An analysis of accuracy and convergence of finite volume methods.

Ali REZGUI[1]

Abstract. This paper presents a general framework for the analysis of finite volume schemes, applied to the discretization of general systems of conservation laws on bidimensional triangular meshes. A clear definition of accuracy in the finite volume sense is given for a large class of schemes. This notion is also compared to the classical truncation-error (or finite difference accuracy) concept. Some error estimate results are established, in the multidimensional linear case, in order to demonstrate the validity and the pertinence of the new concept. Some examples of schemes, for which the convergence analysis of finite difference theory does not apply, are also given and analysed, in the linear one-dimensional case, using the present framework.

1 Introduction

Finite volume methods (FV) are widely used in applications because of their great ability to handle complex geometries. The principle of these methods consists of discretizing a system of conservation laws in its integral form. In two dimensions, for example, a system of conservation laws can be written in the following form :

$$\frac{d}{d\,t}\int_{\Omega} w\,d\Omega + \int_{\partial\Omega} F.n\,d\Gamma = 0 \tag{1}$$

where the unknown $w = w(x,t)$ is a vector taking its values in an open set of $\mathbb{R}^d$, $d\Omega$ (resp. $d\Gamma$) is a measure on $\mathbb{R}^2$ (resp. $\mathbb{R}$) and $F = (f, g)$ is the physical flux which depends on x and t (for a large class of physical problems, the flux F is a function of w also). Note that Ω is an open set of $\mathbb{R}^2$, the boundary of which is $\partial\Omega$ and n is the outward unit normal vector to the boundary of Ω.
Finite difference methods (FD) discretize the system of conservation laws under its differential form :

$$w_t + div\,F = 0 \tag{2}$$

where div is the divergence operator. On uniform cartesian meshes, the two methods (FV and FD) often lead to the same schemes ; this is why the distinction between them is usually not clear. Observe now that, on irregular meshes, there is no equivalence between the two methods. Let us consider, for example, a one-dimensional hyperbolic problem written in its differential form:

$$w_t + f(w)_x = 0 \tag{3}$$

that we discretize on a mesh defined by its interfaces $x_{j+\frac{1}{2}}$. The cells, noted $\Omega_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$, are of measures δx_j and centers x_j. We use the classical finite volume Roe scheme to discretize this problem in space. This scheme is given, by :

$$w_{t|j} + \frac{1}{\delta x_j}(h_{j+\frac{1}{2}} - h_{j-\frac{1}{2}}) = 0 \tag{4}$$

where $w_j = w(x_j, t)$ and the numerical flux is given by :

$$h_{j+\frac{1}{2}} = h^l_{j+\frac{1}{2}} = \frac{1}{2}\left(f_{j+1} + f_j\right) - \frac{1}{2}|A_R|_{j+\frac{1}{2}}\left(w_{j+1} - w_j\right). \tag{5}$$

[1]SINUMEF Lab., ENSAM, 151 Boulevard de l'Hôpital, 75013 Paris, France.
Present adress :14, Place Sugny, 63000 Clermont-Ferrand, France. Tel : (00 33) 4 73 23 99 17.

with $A_{R|j+\frac{1}{2}} = A_R(w_j, w_{j-1}) = A_R(w(w_j, w_{j-1}))$ the Roe average, associated to the Jacobian A of f and the values w_j and w_{j-1} (see [17]), and$|A_R|$ the absolute value of A_R. For sake of simplicity, let us consider the linear case ($f(w) = a\ w,\ a \in$ IR). We can easily show that the scheme is non consistent in the classical finite-difference (or truncation error) sense, which means that the truncation error defined by :

$$e_{t|j} = u_{t|j} + \frac{1}{\delta x_j}(h(u)_{j+\frac{1}{2}} - h(u)_{j-\frac{1}{2}})$$

with u an exact solution of (3) does not go to 0 when $max_j \delta x_j$ goes to 0 (see [19] for example) ; we can hence not deduce convergence from stability and consistency in the finite difference sense.
On the other hand, we know that this scheme is convergent in $L^1_{loc}(\mathbb{R} \times \mathbb{R}^*_+)$ (see Sanders [18] for details). The proof of convergence given by Sanders uses the notions of flux-consistency, conservativity and BV stability. This example shows that the classical finite difference (FD) theory is not sufficient, in general, for the study of FV schemes.

We note also that convergence analyses of FV methods are rare in the literature. We mention however the work of Manteufel and *al.* [13] who give error estimates for some FV schemes using finite difference technics but their method does not seem to generalise to multidimensional problems on curvilinear meshes. The analysis of some cell-vertex methods applied to the advection-diffusion equation can be found in [14] and [1], where the authors use the finite element framework. However, their methods cannot be applied to hyperbolic systems. For hyperbolic problems, many schemes have been proved to be convergent (see [4], [5] for a large review). The difficulties due to the nonlinear nature of the problems allow only the derivation of $d^{1/4}$ or $d^{1/2}$ estimates, where d is the "mesh size". We mention finally the work of Eymard, Gallouët, and Herbin (see [3], [8] and references cited therein and see also [2] for a general review) on the convergence of first-order FV methods for general problems (hyperbolic, elliptic and parabolic systems) ; the important points involved there are the use of fundamental FV methods properties : flux-consistency, conservativity and some "stability" relations.

The aim of this paper is to extend the Lax-equivalence theorem to finite volume schemes, which discretises general systems of conservation laws, by following and extending the Eymard and *al.* work. Following previous works, as Eymard and *al.* [8], [3] and Harten and *al.*[7], we propose a clear definition of accuracy in the finite volume sense, for a large class of physical problems. We then justify this notions by establishing some error estimates in the general linear case. Here, we are essentially interested in the spatial accuracy property, that is why a major part of this paper deals with semi-discrete schemes (time discretization is briefly considered). This paper generalises and completes the results established in [15] and [16] (where the convergence analysis of finite volume schemes is studied only for one dimensional hyperbolic equations). We now turn to the description of this paper.
Section 2 deals with the discretization of bidimensional systems of conservation laws, on unstructured meshes, by FV methods. We present a large class of FV schemes which includes compact, non-compact, cell-center and cell-average schemes. We establish, in Section 3, some error estimate results for bidimensional linear FV methods. We then apply the results proved here to two simple cases. We finally, and briefly, discuss the extension of the results to fully discrete schemes and conclude.

2 Accuracy of finite volume methods

2.1 General framework

In this section, we focus on the discretization of bidimensional systems of conservation laws on triangular meshes. Consider a mesh composed of a family of conforming disjoint triangles of $\mathbb{R}^2$. This means

that the intersection of two different triangles is either the empty set, a common vertex or an entire common edge. Assume also that all angles of the triangles are less than $\pi/2$; this property ensures that the orthogonal bisectors intersect in a point, denoted x_j, which is inside each triangle. The triangles are denoted Ω_j, their measures in $\mathbb{R}^2$ are noted $|\Omega_j|$ and their boundaries $\partial\Omega_j$. The edges of the triangles are denoted Γ, the measures of which in $\mathbb{R}$ are denoted $|\Gamma|$.
Let us consider now a regular vector ν defined on the edges Γ and define the upstream and the downstream cells, $\Gamma,+$ and $\Gamma,-$, on an edge $\Gamma \in \partial\Omega_j \cap \partial\Omega_k$ with respect to ν in the following way. If $|\nu_\Gamma.\overrightarrow{x_j x_k}|$ is greater or equal to 0 (resp. lesser than 0) then take $\Gamma,+ = \Omega_k$ and $\Gamma,- = \Omega_j$ (resp. $\Gamma,+ = \Omega_j$ and $\Gamma,- = \Omega_k$). We put finally $d_\Gamma = |x_j\, x_k|$.

The mesh is supposed to be non degenerate in the following sense :

$$\exists\, c\ \in]0,1[, \quad \exists\, d\ \in\ \mathbb{R}^+ \quad \forall\, j \in\ \mathbb{Z}\ \begin{cases} \forall \Gamma \in \partial\Omega_j\ \ cd\ \leq\ \ |\Gamma|\ \leq d \\ cd^2\ \leq\ \ |\Omega_j|\ \leq d^2. \end{cases}$$

This hypothesis is classical in finite element theory.

Finite volume method consists in discretizing the system of conservation laws, in its integral form, on the mesh cells, *i.e.*

$$\frac{d}{d\,t}\int\limits_{\Omega_j} w\, d\Omega + \sum_{\Gamma\in\partial\Omega_j}\int\limits_{\Gamma} F.n_\Gamma d\Gamma = 0. \tag{6}$$

Since FV accuracy is essentially a spatial property, we only consider semi-discrete schemes of the form :

$$|\Omega_j|\, \mathcal{O}(w_{t|p}, w_{p,p\in D_{\mathcal{O}}})_{|j} + \sum_{\Gamma\in\partial\Omega_j} |\Gamma| h_\Gamma(w_{t|p}, w_{p,p\in D_\Gamma}, n_\Gamma) = 0 \tag{7}$$

where h_Γ is the numerical flux across the cell face Γ, which depends on w_p and $w_{t|p}$, p being a point in some domain noted D_Γ, and $\mathcal{O}$ is a difference operator, the stencil of which is noted $D_{\mathcal{O}}$. Observe that the first term of (7) represents a discretization of the temporal integral and the second term represents a discretization of the integral flux across the cell face. Note also that the discretization of both terms can depend on w and w_t so that our formalism includes compact and non-compact (or classical) schemes (see below for some examples). The scheme (7) is supposed to be conservative, *i.e.*

$$h_\Gamma(w_{t|p}, w_{p,p\in D_\Gamma}, n_\Gamma) = -h_\Gamma(w_{t|p}, w_{p,p\in D_\Gamma}, -n_\Gamma). \tag{8}$$

Before defining consistency, let us decompose our numerical flux in two parts : the first denoted h^e and called *explicit* depends only on w and the second, denoted h^i and called *implicit* is supposed to be linear with respect to w_t. The operator $\mathcal{O}$ is also split in explicit and implicit parts ; the implicit part is again supposed to be linear in w_t. Hence, we can rearrange our scheme and write it in the following form

$$|\Omega_j|\, \mathcal{I} w_{t|j} + \sum_{\Gamma\in\partial\Omega_j} |\Gamma| h^e_\Gamma(w_{p,p\in D_\Gamma}, n_\Gamma) = 0 \tag{9}$$

where $\mathcal{I}$ is the new difference operator built after regrouping the implicit parts which come from the discretization of the temporal and the flux integrals. The numerical flux h^e depends only on w and is supposed to be conservative. The scheme (7) is said to be consistent if :

$$\begin{cases} \forall p \in D_\Gamma\ \ w_p = w\ \text{ then }\ h^e(w_{p,p\in D_\Gamma}, n) = F(w)\,.n \\ \forall p \in D_{\mathcal{I}|j}\ \ w_p = w\ \text{ then }\ \mathcal{I} w_t|_j = w_t. \end{cases} \tag{10}$$

Observe that, in the case of hyperbolic systems -for which conservativity is important-, we do not have to ensure any conservativity property of the operator $\mathcal{I}$ in order to have the convergence of the numerical solution to a weak solution of the system of conservation law. In fact, we have the following result :

Theorem 1 *The semi-discrete scheme (7) is supposed to satisfy the hypotheses of conservativity and consistency (8)-(10) and the operators h^e and $\mathcal{I}$ are supposed to be Lipschitz continuous. Furthermore, we suppose that the numerical solution is bounded and converges, almost every where in $L^1_{loc}(\mathbb{R}^2 \times \mathbb{R}^*_+)$, to some function $w^*(x,t)$ when the mesh size d goes to 0 . Then $w^*(x,t)$ is a weak solution of the system (1).*

Proof : The proof of this result is similar to the proof of Lax-Wendroff theorem [9] and is therefore omitted.

Generally, to construct a finite volume scheme, one has to build an approximation of the physical flux at the cell faces Γ. For instance, we mention the work of Harten and *al.* [6],[7] on ENO schemes on structured meshes. The authors consider cell-average schemes, which means that the unknown w_j is an approximation of the average of w over the cell Ω_j, *i.e.*

$$w_j = \frac{1}{|\Omega_j|} \int\limits_{\Omega_j} w \, d\Omega.$$

On the other hand, Eymard, Gallouët and Herbin, (see [3], [8]) construct FV schemes on unstructured meshes where the unknowns are placed in the centers of the cells (These are called cell-center schemes). In addition, the authors prove the convergence of their schemes, which are first order accurate in the FV sense.

Before extending these notions of accuracy, let us introduce some norms suitable for handling discrete functions. We define a scalar product as following :

$$\forall \ (u)_{j \in \mathbb{Z}}, (v)_{j\in\mathbb{Z}}, \ \ (u,v) = \sum_{j=-\infty}^{\infty} |\Omega_j| u_j v_j .$$

where the discrete functions u_j and v_j are defined on the cells Ω_j (if $u_j = (u_j(1), ..., u_j(d))$ is a vector of $\mathbb{R}^d$ then take $u_j v_j = \sum_k u_j(k) v_j(k)$). The norm associated to this scalar product, namely the L^2 norm, is denoted $|u|_{L^2} = (u,u)^{1/2}$. If a function w is defined by its values on the edges Γ of the triangulation, then its L^2 norm will be given by the following formula :

$$|w|_{L^2} = \left(\sum_{\Gamma} |\Omega_\Gamma| w_\Gamma w_\Gamma \right)^{1/2}$$

where Ω_Γ is the quadrilater having the two vertices of Γ and the centers of $(\Gamma,+)$ and $(\Gamma,-)$ as vertices. Here the summation is taken over all the edges of the triangulation.

We extend now the notions of accuracy introduced above in order to cover compact and non compact schemes and also cell-average and cell-center schemes.

Definition 1 *The scheme (7) (or (9)) is said to be of order r in the finite volume sense if :*

$$\begin{cases} [a] \forall \ j \in \mathbb{Z} \quad (\mathcal{O}(u_{t|p}, u_{p,p\in D_{\mathcal{O}}}))_j = \dfrac{1}{|\Omega_j|} \displaystyle\int\limits_{\Omega_j} u_t d\Omega + d^r R^1_j, \\ [b] \forall \ j \in \mathbb{Z}, \forall \ \Gamma \in \partial\Omega_j \qquad h_\Gamma(u_{t|p}, u_{p,p\in D_\Gamma}, n_\Gamma) = \dfrac{1}{|\Gamma|} \displaystyle\int\limits_{\Gamma} F(u).n d\Gamma + d^r R^2_\Gamma .n_\Gamma \end{cases}$$

where u is an exact regular solution of the system (1) and R^1_j and R^2_Γ are some continuous functions depending on u. These functions are supposed to be in $L^2 \cap L^\infty$.

In the works of Harten and *al.* [6]-[7], the unknowns are the cell-averages and the definition of accuracy is similar to our definition but without the operator $\mathcal{O}$ ($\mathcal{O} = I$). This accuracy, notion called the accuracy in the mean-value sense, is then only characterized by part [b] of definition 1. To justify this notions, the authors suppose that the function $R^2(\Gamma)$ is Lipschitz-continuous with respect to space variables which is equivalent to supposing that the mesh is regular in some sense or that the scheme is consistent in the finite difference sense (see [16]). The justification of this notion therefore seems to be incomplete.
Observe also that the notion of FV accuracy (definition 1) is not equivalent to the classical definition of accuracy in the finite difference sense. In order to see this, consider the scheme (7) as a cell-center scheme. It is then easy to prove that the scheme is at most of order $r-1$ in the classical truncation error (or FD) sense.

2.2 Examples

Let us now give some examples of FV schemes ; in order to avoid long and complicated formulae, we present only one-dimensional schemes. Consider then a one-dimensional mesh defined by its cell faces $x_{j+\frac{1}{2}}$; the cells are denoted $\Omega_j = [x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$; their centers are denoted x_j and their measures δx_j ; we denote also $d_j = x_j - x_{j-1}$. The functions w_j represent the values of w in the center of Ω_j.
We now present three schemes which discretize the hyperbolic system of conservation laws (3) and a scheme discretizing a parabolic problem.

Roe scheme
It is defined by (4). We remind that the numerical flux is given by :

$$h^l_{j+\frac{1}{2}} = \frac{1}{2}\left(f_{j+1} + f_j\right) - \frac{1}{2}|A_R|_{j+\frac{1}{2}}\left(w_{j+1} - w_j\right).$$

On general meshes, this scheme is first-order accurate in the FV sense (Definition 1) and inconsistent in the truncation error sense.

The centered scheme
This scheme can be written in the form (4) with the numerical flux :

$$h^c_{j+\frac{1}{2}} = \frac{\delta x_j f_{j+1} + \delta x_{j+1} f_j}{\delta x_j + \delta x_{j+1}}$$

It is second-order accurate in the FV sense (Definition 1) but only first-order accurate in the truncation error (or FD) sense.

The compact upwind scheme
We proposed this scheme in [10]-[11] and [16] in order to solve the Euler equations. It is defined as follows :

$$\delta x_j w_{t|j} + \frac{\delta x_j^3}{12(d_{j+1} + d_j)}(h^{"}_{j+\frac{1}{2}} - h^{"}_{j-\frac{1}{2}}) + (h_{j+\frac{1}{2}} - h_{j-\frac{1}{2}}) = 0 \tag{11}$$

$$\left\{\begin{array}{rcl} h^{"}_{j+\frac{1}{2}} & = & 2\dfrac{w_{t|j+1} - w_{t|j}}{\delta x_j + \delta x_{j+1}} \\ h_{j+\frac{1}{2}} & = & h^l_{j+\frac{1}{2}} + \dfrac{1}{4}\left[(\delta x_{j+1} - \delta x_j) - sgn(A_R)_{j+\frac{1}{2}}(\delta x_{j+1} + \delta x_j)\right]\dfrac{\delta x_j w_{t|j+1} + \delta x_{j+1} w_{t|j}}{\delta x_j + \delta x_{j+1}} \\ & + & \dfrac{1}{16}\left[(\delta x_{j+1}^2 + \delta x_j^2) - sgn(A_R)(\delta x_{j+1}^2 - \delta x_j^2)\right] h^{"}_{j+\frac{1}{2}}. \end{array}\right. \tag{12}$$

This scheme is third-order accurate in the FV sense (Definition 1) but only second-order accurate in the truncation error (or FD) sense. It is also a 3-point scheme.

Remark 1 *For the first two schemes, the order of accuracy is unchanged if one takes w_j as the cell-average on Ω_j instead of the value at the cell center.*

We also give a classical example which discretize the following parabolic one-dimensional equation :

$$w_t - w_{xx} = 0.$$

The classical scheme is given by :

$$\left\{ \begin{array}{l} \delta x_j w_{t|j} - (h_{j+\frac{1}{2}} - h_{j-\frac{1}{2}}) = 0 \\ h_{j+\frac{1}{2}} = \dfrac{w_{j+1} - w_j}{d_{j+1}} \end{array} \right. \tag{13}$$

where w_j is the value of w at the center x_j. Observe that $h_{j+\frac{1}{2}}$ is an approximation of the flux w_x across the cell-face $x_{j+\frac{1}{2}}$. This scheme is first-order accurate in the FV-sense (Definition 1) on general meshes but it is inconsistent in the truncation error sense.
The following section gives some justifications to the FV-accuracy notion.

3 Error estimate

3.1 Semi-discrete schemes

In this section, we study the convergence of finite volume schemes described above. To justify the new notion of accuracy, one has to give some error estimate of the above methods because the finite difference theory is not able to supply convergence.
We begin by presenting a brief description of the methodology and results of Eymard, Gallouët and Herbin [8], [3] when dealing with steady linear advection-diffusion equations. They discretize the system of conservation laws on the mesh cells and store the unknown in the cell-centers ; fluxes across an edge of the cells are approximated by using a first-order formula. An equation on the error is then formed. Conservativity, consistancy of the fluxes (finite volume accuracy) and some stability relations are then used to establish some error estimates. The schemes studied in their work are first-order accurate in the finite volume sense. They establish error estimate in d. In this paper, we use a similar methdodology to study the family of finite volume schemes considered above.
We now consider a scalar linear bidimensional problem, *i.e.* the functions f and g in (1) are linear with respect to w, which we discretize by a linear FV scheme (which means that the scheme is linear in w_j).
We write the FV scheme (7) in the form :

$$|\Omega_j|\,\mathcal{I}w_{t|j} + \mathcal{E}w_{|j} = 0 \tag{14}$$

where $\mathcal{E}$ is a linear operator and $\mathcal{I}$ is a bounded positive-linear operator.
We introduce $\mathcal{I}^*$ the adjoint of $\mathcal{I}$ in the linear algebra sense and we observe that the function $|u| = (\mathcal{I}^*\mathcal{I}u, u)$ is a norm which is equivalent to the L^2 norm, introduced above. This norm will be widely used in the following sections in order to simplify the presentation.

Classically, stability is defined as follows :

Definition 2 *The scheme (7) is said to be stable in L^2 if there exists a constant C depending on $T > 0$ so that :*

$$\forall\ t \leq T \quad |w(t)| \ \leq \ C|w(t=0)|$$

Multiply (14) by $\mathcal{I}w_{|j}$ and sum over $j \in \mathbb{Z}$, we obtain :

$$\sum_j \Omega_j \mathcal{I}w_{|j}\ \mathcal{I}w_{t|j} + \sum_j \mathcal{I}w_{|j}\ \mathcal{E}w_{|j} = 0$$

or :

$$(\mathcal{I}w, \mathcal{I}w_t) + \sum_j \mathcal{I}w_{|j}\ \mathcal{E}w_{|j} = 0$$

therefore:

$$\frac{1}{2}|w^2|_t + \sum_j \mathcal{I}w_{|j}\ \mathcal{E}w_{|j} = 0.$$

If we suppose that :

$$\sum_j \mathcal{I}w_{|j}\ \mathcal{E}w_{|j}\ \geq\ 0 \tag{15}$$

then

$$\forall\, t\ |w(t)|\ \leq\ |w(t=0)|$$

The consequence of this analysis is that the inequality (15) ensures the stability. This condition is supposed to be satisfied for all FV schemes considered here.
If the order of accuracy of the scheme is greater than one, then we can give the following result, after noting $e = u - w$, the numerical error.

Theorem 2 *The scheme (7) is supposed to be stable in the sense of the condition (15) and of order r with $r \geq 2$ in the finite volume sense (Definition 1). Then we have the following error estimate, if u is regular enough :*

$$\forall\, t \leq T \quad |e(t)|^2\ \leq\ \ C_1|e(t=0)|^2 + C_2 d^{2(r-1)}.$$

The constants C_1 and C_2 depend on u, T and the mesh regularity.

Proof. the scheme (7) is of order r in the FV sense then there exists two continuous bounded functions R^1_j et R^2_Γ of L^2 depending on u such that :

$$\forall j \in \mathbb{Z},\ \ \Gamma \in \partial\Omega_j\ \ \left\{\begin{array}{l} (\mathcal{O}(u_{t|p}, u_{p,p\in D_\mathcal{O}}))_j = \dfrac{1}{|\Omega_j|}\displaystyle\int_{\Omega_j} u_t\, d\Omega + d^r R^1_j \\ h(u_{t|p}, u_{p,p\in D_\Gamma}, n_\Gamma)_\Gamma = \dfrac{1}{|\Gamma|}\displaystyle\int_\Gamma F(u) d\,\Gamma + d^r R^2_\Gamma.n_\Gamma. \end{array}\right.$$

Then u satisfies :

$$|\Omega_j|(\mathcal{O}(u_{t|p}, u_{p,p\in D_\mathcal{O}}))_j + \sum_{\Gamma\in\partial\Omega_j} |\Gamma| h(u_{t|p}, u_{p,p\in D_\Gamma})_\Gamma = d^r|\Omega_j|R^1_j + d^r \sum_{\Gamma\in\partial\Omega_j} |\Gamma| R^2_\Gamma.n_\Gamma.$$

We then use the operators $\mathcal{I}$ and $\mathcal{E}$ introduced earlier in order to write this relation in the following form :

$$|\Omega_j|\mathcal{I}u_{t|j} + \mathcal{E}u_{|j} = d^r|\Omega_j|R^1_j + d^r \sum_{\Gamma\in\partial\Omega_j} |\Gamma| R^2_\Gamma.n_\Gamma. \tag{16}$$

By subtracting (9) from (16), one gets :

$$|\Omega_j|\mathcal{I}e_{t|j} + \mathcal{E}e_{|j} = d^r|\Omega_j|R^1_j + d^r \sum_{\Gamma\in\partial\Omega_j} |\Gamma| R^2_\Gamma.n_\Gamma \tag{17}$$

where $e_j = u_j - w_j$ is the error at the cell j. Multiplying now the equation (17) by $\mathcal{I}e_{|j}$ and summing over j, we obtain :

$$\underbrace{\sum_j |\Omega_j| \mathcal{I}e_{|j}\, \mathcal{I}e_{t|j}}_{T_1} + \underbrace{\sum_j \mathcal{I}e_{|j}\, \mathcal{E}e_{|j}}_{T_2} = \underbrace{\sum_j |\Omega_j| \mathcal{I}e_{|j}\, (d^r R_j^1)}_{T_3} + \underbrace{\sum_j \mathcal{I}e_{|j}\, d^r \sum_{\Gamma \in \partial\Omega_j} |\Gamma| R_\Gamma^2 . n_\Gamma}_{T_4} . \qquad (18)$$

We can easily see that $T_1 = |e|_{|t}^2/2$ and that $T_2 \geq 0$ (this condition is ensured by the inequality of stability (15)). By applying the Young inequality to the term T_3, we obtain :

$$T_3 \leq \frac{1}{2} d^{2r} |R^1|_{L^2}^2 + \frac{1}{2} |\mathcal{I}e|_{L^2}^2 = \frac{1}{2} d^{2r} |R^1|_{L^2}^2 + \frac{1}{2} |e|^2.$$

we can see that the term T_4 can be written, after summing by parts an reordering the summation over the set of edges, as following :

$$T_4 = d^r \sum_j \mathcal{I}e_{|j} \sum_{\Gamma \in \partial\Omega_j} |\Gamma| R_\Gamma^2 . n_\Gamma = d^r \sum_\Gamma |\Gamma| R_\Gamma^2 . n_\Gamma \, (\mathcal{I}e_{\Gamma,+} - \mathcal{I}e_{\Gamma,-}) \qquad (19)$$

where the summation is taken over all the triangulation of Ω and n_Γ denotes the unit normal to Γ outward to $\Gamma, +$.
We then have

$$T_4 \leq \sum_\Gamma d|\Gamma| \, (d^{r-1} |R_\Gamma^2 . n_\Gamma|) \, (|\mathcal{I}e_{\Gamma,+}| + |\mathcal{I}e_{\Gamma,-}|)$$

Apply now Young inequality to obtain

$$T_4 \leq d^{2r-2} \left(\sum_\Gamma d|\Gamma| \, |R_\Gamma^2 . n_\Gamma| \right)^2 + \frac{1}{2} \left(\sum_\Gamma d|\Gamma| \, |\mathcal{I}e_{\Gamma,+}| \right)^2 + \frac{1}{2} \left(\sum_\Gamma d|\Gamma| \, |\mathcal{I}e_{\Gamma,-}| \right)^2$$

The mesh is regular enough so that the two last functions, in this equation, define norms of $\mathcal{I}e$. These norms are equivalent norms to L^2 norm, then we can find a constant b_0 depending on c and bounded independently on d so that

$$T_4 \leq d^{2r-2} |R^2|_{L^2}^2 + \frac{1}{2}(b_0 - 1)|e|^2$$

The functions R^1 and R^2 are in $L^2 \cap L^\infty$ hence the inequality (18) gives :

$$\frac{1}{2}(|e|^2)_t \;\leq\; \frac{1}{2} b_0 |e|^2 + \frac{1}{2} b_1 d^{2r} + \frac{1}{2} b_2 d^{2r-2}$$

where b_1 and b_2 are constants depending on the mesh and the exact solution. Now we have:

$$(|e|^2)_t \;\leq\; b_0 |e|^2 + b_1 d^{2r} + b_2 d^{2r-2}.$$

Multiplying by $exp(-b_0 t)$, one gets :

$$\left(exp(-b_0 t)|e|^2 \right)_t \;\leq\; b_1 exp(-b_0 t) d^{2r} + b_2 exp(-b_0 t) d^{2r-2}.$$

Integrating now on $[0, t]$ with respect to time, it follows that :

$$exp(-b_0 t)|e(t)|^2 - |e(0)|^2 \;\leq$$

$$\sup_{t \leq T} b_1(t) \, [1 - exp(-b_0 T)] \, d^{2r}/2 + \sup_{t \leq T} b_2(t) \, [1 - exp(-b_0 T)] \, d^{2r-2}/2.$$

We finally have to multiply by $exp(b_0 t)$ to get the desired result.□

The error estimate presented above does not apply to first order schemes ($r = 1$). We present then another result valid for all finite volume schemes, including first-order methods.

We define the operators δ_Γ and $\mathcal{K}$ by :

$$\delta_\Gamma f = f_{\Gamma,+} - f_{\Gamma,-} \quad , \qquad \mathcal{K} = (\delta_\Gamma \mathcal{I})^* \frac{1}{d_\Gamma^2} \delta_\Gamma \frac{1}{|\Omega|} \mathcal{E}.$$

Theorem 3 *The scheme (7) is supposed to be stable in the sense of the condition (15) and of order r ($r \geq 1$) in the finite volume sense (Definition 1). Suppose also that the exact solution u is of class $\mathcal{C}^{r+1}$ and that its first derivate is in L^2. Then, if the operator $\mathcal{K}$ is non negative we have the following error estimate :*

$$\forall\, t \leq T \quad |e(t)|^2 \;\leq\; C_1 |e(t=0)|^2 + C_2 d^r.$$

The constants C_1 and C_2 depend on u, T and the mesh regularity.

Proof. There are few differences between this proof and the proof of theorem 2 : the main difference is due to the treatment of the term T_4. Starting from equation (19), we can see that the term T_4 can be written as following :

$$T_4 = d^r \sum_\Gamma |\Gamma| d_\Gamma R_\Gamma^2 . n_\Gamma \frac{\mathcal{I} e_{|\Gamma,+} - \mathcal{I} e_{|\Gamma,-}}{d_\Gamma}$$

where the summation is taken over all the triangulation of Ω and n_Γ denotes the unit normal to Γ outward to $(\Gamma, +)$. Observe that the function :

$$|u|_1 = \left(\sum_\Gamma |\Gamma| d_\Gamma \; u_\Gamma^2 \right)^{1/2} , \quad (u)_{j \in \mathbb{Z}}$$

is a discrete norm equivalent to the L^2 norm, because the mesh is non-degenerate.
We apply the Cauchy-Schwarz inequality to T_4 :

$$T_4 \leq b_3 \, d^r \; |R_\Gamma^2|_{L^2} \; |e^{(1)}|$$

where b_3 is a constant depending on the mesh regularity and the "discrete derivate" of the error $e^{(1)}$ is defined by :

$$e_\Gamma^{(1)} = \frac{\mathcal{I} e_{|\Gamma,+} - \mathcal{I} e_{|\Gamma,-}}{d_\Gamma}$$

It is then clear that if we establish that $e^{(1)}$ is in L^2, independently on the mesh size d, we will obtain the desired result, *i.e.* an error estimate of $|e|$ "in $d^{r/2}$".
We have :

$$e_\Gamma^{(1)} = u_\Gamma^{(1)} - w_\Gamma^{(1)}$$

It is easy to prove that $u_\Gamma^{(1)}$ belongs to L^2 if the first derivate of the exact solution belongs to this space. To prove that $e^{(1)}$ belongs to L^2 as well, it is then sufficient to prove that $w^{(1)}$ belongs to L^2.
From the scheme (14) and for two cells j and k having Γ as a common edge, we can deduce the following relation :

$$w_{t|\Gamma}^{(1)} + \frac{1}{d_\Gamma} \delta_\Gamma \frac{1}{|\Omega|} \mathcal{E} w_{|\Gamma} = 0.$$

Multiplying then by $|\Gamma|\delta_\Gamma \mathcal{I}\ w_{|\Gamma}$ and summing over Γ, then :

$$\frac{1}{2}\left[\left(|w^{(1)}|_1\right)^2\right]_t + \sum_\Gamma |\Gamma|\ \delta_\Gamma \mathcal{I} w_{|\Gamma} \cdot \frac{1}{d_\Gamma}\delta_\Gamma \frac{1}{|\Omega|}\mathcal{E} w_{|\Gamma} = 0.$$

The operator $\mathcal{K}$ is non negative hence :

$$\left[\left(|w^{(1)}|_1\right)^2\right]_t \ \leq\ 0$$

or

$$|w^{(1)}|(t)\ \leq\ |w^{(1)}|(t=0).$$

If the discretization of the initial condition is accurate enough then $w^{(1)}(t=0)$ will be very close to $u^{(1)}(t=0)$; this function belongs to L^2, then $w^{(1)}(t=0)$ belongs to this space too.
We have hence proved that $w^{(1)}$ is in L^2. The reminder of the proof is similar to the proof of theorem 2.
□

3.2 Application

It is not easy to study the convergence of bidimensional high-order FV methods : the main difficulty is obviously the stability property. We then study only the convergence of two simple methods in the one dimensional case.

First we prove that the upwind scheme (4) - (5) is convergent in the linear case ($f(w) = aw$, $a \in \mathbb{R}^+$). This scheme is first-order accurate in the FV sense (Definition 1) and it is stable in the sense of definition 2 ; so it is sufficient to prove that the discrete derivate of this scheme solution ($w^{(1)}$) (positivity of the operator $\mathcal{K}$) is in L^2 to prove an error estimate in $d^{1/2}$.
Starting from (4) - (5) applied for j and $j-1$, we can write the following equation :

$$\frac{(w_j - w_{j-1})_t}{\delta x_j} - \frac{1}{\delta x_j}\left(\frac{w_j - w_{j-1}}{\delta x_j} - \frac{w_{j-1} - w_{j-2}}{\delta x_{j-1}}\right) = 0$$

Multiply now by $w_j - w_{j-1}$ and sum over j :

$$\sum_j \frac{d_{j+\frac{1}{2}}}{\delta x_j} d_{j+\frac{1}{2}} (w^{(1)}_{j+\frac{1}{2}})_t (w^{(1)}_{j+\frac{1}{2}}) - \frac{1}{\delta x_j}\left(\frac{w_j - w_{j-1}}{\delta x_j} - \frac{w_{j-1} - w_{j-2}}{\delta x_{j-1}}\right)(w_j - w_{j-1}) = 0.$$

Thus

$$\sum_j \frac{d_{j+\frac{1}{2}}}{\delta x_j} d_{j+\frac{1}{2}} (w^{(1)}_{j+\frac{1}{2}})_t (w^{(1)}_{j+\frac{1}{2}}) + \frac{1}{2}\sum_j \left(\frac{w_j - w_{j-1}}{\delta x_j}\right)^2 = 0$$

which shows that

$$\sum_j \frac{d_{j+\frac{1}{2}}}{\delta x_j} d_{j+\frac{1}{2}} (w^{(1)}_{j+\frac{1}{2}})^2(t) \leq \sum_j \frac{d_{j+\frac{1}{2}}}{\delta x_j} d_{j+\frac{1}{2}} (w^{(1)}_{j+\frac{1}{2}})^2(t=0)$$

The mesh is non degenerate, hence $w^{(1)}$ is in L^2 and the scheme satisfies an error estimate in $d^{1/2}$.

We now prove the convergence of the scheme (13) which discretizes the diffusion equation. Here too, we observe that this scheme is first-order accurate in the FV sense and it is stable in the sense of definition 2 ; so it is sufficient to prove that the discrete derivate of this scheme solution ($w^{(1)}$) is in L^2 to prove an

error estimate in $d^{1/2}$.
Starting from (13) and following the convergence proof of the first-order scheme, we can write the following relation :

$$d_j w^{(1)}_{j+\frac{1}{2}}|_t + \frac{1}{\delta x_j}\left(w^{(1)}_{j+\frac{1}{2}} - w^{(1)}_{j-\frac{1}{2}}\right) - \frac{1}{\delta x_{j-1}}\left(w^{(1)}_{j-\frac{1}{2}} - w^{(1)}_{j-\frac{3}{2}}\right) = 0.$$

After multiplying by $w^{(1)}_{j+\frac{1}{2}}$ and summing over j, we obtain :

$$\sum_j d_j \left(w^{(1)}_{j+\frac{1}{2}}\right)^2_t + \sum_j \frac{1}{\delta x_j}\left(w^{(1)}_{j+\frac{1}{2}} - w^{(1)}_{j-\frac{1}{2}}\right)^2 = 0,$$

which shows clearly that the $w^{(1)}$ is in L^2 and concludes the proof of the stability and the error estimate.

The study of the scheme (13) is done in the spirit of the error estimate results established before. However, it is easy to have a better error estimate. Following the development of theorem 2, we can establish the following equation for the error :

$$\delta x_j e_{j|t} + \left(e^{(1)}_{j+\frac{1}{2}} - e^{(1)}_{j-\frac{1}{2}}\right) = d^r \delta x_j R^1_j + d^r \left(R^2_{j+\frac{1}{2}} - R^2_{j-\frac{1}{2}}\right)$$

Multiply by e_j, sum over j and sum by part to get :

$$\sum_j \delta x_j \left(e_{j+\frac{1}{2}}\right)^2_t + \frac{1}{2}\sum_j \left(e^{(1)}_{j+\frac{1}{2}}\right)^2 \leq \frac{1}{2} d^r |R^1|^2_{L^2} + \frac{1}{2}|e| + \sum_j d^r d_{j+\frac{1}{2}} R^2_{j+\frac{1}{2}} e^{(1)}_{j+\frac{1}{2}} \qquad (20)$$

Apply now the Young inequality to the last term of the left side of this inequality and follow the proof of theorems 2 and 3 to show that the error is bounded by some constant which multiplies d.
Note that similar results are obtained in [8] for a steady problem. The improvement of this result is due the elliptic nature of the problems treated there.
The analysis of the compact scheme is more difficult because of the great nonlinearity of the scheme with respect to the mesh. We are not able to give simple and general conditions ensuring the stability of these schemes.

3.3 Fully discrete schemes

We start from the semi-discrete scheme (7) or its equivalent form (9). This scheme can be written under the form :

$$\frac{d}{dt} w + \mathcal{L}(w) = 0 \qquad (21)$$

where $w = (..., w_j, w_{j+1}, ...)$ is the vector of unknowns and $\mathcal{L}$ is the space discretization operator (which is function of $\mathcal{I}$ and $\mathcal{E}$). The problem (21) can be discretized at time t^{n+1} by :

$$\mathcal{T}(v_{k|k=n+1-K,n+1}, \Delta t) = 0 \qquad (22)$$

where v^{n+1} is the approximation given by the scheme of the solution w at time t^{n+1}, Δt is the time step and $\mathcal{T}$ is an "abstract" time discretization operator which depends on $\mathcal{I}$ and $\mathcal{E}$. The classical notion of accuracy can be defined as following :

Definition 3 *The scheme (22) is of order s (in time) if, for a regular exact solution w of (21), there exists a function R^t depending on w and its time derivatives and being in $L^2(\mathbb{R}^*_+) \cap L^\infty(\mathbb{R}^*_+)$ such that :*

$$\mathcal{T}(w) = \Delta t^s \, R^t$$

We consider now the linear case. It is well known (Lax equivalence theorem) that if the scheme (22) is stable in the sense of definition 2 then it is convergent and that the L^2 error satisfies :

$$\forall t \leq T \quad |w - v| \leq C_t \, \Delta t^s$$

where C_t depends on T, the initial condition and w (and hence on u, the exact solution of system of conservation laws).
The L^2 error of the completely discrete FV method can be bounded as follows :

$$|e| = |u - v| \leq |u - w| + |w - v|$$

Which shows that it is bounded by the sum of space and time error discretizations. We have then proved :

Theorem 4 *The completely discrete s^{th} order in time and r^{th} order in space FV scheme (7)-(22) is supposed to be stable and to satisfy the hypothesis of theorems 2 and 3. Then the following estimate holds :*

$$\forall\, t \leq T \quad |e(t)|^2 \;\leq\; C_1|e(t=0)|^2 + C_1 d^{2q} + C_3 \Delta t^{2s}, \quad q = \max(r/2, r-1).$$

The constants C_1, C_2 and C_3 depend on u, T, the mesh regularity and the initial conditions.

As an application of this result, we can easily show that the schemes (4) and (13) discretised in time by the Euler-backward scheme are convergent under some restriction on the time step.

4 Conclusion

In this paper we studied the discretization of system of conservation laws, on unstructured meshes, by FV methods. We presented a large class of FV schemes which includes compact, non-compact, cell-center and cell-average schemes. A clear definition of the notion of accuracy in the finite volume sense was given. We established also some error estimate results for linear FV methods in the semi-discrete case. The theorical results obtained were applied to two simple FV methods. We have finally presented the great lines of the extension of the results to completely discrete schemes.

We now remark that the error estimates presented here can be extended to more general meshes : quadrilateral mesh, mesh composed of quadrilaters and triangles,... Obviously, the analysis remains valid for linear systems of equations. We also note that better error estimate could be established for elliptic problems but one has to use more informations on the structure of the spatial discretization operator (see the parabolic problem considered here).

Many directions have to be explored : we will try to find pertinent results about the linear stability of high order schemes on irregular meshes ; it also interesting to study non-linear elliptic and hyperbolic problems. We mention, in the end, that to give a complete convergence analysis, we have to study the influence of boundary conditions.

Adnowlegement. The author is grateful to Professors T. Gallouët and A. Lerat for their advice and comments on this paper.

References

[1] Z. Cai, J. Mandel and S. Mc Cormick, “The finite volume element method for diffusion equation on general triangulation”, *SIAM J. Numer. Anal.*, 28, p. 392-402 (1991).

[2] R. Eymard, T. Gallouët and R. Herbin, "Finite volume methods", Handbook of numerical analysis, Ph. Ciarlet and J. L. Lions eds, North Holland (1999), to appear.

[3] R. Eymard, T. Gallouët and R. Herbin, "Convergence of finite volume schemes for semilinear convection diffusion equations ", *Numer. Math.* 82, 1, p. 91-116 (1999).

[4] E. Godlewski and P. A. Raviart, "Hyperbolic systems of conservation Laws", SMAI (1991).

[5] E. Godlewski and P. A. Raviart, "Numerical approximation of hyperbolic systems of conservation Laws", Springer and Verlag (1996).

[6] A. Harten A. and S. Osher, "Uniformly high-order accurate non-oscillatory schemes I", *SIAM J. Numer. Anal.*, 24, p. 279-309 (1987).

[7] A. Harten, B. Engquist, S. Osher and S. R. Chakravarthy, "Uniformly high-order accurate non-oscillatory schemes III", *J. Comput. Phys.*, 71, p. 231-303 (1987).

[8] R. Herbin, An error estimate for a finite volume scheme for diffusion-convection problems on a triangular mesh, *Num. Meth. P.D.E.*, p. 166-173 (1995).

[9] P.D. Lax and B. Wendroff, "Systems of conservation laws", *Comm. Pure Appl. Math.*, 13, p. 217-237 (1960).

[10] A. Lerat and A. Rezgui, "Third order dissipative schemes for hyperbolic systems", *C. R. Acad. Sci.*, Paris, 323, II b, p. 397-403 (1996).

[11] A. Lerat and A. Rezgui, "Third-order accurate schemes for hyperbolic systems of conservation laws", August 1996, in BW Workshop on Computational Fluid Dynamics Proceed., Beijing, China, ed. F.G. Zhuang, p. 23-32 (1996).

[12] A. Lerat and A. Rezgui, "High-order accurate compact and non compact schemes for compressible flows", 7th Inter. Symp. on Computational Fluid Dynamics, Beijing, China, September 1997, in 7th ISCFD Proceed, Intern. Academic Publi., p. 180-185 (1997).

[13] T.A. Manteufel and A.B. White, "The numerical solution of second order boundary value problems on non uniform meshes", *Math. Comput*, 47, p. 511-536 (1986).

[14] K.W. Morton and M. Stynes, "An analysis of the cell vertex method", *M2AN*, 28, p. 699-724 (1994).

[15] A. Rezgui, "Accuracy and convergence of finite volume methods", *C. R. Acad. Sci.*, Paris, I, 327, p. 421-426 (1998).

[16] A. Rezgui, "Third-order accurate compact and non compact schemes for compressible flow calculations", Thesis, ENSAM, Paris (in french), (1997).

[17] P.L. Roe, "Approximate Riemann solvers, parameter vectors and difference schemes", *J. Comput. Phys* , 43, p. 357-372 (1981).

[18] R. Sanders, "On the convergence of monotone finite-difference schemes with variable space differencing", *Math. Comp.*, 40, p. 91-106 (1983).

[19] Turkel E., Accuracy of schemes with non-uniform meshes for compressible fluid flows, *Appl. Num. Math.*, 2, p. 529-550 (1985).