Application of the Entropy Viscosity Method and the Flux-Corrected Transport Algorithm to Scalar Transport Equations and the Shallow Water Equations

Joshua E. Hansel

ADVISED BY

Jean C. Ragusa

Department of Nuclear Engineering
Texas A&M University

Ph.D. Defense

May 13th, 2016

Radiation Transport

- Consider the following model for radiation transport:

$$\frac{1}{v}\frac{\partial\psi}{\partial t}+\mathbf{\Omega}\cdot\nabla\psi(\mathbf{x},\mathbf{\Omega},E,t)+\Sigma_{\mathrm{t}}(\mathbf{x},E)\psi(\mathbf{x},\mathbf{\Omega},E,t)=Q(\mathbf{x},\mathbf{\Omega},E,t)\,. \quad (1)$$

- For example, $Q\left(\mathbf{x},t\right)$ can include extraneous sources and scattering sources in a source iteration scheme:

$$\frac{1}{v}\frac{\partial\psi}{\partial t}^{(\ell+1)}+\mathbf{\Omega}\cdot\nabla\psi^{(\ell+1)}+\Sigma_{\mathrm{t}}(\mathbf{x})\psi^{(\ell+1)}=Q^{(\ell)}\,, \quad (2)$$

with for example, $Q^{(\ell)}=\frac{Q_{ext}}{4\pi}+\frac{\Sigma_s}{4\pi}\phi^{(\ell)}$.

Scalar Conservation Law Model

- This research considers the following scalar conservation law model:

$$\frac{\partial u}{\partial t} + \nabla \cdot \mathbf{f}(u) + \sigma(\mathbf{x}) u(\mathbf{x}, t) = q(\mathbf{x}, t), \qquad (3)$$

where $\sigma(\mathbf{x}) \geq 0$ and $q(\mathbf{x}, t) \geq 0$.

- Some examples of applicability include

Linear advection: advection of some quantity, such as a tracer	$\mathbf{f}(u) = \mathbf{v}u$
Inviscid Burgers' equation: inviscid fluid flow model capturing some key features of gas dynamics	$\mathbf{f}(u) = \frac{1}{2}u^2\mathbf{1}$
Radiation transport: transport of radiation such as photons or neutrons	$\mathbf{f}(u) = v\mathbf{\Omega}u$

Shallow Water Equations (SWE)

- The shallow water equations are

$$\frac{\partial}{\partial t}\begin{bmatrix} h \\ \mathbf{q} \end{bmatrix} + \nabla \cdot \begin{bmatrix} \mathbf{q} \\ \mathbf{q} \otimes \mathbf{v} + \frac{1}{2} g h^2 \mathbb{I} \end{bmatrix} = \begin{bmatrix} 0 \\ -gh\nabla b \end{bmatrix}. \tag{4}$$

  - $h(\mathbf{x},t) \equiv \int \rho(\mathbf{x},t)\, dz$ is referred to as "height".
  - $\mathbf{q}(\mathbf{x},t) \equiv h(\mathbf{x},t)\,\mathbf{v}(\mathbf{x},t)$ is referred to as "discharge".
  - g is the acceleration due to gravity.
  - $b(\mathbf{x})$ is the bottom topography profile, or "bathymetry".
- Obtained by depth-integrating Navier-Stokes equations with the assumption that the horizontal length scale is much greater than the vertical length scale
- Some important simulations include
  - dam break problems: for example, what wave structures will develop in a dam-break accident?
  - wave/tsunami problems: for example, what height of seawall is necessary to stop a wave of a certain height?

Modeling Conservation Laws

- The strong form of a conservation law PDE

$$u_t + f(u)_x = 0 \tag{5}$$

  does not admit discontinuous solutions, so an integral, or weak form is used:

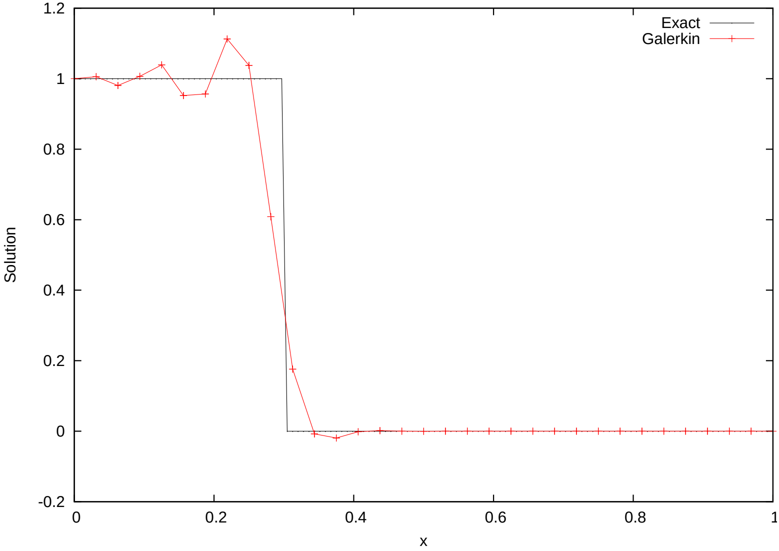
$$\int u_t \phi dx + \int f(u)_x \phi dx = 0\,. \tag{6}$$

- The solution to the weak problem is not necessarily unique.
  - Some physics have been neglected in the hyperbolic PDE model.
  - In particular, hyperbolic PDEs ignore diffusive/viscous effects.
- The physically correct solution is the vanishing viscosity solution, where one takes the limit as $\epsilon \to 0$:

$$u_t^\epsilon + f(u^\epsilon)_x = \epsilon u_{xx}^\epsilon\,. \tag{7}$$

A Simple Motivating Example

Linear Advection of Discontinuous Wave Front



Entropy in Conservation Laws

- Additional conditions/physics are required to filter out spurious weak solutions to obtain the physically correct solution.
  - Gas dynamics has the second law of thermodynamics → entropy conditions.
  - Other PDE systems frequently have similar conditions.
- Scalar conservation law entropy condition:

$$\eta(u)_t + \psi(u)_x \leq 0\,, \tag{8}$$

  for all convex entropy $\eta(u)$ and corresponding flux $\psi(u)$.
- The idea of the entropy viscosity method is to enforce the entropy condition with dissipation proportional to the violation of the condition.

Objectives

- The objectives of this research are the following:
  - Accurately solve conservation law problems using the continuous finite element method (CFEM):
    - 2nd-order-accuracy in space (for smooth problems)
    - convergence to the entropy solution
  - Prevent spurious oscillations.
    - Complete immunity to spurious oscillations remains unattainable for high-order schemes, but quality results are possible in practice.
  - Prevent solution from leaving physically-motivated bounds.
  - Prevent negativities for physically non-negative quantities.

Introduction to Flux-Corrected Transport (FCT)

- FCT initially developed in 1973 for finite difference methods and was more recently applied to CFEM.
- FCT mixes
  - a high-order scheme, which may have spurious oscillations and negativities, and
  - a low-order scheme, which has the desired properties such as monotonicity preservation and positivity preservation.
- The FCT idea: Reverse as much artificial diffusion in the low-order scheme as possible without producing an unphysical solution.
- Summary of the FCT algorithm:
  1. Derive antidiffusion source from low-order scheme to high-order scheme.
  2. Decompose antidiffusion source into antidiffusive fluxes between nodes.
  3. Define physically-motivated solution bounds:

$$u_i^{min} \leq u_i \leq u_i^{max} \quad \forall i\,. \tag{9}$$

  4. Apply a limiter to the antidiffusive fluxes to ensure that solution bounds are not violated.

Outline

- The scalar case
  - Problem formulation
  - Discretization
  - Low-order, DMP-satisfying, positivity-preserving scheme
  - High-order, entropy-based scheme
  - High-order, DMP-satisfying, positivity-preserving FCT scheme
  - Scalar results
- The systems case (using the shallow water equations)
  - Problem formulation
  - Discretization
  - Low-order, domain-invariant, positivity-preserving scheme
  - High-order, entropy-based scheme
  - High-order, positivity-preserving FCT scheme
  - Shallow water equations results
- Conclusions

Scalar Transport

Scalar Problem Formulation

- Consider a linear conservation law model ($\mathbf{f}(u) = \mathbf{v}u$):

$$\frac{\partial u}{\partial t} + \nabla \cdot (\mathbf{v}u) + \sigma(\mathbf{x})u(\mathbf{x}, t) = q(\mathbf{x}, t) \ , \tag{10}$$

  where $\sigma(\mathbf{x}) \geq 0$ and $q(\mathbf{x}, t) \geq 0$.

- Provide initial conditions and some boundary condition, such as Dirichlet:

$$u(\mathbf{x}, 0) = u^0(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{D} \tag{11}$$

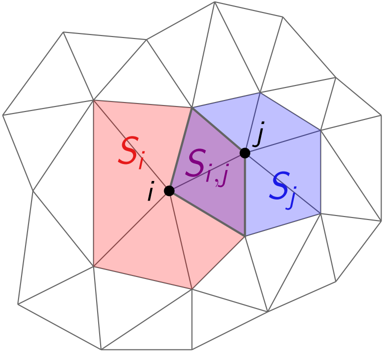
$$u(\mathbf{x}, t) = u^{\text{inc}}(\mathbf{x}) \quad \forall \mathbf{x} \in \partial\mathcal{D}^{\text{inc}} \tag{12}$$

- CFEM solution:

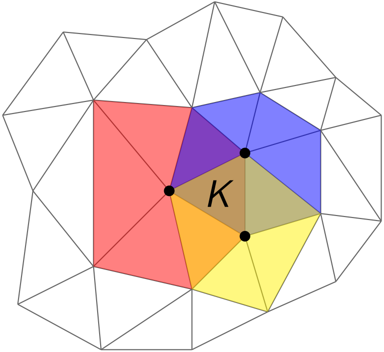
$$\tilde{u}(\mathbf{x}, t) = \sum_{j=1}^{N} U_j(t)\varphi_j(\mathbf{x}) \, , \quad \varphi_j(\mathbf{x}) \in P_h^1 \tag{13}$$

Some Notation

- Let $S_{i,j}$ be the shared support of test functions φ_i and φ_j.
- Let $\mathcal{K}(S_{i,j})$ be the set of indices of cells in $S_{i,j}$.



- Let $\mathcal{I}(K)$ be the set of DoF indices on cell K.
- Let n_K be the number of DoFs on cell K.



Discretization

- Discretizing with forward Euler (FE) in time and Galerkin CFEM in space gives

$$\mathbf{M}^C \frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} + \mathbf{A}\mathbf{U}^n = \mathbf{b}^n \,, \tag{14}$$

where

$$M^C_{i,j} \equiv \int\limits_{S_{i,j}} \varphi_i(\mathbf{x})\varphi_j(\mathbf{x}) d\mathbf{x} \,, \tag{15}$$

$$A_{i,j} \equiv \int\limits_{S_{i,j}} \left(\mathbf{v} \cdot \nabla\varphi_j(\mathbf{x}) + \sigma(\mathbf{x})\varphi_j(\mathbf{x})\right) \varphi_i(\mathbf{x}) d\mathbf{x} \,, \tag{16}$$

$$b^n_i \equiv \int\limits_{S_i} q(\mathbf{x}, t^n)\varphi_i(\mathbf{x}) d\mathbf{x} \,. \tag{17}$$

Temporal Discretization

- Forward Euler (FE):

$$\mathbf{M}^C \frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} + \mathbf{A}\mathbf{U}^n = \mathbf{b}^n \,, \tag{18}$$

- Strong Stability Preserving Runge-Kutta (SSPRK) Methods:

$$\hat{\mathbf{U}}^0 = \mathbf{U}^n \tag{19a}$$

$$\hat{\mathbf{U}}^i = \gamma_i \mathbf{U}^n + \zeta_i \bar{\mathbf{U}}^i \,, \quad \mathbf{M}^C \frac{\bar{\mathbf{U}}^i - \hat{\mathbf{U}}^{i-1}}{\Delta t} + \mathbf{A}\hat{\mathbf{U}}^{i-1} = \mathbf{b}(t_i) \,, \tag{19b}$$

$$\mathbf{U}^{n+1} = \hat{\mathbf{U}}^s \tag{19c}$$

  where $t_i \equiv t^n + c_i \Delta t$, and $\{\gamma_i, \zeta_i, c_i\}_{i=1}^s$ depend on the method.

- Theta Method:

$$\mathbf{M}^C \frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} + \mathbf{A}\mathbf{U}^\theta = \mathbf{b}^\theta \tag{20a}$$

$$\mathbf{U}^\theta \equiv (1-\theta)\mathbf{U}^n + \theta \mathbf{U}^{n+1} \,, \quad \mathbf{b}^\theta \equiv (1-\theta)\mathbf{b}^n + \theta \mathbf{b}^{n+1} \tag{20b}$$

  where $0 \le \theta \le 1$.

Boundary Conditions

- The following 3 methods for imposing incoming flux boundary conditions for node i are considered:
  1. Strongly impose: *replace* equation i with the equation $U_i = u^{\text{inc}}(\mathbf{x}_i)$ (or some multiple of it).
  2. Weakly impose: evaluate incoming boundary fluxes in equation i with values $u^{\text{inc}}(\mathbf{x}_j)$ instead of values U_j.
  3. Weakly impose with boundary penalty: *add* to equation i a multiple α_i of the equation $U_i = u^{\text{inc}}(\mathbf{x}_i)$, where α_i should be large enough such that this *penalty* equation dominates the original equation i.
- The choice of incoming flux boundary condition will later to be shown to have consequences on conservation for FCT.

Low-Order Scheme

- To get the low-order scheme, one does the following:
  - Lumps the mass matrix: $\mathbf{M}^C \rightarrow \mathbf{M}^L$.
  - Adds a low-order diffusion operator: $\mathbf{A} \rightarrow \mathbf{A} + \mathbf{D}^L$.
- This gives the following, where $\mathbf{U}^{L,n+1}$ is the low-order solution:

$$\mathbf{M}^L \frac{\mathbf{U}^{L,n+1} - \mathbf{U}^n}{\Delta t} + (\mathbf{A} + \mathbf{D}^L)\mathbf{U}^n = \mathbf{b}^n \,. \tag{21}$$

- The diffusion matrix $\mathbf{D}^L$ is assembled elementwise, where K denotes an element, using a local bilinear form b_K and a local low-order viscosity ν_K^L:

$$D_{i,j}^L = \sum_{K \in \mathcal{K}(S_{i,j})} \nu_K^L b_K(\varphi_j, \varphi_i) \,. \tag{22}$$

Local Viscous Bilinear Form

- The local viscous bilinear form is defined as follows, where V_K denotes the volume of element K:

$$b_K(\varphi_j,\varphi_i) \equiv \begin{cases} -\frac{1}{n_K-1}V_K & i \neq j, \quad i,j \in \mathcal{I}(K) \\ V_K & i = j, \quad i,j \in \mathcal{I}(K) \\ 0 & i \notin \mathcal{I}(K) \,|\, j \notin \mathcal{I}(K) \end{cases} . \tag{23}$$

- Some properties that result from this definition are

$$b_K(\varphi_i,\varphi_i) > 0\,. \tag{24a}$$

$$b_K(\varphi_j,\varphi_i) < 0 \quad j \neq i\,. \tag{24b}$$

$$\sum_j b_K(\varphi_j,\varphi_i) = 0\,. \tag{24c}$$

- The local viscous matrix in 1-D is

$$\mathbf{D}_K = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \Delta x\,. \tag{25}$$

Low-Order Viscosity

- The low-order viscosity is defined as

$$\nu_K^{L,n} \equiv \max_{i\neq j\in\mathcal{I}(K)} \frac{\max(0, A_{i,j}^n)}{-\sum\limits_{T\in\mathcal{K}(S_{i,j})} b_T(\varphi_j,\varphi_i)}\,. \qquad (26)$$

- This definition is designed to be the smallest number such that the following is guaranteed:

$$D_{i,j}^L \leq -A_{i,j}, \quad j \neq i\,. \qquad (27)$$

- This is used to guarantee that the low-order steady-state matrix $\mathbf{A}^L = \mathbf{A} + \mathbf{D}^L$ is an M-matrix.

M-Matrix Property

- An M-matrix, also known as a monotone matrix, or inverse-positive matrix has the property

$$\mathbf{A}\mathbf{x} = \mathbf{b} \geq 0 \Rightarrow \mathbf{x} \geq 0\,. \tag{28}$$

- Therefore one can prove non-negativity of the solution of a linear system $\mathbf{A}\mathbf{U} = \mathbf{b}$ by proving that $\mathbf{A}$ is an M-matrix and $\mathbf{b}$ is non-negative.
- An M-matrix can be identified by verifying the following properties:

$$A_{i,j} \leq 0 \quad j \neq i \quad \forall i \tag{29a}$$

$$A_{i,i} \geq 0 \quad \forall i \tag{29b}$$

$$\sum_j A_{i,j} \geq 0 \quad \forall i \tag{29c}$$

Local Discrete Maximum Principle for Explicit Euler

- If the following CFL-like condition is satisfied:

$$\Delta t \leq \frac{M^L_{i,i}}{A^L_{i,i}} \quad \forall i \,, \tag{30}$$

then the low-order solution for explicit Euler satisfies the following local discrete maximum principle (DMP):

$$W_i^{\text{DMP},-} \leq U_i^{L,n+1} \leq W_i^{\text{DMP},+} \qquad \forall i \,, \tag{31a}$$

$$W_i^{\text{DMP},\pm} \equiv U^n_{\substack{\max \\ \min},i} \left(1 - \frac{\Delta t}{M^L_{i,i}} \sum_j A^L_{i,j}\right) + \frac{\Delta t}{M^L_{i,i}} b_i^n \,, \tag{31b}$$

$$U^n_{\substack{\max \\ \min},i} = \underset{j\in\mathcal{I}(S_i)}{\substack{\max \\ \min}} U_j^n \,. \tag{31c}$$

Local Discrete Maximum Principle for Theta Method

- If the following CFL-like condition is satisfied:

$$\Delta t \leq \frac{M^L_{i,i}}{(1-\theta)A^L_{i,i}} \quad \forall i\,, \tag{32}$$

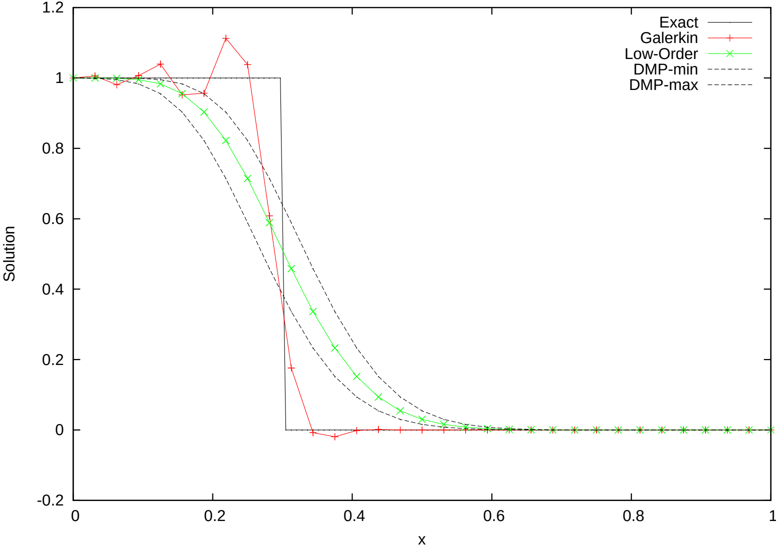
then the low-order theta method solution satisfies the following local DMP, which is implicit:

$$W_i^{\text{DMP},\pm}(\mathbf{U}^{L,n+1}) \equiv \frac{1}{1+\frac{\theta\Delta t}{M^L_{i,i}}A^L_{i,i}}\Bigg[\left(1-\frac{(1-\theta)\Delta t}{M^L_{i,i}}A^L_{i,i}\right)U^n_i - \frac{\Delta t}{M^L_{i,i}}\sum_{j\neq i}A^L_{i,j}\left((1-\theta)U^n_{\substack{\max\\ \min},j\neq i} + \theta U^{L,n+1}_{\substack{\max\\ \min},j\neq i}\right) + \frac{\Delta t}{M^L_{i,i}}b^\theta_i\Bigg]\,, \tag{33a}$$

$$U^n_{\substack{\max\\ \min},j\neq i} \equiv \underset{j\neq i\in\mathcal{I}(S_i)}{\overset{\max}{\min}} U^n_j\,. \tag{33b}$$

Low-Order Scheme Results Example

Linear Advection of Discontinuous Wave Front



Introduction to Entropy Viscosity

- The standard Galerkin CFEM method does not necessarily produce the correct weak solution and may even diverge. Even with FCT, it would not necessarily converge to the correct, physical weak solution, i.e., the *entropy* solution.
- To converge to the entropy solution, one must ensure that an entropy inequality is satisfied:

$$R(u) \equiv \frac{\partial \eta(u)}{\partial t} + \nabla \cdot \mathbf{f}^{\eta}(u) \leq 0 \tag{34}$$

  for any convex entropy $\eta(u)$ and corresponding entropy flux $\mathbf{f}^{\eta}(u)$.
- This *entropy residual* $R(u)$ measures entropy production; where it is positive, the inequality is violated, so the residual should be decreased somehow.
- To enforce the inequality, the entropy viscosity method adds viscosity in proportion to local entropy production, thus decreasing local entropy.

Entropy Viscosity Definition

- One chooses a convex entropy function $\eta(u)$ such as $\eta(u) = \frac{1}{2}u^2$ and manipulates the conservation law equation to get an entropy residual:

$$R(u) = \frac{\partial \eta}{\partial t} + \frac{d\eta}{du}\left(\nabla \cdot (\mathbf{v}u) + \sigma u - q\right) . \tag{35}$$

- Viscosity is set to be proportional to a linear combination of the local entropy residual $R_K(u) = \|R(u)\|_{L^\infty(K)}$ and entropy jumps $J_F(u)$ across the faces:

$$\nu_K^\eta \propto c_R R_K(u_h) + c_J \max_{F \in \partial K} J_F(u_h) . \tag{36}$$

- In practice, the entropy viscosity becomes the following, where the denominator is just a normalization constant:

$$\nu_K^\eta = \frac{c_R R_K(u_h) + c_J \max_{F \in \partial K} J_F(u_h)}{\|\eta(u_h) - \bar{\eta}(u_h)\|_{L^\infty(\mathcal{D})}} . \tag{37}$$

High-Order Scheme

- The high-order viscosity does not need to be any greater than the low-order viscosity:

$$\nu_K^{H,n} = \min(\nu_K^L, \nu_K^{\eta,n}) \, . \tag{38}$$

- For the high-order scheme, the mass matrix is not modified; the only change is the addition of the high-order diffusion operator $\mathbf{D}^{H,n}$: $\mathbf{A} \rightarrow \mathbf{A} + \mathbf{D}^{H,n}$:

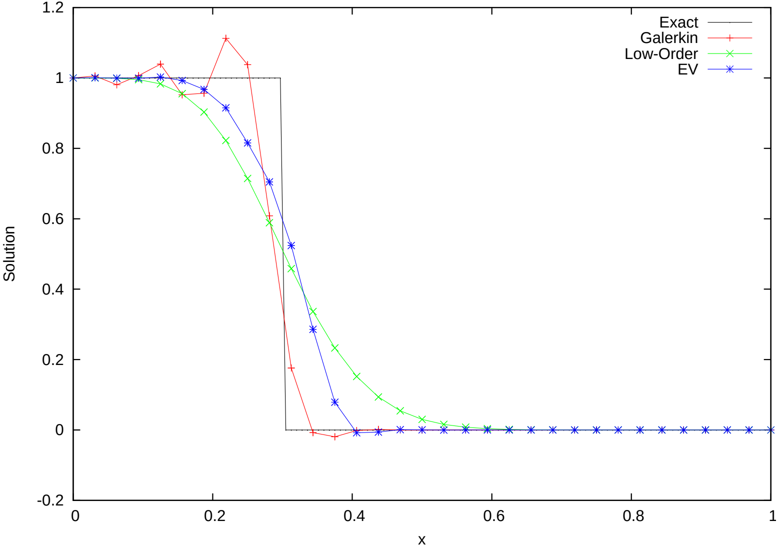
$$\mathbf{M}^C \frac{\mathbf{U}^{H,n+1} - \mathbf{U}^n}{\Delta t} + (\mathbf{A} + \mathbf{D}^{H,n})\mathbf{U}^n = \mathbf{b}^n \, . \tag{39}$$

- The high-order diffusion matrix is computed just as the low-order counterpart, except that $\nu_K^{H,n}$ is used instead of ν_K^L:

$$D_{i,j}^{H,n} = \sum_{K \in \mathcal{K}(S_{i,j})} \nu_K^{H,n} b_K(\varphi_j, \varphi_i) \, . \tag{40}$$

High-Order Scheme Results Example

Linear Advection of Discontinuous Wave Front



FCT Antidiffusive Flux Definition

- Recall that FCT defines antidiffusive correction fluxes from a low-order, monotone scheme to a high-order scheme. Calling these fluxes **p**, this gives

$$\mathbf{M}^L\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+(\mathbf{A}+\mathbf{D}^L)\mathbf{U}^n=\mathbf{b}^n+\mathbf{p}\,. \tag{41}$$

- Subtracting the high-order scheme equation from this gives the definition of **p**:

$$\mathbf{p}\equiv-(\mathbf{M}^C-\mathbf{M}^L)\frac{\mathbf{U}^H-\mathbf{U}^n}{\Delta t}+(\mathbf{D}^L-\mathbf{D}^{H,n})\mathbf{U}^n\,. \tag{42}$$

- Decomposing **p** into internodal fluxes $P_{i,j}$ such that $\sum_j P_{i,j}=p_i$,

$$P_{i,j}=-M^C_{i,j}\left(\frac{U^H_j-U^n_j}{\Delta t}-\frac{U^H_i-U^n_i}{\Delta t}\right)+(D^L_{i,j}-D^{H,n}_{i,j})\left(U^n_j-U^n_i\right)\,. \tag{43}$$

FCT Scheme Overview

- Recall that the objective of FCT is to limit these antidiffusive fluxes to enforce some physical solution bounds:

$$W_i^- \leq U_i^{n+1} \leq W_i^+ \qquad \forall i\,, \tag{44}$$

  for example, the low-order DMP bounds.

- This is achieved by applying a limiting coefficient $L_{i,j}$ to each antidiffusion flux $P_{i,j}$:

$$\mathbf{M}^L \frac{\mathbf{U}^{n+1} - \mathbf{U}^n}{\Delta t} + \mathbf{A}^L \mathbf{U}^n = \mathbf{b} + \bar{\mathbf{p}}\,, \tag{45}$$

  where $\bar{p}_i \equiv L_{i,j} P_{i,j}$.

- Each limiting coefficient is between zero and unity: $0 \leq L_{i,j} \leq 1$.
  - If all $L_{i,j}$ are zero, then the low-order scheme is produced.
  - If all $L_{i,j}$ are one, then the high-order scheme is produced.

Conservation

- FCT scheme is conservative if the net antidiffusion source is zero:

$$\sum_i \bar{p}_i = \sum_i \sum_j L_{i,j} P_{i,j} = 0\,. \tag{46}$$

- The antidiffusive flux decomposition choice yielded $P_{j,i} = -P_{i,j}$ and $P_{i,i} = 0$. Therefore $\sum_i \sum_j P_{i,j} = 0$.
- Then if one enforces symmetry on the limiting coefficients, then conservation is achieved:

$$L_{j,i} = L_{i,j} \quad \Rightarrow \quad \sum_i \bar{p}_i = 0\,. \tag{47}$$

- Caveat: When Dirichlet BC are *strongly* imposed, the FCT scheme is not conservative unless all antidiffusive fluxes from Dirichlet nodes are completely cancelled.
  - The flux decomposition is incorrect in the vicinity of Dirichlet nodes.
  - Equations for Dirichlet nodes overwrite contribution from antidiffusion sources.

Antidiffusion Bounds

- The solution bounds translate to antidiffusion bounds:

$$W_i^- \leq U_i^{n+1} \leq W_i^+ \quad \Rightarrow \quad Q_i^- \leq \bar{p}_i \leq Q_i^+ \,. \tag{48}$$

- For explicit Euler, $Q_i^\pm$ is found to be

$$Q_i^\pm \equiv M_{i,i}^L \frac{W_i^\pm - U_i^n}{\Delta t} + \sum_j A_{i,j}^L U_j^n - b_i^n \,. \tag{49}$$

- Most limiters assume

$$Q_i^+ \geq 0 \,, \quad Q_i^- \leq 0 \quad \forall i \,. \tag{50}$$

  - FCT starts from low-order solution: $\bar{p}_i = 0 \quad \Rightarrow \quad U_i^{FCT} = U_i^L$.
  - Some solution bounds automatically satisfy these requirements; otherwise one must enforce them:

$$Q_i^+ \leftarrow \max(0, Q_i^+) \,, \quad Q_i^- \leftarrow \min(0, Q_i^-) \,. \tag{51}$$

Limiters

- The objective of an FCT limiter is to maximize antidiffusion without violating imposed solution bounds:

  Find $0 \leq \{L_{i,j}\} \leq 1$ *such that* $\sum_i \bar{p}_i$ *is maximized, subject to to the constraints* $Q_i^- \leq \bar{p}_i \leq Q_i^+, \forall i$.

- Zalesak's limiter is the following:

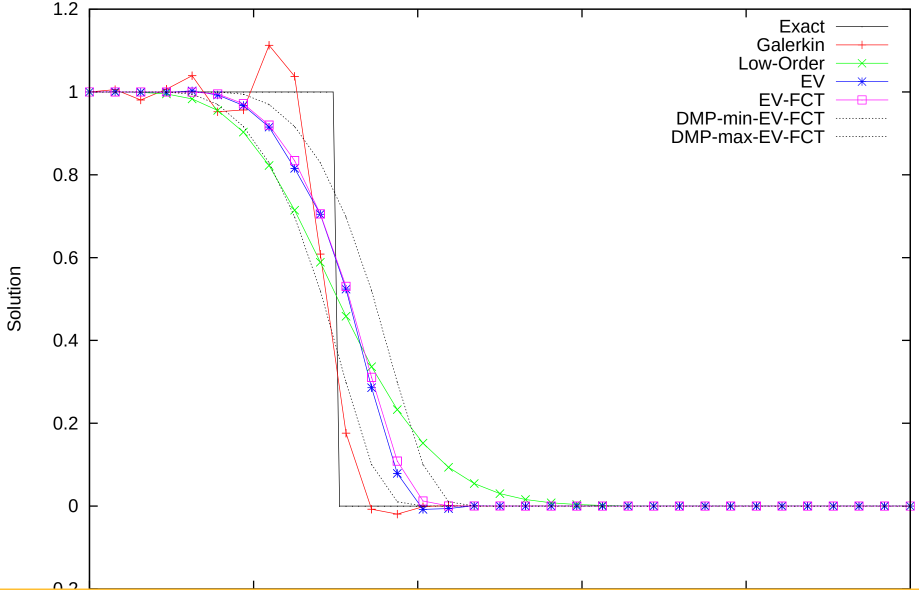
$$L_{i,j} \equiv \begin{cases} \min(L_i^+, L_j^-) & P_{i,j} \geq 0 \\ \min(L_i^-, L_j^+) & P_{i,j} < 0 \end{cases} , \tag{52a}$$

$$L_i^\pm \equiv \begin{cases} 1 & p_i^\pm = 0 \\ \min\left(1, \frac{Q_i^\pm}{p_i^\pm}\right) & p_i^\pm \neq 0 \end{cases} , \tag{52b}$$

$$p_i^- \equiv \sum_{j:P_{i,j}<0} P_{i,j} , \qquad p_i^+ \equiv \sum_{j:P_{i,j}>0} P_{i,j} . \tag{52c}$$

FCT Scheme Results Example

Linear Advection of Discontinuous Wave Front



Analytic Solution Bounds

- Solution bounds can be derived using the method of characteristics (MoC):

$$W_i^{\text{MoC},+} \equiv U^n_{\max,i} e^{-\Delta t \sigma_{\min,i}} + \frac{q_{\max,i}}{\sigma_{\min,i}} (1 - e^{-\Delta t \sigma_{\min,i}}) \,, \tag{53a}$$

$$W_i^{\text{MoC},-} \equiv U^n_{\min,i} e^{-\Delta t \sigma_{\max,i}} + \frac{q_{\min,i}}{\sigma_{\max,i}} (1 - e^{-\Delta t \sigma_{\max,i}}) \,. \tag{53b}$$

$$\sigma_{\max,i} = \max_{\mathbf{x}\in S_i} \sigma(\mathbf{x}) \,, \quad q_{\max,i} = \max_{\mathbf{x}\in S_i} q(\mathbf{x}) \,, \tag{53c}$$

- Recall the low-order DMP bounds for explicit Euler:

$$W_i^{\text{DMP},\pm} \equiv U^n_{\substack{\max\\\min},i} (1 - \Delta t \bar{\sigma}_i) + \Delta t \bar{q}_i \,, \tag{54a}$$

$$\bar{\sigma}_i \equiv \frac{\int\limits_{S_i} \sigma \varphi_i \, d\mathbf{x}}{\int\limits_{S_i} \varphi_i \, d\mathbf{x}} \,, \quad \bar{q}_i \equiv \frac{\int\limits_{S_i} q \varphi_i \, d\mathbf{x}}{\int\limits_{S_i} \varphi_i \, d\mathbf{x}} \tag{54b}$$

Nonlinear Iteration

- Implicit and steady-state time discretizations have nonlinear EV and FCT schemes.
- A fixed-point iteration scheme is used.
- The entropy viscosity scheme for implicit methods must iterate on the nonlinear entropy viscosities in the high-order diffusion matrix:

$$(\mathbf{A} + \mathbf{D}^{H,(\ell)})\mathbf{U}^{(\ell+1)} = \mathbf{b}\,, \tag{55}$$

- The FCT scheme for implicit methods must iterate on the nonlinear limiting coefficients:

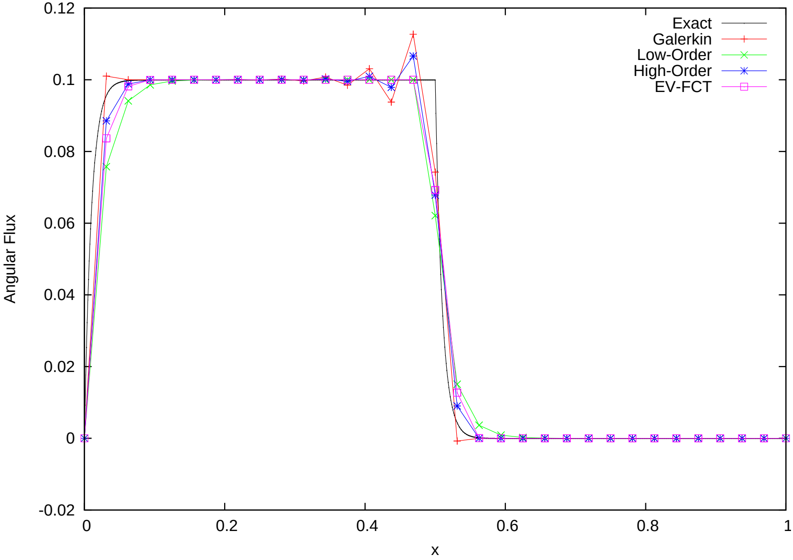
$$(\mathbf{A} + \mathbf{D}^{L})\mathbf{U}^{(\ell+1)} = \mathbf{b} + \bar{\mathbf{p}}^{(\ell)}\,, \tag{56a}$$

$$\bar{p}_i^{(\ell)} = \sum_j L_{i,j}^{(\ell)} P_{i,j}\,, \tag{56b}$$

where $L_{i,j}^{(\ell)}$ depends nonlinearly on $Q_i^{\pm,(\ell)}$, which depends on $W_i^{\pm,(\ell)}$.

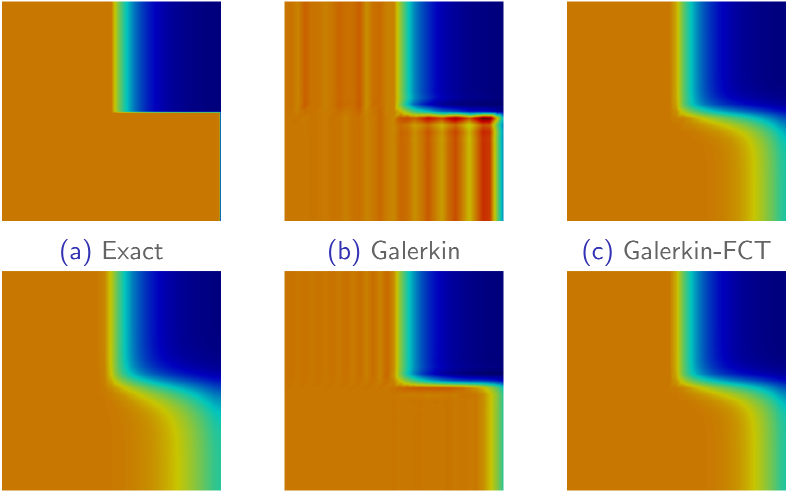
Source-in-Absorber Test Problem

Strong Absorber with Source in Left Half of Domain



2-D Void-to-Absorber Test Problem

Normally-Incident Wave from Void to Absorber Quadrant



(a) Exact

(b) Galerkin

(c) Galerkin-FCT

(d) Low-order

(e) EV

(f) EV-FCT

2-D Void-to-Absorber Test Problem

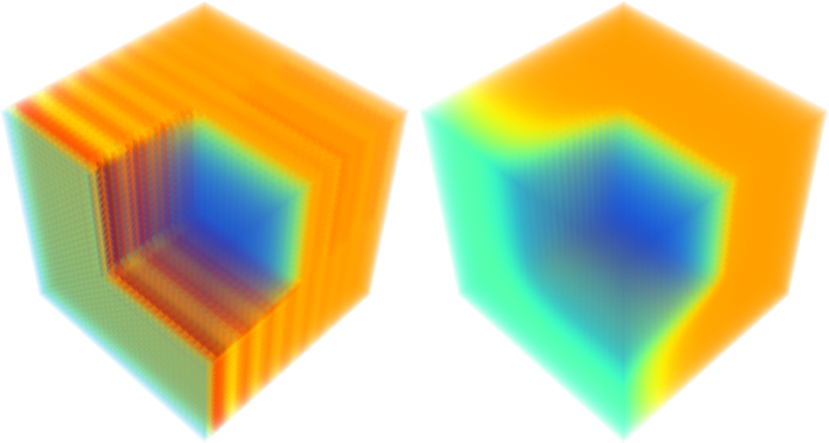
Number of Cells vs. Iterations Study, Steady-State

Table : EV and FCT Iterations Required for EV-FCT Solution

N_{cell}	EV	FCT
64	97	716
256	–	**FAIL**

3-D Void-to-Absorber Test Problem

Normally-Incident Wave from Void to Absorber Octant

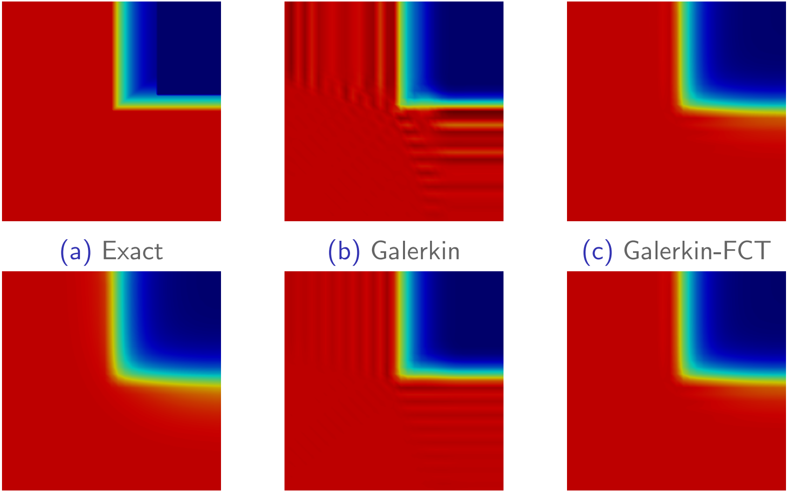


(a) Galerkin

(b) Galerkin-FCT

Skew-Incident Void-to-Absorber Test Problem

Skew-Incident Wave from Void to Absorber Quadrant



(a) Exact

(b) Galerkin

(c) Galerkin-FCT

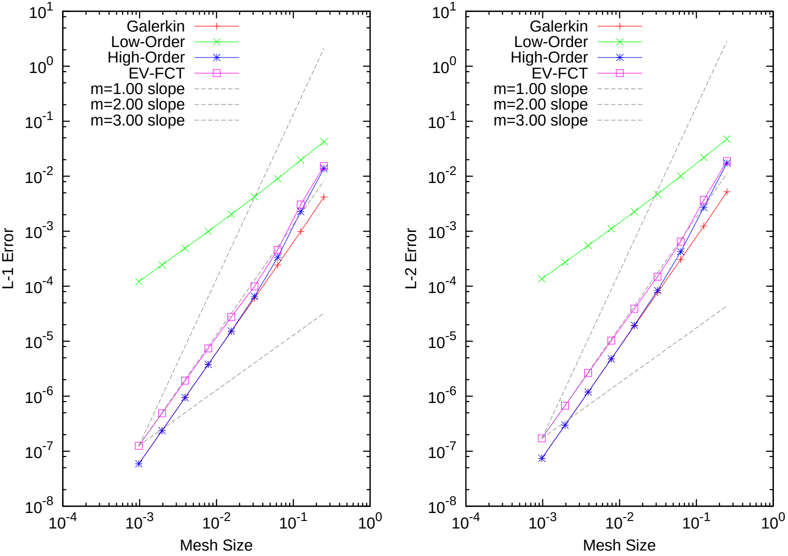
(d) Low-order

(e) EV

(f) EV-FCT

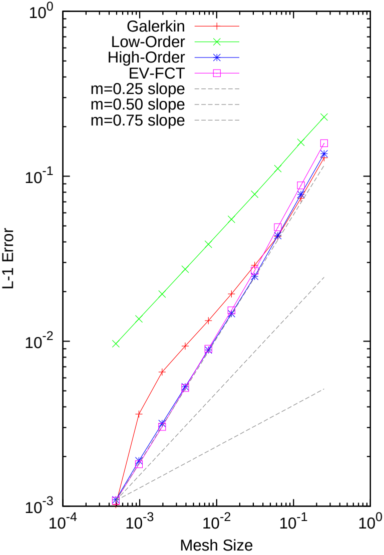
1-D Smooth Problem Convergence Test

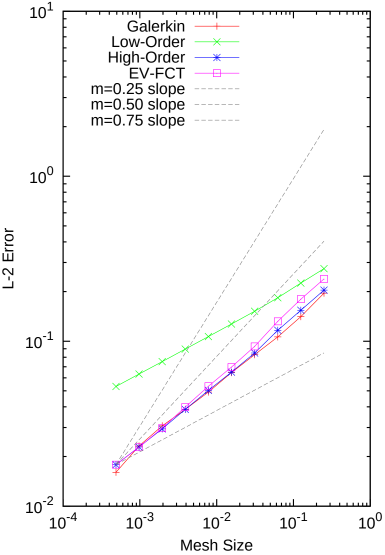
MMS Solution: $u(x,t) = t\sin(\pi x)$, FE



1-D Non-smooth Problem Convergence Test

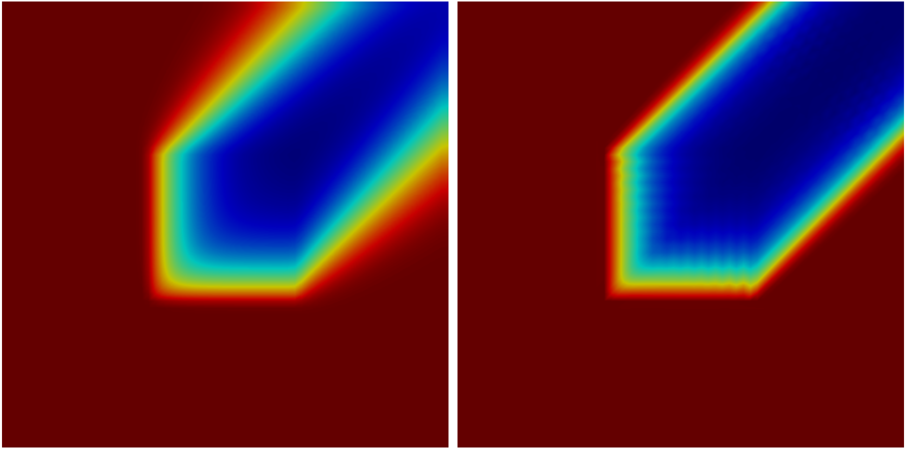
Linear Advection of Discontinuous Wave Front, SSPRK33





Obstruction Test Problem

45°-degree Flux Incident on Square Absorber Region, 4096 cells, FE



(a) Low-order

(b) EV-FCT

Obstruction Test Problem

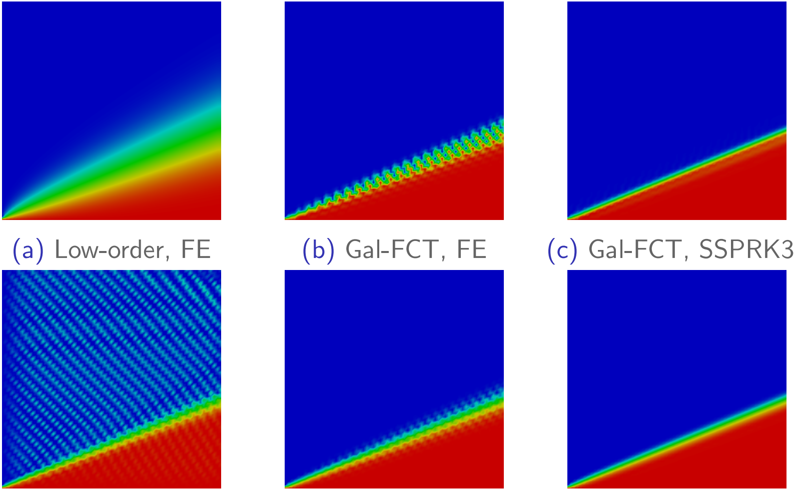
CFL Number vs. Iterations Study, Backward Euler, 256 cells

Table : EV and FCT Iterations Required for EV-FCT Solution

CFL	*EV*		*FCT*	
	Total	*Avg.*	*Total*	*Avg.*
0.1	3999	8.14	3585	7.30
0.5	896	9.05	1499	15.14
1.0	501	10.02	970	19.40
5.0	157	15.70	1130	113.00
10.0	79	15.80	753	150.60
20.0	–	–	**FAIL**	

Glance-in-Void Test Problem

Glancing Incident Flux in a Vacuum, 4096 cells



(a) Low-order, FE

(b) Gal-FCT, FE

(c) Gal-FCT, SSPRK33

(d) EV, FE

(e) EV-FCT, FE

(f) EV-FCT, SSPRK33

Glance-in-Void Test Problem

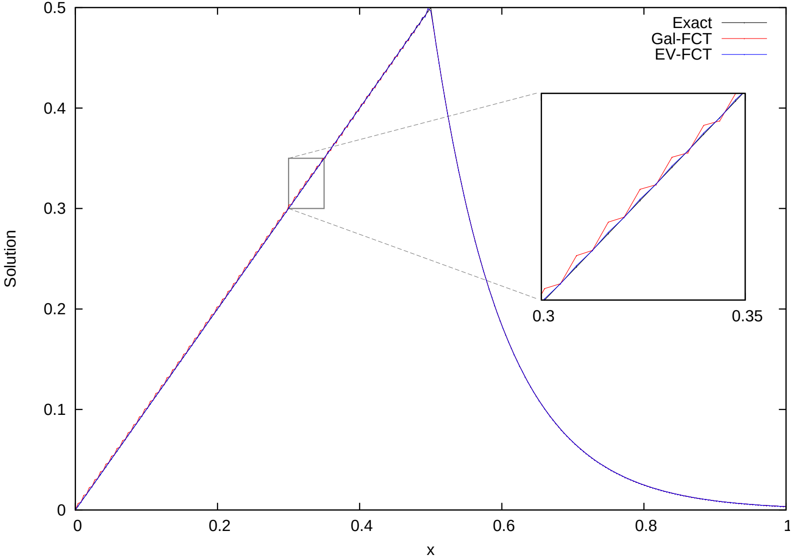
Number of Cells vs. Iterations Study, Steady-State

Table : EV and FCT Iterations Required for EV-FCT Solution

N_{cell}	EV	FCT
64	32	9284
256	59	440
1024	1072	3148
4096	–	**FAIL**

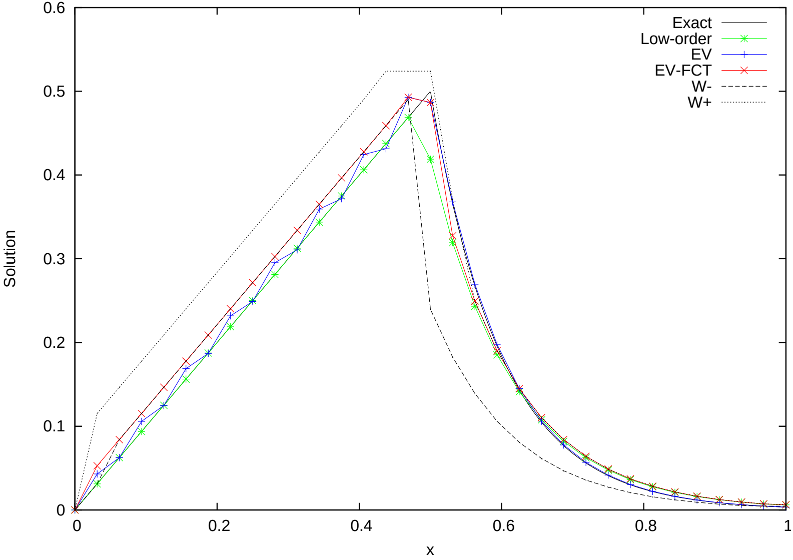
Source-in-Void Test Problem: Galerkin-FCT Vs. EV-FCT

Left Half: Source in Vacuum, Right Half: No Source in Absorber



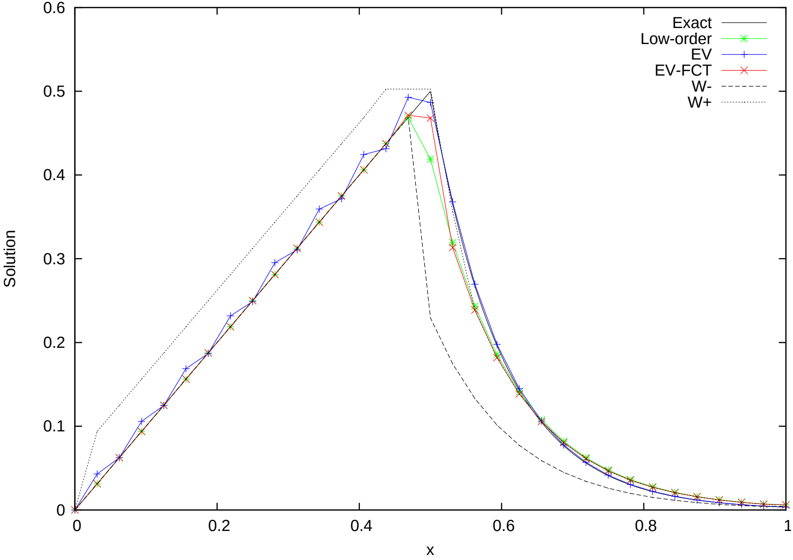
Source-Void-to-Absorber Test Problem

Strongly Imposed Dirichlet BC, $L^+ = L^- = 1$, SS



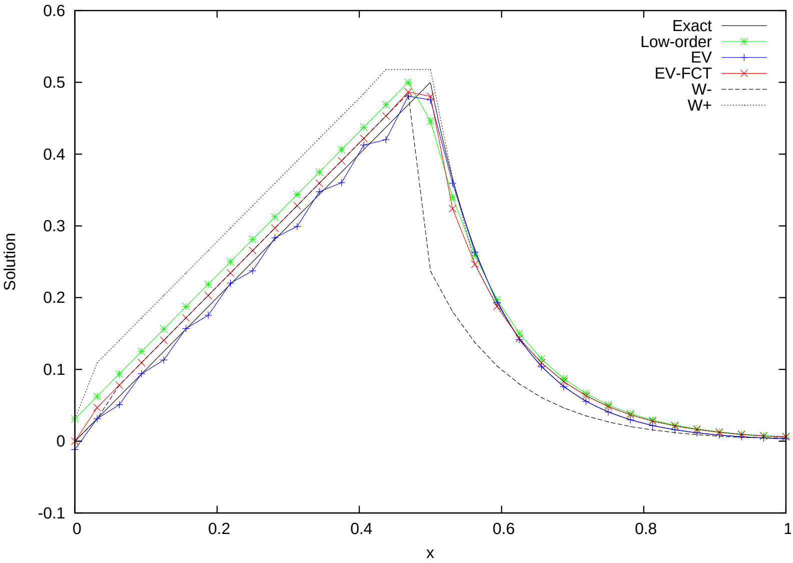
Source-Void-to-Absorber Test Problem

Strongly Imposed Dirichlet BC, $L^+ = L^- = 0$, SS



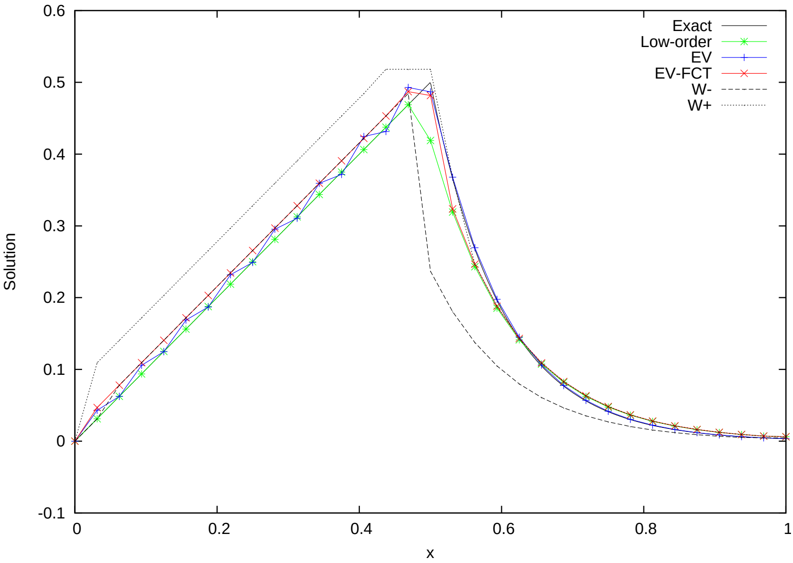
Source-Void-to-Absorber Test Problem

Weakly Imposed Dirichlet BC, SS



Source-Void-to-Absorber Test Problem

Weakly Imposed Dirichlet BC With Boundary Penalty, SS



Source-Void-to-Absorber Test Problem

Number of Cells vs. Iterations Study, Backward Euler, CFL = 1

Table : EV and FCT Iterations Required for EV-FCT Solution

N_{cell}	*EV*		*FCT*	
	Total	*Avg.*	*Total*	*Avg.*
8	661	24.48	244	9.04
16	807	19.21	655	15.60
32	844	11.25	1194	15.92
64	1204	8.72	2024	14.67
128	1752	6.59	3675	13.82
256	2713	5.20	6673	12.78
512	4284	4.14	12098	11.69

Source-Void-to-Absorber Test Problem

CFL Number vs. Iterations Study, Backward Euler, 128 cells

Table : EV and FCT Iterations Required for EV-FCT Solution

		EV		*FCT*		
CFL	N_{step}	*Total*	*Avg.*	*Total*	*Avg.*	L^2 *err.*
0.1	2661	15006	5.64	14036	5.27	3.013×10^{-3}
0.5	533	3445	6.46	5000	9.38	3.033×10^{-3}
1.0	266	1752	6.59	3675	13.82	3.023×10^{-3}
5.0	54	471	8.72	12208	226.07	2.979×10^{-3}
10.0	27	232	8.59	6126	226.89	3.325×10^{-3}
20.0	14	133	9.50	3713	265.21	3.727×10^{-3}
50.0	6	62	10.33	2077	346.17	7.191×10^{-3}

The Shallow Water Equations

Discretization

- A general system of conservation laws, with sources, is

$$\frac{\partial \mathbf{u}}{\partial t} + \nabla \cdot \mathbf{F}(\mathbf{u}) = \mathbf{s}(\mathbf{u})\,. \tag{57}$$

- After doing the following:
  - Substituting the approximate solution $\tilde{\mathbf{u}}\,(\mathbf{x}, t) = \sum_j \mathbf{U}_j(t)\varphi_j(\mathbf{x})$, where $\mathbf{U}_j(t)$ are vector-valued degrees of freedom,
  - Interpolating the flux function at the nodes, $\mathbf{F}(\tilde{\mathbf{u}}) \rightarrow \sum_j \mathrm{F}_j(t)\varphi_j(\mathbf{x})$,
  - Testing with $\varphi_i(\mathbf{x})$, and
  - Evaluating with forward Euler (FE),

  the discrete system becomes

$$\sum_j M^C_{i,j} \frac{\mathbf{U}^{n+1}_j - \mathbf{U}^n_j}{\Delta t} + \sum_j \mathbf{c}_{i,j} \cdot \mathrm{F}^n_j = \mathbf{b}^n_i\,, \tag{58}$$

$$\mathbf{c}_{i,j} \equiv \int_{S_{i,j}} \varphi_i(\mathbf{x}) \nabla \varphi_j(\mathbf{x}) dV\,, \quad \mathbf{b}_i(t) \equiv \int_{S_i} \varphi_i(\mathbf{x}) \mathbf{s}(\tilde{\mathbf{u}}) dV\,. \tag{59}$$

Invariant Domain

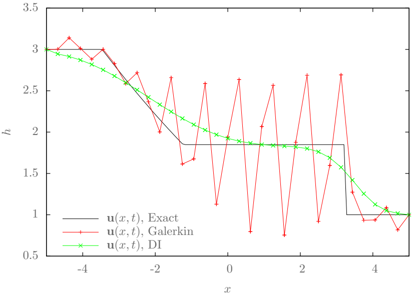
- For the *systems* case, discrete maximum principles no longer apply; the concept of invariant domains becomes the desired tool.
- The objective is to prove that the solution $\tilde{\mathbf{u}}^{n+1} \equiv S(\tilde{\mathbf{u}}^n)$ belongs to an invariant domain with respect to the discrete solution process S.
- First one assumes that the initial data belongs to a convex invariant admissible set: $\mathbf{u}^0 \in A$.
- Next one expresses $S(\tilde{\mathbf{u}}^n)$ as a convex combination of states: $\sum_i a_i \mathbf{b}_i$, and proves the following:
  1. $\sum_i a_i = 1$
  2. $a_i \geq 0 \quad \forall i$
  3. $\mathbf{b}_i \in A \quad \forall i$
- Thus it is proven that $S(\tilde{\mathbf{u}}^n) \in A$ since A is a convex set, and thus A is an invariant domain for S.

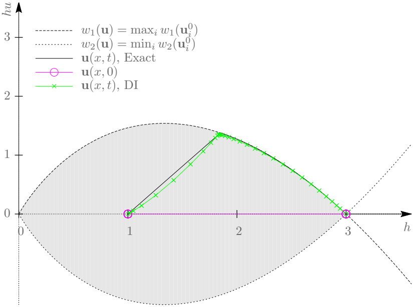
Invariant Domain

- Invariant set definition: A set such that (the entropy solution of the Riemann problem using any 2 elements in the set as left and right states) remains in the set.
- Example: suppose the initial data belongs to the invariant set

$$A_{a,b} \equiv \{(h, hu) \,|\, a \leq w_2(h, hu)\,,\; w_1(h, hu) \leq b\}\,,$$

where $w_1(\mathbf{u})$ and $w_2(\mathbf{u})$ are Riemann invariants.



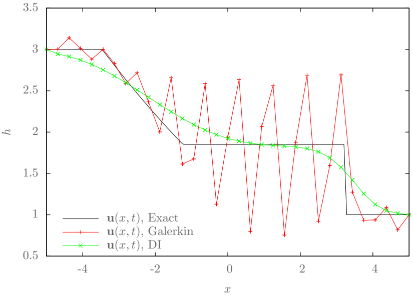


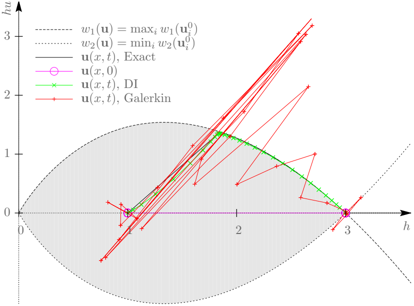
Invariant Domain

- Invariant set definition: A set such that (the entropy solution of the Riemann problem using any 2 elements in the set as left and right states) remains in the set.
- Example: suppose the initial data belongs to the invariant set

$$A_{a,b} \equiv \{(h, hu) \,|\, a \le w_2(h, hu)\,,\; w_1(h, hu) \le b\}\,,$$

where $w_1(\mathbf{u})$ and $w_2(\mathbf{u})$ are Riemann invariants.





Low-Order Scheme

- The domain-invariant low-order scheme lumps the mass matrix and adds a low-order diffusion term:

$$M^L_{i,i}\frac{\mathbf{U}_i^{L,n+1}-\mathbf{U}_i^n}{\Delta t}+\sum_j \mathbf{c}_{i,j}\cdot \mathrm{F}_j^n+\sum_j D_{i,j}^{L,n}\mathbf{U}_j^n=\mathbf{b}_i^n\,, \qquad (60)$$

- The following definition for the low-order diffusion matrix allows the invariant domain property to be proven:

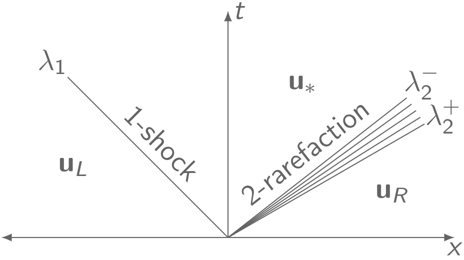
$$D_{i,j}^{L,n}\equiv \max(\lambda_{i,j}^{\max}\left\|\mathbf{c}_{i,j}\right\|_{L^2},\lambda_{j,i}^{\max}\left\|\mathbf{c}_{j,i}\right\|_{L^2})\quad j\neq i\,, \qquad (61)$$

$$D_{i,i}^{L,n}\equiv -\sum_{j\neq i} D_{i,j}^{L,n}\,, \qquad (62)$$

where $\lambda_{i,j}^{\max}\equiv\lambda^{\max}(\mathbf{n}_{i,j},\mathbf{U}_i^n,\mathbf{U}_j^n)$ is the maximum wave speed in the 1-D Riemann problem in the direction $\mathbf{n}_{i,j}\equiv\mathbf{c}_{i,j}/\left\|\mathbf{c}_{i,j}\right\|_{L^2}$ with left state $\mathbf{U}_i^n$ and right state $\mathbf{U}_j^n$.

Maximum Wave Speeds

- Consider an example for the 1-D SWE Riemann problem, where the left wave is a shock, and the right is a rarefaction:



- For a general conservation law system of size m,

$$\lambda^{\max}(\mathbf{u}_L, \mathbf{u}_R) = \max\left(|\lambda_1^-|, |\lambda_m^+|\right) \ . \tag{63}$$

- Then for the example above $\lambda^{\max}(\mathbf{u}_L, \mathbf{u}_R)$ is the maximum of the shock speed $|\lambda_1|$ and rarefaction *head* speed $|\lambda_2^+|$.

Maximum Wave Speeds (Cont.)

- For the SWE, a left shock has the wave speed

$$\lambda_1^-(\mathbf{u}_L, \mathbf{u}_R) = u_L - a_L \left(1 + \left(\frac{(h_* - h_L)(h_* + 2h_L)}{2h_L^2}\right)\right)^{\frac{1}{2}}, \qquad (64)$$

  where $a = \sqrt{gh}$ is the "speed of sound" for the SWE.

- A left rarefaction has the head wave speed

$$\lambda_1^-(\mathbf{u}_L, \mathbf{u}_R) = u_L - a_L \,, \qquad (65)$$

- Note the shock speed depends on the intermediate (star-state) solution for height h_*, which may be computed with the following steps:
  - Apply Rankine-Hugoniot condition to shock(s)
  - Apply Riemann invariant condition to rarefaction(s)
  - Combine resulting expressions and solve nonlinear equation for h_*

High-Order Scheme

- The high-order scheme adds a high-order diffusion term:

$$\sum_j M^C_{i,j}\frac{\mathbf{U}^{H,n+1}_j-\mathbf{U}^n_j}{\Delta t}+\sum_j \mathbf{c}_{i,j}\cdot \mathrm{F}^n_j+\sum_j D^{H,n}_{i,j}\mathbf{U}^n_j=\mathbf{b}^n_i\,, \qquad (66)$$

- The high-order diffusion matrix is proportional to an entropy diffusion matrix and uses the low-order diffusion matrix as an upper bound:

$$D^{H,n}_{i,j}\equiv \min(D^{\eta,n}_{i,j},D^{L,n}_{i,j})\,, \qquad (67)$$

  where the entropy diffusion matrix is proportional to an entropy residual and entropy jump:

$$D^{\eta,n}_{i,j}\equiv\frac{c_{\mathcal{R}}\mathcal{R}^n_{i,j}+c_{\mathcal{J}}\mathcal{J}^n_{i,j}}{\hat{\eta}^n_{i,j}}\,. \qquad (68)$$

Entropy for the Shallow Water Equations

- For the SWE, the entropy is chosen to be the sum of kinetic and potential energy terms:

$$\eta(\mathbf{u}, b) = \frac{1}{2}\frac{\mathbf{q}\cdot\mathbf{q}}{h} + \frac{1}{2}gh\left(h + b\right) . \tag{69}$$

- The entropy flux is

$$\mathbf{f}^{\eta}(\mathbf{u}, b) = g(h + b)\mathbf{q} + \frac{1}{2}\frac{\left(\mathbf{q}\cdot\mathbf{q}\right)\mathbf{q}}{h^2} . \tag{70}$$

- The entropy residual is

$$\mathcal{R}(\mathbf{u}^n, \mathbf{u}^{n-1}) \equiv \frac{\eta(\mathbf{u}^n) - \eta(\mathbf{u}^{n-1})}{\Delta t^{n-1}} + \nabla\cdot\mathbf{f}^{\eta}(\mathbf{u}^n, b) . \tag{71}$$

FCT Antidiffusive Flux Definition

- The antidiffusive fluxes $\mathbf{p}_i$ are defined such that

$$M^L_{i,i}\frac{\mathbf{U}^H_i-\mathbf{U}^n_i}{\Delta t}+\sum_j \mathbf{c}_{i,j}\cdot\mathrm{F}^n_j+\sum_j D^{L,n}_{i,j}\mathbf{U}^n_j=\mathbf{b}^n_i+\mathbf{p}_i\,, \qquad (72)$$

- Subtracting the high-order scheme equation from this gives the definition of $\mathbf{p}$:

$$\mathbf{p}_i\equiv M^L_{i,i}\frac{\mathbf{U}^H_i-\mathbf{U}^n_i}{\Delta t}-\sum_j M^C_{i,j}\frac{\mathbf{U}^H_j-\mathbf{U}^n_j}{\Delta t}+\sum_j(D^{L,n}_{i,j}-D^{H,n}_{i,j})\mathbf{U}^n_j \quad (73)$$

- Decomposing $\mathbf{p}$ into internodal fluxes $\mathbf{P}_{i,j}$ such that $\sum_j \mathbf{P}_{i,j}=\mathbf{p}_i$,

$$\mathbf{P}_{i,j}=-M^C_{i,j}\left(\frac{\mathbf{U}^H_j-\mathbf{U}^n_j}{\Delta t}-\frac{\mathbf{U}^H_i-\mathbf{U}^n_i}{\Delta t}\right)+(D^{L,n}_{i,j}-D^{H,n}_{i,j})(\mathbf{U}^n_j-\mathbf{U}^n_i)\,. \qquad (74)$$

FCT Limitation Process for Systems

- Bounds to impose on FCT solution are unclear for systems case:
  - For scalar case, a DMP was used, but no DMP exists for systems.
  - The invariant domain property is not useful here because while the property is known to hold, the domain itself is not actually known.
- One may want to impose physical bounds on some non-conservative set of variables.
- Limitation of conservative variables may not satisfy these bounds.
- Consider the following sets of variables for the 1-D SWE:
  - Conservative: $\mathbf{u} \equiv [h, hu]^{\mathsf{T}}$
  - Primitive: $\check{\mathbf{u}} \equiv [h, u]^{\mathsf{T}}$
  - Characteristic: $\hat{\mathbf{u}} \equiv [u - 2a, u + 2a]^{\mathsf{T}}$, where $a \equiv \sqrt{gh}$
- Results from literature suggest that limitation on a non-conservative set of variables may produce superior results.

Limitation on Non-Conservative Variables

- Consider some non-conservative set of variables $\hat{\mathbf{u}} = \mathbf{T}^{-1}(\mathbf{u})\mathbf{u}$.
  - For conservative variables, $\mathbf{T}(\mathbf{u}) = \mathbb{I}$.
  - For characteristic variables, $\mathbf{T}(\mathbf{u})$ is the matrix with right eigenvectors of the Jacobian matrix $\partial\mathbf{F}/\partial\mathbf{u}$ as its columns.
- The FCT scheme for limitation of conservative variables is

$$M^L_{i,i}\frac{\mathbf{U}_i^{n+1} - \mathbf{U}_i^n}{\Delta t} + \sum_j \mathbf{c}_{i,j}\cdot\mathrm{F}_j^n + \sum_j D^{L,n}_{i,j}\mathbf{U}_j^n = \mathbf{b}_i^n + \sum_j \mathbf{L}_{i,j}\odot\mathbf{P}_{i,j}\,. \quad (75)$$

- Applying a local transformation $\mathbf{T}^{-1}(\mathbf{U}_i^n)$ gives

$$M^L_{i,i}\frac{\hat{\mathbf{U}}_i^{n+1} - \hat{\mathbf{U}}_i^n}{\Delta t} + \sum_j \mathbf{c}_{i,j}\cdot\hat{\mathrm{F}}_j^n + \sum_j D^{L,n}_{i,j}\hat{\mathbf{U}}_j^n = \hat{\mathbf{b}}_i^n + \sum_j \mathbf{L}_{i,j}\odot\hat{\mathbf{P}}_{i,j}\,, \quad (76)$$

where accents denote transformed quantities, for example,

$$\hat{\mathbf{U}}_j^k = \mathbf{T}^{-1}(\mathbf{U}_i^n)\mathbf{U}_j^k\,, \quad \hat{\mathbf{P}}_{i,j} = \mathbf{T}^{-1}(\mathbf{U}_i^n)\mathbf{P}_{i,j}\,. \quad (77)$$

Limiting Coefficients

- Choose $\mathbf{L}_{i,j}$ to satisfy some solution bounds $\hat{\mathbf{U}}_i^{\pm}$ and then define corresponding antidiffusive flux bounds $\hat{\mathbf{Q}}_i^{\pm}$:

$$\hat{\mathbf{Q}}_i^{-} \leq \sum_j \mathbf{L}_{i,j} \odot \hat{\mathbf{P}}_{i,j} \leq \hat{\mathbf{Q}}_i^{+} \quad \Rightarrow \quad \hat{\mathbf{U}}_i^{-} \leq \hat{\mathbf{U}}_i^{n+1} \leq \hat{\mathbf{U}}_i^{+} \quad \forall i \,. \qquad (78)$$

- Performing some algebra gives the definition

$$\hat{\mathbf{Q}}_i^{\pm} \equiv M_{i,i}^L \frac{\hat{\mathbf{U}}_i^{\pm} - \hat{\mathbf{U}}_i^n}{\Delta t} + \sum_j \mathbf{c}_{i,j} \cdot \hat{\mathrm{F}}_j^n + \sum_j D_{i,j}^{L,n} \hat{\mathbf{U}}_j^n - \hat{\mathbf{b}}_i^n \,. \qquad (79)$$

- As in the scalar case, negative and positive antidiffusive flux sums are required in the limiting coefficient definitions:

$$\hat{p}_i^{p,-} \equiv \sum_{j:\hat{P}_{i,j}^p<0} \hat{P}_{i,j}^p \,, \qquad \hat{p}_i^{p,+} \equiv \sum_{j:\hat{P}_{i,j}^p>0} \hat{P}_{i,j}^p \,. \qquad (80)$$

Limiting Coefficients (Cont.)

- The limiting coefficients are first computed just as in the scalar case:

$$L_i^{p,\pm} \equiv \begin{cases} 1 & \hat{p}_i^{p,\pm} = 0 \\ \min\left(1, \frac{Q_i^{p,\pm}}{\hat{p}_i^{p,\pm}}\right) & \hat{p}_i^{p,\pm} \neq 0 \end{cases} . \tag{81}$$

$$L_{i,j}^p \equiv \begin{cases} \min(L_i^{p,+}, L_j^{p,-}) & \hat{P}_{i,j}^p \geq 0 \\ \min(L_i^{p,-}, L_j^{p,+}) & \hat{P}_{i,j}^p < 0 \end{cases} . \tag{82}$$

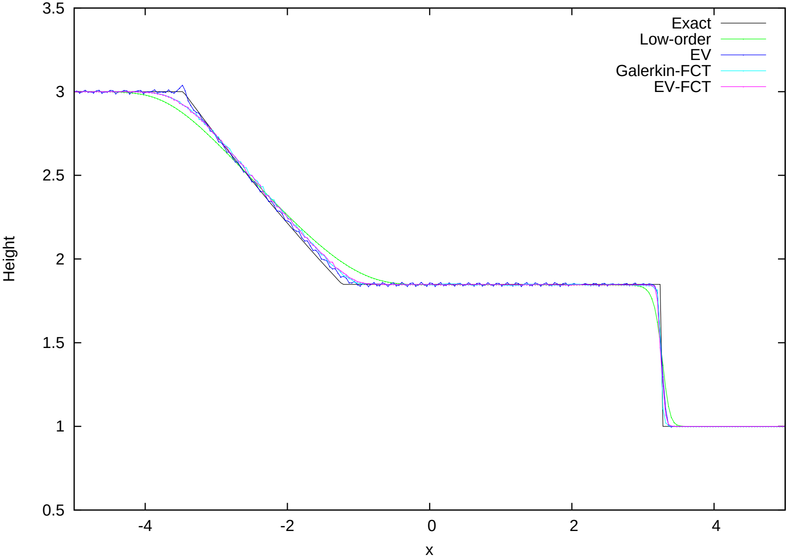
- However, there is a caveat: limiting coefficients may require synchronization, e.g.,

$$L_{i,j}^p \leftarrow \min_k L_{i,j}^k \quad \forall p \, . \tag{83}$$

- Otherwise, antidiffusive fluxes in one component can violate the conditions of another component.

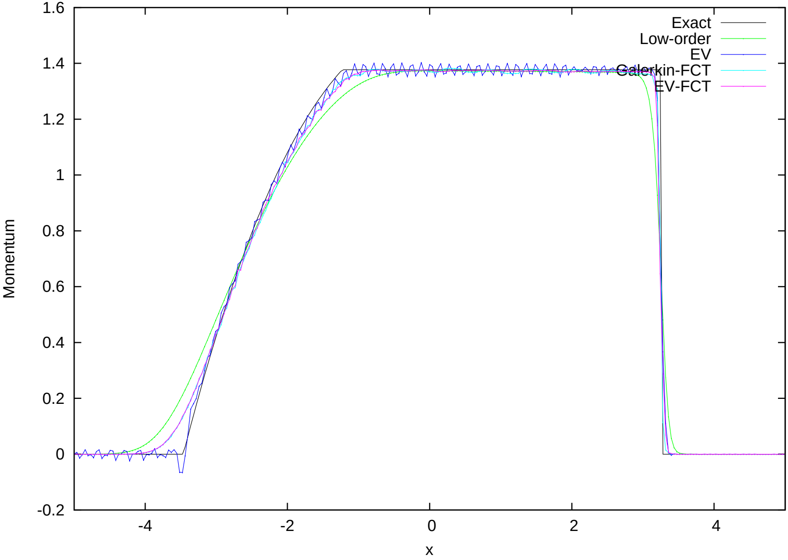
1-D Dam Break Test Problem

Height, 256 cells, FE



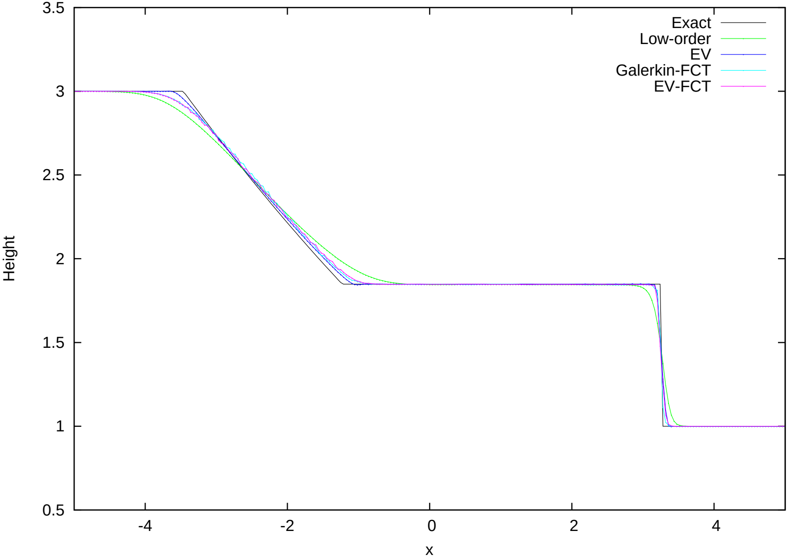
1-D Dam Break Test Problem

Discharge, 256 cells, FE



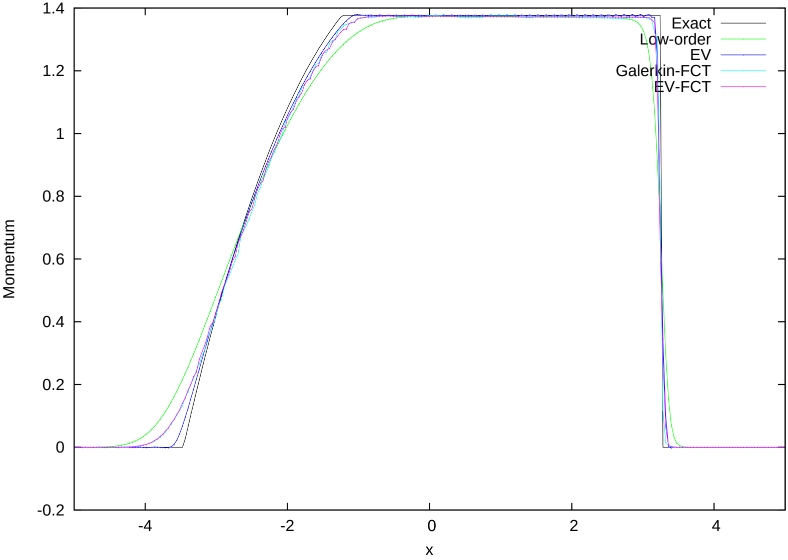
1-D Dam Break Test Problem

Height, 256 cells, SSPRK33



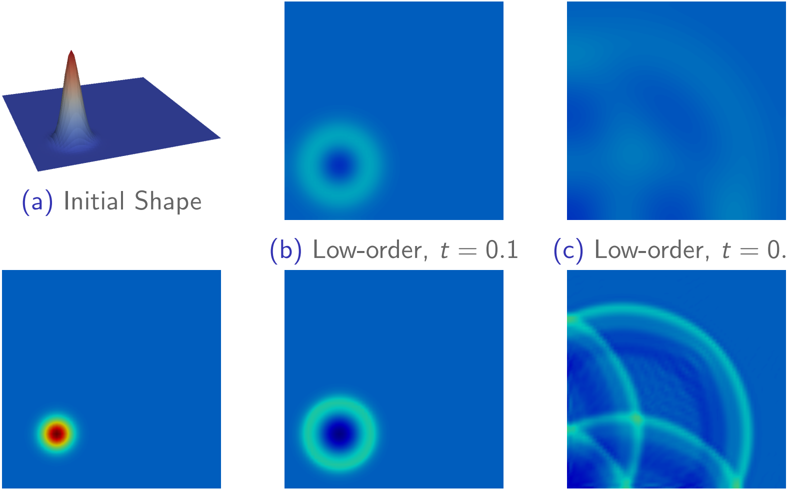
1-D Dam Break Test Problem

Discharge, 256 cells, SSPRK33



Bathtub Test Problem

Height, 4096 cells, FE



(a) Initial Shape

(b) Low-order, $t = 0.1$

(c) Low-order, $t = 0.5$

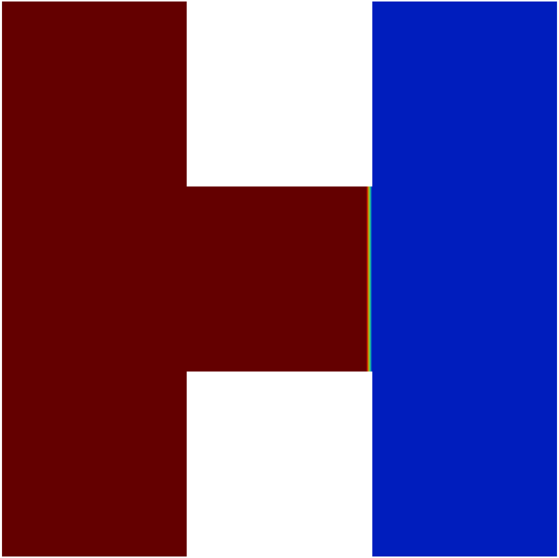
(d) $t = 0$

(e) EV, $t = 0.1$

(f) EV, $t = 0.5$

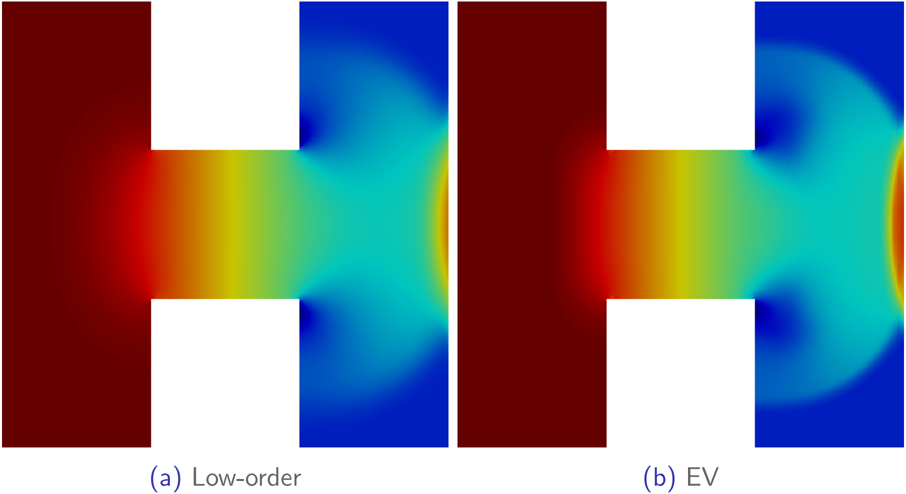
2-D Dam Break Test Problem

Initial Height



2-D Dam Break Test Problem

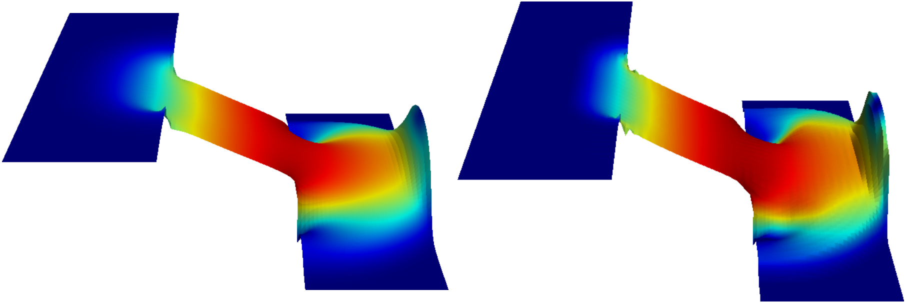
Height, $t = 50$, SSPRK33


(a) Low-order

(b) EV

2-D Dam Break Test Problem

Height Surface, Colored by Discharge, $t = 50$, SSPRK33

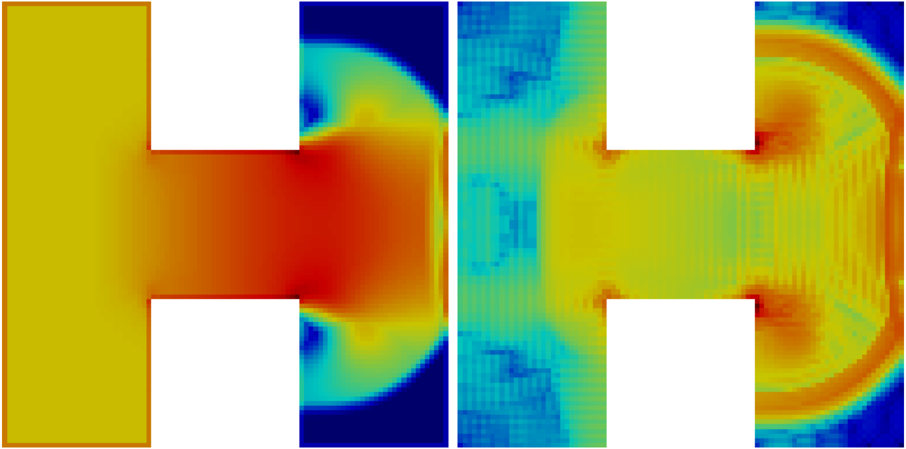


(a) Low-order

(b) EV

2-D Dam Break Test Problem

Viscosity Profiles, $t = 50$, SSPRK33



(a) Low-order

(b) EV

Conclusions

Conclusions

- Both the scalar and systems FCT schemes developed are
  - Non-negativity-preserving
  - Not guaranteed monotone, but work well in practice
  - Entropy-inequality-enforcing
  - 2nd-order-accurate for smooth problems
  - Valid in an arbitrary number of dimensions
  - Valid for unstructured meshes
- The scalar scheme additionally satisfies a discrete maximum principle
- *Stair-stepping*, or *terracing* is observed in some FCT solutions, particularly when using FE
- Severe nonlinear convergence difficulties have been observed for implicit and steady-state FCT
- Characteristic limiting for 1-D SWE was shown to have some success but showed some terracing behavior
- No extension of FCT to 2-D SWE has been developed
- Invariant domain method has not yet been extended to non-flat bottom topography

Acknowledgments

- Dr. Jean Ragusa
- Dr. Jean-Luc Guermond
- Dr. Bojan Popov
- Dr. Marco Delchini
- Dr. Richard Martineau

- This material is based upon work supported under an Integrated University Program Graduate Fellowship.

NEUP Nuclear Energy University Programs

U.S. Department of Energy

Strong-Stability-Preserving Runge-Kutta Methods

- A general s-stage SSPRK method for discretizing a system $\mathbf{M}\frac{d\mathbf{U}}{dt} = \mathbf{r}(\mathbf{U}, t)$ is the following:

$$\mathbf{U}^{n+1} = \hat{\mathbf{U}}^{s}\,, \tag{84}$$

$$\hat{\mathbf{U}}^{i} = \begin{cases} \mathbf{U}^{n} & i = 0 \\ \alpha_i \hat{\mathbf{U}}^{0} + \beta_i \tilde{\mathbf{U}}^{i} & i \neq 0 \end{cases}, \tag{85}$$

$$\mathbf{M}\tilde{\mathbf{U}}^{i} = \mathbf{M}\hat{\mathbf{U}}^{i-1} + \Delta t\, \mathbf{r}(\hat{t}_i, \hat{\mathbf{U}}^{i-1})\,, \tag{86}$$

  where $\hat{t}_i = t^n + c_i \Delta t$,

- An example is the Shu-Osher method (3-stage, 3rd-order-accurate):

$$\alpha = \begin{bmatrix} 0 \\ \frac{3}{4} \\ \frac{1}{3} \end{bmatrix} \qquad \beta = \begin{bmatrix} 1 \\ \frac{1}{4} \\ \frac{2}{3} \end{bmatrix} \qquad c = \begin{bmatrix} 0 \\ 1 \\ \frac{1}{2} \end{bmatrix}. \tag{87}$$

Low-Order Steady-State Matrix Row Sum

- Recall that $A^{L,n}_{i,j} \equiv A^n_{i,j} + D^{L,n}_{i,j}$. Thus, $\sum\limits_j A^{L,n}_{i,j} = \sum\limits_j A^n_{i,j} + \sum\limits_j D^{L,n}_{i,j}$.
- Recall the zero-row-sum property of $\mathbf{D}^{L,n}$:

$$D^{L,n}_{i,i} = -\sum_{j\neq i} D^{L,n}_{i,j} \,, \quad \Rightarrow \quad \sum_j D^{L,n}_{i,j} = 0 \,. \tag{88}$$

- Thus $\sum\limits_j A^{L,n}_{i,j} = \sum\limits_j A^n_{i,j}$:

$$\sum_j A^{L,n}_{i,j} = \sum_j \int\limits_{S_i} \left(\mathbf{f}'(\tilde{u}^n)\cdot\nabla\varphi_j(\mathbf{x}) + \sigma(\mathbf{x})\varphi_j(\mathbf{x})\right)\varphi_i(\mathbf{x})dV \,. \tag{89}$$

- Now, using the fact that $\sum\limits_j \varphi_j(\mathbf{x}) = 1$ (and $\nabla 1 = 0$) gives

$$\sum_j A^{L,n}_{i,j} = \int\limits_{S_i} \sigma(\mathbf{x})\varphi_i(\mathbf{x})dV \,. \tag{90}$$

Limiting Coefficients Definition

- The classic Zalesak limiting strategy starts by separating the negative and positive fluxes:

$$Q_i^- \leq \sum_{j:P_{i,j}<0} L_{i,j} P_{i,j} + \sum_{j:P_{i,j}>0} L_{i,j} P_{i,j} \leq Q_i^+ \,. \tag{91}$$

- Zalesak's limiting coefficients assume that all positive fluxes into a node i have the same limiting coefficient L_i^+ and similarly, negative fluxes have the same limiting coefficient L_i^-:

$$Q_i^- \leq L_i^- p_i^- + L_i^+ p_i^+ \leq Q_i^+ \,. \tag{92}$$

where

$$p_i^- \equiv \sum_{j:P_{i,j}<0} P_{i,j} \,, \qquad p_i^+ \equiv \sum_{j:P_{i,j}>0} P_{i,j} \,. \tag{93}$$

- As a conservative bound for L_i^+, contributions from negative fluxes are ignored (pretending $L_i^- = 0$), giving $L_i^+ \leq \frac{Q_i^+}{p_i^+}$ and similarly for L_i^- and the lower bound.

Limiting Coefficients Definition (Cont.)

- Then, recalling that limiting coefficients are not greater than unity:

$$L_i^{\pm} \equiv \begin{cases} 1 & p_i^{\pm} = 0 \\ \min\left(1, \frac{Q_i^{\pm}}{p_i^{\pm}}\right) & p_i^{\pm} \neq 0 \end{cases} . \tag{94}$$

- However, to limit fluxes conservatively, limited correction fluxes must be equal and opposite:

$$L_{i,j} P_{i,j} = -L_{j,i} P_{j,i} \, . \tag{95}$$

Since $P_{i,j}$ happens to be skew symmetric $(P_{j,i} = -P_{i,j})$ due to the chosen flux decomposition, the limiting coefficients must be symmetric: $L_{j,i} = L_{i,j}$.

Limiting Coefficients Definition (Cont.)

- Thus when deciding the limiting coefficient $L_{i,j}$ for a flux $P_{i,j}$, one must not only consider the bounds for i but also the bounds for j. Specifically, a positive flux $P_{i,j}$ risks violating Q_i^+ and Q_j^-. Putting everything together,

$$L_{i,j} \equiv \begin{cases} \min(L_i^+, L_j^-) & P_{i,j} \geq 0 \\ \min(L_i^-, L_j^+) & P_{i,j} < 0 \end{cases} . \tag{96}$$

Discrete Maximum Principle Bounds for a Theta-Scheme

- Recall the DMP bounds $W_i^{\mathrm{DMP},-} \leq U_i^{L,n+1} \leq W_i^{\mathrm{DMP},+}$ for explicit Euler:

$$W_i^{\pm} \equiv U^n_{\substack{\max\\\min},i} \left(1 - \frac{\Delta t}{M^L_{i,i}} \sum_j A^{L,n}_{i,j}\right) + \frac{\Delta t}{M^L_{i,i}} b_i^n \,. \tag{97}$$

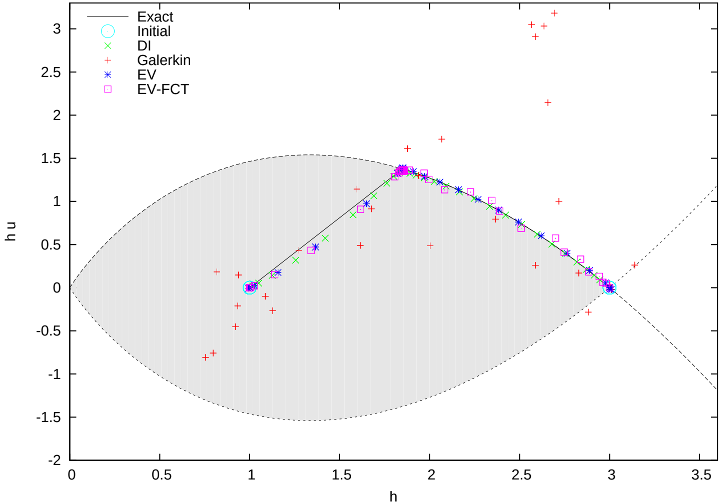
- In contrast, the DMP bounds for the θ scheme are implicit:

$$\begin{aligned} W_i^{\mathrm{DMP},\pm} \equiv \frac{1}{1 + \frac{\theta \Delta t}{M^L_{i,i}} A^{L,n+1}_{i,i}} \Bigg[\left(1 - \frac{(1-\theta)\Delta t}{M^L_{i,i}} \sum_j A^{L,n}_{i,j}\right) U^n_{\substack{\max\\\min},i} \\ - \frac{\theta \Delta t}{M^L_{i,i}} \sum_{j \neq i} A^{L,n+1}_{i,j} U^{L,n+1}_{\substack{\max\\\min},i} + \frac{\Delta t}{M^L_{i,i}} \left((1-\theta) b_i^n + \theta b_i^{n+1}\right) \Bigg] \,, \end{aligned} \tag{98a}$$

$$U^n_{\substack{\max\\\min},i} = \substack{\max\\\min}_{j \in \mathcal{I}(S_i)} U_j^n \,, \quad U^{L,n+1}_{\substack{\max\\\min},i} = \substack{\max\\\min}_{j \in \mathcal{I}(S_i)} U_j^{L,n+1} \,. \tag{98b}$$

1-D Dam Break Test Problem

Phase Diagram, FE, 32 cells


Exact
Initial
DI
Galerkin
EV
EV-FCT
h u
h