# Contribution à l'approximation numérique des systèmes hyperboliques


**Vivien Desveaux**

sous la direction de Christophe Berthon et Yves Coudière

Laboratoire de mathématiques Jean Leray - Université de Nantes


26 novembre 2013

1. Second-order schemes based on dual mesh gradient reconstruction
2. A high-order entropy preserving scheme with *a posteriori* limitation
3. Well-balanced schemes for systems with nonlinear source terms

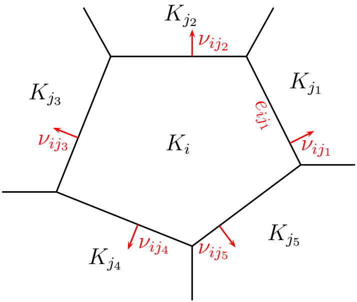
# Introduction

- Hyperbolic system of conservation laws in 2D

$$\partial_t w + \partial_x f(w) + \partial_y g(w) = 0$$

  $w : \mathbb{R}^+ \times \mathbb{R}^2 \to \Omega \subset \mathbb{R}^d$: unknown state vector
  $f, g : \Omega \to \mathbb{R}^d$: flux functions
- $\Omega$ convex set of physical states
- Objective: derive a numerical scheme
  - second order accurate
  - $\Omega$–preserving
  - works on unstructured meshes
  - with an optimized CFL condition

# Mesh notations




Geometry of the cell $K_i$

- polygonal cells $K_i$ (perimeter $\mathcal{P}_i$, area $|K_i|$)
- $\gamma(i)$: index set of the cells neighbouring $K_i$
- $e_{ij}$: common edge between $K_i$ and $K_j$ (length $|e_{ij}|$)
- $\nu_{ij}$: unit outward normal to $e_{ij}$

# First-order scheme

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j \in \gamma(i)} |e_{ij}| \varphi\left(w_i^n, w_j^n, \nu_{ij}\right)$$

- 2D numerical flux: $\varphi$ Godunov-type in each direction $\nu$ [Harten, Lax & van Leer '83]:

$$\varphi(w_L, w_R, \nu) = h_\nu(w_L) + \frac{\delta}{2\Delta t} w_L - \frac{1}{\Delta t} \int_{-\frac{\delta}{2}}^{0} \widetilde{w}_\nu\left(\frac{x}{\Delta t}, w_L, w_R\right) dx$$

  with respect to the CFL condition $\dfrac{\Delta t}{\delta} \max \left|\lambda^\pm(w_L, w_R, \nu)\right| \leq \dfrac{1}{2}$

  - $h_\nu(w) = \nu_x f(w) + \nu_y g(w)$: flux in the $\nu$–direction, with $\nu = (\nu_x, \nu_y)^T$
  - $\widetilde{w}_\nu$ approximate Riemann solver valued in $\Omega$

- Consistency: $\varphi(w, w, \nu) = h_\nu(w)$
- Conservation: $\varphi(w_L, w_R, \nu) = -\varphi(w_R, w_L, -\nu)$

### Theorem (Robustness of the first-order scheme)

*Assume the following CFL condition is satisfied:*

$$\Delta t \frac{\mathcal{P}_i}{|K_i|} \max_{j \in \gamma(i)} \left| \lambda^{\pm}(w_i^n, w_j^n, \nu_{ij}) \right| \leq 1, \quad \forall i \in \mathbb{Z}.$$

*Then the first-order scheme preserves $\Omega$:*

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n+1} \in \Omega.$$

Remark: in [Perthame & Shu '96], they obtain the CFL restriction

$$\Delta t \frac{\mathcal{P}_i}{|K_i|} \max_{j \in \gamma(i)} \left| \lambda^{\pm}(w_i^n, w_j^n \nu_{ij}) \right| \leq \frac{1}{2}, \quad \forall i \in \mathbb{Z}.$$

However this can easily be improved in the Godunov-type framework.

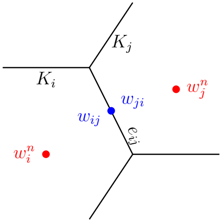
# MUSCL scheme ([van Leer ’79], [Perthame & Shu ’96]...)

First-order scheme on the cell $K_i$

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j \in \gamma(i)} |e_{ij}| \varphi\left(w_i^n, w_j^n, \nu_{ij}\right)$$

Second-order MUSCL scheme on the cell $K_i$

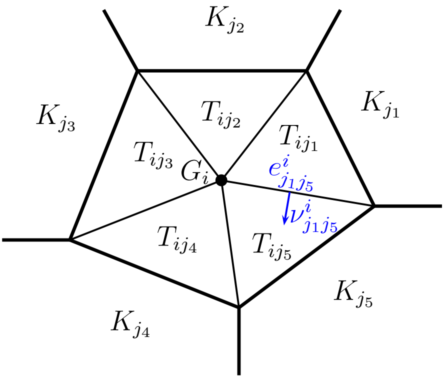
$$w_i^{n+1} = w_i^n - \frac{\Delta t}{|K_i|} \sum_{j \in \gamma(i)} |e_{ij}| \varphi\left(w_{ij}, w_{ji}, \nu_{ij}\right)$$




$w_{ij}$ and $w_{ji}$ are second-order approximations at the interface between $K_i$ and $K_j$

$\rightarrow$ How to compute $w_{ij}$ ?

# Subcells decomposition




Subcells decomposition of the cell $K_i$

- $T_{ij}$: triangle formed by the mass center $G_i$ and the edge $e_{ij}$ (perimeter $\mathcal{P}_{ij}$, area $|T_{ij}|$)
- $\gamma(i,j)$: index set of the two subcells neighbouring $T_{ij}$ in $K_i$
- $e^i_{jk}$: common edge between $T_{ij}$ and $T_{ik}$ (length $|e^i_{jk}|$)
- $\nu^i_{jk}$: unit outward normal to $e^i_{jk}$

## Theorem (Robustness of the MUSCL scheme)

*Assume the following hypotheses:*

(i) *The reconstruction preserves* $\Omega$*:*

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad \forall j \in \gamma(i), \quad w_{ij} \in \Omega.$$

(ii) *The reconstruction satisfies the conservation property*

$$\sum_{j \in \gamma(i)} \frac{|T_{ij}|}{|K_i|} w_{ij} = w_i^n.$$

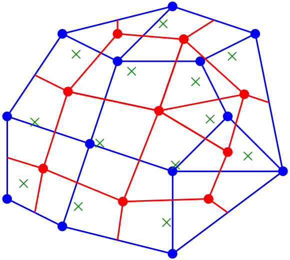
(iii) *The following CFL condition is satisfied for all* $i \in \mathbb{Z}$*:*

$$\Delta t \max_{j \in \gamma(i)} \frac{\mathcal{P}_{ij}}{|T_{ij}|} \max_{k \in \gamma(i,j)} \left| \lambda^{\pm}(w_{ij}, w_{ji}, \nu_{ij}), \lambda^{\pm}(w_{ij}, w_{ik}, \nu_{jk}^{i}) \right| \leq 1$$

*Then the MUSCL scheme preserves* $\Omega$*:*

$$\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n+1} \in \Omega.$$

# The DMGR scheme

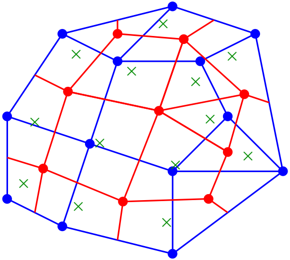


- Vertex of the primal mesh (unknown state)
- Center of a primal cell = Vertex of the dual mesh (known state)
- × Center of a dual cell (known state)

Figure: Primal mesh (blue) and dual mesh (red)

1. We write a MUSCL scheme on both primal and dual meshes
   $\Rightarrow$ At time $t^n$, we know a state at the center of each primal and dual cell

# The DMGR scheme



- Vertex of the primal mesh (unknown state)
- Center of a primal cell = Vertex of the dual mesh (known state)
- × Center of a dual cell (known state)

Figure: Primal mesh (blue) and dual mesh (red)

Assume we have a reconstruction procedure on a generic cell $K$:

$$\begin{cases} \text{state at the center} \\ \text{states at the vertices} \end{cases} \quad \mapsto \quad \text{linear reconstruction } \widetilde{w}$$

This procedure will be detailed after.

# The DMGR scheme

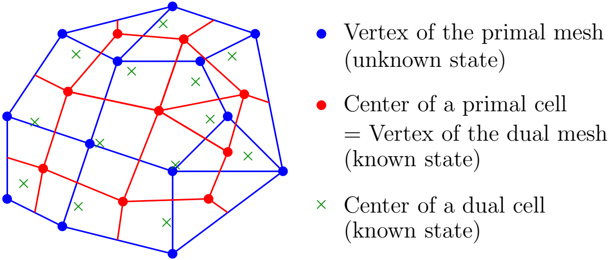
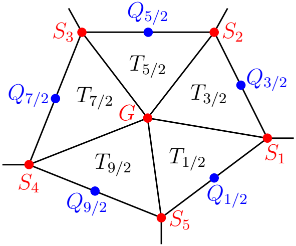


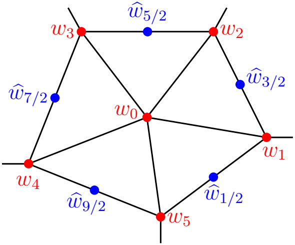

Figure: Primal mesh (blue) and dual mesh (red)

2. We can apply the reconstruction procedure on the dual cells
   $\Rightarrow$ We get a linear function $\widetilde{w}_i^d$ on each dual cell
3. To get the state at the primal vertex $S_i^p$, we take $\widetilde{w}_i^d(S_i^p)$
   $\Rightarrow$ We can apply the reconstruction procedure on the primal cells to get a linear function $\widetilde{w}_i^p$

# Reconstruction procedure




Geometry of the cell $K$




Known states and reconstructed states

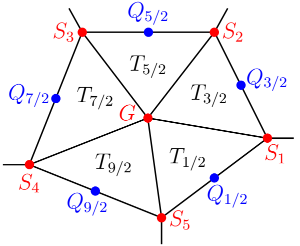
The states $\widehat{w}_{j-1/2}$ have to satisfy:

- $\widehat{w}_{j-1/2} \in \Omega$
- $\displaystyle\sum_j \frac{|T_{j-1/2}|}{|K|} \widehat{w}_{j-1/2} = w_0$

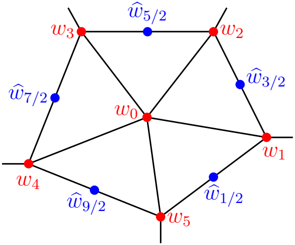
If we take $\widehat{w}_{j-1/2} = \widetilde{w}(Q_{j-1/2})$ with $\widetilde{w}$ a linear function on $K$, we have

$$\sum_j \frac{|T_{j-1/2}|}{|K|} \widehat{w}_{j-1/2} = w_0 \quad \Leftrightarrow \quad \widetilde{w}(G) = w_0$$

# Reconstruction procedure



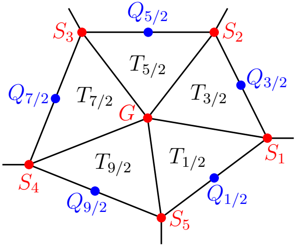

Geometry of the cell $K$



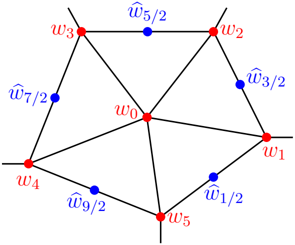

Known states and reconstructed states

1. Gradient reconstruction
   We define a continuous function $\overline{w} : K \to \mathbb{R}^d$ piecewise linear on each triangle $T_{j-1/2}$ and such that $\overline{w}(S_j) = w_j$ and $\overline{w}(G) = w_0$.

# Reconstruction procedure




Geometry of the cell $K$




Known states and reconstructed states

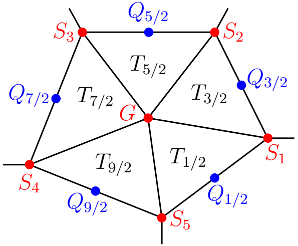
2. Projection

   For a slope $\alpha \in M_{d,2}(\mathbb{R})$, we define $\widetilde{w}_\alpha(X) = w_0 + \alpha \cdot (X - G)$, the linear function whose gradient is $\alpha$.

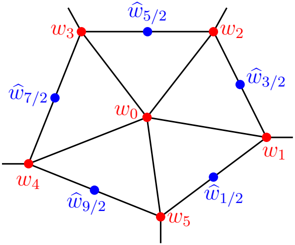
   Let $\mu$ be the slope resulting from the $L^2$–projection of $\overline{w}$:

$$\int_K \left\|\overline{w}(X) - \widetilde{w}_\mu(X)\right\|^2 dX = \min_{\alpha \in M_{d,2}(\mathbb{R})} \int_K \left\|\overline{w}(X) - \widetilde{w}_\alpha(X)\right\|^2 dX.$$

# Reconstruction procedure




Geometry of the cell $K$




Known states and reconstructed states

3. Limitation of the slope $\mu$

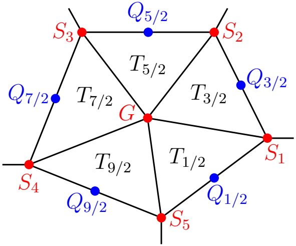
We define the optimal slope limiter by:

$$\alpha_{j-1/2} = \sup \left\{ \theta \in [0,1], \widetilde{w}_{s\mu}(Q_{j-1/2}) \in \Omega, \forall s \in [0, \theta] \right\},$$

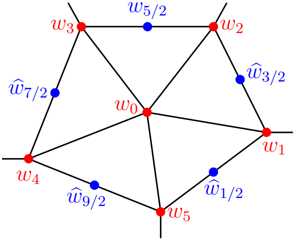
$$\beta = \min_j \alpha_{j-1/2} - \epsilon,$$

where $\epsilon > 0$ is a small paramater s.t. $\widetilde{w}_{\beta\mu}(Q_{j-1/2}) \in \Omega, \quad \forall j.$

# Reconstruction procedure




Geometry of the cell $K$




Known states and reconstructed states

4. Finally, the reconstructed states are given by $\widehat{w}_{j-1/2} = \widetilde{w}_{\beta\mu}(Q_{j-1/2})$.

$$\begin{array}{lcl} \text{Limitation procedure} & \Rightarrow & \widehat{w}_{j-1/2} \in \Omega \\ \widetilde{w}(G) = w_0 & \Rightarrow & \displaystyle\sum_j \frac{|T_{j-1/2}|}{|K|} \widehat{w}_{j-1/2} = w_0 \end{array}$$

$\Rightarrow$ **The DMGR scheme preserves $\Omega$**

# Numerical results: 2D Euler equations

$$\partial_t \begin{pmatrix} \rho \\ \rho u \\ \rho v \\ E \end{pmatrix} + \partial_x \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho u v \\ u(E+p) \end{pmatrix} + \partial_y \begin{pmatrix} \rho v \\ \rho u v \\ \rho v^2 + p \\ v(E+p) \end{pmatrix} = 0$$

- $\rho$: density
- $(u, v)$: velocity
- $E$: total energy
- $p$: pressure given by the ideal gas law

$$p = (\gamma - 1)\left(E - \frac{\rho}{2}\left(u^2 + v^2\right)\right), \quad \text{with } \gamma \in (1, 3]$$

- Set of physical states

$$\Omega = \left\{(\rho, \rho u, \rho v, E) \in \mathbb{R}^4; \quad \rho > 0, \quad E - \frac{\rho}{2}\left(u^2 + v^2\right) > 0\right\}$$

# 2D Riemann problems

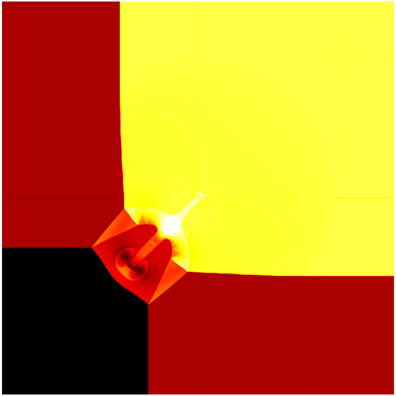


Figure: Four shocks 2D Riemann problem on a Cartesian mesh with $1.5 \times 10^6$ DOF

Figure: Four contact discontinuities 2D Riemann problem on a Cartesian mesh with $1.5 \times 10^6$ DOF

# Double Mach reflection on a ramp

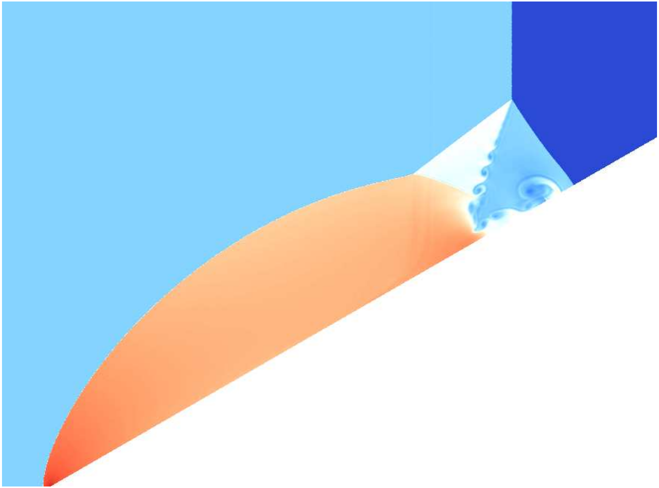


Figure: Double Mach reflection on a ramp on an unstructured mesh with $3 \times 10^6$ DOF

# Mach 3 wind tunnel with a step

Figure: Mach 3 tunnel with a step on an unstructured mesh with $1.5 \times 10^6$ DOF

1. Second-order schemes based on dual mesh gradient reconstruction
   - MUSCL scheme and CFL condition
   - The DMGR scheme
   - Numerical results

2. A high-order entropy preserving scheme with *a posteriori* limitation
   - Motivations
   - From one to all discrete entropy inequalities
   - The e-MOOD scheme for the Euler equations

3. Well-balanced schemes for systems with nonlinear source terms
   - The Ripa model
   - Relaxation models
   - The relaxation scheme
   - Numerical results

# Introduction

- Hyperbolic system of conservation laws in 1D

$$\partial_t w + \partial_x f(w) = 0$$

$w : \mathbb{R}^+ \times \mathbb{R} \to \Omega \subset \mathbb{R}^d$: unknown state vector
$f : \Omega \to \mathbb{R}^d$: flux function

- $\Omega$ convex set of physical states
- Entropy inequalities:

$$\partial_t \eta(w) + \partial_x \mathcal{G}(w) \leq 0,$$

where $w \mapsto \eta(w)$ is convex and $\nabla_w f \nabla_w \eta = \nabla_w \mathcal{G}$

- Objectives:
  - Study the entropy stability of high-order schemes
  - Derive a high-order numerical scheme for the Euler equations which is entropy preserving in the sense of Lax-Wendroff

# Scheme notations

- Space discretization: cells $K_i = [x_{i-1/2}, x_{i+1/2}]$ with constant size $\Delta x = x_{i+1/2} - x_{i-1/2}$
- $w_i^n$: approximate solution at time $t^n$ on the cell $K_i$
- Update at time $t^{n+1} = t^n + \Delta t$ given by

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x} \left( F_{i+1/2}^n - F_{i-1/2}^n \right)$$

  where $F_{i+1/2}^n = F\left( w_{i-s+1}^n, \cdots, w_{i+s}^n \right)$ and $F$ is a consistant numerical flux ($F(w, \cdots, w) = f(w)$)
- We introduce the piecewise constant function

$$w^\Delta(x,t) = w_i^n, \quad \text{for } (x,t) \in K_i \times [t^n, t^{n+1})$$

- The sequence $(\Delta x, \Delta t)$ is devoted to converge to $(0,0)$, the ratio $\frac{\Delta t}{\Delta x}$ being kept constant.

# Lax-Wendroff Theorem

## Theorem

(i) *Assume the following hypotheses:*

- *There exists a compact* $K \subset \Omega$ *such that* $w^\Delta \in K$*;*
- $w^\Delta$ *converges in* $L^1_{loc}(\mathbb{R} \times \mathbb{R}^+; \Omega)$ *to a function* $w$*.*

*Then* $w$ *is a weak solution.*

(ii) *Assume the additional hypothesis:*

- *For all entropy pair* $(\eta, \mathcal{G})$*, there exists an entropy numerical flux* $G$*, consistant with* $\mathcal{G}$ $(G(w, \cdots, w) = \mathcal{G}(w))$*, such that we have the discrete entropy inequality*

$$\frac{\eta(w_i^{n+1}) - \eta(w_i^n)}{\Delta t} + \frac{G^n_{i+1/2} - G^n_{i-1/2}}{\Delta x} \leq 0,$$

*with* $G^n_{i+1/2} = G\left(w^n_{i-s+1}, \cdots, w^n_{i+s}\right)$*.*

*Then* $w$ *is an entropic solution.*

# Example: the MUSCL scheme

- Limited slope $\mu_i^n = L\left(w_i^n - w_{i-1}^n, w_{i+1}^n - w_i^n\right)$, with $L$ a limiter
- The MUSCL flux is defined by

$$F_{i+1/2}^n = F\left(w_i^n + \mu_i^n/2, w_{i+1}^n - \mu_{i+1}^n/2\right)$$

- The known discrete entropy inequalities satisfied by the MUSCL scheme all write

$$\frac{\eta(w_i^{n+1}) - \eta(w_i^n)}{\Delta t} + \frac{G_{i+1/2}^n - G_{i-1/2}^n}{\Delta x} \leq \frac{P_i^n - \eta(w_i^n)}{\Delta t}$$

where $P_i^n = P\left(w_i^n, \mu_i^n, \Delta x, \eta\right)$.

- Examples of operator $P$:

$$P_1(w,\mu,\Delta x,\eta) = \frac{1}{\Delta x}\int_{-\Delta x/2}^{\Delta x/2} \eta\left(w + \frac{x}{\Delta x}\mu\right) dx \quad \text{[Bouchut \textit{et al.} '96]}$$

$$P_2(w,\mu,\Delta x,\eta) = \frac{\eta(w-\mu/2) + \eta(w+\mu/2)}{2} \quad \text{[Berthon '05]}$$

# Convergence study

- The discrete entropy inequality

$$\frac{\eta(w_i^{n+1}) - \eta(w_i^n)}{\Delta t} + \frac{G_{i+1/2}^n - G_{i-1/2}^n,}{\Delta x} \leq \frac{P_i^n - \eta(w_i^n)}{\Delta t}$$

converges weakly to

$$\partial_t \eta(w) + \partial_x \mathcal{G}(w) \leq \delta,$$

where $\delta$ is a positive measure.

## Conjecture (Hou-LeFloch '94)

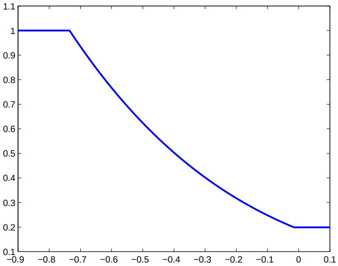
- $\delta = 0$ in the areas where $w$ is smooth
- $\delta > 0$ on the curves of discontinuity of $w$

# Numerical study: test cases (Euler equations)

Entropy error (total mass of the right-hand side):

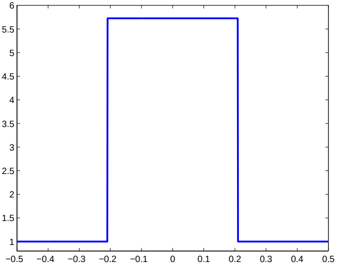
$$I^{\Delta} = \Delta x \sum_{i,n} \left(P_i^n - \eta(w_i^n)\right)$$

1–rarefaction




$I^{\Delta} \overset{?}{\to} 0$

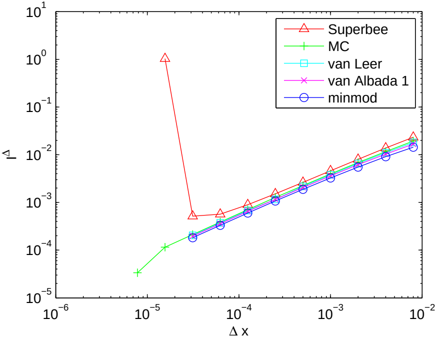
Double shock



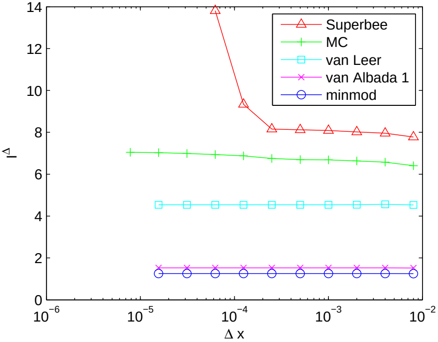

$I^{\Delta} \overset{?}{\to} c > 0$

# Numerical results obtained with a second-order time scheme

## 1-rarefaction




## Double shock



# Conclusion of motivations

- Numerical results confirm the Hou-le Floch conjecture: when the scheme converges, the measure $\delta$ seems to be concentrated on the curves of discontinuity of $w$.
- This does not imply that the limit is not entropic, but the usual discrete entropy inequalities are not relevant to apply the Lax-Wendroff theorem.
- We have to enforce the stronger discrete entropy inequalities

$$\frac{\eta(w_i^{n+1}) - \eta(w_i^n)}{\Delta t} + \frac{G_{i+1/2}^n - G_{i-1/2}^n}{\Delta x} \leq 0.$$

- We suggest to extend the *a posteriori* methods (MOOD) introduced in [Clain, Diot & Loubère '11].

# The family of entropies for the Euler equations

## Lemma

*The entropy pairs* $(\eta, \mathcal{G})$ *of the Euler system write*

$$\eta = \rho\psi(r), \quad \mathcal{G} = \rho\psi(r)u,$$

*where* $r = -\frac{p^{1/\gamma}}{\rho}$ *and* $\psi$ *is a smooth increasing convex function.*

We consider the scheme

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2} - F_{i-1/2}\right),$$

where $w_i^n = (\rho_i^n, \rho_i^n u_i^n, E_i^n)^T$ and $F_{i+1/2} = (F^{\rho}_{i+1/2}, F^{\rho u}_{i+1/2}, F^{E}_{i+1/2})^T$.

We introduce $r_{i+1/2}^n = \begin{cases} r_{i+1}^n & \text{if } F_{i+1/2}^\rho < 0 \\ r_i^n & \text{if } F_{i+1/2}^\rho > 0 \end{cases}.$

### Theorem

*Assume the scheme preserves $\Omega$. Assume the scheme satisfies the specific discrete entropy inequality*

$$\rho_i^{n+1} r_i^{n+1} \leq \rho_i^n r_i^n - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^\rho r_{i+1/2}^n - F_{i-1/2}^\rho r_{i-1/2}^n\right).$$

*Assume the additional CFL like condition*

$$\frac{\Delta t}{\Delta x}\left(\max\left(0, F_{i+1/2}^\rho\right) - \min\left(0, F_{i-1/2}^\rho\right)\right) \leq \rho_i^n.$$

*Then the scheme is entropy preserving: for all smooth increasing convex function $\psi$, we have*

$$\rho_i^{n+1}\psi(r_i^{n+1}) \leq \rho_i^n \psi(r_i^n) - \frac{\Delta t}{\Delta x}\left(F_{i+1/2}^\rho \psi(r_{i+1/2}^n) - F_{i-1/2}^\rho \psi(r_{i-1/2}^n)\right).$$

# First-order scheme

We consider a first-order scheme

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^n, w_{i+1}^n\right) - F\left(w_{i-1}^n, w_i^n\right)\right).$$

For a time step restricted according to the CFL condition

$$\frac{\Delta t}{\Delta x}\max_{i\in\mathbb{Z}}\left|\lambda^{\pm}\left(w_i^n, w_{i+1}^n\right)\right| \leq \frac{1}{2},$$

the first-order scheme is assumed to satisfy:

- Robustness: $\forall i \in \mathbb{Z}, \quad w_i^n \in \Omega \quad \Rightarrow \quad \forall i \in \mathbb{Z}, \quad w_i^{n+1} \in \Omega$
- Stability:

$$\rho_i^{n+1} r_i^{n+1} \leq \rho_i^n r_i^n - \frac{\Delta t}{\Delta x}\Big(F^\rho\left(w_i^n, w_{i+1}^n\right) r_{i+1/2}^n - F^\rho\left(w_{i-1}^n, w_i^n\right) r_{i-1/2}^n\Big).$$

Example: the HLLC/Suliciu relaxation scheme

# Reconstruction procedure

- We consider high-order reconstructed states $w_i^{n,\pm}$ on the cell $K_i$ at the interfaces $x_{i\pm1/2}$.
- These reconstructed states can be obtained by any reconstruction procedure (MUSCL, ENO/WENO, PPM, DMGR...).
- Assumptions:
  - The reconstruction is $\Omega$-preserving: $w_i^{n,\pm} \in \Omega$;
  - The reconstruction is conservative:

$$w_i^n = \frac{1}{2}\left(w_i^{n,-} + w_i^{n,+}\right).$$

# The e-MOOD algorithm

1. Reconstruction step: For all $i \in \mathbb{Z}$, we evaluate high-order reconstructed states $w_i^{n,\pm}$ located at the interfaces $x_{i\pm1/2}$.
2. Evolution step: The solution is evolved as follows:

$$w_i^{n+1,\star} = w_i^n - \frac{\Delta t}{\Delta x}\left(F\left(w_i^{n,+}, w_{i+1}^{n,-}\right) - F\left(w_{i-1}^{n,+}, w_i^{n,-}\right)\right).$$

3. A posteriori limitation step: We have the following alternative:
   - if for all $i \in \mathbb{Z}$, we have

$$\rho^{n+1,\star} r_i^{n+1,\star} \leq \rho_i^n r(w_i^n) - \frac{\Delta t}{\Delta x}\Big(F^\rho\left(w_i^{n,+}, w_{i+1}^{n,-}\right) r_{i+1/2}^n - F^\rho\left(w_{i-1}^{n,+}, w_i^{n,-}\right) r_{i-1/2}^n\Big), \quad (1)$$

     then the solution is valid and the updated solution at time $t^n + \Delta t$ is defined by $w_i^{n+1} = w_i^{n+1,\star}$;
   - otherwise, for all $i \in \mathbb{Z}$ such that (1) is not satisfied, we set $w_i^{n,\pm} = w_i^n$ and we go back to step 2.

## Theorem

*Assume the time step $\Delta t$ is chosen in order to satisfy the two following CFL like conditions:*

$$\frac{\Delta t}{\Delta x}\max_{i\in\mathbb{Z}}\left(\left|\lambda^{\pm}\left(w_i^{n,+},w_{i+1}^{n,-}\right)\right|,\left|\lambda^{\pm}\left(w_i^{n,-},w_i^{n,+}\right)\right|\right)\leq\frac{1}{4},$$

$$\frac{\Delta t}{\Delta x}\left(\max\left(0,F_{i+1/2}^{\rho}\right)-\min\left(0,F_{i-1/2}^{\rho}\right)\right)\leq\rho_i^n.$$

*Then the updated states $w_i^{n+1}$, given by the e-MOOD scheme, belong to $\Omega$. Moreover, for all smooth increasing convex function $\psi$, the e-MOOD scheme satisfies*

$$\frac{1}{\Delta t}\left(\rho_i^{n+1}\psi(r_i^{n+1})-\rho_i^n\psi\left(r_i^n\right)\right)+\frac{1}{\Delta x}\left(F^{\rho}\left(w_i^{n,+},w_{i+1}^{n,-}\right)\psi(r_{i+1/2}^n)\right.$$
$$\left.-F^{\rho}\left(w_{i-1}^{n,+},w_i^{n,-}\right)\psi(r_{i-1/2}^n)\right)\leq 0.$$

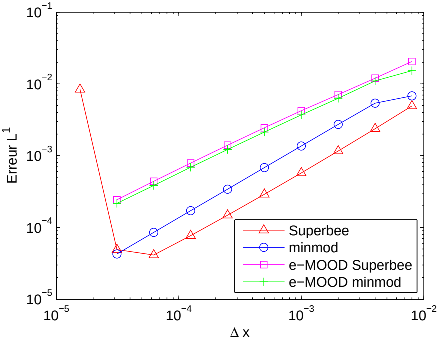
*The e-MOOD scheme is thus entropy preserving.*

# Numerical results obtained with a second-order time scheme

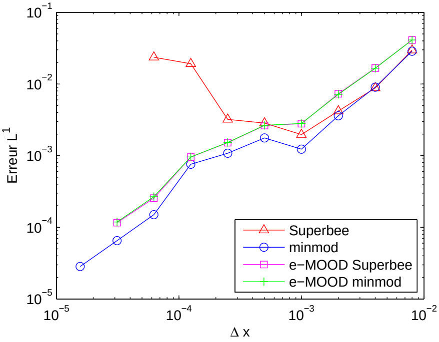
$L^1$ error:

$$\sum_i \left| \rho_i^N - \rho_{ex}(x_i, T) \right|$$

1-rarefaction




Double shock



1. Second-order schemes based on dual mesh gradient reconstruction
   - MUSCL scheme and CFL condition
   - The DMGR scheme
   - Numerical results
2. A high-order entropy preserving scheme with *a posteriori* limitation
   - Motivations
   - From one to all discrete entropy inequalities
   - The e-MOOD scheme for the Euler equations
3. Well-balanced schemes for systems with nonlinear source terms
   - The Ripa model
   - Relaxation models
   - The relaxation scheme
   - Numerical results

# The Ripa model

$$
\begin{cases}
\partial_t h + \partial_x hu = 0 \\
\partial_t hu + \partial_x \left(hu^2 + gh^2\theta/2\right) = -gh\theta\partial_x Z \\
\partial_t h\theta + \partial_x h\theta u = 0
\end{cases}
$$

- $h$: water height
- $u$: velocity
- $\theta$: temperature
- $g$: gravity constant
- $Z(x)$: smooth topography function

# The Ripa model

$$\begin{cases} \partial_t h + \partial_x hu = 0 \\ \partial_t hu + \partial_x \left( hu^2 + gh^2\theta/2 \right) = -gh\theta\partial_x Z \\ \partial_t h\theta + \partial_x h\theta u = 0 \end{cases}$$

- We define
  - the vector of conservative variables $w = (h, hu, h\theta)^T$,
  - the flux function $f(w) = (hu, hu^2 + gh^2\theta/2, h\theta u)^T$,
  - the source term $s(w) = (0, -gh\theta, 0)^T$,

  to rewrite the system into the compact form

$$\partial_t w + \partial_x f(w) = s(w)\partial_x Z.$$

- The set of physical admissible states is

$$\Omega = \left\{ w \in \mathbb{R}^3, \quad h > 0, \quad \theta > 0 \right\}.$$

# The Ripa model

$$\begin{cases} \partial_t h + \partial_x hu = 0 \\ \partial_t hu + \partial_x \left( hu^2 + gh^2\theta/2 \right) = -gh\theta\partial_x Z \\ \partial_t h\theta + \partial_x h\theta u = 0 \end{cases}$$

## Steady states

The steady states at rest are governed by the ODE

$$\begin{cases} u \equiv 0, \\ \partial_x (h^2\theta/2) = -h\theta\partial_x Z. \end{cases}$$

We cannot obtain an explicit expression of all the steady states.

## Lake at rest type solutions

$$\begin{cases} u = 0, \\ \theta = \text{cst}, \\ h + Z = \text{cst}, \end{cases} \qquad \begin{cases} u = 0, \\ z = \text{cst}, \\ h^2\theta = \text{cst}, \end{cases} \qquad \begin{cases} u = 0, \\ h = \text{cst}, \\ z + \frac{h}{2}\ln\theta = \text{cst}. \end{cases}$$

# The relaxation method without source term

- Initial system:

$$\partial_t w + \partial_x f(w) = 0. \tag{1}$$

- Relaxation system:

$$\partial_t W + \partial_x F(W) = \frac{1}{\varepsilon} R(W), \tag{2}$$

  - (2) should formally gives back (1) when $\varepsilon \to 0$.
  - (2) should be “simpler” than (1) (e.g. only linearly degenerate fields)

- The relaxation scheme is based on a splitting strategy:

  Time evolution: We evolve the initial data by the Godunov scheme for the system $\partial_t W + \partial_x F(W) = 0$ (i.e. $\varepsilon = +\infty$).

  Relaxation: We take into account the relaxation source term by solving $\partial_t W = \frac{1}{\varepsilon} R(W)$ then taking the limit for $\varepsilon \to 0$.

# The Suliciu model ([Suliciu '98], [Bouchut '04]...)

$$\begin{cases} \partial_t h + \partial_x hu = 0 \\ \partial_t hu + \partial_x (hu^2 + \pi) = -gh\theta \partial_x Z \\ \partial_t h\theta + \partial_x h\theta u = 0 \\ \partial_t h\pi + \partial_x (u(h\pi + \nu^2)) = \frac{h}{\varepsilon}(gh^2\theta/2 - \pi) \\ \partial_t Z = 0 \end{cases}$$

| Eigenvalues | Riemann invariants |
| --- | --- |
| $u \pm \frac{\nu}{h}$ | $u \pm \frac{\nu}{h}, \quad \pi \mp \nu u, \quad \theta, \quad Z$ |
| $u \ (\times 2)$ | $u, \quad \pi, \quad Z$ |
| $0$ | $hu, \quad \pi + \frac{\nu^2}{h}, \quad \theta, \quad g\theta Z + \frac{u^2}{2} - \frac{\nu^2}{2h^2}$ |

Difficulties to compute the solution of the Riemann problem:

- The order of the eigenvalues is not determined *a priori*.
- There are strong nonlinearities in the Riemann invariants for the eigenvalue 0.

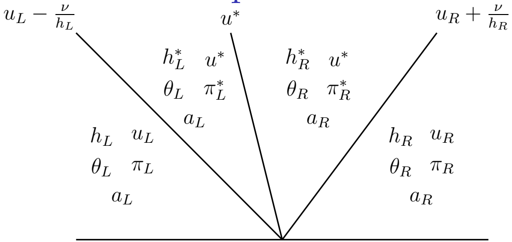
# Relaxation model with moving topography

$$\begin{cases} \partial_t h + \partial_x hu = 0 \\ \partial_t hu + \partial_x (hu^2 + \pi) = -gh\theta\partial_x a \\ \partial_t h\theta + \partial_x h\theta u = 0 \\ \partial_t h\pi + \partial_x (u(h\pi + \nu^2)) = \frac{h}{\varepsilon}(gh^2\theta/2 - \pi) \\ \partial_t a + u\partial_x a = \frac{1}{\varepsilon}(Z - a) \end{cases}$$

| Eigenvalues | Riemann invariants |
|---|---|
| $u \pm \frac{\nu}{h}$ | $u \pm \frac{\nu}{h}$, $\pi \mp \nu u$, $\theta$, $a$ |
| $u$ $(\times 3)$ | $u$ |

- The order of the eigenvalues is fixed: $u - \frac{\nu}{h} < u < u + \frac{\nu}{h}$
- There is a missing invariant for the eigenvalue $u$
  $\Rightarrow$ we need a closure equation

# The Riemann problem




- 5 unknowns: $u^*,\ h^*_{L,R},\ \pi^*_{L,R}$
- 4 equations given by the Riemann invariants $u \pm \frac{\nu}{h},\ \pi \mp \nu u$

Closure equation: $\pi_R^* - \pi_L^* = -g\bar{h}(w_L, w_R)\bar{\theta}(w_L, w_R)(a_R - a_L)$

- $\bar{h}(w_L, w_R)$: $h$-average s.t. $\bar{h}(w, w) = h$ and $\bar{h}(w_L, w_R) = \bar{h}(w_R, w_L)$
- $\bar{\theta}(w_L, w_R)$: $\theta$-average s.t. $\bar{\theta}(w, w) = \theta$ and $\bar{\theta}(w_L, w_R) = \bar{\theta}(w_R, w_L)$

Solution of the Riemann problem

$$u^* = \frac{u_L + u_R}{2} - \frac{\pi_R - \pi_L}{2\nu} - \frac{g}{2\nu}\bar{h}(w_L, w_R)\bar{\theta}(w_L, w_R)(a_R - a_L)$$

$$\pi_L^* = \pi_L + \nu(u_L - u^*) \qquad \pi_R^* = \pi_R + \nu(u^* - u_R)$$

$$\frac{1}{h_L^*} = \frac{1}{h_L} + \frac{u^* - u_L}{\nu} \qquad \frac{1}{h_R^*} = \frac{1}{h_R} + \frac{u_R - u^*}{\nu}$$

# Reformulation into a fully determined model

$$\begin{cases} \partial_t h + \partial_x hu = 0 \\ \partial_t hu + \partial_x (hu^2 + \pi) = -g\bar{h}(X^-, X^+)\bar{\theta}(X^-, X^+)\partial_x a \\ \partial_t h\theta + \partial_x h\theta u = 0 \\ \partial_t h\pi + \partial_x (u(h\pi + \nu^2)) = \frac{h}{\varepsilon}(gh^2\theta/2 - \pi) \\ \partial_t a + u\partial_x a = \frac{1}{\varepsilon}(Z - a) \\ \partial_t X^- + (u-\delta)\partial_x X^- = \frac{1}{\varepsilon}(W - X^-) \\ \partial_t X^+ + (u+\delta)\partial_x X^+ = \frac{1}{\varepsilon}(W - X^+) \end{cases}$$

| Eigenvalues | Riemann invariants |
|---|---|
| $u \pm \frac{\nu}{h}$ | $u \pm \frac{\nu}{h}, \quad \pi \mp \nu u, \quad \theta, \quad a, \quad X^-, \quad X^+$ |
| $u\ (\times 3)$ | $u, \quad \pi + g\bar{h}(X^-, X^+)\bar{\theta}(X^-, X^+)a, \quad X^-, \quad X^+$ |
| $u - \delta$ | $h, \quad u, \quad \theta, \quad \pi, \quad a, \quad X^+$ |
| $u + \delta$ | $h, \quad u, \quad \theta, \quad \pi, \quad a, \quad X^-$ |

Both models lead to the same numerical scheme.

# The relaxation scheme

The relaxation scheme associated with both previous models writes

$$w_i^{n+1} = w_i^n - \frac{\Delta t}{\Delta x}\left(f(w_i^n, w_{i+1}^n) - f(w_{i-1}^n, w_i^n)\right) + \frac{\Delta t}{2}\left(s^+(w_{i-1}^n, w_i^n)\frac{Z_i - Z_{i-1}}{\Delta x} + s^-(w_i^n, w_{i+1}^n)\frac{Z_{i+1} - Z_i}{\Delta x}\right),$$

where the numerical flux is defined by

$$f(w_L,w_R)=\begin{cases} \left(h_L u_L, \quad h_L u_L^2 + g h_L^2 \theta_L/2, \quad h_L \theta_L u_L\right)^T & \text{if } u_L - \frac{\nu}{h_L} > 0, \\ \left(h_L^* u^*, \quad h_L^* (u^*)^2 + \pi_L^*, \quad h_L^* \theta_L u^*\right)^T & \text{if } u_L - \frac{\nu}{h_L} < 0 < u^*, \\ \left(h_R^* u^*, \quad h_R^* (u^*)^2 + \pi_R^*, \quad h_R^* \theta_R u^*\right)^T & \text{if } u^* < 0 < u_R + \frac{\nu}{h_R}, \\ \left(h_R u_R, \quad h_R u_R^2 + g h_R^2 \theta_R/2, \quad h_R \theta_R u_R\right)^T & \text{if } u_R + \frac{\nu}{h_R} < 0, \end{cases}$$

and the numerical source terms are defined by

$$s^+(w_L,w_R)=\left(0, -(\mathrm{sgn}(u^*)+1) g \bar{h}(w_L,w_R)\bar{\theta}(w_L,w_R), 0\right)^T,$$

$$s^-(w_L,w_R)=\left(0, -(1-\mathrm{sgn}(u^*)) g \bar{h}(w_L,w_R)\bar{\theta}(w_L,w_R), 0\right)^T.$$

# Properties of the relaxation scheme (1)

## Theorem (Exact preservation of lake at rest solutions)

*Assume the average functions* $\bar{h}$ *and* $\bar{\theta}$ *are defined by*

$$\bar{h}(w_L, w_R) = \frac{1}{2}(h_L + h_R), \quad \bar{\theta}(w_L, w_R) = \begin{cases} \frac{\theta_R - \theta_L}{\ln(\theta_R) - \ln(\theta_L)} & \text{if } \theta_L \neq \theta_R, \\ \theta_L & \text{if } \theta_L = \theta_R. \end{cases}$$

*Then the relaxation scheme preserves exactly the three lake at rest solutions: if the initial data* $w_i^0$ *is given by*

$$\begin{cases} u_i^0 = 0, \\ \theta_i^0 = \theta, \\ h_i^0 + z_i = H, \end{cases} \quad \text{or} \quad \begin{cases} u_i^0 = 0, \\ z_i = Z, \\ \left(h_i^0\right)^2 \theta_i^0 = P, \end{cases} \quad \text{or} \quad \begin{cases} u_i^0 = 0, \\ h_i^0 = H, \\ z_i + h_i^0 \ln(\theta_i^0)/2 = P, \end{cases}$$

*where* $\theta > 0$, $H > 0$ *and* $P > 0$ *are constants, then the approximate solution* $w_i^n$ *stays at rest:*

$$w_i^n = w_i^0, \quad \forall i \in \mathbb{Z}, \quad \forall n \in \mathbb{N}.$$

# Properties of the relaxation scheme (2)

## Theorem (Well-balancedness)

*Assume the initial data $w_i^0$ satisfies for all $i \in \mathbb{Z}$:*

$$\begin{cases} u_i^0 = 0, \\ (h_{i+1}^0)^2\theta_{i+1}^0/2 - (h_i^0)^2\theta_i^0/2 = -\bar{h}(w_i^0, w_{i+1}^0)\bar{\theta}(w_i^0, w_{i+1}^0)(Z_{i+1} - Z_i). \end{cases}$$

*Then the approximate solution stays at rest: $w_i^n = w_i^0, \forall i \in \mathbb{Z}, \forall n \in \mathbb{N}$.*

## Theorem (Robustness)

*Assume the parameter $\nu$ satisfies the following inequalities:*

$$u_L - \frac{\nu}{h_L} < u^* < u_R + \frac{\nu}{h_R}.$$

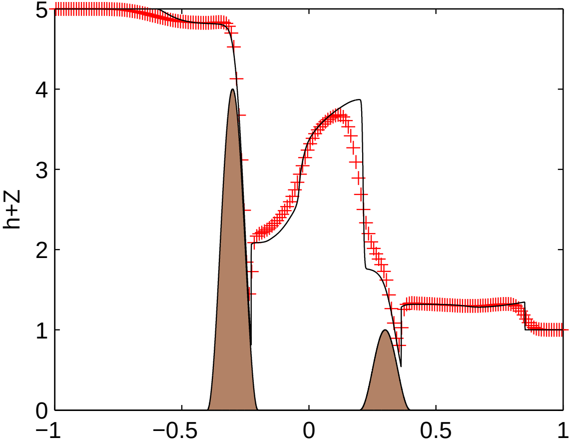
*Assume the following CFL condition is satisfied:*

$$\frac{\Delta t}{\Delta x} \max_{i \in \mathbb{Z}} |u_i^n \pm \nu/h_i^n| \leq \frac{1}{2}$$

*Then the relaxation scheme preserves the set of physical states:*

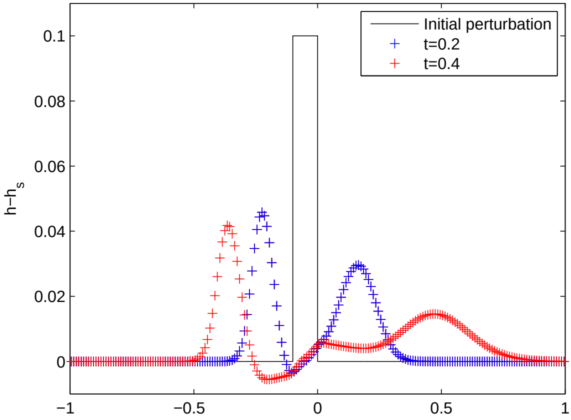
$$\forall i \in \mathbb{Z},\ h_i^n > 0 \text{ and } \theta_i^n > 0 \quad \Rightarrow \quad \forall i \in \mathbb{Z},\ h_i^{n+1} > 0 \text{ and } \theta_i^{n+1} > 0.$$

# Dam break over a non-flat bottom [Chertock, Kurganov & Liu '13]



# Perturbation of a nonlinear steady state

- Topography:
  $Z(x)=-2e^x$
- Steady state solution:
  $(h_s,u_s,\theta_s)(x)=(e^x,0,e^{2x})$
- Initial perturbation:
  $\delta h(x,0)=0.1\chi_{[-0.1,0]}(x)$



# Perspectives

- Extension of the DMGR scheme to higher-order.
- Combination of the DMGR and e-MOOD methods to get high-order entropic schemes in 2D.
- Extension of the e-MOOD scheme to other systems.
- Developement of high-order well-balanced schemes for systems with source terms.
- Does the relaxation schemes for Ripa and Euler with gravity satisfy discrete entropy inequalities ?