

New limiter regions for multidimensional flows

James Woodfield^{a,b,*}, Hilary Weller^c, Colin J. Cotter^a

^a Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom

^b Department of Mathematics and Statistics, University of Reading, Whiteknights, Reading RG6 6AX, United Kingdom

^c Department of Meteorology, University of Reading, Earley Gate, Reading RG6 6BB, United Kingdom

ABSTRACT

Accurate transport algorithms are crucial for computational fluid dynamics and more accurate and efficient schemes are always in development. One dimensional limiting is commonly employed to suppress nonphysical oscillations. However, the application of such limiters can reduce accuracy. It is important to identify the weakest set of sufficient conditions required on the limiter as to allow the development of successful numerical algorithms. The main goal of this paper is to identify new less restrictive sufficient conditions for flux form in-compressible advection to remain monotonic. We identify additional necessary conditions for incompressible flux form advection to be monotonic, demonstrating that the Spekreijse limiter region is not sufficient for incompressible flux form advection to remain monotonic. Then a convex combination argument is used to derive new sufficient conditions that are less restrictive than the Sweby region for a discrete maximum principle. This allows the introduction of two new more general limiter regions suitable for flux form incompressible advection.

1. Introduction

1.1. Historical context and motivation

Higher order accuracy and monotonicity have long been at odds. Godunov [1] showed that the order of a linear monotonicity preserving numerical scheme is necessarily first order. Harten, Hyman, Lax and Keyfitz (HHLK) [2] generalised the Godunov order barrier theorem, and proved that nonlinear schemes with the stronger HHLK-monotone property are necessarily first order. Harten [3] developed nonlinear total variation diminishing (TVD) schemes, satisfying a monotonic property (TVD) weaker than the HHLK-monotone property but stronger than the monotonicity preserving property of Godunov [1]. These TVD schemes bypass the Godunov order barrier, can be second order and have a convenient framework and limiter region introduced by Sweby in 1984 [4].

However, the total variation diminishing framework did not generalise easily for multi-dimensional flows, in part due to Goodman and Leveque showing that a particular two-dimensional definition of TVD would also run into the HHLK order barrier [5]. Nevertheless, in 1987 Spekreijse [6] developed a monotonicity framework suitable for multidimensional flow, it is based on positivity of coefficients and can be used to enforce a discrete local maximum principle (strictly stronger than the total variation diminishing property). Furthermore Spekreijse's [6] positive coefficient framework was applied to a wide class of flux splitting schemes and generalised to unstructured meshes. The corresponding spatial discretisations are referred to as positive coefficient schemes or Local Extrema Diminishing schemes [7]. Spekreijse also showed that less stringent requirements were demanded of the limiter functions, these new weaker set of sufficient conditions required on the limiter function allowed a variety of new flux and slope limiters to be introduced. Of particular note is the development of differentiable limiters contained in the second order region.

* Corresponding author at: Department of Mathematics, Imperial College London, South Kensington Campus, London SW7 2AZ, United Kingdom.
E-mail addresses: James14641@gmail.com, jmw418@ic.ac.uk (J. Woodfield).

Spekreijse's work has since been used to motivate monotonic atmospheric advection: Hundsdorfer et al. [8] proposed using one-dimensional flux limiters in a multidimensional flux form finite difference spatial discretisation, for positive mass preserving transport on the sphere. Hundsdorfer et al. [8] observed preferable accuracy and computational speed using one-dimensional limiters in a method of lines framework as compared with the third order MPDATA scheme in [9]. Wesseling and Zijlema [10] proposes a very similar scheme under a finite volume interpretation, and suggest some limiters in the Spekreijse region could be used to improve accuracy. We will analyse a generalisation of both the numerical method introduced in Wesseling and Zijlema [10] and Hundsdorfer et al. [8], to find the admissible flux limiter region for the incompressible advection equation in flux form. It is known that using one-dimensional limiters in the Spekreijse region can create a one dimensional TVD scheme for uniform flow. It is also believed that the monotonicity properties carry to the multidimensional flux form advection equation case using the positive coefficient framework of Spekreijse [6], some examples of this belief can be found in [11,10,8]. However, it has never been proven and to what extent that this is true has never been resolved.

Whether Spekreijse's limiter region is applicable to multidimensional flux form transport schemes is important as limiters such as OSPRE, van Albada, Albada family, Hemker-Koren, local double-logarithmic reconstruction (LDLR), Čada-Torrilhon, TCDF, MPL2- κ and ENO2 may fail to have a provable discrete maximum principle and consequently could lose positivity which could potentially lead to instability when transporting dense scalars such as liquid water. The region in which limiters are provably discrete maximum principle satisfying is crucial to the successful design of monotone advection algorithms and is the main focus of this paper.

In section 1.2 we define and plot both the Sweby Region, and the Spekreijse limiter Region, in the context of one dimensional constant coefficient advection as to put the rest of the paper in context with the existing literature. In section 2, we introduce the advection equation and the flux form numerical method of interest, a generalisation of the finite volume method in Wesseling and Zijlema [10] and the finite difference method in Hundsdorfer et al. [8]. In section 3.2 we explain why Spekreijse's framework does not necessarily apply to flux form advection schemes, and why the extended admissible limiter region of Spekreijse can fail. Section 3.1 contains the main contribution of this paper, the derivation of two new limiter regions sufficient for the incompressible flux form scheme defined in section 2.2 to remain discrete maximum principle satisfying.

Section 4 contains a wider discussion on the theoretical properties of the numerical scheme introduced in section 2.2, in particular how the method of lines interacts with monotonicity properties, a review of the spatial accuracy a short remark on linear invariance and the introduction of several new limiter functions. In section 5 we include several numerical tests to see to what extent the theoretical results in section 3.2 and section 3.1 are observed in practice. In section 5 we numerically investigate the limiters introduced in section 4.

1.2. Background: different existing limiter regions

A semi discrete flux limited numerical scheme for the constant coefficient $v \in \mathbb{R}^{>0}$ advection equation $u_t + vu_x = 0$ can be written as $\frac{du_i}{dt} + \Delta x^{-1}(F_{i+1/2} - F_{i-1/2}) = 0$, where the flux is given by $F_{i+1/2} = v[u_i + 1/2\psi(R_i)(u_i - u_{i-1})]$, here ψ is the limiter function, and the ratio of successive gradients is given by $R_i = (u_{i+1} - u_i)/(u_i - u_{i-1})$. A similar but different construction is given using the inverse ratio $r_i = 1/R_i$ with flux given by $F_{i+1/2} = v[u_i + 1/2\psi(r_i)(u_{i+1} - u_i)]$. We distinguish these different frameworks by the parameter $\theta = 1, 0$, respectively. The total variation of u is defined as $TV(u) = \sum_{i,j} |u_i - u_{i-1}|$, and serves as a measure of how oscillatory u is. A desirable property of a numerical method is that the total variation does not increase with time. This property is referred to as total variation diminishing (TVD).

In 1984, Sweby [4] introduced sufficient conditions for a flux limited scheme in the $\theta = 0$ formulation to be total variation diminishing,

$$\psi(r) \leq \min(2r, 2), \quad \forall r > 0, \quad (1.1)$$

$$\psi(r) = 0, \quad \forall r < 0. \quad (1.2)$$

These conditions serve as bounds on acceptable limiters in the Sweby diagram, and we shall refer to the set of points $(r, \psi(r))$ for which conditions eqs. (1.1) and (1.2) are satisfied as the Sweby region. However, it is sometimes assumed that these bounds are necessary for schemes to be total variation diminishing [12].

Spekreijse showed that this is not the case in the $\theta = 1$ formulation by finding a more general (TVD) admissible limiter region,

$$\psi(R) \in [\alpha, M], \quad \psi(S)/S \in [-M, 2 + \alpha], \quad M \in (0, \infty), \quad \alpha \in [-\infty, 0], \quad (1.3)$$

for $R, S \in \mathbb{R}$, similarly we shall refer to the set of points $(R, \psi(R))$ for which the conditions in eq. (1.3) are satisfied as the Spekreijse region. Spekreijse showed that flux splitting schemes with limiters in this region satisfy a local discrete maximum principle, which is strictly stronger than the total variation diminishing property in one dimension and is more convenient in multiple dimensions. The use of limiters in the extended region of Spekreijse has been successful in both flux difference splitting [13] and flux vector splitting frameworks [6]. Fig. 1.1 shows both the Sweby region $\mathcal{D}_1$ in Fig. 1.1a, and the Spekreijse region $\mathcal{D}_2$ in Fig. 1.1b. The extended limiter region has allowed a huge variety of new limiters to be introduced into the literature, which can improve accuracy [14], and are not necessarily reduced to the first order upwind scheme at extrema and can be globally smooth [6]. It is worth noting, the $\theta = 0$ framework also possesses a two-parameter generalisation [15,14], which we shall also refer to as the Spekreijse region, or more generally an extended limiter region. In the flux splitting framework, there also exist other extended limiter regions, for a selective literature review see Appendix J. As remarked earlier, the extent to which these extended limiter regions are suitable for flux form incompressible advection has not been resolved.

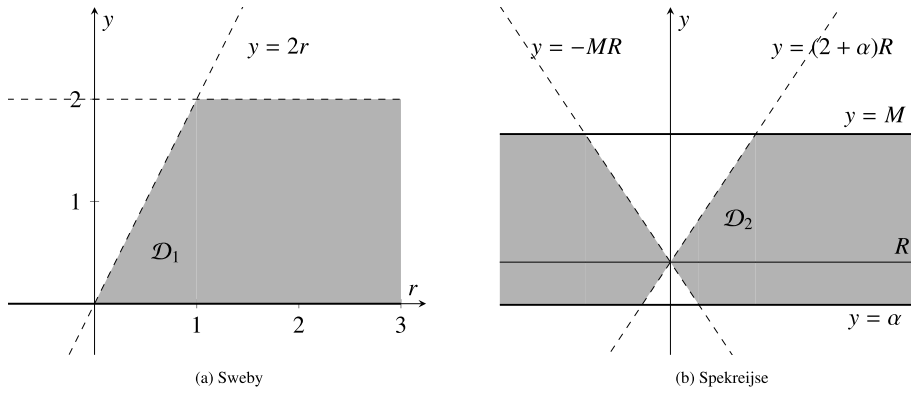


Fig. 1.1. Fig. 1.1a is the plot of the sufficient admissible limiter region in the (r, y) plane as defined by Sweby [4]. Fig. 1.1b is the plot of the admissible limiter region in the (R, y) plane as defined by Spekreijse [6]. The Sweby region D_1 is sufficient for a one-dimensional scheme with flux limiters to be TVD, the Spekreijse region D_2 , which has two free parameters $\alpha \in (-\infty, 0]$, $M \in (0, \infty)$, is also sufficient for the scheme to be TVD.

2. Governing equation and numerical scheme

2.1. Governing equation

We consider the evolution of a bounded initial tracer $u_0(\mathbf{x}) \in L^\infty(\Omega)$ in a arbitrary spatial domain, $\Omega \subseteq \mathbb{R}^d$, $d \geq 1$, over the time interval $[0, T]$ by a bounded $\mathbf{v}(\mathbf{x}, t) \in L^\infty(\Omega \times [0, T])$ divergence free $\text{div}(\mathbf{v}(\mathbf{x}, t)) = 0$ vector field. This is described by the advection equation,

$$\frac{\partial u}{\partial t} + \text{div}(\mathbf{v}u) = 0, \quad \forall (\mathbf{x}, t) \in \Omega \times [0, T], \quad (2.1)$$

$$u(\mathbf{x}, 0) = u_0(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega, \quad (2.2)$$

which is a model for the evolution of densities, and for incompressible flows, scalars.

2.2. Semi discrete scheme

We introduce the notation and describe a generalisation of the flux form finite volume scheme of Wesseling and Zijlema [10] and the finite difference scheme of Hundsdorfer et al. [8].

1. Let $u = u_{i,j}$ denote either the cell mean or pointwise value of a tracer within a cell, where if one or other subscript is missing it is assumed to be at position i or j as appropriate.
2. Reconstruct: A reconstruction operator constructs the values attained to the right, left, up and down of cell (i, j) as follows

$$u_i^R = u_i + \frac{\theta}{2} \psi(R_i)(u_i - u_{i-1}) + \frac{(1-\theta)}{2} \psi\left(\frac{1}{R_i}\right)(u_{i+1} - u_i), \quad (2.3)$$

$$u_i^L = u_i + \frac{\theta}{2} \psi\left(\frac{1}{R_i}\right)(u_i - u_{i+1}) - \frac{(1-\theta)}{2} \psi(R_i)(u_i - u_{i-1}), \quad (2.4)$$

$$u_j^U = u_j + \frac{\theta}{2} \psi(R_j)(u_j - u_{j-1}) + \frac{(1-\theta)}{2} \psi\left(\frac{1}{R_j}\right)(u_{j+1} - u_j), \quad (2.5)$$

$$u_j^D = u_j + \frac{\theta}{2} \psi\left(\frac{1}{R_j}\right)(u_j - u_{j+1}) - \frac{(1-\theta)}{2} \psi(R_j)(u_j - u_{j-1}), \quad (2.6)$$

$$R_i = \frac{u_{i+1} - u_i}{u_i - u_{i-1}}, \quad R_j = \frac{u_{j+1} - u_j}{u_j - u_{j-1}}, \quad (2.7)$$

this reconstruction imposes upwind bias flux limiting. The right, left, up and down of cell (i, j) are denoted with R, L, U, D superscripts. We are using the ratio of successive gradients defined in [6] as R when $\theta = 1$, we call this the Roe gradient. When $\theta = 0$ we are using the inverse ratio defined in [4] $r_i = 1/R_i$, we call this the Sweby gradient. The choice of θ does not matter when the limiter is symmetric $\psi(1/R) = \psi(R)/R$. We only investigate $\theta = \{0, 1\}$ due to the reduced computational cost in evaluating the expressions in eqs. (2.3) to (2.6). When the symmetry condition is broken the two limiter functions should be distinguished differently, as they have their own respective limiter regions, however this should be apparent from the context of the limiter, so we do not include this notationally.

3. Riemann: the donor cell numerical flux function

$$F(u_i^R, u_{i+1}^L, c_{i+0.5}) = c_{i+0.5}^+ u_i^R + c_{i+0.5}^- u_{i+1}^L, \quad (2.8)$$

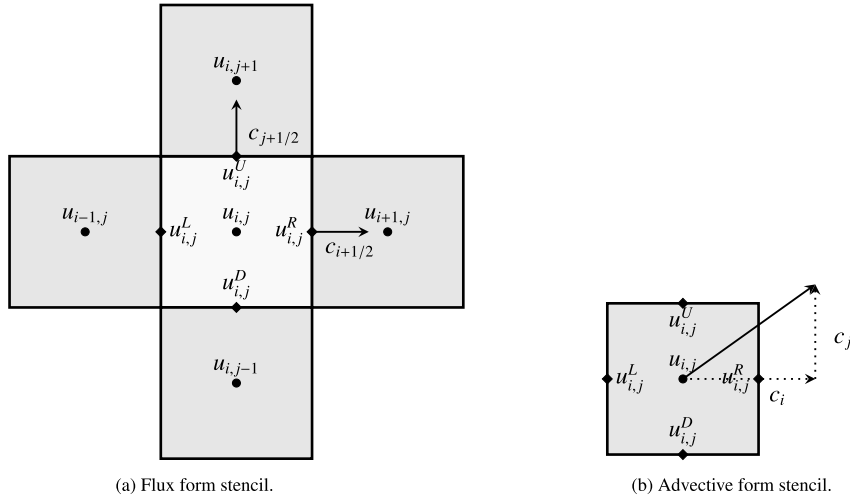


Fig. 2.1. Fig. 2.1a Shows variable placement for a flux form scheme, consistent with numerical method eq. (3.1), where the velocity field is located at cell faces/edges. Fig. 2.1b shows variable placement for an advective form scheme, consistent with eq. (2.10), where the velocity is located at the cell centre. One can use Spekreijse's limiter region using the advective form stencil, but not when using the flux form stencil (Theorem 3.2). When using a flux form scheme, one can instead use the limiter regions presented in this paper Fig. 3.2, Theorem 3.1.

resolves the Riemann problems at the cell boundaries. Here we define the notation $(\cdot)^+ := \max(\cdot, 0)$, $(\cdot)^- := \min(\cdot, 0)$, to mean the positive or negative component of the argument respectively, to be used throughout this paper. $c_{i+1/2}$ denotes the x component of the velocity at position $(i + 1/2, j)$, and $c_{j+1/2}$ denotes the y component of the velocity at position $(i, j + 1/2)$. We often absorb constants like timestep and mesh width into this constant because the numerical flux function is linear. This numerical flux function is consistent, monotone, and Lipschitz continuous as defined in [16]. This type of flux is known as a state interpolated flux as defined in [11,17,8], rather than a flux interpolated flux. In the finite volume setting we approximate the integral of flux through the face by second order Gauss quadrature over each face (midpoint rule), whereas in the finite difference setting we interpret as a point valued flux.

4. Evolve: the semi discrete evolution operator is given by the flux form method

$$\frac{\partial u}{\partial t} + [F_{i+0.5}(u_i^R, u_{i+1}^L, c_{i+0.5}) - F_{i-0.5}(u_{i-1}^R, u_i^L, c_{i-0.5})] + [F_{j+0.5}(u_j^U, u_{j+1}^D, c_{j+0.5}) - F_{j-0.5}(u_{j-1}^U, u_j^D, c_{j-0.5})] = 0. \quad (2.9)$$

We have absorbed the mesh spacing into the coefficients c by linearity of the donor cell flux function, and plan to do so with the time-step Δt .

For context see [6], [10], [8], and for an example of its generalisation to unstructured meshes see [18]. The above scheme is a flux form scheme with velocities defined at cell edges, see Fig. 2.1a. In contrast, an advective form scheme uses a cell-centred velocity c_i, c_j (see Fig. 2.1b) such that the scheme has the following form

$$\frac{\partial u}{\partial t} + [F_{i+0.5}(u_i^R, u_{i+1}^L, c_i) - F_{i-0.5}(u_{i-1}^R, u_i^L, c_i)] + [F_{j+0.5}(u_j^R, u_{j+1}^L, c_j) - F_{j-0.5}(u_{j-1}^R, u_j^L, c_j)]. \quad (2.10)$$

The advective form scheme is provably monotone using the Spekreijse's limiter region and framework [6]. However, this argument is not always valid in the flux form multidimensional case, as will be explained in section 3.2. We present the theoretical results regarding flux form scheme in the next section.

3. Theoretical monotonicity properties of the scheme

We present and derive less restrictive sufficient conditions on the limiters functions ψ to prove that the semidiscrete numerical method (eq. (3.1))

$$\begin{aligned} \frac{\partial u}{\partial t} + [F_{i+0.5}(u_i^R, u_{i+1}^L, c_{i+0.5}) - F_{i-0.5}(u_{i-1}^R, u_i^L, c_{i-0.5})] + [F_{j+0.5}(u_j^U, u_{j+1}^D, c_{j+0.5}) - F_{j-0.5}(u_{j-1}^U, u_j^D, c_{j-0.5})] &= 0, \\ F(u_i^R, u_{i+1}^L, c_{i+0.5}) &= c_{i+0.5}^+ u_i^R + c_{i+0.5}^- u_{i+1}^L, \quad F(u_{i-1}^R, u_i^L, c_{i-0.5}) = c_{i-0.5}^+ u_{i-1}^R + c_{i-0.5}^- u_i^L, \\ F(u_j^U, u_{j+1}^D, c_{j+0.5}) &= c_{j+0.5}^+ u_j^U + c_{j+0.5}^- u_{j+1}^D, \quad F(u_{j-1}^U, u_j^D, c_{j-0.5}) = c_{j-0.5}^+ u_{j-1}^U + c_{j-0.5}^- u_j^D, \\ u_i^R &= u_i + \frac{\theta}{2} \psi(R_i)(u_i - u_{i-1}) + \frac{(1-\theta)}{2} \psi\left(\frac{1}{R_i}\right)(u_{i+1} - u_i), \\ u_i^L &= u_i + \frac{\theta}{2} \psi\left(\frac{1}{R_i}\right)(u_i - u_{i+1}) - \frac{(1-\theta)}{2} \psi(R_i)(u_i - u_{i-1}), \end{aligned} \quad (3.1)$$

$$\begin{aligned}
u_j^U &= u_j + \frac{\theta}{2} \psi(R_j)(u_j - u_{j-1}) + \frac{(1-\theta)}{2} \psi\left(\frac{1}{R_j}\right)(u_{j+1} - u_j), \\
u_j^D &= u_j + \frac{\theta}{2} \psi\left(\frac{1}{R_j}\right)(u_j - u_{j+1}) - \frac{(1-\theta)}{2} \psi(R_j)(u_j - u_{j-1}), \\
R_i &= \frac{u_{i+1} - u_i}{u_i - u_{i-1}}, \quad R_j = \frac{u_{j+1} - u_j}{u_j - u_{j-1}},
\end{aligned}$$

is Local Extrema Diminishing (LED) for incompressible flow for $\theta \in \{0, 1\}$ (We introduced all necessary notation and explanation of this scheme in section 2.2). This can be used to identify less strict conditions in which the forward Euler numerical flow map $u^{n+1} = \text{Forward Euler}(u^n, c^n)$, defined by

$$u_{i,j}^{n+1} = u_{i,j}^n - [F_{i+0.5}(u_i^R, u_{i+1}^L, c_{i+0.5}^n) - F_{i-0.5}(u_{i-1}^R, u_i^L, c_{i-0.5}^n)] - [F_{j+0.5}(u_j^U, u_{j+1}^D, c_{j+0.5}^n) - F_{j-0.5}(u_{j-1}^U, u_j^D, c_{j-0.5}^n)], \quad (3.2)$$

can satisfy a Discrete Local Maximum Principle (DLMP). The time step, Δt , and mesh spacing, Δx , can be absorbed into the face defined velocity field c , as and when appropriate. This scheme generalises the finite volume scheme in [10], and the finite difference scheme in [8], and can be used with a Strong Stability Preserving timestepping scheme to achieve higher order accuracy whilst retaining monotonic properties.

3.1. New limiter regions suitable for incompressible flow

In this section we will show that under new considerations of the limiter function we can put the scheme in a positive coefficient type form Definition 3.1 as defined by Spekreijse [6]. This definition allows use of Lemma 3.2 also given by Spekreijse [6] which can be used to prove a discrete local maximum principle for the forward Euler flow map. And also allows the use of a Lemma by Jameson [7] used to prove the Local Extrema Diminishing (LED) property of the semi-discrete scheme.

Definition 3.1 (Spekreijse, Positive-coefficient type scheme [6]). The semidiscrete scheme given by

$$\frac{\partial u_{i,j}}{\partial t} + A_{i-1/2}(u_i - u_{i-1}) + B_{i+1/2}(u_i - u_{i+1}) + C_{j-1/2}(u_j - u_{j-1}) + D_{j+1/2}(u_j - u_{j+1}) = 0, \quad (3.3)$$

is a positive coefficient scheme when all the nonlinear leading coefficients are non-negative

$$A_{i-1/2}, B_{i+1/2}, C_{j-1/2}, D_{j+1/2} \geq 0. \quad (3.4)$$

Lemma 3.1 (Jameson [7]). Semi-discrete positive coefficient type schemes (satisfying Definition 3.1) are local extrema diminishing [7]

Proof. If $u_{i,j} \geq u_{i+1}, u_{i-1}, u_{j+1}, u_{j-1}$, then $\partial_t u_{i,j} \leq 0$, if $u_{i,j} \leq u_{i+1}, u_{i-1}, u_{j+1}, u_{j-1}$, then $\partial_t u_{i,j} \geq 0$, local maxima do not increase and local minima do not decrease, this is the definition of Local Extrema Diminishing (LED). $\square$

Lemma 3.2 (Spekreijse [6]). When the forward Euler temporal discretisation of a positive coefficient type scheme satisfies the time step restriction,

$$\left(A_{i-1/2}^n + B_{i+1/2}^n + C_{j-1/2}^n + D_{j+1/2}^n \right) \leq 1, \quad (3.5)$$

then $u_{i,j}^{n+1}$ is a convex combination of $u_{i+1}^n, u_{i-1}^n, u_{j+1}^n, u_{j-1}^n$, trivially implying the following local discrete maximum principle with respect to edge sharing neighbours cell mean (or cell centre point-wise) values,

$$u^{n+1} \in [\min\{u_{i+1}^n, u_{i-1}^n, u_{j+1}^n, u_{j-1}^n\}, \max\{u_{i+1}^n, u_{i-1}^n, u_{j+1}^n, u_{j-1}^n\}]. \quad (3.6)$$

This in turn implies weaker properties such as boundedness $\|u^{n+1}\|_\infty \leq \|u^n\|_\infty$, and sign preservation.

Proof. The forward Euler discretisation of eq. (3.3) can be written as

$$u_{i,j}^{n+1} = (1 - [A_{i-1/2}^n + B_{i+1/2}^n + C_{j-1/2}^n + D_{j+1/2}^n])u_{i,j} + A_{i-1/2}^n u_{i-1} + B_{i+1/2}^n u_{i+1} + C_{j-1/2}^n u_{j-1} + D_{j+1/2}^n u_{j+1}. \quad (3.7)$$

Which is a convex combination of neighbouring cells under the condition eq. (3.5). $\square$

Remark. The above formulation of the maximum principle is pessimistic, in practice the upwind biasing imposes a more physically motivated maximum principle by setting some of the coefficients $A_{i-1/2}^n, B_{i+1/2}^n, C_{j-1/2}^n, D_{j+1/2}^n$ to zero depending on the direction of flow.

Theorem 3.1 (Incompressible flow). The Forward Euler discretisation of the numerical method described in section 2.2, can be written as convex combination of neighbour cell mean (or cell centre pointwise) values eq. (3.7) and as a result satisfies the discrete maximum principle eq. (3.6) when the Assumptions 3.1 hold.

Assumptions 3.1 (Theorem 3.1 assumptions).

1. The mesh scaled velocity satisfies a discrete divergence free condition

$$c_{i+1/2} - c_{i-1/2} + c_{j+1/2} - c_{j-1/2} = 0, \quad \forall (i, j). \quad (3.8)$$

2. When $\theta = 1$, the limiter function satisfies

$$\psi(R) \in [0, M_\psi], \quad \psi(S)/S \in [m_\psi, 2], \quad M_\psi \in [0, \infty), \quad m_\psi \in (-\infty, 0]. \quad (3.9)$$

When $\theta = 0$, the limiter function satisfies

$$\psi(1/R) \in [m_\psi, 2], \quad S\psi\left(\frac{1}{S}\right) \in [0, M_\psi], \quad M_\psi \in [0, \infty), \quad m_\psi \in (-\infty, 0]. \quad (3.10)$$

3. The timestep restriction

$$C := c_{i+1/2}^+ - c_{i-1/2}^- + c_{j+1/2}^+ - c_{j-1/2}^- \leq C_{FE} := \frac{2}{2 + M_\psi - m_\psi} \quad (3.11)$$

holds (written in terms of the Courant number).

Proof of Theorem 3.1. Expand the numerical method in terms of their positive and negative components, and taking away the divergence free condition gives

$$\begin{aligned} \frac{\partial u}{\partial t} + c_{i+0.5,j}^+(u_i^R - u_i) + c_{i+0.5}^-(u_{i+1}^L - u_i) - c_{i-0.5}^+(u_{i-1}^R - u_i) - c_{i-0.5}^-(u_i^L - u_i) \\ + c_{j+0.5}^+(u_j^U - u_i) + c_{j+0.5}^-(u_{j+1}^D - u_i) - c_{j-0.5}^+(u_{j-1}^U - u_i) - c_{j-0.5}^-(u_j^D - u_i) = 0. \end{aligned} \quad (3.12)$$

We have rearranged into the following positive coefficient representation only using the definition of R

$$\begin{aligned} \frac{\partial u}{\partial t} + \left(c_{i+1/2}^+ \left[\frac{\theta\psi(R_i)}{2} + \frac{(1-\theta)}{2} R_i \psi\left(\frac{1}{R_i}\right) \right] + c_{i-1/2}^+ \left[1 - \frac{\theta\psi(R_{i-1})}{2R_{i-1}} - \frac{(1-\theta)}{2} \psi\left(\frac{1}{R_{i-1}}\right) \right] \right) [u_i - u_{i-1}] \\ + \left(-c_{i+1/2}^- \left[1 - \frac{\theta}{2} \psi\left(\frac{1}{R_{i+1}}\right) R_{i+1} - \frac{1-\theta}{2} \psi(R_{i+1}) \right] - c_{i-1/2}^- \left[\frac{\theta}{2} \psi\left(\frac{1}{R_i}\right) + \frac{1-\theta}{2} \frac{\psi(R_i)}{R_i} \right] \right) [u_i - u_{i+1}] \\ + \text{y-direction}. \end{aligned} \quad (3.13)$$

When $\theta = 1$ the conditions

$$0 \leq \psi(R) \leq M_\psi, \quad m_\psi \leq \psi(S)/S \leq 2, \quad \forall R, S \in \mathbb{R}, \quad (3.14)$$

are sufficient for the following bounds

$$\begin{aligned} A_{i-1/2} &\in [0, c_{i+0.5}^+ M_\psi/2 + c_{i-0.5}^+(1 - m_\psi/2)], \\ B_{i+1/2} &\in [0, -c_{i+0.5}^- [1 - m_\psi/2] - c_{i-0.5}^- M_\psi/2], \\ C_{j-1/2} &\in [0, c_{j+0.5}^+ M_\psi/2 + c_{j-0.5}^+(1 - m_\psi/2)], \\ D_{j+1/2} &\in [0, -c_{j+0.5}^- [1 - m_\psi/2] - c_{j-0.5}^- M_\psi/2], \end{aligned} \quad (3.15)$$

for some $M_\psi \geq 0$, and $m_\psi \leq 2$. This shows the semi-discrete scheme is written in a positive coefficient representation sufficient for Lemmas 3.1 and 3.2. When $\theta = 0$, the similar conditions

$$m_\psi \leq \psi(1/R) \leq 2, \quad 0 \leq S\psi(1/S) \leq M_\psi, \quad (3.16)$$

are also sufficient for similar bounds. A sufficient timestep restriction $(1 - [A_{i-1/2}^n + B_{i+1/2}^n + C_{j-1/2}^n + D_{j+1/2}^n]) \leq 1$ for both representations to maintain a discrete maximum principle is $C \leq \frac{2}{2 + M_\psi - m_\psi}$, and appears in [4]. The conditions on the timestep in this proof are slightly more convenient than they are sharp, for additional details see Appendix A. $\square$

Remark. We have yet to put some additional sensible constraints on the limiter functions region. $m_\psi \leq 0$ is a design principle that should be adhered to, otherwise $\psi(S) \geq m_\psi S$, runs into conflict with the conditions $\psi(R) \leq M_\psi$, $\psi(R) \leq 0$, when M_ψ is finite and R is large. Good design requires passing through $\psi(1) = 1$ for second order accuracy, so that $M_\psi \in [1, \infty)$ is a sensible construction. We also impose $\psi(0) = 0$ as we have not found a case of a limiter function not satisfying this condition.

Previously we considered sufficient conditions on the limiter function for the forward Euler scheme to have a discrete maximum principle. The next theorem pertains to some necessary conditions on the limiter for the same forward Euler scheme to be positivity preserving.

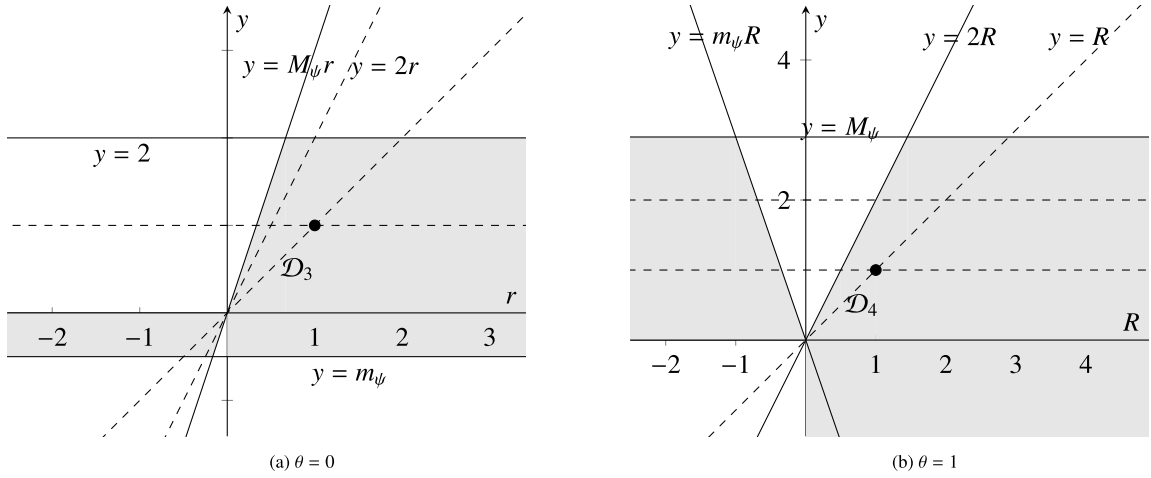


Fig. 3.1. It is necessary (but not proven sufficient), that limiters for incompressible flux form advection are contained in the above regions for positivity preservation. For $\theta = 0$, $\theta = 1$, respectively where $M_\psi \in [0, \infty]$, $m_\psi \in [-\infty, 0]$.

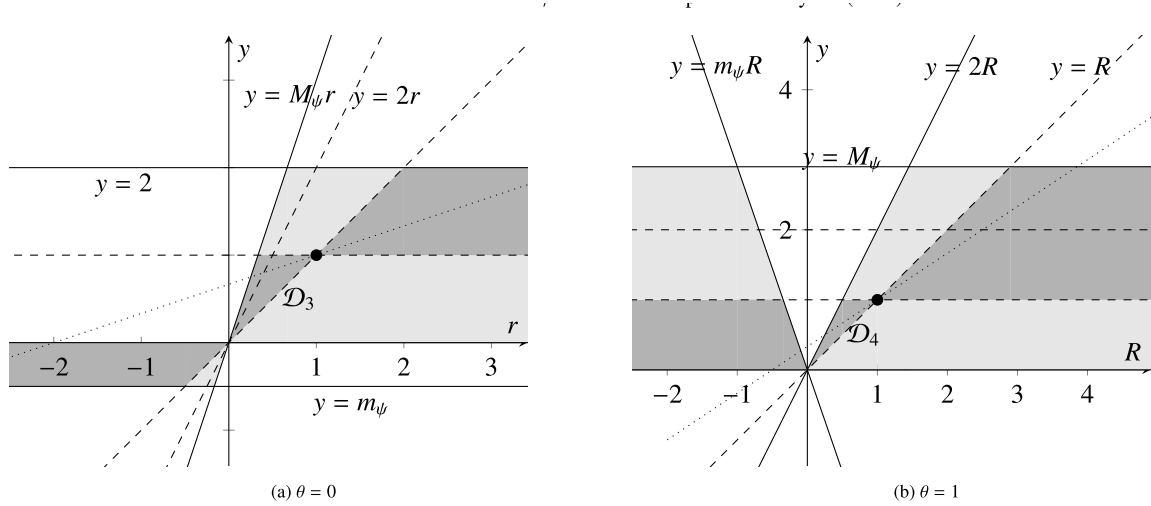


Fig. 3.2. New incompressible flow limiter regions, for $\theta = 0, 1$ respectively. Light gray regions are sufficient for a local maximum principle (bounded by solid lines), and dark gray indicates a desirable second order region (bounded by dashed lines). Light dotted line denotes a 3rd order region in the finite difference sense for uniform flow.

Theorem 3.2 (Necessary conditions). As can be shown through contradiction using case-by-case considerations of the velocity field in the forward Euler representation one can deduce that

$$\text{when } \theta = 1, \quad \psi(R) \geq 0, \quad \frac{\psi(S)}{S} \leq 2, \quad \forall R \in (-\infty, 0], \quad \forall S \in (0, \infty) \quad (3.17)$$

$$\text{when } \theta = 0, \quad \psi(T) \leq 2, \quad \frac{\psi(S)}{S} \geq 0 \quad \forall S \in (-\infty, 0], \quad \forall T \in (0, \infty) \quad (3.18)$$

are necessary conditions for the positivity preservation of the forward Euler scheme. These are plotted in Fig. 3.1.

Proof. Case-by-case considerations are presented in Appendix B. $\square$

Remark. In the $\theta = 1$ framework, the tail (values of ψ , for $R < 0$) has to be non-negative, in the $\theta = 0$ framework the tail (values of ψ , for $r < 0$) has to be non-positive. In the $\theta = 1$ framework in the positive region $R \geq 0$, we must have $\psi(R) \leq 2R$, and for $\theta = 0$ we must have $\psi(t) \leq 2$ in the positive region $t \geq 0$. Fig. 3.1 contains a diagram of some of the necessary conditions.

3.2. Applicability of Spekreijse's limiter region

In this subsection we will discuss why Spekreijse's theory may not always apply to flux form schemes. It is important to note that Spekreijse's theory is correct as a flux splitting framework for hyperbolic PDE's, and if you were to directly use Spekreijse's flux splitting method as defined in [6] on the advection equation, you would be using an advective form method, where velocity is located at the cell centres, (see Fig. 2.1b, eq. (2.10)). In this circumstance, not only will the extended limiter of Spekreijse provably work, but for the advection equation, the monotonicity criteria could be generalised further as the uniform boundedness condition requiring $\psi(R) - \psi(S)/S < 2M$ (eq(2.1) [6]) can be relaxed.

However, using an advective form scheme can lead to the loss of mass conservation for non-uniform flow. We instead investigate whether the positive coefficient argument applies to flux form methods in the multidimensional case where there is an additional dependence on an edge-defined velocity field (see Fig. 2.1a), as local mass conservation is essential for many applications. To use the positive coefficient lemma of Spekreijse, it is natural to write the flux form scheme in the following positive coefficient form

$$\frac{\partial u_{i,j}}{\partial t} + \frac{(u_i^R c_{i+1/2}^+ - u_{i-1}^R c_{i-1/2}^+) u_i^R - u_{i-1}^R}{(u_i^R - u_{i-1}^R)} (u_i - u_{i-1}) - \frac{(u_{i+1}^L c_{i+1/2}^- - u_i^L c_{i-1/2}^-) u_{i+1}^L - u_i^L}{(u_{i+1}^L - u_i^L)} (u_i - u_{i+1}) + y\text{-direction} = 0. \quad (3.19)$$

In order for this to be of positive coefficient type, we require that the leading terms (as defined in Definition 3.1) are positive. This is traditionally done [6] by the use of limiting to establish $(u_i^R - u_{i-1}^R)/(u_i - u_{i-1}) \geq 0$, then this is subsequently used in a mean value theorem on a numerical flux function (typically arising from the flux splitting framework [6]) on a flux f^+ with respect to u to establish $(f^+(u_i^R) - f^+(u_{i-1}^R))/(u_i^R - u_{i-1}^R) \geq 0$. However, we are not in the same flux splitting framework as the numerical flux function here has additional dependence on the velocity field and the mean value theorem won't necessarily apply to a term of the form $(u_i^R c_{i+1/2}^+ - u_{i-1}^R c_{i-1/2}^+)/(u_i^R - u_{i-1}^R)$. It is here one could say that the Spekreijse region is inappropriate for incompressible flow, by choosing an incompressible flow with $(u_i^R, c_{i+1/2}^+, u_{i-1}^R, c_{i-1/2}^+) = (1, 0, 0.5, 1)$ giving a negative leading coefficient. However this argument is only an indication that the theory doesn't hold, it is not sufficient, one would have to rule out the existence of a transform in which the scheme can be put in a different positive coefficient representation. Instead, we will use a counter example described in section 5.1, and rely on a numerical demonstration in Table 1, section 5.2.2, to show that this region is not monotone for the flux form scheme. Analytically, we also prove some additional necessary conditions are required of the limiter function to remain positivity preserving for incompressible flux form advection (in Theorem 3.2), limiters in the Spekreijse limiter region can violate these strictly necessary conditions.

This gives us the first contribution of this paper: the extended region of Spekreijse does not give schemes that have a discrete maximum principle or even positivity preserving when applied to the multidimensional flux form incompressible advection equation.

Remark. Despite this there are circumstances in which you can still use Spekreijse's limiter region. A sufficient condition for the leading coefficients in eq. (3.19) to be positive is that the x -component of velocity v_1 , is independent of x , so that $v_1 = v_1(y, t)$ and the y -component of velocity v_2 , is independent of y so that $v_2 = v_2(x, t)$. This implies that the flow is directionally constant $c_{i+1/2} = c_{i-1/2} = c_i$, $c_{j+1/2} = c_{j-1/2} = c_j$ property, and we can show that the leading terms will be positive directly

$$\frac{(u_i^R c_{i+1/2}^+ - u_{i-1}^R c_{i-1/2}^+)}{(u_i^R - u_{i-1}^R)} = c_i^+ \geq 0. \quad (3.20)$$

This can also be inferred from the mean value theorem, and more generally we will call a flow directional mean value theorem satisfying (MVTS), when one has sufficient conditions on the flow to directionally apply a mean value theorem and attain a positive coefficient representation of the numerical scheme. A flow without this property we will call mean value theorem violating (MVTV).

4. Other theoretical properties of the scheme

4.1. Temporal discretisation

The strong stability literature (e.g. [19–23]) allows boundedness and contractive behaviour in arbitrary seminorms of the semi-discrete/(forward Euler scheme) scheme to be directly translated into a wide variety of SSP methods, allowing the order of the scheme to be increased. The local maximum principle is not preserved in exactly the same way, we actually ensure that each substage satisfies a local maximum principle with respect to the previous substage, this is apparent from the Shu Osher representation, see Appendix E, for a more detailed explanation. We can recall the forward Euler numerical flow map $u^{n+1} = \text{Forward Euler}(u^n, c^n)$, defined as

$$u_{i,j}^{n+1} = u_{i,j}^n - [F_{i+0.5}(u_i^R, u_{i+1}^L, c_{i+0.5}^n) - F_{i-0.5}(u_{i-1}^R, u_i^L, c_{i-0.5}^n)] - [F_{j+0.5}(u_{j-1}^U, u_j^D, c_{j+0.5}^n) - F_{j-0.5}(u_{j-1}^U, u_j^D, c_{j-0.5}^n)]. \quad (4.1)$$

Where the time step, Δt , and mesh spacing, Δx , have been absorbed into the face defined velocity c . The SSP33 Runge Kutta scheme, will be used in the numerical experiments and can be implemented in the following memory efficient (2 register) Shu Osher representation

$$k^1 = \text{Forward Euler}(u^n, c^n), \quad (4.2)$$

$$k^2 = 3/4 \cdot u^n + 1/4 \cdot \text{Forward Euler}(k^1, c^{n+1}), \quad (4.3)$$

$$u^{n+1} = 1/3 \cdot u^n + 2/3 \cdot \text{Forward Euler}(k^2, c^{n+1/2}). \quad (4.4)$$

This scheme is a third order, three stage Runge Kutta method with radius of monotonicity of 1. This method preserves convex seminorms (such as $\|\cdot\|_\infty$) under the same timestep restriction forward Euler does. In [11] it is remarked that for small enough Courant numbers the semi-discrete scheme in this paper is monotone when discretised with RK4. The strong stability literature has since shown that the RK4 method has zero radius of monotonicity, in Appendix F we remark on this particular choice of timestepping method.

4.2. Accuracy

The numerical method viewed as a finite volume method approximating the integral form of the equation is second order or less in dimensions two or greater, and third order or less when in one dimension. The numerical method as viewed as a finite difference method approximating the point valued equation is second order, but can be third order for uniform flow. See Appendix G for a more detailed discussion. Spekreijse, showed that sufficient conditions on the limiter function $\psi(R)$ for second order accuracy are $\psi(1) = 1$, $\psi \in C^2$ near 1, [6]. Sufficient and necessary conditions on the limiter function $\psi(r)$ were shown by Hua-mo [14] to be $\psi(1) = 1$, and ψ is Lipschitz continuous. More specific conditions on accuracy of finite difference accuracy of uniform flow in the neighbourhood of extrema can be found in [14]. See Appendix H for a more detailed discussion. It is also often mistaken that TVD schemes are first order at noncritical extrema because of a slight oversight in the truncation analysis in Osher [24]. This is addressed by Hua-mo for $\theta = 1$ [14], but often attributed to [10], who make similar mistakes in the truncation error analysis as Osher, for $\theta = 0$. See Appendix H for a more detailed discussion on this.

4.3. Linear invariance

Linear invariance implies constants are preserved under incompressible flow and essential in atmospheric advection [25] [26]. In [8] it is incorrectly stated that the scheme is not linear invariant whether in state or flux interpolated form, the state interpolated scheme in this paper is linear invariant for incompressible flow after one takes away a discrete incompressibility condition, see Appendix C.

4.4. Symmetric limiters: old and new

We first introduce some important symmetric flux limiter functions in Spekreijse admissible region, but not in the Sweby admissible region

$$\text{van-Albada}(R) = \frac{R^2 + R}{R^2 + 1}, \quad (4.5)$$

$$\text{Ospre}(R) = \frac{3}{2} \frac{R^2 + R}{R^2 + R + 1}, \quad (4.6)$$

$$\text{ENO2}(R) = \begin{cases} R & \text{where } |R| \leq 1, \\ 1 & \text{where } |R| \geq 1. \end{cases} \quad (4.7)$$

These are introduced in the respective papers [27–29] and are plotted in Fig. 4.1a. We also introduce the subscript P limiters, as those same limiters who are “pushed” into being in the Sweby diagram by the removal of the tail, these limiters are designed keep the symmetry property and fix the monotonicity problems associated with using Spekreijse’s region for multidimensional flux form incompressible flow.

$$\text{van-Albada}_P(R) = \begin{cases} \frac{R^2 + R}{R^2 + 1}, & R \geq 0 \\ 0, & R < 0 \end{cases}, \quad (4.8)$$

$$\text{Ospre}_P(R) = \begin{cases} \frac{3}{2} \frac{R^2 + R}{R^2 + R + 1}, & R \geq 0 \\ 0, & R < 0 \end{cases}, \quad (4.9)$$

$$\text{ENO}_P(R) = \text{minmod}(R) = \max(0, \min(R, 1)). \quad (4.10)$$

4.5. Non symmetric limiters and symmetry breaking

We also introduce the Koren [11] limiter,

$$\text{Koren}(R) = \max(0, \min(2, 2R, 2R/3 + 1/3)). \quad (4.11)$$

It is not symmetric, but is an accurate limiter consisting of restricting the third order upwind region to the Sweby region. We introduce a new limiter with free parameters M, m defined by

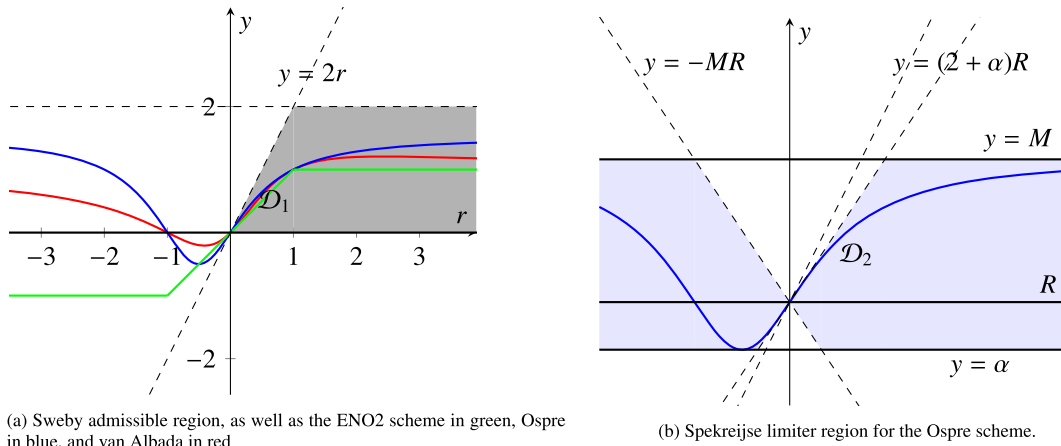


Fig. 4.1. D_1 is the Sweby region, D_2 , is the Spekreijse region, which has two free parameters $\alpha \in [-\infty, 0]$, $M \in (0, \infty)$. We can see that the limiters, van Albada (red) eq. (4.5), Ospre (blue) eq. (4.6), and ENO2 (green) eq. (4.7) are not contained in the Sweby region D_1 (or even the new incompressible flow limiter region D_4), but are in the Spekreijse region D_2 with values for $[M, \alpha]$ given by $[1.5, -0.5]$, $[1/2(1 + \sqrt{2}), 1/2(1 - \sqrt{2})]$, $[1, -1]$ respectively. Only the Ospre scheme and its Spekreijse Region are shown in Fig. 4.1b. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

$$\text{Woodfield}(R, M, m) = \begin{cases} \{R \leq -1 : 0\} \\ \{-1 < R < -\frac{1}{2} : 0\} \\ \{-\frac{1}{2} < R < -\frac{1}{-3m-2} : \frac{2}{3}R + \frac{1}{3}\} \\ \{\frac{1}{-3m-2} < R < 0 : m_\psi R\} \\ \{0 < R < \frac{1}{4} : 2R\} \\ \{\frac{1}{4} < R < \frac{3M-1}{2} : \frac{2}{3}R + \frac{1}{3}\} \\ \{\frac{3M-1}{2} < R : M\} \end{cases} \quad (4.12)$$

consisting of restricting the third order upwind region to the new limiter region. We also introduce the SuperbeeR limiter,

$$\text{SuperbeeR}(R, M, m) = \begin{cases} \max(0, \max(\min(2R, 1), \min(R, M))) & \text{where } R \geq 0 \\ \min(mR, 1) & \text{where } R < 0 \end{cases} \quad (4.13)$$

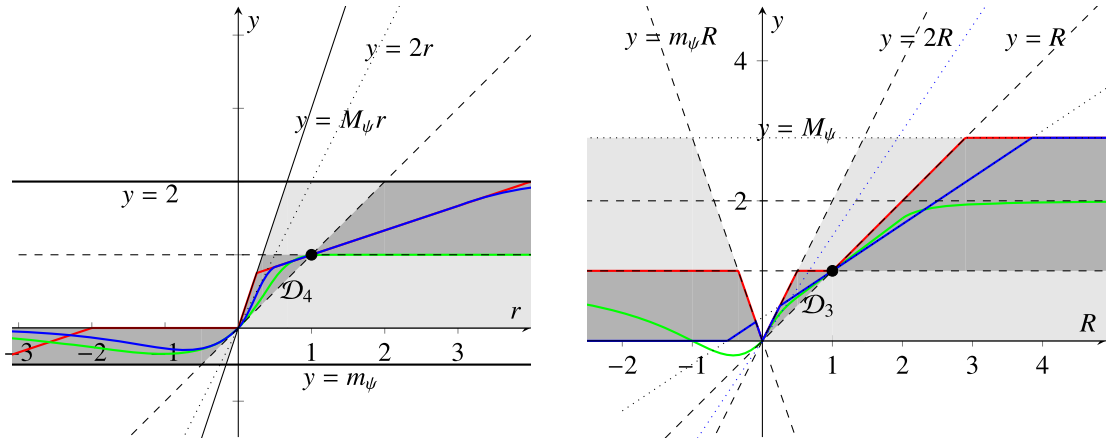
which serves as an extension of traditional Superbee limiter to the new limiter region. This is to test our limiter region but could have application to front tracking. These limiters are plotted in Fig. 4.2b. We define a similar extension to the Koren limiter by restricting the third order upwind region to the $\theta = 1$ maximum principle limiter region for some M, m creating the Woodfield(r, M, m) limiter, plotted in Fig. 4.2a.

In the review paper [30] it is argued on practical experience that using regions in which $R < 0$ makes little difference, a similar sentiment is also found in the review [31]. For example, in [30] a limiter called ultimate-TCDF (Third Order Continuously Differentiable) is introduced, it is defined as

$$\text{UTCDF}(R) = \begin{cases} \frac{R(R+1)}{(R^2+1)}, & \text{where } R < 0, \\ R^3 - 2R^2 + 2R, & 0 \leq R \leq 1/2, \\ 0.75R + 0.25, & (1/2 \leq R < 2), \\ \frac{(2R^2 - 2R - 9/4)}{(R^2 - R - 1)}, & R \geq 2. \end{cases} \quad (4.14)$$

First note the UTCDF limiter is not differentiable at $R = 0$, and more importantly, has no theoretical guarantees. It is not contained within any Spekreijse limiter region,¹ it also leaves the newly defined $\theta = 1$ limiter region, as seen in green in Fig. 4.2. We conjecture that leaving the Spekreijse region may cause problems in practice (fail to be monotonic), in the context of directionally constant flow, as well as flux splitting frameworks. We also conjecture that leaving the newly defined limiter region $\theta = 1$ will cause problems in practice for incompressible flux form advection, in light of Theorem 3.2.

¹ This can be verified since the tail $R < 0$ is identical to the van-Albada limiter so that $\alpha = -0.5$. However, one can see that at $R = 0$, the gradient is $2 > 2 - 0.5$, and for $R \in (0, 0.2928)$ this limiter is not contained within a Spekreijse region.



(a) The new Differentiable(r) limiter in blue, the analytic limiter $\tanh(r)\exp(r)$ in green, and the Woodfield($r, M = 3, -m = -1$) limiter in red.

(b) The Woodfield($R, M = 2.9, m = 2.9$) limiter in blue, the Superbee($R, M = 2.9, -m = -2.9$) limiter in red, and the Ultimate TCDF(R) limiter [30] in green.

Fig. 4.2. New limiters in their respective $\theta = 0, 1$ regions.

Instead one can design limiter functions using the Spekreijse limiter region when in a flux-splitting framework (or when a mean value theorem applies) or using the new $\theta = 0, \theta = 1$ limiter regions in the case of incompressible flux form advection. For example we define the following two limiter functions

$$\text{UTCDF}_P(R) = \begin{cases} 0 & \text{where } R < 0 \\ R^3 - 2R^2 + 2R, & 0 \leq R \leq 1/2, \\ 0.75R + 0.25, & (1/2 \leq R < 2), \\ \frac{(2R^2 - 2R - 9/4)}{(R^2 - R - 1)}, & R \geq 2. \end{cases} \quad (4.15)$$

$$\text{UTCDF}_S(R) = \begin{cases} \frac{R(R+1)}{(R^2+1)}, & \text{where } R < -1, \\ 0 & \text{where } -1 \leq R < 0 \\ R^3 - 2R^2 + 2R, & 0 \leq R \leq 1/2, \\ 0.75R + 0.25, & (1/2 \leq R < 2), \\ \frac{(2R^2 - 2R - 9/4)}{(R^2 - R - 1)}, & R \geq 2. \end{cases} \quad (4.16)$$

The $\text{UTCDF}_P(R)$ limiter is the Ultimate-TCDF restricted to the Sweby region. The $\text{UTCDF}_S(R)$ limiter is formed by restricting the Ultimate-TCDF to the new $\theta = 1$ limiter region. For differentiable limiters suitable for flux splitting, one can design limiters in the Spekreijse region e.g. eq. (4.6). When $\theta = 1$, there are no globally differentiable limiter functions contained within the second order limiter region, that remain suitable for incompressible flow without a mean value theorem assumption of the flow. We introduce the first globally differentiable limiter function contained entirely within the second order region, suitable for incompressible flow by using our new $\theta = 0$ region, it touches the third order region for accuracy.

$$\text{Differentiable}(r) = \begin{cases} \tanh(r)\exp(r), & \text{where } r \leq 0, \\ -8r^3 + 16/3r^2 + r, & 0 < r \leq 1/2 \\ 1/3r + 2/3, & 1/2 < r \leq 3 \\ 1/3 \tanh(r-3) + 5/3, & r > 3 \end{cases} \quad (4.17)$$

plotted in Fig. 4.2a. This limiter is a placeholder that could be improved on/replaced by piecewise polynomials with the same properties.

5. Numerical demonstrations: test setup and results

We have two distinct types of flow (Mean value theorem satisfying (MVTs), and Mean value theorem violating (MVTv)), three different limiter regions (Spekreijse, Sweby, and our new limiter region(s)). We introduce two monotonicity tests:

- Solid body rotation of the LeVeque initial conditions in which a directional mean value theorem holds and all limiter functions should remain monotone [6].

- A sinusoidal deformational flow of the LeVeque initial conditions which is directional mean value theorem violating, used to verify that the Spekreijse limiter region is not monotone for general incompressible flow, but the Sweby region and the new limiter regions are. We also introduce another deformational vector field eq. (5.3) with the same aim.

5.1. Setup: monotonicity tests

The numerical domain is $\Omega = (0, 1] \times (0, 1]$ discretised by 128×128 cells, with periodic boundary conditions. We run the scheme with 4096 timesteps until $T_{max} = 1$ resulting in a maximum cell defined Courant number around 0.2. We locate the stream function $\Psi(x, y)$ at the cell vertices and use a discrete form of the curl operator to create a divergence free vector field (to machine precision) with velocities located at the midpoints of faces, with only normal components. We use the solid body rotation and the sinusoidal deformation defined from the stream functions

$$\Psi = -\pi \left((x - 1/2)^2 + (y - 1/2)^2 \right), \quad (5.1)$$

$$\Psi = \frac{1}{2} \sin(\pi x) \sin(\pi y) \cos(2\pi t/T_{max}), \quad (5.2)$$

$$\Psi = 1/16 \sin(32\pi x) \sin(32\pi y) \cos(2\pi t/T_{max}). \quad (5.3)$$

The solid body rotation flow eq. (5.1) is directional mean value theorem satisfying (with the directionally constant property), the sinusoidal deformation eq. (5.2) does not have a directional mean value property. Equation (5.3) is similar to eq. (5.2), it does not have a mean value theorem, it is used to create larger magnitude monotonicity failures. For the initial condition of the tracer, we use the LeVeque initial conditions [32], defined below

$$u_0 = \begin{cases} 1 & \text{where } \sqrt{(x-0.5)^2 + (y-0.75)^2} \leq 0.15, \text{ and } (x \leq 0.475), \\ 1 & \text{where } \sqrt{(x-0.5)^2 + (y-0.75)^2} \leq 0.15, \text{ and } (x > 0.525), \\ 1 & \text{where } \sqrt{(x-0.5)^2 + (y-0.75)^2} \leq 0.15, (y \geq 0.85), \text{ and } (0.475 < x \leq 0.525), \\ (1 - \frac{R_{cone}}{0.15}) & \text{where } (R_{cone} = \sqrt{(x-0.5)^2 + (y-0.25)^2} \leq 0.15), \\ \frac{1}{2}(1 + \cos(\pi \frac{R_{cos}}{0.15})) & \text{where } (R_{cos} = \sqrt{(x-0.25)^2 + (y-0.5)^2} \leq 0.15), \\ 0 & \text{else.} \end{cases} \quad (5.4)$$

5.2. Numerical results

Section 5.2.1 will show snapshots of the solutions. Positivity will be demonstrated in section 5.2.2 and error norms are presented in section 5.2.3.

5.2.1. Snapshots

Fig. 5.2 and Fig. 5.1 contain the final tracer values of the LeVeque initial conditions after the solid body rotation and the time reversing sinusoidal deformational flow respectively. Fig. 5.3 and Fig. 5.4, contain the horizontal cross section through the final tracer value of the middle of the three advected shapes, for the limiters (eqs. (4.5) to (4.10)) for the MVTs solid body rotation and the MVTV sinusoidal flow respectively. Fig. 5.5 and Fig. 5.6, contains the horizontal cross section through the final tracer value of the middle of the three advected shapes, for the other limiters including (eqs. (4.11), (4.13) and (4.17)), for the MVTs solid body rotation and the MVTV sinusoidal flow respectively.

The final tracer values of the LeVeque initial conditions after the sinusoidal deformational time reversing flow plotted in Fig. 5.1, contain negative values in the final time step for the schemes SSP33:ENO2(R) (fig c), SSP33:Ospre(R) (fig g) and SSP33:VanAlbada(R) (fig k). This observation is predicted (and proven in Theorem 3.2) in section 3.2 regarding the applicability of Spekreijse's limiter region to flux form incompressible advection. The RK4:Koren(R) generated a 10^{-10} negative value over the integration period speculated as caused by the temporal integration as discussed in Appendix F. One observes sharper edges on the slotted cylinder when comparing SSP33:Superbee(R, M, m) (fig i) to SSP33:Superbee(R) (fig h). One observes SSP33:Differentiable(R) (fig l) is more visibly similar to the RK4:Koren(R) scheme, than the other differentiable limiters SSP33:Ospre(R) (fig g), SSP33:VanAlbada(R) (fig k) which are seen to be successively more diffusive than the SSP33:Differentiable(R) scheme, possibly due to the third order region being attained by this limiter as remarked on in eq. (G.3).

The final tracer values of the LeVeque initial conditions after the solid body rotation of Fig. 5.2 using the SSP33 timestepping scheme, contain no significant negative values in the final time step, this aligns with the theoretical predictions when combining by the MVTs property of the flow discussed in section 3.2 and the SSP properties discussed in section 4.1 and Appendix E. The RK4:Koren(R) scheme generated a negative over the integration period as discussed in Appendix F. One observes sharper edges on the slotted cylinder when comparing SSP33:Superbee(R, M, m) (fig i) to SSP33:Superbee(R) (fig h). One observes the SSP33:Differentiable(R) (fig l) limiter is visibly more similar to the RK4:Koren(R) scheme, than the other differentiable limiters SSP33:Ospre(R) (fig g), SSP33:VanAlbada(R) (fig k) which are successively more diffusive.

Fig. 5.3 and Fig. 5.4, indicate that the push of the Ospre, VanAlbada and Eno2 limiters into the Sweby region had little effect on the accuracy of these schemes in the regimes of both sharp and smooth tracer values (for the timescales and flows in these particular experiments). The horizontal cross sections in Fig. 5.5 and Fig. 5.6 one observes that both the Superbee and the Superbee($R, 3, -1$)

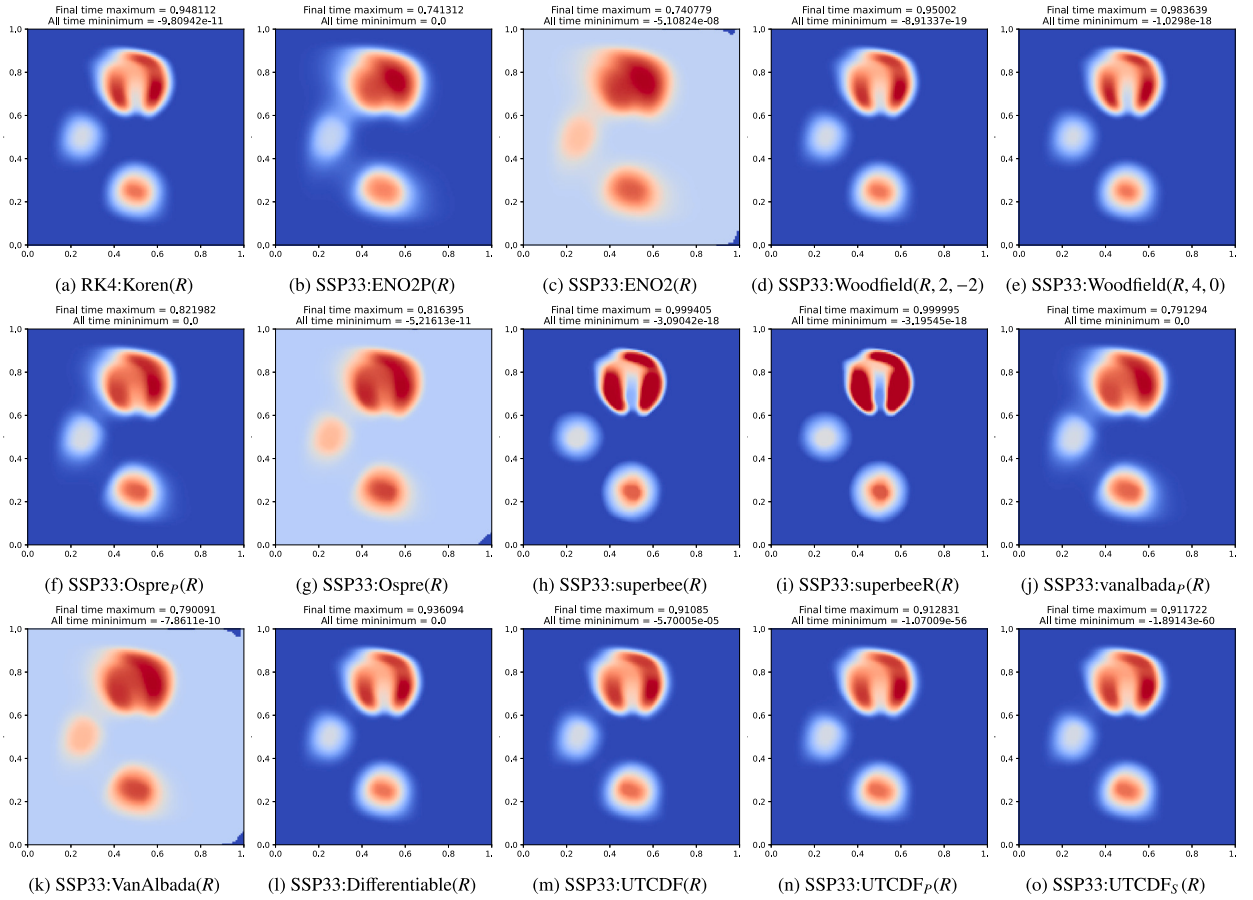


Fig. 5.1. Final state of the time reversing sinusoidal flow of LeVeque initial conditions, any negative values below negative -1×10^{-14} will be plotted as -0.5 and will shift the entire colour range, so that the colour scheme highlights negatives should they appear. Final time maximum and all time minimum values are displayed on the top of each figure.

have unphysical squaring of the cone and cosine lumps peak and more accurate representation of the discontinuous slotted cylinder profile. The Superbee(R , M , m) limiter appears more compressive than Superbee. The horizontal cross sections in Fig. 5.5 and Fig. 5.6 indicate that the Differentiable limiter performs similar to the Koren(R) limiter with a very small increase in diffusion due to the smoothing. The Woodfield(R , 2, -2) performed similar to the Koren(R) limiter, indicating that in the negative gradient region in the $\theta = 1$ region Fig. 3.2 did not provide significant increase in accuracy under these test cases. The Woodfield(R , 4, 0) performed more accurately than the Koren(R) limiter, indicating additional accuracy in the $\theta = 1$ region can be attained using the positive gradient region Fig. 3.2.

5.2.2. Positivity

Table 1 contains the minimum values $\min_{n,i,j} u_{i,j}^n$ attained for both the MVTV sinusoidal time reversing deformational flow eq. (5.2) as well as the MVTs solid body rotation eq. (5.1) of the LeVeque initial conditions eq. (5.4), for all limiters introduced in this paper. The main conclusion of Table 1 is that all limiters not strictly contained in the newly defined limiter regions Fig. 3.2 have not remained positive for the sinusoidal MVTV flow. In particular, the Ospre(R), VanAlbada(R), and ENO2(R), have positivity violations for deformational flow, demonstrating that the Spekreijse limiter region leads to schemes that are not bounded for the flux form advection equation. This is observed in sub-figures (b,g,k) of Fig. 5.1.

Also observed in Table 1 and Fig. 5.1; limiters pushed into the Sweby region Ospre(R), VanAlbada(R), and ENO2(R) remained strictly positive. The new limiters (Woodfield(R , M , m), Differentiable(R), Superbee(R , M , m)) not contained in the Sweby region but contained in the newly defined limiter region, also remained strictly positive. RK4 is not positivity preserving. UTCDF(R) is not positivity preserving, but UTCDFs(R), UTCDFp(R) are. In the second column of Table 1, limiters in Spekreijse's limiter region remained positive for the mean value theorem satisfying flow, as predicted in section 3.2. All results in Table 1 agree with the theoretical contributions and predictions made in this paper up to machine precision. Column 3, is a different mean value theorem satisfying flow, for which significantly larger positivity violations are observed to emphasise the possibility of larger negatives.

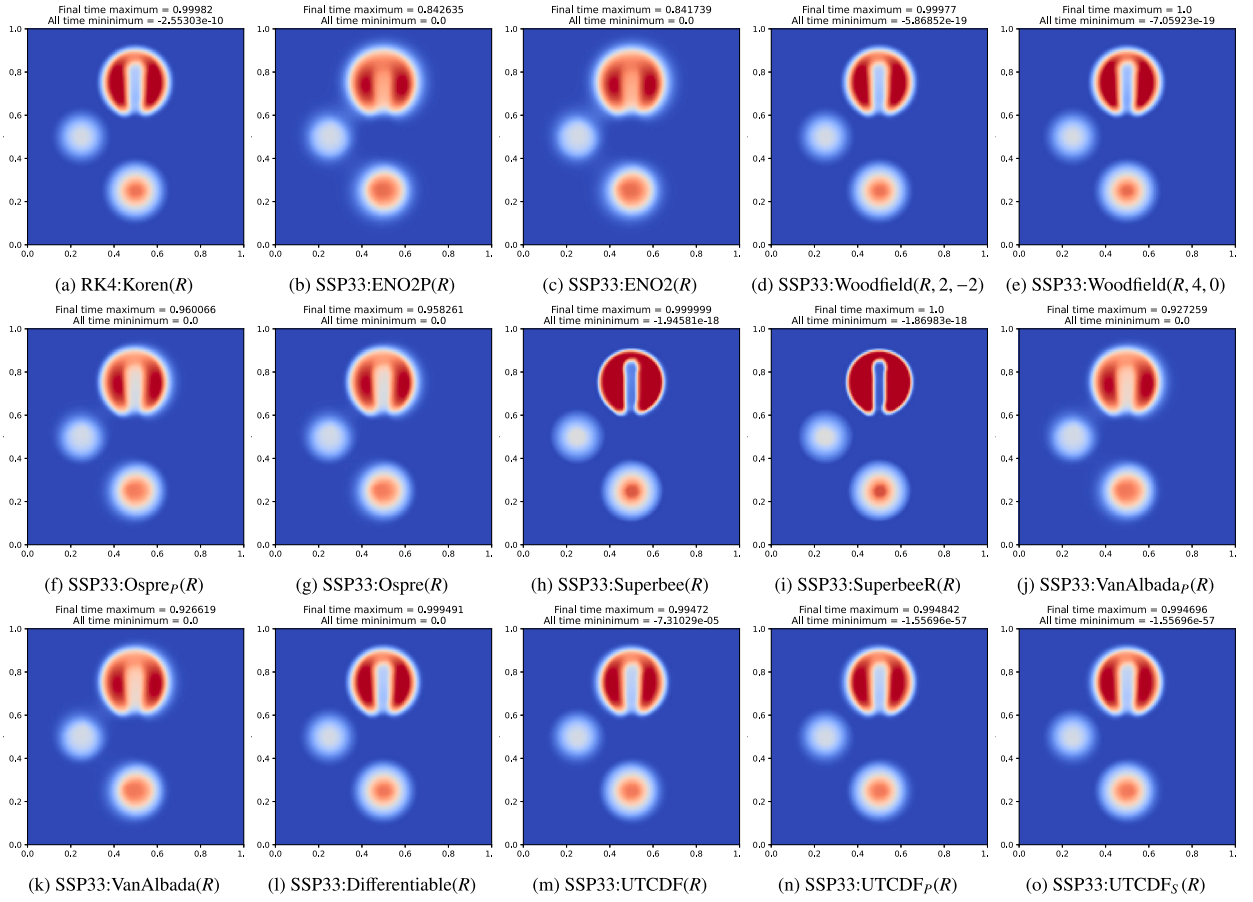


Fig. 5.2. Final state of the solid body rotation of LeVeque initial conditions, any negative values below negative -1×10^{-14} will be plotted as if -0.5 and will shift the entire colour range, so that the colour scheme highlights significant negatives at the final timestep should they appear. The Final time maximum $\max_{i,j} u^n$ and all time minimum values $\max_{i,j,k} u_{i,j}^k$ are displayed on the top of each figure.

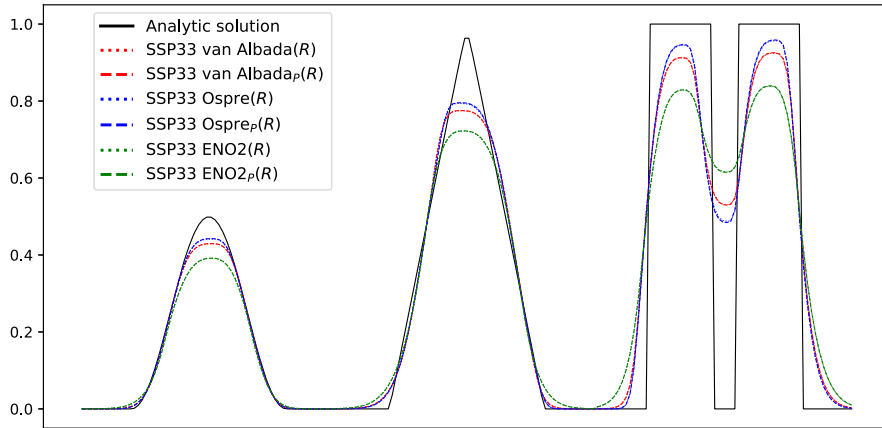


Fig. 5.3. Symmetric Spekreijse limiters and their Symmetric push into the Sweby region. We use the horizontal cross sections through the middle of the three LeVeque shapes; cosine, cone and the Zalesak slotted cylinder after solid body rotation. Conclusion, all the pushed limiters are almost indistinguishable to the eye from the original limiters.

5.2.3. Accuracy

We compute the relative l_2 error norm $re_2(u, \Omega) := \|u - u_e\|_{L^2(\Omega)}^2 / \|u_e\|_{L^2(\Omega)}^2$ in subdomains of the Zalesak slotted cylinder $\Omega_{zal} = \{(y, x) \in [5/8, 7/8] \times [3/8, 5/8]\}$, the Cone $\Omega_{cone} = \{(y, x) \in [3/8, 5/8] \times [1/8, 3/8]\}$ and the Cosine lump $\Omega_{cos} = \{(y, x) \in [3/8, 5/8] \times [1/8, 3/8]\}$. Table 2 and Table 3 show the relative error norms at the final timestep in the regions $\Omega_{zal}, \Omega_{cone}, \Omega_{cos}$, for the solid body

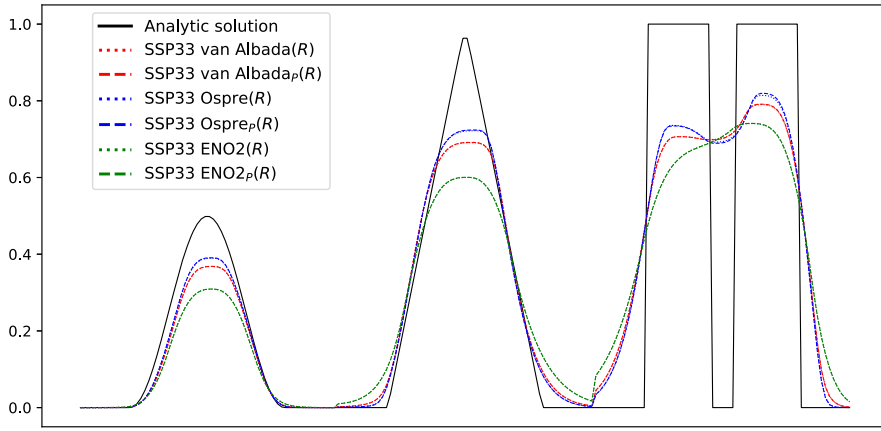


Fig. 5.4. Symmetric Spekreijse limiters and their Symmetric push into the Sweby region. We use the horizontal cross sections through the middle of the three LeVeque shapes; cosine, cone and the Zalesak slotted cylinder after the time reversing sinusoidal flow. Conclusion, all the pushed limiters are almost indistinguishable to the eye from the original limiters.

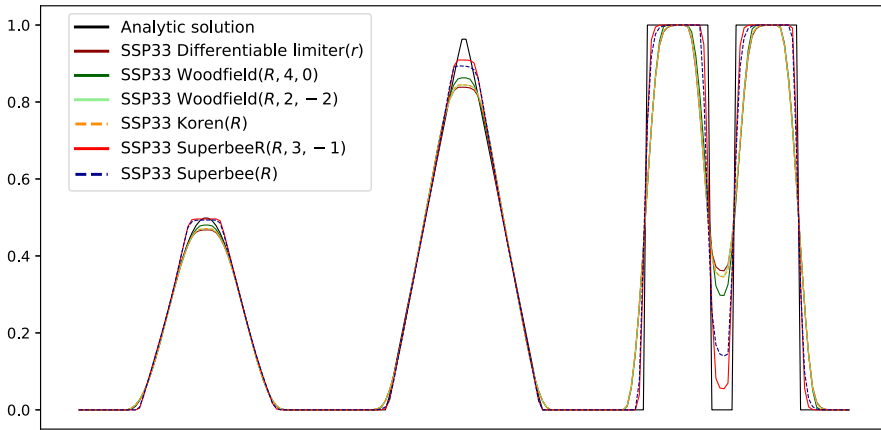


Fig. 5.5. Horizontal cross sections through the middle of the three LeVeque shapes; cosine, cone and the Zalesak slotted cylinder after the solid body rotation. The new differentiable limiter has results very similar to the Koren limiter. The differentiable limiter is more accurate than the Ospre limiter and is suitable for incompressible flow. The Woodfield(R,2,-2) limiter has similar accuracy to the Koren scheme, the Woodfield(R,4,0) limiter is more accurate than the Koren limiter, however it must be run at a reduced timestep. The SuperbeeR(R, M, m) limiter is more compressive than the Superbee(R) limiter.

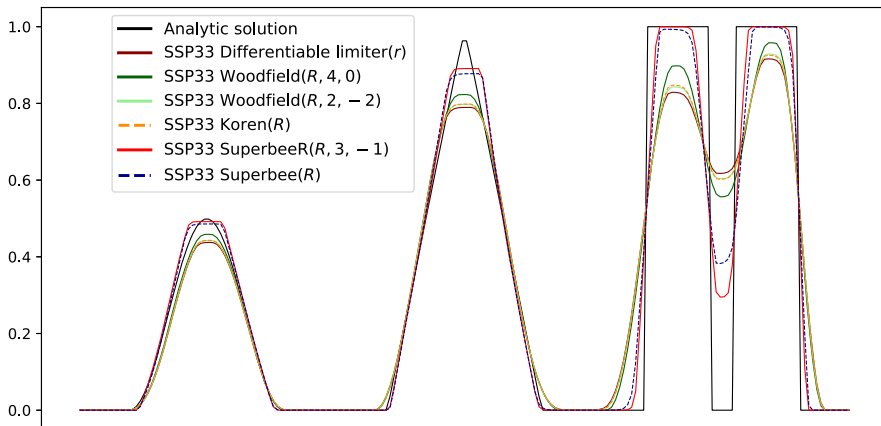


Fig. 5.6. Horizontal cross sections through the middle of the three LeVeque shapes; cosine, cone and the Zalesak slotted cylinder after the time reversing sinusoidal flow. The new differentiable limiter has results very similar to the Koren limiter. The differentiable limiter is more accurate than the Ospre limiter and is suitable for incompressible flow. The Woodfield(R,2,-2) limiter has similar accuracy to the Koren scheme, the Woodfield(R,4,0) limiter is more accurate than the Koren limiter, however it must be run at a reduced timestep. The SuperbeeR(R, M, m) limiter is more compressive than the Superbee(R) limiter.

Table 1

The minimum values attained in the full run of the experiments, bold indicates significant positivity violation. The Mean Value Theorem Violating sinusoidal flow (MVTV) is in the first and third column, the Mean Value Theorem Satisfying (MVTs) rotational flow is in the second column. Only limiters not strictly contained in the new limiter regions Fig. 3.2 are not positivity preserving for the sinusoidal flow. The RK4 scheme and the UTCDF limiter are not positivity preserving.

Scheme	Limiter	MVTV-sin $\min_{\psi_{n,i,j}} u_{i,j}^n$	MVTs-sbr $\min_{\psi_{n,i,j}} u_{i,j}^n$	MVTV-sin 32 $\min_{\psi_{n,i,j}} u_{i,j}^n$
SSP33	ENO2(R)	-5.10824e - 08	0.0	-0.0128304
SSP33	Ospre(R)	-5.21613e - 11	0.0	-0.0450631
SSP33	Vanalbada(R)	-7.8611e - 10	0.0	-1.08423e - 07
SSP33	ENO2 _p (R)	0.0	0.0	0.0
SSP33	Ospre _p (R)	0.0	0.0	0.0
SSP33	Vanalbada _p (R)	0.0	0.0	0.0
SSP33	UTCDF _p (R)	-1.07009e-56	-1.55696e-57	-7.39459e-34
SSP33	Koren(R)	-8.73719e-19	-4.61246e-19	-3.62442e-19
SSP33	Superbee(R)	-3.09042e-18	-1.94581e-18	-1.97676e-18
SSP33	Woodfield(R, 2, -1)	-8.91337e-19	-5.86852e-19	-4.15217e-19
SSP33	Woodfield(R, 4, 0)	-1.0298e-18	-7.05923e-19	-4.22583e-19
SSP33	Differentiable(r)	0.0	0.0	0.0
SSP33	SuperbeeR(R, 3, -1)	-3.19545e-18	-1.86983e-18	-1.8609e-18
SSP33	UTCDF _s (R)	-1.89143e-60	-1.55696e-57	-8.68458e-34
RK4	Koren(R)	-9.80942e - 11	-2.55303e - 10	-3.72077e-19
SSP33	UTCDF(R)	-5.70005e - 05	-7.31029e - 05	-0.000140264

Table 2

Relative L_2 error in different regimes of solid body rotation.

Scheme	Slotted Cylinder $re_2(u, \Omega_{z,sl})$	Cone $re_2(u, \Omega_{cone})$	Cosine bell $re_2(u, \Omega_{cos})$
SSP33 Ospre _p (R)	0.351102	0.0673095	0.0658036
SSP33 Ospre(R)	0.351987	0.0676014	0.0662881
SSP33 van Albada _p (R)	0.371091	0.083057	0.076212
SSP33 van Albada(R)	0.371456	0.0832153	0.0763878
SSP33 ENO2 _p (R)	0.422668	0.148316	0.106093
SSP33 ENO2(R)	0.422894	0.148404	0.106193
SSP33 Woodfield(R, 4, 0)	0.274023	0.01206	0.0286216
SSP33 Koren(R)	0.289004	0.0174684	0.0338112
SSP33 Woodfield(R, 2, -2)	0.290484	0.0181157	0.0342527
SSP33 Differentiable limiter (r)	0.295057	0.0200713	0.0360145
SSP33 SuperbeeR(R, 3, -1)	0.168015	0.0496325	0.0341482
SSP33 Superbee(R)	0.197235	0.0474037	0.0365515

rotation and the sinusoidal flow respectively. The results indicate that the SuperbeeR(R, 3, -1) limiter has the smallest L_2 error in the region of the slotted cylinder, also observed in the horizontal cross section Figs. 5.5 and 5.6. The Woodfield(R, 4, 0) limiter has the smallest L_2 error in the region of the cone, and the cosine bump. The Woodfield(R, 4, 0) scheme was more accurate than Koren limiter in all regions, which in turn was more accurate than the Woodfield(R, 2, -2) limiter in all regions. The Differentiable limiter was more accurate than the Ospre limiter which in turn was more accurate than the VanAlbada limiter observed in Figs. 5.3 to 5.6. Finally pushing the Ospre, VanAlbada and ENO2 limiters into the Sweby region, did not significantly decrease the relative error norm in any particular region, also observed in Figs. 5.3 and 5.4. This is interesting in light of the theoretical considerations in Appendix H since the “pushed” schemes are reduced to the first order upwind scheme at extrema, in particular the ENO2_p(R) no longer satisfies the second order at extrema condition [14] whilst ENO2(R) does.

6. Summary and conclusions

The main contribution of this paper is the derivation of two new limiter regions Fig. 3.2, larger than Sweby’s admissible limiter region and smaller than the Spekreijse limiter region, sufficient for multidimensional incompressible flow to maintain a discrete local maximum principle (Theorem 3.1). We show that the Spekreijse limiter region leads to flux form schemes that are not bounded for the advection equation unless a directional mean value theorem can be proven for each direction, this is demonstrated with a numerical example in Table 1 and by strictly necessary constraints in Theorem 3.2.

In the new limiter regions, we have introduced; the first globally differentiable limiter entirely contained within the second order region capable of being monotone for flux form advection eq. (4.17), a limiter eq. (4.12) that can be more accurate than the Koren

Table 3
Relative L_2 error in different regimes for sinusoidal deformation.

Scheme	Slotted Cylinder $re_2(u, \Omega_{\text{sl}})$	Cone $re_2(u, \Omega_{\text{cone}})$	Cosine bell $re_2(u, \Omega_{\text{cos}})$
SSP33 Ospre _p (R)	0.446784	0.177647	0.129118
SSP33 Ospre(R)	0.447811	0.179705	0.13045
SSP33 van Albada _p (R)	0.463153	0.219316	0.152391
SSP33 van Albada(R)	0.463447	0.220061	0.152813
SSP33 ENO2 _p (R)	0.501686	0.34828	0.234589
SSP33 ENO2(R)	0.501786	0.348622	0.234761
SSP33 Woodfield($R, 4, 0$)	0.373459	0.0762376	0.0640243
SSP33 Koren(R)	0.390221	0.0931876	0.0768807
SSP33 Woodfield($R, 2, -2$)	0.391025	0.0941322	0.0774227
SSP33 Differentiable limiter (r)	0.396675	0.100286	0.0820969
SSP33 SuperbeeR($R, 3, -1$)	0.279732	0.0779715	0.0718248
SSP33 Superbee(R)	0.309864	0.0752206	0.0706076

limiter, and a limiter eq. (4.13) potentially more compressive than the Superbee limiter (more accurate in the regime of tracers with sharp gradients).

The new limiter region provides a framework for proposing schemes that maintain a discrete local maximum principle when solving flux form equations when the flow is incompressible. We also have proven that this particular positive coefficient representation requires all symmetric limiters to be in the Sweby region Appendix A. We conjecture this to be a necessary condition for incompressible flow when symmetry is required on the limiter functions. We have also shown linear in-variance of the numerical scheme. Of practical merit we have numerically tested that pushing three known limiters in the Spekreijse region into the Sweby region had little negative consequence in terms of accuracy, we conjecture this to be related to the accuracy in the local neighbourhood of extrema considerations found in [14].

Future work involves the study of the parameters M, m in which the Woodfield(R, M, m) limiter and the SuperbeeR(R, M, m) limiter are most cost effective when the Courant number restriction is taken into account, and whether this changes when posed in the $\theta = 0$ framework. The design of new limiters in the introduced region Fig. 3.2 is open; accuracy, smoothness, and design criteria making use of less floating point operations would be of practical merit. We have only investigated $\theta = \{0, 1\}$, the generalisation by introducing a free parameter would be interesting and could lead to improved properties of limiting schemes. In the $\theta = 0$ region, there are differentiable limiters in the second order region. In the other region ($\theta = 1$), there are no differentiable limiters in the second order region. One could conjecture that this asymmetry indicates that there may be less restrictive sufficient conditions in the $\theta = 1$ framework than presented here using $\psi(R) < 0$, for $R > 0$.

CRediT authorship contribution statement

James Woodfield: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Hilary Weller:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Colin J. Cotter:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

During this work, the first author has been supported by an EPSRC studentship as part of the Centre for Doctoral Training in the Mathematics of Planet Earth (grant number EP/L016613/1). Rupert Klein and Peter Sweeney for comments leading to the improvement of this document.

Appendix A. Details of proof for Theorem 3.1

Proof. of Theorem 3.1. Expand both the method in section 2.2

$$\begin{aligned} \frac{\partial \bar{u}}{\partial t} + c_{i+0.5,j}^+ u_i^R + c_{i+0.5}^- u_{i+1}^L - c_{i-0.5}^+ u_{i-1}^R - c_{i-0.5}^- u_i^L \\ + c_{j+0.5}^+ u_j^U + c_{j+0.5}^- u_{j+1}^D - c_{j-0.5}^+ u_{j-1}^U - c_{j-0.5}^- u_j^D = 0, \end{aligned} \quad (\text{A.1})$$

and a discrete form of the divergence-free condition

$$\begin{aligned} c_{i+0.5}^+ u + c_{i+0.5}^- u - c_{i-0.5}^+ u - c_{i-0.5}^- u \\ + c_{j+0.5}^+ u + c_{j+0.5}^- u - c_{j-0.5}^+ u - c_{j-0.5}^- u = 0, \end{aligned} \quad (\text{A.2})$$

in terms of their positive and negative components. Where here we adopt the notation where if one or other subscript is missing it is assumed to be at position i or j as appropriate. Taking away the divergence-free condition (A.2) from (A.1) gives

$$\begin{aligned} \frac{\partial \bar{u}}{\partial t} + c_{i+0.5,j}^+ (u_i^R - u_i) + c_{i+0.5}^- (u_{i+1}^L - u_i) - c_{i-0.5}^+ (u_{i-1}^R - u_i) - c_{i-0.5}^- (u_i^L - u_i) \\ + c_{j+0.5}^+ (u_j^U - u_i) + c_{j+0.5}^- (u_{j+1}^D - u_i) - c_{j-0.5}^+ (u_{j-1}^U - u_i) - c_{j-0.5}^- (u_j^D - u_i) = 0. \end{aligned} \quad (\text{A.3})$$

Using $u_{i+1} - u_i = R_i(u_i - u_{i-1})$ and eq. (2.3), one can derive the following identities

$$u_i^R - u_i = \frac{\theta}{2} \psi(R_i)(u_i - u_{i-1}) + \frac{1-\theta}{2} \psi\left(\frac{1}{R_i}\right)(u_{i+1} - u_i), \quad (\text{A.4})$$

$$= \left[\frac{\theta}{2} \psi(R_i) + \frac{1-\theta}{2} R_i \psi\left(\frac{1}{R_i}\right) \right] (u_i - u_{i-1}). \quad (\text{A.5})$$

$$= \left[\frac{\theta}{2} \frac{\psi(R_i)}{R_i} + \frac{1-\theta}{2} \psi\left(\frac{1}{R_i}\right) \right] (u_{i+1} - u_i). \quad (\text{A.6})$$

Using $u_{i-1} - u_{i-2} = \frac{1}{R_{i-1}}(u_i - u_{i-1})$ and eq. (2.3), one can derive the following identities

$$u_{i-1}^R - u_i = (u_{i-1} - u_i) + \frac{\theta}{2} \psi(R_{i-1})(u_{i-1} - u_{i-2}) + \frac{1-\theta}{2} \psi\left(\frac{1}{R_{i-1}}\right)(u_i - u_{i-1}), \quad (\text{A.7})$$

$$= -\left[1 - \frac{\theta}{2} \frac{\psi(R_{i-1})}{R_{i-1}} - \frac{1-\theta}{2} \psi\left(\frac{1}{R_{i-1}}\right) \right] (u_i - u_{i-1}), \quad (\text{A.8})$$

$$= -\left[1 - \frac{\theta}{2} \frac{\psi(R_{i-1})}{R_{i-1}} \right] (u_i - u_{i-1}) + \frac{1-\theta}{2} \psi\left(\frac{1}{R_{i-1}}\right) \frac{1}{R_i} (u_{i+1} - u_i), \quad (\text{A.9})$$

Using $-R_i(u_i - u_{i-1}) = (u_i - u_{i+1})$ and eq. (2.4), one can derive the identities

$$u_i^L - u_i = \frac{\theta}{2} \psi\left(\frac{1}{R_i}\right)(u_i - u_{i+1}) - \frac{1-\theta}{2} \psi(R_i)(u_i - u_{i-1}), \quad (\text{A.10})$$

$$= \left[-\frac{\theta}{2} \psi\left(\frac{1}{R_i}\right) R_i - \frac{1-\theta}{2} \psi(R_i) \right] (u_i - u_{i-1}), \quad (\text{A.11})$$

$$= \left[\frac{\theta}{2} \psi\left(\frac{1}{R_i}\right) + \frac{1-\theta}{2} \frac{\psi(R_i)}{R_i} \right] (u_i - u_{i+1}), \quad (\text{A.12})$$

Using $-R_{i+1}(u_{i+1} - u_i) = (u_{i+1} - u_{i+2})$ and eq. (2.4), one can derive the identities

$$u_{i+1}^L - u_i = (u_{i+1} - u_i) + \frac{\theta}{2} \psi\left(\frac{1}{R_{i+1}}\right)(u_{i+1} - u_{i+2}) - \frac{1-\theta}{2} \psi(R_{i+1})(u_{i+1} - u_i), \quad (\text{A.13})$$

$$= (u_{i+1} - u_i) + \frac{\theta}{2} \psi\left(\frac{1}{R_{i+1}}\right) R_{i+1} (u_{i+1} - u_i) - \frac{1-\theta}{2} \psi(R_{i+1})(u_{i+1} - u_i), \quad (\text{A.14})$$

$$= -\left[1 - \frac{\theta}{2} \psi\left(\frac{1}{R_{i+1}}\right) R_{i+1} - \frac{1-\theta}{2} \psi(R_{i+1}) \right] (u_i - u_{i+1}), \quad (\text{A.15})$$

As highlighted above, there are many achievable representations of the numerical scheme that can be attained using only the definition of R . In this work, we seek a representation with less stringent requirements on the limiter function for the representation to be of positive coefficient type.

The idea behind the representation is to group $u_i^R - u_i$, and $u_{i-1}^R - u_i$, together in terms of $(u_i - u_{i-1})$ only, using eqs. (A.5) and (A.8). And to also group $u_{i+1}^L - u_i$, $u_i^L - u_i$ together in terms of $u_{i+1} - u_i$, using eqs. (A.12) and (A.15).

This way we group like signed terms such as $c_{i+1/2}^+$, and $c_{i-1/2}^+$ together in the same positive coefficient. This, when substituted into eq. (A.3) results in the following representation

$$\begin{aligned} \frac{\partial u}{\partial t} &+ \left(c_{i+1/2}^+ \left[\frac{\theta \psi(R_i)}{2} + \frac{(1-\theta)}{2} R_i \psi\left(\frac{1}{R_i}\right) \right] + c_{i-1/2}^+ \left[1 - \frac{\theta \psi(R_{i-1})}{2R_{i-1}} - \frac{(1-\theta)}{2} \psi\left(\frac{1}{R_{i-1}}\right) \right] \right) [u_i - u_{i-1}] \\ &+ \left(-c_{i+1/2}^- \left[1 - \frac{\theta}{2} \psi\left(\frac{1}{R_{i+1}}\right) R_{i+1} - \frac{1-\theta}{2} \psi(R_{i+1}) \right] - c_{i-1/2}^- \left[\frac{\theta}{2} \psi\left(\frac{1}{R_i}\right) + \frac{1-\theta}{2} \frac{\psi(R_i)}{R_i} \right] \right) [u_i - u_{i+1}] \\ &+ \text{y-direction.} \end{aligned} \quad (\text{A.16})$$

This representation is fundamental to the construction of the limiter frameworks, introduced in this paper. But many different representations of the scheme are possible, and it may be the case that another representation can be proven positive coefficient, perhaps with less demands on the limiter function.

When $\theta = 1$ the representation reduces to

$$\begin{aligned} \frac{\partial u}{\partial t} &+ \left[c_{i+1/2}^+ \frac{\psi(R_i)}{2} + c_{i-1/2}^+ \left[1 - \frac{\psi(R_{i-1})}{2R_{i-1}} \right] \right] (u_i - u_{i-1}) \\ &+ \left[-c_{i+1/2}^- \left[1 - \frac{1}{2} R_{i+1} \psi\left(\frac{1}{R_{i+1}}\right) \right] - c_{i-1/2}^- \frac{1}{2} \psi\left(\frac{1}{R_i}\right) \right] (u_i - u_{i+1}) + \text{y-direction} = 0, \end{aligned} \quad (\text{A.17})$$

where we read off the leading term coefficients

$$A_{i-1/2} = \left(\frac{1}{2} \left[c_{i+0.5}^+ \psi(R_i) \right] + c_{i-0.5}^+ \left[1 - \frac{1}{2} \psi(R_{i-1})/R_{i-1} \right] \right), \quad (\text{A.18})$$

$$B_{i+1/2} = - \left(c_{i+0.5}^- \left[1 - \frac{1}{2} \psi\left(\frac{1}{R_{i+1}}\right) R_{i+1} \right] + \frac{1}{2} \left[c_{i-0.5}^- \psi(R_i^{-1}) \right] \right), \quad (\text{A.19})$$

$$C_{j-1/2} = (i \mapsto j) \circ (A_{i-1/2}), \quad D_{j+1/2} = (i \mapsto j) \circ (B_{i+1/2}). \quad (\text{A.20})$$

We have introduced unorthodox notation $(i \mapsto j) \circ (f(i))$ to mean the same expression but with i replaced with j . Clearly

$$\psi(R) \geq 0 \quad \text{and} \quad \psi(S)/S \leq 2, \quad (\text{A.21})$$

are sufficient, for the scheme to be of positive coefficient type.

It is also imposed as a necessary assumption on the limiter functions to be of this particular positive coefficient representation because of the arbitrary nature of the velocity field. Suppose that $c_{i+1/2} \geq 0$, then for the scheme to be of positive coefficient type at both $u_{i,j}$ and $u_{i+1,j}$ we must require that,

$$1/2 c_{i+1/2}^+ \psi(R_i) \geq 0, \quad c_{i+1/2}^+ (1 - 1/2 \psi(R_i) R_i^{-1}) \geq 0. \quad (\text{A.22})$$

There could always exist a different positive coefficient representation of the scheme under some yet to be found transform or rearrangement.

We now attempt to find a sensible sufficient timestep restriction, the conditions

$$0 \leq \psi(R) \leq M_\psi, \quad m_\psi \leq \psi(S)/S \leq 2, \quad (\text{A.23})$$

$$m_\psi \leq \psi(1/T)T \leq 2, \quad 0 \leq \psi(1/r) \leq M_\psi, \quad \forall R, S, T \in \mathbb{R}, \quad (\text{A.24})$$

are sufficient for the following bounds

$$\begin{aligned} A_{i-1/2} &\in [0, c_{i+0.5}^+ M_\psi/2 + c_{i-0.5}^+ (1 - m_\psi/2)], \\ B_{i+1/2} &\in [0, -c_{i+0.5}^- [1 - m_\psi/2] - c_{i-0.5}^- M_\psi/2], \\ C_{j-1/2} &\in [0, c_{j+0.5}^+ M_\psi/2 + c_{j-0.5}^+ (1 - m_\psi/2)], \\ D_{j+1/2} &\in [0, -c_{j+0.5}^- [1 - m_\psi/2] - c_{j-0.5}^- M_\psi/2], \end{aligned} \quad (\text{A.25})$$

where we are yet to define the constants $M_\psi \geq 0$, and $m_\psi \leq 2$.

The time step restriction is

$$A_{i-1/2} + B_{i+1/2} + C_{i-1/2} + D_{i+1/2} \leq 1, \quad (\text{A.26})$$

which can be satisfied when

$$\begin{aligned} &c_{i+0.5}^+ M/2 + c_{i-0.5}^+ (1 - m/2) - c_{i+0.5}^- (1 - m/2) - c_{i-0.5}^- M/2 \\ &+ c_{j+0.5}^+ M/2 + c_{j-0.5}^+ (1 - m/2) - c_{j+0.5}^- (1 - m/2) - c_{j-0.5}^- M/2 \leq 1. \end{aligned} \quad (\text{A.27})$$

We lose some generality for a more convenient sufficient time step restriction. Define flow in and out Courant numbers by the following definitions

$$C_{i,j}^{in} := c_{i+0.5}^+ - c_{i-0.5}^- + c_{j+0.5}^+ - c_{j-0.5}^-, \quad (\text{A.28})$$

$$C_{i,j}^{out} := -c_{i+0.5}^- + c_{i-0.5}^+ - c_{j+0.5}^- + c_{j-0.5}^+, \quad (\text{A.29})$$

where these definitions are chosen based on how the flows effect the solution du/dt . The time step restriction can be written as

$$C_{i,j}^{in} \frac{M_\psi}{2} + C_{i,j}^{out} (1 - \frac{m_\psi}{2}) \leq 1, \quad (\text{A.30})$$

using incompressibility $C_{i,j}^{in} = C_{i,j}^{out}$ we can write this as the following Courant number restriction

$$C \leq C_{FE} = \frac{1}{(1 + \frac{M_\psi - m_\psi}{2})}. \quad (\text{A.31})$$

So far we have considered the $\theta = 1$ case, the general form is given by

$$\begin{aligned} \frac{\partial u}{\partial t} + & \left(c_{i+1/2}^+ \left[\frac{\theta \psi(R_i)}{2} + \frac{(1-\theta)}{2} R_i \psi\left(\frac{1}{R_i}\right) \right] + c_{i-1/2}^+ \left[1 - \frac{\theta \psi(R_{i-1})}{2R_{i-1}} - \frac{(1-\theta)}{2} \psi\left(\frac{1}{R_{i-1}}\right) \right] \right) [u_i - u_{i-1}] \\ & + \left(-c_{i+1/2}^- \left[1 - \frac{\theta}{2} \psi\left(\frac{1}{R_{i+1}}\right) R_{i+1} - \frac{1-\theta}{2} \psi(R_{i+1}) \right] - c_{i-1/2}^- \left[\frac{\theta}{2} \psi\left(\frac{1}{R_i}\right) + \frac{1-\theta}{2} \frac{\psi(R_i)}{R_i} \right] \right) [u_i - u_{i+1}] \\ & + \text{y-direction}. \end{aligned} \quad (\text{A.32})$$

We now repeat the previous argument for $\theta = 0$, omitting some of the details. The below conditions

$$\frac{1}{2} R_i \psi\left(\frac{1}{R_i}\right) \geq 0, \quad (\text{A.33})$$

$$1 - \frac{1}{2} \psi\left(\frac{1}{R_{i-1}}\right) \geq 0, \quad (\text{A.34})$$

$$1 - \frac{1}{2} \psi(R_{i+1}) \geq 0, \quad (\text{A.35})$$

$$\frac{\psi(R_i)}{2R_i} \geq 0, \quad (\text{A.36})$$

are sufficient for the positivity of the coefficients. This can be written more conveniently as

$$m_\psi \leq \psi(1/R) \leq 2, \quad 0 \leq S\psi\left(\frac{1}{S}\right) \leq M_\psi. \quad (\text{A.37})$$

The timestep restriction for the discrete maximum principle for the $\theta = 0$ scheme is

$$C \leq \frac{2}{2 + M_\psi - m_\psi}. \quad \square \quad (\text{A.38})$$

Appendix B. Necessary conditions for sign preservation

We detail some of the necessary conditions on the limiter function deduced by case-by-case considerations on possible velocity field and tracer values. This is done through means of contradiction-type arguments to attain the necessary conditions for sign preservation.

Necessary. Consider the forward Euler scheme given by

$$\begin{aligned} u_{ij}^{n+1} = & u_{ij}^n - \left(c_{i+1/2}^+ \left[\frac{\theta \psi(R_i)}{2} + \frac{(1-\theta)}{2} R_i \psi\left(\frac{1}{R_i}\right) \right] + c_{i-1/2}^+ \left[1 - \frac{\theta \psi(R_{i-1})}{2R_{i-1}} - \frac{(1-\theta)}{2} \psi\left(\frac{1}{R_{i-1}}\right) \right] \right) [u_i - u_{i-1}] \\ & - \left(-c_{i+1/2}^- \left[1 - \frac{\theta}{2} \psi\left(\frac{1}{R_{i+1}}\right) R_{i+1} - \frac{1-\theta}{2} \psi(R_{i+1}) \right] - c_{i-1/2}^- \left[\frac{\theta}{2} \psi\left(\frac{1}{R_i}\right) + \frac{1-\theta}{2} \frac{\psi(R_i)}{R_i} \right] \right) [u_i - u_{i+1}] \\ & - \left(c_{j+1/2}^+ \left[\frac{\theta \psi(R_j)}{2} + \frac{(1-\theta)}{2} R_j \psi\left(\frac{1}{R_j}\right) \right] + c_{j-1/2}^+ \left[1 - \frac{\theta \psi(R_{j-1})}{2R_{j-1}} - \frac{(1-\theta)}{2} \psi\left(\frac{1}{R_{j-1}}\right) \right] \right) [u_j - u_{j-1}] \\ & - \left(-c_{j+1/2}^- \left[1 - \frac{\theta}{2} \psi\left(\frac{1}{R_{j+1}}\right) R_{j+1} - \frac{1-\theta}{2} \psi(R_{j+1}) \right] - c_{j-1/2}^- \left[\frac{\theta}{2} \psi\left(\frac{1}{R_j}\right) + \frac{1-\theta}{2} \frac{\psi(R_j)}{R_j} \right] \right) [u_j - u_{j+1}]. \end{aligned} \quad (\text{B.1})$$

Let $\theta = 1$, $c_{i+1/2} > 0$, $c_{i-1/2} = 0$, and let $c_{j+1/2}$, $c_{j-1/2}$ be free, such that the divergence free assumption holds. Then let $u_{j+1} = 0$, $u_{j-1} = 0$, $u_{ij}^n = 0$, and then let $u_{i-1}^n > 0$ and let $u_{i+1}^n \geq 0$. The forward Euler scheme reduces to

$$u_{i,j}^{n+1} = 1/2 c_{i+1/2}^+ \psi(R_i) u_{i-1}^n \quad \text{where} \quad R_i = -\frac{u_{i+1}^n}{u_{i-1}^n} \in (-\infty, 0]. \quad (\text{B.2})$$

If one supposes there exists $R < 0$, such that $\psi(R) < 0$, one can find u_{i+1} , u_{i-1} such that $u_{ij}^{n+1} < 0$ a change in sign. Therefore, $\psi(R) \geq 0$ is a necessary assumption when $R < 0$ and $\theta = 1$.

Now let $\theta = 1$, $c_{i+1/2} = 0$, $c_{i-1/2} > 0$, and let $c_{j+1/2}$, $c_{j-1/2}$ be free, such that the divergence free assumption holds. Then let $u_{j+1} = 0$, $u_{j-1} = 0$, $u_{ij}^n = 0$ and then let $u_{i-1}^n > 0$ and let $u_{i-1}^n \geq 0$. The forward Euler scheme reduces to

$$u_{i,j}^{n+1} = c_{i-1/2}^+ \left(1 - \frac{\psi(R_{i-1}^n)}{2R_{i-1}^n} \right) u_{i-1}^n \quad \text{where} \quad R_{i-1} = \frac{-u_{i-1}^n}{u_{i-1}^n - u_{i-2}^n} \in (-\infty, -1] \cup (0, \infty). \quad (\text{B.3})$$

Therefore as can be proven through contradiction, $\psi(R)/R \leq 2$ is a necessary assumption for sign preservation for $\theta = 1$ on the interval $R \in (-\infty, -1] \cup (0, \infty)$.

Now consider $c_{i+1/2} < 0$, $c_{i-1/2} = 0$, and $u_{ij}^n = 0$. Let $c_{j+1/2}$, $c_{j-1/2}$, be free such that the divergence free condition holds. Let $u_j = u_{j-1} = u_{j+1}$, then the scheme reduces to

$$u_{i,j}^{n+1} = (-c_{i-1/2}^-) \left[1 - \frac{1}{2} \psi\left(\frac{1}{R_{i+1}}\right) R_{i+1} \right] u_{i+1} \quad \text{where} \quad R_{i+1} = \frac{u_{i+2} - u_{i+1}}{u_{i+1}} \in [-1, \infty). \quad (\text{B.4})$$

Therefore, using the intermediate variable $S_{i+1} = 1/R_{i+1} \in (-\infty, -1) \cup (0, \infty)$, one can through contradiction show that

$$\frac{\psi(S)}{S} \leq 2, \quad (\text{B.5})$$

is a necessary assumption. Under similar considerations where instead $c_{i+1/2} = 0$, $c_{i-1/2} < 0$ one can show that $\psi(R) \geq 0$ for $R \in (-\infty, 0)$, is a necessary assumption. There is likely more additional necessary assumptions required on the limiter function, that can be derived from additional case-by-case considerations. $\square$

Appendix C. Linear invariance

In this section we prove linear invariance of the numerical scheme.

Definition C.1 (Linear invariant scheme). The map $u_i \mapsto \alpha w_i + \beta$ leaves the numerical method unchanged.

Theorem C.1 (Linear invariance). The method described in section 2.2 is linear invariant for incompressible flow.

Proof. As in [8] the transform $u_i \mapsto \alpha w_i + \beta$ is investigated for linear invariance. However, we do the analysis after the use of a discrete divergence free condition,

$$\frac{\partial u}{\partial t} + \left(\frac{1}{2} \left[c_{i+0.5}^+ \psi(R_i) \right] + c_{i-0.5}^+ \left[1 - \frac{1}{2} \psi(R_{i-1})/R_{i-1} \right] \right) (u_i - u_{i-1}) \quad (\text{C.1})$$

$$- \left(c_{i+0.5}^- \left[1 - \frac{1}{2} \psi\left(\frac{1}{R_{i+1}}\right) R_{i+1} \right] + \frac{1}{2} \left[c_{i-0.5}^- \psi(R_i^{-1}) \right] \right) (u_i - u_{i+1}) \quad (\text{C.2})$$

$$+ \text{same expression with } i \mapsto j = 0, \quad (\text{C.3})$$

and observe

$$R_i = \frac{(u_{i+1} - u_i)}{(u_i - u_{i-1})} \mapsto \frac{(\alpha w_{i+1} + \beta) - (\alpha w_i + \beta)}{(\alpha w_i + \beta) - (\alpha w_{i-1} + \beta)} = \frac{(w_{i+1} - w_i)}{(w_i - w_{i-1})}, \quad (\text{C.4})$$

$$u_i - u_{i-1} \mapsto \alpha(w_i - w_{i-1}), \quad (\text{C.5})$$

$$\frac{\partial u}{\partial t} \mapsto \alpha \frac{\partial w}{\partial t}, \quad (\text{C.6})$$

gives exact linear invariance of the numerical method. Constants are preserved exactly for incompressible flow, and the equation is scaling invariant. This holds for both $\theta \in \{0, 1\}$. $\square$

Appendix D. Time step restrictions

The numerical scheme defined in section 2.2 has a time step restriction for a discrete maximum principle that depends on which limiter is used, whether a directional mean value theorem holds and what temporal discretisation is used. We collate Table D.4 of the maximum cell defined Courant number the Forward Euler scheme has a discrete maximum principle predicted by Theorem 3.1, for some of the limiters introduced in this paper. This table can be used to check whether the scheme should be producing positive or discrete maximum principle satisfying results. The Courant number depends on the parameters M and m , this effects the efficiency of the scheme, and can be made arbitrarily small for injudicious choice of limiter.

Appendix E. Hidden maximum principles from internal strong stability

When one requires that the substages satisfy a discrete maximum principle, and limits at every stage in the Shu Osher representation, one implies substage maximum principles. For example our SSP33 scheme satisfies the (hidden) maximum principles

Table D.4

The sufficient Courant number restrictions for temporal discretisations of the method section 2.2 to satisfy a local discrete maximum principle for different limiters. Column one is the sufficient Courant number limit for incompressible flow and column two is the sufficient Courant number limit for directional mean value theorem satisfying flows. These time stepping criteria hold in more dimensions; however the definition of the Courant number changes.

Scheme	Limiter	Theorem 3.1-C	Spekreijse-MVTS-C
FE	$\psi(R) = 0$	1	1
FE	$\psi(R) = aR + b$	0	0
FE	vanAlbada(R)	0	$2 - \sqrt{2}$
FE	Ospre(R)	0	1/2
FE	ENO2(R)	0	1/2
FE	vanAlbada _p (R)	$\frac{4}{5+\sqrt{2}}$	$\frac{4}{5+\sqrt{2}}$
FE	Ospre _p (R)	4/7	4/7
FE	minmod(R)/ENO2 _p (R)	2/3	2/3
FE	Koren(R)	1/2	1/2
FE	Woodfield(R, M, m)	$2/(2 + M - m)$	$2/(2 + M - m)$
FE	Differentiable(r)	$\frac{2}{4+\sqrt{\frac{5\sqrt{2}}{2}-\frac{11}{2}}} \approx 0.4651$	$\frac{2}{4+\sqrt{\frac{5\sqrt{2}}{2}-\frac{11}{2}}} \approx 0.4651$
RK4	Koren(R)	0	0
FE	UTCDF(R)	0	0
FE	UTCDF _S (R)	1/2	1/2
FE	UTCDF _P (R)	1/2	1/2

$$k^1 \in [\min\{u_{i+1}^n, u_{i-1}^n, u^n, u_{j+1}^n, u_{j-1}^n\}, \max\{u_{i+1}^n, u_{i-1}^n, u^n, u_{j+1}^n, u_{j-1}^n\}], \quad (\text{E.1})$$

$$k^2 \in [\min\{k_{i+1}^1, k_{i-1}^1, k^1, k_{j+1}^1, k_{j-1}^1\}, \max\{k_{i+1}^1, k_{i-1}^1, k^1, k_{j+1}^1, k_{j-1}^1\}], \quad (\text{E.2})$$

$$u^{n+1} \in [\min\{k_{i+1}^2, k_{i-1}^2, k^2, k_{j+1}^2, k_{j-1}^2\}, \max\{k_{i+1}^2, k_{i-1}^2, k^2, k_{j+1}^2, k_{j-1}^2\}], \quad (\text{E.3})$$

instead of the original maximum principle

$$u_{i,j}^{n+1} \in [\min\{u_{i+1}^n, u_{i-1}^n, u^n, u_{j+1}^n, u_{j-1}^n\}, \max\{u_{i+1}^n, u_{i-1}^n, u^n, u_{j+1}^n, u_{j-1}^n\}]. \quad (\text{E.4})$$

This implies a different maximum principle,

$$u_{i,j}^{n+1} \in [\min_{\forall k,l \in \{-3,-2,-1,0,1,2,3\}} u_{i+k,j+l}, \max_{\forall k,l \in \{-3,-2,-1,0,1,2,3\}} u_{i+k,j+l}], \quad (\text{E.5})$$

but enforces positivity and local boundedness of the substages, which could be essential in the context of a numerical method for a coupled system.

Appendix F. Runge Kutta 4

The standard Runge Kutta 4 method well established for both its accuracy and efficiency is proposed for this particular spatial scheme by both [11] and [8]. In [11] it is indicated monotonicity is supposed to be guaranteed by small timesteps, it has since been shown that the RK4 algorithm has no Shu Osher representation (without perturbation techniques and downwind/upwind biasing) and has no(zero) radius of monotonicity. Hundsdorfer et al. [8] notes this but indicates the SSP literature is of little practical importance. In [17], the same author performed similar experiments and quantifies the observation of 10^{-6} negative values [17]. From a practical perspective, the importance of using a SSP timestepping method depends on the range of acceptable monotonicity failure. These negatives would be considered not just significant but very large in our application where the positivity of density should not be negative. We consider roughly 10^{-14} the threshold of acceptable monotonicity failure, due to machine precision error in constructing a divergence free vector field, and other accumulation of machine precision errors at 64 bit arithmetic. At lower precision arithmetic, it may be the case that the SSP literature is of less practical importance. Table 1 contains the minimum values attained by RK4 with the Koren(R) limiter they are positivity violating.

Appendix G. Accuracy of un-limited scheme

Hundsdorfer et al. [8] viewed the semi-discrete method section 2.2 as a finite difference method evolving point-wise values $u_{i,j}$ solving the singular point value equation. Hundsdorfer, and Trompert [17] showed that the third order upwind method solves the following modified equation using backward error analysis

$$\begin{aligned} u_t + (v_1 u)_x = & -\frac{\Delta x^2}{24} [u(v_1)_{xxx} + 3u_x(v_1)_{xx} + 2u_{xx}(v_1)_x] + (\text{y-direction}), \\ & -\frac{\Delta x^3}{12} [u_{xxx}v_1 + u_{xxx}(v_1)_x] + O(\Delta x^4) + (\text{y-direction}). \end{aligned} \quad (\text{G.1})$$

Where v_1 is the x component of the velocity field. So the scheme is a formally third order finite difference method approximating the singular point equation in one and two dimensions for directionally constant flow.

Zijlema and Wesseling [10] proposed the scheme as a finite volume method on cell averaged quantities approximating the integral form of the equation. When the method is interpreted as a finite volume scheme approximating the integral form of the equation, the formal truncation error analysis becomes different see [33–35] for explanation. In which case in one dimension the finite volume interpretation of this scheme becomes truly third order with respect to the integral form of the equation (Aleksandar Donev has uploaded lecture notes proving this fact using the symbolic algebra package Mathematica [36]). However, the scheme becomes second order in two dimensions or more because the one-dimensional flux integral is approximated by second order Gauss quadrature.

The formal truncation analysis of the full method is tedious in the more general setting, and instead one designs limiters using simplifications. If it is assumed the flow is uniformly constant, and the equation is considered in the pointwise finite difference sense. Then the linear scheme associated with

$$\psi(R) = aR + b, \quad (\text{G.2})$$

has the semi-discrete truncation error

$$\frac{u_i^R - u_{i-1}^R}{\Delta x} - (u_x)_i = [a + b - 1] \frac{\Delta x u_{xx}}{2!} + [1 - 3b] \frac{\Delta x^2 u_{xxx}}{3!} + (a + 7b - 1) \frac{\Delta x^3 u_{xxxx}}{4!} + \dots \quad (\text{G.3})$$

So that $u \in C^2$ and $a + b = 1$ is sufficient for second order, such that $\psi(R) = aR + (1 - a)$ passes through $(1, 1)$. $u \in C^3$, $a = 2/3$, $b = 1/3$ is sufficient for third order.

Appendix H. Accuracy of limited scheme

Spekreijse, showed that sufficient conditions on the limiter function $\psi(R)$ for second order accuracy are $\psi(1) = 1$, $\psi \in C^2$ near 1, [6]. Sufficient and necessary conditions on the limiter function $\psi(r)$ were shown by [14] to be $\psi(1) = 1$, and ψ is Lipschitz continuous.

We also want to understand the effect limiting has on the truncation error particularly in the neighbourhood of smooth extrema where the gradient changes. It is often mistaken that TVD schemes are first order at noncritical extrema because of a slight oversight in the truncation analysis in Osher 1984 [24], [37]. However, this is not quite the case as explained by Hua-mo [14] where formal truncation analysis of the $\theta = 0$ scheme is done by expanding the higher order correction in the neighbourhood of a critical point $x_\alpha = x_i + \alpha \Delta x$. We summarise the results from [14] below.

Theorem H.1 (constant flow in one dimension near the critical points $\theta = 0$ [14]). *TVD schemes $\theta = 0$ may have second order accuracy at critical points if $\psi(r = 3) + \psi(r = -1) = 2$. But cannot be uniformly second-order accurate in the neighbourhood of critical points. If $\psi(1) = 1$, then these TVD schemes have second-order accuracy in the region sufficiently far from the critical points of smooth solutions.*

Proof. [14]. $\square$

Similar conclusions for the $\theta = 1$ form are done in [10] however both Theorem 3.1 and proof of Theorem 3.1, have a few technical inaccuracies. Nevertheless, all the arguments are correct in spirit, second order at extrema is technically possible but impossible to achieve in the local neighbourhood of extrema. We recommend following the analysis methods of [14] instead.

Theorem H.2 ([14]-for $\theta = 1$). *Let τ_i denote the truncation error about point x_i , let $x_\xi = x_i + \xi \Delta x$, $\xi \in [0, \infty)$, be a point of interest. TVD schemes may have second order accuracy at critical points if $3\psi(R = 1/3) - \psi(R = -1) = 2$. But cannot be uniformly second-order accurate in the whole neighbourhood of critical points. If $\psi(1) = 1$, then TVD schemes have second-order accuracy in the region sufficiently far from the critical points of smooth solutions.*

Proof. See [14], we include some sketch details of the proof in Appendix I. $\square$

The reason this theoretical work is revisited by us here, is because some authors suggest using the second order at extrema condition in the design of limiters [14,10] and others [28] suggest that satisfying the theoretical second order at extrema condition does not have practical consequences. In our work we have suggested various modifications to limiters including pushing into the Sweby region, this breaks the $3\psi(R = 1/3) - \psi(R = -1) = 2$ condition for the ENO2 limiter.

Appendix I. On the order of total variation diminishing schemes near extrema

In this section we present a sketch of the arguments of [14] but for the $\theta = 0$ case. We use the prime notation u' to denote the spatial derivative u_x for readability. The semi-discrete finite difference local truncation error τ of the spatial derivative u_x at position x_i is

$$\tau_i := \frac{u(x_i) - u(x_{i-1})}{\Delta x} + \frac{\psi(R_u(x_{i+1}, x_i, x_{i-1}))(u(x_i) - u(x_{i-1}))}{2\Delta x} - \frac{\psi(R_u(x_i, x_{i-1}, x_{i-2}))(u(x_{i-1}) - u(x_{i-2}))}{2\Delta x} - u'(x_i), \quad (\text{I.1})$$

where $R_u(a, b, c) := \frac{u(a)-u(b)}{u(b)-u(c)}$. As in [14] we wish to write this local truncation error at x_i as a function of u about an arbitrary point $x_\xi = x_i + \xi \Delta x$, to determine the order of accuracy within a distance of a critical point.

First, Taylor expand the first order upwind term in eq. (I.1) about x_i to cancel with $u_x(x_i)$, then re-expand again about the arbitrary point $x_\xi = x_i + \xi \Delta x$ as in [14]. One can deduce the first order upwind scheme has the following expression of local truncation error

$$\frac{u(x_i) - u(x_{i-1}))}{\Delta x} - u'(x_i) = -\Delta x \frac{u''(x_i)}{2!} + \Delta x^2 \frac{u'''(x_i)}{3!} \dots \quad (\text{I.2})$$

$$= -\frac{\Delta x}{2!} [u''(x_\xi) - \xi \Delta x u'''(x_\xi) + \frac{\xi^2 \Delta x^2 u''''(x_\xi)}{2!} \dots] \quad (\text{I.3})$$

$$+ \frac{\Delta x^2}{3!} [u'''(x_\xi) - \xi \Delta x u''''(x_\xi) \dots] \quad (\text{I.4})$$

$$= -\frac{\Delta x}{2!} u''(x_\xi) + (3\xi + 1) \frac{\Delta x^2 u'''(x_\xi)}{3!} + O(\Delta x^3). \quad (\text{I.5})$$

Now for the higher order flux limited correction in eq. (I.1) use the following Taylor expansions about x_ξ

$$u(x_{i+1}) - u(x_i) = \Delta x u'(x_\xi) - (\xi - 1/2) \Delta x^2 u''(x_\xi) - \frac{(\xi - 1)^3 - \xi^3}{3!} \Delta x^3 u'''(x_\xi) - \dots \quad (\text{I.6})$$

$$u(x_i) - u(x_{i-1}) = \Delta x u'(x_\xi) - (\xi + 1/2) \Delta x^2 u''(x_\xi) - \frac{(\xi)^3 - (\xi + 1)^3}{3!} \Delta x^3 u'''(x_\xi) - \dots \quad (\text{I.7})$$

$$u(x_{i-1}) - u(x_{i-2}) = \Delta x u'(x_\xi) - (\xi + 3/2) \Delta x^2 u''(x_\xi) - \frac{(\xi + 1)^3 - (\xi + 2)^3}{3!} \Delta x^3 u'''(x_\xi) - \dots \quad (\text{I.8})$$

This allows us to write down an expression for the local truncation error in terms of x_ξ

$$\begin{aligned} \tau_i := & -\frac{\Delta x}{2!} u''(x_\xi) + (3\xi + 1) \frac{\Delta x^2 u'''(x_\xi)}{3!} \\ & + \frac{1}{2} \psi(R_i) [u'(x_\xi) - (\xi + 1/2) \Delta x u''(x_\xi) - \frac{(\xi - 1)^3 - \xi^3}{3!} \Delta x^2 u'''(x_\xi)] \\ & - \frac{1}{2} \psi(R_{i-1}) [u'(x_\xi) - (\xi + 3/2) \Delta x u''(x_\xi) - \frac{(\xi + 1)^3 - (\xi + 2)^3}{3!} \Delta x^2 u'''(x_\xi)] + O(\Delta x^3). \end{aligned} \quad (\text{I.9})$$

We collect terms of the appropriate order to define the truncation error as

$$\begin{aligned} \tau_i := & \frac{1}{2} [\psi(R_i) - \psi(R_{i-1})] u'(x_\xi) \\ & + [-1 - \psi(R_i)(\xi + 1/2) + \psi(R_{i-1})(\xi + 3/2)] \frac{\Delta x}{2!} u''(x_\xi) \\ & + \left[(3\xi + 1) - \frac{1}{2} \psi(R_i) [(\xi - 1)^3 - \xi^3] + \frac{1}{2} \psi(R_{i-1}) [(\xi + 1)^3 - (\xi + 2)^3] \right] \frac{\Delta x^2 u'''(x_\xi)}{3!}, \end{aligned} \quad (\text{I.10})$$

here the gradient R is also Taylor expanded about the arbitrary point x_ξ ,

$$R_i = \frac{u'(x_\xi) - (\xi - 1/2) \Delta x u''(x_\xi) - \frac{(\xi - 1)^3 - \xi^3}{3!} \Delta x^2 u'''(x_\xi) - \dots}{u'(x_\xi) - (\xi + 1/2) \Delta x u''(x_\xi) - \frac{(\xi)^3 - (\xi + 1)^3}{3!} \Delta x^2 u'''(x_\xi) - \dots} \quad (\text{I.11})$$

$$R_{i-1} = \frac{u'(x_\xi) - (\xi + 1/2) \Delta x u''(x_\xi) - \frac{(\xi)^3 - (\xi + 1)^3}{3!} \Delta x^2 u'''(x_\xi) - \dots}{u'(x_\xi) - (\xi + 3/2) \Delta x u''(x_\xi) - \frac{(\xi + 1)^3 - (\xi + 2)^3}{3!} \Delta x^2 u'''(x_\xi) - \dots} \quad (\text{I.12})$$

This allows us to state the requirements of the scheme to be first second and third order.

Definition I.1 (First order requirement near x_ξ). First order accuracy requires

$$\left[\frac{\psi(R_i)}{2} - \frac{\psi(R_{i-1})}{2} \right] u_x(x_\xi) = O(\Delta x). \quad (\text{I.13})$$

Definition I.2 (Second order requirement near x_ξ). Second order accuracy requires

$$\begin{aligned} & \left[\frac{\psi(R_i)}{2} - \frac{\psi(R_{i-1})}{2} \right] u_x(x_\xi) \\ & + \left[-1 - (\xi + \frac{1}{2}) \psi(R_i) + (\xi + \frac{3}{2}) \psi(R_{i-1}) \right] \frac{\Delta x u_{xx}(x_\xi)}{2!} = O(\Delta x^2). \end{aligned} \quad (\text{I.14})$$

Definition 1.3 (Third order requirement near x_ξ). Third order accuracy requires

$$\begin{aligned} & \left[\frac{\psi(R_i)}{2} - \frac{\psi(R_{i-1})}{2} \right] u_x(x_\xi) \\ & + \left[-1 - (\xi + \frac{1}{2})\psi(R_i) + (\xi + \frac{3}{2})\psi(R_{i-1}) \right] \frac{\Delta x u_{xx}(x_\xi)}{2!} \\ & + \left[(3\xi + 1) - \frac{\psi(R_i)}{2} [(\xi)^3 - (\xi + 1)^3] + [(\xi + 1)^3 - (\xi + 2)^3] \frac{\psi(R_{i-1})}{2} \right] \frac{\Delta x^2 (u_{xxx})(x_\xi)}{3!} = O(\Delta x^3). \end{aligned} \quad (1.15)$$

There are now three cases to consider. We want to know the error at a critical point x_i , at a non critical point x_i , and the error at x_i within the vicinity of a critical point x_ξ . Where we do not consider inflections as critical. One has to make an assumption on the size of Δx , in order to use the identity $1/(1-a) = 1+a+a^2+\dots$, for $|a| \leq 1$, to approximate the expressions for R_i, R_{i-1} . We will also use the fact that $O(|f(x)|) = O(f(x))$, and the definition of Lipschitz continuity $\forall x, y, \exists K > 0$ s.t. $|f(x) - f(y)| < K|x - y|$. One also has to make technical remarks about the smoothness of u required for derivatives or more sharply, for the mean value theorem to apply (the inclusion of these technical remarks would be a important contribution to the literature). Case 1: We are at a critical point in position i , so that

$$\xi = 0 \quad (1.16)$$

$$u_x(x_\xi) = u_x(x_i) = 0 \quad (1.17)$$

$$R_i = -1 - \frac{2\Delta x}{3} \frac{u_{xxx}(x_i)}{u_{xx}(x_i)} + O(\Delta x^2) \quad (1.18)$$

$$R_{i-1} = \frac{1}{3} + \frac{2\Delta x}{3} \frac{u_{xxx}(x_i)}{u_{xx}(x_i)} + O(\Delta x^2) \quad (1.19)$$

where $u_{xx}(x_i) \neq 0$. Case 2: We are at a non critical point i , so that

$$\xi = 0 \quad (1.20)$$

$$u_x(x_\xi) = u_x(x_i) \neq 0 \quad (1.21)$$

$$R_i = 1 + \Delta x u''(x_i)/u'(x_i) + O(\Delta x^3) \quad (1.22)$$

$$R_{i-1} = 1 + \Delta x u''(x_i)/u'(x_i) + \Delta x^2 (4/3 u_{xxx} x_i - 3 u_{xx}(x_i)^2) + O(\Delta x^3) \quad (1.23)$$

Case 3: We are at x_i in the vicinity of a critical point at x_ξ

$$\xi \neq 0 \quad (1.24)$$

$$u_x(x_\xi) = 0 \quad (1.25)$$

$$R_i = \frac{(\xi - 1/2)}{(\xi + 1/2)} + O(\Delta x) \quad (1.26)$$

$$R_{i-1} = \frac{(\xi + 1/2)}{(\xi + 3/2)} + O(\Delta x) \quad (1.27)$$

We satisfy first order accuracy requirements in case 1 and case 3 trivially because $u_x(x_\xi) = 0$. In case 2 at a non-critical x_i the first order requirement is

$$\psi(1 + \Delta x u''(x_i)/u'(x_i) + O(\Delta x^3)) - \psi(1 + \Delta x u''(x_i)/u'(x_i) + O(\Delta x^2)) = O(\Delta x). \quad (1.28)$$

A sufficient condition for this to be satisfied is Lipschitz continuity of the limiter function.

We first note that all inflection points will be second order because $u_x(x_i), u_{xx}(x_i) = 0$. In case 1 at critical, non inflection points we have second order accuracy when

$$-1 - 1/2\psi(-1 - a\Delta x + O(\Delta x^2)) + 3/2\psi(1/3 + a\Delta x + O(\Delta x^2)) = O(\Delta x), \quad (1.29)$$

where $a = \frac{2}{3} \frac{u_{xxx}(x_i)}{u_{xx}(x_i)}$. If one first assumes

$$1 - 3/2\psi(1/3) + 1/2\psi(-1) = 0 \quad (1.30)$$

and then adds this to eq. (1.29). One can deduce that Lipschitz continuity and $1 - 3/2\psi(1/3) + 1/2\psi(-1) = 0$ are sufficient and necessary for second order, by the use of a contradiction argument for the necessary condition.

In case 2 at a non critical, non inflection point we have second order accuracy when both the conditions

$$\psi(1 + a\Delta x + O(\Delta x^3)) - \psi(1 + a\Delta x + O(\Delta x^2)) = O(\Delta x^2) \quad (1.31)$$

$$-1 - 1/2\psi(1 + a\Delta x + O(\Delta x^3)) + 3/2\psi(1 + a\Delta x + b\Delta x^2) = O(\Delta x) \quad (1.32)$$

are satisfied for $a = u''(x_i)/u'(x_i)$. A sufficient and necessary condition for the first condition to hold is Lipschitz continuity. If this assumption is used on the second condition we can rewrite the second condition as

$$-1 + \psi(1 + a\Delta x + O(\Delta x^2)) = O(\Delta x) \quad (1.33)$$

If one assumes $\psi(1) = 1$ and adds $1 - \psi(1) = 0$ to the above expression eq. (1.33) one can see that Lipschitz continuity and $\psi(1) = 1$ are sufficient and necessary. We now consider second order accuracy in case 3, within a neighbourhood of a critical point which requires

$$-\left[1 - \left(\xi + \frac{1}{2}\right)\psi(R_i) - \left(\xi + \frac{3}{2}\right)\psi(R_{i-1})\right] \frac{u_{xx}(x_\xi)}{2!} = O(\Delta x). \quad (1.34)$$

Leading to the condition

$$\left[1 + \left(\frac{1}{2} + \xi\right)\psi\left(\frac{\xi - 1/2}{\xi + 1/2} + O(\Delta x)\right) - \left(\xi + \frac{3}{2}\right)\psi\left(\frac{\xi + 1/2}{\xi + 3/2} + O(\Delta x)\right)\right] = O(\Delta x), \quad (1.35)$$

assuming differentiability of the limiter then a sufficient condition for the above to hold is the following condition

$$1 + \left(\xi + \frac{1}{2}\right)\psi\left(\frac{\xi - 1/2}{\xi + 1/2}\right) - \left(\xi + \frac{3}{2}\right)\psi\left(\frac{\xi + 1/2}{\xi + 3/2}\right) = 0 \quad (1.36)$$

If the above property eq. (1.36) is assumed and it is subtracted from eq. (1.35), the differentiability is relaxed to the Lipschitz continuity and can be proven both necessary and sufficient.

The second order in the neighbourhood of an extrema condition eq. (1.36) is satisfied for any second order linear scheme $\psi(R) = aR + b$ when $1 - a - b = 0$. But the only solution that satisfies $\psi(0) = 0$ is given by $\psi(R) = aR$ which is not within monotonicity constraints for large gradients [14]. This eq. (1.36) condition reduces to the second order $\psi(1) = 1$ non-extrema condition in the limit $\xi \rightarrow \pm\infty$ sufficiently away from extrema. The eq. (1.36) condition reduces to the previously derived second order at extrema condition $2 = 3\psi(1/3) - \psi(-1)$ in the limit $\xi \rightarrow 0$. Concluding the known accuracy results in the neighbourhood of extrema.

Appendix J. Relationship to other limiter regions

This appendix reviews some other limiter regions, the limiter regions reviewed are different in definition, construction and aims than those presented in this paper. Spekreijse in 1987, introduced a generalised extended limiter region in a flux splitting framework, aligning with the $\theta = 0$ subcell reconstruction choice in this paper, allowing for non zero values of $\psi(R)$ for $R < 0$. This same type of free parameter generalisation was later applied to the Sweby region in the $\theta = 1$ framework in the work by Cahouet-Coquel in 1989 [15] see also [14], in which non zero values are allowed for $\psi(r)$ for $r < 0$. It is worth noting that in both $\theta \in \{0, 1\}$ frameworks, a mean value theorem is always imposed, these extended limiter regions have not been proven positive coefficient for flux form advection for general velocity fields. This is due to their prevalent use in flux splitting or flux vector splitting frameworks. In 2010, Dubois [38] introduced two limiter regions in the context of one-dimensional flows, based on the monotonicity and convexity of the subcell reconstruction. Hyman in 1983 [39] and Huynh in 1993 [40] extended the monotonicity preserving constraints in which the MP-Parabolic construction or a Hermite cubic interpolant is monotone. When written in terms of a limiter function, this is found to

Table J.5

Same as Table 1, but using simple error handling for division, division is defined as $DV(a, b) := (a/(b + 2.2e - 16))$. We observe the same results up to machine precision.

Scheme	Limiter	MVTV-sin $\min_{q_{n,j}} u_{i,j}^n$	MVTS-sbr $\min_{q_{n,j}} u_{i,j}^n$	MVTV-sin 32 $\min_{q_{n,j}} u_{i,j}^n$
SSP33	ENO2(R)	-5.10824e - 08	-1.63505e-16	-0.0128308
SSP33	Ospre(R)	-5.21618e - 11	-4.08984e-16	-0.0450631
SSP33	Vanalbada(R)	-7.8611e - 10	-3.69424e-16	-1.08423e - 07
SSP33	ENO2 _p (R)	-2.52212e-16	-1.76984e-16	-7.11199e-17
SSP33	Ospre _p (R)	-7.20499e-16	-5.87407e-16	-8.58698e-17
SSP33	Vanalbada _p (R)	-3.53086e-16	-4.21421e-16	-7.9581e-17
SSP33	UTCDF _p (R)	-1.15881e-15	-7.18115e-16	-2.96113e-16
SSP33	Koren(R)	-7.50797e-16	-9.29711e-16	-4.70821e-15
SSP33	Superbee(R)	-7.03817e-16	-7.35427e-16	-1.18831e-14
SSP33	SuperbeeR(R, 3, -1)	-1.16908e-15	-8.03216e-16	-9.65484e-16
SSP33	Woodfield(R, 2, -1)	-7.55775e-16	-1.15396e-15	-8.68959e-16
SSP33	Woodfield(R, 4, 0)	-8.93315e-16	-8.78768e-16	-4.14971e-15
SSP33	Differentiable(r)	-5.86937e-16	-2.30743e-16	-1.05367e-14
SSP33	UTCDF _s (R)	-1.26445e-15	-7.46022e-16	-2.3749e-16
RK4	Koren(R)	-9.80942e - 11	-2.55303e - 10	-4.6105e-15
SSP33	UTCDF(R)	-5.70005e - 05	-7.31029e - 05	-0.000140264

be $\psi \in [-\min(3, 3R), \min(3, 3R)]$, $0 \leq \psi(R) \leq 3 \min(1, R)$ respectively. Čada and Torrilhon, in [41] design limiter functions based of the regions defined by both Hyman and Huynh. Typically, limiters deriving from Hyman's work have local monotone properties from the reconstruction, but not formal strict maximum principles (similar to WENO), unless they are further restricted to the Spekreijse limiter region. The above limiter regions are different to those in this paper, in terms of construction, aim and definition, in particular, the extent to which the above limiter regions are proven to have a discrete maximum principle for flux form incompressible advection has not been established.

J.1. Error handling

Here we introduce Table J.5, using the same test case used for Table 1, but with simpler error handling (without conditionals handling case by case considerations in division).

References

- [1] S.K. Godunov, I. Bohachevsky, Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics, *Mat. Sb.* 47 (89) (1959) 271–306.
- [2] A. Harten, J.M. Hyman, P.D. Lax, B. Keyfitz, On finite-difference approximations and entropy conditions for shocks, *Commun. Pure Appl. Math.* 29 (1976) 297–322.
- [3] A. Harten, On a class of high resolution total-variation-stable finite-difference schemes, *SIAM J. Numer. Anal.* 21 (1984) 1–23.
- [4] P.K. Sweby, High resolution schemes using flux limiters for hyperbolic conservation laws, *SIAM J. Numer. Anal.* 21 (1984) 995–1011.
- [5] J. Goodman, R. Leveque, On the accuracy of stable schemes for 2D scalar conservation laws, *Math. Comput.* 45 (1985) 15–21.
- [6] S. Spekreijse, Multigrid solution of monotone second-order discretizations of hyperbolic conservation laws, *Math. Comput.* 49 (1987) 135–155.
- [7] A. Jameson, Positive schemes and shock modelling for compressible flows, *Int. J. Numer. Methods Fluids* 20 (1995) 743–776.
- [8] W. Hundsdorfer, B. Koren, J. Verwer, et al., A positive finite-difference advection scheme, *J. Comput. Phys.* 117 (1995) 35–46.
- [9] P.K. Smolarkiewicz, P.J. Rasch, Monotone advection on the sphere: an Eulerian versus semi-Lagrangian approach, *J. Atmos. Sci.* 48 (1991) 793–810.
- [10] M. Zijlema, P. Wesseling, Higher-order flux-limiting schemes for the finite volume computation of incompressible flow, *Int. J. Comput. Fluid Dyn.* 9 (1998) 89–109.
- [11] B. Koren, A robust upwind discretization method for advection, diffusion and source terms, in: *Centrum voor Wiskunde en Informatica, Amsterdam*, 1993.
- [12] H.K. Versteeg, W. Malalasekera, *An Introduction to Computational Fluid Dynamics: the Finite Volume Method*, Pearson Education, 2007.
- [13] H. Deconinck, P.L. Roe, R. Struijs, A multidimensional generalization of Roe's flux difference splitter for the Euler equations, *Comput. Fluids* 22 (1993) 215–222.
- [14] W. Hua-mo, On the possible accuracy of TVD schemes, *J. Comput. Math.* (1992) 245–252.
- [15] C. Jacques, C. Frederic, Uniformly second order convergent schemes for hyperbolic conservation laws including Leonard's approach, in: *Nonlinear Hyperbolic Equations—Theory, Computation Methods, and Applications: Proceedings of the Second International Conference on Nonlinear Hyperbolic Problems*, Aachen, FRG, March 14 to 18, 1988, Springer, 1989, pp. 51–62.
- [16] R. Eymard, T. Gallouët, R. Herbin, Finite volume methods, in: J.L. Lions, P. Ciarlet (Eds.), *Solution of Equation in R^n (Part 3)*, *Techniques of Scientific Computing (Part 3)*, in: *Handbook of Numerical Analysis*, vol. 7, Elsevier, 2000, pp. 713–1020, <https://hal.archives-ouvertes.fr/hal-02100732>.
- [17] W. Hundsdorfer, R. Trompert, Method of lines and direct discretization: a comparison for linear advection, *Appl. Numer. Math.* 13 (1994) 469–490.
- [18] P. Wilders, G. Fotia, A positive spatial advection scheme on unstructured meshes for tracer transport, *J. Comput. Appl. Math.* 140 (2002) 809–821.
- [19] C. Bolley, M. Crouzeix, Conservation de la positivité lors de la discrétisation des problèmes d'évolution paraboliques, *RAIRO. Anal. Numér.* 12 (1978) 237–245.
- [20] M. Spijker, Contractivity in the numerical solution of initial value problems, *Numer. Math.* 42 (1983) 271–290.
- [21] J.F.B.M. Kraaijevanger, Contractivity of Runge-Kutta methods, *BIT Numer. Math.* 31 (1991) 482–528.
- [22] C.-W. Shu, S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, *J. Comput. Phys.* 77 (1988) 439–471.
- [23] L. Ferracina, M. Spijker, An extension and analysis of the Shu-Osher representation of Runge-Kutta methods, *Math. Comput.* 74 (2005) 201–219.
- [24] S. Osher, S. Chakravarthy, High resolution schemes and the entropy condition, *SIAM J. Numer. Anal.* 21 (1984) 955–984.
- [25] S.-J. Lin, R.B. Rood, Multidimensional flux-form semi-Lagrangian transport schemes, *Mon. Weather Rev.* 124 (1996) 2046–2070.
- [26] B. Leonard, A. Lock, M. MacVean, Conservative explicit unrestricted-time-step multidimensional constancy-preserving advection schemes, *Mon. Weather Rev.* 124 (1996) 2588–2606.
- [27] G.D. Van Albada, B.v. Leer, W. Roberts, A comparative study of computational methods in cosmic gas dynamics, in: *Upwind and High-Resolution Schemes*, Springer, 1997, pp. 95–103.
- [28] N. Waterson, H. Deconinck, A unified approach to the design and application of bounded higher-order convection schemes, *Numerical Methods in Laminar and Turbulent Flow* 9 (1995) 203–214.
- [29] A. Harten, B. Engquist, S. Osher, S.R. Chakravarthy, Uniformly high order accurate essentially non-oscillatory schemes, III, in: *Upwind and High-Resolution Schemes*, Springer, 1987, pp. 218–290.
- [30] D. Zhang, C. Jiang, D. Liang, L. Cheng, A review on TVD schemes and a refined flux-limiter for steady-state calculations, *J. Comput. Phys.* 302 (2015) 114–154.
- [31] F. Kemm, A comparative study of TVD-limiters—well-known limiters and an introduction of new ones, *Int. J. Numer. Methods Fluids* 67 (2011) 404–440.
- [32] R.J. Leveque, High-resolution conservative algorithms for advection in incompressible flow, *SIAM J. Numer. Anal.* 33 (1996) 627–665.
- [33] H. Nishikawa, A truncation error analysis of third-order MUSCL scheme for nonlinear conservation laws, *Int. J. Numer. Methods Fluids* 93 (2021) 1031–1052.
- [34] B. Leonard, Order of accuracy of QUICK and related convection-diffusion schemes, *Appl. Math. Model.* 19 (1995) 640–653.
- [35] H. Nishikawa, The QUICK scheme is a third-order finite-volume scheme with point-valued numerical solutions, *Int. J. Numer. Methods Fluids* 93 (2021) 2311–2338.
- [36] A. Donev, Third order upwind, <https://cims.nyu.edu/~donev/Teaching/CFD/Lectures/ThirdOrderUpwind.pdf>. (Accessed 26 May 2023).
- [37] T. Barth, M. Ohlberger, *Finite volume methods: foundation and analysis*, 2003.
- [38] F. Dubois, Nonlinear interpolation and total variation diminishing schemes, preprint, arXiv:1006.3921, 2010.
- [39] J.M. Hyman, Accurate monotonicity preserving cubic interpolation, *SIAM J. Sci. Stat. Comput.* 4 (1983) 645–654.
- [40] H.T. Huynh, Accurate monotone cubic interpolation, *SIAM J. Numer. Anal.* 30 (1993) 57–100.
- [41] M. Čada, M. Torrilhon, Compact third-order limiter functions for finite volume methods, *J. Comput. Phys.* 228 (2009) 4118–4145.