Comptes Rendus

Mécanique

Robert Eymard, Thierry Gallouët, Raphaele Herbin and Jean-Claude Latché

Finite volume schemes and Lax–Wendroff consistency

Research article / *Article de recherche*

More than a half century of Computational Fluid Dynamics / *Plus d'un demi-siècle de mécanique des fluides numérique*

Finite volume schemes and Lax–Wendroff consistency

Robert Eymard [a], **Thierry Gallouët** [b], **Raphaele Herbin** [*, b] and **Jean-Claude Latché** [c]

[a] LAMA, Université Gustave Eiffel, Marne la Vallée, France

[b] I2M UMR 7373 CNRS & Aix-Marseille Université, France

[c] IRSN Cadarache, France

E-mails: robert.eymard@univ-eiffel.fr (R. Eymard), thierry.gallouet@univ-amu.fr
(T. Gallouët), raphaele.herbin@univ-amu.fr (R. Herbin), jean-claude.latche@irsn.fr
(J.-C. Latché)

*This work is dedicated to the memory of our dear colleague and friend Antonín Novotný,
who suddenly passed away on June 3, 2021*

Abstract. We present a (partial) historical summary of the mathematical analysis of finite difference and finite volume methods, paying special attention to the Lax–Richtmyer and Lax–Wendroff theorems. We then state a Lax–Wendroff consistency result for convection operators on staggered grids (often used in fluid flow simulations), which illustrates a recent generalization of the flux consistency notion designed to cope with general discrete functions.

The finite volume method (FVM) has been used for over 60 years in computational fluid mechanics, and more than 30 years in solid mechanics. However, while the finite element method, also introduced in mechanics, had already been the object of several mathematical works at the turn of this millenium, the American Mathematical Society's (AMS) 2000 Classification only mentions the term "finite volume" in sections 74S10 (Mechanics of deformable solids) and 76M12 (Fluid mechanics); it was not until 2010 that this term appeared in section 65 "Numerical analysis" (65M08, 65N08).

In some respects, the FVM is close to the finite difference method (FDM): for instance it has often been called "conservative finite difference method" by the hyperbolic numerical community, and "finite difference method" in the oil industry. It might be for this reason that the first attempts to show the convergence of the FVM were to try and copy the FDM framework. We show in the first section below that the famous Lax–Richtmyer theorem developed for linear FD schemes fails to give an adequate answer, even for a linear finite volume (FV) scheme. We then turn to

---

* Corresponding author

the Lax–Wendroff theorem that gives two fundamental tools for the analysis of FV schemes, and recapitulate the main steps of the proof of convergence of the schemes that were derived in the 90's. In Section 4, we extend the Lax–Wendroff theorem to the case of staggered meshes and show how it can be used for the celebrated MAC grid.

1. The finite difference method and the Lax–Richtmyer theorem

In classical numerical analysis textbooks, we are taught that, in order to show the convergence of a finite difference (FD) scheme, one should show its stability and its consistency. The founding result in this regard is the Lax–Richtmyer theorem [1] due to P. D. Lax, who exposed it in a seminar at NYU in 1954. The so-called Lax equivalence theorem can be summarized as follows (see for example [2, Theorem 1.5.1]):

Theorem 1.1 (Lax–Richtmyer). *Consider a linear partial differential equation for which the initial value problem is well posed, and a FD scheme consistent for its approximation; then this scheme is convergent if and only if it is stable.*

In some articles and textbooks (see for instance [3, p. 142]), maybe because of the name "equivalence theorem", and because uniform meshes are considered, the Lax–Richtmyer theorem is replaced by the equivalence

$$\text{"consistency} + \text{stability} \Longleftrightarrow \text{convergence"} \tag{1}$$

However, Theorem 1.1 does not state the equivalence (1), which, in fact does not hold in the general case of a PDE which is discretized with a non-constant space step, even in the linear case. As an example, let us consider the approximation on $\mathbb{R} \times]0, T[$, where $T > 0$ is the final time of the study, of the linear transport equation

$$\partial_t u(x, t) + \partial_x(au)(x, t) = 0, \quad x \in \mathbb{R}, \, t \in]0, T[, \tag{2}$$

$$u(x, 0) = u_{\text{ini}}(x), \quad x \in \mathbb{R}, \tag{3}$$

for some given $a > 0$, and initial data $u_{\text{ini}} \in C_c^\infty(\mathbb{R}, \mathbb{R})$, with ∂_t (resp. ∂_x) the time (resp. space) partial derivative. Let $\bar{u}(x, t) = u_{\text{ini}}(x - at)$ be the exact (unique) solution of this problem. The FDM applied to (3) is classically defined by choosing a strictly increasing sequence $(x_i)_{i \in \mathbb{Z}}$ of real numbers, such that $h := \max_{i \in \mathbb{Z}}(x_{i+1} - x_i) < \infty$ and $\underline{h} = \min_{i \in \mathbb{Z}}(x_{i+1} - x_i) > 0$, and a time step $\delta t = T/N$, for $N \in \mathbb{N}$ with $N > 1$ (here a constant time step is considered for simplicity). The initial data is discretized by defining the following quantities that depend on h implicitly:

$$u_i^0 = u_{\text{ini}}(x_i) \quad \text{for any } i \in \mathbb{Z}. \tag{4}$$

Since some upwinding is indeed necessary for stability purposes, the approximation of $\partial_x u$ at the point x_i is upwinded, so that the FD scheme reads

$$\frac{u_i^{n+1} - u_i^n}{\delta t} + a \frac{u_i^n - u_{i-1}^n}{h_{i-1/2}} = 0 \quad \text{for any } i \in \mathbb{Z} \text{ and } n \in [\![0, N-1]\!], \tag{5}$$

with $h_{i-1/2} = x_i - x_{i-1}$. In this context, the definition of the terms *consistency, stability* and *convergence* is the following:

- *Consistency*: it requires two conditions:
  - the consistency of the discretization of the initial condition, that is

$$\lim_{h \to 0} \max_{i \in \mathbb{Z}} |u_i^0 - u_{\text{ini}}(x_i)| = 0. \tag{6}$$

– the consistency of the discretization of the PDE (2); setting $t_n = n\delta t$ for $n \in \mathbb{N}$, it reads:

$$\lim_{\substack{h \to 0 \\ \delta t \to 0}} \max_{i \in \mathbb{Z}, n \in [\![0, N-1]\!]} \left| \frac{u(x_i, t_{n+1}) - u(x_i, t_n)}{\delta t} + a \frac{u(x_i, t_n) - u(x_{i-1}^n, t_n)}{h_{i-1/2}} \right| = 0. \qquad (7)$$

- *Stability*: There exists C depending only on u_{ini} (and thus not on h nor on δt) such that

$$\max_{i \in \mathbb{Z}, n \in [\![0, N-1]\!]} |u_i^n| \le C. \qquad (8)$$

- *Convergence*: $\max_{i \in \mathbb{Z}, n \in [\![0, N-1]\!]} |u_i^n - u(x_i, t_n)| \to 0$ as $h \to 0$.

Clearly, condition (4) ensures the consistency condition (6) on the initial condition. Moreover, the consistency condition (7) on the PDE can be obtained by Taylor expansions. Finally, the stability condition (8) can be shown under the CFL (Courant–Friedrichs–Lewy) condition $a\delta t \le \underline{h}$. The conditions for the Lax–Richtmyer to hold are therefore satisfied, and the convergence of the scheme (4)–(5) is thus proven.

Consider now a variant of the scheme (5) obtained by keeping (4), but replacing $h_{i-1/2}$ by $h_i = (x_{i+1} - x_{i-1})/2$ in (5):

$$\frac{u_i^{n+1} - u_i^n}{\delta t} + a \frac{u_i^n - u_{i-1}^n}{h_i} = 0 \quad \text{for any } i \in \mathbb{Z} \text{ and } n \in [\![0, N-1]\!]. \qquad (9)$$

In the specific case where $x_{2k+1} - x_{2k} = h/2$ and $x_{2k+2} - x_{2k+1} = h$ for all $k \in \mathbb{Z}$, we get that $h_i = 3/4h$ for all $i \in \mathbb{Z}$ and the consistency property no longer holds. Therefore, if the equivalence (1) were true, the scheme (4)–(9) would not be convergent in the above sense. However, let us show that this scheme is in fact convergent in the same sense. Let us write the finite difference scheme obtained at the points $\widetilde{x}_i = (x_i + x_{i+1})/2$ instead of the points x_i, defining the values $\widetilde{u}_i^n$:

$$\widetilde{u}_i^0 = u_{\text{ini}}(\widetilde{x}_i) \quad \text{for any } i \in \mathbb{Z}, \qquad (10)$$

and

$$\frac{\widetilde{u}_i^{n+1} - \widetilde{u}_i^n}{\delta t} + a \frac{\widetilde{u}_i^n - \widetilde{u}_{i-1}^n}{h_i} = 0 \quad \text{for any } i \in \mathbb{Z} \text{ and } n \in [\![0, N-1]\!]. \qquad (11)$$

Then the scheme (10)–(11) again satisfies the consistency condition (6), and satisfies the consistency condition (7) since $h_i = \widetilde{x}_i - \widetilde{x}_{i-1}$. The stability condition (8) can again be shown under the CFL condition $a\delta t \le \underline{h} \le 3/4h$. Hence the Lax–Richtmyer theorem yields the convergence of the values $\widetilde{u}_i^n$ to $u(\widetilde{x}_i, t_n)$. Using $|\widetilde{u}_i^0 - u_i^0| \le h \max|u_{\text{ini}}'|$, the maximum principle applied to the difference $\widetilde{u}_i^n - u_i^n$, solution to the equation obtained by subtracting (9) to (11), shows that

$$\max_{i \in \mathbb{Z}, n \in [\![0, N-1]\!]} \left| \widetilde{u}_i^n - u_i^n \right| \le h \max|u_{\text{ini}}'|,$$

which implies that

$$\max_{i \in \mathbb{Z}, n \in [\![0, N-1]\!]} |u_i^n - u(x_i, t_n)| \le 2h \max|u_{\text{ini}}'| + \max_{i \in \mathbb{Z}, n \in [\![0, N-1]\!]} |\widetilde{u}_i^n - u(\widetilde{x}_i, t_n)|,$$

leading to the convergence of the scheme in the same sense as above. This example shows that on the one hand the direction $\Leftarrow$ of (1) is not true, and that on the other hand the Lax–Richtmyer theorem cannot be applied directly to obtain the convergence of such a scheme, since the scheme is not consistent in the above defined sense. In fact, the scheme (9) is the 1D upwind FV scheme with the control volumes $](x_{i-1} + x_i)/2, (x_i + x_{i+1})/2[$, as shown in the next section. In the scheme (5), the partial derivative $\partial_x u$ is upwinded. In the scheme (9), the unknown u itself is upwinded.

Note that the equivalence (1) also does not hold in the case of an elliptic operator, see the example of a non-uniform 1D mesh in [4] or [5, Section 5.2]. For the analysis of such schemes, other notions must be introduced, and it seems that Peter D. Lax first identified them, in collaboration with B. Wendroff.

2. The finite volume method and the Lax–Wendroff theorem

In a very famous article of 1960, P. D. Lax and B. Wendroff [6] consider discretization schemes for nonlinear hyperbolic systems of conservation laws and show that if a conservative scheme with consistent fluxes, in the sense that they define and which is stated below (see Section 2.2), converges a.e. and boundedly towards a limit, then this limit is necessarily a weak solution of the system. We call this property *Lax–Wendroff (LW) consistency*. The notions of *flux consistency* and *flux conservativity* highlighted in [6] are truly fundamental for the convergence analysis of the FVM for the hyperbolic equations considered in [6], as are their extensions to elliptic and parabolic conservation equations. In order to explain these terms, we consider, again in the 1D case, a general differential form of a conservation law, written on the whole space $\mathbb{R}$ and on a time interval $]0, T[$ where $0 < T < +\infty$ is the final time:

$$\partial_t u(x, t) + \partial_x \mathbb{F}(x, t) = 0, \tag{12}$$

stating the conservation of the quantity u at each point $x \in \mathbb{R}$ and each time $t \in]0, T[$, with $\mathbb{F}$ a vector function depending only on x and t through the unknown u. In addition to this equation, an initial condition is assumed to be given on u. Simple examples of such a conservation law include

- the transport equation $\mathbb{F}(x, t) = au(x, t)$ (linear hyperbolic equation, introduced in the previous section),
- the heat equation $\mathbb{F}(x, t) = -\partial_x u(x, t)$ (linear parabolic equation),
- Burgers' equation $\mathbb{F}(x, t) = u^2(x, t)$ (nonlinear hyperbolic equation),
- the porous media equation $\mathbb{F}(x, t) = -\partial_x u^p(x, t)$, $p > 1$ (nonlinear parabolic equation).

The FVM consists in approximating the integral form of the conservation law, that is to say the balance on the time–space cuboid $]x, x + \delta x[\times]t, t + \delta t[$ (for given $\delta x > 0$ and $\delta t > 0$), rather than the PDE itself (this corresponds to the way such an equation is derived from physical conservation principles). Note that with a non-zero source term on the right-hand side of (12), the equation is then usually renamed "balance law", and this modification poses no problem for a FV discretization. The integral form relative to the differential form (12) of a conservation law reads

$$\int_x^{x+\delta x} (u(x, t + \delta t) - u(x, t)) \, \mathrm{d}x + \int_t^{t+\delta t} \left(\mathbb{F}(x + \delta x, t) - \mathbb{F}(x, t) \right) \mathrm{d}t = 0. \tag{13}$$

Let $(]x_{i-1/2}, x_{i+1/2}[)_{i \in \mathbb{Z}}$ be a family of intervals of $\mathbb{R}$ (also called control volumes or grid cells), with $x_{i-1/2} < x_{i+1/2}$ and such that $\cup_{i \in \mathbb{Z}}]x_{i-1/2}, x_{i+1/2}[= \mathbb{R}$ (or $\Omega \subset \mathbb{R}$), and let $\delta t = T/N$, $N > 1$ (the time step could be taken non-constant). Let x_i be some chosen point in the cell $]x_{i-1/2}, x_{i+1/2}[$ (the choice of this point is constrained by the flux consistency property in the case of elliptic or parabolic problems, but not in the hyperbolic case, contrarily to the FD consistency in the sense of the previous section). The discrete unknowns, $\{u_i^n, i \in \mathbb{Z}, n \in \{0, \dots, N\}\}$ are expected to be approximations of $u(x_i, t_n)$ with $t_n = n\delta t$. The integral form (13) is written on each control volume $]x_{i-1/2}, x_{i+1/2}[$ and time interval $]t_n, t_{n+1}[$, leading to the following FV scheme (with an explicit time scheme for the sake of simplicity):

$$h_i \frac{u_i^{n+1} - u_i^n}{\delta t} + F_{i+1/2}^n - F_{i-1/2}^n = 0 \quad \text{for any } i \in \mathbb{Z} \text{ and } n \in [\![0, N-1]\!], \tag{14}$$

where $h_i = x_{i+1/2} - x_{i+1/2}$, and $F_{i+1/2}^n$ is the numerical flux, which will be expressed in terms of the discrete unknowns to yield a numerical approximation of $\mathbb{F}(x_{i+1/2}, t_n)$. Note that $F_{i+1/2}^n$ is the numerical flux outgoing $]x_{i-1/2}, x_{i+1/2}[$ to the right and its opposite is the numerical flux outgoing $]x_{i+1/2}, x_{i+3/2}[$ to the left: this copies the situation of the exact flux $\mathbb{F}(x_{i+1/2}, t_n)$. This is the well-known "*local conservativity*" or "*flux conservativity*" property, which is important in physical applications, but also fundamental in the mathematical analysis of the FVM. Indeed, it

is thanks to this property that we may hope to prove some convergence properties of the method, both for elliptic or parabolic type and for hyperbolic equations despite the loss of the consistency in the FD sense, as presented in the first section of this paper.

2.1. *Flux conservativity*

Writing the FV scheme in one dimension naturally ensures the numerical flux conservativity, since only one flux is defined: $F_{i+1/2}$ at the interface $x_{i+1/2}$. In a multi-dimensional framework, ($d = 2$ or 3) the PDE (12) is written as $\partial_t u + \text{div}\mathbb{F} = 0$, where $\mathbb{F}$ is a vector function of x and t and div the space divergence operator. The scheme (14) is now written for a control volume K:

$$|K|\frac{u_K^{n+1} - u_K^n}{\delta t} + \sum_{\sigma \subset \partial K} |\sigma| F_{K,\sigma}^n = 0, \quad n \in [\![0, N-1]\!], \tag{15}$$

where $|K|$ (resp. $|\sigma|$) is the volume or surface of K (resp. the surface or length of $|\sigma|$) and $|\sigma| F_{K,\sigma}^n$ is the numerical outgoing flux from K through the face σ; it is an approximation of the ougoing normal flux $\int_\sigma \mathbb{F}(x, t_n) \cdot \boldsymbol{n}_{K,\sigma}$ (where $\boldsymbol{n}_{K,\sigma}$ is the normal vector to σ outward K), which is expressed in terms of the discrete unknowns $(u_M^n)_{M \in \mathcal{M}}$:

$$F_{K,\sigma}^n = \mathscr{F}_{K,\sigma}^n((u_M^n)_{M \in \mathcal{M}}).$$

In 2 or 3D, the numerical flux is defined on either side of the interface σ. Suppose that the interface σ separates the control volumes K and L, which we write as $\sigma = K|L$, then the flux conservativity reads

$$F_{K,\sigma}^n = -F_{L,\sigma}^n. \tag{16}$$

2.2. *Flux consistency*

Let us turn back to the 1D case, for ease of notations. The numerical flux is said to be consistent if for a sequence of time and space discretizations, indexed by m and such that $h^{(m)} \to 0$ and $\delta t^{(m)} \to 0$ as $m \to +\infty$, one has

$$\max_{i \in \mathbb{Z}, n \in [\![0, N-1]\!]} |\mathbb{F}(x_{i+1/2}, t_n) - \widetilde{F}_{i+1/2}^n| \to 0 \quad \text{as } m \to +\infty,$$

where $\widetilde{F}_{i+1/2}^n$ is the quantity obtained from $F_{i+1/2}^n$ when replacing the discrete unknowns by the values of a regular function u:

$$\widetilde{F}_{i+1/2}^n = \mathscr{F}_{i+1/2}^n((u(x_i, t_n))_{i \in \mathbb{Z}}).$$

In the context of nonlinear hyperbolic equations, $\mathbb{F}(x_{i+1/2}, t_n) = f(u(x_{i+1/2}, t_n))$ with $f \in C(\mathbb{R}, \mathbb{R})$, this definition of consistency is equivalent to the usual notion of consistency introduced by Lax (written here for a two-point scheme):

$$\mathscr{F}_{i+1/2}^n(u, u) = f(u)$$

if one assumes $\mathscr{F}_{i+1/2}^n$ to be Lipschitz continuous, or at least, "lip-diag" see [7, Remark 5.2]. In the context of the heat equation ($F(u) = -\nabla u$), the numerical flux

$$F_{i+1/2}^n = \mathscr{F}_{i+1/2}^n(u_i^n, u_{i+1}^n) = -\frac{u_{i+1}^n - u_i^n}{h_{i+1/2}} \quad \text{with } h_{i+1/2} = x_{i+1} - x_i, \tag{17}$$

is consistent in the above sense, since, for a regular function u,

$$\mathbb{F}(x_{i+1/2}, t_n) = -\partial_x u(x_{i+1/2}, t_n)$$

and

$$\widetilde{F}_{i+1/2}^n(u(x_i, t_n), u(x_{i+1}, t_n)) = -\frac{u(x_{i+1}, t_n) - u(x_i, t_n)}{h_{i+1/2}}$$
$$= -\partial_x u(x_{i+1/2}, t_n) - h_{i+1/2}\partial_{xx}^2 u(c, t_n), \ c \in \,]x_i, x_{i+1}[.$$

We notice that in this case the flux depends on the choice of the points x_i.

3. Stability, compactness, convergence

3.1. *Stability*

There are different notions of stability of a numerical scheme. The notion which is of interest in the context of the convergence of a numerical scheme for a general, possibly nonlinear, PDE is an estimate on the approximate solutions, independently of the mesh. For instance, the L^∞ stability of a linear FD scheme for a linear elliptic equation may be obtained by writing the scheme in matrix form and by obtaining a bound of the infinity norm of the inverse of this matrix. Even though the FV approximate solutions are piecewise constant, the estimates are obtained in a norm which is in close relation with the one used for the estimates on the solutions of the continuous problem, and which depends of course on the considered problem; the notion of stability in the FVM is therefore linked to the stability of the continuous problem. Let us take two examples:

- *the heat equation* on $[0, 1]$ with homogeneous Dirichlet conditions: the natural norm for the continuous problem is $L^2(H_0^1)$, and the associated discrete norm corresponding to the choice (17) is $L^2(H_{0,d}^1)$ with

$$\|u\|_{H_{0,d}^1} = \left(\sum_{i=0}^M h_{i+1/2}\left(\frac{u_{i+1} - u_i}{h_{i+1/2}}\right)^2\right)^{1/2},$$

  where M is the number of control volumes and $u_0 = u_{M+1} = 0$ and with $h_{i+1/2} = x_{i+1} - x_i$. We denote by u the weak solution of the heat equation and u_{app} the solution of the time-implicit scheme (implicit schemes are a natural choice for parabolic equations to avoid a condition of type $\delta t \le Ch^2$). The function u_{app} is a piecewise constant function which is equal to u_i^n on the cuboid $]x_{i-1/2}, x_{i+1/2}[\times]t_n, t_{n+1}[$. In the continuous PDE setting, an $L^2(H_0^1)$ estimate on u is obtained by taking u as a test function in the weak formulation of the heat equation, and integrating by parts. Similarly, the $L^2(H_{0,d}^1)$ estimate on u_{app} is obtained by multiplying the i-th discrete equation by $\delta t \, u_i^n$, summing over i and n and performing discrete integrations by parts (obtained by changes of indices in the sums). An $L^2(L^2)$ estimate on the approximate solutions then follows with a discrete Poincaré inequality [5, Lemma 9.1].

- *the transport equation* on $\mathbb{R}^n$, with initial condition $u_{\text{ini}} \in L^\infty$: the natural norm for the continuous problem is L^∞. It is classical and easy to show that the scheme (14), with the upwind flux $F_{i+1/2} = a \, u_i$, is stable, see for example [5, Lemma 20.1], under a CFL [8] condition.

3.2. *The linear case: stability + conservativity + flux consistency $\implies$ convergence*

Convergence for linear operators are often obtained through error estimates (the same technique as in the proof of stability is applied to the error between the approximate solution and the exact one); the compactness analysis of sequences of approximate solutions is another means, which also gives the existence of a solution (see [8] for a seminal paper on this type of proof).

Let us detail this second means, which extends to nonlinear operators. Consider a sequence of approximate solutions, on meshes whose space and time steps tend to 0. The general principle of proof of convergence is then the following. Thanks to the stability, a uniform (with respect to the discretization parameters) estimate on the approximate solutions holds in a Lebesgue space, and so there exists a sub-sequence of this sequence which converges weakly (or $\star$-weakly) in that same Lebesgue space. Each equation of the scheme is then multiplied by the interpolation of a regular test function and by the time step; the summation of all these equations over the time index and space indices, and an integration by parts (using the conservativity of the fluxes) are performed so as to shift the discrete derivatives from the discrete unknowns to the regular test function (the flux consistency which is used on the test functions is thus that given by the dual discrete operator). Since the problem is linear, the weak convergence of the approximate solutions suffices to pass to the limit in all terms; however the parabolic and hyperbolic cases exhibit some different difficulties.

3.2.1. *Elliptic or parabolic case*

The passage to the limit can be performed thanks to the consistency of the flux, see e.g. [5, Theorem 8.1] for the 1D case. In the multidimensional case, the consistency of the flux for the heat equation (Laplace operator) is obtained for meshes which respect an orthogonality condition [5, Definition 9.1]. In this case, the resulting matrix is symmetric and the flux consistency is identical to that given by the dual discrete operator. Anisotropic operators and general meshes have been the object of several different works in the last decades, we refer to [9] for a review of these methods, several of them leading to non-symmetric matrices. In this latter case, one can either prove convergence by an error estimate, using the flux consistency of the primal discrete operator, or a compactness method, using the flux consistency of the dual discrete operator, i.e. on the test functions. In both cases the main difficulty is to establish the stability of the scheme.

3.2.2. *Hyperbolic case*

For stability reasons, the numerical flux is upwinded and introduces an error term whose convergence to 0 must be shown. This term is a sum of products of differences of the values of the solutions in neighbouring cells by discrete derivatives of the test function. On 1D or multiD Cartesian uniform or non-uniform meshes, if $u_{\mathrm{ini}} \in L^{\infty} \cap BV$ (an additional argument is used to handle the case $u_{\mathrm{ini}} \in L^{\infty}$) this term is shown to tend to 0 thanks to a uniform BV estimate on the approximate solutions (and on the continuous solution), see [10, 11] who consider the nonlinear case; this proof uses the "TVD" character (total variation diminishing) of the monotone schemes [12]. Unfortunately, in the case of an unstructured mesh in multiD, even for a linear equation, even if $u_{\mathrm{ini}} \in L^{\infty} \cap BV$, it can be shown that the upwind scheme is not TVD; a counterexample is given in [13]. A "weak BV" estimate based on the numerical diffusion of the scheme is established therein in order to prove convergence.

Let us notice that in both the parabolic and hyperbolic cases, the Lax–Wendroff theorem is not used directly:

- in the parabolic case, because the continuous flux function cannot be applied to the approximate solutions,
- in the hyperbolic case, because the approximate solutions only converge weakly, whereas the Lax–Wendroff theorem supposes a strong convergence.

Nevertheless in both cases, the fundamental notions introduced in [6] are used, namely:

- *flux conservativity*: it is the property that leads to a weak form of the FV scheme by shifting the discrete derivatives of the discrete unknown to the discrete derivatives of the interpolant of the test function,

- *flux consistency*: it is the property that is used to prove the fact that a limit of approximate solutions is a weak solution thanks to the convergence of the discrete derivatives of the interpolant of the test function to the exact derivatives of the same test function.

3.3. *The nonlinear case: more compactness needed*

In the nonlinear case, weak convergence is not sufficient; indeed, if a sequence $(u_n)_{n \in \mathbb{N}}$ converges only weakly in a Lebesgue space to a limit u, there is no reason why the sequence $(f(u_n))_{n \in \mathbb{N}}$ should converge to $f(u)$, even weakly.

3.3.1. *Elliptic and parabolic equations*

In the elliptic or parabolic setting, one of the fundamental tools to obtain more compactness is Kolmogorov's compactness theorem, a consequence of which is that any bounded sequence of L^p, $1 \le p < +\infty$ which is "equicontinuous on average" admits a convergent sub-sequence (see for example [14, Theorem 8.16]). Equicontinuity in mean amounts to showing that the difference between the function and its translates in time and space converges to 0 in the L^p norm, uniformly with respect to the time and space step, see for example [5, Lemma 18.3] in the case of a nonlinear parabolic equation of Stefan type. Once the compactness of the sequence of approximate solutions in $L^2(L^2)$ is proven, we can then exhibit a sub-sequence tending to $\bar{u}$ in $L^2(L^2)$.

By passing to the limit in the "very weak" form of the scheme (that is with the discrete divergence of the discrete normal derivatives of the test functions) we can then show, as in the linear case, that each term (in time and space) converges to the corresponding term in the "very weak" formulation of the continuous problem. In the case of the Stefan problem, namely $\partial_t u - \Delta \varphi(u) = 0$, this gives the convergence of the approximate solutions to the exact solution if φ is an increasing Lipschitz continuous fucntion. There is however an additional difficulty if φ is only non-decreasing. In this case, it is possible to prove compactness in $L^2(L^2)$ of $\varphi(u_{\text{app}})$ (where u_{app} is the approximate solution) but not of u_{app} for which only an $L^2(L^2)$ bound holds. It is possible however to conclude using the Minty trick (see [5, Chapter 4]).

3.3.2. *Hyperbolic equations*

Now consider a nonlinear hyperbolic conservation law of the form

$$\partial_t u(x, t) + \text{div}(f(u(x, t))) = 0, \quad x \in \mathbb{R}^n, \tag{18}$$

$$u(x, 0) = u_{\text{ini}}(x). \tag{19}$$

If $f \in C^1(\mathbb{R}, \mathbb{R}^n)$ and $u_{\text{ini}} \in L^\infty$, there exists a unique entropy solution of this problem [15]. In order to show that a scheme approximates this entropy solution, it is first shown that it satisfies a discrete entropy equation.

The case of Cartesian meshes has been studied independently by Kuznetsov [10] and Crandall and Majda [11]. As in the linear case, if $u_{\text{ini}} \in BV$, a BV estimate on the approximate solutions holds, uniformly with respect to the space and time step; Helly's lemma, which is itself a direct consequence of Kolmogorov's compactness theorem may then be invoked to obtain the convergence of a subsequence of the approximate solutions in $L^1(L^1)$, and one can then use the Lax–Wendroff theorem (which generalizes easily to the entropy formulation) [11], [5, Section 21.5]). It is also possible to handle the case $u_{\text{ini}} \in L^\infty$, using a contraction principle in L^1 for the exact solution and for the approximate solution, see [11] for instance.

In the general case of a non-Cartesian mesh, a suitable BV estimate seems out of reach, and the proof of convergence is performed with the following steps.

- Consider a sequence of space discretizations, indexed by m, and of time steps $\delta t^{(m)}$; for any m, the mesh size is defined as the maximum diameter of the cells and denoted by $h^{(m)}$. Assume that $h^{(m)} \to 0$ and $\delta t^{(m)} \to 0$ as $m \to +\infty$.
- Owing to the L^∞ estimate on the sequence of approximate solutions $(u^{(m)})_{m \in \mathbb{N}}$, there exists a sub-sequence which converges in a "nonlinear weak sense", i.e. there exists $\mu \in L^\infty(\mathbb{R} \times \mathbb{R}_+ \times]0,1[)$ such that for any function $\psi \in C(\mathbb{R}, \mathbb{R})$,

$$\int_{\mathbb{R}} \int_{\mathbb{R}_+} \psi(u^{(m)}) \varphi \to \int_{\mathbb{R}} \int_{\mathbb{R}_+} \mu_\psi \varphi, \forall \varphi \in L^1(L^1) \quad \text{with } \mu_\psi = \int_0^1 \psi\big(\mu(x,t,\alpha)\big) \, d\alpha.$$

  This notion of convergence is equivalent to the convergence to a Young measure [16]; it may seem a little simpler to handle in the sense that it involves a function, μ, rather than a measure; however this function depends on an additional parameter $\alpha \in]0,1[$, which we will have to get rid of in order to reach convergence to an entropy weak solution.
- Using the numerical diffusion of the scheme, we get a uniform weak BV estimate on the sequence of approximate solutions; this estimate is called "weak" for two reasons: on the one hand it involves the differences $|(f(u_K) - f(u_L)) \cdot \boldsymbol{n}_{KL}|$ and not the differences $|u_K - u_L|$, where (K, L) denotes a pair of control volumes sharing a common interface; on the other hand, it only requires that the sum of these differences does not blow up too fast: the difference between the discrete gradient of the interpolated test function and the gradient of the test function itself behaves like the mesh size, and to pass to the limit, we only need that the sum of the differences involving the discrete unknowns be bounded by a term in $C/h^{1-\varepsilon}$ with $\varepsilon > 0$ and $C > 0$ independent of h.
- Using the nonlinear weak convergence and the weak BV estimate, we pass to the limit on the weak form of the discrete entropy and obtain a so-called "process solution" which is an entropy solution up to an integral with respect to the additional parameter α.
- Starting from the discrete entropy inequalities that are verified by the scheme, the process solutions are shown to satisfy an entropy inequality. A uniqueness result on the process solutions can then be obtained thanks to a variable doubling technique "à la Kruskov" [17]; this result differs from Di Perna's [16] in that it takes into account the initial condition in the weak entropy formulation, which allows to avoid the more restrictive conditions on the mesh [18]. The uniqueness of the process solution entails that a process solution is the unique entropy weak solution. It also yields the (strong) convergence, in L^p-spaces, of the approximate solution to the exact solution.

The proof of convergence of the FVM has been obtained for several other problems than the one considered here. However, for systems of PDEs, it is often difficult to obtain compactness results, and LW-consistency then seems an interesting way to make sure that an eventual limit of the scheme is indeed a weak solution of the system. In the next section, we show how this is feasible even on staggered grids, which are often used in the numerical simulation of fluid flows.

4. LW-consistency and staggered grids

Rectangular staggered grids have been used since the sixties in fluid mechanics (in the well-known MAC method [19]) including environmental flows [20], see also [21]. The mathematical analysis of the MAC scheme has been the object of several recent works, see e.g. [22, 23]. Systems of partial differential equations for which no existence or uniqueness is known have also been discretized on staggered grids. Let us mention in particular the compressible Euler equations: one of the advantages of using a staggered grid is to produce a scheme that can be mathematically
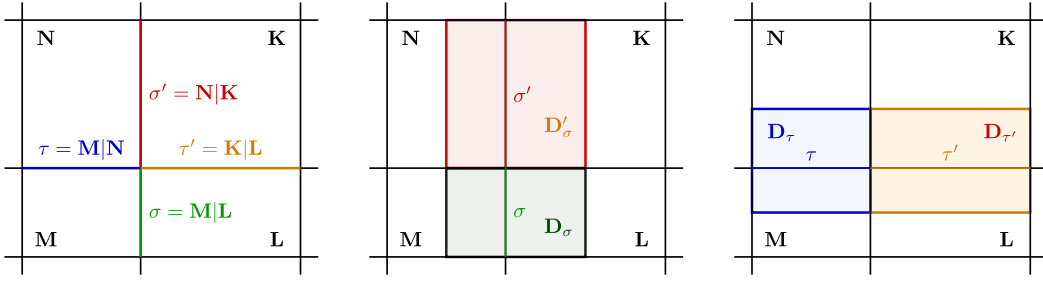
Figure 1. Primal and dual meshes and associated notations for the MAC case. Left: the primal cells; the edges σ and σ' belong to $\mathfrak{F}^{(1)}$ and the edges τ and τ' to $\mathfrak{F}^{(2)}$. Center: the dual cells associated to $\mathfrak{F}^{(1)}$. Right: the dual cells associated to $\mathfrak{F}^{(2)}$.

proven to be asymptotically stable to the incompressible limit, as shown for the isentropic case in [24]. For such systems, a Lax–Wendroff type theorem is very useful; indeed, although true convergence cannot be proven for lack of compactness properties, such a theorem allows to state that if the scheme converges, and provided some bounds on the approximate solutions are satisfied (these bounds are generally not attainable mathematically but can be verified numerically), the limit of the scheme is an entropy weak solution of the system, see [25, 26]. The proof of this result may be obtained thanks to the generalization of the Lax–Wendroff theorem to general grids, which include staggered grids such as the MAC grid [27]. One of the major additional difficulties for staggered grids is that the discrete unknowns are piecewise constant on different grids.

Consider a rectangular domain $\Omega \subset \mathbb{R}^2$ (the 3D case can be tackled in the same way), and a possibly non-uniform rectangular grid. We denote by $\mathfrak{F}$ the set of edges of the mesh, and the internal edge separating the cells K and L is denoted by $\sigma = K|L$ (see Figure 1). This mesh will be referred to in the following as the primal mesh, and denoted by $\mathscr{P}$. Two dual meshes (three in 3D) are considered, each consisting in a partition of Ω indexed by the vertical and horizontal elements of $\mathfrak{F}$, i.e. $\Omega = \cup_{\sigma \in \mathfrak{F}^{(i)}} D_\sigma$, $i = 1, 2$, where $\mathfrak{F}^{(1)}$ (resp. $\mathfrak{F}^{(2)}$) denotes the set of vertical (resp. horizontal) edges. The cells $(D_\sigma)_{\sigma \in \mathfrak{F}}$ are referred to as the dual cells. A half dual cell $D_{K,\sigma}$ is half of the rectangle K with side σ (see Figure 1). For an internal edge $\sigma = K|L$, the dual cell D_σ is the subset of $K \cup L$ defined as $D_\sigma = D_{K,\sigma} \cup D_{L,\sigma}$; for an external edge σ of a cell K, D_σ is the subset $D_{K,\sigma}$ of K.

To illustrate the use of the generalized Lax–Wendroff theorem proven in [27], let us consider as a simple example the mass equation of, say, the compressible Euler equation:

$$\partial_t \rho(\boldsymbol{x}, t) + \operatorname{div}(\rho \boldsymbol{u})(\boldsymbol{x}, t) = 0 \quad (\boldsymbol{x}, t) \in \Omega \times]0, T[, \tag{20}$$

where $\partial_t \rho$ denotes the time derivative of the density ρ, and div the space divergence. The scalar unknown ρ is associated to the primal cells:

$$\rho(\boldsymbol{x}, t) = \rho_K^n \quad \text{for } \boldsymbol{x} \in K, \ K \in \mathscr{P}, \ t \in [t_n, t_{n+1}[, \ n \in [\![0, N-1]\!].$$

The unknowns associated to the i-th component of $\boldsymbol{u}$ are located at the center of the edges of the i-th dual mesh. The associated approximate vector function thus reads: $\boldsymbol{u}(\boldsymbol{x}, t) = (u_1(\boldsymbol{x}, t), u_2(\boldsymbol{x}, t))^t$ where, for $i = 1, 2$,

$$u_i(\boldsymbol{x}, t) = u_\sigma^n, \quad \text{for } \boldsymbol{x} \in D_\sigma, \ \sigma \in \mathfrak{F}^{(i)} \text{ and } t \in [t_n, t_{n+1}[, \ n \in [\![0, N-1]\!].$$

Let $\boldsymbol{e}^{(i)}$ denote the i-th unit vector; the discretization of (20) reads:

$$\mathscr{C}(\rho, \boldsymbol{u})_K^n = (\eth_t \rho)_K^n + \frac{1}{|K|} \sum_{\sigma \in \mathfrak{F}(K)} |\sigma|\, \boldsymbol{F}_\sigma^n \cdot \boldsymbol{n}_{K,\sigma} = 0, \quad \text{with } (\eth_t \rho)_K^n = \frac{\rho_K^{n+1} - \rho_K^n}{t_{n+1} - t_n}$$

$$\text{and } \boldsymbol{F}_\sigma^n = \rho_\sigma^n \boldsymbol{u}_\sigma^n, \text{ where } \boldsymbol{u}_\sigma^n \text{ is defined as } u_\sigma^n\, \boldsymbol{e}^{(i)} \text{ for } \sigma \in \mathfrak{F}^{(i)}, \ i = 1 \text{ or } 2,$$

and, for $\sigma = K|L$, ρ_σ^n stands for a convex combination of ρ_K^n and ρ_L^n (for instance the upwind or a MUSCL choice with respect to u_σ^n). The initial value for the scalar unknown ρ is defined by

$$\rho_K^0 = \frac{1}{|K|} \int_K \rho_0(\boldsymbol{x})\, \mathrm{d}x. \tag{21}$$

For $i = 1, 2$, let $\bar{h}^{(i)} = \max\{|\sigma|,\ \sigma \in \mathfrak{F}^{(i)}\}$ and $\underline{h}^{(i)} = \min\{|\sigma|,\ \sigma \in \mathfrak{F}^{(i)}\}$. We define the space step by $h(\mathscr{P}) = \max(\bar{h}^{(1)}, \bar{h}^{(2)})$, and the time step by $\delta t = \max_{n \in [\![0, N-1]\!]}(t_{n+1} - t_n)$. A sequence of grids is said to be quasi-uniform if the quotients $\bar{h}^{(1)}/\underline{h}^{(2)}$ and $\bar{h}^{(2)}/\underline{h}^{(1)}$ are bounded by a constant, independent of the grid.

Lemma 4.1 (Lax–Wendroff consistency for the mass equation, MAC grid). *Let a sequence of quasi-uniform MAC grids $(\mathscr{P}^{(m)})_{m \in \mathbb{N}}$ and of time discretizations be given, with $h(\mathscr{P}^{(m)})$ and $\delta t^{(m)}$ tending to zero; let $(\rho^{(m)}, \boldsymbol{u}^{(m)})_{m \in \mathbb{N}}$ be the associated sequence of discrete functions.*

We suppose that the sequences $(\rho^{(m)})_{m \in \mathbb{N}}$ and $(\boldsymbol{u}^{(m)})_{m \in \mathbb{N}}$ are bounded in $L^\infty(\Omega \times]0, T[)$ and $L^\infty(\Omega \times]0, T[)^2$ respectively, and that, when m tends to $+\infty$, they converge in $L^p(\Omega \times]0, T[)$ and $L^p(\Omega \times]0, T[)^2$, $1 \le p < +\infty$, to $\rho \in L^\infty(\Omega \times]0, T[)$ and $\boldsymbol{u} \in L^\infty(\Omega \times]0, T[)^2$ respectively. Then $(\rho, \boldsymbol{u})$ is a weak solution of (20), in the sense that, for any function $\varphi \in C_c^\infty(\Omega \times [0, T[)$,

$$\sum_{0 \le n \le N-1} \delta t^{(m)} \sum_{K \in \mathscr{P}^{(m)}} |K|\, \mathscr{C}(\rho, \boldsymbol{u})_K^n \varphi_K^n \rightarrow - \int_\Omega \rho_0(\boldsymbol{x})\, \varphi(\boldsymbol{x}, 0)\, \mathrm{d}x$$

$$- \int_0^T \int_\Omega (\rho(\boldsymbol{x}, t)\partial_t \varphi(\boldsymbol{x}, t) + (\rho \boldsymbol{u})(\boldsymbol{x}, t) \cdot \nabla \varphi(\boldsymbol{x}, t))\, \mathrm{d}x\, \mathrm{d}t \quad as\ m \rightarrow +\infty,$$

where φ_K^n stands for the mean value of $\varphi(\boldsymbol{x}, t_n)$ over the cell K, and so the right-hand side of this assertion vanishes.

The proof of Lemma 4.1 is given in [7] and is quite simple, especially using the *ad hoc* tools developed in [7,27]. However, the developed arguments may be extended to more complex operators, to deal for instance with the momentum and energy balance equations of the compressible Euler equations, even if the proofs are more tricky.

5. Conclusion

In this paper, we have presented some concepts for the mathematical analysis of FV schemes, paying special attention to the flux consistency issue and its most direct consequence, i.e. the LW-consistency of the scheme. We emphasize that we gave here a very partial picture of the mathematical world of finite volumes. In many problems (some of them evoked here), the convergence of the discrete solution to a limit follows by compactness arguments in norms strong enough to "feed" the consistency study; *in fine*, this yields a stronger result than just the scheme consistency, namely the convergence (up to a subsequence if the uniqueness of the solution of the continuous problem is not known) of the numerical solutions to the (a) continuous one (see e.g. [22]). Many parabolic equations enter this framework, including, focusing on fluid flow simulations, incompressible, possibly variable density, or steady barotropic Navier–Stokes equations. In some problems, the continuous solution may be reasonably supposed (or even proven) to be regular, and an error analysis is possible.

The LW-consistency issue is especially important for practical applications in fluid flow simulations. Indeed, in many cases of interest, stronger results are out of reach, and this property is

the only one left to mathematically support the design of schemes. This is for instance the case for multi-dimensional flows governed by hyperbolic systems, as shallow-water equations, Euler equations or models for multi-phase flows. For instance, the study evoked in Section 4 is motivated by such a situation: in the last ten years, a class of staggered schemes has been designed for hyperbolic flow problems [25, 26, 28], and implemented in the open-source software CALIF[3]S developed at IRSN [29]; they are now routinely used for industrial safety applications like hydrogen explosion problems, supposing inviscid or at least vanishing viscosity flows. The accuracy of the numerical schemes involved here is essentially supported by LW-consistency studies [26].

Conflicts of interest

Authors have no conflict of interest to declare.

References

[1] P. Lax, R. Ritchmyer, "Survey of the stability of linear finite difference equations", *Commun. Pure Appl. Math.* **9** (1956), no. 2, p. 267-293.

[2] J. C. Strikwerda, *Finite Difference Schemes and Partial Differential Equations*, 2nd ed., Society for Industrial and Applied Mathematics, Philadelphia, 2004.

[3] R. J. LeVeque, in *Finite Volume Methods for Hyperbolic Problems*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2002.

[4] I. Faille, T. Gallouët, R. Herbin, "Des mathématiciens découvrent les volumes finis", *Matapli* **28** (1991), p. 37-48.

[5] R. Eymard, T. Gallouët, R. Herbin, "Finite volume methods", in *Techniques of Scientific Computing (Part 3)* (J. L. Lions, P. Ciarlet, eds.), Handbook of Numerical Analysis, vol. 7, Elsevier, Amsterdam, 2000, p. 713-1020.

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

[7] T. Gallouët, R. Herbin, J.-C. Latché, "On the weak consistency of finite volume schemes for conservation laws on general meshes", *SeMA J.* **76** (2019), p. 581-594.

[8] R. Courant, K. Friedrichs, H. Lewy, "Über die partiellen Differenzengleichungen der mathematischen Physik", *Math. Ann.* **100** (1928), no. 1, p. 32-74, voir *IBM J. Res. Dev.* **11** (1967), p. 215–234, pour une version en anglais.

[9] J. Droniou, "Finite volume schemes for diffusion equations: introduction to and review of modern methods", *Math. Models Methods Appl. Sci.* **24** (2014), no. 8, p. 1575-1619.

[10] N. N. Kuznecov, S. A. Vološin, "Monotone difference approximations for a first order quasilinear equation", *Dokl. Akad. Nauk SSSR* **229** (1976), no. 6, p. 1317-1320.

[11] M. G. Crandall, A. Majda, "Monotone difference approximations for scalar conservation laws", *Math. Comput.* **34** (1980), no. 149, p. 1-21.

[12] A. Harten, "On a class of high resolution total-variation-stable finite-difference schemes", *SIAM J. Numer. Anal.* **21** (1984), no. 1, p. 1-23, With an appendix by Peter D. Lax.

[13] S. Champier, T. Gallouët, "Convergence d'un schéma décentré amont sur un maillage triangulaire pour un problème hyperbolique linéaire", *Modélisation mathématique et analyse numérique* **26** (1992), no. 7, p. 835-853.

[14] T. Gallouët, R. Herbin, *Mesure, intégration, probabilités*, Ellipses, Paris, 2013, https://hal.archives-ouvertes.fr/cel-00637007v2.

[15] S. N. Kružkov, "First order quasilinear equations with several independent variables", *Math. USSR Sb. (N.S.)* **81** (1970), no. 123, p. 228-255.

[16] R. J. DiPerna, "Measure-valued solutions to conservation laws", *Arch. Ration. Mech. Anal.* **88** (1985), no. 3, p. 223-270.

[17] R. Eymard, T. Gallouët, R. Herbin, "Existence and uniqueness of the entropy solution to a nonlinear hyperbolic equation", *Chin. Ann. Math. Ser. B* **16** (1995), no. 1, p. 1-14, A Chinese summary appears in *Chin. Ann. Math. Ser. A* **16** (1995), no. 1, p. 119.

[18] B. Cockburn, F. Coquel, P. G. LeFloch, "Convergence of the finite volume method for multidimensional conservation laws", *SIAM J. Numer. Anal.* **32** (1995), no. 3, p. 687-705.

[19] F. Harlow, J. Welsh, "Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface", *Phys. Fluids* **8** (1965), p. 2182-2189.

[20] A. Arakawa, V. Lamb, "A potential enstrophy and energy conserving scheme for the shallow water equations", *Mon. Weather Rev.* **109** (1981), p. 18-36.

[21] S. Patankar, *Numerical Heat Transfer and Fluid Flow*, Series in Computational Methods in Mechanics and Thermal Sciences, vol. XIII, Hemisphere Publishing Corporation, McGraw-Hill Book Company, Washington, New York, London, 1980, 197 pages.

[22] T. Gallouët, R. Herbin, J.-C. Latché, K. Mallem, "Convergence of the Marker-And-Cell scheme for the incompressible Navier–Stokes equations on non-uniform grids", *Found. Comput. Math.* **18** (2018), no. 1, p. 249-289.

[23] T. Gallouët, D. Maltese, A. Novotny, "Error estimates for the implicit MAC scheme for the compressible Navier–Stokes equations", *Numer. Math.* **141** (2019), no. 2, p. 495-567.

[24] R. Herbin, J.-C. Latché, K. Saleh, "Low Mach number limit of some staggered schemes for compressible barotropic flows", *Math. Comput.* **90** (2021), no. 329, p. 1039-1087.

[25] R. Herbin, J.-C. Latché, T. Nguyen, "Consistent segregated staggered schemes with explicit steps for the isentropic and full Euler equations", *ESAIM: Math. Model. Numer. Anal.* **52** (2018), no. 3, p. 893-944.

[26] R. Herbin, J.-C. Latché, S. Minjeaud, N. Therme, "Conservativity and weak consistency of a class of staggered finite volume methods for the Euler equations", *Math. Comput.* **90** (2021), no. 329, p. 1155-1177.

[27] T. Gallouët, R. Herbin, J.-C. Latché, "Lax–Wendroff consistency of finite volume schemes for systems of non linear conservation laws: extension to staggered schemes", *SeMA J.* **79** (2022), no. 2, p. 333-354.

[28] R. Herbin, J.-C. Latché, Y. Nasseri, N. Therme, "A consistent quasi–second-order staggered scheme for the two-dimensional shallow water equations", *IMA J. Numer. Anal.* (2021), article no. drab086.

[29] CALIF[3]S, "A software components library for the computation of reactive turbulent flows", https://gforge.irsn.fr/gf/project/isis.