Journal of Computational Physics 229 (2010) 3745–3776

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

Monoslope and multislope MUSCL methods for unstructured meshes

Thierry Buffard [a,*], Stéphane Clain [b]

[a] Université Clermont-Ferrand II, Laboratoire de Mathématiques, UMR CNRS 6620, 63177 Aubière cedex, France
[b] Institut de Mathématiques de Toulouse, UMR CNRS 5219, 118 route de Narbonne, 31062 Toulouse cedex, France

ARTICLE INFO

Article history:
Received 28 November 2008
Received in revised form 14 January 2010
Accepted 18 January 2010
Available online 28 January 2010

Keywords:
High-order scheme
Finite volume
Multislope method
Unstructured mesh
Conservation laws

ABSTRACT

We present new MUSCL techniques associated with cell-centered finite volume method on triangular meshes. The first reconstruction consists in calculating one vectorial slope per control volume by a minimization procedure with respect to a prescribed stability condition. The second technique we propose is based on the computation of three scalar slopes per triangle (one per edge) still respecting some stability condition. The resulting algorithm provides a very simple scheme which is extensible to higher dimensional problems. Numerical approximations have been performed to obtain the convergence order for the advection scalar problem whereas we treat a nonlinear vectorial example, namely the Euler system, to show the capacity of the new MUSCL technique to deal with more complex situations.

1. Introduction

Large numerical simulations in industrial framework require efficient but rather simple numerical methods to face the modelling complexity while making easier the implementation. Flexibility is also required to quickly adapt the computation code to new conditions and models. High-resolution methods such as ENO, WENO or Discontinuous Galerkin methods provide very good accuracy. However, the MUSCL technique is more popular in the industrial context due to its natural simplicity and adaptation capacity to respond to modelling evolutions and complexifications.

Monotone Upstream Scheme for Conservation Law technique (MUSCL technique) has been introduced by Van Leer [27] for one-dimensional hyperbolic problems. The main idea is a piecewise linear reconstruction of the solution to achieve higher accurate schemes still preserving the stability: the maximum principle or the Total Variation Diminishing (TVD) property for instance. Initially elaborated for one-dimensional scalar problems, the MUSCL technique combined with a conservative scheme had to preserve the Total Variation of the solution. To this end, slopes are limited to prevent spurious oscillations or overshooting of the numerical approximations [25] and numerous limiters have been proposed [23] in the one-dimensional framework to achieve high-resolution TVD schemes. A first extension of the MUSCL technique to higher dimensions has been proposed using structured meshes where the MUSCL procedure is applied in each direction [8] but the generalization of the Total Variation Diminishing constraint for higher dimensional geometries makes the scheme to be a first-order method [13]. To get around this negative result, a new class of positive schemes have been introduced [24] which ensures a local maximum principle. The concept of Local Extremum Diminishing was then developed by Jameson [15] where he generalizes the notion of incremental scheme with non-negative coefficients for the multi-dimensional situation. For scalar hyperbolic problem, maximum principle naturally derives from the incremental expression and extensions in the Finite Element context have been proposed by Kuzmin and Turek [18].

* Corresponding author.
E-mail addresses: Thierry.Buffard@univ-bpclermont.fr (T. Buffard), clain@mip.ups-tlse.fr (S. Clain).

0021-9991/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2010.01.026

An other important point that the reconstruction technique has to address concerns the numerical approximations of hyperbolic system solutions. For the Euler system, density and pressure have to be non-negative to be physically admissible and the shallow-water system requires a non-negative height of water. Numerical approximations have to preserve the density and pressure positivity and specific numerical flux have been designed for this purpose [11]. Extension of the positivity preservation criteria both for second-order finite volume schemes have been also developed [22].

To handle more flexible refinements and allow discretization of complex bounded domains, new MUSCL methods for unstructured meshes have been considered based either on the cell-centered representation [16,9,2] or on the vertex centered representation [5,6]. A linear function is constructed on each element using a gradient prediction which should be limited to prevent oscillations of the numerical solutions [10] (see also [12,19,20] for a mathematical study of the high-order schemes).

The classical MUSCL technique consists of two steps. First, a predicted gradient is computed for each element of the mesh using the neighbouring values. Then the gradient is modified to respect some Maximum Principle or Total Variation Diminishing constraint and provide a vectorial slope on the element. New values are therefore computed on each edge of the element using the linear reconstruction. Finally, an approximation of the flux crossing the interface is performed by employing the two reconstructed values situated on both sides of the edge combined with a monotone numerical flux function. To avoid the predictor–corrector algorithm and obtain some optimal reconstruction, we propose to build the vectorial slope on each element by minimizing a convex functional under stability constraints. The idea is to optimize the slope while respecting the Maximum principle or the Total Variation Diminishing property. We intend in this way to produce the best gradient approximation which respects the stability constraint.

The MUSCL method presented above will be referred to as **monoslope method** since the reconstructed values are obtained using the same vectorial slope on each element. We also introduce a new class of MUSCL method named **multislope method** where we use specific scalar slope for each interface. For a given element, we consider a set of normalized vectors and we use the neighbouring values to compute the scalar slopes representing an approximation of the directional derivatives. The slopes are modified afterwards to respect some stability constraint and finally, the reconstructed values are computed on each edge using the corrected slopes. The main advantage of the method is that we only deal with one-dimensional situations and, as we shall show in the following sections, the scalar slopes are very simple to compute even for higher dimensional geometries.

The remainder of the paper is organized as follows. In Section 2, we introduce the notations we shall use in the sequel to describe the finite volume process on triangular meshes for two-dimensional geometries and we review some classical MUSCL-type methods. In particular, we give a precise description of the Maximum Principle domain and the Total Variation Diminishing domain that we employ to keep the stability condition. Section 3 is devoted to a new monoslope MUSCL method while we describe the multislope MUSCL technique in Section 4. Numerical results are presented for the linear advection problem and the Euler system in Section 5.

2. Second-order monoslope MUSCL method

To illustrate the MUSCL reconstruction, we here introduce the classical advection problem but more complex problems such as nonlinear vectorial systems can of course be considered.

Let $\Omega \subset \mathbb{R}^2$, be a polygonal open bounded set of $\mathbb{R}^2, T > 0$. We denote by $\mathbf{V}(t,x)$ a given $\mathbb{R}^2$ vectorial valued function defined on $Q_T = [0,T] \times \overline{\Omega}$. For $t \in [0,T]$, we set

$$\Gamma^-(t) = \{x \in \partial\Omega; \mathbf{V}(t,x) \cdot \mathbf{n}(x) < 0\}, \quad \Gamma^+(t) = \{x \in \partial\Omega; \mathbf{V}(t,x) \cdot \mathbf{n}(x) \geqslant 0\},$$

with $x = (x_1, x_2)$ a generic point of Ω and $\mathbf{n}$ the outwards normal on the boundary $\partial\Omega$.

We consider the advection problem: find $U(t,x)$ a real valued function defined on Q_T such that

$$\begin{aligned} &\partial_t U + \nabla \cdot (\mathbf{V}U) = 0 \quad \text{in }]0,T[\times\Omega, \\ &U(t=0,\cdot) = U_0(\cdot) \quad \text{in } \Omega, \\ &U(t,\cdot) = U_b(t,\cdot) \quad \text{in } \Gamma^-(t),\ t \in]0,T], \end{aligned}$$

where U_0 and U_b are given functions.

To deal with the numerical approximation, we introduce the following ingredients (see Fig. 1). $\mathcal{T}_h$ is a discretization of Ω with triangles K_i of centroid $\mathbf{B}_i, i = 1, \ldots, N$ where N is the number of mesh elements. For a given i, $v(i)$ represents the index set of the common edge elements $K_j \in \mathcal{T}_h, j \in v(i)$ where $S_{ij} = \overline{K}_j \cap \overline{K}_i$ stands for the common edge with midpoint $\mathbf{M}_{ij}$.

We assume furthermore that the mesh satisfies the following hypothesis (see Fig. 2):

$$(\mathcal{H}) \quad \begin{cases} \text{For any } K_i \in \mathcal{T}_h \text{ such that } |v(i)| = 3, \text{ point } \mathbf{B}_i \text{ is strictly} \\ \text{inside the convex set defined by the points } \mathbf{B}_j, j \in v(i). \end{cases}$$

Remark 1. Hypothesis $(\mathcal{H})$ yields that any two of the three vectors $B_iB_j, j \in v(i)$ defines a basis of $\mathbb{R}^2$. Such a property is essential to define the monoslope MUSCL method and it is less restrictive than Hypothesis $(\mathcal{H})$. Nevertheless, the multislope

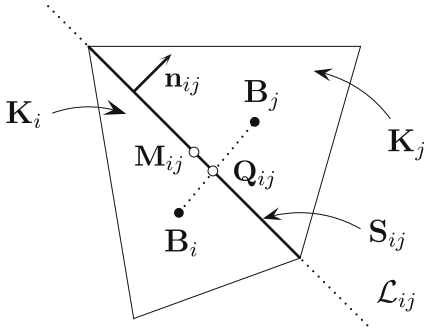


Fig. 1. Notations and conventions of the mesh elements and edges.

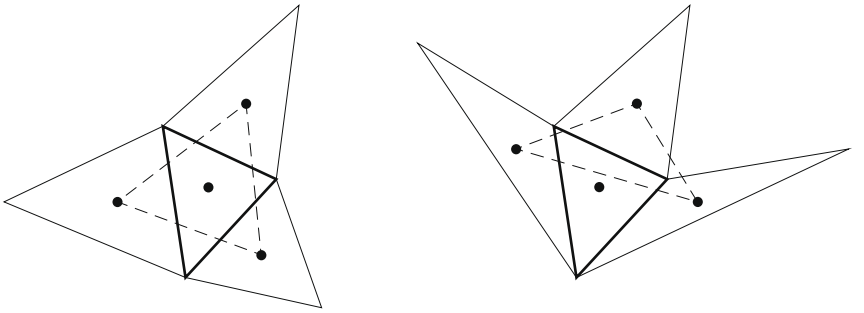


Fig. 2. Configuration satisfying the hypothesis $(\mathcal{H})$ (on left). Configuration which not satisfies the hypothesis $(\mathcal{H})$ (on right).

MUSCL method requires that the B_i point has positive barycentric coordinates with respect to the three other points $B_j, j \in v(i)$. This last point motivates the introduction of Hypothesis $(\mathcal{H})$.

All the Delaunay meshes we have used to perform numerical tests satisfy Hypothesis $(\mathcal{H})$ but we do not manage to relate the Delaunay condition: the triangle circumcircle formed by three nodes does not contain the other nodes.

If $\mathcal{L}_{ij}$ represents the line containing the edge S_{ij}, point $\mathbf{Q}_{ij}$ is defined as the intersection between the segment $[\mathbf{B}_i, \mathbf{B}_j]$ and the line $\mathcal{L}_{ij}$. Note that $\mathbf{Q}_{ij}$ does not belong *a priori* to S_{ij} but only to $\mathcal{L}_{ij}$. For a given edge S_{ij}, $\mathbf{n}_{ij}$ represents the outward normal of K_i pointing to K_j and $\mathbf{n}_{ji} = -\mathbf{n}_{ij}$.

We will use a cell-centered finite volume method where control volumes are the triangles. The sequence $(t^n)_n$ defines a time discretization of $[0, T]$ with $t^{n+1} = t^n + \Delta t$. Let U_i^n stand for an approximation of the mean value of U at time t^n on the element K_i. The conservative first-order finite volume formulation is given by

$$|K_i|\ U_i^{n+1} = |K_i|\ U_i^n - \Delta t \sum_{j \in v(i)} |S_{ij}|\ F_{ij}\left(U_i^n, U_j^n\right), \tag{1}$$

where $F_{ij}(U_i, U_j)$ is a numerical flux from K_i to K_j at interface S_{ij}.

For the advection case, classical numerical flux functions are the Lax–Friedrichs flux or the upwind flux:

$$F_{ij}^{LF}\left(U_i^n, U_j^n\right) = \frac{1}{2}\left(\mathbf{V}(t^n, \mathbf{B}_i) \cdot \mathbf{n}_{ij} U_i^n + \mathbf{V}(t^n, \mathbf{B}_j) \cdot \mathbf{n}_{ij} U_j^n\right) - \lambda\left(U_j^n - U_i^n\right),$$
$$F_{ij}^{upwind}\left(U_i^n, U_j^n\right) = [\mathbf{V}(t^n, \mathbf{M}_{ij}) \cdot \mathbf{n}_{ij}]^+ U_i^n + [\mathbf{V}(t^n, \mathbf{M}_{ij}) \cdot \mathbf{n}_{ij}]^- U_j^n,$$

where $[\cdot]^+$ represents the positive part and λ is a positive constant to guarantee the scheme stability.

2.1. Classical MUSCL methods

First-order schemes give a poor approximation and induce high viscosity effect. A second-order scheme provides a better approximation and manages to reduce the viscous smoothing effect in the vicinity of the shocks.

The popular techniques consist in a local linear reconstruction (see [3,12,26]). Assuming that a constant piecewise approximation $U_h^n = (U_i^n)_i$ of U at time t^n is known, we construct a new linear piecewise approximation $\widetilde{U}_h^n$ in the following way

$$\widetilde{U}_h^n(\mathbf{X}) = U_i^n + \mathbf{a}_i \cdot \mathbf{B}_i\mathbf{X}, \quad \mathbf{X} \in K_i, \tag{2}$$

where $\mathbf{B}_i\mathbf{X}$ stands for the vector $\mathbf{X} - \mathbf{B}_i$, $\mathbf{a}_i \in \mathbb{R}^2$ is the vectorial slope we have to construct, $\mathbf{a}_i \cdot \mathbf{B}_i\mathbf{X}$ is the inner product between $\mathbf{B}_i\mathbf{X}$ and $\mathbf{a}_i$.

Remark that such a linear reconstruction satisfies conservation property

$$\int_{K_i} \widetilde{U}_h^n(\mathbf{X}) dX = |K_i| U_i^n,$$

since the centroid point $\mathbf{B}_i$ is chosen as reference point.

Given a point $\mathbf{X}_{ij}$ on the common edge S_{ij}, we set

$$U_{ij}^n = U_i^n + \mathbf{a}_i \cdot \mathbf{B}_i\mathbf{X}_{ij}, \quad U_{ji}^n = U_j^n + \mathbf{a}_j \cdot \mathbf{B}_j\mathbf{X}_{ij}. \tag{3}$$

We classify this kind of reconstruction as **monoslope method** since we use one slope per element: the same slope $\mathbf{a}_i$ produces all values U_{ij}^n, $j \in v(i)$.

Two useful choices for point $\mathbf{X}_{ij}$ are $\mathbf{Q}_{ij}$ or $\mathbf{M}_{ij}$ (see Fig. 1). The first one is natural from a geometrical point of view since it corresponds to the linear interpolation between $\mathbf{B}_i$ and $\mathbf{B}_j$ whereas the second one is natural from the integration point of view since the numerical integration with the midpoint rule is exact for linear functions along the edge S_{ij}.

To obtain a second-order method, we then substitute the numerical flux $F_{ij}\left(U_i^n, U_j^n\right)$ by $F_{ij}\left(U_{ij}^n, U_{ji}^n\right)$ in relation (1) and obtain:

$$|K_i|\ U_i^{n+1} = |K_i|\ U_i^n - \Delta t \sum_{j \in v(i)} |S_{ij}|\ F_{ij}\left(U_{ij}^n, U_{ji}^n\right). \tag{4}$$

Several slope evaluations have been proposed (see [12,14,3] for an exhaustive list), where two leading requirements have to be satisfied:

(C1) the linearly reconstructed function $\widetilde{U}_h$ satisfies $\widetilde{U}_h = U$ if the function U is linear. In this paper, this property will be referred to as linear consistency of the reconstruction;
(C2) the reconstruction has to respect a maximum principle to avoid overshooting leading to a discrepancy of the numerical approximation.

Remark 2. The case where an element shares a common side with the boundary is treated using ghost cells. Indeed, let us assume that K has a side S on the boundary, we construct a fictitious element $\widetilde{K}$ which shares the same side (by symmetry for example). We then prescribe the boundary condition on $\widetilde{K}$ and we are back in the situation where the element is strictly inside the domain.

2.1.1. Gradient methods

Denote by $K_{j_1}, K_{j_2}, K_{j_3}$ the three adjacent triangles of K_i. We consider the three following hyperplans in the x_1, x_2, U space: hyperplane $\pi_{i,1}$ is defined by the points $\mathbf{B}_i, \mathbf{B}_{j_2}, \mathbf{B}_{j_3}$ with elevations U_i, U_{j_2}, U_{j_3} and $\pi_{i,2}, \pi_{i,3}$ are obtained in the same way. The hyperplane $\pi_{1,2,3}$ is defined by the points $\mathbf{B}_{j_1}, \mathbf{B}_{j_2}, \mathbf{B}_{j_3}$ with elevations $U_{j_1}, U_{j_2}, U_{j_3}$ (see Fig. 3).

For example, $\pi_{i,1}$ is given by equation

$$(u - U_i^n) = \mathbf{G}_{i,1} \cdot \mathbf{B}_i\mathbf{X},$$

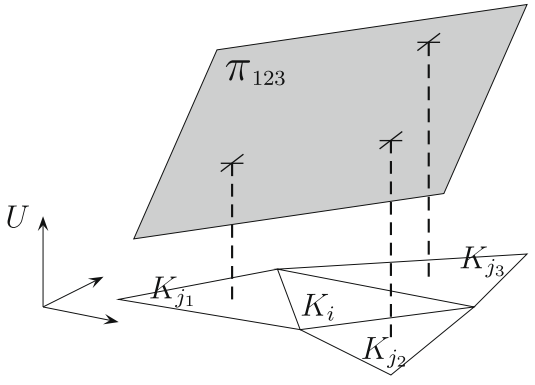


Fig. 3. Plane $\pi_{1,2,3}$ representation.

where $\mathbf{G}_{i,1} \in \mathbb{R}^2$ while $\pi_{1,2,3}$ is given by

$$\left(u - U_{j_1}^n\right) = \mathbf{G}_{1,2,3} \cdot \mathbf{B}_{j_1}\mathbf{X}.$$

A first choice consists in taking the slope $\mathbf{a}_i = \mathbf{G}_{1,2,3}$ and we obtain a linear consistent reconstruction. Other possible choices use a combination of $\mathbf{G}_{i,1}, \mathbf{G}_{i,2}, \mathbf{G}_{i,3}$, setting

$$\mathbf{a}_i = \sigma(\mathbf{G}_{i,1}, \mathbf{G}_{i,2}, \mathbf{G}_{i,3}).$$

The linear consistency is obtained if and only if $\mathbf{a} = \sigma(\mathbf{a}, \mathbf{a}, \mathbf{a})$ for all $\mathbf{a} \in \mathbb{R}^2$.

2.1.2. Minimization method

In Ref. [7], the authors consider the hyperplane minimizing the distance with the four points $(\mathbf{B}_i, U_i), (\mathbf{B}_j, U_j), j \in v(i)$. One has to seek a vector $\mathbf{G}_{LS}$ using a Least Square Method, that is to say which minimizes the functional

$$E(\mathbf{a}) = \sum_{j \in v(i)} \left(U_j^n - \left(U_i^n + \mathbf{a} \cdot \mathbf{B}_i\mathbf{B}_j\right)\right)^2. \tag{5}$$

Existence and uniqueness of the minimum is obvious since the functional is strictly convex.

Moreover, if U is linear, the four points lie in the same hyperplane and the minimum corresponds to the gradient of U, hence we get the linear consistency of the reconstruction.

2.2. The stability conditions

Let us consider two adjacent triangles K_i and K_j. To avoid numerical artefacts in the vicinity of large gradients (overshooting or spurious oscillations), one imposes that the reconstructed values U_{ij} and U_{ji} on S_{ij} satisfy some stability property. To this end, we introduce the following conditions:

(1) The L^∞ stability condition (Maximum Principle constraint or MP constraint):

$$\min\left(U_i^n, U_j^n\right) \leqslant U_{ij}^n, U_{ji}^n \leqslant \max\left(U_i^n, U_j^n\right). \tag{6}$$

(2) The Total Variation Diminishing-like condition (TVD constraint):

$$\text{if } U_i^n \leqslant U_j^n \text{ then } U_i^n \leqslant U_{ij}^n \leqslant U_{ji}^n \leqslant U_j^n. \tag{7}$$

The last condition is named TVD constraint since the property (7) implies the preservation of the BV norm between the initial piecewise constant function and its piecewise linear reconstruction. Moreover, relation (7) implies relation (6) so the Total Variation Diminishing-like condition is a subcase of the L^∞ stability condition.

The slope $\mathbf{a}_i$ provided by one of the above methods does not *a priori* satisfy the stability condition. We impose the stability by multiplying the slope by a limiter $\phi_i \in \mathbb{R}$ such that the values U_{ij}^n and U_{ji}^n obtained with the new slope $\tilde{\mathbf{a}}_i = \phi_i\mathbf{a}_i$ satisfy one of the two stability conditions. In particular, if $\phi_i = 0$, we find again the first-order scheme.

In the case of a linear solution, a predicted slope process which satisfies condition (C1) provides a slope equal to the gradient of the linear function. In this particular case, the limiting procedure has no impact since the predicted slope respects the two stability constraints and one has $\phi_i = 1$. Therefore, it is natural to choose the highest value of $\phi_i \in [0, 1]$ such that the reconstruction satisfies a prescribed stability condition.

2.2.1. The Maximum Principle domain

For a given element K_i, we define the Maximum Principle domain (MP domain) as

$$MP_i = \left\{\mathbf{a} \in \mathbb{R}^2;\ \min\left(U_j^n - U_i^n, 0\right) \leqslant \mathbf{a} \cdot \mathbf{B}_i\mathbf{Q}_{ij} \leqslant \max\left(U_j^n - U_i^n, 0\right),\ j \in v(i)\right\}.$$

If $\mathbf{a}_i \in MP_i$ then $U_{ij}^n = U_i^n + \mathbf{a}_i \cdot \mathbf{B}_i\mathbf{Q}_{ij}$ satisfies stability condition (6) and the converse is also true.

For the sake of simplicity, we introduce a new set of vectors

$$\mathbf{s}_k = \operatorname{sgn}\left(U_{j_k}^n - U_i^n\right)\mathbf{B}_i\mathbf{Q}_{ij_k}, \quad k = 1, 2, 3,$$

where $\operatorname{sgn}(x) = \begin{cases} 1 & \text{for } x \geqslant 0, \\ -1 & \text{for } x < 0. \end{cases}$

The MP_i region is now simply given by

$$MP_i = \{\mathbf{a} \in \mathbb{R}^2;\ 0 \leqslant \mathbf{a}.\mathbf{s}_k \leqslant \gamma_k,\ k = 1, 2, 3\},$$

with $\gamma_k = \left|U_i^n - U_{j_k}^n\right|, k = 1, 2, 3.$

We require that sgn(0) is non-zero in order to extend the equivalence:

$$\mathbf{a}\cdot\mathbf{B}_i\mathbf{Q}_{ij_k}=0 \Longleftrightarrow \mathbf{a}\cdot\mathbf{s}_k=0,$$

to the particular situation $U^n_{j_k}=U^n_i$.

Hypothesis $(\mathcal{H})$ implies that any couple of the three vectors $\mathbf{s}_k, k=1,2,3$ defined a basis of the $\mathbb{R}^2$ space. Therefore we can express one vector from the two others and we have the following unique expansions with non-zero coefficients:

$$\mathbf{s}_1=\alpha_{12}\mathbf{s}_2+\alpha_{13}\mathbf{s}_3, \tag{8}$$
$$\mathbf{s}_2=\alpha_{21}\mathbf{s}_1+\alpha_{23}\mathbf{s}_3, \tag{9}$$
$$\mathbf{s}_3=\alpha_{31}\mathbf{s}_1+\alpha_{32}\mathbf{s}_2. \tag{10}$$

A simple computation gives the following proposition.

Proposition 3. *We have the relations*

$$\alpha_{lm}\ \alpha_{ml}=1, \tag{11}$$
$$\alpha_{lm}\ \alpha_{mk}=-\alpha_{lk} \tag{12}$$

for any circular permutation (l,m,k) of (1,2,3).

Proof. To check properties (11) and (12), let us consider the decomposition of $\mathbf{s}_l$

$$\mathbf{s}_l=\alpha_{lm}\mathbf{s}_m+\alpha_{lk}\mathbf{s}_k.$$

Thanks to hypothesis $(\mathcal{H})$, $\mathbf{s}_l$ is neither collinear to $\mathbf{s}_m$ nor to $\mathbf{s}_k$, hence coefficients α_{lm} and α_{lk} do not vanish. The relation can be rewritten

$$\mathbf{s}_m=\frac{1}{\alpha_{lm}}\mathbf{s}_l-\frac{\alpha_{lk}}{\alpha_{lm}}\mathbf{s}_k=\alpha_{ml}\mathbf{s}_l+\alpha_{mk}\mathbf{s}_k,$$

which gives relations (11) and (12) by identification thanks to the uniqueness of the decomposition. □

We deduce that MP_i domain is characterized only by coefficients α and γ. If at least two of the three γ coefficients vanish, we easily deduce $MP_i=\{(0,0)\}$. We now consider the other situations.

Proposition 4. *Assume that one coefficient, say γ_k, vanishes while the two others, say γ_l and γ_m, are not zero. Then we have*

$$MP_i=\{(0,0)\} \Longleftrightarrow \alpha_{lm}<0.$$

Proof. Let us first remark that if $\mathbf{a}\in\mathbb{R}^2$ with $\mathbf{a}\cdot\mathbf{s}_k=0$, then we have:

$$\mathbf{a}\cdot\mathbf{s}_l=\alpha_{lk}\mathbf{a}\cdot\mathbf{s}_k+\alpha_{lm}\mathbf{a}\cdot\mathbf{s}_m=\alpha_{lm}\mathbf{a}\cdot\mathbf{s}_m. \tag{13}$$

[⇐] Suppose that $\alpha_{lm}<0$ and let $\mathbf{a}\in MP_i$.
Since $\gamma_k=0$, relation (13) is satisfied. From condition $\mathbf{a}\in MP_i$, we have the relations $\mathbf{a}\cdot\mathbf{s}_l\geqslant 0$ and $\mathbf{a}\cdot\mathbf{s}_m\geqslant 0$. It follows that $\mathbf{a}\cdot\mathbf{s}_m=\mathbf{a}\cdot\mathbf{s}_l=0$ since we have $\alpha_{lm}<0$. Hence $\mathbf{a}=(0,0)$.
[⇒] Conversely, suppose that $\alpha_{lm}\geqslant 0$.
Since all the coefficients are non-vanishing, we have $\alpha_{lm}>0$. We shall now construct a non-zero vector of MP_i. To this end, consider $\mathbf{a}\in\mathbb{R}^2$ such that $\mathbf{a}\cdot\mathbf{s}_k=0$ and $\mathbf{a}\cdot\mathbf{s}_l=\min(\gamma_l,\alpha_{lm}\gamma_m)\in]0,\gamma_l]$. We obtain a non-zero vector which satisfies $0<\mathbf{a}\cdot\mathbf{s}_m=\frac{1}{\alpha_{lm}}\ \mathbf{a}\cdot\mathbf{s}_l\leqslant\gamma_m$, then $\mathbf{a}\in MP_i$. □

Remark 5. If only one of the γ_k is zero, the MP_i domain is reduced to the null vector or to a segment.

Proposition 6. *Assume that all the coefficients γ_k are positive, $k=1,2,3$. Then the following assertions are equivalent:*

(i) $\alpha_{12}<0$ and $\alpha_{13}<0$.
(ii) $\alpha_{21}<0$ and $\alpha_{23}<0$.
(iii) $\alpha_{31}<0$ and $\alpha_{32}<0$.
(iv) $MP_i=\{(0,0)\}$.

Proof. Equivalences between (i), (ii) and (iii) derive from relations (11) and (12). It remains to prove the equivalence between (i) and (iv). To this end, let us assume that assertion (i) holds and let $\mathbf{a}\in MP_i$. One has

$$\mathbf{a}\cdot\mathbf{s}_1=\alpha_{12}\mathbf{a}\cdot\mathbf{s}_2+\alpha_{13}\mathbf{a}\cdot\mathbf{s}_3,\quad \text{with } \alpha_{12}\mathbf{a}\cdot\mathbf{s}_2\leqslant 0 \text{ and } \alpha_{13}\mathbf{a}\cdot\mathbf{s}_3\leqslant 0. \tag{14}$$

It follows that $\mathbf{a}\cdot\mathbf{s}_1\leqslant 0$, hence that $\mathbf{a}\cdot\mathbf{s}_1=0$ since $\mathbf{a}\cdot\mathbf{s}_1\geqslant 0$. Relation (14) now gives $\mathbf{a}\cdot\mathbf{s}_2=\mathbf{a}\cdot\mathbf{s}_3=0$ and we conclude that $\mathbf{a}$ is the null vector because $\mathbf{s}_1,\mathbf{s}_2$ is a basis.

Conversely, let us assume that (i) does not hold. We shall construct a non-zero vector $\mathbf{a}$ such that $\mathbf{a} \in MP_i$.
Since assertion (i) is wrong, we have $\alpha_{12} > 0$ or $\alpha_{13} > 0$.
Suppose $\alpha_{12} > 0$ for example. Let $\mathbf{a}$ be the vector of $\mathbb{R}^2$ such that $\mathbf{a} \cdot \mathbf{s}_3 = 0$ and $\mathbf{a} \cdot \mathbf{s}_2 = \min\left(\frac{\gamma_1}{\alpha_{12}}, \gamma_2\right) \in]0, \gamma_2]$. We obtain a non-zero vector which satisfies $0 < \mathbf{a} \cdot \mathbf{s}_1 = \alpha_{12}\, \mathbf{a} \cdot \mathbf{s}_2 \leqslant \gamma_1$, then $\mathbf{a} \in MP_i$. □

Under the same assumption as the above proposition, we have the following corollary using relation (12).

Corollary 7. *Assume that all the coefficients γ_k are positive, $k = 1, 2, 3$. Then the MP_i domain is not reduced to the null vector if and only if one of the three following assertions holds*

(i) $\alpha_{12} > 0$ and $\alpha_{13} > 0$,
(ii) $\alpha_{21} > 0$ and $\alpha_{23} > 0$,
(iii) $\alpha_{31} > 0$ and $\alpha_{32} > 0$.

Proof.

[⇒] We first assume that MP_i domain is not reduced to the null vector. From Proposition 6, we deduce that $\alpha_{12} \geqslant 0$ or $\alpha_{13} \geqslant 0$, hence $\alpha_{12} > 0$ or $\alpha_{13} > 0$ since the coefficients are non-zero. If both coefficients are positive, assertion (i) is right otherwise one of the two coefficients is negative (says $\alpha_{13} < 0$). From relations (11) and (12) we have $\alpha_{21} > 0$ and $\alpha_{23} > 0$ and assertion (ii) holds.
[⇐] Conversely, if for example $\alpha_{12} > 0$ and $\alpha_{13} > 0$ then Proposition 6 immediately implies that MP_i domain is not reduced to the null vector. □

When MP_i domain is not reduce to the null vector, one of the three assertions of Corollary 7 holds. In this case, we adopt the following convention:
Convention. We choose the local indexation such that $\alpha_{31} > 0$ and $\alpha_{32} > 0$.
The MP_i domain is a convex polygonal set (see Fig. 4) which consists in the intersection of the three bands limited by the lines

$$d_k = \{\mathbf{a} \in \mathbb{R}^2;\ \mathbf{a} \cdot \mathbf{s}_k = \gamma_k\}, \quad k = 1, 2, 3, \tag{15}$$

$$\delta_k = \{\mathbf{a} \in \mathbb{R}^2;\ \mathbf{a} \cdot \mathbf{s}_k = 0\}, \quad k = 1, 2, 3. \tag{16}$$

2.2.2. The slope limiter

Let $\mathbf{a}_i$ be a predicted gradient obtained, for example, by one of the methods presented in Section 2.1. The Maximum Principle constraint yields that $\mathbf{a}_i$ has to be in the MP_i domain. If not, we reduce the slope by a limiter $\phi_i \in [0, 1]$ such that $\tilde{\mathbf{a}}_i = \phi_i \mathbf{a}_i \in MP_i$. The most classical limiting procedure (see [12,3]) consists in constructing the three limiters

$$\phi_{i,k} = \begin{cases} \max\left(0, \frac{\gamma_k}{\mathbf{a}_i \cdot \mathbf{s}_k}\right) & \text{if } \mathbf{a}_i \cdot \mathbf{s}_k \neq 0, \\ 1 & \text{if } \mathbf{a}_i \cdot \mathbf{s}_k = 0. \end{cases} \tag{17}$$

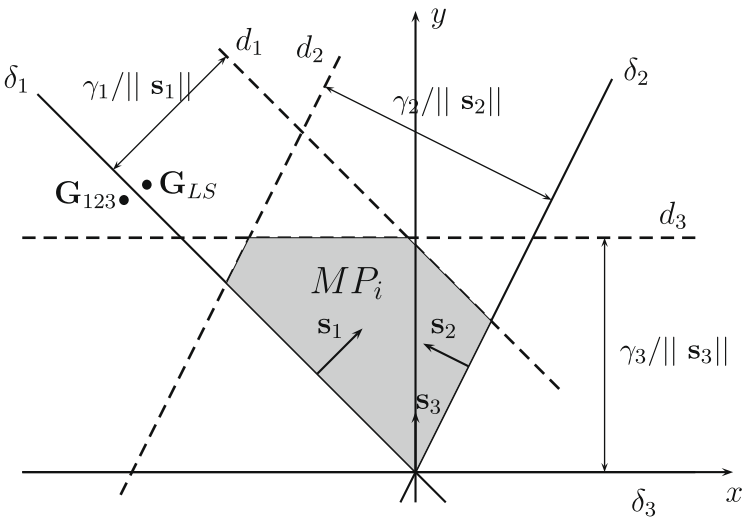


Fig. 4. Maximum principle domain.

Taking $\phi_i = \min(1, \phi_{i,1}, \phi_{i,2}, \phi_{i,3})$, we set $\widetilde{\mathbf{a}}_i = \phi_i \mathbf{a}_i \in MP_i$. If for one local subscript k, $\mathbf{a}_i \cdot \mathbf{s}_k < 0$, the limiter is zero and we obtain a first-order method. Numerical experiments indicate that such a phenomena often occurs resulting in a poor approximation accuracy [7,14]. For example, let us consider a configuration where the predicted slope $\mathbf{G}_{1,2,3}$ is on the left side of line δ_1 while $\mathbf{G}_{LS}$ stands on the right side (see Fig. 4). Applying the limiting procedure (17) yields that $\phi_i > 0$ if we choose $\mathbf{a}_i = \mathbf{G}_{LS}$ whereas $\phi_i = 0$ if we choose $\mathbf{a}_i = \mathbf{G}_{1,2,3}$. In the first case, the resulting slope provides a second-order scheme but the second situation reduces to a first-order scheme.

To avoid the discrepancy, some authors propose to limit the predicted gradient using a orthogonal projection of point $\mathbf{a}_i$ on the boundary of the MP_i domain (see [14]).

2.2.3. The TVD domain

We now consider the more restrictive *TVD* constraint (7) using the same framework introduced for the *MP* constraint. For a given edge S_{ij}, the *TVD* constraint involves the two slopes $\mathbf{a}_i$ and $\mathbf{a}_j$ which also depends on the neighbouring elements leading to a coupling between all the slopes. To avoid the complex interactions between the slopes, we introduce a more restrictive definition of the TVD constraint such that $\mathbf{a}_i$ is computed independently of the other slopes but only depends on the data of the three neighbouring elements. The requirement on slope $\mathbf{a}_i$ is that reconstructed values U_{ij}^n and U_{ji}^n (with $j \in v(i)$) have to satisfy

$$\text{if } U_i^n \leqslant U_j^n \text{ then } U_i^n \leqslant U_{ij}^n \leqslant U_{ij}^{ref} \leqslant U_{ji}^n \leqslant U_j^n, \tag{18}$$

where U_{ij}^{ref} is the reference value at point $\mathbf{Q}_{ij}$ defined by

$$U_{ij}^{ref} = U_i^n + \frac{|\mathbf{B}_i\mathbf{Q}_{ij}|}{|\mathbf{B}_i\mathbf{B}_j|}\left(U_j^n - U_i^n\right) = U_j^n + \frac{|\mathbf{B}_j\mathbf{Q}_{ij}|}{|\mathbf{B}_j\mathbf{B}_i|}\left(U_i^n - U_j^n\right) = U_{ji}^{ref}. \tag{19}$$

We define the TVD_i domain by

$$TVD_i = \left\{\mathbf{a} \in \mathbb{R}^2;\ \min\left(U_{ij}^{ref} - U_i^n, 0\right) \leqslant \mathbf{a}\cdot\mathbf{B}_i\mathbf{Q}_{ij} \leqslant \max\left(U_{ij}^{ref} - U_i^n, 0\right),\ j \in v(i)\right\}.$$

The TVD_i domain is also characterized by

$$TVD_i = \{\mathbf{a} \in \mathbb{R}^2;\ 0 \leqslant \mathbf{a}\cdot\mathbf{s}_k \leqslant \mu_k,\ k = 1,2,3\},$$

with $\mu_k = \left|U_i^n - U_{ij_k}^{ref}\right|, k = 1,2,3$.

The TVD_i domain is a convex polygonal set which consists in the intersection of the three bands limited by the lines

$$d_k = \{\mathbf{a} \in \mathbb{R}^2;\ \mathbf{a}\cdot\mathbf{s}_k = \mu_k\}, \quad k = 1,2,3, \tag{20}$$

$$\delta_k = \{\mathbf{a} \in \mathbb{R}^2;\ \mathbf{a}\cdot\mathbf{s}_k = 0\}, \quad k = 1,2,3. \tag{21}$$

To conclude the section, notice that

$$\mu_k = \frac{|\mathbf{B}_i\mathbf{Q}_{i,j_k}|}{|\mathbf{B}_i\mathbf{B}_{j_k}|}\gamma_k \leqslant \gamma_k,$$

hence, we deduce that the TVD_i domain is a subset of the MP_i domain and the limiting techniques presented for the MP_i domain can directly be adapted to the TVD_i domain using μ_k in place of γ_k.

3. A new monoslope method

All the second-order schemes presented above are developed following two steps: first we compute a predicted slope and, secondly, we use a limiting procedure. We propose here a new method where we build the slope in only one procedure in which we optimize the slope under the *MP* constraint or the *TVD* constraint.

As we state in the convention presented in Section 2.2.1, we choose the local indexation such that the coefficients α_{31} and α_{32} are positive.

3.1. Minimization under the TVD constraint

We only present the construction of the optimized slope respecting the *TVD* constraint. The construction of the optimized slope under the *MP* constraint can also been considered and adapted.

3.1.1. Problem formulation

Let us consider a triangular control volume K_i. It is clear that if U is a linear function defined by $U(\mathbf{X}) = U_0 + \mathbf{L}\cdot\mathbf{X}$, then $U(\mathbf{Q}_{ij}) = U(\mathbf{B}_i) + \mathbf{L}\cdot\mathbf{B}_i\mathbf{Q}_{ij} = U_{ij}^{ref}$ for all $j \in v(i)$. For the general case, we wish to obtain a slope $\mathbf{a}_i$ on K_i for which deviations $U_i + \mathbf{a}_i\cdot\mathbf{B}_i\mathbf{Q}_{ij} - U_{ij}^{ref}$ are as close as possible to 0. Moreover, the slope should provide a reconstruction which respects the stability condition.

We then compute the slope by using a least square method under the *TVD* constraint on element K_i and the optimization problem reads:

find the slope $\tilde{\mathbf{a}}_i$ minimizing the functional

$$E_i(\mathbf{a}) = \sum_{j\in v(i)} \left(U_{ij}^{ref} - (U_i + \mathbf{a}\cdot\mathbf{B}_i\mathbf{Q}_{ij})\right)^2 \quad \text{with } \mathbf{a}\in TVD_i. \tag{22}$$

Using the notations introduced in Section 2.2, we can rewrite the minimization problem as

$$E_i(\mathbf{a}) = \sum_{k=1,2,3} (\mu_k - \mathbf{a}\cdot\mathbf{s}_k)^2 \tag{23}$$

$$\text{with} \quad 0 \leqslant \mathbf{a}\cdot\mathbf{s}_k \leqslant \mu_k, \quad k = 1,2,3. \tag{24}$$

Remark 8. We can also consider another minimization problem using the minimization functional (5). If we add now the *MP* constraint (see [4]), the optimization problem then reads: find the slope $\tilde{\mathbf{a}}_i$ minimizing the functional

$$E_i(\mathbf{a}) = \sum_{j\in v(i)} (U_j - (U_i + \mathbf{a}\cdot\mathbf{B}_i\mathbf{B}_j))^2 \quad \text{with } \mathbf{a}\in MP_i. \tag{25}$$

Note that problem (22) is not equivalent to problem (25).

Since the functional (23) is strictly convex and the domain defined by (24) is convex and bounded, we get the existence and the uniqueness of the minimum $\tilde{\mathbf{a}}_i$. With the slope in hand, we build the new predicted values at any given collocation point $\mathbf{X}_{ij}$

$$U_{ij} = U_i + \tilde{\mathbf{a}}_i\cdot\mathbf{B}_i\mathbf{X}_{ij}, \quad j\in v(i). \tag{26}$$

3.1.2. Computation of the optimal slope

We are now interested in finding the minimum $\tilde{\mathbf{a}}$ of the functional (23) under constraints (24). To simplify the notations, we skip the index i in this subsection.

We first note that $\tilde{\mathbf{a}}$ is obviously the null vector if $TVD = \{(0,0)\}$. Note that if $\mu_3 = 0$, we have $TVD = \{(0,0)\}$ by the indexation convention.

From now on we make the assumption that μ_1, μ_2 and μ_3 are positive. The case where $\mu_1 = 0$ or $\mu_2 = 0$ will also be treated further.

Proposition 9. *Let $\bar{\mathbf{a}}$ be the minimum of $E(\mathbf{a})$ without constraint then $\bar{\mathbf{a}}$ is inside the triangle T_{123} formed by the three lines $d_k, k = 1,2,3$ defined by relation (20). In particular, if the triangle is not reduced to a point, $\bar{\mathbf{a}}$ is strictly inside the triangle.*

Proof. Let us set $\mathbf{G}_1 = d_2 \cap d_3$ (see Fig. 5). We then have

$$\mathbf{G}_1\cdot\mathbf{s}_2 = \mu_2, \quad \mathbf{G}_1\cdot\mathbf{s}_3 = \mu_3.$$

We define in the same way $\mathbf{G}_2 = d_1 \cap d_3$ and $\mathbf{G}_3 = d_1 \cap d_2$ satisfying

$$\mathbf{G}_2\cdot\mathbf{s}_1 = \mu_1, \quad \mathbf{G}_2\cdot\mathbf{s}_3 = \mu_3, \quad \mathbf{G}_3\cdot\mathbf{s}_1 = \mu_1, \quad \mathbf{G}_3\cdot\mathbf{s}_2 = \mu_2.$$

If $\mathbf{G}_1, \mathbf{G}_2$ and $\mathbf{G}_3$ belong to the same line then hypothesis $(\mathcal{H})$ yields $\mathbf{G}_1 = \mathbf{G}_2 = \mathbf{G}_3 = \mathbf{G}$, thus $\bar{\mathbf{a}} = \mathbf{G}$ since $E(\bar{\mathbf{a}}) = 0$ in this exceptional case.

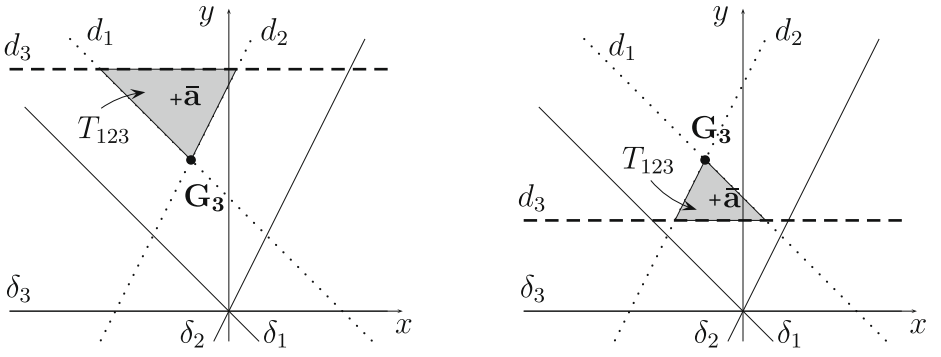


Fig. 5. The triangle T_{123} is above point $\mathbf{G}_3$ (left). The triangle T_{123} is under point $\mathbf{G}_3$ (right).

We now assume that the three points define a non-degenerated triangle T_{123} and we seek $\bar{\mathbf{a}} = \lambda_1\mathbf{G}_1 + \lambda_2\mathbf{G}_2 + \lambda_3\mathbf{G}_3$ using the barycentric coordinates with $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

Existence and uniqueness of the minimum $\bar{\mathbf{a}}$ is clear since $E(\mathbf{a})$ is strictly convex and $\bar{\mathbf{a}}$ has to satisfy the linear system

$$\sum_{k=1,2,3} (\mu_k - \bar{\mathbf{a}} \cdot \mathbf{s}_k)\mathbf{s}_k = 0. \tag{27}$$

Using the barycentric coordinates property and the definition of $\mathbf{G}_k$, we get

$$\sum_{k=1,2,3} \lambda_k(\mu_k - \mathbf{G}_k \cdot \mathbf{s}_k)\mathbf{s}_k = 0.$$

The inner product between the last relation and vector $\mathbf{G}_1$ gives

$$\lambda_1(\mu_1 - \mathbf{G}_1 \cdot \mathbf{s}_1)\mathbf{G}_1 \cdot \mathbf{s}_1 + \lambda_2(\mu_2 - \mathbf{G}_2 \cdot \mathbf{s}_2)\mu_2 + \lambda_3(\mu_3 - \mathbf{G}_3 \cdot \mathbf{s}_3)\mu_3 = 0.$$

Using also vectors $\mathbf{G}_2$ and $\mathbf{G}_3$, we obtain

$$\lambda_1(\mu_1 - \mathbf{G}_1 \cdot \mathbf{s}_1)\mu_1 + \lambda_2(\mu_2 - \mathbf{G}_2 \cdot \mathbf{s}_2)\mathbf{G}_2 \cdot \mathbf{s}_2 + \lambda_3(\mu_3 - \mathbf{G}_3 \cdot \mathbf{s}_3)\mu_3 = 0,$$
$$\lambda_1(\mu_1 - \mathbf{G}_1 \cdot \mathbf{s}_1)\mu_1 + \lambda_2(\mu_2 - \mathbf{G}_2 \cdot \mathbf{s}_2)\mu_2 + \lambda_3(\mu_3 - \mathbf{G}_3 \cdot \mathbf{s}_3)\mathbf{G}_3 \cdot \mathbf{s}_3 = 0.$$

From the three relations, we deduce

$$\lambda_1(\mu_1 - \mathbf{G}_1 \cdot \mathbf{s}_1)^2 = \lambda_2(\mu_2 - \mathbf{G}_2 \cdot \mathbf{s}_2)^2 = \lambda_3(\mu_3 - \mathbf{G}_3 \cdot \mathbf{s}_3)^2. \tag{28}$$

Since triangle T_{123} is not reduced to a point, the quantities $(\mu_k - \mathbf{G}_k \cdot \mathbf{s}_k)^2$ are positive and thus the coordinates λ_k have the same sign. Moreover, the condition $\lambda_1 + \lambda_2 + \lambda_3 = 1$ yields that $\lambda_k > 0$, hence $\bar{\mathbf{a}}$ is strictly inside the triangle. □

Remark 10. An explicit calculation of coefficients λ_k provides an expression independent of U_i and U_j:

$$\lambda_1 = \frac{\alpha_{31}^2}{1 + \alpha_{31}^2 + \alpha_{32}^2}, \quad \lambda_2 = \frac{\alpha_{32}^2}{1 + \alpha_{31}^2 + \alpha_{32}^2}, \quad \lambda_3 = \frac{1}{1 + \alpha_{31}^2 + \alpha_{32}^2}.$$

Remark 11. The exceptional situation where the triangle T_{123} is reduced to a point corresponds to the case where the four points $(\mathbf{B}_i, U_i^n), \left(\mathbf{B}_j, U_j^n\right), j = j_1, j_2, j_3$ lie in the same hyperplane of the (x_1, x_2, U) space. In this case, the optimal slope $\tilde{\mathbf{a}}$ under constraint corresponds to the optimal slope $\bar{\mathbf{a}}$ without constraint and the reconstruction is consistent for linear functions.

Corollary 12. *If triangle T_{123} is not reduced to a point, the minimum without constraint $\bar{\mathbf{a}}$ does not satisfy the TVD constraint. Furthermore the minimum $\tilde{\mathbf{a}}$ with constraint satisfies, at least, one of the six constraints: $\tilde{\mathbf{a}} \cdot \mathbf{s}_\mathbf{k} = 0$ or $\tilde{\mathbf{a}} \cdot \mathbf{s}_\mathbf{k} = \mu_k$ with $k = 1, 2, 3$, i.e. $\tilde{\mathbf{a}} \in \partial TVD$.*

Proof. We notice that $TVD \cap T_{123}$ is reduced to the point $\mathbf{G}_3$ or is included in the segment $[\mathbf{G}_1, \mathbf{G}_2]$ whether d_3 is above ($\mu_3 \geqslant \mathbf{G}_3 \cdot \mathbf{s}_3$, see Fig. 5 *left*) or under ($\mu_3 \leqslant \mathbf{G}_3 \cdot \mathbf{s}_3$, see Fig. 5 *right*) the point $\mathbf{G}_3$. Since $\bar{\mathbf{a}}$ is strictly inside T_{123}, we conclude that $\bar{\mathbf{a}} \notin TVD$.

Finally, if $\tilde{\mathbf{a}}$ is strictly inside the *TVD* domain, then no constraint is saturated and we have $\nabla E(\tilde{\mathbf{a}}) = 0$ thus $\tilde{\mathbf{a}} = \bar{\mathbf{a}}$ which is not possible since $\bar{\mathbf{a}} \notin TVD$. □

Proposition 13. *The minimum under constraint $\tilde{\mathbf{a}}$ belongs to d_1, d_2 or d_3.*

Proof. Let us denote by $\mathbf{s}_1^\perp$ the orthogonal normalized vector to $\mathbf{s}_1$ such that $\mathbf{s}_1^\perp \cdot \mathbf{s}_3 > 0$. Then the half-line on line δ_1 with (0,0) as endpoint which touches the *TVD* domain is characterized by $\lambda\mathbf{s}_1^\perp$ with $\lambda \geqslant 0$ and we have

$$E(\lambda\mathbf{s}_1^\perp) = \mu_1^2 + (\mu_2 - \lambda\mathbf{s}_1^\perp \cdot \mathbf{s}_2)^2 + (\mu_3 - \lambda\mathbf{s}_1^\perp \cdot \mathbf{s}_3)^2.$$

Moreover, we have $\mathbf{s}_1^\perp \cdot \mathbf{s}_3 > 0$ by definition and we also have $\mathbf{s}_1^\perp \cdot \mathbf{s}_2 > 0$ since α_{32} is positive. We deduce that E decreases as λ increases from 0 till $\lambda\mathbf{s}_1^\perp$ reaches the first of the two intersection points $\delta_1 \cap d_2$ or $\delta_1 \cap d_3$. In conclusion, the minimum $\tilde{\mathbf{a}}$ will belong to $\delta_1 \cap TVD$ only if it is equal to intersection point $\delta_1 \cap d_2$ or $\delta_1 \cap d_3$. The same arguments hold using vectors $\mathbf{s}_2^\perp$ and $\tilde{\mathbf{a}}$ will belong to $\delta_2 \cap TVD$ only if it is equal to intersection point $\delta_2 \cap d_1$ or $\delta_2 \cap d_3$. □

Remark 14. If $\mu_1 = 0$ and $\mu_2\mu_3 \neq 0$, the *TVD* domain is reduced to a segment on line δ_1. The previous proof shows in that case that the minimum $\tilde{\mathbf{a}}$ is the intersection point $\delta_1 \cap d_2$ or the intersection point $\delta_1 \cap d_3$. The case $\mu_2 = 0$ and $\mu_1\mu_3 \neq 0$ is similar.

We precise the position of the minimum with constraint in the next proposition.

Proposition 15. *Let $\tilde{\mathbf{a}}$ be the minimum with the TVD constraint. Then we have the following alternative:*

(i) If d_3 is above the intersection point $\mathbf{G}_3$ between d_1 and d_2 ($\mathbf{G}_3 \cdot \mathbf{s}_3 \leqslant \mu_3$) then $\tilde{\mathbf{a}} = \mathbf{G}_3$.

(ii) If d_3 is under the intersection point $\mathbf{G}_3$ between d_1 and d_2 ($\mathbf{G}_3 \cdot \mathbf{s}_3 > \mu_3$) then $\tilde{\mathbf{a}}$ belongs to d_3.

Proof. We first study the situation for the line d_1 where we prove that $\tilde{\mathbf{a}}$ does not belong to d_1 except point $\mathbf{G}_3$. The same argument holds for line d_2. Since $\mathbf{G}_3 = d_1 \cap d_2$, we have $\mathbf{G}_3 \cdot \mathbf{s}_1 = \mu_1$ and $\mathbf{G}_3 \cdot \mathbf{s}_2 = \mu_2$. Consider now a point $\mathbf{a}$ on the segment $d_1 \cap TVD$ (supposed non-empty). Using the parametrization

$$\mathbf{a} = \mathbf{G}_3 + \lambda \mathbf{s}_1^{\perp}, \tag{29}$$

we obtain

$$E(\mathbf{a}) = E(\mathbf{G}_3 + \lambda \mathbf{s}_1^{\perp}) = F(\lambda) = \lambda^2 (\mathbf{s}_1^{\perp} \cdot \mathbf{s}_2)^2 + (\mu_3 - \mathbf{G}_3 \cdot \mathbf{s}_3 - \lambda \mathbf{s}_1^{\perp} \cdot \mathbf{s}_3)^2. \tag{30}$$

We get a convex parabolic curve and the minimum is obtained for λ_0 given by

$$\lambda_0 = \frac{(\mu_3 - \mathbf{G}_3 \cdot \mathbf{s}_3)\ \mathbf{s}_1^{\perp} \cdot \mathbf{s}_3}{(\mathbf{s}_1^{\perp} \cdot \mathbf{s}_3)^2 + (\mathbf{s}_1^{\perp} \cdot \mathbf{s}_2)^2}.$$

Due to the orientation convention $\mathbf{s}_1^{\perp} \cdot \mathbf{s}_3 > 0$, any point $\mathbf{a} \in d_1 \cap TVD$ satisfies $\lambda \leqslant 0$.

Case (i) If d_3 is above $\mathbf{G}_3$, *i.e.* $\mathbf{G}_3 \cdot \mathbf{s}_3 \leqslant \mu_3$ then $\lambda_0 \geqslant 0$ and we deduce that the minimum on the segment $d_1 \cap TVD$ is obtained at point $\lambda = 0$ since λ has to be non-positive.

Case (ii) If d_3 is under $\mathbf{G}_3$, *i.e.* $\mathbf{G}_3 \cdot \mathbf{s}_3 > \mu_3$ then $\lambda_0 < 0$. On the other hand, we can write the point $\mathbf{G}_2 = d_1 \cap d_3$ in the form $\mathbf{G}_2 = \mathbf{G}_3 + v\mathbf{s}_1^{\perp}$ and relation $\mathbf{G}_2 \cdot \mathbf{s}_3 = \mu_3$ leads to

$$v = \frac{(\mu_3 - \mathbf{G}_3 \cdot \mathbf{s}_3)}{\mathbf{s}_1^{\perp} \cdot \mathbf{s}_3} < 0.$$

We obtain

$$\frac{\lambda_0}{v} = \frac{(\mathbf{s}_1^{\perp} \cdot \mathbf{s}_3)^2}{(\mathbf{s}_1^{\perp} \cdot \mathbf{s}_3)^2 + (\mathbf{s}_1^{\perp} \cdot \mathbf{s}_2)^2} < 1.$$

We conclude that $v < \lambda_0 < 0$ and the minimum of E on the segment $d_1 \cap TVD$ occurs for $\lambda = v$, thus the minimum belongs to d_3. □

The first situation corresponds to the choice $\tilde{\mathbf{a}} = \mathbf{G}_3$ whereas the following proposition completes the second assertion.

Proposition 16. *Assume that d_3 is under point $\mathbf{G}_3$. Line d_3 is parted into three pieces: d_3^c is the segment $d_3 \cap TVD$, d_3^- is the left part of d_3 with respect to d_3^c while d_3^+ is the right part of d_3 with respect to d_3^c (see Fig. 6).*

Let $\hat{\mathbf{a}}$ be the minimum of functional $E(\mathbf{a})$ under the constraint $\mathbf{a} \in d_3$. We have the following situations:

- case 1: if $\hat{\mathbf{a}} \in d_3^-$ then $\tilde{\mathbf{a}}$ is the left bound of segment d_3^c.
- case 2: if $\hat{\mathbf{a}} \in d_3^c$ then $\tilde{\mathbf{a}} = \hat{\mathbf{a}}$,
- case 3: if $\hat{\mathbf{a}} \in d_3^+$ then $\tilde{\mathbf{a}}$ is the right bound of segment d_3^c.

Proof. Let us denote by $\mathbf{s}_3^{\perp}$ the orthogonal normalized vector to $\mathbf{s}_3$ such that $\mathbf{s}_3^{\perp}$ goes from the left to the right (see Fig. 6). Line d_3 can be parameterized by using a free parameter λ

$$\mathbf{a} = \widehat{\mathbf{a}} + \lambda \mathbf{s}_3^{\perp}. \tag{31}$$

On line d_3, functional E is then given by

$$E(\mathbf{a}) = E(\widehat{\mathbf{a}} + \lambda \mathbf{s}_3^{\perp}) = F(\lambda),$$

where $F(\lambda)$ is a parabolic function, strictly decreasing for $\lambda < 0$ and strictly increasing for $\lambda > 0$.

If $\hat{\mathbf{a}} \in d_3^c$, then $\hat{\mathbf{a}} \in TVD$. Since Proposition 15 says that $\tilde{\mathbf{a}} \in d_3$, we deduce that $\tilde{\mathbf{a}} = \hat{\mathbf{a}}$.

If $\hat{\mathbf{a}} \in d_3^+$, function $F(\lambda)$ is a decreasing function for λ such that $\mathbf{a} \in TVD$. Therefore, the minimum is obtained at the right bound of segment d_3^c. On the contrary, if $\hat{\mathbf{a}} \in d_3^-$, function $F(\lambda)$ is an increasing function for λ such that $\mathbf{a} \in TVD$. Therefore, the minimum is obtained at the left bound of segment d_3^c. □

We conclude this subsection by a summary of optimal slope computation.

- If at least two of the three μ coefficients vanish, then $\hat{\mathbf{a}} = (0,0)$.
- Assume that all μ coefficients are positive.
  From Proposition 6, $\hat{\mathbf{a}} = (0,0)$ if and only if $\alpha_{21} < 0, \alpha_{31} < 0$ and $\alpha_{32} < 0$.Otherwise, using the convention on the local indexation ($\alpha_{31} > 0$ and $\alpha_{32} > 0$), we derived the following Table from Propositions 15 and 16:

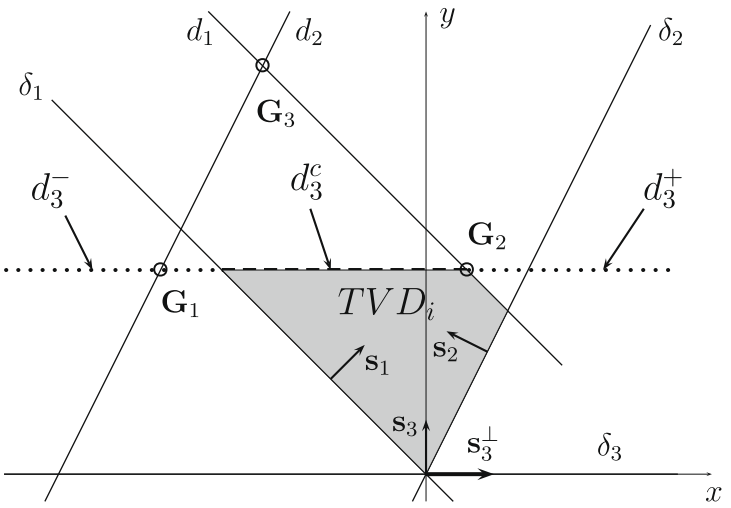


Fig. 6. Partition of the line d_3.

Cases	Optimal slope $\hat{\mathbf{a}}$
case 1: $\psi = \mathbf{G}_3 \cdot \mathbf{s}_3 - \mu_3 = \alpha_{31}\mu_1 + \alpha_{32}\mu_2 - \mu_3 \leqslant 0$	$\mathbf{G}_3$ such that $\mathbf{G}_3 \cdot s_1 = \mu_1$ and $\mathbf{G}_3 \cdot s_2 = \mu_2$
case 2: $\psi > 0$ and $\mu_1 - \frac{\psi}{\alpha_{31}^2+\alpha_{32}^2}\alpha_{31} \leqslant 0$	$\mathbf{P}_1^3$ such that $\mathbf{P}_1^3 \cdot s_1 = 0$ and $\mathbf{P}_1^3 \cdot s_3 = \mu_3$
case 3: $\psi > 0$ and $0 \leqslant \mu_1 - \frac{\psi}{\alpha_{31}^2+\alpha_{32}^2}\alpha_{31} \leqslant \frac{\mu_3}{\alpha_{31}}$	$\hat{\mathbf{a}}$ such that $\hat{\mathbf{a}} \cdot s_i = \mu_i - \frac{\psi}{\alpha_{31}^2+\alpha_{32}^2}\alpha_{3i},\ i = 1,2$ and $\hat{\mathbf{a}} \cdot s_3 = \mu_3$
case 4: $\psi > 0$ and $\frac{\mu_3}{\alpha_{31}} \leqslant \mu_1 - \frac{\psi}{\alpha_{31}^2+\alpha_{32}^2}\alpha_{31}$	$\mathbf{P}_2^3$ such that $\mathbf{P}_2^3 \cdot s_2 = 0$ and $\mathbf{P}_2^3 \cdot s_3 = \mu_3$

Remark 17. If one of $\mu_k = 0$, says μ_1, and $\mu_2\mu_3 \neq 0$, the previous procedure is modified. From Proposition 4 we deduced that $\hat{\mathbf{a}} = (0,0)$ if and only if $\alpha_{32} < 0$. Otherwise, we obtain the following expression for the optimal slope:

$$\begin{aligned} &\text{if } \alpha_{32}\mu_2 \leqslant \mu_3, \quad \hat{\mathbf{a}} = \mathbf{G}_3 \text{ such that } \mathbf{G}_3 \cdot s_1 = \mu_1 = 0 \text{ and } \mathbf{G}_3 \cdot s_2 = \mu_2, \\ &\text{else}, \qquad\quad \hat{\mathbf{a}} = \mathbf{G}_2 \text{ such that } \mathbf{G}_2 \cdot s_1 = \mu_1 = 0 \text{ and } \mathbf{G}_2 \cdot s_3 = \mu_3. \end{aligned}$$

3.2. Q method and M method

The Q method consists in predicting the value U_{ij} using the collocation point $X_{ij} = Q_{ij}$ and we get

$$U_{ij} = U_i + \tilde{\mathbf{a}}_i \cdot \mathbf{B}_i\mathbf{Q}_{ij}, \quad j \in v(i). \tag{32}$$

The reconstruction is consistent with the linear solutions and satisfies *a priori* the stability constraint whether $\tilde{\mathbf{a}}_i \in TVD_i$ or $\tilde{\mathbf{a}}_i \in MP_i$. Nevertheless, the Q method is not optimal. Indeed, flux F_{ij} is an approximation of the exact flux integrated on the edge S_{ij}, therefore numerical integration using the value at the midpoint $\mathbf{M}_{ij}$ provides a better approximation than the value at $\mathbf{Q}_{ij}$. Consequently, we aim to evaluate U_{ij} at point $\mathbf{M}_{ij}$ in place of $\mathbf{Q}_{ij}$ leading to the following M method:

$$U_{ij} = U_i + \tilde{\mathbf{a}}_i \cdot \mathbf{B}_i\mathbf{M}_{ij}, \quad j \in v(i). \tag{33}$$

Note that the reconstruction is still consistent with the linear solutions but does not satisfy *a priori* any stability constraint even if the slope belongs to the *TVD* or *MP* domain. Theoretical stability is lost but as we shall show in the numerical test section, the solution remains L^∞ stable in most of the cases with a better accuracy than the former method using points $\mathbf{Q}_{ij}$.

4. The multislope technique

All the above second-order method are based on the linear reconstruction (2) where the slope $\mathbf{a}_i$ computed on element K_i is used to obtain all the reconstructed values $U_{ij}, j \in v(i)$. A different approach consists in providing three slopes, one for each edge of the element, such that we satisfy the two following basic conditions:

– the reconstruction is consistent for the linear function U, *i.e.* $U_{ij} = U(\mathbf{X}_{ij})$,
– if we have a local extremum at point $\mathbf{B}_i$, we find again a first-order scheme, *i.e.* the slopes vanish.

We call this method a multislope method since each value U_{ij} is obtained using a specific slope for each $j \in v(i)$.

Remark 18. We point out that the multislope reconstruction does not provide any piecewise function whereas the monoslope technique gives a linear function on each cell which satisfies the conservative property.

In the generic finite volume scheme (4), only flux evaluations are of importance but not the shape of the reconstructed function inside the domain whereas, in the finite element approach, shape functions are essential to construct the matrices deriving from the variational formulation. Conservativity property of the reconstruction does not apply in the multislope framework.

4.1. The fundamental decomposition

We first construct the slopes in each direction. To this end, we introduce the normalized vectors

$$\mathbf{t}_k = \mathbf{t}_{ij_k} = \frac{\mathbf{B}_i\mathbf{B}_{j_k}}{|\mathbf{B}_i\mathbf{B}_{j_k}|}, \quad k = 1,2,3.$$

We have the following proposition.

Proposition 19. *Assume that the mesh satisfies hypothesis* $(\mathcal{H})$, *then the following decomposition holds:*

$$\mathbf{t}_1 = \beta_{12}\mathbf{t}_2 + \beta_{13}\mathbf{t}_3, \tag{34}$$
$$\mathbf{t}_2 = \beta_{21}\mathbf{t}_1 + \beta_{23}\mathbf{t}_3, \tag{35}$$
$$\mathbf{t}_3 = \beta_{31}\mathbf{t}_1 + \beta_{32}\mathbf{t}_2, \tag{36}$$

with

$$\beta_{ml}\beta_{lm} = 1, \tag{37}$$
$$\beta_{ml}\beta_{lk} = -\beta_{mk} \tag{38}$$

for any circular permutation (m,l,k) *of (1,2,3) and all the coefficients are negative.*

Proof. Hypothesis $(\mathcal{H})$ reads

$$\mathbf{B}_i = \sum_{k=1,2,3} \rho_k \mathbf{B}_{j_k},$$

with $\rho_k > 0$ and $\rho_1 + \rho_2 + \rho_3 = 1$. We then deduce

$$0 = \sum_{k=1,2,3} \rho_k \mathbf{B}_i\mathbf{B}_{j_k} = \sum_{k=1,2,3} \rho_k |\mathbf{B}_i\mathbf{B}_{j_k}| \mathbf{t}_k.$$

Since $\rho_k|\mathbf{B}_i\mathbf{B}_{j_k}| > 0$, we conclude that all the coefficients β_{ij} are negative. Relations (37), (38) are proved as in Proposition 3. □

4.2. Multislope method with the $\mathbf{Q}_{ij}$ points

To build the multislope method, two sets of slopes are introduced. We define the downstream slopes with respect to point $\mathbf{B}_i$ in direction $\mathbf{t}_{ij_k}$ by

$$p^+_{ij_k} = \frac{U^n_{j_k} - U^n_i}{|\mathbf{B}_i\mathbf{B}_{j_k}|}, \quad k = 1,2,3, \tag{39}$$

and we define the upstream slopes by

$$p^-_{ij_1} = \beta_{12}p^+_{ij_2} + \beta_{13}p^+_{ij_3},$$
$$p^-_{ij_2} = \beta_{21}p^+_{ij_1} + \beta_{23}p^+_{ij_3},$$
$$p^-_{ij_3} = \beta_{31}p^+_{ij_1} + \beta_{32}p^+_{ij_2}.$$

Note that the downstream slopes $p^+_{ij_k}$ correspond to an approximation of the directional derivatives in the $\mathbf{t}_{ij_k}$ directions. We now give a general definition of a limiter to provide L^∞ stability for the reconstruction.

Definition 20. A function $(p,q) \to \theta(p,q)$ is a limiter if it satisfies the properties

$$\theta(p,p) = p, \quad \forall p \in \mathbb{R}, \tag{40}$$
$$\theta(p,q) = 0, \quad \forall p,q \in \mathbb{R} \quad \text{with } pq \leqslant 0, \tag{41}$$
$$\theta(p,q) = \theta(q,p), \quad \forall p,q \in \mathbb{R}. \tag{42}$$

For example the minmod limiter

$$\begin{cases} \theta(p,q) = 0 & pq \leqslant 0, \\ \theta(p,q) = \min(p,q) & p \geqslant 0,\ q \geqslant 0, \\ \theta(p,q) = \max(p,q) & p \leqslant 0,\ q \leqslant 0 \end{cases}$$

satisfies the properties. Other limiters like Van Leer's limiter, superbee limiter also satisfy the properties (40)–(42) (see [21]).

Let us define the limited slopes in the $\mathbf{t}_{ij}$ direction by

$$p_{ij} = \theta\left(p_{ij}^{+}, p_{ij}^{-}\right), \quad j \in v(i). \tag{43}$$

The multislope method reads

$$U_{ij} = U_i + p_{ij}|\mathbf{B}_i\mathbf{Q}_{ij}|, \quad j \in v(i). \tag{44}$$

Proposition 21. *Assume that the mesh satisfies hypothesis* $(\mathcal{H})$. *Then the reconstruction is consistent for the linear solution and we have a first-order scheme at the extrema.*

Proof. To prove the first assertion, let us consider a linear function $U(\mathbf{X}) = U_0 + \mathbf{L}\cdot\mathbf{X}$. The downstream slope is given by

$$p_{ij_k}^{+} = \frac{\mathbf{L}\cdot\mathbf{B}_i\mathbf{B}_{j_k}}{|\mathbf{B}_i\mathbf{B}_{j_k}|} = \mathbf{L}\cdot\mathbf{t}_k$$

and the linearity of function U yields

$$p_{ij_1}^{-} = \beta_{12}p_{ij_2}^{+} + \beta_{13}p_{ij_3}^{+} = \beta_{12}\mathbf{L}\cdot\mathbf{t}_2 + \beta_{13}\mathbf{L}\cdot\mathbf{t}_3 = \mathbf{L}\cdot(\beta_{12}\mathbf{t}_2 + \beta_{13}\mathbf{t}_3) = \mathbf{L}\cdot\mathbf{t}_1 = p_{ij_1}^{+}.$$

We conclude from property (40) that $p_{ij} = p_{ij}^{+}$ and finally we get $U_{ij} = U(\mathbf{Q}_{ij})$.

To prove the second assertion, let assume that U_i is a local minimum. All the slopes p_{ij}^{+} are non-negative since $U_j \geqslant U_i, j \in v(i)$. Under hypothesis $(\mathcal{H})$, coefficients β_{ij} are negative hence p_{ij}^{-} are non-positive. In consequence, property (41) yields $p_{ij} = 0$ and the scheme is reduced to a first-order one. □

Remark 22. The particular choice of the minmod limiter provides a TVD reconstruction in each segment $[\mathbf{B}_i, \mathbf{B}_j]$. Indeed, if $U_i \leqslant U_j$, we have $U_i \leqslant U_{ij} \leqslant U_{ij}^{ref} \leqslant U_{ji} \leqslant U_j$ (see (19) for definition of U_{ij}^{ref}).

4.3. Multislope method with the M_{ij} points

The numerical flux F_{ij} is an approximation of the exact flux integrated on edge S_{ij}. Numerical integration using midpoint $\mathbf{M}_{ij}$ for the quadrature formula provides a second-order approximation. Therefore, better accuracy shall be obtained using $\mathbf{M}_{ij}$ in place of $\mathbf{Q}_{ij}$. We then consider a new set of vectors (Fig. 7 left),

$$\mathbf{r}_k = \mathbf{r}_{ij_k} = \frac{\mathbf{B}_i\mathbf{M}_{ij_k}}{|\mathbf{B}_i\mathbf{M}_{ij_k}|}, \quad k = 1,2,3.$$

As in the previous section, we have the following proposition.

Proposition 23. *Assume that the triangle* $K \in \mathcal{T}_h$ *is not reduced to a segment. Then the non-zero coefficients of the following unique expansions*

$$\mathbf{r}_1 = \delta_{12}\mathbf{r}_2 + \delta_{13}\mathbf{r}_3, \tag{45}$$
$$\mathbf{r}_2 = \delta_{21}\mathbf{r}_1 + \delta_{23}\mathbf{r}_3, \tag{46}$$
$$\mathbf{r}_3 = \delta_{31}\mathbf{r}_1 + \delta_{32}\mathbf{r}_2 \tag{47}$$

satisfy

$$\delta_{ml}\delta_{lm} = 1, \tag{48}$$
$$\delta_{ml}\delta_{lk} = -\delta_{mk} \tag{49}$$

for any circular permutation (m,l,k) *of (1,2,3). Furthermore, since* $\mathbf{B}_i$ *is strictly inside the triangle* $(\mathbf{M}_{ij})_{j\in v(i)}$, *all the coefficients are negative.*

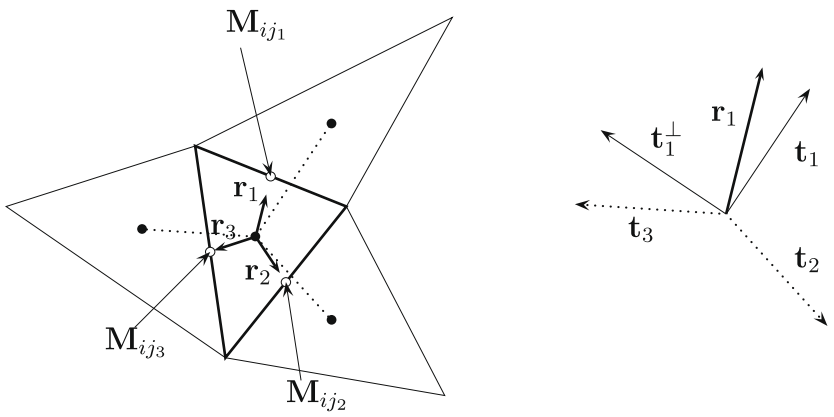


Fig. 7. Vector $\mathbf{r}_k$ (left). Decompositions of vector $\mathbf{r}_1$ in the basis $\mathbf{t}_1, \mathbf{t}_1^\perp$ and vector $\mathbf{t}_1^\perp$ in the basis $\mathbf{t}_2, \mathbf{t}_3$ (right).

4.3.1. Decomposition of $\mathbf{r}$

Natural directions to compute the slopes are $\mathbf{t}_m = \mathbf{t}_{ij_m}, m = 1,2,3$ since basic information (*i.e.* the values of U_i) are given at the centroids. To compute new interpolated values at points $\mathbf{M}_{ij}$, one has to decompose $\mathbf{r}_k$ with respect to the set $(\mathbf{t}_m)_{m=1,2,3}$. Non uniqueness of the decomposition is clear so we propose a decomposition such that we recover the Q method when M_{ij} and Q_{ij} coincide.

Let $\mathbf{t}_k^\perp$ denote a normalized orthogonal vector to $\mathbf{t}_k$.

On the one hand, we consider the unique decomposition of $\mathbf{t}_k^\perp$ in the basis $\{\mathbf{t}_m,\ m \neq k\}$ (Fig. 7, right)

$$\mathbf{t}_1^\perp = \eta_{12}\ \mathbf{t}_2 + \eta_{13}\ \mathbf{t}_3, \tag{50}$$
$$\mathbf{t}_2^\perp = \eta_{21}\ \mathbf{t}_1 + \eta_{23}\ \mathbf{t}_3, \tag{51}$$
$$\mathbf{t}_3^\perp = \eta_{31}\ \mathbf{t}_1 + \eta_{32}\ \mathbf{t}_2. \tag{52}$$

On the other hand, we decompose $\mathbf{r}_k$ as

$$\mathbf{r}_k = (\mathbf{r}_k \cdot \mathbf{t}_k)\mathbf{t}_k + (\mathbf{r}_k \cdot \mathbf{t}_k^\perp)\mathbf{t}_k^\perp. \tag{53}$$

We get the decomposition of $\mathbf{r}_k$ thanks to relations (50)–(53):

$$\mathbf{r}_k = \sum_{m=1,2,3} \xi_{km}\ \mathbf{t}_m, \tag{54}$$

with

$$\xi_{kk} = \mathbf{r}_k \cdot \mathbf{t}_k, \quad \xi_{km} = (\mathbf{r}_k \cdot \mathbf{t}_k^\perp)\eta_{km}, \quad m \neq k.$$

This decomposition satisfies the property:

$$\text{if } \mathbf{r}_k = \mathbf{t}_k \text{ then } \xi_{kk} = 1 \text{ and } \xi_{km} = 0,\ m \neq k.$$

4.3.2. Construction of the slopes

We first define the downstream slopes q_{ij}^+ as

$$q_{ij_k}^+ = \sum_{m=1,2,3} \xi_{km} p_{ij_m}^+, \quad k = 1,2,3. \tag{55}$$

Then we define the upstream slopes

$$q_{ij_1}^- = \delta_{12} q_{ij_2}^+ + \delta_{13} q_{ij_3}^+,$$
$$q_{ij_2}^- = \delta_{21} q_{ij_1}^+ + \delta_{23} q_{ij_3}^+,$$
$$q_{ij_3}^- = \delta_{31} q_{ij_1}^+ + \delta_{32} q_{ij_2}^+.$$

We compute the slopes q_{ij} using the limiter function

$$q_{ij} = \theta\left(q_{ij}^+, q_{ij}^-\right), \quad j \in v(i). \tag{56}$$

We finally define the reconstruction with

$$U_{ij} = U_i + q_{ij} |\mathbf{B}_i \mathbf{M}_{ij}|. \tag{57}$$

Proposition 24. *The reconstruction is consistent for linear functions.*

Proof. Let us consider a function $U(\mathbf{X}) = U_0 + \mathbf{L}.\mathbf{X}$ with $\mathbf{L} \in \mathbb{R}^2$. By construction, we have $p^+_{ij_k} = \mathbf{L}.\mathbf{t}_k$. Relation (55) implies that

$$q^+_{ij_k} = \mathbf{L} \cdot \sum_{m=1,2,3} (\xi_{km}\mathbf{t}_m) = \mathbf{L} \cdot \mathbf{r}_k.$$

Hence we deduce that

$$q^+_{ij_k} = \mathbf{L} \cdot \mathbf{r}_k = \frac{U(\mathbf{M}_{ij_k}) - U(\mathbf{B}_i)}{|\mathbf{B}_i\mathbf{M}_{ij_k}|}.$$

On the other hand, we write for example with $k = 1$

$$q^-_{ij_1} = \delta_{12}q^+_{ij_2} + \delta_{13}q^+_{ij_3} = \mathbf{L} \cdot (\delta_{12}\mathbf{r}_2 + \delta_{13}\mathbf{r}_3) = \mathbf{L} \cdot \mathbf{r}_1 = q^+_{ij_1}.$$

Thanks to property (40), we deduce that $q_{ij_k} = q^+_{ij_k}$ and thus $U_{ij} = U(\mathbf{M}_{ij})$ for all $j \in v(i)$. □

Remark 25. Degeneration to first-order scheme is not guaranteed by the reconstruction at point M if U_i is a local extremum. Indeed, since the point $\mathbf{B}_i$ is strictly inside the triangle with vertices $\mathbf{M}_{ij}, j \in v(i)$, all the coefficients δ_{km} are negative. Therefore if all the slopes $q^+_{ij_k}$ have the same sign, we deduce that $q^-_{ij_k}q^+_{ij_k} < 0$ then $q_{ij_k} = 0$ thanks to relation (41). But if all the slopes p^+_{ij} have the same sign, the slopes $q^+_{ij_k}$ given by relations (55) do not have *a priori* the same sign, hence the slope q_{ij} might be non-zero.

5. Numerical tests

We present in this section, several numerical tests to check the two new MUSCL methods and draw some comparisons with the monoslope technique. The numerical study aims to verify the newly implemented capability and to assess its advantages/disadvantages with regard to the classical discretization methods. A first issue concerns the linear advection problem where we test the numerical schemes using a regular function (say C^3) with and without the limiting procedures to measure the scheme accuracy. We then perform a similar test with an irregular function using the limiters to observe the scheme capability to preserve the discontinuity and to respect the maximum principle. A second issue concerns the Euler system where classical simulations like a one-dimensional Riemann problem, the Mach 3 wind tunnel with step test, and the double shock reflection test propose by Woodward and Colella [28] are performed. In the hydrodynamic context, we make the reconstruction using the primitive variables ρ, u and P to prevent non-positive values of pressure and density approximations generated by the MUSCL technique.

5.1. Solid body rotation: the regular case

The body rotation problem is considered as a challenging test for transport algorithms. We take $\mathbf{V} = (0.5 - y, x - 0.5)$ as the velocity and the unit square as Ω. For the regular case, we convect a regular compact support function given by

$$U_r(x_1, x_2) = \frac{1}{4}(\cos(4\pi r) + 1)^2 \quad \text{if } r < \frac{1}{4}, \quad U_r(x_1, x_2) = 0 \quad \text{if } r > \frac{1}{4}$$

taking $r = \sqrt{(x_1 - 0.3)^2 + (x_2 - 0.3)^2}$. Monoslope and multislope methods using collocation points Q and M are examined from the accuracy point of view. For the sake of simplicity, we name Q-methods (resp. M-methods) all the methods where we use the collocation point Q (resp. point M) for the reconstruction. We evaluate the effective errors in L^1 and L^∞ norms after a complete revolution at time $t = 2\pi$. Three types of meshes are employed: the diagonal mesh, the Scottish mesh and the Delaunay mesh presented in Fig. 8.

For the diagonal and Scottish meshes, we consider several spatial steps $h = 1/10, h = 1/20, h = 1/40, h = 1/80$ and $h = 1/160$ which correspond to $N = 10, 20, \ldots, 160$ with N the number of nodes on each side of the unit square. For the Delaunay mesh, we evaluate the spatial step size with

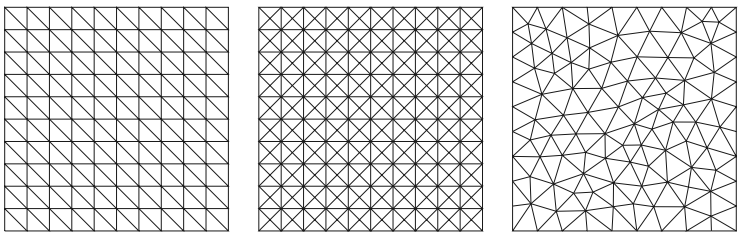


Fig. 8. The three type of mesh employed to evaluate the effective accuracy of the schemes: the diagonal mesh (left), the Scottish mesh (center) and the Delaunay mesh (right).

$$h = \min_{\substack{K_i \in \mathcal{T}_h \\ j \in v(i)}} \frac{|K_i|}{|S_{ij}|}.$$

We also characterize the mesh step by the number N of nodes on each side. Note that $M = Q$ for the diagonal mesh whereas M and Q are different for the Scottish and Delaunay meshes. Consequently, the same method using points Q or M provides the same error with the diagonal meshes.

Since we expect a second-order scheme, we use the third-order TVD Runge–Kutta discretization in time proposed by Jiang and Shu [17] such that numerical errors are only attributed to the spatial discretization.

Table 1
The solid body rotation test with a regular function. Errors and orders in the L^1 norm for MUSCL methods without limiter on diagonal meshes. We recall that the M-methods and the Q-methods are identical since $M = Q$ for this meshes.

N	10		20		40		80		160
First	4.03e−02	0.09	3.79e−02	0.26	3.16e−02	0.43	2.35e−02	0.60	1.55e−02
Mono	3.11e−02	1.13	1.42e−02	2.00	3.55e−03	2.35	6.96e−04	2.23	1.48e−04
Multi	2.98e−02	1.32	1.19e−02	2.11	2.75e−03	1.71	8.42e−04	1.27	3.50e−04

Table 2
The solid body rotation test with a regular function. Errors and orders in the L^∞ norm for MUSCL methods without limiter on diagonal meshes. We recall that the M-methods and the Q-methods are identical since $M = Q$ for this meshes.

N	10		20		40		80		160
First	7.80e−01	−0.09	8.31e−01	0.14	7.52e−01	0.31	6.05e−01	0.49	4.30e−01
Mono	5.37e−01	0.76	3.18e−01	1.78	9.27e−02	2.36	1.80e−02	2.27	3.74e−03
Multi	5.07e−01	0.87	2.77e−01	2.22	5.96e−02	1.97	1.52e−02	1.31	6.15e−03

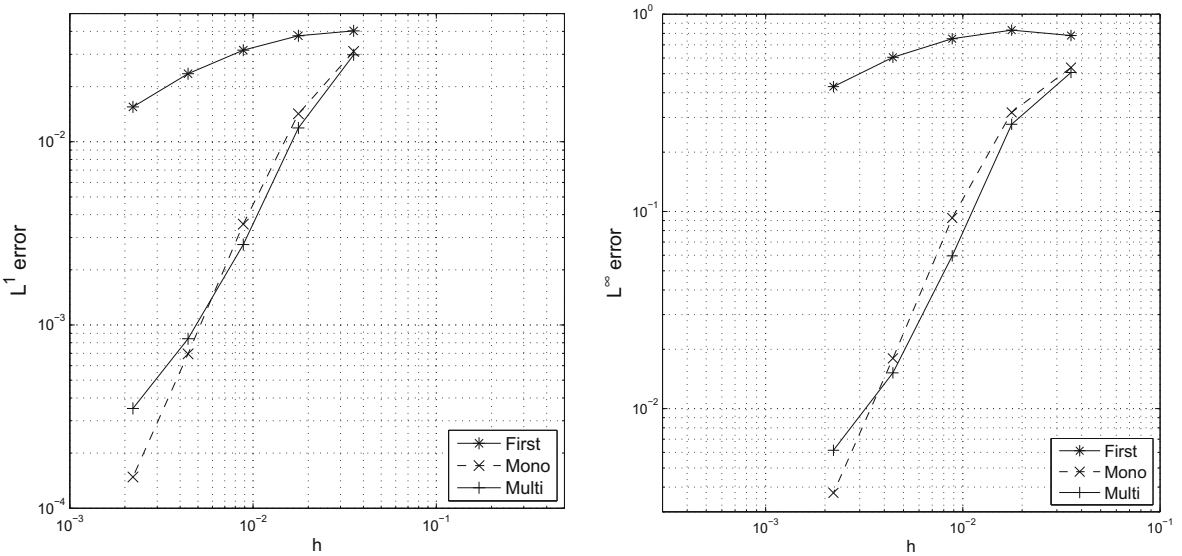


Fig. 9. The solid body rotation test with a regular function. MUSCL methods without limiter on diagonal meshes. Errors in the L^1 norm (left) and the L^∞ norm (right) versus mesh parameter h.

Table 3
The solid body rotation test with a regular function. Errors and orders in the L^1 norm for MUSCL methods without limiter on Scottish meshes.

N	10		20		40		80		160
First	4.04e−02	0.24	3.42e−02	0.41	2.58e−02	0.56	1.75e−02	0.71	1.07e−02
MonoQ	2.49e−02	1.33	9.93e−03	1.48	3.57e−03	1.15	1.61e−03	1.13	7.36e−04
MonoM	4.16e−02	2.64	6.67e−03	2.32	1.34e−03	2.37	2.59e−04	2.15	5.85e−05
MultiQ	2.25e−02	1.30	9.12e−03	1.38	3.51e−03	1.54	1.21e−03	1.43	4.49e−04
MultiM	1.93e−02	2.01	4.78e−03	2.26	1.00e−03	1.50	3.53e−04	1.34	1.39e−04

5.1.1. Reconstruction without limiters

In the present test, we compute the numerical approximation using a reconstruction where the limiters are cancelled. The following table gives the schemes we have considered.

First	First-order scheme
MonoQ	The monoslope scheme evaluated at the Q points
MonoM	The monoslope scheme evaluated at the M points
MultiQ	The multislope scheme evaluated at the Q points
MultiM	The multislope scheme evaluated at the M points

Table 4
The solid body rotation test with the regular function. Errors and orders in the L^∞ norm for MUSCL methods without limiter on Scottish meshes.

N	10		20		40		80		160
First	6.98e−01	-0.08	7.38e−01	0.23	6.28e−01	0.44	4.63e−01	0.63	3.00e−01
MonoQ	4.25e−01	0.69	2.63e−01	1.27	1.09e−01	1.15	4.92e−02	1.02	2.42e−02
MonoM	3.75e−01	1.22	1.61e−01	2.24	3.40e−02	2.45	6.21e−03	2.30	1.26e−03
MultiQ	4.17e−01	0.72	2.54e−01	1.25	1.07e−01	1.63	3.46e−02	1.50	1.22e−02
MultiM	3.20e−01	1.81	9.15e−02	1.80	2.63e−02	1.50	9.32e−03	1.11	4.32e−03

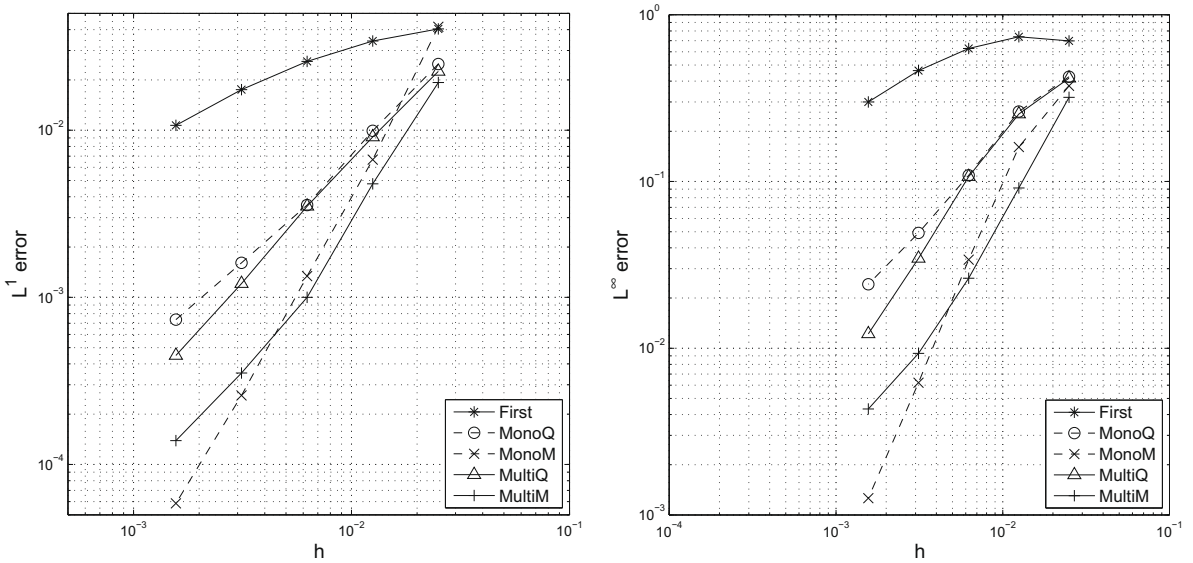


Fig. 10. The solid body rotation test with a regular function. MUSCL methods without limiter on Scottish meshes. Errors in the L^1 norm (left) and the L^∞ norm (right) versus mesh parameter h.

Table 5
The solid body rotation test with a regular function. Errors and orders in the L^1 norm for MUSCL methods without limiter on the Delaunay meshes. Unstable solutions are always obtained for the mesh with 160 nodes per side both for the monoslope and the multislope MUSCL scheme using the M-points.

N	10		20		40		80		160
First	–	–	3.69e−02	0.28	3.04e−02	0.51	2.14e−02	0.64	1.37e−02
MonoQ	2.85e−02	1.36	1.11e−02	1.99	2.80e−03	1.93	7.33e−04	1.50	2.59e−04
MonoM	2.83e−02	1.44	1.04e−02	2.27	2.16e−03	2.30	4.38e−04	–	–
MultiQ	2.67e−02	1.45	9.76e−03	2.02	2.41e−03	1.63	7.79e−04	1.40	2.96e−04
MultiM	2.62e−02	1.65	8.32e−03	2.18	1.83e−03	2.44	3.37e−04	–	–

Table 6
The solid body rotation test with a regular function. Errors and orders in the L^1 norm for MUSCL methods without limiter on the Delaunay meshes.

N	10		20		40		80		160
First	–	–	8.20e−01	0.21	7.10e−01	0.37	5.51e−01	0.54	3.78e−01
MonoQ	6.26e−01	1.20	2.73e−01	1.74	8.18e−02	1.68	2.55e−02	1.19	1.12e−02
MonoM	6.17e−01	1.33	2.45e−01	2.14	5.56e−02	2.27	1.15e−02	–	–
MultiQ	5.90e−01	1.27	2.44e−01	1.89	6.58e−02	1.00	3.28e−02	1.09	1.54e−02
MultiM	5.81e−01	1.65	1.85e−01	2.09	4.36e−02	1.70	1.34e−02	–	–

For the monoslope method, the approximation evaluation of U_{ij} consists in computing an approximation $\mathbf{a}_i$ of ∇U at point B_i using the Least Square method, then we set $U_{ij} = U_i + \mathbf{a}_i \cdot \mathbf{B_i Q}_{ij}$ or $U_{ij} = U_i + \mathbf{a}_i \cdot \mathbf{B_i M}_{ij}$. For the multislope method, the approximation evaluation without limiter consists in combining the downstream slope p_{ij}^+ and the upstream slope p_{ij}^- to constitute an optimal slope

$$p_{ij} = \frac{\chi p_{ij}^- + p_{ij}^+}{1+\chi}. \tag{58}$$

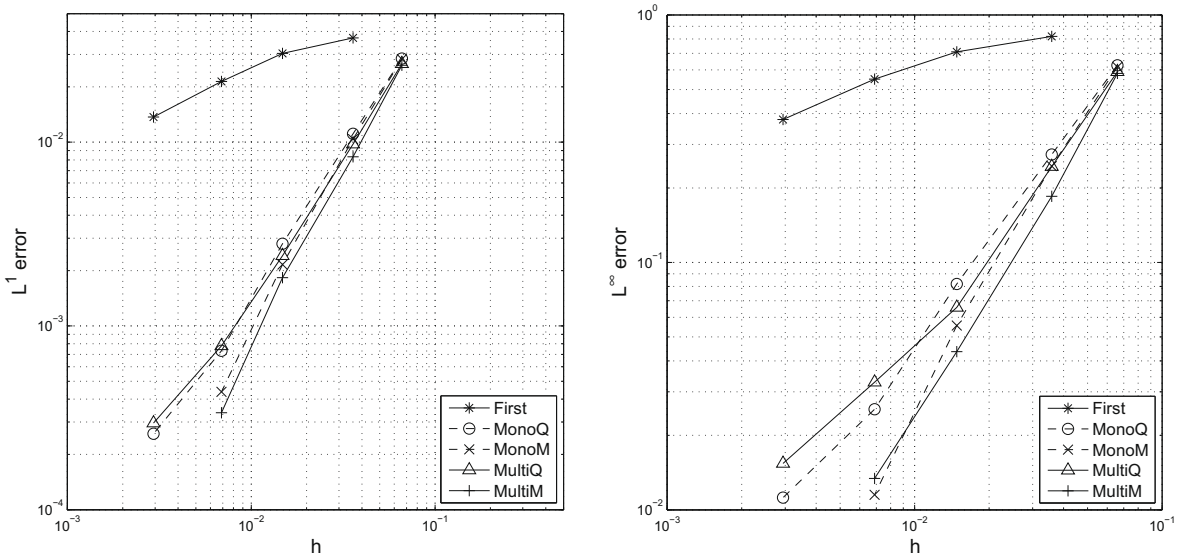


Fig. 11. The solid body rotation test with a regular function. MUSCL methods without limiter on the Delaunay meshes. Errors in the L^1 norm (left) and the L^∞ norm (right) versus mesh parameter h.

Table 7
The solid body rotation test with a regular function. Errors and orders in the L^1 norm for MUSCL methods with limiters on Scottish meshes.

N	10		20		40		80		160
MonoQ	–	–	2.19e−02	0.72	1.33e−02	0.83	7.49e−03	0.93	3.93e−03
MonoQ optTVD	3.24e−02	0.63	2.09e−02	0.76	1.23e−02	0.84	6.85e−03	0.90	3.66e−03
MonoM	–	–	2.55e−02	0.61	1.67e−02	0.75	9.91e−03	1.16	4.43e−03
MonoM optTVD	3.09e−02	0.85	1.72e−02	1.11	7.98e−03	1.08	3.78e−03	0.93	1.99e−03
minmodQ	–	–	2.32e−02	0.70	1.43e−02	0.80	8.19e−03	0.88	4.45e−03
minmodM	2.82e−02	0.93	1.48e−02	1.87	4.04e−03	1.62	1.31e−03	1.76	3.87e−04
Van AlbadaQ	3.34e−02	0.68	2.08e−02	0.83	1.17e−02	0.96	6.03e−03	0.87	3.31e−03
Van AlbadaM	2.58e−02	1.44	9.51e−03	1.49	3.38e−03	1.13	1.54e−03	1.16	6.90e−04
Van LeerQ	3.05e−02	0.82	1.73e−02	0.93	9.11e−03	0.89	4.90e−03	0.99	2.46e−03
Van LeerM	2.44e−02	1.43	9.03e−03	1.20	3.92e−03	1.16	1.75e−03	1.12	8.03e−04

Table 8
The solid body rotation test with a regular function. Errors and orders in the L^∞ norm for MUSCL methods with limiters on Scottish meshes.

N	10		20		40		80		160
MonoQ	–	–	5.69e−01	0.48	4.08e−01	0.65	2.60e−01	0.77	1.52e−01
MonoQ optTVD	6.02e−01	0.16	5.39e−01	0.56	3.65e−01	0.72	2.21e−01	0.85	1.23e−01
MonoM	–	–	6.31e−01	0.40	4.78e−01	0.60	3.15e−01	0.78	1.84e−01
MonoM optTVD	5.75e−01	0.29	4.71e−01	0.85	2.61e−01	1.06	1.25e−01	1.00	6.24e−02
minmodQ	–	–	5.75e−01	0.50	4.07e−01	0.71	2.48e−01	0.84	1.39e−01
minmodM	5.41e−01	0.52	3.76e−01	1.14	1.71e−01	1.22	7.32e−02	1.29	3.00e−02
Van AlbadaQ	6.04e−01	0.18	5.33e−01	0.62	3.46e−01	0.81	1.98e−01	0.87	1.08e−01
Van AlbadaM	5.11e−01	0.87	2.79e−01	1.49	9.95e−02	0.56	6.76e−02	0.94	3.52e−02
Van LeerQ	5.67e−01	0.30	4.60e−01	0.76	2.72e−01	0.84	1.52e−01	0.88	8.25e−02
Van LeerM	4.77e−01	0.95	2.47e−01	1.43	9.18e−02	0.39	7.00e−02	0.97	3.57e−02

The case $p_{ij} = p_{ij}^{+}$ seems to be the natural candidate. Unfortunately, $\chi = 0$ provides an unstable scheme for rough meshes ($N = 10, 20$) while we obtain a stable scheme with χ close to 0 for finer meshes ($N = 80, 160$). We choose coefficient $\chi = 1/3$ which provides a stable solution that satisfies the maximum principle.

Tables 1 and 2 give the L^1 and L^∞ errors for the three schemes using the diagonal meshes and a graphical representation is printed out in Fig. 9. Note that we only present the results obtained with the M-methods since $M = Q$ for such meshes. We observe an asymptotic error of $O(h)$ for the first-order scheme while the second-order scheme with the monoslope reconstruction provides an error of $O(h^2)$. The multislope method does not provide a full second-order convergence and we observe an important accuracy discrepancy with the monoslope situation. We think that relation (58) is responsible of the accuracy reduction and do not provide an efficient approximation of the directional derivative. Coefficients χ would certainly depend on the cell geometry and the adjacent elements in order to obtain a better approximation.

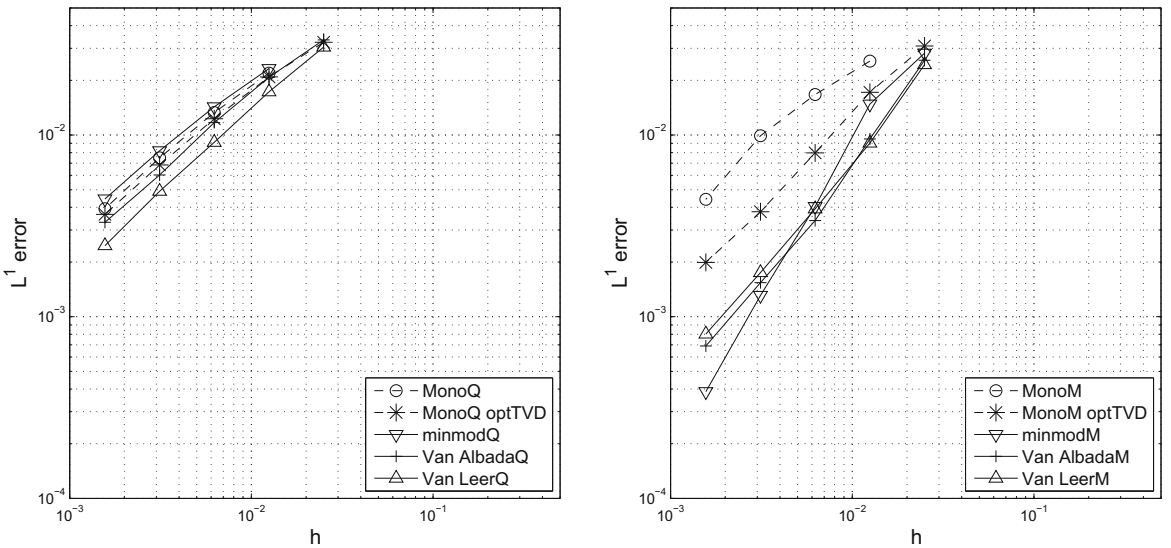


Fig. 12. The solid body rotation test with a regular function. Errors in the L^1 norm versus mesh parameter h for MUSCL methods using limiters on Scottish meshes with edge values evaluated at the Q-points (left) and at the M-points (right).

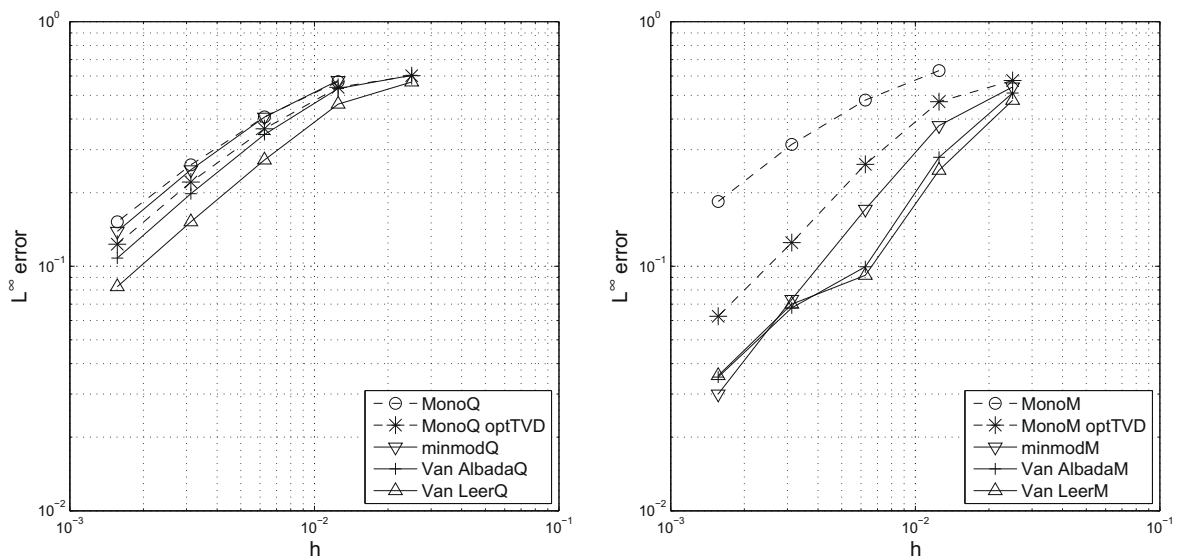


Fig. 13. The solid body rotation test with a regular function. Errors in the L^∞ norm versus mesh parameter h for MUSCL methods using limiters on Scottish meshes with edge values evaluated at the Q-points (left) and at the M-points (right).

We proceed with the Scottish meshes. Tables 3 and 4 give the L^1 and L^∞ errors for the five schemes while the errors are shown in Fig. 10. Clearly, the Q-methods are less accurate than the M-methods since the numerical integration of the flux at points Q is a first-order one whereas we get a second-order numerical integration using points M. For the Q-methods, the multislope method provides a slightly better accuracy but the monoslope method is really more performant at points M and achieves an effective second-order convergence. The multislope reconstruction at points M suffer of the lack of accuracy as in the diagonal mesh case.

We now deal with the unstructured Delaunay meshes using Q-methods and M-methods. Numerical errors and orders are listed in Tables 5 and 6 while the convergence curves are presented in Fig. 11. Like the Scottish mesh case, the collocation points M and Q differ and the M-methods provide the best convergence order. We get the same accuracy between the monoslope and multislope methods both at points Q and M. We obtain an effective second-order scheme with the M-methods. The multislope method works well with anisotropic meshes like Delaunay meshes.

5.1.2. Reconstruction with limiters

We turn to the situation where we compute the numerical approximation of a regular function using the limiting algorithms. The goal is to measure the limiting effect with respect to the unlimiting case. We only consider the Scottish mesh where the Q-methods and the M-methods differ. The following table gives the schemes we have employed in the test.

MonoQ	The monoslope scheme with the Barth limiter evaluated at the Q points
MonoQ opt TVD	The monoslope scheme with the TVD optimized limiter evaluated at the Q points
MonoM	The monoslope scheme with the Barth limiter evaluated at the M points
MonoM opt TVD	The monoslope scheme with the TVD optimized limiter evaluated at the M points
minmodQ	The multislope scheme evaluated at the Q points with the minmod limiter
Van AlbadaQ	The multislope scheme evaluated at the Q points with the Van Albada limiter
Van LeerQ	The multislope scheme evaluated at the Q points with the Van Leer limiter
minmodM	The multislope scheme evaluated at the M points with the minmod limiter
Van AlbadaM	The multislope scheme evaluated at the M points with the Van Albada limiter
Van LeerM	The multislope scheme evaluated at the M points with the Van Leer limiter

Tables 7 and 8 show the L^1 and L^∞ errors while Figs. 12 and 13 prints out the convergence curves obtained using Q-methods and M-methods. We observe the dramatic effect of the limiter for all the considered schemes. The Monoslope schemes using the Barth limiter [1, relation (64)] suffer of a huge accuracy deterioration both for the Q-methods and the M-methods. The TVD optimized monoslope reconstruction at point M gives the best convergence rate of the monoslope family but accuracy is strongly reduced with respect to the unlimiting case. For the multislope reconstructions, numerical results indicate that the limiter choice is very sensitive: for the Q-methods, the less compressive Van Leer function provides the best results

Table 9
The solid body rotation test with a discontinuous function. Errors and orders in the L^1 norm, minimum values and maximum values obtained for MUSCL methods with limiters using the Q-points on Scottish meshes.

N	10		20		40		80		160
First	1.53e−01	0.27	1.27e−01	0.43	9.42e−02	0.44	6.92e−02	0.52	4.82e−02
	0.00		0.00		0.00		0.00		0.00
	0.71		0.94		0.99		0.99		1.00
MonoQ	9.63e−02	0.41	7.24e−02	0.45	5.29e−02	0.45	3.88e−02	0.41	2.93e−02
	0.00		0.00		0.00		0.00		0.00
	0.92		0.94		0.99		1.00		1.00
MonoQ optTVD	9.95e−02	0.45	7.29e−02	0.51	5.13e−02	0.49	3.66e−02	0.48	2.62e−02
	0.00		0.00		0.00		0.00		0.00
	0.94		1.00		1.00		1.00		1.00
minmodQ	1.07e−01	0.45	7.81e−02	0.50	5.51e−02	0.48	3.94e−02	0.46	2.87e−02
	0.00		0.00		0.00		0.00		0.00
	0.92		0.99		1.00		1.00		1.00
Van AlbadaQ	1.05e−01	0.57	7.07e−02	0.48	5.06e−02	0.49	3.60e−02	0.50	2.55e−02
	0.00		0.00		0.00		0.00		0.00
	0.93		0.99		1.00		1.00		1.00
Van LeerQ	9.59e−02	0.49	6.83e−02	0.54	4.69e−02	0.51	3.29e−02	0.38	2.52e−02
	0.00		0.00		0.00		0.00		0.00
	0.97		0.99		1.00		1.00		1.00

Table 10
The solid body rotation test with a discontinuous function. Errors and orders in the L^1 norm, minimum values and maximum values obtained for MUSCL methods with limiters using the M-points on Scottish meshes.

N	10		20		40		80		160
First	1.53e−01	0.27	1.27e−01	0.43	9.42e−02	0.44	6.92e−02	0.52	4.82e−02
	0.00		0.00		0.00		0.00		0.00
	0.71		0.94		0.99		0.99		1.00
MonoM	1.04e−01	0.45	7.62e−02	0.44	5.60e−02	0.44	4.12e−02	0.47	2.97e−02
	0.00		0.00		0.00		0.00		–
	0.92		0.99		1.00		1.00		–
MonoM optTVD	8.93e−02	0.54	6.15e−02	0.57	4.14e−02	0.54	2.85e−02	0.48	2.04e−02
	0.00		0.00		0.00		0.00		0.00
	0.96		1.00		1.00		1.00		1.00
minmodM	8.65e−02	0.61	5.66e−02	0.54	3.90e−02	0.56	2.65e−02	0.50	1.87e−02
	-0.04		0.00		-0.01		-0.01		-0.01
	0.98		1.01		1.00		1.00		1.00
Van AlbadaM	8.16e−02	0.69	5.06e−02	0.56	3.43e−02	0.57	2.31e−02	0.53	1.60e−02
	-0.02		-0.01		0.00		-0.04		0.00
	0.94		1.00		1.00		1.00		1.00
Van LeerM	7.80e−02	0.69	4.84e−02	0.60	3.19e−02	0.58	2.14e−02	0.54	1.47e−02
	-0.05		-0.04		-0.01		-0.01		-0.01
	1.01		1.03		1.01		1.01		1.01

where the L^1 error is cut by 20 with respect to the minmod limiter case. For the M-method, the situation is not so clear, we have obtained the best and surprising result with the minmod limiter. In this test, the multislope methods provide smaller errors than the monoslope methods whatever the limiter choice.

5.2. Solid body rotation: the discontinuous case

We consider the rotation of a cylinder characterized by the discontinuous function

$$U_d(x_1,x_2)=1 \quad \text{if } r<\frac{1}{4}, \quad U_d(x_1,x_2)=0 \text{ if } \quad r>\frac{1}{4},$$

with $r=\sqrt{(x_1-0.5)^2+(x_2-0.3)^2}$. The schemes have to satisfy the maximum principle while reducing the viscosity effect in the vicinity of the discontinuity. Numerical simulations have been performed both with the Scottish meshes and the Delaunay meshes using the following schemes

MonoQ	The monoslope scheme with the Barth limiter evaluated at the Q points
MonoQ opt TVD	The monoslope scheme with the TVD optimized limiter evaluated at the Q points
MonoM	The monoslope scheme with the Barth limiter evaluated at the M points
MonoM opt TVD	The monoslope scheme with the TVD optimized limiter evaluated at the M points
minmodQ	The multislope scheme evaluated at the Q points with the minmod limiter
Van AlbadaQ	The multislope scheme evaluated at the Q points with the Van Albada limiter
Van LeerQ	The multislope scheme evaluated at the Q points with the Van Leer limiter
minmodM	The multislope scheme evaluated at the M points with the minmod limiter
Van AlbadaM	The multislope scheme evaluated at the M points with the Van Albada limiter
Van LeerM	The multislope scheme evaluated at the M points with the Van Leer limiter

The error obtained with the Q-methods and the M-methods using the Scottish meshes are listed in Tables 9 and 10 while the convergence curves are printed out in Fig. 14. On the other hand, we list the L^1 errors obtained with the Q-methods and the M-methods using the Delaunay meshes in Tables 11 and 12 and plot the convergence curves in Fig. 16. We observe that we obtain the same errors with the Scottish and the Delaunay meshes both using point Q or point M. In all the situations, we obtain asymptotically a convergence error of type $Ch^{1/2}$ hence the precision is mainly controlled by the value of the constant C.

Second-order Q-methods improve the approximation accuracy in comparison with the first-order method but the convergence curves are very similar and none of the method has to be distinguished. The convergence order for the M-methods is slightly greater than 1/2 and the multislope methods provide the best accuracy, in particulary when the less compressive limiters are used. The Q-methods and monoslope M-methods satisfy the maximum principle while small over(under)-shoots appear with the multislope M-methods near the discontinuities.

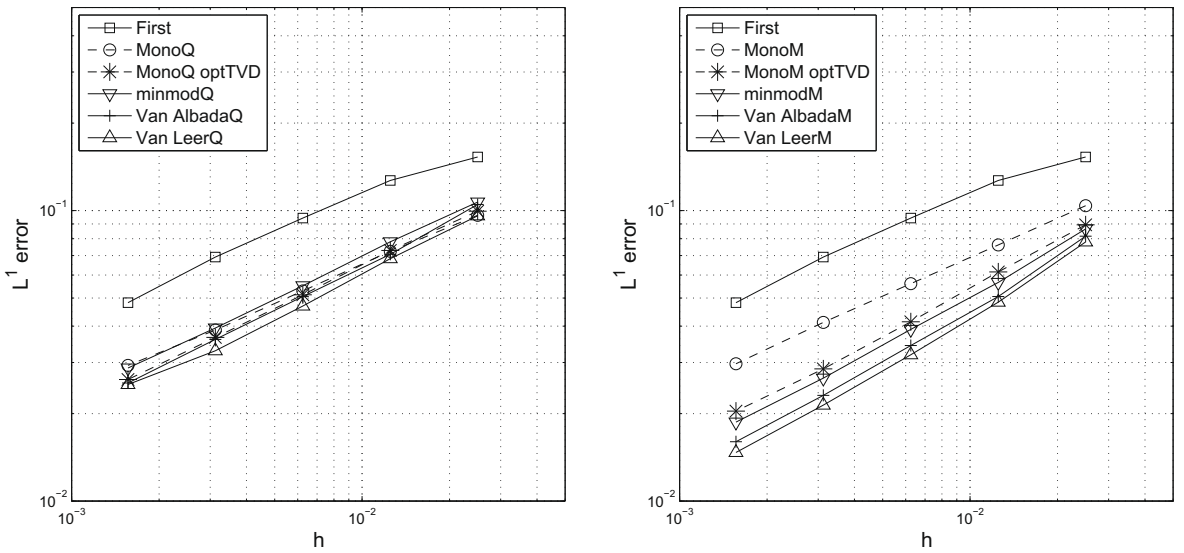


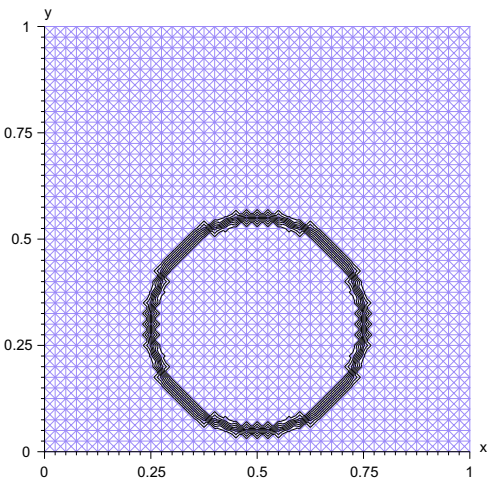
Fig. 14. The solid body rotation with a discontinuous function. Errors in the L^1 norm versus mesh parameter h for MUSCL methods with limiters using the Q-points (left) and the M-points (right) on Scottish meshes.

Table 11
The solid body rotation test with a discontinuous function. Errors and orders in the L^1 norm, minimum values and maximum values obtained for MUSCL methods with limiters using the Q-points on the Delaunay meshes.

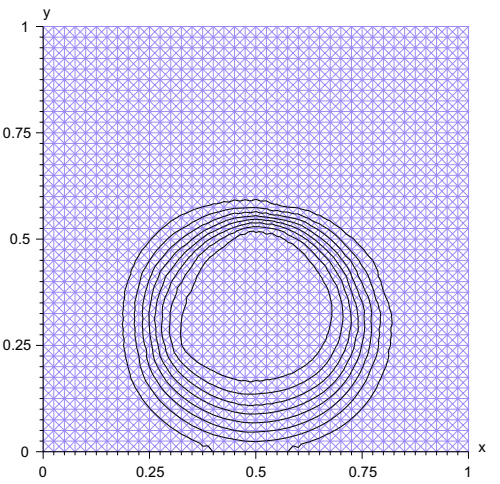
N	10		20		40		80		160
First	1.76e−01	0.36	1.37e−01	0.34	1.08e−01	0.45	7.90e−02	0.49	5.64e−02
	0.00		0.00		0.00		0.00		0.00
	0.60		0.92		0.96		1.00		1.00
MonoQ optTVD	1.10e−01	0.65	7.02e−02	0.52	4.91e−02	0.58	3.28e−02	0.66	2.07e−02
	0.00		0.00		0.00		0.00		0.00
	0.89		1.00		1.00		1.00		1.00
minmodQ	1.17e−01	0.65	7.44e−02	0.51	5.22e−02	0.52	3.65e−02	0.58	2.44e−02
	0.00		-0.05		-0.02		0.00		0.00
	0.84		1.00		1.00		1.00		1.00
Van AlbadaQ	1.12e−01	0.74	6.70e−02	0.54	4.61e−02	0.53	3.20e−02	0.61	2.10e−02
	0.00		0.00		0.00		0.00		0.00
	0.86		1.00		1.00		1.00		1.00
Van LeerQ	1.03e−01	0.81	5.88e−02	0.56	4.00e−02	0.50	2.82e−02	0.59	1.87e−02
	0.00		0.00		0.00		0.00		0.00
	0.92		1.00		1.00		1.00		1.00

Table 12
The solid body rotation test with a discontinuous function. Errors and orders in the L^1 norm, minimum values and maximum values obtained for MUSCL methods with limiters using the M-points on the Delaunay meshes.

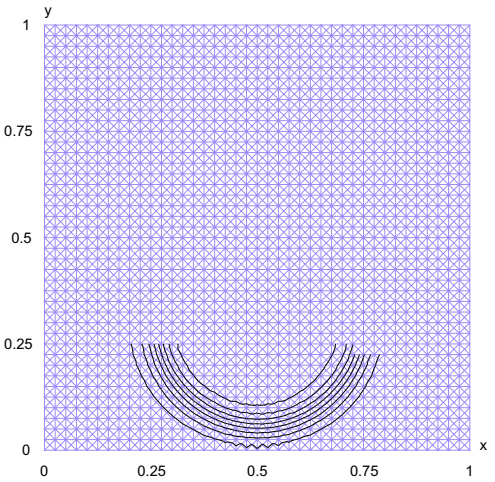
N	10		20		40		80		160
First	1.76e−01	0.36	1.37e−01	0.34	1.08e−01	0.45	7.90e−02	0.49	5.64e−02
	0.00		0.00		0.00		0.00		0.00
	0.60		0.92		0.96		1.00		1.00
MonoM optTVD	1.09e−01	0.71	6.68e−02	0.51	4.69e−02	0.63	3.04e−02	0.67	1.91e−02
	0.00		0.00		0.00		0.00		0.00
	0.89		1.00		1.00		1.00		1.00
minmodM	1.17e−01	0.72	7.09e−02	0.50	5.03e−02	0.53	3.48e−02	0.66	2.20e−02
	0.00		-0.01		-0.01		-0.01		0.00
	0.86		1.00		1.00		1.01		1.00
Van AlbadaM	1.09e−01	0.79	6.32e−02	0.55	4.31e−02	0.54	2.97e−02	0.69	1.84e−02
	0.00		-0.03		0.00		-0.01		0.00
	0.88		1.00		1.00		1.00		1.00
Van LeerM	1.01e−01	0.84	5.63e−02	0.53	3.91e−02	0.69	2.43e−02	0.73	1.46e−02
	-0.01		-0.05		-0.04		-0.03		−0.03
	0.93		1.01		1.02		1.03		1.03


y
1
0.75
0.5
0.25
0
x
0
0.25
0.5
0.75
1

(a)


y
1
0.75
0.5
0.25
0
x
0
0.25
0.5
0.75
1

(b)


y
1
0.75
0.5
0.25
0
x
0
0.25
0.5
0.75
1

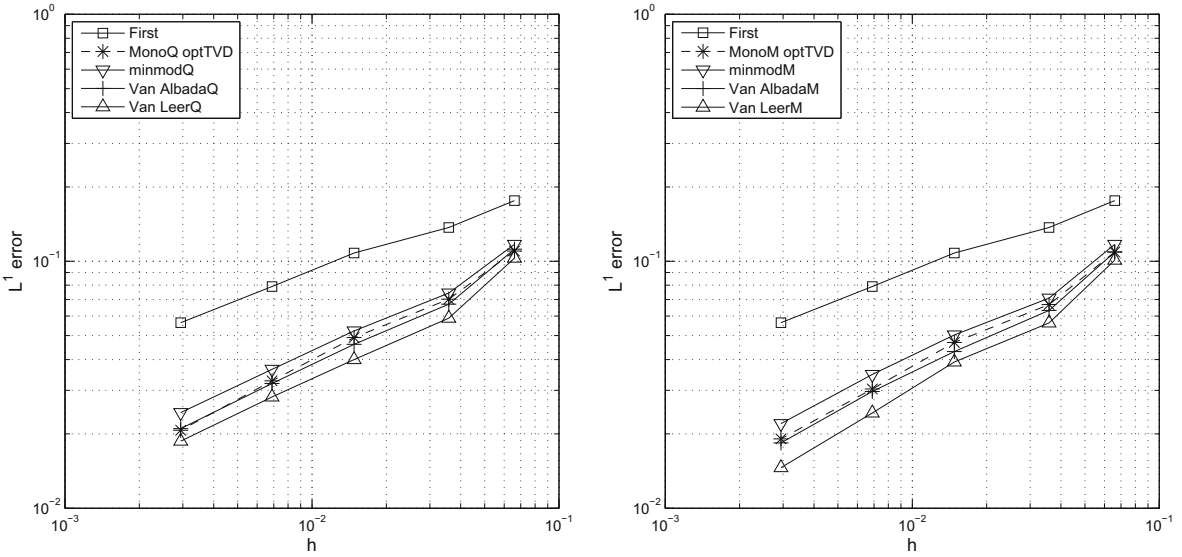


Fig. 16. The solid body rotation with a discontinuous function. Errors in the L^1 norm versus mesh parameter h for MUSCL methods with limiters using the Q-points (left) and the M-points (right) on the Delaunay meshes.

Figs. 15 and 17 show 10 isovalues of the cylinder from 0 to 1 using different limiting strategies with a 40×40 Scottish mesh and a Delaunay mesh respectively. Pictures (a) are the initial function with the two meshes while pictures (b) represent the isovalues after a complete rotation using the classical monoslope MUSCL reconstruction at point M. Pictures (c) and (d) show the isovalue repartition after the revolution using the multislope method at point M with the minmod and the Van leer limiter respectively. At last, pictures (e) present the repartition using the TVD optimized monoslope limiter at point M. Approximations using multislope reconstructions at point M provide the best results but the extrema are not preserved.

5.3. The double rarefaction Riemann problem

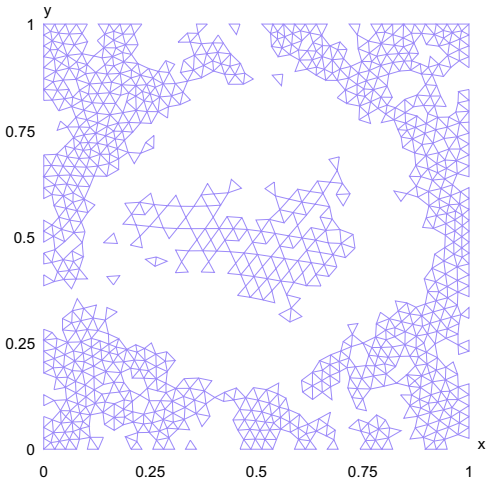
We now deal with the Euler system. We perform the reconstruction with the primitive variables $U = (\rho, u, v, P)$ to prevent non-positive pressure or density approximations. We first consider the double rarefaction Riemann problem which consists of two symmetric rarefaction waves and a trivial contact wave where the intermediate state density is close to zero for assessing the numerical performance of the schemes (see [26]). For the two-dimensional situation, we cut the unit square by the line $x = 0.5$ and we prescribe $U_L = (1, -2.0, 0, 0.4), U_R = (1, +2.0, 0, 0.4)$ on the left and right of the interface such that we recover the one-dimensional situation in the Ox direction. We perform the simulation with the Scottish mesh $(N = 40)$ and the Delaunay mesh until we reach the final time 0.15. The HLLC solver of Toro [26] is used to compute the numerical flux.

Figs. 18 and 19 print out the density and internal energy using the Scottish mesh for the Q-methods and M-methods while Figs. 20 and 21 show the same situation with the Delaunay mesh.

The first-order scheme provide the worst simulations both for the density and the internal energy. Density range is preserved with all the Q-methods and the monoslope M-methods while an overshoot appears with the multislope M-methods. The internal energy presents oscillations in the area where the gas is close to vacuum in particular with the classical monoslope methods. The multislope methods give a good approximation both for the density and the internal energy and we obtain a good result with the Van Leer limiters for the Q-methods family and the optimized monoslope method for the M-methods family.

5.4. The mach 3 wind tunnel with a step

Woodward and Colella propose in [28] two numerical simulations to evaluate the scheme performance to solve the Euler system. A uniform mach 3 flow enters in a tunnel which contains a 0.2 length units step leading to complex shock structures. We use a Delaunay mesh with a number of control volumes equal to 16,714 and close to that of the mesh considered in Woodward and Colella [28, pp. 130–131]. On the other hand, we employed here the HLL solver to compute the numerical flux rather than the HLLC solver to avoid the carbuncle phenomena and no particular treatment is performed for the singularity at the corner of the step. We print out in Fig. 22 the density configuration at the final time $t_f = 4$ units employing (a)


y
1
0.75
0.5
0.25
0
x
0
0.25
0.5
0.75
1

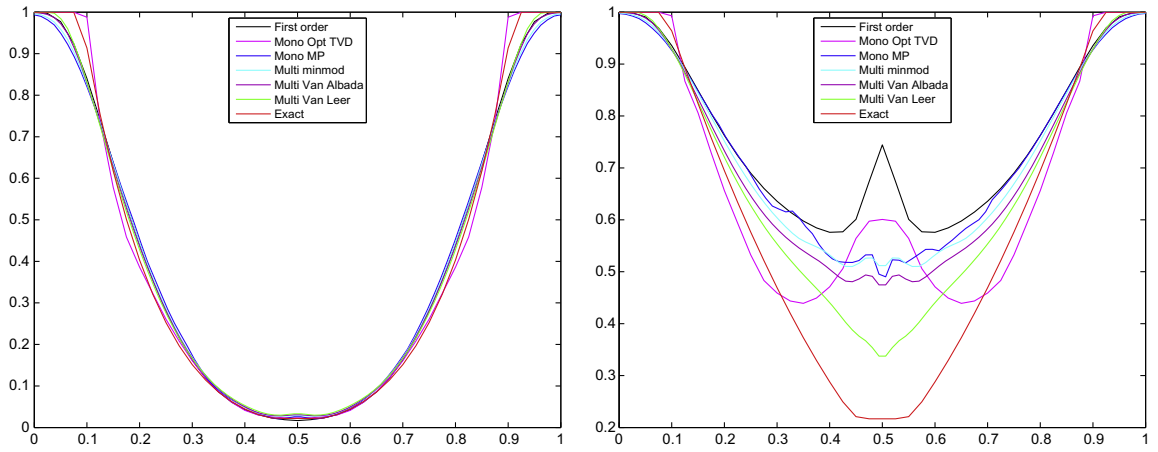


Fig. 18. Solution profiles of the double rarefaction Riemann problem on Scottish mesh using the *Q*-points. Density (left) and internal energy (right) at time $t = 0.15$.

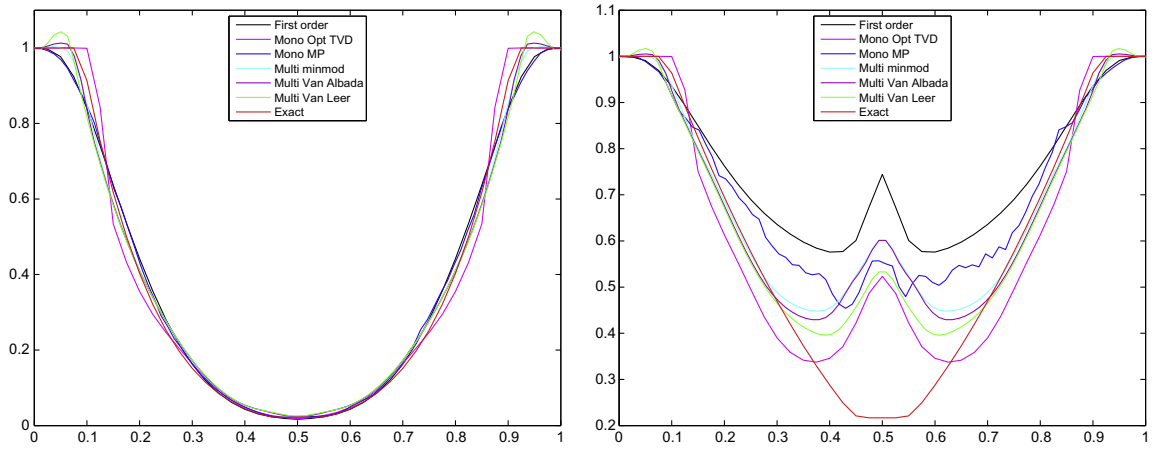


Fig. 19. Solution profiles of the double rarefaction Riemann problem on Scottish mesh using the *M*-points. Density (left) and internal energy (right) at time $t = 0.15$.

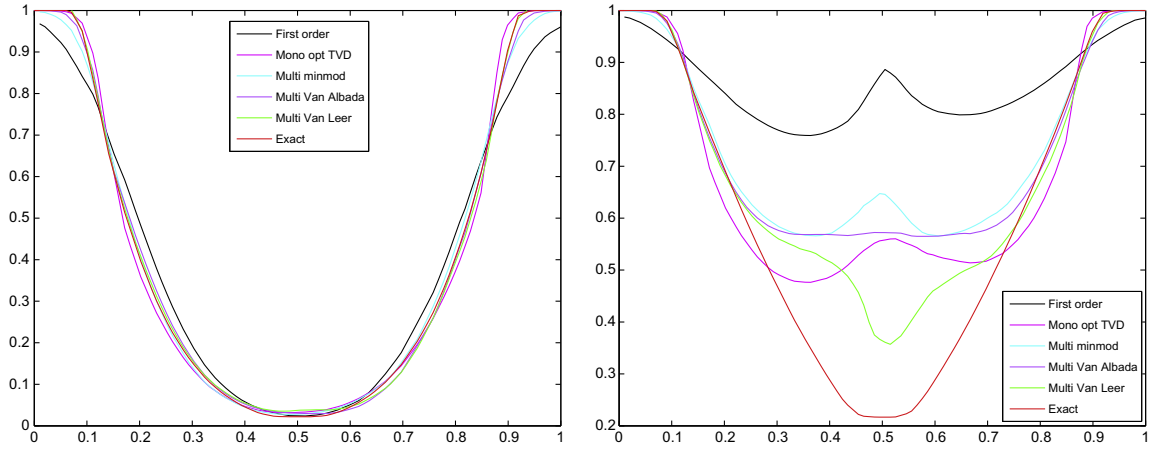


Fig. 20. Solution profiles of the double rarefaction Riemann problem on the Delaunay mesh using the *Q*-points. Density (left) and internal energy (right) at time $t = 0.15$.

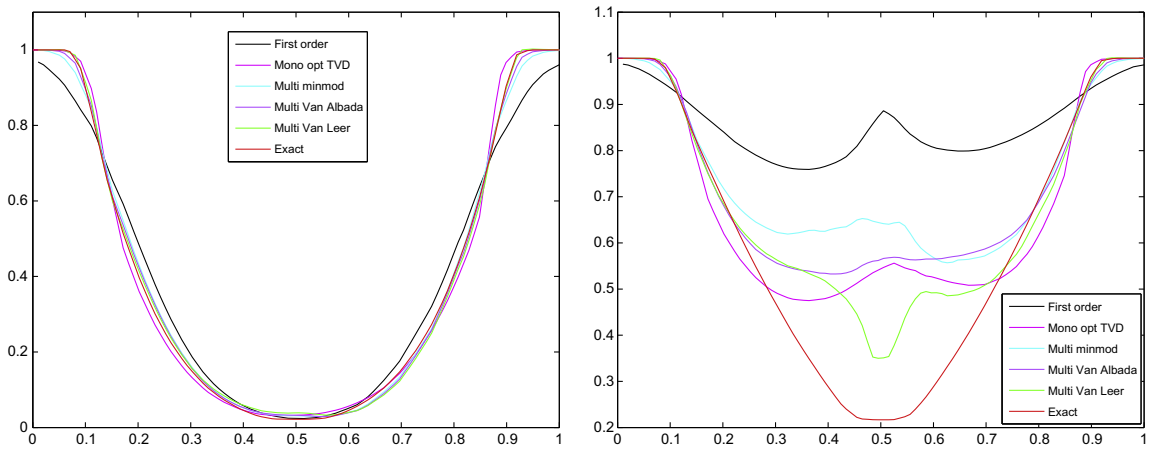
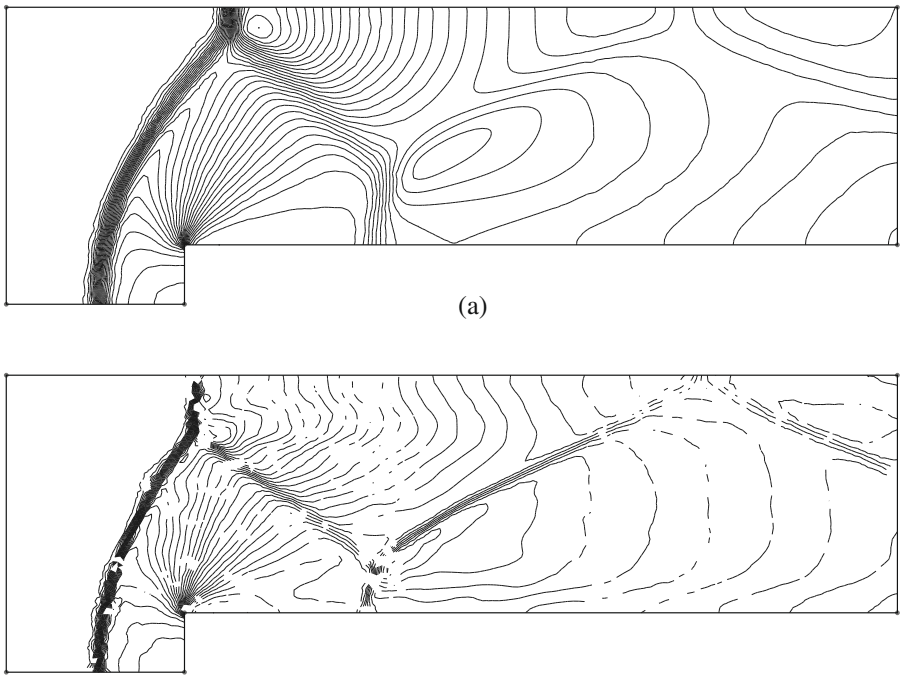
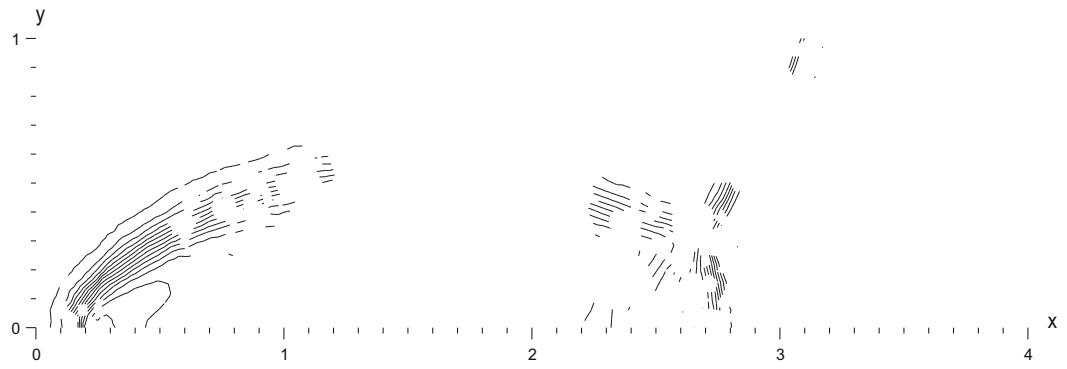


Fig. 21. Solution profiles of the double rarefaction Riemann problem on the Delaunay mesh using the *M*-points. Density (left) and internal energy (right) at time $t = 0.15$.



(a)



the first-order scheme, (b) the monoslope MUSCL scheme at point Q and (c) the multislope MUSCL scheme at point Q with the Van Leer limiter. We have also experiment the M-methods but computations may fail since the density and pressure positivities are not preserved. The multislope method with the less compressive limiter provides sharper shocks, reducing the local numerical diffusion and we obtain a better resolution of the slip lines.

5.5. The double Mach reflection of a strong shock

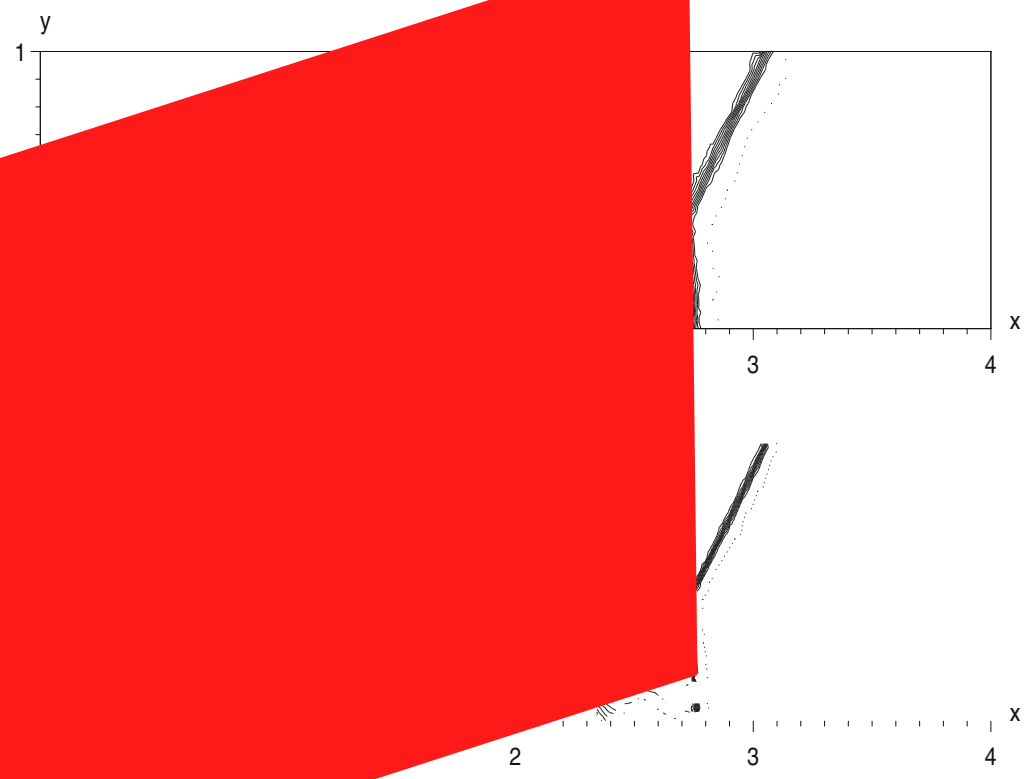
The second popular test proposed by Woodward and Colella for high-resolution schemes corresponds to a planar strong shock meeting a 60° inclinated wall with a wedge (see [28, p. 135] for a detailed description of the physical problem). We use three Delaunay meshes of spatial steps $h = 1/30, h = 1/60, h = 1/120$ to draw comparisons with the numerical experiments of Woodward and Colella [28] and we employed the HLLC solver to compute the numerical flux. In Figs. 23–25 we plot density contours for the three meshes using the first-order method, the monoslope method at point Q and the multislope method at point Q with the Van Leer limiter respectively. The first-order scheme gives a poor

representation of the shock structure and a large amount of viscosity smoothes the shocks whereas the second-order schemes provide cleaner shocks with a reasonable level of numerical dissipation in comparison with the standard schemes [28,17].

6. Conclusions

In this paper, new MUSCL methods have been presented in the context of cell-centered Finite Volume method where control volumes are triangles. First an enhancement of the monoslope MUSCL methods that is to say with one vectorial slope per control volume has been introduced. It is based on a minimization under constraint corresponding to the desired stability condition. Afterwards a new MUSCL method approach has been proposed. It consists in computing three scalar slopes per triangle following the three directions given by the neighbouring triangles.

The methods achieve a better accuracy in comparison with the classical gradient method with a rough limiter. The convergence rate of both the methods are similar but in the multislope case, time consumption is reduced and the implemen-



situation is straig ned with four directions and very few modifications have to be done to adapt MUSCL method, some complementary studies should be considered to provide merical point of view, the choice of the edge midpoint M to compute the interpolate rings higher accuracy but scheme is less stable since the maximum principle is not

Acknowledgments

The authors are grateful to the rev ve comments.

References

[1] T.J. Barth, Numerical methods and error estimation fo on structured and unstructured meshes, VKI Comput. Fluid Dyn., Lecture Notes (2003).
[2] T.J. Barth, D.C. Jespersen, The design and application of upw on unstructured meshes, AIAA Report 89-0366, 1989.
[3] T.J. Barth, M. Ohlberger, Finite Volume Methods: Foundation a s, Encyclopedia of Computational Mechanics, vol. 1, John Wiley & Sons Ltd, 2004 (Chapter 15).

[4] T. Buffard, Analyse de quelques méthodes de volumes finis non structurés pour la résolution des équations d'Euler, Thèse de doctorat de l'Université Paris 6, France, 1993.
[5] E. Bertolazzi, G. Manzini, A cell-centered second-order accurate finite volume method for convection–diffusion problems on unstructured meshes, Math. Models Meth. Appl. Sci. 14 (8) (2004) 1235–1260.
[6] P.-H. Cournède, B. Koobus, A. Dervieux, Positivity statements for a mixed-element-volume scheme on fixed and moving grids, Revue européenne de mécanique numérique 15 (7) (2006) 767–798.
[7] P. Chevrier, H. Galley, A Van Leer finite volume scheme for the Euler equations on unstructured meshes, RAIRO Modèl. Math. Anal. Numér. 27 (2) (1993) 183–201.
[8] P. Colella, Multidimensional upwind methods for hyperbolic conservation laws, J. Comput. Phys. 87 (1) (1990) 171–200.
[9] J.-A. Désidéri, A. Dervieux, Compressible flow solvers using unstructured grids, Von Karman Inst. Fluid Dynamics, Lecture Series 1988-05, 1988.
[10] L.-J. Durlofsky, B. Engquist, S. Osher, Triangle based adaptative stencils for the solution of hyperbolic conservation laws, J. Comput. Phys. 98 (1) (1992) 64–73.
[11] J. Gressier, P. Villedieu, J.-M. Moschetta, Positivity of flux vector splitting schemes, J. Comput. Phys. 155 (1999) 199–220.
[12] E. Godlewski, P.A. Raviart, Numerical approximation of hyperbolic systems of conservation laws, Appl. Math. Sci. 118 (1996).
[13] J.B. Goodman, R.J. LeVeque, On the accuracy of stable schemes for 2D scalar conservation laws, Math. Comput. 45 (171) (1985) 15–21.
[14] M.E. Hubbard, Multidimensional slope limiters for MUSCL-type finite volume schemes on unstructured grids, J. Comput. Phys. 155 (1) (1999) 54–74.
[15] A. Jameson, Artificial diffusion, upwind biasing, limiters and their effect on accuracy and multigrid convergence in transonic and hypersonic flows, in: Proceedings of the 11th AIAA Computational Fluid Dynamics Conference, AIAA Paper 93-3359, 1993.
[16] A. Jameson, D. Mavriplis, Finite volume solution of the two-dimensional Euler equations on a regular triangular mesh, AIAA J. 24 (4) (1986) 611–618.
[17] G.-S. Jiang, C.-W. Shu, Efficient implementation of weighted ENO schemes, J. Comput. Phys. 126 (1996) 202–228.
[18] D. Kuzmin, S. Turek, High-resolution FEM-TVD schemes based on a fully multidimensional flux limiter, J. Comput. Phys. 198 (2004) 131–158.
[19] D. Kröner, Numerical schemes for conservation laws, in: Wiley Teubner (Ed.), Series Advances in Numerical Mathematics, John Wiley & Sons Ltd., 1997.
[20] D. Kröner, S. Noelle, M. Rokyta, Convergence of higher order upwind finite volume schemes on unstructured grids for scalar conservation laws in several space dimensions, Numer. Math. 71 (4) (1995) 527–560.
[21] R.J. LeVeque, Numerical Methods for Conservation Laws, Birkhaüser Verlag, Basel, 1992.
[22] B. Perthame, C.-W. Shu, On positivity preserving finite volume schemes for Euler equations, Numer. Math. 73 (1996) 119–130.
[23] S. Piperno, S. Depeyre, Criteria for the design of limiters yielding efficient high resolution TVD schemes, Comput. Fluids 27 (2) (1998) 183–197.
[24] S.P. Spekreijse, Multigrid solution of monotone second-order discretizations of hyperbolic conservation laws, Math. Comput. 49 (179) (1987) 135–155.
[25] P.K. Sweby, High resolution schemes using flux limiters for hyperbolic conservation laws, SIAM J. Numer. Anal. 21 (5) (1984) 995–1011.
[26] E.F. Toro, Riemann Solvers and Numerical Methods for Fluid Dynamics, A Practical Introduction, Springer-Verlag, Berlin, 1997.
[27] B. Van Leer, Towards the ultimate conservative difference schemes V. A second-order sequel to Godunov's method, J. Comput. Phys. 32 (1) (1979) 101–136.
[28] P. Woodward, P. Colella, The numerical simulation of two-dimensional fluid flow with strong shocks, J. Comput. Phys. 54 (1) (1984) 115–173.