Positivity-preserving DG

September 7, 2023

Contents

Introduction

This writeup describes the positivity-preserving Discontinuous Galerkin method for the Euler equations and related systems of conservation and source-balance laws. We first describe the basis and philosophy of the method and its application to the 1D Euler equations without source or diffusion terms.

1 First things first: 1D Euler equations

See [1] for the details of this section. Here we recap the development of the scheme for the 1D Euler equations with no source terms,

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

where

$$\mathbf{w} = \begin{pmatrix} \rho \\ \rho\mathbf{u} \\ E \end{pmatrix}, \quad \mathbf{f}(\mathbf{w}) = \begin{pmatrix} \rho\mathbf{u} \\ \rho u^2 + p \\ (E+p)\mathbf{u} \end{pmatrix}. \tag{2}$$

The total energy E may be partitioned into the kinetic energy and the internal energy, which we denote

$$e = E - \frac{1}{2}\frac{(\rho u)^2}{\rho}. \tag{3}$$

In MHD, the total energy also includes the magnetic energy. We want to stay in the set of admissible states with positive density and internal energy, defined as

$$G = \left\{ \rho > 0, \quad p = (\gamma - 1)\left(E - \frac{1}{2}\frac{(\rho u)^2}{\rho} \right) > 0 \right\}. \tag{4}$$

We here note, without proof, that G is a convex set, which is of supreme importance.

We start with a first-order scheme,

$$\mathbf{w}_j^{n+1} = \mathbf{w}_j^n - \frac{\Delta t}{\Delta x}\left[\mathbf{h}(\mathbf{w}_j^n, \mathbf{w}_{j+1}^n) - \mathbf{h}(\mathbf{w}_{j-1}^n, \mathbf{w}_j^n) \right]. \tag{5}$$

The function $\mathbf{h}$ is a numerical flux such as the Lax-Friedrichs flux, j indexes the spatial cell and n the timestep. We assume that $\mathbf{h}$ has the property that, if $\mathbf{w}_j^n \in G$, then $\mathbf{w}_j^{n+1} \in G$ so long as the standard CFL condition is satisfied. This first-order scheme will be the basis of the positivity-preserving high-order scheme.

Now consider a first-order in time, high-order DG discretization of the Euler equations on cell j using the same numerical flux function $\mathbf{h}$, where the test functions are denoted ψ_α:

$$\int_{x_{j-1/2}}^{x_{j+1/2}} \frac{\mathbf{w}^{n+1} - \mathbf{w}^n}{\Delta t}\psi_\alpha \, dx - \int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{f}(\mathbf{w}^n)\partial_x \psi_\alpha \, dx = -\left[\mathbf{h}(\mathbf{w}^n)\psi_\alpha \right]_{x_{j-1/2}}^{x_{j+1/2}}. \tag{6}$$

To see what happens to the cell averages under the DG discretization, let $\psi_\alpha = 1$. The volume integral vanishes, leaving

$$\Delta x \bar{\mathbf{w}}^{n+1} = \Delta x \bar{\mathbf{w}}^n - \Delta t \left[\mathbf{h}(\mathbf{w}_{j+1/2}^-, \mathbf{w}_{j+1/2}^+) - \mathbf{h}(\mathbf{w}_{j-1/2}^-, \mathbf{w}_{j-1/2}^+) \right]. \tag{7}$$

We now assume the existence of a quadrature rule which integrates $\mathbf{w}$ exactly on the element, and has all positive weights. That is, we assume that there are nodes $\hat{x}_j^\alpha$ and weights $\hat{w}^\alpha$ such that

$$\bar{\mathbf{w}} = \sum_{\alpha=1}^{N} \hat{w}^\alpha \mathbf{w}(\hat{x}_j^\alpha), \tag{8}$$

and $\hat{w}^\alpha > 0$ for all α.

Zhang and Shu show [1] how we may now rewrite (7) as

$$\bar{\mathbf{w}}_j^{n+1} = \sum_{\alpha=2}^{N-1} \hat{w}_\alpha \mathbf{w}(\hat{x}_j^\alpha) + \hat{w}_N \mathbf{H}_N + \hat{w}_1 \mathbf{H}_1, \tag{9}$$

where

$$\mathbf{H}_1 = \mathbf{w}^+_{j-1/2} - \frac{\Delta t}{\hat{w}_1 \Delta x}\left[\mathbf{h}(\mathbf{w}^+_{j-1/2}, \mathbf{w}^-_{j+1/2}) - \mathbf{h}(\mathbf{w}^-_{j-1/2}, \mathbf{w}^+_{j-1/2})\right], \tag{10}$$

$$\mathbf{H}_N = \mathbf{w}^-_{j+1/2} - \frac{\Delta t}{\hat{w}_N \Delta x}\left[\mathbf{h}(\mathbf{w}^-_{j+1/2}, \mathbf{w}^+_{j+1/2}) - \mathbf{h}(\mathbf{w}^+_{j-1/2}, \mathbf{w}^-_{j+1/2})\right]. \tag{11}$$

The trick is that each of $\mathbf{H}_1$ and $\mathbf{H}_N$ has the form (5)!. We now conclude that, if each of $\mathbf{w}^+_{j-1/2}, \mathbf{w}^-_{j+1/2}$, and $\mathbf{w}(\hat{x}^\alpha_j), 2 \le \alpha \le N-1$ are in G, then $\bar{\mathbf{w}}^{n+1}_j \in G$. This follows because we have written $\bar{\mathbf{w}}^{n+1}_j$ as a convex combination of terms in G, where the weights are just $\hat{w}_\alpha$ which are positive.

In other words, a sufficient condition for the next cell average, $\bar{\mathbf{w}}^{n+1}_j$, to be admissible, is that the previous solution be admissible at N nodal points, namely the interior quadrature nodes and the endpoints. It so happens that in 1D, the LGL nodes are precisely what we need: a quadrature rule with all positive weights which sum to 1.

1.1 The positivity-enforcing limiter

Now that we understand the sufficient condition for cell-average positivity, we can directly enforce it with a solution limiter. A solution limiter, as opposed to a flux limiter, simply modifies the values of the solution "in place". We call this limiter the positivity-*enforcing* limiter, since it can only enforce positivity at the nodes. It must be coupled with appropriate numerical fluxes to obtain a scheme that is overall positivity-preserving.

Begin by defining ϵ_ρ and ϵ_p as very small numbers, which ρ and p, respectively, are supposed to remain above. By default we choose $\epsilon_\rho = \epsilon_p = 10^{-12}$, that is, 4 orders of magnitude or so larger than machine precision. Following [3] and [4], the positivity-enforcing limiter is a simple linear scaling which leaves the cell averages of conserved variables unchanged. It works as follows:

1. Enforce positivity of density: replace $\rho_j(x)$, the density polynomial on cell j, by

$$\hat{\rho}_j(x) = \theta_\rho(\rho_j(x) - \bar{\rho}^n_j) + \bar{\rho}^n_j, \tag{12}$$

where

$$\theta_\rho = \min\left(1, \frac{\bar{\rho}^n_j - \epsilon_\rho}{\bar{\rho}^n_j - \rho_{\min}}\right), \quad \rho_{\min} = \min_\alpha \rho_j(\hat{x}^\alpha_j). \tag{13}$$

2. Enforce positivity of pressure. Define the internal energy as

$$e = (\gamma - 1)^{-1} p = E - \frac{1}{2}\frac{\rho u^2}{\hat{\rho}}. \tag{14}$$

Note the use of the scaled $\hat{\rho}$ in the denominator.

Compute the internal energy of the cell-averaged solution. Note that this is distinct from the cell average of the internal energy, since the average of a nonlinear function is not necessarily equal to the same nonlinear function of the average of its arguments.

$$\overline{e}_j^n = \bar{E}_j^n - \frac{1}{2}\frac{(\overline{\rho u}_j^n)^2}{\bar{\rho}_j^n}. \tag{15}$$

The scaling factor is

$$\theta_p = \min\left(1, \frac{\overline{e}_j^n - \epsilon_p}{\overline{e}_j^n - e_{\min}}\right), \tag{16}$$

where

$$e_{\min} = \min_\alpha e_j(\hat{x}_j^\alpha). \tag{17}$$

We now scale all of the components of $\mathbf{w}$ by a factor designed to enforce positivity of the internal energy:

$$\mathbf{w}_j(x) = \theta_p(\mathbf{w}_j(x) - \bar{\mathbf{w}}_j^n) + \bar{\mathbf{w}}_j^n. \tag{18}$$

Note that because it is a simple scaling, the limiter may be applied directly to the nodal representation of the polynomial. The important point is that we evaluate the minimums in (13) and (**??**) at the positively-weighted quadrature nodes, but the scaling itself may be applied to the original representation.

In WARPXM, the cell average may be taken by contracting a variable's nodal values with the entries of the basis array `LINEAR_AVERAGE`, which may be accessed in C++ code as `getBasisArray_LinearAverage()`.

2 Ideal MHD

The limiter for MHD variables is almost identical to the limiter for the Euler equations, except that the internal energy is defined by subtracting off the magnetic energy as well:

$$e = (\gamma - 1)^{-1}p = E - \frac{1}{2}\frac{\rho u^2}{\rho} - \frac{|B|^2}{2}. \tag{19}$$

See [5] for details.

3 Source terms

We now move to the discussion of the compressible Euler equations with source terms:

$$\partial_t \mathbf{w} + \partial_x \mathbf{f}(\mathbf{w}) = \mathbf{s}(\mathbf{w}, x). \tag{20}$$

The scheme satisfied by the cell averages of the DG solution becomes

$$\bar{\mathbf{w}}_j^{n+1} = \bar{\mathbf{w}}_j^n - \frac{\Delta t}{\Delta x}\left[\mathbf{h}(\mathbf{w}_{j+1/2}^-, \mathbf{w}_{j+1/2}^+) - \mathbf{h}(\mathbf{w}_{j-1/2}^-, \mathbf{w}_{j-1/2}^+)\right] + \frac{\Delta t}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} \mathbf{s}(\mathbf{w}_j, x)\, dx. \tag{21}$$

If we approximate the integral of the source term via a Gauss quadrature rule with nodes x_j^β and weights w_β, then we will have

$$\bar{\mathbf{w}}_j^{n+1} = \bar{\mathbf{w}}_j^n - \frac{\Delta t}{\Delta x}\left[\mathbf{h}(\mathbf{w}_{j+1/2}^-, \mathbf{w}_{j+1/2}^+) - \mathbf{h}(\mathbf{w}_{j-1/2}^-, \mathbf{w}_{j-1/2}^+)\right] + \Delta t \sum_\beta w_\beta s(\mathbf{w}_j(x_j^\beta), x_j^\beta). \tag{22}$$

In [2], the authors show how we may rewrite this expression as

$$\bar{\mathbf{w}}_j^{n+1} = \frac{1}{2}\mathbf{H} + \frac{1}{2}\sum_\beta w_\beta \left(\mathbf{w}_j(x_j^\beta) + 2\Delta t \mathbf{s}(\mathbf{w}_j(x_j^\beta), x_j^\beta)\right). \tag{23}$$

The factor of 2 in front of Δt comes from the need to split the contribution of $\bar{\mathbf{w}}_j^n$ over both terms. The term $\mathbf{H}$ contains the contribution from the numerical flux:

$$\mathbf{H} = \bar{\mathbf{w}}_j^n - \frac{2\Delta t}{\Delta x}\left[\mathbf{h}(\mathbf{w}_{j+1/2}^-, \mathbf{w}_{j+1/2}^+) - \mathbf{h}(\mathbf{w}_{j-1/2}^-, \mathbf{w}_{j-1/2}^+)\right], \tag{24}$$

and we will have $\mathbf{H} \in G$ under a CFL condition twice as stringent as the case with no source terms.

Further, assume that we can choose Δt small enough that if $\mathbf{w} \in G$ then $\mathbf{w} + 2\Delta t \mathbf{s}(\mathbf{w}, x) \in G$. Finally, suppose that $\mathbf{w}_j(x_j^\beta) \in G$ for all the Gauss quadrature points β. In that case, all of the terms in (23) are in G, so we have written $\bar{\mathbf{w}}^{n+1}$ as a convex combination of terms in G, showing that it also lies in the admissible set.

Let's recap how this condition differs from the case without source terms. In addition to requiring that $\mathbf{w}_j(\hat{x}_j^\alpha) \in G$, i.e. that the solution at the positivity-preserving quadrature points is positive, we also need $\mathbf{w}_j(x_j^\beta) \in G$ hold at each of the quadrature points where the source term integral is approximated. This is accompanied by a requirement that

$$\mathbf{w}_j(x_j^\beta) + 2\Delta t \mathbf{s}(\mathbf{w}_j(x_j^\beta), x_j^\beta) \in G. \tag{25}$$

In the 1-dimensional case, it might be typical that the points $\hat{x}_j^\alpha$ are the same as the points x_j^β; however, it is not necessary to the numerical method.

3.1 Timestep limit for source terms

Suppose that we want to evaluate a source term

$$\mathbf{s}(\mathbf{w}) = \dot{\mathbf{w}} = \begin{pmatrix} \dot{\rho} \\ \dot{\rho\mathbf{u}} \\ \dot{E} \end{pmatrix}. \tag{26}$$

The positivity-preserving timestep restriction is that

$$\mathbf{w}^n + 2\Delta t \dot{\mathbf{w}} \in G, \tag{27}$$

or in other words that

$$\rho^n + 2\Delta t \dot{\rho} > 0, \quad p(\mathbf{w}^n + 2\Delta t \dot{\mathbf{w}}) > 0. \tag{28}$$

Meeting this constraint will ensure that the source term not cause the density or pressure to go negative at the next timestep. However, to properly resolve the dynamics associated with that source term, a natural cousin of the positivity-preserving restriction on timestep is the following:

$$\rho^n + \Delta t \dot{\rho} \geq \lambda \rho^n, \quad p(\mathbf{w}^n + \Delta t \dot{\mathbf{w}}) \geq \lambda p(\mathbf{w}^n), \quad 0 < \lambda < 1. \tag{29}$$

This says that neither of the two positive quantities will be allowed to drop below a factor λ of their former values. This ensures that, if the source term represents a decay or loss process, that the timestep respects the characteristic time of that process. However, because pressure is a nonlinear function of the conserved quantities, it will be significantly easier to approximate it with a constraint of the form

$$p(\mathbf{w}^n + 2\Delta t \dot{\mathbf{w}}) \geq (1 - 2(1 - \lambda))p(\mathbf{w}^n), \tag{30}$$

which says exactly the same thing but for allowing the pressure to drop by twice as much over a timestep of twice the length.

Now combine the density constraints by multiplying the second by 2 and taking the maximum of the right hand sides:

$$2\rho^n + 2\Delta t \dot{\rho} > \max(2\lambda \rho^n, \rho^n), \tag{31}$$

or

$$\rho^n + \Delta t \dot{\rho} > \mu \rho^n, \quad \mu = \max\left(\lambda, \frac{1}{2}\right). \tag{32}$$

The constraint on pressure, again taking the maximum of the right-hand sides, is

$$p(\mathbf{w}^n + 2\Delta t \dot{\mathbf{w}}) > \eta p(\mathbf{w}^n), \quad \eta = \max(1 - 2(1 - \lambda), 0). \tag{33}$$

The constraint on ρ is equivalent to

$$\Delta t \leq \frac{(\mu - 1)\rho^n}{\dot{\rho}}. \tag{34}$$

To analyze the pressure constraint, multiply both sides of (33) by $(\gamma - 1)\rho^n(\rho^n + 2\Delta t \dot{\rho})$:

$$\rho^n(\rho^n + 2\Delta t \dot{\rho})(E^n + 2\Delta t \dot{E}) - \rho^n \frac{|(\rho \mathbf{u})^n + 2\Delta t \dot{\rho \mathbf{u}}|^2}{2} > \tag{35}$$

$$\eta \rho^n (\rho^n + 2\Delta t \dot{\rho}) E^n - \eta(\rho^n + 2\Delta t \dot{\rho}) \frac{|\rho \mathbf{u}|^2}{2}. \tag{36}$$

Expanding and collecting factors of $2\Delta t$ gives a quadratic inequality:

$$4\Delta t^2\rho^n\left(\dot{\rho}\dot{E}-\frac{|\dot{\rho\mathbf{u}}|^2}{2}\right)+2\Delta t\left(\rho^nE^n\dot{\rho}+\rho^n\rho^n\dot{E}-\rho^n(\rho\mathbf{u})^n(\dot{\rho\mathbf{u}})-\eta\rho^nE^n\dot{\rho}+\eta\dot{\rho}\frac{|(\rho\mathbf{u})^n|^2}{2}\right) \tag{37}$$

$$+\rho^n\rho^nE^n-\rho^n\frac{|(\rho\mathbf{u})^n|^2}{2}-\eta\rho^n\rho^nE^n+\eta\rho^n\frac{|(\rho\mathbf{u})^n|^2}{2}>0. \tag{38}$$

This can be slightly simplified to

$$4\Delta t^2\rho^n\left(\dot{\rho}\dot{E}-\frac{|\dot{\rho\mathbf{u}}|^2}{2}\right)+2\Delta t\left[\rho^n\left((1-\eta)E^n\dot{\rho}-\rho^n\dot{E}-(\rho\mathbf{u})^n(\dot{\rho\mathbf{u}})\right)+\eta\dot{\rho}\frac{|(\rho\mathbf{u})^n|^2}{2}\right] \tag{39}$$

$$+(1-\eta)\rho^n\left(\rho^nE^n-\frac{|(\rho\mathbf{u})^n|^2}{2}\right)>0. \tag{40}$$

To optimize the pressure constraint, it suffices to find the roots of the polynomial in Δt on the left hand side, and choose the largest one which is still bounded by the Δt we found from the ρ requirement, (34).

4 Two dimensions

In this section we describe the modifications to the 1D story required for making positivity-preserving DG work for the Euler equations and for Ideal MHD in 2D. We are primarily interested in triangular elements. The approach described here is an adaptation of Zhang et al.'s approach [6].

The positivity-preserving DG method in two dimensions is nearly identical to that in 1D. As before, we express the update to the cell average as a flux differencing formula:

$$\bar{\mathbf{w}}^{n+1}=\bar{\mathbf{w}}^n-\frac{\Delta t}{\Delta x}\sum_{l=1}^{3}\int_{\partial\Omega_l}\hat{\mathbf{n}}\cdot\mathbf{h}(\mathbf{w}^-,\mathbf{w}^+)\,\mathrm{d}\mathbf{s}. \tag{41}$$

Here, l indexes the faces of the element, while $\mathbf{h}$ is the numerical flux function. Numerically, the face integrals will be approximated by a quadrature rule, with nodes x^β and weights w^β:

$$\int_{\partial\Omega_l}\hat{\mathbf{n}}\cdot\mathbf{h}(\mathbf{w}^-,\mathbf{w}^+)\,\mathrm{d}\mathbf{s}\approx\sum_{\beta=1}^{N_p}w^{\beta,l}\mathbf{h}(\mathbf{w}^-(x^{\beta,l}),\mathbf{w}^+(x^{\beta,l})). \tag{42}$$

As in one dimension, we suppose that we can decompose the cell average $\bar{\mathbf{w}}^n$ as a convex combination of point values, using a quadrature rule with all positive weights:

$$\bar{\mathbf{w}}^n=\sum_{\alpha=1}^{N}\mathbf{w}(\hat{x}^\alpha)\hat{w}^\alpha. \tag{43}$$

As before, it is critical that the set of points $\hat{x}^\alpha$ contain the face quadrature nodes. Denote the weight for the node $\hat{x}^\alpha = x^{\beta,l}$ by $w^\alpha = w^{\alpha|\beta,l}$. That is, this is the weight for the node $x^{\beta,l}$, but in the volume quadrature rule, *not* the face quadrature rule.

Then we can write

$$\bar{\mathbf{w}}^n = \sum_{l=1}^{3} \sum_{\beta=1}^{N} w^{\alpha|\beta,l} \mathbf{w}(x^{\beta,l}) + \sum_{\alpha=1}^{N_{int}} w^\alpha \mathbf{w}(x^\alpha). \tag{44}$$

The idea is to split up the decomposition of the cell average into contributions from the interior of the cell, and contributions from those quadrature nodes which coincide with a face quadrature node. [1]

Zhang et al. then show how we may express the cell average at the next time step as

$$\bar{\mathbf{w}}^{n+1} = \sum_{\alpha=1}^{N_{int}} w^\alpha \mathbf{w}(x^\alpha) + \sum_{\beta=1}^{N_p} w^{\alpha|\beta,l} [H_{1,\beta} + H_{2,\beta} + H_{3,\beta}]. \tag{45}$$

Each of the $H_{l,\beta}$ have the form

$$H_{l,\beta} = \mathbf{w}(x^{\beta,l}) - \Delta t \frac{w^{\beta,l}}{w^{\alpha|\beta,l}} [\Delta F], \tag{46}$$

where ΔF is a difference of numerical fluxes. Once again, we have deconstructed the cell average of the high-order scheme into a convex combination of point values and expressions of the form (5).

Unlike in one dimension, the total set of nodes at which we must enforce positivity is *not* the same as the LGL collocation nodes. We need positivity at all the nodes $\hat{x}^\alpha$, which as we saw must include the face quadrature nodes, but may also include other nodes in the interior of the element. The basis_definitions.pdf writeup contains details of how this quadrature rule is constructed for the triangular basis elements. Inside of WARPXM, we apply the positivity-enforcing limiter at both the face nodes and the extra interior positivity nodes.

4.1 Cylindrical geometry and source terms

In cylindrical geometry, we typically evaluate source terms at Gaussian Quadrature nodes, rather than the LGL nodes, which appear at $r = 0$, resulting in a singular source term evaluation. When this is done, we must be careful to include the Gaussian quadrature nodes in the set of positivity nodes. This is controlled by the `include_gaussian_quad_nodes` flag in warpy.

[1] Note that if a node is double counted at a vertex, we can pretend that it is two separate nodes which coincide in space, each with half the weight.

References

[1] Xiangxiong Zhang, Chi-Wang Shu. "On positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes". Journal of Computational Physics, 2010.

[2] Xiangxiong Zhang, Chi-Wang Shu. "Positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations with source terms". Journal of Computational Physics, 2011.

[3] Chen Wang, Xiangxiong Zhang, Chi-Wang Shu, Jianguo Ning. "Robust high order discontinuous Galerkin schemes for two-dimensional gaseous detonations". Journal of Computational Physics, 2012.

[4] Xiangxiong Zhang. "On positivity-preserving high order discontinuous Galerkin schemes for compressible Navier–Stokes equations". Journal of Computational Physics, 2017.

[5] Yue Cheng, Fengyan Li, Jianxian Qiu, Liwei Xu. "Positivity-preserving DG and central DG methods for ideal MHD equations". Journal of Computational Pyhysics, 2013.

[6] Xiangxiong Zhang, Yinhua Xia, Chi-Wang Shu. "Maximum-Principle-Satisfying and Positivity-Preserving High Order Discontinuous Galerkin Schemes for Conservation Laws on Triangular Meshes". Journal of Scientific Computation, (2012).