Maximum-principle-satisfying and positivity-preserving high order schemes for conservation laws: Survey and new developments

Xiangxiong Zhang[1] and Chi-Wang Shu[2]

[1]*Department of Mathematics, Brown University, Providence, RI 02912, USA.*
[2]*Division of Applied Mathematics, Brown University, Providence, RI 02912, USA.*

In Zhang & Shu (2010*b*), genuinely high order accurate finite volume and discontinuous Galerkin schemes satisfying a strict maximum principle for scalar conservation laws were developed. The main advantages of such schemes are their provable high order accuracy and their easiness for generalization to multi-dimensions for arbitrarily high order schemes on structured and unstructured meshes. The same idea can be used to construct high order schemes preserving the positivity of certain physical quantities, such as density and pressure for compressible Euler equations, water height for shallow water equations, and density for Vlasov-Boltzmann transport equations. These schemes have been applied in computational fluid dynamics, computational astronomy and astrophysics, plasma simulation, population models and traffic flow models. In this paper, we first review the main ideas of these maximum-principle-satisfying and positivity-preserving high order schemes, then present a simpler implementation which will result in a significant reduction of computational cost especially for weighted essentially nonoscillatory (WENO) finite volume schemes.

1. Introduction

An important property of the unique entropy solution to the scalar conservation law

$$u_t + \nabla \cdot \mathbf{F}(u) = 0, \quad u(\mathbf{x}, 0) = u_0(\mathbf{x}) \tag{1.1}$$

is that it satisfies a strict maximum principle, namely, if $M = \max_{\mathbf{x}} u_0(\mathbf{x})$, $m = \min_{\mathbf{x}} u_0(\mathbf{x})$, then $u(\mathbf{x}, t) \in [m, M]$ for any $\mathbf{x}$ and t. This property is also naturally desired for numerical schemes solving (1.1) since numerical solutions outside of $[m, M]$ often are meaningless physically, such as negative density, or negative percentage or percentage larger than one for a component in a multi-component mixture.

One of the main difficulties in solving (1.1) is that the solution may contain discontinuities even if the initial condition is smooth. Moreover, the weak solutions of (1.1) may not be unique. Therefore, the nonlinear stability and convergence to

Author for correspondence (shu@dam.brown.edu).

Proc. R. Soc. A 1–26; doi: 10.1098/rspa.00000000

the unique entropy solution must be considered for the numerical schemes. Total-variation (TV) stable functions form a compact space, so a conservative TV-stable scheme will produce a subsequence converging to a weak solution by the Lax-Wendroff Theorem. The E-schemes including Godunov, Lax-Friedrichs and Engquist-Osher methods satisfy an entropy inequality and are total-variation-diminishing (TVD) thus maximum-principle-satisfying. However, E-schemes are at most first order accurate. In fact, any TVD scheme in the sense of measuring the variation of grid point values or cell averages will be at most first order accurate around smooth extrema, see Osher & Chakravarthy (1984), although TVD schemes can be designed for any formal order of accuracy for smooth monotone solutions, e.g., the high resolution schemes.

For conventional maximum-principle-satisfying finite difference schemes, the solution is at most second order accurate, for instance, only the second order central scheme was proved to satisfy the maximum principle in Jiang & Tadmor (1998). This fact has a simple proof due to Ami Harten. For simplicity we consider a finite difference scheme, namely u_j^n is the numerical solution approximating the point values $u(x_j, t^n)$ of the exact solution, where n is the time step and j denotes the spatial grid index. Assume the scheme satisfies the maximum principle

$$\max_j u_j^{n+1} \le \max_j u_j^n. \tag{1.2}$$

Consider the linear convection equation $u_t + u_x = 0, u(x,0) = \sin(2\pi x), x \in [0,1]$ with periodic boundary conditions. Set the grid as $x_j = (j - \frac{1}{2})\Delta x$ where $\Delta x = \frac{1}{N}$ and N is a multiple of 4. The numerical initial value is $u_j^0 = \sin(2\pi x_j)$. Without loss of generality, assume $\Delta t = \frac{1}{2}\Delta x$. At the grid point $j = \frac{N}{4} + 1$ and $t = \Delta t$, the exact solution is $\sin(2\pi(x_j - \Delta t)) = \sin(2\pi((\frac{N}{4} + \frac{1}{2})\Delta x - \Delta t)) = \sin(\frac{\pi}{2}) = 1$ and the numerical solution is $u_j^1 \le \max_j u_j^0 = \sin(\frac{\pi}{2} - \frac{\pi}{N})$ by (1.2). The error of the scheme at the grid point $j = \frac{N}{4} + 1$ after one time step is equal to $|1 - u_j^1| = 1 - u_j^1 \ge 1 - \sin(\frac{\pi}{2} - \frac{\pi}{N}) = \frac{\pi^2}{2}\Delta x^2 + O(\Delta x^3)$. That is, even after one time step the scheme is already at most second order accurate. A similar proof also works for finite volume schemes where the numerical solution approximates cell averages of the exact solution.

The simple derivation above implies that (1.2) is too restrictive for the scheme to be higher than second order accurate. A heuristic point of view to understand the restriction is, some high order information of the exact solution is lost since we only measure the total variation or the maximum at the grid points or in cell averages. To overcome this difficulty, Sanders proposed to measure the total variation of the reconstructed polynomials and he succeeded in designing a third order TVD scheme for one-dimensional scalar conservation laws in Sanders (1988), which has been extended to higher order in Zhang & Shu (2010*a*). But it is very difficult to generalize Sanders' scheme to higher space dimension. By measuring the maximum of the reconstructed polynomial, Liu and Osher constructed a third order non-oscillatory scheme in Liu & Osher (1996), which could be generalized to two space dimensions. However, it could be proven maximum-principle-satisfying only for the linear equation. The key step of maximum-principle-satisfying high order schemes above is a high order accurate time evolution which preserves the maximum principle. The exact time

evolution satisfies this property and it was used in Sanders (1988), Zhang & Shu (2010*a*), Liu & Osher (1996). Unfortunately, it is very difficult, if not impossible, to implement such exact time evolution for multi-dimensional nonlinear scalar equations or systems of conservation laws.

Successful high order numerical schemes for hyperbolic conservation laws include, among others, the Runge-Kutta discontinuous Galerkin (RKDG) method with a total variation bounded (TVB) limiter, e.g. in Cockburn & Shu (1989), the essentially non-oscillatory (ENO) finite volume and finite difference schemes, e.g. in Harten *et al.* (1987), Shu & Osher (1988), and the weighted ENO (WENO) finite volume and finite difference schemes, e.g. in Liu *et al.* (1994), Jiang & Shu (1996). Although these schemes are nonlinearly stable in numerical experiments and some of them can be proven to be total variation stable, they do not in general satisfy a strict maximum principle. In Zhang & Shu (2010*b*), we proved a sufficient condition for the cell averages of the numerical solutions in a high order finite volume or a discontinuous Galerkin (DG) scheme with the strong stability preserving (SSP) time discretization, e.g., Shu & Osher (1988), Shu (1988), to be bounded in $[m, M]$ for (1.1). We have also proved that, with a simple scaling limiter introduced in Liu & Osher (1996), this sufficient condition can be enforced and not only the cell averages but also the numerical solution itself can be guaranteed to stay in $[m, M]$ without destroying accuracy for smooth solutions. In other words, we have constructed a high order scheme by adding a simple limiter to a finite volume WENO/ENO or RKDG scheme and it can be proven to be high order accurate and maximum-principle-satisfying. This was the first time that genuinely high order schemes are obtained satisfying a strict maximum principle especially for multidimensional nonlinear problems.

For hyperbolic conservation law systems, the entropy solutions in general do not satisfy the maximum principle. We consider the positivity of some important quantities instead. For instance, density and pressure in compressible Euler equations, and water height in shallow water equations should be nonnegative physically. In practice, failure of preserving positivity of such quantities may cause blow-ups of the computation because the linearized system may become ill-posed. From the point of view of stability, it is highly desired to design schemes which can be proven to be positivity-preserving. Most commonly used high order numerical schemes for solving hyperbolic conservation law systems do not in general satisfy such properties automatically. It is very difficult to design a conservative high order accurate scheme preserving the positivity. In Zhang & Shu (2010*c*, 2011) and Zhang *et al.* (2011), we have generalized the maximum-principle-satisfying techniques to construct conservative positivity-preserving high order finite volume and DG schemes for compressible Euler equations, which could be regarded as an extension of the positivity-preserving schemes in Perthame & Shu (1996).

In this paper, we first review the general framework to construct maximum-principle-satisfying and positivity-preserving schemes of arbitrarily high order accuracy. In §2, we illustrate the main ideas in the context of scalar conservation laws. We then discuss generalizations of this idea to other equations and systems in §3 and §4. In §5, we propose a more efficient implementation of the framework for WENO finite volume schemes, and provide numerical examples to demonstrate their performance. Concluding remarks are given in §6.

2. Maximum-principle-satisfying high order schemes for scalar conservation laws

(*a*) *One-dimensional scalar conservation laws*

We consider the one-dimensional version of (1.1) in this section:

$$u_t + f(u)_x = 0, \qquad u(x,0) = u_0(x). \tag{2.1}$$

(a.1) *The first order schemes*

It is well known that a first order monotone scheme solving (2.1) satisfies the strict maximum principle. A first order monotone scheme has the form

$$u_j^{n+1} = u_j^n - \lambda[\widehat{f}(u_j^n, u_{j+1}^n) - \widehat{f}(u_{j-1}^n, u_j^n)] \equiv H_\lambda(u_{j-1}^n, u_j^n, u_{j+1}^n), \tag{2.2}$$

where $\lambda = \frac{\Delta t}{\Delta x}$ with Δt and Δx being the temporal and spatial mesh sizes (we assume uniform mesh size for the structured mesh cases in this paper for simplicity in presentation, however the methodology does not have a uniform or smooth mesh restriction), and $\widehat{f}(a,b)$ is a monotone flux, namely it is Lipschitz continuous in both arguments, non-decreasing (henceforth referred to as increasing with a slight abuse of the terminology) in the first argument and non-increasing (henceforth referred to as decreasing) in the second argument, and consistent $\widehat{f}(a,a) = f(a)$. Under suitable CFL conditions, typically of the form

$$\alpha\lambda \le 1, \qquad \alpha = \max|f'(u)|, \tag{2.3}$$

for e.g. Lax-Friedrichs scheme and Godunov scheme, one can prove that the function $H_\lambda(a,b,c)$ is increasing in all three arguments, and consistency implies $H_\lambda(a,a,a) = a$. We therefore immediately have the strict maximum principle

$$m = H_\lambda(m,m,m) \le u_j^{n+1} = H_\lambda(u_{j-1}^n, u_j^n, u_{j+1}^n) \le H_\lambda(M,M,M) = M$$

provided $m \le u_{j-1}^n, u_j^n, u_{j+1}^n \le M$.

(a.2) *High order spatial discretization*

Now consider high order finite volume or DG methods, for example, the WENO finite volume method in Liu *et al.* (1994) and the DG method in Cockburn & Shu (1989) solving (2.1). We only discuss the Euler forward temporal discretization in this subsection and leave higher order temporal discretization to section §2 (a.4). The finite volume method or the scheme satisfied by the cell averages in the DG method discretization can be written as:

$$\overline{u}_j^{n+1} = \overline{u}_j^n - \lambda[\widehat{f}(u^-_{j+\frac{1}{2}}, u^+_{j+\frac{1}{2}}) - \widehat{f}(u^-_{j-\frac{1}{2}}, u^+_{j-\frac{1}{2}})] \equiv G_\lambda(\overline{u}_j^n, u^-_{j+\frac{1}{2}}, u^+_{j+\frac{1}{2}}, u^-_{j-\frac{1}{2}}, u^+_{j-\frac{1}{2}}), \tag{2.4}$$

where $\overline{u}_j^n$ is the approximation to the cell averages of $u(x,t)$ in the cell $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ at time level n, $\widehat{f}(\cdot,\cdot)$ is again a monotone flux, and $u^-_{j+\frac{1}{2}}$, $u^+_{j+\frac{1}{2}}$ are the high order approximations of the nodal values $u(x_{j+\frac{1}{2}}, t^n)$ within the cells I_j and I_{j+1} respectively. These values are either reconstructed from the cell averages $\overline{u}_j^n$ in a finite volume method or read directly from the evolved polynomials in a DG method. We assume that there is a polynomial $p_j(x)$ (either reconstructed in

a finite volume method or evolved in a DG method) with degree k, where $k \geq 1$, defined on I_j such that $\overline{u}_j^n$ is the cell average of $p_j(x)$ on I_j, $u^+_{j-\frac{1}{2}} = p_j(x_{j-\frac{1}{2}})$ and $u^-_{j+\frac{1}{2}} = p_j(x_{j+\frac{1}{2}})$.

Given a scheme in the form of (2.4), assuming $\overline{u}_j^n \in [m, M]$ for all j, we would like to derive some sufficient conditions to ensure $\overline{u}_j^{n+1} \in [m, M]$. A very natural first attempt is to see if there is a restriction on λ such that, if all five arguments of G are in $[m, M]$

$$m \leq \overline{u}_j^n, u^-_{j+\frac{1}{2}}, u^+_{j+\frac{1}{2}}, u^-_{j-\frac{1}{2}}, u^+_{j-\frac{1}{2}} \leq M,$$

then we could prove $\overline{u}_j^{n+1} \in [m, M]$. Unfortunately, one can easily build counter examples to show that this cannot be always true. The problem is that the function $G_\lambda(a, b, c, d, e)$ in (2.4) is only monotonically increasing in the first, third and fourth arguments and is monotonically decreasing in the other two arguments. Hence the strategy to prove maximum principle for first order monotone schemes cannot be repeated here. In the literature, many attempts have been made to further limit the four arguments $u^-_{j+\frac{1}{2}}, u^+_{j+\frac{1}{2}}, u^-_{j-\frac{1}{2}}, u^+_{j-\frac{1}{2}}$ (remember the cell average $\overline{u}_j^n$ cannot be changed due to conservation) in the arguments of G_λ in (2.4) to guarantee that $\overline{u}_j^{n+1} \in [m, M]$. However, these limiters always kill accuracy near smooth extrema.

Our approach follows a different strategy. We consider an N-point Legendre Gauss-Lobatto quadrature rule on the interval $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$, which is exact for the integral of polynomials of degree up to $2N-3$. We denote these quadrature points on I_j as

$$S_j = \{x_{j-\frac{1}{2}} = \widehat{x}_j^1, \widehat{x}_j^2, \cdots, \widehat{x}_j^{N-1}, \widehat{x}_j^N = x_{j+\frac{1}{2}}\}. \tag{2.5}$$

Let $\widehat{w}_\alpha$ be the quadrature weights for the interval $[-\frac{1}{2}, \frac{1}{2}]$ such that $\sum_{\alpha=1}^{N} \widehat{w}_\alpha = 1$. Choose N to be the smallest integer satisfying $2N - 3 \geq k$, then

$$\overline{u}_j^n = \frac{1}{\Delta x} \int_{I_j} p_j(x)\, dx = \sum_{\alpha=1}^{N} \widehat{w}_\alpha p_j(\widehat{x}_j^\alpha) = \sum_{\alpha=2}^{N-1} \widehat{w}_\alpha p_j(\widehat{x}_j^\alpha) + \widehat{w}_1 u^+_{j-\frac{1}{2}} + \widehat{w}_N u^-_{j+\frac{1}{2}}. \tag{2.6}$$

We then have the following theorem. We assume that the monotone flux $\widehat{f}$ corresponds to a monotone scheme (2.2) under the CFL condition (2.3).

THEOREM 1. *Consider a finite volume scheme or the scheme satisfied by the cell averages of the DG method (2.4), associated with the approximation polynomials $p_j(x)$ of degree k (either reconstruction or DG polynomials) in the sense that $\overline{u}_j^n = \frac{1}{\Delta x}\int_{I_j} p_j(x)dx$, $u^+_{j-\frac{1}{2}} = p_j(x_{j-\frac{1}{2}})$ and $u^-_{j+\frac{1}{2}} = p_j(x_{j+\frac{1}{2}})$. If $u^-_{j-\frac{1}{2}}$, $u^+_{j+\frac{1}{2}}$ and $p_j(\widehat{x}_j^\alpha)$ $(\alpha = 1, \cdots, N)$ are all in the range $[m, M]$, then $\overline{u}_j^{n+1} \in [m, M]$ under the CFL condition*

$$\lambda a \leq \widehat{w}_1. \tag{2.7}$$

Proof. With (2.6), by adding and subtracting $\widehat{f}\left(u^+_{j-\frac{1}{2}}, u^-_{j+\frac{1}{2}}\right)$, the scheme (2.4) can be rewritten as

$$\begin{aligned}
\overline{u}_j^{n+1} &= \sum_{\alpha=2}^{N-1} \widehat{w}_\alpha p_j(\widehat{x}_j^\alpha) + \widehat{w}_N \left(u^-_{j+\frac{1}{2}} - \frac{\lambda}{\widehat{w}_N} \left[\widehat{f}\left(u^-_{j+\frac{1}{2}}, u^+_{j+\frac{1}{2}}\right) - \widehat{f}\left(u^+_{j-\frac{1}{2}}, u^-_{j+\frac{1}{2}}\right) \right] \right) \\
&\quad + \widehat{w}_1 \left(u^+_{j-\frac{1}{2}} - \frac{\lambda}{\widehat{w}_1} \left[\widehat{f}\left(u^+_{j-\frac{1}{2}}, u^-_{j+\frac{1}{2}}\right) - \widehat{f}\left(u^-_{j-\frac{1}{2}}, u^+_{j-\frac{1}{2}}\right) \right] \right) \\
&= \sum_{\alpha=2}^{N-1} \widehat{w}_\alpha p_j(\widehat{x}_j^\alpha) + \widehat{w}_N H_{\lambda/\widehat{\omega}_N}(u^+_{j-\frac{1}{2}}, u^-_{j+\frac{1}{2}}, u^+_{j+\frac{1}{2}}) + \widehat{w}_1 H_{\lambda/\widehat{\omega}_1}(u^-_{j-\frac{1}{2}}, u^+_{j-\frac{1}{2}}, u^-_{j+\frac{1}{2}}).
\end{aligned} \tag{2.8}$$

Noticing that $\widehat{w}_1 = \widehat{w}_N$ and $H_{\lambda/\widehat{\omega}_1}$ is monotone under the CFL condition (2.7), we can see from (2.8) that $\overline{u}_j^{n+1}$ is a monotonically increasing function of all the arguments involved, namely $u^-_{j-\frac{1}{2}}$, $u^+_{j+\frac{1}{2}}$ and $p_j(\widehat{x}_j^\alpha)$ for $1 \le j \le N$. The same proof for the first order monotone scheme now applies to imply $\overline{u}_j^{n+1} \in [m, M]$. ■

Remark We recall that the CFL condition for linear stability for the DG scheme using polynomial of degree k is $\lambda a \le \frac{1}{2k+1}$ in Cockburn & Shu (1989), which is close to the CFL condition (2.7).

(a.3) *The linear scaling limiter*

Theorem 1 tells us that for the scheme (2.4), we need to modify $p_j(x)$ such that $p_j(x) \in [m, M]$ for all $x \in S_j$ where S_j is defined in (2.5). For all j, assume $\overline{u}_j^n \in [m, M]$, we use the modified polynomial $\widetilde{p}_j(x)$ by the limiter introduced in Liu & Osher (1996), i.e.,

$$\widetilde{p}_j(x) = \theta(p_j(x) - \overline{u}_j^n) + \overline{u}_j^n, \qquad \theta = \min\left\{ \left| \frac{M - \overline{u}_j^n}{M_j - \overline{u}_j^n} \right|, \left| \frac{m - \overline{u}_j^n}{m_j - \overline{u}_j^n} \right|, 1 \right\}, \tag{2.9}$$

with

$$M_j = \max_{x \in I_j} p_j(x), \qquad m_j = \min_{x \in I_j} p_j(x). \tag{2.10}$$

Let $\widetilde{u}^+_{j-\frac{1}{2}} = \widetilde{p}_j(x_{j-\frac{1}{2}})$ and $\widetilde{u}^-_{j+\frac{1}{2}} = \widetilde{p}_j(x_{j+\frac{1}{2}})$. We get the revised scheme of (2.4):

$$\overline{u}_j^{n+1} = \overline{u}_j^n - \lambda[\widehat{f}(\widetilde{u}^-_{j+\frac{1}{2}}, \widetilde{u}^+_{j+\frac{1}{2}}) - \widehat{f}(\widetilde{u}^-_{j-\frac{1}{2}}, \widetilde{u}^+_{j-\frac{1}{2}})]. \tag{2.11}$$

The scheme (2.11) satisfies the sufficient condition in theorem 1. We will show in the next lemma that this limiter does not destroy the uniform high order of accuracy.

LEMMA 1. *Assume $\overline{u}_j^n \in [m, M]$, then (2.9)-(2.10) gives a $(k+1)$-th order accurate limiter.*

Proof. We need to show $\widetilde{p}_j(x) - p_j(x) = O(\Delta x^{k+1})$ for any $x \in I_j$. We only prove the case that $p_j(x)$ is not a constant and $\theta = \left|\frac{M-\overline{u}_j^n}{M_j-\overline{u}_j^n}\right|$, the other cases being similar. Since $\overline{u}_j^n \le M$ and $\overline{u}_j^n \le M_j$, we have $\theta = (M - \overline{u}_j^n)/(M_j - \overline{u}_j^n)$. Therefore,

$$\begin{aligned}
\widetilde{p}_j(x) - p_j(x) &= \theta(p_j(x) - \overline{u}_j^n) + \overline{u}_j^n - p_j(x) \\
&= (\theta - 1)(p_j(x) - \overline{u}_j^n) \\
&= \frac{M - M_j}{M_j - \overline{u}_j^n}(p_j(x) - \overline{u}_j^n) \\
&= (M - M_j)\frac{p_j(x) - \overline{u}_j^n}{M_j - \overline{u}_j^n}.
\end{aligned}$$

By the definition of θ in (2.9), $\theta = \left|\frac{M-\overline{u}_j^n}{M_j-\overline{u}_j^n}\right|$ implies that $\theta = \left|\frac{M-\overline{u}_j^n}{M_j-\overline{u}_j^n}\right| < 1$, i.e. there is an overshoot $M_j > M$, and the overshoot $M_j - M = O(\Delta x^{k+1})$ since $p_j(x)$ is an approximation with error $O(\Delta x^{k+1})$. Thus we only need to prove that $\left|\frac{p_j(x)-\overline{u}_j^n}{M_j-\overline{u}_j^n}\right| \le C_k$, where C_k is a constant depending only on the polynomial degree k. In Liu & Osher (1996), $C_2 = 3$ is proved. We now prove the existence of C_k for any k. Assume $p_j(x) = a_0 + a_1(\frac{x-x_j}{\Delta x}) + \cdots + a_k(\frac{x-x_j}{\Delta x})^k$ and $p(x) = a_0 + a_1 x + \cdots + a_k x^k$, then the cell average of $p(x)$ on $I = [-\frac{1}{2}, \frac{1}{2}]$ is $\overline{p} = \overline{u}_j^n$ and $\max_{x\in I} p(x) = M_j$. So we have

$$\max_{x\in I_j}\left|\frac{p_j(x) - \overline{u}_j^n}{M_j - \overline{u}_j^n}\right| = \max_{x\in I}\left|\frac{p(x) - \overline{p}}{\max_{y\in I} p(y) - \overline{p}}\right|.$$

Let $q(x) = p(x) - \overline{p}$, then it suffices to prove the existence of C_k such that

$$\left|\frac{\min_{x\in I} p(x) - \overline{p}}{\max_{x\in I} p(x) - \overline{p}}\right| = \left|\frac{\min_{x\in I} q(x)}{\max_{x\in I} q(x)}\right| \le C_k.$$

It is easy to check that $|\min_{x\in I} q(x)|$ and $|\max_{x\in I} q(x)|$ are both norms on the finite dimensional linear space consisting of all polynomials of degree k whose averages on the interval I are zero. Any two norms on this finite dimensional space are equivalent, hence their ratio is bounded by a constant C_k. ■

Notice that in (2.10) we need to evaluate the maximum/minimum of a polynomial. We prefer to avoid evaluating the extrema of a polynomial, especially since we will extend the method to two dimensions. Since we only need to control the values at several points, we could replace (2.10) by

$$M_j = \max_{x\in S_j} p_j(x), \qquad m_j = \min_{x\in S_j} p_j(x), \tag{2.12}$$

and the limiter (2.9) and (2.12) is sufficient to enforce $\widetilde{p}_j(x) \in [m, M], \forall x \in S_j$. As to the accuracy, (2.12) is a less restrictive limiter than (2.10), so the accuracy will

not be destroyed. Also, it is a conservative limiter because it does not change the cell average of the polynomial.

For the conservative maximum-principle-satisfying scheme (2.11), it is straightforward to prove the following stability result:

THEOREM 2. *Assuming periodic or zero boundary conditions, then the numerical solution of (2.11) satisfies*

$$\sum_j |\overline{u}_j^{n+1} - m| = \sum_j |\overline{u}_j^n - m|, \quad \sum_j |\overline{u}_j^{n+1} - M| = \sum_j |\overline{u}_j^n - M|.$$

Proof. Taking the sum of (2.11) over j, we obtain $\sum_j \overline{u}_j^{n+1} = \sum_j \overline{u}_j^n$. Since the numerical solutions are maximum-principle-satisfying, namely, $\overline{u}_j^{n+1}, \overline{u}_j^n \in [m, M]$, we have

$$\sum_j |\overline{u}^{n+1} - m| = \sum_j (\overline{u}^{n+1} - m) = \sum_j (\overline{u}^n - m) = \sum_j |\overline{u}^n - m|.$$

The other equality follows similarly. ■

Remark As an easy corollary, if the solution is non-negative, namely if $m \geq 0$, then we have the L^1 stability $\sum_j |\overline{u}_j^{n+1}| = \sum_j |\overline{u}_j^n|$.

(a.4) *High order temporal discretization*

We use strong stability preserving (SSP) high order time discretizations. For more details, see Shu & Osher (1988), Shu (1988). For example, the third order SSP Runge-Kutta method in Shu & Osher (1988) (with the CFL coefficient $c = 1$) is

$$\begin{aligned} u^{(1)} &= u^n + \Delta t F(u^n) \\ u^{(2)} &= \frac{3}{4}u^n + \frac{1}{4}(u^{(1)} + \Delta t F(u^{(1)}) \\ u^{n+1} &= \frac{1}{3}u^n + \frac{2}{3}(u^{(2)} + \Delta t F(u^{(2)})) \end{aligned}$$

where $F(u)$ is the spatial operator, and the third order SSP multi-step method in Shu (1988) (with the CFL coefficient $c = \frac{1}{3}$) is

$$u^{n+1} = \frac{16}{27}(u^n + 3\Delta t F(u^n)) + \frac{11}{27}(u^{n-3} + \frac{12}{11}\Delta t F(u^{n-3})).$$

Here, the CFL coefficient c for a SSP time discretization refers to the fact that, if we assume the Euler forward time discretization for solving the equation $u_t = F(u)$ is stable in a norm or a semi-norm under a time step restriction $\Delta t \leq \Delta t_0$, then the high order SSP time discretization is also stable in the same norm or semi-norm under the time step restriction $\Delta t \leq c\Delta t_0$.

Since a SSP high order time discretization is a convex combinations of Euler forward, the full scheme with a high order SSP time discretization will still satisfy the maximum principle. The limiter (2.9) and (2.12) should be used for each stage in a Runge-Kutta method or each step in a multi-step method. For details of the implementation, see Zhang & Shu (2010*b*).

(b) Two-dimensional extensions

Consider the two-dimensional scalar conservation laws $u_t + f(u)_x + g(u)_y = 0, u(x,y,0) = u_0(x,y)$ with $M = \max_{x,y} u_0(x,y), m = \min_{x,y} u_0(x,y)$. We only discuss the DG method with the Euler forward time discretization in this section, but all the results also hold for the finite volume scheme (e.g. ENO and WENO).

(b.1) *Rectangular meshes*

For simplicity we assume we have a uniform rectangular mesh. At time level n, we have an approximation polynomial $p_{ij}(x,y)$ of degree k with the cell average $\overline{u}_{ij}^n$ on the (i,j) cell $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$. Let $u^+_{i-\frac{1}{2},j}(y), u^-_{i+\frac{1}{2},j}(y)$, $u^+_{i,j-\frac{1}{2}}(x)$, $u^-_{i,j+\frac{1}{2}}(x)$ denote the traces of $p_{ij}(x,y)$ on the four edges respectively. A finite volume scheme or the scheme satisfied by the cell averages of a DG method on a rectangular mesh can be written as

$$\begin{aligned}\overline{u}_{ij}^{n+1} &= \overline{u}_{ij}^n - \frac{\Delta t}{\Delta x \Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \widehat{f}\left[u^-_{i+\frac{1}{2},j}(y), u^+_{i+\frac{1}{2},j}(y)\right] - \widehat{f}\left[u^-_{i-\frac{1}{2},j}(y), u^+_{i-\frac{1}{2},j}(y)\right] dy \\ &\quad - \frac{\Delta t}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \widehat{g}\left[u^-_{i,j+\frac{1}{2}}(x), u^+_{i,j+\frac{1}{2}}(x)\right] - \widehat{g}\left[u^-_{i,j-\frac{1}{2}}(x), u^+_{i,j-\frac{1}{2}}(x)\right] dx,\end{aligned}$$

where $\widehat{f}(\cdot,\cdot)$, $\widehat{g}(\cdot,\cdot)$ are one dimensional monotone fluxes. The integrals can be approximated by quadratures with sufficient accuracy. Let us assume that we use a Gauss quadrature with L points, which is exact for single variable polynomials of degree k. We assume $S_i^x = \{x_i^\beta : \beta = 1, \cdots, L\}$ denote the Gauss quadrature points on $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, and $S_j^y = \{y_j^\beta : \beta = 1, \cdots, L\}$ denote the Gauss quadrature points on $[y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$. For instance, $(x_{i-\frac{1}{2}}, y_j^\beta)$ $(\beta = 1, \cdots, L)$ are the Gauss quadrature points on the left edge of the (i,j) cell. The subscript β will denote the values at the Gauss quadrature points, for instance, $u^+_{i-\frac{1}{2},\beta} = u^+_{i-\frac{1}{2},j}(y_j^\beta)$. Also, w_β denotes the corresponding quadrature weight on interval $[-\frac{1}{2}, \frac{1}{2}]$, so that $\sum_{\beta=1}^L w_\beta = 1$. We will still need to use the N-point Gauss-Lobatto quadrature rule where N is the smallest integer satisfying $2N - 3 \geq k$, and we distinguish the two quadrature rules by adding hats to the Gauss-Lobatto points, i.e., $\widehat{S}_i^x = \{\widehat{x}_i^\alpha : \alpha = 1, \cdots, N\}$ will denote the Gauss-Lobatto quadrature points on $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, and $\widehat{S}_j^y = \{\widehat{y}_j^\alpha : \alpha = 1, \cdots, N\}$ will denote the Gauss-Lobatto quadrature points on $[y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$. Subscripts or superscripts β will be used only for Gauss quadrature points and α only for Gauss-Lobatto points.

Let $\lambda_1 = \frac{\Delta t}{\Delta x}$ and $\lambda_2 = \frac{\Delta t}{\Delta y}$, then the scheme becomes

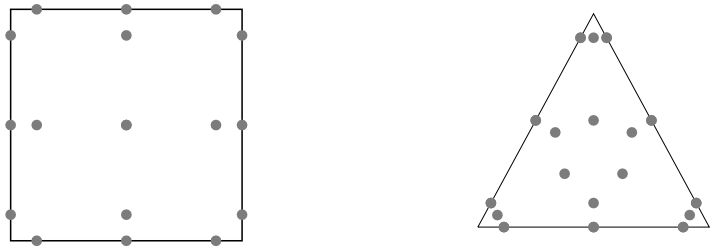
$$\overline{u}_{ij}^{n+1} = \overline{u}_{ij}^n - \lambda_1 \sum_{\beta=1}^L w_\beta \left[\widehat{f}(u^-_{i+\frac{1}{2},\beta}, u^+_{i+\frac{1}{2},\beta}) - \widehat{f}(u^-_{i-\frac{1}{2},\beta}, u^+_{i-\frac{1}{2},\beta})\right]$$

$$-\lambda_2 \sum_{\beta=1}^{L} w_\beta \left[\widehat{g}(u^-_{\beta,j+\frac{1}{2}}, u^+_{\beta,j+\frac{1}{2}}) - \widehat{g}(u^-_{\beta,j-\frac{1}{2}}, u^+_{\beta,j-\frac{1}{2}}) \right]. \tag{2.13}$$

We want to find a sufficient condition for the scheme (2.13) to satisfy $\overline{u}_{ij}^{n+1} \in [m, M]$. We use $\otimes$ to denote the tensor product, for instance, $S_i^x \otimes S_j^y = \{(x, y) : x \in S_i^x, y \in S_j^y\}$. Define the set S_{ij} as

$$S_{ij} = (S_i^x \otimes \widehat{S}_j^y) \cup (\widehat{S}_i^x \otimes S_j^y). \tag{2.14}$$

See figure 1(a) for an illustration for $k = 2$. For simplicity, let $\mu_1 = \frac{\lambda_1 a_1}{\lambda_1 a_1 + \lambda_2 a_2}$ and



(a) S_{ij} in (2.14). (b) S_K^k in (2.20) for $k = 2$.

Figure 1. Points to decompose the cell averages for two-variable quadratic polynomials.

$\mu_2 = \frac{\lambda_2 a_2}{\lambda_1 a_1 + \lambda_2 a_2}$ where $a_1 = \max |f'(u)|$ and $a_2 = \max |g'(u)|$. Notice that $\widehat{w}_1 = \widehat{w}_N$, we have

$$\begin{aligned}
\overline{u}_{ij}^n &= \frac{\mu_1}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} p_{ij}(x, y) dy dx + \frac{\mu_2}{\Delta x \Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} p_{ij}(x, y) dx dy \\
&= \mu_1 \sum_{\beta=1}^{L} \sum_{\alpha=1}^{N} w_\beta \widehat{w}_\alpha p_{ij}(\widehat{x}_i^\alpha, y_j^\beta) + \mu_2 \sum_{\beta=1}^{L} \sum_{\alpha=1}^{N} w_\beta \widehat{w}_\alpha p_{ij}(x_i^\beta, \widehat{y}_j^\alpha) \\
&= \sum_{\beta=1}^{L} \sum_{\alpha=2}^{N-1} w_\beta \widehat{w}_\alpha \left[\mu_1 p_{ij}(\widehat{x}_i^\alpha, y_j^\beta) + \mu_2 p_{ij}(x_i^\beta, \widehat{y}_j^\alpha) \right] \\
&\quad + \sum_{\beta=1}^{L} w_\beta \widehat{w}_1 \left[\mu_1 u^-_{i+\frac{1}{2},\beta} + \mu_1 u^+_{i-\frac{1}{2},\beta} + \mu_2 u^-_{\beta,j+\frac{1}{2}} + \mu_2 u^+_{\beta,j-\frac{1}{2}} \right]
\end{aligned} \tag{2.15}$$

THEOREM 3. *Consider a two-dimensional finite volume scheme or the scheme satisfied by the cell averages of the DG method on rectangular meshes (2.13), associated with the approximation polynomials $p_{ij}(x, y)$ of degree k (either*

reconstruction or DG polynomials). If $u^{\pm}_{\beta,j\pm\frac{1}{2}}$, $u^{\pm}_{i\pm\frac{1}{2},\beta}\in[m,M]$ *and* $p_{ij}(x,y)\in[m,M]$ *(for any* $(x,y)\in S_{ij}$*), then* $\overline{u}^{n+1}_j\in[m,M]$ *under the CFL condition*

$$\lambda_1 a_1+\lambda_2 a_2\le \widehat{w}_1. \tag{2.16}$$

Proof. Plugging (2.15) in, (2.13) can be written as

$$\begin{aligned}
\overline{u}^{n+1}_{ij} &= \sum_{\beta=1}^{L}\sum_{\alpha=2}^{N-1} w_\beta\widehat{w}_\alpha\left[\mu_1 p_{ij}(\widehat{x}^\alpha_i,y^\beta_j)+\mu_2 p_{ij}(x^\beta_i,\widehat{y}^\alpha_j)\right]\\
&\quad+\mu_1\sum_{\beta=1}^{L} w_\beta\widehat{w}_1\left[u^-_{i+\frac{1}{2},\beta}-\frac{\lambda_1}{\mu_1\widehat{w}_1}\left(\widehat{f}(u^-_{i+\frac{1}{2},\beta},u^+_{i+\frac{1}{2},\beta})-\widehat{f}(u^+_{i-\frac{1}{2},\beta},u^-_{i+\frac{1}{2},\beta})\right)\right.\\
&\quad\left.+\,u^+_{i-\frac{1}{2},\beta}-\frac{\lambda_1}{\mu_1\widehat{w}_1}\left(\widehat{f}(u^+_{i-\frac{1}{2},\beta},u^-_{i+\frac{1}{2},\beta})-\widehat{f}(u^-_{i-\frac{1}{2},\beta},u^+_{i-\frac{1}{2},\beta})\right)\right]\\
&\quad+\mu_2\sum_{\beta=1}^{L} w_\beta\widehat{w}_2\left[u^-_{\beta,j+\frac{1}{2}}-\frac{\lambda_2}{\mu_2\widehat{w}_1}\left(\widehat{g}(u^-_{\beta,j+\frac{1}{2}},u^+_{\beta,j+\frac{1}{2}})-\widehat{g}(u^+_{\beta,j-\frac{1}{2}},u^-_{\beta,j+\frac{1}{2}})\right)\right.\\
&\quad\left.+\,u^+_{\beta,j-\frac{1}{2}}-\frac{\lambda_2}{\mu_2\widehat{w}_1}\left(\widehat{g}(u^+_{\beta,j-\frac{1}{2}},u^-_{\beta,j+\frac{1}{2}})-\widehat{g}(u^-_{\beta,j-\frac{1}{2}},u^+_{\beta,j-\frac{1}{2}})\right)\right]
\end{aligned}$$

Following the same arguments as in theorem 1, it is easy to check that the formulation above for $\overline{u}^{n+1}_{ij}$ is a monotonically increasing function with respect to all the arguments $u^{\pm}_{\beta,j\pm\frac{1}{2}}$, $u^{\pm}_{i\pm\frac{1}{2},\beta}$, $p_{ij}(x^\beta_i,\widehat{y}^\alpha_j)$ and $p_{ij}(\widehat{x}^\alpha_i,y^\beta_j)$. ■

To enforce the condition in theorem 3, we can use the following scaling limiter similar to the 1D case. For all i and j, assuming the cell averages $\overline{u}^n_{ij}\in[m,M]$, we use the modified polynomial $\widetilde{p}_{ij}(x,y)$ instead of $p_{ij}(x,y)$, i.e.,

$$\widetilde{p}_{ij}(x,y)=\theta(p_{ij}(x,y)-\overline{u}^n_{ij})+\overline{u}^n_{ij},\qquad \theta=\min\left\{\left|\frac{M-\overline{u}^n_{ij}}{M_{ij}-\overline{u}^n_{ij}}\right|,\left|\frac{m-\overline{u}^n_{ij}}{m_{ij}-\overline{u}^n_{ij}}\right|,1\right\}, \tag{2.17}$$

with

$$M_{ij}=\max_{(x,y)\in S_{ij}} p_{ij}(x,y),\qquad m_{ij}=\min_{(x,y)\in S_{ij}} p_{ij}(x,y). \tag{2.18}$$

It is also straightforward to prove the high order accuracy of this limiter following the proof of lemma 1.

(b.2) *Triangular meshes*

For each triangle K we denote by l^i_K $(i=1,2,3)$ the length of its three edges e^i_K $(i=1,2,3)$, with outward unit normal vector ν^i $(i=1,2,3)$. $K(i)$ denotes the neighboring triangle along e^i_K and $|K|$ is the area of the triangle K. Let $\widehat{F}(u,v,\nu)$ be a one dimensional monotone flux in the ν direction (e.g. Lax-Friedrichs flux), namely $\widehat{F}(u,v,\nu)$ is an increasing function of the first argument and a

decreasing function of the second argument, It satisfies $\widehat{F}(u,v,\nu) = -\widehat{F}(v,u,-\nu)$ (conservativity), and $\widehat{F}(u,u,\nu) = \mathbf{F}(u)\cdot\nu$ (consistency), with $\mathbf{F}(u) = \langle f(u), g(u)\rangle$. The first order monotone scheme can be written as

$$u_K^{n+1} = u_K^n - \frac{\Delta t}{|K|}\sum_{i=1}^{3}\widehat{F}(u_K^n, u_{K(i)}^n, \nu^i) l_K^i = H(u_K^n, u_{K(1)}^n, u_{K(2)}^n, u_{K(3)}^n).$$

Then $H(\cdot,\cdot,\cdot,\cdot)$ is a monotonically increasing function with respect to each argument under the CFL condition $a\frac{\Delta t}{|K|}\sum\limits_{i=1}^{3} l_K^i \le 1$ where $a = \max|\langle f'(u), g'(u)\rangle|$.

A high order finite volume scheme or a scheme satisfied by the cell averages of a DG method, with first order Euler forward time discretization, can be written as

$$\overline{u}_K^{n+1} = \overline{u}_K^n - \frac{\Delta t}{|K|}\sum_{i=1}^{3}\int_{e_K^i}\widehat{F}(u_i^{int(K)}, u_i^{ext(K)}, \nu^i)ds,$$

where $\overline{u}_K^n$ is the cell average over K of the numerical solution, and $u_i^{int(K)}, u_i^{ext(K)}$ are the approximations to the values on the edge e_K^i obtained from the interior and the exterior of K. Assume the DG polynomial on the triangle K is $p_K(x,y)$ of degree k, then in the DG method, the edge integral should be approximated by the $(k+1)$-point Gauss quadrature. The scheme becomes

$$\overline{u}_K^{n+1} = \overline{u}_K^n - \frac{\Delta t}{|K|}\sum_{i=1}^{3}\sum_{\beta=1}^{k+1}\widehat{F}(u_{i,\beta}^{int(K)}, u_{i,\beta}^{ext(K)}, \nu^i) w_\beta l_K^i, \tag{2.19}$$

where w_β denote the $(k+1)$-point Gauss quadrature weights on the interval $[-\frac{1}{2}, \frac{1}{2}]$, so that $\sum\limits_{\beta=1}^{k+1} w_\beta = 1$, and $u_{i,\beta}^{int(K)}$ and $u_{i,\beta}^{ext(K)}$ denote the values of u evaluated at the β-th Gauss quadrature point on the i-th edge from the interior and exterior of the element K respectively.

Motivated by the derivation in the previous subsection, to find a sufficient condition for the scheme (2.19) to satisfy $\overline{u}_K^{n+1} \in [m, M]$, we need to decompose the cell average $\overline{u}_K^n$ by a quadrature rule which include all the Gauss quadrature points for each edge e_K^i with all the quadrature weights being positive. Such a quadrature can be constructed by mapping the Gauss tensor Gauss-Lobatto points on a rectangle to a triangle. Details of the mapping can be found in Zhang *et al.* (2011). In the barycentric coordinates, the set S_K^k of quadrature points for polynomials of degree k on a triangle K can be written as

$$\begin{aligned} S_K^k \;=\; & \left\{\left(\frac{1}{2}+v^\beta, (\frac{1}{2}+\widehat{u}^\alpha)(\frac{1}{2}-v^\beta), (\frac{1}{2}-\widehat{u}^\alpha)(\frac{1}{2}-v^\beta)\right),\right. \\ & \left.\left((\frac{1}{2}-\widehat{u}^\alpha)(\frac{1}{2}-v^\beta), \frac{1}{2}+v^\beta, (\frac{1}{2}+\widehat{u}^\alpha)(\frac{1}{2}-v^\beta)\right),\right. \end{aligned}$$

$$\left((\frac{1}{2}+\widehat{u}^{\alpha})(\frac{1}{2}-v^{\beta}),(\frac{1}{2}-\widehat{u}^{\alpha})(\frac{1}{2}-v^{\beta}),\frac{1}{2}+v^{\beta}\right)\Bigg\} \tag{2.20}$$

where u^{α} $(\alpha=1,\cdots,N)$ and v^{β} $(\beta=1,\cdots,k+1)$ are the Gauss-Lobatto and Gauss quadrature points on the interval $[-\frac{1}{2},\frac{1}{2}]$ respectively. See figure 1(b) for an illustration of S_K^2.

THEOREM 4. *For the scheme (2.19) with the polynomial $p_K(x,y)$ (either reconstruction or DG polynomial) of degree k to satisfy the maximum principle $m\le\overline{u}_K^{n+1}\le M$, a sufficient condition is that each $p_K(x,y)$ satisfies $p_K(x,y)\in[m,M]$, $\forall(x,y)\in S_K^k$ where S_K^k is defined in (2.20), under the CFL condition $a\frac{\Delta t}{|K|}\sum_{i=1}^{3}l_K^i\le\frac{2}{3}\widehat{w}_1$. Here $\widehat{w}_1$ is still the quadrature weight of the N-point Gauss-Lobatto rule on $[-\frac{1}{2},\frac{1}{2}]$ for the first quadrature point.*

The proof is similar to that for the structured mesh cases, see Zhang *et al.* (2011) for the details. We can still use the same scaling limiter to enforce this sufficient condition.

3. Positivity-preserving high order schemes for compressible Euler equations in gas dynamics

(*a*) *Ideal gas*

The one-dimensional Euler system for the perfect gas is given by

$$\mathbf{w}_t+\mathbf{f}(\mathbf{w})_x=0,\quad t\ge0,\ x\in\mathbb{R}, \tag{3.1}$$

where $\mathbf{w}=(\rho,m,E)^T$, $\mathbf{f}(\mathbf{w})=(m,\rho u^2+p,(E+p)u)^T$, $m=\rho u$, $E=\frac{1}{2}\rho u^2+\rho e$, $p=(\gamma-1)\rho e$, ρ is the density, u is the velocity, m is the momentum, E is the total energy, p is the pressure, e is the internal energy, and $\gamma>1$ is a constant ($\gamma=1.4$ for the air). The speed of sound is given by $c=\sqrt{\gamma p/\rho}$ and the three eigenvalues of the Jacobian $\mathbf{f}'(\mathbf{w})$ are $u-c$, u and $u+c$.

Let $p(\mathbf{w})=(\gamma-1)(E-\frac{1}{2}\frac{m^2}{\rho})$ be the pressure function. It can be easily verified that p is a concave function of $\mathbf{w}=(\rho,m,E)^T$ if $\rho\ge0$. Define the set of admissible states by $G=\left\{\mathbf{w}|\,\rho>0\quad\text{and}\quad p=(\gamma-1)\left(E-\frac{1}{2}\frac{m^2}{\rho}\right)>0\right\}$, then G is a convex set. If the density or pressure becomes negative, the system (3.1) will be non-hyperbolic and thus the initial value problem will be ill-posed. In this section we discuss only the perfect gas case, leaving the discussion for general gases to §3 (b).

We are interested in schemes for (3.1) producing the numerical solutions in the admissible set G. We start with a first order scheme

$$\mathbf{w}_j^{n+1}=\mathbf{w}_j^n-\lambda[\widehat{\mathbf{f}}(\mathbf{w}_j^n,\mathbf{w}_{j+1}^n)-\widehat{\mathbf{f}}(\mathbf{w}_{j-1}^n,\mathbf{w}_j^n)], \tag{3.2}$$

where $\widehat{\mathbf{f}}(\cdot,\cdot)$ is a numerical flux. The scheme (3.2) and its numerical flux $\widehat{\mathbf{f}}(\cdot,\cdot)$ are called positivity preserving, if the numerical solution $\mathbf{w}_j^n$ being in the set G for

all j implies the solution $\mathbf{w}_j^{n+1}$ being also in the set G. This is usually achieved under a standard CFL condition

$$\lambda \parallel (|u| + c) \parallel_\infty \leq \alpha_0 \tag{3.3}$$

where α_0 is a constant related to the specific scheme. Examples of positivity preserving fluxes include the Godunov flux, the Lax-Friedrichs flux, the Boltzmann type flux, and the Harten-Lax-van Leer flux, see Perthame & Shu (1996). In Zhang & Shu (2010*c*), we proved that the Lax-Friedrichs flux is positivity preserving with $\alpha_0 = 1$.

In Perthame & Shu (1996), a high order scheme preserving the positivity was proposed, but it is quite difficult to implement the method, especially in multi-dimensions. In Zhang & Shu (2010*c*, 2011), Zhang *et al.* (2011), we generalized the ideas in the previous section to construct high order schemes preserving the positivity of density and pressure for the Euler system.

(a.1) *One-dimensional compressible Euler equations*

First, we consider the first order Euler forward time discretization. A general high order finite volume scheme, or the scheme satisfied by the cell averages of a DG method solving (3.1), has the following form

$$\overline{\mathbf{w}}_j^{n+1} = \overline{\mathbf{w}}_j^n - \lambda \left[\widehat{\mathbf{f}} \left(\mathbf{w}_{j+\frac{1}{2}}^-, \mathbf{w}_{j+\frac{1}{2}}^+ \right) - \widehat{\mathbf{f}} \left(\mathbf{w}_{j-\frac{1}{2}}^-, \mathbf{w}_{j-\frac{1}{2}}^+ \right) \right], \tag{3.4}$$

where $\widehat{\mathbf{f}}$ is a positivity preserving flux under the CFL condition (3.3), $\overline{\mathbf{w}}_j^n$ is the approximation to the cell average of the exact solution $\mathbf{v}(x,t)$ in the cell $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ at time level n, and $\mathbf{w}_{j+\frac{1}{2}}^-$, $\mathbf{w}_{j+\frac{1}{2}}^+$ are the high order approximations of the point values $\mathbf{v}(x_{j+\frac{1}{2}}, t^n)$ within the cells I_j and I_{j+1} respectively. These values are either reconstructed from the cell averages $\overline{\mathbf{w}}_j^n$ in a finite volume method or read directly from the evolved polynomials in a DG method. We assume that there is a polynomial vector $\mathbf{q}_j(x) = (\rho_j(x), m_j(x), E_j(x))^T$ (either reconstructed in a finite volume method or evolved in a DG method) with degree k, where $k \geq 1$, defined on I_j such that $\overline{\mathbf{w}}_j^n$ is the cell average of $\mathbf{q}_j(x)$ on I_j, $\mathbf{w}_{j-\frac{1}{2}}^+ = \mathbf{q}_j(x_{j-\frac{1}{2}})$ and $\mathbf{w}_{j+\frac{1}{2}}^- = \mathbf{q}_j(x_{j+\frac{1}{2}})$. Next, we state a similar result as in the previous section:

THEOREM 5. *For a finite volume scheme or the scheme satisfied by the cell averages of a DG method (3.4), if* $\mathbf{q}_j(\widehat{x}_j^\alpha) \in G$ *for all* j *and* α, *then* $\overline{\mathbf{w}}_j^{n+1} \in G$ *under the CFL condition*

$$\lambda \parallel (|u| + c) \parallel_\infty \leq \widehat{w}_1 \alpha_0.$$

The proof is similar to that for theorem 1 and can be found in Zhang & Shu (2010*c*).

Strong stability preserving high order Runge-Kutta in Shu & Osher (1988) and multi-step in Shu (1988) time discretization will keep the validity of theorem 5 since G is convex. If the numerical solutions have positive density and pressure, it follows that the scheme is L^1 stable for the density ρ and the total energy E due to theorem 2.

(a.2) *A limiter to enforce the sufficient condition*

Given the vector of approximation polynomials $\mathbf{q}_j(x) = (\rho_j(x), m_j(x), E_j(x))^T$, with its cell average $\overline{\mathbf{w}}_j^n = (\overline{\rho}_j^n, \overline{m}_j^n, \overline{E}_j^n)^T \in G$, we would like to modify $\mathbf{q}_j(x)$ into $\widetilde{\mathbf{q}}_j(x)$ such that it satisfies the sufficient condition in theorem 5 without destroying the cell averages and high order accuracy.

Define $\overline{p}_j^n = (\gamma-1)\left(\overline{E}_j^n - \frac{1}{2}(\overline{m}_j^n)^2/\overline{\rho}_j^n\right)$. Then $\overline{\rho}_j^n > 0$ and $\overline{p}_j^n > 0$ for all j. Assume there exists a small number $\varepsilon > 0$ such that $\overline{\rho}_j^n \geq \varepsilon$ and $\overline{p}_j^n \geq \varepsilon$ for all j. For example, we can take $\varepsilon = 10^{-13}$ in the computation.

The first step is to limit the density. Replace $\rho_j(x)$ by

$$\widehat{\rho}_j(x) = \theta_1(\rho_j(x) - \overline{\rho}_j^n) + \overline{\rho}_j^n, \quad \theta_1 = \min\left\{\frac{\overline{\rho}_j^n - \varepsilon}{\overline{\rho}_j^n - \rho_{\min}}, 1\right\}, \quad \rho_{\min} = \min_{\alpha} \rho_j(\widehat{x}_j^\alpha). \tag{3.5}$$

Then the cell average of $\widehat{\rho}_j(x)$ over I_j is still $\overline{\rho}_j^n$ and $\widehat{\rho}_j(\widehat{x}_j^\alpha) \geq \varepsilon$ for all α.

The second step is to enforce the positivity of the pressure. We need to introduce some notations. Let $\widehat{\mathbf{q}}_j(x) = (\widehat{\rho}_j(x), m_j(x), E_j(x))^T$ and $\widehat{\mathbf{q}}_j^\alpha$ denote $\widehat{\mathbf{q}}_j(\widehat{x}_j^\alpha)$. Define $G^\varepsilon = \left\{\mathbf{w} : \rho \geq \varepsilon, p = (\gamma-1)\left(E - \frac{1}{2}\frac{m^2}{\rho}\right) \geq \varepsilon\right\}$, $\partial G^\varepsilon = \left\{\mathbf{w} : \rho \geq \varepsilon, p = \varepsilon\right\}$, and

$$\mathbf{s}^\alpha(t) = (1-t)\overline{\mathbf{w}}_j^n + t\widehat{\mathbf{q}}_j(\widehat{x}_j^\alpha), \quad 0 \leq t \leq 1. \tag{3.6}$$

∂G^ε is a surface and $\mathbf{s}^\alpha(t)$ is the straight line passing through the two points $\overline{\mathbf{w}}_j^n$ and $\widehat{\mathbf{q}}_j(\widehat{x}_j^\alpha)$. If $\widehat{\mathbf{q}}_j(\widehat{x}_j^\alpha) \notin G^\varepsilon$, then the straight line $\mathbf{s}^\alpha(t)$ intersects with the surface ∂G^ε at one and only one point since G^ε is a convex set. If $\widehat{\mathbf{q}}_j(\widehat{x}_j^\alpha) \notin G^\varepsilon$, let t_ε^α denote the parameter in (3.6) corresponding to the intersection point; otherwise let $t_\varepsilon^\alpha = 1$. We only need to solve a quadratic equation to find t_ε^α, see Zhang & Shu (2010*c*) for details. Now we define

$$\widetilde{\mathbf{q}}_j(x) = \theta_2\left(\widehat{\mathbf{q}}_j(x) - \overline{\mathbf{w}}_j^n\right) + \overline{\mathbf{w}}_j^n, \quad \theta_2 = \min_{\alpha=1,2,\cdots,N} t_\varepsilon^\alpha. \tag{3.7}$$

It is easy to check that the cell average of $\widetilde{\mathbf{q}}_j(x)$ over I_j is $\overline{\mathbf{w}}_j^n$ and $\widetilde{\mathbf{q}}_j(\widehat{x}_j^\alpha) \in G$ for all α. See Zhang & Shu (2010*c*) for the proof of the accuracy.

(a.3) *Two-dimensional cases*

In this section we extend our result to finite volume or DG schemes of $(k+1)$-th order accuracy solving two-dimensional Euler equations

$$\mathbf{w}_t + \mathbf{f}(\mathbf{w})_x + \mathbf{g}(\mathbf{w})_y = 0, \quad t \geq 0, (x,y) \in \mathbb{R}^2, \tag{3.8}$$

$$\mathbf{w} = \begin{pmatrix} \rho \\ m \\ n \\ E \end{pmatrix}, \quad \mathbf{f}(\mathbf{w}) = \begin{pmatrix} m \\ \rho u^2 + p \\ \rho uv \\ (E+p)u \end{pmatrix}, \quad \mathbf{g}(\mathbf{w}) = \begin{pmatrix} n \\ \rho uv \\ \rho v^2 + p \\ (E+p)v \end{pmatrix}$$

where $m = \rho u, n = \rho v, E = \frac{1}{2}\rho u^2 + \frac{1}{2}\rho v^2 + \rho e, p = (\gamma-1)\rho e$, and $\langle u, v\rangle$ is the velocity. The eigenvalues of the Jacobian $\mathbf{f}'(\mathbf{w})$ are $u-c$, u, u and $u+c$ and the eigenvalues of the Jacobian $\mathbf{g}'(\mathbf{w})$ are $v-c$, v, v and $v+c$. The pressure

function p is still concave with respect to $\mathbf{w}$ if $\rho \geq 0$ and the set of admissible states $G = \{\mathbf{w} | \rho > 0, p > 0\}$ is still convex.

With the same notions as in §2, a finite volume scheme or the scheme satisfied by the cell averages of a DG method for (3.8) can be written as, on a rectangular mesh,

$$\begin{aligned} \overline{\mathbf{w}}_{ij}^{n+1} &= \overline{\mathbf{w}}_{ij}^{n} - \lambda_1 \sum_{\beta=1}^{L} w_\beta \left[\widehat{\mathbf{f}}\left(\mathbf{w}^-_{i+\frac{1}{2},\beta}, \mathbf{w}^+_{i+\frac{1}{2},\beta} \right) - \widehat{\mathbf{f}}\left(\mathbf{w}^-_{i-\frac{1}{2},\beta}, \mathbf{w}^+_{i-\frac{1}{2},\beta} \right) \right] \\ &\quad -\lambda_2 \sum_{\beta=1}^{L} w_\beta \left[\widehat{\mathbf{g}}\left(\mathbf{w}^-_{\beta,j+\frac{1}{2}}, \mathbf{w}^+_{\beta,j+\frac{1}{2}} \right) - \widehat{\mathbf{g}}\left(\mathbf{w}^-_{\beta,j-\frac{1}{2}}, \mathbf{w}^+_{\beta,j-\frac{1}{2}} \right) \right], \quad (3.9) \end{aligned}$$

or on a triangular mesh,

$$\overline{\mathbf{w}}_K^{n+1} = \overline{\mathbf{w}}_K^{n} - \frac{\Delta t}{|K|} \sum_{i=1}^{3} \sum_{\beta=1}^{k+1} \widehat{\mathbf{F}}(\mathbf{w}_{i,\beta}^{int(K)}, \mathbf{w}_{i,\beta}^{ext(K)}, \nu^i) w_\beta l_K^i. \quad (3.10)$$

Assume at time level n there are approximation polynomials of degree k, $\mathbf{q}_{ij}(x,y)$ with the cell average $\overline{\mathbf{w}}_{ij}^n$ on the (i,j) rectangular cell, or $\mathbf{q}_K(x,y)$ with the cell average $\overline{\mathbf{w}}_K^n$ on the triangle K, let $a_1 = \| (|u| + c) \|_\infty$, $a_2 = \| (|v| + c) \|_\infty$ and $a = \| (|\langle u, v \rangle| + c) \|_\infty$, then we have the following

THEOREM 6. *For a finite volume scheme or the scheme satisfied by the cell averages of a DG method (3.9) on a rectangle, if $\mathbf{q}_{ij}(x,y) \in G$ for all i,j and $(x,y) \in S_{ij}$ defined in (2.14), then $\overline{\mathbf{w}}_{ij}^{n+1} \in G$ under the CFL condition $\lambda_1 a_1 + \lambda_2 a_2 \leq \widehat{w}_1$.*

THEOREM 7. *For a finite volume scheme or the scheme satisfied by the cell averages of a DG method (3.10) on a triangle, if $\mathbf{q}_K(x,y) \in G$ for all K and $(x,y) \in S_K^k$ defined in (2.20), then $\overline{\mathbf{w}}_K^{n+1} \in G$ under the CFL condition $a \frac{\Delta t}{|K|} \sum_{i=1}^{3} l_K^i \leq \frac{2}{3} \widehat{w}_1$.*

We can construct the same type of limiters as in the previous subsection to enforce the sufficient conditions in these two theorems. See Zhang & Shu (2010*c*), Zhang *et al.* (2011) for the proof of the theorems and implementation of limiters.

(*b*) *General equations of state and source terms*

Now we consider the one-dimensional Euler system (3.1) with a general equation of state $E = \rho e(\rho, p) + \frac{1}{2}\rho u^2$ where $e(\rho, p)$ is the internal energy. As we have seen in the previous subsection, to construct high order schemes preserving the positivity of density and pressure, there are four important steps:

1 Prove $G = \{\mathbf{w} : \rho > 0 \quad \text{and} \quad p > 0\}$ is a convex set.

2 Prove the first order scheme (3.2) preserves the positivity.

3 Find a sufficient condition for the Euler forward time discretization as in theorem 5. Then high order SSP Runge-Kutta or multi-step will keep the positivity due to the convexity of G.

4 Construct a limiter to enforce the sufficient condition as in (3.5) and (3.7).

Notice that step 3 and step 4 above both heavily depend on the convexity of G. Therefore, to easily generalize the previous results to general equations of state, we should not give up the convexity. In Zhang & Shu (2011), we proved steps 1 and 2 will hold for any equation of state satisfying $e \geq 0 \Leftrightarrow p \geq 0$ if $\rho \geq 0$. Once step 1 and step 2 are valid, it is very straightforward to complete step 3 and step 4 by following the ideas in theorems 5, (3.5) and (3.7). Two-dimensional extensions are also trivial by following theorem 6 and theorem 7.

For Euler equations with source terms, for instance, the axial symmetry, gravity, chemical reaction or cooling effect, it is still possible to construct positivity-preserving high order schemes. It is straightforward to extend all the previous results to Euler systems with various source terms, see Zhang & Shu (2011).

4. Applications

(*a*) *Maximum-principle-satisfying high order schemes for passive convection equations with a divergence free velocity field*

We will discuss how to take advantage of maximum-principle-satisfying high order schemes for scalar conservation laws to construct such schemes for passive convection equations with a divergence free velocity field. We will explain the main idea for the two-dimensional incompressible Euler equation.

(a.1) *Two-dimensional incompressible Euler equation*

The two dimensional incompressible Euler equations in the vorticity stream-function formulation are given by:

$$\omega_t + (u\omega)_x + (v\omega)_y = 0, \tag{4.1}$$

$$\Delta\psi = \omega, \qquad \langle u, v\rangle = \langle -\psi_y, \psi_x\rangle, \tag{4.2}$$

with suitable initial and boundary conditions. The definition of $\langle u, v\rangle$ in (4.2) gives us the divergence-free condition $u_x + v_y = 0$, which implies (4.1) is equivalent to the non-conservative form

$$\omega_t + u\omega_x + v\omega_y = 0. \tag{4.3}$$

The exact solution of (4.3) satisfies the maximum principle $\omega(x, y, t) \in [m, M]$, for all (x, y, t), where $m = \min_{x,y} \omega_0(x, y)$ and $M = \max_{x,y} \omega_0(x, y)$. For discontinuous solutions or solutions containing sharp gradient regions, it is preferable to solve the conservative form (4.1) rather than the nonconservative form (4.3). However, without the incompressibility condition $u_x + v_y = 0$, the conservative form (4.1) itself does not imply the maximum principle $\omega(x, y, t) \in [m, M]$ for all (x, y, t). This is the main difficulty to get a maximum-principle-satisfying scheme solving the conservative form (4.1) directly.

We recall the high order discontinuous Galerkin method solving (4.1) in Liu & Shu (2000) briefly. For simplicity, we only discuss triangular meshes here. First, solve (4.2) by a standard Poisson solver for the stream-function ψ using continuous

finite elements, then take $u=-\psi_y, v=\psi_x$. Notice that on the boundary of each cell, $\langle u,v\rangle\cdot\nu=\langle-\psi_y,\psi_x\rangle\cdot\nu=\frac{\partial\psi}{\partial\tau}$, which is the tangential derivative. Thus $\langle u,v\rangle\cdot\nu$ is continuous across the cell boundary since the tangential derivative of ψ along each edge is continuous. The cell average scheme with Euler forward in time of the DG method in Liu & Shu (2000) is equivalent to

$$\overline{\omega}_K^{n+1}=\overline{\omega}_K^n-\frac{\Delta t}{|K|}\sum_{i=1}^{3}\sum_{\beta=1}^{k+1}h\left(\omega_{i,\beta}^{int(K)},\omega_{i,\beta}^{ext(K)},\mathbf{u}_\beta\cdot\nu^i\right)w_\beta l_K^i. \tag{4.4}$$

Suppose $\omega_K^n(x,y)$ is the DG polynomial on the triangle K. Then we can show that the right hand side of (4.4) is a monotonically increasing function of the values of $\omega_K^n(x,y)$ evaluated at S_K^k in (2.20). See Zhang *et al.* (2011) for the proof. Therefore, to have $\overline{\omega}_K^{n+1}\in[m,M]$, we only need to show the right hand side of (4.4) is consistent. Namely, it is equal to M if $\omega_K^n(x,y)=M$, $\forall(x,y)\in S_K^k$. This fact was proved in Zhang *et al.* (2011). We therefore have the following theorem.

THEOREM 8. *For a finite volume scheme or the scheme satisfied by the cell averages of a DG method (4.4) solving (4.1) on a triangle, if* $\omega_K^n(x,y)\in[m,M]$, $\forall(x,y)\in S_K^k$ *defined in (2.20), then* $\overline{\omega}_K^{n+1}\in[m,M]$ *under the CFL condition* $a\frac{\Delta t}{|K|}\sum_{i=1}^{3}l_K^i\le\frac{2}{3}\widehat{w}_1$.

Remark 1 The same result on rectangular meshes as in theorem 3 also holds, see Zhang & Shu (2010*b*).

Remark 2 If one chooses another method to solve the velocity field, then the result still holds as long as the quadrature rules are exact for the velocity field in the scheme. This can be easily achieved if we pre-process the divergence-free velocity field so that it is piecewise polynomial of the right degree for accuracy, continuous in the normal component across cell boundaries, and pointwise divergence-free.

(a.2) *The level set equation with a divergence free velocity field*

Let $\phi(t,x,y,z)=0$ define the implicit interface, then the Eulerian formulation of the interface evolution can be written as

$$\phi_t+(u\phi)_x+(v\phi)_y+(w\phi)_z=0,\quad \phi(0,x,y,z)=\phi_0(x,y,z). \tag{4.5}$$

If the velocity field satisfies $u_x+v_y+w_z=0$, then the solution of (4.5) satisfies the maximum principle, i.e., $\phi\in[m,M]$ where m and M are the minimum and maximum of ϕ_0. With the same idea, it is straightforward to construct maximum-principle-satisfying high order finite volume or DG schemes solving (4.5).

(a.3) *Vlasov-Poisson equations*

To describe the evolution of the electron distribution function $f(x,v,t)$ of a collisionless quasi-neutral plasma in one space and one velocity dimension where the ions have been assumed to be stationary, the Vlasov-Possion system is given

by

$$f_t + (vf)_x - (Ef)_v = 0, \tag{4.6}$$

$$E(x,t) = -\phi(x,t)_x, \quad \phi_{xx} = \int_{-\infty}^{\infty} f(x,v,t)\, dv - 1.$$

The exact solution of (4.6) also satisfies the maximum principle, which implies that the exact solution should always be non-negative. The positivity of the numerical solution for solving (4.6) is very difficult to achieve without destroying the conservation and high order accuracy, as indicated in Banks & Hittinger (2010). Since $v_x = E_v = 0$, the equation (4.6) is the same type as (4.1). Thus theorem 8 also applies to (4.6). See figure 2 for the result of the positivity-preserving fifth order finite volume WENO schemes for the two stream instability problem. The implementation detail of positivity-preserving limiter can be found in §5. As we can see, the traditional WENO schemes will produce negative values, which was also reported in Banks & Hittinger (2010). The positivity preserving high order scheme guarantees non-negativity and the result is comparable to those in Banks & Hittinger (2010), Rossmanith & Seal (2011).

Even though theorem 8 is only for the Eulerian schemes solving (4.6), the positivity-preserving techniques can also be extended to semi-Lagrangian schemes, see Rossmanith & Seal (2011), Qiu & Shu (2011).

(*b*) *Shallow water equations*

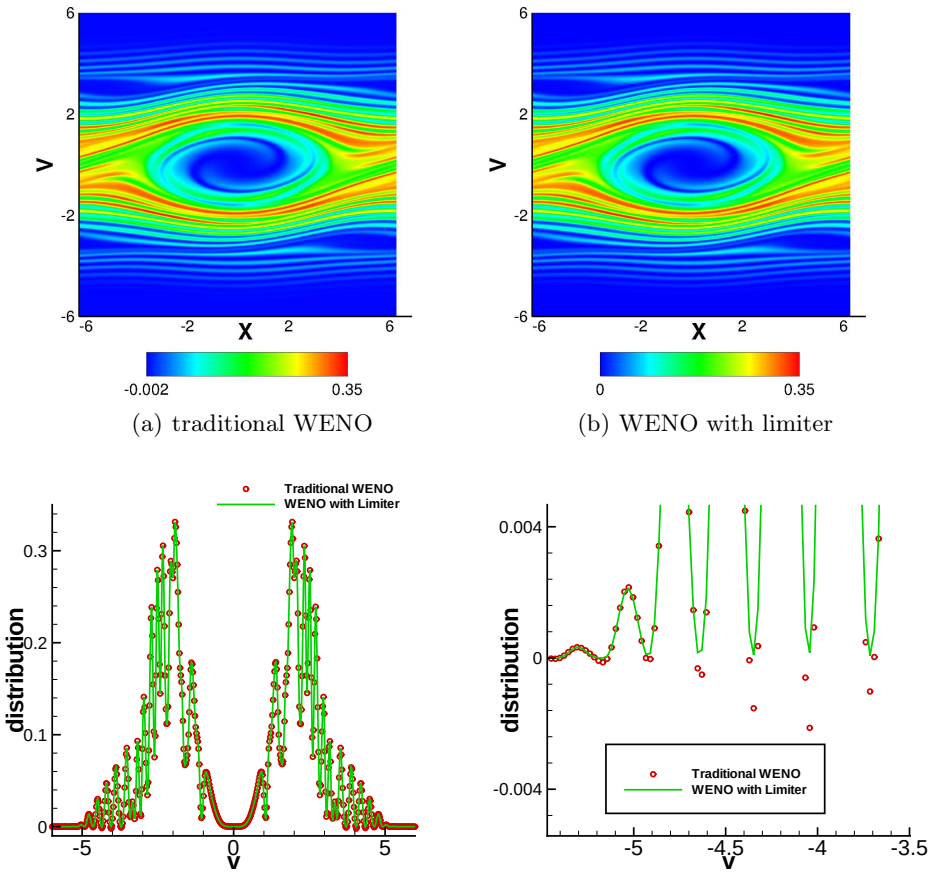
The shallow water equation with a non-flat bottom topography has been widely used to model flows in rivers and coastal areas. The water height is supposed to be non-negative during the time evolution. If it ever becomes negative, the computation will break down quite often since the initial value problem for the linearized system will be ill-posed. The positivity-preserving techniques can be also applied to one or two-dimensional shallow water equations. In Xing *et al.* (2010), we constructed high order DG schemes which preserves the well-balanced property and the non-negativity of the water height.

(*c*) *Vlasov-Boltzmann transport equations*

The Vlasov-Boltzmann transport equations describe the evolution of a probability distribution function $f(x,v,t)$ representing the probability of finding a particle at time t with position at x and phase velocity v. It models a dilute or rarefied gaseous state corresponding to a probabilistic description when the transport is given by a classical Hamiltonian with accelerations component given by the action of a Lorentzian force and particle interactions taken into account as a collision operator. Following the ideas described in previous sections, a high order positivity-preserving DG method was proposed in Cheng *et al.* (2010).

(*d*) *Positivity-preserving schemes for a population model*

When the numerical solutions denote the density or numbers, it is desired to have non-negative solutions. In Zhang *et al.* (2010), a positivity-preserving high order WENO schemes was constructed for a hierarchical size-structured population model, which involve global terms through integrals in the equation and boundary conditions.



(a) traditional WENO

(b) WENO with limiter

(c) Cut along $x = 0$

(d) Cut along $x = 0$ (zoomed)

Figure 2. Vlasov-Poisson: two stream instability at $T = 45$. The third order Runge-Kutta and fifth order finite volume WENO scheme on a 512×512 mesh.

5. A simplified implementation of the maximum-principle-satisfying and positivity-preserving limiter for WENO finite volume schemes

(*a*) *Motivation*

As described in previous sections, the maximum-principle-satisfying and positivity-preserving high order finite volume or discontinuous Galerkin schemes are easy to implement if the approximation polynomials are available. In the DG method, these are simply the DG polynomials. In the finite volume ENO schemes, the polynomials are constructed during the reconstruction procedure. However, the WENO reconstruction returns only some point values rather than approximation polynomials. Therefore, to implement the maximum-principle-satisfying and positivity-preserving high order WENO schemes according to the procedure described in the previous sections, one must first obtain the approximation polynomials beyond the WENO reconstructed point values, for example, by constructing interpolation polynomials as we did in Zhang & Shu (2010*b*). Thus implementation of the limiter for WENO schemes is more expensive and cumbersome especially for multi-dimensional problems. In this section, we will propose an alternative and simpler implementation to achieve the same maximum principle or positivity without using the approximation polynomials explicitly, which results in a reduction of computational cost and complexity of the procedure for WENO schemes and even for the DG method.

Let us revisit maximum-principle-satisfying schemes for the one-dimensional scalar conservation laws in §2. To have $\overline{u}_j^{n+1} \in [m, M]$, $p_j(\widehat{x}_j^\alpha) \in [m, M]$ for all α is sufficient but not necessary. By the mean value theorem, there exists some $x_j^* \in I_j$ such that $p_j(x_j^*) = \frac{1}{1-2\widehat{w}_1} \sum_{\alpha=2}^{N-1} \widehat{w}_\alpha p_j(\widehat{x}_j^\alpha)$. Then (2.8) can be rewritten as

$$\overline{u}_j^{n+1} = (1-2\widehat{w}_1) p_j(x_j^*) + \widehat{w}_N H_{\lambda/\widehat{\omega}_N}(u^+_{j-\frac{1}{2}}, u^-_{j+\frac{1}{2}}, u^+_{j+\frac{1}{2}}) + \widehat{w}_1 H_{\lambda/\widehat{\omega}_1}(u^-_{j-\frac{1}{2}}, u^+_{j-\frac{1}{2}}, u^-_{j+\frac{1}{2}}). \tag{5.1}$$

Therefore, we can have a much weaker sufficient condition.

THEOREM 9. *For the scheme (2.4), if* $p_j(x_j^*), u^\pm_{j\pm\frac{1}{2}}, u^\pm_{j\mp\frac{1}{2}} \in [m, M]$ *then* $\overline{u}_j^{n+1} \in [m, M]$ *under the CFL condition* $\lambda a \le \widehat{w}_1$.

To enforce this new sufficient condition, we can use the same limiter (2.9) with M_j and m_j redefined as

$$M_j = \max\{p_j(x_j^*), u^-_{j+\frac{1}{2}}, u^+_{j-\frac{1}{2}}\}, \quad m_j = \min\{p_j(x_j^*), u^-_{j+\frac{1}{2}}, u^+_{j-\frac{1}{2}}\}. \tag{5.2}$$

(2.9) and (5.2) will not destroy the high order accuracy since it is a less restrictive limiter than (2.9) and (2.10).

Equation (2.6) implies that $p_j(x_j^*) = \frac{\overline{u}_j^n - \widehat{w}_1 u^+_{j-\frac{1}{2}} - \widehat{w}_N u^-_{j+\frac{1}{2}}}{1-2\widehat{w}_1}$. Therefore, θ defined in (2.9) and (5.2) can be calculated without the explicit expression of the approximation polynomial $p_j(x)$ or the location x_j^*. We only need to know the existence of such polynomials to prove the accuracy of the limiter. For WENO schemes, the existence of such approximation polynomials can be established by the interpolation, for example, Hermite interpolation for the one-dimensional case as in Zhang & Shu (2010*b*).

Extensions to two-dimensional cases are straightforward:

THEOREM 10. *Consider the scheme (2.13). There exists some point* (x_i^*, y_j^*) *in the* (i,j) *cell such that*

$$p_{ij}(x_i^*, y_j^*) = \frac{\overline{u}_{ij}^n - \sum_{\beta=1}^{L} w_\beta \widehat{w}_1 \left[\mu_1 \left(u^-_{i+\frac{1}{2},\beta} + u^+_{i-\frac{1}{2},\beta} \right) + \mu_2 \left(u^-_{\beta,j+\frac{1}{2}} + u^+_{\beta,j-\frac{1}{2}} \right) \right]}{1 - 2\widehat{w}_1}. \tag{5.3}$$

If $p_{ij}(x_i^*, y_j^*), u^\pm_{\beta,j\pm\frac{1}{2}}, u^\pm_{i\pm\frac{1}{2},\beta}, u^\pm_{\beta,j\mp\frac{1}{2}}, u^\pm_{i\mp\frac{1}{2},\beta} \in [m, M]$, *then* $\overline{u}_{ij}^{n+1} \in [m, M]$ *under the CFL condition* $\lambda_1 a_1 + \lambda_2 a_2 \le \widehat{w}_1$.

THEOREM 11. *Consider the scheme (2.19). There exists some point* (x_K^*, y_K^*) *in the triangle* K *such that*

$$p_K(x_K^*, y_K^*) = \frac{\overline{u}_K^n - \sum_{i=1}^{3} \sum_{\beta=1}^{k+1} \frac{2}{3} w_\beta \widehat{w}_1 u_{i,\beta}^{int(K)}}{1 - 2\widehat{w}_1}.$$

If $p_K(x_K^*, y_K^*), u_{i,\beta}^{int(K)}, u_{i,\beta}^{ext(K)} \in [m, M]$, *then* $\overline{u}_K^{n+1} \in [m, M]$ *under the CFL condition* $a\frac{\Delta t}{|K|} \sum_{i=1}^{3} l_K^i \le \frac{2}{3}\widehat{w}_1$.

Remark All the results above can also be easily extended to positivity-preserving schemes for compressible Euler equations.

(*b*) *Easy implementation for WENO finite volume schemes*

We only state the algorithm for two-dimensional scalar conservation laws on rectangular meshes, the counterparts for the triangular meshes and compressible Euler equations are similar. For each stage in the SSP Runge-Kutta or each step in the SSP multi-step methods of the finite volume WENO schemes (2.13), the algorithm flowchart of the new limiter is

1. For each rectangle, given $\overline{u}_{ij}^n \in [m, M]$ and $u^\pm_{\beta,j\mp\frac{1}{2}}, u^\pm_{i\mp\frac{1}{2},\beta}$ constructed by the WENO reconstruction, compute $\theta_{ij} = \min\left\{ \left| \frac{M - \overline{u}_{ij}^n}{M_{ij} - \overline{u}_{ij}^n} \right|, \left| \frac{m - \overline{u}_{ij}^n}{m_{ij} - \overline{u}_{ij}^n} \right|, 1 \right\}$ with (5.3) where M_{ij} and m_{ij} are the max and min of $\left\{ p_{ij}(x_i^*, y_j^*), u^\pm_{i\mp\frac{1}{2},\beta}, u^\pm_{\beta,j\mp\frac{1}{2}} \right\}$.
2. Set $\widetilde{u}^\pm_{i\mp\frac{1}{2},\beta} = \theta_{ij}(u^\pm_{i\mp\frac{1}{2},\beta} - \overline{u}_{ij}^n) + \overline{u}_{ij}^n$ and $\widetilde{u}^\pm_{\beta,j\mp\frac{1}{2}} = \theta_{ij}(u^\pm_{\beta,j\mp\frac{1}{2}} - \overline{u}_{ij}^n) + \overline{u}_{ij}^n$.
3. Replace $u^\pm_{i\mp\frac{1}{2},\beta}, u^\pm_{\beta,j\mp\frac{1}{2}}, u^\pm_{i\pm\frac{1}{2},\beta}, u^\pm_{\beta,j\pm\frac{1}{2}}$ by the revised nodal values $\widetilde{u}^\pm_{i\mp\frac{1}{2},\beta}$, $\widetilde{u}^\pm_{\beta,j\mp\frac{1}{2}}$, $\widetilde{u}^\pm_{i\pm\frac{1}{2},\beta}$, $\widetilde{u}^\pm_{\beta,j\pm\frac{1}{2}}$ in the scheme (2.13).

Remark 1 The new algorithm is simpler and less expensive than the implementation in Zhang & Shu (2010*b*), since no extra reconstructions need to be performed for the limiter.

Remark 2 The new algorithm is also cheaper for the DG method because it avoids the evaluation of the point values in S_j, S_{ij} and S_K^k. The algorithm flowchart for the DG method is almost identical to the one described above for the finite volume method, and is therefore omitted to save space.

(*c*) *Numerical tests for the fifth order WENO schemes*

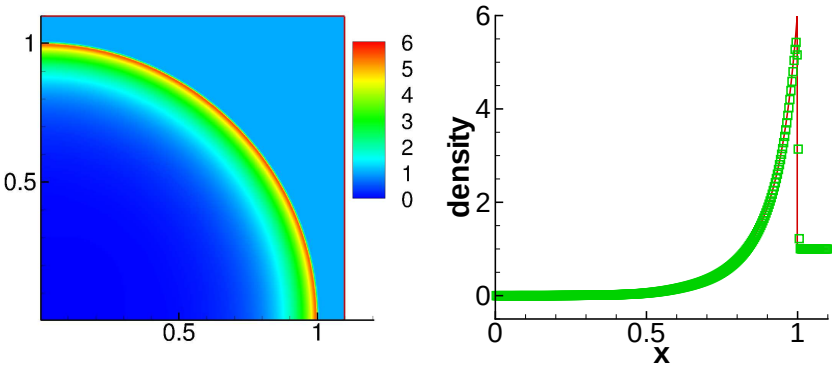
We show some numerical tests for the fifth order finite volume WENO schemes with the simplified implementation of the limiter on rectangular meshes described above. The time discretization is the third order SSP Runge-Kutta and the CFL is taken as (2.16). The algorithm for finite volume WENO schemes on rectangular meshes was described in Shu (2009) and the linear weights can be found in the appendix of Zhang & Shu (2010*b*), where the negative linear weights should be dealt with by the method in Shi *et al.* (2002). Extensive tests for scalar conservation laws were done to test the accuracy for the new limiter mentioned above. The results are similar to those in Zhang & Shu (2010*b*). We will not show the accuracy tests here to save space.

Example 1 (Two stream instability for Vlasov-Poisson equations). The initial and boundary conditions are the same as in Banks & Hittinger (2010). See figure 2 for the results. The numerical solution on the top-right in figure 2 is non-negative everywhere. This can be clearly seen in the cuts along $x = 0$, especially in the zoomed cuts on the bottom-right in figure 2.

Example 2 (Low density or low pressure problems for compressible Euler equations). We consider the two-dimensional Sedov blast wave and ninety-degree shock diffraction problem in Zhang & Shu (2010*c*) where the results of the positivity-preserving third order DG method were reported. Traditional finite volume and finite difference WENO schemes will blow up for such problems. Here we show the results of the fifth order finite volume WENO scheme with the new positivity-preserving limiter. See figures 3 and 4. The results are comparable to those of the DG method.

6. Concluding remarks

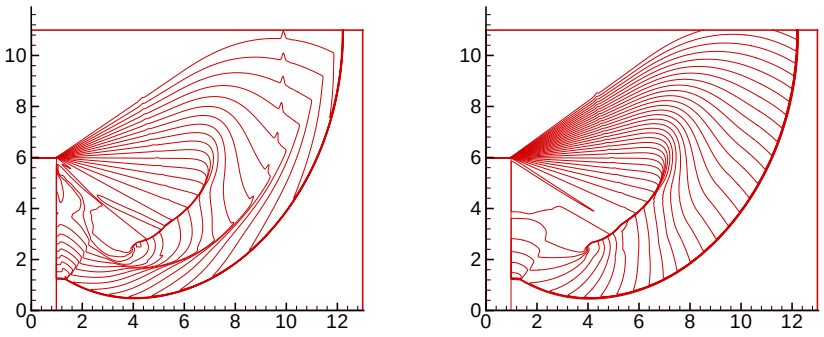
We have given a review of the recently developed maximum-principle-satisfying high order finite volume or DG schemes for scalar conservation laws, including generalizations and applications to two dimensional incompressible Euler equations and passively convection equations with a divergence free velocity field, and positivity-preserving schemes for compressible Euler equations, shallow water equations, Vlasov-Boltzmann transport equations, and a population model. We also propose a simpler and less expensive implementation especially for the finite volume WENO schemes, and provide several numerical examples to demonstrate their performance.



(a) Color contour of density

(b) Cut along $y = 0$. The solid line is the exact solution. Symbols are numerical solutions.

Figure 3. 2D Sedov blast. $T = 1$. $\Delta x = \Delta y = \frac{1.1}{320}$. The third order Runge-Kutta and fifth order finite volume WENO scheme with the positivity-preserving limiter.



(a) Density: 20 equally spaced contour lines from $\rho = 0.066227$ to $\rho = 7.0668$.

(b) Pressure: 40 equally spaced contour lines from $p = 0.091$ to $p = 37$.

Figure 4. Shock diffraction problem. $\Delta x = \Delta y = 1/80$. The third order Runge-Kutta and fifth order finite volume WENO scheme with the positivity-preserving limiter.

Acknowledgment

Support by AFOSR grant FA9550-09-1-0126 and NSF grant DMS-0809086 is acknowledged.

References

Banks, J. W. & Hittinger, J. A. F. 2010 *A new class of nonlinear finite-volume methods for vlasov-simulation*, *IEEE T. Plasma. Sci.*, **38**, No. 9, 2198-2207. (doi:10.1109/TPS.2010.2056937)

Cheng, Y., Gamba, I.M., & Proft, J., 2010 *Positivity-Preserving discontinuous Galerkin schemes for linear Vlasov-Boltzmann transport equations*, *Math. Comput.*, to appear.

Cockburn B., & Shu, C.-W. 1989 *TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws II: general framework*, *Math. Comput.*, **52**, 411-435. (doi:10.1016/0021-9991(89)90183-6)

Harten, A., Engquist, B., Osher, S. & Chakravarthy, S. 1987 *Uniformly high order essentially non-oscillatory schemes, III*, *J. Comput. Phys.*, **71**, 231-303. (doi:10.1006/jcph.1996.5632)

Jiang, G.-S. & Shu, C.-W. 1996 *Efficient implementation of weighted ENO schemes*, *J. Comput. Phys.*, **126**, 202-228. (doi:10.1006/jcph.1996.0130)

Jiang, G.-S. & Tadmor, E. 1998 *Nonoscillatory central schemes for multidimensional hyperbolic conservative laws*, *SIAM J. Sci. Comput.*, **19**, 1892-1917. (doi:10.1137/S106482759631041X)

Liu, J.-G. & Shu, C.-W. 2000 *A high-order discontinuous Galerkin method for 2D incompressible flows*, *J. Comput. Phys.*, **160**, 577-596. (doi:10.1006/jcph.2000.6475)

Liu, X.-D. & Osher, S. 1996 *Non-oscillatory high order accurate self similar maximum principle satisfying shock capturing schemes*, *SIAM J. Numer. Anal.*, **33**, 760-779. (doi:10.1137/0733038)

Liu, X.-D., Osher, S. & Chan, T. 1994 *Weighted essentially non-oscillatory schemes*, *J. Comput. Phys.*, **115**, 200-212. (doi:10.1006/jcph.1994.1187)

Osher, S. & Chakravarthy, S. 1984 *High resolution schemes and the entropy condition*, *SIAM J. Numer. Anal.*, **21**, 955-984.

Perthame, B. & Shu, C.-W. 1996 *On positivity preserving finite volume schemes for Euler equations*, *Numer. Math.*, **73**, 119-130. (doi:10.1007/s002110050187)

Qiu, J.-M. & Shu, C.-W. 2011 *Positivity preserving semi-Lagrangian discontinuous Galerkin formulation: theoretical analysis and application to the Vlasov-Poisson system*, submitted to *J. Comput. Phys.*.

Rossmanith, J. A. & Seal, D. C. 2011 *A positivity-preserving high-order semi-Lagrangian discontinuous Galerkin scheme for the Vlasov-Poisson equations*, submitted to *J. Comput. Phys.*.

Sanders, R. 1988 *A third-order accurate variation nonexpansive difference scheme for single nonlinear conservation law*, *Math. Comput.*, **51**, 535-558. (doi:10.1090/S0025-5718-1988-0935073-3)

Shi, J., Hu, C., & Shu, C.-W. 2001 *A technique of treating negative weights in WENO schemes*, *J. Comput. Phys.*, **175**, 108-127. (doi:10.1006/jcph.2001.6892)

Shu, C.-W. 1988 *Total-Variation-Diminishing time discretizations*, *SIAM J. Sci. Stat. Comp.*, **9**, 1073-1084. (doi:10.1137/0909073)

Shu, C.-W. 2009 *High order weighted essentially non-oscillatory schemes for convection dominated problems*, *SIAM Review*, **51**, 82-126. (doi:10.1137/070679065)

Shu, C.-W. & Osher, S. 1988 *Efficient implementation of essentially non-oscillatory shock-capturing schemes*, *J. Comput. Phys.*, **77**, 439-471. (doi: 10.1016/0021-9991(88)90177-5)

Xing, Y., Zhang, X. and Shu, C.-W. 2010 *Positivity preserving high order well balanced discontinuous Galerkin methods for the shallow water equations*, *Adv. Water Resour.*, **33**, 1476-1493. (doi: 10.1016/j.advwatres.2010.08.005)

Zhang, R., Zhang, M., & Shu, C.-W. 2010 *High order positivity-preserving finite volume WENO schemes for a hierarchical size-structured population model*, submitted to *J. Comput. Appl. Math.*.

Zhang, X. & Shu, C.-W. 2010*a* *A genuinely high order total variation diminishing scheme for one-dimensional scalar conservation laws*, *SIAM J. Numer. Anal.*, **48**, 772-795. (doi:10.1137/090764384)

Zhang, X. & Shu, C.-W. 2010*b* *On maximum-principle-satisfying high order schemes for scalar conservation laws*, *J. Comput. Phys.*, **229**, 3091-3120. (doi:10.1016/j.jcp.2009.12.030)

Zhang, X. & Shu, C.-W. 2010*c* *On positivity preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes*, *J. Comput. Phys.*, **229**, 8918-8934. (doi:10.1016/j.jcp.2010.08.016)

Zhang, X. & Shu, C.-W. 2011 *Positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations with source terms*, *J. Comput. Phys.*, **230**, 1238-1248. (doi:10.1016/j.jcp.2010.10.036)

Zhang, X., Xia, Y., & Shu, C.-W. 2011 *Maximum-principle-satisfying and positivity-preserving high order discontinuous Galerkin schemes for conservation laws on triangular meshes*, *J. Sci. Comput.*, to appear. (doi:10.1007/s10915-011-9472-8)