MATHEMATICS OF COMPUTATION
Volume 73, Number 247, Pages 1203–1234
S 0025-5718(03)01618-1
Article electronically published on November 5, 2003

LAX THEOREM AND FINITE VOLUME SCHEMES

BRUNO DESPRES

ABSTRACT. This work addresses a theory of convergence for finite volume methods applied to linear equations. A non-consistent model problem posed in an abstract Banach space is proved to be convergent. Then various examples show that the functional framework is non-empty. Convergence with a rate $h^{\frac{1}{2}}$ of all TVD schemes for linear advection in 1D is an application of the general result. Using duality techniques and assuming enough regularity of the solution, convergence of the upwind finite volume scheme for linear advection on a 2D triangular mesh is proved in L^α, $2 \leq \alpha \leq +\infty$: provided the solution is in $W^{1,\infty}$, it proves a rate of convergence $h^{\frac{1}{4}-\varepsilon}$ in L^∞.

1. INTRODUCTION

The Lax theorem states that stability and consistency are sufficient conditions for a linear scheme to be convergent. Many numerical examples show that stability is necessary **and** sufficient. However, there are numerical methods, such as finite volume methods in 2D on triangular meshes, which are formally non-consistent even for linear equations. Since numerical evidence shows that finite volume methods for linear equations are convergent, we infer that consistency (that is, consistency in the finite difference setting [28], [13], [27]) is **not necessary**. To our knowledge, a comprehensive explanation of this phenomena is still missing. This work intends to fill the gap, for **non-stationary linear equations**.

Finite volume methods are engineers' methods: in the rest of this paper finite volume stands for P^0 cell-centered numerical schemes. In contrast with the finite element methods [6], there is no functional framework for the introduction of finite volume numerical schemes. Despite this lack of mathematical foundations, finite volume methods are very robust and efficient for practical computations when applied to the direct simulation of complex physics. This is particularly the case in computational fluid dynamics. On the mathematical side, many researchers have stressed that proofs of convergence are very difficult to obtain for finite volume methods: see [20], [26], [7], [12], [15], [11], [30], [2] (based on [22]) and other papers therein; an up-to-date reference is [17]. A rule of thumb for numerical methods is that robustness of a scheme is partially a consequence of its dissipativity. On the mathematical side dissipativity is a standard way to obtain *a priori* estimates, and then to prove convergence. But this paradigm seems not to be true when applied to finite volume methods, mainly due to the formal non-consistency of finite volume methods. Even though it is not the purpose of this work to treat higher

Received by the editor November 28, 2001 and, in revised form, January 10, 2003.
2000 *Mathematics Subject Classification.* Primary 65M12; Secondary 65M15, 65M60.
Key words and phrases. Finite volume schemes, linear advection.

order methods such as discontinuous Galerkin methods (which are a compromise between finite volumes and finite elements, see [19], [8]) for linear problems, we do mention that many mathematical difficulties still exist around all finite volume based numerical schemes on **arbitrary meshes**.

However, some of the issues about convergence of finite volume methods have already been resolved. Among past works to which this one can be related, let us mention the series of papers [31], [32], [33], where the notion of **supraconvergence** has been proposed to explain why formally non-consistent schemes are actually convergent. The same idea is used in [24] and [24] for vertex centered second order and high order schemes on various meshes. This idea (**the structure of the truncation error** has to be taken into account, and not only its magnitude) in conjunction with entropy inequalities has been used in [9], [10] to get a proof of convergence with an optimal rate $h^{\frac{1}{2}}$ of convergence in L^1 for scalar hyperbolic equations in dimension two. Application to scalar linear equations gives the same rate of convergence. Despite this recent and major progress, a general proof with optimal rate of convergence in various L^p spaces for linear finite volume discretization is not reachable by these techniques.

The main idea of this work is the following. If we analyze the truncation error of a finite volume scheme on an arbitrary triangular mesh in 2D, then we are forced to admit that finite volume methods are non-consistent in the finite difference sense: the truncation error is $O(1)$. This motivates the study of an abstract non-consistent model problem posed in a general Banach space: we wonder if it is possible to relax consistency in the Lax theorem, still having convergence of the model problem. **The answer is positive**, based on some simple and formally non-consistent residuals with a vanishing perturbation as the mesh size tends to zero. Destructive interactions in time explain why the error is negligible. Note that we give explicit expansions of the norm of the error, so there is a natural interpretation of our results. This interpretation is the following. Finite volume methods are indeed non-consistent methods: an $O(1)$ numerical error is made at each time step. But this $O(1)$ error is spread over the whole mesh after some time steps, so its norm tends to zero. In some cases, we can prove an $O(\frac{1}{(q+1)^s})$ bound where $s > 0$ and q is the number of time steps **after** the occurrence of the $O(1)$ error: $s = \frac{1}{2}$ in 1D, and $s = \frac{1}{4}$ or $s = \frac{1}{4} - \varepsilon$ in 2D. Finally, it is proved on various examples that the $O(\frac{1}{(q+1)^s})$ decrease of the error implies an explicit and standard $C(T)h^s$ rate of convergence.

This paper is organized as follows. In section 2, we present the model problem, which is an abstract evolution equation posed in a Banach space V. After recalling the Lax theorem in section 3, we study the possibility of getting non-consistent but convergent methods in section 4. In section 5 we prove that there exist non-consistent residuals with a vanishing perturbation when the mesh size tends to 0. Then we turn to applications. We prove that implicit and explicit schemes converge one to the other (section 6). Then, in section 7, we apply the formalism to all TVD schemes for the discretization of linear advection in 1D: we prove the convergence in $L^1(\mathbb{R})$ for all of them. Finally, in sections 8 to 12 we generalize the analysis to finite volume methods for the numerical solution of linear advection on a uniformly regular triangulation: using duality techniques and assuming enough regularity of the solution, we prove convergence in L^α, $2 \leq \alpha \leq +\infty$. Some of these estimates seem to be new: this is particularly the case for the $h^{\frac{1}{4}-\varepsilon}$ L^∞ error estimate on arbitrary regular triangulations. By comparison with [9], [10], one may infer that the $h^{\frac{1}{4}-\varepsilon}$ rate of convergence is probably still sub-optimal.

2. Model problem and notations

Let V be a Banach space equipped with the norm $||.||$ defined by ($L(V)$ is the space of linear continuous operators in V)

$$\forall A \in L(V), \quad ||A|| = \sup_{u \in V, u \neq 0} \frac{||Au||}{||u||}.$$

Let $u(t) \in V$ be the solution of the following abstract and general time evolution problem:

$$\begin{cases} \frac{\partial}{\partial t} u = Au, & t \geq 0, \\ u(0) = u_0 \in V. \end{cases} \tag{2.1}$$

A is a linear operator with a dense domain $D(A) \subset V$. Under very natural hypotheses, this problem is well posed [13], [3]. In the rest of this paper, we assume that the semigroup e^{tA} is bounded as a linear operator in V, namely

$$\exists T > 0 \text{ and } K > 0, \quad ||e^{tA}|| \leq K, \quad \forall t \in [0, T]. \tag{2.2}$$

A convenient abstract framework for the introduction of numerical methods for the numerical solution of problem (2.1) is the following. Let $h > 0$ be a parameter referred to as the mesh size. Studying the convergence of some numerical method when the mesh size tends to 0 consists in studying the limit case $h \to 0^+$.

Let $V_h \subset V$ be some vector subspace of V with finite dimension. V_h is a space of discrete functions. This vector subspace is indexed by the mesh size h. Let Π_h be some approximation operator, $\Pi_h : V \to V_h$. We assume that Π_h is a "good" approximation operator in the sense that

$$\forall u \in D(A), \quad \lim_{h \to 0} ||u - \Pi_h u|| = 0. \tag{2.3}$$

Let A_h be some numerical approximation of the operator A, $A_h : V_h \to V_h$. Let $\Delta t > 0$ be the time step.

Using these notations, a general explicit numerical approximation of (2.1), referred to as the time-explicit scheme (forward Euler), is

$$\begin{cases} \frac{u_h^{n+1} - u_h^n}{\Delta t} = A_h u_h^n, \quad n \geq 0, \\ u_h^0 = \Pi_h u_0. \end{cases} \tag{2.4}$$

In this work we will focus more on the **model problem**

$$\begin{cases} \frac{u_h^{n+1} - u_h^n}{\Delta t} = A_h u_h^n + s_h^n, \quad n \geq 0, \\ u_h^0 = \Pi_h u_0, \end{cases} \tag{2.5}$$

where $s_h^n \in V_h$, $s_h^n = O(1)$. s_h^n incorporates all extra terms due to some non-linear discretization of (2.1), or all the consistency defaults caused by the approximation of A by A_h.

System (2.5) appears naturally when studying the convergence of (2.4) by means of the numerical error e_h^n,

$$e_h^n = v_h^n - u_h^n, \quad \text{where } v_h^n = \Pi_h u(n\Delta t). \tag{2.6}$$

Since v_h^n is solution of

$$\left\{\begin{array}{l} \frac{v_h^{n+1}-v_h^n}{\Delta t} = A_h v_h^n + \left(\frac{v_h^{n+1}-v_h^n}{\Delta t} - A_h v_h^n\right), \quad n \geq 0, \\ v_h^0 = \Pi_h u_0, \end{array}\right. \tag{2.7}$$

then the numerical error e_h^n is the solution of

$$\left\{\begin{array}{l} \frac{e_h^{n+1}-e_h^n}{\Delta t} = A_h e_h^n + s_h^n, \quad n \geq 0, \\ e_h^0 = 0, \end{array}\right. \tag{2.8}$$

where $s_h^n = \frac{v_h^{n+1}-v_h^n}{\Delta t} - A_h v_h^n$. It is clear that (2.7) and (2.8) are particular cases of the model problem (2.5).

The proof of the key estimate will be given for the model problem with variable time steps $t_n \neq t_{n+1}$:

$$\left\{\begin{array}{l} \frac{u_h^{n+1}-u_h^n}{\Delta t_n} = A_h u_h^n + s_h^n, \quad n \geq 0, \\ u_h^0 = \Pi_h u_0, \end{array}\right. \tag{2.9}$$

3. Consistent approximations

Since Lax it is well known that stability and consistency are sufficient to insure the convergence of the numerical solution of the linear scheme (2.4) to the exact solution of (2.1) (see [28], [27], [13]). Let us recall these notions.

Definition 3.1 (Stability)**.** We assume that there exists a function $\tau : \mathbb{R}^+ \to \mathbb{R}^+$, called the maximal time step for a given mesh size h, such that

$$\forall h > 0, \ \forall \Delta t \in \,]0, \tau(h)], \ \forall n, \ \ 0 \leq n\Delta t \leq T,$$

we have

$$||(I + \Delta t A_h)^n|| \leq K, \tag{3.1}$$

where K is given in (2.2).

The inequality

$$\Delta t \leq \tau(h) \tag{3.2}$$

is called the CFL stability condition. In general (that is, when A_h is the numerical discretization of a partial differential operator A) the maximal time step is such that

$$\lim_{h\to 0} \tau(h) = 0. \tag{3.3}$$

A consequence of the CFL inequality (3.3) is then that the finer the mesh and the smaller the time step, the more work one has to do on the computer.

Since the scheme (2.4) is clearly of order one in time, we retain the following simplified definition of consistency.

Definition 3.2 (Consistency)**.** Let $u(t)$ be the solution of (2.1). Let us define the truncation error $r_h^n \in V_h$ as

$$r_h^n = \frac{\Pi_h u((n+1)\Delta t) - \Pi_h u(n\Delta t)}{\Delta t} - A_h \Pi_h u(n\Delta t), \ \ \forall n, h. \tag{3.4}$$

Considering the solution $u(t)$ of (2.1), we assume that there exists $s > 0$ such that

$$\exists C_1 > 0 \text{ and } C_2 > 0, \ \ ||r_h^n|| \leq C_1 h^s + C_2 \Delta t, \ \ \forall n, h. \tag{3.5}$$

The order of the scheme with respect to the space is s, while the order with respect to the time is only 1. Generally speaking, it is necessary to assume that the initial data u_0 is sufficiently regular in order for the consistency inequality to be true: for instance $u_0 \in D(A)$ (see [28], [27]).

Various extensions of this definition of consistency are possible. In particular, full high order schemes such that $||r_h^n|| \leq C_1 h^{s_1} + C_2 \Delta t^{s_2}$ with $s_2 > 1$ are highly desirable for practical numerical computations. However, for the sake of simplicity, we only consider in this paper the case $s_2 = 1$, compatible with (2.4).

Theorem 3.3 (Lax theorem)**.** *Assume that a linear scheme* (2.4) *is stable and consistent. Then, under the CFL condition* (3.3)*, it is convergent. See* [28].

Following (2.6), we define the projection v_h^n of the exact solution and the numerical error e_h^n. Then by definition of the truncation error, we get (2.7) and (2.8). The error is the solution of a non-homogeneous system with zero initial data. Due to the recurrence formula

$$e_h^{n+1} = (I + \Delta t A_h)\, e_h^n + \Delta t\, r_h^n,$$

we get the exact representation formula

$$e_h^n = (I + \Delta t A_h)^n\, e_h^0 + \Delta t \sum_{p=0}^{n-1} (I + \Delta t A_h)^{n-1-p}\, r_h^p, \tag{3.6}$$

that is,

$$e_h^n = \Delta t \sum_{p=0}^{n-1} (I + \Delta t A_h)^{n-1-p}\, r_h^p. \tag{3.7}$$

So

$$||e_h^n|| \leq \Delta t \sum_{p=0}^{n-1} ||\, (I + \Delta t A_h)^{n-1-p}\, ||\; ||r_h^p||;$$

that is, due to (3.1) and (3.5),

$$||e_h^n|| \leq K(n\Delta t)(C_1 h^s + C_2 \Delta t) \leq KT(C_1 h^s + C_2 \tau(h)).$$

Since $C_1 h^s + C_2 \tau(h) \to 0$ when $h \to 0$, it proves an estimate of convergence for the error and ends the proof. The estimate is uniform for $0 \leq n\Delta t \leq T$.

4. Non-consistent approximations

Many numerical methods for the numerical approximation of linear problems, however, do not satisfy the consistency requirement. Examples are known in finite volume methods for linear problems or non-linear numerical methods for linear problems. References may be found in [26], [20] and other papers cited therein. Examples are given in the last sections of this work.

The above fact motivates the study of the model problem

$$\left\{ \begin{array}{l} \frac{u_h^{n+1} - u_h^n}{\Delta t} = A_h u_h^n + s_h^n, n \geq 0, \\ u_h^0 = \Pi_h u_0, \end{array} \right. \tag{4.1}$$

where $u_0 \in V$ and

$$\|s_h^n\| = O(1) \text{ uniformly with respect to } h \to 0 \text{ and } n. \tag{4.2}$$

The solution of (4.1) satisfies

$$u_h^n - (I + \Delta t A_h)^n \, \Pi_h u_0 = \Delta t \sum_{p=0}^{n-1} (I + \Delta t A_h)^{n-1-p} \, s_h^p. \tag{4.3}$$

If we bound the right hand side of this equality using (4.2) and the stability inequality (3.1), we only obtain

$$\|u_h^n - (I + \Delta t A_h)^n \, \Pi_h u_0\| \le K(n\Delta t) \, O(1) \le (KT) \, O(1).$$

The introduction of the extra term s_h^n may induce some discrepancy between u_h^n and the solution of (2.4), that is, $(I + \Delta t A_h)^n \, \Pi_h u_0$. In this case the numerical solution of the scheme (4.1) may be very different from the numerical solution of the scheme (2.4), even for $h \to 0$.

We address the possibility of this extra term becoming negligible after some time steps due to some internal structure of these s_h^n, this internal structure being compatible with the iteration operator $I + \Delta A_h$. In other words, we investigate the possibility that

$$\forall p, \quad \lim_{n \to +\infty} (I + \Delta t A_h)^{n-1-p} \, s_h^p \to 0, \tag{4.4}$$

even if we only have (4.2) and (3.1) (still assuming the CFL inequality). A key issue seems to be obtaining uniform bounds such as

$$|| \, (I + \Delta t A_h)^q \, s_h^p || \le \varepsilon_q \quad \forall p, q, \quad \forall h, \tag{4.5}$$

where the sequence $(\varepsilon_q)_{q \in \mathbb{N}}$ decreases to 0:

$$\varepsilon_q \to 0 \text{ when } q \to +\infty. \tag{4.6}$$

Indeed, a consequence of (4.3)-(4.6) is

$$||u_h^n - (I + \Delta t A_h)^n \, \Pi_h u_0|| \le \Delta t \sum_{p=0}^{n-1} \varepsilon_{n-1-p} = \Delta t \sum_{p=0}^{n-1} \varepsilon_p. \tag{4.7}$$

Since the sum is bounded uniformly with respect to n, i.e.,

$$\Delta t \sum_{p=0}^{n-1} \varepsilon_p \le \Delta t \sum_{p=0}^{n-1} \max_q(\varepsilon_q) = (n\Delta t) \max_q(\varepsilon_q) = T \max_q(\varepsilon_q),$$

since and we have by hypothesis pointwise convergence to 0 (i.e., $\varepsilon_q \to 0$), then the Lebesgue theorem states that

$$\Delta t \sum_{p=0}^{n-1} \varepsilon_p \to 0 \text{ when } n \to 0.$$

In this case we really have $||u_h^n - (I + \Delta t A_h)^n \, \Pi_h u_0|| \to 0$ when $h \to 0$, and for all n, $0 \le n\Delta t \le T$.

In summary, an $O(1)$ perturbation s_h^n such that (4.5) and (4.6) are true has asymptotically a zero perturbation.

5. Convergence for the model problem

In order to apply the previous analysis, the key is to identify some general perturbations s_h^n such that (4.5)-(4.6) is true for the model problem (4.1).

Definition 5.1 (A class of admissible perturbations)**.** An additional term s_h^n such that

$$\exists \overline{C} > 0,\ \forall h, n,\ \exists z_h^n \in V_h \text{ with } ||z_h^n|| \leq \overline{C} \text{ and } s_h^n = (\tau(h)A_h)z_h^n \tag{5.1}$$

is called an admissible perturbation.

In this definition $\tau(h)$ is the maximal time step given by the CFL condition (3.2). Defining

$$T_h = I + \tau(h)A_h \tag{5.2}$$

and using the stability estimate (3.1), we know that there exists a constant $K > 0$ such that $||T_h|| \leq K$ for all h. Thus the operator $\tau(h)A_h$ is bounded:

$$||\tau(h)A_h|| \leq 1 + K, \ \ \forall h,$$

and we only have the bound

$$||s_h^n|| \leq ||\tau(h)A_h||\ ||z_h^n|| \leq (1+K)\overline{C},$$

compatible with (4.2).

We rewrite

$$I + \Delta t A_h = (1-\nu_h)I + \nu_h T_h,\ \ \nu_h = \frac{\Delta t}{\tau_h} \in \,]0,1] \text{ due to (3.1)},$$

and rewrite (4.3) plus (5.1) as

$$u_h^n - (I + \Delta t A_h)^n\, \Pi_h u_0 = \Delta t \sum_{p=0}^{n-1} \left((1-\nu_h)I + \nu_h T_h\right)^{n-1-p} (T_h - I)\, z_h^p. \tag{5.3}$$

Note that all powers of T_h are uniformly bounded due to the stability estimate (3.1):

$$\|(T_h)^n\| \leq K \ \ \forall\, h, n. \tag{5.4}$$

Now we make the remark which is at the base of this work.

In (5.3) $\left((1-\nu_h)I + \nu_h T_h\right)^{n-1-p}$ **is as a polynomial in T_h with non-negative coefficients. Multiplying this polynomial by $T_h - I$ results in cancellation of these coefficients. This cancellation is the key to obtaining** (4.5)-(4.6)**.**

In the next theorem, a bound is given for $B_h^p = \left((1-\nu_h)I + \nu_h T_h\right)^p (T_h - I)$. Note that we state the result in a slightly more general framework, such that the result is also true for non-constant time steps. Of course it also covers the model problem with constant time steps (4.1) or (5.3).

Theorem 5.2. *Let us assume a), b) and c).*

a) *We assume that the CFL inequality* (5.5) *is true* $\forall h > 0$, $\forall n$ *such that*

$$0 \leq \sum_{j=0}^{n-1} \Delta t_j \leq T, \Delta t_j \in \,]0, \tau(h)]. \tag{5.5}$$

b) *We assume that the stability estimate*

$$||\prod_{j=0}^{n-1}(I+\Delta t_j A_h)|| \le K \tag{5.6}$$

holds, where K is given in (2.2) *and* (3.1)*, as a consequence of the CFL inequality.*

c) *We assume that*

$$\exists(\nu^-,\nu^+) \text{ such that } 0<\nu^-\le\nu_{h,j}\le\nu^+<1 \quad \forall j,h, \tag{5.7}$$

where $\nu_{h,j}$ is defined by

$$\nu_{h,j}=\frac{\Delta t_j}{\tau(h)}\in\,]0,1]. \tag{5.8}$$

Let us consider B_h^p given by

$$B_h^p=\prod_{j=0}^{p-1}((1-\nu_{h,j})I+\nu_{h,j}T_h)(T_h-I), \quad B_h^0=T_h-I, \tag{5.9}$$

where T_h is given by (5.2)*. Then there exists $C>0$ such that $\forall h$, $\forall \nu_{h,j}$ satisfying* (5.7)*, $\forall p\ge 0$,*

$$||B_h^p|| \le \left(\max_{\pm}\frac{KC}{\sqrt{(1-\nu^\pm)\nu^\pm}}\right)\frac{1}{\sqrt{p+1}}. \tag{5.10}$$

Here C is a universal constant which does not depend on ν^-,ν^+,p,h.

We drop the index h in the proof since it does not play any role. We study

$$B^p=\prod_{j=0}^{p-1}((1-\nu_j)I+\nu_jT)(T-I), \tag{5.11}$$

where all powers of T are uniformly bounded ($||T^p||\le K$) and (ν_j) is a sequence such that $0<\nu^-\le\nu_j\le\nu^+<1$. Identifying all coefficients in the polynomial expansion

$$\prod_{j=0}^{p-1}((1-\nu_j)I+\nu_jT)=\sum_j \alpha_j^pT^j, \tag{5.12}$$

we get

$$\begin{cases} \alpha_0^0=1, \ \ \alpha_j^0=0 \text{ for } j\ne 0, \\ \alpha_j^{p+1}=(1-\nu_p)\alpha_j^p+\nu_p\alpha_{j-1}^p \ \ \forall j, \ \forall p\ge 0. \end{cases} \tag{5.13}$$

Next we split the rest of the proof in three lemmas.[1]

Lemma 5.3. *Consider* (5.13)*. Then*

$$\forall p\ge 0, \ \ \exists j_0(p)\in\{0,...,p\} \text{ with } \begin{cases} \alpha_j^p-\alpha_{j-1}^p\ge 0, & j\le j_0(p), \\ \alpha_j^p-\alpha_{j-1}^p\le 0, & j>j_0(p). \end{cases} \tag{5.14}$$

[1]Note that (5.13) is the finite difference upwind discretization of

$$\begin{cases} \partial_t\alpha+a\partial_x\alpha=0, \quad a>0, \\ \alpha(0,x)=\delta. \end{cases}$$

The proof of this lemma is by recurrence. Note that $\alpha_j^p \geq 0$ for all j, p.

a) p=0. We define $j_0(0) = 0$, so (5.14) is true.

b) We assume that (5.14) **is true for a given** $p \geq 0$**.** Since

$$\alpha_j^{p+1} - \alpha_{j-1}^{p+1} = (1-\nu_p)(\alpha_j^p - \alpha_{j-1}^p) + \nu_p(\alpha_{j-1}^p - \alpha_{j-2}^p),$$

we deduce that

$$\begin{cases} \forall j \leq j_0(p) & \alpha_j^{p+1} - \alpha_{j-1}^{p+1} \geq 0, \\ \forall j \geq j_0(p)+2 & \alpha_j^{p+1} - \alpha_{j-1}^{p+1} \leq 0. \end{cases}$$

It remains to look at $j = j_0(p)+1$:

$$\begin{cases} \text{If } \alpha_{j_0(p)+1}^{p+1} - \alpha_{j_0(p)}^{p+1} \geq 0, & \text{we define } j_0(p+1) = j_0(p)+1, \\ \text{if } \alpha_{j_0(p)+1}^{p+1} - \alpha_{j_0(p)}^{p+1} < 0, & \text{we define } j_0(p+1) = j_0(p). \end{cases}$$

With this definition of $j_0(p+1)$ property (5.14) is true for $p+1$. This finishes the proof of the lemma.

Lemma 5.4. *Considering B^p given by* (5.11), *one has*

$$||B^p|| \leq 2\alpha_{j_0(p)}^p K, \tag{5.15}$$

where K is the stability constant.

Since $B^p = \sum_j (\alpha_j^p - \alpha_{j-1}^p) T^j$, we get directly

$$||B^p|| \leq K \sum_j |\alpha_j^p - \alpha_{j-1}^p| = K \sum_{j \leq j_0(p)} (\alpha_j^p - \alpha_{j-1}^p) - K \sum_{j > j_0(p)} (\alpha_j^p - \alpha_{j-1}^p),$$

that is, $||B^p|| \leq 2\alpha_{j_0(p)}^p K$.

Lemma 5.5. *There exists $C > 0$ such that*

$$\alpha_{j_0(p)}^p \leq \left(C \max_{\pm} \frac{1}{\sqrt{(1-\nu^\pm)\nu^\pm}} \right) \frac{1}{(p+1)^{\frac{1}{2}}}. \tag{5.16}$$

Substituting the complex number $e^{i\theta}$ in (5.12), we obtain another definition of $\alpha_{j_0(p)}^p$, that is,

$$\alpha_{j_0(p)}^p = \frac{1}{2\pi} \int_0^{2\pi} f^p(\theta) e^{-ij_0(p)\theta} d\theta, \tag{5.17}$$

where the function f^p is defined by $f^p(\theta) = \prod_{j=0}^{p-1} \left((1-\nu_j) + \nu_j e^{i\theta} \right)$. Equality (5.17) means that $\alpha_{j_0(p)}^p$ is the $j_0(p)^{\text{th}}$ Fourier coefficient of f^p. We have

$$|\alpha_{j_0(p)}^p| \leq \frac{1}{2\pi} \int_0^{2\pi} \prod_{j=0}^{p-1} \left| (1-\nu_j) + \nu_j e^{i\theta} \right| d\theta.$$

Since

$$\left| (1-\nu_j) + \nu_j e^{i\theta} \right| = \sqrt{1 - 2\nu_j(1-\nu_j)(1-\cos\theta)} \leq \sqrt{1 - a\sin^2\frac{\theta}{2}},$$

where

$$a = 4\min\left(\nu^-(1-\nu^-), \nu^+(1-\nu^+) \right) \leq 1, \tag{5.18}$$

we get

$$|\alpha^p_{j_0(p)}| \le \frac{1}{2\pi}\int_0^{2\pi} (1-a\sin^2\frac{\theta}{2})^{\frac{p}{2}} d\theta. \tag{5.19}$$

The singularity in this integral is at $\theta = 0$. We isolate it:

$$\int_0^{2\pi} (1-a\sin^2\frac{\theta}{2})^{\frac{p}{2}} d\theta = 2\int_0^{\pi} (1-a\sin^2\frac{\theta}{2})^{\frac{p}{2}} d\theta \le 8\int_0^{\frac{\pi}{4}} (1-a\sin^2\theta)^{\frac{p}{2}} d\theta. \tag{5.20}$$

We use the change of variable $\varphi = p\sin^2\theta$, $\sin\theta = \sqrt{\frac{\varphi}{p}}$. Since $\theta \in [0, \frac{\pi}{4}]$, we have

$$d\theta = \frac{1}{2p\sin\theta\cos\theta} d\varphi \le \frac{1}{2\cos\frac{\pi}{4}} \times \frac{d\varphi}{\sqrt{p\varphi}}.$$

So we obtain

$$\int_0^{\frac{\pi}{4}} (1-a\sin^2\theta)^{\frac{p}{2}} d\theta = \frac{1}{2\cos\frac{\pi}{4}}\frac{1}{\sqrt{p}}\int_0^{\frac{p}{2}} (1-a\frac{\varphi}{p})^{\frac{p}{2}} \frac{d\varphi}{\sqrt{\varphi}}.$$

Finally we note that the function

$$\begin{cases} g_p(\varphi) = \frac{(1-a\frac{\varphi}{p})^{\frac{p}{2}}}{\sqrt{\varphi}} & \text{for } 0 \le \varphi \le \frac{p}{2}, \\ g_p(\varphi) = 0 & \text{in other cases,} \end{cases}$$

is uniformly bounded for all p and $\varphi \ge 0$ by an integrable function $g_p(\varphi) \le g(\varphi)$, where g is defined by

$$g(\varphi) = \frac{e^{-2a\varphi}}{\sqrt{\varphi}} \quad \varphi \ge 0.$$

So

$$\int_0^{\frac{p}{2}} g_p(\varphi)d\varphi = \int_0^{+\infty} g_p(\varphi)d\varphi \le \int_0^{+\infty} g(\varphi)d\varphi = \frac{1}{\sqrt{a}}\int_0^{+\infty} \frac{e^{-2\varphi}}{\sqrt{\varphi}}d\varphi$$

The interesting consequence is that

$$\exists H > 0,\ \forall p, \quad \int_0^{+\infty} g_p(\varphi)d\varphi \le \frac{H}{\sqrt{a}}.$$

So we get that $\exists I > 0$ such that $|\alpha^p_{j_0(p)}| \le \frac{I}{\sqrt{ap}}$, $p \ge 1$, and

$$\exists J > 0 \text{ such that } |\alpha^p_{j_0(p)}| \le \frac{J}{\sqrt{a(p+1)}}, \quad \forall p \ge 0.$$

In conjunction with (5.18), this finishes the proof of the lemma. Finally, Theorem 5.2 is a consequence of (5.16) and (5.10).

Corollary 5.6. *Let us consider an admissible perturbation s_h^n of the form* (5.1)*. We assume that all hypotheses of Theorem 5.2 are true.*

Then we have the property (4.5)*: there exists $C > 0$ such that $\forall p, r, h$*

$$\| \prod_{j=0}^{p-1} ((1-\nu_{h,j})I + \nu_{h,j}T_h)s_h^r \| \le \left(\max_{\pm} \frac{KC}{\sqrt{(1-\nu^{\pm})\nu^{\pm}}} \max_{h,n} ||z_h^n|| \right) \frac{1}{\sqrt{p+1}}. \tag{5.21}$$

Moreover, $\exists \tilde{C} > 0$ such that the solution of the model problem with variable time steps Δt_n,

$$\begin{cases} \frac{u_h^{n+1}-u_h^n}{\Delta t_n} = A_h u_h^n + s_h^n, \quad n \ge 0, \\ u_h^0 = \Pi_h u_0, \end{cases} \tag{5.22}$$

satisfies

$$\begin{aligned}(5.23)\qquad &||u_h^n - \prod_{j=0}^{n-1}((1-\nu_{h,j})I + \nu_{h,j}T_h)\Pi_h u_0|| \\ &\le \left(\max_{\pm} \frac{K\tilde{C}}{\sqrt{(1-\nu^{\pm})\nu^{\pm}}} \max_{h,n} ||z_h^n||\right)\sqrt{T \max_j(\Delta t_j)}.\end{aligned}$$

Here C and $\tilde{C}$ are two universal constants which do not depend on p, n, h, ν^- and ν^+. These estimates are true for $0 \le n\Delta t \le T$.

Inequality (5.21) is a direct consequence of Theorem 5.2.

We detail the rest of the proof only for $\nu^- = \nu^+ = \nu$; the other case is straightforward. Using (4.3) and (5.21), we obtain

$$||u_h^n - (I + \Delta t A_h)^n \Pi_h u_0|| \le \Delta t \sum_{p=0}^{n-1} \frac{KC}{\sqrt{p+1}} \frac{1}{\sqrt{\nu(1-\nu)}} \max_{h,n} ||z_h^n||.$$

Since

$$\sum_{p=0}^{n-1} \frac{1}{\sqrt{p+1}} = 1 + \sum_{p=1}^{n-1} \frac{1}{\sqrt{p+1}} \le 1 + \int_1^n \frac{dx}{\sqrt{x}} = 1 + \frac{\sqrt{n}-1}{2},$$

this proves that

$$\exists F > 0 \text{ such that } \sum_{p=0}^{n-1} \frac{1}{\sqrt{p+1}} \le F\sqrt{n} \ \ \forall n \ge 1.$$

Then

$$\begin{aligned}||u_h^n - (I + \Delta t A_h)^n \Pi_h u_0|| &\le (K\tilde{C} \max_{h,n} ||z_h^n||) \frac{1}{\sqrt{\nu(1-\nu)}} \Delta t \sqrt{n} \\ &\le (K\tilde{C} \max_{h,n} ||z_h^n||) \frac{1}{\sqrt{\nu(1-\nu)}} \Delta t \sqrt{\frac{T}{\Delta t}} = \left(\frac{K\tilde{C}}{\sqrt{\nu(1-\nu)}} \max_{h,n} ||z_h^n||\right)\sqrt{T\Delta t}\end{aligned}$$

for a suitable $\tilde{C} = CF$. This finishes the proof.

Always for $\nu^- = \nu = \nu^+$, a consequence of (5.23) is

$$||u_h^n - (I + \Delta t A_h)^n \Pi_h u_0|| \le \left(\frac{K\tilde{C}}{\sqrt{1-\nu}} \max_{h,n} ||z_h^n||\right)\sqrt{T\tau(h)}.$$

Since the right hand side of this inequality is uniformly bounded for $\Delta t \to 0$, it gives a similar estimate for the continuous-in-time scheme. This means that the lower bound $\nu^- \le \frac{\Delta t}{\tau(h)}$ is not restrictive.

6. Implicit scheme

The next two sections are devoted to showing that the framework developed above is non-empty, and that new results are obtained.

In this section we deal with implicit schemes and prove that, under reasonable assumptions, the difference between the solution of the explicit scheme and the solution of the implicit scheme tends to 0, when $h \to 0$. Let us consider a solution u_h^n of the first order[2] implicit scheme

$$\left\{ \begin{array}{l} \frac{u_h^{n+1}-u_h^n}{\Delta t} = A_h u_h^{n+1}, \quad n \geq 0, \\ u_h^0 = \Pi_h u_0. \end{array} \right. \tag{6.2}$$

The current iteration is

$$(I - \Delta t A_h)\, u_h^{n+1} = u_h^n. \tag{6.3}$$

We assume that (6.2) is unconditionally stable, in the sense that $\forall \Delta t > 0$, the matrix $I - \Delta t A_h$ is non-singular and

$$\forall \Delta t > 0, \ \ \forall h, \ \ \forall p \geq 0, \ \ \| (I - \Delta t A_h)^{-p} \| \leq K, \tag{6.4}$$

where K is the stability constant (2.2), (3.1). We rewrite (6.3) as

$$\frac{u_h^{n+1} - u_h^n}{\Delta t} = A_h u_h^n + (\tau(h) A_h)\, \frac{u_h^{n+1} - u_h^n}{\tau(h)}. \tag{6.5}$$

This means that the implicit scheme may be considered as an explicit scheme plus a perturbation which is admissible (in the sense of Definition 5.1) provided we prove that $\frac{u_h^{n+1}-u_h^n}{\tau(h)}$ is uniformly bounded. Nevertheless, implicit schemes are used mostly with large time steps like $\Delta t > \tau(h)$, so we cannot apply Corollary 5.6 directly to (6.5).

Let us consider a smaller time step

$$\overline{\Delta t} = \frac{\Delta t}{d}, \ \ d \in \mathbb{N}^*, \ \ \text{with } \frac{\overline{\Delta t}}{\tau(h)} \in \,]0, 1[, \tag{6.6}$$

and the linear interpolant $\overline{u}_h^{n,k}$, $k = 0, 1, ..., d$,

$$\overline{u}_h^{n,k} = u_h^n + \frac{k}{d}(u_h^{n+1} - u_h^n), \ \ \overline{u}_h^{n+1,0} = \overline{u}_h^{n,d}.$$

For the sake of simplicity we assume that d is a constant. In other words, d does not depend on the mesh size h.

For $0 \leq k \leq d-1$, this linear interpolant is a solution of

$$\frac{\overline{u}_h^{n,k+1} - \overline{u}_h^{n,k}}{\overline{\Delta t}} = \frac{u_h^{n+1} - u_h^n}{d\overline{\Delta t}} = A_h u_h^{n+1} = A_h \overline{u}_h^{n,k} + s_h^{n,k} \tag{6.7}$$

with

$$s_h^{n,k} = A_h(u_h^{n+1} - \overline{u}_h^{n,k}) = (\tau(h) A_h) \left((1 - \frac{k}{d}) \frac{u_h^{n+1} - u_h^n}{\tau(h)} \right).$$

It remains to obtain some bounds for $\frac{u_h^{n+1}-u_h^n}{\tau(h)}$. We define

$$v_h^n = \frac{u_h^{n+1} - u_h^n}{\Delta t}$$

[2]Generalization of the discussion to the Crank-Nicholson second order implicit scheme (6.1) is straightforward:

$$\left\{ \begin{array}{l} \frac{u_h^{n+1}-u_h^n}{\Delta t} = A_h \frac{u_h^{n+1}+u_h^n}{2}, \quad n \geq 0, \\ u_h^0 = \Pi_h u_0. \end{array} \right. \tag{6.1}$$

and note that v_h^n is the solution of the implicit scheme

$$\left\{\begin{array}{l}\frac{v_h^{n+1}-v_h^n}{\Delta t}=A_h v_h^{n+1}, \quad n\geq 0,\\ \frac{v_h^1-A_h\Pi_h u_0}{\Delta t}=A_h v_h^1.\end{array}\right. \tag{6.8}$$

A consequence of the stability estimate for the implicit scheme is

$$||\frac{u_h^{n+1}-u_h^n}{\Delta t}||=||v_h^n||\leq K||A_h\Pi_h u_0||, \quad \forall n\geq 0.$$

and

$$||\frac{u_h^{n+1}-u_h^n}{\tau(h)}||\leq \frac{\Delta t}{\overline{\overline{\Delta t}}}\frac{\overline{\Delta t}}{\tau(h)}K||A_h\Pi_h u_0||\leq dK||A_h\Pi_h u_0||. \tag{6.9}$$

Under reasonable assumptions, this last term $||A_h\Pi_h u_0||$ is uniformly bounded for many initial data $u_h^0=\Pi_h u_0$, when $u_0\in D(A)$, [13], [27].

So (6.7) enters in the formalism covered by Corollary 5.6. As a consequence we get

Theorem 6.1. *Under all the above hypotheses* (6.4), (6.6) *about the implicit scheme* (6.2), *there exists a constant* $C=C(d,\frac{\overline{\Delta t}}{\tau(h)})>0$ *such that*

$$\|u_h^n-(I+\overline{\Delta t}A_h)^{(dn)}\Pi_h u_0\|\leq (KC||A_h\Pi_h u_0||)\sqrt{T\Delta t}, \quad 0\leq n\Delta t\leq T. \tag{6.10}$$

This result means that the solution of the implicit scheme is asymptotically equal to the solution of the explicit scheme with a smaller time step. Under the hypotheses of the theorem, proving convergence of the explicit scheme is equivalent to proving convergence of the implicit scheme.

7. TVD schemes in 1D

Let us now consider linear advection in 1D,

$$\left\{\begin{array}{ll}\partial_t u=-a\partial_x u, & a>0,\\ u(0,x)=u_0(x).\end{array}\right. \tag{7.1}$$

The solution is

$$u(t,x)=u_0(x-at).$$

We would like to discuss the convergence of TVD schemes for the numerical solution of this problem. The space is $V=L^1(\mathbb{R})$; we assume that $u_0\in L^1(\mathbb{R})\cap BV(\mathbb{R})$.

Let $h>0$ be the mesh size and $\Delta t>0$ the time step. The general form of such a TVD scheme is

$$\frac{u_j^{n+1}-u_j^n}{\Delta t}+a\frac{u_{j+\frac12}^n-u_{j-\frac12}^n}{h}=0, \tag{7.2}$$

with the initial data given by

$$u_j^0=\frac{1}{h}\int_{jh}^{(j+1)h}u_0(x)dx. \tag{7.3}$$

We refer to [26], [20] and the references given therein for an introduction to TVD schemes. It is well known in the theory of TVD schemes that interesting values for the numerical flux are (we give the common name of the limiter and drop the superscript n)

$$\begin{cases} \textbf{upwind}, & u_{j+\frac12} = u_j, \\ \textbf{minmod}, & u_{j+\frac12} = u_j + \frac12(1-\nu)\mathrm{minmod}(1, r_{j+\frac12})(u_{j+1}-u_j), \\ \textbf{superbee}, & u_{j+\frac12} = u_j \\ & +\frac12(1-\nu)\max\left(0, \min(1, 2r_{j+\frac12}), \min(2, r_{j+\frac12})\right)(u_{j+1}-u_j), \\ \textbf{ultrabee}, & u_{j+\frac12} = u_j + (1-\nu)\mathrm{minmod}(\frac{1}{1-\nu}, \frac{r_{j+\frac12}}{\nu})(u_{j+1}-u_j). \end{cases} \tag{7.4}$$

Here $r_{j+\frac12} = \frac{u_j - u_{j-1}}{u_{j+1}-u_j}$ and $\nu = a\frac{\Delta t}{h}$. The minmod function is defined by

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

A common property of all these schemes is the TVD property [20], [26], which we recall now.

Lemma 7.1. *Assume the CFL inequality*

$$a\frac{\Delta t}{h} \le 1,$$

that is, $\tau(h) = \frac{h}{a}$*. Then:*

a) *The linear upwind scheme (i.e.,* (7.2) *with the upwind flux) is* L^1 *stable:*

$$h\sum_j |u_j^{n+1}| \le h \sum_j |u_j^n| \quad (i.e.,\ K = 1). \tag{7.5}$$

b) *The solution of the scheme* (7.2) *with a non-linear TVD flux (examples are given in* (7.4)*) satisfies the TVD inequality*

$$\sum_j |u_j^{n+1} - u_{j-1}^{n+1}| \le \sum_j |u_j^n - u_{j-1}^n|. \tag{7.6}$$

Many other limiters are used in the literature. We note that all of these numerical fluxes are defined as the upwind scheme plus a correction. This correction is non-linear, and corresponds to a limited evaluation of $\frac12(u_{j+1}-u_j)$ for second order fluxes; this is the case for the minmod and superbee limiter formulas, and also for the van Leer limiter formula [20], [26]. The correction is a limited evaluation of $(u_{j+1}-u_j)$ for the ultrabee scheme, which is only first order (see [14] for a discussion of the optimality of the ultrabee limiter for the capture of discontinuous profiles).

We only mention that the numerical solution of all these schemes converges in $L^1([0,T]\times\mathbb{R})$ to the exact solution, based on the TVD property. Concerning convergence and rate of convergence in $L^1(\mathbb{R})$, simple proofs are available for monotone schemes—unfortunately only the upwind scheme is monotone. Convergence and rate of convergence of the minmod, superbee and van Leer limiters are reached using the theory of [17], [4]; we remark that these proves are very technical and never use the linearity of the equation.

As for us, we note that all these TVD schemes may be rewritten as

$$\frac{u_j^{n+1}-u_j^n}{\Delta t} = -a\frac{u_j^n - u_{j-1}^n}{h} - a\frac{s_j^n - s_{j-1}^n}{h}, \tag{7.7}$$

where s_j^n is the non-linear correction

$$s_j^n = u_{j+\frac{1}{2}}^n - u_j^n.$$

Lemma 7.2. *All TVD schemes defined in* (7.4) *satisfy*

$$\sum_j |s_j^n| \le 2(1 - a\frac{\Delta t}{h}) \sum_j |u_j^n - u_{j-1}^n|. \tag{7.8}$$

Due to the definition (7.4), we note that the upwind, minmod and superbee limiters satisfy

$$|s_j^n| \le (1-\nu)|u_{j+1}^n - u_j^n|, \tag{7.9}$$

which proves (7.8). A little algebra is needed for the ultrabee limiter, rewritten as

$$(s_j^n)^{\mathbf{ub}} = \mathrm{minmod}\left(u_{j+1} - u_j, (\frac{1}{\nu} - 1)(u_j - u_{j-1}) \right).$$

Let us assume that $\nu \ge \frac{1}{2}$. Then

$$|(s_j^n)^{\mathbf{ub}}| \le (\frac{1}{\nu} - 1)|u_j - u_{j-1}| \le 2(1-\nu)|u_j - u_{j-1}|.$$

On the contrary if we assume that $\nu < \frac{1}{2}$, then we use

$$|(s_j^n)^{\mathbf{ub}}| \le |(u_{j+1} - u_j| \le 2(1-\nu)|(u_{j+1} - u_j|.$$

In both cases we get (7.8). This ends the proof of the lemma.

Let us define

$$V_h = \{u \in L^1(\mathbb{R})\,;\ u(x) \text{ is constant for } x \in]jh, (j+1)h[\},$$

and

$$u_h^n(x) = u_j^n \quad \forall x \in]jh, (j+1)h[.$$

One has

$$\left\{ \begin{array}{ll} u_h^n = (u_j^n) \in V_h, & \\ (A_h u)_j = -a\frac{u_j - u_{j-1}}{h}, & \forall u \in V_h, \\ z_h^n = (z_{h,j}^n) = (a\frac{s_j^n}{h}) \in V_h. & \end{array} \right. \tag{7.10}$$

Due to (7.8) we note that

$$\|z_h^n\|_{L^1(\mathbb{R})} = h \sum_j |z_{h,j}^n| \le 2a(1 - a\frac{\Delta t}{h}) \sum_j |u_j^0 - u_{j-1}^0|. \tag{7.11}$$

Now we assume that $u_0 \in L^1(\mathbb{R}) \cap BV(\mathbb{R})$. In this case it is well known that the total variation of the discrete profile u^0 is bounded by the total variation of u_0:

$$\sum_j |u_j^0 - u_{j-1}^0| \le TV(u_0). \tag{7.12}$$

For a differentiable u_0, one has

$$TV(u_0) = \int_{\mathbb{R}} |\partial_x u_0(x)| dx.$$

If u_0 is not differentiable, the total variation is defined as

$$TV(u_0) = \max_{\varphi \in C^1(\mathbb{R}), |\varphi(x)| \le 1} \int_{\mathbb{R}} (-\varphi'(x) u_0(x)) dx.$$

What is important here is that z_h^n is bounded in $L^1(\mathbb{R})$ independently of the mesh size h,

$$\|z_h^n\|_{L^1(\mathbb{R})} \le 2a(1-a\frac{\Delta t}{h})TV(u_0). \tag{7.13}$$

So (7.7) may be rewritten as

$$\begin{cases} \frac{u_h^{n+1}-u_h^n}{\Delta t} = A_h u_h^n + (\tau(h)A_h)z_h^n, \quad n \ge 0, \\ u_h^0 = \Pi_h u_0, \end{cases} \tag{7.14}$$

where Π_h is the constant mass approximation (7.3).

We obtain a first result

Lemma 7.3. *Assume the CFL condition*

$$a\frac{\Delta t}{h} < 1, \tag{7.15}$$

and $u_0 \in L^1(\mathbb{R}) \cap BV(\mathbb{R})$. Then $\exists C > 0$ such that for all TVD schemes considered above we have

$$||u_h^n - (I+\Delta t A_h)^n \Pi_h u_0|| \le (C\,TV(u_0))\sqrt{a(1-a\frac{\Delta t}{h})Th}. \tag{7.16}$$

This is true $\forall h$ and $\forall n$, $0 \le n\Delta t \le T$.

The proof is a direct consequence of Corollary 5.6 in the case $\nu^- = \nu^+ = \frac{\Delta t}{\tau(h)}$. The convergence of the upwind scheme may also be proved using our formalism. Let us define the projection of the exact solution

$$v_h^n = \Pi_h u(n\Delta t, .) = \left(\frac{1}{h}\int_{jh}^{(j+1)h} u(n\Delta t, x)dx\right). \tag{7.17}$$

This vector is the solution of

$$\begin{cases} \frac{v_h^{n+1}-v_h^n}{\Delta t} = A_h v_h^n + \tilde{Q}_h^n, \quad n \ge 0, \\ v_h^0 = \Pi_h u_0, \end{cases} \tag{7.18}$$

where

$$\begin{aligned} (\tilde{Q}_h^n)_j = &\frac{1}{\Delta t h}\int_{jh}^{(j+1)h} \left(u((n+1)\Delta t, x) - u(n\Delta t, x)\right)dx \\ &+ \frac{a}{h^2}\left(\int_{jh}^{(j+1)h} u(n\Delta t, x)dx - \int_{(j-1)h}^{jh} u(n\Delta t, x)dx\right). \end{aligned} \tag{7.19}$$

Since

$$\begin{aligned} \int_{jh}^{(j+1)h} \left(u((n+1)\Delta t, x) - u(n\Delta t, x)\right)dx &= \int_{n\Delta t}^{(n+1)\Delta t}\int_{jh}^{(j+1)h} \partial_t u(s,x)dsdx \\ &= -a\int_{n\Delta t}^{(n+1)\Delta t}\int_{jh}^{(j+1)h} \partial_x u(s,x)dsdx \\ &= -a\int_{n\Delta t}^{(n+1)\Delta t} \left(u(s,(j+1)h) - u(s,jh)\right)ds, \end{aligned}$$

we have that

$$(\tilde{Q}_h^n)_j = \frac{a}{h}\left(\left[\frac{1}{h}\int_{jh}^{(j+1)h} u(n\Delta t, x)dx - \frac{1}{\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} u(s,(j+1)h)ds\right] - \left[\frac{1}{h}\int_{(j-1)h}^{jh} u(n\Delta t, x)dx - \frac{1}{\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} u(s,jh)ds\right]\right),$$

that is, $\tilde{Q}_h^n = (\tau(h)A_h)\tilde{z}_h^n$ with

$$(\tilde{z}_h^n)_j = \frac{1}{\tau(h)}\left[\frac{1}{h}\int_{jh}^{(j+1)h} u(n\Delta t, x)dx - \frac{1}{\Delta t}\int_{n\Delta t}^{(n+1)\Delta t} u(s,(j+1)h)ds\right]. \tag{7.20}$$

Note that

$$\int_{n\Delta t}^{(n+1)\Delta t} u(s,(j+1)h)ds = \int_{n\Delta t}^{(n+1)\Delta t} u(n\Delta t,(j+1)h - a(s-n\Delta t))ds = \int_{(j+1)h-a\Delta t}^{(j+1)h} u(n\Delta t, x)\frac{dx}{a},$$

where we have used the change of variable $a(s - n\Delta t) = (j + 1)h - x$. So

$$\begin{aligned}(\tilde{z}_h^n)_j &= \frac{a}{h}\left[\frac{1}{h}\int_{jh}^{(j+1)h} u(n\Delta t, x)dx - \frac{1}{a\Delta t}\int_{(j+1)h-a\Delta t}^{(j+1)h} u(n\Delta t, x)dx\right]\\
&= \frac{a}{h}(1-a\frac{\Delta t}{h})\left[\frac{1}{h-a\Delta t}\int_{jh}^{(j+1)h-a\Delta t} u(n\Delta t, x)dx - \frac{1}{a\Delta t}\int_{(j+1)h-a\Delta t}^{(j+1)h} u(n\Delta t, x)dx\right]\\
&= \frac{a}{h}(1-a\frac{\Delta t}{h})\left[\frac{1}{h-a\Delta t}\int_{jh}^{(j+1)h-a\Delta t} (u(n\Delta t, x)-u(n\Delta t,jh))dx - \frac{1}{a\Delta t}\int_{(j+1)h-a\Delta t}^{(j+1)h} (u(n\Delta t, x)-u(n\Delta t,jh))dx\right].\end{aligned}$$

Since

$$|u(n\Delta t, x) - u(n\Delta t, jh)| \le \int_{jh}^{(j+1)h} |\partial_x u(n\Delta t, x)|dx$$

we obtain

$$|(\tilde{z}_h^n)_j| \le \frac{a}{h}(1-a\frac{\Delta t}{h})\left(2\int_{jh}^{(j+1)h} |\partial_x u(n\Delta t, x)|dx\right).$$

Note that $\tilde{z}_h^n \in V_h$ is uniformly bounded in L^1:

$$||\tilde{z}_h^n|| = h\sum_j |(\tilde{z}_h^n)_j| \le 2a(1-a\frac{\Delta t}{h})TV(u) = 2a(1-a\frac{\Delta t}{h})TV(u_0). \tag{7.21}$$

Next we define the error $e_h^n = v_h^n - u_h^n$. This error is the solution of

$$\begin{cases} \frac{e^{n+1}-e^n}{\Delta t} = A_h e_h^n + (\tau(h)A_h)\tilde{z}_h^n, \quad n \ge 0,\\ e_h^0 = 0.\end{cases} \tag{7.22}$$

So we apply Corollary 5.6 to (7.22) and obtain

Theorem 7.4. *Assume the CFL condition*

$$a\frac{\Delta t}{h} < 1, \tag{7.23}$$

and $u_0 \in L^1(\mathbb{R}) \cap BV(\mathbb{R})$. *Then* $\exists C > 0$ *such that for all TVD schemes considered above we have*

$$||u_h^n - \Pi_h u(n\Delta t)|| \leq (C\, TV(u_0)) \sqrt{a(1-a\frac{\Delta t}{h})Th}. \tag{7.24}$$

This is true $\forall h$ *and* $\forall n$, $0 \leq n\Delta t \leq T$.

Since (7.24) is true for the upwind scheme due to Corollary 5.6 we see that, and since we have Lemma 7.3, it follows that (7.24) is true for all TVD schemes. Note that (7.24) is accurate, in the sense that asymptotic regimes are correctly described. If $a \to 0^+$ (no advection), we indeed find a zero error. If $1 - a\frac{\Delta t}{h} \to 0^+$, we find a zero error compatible with the fact that the scheme is exact for $1 - a\frac{\Delta t}{h} = 0$.

8. Linear advection by finite volume method

The rest of this paper is devoted to the study of linear advection in 2D:

$$\begin{cases} \partial_t u + \vec{a}.\vec{\nabla} u = 0, & (t,x) \in [0,T] \times \Omega, \\ u(t=0,x) = u_0(x), & x \in \Omega = [0,1] \times [0,1]. \end{cases} \tag{8.1}$$

For the sake of simplicity we assume that $\vec{a} \in \mathbb{R}^2$, $\vec{a} \neq 0$, is constant, and supplement (8.1) with periodic boundary conditions

$$\begin{cases} u(t,0,x_2) = u(t,1,x_2), & (t,x_2) \in [0,T] \times [0,1], \\ u(t,x_1,0) = u(t,x_1,1), & (t,x_1) \in [0,T] \times [0,1]. \end{cases} \tag{8.2}$$

Let $(\Omega_j)_{j \in J}$ be a finite triangular mesh of Ω:

$$\begin{cases} \Omega_j \cap \Omega_k = \emptyset, & \forall j,k, \quad j \neq k, \\ \bigcup_{j \in J} \overline{\Omega_j} = \overline{\Omega} = \Omega. \end{cases} \tag{8.3}$$

Two cells are neighboring cells if and only if they have an edge in common (taking periodic boundary conditions into account). Each cell has 3 neighbors: $I(j)$ is the set of indices of the neighbors of the cell j. The outgoing normal from Ω_j on the edge $\overline{\Omega_j} \cap \overline{\Omega_k}$ is denoted as $\vec{n}_{jk}$. Of course the outgoing normal from Ω_j is the opposite of the outgoing normal from Ω_k for $k \in I(j)$,

$$\vec{n}_{jk} + \vec{n}_{kj} = 0. \tag{8.4}$$

We introduce some very natural notation:

$$\begin{cases} l_{jk} = l_{kj} = \mathbb{R}\text{-Lebesgue measure of } \overline{\Omega_j} \cap \overline{\Omega_k}, & \text{a length}, \\ s_j = \mathbb{R}^2\text{-Lebesgue measure of } \Omega_j, & \text{a surface}. \end{cases} \tag{8.5}$$

We also define

$$\begin{cases} I^+(j) = \{k \in I(j); (\vec{a}, \vec{n}_{jk}) > 0\}, \\ I^0(j) = \{k \in I(j); (\vec{a}, \vec{n}_{jk}) = 0\}, \\ I^-(j) = \{k \in I(j); (\vec{a}, \vec{n}_{jk}) < 0\}. \end{cases} \tag{8.6}$$

and

$$m_{jk} = m_{kj} = l_{jk}|(\vec{a}, \vec{n}_{jk})|. \tag{8.7}$$

Here $(.,.)$ denotes the standard scalar product. $I^+(j)$ (resp. $I^-(j)$) is the set of outgoing (resp. incoming) cells from Ω_j. An example is given in Figure 1. With all

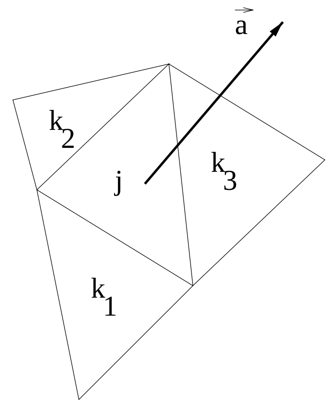

FIGURE 1. $I^+(j) = \{k_2, k_3\}$, $I^-(j) = \{k_1\}$

these notations the finite volume upwind scheme is defined as

$$s_j \frac{u_j^{n+1} - u_j^n}{\Delta t} + \sum_{k \in I^+(j)} m_{jk} u_j^n - \sum_{k \in I^-(j)} m_{jk} u_k^n = 0, \quad \forall j \in J, \tag{8.8}$$

with the constant mass initial approximation

$$u_j^0 = \frac{1}{s_j} \int_{\Omega_j} u(0, x) dx. \tag{8.9}$$

$m_{jk} u_j^n$ is the flux value integrated along the edge $\overline{\Omega_j} \cap \overline{\Omega_k}$, $k \in I^+(j)$.

The following formula will play an important role in the analysis. The proof is left to the reader.

Lemma 8.1. *One has the equality*

$$\sum_{k \in I^+(j)} m_{jk} = \sum_{k \in I^-(j)} m_{jk}, \ \ \forall j. \tag{8.10}$$

Using this formula, we rewrite (8.8) as

$$u_j^{n+1} = \left(1 - \frac{\Delta t}{s_j} \sum_{k \in I^-(j)} m_{jk} \right) u_j^n + \frac{\Delta t}{s_j} \sum_{k \in I^-(j)} m_{jk} u_k^n.$$

Provided the CFL condition is satisfied, i.e.,

$$\frac{\Delta t}{s_j} \sum_{k \in I^-(j)} m_{jk} \leq 1, \tag{8.11}$$

u_j^{n+1} is a convex combination of (u_j^n). As a consequence we get

Lemma 8.2. *Assume* (8.11). *Consider the solution of* (8.8). *Then*

$$\left(\sum_j s_j |u_j^{n+1}|^\alpha \right)^{\frac{1}{\alpha}} \leq \left(\sum_j s_j |u_j^n|^\alpha \right)^{\frac{1}{\alpha}}, \ \forall \alpha, \ 1 \leq \alpha < +\infty, \tag{8.12}$$

and

$$\max_j |u_j^{n+1}| \le \max_j |u_j^n|. \tag{8.13}$$

Now we reformulate (8.8)-(8.9) using the abstract formalism. Let h be a characteristic length of the triangular mesh, which is assumed to be uniformly regular, that is,

$$\exists c_0, c_1 > 0, \quad c_0 h \le l_{jk} \le c_1 h, \ \ \forall j, k, \tag{8.14}$$

or equivalently

$$\exists c_2, c_3 > 0, \quad c_2 h^2 \le s_j \le c_3 h^2, \ \ \forall j. \tag{8.15}$$

It implies (a proof is given in [15])

$$\exists c_4, c_5 > 0, \quad c_4 h \le \sum_{k \in I^-(j)} m_{jk} \le c_5 h, \ \ \forall j, \ \forall h.$$

As a consequence we get

Lemma 8.3. *The maximal time step*

$$\tau(h) = \max_j \frac{s_j}{\sum_{k \in I^-(j)} m_{jk}}$$

is such that

$$\exists C_1 > 0, C_2 > 0, \quad C_1 h \le \tau(h) \le C_2 h. \tag{8.16}$$

In order to simplify the discussion, we assume for the rest of this paper that the CFL condition is bounded away from 0 and 1: The reason is that we desire to use the estimate (5.21) of Corollary 5.6, which is singular at $\nu = 0$ and $\nu = 1$. So we assume that there exist ν^- and ν^+ such that

$$0 < \nu^- \le \frac{\Delta t}{\tau(h)} \le \nu^+ < 1, \ \ \forall h. \tag{8.17}$$

This assumption is not a real restriction since it is in accordance with the practical use which is made of such schemes. Let

$$V^\alpha = L^\alpha(\Omega),$$

and

$$V_h^\alpha = \{u \in V^\alpha; \ u \text{ is constant in } \Omega_j \ \forall j, \text{ that is, } u_j = u_{|\Omega_j}\} \subset V^\alpha.$$

The norm in V^α is

$$||u||_\alpha = \left(\int_\Omega |u|^\alpha dx \right)^{\frac{1}{\alpha}} \ \ \forall u \in V^\alpha,$$

and

$$||u||_\alpha = \left(\sum_j s_j |u_j|^\alpha \right)^{\frac{1}{\alpha}} \ \ \forall u \in V_h^\alpha.$$

Let

$$\begin{cases} \Pi_h : V^\alpha \to V_h^\alpha, \\ (\Pi_h u)_j = \frac{1}{s_j} \int_{\Omega_j} u(x) dx, \end{cases}$$

and

$$\begin{cases} A_h : V_h^\alpha \to V_h^\alpha, \\ (A_h u)_j = \frac{-\sum_{k \in I^+(j)} m_{jk} u_j + \sum_{k \in I^-(j)} m_{jk} u_k}{s_j}. \end{cases} \tag{8.18}$$

With this notation, the upwind finite volume scheme is equivalent to (2.4). Note that the stability inequality may be rewritten as

$$\| (I + \Delta t A_h)^n \|_\alpha \leq 1, \quad \forall n \text{ provided (8.11)} \quad \text{(i.e., } K_\alpha = K = 1). \tag{8.19}$$

In order to study the convergence we define the approximation v_h^n of the exact solution:

$$v_h^n = \Pi_h u(n\Delta t).$$

This vector is a solution of the system

$$\begin{cases} \frac{v_h^{n+1} - v_h^n}{\Delta t} = A_h v_h^n + s_h^n, \quad n \geq 0, \\ v_h^0 = \Pi_h u_0. \end{cases} \tag{8.20}$$

The perturbation s_h^n is given by

$$\begin{aligned}
(s_h^n)_j &= \left(\frac{v_h^{n+1} - v_h^n}{\Delta t} - A_h v_h^n \right)_j \\
&= \frac{1}{s_j} \left(\int_{\Omega_j} \frac{u((n+1)\Delta t) - u(n\Delta t)}{\Delta t} + \sum_{k \in I^+(j)} m_{jk} (v_h^n)_j - \sum_{k \in I^-(j)} m_{jk} (v_h^n)_k \right) \\
&= \frac{1}{s_j} \left(\int_{\Omega_j} \int_{n\Delta t}^{(n+1)\Delta t} \partial_t u(s) ds + \sum_{k \in I^+(j)} m_{jk} (v_h^n)_j - \sum_{k \in I^-(j)} m_{jk} (v_h^n)_k \right) \\
&= \frac{1}{s_j} \left(- \int_{\Omega_j} \int_{n\Delta t}^{(n+1)\Delta t} \vec{a}.\nabla u(s) ds + \sum_{k \in I^+(j)} m_{jk} (v_h^n)_j - \sum_{k \in I^-(j)} m_{jk} (v_h^n)_k \right).
\end{aligned}$$

Finally, after integration by parts,

$$(s_h^n)_j = \frac{1}{s_j} \left(\sum_{k \in I^+(j)} m_{jk} ((v_h^n)_j - u_{h,jk}^n) - \sum_{k \in I^-(j)} m_{jk} ((v_h^n)_k - u_{h,jk}^n) \right), \tag{8.21}$$

where $u_{h,jk}^n$ stands for the time-and-edge average of the solution:

$$u_{h,jk}^n = \frac{1}{\Delta t} \int_{n\Delta t}^{(n+1)\Delta t} \int_{x \in \partial\Omega_j \cap \partial\Omega_k} u(s, x) ds d\sigma.$$

s_h^n given in (8.21) is a function of the difference between the cell average and the edge-time average $((v_h^n)_j - u_{h,jk}^n)$: s_h^n is like a bounded operator

$$\frac{\tau(h)}{s_j} \left(\sum_{k \in I^+(j)} m_{jk}(...) - \sum_{k \in I^-(j)} m_{jk}(...) \right)$$

applied to some bounded quantity (terms like $\frac{(v_h^n)_j - u_{h,jk}^n}{\tau(h)}$). So these s_h^n are $O(1)$. In general (that is, for an arbitrary mesh), there is no chance for this term to be $O(h)$, even for a very smooth solution u. This is precisely the lack of consistency problem that we are addressing in this paper. We rewrite this as

$$(s_h^n)_j = \frac{\tau(h)}{s_j} \left(\sum_{k \in I^+(j)} m_{jk} z_{h,jk}^n - \sum_{k \in I^-(j)} m_{jk} z_{h,jk}^n \right), \tag{8.22}$$

where

$$z_{h,jk}^n = \frac{(v_h^n)_j - u_{h,jk}^n}{\tau(h)}. \tag{8.23}$$

However, a difficulty occurs in dimension greater than 1. For a given h and a given n, $z_h^n = (z_{h,jk}^n)$ lives in a space which has the dimension of the number of edges, greater than the number of cells. In a 2D periodic domain with a triangular mesh, the number of edges is $\frac{3}{2}N$, where N is the number of cells. So there is no chance for z_h^n to belong to V_h^α, $z_h^n \notin V_h^\alpha$. Note that in dimension 1, the number of edges is equal to the number of cells; this is why it is possible to apply the abstract formalism directly only in dimension 1 (as it is done in section 7). In summary, one has

Lemma 8.4. *The truncation error* (8.21) *of the* 2*D advection equation is, in general, not admissible in the sense of Definition 5.1*: $s_h^n \neq \tau(h) A_h z_h$.

This is the classical dimensional obstruction to a simple proof of convergence for finite volume schemes. It means that the proof in dimension 1 needs to be adapted. Let us define

$$W_h^\alpha = \{z = (z_{h,jk});\ z_{h,jk} \text{ is constant in } \partial\Omega_j \cap \partial\Omega_k\ \forall j,k\}.$$

The norm in W_h^α is

$$||z||_\alpha = \left(\sum_{jk} (\tau(h) m_{jk}) |z_{h,jk}|^\alpha \right)^{\frac{1}{\alpha}} \quad \forall z \in W_h^\alpha, \ \ 1 \le \alpha < +\infty,$$

and

$$||z||_\infty = \max_{jk} |z_{h,jk}|.$$

Note that the coefficient $\tau(h) m_{jk} = \tau(h) l_{jk} |(\vec{a}, \vec{n}_{jk})|$ is homogeneous to a surface, $\tau(h) m_{jk} \approx s_j$. The norm in W_h^α is very similar to the norm in V_h^α. It is the reason why we use the same notation $||.||_\alpha$ for the norm in V^α and that in W_h^α.

Lemma 8.5. *Let* $\alpha \in\,]1, +\infty]$. *Assume that* $u_0 \in W_{\text{per}}^{1,\alpha}(\Omega)$ *(so* $u(t) = u_0(x - \vec{a}t) \in W_{\text{per}}^{1,\alpha}(\Omega)$, $\forall t$*). Then* $z_h^n = (z_{h,jk}^n) \in W_h^\alpha$, *and* $\exists C > 0$ *such that*

$$||z_h^n||_\alpha \le C ||\nabla u_0||_{L^\alpha(\Omega)}, \ \ \forall h, n, \ \forall \alpha \in [1, +\infty[. \tag{8.24}$$

Assume that $u_0 \in L^1(\Omega) \cap BV_{\text{per}}(\Omega)$ *(so* $u(t) = u_0(x - \vec{a}t) \in L^1(\Omega) \cap BV_{\text{per}}(\Omega)$, $\forall t$*). Then* $z_h^n \in W_h^1$, *and* $\exists C > 0$ *such that*

$$||z_h^n||_1 \le C ||u_0||_{BV_{\text{per}}(\Omega)}, \ \ \forall h, n. \tag{8.25}$$

The constant C is the same for all $\alpha \in [1, +\infty]$. Such a result is completely standard, so we skip its proof. See [15], where a very similar property is proved. We refer to [3], [18] for an introduction to the functional spaces $W^{1,\alpha}(\Omega)$ and $L^1(\Omega) \cap BV(\Omega)$: $W_{\text{per}}^{1,\alpha}(\Omega) \subset W^{1,\alpha}(\Omega)$ and $L^1(\Omega) \cap BV_{\text{per}}(\Omega) \subset L^1(\Omega) \cap BV(\Omega)$ are the restriction to periodic profiles of these spaces; we leave to the reader the straightforward generalization to other boundary conditions, such as incoming Dirichlet conditions.

9. Convergence via duality estimates

The previous section has shown us that the truncation error is not admissible in 2D, since its structure is not exactly what we have assumed in the general study. So we need another argument to be able to extend the class of perturbations considered in Definition 5.1. Duality is an appropriate tool for this task. The idea, very classical, is to apply the general argument, but for test functions.

The discrete duality product is

$$\langle v_h, w_h \rangle = \sum_j s_j v_{h,j} w_{h,j}, \ \forall v_h \in V_h^\alpha \text{ and } \forall w_h \in V_h^\beta, \quad \frac{1}{\alpha} + \frac{1}{\beta} = 1.$$

The duality product between $z_h \in W_h^\alpha$ and $w_h \in W_h^\beta$ is

$$\langle z_h, w_h \rangle = \sum_{jk} (\tau(h) m_{jk}) z_{h,jk} w_{jk}, \quad \frac{1}{\alpha} + \frac{1}{\beta} = 1.$$

The Hölder inequality [1] implies that

$$\|v_h\|_\alpha = \max_{w_h \in V_h^\beta, ||w_h||_\beta = 1} \langle v_h, w_h \rangle, \ \ \forall v_h \in V_h^\alpha.$$

We define the error

$$e_h^n = v_h^n - u_h^n. \tag{9.1}$$

Lemma 9.1. *The error is bounded by*

$$||e_h^n||_\alpha \le \Delta t \sum_{p=0}^{n-1} || \left(I + \Delta t A_h\right)^{n-1-p} s_h^p ||_\alpha. \tag{9.2}$$

Each term in the sum is

$$|| \left(I + \Delta t A_h\right)^q s_h^p ||_\alpha = \max_{||w_h||_\beta = 1} \left(\sum_{j<k} \sum_{k \in I^+(j)} (\tau(h) m_{jk}) z_{h,jk}^n (w_{h,j}^q - w_{h,k}^q) \right), \tag{9.3}$$

where the test function is $(q = n - 1 - p)$

$$w_h^q = \left(I + \Delta t A_h^*\right)^q w_h. \tag{9.4}$$

Since

$$e_h^n = \Delta t \sum_{p=0}^{n-1} \left(I + \Delta t A_h\right)^{n-1-p} s_h^p, \tag{9.5}$$

then the error is of course bounded by

$$||e_h^n||_\alpha \le \Delta t \sum_{p=0}^{n-1} || \left(I + \Delta t A_h\right)^{n-1-p} s_h^p ||_\alpha, \tag{9.6}$$

where

$$\begin{aligned} || \left(I + \Delta t A_h\right)^q s_h^p ||_\alpha &= \max_{w_h \in V_h^\beta, \ ||w_h||_\beta = 1} \langle w_h, \left(I + \Delta t A_h\right)^q s_h^p \rangle \\ &= \max_{w_h \in V_h^\beta, \ ||w_h||_\beta = 1} \langle \left(I + \Delta A_h^*\right)^q w_h, s_h^p \rangle = \max_{w_h \in V_h^\beta, \ ||w_h||_\beta = 1} \langle w_h^q, s_h^p \rangle . \end{aligned}$$

Here

$$w_h^q = \left(I + \Delta t A_h^*\right)^q w_h \quad \forall w_h \in W_h^\beta,$$

where A_h^* is the adjoint of A_h, defined by

$$\left\{ \begin{array}{l} A_h^* : V_h \to V_h, \\ (A_h u)_j = \frac{-\sum_{k\in I^-(j)} m_{jk}u_j + \sum_{k\in I^+(j)} m_{jk}u_k}{s_j}. \end{array} \right. \tag{9.7}$$

A direct calculation gives (s_h^n is given in (8.22))

$$\langle w_h^q, s_h^p \rangle = \sum_j s_j w_{h,j}^q \frac{\tau(h)}{s_j} \left(\sum_{k\in I^+(j)} m_{jk} z_{h,jk}^n - \sum_{k\in I^-(j)} m_{jk} z_{h,jk}^n \right),$$

where z_h^n is given in (8.23),

$$\langle w_h^q, s_h^p \rangle = \sum_{j<k} \sum_{k\in I^+(j)} (\tau(h) m_{jk}) z_{h,jk}^n (w_{h,j}^q - w_{h,k}^q). \tag{9.8}$$

This expression is the duality product in $W_h^\alpha \times W_h^\beta$ of z_h^n by $(w_{h,j}^q - w_{h,k}^q)$. In summary,

$$|| (I+\Delta t A_h)^q \, s_h^p ||_\alpha = \max_{||w_h||_\beta = 1} \left(\sum_{j<k} \sum_{k\in I^+(j)} (\tau(h) m_{jk}) z_{h,jk}^n (w_{h,j}^q - w_{h,k}^q) \right), \tag{9.9}$$

which ends the proof of the lemma.

The idea is then to prove that $(w_{h,j}^q - w_{h,k}^q) \in W_h^\beta$ in (9.3) is small. We know two *a priori* estimates for the test function $w_h^q \in V_h^\beta$.

a) First *a priori* estimate. The stability inequality (8.19) implies

$$|| (I+\Delta A_h^*)^q ||_\beta \le 1, \forall q \, \forall \beta, \ \forall h.$$

A consequence is of course

$$||w_h^q||_\beta \le ||w_h||_\beta = 1. \tag{9.10}$$

b) Second *a priori* estimate. We use Theorem 5.2 applied to

$$\tau(h) A_h^* w_h^q = (I + \Delta t A_h^*)^q (\tau(h) A_h^*) w_h.$$

Since $||w_h||_\beta = 1$, this proves that

$$\|\tau(h) A_h^* w_h^q\|_\beta \le \frac{C}{(q+1)^{\frac{1}{2}}}, \tag{9.11}$$

for a suitable constant $C > 0$.

Be aware that (9.10) and (9.11) give bounds in V_h^β, and not the bound of $(w_{h,j}^q - w_{h,k}^q) \in W_h^\beta$ we are looking for. In the following we show how to combine (9.10) and (9.11) in order to obtain a bound for $(w_{h,j}^q - w_{h,k}^q) \in W_h^\beta$.

10. The L^2 case

In order to better organize the rest of the proof, we separate the case $\alpha = 2$, treated in this section, and the case $2 < \alpha$, treated in next section. A comment about the case $\alpha < 2$ is given in the conclusion.

Lemma 10.1. *In the case $\alpha = \beta = 2$, there exists $C > 0$ such that the test function defined in* (9.4) *satisfies*

$$||(w^q_{h,j} - w^q_{h,k})||_\beta \leq \frac{C}{(q+1)^{\frac{1}{4}}}. \tag{10.1}$$

Let us compute the scalar product of w^q_h by $-\tau(h)A^*_h w^q_h$:

$$\begin{aligned} -\langle \tau(h)A^*_h w^q_h, w^q_h\rangle &= \tau(h)\sum_j \left(\sum_{k\in I^-(j)} m_{jk} w^q_{h,j} - \sum_{k\in I^+(j)} m_{jk} w^q_{h,k}\right) w^q_{h,j} \\ &= \tau(h)\sum_j \left(\sum_{k\in I^-(j)} m_{jk} (w^q_{h,j})^2 - \sum_{k\in I^+(j)} m_{jk} w^q_{h,j} w^q_{h,k}\right). \end{aligned}$$

But

$$w^q_{h,j} w^q_{h,k} = \frac{1}{2}(w^q_{h,j})^2 + \frac{1}{2}(w^q_{h,k})^2 - \frac{1}{2}(w^q_{h,j} - w^q_{h,k})^2.$$

So

$$\begin{aligned} -\langle \tau(h)A^*_h w^q_h, w^q_h\rangle &= \frac{1}{2}\tau(h)\sum_j \sum_{k\in I^-(j)} m_{jk}(w^q_{h,j})^2 \\ &\quad - \frac{1}{2}\tau(h)\sum_j \sum_{k\in I^+(j)} m_{jk}(w^q_{h,k})^2 + \frac{1}{2}\tau(h)\sum_j \sum_{k\in I^+(j)} m_{jk}(w^q_{h,j} - w^q_{h,k})^2. \end{aligned}$$

Reorganizing the two first terms using (8.10), we get[3]

$$-\langle \tau(h)A^*_h w^q_h, w^q_h\rangle = \frac{1}{2}\tau(h)\sum_j \sum_{k\in I^+(j)} m_{jk}(w^q_{h,j} - w^q_{h,k})^2.$$

Due to (9.10)-(9.11), this becomes

$$\left(\sum_j \sum_{k\in I^+(j)} (\tau(h)m_{jk})(w^q_{h,j} - w^q_{h,k})^2\right)^{\frac{1}{2}} \leq \frac{C}{(q+1)^{\frac{1}{4}}}.$$

This ends the proof of the lemma.

Theorem 10.2. *Assume that the mesh is uniformly regular. Assume the CFL inequality* (8.17)*. Assume that $u \in W^{1,2}_{\rm per}(\Omega)$, i.e., $u_0 \in W^{1,2}_{\rm per}(\Omega)$. Then*

$$\exists C > 0, \ \ ||(I + \Delta t A_h)^q s^p_h||_2 \leq \frac{C}{(q+1)^{\frac{1}{4}}}||\nabla u_0||_{L^2(\Omega)}, \tag{10.2}$$

and

$$\exists \tilde{C} > 0, \ \ ||u^n_h - \Pi_h u(n\Delta t)||_2 \leq \left(\tilde{C}||\nabla u_0||_{L^2(\Omega)}\right) T^{\frac{3}{4}} h^{\frac{1}{4}}. \tag{10.3}$$

Considering (9.8) and (10.1), we get that $\langle w^q_h, s^n_h\rangle \leq \frac{C}{(q+1)^{\frac{1}{4}}}||z^n_h||_2$. Now if the solution u is assumed to be in $W^{1,2}_{\rm per}(\Omega)$, we have (Lemma 8.5) $||z^n_h||_2 \leq C||\nabla u_0||_{L^2(\Omega)}$. So (10.2) is a direct consequence of (10.1). Even if not exactly admissible in the

[3]This inequality expresses the coercivity in L^2 of the operator $-A^*_h$ (resp. $-A_h$). It is the first time in this paper that we use such a property.

sense of definition 3, s_h^n is quasi-admissible with a weaker bound: $O(\frac{1}{(q+1)^{\frac{1}{4}}})$ instead of $O(\frac{1}{(q+1)^{\frac{1}{2}}})$. Using (9.6) and (10.2), we get

$$||e_h^n||_2 \leq (C\Delta t||\nabla u_0||_{L^2(\Omega)}) \sum_{p=0}^{n-1} \frac{1}{(p+1)^{\frac{1}{4}}} \leq (\tilde{C}\Delta t||\nabla u_0||_{L^2(\Omega)}) \times n^{\frac{3}{4}},$$

$$||e_h^n||_2 \leq (\tilde{C}\Delta t||\nabla u_0||_{L^2(\Omega)})(\frac{T}{\Delta t})^{\frac{3}{4}} = \left(\tilde{C}||\nabla u_0||_{L^2(\Omega)}\right) T^{\frac{3}{4}} h^{\frac{1}{4}}.$$

Note that the influence of the CFL condition is embedded in the constant $\tilde{C}$: just recall that $\tilde{C}$ depends on ν^- and ν^+.

11. The L^α case, $\alpha > 2$

Now we study the case $\alpha > 2$. The only difficulty lies in analyzing the "coercivity" in various L^β of the operator $-A_h^*$ (or $-A_h$). We will make use of various Hölder inequalities to get rid of it. The main result of this section is

Theorem 11.1. *Assume that the mesh is uniformly regular. Assume the CFL inequality* (8.17)*. Let* $2 < \alpha < +\infty$*. Assume that* $u_0 \in W_{\text{per}}^{1,\alpha}(\Omega)$*.*

Then there exists $C(\alpha) > 0$ *such that*

$$||u_h^n - \Pi_h u(n\Delta t)||_\alpha \leq (C(\alpha)||\nabla u_0||_\alpha)\, h^{\frac{1}{4}}, \;\; 0 \leq n\Delta t \leq T. \tag{11.1}$$

Let $\alpha = +\infty$*. Assume that* $u_0 \in W_{\text{per}}^{1,\infty}(\Omega)$*. Then* $\forall\varepsilon$*,* $0 < \varepsilon < \frac{1}{4}$*, there exists* $C(\varepsilon) > 0$ *such that*

$$||u_h^n - \Pi_h u(n\Delta t)||_\infty \leq (C(\varepsilon)||\nabla u_0||_\infty)\, h^{\frac{1}{4}-\varepsilon}, \;\; 0 \leq n\Delta t \leq T. \tag{11.2}$$

Let us begin the proof of the theorem by proving

Lemma 11.2. *If* $2 < \alpha < +\infty$ *and* $1 < \beta < 2$*, then there exists* $C(\alpha)$ *such that*

$$\sum_{j<k} \sum_{k \in I^+(j)} (\tau(h) m_{jk}) c_\beta(w_h^q)(w_{h,j}^q - w_{h,k}^q)^2 \leq \frac{C(\alpha)}{(q+1)^{\frac{1}{2}}}, \tag{11.3}$$

where the function c_β *is defined by*

$$c_\beta(a,b) = \int_{s=0}^1 (1-s)\; |a + s(b-a)|^{\beta-2} ds. \tag{11.4}$$

The proof of this lemma is essentially a generalization of the proof in the case $\alpha = \beta = 2$. We already know that

$$\|w_h^q\|_\beta \leq ||w_h||_\beta = 1, \tag{11.5}$$

and

$$\|\tau(h) A_h^* w_h^q\|_\beta \leq \frac{C}{(q+1)^{\frac{1}{2}}}, \tag{11.6}$$

where C is independent of β and w_h^q. Note that $|w_h^q|^{\beta-2} w_h^q \in V_h^\alpha$ with

$$|||w_h^q|^{\beta-2} w_h^q||_\alpha = \left(\sum_j s_j |w_{h,j}^q|^{(\beta-1)\alpha}\right)^{\frac{1}{\alpha}} = \left(\sum_j s_j |w_{h,j}^q|^\beta\right)^{\frac{1}{\alpha}} = ||w_h^q||_\beta^{\frac{\beta}{\alpha}} \leq 1.$$

We compute the duality product

$$
\begin{aligned}
&-\langle \tau(h) A_h^* w_h^q, |w_h^q|^{\beta-2} w_h^q \rangle \\
&= \sum_j \left(\sum_{k\in I^-(j)} m_{jk} w_{h,j}^q - \sum_{k\in I^+(j)} m_{jk} w_{h,k}^q \right) |w_{h,j}^q|^{\beta-2} w_{h,j}^q.
\end{aligned} \tag{11.7}
$$

First,

$$
\begin{aligned}
\left| \langle \tau(h) A_h^* w_h^q, |w_h^q|^{\beta-2} w_h^q \rangle \right| &\le ||\tau(h) A_h^* w_h^q||_\beta \times \|(|w_h^q|^{\beta-2} w_h^q)\|_\alpha \\
&= ||\tau(h) A_h^* w_h^q||_\beta \times \|w_h^q\|_\beta^{\frac{\alpha}{\beta}} \le \frac{C}{(q+1)^{\frac{1}{2}}}
\end{aligned}
$$

using (11.5) and (11.6). Second, we reorganize the right hand side of (11.7):

$$
\begin{aligned}
&\sum_j \left(\sum_{k\in I^-(j)} m_{jk} w_{h,j}^q - \sum_{k\in I^+(j)} m_{jk} w_{h,k}^q \right) |w_{h,j}^q|^{\beta-2} w_{h,j}^q \\
&= \sum_j \left(\sum_{k\in I^+(j)} m_{jk} (|w_{h,j}^q|^\beta - w_{h,k}^q |w_{h,j}^q|^{\beta-2} w_{h,j}^q) \right).
\end{aligned} \tag{11.8}
$$

We use Taylor's formula

$$
f(b) = f(a) + (a-b) f'(a) + \left(\int_0^1 (1-s) f''(a + s(b-a)) ds \right) (b-a)^2,
$$

with $f(a) = |a|^\beta$. So

$$
\begin{aligned}
|w_{h,k}^q|^\beta = {} & |w_{h,j}^q|^\beta + \beta |w_{h,j}^q|^{\beta-2} w_{h,j}^q (w_{h,k}^q - w_{h,j}^q) \\
& + \left(\beta(\beta-1) \int_{s=0}^1 (1-s) |w_{h,j}^q + s(w_{h,k}^q - w_{h,j}^q)|^{\beta-2} ds \right) (w_{h,k}^q - w_{h,j}^q)^2,
\end{aligned}
$$

and

$$
\begin{aligned}
w_{h,k}^q |w_{h,j}^q|^{\beta-2} w_{h,j}^q = {} & \frac{|w_{h,k}^q|^\beta + (\beta-1)|w_{h,j}^q|^\beta}{\beta} \\
& - (\beta-1) c_\beta(w_{h,j}^q, w_{h,k}^q)(w_{h,k}^q - w_{h,j}^q)^2.
\end{aligned}
$$

Plugging this expression into (11.7)-(11.8), we get

$$
\begin{aligned}
&-\langle \tau(h) A_h^* w_h^q, |w_h^q|^{\beta-2} w_h^q \rangle \\
&= \frac{1}{\beta} \sum_j \sum_{k\in I^-(j)} (\tau(h) m_{jk}) (w_{h,j}^q)^\beta - \frac{1}{\beta} \sum_j \sum_{k\in I^+(j)} (\tau(h) m_{jk}) (w_{h,k}^q)^\beta \\
&\quad + (\beta-1) \sum_j \sum_{k\in I^+(j)} (\tau(h) m_{jk}) c_\beta(w_{h,j}^q, w_{h,k}^q)(w_{h,k}^q - w_{h,j}^q)^2.
\end{aligned}
$$

Due to the relation (8.10) we obtain

$$
\begin{aligned}
&-\langle \tau(h) A_h^* w_h^q, |w_h^q|^{\beta-2} w_h^q \rangle \\
&= (\beta-1) \sum_j \sum_{k\in I^+(j)} (\tau(h) m_{jk}) c_\beta(w_{h,j}^q, w_{h,k}^q)(w_{h,k}^q - w_{h,j}^q)^2.
\end{aligned} \tag{11.9}
$$

Finally, using (11.5) and (11.6), we obtain (11.3). This ends the proof of this lemma.

This inequality (11.9) expresses the "coercivity" of $-A_h^*$. We observe that we only obtain a control in a "weighted L^2" of $w^q_{h,k} - w^q_{h,j}$. The rest of the proof is devoted to eliminating the weight $c_\beta(w^q_{h,j}, w^q_{h,k})$, in order to get a more explicit estimate.

Lemma 11.3. *Let $1 < \beta < 2$. There exists $C(\beta) > 0$ such that for all $(w_1, w_2) \in \mathbb{R}^2$, $\max(|w_1|, w_2|) \neq 0$,*

$$1 \leq C(\beta) \times \max(|w_1|^{2-\beta}, |w_2|^{2-\beta}) \times c_\beta(w_1, w_2). \tag{11.10}$$

a) Assume that $\max(|w_1|, |w_2|) = |w_1|$**.** Then

$$\max(|w_1|^{2-\beta}, |w_2|^{2-\beta}) c_\beta(w_1, w_2) = \int_{s=0}^1 (1-s)|1 + s(\frac{w_2}{w_1} - 1)|^{\beta-2} ds.$$

Since $-2 \leq \frac{w_2}{w_1} - 1 \leq 0$ and the function $x \to x^{\beta-2}$ is decreasing for $x > 0$ (recall that $\beta - 2 < 0$), we get

$$\max(|w_1|^{2-\beta}, |w_2|^{2-\beta}) c_\beta(w_1, w_2) \geq \int_{s=0}^{s=\frac{1}{3}} (1-s)(1-2s)^{\beta-2} ds = c_1(\beta) > 0.$$

b) Assume that $\max(|w_1|, |w_2|) = |w_2|$**.** Then

$$\max(|w_1|^{2-\beta}, |w_2|^{2-\beta}) c_\beta(w_1, w_2) = \int_{s=0}^1 (1-s)|1 + (1-s)(\frac{w_1}{w_2} - 1)|^{\beta-2} ds.$$

Since $-2 \leq \frac{w_1}{w_2} - 1 \leq 0$, we get

$$\begin{aligned} &\max(|w_1|^{2-\beta}, |w_2|^{2-\beta}) c_\beta(w_1, w_2) \\ &\qquad \geq \int_{s=\frac{2}{3}}^{s=1} (1-s)(1-2(1-s))^{\beta-2} ds = c_2(\beta) > 0. \end{aligned}$$

c) Defining $C(\beta) = \max(\frac{1}{c_1(\beta)}, \frac{1}{c_2(\beta)})$ ends the proof of the lemma.

Lemma 11.4. *Let $2 < \alpha < +\infty$. Consider the expression* (9.8)*. Then there exists $C(\beta)$ such that the test function w^q defined in* (9.4) *satisfies*

$$\left| \sum_{j<k} \sum_{k \in I^+(j)} (\tau(h) m_{jk}) z^n_{h,jk} (w^q_{h,j} - w^q_{h,k}) \right| \leq ||z^n_h||_\alpha \frac{C(\beta)}{(q+1)^{\frac{1}{4}}}. \tag{11.11}$$

We use

$$\begin{aligned} &\left| \sum_{j<k} \sum_{k \in I^+(j)} (\tau(h) m_{jk}) z^n_{h,jk} (w^q_{h,j} - w^q_{h,k}) \right| \\ &\quad = \left| \sum_{j<k} \sum_{k \in I^+(j)} (\tau(h) m_{jk}) z^n_{h,jk} \frac{1}{\sqrt{c_\beta(w^q_{h,j}, w^q_{h,k})}} \left(\sqrt{c_\beta(w^q_{h,j}, w^q_{h,k})} (w^q_{h,j} - w^q_{h,k}) \right) \right| \\ &\quad \leq C(\beta)^{\frac{1}{2}} \sum_{j<k} \sum_{k \in I^+(j)} (\tau(h) m_{jk}) |z^n_{h,jk}| \max(|w^q_{h,j}|, |w^q_{h,k}|)^{\frac{2-\beta}{2}} \\ &\qquad \times \left(\sqrt{c_\beta(w^q_{h,j}, w^q_{h,k})} |w^q_{h,j} - w^q_{h,k}| \right) \end{aligned}$$

Here we have used Lemma 11.3 to eliminate the weight $\frac{1}{\sqrt{c_\beta(w^q_{h,j}, w^q_{h,k})}}$. Note that the hypothesis of the lemma ($\max(|w_1|, w_2|) \neq 0$) is not needed, since the inequality makes sense even if $\max(|w^q_{h,j}|, |w^q_{h,k}|) = 0$. Using now the Hölder inequality in $W^\alpha_h \times W^\gamma_h \times W^2_h$ (of course $\frac{1}{\alpha} + \frac{1}{\gamma} + \frac{1}{2} = 1$, that is, $\gamma = \frac{\alpha-2}{2\alpha}$), we obtain

$$\begin{aligned}
&\left| \sum_{j<k} \sum_{k \in I^+(j)} (\tau(h) m_{jk}) z^n_{h,jk} (w^q_{h,j} - w^q_{h,k}) \right| \\
&\quad \leq C(\beta)^{\frac{1}{2}} \times ||z^n_h||_\alpha \times || \max(|w^q_{h,j}|, |w^q_{h,k}|)^{\frac{2-\beta}{2}} ||_\gamma \\
&\qquad \times || \sqrt{c_\beta(w^q_{h,j}, w^q_{h,k})} |w^q_{h,j} - w^q_{h,k}| ||_2.
\end{aligned}$$

We already know a bound for the last term (compare with (11.3)). It remains to study

$$|| \max(|w^q_{h,j}|, |w^q_{h,k}|)^{\frac{2-\beta}{2}} ||_\gamma = \left(\sum_{jk} (\tau(h) m_{jk}) \max(|w^q_{h,j}|, |w^q_{h,k}|)^{\frac{(2-\beta)\gamma}{2}} \right)^{\frac{1}{\gamma}}.$$

Since $\frac{1}{\gamma} + \frac{1}{2} = \frac{1}{\beta}$, we get $\frac{(2-\beta)\gamma}{2} = \beta$. So

$$\begin{aligned}
|| \max(|w^q_{h,j}|, |w^q_{h,k}|)^{\frac{2-\beta}{2}} ||_\gamma &= \left(\sum_{jk} (\tau(h) m_{jk}) \max(|w^q_{h,j}|, |w^q_{h,k}|)^\beta \right)^{\frac{1}{\gamma}} \\
&= || \max(|w^q_{h,j}|, |w^q_{h,k}|) ||_\beta^{\frac{\beta}{\gamma}}.
\end{aligned}$$

This expression is a power of the $||.||_\beta$ norm of $(\max(|w^q_{h,j}|, |w^q_{h,k}|)) \in W^\beta_h$. We bound this by the $||.||_\beta$ norm of $w^q_h \in V^\beta_h$, up to some multiplicative factor:

$$\begin{aligned}
&\left(\sum_{jk} (\tau(h) m_{jk}) \max(|w^q_{h,j}|, |w^q_{h,k}|)^\beta \right) \leq \left(\sum_{jk} (\tau(h) m_{jk}) (|w^q_{h,j}|^\beta + |w^q_{h,k}|^\beta) \right) \\
&\quad \leq \left(\sum_j \left(\sum_k (\tau(h) m_{jk}) \right) |w^q_{h,j}|^\beta \right) \leq \left(3 \max_{jk} \frac{\tau(h) m_{jk}}{s_j} \right) \left(\sum_j s_j |w^q_{h,j}|^\beta \right).
\end{aligned}$$

So

$$|| \max(|w^q_{h,j}|, |w^q_{h,k}|)^{\frac{2-\beta}{2}} ||_\gamma \leq \left(3 \max_{jk} \frac{\tau(h) m_{jk}}{s_j} \right)^{\frac{1}{\gamma}} ||w^q_h||_\beta^{\frac{1}{\gamma}} \leq \tilde{C}(\gamma),$$

thanks to the uniform regularity of the mesh and (9.10). So finally we get (11.11).

Final Proof of Theorem 11.1. Inequality (11.2) is a consequence of (9.6), (9.9) and (11.11). Concerning (11.2), we note[4] that $W^{1,\infty}_{\text{per}}(\Omega) \subset W^{1,\alpha}_{\text{per}}(\Omega)$, $\forall 1 \leq \alpha < +\infty$, use (11.1) and let $\alpha \to +\infty$. So

$$\left[\sum_j s_j \left| (u^n_{j,h}) - \Pi_h u(n\Delta t)_j \right|^\alpha \right]^{\frac{1}{\alpha}} \leq C(\alpha) ||\nabla u_0||_\infty h^{\frac{1}{4}}.$$

Since $s_j \geq c_2 h^2$, which is true for uniformly regular meshes, this implies

$$c_2^{\frac{1}{\alpha}} h^{\frac{2}{\alpha}} \left| (u^n_{j,h}) - \Pi_h u(n\Delta t)_j \right| \leq C(\alpha) ||\nabla u_0||_\infty h^{\frac{1}{4}},$$

[4] It is essentially here that we use the boundedness of Ω.

i.e.,

$$\left|(u^n_{j,h}) - \Pi_h u(n\Delta t)_j\right| \leq \frac{C(\alpha)}{c_2^{\frac{1}{\alpha}}} ||\nabla u_0||_\infty h^{\frac{1}{4}-\frac{2}{\alpha}},$$

which is exactly (11.2): $\varepsilon = \frac{2}{\alpha}$. This finishes the proof.

Considering Theorem 6.1, we see that all the above theorems of convergence are also true for the implicit scheme.

12. Conclusion

We have presented an abstract framework for the numerical approximation in general Banach spaces of linear equations by means of finite volume methods. We emphasize an important property of finite volume methods: they are formally non-consistent, so the Lax theorem does not apply. The main result of this work is a proof that cancellation in time of the error is a reason why finite volume methods converge. This result may be considered as a **variation on the Lax theorem**.

Application of finite volume methods for the numerical solution of linear advection on 2D triangular meshes gives some insight on the potentiality of the technique: in this work we have studied convergence only in L^α, $2 \leq \alpha$, but by continuity of all these estimates of convergence with respect to α, it is possible to extend some of them to the case $1 \leq \alpha < 2$.

In this paper we have restricted the discussion to what we think is the core of the method. We will report about some improvements of all these estimates in a forthcoming work. An important issue is to recover the optimal rate of convergence in L^1, proved in [10] for hyperbolic scalar laws in dimension two by means of completely different techniques.

References

[1] Abramowitz M. and Stegun I.A., Handbook of mathematical functions, Dover publications, New York. MR **94b:**00012

[2] R. Botchorishvili, B. Perthame and A. Vasseur, Schémas d'équilibre pour des lois de conservation scalaires avec des termes sources raides, Report 3891, INRIA, France, 2000. Cf. Math. Comp. 72 (2003), 131–157. MR **2003h:**65104

[3] H. Brezis, Analyse Fonctionelle, Masson Paris (1983). MR **85a:**46001

[4] C. Chainais-Hillairet, First and second order schemes for a hyperbolic equation: convergence and error estimate, in *Finite volume for complex applications: Problems and perspectives*, Benkhaldoun and Vilsmeier editors, Hermes Paris (1997), 137–144.

[5] S. Champier, T. Gallouët and R. Herbin, Convergence of an upstream finite volume scheme for a nonlinear hyperbolic equation on a triangular mesh, Numerische Math. 66, (1993), 139–157. MR **95b:**65117

[6] P.G. Ciarlet, The finite element method for elliptic problems, North Holland, Amsterdam (1978). MR **58:**25001

[7] B. Cockburn, F. Coquel and P. Le Floch, An error estimate for finite volume multidimensional conservation laws, Math. Comp., 63, pp 77-103 (1994). MR **95d:**65078

[8] B. Cockburn, Devising discontinuous Galerkin methods for non-linear hyperbolic conservation laws, Journal of Computational and Applied Math, 138, pp 187–204 (2001). MR **2001m:**65127

[9] B. Cockburn and P. A. Gremaud, A priori estimates for numerical methods for scalar conservation laws, part II: multidimensional flux-splitting monotone schemes on non Cartesian grids, SINUM, 35, pp 1775–1803 (1998). MR **99g:**65097

[10] B. Cockburn, P. A. Gremaud and J. X. Yang, A priori estimates for hyperbolic conservation laws, part II: flux-splitting monotone schemes on irregular Cartesian grids, Math. Comp., 66, pp 547–572 (1997). MR **97m:**65173

[11] Y. Coudière, T. Gallouët, and R. Herbin, Discrete Sobolev Inequalities and L^p Error Estimates for Finite Volumes Solutions of Convection Diffusion Equations, M2AN Math. Model. Numer. Anal. 35 (2001), 767–778. MR **2002h:**65167

[12] Y. Coudière, J.P. Vila and P. Villedieu, Convergence d'un schéma volumes finis explicite en temps pour les systèmes hyperboliques linéaires symétriques en domaines bornés, C. R. Acad. Sci. Paris, t. 331, Série I, p. 95-100 (2000). MR **2001d:**65114

[13] R. Dautray and J. L. Lions, Analyse numérique et calcul numérique pour les sciences et les techniques, Masson, 1984. MR **87g:**00002

[14] B. Després and F. Lagoutière, Contact Discontinuity Capturing Schemes for Linear Advection and Compressible Gas Dynamics, J. Sci. Comput. 16 (2001), no. 4, 479–524 (2002). MR **2002k:**65121

[15] B. Després and F. Lagoutière, Generalized Harten formalism and longitudinal variation diminishing schemes for linear advection on arbitrary grids, M2AN Math. Model. Numer. Anal. 35, no. 6, (2001) 1159–1183. MR **2003e:**76079

[16] R.J. DiPerna, Measure value solutions to conservation laws, Arch. Rat. Mech. Anal, 88, (1985), 223–270. MR **86g:**35121

[17] R. Eymard, T. Gallouët and R. Herbin, Handbook of Numerical Analysis, Vol. VII, North-Holland, 2000, p. 723-1020. Editors: P.G. Ciarlet and J.L. Lions. MR **2001h:**65001

[18] E. Giusti, *Minimal surfaces and functions of bounded variation*, Birkhäuser, 1984. MR **87a:**58041

[19] C. Johnson and J. Pitkäranta, An analysis of the discontinuous Galerkin method for a sclar hyperbolic equation, Math. Comp., 46 (173), 1–26 (1986). MR **88b:**65109

[20] R.J. LeVeque, *Numerical methods for conservation laws.* (ETHZ Zurich, Birkhauser, Basel 1992). MR **92m:**65106

[21] R.J. Leveque, High-resolution conservative algorithms for advection in incompressible flows, SIAM J. Numer. Anal., 33 (2), (1996) 627–665. MR **98b:**76049

[22] P.L. Lions, B. Perthame and E. Tadmor, A kinetic formulation of multidimensional scalar conservation laws and related equations, J. Amer. Math. Soc., 7, (1994), 169–191. MR **94d:**35100

[23] P. Lesaint and P. A. Raviart, On a finite element method for solving the neutron transport equation, Mathematical Aspects of Finite Elements in Partial Differential Equations, Academic Press New York, (1974), pp 89-123. MR **58:**31918

[24] T. A. Manteuffel and A. B. White Jr., The numerical solution of second-order boundary value problems on nonuniform meshes. Math. Comp. 47 (1986), no. 176, 511–535. MR **87m:**65116

[25] T. A. Manteuffel and A. B. White Jr., A calculus of difference schemes for the solution of boundary value problems on irregular meshes. SIAM J. Numer. Anal. 29 (1992), no. 5, 1321–1346. MR **94b:**65139

[26] P.A. Raviart and E. Godlewski, Numerical approximation of hyperbolic systems of conservation laws, Appl. Math. Sci., vol. 118, Springer Verlag New York, (1996). MR **98d:**65109

[27] P.A. Raviart and J.M. Thomas, Introduction à l'analyse numérique des équations aux dérivées partielles, Masson Paris (1983). MR **87a:**65001a

[28] R. D. Richtmyer and K. W. Morton, Difference methods for initial-value problems, Interscience Publishers, 1957. MR **20:**438

[29] P.L. Roe and D. Sidilkover, Optimum positive linear schemes for advection in two and three dimensions, SIAM J. Numer. Anal., 29 (6), (1992) 1542–1568. MR **93j:**65137

[30] A. Szepessy, Convergence of a streamline diffusion finite element method for conservation law with boundary conditions, RAIRO Model. Math. et Anal. Num., 25, (1991) 749–783. MR **92g:**65115

[31] B. Wendroff, Supraconvergence in two dimensions, Los Alamos National Laboratory report LA-UR-95-3068, Vieweg (1995).

[32] B. Wendroff and A. B. White, Some supraconvergent schemes for hyperbolic equations on irregular grids, Nonlinear Hyperbolic Equations—Theory, Computation Methods, and Applications (Proc. Second Internat. Conf. Nonlinear Hyperbolic Problems, Aachen, 1988; J. Ballmann and R. Jeltsch, editors), Vieweg, pp 671–677 (1988). MR **89j:**35007

[33] B. Wendroff and A. B. White, A supraconvergent scheme for nonlinear hyperbolic systems, Comput. Math. Appl., 18, pp 761–767 (1989). MR **90g:**65121

Commissariat à l'Energie Atomique, 91680, Bruyères le Chatel, France

Current address: Laboratoire d'analyse numérique, 175 rue du Chevaleret, Université de Paris VI, 75013 Paris, France

E-mail address: `despres@ann.jussieu.fr, bruno.despres@cea.fr`