J Sci Comput (2016) 68:1172–1197
DOI 10.1007/s10915-016-0173-1

An Integrated Linear Reconstruction for Finite Volume Scheme on Unstructured Grids

Li Chen[1] · **Ruo Li**[2]

Received: 14 July 2015 / Revised: 28 December 2015 / Accepted: 22 January 2016 /
Published online: 11 February 2016

Abstract Linear reconstruction based on local cell-averaged values is the most commonly adopted technique to achieve a second-order accuracy when one uses the finite volume scheme on unstructured grids. For solutions with discontinuities appearing in such as conservation laws, a certain limiter has to be applied to the predicted gradient to prevent numerical oscillations. We propose in this paper a new formulation for linear reconstruction on unstructured grids, which integrates the prediction of the gradient and the limiter together. By solving on each cell a tiny linear programming problem without any parameters, the gradient is directly obtained which satisfies the monotonicity condition. It can be shown that the resulting numerical scheme with our new method fulfils a discrete maximum principle with fair relaxed geometric constraints on grids. Numerical results demonstrate that our method achieves satisfactory numerical accuracy with theoretical guarantee of local discrete maximum principle.

Keywords Linear reconstruction · Finite volume method · Simplex method · Linear programming

1 Introduction

When one adopts the finite volume scheme to discretize one-dimensional conservation laws, a second-order accuracy can be achieved by local linear reconstructions. To prevent spurious oscillations where discontinuities may appear, a suitable limiter has to be applied on the gradient given by the linear reconstruction. One common approach in this fold is the Monotonic Upstream-Centered Scheme for Conservation Law (MUSCL) due to van Leer

✉ Ruo Li
rli@math.pku.edu.cn

Li Chen
cheney@pku.edu.cn

[1] School of Mathematical Sciences, Peking University, Beijing 100871, China

[2] CAPT, LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, China

[1]. Later MUSCL-type finite volume schemes were developed for multiple dimension problems on both structured and unstructured grids. In one-dimensional case the TVD criterion can prevent spurious oscillations and achieve second-order accuracy at the meantime, while TVD schemes in multi-dimensional cases have been proved to be at most first-order accurate [2]. Moreover, the implementation of TVD condition on unstructured grids is much less obvious. As an alternative for TVD criterion to prevent spurious oscillations, the local maximum principle was considered [3] which makes it easier to generalize to multi-dimensional problems. To ensure a local maximum principle, a new class of monotone scheme was proposed in [3] based on positivity of coefficients. The positivity of coefficients can be achieved by certain monotonicity condition and some additional geometric constraint on grids. Much subsequent research has been directed towards multi-dimensional numerical schemes which satisfy certain form of monotonicity condition [4–6]. The second-order TVD scheme of [5] was modified in [7] such that the local maximum principle is fulfilled by selecting the gradient which satisfies the monotonicity condition. In [8], an appropriate maximum principle was attained by restricting the gradient to lie within a *maximum principle region*. During the study of the multi-dimensional limiting process (MLP), Park et al. [9] extended the MLP condition from structured to unstructured grids, which also leads to a maximum principle. The MUSCL-type methods were generalized to higher-order reconstruction by introducing the Hessian [10]. The ENO/WENO-type methods are alternative methods to achieve higher-order accuracy. Although these methods have been successfully applied on unstructured grids [11–14], they are not free from parameters.

The linear reconstructions above consist of two steps. First, a predicted gradient is computed for each cell of the mesh using the cell averages in the reconstruction patch of the cell. Then the predicted gradient is limited to give a corrected gradient on the cell to prevent numerical oscillations. There exists a large family of so-called *scalar limiters*, which are implemented by multiplying all components of the gradient by a scalar in [0, 1]. Scalar limiters, though very popular, are far more diffusive, since the constraint triggered by any neighbouring cell will degrade the gradient in one specific direction, resulting limiting in all directions. Several approaches have been suggested to improve the scalar limiter with more consideration on the nature of multi-dimensional problems. Therefore, instead of reducing the magnitude of the predicted gradient, one may try to limit the predicted gradient *componentwisely* [15,16], as is done in the *vector limiters*. One essential difference between the scalar limiter and the vector limiter is that the limited gradient given by the scalar limiter has the same direction with the predicted gradient, while the vector limiter may not only modify the magnitude, but also the direction of the predicted gradient. Therefore, the vector limiter is actually trying to find a gradient in a multi-dimensional cone adjacent to the direction of the predicted gradient. Some authors then turned the vector limiter into the formulation of an optimization problem [17–19]. Taking [17] as an example, the limiter was formulated as a linear programming (LP) problem with the goal of retaining as much of the predicted gradient as possible while fulfilling the monotonicity condition.

Since the direction of the predicted gradient is modified, one actually has no obligation to take the predicted gradient as *a priori* information. Motivated by such an idea, we propose a new formulation of reconstruction. In this formulation, the gradient satisfying the monotonicity condition can be directly obtained without a gradient prediction. The prediction step and limiting step in the traditional reconstruction are therefore *integrated* into a single step. Our construction eventually gives rise to a tiny LP problem similar to that of [19], except with an alternative objective function that does not require gradient prediction any longer. Equipped with an efficient LP solver, the reconstructed gradient can be obtained directly with cheap computational cost. Moreover, the LP problem itself is free from the

grid topology, and thus the method may be applied to those discretization with very flexible grids. More importantly, our method is completely *parameter-free* and can thereby be easily extended to various applications. With some mild assumptions on the grids, we can show that the local maximum principle is preserved on triangular meshes. Similar analysis to alternative meshes can be performed. We present some numerical tests by solving scalar equations to validate our method. It is observed that the discrete scheme achieves satisfactory numerical accuracy without violating the discrete maximum principle. The numerical results for some benchmark problems governed by Euler equations are then provided to demonstrate the capacity of this new exploration.

This paper is organized as follows. At first, we briefly introduce the finite volume methods in Sect. 2, and then in Sect. 3 we present our linear reconstruction procedures in detail. In Sect. 4 we discuss the properties of our reconstruction method. Numerical examples will be given in Sect. 5 to demonstrate the robustness and accuracy of our method. Finally, a short conclusion will be drawn in Sect. 6.

2 MUSCL-Type Finite Volume Method

We briefly introduce the MUSCL-type finite volume methods for hyperbolic conservation laws on unstructured grids below. Let us consider a hyperbolic conservation law as follows

$$\frac{\partial \boldsymbol{u}}{\partial t} + \nabla \cdot \boldsymbol{F}(\boldsymbol{u}) = 0, \tag{1}$$

equipped with appropriate boundary conditions. The problem domain is triangulated into a mesh $\mathcal{T}$ with cells T_i. The cell-centered finite volume discretization is derived by first integrating (1) on the control volume T_i

$$\frac{\mathrm{d}\boldsymbol{u}_i}{\mathrm{d}t} + \frac{1}{|T_i|}\oint_{\partial T_i} \boldsymbol{F}\cdot\boldsymbol{n}\mathrm{d}l = 0, \tag{2}$$

where $\boldsymbol{u}_i(t)$ denotes the cell average of $\boldsymbol{u}$ on the cell T_i and $\boldsymbol{n}$ is the unit outward normal over the boundary ∂T_i, and then approximating the integral on the cell boundary appeared in (2) with the numerical quadrature

$$\oint_{\partial T_i} \boldsymbol{F}\cdot\boldsymbol{n}\mathrm{d}l \approx \sum_{e_{ij}\subset\partial T_i} |e_{ij}| \sum_{q=1}^{Q} w_q \boldsymbol{F}\left(\boldsymbol{u}(\boldsymbol{x}_{ij}^{(q)}, t)\right)\cdot\boldsymbol{n}_{ij}, \tag{3}$$

where e_{ij} represents the common edge between T_i and T_j, $\boldsymbol{x}_{ij}^{(q)}$ and w_q $(q = 1, 2, \cdots, Q)$ are the numerical quadrature points and corresponding weights respectively. Since we are concerned with second-order schemes, we adopt the midpoint quadrature rule and omit the index q hereinafter.

Next we supplant the flux $\boldsymbol{F}(\boldsymbol{u}(\boldsymbol{x}_{ij}, t))\cdot\boldsymbol{n}_{ij}$ at interfaces by a numerical flux function $\mathcal{F}(\boldsymbol{u}_{ij}^-, \boldsymbol{u}_{ij}^+)$, which yields a semi-discrete scheme formulated as

$$\frac{\mathrm{d}\boldsymbol{u}_i}{\mathrm{d}t} + \frac{1}{|T_i|}\sum_{e_{ij}\subset\partial T_i} \mathcal{F}(\boldsymbol{u}_{ij}^-, \boldsymbol{u}_{ij}^+)|e_{ij}| = 0,$$

where $\boldsymbol{u}_{ij}^-$ (resp. $\boldsymbol{u}_{ij}^+$) is the trace value of the numerical solution inside (resp. outside) the cell T_i.

As one of the most popular numerical flux functions, the *Lax-Friedrichs flux* can be expressed as

$$\mathcal{F}(\boldsymbol{u}^-, \boldsymbol{u}^+) = \frac{1}{2}[\boldsymbol{F}(\boldsymbol{u}^-) + \boldsymbol{F}(\boldsymbol{u}^+)] \cdot \boldsymbol{n} - \frac{1}{2}\alpha(\boldsymbol{u}^+ - \boldsymbol{u}^-), \tag{4}$$

where α is taken as an upper bound for the signal speed.

To achieve a higher than first-order numerical accuracy, the trace values $\boldsymbol{u}_{ij}^{\pm}$ are evaluated on a piecewise polynomial numerical solution reconstructed from cell-averaged values $\{\boldsymbol{u}_i\}$. This is the focus of this paper and we will give a method to reconstruct a piecewise linear numerical solution to calculate trace values appeared in the numerical flux.

So far we obtain a system of ordinary differential equations

$$\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = \mathcal{L}(\boldsymbol{u}). \tag{5}$$

The system of ODEs (5) are then discretized by a higher-order TVD Runge–Kutta methods [20]. For example, the first-order TVD Runge–Kutta scheme is simply the forward Euler scheme

$$\boldsymbol{u}^{n+1} = \boldsymbol{u}^n + \Delta t_n \mathcal{L}(\boldsymbol{u}^n), \tag{6}$$

and the second-order TVD Runge–Kutta scheme is given therein as

$$\begin{cases} \boldsymbol{u}^* & = \boldsymbol{u}^n + \Delta t_n \mathcal{L}(\boldsymbol{u}^n), \\ \boldsymbol{u}^{n+1} & = \dfrac{1}{2}\boldsymbol{u}^n + \dfrac{1}{2}\left(\boldsymbol{u}^* + \Delta t_n \mathcal{L}(\boldsymbol{u}^*)\right). \end{cases} \tag{7}$$

To match the second-order spatial discretization, we use the second-order TVD Runge–Kutta scheme (7). The time step length Δt_n can be determined from CFL condition

$$\Delta t_n = \mathrm{CFL} \times \min_{T_i \in \mathcal{T}} \left\{ \frac{h_i}{c_i} \right\},$$

where h_i is the size of the cell T_i and c_i is the maximal signal speed over T_i.

3 Integrated Linear Reconstruction

We present the procedures of our linear reconstruction in this section. The linear reconstruction requires a suitable gradient on each cell. In our method, the gradient is the solution of an LP problem. The LP problem is solved by an efficient algorithm. For simplicity, we will restrict ourselves to two-dimensional cases. The extension to three-dimensional is straightforward.

To start with, let us recall the typical procedure of the linear reconstructions in previous studies. Most of linear reconstructions consist of two steps: first construct a predicted gradient on each cell, and then correct the gradient to enforce some form of monotonicity condition. For clarity, we temporarily rename the relevant cell indices of a given cell within this section: the index 0 here stands for the current cell, whereas indices $1, 2, \ldots, m$ stand for its neighbouring cells. The neighbourhood may consist of either cells sharing the edges of the given cell, or cells sharing at least one vertex of the given cell. We refer to the former as *edge neighbours* (*EN*) (Fig. 1), whereas the latter as *vertex neighbours* (*VN*) (Fig. 2). Note that both definitions exclude the cell 0 itself.

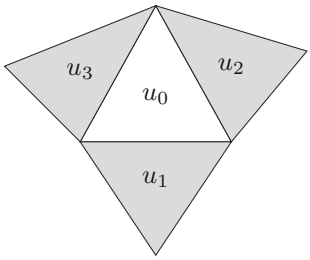

Fig. 1 Edge neighbours (EN)

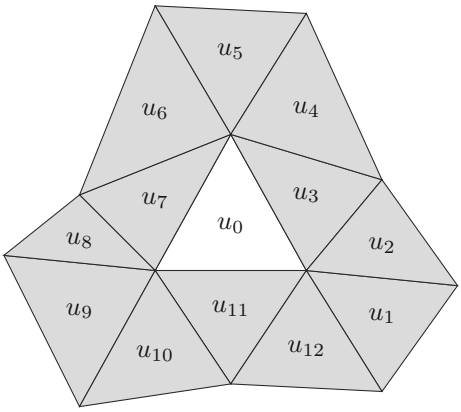

Fig. 2 Vertex neighbours (VN)

Let u_i and $\boldsymbol{x}_i = (x_i, y_i)$ be the cell-averaged value and mass center of T_i respectively $(i = 0, 1, \ldots, m)$. It is reasonable to assume that the mass centers $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are *non-colinear*. The reconstructed linear function on the cell T_0 can be formulated as

$$u(x, y) = u_0 + L_x(x - x_0) + L_y(y - y_0), \tag{8}$$

where $\boldsymbol{L} = [L_x, L_y]^\top$ is a limited form of the predicted gradient $\boldsymbol{g} = [g_x, g_y]^\top$. The predicted gradient is taken as the Green–Gauss or least-squares approximations etc. [4,8,9]. For example, the least-square reconstruction gives a gradient prediction by fitting the cell-averaged values of the cell and its neighbourhood

$$\begin{pmatrix} x_1 - x_0 & y_1 - y_0 \\ \vdots & \vdots \\ x_m - x_0 & y_m - y_0 \end{pmatrix} \begin{bmatrix} g_x \\ g_y \end{bmatrix} = \begin{bmatrix} u_1 - u_0 \\ \vdots \\ u_m - u_0 \end{bmatrix}. \tag{9}$$

The solution of (9) exists in the least-squares sense, i.e.

$$\begin{bmatrix} g_x \\ g_y \end{bmatrix} = \frac{1}{d_{xx}d_{yy} - d_{xy}^2} \begin{bmatrix} d_{yy}p_x - d_{xy}p_y \\ d_{xx}p_y - d_{xy}p_x \end{bmatrix}, \tag{10}$$

where

$$d_{xx} = \sum_{i=1}^{m}(x_i - x_0)^2, \quad d_{yy} = \sum_{i=1}^{m}(y_i - y_0)^2, \quad d_{xy} = \sum_{i=1}^{m}(x_i - x_0)(y_i - y_0),$$
$$p_x = \sum_{i=1}^{m}(x_i - x_0)(u_i - u_0), \quad p_y = \sum_{i=1}^{m}(y_i - y_0)(u_i - u_0).$$

Next a monotonicity condition is enforced to prevent spurious oscillations. We present scalar limiters with two versions of monotonicity conditions [19] for comparison with our method. Following the construction of MUSCL schemes, the gradient is limited such that the reconstructed value evaluated at the mass center of any neighbouring cell is bounded by u_0 and the cell-averaged value of that cell, i.e.

$$\min\{u_0, u_i\} \le u(x_i, y_i) \le \max\{u_0, u_i\}, \quad i = 1, 2, \ldots, m. \tag{11}$$

The constraints (11) may be viewed as a two-dimensional generalization of the *minmod* limiter for one-dimensional case. Later in Sect. 3.1 we will give a further remark on this point.

Instead of limiting using two cell-averaged values, a less diffusive approach is to limit the predicted gradient using the maximum and minimum over all neighbouring cells and the cell itself, i.e.

$$\min\{u_0, u_1, \ldots, u_m\} \le u(x_i, y_i) \le \max\{u_0, u_1, \ldots, u_m\}, \quad i = 1, 2, \ldots, m. \tag{12}$$

Following the terminology of [19], we refer to the monotonicity condition (11) as the *standard formulation*, whereas (12) as the *relaxed formulation*. As will be shown in Sect. 4, either (11) or (12) leads to a certain discrete maximum principle and subsequently $\mathbb{L}^\infty$-stability with an additional geometric hypothesis on the grid.

The predicted gradient is then multiplied by a scalar factor $\phi \in [0, 1]$ as large as possible without violating the monotonicity condition

$$\boldsymbol{L} = \phi \boldsymbol{g}. \tag{13}$$

Explicit expressions are available for the scalar limiter by substituting (8) (13) into (11) or (12). The standard form of monotonicity condition (11) leads to the following upper bound of ϕ

$$\Phi_i = \begin{cases} \dfrac{v_i^+}{\boldsymbol{g} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_0)}, & \boldsymbol{g} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_0) > v_i^+, \\ \dfrac{v_i^-}{\boldsymbol{g} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_0)}, & \boldsymbol{g} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_0) < v_i^-, \\ 1, & \text{otherwise}, \end{cases} \tag{14}$$

for $i = 1, 2, \ldots, m$, while the relaxed form (12) leads to the following upper bound of ϕ

$$\Phi_i = \begin{cases} \dfrac{v_{\max}^+}{\boldsymbol{g} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_0)}, & \boldsymbol{g} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_0) > v_{\max}^+, \\ \dfrac{v_{\min}^-}{\boldsymbol{g} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_0)}, & \boldsymbol{g} \cdot (\boldsymbol{x}_i - \boldsymbol{x}_0) < v_{\min}^-, \\ 1, & \text{otherwise}, \end{cases} \tag{15}$$

for $i = 1, 2, \ldots, m$, where

$$v_i = u_i - u_0, \quad v_i^+ = \max\{v_i, 0\}, \quad v_i^- = \min\{v_i, 0\}, \quad i = 1, 2, \ldots, m,$$
$$v_{\max}^+ = \max\{0, v_1, v_2, \ldots, v_m\}, \quad v_{\min}^- = \min\{0, v_1, v_2, \ldots, v_m\}.$$

Finally the scalar factor ϕ is selected as the minimum of all Φ_i's

$$\phi = \min\{\Phi_1, \Phi_2, \ldots, \Phi_m\}.$$

Instead of a single factor in the scalar limiter, the vector limiter is implemented by multiplying different factors to each component of the predicted gradient. Since the direction of

the predicted gradient is modified, one may be persuaded that the prediction of the gradient is actually unnecessary. As a further exploration of our method, the gradient is obtained without prediction at all. Since our reconstruction approach is not split into the prediction and limiting steps any more, we will refer it as the *integrated linear reconstruction (ILR)* from now on.

3.1 Reconstruction as an LP Problem

For better numerical accuracy and less numerical diffusion, the reconstructed values should fit the cell-averaged values as best as possible without violating any monotonicity conditions. In the following construction we will adopt the standard form of the monotonicity condition (11). Denote by $\hat{u}_i = u_0 + L_x(x_i - x_0) + L_y(y_i - y_0)$ the reconstructed value at the mass center of neighbouring cell T_i $(i = 1, 2, \ldots, m)$, then the inequality (11) becomes

$$v_i^- \leq (x_i - x_0)L_x + (y_i - y_0)L_y \leq v_i^+, \quad i = 1, 2, \ldots, m. \tag{16}$$

We use the l^1-norm to measure the difference between the reconstructed and cell-averaged values when (16) holds. Define the total gaps δ by

$$\delta(L_x, L_y) := \sum_{i=1}^{m} |u_i - \hat{u}_i| = \sum_{i=1}^{m} |v_i - \hat{v}_i|. \tag{17}$$

where $\hat{v}_i = \hat{u}_i - u_0 = (x_i - x_0)L_x + (y_i - y_0)L_y$ $(i = 1, 2, \ldots, m)$.

Our goal is to minimize the total gaps δ subjected to the monotonicity conditions (16) (see Fig. 3). This seems a non-linear objective function at first glance. However, one observes that, the difference $v_i - \hat{v}_i$ has the same sign with v_i, and consequently

$$\begin{aligned} \delta &= \sum_{i=1}^{m} \operatorname{sgn}(v_i)(v_i - \hat{v}_i) \\ &= -\sum_{i=1}^{m} \operatorname{sgn}(v_i)\hat{v}_i + \text{const} \end{aligned}$$

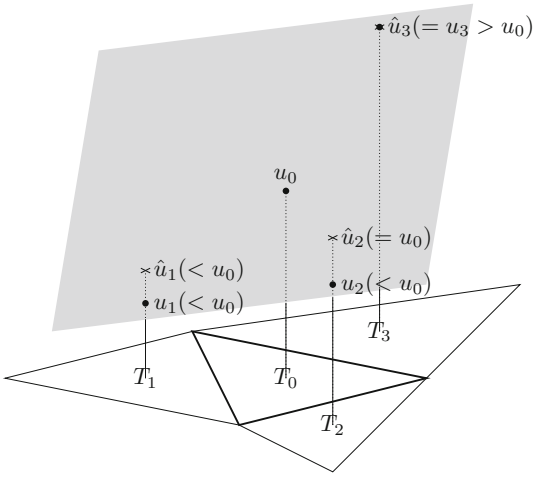

Fig. 3 Illustration of the integrated linear reconstruction: the *shaded plane* represents linear reconstruction on cell T_0; the *cross symbols* represent reconstructed values, while the *dot symbols* represent cell-averaged values

$$= -\sum_{i=1}^{m} \text{sgn}\,(v_i)[(x_i - x_0)L_x + (y_i - y_0)L_y] + \text{const}$$
$$= -\left[\sum_{i=1}^{m} \text{sgn}\,(v_i)(x_i - x_0)\right] L_x - \left[\sum_{i=1}^{m} \text{sgn}\,(v_i)(y_i - y_0)\right] L_y + \text{const}.$$

As a result, δ is actually an *affine* function of the limited gradient (L_x, L_y).

So far we have formulated the linear reconstruction in the following LP problem

$$\begin{aligned} &\max \left[\sum_{i=1}^{m} \text{sgn}\,(v_i)(x_i - x_0)\right] L_x + \left[\sum_{i=1}^{m} \text{sgn}\,(v_i)(y_i - y_0)\right] L_y \\ &\text{s.t. } v_i^- \le (x_i - x_0)L_x + (y_i - y_0)L_y \le v_i^+, \quad i = 1, 2, \ldots, m. \end{aligned} \tag{18}$$

In particular the solution of (18) exactly recovers the linear data, since the exact gradient satisfies all the constraints and achieves minimal total gaps $\delta = 0$.

Remark 1 In one-dimensional case, our approach actually degenerates to the *minmod slope*. Indeed, consider two neighbours assigned with labels 1 and 2 such that $x_1 < x_0 < x_2$. Then the LP problem (18) reduces to

$$\begin{aligned} \max \quad & [\text{sgn}\,(v_1)(x_1 - x_0) + \text{sgn}\,(v_2)(x_2 - x_0)]L_x \\ \text{s.t.} \quad & v_1^- \le (x_1 - x_0)L_x \le v_1^+ \text{ and } v_2^- \le (x_2 - x_0)L_x \le v_2^+. \end{aligned} \tag{19}$$

It is easy to derive the optimal solution of (19)

$$L_x = \text{minmod}\left(\frac{u_1 - u_0}{x_1 - x_0}, \frac{u_2 - u_0}{x_2 - x_0}\right),$$

where the minmod function is defined by

$$\text{minmod}\,(a, b) = \begin{cases} a, & \text{if } ab > 0 \text{ and } |a| < |b|, \\ b, & \text{if } ab > 0 \text{ and } |a| \ge |b|, \\ 0, & \text{if } ab \le 0. \end{cases}$$

A subsequent question is about the existence of the solution for LP problem (18). Since each double-sided constraint of (16) corresponds to a strip in the (L_x, L_y)-plane, the feasible region of (18) is the intersection of m strips (see Fig. 4). Consider the region enclosed by any two of the m constraints (16), saying $i = 1$ and 2 for instance

$$\begin{cases} v_1^- \le (x_1 - x_0)L_x + (y_1 - y_0)L_y \le v_1^+, \\ v_2^- \le (x_2 - x_0)L_x + (y_2 - y_0)L_y \le v_2^+. \end{cases}$$

This region is bounded as long as $\boldsymbol{x}_0$, $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are non-colinear. Since $\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ are non-colinear, the feasible region of (18) is always bounded. Another observation is that $(L_x, L_y) = (0, 0)$ must be a feasible point. From the theory of LP, there exists a bounded optimal solution for (18).

3.2 LP Solver

The efficiency of LP solver is crucial since we need to solve an LP problem on *each* cell. Following [19], the *all-inequality simplex method* is taken as our LP solver. This variant of simplex method applies to LP problems consisting exclusively of inequality constraints. The

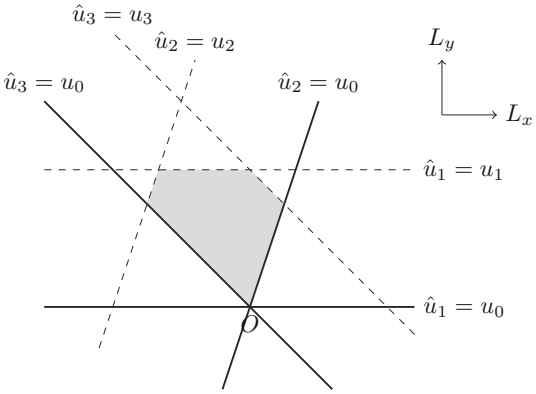

Fig. 4 Feasible region when $m = 3$: the point (L_x, L_y) on the *solid lines* and *dashed lines* satisfies $u_0+(x_i-x_0)L_x+(y_i-y_0)L_y = u_0$ and $u_0+(x_i-x_0)L_x+(y_i-y_0)L_y = u_i$ respectively $(i = 1, 2, 3)$

idea is the same as the standard simplex algorithm: if there is an optimal solution for the LP, then there exists at least one vertex in the feasible region that achieves optimality. We search this vertex in an iterative manner. Compared to traditional simplex methods, this method is especially useful if the number of inequality constraints is much greater than the number of variables, which is the very case here. The algorithm of the all-inequality simplex method for the following LP problem

$$\max \boldsymbol{c}^\top \boldsymbol{x} \quad \text{s.t. } \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}, \tag{20}$$

where $\boldsymbol{c}, \boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{A} \in \mathbb{R}^{N\times d}$ and $\boldsymbol{b} \in \mathbb{R}^N$, can be found in "Appendix".

In our case $d = 2$ and $N = 2m$. The corresponding variables become

$$\boldsymbol{c} = \left[\sum_{i=1}^{m} \operatorname{sgn}(v_i)(x_i - x_0), \sum_{i=1}^{m} \operatorname{sgn}(v_i)(y_i - y_0)\right]^\top, \quad \boldsymbol{x} = \boldsymbol{L} = [L_x, L_y]^\top,$$

$$\boldsymbol{A} = \begin{pmatrix} x_1 - x_0 & y_1 - y_0 \\ \vdots & \vdots \\ x_m - x_0 & y_m - y_0 \\ -(x_1 - x_0) & -(y_1 - y_0) \\ \vdots & \vdots \\ -(x_m - x_0) & -(y_m - y_0) \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} v_1^+ \\ \vdots \\ v_m^+ \\ -v_1^- \\ \vdots \\ -v_m^- \end{pmatrix}.$$

Besides, it is known *a priori* that $\boldsymbol{L} = [0, 0]^\top$ is a valid starting point. We initialize the active matrix $\boldsymbol{M}$ with the rows of $\boldsymbol{A}$ corresponding to the following two constraints

$$(x_1 - x_0)L_x + (y_1 - y_0)L_y \lesseqgtr 0 \quad \text{and} \quad (x_2 - x_0)L_x + (y_2 - y_0)L_y \lesseqgtr 0.$$

The cost of the reconstruction is largely affected by the number of iterations needed to solve the LP. In all the two-dimensional cases we have run so far, the maximum number of iterations was 5. Moreover, one iteration of this simplex method is also very cheap, which requires only the solution of two 2×2 linear systems to determine the Lagrange multiplier and search direction.

4 Discrete Maximum Principle

In this section we first give an *a posteriori* estimation of the gradient reduction, and then prove the local maximum principle for our ILR.

In [19], May et al. was aiming at retaining as much of the unlimited gradient as possible without violating the monotonicity conditions. The objective function then equals to the gradient reduction in l^1-norm: $\|\boldsymbol{g}-\boldsymbol{L}\|_1 = |g_x - L_x| + |g_y - L_y|$. Although optimal solutions of these two problems do not coincide, we claim that our optimal solution actually achieves *quasi-optimality* in the sense of minimal gradient reduction. This statement is made more clear in the following Theorem 1.

Theorem 1 (*A posteriori* estimation of gradient reduction). *Let $\boldsymbol{g}$ be the least-square predicted gradient given by* (10) *and $\boldsymbol{L}$ be any feasible gradient satisfying* (16)*, then the following l^1-estimation holds*

$$\|\boldsymbol{g}-\boldsymbol{L}\|_1 \le C\delta, \tag{21}$$

where

$$C = \frac{1}{d_{xx}d_{yy}-d_{xy}^2} \max_{1\le i\le m} \left[|d_{yy}(x_i-x_0)-d_{xy}(y_i-y_0)|+|d_{xx}(y_i-y_0)-d_{xy}(x_i-x_0)|\right] \tag{22}$$

is a geometry-dependent constant and δ is the total gaps (17)*.*

Proof Denote

$$\boldsymbol{X} = \begin{pmatrix} x_1-x_0 & \dots & x_m-x_0 \\ y_1-y_0 & \dots & y_m-y_0 \end{pmatrix}, \quad \boldsymbol{v} = [v_1,\dots,v_m]^\top, \quad \hat{\boldsymbol{v}} = [\hat{v}_1,\dots,\hat{v}_m]^\top,$$

then we have the following two identities

$$\boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{g} = \boldsymbol{X}\boldsymbol{v} \quad \text{and} \quad \boldsymbol{X}\boldsymbol{X}^\top\boldsymbol{L} = \boldsymbol{X}\hat{\boldsymbol{v}}.$$

Subtracting these two identities yields

$$\boldsymbol{X}\boldsymbol{X}^\top(\boldsymbol{g}-\boldsymbol{L}) = \boldsymbol{X}(\boldsymbol{v}-\hat{\boldsymbol{v}}).$$

By assumption $\boldsymbol{X}\boldsymbol{X}^\top$ is invertible. Therefore

$$\|\boldsymbol{g}-\boldsymbol{L}\|_1 \le \|(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}\|_1\|\boldsymbol{v}-\hat{\boldsymbol{v}}\|_1 = C\delta,$$

where C, given in (22), is the l^1-norm of the matrix $(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{X}$. □

From Theorem 1 we conclude that by minimizing the total gaps δ, the quasi-optimality of the reduction of the resulting gradient $\boldsymbol{L}$ with respect to $\boldsymbol{g}$ is enforced.

We establish below the discrete maximum principle of linear reconstruction with either EN or VN for scalar conservation laws. The whole proof is sketched out in the following three stages: first we show that the solution is bounded from above and below by interpolants at the previous time level (Lemma 4); next we propose technical conditions of values at quadrature points that lead to a maximum principle (Theorem 5); finally we demonstrate that the monotonicity condition along with some geometric hypothesis results in the proposed technical conditions (Theorems 6, 7). The following notations of neighbouring indices are introduced to simplify the expressions (see Figs. 1, 2)

$$EN(i) := \{j | e_{ij} \subset \partial T_i\},$$
$$VN(i) := \{j | T_i \cap T_j \neq \varnothing, j \neq i\},$$
$$\overline{EN}(i) := EN(i) \cup \{i\},$$
$$\overline{VN}(i) := VN(i) \cup \{i\}.$$

We also define the extremal trace values and extremal interior values at the quadrature points through

$$U_i^{\min} := \min_{j \in EN(i)} \{u_{ij}^{\pm}\}, \quad U_i^{\max} := \max_{j \in EN(i)} \{u_{ij}^{\pm}\},$$
$$\mathcal{U}_i^{\min} := \min_{j \in EN(i)} \{u_{ij}^{-}\}, \quad \mathcal{U}_i^{\max} := \max_{j \in EN(i)} \{u_{ij}^{-}\}.$$

In the following argument we consider the fully discrete finite volume scheme

$$u_i^{n+1} = u_i^n - \frac{\Delta t_n}{|T_i|} \sum_{j \in EN(i)} \mathcal{F}(u_{ij}^-, u_{ij}^+)|e_{ij}|. \tag{23}$$

Definition 2 We say that the finite volume scheme (23) fulfils *the local maximum principle*, if the solution at a given cell is bounded by the minimum and maximum values within this cell and its VN at the previous time step, i.e.

$$u_i^{\min} \leq u_i^{n+1} \leq u_i^{\max}, \tag{24}$$

where the local extrema

$$u_i^{\min} := \min_{j \in \overline{VN}(i)} \{u_j^n\}, \quad u_i^{\max} := \max_{j \in \overline{VN}(i)} \{u_j^n\}.$$

And we say that the finite volume scheme (23) fulfils *the global maximum principle*, if the global solution maximum is non-increasing and the global solution minimum is non-decreasing between two successive time levels, i.e.

$$u^{\min,n+1} \geq u^{\min,n}, \quad u^{\max,n+1} \leq u^{\max,n}, \tag{25}$$

where the global extrema

$$u^{\min,n} := \min_{T_i \in \mathcal{T}} \{u_i^n\}, \quad u^{\max,n} := \max_{T_i \in \mathcal{T}} \{u_i^n\}.$$

Of course the local maximum principle implies the global maximum principle, but the converse is not true. Before the proof of maximum principle, let us investigate some basic properties of linear reconstruction.

Lemma 3 *With the notations defined above and a monotone Lipschitz continuous numerical flux function $\mathcal{F}$, the fully discrete finite volume scheme* (23) *satisfies the following inequality*

$$\frac{L_i}{|T_i|} M(U_i^{\min} - \mathcal{U}_i^{\max}) \leq \frac{u_i^{n+1} - u_i^n}{\Delta t_n} \leq \frac{L_i}{|T_i|} M(U_i^{\max} - \mathcal{U}_i^{\min}), \tag{26}$$

where L_i denotes the perimeter of T_i and M is the Lipschitz constant of $\mathcal{F}$ with respect to the second argument, i.e.

$$|\mathcal{F}(u^-, u^+) - \mathcal{F}(u^-, \tilde{u}^+)| \leq M|u^+ - \tilde{u}^+| \quad \forall u^-, u^+, \tilde{u}^+ \in [U_i^{\min}, U_i^{\max}].$$

Proof Utilizing the consistency and monotonicity of numerical flux function, we have

$$\begin{aligned}
\frac{u_i^{n+1}-u_i^n}{\Delta t_n} &= -\frac{1}{|T_i|}\sum_{j\in EN(i)}\mathcal{F}(u_{ij}^-,u_{ij}^+)|e_{ij}| \\
&= -\frac{1}{|T_i|}\sum_{j\in EN(i)}\Big[\mathcal{F}(\mathcal{U}_i^{\min},U_i^{\max})|e_{ij}|+(\mathcal{F}(u_{ij}^-,u_{ij}^+)-\mathcal{F}(\mathcal{U}_i^{\min},u_{ij}^+))|e_{ij}| \\
&\qquad +(\mathcal{F}(\mathcal{U}_i^{\min},u_{ij}^+)-\mathcal{F}(\mathcal{U}_i^{\min},U_i^{\max}))|e_{ij}|\Big] \\
&\le -\frac{1}{|T_i|}\sum_{j\in EN(i)}\mathcal{F}(\mathcal{U}_i^{\min},U_i^{\max})|e_{ij}| \\
&= -\frac{1}{|T_i|}\sum_{j\in EN(i)}\big[\mathcal{F}(\mathcal{U}_i^{\min},U_i^{\max})-\mathcal{F}(\mathcal{U}_i^{\min},\mathcal{U}_i^{\min})\big]|e_{ij}| \\
&\le \frac{1}{|T_i|}M(U_i^{\max}-\mathcal{U}_i^{\min})\sum_{j\in EN(i)}|e_{ij}| \\
&= \frac{L_i}{|T_i|}M(U_i^{\max}-\mathcal{U}_i^{\min}),
\end{aligned}$$

In a similar manner we can prove the left hand side of the inequality (26). □

Next we attempt to eliminate the extremal interior value terms $\mathcal{U}_i^{\min}$ and $\mathcal{U}_i^{\max}$ appeared in (26). In fact, utilizing the interpretation of Barth's geometric shape parameter [21] one can prove that for linear reconstruction there exists a constant $\Gamma\in(1,+\infty)$ such that

$$\mathcal{U}_i^{\max}-\mathcal{U}_i^{\min}\le\Gamma(\mathcal{U}_i^{\max}-u_i^n)\quad\text{and}\quad\mathcal{U}_i^{\max}-\mathcal{U}_i^{\min}\le\Gamma(u_i^n-\mathcal{U}_i^{\min}). \tag{27}$$

For triangular control volumes, the geometric shape parameter $\Gamma=3$. As a result, the updated solution is bounded by interpolants of extremal trace values and the current solution. Precisely we have the following result.

Lemma 4 *The fully discrete finite volume scheme* (23) *constructed with a monotone Lipschitz continuous numerical flux function and linear reconstruction exhibits the local interpolated interval bound*

$$\sigma_i U_i^{\min}+(1-\sigma_i)u_i^n\le u_i^{n+1}\le\sigma_i U_i^{\max}+(1-\sigma_i)u_i^n,$$

where the time-step-proportional interpolation parameter σ_i *is defined by*

$$\sigma_i=\frac{L_i}{|T_i|}M\Gamma\Delta t_n.$$

Proof By definition $U_i^{\max}\ge\mathcal{U}_i^{\max}\ge\mathcal{U}_i^{\min}\ge U_i^{\min}$. Applying the inequalities (27) we conclude that

$$\begin{aligned}
&U_i^{\max}-\mathcal{U}_i^{\min}-\Gamma(U_i^{\max}-u_i^n) \\
&\quad=(\mathcal{U}_i^{\max}-\mathcal{U}_i^{\min})-\Gamma(\mathcal{U}_i^{\max}-u_i^n)-(\Gamma-1)(U_i^{\max}-\mathcal{U}_i^{\max})\le 0, \\
&U_i^{\min}-\mathcal{U}_i^{\max}-\Gamma(U_i^{\min}-u_i^n) \\
&\quad=\Gamma(u_i^n-\mathcal{U}_i^{\min})-(\mathcal{U}_i^{\max}-\mathcal{U}_i^{\min})+(\Gamma-1)(\mathcal{U}_i^{\min}-U_i^{\min})\ge 0.
\end{aligned}$$

Consequently,

$$U_i^{\max} - \mathcal{U}_i^{\min} \le \Gamma(U_i^{\max} - u_i) \quad \text{and} \quad U_i^{\min} - \mathcal{U}_i^{\max} \ge \Gamma(U_i^{\min} - u_i). \tag{28}$$

Inserting (28) into (26) and rearranging terms give rise to the desired results. □

Given the two-sided bound, a discrete maximum principle can be established under a CFL-like time step restriction provided that the extremal trace values $U_i^{\max}$ and $U_i^{\min}$ can be bounded from both above and below by the neighbouring cell averages. Actually we have the following Theorem 5.

Theorem 5 *The fully discrete finite volume scheme* (23) *constructed with a monotone Lipschitz continuous numerical flux function and linear reconstruction fulfils the global maximum principle* (25) *if on each cell T_i the following condition holds*

$$u^{\min,n} \le u_{ij}^- \le u^{\max,n}, \quad \forall j \in EN(i). \tag{29}$$

Furthermore, the scheme fulfils the local maximum principle (24) *if*

$$\max\{u_i^{\min}, u_j^{\min}\} \le u_{ij}^- \le \min\{u_i^{\max}, u_j^{\max}\}, \quad \forall j \in EN(i). \tag{30}$$

In particular, the local maximum principle is satisfied if

$$\min\{u_i^n, u_j^n\} \le u_{ij}^- \le \max\{u_i^n, u_j^n\}, \quad \forall j \in EN(i). \tag{31}$$

Proof Assuming the time step restriction

$$\Delta t_n \le \min_{T_i \in \mathcal{T}} \left\{ \frac{|T_i|}{M\Gamma L_i} \right\},$$

we have $0 \le \sigma_i \le 1$ and consequently

$$u^{\min,n} = \sigma_i u^{\min,n} + (1-\sigma_i)u^{\min,n} \le u_i^{n+1} \le \sigma_i u^{\max,n} + (1-\sigma_i)u^{\max,n} = u^{\max,n},$$

when (29) holds. The global maximum principle then follows.

Now we rewrite the condition (30) as

$$u_i^{\min} \le u_{ij}^{\pm} \le u_i^{\max}.$$

Then the extremal trace values can be bounded from above and below by the local extrema

$$u_i^{\min} \le U_i^{\min} \le U_i^{\max} \le u_i^{\max},$$

and as a result,

$$u_i^{\min} = \sigma_i u_i^{\min} + (1-\sigma_i)u_i^{\min} \le u_i^{n+1} \le \sigma_i u_i^{\max} + (1-\sigma_i)u_i^{\max} = u_i^{\max},$$

which leads to the local maximum principle (24). □

Let us recall the monotonicity conditions in Sect. 3 in a formal notation: the standard form

$$\min\{u_j^n, u_i^n\} \le u(\boldsymbol{x}_j) \le \max\{u_j^n, u_i^n\}, \tag{32}$$

and the relaxed form

$$\min_j\{u_j^n\} \le u(\boldsymbol{x}_j) \le \max_j\{u_j^n\}, \tag{33}$$

where either $j \in \overline{EN}(i)$ or $j \in \overline{VN}(i)$.

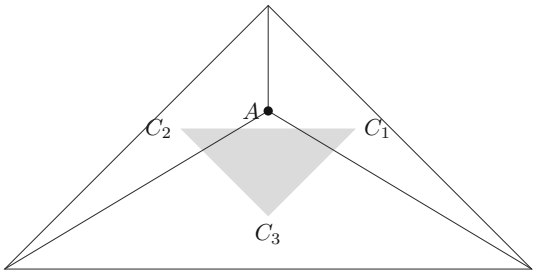

Fig. 5 Violation of Hypothesis $\mathcal{H}_1$: vertex A is not contained in the convex hull spanned by mass centers of cells from $N(A)$

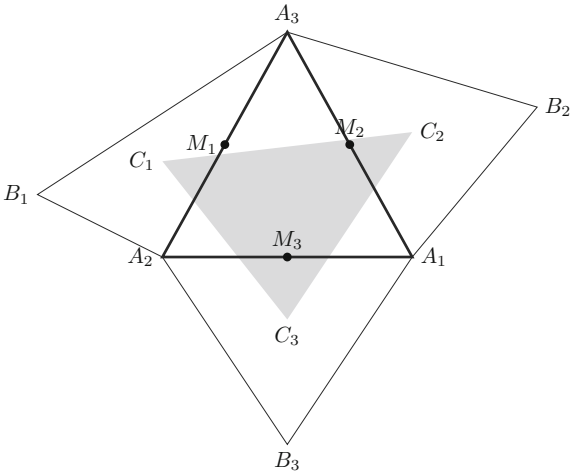

Fig. 6 Violation of Hypothesis $\mathcal{H}_2$: midpoint M_1 is not contained in the triangle $\Delta C_1C_2C_3$

It is worth mentioning that these conditions all restrict values at *the adjacent mass centers* rather than *the quadrature points*. However, with an additional geometric hypothesis the quadrature points will be contained in the convex hull spanned by the adjacent mass centers, and the restriction on the quadrature points can thereby be obtained. Our results below rely on the following Hypothesis $\mathcal{H}_1$ or $\mathcal{H}_2$

- Hypothesis $\mathcal{H}_1$: Any vertex of the grid falls in the convex hull spanned by mass centers of cells sharing this vertex (Fig. 5);
- Hypothesis $\mathcal{H}_2$: For each cell of the grid, the edge midpoints fall in the convex hull spanned by mass centers of EN (Fig. 6).

Hypothesis $\mathcal{H}_1$ is required in the proof of maximum principle for VN, while Hypothesis $\mathcal{H}_2$ is required in the proof of maximum principle for EN. Here we present two general results for linear reconstruction in Theorems 6 and 7, from which it can be shown that our ILR satisfies the local maximum principle.

Theorem 6 (Maximum principle for vertex neighbours). *Consider the fully discrete finite volume scheme* (23) *with a monotone Lipschitz continuous numerical flux function and linear reconstruction from VN on a grid satisfying Hypothesis* $\mathcal{H}_1$.

1. *If the linear reconstruction satisfies the standard form* (11) *of monotonicity condition, then the scheme satisfies a local maximum principle* (24) *under a proper CFL restriction.*
2. *If the linear reconstruction satisfies the relaxed form* (12) *of monotonicity condition, then only a global maximum principle* (25) *can be guaranteed under a proper CFL restriction.*

Proof (1) Consider two adjacent cells T_i and T_j with a common edge $e_{ij} = \overline{AB}$. From Hypothesis $\mathcal{H}_1$ we know that the vertex A falls in the convex hull spanned by mass centers of cells from $N(A)$, the set of indices of cells sharing the vertex A. As a result,

$$\min_{k\in N(A)}\{u(\boldsymbol{x}_k)\} \le u(A) \le \max_{k\in N(A)}\{u(\boldsymbol{x}_k)\}.$$

Utilizing the standard form of monotonicity condition

$$\min\{u_k^n, u_i^n\} \le u(\boldsymbol{x}_k) \le \max\{u_k^n, u_i^n\}, \quad \forall k \in N(A), \tag{34}$$

we conclude that

$$\min_{k\in N(A)}\{u_k^n\} \le u(A) \le \max_{k\in N(A)}\{u_k^n\}.$$

The inclusion relations $N(A) \subset \overline{VN}(i)$ and $N(A) \subset \overline{VN}(j)$ lead to

$$\min_{k\in \overline{VN}(i)}\{u_k^n\} \le u(A) \le \max_{k\in \overline{VN}(i)}\{u_k^n\}, \quad \min_{k\in \overline{VN}(j)}\{u_k^n\} \le u(A) \le \max_{k\in \overline{VN}(j)}\{u_k^n\}, \tag{35}$$

therefore

$$\max\{u_i^{\min}, u_j^{\min}\} \le u(A) \le \min\{u_i^{\max}, u_j^{\max}\} \tag{36}$$

We can obtain the same bounds for vertex B. Therefore the interpolation u_{ij}^- satisfies (30) and the local maximum principle (24) follows.

(2) If we are using the relaxed form of monotonicity condition, then (34) should be replaced with

$$u_i^{\min} \le u(\boldsymbol{x}_k) \le u_i^{\max}, \quad \forall k \in N(A). \tag{37}$$

Therefore,

$$u_i^{\min} \le u(A) \le u_i^{\max}.$$

Similarly,

$$u_i^{\min} \le u(B) \le u_i^{\max}.$$

As a result,

$$u_i^{\min} \le u_{ij}^- \le u_i^{\max}.$$

Note that generally the inequality $u_j^{\min} \le u_{ij}^- \le u_j^{\max}$ does not hold. Nevertheless we can still obtain a global maximum principle from (29). □

Theorem 7 (Maximum principle for edge neighbours). *Consider the fully discrete finite volume scheme* (23) *with a monotone Lipschitz continuous numerical flux function and linear reconstruction from EN on a grid satisfying Hypothesis* $\mathcal{H}_2$. *If the linear reconstruction satisfies either standard form* (11) *or relaxed form* (12) *of the monotonicity condition, then the scheme exhibits a local maximum principle* (24) *under a proper CFL restriction.*

Proof Again we consider two adjacent cells T_i and T_j with a common edge. From Hypothesis $\mathcal{H}_2$ we know that

$$\min_{k\in EN(i)}\{u(\boldsymbol{x}_k)\} \le u_{ij}^- \le \max_{k\in EN(i)}\{u(\boldsymbol{x}_k)\}.$$

Moreover, for any $k \in EN(i)$, both standard and relaxed forms lead to

$$\min_{l \in \overline{EN}(i)} \{u_l^n\} \le u(\boldsymbol{x}_k) \le \max_{l \in \overline{EN}(i)} \{u_l^n\}.$$

Note that $\overline{EN}(i) \subset \overline{VN}(i)$ and $\overline{EN}(i) \subset \overline{VN}(j)$. As a result,

$$\max\left\{\min_{l \in \overline{VN}(i)} \{u_l^n\}, \min_{l \in \overline{VN}(j)} \{u_l^n\}\right\} \le u_{ij}^- \le \min\left\{\max_{l \in \overline{VN}(i)} \{u_l^n\}, \max_{l \in \overline{VN}(j)} \{u_l^n\}\right\}.$$

The local maximum principle then follows from (30). □

The geometry of the grid has to be examined for Hypothesis $\mathcal{H}_2$. Here we give a sufficient condition for Hypothesis $\mathcal{H}_2$ on a triangular grid, which is the case in our numerical experiments.

Lemma 8 *For any triangle $\Delta A_1A_2A_3$ in the grid. Let $\Delta B_1A_2A_3$, $\Delta A_1B_2A_3$ and $\Delta A_1A_2B_3$ be the edge neighbours of $\Delta A_1A_2A_3$* (see Fig. 6) *and assume the following constraint on the triangulation*

$$-\frac{4}{3} \le \lambda_j(B_i) \le -\frac{2}{3} \quad (i = j),$$
$$\frac{2}{3} \le \lambda_j(B_i) \le \frac{4}{3} \quad (i \ne j),$$

where λ_j denotes the barycentric coordinate with respect to $A_j (j = 1, 2, 3)$. Then the grid satisfies Hypothesis $\mathcal{H}_2$.

Proof Let C_1, C_2 and C_3 be the mass centers of three EN accordingly. Denote by $\mathcal{A}$ and $\mathcal{C}$ the barycentric coordinate systems of $\Delta A_1A_2A_3$ and of $\Delta C_1C_2C_3$. Then the matrix formed by $\mathcal{A}$-coordinates of edge midpoints is

$$\boldsymbol{M} = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}.$$

Let $\boldsymbol{B}$ be the matrix formed by $\mathcal{A}$-coordinates of B_1, B_2 and B_3. Then the matrix formed by $\mathcal{A}$-coordinates of C_1, C_2 and C_3 is $(\boldsymbol{B} + 2\boldsymbol{M})/3$. Furthermore, let $\boldsymbol{P}$ be the matrix formed by $\mathcal{C}$-coordinates of edge midpoints, then we have

$$\boldsymbol{M} = \frac{1}{3}(\boldsymbol{B} + 2\boldsymbol{M})\boldsymbol{P}.$$

Now Hypothesis $\mathcal{H}_2$ is equivalent to the entrywise non-negativity of matrix $\boldsymbol{P}$. By assumption, the entry of matrix $\boldsymbol{K} = \boldsymbol{B} + 2\boldsymbol{M}$ satisfies

$$-\frac{4}{3} \le K_{ij} \le -\frac{2}{3} \quad (i = j),$$
$$\frac{5}{3} \le K_{ij} \le \frac{7}{3} \quad (i \ne j).$$

By carrying out an algebraic manipulation we can show that all entries of $\boldsymbol{P} = 3\boldsymbol{K}^{-1}\boldsymbol{M}$ are non-negative. □

For grids with alternative cell geometry, similar conditions may be required.

Table 1 Results for the advection of double sine wave

	Cells	$\mathbb{L}^1$ error	Order	$\mathbb{L}^\infty$ error	Order	CPU time (s)
Unlimited	992	2.23e−2	1.99	5.13e−2	2.01	0.7
	3968	5.59e−3	2.00	1.25e−2	2.04	6.0
	15,872	1.40e−3	2.00	3.06e−3	2.03	49.5
	63488	3.51e−4	2.00	7.66e−4	2.00	397.9
SSL	992	1.73e−1	0.78	4.79e−1	0.67	0.9
	3968	8.62e−2	1.01	2.66e−1	0.85	7.4
	15,872	4.44e−2	0.96	1.37e−1	0.95	60.3
	63488	2.41e−2	0.88	7.38e−2	0.90	481.7
RSL	992	6.14e−2	1.66	2.65e−1	1.09	1.0
	3968	1.94e−2	1.66	1.18e−1	1.17	8.0
	15,872	5.52e−3	1.82	5.02e−2	1.23	61.7
	63488	1.46e−3	1.92	2.07e−2	1.27	503.9
ILR	992	7.02e−2	1.55	2.86e−1	1.05	1.0
	3968	2.21e−2	1.67	1.34e−1	1.10	8.6
	15,872	5.99e−3	1.88	5.80e−2	1.20	68.7
	63488	1.58e−3	1.93	2.45e−2	1.25	553.9

5 Numerical Results

We present several numerical examples to demonstrate the numerical performance of our ILR on unstructured grids. Here the comparison is made with standard scalar limiter (SSL) [cf. (14)] and relaxed scalar limiter (RSL) [cf. (15)].

5.1 Linear Advection Equations

We first consider the linear advection equation

$$\frac{\partial u}{\partial t} + \boldsymbol{a} \cdot \nabla u = 0, \tag{38}$$

with constant wave velocity $\boldsymbol{a} = (1, 2)$. The boundary conditions are periodic. The mesh is generated by a global refinement on an irregular Delaunay triangulation, with the ratio of maximum to minimum size being 1.47. The Lax-Friedrichs flux (4) is adopted as the numerical flux and CFL number is 0.4.

Double sine wave: The initial profile is given by the double sine wave function [8,19]

$$u_0(x, y) = \sin(2\pi x) \sin(2\pi y),$$

on the domain $[0, 1] \times [0, 1]$. Errors in $\mathbb{L}^1$ and $\mathbb{L}^\infty$ norms for the solution at $t = 1$ as well as the CPU time are shown in Table 1. The results of ILR and RSL almost exhibit second-order accuracy, in contrast with the first-order accuracy of SSL. Although ILR seems slightly more time-consuming than SSL or RSL, it is still very efficient. In fact, the average number of iterations is approximately two. Moreover, ILR is free from the computation of predicted gradient. As a result, there is almost no extra cost in solving LP problems. Figure 7 shows

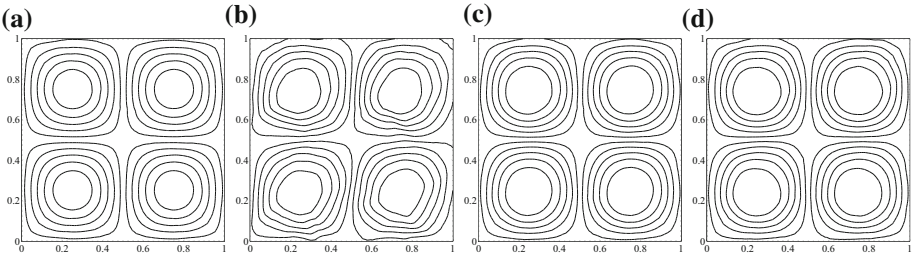

Fig. 7 Contours of the double sine wave at $t = 1$ with 3968 cells. **a** Unlimited. **b** SSL. **c** RSL. (d) ILR

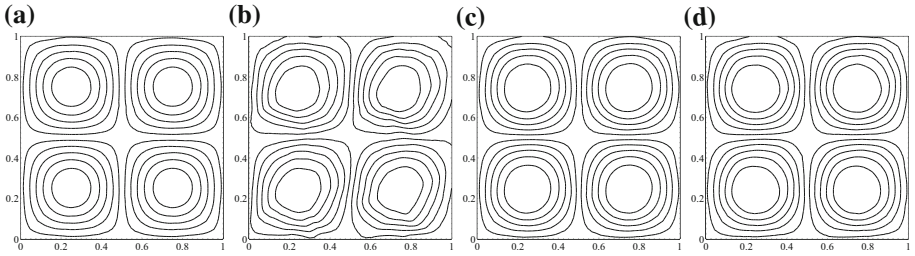

Fig. 8 Contours of the square wave at $t = 1$ with 3968 cells. **a** Unlimited. **b** SSL. **c** RSL. **d** ILR

Table 2 Global maximum and minimum values for the advection of square wave at $t = 1$

	3968 cells		15,872 cells	
	Min	Max	Min	Max
Unlimited	−0.164	1.2307	−0.185	1.2449
SSL	3.21e−5	0.9836	5.11e−9	0.9995
RSL	4.43e−7	0.9993	1.42e−12	1.0000
ILR	7.52e−7	0.9988	8.54e−12	1.0000

the contour lines for the double sine wave with 3968 cells. Compared to SSL, both ILR and RSL are less sensitive to the grid orientation.

Square wave: Instead of a smooth function, we use the following square wave for initial profile [6]

$$u_0(x, y) = \begin{cases} 1, & \text{if } 0.25 < x, y < 0.75, \\ 0, & \text{otherwise.} \end{cases}$$

Figure 8 shows the advection of square wave at $t = 1$. In Table 2 we examine the global maximum and minimum values on two levels of grids. One may see that ILR is, again similar as RSL, less dissipative than SSL, and resolves the discontinuity without spurious oscillations.

5.2 Solid Body Rotation

Another test to assess the capacity of the scheme to preserve the shape is the solid body rotation [22]. Consider the advection (38) with velocity $\boldsymbol{a} = (-(y - 0.5), x - 0.5)$ in the domain $[0, 1] \times [0, 1]$. The initial profile consists of three geometry shapes. Each shape

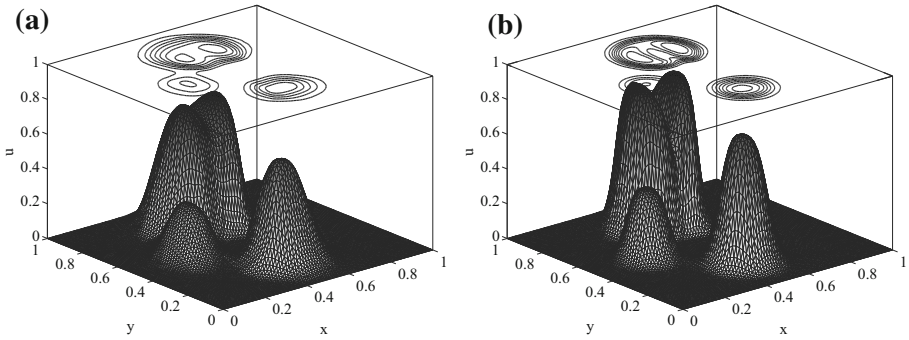

Fig. 9 Contours of the solid body rotation at $t = 2\pi$ with 15,872 cells. **a** SSL. **b** ILR

is located within the circle of radius $r_0 = 0.15$ whose center is (x_0, y_0). Moreover, let $r(x, y) = \frac{1}{r_0}\sqrt{(x - x_0)^2 + (y - y_0)^2}$ be the normalized distance. The slotted cylinder is centered at $(x_0, y_0) = (0.5, 0.75)$ and

$$u_0(x, y) = \begin{cases} 1, & \text{if } |x - x_0| \geq 0.025 \text{ or } y \geq 0.85, \\ 0, & \text{otherwise.} \end{cases}$$

The sharp cone is centered at $(x_0, y_0) = (0.5, 0.25)$ and

$$u_0(x, y) = 1 - r(x, y).$$

The smooth hump is centered at $(x_0, y_0) = (0.25, 0.5)$ and

$$u_0(x, y) = \frac{1 + \cos(\pi r(x, y))}{4}.$$

For the rest of the domain, the initial value is zero. The homogeneous Dirichlet boundary conditions are prescribed. The snapshots presented in Fig. 9 show the shape of the solution on a Delaunay mesh with 15,872 cells at $t = 2\pi$, which corresponds to one full rotation. Due to excessive numerical dissipation of SSL, the slotted cylinder is significantly smeared and distorted. On the other hand, the result of ILR almost preserves the shape of slotted cylinder without much distortion.

5.3 Euler Equations

For system of conservation laws, we apply the reconstruction directly to the conservative variables, mainly for the purposes of speed and simplicity, though in many aspects the characteristic limiting seems to be the most natural choice [8]. For Euler equations, the HLL flux is adopted as the numerical flux.

5.3.1 Shock Tube Problems

These tests are to examine the capability to resolve various linear and non-linear waves on unstructured grids. The computational domain is $[0, 1] \times [0, 0.1]$ with an irregular triangulation of 101 nodes in the x-direction and 11 nodes in the y-direction on the boundary. Initially the discontinuity is located at $x = 0.5$ with the prescribed left and right states as follows

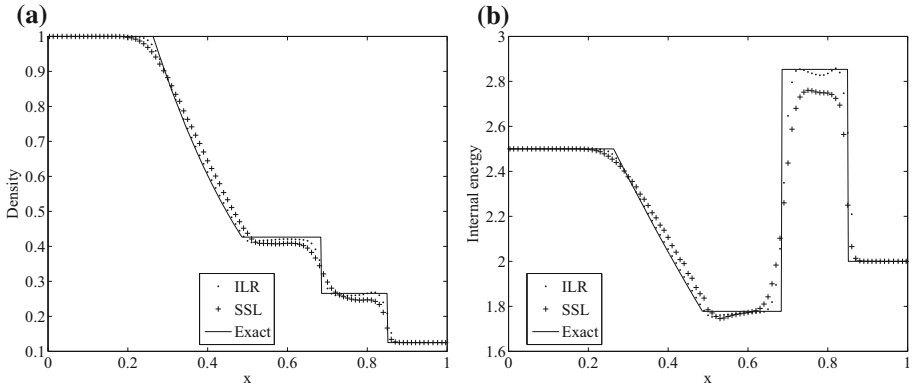

Fig. 10 Sod's problem: density and internal energy at $t = 0.2$

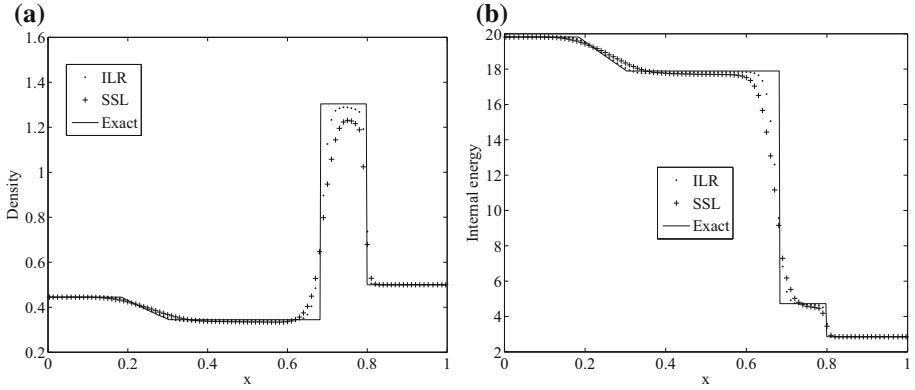

Fig. 11 Lax's problem: density and internal energy at $t = 0.12$

Sod's problem

$$(\rho_{\mathrm{L}}, u_{\mathrm{L}}, v_{\mathrm{L}}, p_{\mathrm{L}}) = (1, 0, 0, 1),$$
$$(\rho_{\mathrm{R}}, u_{\mathrm{R}}, v_{\mathrm{R}}, p_{\mathrm{R}}) = (0.125, 0, 0, 0.1).$$

Lax's problem

$$(\rho_{\mathrm{L}}, u_{\mathrm{L}}, v_{\mathrm{L}}, p_{\mathrm{L}}) = (0.445, 0.698, 0, 3.528),$$
$$(\rho_{\mathrm{R}}, u_{\mathrm{R}}, v_{\mathrm{R}}, p_{\mathrm{R}}) = (0.5, 0, 0, 0.571).$$

123 problem

$$(\rho_{\mathrm{L}}, u_{\mathrm{L}}, v_{\mathrm{L}}, p_{\mathrm{L}}) = (1, -2, 0, 0.4),$$
$$(\rho_{\mathrm{R}}, u_{\mathrm{R}}, v_{\mathrm{R}}, p_{\mathrm{R}}) = (1, 2, 0, 0.4).$$

Figures 10, 11 and 12 show the density and internal energy distributions for these test cases, which confirm the better performance of ILR. For 123 problem, the result of ILR captures expansion waves more accurately than SSL, and reduces the non-physical peak that appears in the middle of the domain considerably.

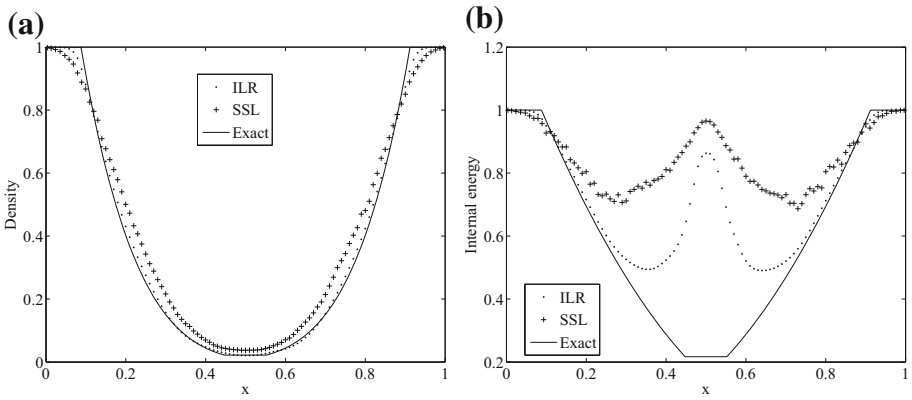

Fig. 12 123 problem: density and internal energy at $t = 0.15$

5.3.2 Isentropic Vortex Evolution

This test is to assess the performance of the proposed schemes for evolution of a two-dimensional inviscid isentropic vortex in a free stream [9]. The exact solution is simply a passive advection of the initial vortex with mean velocity. Assuming the free stream conditions $\rho_\infty = p_\infty = T_\infty = 1$ and $(u_\infty, v_\infty) = (1, 1)$, the initial condition is given by the mean flow field perturbed by

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

where the vortex strength $\beta = 5.0$ and adiabatic index $\gamma = 1.4$. The shifted coordinates $(\overline{x}, \overline{y}) = (x - x_0, y - y_0)$, where $(x_0, y_0) = (5, 5)$ is the coordinates of the vortex center initially and $r^2 = \overline{x}^2 + \overline{y}^2$. The entire flow field is required to be isentropic so, for an ideal gas, $p/\rho^\gamma = 1$, and consequently, the resulting state is given by

$$\rho = T^{1/(\gamma-1)} = (T_\infty + \delta T)^{1/(\gamma-1)} = \left[1 - \frac{(\gamma-1)\beta^2}{8\gamma\pi^2} e^{1-r^2}\right]^{1/(\gamma-1)},$$

$$\rho u = \rho(u_\infty + \delta u) = \rho\left[1 - \frac{\beta}{2\pi} e^{(1-r^2)/2}\overline{y}\right],$$

$$\rho v = \rho(v_\infty + \delta v) = \rho\left[1 + \frac{\beta}{2\pi} e^{(1-r^2)/2}\overline{x}\right],$$

$$p = \rho^\gamma \quad \text{and} \quad E = \frac{p}{\gamma-1} + \frac{1}{2}\rho(u^2 + v^2).$$

The computational domain is $[0, 10] \times [0, 10]$ where the periodic boundary condition is applied. The vortex moves in the right-up direction with the free stream, and returns back to the initial location at every time interval $\Delta T = 10$.

Figure 13 shows the density contours on an irregular Delaunay mesh with 15,872 cells. Due to the numerical diffusion of the scalar limiter, the vortex shape is significantly distorted. However, ILR almost keeps the initial vortex shape. Figure 14 gives the comparison of density distribution across the vortex center. Again ILR exhibits less dissipation than SSL.

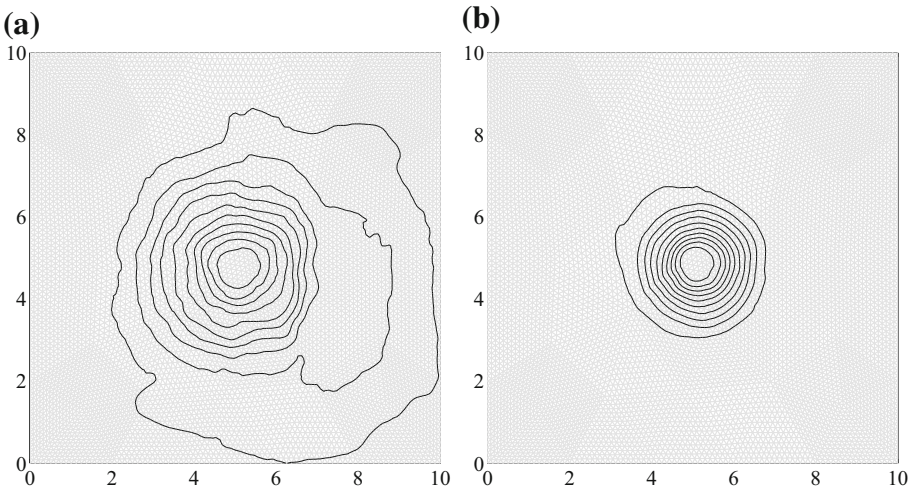

Fig. 13 Density contours of the isentropic vortex at $t = 30$ with 15,872 cells. **a** SSL. **b** ILR

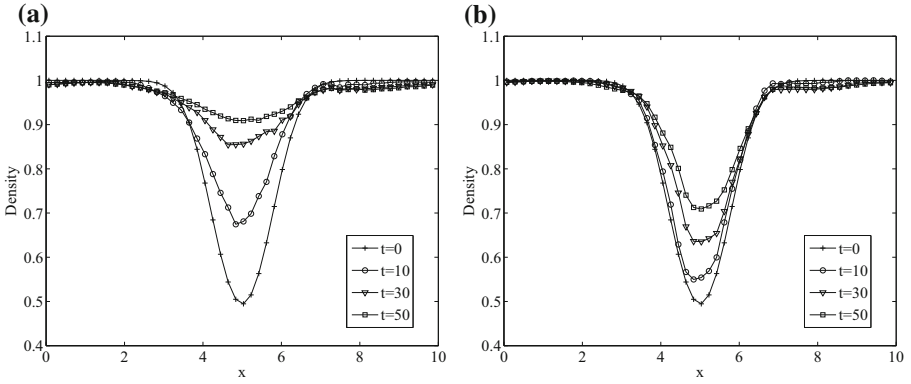

Fig. 14 Density distributions across the vortex center line at $t = 0$, 10, 30 and 50. **a** SSL. **b** ILR

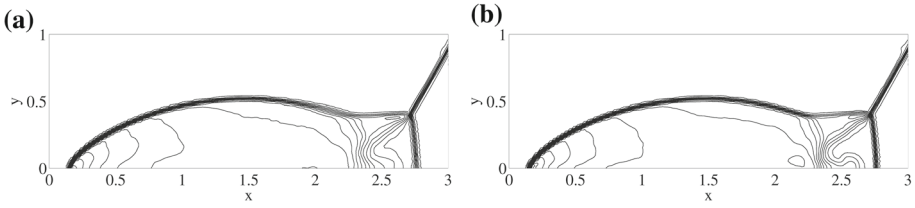

Fig. 15 Density contours for double Mach reflection with 23,367 cells. Thirty equally spaced contours from $\rho = 1.5$ to $\rho = 23.5$ are plotted. **a** 23,367 cells, SSL. **b** 23,367 cells, ILR

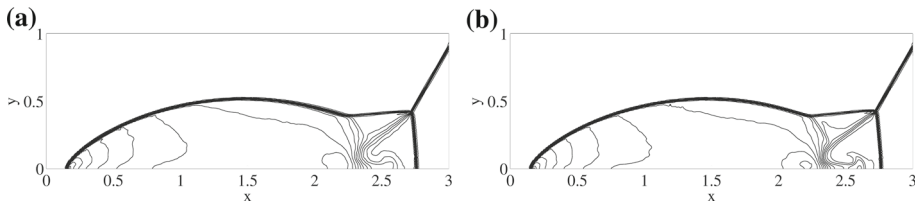

Fig. 16 Density contours of the double Mach reflection with 93,468 cells. Thirty equally spaced contours from $\rho = 1.5$ to $\rho = 23.5$ are plotted. **a** 93,468 cells, SSL. **b** 93,468 cells, ILR

5.3.3 Double Mach Reflection

The double Mach reflection problem comes from [23]. Results at $t = 0.2$ are plotted in Figs. 15 and 16 on two levels of Delaunay meshes. All pictures are the density contours with 30 equally spaced contour lines from 1.5 to 23.5. We can see that ILR captures the complicated flow structure under the Mach stem much better than SSL does, with a reasonable level of numerical diffusion.

In conclusion, ILR achieves better numerical resolution than SSL. While for RSL, the numerical dissipation may be on the similar level as ILR. Nevertheless, ILR satisfies the local maximum principle, while RSL only leads to a global maximum principle, which is much less sufficient to rule out spurious oscillations. Since we have not found any example in the literature where RSL produces numerical oscillations, the advantage of our method is mainly in the theoretical fold during the current stage.

6 Conclusion

We present an ILR under the framework of MUSCL-type schemes on unstructured grids. The reconstruction is formulated as an LP problem in every cell and contains no parameters at all. And the all-inequality simplex method used in this paper appears to be a very efficient algorithm to solve these LP problems. Moreover, the resulting schemes satisfy a local discrete maximum principle. Extension to higher-order reconstruction is possible, by increasing the number of variables and constraints in the LP problems. Additional constraints may be required to guarantee the local maximum principle. Fortunately, our framework is general enough to implement this with little extra cost.

Acknowledgments The authors appreciate the financial supports provided by the National Natural Science Foundation of China (Grant Numbers 91330205 and 11325102).

Appendix: All-Inequality Simplex Method

Consider a general LP problem in the all-equality form (20). The Lagrangian of (20) is

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{c}^\top \boldsymbol{x} + \boldsymbol{\lambda}^\top (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) = \boldsymbol{b}^\top \boldsymbol{\lambda} + \boldsymbol{x}^\top (\boldsymbol{c} - \boldsymbol{A}^\top \boldsymbol{\lambda}),$$

and the dual problem of (20) is

$$\min \boldsymbol{b}^\top \boldsymbol{\lambda} \quad \text{s.t. } \boldsymbol{A}^\top \boldsymbol{\lambda} = \boldsymbol{c} \text{ and } \boldsymbol{\lambda} \geq 0. \tag{39}$$

By the *weak duality theorem*, one knows that if $\boldsymbol{x}$ is feasible for (20) and $\boldsymbol{\lambda}$ is feasible for (39), then

$$\boldsymbol{c}^\top \boldsymbol{x} \leq \boldsymbol{b}^\top \boldsymbol{\lambda}.$$

Moreover, $\boldsymbol{x}$ is optimal for (20) provided that the equality above holds. Now we start from a vertex, and go from vertex to vertex until this equality holds. The iteration takes the form

$$\boldsymbol{x} := \boldsymbol{x} + \alpha \boldsymbol{p},$$

where the step length α and the searching direction $\boldsymbol{p}$ need to be determined at each step.

Denote by $\boldsymbol{a}_i^\top$ the i-th row of $\boldsymbol{A}$. The *active matrix* $\boldsymbol{M}$ consists of d rows of the matrix $\boldsymbol{A}$, indicating that corresponding constraints at the current searching point $\boldsymbol{x}$ are active. The active matrix $\boldsymbol{M}$ needs to be non-singular.

Suppose that at a given step the active Lagrange multipliers $\tilde{\boldsymbol{\lambda}}$ satisfy

$$\boldsymbol{M}^\top \tilde{\boldsymbol{\lambda}} = \boldsymbol{c} \text{ and } \tilde{\boldsymbol{\lambda}} \geq 0,$$

then $\boldsymbol{c}^\top \boldsymbol{x} = \boldsymbol{b}^\top \boldsymbol{\lambda}$ with all the inactive Lagrange multipliers vanishing, and we end up with an optimal point $\boldsymbol{x}$. Otherwise, we remove the constraint j such that the component $\lambda_j < 0$ and determine the searching direction $\boldsymbol{p}$ through

$$\boldsymbol{M}\boldsymbol{p} = -\boldsymbol{e}_j,$$

where $\boldsymbol{e}_j$ is the j-th coordinate vector.

Assuming that the point in the next step $\boldsymbol{x} + \alpha_i \boldsymbol{p}$ satisfies the constraint i, then we have

$$\boldsymbol{a}_i^\top (\boldsymbol{x} + \alpha_i \boldsymbol{p}) = b_i,$$

which yields the step length with regard to i to be

$$\alpha_i = \frac{b_i - \boldsymbol{a}_i^\top \boldsymbol{x}}{\boldsymbol{a}_i^\top \boldsymbol{p}}.$$

Since $b_i - \boldsymbol{a}_i^\top \boldsymbol{x} \geq 0$, it suffices to consider those constraints satisfying $\boldsymbol{a}_i^\top \boldsymbol{p} > 0$. The actual step length α is then taken as the minimum of all the estimated step lengths

$$\alpha = \min_{i \in D} \{\alpha_i\}.$$

Finally we choose an index k satisfying $\alpha_k = \alpha$ and update the active matrix by replacing j-th row of $\boldsymbol{M}$ with k-th row of $\boldsymbol{A}$. So far we finish one loop. The whole algorithm is illustrated in Algorithm 1. Note that we use the negation of both the searching direction $\boldsymbol{p}$ and step length α.

Algorithm 1 All-inequality simplex method

1: Initialize the point $\boldsymbol{x} \in \mathbb{R}^d$ and active matrix $\boldsymbol{M} \in \mathbb{R}^{d\times d}$.
2: Calculate the Lagrange multiplier $\boldsymbol{\lambda} \in \mathbb{R}^d$ by solving $\boldsymbol{M}^\top \boldsymbol{\lambda} = \boldsymbol{c}$.
3: **if** $\boldsymbol{\lambda} \geq 0$, **then**
4: STOP. The point $\boldsymbol{x}$ is optimal.
5: **else**
6: Set $j := \underset{j}{\operatorname{argmin}}\, \lambda_j$.
7: **end if**
8: Calculate the searching direction $\boldsymbol{p}$ from $\boldsymbol{Mp} = \boldsymbol{e}_j$.
9: Find all indices of decreasing constraints along $\boldsymbol{p}$ as

$$D := \{i : \boldsymbol{a}_i^\top \boldsymbol{p} < 0\}.$$

10: For all $i \in D$, calculate the step length α_i one can take before violating constraint i

$$\alpha_i := \frac{b_i - \boldsymbol{a}_i^\top \boldsymbol{x}}{\boldsymbol{a}_i^\top \boldsymbol{p}}, \quad \forall i \in D.$$

11: Find the step length as large as possible without violating any constraints

$$k := \underset{i\in D}{\operatorname{argmax}}\, \alpha_i.$$

12: Update the point

$$\boldsymbol{x} := \boldsymbol{x} + \alpha_k \boldsymbol{p}.$$

13: Replace j-th row of $\boldsymbol{M}$ with k-th row of $\boldsymbol{A}$.
14: Go to Step 2.

References

1. van Leer, B.: Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method. J. Comput. Phys. **32**(1), 101–136 (1979)
2. Goodman, J.B., LeVeque, R.J.: On the accuracy of stable schemes for 2d scalar conversation laws. Math. Comput. **45**(171), 15–21 (1985)
3. Spekreijse, S.: Multigrid solution of monotone second-order discretizations of hyperbolic conservation laws. Math. Comput. **49**(179), 135–155 (1987)
4. Barth, T.J., Jespersen, D.C.: The design and application of upwind schemes on unstructured meshes. In: 27th AIAA Aerospace Sciences Meeting, vol. 366, Jan 1989
5. Durlofsky, L.J., Engquist, B., Osher, S.: Triangle based adaptive stencils for the solution of hyperbolic conservation laws. J. Comput. Phys. **98**(1), 64–73 (1992)
6. Batten, P., Lambert, C., Causon, D.M.: Positively conservative high-resolution convection schemes for unstructured elements. Int. J. Numer. Methods Eng. **39**(11), 1821–1838 (1996)
7. Liu, X.D.: A maximum principle satisfying modification of triangle based adapative stencils for the solution of scalar hyperbolic conservation laws. SIAM J. Numer. Anal. **30**(3), 701–716 (1993)
8. Hubbard, M.E.: Multidimensional slope limiters for MUSCL-type finite volume schemes on unstructured grids. J. Comput. Phys. **155**(1), 54–74 (1999)
9. Park, J.S., Yoon, S.H., Kim, C.: Multi-dimensional limiting process for hyperbolic conservation laws on unstructured grids. J. Comput. Phys. **229**(3), 788–812 (2010)
10. Barth, T.J., Frederickson, P.O.: Higher order solution of the Euler equations on unstructured grids using quadratic reconstruction. In: 28th AIAA Aerospace Sciences Meeting, vol. 13, Jan 1990
11. Abgrall, R.: On essentially non-oscillatory schemes on unstructured meshes: analysis and implementation. J. Comput. Phys. **114**(1), 45–58 (1994)
12. Friedrich, O.: Weighted essentially non-oscillatory schemes for the interpolation of mean values on unstructured grids. J. Comput. Phys. **144**(1), 194–212 (1998)

13. Hu, C., Shu, C.W.: Weighted essentially non-oscillatory schemes on triangular meshes. J. Comput. Phys. **150**(1), 97–127 (1999)
14. Liu, Y., Zhang, Y.T.: A robust reconstruction for unstructured WENO schemes. J. Sci. Comput. **54**(2–3), 603–621 (2013)
15. Jawahar, P., Kamath, H.: A high-resolution procedure for Euler and Navier–Stokes computations on unstructured grids. J. Comput. Phys. **164**(1), 165–203 (2000)
16. Li, W., Ren, Y.X., Lei, G., Luo, H.: The multi-dimensional limiters for solving hyperbolic conservation laws on unstructured grids. J. Comput. Phys. **230**(21), 7775–7795 (2011)
17. Berger, M., Aftosmis, M.J., Murman, S.M.: Analysis of slope limiters on irregular grids. In: 43rd AIAA Aerospace Sciences Meeting, vol. 490, May 2005
18. Buffard, T., Clain, S.: Monoslope and multislope MUSCL methods for unstructured meshes. J. Comput. Phys. **229**(10), 3745–3776 (2010)
19. May, S., Berger, M.: Two-dimensional slope limiters for finite volume schemes on non-coordinate-aligned meshes. SIAM J. Sci. Comput. **35**(5), A2163–A2187 (2013)
20. Shu, C.W., Osher, S.: Efficient implementation of essentially non-oscillatory shock-capturing schemes. J. Comput. Phys. **77**(2), 439–471 (1988)
21. Barth, T.J., Ohlberger, M.: Finite volume methods: foundation and analysis. In: Encyclopedia of Computational Mechanics. Citeseer, (2004)
22. LeVeque, R.J.: High-resolution conservative algorithms for advection in incompressible flow. SIAM J. Numer. Anal. **33**(2), 627–665 (1996)
23. Woodward, P., Colella, P.: The numerical simulation of two-dimensional fluid flow with strong shocks. J. Comput. Phys. **54**(1), 115–173 (1984)