^{nc}$. By introducing a ghost value U^R at the boundary, we can use the same expressions also at a boundary face. An expression for $\tilde{H}^{nc}(U^L, U^R, v, \bar{n}^L)$ is derived in Section 4, such that it reduces to the deviation of the numerical flux from the central flux in the conservative case, \tilde{F}_{ik} . Finally, the space-time DGFEM weak formulation for nonconservative partial differential equations (3) is:

Find a $U \in V_h$ such that for all $V \in V_h$:

$$\begin{aligned}
0 = & \sum_{\mathcal{K} \in \mathcal{T}_h^n} \int_{\mathcal{K}} \left(-V_{i,0} U_i + V_i G_{ikr} U_{r,k} \right) d\mathcal{K} \\
& + \sum_{\mathcal{K} \in \mathcal{T}_h^n} \left(\int_{K(t_{n+1}^-)} V_i^L U_i^L dK - \int_{K(t_n^+)} V_i^L U_i^R dK \right) \\
& + \sum_{\mathcal{S} \in \mathcal{S}^n} \int_{\mathcal{S}} \llbracket V_i \rrbracket_k \tilde{H}_{ik}^{nc} d\mathcal{S} \\
& + \sum_{\mathcal{S} \in \mathcal{S}^n} \int_{\mathcal{S}} \{V_i\} \left(\int_0^1 G_{ikr}(\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau}(\tau; U^L, U^R) d\tau \bar{n}_k^L \right) d\mathcal{S}.
\end{aligned} \tag{27}$$

Note that due to the introduction of the upwind flux at the time faces, each space-time slab only depends on the previous space-time slab so that the summation over all space-time slabs could be dropped.

4 The NCP numerical flux

In Section 3 we derived a weak formulation for space-time DGFEM for systems of equations containing a nonconservative product. To obtain an expression for the flux $\tilde{H}_{ik}^{nc}(U^L, U^R, v, \bar{n}^L)$ in (27), we first discuss the numerical flux \hat{U} , and then derive this numerical NCP flux for NonConservative Products.

Consider the following nonconservative hyperbolic system:

$$\partial_t U + G(U) \partial_x U = 0, \tag{28}$$

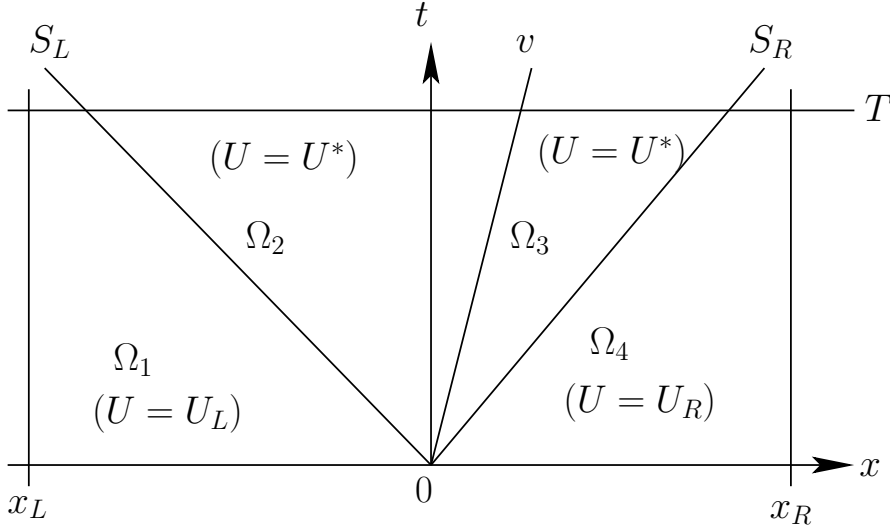


Fig. 1. Wave pattern of the solution for the Riemann problem. Here S_L and S_R are the fastest left and right moving signal velocities and v is the velocity of the element boundary point.

where $U \in \mathbb{R}^m$, with m the number of components of U , similarly $G(U) \in \mathbb{R}^{m \times m}$ and $x \in \mathbb{R}$ is along the normal of the face. To approximate the Riemann solution of (28) we consider only the fastest left and right moving waves of the system with velocities S_L and S_R and the grid velocity. In the star region (see Figure 1), which is the domain enclosed by the waves S_L and S_R , the averaged exact solution \bar{U}^* is defined as:

$$\bar{U}^* = \frac{1}{T(S_R - S_L)} \int_{TS_L}^{TS_R} U(x, T) dx. \quad (29)$$

In what follows we obtain a relation for \bar{U}^* from the weak formulation of (28). Integration by parts with respect to t over the control volume $\Omega_1 \cup \Omega_2$ yields:

$$\int_{x_L}^{S_L T} U_L dx + \int_{S_L T}^{vT} U(x, T) dx = \int_{x_L}^0 U(x, 0) dx - \int_{\Omega_2} G(U) \partial_x U dx dt - \int_0^T \int_0^1 G(\phi_{LL^*}(\tau; U_L, U_L^*)) \frac{\partial \phi_{LL^*}}{\partial \tau}(\tau; U_L, U_L^*) d\tau dt, \quad (30)$$

where $U_L^* = \lim_{s \downarrow S_L} U^*(st, t)$ is the trace of U^* taken from the interior of Ω_2 , which is constant along the wave S_L due to the self similarity of the solution in the star region. Replace the exact integrand in the second integral on the left hand side of (30) with the approximate solution \bar{U}^* . Furthermore, using

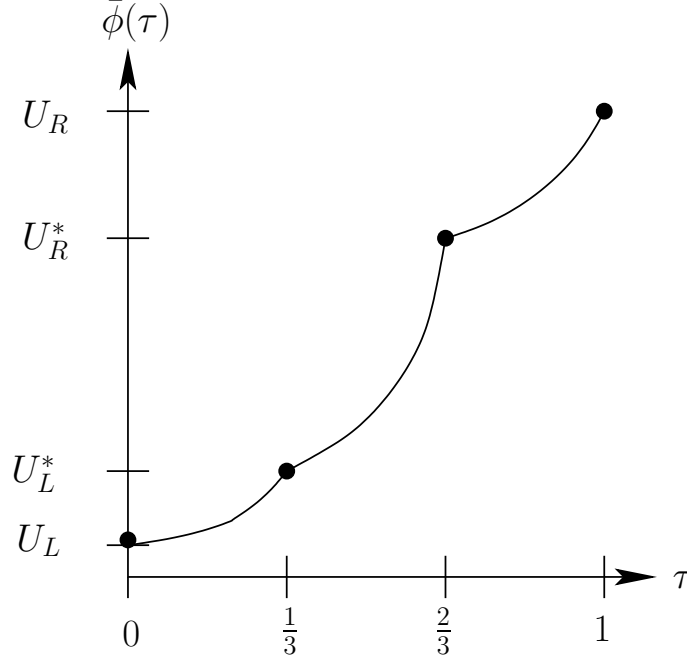


Fig. 2. Combining the paths to form $\bar{\phi}_{LR}(\tau; U_L, U_R) = \phi_{LL^*} \cap \phi_{L^*v} \cap \phi_{vR^*} \cap \phi_{R^*R}$.

the self similarity of the solution in the star region [3], we obtain:

$$\begin{aligned}
\int_{\Omega_2} G(U) \partial_x U \, dx dt &= \int_{t=0}^T \int_{x=S_L t}^{vt} G(U) \partial_x U \, dx dt \\
&= \int_{t=0}^T \int_{S_L}^v G(U^*) \partial_s U^* \partial_x s |J| \, ds dt \\
&= T \int_{S_L}^v G(U^*) \partial_s U^* \, ds,
\end{aligned} \tag{31}$$

where we used the coordinate transformation $x = st$, $t = t$, which has a Jacobian $|J| = t$. Introduce the trace of U^* taken from the interior of Ω_2 along the line $x = vt$ as: $U_{Lv}^* = \lim_{s \uparrow v} U^*(st, t)$ and the path $\phi_{L^*v} : [0, 1] \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ with:

$$\phi_{L^*v}(\tau; U_L^*, U_{Lv}^*) = U^*(s), \quad \text{if } S_L < s < v.$$

By connecting these two paths into the path $\phi_{Lv} : [0, 1] \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$, such that $\phi_{Lv}(\tau; U_L, U_{Lv}^*) = \phi_{LL^*} \cup \phi_{L^*v}$, redefining τ and using (31), the integral contributions on the righthand side of (30) can be combined, resulting in:

$$S_L U_L + (v - S_L) \bar{U}^* = - \int_0^1 G(\phi_{Lv}(\tau; U_L, U_{Lv}^*)) \frac{\partial \phi_{Lv}}{\partial \tau}(\tau; U_L, U_{Lv}^*) \, d\tau \tag{32}$$

(see Figure 2). Similarly, integration by parts with respect to t for the control volume $\Omega_3 \cup \Omega_4$ yields:

$$\int_{vT}^{S_{RT}} U(x, T) \, dx + \int_{S_{RT}}^{x_R} U_R \, dx = \int_0^{x_R} U(x, 0) \, dx - \int_{\Omega_3} G(U) \partial_x U \, dx \, dt$$

$$- \int_0^T \int_0^1 G(\phi_{R^*R}(\tau; U_R^*, U_R)) \frac{\partial \phi_{R^*R}}{\partial \tau}(\tau; U_R^*, U_R) d\tau dt, \quad (33)$$

where $U_R^* = \lim_{s \uparrow S_R} U^*(st, t)$ is the trace of U^* taken from the interior of Ω_3 , which is constant along the wave S_R . Furthermore, denote the trace of U^* taken from the interior of Ω_3 along the line $x = vt$ as: $U_{Rv}^* = \lim_{s \downarrow v} U^*(st, t)$. Replace the exact integrand in the second integral on the left hand side of (33) with the average of the exact solution \bar{U}^* . Introduce the path $\phi_{vR^*} : [0, 1] \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ with:

$$\phi_{vR^*}(\tau; U_{Rv}^*, U_R^*) = U^*(s), \quad \text{if } v < s < S_R,$$

and the path $\phi_{vR} : [0, 1] \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ such that $\phi_{vR}(\tau; U_{Rv}^*, U_R) = \phi_{R^*R} \cup \phi_{vR^*}$ after redefining τ . Using the self similarity of the solution in the star region Ω_3 , similar to (31), the integral contributions on the righthand side of (33) can be combined, resulting in:

$$(S_R - v)\bar{U}^* - S_R U_R = - \int_0^1 G(\phi_{vR}(\tau; U_{Rv}^*, U_R)) \frac{\partial \phi_{vR}}{\partial \tau}(\tau; U_{Rv}^*, U_R) d\tau. \quad (34)$$

Note that $U_{Lv}^* = U_{Rv}^*$ since the solution U is smooth across $\partial\bar{\Omega}_2 \cap \partial\bar{\Omega}_3$, where $\bar{\Omega}_2$ and $\bar{\Omega}_3$ are the closures of Ω_2 and Ω_3 . Now, introduce the path $\bar{\phi} : [0, 1] \times \mathbb{R}^m \times \mathbb{R}^m \rightarrow \mathbb{R}^m$ and redefine τ such that $\bar{\phi}(\tau; U_L, U_R) = \phi_{Lv} \cup \phi_{vR}$ then, by adding (32) and (34) and rearranging terms, we obtain:

$$\bar{U}^* = \frac{S_R U_R - S_L U_L}{S_R - S_L} - \frac{1}{S_R - S_L} \int_0^1 G(\bar{\phi}(\tau; U_L, U_R)) \frac{\partial \bar{\phi}}{\partial \tau}(\tau; U_L, U_R) d\tau. \quad (35)$$

This equation is still exact if we would know the path $\bar{\phi}$. Note from Figure 1 that outside the star region the solution is still at its initial values at $t = 0$ denoted by U_L and U_R . Within the star region bounded by the slowest and fastest signal speed S_L and S_R , respectively, an averaged star state solution \bar{U}^* is assumed. We define the numerical flux for U as:

$$\hat{U} = \begin{cases} U_L, & \text{if } v \leq S_L, \\ \bar{U}^*, & \text{if } S_L < v < S_R, \\ U_R, & \text{if } v \geq S_R, \end{cases} \quad (36)$$

where the averaged star state solution \bar{U}^* is given by (35) and v is the velocity of the element boundary point.

We now continue to derive an expression for $H^{nc}(U_L, U_R, v, \bar{n}^L)$. Define

$$\int_0^\tau G(\bar{\phi}(\tilde{\tau}; U_1, U_2)) \frac{\partial \bar{\phi}}{\partial \tilde{\tau}}(\tilde{\tau}; U_1, U_2) d\tilde{\tau} \equiv \int_0^\tau d\mathcal{G}(\bar{\phi}(\tilde{\tau}; U_1, U_2))$$

so that:

$$\int_0^1 G(\bar{\phi}(\tilde{\tau}; U_1, U_2)) \frac{\partial \bar{\phi}}{\partial \tilde{\tau}}(\tilde{\tau}; U_1, U_2) d\tilde{\tau} = \mathcal{G}(U_2) - \mathcal{G}(U_1),$$

using conditions H1-H4. Denote $\mathcal{G}(U_k) = \mathcal{G}_k$ and introduce $\tilde{\mathcal{G}}_k = \mathcal{G}_k - \{\{\mathcal{G}\}\}$, for $k = 1, 2$ with $\{\{\mathcal{G}\}\} = (\mathcal{G}_1 + \mathcal{G}_2)/2$. Note that $\mathcal{G}_2 - \mathcal{G}_1 = \tilde{\mathcal{G}}_2 - \tilde{\mathcal{G}}_1$. From (32) and (34), the definition of the paths, conditions H1-H4 and assuming $U_{Lv}^* = U_{Rv}^* = \bar{U}^*$, we then obtain:

$$S_L U_L + (v - S_L) \bar{U}^* = -\tilde{\mathcal{G}}^* + \tilde{\mathcal{G}}_L, \quad (37)$$

and:

$$(S_R - v) \bar{U}^* - S_R U_R = -\tilde{\mathcal{G}}_R + \tilde{\mathcal{G}}^*, \quad (38)$$

where $\mathcal{G}_L = \mathcal{G}(U_L)$, $\mathcal{G}_R = \mathcal{G}(U_R)$ and $\mathcal{G}^* = \mathcal{G}(\bar{U}^*)$. Subtracting (38) from (37) and rearranging the terms, we obtain:

$$\tilde{\mathcal{G}}^* = \{\{\tilde{\mathcal{G}}\}\} + \frac{1}{2} \left((S_R - v) \bar{U}^* + (S_L - v) \bar{U}^* - S_L U_L - S_R U_R \right) \quad (39)$$

with $\{\{\tilde{\mathcal{G}}\}\} \equiv (\tilde{\mathcal{G}}_L + \tilde{\mathcal{G}}_R)/2 = 0$. Similarly, by adding (37) and (38) together and rearranging terms, we obtain:

$$\tilde{\mathcal{G}}_L = \frac{1}{2} \int_0^1 G(\bar{\phi}(\tau; U_R, U_L)) \frac{\partial \phi}{\partial \tau}(\tau; U_R, U_L) d\tau \quad (40)$$

and:

$$\tilde{\mathcal{G}}_R = \frac{1}{2} \int_0^1 G(\bar{\phi}(\tau; U_L, U_R)) \frac{\partial \phi}{\partial \tau}(\tau; U_L, U_R) d\tau \quad (41)$$

The NCP numerical flux $\tilde{H}(U_L, U_R, v, \bar{n}^L)$ is thus given by:

$$\tilde{H}_{ik}^{nc}(U_L, U_R, v, \bar{n}^L) = \begin{cases} \frac{1}{2} \int_0^1 G_{ikr}(\bar{\phi}(\tau; U_R, U_L)) \frac{\partial \bar{\phi}_r}{\partial \tau}(\tau; U_R, U_L) d\tau & \text{if } S_L > v, \\ \frac{1}{2} \left((S_R - v) \bar{U}_i^* + (S_L - v) \bar{U}_i^* - S_L U_i^L - S_R U_i^R \right) \bar{n}_k & \text{if } S_L < v < S_R, \\ \frac{1}{2} \int_0^1 G_{ikr}(\bar{\phi}(\tau; U_L, U_R)) \frac{\partial \bar{\phi}_r}{\partial \tau}(\tau; U_L, U_R) d\tau & \text{if } S_R < v. \end{cases} \quad (42)$$

Note that if G is the Jacobian of some flux function F , then $\tilde{H}^{nc}(U_L, U_R, v, \bar{n}^L)$ is exactly the HLL version, minus the central flux $(F_L + F_R)/2$, of the numerical flux derived for moving grids in van der Vegt and van der Ven [18]. Take, e.g., the case $S_L > v$ in (42). In the conservative case, this reduces readily to $(F_L - F_R)/2$ which does equal $F_L - (F_L + F_R)/2$.

5 Test cases

5.1 The shallow water equations with topography

We consider the non-dimensional form of the shallow water system with topography. The system reads:

$$U_{i,0} + G_{ij}U_{j,1} = 0, \text{ for } i, j = 1, 2, 3 \quad (43)$$

with:

$$U = \begin{bmatrix} b \\ h \\ hu \end{bmatrix}, \text{ and } G(U) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ F^{-2}h & F^{-2}h - u^2 & 2u \end{bmatrix}. \quad (44)$$

Here b is the topography, h the water depth, u the flow velocity and F the Froude number defined as $F = u_0^*/\sqrt{g^*h_0^*}$, where the starred values denote reference values. The eigenvalues of $G(U)$ are given by:

$$\lambda_1 = u - \sqrt{F^{-2}h}, \quad \lambda_2 = 0, \quad \lambda_3 = u + \sqrt{F^{-2}h}. \quad (45)$$

In the numerical flux, as derived in Section 4, we take:

$$S_L = \min(u_L - \sqrt{F^{-2}h_L}, u_R - \sqrt{F^{-2}h_R}) \text{ and} \\ S_R = \max(u_L + \sqrt{F^{-2}h_L}, u_R + \sqrt{F^{-2}h_R}).$$

Test cases 1 and 2: rest flow

For test cases 1 and 2 we only consider the solution determined with space-DGFEM calculations using the linear path $\phi = U_L + \tau(U_R - U_L)$. Consider flow at rest over a discontinuous topography with initial conditions:

- Test case 1:

$$b(x) = \begin{cases} 1 & \text{for } x < 10 \\ 0 & \text{for } x > 10 \end{cases}, \quad h(x) + b(x) = 2, \quad h(x)u(x) = 0.$$

- Test case 2:

$$b(x) = \begin{cases} 0 & \text{for } x < 10 \\ 1 & \text{for } x > 10 \end{cases}, \quad h(x) + b(x) = 2, \quad h(x)u(x) = 0.$$

In Figures 3 and 4 we show the initial condition and the solution at $t = 200$, calculated using a time step $\Delta t = 0.01$ on a grid with 80 cells. The solution

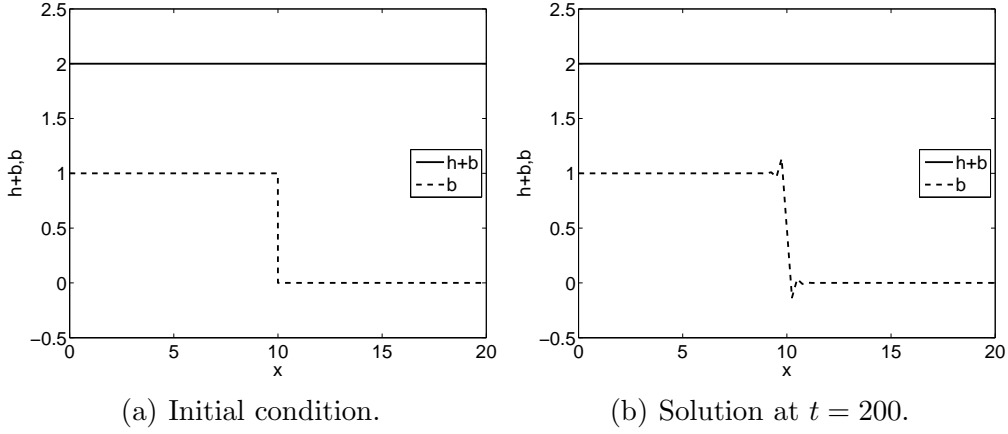


Fig. 3. Test case 1: flow at rest over a discontinuous topography. $F = 0.2$, 80 cells, $\Delta t = 0.01$.

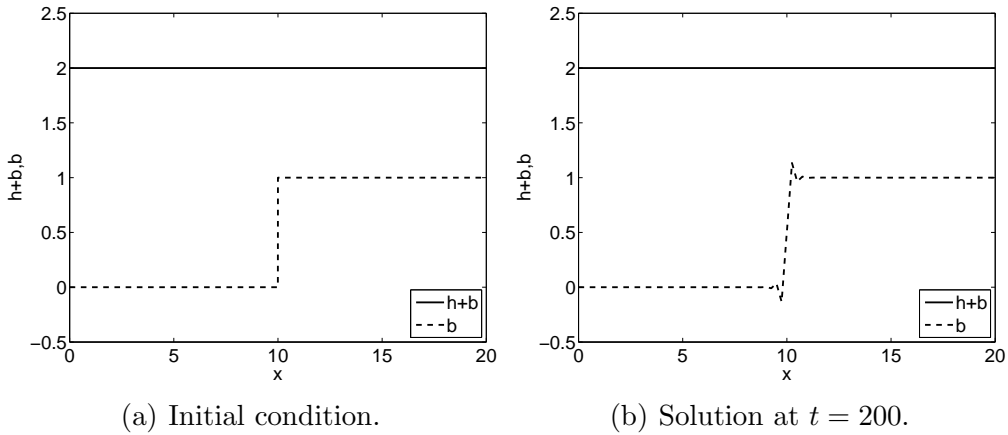


Fig. 4. Test case 2: flow at rest over a discontinuous topography. $F = 0.2$, 80 cells, $\Delta t = 0.01$.

does not change from $t = 0$ to 200. Note that the small over- and undershoots in the topography are a result of the first time step in which the discontinuous topography is forced to become continuous while the flow remains at rest. At $t = 200$ the absolute differences in the nodes between the numerical solution of $h + b$ with the exact solution are less than 10^{-10} . The absolute values of the error in the velocities (not shown here) is less than 10^{-8} .

We can also prove this theoretically when using linear basis functions and taking the path $\phi = U_L + \tau(U_R - U_L)$. Consider the one dimensional version

of the space DGFEM weak formulation (A.9) for the shallow water equations:

$$\begin{aligned}
0 &= \sum_k \int_{K_k} V_i(U_{i,0} + G_{ij}U_{j,1}) dK \\
&+ \sum_{S \in \mathcal{S}_I} \int_S \{V_i\} \left(\int_0^1 G_{ij}(\phi(\tau; U_L, U_R)) \frac{\partial \phi}{\partial \tau}(\tau; U_L, U_R) d\tau \right) \bar{n}^L dS \\
&+ \sum_{S \in \mathcal{S}_I} \int_S [V_i] \tilde{H}_i^{nc} dS.
\end{aligned}$$

We only consider cell K_k where the contributions satisfy:

$$\begin{aligned}
0 &= \int_{K_k} V_i(U_{i,0} + G_{ij}U_{j,1}) dK \\
&+ \int_{S_{k+1}} \frac{1}{2} V_i^L \left(\int_0^1 G_{ij}(\phi(\tau; U_L, U_R)) \frac{\partial \phi}{\partial \tau}(\tau; U_L, U_R) d\tau \right) \bar{n}^L + V_i^L \tilde{H}_i^{nc} dS \\
&+ \int_{S_k} \frac{1}{2} V_i^R \left(\int_0^1 G_{ij}(\phi(\tau; U_L, U_R)) \frac{\partial \phi}{\partial \tau}(\tau; U_L, U_R) d\tau \right) \bar{n}^L - V_i^R \tilde{H}_i^{nc} dS.
\end{aligned} \tag{46}$$

For the numerical flux we take the star-state solution given by (35). For rest flow, using $\phi = U_L + \tau(U_R - U_L)$ and $h_L + b_L = h_R + b_R$ the star-state solution is given by:

$$\bar{U}^* = \frac{1}{S_R - S_L} \begin{bmatrix} S_R b_R - S_L b_L \\ S_R h_R - S_L h_L \\ 0 \end{bmatrix}, \tag{47}$$

so that the numerical flux $\tilde{H}^{nc} = \frac{1}{2}(S_L(\bar{U}^* - U_L) + S_R(\bar{U}^* - U_R))$ is given by:

$$\tilde{H}^{nc} = \frac{S_L S_R}{S_R - S_L} \begin{bmatrix} b_R - b_L \\ h_R - h_L \\ 0 \end{bmatrix}. \tag{48}$$

Also, using $\phi = U_L + \tau(U_R - U_L)$ and $h_L + b_L = h_R + b_R$ we can show that:

$$\int_0^1 G_{ij}(\phi(\tau; U_L, U_R)) \frac{\partial \phi}{\partial \tau}(\tau; U_L, U_R) d\tau = [0 \ 0 \ 0]^T.$$

We can write (46) now as:

$$0 = \int_{K_k} V_i(U_{i,0} + G_{ij}U_{j,1}) dK + \int_{S_{k+1}} V_i^L \tilde{H}_i^{nc} dS - \int_{S_k} V_i^R \tilde{H}_i^{nc} dS. \tag{49}$$

Using linear basis functions we can evaluate the integrals as follows:

$$\int_{K_k} V_i U_{i,0} dK = \Delta x \bar{V}_i|_{K_k} \partial_t \bar{U}_i|_{K_k} + \frac{\Delta x}{3} \hat{V}_i|_{K_k} \partial_t \hat{U}_i|_{K_k}, \tag{50a}$$

$$\int_{K_k} V_i G_{ij} U_{j,1} dK = \int_{-1}^1 (\bar{V}_i|_{K_k} + \widehat{V}_i|_{K_k} \xi) G(\bar{U}_i|_{K_k} + \widehat{U}_i|_{K_k} \xi) \widehat{U}_i|_{K_k} d\xi = 0, \quad (50b)$$

$$\int_{S_{k+1}} V^L \tilde{H}^{nc} dS = (\bar{V}|_{K_k} + \widehat{V}|_{K_k}) \frac{S_{k+1}^L S_{k+1}^R}{S_{k+1}^R - S_{k+1}^L} \begin{bmatrix} b_{k+1}^R - b_{k+1}^L \\ h_{k+1}^R - h_{k+1}^L \\ 0 \end{bmatrix}, \quad (50c)$$

$$\int_{S_k} V^R \tilde{H}^{nc} dS = (\bar{V}|_{K_k} - \widehat{V}|_{K_k}) \frac{S_k^L S_k^R}{S_k^R - S_k^L} \begin{bmatrix} b_k^R - b_k^L \\ h_k^R - h_k^L \\ 0 \end{bmatrix}, \quad (50d)$$

where $\bar{(\cdot)}$ and $\widehat{(\cdot)}$ are the means and slopes, respectively, of the approximation for U and V . The second integral (50b) is zero using $h_L + b_L = h_R + b_R$ and the fact that the slope of $h + b = 0$ (so $\widehat{U}|_{K_k} = (-\hat{h}_k, \hat{h}_k, 0)$). Note that in (50c) and (50d) we have $b_{k+1}^R - b_{k+1}^L + h_{k+1}^R - h_{k+1}^L = 0$ and $b_k^R - b_k^L + h_k^R - h_k^L = 0$, respectively so that:

$$\partial_t(\bar{h}_k + \bar{b}_k) = 0, \quad \partial_t(\hat{h}_k + \hat{b}_k) = 0, \quad \partial_t \overline{hu}_k = 0, \quad \partial_t \widehat{hu}_k = 0,$$

meaning that for rest flow $h + b$ remains constant.

Test case 3: Subcritical flow over a bump

We now consider subcritical flow with a Froude number of $F = 0.2$ over a bump. The topography reads:

$$b(x) = \begin{cases} a(b - (x - x_p))(b + (x - x_p))b^{-2} & \text{for } |x - x_p| \leq b, \\ 0 & \text{otherwise.} \end{cases} \quad (51)$$

We use $x_p = 10$, $a = 0.5$ and $b = 2$ as in [14]. The exact steady state solution for this test case is found by solving the following third order equation in u [6,14]:

$$F^2 u^3 / 2 + (b + F^2 / 2 - 1)u + 1 = 0 \text{ with } hu = 1. \quad (52)$$

The domain $x \in [0, 20]$ is divided into 40, 80, 160 and 320 cells. We consider DGFEM and STDGFEM calculations using the linear path $\phi = U_L + \tau(U_R - U_L)$. The final solution for the DGFEM calculations is considered at time $t = 300$. Time steps of $\Delta t = 0.01$ are made, except for the solutions on a grid with 320 cells where $\Delta t = 0.005$. For STDGFEM calculations we consider the solution after one physical time step of $\Delta t = 10^{21}$. We can do this because we want to consider the steady state solution. We solve the system of non-linear equations using a pseudo time stepping integration method (see van der Vegt and van der Ven. [18]). As stopping criterium in the pseudo time-stepping calculation we take that the maximum residual must be smaller than 10^{-8} . A

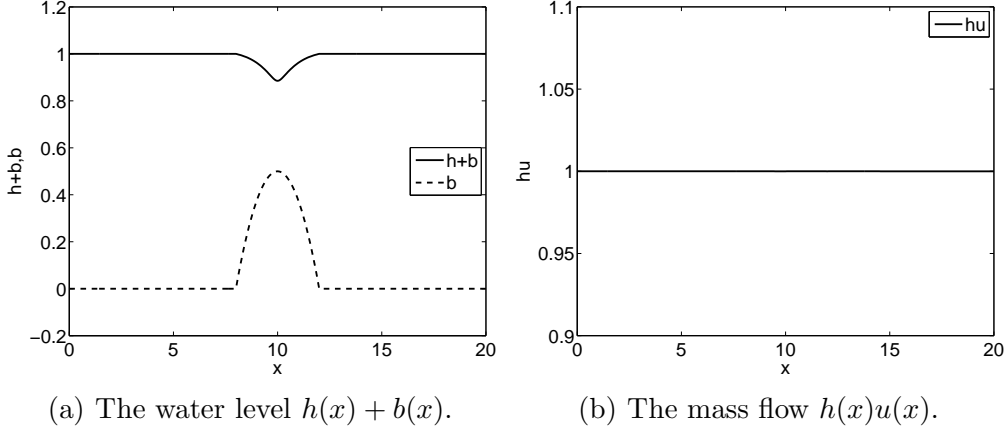


Fig. 5. Test case 3: steady-state solution calculated using space DGFEM, $F = 0.2$, 320 cells.

pseudo time stepping CFL number of $CFL^{pseudo} = 0.8$ is used.

The initial condition is $h + b = 1$ and $hu = 1$. To implement the boundary we force the mean of hu to be 1 in the first cell and the slope of hu to be 0; and in the last cell we force the mean of $h + b$ to be 1 and the slopes of $h + b$ to be 0. The steady state solution is given in Figure 5. The order convergence is determined by looking at the L^2 and the L^{max} norm of the numerical error in $h + b$ and hu with respect to the exact solution:

$$\|(h + b)_{num} - (h + b)_{exact}\|_2 = \left(\sum_{k=1}^{N_{cells}} \int_{K_k} ((h + b)_{K_k}^{num} - (h + b)_{K_k}^{exact})^2 \right)^{1/2}, \quad (53)$$

and:

$$\|(h + b)_{num} - (h + b)_{exact}\|_{max} = \max\{|(h + b)_{num}^i - (h + b)_{exact}^i| : 1 \leq i \leq N_{cells}\}. \quad (54)$$

The order of convergence using DGFEM and STDGFEM is given in Table 1. For both the space-DGFEM calculations and the space-time DGFEM calculations second order convergence is found.

Test case 4: Supercritical flow over a bump

Next, we consider supercritical flow with a Froude number of $F = 1.9$ over a bump. We use the same topography (51) and the exact solution can be found by solving (52). The domain $x \in [0, 20]$ is again divided into 40, 80, 160 and 320 cells and we consider DGFEM and STDGFEM calculations using the linear path $\phi = U_L + \tau(U_R - U_L)$. The final solution for the space-DGFEM calculations is considered at time $t = 300$. Time steps of $\Delta t = 0.01$ are made,

DGFEM

N_{cells}	$h + b$				hu			
	L^2 error	p	L^{max} error	p	L^2 error	p	L^{max} error	p
40	$0.1195 \cdot 10^{-2}$	-	$0.6702 \cdot 10^{-2}$	-	$0.7448 \cdot 10^{-3}$	-	$0.2281 \cdot 10^{-2}$	-
80	$0.3240 \cdot 10^{-3}$	1.9	$0.2408 \cdot 10^{-2}$	1.5	$0.1534 \cdot 10^{-3}$	2.3	$0.8437 \cdot 10^{-3}$	1.4
160	$0.8422 \cdot 10^{-4}$	1.9	$0.7009 \cdot 10^{-3}$	1.8	$0.2344 \cdot 10^{-4}$	2.7	$0.1463 \cdot 10^{-3}$	2.5
320	$0.2055 \cdot 10^{-4}$	2.0	$0.1786 \cdot 10^{-3}$	2.0	$0.7622 \cdot 10^{-5}$	1.6	$0.3307 \cdot 10^{-4}$	2.1

STDGFEM

N_{cells}	$h + b$				hu			
	L^2 error	p	L^{max} error	p	L^2 error	p	L^{max} error	p
40	$0.1202 \cdot 10^{-2}$	-	$0.6750 \cdot 10^{-2}$	-	$0.7380 \cdot 10^{-3}$	-	$0.2263 \cdot 10^{-2}$	-
80	$0.3246 \cdot 10^{-3}$	1.9	$0.2412 \cdot 10^{-2}$	1.5	$0.1530 \cdot 10^{-3}$	2.3	$0.8481 \cdot 10^{-3}$	1.4
160	$0.8425 \cdot 10^{-4}$	1.9	$0.7021 \cdot 10^{-3}$	1.8	$0.2327 \cdot 10^{-4}$	2.7	$0.1413 \cdot 10^{-3}$	2.6
320	$0.2127 \cdot 10^{-4}$	2.0	$0.1851 \cdot 10^{-3}$	1.9	$0.3313 \cdot 10^{-5}$	2.8	$0.2744 \cdot 10^{-4}$	2.4

Table 1

L^2 and L^{max} error for $h + b$ and hu using DGFEM and STDGFEM for test case 3. Convergence rates are shown for $F = 0.2$.

except for the solutions on a grid with 320 cells where $\Delta t = 0.005$.

For the STDGFEM calculation we consider again the solution after one physical time step of $\Delta t = 10^{21}$. The same stopping criteria are used as in the subcritical flow case. The initial condition is $h + b = 1$ and $hu = 1$. To implement the boundary condition on the left we force the mean of hu to be 1 and the slope of hu to be 0 in the first cell. At the right boundary $x = 20$, we set the right trace of all the variables equal to the left trace and we let the numerical flux decide what to do with this information. The steady-state solution is given in Figure 6. The order convergence is again determined by computing the L^2 and the L^{max} norm of the numerical error in $h + b$ and hu with respect to the exact solution as defined in (53) and (54). The order of convergence using DGFEM and STDGFEM is given in Table 2.

We see that the space- and space-time DGFEM calculations results in second order convergence for $h + b$. We do not show the order of convergence for hu because the error for hu is of the order of machine precision on all meshes for the space DGFEM calculations and stabilizes around 10^{-8} for the space-time DGFEM calculations.

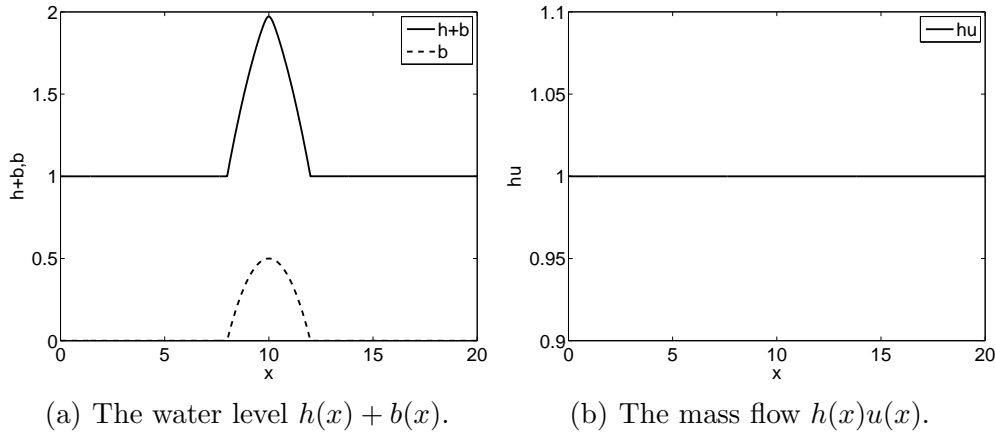


Fig. 6. Test case 4: steady-state solution calculated using space DGFEM, $F = 1.9$, 320 cells.

N_{cells}	DGFEM $h + b$				STDGFEM $h + b$			
	L^2 error	p	L^{max} error	p	L^2 error	p	L^{max} error	p
40	$0.7562 \cdot 10^{-2}$	-	$0.4630 \cdot 10^{-1}$	-	$0.7562 \cdot 10^{-2}$	-	$0.4630 \cdot 10^{-1}$	-
80	$0.1281 \cdot 10^{-2}$	2.6	$0.9410 \cdot 10^{-2}$	2.3	$0.1281 \cdot 10^{-2}$	2.6	$0.9410 \cdot 10^{-2}$	2.3
160	$0.3188 \cdot 10^{-3}$	2.0	$0.2615 \cdot 10^{-2}$	1.8	$0.3188 \cdot 10^{-3}$	2.0	$0.2615 \cdot 10^{-2}$	1.8
320	$0.7914 \cdot 10^{-4}$	2.0	$0.6883 \cdot 10^{-3}$	1.9	$0.7914 \cdot 10^{-4}$	2.0	$0.6883 \cdot 10^{-3}$	1.9

Table 2
 L^2 and L^{max} error for $h + b$ using DGFEM and STDGFEM for test case 4. Convergence rates are shown for $F = 1.9$.

Conclusions

For the shallow water equations with topography we showed numerical results of four test cases calculated using the space- and space-time DGFEM discretizations we developed for nonconservative hyperbolic partial differential equations. For all test cases we obtained good results. For test cases 1 and 2 we showed that rest flow remained unchanged despite having discontinuities in the topography. We do not need any special tricks to deal with these discontinuities. In test cases 3 and 4 we solved subcritical and supercritical flow over a bump. We demonstrated that the scheme is second order accurate. We were also able to solve transcritical flow over a bump leading to more complex flow, but we do not show results here. See Houghton and Kasahara [6] for the analysis of these test cases.

5.2 The depth averaged two-fluid model

In this section we consider two fluid models (also known as Eulerian models) in which the particle phase is treated as a continuum by averaging over individual particles. Two systems of frequently used governing equations that follow the Eulerian approach, are the systems derived by Anderson and Jackson [1], and the systems derived by Drew and Lahey [4] and Enwald et al. [5]. Apart from their derivation, the difference between these systems of equations is how the fluid-phase shear stress (if included) is multiplied by the solid volume fraction in the momentum equations (see also van Wachem et al. [17]). In the limiting case that pressure is the only fluid stress, both formulations are equal.

We will consider a simplification of these sets of equations, namely the depth-averaged two fluid model as derived by Pitman and Le [12]. They start with the system derived by Anderson and Jackson [1] and use the shallow flow assumption, $H/L \ll 1$, to obtain, in a similar way as the shallow water equations are derived from the Navier-Stokes equations, the depth averaged two-fluid model. Since the pressure is the only fluid stress, the same depth-averaged two fluid model also follows from the system derived by Drew and Lahey [4] and Enwald et al. [5].

The dimensionless depth-averaged two fluid model as derived by Pitman and Le [12], but ignoring source terms for simplicity, can be written as:

$$U_{i,0} + G_{ij}U_{j,1} = 0, \text{ for } i, j = 1, 2, 3, 4, \quad (55)$$

where:

$$U = [h(1 - \alpha), h\alpha, hu(1 - \alpha), hav, b]^T$$

$$G(U) = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ u^2(1 - \alpha) - 2u^2 + 2c_2h(1 - \alpha) & u^2(1 - \alpha) + 2c_2h(1 - \alpha) & (2 + \alpha)u & -u(1 - \alpha) & (1 - \alpha)hc_5 \\ (c_1 + c_3)h\alpha & c_1(1 + \alpha)h + c_3h\alpha - v^2 & 0 & 2v & hac_4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (56)$$

Here, we have defined the following variables:

$$c_1 = \frac{1}{2}\varepsilon(1 - \rho_f/\rho_s)\alpha_{xx}g^z, \quad c_2 = \frac{1}{2}\varepsilon g^z, \quad c_3 = \varepsilon g^z \rho_f/\rho_s, \\ c_4 = (1 - \rho_f/\rho_s)\varepsilon\alpha_{xx}g^z + \varepsilon g^z \rho_f/\rho_s, \quad c_5 = \varepsilon g^z,$$

Again we have taken the topography as an unknown. The meaning of the different symbols is:

$h(x, t)$: depth of the flow,
 $v(x, t)$: velocity of the solid phase,

$u(x, t)$: velocity of the fluid phase,
 $\alpha(x, t)$: volume fraction of the solid phase,
 $b(x)$: topography term,
 $\varepsilon = H/L$, where H is the characteristic length in the z -direction and L the characteristic length in the y -direction,
 ρ_f : the fluid density,
 ρ_s : the solid density,
 $\alpha_{xx} = k_{ap}$, where k_{ap} is the Earth pressure coefficient,
 g^z : z -component of the scaled gravity.

Note that in the limit $\alpha \rightarrow 0$, this model reduces to the shallow water equations:

$$\begin{aligned} \partial_t h + \partial_x(hu) &= 0, \\ \partial_t(hu) + \partial_x(hu^2 + \frac{1}{2}\varepsilon g^z h^2) &= -\varepsilon g^z h \partial_x b. \end{aligned} \tag{57}$$

In the limit $\alpha \rightarrow 1$, the depth-averaged two-fluid model reduces to:

$$\begin{aligned} \partial_t h + \partial_x(hv) &= 0, \\ \partial_t(hv) + \partial_x(hv^2 + \frac{1}{2}\varepsilon k_{ap} g^z h^2) &= -\varepsilon k_{ap} g^z h \partial_x b, \end{aligned} \tag{58}$$

which is the Savage-Hutter model without source terms, a model that simulates avalanches of dry granular matter [7].

In our simulations, we set the Earth pressure coefficient to be $\alpha_{xx} = 1$ and take $\varepsilon = 1$. For notational convenience, we set $\rho_f/\rho_s = \rho$ and $g^z = g$. To compute the eigenvalues of $G(U)$, we use the LAPACK package. The biggest eigenvalue is used for S_R and the smallest eigenvalue is used for S_L in the NCP numerical flux.

Test case 5: Two-phase subcritical flow

As in the case of the shallow water equations with topography, also for the two-phase flow model we consider the steady state solution for subcritical flow over a bump. We consider the same topography as given in (51). The reference solution for this problem is found by solving the following ODE:

$$\partial_x U = A^{-1} S, \tag{59}$$

where U , A and S are given by:

$$\begin{aligned}
 U &= [h(1 - \alpha), \quad h\alpha]^T, \quad S = \begin{bmatrix} -(1 - \alpha)hg\partial_x b \\ -gh\alpha\partial_x b \end{bmatrix} \\
 A &= \begin{bmatrix} u^2(1 - \alpha) - 2u^2 + gh(1 - \alpha) & u^2(1 - \alpha) + gh(1 - \alpha) \\ \frac{1}{2}(1 + \rho)gh\alpha & \frac{1}{2}(1 - \rho)g(1 + \alpha)h + g\rho h\alpha - v^2 \end{bmatrix},
 \end{aligned} \tag{60}$$

with the topography derivative a known function and steady state discharges:

$$hu(1 - \alpha) = q_1, \quad hv\alpha = q_2, \tag{61}$$

with q_1 and q_2 integration constants. For this test case we take $q_1 = 0.2$, $q_2 = 0.1$, $g = 1$ and $\rho = 0.5$. As initial condition we take $h(1 - \alpha) = 1$, $h\alpha = 0.6$, $hu(1 - \alpha) = 0.2$ and $hv\alpha = 0.1$. We use the STDGFEM formulation to calculate the solution. We consider one physical time step of $\Delta t = 10^{21}$ and use a pseudo time stepping integration method to solve the system of non-linear equations. We determine the solution on a domain $x \in [0, 20]$ divided into 40, 80, 160 and 320 cells. As stopping criterium in the pseudo time-stepping method we take that the maximum residual must be smaller than 10^{-8} . A pseudo time stepping CFL number of $CFL^{pseudo} = 0.1$ is used. At the boundaries, we define the exterior trace to be the same as the initial condition. The numerical flux decides then what to do with this information. The steady state solution is given in Figure 7. The order convergence is determined by computing the L^2 and the L^{max} norm of the error, similar as to what is done in (53) and (54). The order of convergence is given in Table 3. We clearly see second order convergence as is expected.

Test case 6: Two-phase supercritical flow

We will now consider the steady state solution of two-phase supercritical flow over a bump. We consider again the topography given in (51). The exact solution is found by solving (59)-(61). For this test case we now take $q_1 = 4$ and $q_2 = 2$. Other constants remain the same as in test case 5. We again use the same solution strategy as in test case 5. The steady state solution is given in Figure 8 and the order convergence is given in Table 4. Again we clearly see second order convergence for the variables $h(1 - \alpha) + b$ and $h\alpha + b$. We do not see second order convergence for the variables $hu(1 - \alpha)$ and $hv\alpha$ because the error for these solutions stabilizes around 10^{-8} .

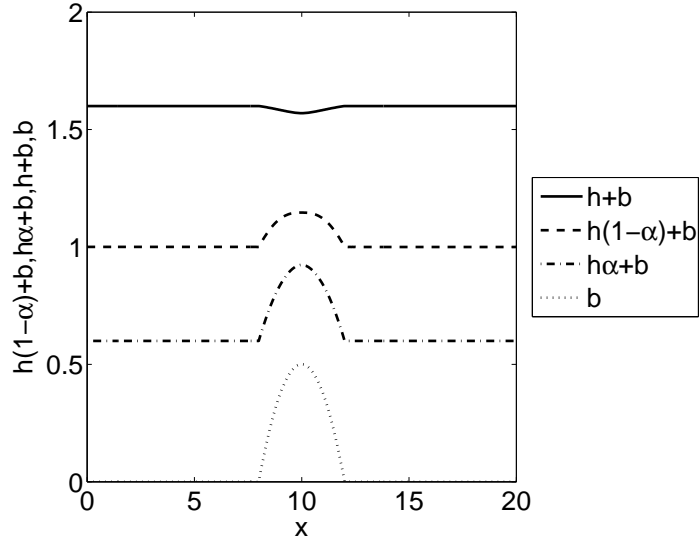


Fig. 7. Test case 5: steady-state solution for a subcritical two-phase flow calculated with STDGFEM using 320 cells. Shown are the total flow height $h + b$, the flow height due to the fluid phase $h(1 - \alpha)$, the flow height due to the solids phase $h\alpha$ and the topography b .

STDGFEM									
N_{cells}	$h(1 - \alpha) + b$				$h\alpha + b$				
	L^2 error	p	L^{max} error	p	L^2 error	p	L^{max} error	p	
40	$0.8171 \cdot 10^{-3}$	-	$0.2308 \cdot 10^{-2}$	-	$0.1404 \cdot 10^{-2}$	-	$0.4194 \cdot 10^{-2}$	-	
80	$0.2025 \cdot 10^{-3}$	2.0	$0.5584 \cdot 10^{-3}$	2.0	$0.3537 \cdot 10^{-3}$	2.0	$0.9903 \cdot 10^{-3}$	2.1	
160	$0.4871 \cdot 10^{-4}$	2.1	$0.1322 \cdot 10^{-3}$	2.1	$0.8511 \cdot 10^{-4}$	2.1	$0.2306 \cdot 10^{-3}$	2.1	
320	$0.9789 \cdot 10^{-5}$	2.3	$0.2651 \cdot 10^{-4}$	2.3	$0.1712 \cdot 10^{-4}$	2.3	$0.4597 \cdot 10^{-4}$	2.3	
N_{cells}	$hu(1 - \alpha)$				$hv(\alpha)$				
	L^2 error	p	L^{max} error	p	L^2 error	p	L^{max} error	p	
40	$0.3672 \cdot 10^{-4}$	-	$0.1442 \cdot 10^{-3}$	-	$0.1212 \cdot 10^{-4}$	-	$0.3409 \cdot 10^{-4}$	-	
80	$0.5911 \cdot 10^{-5}$	2.6	$0.3448 \cdot 10^{-4}$	2.1	$0.1791 \cdot 10^{-5}$	2.8	$0.8054 \cdot 10^{-5}$	2.1	
160	$0.1049 \cdot 10^{-5}$	2.5	$0.8471 \cdot 10^{-5}$	2.0	$0.3807 \cdot 10^{-6}$	2.2	$0.2048 \cdot 10^{-5}$	2.0	
320	$0.1723 \cdot 10^{-6}$	2.6	$0.2078 \cdot 10^{-5}$	2.0	$0.5115 \cdot 10^{-7}$	2.9	$0.4861 \cdot 10^{-6}$	2.1	

Table 3
 L^2 and L^{max} error for $h(1 - \alpha) + b$, $h\alpha + b$, $hu(1 - \alpha)$ and $hv\alpha$ using STDGFEM for test case 5.

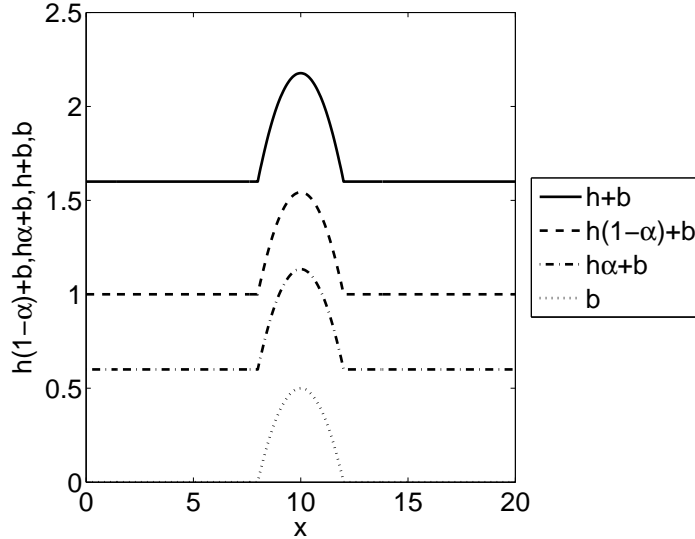


Fig. 8. Test case 6: steady-state solution for a supercritical two-phase flow calculated using STDGFEM using 320 cells. Shown are the total flow height $h + b$, the flow height due to the fluid phase $h(1 - \alpha)$, the flow height due to the solids phase $h\alpha$ and the topography b .

STDGFEM									
N_{cells}	$h(1 - \alpha) + b$				$h\alpha + b$				
	L^2 error	p	L^{max} error	p	L^2 error	p	L^{max} error	p	
40	$0.2400 \cdot 10^{-2}$	-	$0.5674 \cdot 10^{-2}$	-	$0.2359 \cdot 10^{-2}$	-	$0.5575 \cdot 10^{-2}$	-	
80	$0.6060 \cdot 10^{-3}$	2.0	$0.1402 \cdot 10^{-2}$	2.0	$0.5958 \cdot 10^{-3}$	2.0	$0.1378 \cdot 10^{-2}$	2.0	
160	$0.1459 \cdot 10^{-3}$	2.1	$0.3339 \cdot 10^{-3}$	2.1	$0.1434 \cdot 10^{-3}$	2.1	$0.3280 \cdot 10^{-3}$	2.1	
320	$0.2933 \cdot 10^{-4}$	2.3	$0.6678 \cdot 10^{-4}$	2.3	$0.2884 \cdot 10^{-4}$	2.3	$0.6561 \cdot 10^{-4}$	2.3	

Table 4
 L^2 and L^{max} error for $h(1 - \alpha) + b$, $h\alpha + b$, $hu(1 - \alpha)$ and $hv\alpha$ using STDGFEM for test case 6.

Test case 7: A two-phase dam break problem

For the depth-averaged two-phase flow model we consider a dam break type test case. Consider two mixtures separated by a membrane. The left mixture has a solid volume fraction of $\alpha = 0.4$ and the right mixture has a solid volume fraction of $\alpha = 0.6$. At time $t = 0$ we remove the membrane. We want to know how the mixtures behave. We consider the solution on the domain $[0, 1]$. As initial condition we take:

$$U(x, 0) = \begin{cases} U_L, & \text{if } x < 0.5, \\ U_R, & \text{if } x > 0.5, \end{cases}$$

where:

$$U_L = [1.8, 1.2, 0, 0, 0]^T \text{ and } U_R = [1.2, 1.8, 0, 0, 0]^T.$$

The constants in the computation are taken as $g = 1$ and $\rho = 0.5$. We compute the solution on a domain with 16, 32, 64, 128, 256, 512 or 1024 elements. We consider DGFEM calculations using the linear path $\phi = U_L + \tau(U_R - U_L)$. The solution is determined at $t = 0.175$ using a time step of $\Delta t = 0.0001$. To be able to compute this test case we use slope limiters to deal with overshoots and undershoots, otherwise the computation crashes. The solution in a cell is given by:

$$U_k = \bar{U}_k + \psi(x)m(\hat{U}_k, \bar{U}_{k+1} - \bar{U}_k, \bar{U}_k - \bar{U}_{k-1}), \quad (62)$$

where the minmod function m is defined as:

$$m(a_1, a_2, a_3) = \begin{cases} s \min_{1 \leq n \leq 3} |a_n| & \text{if } s = \text{sign}(a_1) = \text{sign}(a_2) = \text{sign}(a_3) \\ 0 & \text{otherwise.} \end{cases} \quad (63)$$

The solutions of $h(1-\alpha)$, $h\alpha$, b and h computed on a mesh with 1024 elements are depicted in Figure 9a, the solutions of $hu(1-\alpha)$ and $h\nu\alpha$ are depicted in Figure 9b and the solution of α is depicted in Figure 9c. As expected, the particles in the mixture with the larger solid volume fraction are moving to the mixture with the lower solid volume fraction. In the middle of the domain an average state is reached where the solid volume fraction stays approximately 0.5. Since we do not have an exact solution, we compute the order behavior using the following approach:

$$\frac{\|U_N - U_{2N}\|_2}{\|U_{2N} - U_{4N}\|_2} = 2^p, \quad (64)$$

where p is the order of convergence, U_N the solution on a mesh consisting of N cells, and $\|\cdot\|_2$ is the L^2 norm. The order behavior is shown in Table 5. Due to the presence of shocks we cannot obtain second order accuracy. Instead we obtain a convergence rate of approximately $\mathcal{O}(h^{1/2})$.

6 Effect of the path in phase space on numerical solution

6.1 Polynomial paths

In the numerical test cases discussed in the previous section a linear path was taken: $\phi = U^L + \tau(U^R - U^L)$. In this section, we will investigate the effect of different paths on our numerical results. To determine this effect we again consider test case 7 in Section 5.2 for which we expect to find the biggest effect

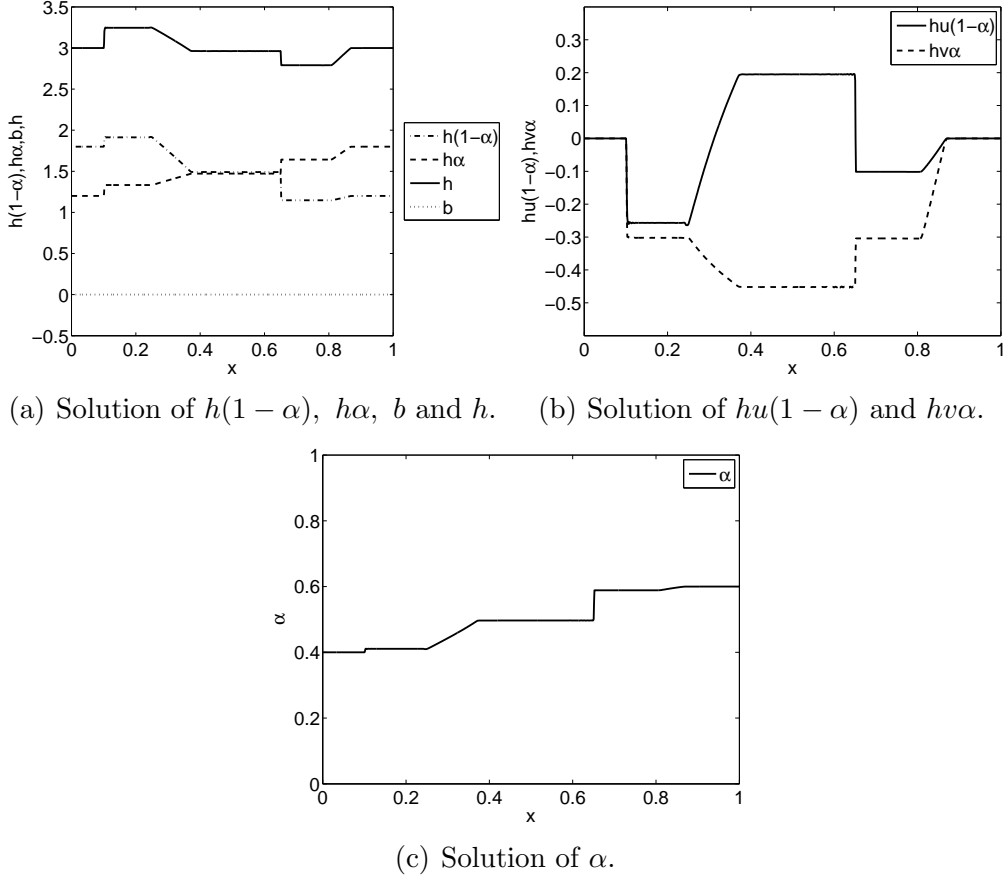
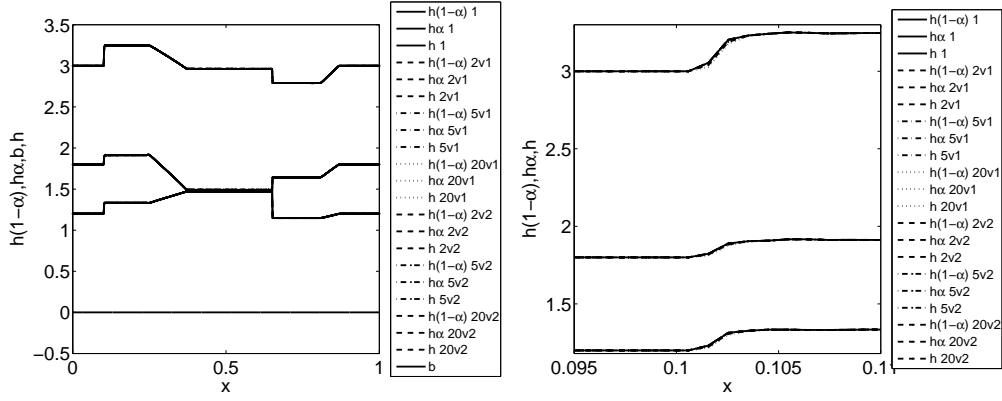


Fig. 9. Test case 7. The solution computed on a mesh with 1024 elements at time $t = 0.175$ using DGFEM.

DGFEM									
N_{cells}	L^2 of $h(1 - \alpha)$	p	L^2 of $h\alpha$	p	L^2 of $hu(1 - \alpha)$	p	L^2 of $h\nu\alpha$	p	
32	$0.1238 \cdot 10^{-1}$	-	$0.7030 \cdot 10^{-2}$	-	$0.1263 \cdot 10^{-1}$	-	$0.1384 \cdot 10^{-1}$	-	
64	$0.1125 \cdot 10^{-1}$	0.1	$0.5780 \cdot 10^{-2}$	0.3	$0.1155 \cdot 10^{-1}$	0.1	$0.8164 \cdot 10^{-2}$	0.8	
128	$0.6231 \cdot 10^{-2}$	0.9	$0.3391 \cdot 10^{-2}$	0.8	$0.7114 \cdot 10^{-2}$	0.7	$0.4465 \cdot 10^{-2}$	0.9	
256	$0.4379 \cdot 10^{-2}$	0.5	$0.2751 \cdot 10^{-2}$	0.3	$0.4494 \cdot 10^{-2}$	0.7	$0.3828 \cdot 10^{-2}$	0.2	
512	$0.3085 \cdot 10^{-2}$	0.5	$0.1875 \cdot 10^{-2}$	0.6	$0.3536 \cdot 10^{-2}$	0.3	$0.3275 \cdot 10^{-2}$	0.2	

Table 5
 L^2 error and convergence rate for $h(1 - \alpha)$, $h\alpha$, $hu(1 - \alpha)$ and $h\nu\alpha$ using DGFEM for test case 7. The convergence rates are shown for the solution at $t = 0.175$. With L^2 of U we mean $\|U_N - U_{2N}\|_2$.

of the path due to the shock waves in the solution. We use the following paths



(a) The solution on the whole domain. (b) The solution zoomed in on the left shock wave.

Fig. 10. Solution of $h(1 - \alpha)$, $h\alpha$, b and h calculated on a mesh with 1024 elements at time $t = 0.175$ using the paths defined in (65).

and note that in one dimension property (H4) can be neglected:

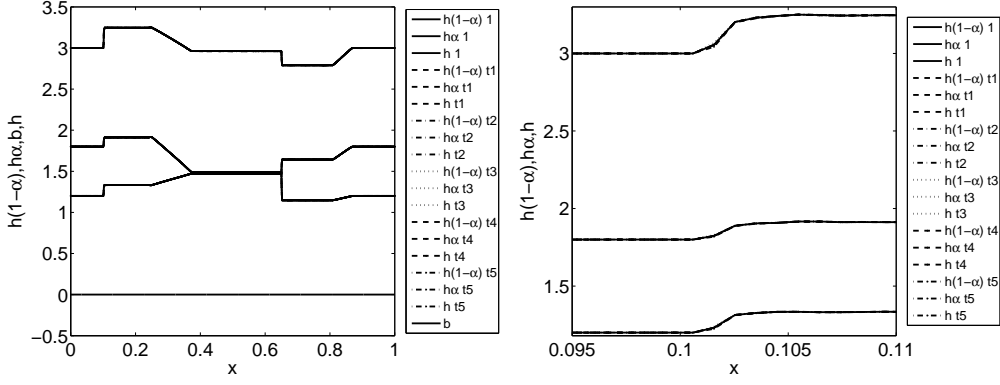
$$\begin{aligned}
 \phi_{2v1} &= U^L + \tau^2(U^R - U^L), & \phi_{2v2} &= U^R + (1 - \tau)^2(U^L - U^R), \\
 \phi_{5v1} &= U^L + \tau^5(U^R - U^L), & \phi_{5v2} &= U^R + (1 - \tau)^5(U^L - U^R), \\
 \phi_{20v1} &= U^L + \tau^{20}(U^R - U^L), & \phi_{20v2} &= U^R + (1 - \tau)^{20}(U^L - U^R).
 \end{aligned} \tag{65}$$

In Figure 10, $h(1 - \alpha)$, $h\alpha$, b and h are shown on the whole domain and also zoomed in on the left shock wave. The deviations shown in these figures are approximately also seen in the mass flow variables and the void fraction.

In these computations it is important to have a good numerical integration scheme to approximate the path integral. Incorrectly approximating the path integral results in solutions having incorrect faster or slower shock speeds. A two-point Gauss integration scheme is sufficient when taking ϕ linear or when using ϕ_{2v1} and ϕ_{2v2} . For the other paths we split the domain $[0, 1]$ into 8 nonintersecting uniform intervals and within each interval we evaluate the integral in the two Gauss points corresponding to that particular interval. To conclude for this test case, when properly integrated any choice of paths in (65) leads to the same numerical solution with only minor differences.

6.2 Toumi paths

In this section we will consider paths similar to those chosen in Toumi [15]. We will compare the solutions determined with the following five paths with



(a) The solution on the whole domain. (b) The solution zoomed in on the left shock wave.

Fig. 11. Solution of $h(1 - \alpha)$, $h\alpha$, b and h calculated on a mesh with 1024 elements at time $t = 0.175$ using the paths defined in (66).

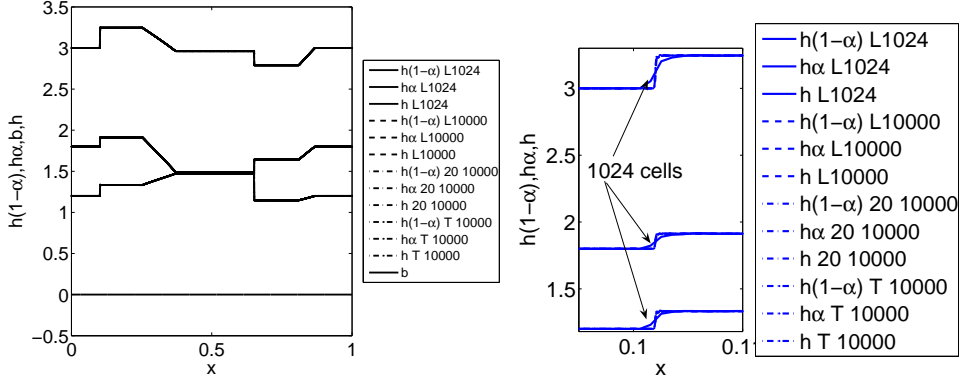
the solution determined with a linear path:

$$\begin{aligned}
\phi_{T1}(\tau; U^L, U^R) &= \begin{cases} (U_1^L + 2\tau(U_1^R - U_1^L), U_2^L, U_3^L, U_4^L, U_5^L), & \text{for } \tau \in [0, \frac{1}{2}] \\ (U_1^R, U_2^L + (2\tau - 1)(U_2^R - U_2^L), U_3^L + (2\tau - 1)(U_3^R - U_3^L), \\ U_4^L + (2\tau - 1)(U_4^R - U_4^L), U_5^L + (2\tau - 1)(U_5^R - U_5^L)), & \text{for } \tau \in [\frac{1}{2}, 1] \end{cases} \\
\phi_{T2}(\tau; U^L, U^R) &= \begin{cases} (U_1^L, U_2^L + 2\tau(U_2^R - U_2^L), U_3^L, U_4^L, U_5^L), & \text{for } \tau \in [0, \frac{1}{2}] \\ (U_1^L + (2\tau - 1)(U_1^R - U_1^L), U_2^R, U_3^L + (2\tau - 1)(U_3^R - U_3^L), \\ U_4^L + (2\tau - 1)(U_4^R - U_4^L), U_5^L + (2\tau - 1)(U_5^R - U_5^L)), & \text{for } \tau \in [\frac{1}{2}, 1] \end{cases} \\
\phi_{T3}(\tau; U^L, U^R) &= \begin{cases} (U_1^L, U_2^L, U_3^L + 2\tau(U_3^R - U_3^L), U_4^L, U_5^L), & \text{for } \tau \in [0, \frac{1}{2}] \\ (U_1^L + (2\tau - 1)(U_1^R - U_1^L), U_2^L + (2\tau - 1)(U_2^R - U_2^L), U_3^R, \\ U_4^L + (2\tau - 1)(U_4^R - U_4^L), U_5^L + (2\tau - 1)(U_5^R - U_5^L)), & \text{for } \tau \in [\frac{1}{2}, 1] \end{cases} \\
\phi_{T4}(\tau; U^L, U^R) &= \begin{cases} (U_1^L, U_2^L, U_3^L, U_4^L + 2\tau(U_4^R - U_4^L), U_5^L), & \text{for } \tau \in [0, \frac{1}{2}] \\ (U_1^L + (2\tau - 1)(U_1^R - U_1^L), U_2^L + (2\tau - 1)(U_2^R - U_2^L), \\ U_3^L + (2\tau - 1)(U_3^R - U_3^L), U_4^R, U_5^L + (2\tau - 1)(U_5^R - U_5^L)), & \text{for } \tau \in [\frac{1}{2}, 1] \end{cases} \\
\phi_{T5}(\tau; U^L, U^R) &= \begin{cases} (U_1^L, U_2^L, U_3^L, U_4^L, U_5^L + 2\tau(U_5^R - U_5^L)), & \text{for } \tau \in [0, \frac{1}{2}] \\ (U_1^L + (2\tau - 1)(U_1^R - U_1^L), U_2^L + (2\tau - 1)(U_2^R - U_2^L), \\ U_3^L + (2\tau - 1)(U_3^R - U_3^L), U_4^L + (2\tau - 1)(U_4^R - U_4^L), U_5^R), & \text{for } \tau \in [\frac{1}{2}, 1] \end{cases}
\end{aligned} \tag{66}$$

In the implementation the integrals are computed using a two-point Gauss integration rule. In Figure 11, $h(1 - \alpha)$, $h\alpha$, b and h are shown on the whole domain and also zoomed in on the left shock wave. The deviations shown in these figures are approximately also seen in the mass flow variables and the void fraction. We see that the final solution determined with the paths given in (66) are all very similar. The choice of one of these paths does not have a big effect on the final solution compared to the linear path.

6.3 Refining the mesh

As a final check we further refine our mesh. We will calculate the solution on a mesh with 10000 elements. We only do this for the linear path, ϕ_{20v1}



(a) The solution on the whole domain. (b) The solution zoomed in on the left shock wave.

Fig. 12. Solution of $h(1 - \alpha)$, $h\alpha$, b and h calculated on a mesh with 10000 elements at time $t = 0.175$ using the linear path, ϕ_{20v1} and ϕ_{T1} .

(see (65)) and ϕ_{T1} (see (66)) and compare these solutions with the numerical solution determined with the linear path on a mesh with 1024 elements. In Figure 12, $h(1 - \alpha)$, $h\alpha$, b and h are shown on the whole domain and also zoomed in on the left shock wave. The deviations shown in these figures are approximately also seen in the mass flow variables and the void fraction. To obtain these figures, the integral of the nonconservative product for each path was evaluated differently. For the linear path a two point Gauss integration scheme was used for the whole domain $[0, 1]$. For the path ϕ_{20v1} we divided the domain $[0, 1]$ into 16 nonintersecting uniform domains and within each domain we used again a two point Gauss integration scheme. For the path ϕ_{T1} the domain $[0, 1]$ was divided into 8 nonintersecting uniform domains and within each domain we used a two point Gauss integration scheme. As we see in these figures, the differences in the numerical solution for all the paths are minimal. The slight differences in the shock speed are more likely to be caused by the numerical integration scheme than the difference in the path. If we were to determine the numerical solution using the path ϕ_{20v1} by dividing the domain $[0, 1]$ into 8 nonintersecting uniform domains instead of 16, the differences in shock speed in comparison to the other paths will increase, so it is important to have a good approximation for the integral of the nonconservative product. We conclude that it is important to have a good numerical integration scheme to approximate the path integral. Using a linear path, a two points Gauss integration scheme, without refinement, suffices. We saw that it does not matter which path is chosen, but choosing the linear path, due to the simple integration scheme, is by far the cheapest and easiest choice.

7 Conclusions

In this article we have derived weak formulations for space- and space-time DGFEM for nonconservative hyperbolic partial differential equations. We also introduced the numerical flux for systems with nonconservative products (NCP flux) suitable for DGFEM thereof.

As test cases we considered the shallow water equations with topography and a simplified depth-averaged two-phase flow model. For the shallow water equations we considered rest flow over discontinuous topography and showed, both numerically and theoretically, that the rest flow is preserved. We also considered subcritical and supercritical flow over a bump. For these test cases we obtained second order accuracy in the L^2 norm for linear basis functions.

For the simplified depth-averaged two-phase flow model we also considered subcritical and supercritical flow over a bump and again obtained second order accuracy using linear basis functions. We also considered a dam-break type test case. This test case we further used to investigate the effect of the path on the numerical solution. The effect of the path was very small in the numerical solutions. Taking different paths did not lead to relevant changes in the final solution. We did see, however, that for certain paths it is not sufficient to simply use a two-point Gauss integration scheme over the whole domain of integration for the path integral, but higher order integration rules were required. This resulted in significantly larger computational cost which is undesirable. We therefore concluded that the effect of the path on the numerical solution is very small. In future computations, the simple linear path therefore seems to suffice.

A Derivation of the weak formulation for space DGFEM

In this section we derive a space DGFEM weak formulation for hyperbolic nonconservative partial differential equations. As opposed to the derivation of the weak formulation for space-time DGFEM, we now only consider fixed grids. We first introduce the function spaces and basis functions after which we derive the weak formulation.

Let $\Omega \subset \mathbb{R}^q$ be the bounded flow domain approximated by Ω_h such that $\Omega_h \rightarrow \Omega$ as $h \rightarrow 0$, with h the radius of the smallest sphere completely containing the largest element K_j . Consider approximations of $U(x, t)$ and the test function $V(x, t)$ in the finite element space defined as:

$$W_h = \left\{ V \in (L^2(\Omega_h))^m : V|_{K_j} \circ F_K \in (P^p(\hat{K}))^m \right\}, \quad (\text{A.1})$$

where m denotes the dimension of U . Polynomial approximations for the trial function U and the test function V in each element K_j are introduced:

$$U(t, \bar{x})|_{K_j} = \hat{U}_m \psi_m(\bar{x}), \text{ and } V(t, \bar{x})|_{K_j} = \hat{V}_l \psi_l(\bar{x})$$

for $m, l = 0, 1, 2, \dots, M$, where M depends on the order of accuracy and the space dimension, and where the basis functions ψ are given by:

$$\psi_m = \begin{cases} 1 & \text{for } m = 0 \\ \varphi_m(\bar{x}) - \frac{1}{|K_j|} \int_{K_j} \varphi_m(\bar{x}) dK & \text{for } m = 1, 2, \dots, M \end{cases}$$

where the functions φ in element K_j are related to the basis functions $\hat{\varphi}$ on the master element \hat{K} through the mapping F :

$$\varphi_m = \hat{\varphi}_m \circ F_K^{-1}$$

with $\hat{\varphi}_m(\xi) \in P^p(\hat{K})$ and ξ the local coordinates in the master element \hat{K} .

The weak formulation for space DGFEM can be derived in a similar manner as that for space-time DGFEM, except that now we consider fixed grids. Before discussing the space DGFEM weak formulation for equations containing nonconservative products, we first introduce as a reference the space DGFEM weak formulation for equations in conservative form (see e.g. Tassi, Bokhove and Vionnet [14]).

Consider partial differential equations in conservative form:

$$U_{i,0} + F_{ik,k} = 0, \quad \bar{x} \in \mathbb{R}^q, \quad t > 0, \quad (\text{A.2})$$

where $U \in \mathbb{R}^m$ and $F \in \mathbb{R}^m \times \mathbb{R}^q$. Using the approach discussed in Tassi, Bokhove and Vionnet [14], the space DG formulation for (A.2) can be stated as:

Find a $U \in W_h$ such that for all $V \in W_h$:

$$0 = \sum_j \int_{K_j} (V_i U_{i,0} - V_{i,k} F_{ik}) dK + \sum_{S \in \mathcal{S}_I} \int_S \llbracket V_i \rrbracket_k \{ \{ F_{ik} \} \} dS + \sum_{S \in \mathcal{S}_B} \int_S V_i^L F_{ik}^L \bar{n}_k^L dS. \quad (\text{A.3})$$

Note that at this point no numerical fluxes have been introduced yet into the DG formulation. We now continue with equations containing nonconservative products. Let $U \in W_h$ (see (A.1)). We know that the numerical solution is continuous on an element and discontinuous across a face, so, using Theorem 2, U is a weak solution to (A.2) if:

$$0 = \int_{\Omega_h} V_i U_{i,0} dK + \int_{\Omega_h} V_i d\bar{\mu}_i \quad (\text{A.4a})$$

$$\begin{aligned}
&= \sum_j \int_{K_j} V_i (U_{i,0} + G_{ikr} U_{r,k}) dK \\
&\quad + \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i \left(\int_0^1 G_{ikr}(\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau}(\tau; U^L, U^R) d\tau \bar{n}_k^L \right) d\mathcal{S}, \quad (\text{A.4b})
\end{aligned}$$

where $V \in W_h$ is an arbitrary test function. Furthermore, \widehat{V} is the value (numerical flux) of the test function V on a face \mathcal{S} . Note that Theorem 2 is applied to nonconservative products in space-time where space and time variables are not explicitly distinguished. In space DGFEM this is the case and we only need the space part of the measure in Theorem 2. This measure is denoted in (A.4a) as $\bar{\mu}_i$. The crucial point in obtaining the DG formulation is the choice of the numerical flux for the test function V . We choose the numerical flux for V such that if there exists an F such that $G_{ikr} = \partial F_{ik} / \partial U_r$, then the DG formulation for the system containing a nonconservative products reduces to the conservative space DGFEM weak formulation given by (A.3).

Theorem 4 *If the numerical flux \widehat{V} for the test function V in (A.4b) is defined as $\widehat{V} = \{\{V\}\}$, then the weak formulation (A.4b) will reduce to the conservative space DGFEM formulation (A.3) when there exists an F such that $G_{ikr} = \partial F_{ik} / \partial U_r$.*

Proof Assume there is an F such that $G_{ikr} = \partial F_{ik} / \partial U_r$. We immediately see:

$$\int_0^1 G_{ikr}(\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau}(\tau; U^L, U^R) d\tau \bar{n}_k^L = -\llbracket F_{ik} \rrbracket_k. \quad (\text{A.5})$$

Integrating by parts the volume integral in (A.4b) we obtain:

$$\begin{aligned}
0 &= \sum_k \int_{K_k} V_i U_{i,0} - V_{i,k} F_{ik} dK + \sum_k \int_{\partial K_k} V_i^L F_{ik}^L \bar{n}_k^L d(\partial K) \\
&\quad - \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i \llbracket F_{ik} \rrbracket_k d\mathcal{S}. \quad (\text{A.6})
\end{aligned}$$

Use relations (13) and (14) to write the element boundary integrals as face integrals:

$$\begin{aligned}
\sum_j \int_{\partial K_j} V_i^L F_{ik}^L \bar{n}_k^L d(\partial K) &= \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \llbracket V_i F_{ik} \rrbracket_k d\mathcal{S} + \sum_{\mathcal{S} \in \mathcal{S}_B} \int_{\mathcal{S}} V_i^L F_{ik}^L \bar{n}_k^L d\mathcal{S} \\
&= \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \left(\{\{V_i\}\} \llbracket F_{ik} \rrbracket_k + \llbracket V_i \rrbracket_k \{\{F_{ik}\}\} \right) d\mathcal{S} \\
&\quad + \sum_{\mathcal{S} \in \mathcal{S}_B} \int_{\mathcal{S}} V_i^L F_{ik}^L \bar{n}_k^L d\mathcal{S}. \quad (\text{A.7})
\end{aligned}$$

Combining (A.6) and (A.7) we obtain:

$$\begin{aligned}
0 = \sum_j \int_{K_j} V_i U_{i,0} - V_{i,k} F_{ik} dK + \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \left(\{\{V_i\}\} [F_{ik}]_k + [V_i]_k \{\{F_{ik}\}\} \right) d\mathcal{S} \\
+ \sum_{\mathcal{S} \in \mathcal{S}_B} \int_{\mathcal{S}} V_i^L F_{ik}^L \bar{n}_k^L d\mathcal{S} - \sum_{\mathcal{S} \in \mathcal{S}_I} \int_{\mathcal{S}} \widehat{V}_i [F_{ik}]_k d\mathcal{S}. \quad (\text{A.8})
\end{aligned}$$

The term $\{\{V_i\}\} [F_{ik}]_k$ is set to zero in the space DG formulation for conservative systems arguing that the formulation must be conservative. For a general nonconservative system we can not use this argument. Instead, we note that by taking $\widehat{V} = \{\{V\}\}$ on the faces \mathcal{S} , the contribution $\int_{\mathcal{S}} \{\{V_i\}\} [F_{ik}]_k d\mathcal{S}$ cancels with $-\int_{\mathcal{S}} \widehat{V}_i [F_{ik}]_k d\mathcal{S}$. We now obtain the weak formulation given by (A.3). \square

Theorem 4 allows us to finalize the derivation of the DGFEM weak formulation, similar to the space-time DG formulation, to:

Find a $U \in W_h$ such that for all $V \in W_h$:

$$\begin{aligned}
0 = \sum_j \int_{K_j} V_i (U_{i,0} + G_{ikr} U_{r,k}) dK + \sum_{\mathcal{S}} \int_{\mathcal{S}} [V_i]_k \tilde{H}_{ik}^{nc} d\mathcal{S}, \\
+ \sum_{\mathcal{S}} \int_{\mathcal{S}} \{\{V_i\}\} \left(\int_0^1 G_{ikr} (\phi_r(\tau; U^L, U^R)) \frac{\partial \phi_r}{\partial \tau}(\tau; U^L, U^R) d\tau \bar{n}_k^L \right) d\mathcal{S}. \quad (\text{A.9})
\end{aligned}$$

Note that we combined the fluxes at interior and boundary faces by using a ghost value U^R at the boundary.

Acknowledgments

We would like to thank C. M. Klaij (Twente) who contributed to the development of the numerical flux for systems with nonconservative products. O.B. acknowledges a fellowship from The Royal Netherlands Academy of Arts and Sciences (KNAW), and S.R. support from the Institute of Mechanics, Processes and Control Twente. For J.V. this research was partly funded by the ICT project BRICKS (<http://www.bsik-bricks.nl>), theme MSV1.

References

- [1] T.B. Anderson and R. Jackson, A fluid mechanical description of fluidized beds, *Ind. Eng. Chem. Fundam.* **6**, 527 (1967).
- [2] M. Castro, J.M. Gallardo and C. Parés, High order finite volume schemes based on reconstruction of states for solving hyperbolic systems with nonconservative products. Application to shallow-water systems, *Math. Comput.* **75**, 1103 (2006).

- [3] G. Dal Maso, P. G. LeFloch and F. Murat, Definition and weak stability of nonconservative products, *J. Math. Pures Appl.* **74**, 483 (1995).
- [4] D.A. Drew and R.T. Lahey, Analytical modeling of multiphase flow, in *Particulate Two-Phase flows*, edited by M.C. Roco, Butterworth-Heinemann (Boston, 1993).
- [5] H. Enwald, E. Peirano and A.E. Almstedt, Eulerian Two-phase Flow Theory Applied to Fluidization, *Int. J. Multiphase Flow* **22**, 21 (1996).
- [6] D.D. Houghton and A. Kasahara, Nonlinear Shallow Fluid Flow Over an Isolated Ridge, *Comm. on Pure and Applied Math.* **21**, 1 (1968).
- [7] K. Hutter, Y. Wang and S.P. Pudasaini, The Savage-Hutter avalanche model: how far can it be pushed?, *Phil. Trans. R. Soc. A* **363**, 1507 (2005).
- [8] C.M. Klaij, J.J.W. van der Vegt and H. van der Ven, Space-time discontinuous Galerkin method for the compressible Navier-Stokes equations, *J. Comput. Phys.*, in press.
- [9] P.G. LeFloch, Shock waves for nonlinear hyperbolic systems in nonconservative form, Report 593, Institute for Mathematics and its Applications, Minneapolis, MN, (1989).
- [10] C. Parés, Numerical methods for nonconservative hyperbolic systems: a theoretical framework, *SIAM J. Numer. Anal.* **44**, 300 (2006).
- [11] C. Parés and M. Castro, On the well-balance property of Roe's method for nonconservative hyperbolic systems. Applications to shallow-water systems, *ESIAM: Mathematical Modeling and Numerical Analysis* **38**, 821 (2004).
- [12] E.B. Pitman and L. Le, A two-fluid model for avalanche and debris flows, *Phil. Trans. R. Soc. A* **363**, 1573 (2005).
- [13] R. Saurel and R. Abgrall, A Multiphase Godunov Method for Compressible Multifluid and Multiphase Flows, *J. Comput. Phys.* **150**, 425 (1999).
- [14] P.A. Tassi, O. Bokhove and C.A. Vionnet, Space discontinuous Galerkin method for shallow water flows - kinetic and HLLC flux, and potential vorticity generation, *Advances in Water Resources*, in press.
- [15] I. Toumi, A Weak Formulation of Roe's Approximate Riemann Solver, *J. Comput. Phys.* **102**, 360 (1992).
- [16] I. Toumi and A. Kumbaro, An Approximate Linearized Riemann Solver for a Two-Fluid Model, *J. Comput. Phys.* **124**, 286 (1996).
- [17] B.G.M. van Wachem, J.C. Schouten and C.M. van den Bleek, Comparative Analysis of CFD Models of Dense Gas-Solid Systems, *AIChE Journal* **47**, 1035 (2001).
- [18] J.J.W. van der Vegt and H. van der Ven, Space-Time Discontinuous Galerkin Finite Element Method with Dynamic Grid Motion for Inviscid Compressible Flows I. General Formulation, *J. Comput. Phys.* **182**, 546 (2002).

- [19] W.P. Ziemer, Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation, Springer-Verlag (New York Inc, 1989).