Numer. Math. (2012) 121:545–563
DOI 10.1007/s00211-011-0443-7

Numerische Mathematik

A minimum entropy principle of high order schemes for gas dynamics equations

Xiangxiong Zhang · Chi-Wang Shu

Received: 17 July 2011 / Revised: 15 September 2011 / Published online: 22 December 2011

Abstract The entropy solutions of the compressible Euler equations satisfy a minimum principle for the specific entropy (Tadmor in Appl Numer Math 2:211–219, 1986). First order schemes such as Godunov-type and Lax-Friedrichs schemes and the second order kinetic schemes (Khobalatte and Perthame in Math Comput 62:119–131, 1994) also satisfy a discrete minimum entropy principle. In this paper, we show an extension of the positivity-preserving high order schemes for the compressible Euler equations in Zhang and Shu (J Comput Phys 229:8918–8934, 2010) and Zhang et al. (J Scientific Comput, in press), to enforce the minimum entropy principle for high order finite volume and discontinuous Galerkin (DG) schemes.

Mathematics Subject Classification (2010) 65M60 · 76N15

1 Introduction

The one dimensional version of the compressible Euler equations for the perfect gas in gas dynamics is given by

Research supported by AFOSR grant FA9550-09-1-0126 and NSF grant DMS-1112700.

X. Zhang · C.-W. Shu (✉)
Division of Applied Mathematics, Brown University,
Providence, RI 02912, USA
e-mail: shu@dam.brown.edu

Present Address:
X. Zhang
Department of Mathematics, MIT,
Cambridge, MA 02139, USA
e-mail: zhangxx@dam.brown.edu

$$\mathbf{w}_t + \mathbf{f}(\mathbf{w})_x = 0, \quad t \geq 0, \ x \in \mathbb{R}, \tag{1.1}$$
$$\mathbf{w} = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix}, \quad \mathbf{f}(\mathbf{w}) = \begin{pmatrix} m \\ \rho u^2 + p \\ (E + p)u \end{pmatrix}$$

with

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

where ρ is the density, u is the velocity, m is the momentum, E is the total energy and p is the pressure. We consider the initial value problem for system (1.1) with the initial data $\mathbf{w}_0(x)$.

It is well known that entropy inequalities should be considered for general hyperbolic conservation laws. The generalized entropy function for (1.1) is a smooth convex function $U(\mathbf{w})$ with an entropy flux $F(\mathbf{w})$ such that the following relation holds:

$$U_{\mathbf{w}}^T \mathbf{f}_{\mathbf{w}} = F_{\mathbf{w}}.$$

Entropy solutions of (1.1) are weak solutions which in addition satisfy $U(\mathbf{w})_t + F(\mathbf{w})_x \leq 0$ in the sense of distributions for all the entropy pairs (U, F).

In [11], a minimum principle of the specific entropy $S(x, t) = \ln \frac{p}{\rho^\gamma}$ was proved for the entropy solutions:

$$S(x, t + h) \geq \min\{S(y, t) : |y - x| \leq ||u||_\infty h\}.$$

The first order schemes including Godunov and Lax-Friedrichs schemes preserve a similar discrete property [11]. In [6], a first order kinetic scheme for multi-dimensional cases on a general mesh and a second order kinetic scheme satisfying the same property were discussed. However, it seems difficult to construct higher order minimum-entropy-principle-satisfying schemes. The minimum principle of specific entropy in [11] is so far the best pointwise estimate of entropy for gas dynamics equations, which is different from any estimate of total entropy. In particular, it was reported in [6] that enforcing this minimum entropy principle numerically might damp oscillations in numerical solutions.

In this paper, we will discuss the minimum entropy principle of an arbitrarily high order scheme on a rectangular or unstructured triangular mesh. To have the specific entropy well-defined, the very first step is to guarantee the positivity of density and pressure of the numerical solution, which can be done for a high order finite volume or a discontinuous Galerkin (DG) scheme following [7,14–16]. The main idea of positivity-preserving techniques for high order schemes in [14] is to find a sufficient condition to preserve the positivity of the cell averages by repeated convex combinations, namely,

1. Use strong stability preserving (SSP) high order time discretizations which are convex combinations of forward Euler. For more details, see [3,4,8,9]. Then it

suffices to find a way to preserve the positivity for the forward Euler time discretization since the set of states with positive density and positive pressure is convex.
2. Use first order schemes which can keep the positivity of density and pressure as building blocks. High order spatial discretization with forward Euler is equivalent to a convex combination of formal first order schemes, thus will keep the positivity provided a certain sufficient condition is satisfied.
3. A simple conservative limiter can enforce the sufficient condition without destroying accuracy for smooth solutions.

In fact, the methodology above can be used to enforce any property for high order schemes as long as the states satisfying this property form a convex set. In particular, we will show in Sect. 2 that the specific entropy function is quasi-concave, thus the following set is convex,

$$G = \left\{ \mathbf{w} = \begin{pmatrix} \rho \\ m \\ E \end{pmatrix} \middle| \rho > 0, \quad p > 0, \quad \text{and} \quad S \geq S_0 = \min_x S(\mathbf{w}_0(x)) \right\}. \quad (1.2)$$

Therefore, we can easily derive a sufficient condition for a high order scheme to keep numerical solutions lie in G, i.e., the minimum of the specific entropy at any later time will be bounded from below by the initial minimum. Then a straightforward extension of the limiter in [14] can enforce this condition without destroying conservation. This limiter will not destroy accuracy for generic smooth solutions, to be explained in Sect. 2.

The conclusion of this paper is, by adding a simple limiter which will be specified later to a high order accurate finite volume scheme, e.g., the essentially non-oscillatory (ENO) and the weighted ENO (WENO) finite volume schemes, or a discontinuous Galerkin scheme solving one or multi-dimensional Euler equations, with the time evolution by a SSP Runge–Kutta or multi-step method, the final scheme satisfies the minimum entropy principle and remains high order accurate for generic smooth solutions.

The paper is organized as follows. We first describe the one-dimensional case in Sect. 2. Then we discuss the two-dimensional cases in Sect. 3. In Sect. 4, we show the numerical tests for high order DG schemes. Concluding remarks are given in Sect. 5.

2 The one-dimensional case

2.1 Preliminaries

Lemma 2.1 *$S(\mathbf{w}) = \ln \frac{p}{\rho^\gamma}$ is a quasi-concave function, namely, the following inequality holds,*

$$S(\lambda_1 \mathbf{w}_1 + \lambda_2 \mathbf{w}_2) > \min\{S(\mathbf{w}_1), S(\mathbf{w}_2)\}, \quad \textit{if} \quad \rho_1, \rho_2 > 0,$$

where $\mathbf{w}_1 \neq \mathbf{w}_2$ $\lambda_1, \lambda_2 > 0$ and $\lambda_1 + \lambda_2 = 1$.

Proof Let $U(\mathbf{w}) = -\rho h(S(\mathbf{w}))$, then $U_{\mathbf{ww}}$ is positive definite if and only if $\rho(h'(S) - \gamma h''(S)) > 0$ and $h'(S) > 0$, see [5]. In particular, we can take $h(S) = S$. Let $\bar{\mathbf{w}} = \lambda_1 \mathbf{w}_1 + \lambda_2 \mathbf{w}_2$ and $S^* = \min\{S(\mathbf{w}_1), S(\mathbf{w}_2)\}$. $U_{\mathbf{ww}} > 0$ implies $U(\bar{\mathbf{w}}) < \lambda_1 U(\mathbf{w}_1) + \lambda_2 U(\mathbf{w}_2)$. So

$$-\bar{\rho} S(\bar{\mathbf{w}}) < -\lambda_1 \rho_1 S(\mathbf{w}_1) - \lambda_2 \rho_2 S(\mathbf{w}_2) \le -\lambda_1 \rho_1 S^* - \lambda_2 \rho_2 S^* = -\bar{\rho} S^*.$$

Thus we have $S(\bar{\mathbf{w}}) > S^* = \min\{S(\mathbf{w}_1), S(\mathbf{w}_2)\}$. □

Lemma 2.2 *For a vector valued function* $\mathbf{w}(x) = (\rho(x), m(x), E(x))^T$ *defined on an interval* $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ *satisfying* $\rho(x) > 0$ *for all* $x \in I_j$, *we have*

$$S\left(\frac{1}{\Delta x}\int_{I_j} \mathbf{w}(x)\,dx\right) \ge \min_{x\in I_j} S(\mathbf{w}(x)),$$

where $\Delta x = x_{j+\frac{1}{2}} - x_{j-\frac{1}{2}}$.

Proof Define $U(\mathbf{w}) = -\rho S$ then U is a convex function. Let $\bar{\rho} = \frac{1}{\Delta x}\int_{I_j} \rho(x)dx$ and $\bar{\mathbf{w}} = \frac{1}{\Delta x}\int_{I_j} \mathbf{w}(x)dx$. Jensen's inequality implies

$$\begin{aligned} -\bar{\rho} S(\bar{\mathbf{w}}) &= U\left(\frac{1}{\Delta x}\int_{I_j} \mathbf{w}(x)\,dx\right) \le \frac{1}{\Delta x}\int_{I_j} U\left(\mathbf{w}(x)\right)dx \\ &= -\int_{I_j} \frac{1}{\Delta x}\rho(x) S(\mathbf{w}(x))\,dx \le -\bar{\rho}\min_{x\in I_j} S(\mathbf{w}(x)). \end{aligned}$$

□

Therefore, G defined in (1.2) is a convex set by the concavity of pressure and Lemma 2.1. The entropy solutions are in the set G, see [11].

Consider a first order scheme for (1.1)

$$\mathbf{w}_j^{n+1} = \mathbf{w}_j^n - \lambda[\widehat{\mathbf{f}}(\mathbf{w}_j^n, \mathbf{w}_{j+1}^n) - \widehat{\mathbf{f}}(\mathbf{w}_{j-1}^n, \mathbf{w}_j^n)], \tag{2.1}$$

where $\widehat{\mathbf{f}}(\cdot,\cdot)$ is a numerical flux, n refers to the time step and j to the spatial cell (we assume uniform mesh size only for simplicity), and $\lambda = \frac{\Delta t}{\Delta x}$ is the ratio of time and space mesh sizes. $\mathbf{w}_j^n$ is the approximation to the cell average of the exact solution $\mathbf{v}(x,t)$ in the cell $I_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$, or the point value of the exact solution $\mathbf{v}(x,t)$ at x_j, at time level n. For Godunov, Lax–Friedrichs and kinetic type fluxes [6], the scheme (2.1) satisfies that $\mathbf{w}_j^n$ 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

$$\lambda \parallel (|u| + c) \parallel_\infty \le \alpha_0, \tag{2.2}$$

where α_0 is a constant depending on the flux.

Recall that the numerical solutions of Godunov scheme are the cell averages of the exact solution if $\lambda \parallel (|u|+c) \parallel_\infty \leq 1$. Thus Lemma 2.2 implies $\alpha_0 = 1$ for the Godunov flux. Following the same proof as that in the Appendix of [7], it is straightforward to check that $\alpha_0 = \frac{1}{2}$ for the Lax–Friedrichs flux.

2.2 High order schemes

We now consider a general high order finite volume scheme, or the scheme satisfied by the cell averages of a DG method solving (1.1), with forward Euler time discretization, which has the following form

$$\overline{\mathbf{w}}_j^{n+1} = \overline{\mathbf{w}}_j^n - \lambda \left[\widehat{\mathbf{f}} \left(\mathbf{w}_{j+\frac{1}{2}}^-, \mathbf{w}_{j+\frac{1}{2}}^+ \right) - \widehat{\mathbf{f}} \left(\mathbf{w}_{j-\frac{1}{2}}^-, \mathbf{w}_{j-\frac{1}{2}}^+ \right) \right], \tag{2.3}$$

where $\widehat{\mathbf{f}}$ is Godunov, Lax–Friedrichs or kinetic type flux, $\overline{\mathbf{w}}_j^n$ 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 $\mathbf{w}_{j+\frac{1}{2}}^- = \mathbf{q}_j(x_{j+\frac{1}{2}})$.

We need the 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 would need to choose N such that $2N-3 \geq k$. Denote these quadrature points on I_j as

$$\left\{ x_{j-\frac{1}{2}} = \widehat{x}_j^1, \widehat{x}_j^2, \ldots, \widehat{x}_j^{N-1}, \widehat{x}_j^N = x_{j+\frac{1}{2}} \right\}. \tag{2.4}$$

Let $\widehat{w}_\alpha$ be the quadrature weights for the interval $[-\frac{1}{2}, \frac{1}{2}]$ such that $\sum_{\alpha=1}^N \widehat{w}_\alpha = 1$.

Theorem 2.3 *The high order scheme* (2.3) *satisfies a minimum entropy principle, i.e., assuming the numerical solution at time level n has positive density and positive pressure, then* $\overline{\mathbf{w}}_j^{n+1}$ *has positive density and positive pressure, and*

$$S(\overline{\mathbf{w}}_j^{n+1}) \geq \min \left\{ \min_\alpha S(\mathbf{q}_j(\widehat{x}_j^\alpha)), S(\mathbf{w}_{j+\frac{1}{2}}^+), S(\mathbf{w}_{j-\frac{1}{2}}^-) \right\},$$

under the CFL condition

$$\lambda \parallel (|u|+c) \parallel_\infty \leq \widehat{w}_1 \alpha_0. \tag{2.5}$$

In particular, if $\mathbf{q}_j(\widehat{x}_j^\alpha) \in G$ *for all* j *and* α, *then* $\overline{\mathbf{w}}_j^{n+1} \in G$.

Proof The positivity of density and pressure of $\overline{\mathbf{w}}_j^{n+1}$ was proved in Theorem 2.1 of [14]. Thus $S(\overline{\mathbf{w}}_j^{n+1})$ is well-defined. The exactness of the quadrature rule for polynomials of degree k implies

$$\overline{\mathbf{w}}_j^n = \frac{1}{\Delta x}\int_{I_j} \mathbf{q}_j(x)\,dx = \sum_{\alpha=1}^{N} \widehat{w}_\alpha \mathbf{q}_j(\widehat{x}_j^\alpha).$$

By adding and subtracting $\widehat{\mathbf{f}}\left(\mathbf{w}^+_{j-\frac{1}{2}}, \mathbf{w}^-_{j+\frac{1}{2}}\right)$, the scheme (2.3) becomes

$$\begin{aligned}
\overline{\mathbf{w}}_j^{n+1} &= \sum_{\alpha=1}^{N} \widehat{w}_\alpha \mathbf{q}_j(\widehat{x}_j^\alpha) - \lambda\left[\widehat{\mathbf{f}}\left(\mathbf{w}^-_{j+\frac{1}{2}}, \mathbf{w}^+_{j+\frac{1}{2}}\right) - \widehat{\mathbf{f}}\left(\mathbf{w}^+_{j-\frac{1}{2}}, \mathbf{w}^-_{j+\frac{1}{2}}\right)\right.\\
&\quad \left. + \widehat{\mathbf{f}}\left(\mathbf{w}^+_{j-\frac{1}{2}}, \mathbf{w}^-_{j+\frac{1}{2}}\right) - \widehat{\mathbf{f}}\left(\mathbf{w}^-_{j-\frac{1}{2}}, \mathbf{w}^+_{j-\frac{1}{2}}\right)\right]\\
&= \sum_{\alpha=2}^{N-1} \widehat{w}_\alpha \mathbf{q}_j(\widehat{x}_j^\alpha) + \widehat{w}_N\left(\mathbf{w}^-_{j+\frac{1}{2}} - \frac{\lambda}{\widehat{w}_N}\left[\widehat{\mathbf{f}}\left(\mathbf{w}^-_{j+\frac{1}{2}}, \mathbf{w}^+_{j+\frac{1}{2}}\right) - \widehat{\mathbf{f}}\left(\mathbf{w}^+_{j-\frac{1}{2}}, \mathbf{w}^-_{j+\frac{1}{2}}\right)\right]\right)\\
&\quad + \widehat{w}_1\left(\mathbf{w}^+_{j-\frac{1}{2}} - \frac{\lambda}{\widehat{w}_1}\left[\widehat{\mathbf{f}}\left(\mathbf{w}^+_{j-\frac{1}{2}}, \mathbf{w}^-_{j+\frac{1}{2}}\right) - \widehat{\mathbf{f}}\left(\mathbf{w}^-_{j-\frac{1}{2}}, \mathbf{w}^+_{j-\frac{1}{2}}\right)\right]\right)\\
&= \sum_{\alpha=2}^{N-1} \widehat{w}_\alpha \mathbf{q}_j(\widehat{x}_j^\alpha) + \widehat{w}_N \mathbf{H}_N + \widehat{w}_1 \mathbf{H}_1,
\end{aligned}$$

where

$$\mathbf{H}_1 = \mathbf{w}^+_{j-\frac{1}{2}} - \frac{\lambda}{\widehat{w}_1}\left[\widehat{\mathbf{f}}\left(\mathbf{w}^+_{j-\frac{1}{2}}, \mathbf{w}^-_{j+\frac{1}{2}}\right) - \widehat{\mathbf{f}}\left(\mathbf{w}^-_{j-\frac{1}{2}}, \mathbf{w}^+_{j-\frac{1}{2}}\right)\right] \tag{2.6}$$

$$\mathbf{H}_N = \mathbf{w}^-_{j+\frac{1}{2}} - \frac{\lambda}{\widehat{w}_N}\left[\widehat{\mathbf{f}}\left(\mathbf{w}^-_{j+\frac{1}{2}}, \mathbf{w}^+_{j+\frac{1}{2}}\right) - \widehat{\mathbf{f}}\left(\mathbf{w}^+_{j-\frac{1}{2}}, \mathbf{w}^-_{j+\frac{1}{2}}\right)\right]. \tag{2.7}$$

Notice that (2.6) and (2.7) are both of the type (2.1), and $\widehat{w}_1 = \widehat{w}_N$, therefore $\mathbf{H}_1$ and $\mathbf{H}_N$ satisfy the minimum entropy principle under the CFL condition (2.5). Since $\overline{\mathbf{w}}_j^{n+1}$ is a convex combination of $\mathbf{H}_1, \mathbf{H}_N$ and $\mathbf{q}_j(\widehat{x}_j^\alpha)$, by Lemma 2.1, we get the minimum entropy principle for $\overline{\mathbf{w}}_j^{n+1}$. □

The high order SSP time discretizations will keep the validity of Theorem 2.3 since they are convex combinations of forward Euler.

Theorem 2.3 implies that, to have the minimum principle for $\overline{\mathbf{w}}_j^{n+1} \in G$, we need to enforce $\mathbf{q}_j(\widehat{x}_j^\alpha) \in G$. The positivity of density and pressure of $\mathbf{q}_j(\widehat{x}_j^\alpha) \in G$ was discussed in [14]. Thus here we only show how to enforce the entropy part.

At time level n, given $\overline{\mathbf{w}}_j^n \in G$, assume $\mathbf{q}_j(\widehat{x}_j^\alpha)$ $(\alpha = 1, \ldots, N)$ have positive density and pressure. Define $\partial G = \{\mathbf{w} : \rho, p > 0, S = S_0 = \min_x S(\mathbf{w}_0(x))\}$, and

$$\mathbf{L}(t) = (1-t)\overline{\mathbf{w}}_j^n + t\mathbf{q}_j(x), \quad 0 \le t \le 1. \tag{2.8}$$

∂G is a surface and $\mathbf{L}(t)$ is the line segment connecting the two points $\overline{\mathbf{w}}_j^n$ and $\mathbf{q}_j(x)$, where t is a parameter. If $S(\mathbf{q}_j(x)) < S_0$, then the line segment $\mathbf{L}(t)$ ($t \in [0,1]$) intersects with the surface ∂G at one and only one point since G is a convex set. If $S(\mathbf{q}_j(x)) < S_0$, let $t(x)$ denote the parameter in (2.8) corresponding to the intersection point; otherwise let $t(x) = 1$. In practice, we can find $t(x)$ by using Newton iteration to solve $S(\mathbf{L}(t(x))) = S_0$. Now we define

$$\widetilde{\mathbf{q}}_j(x) = \theta_j(\mathbf{q}_j(x) - \overline{\mathbf{w}}_j^n) + \overline{\mathbf{w}}_j^n, \quad \theta_j = \min_{x \in \{\widehat{x}_j^1, \ldots, \widehat{x}_j^N\}} t(x). \tag{2.9}$$

The limiter (2.9) should be used for each stage in a SSP Runge–Kutta method or each step in a SSP multi-step method. It is easy to check that the cell average of $\widetilde{\mathbf{q}}_j(x)$ over I_j is $\overline{\mathbf{w}}_j^n$.

Lemma 2.4 *$\widetilde{\mathbf{q}}_j(x)$ defined in (2.9) satisfies $\widetilde{\mathbf{q}}_j(\widehat{x}_j^\alpha) \in G$ for all α.*

Proof First notice that $\widetilde{\mathbf{q}}_j(x)$ is a convex combination of $\mathbf{q}_j(x)$ and $\overline{\mathbf{w}}_j^n$, thus $\widetilde{\mathbf{q}}_j(\widehat{x}_j^\alpha)$ still have positive density and pressure since pressure is a concave function. We only need to prove $S(\widetilde{\mathbf{q}}_j(\widehat{x}_j^\alpha)) \geq S_0$ for the case that $S(\mathbf{q}_j(\widehat{x}_j^\alpha)) < S_0$.

If $S(\mathbf{q}_j(\widehat{x}_j^\alpha)) < S_0$, then $S(\mathbf{L}(t(\widehat{x}_j^\alpha))) = S_0$ and

$$\begin{aligned}
\widetilde{\mathbf{q}}_j(\widehat{x}_j^\alpha) &= \theta_j(\widetilde{\mathbf{q}}_j(\widehat{x}_j^\alpha) - \overline{\mathbf{w}}_j^n) + \overline{\mathbf{w}}_j^n \\
&= \frac{\theta_j}{t(\widehat{x}_j^\alpha)}[t(\widehat{x}_j^\alpha)(\widetilde{\mathbf{q}}_j(\widehat{x}_j^\alpha) - \overline{\mathbf{w}}_j^n) + \overline{\mathbf{w}}_j^n] + \left(1 - \frac{\theta_j}{t(\widehat{x}_j^\alpha)}\right)\overline{\mathbf{w}}_j^n \\
&= \frac{\theta_j}{t(\widehat{x}_j^\alpha)}\mathbf{L}(t(\widehat{x}_j^\alpha)) + \left(1 - \frac{\theta_j}{t(\widehat{x}_j^\alpha)}\right)\overline{\mathbf{w}}_j^n,
\end{aligned}$$

So $\widetilde{\mathbf{q}}_j(\widehat{x}_j^\alpha)$ is a convex combination of $\mathbf{L}(t(\widehat{x}_j^\alpha))$ and $\overline{\mathbf{w}}_j^n$, thus $S(\widetilde{\mathbf{q}}_j(\widehat{x}_j^\alpha)) \geq S_0$. □

We refer to a generic smooth solution as a smooth solution $\mathbf{v}(x, t^n) \in G$ satisfying $\min S(\mathbf{v}(x, t^n)) = S_0$ and the second order derivative of $S(\mathbf{v}(x, t^n))$ with respect to x does not vanish at the global minimum. For such generic smooth solutions, the limiter (2.9) does not affect the high order accuracy of the original scheme. Assume $\mathbf{q}_j(x)$ is a $(k+1)$-th order accurate approximation $\mathbf{q}_j(x) - \mathbf{v}(x, t^n) = O(\Delta x^{k+1})$. Without loss of generality, assume $\theta_j = t(\widehat{x}_j^\beta)$ for some β. Since $\mathbf{q}_j(\widehat{x}_j^\beta)$, $\mathbf{L}(t(\widehat{x}_j^\beta))$ and $\overline{\mathbf{w}}_j^n$ lie on the same line, we have $\theta_j - 1 = -\frac{\|\mathbf{q}_j(\widehat{x}_j^\beta) - \mathbf{L}(t(\widehat{x}_j^\beta))\|}{\|\mathbf{q}_j(\widehat{x}_j^\beta) - \overline{\mathbf{w}}_j^n\|}$. Thus,

$$\begin{aligned}
\widetilde{\mathbf{q}}_j(x) - \mathbf{q}_j(x) &= \theta_j(\mathbf{q}_j(x) - \overline{\mathbf{w}}_j^n) + \overline{\mathbf{w}}_j^n - \mathbf{q}_j(x) \\
&= (\theta_j - 1)(\mathbf{q}_j(x) - \overline{\mathbf{w}}_j^n) \\
&= -\frac{\| \mathbf{q}_j(\widehat{x}_j^\beta) - \mathbf{L}(t(\widehat{x}_j^\beta)) \|}{\| \mathbf{q}_j(\widehat{x}_j^\beta) - \overline{\mathbf{w}}_j^n \|}(\mathbf{q}_j(x) - \overline{\mathbf{w}}_j^n).
\end{aligned}$$

Define $d(\mathbf{z}, \partial G) = \min_{\mathbf{w}\in\partial G} \parallel \mathbf{z} - \mathbf{w} \parallel$. Since S_0 is the minimum of $S(\mathbf{v}(x, t^n))$, there is at least one $\widehat{x}_j^\alpha$ such that $S_0 - S(\mathbf{v}(\widehat{x}_j^\alpha, t^n)) \geq C\Delta x^2$ where C is a nonzero constant depending on the derivatives of $S(\mathbf{v}(x, t^n))$. This implies $d(\mathbf{v}(\widehat{x}_j^\alpha, t^n), \partial G) \geq O(\Delta x^2)$, thus $d(\mathbf{q}_j(\widehat{x}_j^\alpha), \partial G) \geq O(\Delta x^2)$. $\overline{\mathbf{w}}_j^n$ is the cell average of $\mathbf{q}_j(x)$ implies $\overline{\mathbf{w}}_j^n = \sum_{\alpha=1}^N \widehat{w}_\alpha \mathbf{q}_j(\widehat{x}_j^\alpha)$. So $d(\overline{\mathbf{w}}_j^n, \partial G) \geq O(\Delta x^2)$.

On the other hand, $\theta_j = t(\widehat{x}_j^\beta)$ implies $S(\mathbf{q}_j(\widehat{x}_j^\beta)) \notin G$, so $\parallel \mathbf{q}_j(\widehat{x}_j^\beta) - \overline{\mathbf{w}}_j^n \parallel > d(\overline{\mathbf{w}}_j^n, \partial G) \geq O(\Delta x^2)$. The overshoot is small $\parallel \mathbf{q}_j(\widehat{x}_j^\beta) - \mathbf{L}(t(\widehat{x}_j^\beta)) \parallel = O(\Delta x^{k+1})$ since $\mathbf{q}_j(x)$ is an accurate approximation to $\mathbf{v}(x, t^n) \in G$.

Finally, notice that $\parallel \mathbf{q}_j(x) - \overline{\mathbf{w}}_j^n \parallel = O(\Delta x)$, we get that $\parallel \widetilde{\mathbf{q}}_j(x) - \mathbf{q}_j(x) \parallel = O(\Delta x^k), \forall x \in I_j$.

We remark that for the non-generic situation that the second derivative of $S(\mathbf{v}(x, t^n))$ with respect to x does vanish at the global minimum, it seems difficult to design a conservative limiter which can be proved not to destroy accuracy. On the other hand, the fact that our limiter is easy to implement also for multi-dimensional cases (see next section) and that it maintains high order accuracy for generic smooth solutions makes it a good technique to adopt for high order schemes.

3 The 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 with initial data $\mathbf{w}_0(x, y)$

$$\mathbf{w}_t + \mathbf{f}(\mathbf{w})_x + \mathbf{g}(\mathbf{w})_y = 0, \quad t \geq 0, (x, y) \in \mathbb{R}^2, \tag{3.1}$$

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

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 specific entropy $S = \ln \frac{p}{\rho^\gamma}$ is quasi-concave with respect to $\mathbf{w}$ if $\rho > 0$ and the set of admissible states $G = \{\mathbf{w} | \rho > 0, p > 0, S \geq S_0 = \min_{x,y} S(\mathbf{w}_0(x, y))\}$ is still convex.

3.1 Rectangular meshes

For simplicity we assume we have a uniform rectangular mesh. At time level n, we have an approximation polynomial $\mathbf{q}_{ij}(x, y)$ of degree k with the cell average $\overline{\mathbf{w}}_{ij}^n$ 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 $\mathbf{w}^+_{i-\frac{1}{2},j}(y)$, $\mathbf{w}^-_{i+\frac{1}{2},j}(y)$, $\mathbf{w}^+_{i,j-\frac{1}{2}}(x)$, $\mathbf{w}^-_{i,j+\frac{1}{2}}(x)$ denote the traces of $\mathbf{q}_{ij}(x, y)$ 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

$$\overline{\mathbf{w}}_{ij}^{n+1} = \overline{\mathbf{w}}_{ij}^{n} - \frac{\Delta t}{\Delta x \Delta y} \int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}} \widehat{\mathbf{f}}\left[\mathbf{w}^{-}_{i+\frac{1}{2},j}(y), \mathbf{w}^{+}_{i+\frac{1}{2},j}(y)\right] - \widehat{\mathbf{f}}\left[\mathbf{w}^{-}_{i-\frac{1}{2},j}(y), \mathbf{w}^{+}_{i-\frac{1}{2},j}(y)\right] dy$$

$$- \frac{\Delta t}{\Delta x \Delta y} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \widehat{\mathbf{g}}\left[\mathbf{w}^{-}_{i,j+\frac{1}{2}}(x), \mathbf{w}^{+}_{i,j+\frac{1}{2}}(x)\right] - \widehat{\mathbf{g}}\left[\mathbf{w}^{-}_{i,j-\frac{1}{2}}(x), \mathbf{w}^{+}_{i,j-\frac{1}{2}}(x)\right] dx,$$

where $\widehat{\mathbf{f}}(\cdot,\cdot)$, $\widehat{\mathbf{g}}(\cdot,\cdot)$ are one dimensional 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 $S_i^x = \{x_i^\beta : \beta = 1, \ldots, L\}$ denote the Gauss quadrature points on $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, and $S_j^y = \{y_j^\beta : \beta = 1, \ldots, L\}$ denote the Gauss quadrature points on $[y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$. For instance, $(x_{i-\frac{1}{2}}, y_j^\beta)$ $(\beta = 1, \ldots, L)$ 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, $\mathbf{w}^{+}_{i-\frac{1}{2},\beta} = \mathbf{w}^{+}_{i-\frac{1}{2},j}(y_j^\beta)$. 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., $\widehat{S}_i^x = \{\widehat{x}_i^\alpha : \alpha = 1, \ldots, N\}$ will denote the Gauss–Lobatto quadrature points on $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$, and $\widehat{S}_j^y = \{\widehat{y}_j^\alpha : \alpha = 1, \ldots, N\}$ 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

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

$$- \lambda_2 \sum_{\beta=1}^{L} w_\beta \left[\widehat{\mathbf{g}}\left(\mathbf{w}^{-}_{\beta,j+\frac{1}{2}}, \mathbf{w}^{+}_{\beta,j+\frac{1}{2}}\right) - \widehat{\mathbf{g}}\left(\mathbf{w}^{-}_{\beta,j-\frac{1}{2}}, \mathbf{w}^{+}_{\beta,j-\frac{1}{2}}\right)\right]. \quad (3.2)$$

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

$$S_{ij} = (S_i^x \otimes \widehat{S}_j^y) \cup (\widehat{S}_i^x \otimes S_j^y). \quad (3.3)$$

For simplicity, let $\mu_1 = \frac{\lambda_1 a_1}{\lambda_1 a_1 + \lambda_2 a_2}$ and $\mu_2 = \frac{\lambda_2 a_2}{\lambda_1 a_1 + \lambda_2 a_2}$ where $a_1 = \| (|u| + c) \|_\infty$, $a_2 = \| (|v| + c) \|_\infty$. Notice that $\widehat{w}_1 = \widehat{w}_N$, we have

$$\begin{aligned}\overline{\mathbf{w}}_{ij}^{n} &= \frac{\mu_1}{\Delta x\Delta y}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}\mathbf{q}_{ij}(x,y)\,dydx+\frac{\mu_2}{\Delta x\Delta y}\int_{y_{j-\frac{1}{2}}}^{y_{j+\frac{1}{2}}}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}\mathbf{q}_{ij}(x,y)\,dxdy\\ &=\mu_1\sum_{\beta=1}^{L}\sum_{\alpha=1}^{N}w_\beta\widehat{w}_\alpha\mathbf{q}_{ij}(\widehat{x}_i^{\alpha},y_j^{\beta})+\mu_2\sum_{\beta=1}^{L}\sum_{\alpha=1}^{N}w_\beta\widehat{w}_\alpha\mathbf{q}_{ij}(x_i^{\beta},\widehat{y}_j^{\alpha})\\ &=\sum_{\beta=1}^{L}\sum_{\alpha=2}^{N-1}w_\beta\widehat{w}_\alpha\left[\mu_1\mathbf{q}_{ij}(\widehat{x}_i^{\alpha},y_j^{\beta})+\mu_2\mathbf{q}_{ij}(x_i^{\beta},\widehat{y}_j^{\alpha})\right]\\ &\quad+\sum_{\beta=1}^{L}w_\beta\widehat{w}_1\left[\mu_1\mathbf{w}^{-}_{i+\frac{1}{2},\beta}+\mu_1\mathbf{w}^{+}_{i-\frac{1}{2},\beta}+\mu_2\mathbf{w}^{-}_{\beta,j+\frac{1}{2}}+\mu_2\mathbf{w}^{+}_{\beta,j-\frac{1}{2}}\right]\end{aligned}\tag{3.4}$$

Theorem 3.1 *Consider a two-dimensional finite volume scheme or the scheme satisfied by the cell averages of the DG method on rectangular meshes (3.2), associated with the approximation polynomials $\mathbf{q}_{ij}(x,y)$ of degree k (either reconstruction or DG polynomials). If $\mathbf{w}^{\pm}_{\beta,j\pm\frac{1}{2}},\mathbf{w}^{\pm}_{i\pm\frac{1}{2},\beta}\in G$ and $\mathbf{q}_{ij}(x,y)\in G$ (for any $(x,y)\in S_{ij}$), then $\overline{\mathbf{w}}_j^{n+1}\in G$ under the CFL condition*

$$\lambda_1 a_1+\lambda_2 a_2\le\widehat{w}_1\alpha_0.\tag{3.5}$$

Proof Plugging (3.4) in, (3.2) can be written as

$$\begin{aligned}\overline{\mathbf{w}}_{ij}^{n+1} &=\sum_{\beta=1}^{L}\sum_{\alpha=2}^{N-1}w_\beta\widehat{w}_\alpha\left[\mu_1\mathbf{q}_{ij}(\widehat{x}_i^{\alpha},y_j^{\beta})+\mu_2\mathbf{q}_{ij}(x_i^{\beta},\widehat{y}_j^{\alpha})\right]\\ &\quad+\mu_1\sum_{\beta=1}^{L}w_\beta\widehat{w}_1\left[\mathbf{w}^{-}_{i+\frac{1}{2},\beta}-\frac{\lambda_1}{\mu_1\widehat{w}_1}\left(\widehat{\mathbf{f}}(\mathbf{w}^{-}_{i+\frac{1}{2},\beta},\mathbf{w}^{+}_{i+\frac{1}{2},\beta})-\widehat{\mathbf{f}}(\mathbf{w}^{+}_{i-\frac{1}{2},\beta},\mathbf{w}^{-}_{i+\frac{1}{2},\beta})\right)\right.\\ &\quad\left.+\mathbf{w}^{+}_{i-\frac{1}{2},\beta}-\frac{\lambda_1}{\mu_1\widehat{w}_1}\left(\widehat{\mathbf{f}}(\mathbf{w}^{+}_{i-\frac{1}{2},\beta},\mathbf{w}^{-}_{i+\frac{1}{2},\beta})-\widehat{\mathbf{f}}(\mathbf{w}^{-}_{i-\frac{1}{2},\beta},\mathbf{w}^{+}_{i-\frac{1}{2},\beta})\right)\right]\\ &\quad+\mu_2\sum_{\beta=1}^{L}w_\beta\widehat{w}_N\left[\mathbf{w}^{-}_{\beta,j+\frac{1}{2}}-\frac{\lambda_2}{\mu_2\widehat{w}_1}\left(\widehat{\mathbf{g}}(\mathbf{w}^{-}_{\beta,j+\frac{1}{2}},\mathbf{w}^{+}_{\beta,j+\frac{1}{2}})-\widehat{\mathbf{g}}(\mathbf{w}^{+}_{\beta,j-\frac{1}{2}},\mathbf{w}^{-}_{\beta,j+\frac{1}{2}})\right)\right.\\ &\quad\left.+\mathbf{w}^{+}_{\beta,j-\frac{1}{2}}-\frac{\lambda_2}{\mu_2\widehat{w}_1}\left(\widehat{\mathbf{g}}(\mathbf{w}^{+}_{\beta,j-\frac{1}{2}},\mathbf{w}^{-}_{\beta,j+\frac{1}{2}})-\widehat{\mathbf{g}}(\mathbf{w}^{-}_{\beta,j-\frac{1}{2}},\mathbf{w}^{+}_{\beta,j-\frac{1}{2}})\right)\right]\end{aligned}$$

Following the same arguments as in Theorem 2.3, we conclude $\overline{\mathbf{w}}_j^{n+1}\in G$. □

The limiter in the previous section can be extended easily to two-dimensional cases. At time level n, given $\overline{\mathbf{w}}_{ij}^{n}\in G$, do the following modification

$$\widetilde{\mathbf{q}}_{ij}(x,y)=\theta_{ij}(\mathbf{q}_{ij}(x,y)-\overline{\mathbf{w}}_{ij}^{n})+\overline{\mathbf{w}}_{ij}^{n},\quad\theta_{ij}=\min_{(x,y)\in S_{ij}}t(x,y),\tag{3.6}$$

where $t(x, y)$ is the parameter corresponding to the intersection point of the surface ∂G and the line segment $\mathbf{L}(t) = (1-t)\overline{\mathbf{w}}_{ij}^n + t\mathbf{q}_{ij}(x, y)$ if $\mathbf{q}_{ij}(x, y) \notin G$; $t(x, y) = 1$ otherwise.

3.2 Triangular meshes

For simplicity, we only discuss DG schemes in this subsection. All the conclusions will also hold for a high order finite volume scheme.

For each triangle K we denote by l_K^i $(i = 1, 2, 3)$ the length of its three edges e_K^i $(i = 1, 2, 3)$, with outward unit normal vector ν^i $(i = 1, 2, 3)$. $K(i)$ denotes the neighboring triangle along e_K^i and $|K|$ is the area of the triangle K. Let $\widehat{\mathbf{F}}(\mathbf{w}, \mathbf{v}, \nu)$ be a one dimensional monotone flux in the ν direction satisfying $\widehat{\mathbf{F}}(\mathbf{w}, \mathbf{v}, \nu) = -\widehat{\mathbf{F}}(\mathbf{v}, \mathbf{w}, -\nu)$ (conservation), and $\widehat{\mathbf{F}}(\mathbf{w}, \mathbf{w}, \nu) = \mathbf{F}(\mathbf{w}) \cdot \nu$ (consistency), with $\mathbf{F}(\mathbf{w}) = \langle \mathbf{f}(\mathbf{w}), \mathbf{g}(\mathbf{w}) \rangle$. For example, the Lax-Friedrichs flux is defined as

$$\widehat{\mathbf{F}}(\mathbf{w}, \mathbf{v}, \nu) = \frac{1}{2}(\mathbf{F}(\mathbf{w}) \cdot \nu + \mathbf{F}(\mathbf{v}) \cdot \nu - a(\mathbf{v} - \mathbf{w})), \quad a = \| \, |\langle u, v \rangle| + c \, \|_\infty .$$

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

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

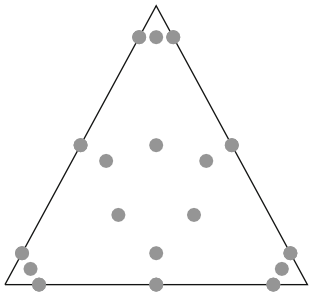
where $\overline{\mathbf{w}}_K^n$ is the cell average over K of the numerical solution, and $\mathbf{w}_i^{int(K)}$, $\mathbf{w}_i^{ext(K)}$ are the approximations to the values on the edge e_K^i obtained from the interior and the exterior of K. Assume the DG polynomial on the triangle K is $\mathbf{q}_K(x,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

$$\overline{\mathbf{w}}_K^{n+1} = \overline{\mathbf{w}}_K^n - \frac{\Delta t}{|K|} \sum_{i=1}^{3} \sum_{\beta=1}^{k+1} \widehat{\mathbf{F}}(\mathbf{w}_{i,\beta}^{int(K)}, \mathbf{w}_{i,\beta}^{ext(K)}, \nu^i) w_\beta l_K^i, \tag{3.7}$$

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

We need the quadrature rule introduced in [15] for $\mathbf{q}_K(x, y)$ on K. 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

Fig. 1 Points in (3.8) for $k = 2$



$$S_K^k = \left\{ \left(\frac{1}{2} + v^\beta, \left(\frac{1}{2} + \widehat{u}^\alpha\right)\left(\frac{1}{2} - v^\beta\right), \left(\frac{1}{2} - \widehat{u}^\alpha\right)\left(\frac{1}{2} - v^\beta\right)\right), \right.$$
$$\left(\left(\frac{1}{2} - \widehat{u}^\alpha\right)\left(\frac{1}{2} - v^\beta\right), \frac{1}{2} + v^\beta, \left(\frac{1}{2} + \widehat{u}^\alpha\right)\left(\frac{1}{2} - v^\beta\right)\right), \tag{3.8}$$
$$\left.\left(\left(\frac{1}{2} + \widehat{u}^\alpha\right)\left(\frac{1}{2} - v^\beta\right), \left(\frac{1}{2} - \widehat{u}^\alpha\right)\left(\frac{1}{2} - v^\beta\right), \frac{1}{2} + v^\beta\right)\right\}$$

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

Following Theorem 5.1 in [15] and the same arguments as in Theorem 2.3, we have

Theorem 3.2 *For the scheme* (3.7) *with the polynomial* $\mathbf{q}_K(x, y)$ *of degree* k, *if* $\mathbf{w}_{i,\beta}^{ext(K)} \in G$ *and* $\mathbf{q}_K(x, y) \in G, \forall (x, y) \in S_K^k$ *where* S_K^k *is defined in* (3.8), *then* $\overline{\mathbf{w}}_K^{n+1} \in G$ *under the CFL condition a* $\frac{\Delta t}{|K|} \sum_{i=1}^3 l_K^i \leq \frac{2}{3}\widehat{w}_1\alpha_0$.

The the same limiter can be used to enforce the sufficient condition. At time level n, given $\overline{\mathbf{w}}_K^n \in G$, do the following modification

$$\widetilde{\mathbf{q}}_K(x, y) = \theta_K\left(\mathbf{q}_K(x, y) - \overline{\mathbf{w}}_K^n\right) + \overline{\mathbf{w}}_K^n, \quad \theta_K = \min_{(x,y)\in S_K} t(x, y), \tag{3.9}$$

where $t(x, y)$ is the parameter corresponding to the intersection point of the surface ∂G and the line segment $\mathbf{L}(t) = (1-t)\overline{\mathbf{w}}_K^n + t\mathbf{q}_K(x, y)$ if $\mathbf{q}_K(x, y) \notin G$; $t(x, y) = 1$ otherwise.

4 Numerical tests

In this section, $\gamma = 1.4$ for all the examples.

Example 4.1 Accuracy tests.

We first test the accuracy of the entropy limiter (2.9). The initial condition is $\rho_0(x, y) = 1 + \frac{1}{2}\sin(2\pi x)$, $u_0(x) = 1$, $p_0(x) = 1$. The domain is [0, 1] and the boundary condition is periodic. The exact solution is $\rho(x, y, t) = 1 + \frac{1}{2}\sin(2\pi(x - t))$, $u(x, t) =$

1, $p(x,t) = 1$. The numerical schemes are the third order and fourth order DG schemes with Lax–Friedrichs flux [2] and the third order SSP time discretizations with the entropy limiter (2.9) used at each time stage or each time step.

The third order SSP Runge–Kutta method in [9] (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 [8] (with the CFL coefficient $c = \frac{1}{3}$) is

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

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

For $k = 2$ and $k = 3$, $\widehat{w}_1 = \frac{1}{6}$ and $\alpha_0 = \frac{1}{2}$, thus the time step (2.5) for P^2 DG with Runge–Kutta is taken as $\Delta t = \frac{1}{12}\frac{\Delta x}{\|(|u|+c)\|_\infty}$. Since the CFL coefficient $c = \frac{1}{3}$ for the third order SSP multi-step method, the time step is taken as $\Delta t = \frac{1}{36}\frac{\Delta x}{\|(|u|+c)\|_\infty}$. For the fourth order scheme, in order to make the error from spatial discretizations dominant, we replace Δx by $\Delta x^{\frac{4}{3}}$.

The accuracy result is listed in Table 1. We observe that, for Runge–Kutta, the accuracy degenerates when the mesh is fine enough. This is due to the lower order accuracy in the intermediate stages of the Runge–Kutta method. In particular, recall that the limiter (2.9) does not destroy accuracy for generic smooth solutions only if the polynomial $\mathbf{q}_j(x)$ is a $(k+1)$th accurate approximation to the exact solution. The DG polynomials $\mathbf{q}_j(x)$ in the intermediate stages of a Runge–Kutta methods are in general not $(k+1)$th order accurate, therefore the limiter (2.9) may kill the accuracy when it is imposed in the intermediate stages. The same phenomenon exists for the high order maximum-principle-satisfying schemes, see [13]. A similar phenomenon of the Runge–Kutta method in the context of boundary conditions was pointed out in [1].

The full accuracy order is observed for the multi-step time discretization, which justifies that the limiter itself does not kill accuracy. Since accuracy degeneracy is usually only observed on very fine meshes for Runge–Kutta methods, in applications it is often acceptable to use the Runge–Kutta methods, similar to the conclusions in [1,13].

Example 4.2 The Lax shock tube problem.

For high order DG schemes solving the compressible Euler equations, even though the characteristicwise TVB limiter in [2] can kill oscillations, it is not sufficient to

Table 1 Third order SSP time discretizations and high order DG spatial discretizations with the entropy limiter (2.9), $\Delta x = \frac{1}{N}$, $t = 0.1$

	N	SSP Runge–Kutta				SSP multi-step			
		L^1 error	Order	L^∞ error	Order	L^1 error	Order	L^∞ error	Order
P^2DG		$\Delta t = \frac{1}{12}\frac{\Delta x}{\|(\|u\|+c)\|_\infty}$				$\Delta t = \frac{1}{36}\frac{\Delta x}{\|(\|u\|+c)\|_\infty}$			
	8	1.49E−3	–	5.23E−3	–	1.58E−3	–	5.79E−3	–
	16	1.64E−4	3.17	8.46E−4	2.62	1.65E−4	3.25	8.77E−4	2.72
	32	2.07E−5	2.98	9.07E−5	3.22	2.06E−5	3.00	9.07E−5	3.27
	64	2.62E−6	2.97	1.17E−5	2.95	2.60E−6	2.98	1.17E−5	2.95
	128	3.30E−7	2.98	1.47E−6	2.98	3.25E−7	2.99	1.47E−6	2.98
	256	4.16E−8	2.99	1.84E−7	2.99	4.07E−8	2.99	1.84E−7	2.99
	512	5.23E−9	2.99	3.51E−8	2.39	5.09E−9	3.00	2.31E−8	3.00
	1,024	6.60E−10	2.98	1.02E−8	1.78	6.36E−10	3.00	2.88E−9	3.00
P^3DG		$\Delta t = \frac{1}{12}\frac{\Delta x^{\frac{4}{3}}}{\|(\|u\|+c)\|_\infty}$				$\Delta t = \frac{1}{36}\frac{\Delta x^{\frac{4}{3}}}{\|(\|u\|+c)\|_\infty}$			
	4	5.11E−4	–	4.52E−3	–	5.19E−4	–	2.09E−3	–
	8	2.45E−5	4.38	1.05E−4	4.31	2.46E−5	4.39	1.05E−4	4.31
	16	1.40E−6	4.12	5.82E−6	4.18	1.40E−6	4.13	5.34E−6	4.30
	32	9.02E−8	3.96	5.61E−7	3.38	8.41E−8	4.06	3.74E−7	3.83
	64	6.66E−9	3.75	1.24E−7	2.16	5.21E−9	4.01	2.23E−8	4.06
	128	5.29E−10	3.65	1.83E−8	2.77	3.24E−10	4.00	1.42E−9	3.97
	256	4.37E−11	3.59	3.39E−9	2.43	2.02E−11	4.00	9.00E−11	3.98
	512	4.14E−12	3.39	5.61E−10	2.59	1.26E−12	3.99	5.55E−12	4.01

stabilize the scheme for problems with low densities or low pressures. In [14], it was reported that high order RKDG scheme with both the positivity-preserving limiter and the TVB limiter worked fine for a lot of demanding problems. Recent study reveals that the third order RKDG scheme with only the positivity-preserving limiter is stable even for strong shocks, see [12], which is not surprising since a conservative positivity-preserving scheme is L^1 stable [16]. However, the positivity-preserving limiter alone can not kill oscillations for high order DG schemes, and the oscillations are much more prominent in the fourth order DG scheme than in the third order scheme. See Fig. 2 for the result of the fourth order DG scheme with P^3 element for the Lax shock tube problem. We can see that the result with only the positivity-preserving limiter is oscillatory and the result with the positivity-preserving limiter and (2.9) has much less oscillations. In other words, enforcing the minimum entropy principle will damp the oscillations in high order schemes, which was pointed out in [6].

Example 4.3 Double rarefactions with low densities and low pressures.

Consider the following one-dimensional Riemann problem with initial data

$$\rho_L = \rho_R = p_L = p_R = 1, \quad u_R = -u_L = v_0 \geq 0.$$

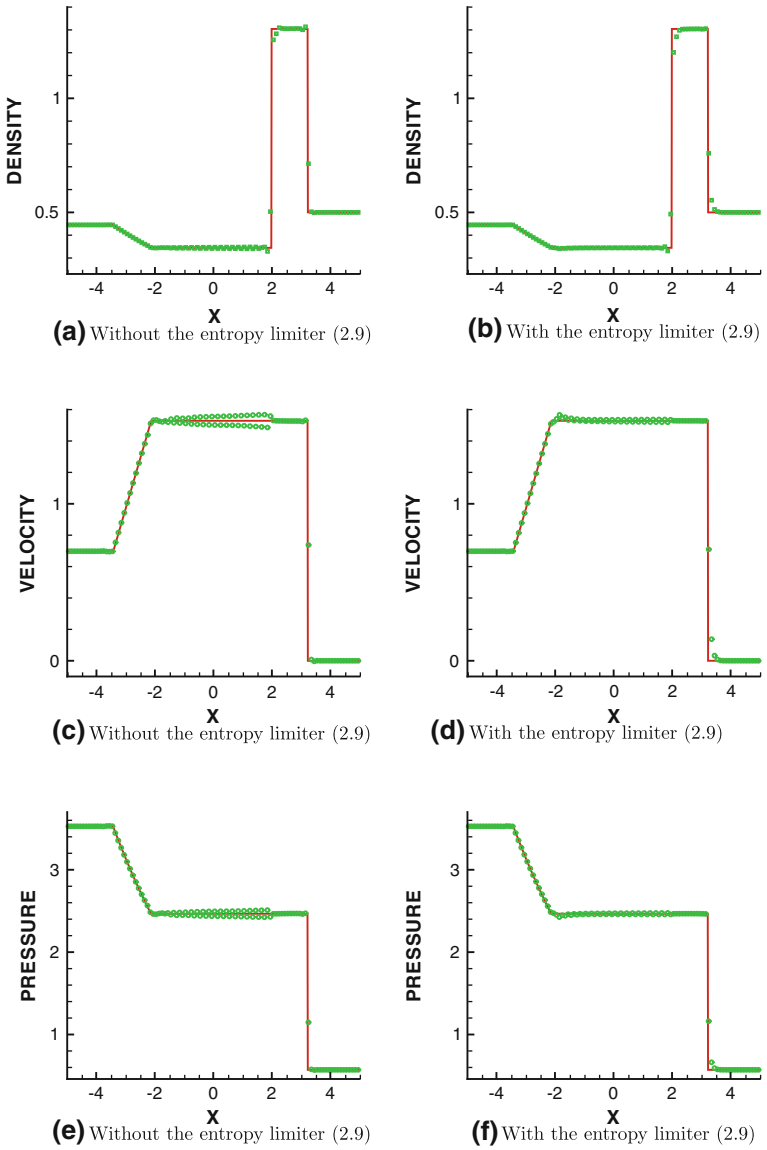


Fig. 2 Lax shock tube problem. P^3 element DG with the positivity-preserving limiter without any TVD or TVB limiter 100 cells. The *solid lines* are the exact solutions. The *symbols* are the numerical solutions

We first test the entropy limiter for the case $v_0 = 4$, in which the lowest pressure of the exact solution is around 0.0034. In Fig. 3, we can see that the DG method with only the positivity limiter works well and adding the entropy limiter does not destroy the good performance.

Then we test the entropy limiter for the case $v_0 = 12$, in which vacuum is present. In Fig. 4, we observe that the entropy limiter indeed damps the overshoots near $x = \pm 4$.

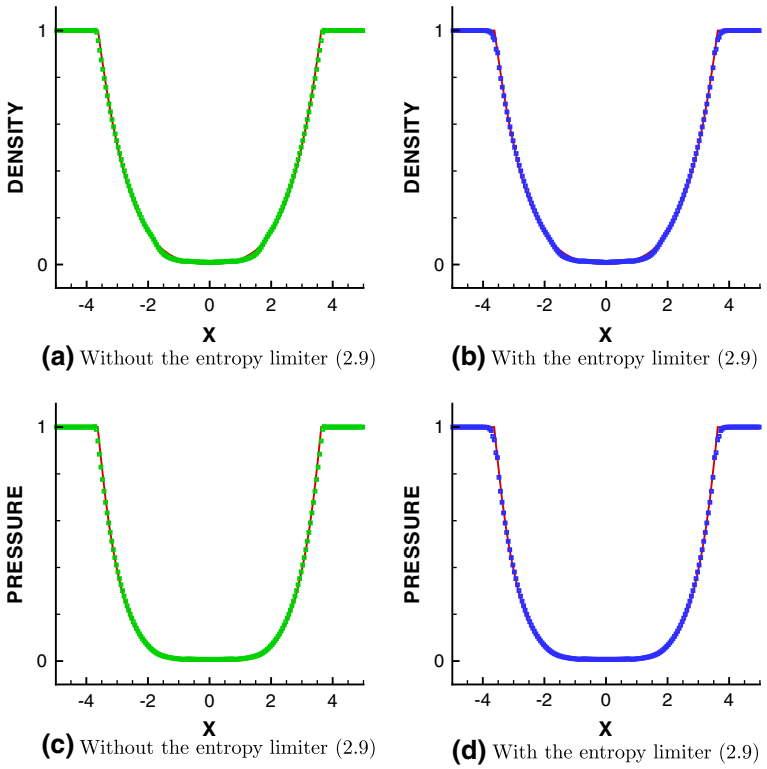


Fig. 3 Example 4.3 with $v_0 = 4$. P^3 element DG with the positivity-preserving limiter without any TVD or TVB limiter. The *solid lines* are the exact solutions. The *symbols* are the numerical solutions on 200 cells at time $T = 0.7$

For this particular problem, the exact solution is isentropic. We plot the history of minimum and maximum specific entropy values of the numerical solutions in Fig. 5 for the P^3 element DG scheme with the positivity-preserving limiter and the entropy limiter. We observe that the minimum stays the same, which means the entropy limiter does its job to keep the minimum principle. On the other hand, the maximum is not in control, which however does not affect the performance of the scheme as Figs. 3 and 4 show.

Example 4.4 The Shu–Osher problem.

We consider the problem of shock wave interacting with sine waves in density, proposed in [10]. Initially,

$$\begin{aligned} &\rho = 3.857143, \quad u = 2.629369, \quad p = 10.33333, \quad \text{if} \quad x < -4; \\ &\rho = 1 + 0.2 \sin 5x, \quad u = 0, \quad p = 1, \quad \text{if} \quad x \geq -4. \end{aligned}$$

See Fig. 6 for the good performance of the entropy limiter.

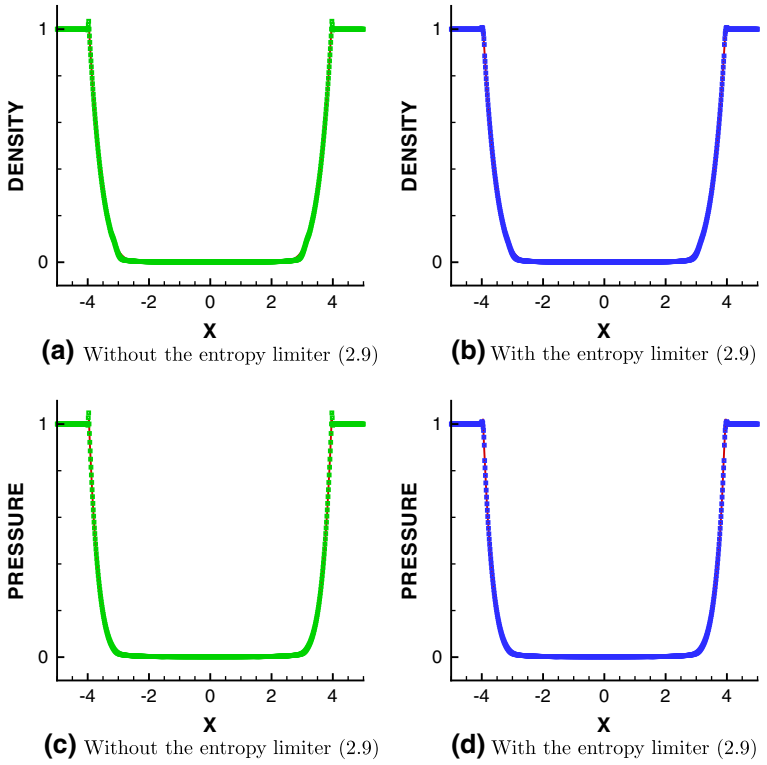


Fig. 4 Example 4.3 with $v_0 = 12$. P^3 element DG with the positivity-preserving limiter without any TVD or TVB limiter. The *solid lines* are the reference solutions obtained by the P^3 element DG with only the positivity-preserving limiter on 10,000 cells. The *symbols* are the numerical solutions on 800 cells at time $T = 0.3$

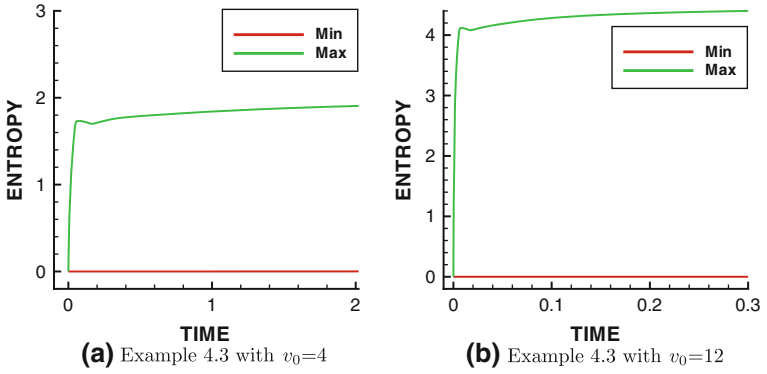
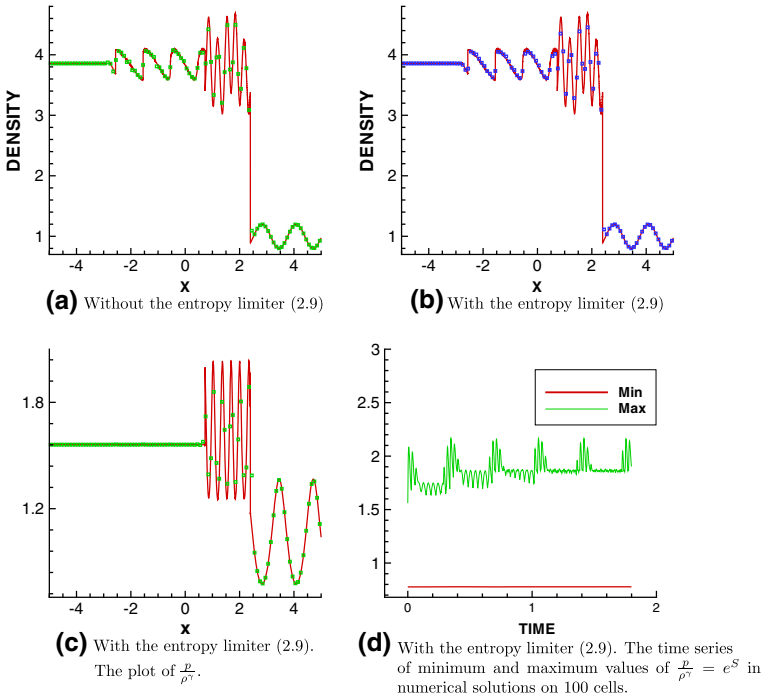


Fig. 5 The minimum and maximum values of specific entropy in the solutions of the P^3 element DG scheme with the positivity-preserving limiter and the entropy limiter



(a) Without the entropy limiter (2.9)

(b) With the entropy limiter (2.9)

(c) With the entropy limiter (2.9). The plot of $\frac{p}{\rho^\gamma}$.

(d) With the entropy limiter (2.9). The time series of minimum and maximum values of $\frac{p}{\rho^\gamma} = e^S$ in numerical solutions on 100 cells.

Fig. 6 The Shu–Osher problem. P^3 element DG with the positivity-preserving limiter without any TVD or TVB limiter. The *solid lines* are the reference solutions obtained by the P^3 element DG with only the positivity-preserving limiter on 10,000 cells. The *symbols* are the numerical solutions on 100 cells at time $T = 1.8$

5 Concluding remarks

In this paper, we have discussed the minimum entropy principle for high order schemes solving the compressible Euler equations in gas dynamics. An extension of positivity-preserving limiter in [14] can be used to enforce the minimum entropy principle. The generalizations to higher dimension are straightforward. Numerical tests imply that enforcing the minimum entropy principle may damp the oscillations in high order schemes.

Acknowledgments The authors would like to thank Eitan Tadmor for helpful discussions about the minimum entropy principle for gas dynamics equations.

References

1. Carpenter, M.H., Gottlieb, D., Abarbanel, S., Don, W.-S.: The theoretical accuracy of Runge–Kutta time discretizations for the initial boundary value problem: a study of the boundary error. SIAM J. Scientific Comput. **16**, 1241–1252 (1995)
2. Cockburn, B., Lin, S.-Y., Shu, C.-W.: TVB Runge–Kutta local projection discontinuous Galerkin finite element method for conservation laws III: one dimensional systems. J. Comput. Phys. **84**, 90–113 (1989)

3. Gottlieb, S., Ketcheson, D.I., Shu, C.-W.: High order strong stability preserving time discretizations. J. Scientific Comput. **38**, 251–289 (2009)
4. Gottlieb, S., Shu, C.-W., Tadmor, E.: Strong stability preserving high order time discretization methods. SIAM Rev. **43**, 89–112 (2001)
5. Harten, A.: On the symmetric form of systems of conservation laws with entropy. J. Comput. Phys. **49**, 151–164 (1983)
6. Khobalatte, B., Perthame, B.: Maximum principle on the entropy and second-order kinetic schemes. Math. Comput. **62**, 119–131 (1994)
7. Perthame, B., Shu, C.-W.: On positivity preserving finite volume schemes for Euler equations. Numerische Mathematik **73**, 119–130 (1996)
8. Shu, C.-W.: Total-variation-diminishing time discretizations. SIAM J. Scientific Stat. Comput. **9**, 1073–1084 (1988)
9. Shu, C.-W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys. **77**, 439–471 (1988)
10. Shu, C.-W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes, II. J. Comput. Phys. **83**, 32–78 (1989)
11. Tadmor, E.: A minimum entropy principle in the gas dynamics equations. Appl. Numer. Math. **2**, 211–219 (1986)
12. Wang, C., Zhang, X., Shu, C.-W., Ning, J.: Robust high order discontinuous Galerkin schemes for two-dimensional gaseous detonations. J. Comput. Phys. **231**, 653–665 (2012)
13. Zhang, X., Shu, C.-W.: On maximum-principle-satisfying high order schemes for scalar conservation laws. J. Comput. Phys. **229**, 3091–3120 (2010)
14. Zhang, X., Shu, C.-W.: On positivity preserving high order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes. J. Comput. Phys. **229**, 8918–8934 (2010)
15. Zhang, X., Xia, Y., Shu, C.-W.: Maximum-principle-satisfying and positivity-preserving high order discontinuous Galerkin schemes for conservation laws on triangular meshes. J. Scientific Comput. (in press)
16. Zhang, X., Shu, C.-W.: Maximum-principle-satisfying and positivity-preserving high order schemes for conservation laws: survey and new developments. Proc Royal Soc. A **467**, 2752–2776 (2011)