Notes on conservative, bounded, advection schemes

Colin Cotter

January 1, 2023

As described in [1], the conservation of a tracer q can be most easily derived from the conservative form of the advection equations

$$(\rho q)_t + \nabla \cdot (\boldsymbol{u}\rho q) = 0, \tag{1}$$

where the density satisfies the continuity equation

$$\rho_t + \nabla \cdot (\boldsymbol{u}\rho) = 0. \tag{2}$$

On the other hand, the L_∞ bounds for q are most easily derived from the characteristic form

$$q_t + \boldsymbol{u} \cdot \nabla q = 0. \tag{3}$$

1 Continuity equation

The first step is that we need a positive scheme for ρ, *i.e.*, $\rho(x,t) > 0$ if $\rho(x,0) > 0$.

The time-continuous discontinuous Galerkin (DG) scheme for $\rho \in DG(n)$ is defined on an element e with boundary ∂e by

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\int_e \phi\rho \,\mathrm{d}\,x - \int_e \nabla\phi \cdot \boldsymbol{u}\rho \,\mathrm{d}\,x + \int_{\partial e} \phi \boldsymbol{u} \cdot \boldsymbol{n}\tilde{\rho} \,\mathrm{d}\,s = 0, \quad \forall \phi \in DG(n), \tag{4}$$

where $\boldsymbol{n}$ is the normal to the boundary, and $\tilde{\rho}$ is the value of ρ on the upwind side of ∂e.

Here we consider the forward Euler time discretisation, given by

$$\int_e \phi\rho^{n+1} \,\mathrm{d}\,x = \int_e \phi\rho^n \,\mathrm{d}\,x + \Delta t\int_e \nabla\phi \cdot \boldsymbol{u}\rho^n \,\mathrm{d}\,x - \Delta t\int_{\partial e} \phi \boldsymbol{u} \cdot \boldsymbol{n}\tilde{\rho}^n \,\mathrm{d}\,s = 0, \quad \forall \phi \in DG(n). \tag{5}$$

Forward Euler steps can be concatenated in SSPRK schemes to obtain higher-order time discretisations.

The above linear scheme is not positivity-preserving. To obtain a positivity-preserving scheme, we take the following steps:

1. A pre-processing limiter applied to ρ, such as the Barth-Jespersen limiter or the Kuzmin vertex-based limiter, that sets bounds $(\rho_{\min}, \rho_{\max})$ on each element e, based on the mean value $\overline{\rho}_e^n$, defined by

   $$\overline{\rho}^n = \frac{1}{|e|}\int_e \rho^n \,\mathrm{d}\,x, \tag{6}$$

   and the value in selected surrounding elements. The limiter is then defined by

   $$\Pi_1\rho^n|_e = \overline{\rho}^n + \alpha(\rho^n - \overline{\rho}^n), \tag{7}$$

   with $0 \leq \alpha \leq 1$ chosen as the maximum value such that $\Pi\rho^n$ lies within the bounds in element e. Note that such a limiter does not alter the element integral of ρ, and so does not violate conservation.

   If the element mean values of ρ^n are all positive, then so are the bounds, and therefore $\Pi\rho^n$ is positive everywhere.

2. A second pre-processing limiter Π_2 to guarantee that the element means $\overline{\rho}^{n+1}$ will all be positive following the Euler step. The goal of this section is to introduce such a limiter.

3. The application of the forward Euler step,

$$\int_e \phi\rho^{n+1}\,\mathrm{d}\,x = \int_e \phi\rho^*\,\mathrm{d}\,x + \Delta t\int_e \nabla\phi\cdot\boldsymbol{u}\rho^*\,\mathrm{d}\,x - \Delta t\int_{\partial e}\phi\boldsymbol{u}\cdot\boldsymbol{n}\rho^*\,\mathrm{d}\,s = 0, \quad \forall\phi\in DG(n), \tag{8}$$

where $\rho^* = \Pi_2\hat{\rho} = \Pi_2\Pi_1\rho^n$.

We now develop the limiter Π_2.

If we choose ϕ as the indicator function for element e, *i.e.*,

$$\phi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x}\in e, \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

then Equation (8) becomes

$$\int_e \rho^{n+1}\,\mathrm{d}\,x = \int_e \rho^*\,\mathrm{d}\,x - \Delta t\int_{\partial e}\boldsymbol{u}\cdot\boldsymbol{n}\rho^*\,\mathrm{d}\,s = 0. \tag{10}$$

This may be written in the form

$$\overline{\rho}^{n+1} = \overline{\rho}^{n}\left(1 + c^- - c^+\right), \tag{11}$$

where

$$c^{\pm} = \pm\frac{\Delta t}{|e|}\int_{\partial e^{\pm}}\boldsymbol{u}\cdot\boldsymbol{n}\frac{\tilde{\rho}^*}{\overline{\rho}^n}\,\mathrm{d}\,s, \tag{12}$$

where ∂e^+, ∂e^- are the parts of ∂e where $\boldsymbol{u}\cdot\boldsymbol{n}$ is positive and negative respectively. Both c^+ and c^- are positive if ρ is positive everywhere.

For $\overline{\rho}^{n+1}$ to be positive, we need

$$c^+ < 1 + c^-. \tag{13}$$

This will always be satisfied for sufficiently small Δt. However, it may be violated if there are dramatic jumps in ρ, even for moderate values of the upwind/downwind Courant numbers

$$\tilde{c}^{\pm} = \pm\frac{\Delta t}{|e|}\int_{\partial e^{\pm}}\boldsymbol{u}\cdot\boldsymbol{n}\,\mathrm{d}\,s. \tag{14}$$

Hence, we use a failsafe limiter to adjust the slope in each element where Equation (13) is not satisfied, to ensure positivity.

Since c^- depends on values of ρ from outside element e, it will not be altered by limiting the slope in element e. Hence, if the positivity condition (13) is not satisfied, applying a limiter

$$\rho^* = \overline{\tilde{\rho}} + \beta\left(\tilde{\rho} - \overline{\tilde{\rho}}\right), \tag{15}$$

where $0\le\beta\le 1$, results in replacing c^+ by

$$\tilde{c}^+ + \beta(c^+ - \tilde{c}^+). \tag{16}$$

Hence, the positivity condition becomes

$$\tilde{c}^+ + \beta(c^+ - \tilde{c}^+) < 1 + c^-. \tag{17}$$

It is always possible to achieve this condition by setting $\beta = 0$, in which case the condition becomes

$$\tilde{c}^+ < 1 + c^-, \tag{18}$$

for which the CFL condition

$$\tilde{c}^+ < 1, \tag{19}$$

is sufficient. Hence, we choose $\beta\le 1$ to be as large as possible provided that Equation (17) is satisfied, *i.e.*,

$$\beta = \left\{\ \max\left(0, \min\left(1, \tfrac{1+c^- - \tilde{c}^+}{c^+ - \tilde{c}^+}\right)\right)\right. . \tag{20}$$

2 Reconstruction of flux

3 Bounded advection

The forward Euler conservative DG discretisation for the transport equation for q is,

$$\int_e \phi\rho^{n+1}q^{n+1}\,\mathrm{d}\,x = \int_e \phi\rho^n q^n\,\mathrm{d}\,x + \Delta t\int_e \nabla\phi\cdot\boldsymbol{F}^n q^n\,\mathrm{d}\,x - \Delta t\int_{\partial e}\phi\boldsymbol{F}^n\cdot\boldsymbol{n}\tilde{q}^n\,\mathrm{d}\,s = 0, \quad \forall\phi\in DG(n), \tag{21}$$

where $\tilde{q}^n$ is the value of q^n on the upwind side of ∂e.

Since this linear scheme does not guarantee bounds on q^{n+1}, we also apply limiters to ensure this, following the same two step strategy, first ensuring that all values of q lie between bounds defined from element mean values, then adjusting the slope so that the outflow values guarantee positive element mean values in the next timestep. We use ρ-weighted element mean values defined by

$$\overline{q} = \frac{1}{\overline{\rho^n}}\int_e \rho^n q\,\mathrm{d}\,x, \tag{22}$$

so that

$$\overline{\rho^n}\overline{q} = \int_e \rho^n q\,\mathrm{d}\,x. \tag{23}$$

In this case, replacing q by the limited function

$$\Pi_1 q = \overline{q} + \alpha(q-\overline{q}), \tag{24}$$

does not violate conservation, since

$$\int_e \rho^n\Pi_1 q\,\mathrm{d}\,x = \overline{q}\int_e \rho^n\,\mathrm{d}\,x + \alpha\left(\int_e \rho^n q\,\mathrm{d}\,x - \overline{q}\int_e \rho^n\,\mathrm{d}\,x\right), \tag{25}$$

$$= \overline{\rho^n}\overline{q} + \alpha\left(\overline{\rho^n}\overline{q} - \overline{\rho^n}\overline{q}\right) = \overline{\rho^n}\overline{q}, \tag{26}$$

as required.

The forward Euler timestep for q is then as follows:

1. A pre-processing limiter applied to q, such as the Barth-Jespersen limiter or the Kuzmin vertex-based limiter, that sets bounds $(q_{\min}, q_{\max})$ on each element e, based on the mean value $\overline{q}^n_e$, and the value in selected surrounding elements. The limiter is then defined by

$$\Pi_1 q^n|_e = \overline{q}^n + \alpha(q^n - \overline{q}^n), \tag{27}$$

   with $0\leq\alpha\leq 1$ chosen as the maximum value such that Πq^n lies within the bounds in element e.

   If the element mean values of q^n are all positive, then so are the bounds, and therefore Πq^n is positive everywhere.

2. A second pre-processing limiter Π_2 to guarantee that the element means $\overline{q}^{n+1}$ will lie within their bounds at the following the Euler step. The goal of this section is to introduce such a limiter.

3. The application of the forward Euler step,

$$\int_e \phi\rho^{n+1}q^{n+1}\,\mathrm{d}\,x = \int_e \phi\rho^n q^*\,\mathrm{d}\,x + \Delta t\int_e \nabla\phi\cdot\boldsymbol{u}\rho^n q^*\,\mathrm{d}\,x - \Delta t\int_{\partial e}\phi\boldsymbol{u}\cdot\boldsymbol{n}\rho^n q^*\,\mathrm{d}\,s = 0, \quad \forall\phi\in DG(n), \tag{28}$$

   where $q^* = \Pi_2\hat{q} = \Pi_2\Pi_1 q^n$.

We now develop the limiter Π_2 for q, assuming that q already satisfies the bounds in each element.

If we choose ϕ as the indicator function for element e, then Equation (28) becomes

$$\int_e \rho^{n+1}q^{n+1}\,\mathrm{d}\,x = \int_e \rho^n q^*\,\mathrm{d}\,x - \Delta t\int_{\partial e}\boldsymbol{u}\cdot\boldsymbol{n}\rho^n q^*\,\mathrm{d}\,s = 0. \tag{29}$$

This may be written in the form

$$\overline{\rho}^{n+1}\overline{q^{n+1}} = \overline{\rho}^n\left(q^n + c^- q^- - c^+ q^+\right), \tag{30}$$

where

$$\overline{q^{n+1}} = \frac{1}{\overline{\rho^{n+1}}} \int_e \rho^{n+1} q^{n+1} \,\mathrm{d}\, x, \tag{31}$$

$$q^{\pm} = \pm \frac{1}{c^{\pm}} \int_{\partial e^{\pm}} \boldsymbol{F} \cdot \boldsymbol{n} \tilde{q} \,\mathrm{d}\, s, \tag{32}$$

and $c^{\pm} > 0$ are defined as before. Finally, making use of Equation (11), we obtain

$$(1 + c^{-} - c^{+})\overline{q^{n+1}} = \bar{q}^{n} + c^{-} q^{-} - c^{+} q^{+}, \tag{33}$$

where positivity ensures that $1 + c^{-} - c^{+} > 0$. We see that $\bar{q}^{n+1}$ might exceed the bounds if q^{+} is too far below $\bar{q}^{n}$.

If our limiter scheme has an imposed maximum bound $\bar{q}^{n+1} \leq q^{\max}$, we now need to adjust the slope in q^{n},

$$q^{n} \mapsto \bar{q}^{n} + \beta(q^{n} - \bar{q}^{n}), \tag{34}$$

where $0 \leq \beta \leq 1$, in order to replace

$$q^{+} \mapsto \bar{q}^{n} + \beta(q^{+} - \bar{q}^{n}). \tag{35}$$

Since $q^{-} \leq q^{\max}$, since we have already limited the slope so that q satisfies the element bounds, which are the same on each side of each element boundary. Therefore, if we choose $\beta = 0$, the update becomes

$$(1 + c^{-} - c^{+})\bar{q}^{n+1} = \bar{q}^{n}(1 - c^{-}) + c^{-} q^{-} < q^{\max}(1 + c^{-} - c^{+}), \tag{36}$$

and the bound will be satisfied as required. We choose the maximum value of $\beta \leq 1$ such that

$$\bar{q}^{n}(1 - c^{+}) + c^{-} q^{-} - c^{+} \beta(q^{+} - \bar{q}^{n}) \leq (1 + c^{-} - c^{+}) q^{\max}, \tag{37}$$

i.e.,

$$\beta = \min\left(1, \frac{(1 + c^{-} - c^{+}) q^{\max} - (\bar{q}^{n}(1 - c^{+}) + c^{-} q^{-})}{c^{+}(\bar{q}^{n} - q^{+})}\right). \tag{38}$$

References

[1] John Thuburn. Multidimensional flux-limited advection schemes. *J. Comput. Phys.*, 123(1):74–83, jan 1996.