Positivity-preserving high order finite volume HWENO schemes for compressible Euler equations[*]

Xiaofeng Cai[†], Xiangxiong Zhang[‡] and Jianxian Qiu[§]

Abstract: In this paper, we present a positivity-preserving high order finite volume Hermite weighted essentially non-oscillatory (HWENO) scheme for compressible Euler equations based on the framework for constructing uniformly high order accurate positivity-preserving discontinuous Galerkin and finite volume schemes for Euler equations proposed in [20]. The major advantages of the HWENO schemes is their compactness in the spacial field because the function and its first derivative are evolved in time and used in the reconstructions. On the other hand, the HWENO reconstruction tends to be more oscillatory than those of conventional WENO schemes. Thus positivity preserving techniques are more needed in HWENO schemes for the sake of stability. Numerical tests will be shown to demonstrate the robustness and high-resolution of the schemes.

Keywords: Positivity preserving; high order accuracy; Hermite weighted essentially non-oscillatory scheme; finite volume scheme; compressible Euler equations.

[†]School of Mathematical Sciences, Xiamen University, Xiamen, Fujian, 361005, P.R. China. E-mail: xfcai89@126.com.

[‡]Department of Mathematics, Purdue University, West Lafayette, IN 47907-2067, USA. E-mail: zhan1966@purdue.edu.

[§]School of Mathematical Sciences and Fujian Provincial Key Laboratory of Mathematical Modeling & High-Performance Scientific Computing, Xiamen University, Xiamen, Fujian, 361005, P.R. China. E-mail: jxqiu@xmu.edu.cn.

1 Introduction

When solving gas dynamics equations with conservative Eulerian schemes, high order schemes can produce numerical solutions with better resolution than first order and second order ones. To render high order schemes stable for simulating shock waves, limiters or nonlinear non-oscillatory reconstructions must be used, e.g., a total-variation-bounded (TVB) limiter for high order discontinuous Galerkin method [2] and weighted essentially non-oscillatory (WENO) type schemes such as finite difference WENO schemes [6, 24] and finite volume Hermite WENO (HWENO) schemes [11, 12, 28]. Even though these schemes have been demonstrated to perform well for a wide range of problems, in practice they may still be unstable due to loss of positivity for low density or low pressure problems.

For resolving this difficulty, a series of high order positivity-preserving schemes such as discontinuous Galerkin method, finite volume and finite difference schemes WENO were developed recently in [19, 20, 24] following a general methodology as reviewed in [22]. This method for finite volume schemes can be easily implemented as a post processing step to limit the high order reconstruction polynomials or high order reconstructed point values without destroying accuracy and can be easily generalized, for instance, to general equations of state and Euler system with source terms [21], to controlling the physical entropy in gas dynamics [27], to unstructured meshes [23], to convection-diffusion equations [25, 26] and to the shallow water equations [17, 18].

Finite volume HWENO schemes was proposed in [11, 12, 28, 1]. There are quite a few advantages using a compact numerical stencil: first, it is easier to deal with boundary conditions and complex geometries; second, for the same formal accuracy, compact stencils are known to exhibit more resolution of the smaller scales by improving the dispersive and the dissipative properties of the numerical scheme [15, 8].

However, in practice HWENO schemes are less robust than conventional WENO schemes. The HWENO schemes are more unstable numerically for low pressure or low density problems, in which positivity preserving is a crucial property for the sake of stability. For instance,

the flux limiter in [5] was applied to stabilize the finite difference HWENO schemes [9]. The method to construct positivity-preserving limiters for finite volume schemes in [19, 20] can be applied to finite volume HWENO scheme to achieve positivity without losing conservation. The limiter will not destroy the high order accuracy of the HWENO scheme for smooth solutions without vacuum. However, the positivity-preserving limiter in [19, 20] is defined for reconstruction polynomials, which are not available in WENO and HWENO reconstructions. Such polynomials can be obtained by interpolating reconstructed point values in WENO [19], but the interpolation step is computationally inefficient especially in high dimensions. A simpler and more efficient implementation of the positivity-preserving limiter for WENO reconstruction was discussed in [22]. In this paper, we follow [22, 16] to implement an efficient and robust positivity-preserving limiter for finite volume HWENO schemes. The HWENO scheme with this simple limiter will be much more robust for compressible Euler equations.

The paper is organized as follows. In Section 2, we briefly review the finite volume HWENO schemes in [11, 12, 28]. In Section 3, we introduce positivity-preserving finite volume HWENO schemes in one dimension and two dimensions for the perfect gas. In Section 4, numerical tests of the fifth order HWENO schemes for one dimensional Euler equations and the fourth order finite volume HWENO schemes in two dimensional case are shown. Concluding remarks are given in Section 5.

2 Description of finite volume Hermite WENO schemes

We briefly review the construction of finite volume HWENO schemes for solving conservation laws

$$\begin{cases} q_t + \nabla \cdot F(q) = 0, \\ q(x, 0) = q_0(x), \end{cases} \tag{2.1}$$

in [11, 12, 28]. Taking the gradient with respect to the spatial variables in (2.1), we have

$$(\nabla q)_t^T + \nabla^T (\nabla \cdot F(q)) = 0, \tag{2.2}$$

3

thus

$$(\nabla q)_t^T + \nabla \cdot (\nabla \otimes F(q)) = 0, \tag{2.3}$$

where $\otimes$ is a tensor product. For instance, the tensor product of two vectors $a = (a_1 \ a_2)$ and $b = (b_1 \ b_2)$ is

$$a \otimes b = (a_1 \ a_2)^T (b_1 \ b_2) = \begin{pmatrix} a_1 b_1 & a_1 b_2 \\ a_2 b_1 & a_2 b_2 \end{pmatrix}. \tag{2.4}$$

For using a Hermite interpolation procedure, both the function and its derivative are needed during the evolution in time. The finite volume HWENO schemes are defined for the equations:

$$U_t + \nabla \cdot \mathcal{F}(U) = 0, \tag{2.5}$$

where $U = (q, \nabla q)^T$ and $\mathcal{F}(U) = \begin{pmatrix} F(q) \\ \nabla \otimes F(q) \end{pmatrix}$. We integrate the system (2.5) on a control volume Ω_j, which is an interval $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ in one dimensional case or a rectangle $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$ in two dimensional cases. After integration by parts, we obtain the integral form of the equation as :

$$\frac{d}{dt}\overline{U}_{\Omega_j} = -\frac{1}{|\Omega_j|} \int_{\partial\Omega_j} \mathcal{F}(U) \cdot n \, ds \tag{2.6}$$

where $|\Omega_j|$ is the volume of the control volume Ω_j and n represents the outward unit normal vector to the boundary of the control volume $\partial\Omega_j$. The line integral in (2.6) can be approximated by a L-point Gaussian quadrature on each side of $\partial\Omega_j = \bigcup_{s=1}^{S} \partial\Omega_{js}$:

$$\int_{\partial\Omega_j} \mathcal{F}(U) \cdot n \, ds \approx \sum_{s=1}^{S} |\partial\Omega_{js}| \sum_{l=1}^{L} \omega_l \mathcal{F}(U(G_{sl}, t)) \cdot n, \tag{2.7}$$

where G_{sl} and ω_l are Gaussian quadrature points on $\partial\Omega_{js}$ and weights respectively. The flux $\mathcal{F}(U(G_{sl}, t)) \cdot n$ at Gaussian quadrature point is replaced by a numerical flux (approximate or exact Riemann solvers). For example, one could use the simple Lax-Friedrichs flux, which is given by

$$\mathcal{F}(U(G_{sl}, t)) \cdot n \approx \frac{1}{2}[\mathcal{F}(U^-(G_{sl}, t)) + \mathcal{F}(U^+(G_{sl}, t))] \cdot n - \alpha(U^+(G_{sl}, t) - U^-(G_{sl}, t)), \tag{2.8}$$

where α is taken as an upper bound for the eigenvalues of the Jacobian along the direction n, and U^- and U^+ are the reconstructed values of U at the Gaussian point G_{sl} in the

4

inside and the outside Ω_j. The procedures of finite volume Hermite WENO reconstruction of $U^{\pm}(G_{sl}, t)$ in one-dimensional and two-dimensional cases are given in detail in [11, 28], respectively. Finally, the semi-discretization HWENO scheme (2.6) can be written in the following ODE form:

$$\frac{d}{dt}\overline{U}_{\Omega_j} = \mathcal{L}(\overline{U})_{\Omega_j}. \tag{2.9}$$

The method of lines ODE (2.9) is then discretized in time by a SSP Runge-Kutta method in [14]. The third-order version in [14] is used in this paper,

$$\begin{aligned}
\overline{U}^{(1)} &= \overline{U}^n + \Delta t \mathcal{L}(\overline{U}^n), \\
\overline{U}^{(2)} &= \tfrac{3}{4}\overline{U}^n + \tfrac{1}{4}(\overline{U}^{(1)} + \Delta t \mathcal{L}(\overline{U}^{(1)})), \\
\overline{U}^{n+1} &= \tfrac{1}{3}\overline{U}^n + \tfrac{2}{3}(\overline{U}^{(2)} + \Delta t \mathcal{L}(\overline{U}^{(2)})).
\end{aligned} \tag{2.10}$$

3 Positivity-preserving limiters for finite volume HWENO schemes

We present positivity-preserving limiters for finite volume HWENO schemes based on the work in [20, 22] for compressible Euler equations and their improvement in [16] for reactive Euler equations. We apply the method in [20, 22, 16] to (2.1) and leave derivative terms unchanged, since derivative terms do not affect the positivity in this method.

3.1 Positivity-preserving high order finite volume HWENO schemes for solving one-dimensional Euler equations

Consider the one dimensional Euler equations for the perfect gas being given by

$$q_t + f(q)_x = 0, \quad t \geqslant 0, x \in \mathbb{R}, \tag{3.1}$$

$$q = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix}, f(q) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ (E + p)u \end{pmatrix} \tag{3.2}$$

with

$$m = \rho u, \quad E = \frac{1}{2}\rho u^2 + \rho e, \quad p = (\gamma - 1)\rho e,$$

where 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 $f'(q)$ are $u - c, u, u + c$.

Let $p(q) = (\gamma - 1)\left(E - \frac{1}{2}\frac{m^2}{\rho}\right)$ be the pressure function. It can be easily verified that p is a concave function of $q = (\rho, m, E)^T$ if $\rho > 0$. For $q_1 = (\rho_1, m_1, E_1)^T$ and $q_2 = (\rho_2, m_2, E_2)^T$, Jensen's inequality implies, for $0 \leqslant s \leqslant 1$,

$$p(sq_1 + (1 - s)q_2) \geqslant sp(q_1) + (1 - s)p(q_2), \quad \text{if} \quad \rho_1 \geqslant 0, \quad \rho_2 \geqslant 0. \tag{3.3}$$

Define the set of admissible states by

$$G = \left\{ q = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix} \middle| \rho > 0 \quad \text{and} \quad p = (\gamma - 1)\left(E - \frac{1}{2}\frac{m^2}{\rho}\right) > 0 \right\}, \tag{3.4}$$

then G is a convex set.

The time discretization is used in HWENO schemes as the high order strong stability preserving (SSP) methods which are convex combinations of Euler forward. Due to the convexity of G, if Euler forward is positivity-preserving, then so are the high order SSP time discretizations. Thus we only need to discuss the Euler forward in time.

Let $\xi = q_x$. Taking the derivative of (3.1), we obtain

$$\xi_t + \mathcal{H}(q, \xi)_x = 0,$$

where $\mathcal{H}(q, \xi) = f'(q)\xi$.

Then the Euler forward temporal discretization for the the semi-discretization HWENO scheme of (2.9) can be written as

$$\begin{cases} \overline{q}_j^{n+1} = \overline{q}_j^n - \frac{\Delta t}{\Delta x}\left[\widehat{f}\left(q_{j+\frac{1}{2}}^-, q_{j+\frac{1}{2}}^+\right) - \widehat{f}\left(q_{j-\frac{1}{2}}^-, q_{j-\frac{1}{2}}^+\right)\right], \\ \overline{\xi}_j^{n+1} = \overline{\xi}_j^n - \frac{\Delta t}{\Delta x}[\widehat{\mathcal{H}}(q_{j+\frac{1}{2}}^-, q_{j+\frac{1}{2}}^+; \xi_{j+\frac{1}{2}}^-, \xi_{j+\frac{1}{2}}^+) - \widehat{\mathcal{H}}(q_{j-\frac{1}{2}}^-, q_{j-\frac{1}{2}}^+; \xi_{j-\frac{1}{2}}^-, \xi_{j-\frac{1}{2}}^+)], \end{cases} \tag{3.5}$$

where $\overline{q}_j^n$ is the approximation of the cell average of the exact solution $q(x, t)$ in the cell $I_j = \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$ at time level n, and $q_{j+\frac{1}{2}}^-, q_{j+\frac{1}{2}}^+$ are the high order approximations of the point values $\mathbf{q}\left(x_{j+\frac{1}{2}}, t^n\right)$ within the cells I_j and I_{j+1} respectively. These values are reconstructed from $\overline{q}_j^n$ and $\overline{\xi}_j^n$ by the HWENO reconstruction. We assume that there is a

vector of degree k polynomials $q_j(x) = (\rho_j(x), m_j(x), E_j(x))^T$ which are $(k+1)$-th order accurate approximations to smooth exact solutions $q(x, t)$ on I_j, and satisfy that $\bar{q}_j^n$ is the cell average of $q_j(x)$ on I_j, $q_{j-\frac{1}{2}}^+ = q_j\left(x_{j-\frac{1}{2}}\right)$ and $q_{j+\frac{1}{2}}^- = q_j\left(x_{j+\frac{1}{2}}\right)$. The existence of such polynomials can be established by the interpolation involving $q_{j+\frac{1}{2}}^-$, $q_{j-\frac{1}{2}}^+$ and $\bar{q}_j^n$.

The numerical fluxes $\widehat{f}(a, b)$ and $\widehat{\mathcal{H}}(a, b; c, d)$ used in this paper are the following global Lax-Friedrichs fluxes:

$$
\begin{aligned}
\widehat{f}(a, b) &= \tfrac{1}{2}[f(a) + f(b) - \alpha(b - a)], \\
\widehat{\mathcal{H}}(a, b; c, d) &= \tfrac{1}{2}[f'(a)c + f'(b)d - \alpha(d - c)].
\end{aligned}
\tag{3.6}
$$

where $\alpha = \|(|u| + c)\|_\infty$.

We need the N-point Legendre Gauss-Lobatto quadrature rule on the interval $I_j = \left[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}\right]$, which is exact for the integral of polynomials of degree up to $2N - 3$. We would need to choose N such that $2N - 3 \geqslant k$. The smallest $N = 4$ is chosen for satisfying $2N - 3 \geq 4$. Denote these quadrature points on I_j as

$$
S_j = \{x_{j-\frac{1}{2}} = \widehat{x}_j^1, \widehat{x}_j^2, \ldots, \widehat{x}_j^{N-1}, \widehat{x}_j^N = x_{j+\frac{1}{2}}\}.
\tag{3.7}
$$

Let $\widehat{\omega}_j$ denote weights in N-point Gauss-Lobatto quadrature for the reference cell $[-\frac{1}{2}, \frac{1}{2}]$, then the smallest weight for $N = 4$ is $\widehat{\omega}_1 = \widehat{\omega}_4 = \frac{1}{12}$.

A general framework to construct a high order positivity preserving finite volume scheme for the Euler equations was introduced in [20], in which a sufficient condition for $\bar{q}_j^{n+1} \in G$ is that, all the nodal value $q_j(\widehat{x}_j^\alpha) \in G$ for all j and α under suitable CFL conditions

$$
\lambda \|(|u| + c)\|_\infty \leq \widehat{\omega}_1 \alpha_0,
\tag{3.8}
$$

where $\alpha_0 = 1$ for the Lax-Friedrichs flux. The positivity-preserving limiter in [20] can enforce this sufficient condition without destroying conservation and accuracy. For more detailed description of such a positivity-preserving method, see [20, 22]. However, neither the polynomial $q_j(x)$ nor the nodal values $q_j(\widehat{x}_j^\alpha)$ ($\alpha = 2, \cdots, N - 1$) are available from the WENO type reconstruction, which poses additional chanllenge to implementing an efficient limiter. In [22, 16], simpler implementations of the limiter were discussed.

Following Section 5 in [22], for preserving positivity in the finite volume HWENO scheme (3.5), we have the following weaker sufficient condition which can be enforced without constructing the polynomials $q_j(x)$.

THEOREM 1. *Let $q_j(x) = (\rho_j(x), m_j(x), E_j(x))^T$ be the approximation polynomials approximating q in the HWENO scheme (3.5). For the scheme (3.5), if*

$$\frac{\overline{q}_j^n - \widehat{\omega}_1 q_{j-\frac{1}{2}}^+ - \widehat{\omega}_N q_{j+\frac{1}{2}}^-}{1 - 2\widehat{\omega}_1} \in G, \quad and \quad q_{j\pm\frac{1}{2}}^\pm \in G, \tag{3.9}$$

then $\overline{q}_j^{n+1} \in G$ under the CFL condition $\lambda\|(|u| + c)\|_\infty \leqslant \widehat{\omega}_1 \alpha_0$, where $\alpha_0 = 1$ for the Lax-Friedrichs flux.

REMARK 1. *In HWENO reconstruction, only point values $q_{j\pm\frac{1}{2}}^\pm$ and $q_{j\mp\frac{1}{2}}^\pm$ are reconstructed. In Theorem 1, we need the existence of the approximation polynomials $q_j(x)$ which is a high order accurate approximation to q and has cell average $\overline{q}_j^n$ and cell end values $q_{j-\frac{1}{2}}^+$, $q_{j+\frac{1}{2}}^-$. The existence of such polynomials can be established by interpolation, see [19].*

REMARK 2. *Notice that $\widehat{\omega}_1 = \widehat{\omega}_N$ and $\overline{q}_j^n = \frac{1}{\Delta x_j} \int_{I_j} q_j(x)\,dx = \sum_{\alpha=1}^N q_j(\widehat{x}_j^\alpha)\widehat{\omega}_j$, thus we have*

$$\frac{\overline{q}_j^n - \widehat{\omega}_1 q_{j-\frac{1}{2}}^+ - \widehat{\omega}_N q_{j+\frac{1}{2}}^-}{1 - 2\widehat{\omega}_1} = \sum_{\alpha=2}^{N-1} \frac{\widehat{\omega}_j}{1 - 2\widehat{\omega}_1} q_j(\widehat{x}_j^\alpha),$$

which is a convex combination of $q_j(x^\alpha)$ ($\alpha = 2, \cdots, N-1$). By the mean value theorem, there exist some points x_j^1, x_j^2, x_j^3 in I_j such that

$$(\rho_j(x_j^1), m_j(x_j^2), E_j(x_j^3))^T = \sum_{\alpha=2}^{N-1} \frac{\widehat{\omega}_j}{1 - 2\widehat{\omega}_1} q_j(\widehat{x}_j^\alpha) = \frac{\overline{q}_j^n - \widehat{\omega}_1 q_{j-\frac{1}{2}}^+ - \widehat{\omega}_N q_{j+\frac{1}{2}}^-}{1 - 2\widehat{\omega}_1}.$$

Even though the points x_j^1, x_j^2, x_j^3 are three different points, for convenience we will treat $(\rho_j(x_j^1), m_j(x_j^2), E_j(x_j^3))^T$ as if it is $q_j(x)$ evaluated at one point in the following discussion.

For each cell I_j, given $\overline{q}_j^n \in G$ and nodal values $q_{j\pm\frac{1}{2}}^\mp$ constructed in HWENO procedure, the following limiter can be used to enforce the sufficient condition (3.9).

1. Set up a small number $\varepsilon = \min_j\{10^{-13}, \overline{\rho}_j^n, p(\overline{q}_j^n)\}$.

2. In each cell, evaluate $\theta_1 = \min\left\{\left|\frac{\overline{\rho}_j^n - \varepsilon}{\overline{\rho}_j^n - \rho_{\min}}\right|, 1\right\}$ where

$$\rho_{\min} = \min\{\rho_{j+\frac{1}{2}}^-, \rho_{j-\frac{1}{2}}^+, \rho_j(x_j^1)\}, \quad \rho_j(x_j^1) = \frac{\overline{\rho}_j^n - \widehat{\omega}_1\rho_{j-\frac{1}{2}}^+ - \widehat{\omega}_N\rho_{j+\frac{1}{2}}^-}{1 - 2\widehat{\omega}_1}.$$

3. Modify the density first: set

$$\widehat{\rho}_{j+\frac{1}{2}}^- = \theta_1(\rho_{j+\frac{1}{2}}^- - \overline{\rho}_j^n) + \overline{\rho}_j^n, \quad \widehat{\rho}_{j-\frac{1}{2}}^+ = \theta_1(\rho_{j-\frac{1}{2}}^+ - \overline{\rho}_j^n) + \overline{\rho}_j^n.$$

Then denote

$$\widehat{q}_j^1 = \widehat{q}_{j+\frac{1}{2}}^- = \left(\widehat{\rho}_{j+\frac{1}{2}}^-, m_{j+\frac{1}{2}}^-, E_{j+\frac{1}{2}}^-\right)^T, \widehat{q}_j^2 = \widehat{q}_{j-\frac{1}{2}}^+ = \left(\widehat{\rho}_{j-\frac{1}{2}}^+, m_{j-\frac{1}{2}}^+, E_{j-\frac{1}{2}}^+\right)^T,$$

and $\widehat{q}_j^3 = (\widehat{\rho}_j(x_j^1), m_j(x_j^2), E_j(x_j^3))^T = \frac{\overline{q}_j^n - \widehat{\omega}_1\widehat{q}_{j-\frac{1}{2}}^+ - \widehat{\omega}_N\widehat{q}_{j+\frac{1}{2}}^-}{1-2\widehat{\omega}_1}$. Here $\widehat{\rho}_j(x)$ denotes the polynomial $\widehat{\rho}_j(x) = \theta_1(q_j(x) - \overline{\rho}_j^n) + \overline{\rho}_j^n$. Notice that we only need to compute three nodal values of $\widehat{\rho}_j(x)$.

4. Then modify the pressure: for $l = 1, 2, 3$, if $p(\widehat{q}_j^l) < \varepsilon$, then solve the following quadratic equation for $t_\varepsilon^l \in [0, 1]$,

$$p[(1 - t_\varepsilon^l)\overline{q}_j^n + t_\varepsilon^l\widehat{q}_j^\alpha] = \varepsilon. \tag{3.10}$$

The convexity of the set $\{(\rho, m, E)^T : \rho > 0, p \geq \varepsilon\}$ implies the solution to this quadratic equation is unique in the interval $[0, 1]$. If $p(\widehat{q}_j^l) \geq \varepsilon$, then set $t_\varepsilon^l = 1$. Set $\theta_2 = \min\limits_{\alpha=1,2,3} t_\varepsilon^l$.

5. For modifying the pressure, an easier yet more robust alternative was introduced in [16]: for $l = 1, 2, 3$, if $p(\widehat{q}_j^l) < \varepsilon$, set

$$t_\varepsilon^l = \frac{p(\overline{q}_j^n) - \varepsilon}{p(\overline{q}_j^n) - p(\widehat{q}_j^l)}. \tag{3.11}$$

if $p(\widehat{q}_j^l) \geq \varepsilon$, then set $t_\varepsilon^l = 1$. Set $\theta_2 = \min\limits_{l=1,2,3} t_\varepsilon^l$. By Jensen's inequality on the concave pressure function, we can see that t_ε^l computed in (3.11) is smaller than the one computed in (3.10). In other words, (3.10) results in less modification than (3.11). Nonetheless, both approaches are accurate modifications for smooth solutions without vacuum.

9

6. Compute

$$\widetilde{q}^-_{j+\frac{1}{2}} = \theta_2(\widehat{q}^-_{j+\frac{1}{2}} - \overline{q}^n_j) + \overline{q}^n_j, \quad \widetilde{q}^+_{j-\frac{1}{2}} = \theta_2(\widehat{q}^+_{j-\frac{1}{2}} - \overline{q}^n_j) + \overline{q}^n_j. \tag{3.12}$$

It is straightforward to check that $\widetilde{q}^\pm_{j\pm\frac{1}{2}}$ and $\overline{q}^n_j$ satisfy the condition (3.9).

7. Use $\widetilde{q}^\pm_{j\pm\frac{1}{2}}$ instead of $q^\pm_{j\pm\frac{1}{2}}$, in the scheme (3.5) with the CFL condition (3.8), where $\||u| + c\|_\infty$ is the maximum of eigenvalues of the Jacobian $f'(q)$ over the set consisting of $\overline{q}^n_j, \widetilde{q}^-_{j+\frac{1}{2}}, \widetilde{q}^+_{j-\frac{1}{2}}$ for all j.

To see why the simplified limiter discussed above is still an accurate modification, we can compare it to the one in [20] enforcing the stronger condition $q_j(\widehat{x}^\alpha_j) \in G$ for all α. For simplicity, we omit the subscript j here. We only discuss the case for enforcing positivity of pressure since the discussion for density will be similar. Given the cell average $\overline{q} \in G$ satisfying $p[\overline{q}] \geq \varepsilon$ and the high order accurate reconstructed point values q^+ and q^- on the cell I, we assume the density of q^+ and q^- are already positive. Let $q(x)$ be a high order accurate approximation polynomial satisfying that $\overline{q}$ is its cell average on I and q^+ and q^- are its left and right end point values on I. We compare the following two limiters:

1. Let $\widehat{\theta} = \min\limits_{\alpha=1,\cdots,N} t^\alpha_\varepsilon$ where $t^\alpha_\varepsilon = 1$ if $p[q(\widehat{x}^\alpha)] \geq \varepsilon$ and $t^\alpha_\varepsilon \in [0,1]$ solves the quadratic equation $p[\theta(q(\widehat{x}^\alpha) - \overline{q}) + \overline{q}] = \varepsilon$ for the unknown $\theta \in [0,1]$ otherwise.

2. Let $\overline{\theta} = \min\limits_{l=1,2,3} t^l_\varepsilon$ where $t^l_\varepsilon = 1$ if $p[q^l] \geq \varepsilon$ and $t^l_\varepsilon \in [0,1]$ solves the quadratic equation $p[\theta(q^l - \overline{q}) + \overline{q}] = \varepsilon$ for the unknown $\theta \in [0,1]$ otherwise. Here $q^1 = q^- = q(\widehat{x}^N)$, $q^2 = q^+ = q(\widehat{x}^1)$ and $q^3 = \frac{\overline{q} - \widehat{\omega}_1 q^+ - \widehat{\omega}_N q^-}{1 - 2\widehat{\omega}_1}$.

Define

$$q^\theta(x) = \theta(q(x) - \overline{q}) + \overline{q}, \quad \theta \in [0,1].$$

Then $\theta = \overline{\theta}$ is the largest number in $[0,1]$ such that the following inequalities hold,

$$p\left[q^\theta(\widehat{x}^1)\right] \geq \varepsilon, \quad p\left[q^\theta(\widehat{x}^N)\right] \geq \varepsilon, \quad p\left[\frac{\overline{q} - \widehat{\omega}_1 q^\theta(\widehat{x}^1) - \widehat{\omega}_N q^\theta(\widehat{x}^N)}{1 - 2\widehat{\omega}_1}\right] \geq \varepsilon. \tag{3.13}$$

10

On the other hand, $\theta = \hat{\theta}$ also satisfies (3.13) due to the Jensen's inequality for the concave pressure function and the fact that the following convex combination holds for any θ,

$$\frac{\overline{q} - \widehat{\omega}_1 q^\theta(\widehat{x}^1) - \widehat{\omega}_N q^\theta(\widehat{x}^N)}{1 - 2\widehat{\omega}_1} = \sum_{\alpha=2}^{N-1} \frac{\widehat{\omega}_j}{1 - 2\widehat{\omega}_1} q^\theta(\widehat{x}^\alpha_j).$$

Therefore we have shown that $\overline{\theta} \geq \hat{\theta}$ thus the simplified limiter results in less modification than the limiter in [20] enforcing the stronger sufficient condition $q(\widehat{x}^\alpha) \in G$ for all α.

3.2 Positivity-preserving limiter for finite volume HWENO schemes in Two dimensions on Cartesian meshes

In this section section we consider two dimensional Euler equations

$$q_t + f(q)_x + g(q)_y = 0, \quad t \geq 0, \ (x, y) \in \mathbb{R}^2, \tag{3.14}$$

where

$$q = \begin{pmatrix} \rho \\ m \\ n \\ E \end{pmatrix}, \ f(q) = \begin{pmatrix} \rho u \\ \rho u^2 + p \\ \rho uv \\ (E + p)u \end{pmatrix}, \ g(q) = \begin{pmatrix} \rho v \\ \rho uv \\ \rho v^2 + p \\ (E + p)v \end{pmatrix} \tag{3.15}$$

with

$$m = \rho u, \ n = \rho v, \ E = \frac{1}{2}\rho u^2 + \frac{1}{2}\rho v^2 + \rho e, \ p = (\gamma - 1)\rho e,$$

where ρ is the density, u is the velocity in x direction, v is the velocity in y direction, m and n are the momenta, E is the total energy, p is the pressure, e is the internal energy. The speed of sound is given by $c = \sqrt{(\gamma p/\rho)}$. The eigenvalues of the Jacobian $f'(q)$ are $u - c, u, u$ and $u + c$ and the eigenvalues of the Jacobian $g'(q)$ are $v - c, v, v$ and $v + c$. The pressure function p is concave with respect to q if $\rho \geq 0$. Thus the set of admissible states

$$G = \left\{ q = \begin{pmatrix} \rho \\ m \\ n \\ E \end{pmatrix} \middle| \ \rho > 0 \text{ and } p = (\gamma - 1)\left(E - \frac{1}{2}\frac{m^2}{\rho} - \frac{1}{2}\frac{n^2}{\rho}\right) > 0 \right\}$$

is still convex.

For simplicity we assume we have a uniform rectangular mesh. At time level n, we have a vector of approximation polynomials of degree k, $q_{ij}(x, y) = (\rho_{ij}(x, y), m_{ij}(x, y), n_{ij}(x, y), E_{ij}(x, y))^T$

11

with the cell average $\overline{q}_{ij}^m = (\overline{\rho}_{ij}^n, \overline{m}_{ij}^n, \overline{n}_{ij}^n, \overline{E}_{ij}^n)^T$ on the (i,j) cell denoted by $I_{ij} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$. Let $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$ and $\Delta y = y_{j+\frac{1}{2}} - y_{j-\frac{1}{2}}$.

Let $\xi = \frac{\partial q}{\partial x}$, $\eta = \frac{\partial q}{\partial y}$. Taking the derivatives of (3.14), we obtain

$$\xi_t + \mathcal{H}(q,\xi)_x + \mathcal{R}(q,\xi)_y = 0, \tag{3.16}$$

$$\eta_t + \mathcal{K}(q,\eta)_x + \mathcal{S}(q,\eta)_y = 0, \tag{3.17}$$

where $\mathcal{H}(q,\xi) = f'(q)\xi, \mathcal{R}(q,\xi) = g'(q)\xi, \mathcal{K}(q,\eta) = f'(q)\eta, \mathcal{S}(q,\eta) = g'(q)\eta$.

We first integrate the equations (3.14), (3.16) and (3.17) on I_{ij}:

$$\frac{d}{dt}\overline{q}_{ij} = -\frac{1}{\Delta x \Delta y} \int_{\partial I_{ij}} F \cdot n ds, \tag{3.18}$$

$$\frac{d}{dt}\overline{\xi}_{ij} = -\frac{1}{\Delta y} \int_{\partial I_{ij}} H \cdot n ds, \tag{3.19}$$

$$\frac{d}{dt}\overline{\eta}_{ij} = -\frac{1}{\Delta x} \int_{\partial I_{ij}} Q \cdot n ds, \tag{3.20}$$

where

$$\overline{q}_{ij} = \frac{1}{\Delta x \Delta y} \int_{I_{ij}} q dx dy, \quad \overline{\xi}_{ij} = \frac{1}{\Delta y} \int_{I_{ij}} \frac{\partial q}{\partial x} dx dy, \quad \overline{\eta}_{ij} = \frac{1}{\Delta x} \int_{I_{ij}} \frac{\partial q}{\partial y} dx dy$$

and $F = (f,g)^T, H = (\mathcal{H}, \mathcal{R})^T, Q = (\mathcal{K}, \mathcal{S})^T$.

The line integral in (3.18)-(3.20) are approximated by a L-point Gauss quadrature on each side of $\partial I_{ij} = \bigcup_{s=1}^4 \partial I_{ij\,s}$,

$$\int_{\partial I_{ij}} F \cdot n ds \approx \sum_{s=1}^4 |\partial I_{ij\,s}| \sum_{l=1}^L \omega_l F(q(G_{sl}, t)) \cdot n, \tag{3.21}$$

$$\int_{\partial I_{ij}} H \cdot n ds \approx \sum_{s=1}^4 |\partial I_{ij\,s}| \sum_{l=1}^L \omega_l H(q(G_{sl}, t), \xi(G_{sl}, t)) \cdot n, \tag{3.22}$$

$$\int_{\partial I_{ij}} Q \cdot n ds \approx \sum_{s=1}^4 |\partial I_{ij\,s}| \sum_{l=1}^L \omega_l Q(q(G_{sl}, t), \eta(G_{sl}, t)) \cdot n, \tag{3.23}$$

where ω_l $(l = 1, \cdots, L)$ denote the Gauss quadrature weights for the reference interval $[-\frac{1}{2}, \frac{1}{2}]$.

Since we are constructing schemes up to fourth-order accuracy, two-point Gaussian will be used in each line integration, and $F(q(G_{sl},t)) \cdot n, H(q(G_{sl},t), \xi(G_{sl},t)) \cdot n, Q(q(G_{sl},t), \eta(G_{sl},t)) \cdot$

n are replaced by numerical fluxes such as the global Lax-Friedrichs fluxes:

$$\widehat{f}(q^-(G_{sl},t),q^+(G_{sl},t)) = \frac{1}{2}[f(q^-(G_{sl},t))+f(q^+(G_{sl},t))-\alpha(q^+(G_{sl},t)-q^-(G_{sl},t))], \quad (3.24)$$

$$\widehat{\mathcal{H}}(q^-(G_{sl},t),q^+(G_{sl},t);\xi^-(G_{sl},t),\xi^+(G_{sl},t)) = \frac{1}{2}[\mathcal{H}(q^-(G_{sl},t),\xi^-(G_{sl},t))$$
$$+\mathcal{H}(q^+(G_{sl},t),\xi^+(G_{sl},t))-\alpha(\xi^+(G_{sl},t)-\xi^-(G_{sl},t))], \quad (3.25)$$

$$\widehat{\mathcal{K}}(q^-(G_{sl},t),q^+(G_{sl},t);\eta^-(G_{sl},t),\eta^+(G_{sl},t)) = \frac{1}{2}[\mathcal{K}(q^-(G_{sl},t),\eta^-(G_{sl},t))$$
$$+\mathcal{K}(q^+(G_{sl},t),\eta^+(G_{sl},t))-\alpha(\eta^+(G_{sl},t)-\eta^-(G_{sl},t))], \quad (3.26)$$

where $\alpha = \max\{\|(|u|+c)\|_\infty, \|(|v|+c)\|_\infty\}$ and for $G_{sl} = (x_{i\pm\frac{1}{2}}, y_{j\pm\sqrt{3}/6})$, $q^\pm(G_{sl},t), \xi^\pm(G_{sl},t)$, $\eta^\pm(G_{sl},t)$ are the left and right limits of the solutions u, v, w at the cell interface G_{sl} respectively; and

$$\widehat{g}(q^-(G_{sl},t),q^+(G_{sl},t)) = \frac{1}{2}[g(q^-(G_{sl},t))+g(q^+(G_{sl},t))-\alpha(q^+(G_{sl},t)-q^-(G_{sl},t))], \quad (3.27)$$

$$\widehat{\mathcal{R}}(q^-(G_{sl},t),q^+(G_{sl},t);\xi^-(G_{sl},t),\xi^+(G_{sl},t)) = \frac{1}{2}[\mathcal{R}(q^-(G_{sl},t),\xi^-(G_{sl},t))$$
$$+\mathcal{R}(q^+(G_{sl},t),\xi^+(G_{sl},t))-\alpha(\xi^+(G_{sl},t)-\xi^-(G_{sl},t))], \quad (3.28)$$

$$\widehat{\mathcal{S}}(q^-(G_{sl},t),q^+(G_{sl},t);\eta^-(G_{sl},t),\eta^+(G_{sl},t)) = \frac{1}{2}[\mathcal{S}(q^-(G_{sl},t),\eta^-(G_{sl},t))$$
$$+\mathcal{S}(q^+(G_{sl},t),\eta^+(G_{sl},t))-\alpha(\eta^+(G_{sl},t)-\eta^-(G_{sl},t))], \quad (3.29)$$

where $G_{sl} = (x_{i\pm\sqrt{3}/6}, y_{j\pm\frac{1}{2}})$, $q^\pm(G_{sl},t), \xi^\pm(G_{sl},t), \eta^\pm(G_{sl},t)$ are the bottom and top limits of the solutions u, v, w at the cell interface G_{sl} respectively. The procedure of reconstruction of $q^\pm(G_{sl},t), \xi^\pm(G_{sl},t), \eta^\pm(G_{sl},t)$ from $\overline{q}_{ij}, \overline{\xi}_{ij}, \overline{\eta}_{ij}$ is given in detail in [12, 28].

Let $\lambda_1 = \frac{\Delta t}{\Delta x}$ and $\lambda_2 = \frac{\Delta t}{\Delta y}$, then Euler forward temporal discretization for the the semi-discretization HWENO scheme of (3.18)-(3.20) associated with approximation polynomials $\mathbf{q}_{ij}(x,y) = (\rho_{ij}(x,y), m_{ij}(x,y), n_{ij}(x,y), E_{ij}(x,y))^T$ becomes

$$\overline{q}_{ij}^{n+1} = \overline{q}_{ij}^n - \lambda_1 \sum_{\beta=1}^L \omega_\beta \left[\widehat{f}\left(q^-_{i+\frac{1}{2},\beta}, q^+_{i+\frac{1}{2},\beta}\right) - \widehat{f}\left(q^-_{i-\frac{1}{2},\beta}, q^+_{i-\frac{1}{2},\beta}\right)\right]$$
$$-\lambda_2 \sum_{\beta=1}^L \omega_\beta \left[\widehat{g}\left(q^-_{\beta,j+\frac{1}{2}}, q^+_{\beta,j+\frac{1}{2}}\right) - \widehat{g}\left(q^-_{\beta,j-\frac{1}{2}}, q^+_{\beta,j-\frac{1}{2}}\right)\right], \quad (3.30)$$

$$\overline{\xi}_{ij}^{n+1} = \overline{\xi}_{ij}^{n} - \lambda_1 \sum_{\beta=1}^{L} \omega_\beta \left[\widehat{\mathcal{H}} \left(q_{i+\frac{1}{2},\beta}^{-}, q_{i+\frac{1}{2},\beta}^{+}; \xi_{i+\frac{1}{2},\beta}^{-}, \xi_{i+\frac{1}{2},\beta}^{+} \right) - \widehat{\mathcal{H}} \left(q_{i-\frac{1}{2},\beta}^{-}, q_{i-\frac{1}{2},\beta}^{+}; \xi_{i-\frac{1}{2},\beta}^{-}, \xi_{i-\frac{1}{2},\beta}^{+} \right) \right]$$

$$-\lambda_2 \sum_{\beta=1}^{L} \omega_\beta \left[\widehat{\mathcal{R}} \left(q_{\beta,j+\frac{1}{2}}^{-}, q_{\beta,j+\frac{1}{2}}^{+}; \xi_{\beta,j+\frac{1}{2}}^{-}, \xi_{\beta,j+\frac{1}{2}}^{+} \right) - \widehat{\mathcal{R}} \left(q_{\beta,j-\frac{1}{2}}^{-}, q_{\beta,j-\frac{1}{2}}^{+}; \xi_{\beta,j-\frac{1}{2}}^{-}, \xi_{\beta,j-\frac{1}{2}}^{+} \right) \right],$$

$$(3.31)$$

$$\overline{\eta}_{ij}^{n+1} = \overline{\eta}_{ij}^{n} - \lambda_1 \sum_{\beta=1}^{L} \omega_\beta \left[\widehat{\mathcal{K}} \left(q_{i+\frac{1}{2},\beta}^{-}, q_{i+\frac{1}{2},\beta}^{+}; \eta_{i+\frac{1}{2},\beta}^{-}, \eta_{i+\frac{1}{2},\beta}^{+} \right) - \widehat{\mathcal{K}} \left(q_{i-\frac{1}{2},\beta}^{-}, q_{i-\frac{1}{2},\beta}^{+}; \eta_{i-\frac{1}{2},\beta}^{-}, \eta_{i-\frac{1}{2},\beta}^{+} \right) \right]$$

$$-\lambda_2 \sum_{\beta=1}^{L} \omega_\beta \left[\widehat{\mathcal{S}} \left(q_{\beta,j+\frac{1}{2}}^{-}, q_{\beta,j+\frac{1}{2}}^{+}; \eta_{\beta,j+\frac{1}{2}}^{-}, \eta_{\beta,j+\frac{1}{2}}^{+} \right) - \widehat{\mathcal{S}} \left(q_{\beta,j-\frac{1}{2}}^{-}, q_{\beta,j-\frac{1}{2}}^{+}; \eta_{\beta,j-\frac{1}{2}}^{-}, \eta_{\beta,j-\frac{1}{2}}^{+} \right) \right],$$

$$(3.32)$$

Let $\mu_1 = \frac{\frac{\Delta t}{\Delta x}\alpha}{\frac{\Delta t}{\Delta x}\alpha + \frac{\Delta t}{\Delta y}\alpha} = \frac{\frac{\Delta t}{\Delta x}}{\frac{\Delta t}{\Delta x} + \frac{\Delta t}{\Delta y}}$ and $\mu_2 = \frac{\frac{\Delta t}{\Delta y}\alpha}{\frac{\Delta t}{\Delta x}\alpha + \frac{\Delta t}{\Delta y}\alpha} = 1 - \mu_1$. The extension of the discussion in previous subsection to the two-dimensional case is straightforward. Following Theorem 1 in previous subsection and Theorem 10 in [22], we have

THEOREM 2. *Let* $q_{ij}(x,y) = (\rho_{ij}(x,y), m_{ij}(x,y), n_{ij}(x,y), E_{ij}(x,y))^T$ *be the approximation polynomials in* (i,j) *cell approximating* q *in the HWENO scheme* (3.30). *By the mean value theorem, there exist some points* $(x_i^1, y_j^1), (x_i^2, y_j^2), (x_i^3, y_j^3), (x_i^4, y_j^4)$ *in* (i,j) *cell such that*

$$(\rho_{ij}(x_i^1, y_j^1), m_{ij}(x_i^2, y_j^2), n_{ij}(x_i^3, y_j^3), E_{ij}(x_i^4, y_j^4))^T$$

$$= \frac{\overline{q}_{ij}^{n} - \sum_{\beta=1}^{L} \omega_\beta \widehat{\omega}_1 \left[\mu_1 \left(q_{i+\frac{1}{2},\beta}^{-} + q_{i-\frac{1}{2},\beta}^{+} \right) + \mu_2 \left(q_{\beta,j+\frac{1}{2}}^{-} + q_{\beta,j-\frac{1}{2}}^{+} \right) \right]}{1 - 2\widehat{\omega}_1}.$$

For the scheme (3.30), *if*

$$(\rho_{ij}(x_i^1, y_j^1), m_{ij}(x_i^2, y_j^2), n_{ij}(x_i^3, y_j^3), E_{ij}(x_i^4, y_j^4))^T, q_{\beta,j\pm\frac{1}{2}}^{\pm}, q_{i\pm\frac{1}{2},\beta}^{\pm}, q_{\beta,j\mp\frac{1}{2}}^{\pm}, q_{i\mp\frac{1}{2},\beta}^{\pm} \in G, \quad (3.33)$$

then $\overline{q}_{ij}^{n+1} \in G$ *under the CFL condition* $(\lambda_1 + \lambda_2) \max\{\|(|u|+c)\|_\infty, \|(|v|+c)\|_\infty\} \leqslant \widehat{\omega}_1 \alpha_0$, *where* $\alpha_0 = 1$ *for the Lax-Friedrichs flux.*

For each cell (i,j), given $\overline{q}_{ij}^{n} \in G$ and nodal values $q_{\beta,j\pm\frac{1}{2}}^{\pm}, q_{i\pm\frac{1}{2},\beta}^{\pm}, q_{\beta,j\mp\frac{1}{2}}^{\pm}, q_{i\mp\frac{1}{2},\beta}^{\pm}$ constructed in HWENO procedure, the following limiter can be used to enforce the sufficient condition (3.33).

1. Set up a small number $\varepsilon = \min_{i,j}\{10^{-13}, \overline{\rho}_{ij}^n, p(\overline{q}_{ij}^n)\}$.

2. In each cell, evaluate

$$\theta_1 = \min\left\{\left|\frac{\overline{\rho}_{ij}^n - \varepsilon}{\overline{\rho}_{ij}^n - \rho_{\min}}\right|, 1\right\}, \quad \rho_{\min} = \min\left\{\rho_{\beta,j\pm\frac{1}{2}}^{\pm}, \rho_{i\pm\frac{1}{2},\beta}^{\pm}, \rho_{\beta,j\mp\frac{1}{2}}^{\pm}, \rho_{i\mp\frac{1}{2},\beta}^{\pm}, \rho_{ij}\left(x_i^1, y_j^1\right)\right\}$$

$$(3.34)$$

3. Modify the density first: set

$$\widehat{\rho}_{\beta,j-\frac{1}{2}}^{+} = \theta_1(\rho_{\beta,j-\frac{1}{2}}^{+} - \overline{\rho}_{ij}^n) + \overline{\rho}_{ij}^n,$$

$$\widehat{\rho}_{i-\frac{1}{2},\beta}^{+} = \theta_1(\rho_{i-\frac{1}{2},\beta}^{+} - \overline{\rho}_{ij}^n) + \overline{\rho}_{ij}^n.$$

$$\widetilde{\rho}_{\beta,j+\frac{1}{2}}^{-} = \theta_1(\rho_{\beta,j+\frac{1}{2}}^{-} - \overline{\rho}_{ij}^n) + \overline{\rho}_{ij}^n,$$

$$\widetilde{\rho}_{i+\frac{1}{2},\beta}^{-} = \theta_1(\rho_{i+\frac{1}{2},\beta}^{-} - \overline{\rho}_{ij}^n) + \overline{\rho}_{ij}^n.$$

Then denote

$$\widehat{q}_{ij}^1 = \widehat{q}_{i-\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+} = \left(\widehat{\rho}_{i-\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+}, m_{i-\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+}, n_{i-\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+}, E_{i-\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+}\right)^T,$$

$$\widehat{q}_{ij}^2 = \widehat{q}_{i+\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+} = \left(\widehat{\rho}_{i+\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+}, m_{i+\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+}, n_{i+\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+}, E_{i+\frac{\sqrt{3}}{6},j-\frac{1}{2}}^{+}\right)^T,$$

$$\widehat{q}_{ij}^3 = \widehat{q}_{i-\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{+} = \left(\widehat{\rho}_{i-\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{+}, m_{i-\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{+}, n_{i-\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{+}, E_{i-\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{+}\right)^T,$$

$$\widehat{q}_{ij}^4 = \widehat{q}_{i-\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{+} = \left(\widehat{\rho}_{i-\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{+}, m_{i-\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{+}, n_{i-\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{+}, E_{i-\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{+}\right)^T,$$

$$\widehat{q}_{ij}^5 = \widetilde{q}_{i-\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-} = \left(\widetilde{\rho}_{i-\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-}, m_{i-\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-}, n_{i-\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-}, E_{i-\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-}\right)^T,$$

$$\widehat{q}_{ij}^6 = \widetilde{q}_{i+\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-} = \left(\widetilde{\rho}_{i+\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-}, m_{i+\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-}, n_{i+\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-}, E_{i+\frac{\sqrt{3}}{6},j+\frac{1}{2}}^{-}\right)^T,$$

$$\widehat{q}_{ij}^7 = \widetilde{q}_{i+\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{-} = \left(\widetilde{\rho}_{i+\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{-}, m_{i+\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{-}, n_{i+\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{-}, E_{i+\frac{1}{2},j-\frac{\sqrt{3}}{6}}^{-}\right)^T,$$

$$\widehat{q}_{ij}^8 = \widetilde{q}_{i+\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{-} = \left(\widetilde{\rho}_{i+\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{-}, m_{i+\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{-}, n_{i+\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{-}, E_{i+\frac{1}{2},j+\frac{\sqrt{3}}{6}}^{-}\right)^T,$$

and

$$\widehat{q}_{ij}^{9} = (\widehat{\rho}_{ij}(x_i^1, y_j^1), m_{ij}(x_i^2, y_j^2), n_{ij}(x_i^3, y_j^3), E_{ij}(x_i^4, y_j^4))^T$$

$$= \frac{\overline{q}_{ij}^n - \sum\limits_{\beta=1}^{L} \omega_\beta \widehat{\omega}_1 \left[\mu_1 \left(\widehat{q}_{i+\frac{1}{2},\beta}^- + \widehat{q}_{i-\frac{1}{2},\beta}^+ \right) + \mu_2 \left(\widehat{q}_{\beta,j+\frac{1}{2}}^- + \widehat{q}_{\beta,j-\frac{1}{2}}^+ \right) \right]}{1 - 2\widehat{\omega}_1}$$

4. Then modify the pressure: for $l = 1, 2, \cdots, 9$, if $p(\widehat{q}_{ij}^l) < \varepsilon$, set

$$t_\varepsilon^l = \frac{p(\overline{q}_{ij}^n) - \varepsilon}{p(\overline{q}_{ij}^n) - p(\widehat{q}_{ij}^l)}. \tag{3.35}$$

if $p(\widehat{q}_{ij}^l) \geqslant \varepsilon$, then set $t_\varepsilon^l = 1$. $\theta_2 = \min\limits_{l=1,2,\cdots,9} t_\varepsilon^l$.

Get

$$\widetilde{q}_{\beta,j+\frac{1}{2}}^- = \theta_2(\widehat{q}_{\beta,j+\frac{1}{2}}^- - \overline{q}_{ij}^n) + \overline{q}_{ij}^n, \quad \widetilde{q}_{\beta,j-\frac{1}{2}}^+ = \theta_2(\widehat{q}_{\beta,j-\frac{1}{2}}^+ - \overline{q}_{ij}^n) + \overline{q}_{ij}^n$$

$$\widetilde{q}_{i+\frac{1}{2},\beta}^- = \theta_2(\widehat{q}_{i+\frac{1}{2},\beta}^- - \overline{q}_{ij}^n) + \overline{q}_{ij}^n, \quad \widetilde{q}_{i-\frac{1}{2},\beta}^+ = \theta_2(\widehat{q}_{i-\frac{1}{2},\beta}^+ - \overline{q}_{ij}^n) + \overline{q}_{ij}^n$$

5. Use $\widetilde{q}_{\beta,j\pm\frac{1}{2}}^\pm, \widetilde{q}_{i\pm\frac{1}{2},\beta}^\pm$ instead of $q_{\beta,j\pm\frac{1}{2}}^\pm, q_{i\pm\frac{1}{2},\beta}^\pm$, in the scheme (3.30) with the CFL condition $(\lambda_1 + \lambda_2) \max\{\|(|u| + c)\|_\infty, \|(|v| + c)\|_\infty\} \leqslant \frac{1}{12}$, where the maximum is taken over $\overline{q}_{ij}^n, \widetilde{q}_{\beta,j\pm\frac{1}{2}}^\pm, \widetilde{q}_{i\pm\frac{1}{2},\beta}^\pm$ for all i, j.

3.3 The algorithm for SSP Runge-Kutta time discretization

By Theorem 1, the fourth order finite volume HWENO scheme is positivity-preserving under the suitable CFL condition $\frac{\Delta t}{\Delta x} \||u| + c\|_\infty \leq \frac{1}{12}$. But this CFL condition should be satisfied for each time stage of Runge-Kutta thus we need $\||u| + c\|_\infty$ in each time stage. Given solutions at time step n, it is hard to accurately estimate $\||u| + c\|_\infty$ in the two inner time stages in the third order SSP Runge-Kutta. On the other hand, the CFL condition $\frac{\Delta t}{\Delta x} \||u| + c\|_\infty \leq \frac{1}{12}$ is sufficient rather than necessary for the sake of preserving the positivity of cell averages. To evolve to time step $n+1$, we can start with a larger CFL (e.g., $\frac{\Delta t}{\Delta x} \||u| + c\|_\infty \leq \frac{1}{3}$ for a fourth order finite volume scheme) in time step n to save computational cost. If negative cell averages emerge in any of three stage in one step of Runge-Kutta, then return

16

to stage one in time step n and restart the computation with the stringent CFL condition. Such an efficient algorithm was used in [16].

This algorithm for the scheme in Section 3.1 with the third order SSP Runge-Kutta method (2.10) can be implemented as follows:

1. At time level n, in each cell I_j, given $\overline{q}_j^n \in G$ and reconstructed point values $q_{j+\frac{1}{2}}^-$ and $q_{j-\frac{1}{2}}^+$, apply the limiter to obtain $\widetilde{q}_{j+\frac{1}{2}}^-, \widetilde{q}_{j-\frac{1}{2}}^+ \in G$.

2. Compute $\max \||u| + c\|_\infty$ by taking the maximum over $\overline{q}_j^n, \widetilde{q}_{j+\frac{1}{2}}^-$ and $\widetilde{q}_{j-\frac{1}{2}}^+$ for all j.

3. Set the time step $\Delta t = a \frac{\Delta x}{\max \||u|+c\|_\infty}$ where a is a linearly stable CFL number for a fourth order finite volume scheme. For example, we can use $a = \frac{1}{3}$ in practice.

4. Compute the first stage, denoted by $q_j^{(1)}$.

   - If the cell averages $\overline{q}_j^{(1)}$ are positive, then proceed to the next step.

   - If the cell averages $\overline{q}_j^{(1)}$ contain negative density or pressure, then recompute the first stage with the stringent CFL, $\Delta t = \frac{1}{12} \frac{\Delta x}{\max \||u|+c\|_\infty}$. Notice that Theorem 1 guarantees that the cell averages $\overline{q}_j^{(1)}$ will be positive with this CFL.

5. Given $\overline{q}_j^{(1)} \in G$ and reconstructed point values $q_{j+\frac{1}{2}}^{-,(1)}$ and $q_{j-\frac{1}{2}}^{+,(1)}$, apply the limiter to obtain $\widetilde{q}_{j+\frac{1}{2}}^{-,(1)}, \widetilde{q}_{j-\frac{1}{2}}^{+,(1)} \in G$. Compute the second stage, denoted by $q_j^{(2)}$.

   - If the cell averages $\overline{q}_j^{(2)}$ are positive, then proceed to the next step.

   - If the cell averages $\overline{q}_j^{(2)}$ contain negative density or pressure, then return to Step 4 and restart the computation with half time step. Notice that even if the time step is already $\Delta t = \frac{1}{12} \frac{\Delta x}{\max \||u|+c\|_\infty}$ in Step 4, there is no guarantee that $\overline{q}_j^{(2)}$ should be positive because the wave speed $\max \||u| + c\|_\infty$ was computed based on q_j^n rather than $q_j^{(1)}$.

6. Given $\overline{q}_j^{(2)} \in G$ and reconstructed point values $q_{j+\frac{1}{2}}^{-,(2)}$ and $q_{j-\frac{1}{2}}^{+,(2)}$, apply the limiter to obtain $\widetilde{q}_{j+\frac{1}{2}}^{-,(2)}, \widetilde{q}_{j-\frac{1}{2}}^{+,(2)} \in G$. Compute q_j^{n+1}.

- If the cell averages $\overline{q}_j^{n+1}$ are positive, then computation to time step $n+1$ is done.

- If the cell averages $\overline{q}_j^{n+1}$ contain negative density or pressure, then return to Step 4 and restart the computation with half time step.

REMARK 3. *Theorem 1 implies that the implementation above will not result in any endless loop. When time step is small enough such that $\Delta t \leq \frac{1}{12} \frac{\Delta x}{\alpha^*}$ where α^* is larger than or equal to the maximum wave speed in all inner stages, $\overline{q}_j^{n+1}$ will be positive.*

4   Numerical tests for the Ideal Gas

In this section, we perform a detailed comparison of the fifth order finite volume HWENO scheme with the positivity-preserving limiter (HWENO5) and the fifth order finite volume WENO scheme with the positivity-preserving limiter (WENO5) in 1D dimensional cases and show some results of the fourth order finite volume HWENO scheme in two dimensional case with the positivity-preserving limiter (HWENO4) for several demanding examples. The HWENO schemes without the positivity-preserving limiter will blow up for most examples in this section.

EXAMPLE 1. *Accuracy test.*

Consider the vortex evolution problem [24] for (2.1). The mean flow is $\rho = p = u = v = 1$. Add to the mean flow an isentropic vortex perturbation centered at (x_0, y_0) in (u, v) and $T = p/\rho$, no perturbation in $S = p/\rho^\gamma$,

$$(\delta u, \delta v) = \frac{\epsilon}{2\pi} e^{0.5(1-r^2)}(-\overline{y}, \overline{x}), \ \delta T = \frac{(\gamma-1)\epsilon^2}{8\gamma\pi^2} e^{1-r^2},$$

where $(\overline{x}, \overline{y}) = (x - x_0, y - y_0), r^2 = \overline{x}^2 + \overline{y}^2$. The exact solution is the passive convection of the vortex with the mean velocity.

The domain is taken as $[-5, 15] \times [-5, 15]$ and $(x_0, y_0) = (5, 5)$. The boundary condition is periodic. We set $\gamma = 1.4$ and the vortex strength $\epsilon = 10.0828$ such that the lowest density and lowest pressure of the exact solution are 7.8×10^{-15} and 1.7×10^{-20}. We test the accuracy

of positivity-preserving limiter on the fourth order finite volume HWENO scheme with the third order SSP Runge-Kutta at $T = 0.01$ with $\Delta x = \Delta y$ under $CFL = 0.2$.

In Table 4.1, we find the finite volume HWENO scheme with positivity-preserving limiter can ensure the positivity of density ρ and pressure p. To see how many positivity limiters were actually used in this example, we recorded the number of cells where positivity-preserving limiter was activated (namely, $\theta < 1$). For each time stage, the percentage of the such cells is listed as well. We clearly observe the accuracy of the HWENO scheme with limiters is formally fourth order in both L_1 norm and L_∞ norm.

Table 4.1: Fourth order finite volume HWENO scheme with the positivity-preserving limiter, for the vortex evolution problem, $T = 0.01$, and $\Delta x = \Delta y$. ρ_{min} and p_{min} are minimum density and pressure of the numerical solution respectively.

N	L_1 error	Order	L_∞ error	Order	ρ_{min}	p_{min}	limited
20	2.68E-04	-	1.71E-02	-	1.96E-01	1.43E-01	$\leq 22\%$
40	5.29E-05	2.34	3.42E-03	2.32	1.60E-02	5.29E-03	$\leq 8.31\%$
80	5.69E-06	3.22	1.09E-03	1.65	8.11E-04	1.23E-04	$\leq 3.95\%$
160	6.43E-07	3.14	2.07E-04	2.40	3.90E-05	2.00E-06	$\leq 1.93\%$
320	3.23E-08	4.32	2.23E-05	3.21	1.13E-06	9.97E-08	$\leq 0.94\%$
640	1.62E-09	4.32	1.05E-06	4.41	6.78E-08	1.99E-9	$\leq 0.47\%$

EXAMPLE 2. *Sedov blast waves.*

The Sedov point-blast wave is a typical low density and low pressure problem involving shocks. The exact solution formula can be found in [13, 7].

The computational domain is taken to be $[-2, 2]$. The boundary condition is outflow. For the initial condition, the density is 1, velocity is zero, total energy is 10^{-13} everywhere except that the energy in the center cell is the constant $\frac{E_0}{\Delta x}$ with $E_0 = 3200000$ (emulating a δ-function at the center). $\gamma = 1.4$. The computational results with the positivity-preserving limiter at $T = 0.001$ on a mesh size of $\Delta x = 0.01$ are shown in Figure 4.1. For each time stage, the percentage of cells where positivity limiter was activated is less than 0.5%. By comparing with WENO5, we can observe that a slightly sharper blast wave is obtained by using the HWENO5 scheme.

(a) pressure

(b) velocity

(c) density
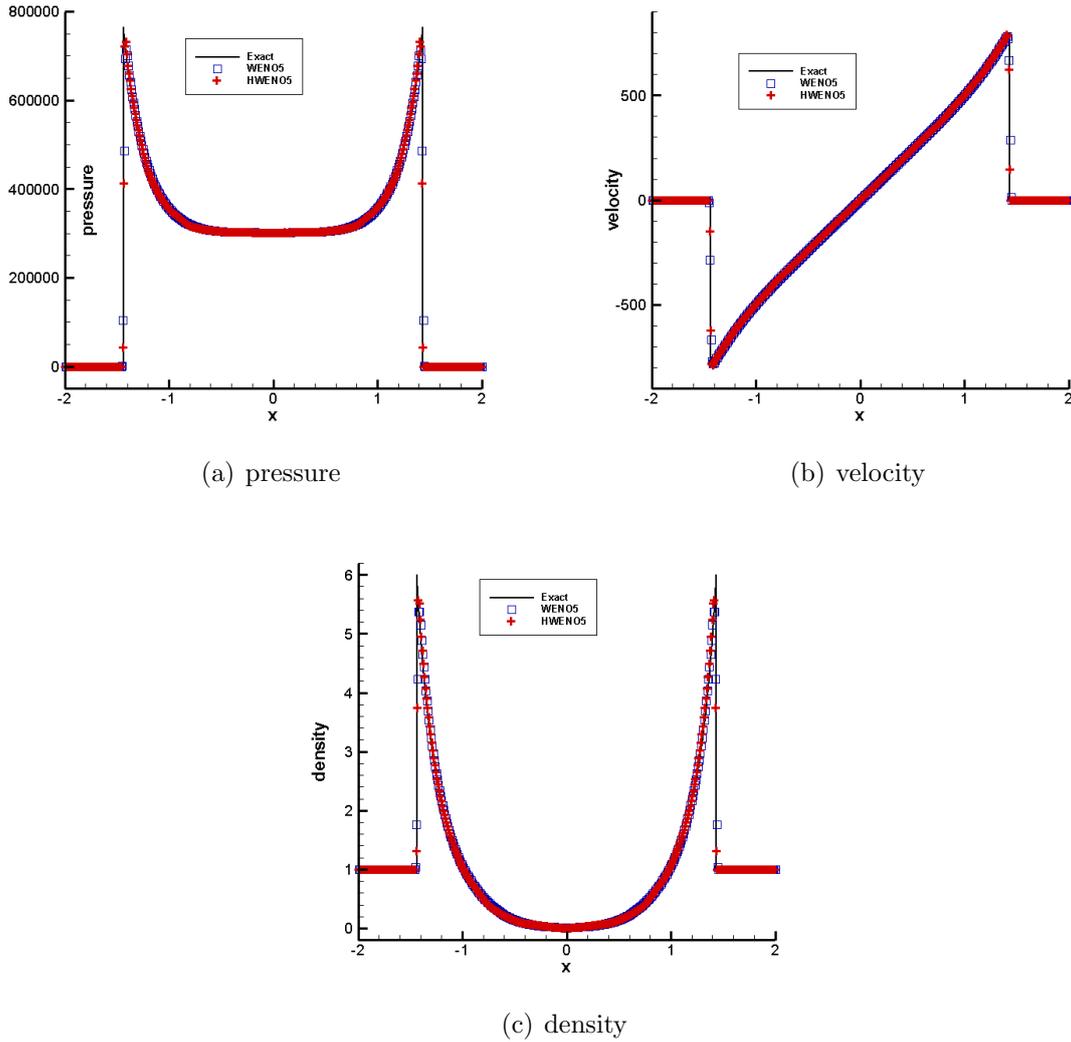
Figure 4.1: 1D Sedov blast. T=0.001. $\Delta x = 0.01$. HWENO5 and WENO5. The solid line is the exact solution. Squares: numerical solution of WENO5; Pluses: numerical solution of HWENO5.

EXAMPLE 3. *The extreme one-dimensional double rarefaction problem.*

The initial condition is $\rho_L = \rho_R = 7, u_L = -1, u_R = 1, p_L = p_R = 0.2, \gamma = 1.4$. The computational domain is taken to be $[-1, 1]$. The boundary condition is outflow.

The exact solution contains vacuum. The results of positivity-preserving fifth order HWENO schemes and positivity-preserving fifth order WENO schemes at $T = 0.6$ on a mesh size of $\Delta x = 1/200$ under $CFL = 0.5$ are shown in Figure 4.2. For each time stage, the percentage of cells where positivity limiter applied is less than 0.5%. In the right panel

20

of Figure 4.2, we can observe that both schemes preserves positive density and pressure and capture shocks well.

EXAMPLE 4. *A two-dimensional double rarefaction with the initial condition* $\rho_L = \rho_R = 7, u_L = -1, u_R = 1, v_L = v_R = 0, p_L = p_R = 0.2.$

The computational domain is $[-1, 1] \times [-1, 1]$. The boundary condition is outflow. The results of cutting at $y = 0$ of the problem at $T = 0.6$ with $\Delta x = 1/200$ are presented. The results of HWENO4 are very well. The maximum percentage of limited cells is 1%.

EXAMPLE 5. *Shock diffraction problem.*

The setup is the following: the computational domain is the union of $[0, 1] \times [6, 11]$ and $[1, 13] \times [0, 11]$; the initial condition is a pure right-moving shock of $Mach = 5.09$, initially located at $x = 0.5$ and $6 \leqslant y \leqslant 11$, moving into undisturbed air ahead of the shock with a density of 1.4 and pressure of 1. The boundary conditions are inflow at $x = 0, 6 \leqslant y \leqslant 11$, outflow at $x = 13, 0 \leqslant y \leqslant 11, 1 \leqslant x \leqslant 13, y = 0$ and $0 \leqslant x \leqslant 13, y = 11$, and reflective at the walls $0 \leqslant x \leqslant 1, y = 6$ and $x = 1, 0 \leqslant y \leqslant 6. \lambda = 1.4$. The density and pressure at $t = 2.3$ are presented in Figure 4.4 and Figure 4.5. For each time stage, if the positivity-preserving finite volume HWENO scheme with $\Delta x = \Delta y = \frac{1}{32}$ applied to solve the problem, the percentage of the cells that need the usage of the positivity-preserving limiter is less than 1.1%. Such ratio is 0.55% in case of the positivity-preserving finite volume HWENO scheme with $\Delta x = \Delta y = \frac{1}{64}$.

The results are comparable to those of positivity-preserving DG method [20], finite volume WENO scheme [22] and finite difference WENO scheme [24].

5   Conclusions

In this paper, we have proposed the positivity-preserving finite volume HWENO schemes in both one dimension and two dimension based on a general framework to construct arbitrarily high order schemes which can preserve the positivity of density and pressure for

conservation laws in [20]. The present schemes keep the essentially non-oscillatory properties for low density and low pressure problems. Compared to positivity-preserving finite volume WENO schemes in one-dimensional case, positivity-preserving finite volume HWENO schemes can produce better resolutions in several examples due to its compactness properties. Extensions of our HWENO scheme for Euler equations with a source term constitute our future work.

References

[1] Capdeville, G, A Hermite upwind WENO scheme for solving hyperbolic conservation laws, Journal of Computational Physics, 227 (2008), 2430-2454.

[2] B. Cockburn and C.-W. Shu, Runge-Kutta discontinuous Galerkin method for convection-dominated problems, Journal of Scientific Computing, 16 (2001), 173-261.

[3] Y. Ha and C. L. Gardner, Positive scheme numerical simulation of high mach number astrophysical jets, Journal of Scientific Computing, 34 (2008), 247-259.

[4] Y. Ha, C. Gardner, A. Gelb and C.-W. Shu, Numerical simulation of high Mach number astrophysical jets with radiative cooling, Journal of Scientific Computing, 24 (2005), 597-612.

[5] X.Y. Hu, N.A. Adams, C.-W. Shu, Positivity-preserving method for high-order conservative schemes solving compressible Euler equations. Journal of Computational Physics, 242 (2013), 169-180.

[6] G. Jiang, C.-W. Shu, Efficient implementation of weighted ENO schemes Journal of Computational Physics, 126 (1996), 202-228.

[7] V.P. Korobeinikov, Problems of Point-Blast Theory, American Institute of Physics,1991.

[8] S.K. Lele, Compact finite-difference schemes with spectra-like resolution, Journal of Computational Physics, 103 (1992), 16-42.

[9] H. Liu and J. Qiu, Finite Difference Hermite WENO schemes for conservation laws, Journal of Scientific Computing, 63 (2015), 548-572.

[10] B.Perthame and C.-W. Shu, On positivity preserving finite volume schemes for Euler equations, Numerische Mathematik, 73 (1996), 119-130.

[11] J. Qiu and C.-W. Shu, Hermite WENO schemes and their application as limiters for Runge-Kutta discontinuous Galerkin method: one-dimensional case, Journal of Computational Physics, 193 (2004) 115-135.

[12] J. Qiu and C.-W. Shu, Hermite WENO schemes and their application as limiters for Runge-Kutta discontinuous Galerkin method II: Two dimensional case, Computers & Fluids , 34 (2005) 642-663.

[13] L.I. Sedov, Similarity and Dimensional Methods in Mechanics, Academic Press, New York, 1959.

[14] C.-W. Shu, S. Osher, Efficient implementation of essentially non-oscillatory shock-capturing schemes, Journal of Computational Physics 77 (1988) 439-471.

[15] B. Van-Leer, Towards the ultimate conservative difference scheme: III. A new approach to numerical convection, Journal of Computational Physics. 23 (1977) 276-299.

[16] C. Wang, X. Zhang, C.-W. Shu and J. Ning, Robust high order discontinuous Galerkin schemes for two-dimensional gaseous detonations, Journal of Computational Physics, 231 (2012), 653-665.

[17] Y. Xing, X. Zhang, and Chi-Wang Shu, Positivity-preserving high order well-balanced discontinuous Galerkin methods for the shallow water equations. Advances in Water Resources 33 (2010), 1476-1493.

[18] Y. Xing and X. Zhang, Positivity-preserving well-balanced discontinuous Galerkin methods for the shallow water equations on unstructured triangular meshes, Journal of Scientific Computing, 57 (2013), 19-41.

[19] X. Zhang and C.-W. Shu, On maximum-principle-satisfying high order schemes for scalar conservation laws , Journal of Computational Physics, 229 (2010), 3091-3120.

[20] X. Zhang and C.-W. Shu, On positivity preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes, Journal of Computational Physics, 229 (2010), 8918-8934.

[21] X. Zhang and C.-W. Shu, Positivity preserving high order discontinuous Galerkin schemes for compressible Euler equations with source terms, Journal of Computational Physics, 230 (2011), 1238-1248.

[22] X. Zhang and C.-W. Shu, Maximum-principle-satisfying and positivity-preserving high order schemes for conservation laws: Survey and new developments, Proceedings of the Royal Society A, 467 (2011), 2752-2776.

[23] X. Zhang, Y. Xia and C.-W. Shu, Maximum-principle-satisfying and positivity-preserving high order discontinuous Galerkin schemes for conservation laws on triangular meshes, Journal of Scientific Computing, 50 (2012), 29-62.

[24] X. Zhang and C.-W. Shu, Positivity-preserving high order finite difference WENO schemes for compressible Euler equations, Journal of Computational Physics, 231 (2012), 2245-2258.

[25] X. Zhang, Y.-Y. Liu and C.-W. Shu, Maximum-principle-satisfying high order finite volume WENO schemes for convection-diffusion equations, SIAM Journal on Scientific Computing, 34 (2012), A627-A658.

[26] Y. Zhang, X. Zhang and C.-W. Shu, Maximum-principle-satisfying second order discontinuous Galerkin schemes for convection-diffusion equations on triangular meshes, Journal of Computational Physics, 234 (2013), 295-316.

[27] X. Zhang and C.-W. Shu, A minimum entropy principle of high order schemes for gas dynamics equations, Numerische Mathematik, 121 (2012), 545-563.

[28] J. Zhu and J. Qiu, A Class of Fourth order Finite Volume Hermite Weighted Essentially Non-oscillatory Schemes, Science in China, Series A–Mathematics, 51 (2008), 1549-1560.
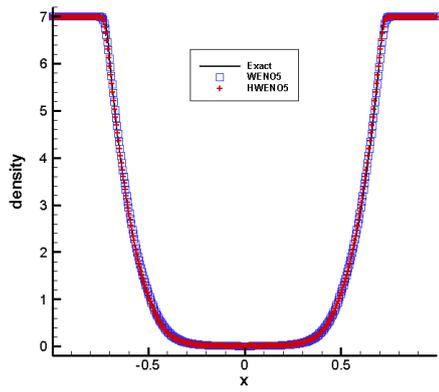
(a) density
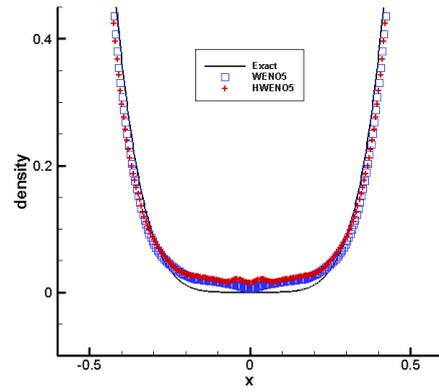
(b) density
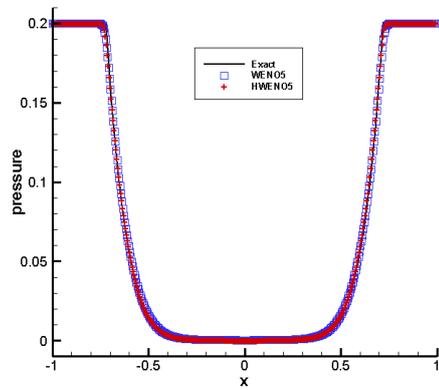
(c) pressure
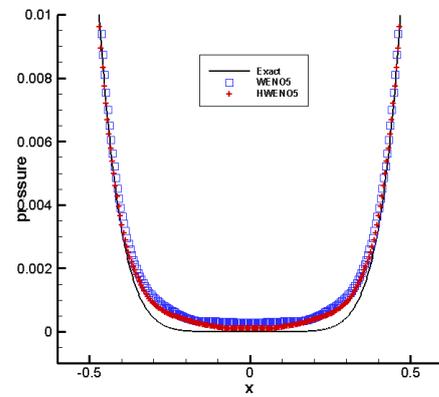
(d) pressure
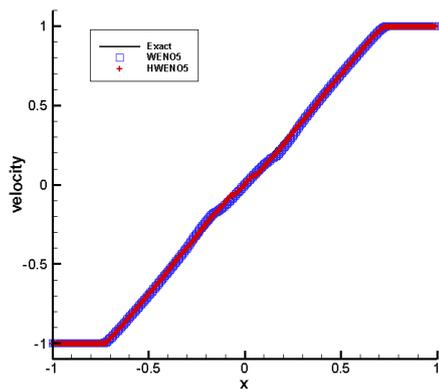
(e) velocity
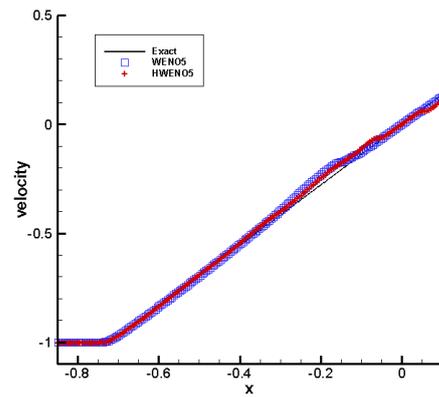
(f) velocity

Figure 4.2: Double rarefaction. T=0.6. left: 1D problem. Right: the zoom. $\Delta x = 1/200$. HWENO5 and WENO5. The solid line is the exact solution. Squares: numerical solution of WENO5; Pluses: numerical solution of HWENO5.
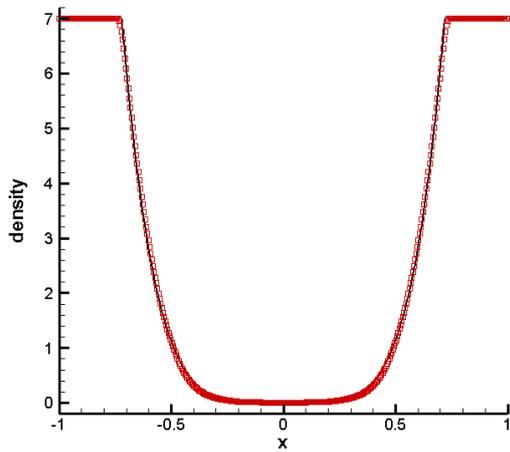
(a) density
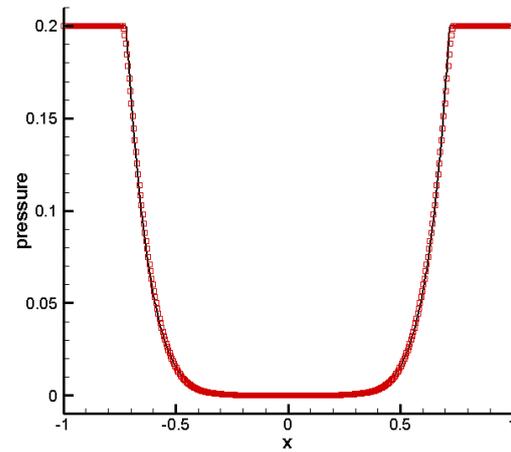
(b) pressure
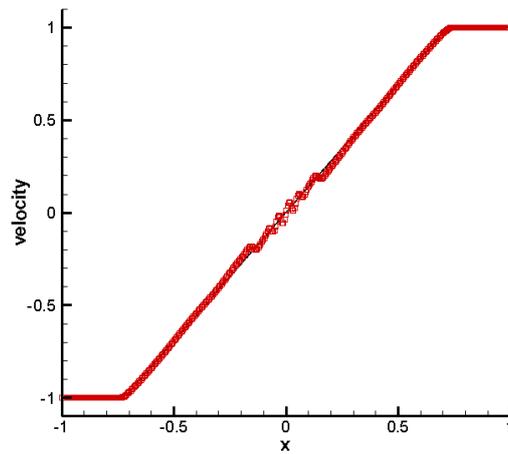
(c) velocity

Figure 4.3: Double rarefaction. T=0.6. Cut at $y = 0$. $\Delta x = 1/200$. HWENO scheme with the positive-preserving limiter. The solid line is the exact solution. Symbols are numerical solutions

27
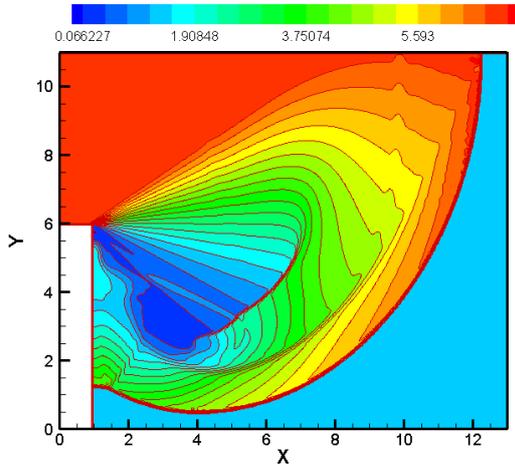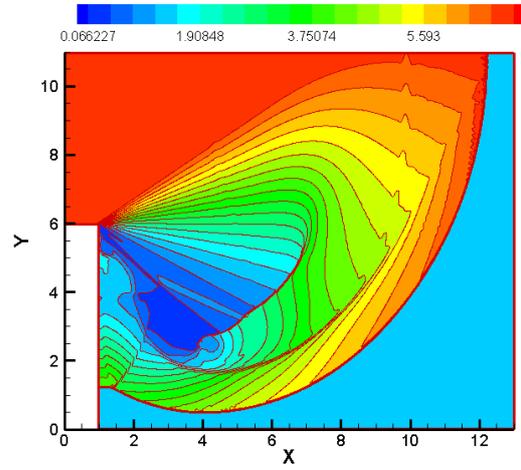
(a) $\Delta x = \Delta y = 1/32$ (b) $\Delta x = \Delta y = 1/64$

Figure 4.4: Shock diffraction problem. $T = 2.3$. Density, 20 equally spaced contour lines from $\rho = 0.066227$ to $\rho = 7.0668$.
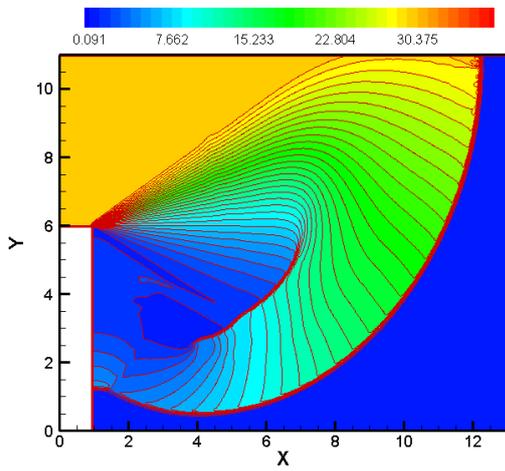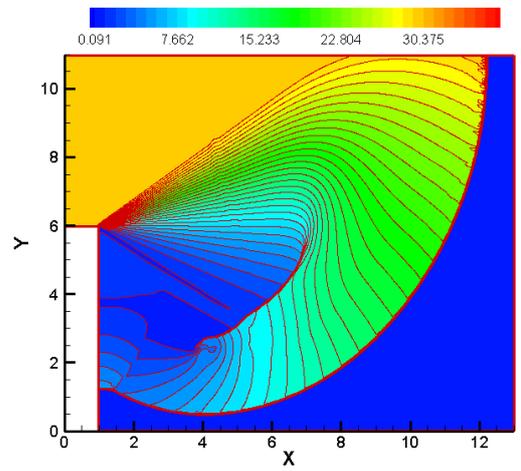
(a) $\Delta x = \Delta y = 1/32$ (b) $\Delta x = \Delta y = 1/64$

Figure 4.5: Shock diffraction problem. $T = 2.3$. Pressure, 40 equally spaced contour lines from $p = 0.091$ to $p = 37$.