High order fluctuation schemes on triangular meshes

R. Abgrall† and P.L. Roe‡
† Mathématiques Appliquées de Bordeaux,
Université Bordeaux I, 33 405 Talence Cedex, France
and Institut Universitaire de France
‡ W. M. Keck Laboratory for Computational Fluid Dynamics,
Department of Aerospace Engineering,
University of Michigan, Ann Arbor, MI 48109, USA

Draft May 3, 2002

Abstract

We develop a new class of schemes for the numerical solution of first-order steady conservation laws. The schemes are of the residual distribution, or fluctuation-splitting type. These schemes have mostly been developed in the context of triangular or tetrahedral elements whose degrees of freedom are their nodal values. We work here with more general elements that allow high-order accuracy. We introduce, for an arbitrary number of degrees of freedom, a simple mapping from a low-order monotone scheme to a monotone scheme that is as accurate as the degrees of freedom will allow. Proofs of consistency, convergence and accuracy are presented, and numerical examples from second, third and fourth-order schemes.

1 Introduction

In this paper, we consider high order discretisations of the problem

$$\text{div } (f(u)) = 0 \quad x \in \Omega \tag{1}$$

with boundary conditions on the inflow part of the domain $\Omega \subset \mathbb{R}^d$. Here, f is a C^2 real valued function, and u is a vector valued function. We consider entropy weak solutions of (1). An approximate solution of (1) is sought on a triangular mesh τ_h. The nodes or τ_h are denoted by $\{M_i\}$ and T is a generic triangle.

There are many schemes for approximating (1), but here we consider upwind residual schemes, as in [12, 4, 7]. These schemes are distributive schemes; in their simplest forms they can be written, in terms of an iteration counter n, as

$$|C_i|\frac{u_i^{n+1}-u_i^n}{\Delta t} + \sum_{T, M_i\in T} \Phi_i^T = 0 \tag{2}$$

and in this paper we seek steady solutions independent of n. Here, $|C_i|$ is the area of the dual cell associated with M_i. The sum is over those triangles T that share the vertex M_i and on each triangle T we define a residual

$$\Phi^T = \int_T \text{div } f^h(u_h) dx. \tag{3}$$

We also define Φ_i^T as the amount of Φ^T associated with vertex M_i, such that a conservation property is satisfied

$$\sum_{i, M_i\in T} \Phi_i^T = \Phi^T \tag{4}$$

Roe [11] refers to Φ^T as the fluctuation in triangle T and to the quantities Φ_i^T as signals sent to the vertices. One major objective of this paper is to generalize this form. Each triangle may be equipped with additional degrees of freedom in addition to the values at the vertices. For example, values can be assigned to the mid-points of the sides. This will be explained more fully in Section 2.1

In this paper, we give precise conditions on f^h, $\{\Phi_j^T\}_{M_j,T}$ for which, under the conditions of a Lax Wendroff theorem, the limit solution is a weak solution of (1). Then, we provide conditions for which the solution is formally high order accurate. In a second part, we consider from this perspective some known examples having second order accuracy, including some that are monotonicity preserving. Then we show a stability property of these schemes, in the case $f(u) = \vec{\lambda} u$, that yields an error estimate in some cases. Using these principles, we construct third- and fourth- order monotonicity preserving schemes and give some numerical applications. This construction can be formally extended to any order of accuracy, so these schemes can be seen as an alternative to the DG schemes. However, we believe that our schemes will prove less expensive, especially for very high-order and/or three-dimensional cases.

2 Notations and basic concepts

2.1 Notation and degrees of freedom

We explain our notation, and specify the degrees of freedom. In a mesh τ whose triangles are denoted by T, and whose vertices by M_j, we seek a solution that is piecewise polynomial of degree k in each triangle, and therefore need to provide each T with $(k+1)(k+2)/2$ degrees of freedom. A simple way to do this is to use P^k elements whose degrees of freedom are the values of the solution at the points whose barycentric coordinates within T are $\{i/k, j/k, (k-i-j)/k\}$ for $0 \leq i, j \leq k$. This creates k^2 subtriangles within T; denote a typical subtriangle by T'. See Figure 1 for the cases $k=1$ and $k=2$.

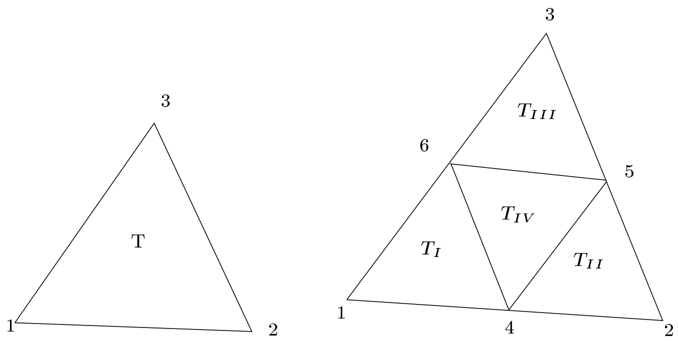


Figure 1: Sub–triangles of P^1 and P^2 interpolation.

This way of adding points in T provide a natural sub–triangulation of T, and we denote by T'_T a generic sub–triangle. The collection of all these sub–triangles furnishes a natural conformal refined triangulation of Ω.

We denote by σ a point that may belong to the set M_j, or which may be one of the additional points introduced by the sub–triangulation. The schemes that we study here are of the following form : for any degree of freedom σ

$$\begin{aligned} u_\sigma^{n+1} &= u_\sigma^n - \frac{\Delta t}{|C'_\sigma|} \textstyle\sum_{T,\sigma\in T} \Psi_\sigma^T \\ \Psi_\sigma^T &= \sum_{T'_T\subset T, \sigma\in T'_T} \Phi_\sigma^{T'_T} \end{aligned} \tag{5}$$

Once more, C'_σ is the dual control volume associated with any new or original point σ of the refined mesh. The residuals $\{\Phi_\sigma^{T'_T}\}$ satisfy the conservation relation

$$\sum_{T'_T\subset T, \sigma\in T} \Phi_\sigma^{T'_T} = \int_T \text{div } f^h(u^h)dx.$$

For k=1, the only subtriangle of T is t itself; for $k=2$ the subtriangles are $T_I, T_{II}, T_{III}, T_{IV}$, and so on.

2.2 Consistency, accuracy

Very natural questions are then

- Under which assumptions do we have consistency, and can we exhibit a Lax–Wendroff theorem under reasonable assumptions,
- What accuracy can we reach in principle and how could this be achieved in practice?

In this section, we assume that the problem

$$\frac{\partial u}{\partial t} + \operatorname{div} f(u) = 0 \tag{6}$$

is vector valued : $u \in \mathbb{R}^p$ is defined on $\mathbb{R}^d$. The flux $f = (f_1, \cdots, f_d)$ is also defined on $\mathbb{R}^d$, and $f_i \in \mathbb{R}^p$.

2.2.1 A Lax Wendroff theorem

Under which reasonable conditions can the scheme (5) converge towards a weak solution of (6) ? To answer this question, we give a Lax Wendroff–like theorem. Let us give some conditions on the mesh, the fluctuations $\Phi_\sigma^{T'_T}$, and define some notations and functional spaces.

Assumption 2.1 (H0). *The mesh is is conformal and regular. By regular we mean that all triangles are roughly the same size, more precisely that there exist constants C_1 and C_2 such that for any triangle*

$$T,\ \ C_1 \leq \sup_{T\in\tau_h} \frac{h^2}{|T|} \leq C_2$$

where for this purpose the triangles considered are those at the finest level of subtriangulation. Of course, if the original triangulation is regular in this sense, so is the subtriangulation. We say that two triangles are neighbors if they have a common edge

Let τ_h be a triangulation satisfying (H0), and $\mathcal{C}_h$ be a set of dual volumes associated with the vertices of τ_h. Then we define the following subspaces,

$$V_h^k = \{v_h \in C^0(\mathbb{R}^2)^p; v_{h|T} \text{ polynomial of degree k }, \forall\ T \in \tau_h\}$$

$$X^h = \{v_h; v_{h|C} \text{ constant } \in \mathbb{R}^{dp}, \forall\ C \in \mathcal{C}_h\}.$$

Here, $f_{|T}$ denotes the restriction of f to T.

We denote by $\pi_h v$ the piecewise linear interpolation of a continuous function. Let $L_h\ : V_h^k \to X^h$ be the mass lumping operator, $L_h(v) = \sum_i v(x_\sigma)\chi_\sigma$ where χ_σ is the characteristic function of the cell C_σ.

We assume that the residual $\Phi_\sigma^{T'_T}$ and the numerical solution satisfy the following conditions.

Assumption 2.2 (H1). *Let τ_h be a triangulation satisfying (H0). For any $C \in \mathbb{R}^+$, there exists $C'(C,\tau_h) \in \mathbb{R}^+$ which depends only on C and τ_h such that for any $u \in (X^h)^2$, with $||u||_{L^\infty(\mathbb{R}^2)} \leq C$ we have*

$$\forall T, \forall\sigma, x_\sigma \in T, \left|\left|\Phi_\sigma^{T'_T}\right|\right| \leq C'(C,\tau_h)\ h \sum_{\sigma'\in T} ||u_{\sigma'} - u_\sigma|| \,. \tag{7}$$

Remark 2.1. *One should see this assumption as asserting continuity of the residual components (or signals) $\Phi_\sigma^{T'_T}$ with respect of the nodal values of u ; in particular, when u is constant, $\Phi_\sigma^{T'_T} = 0$. The proof of theorem 2.2 when $\Phi_\sigma^{T'_T}$ satisfy (H1) is still valid if the number of arguments in $\Phi_\sigma^{T'_T}$ is bounded independently of h and the triangle T. In practice, this is always true if the triangulation is uniform, since the arguments of $\Phi_\sigma^{T'_T}$ are contained in some neighborhood of σ comprising a finite number of points.*

Assumption 2.3 (H2). *There exists an approximation f^h of the flux f such that*

(i) $\forall u^h \in X^h$, $\Phi^T := \int_T \text{div } f^h(u^h)dx = \sum_{T'_T \subset T} \sum_{\sigma \in T'_T} \Phi_\sigma^{T'_T}(u^h)$,

(ii) $\forall u^h \in X^h$, $\forall T_1, T_2$ *neigbors,*

$$f^h(u^h)_{|T_1} \, . \, \vec{n} = f^h(u^h)_{|T_2} \, . \, \vec{n} \text{ p.p. on } T_1 \cap T_2$$

where $\vec{n}$ is a normal of $T_1 \cap T_2$.

(iii) For any $C > 0$, there exists $C'(C)$ such that for any $u^h \in X^h$ with $||u^h||_{L^\infty(\mathbb{R}^2)} \leq C$, one has for $T \in \tau_h$ and $f_T^h = f_{|T}^h$, $|| \text{ div } f_T^h(u^h)|| \leq \frac{C'}{h} \sum_{\sigma, \sigma'} ||u_\sigma^h - u_{\sigma'}^h||$ *a.e. on T.*

(iv) For any sequence $(u^h)_h$ such that (u^h) is bounded in $L^\infty(\mathbb{R}^2 \times \mathbb{R}^+)^p$ independently of h and convergent in $L^2_{loc}(\mathbb{R}^2 \times \mathbb{R}^+)^p$ to u, we have

$$\lim_h ||f^h(u^h) - f(u)||_{L^1_{loc}(\mathbb{R}^2 \times \mathbb{R}^+)^p} = 0.$$

Later, we give examples of schemes that satisfy the previous assumptions. We have the following result,

Theorem 2.2. *Let be $u_0 \in L^\infty(\mathbb{R}^2)^p$ and u^h the approximation given by (5). We assume that the scheme satisfies the assumptions (H1) and (H2). We also assume there exists a constant C that depends only on C_1, C_2 and u_0 and a function $u \in (L^2(\mathbb{R}^2 \times \mathbb{R}^+))^p$ such that*

$$\sup_h \sup_{x,y,t} |u^h(x,y,t)| \leq C$$

$$\lim_h ||u - u_h||_{L^2_{loc}(\mathbb{R}^2 \times \mathbb{R}^+)^p} = 0$$

Then u is a weak solution of (6).

The proof of Theorem 2.2 is given in Appendix A.

2.2.2 Accuracy

Here, we consider the steady case

$$\begin{array}{ll} \text{div } f(u) = 0 & x \in \Omega \\ \text{Boundary conditions on the inflow boundary} & \end{array} \tag{8}$$

and derive a formal sufficient condition that ensures the scheme is $r+1$–th order.

In appendix A, it is shown that for any test function $\varphi \in C^1(\mathbb{R})$, we have

$$\begin{aligned} \sum_\sigma \varphi_\sigma \Big(\sum_{T'_T, \sigma \in T'_T} \Phi_\sigma^{T'_T} \Big) &= \int \varphi^h \text{ div } f^h(u^h) \\ &+ \sum_T \Big(\sum_{\sigma \neq \sigma'} (\varphi_\sigma - \varphi_{\sigma'}) (\Phi_\sigma^{T'_T} - \Phi_\sigma^T) \Big) \end{aligned}$$

Here, $\Phi_\sigma^T = \int_T \psi_\sigma \text{div } f^h(u^h)$ where ψ_σ is any of the $(r+1)(r+2)/2$ P_r basis functions.

Then we notice that $\int_T \text{div} \psi_\sigma \mathbf{f}(u^h) = \mathcal{O}(h^{r+2})$ for a steady problem if u^h is the interpolant of a regular enough function, say $u \in H^s$ for s large enough . In fact, if u^h represents the interpolant of such a function u, then

$$\begin{aligned} \int_T \text{div} \psi_\sigma F(u^h) dx &= \int_{\partial T} \psi_\sigma F(u^h) \cdot \vec{n} dl + \int_T \nabla \psi_\sigma F(u^h) \\ &= \int_{\partial T} \big(F(u^h) - F(u)\big) \cdot \vec{n} dl + \int_T \nabla \psi_\sigma \big(F(u^h) - F(u)\big) \\ &= \mathcal{O}(h) \times \mathcal{O}(h^{r+1}) \end{aligned}$$

If we assume that $\Phi_j^q = \mathcal{O}(h^{r+2})$ for a steady state solution, it is clear that the term II is $\mathcal{O}(h^{r+3})$, and proceeding as in [1], we conclude that

$$\sum_j \sum_{T, M_j \in T} \varphi(x_j)\Phi_j = \int_\Omega \varphi(x) \text{ div } F(u)dx + \mathcal{O}(h^{r+1})$$

for a regular mesh.

In section 3, we recall schemes that are second order (for steady state) and monotonicity preserving. In section 5, we construct similar schemes for $r \geq 3$ and provide a general construction method.

3 Examples of schemes

We recall the N schemes, LDA and PSI schemes developed by Roe, Deconinck and coworkers in the context of non–refined grids and P_1 elements. [4] They are examples of upwind schemes, meaning that the residuals ϕ^T are split into signals ϕ_i^T in a manner that reflects the local flow of information by enhancing those signals when M_i is a downwind node of T. These schemes satisfy the technical requirements of the previous sections. The PSI scheme is the only one that is both second-order and monotone. Schemes having both of these properties must be constructed in an essentially nonlinear way to avoid the restrictions of Godunov's Theorem , that linear monotone schemes are only first order. See for example [9].

We begin this section by a systematic construction of high order monotone scheme. Then we provide examples for linear and non linear problems.

3.1 Monotone Schemes

To begin with, assume that Φ_j^T, the part of the fluctuation assigned to the jth vertex, can be written as

$$\Phi_j^T = \sum_{j, M_j \in T} c_{ij}^T (u_i - u_j).$$

The coefficients c_{ij}^T may be independent of the data (linear schemes) or may depend on the data (non linear schemes.) The scheme is monotonicity preserving provided $c_{ij} \geq 0$ under the CFL like condition

$$\Delta t \left[\frac{\displaystyle\sum_{T, M_i \in T} c_{ij}^T}{|C_i|} \right] \leq 1 \text{ for any } M_i.$$

It is possible to construct monotone second-order schemes for steady problems, but we have already noted that Godunov's theorem states that they cannot be linear. A systematic construction is to begin with a monotone, linear, and therefore first-order scheme, and then to modify it so that it achieves higher-order accuracy.

Denote by Φ_j^M the part of the fluctuation assigned by some monotone scheme to the jth vertex, and by Φ_j^H the part assigned by a nonlinear high-order scheme. The requirements are that

$$\begin{array}{ll} \displaystyle\sum_j \Phi_j^H = \sum_j \Phi_j^M = \Phi & \text{Conservation} \\ \Phi_j^H \Phi_j^M \geq 0 & \text{Monotonicity} \end{array}$$

We get second order accuracy under the following conditions

$$\begin{array}{ll} \Phi_j^H = \mathcal{O}(\Phi) & \\ \text{and} & \\ \Phi = \mathcal{O}(h^3) & \text{at steady state for smooth solutions.} \end{array}$$

The two conditions are important from the analysis of section 2.2.2 ; in particular they impose second order accuracy constraints on the flux approximation $f^h(u^h)$.

After introducing

$$\beta_j = \frac{\phi_j^M}{\phi}, \qquad \hat{\beta}_j = \frac{\phi_j^H}{\phi}$$

these conditions can be reformulated as constraints on the weights with which the residuals are distributed to the nodes

$$\begin{array}{ll} \sum_{j=1}^{N} \beta_j = \sum_{j=1}^{N} \hat{\beta}_j = 1 & \text{Conservation} \\ \beta_j \hat{\beta}_j \geq 0 & \text{Monotonicity} \\ \hat{\beta}_j \text{ is bounded} & \text{High-order accuracy if } \phi^T \text{ is small enough} \end{array}$$

Our procedure is to impose these constraints on a mapping, which takes a set of weights $\{\beta_j\}$ corresponding to a monotone scheme (which may only be first-order accurate) to a set of weights $\{\hat{\beta}_j\}$ corresponding to a scheme that is both monotone and of maximum accuracy. In reformulating, we allow for an arbitrary number N of degrees of freedom per triangle. The mapping from the N-vector $\{\beta_j\}$ to the N-vector$\{\hat{\beta}_j\}$ cannot be linear, because of the Godunov Theorem, but there are many such nonlinear mappings, They are the truly multidimensional analogues of limiter functions [6] and we expect their thorough investigation to take considerable time.

Here, we simply offer two examples for the case $N = 3$, presenting the vectors $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ and $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ as the barycentric coordinates of a point in space with respect to an equilateral triangle. To ensure boundedness, we insist that, for all j, $0 \leq \hat{\beta}_j \leq 1$, so that the point $\boldsymbol{\beta}$ lies within the triangle, or on its boundary. A weaker condition, constraining the point $\boldsymbol{\beta}$ to a finite neighborhood of the triangle, seems possible but has not yet been explored.

If the monotone weights are all positive, then $\boldsymbol{\beta}$ already lies within the triangle, and it is natural to take simply $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$. If $\boldsymbol{\beta}$ lies outside the triangle, one possibility is simply to project $\boldsymbol{\beta}$ onto the boundary of the triangle. For example, we may take

$$\hat{\beta}_j = \frac{\beta_j^+}{\sum_j \beta_j^+} \tag{9}$$

In Figure 2 this is shown geometrically on the left. The mapping from $\boldsymbol{\beta}$ to $\hat{\boldsymbol{\beta}}$ is always a translation toward one of the vertices.

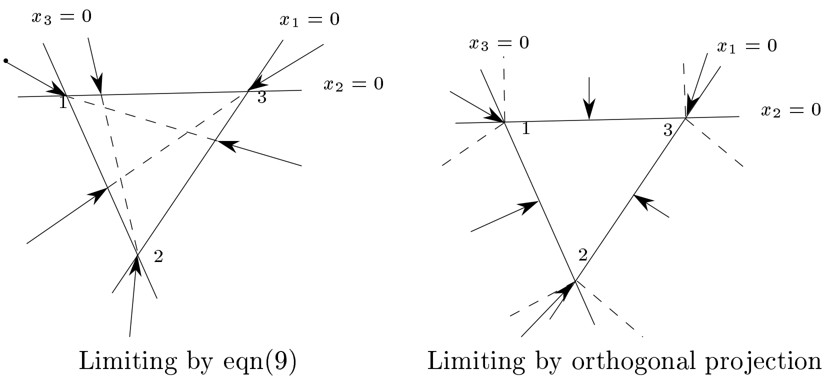


Figure 2: Geometrical illustration of the two limiters

An alternative is to take $\hat{\boldsymbol{\beta}}$ as the point on the boundary of the triangle that is closest to $\boldsymbol{\beta}$. The following logic accomplishes that

```
IF (B1.GE.0.) AND (B2.GE.0.) AND (B3.GE.0.) GOTO END
BS1 = B1+B2/2, BS2 = B2+B3/2, BS3 = B3+B1/2
B1 = MAX(0.,MIN(BS1,1.-BS2,1.))
B2 = MAX(0.,MIN(BS2,1.-BS3,1.))
B3 = MAX(0.,MIN(BS3,1.-BS1,1.))
END
```

Since $\hat{\boldsymbol{\beta}}_j$ is homogeneous of degree 1 in x_j, this can be coded without division. In Figure 2 this map is shown geometrically on the right. Extension to $N \geq 4$ is straightforward for both maps.

3.2 Case of a linear flux

We start by $f(u) = \vec{\lambda} u$. If u^h is a piecewise linear function, the residual Φ_T can be written

$$\Phi^T = \sum_{M_i \in T} k_i u_i \tag{10}$$

where

$$k_i = \int_T \nabla \Lambda_i \cdot \vec{\lambda} dx = \frac{1}{2} \vec{\lambda} \cdot \vec{n}_i.$$

The vector $\vec{n}_i$ is the inward normal vector to T opposite the vectex M_i, see Figure 3

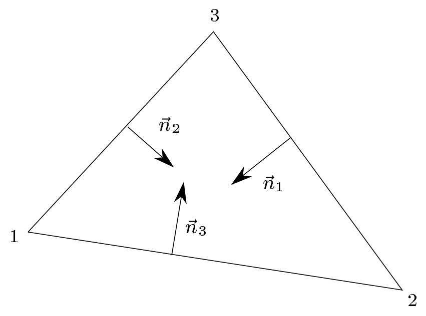


Figure 3: Inward normal vectors $\vec{n}_i$.

3.2.1 The N scheme

The N scheme may be written in several ways, one of which is

$$\Phi_i^T = k_i^+ \big(u_i - \widetilde{u}\big) \tag{11}$$

where $k_i^+ = \max(0., k_i)$. For consistency with the definition (10) we must have

$$\sum_j k_j^+ u_j - \big(\sum_j k_j^+\big) \widetilde{u} = \sum_j k_j u_j$$

so that

$$-\big(\sum_j k_j^+\big) \widetilde{u} = \sum_j k_j^- u_j$$

because $k_j = k_j^+ + k_j^-$. Since $\sum_j k_j = 0$ because $\sum_j n_j = 0$, we have

$$\widetilde{u} = \Big(\sum_j k_j^-\Big)^{-1} \Big(\sum_j k_j^- u_j\Big). \tag{12}$$

The N scheme is defined by equations (11) and (12). We define

$$n = \left(\sum_j k_j^-\right)^{-1} \leq 0$$

and note that yet another definition of the N scheme would be

$$\Phi_i^T = \sum_j k_i^+ n k_j^- (u_i - u_j) = \sum_j c_{ij}(u_i - u_j). \tag{13}$$

where $c_{ij} = k_i^+ n k_j^-$. Since $c_{ij} \geq 0$, the N scheme is monotonicity preserving under the constraint

$$\frac{\Delta t}{|T|} \max_i \sum_{j \neq i} c_{ij} \leq 1.$$

There are two types of triangles : those for which only one k_j is positive, and those for which two of the k_js are positive. In the first case, the triangle is called one target ; in the second case, it is a two target triangle. For a one target triangle, assume $k_1 > 0$, $k_2, k_3 \leq 0$, so that

$$\Phi_1^T = \Phi^T,\ \Phi_2^T = 0,\ \Phi_3^T = 0.$$

In this case the point $\boldsymbol{\beta}$ lies at one vertex of the triangle; either of the above mappings will leave it there . In a two target triangle, say $k_1, k_2 > 0$ and $k_3 < 0$. We have

$$\Phi_1^T = k_1(u_1 - u_3),\ \Phi_2^T = k_2(u_2 - u_3),\ \Phi_3^T = 0.$$

In this case $\boldsymbol{\beta}$ lies at one side of the triangle, but not necessarily between the vertices. In fact, in a two target case, we have, in a smooth steady state, $\Phi^T = \mathcal{O}(h^3)$, but Φ_1^T , $\Phi_2^T = \mathcal{O}(h^2)$. so the point $\boldsymbol{\beta}$ may lie arbitrarily far from the triangle. This example shows that the scheme is only first order accurate at steady state, and this is confirmed by numerical experiments.

3.2.2 The LDA scheme

A second order upwind scheme is the LDA (Low Diffusion Advection) scheme defined by

$$\Phi_i^T = -nk_i^+ \Phi^T. \tag{14}$$

Since $\Phi_i^T = \sum_j -(k_i^+ n k_j)(u_i - u_j) = \sum_j c_{ij}(u_i - u_j)$ where $c_{ij} = -k_i^+ n k_j$, we see that c_{ij} may be either positive or negative, so that the LDA is not a monotonicity preserving scheme in general. However, (14) shows that each Φ_i^T is of the same order as Φ^T, and the scheme should be second-order in the steady state. Again this is confirmed by numerical experiments. However, we see that in the one target case, the two schemes coincide.

3.2.3 The PSI scheme

The PSI (Positive Streamwise Invariant) scheme of Struijs is constructed from the N scheme using one of the limiter presented in paragraph 3.1. Remarkably the scheme does not depend on which of the two limiters is chosen ! This is because, in the two target case, $\boldsymbol{\beta}$ lies on an edge, and if it is outside the triangle, either limiter will move it to the nearest vertex. This gives precisely the successful PSI scheme of Struijs [4].

3.3 Case of a non linear flux

Similar schemes can be constructed for the problem

$$\text{div } f(u) = 0$$

provided we can find, for each element T, a *local linearization*, that is to say, a vector $\widehat{\vec{\lambda}}$ such that, for a piecewise interpolant u^h,

$$\int_T \text{div } f(u^h)dx = \int_T \widehat{\vec{\lambda}} \cdot \nabla u^h dx.$$

In the case of Burgers' equation $f(u) = (u^2/2, u)$, which we employ later as a test problem in numerical examples, a solution is given by $\widehat{\vec{\lambda}} = (\frac{u_1+u_2+u_3}{3}, 1)$, (u_1, u_2, u_3) being the values at the three vertices.

4 Analysis of the PSI and LDA schemes

We start again from the variational formulation for the discretisation of

$$\begin{array}{ll} \vec{\lambda} \cdot \nabla u = f & x \in \Omega \\ u = 0 & \text{on } \Gamma^-. \end{array}$$

with $\vec{\lambda}$ a consant vector. The inflow boundary Γ^- is defined as

$$\Gamma^- = \{x \in \partial\Omega, \vec{n(x)} \cdot \vec{\lambda} > 0\}$$

where $\vec{n(x)}$ is the inward normal vector at $x \in \partial\Gamma$. For the sake of simplicity, we assume that Ω is polygonal, and that the mesh is adapted to the boundary ($\partial\Omega = \partial\bigg(\cup_{T \in \mathcal{T}_h} T\bigg)$).

Taking φ piecewise linear, we have

$$\begin{array}{ll} \displaystyle\sum_i \varphi(x_i) \sum_{T, M_i \in T} \Phi_i = 0, u_i & \text{for } M_i \notin \Gamma^-, \\ u_i = 0 & \text{if } M_i \in \Gamma^- \end{array} \tag{15}$$

We assume that the residual Φ_i is

$$\Phi_i = \beta_i \Phi, \qquad \Phi = \int_T (\vec{\lambda} \cdot \nabla u - f)dx$$

with $\displaystyle\sum_j \beta_j = 1$ and β_j bounded. The relation (15) can be rewritten as

$$\sum_i \varphi(x_i) \sum_{T, M_i \in T} \Phi_i = \sum_T \bigg(\sum_{M_j \in T} \beta_j \varphi(x_j) \bigg) \Phi$$

If the flux is linear, we have

$$\sum_T \bigg(\sum_{M_j \in T} \beta_j \varphi(x_i) \bigg) \Phi = \int_\Omega \ell(\varphi)(\vec{\lambda} \cdot \nabla u^h - f)dx \tag{16}$$

where, for any triangle T,

$$\ell(\varphi)_{|T} = \sum_{M_j \in T} \beta_j \varphi(x_j).$$

We denote by a the bilinear form on $V_h^1 \times V_h^1$ defined by

$$a(u^h, v^h) = \int_\Omega \ell(v^h) \vec{\lambda} \cdot \nabla u^h dx.$$

Consider the following spaces

- $V_h^1=\{u^h$ piecewise linear and continuous $, u^h_{|\Gamma^-}=0\}$ equipped with the graph norm

$$||u^h||^2_{V_h^1}=\int_\Omega \big(\vec{\lambda}\cdot\nabla u^h\big)^2.$$

- $X^h=\{v^h\in L^2(\Omega), v^h$ constant in each $T\}$ equipped with the L^2 norm.

We seek the solution of (15) in V_h^1, and if $u^h\in V_h^1$, $\ell(u^h)\in X^h$ when the β_is are uniformly bounded.

We specialise now to the PSI scheme. The case of the LDA scheme is similar. The first remark is that for any vertex M_i, there is at least one triangle having M_i as vertex that is one target on M_i

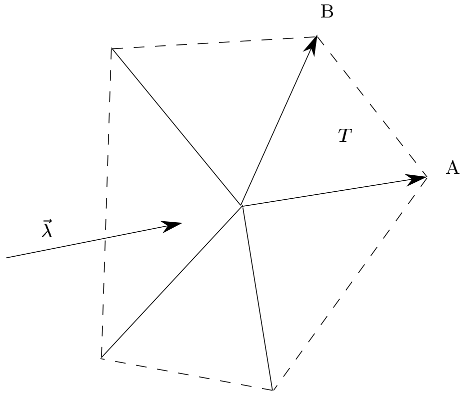


Figure 4: There is at least one one–target triangle.

To prove that, consider all the triangle T sharing M_i. As in Figure 4, we number the triangles in the positive orientation. Consider T a triangle, its vertices are $\{M_i, A, B\}$. The triangle T is one target to M_i if $\vec{\lambda}$ belongs to the angular sector defined by $\vec{M_iA}$, $\vec{M_iB}$, that is if $\det(\vec{\lambda},\vec{M_iA})>0$ and $\det(\vec{\lambda},\vec{M_iB})<0$. There is clearly at least one one-target triangle because as a point A rotates completely around M_i, the value of $\det(\vec{\lambda},\vec{M_iA})$ must change sign twice (when $\vec{M_iA}$ is parallel or antipallel to $\vec{\lambda}$). One of the sectors must span the transition from positive to negative. The proof extends to 3D.

The consequence of that is, denoting by T_{M_i} the one–target triangle atM_i, that

$$\int_\Omega \ell(v^h)^2dx\geq \sum_i |T_{M_i}|(v_i^h)^2.$$

If the mesh is regular, there exist a positive constant C such that

$$\sum_i |T_{M_i}|(v_i^h)^2\geq C\int_\Omega (v^h)^2dx$$

(here we use the fact that $v_i=0$ on Γ^-), so we get

$$\int_\Omega \ell(v^h)^2dx\geq C\int_\Omega (v^h)^2dx. \tag{17}$$

Using this remark, we show in Appendix B that there exists $C>0$ such that

$$\sup_{v_h\in V_h^1}\frac{\displaystyle\int_\Omega \ell(v^h)\vec{\lambda}\cdot\nabla u^h}{\sqrt{\displaystyle\int_\Omega (v^h)^2}}\geq C\sqrt{\int_\Omega \big(\vec{\lambda}\cdot\nabla u^h\big)^2}. \tag{18}$$

The inequality (18) shows that we have a uniform inf–sup condition, that is we have stability of the scheme (see [5]).

Moreover, it is possible to derive an error estimate. First, we notice that

$$a(u^h, v^h) \leq ||u^h||_{V_h^1} \, ||\ell(v^h)||_{L^2} \leq ||u^h||_{V_h^1} ||v^h||_{V_h^1}.$$

The last inequality comes from the fact $0 \leq \beta_i \leq 1$.

By Céa's lemma, we have the inequality

$$||u^h - u||_{V_h^1} \leq \big(1 + \frac{1}{C}\big) \inf_{w^h \in W_h} ||u - w^h||_{W_h}||. \tag{19}$$

In particular, if $u \in H^1(\Omega)$, we have the existence of C' such that

$$||u^h - u||_{V_h^1} \leq C'h. \tag{20}$$

The important fact are the relation (16) and the inequality (17). This inequality comes from the upwind property of the scheme and enables us to show the coercivity of the scheme. We note that this proof works also for the LDA scheme.

5 Construction of high order upwind monotonicity preserving residual distributive scheme

We show the explicit construction of a third order scheme. The extension to higher order accuracy is straightforward.

Denote V_h^2 the space of continuous functions that are quadratic in each triangle T equipped with the standard H^2 norm. We want to construct a scheme whose variational formulation reads : find $u^h \in V_h^2$ such that

$$\forall v^h \in V_h^2 \int_\Omega \ell(v^h) \lambda \cdot \nabla u^h = 0.$$

We need to specify the space V^h as well as the mapping ℓ. The key is to have an inequality of the type (18). If so, the error estimate (19) is obvious, and (16) becomes

$$||u^h - u||_{V^h} \leq C'h^2. \tag{21}$$

Consider T a triangle, and T_I, T_{II}, T_{III} and T_{IV} as in Figure 1, and we still denote by σ the degrees of freedom of u^h. We can construct the N scheme on the sub–triangles of T. We denote by $\Phi_\sigma^{T_\xi}$ for $\xi = I, II, III, IV$ the corresponding signals sent to the points contained in T. Then we consider

$$\Phi_\xi = \int_{T_\xi} \vec{\lambda} \cdot \nabla u^h.$$

It is clear that

$$\Phi_\xi \neq \sum_{\sigma \in T_\xi} \Phi_\sigma^{T_\xi},$$

but we still can use the N scheme for a comparison purpose. We insist that the signals sent by the subtriangles to their vertices be limited to the same order of magnitude as the restriction of the high-order residual Φ_ξ. Thus we define the scheme by a set of modified signals

$$\Psi_\sigma^{T_\xi} = \beta_\sigma^\xi \Phi_\xi \tag{22}$$

with

$$\beta_\sigma^\xi = \frac{\left(\dfrac{\Phi_\sigma^\xi}{\Phi_\xi}\right)^+}{\sum\limits_{\sigma' \in T^\xi} \left(\dfrac{\Phi_{\sigma'}^\xi}{\Phi_\xi}\right)^+} \tag{23}$$

and

$$\Psi_\sigma = \sum_{T_\xi, \sigma\in T_\xi} \Psi_\sigma^{T_\xi}. \tag{24}$$

The conservation relation is

$$\sum_{\sigma\in T} \Psi_\sigma = \sum_{T_\xi} \int_{T_\xi} \vec{\lambda}\cdot\nabla u^h = \int_T \vec{\lambda}\cdot\nabla u^h.$$

Provided the stability conditions of the Lax Wendroff condition are true, the scheme will be convergent : u^h is continuous accross the edges of T, thanks to the continuity of u^h.

Let us come back to the stability property of the scheme. The variational formulation is given with

$$\ell(v^h)_{T_\xi} = \sum_{\sigma\in T_\xi} \beta_\sigma^\xi v^h(x_\sigma).$$

The series of inequality (17–18) works, so that we get the stability and the error estimate (21). The proof is given in Appendix B.

An extension of the LDA scheme (which has been tested and provide similar error curves, but the results are not reported here) is the following. Consider $T'_\xi \subset T$, define $k_\sigma^\xi = \vec{\lambda}\cdot\vec{n}_\sigma^\xi$ where $\vec{n}_\sigma$ is the inward normal to T'_ξ opposite to σ. In (22), we replace β_σ^ξ by

$$\beta_\sigma^\xi = -(k_\sigma^\xi)^+\Big(\sum_{\sigma'\in T'_\xi}(k_{\sigma'}^\xi)^+\Big)^{-1}.$$

This defines a high order LDA scheme. It is not monotone.

6 Some numerical results

6.1 Linear advection

We consider the problem

$$\begin{array}{ll} -y\dfrac{\partial u}{\partial x} + x\dfrac{\partial u}{\partial x} = 0 & (x,y)\in[-1,1]\times[0,1] \\ u(x,0) = & \begin{cases} 0 & \text{if } x\notin[0.1,0.7] \\ \sin\left(\pi\frac{x-0.1}{0.6}\right) & x\in[0.1,0.7]\end{cases}\end{array} \tag{25}$$

The exact solution is, with $r=\sqrt{x^2+y^2}$, $u(x,y) = \sin\left(\pi\frac{r-0.1}{0.6}\right)$ if $r\in[0.1,0.7]$ and 0 elsewhere. The numerical solution is computed with a pseudo–time marching algorithm where the initial solution is $u_0 = 2$.

On each sub–triangle T'_T, we employ the scalar N scheme. The average velocity is defined by

$$(-\bar{y},\bar{x}) = \int_T (-y,x)dxdy.$$

Three schemes are compared : the standard second order PSI scheme, a third order scheme and a fourth order one. The high order schemes are constructed using P_3 and P_4 interpolation with the first limiter of section 5. The mesh has 628 vertices and 1162 elements and is displayed on Figure 5.

The isolines of the solutions are displayed on Figure 6. The circular shape of the solution is perfectly well respected in each case. What is more interesting is a plot in the exit section, i.e. for $x=0$ and $y\leq -1$. This is given in Figure 7 with a comparison with the exact one. We also provide a zoom of the solution around $(x,y) = (0,-0.5)$ where the solution is maximum. Because of the time–marching algorithm and the initial solution we have chosen, this Figure shows clearly that an increase of the formal accuracy of the scheme does improve its effective accuracy. The fourth order scheme is almost perfect.

Last we provide errors for the advection problem $\vec{\lambda} = (0,1)^T$ and the boundary condition $u(x,y) = \cos\pi x$ on $[-1,1]\times[0,1]$ for the second order PSI scheme (Figure 8-a), the third order one (Figure 8-b) and the

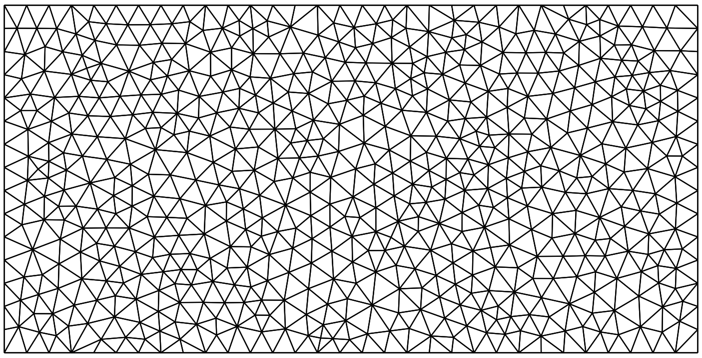
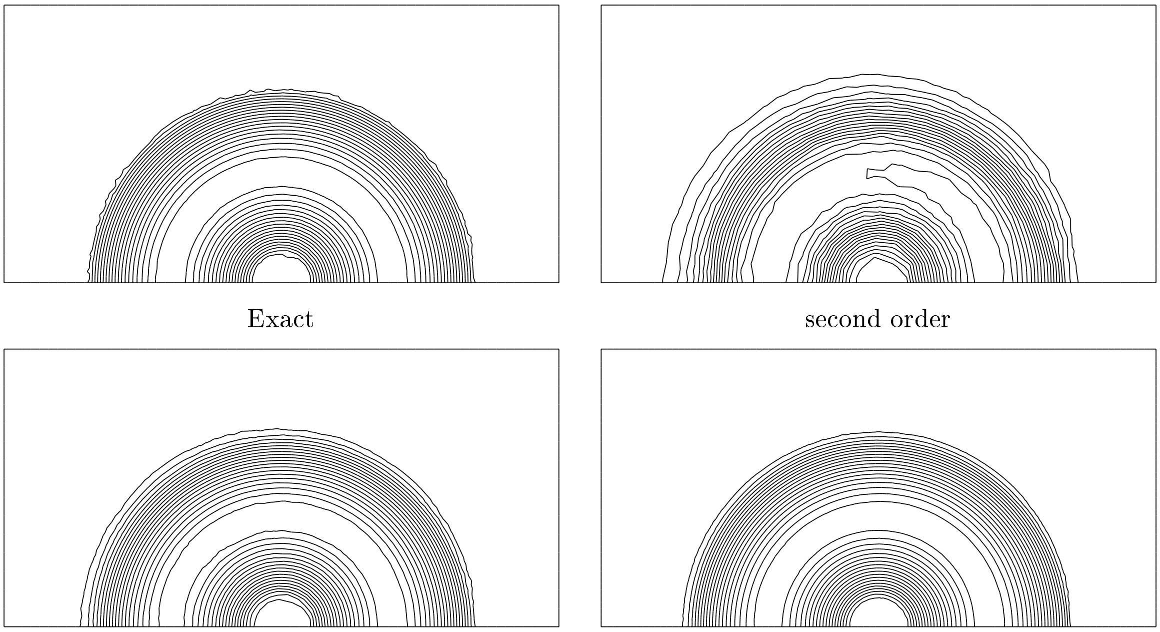


Figure 5: Mesh for the linear circular advection problem.



Exact

second order

Third order

Fourth order

Figure 6: Isolines of the exact and computed solutions

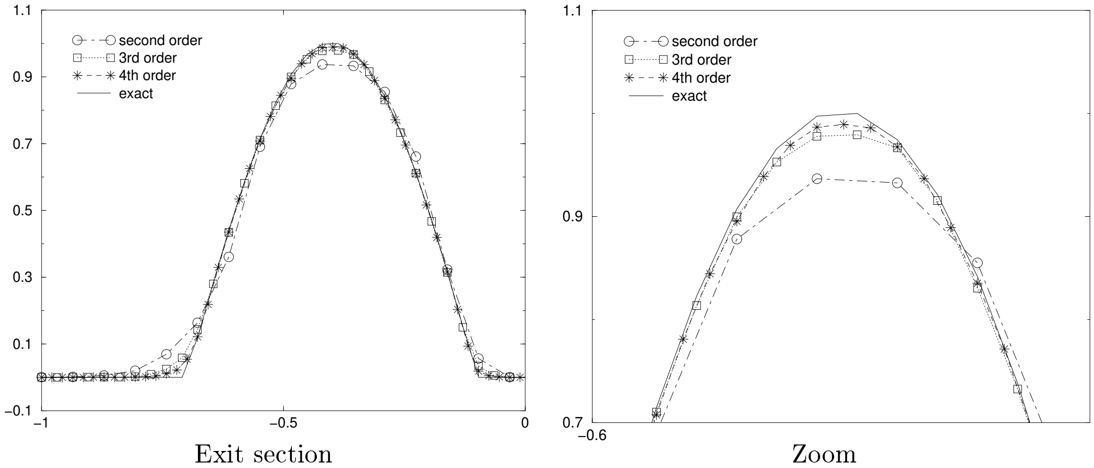


Figure 7: Plot of the solutions in the interval $[-1, 0]$, and zoom around the maximum

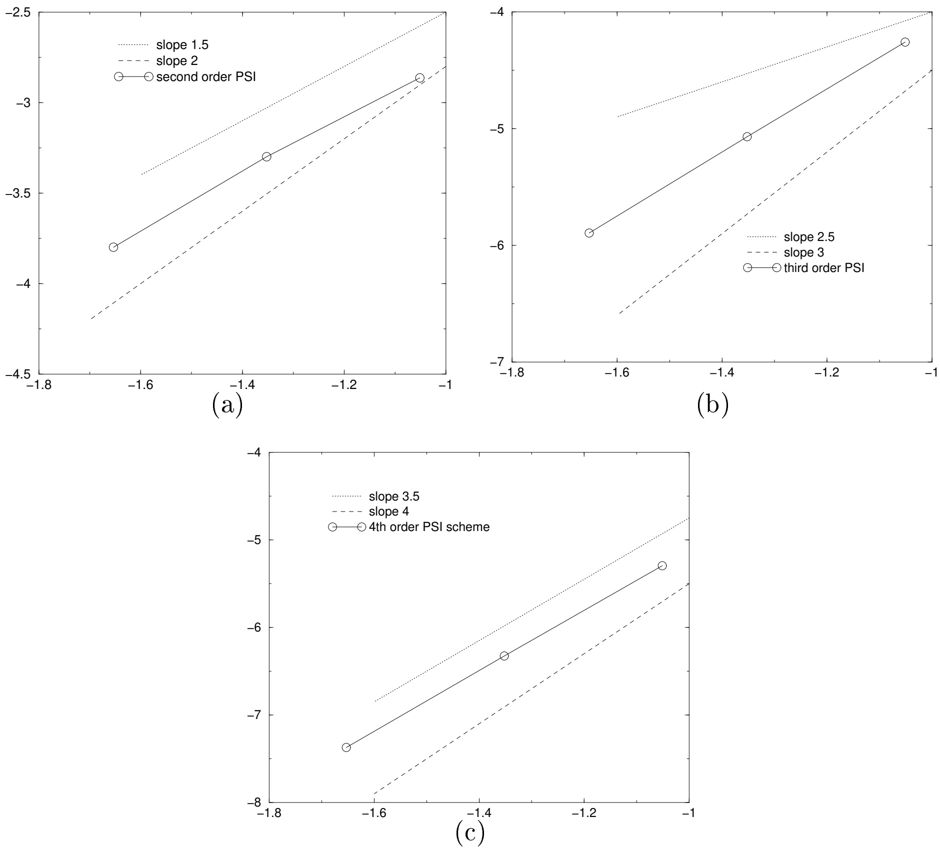


Figure 8: Numerical error of the : (a) second order, (b) third order and (c) fourth order PSI schemes

fourth order one (Figure 8-c). We see that the error slope of the k+1 th order scheme is between $k+1/2$ and $k+1$. This is not in contradiction with the expected theoretical $k+1/2$ order of accuracy.

To obtain these results, a careful implementation of the limiter function

$$\hat{\beta}_i = \frac{\left(\frac{\Phi_i}{\Phi}\right)^+}{\sum_{j=1}^{3}\left(\frac{\Phi_j}{\Phi}\right)^+}$$

is needed. The first numerical results did not quite deliver the expected accuracy. It seems more important for the higher order schemes than for the the second order PSI scheme that $\sum_j \hat{\beta}_j = 1$ exactly. We have chosen to compute the revised weights (in pseudo fortran) as

$$\hat{\beta}_i = \frac{\Big(1+\text{sign}(1., \Phi_i\Phi)\Big)\Phi_i + \epsilon}{\sum_{j=1}^{3}\Big(1+\text{sign}(1., \Phi_j\Phi)\Big)\Phi_j + 3\epsilon}$$

with $\epsilon = 10^{-10}$ in double precision.

6.2 Burger equation

We consider the Burger equation

$$\begin{aligned} &\frac{1}{2}\frac{\partial u^2}{\partial x} + \frac{\partial u}{\partial y} = 0, \qquad x \in [0,1]\times[0,1] \\ &u(x,y) = 1.5 - 2x \qquad \text{on the inflow boundary} \end{aligned}$$

The exact solution is

$$u(x,y) = \begin{cases} -0.5 & \text{if } y \le 0.5 \text{ and } -2\big(x-\frac{3}{4}\big) + y - \frac{1}{2} \le 0 \\ 1.5 & \text{if } y \le 0.5 \text{ and } -2\big(x-\frac{3}{4}\big) + y - \frac{1}{2} \ge 0 \\ \max\Big(-0.5, \min\Big(1.5, \frac{x-\frac{3}{4}}{y-1./2.}\Big)\Big) & \text{else} \end{cases}$$

Once more, the solution is computed with a pseudo–time marching algorithm, the initial condition is set to $u=-0.5$. We represent the solutions computed for the mesh represented on Figure 9 with 1041 vertices and 1960 triangles.

The isolines of the solutions are displayed on Figure 10.

We also display cross sections for $y = 0.75$ accross the discontinuity (Figure 11) and $y = 0.25$ accross the fan (Figure 12).

Last, we give a zoom of the cross–section $y = 0.25$ around the corner of the fan on Figure 13. We see a very clear improvement of the quality of the solution with the increase of the accuracy order.

7 Conclusions

We have presented and analysed a class of schemes for hyperbolic problems, with a particular emphasis on scalar problems. We have provided a very general condition that ensures that the limit solution, if it exists, is a weak solution of the problem. Then we have discussed the construction of second-order monotone schemes from a new perspective that extends very naturally to higher-order. It is, however, not the only possible extension; it should be compared with proposals recently put forward by Caraeni [3] (see also [10]).

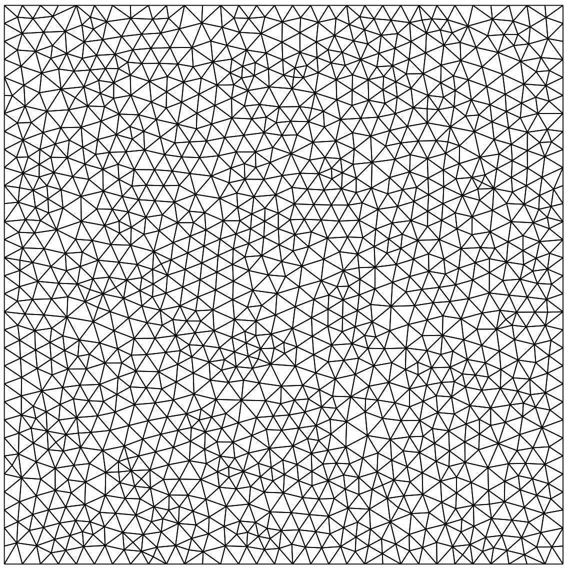
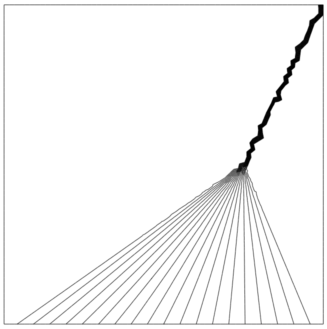
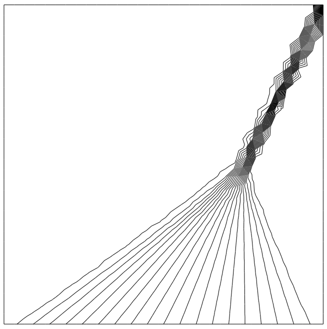


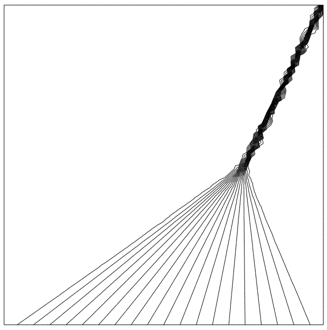
Figure 9: Mesh for the Burger's equation



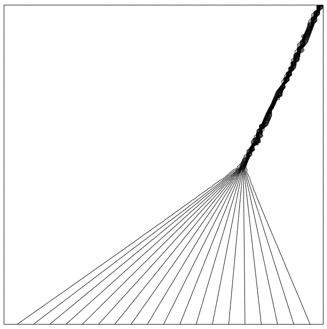
Exact



second order

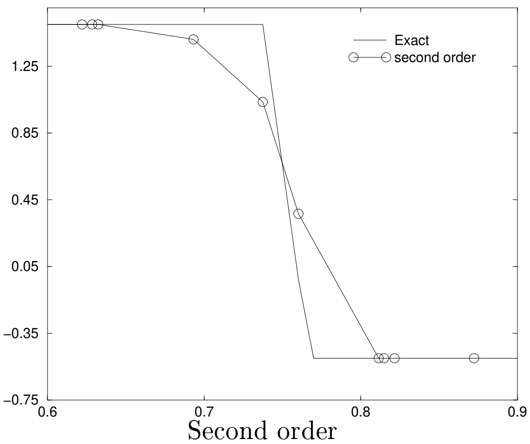


Third order

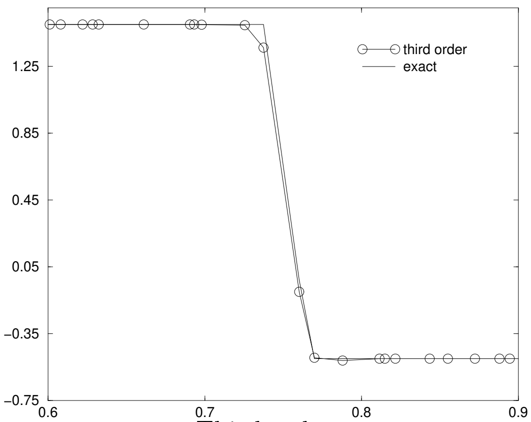


Fourth order

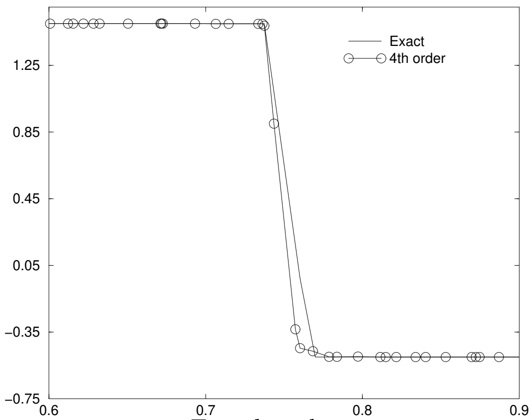
Figure 10: Isolines of the exact and computed solutions



Second order

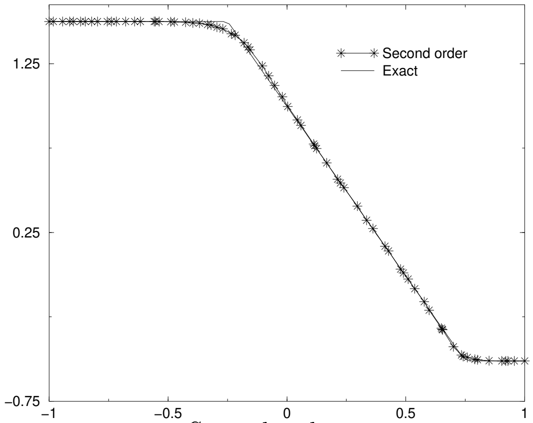


Third order

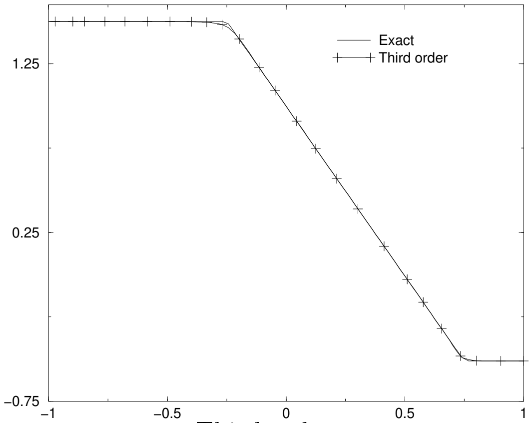


Fourth order

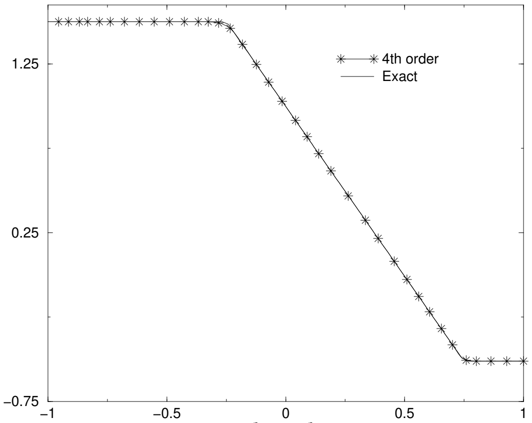
Figure 11: Plot of the solutions accross the shock, $y = 0.75$



Second order



Third order



Fourth order

Figure 12: Plot of the solutions accross the fan, $y = 0.25$

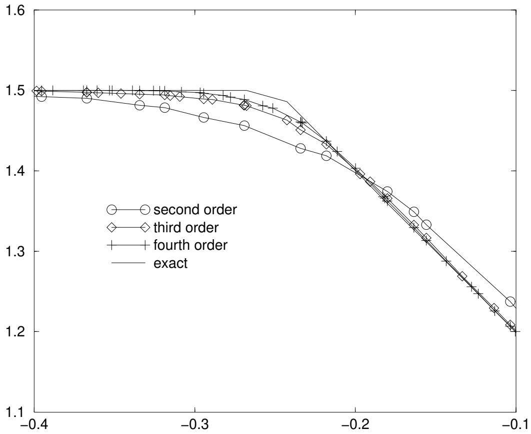


Figure 13: Zoomm accross the fan, around the corner

It also has many points of similarity with ideas being developped in a finite-volume context by Wang [14]. Many variations of the present method are also possible, in particular the use of other function spaces for representing the solution. The extension to quadrangular meshes can be conducted by cutting every quadrangle into two triangles. Exploring these possibilities may lead to robust, high-order, methods that are less costly than the currently popular Discontinuous Galerkin methods.

The extension to systems can proceed in several ways. The underlying N scheme can be extended formally by treating the coefficient k_j in Section 3 as matrix-valued [13]. Another method is to decompose a vector-valued residual into components associated with scalar advection problems and Cauchy-Riemann systems, treating each of these components independently, as discussed in [10].

Acknowledgements

This work was started during a visit to Bordeaux University, in June 2001, by the second author, who gratefully appreciated the warm hospitality shown on that occasion by members of the Département de Mathématiques Appliquées. Katherine Mer–Nkonga (CEA CESTA, France) is also acknowledged for her help in a very early draft of the paper [2].

References

[1] R. Abgrall, *Toward the ultimate conservative scheme : Following the quest*, J. Comput. Phys., (accepted for publication).

[2] R. Abgrall and K. Mer, *Un théorème de type Lax–Wendroff pour les schémas distributifs*, Tech. Rep. 98010, Mathématiques Appliquées de Bordeaux, Mars 1998.

[3] D. Caraeni, M. Caraeni, and L. Fuchs, *A parallel multidimensional upwind algorithm.* AIAA Paper 201-2547, 2001. 15th AIAA CFD Meeting, Anaheim.

[4] H. Deconinck, R. Struijs, and P. L. Roe, *Compact advection schemes on unstructured grids*, tech. rep., VKI, 1993. VKI LS 1993-04, Computational Fluid Dynamics.

[5] A. Ern and J. Guermond, *Eléments finis : théorie, applications, mise en oeuvre*, vol. 36 of Mathématiques et Applications, Springer, 2002.

[6] E. Godlewski and P. Raviart, *Hyperbolic systems of conservations laws*, Applied Mathematical Sciences, Springer Verlag, 1995.

[7] T. J. R. Hughes and M. Mallet, *A new finite element formulation for CFD: III. the generalized streamline operator for multidimensional advective-diffusive systems*, Comp. Meth. Appl. Mech. Engrg., 58 (1986), pp. 305–328.

[8] D. Kröner, M. Rokyta, and M. Wierse, *A Lax-Wendroff type theorem for upwind finite volume schemes in* 2*-d*, East-West J. Numer. Math., 4 (1996), pp. 279–292.

[9] R. Leveque, *Numerical methods for conservation laws*, Birkhauser, 1992. Zürich.

[10] H. Nishikawa, M. Rad, and P. Roe, *A third-order fluctuation splitting scheme that preserves potential flow.* AIAA Paper 2001-2595, 2001. 15th AIAA CFD Meeting, Anaheim.

[11] P. Roe, *Fluctuations and signals - a framework for numerical evolution problems.*, in Numerical methods for fluid dynamics, Proc. Conf., Reading/U.K., 1982, pp. 219–257.

[12] R. Struijs, H. Deconinck, and P. L. Roe, *Fluctuation splitting schemes for the 2d Euler equations*, tech. rep., VKI, 1991. VKI LS 1991-01, Computational Fluid Dynamics.

[13] E. van der Weide and H. Deconinck, *Positive matrix distribution schemes for hyperbolic systems*, in Computational Fluid Dynamics '96, Wiley, 1996, pp. 747–753.

[14] Z. WANG, *Spectral (finite) volume method for conservation laws on unstructured grids: Basic formulation*, J. Comput. Physics, (in press).

A Proof of Theorem 2.2

We start by a lemma that is inspired from [8].

Lemma A.1. *Let $T>0$ and N the integer part of $\frac{T}{\Delta t}$. We consider $\mathcal{Q}\subset\mathbb{R}^2$, a bounded domain. Let $(u^h)_h$ be a sequence such that $u^h(\,.\,,t_n)\in(X_h)^m$ for any $n\leq N$. We assume there exists a constant C independant of h and $u\in L^2_{loc}(\mathcal{Q}\times[0,T])^m$ such that*

$$\sup_h\sup_{x,y,t}|u^h(x,y,t)|\leq C,\quad \lim_h \left|\left|u^h-u\right|\right|_{L^2(\mathcal{Q}\times[0,T])}=0.$$

Then

1. $\lim_h\sum_{n=0}^{N}\Delta t\sum_{T\subset\mathcal{Q}}\sum_{T'_T\subset T}|T'_T|\sum_{\sigma,\sigma'\in T}\left|\left|u^h_\sigma-u^h_{\sigma'}\right|\right|=0.$
2. $\tilde{u}^h=\pi_h u^h$ *satisfies* $\lim_h h\,\left|\left|\nabla\tilde{u}^h\right|\right|_{L^2(\mathcal{Q}\times[0,T])}=0.$

Proof. It is enough to prove the lemma for real valued functions v that satisfies its assumptions.

First claim : For any sub–triangle $T'T$, we define (ω beeing a cycle of $\{1,2,3\}$)

$$v^h_{|T}=\sum_{\sigma\in T}v^h_\sigma\chi_{C_\sigma\cap T}$$

$$w^h_{|T}=\sum_{x_\sigma\in T}v^h_{\omega(\sigma)}\chi_{C_\sigma\cap T}.$$

This enables to define two functions on $\mathbb{R}^2\times\mathbb{R}^+$ that are bounded independantly of h. More over,

$$|T|\sum_{\sigma\neq\sigma'}\left|v^h_\sigma-v^h_{\sigma'}\right|=3\int_T\left|v^h-w^h\right|dx.$$

The "3" comes form the definition of the dual cells,

We have

$$\sum_{n=0}^{N}\Delta t\sum_{T\subset\mathcal{Q}}\sum_{T'_T\subset T}|T'_T|\sum_{\sigma,\sigma'\in T'_T}\left|(v^h_\sigma)^n-(v^h_{\sigma'})^n\right|=\int_0^{t_N}\int_{\cup T\subset\mathcal{Q}}\left|v^h-w^h\right|dxdt\leq\int_0^T\int_{\cup T\subset\mathcal{Q}}\left|v^h-w^h\right|dxdt.$$

Since (v^h) is bounded, there exists $v'\in L^\infty(\mathcal{Q}\times[0,T])$ such that $v^h\to v'$ for the weak star topology. Similarly, the exists $w\in L^\infty(\mathcal{Q}\times[0,T])$, with $w^h\to w$ for the weak star topology.

Since $v_h\to v$ in L^2_{loc}, we have $v'=v$ because $\mathcal{Q}\times[0,T]$ is bounded and $C_0^\infty(\mathcal{Q}\times[0,T])$ is dense in $L^1(\mathcal{Q}\times[0,T])$.

Let us show that $w=v$. Let be $\phi\in C_0^\infty(R^2\times\mathbb{R}^+)$ We have

$$\begin{aligned}
\int_0^T\int_{\mathcal{Q}}v_h\phi dxdt &= \int_0^T\sum_{T,T\cap\mathcal{Q}\neq\emptyset}\sum_{T'_T\subset T}\sum_{\sigma\in T'_T}v^h_\sigma\int_{C_\sigma\cap T'_T}\phi(x,t)dxdt\\
&= \sum_{\sigma\in T}w^h_\sigma\int_{C_{\omega^{-1}(\sigma)}\cap T}\phi(x,t)dxdt\\
&= \int_0^T\int_{\mathcal{Q}}w_h\phi dxdt+\sum_{\sigma\in T}w^h_\sigma\left(\int_{C_{\omega^{-1}(\sigma)}\cap T}\phi(x,t)dxdt-\int_{C_\sigma\cap T}\phi(x,t)dxdt\right)
\end{aligned}$$

We can also write

$$\int_{C_\sigma \cap T'_T} \phi dxdt = |C_\sigma \cap T'_T| \phi(\sigma), \int_{C_{\omega^{-1}(\sigma)} \cap T} \phi dxdt = |C_\sigma \cap T| \phi(\sigma')$$

for $\sigma \in C_\sigma \cap T$ et $\sigma' \in C_{\omega^{-1}(\sigma) \cap T}$ well chosen. Since $\nabla\phi$ is bounded on $\mathcal{Q} \times [0,T]$ and v^h is bounded, since $|C_\sigma \cap T| = |C_{\omega^{-1}(\sigma)} \cap T|$[1], we have

$$\left| \int_0^T \int_{\mathcal{Q}} v_h \phi dxdt - \int_0^T \int_{\mathcal{Q}} w_h \phi dxdt \right| \leq Ch$$

where C is independant of h. Hence, $w = v' = v$.

The same technique shows that (v_h^2) and (w_h^2) have the same weak star limit. Let us show it is v^2. By density of $C_0^\infty(\mathcal{Q} \times [0,T])$ in $L^1(\mathcal{Q} \times [0,T])$ and since v_h^2 is bounded independamment of h, we may choose test functions ϕ in $C_0^\infty(\mathcal{Q} \times [0,T])$. The function ϕ is bounded in $\mathcal{Q} \times [0,T]$. Hence, we have

$$\int_{\mathcal{Q} \times [0,T]} |v - v_h|^2 \phi dxdt \to 0,$$

and then

$$\int_{\mathcal{Q} \times [0,T]} v^2 \phi dxdt - 2 \int_{\mathcal{Q} \times [0,T]} v v_h \phi dxdt + \int_{\mathcal{Q} \times [0,T]} v_h^2 \phi dxdt \to 0.$$

By the Cauchy Schwarz inequality, $v\phi \in L^1(\mathcal{Q} \times [0,T])$; the second term tends towards

$$\int_{\mathcal{Q} \times [0,T]} v^2 \phi dxdt.$$

and the $v_h^2 \to v^2$ in L^∞ weak star.

Last, by the same argument for $\phi = 1$, since $w_h^2 \to v^2$ in L^∞ weak star, we obtain

$$\int_{\mathcal{Q} \times [0,T]} |w_h - v|^2 dxdt \to 0,$$

and finaly,

$$\int_{\mathcal{Q} \times [0,T]} |w_h - v_h|^2 dxdt \to 0.$$

The conclusion comes because $\mathcal{Q} \times [0,T]$ is bounded, hence $L^1(\mathcal{Q} \times [0,T]) \subset L^2(\mathcal{Q} \times [0,T])$

Second claim : We have

$$\nabla \tilde{v}_h^{|T'_T} = \frac{1}{2|T'_T|} \left(u_{\sigma_1} \vec{n}_\sigma^T + u_{\sigma_2} \vec{n}_{\sigma_2}^T + u_{\sigma_3} \vec{n}_{\sigma_3}^T \right) = \frac{1}{2|T|} \left((u_{\sigma_2} - u_{\sigma_1}) \vec{n}_{\sigma_2}^T + (u_{\sigma_3} - u_{\sigma_1}) \vec{n}_{\sigma_3}^T \right),$$

thus

$$\Delta t \sum_0^N \int_{\mathcal{Q}} ||\nabla \tilde{v}_h^n||^2 dxdt = \Delta t \sum_0^N \sum_{T, T \subset \mathcal{Q}} |T| \sum_{T'_T \subset T} \left|\left| (\nabla \tilde{v}_h^n)_{|T'_T} \right|\right|^2 \leq \Delta t \sum_0^N \sum_{T, T \subset \mathcal{Q}} \sum_{T'_T \subset T} \sum_{\sigma, \sigma' \in T'_T} |v_\sigma^n - v_{\sigma'}^n|^2$$

because the triangulation is regular (assumption H0). We conclude by the same arguments as before. ☐

Remark A.2. *One can consider distributive schemes that are formulated on non triangular meshes. In this case, the arguments remain similar, as soon as we assume that the elements in the mesh have a number of edges that is bounded independantly of h, and that the area of the elements satisfies for any Ω bounded*

$$\sum_{T_i \cap \Omega \neq \emptyset} \sum_{T_{i_l} \text{ neighbor of } T_i} \big|\, |T_i| - |T_{i_l}| \,\big| \to 0$$

when $h \to 0$.

[1] See remark A.2

Lemma A.3. *Let $\varphi \in C_0^1(\mathbb{R}^2 \times \mathbb{R}^+)$. With the assumptions of Theorem 2.2, one has*

$$\sum_\sigma \Delta t \sum_\sigma |C_\sigma|(u_\sigma^{n+1} - u_\sigma^n)\varphi(\sigma, t_n) + \int_{\mathbb{R}^2\times\mathbb{R}^+} u\frac{\partial \varphi}{\partial t} dxdt + \int_{\mathbb{R}^2} u_0(x)\varphi(x,0) \to 0$$

when $h \to 0$.

The proof is classical : it is the same as in the proof of the classical Lax Wendroff theorem.

Let us introduce ψ_σ the basis function at σ and Ψ_σ^T the fluctuation

$$\Psi_\sigma^T = \int_\Omega \psi_\sigma \text{div } f^h(u^h).$$

We have

Lemma A.4. *Let $\varphi \in C_0^1(\mathbb{R}^2 \times \mathbb{R}^+)$, $f \in (C^1(\mathbb{R}^m))^{2m}$. Assume that f^h satisfies the assumption (H2)–(iii), and that u_h satisfies those of Theorem 2.2.*

$$\Delta t \sum_{n,T} \sum_T \sum_{\sigma \in T_T' \subset T} \varphi_\sigma^n \Psi_i^T(u_h^n) + \int_{R^2\times\mathbb{R}^+} f(u(x,t)) \ div\ \varphi(x,t) dxdt \to 0$$

when $h \to 0$.

In order to simplify the proof of this result, the computations are done within one triangle T.

Proof. Let $\varphi \in C_0^1(\mathbb{R}^2 \times [0, +\infty[)$. Let Ω and τ such that $\text{supp}(\varphi) \subset \Omega \times [0, \tau]$. Consider $T \in \tau_h$ Now $\pi\varphi$ is the interpolation of φ of degree k. We have :

$$\Delta t \sum_{n,T} \sum_{\sigma \in T_T'} \varphi_\sigma^n \Psi_i^T(u_h^n) = \sum_{n,T} \int_{t^n}^{t^{n+1}} \int_T \pi\varphi \text{ div } f_T^h(u_h^n) dxdt$$

One can apply the Green formula,

$$\begin{aligned}\sum_T \int_T \pi_h\varphi(x,t^n) \text{ div } f_T^h(u_h^n)dxdt &= -\sum_T \int_T \nabla(\pi_h\varphi) \ . \ f_T^h(u_h^n)dx \\ &+ \sum_T \int_{\partial T} \pi_h \varphi(x,t^n) f_T^h(u_h^n) \ .\vec{n}dx \ \ .\end{aligned}$$

Since $\pi_h\varphi$ is continuous and from (H2)-*(iii)*, the second sum on the right hand side cancels.

Moreover, φ is C^1 and the triangulation is regular : $\nabla(\pi_h\varphi)$ is uniformly bounded by a constant C independant of h. Thus we have

$$\begin{aligned}\Delta t \left| \sum_{n,T} \int_{t^n}^{t^{n+1}} \int_T \nabla(\pi_h\varphi) \ . \ f_T^h(u_h^n)dx \right. &- \left. \sum_{n,T} \int_{t^n}^{t^{n+1}} \int_T \nabla\varphi \ . \ f(u)dx \right| \leq \\ &C \quad \Delta t \sum_{n,T} \int_{T\times[t^n,t^{n+1}]} \left|\left| f^h(u_h) - f(u) \right|\right| dxdt \\ &+ \quad \Delta t \sum_{n,T} \int_{T\times[t^n,t^{n+1}]} ||\nabla\pi_h\pi - \nabla\varphi|| \ ||f(u_h)|| \, dxdt\end{aligned}$$

The first sum is less than $\left|\left|f^h(u_h) - f(u)\right|\right|_{L^1(\Omega\times[0,T])}$ and tends to 0 because $||u_h||_\infty$ is bounded independantly of h, f is C^1 and $u_h \to u$ in L^2_{loc}.

Similarly, since u_h is bounded and f is continuous, $f(u_h)$ is bounded uniformly in h by a constant C. The second term of the right hand side sum is bounded by the L^1 norm of $\nabla\pi_h\varphi - \nabla\varphi$ that tends to 0 since the triangulation is uniform.

□

Proof of Theorem 2.2: We multiply (5) by $\varphi(\sigma, t^n)|C_\sigma|$, where the test function φ belongs to $C_0^1\left(\mathbb{R}^2 \times [0, +\infty[\right)$, such that $\text{supp}(\varphi) \subset \Omega \times [0, T]$ for Ω and T well chosen ($n\Delta t \leq T$). Then we sum up on the indices $n = 0, N$ and $\sigma \in J$ such that $\sigma \in \Omega$:

$$\sum_{\sigma \in J} \sum_{n=0,N} |C_\sigma| \left(u_\sigma^{n+1} - u_\sigma^n\right) \varphi(x_\sigma, t^n) \; + \; \Delta t \sum_{\sigma \in J} \sum_{n=0,N} \sum_{T; \sigma \in T} \phi_\sigma^T(u_h^n) \varphi(\sigma, t^n) = 0,$$

From Lemma A.3, we have

$$\sum_{\sigma \in J} \sum_{n=0,N} |C_\sigma| \left(u_\sigma^{n+1} - u_\sigma^n\right) \varphi(\sigma, t^n) = -\int_{\mathbb{R}^2 \times \mathbb{R}^+} u \frac{\partial \varphi}{\partial t} - \int_{\mathbb{R}^2} u_0 \varphi(., 0) + o(1).$$

For the space term, we write :

$$\begin{aligned}
\sum_{\sigma \in J} \sum_{T; \sigma \in T} \Phi_\sigma^{T,n} \varphi_\sigma^n \quad &= \sum_{T \subset \Omega} \sum_{\sigma} \Psi_\sigma^{T,n} \varphi_\sigma^n \\
&= \sum_{T \subset \Omega} \left(\sum_{\sigma \in T} \varphi_\sigma^n \left(\Phi_\sigma^{T,n} - \Psi_\sigma^{T,n}\right) \right),
\end{aligned}$$

By the conservation relation, we have

$$\sum_{\sigma \in T} \left(\Phi_\sigma^{T,n} - \Psi_\sigma^{T,n}\right) = 0,$$

so we can rewrite the second term of the previous sum as

$$\sum_{T \subset \Omega} \left(\sum_{\sigma\sigma' \in T} \left(\varphi_\sigma^n - \varphi_{\sigma'}^{n,T}\right) \left(\Phi_\sigma^{T,n} - \Psi_\sigma^{T,n}\right) \right)$$

By Lemma A.4, we have

$$\Delta t \sum_{n=0,N} \sum_{\sigma \in J} \sum_{T; \sigma \in T} \Phi_\sigma^{T,n} \varphi_\sigma^n = -\int_{\mathbb{R}^2 \times \mathbb{R}^+} f(u(x,t)) \;\; \text{div} \;\; \varphi(x,t) dx dt \; + \; o(1).$$

Last, from (H1) and Lemma A.1, setting $\Phi_\sigma^{T,n} - \Psi_\sigma^{T,n} = \mathcal{S}_\sigma^{T,n}$ to simplify, we have

$$\begin{aligned}
\Delta t \left| \sum_{n,T} \sum_{\sigma \neq \sigma' \in T_T'} \mathcal{S}_\Sigma^{T,n} \left(\varphi_\sigma^n - \varphi_{\sigma'}^n\right) \right| \quad &\leq C h \Delta t \sum_{n=0,N} \sum_{T \subset \Omega} \left|\mathcal{S}_\sigma^T(u_h^n)\right| \\
&\leq C h^2 \Delta t \sum_{n=0,N} \sum_{T \subset \Omega} \sum_{\sigma, \sigma' \in T} |u_\sigma^n - u_{\sigma'}^n| \longrightarrow 0,
\end{aligned}$$

because both Ψ_σ and Φ_σ satisfy (H1). This ends the proof $\square$

Remark A.5. *If we assume only the continuity of the fluctuations and a regular mesh, the last part of the proof remains identical since it is suffisant to show that the translated of u_h^n which come in the computation of ϕ_σ^T converge to u at least for a subsequence and almost everywhere. As for the classical Lax–Wendroff theorem, this is proved via Lemma lemme A.1.*

B Stability proof

Define V_h^k beeing the space of continuous functions made of piecewise polynomials of degree k on each triangle T. We consider the following bilinear form defined on $V_h^k \times V_h^k$

$$a(v_h, u_h) = \int_\Omega \ell(v_h) \vec{\lambda} \cdot \nabla u_h dx,$$

where, following the notations of section 5, we have defined

$$\ell_{T_\xi}(v_h) = \sum_{\sigma \in T_\xi} \beta^\xi_\sigma v^h_\sigma.$$

We assume that for any degree of freedom, there is at least one sub–triangle T_T for which $\beta^\xi_{\sigma'} = 0$ except for σ where $\beta_\sigma = 1$. We note that ℓ maps V_h^k on the set of L^2 functions that are constant on each T'_T.

We know that there exists $c > 0$ such that

$$\int_\Omega \ell(v^h)^2 dx \geq c^2 \int_\Omega (v^h)^2.$$

In this appendix, we show that

$$\inf_{u^h \in V_h^k} \sup_{v^h \in V_h^k} \frac{a(v^h, u^h)}{\sqrt{\int_\Omega (v^h)^2 dx} \sqrt{\int_\Omega \left(\vec{\lambda} \cdot \nabla v_h\right)^2}} \geq C_k > 0.$$

Proof. We denote by $|| \,.\, ||_{0,\Omega}$ the L^2 norm. First, we have

$$\sup_{v^h \in V_h^k} \frac{a(v^h, u^h)}{||v^h||_{0,\Omega}} = \sup_{v^h \in V_h^k} \left(\frac{a(v^h, u^h)}{||\ell(v^h)||_{0,\Omega}} \frac{||\ell(v^h)||_{0,\Omega}}{||v^h||_{0,\Omega}} \right) \geq c \sup_{v^h \in V_h^k} \frac{a(v^h, u^h)}{||\ell(v^h)||_{0,\Omega}} = ||\Pi_\ell(\vec{\lambda} \cdot \nabla u^h)||$$

where Π_ℓ denote the L^2 projection on the space $\ell(V_h^k) \subset L^2(\Omega)$.

The problem is to estimate $||\Pi_\ell(\vec{\lambda} \cdot \nabla u^h)||$, that is

$$\sup_{v^h \in V_h^k} \frac{\int_\Omega \ell(v^h) \vec{\lambda} \cdot \nabla u^h}{||\ell(v^h)||_{0,\Omega}}.$$

We first note that for any $v_h \in V_h^k$, we have

$$\int_\Omega \ell(v_h) \vec{\lambda} \cdot \nabla u^h dx = \int_\Omega \ell(v_h) U dx$$

where U is the piecewise constant function defined on each T'_T by

$$U_{|T'_T} := \frac{\int_{T'_T} \vec{\lambda} \cdot \nabla u dx}{|T'_T|}.$$

The second thing is to note that

$$\int_\Omega \ell(v_h) U dx = \sum_T \sum_{T'_T \subset T} \int_{T'_T} \ell(v^h) U_{T'_T} dx = \sum_T \sum_{T'_T \subset T} \int_{T'_T} \ell(v^h_1) U_{T'_T} dx$$

where v_1^h is the continuous function, linear on each sub–triangle T'_T, that interpolate $\{v_\sigma\}_{\sigma \in T'_T}$. The remark enable to reduce the problem to the case $k = 1$.

Then, we show the existence of $C > 0$ such that for any v^h piecewise linear and continuous,

$$\sup_{v_h \in V_h^1} \frac{\int_\Omega \ell(v_h) \vec{\lambda} \cdot \nabla u^h dx}{||v_h||_{0,\Omega}} \geq C ||\vec{\lambda} \cdot \nabla u^h||_{0,\Omega}.$$

We have, for any U piecewise constant on each T,

$$\frac{\displaystyle\int_\Omega \ell(v^h)U}{\sqrt{\displaystyle\int_\Omega \ell(v^h)^2 dx}} = \frac{\displaystyle\sum_T |T|\ell(v)_{|T}U_T}{\sqrt{\displaystyle\sum_T |T|\ell(v)^2_{|T}}}.$$

We have also

$$\ell(v)^2_{|T} \leq \big(\sum_{\sigma\in T}\beta_\sigma^2\big)\big(\sum_{\sigma\in T} v_\sigma^2\big) \leq \sum_{\sigma\in T} v_\sigma^2$$

because $\beta_\sigma \geq 0$. Thus we have

$$\sum_T |T|\big(\sum_{\sigma\in T} v_\sigma^2\big) \leq \int_\Omega \ell(v^h)^2 dx,$$

and then

$$\frac{\displaystyle\int_\Omega \ell(v^h)U}{\sqrt{\displaystyle\int_\Omega \ell(v^h)^2 dx}} \geq \frac{\displaystyle\sum_T |T|\big(\sum_\sigma (U_T\beta_\sigma)v_\sigma\big)}{\displaystyle\sum_T |T|\big(\sum_{\sigma\in T} v_\sigma^2\big)}.$$

Taking the supremum of these inequalities, we get

$$\sup_{v^h\in V_h^k} \frac{a(v^h,u^h)}{||v^h||_{0,\Omega}} \geq \sqrt{\sum_T |T|U_T^2\big(\sum_\sigma \beta_\sigma^2\big)}.$$

Since $\displaystyle\sum_\sigma \beta_\sigma = 1$ and $\beta_\sigma \geq 0$, there exists a constant $\alpha > 0$ such that

$$\sup_{v^h\in V_h^k} \frac{a(v^h,u^h)}{||v^h||_{0,\Omega}} \geq \alpha\sqrt{\sum_T |T|U_T^2}.$$

Coming back to the original problem, we have shown the existence of $C > 0$ such that

$$C\sqrt{\sum_T \sum_{T'_T\subset T} \frac{\left(\displaystyle\int_{T'_T} \vec{\lambda}\cdot\nabla u^h\right)^2}{T'_T}} \leq ||\Pi_\ell(\vec{\lambda}\cdot\nabla u^h)||_{0,\Omega}. \tag{26}$$

If $k = 1$, we have $U_T = \vec{\lambda}\cdot\nabla u^h$, so (26) is nothing more than

$$C||\vec{\lambda}\cdot\nabla u^h||_{0,\Omega} \leq ||\Pi_\ell(\vec{\lambda}\cdot\nabla u^h)||_{0,\Omega}.$$

In all cases $k > 1$, the left hand side of (26) is quadratic in u^h, and it is easy to prove that if

$$\sum_{T'_T\subset T} \frac{\left(\displaystyle\int_{T'_T} \vec{\lambda}\cdot\nabla u^h\right)^2}{T'_T} = 0,$$

the $k^2 > (k+1)(k+2)/2$ terms of the sum are all vanishing so that $\vec{\lambda}\cdot\nabla u^h = 0$ in the vector space of polynomials of degree k on T, denoted $P_k(T)$.

This shows that on the finite dimensional space $P_k(T)$, the two quadratic forms

$$\int_T \big(\vec{\lambda}\cdot\nabla u^h\big)^2$$

and

$$\sum_{T'_T \subset T} \frac{\left(\displaystyle\int_{T'_T} \vec{\lambda} \cdot \nabla u^h\right)^2}{T'_T}$$

are equivalent. Hence, there exists $c_k > 0$ such that

$$\sum_{T'_T \subset T} \frac{\left(\displaystyle\int_{T'_T} \vec{\lambda} \cdot \nabla u^h\right)^2}{T'_T} \geq c_k \int_T \left(\vec{\lambda} \cdot \nabla u^h\right)^2.$$

This ends the proof : that there exists $C_k > 0$ such that

$$\inf_{u^h \in V_h^k} \sup_{v^h \in V_h^k} \frac{a(v^h, u^h)}{\sqrt{\displaystyle\int_\Omega (v^h)^2 dx} \sqrt{\displaystyle\int_\Omega \left(\vec{\lambda} \cdot \nabla v_h\right)^2}} \geq C_k > 0.$$

$\square$