L^∞ stability of the MUSCL methods

Stéphane Clain, Vivien Clauzon

L^∞ stability of the MUSCL methods

Stéphane Clain[1], **Vivien Clauzon**[2]

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

Received: date / Revised version: date

Summary We present a general L^∞ stability result for a generic finite volume method for hyperbolic scalar equations coupled with a large class of reconstruction. We show that the stability is obtained if the reconstruction respects two fundamental properties: the convexity property and the sign inversion property. We also introduce a new MUSCL technique, the multislope MUSCL technique, based on the approximations of the directional derivative in contrast to the classical piecewise reconstruction, the monoslope MUSCL technique, based on the gradient reconstruction. We show that under specific constraints we shall detail, the two MUSCL reconstructions satisfy the convexity and sign inversion properties and we prove the L^∞ stability.

Mathematics Subject Classification (2000) 65N12-65N20

Key words Finite volume, MUSCL method, L^∞ stability, reconstruction method

1 Introduction

L^∞ stability plays a fundamental role to provide suitable numerical approximations computed by a finite volume scheme. For hyperbolic scalar problem, the L^∞ stability is required to prove the convergence of approximations to the entropy solution when the mesh step goes to zero. It is well-known that explicit first-order schemes using a monotonone flux function are stable [9] p. 383, [13] p. 174, but the situation becomes more complex

for second-order schemes using a reconstruction, for example the popular MUSCL technique coupled with a cell centered finite volume scheme (see [9] p. 405, [13] p. 212, [11]). For the one dimensional situation, the L^∞ stability and the Total Variation stability for second-order scheme using a MUSCL reconstruction have been proved [17], [15]. For higher dimensions, the TVD stability condition reduces the method to a first-order scheme for uniform cartesian meshes [10] while the TVD stability no longer holds for unstructured meshes [8]. To deal with stability in dimension two or greater, a generalisation of the one dimensional incremental scheme is introduced for the cartesian grid: the positive coefficient scheme [16], and generalized to unstructured meshes [12,7] based on the Local Extremum Diminishing concept. To prove L^∞ stability, one has to rewrite the finite volume scheme as a positive scheme where the coefficients belong to the interval $[0,1]$. Under an appropriate CFL condition, the L^∞ stability derives from a convexity argument (see [1] and the references therein).

In this paper, we intend to generalize the L^∞ stability result for schemes coupled with a reconstruction method. We first propose a different definition of the reconstruction where we only deal with the reconstructed values instead of considering the whole reconstructed function. We then introduce two fundamental proprerties: the convexity property and the sign inversion property. Closely concepts related to the inversion sign property have been also introduced by [14] using geometrical arguments (see also [3]) but we manage to avoid the geometrical aspect working directly with the reconstructed values instead of the collocation points. More recently, a close version of the property has been proposed by [3] in the context of the classical MUSCL method. We prove that, under an appropriate CFL condition, a finite volume scheme with a monotone numerical flux coupled with a reconstruction satisfying the two properties is L^∞ stable.

In a second step, we prove that the two properties are satisfied by a large class of MUSCL methods: the monoslope (classical) MUSCL and the multislope MUSCL methods. The classical MUSCL technique consists in two steps: a predicted gradient is computed for each element of the mesh using the neighbouring values then it is modified to respect some Maximum Principle or Total Variation Diminishing constraint [2,11]. The MUSCL method is referred to as monoslope method since the reconstructed values at each interfaces are obtained using the same vectorial slope evaluated on the cell. The new MUSCL method named multislope method consists in using a specific scalar slope for each interface which corresponds to an approximation of the directional derivative instead of an approximation of the gradient [2,4,5]. For a given element, we consider a set of normalized vectors and we use the neighbouring values to compute the scalar slopes in each direction which are modified afterwards to respect some stability

constraint. The main advantage of the mulstislope method is that we only deal with one-dimensional problems independently of the space dimension where Ω belongs to.

The organization of the paper is as follows. In section 2, we present a general result for L^∞ stability for scalar hyperbolic equations where the two fundamental properties are introduced. The results are given for schemes based on a Euler forward discretization in time but all the results hold if one only consider the semidiscrete approximation in space. Section 3 is dedicated to the multislope MUSCL method where we prove the L^∞ stability. In particular, we introduce two new definitions, the α-convexity of the triangulation and the τ_{lim} parameter which are crucial to control the reconstruction and the limiter. Section 4 deals with the L^∞ stability for the monoslope MUSCL method. We end the section with a comparison between the monoslope and multislope method in order to show that the multislope reconstruction is less sensitive to the stability constraint: for a given configuration, the multislope method provides non zero slopes while the monoslope method is reduced to a first-order one.

2 Nonlinear stability : a general result

2.1 Notations and geometrical ingredients

Let Ω be an open bounded polygonal set of $\mathbb{R}^2$, we denote by $\mathcal{T}_h$ an unstructured mesh of Ω composed of close triangles (cells, control volumes or elements). We denote by K_i, $i = 1, \ldots, I$, the elements of centroid (gravity center) $B_i \in K_i$ where I represents the number of cells. To handle the boundary conditions, we add ghost cells in the following way: if triangle K_i has an edge e on $\partial\Omega$, we construct the symmetrical triangle K_j and we denote by $\widetilde{\mathcal{T}_h}$ the mesh completed with the ghost cells.

For a given element $K_i \in \mathcal{T}_h$, $\nu(i)$ is the index set of the neighbouring elements $K_j \in \widetilde{\mathcal{T}_h}$ which share a common edge and we define the sides of the mesh by

$$S_{ij} = K_i \cap K_j \neq \emptyset, \quad j \in \nu(i)$$

where $n_{ij} = (n_{ij,1}, n_{ij,2})$ is the outward normal vector. Note that $n_{ij} = -n_{ji}$ so the index order is of importance. In the sequel, $|K_i|$ represents the surface of the element while $|S_{ij}|$ or $|B_iB_j|$ are the length of the side or the segment $[B_i, B_j]$.
We also denote by L_{ij} the line such that $S_{ij} \subset L_{ij}$ and Q_{ij} the intersection point between $[B_i, B_j]$ and L_{ij} (see figure 1). Note that *a priori*, Q_{ij} is not necessary a point of S_{ij}.

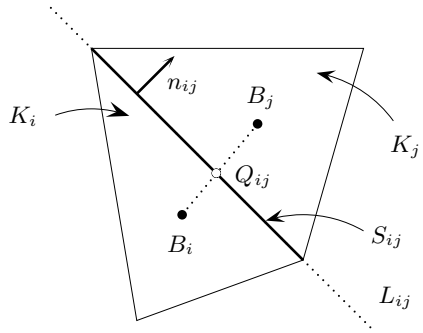

Fig. 1. Notations and conventions of the mesh elements and edges.

Remark 1 We easily extend the notation for the three dimensional geometries where the mesh is composed of tetrahedron and $|K_i|$, $|S_{ij}|$, $|B_iB_j|$ stand for the volume, the surface and the length respectively.

We consider a general scalar hyperbolic problem cast in the conservative form

$$\partial_t u + \partial_{x_1} f_1(u) + \partial_{x_2} f_2(u) = 0 \tag{1}$$

where f_1 and f_2 are C^1 real value functions defined on $\mathbb{R}$. Of course the definition domain of f_1 and f_2 can be reduced to the admissible domain of the solution u but we skip this point for the sake of simplicity.
In the sequel, we shall only consider the reflexion boundary condition

$$\partial_n u = 0 \quad \text{ on } \partial\Omega, \tag{2}$$

while we assume $u(.,t=0) = u^0$ on Ω where u^0 stands for the solution at the initial time $t = 0$.

2.2 Generic first-order monotone scheme

For a given time t^n and a cell $K_i \in \mathcal{T}_h$, we denote by

$$u_i^n \approx \frac{1}{|K_i|} \int_{K_i} u(.,t^n)\, dx$$

an approximation of the mean value of u on cell K_i at time t^n. For any ghost cell K_j which shares the side $e \in \partial\Omega$ with triangle $K_i \in \mathcal{T}_h$, we set $u_j^n = u_i^n$ in order to satisfy the reflexion boundary condition (2). It results that

$$\{u_i;\ K_i \in \widetilde{\mathcal{T}_h}\} = \{u_i;\ K_i \in \mathcal{T}_h\}. \tag{3}$$

For any side S_{ij}, we denote by $g(\alpha,\beta,n_{ij})$, $\alpha,\beta \in \mathbb{R}$ the numerical flux accross S_{ij} in the direction n_{ij}. We detail the conditions required by the numerical flux:

(a) function g is continuous, differentiable with respect to the first and the second argument and $\partial_1 g$, $\partial_2 g$ are continuous functions;
(b) the numerical flux is consistent with the physical flux (f_1, f_2):

$$g(\alpha, \alpha, n_{ij}) = f_1(\alpha) n_{ij,1} + f_2(\alpha) n_{ij,2}; \tag{4}$$

(c) the numerical flux is monotone:

$$\partial_1 g(\alpha, \beta, n_{ij}) \geq 0, \qquad \partial_2 g(\alpha, \beta, n_{ij}) \leq 0. \tag{5}$$

Note that the consistancy implies the conservation property

$$\sum_{j \in \nu(i)} \frac{|S_{ij}|}{|K_i|} g(\alpha, \alpha, n_{ij}) = 0. \tag{6}$$

The flux conservation accross the interface is usualy satisfied by the numerical flux:

$$g(\alpha, \beta, n_{ij}) = -g(\beta, \alpha, n_{ji})$$

but this last condition is not necessary to provide the stability of the scheme and only relation (6) is required. Non-conservative numerical flux may be also considered.

If we have an approximation $u_h^n = \sum_{K_i \in \mathcal{T}_h} u_i^n 1_{K_i}$ at time t^n, the generic first order explicite finite volume scheme provides an approximation at time $t^{n+1} = t^n + \Delta t$ by

$$u_i^{n+1} = u_i^n - \Delta t \sum_{j \in \nu(i)} \frac{|S_{ij}|}{|K_i|} g(u_i^n, u_j^n, n_{ij}). \tag{7}$$

2.3 Reconstruction

Nonlinear stability *i.e.* L^∞ stability for first-order scheme is well-established if one satisfies an appropriate CFL condition [9, 13]. We intend to define a general framework to obtain the nonlinear stability when we applied a reconstruction procedure to enhance the approximation accuracy. To this end we use a more general definition of the reconstruction operator where we only focus on the reconstructed values on the sides instead of providing a complete reconstruction (classically a linear piecewise reconstruction) on the whole domain.

A reconstruction is an operator which gives new values on both sides of S_{ij} using the values u_i on the cells $K_i \in \widetilde{\mathcal{T}_h}$. Formally, we define the reconstruction operator $\mathcal{R}$ by

$$(u_i)_{K_i \in \widetilde{\mathcal{T}_h}} \overset{\mathcal{R}}{\to} (u_{ij})_{K_i \in \mathcal{T}_h,\, j \in \nu(i)} \tag{8}$$

Usually, the values u_{ij} and u_{ji} correspond to an approximation of u at a collocation point X_{ij} on the side S_{ij}. Note that we employ the ghost cells in order to realise the reconstruction on the boundary sides of Ω.

Assume that we have an approximation u_i^n for all $K_i \in \mathcal{T}_h$, we extend the solution on the ghost cells as we define in paragraph 2.2 and the reconstruction provides u_{ij}^n and u_{ji}^n on each side S_{ij}. We then compute an approximation u_i^{n+1} with the generic second-order scheme

$$u_i^{n+1} = u_i^n - \Delta t \sum_{j\in\nu(i)} \frac{|S_{ij}|}{|K_i|} g(u_{ij}^n, u_{ji}^n, n_{ij}). \tag{9}$$

We now introduce the two fondamental assumptions on the reconstruction operator to obtain the L^∞ stability. We only present the situation for the two-dimensional case but the extension to the multi-dimensional case is straigthforward. In the sequel, we use the following notations :

$$\Delta_{ij}u = u_{ij} - u_i, \quad \Delta_{ji}u = u_{ji} - u_j \tag{10}$$

$$\widetilde{\Delta_{ij}}u = u_{ij} - u_j, \quad \widetilde{\Delta_{ji}}u = u_{ji} - u_i \tag{11}$$

Note that we have the identity $u_i + \Delta_{ij}u = u_j + \widetilde{\Delta_{ij}}u$

Definition 1 (Convexity property) *The reconstruction has the convexity property if for any $K_i \in \mathcal{T}_h$ and $j \in \nu(i)$, there exists $\theta_{ij} \in [0,1]$ such that*

$$u_{ij} = (1-\theta_{ij})u_i + \theta_{ij}u_j. \tag{12}$$

Using definition of $\Delta_{ij}u$ and $\widetilde{\Delta_{ij}}u$ we get

$$\Delta_{ij}u = \theta_{ij}(u_j - u_i), \qquad \widetilde{\Delta_{ij}}u = (1-\theta_{ij})(u_i - u_j). \tag{13}$$

In particular, we deduce that if $u_i = u_j$, the convex reconstruction assumption yields $u_{ij} = u_i$. Note that $\theta_{ij} = 0$ corresponds to a first-order reconstruction.

Remark 2 Relation (12) does not implies that u_{ij} only depends on u_i and u_j. Indeed, as we shall see in the sequel, coefficients θ_{ij} depend on the other values of u_k, $k \in \nu(i) \cup \nu(j)$.

For any $K_i \in \mathcal{T}_h$, we say that u_i is a discrete local maximum (resp. discrete local minimum) if $u_i \geq u_j$, $\forall j \in \nu(i)$ (resp. $u_i \leq u_j$, $\forall j \in \nu(i)$). We introduce a complementary definition to the convexity property.

Definition 2 (Degeneracy at the Extrema property) *The reconstruction degenerates at the discrete local extrema if coefficients θ_{ij} satisfy the condition : if u_i is a discrete local extremum then $\theta_{ij} = 0$ for all $j \in \nu(i)$ i. e. we find again a first-order scheme at the extrema.*

We recall that the scheme (9) is a Local Diminishing Extrema scheme (LED scheme) if we have (see [12, 7]):

- if u_i^n is a discrete local maximum then $u_i^{n+1} \leq u_i^n$;
- if u_i^n is a discrete local minimum then $u_i^{n+1} \geq u_i^n$.

The LED property means that the maximum (resp. minimum) can not increase (resp. decrease). It is not enough to prove the L^∞ stability for a total discretizated scheme but a scheme which does not satisfy the LED property is disqualified since the L^∞ stability implies the LED property.

Proposition 1 *If $\mathcal{R}$ is a reconstruction which satisfies the convexity property and degenerates at the extrema then the scheme (9) equipped with a monotone flux is LED.*

Proof We write the genereric scheme in the following way

$$u_i^{n+1} = u_i^n - \Delta t \sum_{j \in \nu(i)} \frac{|S_{ij}|}{|K_i|} g(u_i^n + \Delta_{ij}^n u, u_i^n + \widetilde{\Delta_{ji}^n} u, n_{ij}).$$

On the other hand, the conservation property (6) yields

$$\sum_{j \in \nu(i)} \frac{|S_{ij}|}{|K_i|} g(u_i^n, u_i^n, n_{ij}) = 0.$$

We introduce the function

$$h_{ij}(\xi) = g(u_i^n + \xi \Delta_{ij}^n u, u_i^n + \xi \widetilde{\Delta_{ji}^n} u, n_{ij}).$$

Function h_{ij} in continuous differentiable on the intervalle $[0, 1]$ with $h_{ij}(0) = g(u_i^n, u_i^n, n_{ij})$ and $h_{ij}(1) = g(u_i^n + \Delta_{ij}^n u, u_i^n + \widetilde{\Delta_{ji}^n} u, n_{ij})$. It results that there exists an intermediate value ξ_0 such that

$$g(u_i^n + \Delta_{ij}^n u, u_i^n + \widetilde{\Delta_{ji}^n} u, n_{ij}) - g(u_i^n, u_i^n, n_{ij}) = h'_{ij}(\xi_0).$$

We then obtain

$$\begin{aligned} &g(u_i^n + \Delta_{ij}^n u, u_i^n + \widetilde{\Delta_{ji}^n} u, n_{ij}) - g(u_i^n, u_i^n, n_{ij}) = \\ &\partial_1 g(\hat{u}_{ij}^n, \bar{u}_{ji}^n, n_{ij}) \Delta_{ij}^n u + \partial_2 g(\hat{u}_{ij}^n, \bar{u}_{ji}^n, n_{ij}) \widetilde{\Delta_{ji}^n} u, \end{aligned} \tag{14}$$

where $\partial_1 g$ and $\partial_2 g$ stand for the partial derivatives in function of the first and second argument while $\hat{u}_{ij}^n = u_i^n + \xi_0 \Delta_{ij}^n u$ and $\bar{u}_{ji}^n = u_i^n + \xi_0 \widetilde{\Delta_{ji}^n} u$.

We give the proof for a local maximum and we assume that $u_i^n \geq u_j^n$ for all $j \in \nu(i)$. Then we have $\theta_{ij}^n = 0$ hence $\Delta_{ij}^n u = 0$ thanks to relation

(13). Note that *a priori* coefficients θ_{ji} do not vanish and we still have $\widetilde{\Delta_{ji}^n u} = (1-\theta_{ji}^n)(u_j - u_i)$. The generic scheme then rewrites

$$u_i^{n+1} = u_i^n - \Delta t \sum_{j\in\nu(i)} \frac{|S_{ij}|}{|K_i|} \partial_2 g(\hat{u}_{ij}^n, \bar{u}_{ji}^n, n_{ij})(1-\theta_{ji}^n)(u_j - u_i).$$

The monotony of the flux implies that $\partial_2 g(\hat{u}_{ij}, \bar{u}_{ji}, n_{ij}) \leq 0$ while the convexity of the reconstruction yields $1-\theta_{ji}^n \geq 0$. It results from the discrete local maximum assumption that

$$\partial_2 g(\hat{u}_{ij}^n, \bar{u}_{ji}^n, n_{ij})(1-\theta_{ji})(u_j - u_i) \geq 0.$$

Hence $u_i^{n+1} \leq u_i^n$. $\square$

We now introduce the second fundamental assumption on the reconstruction operator first introduced by [3] for the particular case of the piecewise linear reconstruction.

Definition 3 (Inversion Sign property) *The reconstruction $\mathcal{R}$ has the sign inversion property if there exist a constant C_ω and coefficients $\omega_{ijk} \geq 0$, for any $K_i \in \mathcal{T}_h$, $j \in \nu(i)$, $k \in \nu(i)$ which realize*

$$u_{ij} - u_i = \Delta_{ij} u = -\sum_{k\in\nu(i)} \omega_{ijk}(u_k - u_i), \tag{15}$$

with

$$\sum_{j\in\nu(i)} \omega_{ijk} \leq C_\omega. \tag{16}$$

The expression "sign inversion" is motivated by the change of sign between the left-hand side term $u_{ij} - u_i$ and the quantities $\omega_{ijk}(u_k - u_i)$ with $\omega_{ijk} \geq 0$ in the right-hand side term.

Remark 3 A similar idea has been introduced by [14] but the coefficients α and β (we use the notations of [14], p 532) are defined in function of the collocation points x_{ijl} where u_{ijl} are supposed to be approximated. In our presentation, coefficient ω_{ijk} do not necessarily depend on the mesh even if in practice there are strongly linked to the geometry. Furthermore, we do not require that the reconstruction is a second-order (or more) method since we only deal with the stability. The sign inversion property we use here has been first proposed by [3] in the monoslope MUSCL reconstruction context.

Proposition 2 *If the reconstruction operator $\mathcal{R}$ satisfie the convexity and the sign inversion properties, then the reconstruction degenerates at the discrete local extrema.*

Proof Assume that u_i is a discrete local maximum, then $u_k - u_i \leq 0$ for all $k \in \nu(i)$. It results from relation (15) that $u_{ij} - u_i \geq 0$ for any $j \in \nu(i)$. On the other side, the convex property yields $u_{ij} = (1-\theta_{ij})u_i + \theta_{ij}u_j$ hence $u_{ij} - u_i = \theta_{ij}(u_j - u_i) \leq 0$ since $u_j - u_i \leq 0$. Consequently, we have $u_{ij} = u_i$ and $\theta_{ij} = 0$. From proposition 1, we deduce that the scheme has the LED property. □

We now deal with the definition of a positive (coefficients) scheme [16, 12, 7]. The generic scheme (9) can be written as a positive scheme if we have

$$u_i^{n+1} = u_i^n + \Delta t \sum_{j \in \nu(i)} \alpha_{ij}^n (u_j^n - u_i^n), \tag{17}$$

where α_{ij}^n are non negative coefficients depending on the approximation u_h^n at time t^n and the mesh characteristics. If one can prove that the coefficients α_{ij}^n are uniformly bounded by a constant, we shall see that under an appropriate CFL condition, u_i^{n+1} is obtained as a convex combinaison of u_j^n and u_i^n and we get the L^∞ stability of the scheme. In the sequel, we define the characteristic length of the mesh $\mathcal{T}_h$ by

$$h = \min_{\substack{K_i \in \mathcal{T}_h \\ j \in \nu(i)}} \frac{|K_i|}{|S_{ij}|}. \tag{18}$$

The following proposition gives an estimate for coefficients α_{ij}^n.

Proposition 3 *We suppose that $\mathcal{R}$ is a reconstruction which satisfies the convexity and the sign inversion properties. We also suppose that g is a numerical flux satisfying properties (a)-(d) of the subsection 2.2. Assume that $|u_i^n| \leq M$ for all $K_i \in \mathcal{T}_h$ then there exists a constant $K(M)$ and coefficients $\alpha_{ij}^n \geq 0$ with*

$$\alpha_{ij}^n \leq \frac{K(M)}{h}(1 + C_\omega), \quad \forall K_i \in \mathcal{T}_h,\ j \in \nu(i) \tag{19}$$

such that we can write the scheme as a positive scheme (17).

Proof Using the expression (14) introduced in proposition 1, we can rewrite the scheme as

$$u_i^{n+1} = u_i^n - \Delta t \sum_{j \in \nu(i)} \frac{|S_{ij}|}{|K_i|} \left(\partial_1 g(\hat{u}_{ij}^n, \bar{u}_{ji}^n, n_{ij}) \Delta_{ij}^n u + \partial_2 g(\hat{u}_{ij}^n, \bar{u}_{ji}^n, n_{ij}) \widetilde{\Delta_{ji}^n} u \right).$$

For the sake of simplicity, we introduce the notation

$$A_{ij} = \partial_1 g(\hat{u}_{ij}^n, \bar{u}_{ji}^n, n_{ij}), \quad B_{ij} = \partial_2 g(\hat{u}_{ij}^n, \bar{u}_{ji}^n, n_{ij})$$

where we skip the time index. Using the convexity property and the sign inversion property of the reconstruction, we write

$$u_i^{n+1} = u_i^n - \Delta t \sum_{j\in\nu(i)} \frac{|S_{ij}|}{|K_i|} \left(-A_{ij} \sum_{k\in\nu(i)} \omega_{ijk}(u_k^n - u_i^n) + B_{ij}(1-\theta_{ji})(u_j^n - u_i^n) \right).$$

After permutation of the summation and the indexes j and k for A_{ij} and ω_{ijk}, we get

$$u_i^{n+1} = u_i^n + \Delta t \sum_{j\in\nu(i)} (u_j^n - u_i^n) \left[\sum_{k\in\nu(i)} \frac{|S_{ik}|}{|K_i|} A_{ik}\omega_{ikj} \right] - \Delta t \sum_{j\in\nu(i)} (u_j^n - u_i^n) \frac{|S_{ij}|}{|K_i|} B_{ij}(1-\theta_{ji}).$$

We obtain the scheme (17) setting

$$\alpha_{ij}^n = \left[\sum_{k\in\nu(i)} \frac{|S_{ik}|}{|K_i|} A_{ik}\omega_{ikj} \right] - \frac{|S_{ij}|}{|K_i|} B_{ij}(1-\theta_{ji}). \tag{20}$$

Since $\omega_{ikj} \geq 0$ and $\theta_{ji} \in [0,1]$, we deduce from the flux monotony assumption that $A_{ij} \geq 0$ and $B_{ij} \leq 0$, hence $\alpha_{ij}^n \geq 0$. We now have to produce a uniform estimation for the coefficients. Since the approximations u_i^n are uniformly bounded by M, the values $\hat{u}_{ij}^n$ and $\bar{u}_{ji}^n$ are also bounded by M and the continuity of $\partial_1 g(.,.,n_{ij})$ and $\partial_2 g(.,.,n_{ij})$ with the monotony yield

$$0 \leq A_{ik} \leq K(M), \qquad 0 \leq -B_{ij} \leq K(M)$$

with

$$K(M) = \sup_{\substack{-M\leq\alpha,\beta\leq M \\ |\mathbf{n}|=1}} \Big(|\partial_1 g(\alpha,\beta,\mathbf{n})|, |\partial_2 g(\alpha,\beta,\mathbf{n})| \Big). \tag{21}$$

We then deduce using (16) that $\alpha_{ij}^n \leq \dfrac{K(M)}{h}(1+C_\omega)$. □

We conclude the section with the main theorem

Theorem 1 *Let u_i^0 be the initial approximation at time $t=0$ such that $|u_i^0| \leq M$ uniformly and assume that the following CFL condition holds*

$$\frac{\Delta t}{h} \leq \frac{1}{3K(M)(1+C_\omega)} \tag{22}$$

where $K(M)$ is given by relation (21) and C_ω by relation (16). Then u_i^n is uniformly bounded by M at any time t^n.

Proof We prove the theorem by induction. At $t = t^0 = 0$, the property holds by assumption. Let now consider that u_i^n is uniformly bounded by M at time t^n, we show that the same property holds for u_i^{n+1}. Indeed, scheme (17) can be written in the following form

$$u_i^{n+1} = \left(1 - \Delta t \sum_{j\in\nu(i)} \alpha_{ij}^n\right) u_i^n + \sum_{j\in\nu(i)} \Delta t \alpha_{ij}^n u_j^n.$$

Since $\#\nu(i) = 3$, we have

$$0 \leq \Delta t \sum_{j\in\nu(i)} \alpha_{ij}^n \leq 3\Delta t \frac{K(M)}{h}(1 + C_\omega) \leq 1.$$

We obtain a similar inequality for $\Delta t\alpha_{ij}^n$ and we deduce that u_i^{n+1} is a convex reconstruction of the neighbouring values at time t^n, hence u_i^{n+1} is bounded by M. □

Remark 4 Extension of the theroem to the three-dimensional case with tetrahedron is straigthforward and the CFL condition becomes

$$\frac{\Delta t}{h} \leq \frac{1}{4K(M)(1 + C_\omega)}$$

since $\#\nu(i) = 4$.

3 Multislope Muscl reconstruction

We present a new MUSCL technique based on the approximations of the directional derivatives. The main ingerdient is the barycentric coordinates which provide a powerful tool to manipulate the geometrical data of the mesh such as the centroïd points.

3.1 The fundamental decomposition

We introduce an assumption on the mesh which guarantees that an admissible reconstruction.

Definition 4 *Let $\alpha \in [0, 1]$, the triangulation is α-convex if for any $K_i \in \mathcal{T}_h$, there exist barycentric coordinates ρ_{ij}, $j \in \nu(i)$*

$$\sum_{j\in\nu(i)} \rho_{ij} = 1, \qquad B_i = \sum_{j\in\nu(i)} \rho_{ij} B_j \tag{23}$$

such that $\alpha \leq \rho_{ij}$.

Remark 5 The definition means that B_i lies in the triangle formed by the three points B_j and α controls the distance between B_i and the triangle edges. An extension for the three-dimensional geometries is straigthforward if we employ tetrahedron cells. The notion of α-convexity is similar to the definition of a $B-$uniform triangulation (see [6], [14], p 533) but the mesh characterization we use here is based on the barycentric coordinates.

Assuming that there is no degenerated element which is the case if $\alpha > 0$, we can define the normalized vectors (see figure 2)

$$t_{ij} = \frac{B_i B_j}{|B_i B_j|}. \tag{24}$$

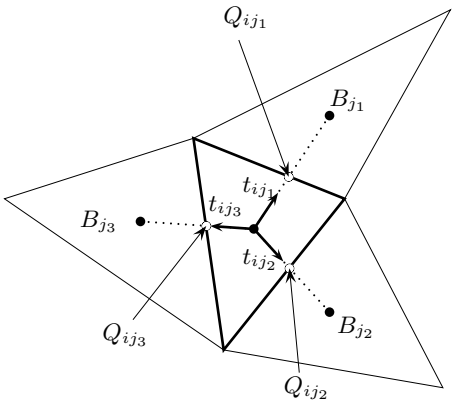

Fig. 2. Definition of vector t_{ij}.

We now introduce the fundamental decomposition which is the main tool to construct the multislope method.

Proposition 4 *Let $\mathcal{T}_h$ be a α-convex triangulation, then we have the fundamental decomposition*

$$t_{ij} = \sum_{\substack{k\in\nu(i)\\ k\neq j}} \beta_{ijk} t_{ik} \tag{25}$$

with

$$\beta_{ijk} = -\frac{\rho_{ik}|B_i B_k|}{\rho_{ij}|B_i B_j|}, \ \ \textit{if } k \neq j. \tag{26}$$

Proof Using the barycentric coordinates we have $\sum_{j\in\nu(i)} \rho_{ij} B_i B_j = 0$ hence, for $j \in \nu(i)$ fixed

$$\rho_{ij} B_i B_j = -\sum_{\substack{k\in\nu(i)\\ k\neq j}} \rho_{ik} B_i B_k.$$

Using the definition of t_{ij}, we deduce immediately

$$\rho_{ij}|B_iB_j|t_{ij} = -\sum_{\substack{k\in\nu(i)\\k\neq j}} \rho_{ik}|B_iB_k|t_{ik}.$$

Divided by $\rho_{ij}|B_iB_j|$ and the proposition is proved. □
Note that $\alpha \leq \rho_{ij}$ and $0 \leq \rho_{ik} \leq 1-\alpha$ so we get the following estimation

$$0 < -\beta_{ijk} \leq \frac{1-\alpha}{\alpha}\frac{|B_iB_k|}{|B_iB_j|}. \tag{27}$$

Remark 6 Coefficients β_{ijk} are similar to coefficients d_{ijk} proposed in [14], p. 532 but we here use the normalized directions t_{ij} instead of the vectors B_iB_j because vectors t_{ij} are more suitable to build the multislope MUSCL method.

The now introduce the following mesh parameter.

Definition 5 *Let $\mathcal{T}_h$ be a triangulation, we define*

$$\tau_{lim} = \inf_{\substack{K_i\in\mathcal{T}_h\\ j\in\nu(i)}} \frac{|B_iB_j|}{|B_iQ_{ij}|}. \tag{28}$$

Proposition 5 *We have $\tau_{lim} \in [1,2]$. Moreover, if $\tau_{lim} > 1$ then*

$$\frac{|B_iB_j|}{|B_iQ_{ij}|} \leq \frac{\tau_{lim}}{\tau_{lim}-1}. \tag{29}$$

Proof Since $Q_{ij} \in [B_i, B_j]$ we have $|B_iQ_{ij}| + |Q_{ij}B_j| = |B_iB_j|$ hence

$$\frac{1}{2}|B_iB_j| \leq \max(|B_iQ_{ij}|, |Q_{ij}B_j|) \leq |B_iB_j|.$$

It results that

$$1 \leq \min\left(\frac{|B_iB_j|}{|B_iQ_{ij}|}, \frac{|B_iB_j|}{|B_jQ_{ij}|}\right) \leq 2.$$

Taking the minimum over all the sides and we get the estimation $\tau_{lim} \in [1,2]$.

Assume now that $\tau_{lim} > 1$, then $|B_iQ_{ij}| \neq 0$ for all $K_i \in \mathcal{T}_h$, $j \in \nu(i)$. On the other hand we have $\tau_{lim}|B_iQ_{ij}| + \tau_{lim}|Q_{ij}B_j| = \tau_{lim}|B_iB_j|$ and using the fact that $\tau_{lim}|Q_{ij}B_j| \leq |B_iB_j|$ we get

$$\tau_{lim}|B_iQ_{ij}| + |B_iB_j| \geq \tau_{lim}|B_iB_j|.$$

We deduce the estimation (29) dividing by $|B_iQ_{ij}|$. □

3.2 The limiters

We recall the notion of limiter functions following the Jameson notation [12] adding some slight modifications.

Definition 6 *A limiter is a function $L : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that*

1. *for all $p, q \in \mathbb{R}$, if $pq \leq 0$ then $L(p, q) = 0$;*
2. *for all $p \in \mathbb{R}$, $L(p, p) = p$;*
3. *for all $p, q, \alpha \in \mathbb{R}$, $L(\alpha p, \alpha q) = \alpha L(p, q)$.*

Since the limiter L is a $1-$homogeneous function, there exists a function $\psi : \mathbb{R} \to \mathbb{R}$ such that

$$\forall p, q \in \mathbb{R},\ q \neq 0 \quad L(p, q) = \psi\left(\frac{p}{q}\right) q. \tag{30}$$

and we have

$$L(p, q) = qL\left(\frac{p}{q}, 1\right) = pL\left(1, \frac{q}{p}\right).$$

Moreover the condition $L(p, q) = 0$ if $pq \leq 0$ leads to $\psi(r) = 0$ if $r < 0$. At last the condition $L(p, p) = p$ leads to $\psi(1) = 1$.

Definition 7 *We say that the limiter is symmetric if*

$$\forall p, q \in \mathbb{R}, \quad L(p, q) = L(q, p). \tag{31}$$

We say that the limiter is bounded by $C \geq 0$ if

$$\forall p, q \in \mathbb{R}, \quad |L(p, q)| \leq C \min(|p|, |q|). \tag{32}$$

Assume that L is symmetric then we have

$$\psi\left(\frac{p}{q}\right) q = L(p, q) = L(q, p) = \psi\left(\frac{q}{p}\right) p.$$

Setting $r = p/q$ and we get the relation

$$r\psi\left(\frac{1}{r}\right) = \psi(r) \tag{33}$$

Proposition 6 *Assume that L is bounded then we have*

$$0 \leq \psi(r) \leq C, \quad \textit{and} \quad 0 \leq \psi(r) \leq Cr. \tag{34}$$

Proof From (32), we deduce with relation (30) that for all p, q

$$\psi\left(\frac{p}{q}\right)|q| = |L(p,q)| \leq C|q|.$$

Hence we deduce with $r = p/q$ that $\psi(r) \leq C$. On the other hand we can write

$$\psi\left(\frac{p}{q}\right)|q| = |L(p,q)| \leq C|p|$$

and dividing by q we get the second estimation.
Note that in the case that L is symmetric, the condition $\psi(r) \leq C$ implies directly the estimation $\psi(r) \leq Cr$ thanks to relation (33). □

Remark 7 We do not assume that the limiter is *a priori* symmetric. Indeed, we shall see that, in the limiting routine, the slope p has to be favoured. To produce such a behaviour, the limiter has to be asymmetric.

Remark 8 If L is a bounded limiter, the associated function ψ has to satisfy the properties

$$0 \leq \psi(r) \leq \min(C, Cr) \quad \text{and} \quad \psi(1) = 1.$$

It results that ψ belongs to a specific domain (Harten, Sweby, Van-Leer domain) controlled by the constant C. We shall present in the next section some useful limiters.

3.3 The limited slopes

The multislope MUSCL method consists in computing reconstructed values u_{ij} and u_{ji} at the collocation point Q_{ij}:

$$u_{ij} = u_i + p_{ij}|B_iQ_{ij}|, \quad u_{ji} = u_j + p_{ji}|B_jQ_{ij}|,$$

where p_{ij} and p_{ji} are approximations of the directional derivative of u following the directions t_{ij} and t_{ji} respectively. Slopes p_{ij} and p_{ji} have to be designed such that we maintain the L^∞ stability. We present here the construction of the scalar slopes.

Definition 8 (downstream and upstream slopes) *We define the downstream slopes from point B_i in direction t_{ij},, $j \in \nu(i)$ by*

$$p_{ij}^+ = \frac{u_j - u_i}{|B_iB_j|} \tag{35}$$

and the upstream slopes by

$$p_{ij}^- = \sum_{\substack{k\in\nu(i)\\ k\neq j}} \beta_{ijk}\, p_{ik}^+. \tag{36}$$

Note that the definition can be easily extended to the three-dimensional case if we employ tetrahedron cells.

Remark 9 Let us consider an observer located at point B_i looking in direction t_{ij}. The value p_{ij}^+ represents the coefficient of the slope that the observer can see in front of him (the downstream slope) while p_{ij}^- represents the slope behind him (the upstream slope).

Let L be a bounded limiter we then define the reconstruction $\mathcal{R}$ by

$$p_{ij} = L(p_{ij}^+, p_{ij}^-), \quad u_{ij} = u_i + p_{ij}|B_iQ_{ij}|. \tag{37}$$

We have a second-order reconstruction in the following sense.

Proposition 7 *The reconstruction is consistent for linear function : if $u \in \mathbb{P}_1$ is a first-order polynomial function and define $u_i = u(B_i)$ for all $K_i \in \widetilde{\mathcal{T}_h}$ then $u_{ij} = u(Q_{ij})$ for all $K_i \in \mathcal{T}_h$ and $j \in \nu(i)$.*

Proof Function u write $u(X) = u(B_i) + a.B_iX$ for any $X \in \mathbb{R}$. The downstream slopes then are

$$p_{ij}^+ = \frac{u_j - u_i}{|B_iB_j|} = \frac{a.B_iB_j}{|B_iB_j|} = a.t_{ij}.$$

On the other hand, we have for the upstream slopes using the fundamental decomposition (25)

$$p_{ij}^- = \sum_{\substack{k\in\nu(i)\\k\neq j}} \beta_{ijk}\, p_{ik}^+ = a. \sum_{\substack{k\in\nu(i)\\k\neq j}} \beta_{ijk}\, t_{ik} = a.t_{ij}.$$

It results that $p_{ij} = p_{ij}^+ = p_{ij}^- = a.t_{ij}$ and we get

$$u_{ij} = u(B_i) + a.t_{ij}|B_iQ_{ij}| = u(B_i) + a.B_iQ_{ij}.$$

Hence $u_{ij} = u(Q_{ij})$. □

Remark 10 The multislope method is based on the two slopes p_{ij}^+ and p_{ij}^- but they do not play the same role. p_{ij}^+ is the predicted slope that should be used if no limiter is employed whereas p_{ij}^- is used to modify the predicted slope in order to preserve the stability therefore p_{ij}^+ and p_{ij}^- do not play a symmetric role. Moreover, since Q_{ij} belongs to the segment $[B_i, B_j]$, we obtain a better approximation computing u_{ij} with p_{ij}^+ than p_{ij}^-. These reasons motivate the interest to an asymmetric limiter $L(p_{ij}^+, p_{ij}^-)$ which favoures p_{ij}^+.

Proposition 8 *Let $\mathcal{T}_h$ be a α-convex triangulation and assume that L is a bounded limiter with the bound $C = \tau_{lim}$. Then the reconstruction has the convexity and sign inversion properties with*

$$C_\omega = \frac{2}{\alpha}.$$

Proof To prove the convexity property, we have to show that

$$u_{ij} \in [\min(u_i, u_j), \max(u_i, u_j)].$$

Let p_{ij}^+, p_{ij}^- be given by relations (35) and (36). If $p_{ij}^+ p_{ij}^- \leq 0$, the limiting procedure yields $p_{ij} = 0$ hence $u_{ij} = u_i$. Let us now consider the other situation where $p_{ij}^+ p_{ij}^- > 0$ and assume, for example, that $u_i > u_j$ such that $p_{ij}^+ > 0$. By definition we write $u_{ij} = u_i + p_{ij}|B_iQ_{ij}| \geq u_i$. On the other hand, using the fact that L is bounded by τ_{lim}, we can write $0 \leq p_{ij} \leq \tau_{lim} p_{ij}^+$ hence

$$\begin{aligned} u_{ij} = u_i + p_{ij}\frac{|B_iQ_{ij}|}{|B_iB_j|}|B_iB_j| &\leq u_i + p_{ij}^+|B_iB_j|\tau_{lim}\frac{|B_iQ_{ij}|}{|B_iB_j|} \\ &\leq u_i + p_{ij}^+|B_iB_j| = u_j. \end{aligned}$$

We now deal with the sign inversion property. One more time, if $p_{ij}^+ p_{ij}^- \leq 0$, the limiting procedure yields $p_{ij} = 0$ hence we have $u_{ij} = u_i$ and we set $\omega_{ijk} = 0$ for all $k \in \nu(i)$. Assume now that $p_{ij}^+ p_{ij}^- > 0$ and consider the case where $p_{ij}^+ > 0$, we can write

$$u_{ij} = u_i + p_{ij}^- \psi\left(\frac{p_{ij}^+}{p_{ij}^-}\right)|B_iQ_{ij}| = u_i + \psi\left(\frac{p_{ij}^+}{p_{ij}^-}\right)\sum_{\substack{k\in\nu(i)\\k\neq j}} \beta_{ijk} p_{ik}^+ . |B_iQ_{ij}|.$$

Setting $\psi_{ij} = \psi\left(\dfrac{p_{ij}^+}{p_{ij}^-}\right)$, we have noting that $u_k - u_i = p_{ik}^+|B_iB_k|$

$$u_{ij} = u_i + \psi_{ij}\sum_{\substack{k\in\nu(i)\\k\neq j}} \beta_{ijk}\frac{|B_iQ_{ij}|}{|B_iB_k|}p_{ik}^+ . |B_iB_k| = u_i - \sum_{\substack{k\in\nu(i)\\k\neq j}} \omega_{ijk}(u_k - u_i)$$

where we have set

$$\omega_{ijk} = -\psi_{ij}\beta_{ijk}\frac{|B_iQ_{ij}|}{|B_iB_k|}.$$

Since we have $\beta_{ijk} \leq 0$ and $\psi_{ij} \geq 0$, we deduce that $\omega_{ijk} \geq 0$. Furthermore, we have with relation (26)

$$\sum_{j\in\nu(i)} \omega_{ijk} = \sum_{\substack{j\in\nu(i)\\j\neq k}} \psi_{ij}\frac{\rho_{ik}|B_iB_k|}{\rho_{ij}|B_iB_j|}\frac{|B_iQ_{ij}|}{|B_iB_k|} = \sum_{\substack{j\in\nu(i)\\j\neq k}} \psi_{ij}\frac{\rho_{ik}}{\rho_{ij}}\frac{|B_iQ_{ij}|}{|B_iB_j|}.$$

where we state $\omega_{ijj}=0$ by convention.
The limiter is bounded by τ_{lim}, then we get

$$\sum_{j\in\nu(i)}\omega_{ijk}\leq\sum_{\substack{j\in\nu(i)\\ j\neq k}}\frac{\rho_{ik}}{\rho_{ij}}\leq\frac{2}{\alpha}$$

because $\alpha\leq\rho_{ij}$, $\rho_{ik}\leq 1$ and $\#\nu(i)=3$. Hence we get the sign inversion property with $C_\omega=\dfrac{2}{\alpha}$. $\square$

Remark 11 A similar proof is obtained for the three-dimensional situation and we have the estimate

$$C_\omega=\frac{3}{\alpha}.$$

Theorem 2 *Let $\mathcal{T}_h$ be a α-convex triangulation. Then the finite volume scheme (9) based on the reconstruction (37) and a limiter (30) bounded by τ_{lim} is L^∞ stable under the CFL condition*

$$\frac{\Delta t}{h}\leq\frac{\alpha}{6K(M)}\tag{38}$$

Proof The reconstruction satisfies the convexity and sign inversion properties with $C_\omega=\frac{2}{\alpha}$. Theorem (1) yields that the scheme is L^∞ stable under the CFL constraint

$$\frac{\Delta t}{h}\leq\frac{1}{3K(M)(1+C_\omega)}$$

Using $C_\omega=\frac{2}{\alpha}$ and we get estimation (38). $\square$
In the three-dimensional case, we obtain the CFL condition

$$\frac{\Delta t}{h}\leq\frac{\alpha}{12K(M)}$$

Corollary 1 *Let $\mathcal{R}$ be a reconstruction and assume that there exists a limiter L bounded by $\tau_{\rm lim}$ such that we can write*

$$u_{ij}=u_i+L(p_{ij}^+p_{ij}^-)|B_iQ_{ij}|.$$

Then the finite volume scheme (9) is L^∞ stable under the CFL condition (38).

To end this section, we propose some limiters which are bounded by τ_{lim}. The first one is the minmod limiter, symmetric bounded by $C=1$:

$$\text{minmod}(p,q)=\begin{cases}0, & \text{if } pq\leq 0,\\ \text{sign}(p)\min(|p|,|q|), & \text{if } pq>0.\end{cases}\tag{39}$$

We define a class of τ-limiter for $\tau \in]1,2]$ which are an extension of classical limiter for the one-dimensional problem:

the τ-Bee limiter (see [12]) is a symmetric limiter bounded by τ

$$L_{SB}(p,q) = q \times \max(0, \min(1, \tau r), \min(r, \tau)), \quad r = \frac{p}{q};$$

the τ-Van Leer limiter is a symmetric limiter bounded by τ

$$L_{VL}(p,q) = q \times \begin{cases} \dfrac{r + (\tau-1)r}{(\tau-1)+r}, & \text{if } r \geq 1, \\ \dfrac{r + (\tau-1)r}{1+(\tau-1)r}, & \text{if } 0 \leq r \leq 1, \\ 0, & \text{if } r \leq 0; \end{cases}$$

the τ-Van Alabada limiter is a symmetric limiter bounded by τ

$$L_{VA}(p,q) = q \times \begin{cases} \dfrac{r + (\tau-1)r^2}{(\tau-1)+r^2}, & \text{if } r \geq 1, \\ \dfrac{r + (\tau-1)r^2}{1+(\tau-1)r^2}, & \text{if } 0 \leq r \leq 1, \\ 0, & \text{if } r \leq 0; \end{cases}$$

the τ-minmod limiter is an asymmetric limiter bounded by τ

$$\tau\text{-minmod}(p,q) = q \times \begin{cases} 0, & \text{if } r \leq 0, \\ \min(\tau, r), & \text{if } r > 0. \end{cases}$$

Note that all the τ-limiters converge to the minmod limiter when τ converges to 1. The minmod limiter does not depend of τ and it is a good candidate to provide stability when the mesh is strongly deformed.

4 Monoslopes Muscl reconstruction

We now deal with the classical linear piecewise reconstruction. For a given set $(u_i)_{K_i \in \widetilde{\mathcal{T}_h}}$, we define on each $K_i \in \mathcal{T}_h$ the function $u_h(X) = u_i + a_i.B_iX$ for all $X \in K_i$ where $a_i \in \mathbb{R}^2$ depends on u_i and u_j, $j \in \nu(i)$. We then define the reconstruction $\mathcal{R}$ by the operator

$$(u_i)_{K_i \in \widetilde{\mathcal{T}_h}} \xrightarrow{\mathcal{R}} (u_{ij})_{K_i \in \mathcal{T}_h, j \in \nu(i)}$$

where u_{ij} are defined by

$$u_{ij} = u_i + a_i.B_iQ_{ij}. \tag{40}$$

To provide the stability, slopes a_i have to satisfy a specific constraint to keep the scheme from producing oscillations. To this end, we introduce two close subsets of $\mathbb{R}^2$ named the stability domains (TVD and MP domain) which characterize the stability condition that the slopes have to respect.

Definition 9 *Let $K_i \in \mathcal{T}_h$ and $u_i, u_j \in \mathbb{R}$, $j \in \nu(i)$.*
We define the Total Variation Diminishing domain for element K_i (TVD_i *domain)*

$$\text{TVD}_i = \{a \in \mathbb{R}^2;\ \min(u_i, u_j) \leq u_i + a.B_iB_j \leq \max(u_i, u_j), j \in \nu(i)\}. \tag{41}$$

We define the Maximum Principle domain for element K_i (MP_i *domain)*

$$\text{MP}_i = \{a \in \mathbb{R}^2;\ \min(u_i, u_j) \leq u_i + a.B_iQ_{ij} \leq \max(u_i, u_j), j \in \nu(i)\}. \tag{42}$$

An extension to the three-dimensional situation is clear. Moreover, we have $\text{TVD}_i \subset \text{MP}_i$.

Proposition 9 *If $a_i \in \text{TVD}_i$ or $a_i \in \text{MP}_i$ then the reconstruction has the convexity property.*

Proof Let $a_i \in \text{TVD}_i$, since $|a_i.B_iQ_{ij}| \leq |a_i.B_iB_j|$ we deduce that

$$u_{ij} \in [\min(u_i, u_j), \max(u_i, u_j)].$$

If $a_i \in \text{MP}_i$, we have immediately by definition $u_{ij} \in [\min(u_i, u_j), \max(u_i, u_j)]$. Hence u_{ij} can be written as a convex combinaison between u_i and u_j. □

We now prove the invesion sign property. We first deal with the TVD domain.

Proposition 10 *Let $\mathcal{T}_h$ be a α-convex triangulation. If $a_i \in \text{TVD}_i$ then the reconstruction has the sign inversion property with*

$$C_\omega = \frac{2}{\alpha}.$$

Proof Using the fundamental decomposition (25), we write

$$\begin{aligned} u_{ij} = u_i + a_i.|B_iQ_{ij}|t_{ij} &= u_i + a_i.\left(\sum_{\substack{k\in\nu(i)\\k\neq j}} \beta_{ijk}|B_iQ_{ij}|t_{ik}\right) \\ &= u_i + \sum_{\substack{k\in\nu(i)\\k\neq j}} \beta_{ijk}\frac{|B_iQ_{ij}|}{|B_iB_k|}a_i.B_iB_k. \end{aligned}$$

Since $a_i \in \text{TVD}_i$ we have

$$\min(0, u_k - u_i) \leq a_{ik}.B_iB_k \leq \max(0, u_k - u_i), \forall k \in \nu(i).$$

Assume first that $u_k \neq u_i$, we then have

$$0 \leq \frac{a_{ik}.B_iB_k}{u_k - u_i} \leq 1.$$

On the other hand, if $u_k = u_i$, we have $a_{ik}.B_iB_k = u_k - u_i = 0$ and we adopt the convention

$$\frac{a_{ik}.B_iB_k}{u_k - u_i} = 1.$$

We then deduce with relation (26)

$$u_{ij} = u_i - \sum_{\substack{k \in \nu(i) \\ k \neq j}} \omega_{ijk}(u_k - u_i).$$

where

$$\omega_{ijk} = \frac{\rho_{ik}|B_iB_k|}{\rho_{ij}|B_iB_j|} \frac{|B_iQ_{ij}|}{|B_iB_k|} \frac{a_i.B_iB_k}{u_k - u_i} = \frac{\rho_{ik}}{\rho_{ij}} \frac{|B_iQ_{ij}|}{|B_iB_j|} \frac{a_i.B_iB_k}{u_k - u_i} \geq 0.$$

Since $\mathcal{T}_h$ is a α-convex triangulation and $|B_iQ_{ij}| \leq |B_iB_j|$, we have

$$\sum_{\substack{j \in \nu(i) \\ j \neq k}} \omega_{ijk} \leq \sum_{\substack{j \in \nu(i) \\ j \neq k}} \frac{\rho_{ik}}{\rho_{ij}} \leq \frac{2}{\alpha} = C_\omega. \qquad \square$$

To prove the sign inversion property with the MP domain we need a restriction on the mesh.

Proposition 11 *Let $\mathcal{T}_h$ be a α-convex triangulation and assume that $\tau_{lim} > 1$. If $a_i \in \mathrm{MP}_i$ then the reconstruction has the sign inversion property with*

$$C_\omega = \frac{2}{\alpha(\tau_{lim} - 1)}.$$

Proof As in propostion (10), we write

$$u_{ij} = u_i - \sum_{k \in \nu(i)} \omega_{ijk}(u_k - u_i).$$

where

$$\omega_{ijk} = \frac{\rho_{ik}|B_iB_k|}{\rho_{ij}|B_iB_j|} \frac{|B_iQ_{ij}|}{|B_iQ_{ik}|} \frac{a_i.B_iQ_{ik}}{u_k - u_i} \geq 0.$$

On one hand, since $a_i \in \mathrm{MP}_i$ we have

$$0 \leq \frac{a_i.B_iQ_{ik}}{u_k - u_i} \leq 1.$$

On the other hand we have with relation (29)

$$\frac{|B_iB_k|}{|B_iQ_{ik}|} \leq \frac{\tau_{lim}}{\tau_{lim}-1}, \qquad \tau_{lim}\frac{|B_iQ_{ij}|}{|B_iB_j|} \leq 1,$$

and we get

$$\omega_{ijk} \leq \frac{\rho_{ik}}{\rho_{ij}}\frac{1}{\tau_{lim}-1}.$$

Using the α-convexity of the triangulation, we obtain estimation (11). □
We now give the two stability results where we use the TVD domain or the MP domain

Theorem 3 *Let $\mathcal{T}_h$ be a α-convex triangulation. Then the finite volume scheme (9) based on the reconstruction (40) such that a_i satisfies (41) is L^∞ stable under the CFL condition*

$$\frac{\Delta t}{h} \leq \frac{\alpha}{6K(M)}. \tag{43}$$

Theorem 4 *Let $\mathcal{T}_h$ be a α-convex triangulation. Then the finite volume scheme (9) based on the reconstruction (40) such that a_i satisfies (42) is L^∞ stable under the CFL condition*

$$\frac{\Delta t}{h} \leq \frac{\alpha}{6K(M)}(\tau_{lim}-1). \tag{44}$$

We give here a classical example of reconstruction using the monoslope MUSCL technique where the slopes a_i belong to MP_i or TVD_i (see [11] for an overview of the classical MUSCL reconstruction).
In $\mathbb{R}^3$, we consider the four points (B_i, u_i) and (B_j, u_j), $j \in \nu(i)$. We define the hyperplane $z = \pi(x_1, x_2)$ which contain the three points (B_j, u_j), $j \in \nu(i)$ and denote by $\widetilde{a}_i \in \mathbb{R}^2$ the gradient of π. Since $\widetilde{a}_i$ does not belong to MP_i *a priori*, we modify the slope in the following way. We first compute a limiter in each direction

$$\phi_{ij} = \begin{cases} 0, & \text{if } \widetilde{a}_i.B_iB_j(u_j-u_i) \leq 0, \\ \min\left(1, \dfrac{\widetilde{a}_i.B_iB_j}{u_j-u_i}\right), & \text{if } \widetilde{a}_i.B_iB_j(u_j-u_i) > 0. \end{cases} \tag{45}$$

We then set

$$a_i = \min_{j\in\nu(i)}(\phi_{ij})\widetilde{a}_i \in \text{MP}_i. \tag{46}$$

4.1 A comparison between the monoslope and the multislope MUSCL methods

We present here a comparison between the reconstruction (40) using the monoslope method (45,46) and the reconstruction (37) using the multislope method with the minmod limiter (39). We consider the simple mesh $\mathcal{T}_h$ compose of three equilateral triangles where K_1 is the central element while K_2, K_3 and K_4 are the three neighbouring elements with common sides S_{12}, S_{13}, S_{14} respectively (see figure 3). We also assume that $|B_1B_j| = 1$ for all $j \in \nu(1) = \{2, 3, 4\}$. We denote by $u_1, \ldots, u_4$ the approximations of u on each cell.

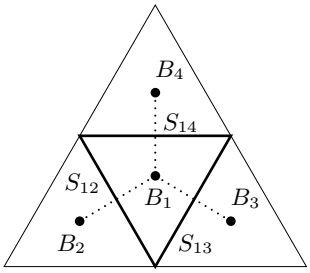

Fig. 3. The mesh constituted with four equilateral triangles.

We first deal with the monoslope reconstruction. For element K_1, the predicted slope $\widetilde{a}_1 \in \mathbb{R}^2$ satisfies

$$\widetilde{a}_1.B_jB_k = u_k - u_j, \quad j, k \in \nu(1).$$

Let $j \in \nu(1)$, using the barycentric coordinates, we write

$$B_1B_j = \sum_{\substack{k\in\nu(1)\\ k\neq k}} \rho_{1k} B_kB_j$$

and we deduce

$$\widetilde{a}_1.B_1B_j = \sum_{\substack{k\in\nu(1)\\ k\neq j}} \rho_{1k}(u_j - u_k) = \sum_{k\in\nu(1)} \rho_{1k}(u_j - u_k).$$

For $j \in \nu(1)$, let ϕ_{1j} be defined by relation (45) and consider the quantities

$$D_{1j} = \widetilde{a}_1.B_1B_j(u_j - u_1) = \sum_{k\in\nu(1)} \rho_{1k}(u_j - u_k)(u_j - u_1).$$

Relation (45) yields that $\phi_{1j} = 0$ if and only if $\widetilde{a}_1.B_1B_j(u_j - u_1) \leq 0$ which corresponds to $D_{1j} \leq 0$.
In the particular case where the triangles are equilateral we have $\rho_{1k} = \frac{1}{3}$ and we get

$$D_{1j} = \frac{1}{3} \sum_{k \in \nu(1)} (u_j - u_k)(u_j - u_1).$$

We now proceed using particular values for u_j setting $u_1 = 0$, $u_2 = \alpha > 0$, $u_3 = -\alpha$ and $u_4 = \beta$. A short calculation gives

$$D_{12} = \frac{1}{3}(3\alpha - \beta)\alpha.$$

We then deduce that $D_{12} \leq 0$ if $\beta \geq 3\alpha$. It result that $\phi_{12} = 0$ hence $\phi_i = 0$. Vector a_i is the null vector and the scheme is reduced to a first-order one.

We now consider the multislope reconstruction. The downstream slopes are given by

$$p_{12}^+ = \alpha, \quad p_{13}^+ = -\alpha, \quad p_{14}^+ = \beta$$

while the upstream slopes are given by

$$p_{12}^- = \alpha - \beta, \quad p_{13}^- = -\alpha - \beta, \quad p_{14}^- = 0.$$

Using the minmod limiter, we found if $\beta \geq 3\alpha$

$$p_{12} = 0, \quad p_{13} = -\alpha, \quad p_{14} = 0.$$

The scheme does not degenerate since $p_{13} \neq 0$. In conclusion we have obtain a configuration where the monoslope MUSCL method degenerates *i.e.* the slope is reduced to zero while the multislope MUSCL method is still efficient.

5 Conclusion

We have considered a generic finite volume method for hyperbolic scalar equations coupled with a large class of reconstruction where the numerical flux across the interface is computed using the reconstructed values to enhance the scheme and produce more accurate approximations. Based on two fundamental assumptions, namely the convexity and the sign inversion properties, we have obtained the L^∞ stability when a monotone numerical flux function is employed.

Two applications of the stability result have been proposed. We have first introduced the multislope MUSCL technique and show that the two fondamental properties are satisfied under a specific constraint for the mesh: the α-convexity. We also show the stability of the monoslope (classical)

MUSCL methods under a condition on the reconstruction: the MP or TVD constraints. The principle ingredient employed in the reconstructions is the use of the barycentric coordinates which are a powerful tool to manipulate the geometrical data of the mesh. To obtain L^∞ stable MUSCL methods, two characteristic mesh parameters have been brought to the fore: the α parameter which controls the regularity of the mesh and the CFL condition; the τ_{lim} parameter which controls the limiters. For the two dimensional situation, it is rather easy to produce a triangulation which satisfies the α-convexity (Delaunay triangulation for instance) but the three-dimensional situation is more complex. All the meshes we have experimented does not satisfy the α-convexity. Indeed, there always exists a very small number of cells which provide negative barycentric coordinates. In order to use a second-order scheme even if the mesh is not α-convex, we cancel the reconstruction for the cells over which the barycentric coordinates are lower than a prescribed value of α and the L^∞ stability is then preserved. In practice, less that 1% of the cells are affected.

References

1. T. J. Barth, M. Ohlberger, Finite volume methods: foundation and analysis, Volume 1, chapter 15, Encyclopedia of Computational Mechanics, John Wiley & Sons Ltd, (2004).
2. T. Buffard, S. Clain, Monoslope and Multislope MUSCL Methods for unstructured meshes, preprint available at `http://hal.archives-ouvertes.fr/hal-00323691/fr/`, (2008).
3. C. Chainais-Hillairet , Second order finite volume schemes for a nonlinear hyperbolic equation : error estimate, M2AS **23**, 467–490 (2000).
4. S. Clain, V. Clauzon, The multislope MUSCL method, Proceeding in the Finite Volumes for Complex Application 5, Wiley, 297–304 (2008).
5. V. Clauzon, Analyse de Schémas d'ordre élevé pour les écoulements compressibles. Application à la simulation numérique d'une torche à plasma, PhD Thesis, Blaise Pascal University, Clermont Ferrand, Thesis available at `http://tel.archives-ouvertes.fr/tel-00235951/fr/` (2008).
6. B. Cockburn, , Hou S., Shu C.W., The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation law IV: the multidimensional cas, Math. of Compt. **54**, 545–581 (1990).
7. P.-H. Cournède, C. Debiez, A. Dervieux, A positive MUSCL scheme for triangulations, INRIA report 3465 (1998).
8. R. Eymard, T. Galloüet, R. Herbin, Finite Volume Methods, Handbook of Numerical Analysis, Vol. VII, Editors: P.G. Ciarlet and J.L. Lions, North Holland, 713–1020 (2000).
9. E. Godlewski and P.-A. Raviart, *Numerical Approximation of hyperbolic systems of conservation law*, Applied Mathematical sciences Springer **118** (1996).
10. J. B. Goodman, R. J. Leveque, On the accuracy of stable schemes for 2D scalar conversation laws, Mathematics of computation **45**(171), 15–21 (1985).
11. M. E. Hubbard, Multidimensional slope limiters for MUSCL-type finite volume schemes on unstructured grids, J. Comput. Phys. **155**(1), 54–74 (1999).

12. A. Jameson, Artificial diffusion, upwind biasing, limiters and their effect on accuracy and multigrid convergence in transonic and hypersonic flows, AIAA paper No 3359 (1993).
13. D. Kröner, *Numerical schemes for conservation lows* (Wiley Teubner 1997)
14. 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), 527–560 (1995).
15. S. Osher, Convergence of generalized MUSCL schemes, SIAM J. Numer. Anal. **22**(5), 947–961 (1985).
16. S. P. Spekreijse, Multigrid solution of monotone second-order discretizations of hyperbolic conservation laws, Math. Comp. **49**(179), 135–155 (1987).
17. B. Van Leer, Towards the ultimate conservative difference scheme II, Monotonicity and conservation combined in a second order scheme, J. Comp. Phys. **14** 361–376 (1974).