Well-balanced second-order finite element approximation of the shallow water equations with friction

Jean-Luc Guermond

Department of Mathematics
Texas A&M University

NUMWAVE
Montpellier, France
Dec 11–13, 2017

Co-Authors and acknowledgments

Work done in collaboration with:
Pascal Azerad (Institut de Modélisation Mathématique de Montpellier)
Matthew Farthing (ERDC, Vicksburg, MI)
Chris Kees (ERDC, Vicksburg, MI)
Bojan Popov (Dept. Math., TAMU, TX)
Manuel Quezada (TAMU → ERDC, Vicksburg, MI)

Support:

AFOSR AIR FORCE OFFICE OF SCIENTIFIC RESEARCH ARL Lawrence Livermore National Laboratory NSF

Long term research project on hyperbolic systems (JLG+BP)

Major goals and objectives of the project

- Develop and analyze robust numerical methods for solving nonlinear phenomena such as nonlinear conservation laws, advection-dominated multi-phase flows, and free-boundary problems.
- Construct methods that guarantee some sort of maximum principle (or invariant domain property for systems), have built-in entropy dissipation, run with optimal CFL, work on arbitrary meshes in any space dimension, and are high-order accurate for smooth solutions.
- New methods must be cost-efficient and easily parallelizable.
- The above objectives must be reached by stating precise mathematical statements supported either by proofs or very strong numerical evidences.

Outline

Hyperbolic systems

Hyperbolic systems

The PDEs

- Hyperbolic system

$$\partial_t \mathbf{u} + \nabla\cdot \mathbf{f}(\mathbf{u}) = 0, \qquad (\mathbf{x}, t) \in D\times\mathbb{R}_+.$$
$$u(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \qquad \mathbf{x} \in D.$$

- D open polyhedral domain in $\mathbb{R}^d$.
- $\mathbf{f} \in \mathcal{C}^1(\mathbb{R}^m; \mathbb{R}^{m\times d})$, the flux.
- $\mathbf{u}_0$, admissible initial data.
- Periodic BCs or $\mathbf{u}_0$ has compact support (to simplify BCs)

Hyperbolic systems

The PDEs

- Hyperbolic system

$$\partial_t \mathbf{u} + \nabla\cdot\mathbf{f}(\mathbf{u}) = 0, \qquad (\mathbf{x}, t) \in D\times\mathbb{R}_+.$$
$$u(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \qquad \mathbf{x} \in D.$$

- D open polyhedral domain in $\mathbb{R}^d$.
- $\mathbf{f} \in \mathcal{C}^1(\mathbb{R}^m; \mathbb{R}^{m\times d})$, the flux.
- $\mathbf{u}_0$, admissible initial data.
- Periodic BCs or $\mathbf{u}_0$ has compact support (to simplify BCs)

Hyperbolic systems

The PDEs

- Hyperbolic system

$$\partial_t \mathbf{u} + \nabla\cdot\mathbf{f}(\mathbf{u}) = 0, \qquad (\mathbf{x}, t) \in D\times\mathbb{R}_+.$$
$$u(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \qquad \mathbf{x} \in D.$$

- D open polyhedral domain in $\mathbb{R}^d$.
- $\mathbf{f} \in \mathcal{C}^1(\mathbb{R}^m; \mathbb{R}^{m\times d})$, the flux.
- $\mathbf{u}_0$, admissible initial data.
- Periodic BCs or $\mathbf{u}_0$ has compact support (to simplify BCs)

Hyperbolic systems

The PDEs

- Hyperbolic system

$$\partial_t \mathbf{u} + \nabla\cdot\mathbf{f}(\mathbf{u}) = 0, \qquad (\mathbf{x}, t) \in D\times\mathbb{R}_+.$$
$$u(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \qquad \mathbf{x} \in D.$$

- D open polyhedral domain in $\mathbb{R}^d$.
- $\mathbf{f} \in \mathcal{C}^1(\mathbb{R}^m; \mathbb{R}^{m\times d})$, the flux.
- $\mathbf{u}_0$, admissible initial data.
- Periodic BCs or $\mathbf{u}_0$ has compact support (to simplify BCs)

Hyperbolic systems

The PDEs

- Hyperbolic system

$$\partial_t \mathbf{u} + \nabla\cdot\mathbf{f}(\mathbf{u}) = 0, \qquad (\mathbf{x}, t) \in D\times\mathbb{R}_+.$$
$$u(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x}), \qquad \mathbf{x} \in D.$$

- D open polyhedral domain in $\mathbb{R}^d$.
- $\mathbf{f} \in \mathcal{C}^1(\mathbb{R}^m; \mathbb{R}^{m\times d})$, the flux.
- $\mathbf{u}_0$, admissible initial data.
- Periodic BCs or $\mathbf{u}_0$ has compact support (to simplify BCs)

Formulation of the problem

Assumptions

- $\exists$ admissible set $\mathcal{A}$ s.t. for all $(\mathbf{u}_l, \mathbf{u}_r) \in \mathcal{A}$ the 1D Riemann problem

$$\partial_t \mathbf{v} + \partial_x(\mathbf{n}\cdot\mathbf{f}(\mathbf{v})) = 0, \quad \mathbf{v}(x,0) = \begin{cases} \mathbf{u}_l & \text{if } x < 0 \\ \mathbf{u}_r & \text{if } x > 0. \end{cases}$$

  has a unique "entropy" solution $\mathbf{u}(\mathbf{u}_l, \mathbf{u}_r)(x,t)$ for all $\mathbf{n} \in \mathbb{R}^d$, $\|\mathbf{n}\|_{\ell^2} = 1$.
- There exists an invariant set $A \subset \mathcal{A}$, i.e.,

$$\mathbf{u}(\mathbf{u}_l, \mathbf{u}_r)(x,t) \in A, \quad \forall t \geq 0, \forall x \in \mathbb{R}, \quad \forall \mathbf{u}_l, \mathbf{u}_r \in A.$$

- A is convex. (Hoff (1979, 1985), Chueh, Conley, Smoller (1973))

Formulation of the problem

Assumptions

- $\exists$ admissible set $\mathcal{A}$ s.t. for all $(\mathbf{u}_l, \mathbf{u}_r) \in \mathcal{A}$ the 1D Riemann problem

$$\partial_t \mathbf{v} + \partial_x(\mathbf{n}\cdot\mathbf{f}(\mathbf{v})) = 0, \quad \mathbf{v}(x,0) = \begin{cases} \mathbf{u}_l & \text{if } x < 0 \\ \mathbf{u}_r & \text{if } x > 0. \end{cases}$$

  has a unique "entropy" solution $\mathbf{u}(\mathbf{u}_l, \mathbf{u}_r)(x,t)$ for all $\mathbf{n} \in \mathbb{R}^d$, $\|\mathbf{n}\|_{\ell^2} = 1$.
- There exists an invariant set $A \subset \mathcal{A}$, i.e.,

$$\mathbf{u}(\mathbf{u}_l, \mathbf{u}_r)(x,t) \in A, \quad \forall t \geq 0, \forall x \in \mathbb{R}, \quad \forall \mathbf{u}_l, \mathbf{u}_r \in A.$$

- A is convex. (Hoff (1979, 1985), Chueh, Conley, Smoller (1973))

Formulation of the problem

Assumptions

- $\exists$ admissible set $\mathcal{A}$ s.t. for all $(\mathbf{u}_l, \mathbf{u}_r) \in \mathcal{A}$ the 1D Riemann problem

$$\partial_t \mathbf{v} + \partial_x(\mathbf{n}{\cdot}\mathbf{f}(\mathbf{v})) = 0, \quad \mathbf{v}(x,0) = \begin{cases} \mathbf{u}_l & \text{if } x < 0 \\ \mathbf{u}_r & \text{if } x > 0. \end{cases}$$

  has a unique "entropy" solution $\mathbf{u}(\mathbf{u}_l, \mathbf{u}_r)(x,t)$ for all $\mathbf{n} \in \mathbb{R}^d$, $\|\mathbf{n}\|_{\ell^2} = 1$.
- There exists an invariant set $A \subset \mathcal{A}$, i.e.,

$$\mathbf{u}(\mathbf{u}_l, \mathbf{u}_r)(x,t) \in A, \quad \forall t \geq 0, \forall x \in \mathbb{R}, \quad \forall \mathbf{u}_l, \mathbf{u}_r \in A.$$

- A is convex. (**Hoff (1979, 1985)**, **Chueh, Conley, Smoller (1973)**)

Approximation (time and space)

FE space/Shape functions

- $\{\mathcal{T}_h\}_{h>0}$ shape regular conforming mesh sequence
- $\{\varphi_1, \ldots, \varphi_I\}$, positive + partition of unity ($\sum_{j\in\{1:I\}} \varphi_j = 1$)
- Ex: $\mathbb{P}_1$, $\mathbb{Q}_1$, Bernstein polynomials (any degree)
- $m_i := \int_D \varphi_i \,\mathrm{d}\mathbf{x}$, lumped mass matrix ($m_i = \sum_{j\in\mathcal{I}(S_i)} \int_D \varphi_i\varphi_j \,\mathrm{d}\mathbf{x}$)

Approximation (time and space)

FE space/Shape functions

- $\{\mathcal{T}_h\}_{h>0}$ shape regular conforming mesh sequence
- $\{\varphi_1, \ldots, \varphi_I\}$, positive + partition of unity ($\sum_{j\in\{1:I\}} \varphi_j = 1$)
- Ex: $\mathbb{P}_1$, $\mathbb{Q}_1$, Bernstein polynomials (any degree)
- $m_i := \int_D \varphi_i \,\mathrm{d}\mathbf{x}$, lumped mass matrix ($m_i = \sum_{j\in\mathcal{I}(S_i)} \int_D \varphi_i\varphi_j \,\mathrm{d}\mathbf{x}$)

Approximation (time and space)

FE space/Shape functions

- $\{\mathcal{T}_h\}_{h>0}$ shape regular conforming mesh sequence
- $\{\varphi_1, \ldots, \varphi_I\}$, positive + partition of unity $(\sum_{j\in\{1:I\}} \varphi_j = 1)$
- Ex: $\mathbb{P}_1$, $\mathbb{Q}_1$, Bernstein polynomials (any degree)
- $m_i := \int_D \varphi_i \,\mathrm{d}\mathbf{x}$, lumped mass matrix $(m_i = \sum_{j\in\mathcal{I}(S_i)} \int_D \varphi_i\varphi_j \,\mathrm{d}\mathbf{x})$

Approximation (time and space)

FE space/Shape functions

- $\{\mathcal{T}_h\}_{h>0}$ shape regular conforming mesh sequence
- $\{\varphi_1, \ldots, \varphi_I\}$, positive + partition of unity $(\sum_{j\in\{1:I\}} \varphi_j = 1)$
- Ex: $\mathbb{P}_1$, $\mathbb{Q}_1$, Bernstein polynomials (any degree)
- $m_i := \int_D \varphi_i \,\mathrm{d}\mathbf{x}$, lumped mass matrix $(m_i = \sum_{j\in\mathcal{I}(S_i)} \int_D \varphi_i\varphi_j \,\mathrm{d}\mathbf{x})$

Approximation (time and space)

Algorithm: Galerkin

- Set $u_h(\mathbf{x}, t) = \sum_{j=1}^{I} \mathbf{U}_j(t)\varphi_j(\mathbf{x})$.
- Galerkin + lumped mass matrix

$$m_i \frac{\partial \mathbf{U}_i}{\partial t} + \int_D \nabla\cdot(\mathbf{f}(\mathbf{u}_h))\varphi_i \,\mathrm{d}\mathbf{x} = 0$$

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Approximate $\frac{\partial \mathbf{U}_i}{\partial_t}$ by $\frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t}$
- Approximate $\mathbf{f}(\mathbf{u}_h)$ by $\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j$

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} + \int_D \nabla\cdot\left(\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j\right)\varphi_i \,\mathrm{d}\mathbf{x} + \sum_{j\in\mathcal{I}(S_i)} d_{ij}^{\mathrm{V},n}(\mathbf{U}_i^n - \mathbf{U}_j^n) = 0.$$

- How should we choose artificial viscosity $d_{ij}^{\mathrm{V},n}$?

Approximation (time and space)

Algorithm: Galerkin

- Set $u_h(\mathbf{x},t) = \sum_{j=1}^{I} \mathbf{U}_j(t)\varphi_j(\mathbf{x})$.
- Galerkin + lumped mass matrix

$$m_i \frac{\partial \mathbf{U}_i}{\partial t} + \int_D \nabla\cdot(\mathbf{f}(\mathbf{u}_h))\varphi_i \,\mathrm{d}\mathbf{x} = 0$$

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Approximate $\frac{\partial \mathbf{U}_i}{\partial_t}$ by $\frac{\mathbf{U}_i^{n+1}-\mathbf{U}_i^n}{\Delta t}$
- Approximate $\mathbf{f}(\mathbf{u}_h)$ by $\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j$

$$m_i \frac{\mathbf{U}_i^{n+1}-\mathbf{U}_i^n}{\Delta t} + \int_D \nabla\cdot\left(\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j\right)\varphi_i \,\mathrm{d}\mathbf{x} + \sum_{j\in\mathcal{I}(S_i)} d_{ij}^{\mathrm{V},n}(\mathbf{U}_i^n - \mathbf{U}_j^n) = 0.$$

- How should we choose artificial viscosity $d_{ij}^{\mathrm{V},n}$?

Approximation (time and space)

Algorithm: Galerkin

- Set $u_h(\mathbf{x},t) = \sum_{j=1}^{I} \mathbf{U}_j(t)\varphi_j(\mathbf{x})$.
- Galerkin + lumped mass matrix

$$m_i \frac{\partial \mathbf{U}_i}{\partial t} + \int_D \nabla\cdot(\mathbf{f}(\mathbf{u}_h))\varphi_i \,\mathrm{d}\mathbf{x} = 0$$

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Approximate $\frac{\partial \mathbf{U}_i}{\partial_t}$ by $\frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t}$
- Approximate $\mathbf{f}(\mathbf{u}_h)$ by $\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j$

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} + \int_D \nabla\cdot\left(\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j\right)\varphi_i \,\mathrm{d}\mathbf{x} + \sum_{j\in\mathcal{I}(S_i)} d_{ij}^{\mathrm{V},n}(\mathbf{U}_i^n - \mathbf{U}_j^n) = 0.$$

- How should we choose artificial viscosity $d_{ij}^{\mathrm{V},n}$?

Approximation (time and space)

Algorithm: Galerkin

- Set $u_h(\mathbf{x},t) = \sum_{j=1}^{I} \mathbf{U}_j(t)\varphi_j(\mathbf{x})$.
- Galerkin + lumped mass matrix

$$m_i \frac{\partial \mathbf{U}_i}{\partial t} + \int_D \nabla\cdot(\mathbf{f}(\mathbf{u}_h))\varphi_i \,\mathrm{d}\mathbf{x} = 0$$

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Approximate $\frac{\partial \mathbf{U}_i}{\partial_t}$ by $\frac{\mathbf{U}_i^{n+1}-\mathbf{U}_i^n}{\Delta t}$
- Approximate $\mathbf{f}(\mathbf{u}_h)$ by $\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j$

$$m_i \frac{\mathbf{U}_i^{n+1}-\mathbf{U}_i^n}{\Delta t} + \int_D \nabla\cdot\left(\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j\right)\varphi_i\,\mathrm{d}\mathbf{x} + \sum_{j\in\mathcal{I}(S_i)} d_{ij}^{\mathrm{V},n}(\mathbf{U}_j^n-\mathbf{U}_i^n) = 0.$$

- How should we choose artificial viscosity $d_{ij}^{\mathrm{V},n}$?

Approximation (time and space)

Algorithm: Galerkin

- Set $u_h(\mathbf{x}, t) = \sum_{j=1}^{I} \mathbf{U}_j(t)\varphi_j(\mathbf{x})$.
- Galerkin + lumped mass matrix

$$m_i \frac{\partial \mathbf{U}_i}{\partial t} + \int_D \nabla\cdot(\mathbf{f}(\mathbf{u}_h))\varphi_i \,\mathrm{d}\mathbf{x} = 0$$

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Approximate $\frac{\partial \mathbf{U}_i}{\partial_t}$ by $\frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t}$
- Approximate $\mathbf{f}(\mathbf{u}_h)$ by $\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j$

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} + \int_D \nabla\cdot\left(\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j\right)\varphi_i \,\mathrm{d}\mathbf{x} + \sum_{j\in\mathcal{I}(S_i)} d_{ij}^{\mathrm{V},n}(\mathbf{U}_i^n - \mathbf{U}_j^n) = 0.$$

- How should we choose artificial viscosity $d_{ij}^{\mathrm{V},n}$?

Approximation (time and space)

Algorithm: Galerkin

- Set $u_h(\mathbf{x}, t) = \sum_{j=1}^{I} \mathbf{U}_j(t)\varphi_j(\mathbf{x})$.
- Galerkin + lumped mass matrix

$$m_i \frac{\partial \mathbf{U}_i}{\partial t} + \int_D \nabla\cdot(\mathbf{f}(\mathbf{u}_h))\varphi_i \, \mathrm{d}\mathbf{x} = 0$$

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Approximate $\frac{\partial \mathbf{U}_i}{\partial_t}$ by $\frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t}$
- Approximate $\mathbf{f}(\mathbf{u}_h)$ by $\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j$

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} + \int_D \nabla\cdot\left(\sum_{j\in\mathcal{I}(S_i)}(\mathbf{f}(\mathbf{U}_j^n))\varphi_j\right)\varphi_i \, \mathrm{d}\mathbf{x} + \sum_{j\in\mathcal{I}(S_i)} d_{ij}^{\mathrm{V},n}(\mathbf{U}_i^n - \mathbf{U}_j^n) = 0.$$

- How should we choose artificial viscosity $d_{ij}^{\mathrm{V},n}$?

Approximation (time and space)

Algorithm: Galerkin

- Set $u_h(\mathbf{x},t) = \sum_{j=1}^{I} \mathbf{U}_j(t)\varphi_j(\mathbf{x})$.
- Galerkin + lumped mass matrix

$$m_i \frac{\partial \mathbf{U}_i}{\partial t} + \int_D \nabla\cdot(\mathbf{f}(\mathbf{u}_h))\varphi_i \,\mathrm{d}\mathbf{x} = 0$$

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Approximate $\frac{\partial \mathbf{U}_i}{\partial_t}$ by $\frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t}$
- Approximate $\mathbf{f}(\mathbf{u}_h)$ by $\sum_{j\in\mathcal{I}(S_i)} (\mathbf{f}(\mathbf{U}_j^n))\varphi_j$

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} + \int_D \nabla\cdot\left(\sum_{j\in\mathcal{I}(S_i)} (\mathbf{f}(\mathbf{U}_j^n))\varphi_j\right)\varphi_i \,\mathrm{d}\mathbf{x} + \sum_{j\in\mathcal{I}(S_i)} d_{ij}^{\mathrm{V},n}(\mathbf{U}_i^n - \mathbf{U}_j^n) = 0.$$

- How should we choose artificial viscosity $d_{ij}^{\mathrm{V},n}$?

Local Extrema Diminishing (LED) method (easy to get it wrong)

LED is an easy (first-order) algebraic method to estimate $d_{ij}^{\mathrm{V},n}$ to achieve maximum principle; **Roe (1981) p. 361**, **Jameson (1995) §2.1** and others, see e.g., **Kuzmin et al. (2005), p. 163**, **Kuzmin, Turek (2002) Eq. (32)-(33)**.

Theorem

- *LED is exactly equivalent to choosing Roe's average velocity for scalar equations.*
- *There exist C^∞ fluxes and piecewise smooth initial data such that the LED approximation does not converge to the unique entropy solution.*

Local Extrema Diminishing (LED) method (easy to get it wrong)

LED is an easy (first-order) algebraic method to estimate $d_{ij}^{\mathrm{V},n}$ to achieve maximum principle; **Roe (1981) p. 361**, **Jameson (1995) §2.1** and others, see e.g., **Kuzmin et al. (2005), p. 163**, **Kuzmin, Turek (2002) Eq. (32)-(33)**.

Theorem

- *LED is exactly equivalent to choosing Roe's average velocity for scalar equations.*
- *There exist C^∞ fluxes and piecewise smooth initial data such that the LED approximation does not converge to the unique entropy solution.*

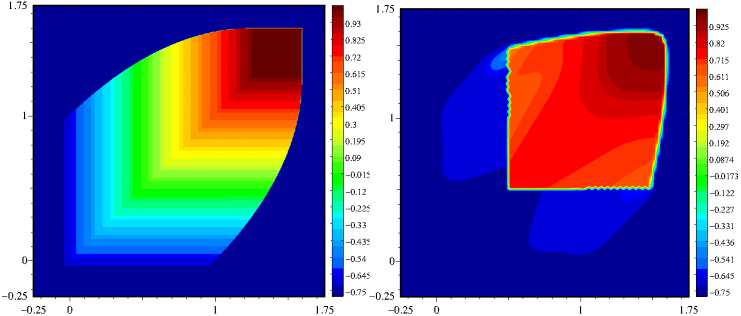
Local Extrema Diminishing (LED) method (easy to get it wrong)

LED is an easy (first-order) algebraic method to estimate $d_{ij}^{\mathrm{V},n}$ to achieve maximum principle; **Roe (1981) p. 361**, **Jameson (1995) §2.1** and others, see e.g., **Kuzmin et al. (2005), p. 163**, **Kuzmin, Turek (2002) Eq. (32)-(33)**.

Theorem

- *LED is exactly equivalent to choosing Roe's average velocity for scalar equations.*
- *There exist C^∞ fluxes and piecewise smooth initial data such that the LED approximation does not converge to the unique entropy solution.*

Local Extrema Diminishing (LED) method (easy to get it wrong)



Burger's equation. Left: $\mathbb{P}_1$ interpolant of the exact solution at $t = 0.75$; Right: piecewise linear approximation of the solution using the first-order LED scheme with 7543 grid points.

Approximation (time and space)

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Introduce

$$\mathbf{c}_{ij} = \int_D \varphi_i(\mathbf{x})\nabla\varphi_j(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

- Then

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} = \sum_j \left(-\mathbf{c}_{ij}\cdot\mathbf{f}(\mathbf{U}_j) + d_{ij}^{\mathrm{V},n}(\mathbf{U}_j - \mathbf{U}_i) \right).$$

- Observe that partition of unity implies conservation: $\sum_j \mathbf{c}_{ij} = 0$.
- We define $d_{ij}^{\mathrm{V},n}$ such that $\sum_j d_{ij}^{\mathrm{V},n} = 0$, (conservation).

Remark

- *Rest of the talk applies to any method that can be formalized as above. (FV, DG, FD, etc.)*

Approximation (time and space)

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Introduce

$$\mathbf{c}_{ij} = \int_D \varphi_i(\mathbf{x})\nabla\varphi_j(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

- Then

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} = \sum_j \left(-\mathbf{c}_{ij}\cdot\mathbf{f}(\mathbf{U}_j) + d_{ij}^{\mathrm{V},n}(\mathbf{U}_j - \mathbf{U}_i)\right).$$

- Observe that partition of unity implies conservation: $\sum_j \mathbf{c}_{ij} = 0$.
- We define $d_{ij}^{\mathrm{V},n}$ such that $\sum_j d_{ij}^{\mathrm{V},n} = 0$, (conservation).

Remark

- *Rest of the talk applies to any method that can be formalized as above. (FV, DG, FD, etc.)*

Approximation (time and space)

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Introduce

$$\mathbf{c}_{ij} = \int_D \varphi_i(\mathbf{x})\nabla\varphi_j(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

- Then

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} = \sum_j \left(-\mathbf{c}_{ij}\cdot\mathbf{f}(\mathbf{U}_j) + d_{ij}^{\mathrm{V},n}(\mathbf{U}_j - \mathbf{U}_i)\right).$$

- Observe that partition of unity implies conservation: $\sum_j \mathbf{c}_{ij} = 0$.
- We define $d_{ij}^{\mathrm{V},n}$ such that $\sum_j d_{ij}^{\mathrm{V},n} = 0$, (conservation).

Remark

- *Rest of the talk applies to any method that can be formalized as above. (FV, DG, FD, etc.)*

Approximation (time and space)

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Introduce

$$\mathbf{c}_{ij} = \int_D \varphi_i(\mathbf{x}) \nabla \varphi_j(\mathbf{x}) \, \mathrm{d}\mathbf{x}.$$

- Then

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} = \sum_j \left(-\mathbf{c}_{ij} \cdot \mathbf{f}(\mathbf{U}_j) + d_{ij}^{\mathrm{V},n} (\mathbf{U}_j - \mathbf{U}_i) \right).$$

- Observe that partition of unity implies conservation: $\sum_j \mathbf{c}_{ij} = 0$.
- We define $d_{ii}^{\mathrm{V},n}$ such that $\sum_j d_{ij}^{\mathrm{V},n} = 0$, (conservation).

Remark

- *Rest of the talk applies to any method that can be formalized as above. (FV, DG, FD, etc.)*

Approximation (time and space)

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Introduce

$$\mathbf{c}_{ij} = \int_D \varphi_i(\mathbf{x}) \nabla \varphi_j(\mathbf{x}) \, \mathrm{d}\mathbf{x}.$$

- Then

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} = \sum_j \left(-\mathbf{c}_{ij} \cdot \mathbf{f}(\mathbf{U}_j) + d_{ij}^{\mathrm{V},n} (\mathbf{U}_j - \mathbf{U}_i) \right).$$

- Observe that partition of unity implies conservation: $\sum_j \mathbf{c}_{ij} = 0$.
- We define $d_{ii}^{\mathrm{V},n}$ such that $\sum_j d_{ij}^{\mathrm{V},n} = 0$, (conservation).

Remark

- *Rest of the talk applies to any method that can be formalized as above. (FV, DG, FD, etc.)*

Approximation (time and space)

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Define $\mathbf{n}_{ij} = \mathbf{c}_{ij}/\|\mathbf{c}_{ij}\|_{\ell^2} \in \mathbb{R}^d$, (unit vector).
- $\mathbf{f}_{ij}(\mathbf{U}) := \mathbf{n}_{ij}\cdot\mathbf{f}(\mathbf{U})$ is an hyperbolic flux by definition of hyperbolicity!
- Then define

$$\overline{\mathbf{U}}(\mathbf{U}_i, \mathbf{U}_j) := \frac{1}{2}(\mathbf{U}_i + \mathbf{U}_j) + \frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}(\mathbf{f}_{ij}(\mathbf{U}_i) - \mathbf{f}_{ij}(\mathbf{U}_j)).$$

- Observe that $\mathbf{U}_i^{n+1} \in \mathrm{Conv}\{\overline{\mathbf{U}}(\mathbf{U}_i^n, \mathbf{U}_j^n) \mid j \in \mathcal{I}(S_i)\}$ up to CFL condition.

Approximation (time and space)

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Define $\mathbf{n}_{ij} = \mathbf{c}_{ij}/\|\mathbf{c}_{ij}\|_{\ell^2} \in \mathbb{R}^d$, (unit vector).
- $\mathbf{f}_{ij}(\mathbf{U}) := \mathbf{n}_{ij}\cdot\mathbf{f}(\mathbf{U})$ is an hyperbolic flux by definition of hyperbolicity!
- Then define

$$\overline{\mathbf{U}}(\mathbf{U}_i, \mathbf{U}_j) := \frac{1}{2}(\mathbf{U}_i + \mathbf{U}_j) + \frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}(\mathbf{f}_{ij}(\mathbf{U}_i) - \mathbf{f}_{ij}(\mathbf{U}_j)).$$

- Observe that $\mathbf{U}_i^{n+1} \in \mathrm{Conv}\{\overline{\mathbf{U}}(\mathbf{U}_i^n, \mathbf{U}_j^n) \mid j \in \mathcal{I}(S_i)\}$ up to CFL condition.

Approximation (time and space)

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Define $\mathbf{n}_{ij} = \mathbf{c}_{ij}/\|\mathbf{c}_{ij}\|_{\ell^2} \in \mathbb{R}^d$, (unit vector).
- $\mathbf{f}_{ij}(\mathbf{U}) := \mathbf{n}_{ij}\cdot\mathbf{f}(\mathbf{U})$ is an hyperbolic flux by definition of hyperbolicity!
- Then define

$$\overline{\mathbf{U}}(\mathbf{U}_i, \mathbf{U}_j) := \frac{1}{2}(\mathbf{U}_i + \mathbf{U}_j) + \frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}(\mathbf{f}_{ij}(\mathbf{U}_i) - \mathbf{f}_{ij}(\mathbf{U}_j)).$$

- Observe that $\mathbf{U}_i^{n+1} \in \mathrm{Conv}\{\overline{\mathbf{U}}(\mathbf{U}_i^n, \mathbf{U}_j^n) \mid j \in \mathcal{I}(S_i)\}$ up to CFL condition.

Approximation (time and space)

Algorithm: Galerkin + First-order viscosity + Explicit Euler

- Define $\mathbf{n}_{ij} = \mathbf{c}_{ij}/\|\mathbf{c}_{ij}\|_{\ell^2} \in \mathbb{R}^d$, (unit vector).
- $\mathbf{f}_{ij}(\mathbf{U}) := \mathbf{n}_{ij}\cdot\mathbf{f}(\mathbf{U})$ is an hyperbolic flux by definition of hyperbolicity!
- Then define

$$\overline{\mathbf{U}}(\mathbf{U}_i, \mathbf{U}_j) := \frac{1}{2}(\mathbf{U}_i + \mathbf{U}_j) + \frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}(\mathbf{f}_{ij}(\mathbf{U}_i) - \mathbf{f}_{ij}(\mathbf{U}_j)).$$

- Observe that $\mathbf{U}_i^{n+1} \in \mathrm{Conv}\{\overline{\mathbf{U}}(\mathbf{U}_i^n, \mathbf{U}_j^n) \mid j \in \mathcal{I}(S_i)\}$ up to CFL condition.

Approximation (time and space)

Lemma

- *Consider the fake 1D Riemann problem!*

$$\partial_t \mathbf{v} + \partial_x(\mathbf{n}_{ij}\cdot\mathbf{f}(\mathbf{v})) = 0, \quad \mathbf{v}(x,0) = \begin{cases} \mathbf{U}_i & \text{if } x < 0 \\ \mathbf{U}_j & \text{if } x > 0. \end{cases}$$

- *Let $\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)$ be maximum wave speed in 1D Riemann problem*
- *Then $\overline{\mathbf{U}}(\mathbf{U}_i, \mathbf{U}_j) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{v}(x,t)\,\mathrm{d}x$ with fake time $t = \frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}$, provided*

$$\boxed{\frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) = t\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) \leq \frac{1}{2}}$$

- Define viscosity coefficient

$$\boxed{d_{ij}^{\mathrm{V},n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i.}$$

Approximation (time and space)

Lemma

- *Consider the fake 1D Riemann problem!*

$$\partial_t \mathbf{v} + \partial_x(\mathbf{n}_{ij}\cdot\mathbf{f}(\mathbf{v})) = 0, \quad \mathbf{v}(x,0) = \begin{cases} \mathbf{U}_i & \text{if } x < 0 \\ \mathbf{U}_j & \text{if } x > 0. \end{cases}$$

- *Let $\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)$ be maximum wave speed in 1D Riemann problem*
- *Then $\overline{\mathbf{U}}(\mathbf{U}_i, \mathbf{U}_j) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{v}(x,t)\,\mathrm{d}x$ with fake time $t = \frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}$, provided*

$$\boxed{\frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}} \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) = t\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) \leq \frac{1}{2}}$$

- Define viscosity coefficient

$$\boxed{d_{ij}^{\mathrm{V},n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i.}$$

Approximation (time and space)

Lemma

- *Consider the fake 1D Riemann problem!*

$$\partial_t \mathbf{v} + \partial_x(\mathbf{n}_{ij}\cdot\mathbf{f}(\mathbf{v})) = 0, \quad \mathbf{v}(x,0) = \begin{cases} \mathbf{U}_i & \text{if } x < 0 \\ \mathbf{U}_j & \text{if } x > 0. \end{cases}$$

- *Let* $\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)$ *be maximum wave speed in 1D Riemann problem*
- *Then* $\overline{\mathbf{U}}(\mathbf{U}_i, \mathbf{U}_j) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{v}(x,t)\,\mathrm{d}x$ *with fake time* $t = \frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}$, *provided*

$$\boxed{\frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) = t\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) \leq \frac{1}{2}}$$

- Define viscosity coefficient

$$\boxed{d_{ij}^{\mathrm{V},n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i.}$$

Approximation (time and space)

Lemma

- *Consider the fake 1D Riemann problem!*

$$\partial_t \mathbf{v} + \partial_x(\mathbf{n}_{ij}\cdot\mathbf{f}(\mathbf{v})) = 0, \quad \mathbf{v}(x,0) = \begin{cases} \mathbf{U}_i & \text{if } x < 0 \\ \mathbf{U}_j & \text{if } x > 0. \end{cases}$$

- *Let* $\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)$ *be maximum wave speed in 1D Riemann problem*
- *Then* $\overline{\mathbf{U}}(\mathbf{U}_i, \mathbf{U}_j) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathbf{v}(x,t)\,\mathrm{d}x$ *with fake time* $t = \frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}$, *provided*

$$\boxed{\frac{\|\mathbf{c}_{ij}\|_{\ell^2}}{2d_{ij}^{\mathrm{V},n}}\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) = t\lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) \le \frac{1}{2}}$$

- Define viscosity coefficient

$$\boxed{d_{ij}^{\mathrm{V},n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \ne i.}$$

Approximation (time and space)

Theorem

Provided CFL condition, $(1 - 2\frac{\Delta t}{m_i}|D_{ii}|) \geq 0$.

- *Local invariance:* $\mathbf{U}_i^{n+1} \in Conv\{\overline{\mathbf{U}}(\mathbf{U}_i^n, \mathbf{U}_j^n) \mid j \in \mathcal{I}(S_i)\}$.
- *Global invariance: The scheme preserves all the convex invariant sets. (Let A be a convex invariant set, assume $\mathbf{U}_0 \in A$, then $\mathbf{U}_i^{n+1} \in A$ for all $n \geq 0$.)*
- *Discrete entropy inequality for all the entropy pairs (η, q):*

$$\frac{m_i}{\Delta t}(\eta(\mathbf{U}_i^{n+1}) - \eta(\mathbf{U}_i^n)) + \int_D \nabla\cdot(\Pi_h \mathbf{q}(\mathbf{u}_h^n))\varphi_i \,\mathrm{d}x + \sum_{i \neq j \in \mathcal{I}(S_i)} d_{ij}\eta(\mathbf{U}_j^n) \leq 0.$$

Approximation (time and space)

Theorem

Provided CFL condition, $(1 - 2\frac{\Delta t}{m_i}|D_{ii}|) \geq 0$.

- *Local invariance:* $\mathbf{U}_i^{n+1} \in Conv\{\overline{\mathbf{U}}(\mathbf{U}_i^n, \mathbf{U}_j^n) \mid j \in \mathcal{I}(S_i)\}$.
- *Global invariance: The scheme preserves all the convex invariant sets.*
  (Let A be a convex invariant set, assume $\mathbf{U}_0 \in A$, then $\mathbf{U}_i^{n+1} \in A$ for all $n \geq 0$.)
- *Discrete entropy inequality for all the entropy pairs (η, q):*

$$\frac{m_i}{\Delta t}(\eta(\mathbf{U}_i^{n+1}) - \eta(\mathbf{U}_i^n)) + \int_D \nabla\cdot(\Pi_h \mathbf{q}(\mathbf{u}_h^n))\varphi_i \,\mathrm{d}x + \sum_{i\neq j\in\mathcal{I}(S_i)} d_{ij}\eta(\mathbf{U}_j^n) \leq 0.$$

Approximation (time and space)

Theorem

Provided CFL condition, $(1 - 2\frac{\Delta t}{m_i}|D_{ii}|) \geq 0$.

- *Local invariance:* $\mathbf{U}_i^{n+1} \in Conv\{\overline{\mathbf{U}}(\mathbf{U}_i^n, \mathbf{U}_j^n) \mid j \in \mathcal{I}(S_i)\}$.
- *Global invariance: The scheme preserves all the convex invariant sets.* *(Let A be a convex invariant set, assume $\mathbf{U}_0 \in A$, then $\mathbf{U}_i^{n+1} \in A$ for all $n \geq 0$.)*
- *Discrete entropy inequality for all the entropy pairs (η, q):*

$$\frac{m_i}{\Delta t}(\eta(\mathbf{U}_i^{n+1}) - \eta(\mathbf{U}_i^n)) + \int_D \nabla\cdot(\Pi_h \mathbf{q}(\mathbf{u}_h^n))\varphi_i \,\mathrm{d}x + \sum_{i \neq j \in \mathcal{I}(S_i)} d_{ij}\eta(\mathbf{U}_j^n) \leq 0.$$

Approximation (time and space)

Theorem

Provided CFL condition, $(1 - 2\frac{\Delta t}{m_i}|D_{ii}|) \geq 0$.

- *Local invariance:* $\mathbf{U}_i^{n+1} \in Conv\{\overline{\mathbf{U}}(\mathbf{U}_i^n, \mathbf{U}_j^n) \mid j \in \mathcal{I}(S_i)\}$.
- *Global invariance: The scheme preserves all the convex invariant sets.* *(Let* A *be a convex invariant set, assume* $\mathbf{U}_0 \in A$*, then* $\mathbf{U}_i^{n+1} \in A$ *for all* $n \geq 0$*.)*
- *Discrete entropy inequality for all the entropy pairs* (η, q)*:*

$$\frac{m_i}{\Delta t}(\eta(\mathbf{U}_i^{n+1}) - \eta(\mathbf{U}_i^n)) + \int_D \nabla\cdot(\Pi_h \mathbf{q}(\mathbf{u}_h^n))\varphi_i \,\mathrm{d}x + \sum_{i \neq j \in \mathcal{I}(S_i)} d_{ij}\eta(\mathbf{U}_j^n) \leq 0.$$

Approximation (time and space)

Theorem

Provided CFL condition, $(1 - 2\frac{\Delta t}{m_i}|D_{ii}|) \geq 0$.

- *Local invariance:* $\mathbf{U}_i^{n+1} \in Conv\{\overline{\mathbf{U}}(\mathbf{U}_i^n, \mathbf{U}_j^n) \mid j \in \mathcal{I}(S_i)\}$.
- *Global invariance: The scheme preserves all the convex invariant sets.* *(Let A be a convex invariant set, assume $\mathbf{U}_0 \in A$, then $\mathbf{U}_i^{n+1} \in A$ for all $n \geq 0$.)*
- *Discrete entropy inequality for all the entropy pairs* (η, q)*:*

$$\frac{m_i}{\Delta t}(\eta(\mathbf{U}_i^{n+1}) - \eta(\mathbf{U}_i^n)) + \int_D \nabla\cdot(\Pi_h \mathbf{q}(\mathbf{u}_h^n))\varphi_i \,\mathrm{d}x + \sum_{i \neq j \in \mathcal{I}(S_i)} d_{ij}\eta(\mathbf{U}_j^n) \leq 0.$$

Approximation (time and space)

Is it new?

- Loose extension of non-staggered Lax-**Friedrichs (1954)** to FE.
- Similar results proved by **Hoff (1979, 1985)**, **Perthame-Shu (1996)**, **Frid (2001)** in FV context and compressible Euler.
- Some relation with flux vector splitting theory of **Bouchut-Frid (2006)**.
- Not aware of similar results for arbitrary hyperbolic systems and continuous FE.

Approximation (time and space)

Is it new?

- Loose extension of non-staggered Lax-**Friedrichs (1954)** to FE.
- Similar results proved by **Hoff (1979, 1985)**, **Perthame-Shu (1996)**, **Frid (2001)** in FV context and compressible Euler.
- Some relation with flux vector splitting theory of **Bouchut-Frid (2006)**.
- Not aware of similar results for arbitrary hyperbolic systems and continuous FE.

Approximation (time and space)

Is it new?

- Loose extension of non-staggered Lax-**Friedrichs (1954)** to FE.
- Similar results proved by **Hoff (1979, 1985)**, **Perthame-Shu (1996)**, **Frid (2001)** in FV context and compressible Euler.
- Some relation with flux vector splitting theory of **Bouchut-Frid (2006)**.
- Not aware of similar results for arbitrary hyperbolic systems and continuous FE.

Approximation (time and space)

Is it new?

- Loose extension of non-staggered Lax-**Friedrichs (1954)** to FE.
- Similar results proved by **Hoff (1979, 1985)**, **Perthame-Shu (1996)**, **Frid (2001)** in FV context and compressible Euler.
- Some relation with flux vector splitting theory of **Bouchut-Frid (2006)**.
- Not aware of similar results for arbitrary hyperbolic systems and continuous FE.

High-order extension

Higher-order in time

- Use SSP method to get higher-order in time.
- Strong Stability Preserving methods (SSP), **Kraaijevanger (1991)**, **Gottlieb-Shu-Tadmor (2001)**, **Spiteri-Ruuth (2002)** **Ferracina-Spijker (2005)**, **Higueras (2005)**, **etc.**:

High-order extension

Higher-order in time

- Use SSP method to get higher-order in time.
- Strong Stability Preserving methods (SSP), **Kraaijevanger (1991)**, **Gottlieb-Shu-Tadmor (2001), Spiteri-Ruuth (2002) Ferracina-Spijker (2005), Higueras (2005), etc.**:

References

- Jean-Luc Guermond and Bojan Popov. Invariant domains and first-order continuous finite element approximation for hyperbolic systems. *SIAM J. Numer. Anal.*, 54(4):2466–2489, 2016
- Jean-Luc Guermond and Bojan Popov. Fast estimation from above of the maximum wave speed in the Riemann problem for the Euler equations. *J. Comput. Phys.*, 321:908–926, 2016
- Jean-Luc Guermond and Bojan Popov. Error Estimates of a First-order Lagrange Finite Element Technique for Nonlinear Scalar Conservation Equations. *SIAM J. Numer. Anal.*, 54(1):57–85, 2016

A priori error estimate for scalar equations: A useful lemma

Lemma (Guermond, Popov (2014-16))

Assume $u_0 \in BV(\Omega)$. Let $\widetilde{u}_h : D\times[0,T] \longrightarrow \mathbb{R}$ be any approximate solution. Assume that there is Λ a bounded functional on Lipschitz functions so that $\forall k \in [u_{\min}, u_{\max}]$, $\forall \psi \in W^{1,\infty}_c(D\times[0,T];\mathbb{R}^+)$:

$$-\int_0^T\int_D \left(|\widetilde{u}_h - k|\partial_t\psi + sgn(\widetilde{u}_h - k)(\mathbf{f}(\widetilde{u}_h) - \mathbf{f}(k))\cdot\nabla\psi\right) \mathrm{d}\mathbf{x}\,\mathrm{d}t$$
$$+\|\pi_h((\widetilde{u}_h(T) - k)\bar{\pi}_h\psi(\cdot,\mathcal{T}_h))\|_{\ell^1_h} - \|\pi_h((\widetilde{u}_h(0) - k)\bar{\pi}_h\psi(\cdot,\sigma_h))\|_{\ell^1_h} \leq \Lambda(\psi),$$

where $\|\cdot\|_{\ell^1_h}$ is the discrete L^1-norm and $|T - \mathcal{T}_h| \leq \gamma\Delta t$, $|0 - \sigma_h| \leq \gamma\Delta t$, $\gamma > 0$ is a uniform constant. Then the following estimate holds

$$\boxed{\|u(\cdot,T) - \tilde{u}_h(\cdot,T)\|_{L^1(\Omega)} \leq c\left((\epsilon + h)|u_0|_{BV(\Omega)} + \Lambda^*\right)}$$

where $\Lambda^* := \sup_{0\leq t\leq T} \frac{\int_0^t\int_D \Lambda(\phi)\,\mathrm{d}y\,\mathrm{d}s}{\Gamma_\delta(t)}$, where ϕ Kruskov's kernel.

- Generalization of results by Cockburn-Gremaud (1996) and Bouchut-Perthame (1998) based on Kruskov (1970), Kuznecov (1976).

A priori error estimate for scalar equations: A useful lemma

Lemma (Guermond, Popov (2014-16))

Assume $u_0 \in BV(\Omega)$. Let $\widetilde{u}_h : D\times[0,T] \longrightarrow \mathbb{R}$ be any approximate solution. Assume that there is Λ a bounded functional on Lipschitz functions so that $\forall k \in [u_{\min}, u_{\max}]$, $\forall \psi \in W^{1,\infty}_c(D\times[0,T];\mathbb{R}^+)$:

$$-\int_0^T\int_D \left(|\widetilde{u}_h - k|\partial_t\psi + sgn(\widetilde{u}_h - k)(\mathbf{f}(\widetilde{u}_h) - \mathbf{f}(k))\cdot\nabla\psi\right) \mathrm{d}\mathbf{x}\,\mathrm{d}t$$
$$+ \|\pi_h((\widetilde{u}_h(T) - k)\bar{\pi}_h\psi(\cdot,\mathcal{T}_h))\|_{\ell^1_h} - \|\pi_h((\widetilde{u}_h(0) - k)\bar{\pi}_h\psi(\cdot,\sigma_h))\|_{\ell^1_h} \le \Lambda(\psi),$$

where $\|\cdot\|_{\ell^1_h}$ is the discrete L^1-norm and $|T - \mathcal{T}_h| \le \gamma\Delta t$, $|0 - \sigma_h| \le \gamma\Delta t$, $\gamma > 0$ is a uniform constant. Then the following estimate holds

$$\boxed{\|u(\cdot,T) - \tilde{u}_h(\cdot,T)\|_{L^1(\Omega)} \le c\left((\epsilon + h)|u_0|_{BV(\Omega)} + \Lambda^*\right)}$$

where $\Lambda^* := \sup_{0\le t\le T} \frac{\int_0^t\int_D \Lambda(\phi)\,\mathrm{d}\mathbf{y}\,\mathrm{d}s}{\Gamma_\delta(t)}$, where ϕ Kruskov's kernel.

- Generalization of results by Cockburn-Gremaud (1996) and Bouchut-Perthame (1998) based on Kruskov (1970), Kuznecov (1976).

A priori error estimate for scalar equations: A useful lemma

Lemma (Guermond, Popov (2014-16))

Assume $u_0 \in BV(\Omega)$. Let $\widetilde{u}_h : D\times[0,T] \longrightarrow \mathbb{R}$ be any approximate solution. Assume that there is Λ a bounded functional on Lipschitz functions so that $\forall k \in [u_{\min}, u_{\max}]$, $\forall \psi \in W^{1,\infty}_c(D\times[0,T];\mathbb{R}^+)$:

$$-\int_0^T\int_D \left(|\widetilde{u}_h - k|\partial_t\psi + sgn(\widetilde{u}_h - k)(\mathbf{f}(\widetilde{u}_h) - \mathbf{f}(k))\cdot\nabla\psi\right) \mathrm{d}\mathbf{x}\,\mathrm{d}t$$
$$+ \|\pi_h((\widetilde{u}_h(T) - k)\bar{\pi}_h\psi(\cdot,\mathcal{T}_h))\|_{\ell^1_h} - \|\pi_h((\widetilde{u}_h(0) - k)\bar{\pi}_h\psi(\cdot,\sigma_h))\|_{\ell^1_h} \leq \Lambda(\psi),$$

where $\|\cdot\|_{\ell^1_h}$ is the discrete L^1-norm and $|T - \mathcal{T}_h| \leq \gamma\Delta t$, $|0 - \sigma_h| \leq \gamma\Delta t$, $\gamma > 0$ is a uniform constant. Then the following estimate holds

$$\boxed{\|u(\cdot,T) - \tilde{u}_h(\cdot,T)\|_{L^1(\Omega)} \leq c\left((\epsilon + h)|u_0|_{BV(\Omega)} + \Lambda^*\right)}$$

where $\Lambda^* := \sup_{0\leq t\leq T} \frac{\int_0^t\int_D \Lambda(\phi)\,\mathrm{d}\mathbf{y}\,\mathrm{d}s}{\Gamma_\delta(t)}$, where ϕ Kruskov's kernel.

- Generalization of results by **Cockburn-Gremaud (1996)** and **Bouchut-Perthame (1998)** based on **Kruskov (1970)**, **Kuznecov (1976)**.

A priori error estimate for scalar equations

English translation

Control on all the Kruskov entropies $\Rightarrow$ Convergence estimate.

Theorem (Guermond, Popov (2014-16))

Assume $u_0 \in BV$ and $\mathbf{f}$ Lipschitz. Let u_h be the first-order viscosity solution. Then there is c_0, uniform, such that the following holds if CFL$\leq c_0$:

(i) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{2}}$ if a priori BV estimate on u_h.

(ii) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{4}}$ otherwise.

- BV estimate is trivial in 1D (Harten's lemma).
- BV estimate can be proved in nD on special meshes.
- Similar results for FV Chainais-Hillairet (1999), Eymard et al (1998)
- First error estimates for explicit continuous FE method (as far as we know).

A priori error estimate for scalar equations

English translation

Control on all the Kruskov entropies $\Rightarrow$ Convergence estimate.

Theorem (Guermond, Popov (2014-16))

Assume $u_0 \in BV$ and $\mathbf{f}$ Lipschitz. Let u_h be the first-order viscosity solution. Then there is c_0, uniform, such that the following holds if CFL$\leq c_0$:

(i) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{2}}$ if a priori BV estimate on u_h.

(ii) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{4}}$ otherwise.

- BV estimate is trivial in 1D (Harten's lemma).
- BV estimate can be proved in nD on special meshes.
- Similar results for FV **Chainais-Hillairet (1999), Eymard et al (1998)**
- First error estimates for explicit continuous FE method (as far as we know).

A priori error estimate for scalar equations

English translation

Control on all the Kruskov entropies $\Rightarrow$ Convergence estimate.

Theorem (Guermond, Popov (2014-16))

Assume $u_0 \in BV$ and $\mathbf{f}$ Lipschitz. Let u_h be the first-order viscosity solution. Then there is c_0, uniform, such that the following holds if CFL$\leq c_0$:

(i) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{2}}$ if a priori BV estimate on u_h.

(ii) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{4}}$ otherwise.

- BV estimate is trivial in 1D (Harten's lemma).
- BV estimate can be proved in nD on special meshes.
- Similar results for FV Chainais-Hillairet (1999), Eymard et al (1998)
- First error estimates for explicit continuous FE method (as far as we know).

A priori error estimate for scalar equations

English translation

Control on all the Kruskov entropies $\Rightarrow$ Convergence estimate.

Theorem (Guermond, Popov (2014-16))

Assume $u_0 \in BV$ and $\mathbf{f}$ Lipschitz. Let u_h be the first-order viscosity solution. Then there is c_0, uniform, such that the following holds if CFL$\leq c_0$:

(i) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{2}}$ if a priori BV estimate on u_h.

(ii) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{4}}$ otherwise.

- BV estimate is trivial in 1D (Harten's lemma).
- BV estimate can be proved in nD on special meshes.
- Similar results for FV **Chainais-Hillairet (1999), Eymard et al (1998)**
- First error estimates for explicit continuous FE method (as far as we know).

A priori error estimate for scalar equations

English translation

Control on all the Kruskov entropies $\Rightarrow$ Convergence estimate.

Theorem (Guermond, Popov (2014-16))

Assume $u_0 \in BV$ and $\mathbf{f}$ Lipschitz. Let u_h be the first-order viscosity solution. Then there is c_0, uniform, such that the following holds if CFL$\leq c_0$:

(i) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{2}}$ if a priori BV estimate on u_h.

(ii) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{4}}$ otherwise.

- BV estimate is trivial in 1D (Harten's lemma).
- BV estimate can be proved in nD on special meshes.
- Similar results for FV **Chainais-Hillairet (1999), Eymard et al (1998)**
- First error estimates for explicit continuous FE method (as far as we know).

A priori error estimate for scalar equations

English translation

Control on all the Kruskov entropies $\Rightarrow$ Convergence estimate.

Theorem (Guermond, Popov (2014-16))

Assume $u_0 \in BV$ and $\mathbf{f}$ Lipschitz. Let u_h be the first-order viscosity solution. Then there is c_0, uniform, such that the following holds if CFL$\leq c_0$:

(i) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{2}}$ if a priori BV estimate on u_h.

(ii) $\|u(T) - u_h(T)\|_{L^\infty((0,T);L^1)} \leq ch^{\frac{1}{4}}$ otherwise.

- BV estimate is trivial in 1D (Harten's lemma).
- BV estimate can be proved in nD on special meshes.
- Similar results for FV **Chainais-Hillairet (1999), Eymard et al (1998)**
- First error estimates for explicit continuous FE method (as far as we know).

Second-order extensions, scalar equations

Second-order, scalar equations

Scalar conservation equations (it's easy to get it wrong)

Naive approach: FCT+Galerkin

- One could think of using Flux Corrected Transport (FCT) with
  - Galerkin (high-orer)
  - above method (low-order).
- Method is high-order and satisfies the maximum principle locally.

Lemma

There exist C^∞ fluxes and piecewise smooth initial data such that under the CFL condition $1 + 2\Delta t \frac{d_{ii}^n}{m_i} \geq 0$ the approximate sequence given by Galerkin with the FCT does not converge to the unique entropy solution.

Scalar conservation equations (it's easy to get it wrong)

Naive approach: FCT+Galerkin

- One could think of using Flux Corrected Transport (FCT) with
  - Galerkin (high-orer)
  - above method (low-order).
- Method is high-order and satisfies the maximum principle locally.

Lemma

There exist C^∞ fluxes and piecewise smooth initial data such that under the CFL condition $1 + 2\Delta t \frac{d_{ii}^n}{m_i} \geq 0$ the approximate sequence given by Galerkin with the FCT does not converge to the unique entropy solution.

Scalar conservation equations (it's easy to get it wrong)

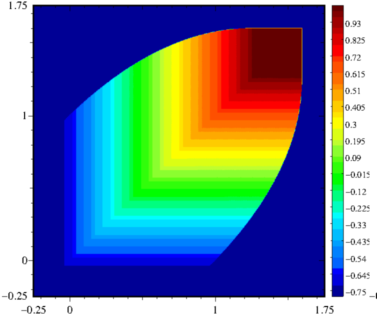
Naive approach: FCT+Galerkin

- One could think of using Flux Corrected Transport (FCT) with
  - Galerkin (high-orer)
  - above method (low-order).
- Method is high-order and satisfies the maximum principle locally.

Lemma

There exist C^∞ fluxes and piecewise smooth initial data such that under the CFL condition $1 + 2\Delta t \frac{d_{ii}^n}{m_i} \geq 0$ the approximate sequence given by Galerkin with the FCT does not converge to the unique entropy solution.

Scalar conservation equations (it's easy to get it wrong)



Burger's equation. Left: $\mathbb{P}_1$ interpolant of the exact solution at $t = 0.75$; Right: piecewise linear approximation of the solution using Galerkin+FCT with 474189 grid points.

Scalar conservation equations

Smoothness indicator

- Let $\beta_{ij} > 0$ (arbitrary for the time being).
- Define $\alpha_i^n \in [0,1]$

$$\alpha_i^n := \frac{\left|\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}(\mathsf{U}_j^n - \mathsf{U}_i^n)\right|}{\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}|\mathsf{U}_j^n - \mathsf{U}_i^n|},$$

- Let $d_{ij}^{V,n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i.$
- Set $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_j^n))d_{ij}^{V,n}$, $\psi \in C^{0,1}([0,1];[0,1])$ arbitrary with $\psi(1) = 1$.

literature

Smoothness indicator idea can be found in **Jameson et al., (1981) Eq. (12)**, **Jameson, (2017) p. 6**, **Burman, (2007), Thm. 4.1**

Scalar conservation equations

Smoothness indicator

- Let $\beta_{ij} > 0$ (arbitrary for the time being).
- Define $\alpha_i^n \in [0,1]$

$$\alpha_i^n := \frac{\left|\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}(\mathsf{U}_j^n - \mathsf{U}_i^n)\right|}{\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}|\mathsf{U}_j^n - \mathsf{U}_i^n|},$$

- Let $d_{ij}^{V,n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i.$
- Set $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_j^n)) d_{ij}^{V,n}$, $\psi \in C^{0,1}([0,1];[0,1])$ arbitrary with $\psi(1) = 1$.

literature

Smoothness indicator idea can be found in **Jameson et al., (1981) Eq. (12)**, **Jameson, (2017) p. 6**, **Burman, (2007), Thm. 4.1**

Scalar conservation equations

Smoothness indicator

- Let $\beta_{ij} > 0$ (arbitrary for the time being).
- Define $\alpha_i^n \in [0,1]$

$$\alpha_i^n := \frac{\left|\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}(\mathsf{U}_j^n - \mathsf{U}_i^n)\right|}{\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}|\mathsf{U}_j^n - \mathsf{U}_i^n|},$$

- Let $d_{ij}^{V,n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i.$
- Set $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_j^n))d_{ij}^{V,n}$, $\psi \in C^{0,1}([0,1];[0,1])$ arbitrary with $\psi(1) = 1$.

literature

Smoothness indicator idea can be found in **Jameson et al., (1981) Eq. (12)**, **Jameson, (2017) p. 6**, **Burman, (2007), Thm. 4.1**

Scalar conservation equations

Smoothness indicator

- Let $\beta_{ij} > 0$ (arbitrary for the time being).
- Define $\alpha_i^n \in [0,1]$

$$\alpha_i^n := \frac{\left|\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}(\mathsf{U}_j^n - \mathsf{U}_i^n)\right|}{\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}|\mathsf{U}_j^n - \mathsf{U}_i^n|},$$

- Let $d_{ij}^{V,n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i.$
- Set $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_i^n))d_{ij}^{V,n}$, $\psi \in C^{0,1}([0,1];[0,1])$ arbitrary with $\psi(1) = 1$.

literature

Smoothness indicator idea can be found in **Jameson et al., (1981) Eq. (12)**, **Jameson, (2017) p. 6**, **Burman, (2007), Thm. 4.1**

Scalar conservation equations

Smoothness indicator

- Let $\beta_{ij} > 0$ (arbitrary for the time being).
- Define $\alpha_i^n \in [0,1]$

$$\alpha_i^n := \frac{\left|\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}(\mathsf{U}_j^n - \mathsf{U}_i^n)\right|}{\sum_{j\in\mathcal{I}(S_i)} \beta_{ij}|\mathsf{U}_j^n - \mathsf{U}_i^n|},$$

- Let $d_{ij}^{V,n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i.$
- Set $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_j^n))d_{ij}^{V,n}$, $\psi \in C^{0,1}([0,1];[0,1])$ arbitrary with $\psi(1) = 1$.

literature

Smoothness indicator idea can be found in **Jameson et al., (1981) Eq. (12)**, **Jameson, (2017) p. 6**, **Burman, (2007), Thm. 4.1**

Scalar conservation equations

Heuristics

- α_i should be $\mathcal{O}(h^2)$ (away from extremas).

Generalized barycentric coordinates

- In 1D, $h_i = x_i - x_{i-1}$, $h_{i+1} = x_{i+1} - x_i$ and set $\beta_{i,i-1} = \frac{h_i}{h_i + h_{i+1}}$ and $\beta_{i,i+1} = \frac{h_{i+1}}{h_i + h_{i+1}}$.
- Multi-D, $\{\beta_{ij}\}_{j \in \mathcal{I}(S_i)}$ are generalized barycentric coordinates (**Floater (2015)**, **Warren et al. (2007)**).

Scalar conservation equations

Heuristics

- α_i should be $\mathcal{O}(h^2)$ (away from extremas).

Generalized barycentric coordinates

- In 1D, $h_i = x_i - x_{i-1}$, $h_{i+1} = x_{i+1} - x_i$ and set $\beta_{i,i-1} = \frac{h_i}{h_i + h_{i+1}}$ and $\beta_{i,i+1} = \frac{h_{i+1}}{h_i + h_{i+1}}$.
- Multi-D, $\{\beta_{ij}\}_{j \in \mathcal{I}(S_i)}$ are generalized barycentric coordinates (**Floater (2015)**, **Warren et al. (2007)**).

Scalar conservation equations

Heuristics

- α_i should be $\mathcal{O}(h^2)$ (away from extremas).

Generalized barycentric coordinates

- In 1D, $h_i = x_i - x_{i-1}$, $h_{i+1} = x_{i+1} - x_i$ and set $\beta_{i,i-1} = \frac{h_i}{h_i + h_{i+1}}$ and $\beta_{i,i+1} = \frac{h_{i+1}}{h_i + h_{i+1}}$.
- Multi-D, $\{\beta_{ij}\}_{j \in \mathcal{I}(S_i)}$ are generalized barycentric coordinates (**Floater (2015)**, **Warren et al. (2007)**).

Scalar conservation equations

Theorem

Let $\psi \in C^{0,1}([0,1];[0,1])$ *be any positive function such that* $\psi(1) = 1$. *Then, the scheme using* $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_j^n)) d_{ij}^{V,n}$ *is locally maximum principle preserving under a local CFL condition that depends on the Lipschitz constant of* ψ.

Extension to hyperbolic systems: positivity

Mass, energy, water height, etc

- Assume hyperbolic system has an equation like $\partial_t \rho + \nabla \cdot \mathbf{q} = 0$.
- Assume that the PDE enforces $\mathbf{q}/\rho < \infty$.
- Set $d_{ij}^{V,n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i$.
- Then define $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_j^n)) d_{ij}^{V,n}$, $\psi \in C^{0,1}([0,1];[0,1])$ arbitrary with $\psi(1) = 1$.

Extension to hyperbolic systems: positivity

Mass, energy, water height, etc

- Assume hyperbolic system has an equation like $\partial_t \rho + \nabla \cdot \mathbf{q} = 0$.
- Assume that the PDE enforces $\mathbf{q}/\rho < \infty$.
- Set $d_{ij}^{V,n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) \|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i.$
- Then define $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_j^n)) d_{ij}^{V,n}$, $\psi \in C^{0,1}([0,1];[0,1])$ arbitrary with $\psi(1) = 1$.

Extension to hyperbolic systems: positivity

Mass, energy, water height, etc

- Assume hyperbolic system has an equation like $\partial_t \rho + \nabla \cdot \mathbf{q} = 0$.
- Assume that the PDE enforces $\mathbf{q}/\rho < \infty$.
- Set $d_{ij}^{V,n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j) \|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i$.
- Then define $d_{ij}^{n} = \max(\psi(\alpha_i^n), \psi(\alpha_j^n)) d_{ij}^{V,n}$, $\psi \in C^{0,1}([0,1];[0,1])$ arbitrary with $\psi(1) = 1$.

Extension to hyperbolic systems: positivity

Mass, energy, water height, etc

- Assume hyperbolic system has an equation like $\partial_t \rho + \nabla{\cdot}\mathbf{q} = 0$.
- Assume that the PDE enforces $\mathbf{q}/\rho < \infty$.
- Set $d_{ij}^{V,n} := \lambda_{\max}(\mathbf{f}, \mathbf{n}_{ij}, \mathbf{U}_i, \mathbf{U}_j)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad j \neq i$.
- Then define $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_j^n)) d_{ij}^{V,n}$, $\psi \in C^{0,1}([0,1];[0,1])$ arbitrary with $\psi(1) = 1$.

Hyperbolic systems, positivity

Theorem

Let $\psi \in C^{0,1}([0,1];[0,1])$ be any positive function such that $\psi(1) = 1$. The scheme using $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_j^n))d_{ij}^{V,n}$ is positivity preserving preserving under a local CFL condition.

Mass, energy, water height, etc

- Euler: density, (internal energy).
- Shallow water: water height.

Hyperbolic systems, positivity

Theorem

Let $\psi \in C^{0,1}([0,1];[0,1])$ be any positive function such that $\psi(1) = 1$. The scheme using $d_{ij}^n = \max(\psi(\alpha_i^n), \psi(\alpha_i^n)) d_{ij}^{V,n}$ is positivity preserving preserving under a local CFL condition.

Mass, energy, water height, etc

- Euler: density, (internal energy).
- Shallow water: water height.

References

- J.-L. Guermond and Bojan Popov. Invariant domains and second-order continuous finite element approximation for scalar conservation equations. *SIAM J. Numer. Anal.* In press
- J.-L. Guermond, Bojan Popov, , M. Nazarov, and Ignacio Tomas. Second-order invariant domain preserving approximation of the euler equations using convex limiting. Submitted SIAM SISC

Outline

Shallow water

Shallow water

Shallow water equations

- Conservation of mass and momentum:

$$\partial_t \mathbf{u} + \nabla\cdot\mathbf{f}(\mathbf{u}) + \begin{pmatrix} 0 \\ g\mathrm{h}\nabla z \end{pmatrix} = \mathbf{S}(\mathbf{u}), \quad x \in D, t \in \mathbb{R}_+$$

  where $z : D \to \mathbb{R}$ is the bathymetry, g is the gravity constant, h is the water height,

- Flux

$$\mathbf{f}(\mathbf{u}) = \begin{pmatrix} \mathbf{q}^{\mathsf{T}} \\ \frac{1}{\mathrm{h}}\mathbf{q}\otimes\mathbf{q} + \frac{1}{2}g\mathrm{h}^2\mathbb{I}_d \end{pmatrix} \in \mathbb{R}^{(1+d)\times d},$$

  $\mathbf{q}$ is the discharge.

- Manning friction

$$\mathbf{S}(\mathbf{u}) = -gn^2\mathrm{h}^{-\gamma}\mathbf{q}\|\mathbf{v}\|_{\ell^2}.$$

  We take $\gamma = \frac{4}{3}$.

Shallow water

Shallow water equations

- Conservation of mass and momentum:

$$\partial_t \mathbf{u} + \nabla\cdot\mathbf{f}(\mathbf{u}) + \begin{pmatrix} 0 \\ g\mathrm{h}\nabla z \end{pmatrix} = \mathbf{S}(\mathbf{u}), \quad x \in D, t \in \mathbb{R}_+$$

  where $z : D \to \mathbb{R}$ is the bathymetry, g is the gravity constant, h is the water height,

- Flux

$$\mathbf{f}(\mathbf{u}) = \begin{pmatrix} \mathbf{q}^{\mathrm{T}} \\ \frac{1}{\mathrm{h}}\mathbf{q}\otimes\mathbf{q} + \frac{1}{2}g\mathrm{h}^2\mathbb{I}_d \end{pmatrix} \in \mathbb{R}^{(1+d)\times d},$$

  $\mathbf{q}$ is the discharge.

- Manning friction

$$\mathbf{S}(\mathbf{u}) = -gn^2\mathrm{h}^{-\gamma}\mathbf{q}\|\mathbf{v}\|_{\ell^2}.$$

  We take $\gamma = \frac{4}{3}$.

Shallow water

Shallow water equations

- Conservation of mass and momentum:

$$\partial_t \mathbf{u} + \nabla\cdot\mathbf{f}(\mathbf{u}) + \begin{pmatrix} 0 \\ g\mathrm{h}\nabla z \end{pmatrix} = \mathbf{S}(\mathbf{u}), \quad x \in D, t \in \mathbb{R}_+$$

  where $z : D \to \mathbb{R}$ is the bathymetry, g is the gravity constant, h is the water height,

- Flux

$$\mathbf{f}(\mathbf{u}) = \begin{pmatrix} \mathbf{q}^{\mathrm{T}} \\ \frac{1}{\mathrm{h}}\mathbf{q}\otimes\mathbf{q} + \frac{1}{2}g\mathrm{h}^2\mathbb{I}_d \end{pmatrix} \in \mathbb{R}^{(1+d)\times d},$$

  $\mathbf{q}$ is the discharge.

- Manning friction

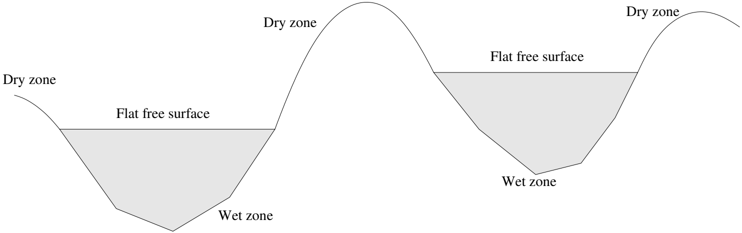
$$\mathbf{S}(\mathbf{u}) = -gn^2\mathrm{h}^{-\gamma}\mathbf{q}\|\mathbf{v}\|_{\ell^2}.$$

  We take $\gamma = \frac{4}{3}$.

Shallow water

Definition (Well balancing at rest)

A numerical scheme is said to be well-balanced at rest if rest states are invariant by the scheme, (i.e., rest is exactly preserved).



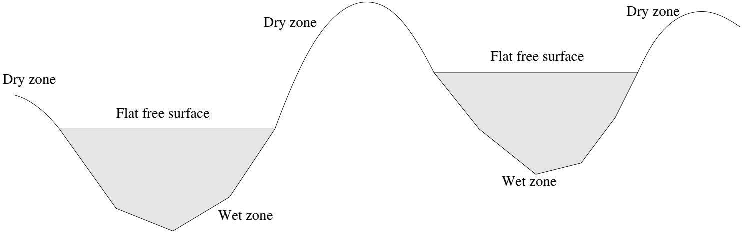
English translation

If discharge is zero and $h + z = \text{const}$, nothing should move.

Shallow water

Definition (Well balancing at rest)

A numerical scheme is said to be well-balanced at rest if rest states are invariant by the scheme, (i.e., rest is exactly preserved).



English translation

If discharge is zero and $\mathrm{h} + z = \text{const}$, nothing should move.

Well-balancing

Definition (Well balancing for sliding steady state)

- Assume that the bottom is a plane with two tangent orthonormal vectors $\mathbf{t}_1, \mathbf{t}_2$ with $\mathbf{t}_2$ being horizontal and $\mathbf{t}_1$ pointing downward, i.e., $\nabla z = -b\mathbf{t}_1$ with $b > 0$.
- A numerical scheme is said to be well-balanced for sliding steady states if it preserves the following steady state solution:

$$\mathbf{q}(\mathbf{x}, t)\cdot\mathbf{t}_2 = 0, \quad \mathbf{q}(\mathbf{x}, t)\cdot\mathbf{t}_1 = q_0, \quad \mathrm{h}(\mathbf{x}, t) = \mathrm{h}_0 := \left(\frac{n^2 q_0^2}{b}\right)^{\frac{1}{2+\gamma}}.$$

Bermudez, Vazquez (1994), **Greenberg, Leroux (1996)**

Key questions

Key questions

- Preserving rest states and sliding states (well-balancing)
- Water height h must stay positive.
- $\mathbf{S}(\mathbf{u})$ is singular h $\to 0$. How this term should be handled to make a positive explicit scheme?
- Can significant results produced by the abundant FV and DG literature be reproduced with continuous FE without invoking ad hoc linear stabilization (à la SUPG, egde stabilization, subgrid stabilization, etc.), i.e., without ad hoc parameters.

Key questions

Key questions

- Preserving rest states and sliding states (well-balancing)
- Water height h must stay positive.
- $\mathbf{S}(\mathbf{u})$ is singular $\mathrm{h} \to 0$. How this term should be handled to make a positive explicit scheme?
- Can significant results produced by the abundant FV and DG literature be reproduced with continuous FE without invoking ad hoc linear stabilization (à la SUPG, egde stabilization, subgrid stabilization, etc.), i.e., without ad hoc parameters.

Key questions

Key questions

- Preserving rest states and sliding states (well-balancing)
- Water height h must stay positive.
- $\mathbf{S}(\mathbf{u})$ is singular $h \to 0$. How this term should be handled to make a positive explicit scheme?
- Can significant results produced by the abundant FV and DG literature be reproduced with continuous FE without invoking ad hoc linear stabilization (à la SUPG, egde stabilization, subgrid stabilization, etc.), i.e., without ad hoc parameters.

Key questions

Key questions

- Preserving rest states and sliding states (well-balancing)
- Water height h must stay positive.
- $\mathbf{S}(\mathbf{u})$ is singular $\mathrm{h} \to 0$. How this term should be handled to make a positive explicit scheme?
- Can significant results produced by the abundant FV and DG literature be reproduced with continuous FE without invoking ad hoc linear stabilization (à la SUPG, egde stabilization, subgrid stabilization, etc.), i.e., without ad hoc parameters.

Well-balancing

Well-balanced, positivity preserving, second-order FE scheme

$$
\begin{aligned}
m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} &= \sum_{j \in \mathcal{I}(D_i)} -\mathbf{g}(\mathbf{U}_j^n) \cdot \mathbf{c}_{ij} - \begin{pmatrix} 0 \\ g\mathrm{H}_i^n(\mathrm{H}_j^n + Z_j)\mathbf{c}_{ij} \end{pmatrix} \\
&- \left(0, \frac{2gn^2 \mathbf{Q}_i^n \|\mathbf{V}_i\|_{\ell^2} m_i}{(\mathrm{H}_i^n)^\gamma + \max((\mathrm{H}_i^n)^\gamma, 2gn^2 \Delta t \|\mathbf{V}_i\|_{\ell^2})}\right)^{\mathrm{T}} \\
&+ \sum_{i \neq j \in \mathcal{I}(D_i)} ((d_{ij}^n - \mu_{ij}^n)(\mathbf{U}_j^{*,i,n} - \mathbf{U}_i^{*,j,n}) + \mu_{ij}^n(\mathbf{U}_j^n - \mathbf{U}_i^n))
\end{aligned}
$$

Hydrostatic reconstruction (Audusse et al. (2004))

$$
\begin{aligned}
\mathrm{H}_i^{*,j} &:= \max(0, \mathrm{H}_i + \mathrm{Z}_i - \max(\mathrm{Z}_i, \mathrm{Z}_j)), \\
\mathbf{U}_j^{*,i,n} &:= \mathbf{U}_j \frac{\mathrm{H}_j^{*,i}}{\mathrm{H}_j}.
\end{aligned}
$$

Well-balancing

Well-balanced, positivity preserving, second-order FE scheme

$$m_i \frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} = \sum_{j \in \mathcal{I}(D_i)} -\mathbf{g}(\mathbf{U}_j^n) \cdot \mathbf{c}_{ij} - \begin{pmatrix} 0 \\ g\mathsf{H}_i^n(\mathsf{H}_j^n + Z_j)\mathbf{c}_{ij} \end{pmatrix}$$

$$- \left(0, \frac{2gn^2 \mathbf{Q}_i^n \|\mathbf{V}_i\|_{\ell^2} m_i}{(\mathsf{H}_i^n)^\gamma + \max((\mathsf{H}_i^n)^\gamma, 2gn^2 \Delta t \|\mathbf{V}_i\|_{\ell^2})}\right)^{\mathrm{T}}$$

$$+ \sum_{i \neq j \in \mathcal{I}(D_i)} \left((d_{ij}^n - \mu_{ij}^n)(\mathbf{U}_j^{*,i,n} - \mathbf{U}_i^{*,j,n}) + \mu_{ij}^n (\mathbf{U}_j^n - \mathbf{U}_i^n)\right)$$

Hydrostatic reconstruction (Audusse et al. (2004)

$$\mathsf{H}_i^{*,j} := \max(0, \mathsf{H}_i + \mathsf{Z}_i - \max(\mathsf{Z}_i, \mathsf{Z}_j)),$$

$$\mathbf{U}_j^{*,i,n} := \mathbf{U}_j \frac{\mathsf{H}_j^{*,i}}{\mathsf{H}_j}.$$

Well-balancing

Theorem

- *The scheme is well-balanced w.r.t. rest states.*
- *The scheme is well-balanced w.r.t. sliding states if the following alternative holds:* (i) *mesh is centro-symmetric and fine enough, and the artificial viscosity is defined so that* d_{ij}^n *and* μ_{ij}^n *are constant when* $\mathbf{U}_j^n = \mathbf{U}_i^n$ *for all* $j \in \mathcal{I}(D_i)$*; or* (ii) *the mesh is non-uniform but the artificial viscosity is defined so that* $d_{ij}^n = 0$ *and* $\mu_{ij}^n = 0$ *when* $\mathbf{U}_i^n = \mathbf{U}_j^n$ *for all* $j \in \mathcal{I}(D_i)$ *and all* $i \in \{1{:}I\}$.

Well-balancing

Theorem

- *The scheme is well-balanced w.r.t. rest states.*
- *The scheme is well-balanced w.r.t. sliding states if the following alternative holds:* (i) *mesh is centro-symmetric and fine enough, and the artificial viscosity is defined so that* d_{ij}^n *and* μ_{ij}^n *are constant when* $\mathbf{U}_j^n = \mathbf{U}_i^n$ *for all* $j \in \mathcal{I}(D_i)$*; or* (ii) *the mesh is non-uniform but the artificial viscosity is defined so that* $d_{ij}^n = 0$ *and* $\mu_{ij}^n = 0$ *when* $\mathbf{U}_i^n = \mathbf{U}_j^n$ *for all* $j \in \mathcal{I}(D_i)$ *and all* $i \in \{1{:}I\}$.

Smoothness-based viscosities

- Conservation of mass:

$$\partial_t h + \nabla\cdot\mathbf{q} = 0.$$

  and $\mathbf{q}/h$ is bounded.
- Use smoothness-based viscosity defined for scalar conservation equations:

$$\mu_{ij}^n := \max(\psi_i^n, \psi_j^n) \max((\mathbf{V}_i\cdot\mathbf{n}_{ij})_-, (\mathbf{V}_j\cdot\mathbf{n}_{ij})_+)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad i \neq j.$$

$$d_{ij}^n = \max(\psi_i^n, \psi_j^n) d_{ij}^V$$

Smoothness-based viscosities

- Conservation of mass:

$$\partial_t h + \nabla\cdot\mathbf{q} = 0.$$

  and $\mathbf{q}/h$ is bounded.
- Use smoothness-based viscosity defined for scalar conservation equations:

$$\mu_{ij}^n := \max(\psi_i^n, \psi_j^n) \max((\mathbf{V}_i\cdot\mathbf{n}_{ij})_-, (\mathbf{V}_j\cdot\mathbf{n}_{ij})_+)\|\mathbf{c}_{ij}\|_{\ell^2}, \quad i \neq j.$$
$$d_{ij}^n = \max(\psi_i^n, \psi_j^n) d_{ij}^v$$

Smoothness-based viscosities

Theorem

The Scheme is positive.

The scheme is also formally second-order in space.

Smoothness-based viscosities

Theorem

The Scheme is positive.

The scheme is also formally second-order in space.

Entropy-viscosity for hyperbolic systems

- Smoothness-based viscosity limited to second-order.

- Use viscosity based on entropy residual or commutator.
- Let $(\eta, \mathbf{F})$ be an entropy pair.
- Set $\epsilon = 10^{-\frac{16}{2}}$ and define

$$\eta_i^{\min} := \min_{j\in\mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \eta_i^{\max} := \max_{j\in\mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \epsilon_i := \epsilon \max(|\eta_i^{\max}|, |\eta_i^{\min}|).$$

- Define entropy residual

$$R_i^n = \frac{1}{\max(|\eta_i^{\max} - \eta_i^{\min}|, \epsilon_i)} \sum_{j\in\mathcal{I}(D_i)} (\mathbf{F}(\mathbf{U}_j^n) - \eta'(\mathbf{U}_i^n)^{\mathrm{T}}\mathbf{f}(\mathbf{U}_j^n))\cdot\mathbf{c}_{ij}. \tag{1}$$

- R_i^n is a commutator that measures smoothness of entropy.
- $(\eta, \mathbf{F})$ does not need to be an entropy pair for the system with source.

Entropy-viscosity for hyperbolic systems

- Smoothness-based viscosity limited to second-order.

- Use viscosity based on entropy residual or commutator.
- Let $(\eta, \mathbf{F})$ be an entropy pair.
- Set $\epsilon = 10^{-\frac{16}{2}}$ and define

$$\eta_i^{\min} := \min_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \eta_i^{\max} := \max_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \epsilon_i := \epsilon \max(|\eta_i^{\max}|, |\eta_i^{\min}|).$$

- Define entropy residual

$$R_i^n = \frac{1}{\max(|\eta_i^{\max} - \eta_i^{\min}|, \epsilon_i)} \sum_{j \in \mathcal{I}(D_i)} (\mathbf{F}(\mathbf{U}_j^n) - \eta'(\mathbf{U}_i^n)^{\mathrm{T}} \mathbf{f}(\mathbf{U}_j^n)) \cdot \mathbf{c}_{ij}. \qquad (1)$$

- R_i^n is a commutator that measures smoothness of entropy.
- $(\eta, \mathbf{F})$ does not need to be an entropy pair for the system with source.

Entropy-viscosity for hyperbolic systems

- Smoothness-based viscosity limited to second-order.

- Use viscosity based on entropy residual or commutator.
- Let $(\eta, \mathbf{F})$ be an entropy pair.
- Set $\epsilon = 10^{-\frac{16}{2}}$ and define

$$\eta_i^{\min} := \min_{j\in\mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \eta_i^{\max} := \max_{j\in\mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \epsilon_i := \epsilon \max(|\eta_i^{\max}|, |\eta_i^{\min}|).$$

- Define entropy residual

$$R_i^n = \frac{1}{\max(|\eta_i^{\max} - \eta_i^{\min}|, \epsilon_i)} \sum_{j\in\mathcal{I}(D_i)} (\mathbf{F}(\mathbf{U}_j^n) - \eta'(\mathbf{U}_i^n)^{\mathrm{T}} \mathbf{f}(\mathbf{U}_j^n)) \cdot \mathbf{c}_{ij}. \qquad (1)$$

- R_i^n is a commutator that measures smoothness of entropy.
- $(\eta, \mathbf{F})$ does not need to be an entropy pair for the system with source.

Entropy-viscosity for hyperbolic systems

- Smoothness-based viscosity limited to second-order.

- Use viscosity based on entropy residual or commutator.
- Let $(\eta, \mathbf{F})$ be an entropy pair.
- Set $\epsilon = 10^{-\frac{16}{2}}$ and define

$$\eta_i^{\min} := \min_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \eta_i^{\max} := \max_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \epsilon_i := \epsilon \max(|\eta_i^{\max}|, |\eta_i^{\min}|).$$

- Define entropy residual

$$R_i^n = \frac{1}{\max(|\eta_i^{\max} - \eta_i^{\min}|, \epsilon_i)} \sum_{j \in \mathcal{I}(D_i)} (\mathbf{F}(\mathbf{U}_j^n) - \eta'(\mathbf{U}_i^n)^{\mathrm{T}} \mathbf{f}(\mathbf{U}_j^n)) \cdot \mathbf{c}_{ij}. \qquad (1)$$

- R_i^n is a commutator that measures smoothness of entropy.
- $(\eta, \mathbf{F})$ does not need to be an entropy pair for the system with source.

Entropy-viscosity for hyperbolic systems

- Smoothness-based viscosity limited to second-order.

- Use viscosity based on entropy residual or commutator.
- Let $(\eta, \mathbf{F})$ be an entropy pair.
- Set $\epsilon = 10^{-\frac{16}{2}}$ and define

$$\eta_i^{\min} := \min_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \eta_i^{\max} := \max_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \epsilon_i := \epsilon \max(|\eta_i^{\max}|, |\eta_i^{\min}|).$$

- Define entropy residual

$$R_i^n = \frac{1}{\max(|\eta_i^{\max} - \eta_i^{\min}|, \epsilon_i)} \sum_{j \in \mathcal{I}(D_i)} (\mathbf{F}(\mathbf{U}_j^n) - \eta'(\mathbf{U}_i^n)^{\mathrm{T}} \mathbf{f}(\mathbf{U}_j^n)) \cdot \mathbf{c}_{ij}. \qquad (1)$$

- R_i^n is a commutator that measures smoothness of entropy.
- $(\eta, \mathbf{F})$ does not need to be an entropy pair for the system with source.

Entropy-viscosity for hyperbolic systems

- Smoothness-based viscosity limited to second-order.

- Use viscosity based on entropy residual or commutator.
- Let $(\eta, \mathbf{F})$ be an entropy pair.
- Set $\epsilon = 10^{-\frac{16}{2}}$ and define

$$\eta_i^{\min} := \min_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \eta_i^{\max} := \max_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \epsilon_i := \epsilon \max(|\eta_i^{\max}|, |\eta_i^{\min}|).$$

- Define entropy residual

$$R_i^n = \frac{1}{\max(|\eta_i^{\max} - \eta_i^{\min}|, \epsilon_i)} \sum_{j \in \mathcal{I}(D_i)} (\mathbf{F}(\mathbf{U}_j^n) - \eta'(\mathbf{U}_i^n)^{\mathrm{T}} \mathbf{f}(\mathbf{U}_j^n)) \cdot \mathbf{c}_{ij}. \tag{1}$$

- R_i^n is a commutator that measures smoothness of entropy.
- $(\eta, \mathbf{F})$ does not need to be an entropy pair for the system with source.

Entropy-viscosity for hyperbolic systems

- Smoothness-based viscosity limited to second-order.

- Use viscosity based on entropy residual or commutator.
- Let $(\eta, \mathbf{F})$ be an entropy pair.
- Set $\epsilon = 10^{-\frac{16}{2}}$ and define

$$\eta_i^{\min} := \min_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \eta_i^{\max} := \max_{j \in \mathcal{I}(D_i)} \eta(\mathbf{U}_j^n), \qquad \epsilon_i := \epsilon \max(|\eta_i^{\max}|, |\eta_i^{\min}|).$$

- Define entropy residual

$$R_i^n = \frac{1}{\max(|\eta_i^{\max} - \eta_i^{\min}|, \epsilon_i)} \sum_{j \in \mathcal{I}(D_i)} (\mathbf{F}(\mathbf{U}_j^n) - \eta'(\mathbf{U}_i^n)^{\mathrm{T}} \mathbf{f}(\mathbf{U}_j^n)) \cdot \mathbf{c}_{ij}. \tag{1}$$

- R_i^n is a commutator that measures smoothness of entropy.
- $(\eta, \mathbf{F})$ does not need to be an entropy pair for the system with source.

Entropy-viscosity for hyperbolic systems

Example 1

$$\eta(\mathbf{u}) = g(\tfrac{1}{2}h^2 + hz) + \tfrac{1}{2}h\|\mathbf{v}\|_{\ell^2}^2, \quad F(\mathbf{u}) = (\tfrac{1}{2}h\|\mathbf{v}\|_{\ell^2}^2 + g(h^2 + hz))\mathbf{v},$$

$$|R_i^n| := \frac{1}{\Delta\eta_i^n} \sum_{j\in\mathcal{I}(D_i)} \Big(\mathbf{F}(\mathbf{U}_j^n) - \nabla\eta(\mathbf{U}_i^n)\cdot\mathbf{f}(\mathbf{U}_j^n)\Big)\cdot\mathbf{c}_{ij} - g\mathrm{H}_i^n\mathrm{Z}_j\mathbf{V}_j^n\cdot\mathbf{c}_{ij}.$$

Example 2, flat bottom

$$\eta^{\text{flat}}(\mathbf{u}) = g\tfrac{1}{2}h^2 + \tfrac{1}{2}h\|\mathbf{v}\|_{\ell^2}^2, \quad \mathbf{F}^{\text{flat}}(\mathbf{u}) = (\tfrac{1}{2}h\|\mathbf{v}\|_{\ell^2}^2 + gh^2)\mathbf{v},$$

$$|R_i^n| := \frac{1}{2\Delta\eta_i^{\text{flat},n}} \sum_{j\in\mathcal{I}(D_i)} \Big(\mathbf{F}^{\text{flat}}(\mathbf{U}_j^n) - \nabla\eta^{\text{flat}}(\mathbf{U}_i^n)\cdot\mathbf{f}(\mathbf{U}_j^n)\Big)\cdot\mathbf{c}_{ij}.$$

Then the numerical (E)ntropy (V)iscosities are defined as follows:

$$\mu_{ij}^{\text{EV},n} := \min\left(\mu_{ij}^{\text{V},n}, \max(|R_i^n|, |R_j^n|)\right),$$

$$d_{ij}^{\text{EV},n} := \min\left(d_{ij}^{\text{V},n}, \max(|R_i^n|, |R_j^n|)\right),$$

Entropy-viscosity for hyperbolic systems

Example 1

$$\eta(\mathbf{u}) = g(\tfrac{1}{2}h^2 + hz) + \tfrac{1}{2}h\|\mathbf{v}\|^2_{\ell^2}, \quad F(\mathbf{u}) = (\tfrac{1}{2}h\|\mathbf{v}\|^2_{\ell^2} + g(h^2 + hz))\mathbf{v},$$

$$|R_i^n| := \frac{1}{\Delta\eta_i^n} \sum_{j\in\mathcal{I}(D_i)} \left(\mathbf{F}(\mathbf{U}_j^n) - \nabla\eta(\mathbf{U}_i^n)\cdot\mathbf{f}(\mathbf{U}_j^n)\right)\cdot\mathbf{c}_{ij} - g\mathrm{H}_i^n\mathrm{Z}_j\mathbf{V}_j^n\cdot\mathbf{c}_{ij}.$$

Example 2, flat bottom

$$\eta^{\text{flat}}(\mathbf{u}) = g\tfrac{1}{2}h^2 + \tfrac{1}{2}h\|\mathbf{v}\|^2_{\ell^2}, \quad \mathbf{F}^{\text{flat}}(\mathbf{u}) = (\tfrac{1}{2}h\|\mathbf{v}\|^2_{\ell^2} + gh^2)\mathbf{v},$$

$$|R_i^n| := \frac{1}{2\Delta\eta_i^{\text{flat},n}} \sum_{j\in\mathcal{I}(D_i)} \left(\mathbf{F}^{\text{flat}}(\mathbf{U}_j^n) - \nabla\eta^{\text{flat}}(\mathbf{U}_i^n)\cdot\mathbf{f}(\mathbf{U}_j^n)\right)\cdot\mathbf{c}_{ij}.$$

Then the numerical (E)ntropy (V)iscosities are defined as follows:

$$\mu_{ij}^{\mathrm{EV},n} := \min\left(\mu_{ij}^{\mathrm{V},n}, \max(|R_i^n|, |R_j^n|)\right),$$

$$d_{ij}^{\mathrm{EV},n} := \min\left(d_{ij}^{\mathrm{V},n}, \max(|R_i^n|, |R_j^n|)\right),$$

Entropy-viscosity for hyperbolic systems

Example 1

$$\eta(\mathbf{u}) = g(\tfrac{1}{2}h^2 + hz) + \tfrac{1}{2}h\|\mathbf{v}\|_{\ell^2}^2, \quad F(\mathbf{u}) = (\tfrac{1}{2}h\|\mathbf{v}\|_{\ell^2}^2 + g(h^2 + hz))\mathbf{v},$$

$$|R_i^n| := \frac{1}{\Delta\eta_i^n} \sum_{j\in\mathcal{I}(D_i)} \Big(\mathbf{F}(\mathbf{U}_j^n) - \nabla\eta(\mathbf{U}_i^n)\cdot\mathbf{f}(\mathbf{U}_j^n)\Big)\cdot\mathbf{c}_{ij} - g\mathrm{H}_i^n\mathrm{Z}_j\mathbf{V}_j^n\cdot\mathbf{c}_{ij}.$$

Example 2, flat bottom

$$\eta^{\text{flat}}(\mathbf{u}) = g\tfrac{1}{2}h^2 + \tfrac{1}{2}h\|\mathbf{v}\|_{\ell^2}^2, \quad \mathbf{F}^{\text{flat}}(\mathbf{u}) = (\tfrac{1}{2}h\|\mathbf{v}\|_{\ell^2}^2 + gh^2)\mathbf{v},$$

$$|R_i^n| := \frac{1}{2\Delta\eta_i^{\text{flat},n}} \sum_{j\in\mathcal{I}(D_i)} \Big(\mathbf{F}^{\text{flat}}(\mathbf{U}_j^n) - \nabla\eta^{\text{flat}}(\mathbf{U}_i^n)\cdot\mathbf{f}(\mathbf{U}_j^n)\Big)\cdot\mathbf{c}_{ij}.$$

Then the numerical (E)ntropy (V)iscosities are defined as follows:

$$\mu_{ij}^{\mathrm{EV},n} := \min\Big(\mu_{ij}^{\mathrm{V},n}, \max(|R_i^n|, |R_j^n|)\Big),$$

$$d_{ij}^{\mathrm{EV},n} := \min\Big(d_{ij}^{\mathrm{V},n}, \max(|R_i^n|, |R_j^n|)\Big),$$

Limiting

- What should be limited?
- What should be the bounds?
- How can that be done?

Limiting

- Principle: high-order minus low-order solution gives

$$m_i(\mathbf{U}_i^{\mathrm{H},n+1} - \mathbf{U}_i^{\mathrm{L},n+1}) = \sum_{j\in\mathcal{I}(D_i)} \mathbf{A}_{ij}.$$

  Observe that $\mathbf{A}_{ij} = -\mathbf{A}_{ji}$. This implies mass conservation: $\sum_{i\in\{1:I\}} m_i \mathbf{U}_i^{\mathrm{H},n+1} = \sum_{i\in\{1:I\}} m_i \mathbf{U}_i^{\mathrm{L},n+1}$.

- Introduce limiter $0 \le \ell_{ij} = \ell_{ji} \le 1$

$$m_i(\mathbf{U}_i^{n+1} - \mathbf{U}_i^{\mathrm{L},n+1}) = \sum_{j\in\mathcal{I}(D_i)} \ell_{ij}\mathbf{A}_{ij}.$$

  Symmetry $\ell_{ij} = \ell_{ji}$ implies mass conservation: $\sum_{i\in\{1:I\}} m_i \mathbf{U}_i^{n+1} = \sum_{i\in\{1:I\}} m_i \mathbf{U}_i^{\mathrm{L},n+1}$.

Limiting

- Principle: high-order minus low-order solution gives

$$m_i(\mathbf{U}_i^{\mathrm{H},n+1} - \mathbf{U}_i^{\mathrm{L},n+1}) = \sum_{j\in\mathcal{I}(D_i)} \mathbf{A}_{ij}.$$

  Observe that $\mathbf{A}_{ij} = -\mathbf{A}_{ji}$. This implies mass conservation: $\sum_{i\in\{1:I\}} m_i\mathbf{U}_i^{\mathrm{H},n+1} = \sum_{i\in\{1:I\}} m_i\mathbf{U}_i^{\mathrm{L},n+1}$.

- Introduce limiter $0 \le \ell_{ij} = \ell_{ji} \le 1$

$$m_i(\mathbf{U}_i^{\mathrm{n}+1} - \mathbf{U}_i^{\mathrm{L},n+1}) = \sum_{j\in\mathcal{I}(D_i)} \ell_{ij}\mathbf{A}_{ij}.$$

  Symmetry $\ell_{ij} = \ell_{ji}$ implies mass conservation: $\sum_{i\in\{1:I\}} m_i\mathbf{U}_i^{n+1} = \sum_{i\in\{1:I\}} m_i\mathbf{U}_i^{\mathrm{L},n+1}$.

Limiting with exact bounds

- Set

$$\mathbf{S}_i^n := \left(0, \frac{-2gn^2\mathbf{Q}_i^n\|\mathbf{V}_i\|_{\ell^2} m_i}{(\mathrm{H}_i^n)^\gamma + \max((\mathrm{H}_i^n)^\gamma, 2gn^2\Delta t\|\mathbf{V}_i\|_{\ell^2})} + \sum_{j\in\mathcal{I}(D_i)} g(-\mathrm{H}_i^n Z_j + \frac{(\mathrm{H}_j-\mathrm{H}_i)^2}{2})\mathbf{c}_{ij}\right)^{\mathrm{T}},$$

- Define auxiliary states:

$$\overline{\mathbf{U}_{ij}^n} = -\frac{\mathbf{c}_{ij}}{2d_{ij}^{\mathrm{V},n}}\cdot(\mathbf{g}(\mathbf{U}_j^n) - \mathbf{g}(\mathbf{U}_i^n)) + \frac{1}{2}(\mathbf{U}_j^n + \mathbf{U}_i^n),$$

$$\widetilde{\mathbf{U}_{ij}^n} = \frac{d_{ij}^{\mathrm{V},n} - \mu_{ij}^{\mathrm{V},n}}{2d_{ij}^{\mathrm{V},n}}(\mathbf{U}_j^{*,i,n} - \mathbf{U}_j^n - (\mathbf{U}_i^{*,j,n} - \mathbf{U}_i^n)).$$

Lemma

Let $\mathbf{W}_i^{\mathrm{L},n+1} := \mathbf{U}_i^{\mathrm{L},n+1} - \frac{\Delta t}{m_i}\mathbf{S}_i^n$, then, if $1 - \frac{2\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} d_{ij}^{\mathrm{V},n} \geq 0$, the following convex combination holds true:

$$\mathbf{W}_i^{\mathrm{L},n+1} = \mathbf{U}_i^n\left(1 - \frac{\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} 2d_{ij}^{\mathrm{V},n}\right) + \frac{\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} 2d_{ij}^{\mathrm{V},n}\left(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}\right).$$

Furthermore we have $\overline{\mathrm{H}_{ij}^n} + \widetilde{\mathrm{H}_{ij}^n} \geq 0$ for all $j \in \mathcal{I}(D_i)$.

Limiting with exact bounds

- Set

$$\mathbf{S}_i^n := \left(0, \frac{-2gn^2\mathbf{Q}_i^n\|\mathbf{V}_i\|_{\ell^2} m_i}{(\mathrm{H}_i^n)^\gamma + \max((\mathrm{H}_i^n)^\gamma, 2gn^2\Delta t\|\mathbf{V}_i\|_{\ell^2})} + \sum_{j\in\mathcal{I}(D_i)} g(-\mathrm{H}_i^n Z_j + \frac{(\mathrm{H}_j - \mathrm{H}_i)^2}{2})\mathbf{c}_{ij}\right)^{\mathrm{T}},$$

- Define auxiliary states:

$$\overline{\mathbf{U}_{ij}^n} = -\frac{\mathbf{c}_{ij}}{2d_{ij}^{\mathrm{V},n}} \cdot (\mathbf{g}(\mathbf{U}_j^n) - \mathbf{g}(\mathbf{U}_i^n)) + \frac{1}{2}(\mathbf{U}_j^n + \mathbf{U}_i^n),$$

$$\widetilde{\mathbf{U}_{ij}^n} = \frac{d_{ij}^{\mathrm{V},n} - \mu_{ij}^{\mathrm{V},n}}{2d_{ij}^{\mathrm{V},n}}(\mathbf{U}_j^{*,i,n} - \mathbf{U}_j^n - (\mathbf{U}_i^{*,j,n} - \mathbf{U}_i^n)).$$

Lemma

Let $\mathbf{W}_i^{\mathrm{L},n+1} := \mathbf{U}_i^{\mathrm{L},n+1} - \frac{\Delta t}{m_i}\mathbf{S}_i^n$, then, if $1 - \frac{2\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} d_{ij}^{\mathrm{V},n} \geq 0$, the following convex combination holds true:

$$\mathbf{W}_i^{\mathrm{L},n+1} = \mathbf{U}_i^n\left(1 - \frac{\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} 2d_{ij}^{\mathrm{V},n}\right) + \frac{\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} 2d_{ij}^{\mathrm{V},n}\left(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}\right).$$

Furthermore we have $\overline{\mathrm{H}_{ij}^n} + \widetilde{\mathrm{H}_{ij}^n} \geq 0$ for all $j \in \mathcal{I}(D_i)$.

Limiting with exact bounds

- Set

$$\mathbf{S}_i^n := \left(0, \frac{-2gn^2\mathbf{Q}_i^n\|\mathbf{V}_i\|_{\ell^2}m_i}{(\mathrm{H}_i^n)^\gamma + \max((\mathrm{H}_i^n)^\gamma, 2gn^2\Delta t\|\mathbf{V}_i\|_{\ell^2})} + \sum_{j\in\mathcal{I}(D_i)} g(-\mathrm{H}_i^n Z_j + \frac{(\mathrm{H}_j - \mathrm{H}_i)^2}{2})\mathbf{c}_{ij}\right)^{\mathrm{T}},$$

- Define auxiliary states:

$$\overline{\mathbf{U}_{ij}^n} = -\frac{\mathbf{c}_{ij}}{2d_{ij}^{\mathrm{V},n}}\cdot(\mathbf{g}(\mathbf{U}_j^n) - \mathbf{g}(\mathbf{U}_i^n)) + \frac{1}{2}(\mathbf{U}_j^n + \mathbf{U}_i^n),$$

$$\widetilde{\mathbf{U}_{ij}^n} = \frac{d_{ij}^{\mathrm{V},n} - \mu_{ij}^{\mathrm{V},n}}{2d_{ij}^{\mathrm{V},n}}(\mathbf{U}_j^{*,i,n} - \mathbf{U}_j^n - (\mathbf{U}_i^{*,j,n} - \mathbf{U}_i^n)).$$

Lemma

Let $\mathbf{W}_i^{\mathrm{L},n+1} := \mathbf{U}_i^{\mathrm{L},n+1} - \frac{\Delta t}{m_i}\mathbf{S}_i^n$, *then, if* $1 - \frac{2\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} d_{ij}^{\mathrm{V},n} \geq 0$, *the following convex combination holds true:*

$$\mathbf{W}_i^{\mathrm{L},n+1} = \mathbf{U}_i^n\Big(1 - \frac{\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} 2d_{ij}^{\mathrm{V},n}\Big) + \frac{\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} 2d_{ij}^{\mathrm{V},n}\left(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}\right).$$

Furthermore we have $\overline{\mathrm{H}_{ij}^n} + \widetilde{\mathrm{H}_{ij}^n} \geq 0$ *for all* $j \in \mathcal{I}(D_i)$.

Limiting with exact bounds

Corollary (Water height)

The following holds true:

$$0 \le \mathsf{H}_i^{\min} := \min_{j\in\mathcal{I}(D_i)} \left(\overline{\mathsf{H}_{ij}^n} + \widetilde{\mathsf{H}_{ij}^n}\right) \le \mathsf{H}_i^{\mathsf{L},n+1} \le \max_{j\in\mathcal{I}(D_i)} \left(\overline{\mathsf{H}_{ij}^n} + \widetilde{\mathsf{H}_{ij}^n}\right) =: \mathsf{H}_i^{\max},$$

Corollary (Convex constraint)

The following holds true for any quasiconcave function Ψ:

$$\min_{j\in\mathcal{I}(D_i)} \Psi\left(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}\right) \le \Psi(\mathbf{U}_i^{\mathsf{L},n+1}).$$

Limiting with exact bounds

Corollary (Water height)

The following holds true:

$$0 \le \mathsf{H}_i^{\min} := \min_{j \in \mathcal{I}(D_i)} \left(\overline{\mathsf{H}_{ij}^n} + \widetilde{\mathsf{H}_{ij}^n} \right) \le \mathsf{H}_i^{\mathrm{L},n+1} \le \max_{j \in \mathcal{I}(D_i)} \left(\overline{\mathsf{H}_{ij}^n} + \widetilde{\mathsf{H}_{ij}^n} \right) =: \mathsf{H}_i^{\max},$$

Corollary (Convex constraint)

The following holds true for any quasiconcave function Ψ:

$$\min_{j \in \mathcal{I}(D_i)} \Psi\left(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n} \right) \le \Psi(\mathbf{U}_i^{\mathrm{L},n+1}).$$

Limiting with exact bounds

- Dry state indicator to detect these regions:
  $\mathsf{H}_i^{\mathrm{dry}} := \mathsf{H}_i^{\mathrm{L},n} - \frac{1}{2}(\max_{j\in\mathcal{I}(D_i)} \mathsf{H}_j^n - \min_{j\in\mathcal{I}(D_i)} \mathsf{H}_j^n)$.
- Limit the water height as follows:

$$Q_i^- := m_i(\mathsf{H}_i^{\min} - \mathsf{H}_i^{\mathrm{L},n+1}), \qquad Q_i^+ := m_i(\mathsf{H}_i^{\max} - \mathsf{H}_i^{\mathrm{L},n+1}),$$

$$P_i^- := \sum_{i\neq j\in\mathcal{I}(D_i)} (\mathbf{A}_{ij}^{\mathrm{h}})_-, \qquad P_i^+ := \sum_{i\neq j\in\mathcal{I}(D_i)} (\mathbf{A}_{ij}^{\mathrm{h}})_+,$$

$$R_i^- := \begin{cases} 0 & \text{if } \mathsf{H}_i^{\mathrm{dry}} \le 0, \\ 1 & \text{if } P_i = 0,\ \mathsf{H}_i^{\mathrm{dry}} > 0, \\ \frac{Q_i^-}{P_i^-} & \text{if } P_i \neq 0,\ \mathsf{H}_i^{\mathrm{dry}} > 0. \end{cases} \qquad R_i^+ := \begin{cases} 0 & \text{if } \mathsf{H}_i^{\mathrm{dry}} \le 0, \\ 1 & \text{if } P_i = 0,\ \mathsf{H}_i^{\mathrm{dry}} > 0, \\ \frac{Q_i^+}{P_i^+} & \text{if } P_i \neq 0,\ \mathsf{H}_i^{\mathrm{dry}} > 0. \end{cases}$$

$$\ell_{ij} := \min(R_i^+, R_j^-),\ \text{if } \mathbf{A}_{ij}^{\mathrm{h}} \ge 0, \qquad \ell_{ij} := \min(R_i^-, R_j^+),\ \text{if } \mathbf{A}_{ij}^{\mathrm{h}} < 0.$$

Lemma

Under the CFL condition $1 - \frac{2\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} d_{ij}^{\mathrm{L},n} \ge 0$, *the update satisfies the bounds* $0 \le \mathsf{H}_i^{\min} \le \mathsf{H}_i^{n+1} \le \mathsf{H}_i^{\max}$.

Limiting with exact bounds

- Dry state indicator to detect these regions: $\mathsf{H}_i^{\mathrm{dry}} := \mathsf{H}_i^{\mathrm{L},n} - \frac{1}{2}(\max_{j\in\mathcal{I}(D_i)} \mathsf{H}_j^n - \min_{j\in\mathcal{I}(D_i)} \mathsf{H}_j^n)$.
- Limit the water height as follows:

$$Q_i^- := m_i(\mathsf{H}_i^{\min} - \mathsf{H}_i^{\mathrm{L},n+1}), \qquad Q_i^+ := m_i(\mathsf{H}_i^{\max} - \mathsf{H}_i^{\mathrm{L},n+1}),$$

$$P_i^- := \sum_{i\neq j\in\mathcal{I}(D_i)} (\mathbf{A}_{ij}^{\mathrm{h}})_-, \qquad P_i^+ := \sum_{i\neq j\in\mathcal{I}(D_i)} (\mathbf{A}_{ij}^{\mathrm{h}})_+,$$

$$R_i^- := \begin{cases} 0 & \text{if } \mathsf{H}_i^{\mathrm{dry}} \le 0, \\ 1 & \text{if } P_i = 0,\ \mathsf{H}_i^{\mathrm{dry}} > 0, \\ \frac{Q_i^-}{P_i^-} & \text{if } P_i \neq 0,\ \mathsf{H}_i^{\mathrm{dry}} > 0. \end{cases} \qquad R_i^+ := \begin{cases} 0 & \text{if } \mathsf{H}_i^{\mathrm{dry}} \le 0, \\ 1 & \text{if } P_i = 0,\ \mathsf{H}_i^{\mathrm{dry}} > 0, \\ \frac{Q_i^+}{P_i^+} & \text{if } P_i \neq 0,\ \mathsf{H}_i^{\mathrm{dry}} > 0. \end{cases}$$

$$\ell_{ij} := \min(R_i^+, R_j^-), \text{ if } \mathbf{A}_{ij}^{\mathrm{h}} \ge 0, \qquad \ell_{ij} := \min(R_i^-, R_j^+), \text{ if } \mathbf{A}_{ij}^{\mathrm{h}} < 0.$$

Lemma

Under the CFL condition $1 - \frac{2\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} d_{ij}^{\mathrm{L},n} \ge 0$*, the update satisfies the bounds* $0 \le \mathsf{H}_i^{\min} \le \mathsf{H}_i^{n+1} \le \mathsf{H}_i^{\max}$.

Limiting with exact bounds

- Dry state indicator to detect these regions:
  $\mathsf{H}_i^{\text{dry}} := \mathsf{H}_i^{\mathsf{L},n} - \frac{1}{2}(\max_{j\in\mathcal{I}(D_i)} \mathsf{H}_j^n - \min_{j\in\mathcal{I}(D_i)} \mathsf{H}_j^n)$.
- Limit the water height as follows:

$$Q_i^- := m_i(\mathsf{H}_i^{\min} - \mathsf{H}_i^{\mathsf{L},n+1}), \qquad Q_i^+ := m_i(\mathsf{H}_i^{\max} - \mathsf{H}_i^{\mathsf{L},n+1}),$$

$$P_i^- := \sum_{i\neq j\in\mathcal{I}(D_i)} (\mathbf{A}_{ij}^{\mathsf{h}})_-, \qquad P_i^+ := \sum_{i\neq j\in\mathcal{I}(D_i)} (\mathbf{A}_{ij}^{\mathsf{h}})_+,$$

$$R_i^- := \begin{cases} 0 & \text{if } \mathsf{H}_i^{\text{dry}} \leq 0, \\ 1 & \text{if } P_i = 0,\ \mathsf{H}_i^{\text{dry}} > 0, \\ \frac{Q_i^-}{P_i^-} & \text{if } P_i \neq 0,\ \mathsf{H}_i^{\text{dry}} > 0. \end{cases} \qquad R_i^+ := \begin{cases} 0 & \text{if } \mathsf{H}_i^{\text{dry}} \leq 0, \\ 1 & \text{if } P_i = 0,\ \mathsf{H}_i^{\text{dry}} > 0, \\ \frac{Q_i^+}{P_i^+} & \text{if } P_i \neq 0,\ \mathsf{H}_i^{\text{dry}} > 0. \end{cases}$$

$$\ell_{ij} := \min(R_i^+, R_j^-),\ \text{if } \mathbf{A}_{ij}^{\mathsf{h}} \geq 0, \qquad \ell_{ij} := \min(R_i^-, R_j^+),\ \text{if } \mathbf{A}_{ij}^{\mathsf{h}} < 0.$$

Lemma

Under the CFL condition $1 - \frac{2\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} d_{ij}^{\mathsf{L},n} \geq 0$, *the update satisfies the bounds* $0 \leq \mathsf{H}_i^{\min} \leq \mathsf{H}_i^{n+1} \leq \mathsf{H}_i^{\max}$.

Limiting with exact bounds

- One can limit kinetic energy (for instance) $\psi(\mathbf{W}) := \frac{1}{2}\frac{1}{\mathsf{H}(\mathbf{W})}\|\mathbf{Q}(\mathbf{W})\|^2_{\ell^2}$.
- The following holds:

$$\psi(\mathbf{W}_i^{\mathsf{L},n+1}) \le \max_{j\in\mathcal{I}(D_i)} \psi(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}) =: \mathsf{K}_i^{\max}.$$

- Let $\lambda_j := \frac{1}{\mathrm{card}(\mathcal{I}(D_i))-1}$, $j \in \mathcal{I}(D_i)\setminus\{i\}$. Set

$$\mathsf{H}_i^{\mathsf{W},\mathsf{L}} := \mathsf{H}(\mathbf{W}_i^{\mathsf{L},n+1}), \qquad \mathbf{Q}_i^{\mathsf{W},\mathsf{L}} := \mathbf{Q}(\mathbf{W}_i^{\mathsf{L},n+1}), \tag{2}$$

$$\mathbf{P}_{ij} := (\mathbf{P}_{ij}^{\mathsf{h}}, \mathbf{P}_{ij}^{\mathsf{q}})^{\mathsf{T}} := \frac{1}{m_i\lambda_j}\mathbf{A}_{ij}, \qquad a := -\frac{1}{2}\|\mathbf{P}_{ij}^{\mathsf{h}}\|^2_{\ell^2}, \tag{3}$$

$$b := \mathsf{K}_i^{\max}\mathbf{P}_{ij}^{\mathsf{h}} - 2\mathbf{Q}_i^{\mathsf{W},\mathsf{L}}\cdot\mathbf{P}_{ij}^{\mathsf{q}}, \qquad c := \mathsf{K}_i^{\max}\mathsf{H}_i^{\mathsf{W},\mathsf{L}} - \frac{1}{2}\|\mathbf{Q}_i^{\mathsf{W},\mathsf{L}}\|^2_{\ell^2}. \tag{4}$$

- Let r largest positive root of $ax^2 + bx + c = 0$ with $r = 1$ if no positive root.
- Let ℓ_{ij}^{h} water height limiter. Then set

$$\ell_j^{i,\mathsf{K}} := \min(r, \ell_{ij}^{\mathsf{h}}), \qquad \ell_{ij} = \min(\ell_j^{i,\mathsf{K}}, \ell_i^{j,\mathsf{K}}).$$

Lemma

Under the CFL condition $1 - \frac{2\Delta t}{m_i}\sum_{i\ne j\in\mathcal{I}(D_i)} d_{ij}^{\mathsf{L},n} \ge 0$, the update $\mathbf{U}_i^{n+1}$ with the above limiting satisfies the bound $\psi(\mathbf{U}_i^{n+1} - \frac{\Delta t}{m_i}\mathbf{S}_i^n) \le \mathsf{K}_i^{\max}$.

Limiting with exact bounds

- One can limit kinetic energy (for instance) $\psi(\mathbf{W}) := \frac{1}{2}\frac{1}{\mathsf{H}(\mathbf{W})}\|\mathbf{Q}(\mathbf{W})\|_{\ell^2}^2$.
- The following holds:

$$\psi(\mathbf{W}_i^{\mathsf{L},n+1}) \le \max_{j\in\mathcal{I}(D_i)} \psi(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}) =: \mathsf{K}_i^{\max}.$$

- Let $\lambda_j := \frac{1}{\mathrm{card}(\mathcal{I}(D_i))-1}$, $j \in \mathcal{I}(D_i)\setminus\{i\}$. Set

$$\mathsf{H}_i^{\mathsf{W},\mathsf{L}} := \mathsf{H}(\mathbf{W}_i^{\mathsf{L},n+1}), \qquad \mathbf{Q}_i^{\mathsf{W},\mathsf{L}} := \mathbf{Q}(\mathbf{W}_i^{\mathsf{L},n+1}), \tag{2}$$

$$\mathbf{P}_{ij} := (\mathbf{P}_{ij}^{\mathsf{h}}, \mathbf{P}_{ij}^{\mathsf{q}})^{\mathsf{T}} := \frac{1}{m_i\lambda_j}\mathbf{A}_{ij}, \qquad a := -\frac{1}{2}\|\mathbf{P}_{ij}^{\mathsf{h}}\|_{\ell^2}^2, \tag{3}$$

$$b := \mathsf{K}_i^{\max}\mathbf{P}_{ij}^{\mathsf{h}} - 2\mathbf{Q}_i^{\mathsf{W},\mathsf{L}}\cdot\mathbf{P}_{ij}^{\mathsf{q}}, \qquad c := \mathsf{K}_i^{\max}\mathsf{H}_i^{\mathsf{W},\mathsf{L}} - \frac{1}{2}\|\mathbf{Q}_i^{\mathsf{W},\mathsf{L}}\|_{\ell^2}^2. \tag{4}$$

- Let r largest positive root of $ax^2+bx+c=0$ with $r=1$ if no positive root.
- Let ℓ_{ij}^{h} water height limiter. Then set

$$\ell_j^{i,\mathsf{K}} := \min(r, \ell_{ij}^{\mathsf{h}}), \qquad \ell_{ij} = \min(\ell_j^{i,\mathsf{K}}, \ell_i^{j,\mathsf{K}}).$$

Lemma

Under the CFL condition $1 - \frac{2\Delta t}{m_i}\sum_{i\ne j\in\mathcal{I}(D_i)} d_{ij}^{\mathsf{L},n} \ge 0$, the update $\mathbf{U}_i^{n+1}$ with the above limiting satisfies the bound $\psi(\mathbf{U}_i^{n+1} - \frac{\Delta t}{m_i}\mathbf{S}_i^n) \le \mathsf{K}_i^{\max}$.

Limiting with exact bounds

- One can limit kinetic energy (for instance) $\psi(\mathbf{W}) := \frac{1}{2}\frac{1}{\mathsf{H}(\mathbf{W})}\|\mathbf{Q}(\mathbf{W})\|_{\ell^2}^2$.
- The following holds:

$$\psi(\mathbf{W}_i^{\mathrm{L},n+1}) \le \max_{j\in\mathcal{I}(D_i)} \psi(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}) =: \mathsf{K}_i^{\max}.$$

- Let $\lambda_j := \frac{1}{\mathrm{card}(\mathcal{I}(D_i))-1}$, $j \in \mathcal{I}(D_i)\setminus\{i\}$. Set

$$\mathsf{H}_i^{\mathrm{W,L}} := \mathsf{H}(\mathbf{W}_i^{\mathrm{L},n+1}), \qquad \mathbf{Q}_i^{\mathrm{W,L}} := \mathbf{Q}(\mathbf{W}_i^{\mathrm{L},n+1}), \tag{2}$$

$$\mathbf{P}_{ij} := (\mathbf{P}_{ij}^{\mathsf{h}}, \mathbf{P}_{ij}^{\mathbf{q}})^{\mathrm{T}} := \frac{1}{m_i\lambda_j}\mathbf{A}_{ij}, \qquad a := -\frac{1}{2}\|\mathbf{P}_{ij}^{\mathsf{h}}\|_{\ell^2}^2, \tag{3}$$

$$b := \mathsf{K}_i^{\max}\mathbf{P}_{ij}^{\mathsf{h}} - 2\mathbf{Q}_i^{\mathrm{W,L}}\cdot\mathbf{P}_{ij}^{\mathbf{q}}, \qquad c := \mathsf{K}_i^{\max}\mathsf{H}_i^{\mathrm{W,L}} - \frac{1}{2}\|\mathbf{Q}_i^{\mathrm{W,L}}\|_{\ell^2}^2. \tag{4}$$

- Let r largest positive root of $ax^2 + bx + c = 0$ with $r = 1$ if no positive root.
- Let ℓ_{ij}^{h} water height limiter. Then set

$$\ell_j^{i,\mathsf{K}} := \min(r, \ell_{ij}^{\mathsf{h}}), \qquad \ell_{ij} = \min(\ell_j^{i,\mathsf{K}}, \ell_i^{j,\mathsf{K}}).$$

Lemma

Under the CFL condition $1 - \frac{2\Delta t}{m_i}\sum_{i\ne j\in\mathcal{I}(D_i)} d_{ij}^{\mathrm{L},n} \ge 0$, the update $\mathbf{U}_i^{n+1}$ with the above limiting satisfies the bound $\psi(\mathbf{U}_i^{n+1} - \frac{\Delta t}{m_i}\mathbf{S}_i^n) \le \mathsf{K}_i^{\max}$.

Limiting with exact bounds

- One can limit kinetic energy (for instance) $\psi(\mathbf{W}) := \frac{1}{2}\frac{1}{\mathsf{H}(\mathbf{W})}\|\mathbf{Q}(\mathbf{W})\|_{\ell^2}^2$.
- The following holds:

$$\psi(\mathbf{W}_i^{\mathrm{L},n+1}) \le \max_{j\in\mathcal{I}(D_i)} \psi(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}) =: \mathsf{K}_i^{\max}.$$

- Let $\lambda_j := \frac{1}{\mathrm{card}(\mathcal{I}(D_i))-1}$, $j \in \mathcal{I}(D_i)\setminus\{i\}$. Set

$$\mathsf{H}_i^{\mathrm{W,L}} := \mathsf{H}(\mathbf{W}_i^{\mathrm{L},n+1}), \qquad \mathbf{Q}_i^{\mathrm{W,L}} := \mathbf{Q}(\mathbf{W}_i^{\mathrm{L},n+1}), \tag{2}$$

$$\mathbf{P}_{ij} := (\mathbf{P}_{ij}^{\mathsf{h}}, \mathbf{P}_{ij}^{\mathbf{q}})^{\mathrm{T}} := \frac{1}{m_i\lambda_j}\mathbf{A}_{ij}, \qquad a := -\frac{1}{2}\|\mathbf{P}_{ij}^{\mathsf{h}}\|_{\ell^2}^2, \tag{3}$$

$$b := \mathsf{K}_i^{\max}\mathbf{P}_{ij}^{\mathsf{h}} - 2\mathbf{Q}_i^{\mathrm{W,L}}\cdot\mathbf{P}_{ij}^{\mathbf{q}}, \qquad c := \mathsf{K}_i^{\max}\mathsf{H}_i^{\mathrm{W,L}} - \frac{1}{2}\|\mathbf{Q}_i^{\mathrm{W,L}}\|_{\ell^2}^2. \tag{4}$$

- Let r largest positive root of $ax^2 + bx + c = 0$ with $r = 1$ if no positive root.
- Let ℓ_{ij}^{h} water height limiter. Then set

$$\ell_j^{i,\mathsf{K}} := \min(r, \ell_{ij}^{\mathsf{h}}), \qquad \ell_{ij} = \min(\ell_j^{i,\mathsf{K}}, \ell_i^{j,\mathsf{K}}).$$

Lemma

Under the CFL condition $1 - \frac{2\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} d_{ij}^{\mathrm{L},n} \ge 0$, *the update* $\mathbf{U}_i^{n+1}$ *with the above limiting satisfies the bound* $\psi(\mathbf{U}_i^{n+1} - \frac{\Delta t}{m_i}\mathbf{S}_i^n) \le \mathsf{K}_i^{\max}$.

Limiting with exact bounds

- One can limit kinetic energy (for instance) $\psi(\mathbf{W}) := \frac{1}{2}\frac{1}{\mathsf{H}(\mathbf{W})}\|\mathbf{Q}(\mathbf{W})\|_{\ell^2}^2$.
- The following holds:

$$\psi(\mathbf{W}_i^{\mathsf{L},n+1}) \leq \max_{j\in\mathcal{I}(D_i)} \psi(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}) =: \mathsf{K}_i^{\max}.$$

- Let $\lambda_j := \frac{1}{\mathrm{card}(\mathcal{I}(D_i))-1}$, $j \in \mathcal{I}(D_i)\setminus\{i\}$. Set

$$\mathsf{H}_i^{\mathsf{W,L}} := \mathsf{H}(\mathbf{W}_i^{\mathsf{L},n+1}), \qquad \mathbf{Q}_i^{\mathsf{W,L}} := \mathbf{Q}(\mathbf{W}_i^{\mathsf{L},n+1}), \tag{2}$$

$$\mathbf{P}_{ij} := (\mathbf{P}_{ij}^{\mathsf{h}}, \mathbf{P}_{ij}^{\mathbf{q}})^{\mathrm{T}} := \frac{1}{m_i\lambda_j}\mathbf{A}_{ij}, \qquad a := -\frac{1}{2}\|\mathbf{P}_{ij}^{\mathsf{h}}\|_{\ell^2}^2, \tag{3}$$

$$b := \mathsf{K}_i^{\max}\mathbf{P}_{ij}^{\mathsf{h}} - 2\mathbf{Q}_i^{\mathsf{W,L}}\cdot\mathbf{P}_{ij}^{\mathbf{q}}, \qquad c := \mathsf{K}_i^{\max}\mathsf{H}_i^{\mathsf{W,L}} - \frac{1}{2}\|\mathbf{Q}_i^{\mathsf{W,L}}\|_{\ell^2}^2. \tag{4}$$

- Let r largest positive root of $ax^2 + bx + c = 0$ with $r = 1$ if no positive root.
- Let ℓ_{ij}^{h} water height limiter. Then set

$$\ell_j^{i,\mathsf{K}} := \min(r, \ell_{ij}^{\mathsf{h}}), \qquad \ell_{ij} = \min(\ell_j^{i,\mathsf{K}}, \ell_i^{j,\mathsf{K}}).$$

Lemma

Under the CFL condition $1 - \frac{2\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} d_{ij}^{\mathsf{L},n} \geq 0$, the update $\mathbf{U}_i^{n+1}$ with the above limiting satisfies the bound $\psi(\mathbf{U}_i^{n+1} - \frac{\Delta t}{m_i}\mathbf{S}_i^n) \leq \mathsf{K}_i^{\max}$.

Limiting with exact bounds

- One can limit kinetic energy (for instance) $\psi(\mathbf{W}) := \frac{1}{2}\frac{1}{\mathsf{H}(\mathbf{W})}\|\mathbf{Q}(\mathbf{W})\|_{\ell^2}^2$.
- The following holds:

$$\psi(\mathbf{W}_i^{\mathrm{L},n+1}) \leq \max_{j\in\mathcal{I}(D_i)} \psi(\overline{\mathbf{U}_{ij}^n} + \widetilde{\mathbf{U}_{ij}^n}) =: \mathsf{K}_i^{\max}.$$

- Let $\lambda_j := \frac{1}{\mathrm{card}(\mathcal{I}(D_i))-1}$, $j \in \mathcal{I}(D_i)\setminus\{i\}$. Set

$$\mathsf{H}_i^{\mathrm{W,L}} := \mathsf{H}(\mathbf{W}_i^{\mathrm{L},n+1}), \qquad \mathbf{Q}_i^{\mathrm{W,L}} := \mathbf{Q}(\mathbf{W}_i^{\mathrm{L},n+1}), \tag{2}$$

$$\mathbf{P}_{ij} := (\mathbf{P}_{ij}^{\mathrm{h}}, \mathbf{P}_{ij}^{\mathbf{q}})^{\mathrm{T}} := \frac{1}{m_i\lambda_j}\mathbf{A}_{ij}, \qquad a := -\frac{1}{2}\|\mathbf{P}_{ij}^{\mathrm{h}}\|_{\ell^2}^2, \tag{3}$$

$$b := \mathsf{K}_i^{\max}\mathbf{P}_{ij}^{\mathrm{h}} - 2\mathbf{Q}_i^{\mathrm{W,L}}\cdot\mathbf{P}_{ij}^{\mathbf{q}}, \qquad c := \mathsf{K}_i^{\max}\mathsf{H}_i^{\mathrm{W,L}} - \frac{1}{2}\|\mathbf{Q}_i^{\mathrm{W,L}}\|_{\ell^2}^2. \tag{4}$$

- Let r largest positive root of $ax^2+bx+c=0$ with $r=1$ if no positive root.
- Let ℓ_{ij}^{h} water height limiter. Then set

$$\ell_j^{i,\mathrm{K}} := \min(r, \ell_{ij}^{\mathrm{h}}), \qquad \ell_{ij} = \min(\ell_j^{i,\mathrm{K}}, \ell_i^{j,\mathrm{K}}).$$

Lemma

Under the CFL condition $1 - \frac{2\Delta t}{m_i}\sum_{i\neq j\in\mathcal{I}(D_i)} d_{ij}^{\mathrm{L},n} \geq 0$, *the update* $\mathbf{U}_i^{n+1}$ *with the above limiting satisfies the bound* $\psi(\mathbf{U}_i^{n+1} - \frac{\Delta t}{m_i}\mathbf{S}_i^n) \leq \mathsf{K}_i^{\max}$.

Well-balancing, sliding state

h_0 (m)	q_0 (m^2s^{-1})	n ($m^{-1/3}s$)	b	Error
5.7708E-01	2.0E-00	2.0E-02	-1.E-02	4.26E-14
9.5635E-02	1.0E-01	2.0E-02	-1.E-02	1.82E-15
2.5119E-01	1.0E-01	1.0E-01	-1.E-02	9.04E-15
2.4022E-02	2.0E-03	1.0E-01	-1.E-02	1.49E-14
4.4894E-01	2.0E-00	1.0E-01	$-1/\sqrt{3}$	1.86E-14

Table: Well-balancing tests, TAMU code (EV-α^2, consistent), $\gamma = \frac{4}{3}$.

Limiting

Hydraulic jump.

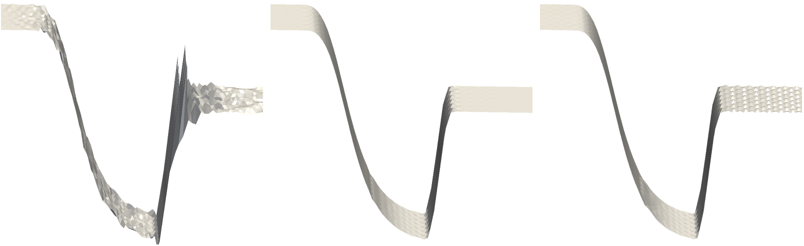


Figure: Galerkin + full limiting; EV + full limiting; EV without any limiting.

2D paraboloid with friction

Proteus				
I	L^1-error	Rate	L^∞-error	Rate
441	4.58E-02		8.41E-02	
1681	1.45E-02	1.72	4.07E-02	1.08
6561	6.30E-03	1.22	2.64E-02	0.63
25921	2.24E-03	1.50	1.34E-02	0.99
103041	7.52E-04	1.58	6.46E-03	1.06
TAMU				
I	L^1-error	Rate	L^∞-error	Rate
508	4.70E-02		8.12E-02	
1926	1.95E-02	1.32	4.55E-02	0.87
7553	7.67E-03	1.37	1.98E-02	1.22
29870	2.91E-03	1.41	1.11E-02	0.84
118851	1.09E-03	1.43	6.21E-03	0.85

Table: L^1 convergence; 2D paraboloid with friction.

Dam break over three bumps

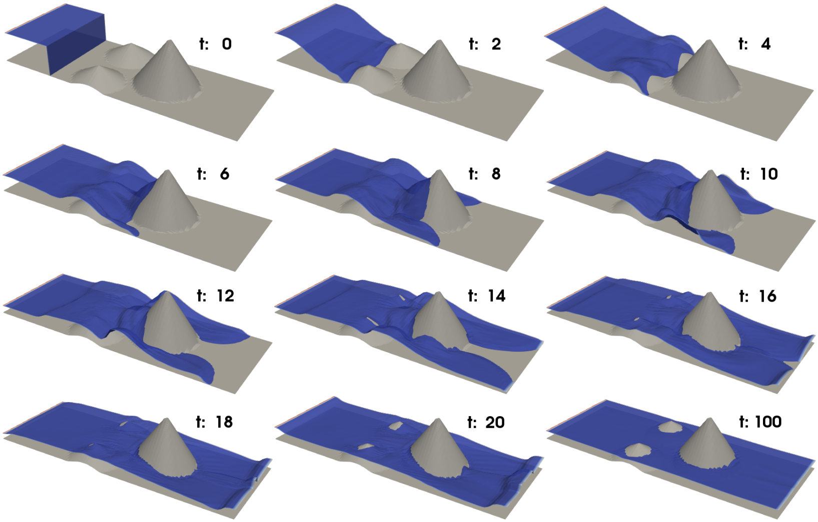
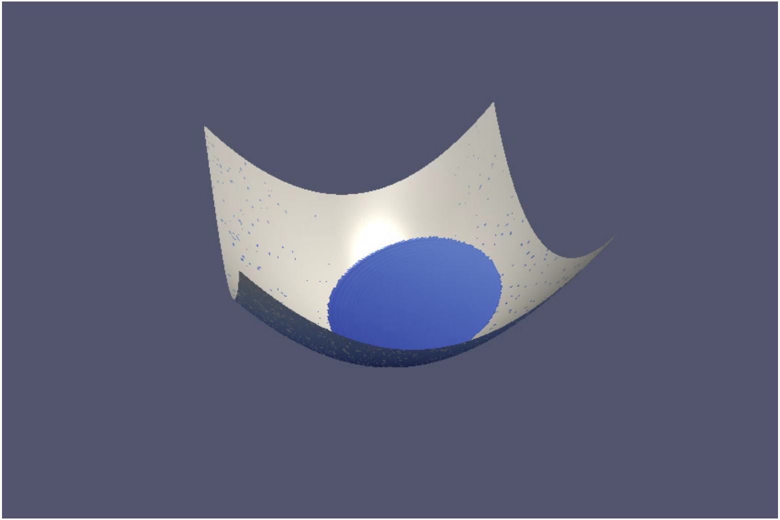


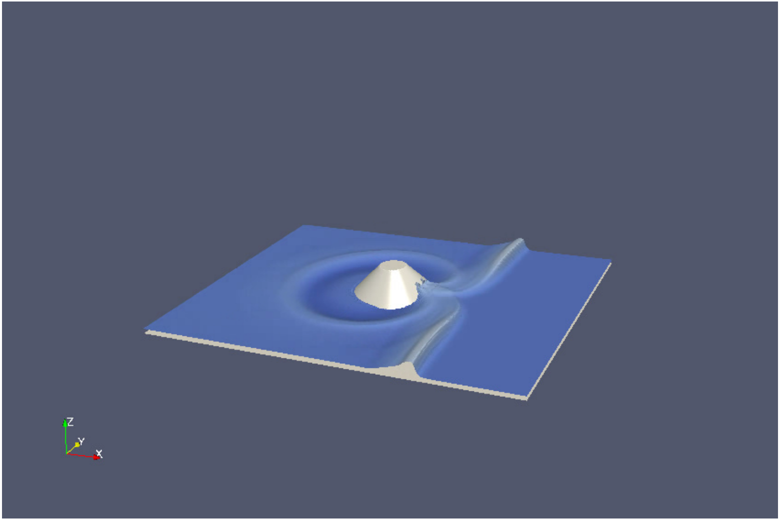
Figure: Surface plot of the water elevation $h(\mathbf{x}, t) + z(\mathbf{x})$ at different times.

Shallow water



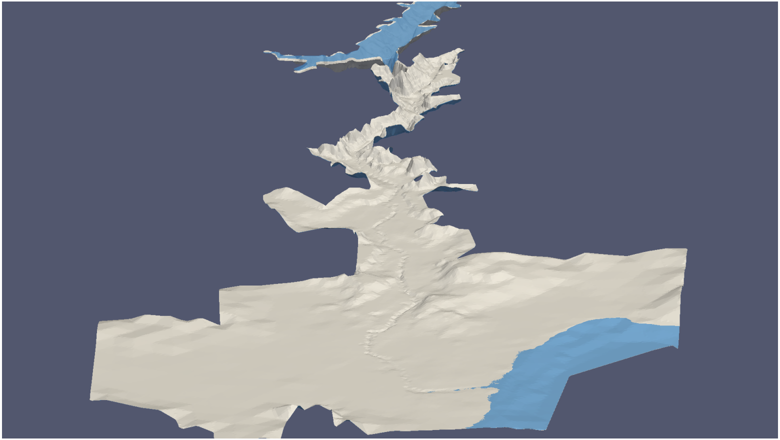
Flat free surface rotating in a paraboloid, no friction (wine/water in a glass).

Overtopping of island



(87767 $\mathbb{P}_1$ nodes)

Malpasset



(29381 $\mathbb{P}_1$ nodes)

The Malpasset Dam was an arch dam on the Reyran River, located approximately 7 km north of Fréjus on Côte d'Azur (French Riviera), southern France. The dam collapsed on December 2, 1959, killing 423 people in the resulting flood.

Malpasset

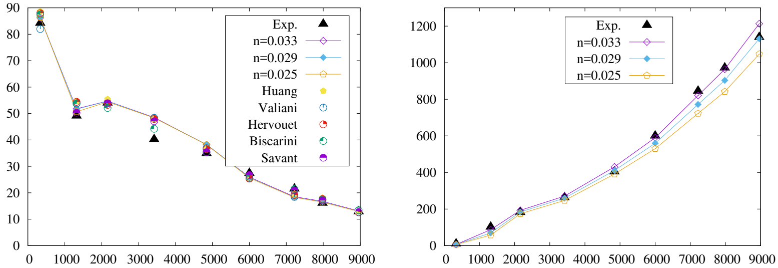


Figure: Maximum water elevation (left). Arrival time (right).

Shallow water

- R. Pasquetti, J. L. Guermond, and B. Popov. *Stabilized Spectral Element Approximation of the Saint Venant System Using the Entropy Viscosity Technique*, pages 397–404.
  Springer International Publishing, Cham, 2015
- Pascal Azerad, J.-L. Guermond, and Bojan Popov. Well-balanced second-order approximation of the shallow water equations with continuous finite elements.
  SIAM J. Numer. Anal.
  In press
- Guermond J.-L., M. Quezada de Luna, C. Kees, B. Popov, and M. Farthing. Well-balanced second-order fe approximation of the shallow water equations with friction.
  Submitted SIAM SISC

Conclusions

Continuous finite elements

- Continuous FE are viable tools to solve hyperbolic systems.
- Continuous FE are viable alternatives to DG and FV.
- Continuous FE are easy to implement and parallelize.
- Exa-scale computing will need simple, robust, methods.

Current and future work

- Convergence analysis, error estimates beyond first-order.
- Extension to higher-order polynomials for scalar equations (order 3 and higher).
- Beyond positivity: Second-order invariant domain preserving techniques to systems (Shallow water, Euler).
- Extension to equations with source terms (Radiative transport, Radiative hydrodynamics).

Conclusions

Continuous finite elements

- Continuous FE are viable tools to solve hyperbolic systems.
- Continuous FE are viable alternatives to DG and FV.
- Continuous FE are easy to implement and parallelize.
- Exa-scale computing will need simple, robust, methods.

Current and future work

- Convergence analysis, error estimates beyond first-order.
- Extension to higher-order polynomials for scalar equations (order 3 and higher).
- Beyond positivity: Second-order invariant domain preserving techniques to systems (Shallow water, Euler).
- Extension to equations with source terms (Radiative transport, Radiative hydrodynamics).

Conclusions

Continuous finite elements

- Continuous FE are viable tools to solve hyperbolic systems.
- Continuous FE are viable alternatives to DG and FV.
- Continuous FE are easy to implement and parallelize.
- Exa-scale computing will need simple, robust, methods.

Current and future work

- Convergence analysis, error estimates beyond first-order.
- Extension to higher-order polynomials for scalar equations (order 3 and higher).
- Beyond positivity: Second-order invariant domain preserving techniques to systems (Shallow water, Euler).
- Extension to equations with source terms (Radiative transport, Radiative hydrodynamics).

Conclusions

Continuous finite elements

- Continuous FE are viable tools to solve hyperbolic systems.
- Continuous FE are viable alternatives to DG and FV.
- Continuous FE are easy to implement and parallelize.
- Exa-scale computing will need simple, robust, methods.

Current and future work

- Convergence analysis, error estimates beyond first-order.
- Extension to higher-order polynomials for scalar equations (order 3 and higher).
- Beyond positivity: Second-order invariant domain preserving techniques to systems (Shallow water, Euler).
- Extension to equations with source terms (Radiative transport, Radiative hydrodynamics).

Conclusions

Continuous finite elements

- Continuous FE are viable tools to solve hyperbolic systems.
- Continuous FE are viable alternatives to DG and FV.
- Continuous FE are easy to implement and parallelize.
- Exa-scale computing will need simple, robust, methods.

Current and future work

- Convergence analysis, error estimates beyond first-order.
- Extension to higher-order polynomials for scalar equations (order 3 and higher).
- Beyond positivity: Second-order invariant domain preserving techniques to systems (Shallow water, Euler).
- Extension to equations with source terms (Radiative transport, Radiative hydrodynamics).

Conclusions

Continuous finite elements

- Continuous FE are viable tools to solve hyperbolic systems.
- Continuous FE are viable alternatives to DG and FV.
- Continuous FE are easy to implement and parallelize.
- Exa-scale computing will need simple, robust, methods.

Current and future work

- Convergence analysis, error estimates beyond first-order.
- Extension to higher-order polynomials for scalar equations (order 3 and higher).
- Beyond positivity: Second-order invariant domain preserving techniques to systems (Shallow water, Euler).
- Extension to equations with source terms (Radiative transport, Radiative hydrodynamics).

Conclusions

Continuous finite elements

- Continuous FE are viable tools to solve hyperbolic systems.
- Continuous FE are viable alternatives to DG and FV.
- Continuous FE are easy to implement and parallelize.
- Exa-scale computing will need simple, robust, methods.

Current and future work

- Convergence analysis, error estimates beyond first-order.
- Extension to higher-order polynomials for scalar equations (order 3 and higher).
- Beyond positivity: Second-order invariant domain preserving techniques to systems (Shallow water, Euler).
- Extension to equations with source terms (Radiative transport, Radiative hydrodynamics).

Conclusions

Continuous finite elements

- Continuous FE are viable tools to solve hyperbolic systems.
- Continuous FE are viable alternatives to DG and FV.
- Continuous FE are easy to implement and parallelize.
- Exa-scale computing will need simple, robust, methods.

Current and future work

- Convergence analysis, error estimates beyond first-order.
- Extension to higher-order polynomials for scalar equations (order 3 and higher).
- Beyond positivity: Second-order invariant domain preserving techniques to systems (Shallow water, Euler).
- Extension to equations with source terms (Radiative transport, Radiative hydrodynamics).

References I

[1] Pascal Azerad, J.-L. Guermond, and Bojan Popov. Well-balanced second-order approximation of the shallow water equations with continuous finite elements. *SIAM J. Numer. Anal.* In press.

[2] J.-L. Guermond and Bojan Popov. Invariant domains and second-order continuous finite element approximation for scalar conservation equations. *SIAM J. Numer. Anal.* In press.

[3] J.-L. Guermond, Bojan Popov, , M. Nazarov, and Ignacio Tomas. Second-order invariant domain preserving approximation of the euler equations using convex limiting. Submitted SIAM SISC.

[4] Jean-Luc Guermond and Bojan Popov. Error Estimates of a First-order Lagrange Finite Element Technique for Nonlinear Scalar Conservation Equations. *SIAM J. Numer. Anal.*, 54(1):57–85, 2016.

[5] Jean-Luc Guermond and Bojan Popov. Fast estimation from above of the maximum wave speed in the Riemann problem for the Euler equations. *J. Comput. Phys.*, 321:908–926, 2016.

[6] Jean-Luc Guermond and Bojan Popov. Invariant domains and first-order continuous finite element approximation for hyperbolic systems. *SIAM J. Numer. Anal.*, 54(4):2466–2489, 2016.

[7] Guermond J.-L., M. Quezada de Luna, C. Kees, B. Popov, and M. Farthing. Well-balanced second-order fe approximation of the shallow water equations with friction. Submitted SIAM SISC.

References II

[8] R. Pasquetti, J. L. Guermond, and B. Popov. *Stabilized Spectral Element Approximation of the Saint Venant System Using the Entropy Viscosity Technique*, pages 397–404. Springer International Publishing, Cham, 2015.