Nodal Discontinuous Galerkin Method for the Euler Equations in GR

Samuel J Dunham[1], Eirik Endeve[2], et al.

Contents

[2] Department of Astronomy, Vanderbilt University, 6301 Stevenson Science Center, Nashville TN, 37212, USA; samuel.j.dunham@vanderbilt.edu

[2] Computational and Applied Mathematics Group, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6354, USA; endevee@ornl.gov

1. Discontinuous Galerkin Scheme

We assume a spacetime metric

$$ds^2 = -\alpha^2\,dt^2 + \gamma_{ij}\,dx^i\,dx^j, \tag{1}$$

and consider the system of conservation laws with sources

$$\partial_t\big(\sqrt{\gamma}\,\boldsymbol{U}\big) + \sum_{i=1}^{d}\partial_i\big(\alpha\,\sqrt{\gamma}\,\boldsymbol{F}^i(\boldsymbol{U})\big) = \alpha\,\sqrt{\gamma}\,\boldsymbol{G}(\boldsymbol{U}), \tag{2}$$

where

$$\boldsymbol{U} = \big(D,\,S_j,\,\tau\big)^{\mathrm{T}} = \big(\rho\,W,\,\rho\,h\,W^2\,v_j,\,\rho\,W\,(h\,W-1)-p\big)^{\mathrm{T}}, \tag{3}$$

$$\boldsymbol{F}^i(\boldsymbol{U}) = \big(D\,v^i,\,\big)^{\mathrm{T}} \tag{4}$$

2. Bound-Preserving Methods Using First-Order DG Scheme

2.1. Cartesian Coordinates

This section closely follows Qin et al. (2016).

2.1.1. Set of Admissible States

We consider a one-dimensional system of conservation laws:

$$\partial_t\,\boldsymbol{U} + \partial_x\,\boldsymbol{F}\left(\boldsymbol{U}\right) = \boldsymbol{0}, \tag{5}$$

where $\boldsymbol{U}$ is a vector of conserved variables, defined as:

$$\boldsymbol{U} \longrightarrow \begin{pmatrix} D \\ S \\ \tau \end{pmatrix} = \begin{pmatrix} \rho\,W \\ \rho\,h\,W^2\,v \\ \rho\,W\,(h\,W-1)-p \end{pmatrix}, \tag{6}$$

and $\boldsymbol{F}\left(\boldsymbol{U}\right)$ are the fluxes of those conserved quantities:

$$\boldsymbol{F}\left(\boldsymbol{U}\right) \longrightarrow \begin{pmatrix} \rho\,W\,v \\ \rho\,h\,W^2\,v^2+p \\ \rho\,h\,W^2\,v - D\,v \end{pmatrix}. \tag{7}$$

The physics leads us to define a set of admissible states, $\mathcal{G}_p$ (the subscript p stands for primitive), as:

$$\mathcal{G}_p \equiv \left\{\boldsymbol{U}\middle|\rho>0,\,p>0,\,v^2<1\right\}. \tag{8}$$

It is shown in Mignone & Bodo (2005) that $\mathcal{G}$ is a convex set[3] and can equivalently be written in terms of the conserved variables as:

$$\mathcal{G} \equiv \left\{\boldsymbol{U}\middle|D>0,\,\tau+D>\sqrt{D^2+S^2}\right\}. \tag{9}$$

[3] Convex in the sense that if $\boldsymbol{U}_1 \in \mathcal{G}$ and $\boldsymbol{U}_2 \in \mathcal{G}$, then $\alpha_1\,\boldsymbol{U}_1 + \alpha_2\,\boldsymbol{U}_2 \in \mathcal{G}$, where $\alpha_1,\,\alpha_2 \in [0,1]$ and $\alpha_1+\alpha_2=1$.

2.1.2. Time-Step Derivation/CFL Condition

For the first-order DG method using forward-Euler time-stepping, we evolve the vector of conserved variables as:

$$\overline{\boldsymbol{U}}_i^{n+1} = \overline{\boldsymbol{U}}_i^n - \eta_i \left[\hat{\boldsymbol{F}} \left(\overline{\boldsymbol{U}}_i^n, \overline{\boldsymbol{U}}_{i+1}^n \right) - \hat{\boldsymbol{F}} \left(\overline{\boldsymbol{U}}_{i-1}^n, \overline{\boldsymbol{U}}_i^n \right) \right], \tag{10}$$

where

$$\overline{\boldsymbol{U}}_i \equiv \frac{1}{\Delta x_i} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \boldsymbol{U}_i \, dx, \tag{11}$$

$\eta_i \equiv \Delta t_i / \Delta x_i$, and $\hat{\boldsymbol{F}}$ is the numerical flux. In this document we use the local Lax-Friedrichs flux, defined as:

$$\hat{\boldsymbol{F}} \left(a, b \right) = \frac{1}{2} \left[\boldsymbol{F} \left(a \right) + \boldsymbol{F} \left(b \right) - \alpha_{ab} \left(b - a \right) \right], \tag{12}$$

where a and b represent the state of the fluid in two different elements, α_{ab} is an estimate for the wave-speed:

$$\alpha_{ab} = \max \left[\alpha \left(a \right), \alpha \left(b \right) \right], \tag{13}$$

and α is the largest (in absolute value) eigenvalue of the flux-Jacobian:

$$\alpha = \left| \left| \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{U}} \right| \right|. \tag{14}$$

Using this we define the following variables:

$$\alpha_{i+\frac{1}{2}} = \max \left[\alpha \left(\overline{\boldsymbol{U}}_i \right), \alpha \left(\overline{\boldsymbol{U}}_{i+1} \right) \right], \qquad \alpha_{i-\frac{1}{2}} = \max \left[\alpha \left(\overline{\boldsymbol{U}}_{i-1} \right), \alpha \left(\overline{\boldsymbol{U}}_i \right) \right]. \tag{15}$$

Substituting (12) with (15) into (10):

$$\begin{aligned} \overline{\boldsymbol{U}}_i^{n+1} &= \overline{\boldsymbol{U}}_i^n - \frac{\eta_i}{2} \left[\boldsymbol{F} \left(\overline{\boldsymbol{U}}_i^n \right) + \boldsymbol{F} \left(\overline{\boldsymbol{U}}_{i+1}^n \right) - \alpha_{i+\frac{1}{2}} \left(\overline{\boldsymbol{U}}_{i+1}^n - \overline{\boldsymbol{U}}_i^n \right) \right. \\ &\qquad \left. - \boldsymbol{F} \left(\overline{\boldsymbol{U}}_i^n \right) - \boldsymbol{F} \left(\overline{\boldsymbol{U}}_{i-1}^n \right) + \alpha_{i-\frac{1}{2}} \left(\overline{\boldsymbol{U}}_i^n - \overline{\boldsymbol{U}}_{i-1}^n \right) \right] \\ &= \left[1 - \frac{\eta_i}{2} \left(\alpha_{i+\frac{1}{2}} + \alpha_{i-\frac{1}{2}} \right) \right] \overline{\boldsymbol{U}}_i^n + \frac{\eta_i}{2} \alpha_{i+\frac{1}{2}} \left[\overline{\boldsymbol{U}}_{i+1}^n - \frac{1}{\alpha_{i+\frac{1}{2}}} \boldsymbol{F} \left(\overline{\boldsymbol{U}}_{i+1}^n \right) \right] \\ &\qquad + \frac{\eta_i}{2} \alpha_{i-\frac{1}{2}} \left[\overline{\boldsymbol{U}}_{i-1}^n + \frac{1}{\alpha_{i-\frac{1}{2}}} \boldsymbol{F} \left(\overline{\boldsymbol{U}}_{i-1}^n \right) \right] \\ &= \left[1 - \frac{\eta_i}{2} \left(\alpha_{i+\frac{1}{2}} + \alpha_{i-\frac{1}{2}} \right) \right] \overline{\boldsymbol{U}}_i^n + \frac{\eta_i}{2} \alpha_{i+\frac{1}{2}} \overline{\boldsymbol{H}}^- \left(\overline{\boldsymbol{U}}_{i+1}^n, \alpha_{i+\frac{1}{2}} \right) + \frac{\eta_i}{2} \alpha_{i+\frac{1}{2}} \overline{\boldsymbol{H}}^+ \left(\overline{\boldsymbol{U}}_{i-1}^n, \alpha_{i-\frac{1}{2}} \right), \end{aligned} \tag{16}$$

where

$$\overline{\boldsymbol{H}}^{\pm} \left(\overline{\boldsymbol{U}}, \alpha \right) \equiv \overline{\boldsymbol{U}} \pm \frac{1}{\alpha} \boldsymbol{F} \left(\overline{\boldsymbol{U}} \right). \tag{17}$$

The proof that $\overline{\boldsymbol{H}}^{\pm} \in \mathcal{G}$ is given in Qin et al. (2016). Therefore, we see that with a restriction on $\alpha_{i\pm\frac{1}{2}}$ that (16) is a convex combination. The restriction is (recalling that $\eta_i = \Delta t_i / \Delta x_i$):

$$1 - \frac{\eta_i}{2} \left(\alpha_{i+\frac{1}{2}} + \alpha_{i-\frac{1}{2}} \right) > 0 \implies \frac{\eta_i}{2} \left(\alpha_{i+\frac{1}{2}} + \alpha_{i-\frac{1}{2}} \right) < 1 \implies \Delta t_i < \frac{2 \, \Delta x_i}{\alpha_{i+\frac{1}{2}} + \alpha_{i-\frac{1}{2}}} \leq \frac{\Delta x_i}{\max \left(\alpha_{i\pm\frac{1}{2}} \right)}. \tag{18}$$

We want a time-step that is the same for all elements at a given time, so we tighten the restriction to:

$$\Delta t < \text{min}_i \left(\frac{\Delta x_i}{\text{max}\left(\alpha_{i\pm\frac{1}{2}}\right)} \right) = \frac{\Delta x}{\text{max}_i \left(\alpha_{i\pm\frac{1}{2}}\right)}, \tag{19}$$

where the equality follows for a uniform mesh, i.e. $\Delta x_i = \Delta x \, \forall i$.

2.2. Curvilinear Coordinates in 1-D

NOTE: We assume a conformally-flat, time-independent spatial three-metric:

$$\gamma_{ij}\left(x^k,t\right) \longrightarrow \psi^4\left(x^k\right)\,\overline{\gamma_{ii}}\left(x^k\right), \tag{20}$$

where $\psi\left(x^k\right)$ is the conformal factor and $\overline{\gamma}_{ii}$ is the flat-space metric.

2.2.1. Set of Admissible States

We again consider a one-dimensional system of conservation laws, but this time with a curvilinear metric:

$$\partial_t(\sqrt{\gamma}\,\boldsymbol{U}) + \partial_i\left(\sqrt{\gamma}\,\boldsymbol{F}^i\right) = \sqrt{\gamma}\,\boldsymbol{Q}, \quad \text{(no sum on } i\text{)}, \tag{21}$$

where $\boldsymbol{U}$ is given by:

$$\boldsymbol{U} \longrightarrow \begin{pmatrix} D \\ S_j \\ \tau \end{pmatrix} = \begin{pmatrix} \rho\,W \\ \rho\,h\,W^2\,v_j \\ \rho\,W\,(h\,W-1)-p \end{pmatrix} = \begin{pmatrix} \rho\,W \\ \rho\,h\,W^2\,\gamma_{jk}\,v^k \\ \rho\,W\,(h\,W-1)-p \end{pmatrix}, \tag{22}$$

$\boldsymbol{F}^i\left(\boldsymbol{U}\right)$ are the fluxes in the x^i-direction of those conserved quantities:

$$\boldsymbol{F}^i\left(\boldsymbol{U}\right) \longrightarrow \begin{pmatrix} D\,v^i \\ S^i\,v_j + p\,\delta^i{}_j \\ S^i - D\,v^i \end{pmatrix} = \begin{pmatrix} \rho\,W\,v^i \\ \rho\,h\,W^2\,v^i\,v_j + p\,\delta^i{}_j \\ \rho\,h\,W^2\,v^i - D\,v^i \end{pmatrix} = \begin{pmatrix} \rho\,W\,v^i \\ \rho\,h\,W^2\,\gamma_{jk}\,v^i\,v^k + p\,\delta^i{}_j \\ \rho\,h\,W^2\,v^i - D\,v^i \end{pmatrix}, \tag{23}$$

and $\boldsymbol{Q}$ is a source term:

$$\boldsymbol{Q} \longrightarrow \begin{pmatrix} 0 \\ \frac{1}{2}\,P^{ik}\,\partial_j\,\gamma_{ik} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{2}\left[P^{11}\,\partial_j\,\gamma_{11} + P^{22}\,\partial_j\,\gamma_{22} + P^{33}\,\partial_j\,\gamma_{33}\right] \\ 0 \end{pmatrix} \tag{24}$$

$$= \begin{pmatrix} 0 \\ P^{11}\,h_1\,\partial_j\,h_1 + P^{22}\,h_2\,\partial_j\,h_2 + P^{33}\,h_3\,\partial_j\,h_3 \\ 0 \end{pmatrix}, \tag{25}$$

where we have used the fact that $\gamma_{kk} = \left(h_k\right)^2$. The P^{ik} are components of the pressure tensor:

$$P^{ik} = S^i\,v^k + p\,\gamma^{ik} = \gamma^{i\ell}\,S_\ell\,v^k + p\,\gamma^{ik} = \gamma^{i\ell}\,S_\ell\,v^k + p\,\gamma^{i\ell}\,\delta^k_{\ \ell} = \gamma^{i\ell}\left(S_\ell\,v^k + p\,\delta^k_{\ \ell}\right). \tag{26}$$

Since the spatial three-metric is diagonal the only non-zero term is that with $\ell = i$. We can therefore simplify further:

$$P^{ik} = \gamma^{ii}\left(S_i\,v^k + p\,\delta^k_{\ i}\right) = \frac{1}{\gamma_{ii}}\,\left(S_i\,v^k + p\,\delta^k_{\ i}\right), \quad \text{(no sum on } i\text{)}. \tag{27}$$

For the source-term sum we then have:

$$Q_j = \frac{1}{2}\,P^{ik}\,\partial_j\,\gamma_{ik} = \frac{1}{2}\,P^{kk}\,\partial_j\,\gamma_{kk} = \frac{1}{2}\,P^{kk}\,\partial_j\left(h_k\right)^2 = P^{kk}\,h_k\,\partial_j\,h_k. \tag{28}$$

These definitions lead us to define the same set of admissible states as before, namely:

$$\mathcal{G}_p \equiv \left\{\boldsymbol{U}\middle|\rho > 0,\, p > 0,\, v^2 < 1\right\}, \tag{29}$$

the only difference being that v^2 now involves the metric:

$$v^2 = v^j \, v_j = \gamma_{kj} \, v^k \, v^j \, . \tag{30}$$

Before continuing, we show that the introduction of the metric doesn't affect the translation between $\mathcal{G}_p$ and $\mathcal{G}$...(SD: I've shown this, just need to TeX it up)

2.2.2. Time-Step Derivation/CFL Condition

We start by integrating both sides of (21) over dx^i and dividing by the volume of the K^{th} element, ΔV_K (recalling that there is no sum on i):

$$\frac{1}{\Delta V_K} \int_{x^i_L}^{x^i_H} \partial_t (\sqrt{\gamma} \boldsymbol{U}) dx^i + \frac{1}{\Delta V_K} \int_{x^i_L}^{x^i_H} \partial_i \big(\sqrt{\gamma} \boldsymbol{F}^i \left(\boldsymbol{U}\right)\big) dx^i = \frac{1}{\Delta V_K} \int_{x^i_L}^{x^i_H} \sqrt{\gamma} \, \boldsymbol{Q} \, dx^i , \tag{31}$$

where:

$$\Delta V_K = \int_{x^i_L}^{x^i_H} dV = \int_{x^i_L}^{x^i_H} \sqrt{\gamma} \, dx^i . \tag{32}$$

By defining the cell-average as:

$$\boldsymbol{W}_K \equiv \frac{1}{\Delta V_K} \int_{x^i_L}^{x^i_H} \boldsymbol{W} \, dV , \tag{33}$$

we have:

$$\frac{d \, \boldsymbol{U}_K}{dt} + \frac{1}{\Delta V_K} \left(\sqrt{\gamma} \, \hat{\boldsymbol{F}} \left(\boldsymbol{U}\right) \right) \Big|_{x^i_L}^{x^i_H} = \boldsymbol{Q}_K . \tag{34}$$

Now, using the common notation of the time step being represented as a superscript n:

$$\boldsymbol{U}^{n+1}_K = \boldsymbol{U}^n_K - \frac{\Delta t^n_K}{\Delta V_K} \left[\sqrt{\gamma}_H \, \hat{\boldsymbol{F}}^n_H - \sqrt{\gamma}_L \, \hat{\boldsymbol{F}}^n_L \right] + \Delta t^n_K \, \boldsymbol{Q}^n_K . \tag{35}$$

Now we define a parameter a la Zhang & Shu (2011): $\varepsilon \in (0,1)$, such that (NOTE: Zhang & Shu (2011) set $\varepsilon = 1/2$):

$$\boldsymbol{U}^n_K = \varepsilon \, \boldsymbol{U}^n_K + (1 - \varepsilon) \, \boldsymbol{U}^n_K . \tag{36}$$

We can use the first term to balance out the term in the square brackets and the second term to balance out the source term.

So, we get:

$$\begin{aligned}
\boldsymbol{U}^{n+1}_K &= \varepsilon \left\{ \boldsymbol{U}^n_K - \frac{\Delta t^n_K}{\varepsilon \, \Delta V_K} \left[\sqrt{\gamma}_H \, \hat{\boldsymbol{F}}^n_H - \sqrt{\gamma}_L \, \hat{\boldsymbol{F}}^n_L \right] \right\} + (1 - \varepsilon) \, \boldsymbol{U}^n_K + \Delta t^n_K \, \boldsymbol{Q}^n_K & (37) \\
&= \varepsilon \left\{ \boldsymbol{U}^n_K - \eta^n_K \left[\sqrt{\gamma}_H \, \hat{\boldsymbol{F}} \left(\boldsymbol{U}^n_{K+1}, \boldsymbol{U}^n_K \right) - \sqrt{\gamma}_L \, \hat{\boldsymbol{F}} \left(\boldsymbol{U}^n_K, \boldsymbol{U}^n_{K-1} \right) \right] \right\} + (1 - \varepsilon) \, \boldsymbol{U}^n_K + \Delta t^n_K \, \boldsymbol{Q}^n_K & (38) \\
&= \varepsilon \, \boldsymbol{H}_{K,1} + (1 - \varepsilon) \, \boldsymbol{H}_{K,2} , & (39)
\end{aligned}$$

where

$$\boldsymbol{H}_{K,1} \equiv \boldsymbol{U}^n_K - \eta^n_K \left[\sqrt{\gamma}_H \, \hat{\boldsymbol{F}} \left(\boldsymbol{U}^n_{K+1}, \boldsymbol{U}^n_K \right) - \sqrt{\gamma}_L \, \hat{\boldsymbol{F}} \left(\boldsymbol{U}^n_K, \boldsymbol{U}^n_{K-1} \right) \right] , \tag{40}$$

$$\boldsymbol{H}_{K,2} \equiv \boldsymbol{U}_K^n + \frac{\Delta t_K^n}{1-\varepsilon}\,\boldsymbol{Q}_K^n, \tag{41}$$

and

$$\eta_K^n \equiv \frac{\Delta t_K^n}{\varepsilon\,\Delta V_K}. \tag{42}$$

We proceed by focusing on each term individually, starting with the numerical flux term, $\boldsymbol{H}_{K,1}$.

2.2.3. *Numerical flux term*

We have to show that $\boldsymbol{H}_{K,1} \in \mathcal{G}$. We again we use the Local-Lax-Friedrichs flux, (12), yielding for $\boldsymbol{H}_{K,1}$:

$$\boldsymbol{U}_K^n - \frac{\eta_K^n}{2}\Big\{\sqrt{\gamma}_H\left[\boldsymbol{F}\left(\boldsymbol{U}_{K+1}^n\right) + \boldsymbol{F}\left(\boldsymbol{U}_K^n\right) - \alpha_H^n\left(\boldsymbol{U}_{K+1}^n - \boldsymbol{U}_K^n\right)\right] \tag{43}$$

$$- \sqrt{\gamma}_L\left[\boldsymbol{F}\left(\boldsymbol{U}_K^n\right) + \boldsymbol{F}\left(\boldsymbol{U}_{K-1}^n\right) - \alpha_L^n\left(\boldsymbol{U}_K^n - \boldsymbol{U}_{K-1}^n\right)\right]\Big\} \tag{44}$$

$$= \left(1 - \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_H\,\alpha_H^n - \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_L\,\alpha_L^n\right)\boldsymbol{U}_K^n \tag{45}$$

$$- \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_H\,\boldsymbol{F}\left(\boldsymbol{U}_K^n\right) + \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_L\,\boldsymbol{F}\left(\boldsymbol{U}_K^n\right) \tag{46}$$

$$+ \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_L\,\alpha_L^n\left[\boldsymbol{U}_{K-1}^n + \frac{1}{\alpha_L^n}\boldsymbol{F}\left(\boldsymbol{U}_{K-1}^n\right)\right] + \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_H\,\alpha_H^n\left[\boldsymbol{U}_{K+1}^n - \frac{1}{\alpha_H^n}\boldsymbol{F}\left(\boldsymbol{U}_{K+1}^n\right)\right]. \tag{47}$$

Now we add and subtract $\frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_H\,\alpha_H^n\,\boldsymbol{U}_K^n$ and $\frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_L\,\alpha_L^n\,\boldsymbol{U}_K^n$, yielding:

$$\left(1 - \eta_K^n\,\sqrt{\gamma}_H\,\alpha_H^n - \eta_K^n\,\sqrt{\gamma}_L\,\alpha_L^n\right)\boldsymbol{U}_K^n \tag{48}$$

$$+ \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_H\,\alpha_H^n\left[\boldsymbol{U}_K^n - \frac{1}{\alpha_H^n}\boldsymbol{F}\left(\boldsymbol{U}_K^n\right)\right] + \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_L\,\alpha_L^n\left[\boldsymbol{U}_K^n + \frac{1}{\alpha_L^n}\boldsymbol{F}\left(\boldsymbol{U}_K^n\right)\right] \tag{49}$$

$$+ \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_L\,\alpha_L^n\left[\boldsymbol{U}_{K-1}^n + \frac{1}{\alpha_L^n}\boldsymbol{F}\left(\boldsymbol{U}_{K-1}^n\right)\right] + \frac{1}{2}\,\eta_K^n\,\sqrt{\gamma}_H\,\alpha_H^n\left[\boldsymbol{U}_{K+1}^n - \frac{1}{\alpha_H^n}\boldsymbol{F}\left(\boldsymbol{U}_{K+1}^n\right)\right]. \tag{50}$$

All of the terms in square brackets are similar to the $\boldsymbol{H}_K$ quantities in Qin et al. (2016), and are therefore in $\mathcal{G}$. It can easily be seen that the sum of the coefficients is unity. The final condition is that the coefficient of $\boldsymbol{U}_K^n > 0$, or (recalling that $\eta_K^n = \Delta t_K / \left(\varepsilon\,\Delta V_K\right)$):

$$1 - \eta_K^n\,\sqrt{\gamma}_H\,\alpha_H^n - \eta_K^n\,\sqrt{\gamma}_L\,\alpha_L^n > 0 \implies \eta_K^n\left(\sqrt{\gamma}_H\,\alpha_H^n + \sqrt{\gamma}_L\,\alpha_L^n\right) < 1 \tag{51}$$

$$\implies \Delta t_K^n < \frac{\varepsilon\,\Delta V_K}{\sqrt{\gamma}_H\,\alpha_H^n + \sqrt{\gamma}_L\,\alpha_L^n} \leq \frac{\varepsilon\,\Delta V_K}{2\max\left(\sqrt{\gamma}_{K\pm\frac{1}{2}}\,\alpha_{K\pm\frac{1}{2}}^n\right)}. \tag{52}$$

Again we want a time-step that is the same for all elements at a given time, so:

$$\Delta t^n < \min_K\left(\frac{\varepsilon\,\Delta V_K}{2\max\left(\sqrt{\gamma}_{K\pm\frac{1}{2}}\,\alpha_{K\pm\frac{1}{2}}^n\right)}\right). \tag{53}$$

We close the numerical flux section by writing the explicit form of the time-step for spherical-polar coordinates.

Time-step for Spherical-Polar Coordinates

For spherical-polar coordinates in 1-D we have that $\Delta V_K = 1/3\left(r_H^3 - r_L^3\right)$, and (assuming $\alpha_{K\pm\frac{1}{2}} = 1 \;\forall\; i$) max $\left(\sqrt{\gamma}_{K\pm\frac{1}{2}}\,\alpha_{K\pm\frac{1}{2}}\right) = r_H^2$, so:

$$\Delta t < \min_i\left\{\frac{\varepsilon\, 1/3\left[r_H^3 - r_L^3\right]}{2\, r_H^2}\right\} \tag{54}$$

$$= \min_i\left\{\frac{\varepsilon}{6}\, r_H\left[1 - \frac{r_L^3}{r_H^3}\right]\right\} \tag{55}$$

$$= \min_i\left\{\frac{\varepsilon}{6}\, r_H\left[1 - \left(1 - \frac{\Delta r_i}{r_H}\right)^3\right]\right\} \tag{56}$$

$$= \min_i\left\{\frac{\varepsilon}{6}\, r_H\left[1 - \left(1 + \left(\frac{\Delta r_i}{r_H}\right)^2 - 2\frac{\Delta r_i}{r_H}\right)\left(1 - \frac{\Delta r_i}{r_H}\right)\right]\right\} \tag{57}$$

$$= \min_i\left\{\frac{\varepsilon}{6}\, r_H\left[\left(\frac{\Delta r_i}{r_H}\right)^3 - 3\left(\frac{\Delta r_i}{r_H}\right)^2 + 3\frac{\Delta r_i}{r_H}\right]\right\} \tag{58}$$

$$= \min_i\left\{\frac{\varepsilon}{6}\, \Delta r_i\left[\left(\frac{\Delta r_i}{r_H}\right)^2 - 3\left(\frac{\Delta r_i}{r_H}\right) + 3\right]\right\}. \tag{59}$$

We know that $\Delta r_i / r_H \in [0,1]$; the minimum value of the quadratic function in this domain is unity. So, we have that for spherical-polar coordinates:

$$\Delta t < \frac{\varepsilon}{6}\min\left(\Delta r_i\right). \tag{60}$$

Next we handle the source term.

2.2.4. Source term

For this section we drop the subscript K and the superscript n (but keep in mind that all quantities are still cell-averages). We have to show that $\boldsymbol{H}_2 \in \mathcal{G}$, where

$$\boldsymbol{H}_2 = \begin{pmatrix} D \\ S_j + \frac{\Delta t}{1-\varepsilon}\, Q_j \\ \tau \end{pmatrix}, \quad \left(H_2\right)_1 > 0, \quad \left(H_2\right)_5 + \left(H_2\right)_1 > \sqrt{\left(H_2\right)_1\left(H_2\right)_1 + \left(H_2\right)_j\left(H_2\right)^j}. \tag{61}$$

It is clear that the first requirement for $\boldsymbol{H}_2$ is met, i.e. $D > 0$. The second requirement is:

$$D + \tau > \sqrt{D^2 + \left[S_j + \frac{\Delta t}{1-\varepsilon}Q_j\right]\left[S^j + \frac{\Delta t}{1-\varepsilon}Q^j\right]} \tag{62}$$

$$= \sqrt{D^2 + S_j\, S^j + \frac{\Delta t}{1-\varepsilon}\left(S_j\, Q^j + S^j\, Q_j\right) + \left(\frac{\Delta t}{1-\varepsilon}\right)^2 Q_j\, Q^j} \tag{63}$$

Now we square both sides:

$$D^2 + \tau^2 + 2\,D\,\tau > D^2 + S_j\,S^j + \frac{\Delta t}{1-\varepsilon}\left(S_j\,Q^j + S^j\,Q_j\right) + \left(\frac{\Delta t}{1-\varepsilon}\right)^2 Q_j\,Q^j \tag{64}$$

$$\Longrightarrow \tau\,(\tau + 2\,D) > \gamma^{jk}\,S_j\,S_k + \frac{2\,\Delta t}{1-\varepsilon}\,\gamma^{jk}\,S_j\,Q_k + \left(\frac{\Delta t}{1-\varepsilon}\right)^2 \gamma^{jk}\,Q_j\,Q_k \tag{65}$$

$$\Longrightarrow \tau\,(\tau + 2\,D) > \frac{S_j\,S_k}{\gamma_{jk}} + \frac{2\,\Delta t}{1-\varepsilon}\,\frac{S_j\,Q_k}{\gamma_{jk}} + \left(\frac{\Delta t}{1-\varepsilon}\right)^2 \frac{Q_j\,Q_k}{\gamma_{jk}} \tag{66}$$

$$\Longrightarrow a\,(\Delta t)^2 + b\,\Delta t + c < 0, \tag{67}$$

where:

$$a = \frac{1}{(1-\varepsilon)^2}\,\overrightarrow{Q}\cdot\overrightarrow{Q} = \frac{1}{(1-\varepsilon)^2}\,\frac{Q_j\,Q_k}{\gamma_{jk}} = \frac{1}{(1-\varepsilon)^2}\sum_{k=1}^{3}\frac{(Q_k)^2}{\gamma_{kk}} \tag{68}$$

$$b = \frac{2}{1-\varepsilon}\,\overrightarrow{S}\cdot\overrightarrow{Q} = \frac{2}{1-\varepsilon}\,\frac{S_j\,Q_k}{\gamma_{jk}} = \frac{2}{1-\varepsilon}\sum_{k=1}^{3}\frac{S_k\,Q_k}{\gamma_{kk}} \tag{69}$$

$$c = -\tau\,(\tau + 2\,D) + \overrightarrow{S}\cdot\overrightarrow{S} = -\tau\,(\tau + 2\,D) + \sum_{k=1}^{3}\frac{(S_k)^2}{\gamma_{kk}}. \tag{70}$$

We want to make sure that our function has at least one real root, which means we must have that $b^2 - 4\,a\,c \geq 0$:

$$b^2 - 4\,a\,c = \frac{4}{(1-\varepsilon)^2}\left(\overrightarrow{S}\cdot\overrightarrow{Q}\right)^2 - \frac{4}{(1-\varepsilon)^2}\,\overrightarrow{Q}\cdot\overrightarrow{Q}\left[-\tau\,(\tau + 2\,D) + \overrightarrow{S}\cdot\overrightarrow{S}\right] \tag{71}$$

$$= \frac{4}{(1-\varepsilon)^2}\left[\left(\overrightarrow{S}\cdot\overrightarrow{Q}\right)^2 - \left(\overrightarrow{Q}\cdot\overrightarrow{Q}\right)\left(\overrightarrow{S}\cdot\overrightarrow{S}\right) + \tau\,(\tau + 2\,D)\,\overrightarrow{Q}\cdot\overrightarrow{Q}\right] \tag{72}$$

$$= \frac{4}{(1-\varepsilon)^2}\left[\left|\overrightarrow{S}\right|^2\left|\overrightarrow{Q}\right|^2\cos^2\theta_{SQ} - \left|\overrightarrow{Q}\right|^2\left|\overrightarrow{S}\right|^2 + \tau\,(\tau + 2\,D)\left|\overrightarrow{Q}\right|^2\right] \tag{73}$$

$$= \frac{4}{(1-\varepsilon)^2}\left|\overrightarrow{Q}\right|^2\left[\tau\,(\tau + 2\,D) - \left|\overrightarrow{S}\right|^2\left(1 - \cos^2\theta_{SQ}\right)\right] \tag{74}$$

$$= \frac{4}{(1-\varepsilon)^2}\left|\overrightarrow{Q}\right|^2\left[\tau\,(\tau + 2\,D) - \left|\overrightarrow{S}\right|^2\sin^2\theta_{SQ}\right], \tag{75}$$

where θ_{SQ} is the angle between the momentum-density vector and the source-term vector. To guarantee at least one real root we must have that

$$\tau\,(\tau + 2\,D) \geq \left|\overrightarrow{S}\right|^2\sin^2\theta_{SQ}. \tag{76}$$

$$b^2 - 4\,a\,c' = \frac{4\,(S_1)^2}{\gamma_{11}} a - 4\,a\left(S_1\,S^1 - \tau^2 - 2\,D\,\tau\right) = 4\,a\,\tau\,(\tau + 2\,D)\,. \tag{77}$$

Since $\tau \geq 0$, we must have that $\tau > -2\,D$. But, from condition two for $\boldsymbol{H}_{K,2}$ we have that $\tau > -D$, so this condition is automatically satisfied.

The solutions to this quadratic equation are:

$$\Delta t = \frac{-b}{2\,a} \pm \frac{1}{2\,a}\sqrt{b^2 - 4\,a\,c'} = \frac{-S_1}{\sqrt{\gamma_{11}}\,\sqrt{a}} \pm \frac{1}{2\,a}\sqrt{\frac{4S_1^2}{\gamma_{11}} a - 4\,a\,c'} \tag{78}$$

$$= \frac{-S_1}{\sqrt{\gamma_{11}}\,\sqrt{a}} \pm \frac{1}{\sqrt{\gamma_{11}}\,\sqrt{a}}\sqrt{S_1^2 - c'\,\gamma_{11}} = \frac{1}{\sqrt{\gamma_{11}}\,\sqrt{a}}\left[-S_1 \pm \sqrt{S_1^2 - c'\,\gamma_{11}}\right] \tag{79}$$

$$= \frac{1}{\sqrt{\gamma_{11}}\,\sqrt{a}}\left[-S_1 \pm \sqrt{\gamma_{11}\,(\tau^2 + 2\,D\,\tau)}\right] \tag{80}$$

$$= \frac{2\,(1-\varepsilon)}{P^{kk}\,\partial_1\,\gamma_{kk}}\left[-S_1 \pm \sqrt{\gamma_{11}\,(\tau^2 + 2\,D\,\tau)}\right] \tag{81}$$

$$= \frac{2\,(1-\varepsilon)}{P^{kk}\,\partial_1\,\gamma_{kk}}\left[-S_1 \pm \sqrt{\gamma_{11}\,\tau\,(\tau + 2\,D)}\right]. \tag{82}$$

So, we end up with:

$$\Delta t < \min\left\{\min_i\left(\frac{\varepsilon\,\Delta V_K}{2\max\left(\sqrt{\gamma}_{i\pm\frac{1}{2}}\,\alpha_{i\pm\frac{1}{2}}\right)}\right), \min_i^n\left(\frac{2\,(1-\varepsilon)}{P^{kk}\,\partial_1\,\gamma_{kk}}\left[-S_1 \pm \sqrt{\gamma_{11}\,\tau\,(\tau + 2\,D)}\right]\right)\right\}. \tag{83}$$

2.3. Demanding that $q > 0$

Sometimes it happens that the cell-average of q, $q_K \equiv q\,(\boldsymbol{U}_K) < 0$, so our positivity limiter will fail. To get around this we modify the conserved energy, τ, to demand that $q = \varepsilon$, where $0 < \varepsilon \ll 1$. The transformation we make is:

$$\tau_K \longrightarrow \alpha\,\tau_K, \quad \alpha > 1. \tag{84}$$

This modifies the definition of q_K from:

$$q_K = \tau_K + D_K - \sqrt{D_K^2 + S_K^2 + \varepsilon} < 0, \tag{85}$$

to:

$$\varepsilon = \alpha\,\tau_K + D_K - \sqrt{D_K^2 + S_K^2 + \varepsilon}. \tag{86}$$

Solving this for α, we get:

$$\alpha = \tau_K^{-1}\left[\varepsilon - D_K + \sqrt{D_K^2 + S_K^2 + \varepsilon}\right]. \tag{87}$$

3. Computing the Time-Step Using Higher-Order DG Schemes

When making the jump to higher-order DG schemes, we can simply do the same as in the first-order scheme, except we compute the quantities in all of the nodal points instead of using a cell-average. This is valid because the cell-average is a convex combination...(SD: Need to expand on this). The proof starts with the discretized equation valid at each quadrature point, q:

$$\boldsymbol{U}_q^{n+1} = \boldsymbol{U}_q^n + \Delta t\, \mathcal{L}_q^n, \tag{88}$$

where $\mathcal{L}_q^n$ is a general form of the RHS at time t^n. If we define a vector $\overline{\boldsymbol{U}} \equiv (\boldsymbol{U}_1, \cdots, \boldsymbol{U}_q, \cdots, \boldsymbol{U}_Q)^T$, where Q is the total number of quadrature points, and $\overline{\boldsymbol{W}} \equiv (\boldsymbol{W}_1, \cdots, \boldsymbol{W}_q, \cdots, \boldsymbol{W}_Q)^T$ as a vector of quadrature weights, then we can write the cell-average of $\boldsymbol{U}$ as:

$$\boldsymbol{U}_K \equiv \overline{\boldsymbol{W}}^T \overline{\boldsymbol{U}}. \tag{89}$$

If we then compute the cell-average of the above equation, we get:

$$\boldsymbol{U}_K^{n+1} = \boldsymbol{U}_K^n + \Delta t\, \overline{\boldsymbol{W}}^T\, \overline{\mathcal{L}}_q^n = \overline{\boldsymbol{W}}^T \left(\overline{\boldsymbol{U}}^n + \Delta t\, \overline{\mathcal{L}}^n \right) \tag{90}$$

4. Recovery of Primitive Variables

In order to recover the primitive from the conserved variables we need to solve the nonlinear equation:

$$f\,(p) = p - \overline{p}\,(p) = 0, \tag{91}$$

where $\overline{p}\,(p)$ is the pressure as obtained via the ideal gas equation of state with an initial guess, p:

$$\overline{p} = (\Gamma - 1)\,\rho\,\epsilon, \tag{92}$$

where

$$\rho = \rho\,(\boldsymbol{U}, p)\,, \quad \epsilon = \epsilon\,(\boldsymbol{U}, p)\,. \tag{93}$$

In order to solve this equation we make use of the bisection method, and therefore need bounds on our initial guess for the pressure.

4.1. Upper and Lower Bounds for Pressure

We obtain a lower bound for the pressure with:

$$\tau = D\,(h\,W - 1) - p \implies p = -\,(\tau + D) + D\,h\,W \geq -\,(\tau + D) + D\,h\,W\,\sqrt{v^i\,v_i} = -\,(\tau + D) + \sqrt{S^i\,S_i}. \tag{94}$$

So, since the pressure must be non-negative, we have:

$$p \geq \text{MAX}\left[-\,(\tau + D) + \sqrt{S^i\,S_i}, \text{SqrtTiny}\right]. \tag{95}$$

For an upper bound, we first note that:

$$h = 1 + \frac{e + p}{\rho} = 1 + \frac{\Gamma}{\Gamma - 1}\frac{p}{\rho} = 1 + \frac{\Gamma}{\Gamma - 1}\frac{p\,W}{D}, \tag{96}$$

so,

$$\tau = D\left(W + \frac{\Gamma}{\Gamma - 1}\frac{p\,W^2}{D} - 1\right) - p = D\,(W - 1) + p\left(\frac{\Gamma}{\Gamma - 1}W^2 - 1\right) > p\left(\frac{\Gamma}{\Gamma - 1} - 1\right) = \frac{p}{\Gamma - 1}. \tag{97}$$

So,

$$p < (\Gamma - 1)\,\tau. \tag{98}$$

Typically, $\Gamma - 1 < 3$. So, we end up with:

$$p < 3\,\tau. \tag{99}$$

5. Error Introduced by Linear Interpolation

We obtain the initial conditions for the standing accretion shock problem with an external code. We then have to interpolate these data onto the grid for use in `thornado`. We start by proving that linear interpolation is a convex combination of the solution at the boundary points.

5.1. Proof that Linear Interpolation is a Convex Combination of Boundary-Points

Linear interpolation of a function, f, of a variable, r, bounded by two points r_L and r_H, with $r_H > r_L$, can be written as:

$$f\left(r\right) = f\left(r_L\right) + \frac{f\left(r_H\right) - f\left(r_L\right)}{r_H - r_L}\left(r - r_L\right). \tag{100}$$

By making a change of variables from r to η, where:

$$r\left(\eta\right) = \eta\,r_H + \left(1-\eta\right)r_L = r_L + \eta\,\Delta r \implies \eta = \frac{r - r_L}{\Delta r}, \quad \eta \in \left[0,1\right], \tag{101}$$

such that $r\left(\eta=0\right) = r_L$ and $r\left(\eta=1\right) = r_H$, we have that:

$$f\left(\eta\right) = f\left(0\right) + \frac{f\left(1\right) - f\left(0\right)}{1-0}\left(\eta - 0\right) = f\left(0\right) + \left(f\left(1\right) - f\left(0\right)\right)\eta \tag{102}$$

$$\implies f\left(\eta\right) = \eta\,f\left(1\right) + \left(1-\eta\right)f\left(0\right). \tag{103}$$

So when we interpolate a fluid variable, say the mass-density ρ, we have:

$$\rho\left(r\right) = \rho_L + \frac{\Delta\rho}{\Delta r}\left(r - r_L\right) \longrightarrow \rho\left(\eta\right) = \eta\,\rho_H + \left(1-\eta\right)\rho_L, \tag{104}$$

where $\rho_L \equiv \rho\left(r_L\right)$ and $\rho_H = \rho\left(r_H\right)$.

5.2. Example: $f\left(x\right) = x^2$

We can show this directly for simple functions. We take as an example $f\left(x\right) = x^2$:

$$f\left(x\right) = x^2 = \left[x_L + \eta\left(x_R - x_L\right)\right]^2 = x_L^2 + \eta^2\left(x_R - x_L\right)^2 + 2\,x_L\,\eta\left(x_R - x_L\right) \tag{105}$$

$$= x_L^2 + \eta^2\,x_R^2 + \eta^2\,x_L^2 - 2\,\eta^2\,x_L\,x_R + 2\,\eta\,x_L\,x_R - 2\,\eta\,x_L^2. \tag{106}$$

Now we add and subtract the interpolated solution: $\eta\,x_H^2 + \left(1-\eta\right)x_L^2$. This yields:

$$x^2 = x_L^2 + \eta^2\,x_R^2 + \eta^2\,x_L^2 - 2\,\eta^2\,x_L\,x_R + 2\,\eta\,x_L\,x_R - 2\,\eta\,x_L^2 + \eta\,x_H^2 + \left(1-\eta\right)x_L^2 - \eta\,x_H^2 - \left(1-\eta\right)x_L^2 \tag{107}$$

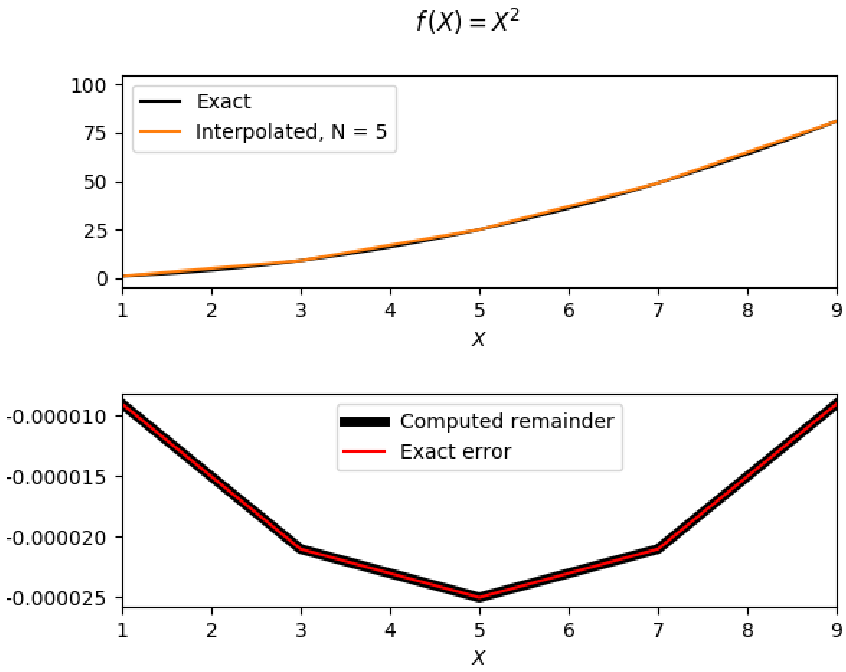
$$= \left[\eta\,x_H^2 + \left(1-\eta\right)x_L^2\right] + \left[-\eta\left(1-\eta\right)x_L^2 - \eta\left(1-\eta\right)x_H^2 + 2\,\eta\,x_L\,x_H\left(1-\eta\right)\right] \tag{108}$$

$$= \left[\eta\,x_H^2 + \left(1-\eta\right)x_L^2\right] - \eta\left(1-\eta\right)\left[x_L^2 + x_H^2 - 2\,x_L\,x_H\right] \tag{109}$$

$$= \left[\eta\,x_H^2 + \left(1-\eta\right)x_L^2\right] - \eta\left(1-\eta\right)\left(x_H - x_L\right)^2 \tag{110}$$

$$= \left[\eta\,x_H^2 + \left(1-\eta\right)x_L^2\right] - \eta\left(1-\eta\right)\left(\Delta x\right)^2. \tag{111}$$

The term in square brackets is the interpolated solution, and the second term is the remainder, i.e. the error in the interpolated solution. We show the results of this in the following figure:



5.3. Mass Constant

The mass constant, C_D, from the 3+1 GR hydrodynamics equations is:

$$C_D\,(r) = \psi\,(r)^6\;\alpha\,(r)\;r^2\,\rho\,(r)\;W\,[v\,(r)]\;v\,(r)\,. \tag{112}$$

From the data we have this exact value at the two endpoints:

$$C_D\,(r_L) \equiv C_D^L = \psi_L^6\,\alpha_L\,r_L^2\,\rho_L\,W_L\,v_L \tag{113}$$

$$C_D\,(r_H) \equiv C_D^H = \psi_H^6\,\alpha_H\,r_H^2\,\rho_H\,W_H\,v_H\,. \tag{114}$$

REFERENCES

Mignone, A., & Bodo, G. 2005, Monthly Notices of the Royal Astronomical Society, 364, 126

Qin, T., Shu, C.-W., & Yang, Y. 2016, Journal of Computational Physics, 315, 323

Zhang, X., & Shu, C.-W. 2011, Journal of Computational Physics, 230, 1238

This preprint was prepared with the AAS LaTeX macros v5.0.