# Finite Volume Methods: Foundation and Analysis

Timothy Barth[1], Raphaèle Herbin[2] and Mario Ohlberger[3]

[1] *NASA Ames Research Center, Moffett Field, CA, USA*
[2] *Aix-Marseille Université, CNRS, Centrale Marseille, Marseille, France*
[3] *Applied Mathematics Münster, CeNoS, and CMTC, University of Münster, Münster, Germany*

ABSTRACT

Finite volume methods are a class of discretization schemes resulting from the decomposition of a problem domain into nonoverlapping control volumes. Degrees of freedom are assigned to each control volume that determine local approximation spaces and quadratures used in the calculation of control volume surface fluxes and interior integrals. An imposition of conservation and balance law statements in each and every control volume constrains surface fluxes and results in a coupled system of equations for the unknown degrees of freedom that must be solved by a numerical method.

Finite volume methods have proved highly successful in approximating the solution to a wide variety of conservation and balance laws. They are extensively used in fluid mechanics, meteorology, electromagnetics, semiconductor device simulation, materials modeling, heat transfer, models of biological processes, and many other engineering problems governed by conservation and balance laws that may be written in integral control volume form.

This chapter reviews elements of the foundation and analysis of modern finite volume methods for approximating hyperbolic, elliptic, and parabolic partial differential equations. These different equations have markedly different continuous problem regularity and function spaces (e.g., $L^\infty$, $L^2$, and $H_0^1$) that must be adequately represented in finite-dimensional discretizations. Particular attention is given to finite volume discretizations yielding numerical solutions that inherit properties of the underlying continuous solutions such as maximum (minimum) principles, total variation control, $L^2$ stability, global entropy decay, and local balance law conservation while also having favorable accuracy and convergence properties on structured and unstructured meshes.

As a starting point, a review of scalar nonlinear hyperbolic conservation laws and the development of high-order accurate schemes for discretizing them is presented. A key tool in the design and analysis of finite volume schemes suitable for discontinuity capturing is discrete maximum principle analysis. A number of mathematical and algorithmic developments used in the construction of numerical schemes possessing local discrete maximum principles are reviewed in one and several space dimensions. These developments include monotone fluxes, TVD discretization, positive coefficient discretization, nonoscillatory reconstruction, slope limiters, strong stability preserving time integrators, and so on. When available, theoretical results concerning *a priori* and *a posteriori* error estimates and convergence to entropy weak solutions are given.

A review of the discretization of elliptic and parabolic problems is then presented. The tools needed

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

for the theoretical analysis of the two point flux approximation scheme for the convection diffusion equation are described. Such schemes require an orthogonality condition on the mesh in order for the numerical fluxes to be consistent. Under this condition, the scheme may be shown to be monotone. A weak formulation of the scheme is derived, which facilitates obtaining stability, convergence, and error estimate results. The discretization of anisotropic problems is then considered and a review is given of some of the numerous schemes that have been designed in recent years, along with their properties. Parabolic problems are then addressed, both in the linear and nonlinear cases.

A discussion of further advanced topics is then given including the extension of the finite volume method to systems of hyperbolic conservation laws. Numerical flux functions based on an exact or approximate solution of the Riemann problem of gas dynamics are discussed. This is followed by the review of another class of numerical flux functions for symmetrizable systems of conservation laws that yield finite volume solutions with provable global decay of the total mathematical entropy for a closed entropy system, often referred to as entropy stability.

Finally, a detailed review of the discretization of the steady-state incompressible Navier–Stokes equations using the Marker-And-Cell (MAC) finite volume method is then presented. The MAC scheme uses a staggered mesh discretization for pressure and velocities on primal and dual control volumes. After reformulating the MAC scheme in weak form, analysis results concerning stability, weak consistency, and convergence are given.

KEY WORDS: finite volume methods, conservation laws, elliptic and parabolic equations, nonoscillatory approximation, discrete maximum principles, higher order schemes

*Encyclopedia of Computational Mechanics*. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Contents



*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

## 1. Introduction

Finite volume methods (FVMs) are a popular class of discretization schemes that are well suited to approximating conservation and balance laws. These laws may yield partial differential equations (PDEs) of different type (hyperbolic, elliptic, or parabolic) as well as coupled systems of equations with individual equations of different type. Consequently, the regularity of the solution to these equations may be quite different from one another and so too the functional spaces in which the solutions of the continuous problems are sought, viz. $L^\infty$, $H_0^1$, $L^2$, and so on. While the piecewise approximation spaces used in FVMs (e.g., piecewise constant spaces for the simplest FVMs) are natural candidates for hyperbolic problems, these approximation spaces are not natural candidates for elliptic problems in $H_0^1$. But as is revealed by mathematical analysis, these piecewise approximation spaces, in particular piecewise constant spaces, are still provably viable candidates for problems in $H_1^0$ when the mesh satisfies certain technical requirements.

A question that is often asked by a non-expert concerns the differences between the finite volume method, finite element method (FEM), and the finite difference method (FDM). The answer truly lies in the concepts of the methods, but in some cases, these methods do yield similar schemes. This similarity may be seen using the simple example $u'' = f$ discretized by all three of the methods using a constant mesh spacing on the unit interval $[0, 1]$. Roughly speaking, one could say that the FEM is based on a weak formulation coupled with a convenient finite-dimensional approximation of the infinite-dimensional function spaces. The FDM relies on an approximation of the differential operators using Taylor expansions. The FVM is constructed from a balance equation, rather than the PDE itself, with a consistent approximation of the fluxes defined on the boundary of the control volume on which the balance equation is written.

Confusion between the FVM and the FDM arises from the fact that the FVM is sometimes called an FDM when the fluxes on the boundary of each control volume are approximated by finite differences. This is sometimes the case, for instance, in oil reservoir simulations utilizing isotropic diffusion models that are discretized on Cartesian grids such that the diffusion flux can

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

be easily approximated by a simple difference quotient. Moreover, numerous schemes have been designed for hyperbolic problems that are called FDMs, although they can also be interpreted as FVMs with suitable approximation of the fluxes at the interfaces of the discretization control volume (also sometimes referred to as a "cell"). Links between the FVM and the finite element method (FEM) can also be found. In certain instances, the FVM can be interpreted as an FEM using a particular integration rule. Conversely, there are instances where the FEM can be interpreted as FVM. For example, the piecewise linear FEM discretization of the Laplace operator on a triangular mesh satisfying the weak Delaunay triangulation condition yields a matrix that is the same as that of the FVM on the dual Voronoï mesh; see Eymard *et al.*, 2000 for details. As another example, the FVM is sometimes presented as a discontinuous Galerkin method (DGM) of lowest order that uses a finite-dimensional approximation of the continuous space that is nonconforming. This is mathematically insightful, but the tools used to analyze DGMs of higher order accuracy do not seem to directly apply to FVMs of higher order accuracy. Other families of FVMs have been developed such as vertex-centered schemes, box or covolume schemes, and finite volume element (FVE) methods that facilitate the compact discretization of various differential operators. Particular attention is given in this chapter to cell centered schemes because these schemes are widely used in industrial codes and are well suited to the discretization of conservation laws of the general form

$$\partial_t u + \nabla \cdot \boldsymbol{f}(u, \nabla u) + s(u) = 0 \tag{1}$$

where $u$ is a function of space and time, $\boldsymbol{f} \in C^1(\mathbb{R}\times\mathbb{R}^d, \mathbb{R}^d)$ is the flux function, $s \in C(\mathbb{R}, \mathbb{R})$ is a low-order source term, and $d$ is the space dimension. This conservation law may be obtained from the following balance equation written for a control volume $K$ with exterior boundary normal $\boldsymbol{n}_K$

$$\int_K \partial_t u \mathrm{d}\boldsymbol{x} + \int_{\partial K} \boldsymbol{f}(u, \nabla u) \cdot \boldsymbol{n}_K \mathrm{d}s + \int_K s(u) \mathrm{d}\boldsymbol{x} = 0 \tag{2}$$

by letting the size of $K$ tend to zero. In the above integrals and in the sequel, $\mathrm{d}\boldsymbol{x}$ represents the integration symbol on a $d$-dimensional subset of $\mathbb{R}^d$ and $\mathrm{d}s$ on a $d-1$-dimensional subset of $\mathbb{R}^d$. Note that conversely, the balance equation (2) may be obtained from the conservation law (1) using integration over a control volume $K$ and applying the Stokes formula. If the control volume $K$ is a polytope (a polygon in 2D or a polyhedron in 3D), then the boundary is the union of faces (or edges in 2D), denoted here by $\sigma$, so that (2) may be written as

$$\int_K \partial_t u \mathrm{d}\boldsymbol{x} + \sum_{\sigma\subset\partial K} \int_\sigma \boldsymbol{f}(u, \nabla u) \cdot \boldsymbol{n}_K \mathrm{d}s + \int_K s(u) \mathrm{d}\boldsymbol{x} = 0 \tag{3}$$

Replacing the continuous time derivative with an explicit Euler time discretization with uniform time step $\delta t$ yields

$$\frac{1}{\delta t}\int_K (u^{n+1} - u^n) \mathrm{d}\boldsymbol{x} + \sum_{\sigma\subset\partial K} \int_\sigma \boldsymbol{f}(u^n, \nabla u^n) \cdot \boldsymbol{n}_K \mathrm{d}s + \int_K s(u^n) \mathrm{d}\boldsymbol{x} = 0$$

where $u^n$ denotes an approximation of $u$ at time $t_n = n\delta t$. For each time $t_n$ and control volume $K$, the discrete unknown $u_K^n$ approximates $u$ in the control volume $K$ at time $t_n = n\delta t$. To obtain the approximate equations needed to solve for $u_K^n$ (which defines the numerical scheme), the flux integrals $\int_\sigma \boldsymbol{f}(u^n, \nabla u^n) \cdot \boldsymbol{n}_K \mathrm{d}s$ must be discretized. Let $F_{K,\sigma}(u^n)$ denote a numerical flux that approximates $\boldsymbol{f}$. A nontrivial task in developing a new FVM scheme is to devise

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

a numerical flux so that properties such as discrete conservation, consistency, accuracy, and convergence (discussed later) are obtained from the resulting discretization. To illustrate the task of devising a numerical flux, consider the linear convection equation that is obtained from (2) by setting $\boldsymbol{f}(u,\nabla u)=\boldsymbol{v}u$, where $\boldsymbol{v}$ is a constant vector of $\mathbb{R}^d$ and $s(u)=0$. The balance equation then reduces to the following simple form

$$\int_K \partial_t u \mathrm{d}\boldsymbol{x} + \int_{\partial K} u\boldsymbol{v}\cdot\boldsymbol{n}_K \mathrm{d}s = 0 \tag{4}$$

In order to approximate the flux $u\boldsymbol{v}\cdot\boldsymbol{n}_K$ on the faces of each control volume, one needs to approximate the value of $u$ on these edges as a function of the discrete unknowns $u_K$ associated to each control volume K. This may be done in several ways. A straightforward choice is to approximate the value of $u$ on the face $\sigma=\sigma_{KL}$ separating the control volumes $K$ and $L$ by the mean value $\frac{1}{2}(u_K+u_L)$. This yields the so-called "centered" numerical flux

$$F_{K,\sigma}^{(cv,c)} = \frac{1}{2} v_{K,\sigma}(u_K+u_L)$$

where $v_{K,\sigma}=\int_\sigma \boldsymbol{v}\cdot\boldsymbol{n}_K \mathrm{d}s$. This centered choice is known to lead to stability problems and is therefore not used in practice. A popular choice is the so-called "upwind" numerical flux given by

$$F_{K,\sigma}^{(cv,c)} = v_{K,\sigma}^{+}u_K - v_{K,\sigma}^{-}u_L$$

where $x^+=\max(x,0)$ and $x^-=-\min(x,0)$. Note that this formula is equivalent to

$$F_{K,\sigma}^{(cv,c)} = \begin{cases} v_{K,\sigma}u_K & \text{if } v_{K,\sigma}\geq 0 \\ \\ v_{K,\sigma}u_L & \text{otherwise} \end{cases}$$

This numerical flux results in schemes satisfying the desired properties mentioned above. A linear convection diffusion reaction balance equation can be obtained from (2) by setting $f(u,\nabla u)=\nabla u+\boldsymbol{v}u$, $\boldsymbol{v}\in\mathbb{R}^d$ and $s(u)=bu$, $b\in\mathbb{R}$,

$$\partial_t u - \Delta u + \mathrm{div}(\boldsymbol{v}u) + bu = 0 \quad \text{ on } \Omega$$

The flux through a given edge is then given by

$$\int_\sigma \boldsymbol{f}(u)\cdot\boldsymbol{n}_{K,\sigma}\mathrm{d}s = \int_\sigma \left(-\nabla u\cdot\boldsymbol{n}_{K,\sigma} + \boldsymbol{v}\cdot\boldsymbol{n}_{K,\sigma}u\right)\mathrm{d}s$$

so that the additional diffusion term $\int_\sigma -\nabla u\cdot\boldsymbol{n}_{K,\sigma}\mathrm{d}s$, involving the normal derivative to the boundary of a control volume, must now be discretized. On a Cartesian grid, a possible simple discretization is obtained using the difference quotient between the value of $u$ in $K$ and an adjacent control volume $L$, that is,

$$F_{K,\sigma}^{(d)} = -\frac{|\sigma|}{d_{KL}}(u_L-u_K) \tag{5}$$

where $|\sigma|$ stands for the $(d-1)$dimensional Lebesgue measure of $\sigma$ (area if $d=3$, length if $d=2$) and $d_{KL}$ is the distance between some (well-chosen) points of $K$ and $L$. The numerical flux is known in the porous media community as the "two point" (TP) flux, and the resulting scheme as the "two point flux approximation" (TPFA) scheme. If the points that are used to compute

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

the distance $d_{KL}$ are carefully chosen, then the resulting diffusion flux (5) is consistent in the finite difference sense (note, however, that the resulting approximation of the second-order diffusion operator may not necessarily be consistent in the finite difference sense). Results presented in Section 5 reveal that TP fluxes for the discretization of the diffusion flux yield accurate results if the mesh satisfies an orthogonality condition. This orthogonality condition requires that there exists a family of points $x_K$, such that for a given interface $\sigma_{KL}$ between the control volumes $K$ and $L$, the line segment $x_K x_L$ is orthogonal to the edge $\sigma_{KL}$. The length $d_{KL}$ is then defined as the distance between $x_K$ and $x_L$. Such a family of points exists, for instance, in the case of triangles, rectangles, or Voronoï meshes, but not for general meshes. For general meshes and for anisotropic diffusion problems, a wide variety of schemes have been introduced in the recent years that are reviewed in Section 5.

The convergence analysis of FVMs is rather recent. Since these methods are currently employed in nonlinear problems where the regularity of the solution is not clear, one would like to obtain theoretical results on the convergence of the scheme without (nonphysical) regularity assumptions on the data or solution. The usual path to the proof of convergence of FVMs, which was initiated in Eymard, Gallouët, and Herbin (1999), Champier *et al.*, 1993, and now been used in a wide number of FVM papers and also adapted to other schemes, is based on establishing the following set of theoretical results:

1. Establish *a priori* estimates on the solution to the scheme in a mesh-dependent norm and deduce the existence of a solution to the scheme.
2. Prove a compactness result.
3. Prove a realistic regularity property of any possible limit.
4. By a passage to the limit in the scheme, prove that any possible limit satisfies a weak form of the original PDE.

Note that a by-product of this approach is an existence proof for the original PDE. Even though the existence is sometimes known before attacking the discretization of the problem, it can be and sometimes has been proved by a numerical approximation technique. This is the case for the now historical result on the existence of the Laplace and biharmonic equations by the convergence of a finite difference approximation in the famous paper Courant *et al.* (1928) (see Courant, Friedrichs and Lewy (1967) for its English version). The next natural question is the rate of convergence of the scheme. Theoretical results on error estimates are often linked to a uniqueness result and are therefore not always accessible.

## 2. Scalar Nonlinear Hyperbolic Conservation Laws

Many problems arising in science and engineering lead to the study of nonlinear hyperbolic conservation laws. Some examples include fluid mechanics, meteorology, electromagnetics, semiconductor device simulation, and numerous models of biological processes. As a prototype conservation law, consider a flux function $\boldsymbol{f}$ depending only on $u$, and the Cauchy initial value

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

problem

$$\partial_t u + \nabla \cdot \boldsymbol{f}(u) \quad = 0 \quad \text{in} \quad \mathbb{R}^d \times \mathbb{R}^+ \tag{6a}$$

$$u(\boldsymbol{x}, 0) \quad = u_0(\boldsymbol{x}) \quad \text{in} \quad \mathbb{R}^d \tag{6b}$$

Here $u(\boldsymbol{x}, t)\colon \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}$ denotes the dependent solution variable, $\boldsymbol{f} \in C^1(\mathbb{R}, \mathbb{R}^d)$ denotes the flux function, and $u_0(\boldsymbol{x})\colon \mathbb{R}^d \to \mathbb{R}$ the initial data.

The function $u$ is a *classical solution* of the scalar initial value problem if $u \in C^1(\mathbb{R}^d \times \mathbb{R}^+)$ satisfies (6) pointwise. An essential feature of nonlinear conservation laws is that, in general, gradients of $u$ blow up in finite time, even when the initial data $u_0$ is arbitrarily smooth. Beyond some critical time $t_0$ classical solutions of (6) do not exist. This behavior will be demonstrated shortly using the method of characteristics. By introducing the notion of weak solutions of (6) together with an entropy condition, it then becomes possible to define a class of solutions where existence and uniqueness are guaranteed for times greater than $t_0$. These are precisely the solutions that are numerically sought in the finite volume method.

### *2.1. The method of characteristics*

Let $u$ be a classical solution of (6). Further, define the vector

$$\boldsymbol{a}(u) = \boldsymbol{f}'(u) = (f_1'(u), \ldots, f_d'(u))^T$$

A characteristic $\Gamma_{\boldsymbol{\xi}}$ is a curve $(\boldsymbol{x}(t), t)$ such that

$$\begin{aligned} \boldsymbol{x}'(t) &= \boldsymbol{a}(u(\boldsymbol{x}(t), t)) \quad \text{for} \quad t > 0 \\ \boldsymbol{x}(0) &= \boldsymbol{\xi} \end{aligned}$$

Since $u$ is assumed to be a classical solution, it is readily verified that

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} u(\boldsymbol{x}(t), t) &= \partial_t u + \boldsymbol{x}'(t) \cdot \nabla u \\ &= \partial_t u + \boldsymbol{a}(u) \cdot \nabla u = \partial_t u + \nabla \cdot \boldsymbol{f}(u) = 0 \end{aligned}$$

Therefore, $u$ is constant along a characteristic curve and $\Gamma_\xi$ is a straight line since

$$\begin{aligned} \boldsymbol{x}'(t) &= \boldsymbol{a}(u(\boldsymbol{x}(t), t)) = \boldsymbol{a}(u(\boldsymbol{x}(0), 0)) \\ &= \boldsymbol{a}(u(\boldsymbol{\xi}, 0)) = \boldsymbol{a}(u_0(\boldsymbol{\xi})) = \text{constant} \end{aligned}$$

In particular, $\boldsymbol{x}(t)$ is given by

$$\boldsymbol{x}(t) = \boldsymbol{\xi} + t\boldsymbol{a}(u_0(\xi)) \tag{7}$$

This important property may be used to construct classical solutions. If $\boldsymbol{x}$ and $t$ are fixed and $\boldsymbol{\xi}$ determined as a solution of (7), then

$$u(\boldsymbol{x}, t) = u_0(\boldsymbol{\xi})$$

This procedure is the basis of the so-called method of characteristics. On the other hand, this construction shows that the intersection of any two straight characteristic lines leads to a contradiction in the definition of $u(\boldsymbol{x}, t)$. Thus, classical solutions can only exist up to the first time $t_0$ at which any two characteristics intersect.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

### 2.2. Weak solutions

Since, in general, classical solutions only exist for a finite time $t_0$, it is necessary to introduce the notion of weak solutions that are well defined for times $t > t_0$.

**Definition 1.** *(Weak solution) Let $u_0 \in L^\infty(\mathbb{R}^d)$. Then, $u$ is a weak solution of (6) if $u \in L^\infty(\mathbb{R}^d \times \mathbb{R}^+)$ and (6) holds in the distributional sense, that is,*

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^+} (u\partial_t\phi + \boldsymbol{f}(u)\cdot\nabla\phi)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{x} + \int_{\mathbb{R}^d} u_0\phi(\boldsymbol{x},0)\,\mathrm{d}\boldsymbol{x} = 0 \quad \textit{for all} \quad \phi \in C_0^1(\mathbb{R}^d\times\mathbb{R}^+) \tag{8}$$

Note that classical solutions are weak solutions and weak solutions that lie in $C^1(\mathbb{R}^d \times \mathbb{R}^+)$ satisfy (6) in the classical sense.

It can be shown (Kruzkov, 1970; Oleinik, 1963) that there always exists at least one weak solution to (6) if the flux function $f$ is at least Lipschitz continuous. Nevertheless, the class of weak solutions is too large to ensure uniqueness of solutions. An important class of solutions are piecewise classical solutions with discontinuities separating the smooth regions. The following lemma gives a necessary and sufficient condition imposed on these discontinuities such that the solution is a weak solution; see, for example, Godlewski and Raviart (1991) and Kröner (1997). Later a simple example is given where infinitely many weak solutions exist.

**Lemma 1.** *(Rankine–Hugoniot jump condition) Assume that the space-time domain $\mathbb{R}^d \times \mathbb{R}^+$ is separated by a smooth hypersurface $S$ into two parts $Q_l$ and $Q_r$. Furthermore, assume $u$ is a $C^1$-function on $Q_l$ and $Q_r$, respectively. Then, $u$ is a weak solution of (6) if and only if the following two conditions hold:*

1. *$u$ is a classical solution in $Q_l$ and $Q_r$.*
2. *$u$ satisfies the Rankine–Hugoniot jump condition, that is,*

$$[u]s = [\boldsymbol{f}(u)]\cdot\boldsymbol{n} \quad \textit{on} \quad S \tag{9}$$

*Here, $(\boldsymbol{n}, -s)^T$ denotes a unit normal vector for the (space-time) hypersurface $S$ and $[u]$ denotes the jump in $u$ across the hypersurface $S$.*

In one space dimension (i.e., $\boldsymbol{f} = f$ is a scalar function), it may be assumed that $S$ is parameterized by $(\sigma(t), t)$ such that $s = \sigma'(t)$ and $\boldsymbol{n} = 1$. The Rankine–Hugoniot jump condition then reduces to

$$s = \frac{[f(u)]}{[u]} \quad \text{on} \quad S \tag{10}$$

**Example 1.** *(Non-uniqueness of weak solutions)* Consider the one-dimensional Burgers' equation, $f(u) = u^2/2$, with Riemann data: $u_0(x) = u_l$ for $x < 0$ and $u_0(x) = u_r$ for $x \geq 0$. Then, for any $a \geq \max(u_l, -u_r)$ a function $u$ given by

$$u(x,t) = \begin{cases} u_l, & x < s_1 t \\ -a, & s_1 t < x < 0 \\ a, & 0 < x < s_2 t \\ u_r, & s_2 t < x \end{cases} \tag{11}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

is a weak solution if $s_1 = (u_l - a)/2$ and $s_2 = (a + u_r)/2$. This is easily checked since $u$ is piecewise constant and satisfies the Rankine–Hugoniot jump condition. This elucidates a one-parameter family of weak solutions. In fact, there is also a classical solution whenever $u_l \leq u_r$. In this case, the characteristics do not intersect and the method of characteristics yields the classical solution

$$u(x,t) = \begin{cases} u_l, & x < u_l t \\ x/t, & u_l t < x < u_r t \\ u_r, & u_r t < x \end{cases} \tag{12}$$

This solution is the unique classical solution but not the unique weak solution. Consequently, additional conditions must be introduced in order to single out one solution within the class of weak solutions. These additional conditions give rise to the notion of a unique entropy weak solution.

### *2.3. Entropy weak solutions and vanishing viscosity*

In order to introduce the notion of entropy weak solutions, it is useful to first demonstrate that there is a class of additional conservation laws for any classical solution of (6). Let $u$ be a classical solution and $\eta\colon \mathbb{R} \to \mathbb{R}$ a smooth function. Multiplying (6a) by $\eta'(u)$, one obtains

$$0 = \eta'(u)\partial_t u + \eta'(u)\nabla \cdot \boldsymbol{f}(u) = \partial_t \eta(u) + \nabla \cdot \boldsymbol{F}(u) \tag{13}$$

where $\boldsymbol{F}$ is any primitive of $\eta' \boldsymbol{f}'$. This reveals that for a classical solution $u$, the quantity $\eta(u)$, henceforth called an entropy function, is a conserved quantity.

**Definition 2.** *(Entropy–entropy flux pair) Let $\eta\colon \mathbb{R} \to \mathbb{R}$ be a smooth convex function and $F\colon \mathbb{R} \to \mathbb{R}^d$ a smooth function such that*

$$\boldsymbol{F}' = \eta' \boldsymbol{f}' \tag{14}$$

*in (13). Then $(\eta, \boldsymbol{F})$ is called an entropy–entropy flux pair or more simply an entropy pair for the equation (6a).*

**Note 1.** *(Kruzkov entropies) The family of smooth convex entropies $\eta$ may be equivalently replaced by the nonsmooth family of the so-called Kruzkov entropies, that is, $\eta_\kappa(u) \equiv |u - \kappa|$ for all $\kappa \in \mathbb{R}$. The associated entropy flux is then $\boldsymbol{F}_\kappa(u) = (\boldsymbol{F}(u) - \boldsymbol{F}(\kappa))\mathrm{sg}(u - \kappa)$, where* sg *denotes the sign function (see e.g., Kröner, 1997).*

Unfortunately, the relation (13) cannot be fulfilled for weak solutions in general, as it would lead to additional jump conditions that would contradict the Rankine–Hugoniot jump condition lemma. Rather, a weak solution may satisfy the relation (13) in the distributional sense with inequality. To see that this concept of entropy effectively selects a unique, physically relevant solution among all weak solutions, consider the viscosity-perturbed equation

$$\partial_t u_\epsilon + \nabla \cdot \boldsymbol{f}(u_\epsilon) = \epsilon \Delta u_\epsilon \tag{15}$$

with $\epsilon > 0$. For this parabolic problem, it may be assumed that a unique smooth solution $u_\epsilon$ exists. Multiplying by $\eta'$ and rearranging terms yields the additional equation

$$\partial_t \eta(u_\epsilon) + \nabla \cdot \boldsymbol{F}(u_\epsilon) = \epsilon \Delta \eta(u_\epsilon) - \epsilon \eta''(u_\epsilon)|\nabla u|^2$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Furthermore, since $\eta$ is assumed convex ($\eta'' \geq 0$), the following inequality is obtained

$$\partial_t \eta(u_\epsilon) + \nabla \cdot \boldsymbol{F}(u_\epsilon) \leq \epsilon \Delta \eta(u_\epsilon)$$

Taking the limit $\epsilon \to 0$ establishes (Málek, Nečas, Rokyta and Røužička, 1996) that $u_\epsilon$ converges toward some $u$ a.e. in $\mathbb{R}^d \times \mathbb{R}^+$ where $u$ is a weak solution of (6) and satisfies the entropy condition

$$\partial_t \eta(u) + \nabla \cdot \boldsymbol{F}(u) \leq 0 \tag{16}$$

in the sense of distributions on $\mathbb{R}^d \times \mathbb{R}^+$.

By this procedure, a unique weak solution has been identified as the limit of the approximating sequence $u_\epsilon$. The obtained solution $u$ is called the vanishing viscosity weak solution of (6). Motivated by the entropy inequality (16) of the vanishing viscosity solution, it is now possible to introduce the notion of entropy weak solutions. This notion is weak enough for the existence and strong enough for the uniqueness of solutions to (6).

**Definition 3.** *(Entropy weak solution) Let $u$ be a weak solution of* (6)*. Then, $u$ is called an entropy weak solution if $u$ satisfies for all entropy pairs $(\eta, \boldsymbol{F})$*

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^+} (\eta(u)\partial_t \phi + \boldsymbol{F}(u) \cdot \nabla \phi) \, \mathrm{d}t \, \mathrm{d}\boldsymbol{x} + \int_{\mathbb{R}^d} \eta(u_0)\phi(\boldsymbol{x}, 0) \, \mathrm{d}\boldsymbol{x} \geq 0 \tag{17}$$

*for all $\phi \in C_0^1(\mathbb{R}^d \times \mathbb{R}^+, \mathbb{R}^+)$.*

From the vanishing viscosity method, it is known that entropy weak solutions exist. The following $L^1$ contraction principle guarantees that entropy solutions are uniquely defined; see Kruzkov (1970).

**Theorem 1.** *($L^1$-contraction principle) Let $u$ and $v$ be two entropy weak solutions of* (6) *with respect to initial data $u_0$ and $v_0$. Then, the following $L^1$-contraction principle holds*

$$\|u(\cdot, t) - v(\cdot, t)\|_{L^1(\mathbb{R}^d)} \leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} \tag{18}$$

*for almost every $t > 0$.*

This principle demonstrates a continuous dependence of the solution on the initial data and consequently the uniqueness of entropy weak solutions. Finally, note that an analog of the Rankine–Hugoniot condition exists (with inequality) in terms of the entropy pair for all entropy weak solutions

$$[\eta(u)]s \geq [\boldsymbol{F}(u)] \cdot \boldsymbol{n} \quad \text{on} \quad S \tag{19}$$

### *2.4. Measure-valued or entropy process solutions*

The numerical analysis of conservation laws is facilitated by an even weaker formulation of solutions to (6). For instance, the convergence analysis of finite volume schemes makes it necessary to introduce the so-called measure-valued or entropy process solutions; see DiPerna (1985) and Eymard *et al.* (2000).

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

**Definition 4.** *(Entropy process solution) A function* $\mu(x,t,\alpha) \in L^\infty(\mathbb{R}^d \times \mathbb{R}^+ \times (0,1))$ *is called an entropy process solution of* (6) *if* $u$ *satisfies for all entropy pairs* $(\eta, F)$

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^+} \int_0^1 \eta(\mu) \partial_t \left(\phi + \boldsymbol{F}(\mu) \cdot \nabla \phi\right) \, \mathrm{d}\alpha \, \mathrm{d}t \, \mathrm{d}\boldsymbol{x} + \int_{\mathbb{R}^d} \eta(u_0)\phi(x,0) \, \mathrm{d}\boldsymbol{x} \geq 0$$

*for all* $\phi \in C_0^1(\mathbb{R}^d \times \mathbb{R}^+, \mathbb{R}^+)$.

The most important property of such entropy process solutions is the following uniqueness and regularity result (see Eymard *et al.*, 2000, Theorem 6.3).

**Theorem 2.** *(Uniqueness of entropy process solutions) Let* $u_0 \in L^\infty(\mathbb{R}^d)$ *and* $f \in C^1(\mathbb{R})$. *The entropy process solution* $\mu$ *of problem* (6) *is unique. Moreover, there exists a function* $u \in L^\infty(\mathbb{R}^d \times \mathbb{R}^+)$ *such that* $u(x,t) = \mu(x,t,\alpha)$ *a.e. for* $(x,t,\alpha) \in \mathbb{R}^d \times \mathbb{R}^+ \times (0,1)$ *and* $u$ *is the unique entropy weak solution of* (6).

## 3. Finite Volume Methods for Nonlinear Hyperbolic Conservation Laws

In the FVM for hyperbolic conservation laws, the computational domain, $\Omega \subset \mathbb{R}^d$, is first tessellated into a collection of nonoverlapping control volumes that completely cover the domain. Notationally, let $\mathcal{T}$ denote a tessellation of the domain $\Omega$ with control volumes $K \in \mathcal{T}$ such that $\cup_{K \in \mathcal{T}} \overline{K} = \overline{\Omega}$. Let $h_K$ denote a length scale associated with each control volume $K$, for example, $h_K \equiv \mathrm{diam}(K)$. For two distinct control volumes $K$ and $L$ in $\mathcal{T}$, the intersection is either an oriented edge (2-D) or face (3-D) denoted by $\sigma_{K,L}$ with oriented normal $\boldsymbol{n}_{K,L}$ or else a set of measure at most $d-2$. For simplicity, it is assumed throughout that control volumes $K$ are time invariant (unchanging in time). As mentioned in the introduction (see (2)), for each control volume, an *integral conservation law* statement holds.

**Definition 5.** *(Integral conservation law) An integral conservation law asserts that the rate of change of the total amount of a substance with density* $u$ *in a time-invariant control volume* $K$ *is equal to the total flux of the substance through the boundary* $\partial K$

$$\frac{\mathrm{d}}{\mathrm{d}t} \int_K u \, \mathrm{d}\boldsymbol{x} + \int_{\partial K} \boldsymbol{f}(u) \cdot \boldsymbol{n} \, \mathrm{d}s = 0 \tag{20}$$

This integral conservation law statement is readily obtained upon spatial integration of the divergence equation (6a) in the region $K$ and application of the divergence theorem. The choice of control volume tessellation is flexible in the FVM. For example, Figure 1 depicts a 2-D triangle complex and two typical control volume tessellations (among many others) used in the FVM. In the *cell-centered* FVM shown in Figure 1(a), the triangles themselves serve as control volumes with solution unknowns (degrees of freedom) stored on a per triangle basis. In the *vertex-centered* FVM shown in Figure 1(b), control volumes are formed as a geometric dual to the triangle complex and solution unknowns stored on a per triangulation vertex basis.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

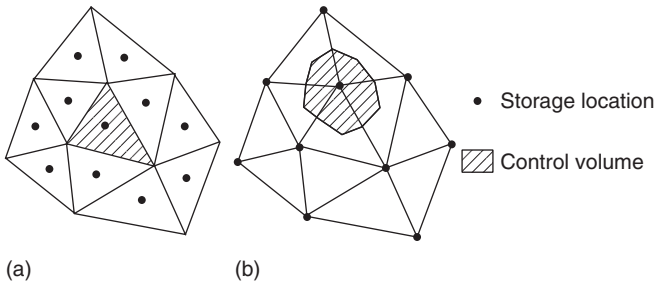


Figure 1. Control volume variants used in the finite volume method: (a) cell-centered and (b) vertex-centered control volume tessellation.

### *3.1. Finite volume discretization from cell averages via exact or approximate Riemann problems*

An important class of FVMs for hyperbolic problems comes from the introduction of control volume cell averages

$$u_K(t) \equiv \frac{1}{|K|}\int_K u(\boldsymbol{x},t)\,\mathrm{d}\boldsymbol{x}, \quad \forall K \in \mathcal{T} \tag{21}$$

as unknowns in the numerical solution procedure. Recall that for simplicity, the control volume $K$ is assumed to be unchanging in time unless otherwise stated. The FVM can then be interpreted as producing evolution equations for approximations of cell average unknowns

$$\frac{\mathrm{d}}{\mathrm{d}t}|K|\,u_K + \int_{\partial K} \boldsymbol{f}(u)\cdot\boldsymbol{n}\,\mathrm{d}s = 0, \quad \forall K \in \mathcal{T} \tag{22}$$

if the flux integral can somehow be evaluated. In the context of the gas dynamic equations, Godunov (1959) pursued an interpretation of the cell averages as a piecewise constant representation of the numerical solution. This interpretation readily applies to scalar hyperbolic conservation laws as well. The piecewise constant representation renders the numerical solution multivalued at control volume interfaces. It then becomes unclear how the flux integral appearing in (22) should be interpreted so that discrete conservation is maintained. Discrete conservation in the finite volume method demands that for two control volumes $K$ and $L$ that share an interface $\sigma_{K,L}$, the amount of substance that fluxes out of $K$ through $\sigma_{K,L}$ must exactly equal the amount of substance that fluxes into $L$ through $\sigma_{L,K}$. Godunov solved the problems of multivalued solution states and discrete conservation by finding a single solution state $u^*(u_K,u_L)$ with symmetry $u^*(u_K,u_L) = u^*(u_L,u_K)$ so that discrete conservation is automatically satisfied

$$\int_{\sigma_{K,L}} \boldsymbol{f}(u^*(u_K,u_L))\cdot\boldsymbol{n}\,\mathrm{d}s = -\int_{\sigma_{L,K}} \boldsymbol{f}(u^*(u_L,u_K))\cdot\boldsymbol{n}\,\mathrm{d}s \tag{23}$$

In Godunov's original work, this single solution state was obtained by solving the one-dimensional Riemann problem of gas dynamics. For the scalar hyperbolic problem (6) this amounts to finding the state $u^*(u_K,u_L)$ equal to the Riemann state $u^R(u_K,u_L;\boldsymbol{n}_{K,L})$ described in the following definition.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

**Definition 6.** *(Scalar Riemann state) Given two solution states $u$ and $v$ and normal vector $\boldsymbol{n}$, let $h_{\boldsymbol{n}}(u) \equiv \boldsymbol{f}(u)\cdot\boldsymbol{n}$ denote the flux in* (6) *projected in the direction $\boldsymbol{n}$. Calculate the Riemann state $u^R(u,v;\boldsymbol{n})$ by first solving the one-dimensional Riemann problem for $w(\xi,\tau;u,v)\colon \mathbb{R}\times\mathbb{R}^+\mapsto\mathbb{R}$*

$$\frac{\partial}{\partial\tau}w+\frac{\partial}{\partial\xi}h_{\boldsymbol{n}}(w)=0 \tag{24}$$

*subject to initial data*

$$w(\xi,0;u,v)=\begin{cases}u & \text{if } \xi<0\\ v & \text{if } \xi>0\end{cases} \tag{25}$$

*Next, determine the desired Riemann state from the Riemann problem solution $w(\xi,\tau;u,v)$ by evaluating it at $\xi=0$ for any positive time, that is,*

$$u^R(u,v;\boldsymbol{n})=w(0,\tau>0;u,v) \tag{26}$$

*Owing to the self-similarity of the Riemann problem in the single parameter $\xi/\tau$, this state value is independent of time for $\tau>0$.*

More generally, given two solution states $u$ and $v$ that share an interface with normal $\boldsymbol{n}$, variants of the Godunov approach are obtained by supplanting the true flux at this interface by a *numerical flux function* $g(u,v;\boldsymbol{n})\colon\mathbb{R}\times\mathbb{R}\mapsto\mathbb{R}$, a Lipschitz continuous function of the two interface states $u$ and $v$ for a given normal $\boldsymbol{n}$. In higher space dimensions, the flux integral appearing in (20) is then approximated by

$$\int_{\partial K}\boldsymbol{f}(u)\cdot\boldsymbol{n}\,\mathrm{d}s\approx\sum_{\sigma_{K,L}\subset\partial K}g(u_K,u_L;\boldsymbol{n}_{K,L})|\sigma_{K,L}| \tag{27}$$

The numerical flux is assumed to satisfy the properties:

- *Conservation*: This property ensures that fluxes from adjacent control volumes sharing a mutual interface exactly cancel when summed. This is achieved if the numerical flux satisfies the identity
$$g(u,v;\boldsymbol{n})=-g(v,u;-\boldsymbol{n}) \tag{28a}$$
- *Consistency*: Consistency is obtained if the numerical flux with identical state arguments reduces to the true flux of that same state projected in the direction $\boldsymbol{n}$, that is,
$$g(u,u;\boldsymbol{n})=\boldsymbol{f}(u)\cdot\boldsymbol{n} \tag{28b}$$

Combining (22) and (27) yields the following semi-discrete FVM shown here for a time invariant (fixed) mesh.

**Definition 7.** *(Semi-discrete FVM for hyperbolic problems) The semi-discrete finite volume approximation of the hyperbolic conservation law problem* (6) *using cell averaged solution unknowns (21) with continuous in time derivatives (22) and numerical flux function quadrature in space (27) for time-invariant control volumes $K$ is given by*

$$\frac{\mathrm{d}}{\mathrm{d}t}u_K+\frac{1}{|K|}\sum_{\sigma_{K,L}\subset\partial K}g(u_K,u_L;\boldsymbol{n}_{K,L})\,|\sigma_{K,L}|=0\quad\forall K\in\mathcal{T} \tag{29}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

This system of ordinary differential equations can be marched forward in time using a variety of explicit and implicit time integration methods; see Section 4.4.

Let $u_K^n$ denote a numerical approximation of the cell average solution in the control volume $K$ at time $t^n \equiv n\Delta t$. A particularly simple time integration method is the forward Euler scheme

$$\frac{\mathrm{d}}{\mathrm{d}t}u_K \approx \frac{u_K^{n+1} - u_K^n}{\Delta t} \tag{30}$$

which can be viewed as resulting from a piecewise constant representation of the solution in time. Using the explicit time advancement formula together with the flux quadrature (27) yields the following fully discrete finite volume formulation:

**Definition 8.** *(Fully discrete FVM for hyperbolic problems) The fully discrete finite volume approximation of the hyperbolic conservation law problem* (6) *using Euler explicit time advancement (30) and numerical flux function quadrature in space (27) for time-invariant control volumes $K$ is given by*

$$u_K^{n+1} = u_K^n - \frac{\Delta t}{|K|}\sum_{\sigma_{K,L}\subset\partial K} g(u_K^n, u_L^n; \boldsymbol{n}_{K,L})\,|\sigma_{K,L}| \quad \forall K \in \mathcal{T} \tag{31}$$

Unfortunately, the numerical flux conditions (28a) and (28b) are insufficient to guarantee that stable numerical solutions will be produced that converge to entropy satisfying weak solutions (17). Consequently, additional numerical flux restrictions are necessary. In the following section, an early result concerning monotone schemes is presented that addresses the question of convergence to entropy weak solutions. These results motivate the construction of monotone flux and E-flux functions that guarantee a local maximum principle and global maximum norm stability.

### *3.2. Monotone fluxes and E-flux functions*

An early result by Harten, Hyman and Lax (1976) addresses the question of convergence of the fully discrete one-dimensional finite volume scheme to weak entropy satisfying solutions. For ease of notation in describing stencil operators, the shorthand notation $u_j^n \equiv u_{K_j}^n$ has been adopted.

**Theorem 3.** *(Monotone schemes and weak solutions) Consider a 1-D finite volume discretization of* (6) *with $2k+1$ stencil on a uniformly spaced mesh in both time and space with corresponding mesh spacing parameters $\Delta t$ and $\Delta x$*

$$\begin{aligned} u_j^{n+1} &= H_j(u_{j+k}^n, \ldots, u_j^n, \ldots, u_{j-k}^n) \\ &= u_j^n - \frac{\Delta t}{\Delta x}(g_{j+1/2}^n - g_{j-1/2}^n) \end{aligned} \tag{32}$$

*and consistent numerical flux of the form*

$$g_{j+1/2} = g(u_{j+k}, \ldots, u_{j+1}, u_j, \ldots, u_{j-k+1}) \tag{33}$$

*that is monotone in the sense*

$$\frac{\partial H_j}{\partial u_{j+l}} \geq 0, \quad \forall\, |l| \leq k \tag{34}$$

*Assume that* $u_j^n$ *converges boundedly almost everywhere to some function* $u(x,t)$*; then as* $\Delta t$ *and* $\Delta x$ *tend to zero with* $\Delta t/\Delta x = \text{constant}$*, this limit function* $u(x,t)$ *is an entropy satisfying weak solution of* (6)*.*

Note that this theorem assumes convergence in the limit, which was later proved to be the case in multidimensions by Crandall and Majda (1980).

The monotone scheme conditions (34) may be achieved by the use of Lipschitz continuous *monotone fluxes* satisfying the following conditions

$$\frac{\partial g_{j+1/2}}{\partial u_l} \quad \geq 0 \quad \text{if} \quad l = j \tag{35a}$$

$$\frac{\partial g_{j+1/2}}{\partial u_l} \quad \leq 0 \quad \text{if} \quad l \neq j \tag{35b}$$

together with a CFL (Courant–Friedrichs–Lewy) like condition

$$1 - \frac{\Delta t}{\Delta x}\left(\frac{\partial g_{j+1/2}}{\partial u_j} - \frac{\partial g_{j-1/2}}{\partial u_j}\right) \geq 0 \tag{36}$$

These monotone flux functions are considered further in the following section for the semi-discrete FVM (29) and the fully discrete FVM (31). Example monotone flux functions are then presented.

*3.2.1. Monotone flux functions.* The numerical flux function (33) accommodates large mesh stencils unlike the numerical flux functions considered in Section 3.1 that only utilize two states. Given two states $u$ and $v$ that share an interface with normal $\boldsymbol{n}$, the flux monotonicity conditions (35a) and (35b) reduce to

$$\frac{\partial g(u,v;\boldsymbol{n})}{\partial u} \quad \geq 0 \tag{37a}$$

$$\frac{\partial g(u,v;\boldsymbol{n})}{\partial v} \quad \leq 0 \tag{37b}$$

Some examples of two state monotone fluxes for the hyperbolic problem (6a) with convex flux, $f'' > 0$, are

- (Riemann flux), see Section 3.1

$$g^{\mathrm{R}}(u,v;\boldsymbol{n}) = \begin{cases} \min\limits_{w\in[u,v]} \boldsymbol{f}(w)\cdot\boldsymbol{n} & \text{if } u < v\ , \\ \max\limits_{w\in[u,v]} \boldsymbol{f}(w)\cdot\boldsymbol{n} & \text{if } u > v\ . \end{cases} \tag{38}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

- ("local" Lax–Friedrichs flux), see Shu and Osher (1988)

$$g^{\mathrm{LF}}(u,v;\boldsymbol{n}) = \frac{1}{2}(\boldsymbol{f}(u)+\boldsymbol{f}(v))\cdot\boldsymbol{n} - \frac{1}{2}\sup_{w\in[u,v]}|\boldsymbol{f}'(w)\cdot\boldsymbol{n}|(v-u) \tag{39}$$

- (Upwind flux with sonic point modification), see Godlewski and Raviart (1991)

$$g^{\mathrm{upwind}}(u,v;\boldsymbol{n}) = \begin{cases} \boldsymbol{f}(u)\cdot\boldsymbol{n} & \text{if } \boldsymbol{f}'(w)\cdot\boldsymbol{n}\geq 0, \forall w\in[u,v]\ , \\ \boldsymbol{f}(v)\cdot\boldsymbol{n} & \text{if } \boldsymbol{f}'(w)\cdot\boldsymbol{n}\leq 0, \forall w\in[u,v]\ , \\ g^{\mathrm{LF}}(u,v;\boldsymbol{n}) & \text{otherwise .} \end{cases} \tag{40}$$

Note that this flux is not monotone in the usual sense introduced above, since the function $g^{\mathrm{upwind}}$ is not continuous at the sonic point. However, the notion of monotone flux can be extended to noncontinuous functions in the following way:

**Definition 9 (Monotone numerical flux.)** *Eymard et al., 2016* Let $f$ be a locally Lipschitz continuous function on $\mathbb{R}$, and let $A, B \in \mathbb{R}$ such that $A \leq B$. A function $g : \mathbb{R}^2 \to \mathbb{R}$ is said to be a monotone numerical flux for $f$ on $[A, B]$ if it satisfies the following assumptions:

- There exists $L_g \in \mathbb{R}_+$ such that

$$\forall (a,b)\in(A,B)^2, \quad \begin{cases} |g(a,b)-f(a)| \leq L_g|a-b| \\ |g(b,a)-f(a)| \leq L_g|a-b| \end{cases} \tag{41}$$

- $g(s,s) = f(s)$, for all $s \in [A,B]$,
- the function $g : (a,b) \mapsto g(a,b)$, from $[A,B]^2$ to $\mathbb{R}$, is nondecreasing with respect to $a$ and nonincreasing with respect to $b$.

The flux $g^{\mathrm{upwind}}(\cdot,\cdot;\boldsymbol{n})$ defined by (40) can be shown to be monotone for the function $f = \boldsymbol{f}\cdot\boldsymbol{n}$. It is then easy to show that proofs of convergence that are written for Lipschitz continuous monotone fluxes are also valid for monotone fluxes in the sense of Definition 41.

Monotone flux functions have played an enormously important role in the development of FVMs for hyperbolic problems but can sometimes be difficult to construct or evaluate. In the following section, a slightly weaker condition called the E-flux condition is presented. Monotone fluxes satisfy this E-flux condition.

*3.2.2. E-flux functions.* Another class of numerical fluxes arising frequently in analysis and practical implementations was introduced by Osher (1984). These fluxes are called E-fluxes due to the relationship to Oleinik's well-known E-condition, which characterizes entropy satisfying discontinuities. E-fluxes satisfy the inequality

$$\frac{g^{\mathrm{E}}(u,v;\boldsymbol{n}) - \boldsymbol{f}(w)\cdot\boldsymbol{n}}{v-u} \leq 0, \quad \forall w\in[u,v] \tag{42}$$

E-fluxes can be characterized by their relationship to the Riemann flux. Specifically, E-fluxes are those fluxes such that

$$g^{\mathrm{E}}(u,v;\boldsymbol{n}) \leq g^{\mathrm{R}}(u,v;\boldsymbol{n}) \quad \text{if } v<u \tag{43a}$$

$$g^{\mathrm{E}}(u,v;\boldsymbol{n}) \geq g^{\mathrm{R}}(u,v;\boldsymbol{n}) \quad \text{if } v>u \tag{43b}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Viewed another way, note that any numerical flux can be written in the form

$$g(u,v;\boldsymbol{n}) = \frac{1}{2}(\boldsymbol{f}(u)+\boldsymbol{f}(v))\cdot\boldsymbol{n} - \frac{1}{2}Q(u,v;\boldsymbol{n})(v-u) \tag{44}$$

where $Q(\cdot)$ denotes a viscosity for the scheme. When written in this form, E-fluxes are those fluxes that contribute at least as much viscosity as the Riemann flux, that is,

$$Q^{\mathrm{R}}(u,v;\boldsymbol{n}) \leq Q^{\mathrm{E}}(u,v;\boldsymbol{n}) \tag{45}$$

The most prominent E-flux is the Enquist–Osher flux

$$g^{\mathrm{EO}}(u,v;\boldsymbol{n}) = \frac{1}{2}\left(\boldsymbol{f}(u)+\boldsymbol{f}(v)\right)\cdot\boldsymbol{n} - \frac{1}{2}\int_u^v |\boldsymbol{f}'(w)\cdot\boldsymbol{n}|\,\mathrm{d}w \tag{46}$$

which was generalized to systems of conservation laws in Osher and Solomon (1982), see Sect. 6.1.2.

Monotone flux and E-flux functions provide the needed properties for proving discrete maximum principles in the FVM. These results are presented in the following section. Section 6.1.3 shows that the E-flux condition (42) has a natural extension to systems of symmetrizable hyperbolic conservation laws and plays an important role in proving energy/entropy stability of those systems using the FVM.

*3.2.3. Discrete maximum principles and stability using monotone flux and E-flux functions.* A compelling motivation for the use of monotone flux and E-flux functions in the FVMs (29) and (31) is to obtain discrete maximum principles in the resulting numerical solutions. A standard analysis technique is to first construct local discrete maximum principles that can then be applied successively to obtain global maximum principles and maximum norm stability results.

The following two lemmas concern the boundedness of local extrema and a discrete maximum principle for FVMs that can be written in nonnegative coefficient form. The first lemma addresses the evolution of local extrema using the semi-discrete finite volume discretization (29) when rewritten in nonnegative coefficient form.

**Lemma 2.** *(Semi-discrete LED property) The semi-discrete scheme for each* $K \in \mathcal{T}$

$$\frac{\mathrm{d}u_K}{\mathrm{d}t} = \frac{1}{|K|}\sum_{\sigma_{K,L}\subset\partial K} C_{K,L}(u_h)(u_L - u_K) \tag{47}$$

*with* $u_h \equiv \{u_{K_1}, u_{K_2}, \ldots\}$ *is local extremum diminishing (LED), that is, local maxima are nonincreasing and local minima are nondecreasing, if*

$$C_{K,L}(u_h) \geq 0, \quad \forall \sigma_{K,L} \subset \partial K \tag{48}$$

The second lemma addresses a local maximum principle using the fully discrete finite volume discretization (31) when rewritten in nonnegative coefficient form.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

**Lemma 3.** *(Fully discrete local maximum principle) The fully discrete scheme for the time slab increment* $[t^n, t^{n+1}]$ *and each* $K \in \mathcal{T}$

$$u_K^{n+1} = u_K^n + \frac{\Delta t}{|K|} \sum_{\sigma_{K,L} \subset \partial K} C_{K,L}(u_h^n)(u_L^n - u_K^n) \tag{49}$$

*with* $u_h \equiv \{u_{K_1}, u_{K_2}, \ldots\}$ *exhibits a local discrete maximum principle for each* $n = 0, 1, 2, \ldots$

$$\min_{\sigma_{K,L} \subset \partial K} (u_L^n, u_K^n) \leq u_K^{n+1} \leq \max_{\sigma_{K,L} \subset \partial K} (u_L^n, u_K^n) \tag{50}$$

*if*

$$C_{K,L}(u_h^n) \geq 0, \quad \forall \sigma_{K,L} \subset \partial K \tag{51}$$

*and* $\Delta t$ *is chosen such that the CFL-like condition is satisfied*

$$1 - \frac{\Delta t}{|K|} \sum_{\sigma_{K,L} \subset \partial K} C_{K,L}(u_h^n) \geq 0 \tag{52}$$

The results of lemmas 2 and 3 require showing that the semi-discrete finite volume scheme (29) and fully discrete finite volume scheme (31) can be placed in nonnegative coefficient form. This can be easily shown when monotone flux and E-flux functions are used by rewriting the flux divergence terms in the following equivalent forms

$$\sum_{\sigma_{K,L} \subset \partial K} g(u_K, u_L; \boldsymbol{n}_{K,L}) \, |\sigma_{K,L}| = \sum_{\sigma_{K,L} \subset \partial K} C_{K,L}(u_h) \, (u_L - u_K) \tag{53a}$$

$$= \sum_{\sigma_{K,L} \subset \partial K} \frac{g(u_K, u_L; \boldsymbol{n}_{K,L}) - \boldsymbol{f}(u_K) \cdot \boldsymbol{n}_{K,L}}{u_L - u_K} \, |\sigma_{K,L}| \, (u_L - u_K) \tag{53b}$$

$$= \sum_{\sigma_{K,L} \subset \partial K} \frac{\partial g(u_K, \tilde{u}_{K,L})}{\partial u_K} \, |\sigma_{K,L}| \, (u_L - u_K) \tag{53c}$$

for appropriately chosen $\tilde{u}_{K,L} \in [u_K, u_L]$. Nonnegativity of the coefficients $C_{K,L}$ in the (53a) right-hand-side summands is achieved in the (53b) right-hand-side summands whenever the numerical flux satisfies the E-flux condition (42) and similarly in (53c) whenever the numerical flux satisfies the monotonicity conditions (37). These results are then used to make a statement concerning stability in a maximum norm.

**Theorem 4.** *($L^\infty$-stability) The fully discrete finite volume scheme (31) utilizing either the monotone flux of Section 3.2.1 or the E-flux functions of Section 3.2.2 subject to a local CFL-like condition as given in lemma 3 for each time slab increment* $[t^n, t^{n+1}]$ *is* $L^\infty$*-stable in the following sense*

$$\inf_{\boldsymbol{x} \in \mathbb{R}^d} u_0(\boldsymbol{x}) \leq u_K^n \leq \sup_{\boldsymbol{x} \in \mathbb{R}^d} u_0(\boldsymbol{x}) \tag{54}$$

*for all* $K \in \mathcal{T}$ *and time step* $t^n$, $n = 0, 1, \ldots$.

The LED and local maximum principles discussed in lemmas 2 and 3 preclude the introduction of spurious extrema and $\mathcal{O}(1)$ Gibbs-like oscillations that occur near solution

discontinuities computed using many numerical methods (even in the presence of grid refinement). For this reason, FVMs for hyperbolic conservation laws that possess these LED and discrete maximum principle properties have proved highly successful in practical calculations.

### *3.3. Stability, convergence, and error estimates*

Several stability results have been presented in Section 3.2.3 that originate from discrete maximum principle analysis and are straightforwardly stated in the multidimensional setting and on general unstructured meshes. In presenting results concerning convergence and error estimates, a notable difference arises between one and several space dimensions. This is due to the lack of a BV bound on the approximate solution in the multidimensional setting, except in the case of Cartesian meshes.

*3.3.1. Convergence results.* The $L^\infty$-stability bound (54) is an essential ingredient in the proof of convergence of the fully discrete finite volume scheme (31). This bound permits the extraction of a subsequence that converges against some limit in the $L^\infty$ weak-starred sense. The primary task that then remains is to identify this limit with the unique solution of the problem. To this end stronger estimates are needed, both for convergence and for deriving convergence rates.

Let BV denote the space of functions with bounded variation, that is,

$$\mathrm{BV} = \left\{ g \in L^1(\mathbb{R}^d) \middle| |g|_{\mathrm{BV}} < \infty \right\}$$

with

$$|g|_{\mathrm{BV}} = \sup_{\substack{\varphi \in C_c^1(\mathbb{R}^d)^d \\ \|\varphi\|_\infty \leq 1}} \int_{\mathbb{R}^d} g \nabla \cdot \varphi \, \mathrm{d}\boldsymbol{x}$$

From the theory of scalar conservation laws, it is known that, provided the initial data is in BV, the solution remains in BV for all times. Therefore, it is desirable to have an analog of this property for the approximate solution as well. Unfortunately, it is known that finite volume schemes for hyperbolic problems on unstructured meshes may be non-total variation diminishing (TVD), and in fact, the counterexample of Després, 2004 shows that the approximate solutions of an upwind finite volume scheme on triangles may blow up in the BV norm. However, the approximate finite volume solutions can be shown to fulfill a weaker estimate, called a *weak BV estimate*; see Champier and Gallouët (1992), Champier *et al.* (1993), Vila (1994), Cockburn, Coquel and Lefloch (1994), and Eymard *et al.* (1998).

**Theorem 5.** *(Weak BV estimate) Let $\mathcal{T}$ be a regular triangulation, and let $J$ be a uniform partition of $[0,\tau]$, for example, $\Delta t^n \equiv \Delta t$. Assume that there exists some $\alpha > 0$ such that $\alpha h^2 \leq |K|$, $\alpha |\partial K| \leq h$. Let $g$ be a monotone flux function, that is, a Lipschitz-continuous function satisfying* (35)*. Assume the following CFL-like condition for a given $\xi \in (0,1)$*

$$\Delta t \leq \frac{(1-\xi)\alpha^2 h}{L_g}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

*where $L_g$ is the Lipschitz constant of the numerical flux function $g$. Furthermore, let $u_0 \in L^\infty(\mathbb{R}^d) \cap BV(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Then, the numerical solution of the fully discrete solution (31) fulfills the following estimate*

$$\sum_n \Delta t \sum_{\sigma_{KL} \in \mathcal{E}} \chi_{KL} h |u_K^n - u_L^n| Q_{KL}(u_K^n, u_L^n) \leq C\sqrt{T|B_{R+h}(0)|}\sqrt{h} \tag{55}$$

*where $C$ only depends on $\alpha$, $L_g$, $\xi$, and the initial function $u_0$. In this formula $Q_{KL}$ is defined as*

$$Q_{KL}(u,v) \equiv \frac{2g_{KL}(u,v) - g_{KL}(u,u) - g_{KL}(v,v)}{u-v}$$

*and $\chi_{KL}$ denotes the discrete cutoff function on $B_R(0) \subset \mathbb{R}^d$, that is,*

$$\chi_{KL} = \begin{cases} 1 & \text{if } (K \cup L) \cap B_R(0) \neq \emptyset \ , \\ 0 & \text{else} \end{cases}$$

Note that in the case of a strong BV estimate (obtained in the case of Cartesian meshes), the right-hand side of (55) would be $\mathcal{O}(h)$ instead of $\mathcal{O}(\sqrt{h})$. In the case of linear advection problems, a somewhat stronger estimate was proven in Després and Lagoutière, 2010, which implies the above weak BV estimate.

Another important property of monotone finite volume schemes is that they preserve the $L^1$-contraction property (Theorem 1).

**Theorem 6.** *($L^1$-contraction property and Lipschitz estimate in time) Let $u_h, v_h \in V_h^0$ be the approximate monotone finite volume solutions corresponding to initial data $u_0$, $v_0$ assuming that the CFL-like condition for stability has been fulfilled. Then the following discrete $L^1$-contraction property holds*

$$\|u_h(\cdot, t+\tau) - v_h(\cdot, t+\tau)\|_{L^1(\mathbb{R}^d)} \leq \|u_h(\cdot,t) - v_h(\cdot,t)\|_{L^1(\mathbb{R}^d)}$$

*Furthermore, a discrete Lipschitz estimate in time is obtained*

$$\sum_{K \in \mathcal{T}} |K||u_K^{n+1} - u_K^n| \leq L_g \Delta t^n \sum_{K \in \mathcal{T}} \sum_{\sigma_{KL} \in \mathcal{E}_K} |\sigma_{KL}||u_K^0 - u_L^0|$$

The principal ingredients of the convergence theory for scalar nonlinear conservation laws are compactness of the family of approximate solutions and the passage to the limit within the entropy inequality (17). In dealing with nonlinear equations, strong compactness is needed in order to pass to the limit in (17). In one space dimension or in the case of multidimensional Cartesian meshes, due to the BV estimate and the selection principle of Helly, strong compactness is ensured and the passage to the limit is summarized in the well-known Lax–Wendroff theorem; see Lax and Wendroff (1960).

**Theorem 7.** *(Lax–Wendroff theorem) Let $(u_m)_{m \in \mathbb{N}}$ be a sequence of discrete solutions defined by the finite volume scheme in one space dimension with respect to initial data $u_0$. Assume that $(u_m)_{m \in \mathbb{N}}$ is uniformly bounded with respect to $m$ in $L^\infty$ and $u_m$ converges almost everywhere in $\mathbb{R} \times \mathbb{R}^+$ to some function $u$. Then $u$ is the uniquely defined entropy weak solution of* (6).

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

With the lack of a BV estimate for the approximate solution in multiple space dimensions, one cannot expect a passage to the limit of the nonlinear terms in the entropy inequality in the classical sense. Nevertheless, the weak compactness obtained by the $L^\infty$-estimate is enough to obtain a measure-valued or entropy process solution in the limit. The convergence of a subsequence of approximate solutions to the entropy weak solution is then obtained thanks to the uniqueness result of Eymard *et al.*, 1995. Note that this approach has been adapted for a constrained hyperbolic equation in Andreianov *et al.*, 2011.

The key theorem for this convergence result is the following compactness theorem of Eymard *et al.* (2000), which is a convenient rewriting of a fundamental theorem on Young measures for PDEs; see Tartar (1979) and Ball (1989).

**Theorem 8.** *Let $(u_m)_{m\in\mathbb{N}}$ be a family of bounded functions in $L^\infty(\mathbb{R}^d)$. Then, there exists a subsequence $(u_m)_{m\in\mathbb{N}}$, and a function $u \in L^\infty(\mathbb{R}^d \times (0,1))$ such that for all functions $g \in C(\mathbb{R})$ the weak-$\star$ limit of $g(u_m)$ exists and*

$$\lim_{m\to\infty} \int_{\mathbb{R}^d} g(u_m(\boldsymbol{x}))\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_0^1 \int_{\mathbb{R}^d} g(u(\boldsymbol{x},\alpha))\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\,\mathrm{d}\alpha, \quad \textit{for all} \;\; \phi \in L^1(\mathbb{R}^d) \tag{56}$$

In order to prove the convergence of a FVM, it now remains to be shown that the residual of the entropy inequality (17) for the approximate solution $u_h$ tends to zero if $h$ and $\Delta t$ tend to zero. Before presenting this estimate for the finite volume approximation, a general convergence theorem is given, which can be viewed as a generalization of the classical Lax–Wendroff result; see Eymard *et al.* (2000).

**Theorem 9.** *(Sufficient condition for convergence) Let $u_0 \in L^\infty(\mathbb{R}^d)$ and $\boldsymbol{f} \in C^1(\mathbb{R})$. Further, let $(u_m)_{m\in\mathbb{N}}$ be any family of uniformly bounded functions in $L^\infty(\mathbb{R}^d \times \mathbb{R}^+)$ that satisfies the following estimate for the residual of the entropy inequality using the class of Kruzkov entropy pairs $(\eta_\kappa, \boldsymbol{F}_\kappa)$ (see Note 1).*

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^+} (\eta_\kappa(u_m)\partial_t\phi + \boldsymbol{F}_\kappa(u_m)\cdot\nabla\phi)\,\mathrm{d}t\,\mathrm{d}\boldsymbol{x} + \int_{\mathbb{R}^d} \eta_\kappa(u_0)\phi(x,0)\,\mathrm{d}\boldsymbol{x} \geq -R(\kappa,u_m,\phi) \tag{57}$$

*for all $\kappa \in \mathbb{R}$ and $\phi \in C_0^1(\mathbb{R}^d \times \mathbb{R}^+, \mathbb{R}^+)$ where the residual $R(\kappa,u_m,\phi)$ tends to zero for $m \to \infty$ uniformly in $\kappa$. Then, $u_m$ converges strongly to the unique entropy weak solution of (6) in $L^p_{\mathrm{loc}}(\mathbb{R}^d \times \mathbb{R}^+)$ for all $p \in [1,\infty)$.*

**Theorem 10.** *(Estimate on the residual of the entropy inequality) Let $(u_m)_{m\in\mathbb{N}}$ be a sequence of monotone finite volume approximations satisfying a local CFL-like condition as given in (52) such that $h, \Delta t$ tend to zero for $m \to \infty$. Then, there exist measures $\mu_m \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^+)$ and $\nu_m \in \mathcal{M}(\mathbb{R}^d)$ such that the residual $R(\kappa,u_m,\phi)$ of the entropy inequality is estimated by*

$$R(\kappa,u_m,\phi) \leq \int_{\mathbb{R}^d}\int_{\mathbb{R}^+} (|\partial_t\phi(\boldsymbol{x},t)| + |\nabla\phi(\boldsymbol{x},t)|)\,\mathrm{d}\mu_m(\boldsymbol{x},t) + \int_{\mathbb{R}^d} \phi(\boldsymbol{x},0)\,\mathrm{d}\nu_m(\boldsymbol{x})$$

*for all $\kappa \in \mathbb{R}$ and $\phi \in C_0^1(\mathbb{R}^d \times \mathbb{R}^+, \mathbb{R}^+)$. The measures $\mu_m$ and $\nu_m$ satisfy the following properties:*

1. *For all compact subsets $\Omega \subset \mathbb{R}^d \times \mathbb{R}^+$, $\lim_{m\to\infty} \mu_m(\Omega) = 0$.*

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

*2. For all $g \in C_0(\mathbb{R}^d)$ the measure $\nu_m$ is given by $\langle \nu_m, g\rangle = \int_{\mathbb{R}^d} g(\boldsymbol{x})|u_0(\boldsymbol{x}) - u_m(\boldsymbol{x},0)|\,\mathrm{d}\boldsymbol{x}$.*

These theorems are sufficient for establishing convergence of monotone finite volume schemes.

**Corollary 1.** *(Convergence theorem) Let $(u_m)_{m\in\mathbb{N}}$ be a sequence of monotone finite volume approximations satisfying the assumptions of Theorem 10. Then, $u_m$ converges strongly to the unique entropy weak solution of (6) in $L^p_{\mathrm{loc}}(\mathbb{R}^d \times \mathbb{R}^+)$ for all $p \in [1,\infty)$.*

Convergence of higher order finite volume schemes can also be proved within the given framework as long as they are $L^\infty$-stable and allow for an estimate on the entropy residual in the sense of Theorem 10; for details see Kröner, Noelle and Rokyta (1995) and Chainais-Hillairet (2000).

*3.3.2. Error estimates and convergence rates.* There are two primary approaches taken to obtaining error estimates for approximations of scalar nonlinear conservation laws. One approach is based on the ideas of Oleinik and is applicable only in one space dimension; see Oleinik (1963) and Tadmor (1991). The second approach, which is widely used in the numerical analysis of conservation laws, is based on the doubling of variables technique of Kruzkov; see Kruzkov (1970) and Kuznetsov (1976). In essence, this technique enables one to estimate the error between the exact and approximate solution of a conservation law in terms of the entropy residual $R(\kappa, u_m, \Phi)$ introduced in (57). Thus, an *a posteriori* error estimate is obtained. Using *a priori* estimates of the approximate solution (see Section 3.2.3, and Theorems 5, 6), a convergence rate or an *a priori* error estimate is then obtained. The following theorem gives a fundamental error estimate for conservation laws independent of the particular finite volume scheme; see Eymard *et al.* (1998), Eymard *et al.* (2000), Chainais-Hillairet (1999), and Kröner and Ohlberger (2000).

**Theorem 11.** *(Fundamental error estimate) Let $u_0 \in BV(\mathbb{R}^d)$ and let $u$ be an entropy weak solution of (6). Furthermore, let $v \in L^\infty(\mathbb{R}^d \times \mathbb{R}^+)$ be a solution of the following entropy inequalities with residual term $R$:*

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^+} (\eta_\kappa(v)\partial_t\phi + F_\kappa(v)\cdot\nabla\phi)\,\mathrm{d}\boldsymbol{x} + \int_{\mathbb{R}^d} \eta_\kappa(u_0)\phi(\cdot,0)\mathrm{d}\boldsymbol{x} \geq -R(\kappa, v, \phi) \tag{58}$$

*for all $\kappa \in \mathbb{R}$ and $\phi \in C_0^1(\mathbb{R}^d \times \mathbb{R}^+, \mathbb{R}^+)$. Suppose that there exist measures $\mu_v \in \mathcal{M}(\mathbb{R}^d \times \mathbb{R}^+)$ and $\nu_v \in \mathcal{M}(\mathbb{R}^d)$ such that $R(\kappa, v, \phi)$ can be estimated independently of $\kappa$ by*

$$R(\kappa, v, \phi) \leq \langle |\partial_t\phi| + |\nabla\phi|, \mu_v\rangle + \langle |\phi(\cdot,0)|, \nu_v\rangle \tag{59}$$

*Let $G \subset\subset \mathbb{R}^d \times \mathbb{R}^+$, $\omega \equiv Lip(f)$, and choose $T$, $R$ and $\boldsymbol{x}_0$ such that $T \in ]0, (R/\omega)[$ and $G$ lies within its cone of dependence $D_0$, that is, $G \subset D_0$ where $D_\delta$ is given as*

$$D_\delta := \bigcup_{0\leq t\leq T} B_{R-\omega t+\delta}(\boldsymbol{x}_0)\times\{t\} \tag{60}$$

*Then, there exists a $\delta \geq 0$ and positive constants $C_1$, $C_2$ such that $u$, $v$ satisfy the following error estimate*

$$\|u - v\|_{L^1(G)} \quad \leq T\left(\nu_v(B_{R+\delta}(x_0)) + C_1\mu_v(D_\delta) + C_2\sqrt{\mu_v(D_\delta)}\right) \tag{61}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

This estimate can be used either as an *a posteriori* control of the error, as the right-hand side of the estimate (61) only depends on $v$, or it can be used as an *a priori* error bound if one is able to estimate further the measures $\mu_v$ and $\nu_v$ using some *a priori* bounds on $v$. Finally, note that comparable estimates to (61) are obtainable in an $L^\infty(0,T;L^1(\mathbb{R}^d))$-norm; see Cockburn and Gau (1995) and Bouchut and Perthame (1998).

*3.3.3. A posteriori error estimate.* Based on the fundamental error estimate in Theorem 11, the following theorem states an *a posteriori* error estimate that can be used to design self-adaptive variants of finite volume schemes; see Kröner and Ohlberger (2000) and the review article Ohlberger (2009).

**Theorem 12.** *(A posteriori error estimate) Assume the conditions and notations as in Theorem 11. Let $v = u_h$ be a numerical approximation to* (6) *obtained from a monotone finite volume scheme that satisfies a local CFL-like condition as given in (52). Then the following error estimate holds*

$$\int_G |u - u_h| \mathrm{d}\boldsymbol{x} \leq T\big(\|u_0 - u_h(\cdot,0)\|_{L^1(B_{R+h}(x_0))} + C_1\eta + C_2\sqrt{\eta}\big) \tag{62}$$

*where*

$$\begin{aligned} \eta &\equiv \sum_{n\in I_0}\sum_{K\in M(t^n)} |u_K^{n+1} - u_K^n| \Delta t^n h_K^d \\ &+ 2\sum_{n\in I_0}\sum_{\sigma_{KL}\in\mathcal{E}(t^n)} \Delta t^n(\Delta t^n + h_{KL}) \times Q_{KL}(u_K^n, u_L^n)|u_K^n - u_L^n| \end{aligned} \tag{63}$$

*with*

$$Q_{KL}(u,v) \equiv \frac{2g_{KL}(u,v) - g_{KL}(u,u) - g_{KL}(v,v)}{u - v} \tag{64}$$

*and the index sets $I_0, M(t), E(t)$ are given by*

$$\begin{aligned} I_0 &\equiv \left\{ n \,|\, 0 \leq t^n \leq \min\left\{ \frac{R+\delta}{\omega}, T \right\} \right\}, \\ M(t) &\equiv \{K \,|\, \textit{there exists} \quad \boldsymbol{x} \in K \ \textit{such that} \ (\boldsymbol{x},t) \in D_{R+\delta}\}, \\ E(t) &\equiv \{\sigma_{KL} \,|\, \textit{there exists} \quad \boldsymbol{x} \in K \cup L \ \textit{such that} \quad (\boldsymbol{x},t) \in D_{R+\delta}\} \end{aligned}$$

*Furthermore, the constants $C_1$, $C_2$ only depend on $T$, $\omega$, $\|u_0\|_{BV}$ and $\|u_0\|_{L^\infty}$.*

Note that this *a posteriori* error estimate is local, since the error on a compact set $K$ is estimated by discrete quantities that are supported in the cone of dependence $D_{R+\delta}$. Similar results have been obtained for conservation laws on bounded domains Ohlberger and Vovelle (2006) and for higher order discontinuous Galerkin generalizations Dedner *et al.* (2007).

*3.3.4. A priori error estimate.* Using the weak BV estimate (Theorem 5) and the Lipschitz estimate in time (Theorem 6), the right-hand side of the *a posteriori* error estimate (62) can

be further estimated. This yields an *a priori* error estimate as stated in the following theorem; for details, see Cockburn and Gremaud (1996, 1997, 1998a), Eymard *et al.* (1998), Chainais-Hillairet (1999), and Eymard *et al.* (2000).

**Theorem 13.** *(A priori error estimate) Assume the conditions and notations as in Theorem 11 and let $v = u_h$ be the approximation to* (6) *given by a monotone finite volume scheme that satisfies a local CFL-like condition as given in (52). Then there exists $C \geq 0$ such that*

$$\int_G |u - u_h| \, \mathrm{d}\boldsymbol{x} \leq C h^{1/4} \tag{65}$$

*Moreover, in the one-dimensional case or Cartesian multidimensional case, the optimal convergence rate of $h^{1/2}$ is obtained.*

## 4. Higher Order Accurate Finite Volume Methods for Hyperbolic Problems

The FVMs for hyperbolic problems described in Section 3 are a powerful methodology that has been used extensively in practical calculations. The error estimates of Section 3.3 predict a deterioration from first-order accuracy for general problems and this deterioration can be observed in some practical calculations. Nevertheless, for smooth solutions on smoothly varying meshes, first-order accuracy is routinely observed. For problems that require very high solution accuracy, first-order accuracy may still require a prohibitively large number of finite volume cells. This has motivated the development of higher order accurate FVMs that require far fewer finite volume cells while still providing a nonoscillatory resolution of discontinuities and steep solution gradients. Unfortunately, Godunov (1959) has shown that all *linear* schemes that preserve solution monotonicity are at most first-order accurate. Thus, higher order accurate methods must utilize essential *nonlinearity* so that non-oscillatory resolution of discontinuities and high-order accuracy away from discontinuities are simultaneously attained. These methods are described below using both structured and unstructured meshes.

### *4.1. Higher order accurate finite volume methods for hyperbolic problems in one dimension*

A significant step forward in the generalization of FVMs to higher order accuracy for hyperbolic problems is due to van Leer (1979). In this work, van Leer generalized Godunov's method by employing linear solution *reconstruction* in each cell from given cell averages (Figures 2b). This reconstruction yields a piecewise linear representation of the solution while still retaining cell averages equal to the given data. The concept extends to higher order polynomials such as depicted in Figure 2(c) using piecewise quadratic reconstruction; see Colella and Woodward (1984). A close examination of Figure 2(b) reveals that this particular piecewise linear reconstruction contains more local extrema than the underlying data. These spurious extrema (oscillations) can seriously degrade the accuracy of a numerical solution in both space and time. One approach for removing these spurious extrema during the reconstruction process

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

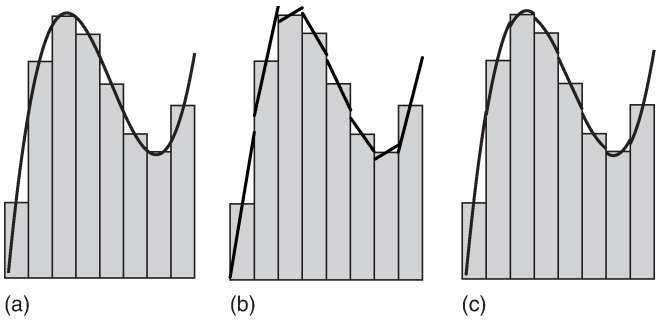


Figure 2. Piecewise polynomial approximation used in the finite volume method (a) cell averaging of analytic data, (b) piecewise linear reconstruction from cell averages, and (c) piecewise quadratic reconstruction from cell averages.

is to alter the slope of the reconstructed data in some or all of the cells. Another approach is based on the idea of finding the smoothest piecewise polynomial reconstructions among many possible candidates. The underlying mathematical principles and techniques used in the development of higher order accurate finite volume approximations that control spurious extrema are important technical contributions that are discussed in the remainder of this section.

*4.1.1. Total variation diminishing (TVD) finite volume methods.* In considering the scalar nonlinear conservation law (6) in one space dimension and time, Lax (1973) made the following basic observation:

> *the total increasing and decreasing variations of a differentiable solution between any pair of characteristics are conserved.*

Furthermore, in the presence of shock wave discontinuities, information is lost and the total variation *decreases*. For the 1-D nonlinear conservation law with compactly supported or periodic solution data $u(x,t)$, integrating along the constant time spatial coordinate at times $t_1$ and $t_2$ yields

$$\int_{-\infty}^{\infty} |\,\mathrm{d}u(x,t_2)| \leq \int_{-\infty}^{\infty} |\,\mathrm{d}u(x,t_1)|, \quad t_2 \geq t_1 \tag{66}$$

This total variation property for scalar conservation laws motivated Harten (1983a) to consider the discrete total variation in the design of numerical methods. Using a one-dimensional mesh (assumed periodic here) with solution data $u_h \equiv \{u_1, u_2, \ldots, u_N\}$ at cell centroids of the intervals $[x_{j-1/2}, x_{j+1/2}]$, $j = 1, \ldots, N$, Harten considered the discrete total variations

$$\mathrm{TV}(u_h) \equiv \sum_j |\Delta_{j+1/2} u_h|, \quad \Delta_{j+1/2} u_h \equiv u_{j+1} - u_j$$

*Encyclopedia of Computational Mechanics*. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

and imposed the discrete total variation nonincreasing (TVNI) bound counterpart to (66)

$$\mathrm{TV}(u_h^{n+1}) \leq \mathrm{TV}(u_h^n) \tag{67}$$

in the design of numerical discretizations for nonlinear conservation laws. A number of theoretical results relating TVNI schemes and monotone schemes follow from analysis.

**Theorem 14.** *(TVNI and monotone scheme properties, Harten, 1983a) (i) Monotone schemes are TVNI. (ii) TVNI schemes are monotonicity preserving, that is, the number of solution extrema is preserved in time.*

Using the notion of discrete total variation, Harten (1983a) then constructed sufficient algebraic conditions for achieving the TVNI inequality (67) in a fully discrete numerical method.

**Theorem 15.** *(Harten's explicit TVD criteria) The fully discrete 1-D discretization*

$$u_j^{n+1} = u_j^n + \Delta t \left( C_{j+1/2}(u_h^n) \Delta_{j+1/2} u_h^n + D_{j+1/2}(u_h^n) \Delta_{j-1/2} u_h^n \right), \quad j = 1, \ldots, N \tag{68}$$

*is TVNI if for each j*

$$C_{j+1/2} \geq 0 \tag{69a}$$

$$D_{j+1/2} \leq 0 \tag{69b}$$

$$1 - \Delta t \left( C_{j-1/2} - D_{j+1/2} \right) \geq 0 \tag{69c}$$

Note that although the inequality constraints (69) in Theorem 15 insure that the total variation is nonincreasing, these conditions are often referred to as total variation diminishing (TVD) conditions. Also note that inequality (69c) implies a CFL-like time step restriction that may be different from the time step required for stability of the numerical method. The TVD conditions are easily generalized to wider support stencils written in incremental form; see, for example, Jameson and Lax (1986) and their corrected result in Jameson and Lax (1987).

**Theorem 16.** *(Generalized explicit TVD criteria) The fully discrete explicit 1-D scheme*

$$u_j^{n+1} = u_j^n + \Delta t \sum_{l=-k}^{k-1} C_{j+1/2}^{(l)}(u_h^n) \Delta_{j+l+1/2} u_h^n, \ j = 1, \ldots, N \tag{70}$$

*with integer stencil width parameter $k > 0$ is TVNI if for each j*

$$C_{j+1/2}^{(k-1)} \geq 0 \tag{71a}$$

$$C_{j+1/2}^{(-k)} \leq 0 \tag{71b}$$

$$C_{j+1/2}^{(l-1)} - C_{j-1/2}^{(l)} \geq 0, \ -k+1 \leq l \leq k-1, \ l \neq 0 \tag{71c}$$

$$1 - \Delta t \left( C_{j-1/2}^{(0)} - C_{j+1/2}^{(-1)} \right) \geq 0 \tag{71d}$$

While this simple Euler explicit time integration scheme may seem inadequate for applications requiring true high-order space–time accuracy, special attention and analysis is given to this

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

fully discrete form because it serves as a fundamental building block for an important class of high-order accurate Runge–Kutta time integration techniques discussed in Section 4.4 that, by construction, inherit TVD (and later maximum principle) properties of the fully discrete scheme (70).

The extension to implicit methods follows immediately upon rewriting the implicit scheme in terms of the solution spatial increments $\Delta_{j+l+1/2}u_h$ and imposing sufficient algebraic conditions such that the implicit matrix acting on spatial increments has a nonnegative inverse.

**Theorem 17.** *(Generalized implicit TVD criteria) The fully discrete implicit 1-D scheme*

$$u_j^{n+1} - \Delta t \sum_{l=-k}^{k-1} C_{j+1/2}^{(l)}(u_h^{n+1})\Delta_{j+l+1/2}u_h^{n+1} = u_j^n, \quad j = 1, \ldots, N \tag{72}$$

*with integer stencil width parameter $k > 0$ is TVNI if for each $j$*

$$C_{j+1/2}^{(k-1)} \geq 0 \tag{73a}$$

$$C_{j+1/2}^{(-k)} \leq 0 \tag{73b}$$

$$C_{j+1/2}^{(l-1)} - C_{j-1/2}^{(l)} \geq 0, \quad -k+1 \leq l \leq k-1,\ l \neq 0 \tag{73c}$$

Theorems 16 and 17 provide sufficient conditions for nonincreasing total variation of explicit (70) or implicit (72) numerical schemes written in incremental form. These incremental forms do not imply *discrete conservation* unless additional constraints are imposed on the discretizations. A sufficient condition for discrete conservation of the discretizations (70) and (72) is that these discretizations can be written in a finite volume flux balance form

$$g_{j+1/2} - g_{j-1/2} = \sum_{l=-k}^{k-1} C_{j+1/2}^{(l)}(u_h)\Delta_{j+l+1/2}u_h$$

where $g_{j\pm1/2}$ are the usual numerical flux functions. Section 4.1.2 provides an example of how the discrete TVD conditions and discrete conservation can be simultaneously achieved. A more comprehensive overview of finite volume numerical methods based on TVD constructions can be found the books by Godlewski and Raviart (1991) and LeVeque (2002).

*4.1.2. MUSCL finite volume methods.* A general family of TVD FVMs with five-cell stencil is the monotone upstream-centered scheme for conservation laws (MUSCL) discretization of van Leer (1979) and van Leer (1985). MUSCL methods utilize a $\kappa$-parameter family of interpolation formulas with *slope limiter function* $\Psi(R)\colon \mathbb{R} \mapsto \mathbb{R}$

$$\begin{aligned} u_{j+1/2}^- &= u_j + \frac{1+\kappa}{4}\Psi(R_i)\Delta_{j-1/2}u_h + \frac{1-\kappa}{4}\Psi\left(\frac{1}{R_j}\right)\Delta_{j+1/2}u_h \\ u_{j-1/2}^+ &= u_j - \frac{1+\kappa}{4}\Psi\left(\frac{1}{R_j}\right)\Delta_{j+1/2}u_h - \frac{1-\kappa}{4}\Psi(R_j)\Delta_{j-1/2}u_h \end{aligned} \tag{74}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

where $R_j$ is a ratio of successive solution increments

$$R_j \equiv \frac{\Delta_{j+1/2} u_h}{\Delta_{j-1/2} u_h} \tag{75}$$

These formulas are self-consistent with a piecewise linear representation of a solution reconstructed from cell average values $u_j$. When this reconstructed solution is cell averaged, the linear corrections vanish identically (regardless of the value of the slope limiter $\Psi(\cdot)$) and the cell average values $u_j$ are obtained. The technique of incorporating limiter functions to obtain nonoscillatory resolution of discontinuities and steep gradients can be found in the flux corrected transport method developed earlier by Boris and Book (1973). For convenience, the interpolation formulas (74) have been written for a uniformly spaced mesh, although the extension to irregular mesh spacing is straightforward. The unlimited form of this interpolation is obtained by setting $\Psi(R) = 1$. In this unlimited case, the truncation error for the conservation law divergence in (6a) is given by

$$\text{Truncation Error} = -\frac{(\kappa - 1/3)}{4}(\Delta x)^2 \frac{\partial^3}{\partial x^3} f(u)$$

Using the MUSCL interpolation formulas given in (74), sufficient conditions to be imposed on the limiter function $\Psi(\cdot)$ to obtain the discrete TVD property are readily obtained.

**Theorem 18.** *(MUSCL TVD FVM) The fully discrete 1-D scheme*

$$u_j^{n+1} = u_j^n - \frac{\Delta t}{\Delta x_j}(g_{j+1/2}^n - g_{j-1/2}^n), \quad j = 1, \ldots, N$$

*with monotone Lipschitz continuous numerical flux function*

$$g_{j+1/2} = g(u_{j+1/2}^-, u_{j+1/2}^+)$$

*utilizing the $\kappa$-parameter family of MUSCL interpolation formulas (74) and (75) is TVNI if there exists a $\Psi(R)$ such that $\forall R \in \mathbb{R}$*

$$0 \le \Psi(R) \le \frac{3 - \kappa}{1 - \kappa} - (1 + \alpha)\frac{1 + \kappa}{1 - \kappa} \tag{76a}$$

*and*

$$0 \le \frac{\Psi(R)}{R} \le 2 + \alpha \tag{76b}$$

*with $\alpha \in [-2, 2\,(1 - \kappa)/(1 + \kappa)]$ under the time step restriction*

$$1 - \frac{\Delta t}{\Delta x_j}\frac{2 - (2 + \alpha)\kappa}{1 - \kappa}\left|\frac{\partial g}{\partial u}\right|_j^{\max} \ge 0$$

*where*

$$\left|\frac{\partial g}{\partial u}\right|_j^{\max} \equiv \sup_{\substack{\tilde{u} \in [u_{j-1/2}^-, u_{j+1/2}^-] \\ \tilde{\tilde{u}} \in [u_{j-1/2}^+, u_{j+1/2}^+]}} \left( \tfrac{\partial g}{\partial \tilde{u}}(\tilde{u}, u_{j+1/2}^+) - \tfrac{\partial g}{\partial \tilde{\tilde{u}}}(u_{j-1/2}^-, \tilde{\tilde{u}}) \right)$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Table 1. Members of the MUSCL TVD family of schemes.

| $\kappa$ | Unlimited scheme | $\beta_{\max}$ | Truncation error |
|---|---|---|---|
| $1/3$ | Third-order | 4 | 0 |
| $-1$ | Fully upwind | 2 | $\frac{1}{3}(\Delta x)^2 \frac{\partial^3}{\partial x^3} f(u)$ |
| 0 | Fromm's | 3 | $\frac{1}{12}(\Delta x)^2 \frac{\partial^3}{\partial x^3} f(u)$ |
| $1/2$ | Low truncation error | 5 | $-\frac{1}{24}(\Delta x)^2 \frac{\partial^3}{\partial x^3} f(u)$ |

From accuracy considerations away from extrema, it is desirable that the unlimited form of the discretization is obtained. Consequently, the constraint $\Psi(1) = 1$ is also imposed upon the limiter function. This constraint, together with the algebraic conditions (76a) and (76b), is readily achieved using the well-known *MinMod* limiter with compression parameter $\beta$ determined from the TVD analysis

$$\Psi^{\mathrm{MM}}(R) = \max(0, \min(R, \beta)), \quad \beta \in \left[1, \frac{(3-\kappa)}{(1-\kappa)}\right] \tag{77}$$

Table 1 summarizes the MUSCL TVD scheme and maximum compression parameter $\beta$ for a number of values of $\kappa$. Another limiter due to van Leer that meets the technical conditions of Theorem 18 and also satisfies $\Psi(1) = 1$ is given by

$$\Psi^{\mathrm{VL}}(R) = \frac{R + |R|}{1 + |R|} \tag{78}$$

This limiter exhibits differentiability away from $R = 0$, which improves the iterative convergence to steady state for many algorithms. Numerous other limiter functions are considered and analyzed in Sweby (1984).

Unfortunately, TVD schemes locally degenerate to piecewise constant approximations at smooth extrema, which locally degrades the accuracy. This is an unavoidable consequence of the strict TVD condition.

**Theorem 19.** *(TVD critical point accuracy, Osher, 1984) The TVD discretizations (68), (70) and (72) all reduce to at most first-order accuracy at nonsonic critical points, that is, points $u^*$ at which $f'(u^*) \neq 0$ and $u^*_x = 0$.*

The next class of finite volume methods addresses this deterioration at smooth extrema.

*4.1.3. ENO/WENO finite volume methods.* To address the degradation in accuracy of TVD methods at critical points, Harten *et al.*, 1986 proposed a new class of finite volume

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

discretizations in 1-D based on a weaker form of total variation control, see also Harten *et al.*, 1987; Harten, 1989. These discretizations utilize polynomial reconstruction from cell averages. Let $\mathcal{P}_p(K)$ denote polynomials of degree at most $p$ in the control volume $K$ and $V_h^p$ the broken space of piecewise $p$-order polynomials for a tesselation $\mathcal{T}$, that is,

$$V_h^p = \{ v \,\big|\, v|_K \in \mathcal{P}_p(K), \quad \forall K \in \mathcal{T} \} \tag{79}$$

Let $R_p^0 \colon V_h^0 \mapsto V_h^p$ denote a polynomial reconstruction operator that maps cell averages to the broken space of $p$-order polynomials. The evolution of solution total variation in the FVM may be understood by first constructing the following operator composition form of the FVM for a time slab increment $[t^n, t^{n+1}]$

$$u_h^{n+1} = A \cdot E(\tau) \cdot R_p^0 \, u_h^n \tag{80}$$

In this equation, $E(\tau)$ is the evolution operator for the PDE, and $A$ is the cell averaging operator. Since $A$ is a nonnegative operator and $E(\tau)$ represents exact evolution over short time, the evolution of the solution total variation satisfies

$$\mathrm{TV}(u_h^{n+1}) = \mathrm{TV}(A \cdot E(\tau) \cdot R_p^0 \, u_h^n) \leq \mathrm{TV}(R_p^0 \, u_h^n)$$

This indicates that the control of solution total variation depends entirely on total variation properties of the reconstruction operator. The requirements of non-oscillatory high-order accuracy for smooth solutions and discrete conservation suggests the following additional design objectives for the reconstruction operator

- $R_p^0(x; u_h) = u(x) + e(x)\Delta x^{p+1} + O(\Delta x^{p+2})$ (81a)
  to insure accuracy whenever the infinite-dimensional solution $u$ is smooth,
- $A|_K R_p^0 \, u_h = u_h|_K = u_K$ (81b)
  to insure discrete conservation,
- $\mathrm{TV}(R_p^0 \, u_h^n) \leq \mathrm{TV}(u_h^n) + O(\Delta x^{p+1})$ (81c)
  to insure an essentially nonoscillatory (ENO) reconstruction.

The last design objective (81c) permits a small $O(\Delta x^{p+1})$ increase in total variation, thus yielding

$$\mathrm{TV}(u_h^{n+1}) \leq \mathrm{TV}(R_p^0 \, u_h^n) \leq \mathrm{TV}(u_h^n) + O(\Delta x^{p+1})$$

which motivates the ENO name given to the reconstruction. Harten points out that the error function $e(x)$ may not be Lipschitz continuous at certain points so that the cumulative error in the scheme is $O(\Delta x^p)$ in a maximum norm but remains $O(\Delta x^{p+1})$ in an $L_1$-norm. To achieve the ENO reconstruction design criterion (81), Harten and coworkers considered breaking the task into two parts:

1. Polynomial reconstruction from a given stencil of cell averages
2. Construction of the "smoothest" polynomial approximation by an adaptive stencil selection algorithm

The resulting ENO reconstruction is then used to calculate numerical flux function states $u^{\pm}_{j+1/2}$ and $u^{\pm}_{j-1/2}$ in the finite volume discretization

$$\frac{d}{\mathrm{d}t} u_j = -\frac{1}{\Delta x_j} (g(u^-_{j+1/2}, u^+_{j+1/2}) - g(u^-_{j-1/2}, u^+_{j-1/2})), \quad j = 1, \ldots, N$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

The semi-discrete equations are then evolved forward in time using some form of high-order accurate time stepping that also prevents the introduction of spurious oscillations (Section 4.4).

In the following section, a commonly used reconstruction technique from cell averages is presented. This is followed by a description of the adaptive stencil algorithm for obtaining the smoothest polynomial approximation.

*4.1.4. ENO reconstruction via primitive function.* Given cell averages $u_j$ of a piecewise smooth function $u(x)$, one can readily evaluate pointwise values using the notion of a *primitive function* $U(x)$ given by

$$U(x) = \int_{x_0}^{x} u(\xi)\,\mathrm{d}\xi$$

by exploiting the relationship

$$\sum_{i=1}^{j} \Delta x_i u_i = U(x_{j+1/2}) = U_{j+1/2}$$

Let $H_{p+1}(x;U)$ denote a $(p+1)$-order polynomial interpolant of a function $U$ defined piecewise for intervals $[x_{j-1/2}, x_{j+1/2}], j = 1, \ldots, N$. Since

$$u(x) \equiv \frac{\mathrm{d}}{\mathrm{d}x} U(x)$$

a $p$-order reconstruction operator is obtained from

$$R_p^0(x;u_h) = \frac{\mathrm{d}}{\mathrm{d}x} H_{p+1}(x;U(u_h))$$

and consequently for smooth data

$$\frac{\mathrm{d}^l}{\mathrm{d}x^l} R_p(x;u_h) = \frac{\mathrm{d}^l}{\mathrm{d}x^l} u(x) + O(\Delta x^{p+1-l})$$

By virtue of the use of the primitive function $U(x)$, it follows that

$$A|_K R_p^0(x;u_h) = u_K$$

and from the polynomial interpolation problem for smooth data

$$R_p^0(x;u_h) = u(x) + O(\Delta x^{p+1})$$

as desired.

*4.1.5. ENO smoothest polynomial approximation.* The piecewise polynomial interpolants $H_{p+1}(x;U)$ described in Section 4.1.4 depend on the particular choice of stencil for pointwise values of the primitive function $U$ for each cell $[x_{j-1/2}, x_{j+1/2}]$. Consequently, the resulting reconstructions will not generally satisfy the oscillation requirement (81c). This motivated Harten and coworkers to consider a new algorithm for smoothest polynomial interpolation

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

using adaptive stencils. When used to interpolate pointwise values of the primitive function, the resulting reconstruction satisfies (81a–c). Specifically, a high-order accurate interpolant is constructed

$$\frac{\mathrm{d}^k}{\mathrm{d}x^k} R_p^0(x;u_h) = \frac{\mathrm{d}^k}{\mathrm{d}x^k} u(x) + O(\Delta x^{p+1-k}) \ , \quad 0 \le k \le p$$

which avoids having Gibbs oscillations at discontinuities in the sense

$$TV(R_p^0(x;u_h)) \le TV(u) + \mathcal{O}(\Delta x^{p+1})$$

The strategy pursued by Harten and coworkers was to construct an ENO polynomial $H_{p+1}^{\mathrm{ENO}}(x;U(u_h))$ in each interval $[x_{j-1/2}, x_{j+1/2}], j = 1,\ldots,N$, which interpolates the primitive function $U(x)$ at the $p+2$ successive points $\{x_{i-1/2}\}$, $i_p(j) \le i \le i_p(j)+p+1$ that include $x_{j-1/2}$ and $x_{j+1/2}$. This describes $p+1$ possible polynomials depending on the choice of $i_p(j)$ for an interval $[x_{j-1/2}, x_{j+1/2}]$. The ENO strategy selects the value $i_p(j)$ for each interval that produces the "smoothest" polynomial interpolant for a given input data. Information about smoothness of $U(x)$ is extracted from a table of divided differences defined recursively

$$\begin{aligned}
U[x_{i-1/2}] &= U(x_{i-1/2}) \\
U[x_{i-1/2}, x_{i+1/2}] &= \frac{U[x_{i+1/2}] - U[x_{i-1/2}]}{x_{i+1/2} - x_{i-1/2}} \\
&\vdots \\
U[x_{i-1/2}, \ldots, x_{i+k-1/2}] &= \frac{U[x_{i+1/2}, \ldots, x_{i+k-1/2}] - U[x_{i-1/2}, \ldots, x_{i+k-3/2}]}{x_{i+k-1/2} - x_{i-1/2}}
\end{aligned}$$

The stencil producing the smoothest interpolant is then chosen hierarchically by setting

$$i_1(j) = j$$

and for $k = 1,\ldots,p$

$$i_{k+1}(j) = \begin{cases} i_k(j) - 1 & \text{if } |U[x_{i_k(j)-3/2}, \ldots, x_{i_k(j)+k-1/2}]| < |U[x_{i_k(j)-1/2}, \ldots, x_{i_k(j)+k+1/2}]| \\ i_k(j) & \text{otherwise} \end{cases} \tag{82}$$

Harten *et al.* (1986) demonstrated that this interpolant is monotone in any cell interval containing a discontinuity and the resulting reconstruction satisfies the design objectives (81a-c).

*4.1.6. WENO reconstruction.* The solution adaptive nature of the ENO stencil selection algorithm (82) yields nondifferentiable fluxes that may impede the performance of solution algorithms Jiang and Shu (1996). Recall that the $p$-order ENO reconstruction considers $p+1$ possible polynomial stencils. The stencil selection algorithm chooses only one of these possible stencils and other slightly less smooth stencils may give similar $\mathcal{O}(\Delta x^{p+1})$ accuracy. However, using a linear combination of *all* $p+1$ possible polynomials $\{P_p^{(0)}, P_p^{(2)}, \ldots, P_p^{(p)}\}$ with optimized weights $\omega_k, k = 0,\ldots,p$ potentially yields a more accurate $\mathcal{O}(\Delta x^{2p+1})$ interpolant for smooth enough data

$$P_p(x) = \sum_{k=0}^{p} \omega_k P_p^{(k)}(x) + \mathcal{O}(\Delta x^{2p+1}) \quad \sum_{k=0}^{p} \omega_k = 1$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

For example, optimized weights for $p = 0, 1, 2$ are readily computed

$$\begin{aligned} p &= 0: \quad \omega_0 = 1 \\ p &= 1: \quad \omega_0 = \frac{2}{3}, \omega_1 = \frac{1}{3} \\ p &= 2: \quad \omega_0 = \frac{3}{10}, \omega_1 = \frac{3}{5}, \omega_2 = \frac{1}{10} \end{aligned}$$

In the weighted essentially nonoscillatory (WENO) schemes of Jiang and Shu (1996) and Shu (1999), approximate weights $\widetilde{\omega}_k, k = 0, \ldots, p$ are devised such that for smooth solutions

$$\widetilde{\omega}_k = \omega_k + \mathcal{O}(\Delta x^p)$$

so that the $\mathcal{O}(\Delta x^{2p+1})$ accuracy is still retained using these approximations

$$P_p(x) = \sum_{k=0}^{p} \widetilde{\omega}_k P_p^{(k)}(x) + \mathcal{O}(\Delta x^{2p+1}) \ , \quad \sum_{k=0}^{p} \widetilde{\omega}_k = 1$$

The approximate weights are constructed using the *ad hoc* formulas for $k = 0, \ldots, p$

$$\alpha_k = \frac{\omega_k}{(\epsilon + \beta_k)^2}, \quad \widetilde{\omega}_k = \frac{\alpha_k}{\sum_{l=0}^{p} \alpha_l}$$

where $\epsilon$ is an approximation to the square root of the machine precision and $\beta_k$ is a smoothness indicator

$$\beta_k = \sum_{l=1}^{p} \int_{x_{j-1/2}}^{x_{j+1/2}} \Delta x^{2l-1} \left( \frac{d^l P_p^{(k)}(x)}{\partial x^l} \right)^2 \mathrm{d}x$$

For a sequence of smooth solutions with decreasing smoothness indicator $\beta_k$, these formulas approach the optimized weights, $\widetilde{\omega}_k \to \omega_k$. These formulas also yield vanishing weights $\widetilde{\omega}_k \to 0$ for stencils with large values of the smoothness indicator such as those encountered at discontinuities. In this way, the WENO construction retains some of the attributes of the original ENO formulation but with increased accuracy in smooth solution regions and improved differentiability often yielding superior robustness for steady-state calculations.

### *4.2. Higher order accurate finite volume methods for hyperbolic problems in multiple dimensions*

Although the one-dimensional finite volume discretizations given in previous sections may be readily applied in multidimensions on a dimension-by-dimension basis, a result of Goodman and LeVeque (1985) shows that TVD schemes in two or more space dimensions are only first-order accurate.

**Theorem 20.** *(Accuracy of TVD finite volume discretization on multidimensional Cartesian meshes) Any two-dimensional finite volume scheme of the form*

$$u_{i,j}^{n+1} = u_{i,j}^n - \frac{\Delta t}{|K_{i,j}|}(g_{i+1/2,j}^n - g_{i-1/2,j}^n) - \frac{\Delta t}{|K_{i,j}|}(h_{i,j+1/2}^n - h_{i,j-1/2}^n), \quad 1 \le i \le M, \ 1 \le j \le N$$

*with Lipschitz continuous numerical fluxes for integers $p, q, r, s$*

$$\begin{aligned} g_{i+1/2,j} &= g(u_{i-p,j-q}, \ldots, u_{i+r,j+s}) \\ h_{i,j+1/2} &= h(u_{i-p,j-q}, \ldots, u_{i+r,j+s}) \end{aligned}$$

*that is TVNI in the sense*

$$TV(u_h^{n+1}) \leq TV(u_h^n)$$

*where*

$$TV(u) \equiv \sum_{i,j} \Big[ \Delta y_{i+1/2,j} |u_{i+1,j} - u_{i,j}| + \Delta x_{i,j+1/2} |u_{i,j+1} - u_{i,j}| \Big]$$

*is at most first order accurate.*

Motivated by the negative results of Goodman and LeVeque, weaker conditions yielding solution monotonicity preservation have been developed from discrete maximum principle analysis. These alternative constructions have the positive attribute that they extend to unstructured meshes as well.

*4.2.1. Positive coefficient FVMs on structured meshes.* Lemma 3 considers FVMs of the form

$$u_K^{n+1} = u_K^n + \frac{\Delta t}{|K|} \sum_{\sigma_{K,L} \subset K} C_{K,L}(u_h^n)(u_L^n - u_K^n)$$

with $u_h = \{u_{K_1}, u_{K_2}, \ldots\}$ and establishes a local discrete maximum principle

$$\min_{\sigma_{K,L} \subset K} (u_K^n, u_L^n) \leq u_K^{n+1} \leq \max_{\sigma_{K,L} \subset K} (u_K^n, u_L^n)$$

for each $K \in \mathcal{T}$ and $n = 0, 1, 2, \ldots$ under a CFL-like condition on the time step parameter if all coefficients $C_{K,L}(u_h)$ are nonnegative. Discretizations of this type are often called *positive coefficient* discretizations or more simply *positive* discretizations. To circumvent the negative result of Theorem 20, Spekreijse (1987) developed a family of high-order accurate positive coefficient discretizations on two-dimensional structured meshes. For purposes of positivity analysis, these methods are written in incremental form on a $M \times N$ logically rectangular 2-D mesh

$$\begin{aligned} u_{i,j}^{n+1} &= u_{i,j}^n + \Delta t \Big( A_{i+1,j}^n (u_{i+1,j}^n - u_{i,j}^n) + B_{i,j+1}^n (u_{i,j+1}^n - u_{i,j}^n) \\ &\quad + C_{i-1,j}^n (u_{i-1,j}^n - u_{i,j}^n) + D_{i,j-1}^n (u_{i,j-1}^n - u_{i,j}^n) \Big), \; 1 \leq i \leq M, \; 1 \leq j \leq N \end{aligned} \quad (83)$$

where the coefficients are nonlinear functions of the solution

$$\begin{aligned} A_{i+1,j}^n &= A(\ldots, u_{i-1,j}^n, u_{i,j}^n, u_{i+1,j}^n, \ldots) \\ B_{i,j+1}^n &= B(\ldots, u_{i,j-1}^n, u_{i,j}^n, u_{i,j+1}^n, \ldots) \\ C_{i-1,j}^n &= C(\ldots, u_{i-1,j}^n, u_{i,j}^n, u_{i+1,j}^n, \ldots) \\ D_{i,j-1}^n &= D(\ldots, u_{i,j-1}^n, u_{i,j}^n, u_{i,j+1}^n, \ldots) \end{aligned}$$

Once written in incremental form, the following lemma follows from standard positive coefficient maximum principle analysis.

**Lemma 4.** *(Positive coefficient FVMs in multidimensions) The discretization (83) is a positive coefficient FVM if for each* $1 \le i \le M,\ 1 \le j \le N$ *and time slab increment* $[t^n, t^{n+1}]$, $n = 0, 1, 2, \ldots$

$$A^n_{i+1,j} \ge 0,\ B^n_{i,j+1} \ge 0,\ C^n_{i-1,j} \ge 0,\ D^n_{i,j-1} \ge 0 \tag{84}$$

*and*

$$1 - \Delta t\ (A^n_{i+1,j} + B^n_{i,j+1} + C^n_{i-1,j} + D^n_{i,j-1}) \ge 0 \tag{85}$$

*with discrete maximum principle*

$$\min(u^n_{i,j}, u^n_{i-1,j}, u^n_{i+1,j}, u^n_{i,j-1}, u^n_{i,j+1}) \le u^{n+1}_{i,j} \le \max(u^n_{i,j}, u^n_{i-1,j}, u^n_{i+1,j}, u^n_{i,j-1}, u^n_{i,j+1})$$

Using a procedure similar to that used in the development of MUSCL TVD FVMs in 1-D, Spekreijse (1987) developed a family of monotonicity preserving MUSCL interpolations from the positivity conditions of Lemma 4.

**Theorem 21.** *(MUSCL positive coefficient FVM) Assume a fully discrete 2-D FVM*

$$u^{n+1}_{i,j} = u^n_{i,j} - \frac{\Delta t}{|K_{i,j}|}(g^n_{i+1/2,j} - g^n_{i-1/2,j}) - \frac{\Delta t}{|K_{i,j}|}(h^n_{i,j+1/2} - h^n_{i,j-1/2}), \quad 1 \le i \le M,\ 1 \le j \le N$$

*for* $n = 0, 1, 2, \ldots$ *utilizing monotone Lipschitz continuous numerical flux functions*

$$\begin{aligned} g_{i+1/2,j} &= g(u^-_{i+1/2,j}, u^+_{i+1/2,j}) \\ h_{i,j+1/2} &= h(u^-_{i,j+1/2}, u^+_{i,j+1/2}) \end{aligned}$$

*and MUSCL extrapolation formulas*

$$\begin{aligned} u^-_{i+1/2,j} &= u_{i,j} + \frac{1}{2}\Psi(R_{i,j})(u_{i,j} - u_{i-1,j}) \\ u^+_{i-1/2,j} &= u_{i,j} - \frac{1}{2}\Psi\left(\frac{1}{R_{i,j}}\right)(u_{i+1,j} - u_{i,j}) \\ u^-_{i,j+1/2} &= u_{i,j} + \frac{1}{2}\Psi(S_{i,j})(u_{i,j} - u_{i,j-1}) \\ u^+_{i,j-1/2} &= u_{i,j} - \frac{1}{2}\Psi\left(\frac{1}{S_{i,j}}\right)(u_{i,j+1} - u_{i,j}) \end{aligned}$$

*with*

$$R_{i,j} \equiv \frac{u_{i+1,j} - u_{i,j}}{u_{i,j} - u_{i-1,j}}, \quad S_{i,j} \equiv \frac{u_{i,j+1} - u_{i,j}}{u_{i,j} - u_{i,j-1}}$$

*This finite volume discretization satisfies the local maximum principle properties of Lemma 4 and is second-order accurate if the limiter* $\Psi = \Psi(R)$ *has the properties that there exist constants* $\beta \in (0, \infty)$, $\alpha \in [-2, 0]$ *such that* $\forall R \in \mathbb{R}$

$$\alpha \le \Psi(R) \le \beta, \quad -\beta \le \frac{\Psi(R)}{R} \le 2 + \alpha \tag{86}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

*with the constraint* $\Psi(1) = 1$ *and the smoothness condition* $\Psi(R) \in C^2$ *near* $R = 1$ *together with a time step restriction for stability*

$$1 - (1+\beta)\frac{\Delta t}{|K_{i,j}|}\left(\left|\frac{\partial g}{\partial u}\right|_{i,j}^{n,\max} + \left|\frac{\partial h}{\partial u}\right|_{i,j}^{n,\max}\right) \geq 0$$

*where*

$$\left|\frac{\partial g}{\partial u}\right|_{i,j}^{\max} \equiv \sup_{\substack{\tilde{u}\in[u^-_{i-1/2,j},u^-_{i+1/2,j}]\\ \tilde{\tilde{u}}\in[u^+_{i-1/2,j},u^+_{i+1/2,j}]}} \left(\tfrac{\partial g}{\partial \tilde{u}}(\tilde{u}, u^+_{i+1/2,j}) - \tfrac{\partial g}{\partial \tilde{\tilde{u}}}(u^-_{i-1/2,j}, \tilde{\tilde{u}})\right) \geq 0$$

$$\left|\frac{\partial h}{\partial u}\right|_{i,j}^{\max} \equiv \sup_{\substack{\hat{u}\in[u^-_{i,j-1/2},u^-_{i,j+1/2}]\\ \hat{\hat{u}}\in[u^+_{i,j-1/2},u^+_{i,j+1/2}]}} \left(\tfrac{\partial h}{\partial \hat{u}}(\hat{u}, u^+_{i,j+1/2}) - \tfrac{\partial h}{\partial \hat{\hat{u}}}(u^-_{i,j-1/2}, \hat{\hat{u}})\right) \geq 0$$

Many limiter functions satisfy the technical conditions (86) of Theorem 21. Some examples include

- the van Leer limiter (78)
$$\Psi^{\mathrm{VL}}(R) = \frac{R + |R|}{1 + |R|}$$
- Koren limiter Koren (1988)
$$\Psi^{\mathrm{K}}(R) = \frac{R + 2R^2}{2 - R + 2R^2}$$

For smooth solutions, the Koren limiter results in state extrapolations identical to the most accurate $\kappa = 1/3$ MUSCL interpolations of Section 4.1.2.

### *4.3. Higher order accurate finite volume methods for hyperbolic problems on unstructured meshes*

FVMs for hyperbolic problems have been extended to unstructured meshes using data reconstruction from cell averages. The polynomial reconstruction operator in the FVM maps cell-averaged data into the broken space (79) consisting of $p$-order piecewise polynomials in each control volume, $R_p^0 \colon V_h^0 \mapsto V_h^p$, while preserving the cell average value, $\int_K R_p^0(x; u_h)\, dx = u_K|K|$ for all $K \in \mathcal{T}$. Using this reconstruction operator, higher order accurate finite volume discretizations on unstructured meshes using polynomial data reconstruction are of the general form for time-invariant control volumes $K$

$$\frac{\mathrm{d}u_K}{\mathrm{d}t} = -\frac{1}{|K|}\sum_{\substack{\sigma\subset\partial K\\ 1\leq q\leq Q}} \omega_q\, g(u_\sigma^-(x_{q,\sigma}; u_h), u_\sigma^+(x_{q,\sigma}; u_h); \boldsymbol{n}_\sigma)\,|\sigma|, \quad \forall K \in \mathcal{T} \tag{87}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

using $Q$-point quadrature with positive weights $\omega_q \in \mathbb{R}^+$ and locations $x_{q,\sigma} \in \sigma$ for $q = 1, \ldots, Q$. In this formula, each $\sigma$ interface is assumed planar with unique normal $\boldsymbol{n}_\sigma$ and $g(u, v; \boldsymbol{n})$ denotes any of the numerical fluxes described in Section 3. The $\sigma$ interface states used in the numerical flux quadrature are calculated from the reconstruction $R_p^0(x; u_h)$

$$u_\sigma^\pm(x; u_h) \equiv \lim_{\epsilon \downarrow 0} R_p^0(x \pm \epsilon \boldsymbol{n}_\sigma; u_h) \tag{88}$$

Due to the cell-wise piecewise structure of the reconstruction operator, the two states $u_\sigma^\pm(x; u_h)$ are generally distinct.

*4.3.1. General p-exact reconstruction operators on unstructured meshes.* The reconstruction operator discussed in Section 4.1.4 exploits properties of a primitive function in 1-D. This approach is problematic to extend to general unstructured meshes. Consequently, more general reconstruction design principles have been developed and used in implementations. Abstractly, the reconstruction operator $R_p^0$ appearing in (88) serves as a finite-dimensional pseudoinverse of the cell averaging operator $A$. The development of a general polynomial reconstruction operator, $R_p^0$, that reconstructs $p$-degree polynomials from cell-averaged data on unstructured meshes follows from the application of a small number of design principles:

1. (Conservation of the mean) Given solution cell averages $u_h \in V_h^0$, the reconstruction $R_p^0 u_h$ is required to have the correct cell average, that is,

   $$\text{if } \; v = R_p^0 u_h \text{ then } \; u_h = Av$$

   More concisely,

   $$AR_p^0 = I$$

   so that $R_p^0$ is a right inverse of the averaging operator $A$.
2. ($p$-exactness) A reconstruction operator $R_p^0$ is $p$-exact if $R_p^0 A$ reconstructs polynomials of degree $p$ or less exactly, that is,

   $$\text{if } \; u \in \mathcal{P}_p \text{ and } \; v = Au \text{ then } \; R_p^0 v = u$$

   This can be written succinctly as

   $$R_p^0 A|_{\mathcal{P}_p} = I$$

   so that $R_p^0$ is a left inverse of the averaging operator $A$ restricted to the space of polynomials of degree at most $p$.
3. (Compact support) The reconstruction in a control volume $K$ should only depend on cell averages in a relatively small neighborhood surrounding $K$. Recall that a polynomial of degree $p$ in $\mathbb{R}^d$ contains $\binom{p+d}{d}$ degrees of freedom. The support set for $K$ is required to contain at least this number of neighbors. As the support set becomes even larger for fixed $p$, not only does the computational cost increase, but eventually, the accuracy decreases as less valid data from further away is brought into the calculation.

Practical implementations of general $p$-order polynomial reconstruction operators fall into two classes:

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

- *Fixed support stencil reconstructions.* These methods choose a fixed support set as a preprocessing step. Various limiting strategies are then employed to obtain nonoscillatory approximation; see, for example, Barth and Frederickson (1990) and Delanaye (1996) for further details.
- *Adaptive support stencil reconstructions.* These ENO-like methods dynamically choose reconstruction stencils based on solution smoothness criteria; see, for example, Harten and Chakravarthy (1991), Vankeirsblick (1993), Abgrall (1994), Sonar (1997), and Sonar (1998) for further details.

*4.3.2. Higher order finite volume methods for hyperbolic problems on unstructured meshes using linear reconstruction.* Considerable simplification is possible when only linear reconstruction polynomials are sought, $R_1^0(x; u_h)$. The maximum principle analysis presented next not only gives sufficient conditions for a discrete maximum principle but also explicitly exposes the dependence of cell shape with respect to maximum allowable time step for the fully discrete approximation. This geometrical shape parameter for a control volume $K$ can be calculated from the formula

$$\Gamma_K^{\text{geom}} = \sup_{0 \le \theta \le 2\pi} \alpha_K^{-1}(\theta) \tag{89}$$

where $0 < \alpha_K(\theta) < 1$ represents the smallest fractional perpendicular distance from the centroid to one of two minimally separated parallel hyperplanes with orientation $\theta$ and hyperplane location such that all quadrature points in the control volume lie between or on the hyperplanes as shown in Figure 3. Table 2 lists $\Gamma^{\text{geom}}$ values for various control volume shapes

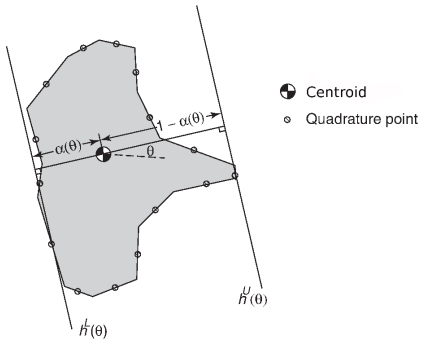


Figure 3. Minimally separated hyperplanes $h^L(\theta)$ and $h^U(\theta)$ and the fractional distance ratio $\alpha(\theta)$ for use in the calculation of $\Gamma^{\text{geom}}$.

in $\mathbb{R}^1$, $\mathbb{R}^2$, $\mathbb{R}^3$, and $\mathbb{R}^d$. As might be expected, those geometries that have exact quadrature point symmetry with respect to the control volume centroid have geometric shape parameters $\Gamma^{\text{geom}}$ equal to 2 regardless of the number of space dimensions involved.

Using standard maximum principle analysis, a discrete maximum principle is obtained under a CFL-like time step restriction if the solution reconstruction in each and every control volume

Table 2. Reconstruction geometry factors for various control volume shapes utilizing midpoint flux quadrature rule.

| Control volume shape | Space dimension | $\Gamma^{\text{geom}}$ |
|---|---|---|
| Segment | 1 | 2 |
| Triangle | 2 | 3 |
| Parallelogram | 2 | 2 |
| Tetrahedron | 3 | 4 |
| parallelepiped | 3 | 2 |
| Simplex | d | d + 1 |
| Hyper-parallelepiped | d | 2 |

$K$ can be bounded from above and below respectively by the neighboring cell averages. This restriction is stated more precisely in the following theorem:

**Theorem 22.** *(Finite volume maximum principle on unstructured meshes, $R_1^0$) Let $u_K^{\min}$ and $u_K^{\max}$ denote the minimum and maximum value of solution cell averages for a given cell $K$ and all adjacent cell neighbors, that is,*

$$u_K^{\min} \equiv \min_{\sigma_{K,L} \subset \partial K} (u_K, u_L) \text{ and } \quad u_K^{\max} \equiv \max_{\sigma_{K,L} \subset \partial K} (u_K, u_L) \tag{90}$$

*The fully discrete finite volume scheme*

$$u_K^{n+1} = u_K^n - \frac{\Delta t}{|K|} \sum_{\substack{\sigma \subset \partial K \\ 1 \le q \le Q}} \omega_q \, g(u_\sigma^-(x_{q,\sigma}; u_h), u_\sigma^+(x_{q,\sigma}; u_h); \boldsymbol{n}_\sigma) \, |\sigma|, \quad \forall K \in \mathcal{T} \tag{91}$$

*with monotone Lipschitz continuous numerical flux function, $Q$-point quadrature with nonnegative weights, and linear reconstructions*

$$u_\sigma^{\pm}(x; u_h) \equiv \lim_{\epsilon \downarrow 0} R_1^0(x \pm \epsilon \boldsymbol{n}_\sigma; u_h) \tag{92}$$

*exhibits the local maximum principle for each $K \in \mathcal{T}$ and $n = 0, 1, 2, \ldots$*

$$u_K^{\min,n} \le u_K^{n+1} \le u_K^{\max,n}$$

*under the time step restriction*

$$1 - \frac{\Delta t}{|K|} \Gamma_K^{\text{geom}} \sum_{\substack{\sigma \subset \partial K \\ 1 \le q \le Q}} \sup_{\substack{\tilde{u} \in [u_K^{\min,n}, u_K^{\max,n}] \\ \tilde{\tilde{u}} \in [u_K^{\min,n}, u_K^{\max,n}]}} \left| \frac{\partial g}{\partial \widetilde{\widetilde{u}}} (\widetilde{u}, \widetilde{\widetilde{u}}; \boldsymbol{n}_\sigma) |\sigma| \right| \ge 0 \tag{93}$$

*if the linear reconstruction evaluated at interface quadrature points satisfies*

$$\max(u_K^{\min,n}, u_L^{\min,n}) \le u_\sigma^{-,n}(x_{q,\sigma}) \le \min(u_K^{\max,n}, u_L^{\max,n}), \quad \forall \sigma_{K,L} \subset \partial K \tag{94}$$

*for all $x_{q,\sigma} \in \sigma$, $q = 1, \ldots, Q$.*

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Equation (94) can be interpreted as stating that the data reconstruction in a cell, when evaluated at quadrature points, should be bounded from above and below by cell-averaged values of adjacent neighbors (including itself). This is completely consistent with the interpretation originally given in van Leer (1979).

Comparing the maximum allowable time step using piecewise constant approximation (52) with piecewise linear approximation (93) indicates that the maximum time step using piecewise linear approximation is reduced by a factor of approximately $1/\Gamma_K^{\text{geom}}$. Using shape regular quadrilaterals (2-D) or hexahedron (3-D), this time step reduction is approximately 1/2. Using triangles (2-D) or tetrahedra (3-D), this time step reduction is approximately 1/3 and 1/4, respectively.

*4.3.3. Slope limiters for linear reconstruction.* Given a linear reconstruction $R_1^0(x;u_h)$ that does not necessarily satisfy the requirements of Theorem 22, it is possible to modify the reconstruction so that the modified reconstruction does satisfy the requirements of Theorem 22. For each control volume $K \in \mathcal{T}$, assume a modified local reconstruction operator $\widetilde{R}_1^0(x;u_h)|_K$ of the form

$$\widetilde{R}_1^0(x;u_h)|_K = u_K + \alpha_K(R_1^0(x;u_h)|_K - u_K)$$

for $\alpha_K \in [0,1]$. By construction, this modified reconstruction correctly reproduces the control volume cell average for all values of $\alpha_K$, that is,

$$\frac{1}{|K|}\int_K \widetilde{R}_1^0(x;u_h)\,\mathrm{d}x = u_K \tag{95}$$

The most restrictive value of $\alpha_K$ for each control volume $K$ is then computed on the basis of the Theorem 22 constraint (94)

$$\alpha_K^{\text{MM}} = \min_{\substack{\sigma_{K,L}\subset\partial K\\ 1\le q\le Q}} \begin{cases} \frac{\min(u_K^{\max},u_L^{\max})-u_K}{R_1^0(x_{q,\sigma};u_h)|_K-u_K} & \text{if } R_1^0(x_{q,\sigma};u_h)|_K > \min(u_K^{\max},u_L^{\max}) \\ \frac{\max(u_K^{\min},u_L^{\min})-u_K}{R_1^0(x_{q,\sigma};u_h)|_K-u_K} & \text{if } R_1^0(x_{q,\sigma};u_h)|_K < \max(u_K^{\min},u_L^{\min}) \\ 1 & \text{otherwise} \end{cases} \tag{96}$$

where $u^{\max}$ and $u^{\min}$ are defined in (90). When the resulting modified reconstruction operator is used in the extrapolation formulas (92), the discrete maximum principle of Theorem 22 is attained under a CFL-like time step restriction. Utilizing the inequalities

$$\max(u_K,u_L) \le \min(u_K^{\max},u_L^{\max}) \;\text{ and }\; \min(u_K,u_L) \ge \max(u_K^{\min},u_L^{\min})$$

it is straightforward to construct a more restrictive limiter function

$$\alpha_K^{\text{LM}} = \min_{\substack{\sigma_{K,L}\subset\partial K\\ 1\le q\le Q}} \begin{cases} \frac{\max(u_K,u_L)-u_K}{R_1^0(x_{q,\sigma};u_h)|_K-u_K} & \text{if } R_1^0(x_{q,\sigma};u_h)|_K > \max(u_K,u_L) \\ \frac{\min(u_K,u_L)-u_K}{R_1^0(x_{q,\sigma};u_h)|_K-u_K} & \text{if } R_1^0(x_{q,\sigma};u_h)|_K < \min(u_K,u_L) \\ 1 & \text{otherwise} \end{cases} \tag{97}$$

that yields modified reconstructions satisfying the technical conditions of Theorem 22. This simplified limiter (97) introduces additional slope reduction when compared to (96). This can

be detrimental to the overall accuracy of the discretization. The limiter strategy (97) and other variants for simplicial control volumes are discussed further in Liu (1993), Wierse (1994), and Batten *et al.* (1996). Note that in practical implementations, both limiters (96) and (97) require some modification to prevent division by zero for constant solution data.

*4.3.4. Linear reconstruction on simplicial control volumes.* Linear reconstruction operators on general control volumes that satisfy the cell averaging requirement often exploit the fact that the cell average is also a pointwise value of any valid linear reconstruction evaluated at the centroid of the control volume. This reduces the reconstruction problem to that of gradient estimation given pointwise samples at the centroids. In this case, it is convenient to express the reconstruction in the form

$$R_1^0(x; u_h)|_K = u_K + (\nabla u)_K \cdot (x - x_K^{\bullet}) \tag{98}$$

where $x_K^{\bullet}$ denotes the centroid for the control volume $K$ and $(\nabla u)_K$ is the gradient to be determined. Figure 4 depicts a 2-D simplex $K_O$ and three adjacent neighboring simplices. Also shown are the corresponding four pointwise solution values $\{u_A, u_B, u_C, u_O\}$ located at centroids of each simplex. By selecting any three of the four pointwise solution

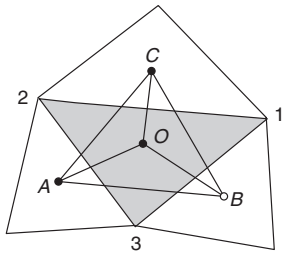


Figure 4. Triangle control volume $K_O$ (shaded) with three adjacent cell neighbors.

values, a set of four possible gradients are uniquely determined, that is, $\{\nabla(u_A, u_B, u_C)$, $\nabla(u_A, u_B, u_O)$,$\nabla(u_B, u_C, u_O)$,$\nabla(u_C, u_A, u_O)\}$. A number of slope limited reconstruction techniques are possible for use in the finite volume scheme (91) that meet the technical conditions of Theorem 22.

1. Choose $(\nabla u)_{K_O} = \nabla(u_A, u_B, u_C)$ and limit the resulting reconstruction using (96) or (97).
2. Limit four reconstructions corresponding to $(\nabla u)_{K_O}$ equal to $\nabla(u_A, u_B, u_C)$, $\nabla(u_A, u_B, u_O)$, $\nabla(u_B, u_C, u_O)$, and $\nabla(u_C, u_A, u_O)$ using (96) or (97) and choose the limited reconstruction with largest gradient magnitude. This technique is a generalization of that described in Batten *et al.* (1996) wherein limiter (97) is used.
3. Evaluate four reconstructions corresponding to $(\nabla u)_{K_O}$ equal to $\nabla(u_A, u_B, u_C)$, $\nabla(u_A, u_B, u_O)$, $\nabla(u_B, u_C, u_O)$, and $\nabla(u_C, u_A, u_O)$ and choose the largest gradient magnitude that satisfies the maximum principle reconstruction bound inequality (94). If all reconstructions fail the bound inequality, the reconstruction gradient is set equal to zero; see Liu (1993).

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

*4.3.5. Linear reconstruction on general control volume shapes.* It is again convenient to express the linear reconstruction in the form

$$R_1^0(x; u_h)|_K = u_K + (\nabla u)_K \cdot (x - x_K^{\otimes}) \tag{99}$$

but now the shape and number of adjacent neighboring cells is irregular. Two common techniques for simplified linear reconstruction include a simplified least squares technique and a Green-Gauss quadrature technique.

*Least squares linear reconstruction* Again exploiting the fact that the cell average value is also a pointwise value of the linear reconstruction evaluated at the centroid of a general control volume shape, the task of linear reconstruction reduces to the problem of gradient estimation given pointwise values. In the simplified least squares reconstruction technique, a triangulation (2-D) or tetrahedralization (3-D) of centroids is constructed as shown in Figure 5. Referring

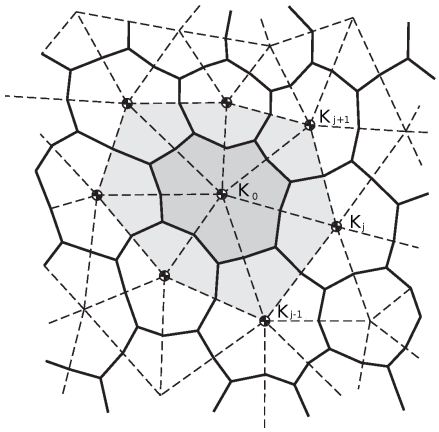


Figure 5. Triangulation of centroid locations showing a typical control volume $K_0$ associated with cyclically indexed graph neighbors $K_j, j = 1, \ldots, N_0$.

to this figure, for each of the $N_0$ edges of the simplex mesh incident to $K_0$, an edge projected gradient constraint equation is constructed subject to a specified nonzero scaling $w_{K_0}$ for $j = 1, \ldots, N_0$

$$w_j(\nabla u)_{K_0} \cdot (x_j^{\otimes} - x_0^{\otimes}) = w_k(u_{K_j} - u_{K_0})$$

The number of edges incident to a simplicial mesh vertex in $\mathbb{R}^d$ is greater than $d$ away from mesh boundaries thereby producing the following nonsquare matrix of constraint equations

$$\begin{bmatrix} w_1 \Delta x_1^{\otimes} & w_1 \Delta y_1^{\otimes} \\ \vdots & \vdots \\ w_{N_0} \Delta x_{N_0}^{\otimes} & w_{N_0} \Delta y_{N_0}^{\otimes} \end{bmatrix} (\nabla u)_{K_0} = \begin{pmatrix} w_1(u_{K_1} - u_{K_0}) \\ \vdots \\ w_{N_0}(u_{K_{N_0}} - u_{K_0}) \end{pmatrix}$$

or in abstract form

$$\begin{bmatrix} \vec{L}_1 & \vec{L}_2 \end{bmatrix} \nabla u = \vec{f}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

This abstract form can be solved in a least squares sense using a variety of techniques, for example, Gram-Schmidt, modified Gram-Schmidt, Householder rotations, SVD, and so on. When least squares matrix conditioning is not an issue, the following symbolic solution may be used

$$\nabla u = \frac{1}{l_{11}l_{22} - l_{12}^2} \begin{pmatrix} l_{22}(\vec{L}_1 \cdot \vec{f}) - l_{12}(\vec{L}_2 \cdot \vec{f}) \\ l_{11}(\vec{L}_2 \cdot \vec{f}) - l_{12}(\vec{L}_1 \cdot \vec{f}) \end{pmatrix} \tag{100}$$

with $l_{ij} = \vec{L}_i \cdot \vec{L}_j$. The form of this solution in terms of scalar dot products over incident edges suggests that the least squares linear reconstruction can be efficiently computed via an edge data structure without the need for storing a nonsquare matrix.

*Green–Gauss linear reconstruction* Again referring to Figure 5, reconstruction gradients may be approximated from mean value approximation and application of the Green–Gauss identity

$$|K_0|(\nabla u)_{K_0} \approx \int_{K_0} \nabla u \, \mathrm{d}x = \int_{\partial K_0} u \, \boldsymbol{n} \, \mathrm{d}x = \sum_{j=1}^{N_0} \int_{\sigma_{K_0,K_j}} u \, \boldsymbol{n} \, \mathrm{d}x \approx \sum_{j=1}^{N_0} \frac{1}{2}(u_{K_0} + u_{K_j}) \, \boldsymbol{\nu}_{0j} \tag{101}$$

In this formula, $\boldsymbol{\nu}_{0j} \equiv \int_{\sigma_{K_0,K_j}} \boldsymbol{n} \, dx$ which is independent of the shape of $\sigma_{K_0,K_j}$ and only depends on its boundary shape (end points in 2-d). A notable property of this formula is that the gradient calculation is exact whenever the numerical solution varies linearly over the support of the reconstruction.

The slope limiting procedures (96) or (97) may then be applied to the least squares or Gauss-Gauss reconstruction so that when used in the finite volume discretization (91) the discrete maximum principle of Theorem 22 is obtained. Gradient reconstruction together with a slope limitation on general meshes can be found in Buffard and Clain, 2010; Calgaro *et al.*, 2010. In these works, the limitation procedure is presented as a limitation of the slope defined by the cell and face values on the basis of its comparison with other slopes defined by the values taken by the solution in the neighborhood. Then, under geometric assumptions for the mesh, this limitation may be shown to imply some conditions for the approximation at the face, which again ensure, at least for pure convection problems, a local maximum principle Clain and Clauzon, 2010; Clain, 2013; Berthon *et al.*, 2014.

Another novel approach found in Tran, 2008; Michel *et al.*, 2010 and Piar *et al.*, 2013; Therme, 2015 is based on the observation that for a linear convection term, stability conditions may be exploited in the calculation of an admissible interval for the state value at the face. This suggests a crude limitation process, which does not use any slope computation and simply consists of performing a (one-dimensional) projection of the tentative affine reconstructed face value on this interval. In addition, stability conditions are purely algebraic (in the sense that they do not require any geometric computation) and thus work with arbitrary meshes.

*4.3.6. Positive coefficient finite volume methods on unstructured meshes.* Several related positive coefficient schemes have been proposed for multidimensional simplicial meshes using one-dimensional interpolation. The simplest example is the *upwind triangle scheme* as introduced by Billey et al. (1987), Desideri and Dervieux (1988) and Rostand and Stoufflet

(1988) with later improved variants given by Jameson (1993) and Cournède, Debiez and Dervieux (1998). These schemes are not finite volume methods in the sense described in Section 4.3.2 owing to the fact that a single multidimensional gradient is not obtained in each control volume. Referring to Figure 6, the starting point for these methods is the semidiscrete FVM

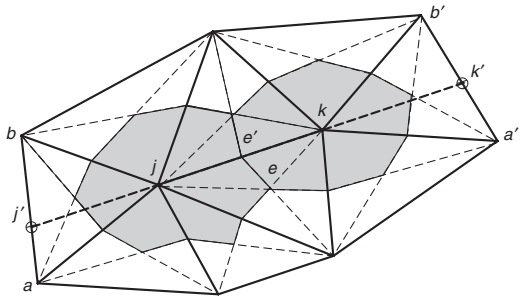


Figure 6. Triangle complex used in the upwind triangle schemes showing the linear extension of $e_{jk}$ into neighboring triangle for the determination of points $x_{j'}$ and $x_{k'}$.

(29) from Section 3.1

$$\frac{d}{\mathrm{d}t}u_j = -\frac{1}{|K_j|}\sum_{\sigma_{j,k}\subset\partial K_j} g(u_j, u_k; \boldsymbol{n}_{j,k})\,|\sigma_{j,k}| \tag{102}$$

for each $K_j \in \mathcal{T}$ with numerical flux function

$$g(u, v; \boldsymbol{n}) = \frac{1}{2}(\boldsymbol{f}(u) + \boldsymbol{f}(v)) \cdot \boldsymbol{n} - \frac{1}{2}|a(u, v; \boldsymbol{n})|(v - u) \tag{103}$$

utilizing the mean value speed satisfying

$$(\boldsymbol{f}(v) - \boldsymbol{f}(u)) \cdot \boldsymbol{n} = a(u, v; \boldsymbol{n})(v - u) \tag{104}$$

With further modifications at sonic points, the modified numerical flux can be shown to be an E-flux. As discussed earlier, this FVM is at most first order accurate for hyperbolic conservation laws. The main idea in the upwind triangle scheme is to add antidiffusion terms to the numerical flux function (103) such that the sum of added diffusion and antidiffusion terms in the numerical flux function vanishes entirely whenever the numerical solution varies linearly over the support of the flux function. The amount of added antidiffusion is determined from maximum principle analysis. The resulting method and maximum principle results are summarized in the following theorem:

**Theorem 23.** *(LED and maximum principles for the upwind triangle scheme) Referring to simplex configuration in Figure 6, let $u_j$ denote the nodal solution value at a simplex vertex $v_j$ in one-to-one correspondence with control volumes $K_j \in \mathcal{T}$. Let $g(u_{j'}, u_j, u_k, u_{k'}; \boldsymbol{n}_{jk})$ denote*

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

*the numerical flux function with limiter function* $\Psi(\cdot)\colon \mathbb{R} \mapsto \mathbb{R}$

$$\begin{aligned} g(u_{j'}, u_j, u_k, u_{k'}) \quad \equiv \quad & \frac{1}{2}(\boldsymbol{f}(u_j) + \boldsymbol{f}(u_k)) \cdot \boldsymbol{n}_{jk} \\ & - \frac{1}{2} a^+(u_j, u_k; \boldsymbol{n}_{jk}) \left(1 - \Psi\left(\frac{h_{jk}\Delta_{j'j}u}{h_{j'j}\Delta_{jk}u}\right)\right)(u_k - u_j) \\ & + \frac{1}{2} a^-(u_j, u_k; \boldsymbol{n}_{jk}) \left(1 - \Psi\left(\frac{h_{jk}\Delta_{kk'}u}{h_{kk'}\Delta_{jk}u}\right)\right)(u_k - u_j) \end{aligned}$$

*utilizing the mean value speed* $a(u_j, u_k; \boldsymbol{n}_{jk})$ *satisfying (104) and variable spacing parameter* $h_{jk} = |x_k - x_j|$. *The semidiscrete FVM*

$$\frac{d}{dt}u_j = -\frac{1}{|K_j|} \sum_{\sigma_{j,k} \subset \partial K_j} g(u_{j'}, u_j, u_k, u_{k'}; \boldsymbol{n}_{j,k})\, |\sigma_{j,k}|, \quad \forall K_j \in \mathcal{T}$$

*with linearly interpolated values* $u_{j'}$ *and* $u_{k'}$ *as depicted in Figure 6 is local extremum diminishing (LED) in the sense of Lemma 2 and exhibits the local spatial maximum principle at steady state* $u^*$

$$\min_{\sigma_{j,k} \subset \partial K_j} u_k^* \le u_j^* \le \max_{\sigma_{j,k} \subset \partial K_j} u_k^*$$

*if the limiter* $\Psi(R)$ *satisfies* $\forall R \in \mathbb{R}$

$$0 \le \frac{[\Psi(R)]}{R}, \quad 0 \le \Psi(R) \le 2$$

Some standard limiter functions that satisfy the requirements of Theorem 23 include

- the MinMod limiter with maximum compression parameter equal to 2
$$\Psi^{\mathrm{MM}}(R) = \max(0, \min(R, 2))$$
- the van Leer limiter
$$\Psi^{\mathrm{VL}}(R) = \frac{R + |R|}{1 + |R|}$$

Other limiter formulations involving three successive one-dimensional slopes are given in Jameson (1993) and Cournède, Debiez and Dervieux (1998).

### *4.4. Higher order accurate time integration schemes*

The derivation of finite volume schemes in Section 3 began with a semidiscrete formulation (29) that was later extended to a fully discrete formulation (31) by the introduction of first-order accurate forward Euler time integration. These latter schemes were then subsequently extended to higher order accuracy in space using a variety of techniques. For many computing problems of interest, first-order accuracy in time is then no longer enough. To overcome this

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

low-order accuracy in time, a general class of higher order accurate time integration methods was developed that preserve stability properties of the fully discrete scheme with forward Euler time integration. Following Gottlieb, Shu and Tadmor (2001) and Shu (2002), these methods will be referred to as *strong stability preserving* (SSP) time integration methods.

Explicit SSP Runge–Kutta methods were originally developed by Shu (1988), Shu and Osher (1988) and Gottlieb and Shu (1998) and called *TVD Runge–Kutta time discretizations*. In a slightly more general approach, total variation bounded (TVB) Runge–Kutta methods were considered by Cockburn *et al.* (1989), Cockburn *et al.* (1989), Cockburn, Hou and Shu (1990) and Cockburn and Shu (1998b) in combination with the discontinuous Galerkin discretization in space. Küther (2000) later gave error estimates for second-order TVD Runge–Kutta finite volume approximations of hyperbolic conservation laws.

To present the general framework of SSP Runge–Kutta methods, consider writing the semidiscrete FVM in the following form

$$\frac{\mathrm{d}}{\mathrm{d}t}U(t) = L(U(t)) \tag{105}$$

where $U = U(t)$ denotes the solution vector of the semidiscrete FVM. Using this notation together with forward Euler time integration yields the fully discrete form

$$U^{n+1} = U^n - \Delta t\ L(U^n) \tag{106}$$

where $U^n$ is now an approximation of $U(t^n)$. As demonstrated in Section 3.3, the forward Euler time discretization is stable with respect to the $L^\infty$-norm, that is,

$$\|U^{n+1}\|_\infty \leq \|U^n\|_\infty \tag{107}$$

subject to a CFL-like time step restriction

$$\Delta t \leq \Delta t_0 \tag{108}$$

With this assumption, a time integration method is said to be SSP (Gottlieb, Shu and Tadmor, 2001) if it preserves the stability property (107), albeit with perhaps a slightly different restriction on the time step

$$\Delta t \leq c\ \Delta t_0 \tag{109}$$

where $c$ is called the CFL coefficient of the SSP method. In this framework, a general objective is to find SSP methods that are higher order accurate, have low computational cost and storage requirements, and have preferably a large CFL coefficient. Note that the TVB Runge-Kutta methods can be embedded into this class if the following relaxed notion of stability is assumed

$$\|U^{n+1}\|_\infty \leq (1 + \mathcal{O}(\Delta t))\ \|U^n\|_\infty \tag{110}$$

*4.4.1. Explicit SSP Runge–Kutta methods.* Following Shu and Osher (1988) and the review articles by Gottlieb, Shu and Tadmor (2001) and Shu (2002), a general $m$-stage Runge–Kutta

*Encyclopedia of Computational Mechanics*. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

method for integrating (105) in time can be algorithmically represented as

$$\begin{aligned} \tilde{U}^0 &:= U^n \\ \tilde{U}^l &:= \sum_{k=0}^{l-1} \big(\alpha_{lk}\tilde{U}^k + \beta_{lk}\Delta t L(\tilde{U}^k)\big), \quad \alpha_{lk} \geq 0,\ l = 1, \ldots, m \\ U^{n+1} &:= \tilde{U}^m \end{aligned} \tag{111}$$

To ensure consistency, the additional constraint $\sum_{k=0}^{l-1} \alpha_{lk} = 1$ is imposed. If, in addition, all $\beta_{lk}$ are assumed to be nonnegative, it is straightforward to see that the method can be written as a convex (positive weighted) combination of simple forward Euler steps with $\Delta t$ replaced by $(\beta_{lk}/\alpha_{lk})\Delta t$. From this property, Shu and Osher (1988) concluded the following lemma:

**Lemma 5.** *If the forward Euler method (106) is $L^\infty$-stable subject to the CFL condition (108), then the Runge–Kutta method (111) with $\beta_{lk} \geq 0$ is SSP, that is, the method is $L^\infty$-stable under the time step restriction (109) with CFL coefficient*

$$c = \min_{l,k} \frac{\beta_{lk}}{\alpha_{lk}} \tag{112}$$

In the case of negative $\beta_{lk}$, a similar result can be proven, see (Shu and Osher, 1988).

*4.4.2. Optimal second- and third-order nonlinear SSP Runge–Kutta methods.* Gottlieb, Shu and Tadmor (2001) (Proposition 3.1) show that the maximal CFL coefficient for any $m$-stage, $m$th order accurate SSP Runge–Kutta methods is $c = 1$. Therefore, SSP Runge–Kutta methods that achieve $c = 1$ are termed 'optimal'. Note that this restriction is not true if the number of stages is higher than the order of accuracy; see Shu (1988).

Optimal second- and third-order nonlinear SSP Runge–Kutta methods are given in Shu and Osher (1988). The optimal second-order, two-stage nonlinear SSP Runge–Kutta method is given by

$$\begin{aligned} \tilde{U}^0 &:= U^n \\ \tilde{U}^1 &:= \tilde{U}^0 + \Delta t\ L(\tilde{U}^0) \\ U^{n+1} &:= \frac{1}{2}\tilde{U}^0 + \frac{1}{2}\tilde{U}^1 + \frac{1}{2}\Delta t\ L(\tilde{U}^1) \end{aligned}$$

This method corresponds to the well-known method of Heun. Similarly, the optimal third-order, three-stage nonlinear SSP Runge–Kutta method is given by

$$\begin{aligned} \tilde{U}^0 &:= U^n \\ \tilde{U}^1 &:= \tilde{U}^0 + \Delta t\ L(\tilde{U}^0) \\ \tilde{U}^2 &:= \frac{3}{4}\tilde{U}^0 + \frac{1}{4}\tilde{U}^1 + \frac{1}{4}\Delta t\ L(\tilde{U}^1) \\ U^{n+1} &:= \frac{1}{3}\tilde{U}^0 + \frac{2}{3}\tilde{U}^2 + \frac{2}{3}\Delta t L(\tilde{U}^2) \end{aligned}$$

Further methods addressing even higher-order accuracy or lower storage requirements are given in the review articles of Gottlieb, Shu and Tadmor (2001) and Shu (2002) where SSP multistep methods are also discussed.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

## 5. Finite Volume Methods for Elliptic and Parabolic Problems

As discussed in Section 1, hyperbolic conservation laws are often approximations to physical problems with small or nearly vanishing viscosity. For some problems, the quantitative solution effects of these small viscosity or diffusion terms are actually sought. For other problems (e.g., porous media problems) the diffusive terms are actually large and the diffusion coefficients (or permeabilities) are often highly spatially dependent. For a linear convection diffusion problem, the flux function $\boldsymbol{f}$ in (1) now becomes $\boldsymbol{f}(u, \nabla u) = -\lambda \nabla u + \boldsymbol{v}u$, $\boldsymbol{v} \in \mathbb{R}^d$ where $\lambda > 0$ is the diffusion coefficient, which may be small or large depending on the problem, may depend on $\boldsymbol{x}$ as in the case of heterogeneous media, or may be replaced by a matrix as in the case of anisotropic media. The flux through a given edge $\sigma$ of a control volume $K$ becomes

$$\int_\sigma \boldsymbol{f}(u) \cdot \boldsymbol{n}_{K,\sigma} \, \mathrm{d}s = \int_\sigma (-\lambda \nabla u \cdot \boldsymbol{n}_{K,\sigma} + \boldsymbol{v} \cdot \boldsymbol{n}_{K,\sigma} u) \, \mathrm{d}s \tag{113}$$

so that the additional term $\int_\sigma -\lambda \nabla u \cdot \boldsymbol{n}_{K,\sigma} \, \mathrm{d}s$ must now be discretized.

Emphasis herein is given to analysis of TPFA methods. On the one hand, these methods have specific monotonicity properties that allow one to cope with noncoercive problems and irregular data. On the other hand, they assume a discretization of the Laplace operator on so-called $\Delta$-admissible meshes. Such meshes include the Voronoï tessellation obtained from a set of vertices of a mesh as depicted in Figure 7. Other methods that apply on general grids

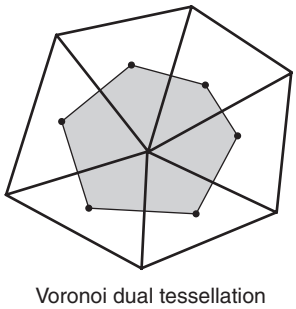


Figure 7. Control volume obtained by the Voronoï dual mesh. Edges of the Voronoï dual are perpendicular to the edges of the triangulation.

for anisotropic and heterogeneous diffusion operators are outlined in Section 5.2.

### *5.1. Convergence analysis for the steady state reaction convection diffusion equation*

*5.1.1. The continuous and discrete problems.* Let $\Omega$ be an open bounded polygonal subset of $\mathbb{R}^d$, $d = 2$ or 3, $f \in L^2(\Omega)$, $\boldsymbol{v} \in \mathbb{R}^d$ and $b \in \mathbb{R}$, and consider the following steady state linear reaction convection diffusion equation:

$$-\Delta u + \mathrm{div}(\boldsymbol{v}u) + bu = f \text{ on } \Omega \tag{114}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

together with homogeneous boundary conditions on $\partial\Omega$. A weak formulation of this problem is given by

$$\left\{ \begin{array}{l} \text{Find } u \in H_0^1(\Omega) \text{ such that} \\ \displaystyle\int_\Omega \nabla u \cdot \nabla \phi \, \mathrm{d}\boldsymbol{x} + \int_\Omega \mathrm{div}(\boldsymbol{v}u)\phi \, \mathrm{d}\boldsymbol{x} + \int_\Omega bu\phi \, \mathrm{d}\boldsymbol{x} = \int_\Omega f\phi \, \mathrm{d}\boldsymbol{x} \quad \forall \phi \in H_0^1(\Omega) \end{array} \right. \tag{115}$$

In order to discretize the above reaction convection diffusion equation, a "$\Delta$-admissible" mesh of $\Omega$ is used in discretizing the Laplace operator. A $\Delta$-admissible mesh $\mathcal{T}$ is composed of control volumes with faces (or edges) denoted by $\mathcal{E}$ and a set of points $\mathcal{P} = (x_K)_{K\in\mathcal{T}}$ chosen such that the following orthogonality condition holds: for an interface $\sigma_{KL}$ separating the cells $K$ and $L$, the line segment $x_K x_L$ is orthogonal to this interface; see Figure 8. This orthogonality condition on the mesh is used to prove the consistency of the flux as detailed in the following section. Such a family of orthogonal points exists, for instance, in the case of triangles, rectangles, or Voronoï tessellations; see Eymard *et al.* (2000) for more details. Unfortunately, this orthogonality condition is not always satisfied for general meshes so that certain theoretical results given below are not valid. To address this problem, some schemes designed specially for general meshes are discussed in Section 5.2.

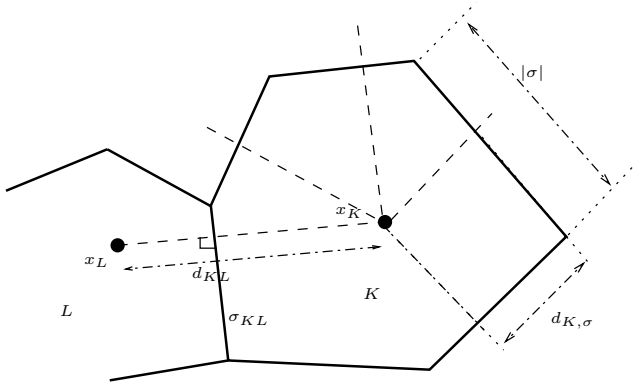


Figure 8. Notations for a control volume.

The classical flux form of the finite volume scheme is

$$\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(u_\mathcal{T}) + b|K|u_K = |K|f_K \quad \forall K \in \mathcal{T} \tag{116}$$

where $|K|$ denotes the $d$-dimensional Lebesgue measure of $K$, $f_K$ is the mean value of $f$ over

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

$K$, and $F_{K,\sigma}(u_\mathcal{T})$ is the numerical flux given by

$$F_{K,\sigma}(u_\mathcal{T}) = \begin{cases} -\frac{|\sigma|}{d_{KL}}(u_L - u_K) + v^+_{K,\sigma}u_K - v^-_{K,\sigma}u_L & \text{if } \sigma = \sigma_{KL} \\ -\frac{|\sigma|}{d_{K,\sigma}}(-u_K) + v^+_{K,\sigma}u_K & \text{if } \sigma \text{ is an edge of } K \text{ on } \partial\Omega \end{cases} \tag{117}$$

with $v_{K,\sigma} = |\sigma|\boldsymbol{v}\cdot\boldsymbol{n}_{K,\sigma}$, $v^+_{K,\sigma} = \max(v_{K,\sigma}, 0)$ and $v^-_{K,\sigma} = -\min(v_{K,\sigma}, 0)$. One important feature of FVMs is numerical flux consistency, namely that $F_{K,\sigma}(u_\mathcal{T})$ is a consistent approximation of the exact flux, $\bar{F}_{K,\sigma}(u) = \int_\sigma \nabla u \cdot \boldsymbol{n}_{K,\sigma}$. Assuming a regular exact solution $u$, let $F^*_{K,\sigma}(u)$ denote the flux obtained by replacing the discrete unknowns $u_K$ by the values of the exact solution $u(\boldsymbol{x}_K)$. Then, the numerical flux (117) is consistent in the sense

$$|F^*_{K,\sigma}(u) - \bar{F}_{K,\sigma}(u)| \to 0 \text{ as } h_\mathcal{T} \to 0$$

with $h_\mathcal{T} = \max_{K\in\mathcal{T}} \mathrm{diam}K$. Note that the discrete Laplace operator $\Delta_\mathcal{T} u_\mathcal{T}$ resulting from these numerical diffusion fluxes (see the precise formulas in (120) below) is not consistent in a finite difference sense. This is demonstrated using a simple Laplace operator example with a nonuniform one-dimensional mesh in Eymard *et al.*, 2000, Chapter 2, Example 2.1. Even so, this discretization is consistent in a weak sense. Specifically, thanks to conservation and the consistency of the flux, it can be proved that the consistency error of the Laplace operator tends to zero in a $L^\infty$ weak-star topology; see Eymard *et al.*, 2000, Chapter 2, Remark 2.9. The convergence analysis relies on the fact that the discrete Laplace operator is consistent for some discrete dual $H^1_0$ norm as addressed in Lemma 8.

Note that the system (116) and (117) always leads to a linear system containing $N$ equations and $N$ unknowns with $N = \mathrm{card}(\mathcal{T})$. This linear system may be written as

$$\sum_{L\in\mathcal{T}} A_{K,L}u_L = |K|f_K \text{ for all } K \in \mathcal{T} \tag{118}$$

with

$$\begin{aligned} A_{K,K} &= \sum_{\sigma\in\mathcal{E}_K} \left(\frac{|\sigma|}{d_\sigma} + |\sigma|v^+_{K,\sigma}\right) + b|K| \\ A_{K,L} &= -\frac{|\sigma|}{d_\sigma} - |\sigma|v^-_{K,\sigma} \text{ with } \sigma = K|L \\ A_{K,L} &= 0 \text{ if } K \text{ and } L \text{ do not share an interface} \end{aligned}$$

The finite volume scheme may also be written in the following equivalent weak form:

$$\begin{cases} \text{Find } u_\mathcal{T} \in H_\mathcal{T}(\Omega) \text{ such that} \\ \qquad [u_\mathcal{T}, \phi]_\mathcal{T} + c_\mathcal{T}(u_\mathcal{T}, \phi) + \displaystyle\int_\Omega b u_\mathcal{T}\phi \,\mathrm{d}\boldsymbol{x} = \int_\Omega f\phi \,\mathrm{d}\boldsymbol{x},\ \forall \phi \in H_\mathcal{T}(\Omega) \end{cases} \tag{119}$$

where $H_\mathcal{T}(\Omega)$ is the space of piecewise constant functions on the control volumes of $\mathcal{T}$. The inner product $[\cdot,\cdot]_\mathcal{T}$ is defined by

$$[u, \phi]_\mathcal{T} = \sum_{\sigma_{KL}\in\mathcal{E}_{\mathrm{int}}} \frac{|\sigma_{KL}|}{d_{KL}}(u_L - u_K)(\phi_L - \phi_K) + \sum_{\sigma\in\mathcal{E}_{\mathrm{ext}}} \frac{|\sigma|}{d_{K,\sigma}}\, u_K\ \phi_K$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

where $\mathcal{E}_{\rm int}$ (resp. $\mathcal{E}_{\rm ext}$, $\mathcal{E}_K$) denotes the set of edges (or faces) included in $\Omega$ (resp. $\partial\Omega$, $\partial K$), $|\sigma|$ the $(d-1)$–dimensional Lebesgue measure of $\sigma$, $d_{KL}$ the distance between $x_K$ and $x_L$ (Figure 8), and $d_{K,\sigma}$ the distance between $x_K$ and $\sigma$. In the first summation, $\sigma_{KL}$ denotes the edge separating the control volumes $K$ and $L$, and in the last summation, the volume $K$ is the unique volume in which $\sigma$ is an edge. To complete the weak formulation, the bilinear convective form is defined by

$$c_{\mathcal{T}}(u_{\mathcal{T}},\phi)=\sum_{K\in\mathcal{T}}\phi_K\left[\sum_{\sigma_{KL}\in\mathcal{E}_K}(v^+_{K,\sigma_{KL}}u_K-v^-_{K,\sigma_{KL}}u_L)+\sum_{\sigma\in\mathcal{E}_K\cap\mathcal{E}_{\rm ext}}v^+_{K,\sigma}u_K\right]$$

Taking $\phi=1_K$ in (119), it is easily seen that (119) implies (116). Conversely, let $\phi\in H_{\mathcal{T}}(\Omega)$. Multiplying (116) by $\phi_K$, summing the resulting equations for all $K\in\mathcal{T}$, and reordering the summations yields (119).

One may also define a discrete Laplace operator in $H_{\mathcal{T}}$ in the following way. For $\phi\in H_{\mathcal{T}}$, let $\Delta_{\mathcal{T}}\phi\in H_{\mathcal{T}}$ be defined by

$$(\Delta_{\mathcal{T}}\phi)_K=-\frac{1}{|K|}\sum_{\sigma\in\mathcal{E}_K}F^{(d)}_{K,\sigma}(\phi) \tag{120a}$$

$$F^{(d)}_{K,\sigma}(\phi)=\begin{cases}\frac{|\sigma|}{d_{KL}}(\phi_K-\phi_L) & \text{if } \sigma=\sigma_{KL}\\ \frac{|\sigma|}{d_{KL}}(\phi_K) & \text{if } \sigma\subset\partial\Omega\end{cases} \tag{120b}$$

Then, using the conservation property of the flux ($F_{K,\sigma}=-F_{L,\sigma}$ if $\sigma=\sigma_{KL}$), it follows that

$$[u,\phi]_{\mathcal{T}}=-\int_\Omega\Delta_{\mathcal{T}}u\;\phi\,{\rm d}\boldsymbol{x}=-\int_\Omega u\;\Delta_{\mathcal{T}}\phi\,{\rm d}\boldsymbol{x},\;\forall\;u,\phi\in H_{\mathcal{T}}(\Omega) \tag{121}$$

Using the following Poincaré inequality that holds for $u\in H_{\mathcal{T}}$ (see e.g. Eymard *et al.*, 2000, Lemma 9.1):

$$\|u\|_{L^2(\Omega)}\le{\rm diam}(\Omega)\;||u||_{1,\mathcal{T}} \tag{122}$$

a mesh dependent "discrete $H^1_0$ norm" may be defined using the inner product introduced above, that is,

$$\|u\|_{1,\mathcal{T}}=([u,u]_{\mathcal{T}})^{1/2}=\Big(\sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}}\frac{|\sigma|}{d_{KL}}(u_L-u_K)^2+\sum_{\sigma\in\mathcal{E}_{\rm ext}}\frac{|\sigma|}{d_{K,\sigma}}\;u_K^2\Big)^{1/2} \tag{123}$$

*5.1.2. Monotonicity of the scheme.* Elliptic PDEs are known to satisfy certain maximum principle properties. An important property is positivity of the solution for a convection diffusion problem obtained from Problem (114) with $b\ge 0$ and $f\ge 0$ together with a homogeneous Dirichlet boundary condition. Let $u$ be the solution to this problem; then $u\ge 0$. Note that this has been proved even in the noncoercive case; see Droniou, 2002. Moreover, a more classical result states that if $u$ is the solution to Problem (114) with $b=0$, $f=0$, and ${\rm div}\,\boldsymbol{v}=0$ (in that case the problem is coercive) with the nonhomogeneous Dirichlet boundary condition

$$u=u_b\text{ on }\partial\Omega \tag{124}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

where $u_b$ is a smooth function from $\Omega$ to $\mathbb{R}$, then $\inf_{\partial\Omega} u_b \leq u \leq \sup_{\partial\Omega} u_b$, see e.g. Evans, 2010, Chapter 6 .

These properties are important in physical applications; for such applications, numerical schemes are sought that preserve these properties. This is indeed the case for the associated finite volume scheme, which is given by (116) with the following modified numerical fluxes (because of the nonhomogeneous boundary condition):

$$F_{K,\sigma}(u_{\mathcal{T}}) = \begin{cases} -\frac{|\sigma|}{d_{KL}}(u_L - u_K) + v^+_{K,\sigma} u_K - v^-_{K,\sigma} u_L & \text{if } \sigma = \sigma_{KL} \\ -\frac{|\sigma|}{d_{K,\sigma}}(u_{b,\sigma} - u_K) + v^+_{K,\sigma} u_K & \text{if } \sigma \text{ is an edge of } K \text{ on } \partial\Omega \end{cases} \tag{125}$$

where $u_{b,\sigma} = u_b(\boldsymbol{x}_\sigma)$.

**Lemma 6 (Positivity)** *If $b \geq 0$, then the matrix of the scheme* (116)-(125) *is an inverse-positive matrix, that is, it is invertible and its inverse is nonnegative (all its coefficients are nonnegative). As a consequence, if $f_K \geq 0$ for all $K \in \mathcal{T}$, and $u_{b,\sigma} \geq 0$ for all $\sigma \in \mathcal{E}_{\rm ext}$, then the solution $(u_K)_{K\in\mathcal{T}}$ of* (116) *and* (117) *(with $b \geq 0$) satisfies $u_K \geq 0$ for all $K \in \mathcal{T}$.*

The proof of this result can be found in Eymard, Gallouët and Herbin (2002) in the case $\operatorname{div} \boldsymbol{v} = 0$. In the case where the sign of $\operatorname{div} \boldsymbol{v}$ is not known (and the problem is therefore not coercive), the result is still true, and relies on the fact that the matrix $A$ of the scheme is irreducible and diagonally dominant by column, see for example, Fettah and Gallouët, 2013.

The positivity property immediately yields the existence and uniqueness of the solution of the numerical scheme (119), which can also be proved directly thanks to an $L^2$ *a priori* estimate on the approximate solution.

**Lemma 7 (Discrete maximum principle)** *If $f_K = 0$ for all $K \in \mathcal{T}$, $b = 0$ and $\operatorname{div} \boldsymbol{v} = 0$, then the solution $(u_K)_{K\in\mathcal{T}}$ of* (116)–(125) *satisfies*

$$\min_{\sigma\in\mathcal{E}_{\rm ext}} u_{b,\sigma} \leq u_K \leq \max_{\sigma\in\mathcal{E}_{\rm ext}} u_{b,\sigma} \text{ for all } K \in \mathcal{T}$$

The maximum principle can be deduced from the fact that the matrix $A$ of the system is diagonal dominant by row (note that this is only true if $\operatorname{div} \boldsymbol{v} = 0$).

*5.1.3. Convergence results.* The mathematical analysis of any numerical scheme must address the question of existence of a solution (which is rather straightforward here since the problem is linear) and the question of convergence (i.e., "does the approximate solution converge to the solution of the continuous problem as the mesh size tends to 0?"). A related question concerns obtaining a rate of convergence through error estimates that are usually dependent on regularity assumptions on the continuous solution. The proof of the convergence of the finite volume scheme for a semilinear equation generalizing (114) was first proven in Eymard, Gallouët, and Herbin, 1999 (see also Eymard *et al.* (2000)). The result will be stated here for the linear case and the main steps of the proof explained, since the techniques extend to nonlinear problems.

Under the assumptions of Lemma 6, it is easily seen that the system (119) (resp. (116)) has a unique solution $u_{\mathcal{T}} \in H_{\mathcal{T}}$ (resp. $(u_K)_{K\in\mathcal{T}}$). Let $(\mathcal{T}_n)_{n\in\mathbb{N}}$ be a sequence of finite volume discretizations satisfying the orthogonality condition and let $h_{\mathcal{T}_n}$ be the size of the mesh $\mathcal{T}_n$, that is the maximum of the diameters of the control volumes of $\mathcal{T}_n$. In this case, assuming that $h_{\mathcal{T}_n} \to 0$ as $n \to +\infty$, the corresponding sequence $(u_{\mathcal{T}_n})_{n\in\mathbb{N}}$ can be shown to converge in $L^2(\Omega)$ to the unique solution of (115). The proof of this result may be decomposed into four steps:

1. *A priori* estimates on the approximate solution in the $H_{\mathcal{T}}$ norm and the $L^2$ norm are obtained which yield existence (and uniqueness) of $u_{\mathcal{T}}$ solution of the scheme. These estimates entail the weak convergence of $(u_{\mathcal{T}_n})_{n\in\mathbb{N}}$ in $L^2(\Omega)$, up to a subsequence, to some $\bar{u} \in L^2(\Omega)$.
2. Strong convergence and regularity of the limit, that is $\bar{u} \in H_0^1(\Omega)$, are obtained through a discrete Rellich theorem, described below.
3. The fact that the limit $\bar{u}$ is a weak solution of the continuous problem is obtained by a passage to the limit in the scheme as $h_{\mathcal{T}}$ tends to zero.
4. A classical uniqueness argument is then used to show that the whole sequence converges.

Note that there is no need to assume the existence of the solution to the continuous problem as it is obtained as a by-product of the convergence of the scheme. In the present linear case, this is not necessary, since existence is well known even in the non-coercive case; see Droniou, 2002. For more complicated nonlinear problems, obtaining the existence of the solution *via* the convergence of the numerical scheme can be useful (see e.g. Bouillard *et al.*, 2007).

For the sake of simplicity, these four steps will be detailed in the following paragraphs for the pure diffusion operator. A sketch of the proof is provided for order $h$ convergence in $L^2$ and $H_{\mathcal{T}}$ norms assuming regularity conditions on the solution, namely $u \in H^2(\Omega)$. Note that the upwind scheme for the convection flux does not lead to any additional difficulty; see Eymard, Gallouët, and Herbin (1999); Gallouët, Herbin and Vignal (2000). Order 2 convergence in the $L^2$ norm may be proved for the pure diffusion operator on uniform grids. However, the same result on triangular meshes, which is observed in numerical experiments, remains an open problem. Recall that higher convergence rates in weaker norms (including this special case) are known and proved for most Galerkin methods via duality arguments (the so-called Aubin-Nitsche lemma, Ciarlet (1991)).

*5.1.4. A priori estimate.*

**Definition 10 (Discrete $H^{-1}$ norm)** *Let $\psi \in H_{\mathcal{T}}(\Omega)$, then its discrete $H^{-1}$ norm is defined by*

$$\|\psi\|_{-1,\mathcal{T}} = \sup_{v\in H_{\mathcal{T}}(\Omega), v\neq 0} \frac{\displaystyle\int_\Omega \psi v \,\mathrm{d}\boldsymbol{x}}{\|v\|_{1,\mathcal{T}}} \tag{126}$$

*Note that, by the discrete Poincaré inequality* (122)*,*

$$\|\psi\|_{-1,\mathcal{T}} \leq \operatorname{diam}(\Omega)\ \|\psi\|_{L^2(\Omega)}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Using the formulation (119) with $\boldsymbol{v} = 0$ and $b = 0$, the finite volume scheme can written as

$$[u_{\mathcal{T}}, v]_{\mathcal{T}} = \int_{\Omega} fv \, \mathrm{d}\boldsymbol{x}, \; \forall v \in H_{\mathcal{T}}(\Omega)$$

Choosing $v = u_{\mathcal{T}}$, by definition (126), one obtains

$$||u_{\mathcal{T}}||_{1,\mathcal{T}} \leq \|f\|_{-1,\mathcal{T}} \tag{127}$$

Taking $f = 0$ yields the uniqueness (and therefore the existence) of the discrete solution. This estimate also yields weak convergence of a subsequence of approximate solutions in $L^2(\Omega)$.

*5.1.5. Convergence theorem.* In order to prove strong convergence of the approximate solutions, some control on solution oscillations is needed. In the finite element framework, the family of approximate solutions is bounded in $H^1(\Omega)$ so that one may use the Rellich theorem to obtain compactness in $L^2(\Omega)$. This is too restrictive and not used here. However, the Rellich theorem derives from the Kolmogorov theorem, which gives a necessary and sufficient condition for a bounded family of $L^p(\Omega)$, $p < +\infty$, to be relatively compact. Thus, the Kolmogorov theorem is an adequate tool in finite volume analysis. In order to use it, some estimates on the translates of functions of $H_{\mathcal{T}}(\Omega)$ are needed. Indeed, one may show in a way that is similar to that of the continuous case (replacing the derivatives by differences) that for any function $v \in H_{\mathcal{T}}(\Omega)$

$$\|v(\cdot + \eta) - v\|^2_{L^2(\Omega)} \leq |\eta| \;\; (|\eta| + 4h_{\mathcal{T}}))\, ||v||^2_{1,\mathcal{T}} \quad \forall \eta \in \mathbb{R}^d$$

This estimate is then used in obtaining the following result:

**Theorem 24 (Discrete Rellich theorem)** *Let $(\mathcal{T}_n)_{n\in\mathbb{N}}$ be a sequence of finite volume discretizations satisfying the orthogonality condition, such that $h_{\mathcal{T}_n} \to 0$. Let $(u_n)_{n\in\mathbb{N}} \subset L^2(\Omega)$ such that $u_n \in H_{\mathcal{T}_n}$ and $\|u_n\|_{1,\mathcal{T}_n} \leq C$ where $C \in \mathbb{R}$. Then, there exists a subsequence $(u_n)_{n\in\mathbb{N}}$ and $\bar{u} \in H^1_0(\Omega)$ such that $u_n \to \bar{u}$ in $L^2(\Omega)$ as $n \to +\infty$.*

From the discrete $H^1$ estimate (127), the above theorem yields the strong convergence of a subsequence of the approximate solutions in $L^2(\Omega)$ to some function $\bar{u} \in H^1_0(\Omega)$.

*5.1.6. Passage to the limit in the scheme.* One now needs to show that the limit $\bar{u}$ is a solution to the continuous variational problem. Let $(\mathcal{T}_n)$ be a sequence of discretizations such that $h_{\mathcal{T}_n} \to 0$. For each mesh $\mathcal{T}_n$, the finite volume scheme is given by

$$[u_{\mathcal{T}_n}, v]_{\mathcal{T}_n} = \int_{\Omega} fv \, \mathrm{d}\boldsymbol{x}, \; \forall v \in H_{\mathcal{T}_n}(\Omega) \tag{128}$$

**Lemma 8 (Consistency of the discrete Laplace operator)** *Let $\mathcal{T}$ be a finite volume mesh satisfying the orthogonality condition. Denote by $P_{\mathcal{T}}$ and $\Pi_{\mathcal{T}}$ the following interpolation operators:*

$$P_{\mathcal{T}} \; : \; C(\Omega) \to H_{\mathcal{T}}(\Omega), \quad P_{\mathcal{T}}\varphi(x) = \varphi(x_K), \quad \forall x \in K, \forall K \in \mathcal{T} \tag{129}$$

$$\Pi_{\mathcal{T}} \; : \; L^2(\Omega) \to H_{\mathcal{T}}(\Omega), \quad \Pi_{\mathcal{T}}\varphi(\boldsymbol{x}) = \frac{1}{|K|}\int_K \varphi \, \mathrm{d}\boldsymbol{x}, \quad \forall \boldsymbol{x} \in K, \forall K \in \mathcal{T} \tag{130}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

*For $\varphi \in C_c^\infty(\Omega)$, define the consistency error $R_{\Delta,\mathcal{T}}(\varphi) \in H_\mathcal{T}(\Omega)$ on the discrete Laplace operator by*

$$R_{\Delta,\mathcal{T}}(\varphi) = \Delta_\mathcal{T} P_\mathcal{T}\varphi - \Pi_\mathcal{T}(\Delta\varphi)$$

*Then, there exists $C_\varphi$ depending only on $\varphi$ such that*

$$\|R_{\Delta,\mathcal{T}}(\varphi)\|_{-1,\mathcal{T}} \leq C_\varphi h_\mathcal{T} \tag{131}$$

*for any $h_\mathcal{T}$ sufficiently small (i.e., to say, smaller than the distance between the support of $\varphi$ and the boundary $\partial\Omega$).*

**Proof:** For $\varphi \in C_c^\infty(\Omega)$, one has

$$\|R_{\Delta,\mathcal{T}}(\varphi)\|_{-1,\mathcal{T}} = \sup_{v\in H_\mathcal{T}(\Omega),\|v\|_{1,\mathcal{T}}=1} X(v)$$

with

$$X(v) = \sum_{K\in\mathcal{T}} |K| \left[(\Delta_\mathcal{T} P_\mathcal{T}\varphi)_K\, v_K - (\Pi_\mathcal{T}(\Delta\varphi))_K\, v_K\right]$$

For $h_\mathcal{T}$ small enough, $\varphi$ vanishes in all the control volumes having an edge on the boundary of the domain so that, by the definition of $\Delta_\mathcal{T}$, $P_\mathcal{T}$, and $\Pi_\mathcal{T}$

$$\begin{aligned} X(v) &= \sum_{K\in\mathcal{T}} v_K \left[\sum_{\sigma\in\mathcal{E}_K} F_{K,\sigma}(P_\mathcal{T}\varphi) - \int_\sigma \nabla\varphi\cdot\boldsymbol{n}_{K,\sigma}\,\mathrm{d}s\right] \\ &= \sum_{\sigma_{KL}\in\mathcal{E}_{\rm int}} |\sigma|\; R_{K,\sigma}(\varphi)\;(v_K - v_L) \end{aligned} \tag{132}$$

where $R_{K,\sigma}(\varphi)$ is the consistency error on the fluxes, defined by

$$R_{K,\sigma}(\varphi) = \frac{1}{|\sigma|}\big(F_{K,\sigma}(P_\mathcal{T}\varphi) - \int_\sigma \nabla\varphi\cdot\boldsymbol{n}_{K,\sigma}\,\mathrm{d}s\big)$$

The property of consistency of the fluxes states that for a regular function $\varphi$, there exists $c_\varphi \in \mathbb{R}$ depending only on $\varphi$ such that

$$|R_{K,\sigma}(\varphi)| \leq c_\varphi h_\mathcal{T}$$

This result, proved in Eymard *et al.* (2000), is a central argument of the proof. It relies on the orthogonality condition for the mesh, and is obtained by Taylor expansions. From the Cauchy–Schwarz inequality and (132), it follows that

$$X(v) \leq C_\varphi h_\mathcal{T} \|v\|_{1,\mathcal{T}}$$

which concludes the proof.

An immediate consequence is the following corollary:

**Corollary 2.** *Let $(\mathcal{T}_n)_{n\in\mathbb{N}}$ be a family of meshes satisfying the orthogonality property such that $h_{\mathcal{T}_n} \to 0$. Let $(u_{\mathcal{T}_n})_{n\in\mathbb{N}} \subset L^2(\Omega)$ and $\bar{u} \in H^1(\Omega)$ such that $\|u_{\mathcal{T}_n}\|_{1,\mathcal{T}} \leq C$, where $C \in \mathbb{R}_+$ and $u_{\mathcal{T}_n} \to \bar{u}$ in $L^2(\Omega)$ as $n \to +\infty$, then*

$$\int_\Omega u_{\mathcal{T}_n}\; \Delta_{\mathcal{T}_n}(P_{\mathcal{T}_n}\varphi)\,\mathrm{d}\boldsymbol{x} \to \int_\Omega \bar{u}\;\Delta\varphi\,\mathrm{d}\boldsymbol{x} \text{ as } n\to+\infty,\ \forall\ \varphi\in C_c^\infty(\Omega)$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

A sketch of the proof of convergence of the scheme is now given. Set $v = P_{\mathcal{T}_n}\varphi$ in (128). From (121), it follows that

$$-\int_\Omega u_{\mathcal{T}_n}\Delta_{\mathcal{T}_n}(P_{\mathcal{T}_n}\varphi)\,\mathrm{d}\boldsymbol{x} = \int_\Omega f P_{\mathcal{T}_n}\varphi\,\mathrm{d}\boldsymbol{x}$$

Using Corollary 2 and the fact that the right-hand side converges to $\int_\Omega \varphi\,\mathrm{d}\boldsymbol{x}$, passing to the limit as $n \to +\infty$ yields

$$-\int_\Omega \bar{u}\Delta\varphi\,\mathrm{d}\boldsymbol{x} = \int_\Omega f\varphi\,\mathrm{d}\boldsymbol{x}$$

Using the discrete Rellich theorem, it was shown previously that $\bar{u} \in H_0^1(\Omega)$; thus it can be concluded that $\bar{u}$ is the solution to (115).

*5.1.7. Error analysis.* An error estimate for convection diffusion equations was first obtained in Herbin, 1995 in the case of continuous data and triangular meshes. It was extended to $L^2$ data, general admissible meshes, and general boundary conditions in Gallouët, Herbin and Vignal (2000). The key argument for the error analysis is the fact that Lemma (8) still holds under regularity assumptions for the mesh for $\phi$ in $H^2(\Omega)$. From the variational form of the scheme (128), it follows that

$$[u_{\mathcal{T}_n} - P_{\mathcal{T}_n}u, v]_{\mathcal{T}_n} = \int_\Omega f v\,\mathrm{d}\boldsymbol{x} - [P_{\mathcal{T}_n}u, v]_{\mathcal{T}_n} \quad \forall v \in H_{\mathcal{T}_n}(\Omega)$$

where $u$ is the solution to the continuous problem. Integrating the continuous equation $-\Delta u = f$ over each control volume, in order to replace the first term of the right-hand side of the above relation, yields

$$[u_{\mathcal{T}_n} - P_{\mathcal{T}_n}u, v]_{\mathcal{T}_n} = \int_\Omega R_{\Delta,\mathcal{T}_n}(u) v\,\mathrm{d}\boldsymbol{x} \quad \forall v \in H_{\mathcal{T}_n}(\Omega)$$

A first-order convergence result in the $H_{\mathcal{T}}$ norm then follows from the stability estimate (127). First-order convergence is also obtained in the $L^2$ norm, thanks to the discrete Poincaré inequality. Numerical evidence shows that second-order convergence is obtained for several types of mesh. The mathematical proof is straightforward in the case of rectangular meshes, since in this case, the consistency error on the flux is of order two. Second-order convergence for general meshes remains an open question and is the object of ongoing works; see Omnes, 2011 and Droniou and Nataraj, 2016.

### *5.2. Discretization of anisotropic elliptic problems*

A wide variety of schemes has been developed in the last few years for the numerical simulation of anisotropic diffusion equations on general meshes, see Herbin and Hubert, 2008; Eymard *et al.*, 2011; Droniou, 2014; Lipnikov *et al.*, 2014; Droniou *et al.*, 2016 and references therein. The rigorous analysis of these methods is useful to avoid designing numerical schemes that seem to be well defined and robust, but which in the end do not converge to solutions of the proper model. Such an example of a nonconverging TP flux method may be found in Faille, 1992b, Chapter III, Section 3.2 in the context of porous media.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

For simplicity, the following standard elliptic problem is discretized over a convex polygonal domain $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3:

$$-\nabla \cdot \Lambda \nabla u \quad = f \text{ in } \Omega \tag{133a}$$

$$u(x) \quad = 0 \text{ on } \partial\Omega \tag{133b}$$

for $\Lambda \in \mathbb{R}^{d\times d}$, a symmetric positive definite (SPD) matrix (assumed constant for the sake of simplicity) and $f \in L^2(\Omega)$. The flux balance form of problem (133a) is given by

$$-\int_{\partial K} \Lambda \nabla u \cdot \boldsymbol{n}_K \, \mathrm{d}s = \int_K f \, \mathrm{d}\boldsymbol{x} \tag{134}$$

The solution to problem (133) is understood here in a weak sense. Specifically, there exists a unique weak solution $u \in H_0^1(\Omega)$ satisfying

$$\int_\Omega \Lambda \nabla u(\boldsymbol{x}) \cdot \nabla v(\boldsymbol{x}) = \int_\Omega g(\boldsymbol{x}) v(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \quad \forall v \in H_0^1(\Omega) \tag{135}$$

For heterogeneous anisotropic problems such as (133), a variety of recently developed schemes are based on the definition of a numerical flux which uses the cell centered unknowns together with some other unknowns that are located at the edges of the mesh. Some of these methods permit the explicit elimination of the edge unknowns while ensuring the following properties:

(P1) The schemes must apply on any type of grid: conforming or nonconforming, 2D or 3D (see for instance the frameworks of kinetic formulations or financial mathematics), and made with control volumes that are only assumed to be polyhedral (the boundary of each control volume is a finite union of subsets of hyperplanes).
(P2) The matrices of the generated linear systems are expected to be sparse, symmetric, and positive definite.
(P3) One should able to prove the convergence of the discrete solution and an associated gradient to the solution of the continuous problem and its gradient with no regularity assumption on the solution of the continuous problem as well as show error estimates if the continuous solution is regular enough.

The idea of some of the schemes presented below is to find an approximation of the solution of (133) by setting up a system of discrete equations for a family of values $((u_K)_{K\in\mathcal{T}}, (u_\sigma)_{\sigma\in\mathcal{E}})$ in the control volumes and on the interfaces. The values $u_\sigma$ on the interfaces are introduced so as to allow for a natural consistent approximation of the normal fluxes in the case of an anisotropic operator and a general (possibly nonconforming) mesh. Counting both cell and interface unknowns yields $\mathrm{card}(\mathcal{T}) + \mathrm{card}(\mathcal{E})$ as the total number of unknowns. But in some cases, the unknowns $(u_\sigma)_{\sigma\in\mathcal{E}}$ can be eliminated. In fact, the edge unknowns can always be eliminated by a suitable procedure (discussed below), but the resulting scheme is not always optimal in terms of flux accuracy or robustness.

Following the idea of the finite volume framework, consider the flux balance form (134). The integral boundary $\partial K$ is decomposed as the sum of integrals over the interfaces of $\mathcal{E}$ which

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

compose $\partial K$

$$\sum_{\sigma\in\mathcal{E}_K}\left(-\int_\sigma \Lambda\nabla u(\boldsymbol{x})\cdot\boldsymbol{n}_{K,\sigma}\right)\mathrm{d}s=\int_K f\mathrm{d}\boldsymbol{x}$$

The flux $-\int_\sigma \Lambda\nabla u(\boldsymbol{x})\cdot\boldsymbol{n}_{K,\sigma}\,\mathrm{d}s$ is approximated by a function $F_{K,\sigma}(u)$ of the values $((u_K)_{K\in\mathcal{T}},(u_\sigma)_{\sigma\in\mathcal{E}})$ at the centers of the control volumes and on interfaces.

*5.2.1. TPFA scheme.* Assume first that $\Lambda=\lambda\,\mathrm{Id}$ where $\lambda$ is a given scalar function so that the diffusion is heterogeneous and isotropic. This is a problem of great interest in petroleum reservoir engineering where the permeability of the medium is generally constant on each cell of the grid (but may vary a lot from one cell to another). The so-called TPFA is a straightforward extension of the finite volume scheme defined by the numerical fluxes (120b) for the isotropic homogeneous problem to the heterogeneous diffusion problem. Begin by setting

$$F_{K,\sigma}(u)=-|\sigma|\lambda_K\frac{u_\sigma-u_K}{d_{K,\sigma}}\quad\forall K\in\mathcal{T},\ \forall\sigma\in\mathcal{E}_K,\tag{136}$$

where $\lambda_K$ is the mean value of $\lambda$ on $K$. The unknowns $u_\sigma$ may be eliminated by using conservation of the numerical flux, that is, $F_{L,\sigma}=-F_{K,\sigma}$ if $\sigma=\sigma_{KL}\subset\Omega$. If $\boldsymbol{x}_L\notin\sigma$; this yields the following value of $u_\sigma$

$$u_\sigma=\frac{1}{\frac{\lambda_{K,\sigma}}{d_{K,\sigma}}+\frac{\lambda_{L,\sigma}}{d_{L,\sigma}}}\Big(\frac{\lambda_{K,\sigma}}{d_{K,\sigma}}u_K+\frac{\lambda_{L,\sigma}}{d_{L,\sigma}}u_L\Big)$$

If $\boldsymbol{x}_L\in\sigma$, then $u_\sigma=u_K$. Plugging these expressions in the flux (136) yields the following numerical flux involving the harmonic mean of $\lambda$:

$$F_{K,\sigma}=-\tau_\sigma(u_L-u_K)\quad\text{if }\sigma\in\mathcal{E}_{\mathrm{int}},\ \sigma=\sigma_{KL}$$

where

$$\tau_\sigma=|\sigma|\frac{\lambda_{K,\sigma}\lambda_{L,\sigma}}{\lambda_{K,\sigma}d_{L,\sigma}+\lambda_{L,\sigma}d_{K,\sigma}}\quad\text{if }\overline{\boldsymbol{x}}_\sigma\neq\boldsymbol{x}_K\text{ and }\overline{\boldsymbol{x}}_\sigma\neq\boldsymbol{x}_L$$

and

$$\tau_\sigma=|\sigma|\frac{\lambda_{K,\sigma}}{d_{K,\sigma}}\quad\text{if }\overline{\boldsymbol{x}}_\sigma\neq\boldsymbol{x}_K\text{ and }\overline{\boldsymbol{x}}_\sigma=\boldsymbol{x}_L$$

In practice using this harmonic mean $\lambda$ in the expression of the flux gives very good results.

This scheme is ideal for the $\Delta$-admissible meshes introduced in Section 5.1. For instance, in the case of rectangles, the properties obtained for the Laplace operator remain valid, that is, the scheme is robust and converges to the correct solution of the problem. However, geological grids often consist of very distorted (and sometimes degenerate) parallelograms for which the orthogonality condition fails to hold. In this case, the numerical diffusion flux becomes inconsistent. If the number of interfaces for which the orthogonality condition fails is large, the approximate solution may be very far from the solution of the problem. In this case, a natural idea to get a consistent numerical flux at an interface is to compute it by using more points than just the two points of the neighboring cells. This is the principle used in the methods developed in the works of Eymard *et al.*, 2000 (see Section 3.1.1) and Faille, 1992a

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

where the values at the points needed to define the normal gradient were interpolated from the cell unknowns. Alternatively, in (Coudière *et al.*, 1999) a gradient is reconstructed on diamond cells from the cell unknowns. In both cases, the scheme can only be proved to be stable under some geometrical assumptions on the mesh. Consequently, several other methods have been designed in recent years. Three of them are described below and a brief overview of some others is also given.

*5.2.2. Multiple Point Flux Approximation schemes.* The class of multiple point flux approximation (MPFA) methods was initiated in the works of Edwards and Rogers, 1994; Edwards and Rogers, 1998; Aavatsmark *et al.*, 1998a; Aavatsmark *et al.*, 1998b. The idea is to compute consistent fluxes by introducing an approximate gradient on subcells around each vertex using some additional edge unknowns, which are then eliminated by exploiting flux conservation. Numerous variants of MPFA schemes have been derived. The MPFA "O-scheme" for a rectangular mesh is particularly simple to describe. For $K \in \mathcal{T}$ and $\mathsf{v}$ a vertex

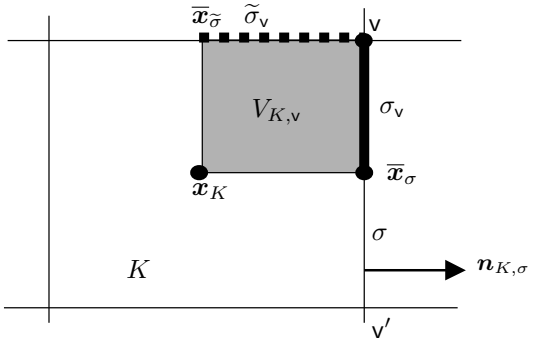


Figure 9. Notations for MPFA-O schemes defined on a rectangular mesh.

of $K$, let $V_{K,\mathsf{v}}$ be the parallelepiped polyhedron whose faces are parallel to the faces of $K$ and that has $\boldsymbol{x}_K$, $\overline{\boldsymbol{x}}_\sigma$ and $\mathsf{v}$ as vertices, where $\overline{\boldsymbol{x}}_\sigma$ denotes the centroid of $\sigma \in \mathcal{E}$. The discrete flux balance over a cell $K$ is established, but instead of computing the numerical flux over the whole interface $\sigma$, two numerical fluxes $F_{K,\sigma,\mathsf{v}}(u)$ and $F_{K,\sigma,\mathsf{w}}(u)$ over the half edges $\sigma_\mathsf{v}$ and $\sigma_\mathsf{w}$ (see Figure 9) are introduced. The discrete balance equation thus reads

$$\sum_{\sigma\in\mathcal{E}_K}\sum_{\mathsf{v}\in\mathcal{V}_\sigma} F_{K,\sigma,\mathsf{v}}(u) = \int_K f(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

The numerical fluxes over these half edges are computed by assuming an approximate solution $u$, which is piecewise affine on the subcells $V_{K,\mathsf{v}}$ so that its gradient $\nabla_{V_{K,\mathsf{v}}}u$ is piecewise constant. Then, observing that $\nabla_{V_{K,\mathsf{v}}}u\cdot(\boldsymbol{x}_K-\overline{\boldsymbol{x}}_\sigma) = u_K - u_\sigma$ and $\nabla_{V_{K,\mathsf{v}}}u\cdot(\boldsymbol{x}_K-\overline{\boldsymbol{x}}_{\widetilde{\sigma}}) = u_K - u_{\widetilde{\sigma}}$, the constant gradient of $u$ is computed as

$$\nabla_{V_{K,\mathsf{v}}}u = \frac{1}{|V_{K,\mathsf{v}}|}\sum_{\sigma\in\mathcal{E}_{K,\mathsf{v}}} |\sigma_\mathsf{v}|(v_\sigma - v_K)\boldsymbol{n}_{K,\sigma}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

where $\mathcal{E}_{K,\mathsf{v}}$ is the set of interfaces so that the numerical flux through the half edge $\sigma_\mathsf{v}$ can be computed as

$$F_{K,\sigma,\mathsf{v}}(u) = |\sigma_\mathsf{v}|\Lambda\nabla_{V_{K,\mathsf{v}}}u \cdot \boldsymbol{n}_{K,\sigma} = \frac{1}{|V_{K,\mathsf{v}}|}|\sigma_\mathsf{v}|^2(v_\sigma - v_K)A\boldsymbol{n}_{K,\sigma} \cdot \boldsymbol{n}_{K,\sigma}$$

In order to recover the complete scheme written only in terms of the cell unknowns $(u_K)_{K\in\mathcal{T}}$, it remains to eliminate the face unknowns $(u_\sigma)_{\sigma\in\mathcal{E}_{\rm int}}$. This is accomplished by requiring the conservation of the numerical fluxes $F_{K,\sigma,\mathsf{v}}(u)$ through the half-edges $\sigma_\mathsf{v}$. The following linear system then needs to be solved for each vertex of the mesh:

$$F_{K,\sigma,\mathsf{v}}(u) = -F_{L,\sigma,\mathsf{v}}(u),\ \forall\sigma \in \mathcal{E}_\mathsf{v},\ \text{for any } \sigma = K|L$$

Note that this local system is always invertible in the case of rectangular meshes considered here, but not so for general distorted meshes. The MPFA O-scheme has been introduced and used in the context of oil reservoir simulation. However, for highly nonconforming meshes, the local systems may be difficult to invert and the numerical fluxes may become unrealistic.

The MPFA O-scheme may be constructed for general meshes; see Aavatsmark *et al.*, 1998b. Other variants have been proposed such as the so-called L scheme in Aavatsmark *et al.*, 2008, the U scheme in Aavatsmark *et al.*, 1998b, and the scheme given in Agelas *et al.*, 2010. See Droniou, 2014 for a review of these methods and their properties.

**Remark** 1.[Gradient schemes] The numerical fluxes of the O-scheme were computed by constructing an approximate gradient on subcells on the grid. The idea of using a discrete gradient can be pursued further by using it in an approximate weak formulation of the scheme. These so-called gradient schemes have been designed to deal with anisotropic elliptic problems on polytope meshes in Eymard *et al.*, 2012. The idea is to discretize the weak form of equation (145a), thanks to a gradient discretization, which consists of a set of discrete unknowns, a reconstruction operator, and a discrete gradient, which are both constructed from the set of discrete unknowns such that the $L^2$ norm of the discrete gradient is a norm on the set of discrete unknowns. In order for the gradient scheme to be convergent for the linear problem, the gradient discretization should satisfy the following properties:

- *Coercivity* ensures uniform discrete Poincaré inequalities for the family of gradient discretizations. This is essential in obtaining *a priori* estimates for the solutions to gradient schemes.
- *Consistency* ensures that the family of gradient discretizations "covers" the whole energy space of the model (e.g., $H_0^1(\Omega)$ for the linear equation (133)).
- *Limit-conformity* ensures that the family of gradient and function reconstructions asymptotically satisfies the Stokes formula.

See Eymard *et al.*, 2012; Droniou *et al.*, 2010; Droniou *et al.*, 2016; Droniou *et al.*, 2014 for more precise definitions. Even though gradient schemes are not, in general, finite volume schemes (in fact, they include conforming and nonconforming finite element schemes, mimetic methods, and other schemes), some particular finite volume schemes are gradient schemes. Furthermore,

several recently developed schemes are in fact gradient schemes. This is the case for the O-scheme on rectangular and simplicial meshes. This is also the case for some forms of thediscrete duality finite volume (DDFV) scheme discussed next.

*5.2.3. DDFV schemes.* The design principle of DDFV schemes Hermeline (2000), Hermeline, 2003; Domelevo and Omnes, 2005; Andreianov *et al.*, 2007; Boyer and Hubert, 2008; Hermeline, 2007; Andreianov *et al.*, 2008; Coudière and Hubert, 2011 is again to introduce some discrete gradients to compute the numerical diffusion flux on a face (or edge) of the mesh. However, now the additional unknowns that are used to compute the fluxes are located at the vertices rather than at the faces (or edges of the mesh). The approximate gradients are chosen piecewise constant on the so-called "diamond cells" (dotted area in Figure 10) and they are computed using the directions given by the vertices of the diamond cell in Figure 10, which are linearly independent. The discrete gradient on the diamond cell $D_\sigma$ associated with the interface

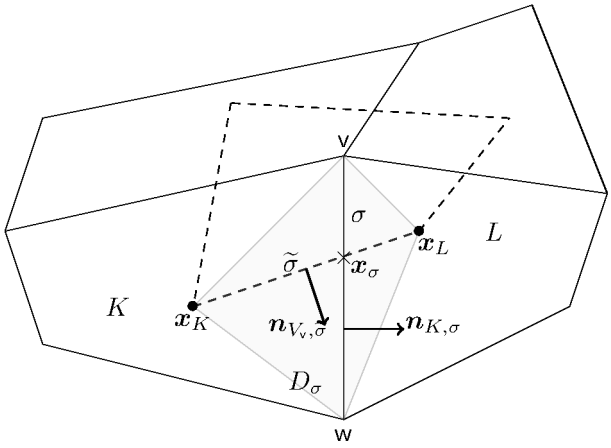


Figure 10. 2D DDFV: Control volumes $K$ and $L$ and diamond cell $D_\sigma$ for the gradient reconstruction.

$\sigma = \sigma_{KL}$, whose vertices are $\mathsf{v}$ and $\mathsf{w}$, is then computed such that

$$\nabla_\sigma u \cdot (\boldsymbol{x}_K - \boldsymbol{x}_L) = u_K - u_L \text{ and } \nabla_\sigma u \cdot (\mathsf{v} - \mathsf{w}) = u_\mathsf{v} - u_\mathsf{w}$$

The approximate numerical flux through an interface $\sigma$ belonging to the boundary of $K$ can then be defined as

$$F_{K,\sigma}(u) = -|\sigma|\Lambda_\sigma \nabla_\sigma u \cdot \boldsymbol{n}_{K,\sigma}$$

where $\Lambda_\sigma$ is the mean value of $\Lambda$ over the diamond cell $D_\sigma$. The discrete unknowns are the values $(u_K)_{K\in\mathcal{T}}$ and $(u_\mathsf{v})_{\mathsf{v}\in\mathcal{V}}$ where $\mathcal{V}$ is the set of vertices of the mesh.

In order to have as many equations as unknowns, the discrete balance equations are written on the primal cells $K$ as in the TP scheme and also on dual cells $P_\mathsf{v}$ associated with the vertices, such as shown in Figure 10. The flux through the edge $\widetilde{\sigma} : \boldsymbol{x}_K\boldsymbol{x}_L$ is easily computed

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

using the discrete gradient

$$F_{P_{\mathsf{v}},\widetilde{\sigma}}(u) = -|\widetilde{\sigma}|\Lambda_\sigma \nabla_\sigma u \cdot \boldsymbol{n}_{P_{\mathsf{v}},\widetilde{\sigma}}$$

The linear system to be solved is given by

$$\sum_{\sigma\in\mathcal{T}} F_{K,\sigma}(u) = \int_K f \mathrm{d}\boldsymbol{x} \quad \forall K \in \mathcal{T} \text{ and } \sum_{\mathsf{v}\in\mathcal{V}} F_{P_{\mathsf{v}},\widetilde{\sigma}}(u) = \int_{P_{\mathsf{v}}} f \mathrm{d}\boldsymbol{x} \quad \forall \mathsf{v} \in \mathcal{V} \tag{137}$$

and is of order $N_\mathcal{T} + N_\mathcal{V}$, where $N_\mathcal{T}$ is the number of cells of the mesh and $N_\mathcal{V}$ the number of vertices. The name DDFV was chosen as a reminder to the fact that the system may also be written using a discrete divergence operator that can be deduced from the above discrete gradient using a discrete Stokes formula. This property is used in the convergence analysis that was performed in two space dimensions for a number of problems in Domelevo and Omnes, 2005; Boyer and Hubert, 2008; Andreianov *et al.*, 2007.

The extension of the DDFV method to the three-dimensional setting was developed more recently. Two main approaches exist: (i) the CeVe-DDFV method, which uses cell and vertex unknowns as described in Hermeline, 2009; Coudière *et al.*, 2009; Andreianov *et al.*, 2010, and (ii) the CeVeFE-DDFV method, which uses cell, vertex, faces and edges unknowns as described in Coudière and Hubert, 2011; Coudière *et al.*, 2011. The coercivity properties of the two methods differ: the CeVe-DDFV method does not seem to be unconditionally coercive on generic meshes, whereas the CeVeFE-DDFV method is unconditionally coercive; see Droniou, 2014. In fact, this latter method is a gradient scheme as described in Remark 1. The original construction of this scheme necessitates four meshes. However, it can also be described using only one mesh; see for example Droniou *et al.*, 2016 for such a description, which relies on the definition of a discrete gradient on the octahedral cells $V_{K,\mathsf{v}}$ depicted in Figure 11. As shown in

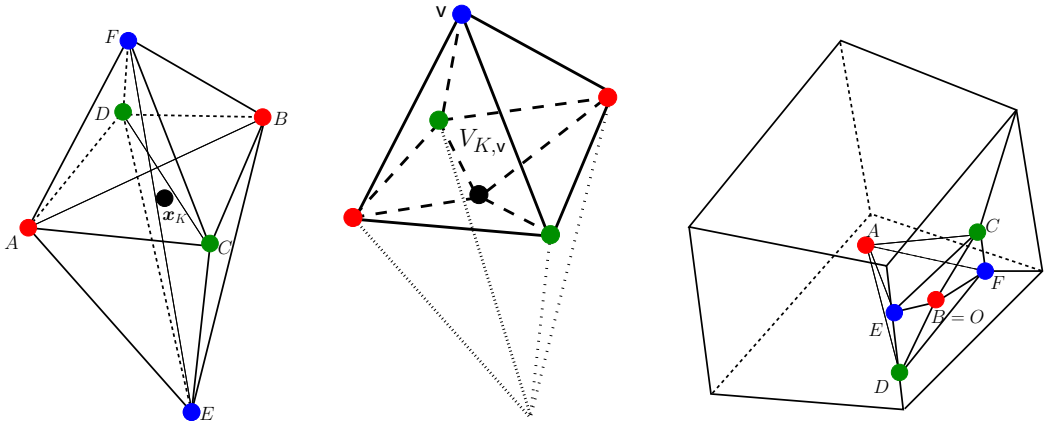


Figure 11. (a) Octahedral cell $K$ for the CeVeFE-DDFV scheme. (b) Illustration of $V_{K,\mathsf{v}}$. (c) Construction of a degenerate octahedron from a non-conforming hexahedral mesh in a heterogeneous medium ($CDEF$ is the intersection of the boundaries of two nonmatching hexahedral cells).

Andreianov *et al.*, 2007 (Section IX.B) for 2D DDFV methods, the other three meshes may be reconstructed from the "diamond" mesh, but they are not really needed to define the scheme

(nor to implement it). Note that the octahedral cells may be degenerate. This is the case, for instance, when working with nonconforming hexahedral meshes used when modeling flows in heterogeneous porous media.

*5.2.4. HHM schemes.* In the past decades, several schemes have been developed for elliptic equations so as to satisfy some form of calculus formula at the discrete level. These schemes are called mimetic finite difference (MFD) or compatible discrete operator (CDO) schemes. See Lipnikov *et al.*, 2014 for a review of MFD methods and Bonelle and Ern, 2014; Bonelle *et al.*, 2015 and references therein for CDO methods. Contrary to DDFV methods, which construct discrete operators and duality products to satisfy fully discrete calculus formulas, MFD/CDO methods are based on discrete operators satisfying a Stokes formula, which involves both continuous and discrete functions. Depending on the choice of the location of the main geometrical entities attached to the degrees of freedom (faces or vertices), two different MFD/CDO families exist.

A first MFD method, hereafter called hybrid mimetic finite difference (HMFD), is obtained by using a mixed form of (133) with fluxes through the mesh faces as initial unknowns, that is, by introducing $\boldsymbol{q} = \Lambda\nabla u$ so that $-\nabla\cdot\boldsymbol{q} = f$, and then discretizing this set of two equations. The resulting scheme was proved in Droniou *et al.*, 2010 to be embedded in a slightly larger family, which also contains the hybrid finite volume (HFV) method in Eymard *et al.*, 2010a and the mixed finite volume (MFV) method in Droniou and Eymard, 2006. The schemes of this family are called hybrid mimetic mixed (HMM) schemes. Each scheme in this family can be written in three different ways depending on the approach considered: HMFD, HFV, or MFV. The HFV formulation of a HMM scheme is very close to the weak formulation (135) of the elliptic PDE. It consists of a weak formulation with a discrete gradient and a stabilization term (bilinear form on $(u, v)$), see Eymard *et al.*, 2010a. It was proved in Droniou *et al.*, 2013 that the discrete gradient can be modified to include the stabilization terms. Thus, all HMM methods (and therefore also all HMFD methods) are part of the gradient scheme family described in Remark 1.

*5.2.5. Other schemes and topics of interest.* Another type of scheme for diffusion problems is the FVE method, which is based on finite element spaces with vertex unknowns and flux balances on dual meshes around vertices; see Cai and McCormick, 1990; Cai, 1991; Cai *et al.*, 1991; Süli, 1991; Lazarov, Michev and Vassilevsky, 1996; Chatzipantelidis and Lazarov, 2005; Chatzipantelidis *et al.*, 2013 and references therein. The FVE method can be recast in Petrov–Galerkin form using a piecewise constant test space together with a conforming trial space. To formulate and analyze the Petrov–Galerkin representation, two tessellations of $\Omega$ are considered: a triangulation $\mathcal{T}$ with simplicial elements $K \in \mathcal{T}$ and a dual tessellation $\mathcal{T}^*$ with control volumes $T \in \mathcal{T}^*$. In the class of conforming trial space methods such as the FVE method, a globally continuous, piecewise $p$-order polynomial trial space with zero trace value on the physical domain boundary is constructed

$$X_h = \{v \in C^0(\Omega) \Big| v|_K \in \mathcal{P}_p(K),\ \forall K \in \mathcal{T} \quad \text{and} \quad v|_{\partial\Omega} = 0\}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

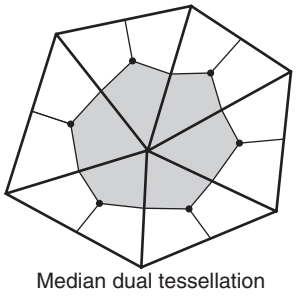


Figure 12. Control volume for the finite volume element method formed from median dual segments in each triangle.

using nodal Lagrange elements on the simplicial mesh. A dual tessellation $\mathcal{T}^*$ of the Lagrange element is then constructed. See Figure 12 which shows a linear Lagrange element with dual tessellation. These dual tessellated regions form control volumes for the FVM. The tessellation technique extends to higher-order Lagrange elements in a straightforward manner. A piecewise constant test space is then constructed using $\mathcal{T}^*$

$$Y_h = \{v \,\big|\, v|_T \in \chi(T),\ \forall T \in \mathcal{T}^*\}$$

where $\chi(T)$ is a characteristic function in the control volume $T$. The FVE discretization of (133a) then yields the following Petrov–Galerkin formulation: Find $u_h \in X_h$ such that

$$\sum_{\forall T \in \mathcal{T}^*} \left( \int_{\partial T} w_h \Lambda \nabla u_h \cdot \boldsymbol{n} \, \mathrm{d}s + \int_T w_h f \, \mathrm{d}x \right) = 0, \quad \forall w_h \in Y_h \tag{138}$$

The analysis of (138) by Ewing, Lin and Lin (2002) using linear elements gives an *a priori* estimate in an $L^2$ norm that requires the least amount of solution regularity when compared to previous methods of analysis.

**Theorem 25.** *(FVE a priori error estimate, Ewing, Lin and Lin (2002)) Assume a 2-D quasi-uniform triangulation $\mathcal{T}$ with dual tessellation $\mathcal{T}^*$ such that $\exists C > 0$ satisfying*

$$C^{-1}h^2 \le |T| \le Ch^2,\ \forall T \in \mathcal{T}^*$$

*Assume that $u$ and $u_h$ are solutions of (133a) and (138) respectively with $u \in H^2(\Omega)$, $f \in H^\beta$, $(0 \le \beta \le 1)$. Then $\exists C' > 0$ such that the a priori estimate holds*

$$\|u - u_h\|_{L^2(\Omega)} \le C' \left( h^2 \|u\|_{H^2(\Omega)} + h^{1+\beta} \|f\|_{H^\beta(\Omega)} \right) \tag{139}$$

Unlike the finite element method, the error estimate (139) reveals that optimal order convergence is obtained only if $f \in H^\beta$ with $\beta \ge 1$. Moreover, numerical results show that the source term regularity cannot be reduced without deteriorating the measured convergence rate. Optimal convergence rates are also shown for the nonconforming Crouzeix–Raviart element based FVM analyzed by Chatzipantelidis (1999) for $u \in H^2(\Omega)$ and $f \in H^1(\Omega)$.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Other FVMs or related schemes for elliptic boundary value problems have been proposed and analyzed under a variety of names: box methods in Bank and Rose (1987), Bank and Rose (1987),Croisille and Greff, 2005, Greff, 2007, covolume methods in Chou and Li (2000), diamond cell methods in Coudière *et al.*, 1999; Kútik and Mikula, 2015, and finite volume based on the Crouzeix–Raviart element in Chatzipantelidis (1999); Ewing, Lin and Lin (2002).

*A posteriori* error estimates have been obtained for a number of the above mentioned schemes; see for example, Lazarov and Tomov, 2002; Omnes, 2008; Di Pietro *et al.*, 2011; Ern and Vohralík, 2011; Chen and Wang, 2013; Cancès *et al.*, 2014; Chen and Gunzburger, 2014; Arbogast *et al.*, 2014; Erath, 2015. See the review article by Di Pietro and Vohralík, 2014 and references therein for more on this subject.

A number of recent methods for anisotropic problems can be found in Herbin and Hubert, 2008; Eymard *et al.*, 2011 which give a numerical comparison of the methods for several 2D and 3D application benchmarks using various types of meshes. See also the recent review articles in Droniou, 2014; Droniou, 2014; Di Pietro and Vohralík, 2014.

*5.2.6. The transmissivity structure.* In the theoretical study of linear or nonlinear PDEs, it is sometimes useful to use a nonlinear function of the unknowns as a test function in the weak formulation. This may be the case, for instance, to establish some of the properties of the solutions or in order to show the existence of solutions to the systems. Here are four examples where this approach is used to show

1. positivity of the solution for a linear elliptic problem with nonnegative right-hand side;
2. $L^\infty$ estimates for the solution of an elliptic problem for a right-hand side in $L^p$, $p > \frac{d}{2}$;
3. existence of solutions for elliptic problems in a nonvariational framework, for instance, if the right-hand side is only integrable; see Boccardo and Gallouët, 1989;
4. existence of solutions for noncoercive convection diffusion problems (and positivity if the right-hand-side is nonnegative); see Droniou, 2002.

See Gallouët, 2007 for more details and references on these four examples. In the theoretical study of discretization schemes for the same equations, the same technique (i.e., using a nonlinear function of the unknown in a weak formulation) is also often used. However, in the current state of the art, this technique is successful only if the schemes are of the form

$$\sum_{L\in\mathbb{N}(K)} \tau_{K,L}(u_K - u_L) = \text{rhs}$$

with $\tau_{K,L} \geq 0$. This is the case for the TPFA scheme on admissible meshes (see e.g., Gallouët and Herbin, 1999; Droniou and Gallouët, 2002; Droniou, 2003; Droniou *et al.*, 2003; Chainais-Hillairet and Droniou, 2011) for noncoercive problems and/or problems with nonregular right-hand-side. For general schemes, this is an open problem.

Some results (e.g., positivity, maximum principle) have been obtained for finite volume schemes using more general meshes but the price to pay is that the scheme is nonlinear, even

in the case of a linear problem; see for example Le Potier, 2005; Le Potier, 2008; Le Potier, 2009; Genty and Le Potier, 2011; Droniou and Le Potier, 2011; Cancès *et al.*, 2013 for the original schemes and Droniou, 2014 for a review.

### *5.3. The parabolic case*

*5.3.1. The continuous problem.* Next, consider a time-dependent convection diffusion equation. Let $T > 0$, $u_0 \in L^2(\Omega)$ and $\boldsymbol{v} \in \mathbb{R}^d$ be given. The time-dependent convection diffusion problem, $u : \Omega \times [0,T] \to \mathbb{R}$, is given by

$$\begin{cases} \partial_t u + \mathrm{div}(\boldsymbol{v}u) - \Delta u = 0 \text{ in } \Omega \times (0,T) \\ u = 0 \text{ in } \partial\Omega \times (0,T) \\ u(.,0) = u_0 \text{ in } \Omega \end{cases} \tag{140}$$

For the sake of simplicity, consider the case $\boldsymbol{v} = 0$. A weak formulation of this problem is

$$\begin{cases} \text{Find } u \in L^2(0,T;\, H_0^1(\Omega)) \text{ such that } \partial_t u \in L^2(0,T;\, H^{-1}(\Omega)) \text{ and} \\ \quad < \partial_t u, \varphi >_{H^{-1},H_0^1} + \displaystyle\int_\Omega \nabla u(\boldsymbol{x},\cdot) \cdot \nabla \varphi(\boldsymbol{x},\cdot) \mathrm{d}\boldsymbol{x} = 0 \quad \forall\, \varphi \in H_0^1(\Omega) \text{ a.e. in } (0,T) \\ \quad u(\cdot,0) = u_0 \end{cases} \tag{141}$$

Here, the duality product $< \cdot, \cdot >_{H^{-1},H_0^1}$ is defined by

$$< \partial_t u, \varphi >_{H^{-1},H_0^1} = -\int_0^T u(\cdot,t)\partial_t(\cdot,t)\varphi \mathrm{d}t \in H_0^1(\Omega)$$

Hence, identifying the space $L^2(\Omega)$ with its dual space $(L^2(\Omega))'$, the statement $\partial_t u \in L^2(0,T;\, H^{-1}(\Omega))$ is to be understood as requiring that there exists $v \in L^2(0,T;\, H^{-1}(\Omega))$ such that

$$\underbrace{-\int_0^T u(\cdot,t)\partial_t(\cdot,t)\varphi \mathrm{d}t}_{\in H_0^1(\Omega)} = \underbrace{\int_0^T v(\cdot,t)\varphi(\cdot,t)}_{\in H^{-1}(\Omega)}$$

which makes sense since $H_0^1(\Omega)$ can then be identified to its dual space $H^{-1}(\Omega)$. Note that if $u$ satisfies (141), then $u \in C([0,T], L^2(\Omega))$.

As in the steady state case, some estimates on $u$ are obtained in order to get compactness properties despite the lack of regularity of the approximate finite volume solution. In the continuous framework, the natural estimates are in $L^2(0,T;\, H_0^1(\Omega))$ for $u$ and in $L^2(0,T;\, H^{-1}(\Omega))$ for $\partial_t u$. These estimates give compactness in $L^2(0,T;\, L^2(\Omega))$ for a sequence of approximate solutions (using, for instance, the Faedo–Galerkin approximation), which proves the existence of a solution to (141). It is reasonable to therefore look for the same kind of estimates in the discrete framework, which will also yield the compactness in $L^2(0,T;\, L^2(\Omega))$ of a sequence of approximate finite volume solutions.

*5.3.2. The finite volume scheme.* Let $\mathcal{T}$ be an admissible mesh of $\Omega$, in the sense introduced in Section 5.1, and let $\delta t = T/M$ be the (uniform) time step. The finite volume scheme with

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

implicit Euler discretization in time is given by

$$\begin{cases} |K|\dfrac{u_K^{n+1}-u_K^n}{\delta t}+\displaystyle\sum_{\sigma\in\mathcal{E}_K}F_{K,\sigma}(u_{\mathcal{T}}^{n+1})=0 \quad 0\le n\le M-1 \\ u_K^0=\dfrac{1}{|K|}\displaystyle\int_K u_0(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \end{cases} \tag{142}$$

with $F_{K,\sigma}(u_{\mathcal{T}}^{n+1})=-\dfrac{|\sigma|}{d_{KL}}(u_L^{n+1}-u_K^{n+1})+v_{K,\sigma}^+u_K^{n+1}-v_{K,\sigma}^-u_L^{n+1}$.

The existence and uniqueness of a solution $(u_K^n)_{n\in\mathbb{N}}$ to (142) is easily deduced from the steady-state case. Denote by $H_{\mathcal{D}}(\Omega\times(0,T))$ the set of functions of $L^2(\Omega\times(0,T))$, which are piecewise constant on the subsets $K\times[t_n,t_{n+1})$. Define an approximate solution $u_{\mathcal{D}}\in H_{\mathcal{D}}(\Omega\times(0,T))$ by $u_{\mathcal{D}}(x,t)=u_K^n,\ \forall x\in K,\ \forall t\in[t_n,t_{n+1})$. Using a variational technique similar to the way the estimate (127) is established in the steady-state case, the following *a priori* estimates on $u_{\mathcal{D}}$ may be obtained:

$$\|u_{\mathcal{D}}\|_{L^\infty(0,T;\,L^2(\Omega))}\le C \tag{143}$$

and

$$\delta t\sum_{n=1}^{M}\|u_{\mathcal{D}}(\cdot,t_n)\|_{1,\mathcal{T}}^2\le C \tag{144}$$

where $\|\cdot\|_{1,\mathcal{T}}$ is the discrete $H_0^1$ norm defined by (123) and $C$ depends only on the initial condition. Using equation (142), a discrete $H^{-1}$ estimate on the discrete time derivative is derived. For $n=0,\dots,M-1$, let $\delta_t^n u=\frac{u_{n+1}-u_n}{\delta t}\in H_{\mathcal{T}}(\Omega)$; then the following estimate is obtained:

$$\delta t\sum_{n=1}^{M-1}\|\delta_t^n u\|_{-1,\mathcal{T}}\le C$$

where the discrete dual norm $\|\cdot\|_{-1,\mathcal{T}}$ is defined by (126), and $C$ depends only on the initial condition. The compactness of a sequence of approximations may then be deduced from a generalization of the Aubin–Simon lemma for sequences of embedded subspaces that can be stated in the general $L^p$ setting and used in the context of less regular solutions or nonlinear problems.

**Lemma 9 (Generalized Aubin–Simon lemma)** *Let $1\le p<+\infty$. Let $B$ be a Banach space, $(X_\ell)_{\ell\in\mathbb{N}}$ be a sequence of Banach spaces included in $B$, and $(Y_\ell)_{\ell\in\mathbb{N}}$ be a sequence of Banach spaces. Assume that the sequence $(X_\ell,Y_\ell)_{\ell\in\mathbb{N}}$ satisfies the following conditions:*

1. *The sequence $(X_\ell)_{\ell\in\mathbb{N}}$ is such that any sequence $(u_\ell)_{\ell\in\mathbb{N}}$, satisfying $u_\ell\in X_\ell$ (for all $\ell\in\mathbb{N}$) and $(\|u_\ell\|_{X_\ell})_{\ell\in\mathbb{N}}$ bounded, is relatively compact in $B$.*
2. *$X_\ell\subset Y_\ell$ (for all $\ell\in\mathbb{N}$) and if the sequence $(u_\ell)_{\ell\in\mathbb{N}}$ is such that $u_\ell\in X_\ell$ (for all $\ell\in\mathbb{N}$), $(\|u_\ell\|_{X_\ell})_{\ell\in\mathbb{N}}$ bounded and $\|u_\ell\|_{Y_\ell}\to 0$ (as $\ell\to+\infty$), then any subsequence converging in $B$ converges (in $B$) to $0$.*

*Let $T>0$ and $(u_\ell)_{\ell\in\mathbb{N}}$ be a sequence of $L^p((0,T),B)$ satisfying the following conditions:*

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

1. *For all* $\ell \in \mathbb{N}$, *there exists* $N \in \mathbb{N}^\star$ *and* $k_1, \ldots, k_N$ *in* $\mathbb{R}_+^\star$ *such that* $\sum_{i=1}^N k_i = T$ *and* $u_\ell(t) = v_i$ *for* $t \in (t_{i-1}, t_i)$, $i \in \{1, \ldots, N\}$, $t_0 = 0$, $t_i = t_{i-1} + k_i$, $v_i \in X_\ell$. *(Of course, the values* $N$, $k_i$ *and* $v_i$ *depend on* $\ell$*.)*
2. *The sequence* $(u_\ell)_{\ell \in \mathbb{N}}$ *is bounded in* $L^p((0,T), B)$.
3. *The sequence* $(\|u_\ell\|_{L^1((0,T),X_\ell)})_{\ell \in \mathbb{N}}$ *is bounded.*
4. *The sequence* $(\|\delta_t u_\ell\|_{L^p((0,T),Y_\ell)})_{\ell \in \mathbb{N}}$ *is bounded, where the function* $\delta_t u_\ell$ *is defined a.e. by*

$$\delta_t u_\ell(t) = \frac{v_i - v_{i-1}}{k_i} \text{ for } t \in (t_{i-1}, t_i)$$

*Then there exists* $u \in L^p((0,T), B)$ *such that, up to a subsequence,* $u_\ell \to u$ *in* $L^p((0,T), B)$.

See, for example, Chénier *et al.*, 2015 for its proof. The convergence in $L^2(0,T; L^2(\Omega))$ of $u_\mathcal{D}$ to some function $\bar{u} \in L^2(0,T; H_0^1(\Omega))$ is then obtained. As in the elliptic case, a passage to the limit in the scheme then yields that $\bar{u} = u$, weak solution of (141). This analysis may be generalized to the case of nonhomogeneous Dirichlet boundary conditions; see Bouillard *et al.*, 2007.

Error estimates for finite volume schemes applied to linear parabolic equations may also be obtained; see for example, Bradji, 2008; Bradji and Fuhrmann, 2010; Chatzipantelidis *et al.*, 2013. Note that in the case of parabolic equations with $L^1$ data, only the TPFA scheme has been analyzed in Gallouët *et al.*, 2012. As in the elliptic case, the analysis is based upon some test functions that are nonlinear functions of the unknowns, as in the continuous case. As already mentioned in Section 5.2.6, the extension of these results to numerical schemes on general grids is still an open problem. Nonlinear parabolic systems have also been analyzed, with several areas of applications: electrochemistry in Bradji and Herbin, 2008, porous media in Bouillard *et al.*, 2007; Chainais-Hillairet and Droniou, 2007, image processing in Mikula and Ramarosy, 2001; Frolkovič and Mikula, 2007; Drblíková and Mikula, 2007; Drblíková *et al.*, 2009, to cite only a few.

*5.3.3. Nonlinear conservation laws including diffusion terms.* As a final scalar PDE model, consider the nonlinear first-order conservation law with an added second-order Laplacian diffusion term

$$\partial_t u + \nabla \cdot \boldsymbol{f}(u) - \varepsilon \triangle u = 0 \text{ in } \mathbb{R}^d \times \mathbb{R}^+ \tag{145a}$$

$$u(x,0) = u_0 \text{ in } \mathbb{R}^d \tag{145b}$$

Here, $u(x,t)\colon \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}$ denotes the dependent solution variable, $\boldsymbol{f} \in C^1(\mathbb{R}, \mathbb{R}^d)$ the hyperbolic flux, and $\varepsilon \geq 0$ a small diffusion coefficient. An application of the divergence and Gauss theorems to (145a) integrated in a region $K$ yields the following integral conservation law form

$$\frac{\partial}{\partial t} \int_K u \, \mathrm{d}x + \int_{\partial K} f(u) \cdot \boldsymbol{n}_K \, \mathrm{d}s - \int_{\partial K} \varepsilon \nabla u \cdot \boldsymbol{n}_K \, \mathrm{d}s = 0 \tag{146}$$

Using the finite volume approximation of the diffusion operator of the previous section, the fully discrete form (31) of Section 3 is extended to the integral conservation law (146) by the

introduction of a numerical diffusion flux function $F_{K,\sigma}(u_h)$ for a control volume $K \in \mathcal{T}$ such that

$$\int_{\partial K} \varepsilon \nabla u \cdot \boldsymbol{n}_K \, \mathrm{d}s \approx \sum_{\sigma \in \mathcal{E}_K} F_{K,\sigma}(u_h)$$

where the discrete diffusive flux is chosen similar to the linear parabolic case, as $F_{K,\sigma}(u_h^m) := \frac{|\sigma|}{d_{KL}}(u_L^m - u_K^m)$. When combined with the general finite volume formulation (31) for hyperbolic conservation laws, the following fully discrete scheme is produced

$$u_K^{n+1} = u_K^n - \frac{\Delta t^n}{|K|} \sum_{\sigma \in \mathcal{E}_K} \left( g(u_K^n, u_L^n; \sigma) \, |\sigma| - F_{K,\sigma}(u_h^m) \right), \ \forall K \in \mathcal{T} \tag{147}$$

In this equation, the index $m$ may be chosen either as $n$ or $n+1$, corresponding to an explicit or implicit discretization.

Stability analysis reveals a CFL-like stability condition for the explicit scheme [choice $m = n$ in (147) ]

$$\Delta t^n \leq \frac{\alpha^3 (h_{\min}^n)^2}{\alpha L_g h_{\min}^n + \varepsilon}$$

where $L_g$ denotes the Lipschitz constant of the hyperbolic numerical flux, $\alpha$ is a positive mesh-dependent parameter, and $\epsilon$ is the diffusion coefficient. In constructing this bound, a certain form of shape regularity is assumed such that there exists an $\alpha > 0$ such that for all $j, k$ with $h_K \equiv \mathrm{diam}(K)$

$$\alpha h_K^2 \leq |K| \qquad \alpha |\partial K| \leq h_K \qquad \alpha h_K \leq d_{KL} \tag{148}$$

Thus, $\Delta t^n$ is of the order $h^2$ for large $\varepsilon$ and of the order $h$ for $\varepsilon \leq h$. In cases where the diffusion coefficient is larger than the mesh size, it is advisable to use an implicit scheme ($m = n + 1$). In this latter situation, no time step restriction has to be imposed; see Eymard *et al.* (2002) and Ohlberger (2001b).

In order to demonstrate the main difficulties when analyzing convection-dominated problems, consider the following result from Feistauer *et al.* (1999) for a homogeneous diffusive boundary value problem. In this work, a MFV, finite element method sharing similarities with the methods described above is used to obtain a numerical approximation $u_h$ of the exact solution $u$. Using typical energy-based techniques, they prove the following error bound.

**Theorem 26.** *For initial data $u_0 \in L^\infty(\mathbb{R}^2) \cap W^{1,2}(\mathbb{R}^2)$ and $\tau > 0$ there exist constants $c_1, c_2 > 0$ independent of $\varepsilon$ such that*

$$\|u(\cdot, \tau) - u_h(\cdot, \tau)\|_{L^2(\Omega)} \leq c_1 h e^{c_2 \tau / \epsilon} \tag{149}$$

This estimate is fundamentally different from estimates for the purely hyperbolic problems of Sections 3 and 4. Specifically, this result shows how the estimate strongly depends on the small parameter $\varepsilon$, ultimately becoming unbounded as $\varepsilon$ tends to zero.

In the context of convection dominated or degenerate parabolic equations, Kruzkov techniques have been used by Carrillo (1999) and Karlsen and Risebro (2000) in proving the

uniqueness and stability of solutions. Utilizing these techniques, convergence of finite volume schemes (uniform with respect to $\varepsilon \to 0$) was proved in Eymard *et al.* (2002) and error estimates were obtained for viscous approximations in Evje and Karlsen (2002) and Eymard, Gallouët and Herbin (2002). In Ohlberger (2001a, 2001b) uniform *a posteriori error* estimates suitable for adaptive meshing are given, see also Ohlberger and Rohde, 2002 for the case of weakly coupled systems. Using the theory of nonlinear semigroups, continuous dependence results were also obtained in Cockburn and Gripenberg (1999) (see Cockburn (2003) for a review). FVMs for nonlinear systems of parabolic equations and degenerate parabolic equations have received a lot of attention in the recent years in a wide area of applications such as porous media, corrosion models, population models, and so on as discussed in Chainais-Hillairet and Peng, 2004; Chainais-Hillairet and Droniou, 2007; Andreianov *et al.*, 2010; Andreianov *et al.*, 2011; Chainais-Hillairet *et al.*, 2011; Andreianov *et al.*, 2011; Bessemoulin and Filbet, 2012; Andreianov *et al.*, 2010; Chainais-Hillairet *et al.*, 2013; Bessemoulin *et al.*, 2014; Chainais-Hillairet *et al.*, 2015 and references therein.

## 6. Advanced Topics

### *6.1. Extension to systems of nonlinear hyperbolic conservation laws*

The early widespread use of FVMs comes, in part, from the relative ease in algorithmically extending the numerical discretization of Section 3 to systems of nonlinear hyperbolic conservation laws of the form

$$\partial_t \boldsymbol{u} + \nabla \cdot \boldsymbol{f}(\boldsymbol{u}) = 0 \text{ in } \mathbb{R}^d \times \mathbb{R}^+ \tag{150a}$$

$$\boldsymbol{u}(x,0) = \boldsymbol{u}_0(x) \text{ in } \mathbb{R}^d \tag{150b}$$

In these equations, $\boldsymbol{u}(x,t)\colon \mathbb{R}^d \times \mathbb{R}^+ \to \mathbb{R}^m$ denotes the vector of dependent solution variables, $\boldsymbol{f}\colon \mathbb{R}^m \to \mathbb{R}^{m\times d}$ denotes the flux vector, and $\boldsymbol{u}_0\colon \mathbb{R}^d \to \mathbb{R}^m$ denotes the initial data vector function. It is also assumed that the conservation law system (150) possesses an additional scalar entropy equation in divergence form with convex entropy–entropy flux pair $\{U, F\}$, $U\colon \mathbb{R}^m \mapsto \mathbb{R}$ convex and $F\colon \mathbb{R}^m \to \mathbb{R}^{1\times d}$

$$(U(\boldsymbol{u}))_t + \nabla \cdot (F(\boldsymbol{u})) \leq 0 \tag{151}$$

with the right-hand side equal to zero for classical (smooth) solutions. As discussed further in Section 6.1.3, the existence of this entropy extension equation with convex entropy function $U$ is sufficient to deduce that (150a) can be recast in symmetric hyperbolic form under a change of variables $\boldsymbol{u} \mapsto \boldsymbol{v}$ where $\boldsymbol{v} \in \mathbb{R}^m$ are the so-called entropy variables for the conservation law system. This also implies hyperbolicity in the sense that eigenvalues of the flux jacobian matrix $A(\boldsymbol{u};\boldsymbol{n}) \equiv \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}} \cdot \boldsymbol{n}$ are real (but not necessarily distinct). This eigenstructure is used in the construction of upwind numerical flux functions in Sections 6.1.1 and 6.1.2.

*6.1.1. Numerical flux functions for linear hyperbolic systems.* The main task in extending FVMs to systems of nonlinear conservation laws is the construction of a suitable numerical flux

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

function. To gain insight into this task, consider the one-dimensional *linear* Cauchy problem

$$\begin{aligned} \partial_t \boldsymbol{u} + \partial_x (A\,\boldsymbol{u}) &= 0 \quad \text{in } \mathbb{R}\times\mathbb{R}^+ \\ \boldsymbol{u}(x,0) &= \boldsymbol{u}_0(x) \text{ in } \mathbb{R} \end{aligned} \tag{152}$$

where $A \in \mathbb{R}^{m\times m}$ is a *constant* matrix. Assume the matrix $A$ has $m$ has real eigenvalues, $\lambda_1 \le \lambda_1 \le \cdots \le \lambda_m$, and a complete set of right and left eigenvectors denoted by $\boldsymbol{r}_k \in \mathbb{R}^m$ and $\boldsymbol{l}_k \in \mathbb{R}^m$, respectively for $k = 1,\ldots,m$. Furthermore, let $X \in \mathbb{R}^{m\times m}$ denote the matrix of right eigenvectors, $X = [\boldsymbol{r}_1,\ldots,\boldsymbol{r}_m]$, and $\Lambda \in \mathbb{R}^{m\times m}$ the diagonal matrix of eigenvalues, $\Lambda = \text{diag}(\lambda_1,\ldots,\lambda_m)$ so that $A = X\Lambda X^{-1}$. The one-dimensional system (152) is readily decoupled into scalar equations via the transformation to the characteristic variables $\boldsymbol{\alpha} \equiv X^{-1}\boldsymbol{u}$, solved as scalar equations, and recomposed into system form yielding

$$\boldsymbol{u}(x,t) = \sum_{k=1}^{m} \boldsymbol{l}_k \cdot \boldsymbol{u}_0(x-\lambda_k t)\,\boldsymbol{r}_k$$

Using this solution form, it is straightforward to solve the associated Riemann problem for $\boldsymbol{w}(\xi,\tau;\boldsymbol{u},\boldsymbol{v}) \in R^m$

$$\partial_\tau \boldsymbol{w} + \partial_\xi (A\boldsymbol{w}) = 0 \quad \text{in } \mathbb{R}\times\mathbb{R}^+$$

with initial data

$$\boldsymbol{w}(\xi,0;\boldsymbol{u},\boldsymbol{v}) = \begin{cases} \boldsymbol{u} & \text{if } \xi < 0 \\ \boldsymbol{v} & \text{if } \xi > 0 \end{cases}$$

thereby producing the following Riemann-like numerical flux function

$$\begin{aligned} \boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) &= A\boldsymbol{w}(0,\tau>0;\boldsymbol{u},\boldsymbol{v}) \\ &= \frac{1}{2}(A\boldsymbol{u}+A\boldsymbol{v}) - \frac{1}{2}|A|\,(\boldsymbol{v}-\boldsymbol{u}) \end{aligned} \tag{153}$$

with $|A| \equiv X|\Lambda|X^{-1}$. When this numerical flux function is used in the one-dimensional fully discrete finite volume discretization with numerical solution $u_j^n$ at $x_j = j\Delta x$ and $t^n = n\Delta t$

$$\begin{aligned} \boldsymbol{u}_j^{n+1} &= \boldsymbol{u}_j^n - \frac{\Delta t}{\Delta x}(\boldsymbol{g}(\boldsymbol{u}_{j+1}^n,\boldsymbol{u}_j^n) - (\boldsymbol{g}(\boldsymbol{u}_j^n,\boldsymbol{u}_{j-1}^n)) \\ &= \boldsymbol{u}_j^n - \frac{\Delta t}{\Delta x}\left(A^+(\boldsymbol{u}_j^n - \boldsymbol{u}_{j-1}^n) + A^-(\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_j^n)\right) \qquad A^\pm \equiv X\Lambda^\pm X^{-1} \end{aligned} \tag{154}$$

the resulting discretization reproduces the Courant–Isaacson–Rees (CIR) upwind discretization for linear systems of hyperbolic equations given in Courant *et al.*, 1952.

*6.1.2. Numerical flux functions for nonlinear systems of conservation laws.* In Godunov, 1959, exact solutions of the one-dimensional nonlinear Riemann problem of gas dynamics were used in the construction of a numerical flux function

$$\boldsymbol{g}^{\mathrm{R}}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{f}(\boldsymbol{w}(0,t>0;\boldsymbol{u},\boldsymbol{v})) \tag{155}$$

where $\boldsymbol{w}(\xi,\tau;\boldsymbol{u},\boldsymbol{v}) \in \mathbb{R}^m$ is a solution of a nonlinear Riemann problem

$$\partial_\tau \boldsymbol{w} + \partial_\xi \boldsymbol{f}(\boldsymbol{w}) = 0 \quad \text{in } \mathbb{R}\times\mathbb{R}^+ \tag{156}$$

with initial data

$$\boldsymbol{w}(\xi, 0; \boldsymbol{u}, \boldsymbol{v}) = \begin{cases} \boldsymbol{u} & \text{if } \xi < 0 \\ \boldsymbol{v} & \text{if } \xi > 0 \end{cases}$$

Recall that solutions of the Riemann problem for gas dynamic systems with ideal gas law are a composition of shock, contact, and rarefaction wave family solutions. For the gas dynamic equations considered by Godunov, a unique solution of the Riemann problem exists for general states $\boldsymbol{u}$ and $\boldsymbol{v}$ except those states producing a vacuum. Even so, the solution of the Riemann problem is both mathematically and computationally nontrivial. Consequently, a number of alternative numerical fluxes have been proposed that are more computationally efficient. These alternative numerical fluxes can be sometimes interpreted as approximate Riemann solvers. A partial list of alternative numerical fluxes is given here. A more detailed treatment of this subject is given in Godlewski and Raviart (1991), Kröner (1997), and LeVeque (2002). In describing these fluxes, it is assumed that the flux jacobian matrix $A(\boldsymbol{u}; \boldsymbol{n}) \equiv \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}} \cdot \boldsymbol{n}$ is diagonalizable via the matrix of right eigenvectors denoted by $X \in \mathbb{R}^{m \times m}$ and the diagonal matrix of real eigenvalues $\Lambda \equiv \text{diag}[\lambda_1, \ldots, \lambda_m]$, so that $A \equiv X \Lambda X^{-1}$.

- *Osher–Solomon flux (Osher and Solomon, 1982).* This numerical flux is a system generalization of the Enquist–Osher flux (46) of Section 3. All wave families are approximated in state space as rarefaction or inverted rarefaction waves with Lipschitz continuous partial derivatives. The Osher-Solomon numerical flux is of the form

$$\boldsymbol{g}^{\text{OS}}(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{n}) = \frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}) + \boldsymbol{f}(\boldsymbol{v})) \cdot \boldsymbol{n} - \frac{1}{2} \int_{\boldsymbol{u}}^{\boldsymbol{v}} |A(\boldsymbol{w}; \boldsymbol{n})| \, \mathrm{d}\boldsymbol{w} \tag{157}$$

  where $|A|$ denotes the matrix absolute value via eigenvalues and eigenvectors. Integrating on $m$ rarefaction wave integral subpaths that are each parallel to a right eigenvector, a system decoupling occurs on each subpath integration. Furthermore, for the gas dynamic equations with ideal gas law, it is straightforward to construct $m-1$ Riemann invariants on each subpath thereby eliminating the need for path integration altogether. This reduces the numerical flux calculation to purely algebraic computations with special care taken at sonic points; see Osher and Solomon (1982).
- *Roe flux (Roe, 1981).* Roe's numerical flux can be interpreted as approximating all wave families as discontinuities. The numerical flux is of the form

$$\boldsymbol{g}^{\text{Roe}}(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{n}) \quad = \quad \frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}) + \boldsymbol{f}(\boldsymbol{v})) \cdot \boldsymbol{n} - \frac{1}{2} |A(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{n})| (\boldsymbol{v} - \boldsymbol{u})$$

  where $A(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{n})$ is the "Roe matrix" satisfying the matrix mean value identity

$$(\boldsymbol{f}(\boldsymbol{v}) - \boldsymbol{f}(\boldsymbol{u})) \cdot \boldsymbol{n} = A(\boldsymbol{u}, \boldsymbol{v}; \boldsymbol{n})(\boldsymbol{v} - \boldsymbol{u}) \tag{158}$$

  with $A(\boldsymbol{u}, \boldsymbol{u}; \boldsymbol{n}) = A(\boldsymbol{u}; \boldsymbol{n})$. For the equations of gas dynamics with ideal gas law, the Roe matrix takes a particularly simple form. Steady discrete mesh-aligned shock profiles are resolved with at most one intermediate point. The Roe flux does not preclude the formation of entropy violating expansion shocks unless additional steps are taken near sonic points.
- *VFRoe flux (Gallouët and Masella, 1996).* It is not always straightforward to find matrices that satisfy the Roe condition (158). Another approach, using a linearized

problem but not requiring the Roe condition, is the so-called "VFRoe" scheme. The idea is to use the Godunov flux (155) but with $\boldsymbol{w}$ as the solution of a linearized Riemann problem rather than the exact Riemann problem (156). The VFRoe flux is given by

$$\boldsymbol{g}^{\mathrm{VFRoe}}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{f}(\boldsymbol{w}(0,t>0;\boldsymbol{u},\boldsymbol{v}))$$

where $\boldsymbol{w}(\xi,\tau;\boldsymbol{u},\boldsymbol{v}) \in \mathbb{R}^m$ is a solution of a linearized Riemann problem

$$\partial_\tau \boldsymbol{w} + A(\boldsymbol{u},\boldsymbol{v})\partial_\xi \boldsymbol{w} = 0 \quad \text{in} \quad \mathbb{R}\times\mathbb{R}^+ \tag{159}$$

with initial data

$$\boldsymbol{w}(\xi,0;\boldsymbol{u},\boldsymbol{v}) = \begin{cases} \boldsymbol{u} & \text{if } \xi < 0 \\ \boldsymbol{v} & \text{if } \xi > 0 \end{cases}$$

where $A(\boldsymbol{u},\boldsymbol{v})$ is an approximate jacobian matrix. Several choices of such matrices exist as discussed in Masella *et al.*, 1999; Brun *et al.*, 2000; Buffard *et al.*, 2000; Gallouët *et al.*, 2002; Gallouët *et al.*, 2003. As in the case of the Roe scheme, solutions violating an entropy inequality may occur. Entropy correction techniques have been designed to avoid this problem (Helluy *et al.*, 2010).

- *Steger–Warming flux vector splitting (Steger and Warming, 1981).* Steger and Warming considered a splitting of the flux vector for the gas dynamic equations with ideal gas law that exploited the fact that the flux vector is homogeneous of degree one in the conserved variables. Euler's identity for homogeneous functions of degree one then yields $\boldsymbol{f}(\boldsymbol{u})\cdot\boldsymbol{n} = A(\boldsymbol{u};\boldsymbol{n})\boldsymbol{u}$. Steger and Warming then considered the matrix splitting

$$A = A^+ + A^-, \quad A^\pm \equiv X\Lambda^\pm X^{-1}$$

where $\Lambda^\pm$ is computed componentwise. From this matrix splitting, the final upwind numerical flux function was constructed as

$$\boldsymbol{g}^{\mathrm{SW}}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{n}) = A^+(\boldsymbol{u};\boldsymbol{n})\,\boldsymbol{u} + A^-(\boldsymbol{v};\boldsymbol{n})\boldsymbol{v}$$

Although not part of their explicit construction, for the gas dynamic equations with ideal gas law, the jacobian matrix $\partial\boldsymbol{g}^{\mathrm{SW}}/\partial\boldsymbol{u}$ has eigenvalues that are all nonnegative and the jacobian matrix $\partial\boldsymbol{g}^{\mathrm{SW}}/\partial\boldsymbol{v}$ has eigenvalues that are all nonpositive whenever the ratio of specific heats $\gamma$ lies in the interval $[1, 5/3]$. The matrix splitting leads to numerical fluxes that do not vary smoothly near sonic and stagnation points. Use of the Steger–Warming flux splitting in the schemes of Sections 3 and 4 results in rather poor resolution of linearly degenerate contact waves and velocity slip surfaces due to the introduction of excessive artificial diffusion for these wave families.

- *van Leer flux vector splitting.* van Leer (1982) provided an alternative flux splitting for the gas dynamic equations that produces a numerical flux of the form

$$\boldsymbol{g}^{\mathrm{VL}}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{n}) = \boldsymbol{f}^-(\boldsymbol{u};\boldsymbol{n}) + \boldsymbol{f}^+(\boldsymbol{v};\boldsymbol{n})$$

using special Mach number polynomials to construct fluxes that remain smooth near sonic and stagnation points. As part of the splitting construction, the jacobian matrix $\partial\boldsymbol{g}^{\mathrm{SW}}/\partial\boldsymbol{u}$ has eigenvalues that are all nonnegative and the matrix $\partial\boldsymbol{g}^{\mathrm{SW}}/\partial\boldsymbol{v}$ has eigenvalues that are all nonpositive. The resulting expressions for the flux splitting are somewhat simpler when compared to the Steger–Warming splitting. The van Leer splitting also introduces excessive diffusion in the resolution of linearly degenerate contact waves and velocity slip surfaces.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

- *System local Lax–Friedrichs flux.* This numerical flux is the system equation counterpart of the scalar local Lax–Friedrichs flux (39). For systems of conservation laws, the Lax–Friedrichs flux is given by

$$\boldsymbol{g}^{\mathrm{LF}}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{n}) = \frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}) + \boldsymbol{f}(\boldsymbol{v}))\cdot\boldsymbol{n} - \frac{1}{2}\alpha_{\max}(\boldsymbol{v}-\boldsymbol{u})$$

  where $\alpha_{max}$ is calculated from

$$\alpha_{\max} = \max_{1\le k\le m}\sup_{\boldsymbol{w}\in[\boldsymbol{u},\boldsymbol{v}]}|\lambda_k(\boldsymbol{w};\boldsymbol{n})|$$

  The system Lax-Friedrichs flux is usually not applied on the boundary of domains since it generally requires an overspecification of boundary data. The system Lax–Friedrichs flux introduces a relatively large amount of artificial diffusion when used in the schemes of Section 3. Consequently, this numerical flux is typically only used together with relatively high-order reconstruction schemes where the detrimental effects of excessive artificial diffusion are mitigated.
- *Harten–Lax–van Leer flux (Harten, Lax and van Leer, 1983).* The Harten–Lax–van Leer numerical flux originates from a simplified two wave model of more general $m$ wave systems such that waves associated with the smallest and largest characteristic speeds of the $m$ wave system are always accurately represented in the two-wave model. The following numerical flux results from this simplified two-wave model

$$\begin{aligned}\boldsymbol{g}^{\mathrm{HLL}}(\boldsymbol{u},\boldsymbol{v};\boldsymbol{n}) &= \frac{1}{2}(\boldsymbol{f}(\boldsymbol{u})+\boldsymbol{f}(\boldsymbol{v}))\cdot\boldsymbol{n}\\ &- \frac{1}{2}\frac{\alpha_{\max}+\alpha_{\min}}{\alpha_{\max}-\alpha_{\min}}(\boldsymbol{f}(\boldsymbol{v})-\boldsymbol{f}(\boldsymbol{u}))\cdot\boldsymbol{n} + \frac{\alpha_{\max}\alpha_{\min}}{\alpha_{\max}-\alpha_{\min}}(\boldsymbol{v}-\boldsymbol{u})\end{aligned}$$

  where

$$\alpha_{\max} = \max_{1\le k\le m}(0,\sup_{\boldsymbol{w}\in[\boldsymbol{u},\boldsymbol{v}]}\lambda_k(\boldsymbol{w};\boldsymbol{n})),\quad \alpha_{\min} = \min_{1\le k\le m}(0,\inf_{\boldsymbol{w}\in[\boldsymbol{u},\boldsymbol{v}]}\lambda_k(\boldsymbol{w};\boldsymbol{n}))$$

  Using this flux, full upwinding is obtained for supersonic flow. Modifications of this flux are suggested in Einfeldt *et al.* (1998) to improve the resolution of intermediate waves as well.
- *Entropy stable fluxes.* An important class of numerical fluxes are motivated from energy analysis of the FVM for symmetrizable systems of conservation laws. A brief introduction to this theory and resulting flux functions are described in Section 6.1.3. Use of these fluxes in the FVM guarantees a form of global entropy stability as characterized in Theorem 27.
- *Relaxation schemes (Jin and Xin, 1995).* Another way to avoid solving the Riemann problem (either because it is time consuming or because it not solvable) is to transform the hyperbolic system into a simpler one, for instance, into a linear system with relaxation term such as in Jin and Xin, 1995. The relaxation scheme can be reinterpreted as defining a particular approximate Riemann solver for the original system conservation laws, see LeVeque and Pelanti, 2001. Relaxation has also been used to replace a real equation of state (EOS) by the perfect gas EOS in the Euler system in Coquel and Perthame, 1998. Relaxation schemes are used in a wide area of applications such as pressureless gases in Berthon *et al.*, 2006, two-phase flows in Pelanti, 2011, discontinuous fluxes in Karlsen

*et al.*, 2003, multicomponent flows in Dellacherie, 2003, Gallouët *et al.*, 2010, Dorogan *et al.*, 2012, magnetohydrodynamics in Bouchut *et al.*, 2009, or kidney physiology in Perthame, 2015, to cite only a few.

- *Well-balanced schemes (Greenberg and LeRoux, 1996).* The notion of well-balanced schemes was introduced by Greenberg and LeRoux, 1996 and Gosse, 1996 to deal with source terms in hyperbolic systems in such a way that the steady state solutions are correctly approximated. Well-balanced schemes have been used in several applications; see Bouchut and Zeitlin, 2008 and Bouchut, 2004 for a thorough presentation of the schemes and Berthon, 2016; Amadori and Gosse, 2016; Desprès, 2016 for some recent work on the subject.

Further examples of numerical fluxes include the kinetic flux vector splitting due to Deshpande (1986), the convective upwind and split pressure (CUSP) flux of Jameson (1993) and Tatsumi, Martinelli and Jameson (1994), the advection upstream spitting method(AUSM) flux of Liou and Steffen (1993), and the related internal energy-based solvers on staggered grids of Herbin *et al.*, 2013; Herbin *et al.*, 2016b. Note that these latter staggered schemes have also been implemented in implicit and semi-implicit time discretizations using staggered meshes in Gastaldo *et al.*, 2011 or collocated meshes in Herbin *et al.*, 2016a in coping with low-Mach number flows.

*6.1.3. Entropy stability of the finite volume method for systems of hyperbolic conservation laws.* In the symmetrization theory for first-order conservation laws (see Godunov, 1961; Mock, 1980; Tadmor, 1987; Harten, 1983a) one seeks a locally invertible mapping $\boldsymbol{u}(\boldsymbol{v}) : \mathbb{R}^m \mapsto \mathbb{R}^m$ applied to (150a) so that when transformed

$$\underbrace{\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{v}}}_{\text{SPD}} \frac{\partial \boldsymbol{v}}{\partial t} + \sum_{i=1}^{d} \underbrace{\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{v}}}_{\text{symmetric}} \frac{\partial \boldsymbol{v}}{\partial x_i} = 0$$

the matrix $\frac{\partial \boldsymbol{u}}{\partial \boldsymbol{v}}$ is SPD and the matrices $\frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{v}}$ are symmetric. Mock, 1980 has proved that the existence of an entropy equation (151) with convex entropy function $U$ is sufficient to guarantee that the system can be symmetrized via a change of variable with the new variables $\boldsymbol{v}$ (commonly referred to as entropy variables) calculated from $\boldsymbol{v}^T = \frac{\partial U}{\partial \boldsymbol{u}}$. The entropy variables also serve as nonlinear weights such that the weighted combination of conservation law equations equals the entropy extension equation, that is,

$$\boldsymbol{v} \cdot (\boldsymbol{u}_t + \nabla \cdot \boldsymbol{f}) = U_t + \nabla \cdot F$$

with the right-hand side equal to zero for smooth solutions but not necessarily so for nonsmooth solutions. Specifically, the inequality in the entropy equation (151) repeated here

$$U_t + \nabla \cdot F \leq 0$$

implies a form of stability and decay in a closed entropy system:

- The total mathematical entropy is nonincreasing (macroscopic Boltzmann H-theorem)

$$\frac{d}{\mathrm{d}t} \int_{\Omega} U(\boldsymbol{u}(x,t)) \, \mathrm{d}x \leq 0 \tag{160}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

- A two-sided bound on the total mathematical entropy for $t_0 \leq t$

$$\int_\Omega U(\boldsymbol{u}^*)\,\mathrm{d}x \leq \int_\Omega U(\boldsymbol{u}(x,t))\,\mathrm{d}x \leq \int_\Omega U(\boldsymbol{u}(x,t_0))\,\mathrm{d}x \tag{161}$$

  where $\boldsymbol{u}^*$ is the minimum mathematical entropy state (maximum physical entropy state)

$$\boldsymbol{u}^* = \frac{1}{\mathrm{meas}(\Omega)} \int_\Omega \boldsymbol{u}(x,t_0)\ \mathrm{d}x$$

  that is invariant in time for a closed entropy system.
- System stability for $t_0 \leq t$

$$\|\boldsymbol{u}(x,t) - \boldsymbol{u}^*\|_{L^2(\Omega)} \leq (c_0^{-1} C_0)^{1/2}\, \|\boldsymbol{u}(x,t_0)) - \boldsymbol{u}^*\|_{L_2(\Omega)} \tag{162}$$

  where $C_0$ and $c_0$ are $L^2$-norm induced positive upper and lower bounds of the transformation jacobian $\frac{\partial \boldsymbol{v}}{\partial \boldsymbol{u}}$.

The last system stability statement is readily obtained from convexity of the entropy function after first rewriting the two-sided total entropy bound using a Taylor series with integral remainder formula

$$\int_\Omega \int_0^1 (1-\theta)(\boldsymbol{u}(x,t) - \boldsymbol{u}^*) \cdot \frac{\partial^2 U}{\partial \boldsymbol{u}^2}(\overline{\boldsymbol{u}}(\theta))\,(\boldsymbol{u}(x,t) - \boldsymbol{u}^*)\,\mathrm{d}\theta\,\mathrm{d}x$$
$$\leq \int_\Omega \int_0^1 (1-\theta)(\boldsymbol{u}(x,t_0) - \boldsymbol{u}^*) \cdot \frac{\partial^2 U}{\partial \boldsymbol{u}^2}(\overline{\boldsymbol{u}}_0(\theta))\,(\boldsymbol{u}(x,t_0) - \boldsymbol{u}^*)\,\mathrm{d}\theta\,\mathrm{d}x$$

where $\overline{\boldsymbol{u}}(\theta) \equiv \boldsymbol{u}^* + \theta\,(\boldsymbol{u}(x,t) - \boldsymbol{u}^*)$ and $\overline{\boldsymbol{u}}_0(\theta) \equiv \boldsymbol{u}^* + \theta\,(\boldsymbol{u}(x,t_0) - \boldsymbol{u}^*)$. This shows how one can go from a statement of boundedness for the scalar convex entropy function $U$ to a statement of boundedness for the vector of conservation variables $\boldsymbol{u}$.

These stability results have motivated the construction of numerical discretizations that discretely inherit or mimic these forms of entropy stability. The following theorem considers semidiscrete and fully discrete finite volume discretizations and provides sufficient conditions to be imposed on the numerical flux function so that discrete counterparts of the total entropy bounds (160) and (161) are obtained.

**Theorem 27.** *(Discrete entropy stability) Consider the hyperbolic conservation law system (150) with convex entropy extension equation (151) that is symmetrizable via the locally invertible change of variables $\boldsymbol{u} \mapsto \boldsymbol{v}$ and cell-averages $\boldsymbol{u}_K \equiv \boldsymbol{u}(\boldsymbol{v}_K)$ for a problem domain representing a closed entropy system.*

1. *The semidiscrete finite volume discretization*

$$\frac{d}{dt}\boldsymbol{u}_K\,|K| + \sum_{\sigma_{K,L} \subset \partial K} g(\boldsymbol{u}_K, \boldsymbol{u}_L; \boldsymbol{n}_{K,L})\,|\sigma_{K,L}| = 0, \quad \forall K \in \mathcal{T} \tag{163}$$

   *exhibits nonincreasing discrete total entropy*

$$\frac{d}{dt} \sum_{K \subset \mathcal{T}} U(\boldsymbol{u}_K)|K| \leq 0 \tag{164}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

*2. The fully discrete finite volume discretization*

$$(\boldsymbol{u}_K^{n+1} - \boldsymbol{u}_K^n)\,|K| + \sum_{\sigma_{K,L}\subset\partial K} g(\boldsymbol{u}_K^{n+1}, \boldsymbol{u}_L^{n+1}; \boldsymbol{n}_{K,L})\,|\sigma_{K,L}| = 0, \quad \forall K \in \mathcal{T},\ n = 0, 1, \ldots \tag{165}$$

*satisfies the two-sided discrete total entropy bound*

$$\sum_{K\subset\mathcal{T}} U(\boldsymbol{u}^*)|K| \leq \sum_{K\subset\mathcal{T}} U(\boldsymbol{u}_K^n)|K| \leq \sum_{K\subset\mathcal{T}} U(\boldsymbol{u}_K^0)|K| \tag{166}$$

*with*

$$\boldsymbol{u}^* = \frac{1}{\text{meas}(\Omega)} \sum_{K\subset\mathcal{T}} \boldsymbol{u}_K^0 |K| \tag{167}$$

*if the numerical flux function satisfies any of the following related sufficient conditions:*

- *The numerical flux function is of the following path integration form; see Barth, 1998*

$$\boldsymbol{g}(\boldsymbol{u}_-, \boldsymbol{u}_+; \boldsymbol{n}) = \frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}_-) + \boldsymbol{f}(\boldsymbol{u}_+)) \cdot \boldsymbol{n} - \frac{1}{2}\int_0^1 |A(\boldsymbol{u}(\overline{\boldsymbol{v}}(\theta)))|_{\boldsymbol{u_v}} [\boldsymbol{v}]_-^+ \, d\theta \tag{168}$$

  *where $|A|_{\boldsymbol{u_v}}$ is the matrix absolute value with respect to the matrix $\boldsymbol{u_v}$ (see remark) and $\overline{\boldsymbol{v}}(\theta) \equiv \boldsymbol{v}_- + \theta\,[\boldsymbol{v}]_-^+$.*
- *The numerical flux function is of the following form for $Q \in \mathbb{R}^{m\times m}$; see Tadmor, 2004; Fjordholm et al., 2012*

$$\boldsymbol{g}(\boldsymbol{u}_-, \boldsymbol{u}_+; \boldsymbol{n}) = \boldsymbol{g}^*(\boldsymbol{u}_-, \boldsymbol{u}_+; \boldsymbol{n}) - Q(\boldsymbol{u}_-, \boldsymbol{u}_+; \boldsymbol{n})[\boldsymbol{v}]_-^+ \tag{169}$$

  *where $\boldsymbol{g}^*(\boldsymbol{u}_-, \boldsymbol{u}_+; \boldsymbol{n})$ satisfies*

$$[\boldsymbol{v}]_-^+ \cdot \boldsymbol{g}^*(\boldsymbol{u}_-, \boldsymbol{u}_+; \boldsymbol{n}) = [F^* \cdot \boldsymbol{n}]_-^+ \tag{170}$$

  *with $F^*$ a primitive function that satisfies $\boldsymbol{f}^T = \frac{\partial F^*}{\partial \boldsymbol{v}}$ and $Q$ is any dissipation matrix such that (see remarks)*

$$[\boldsymbol{v}]_-^+ \cdot Q[\boldsymbol{v}]_-^+ \geq 0 \tag{171}$$

- *The numerical flux function satisfies the system generalization of Osher's E-flux condition (42) for scalar conservation laws, see Barth, 2006*

$$[\boldsymbol{v}]_-^+ \cdot (\boldsymbol{g}(\boldsymbol{u}_-, \boldsymbol{u}_+; \boldsymbol{n}) - \boldsymbol{f}(\boldsymbol{u}(\overline{\boldsymbol{v}}(\theta))) \cdot \boldsymbol{n}) \leq 0, \quad \forall \theta \in [0, 1] \tag{172}$$

  *with $\overline{\boldsymbol{v}}(\theta) \equiv \boldsymbol{v}_- + \theta\,[\boldsymbol{v}]_-^+$.*

*In these equations, $[\boldsymbol{v}]_-^+ \equiv \boldsymbol{v}_+ - \boldsymbol{v}_-$ and the shorthand notation $\boldsymbol{u}_\pm \equiv \boldsymbol{u}(\boldsymbol{v}_\pm)$ have been used.*

**Remarks** 1.*(Theorem 27)*

- Observe that the formulations (163) and (165) contain nonlinearity, not only from the use of numerical flux functions $\boldsymbol{g}(\cdot,\cdot;\cdot)$ but also from the use of entropy variables in $\boldsymbol{u}_K \equiv \boldsymbol{u}(\boldsymbol{v}_K)$.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

- The sufficient conditions (168) and (172) are those originally obtained for the discontinuous Galerkin finite element method that have been reduced to the present FVM by choosing finite-dimensional trial and test spaces equal to piecewise constant polynomials.
- The matrix absolute value $|A|_{\boldsymbol{u_v}}$ with respect to the matrix $\boldsymbol{u_v}$ appearing in the path integrated flux (168) can be efficiently computed as

$$|A|_{\boldsymbol{u_v}} = X|\Lambda|X^T \tag{173}$$

  where $X$ is the matrix of scaled right eigenvectors that diagonalizes $A$ such that

$$A = X\Lambda X^{-1}, \qquad \boldsymbol{u_v} = XX^T$$

  using the constructive proof given in Barth, 1998.
- These entropy stability results have been extended to finite volume methods with higher order reconstruction together with specific forms of the (169) dissipation matrix $Q$ discussed in Fjordholm *et al.*, 2012 including the obvious choice $|A|_{\boldsymbol{u_v}}$

$$Q = X|\Lambda|X^T \tag{174}$$

- It is well known that the assumption of the existence of $F^*$ satisfying $\boldsymbol{f}_i^T = \frac{\partial F_i^*}{\partial \boldsymbol{v}}$ used in (170) is not always generically valid. For example, in ideal compressible magnetohydrodynamics, $\boldsymbol{f}^T = \frac{\partial F_i^*}{\partial \boldsymbol{v}} - \frac{\partial \phi}{\partial \boldsymbol{v}}\boldsymbol{B}_i$ where $\boldsymbol{B}$ is the magnetic induction field and $\frac{\partial \phi}{\partial \boldsymbol{v}}$ are involution multipliers, so that sufficient conditions resulting from entropy stability analysis are significantly different; see Barth, 2007.

### *6.2. The Marker-and-Cell scheme for fluid flows*

The Marker-and-Cell (MAC) scheme, introduced in the mid-60s by Harlow and Welch, 1965, is one of the most popular engineering methods for approximating the Navier-Stokes equations. This is primarily due to the simplicity, efficiency and remarkable mathematical properties of the method; see for example Patankar, 1980; Wesseling, 2001. The first error analysis seems to be that of Porsching, 1978 in the case of the time-dependent Stokes equations on uniform square grids. The mathematical analysis of the scheme was performed for the steady state Stokes equations and Navier–Stokes equations in Nicolaïdes, 1992; Nicolaïdes and Wu, 1996 for uniform rectangular meshes with $H^2$ regularity assumption on the pressure. In the 90s, using the tools that were developed for the finite volume theory that can be found in Eymard *et al.*, 2000, an order 1 error estimate for nonuniform meshes was obtained in Blanc, 1999, with order 2 convergence for uniform meshes, under the usual regularity assumptions ($H^2$ for the velocities, $H^1$ for the pressure). The convergence of the MAC scheme for the Stokes equations with a right-hand side in $H^{-1}(\Omega)$ was later proved in Blanc, 2005.

*6.2.1. The MAC scheme for the steady-state Navier–Stokes equations.* Consider the incompressible steady-state Navier-Stokes equations with Dirichlet boundary conditions on

a bounded domain $\Omega$ of $\mathbb{R}^d$, $d = 2$ or 3. Denoting by $\bar{\boldsymbol{u}}$ the velocity and by $\bar{p}$ the pressure, the governing equations are given by

$$\operatorname{div} \bar{\boldsymbol{u}} = 0 \qquad \text{in } \Omega \tag{175a}$$

$$-\Delta \bar{\boldsymbol{u}} + (\bar{\boldsymbol{u}} \cdot \nabla)\bar{\boldsymbol{u}} + \boldsymbol{\nabla} \bar{p} = \boldsymbol{f} \qquad \text{in } \Omega \tag{175b}$$

$$\bar{\boldsymbol{u}} = 0 \qquad \text{on } \partial\Omega \tag{175c}$$

In this section, the notation $(\bar{\boldsymbol{u}}, \bar{p})$ denotes a solution to the continuous problem, while the notation $(\boldsymbol{u}, p)$ denotes a solution to the approximate problem obtained from the MAC discretization. Writing the balance form of these equations for a control volume $K$ for the incompressibility condition (175a) and a control volume $D$ for the momentum equation (175a) (both included in $\Omega$) yields

$$\int_{\partial K} \bar{\boldsymbol{u}} \cdot \boldsymbol{n} = 0 \tag{176a}$$

$$-\int_{\partial D} \boldsymbol{\nabla} \bar{\boldsymbol{u}} \cdot \boldsymbol{n} + \int_{\partial D} \bar{\boldsymbol{u}}(\bar{\boldsymbol{u}} \cdot \boldsymbol{n}) + \int_D \boldsymbol{\nabla} \bar{p} = \boldsymbol{f} \tag{176b}$$

where $\boldsymbol{f} \in L^2(\Omega)^d$. The balance form (176a) of the incompressibility condition is also given by

$$\sum_{\sigma \subset \partial K} \int_\sigma \bar{\boldsymbol{u}} \cdot \boldsymbol{n}_{K,\sigma} = 0$$

where $\boldsymbol{n}_{K,\sigma}$ denotes the unit vector normal to $\sigma$ and outward $K$.

In the MAC scheme, a Cartesian mesh $\mathcal{T}$ of the domain is used with the set of edges (or faces in 3D) denoted by $\mathcal{E}$. The discrete velocity unknowns $u_\sigma$ are located on the edges (or faces) $\sigma$ of the mesh. These unknowns are approximations of $\bar{\boldsymbol{u}} \cdot \boldsymbol{e}_i$ on the faces (or edges) that are orthogonal to $\boldsymbol{e}_i$, the $i$th vector of the canonical basis of $\mathbb{R}^d$. This is illustrated in Figure 13 which depicts a 2-D MAC grid where the horizontal components of the velocities are located at the $\times$ symbols on the vertical edges and the vertical components of the velocities are located at the $\square$ symbols on the horizontal edges. Taking for $K$ a pressure cell (dotted cell on Figure 13) with edges (or faces) orthogonal to the vectors of the canonical basis, a natural discretization of the flux $\int_\sigma \boldsymbol{u} \cdot \boldsymbol{n}_{K,\sigma}$ is $|\sigma| u_\sigma \epsilon_{K,\sigma}$ where $\epsilon_{K,\sigma} = \boldsymbol{n}_{K,\sigma} \cdot \boldsymbol{e}_i = \pm 1$ depending on the direction of $\boldsymbol{n}_{K,\sigma}$ and $u_\sigma$ is the discrete unknown associated with the face $\sigma$. If $\sigma$ belongs to the set $\mathcal{E}^{(i)}$ of edges that are orthogonal to the $i$th unit vector of the canonical basis, then $u_\sigma$ is an approximation of the mean value of $\boldsymbol{u} \cdot \boldsymbol{e}_i$ over $\sigma$. The balance form (176a) of the incompressibility condition may then be discretized as

$$\sum_{\sigma \in \mathcal{E}_K} |\sigma| u_\sigma \epsilon_{K,\sigma} = 0$$

where $\mathcal{E}_K$ is the set of edges (or faces) of $K$. Since the discrete unknowns are approximations of the components of $\bar{\boldsymbol{u}}$ on the corresponding edges (or faces), it is natural to introduce the spaces $H_{\mathcal{E}}^{(i)}$, $i = 1, \ldots, d$ for the discrete velocity unknowns as the set of piecewise constant functions on the velocity control volumes that are centered on the edges (or faces) $\sigma \in \mathcal{E}^{(i)}$ of the mesh. The resulting velocity control volumes are the North-West ($i = 1$) and North-East

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

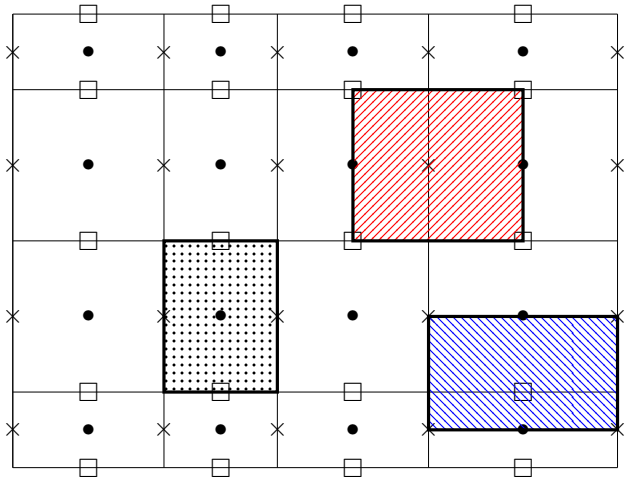

Figure 13. MAC unknowns: •, pressure; ×, horizontal velocity; □, vertical velocity. Control volumes: pressure cell, dotted area; horizontal and vertical velocities, striped areas.

$(i = 2)$ dashed cells in Figure 13. In order to take (partially) into account the homogeneous Dirichlet boundary conditions, the spaces

$$H_{\mathcal{E},0}^{(i)} = \left\{ u \in H_{\mathcal{E}}^{(i)},\ u(\boldsymbol{x}) = 0\ \forall \boldsymbol{x} \in D_\sigma,\ \sigma \in \widetilde{\mathcal{E}}_{\text{ext}}^{(i)},\ i = 1, \ldots, d \right\}$$

are defined. By introducing $\mathbf{H}_{\mathcal{E},0} = \prod_{i=1}^{d} H_{\mathcal{E},0}^{(i)}$, a vector function $\boldsymbol{u} \in \mathbf{H}_{\mathcal{E},0}$ can be defined as $\boldsymbol{u} = (u_1, \ldots, u_d)$ with $u_i = \sum_{\sigma \in \mathcal{E}^{(i)}} u_\sigma \mathbb{1}_{D_\sigma}$, where $\mathbb{1}_{D_\sigma}$ is the characteristic function of $D_\sigma$, that is $\mathbb{1}_{D_\sigma}(\boldsymbol{x}) = 1$ if $\boldsymbol{x} \in D_\sigma$ and zero otherwise. A discrete divergence $\text{div}_K \boldsymbol{u}$ of $\boldsymbol{u} \in \mathbf{H}_{\mathcal{E},0}$ can then be defined on the cell $K$

$$\text{div}_K \boldsymbol{u} = \frac{1}{|K|} \sum_{\sigma \in \mathcal{E}_K} |\sigma| u_\sigma \epsilon_{K,\sigma} \tag{177}$$

The discrete divergence of $\boldsymbol{u} = (u_1, \ldots, u_d) \in \mathbf{H}_{\mathcal{E},0}$ may also be written as

$$\text{div}_{\mathcal{T}}(\boldsymbol{u}) = \sum_{i=1}^{d} (\delta_i u_i)_K \mathbb{1}_K \tag{178}$$

where $\mathbb{1}_K$ is the characteristic function of the control volume $K$ and $(\delta_i u_i)_K$ is the discrete derivative of $u_i$ on $K$ defined by

$$(\delta_i u_i)_K = \frac{|\sigma|}{|K|}(u_{\sigma'} - u_\sigma) \text{ with } K = [\overrightarrow{\sigma\sigma'}] \text{ and } \sigma, \sigma' \in \mathcal{E}^{(i)} \tag{179}$$

where $K = [\overrightarrow{\sigma\sigma'}]$ means that $\sigma$ and $\sigma'$ are faces of $K$ that are parallel and oriented. The discrete derivatives and divergence are consistent in the sense that if $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_d)$ is a

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

smooth vector function over $\Omega$ and $\Pi_{\mathcal{E},i}$ are "reasonable" interpolators (for $i = 1, \ldots, d$), then $\delta_i(\Pi_{\mathcal{E}}^{(i)}\varphi_i)$ tends to $\partial_i\varphi$ and $\mathrm{div}_{\mathcal{T}}(\Pi_{\mathcal{E}}\boldsymbol{\varphi})$ tends to $\mathrm{div}\boldsymbol{\varphi}$ as the mesh size tends to 0. This is stated precisely in the following lemma.

**Lemma 10 (Discrete derivative and divergence consistency)** *Let $\mathcal{D} = (\mathcal{T}, \mathcal{E})$ be an MAC grid, and let $\Pi_{\mathcal{E}}$ be an interpolator from $C_c^\infty(\Omega)^d$ to $\mathbf{H}_{\mathcal{E},0}$ such that, for any $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_d)^t \in (C_c^\infty(\Omega))^d$, there exists $C_{\boldsymbol{\varphi}} \geq 0$ depending only on $\boldsymbol{\varphi}$ such that*

$$\begin{aligned} &\Pi_{\mathcal{E}}\boldsymbol{\varphi} = \left(\Pi_{\mathcal{E}}^{(1)}\varphi_1, \cdots, \Pi_{\mathcal{E}}^{(d)}\varphi_d\right) \in H_{\mathcal{E},0}^{(1)} \times \cdots \times H_{\mathcal{E},0}^{(d)}, \text{ where} \\ &|\Pi_{\mathcal{E}}^{(i)}\varphi_i(\boldsymbol{x}) - \varphi_i(\boldsymbol{x}_\sigma)| \leq C_{\boldsymbol{\varphi}} h_{\mathcal{T}}^2, \ \forall \boldsymbol{x} \in D_\sigma, \ \forall \sigma \in \mathcal{E}^{(i)}, \ \forall i = 1, \cdots, d \end{aligned} \tag{180}$$

*Let $\eta$ denote the regularity of the mesh*

$$\eta_{\mathcal{T}} = \max\left\{\frac{|\sigma|}{|\sigma'|}, \ \sigma \in \mathcal{E}^{(i)}, \forall \sigma' \in \mathcal{E}^{(j)}, i, j \in 1, \ldots, d, i \neq j\right\} \tag{181}$$

*Then there exists $C_{\boldsymbol{\varphi},\eta} \geq 0$ such that*

$$|\delta_i \Pi_{\mathcal{E}}^{(i)}\varphi_i(\boldsymbol{x}) - \partial_i\varphi_i(\boldsymbol{x})| \leq C_{\boldsymbol{\varphi},\eta} h_{\mathcal{T}}$$

*for a.e. $\boldsymbol{x} \in \Omega$. As a consequence, if $(\mathcal{D}_n)_{n\in\mathbb{N}} = (\mathcal{T}_n, \mathcal{E}_n)_{n\in\mathbb{N}}$ is a sequence of MAC grids such that $\eta_n \leq \eta$ for all $n$ and $h_{\mathcal{T}_n} \to 0$ as $n \to +\infty$, then $\mathrm{div}_{\mathcal{T}_n}(\Pi_{\mathcal{E}_n}\boldsymbol{\varphi}) \to \mathrm{div}\boldsymbol{\varphi}$ uniformly as $n \to +\infty$.*

Examples of interpolators satisfying (180) are, for instance, the mean value over an edge (or face) or the value of a function at the centroid of the edge (or face).

For the momentum equation, there is a need to distinguish the various components of the velocity and momentum as well as associate a control volume with each unknown. To illustrate this, Figure 13 shows a cell associated with the horizontal velocity (North-West dashed lines, $\times$ in the center) centered on a vertical edge belonging to $\mathcal{E}_1$ and a cell associated with the vertical velocities (North-East dashed lines, $\square$ in the center) centered on a vertical edge belonging to $\mathcal{E}_2$. The set $\mathcal{E}$ of faces that are orthogonal to the $i$th unit vector $\boldsymbol{e}_i$ can be decomposed as $\mathcal{E}^{(i)} = \mathcal{E}_{\mathrm{int}}^{(i)} \cup \mathcal{E}_{\mathrm{ext}}^{(i)}$ where $\mathcal{E}_{\mathrm{int}}^{(i)}$ (resp. $\mathcal{E}_{\mathrm{ext}}^{(i)}$) are the edges of $\mathcal{E}^{(i)}$ that lie in the interior (resp. on the boundary) of the domain $\Omega$. Next, take for the control volume $D$ in (176b) a velocity cell $D_\sigma, \sigma \in \mathcal{E}^{(i)}$, corresponding to the $i$th velocity component. Each velocity grid consisting of such cells is an admissible mesh for the Laplace operator. Therefore, the finite volume discretization given in Section (5) can be directly applied. The diffusion fluxes as in (120) for each velocity grid $i = 1, \ldots, d$ (replacing the cell $K$ by a cell $D_\sigma$ and the edge $\sigma$ by an edge $\epsilon$ of the set $\widetilde{\mathcal{E}}(D_\sigma)$ of edges (or faces) of the velocity cell $D_\sigma$ are defined by

$$F_{\sigma,\epsilon}^{(d,i)}(u_i) = \begin{cases} -\frac{|\epsilon|}{d_\epsilon}(u_{\sigma'} - u_\sigma) & \text{if } \epsilon = D_\sigma | D_{\sigma'} \\ -\frac{|\epsilon|}{d_\epsilon}(-u_\sigma) & \text{if } \epsilon \subset \partial\Omega \end{cases} \tag{182}$$

The numerical nonlinear convection flux $F_{\sigma,\epsilon}^{(c,i)}(\boldsymbol{u})(u_i)$ of the $i$th velocity component through an edge $\epsilon = \sigma|\sigma'$ separating the velocity cells $D_\sigma$ and $D_{\sigma'}$, $\sigma, \sigma' \in \mathcal{E}_{\mathrm{int}}^{(i)}$, which is an approximation of $\int_\epsilon u_i \boldsymbol{u} \cdot \boldsymbol{n}_{\sigma,\epsilon}$, can be defined by

$$F_{\sigma,\epsilon}^{(c,i)}(\boldsymbol{u})(u_i) = |\epsilon| u_{\sigma,\epsilon} \frac{u_\sigma + u_{\sigma'}}{2}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

where $u_{\sigma,\epsilon}$ is an approximation of the velocity flux $\boldsymbol{u}\cdot\boldsymbol{n}$ through the edge $\epsilon$. This flux must be chosen carefully to obtain the $L^2$ stability of the scheme. More precisely, a discrete counterpart of the free divergence of $\boldsymbol{u}$ must be satisfied also on the dual cells. Two cases can be distinguished:

- First case – the vector $\boldsymbol{e}_i$ is normal to $\epsilon$ and $\epsilon$ is included in a primal cell $K$ with $\mathcal{E}^{(i)}(K)=\{\sigma,\sigma'\}$. The mass flux through $\epsilon=\sigma|\sigma'$ is then given by

$$|\epsilon|u_{\sigma,\epsilon}=\frac{1}{2}\left(-|\sigma|u_{K,\sigma}+|\sigma'|u_{K,\sigma'}\right) \tag{183}$$

- Second case – the vector $\boldsymbol{e}_i$ is tangent to $\epsilon$ and $\epsilon$ is the union of the halves of two primal faces $\tau$ and $\tau'$ such that $\sigma=K|L$ with $\tau\in\mathcal{E}(K)$ and $\tau'\in\mathcal{E}(L)$. The mass flux through $\epsilon$ is then given by

$$|\epsilon|u_{\sigma,\epsilon}=\frac{1}{2}\left(|\tau|u_{K,\tau}+|\tau'|u_{L,\tau'}\right) \tag{184}$$

Using this definition, the usual finite volume property of local conservativity of the flux through an interface $\overrightarrow{\sigma|\sigma'}$ is obtained, that is,

$$|\epsilon|u_{\sigma,\epsilon}=-|\epsilon|u_{\sigma',\epsilon} \tag{185}$$

together with the following discrete free divergence condition on the dual cells:

$$\sum_{\epsilon\in\mathcal{E}(D_\sigma)}|\epsilon|u_{\sigma,\epsilon}=\frac{1}{2}\sum_{\sigma\in\mathcal{E}(K)}|\sigma|u_{K,\sigma}+\frac{1}{2}\sum_{\sigma\in\mathcal{E}(L)}|\sigma|u_{L,\sigma}=0 \tag{186}$$

Note that $u_{\sigma,\epsilon}=0$ if $\epsilon\subset\partial\Omega$, which is consistent with the boundary conditions (175c). A discretization of each component of (176b) may now be written as

$$\sum_{\epsilon\in\widetilde{\mathcal{E}}(D_\sigma)}F^{(d,i)}_{\sigma,\epsilon}(u_i)+\sum_{\epsilon\in\widetilde{\mathcal{E}}(D_\sigma)}C^{(i)}_{\mathcal{E}}(\boldsymbol{u})u_i+|D_\sigma|\delta_i p=|D_\sigma|f_{i,\sigma},\, i=1,\dots,d$$

where $\delta_i p$ denotes the discrete derivative of $p$ given by the following (natural) differential quotient for $i=1,\dots,d$

$$(\delta_i p)_\sigma=\frac{1}{d_{K,L}}(p_L-p_K)\quad\text{for }\sigma=K|L$$

with $K,L$ chosen such that $\boldsymbol{n}_{K,\sigma}\cdot\boldsymbol{e}_i=1$. Let $L_{\mathcal{T}}$ denote the space of functions that are piecewise constant on the pressure control volumes and let $p=\sum_{K\in\mathcal{T}}p_K\mathbb{1}_K$. The discrete gradient of $p$ may be defined as $\nabla_{\mathcal{E}}p=(\delta_1 p,\dots,\delta_d p)^t$ where $\delta_i p=\sum_{\sigma\in\mathcal{E}^{(i)}_{\rm int}}(\delta_i p)_\sigma\mathbb{1}_{D_\sigma}$. Using this notation, the following discrete duality property can be succinctly written:

$$\int_\Omega q\ \mathrm{div}_{\mathcal{T}}\boldsymbol{v}+\int_\Omega\nabla_{\mathcal{E}}q\cdot\boldsymbol{v}=0,\ \ \forall q\in L_{\mathcal{T}}\quad\forall\boldsymbol{v}\in\mathbf{H}_{\mathcal{E},0} \tag{187}$$

In order to finish writing the scheme, let $L_{\mathcal{T},0}$ denote the set of functions that have zero mean value and are piecewise constant on the pressure control volumes (i.e., the control volumes of

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

the rectangular mesh $\mathcal{T}$, pointed cell in Figure 13). The $i$th component of the discrete Laplace operator is classically defined as

$$-\Delta_{\mathcal{E}}^{(i)}: \quad \left| \begin{array}{l} H_{\mathcal{E},0}^{(i)} \longrightarrow H_{\mathcal{E},0}^{(i)} \\ u_i \longmapsto -\Delta_{\mathcal{E}} u_i = -\displaystyle\sum_{\sigma\in\mathcal{E}^{(i)}} \frac{1}{|D_\sigma|} \sum_{\epsilon\in\tilde{\mathcal{E}}(D_\sigma)} F_{\sigma,\epsilon}^{(d,i)}(u_i)\, \mathbb{1}_{D_\sigma} \end{array} \right. \tag{188}$$

where $\tilde{\mathcal{E}}(D_\sigma)$ denotes the faces of $D_\sigma$. Then the discrete Laplace operator of the full velocity vector is defined by

$$\begin{array}{ll} -\Delta_{\mathcal{E}}: & \mathbf{H}_{\mathcal{E},0} \longrightarrow \mathbf{H}_{\mathcal{E},0} \\ & \boldsymbol{v} \mapsto -\Delta_{\mathcal{E}}\boldsymbol{v} = (-\Delta_{\mathcal{E}}^{(1)} u_1, \ldots, -\Delta_{\mathcal{E}}^{(d)} u_d)^t \end{array} \tag{189}$$

Finally, define the $i$th component $C_{\mathcal{E}}^{(i)}(\boldsymbol{u})$ of the non linear convection operator by

$$C_{\mathcal{E}}^{(i)}(\boldsymbol{u}): \quad \left| \begin{array}{l} H_{\mathcal{E},0}^{(i)} \longrightarrow H_{\mathcal{E},0}^{(i)} \\ v_i \longmapsto C_{\mathcal{E}}^{(i)}(\boldsymbol{u}) v_i = \displaystyle\sum_{\sigma\in\tilde{\mathcal{E}}_{\rm int}^{(i)}} \frac{1}{|D_\sigma|} \sum_{\substack{\epsilon\in\tilde{\mathcal{E}}(D_\sigma) \\ \epsilon=\sigma|\sigma'}} F^{(c,i)}(\boldsymbol{u})_{\sigma,\epsilon}(v_i)\mathbb{1}_{D_\sigma} \end{array} \right. \tag{190}$$

and the full discrete convection operator $\boldsymbol{C}_{\mathcal{E}}(\boldsymbol{u}),\ \mathbf{H}_{\mathcal{E},0} \longrightarrow \mathbf{H}_{\mathcal{E},0}$ by

$$\boldsymbol{C}_{\mathcal{E}}(\boldsymbol{u})\boldsymbol{v} = (C_{\mathcal{E}}^{(1)}(\boldsymbol{u})v_1, \ldots, C_{\mathcal{E}}^{(d)}(\boldsymbol{u})v_d)^t$$

With these notations, the MAC scheme for the discretization of (176) on a rectangular (non uniform) staggered grid $\mathcal{T}$ can be written in following compact form for $\boldsymbol{u} \in \mathbf{H}_{\mathcal{E},0},\ p \in L_{\mathcal{T},0}$:

$$-\Delta_{\mathcal{E}}\boldsymbol{u} + \boldsymbol{C}_{\mathcal{E}}(\boldsymbol{u})\boldsymbol{u} + \boldsymbol{\nabla}_{\mathcal{E}} p = \boldsymbol{f}_{\mathcal{E}} \tag{191a}$$

$$\mathrm{div}_{\mathcal{T}}\boldsymbol{u} = 0 \tag{191b}$$

where $\boldsymbol{f}_{\mathcal{E}} = (f_{1,\mathcal{E}_1}, \ldots, f_{d,\mathcal{E}_d})$ and $f_{i,\mathcal{E}_i}$ is the $L^2$ projection on $H_{\mathcal{E}}^{(i)}$ defined by

$$f_{i,\mathcal{E}_i} = \sum_{\sigma\in\mathcal{E}_{\rm int}^{(i)}} \left( \frac{1}{|D_\sigma|} \int_{D_\sigma} f_i(\boldsymbol{x}) \right) \mathbb{1}_{D_\sigma}$$

When performing the convergence analysis of the scheme (191), it can be shown that sequences of approximate solutions tend to a weak solution of (176) as the mesh size tends to zero. A weak formulation of (176) is given by

$$\begin{array}{l} \text{Find } (\bar{\boldsymbol{u}}, \bar{p}) \in H_0^1(\Omega)^d \times L_0^2(\Omega) \ \text{ such that } \ \forall (\boldsymbol{v}, q) \in H_0^1(\Omega)^d \times L_0^2(\Omega) \\ \displaystyle\int_\Omega \nabla\bar{\boldsymbol{u}} : \nabla\boldsymbol{v} + \int_\Omega ((\bar{\boldsymbol{u}}\cdot\nabla)\bar{\boldsymbol{u}})\cdot\boldsymbol{v} - \int_\Omega \bar{p}\,\mathrm{div}\boldsymbol{v} = \int_\Omega \boldsymbol{f}\cdot\boldsymbol{v} \qquad (192a) \\ \displaystyle\int_\Omega q\ \mathrm{div}\bar{\boldsymbol{u}} = 0 \qquad (192b) \end{array}$$

where $L_0^2(\Omega)$ stands for the subspace of $L^2(\Omega)$ of zero mean-valued functions. Similarly, a weak form of the scheme equivalent to (191) is convenient for use in convergence analysis. As

in Section 5, a discrete $H_0^1$ inner product is defined; however, here it must be defined on the velocity grid. This inner product satisfies $\forall(\boldsymbol{u},\boldsymbol{v}) \in {\mathbf{H}_{\mathcal{E},0}}^2$

$$\int_\Omega -\Delta_{\mathcal{E}}\boldsymbol{u}\cdot\boldsymbol{v} = [\boldsymbol{u},\boldsymbol{v}]_{1,\mathcal{E},0} = \sum_{i=1}^{d}[u_i,v_i]_{1,\mathcal{E}^{(i)},0} \tag{193}$$

with

$$[u_i,v_i]_{1,\mathcal{E}^{(i)},0} = \sum_{\substack{\epsilon\in\widetilde{\mathcal{E}}^{(i)}_{\text{int}}\\ \epsilon=\overrightarrow{\sigma|\sigma'}}} \frac{|\epsilon|}{d_\epsilon}\,(u_\sigma - u_{\sigma'})\,(v_\sigma - v_{\sigma'}) + \sum_{\substack{\epsilon\in\widetilde{\mathcal{E}}^{(i)}\\ \epsilon\subset\partial(D_\sigma)}} \frac{|\epsilon|}{d_\epsilon}\,u_\sigma\,v_\sigma$$

The bilinear forms $\left|\begin{array}{l} H^{(i)}_{\mathcal{E},0}\times H^{(i)}_{\mathcal{E},0}\to\mathbb{R}\\ (u,v)\mapsto [u_i,v_i]_{1,\mathcal{E}^{(i)},0}\end{array}\right.$ and $\left|\begin{array}{l}\mathbf{H}_{\mathcal{E},0}\times\mathbf{H}_{\mathcal{E},0}\to\mathbb{R}\\ (\boldsymbol{u},\boldsymbol{v})\mapsto[\boldsymbol{u},\boldsymbol{v}]_{1,\mathcal{E},0}\end{array}\right.$ are inner products on $H^{(i)}_{\mathcal{E},0}$ and $\mathbf{H}_{\mathcal{E},0}$ respectively, which induce the following scalar and vector discrete $H_0^1$ norms:

$$\|u_i\|^2_{1,\mathcal{E}^{(i)},0} = [u_i,u_i]_{1,\mathcal{E}^{(i)},0} = \sum_{\substack{\epsilon\in\widetilde{\mathcal{E}}^{(i)}_{\text{int}}\\ \epsilon=\overrightarrow{\sigma|\sigma'}}} \frac{|\epsilon|}{d_\epsilon}\,(u_\sigma - u_{\sigma'})^2 + \sum_{\substack{\epsilon\in\widetilde{\mathcal{E}}^{(i)}_{\text{ext}}\\ \epsilon\subset\partial(D_\sigma)}} \frac{|\epsilon|}{d_\epsilon}\,u_\sigma^2 \text{ for } i=1,\dots,d \tag{194a}$$

and (194b)

$$\|\boldsymbol{u}\|^2_{1,\mathcal{E},0} = [\boldsymbol{u},\boldsymbol{u}]_{1,\mathcal{E},0} = \sum_{i=1}^{d}\|u_i\|^2_{1,\mathcal{E}^{(i)},0} \tag{194c}$$

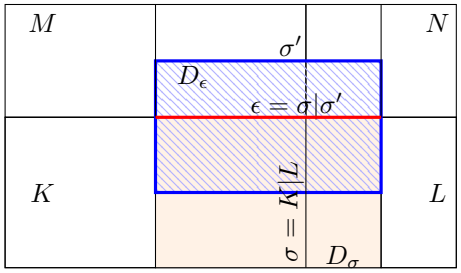


Figure 14. Full grid for definition of the derivative of the velocity.

When utilizing Cartesian grids, this inner product may be formulated as the $L^2$ inner product of discrete gradients. Consider the following discrete gradient of each velocity component $u_i$:

$$\nabla_{\mathcal{E}^{(i)}}u_i = (\delta_1 u_i,\dots,\delta_d u_i) \text{ with } \delta_j u_i = \sum_{\substack{\epsilon\in\widetilde{\mathcal{E}}^{(i)}\\ \epsilon\perp\boldsymbol{e}_j}} (\delta_j u_i)_{D_\epsilon}\,\mathbb{1}_{D_\epsilon} \tag{195}$$

where $(\delta_j u_i)_{D_\epsilon} = \dfrac{u_{\sigma'}-u_\sigma}{d_\epsilon}$ with $\epsilon = \overrightarrow{\sigma|\sigma'}$, and $D_\epsilon = \epsilon\times\boldsymbol{x}_\sigma\boldsymbol{x}_{\sigma'}$; see Figure 14. This definition is compatible with the definition of the discrete derivative $(\delta_i u_i)_K$ given by (179), since, if

*Encyclopedia of Computational Mechanics*. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

$\epsilon \subset K$ then $D_\epsilon = K$. With this definition, it follows that

$$\int_\Omega \nabla_{\mathcal{E}^{(i)}} u \cdot \nabla_{\mathcal{E}^{(i)}} v = [u, v]_{1,\mathcal{E}^{(i)},0}, \quad \forall u, v \in H^{(i)}_{\mathcal{E},0}, \forall i = 1, \ldots, d \tag{196}$$

where $[u, v]_{1,\mathcal{E}^{(i)},0}$ is the discrete $H^1_0$ inner product defined by (193). Defining the following gradient for $\boldsymbol{u}$:

$$\nabla_{\mathcal{E}} \boldsymbol{u} = (\nabla_{\mathcal{E}^{(1)}} u_1, \ldots, \nabla_{\mathcal{E}^{(d)}} u_d)$$

it follows from the component formulas that

$$\int_\Omega \nabla_{\mathcal{E}} \boldsymbol{u} : \nabla_{\mathcal{E}} \boldsymbol{v} = [\boldsymbol{u}, \boldsymbol{v}]_{1,\mathcal{E},0}$$

With this formulation, the MAC scheme for the linear Stokes problem can be interpreted as a gradient scheme in the sense introduced in Eymard *et al.*, 2012 (see Eymard *et al.*, 2015 for more details on the generalization of this formulation to other schemes). Thanks to this result, (strong) convergence of this discrete gradient to the gradient of the exact velocity as well as strong convergence of the pressure can be shown.

Defining a weak form $b_{\mathcal{E}}$ of the nonlinear convection operator

$$b_{\mathcal{E}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = \sum_{i=1}^{d} b^{(i)}_{\mathcal{E}}(\boldsymbol{u}, v_i, w_i), \quad \forall (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \in {\mathbf{H}_{\mathcal{E},0}}^3 \tag{197}$$

where for $i = 1, \ldots, d$

$$b^{(i)}_{\mathcal{E}}(\boldsymbol{u}, v_i, w_i) = \int_\Omega C^{(i)}_{\mathcal{E}}(\boldsymbol{u}) v_i \; w_i$$

it is now possible to introduce a weak formulation

$$\text{Find } (\boldsymbol{u}, p) \in \mathbf{H}_{\mathcal{E},0} \times L_{\mathcal{T},0} \text{ such that } \forall (\boldsymbol{v}, q) \in \mathbf{H}_{\mathcal{E},0} \times L_{\mathcal{T}}$$

$$\int_\Omega \nabla_{\mathcal{E}} \boldsymbol{u} : \nabla_{\mathcal{E}} \boldsymbol{v} + b_{\mathcal{E}}(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{v}) - \int_\Omega p \, \mathrm{div}_{\mathcal{T}}(\boldsymbol{v}) = \int_\Omega \boldsymbol{f}_{\mathcal{E}} \cdot \boldsymbol{v} \tag{198a}$$

$$\int_\Omega \mathrm{div}_{\mathcal{T}} \boldsymbol{u} \, q = 0 \tag{198b}$$

which is equivalent to the MAC scheme (191).

*6.2.2. Convergence analysis of the MAC scheme.* The proof of convergence of the MAC scheme using this latter weak form then closely follows the proof of existence of a solution to the Navier-Stokes equations; see for example Boyer and Fabrie, 2013. This analysis requires estimates on the trilinear form $b_{\mathcal{E}}$.

**Lemma 11 (Estimates on $b_{\mathcal{E}}$)** *Let $\mathcal{D} = (\mathcal{T}, \mathcal{E})$ be a MAC grid and let $b_{\mathcal{E}}$ be defined by* (197)*. For $d = 3$, there exists $C_{\eta_{\mathcal{T}}} > 0$, depending only on the regularity $\eta_{\mathcal{T}}$ of the mesh defined by* (181)*, such that*

$$|b_{\mathcal{E}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})| \leq C_{\eta_{\mathcal{T}}} \; \|\boldsymbol{u}\|_{L^4(\Omega)^d} \|\boldsymbol{v}\|_{1,\mathcal{E},0} \; \|\boldsymbol{w}\|_{L^4(\Omega)^d}, \quad \forall (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \in \boldsymbol{E}_{\mathcal{E}} \times {\mathbf{H}_{\mathcal{E},0}}^2 \tag{199}$$

*and*

$$|b_{\mathcal{E}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w})| \leq C_{\eta_{\mathcal{T}}} \; \|\boldsymbol{u}\|_{1,\mathcal{E},0} \; \|\boldsymbol{v}\|_{1,\mathcal{E},0} \|\boldsymbol{w}\|_{1,\mathcal{E},0}, \quad \forall (\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \in \boldsymbol{E}_{\mathcal{E}} \times {\mathbf{H}_{\mathcal{E},0}}^2 \tag{200}$$

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

An important property needed to obtain some *a priori* estimates on the velocity is that the nonlinear convection term vanishes when taking $\boldsymbol{u}$ as test function in (192). This is also the case for its discrete counterpart, as stated in the next lemma.

**Lemma 12 ($b_{\mathcal{E}}$ is skew-symmetric)** *Let $(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) \in \boldsymbol{E}_{\mathcal{E}} \times \mathbf{H}_{\mathcal{E},0} \times \mathbf{H}_{\mathcal{E},0}$, then*

$$b_{\mathcal{E}}(\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}) = -b_{\mathcal{E}}(\boldsymbol{u}, \boldsymbol{w}, \boldsymbol{v}) \tag{201}$$

*and therefore*

$$b_{\mathcal{E}}(\boldsymbol{u}, \boldsymbol{u}, \boldsymbol{u}) = 0 \quad \forall \boldsymbol{u} \in \boldsymbol{E}_{\mathcal{E}} \tag{202}$$

From this property, using a so-called Fortin operator to obtain a discrete divergence free test function, following estimates and existence result are obtained:

**Theorem 28 (Existence and estimates)** *There exists a solution to* (198) *and $C_{\eta_{\mathcal{T}}} > 0$, depending only on the regularity $\eta_{\mathcal{T}}$ of the mesh and $\Omega$, such that any solution of* (198) *satisfies the stability estimate*

$$\|\boldsymbol{u}\|_{1,\mathcal{E},0} + \|p\|_{L^2(\Omega)} \leq C_{\eta_{\mathcal{T}}} \, \|\boldsymbol{f}\|_{L^2(\Omega)^d} \tag{203}$$

A challenging part of the convergence analysis is the study of the nonlinear convection term. One approach is to reconstruct a full grid velocity for each component, which converges as the component itself.

**Lemma 13 (Weak consistency of the nonlinear convection term)** *Let $(\mathcal{D}_n)_{n\in\mathbb{N}}$ with $\mathcal{D}_n = (\mathcal{T}_n, \mathcal{E}_n)$ be a sequence of meshes such that $h_{\mathcal{T}_n} = \max_{K\in\mathcal{T}_n} \operatorname{diam}(K) \to 0$ as $n \to +\infty$. Further, assume that there exists $\eta > 0$ such that $\eta_{\mathcal{T}_n} \leq \eta$ for any $n \in \mathbb{N}$ (with $\eta_{\mathcal{T}_n}$ defined by* (181)*). Let $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{w}_n)_{n\in\mathbb{N}}$ be two sequences of functions such that*

- *$\boldsymbol{v}_n \in \mathbf{H}_{\mathcal{E}_n,0}$ and $\boldsymbol{w}_n \in \mathbf{H}_{\mathcal{E}_n,0}$,*
- *the sequences $(\boldsymbol{v}_n)_{n\in\mathbb{N}}$ and $(\boldsymbol{w}_n)_{n\in\mathbb{N}}$ converge in $L^2(\Omega)^d$ to $\bar{\boldsymbol{v}} \in L^2(\Omega)^d$ and $\bar{\boldsymbol{w}} \in L^2(\Omega)^d$ respectively.*

*Let $(\Pi_{\mathcal{E}_n})_{n\in\mathbb{N}}$ be a family of interpolators satisfying* (180) *and let $\boldsymbol{\varphi} \in C_c^\infty(\Omega)^d$. Then $b_{\mathcal{E}}(\boldsymbol{v}_n, \boldsymbol{w}_n, \Pi_{\mathcal{E}_n}\boldsymbol{\varphi}) \to b(\bar{\boldsymbol{v}}, \bar{\boldsymbol{w}}, \boldsymbol{\varphi})$ as $n \to +\infty$.*

The above consistency result together with the estimates on the velocity and the pressure yield the following convergence result:

**Theorem 29 (Convergence of the scheme)** *Let $(\mathcal{D}_n)_{n\in\mathbb{N}}$ with $\mathcal{D}_n = (\mathcal{T}_n, \mathcal{E}_n)$ be a sequence of meshes such that $h_{\mathcal{T}_n} = \max_{K\in\mathcal{T}_n} \operatorname{diam}(K) \to 0$ as $n \to +\infty$ . Further, assume that there exists $\eta > 0$ such that $\eta_{\mathcal{T}_n} \leq \eta$ for any $n \in \mathbb{N}$ [with $\eta_{\mathcal{T}_n}$ defined by* (181)*]. Let $(\boldsymbol{u}_n, p_n)$ be a solution to the MAC scheme* (191) *or its weak form* (198)*, for $\mathcal{D} = \mathcal{D}_n$, then there exists $\bar{\boldsymbol{u}} \in H_0^1(\Omega)^d$ and $\bar{p} \in L^2(\Omega)$ such that, up to a subsequence*

- *the sequence $(\boldsymbol{u}_n)_{n\in\mathbb{N}}$ converges to $\bar{\boldsymbol{u}}$ in $L^2(\Omega)^d$,*
- *the sequence $(\nabla_n \boldsymbol{u}_n)_{n\in\mathbb{N}}$ converges to $\nabla\bar{\boldsymbol{u}}$ in $L^2(\Omega)^{d\times d}$,*

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

- *the sequence* $(p_n)_{n\in\mathbb{N}}$ *converges to* $\bar{p}$ *in* $L^2(\Omega)$,
- $(\bar{\boldsymbol{u}}, \bar{p})$ *is a solution to the weak formulation* (192).

The proof of this result is obtained by taking interpolates of smooth functions as test functions in the weak form of the scheme (198) and passing to the limit in the weak form of the scheme Herbin *et al.*, 2014; Gallouët *et al.*, 2015a.

*6.2.3. Further developments in the incompressible and compressible case.* The time dependent Navier–Stokes equations can also be easily discretized with the MAC scheme, either with a time implicit scheme or else with a pressure correction scheme. In the case of the implicit scheme, the scheme is proven to be convergent in Gallouët *et al.*, 2015a. The case of the pressure correction scheme remains an open question.

The MAC scheme can also be easily written and implemented for compressible flows. Some theoretical results have been proven in the case of a perfect gas and the implicit Euler scheme for the compressible Stokes equations (Eymard *et al.*, 2010b) and in the case of the semi-stationary Navier–Stokes equations (Gallouët *et al.*, 2015c). Convergence results and error estimates were also obtained for the same type of schemes on simplicial and quadrilateral staggered grids using the Crouxeix–Raviart and the Rannacher–Turek finite element spaces for the discretization of the diffusion operator; see Gallouët *et al.*, 2015b,Gallouët *et al.*, 2015d.

In the case of the Euler and Navier–Stokes equations, the convergence remains an open question because of the lack of estimates. However, stability results exist. In particular, the conservation of the discrete kinetic equation can be obtained (Gallouët *et al.*, 2010; Herbin and Latché, 2010). The weak consistency of the scheme has been proved for a decoupled scheme in the case of the Euler (Herbin *et al.*, 2013; Herbin *et al.*, 2016b). Weak consistency in this context means that if some estimates on the approximate solutions are assumed, then a sequence of approximate solutions can be shown to converge to a weak solution of the system as the mesh and time steps tend to zero (under appropriate CFL conditions). The limit of the scheme may also be shown to be an entropy weak solution of the Euler system (for the perfect gas EOS). Note that similar type schemes on simplicial and quadrilateral staggered grids have been developed and studied using the Crouxeix–Raviart and the Rannacher–Turek finite element spaces for the discretization of the diffusion operator. Both an implicit scheme and a pressure correction scheme have been studied in Gallouët *et al.*, 2008; Babik *et al.*, 2011; Gastaldo *et al.*, 2011.

## 7. Related Chapters

(*See also* Finite Element Methods, Discontinuous Galerkin Methods for Computational Fluid Dynamics, Aerodynamics)

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

REFERENCES

I. Aavatsmark, T. Barkve, O. Boe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. II. Discussion and numerical results. *SIAM J. Sci. Comput.* 1998a; **19**(5):1717–1736.

I. Aavatsmark, T. Barkve, O. Boe, and T. Mannseth. Discretization on unstructured grids for inhomogeneous, anisotropic media. Part I: Derivation of the methods. *SIAM Journal on Sci. Comp.* 1998b; **19**:1700–1716.

I. Aavatsmark, E. Reiso, and E. Teigland. Control volume discretization for quadrilateral grids with faults and local refinements. *Comput. Geosci.* 2001; **5**:1–23.

I. Aavatsmark, G. T. Eigestad, B. T. Mallison, and J. M. Nordbotten. A compact multipoint flux approximation method with improved robustness. *Numer. Methods Partial Differential Equations* 2008; **24**(5):1329–1360.

R. Abgrall. On essentially non-oscillatory schemes on unstructured meshes: analysis and implementation. *J. Comput. Phys.* 1994; **114**:45–58.

L. Agelas, D. A. Di Pietro, and J. Droniou. The G method for heterogeneous anisotropic diffusion on general meshes. *M2AN Math. Model. Numer. Anal.* 2010; **44**(4):597–625.

D. Amadori and L. Gosse. Trror Estimates for well-balanced and time-split schemes on a locally damped wave equation. *Math. Comp.* 2016; **85**(298):601–633.

B. Andreianov, M. Bendahmane, and K. Karlsen. A gradient reconstruction formula for finite-volume schemes and discrete duality. In *Finite volumes for complex applications V*, ISTE, London 2008, pp. 161–168.

B. Andreianov, F. Boyer, and F. Hubert. Discrete duality finite volume schemes for Leray-Lions-type elliptic problems on general 2D meshes. *Numer. Methods Partial Differential Equations* 2007; **23**(1):145–195.

B. Andreianov, M. Bendahmane, and K. H. Karlsen. Discrete duality finite volume schemes for doubly nonlinear degenerate hyperbolic-parabolic equations. *J. Hyperbolic Differ. Eq.* 2010; **7**(1):1–67.

B. Andreianov, R. Eymard, M. Ghilani, and. N Marhraoui. Finite volume approximation of degenerate two-phase flow model with unlimited air mobility. *Numer. Methods Partial Differential Equations* 2013; **2**(29):441–474.

B. Andreianov, M. Bendahmane, and M. Saad. Finite volume methods for degenerate chemotaxis model. *J. Comput. Appl. Math.* 2011; **235**(14):4015–4031.

B. Andreianov, M. Bendahmane, and R. Ruiz-Baier. Analysis of a finite volume method for a cross-diffusion model in population dynamics. *J. Comput. Appl. Math.* 2011; **235**(2):307–344.

B. Andreianov, P. Goatin, and N. Seguin. Finite volume schemes for locally constrained conservation laws. *Numer. Math.* 2010; **115**(4):609–645.

T. Arbogast, D. Estep, B. Sheehan, and S. Tavener. *A posteriori* error estimates for mixed finite element and finite volume methods for problems coupled through a boundary with nonmatching grids. *IMA J. Numer. Anal.* 2014; **34**(4):1625–1653.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

F. Archambeau, J.-M. Hérard, and J. Laviéville. Comparative study of pressure-correction and Godunov-type schemes on unsteady compressible cases. *Comput. & Fluids* 2009; **38**(8):1495–1509.

F. Babik, R. Herbin, W. Kheriji, and J.-C. Latché. Discretization of the viscous dissipation term with the MAC scheme. In *Finite volumes for complex applications. VI. Problems & perspectives. Volume 1, 2*, vol. 4 of *Springer Proc. Math.*, Springer, Heidelberg 2011, pp. 571–579.

J. Ball. *A version of the fundamental theorem for Young measures*, *PDEs and continuum models of phase transitions (Nice, 1988). Lecture Notes in Phys.* 1989; **344**:207-215, Springer, Berlin, 1989.

R. Bank and D. J. Rose. Some error estimates for the box method. *SIAM J. Numer. Anal.* 1987; **24**:777–787.

T. J. Barth and P. Frederickson 1990. *Higher order solution of the Euler equations on unstructured grids using quadratic reconstruction.* Report 90-0013, American Institute for Aeronautics and Astronautics, 1990.

T. J. Barth and D. C. Jespersen 1989. *The design and application of upwind schemes on unstructured meshes.* Report 89-0366, American Institute for Aeronautics and Astronautics, 1989.

T. J. Barth. Numerical methods for gasdynamic systems on ynstructured meshes. In Kröner, Ohlberger, & Rohde (eds.), *An Introduction to Recent Developments in Theory and Numerics for Conservation Laws*, vol. 5 of *Lecture Notes in Computational Science and Engineering*, pp. 195–285. Springer-Verlag, Heidelberg 1998.

T. J. Barth. On discontinuous Galerkin approximations of Boltzmann moment systems with Levermore closure. *Comput. Meth. Appl. Mech. Engrg.* 2006; **195**:3311–3330.

T. J. Barth. On the role of involutions in the discontinuous Galerkin discretization of Maxwell and magnetohydrodynamics systems. *IMA Vol. Math. Appl.* 2007; Springer, **142**:69–88.

P. Batten, C. Lambert, and D. M. Causon. Positively conservative high-resolution convection schemes for unstructured elements. *Int. J. Numer. Meth. Engrg.* 1996; **39**:1821–1838.

A. Berman and R. J. Plemmons. *Nonnegative matrices in the mathematical sciences*, volume 9 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. Revised reprint of the 1979 original.

C. Berthon, Y. Coudière, and V. Desveaux. Second-order MUSCL schemes based on Dual Mesh Gradient Reconstruction (DMGR). *Mathematical Modelling and Numerical Analysis* 2014; **48**:583–602.

C. Berthon, M. Breuß, and M.-O. Titeux. A relaxation scheme for the approximation of the pressureless Euler equations. *Numer. Methods Partial Differential Equations* 2006; **22**(2):484–505.

C. Berthon. A fully well-balanced, positive and entropy-satisfying Godunov-type method for the shallow-water equations. *Math. Comp.*, 85(299):1281–1307, 2016.

M. Bessemoulin-Chatard, C. Chainais-Hillairet, and F. Filbet. On discrete functional inequalities for some finite volume schemes. *IMA J. Numer. Anal.* 2015; **35**(3):1125–1149.

M. Bessemoulin-Chatard, C. Chainais-Hillairet, and M.H. Vignal. Study of a finite volume scheme for the drift-diffusion system. Asymptotic behavior in the quasi-neutral limit. *SIAM J. Numer. Anal.* 2014; **52**(4):1666–1691.

M. Bessemoulin-Chatard, C. Chainais-Hillairet, and F. Filbet. A finite volume scheme for nonlinear degenerate parabolic equations. *SIAM J. Sci. Comput.* 2012; **34**(5):B559–B583, 2012.

V. Billey, J. Périaux, P. Perrier, and B. Stoufflet. 2-D and 3-D Euler computations with finite element methods in aerodynamics. vol. 1270 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1987.

P. Blanc. Error estimate for a finite volume scheme on a MAC mesh for the Stokes problem. In *Finite volumes for complex applications II*, Hermes Sci. Publ., Paris, 1999, pp. 117–124.

P. Blanc. Convergence of a finite volume scheme on a MAC mesh for the Stokes problem with right-hand side in $H^{-1}$. In *Finite volumes for complex applications IV*, ISTE, London, 2005, pp. 133–142.

L. Boccardo and T. Gallouët. Nonlinear elliptic and parabolic equations involving measure data. *J. Funct. Anal.* 1989; **87**(1):149–169.

J. Bonelle and A. Ern. Analysis of compatible discrete operator schemes for elliptic problems on polyhedral meshes. *ESAIM Math. Model. Numer. Anal.* 2014; **48**(2):553–581.

J. Bonelle, D. A. Di Pietro, and A. Ern. Low-order reconstruction operators on polyhedral meshes: application to compatible discrete operator schemes. *Comput. Aided Geom. Design* 2015; **35/36**:27–41.

J. Boris and D. Book. Flux corrected transport: SHASTA, a fluid transport algorithm that works. *J. Comp. Phys.* 1973; **11**:38–69.

F. Bouchut & B. Perthame. Kruzkov's estimates for scalar conservation laws revisited. *Trans. Am. Math. Soc.* 1998; **350**(7):2847–2870.

F. Bouchut. *Nonlinear stability of finite volume methods for hyperbolic conservation laws and well-balanced schemes for sources.* Frontiers in Mathematics. Birkhäuser Verlag, Basel, 2004.

F. Bouchut, Christian Klingenberg, and Knut Waagan. An approximate Riemann solver for ideal MHD based on relaxation. In *Hyperbolic problems: theory, numerics and applications*, volume 67 of *Proc. Sympos. Appl. Math.*, pages 439–443. Amer. Math. Soc., Providence, RI, 2009.

F. Bouchut and Vladimir Zeitlin. A robust well-balanced scheme for multi-layer shallow water equations. *Discrete Contin. Dyn. Syst. Ser. B* 2010; **13**(4):739–758.

F. Bouchut and Vladimir Zeitlin. Finite volume schemes for the approximation via characteristics of linear convection equations with irregular data. *J. Evol. Equ.* 2011; **113**(4):687–724.

N. Bouillard, R. Eymard, R. Herbin, and Ph. Montarnal. Diffusion with dissolution and precipitation in a porous medium: mathematical analysis and numerical approximation of a simplified model. *M2AN Math. Model. Numer. Anal.* 2007; **41**(6):975–1000.

F. Boyer and F. Hubert. Finite volume method for 2D linear and nonlinear elliptic problems with discontinuities. *SIAM Journal on Numerical Analysis* 2008; **46**(6):3032–3070.

F. Boyer and P. Fabrie. *Mathematical tools for the study of the incompressible Navier-Stokes equations and related models*, vol. 183 of *Applied Mathematical Sciences*. Springer, New York, 2013.

A. Bradji. Some simple error estimates of finite volume approximate solution for parabolic equations. *Comptes Rendus Mathematique* 2008; **346**(9):571–574.

A. Bradji and R. Herbin. Discretization of coupled heat and electrical diffusion problems by finite-element and finite-volume methods. *IMA journal of numerical analysis* 2008; **28**(3):469–495.

A. Bradji and J. Fuhrmann. Error estimates of the discretization of linear parabolic equations on general nonconforming spatial grids. *Comptes Rendus Mathematique* 2010; **348**(19):1119–1122.

G. Brun, J.-M. Hérard, D. Jeandel, and M. Uhlmann. An approximate Roe-type Riemann solver for a class of realizable second order closures. *Int. J. Comput. Fluid Dyn.* 2000; **13**(3):233–249.

T. Buffard and S. Clain. Monoslope and multislope MUSCL methods for unstructured meshes. *Journal of Computational Physics* 2010; **229**:3745–3776.

T. Buffard, T. Gallouët, and J.-M. Hérard. A sequel to a rough Godunov scheme: application to real gases. *Comput. & Fluids* 2000; **29**(7):813–847.

Z. Q. Cai. On the finite volume element method. *Numer. Math.* 1991; **58**(7):713–735.

Z. Q. Cai and S. McCormick. On the accuracy of the finite volume element method for diffusion equations on composite grids. *SIAM J. Numer. Anal.* 1990; **27**(3):636–655.

Z. Q. Cai, J. Mandel, and S. McCormick. The finite volume element method for diffusion equations on general triangulations. *SIAM J. Numer. Anal.* 1991; **28**(2):392–402.

C. Calgaro, E. Chane-Kane, E. Creusé, and T. Goudon. $L^\infty$-stability of vertex-based MUSCL finite volume schemes on unstructured grids: Simulation of incompressible flows with high density ratios. *Journal of Computational Physics* 2010; **229**:6027–6046.

C. Cancès, M. Cathala, and C. Le Potier. Monotone corrections for generic cell-centered finite volume approximations of anisotropic diffusion equations. *Numer. Math.* 2013; **125**(3):387–417.

C. Cancès, I. S. Pop, and M. Vohralík. An a posteriori error estimate for vertex-centered finite volume discretizations of immiscible incompressible two-phase flow. *Math. Comp.* 2014; **83**(285):153–188.

J. Carrillo. Entropy solutions for nonlinear degenerate problems. *Arch. Ration. Mech. Anal.* 1999; **147**:269–361.

C. Chalons, and F. Coquel. Modified Suliciu relaxation system and exact resolution of isolated shock waves. *Math. Models Methods Appl. Sci.* 2014; **24**(5):937–971.

C. Chainais-Hillairet. Finite volume schemes for a nonlinear hyperbolic equation: convergence towards the entropy solution and error estimates. *M2AN Math. Model. Numer. Anal.* 1999; **33**:129–156.

C. Chainais-Hillairet. Second-order finite-volume schemes for a non-linear hyperbolic equation: error estimates. *Math. Methods Appl. Sci.* 2000; **23**(5):467–490.

C. Chainais-Hillairet and Y.-J. Peng. Finite volume approximation for degenerate drift-diffusion system in several space dimensions. *SMath. Methods Appl. Sci.* 2004; **14**(3):461–481.

C. Chainais-Hillairet and J. Droniou. Convergence analysis of a mixed finite volume scheme for an elliptic-parabolic system modeling miscible fluid flows in porous media. *SIAM J. Numer. Anal.* 2007; **45**(5):2228–2258.

C. Chainais-Hillairet and F. Filbet. Asymptotic behaviour of a finite-volume scheme for the transient drift-diffusion model. *IMA J. Numer. Anal.* 2007; **27**(45):689–716.

C. Chainais-Hillairet and J. Droniou. Finite-volume schemes for noncoercive elliptic problems with Neumann boundary conditions. *IMA J. Numer. Anal.* 2011; **31**(1):61–85.

C. Chainais-Hillairet and M. Gisclon and A. Jüngel. A finite-volume scheme for the multidimensional quantum drift-diffusion model for semiconductors. *Numer. Meth. Partial Diff. Eq.* 2011; **27**(6):1483–1510.

C. Chainais-Hillairet, S.. Krell and A. Mouton. Study of discrete duality finite volume schemes for the Peaceman model. *SIAM J. Sci. Comput.* 2013; **35**(6):A2928–A2952.

C. Chainais-Hillairet, S.. Krell and A. Mouton. Convergence analysis of a DDFV scheme for a system describing miscible fluid flows in porous media. *Numer. Methods Partial Differential Equations* 2015; **31**(3):723–760.

S Champier and T Gallouët. Convergence d'un schéma décentré amont sur un maillage triangulaire pour un problème hyperbolique linéaire. *Modélisation mathématique et analyse numérique* 1992; **26**(7):835–853.

S. Champier, T. Gallouët, and R. Herbin. Convergence of an upstream finite volume scheme for a nonlinear hyperbolic equation on a triangular mesh. *Numer. Math.* 1993; **66**(2):139–157.

P. Chatzipantelidis. A Finite volume method based on the Crouzeix-Raviart element for elliptic problems. *Numer. Math.* 1999; **82**:409–432.

P. Chatzipantelidis and R. D. Lazarov. Error estimates for a finite volume element method for elliptic PDEs in nonconvex polygonal domains. *SIAM J. Numer. Anal.* 2005; **42**(5):1932–1958.

P. Chatzipantelidis, R. Lazarov, and V. Thomée. Some error estimates for the finite volume element method for a parabolic problem. *Comput. Methods Appl. Math.* 2013; **13**(3):251–279.

L. Chen and M. Wang. Cell conservative flux recovery and a posteriori error estimate of vertex-centered finite volume methods. *Adv. Appl. Math. Mech.* 2013; **5**(5):705–727.

Q. Chen and M. Gunzburger. Goal-oriented a posteriori error estimation for finite volume methods. *J. Comput. Appl. Math.* 2014; **265**:69–82.

E. Chénier, R. Eymard, T. Gallouët and R. Herbin. An extension of the MAC scheme to locally refined meshes : convergence analysis for the full tensor time-dependent Navier-Stokes equations. *Calcolo* 2015; **52**:69-107.

S. Chou and Q. Li. Error estimates in $L^2$, $H^1$ and $L^\infty$ in covolume methods for elliptic and parabolic problems: a unified approach. *Math. Comp.* 2000; **69**:103–120.

P. G. Ciarlet. Basic error estimates for elliptic problems. In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis, Volume II*, volume 2, pages 17–351. North Holland, 1991.

S. Clain and V. Clauzon. $L^\infty$ stability of the MUSCL methods. *Numerische Mathematik* 2010; **116**:31–64.

S. Clain. Finite volume maximum principle for hyperbolic scalar problems. *SIAM J. Numer. Anal.* 2013; **51**(1):467–490.

B. Cockburn, S. Y. Lin, and C. W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. III. One-dimensional systems. *J. Comp. Phys.* 1989; **84**:90–113.

B. Cockburn and C. W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. II. General framework. *J. Comp. Phys.* 1989; **52**:411–435.

B. Cockburn, S. Hou, and C. W. Shu. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws. IV. The multidimensional case. *Math. Comp.* 1990; **54**(190):545–581.

B. Cockburn, F. Coquel, and P. G. Lefloch. An error estimate for finite volume methods for multidimensional conservation laws. *Math. Comp.* 1994; **63**:77-103.

F. Coquel, J.-M. Hérard, K. Saleh, and N. Seguin. A robust entropy-satisfying finite volume scheme for the isentropic Baer-Nunziato model. *ESAIM Math. Model. Numer. Anal.* 2014; **48**(1):165–206.

F. Coquel, K. Saleh, and N. Seguin. A robust and entropy-satisfying numerical scheme for fluid flows in discontinuous nozzles. *Math. Models Methods Appl. Sci.* 2014; **24**(10):2043–2083.

B. Cockburn and H. Gau. *A posteriori* error estimates for general numerical methods for scalar conservation laws. *Comput. Appl. Math.* 1995; **14**:37–47.

B. Cockburn and P.-A. Gremaud. A priori error estimates for numerical methods for scalar conservation laws. Part 1: The general approach. *Math. Comput.* 1996a; **65**:533–573.

B. Cockburn and P.-A. Gremaud. *A Priori* error estimates for numerical methods for scalar conservation laws. Part 2: Flux splitting montone schemes on irregular cartesian grids. *Math. Comput.* 1996b; **66**:547–572.

B. Cockburn and P.-A. Gremaud. *A Priori* error estimates for numerical methods for scalar conservation laws. Part 3: Multidimensional flux-splitting monotone schemes on non-cartesian grids. *SIAM J. Numer. Anal.* 1998a; **35**:1775–1803.

B. Cockburn and C. W. Shu. The Runge-Kutta discontinuous Galerkin method for conservation laws. V. Multidimensional systems. *J. Comput. Phys.* 1998b; **141**(2):199–224.

B. Cockburn and G. Gripenberg. Continuous dependence on the nonlinearities of solutions of degenerate parabolic equations. *J. Diff. Equations* 1999; **151**(2):231–251.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

B. Cockburn. Continuous dependence and error estimates for viscosity methods. *Acta Numer.* 2003; **12**:127–180.

P. Colella and P. Woodward. The piecewise parabolic methods for gas-dynamical simulations. *J. Comp. Phys.* 1984; **54**:174–201.

F. Coquel and B. Perthame. Relaxation of energy and approximate Riemann solvers for general pressure laws in fluid dynamics. *SIAM J. Numer. Anal.* 1998; **35**(6):2223–2249.

Y. Coudière and F. Hubert. A 3D discrete duality finite volume method for nonlinear elliptic equations. *SIAM Journal on Scientific Computing* 2011; **33**(4):1739–1764.

Y. Coudière, F. Hubert, and G. Manzini. A CeVeFE DDFV scheme for discontinuous anisotropic permeability tensors. In *Finite volumes for complex applications VI*, vol. 4 of *Springer Proc. Math.*, pp. 283–291. Springer, Heidelberg 2011.

Y. Coudière, C. Pierre, O. Rousseau, and R. Turpault. A 2D/3D discrete duality finite volume scheme. Application to ECG simulation. *Int. J. Finite Vol.* 2009; **6**(1).

Y. Coudière, J.-P. Vila, and P. Villedieu. Convergence rate of a finite volume scheme for a two-dimensional convection-diffusion problem. *M2AN Math. Model. Numer. Anal.* 1999; **33**(3):493–516.

R. Courant, K. Friedrichs, and H. Lewy. Über die partiellen Differenzengleichungen der mathematischen Physik. *Math. Ann.* 1928; **100**(1):32–74.

R. Courant, E. Isaacson, and M. Rees. On the solution of nonlinear hyperbolic differential equations by finite differences. *Comm. Pure Appl. Math.* 1952; **5**:243–255.

R. Courant, K. Friedrichs, and H. Lewy. On the partial difference equations of mathematical physics. *IBM J. Res. Develop.* 1967; **11**:215–234.

P.-H. Cournède, C. Debiez, and A. Dervieux. *A positive MUSCL scheme for triangulations.* Report 3465, Institut National De Recherche En Informatique Et En Automatique (INRIA), 1998.

M. Crandall and A. Majda. Monotone difference approximations of scalar conservation laws. *Math. Comp.* 1980; **34**:1–21.

J.-P. Croisille and I. Greff. An efficient box-scheme for convection-diffusion equations with sharp contrast in the diffusion coefficients. *Comput. & Fluids* 2005; **34**(4-5):461–489.

A. Dedner, C. Makridakis, and M. Ohlberger. Error control for a class of Runge-Kutta discontinuous Galerkin methods for nonlinear conservation laws. *SIAM J. Numer. Anal.* 2007; **45**(2):514–538.

M. Delanaye. *Polynomial reconstruction finite volume schemes for the compressible Euler and Navier-Stokes equations on unstructured adaptive grids.* Ph.D. thesis, University of Liége, Belgium.

S. Dellacherie. Relaxation schemes for the multicomponent Euler system. *M2AN Math. Model. Numer. Anal.* 2003; **37**(6):909–936.

S. M. Deshpande. *On the Maxwellian distribution, symmetric form, and entropy conservation for the Euler equations.* Report TP-2613, NASA Langley, 1986.

J. A. Desideri and A. Dervieux. Compressible flow solvers using unstructured grids. von Karman Institute Lecture Series 1988-05, von Karman Institute, Brussels, 1988.

B. Desprès. An explicit a priori estimate for a finite volume approximation of linear advection on non Cartesian grids. *SIAM J. Numer. Anal.* 2004; **42**(2):484–504.

B. Desprès and C. Buet. The structure of well-balanced schemes for Friedrichs systems with linear relaxation. *Appl. Math. Comput.* 2016; **272**(2):440–459.

B. Desprès and F. Lagoutière. Generalized Harten formalism and longitudinal variation diminishing schemes for linear advection on arbitrary grids. *ESAIM: Mathematical Modelling and Numerical Analysis* 2010; **35**(62):1149–1183.

D. A. Di Pietro, M. Vohralík, and C. Widmer. An a posteriori error estimator for a finite volume discretization of the two-phase flow. In *Finite volumes for complex applications. VI. Problems & perspectives. Volume 1, 2*, vol. 4 of *Springer Proc. Math.*, pp. 341–349. Springer, Heidelberg 2011.

D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*, **69**, Springer, 2012.

D. A. Di Pietro and M. Vohralík. A review of recent advances in discretization methods, *a posteriori* error analysis, and adaptive algorithms for numerical modeling in geosciences. *Oil and Gas Science and Technology* 2014; **69**(4):701–730.

R. J. DiPerna. Measure-valued solutions to conservation laws. *Arch. Rational Mech. Anal.* 1985; **88**(3):223–270.

K. Domelevo and P. Omnes. A finite volume method for the Laplace equation on almost arbitrary two-dimensional grids. *M2AN Math. Model. Numer. Anal.* 2005; **39**(6):1203–1249.

K. Dorogan and J.-Marc. Hérard. A two-dimensional relaxation scheme for the hybrid modelling of gas-particle two-phase flows. *Int. J. Finite Vol.*, 8:30, 2012.

O. Drblíková, A. Handlovičová, and K. Mikula. Error estimates of the finite volume scheme for the nonlinear tensor-driven anisotropic diffusion. *Appl. Numer. Math.* 2009; **59**(10):2548–2570.

O. Drblíková and K. Mikula. Convergence analysis of finite volume scheme for nonlinear tensor anisotropic diffusion in image processing. *SIAM J. on Numer. Anal.* 2007; **46**(1):37–60.

J. Droniou. Non-coercive linear elliptic problems. *Potential Analysis* 2002; **17**(2):181–203.

J. Droniou. Error estimates for the convergence of a finite volume discretization of convection-diffusion equations. *J. Numer. Math.* 2003; **11**(1):1–32.

J. Droniou. Finite volume schemes for diffusion equations: introduction to and review of modern methods. *Math. Models Methods Appl. Sci.* 2014; **24**(8):1575–1619.

J. Droniou and R. Eymard. A mixed finite volume scheme for anisotropic diffusion problems on any grid. *Numer. Math.* 2006; **105**(1):35–71.

J. Droniou, R. Eymard, T. Gallouët, C. Guichard, and R. Herbin. *Gradient Schemes for elliptic and parabolic problems.* in preparation.

J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. A unified approach to mimetic finite difference, hybrid finite volume and mixed finite volume methods. *Math. Models Methods Appl. Sci.* 2010; **20**(2):265–295.

J. Droniou, R. Eymard, T. Gallouët, and R. Herbin. Gradient schemes: a generic framework for the discretisation of linear, nonlinear and nonlocal elliptic and parabolic equations. *Math. Models Methods Appl. Sci.* 2013; **23**(13):2395–2432.

J. Droniou, R. Eymard, and R. Herbin. Gradient schemes: generic tools for the numerical analysis of diffusion equations. *M2AN Math. Model. Numer. Anal.*, to appear.

J. Droniou and T. Gallouët. Finite volume methods for convection-diffusion equations with right-hand side in $H^{-1}$. *M2AN Math. Model. Numer. Anal.* 2002; **36**(4):705–724.

J. Droniou, T. Gallouët, and R. Herbin. A finite volume scheme for a noncoercive elliptic equation with measure data. *SIAM J. Numer. Anal.* 2003; **41**(6):1997–2031.

J. Droniou and C. Le Potier. Construction and convergence study of schemes preserving the elliptic local maximum principle. *SIAM J. Numer. Anal.* 2011; **49**(2):459–490.

J. Droniou and N. Nataraj in preparation. Improved $L^2$ estimate for gradient schemes, and super-convergence of finite volume methods. In preparation.

M. G. Edwards and C. F. Rogers 1994. A flux continuous scheme for the full tensor pressure equation. In *4th European Conference on the Mathematics of Oil Recovery* 1994.

M. G. Edwards and C. F. Rogers. Finite volume discretization with imposed flux continuity for the general tensor pressure equation. *Computational Geosciences* 1998; **2**(4):259–290.

B. Einfeldt, C. Munz, P. L. Roe, and B. Sjögreen. On Godunov-type methods near low densities. *J. Comp. Phys.* 1992; **92**:273–295.

C. Erath. A nonconforming a posteriori estimator for the coupling of cell-centered finite volume and boundary element methods. *Numer. Math.* 2015; **131**(3):425–451.

A. Ern and M. Vohralík. A unified framework for a posteriori error estimation in elliptic and parabolic problems with application to finite volumes. In *Finite volumes for complex applications. VI. Problems & perspectives. Volume 1, 2*, vol. 4 of *Springer Proc. Math.*, pp. 821–837. Springer, Heidelberg 2011.

L. C. Evans. *Partial differential equations*, vol. 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second ed. 2010.

S. Evje and K. H. Karlsen. An error estimate for viscous approximate solutions of degenerate parabolic equations. *J. Nonlin. Math. Phys.* 2002; **9**(3):262–281.

R. E. Ewing, T. Lin, and Y. Lin. On the accuracy of the finite volume element method based on piecwise linear polynomials. *SIAM J. Numer. Anal.* 2002; **39**(6):1865–1888.

R. Eymard, T. Gallouët, J. Vovelle. Limit boundary conditions for finite volume approximations of some physical problems. *J. Comput. Appl. Math.* 2003; **161**(2):349–369.

R. Eymard, T. Gallouët, M. Ghilani, and R. Herbin. Error estimates for the approximate solution of a nonlinear hyperbolic equation given by finite volume schemes. *IMA J. of Numer. Anal.* 1998; **18**:563–594.

R. Eymard, T. Gallouët, R. Herbin, and A. Michel. Convergence of a finite volume scheme for nonlinear degenerate parabolic equations. *Numer. Math.* 2002; **92**(1):41–82.

R. Eymard, T. Gallouët and R. Herbin. Finite volume methods. In P. G. Ciarlet & J.-L. Lions (eds.), *Techniques of Scientific Computing, Part III*, Handbook of Numerical Analysis, VII, North-Holland, Amsterdam 2000, pp. 713–1020.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

R. Eymard, T. Gallouët and R. Herbin. Existence and uniqueness of the entropy solution to a nonlinear hyperbolic equation. In *Chinese Ann. Math. Ser. B*,**16**(1):1–14.

R. Eymard, T. Gallouët, and R. Herbin. Finite volume approximation of elliptic problems and convergence of an approximate gradient. *Appl. Numer. Math.* 2001; **37**(1–2):31–53.

R. Eymard, T. Gallouët, and R. Herbin. Error estimates for approximate solutions of a nonlinear convection-diffusion problem. *Adv. Diff. Equations* 2002; **7**(4):419–440.

R. Eymard, T. Gallouët, and R. Herbin. Convergence of finite volume schemes for semilinear convection diffusion equations. *Numer. Math.*, 82(1):91–116, 1999.

R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes SUSHI: a scheme using stabilization and hybrid interfaces. *IMA J. Numer. Anal.* 2010a; **30**(4):1009–1043.

R. Eymard, T. Gallouët, R. Herbin, and J.-C. Latché. Convergence of the MAC scheme for the compressible Stokes equations. *SIAM J. Numer. Anal.* 2010b; **48**(6):2218–2246.

R. Eymard, G. Henry, R. Herbin, F. Hubert, R. Klofkorn, and G. Manzini 2011. 3D Benchmark on Discretization Schemes for Anisotropic Diffusion Problems on General Grids. In *Proceedings of Finite Volumes for Complex Applications VI*, Praha. Springer, Springer 2011, pp. 895–930.

R. Eymard, C. Guichard, and R. Herbin. Small-stencil 3D schemes for diffusive flows in porous media. *M2AN Math. Model. Numer. Anal.* 2012; **46**:265–290.

R. Eymard, P. Féron, and C. Guichard 2015. Gradient schemes for the incompressible steady Navier-Stokes problem. In *6th International conference on Approximation Methods and Numerical Modelling in Environment and Natural Resources*. Université de Pau 2015.

R. Eymard, T. Gallouët, and R. Herbin 2016. *Finite volume methods: schemes and analysis*. in preparation.

I. Faille. A control volume method to solve an elliptic equation on a two-dimensional irregular mesh. *Comput. Methods Appl. Mech. Engrg.* 1992a; **100**(2):275–290.

I. Faille 1992b. *Modélisation bidimensionnelle de la genèse et de la migration des hydrocarbures dans un bassin sédimentaire*. Ph.D. thesis, University Joseph Fourier, Grenoble I, 1992b.

M. Feistauer, J. Felcman, M. Lukáčová-Medvid'ová, and G. Warnecke. Error estimates for a combined finite volume-finite element method for nonlinear convection-diffusion problems. *SIAM J. Numer. Anal.* 1999; **36**(5):1528–1548.

A. Fettah and T. Gallouët. Numerical approximation of the general compressible Stokes problem. *IMA J. Numer. Anal.* 2013; **33**(3):922–951.

F. Filbet. Convergence of a finite volume scheme for the Vlasov-Poisson system. *SIAM J. Numer. Anal.* 2001; **39**(4):1146–1169.

F. Filbet. A finite volume scheme for the Patlak-Keller-Segel chemotaxis model. *Numer. Math.* 2006; **104**(4):457–488.

U. Fjordholm, S. Mishra, and E. Tadmor. Arbitrarily high-order accurate entropy stable essentially nonoscillatory schemes for systems of conservation laws. *SIAM J. Num. Anal.* 2012; **50**:544–573.

P. Frolkovič and K. Mikula. Flux-based level set method: a finite volume method for evolving interfaces. *Applied numerical mathematics* 2007; **57**(4):436–454

T. Gallouët 2007. Nonlinear methods for linear equations. In T. Aliziane, K. Lemrabet, A. Mokrane, & D. Teniou (eds.), *Actes du 3eme colloque sur les Tendandes des Applications Mathématiques en Tunise, Algérie, Maroc (14-18 avril 2007)*, AMNEDP-USTHB 2007, pp. 17–22.

T. Gallouët, Ph. Helluy, J.-M. Hérard, and Julien Nussbaum. Hyperbolic relaxation models for granular flows. *M2AN Math. Model. Numer. Anal.* 2010; **44**(2):371–400.

T. Gallouët, L. Gastaldo, R. Herbin, and J.-C. Latché. An unconditionally stable pressure correction scheme for the compressible barotropic Navier-Stokes equations. *M2AN Math. Model. Numer. Anal.* 2008; **42**(2):303–331.

T. Gallouët, J.-M. Hérard, and N. Seguin. Some approximate Godunov schemes to compute shallow-water equations with topography. *Comput. & Fluids*, 32(4):479–513, 2003.

T. Gallouët, J.-M. Hérard, and N. Seguin. Some recent finite volume schemes to compute Euler equations using real gas EOS. *Internat. J. Numer. Methods Fluids* 2002; **39**(12):1073–1138.

T. Gallouët and R. Herbin. Finite volume approximation of elliptic problems with irregular data. In *Finite volumes for complex applications II*, Hermes Sci. Publ., Paris 1999, pp. 155–162.

T. Gallouët, R. Herbin, J.-C. Latché, and K. Mallem. Convergence of the Marker-and-Cell scheme for the incompressible Navier-Stokes equations. submitted.

T. Gallouët, R. Herbin, J.-C. Latché, and D. Maltese. Convergence of the Marker-and-Cell scheme for the compressible stationary Navier-Stokes equations. submitted.

T. Gallouët, R. Herbin, and J.-C. Latché. Kinetic energy control in explicit finite volume discretizations of the incompressible and compressible Navier-Stokes equations. *Int. J. Finite Vol.* 2010; **7**(2):6.

T. Gallouët, R. Herbin, and M. H. Vignal. Error estimates on the approximate finite volume solution of convection diffusion equations with general boundary conditions. *SIAM J. Numer. Anal.* 2000; **37**(6):1935–1972.

T. Gallouët, R. Herbin, D. Maltese, and A. Novotny. Convergence of the Marker-and-Cell scheme for a semi-stationary compressible Navier-Stokes problem. under revision.

T. Gallouët, R. Herbin, D. Maltese, and A. Novotny. Error estimates for a numerical approximation to the compressible barotropic NavierStokes equations. *IMA Journal of Numerical Analysis*, 2015.

T. Gallouët, A. Larcher, and J. Latché. Convergence of a finite volume scheme for the convection-diffusion equation with $L^1$ data. *Mathematics of Computation* 2012; **81**(279):1429–1454.

T. Gallouët and J.-M. Masella. Un schéma de Godunov approché. *C. R. Acad. Sci. Paris Sér. I Math.* 1996; **323**(1):77–84.

L. Gastaldo, R. Herbin, W. Kheriji, C. Lapuerta, and J.-C. Latché. Staggered discretizations, pressure correction schemes and all speed barotropic flows. In *Finite volumes for complex*

*applications. VI. Problems & perspectives. Volume 1, 2*, vol. 4 of *Springer Proc. Math.*, Springer, Heidelberg 2011, pp. 839–855.

A. Genty and C. Le Potier. Maximum and minimum principles for radionuclide transport calculations in geological radioactive waste repository: comparison between a mixed hybrid finite element method and finite volume element discretizations. *Transp. Porous Media* 2011; **88**(1):65–85.

E. Godlewski and P.-A. Raviart. Hyperbolic Systems of Conservation Laws. *Mathematiques & Applications*, Ellipses, Paris, 1991.

S. K. Godunov. A finite difference method for the numerical computation of discontinuous solutions of the equations of fluid dynamics. *Mat. Sb.* 1959; **47**:271–290.

S. K. Godunov. An interesting class of quasilinear systems. *Dokl. Akad. Nauk. SSSR* 1961; **139**:521–523.

J. D. Goodman and R. J. L. Veque. On the accuracy of stable schemes for 2D conservation laws. *Math. Comp.* 1985; **45**(171):15–21.

S. Gottlieb and C.-W. Shu. Total variation diminishing Runge-Kutta schemes. *Math. Comput.* 1998; **67**(221):73–85.

L. Gosse. A well-balanced scheme able to cope with hydrodynamic limits for linear kinetic models. *Appl. Math. Lett.* 2015; **42**(3):15–21.

L. Gosse. Un schéma-équilibre adapté aux lois de conservation scalaires non-homogènes. *C. R. Acad. Sci. Paris Sér. I Math.* 1996; **323**(5):543–546.

S. Gottlieb, C. W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Rev.* 2001; **43**(1):89–112.

I. Greff. Nonconforming box-schemes for elliptic problems on rectangular grids. *SIAM J. Numer. Anal.* 2007; **45**(3):946–968.

J. M. Greenberg and A. Y. Leroux. A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. *SIAM J. Numer. Anal.* 1996; **33**(1):1–16.

F. Harlow and J. Welch. Numerical calculation of time-dependent viscous incompressible flow of fluid with a free surface. *Physics of Fluids* 1965; **8**:2182–2189.

A. Harten. High resolution schemes for hyperbolic conservation laws. *J. Comp. Phys.* 1983a; **49**:151–164.

A. Harten. On the symmetric form of systems of conservation laws with entropy. *J. Comp. Phys.* 1983b; **49**:357–393.

A. Harten. ENO schemes with subcell resolution. *J. Comp. Phys.* 1989; **83**:148–184.

A. Harten, S. Osher, B. Engquist, and S. Chakravarthy. Some results on uniformly high order accurate essentially non-oscillatory schemes. *Appl. Num. Math.* 1986; **2**:347–377.

A. Harten, S. Osher, B. Engquist, and S. Chakravarthy. Uniformly high-order accurate essentially nonoscillatory schemes III. *J. Comput. Phys.* 1987; **71**(2):231–303.

A. Harten and S. Chakravarthy. Multi-dimensional ENO schemes for general geometries. Report ICASE-91-76, Institute for Computer Applications in Science and Engineering (ICASE), 1991.

A. Harten, J. M. Hyman, and P. D. Lax. On finite-difference approximations and entropy conditions for shocks. *Comm. Pure and Appl. Math.* 1976; **29**:297–322.

A. Harten, P. D. Lax, and B. van Leer. On upstream differencing and Godunov-type schemes for hyperbolic conservation laws. *SIAM Rev.* 1983; **25**:35–61.

P. Helluy, J.-M. Hérard, H. Mathis, and S. Müller. A simple parameter-free entropy correction for approximate Riemann solvers. *C. R. Méc. Acad. Sci. Paris* 2010; **338**(9):493–498.

R. Herbin. An error estimate for a finite volume scheme for a diffusion-convection problem on a triangular mesh. *Numer. Methods Partial Differential Equations* 1995; **11**(2):165–173.

R. Herbin and M. Ohlberger. A posteriori error estimate for finite volume approximations of convection diffusion problems. *proceedings: Finite volumes for complex applications–problems and perspectives, Porquerolles*, Hermes Science Publications: Paris, 2002; 753-760.

R. Herbin and F. Hubert 2008. Benchmark on Discretization Schemes for Anisotropic Diffusion Problems on General Grids for anisotropic heterogeneous diffusion problems. In R. Eymard & J.-M. Hérard (eds.), *Finite Volumes for Complex Applications V*, Wiley, 2008, pp. 659–692.

R. Herbin and J.-C. Latché. Kinetic energy control in the MAC discretization of the compressible Navier-Stokes equations. *Int. J. Finite Vol.* 2010; **7**(2):6.

R. Herbin, J.-C. Latché, and T. T. Nguyen. Explicit staggered schemes for the compressible Euler equations. In *Applied mathematics in Savoie—AMIS 2012: Multiphase flow in industrial and environmental engineering*, vol. 40 of *ESAIM Proc.*, EDP Sci., Les Ulis 2013, pp. 83–102.

R. Herbin, J.-C. Latché, and C. Zaza. A cell-centered pressure-correction scheme for the compressible Euler equations. *M3AS*, under revision.

R. Herbin, J.-C. Latché and T. Nguyen. On some consistent explicit staggered schemes for the shallow water and Euler equations. *ESAIM: Mathematical Modelling and Numerical Analysis*, under revision.

R. Herbin, J.-C. Latché, and K. Mallem. Convergence of the MAC Scheme for the steady-state incompressible Navier-Stokes equations on non-uniform grids. In J. Fuhrmann, M. Ohlberger, & C. Rohde (eds.), *Finite Volumes for Complex Applications VII-Methods and Theoretical Aspects*, vol. 77 of *Springer Proceedings in Mathematics & Statistics*, Springer International Publishing 2014, pp. 343–351.

F. Hermeline. A finite volume method for the approximation of diffusion operators on distorted meshes. *J. Comput. Phys.* 2000; **160**(2):481–499.

F. Hermeline. Approximation of diffusion operators with discontinuous tensor coefficients on distorted meshes. *Computer methods in applied mechanics and engineering* 2003; **192**(16):1939–1959.

F. Hermeline. Approximation of 2-D and 3-D diffusion operators with variable full tensor coefficients on arbitrary meshes. *Comput. Meth. Appl. Mech. Engin.* 2007; **196**(21):2497–2526.

F. Hermeline. A finite volume method for approximating 3D diffusion operators on general meshes. *J. Comp. Phys.* 2009; **228**(16):5763–5786.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

A. Jameson and P. D. Lax. Conditions for the construction of multipoint variation diminishing difference schemes. *Appl. Numer. Math.* 1986; **2**(3-5):235–345.

A. Jameson and P. D. Lax. Corrigendum: conditions for the construction of multipoint variation diminishing difference schemes. *Appl. Numer. Math.* 1987; **3**(3):289.

A. Jameson. *Artificial diffusion, upwind biasing, limiters and their effect on accuracy and convergence in transonic and hypersonic flows*. Report AIAA-93-3359, American Institute for Aeronautics and Astronautics, 1993.

G. Jiang and C. W. Shu. Efficient implementation of weighted ENO schemes. *J. Comp. Phys.* 1996; **126**:202–228.

Shi Jin and Zhou Ping Xin. The relaxation schemes for systems of conservation laws in arbitrary space dimensions. *Comm. Pure Appl. Math.* 1995; **48**(3):235–276.

M. Küther. Error estimates for second order finite volume schemes using a TVD-Runge-Kutta time discretization for a nonlinear scalar hyperbolic conservation law. *East-West J. Numer. Math.* 2000; **8**(4):299–322.

K. H. Karlsen and N. H. Risebro. On the uniqueness and stability of entropy solutions of nonlinear degenerate parabolic equations with rough coefficients. Preprint 143, Department of Mathematics, University of Bergen, 2000.

K. H. Karlsen, C. Klingenberg, and N. H. Risebro. Relaxation schemes for conservation laws with discontinuous coefficients. In *Hyperbolic problems: theory, numerics, applications*, Springer, Berlin, 2003, pp. 611–620.

B. Koren. Upwind schemes for the Navier-Stokes equations. In *Proceedings of the Second International Conference on Hyperbolic Problems*, Vieweg, Braunschweig, 1988.

D. Kröner. *Numerical Schemes for Conservation Laws*. Wiley–Teubner, Stuttgart, 1997.

D. Kröner and M. Ohlberger. *A posteriori* error estimates for upwind finite volume schemes for nonlinear conservation laws in multidimensions. *Math. Comput.* 2000; **69**:25–39.

D. Kröner, S. Noelle, and M. Rokyta. Convergence of higher order upwind finite volume schemes on unstructured grids for conservation laws in several space dimensions. *Numer. Math.* 1995; **71**:527–560.

S. N. Kruzkov. First order quasilinear equations in several independent variables. *Math. USSR Sbornik* 1970; **10**:217–243.

P. Kútik and K. Mikula. Diamond-cell finite volume scheme for the Heston model. *Discrete Contin. Dyn. Syst. Ser. S* 2015; **8**(5):913–931.

N. N. Kuznetsov. Accuracy of some approximate methods for computing the weak solutions of a first-order quasi-linear equation. *USSR, Comput. Math. and Math. Phys.* 1976; **16**(6):159–193.

P. D. Lax and B. Wendroff. Systems of conservation laws. *Comm. Pure Appl. Math.* 1960; **13**:217–237.

P. D. Lax. *Hyperbolic Systems of Conservation Laws and the Mathematical Theory of Shock Waves*. SIAM, Philadelphia, Penn., 1973.

R.D. Lazarov, I.D. Michev, and P.S. Vassilevsky. Finite volume methods for convection-diffusion problems. *SIAM J. Numer. Anal.* 1996; **33**:31–35.

*Encyclopedia of Computational Mechanics*. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

R. Lazarov and S. Tomov. A posteriori error estimates for finite volume element approximations of convection-diffusion-reaction equations. *Comput. Geosci.* 2002; **6**(3-4):483–503.

C. Le Potier. Schéma volumes finis monotone pour des opérateurs de diffusion fortement anisotropes sur des maillages de triangles non structurés. *Comptes Rendus Mathématique* 2005; **341**(12):787–792.

C. Le Potier. Finite volume scheme satisfying maximum and minimum principles for anisotropic diffusion operators. In *Finite volumes for complex applications V*, ISTE, London 2008, pp. 103–118.

C. Le Potier. A nonlinear finite volume scheme satisfying maximum and minimum principles for diffusion operators. *Int. J. Finite Vol.* 2009; **6**(2).

R. LeVeque. *Finite Volume Methods for Hyperbolic Problems.* Cambridge University Press 2002.

R. . J. LeVeque and M. Pelanti. A class of approximate Riemann solvers and their relation to relaxation schemes. *J. Comput. Phys.* 2001; **172**(2):572–591.

M. S. Liou and C. J. Steffen. A new flux-splitting scheme. *J. Comput. Phys.* 1993; **107**:23–39.

K. Lipnikov, G. Manzini, and M. Shaskov. Mimetic finite difference method. *J. Comput. Phys.* 2014; **257**(part B):1163–1227.

X.-D. Liu. A maximum principle satisfying modification of triangle based adaptive stencils for the solution of scalar hyperbolic conservation laws. *SIAM J. Numer. Anal.* 1993; **30**:701–716.

J. Málek, J. Nečas, M. Rokyta, and M. Røužička. Weak and measure-valued solutions to evolutionary PDEs. *Applied Mathematics and Mathematical Computation*, vol. 13. Chapman and Hall: London, 1968; 44-177.

S. Martin and J. Vovelle. Convergence of implicit finite volume methods for scalar conservation laws with discontinuous flux function. *M2AN Math. Model. Numer. Anal.* 2008; **42**(5):699–727

J.-M. Masella, I. Faille, and T. Gallouët. On an approximate Godunov scheme. *Int. J. Comput. Fluid Dyn.* 1999; **12**(2):133–149.

B. Merlet $L^\infty$- and $L^2$-Error estimates for a finite volume approximation of linear advection *SIAM J. Numer. Anal.* 2007; **46**(1):124–150.

A. Michel, Q.H. Tran and G. Favennec. A genuinely one-dimensional upwind Scheme with accuracy enhancement for multidimensional advection problems. *ECMOR XII 12 th European Conference on the Mathematics of Oil Recovery*, 2010.

K. Mikula and N. Ramarosy. Semi-implicit finite volume scheme for solving nonlinear diffusion equations in image processing. *Numerische Mathematik* 2001; **89**(3):561–590.

M. S. Mock. Systems of conservation laws of mixed type. *J. Diff. Eqns.* 1980; **37**:70–88.

R. Nicolaïdes. Analysis and convergence of the MAC scheme I: The linear problem. *SIAM J. Numer. Anal.* 1992; **29**:1579–1591.

R. Nicolaïdes and X. Wu. Analysis and convergence of the MAC scheme II, Navier-Stokes equations. *Math. Comp.* 1996; **65**:29–44.

M. Ohlberger. *A posteriori* error estimates for finite volume approximations to singularly perturbed nonlinear convection-diffusion equations. *Numer. Math.* 2001a; **87**(4):737–761.

M. Ohlberger. *A posteriori* error estimates for vertex centered finite volume approximations of convection-diffusion-reaction equations. *M2AN Math. Modell. Numer. Anal.* 2001b; **35**(2):355–387.

M. Ohlberger and C. Rohde. Adaptive finite volume approximations for weakly coupled convection dominated parabolic systems. *IMA J. Numer. Anal.* 2002; **22**(2):253–280.

M. Ohlberger and J. Vovelle. Error estimate for the approximation of nonlinear conservation laws on bounded domains by the finite volume method. *Math. Comp.* 2006; **75**(253):113–150 (electronic).

M. Ohlberger. A review of a posteriori error control and adaptivity for approximations of non-linear conservation laws. *Internat. J. Numer. Methods Fluids* 2009; **59**(3):333–354.

O. A. Oleinik. Discontinuous solutions of non-linear differential equations. *Amer. Math. Soc. Transl.(2)* 1963; **26**:95–172.

P. Omnes. An *a posteriori* error bound for the discrete duality finite volume discretization of the Laplace equation and application to the adaptive simulation in a domain with a crack. newblock In *Finite volumes for complex applications V* ISTE, London 2008, pp. 617–624.

P. Omnes. On the second-order convergence of a function reconstructed from finite volume approximations of the Laplace equation on Delaunay-Voronoi meshes. *ESAIM Math. Model. Numer. Anal.* 2011; **45**(4):627–650.

S. Osher and F. Solomon. Upwind difference schemes for hyperbolic systems of conservation laws. *Math. Comput.* 1982; **38**(158):339–374.

S. Osher. Riemann solvers, the entropy condition, and difference approximations. *SIAM J. Numer. Anal.* 1984; **21**(2):217–235.

S. Patankar. *Numerical heat transfer and fluid flow. Series in Computational Methods in Mechanics and Thermal Sciences*, vol. XIII. Washington - New York - London: Hemisphere Publishing Corporation; New York. McGraw-Hill Book Company 1980.

M. Pelanti, F. Bouchut, and A. Mangeney. A Riemann solver for single-phase and two-phase shallow flow models based on relaxation. Relations with Roe and VFRoe solvers. *J. Comput. Phys.*, 230(3):515–550, 2011.

B. Perthame, N. Seguin, Nicolas and M. Tournus. A simple derivation of BV bounds for inhomogeneous relaxation systems. *Commun. Math. Sci.*, 13(2):577–586, 2015.

L. Piar, F. Babik, R. Herbin, and J.-C. Latché. A formally second order cell centered scheme for convection-diffusion equations on general grids. *International Journal for Numerical Methods in Fluids* 2013; **71**:873–890.

T. A. Porsching. Error estimates for MAC-like approximations to the linear Navier-Stokes equations. *Numer. Math.* 1978; **29**(3):291–306.

P. L. Roe. Approximate Riemann solvers, parameter vectors, and difference schemes. *J. Comput. Phys.* 1981; **43**:357–372.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

P. Rostand and B. Stoufflet. TVD schemes to compute compressible viscous flows on unstructured meshes. *Proceedings of the Second International Conference on Hyperbolic Problems*. Vieweg: Braunschweig, 1988.

E. Süli. Convergence of finite volume schemes for Poisson's equation on nonuniform meshes. *SIAM J. Numer. Anal.* 1991; **28**:1419–1430.

C. W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Comput. Phys.* 1988; **77**:439–471.

C. W. Shu. Total-variation-diminishing time discretizations. *SIAM J. Sci. Statist. Comput.* 1988; **9**:1073–1084.

C. W. Shu. *High Order ENO and WENO Schemes*, *Lecture Notes in Computational Science and Engineering*, vol. 9. Springer-Verlag: Heidelberg, 1999.

C. W. Shu. A survey of strong stability preserving high order time discretizations. *Collected lectures on the preservation of stability under discretization* (Fort Collins, CO, 2001), SIAM: Philadelphia, 2002; 51-65.

T. Sonar. On the construction of essentially non-oscillatory finite volume approximations to hyperbolic conservation laws on general triangulations: polynomial recovery, accuracy, and stencil selection. *Comput. Methods Appl. Mech. Eng.* 1997; **140**:157–181.

T. Sonar. On families of pointwise optimal finite volume ENO approximations. *SIAM J. Numer. Anal.* 1998; **35**(6):2350–2379.

S. P. Spekreijse. Multigrid solution of monotone second-order discretizations of hyperbolic conservation laws. *Math. Comput.* 1987; **49**:135–155.

J. L. Steger and R. F. Warming. Flux vector splitting of the inviscid gasdynamic equations with application to finite difference methods. J. Comput. Phys. 1981; **40**:263–293.

P. K. Sweby. High resolution schemes using flux limiters for hyperbolic conservation laws. *SIAM J. Numer. Anal.* 1984; **21**(5):995–1011.

A. Szepessy. An existence result for scalar conservation laws using measure valued solutions. *Commun. Partial Diff. Equations* 1989; **14**:1329–1350.

E. Tadmor. Entropy functions for symmetric systems of conservation laws. *J. Math. Anal. Appl.* 1987; **122**:355–359.

E. Tadmor. Local error estimates for discontinuous solutions of nonlinear hyperbolic equations. *SIAM J. Numer. Anal.* 1991; **28**:891–906.

E. Tadmor. Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems. *Acta Numerica* 2004; **13**:451–512.

L. Tartar. Compensated compactness and applications to partial differential equations, Nonlinear analysis and mechanics: Heriot-Watt Symposium, Vol. IV, Res. Notes in Math.. Pitman: Boston, Mass.-London, 1979.

S. Tatsumi, L. Martinelli, and A. Jameson. *Design, Implementation, and Validation of Flux Limited Schemes for the Solution of the Compressible Navier-Stokes Equations*. Report AIAA-94-0647, American Institute for Aeronautics and Astronautics, 1994.

N. Therme 2015. *Schémas numériques pour la simulation de l'explosion*. Ph.D. thesis, Aix-Marseille Université.

*Encyclopedia of Computational Mechanics*. Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.

Q. Tran. Second-order slope limiters for the simultaneous linear advection of (not so) independent variables. *Commun. Math. Sci.* 2008; **6**(3):569–593.

B. van Leer. Towards the ultimate conservative difference schemes V. A second order sequel to Godunov's method. *J. Comput. Phys.* 1979; **32**:101–136.

B. van Leer. *Flux-Vector Splitting for the Euler Equations.* Report ICASE-82-30, Institute for Computer Applications in Science and Engineering, NASA Langley, 1982.

B. van Leer. *Upwind-Difference Schemes for Aerodynamics Problems Governed by the Euler Equations*, *Lectures in Applied Mathematics 22*, AMS: Providence, Rhode Island, 1985.

P. Vankeirsblick. *Algorithmic Developments for the Solution of Hyperbolic Conservation Laws on Adaptive Unstructured Grids.* PhD thesis, Katholieke Universiteit Leuven, Belgium, 1993.

J. P. Vila. Convergence and error estimates in finite volume schemes for general multi-dimensional scalar conservation laws I: Explicit monotone schemes. *RAIRO, Model. Math. Anal. Numer.* 1994; **28**:267-295.

C. Viozat, C. Held, K. Mer, and A. Dervieux. *On Vertex-Center Unstructured Finite-Volume Methods for Stretched Anisotropic Triangulations.* Report 3464, Institut National De Recherche En Informatique Et En Automatique (INRIA), 1998.

J. Vovelle. Convergence of finite volume monotone schemes for scalar conservation laws on bounded domains. *Numer. Math.* 2002; **90**(3):563–596.

P. Wesseling. *Principles of Computational Fluid Dynamics.* Springer 2001.

M. Wierse. *Higher Order Upwind Scheme on Unstructured Grids for the Compressible Euler Equations in Time Dependent Geometries in 3D.* PhD thesis, University of Freiburg, Germany, 1994.

*Encyclopedia of Computational Mechanics.* Edited by Erwin Stein, René de Borst and Thomas J.R. Hughes.